Mobilidade num ambiente de serviços diferenciados *

Saulo V. de Vasconcellos e José F. de Rezende
Grupo de Teleinformática e Automação (GTA)
COPPE/Programa de Engenharia Elétrica
Universidade Federal do Rio de Janeiro
{vaz,rezende}@gta.ufrj.br
http://www.gta.ufrj.br

Resumo

As comunicações móveis vem apresentando um grande crescimento no mercado de telecomunicações. Usuários começam a buscar serviços de dados multimídia e acesso a Internet, demandando garantias de qualidade. Este artigo faz então um estudo de qualidade de serviço para redes móveis. Um cenário típico de interconexão de redes móveis com a infra-estrutura global IP é mostrado, assim como suas principais dificuldades. São sugeridas classes de serviço que se adequam às características de mobilidade e um mapeamento delas no modelo de diferenciação de serviços do IETF é realizado. Mecanismos são apresentados para dar suporte a uma das classes sugeridas e uma avaliação de desempenho destes é realizada.

Palavras chave: comunicação sem fio, qualidade de serviço (QoS), differenciação de servicos, Internet.

Abstract

Nowadays, mobile communications are an important slice of the telecommunications market. Users are looking for multimedia services and Internet access, needing service quality. This paper is a study of quality of service guarantees to mobile environments. A typical scenario, for internetworking mobile and fixed networks is presented, as well as its main challenges. Service classes concerned with mobility are suggested and a mapping from these classes to the IETF's quality of service model is carried through. Mechanisms intended to support the proposed classes are presented and simulated.

Keywords: wireless communication, quality of service (QoS), service differentiation, Internet.

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

1 Introdução

O setor de comunicações móveis possui atualmente uma grande penetração no mercado de telecomunicações. Juntamente com a Internet, este é o setor de mercado que vem apresentando o maior crescimento [1]. Os custos dos serviços vem baixando e usuários já chegam, em alguns casos, a optar pela comodidade da conectividade oferecida nos serviços móveis. Com o surgimento de redes de pacotes sem fio, os custos tendem a diminuir ainda mais, uma vez que os recursos disponíveis passam a ser melhor aproveitados. Para exemplificar o melhor aproveitamento do meio nas redes orientadas a pacotes, observa-se que a telefonia através de comutação de circuitos chega a ocupar canais de até 64 kbps, contra os 6-8 kbps necessários na telefonia sobre pacotes [2]. Assim, com a diminuição dos custos, a popularização do serviço só tende a aumentar ainda mais.

Outro fator que faz crescer a popularidade das comunicações móveis por pacotes é a possibilidade de fusão entre comunicação e computação. A disponibilização no mercado de PDA's (*Personal Digital Assistants*) e celulares "inteligentes" a preços cada vez mais acessíveis, traz serviços atraentes pela extrema flexibilidade. A comodidade de acessar serviços de dados e multimídia em terminais móveis tendo a Internet como infra-estrutura exige, entretanto, todo o planejamento de esquemas de suporte adequados. As características do tráfego multimídia e do tráfego de dados devem ser levados em consideração, assim como as particularidades das redes sem fio. O tráfego de voz, por exemplo, exige baixas latências e *jitter*, podendo sofrer perdas, enquanto que o tráfego de dados necessita de confiabilidade. As redes sem fio acrescentam novos fatores como limitação de vazão e altas taxas de perdas [3], além de problemas inerentes a mobilidade dos nós. Os modelos atualmente utilizados para diferenciação de serviços nem sempre irão se adequar às características desse tipo de tráfego.

Alguns trabalhos anteriores abordam a questão de mobilidade em ambientes de diferenciação de serviços [4, 5]. O primeiro analisa aspectos da mobilidade que influenciariam o ambiente de diferenciação de serviços e o segundo sugere uma arquitetura que traz como ponto fundamental a utilização de mecanismos como controle de admissão e sinalização explícita. Nenhum deles buscou aplicar a arquitetura de diferenciação de serviços na Internet tal como ela é. O enfoque principal deste trabalho é então a proposta e a avaliação de desempenho de mecanismos de suporte a uma classe de serviço voltada para a mobilidade num ambiente de serviços diferenciados. Além disso, este artigo tem como objetivo estudar um cenário típico de interconexão de redes móveis com a infra-estrutura global IP, apresentar classes de serviço que melhor se adequam às características de mobilidade e discutir sobre o mapeamento dessas classes no modelo de diferenciação de serviço do IETF (*Internet Engineering Task Force*) [6].

O restante desse artigo encontra-se organizado da seguinte forma: A seção 2 traz um cenário típico de redes móveis conectadas à infra-estrutura IP. As dificuldades encontradas nesse cenário e causadas pela mobilidade são também apresentadas nessa seção. Em seguida, na seção 3, algumas classes de serviço que incluem características de mobilidade são definidas. As principais propostas de qualidade de serviço na Internet são então abordadas. Um mapeamento entre as classes e PHB's também é apresentado. Os mecanismos de suporte necessários para uma classe de serviço de interesse são sugeridos na seção 4. A seção seguinte apresenta simulações que permitem a avaliação dos mecanismos sugeridos. Por fim, as conclusões desse trabalho são apresentadas na seção 6.

2 Cenário de estudo

Existem algumas características importantes que devem estar presentes quando se buscam esquemas de transmissão de dados em redes de pacotes sem fio. A independência da tecnologia de rede é um exemplo. Mesmo num ambiente de redes de acesso sem fio, a interconexão com o restante da Internet deve ser feita de forma transparente. Assim, ao se buscar soluções, é interessante que se pense em termos de redes IP, uma vez que toda a Internet é baseada nessa tecnologia. Os esquemas de endereçamento, roteamento, qualidade de serviço, entre outros, são amplamente estudados e conhecidos nessas redes. Por esses motivos, existe uma tendência para que a terceira geração de sistemas de comunicações móveis venham a utilizar tecnologia IP fim-a-fim [7].

Um cenário típico para redes sem fio é apresentado na figura 1, onde um provedor possui uma rede sem fio cujas estações radio-base (ERB's) estão conectadas a um roteador de borda de uma infra-estrutura IP [8]. Assim, as redes sem fio funcionam como redes de acesso. Diferentes provedores de acesso podem se conectar a rede fixa. Os usuários de terminais móveis têm então a possibilidade de se mover dentro da rede sem fio de seu provedor assim como nos demais provedores de acesso, através de esquemas de *roaming*. As redes sem fio podem envolver diferentes tecnologias. Elas podem ser redes locais sem fio adequadas para uso dentro de prédios, redes celulares sem fio para regiões metropolitanas ou redes utilizando satélites para áreas mais amplas.

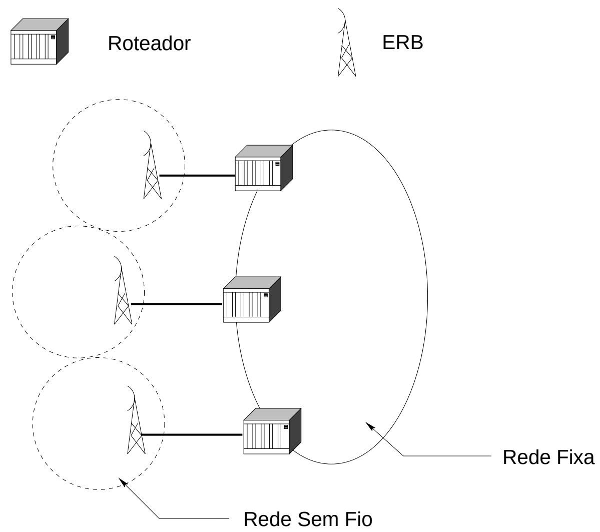


Figura 1: Arquitetura típica.

De acordo com a figura 1, observa-se a possibilidade de um nó, ao se mover, passar de um domínio de acesso para outro adjacente durante a transmissão. Essas mudanças são denominadas de *handovers*. Ao se passar de um domínio que utiliza uma determinada tecnologia de transmissão sem fio para domínios que utilizem outro tipo de tecnologia, como o que ocorre no caso da saída do interior de um prédio (ocorrendo a modificação da conexão de uma rede local sem fio para uma rede metropolitana sem fio), é dito que se realiza um *handover* vertical. No caso das tecnologias serem equivalentes, como no caso de uma rede celular, onde existem várias estações rádio base com mesma tecnologia de transmissão, os *handovers* são ditos horizontais [9]. Nos *handovers* verticais podem ocorrer mudanças abruptas em características do enlace devido às mudanças de tecnologia[1].

Os *handovers* podem causar problemas aos roteadores de borda. Conforme foi colocado inicialmente, as redes móveis são redes de acesso conectados através de um roteador de borda a uma infra-estrutura de rede fixa (figura 1). Já que os nós são móveis, podem ocorrer situações onde um grande número de nós se encaminhe para uma determinada rede de acesso, gerando uma sobrecarga em seu roteador de borda. Essa incerteza com relação ao volume de tráfego traz maior complexidade à tarefa de realizar um provisionamento que garanta a qualidade de serviço aos usuários.

3 Arquitetura de QoS para mobilidade

A terceira geração de sistemas móveis visa dar suporte a tráfego multimídia através de rede de pacotes [8]. Assim, torna-se necessário o estudo de esquemas de qualidade de serviço (QoS) para permitir o funcionamento adequado de tais aplicações. O bom funcionamento de propostas de QoS na Internet depende muito do conhecimento das características de tráfego, incluindo seu volume através dos nós da rede [12]. Através desses dados é possível que se realize um provisionamento adequado. No caso de clientes móveis essa previsão de tráfego fica muito prejudicada, por causa do fator dinâmico que é acrescentado com o movimento dos nós. Dessa forma é importante a existência de mecanismos que possam lidar com essa "dinamicidade", sem deixar de oferecer determinados graus de serviço aos usuários, mesmo em situações críticas. As características de tráfego para as diferentes aplicações podem então ser exploradas para que cada uma possa ser atendida da melhor forma. Nesse momento deve ser identificado o que é mais importante para cada tipo de aplicação: a confiabilidade, a disponibilidade/conectividade do serviço, *etc.*.

Os parâmetros de QoS normalmente abordados em redes fixas tradicionais são atraso, variação no atraso (*delay jitter*), vazão e taxa de perdas. Quando se fala em redes móveis, entretanto, novos fatores de QoS que advêm da mobilidade e de sua imprevisibilidade se tornam importantes. A garantia de serviços sem interrupções é uma delas. O fato de estar continuamente conectado pode ser mais importante do que a ausência total de perdas, ou até mesmo do que a presença de uma alta vazão. Isto é verdade para tráfegos de voz por exemplo. Um usuário pode preferir uma diminuição na qualidade da voz em vez de ter sua chamada interrompida de forma abrupta. Assim, deve se levar em consideração a degradação suave do serviço como uma opção de QoS. Neste contexto aparecem como parâmetros de QoS a probabilidade de conexão contínua e o perfil de perdas [11, 13]. A partir dos parâmetros de QoS apresentados e das características das aplicações foram especificadas classes de serviço voltadas para mobilidade, apresentadas na próxima seção.

3.1 Classes de serviço

Especialmente nos ambientes de comunicações móveis, devido à suas características limitantes, classes de serviço que permitissem certos graus de flexibilidade seriam interessantes. Tendo como ponto de partida problemas causados pela mobilidade e características particulares de cada aplicação, diferentes classes podem ser visualizadas. Os autores de [7]

[1]No caso de um nó móvel sair de um prédio para ambiente externo, ele poderia estar trocando a conexão com uma rede local sem fio como uma WaveLAN de 2 Mbps por uma rede metropolitana baseada em GPRS [10] de 160 kbps. Dessa forma, nota-se que as variações podem ser muito grandes, além de freqüentes [11].

Tabela 1: Classes de serviço.

Classe	Aplicação	Taxa de perdas	Atraso	Banda passante	*Jitter*	Probabilidade de desconexão
A	controle	nula	baixo	–	–	nula
B	multimídia em tempo real	baixa-média	baixo	média-alta	baixo	baixa
C	login remoto	baixa	baixo	baixa	–	nula-baixa
D	FTP	baixa-média	alto	alta	–	média-alta
E	*web*	baixa-média	médio	média	–	baixa
F	*e-mail, news*	baixa	alto	baixa	–	média-alta

sugerem uma classificação baseada em apenas dois parâmetros: a taxa de perdas e o atraso. Ao se considerar os quatro parâmetros clássicos de QoS e ainda o fator de conectividade, podem ser derivadas novas classes, atendendo não só as características das aplicações mas também as novas necessidades impostas pelo ambiente de mobilidade. A tabela 1 apresenta essas classes.

A classe para tráfego de controle, de gerenciamento de rede e de informações de emergência em geral (classe A) é uma classe crítica onde não são admitidas perdas ou atrasos significativos. A banda passante necessária, além de ser uma pequena fração do total disponível, é em geral bem conhecida, uma vez que tal tráfego é previsível [7]. A urgência desse tipo de tráfego exige alta conectividade, levando as probabilidades de desconexão a serem nulas, no que depender de provisionamento. A classe B, para aplicações interativas e de tempo real como áudio, vídeo, telefonia e vídeo conferências, tem como requisito a baixa latência. Essas aplicações são, em geral, robustas a determinados níveis de taxa de erros. A banda passante utilizada pode ser alta, em casos de vídeo por exemplo, ou média como para telefonia. O *jitter* deve ser mantido pequeno para que se evitem *buffers* de recepção muito grandes. A probabilidade de desconexão deve ser mantida baixa, evitando-se interrupções no serviço. Ao se pensar em aplicações como login remoto (classe C) por exemplo, obtém-se uma classe onde atrasos moderados são aceitáveis, assim como taxas de perda medianas. Tais aplicações tem como característica a utilização de pouca banda passante. A conectividade do serviço deve ser alta, para o conforto do usuário. Já as aplicações de transferência de arquivos (classe D), apresentam características de altas vazões, com menos importância para atraso. Se existirem esquemas de recuperação de transmissões interrompidas, a conectividade deixa de ser um fator primordial e as probabilidades de desconexão podem ser mais elevadas. Uma classe E, adequada para aplicativos de navegação na *Web*, por sua vez, exige banda passante moderada, mantendo as necessidades de interatividade. Aplicações de pequena vazão e que não exigem interatividade, como *e-mail* e *news*, podem pertencer a uma outra classe de serviço, (classe F), que exige confiabilidade. Mais uma vez, a probabilidade de desconexão alta não influenciará na satisfação do usuário.

3.2 Diferenciação de Serviços na Internet

A definição das classes de serviço pode ser vista como um primeiro passo na busca de uma estrutura de QoS para as redes móveis. Uma outra preocupação é a arquitetura de QoS a ser empregada. Como sugerido anteriormente, é interessante que se procure esquemas que permitam a interoperação transparente com o restante da Internet, que é a rede de dados mais representativa nos dias de hoje.

Uma vez que os serviços oferecidos pela rede móvel podem variar abruptamente, de acordo com o apresentado na seção 2, é importante que os mecanismos de QoS sejam adaptativos, ágeis às mudanças nas características oferecidas. Para tornar isso possível alguns autores frisam a importância da existência de esquemas de sinalização [5, 7]. De qualquer forma, as garantias oferecidas não podem ser rígidas como nos casos de redes fixas.

Uma solução possível para tornar adaptativo o provisionamento de QoS é a utilização de um agente central para o gerenciamento de recursos. Uma arquitetura assim é sugerida em [7]. Nela, o agente central obtém informações dos roteadores de borda através de protocolos de sinalização. Possuindo informações de todo o domínio ele pode realizar procedimentos de AAA (*Authentication, Authorization and Acounting*) e de gerenciamento de QoS. Uma desvantagem dessa proposta consiste na complexidade adicionada pela existência do agente centralizador e da sinalização explícita, que gera tráfego para fins diferentes da transmissão de dados (*overhead*). Além disso, o esquema apresentado não é baseado nos paradigmas da proposta de Diferenciação de Serviços (DiffServ) [6, 14], que parece ser a solução mais promissora para a Internet atualmente.

A arquitetura de QoS com DiffServ utiliza o campo TOS do cabeçalho IP para diferenciar os fluxos e os trata como agregados. No interior do domínio DiffServ os agregados de fluxos de cada classe são tratados e encaminhados nó a nó de acordo com seus PHB's (*Per Hop Behavior*). Atualmente a arquitetura DiffServ oferece dois PHB's padronizados pelo IETF: O Encaminhamento Expresso (EF - *Expedited Forwarding* [15]) e o Encaminhamento Assegurado (AF - *Assured Forwarding* [16]).

Através da utilização do PHB EF, pode ser obtido o serviço de "linha privativa virtual", que oferece garantias de baixo atraso, baixo *jitter* e pequena taxa de perdas, oferecendo uma banda passante assegurada ao agregado. Para realizar isso, é necessária a garantia de que a taxa de serviço nos nós do domínio EF será superior à taxa de chegada. Dessa forma, as filas nos roteadores estarão sempre minimizadas. Para tanto, esse serviço emprega condicionadores de tráfego nos nós de ingresso, de forma a garantir a correta relação entre as taxas. O EF é então interessante para aplicações de tempo real. O AF, por sua vez, procura garantir vazão dos dados, permitindo a ocorrência de rajadas de pacotes, desde que estas sejam de curta duração. Desta forma congestionamentos de curta duração são permitidos nas filas dos nós AF. Os congestionamentos longos, entretanto, são sempre evitados. Para isso, mecanismos de gerenciamento ativo de filas, como RED [17] são utilizados.

Uma forma de se obter a flexibilidade necessária para o provisionamento adequado de QoS em redes móveis seguindo a arquitetura DiffServ deve ser estudada. Uma opção é deixar essa flexibilidade implícita na escolha de PHB's menos rígidos. Assim, em casos críticos, onde a capacidade da infra-estrutura é excedida, diferentes tipos de degradação de serviços podem ser oferecidos aos nós móveis. Em casos onde a degradação não fosse permitida, algum método de controle de admissão se faria necessário, negando o acesso caso a demanda ultrapassasse certos níveis, onde o provisionamento passasse a ser inadequado. Em situações menos rígidas, a banda passante disponível poderia diminuir até níveis mínimos. Opcionalmente, outro tipo de serviço poderia sofrer degradação através de maiores taxas de perdas, ou até mesmo maiores atrasos e *jitter*. O fato do terminal estar parado ou se movendo também pode ser um fator para a degradação do serviço prestado [18], uma vez que garantir QoS após um *handover* é uma tarefa difícil, que envolve reservas de recursos. As reservas são recursos que ficarão na maior parte do tempo inutilizados, não sendo interessantes para o provedor de serviço.

3.3 Mapeamento das classes

Uma outra etapa da definição de um esquema de QoS inclui o mapeamento entre classes de serviços requisitadas e possíveis PHB's para atendê-las. O mapeamento do serviço da classe A, por exemplo, é equivalente ao mapeamento de um serviço de "linha privada virtual" oferecido pelo pelo PHB EF. Esse serviço teria que ter tráfego pequeno, onde o seu máximo pudesse ser previsível. Desse modo seria possível garantir o correto provisionamento de recursos, incluindo reservas adequadas para que mesmo com a realização de *handovers* parâmetros de QoS, como conectividade, continuem garantidos.

A classe B oferece características de tempo real, com baixo atraso. Por causa de características da mobilidade, é necessário que exista um serviço mais flexível. Dessa forma essa classe será dividida em duas subclasses, ambas mantendo as características de baixos atrasos, porém uma delas apresentando maior flexibilidade em termos de taxa de perdas. Um PHB EFD (*Expedited Forwarding with Dropping*), sugerido em [18] poderia ser empregado. Essa classe será estudada detalhadamente na seção 3.4.

O tráfego para classe C é parecido com o da classe B, exceto pela menor taxa de perdas e menor banda passante. Uma solução seria incluir esta classe para ser tratada pelo PHB EFD, acrescentando-se mecanismos de prioridade, para diminuir a taxa de perdas. Como a vazão exigida é pequena, talvez esse tráfego não interfira significativamente no desempenho do tráfego de tempo real. Para prover as características da classe D, os serviços oferecidos pelo AF seriam adequados. Nessa classe, a conectividade não é um fator importante. Dessa forma, a conexão poderia ser aceita somente se houvesse taxas mínimas garantidas. Aplicativos de navegação na *Web*, pertencentes a classe E, possuem característica de tráfego de curta duração. Assim o PHB AF, devidamente sintonizado para tráfegos curtos, seria adequado. A classe F poderia ser atendida por mecanismos de melhor esforço, por não ser exigente em relaçãp aos parâmetros de QoS.

3.4 Classe B: Tempo real em mobilidade

Para o serviço B, que exige tempo real porém onde o tráfego não é tão previsível como na classe A, a mobilidade do nó é um fator que influencia a qualidade de serviço oferecida. A característica de baixos atrasos deve sempre ser mantida. A qualidade em termos de taxa de perdas, entretanto, pode ser modificada. Assim, caso o nó esteja parado em uma célula, seu serviço de tráfego sem perdas e com baixo atraso estaria garantido. Ao realizar *handover*, entretanto, o nó móvel enfrentaria uma probabilidade de mudar de serviço. Essa probabilidade iria depender da quantidade de recursos reservados para fins de *handover*.

Desta forma, esse artigo propõe que clientes do serviço B que cheguem a uma região com escassez de recursos fiquem temporariamente utilizando uma sub-classe B1, com maiores taxas de perdas. Essa classe seria mais flexível, permitindo que um nó, mesmo estando parado, sofra degradação até determinados níveis em sua qualidade. Essa degradação ocorreria por causa do aumento do número de nós dentro de uma dada região. Durante esses períodos de sobrecarga, clientes da classe B que já estivessem na célula antes dos recursos se tornarem escassos ficariam temporariamente pertencendo a uma sub-classe prioritária B0, mantendo as características da classe B incluindo baixas perdas.

Para o atendimento de classes de tempo real com perdas o PHB apresentado em [18] pode ser uma solução. O PHB sugerido é o encaminhamento expresso com perdas (EFD - *Expedited Forwarding with Dropping*). Similarmente ao EF, o EFD garante atrasos pequenos. Isso se dá porque o EFD utiliza os recursos reservados para o tráfego EF, que entretanto não foram utilizados. Dessa forma, os pacotes EFD ocupam a mesma fila dos pacotes EF. Se a fila não estiver cheia, não haverá problemas. Caso contrário, o transbordo do *buffer* estará sendo ocasionado pelos pacotes EFD, uma vez que o tráfego EF é condicionado de modo a não sobrecarregar as filas dos roteadores. Nessa situação de chegada de pacotes na fila cheia, os pacotes EFD serão descartados preferencialmente, ou seja, com a chegada de novos pacotes EF, pacotes EFD são obrigatoriamente descartados de modo a oferecer espaço nos *buffers* do nó. Técnicas de descarte dos pacotes da fila (*push-out*) podem influenciar no padrão de perdas obtido, evitando longos períodos sem pacotes por exemplo. Isso seria interessante no caso de tráfego de voz e vídeo, onde longas rajadas influenciam na qualidade percebida pelo usuário.

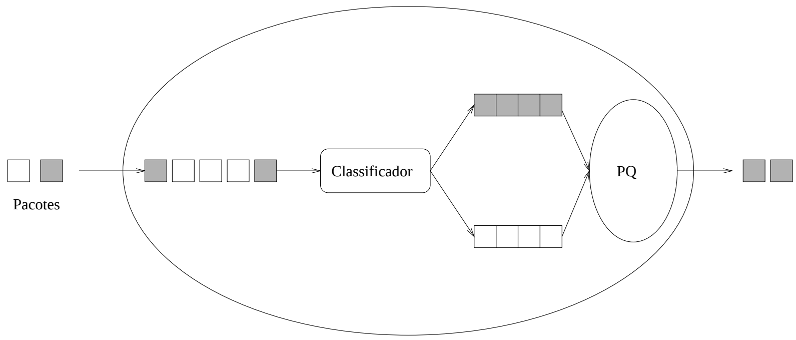
Uma opção para a implementação de classes de tempo real com perdas concorrendo com tráfego prioritário, sem perdas, é a utilização de políticas protetoras para o gerenciamento de *buffers*. Tais políticas garantem o serviço da classe prioritária oferecendo porém garantias para a classe inferior. Um exemplo de tais políticas é apresentado em [19] e será discutido na seção 4.

4 Mecanismos de suporte

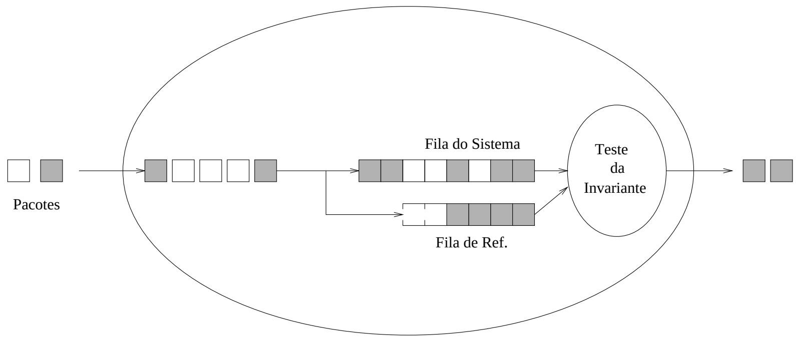
Ao se especificar uma classe de serviço, deixa-se em aberto as diversas possibilidades de sua implementação. Diferentes mecanismos de suporte para realizar os esquemas de diferenciação de tráfego podem ser arquitetados, desde que estes consigam prover as necessidades especificadas em cada classe. Nessa seção serão estudados dois mecanismos diferentes para fornecer os serviços de tempo real das classes B0 e B1, como especificadas na seção anterior. Ambas as classes devem fornecer um serviço de baixo atraso, sendo que a classe B1 apresenta maior tolerância a perdas.

O primeiro mecanismo a ser avaliado é baseado em múltiplas filas, sendo escalonadas de acordo com a prioridade do tráfego. A figura 2(a) apresenta o esquema descrito. Os tráfegos da classe B0 e B1 são marcados na entrada do domínio DiffServ. Cada uma das agregações será então classificada para a fila correspondente no roteador. Ao transmitir, um escalonador com prioridade (PQ-*Priority Queue*) escolhe de qual fila será enviado o próximo pacote. Assim, sempre que houver pacote de tráfego B0, este será enviado, e os pacotes de tráfego B1 somente poderão ser transmitidos caso a fila prioritária esteja desocupada. Este esquema possui como principal vantagem a simplicidade de implementação, podendo entretanto causar grandes atrasos ou até mesmo *starvation* na transmissão de pacotes menos prioritários. Essa solução, portanto, garante o desempenho do tráfego prioritário, sem oferecer garantias ao tráfego de menor prioridade.

Uma outra solução seria a utilização de políticas protetoras, como a SPP (*Simulated Protective Policy*), apresentada em [19]. Uma política é dita protetora quando ela conse-



(a) Nó com duas filas.



(b) Nó com SPP.

Figura 2: Mecanismos para diferenciação de classes.

gue manter as garantias do tráfego prioritário, como taxa de perdas, independentemente da carga e dos padrões de chegada do tráfego de menor prioridade. Tais políticas são importantes nas condições em que torna-se difícil a tarefa de prever o comportamento do tráfego menos prioritário, como no caso das redes móveis. O mecanismo de SPP possui ainda uma característica que a faz importante para o cenário estudado, onde o desempenho do tráfego B1 não deve ser deixado de lado. O SPP é uma política que garante o desempenho mínimo para o tráfego mais prioritário (B0, no caso de redes móveis aqui apresentado) mas também otimiza o desempenho do tráfego menos prioritário B1.

O SPP funciona simulando uma fila de referência que representa uma fração do *buffer* do nó. Os pacotes mais prioritários que chegam são armazenados no *buffer* e também na fila simulada. Os pacotes menos prioritários só podem ocupar o *buffer*, não entrando na fila simulada. Caso o *buffer* esteja cheio e um pacote mais prioritário chegue, um menos prioritário deverá ser descartado caso ainda haja espaço na fila simulada. No momento do envio de pacotes uma invariante também precisa ser respeitada: o número de espaços disponíveis no *buffer* para pacotes de alta prioridade deve sempre ser maior ou igual ao espaço disponível na fila de referência. Deve se notar que os espaços ocupados por pacotes de baixa prioridade contam como espaços vazios para os de maior prioridade, uma vez que são descartados em caso de necessidade (através de técnicas de *push-out*). Essa invariante deve ser verificada ao se tentar servir pacotes de menor prioridade. Caso ela seja violada, o pacote de baixa prioridade que seria enviado deve ser descartado.

A política SPP pode então ser utilizada para prover os serviços desejados pelas classes

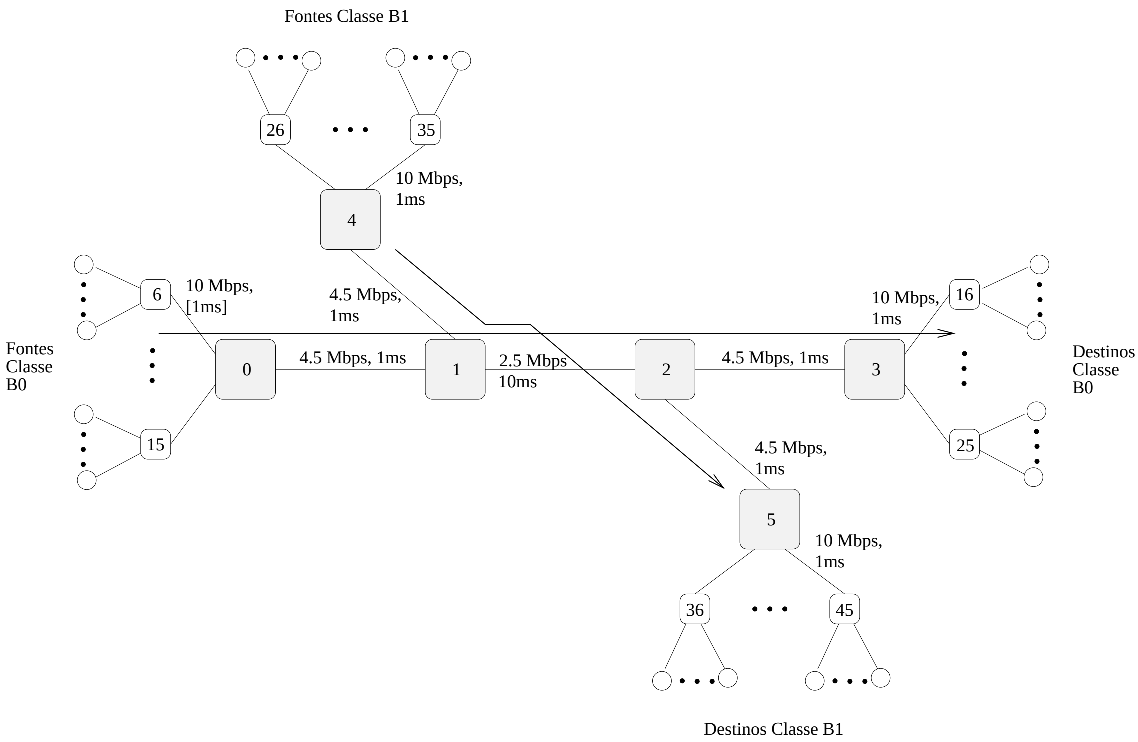


Figura 3: Topologia.

B0 e B1, evitando perdas para a classe B0 e buscando otimizar o atendimento da classe B1. A figura 2(b) apresenta este esquema. Os pacotes das diferentes classes são marcados na entrada do domínio DiffServ. A fila SPP trata diferenciadamente os pacotes.

Uma avaliação do desempenho de ambas as soluções foi realizada através de simulações que serão apresentadas na próxima seção.

5 Avaliação

A avaliação dos mecanismos propostos na seção 4 foi realizada através de simulações. O simulador utilizado foi o ns-2 (*network simulator-2* [20]). Extensões para dar suporte aos mecanismos de DiffServ foram implementadas no GTA-COPPE [21, 22].

A topologia simulada está apresentada na figura 3. Nesse estudo a preocupação principal não foi modelar a rede sem fio, mas sim avaliar os mecanismos propostos quanto a garantias de atraso e o grau de diferenciação obtido, levando-se em consideração a chegada de novas fontes, devido à mobilidade. O atraso foi escolhido como métrica porque ele é o fator mais importante para as aplicações multimídia de tempo real, as quais são focadas pelas classes B0 e B1 simuladas. A taxa de perdas também é parâmetro de QoS para as classes simuladas, devendo ocorrer proteção da classe B0.

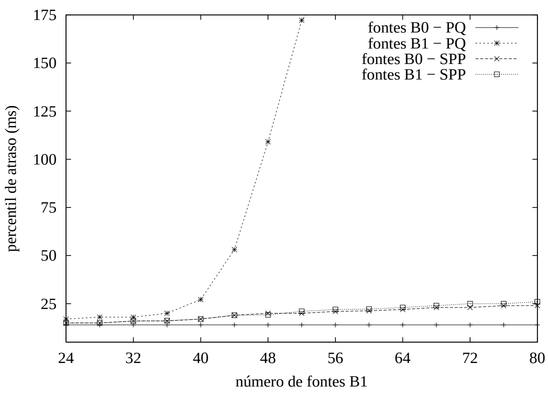
A modelagem de tráfego de tempo real foi realizada utilizando-se fontes de tráfego *on-off* [23]. Essas fontes apresentam intervalos de atividade exponencialmente distribuídos com média de 400 ms e intervalos de inatividade com distribuição equivalente, porém com média de 600 ms. A taxa de transmissão nos intervalos de atividade é de 64 kbps. Essas fontes procuram representar tráfego de voz codificado por PCM com supressão de silêncios, sendo transmitidas pela rede IP através do protocolo de transporte UDP. O tráfego gerado por cada fonte apresenta então uma taxa média de 25,6 kbps.

As fontes foram divididas entre prioritárias, da classe B0, e menos prioritárias, da classe B1. As fontes B0, conforme descrito na seção 3.4, representam nós que já estavam acessando a rede móvel antes da situação de sobrecarga. Os novos nós que chegam, causando a sobrecarga, são pertencentes a classe B1. Inicialmente, 40 fontes da classe B0 encontram-se transmitindo na rede. Essas fontes geram em conjunto uma taxa média de 1,024 Mbps, representando 40,96% do enlace de gargalo. Este número permanece constante e as novas fontes que chegam são encaminhadas para a classe B1.

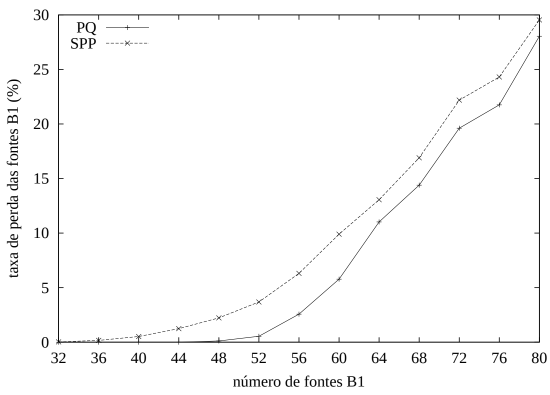
As fontes B0 estão conectadas através dos nós de 6 a 15, sendo endereçadas aos nós 16 ao 25. As fontes B1 se conectam através dos nós 26 a 35 e destinam-se aos nós 36 a 45. O enlace de gargalo, é aquele entre os nós 1 e 2. Os mecanismos descritos na seção anterior atuam no nó 1, de entrada do gargalo. Os nós 0 e 4 poderiam ser vistos como um único nó, representando o roteador que conecta a rede sem fio à infra-estrutura fixa. As características de vazão e atraso dos enlaces estão indicadas na figura.

O tamanho total do *buffer* do nó 1 é constante, tanto para o caso de uma fila (SPP) como para o esquema com duas filas (PQ). Nesse último, ambas as filas têm o mesmo tamanho, igual a metade da fila única do SPP. Esse fato representaria um roteador com uma quantidade limitada e fixa de memória, sendo configurado para um dos dois esquemas.

A figura 4 apresenta os resultados da simulação para atraso e taxa de perdas obtidos na rede. Os resultados apresentados no gráfico 4(a) mostram o percentil-95 de atraso em ambas as classes, com o aumento do número de fontes B1 para cada mecanismo simulado. A taxa de perdas para a classe B1 é apresentada na figura 4(b). As perdas da classe B0 para ambos os mecanismos ficaram muito baixas, em torno de zero, e por isso não são mostradas.



(a) Percentil-95 para atraso.



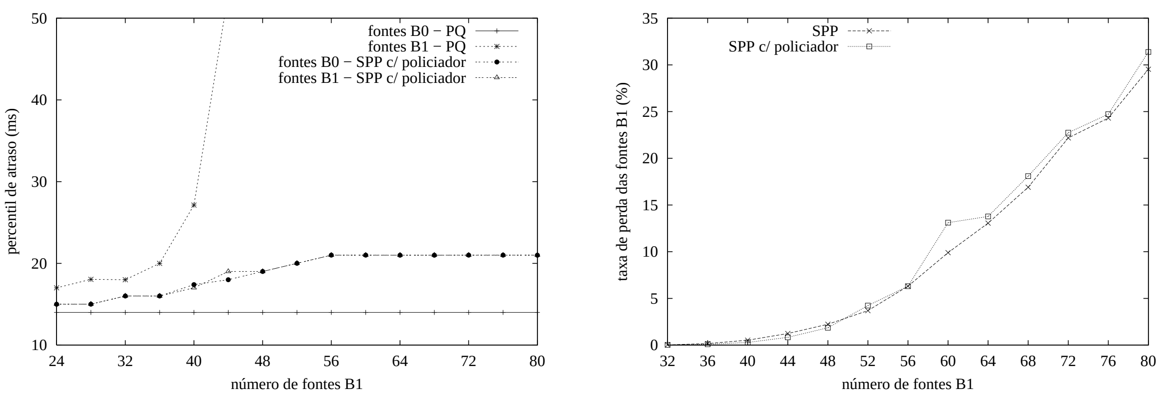
(b) Perdas da classe B1.

Figura 4: Comportamento do atraso e taxa de perdas com o aumento de fontes menos prioritárias. SPP e duas filas (PQ).

Observa-se que o menor atraso é obtido para o tráfego B0 do mecanismo de duas filas (PQ). Em contrapartida, a classe B1 desse mecanismo não oferece garantias quanto a atraso. Isso pode ser percebido pelo grande aumento do percentil-95 com a chegada de clientes dessa classe. A curva do percentil para o SPP, por sua vez, é mantido baixo (menor que 25 ms) para ambas as classes. Conforme colocado anteriormente, a taxa de perdas da classe B0 se mantém próxima de zero, como desejado para ambos os mecanismos. A classe B1, entretanto, apresenta crescimento significativo na taxa de perdas, de acordo com a figura 4(b). Esse comportamento era esperado, estando previsto na especificação da classe. A taxa de perdas oferecida à classe B1 pelo mecanismo PQ fica ligeiramente abaixo da observada com o SPP, entretanto o grande crescimento no atraso para a classe menos prioritária nesse mecanismo inviabiliza sua utilização em tráfego de tempo real[2]. A principal desvantagem observada no SPP é o crescimento de seu percentil de atraso com o aumento do número de fontes, fato que é notado pela derivada positiva da curva do percentil. Isso mostra que em termos de atraso o tráfego B0 é influenciado pela chegada de fontes B1, o que não é desejável.

O fato do atraso no tráfego da classe B0 crescer de acordo com o aumento das fontes de classe B1 demonstra também pouco isolamento entre classes. Esse aumento ocorre porque no SPP ambas as classes ocupam a mesma fila. Assim, o crescimento na ocupação dessa fila traz aumentos no atraso para as duas classes. Uma forma de se estabelecer um limite para a interferência causada pelo tráfego B1 no tráfego B0 é através da utilização de um mecanismo de controle de admissão para o primeiro. Esquemas de controle de admissão por conexão não seriam interessantes por causa da complexidade adicionada e por fugir da proposta de DiffServ. Dessa forma, um método de controle de admissão em nível de pacotes (um policiador) foi considerado. Um *token bucket* [24] foi escolhido para exercer o policiamento de tráfego, descartando os pacotes da classe B1 quando seu agregado ultrapassar uma taxa limite. Dessa maneira, o atraso global é mantido baixo, uma vez que as filas não são mantidas excessivamente cheias. Nesse caso, uma opção pela diminuição do atraso, em detrimento da taxa de perdas, foi realizada.

O policiamento foi realizado no nó de borda número 4, mantendo a taxa do agregado de fontes B1 limitada. A taxa limite foi escolhida de modo a permitir uma taxa média equivalente àquela gerada por 60 fontes. Quando a taxa média ultrapassa esse valor, pacotes passam a ser descartados. Essa simulação é apresentada na figura 5. É importante observar que nesse caso o policiador foi utilizado apenas em conjunto com a política SPP. A política PQ não utilizou policiadores.



(a) Percentil-95 para atraso.

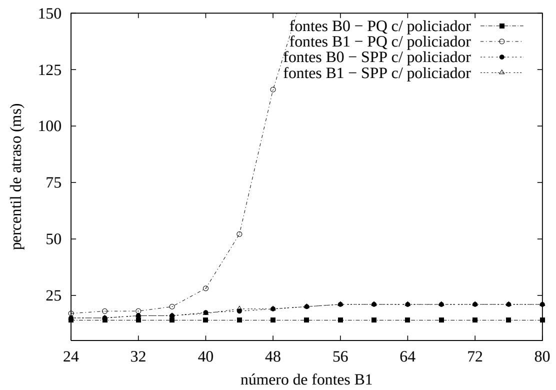
(b) Perdas do SPP com e sem o policiador.

Figura 5: Acrescentando o policiador para SPP.

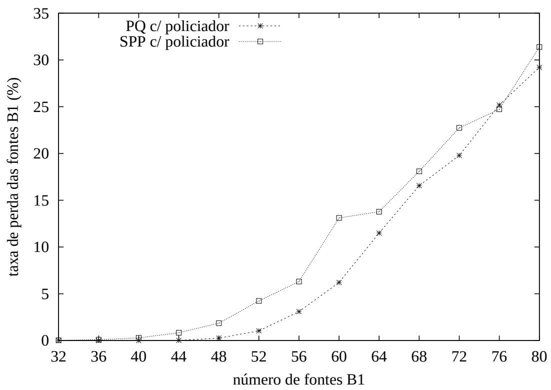
Através do gráfico 5(a) pode ser observada a atuação do policiador. A curva referente

[2]Em telefonia, por exemplo, atrasos de até 150 ms são aceitáveis.

às classes do SPP cresce de forma aproximadamente linear com o aumento do número de fontes. A partir da taxa equivalente a 60 fontes, entretanto, a curva passa a se manter constante. A taxa de perdas para o SPP, observada na figura 5(b), é ligeiramente maior que a observada no caso anterior, sem policiador. Esse fato era esperado, uma vez que o policiador descarta o tráfego excedente. As perdas para a classe B1 do SPP ocorrem agora tanto no núcleo da rede, através de perdas na fila do nó, como também na borda, através do policiador.



(a) Percentil-95 para atraso.



(b) Perdas da classe B1.

Figura 6: Acrescentando o policiador para os dois mecanismos.

Uma nova simulação foi realizada para garantir uma comparação justa entre os mecanismos. Para tanto, incluiu-se o policiador para o tráfego B1 no esquema do PQ. Como observado na figura 6, não houve mudanças significativas. A alta taxa de crescimento na curva do percentil de tráfego B0, para o esquema com duas filas, se manteve, mesmo com a limitação de taxa imposta pelo policiador. Esse fato ocorre por causa de características particulares do escalonador de prioridade, que não serve tráfego B1 enquanto houver pacotes B0 na fila prioritária. A taxa de perdas do tráfego B1 para o esquema de duas filas sofre um ligeiro aumento, causado pelo policiador.

As simulações mostraram que o mecanismo PQ, apesar de oferecer os melhores resultados de atraso e perdas para a classe B0, não oferece nenhum tipo de garantia em termos de atraso para a classe B1. Assim, tal mecanismo não é adequado para um esquema de suporte e diferenciação das classes B0 e B1. Isso porque a classe B1 tem como objetivo dar suporte a aplicações multimídia de tempo real. Desse modo, mesmo sendo de menor prioridade com relação a perdas, ela deve receber serviço de baixos atrasos, de modo similar ao da classe B0. O mecanismo de SPP, quando aplicado em conjunto com um policiador do tráfego B1, conseguiu fornecer um serviço com atraso baixo e limitado para ambas as classes. Além disso, as condições de taxas de perda próximas de zero para o tráfego B0 foram satisfeitas. A taxa de perdas para a classe B1, entretanto, ficou mais alta que a encontrada com o mecanismo PQ.

6 Conclusão

Este artigo apresentou uma estrutura de diferenciação de serviços voltada para o ambiente de mobilidade. Foi abordado um cenário típico de interconexão entre redes móveis e fixas. Problemas ocasionados pela mobilidade dos nós foram explorados. Foram propostas classes de serviço a partir de características das redes sem fio e das aplicações que demandam QoS. Mecanismos apresentados para dar suporte a uma das classes em questão foram avaliados através de simulações.

A partir das simulações, observou-se que o mecanismo de suporte baseado na política protetora SPP foi mais adequado que o esquema baseado em duas filas. Ambas as políticas ofereceram bom isolamento entre classes em termos de perda. A classe prioritária B0 manteve-se protegida, não apresentando perdas significativas.

A política baseada em duas filas ofereceu os melhores resultados de atraso para classe prioritária B0, porém não ofereceu garantias à classe B1. Por sua vez, o mecanismo SPP manteve o atraso para ambas as classes pequeno, apesar de sua curva de percentil aumentar de forma aproximadamente linear, aparentemente sem garantias de limitação. A partir dessa característica, verificou-se a necessidade de controle de admissão para garantir o atraso máximo dos tráfegos, independentemente do volume de tráfego B1. O mecanismo empregado foi de controle de admissão em nível de pacotes, através do algoritmo de *token bucket*. Com o mecanismo de controle de admissão integrado ao esquema do SPP foi possível fornecer o serviço desejado, garantindo atraso baixo e limitado para ambas as classes.

Como continuação desse trabalho, pretende-se investigar novos mecanismos que possam dar suporte a novas classes de serviço voltadas para a mobilidade. Políticas que combinem propriedades de prioridade tanto espacial como temporal podem oferecer diferenciação de atraso e perdas [25], podendo ser empregadas para a criação de mecanismos de suporte às classes. Para isso, políticas que permitam incorporar o parâmetro de QoS "padrão de perdas" às classes também serão estudados. Esquemas de *push-out* são uma opção que merecem investigação. Cenários de simulação que modelem de maneira mais realista as redes móveis também deverão ser realizadas.

Referências

[1] I. Guardini, P. D'Urso, and P. Fasano, "The Role of Internet Technology in Future Mobile Data Systems," *IEEE Communications Magazine*, vol. 38, no. 11, pp. 68–72, 2000.

[2] B. Li, M. Hamdi, D. Jiang, X.-R. Cao, and Y. T. Hou, "QoS-Enabled Voice Support in the Next-Generation Internet: Issues, Existing Approaches and Challenges," *IEEE Communications Magazine*, vol. 38, pp. 54–61, Apr. 2000.

[3] G. H. Forman and J. Zahorjan, "The Challenges of Mobile Computing," tech. rep., CSE - University of Washington, Mar. 1994. `URL: ftp://ftp.cs.washington.edu`.

[4] T. Braun, C. Castelluccia, and G. Stattenberger, "An Analysis of the DiffServ Approach in Mobile Environments," in *First Workshop on IP Quality of Service for Wireless and Mobile Networking, IQWiM'99*, (Germany), Apr. 1999.

[5] I. Mahadevan and K. M. Sivalingam, "Architecture and Experimental Framework for Supporting QoS in Wireless Networks Using Differentiated Services," *ACM/Baltzer Mobile Networks and Applications Journal*, May 2000.

[6] S. Blake, D. L. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An Architecture for Differentiated Services," *Internet RFC 2475*, Dec. 1998.

[7] J.-C. Chen, A. McAuley, A. Caro, S. Baba, Y. Ohba, and P. Ramanathan, "QoS Architecture Based on Differentieted Services for Next Generation Wireless IP Networks," *Internet Draft*, July 2000.

[8] A. Fieger, M. Zitterbart, R. Keller, and J. Diederich, "Towards QoS-support in the Presence of Handover," in *First Workshop on IP Quality of Service for Wireless and Mobile Networking, IQWiM'99*, (Germany), Apr. 1999.

[9] E. A. Brewer and R. H. Katz, "A Network Architecture for Heterogeneous Mobile Computing," *IEEE Personal Communications*, Oct. 1998.

[10] R. Kalden, I. Meirik, and M. Meyer, "Wireless Internet Access Based on GPRS," *IEEE Personal Communications*, vol. 7, pp. 8–18, Apr. 2000.

[11] S. Singh, "Quality of Service Guarantees in Mobile Computing," *Journal of Computer Communications*, vol. 19, no. 4, pp. 359–371, 1996.

[12] L. Mathy, C. Edwards, and D. Hutchison, "The Internet: A Global Telecommunications Solution?," *IEEE Network Magazine*, vol. 14, pp. 46–57, July 2000.

[13] K. Seal and S. Singh, "Loss profiles: A Quality of Service Measure in Mobile Computing," *Journal of Wireless Networks*, vol. 2, no. 1, pp. 45–61, 1996.

[14] K. Nichols, S. Blake, F. Baker, and D. L. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers," *Internet RFC 2474*, Dec. 1998.

[15] V. Jacobson, K. Nichols, and K. Poduri, "An Expedited Forwarding PHB," *Internet RFC 2598*, June 1999.

[16] J. Heinanen, F. Baker, W. Weiss, and J. Wroclawski, "Assured Forwarding PHB group," *Internet RFC 2597*, June 1999.

[17] S. Floyd and V. Jacobson, "Random Early Detection Gateways for Congestion Avoidance," *IEEE/ACM Transactions on Networking*, vol. 1, pp. 397–413, Aug. 1993.

[18] J. Diederich and M. Zitterbart, "An Expedited Forwarding with dropping PHB," *Internet Draft*, Oct. 1999.

[19] I. Cidon, R. Guérin, and A. Khamisy, "On protective buffer policies," in *Proceedings of INFOCOM '93*, pp. 1051–1058, 1993.

[20] K. Fall and K. Varadhan, "Network simulator - NS (version 2) notes and documentation," tech. rep., The VINT Project, July 1999. `URL: http://www-mash.cs.berkeley.edu/ns/`.

[21] A. Ziviani, J. F. de Rezende, and O. C. M. B. Duarte, "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), pp. 354–367, maio 1999.

[22] J. F. de Rezende, "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), pp. 339–353, maio 1999.

[23] A. Ziviani, J. F. de Rezende, and O. C. M. B. Duarte, "Avaliando o Suporte ao Tráfego de Voz em um Ambiente de Diferenciação de Serviços," in *VI Simpósio Brasileiro de Multimídia e Sistemas Hipermídia - SBMIDIA'2000*, (Brasil), pp. 341–356, Junho 2000.

[24] A. S. Tanenbaum, *Computer Networks*. New Jersey: Prentice Hall, 3rd. ed., 1996.

[25] A. Gravey and H. Hébuterne, "Mixing Time and Loss Priorities in a Single Server Queue," in *Proceedings of ITC-13*, (Denmark), pp. 147–152, June 1990.