COPPE/UFRJ

MODELAGEM ANALÍTICA E AVALIAÇÃO DO RETARDO DAS MENSAGENS NO PROTOCOLO DE ACESSO AO MEIO DO PADRÃO IEEE 802.16

Danielle Lopes Ferreira Gonçalves Vieira

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia de Sistemas e Computação.

Orientador(es): Luís Felipe Magalhães de Moraes

Rio de Janeiro

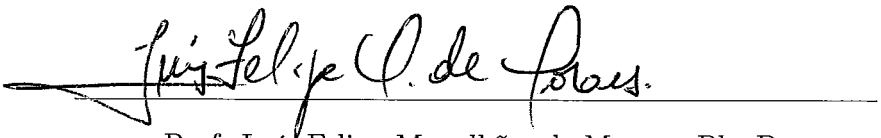
Setembro de 2008

MODELAGEM ANALÍTICA E AVALIAÇÃO DO RETARDO DAS MENSAGENS NO PROTOCOLO DE ACESSO AO MEIO DO PADRÃO IEEE 802.16

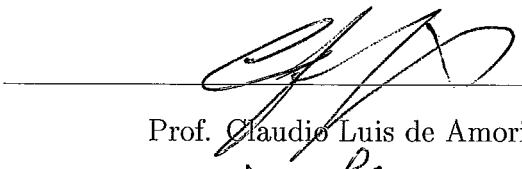
Danielle Lopes Ferreira Gonçalves Vieira

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

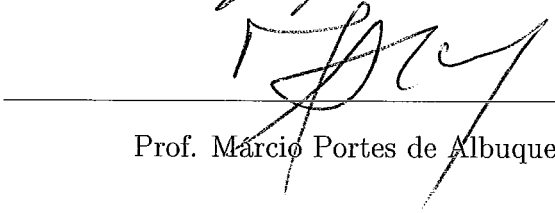
Aprovada por:



Prof. Luís Felipe Magalhães de Moraes, Ph. D.



Prof. Claudio Luis de Amorim, Ph. D.



Prof. Márcio Portes de Albuquerque, Dr.

RIO DE JANEIRO, RJ - BRASIL

SETEMBRO DE 2008

Vieira, Danielle Lopes Ferreira Gonçalves

Modelagem Analítica e Avaliação do Retardo das Mensagens no Protocolo de Acesso ao Meio do Padrão IEEE 802.16/Danielle Lopes Ferreira Gonçalves Vieira. - Rio de Janeiro: UFRJ/COPPE, 2008.

XVIII, 93 p.: il.; 29,7 cm.

Orientador: Luís Felipe Magalhães de Moraes

Dissertação (mestrado) - UFRJ/COPPE/Programa de Engenharia de Sistemas e Computação, 2008.

Referências Bibliográficas: p. 84-88.

1. Redes Metropolitanas Sem Fio. 2. Protocolos de Acesso ao Meio. 3. Qualidade de Serviço. 4. Avaliação de Desempenho I. Moraes, Luís Felipe Magalhães de II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

Dedicatória

Dedico esse trabalho as minhas lindas e admiráveis filhas Eduarda e Pietra, que são a razão de minha vida. Desculpem as horas roubadas de nosso convívio, pelo tempo dedicado ao estudo, concedendo a mim a oportunidade de me realizar ainda mais.

Agradecimentos

Primeiramente, gostaria de agradecer a Deus pelo dom da vida e pela graça de poder realizar este trabalho.

Ao meu pai Jorge, pelas palavras sábias e incentivadoras nos momentos difíceis, que me deram força e coragem para continuar. Agradecimento mais que especial a minha mãe, pelo apoio, suporte, amor e preocupação, sempre. Sem você eu não teria conseguido. Carinhosamente ao meu marido Edward. Aos meus irmãos Jefferson, Vinícius e Débora (em especial a querida irmã e amiga Débora, pelo apoio incondicional ao meu estudo), as minhas filhas Eduarda e Pietra, ao cunhado Serginho, bem como, toda a minha família, por todo amor e carinho, não só durante a realização deste trabalho, como na vida inteira.

Agradeço ao meu orientador, Prof. Luís Felipe, pelos grandes ensinamentos e pelo total apoio desde o início do meu trabalho e aos demais integrantes da banca, os Professores Claudio Amorim e Márcio Portes, pela valiosa ajuda nesta fase final.

Agradeço a todos os amigos que conheci durante este período, em particular: Tiago, Rafael Fernandes, Rafael Bezerra, Júlio, Paulo, Bruno, Eduardo, Airon, Cláudia, Diogo, Verissimo, Schiller, Michelini e todos os outros, que por ventura eu tenha esquecido. Agradecimento especial ao amigo Gustavo, pela ajuda, contribuição, companheirismo e amizade. Agradecimento muito especial ao amigo Beto, pela amizade, pela ajuda e pelas grandes contribuições na reta final de elaboração dessa dissertação.

Ao Programa de Engenharia de Sistemas e Computação (PESC/COPPE/UFRJ), pelo apoio operacional.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

MODELAGEM ANALÍTICA E AVALIAÇÃO DO RETARDO DAS MENSAGENS NO PROTOCOLO DE ACESSO AO MEIO DO PADRÃO IEEE 802.16

Danielle Lopes Ferreira

Setembro/2008

Orientador: Luís Felipe Magalhães de Moraes

Programa: Engenharia de Sistemas e Computação

O padrão IEEE 802.16 surgiu como uma solução para o acesso à banda larga através de meios sem fio, sendo desenvolvido para transmitir dados e aplicações multimídia com diferentes requisitos de qualidade de serviço (QoS). Visando alcançar o nível de QoS desejado para aplicações multimídia, esse padrão fornece diferentes mecanismos de escalonamento. Vários trabalhos na literatura investigam o impacto dos mecanismos de escalonamento no desempenho dessas redes. A maioria desses trabalhos apresentam resultados de avaliações de desempenho obtidos através de simulação ou propõem modelos analíticos para essa avaliação, mas com alguma alteração do protocolo de acesso ao meio (MAC) do padrão IEEE 802.16. Dentro desse contexto, este trabalho modela o retardo total das mensagens de aplicações em tempo real e outros, transmitidas pelo protocolo MAC do padrão IEEE 802.16. O retardo total das mensagens é obtido através de dois modelos distintos: um para a fase de contenção (disputa pelo meio físico) aplicado somente ao tráfego de tempo não-real; e outro para a fase de alocação de dados, onde o modelo suporta duas classes de prioridade de tráfego, uma para serviço de tempo real e outra para tempo não-real. Além disso, foi realizada uma avaliação da solução proposta, onde foi analisado o retardo total das mensagens para dois cenários, na qual a carga de cada tipo varia. Também foi investigado a influência de alguns parâmetros, do mecanismo de acesso aleatório, no desempenho destas redes. Após a avaliação da solução proposta, os resultados obtidos com o modelo analítico são comparados com resultados obtidos através de simulação. Onde observou-se que o modelo analítico proposto representa o comportamento do protocolo da camada MAC do padrão IEEE 802.16, considerando como métrica de desempenho o retardo total do sistema.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

ANALYTICAL MODELLING AND EVALUATION OF THE DELAY OF THE MESSAGES IN THE PROTOCOL OF ACCESS TO IEEE 802.16 STANDARD

Danielle Lopes Ferreira

September/2008

Advisor: Luís Felipe Magalhães de Moraes

Department: Systems Engineering and Computer Science

The IEEE 802.16 standard came up as a solution to broadband access through wireless media, being developed to transmit data and multimedia applications with different quality of service (QoS) requirements. In order to achieve the QoS requirements of multimedia applications, the IEEE 802.16 standard offers different scheduling schemes. Thus, lots of work found in literature investigates the impact of scheduling mechanisms on the performance of these networks, however these works present the performance evaluations through simulation. Moreover, some works propose analytical models to perform this evaluation, but with some change in the IEEE 802.16 standard media access protocol (MAC). Thus, this work models the total delay of messages for the real and non-real time traffic transmitted by the IEEE 802.16 standard. The total delay of messages is obtained through two distinct models: one for contention phase (disputation by physical means), applied only for the non-real time traffic; and another model for the data allocation phase, where the model support two classes of traffic priority, one for real time service and another for non-real time. Moreover, an evaluation of the presented solution was accomplished, where the total delay of messages was analyzed for two scenaries, which the load of each type varies. Besides, the influence of some parameters was investigated, of the random access mechanism, in performance of the networks. After an evaluation of the presented solution, the results obtained with the analytical model are compared with the simulation results. Where it was observed that the proposed analytical model represents the behavior of the IEEE 802.16 standard MAC layer protocol, as a function of the systems end-to-end delay.

Conteúdo

Lista de Figuras

Lista de Tabelas

Lista de Acrônimos

ATM	:	*Asynchronous Transfer Mode*;
BS	:	*Base Station*;
DL	:	*Downlink*;
DSL	:	*Digital Subscriber Line*;
ETSI	:	*European Telecommunications Standards Institute*;
FDD	:	*Frequency Division Duplex*;
FIFO	:	*First In, First Out*;
FCFS	:	*First Come, First Serverd*;
HOL	:	*Head of the Line*;
IE	:	*Information Element*;
IEEE	:	*Institute of Electrical and Electronic Engineers*;
MAC	:	*Medium Access Control*;
PHY	:	*Physical Layer*;
PMP	:	*Point Multi-Point*;
QoS	:	*Quality of Service*;
SS	:	*Subscriber Station*;
TDD	:	*Time-Division Duplex*;
TDM	:	*Time-Division Multiplexing*;
TDMA	:	*Time-Division Multiple Access*;
UL	:	*Uplink*;
WLAN	:	*Wireless Local Area Network*;
WMAN	:	*Wireless Metropolitan Area Network*;
WPAN	:	*Wireless Personal Area Network*;

Notações do Modelo Analítico

N	:	*Número de estações no sistema*;
BE	:	*Expoente de backoff*;
N_B	:	*Número de tentativas de backoff*;
m	:	*estágio máximo de backoff*;
R	:	*Número de tentativas no estágio máximo de backoff*;
$B(t)$	:	*Processo estocástico que representa o tamanho do contador de tempo de backoff para uma determinada estação*;
$S(t)$	:	*Processo estocástico que representa o estágio de backoff no tempo t*;
τ	:	*Probabilidade estacionária de uma estação transmitir uma mensagem num segmento de tempo aleatório*;
p	:	*Probabilidade de colisão de uma mensagem de requisição de largura de banda*;
P_{tr}	:	*Probabilidade de pelo menos uma estação transmitir durante um segmento escolhido aleatoriamente*;
P_s	:	*Probabilidade de transmissão com sucesso*;
W_{min}	:	*Janela mínima de backoff*;
W_{max}	:	*Janela máxima de backoff*;
i	:	*Estágio de backoff*;
k	:	*Contador de backoff - número de oportunidades de transmissão para o qual uma estação deve esperar antes de iniciar a sua transmissão*;

$\{S(t), B(t)\}$	:	*Processo bi-dimensional representando a cadeia de Markov de tempo discreto do algoritmo de backoff*;
$b_{i,k}$	:	*Distribuição estacionária das probabilidades de transição da cadeia de Markov bi-dimensional*;
D_r	:	*Variável aleatória que representa o retardo médio das mensagens de requisição de largura de banda*;
D_1	:	*Variável aleatória que representa o retardo médio para as mensagens de tempo real*;
D_2	:	*Variável aleatória que representa o retardo médio para as mensagens de tempo não-real*;
N_c	:	*Variável aleatória que representa o número médio de colisões de uma mensagem de requisição de largura de banda*;
T_c	:	*Tempo de duração de uma colisão da mensagem de requisição de largura de banda*;
T_s	:	*Tempo gasto para uma transmissão com sucesso*;
β	:	*Variável aleatória que representa o retardo médio do contador de backoff*;
Φ	:	*Variável aleatória que representa o tempo pelo qual o contador da estação permanece congelado devido a transmissão de outras estações*;
N_q	:	*Variável aleatória que representa o número de vezes que uma estação deve esperar pela oportunidade de transmissão de outras estações antes do seu contador alcançar 0*;
μ	:	*Tamanho do período de contenção*;
λ_1	:	*Taxa de chegada das requisições virtuais de tempo real*;
λ_2	:	*Taxa de chegada das requisições de tempo não-real*;
ν_1	:	*Tempo de serviço médio das requisições virtuais de tempo real*;
ν_2	:	*Tempo de serviço médio das requisições de tempo não-real*;
l	:	*Número de requisições para o tráfego de tempo não-real*;

g : *Número de requisições de tempo real que chegam antes da i-ésima requisição de tempo não-real ser servida*;

S_i : *Variável aleatória que representa o tempo de serviço para a i-ésima requisição para o tráfego de tempo não-real*;

V_j : *Variável aleatória que representa o tempo de serviço para a j-ésima requisição virtual para o tráfego de tempo real*;

t_2 : *Tempo do início do próximo MAP*;

t' : *Tempo de chegada da i-ésima requisição para o tráfego de tempo não-real*;

T_{MAP} : *variável aleatória que representa o tempo médio definido pelo MAP*;

L : *Variável aleatória que representa o número de requisições virtuais para o tráfego de tempo real chegando durante o tempo do próximo MAP*;

G : *Variável aleatória que representa o número de requisições para o tráfego de tempo não-real chegando de* t_1 *até* t_2;

ρ_1 : *Carga oferecida do tráfego de tempo real*;

ρ_2 : *Carga oferecida do tráfego de tempo não-real*;

ρ : *Carga total oferecida*;

δ : *Variável aleatória que representa o retardo do contador de backoff de uma estação antes de acessar o canal em condições ocupadas*;

X : *Variável aleatória*;

$E[X]$: *Média da variável aleatória X*

$E[X^2]$: *Segundo momento da variável aleatória X*

Capítulo 1

Introdução

A RÁPIDA proliferação dos dispositivos móveis (como *laptops*, *handhelds* e PDAs) conduziu a uma mudança revolucionária na área da computação nos últimos anos. As redes sem fio formam atualmente uma grande vertente tecnológica, justificada pela busca de praticidade e acessibilidade aos meios de comunicação. Este capítulo introdutório apresenta os conceitos básicos inerentes às comunicações sem fio. A primeira seção apresenta uma visão geral de redes sem fio através das padronizações existentes para esta tecnologia. A motivação e os objetivos do trabalho são expostos nas próximas seções. Além disso, as principais contribuições alcançadas são apresentadas. Por fim, a estrutura do trabalho está descrita na última seção.

1.1 Redes sem Fio

A implantação da infraestrutura necessária à comunicação sem fio através da instalação de pontos de acesso à rede sem fio, nos quais pode-se acessar a Internet, nos principais grandes centros, são uma grande tendência e vêm apresentando um crescimento extremamente rápido. O elevado aumento no número de dispositivos de computação móveis, como *laptops*, *handhelds* e PDAs, conduziu a uma mudança revolucionária na computação nos últimos anos. A era do computador pessoal (um computador por pessoa) está perdendo espaço para a era da "computação úbiqua", na qual usuários utilizam, simultaneamente, vários aparelhos eletrônicos através dos quais podem acessar todas as informações necessárias a qualquer hora e em qualquer lugar, fazendo com que a comunicação, através de redes sem fio, seja a solução mais simples para os interconectar.

Uma rede sem fio pode oferecer conexão aos serviços fornecidos na Internet além de permitir a conectividade sem fio de estações fixas, móveis e portáteis, dentro de uma determinada região.

Classificação, Padrões e Tecnologias das Redes sem fio

- WPAN - *Wireless Personal Area Network* ou rede pessoal sem fio. Abrange o ambiente que cerca o usuário, com um alcance de aproximadamente 10 metros. É em geral utilizada para interligar dispositivos eletrônicos fisicamente próximos, eliminando os cabos usualmente utilizados para interligar teclados, impressoras, telefones móveis, agendas eletrônicas, computadores de mão, mouses e outros. Alcança taxas de alguns kbps até alguns Mbps. Como exemplo de tecnologia pode-se citar o *bluetooth*, padrão IEEE 802.15 [1], para estabelecer esta comunicação.

- WLAN - *Wireless Local Area Network* ou rede local sem fio. É utilizada para interligar dispositivos sem fio, cujo alcance de transmissão chega a algumas centenas de metros. As taxas de transmissão são da ordem de dezenas de Mbps. Atualmente existem dois padrões bem definidos para redes locais sem

fio: o padrão IEEE 802.11 [2] e o padrão ETSI HIPERLAN [3]. O padrão IEEE 802.11, também conhecido com WiFi, é mais difundido e faz parte da aliança internacional de fabricantes WECA (*Wireless Ethernet Compatibility Alliance*).

- WMAN - *Wireless Metropolitan Area Network* ou redes metropolitanas sem fio. Tem o alcance maior que as WLANs, chegando a dezenas de quilômetros e taxas de transmissão da ordem de centenas de Mbps. Há duas padronizações para as WMANs: o padrão IEEE 802.16 [4] e o padrão ETSI HIPERMAN [5]. O padrão IEEE 802.16 faz parte de uma aliança internacional denominada WiMax [6], tendo como objetivo garantir a interoperabilidade entre os dispositivos de diferentes fabricantes.

Dentre as redes sem fio, a crescente demanda por acesso à Internet com alta velocidade e serviços de multimídia, proporcionou o rápido desenvolvimento do acesso sem fio para WMAN. O chamado *Broadband Wireless Access System* (BWA) surgiu como a "última milha" para acesso à banda-larga e apresenta várias vantagens em relação aos sistemas via cabo e DSL (*Digital Subscriber Line*), dentre as quais: rápida implantação, alta escalabilidade, baixo custo de atualização e manutenção. O padrão IEEE 802.16 [4], surgiu com o intuito de prover um sistema de acesso sem fio de alta velocidade e de alto desempenho, com diferenciação de serviços para tipos de tráfego com diferentes requisitos de qualidade de serviço (*Quality of Service* - QoS).

Para fornecer essa qualidade de serviço, o protocolo de controle de acesso ao meio (MAC) do padrão IEEE 802.16 oferece diferentes formas de compartilhamento do meio sem fio, para os diferente tipos de tráfego. Porém, sabe-se que um dos grandes problemas que surgem nas redes de comunicação sem fio é, como encontrar uma forma econômica e eficiente de compartilhar o recurso mais caro e escasso de uma rede de telecomunicações, o meio de transmissão. O problema neste caso é como controlar o acesso a este canal compartilhado de forma que a faixa de freqüência da transmissão seja dividida eficientemente entre os usuários. A solução mais adequada depende das características do ambiente em questão e dos requisitos a serem atendidos.

A parte mais crítica da camada MAC de uma tecnologia sem fio infraestruturada é o múltiplo acesso no canal de subida (*uplink*). No canal de descida (*downlink*), a estação base tem completo conhecimento da demanda de largura de banda corrente, isto é, das mensagens armazenados no seu *buffer* e está apto a escalonar a transmissão. No *uplink*, a estação base não conhece o conteúdo do *buffer* das estações. Há duas soluções extremas para o múltiplo acesso no *uplink*. A primeira é alocar recursos para cada estação, tendo ou não dados a enviar. A outra possibilidade é garantir recursos para a estação apenas quando ela tiver dados prontos para enviar e requisitar largura de banda explicitamente. A vantagem do primeiro extremo é o pequeno retardo de acesso, a desvantagem é o desperdício do recurso se não houver dados a transmitir na estação. A desvantagem do segundo extremo é o retardo de acesso adicional e o recurso adicional para transmitir requisição de largura de banda, a vantagem é a boa utilização dos recursos. Como será apresentado no próximo capítulo, o protocolo da camada MAC do padrão IEEE 802.16 faz o uso destas duas soluções extremas para alocar recursos para os diferentes tipos de tráfego.

Portanto, a avaliação de desempenho do protocolo da camada MAC do padrão IEEE 802.16 torna-se de extrema importância e grande utilidade. Após a contextualização do presente trabalho serão descritas, nas seções abaixo, a motivação e definição do problema, os objetivos e as contribuições deste trabalho.

1.2 Motivação

A busca de soluções para a avaliação de desempenho das redes do padrão IEEE 802.16, é alvo de esforços da comunidade científica, como identificado na literatura. Diferentes técnicas, metodologias e teorias são conhecidas para avaliação de desempenho de sistemas computacionais. Cada uma apresenta possibilidades, vantagens e limitações, e são aplicáveis em diferentes contextos e com custo de avaliação diverso. A seguir, essas técnicas serão descritas e contextualizadas.

1. Medição - é uma técnica de medição de desempenho de sistemas reais e consiste em monitorar o sistema enquanto ele está sendo submetido a uma carga

em particular. Esta técnica é aplicada em um sistema que está em um estágio de desenvolvimento pós-protótipo. A ferramenta desta técnica é a instrumentação do sistema e apresenta uma precisão variável, além de um alto custo de implementação [7]. Um monitor é uma ferramenta utilizada para observar as atividades de um sistema. Em geral, os monitores coletam resultados de desempenho, produzem estatísticas e apresentam os resultados, além de identificarem áreas com problemas e sugerirem correções. Monitores são utilizados não apenas por analistas de desempenho, mas também por programadores e gerentes de sistemas. Monitoramento é o primeiro passo em medições de desempenho. A técnica de experimentação tem grande importância prática por identificar problemas correntes, como a necessidade de ajustes de parâmetros, além de prever potenciais problemas futuros. A principal vantagem é obter o desempenho do sistema real ao invéz do desempenho do modelo do sistema, pois as interações que afetam o desempenho do sistema real podem ser difíceis de se captar no modelo do sistema. Como desvantagem pode-se citar a necessidade de ter um sistema em execução e de instrumentar o sistema. Além disso, é difícil estimar o tempo gasto para instrumentar, realizar as medidas e modificar o sistema para estudar o efeito das alterações.

2. Modelos de simulação - a simulação pode ser utilizada para avaliar e modelar o desempenho de um sistema computacional e é uma técnica aplicada em qualquer estágio do ciclo de vida de um sistema, o tempo exigido para sua aplicação é médio. Esta técnica utiliza, como ferramenta, linguagens de programação e apresenta uma precisão moderada e um médio custo de implementação [7]. Um simulador é construído a partir de um modelo de desempenho do sistema. Entretanto, modelos de simulação podem falhar e muito tempo pode ser gasto em seu desenvolvimento. Escolher a linguagem é um importante passo no processo de desenvolvimento de um modelo de simulação [7]. Existem quatro opções: linguagens de simulação, linguagens de propósito geral, extensões de linguagens de propósito geral e pacotes de simulação [8]. Cada uma delas apresenta vantagens e desvantagens em relação ao consumo de tempo, facilidades embutidas na linguagem, até a familiaridade do progra-

mador e do analista com a linguagem. A simulação é uma ferramenta versátil, poderosa e extremamente útil na avaliação de desempenho. A sua principal vantagem é possibilitar que os modelos de simulação sejam construídos com níveis arbitrários de detalhes, permitindo simular situações complexas que são analiticamente intratáveis. Como desvantagens pode-se citar a complexidade no desenvolvimento do simulador e o tempo de execução da simulação. Um simulador pode utilizar números aleatórios para gerar variáveis aleatórias, que representam tempos de chegada e de serviços no sistema de acordo com distribuições de probabilidade. Com base nessas variáveis, questões de desempenho podem ser respondidas utilizando-se técnicas estatísticas para fornecer valores estimados. Para avaliar o desempenho de um sistema, uma vez construído o modelo probabilístico, asserções são feitas sobre o processo de chegada e o tempo de serviço de tarefas no sistema, as respostas às questões de desempenho podem, em teoria, ser analiticamente determinadas. Entretanto, na prática, estas questões são muito difíceis de serem determinadas analiticamente e as respostas a elas podem ser realizadas por um estudo de simulação [8]. Além disso, a representação por simulação do funcionamento de um determinado comportamento ou sistema é, na maioria das vezes, uma aproximação dos sistemas reais desenvolvida sinteticamente e que pode ser incompleta devido as simplificações realizadas na representação desse comportamento ou sistema. Assim, a confiabilidade dos resultados obtidos através dessa técnica pode ser, as vezes, variável ou até mesmo questionável.

3. Modelos analíticos - em sistemas computacionais muitas tarefas compartilham recursos tais como CPU, discos e outros dispositivos. Como, normalmente, somente uma tarefa por vez pode utilizar o recurso, todas as outras tarefas ficam esperando em filas por aquele recurso. O conjunto de recursos dá origem a uma rede de filas. Uma das formas mais conhecidas para a construção de modelos analíticos de filas é empregando a teoria de filas [7]. O sistema pode ser descrito por equações matemáticas, cujas soluções consistem na resolução do modelo. A teoria de filas é uma ferramenta matemática empregada para realizar análise de desempenho de um sistema o qual pode ser modelado como

uma rede de filas. Esta teoria ajuda a determinar, por exemplo, o tempo que as tarefas gastaram em várias filas dentro do sistema computacional. Estes tempos podem então ser combinados para predizer o tempo de resposta, o qual corresponde basicamente ao tempo total que a tarefa gastou dentro do sistema, incluindo o tempo de serviço. Além dessa ferramenta matemática pode-se citar outras, como: processos estocásticos, teoria da renovação, teoria dos grafos, geometria analítica, cálculo diferencial e integral, probabilidade e estatística, entre outras. Um modelo analítico pode ser aplicado em qualquer estágio do ciclo de vida de um sistema e o tempo exigido para sua aplicação é pequeno [7]. Essa ferramenta possibilita uma representação do funcionamento e/ou uma análise numérica do problema representado, permitindo assim a realização de uma avaliação de desempenho consistente e rápida. Além disso, quando o sistema a ser representado é muito complexo, essa técnica possibilita uma representação aproximada desse sistema, devido à necessidade de simplificações nessa representação para torná-la numericamente tratável. Geralmente, essas simplificações tornam a utilização da técnica analítica distante do comportamento real do sistema modelado e, portanto, pode ser necessária uma validação dos resultados, obtidos analiticamente, através das outras técnicas: simulação e/ou medição, descritos anteriormente.

Como foi já explicitado, uma modelagem analítica é muito importante para que se tenha uma adequada avaliação do protocolo da camada MAC do padrão IEEE 802.16. Dessa forma, o contexto dessa dissertação é a elaboração de um modelo analítico utilizando a teoria de filas e cadeia de Markov para representar o comportamento do protocolo da camada MAC do padrão IEEE 802.16 em termos do retardo fim-a-fim do sistema.

1.3 Objetivo do Trabalho

O objetivo do presente trabalho é avaliar o desempenho do protocolo MAC do padrão IEEE 802.16 em termos do retardo total. Para tal, será elaborado um modelo

analítico que represente as características do protocolo. Essa avaliação analítica é motivada pelo seguinte:

- A importância de se ter um método de avaliação de desempenho analítico para o protocolo MAC do padrão IEEE 802.16;
- A necessidade de verificar o nível de proximidade ou adequação dos modelos simulados utilizados nas pesquisas comparando-os com resultados obtidos analiticamente.

Para isso, será realizada uma análise, levando em consideração o comportamento dos dois extremos do protocolo de múltiplo acesso proposto pelo padrão IEEE 802.16. Para o esquema de alocação fixa, utiliza-se a teoria de filas e o esquema com acesso aleatório será modelado através de uma cadeia de Markov. Para a alocação dos dados e dos recursos pré-alocados para o canal compartilhado, a modelagem se dará através de um modelo *leaky-bucket* [1] com prioridade.

De forma mais objetiva, esse trabalho busca responder ou dar início a respostas para as seguintes perguntas em aberto:

1. Qual é o retardo fim-a-fim dos usuários de uma rede IEEE 802.16 onde existe a presença de diferentes tipos de tráfego?
2. Qual é o impacto dos diversos parâmetros do mecanismo de acesso aleatório no desempenho dos usuários que fazem uso dessa forma de acesso ao meio e da rede como um todo?
3. Qual o impacto do retardo do escalonamento das mensagens de dados, ou seja, após os recursos serem alocados no retardo total das mensagens da rede?
4. Qual o nível de proximidade ou semelhança do modelo analítico com os resultados de simulação?

Buscando-se alcançar esses objetivos e responder essas perguntas, na próxima seção as contribuições deste trabalho serão descritas.

[1] Sistema de enfileiramento com um único servidor que tem tempo de serviço constante.

1.4 Contribuições do Trabalho

Dentre os principais resultados alcançados com a elaboração deste trabalho, as seguintes contribuições podem ser relacionadas:

- A elaboração de um modelo analítico, que leva em consideração as características do padrão IEEE 802.16 para usuários com diferentes requisitos de serviços para tipos de tráfego que requerem qualidade de serviço distintas;
- A investigação da influência de alguns parâmetros, tais como janela de contenção inicial e máxima e o número de retransmissões no mecanismo de acesso aleatório do protocolo MAC e no desempenho destas redes;
- A análise do retardo médio do escalonamento das mensagens com recursos pré-alocados;
- A avaliação de desempenho do padrão IEEE 802.16 sob a métrica do atraso total das mensagens;
- A validação da modelagem analítica com modelos simulados.

1.5 Organização do Trabalho

Para um melhor entendimento do restante deste trabalho, segue abaixo a estrutura da dissertação, indicando como esta encontra-se organizada.

No capítulo 2, é apresentado uma revisão dos conceitos básicos da teoria necessários ao entendimento desse trabalho. Além disso, são descritos os protocolos de múltiplo acesso e os aspectos da qualidade de serviço em redes sem fio.

No Capítulo 3, uma breve descrição do padrão IEEE 802.16 é apresentada. Além disso, é detalhado o funcionamento das camadas física e MAC, bem como a arquitetura de QoS definida pelo padrão. Finalmente, os trabalhos relacionados ao tema dessa dissertação são contextualizados em relação às redes metropolitanas sem fio.

O Capítulo 4 apresenta uma modelagem analítica para o retardo total das mensagens transmitidas no sub-canal de subida *uplink* para a avaliação de desempenho do padrão IEEE 802.16.

No Capítulo 5, os resultados numéricos obtidos através do modelo analítico descrito no capítulo anterior são apresentados. Além disso, este capítulo apresenta uma comparação entre os resultados obtidos analiticamente e por simulação.

O Capítulo 6 finaliza este trabalho consolidando os resultados apresentados no capítulo anterior através das conclusões e observações relevantes. Além disso, algumas perspectivas para trabalhos futuros são sugeridas.

Capítulo 2

Referencial Teórico

Este capítulo apresenta uma revisão dos conceitos básicos da teoria necessários ao entendimento desse trabalho. São descritos, em linhas gerais, os conceitos básicos dos protocolos de múltiplo acesso e os aspectos da qualidade de serviço em redes sem fio.

2.1 Protocolos de Múltiplo Acesso

Existe a necessidade de utilizar protocolos de múltiplo acesso para coordenar o compartilhamento do meio, quando um recurso é compartilhado por vários usuários independentes. Nesta dissertação, o recurso que deseja-se compartilhar é o canal de comunicação sem fio entre as estações. Nesse tipo de ambiente dinâmico, é necessário encontrar uma forma de compartilhar o canal de maneira adaptativa. Além das questões relacionadas à teoria de filas [9] devido à natureza aleatória das demandas, alocar o canal para um conjunto de demandas geograficamente distribuídas (e possivelmente móveis) é um sério problema e tem um custo associado. Como exemplo, de custo, pode-se citar:

- colisões, devido a um fraco (ou nenhum) controle;
- capacidade de transmissão desperdiçada por causa de um controle muito rígido;
- sobrecarga adicional no tráfego do sistema em função de um controle dinâmico [10].

No intuito de permitir o acesso eficiente ao recurso disponível, existe desperdício devido ao custo de organizar as demandas em algum tipo de fila cooperativa.

Devido ao exposto acima, um grande problema que recai em análise de fila acontece quando clientes (usuários) competem pelo acesso a um recurso limitado. Este é o problema clássico do compartilhamento e da alocação de recursos, que podem ser de vários tipos, como: a capacidade de processamento de uma CPU, a utilização de uma memória compartilhada, a capacidade de armazenamento em disco e, no caso das redes sem fio, o canal de comunicação. A análise e a aplicação da **teoria de sistemas de filas** [9] podem ser utilizadas em várias áreas, inclusive no campo de sistemas de computação.

Muitas questões envolvendo redes de computadores tratam da alocação eficiente do canal de comunicação entre demandas competitivas. Existe uma grande vantagem no tratamento ao compartilhamento de recursos. Por exemplo, há duas soluções para

o acesso ao canal de comunicação: a solução clássica provê um canal dedicado para cada usuário que necessita acessar o meio pelo tempo que for necessário; a outra solução é prover um único canal de alta velocidade para ser compartilhado por um grande número de usuários. Esta vantagem vem da **lei dos grandes números** [9] a qual declara que, com uma alta probabilidade, a demanda em qualquer instante será muito próxima a soma média das demandas daquela população de usuários.

Classificação

Devido a grande variedade de funcionalidades em relação a alocação estática ou dinâmica do canal, do mecanismo de controle centralizado ou distribuído e o comportamento adaptativo do algoritmo de controle existentes nos diversos protocolos de controle de acesso ao meio (*Media Access Control* - MAC) estes protocolos podem ser classificados de acordo com a figura 2.1 [11] [1]. Não há nenhum protocolo que se sobressaia aos outros sob todos os aspectos de desempenho, cada classe tem suas próprias vantagens e desvantagens.

De acordo com o esquema de controle do protocolo de múltiplo acesso, estes protocolos podem ser classificados em três categorias [15]:

- Protocolos de Alocação Fixa (canal ocioso) - Os protocolos de alocação fixa caracterizam-se por atribuir uma parte do canal para cada estação, de maneira fixa. O TDMA (*Time-Division Multiple Access*) e o FDMA (*Frequency-Division Multiple Access*) [13] são exemplos deste tipo de protocolo. São extremamente fáceis de implementar, porém, quando uma estação não tem mensagens para enviar durante o período de tempo em que o meio está alocado para ela, o canal ficará ocioso, ou seja, sem transmissão de dados enquanto outros terminais poderiam utilizá-lo.

- Protocolos de Acesso Aleatório (colisões) - Os protocolos de acesso aleatório não possuem um controle rígido para alocação do canal, também são relativamente simples de implementar, porém, a possibilidade de ocorrer colisões

[1] Para um estudo mais detalhado sobre protocolos de múltiplo acesso, recomenda-se [12, 13, 14].

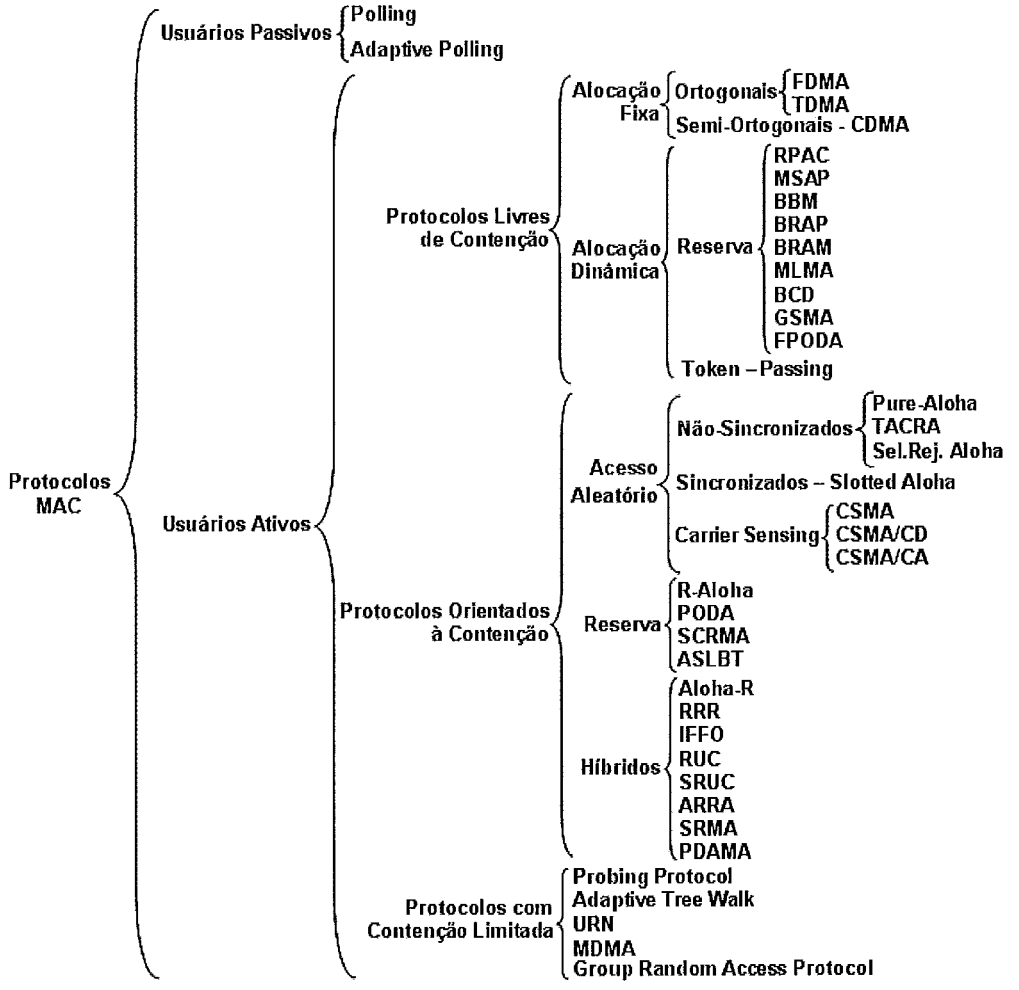


Figura 2.1: Classificação dos protocolos MAC.

quando duas ou mais estações tentam transmitir ao mesmo tempo provoca desperdício no canal de transmissão. Pode-se citar o ALOHA [16] e o CSMA (*Carrier Sense Multiple Access*) [17] como exemplos desta categoria.

- Protocolos de Alocação Dinâmica (sobrecarga) - Nos protocolos de alocação dinâmica o canal é alocado de acordo com as necessidades das estações. Existem algumas formas de implementar este controle, como por exemplo: na forma de *polling* [18] (onde uma estação espera ser "questionada" se necessita acessar o canal), ou na forma de reservas explícitas, como no RPAC (*Reservation-Priority Access Control*) [19]. Com a utilização destes protocolos há uma sobrecarga na rede devido aos sinais de controle. Porém, não existe a possibilidade de colisões e o canal é alocado sob demanda, evitando períodos de

tempo ociosos no canal.

Contudo, existe a possibilidade de combinar alguns destes protocolos para formar métodos de acesso híbridos [13]. Estes esquemas sofrem uma combinação dos custos relacionados ao desperdício do canal.

Uma característica interessante é que alguns protocolos de múltiplo acesso já incorporam mecanismos de escalonamento das mensagens através de prioridades, possibilitando uma diferenciação no tratamento de classes de tráfego distintas. Esta diferenciação é fundamental para transmissão de dados e serviços multimídia com diferentes requisitos de qualidade de serviço (QoS). Dentro desse contexto será investigada, nessa dissertação, esta característica no protocolo MAC do padrão IEEE 802.16 que será apresentado no capítulo 3.

2.2 Qualidade de Serviço

PARA dar suporte à grande diversidade de aplicações disponíveis na Internet, tais como serviços de voz, vídeo, multimídia e transferência de arquivos, garantir a qualidade de serviço (QoS) em redes de computadores tornou-se uma necessidade básica. Para que essa tarefa seja realizada eficientemente, deve-se projetar e implementar um conjunto de mecanismos que incluem policiamento, moldagem do tráfego, controle de admissão e escalonamento. Essa seção apresenta os conceitos básicos inerentes à qualidade de serviço em redes de computadores.

Algumas aplicações são relativamente insensíveis à degradação transitória da qualidade de serviço oferecida pela rede. Por exemplo, num serviço de transferência de arquivos, a redução da largura de banda disponível ou o aumento do atraso dos pacotes podem afetar o desempenho da aplicação, mas não comprometem a sua operação. Devido à capacidade de adaptação e as variações na disponibilidade de recursos, essas aplicações contentam-se com um serviço do tipo melhor esforço, no qual a rede compromete-se apenas em tentar transmitir o tráfego gerado pela aplicação, sem no entanto oferecer garantias de desempenho. Já no caso de aplicações

em tempo real, a diminuição da largura de banda disponível ou o aumento do atraso podem inviabilizar a sua operação. Neste caso, a rede necessita reservar recursos para as aplicações de modo que, mesmo em momentos de maior carga na rede, os requisitos mínimos de desempenho destas aplicações sejam atendidos, ou seja, a rede deve fornecer um serviço com garantias de qualidade de serviço (QoS).

A forma de se obter qualidade de serviço, envolve a inclusão de mecanismos que buscam racionalizar o uso dos recursos disponíveis na rede. Esses mecanismos estabelecem níveis de serviço e permitem a convivência na mesma rede de tráfegos com requisitos distintos de qualidade. Tráfegos pertencentes a níveis de serviço diferentes são tratados de forma que o nível mais prioritário possa sempre dispor dos recursos de que necessita, ainda que em detrimento dos níveis menos prioritários. Ao mesmo tempo, tráfegos pertencentes a um mesmo nível de serviço são tratados de maneira que suas demandas sejam atendidas de forma justa.

Dentre as métricas de QoS mais utilizadas na literatura destacam-se:

- retardo médio - aplicações de tempo real exigem rígidos requisitos de tempo na transmissão dos dados. Longos atrasos tornam estas aplicações menos "realísticas". Contudo, mesmo para aplicações que não possuem estas exigências, pequenos atrasos são sempre melhores;
- *jitter* - variação do atraso. Aplicações com taxa de bit constante são bastante suscetíveis ao *jitter*;
- taxa de perda - algumas aplicações tais como correio eletrônico, transferência de arquivos e transferência de documentos Web etc, necessitam de uma transmissão completamente confiável, ou seja, sem nenhuma perda de dados. Por outro lado, aplicações multimídia como áudio ou vídeo em tempo real, podem suportar um certo nível de perda;
- banda obtida - Da mesma forma, algumas aplicações necessitam de uma quantidade mínima de largura de banda para transmitir dados.

Em [20], os autores identificam quatro princípios básicos para prover qualidade

de serviço em aplicações multimídia:

- Classificação de pacotes - permite a distinção entre pacotes pertencentes a diferentes classes de tráfego e possibilita um tratamento diferenciado para cada pacote. Entretanto, simplesmente classificar os pacotes não garante que eles recebam um serviço com a QoS desejada. A classificação é apenas um mecanismo para distingüi-los;

- Escalonamento e policiamento de tráfego - trata a diferenciação no tratamento dos pacotes de diferentes classes de tráfego. É desejável que haja um grau de isolamento entre fluxos distintos de tráfego, para que um mal comportamento de um determinado fluxo não afete os demais. Se um fluxo específico deve seguir certos critérios (como por exemplo, não exceder alguma taxa pré-estabelecida), um mecanismo de policiamento pode ser empregado para garantir que estes parâmetros sejam observados. Uma outra alternativa para o isolamento do tráfego é a utilização de um escalonador de pacotes. Por exemplo, o escalonador pode alocar uma quantidade fixa da largura de banda do canal para cada fluxo;

- Eficiência - A eficiência de um protocolo de controle de acesso ao meio é uma medida do aproveitamento da largura de banda disponível, sendo normalmente expressa pela razão entre a taxa útil e a capacidade do canal. Um protocolo de controle de acesso ao meio deve procurar maximizar a eficiência sem comprometer a qualidade de serviço oferecida às conexões. Para isso, deve procurar minimizar o *overhead* que introduz.

- Controle de admissão - restringe o número de usuários simultaneamente presentes na rede de forma a evitar a saturação do enlace sem fio e, conseqüentemente, a violação dos requisitos de QoS.

As seções seguintes provêem uma visão geral de vários mecanismos de implementação dos quatro princípios listados acima, considerando as principais disciplinas de filas utilizadas.

2.2.1 Tipos de Tráfego

O primeiro passo para prover diferentes níveis de serviços, para suportar as aplicações multimídia, é através da classificação de pacotes. Os pacotes que trafegam na rede podem ser divididos em três tipos básicos: voz, vídeo e dados. Voz e vídeo são exemplos de tráfego em tempo real, onde os bits são gerados periodicamente, formando um fluxo constante de dados. Se nenhum esquema de compressão é utilizado, este fluxo é chamado de tráfego com taxa constante de bits (*Cosntant Bit Rate* - CBR). Entretanto, esquemas de compressão convertem este tipo de tráfego para uma taxa variável de bits (*Variable Bit Rate* - VBR). Esse tipo de tráfego não suporta grandes variações no atraso (*jitter*) durante as transmissões. Por outro lado, aplicações de dados, que não exigem tempo real, não possuem fortes restrições com relação ao atraso nas transmissões e, além disso, possuem taxa variável de bits (VBR) [21].

A utilização de modelos de tráfego para avaliação de desempenho em redes de computadores é de extrema importância. Em particular, a distribuição de Poisson tem sido bastante utilizada para esta finalidade [9]. Porém, no contexto das aplicações multimídia, onde existe a integração dos tráfegos de dados, voz e vídeo, modelos mais elaborados para caracterizar cada tipo de aplicação são necessários.

2.2.2 Escalonamento de Pacotes

A disciplina de escalonamento diz respeito à política de transmissão de pacotes utilizada. Uma das principais formas de prover qualidade de serviços numa rede de computadores é incorporar disciplinas de escalonamento ao protocolo de acesso ao meio, para que seja obtida a diferenciação de serviços na camada MAC. Dentro desse contexto, serão apresentadas duas importantes disciplinas de escalonamento de pacotes estudadas na literatura: *first-in-first-out* e fila de prioridades.

First-In-First-Out (FIFO)

A disciplina ***First-In, First-Out*** **(FIFO)** é a abordagem mais simples para gerenciar o escalonamento de pacotes, onde todos os pacotes que chegam são colocados em uma fila comum e servidos pela ordem de chegada, como ilustra a Figura 2.2. Quando a fila está cheia, ocorre o descarte de pacotes (*packet loss*). Além disso, não é possível prover diferentes níveis de QoS para fluxos distintos, visto que a disciplina FIFO trata todos os pacotes de maneira igual. Resta ainda comentar que quando um usuário envia pacotes, utilizando essa disciplina, a uma alta taxa, ele ocupa todo o sistema, impedindo outros usuários de acessá-lo - *hogging* [22].

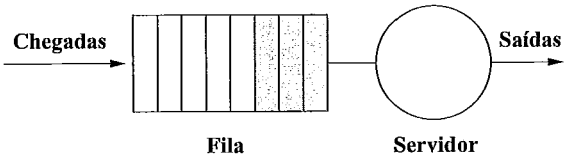


Figura 2.2: Abstração de um fila FIFO.

Fila de Prioridades (HOL - *Head-Of-the-Line*)

Nesta disciplina, uma fila separada é mantida para cada classe e os pacotes que chegam ao sistema são classificados em duas ou mais classes de prioridade, sendo transmitido sempre o pacote que estiver na "cabeça" da fila de maior prioridade, que não se encontra vazia. E os pacotes que pertencem a mesma classe de prioridade, podem ser servidos como uma disciplina FIFO. A figura 2.3, ilustra a fila de prioridades.

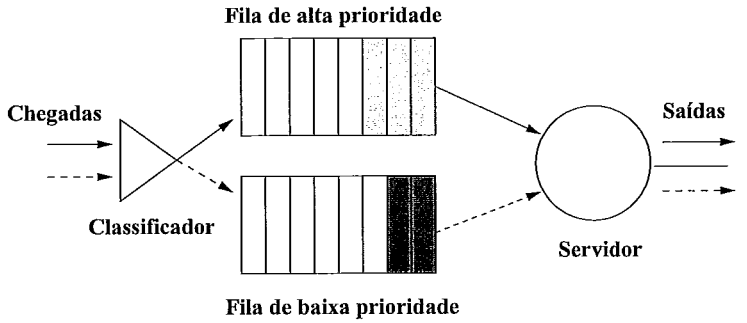


Figura 2.3: Modelo de fila com prioridades.

2.2.3 Controle de Admissão

O controle de admissão tem como principal função decidir adequadamente se um canal de comunicação pode ou não aceitar uma nova conexão. A nova conexão será aceita, se a qualidade de serviço para todas as fontes que compartilham o mesmo canal (incluindo a nova) for satisfeita, caso contrário, essa conexão será rejeitada. Os índices de QoS podem ser expressos em termos de atraso máximo, probabilidade de perda, *jitter* e outras métricas de desempenho.

Cada fonte deve especificar o seu fluxo através de um conjunto de parâmetros conhecidos como **descritores de tráfego** para que este controle determine se a QoS no sistema será mantida e para poder calcular a quantidade de banda que deve ser reservada para cada recurso. Estes descritores devem caracterizar os fluxos de tráfego de maneira compacta e eficiente.

Como exemplo de uma solução bastante utilizada, pode-se citar o conceito de **banda efetiva** [21] de cada fonte, que é um mecanismo de controle de admissão estatístico. O cálculo consiste em alocar uma quantidade de banda entre a taxa média e a taxa de pico de uma determinada fonte. Existe uma grande variedade de mecanismos propostos na literatura para controle de admissão e para um estudo mais detalhado sobre o assunto recomenda-se [23].

2.2.4 Policiamento de Tráfego

Como foi estudado, uma vez que o controle de admissão aceite uma nova conexão, a qualidade de serviço será garantida se a fonte obedecer os descritores de tráfego que são especificados durante o estabelecimento desta conexão. Entretanto, se o fluxo de tráfego violar o "contrato" inicial, a rede poderá não suportar um desempenho aceitável. Assim, para impedir a violação dos contratos estabelecidos, deve existir algum mecanismo de policiamento do tráfego na rede.

O algoritmo balde furado (*leaky bucket*) tem sido muito utilizado como mecanismo de policiamento de tráfego. Através dele, pode-se garantir: a taxa média de

pacotes que um fluxo pode enviar na rede, a taxa de pico para este determinado fluxo e o tamanho máximo da rajada, ou seja, o número máximo de pacotes enviados em um curto período de tempo. O balde furado consiste em uma fila finita. Quando chega um pacote, se houver espaço na fila, ele é incluído na fila, caso contrário, ele é descartado. A cada unidade de tempo, um pacote é transmitido (a menos que a fila esteja vazia). Pode-se ainda, utilizar este mecanismo em conjunto com a disciplina de escalonamento HOL vista anteriormente [20].

2.3 Considerações Finais

Neste capítulo foram introduzidos os principais conceitos relacionados a essa dissertação. Onde foram apresentadas as principais características dos protocolos de múltiplo acesso e os principais conceitos referentes a qualidade de serviço em redes sem fio foram abordados. No próximo capítulo será apresentado o padrão IEEE 802.16 e os trabalhos relacionados ao tema dessa dissertação.

Capítulo 3

Padrão IEEE 802.16 e Trabalhos Relacionados

ESTE capítulo descreve de maneira abrangente os aspectos significativos do padrão IEEE 802.16 e realiza uma sucinta descrição da camada física e MAC, bem como, da arquitetura de QoS especificada pelo padrão. O escopo do padrão é especificar a interface aérea, incluindo a camada de acesso ao meio (MAC) e a camada física (PHY), para redes metropolitanas sem fio (WMAN), com diferenciação de serviços. Além disso, serão apresentados os trabalhos relacionados ao tema dessa dissertação.

3.1 Visão Geral

O padrão IEEE 802.16 [4] especifica uma interface sem fio para redes metropolitanas (WMAN) e vem sendo desenvolvido com a finalidade de padronizar a tecnologia de acesso sem fio à banda larga. O padrão define a interface aérea e o protocolo de acesso ao meio para redes metropolitanas sem fio fornecendo altas taxas de transmissão para o acesso comercial e residencial à Internet.

Para promover e certificar a compatibilidade e interoperabilidade entre os equipamentos de acesso sem fio à banda larga, que estejam em conformidade com o padrão IEEE 802.16, foi formado o WiMAX Forum. O WiMAX (*Worldwide Interoperability for Microwave Access*) foi definido pelo WiMAX Forum como uma tecnologia que possibilite uma alternativa a conexão a cabo ou DSL (*Digital Subscriber Line*). Esse fórum tem como função certificar que os equipamentos industriais e produtos comerciais estejam em conformidade entre si, além de promover o uso desta tecnologia.

A arquitetura de uma rede que utiliza o padrão IEEE 802.16 possui dois elementos principais: Estação Base (*Base Station* - BS) e Estação Cliente (*Subscriber Station* - SS), como mostra a Figura 3.1. A BS é o nó central que coordena toda a comunicação e as SSs se localizam a diferentes distâncias da BS, em uma topologia Ponto-Multiponto (PMP). Além disso, todo o tráfego de dados da rede passa pela BS, ou seja, não existe comunicação direta entre as SSs. A estação base pode estar conectada a uma outra infra-estrutura de rede (como por exemplo, a Internet), possibilitando uma extensão dos serviços oferecidos aos usuários. Da mesma forma, as estações clientes podem oferecer serviços diferenciados para usuários conectados através de uma rede local cabeada, ou sem fio.

O padrão também permite topologia *Mesh* (opcional). A principal diferença entre essas topologias, *Mesh* e PMP, está no fato de que em uma rede PMP o tráfego flui apenas entre a BS e as SSs, enquanto que no modo *Mesh*, o tráfego pode ser roteado através das SSs e pode ocorrer diretamente entre duas SSs, como mostra a figura 3.2. Este trabalho concentra-se nas redes com topologia PMP.

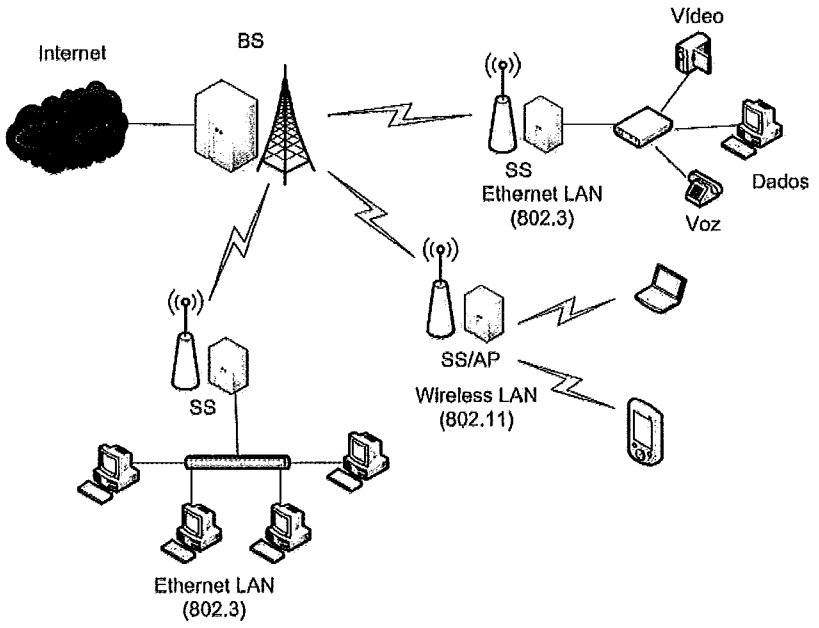
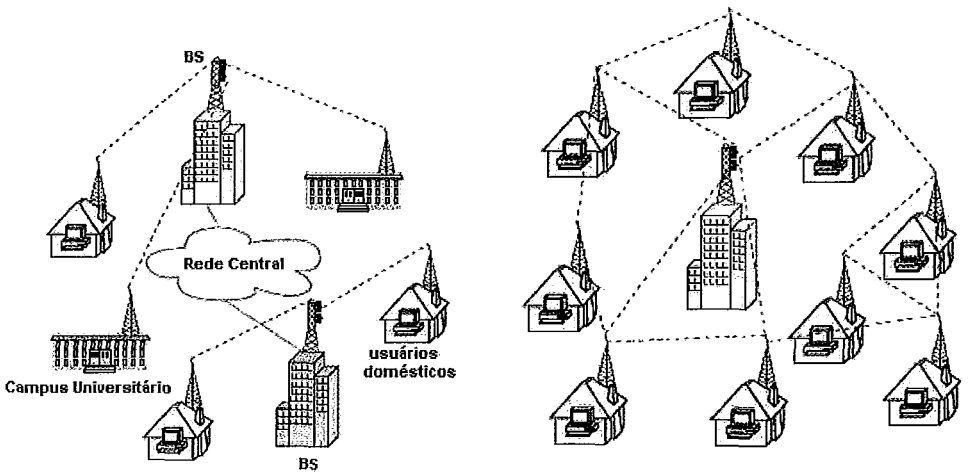


Figura 3.1: Arquitetura básica do sistema BWA.

Esta tecnologia foi desenvolvida para alavancar o acesso sem fio à banda larga em redes metropolitanas (MANs), oferecendo desempenho comparável às tradicionais tecnologias a cabo e DSL. Entretanto, as principais vantagens desta tecnologia são: a habilidade de prover serviços rapidamente, mesmo em áreas de difícil implantação de infra-estrutura; evitar gastos desnecessários com custos de instalações; a capacidade de ultrapassar limites físicos, como paredes ou prédios; alta escalabilidade; baixo custo de atualização e manutenção; dentre outros.

3.2 Camada PHY e MAC

Com o intuito de especificar formalmente as redes sem fio banda larga que cobrissem áreas metropolitanas, em 1999, foi criado pelo *Institute of Eletrical and Eletronics Engineers* o *Broadband Wireless Access* (BWA) *Working Group*, ou IEEE 802.16. Em sua primeira versão, em 2001, o padrão IEEE 802.16 operava em um intervalo de freqüência licenciada entre 10 e 66 GHz, onde era necessário o uso de antenas direcionais para obter desempenho satisfatório. Na área metropolitana, entretanto, não se pode assegurar a operação com linha de visada, devido a presença



(a) Ponto-Multiponto (b) Mesh

Figura 3.2: Topologias permitidas pelo padrão IEEE 802.16

de obstáculos, como: edificações, árvores, etc. Assim, o padrão foi estendido com as versões 802.16a, 802.16b e 802.16c, abordando, respectivamente, problemas relacionados ao espectro de freqüências, a qualidade de serviço e a interoperabilidade do padrão. Outras emendas foram autorizadas posteriormente, o 802.16e fornece suporte a mobilidade, o 802.16f tem o objetivo de melhorar a funcionalidade de múltiplos saltos e o 802.16g prevê melhorar o *handover* e a QoS.

A tecnologia WiMAX pode alcançar, teoricamente uma área de cobertura de 50 Km [24]. As taxas de transmissão de dados vão de 50 à 150 Mbps, dependendo da largura de freqüência do canal e do tipo de modulação [25]. As transmissões ocorrem em dois canais diferentes: um canal de descida (*downlink* - DL), com o fluxo de dados direcionado da BS para as SSs, e outro de subida (*uplink* - UL), com o fluxo de dados direcionado das SSs para a BS. No *downlink*, os dados são transmitidos por difusão, enquanto no *uplink* o meio é compartilhado através de múltiplo acesso. A duplexação entre eles pode ser feita de duas maneiras: duplexação por divisão de freqüências (*Frequency-Division Duplexing* - FDD) e duplexação por divisão do tempo (*Time-Division Duplexing* - TDD). Basicamente, no FDD, os dois canais compartilham o mesmo tempo e os dados são transmitidos em freqüências diferentes, como a transmissão no canal de *downlink* pode ser feita em rajadas, existe suporte

a estações tanto *full-duplex* quanto *half-duplex*. Já no TDD, os canais são divididos no tempo, e utilizam a mesma freqüência. Embora o quadro possua tamanho fixo na duplexação por tempo, a divisão entre os tempos fornecidos para *downlink* e para o *uplink* pode ser desigual. As Figuras 3.3 e 3.4 mostram a estrutura do quadro PHY com TDD e FDD.

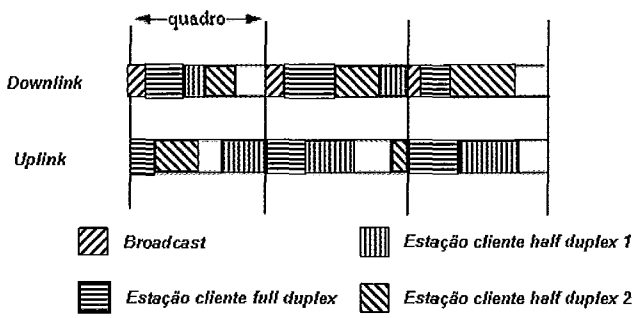


Figura 3.3: Estrutura do Quadro FDD

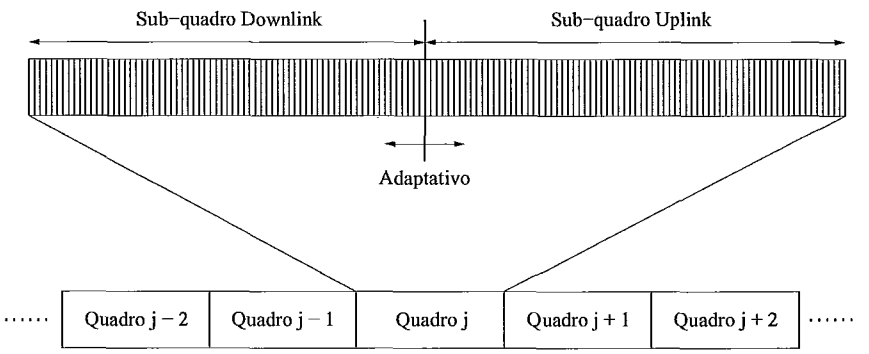


Figura 3.4: Estrutura do Quadro TDD

Como foi visto, na duplexação por divisão de freqüência, FDD, o *downlink* suporta estações cliente *full-duplex* e *half-duplex*, e é possível que estas últimas estejam programadas para transmitir posteriormente no mesmo quadro. Há uma seção de controle de quadro, que possui os DL-MAP e UL-MAP, utilizados para indicar os segmentos físicos nos quais as rajadas devem começar, e uma seção TDMA (*Time-Division Multiple Access*), controlada pelo DL-MAP, que permite a decodificação de somente algumas regiões específicas do sub-quadro, o que é utilizado por estações clientes *half-duplex* que necessitem transmitir antes de receber o *downlink* completo.

Na duplexação por divisão no tempo, durante o *downlink* os pacotes de dados são transmitidos por difusão pela BS para todas as SSs, que por sua vez, capturam apenas os pacotes destinados a elas, portanto, a transmissão é relativamente simples pois somente a BS transmite neste sub-quadro. Durante o *uplink*, através da mensagem UL-MAP no começo de cada quadro, a BS transmite por difusão o número de segmentos que será atribuído para cada SS dentro do sub-quadro. A UL-MAP contém informações específicas (*Information Element* - IE) que incluem as oportunidades de transmissão, ou seja, os segmentos de tempo durante os quais a SS pode transmitir durante o sub-quadro *uplink*. Portanto, após receber a UL-MAP, as estações transmitem os dados em segmentos de tempo pré-definidos como indicados no IE. Na BS, é necessário um módulo de escalonamento do *uplink* para determinar as oportunidades de transmissão (IEs) utilizando as requisições (BW-Request) enviadas pelas SSs. A Figura 3.5 ilustra a estrutura do quadro MAC no esquema de alocação TDD.

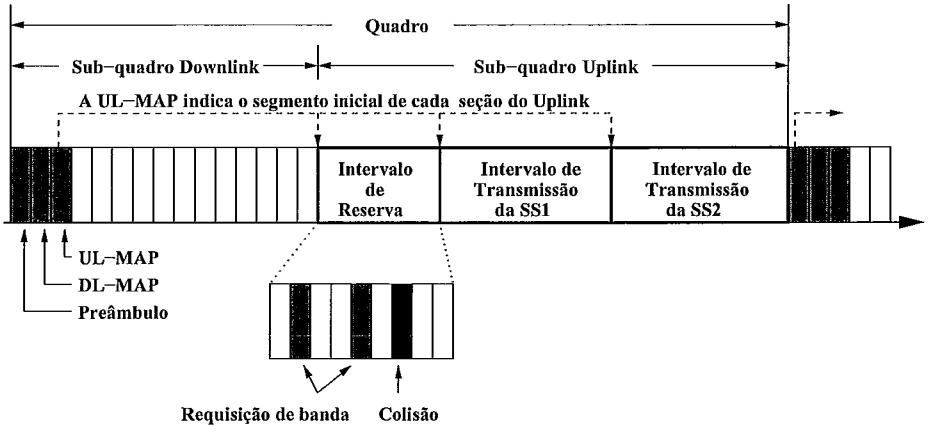


Figura 3.5: Estrutura do quadro MAC no esquema TDD.

Para enviar requisições de oportunidades de transmissão para a BS, as SSs [4] utilizam acesso aleatório e *piggybacking*[1] no sub-quadro de *uplink*. A BS é responsável por estabelecer um intervalo de reserva no início do sub-quadro de *uplink* para que as SSs possam requisitar as oportunidades de transmissões no próximo sub-quadro de *uplink*, ou em algum mais a frente, dependendo da ocorrência ou não de colisões. É importante notar que o IEEE 802.16 utiliza um protocolo de acesso ao meio ba-

[1]Requisições enviadas pelas SSs no final do quadro de dados, transmitidas durante o *uplink*.

seado em alocação dinâmica, onde o período de reserva, que serve para identificar as demandas dos usuários, utiliza acesso aleatório. Depois de enviar a requisição de banda para a BS, a estação aguarda ser escalonada em algum sub-quadro *uplink* mais a frente, como indica a Figura 3.6. Para a resolução de colisões neste intervalo, o padrão define o algoritmo *binary truncated exponential backoff*, onde uma SS detecta a ocorrência de colisão caso a UL-MAP do próximo quadro não contenha nenhuma oportunidade de transmissão destinada a ela.

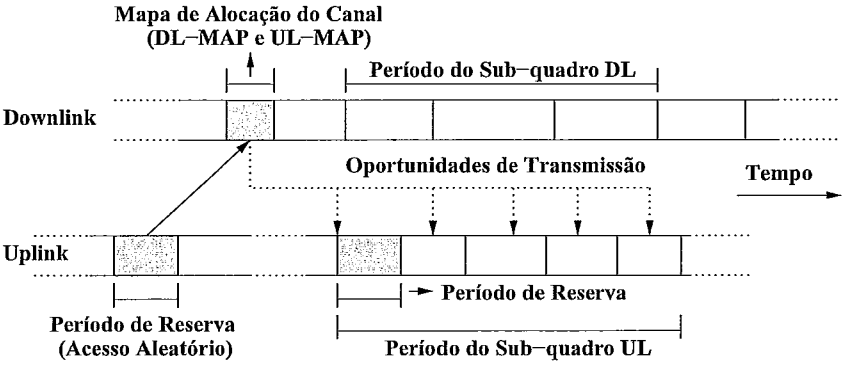


Figura 3.6: Estrutura de alocação do IEEE 802.16.

A camada de acesso ao meio (*Medium Access Control layer - MAC*) é orientada à conexão. Cada conexão é identificada por um identificador (*Connection Identifier - CID*) de 16 bits e cada SS tem um endereço MAC único que a identifica e é utilizado para registrá-la e autenticá-la na rede. Todo tráfego, incluindo o tráfego não orientado à conexão, é mapeado para uma conexão. Além do gerenciamento das conexões, a camada MAC é responsável pelo controle de acesso ao meio e pela alocação de banda.

A alocação dos recursos para as SSs é realizada sob demanda. Quando uma SS precisa de largura de banda para uma conexão, ela envia uma mensagem de requisição para a BS. Uma requisição de banda pode ser enviada como um pacote individual em um *grant* (garantia da oportunidade de transmissão) reservado para esse fim, ou pode ser enviada juntamente com um pacote de dados (*Piggybacking*). A requisição de largura de banda pode ser incremental, indicando a largura de banda adicional que a estação precisa, ou agregada, indicando a largura de banda total requisitada pela SS. Para a SS, as requisições de banda sempre são referentes a uma