

Uma Arquitetura de Escalonamento Hierárquica para Transmissões *Uplink* em Redes WiMAX Baseadas em OFDMA

Eliane Gobetti de Camargo, Cristiano Bonato Both, Rafael Kunst,
Lisandro Zambenedetti Granville, Juergen Rochol¹

¹Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Porto Alegre – RS – Brazil

{egcamargo, cbboth, rkunst, granville, juergen}@inf.ufrgs.br

Abstract. *IEEE 802.16e is an emerging next generation network for providing access with QoS guarantees for both, fixed and mobile users. In the context of mobile applications, OFDMA should be applied as channel multiplexing technique, since it is able to carry data from many users within a transmission symbol. However, the resources allocation in OFDMA frame is not defined by IEEE 802.16e standard. Most of the current researches on OFDMA scheduling are focused in downlink transmissions or, otherwise, do not consider the frame structure of the technology. Therefore, this paper presents an uplink hierarchical scheduling architecture for OFDMA-based WiMAX networks. Moreover, we analyze the performance of the proposed architecture comparing the results with related works.*

Resumo. *A tecnologia IEEE 802.16e é um projeto emergente para redes de próxima geração com suporte à qualidade de serviço para usuários fixos e móveis. No contexto de aplicações móveis, a técnica OFDMA deve ser empregada para multiplexação de canais de rádio frequência, suportando múltiplos usuários em um mesmo símbolo. Entretanto, a alocação de recursos considerando a estrutura do quadro OFDMA não é definida pelo padrão IEEE 802.16e. A maioria das pesquisas atuais sobre arquiteturas de escalonamento baseadas em OFDMA estão focadas na transmissão downlink ou não consideram a estrutura do quadro definida para essa tecnologia. Sendo assim, neste artigo é apresentada uma arquitetura de escalonamento hierárquica para transmissões uplink em redes WiMAX baseadas em OFDMA. Finalmente, analisa-se o desempenho da arquitetura proposta, comparando com trabalhos relacionados.*

1. Introdução

IEEE 802.16e [IEEE 2005], também conhecido como WiMAX (*Worldwide Interoperability for Microwave Access*), é uma tecnologia emergente de próxima geração para redes sem fio, que suporta uma grande quantidade de usuários móveis e nomádicos (fixos) distribuídos através de uma extensa área geográfica. Além disso, esta tecnologia provê garantias estritas de QoS (*Quality of Service*) para aplicações de dados, voz e vídeo. Neste contexto, para garantir QoS, é necessário disponibilizar maiores quantidades de recursos, o que pode ser melhorado pela utilização da técnica de transmissão OFDMA (*Orthogonal Frequency Division Multiple Access*).

A implementação da técnica OFDMA é obrigatória em redes WiMAX móveis. Essa técnica se aplica bem a cenários com mobilidade, pois utiliza subportadoras ortogonais, tornando a tecnologia menos suscetível a ruídos e interferências do meio de transmissão [Berardinelli et al. 2008]. Além disso, a tecnologia permite que múltiplos usuários transmitam em um mesmo símbolo OFDMA, provendo uma melhora na distribuição da largura de banda do canal de rádio frequência. A organização do quadro OFDMA é realizada nos domínios tempo e frequência, dividindo o quadro em *downlink* (DL) e *uplink* (UL). As oportunidades de transmissão, no subquadro UL, utilizam o meio compartilhado e são divididas em duas partes: alocação de rajadas de dados e alocação de rajadas de requisições de banda. A organização de cada região do quadro é responsabilidade do escalonador e deve considerar os recursos físicos disponíveis, bem como os requisitos de QoS das aplicações. O padrão IEEE 802.16e define que uma arquitetura de escalonamento UL deve ser implementada para tratar as oportunidades de transmissão. Entretanto, a estrutura do escalonador e os algoritmos para atendimento das requisições não são definidos.

Este artigo propõe uma arquitetura de escalonamento hierárquica para transmissões UL em redes WiMAX baseadas em OFDMA. A proposta é organizada em dois níveis de escalonamento. No primeiro nível, trata-se o problema da alocação de recursos para requisição de banda e transmissão de dados, considerando a estrutura do subquadro UL. Para isso, é aplicado o algoritmo PQ (*Priority Queuing*). O segundo nível é responsável pelo gerenciamento das filas de requisições de banda recebidas em quadros anteriores. A implementação deste segundo nível combina os conceitos de WFQ (*Weighted Fair Queuing*) e EDF (*Earliest Deadline First*) para garantir os requisitos de QoS das aplicações.

O desempenho da arquitetura proposta é analisado considerando a transmissão de tráfego com requisitos de tempo real e BE (*Best Effort*). Os resultados mostram o atraso médio no atendimento das requisições de banda, o número de requisições atendidas e também o número de requisições descartadas. Finalmente, a partir das investigações realizadas pode-se observar o ganho de desempenho comparado com trabalhos relacionados.

O restante deste artigo está organizado da seguinte forma. Na Seção 2, descreve-se a fundamentação teórica sobre a estrutura do quadro OFDMA, além de apresentar alguns trabalhos relacionados ao escalonamento na tecnologia IEEE 802.16e. Na Seção 3, é apresentada a proposta de uma arquitetura de escalonamento hierárquica para transmissões UL. Na Seção 4, discute-se os resultados obtidos em relação aos trabalhos relacionados. Finalmente, na Seção 5, apresenta-se algumas conclusões e aponta-se direções para futuras investigações.

2. Fundamentação Teórica

O padrão IEEE 802.16e especifica as camadas física e de controle de acesso ao meio da tecnologia WiMAX, que oferece uma alternativa aos sistemas de acesso tradicionais, como por exemplo, *cable modem* e DSL (*Digital Subscriber Line*). Nesta seção são descritas as fundamentações teóricas sobre as características do quadro OFDMA, bem como trabalhos relacionados.

2.1. Estrutura do quadro OFDMA

A estrutura do quadro OFDMA precisa ser considerada em uma arquitetura de escalonamento, pois a composição dinâmica desse quadro é um recurso importante para o desempenho no atendimento de diferentes classes de serviços. Além disso, é importante ressaltar que outras técnicas de multiplexação, como por exemplo OFDM (*Orthogonal Frequency Division Multiplexing*), possuem uma estrutura de quadro diferente.

OFDMA é uma técnica de multiplexação de canal que possui como principal característica o suporte a múltiplos usuários em um mesmo símbolo OFDMA, tal característica é obtida pela divisão do canal sem fio em subcanais. Segundo o padrão IEEE 802.16e [IEEE 2005], os subcanais são formados por conjuntos de subportadoras. Esse processo de multiplexação é conhecido como subcanalização ou zonas de permutação. O padrão especifica 7 métodos de permutação das subportadoras, mas apenas o método PUSC (*Partially Used Subchannelization*) é de implementação obrigatória. Por isso, somente este método é considerado neste trabalho. O processo de subcanalização torna o mecanismo de alocação de banda mais flexível, porque as alocações passam a ser realizadas nos domínios de tempo (símbolos OFDMA) e frequência (subcanais), utilizando uma unidade de alocação mínima chamada *slot*. Na transmissão *downlink*, ou DL PUSC, cada *slot* é composto por 2 símbolos OFDMA e 1 subcanal, sendo que cada subcanal utiliza 24 subportadoras de dados. Já, na transmissão *uplink*, ou UL PUSC, cada *slot* é formado por 3 símbolos OFDMA que utilizam 48 subportadoras de dados, totalizando 16 subportadoras de dados por subcanal. A subcanalização e o mecanismo de alocação de banda determinam a estrutura do quadro OFDMA, que pode ser observada na Figura 1.

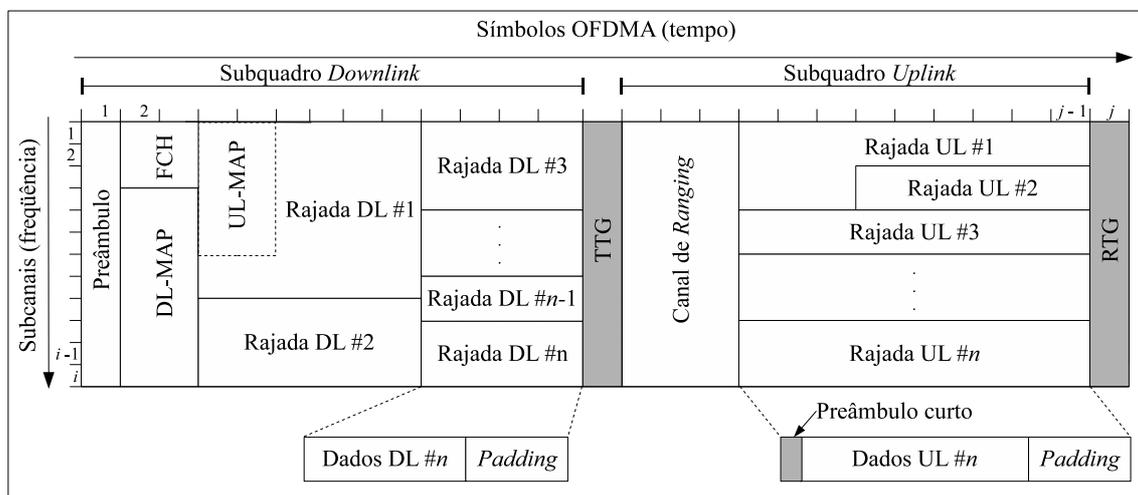


Figura 1. Estrutura do quadro OFDMA

O quadro OFDMA pode ser representado como uma matriz bidimensional, na qual os eixos x e y representam os domínios de tempo e frequência, respectivamente. A construção dos subquadros deve considerar, principalmente, as características de acesso ao meio. No subquadro DL apenas a BS (*Base Station*) transmite informações para todas as SSs (*Subscriber Station*), iniciando com um preâmbulo de sincronização formado por 1 símbolo de duração e transmitido em todos os subcanais. No segundo símbolo inicia a transmissão do FCH (*Frame Control Header*), utilizando 4 subcanais, como pode

ser observado na Figura 1. O FCH contém informações sobre a localização e duração da mensagem DL-MAP. Esse mapa contém informações sobre a localização, duração e configuração AMC (*Adaptive Modulation and Coding*) das rajadas DL. Uma configuração AMC indica o tipo de modulação e a razão de codificação utilizada na transmissão de uma determinada rajada de dados.

Já no subquadro UL, o meio de transmissão é compartilhado por todas as SSs que desejam transmitir informações para a BS. Por isso, o acesso das SSs ao subquadro UL precisa ser coordenado pela BS, através de um mecanismo de concessão de oportunidades de transmissão. Se um subquadro UL for alocado, a primeira rajada DL deve conter um mapa chamado UL-MAP. Esse mapa contém o mesmo tipo de informações que o DL-MAP, mas referente as rajadas UL. Cada rajada UL é formada por informações enviadas por apenas uma SS. Essas informações podem conter dados e nesse caso, a rajada UL é uma oportunidade de transmissão de dados concedida a uma determinada SS. Por outro lado, essa informação pode ser uma requisição de banda e dessa forma, a rajada UL é uma oportunidade de requisição de banda concedida a SS. Além das rajadas, um subquadro UL pode conter um canal de *ranging*, como no exemplo da Figura 1. Este canal, quando alocado no início do subquadro UL, conforme You *et al.* [You et al. 2005], minimiza a interferência na transmissão dos dados.

O canal de *ranging* é compartilhado por todas as SSs dentro da área de cobertura da BS, sendo utilizado para informação de códigos CDMA (*Code Division Multiple Access*). Uma SS informa um código CDMA no canal de *ranging* para alertar a BS que necessita realizar um dos quatro processos de *ranging*: (i) *ranging* inicial, (ii) *ranging* periódico, (iii) *handover* ou (iv) requisição de banda. A BS verifica a qual desses 4 grupos pertence o código CDMA informado e adota o procedimento adequado para atender a solicitação da SS. Por exemplo, se o código informado pertencer ao grupo de requisição de banda, a BS deverá alocar uma oportunidade, em um dos próximos subquadros UL, para que a SS envie seu pedido. Como o canal de *ranging* é compartilhado pelas SSs, podem ocorrer colisões. Por isso, este processo é conhecido como período de contenção. Além da contenção, os pedidos de banda podem ser enviados em oportunidades alocadas pela BS para uma determinada SS, sem a necessidade desta requisitar tal oportunidade e sem a probabilidade de ocorrerem colisões. Esse processo é conhecido como período de *polling* e destina-se aos fluxos de informações com altos requisitos de QoS.

Quando não houver informação suficiente para preencher completamente os subquadros DL ou UL, os mesmos serão preenchidos com *padding*. Além disso, entre cada subquadro existe um TTG (*Transmit/Receive Transition Gap*) alocado para permitir que a BS altere do modo de transmissão para o de recepção. No final do subquadro UL, é alocado um RTG (*Receive/Transmit Transition Gap*) para que a BS retorne ao modo de transmissão.

Considerando o processo de subcanalização e o mecanismo de alocação de banda descritos anteriormente, uma arquitetura de escalonamento UL deve suportar, principalmente, três funções: (i) prover oportunidades de requisição de banda pelas SSs, (ii) gerenciar as filas de requisições de banda para transmissão de dados UL, e (iii) construir os subquadros UL. Na Subseção 2.2, são descritos trabalhos relacionados que propõem soluções para algumas características do escalonamento na tecnologia IEEE 802.16e.

2.2. Trabalhos Relacionados

Pesquisas recentes têm sido realizadas sobre escalonamento em OFDMA para redes WiMAX, abordando parcialmente as funções citadas na Subseção 2.1. Em Lin *et al.* [Lin *et al.* 2008] é proposto o algoritmo HUF (*Highest Urgency First*), projetado para alocar banda considerando um critério de urgência no atendimento de diferentes requisições. Além disso, o algoritmo considera as condições de propagação de um canal sem fio, através da análise da configuração AMC utilizada na transmissão. Jae-Woo So [So 2008], propõe uma arquitetura de escalonamento e, através de um modelo analítico, realiza uma análise de desempenho da sobrecarga de sinalização inserida para serviços VoIP em transmissões DL e UL. Einhaus e Klein [Einhaus e Klein 2006] conceberam um algoritmo de escalonamento básico para o desenvolvimento de um sistema de controle centralizado, considerando os efeitos do OFDMA. Pang *et al.* [Pang *et al.* 2008] apresentam um esquema para alocação dos recursos *uplink* de baixa complexidade. Entretanto, os trabalhos de Pang *et al.* e Einhaus e Klein não consideram a estrutura do quadro OFDMA.

As interfaces físicas previstas para redes WiMAX e a estrutura do quadro OFDMA são descritas por So-In *et al.* [So-In *et al.* 2009]. Os autores descrevem como essas características influenciam na proposta de uma estratégia de escalonamento adequada para a tecnologia. Além disso, são reunidas algumas propostas de mecanismos de escalonamento para o padrão IEEE 802.16e disponíveis na literatura. Da mesma forma, Lei *et al.* [Lei *et al.* 2008] apresentam um algoritmo de escalonamento oportunístico que pode ser adaptado para OFDM ou OFDMA. Porém, o algoritmo proposto precisa ser adaptado de acordo com a interface física empregada.

Outras abordagens propõem o gerenciamento dos recursos apenas para a transmissão DL. O trabalho proposto por Cohen e Katzir [Cohen e Katzir 2008] consiste em um algoritmo que busca resolver o problema do escalonamento DL dividindo-o em dois subproblemas: macro e micro escalonamento. Além disso, Wang e Iversen [Wang *et al.* 2008] sugerem um modelo para alocação dinâmica de recursos e para controle de admissão de conexões.

Analisando as pesquisas atuais, percebe-se que existem questões em aberto, principalmente porque há poucas investigações no sentido de propor uma solução para o problema do escalonamento de recursos em transmissões UL, considerando a estrutura do quadro OFDMA.

3. Proposta de uma Arquitetura de Escalonamento

Uma arquitetura de escalonamento hierárquica para transmissões UL em redes WiMAX baseadas em OFDMA, que atende as funções de escalonamento UL citadas na Subseção 2.1, é apresentada nesta seção. Essa arquitetura é voltada para a topologia PMP (*Point to Multipoint*), utilizando TDD (*Time Division Duplexing*) e deve ser implementada na BS, como pode ser observado na Figura 2. A proposta é composta de dois níveis hierárquicos para dividir o problema do escalonamento em dois subproblemas: (i) alocação dos recursos considerando a estrutura do quadro OFDMA, chamada de *Nível 1*, e (ii) gerenciamento das filas de requisição de banda, chamada de *Nível 2*.

O *Nível 1* aborda o problema da alocação de recursos para requisição de banda e transmissão de dados, considerando a estrutura do subquadro UL. Na Subseção 3.1

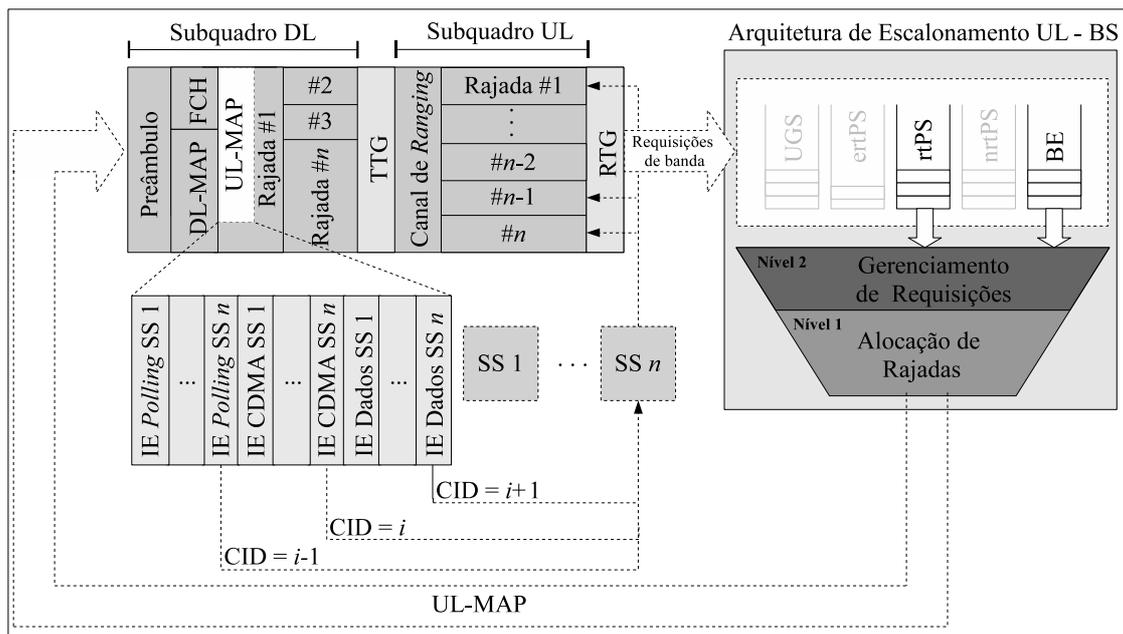


Figura 2. Estrutura do quadro OFDMA

são descritos o mecanismo e o algoritmo utilizados para garantir o funcionamento desse primeiro nível da arquitetura. Por outro lado, o *Nível 2* da arquitetura, apresentado na Subseção 3.2, é responsável pelo gerenciamento das filas de requisições de banda recebidas em quadros anteriores. No padrão IEEE 802.16e são definidas as filas: UGS (*Unsolicited Grant Service*), ertPS (*extended real-time Polling Service*), rtPS (*real-time Polling Service*), nrtPS (*non-real-time Polling Service*) e BE.

O objetivo da arquitetura proposta é a composição do UL-MAP referente ao próximo subquadro UL a ser transmitido. Esse subquadro possuirá as requisições de banda enviadas pelas SSs, que serão inseridas nas filas da BS. O UL-MAP é formado por estruturas de dados do tipo IE (*Information Element*). Na presente proposta, são considerados três tipos de IEs: IE *Polling*, IE CDMA e IE Dados que representam as oportunidades de transmissão concedidas a uma determinada conexão, de acordo com seu CID (*Connection Identifier*).

3.1. Alocação de Rajadas

O *Nível 1* da arquitetura proposta é responsável pela organização das rajadas a serem transmitidas no subquadro UL. Essa organização é realizada utilizando *slots* como unidade de alocação. A capacidade de transmissão de informações de cada *slot* é definida a partir dos símbolos OFDMA transmitidos em cada quadro. O número de símbolos OFDMA em cada quadro (N_{simb}) é calculado considerando a duração total de um símbolo (T_s), e a duração do quadro (T_q) [Both et al. 2008]. Sendo assim, N_{simb} é definido pela expressão 1.

$$N_{simb} = \left\lfloor \frac{T_q}{T_s} \right\rfloor \quad (1)$$

O número de bits transmitidos em cada símbolo OFDMA é obtido a partir da

vazão total da rede (V), calculada pela expressão 2 [Nuaymi e Noun 2006]. O valor de V é obtido em função da largura de banda disponível (B_W), do fator de amostragem empregado (n), do número total de subportadoras (N_{FFT}), da quantidade de subportadoras utilizadas para transmissão de dados (N_{dados}) e do percentual de informações redundantes inserido no símbolo OFDMA pelo prefixo cíclico (G). Além disso, a vazão é afetada pela configuração AMC utilizada, através da variação da razão de codificação do canal (R_{cod}) e da quantidade de bits associada a cada símbolo de modulação (M).

$$V = \left\lfloor \frac{B_W \cdot n}{8000} \right\rfloor \cdot 8000 \cdot \frac{N_{dados}}{N_{FFT}} \cdot \frac{1}{1 + G} \cdot R_{cod} \cdot M \quad (2)$$

A partir do valor de V , dado em bits por segundo, pode-se calcular a quantidade de bits transmitida em um quadro OFDMA (b_q), pela expressão 3, onde T_q é medido em milissegundos.

$$b_q = V \cdot T_q \cdot 10^{-3} \quad (3)$$

Com os valores de b_q e N_{simb} , é possível obter a quantidade de bits transmitida em cada símbolo OFDMA (b_{simb}), conforme a expressão 4.

$$b_{simb} = \left\lfloor \frac{b_q}{N_{simb}} \right\rfloor \quad (4)$$

Um símbolo OFDMA compreende as informações transportadas pelos símbolos discretos de todas as subportadoras. O padrão IEEE 802.16e define que cada *slot* UL é composto por 3 símbolos OFDMA, transmitidos utilizando apenas as subportadoras de 1 subcanal. O número de subcanais utilizados em cada subquadro UL (N_{sc}), é dado pela expressão 5. Nesta expressão, o valor de α depende do método de subcanalização utilizado. No método UL PUSC, considerado neste trabalho, α é igual a 16 subportadoras de dados por subcanal.

$$N_{sc} = \left\lfloor \frac{N_{dados}}{\alpha} \right\rfloor \quad (5)$$

Considerando as expressões 4 e 5, pode-se calcular a capacidade de transmissão do *slot* UL ($C_{slot-ul}$), em bits, através da expressão 6.

$$C_{slot-ul} = \left\lfloor \frac{3 \cdot b_{simb}}{N_{sc}} \right\rfloor \quad (6)$$

A partir de $C_{slot-ul}$, o *Nível 1* da arquitetura proposta pode definir as áreas do subquadro UL utilizadas para requisição de banda e transmissão de rajadas de dados. Porém, as áreas para requisição de banda subdividem-se em: canal de *ranging*, requisições de banda utilizando *polling* e requisições de banda referentes aos códigos CDMA informados no canal de *ranging*, conforme pode ser visto na Figura 2. Os recursos destinados a cada uma dessas regiões são definidos pelo algoritmo de escalonamento PQ, conforme a

seguinte ordem de prioridade: (i) canal de *ranging*, (ii) IE *Polling*, (iii) IE CDMA e (iv) oportunidades para transmissão de dados.

Finalmente, o *Nível 1* deve fornecer ao *Nível 2* a quantidade de *slots* disponíveis para alocação de oportunidades de transmissão de dados ($N_{slot-dados}$), que serão distribuídas entre as SSs. Esse valor é calculado pela expressão 7.

$$N_{slot-dados} = \left\lfloor \frac{b_q}{C_{slot-ul}} \right\rfloor - L_{ranging} - 2 \cdot (L_{polling} + L_{CDMA}) \quad (7)$$

A quantidade de *slots* atribuídos ao canal de *ranging* ($L_{ranging}$) foi fixada em 12, distribuídos em 3 símbolos OFDMA e 12 subcanais. Cada *slot* do canal de *ranging* é capaz de receber a informação de até 30 códigos CDMA diferentes, porém, existe a probabilidade de que mais de uma SS informe o mesmo código utilizando a mesma oportunidade, causando colisões [You et al. 2005]. Na expressão 7, $L_{polling}$ representa o número de conexões transmitindo fluxos de dados baseados em *polling* e L_{CDMA} corresponde ao número de códigos CDMA que estão na fila da BS aguardando por uma oportunidade de requisição de banda. Os valores $L_{polling}$ e L_{CDMA} devem ser multiplicados por 2, tendo em vista que o atendimento de cada requisição é realizado em 2 *slots*: o primeiro contendo um preâmbulo e o segundo, o cabeçalho de requisição de banda.

3.2. Gerenciamento de Requisições

O *Nível 2* da arquitetura proposta foi projetado para definir quais requisições de banda devem ser atendidas no próximo subquadro UL. Essa definição é baseada na classificação dos tráfegos de acordo com as filas de requisição de banda, correspondendo às classes de serviços do padrão IEEE 802.16e. A arquitetura foi definida para suportar uma fila para cada classe de serviço. Neste artigo, considera-se as filas rtPS e BE, que correspondem aos tráfegos com maior e menor requisitos de QoS, dentre as classes de serviço que não possuem largura de banda garantida.

O Gerenciamento de Requisições recebe do *Nível 1* a quantidade de recursos disponíveis para alocação de oportunidades de transmissão de dados e, posteriormente, calcula a quantidade de *slots* necessária para cada transmissão de dados ($N_{slot-req}$). Esse cálculo é realizado considerando a quantidade de bits solicitada a cada requisição de banda (b_{req}), conforme a expressão 8.

$$N_{slot-req} = \left\lceil \frac{b_{req}}{C_{slot-ul}} \right\rceil \quad (8)$$

A seleção de qual fila deve ser atendida, dentre as que estão sendo utilizadas, é baseada na aplicação do algoritmo WFQ. A escolha desse algoritmo prioriza o tráfego rtPS, que demanda mais restrições de QoS em relação as aplicações com tráfego BE, além de evitar postergação infinita desse tráfego. A priorização é realizada pela atribuição de peso 4 para a fila de requisições rtPS, enquanto a fila de requisições BE possui peso 1. Dessa forma, 4 requisições rtPS serão atendidas para cada requisição BE. Esse algoritmo é aplicado enquanto existirem *slots* disponíveis em um subquadro UL para a transmissão de dados pelas SSs, respeitando a inequação 9, onde δ representa o número de requisições que estão aguardando atendimento em determinada fila da BS.

$$\left(\sum_{i=1}^{i=\delta_{rtPS}} \left\lceil \frac{b_{req_i}}{C_{slot-ul}} \right\rceil \right) + \left(\sum_{j=1}^{j=\delta_{BE}} \left\lceil \frac{b_{req_j}}{C_{slot-ul}} \right\rceil \right) \leq N_{slot-dados} \quad (9)$$

Além disso, a priorização das requisições de banda dentro da fila rtPS é obtida pelo emprego do algoritmo EDF, para garantir os requisitos de tempo real dessa classe de serviço. O algoritmo EDF prioriza as requisições pelo estabelecimento de prazos para atendimento (P_a), definidos pela expressão 10, que considera o atraso máximo em que uma requisição rtPS pode ser atendida (A_{max}) e a duração do quadro (T_q).

$$P_a = \frac{A_{max}}{T_q} \quad (10)$$

A cada requisição rtPS é associado um P_a , assim que esta é inserida na fila do escalonador. Após cada T_q em que a requisição permanece nessa fila, o prazo de atendimento é decrementado em uma unidade. Quando $P_a = 0$, a requisição é descartada, pois sua validade foi superada. Por outro lado, na fila BE, as requisições são classificadas utilizando o algoritmo FIFO (*First In First Out*) para garantir justiça no atendimento dos fluxos desse serviço.

Os dois níveis da arquitetura de escalonamento proposta são projetados para cobrir os aspectos que não são definidos pelo padrão IEEE 802.16e para uma transmissão UL sobre um canal sem fio baseado em OFDMA. Na Seção 4, é analisado o desempenho da arquitetura, discutindo aspectos como atraso e número de requisições atendidas.

4. Resultados

O desempenho da arquitetura proposta é analisado em comparação com o algoritmo HUF [Lin et al. 2008] e com o trabalho de Jae-Woo So [So 2008], uma vez que essas abordagens referem-se ao escalonamento UL e consideram aspectos dos dois níveis da arquitetura proposta. Os resultados são obtidos através de simulações implementadas no ambiente de programação Labview, levando em consideração o modelo apresentado na Seção 3. Como estudo de caso, considera-se as classes de serviço rtPS e BE, que representam tráfegos com e sem requisitos de QoS, respectivamente. O tráfego rtPS foi modelado por pacotes VoIP de 160 bytes, utilizando o CODEC G.711, gerados a cada 20ms. Além disso, foi adicionado tráfego de baixa prioridade composto por pacotes BE de 256 bytes, em uma proporção de 50% das conexões, utilizada em todos os cenários analisados, como sugerido por Both *et. al.* [Both et al. 2008]. Os demais parâmetros de simulação são apresentados na Tabela 1.

O primeiro aspecto considerado para análise de desempenho é o percentual de requisições rtPS atendidas por uma BS, em relação ao número de conexões ativas na rede, como pode ser verificado na Figura 3. O gráfico apresenta a diferença no percentual de atendimento das requisições entre a arquitetura proposta e as demais abordagens analisadas. Percebe-se que o percentual de requisições atendidas é maior quando uma pequena quantidade de conexões está ativa. A medida que aumenta o número de conexões, o percentual de requisições atendidas é reduzido. Esse comportamento ocorre porque havendo um maior número de conexões, mais recursos precisam ser destinados às oportunidades

Tabela 1. Parâmetros de Simulação

Parâmetro	Valor
Largura de banda (B_w)	3,5MHz
Taxa de amostragem (n)	8/7
Subportadoras total (N_{FFT})	512
Subportadoras de dados (N_{dados})	360
Subportadoras de dados por subcanal UL (α)	16
Duração do prefixo cíclico (G)	1/16 ($8\mu s$)
Duração do símbolo OFDMA (T_s)	$136\mu s$
Duração do quadro (T_q)	20ms
Configuração AMC	16-QAM 3/4 ($M = 4, R_{Cod} = 3/4$)
Proporção do quadro (DL:UL)	1:1
Atraso máximo para atendimento rtPS (A_{max})	155ms
Intervalo de confiança	95%

de requisição de banda. Conseqüentemente, mais requisições são recebidas pela BS, que necessita priorizar o atendimento, deixando de servir a totalidade das requisições no subquadro seguinte.

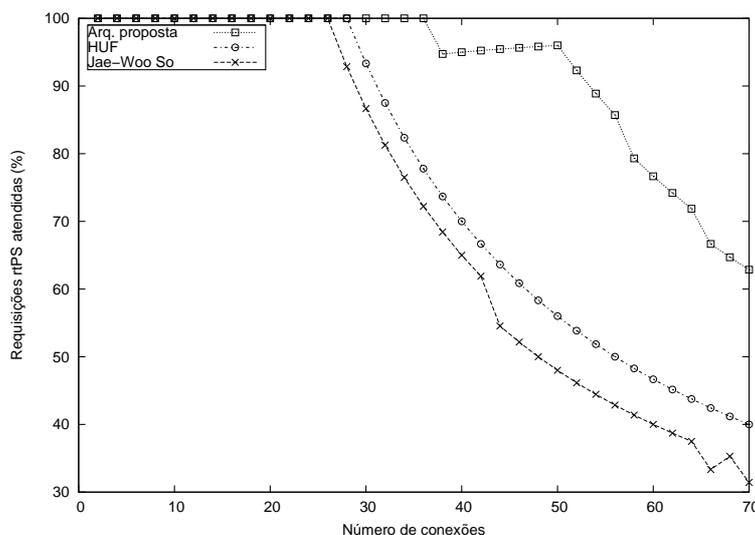


Figura 3. Requisições rtPS atendidas

A arquitetura proposta mantém o desempenho ótimo até aproximadamente 36 conexões, ou seja, todas as requisições rtPS recebidas em um subquadro são atendidas no próximo subquadro. Esse valor representa um ganho de aproximadamente 22% em relação ao algoritmo HUF e cerca de 28% sobre a abordagem de Jae-Woo So. Além disso, a maior vantagem é observada quando há 50 conexões, situação em que 96% das requisições são atendidas na arquitetura proposta, o que representa uma melhora em torno de 42% considerando HUF e 50% em comparação a Jae-Woo So. O pior caso apresentado ocorre quando há 70 conexões ativas. Nesse caso, a arquitetura proposta atende aproximadamente 63% das requisições, apresentando vantagens em torno de 36% e 50%, em relação às propostas HUF e Jae-Woo So, respectivamente.

O número de requisições atendidas é uma métrica importante para avaliação de desempenho. Entretanto, ela deve ser analisada em conjunto com a quantidade de requisições descartadas. O percentual de requisições rtPS descartadas, em relação ao número de conexões ativas, é mostrado na Figura 4. Nesse cenário, não é considerada a abordagem de Jae-Woo So, pois a mesma não define um critério para descarte de requisições. Para as demais abordagens, é utilizado o parâmetro de prazo para atendimento, calculado em função de A_{max} .

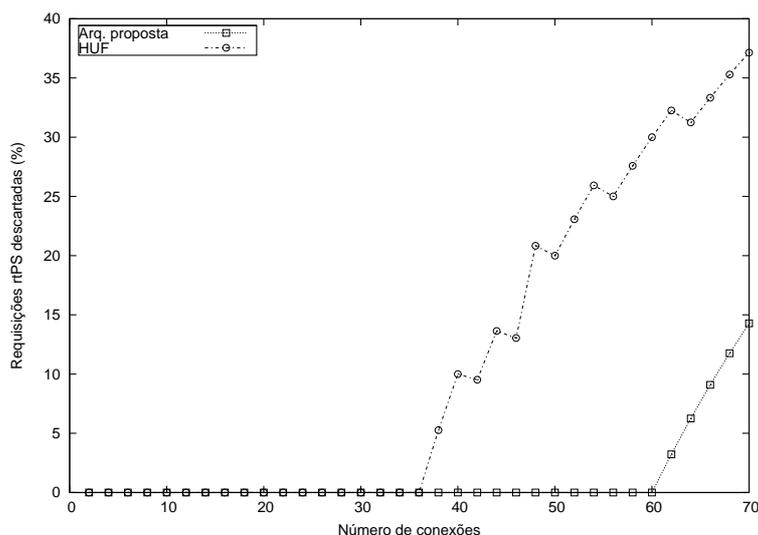


Figura 4. Requisições rtPS descartadas

A Figura 4 mostra que, até 36 conexões, todas as requisições são atendidas, nos dois casos analisados. Após esse ponto, o algoritmo HUF começa a descartar requisições, enquanto a arquitetura proposta só inicia esse processo a partir de 60 conexões, representando um ganho de 40%. O descarte ocorre pelo não atendimento das requisições rtPS dentro do atraso máximo, fixado em 155ms, conforme sugerido por Ali [Ali et al. 2008]. Além disso, percebe-se que o comportamento das curvas, após o primeiro descarte, possui um crescimento acentuado. Por exemplo, quando há 70 conexões, a arquitetura proposta mantém um desempenho cerca de 62% superior ao HUF.

Outra investigação realizada refere-se ao atraso médio no atendimento de requisições rtPS em relação ao número de conexões. Esse comportamento é apresentado na Figura 5. Pode-se observar que até 26 conexões ativas, o atraso é mínimo nos três casos, pois as requisições recebidas são atendidas no quadro seguinte. Após esse ponto, ocorre um crescimento no atraso médio das requisições.

Analisando a Figura 5, percebe-se que a arquitetura Jae-Woo So, após 26 conexões, apresenta um atraso médio sempre crescente. Esse comportamento ocorre devido a ausência de um critério de atraso máximo. Entretanto, nas outras propostas, o atraso médio, tende a convergir para o limite de 155ms, sendo que a proposta deste trabalho converge mais lentamente que o algoritmo HUF. A arquitetura apresentada neste artigo, mantém o atraso mínimo até 50 conexões, o que representa um ganho de aproximadamente 42% em comparação com o desempenho de HUF. Cabe ressaltar que o atraso apresentado no gráfico é inferior ao limite máximo definido, pois trata-se de um valor

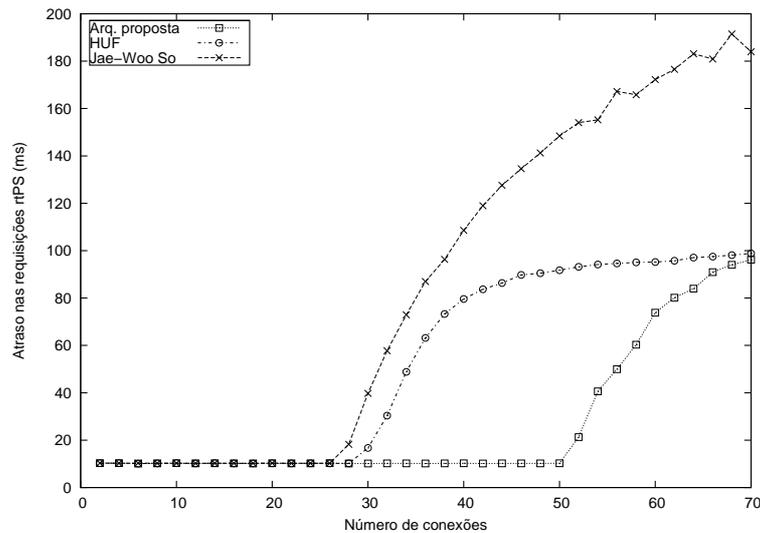


Figura 5. Atraso médio no atendimento das requisições rtPS

médio.

A última investigação realizada diz respeito à comparação do desempenho entre alocações de oportunidades de transmissão de dados das classes rtPS e BE, considerando o número de conexões ativas, como pode ser observado na Figura 6. Nesse caso, o comportamento das alocações para transmissão de dados rtPS é apresentado na Figura 6 (a), enquanto as curvas relativas à classe de serviço BE são mostradas na Figura 6 (b).

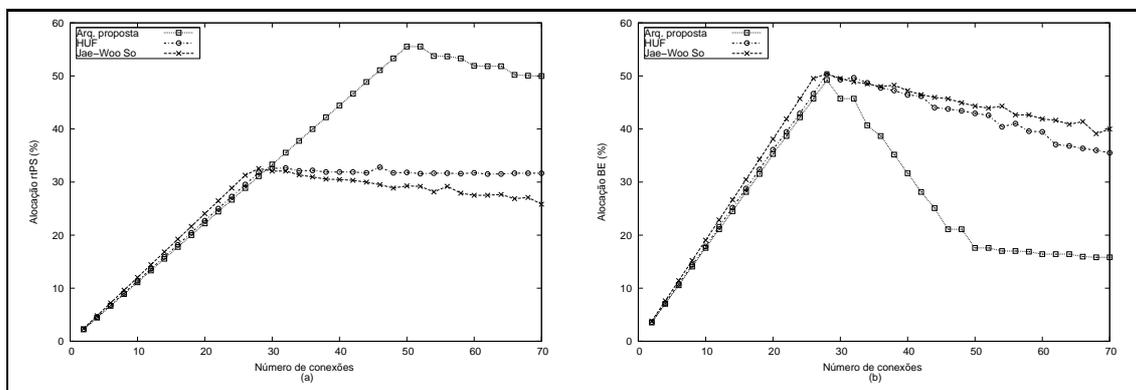


Figura 6. Percentual do subquadro UL alocado para transmissões rtPS e BE

O comportamento das três propostas é semelhante quando o número de conexões é inferior a 30. Nesse caso, observa-se que os recursos disponíveis na rede são suficientes para atender a todas as oportunidades de requisição rtPS e BE no quadro seguinte. Pode-se constatar que a arquitetura proposta prioriza a classe de tráfego de tempo real, diferentemente das demais abordagens. Jae-Woo So não diferencia as classes de serviço, enquanto o algoritmo HUF utiliza como um dos parâmetros para priorização de requisições o tamanho do pacote a ser transmitido. Analisando-se a Figura 6 (a), percebe-se que a arquitetura proposta possui um ganho de até 43% em relação a HUF e 47% se comparado com Jae-Woo So.

Nessas investigações, considerou-se uma largura de banda de 3,5MHz em todos os cenários. Por isso, quando se prioriza uma classe de serviço, ocorre uma penalização da outra classe, de forma proporcional. A arquitetura proposta não prioriza o atendimento de requisições BE, pois essas não possuem requisitos de QoS, diferentemente das demais abordagens, como pode ser visto na Figura 6 (b).

O percentual do subquadro UL alocado para fluxos de dados rtPS e BE é menor do que 100%. Isso ocorre porque parte do subquadro é alocada para canal de *ranging* e oportunidades de requisição de banda. Cabe salientar, que o número de requisições cresce com o aumento do número de conexões. Sendo assim, se um mecanismo de controle de admissão não for implementado para limitar o número de conexões, a tendência é que todo o subquadro UL seja ocupado por oportunidades de requisição de banda, não ocorrendo transmissão de dados.

5. Conclusões e Trabalhos Futuros

Neste artigo, foi proposta uma arquitetura de escalonamento hierárquica para transmissões UL em redes WiMAX baseadas em OFDMA. O principal objetivo da arquitetura é o escalonamento UL, considerando requisições de banda para priorização de tráfego, a fim de prover estritos requisitos de QoS. Sendo assim, foi proposta uma hierarquia composta por dois níveis: o *Nível 1* é responsável pela alocação dos recursos considerando a estrutura do quadro OFDMA, enquanto o *Nível 2* é responsável pelo gerenciamento das filas de requisição de banda.

O desempenho da arquitetura proposta foi analisado, considerando as métricas de atraso médio no atendimento das requisições de banda, número de requisições atendidas e quantidade de requisições descartadas. Os resultados foram comparados com dois trabalhos relacionados, mostrando que o desempenho da arquitetura proposta é superior às demais abordagens considerando tráfego rtPS. Esse desempenho é alcançado porque os critérios utilizados na priorização de tráfego são otimizados para transmissões de tempo real. Além disso, este artigo demonstra a importância de considerar os detalhes da estrutura do quadro OFDMA no processo de escalonamento.

Direções para futuras investigações devem envolver a análise de desempenho da arquitetura proposta considerando as classes de serviço que não foram abordadas neste trabalho. Outra característica importante, que pode ser integrada à arquitetura proposta, é um algoritmo de escalonamento para transmissões DL que aloque as rajadas de dados do subquadro DL. Finalmente, sugere-se a definição de um mecanismo de controle de admissão que leve em conta a alocação de recursos em um quadro OFDMA, com o objetivo de evitar que todo o quadro seja utilizado para requisições de banda.

Agradecimentos

Os autores gostariam de agradecer à Universidade Federal do Rio Grande do Sul (UFRGS) e à Universidade de Santa Cruz do Sul (UNISC) pelo apoio financeiro.

Referências

- Ali, N., Hayajneh, M., and Hassanein, H. (2008). Cross layer scheduling algorithm for IEEE 802.16. In *Proceedings of IEEE International Conference on Communications, 2008. ICC '08*.

- Berardinelli, G., Temino, L. A. M. R. D., Frattasi, S., Rahman, M. I., and Mogensen, P. (2008). OFDMA vs. SC-FDMA: performance comparison in local area IMT-A scenarios. *IEEE Wireless Communications*, 15(5):64–72.
- Both, C. B., de Camargo, E. G., Kunst, R., Granville, L. Z., and Rochol, J. (2008). Analysis of WiMAX bandwidth allocation mechanism considering physical conditions. In *Proceedings of the IEEE International Symposium on Wireless Communication Systems 2008 - ISWCS*.
- Cohen, R. and Katzir, L. (2008). Computational analysis and efficient algorithms for micro and macro OFDMA scheduling. In *Proceedings of the IEEE INFOCOM - Conference on Computer Communications*.
- Einhaus, M. and Klein, O. (2006). Performance evaluation of a basic OFDMA scheduling algorithm for packet data transmissions. In *Proceedings of the 11th IEEE Symposium on Computers and Communications (ISCC)*.
- IEEE (2005). IEEE standard for local and metropolitan area networks - part 16: Air interface for fixed and mobile broadband wireless access systems - IEEE std. 802.16-2005.
- Lei, L., Lin, C., Cai, J., and Shen, X. S. (2008). Flow-level performance of opportunistic OFDM-TDMA and OFDMA networks. *IEEE Transactions on Wireless Communications*, 7(12):5461–5472.
- Lin, Y., Wu, C., Lin, Y., and Lai, Y. (2008). A latency and modulation aware bandwidth allocation algorithm for WiMAX base stations. In *Proceedings of the WCNC - Wireless Communications and Networking Conference*.
- Nuaymi, L. and Noun, Z. (2006). Enabling large-scale wireless broadband: The case for TAPs. In *Proceedings of 17th Annual IEEE Symposium of Personal, Indoor and Mobile Radio Communications*, Finland.
- Pang, D., Hu, J., Zhou, J., Shi, J., and Dutkiewicz, E. (2008). An uplink resource allocation scheme for SDMA-based IEEE 802.16 MIMO-OFDMA systems. In *Proceedings of the IEEE GLOBECOM - Global Communications Conference*.
- So, J.-W. (2008). Performance analysis of VoIP services in the IEEE 802.16e OFDMA system with inband signaling. *IEEE Transactions on Vehicular Technology*.
- So-In, C., Jain, R., and Tamimi, A.-K. (2009). Scheduling in IEEE 802.16e mobile WiMAX networks: key issues and a survey. *IEEE Journal on Selected Areas in Communications*, 27(2):156–171.
- Wang, H., , and Iversen, V. B. (2008). Hierarchical downlink resource management framework for OFDMA based WiMAX systems. In *Proceedings of the WCNC - Wireless Communications and Networking Conference*.
- You, J., Kim, K., and Kim, K. (2005). Capacity evaluation of the OFDMA-CDMA ranging subsystem in IEEE 802.16-2004. In *Proceedings of the IEEE International Conference on Wireless And Mobile Computing, Networking And Communications - WiMob*.