

Proposta e Avaliação de um Novo Protocolo de Acesso ao Meio para *Broadband Wireless Access*

Luís Felipe Magalhães de Moraes e Paulo Ditarso Maciel Júnior

Resumo— Os Sistemas BWA (*Broadband Wireless Access*) surgiram como uma solução para o acesso à banda-larga através de rede sem fio. Estes sistemas foram desenvolvidos para transmitir dados e serviços multimídia com diferentes requisitos de qualidade de serviço (QoS). O IEEE 802.16 especifica a camada PHY e MAC para sistemas BWA. Porém, o padrão prevê apenas o suporte a QoS e não define como escalonar os diferentes tipos de tráfego. Neste artigo é proposto um novo protocolo MAC para BWA que incorpora um mecanismo de escalonamento de tráfego com prioridades baseadas em mensagens e/ou em estações. Além disso, um modelo analítico para o tempo de espera das mensagens é apresentado e, através deste modelo, alguns resultados numéricos são obtidos.

Palavras-Chave— Redes sem fio, protocolo de múltiplo acesso, Qualidade de Serviço (QoS).

Abstract— The *Broadband Wireless Access Systems (BWA)* appeared as a solution for broadband access through wireless network. This systems was developed to transmit data and multimedia services with distincts quality-of-service (QoS) requirements. IEEE 802.16 Standard specifies the PHY and MAC layers for BWA systems. However, the standard defines only QoS suport but not how to schedule different types of traffic. This article propose a new MAC protocol for BWA systems that incorporates a traffic scheduling mecanism based on message and/or station priorities. Moreover, an analitic model of the messages waiting time and some numeric results are presented.

Keywords— Wireless Networks, multiple access protocol, Quality-of-Service (QoS).

I. INTRODUÇÃO

O padrão IEEE 802.16 [1] surgiu como solução para o acesso sem fio à banda-larga (o chamado *Broadband Wireless Access System - BWA*) 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*).

Ao mesmo tempo que o padrão IEEE 802.16 suporta apenas mecanismos para prover QoS, ele não inclui uma solução completa para fornecer garantias a várias aplicações e não sugere como escalonar eficientemente o tráfego para satisfazer tais requisitos. Diversos mecanismos de escalonamento de tipos diferentes de tráfego têm sido propostos na literatura [2]-[6], já que o 802.16 fornece apenas os mecanismos de sinalização, e não especifica nenhum algoritmo de escalonamento e nem de controle de admissão. Porém, estas propostas abordam apenas algoritmos para o escalonamento do tráfego e poucos trabalhos têm sido apresentados envolvendo alterações no protocolo de acesso ao meio, para que o tráfego escalonado possa ser eficientemente escoado. Dentro desse contexto, este artigo apresenta um novo protocolo MAC para Sistemas BWA que contém um mecanismo de escalonamento de tráfego baseado em regras de prioridade e que pode ser facilmente implementado no padrão IEEE 802.16 [1].

O artigo está organizado da seguinte maneira. A Seção II apresenta uma breve descrição do padrão IEEE 802.16. Na Seção III o novo protocolo de acesso ao meio é proposto. Uma modelagem analítica

Luís Felipe Magalhães de Moraes e Paulo Ditarso Maciel Júnior, Laboratório de Redes de Alta Velocidade (RAVEL), Programa de Engenharia de Sistemas e Computação - COPPE/UF RJ, Caixa Postal: 68.511 - 21941-972 - Rio de Janeiro, RJ. E-mails: {moraes,paulo}@ravel.ufjr.br. Este trabalho foi financiado pelo FINEP (1956/04).

para o tempo de espera das mensagens está descrita na Seção IV. Em seguida, na Seção V, alguns resultados numéricos são apresentados. Por fim, conclusões e trabalhos futuros estão presentes na Seção VI.

II. PADRÃO IEEE 802.16

Nesta seção será apresentada uma sucinta descrição da camada MAC e da arquitetura de QoS especificada pelo padrão IEEE 802.16. Como será visto, o 802.16 fornece apenas o suporte para a implementação dessa arquitetura, possibilitando que cada fabricante introduza a sua própria solução.

A. Camada MAC

A arquitetura básica consiste de uma estação base (*Base Station - BS*) e uma ou mais estações cliente (*Subscriber Station - SS*). A BS é o nó central que coordena toda a comunicação e as SSs se localizam a diferentes distâncias da BS, e todo o tráfego de dados da rede passa pela BS, ou seja, não existe comunicação direta entre as SSs. As transmissões ocorrem em dois canais: 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 DL, os dados são transmitidos por difusão, enquanto no UL o meio é compartilhado através de múltiplo acesso.

Durante o DL a transmissão é relativamente simples pois somente a BS transmite neste sub-quadro. Para o UL, a BS determina o número de segmentos que será atribuído para cada SS dentro do sub-quadro. Esta informação é transmitida por difusão pela BS através da mensagem UL-MAP no começo de cada 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 UL. Após receber a UL-MAP, as estações transmitem os dados em segmentos de tempo pré-definidos como indicados no IE. É necessário um módulo de escalonamento do UL na BS para determinar as oportunidades de transmissão (IEs) utilizando as requisições de banda (BW-Request) enviadas pelas SSs. A Figura 1 ilustra a estrutura do quadro MAC.

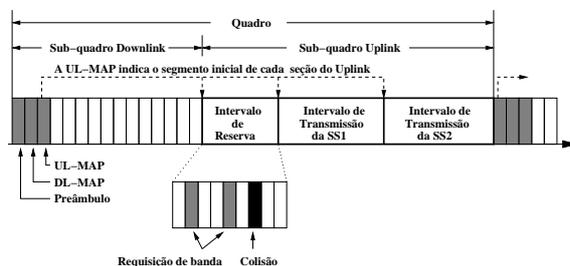


Fig. 1. Estrutura do quadro MAC.

O IEEE 802.16 [1] utiliza acesso aleatório e *piggybacking*¹ no sub-quadro UL para enviar requisições de oportunidades de transmissão para a BS. Esta é responsável por estabelecer um intervalo de reserva no início do UL para que as SSs possam requisitar as oportunidades de transmissões no próximo sub-quadro UL (ou em algum mais a

¹Requisições enviadas pelas SSs no final do quadro de dados, que são transmitidos durante o UL.

frente, dependendo da ocorrência ou não de colisões). O padrão define o algoritmo *binary truncated exponential backoff* para resolução de colisões neste intervalo. Uma SS detecta a ocorrência de colisão caso a UL-MAP do próximo quadro não contenha nenhuma oportunidade de transmissão destinada para ela. Porém, o 802.16 define apenas os mecanismos para sinalização de QoS, tais como BW-Request e UL-MAP, mas não define o escalonador de UL, ou seja, o mecanismo que determina as IEs na UL-MAP. Uma outra característica do padrão é o suporte a requisição de oportunidades de transmissão baseada em conexão (*Grants per Connection - GPC*) ou por estação (*Grants per Subscriber Station - GPSS*). Na GPSS, a estação requisita oportunidades de transmissão como um pacote para todos os serviços que ela mantém, e esta SS é responsável por alocar as oportunidades recebidas entre os diferentes tipos de fluxos.

B. Arquitetura de QoS

O IEEE 802.16 suporta vários tipos de tráfego (dados, voz, vídeo) com diferentes requisitos de QoS. Dentro desse contexto, a camada MAC define mecanismos de sinalização de QoS e funções para controlar a transmissão de dados entre a BS e as SSs. Além disso, o padrão define quatro tipos de fluxos de dados [1], cada um com diferentes requisitos de QoS e uma correspondente política para o escalonador do UL:

- 1) *Unsolicited Grant Service (UGS)*: estas aplicações requerem uma constante alocação de banda.
- 2) *Real-Time Polling Service (rtPS)*: estas aplicações possuem requisitos específicos de banda, bem como um máximo atraso tolerável.
- 3) *Non-Real-Time Polling Service (nrtPS)*: estas aplicações são insensíveis ao atraso no tempo e requerem um mínimo de alocação de banda.
- 4) *Best Effort Service (BE)*: estas aplicações recebem banda disponível após a alocação dos três fluxos anteriores.

A Figura 2 exibe a arquitetura de QoS existente no 802.16. O módulo de escalonamento de pacotes do UL (*Uplink Packet Scheduling - UPS*) encontra-se na BS e controla todas as transmissões de pacotes no UL. Como o protocolo é orientado à conexão, a aplicação deve estabelecer uma conexão entre a BS e o fluxo de serviço associado (UGS, rtPS, nrtPS ou BE). A BS identifica as conexões com um CID (*Connection ID*) único para cada uma. O 802.16 define o processo de sinalização para o estabelecimento de uma conexão (*Connection Request, Response*) entre SS e BS, mas não especifica o processo de controle de admissão.

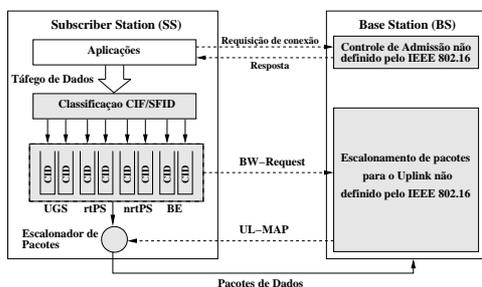


Fig. 2. Arquitetura de QoS do IEEE 802.16.

Em resumo, o IEEE 802.16 especifica: o mecanismo de sinalização para troca de informações entre a BS e as SSs, como a configuração de conexões, BW-Request e UL-MAP; e o escalonamento do UL para tráfego UGS. O padrão não define: o escalonamento do UL para serviços rtPS, nrtPS e BE; controle de admissão e o policiamento do tráfego.

III. PROTOCOLO PROPOSTO

O protocolo proposto utiliza o método de acesso RPAC (*Reservation-Priority Access Control*) descrito em [7], onde o período de reserva é governado pelo esquema TDMA com um segmento reservado para cada estação na rede de forma fixa e, após este período, as estações transmitem suas mensagens de acordo com as regras de prioridade estabelecidas. Porém, diferente do RPAC, estas requisições são processadas de maneira centralizada pela BS, que coordena o acesso ao canal de transmissão. Outra característica relevante desse protocolo é a incorporação de um escalonador de tráfego que utiliza regras de prioridade, possibilitando assim, o suporte à QoS através da diferenciação de serviços baseado em mensagens ou em estações, como será visto adiante.

Após a apresentação das características do protocolo, será descrito em detalhes o seu funcionamento. Considera-se um sistema em que o canal de comunicação é totalmente sincronizado pela BS através da segmentação do tempo em intervalos de duração fixa de τ segundos, onde as transmissões só ocorrem no começo de cada segmento. No intervalo de reserva é atribuído um segmento de tempo a cada estação para requisição de banda. A estação envia para a BS no seu segmento pré-alocado todas as informações necessárias para transmitir as mensagens que estão armazenadas no seu *buffer*.

A estrutura do quadro MAC para o protocolo proposto está ilustrada na Figura 3. Diferente do padrão 802.16, propõem-se um canal onde os quadros não possuem tamanho fixo. Basicamente, o tamanho do quadro depende da quantidade de pacotes que chegaram no quadro anterior. No período de reserva, que localiza-se no final do sub-quadro UL e não no início como no 802.16, a estação informa para a BS o tipo dos serviços para os quais está requisitando banda e a quantidade de pacotes que chegaram para cada serviço. Após processar todas as requisições, a BS envia, no sub-quadro DL do próximo quadro, a UL-MAP com as oportunidades de transmissões para todas as estações.

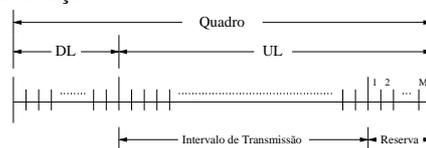


Fig. 3. Estrutura do quadro MAC para o protocolo proposto.

Para efeito de análise do protocolo, a atividade no canal pode ser vista como uma seqüência de intervalos de reserva, *downlink* e de transmissão onde cada seqüência constitui um **ciclo de transmissão**, como ilustrado na Figura 4. É importante notar que existe uma diferença entre o quadro MAC e o ciclo de transmissão, apesar de ambos terem um tamanho igual pois o intervalo de reserva é fixo.

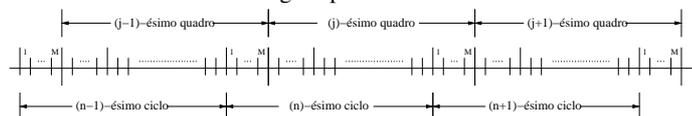


Fig. 4. Ciclos consecutivos de transmissão.

Na realidade, o n -ésimo ciclo é formado pelo intervalo de reserva do $(j - 1)$ -ésimo quadro, mais o intervalo de *downlink* e o intervalo de transmissão (parte do intervalo de *uplink*) do j -ésimo quadro. Com esta definição dos ciclos de transmissões pode-se utilizar a abordagem analítica descrita em [7] para análise do tempo de espera das mensagens, como será visto na Seção IV. Pela figura, pode-se definir L_n^R como o tamanho do intervalo de reserva do n -ésimo ciclo, L_n^{DL} como o tamanho do *downlink* do n -ésimo ciclo e L_n^T como o tamanho do n -ésimo intervalo de transmissão, todos medidos em segmentos. Assim, $L_n = L_n^R + L_n^{DL} + L_n^T$ é o tamanho total do n -ésimo ciclo. Seguindo o protocolo TDMA, cada intervalo de reserva

é composto de M segmentos ($L_n^R = M; n = 1, 2, 3, \dots$), onde M é o número de estações na rede.

Propõem-se um protocolo de acesso ao meio com prioridades baseadas em mensagens ou em estações, em conformidade com o protocolo 802.16 que utiliza admissões GPC ou GPSS. Assim, de acordo com as regras de prioridades utilizadas para determinar a ordem na qual as mensagens devem ser transmitidas durante o intervalo de transmissão, as seguintes versões foram consideradas:

- **Versão I**, sob a qual, para quaisquer $p, q \in \{1, \dots, P\}$ tal que $p < q$, todas as mensagens de classe p são transmitidas antes de qualquer mensagem de classe q , independente da estação a qual pertença. Para mensagens pertencentes a mesma classe mas em estações distintas, a ordem de transmissão é de acordo com a ordem na qual as estações acessam o canal (primeiro a estação 1 e por último a estação M). Para mensagens na mesma estação com a mesma classe de prioridade, as transmissões ocorrem por ordem de chegada.
- **Versão II**, sob a qual, para quaisquer $i, j \in \{1, \dots, M\}$ tal que $i < j$, todas as mensagens no terminal i são transmitidas antes de qualquer mensagem no terminal j , independente da sua classe de prioridade. Em qualquer terminal, as mensagens são transmitidas de acordo com suas prioridades e em ordem de chegada caso elas pertençam a mesma classe, ou seja, em cada terminal, a disciplina de prioridade HOL (*Head-of-Line*) é aplicada com a maior prioridade atribuída para a classe 1 e a menor para a classe P .

O comportamento do canal de acordo com as versões I e II do protocolo proposto está ilustrado nas Figuras 5 e 6 respectivamente.

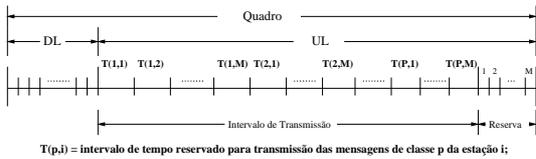


Fig. 5. Comportamento da Versão I do protocolo proposto.

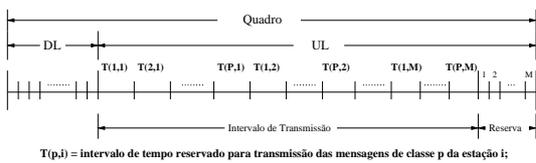


Fig. 6. Comportamento da Versão II do protocolo proposto.

IV. MODELAGEM ANALÍTICA

A técnica utilizada para obter o tempo médio de espera para as mensagens é similar ao método empregado em [7] e [8]. Primeiramente, será detalhada a análise para a Versão I e em seguida, uma adaptação do resultado obtido para a Versão II é apresentada.

A. Modelagem da Versão I

Considerando um sistema com uma BS e M ($M \geq 1$) estações clientes (SSs), todas já associadas com a estação base, onde cada estação possui *buffer* infinito. O canal de transmissão tem uma taxa de C bits/s é considerado sem erro. As mensagens geradas em cada estação são compostas de unidades fixas de dados (pacotes) e o tempo de transmissão de cada pacote é igual a um segmento de tempo (τ). As mensagens são compostas por um número aleatório de pacotes, onde cada pacote contém μ^{-1} bits. Como o tempo de transmissão de um pacote é considerado igual ao tempo de um segmento, temos que $\tau = (\mu C)^{-1}$ (exatamente um segmento para transmitir um pacote).

As mensagens que chegam em cada estação pertencem a uma das P diferentes classes. Assume-se que a mensagem da classe 1 possui

uma maior prioridade e a mensagem da classe P possui a menor prioridade. Em cada estação, a chegada de mensagens é caracterizada por um processo de Poisson tal que, λ_i^p (mensagens por segmento) é a taxa média de chegada das mensagens de classe p na estação i . O número de pacotes que compõem a m -ésima mensagem de classe p é denotado por $B_{i,m}^p$ ($i = 1, 2, \dots, M; p = 1, 2, \dots, P$). Assumindo que $\{B_{i,m}^p; m \geq 1\}$ é uma seqüência de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) para cada p com média b_i^p e segundo momento $b_{2,i}^p$. Como definido na Seção III, $L_n = L_n^R + L_n^{DL} + L_n^T$ é o tamanho do n -ésimo ciclo e assume-se que $L_n^R = M$, com duração de $M\tau$ segundos.

$W_{i,m}^p$ representa o tempo de espera (medido em segmentos) para a m -ésima mensagem de classe p chegando na estação i . O número de mensagens de classe p armazenadas no *buffer* da estação i no começo do n -ésimo ciclo é representado por $Z_{i,n}^p$. $N_{i,k,n}^p$ é o número de mensagens de classe p chegando no terminal i durante o k -ésimo segmento do n -ésimo ciclo; e $B_{i,m,n}^p$ é o número de pacotes contidos na m -ésima mensagem de classe p transmitida pelo terminal i durante o n -ésimo ciclo. A Cadeia de Markov vetorial $\underline{Y}_i^p = \{Y_{i,n}^p; n \geq 1\}$, representa o processo de estado do canal onde $\underline{Y}_{i,n}^p = \{L_n, N_{i,k,n}^p, B_{i,m,n}^p\}$. Para análise do tempo de espera das mensagens define-se as seguintes funções de $\underline{Y}_{i,n}^p$:

- $N_i^p(\underline{Y}_{i,n}^p)$ = ao número total de mensagens de classe p transmitidas na estação i , durante o $(n+1)$ -ésimo ciclo, dado $\underline{Y}_{i,n}^p$;
- $W_i^p(\underline{Y}_{i,n}^p)$ = a soma das componentes do tempo de espera de todas as mensagens de classe p servidas no terminal i no $(n+1)$ -ésimo ciclo, dado $\underline{Y}_{i,n}^p$.

A partir das definições acima, têm-se o cálculo de $N_i^p(\underline{Y}_{i,n}^p)$ e $W_i^p(\underline{Y}_{i,n}^p)$:

$$N_i^p(\underline{Y}_{i,n}^p) = Z_{i,n+1}^p; \quad i = 1, 2, \dots, M; p = 1, 2, \dots, P; \quad (1)$$

$$\begin{aligned} e \quad W_i^p(\underline{Y}_{i,n}^p) &= \sum_{k=1}^{L_n-1} N_{i,k,n}^p (L_n - k) + (M + L_n^{DL}) Z_{i,n+1}^p + \\ &+ \left(\sum_{j=1}^{p-1} \sum_{g=1}^M \sum_{k=1}^{Z_{g,n+1}^j} B_{g,k,n+1}^j \right. \\ &+ \left. \sum_{g=1}^{i-1} \sum_{k=1}^{Z_{g,n+1}^p} B_{g,k,n+1}^p \right) Z_{i,n+1}^p + \\ &+ \left(\sum_{k=1}^{Z_{i,n+1}^p-1} B_{i,k,n+1}^p (Z_{i,n+1}^p - k) \right). \end{aligned} \quad (2)$$

O primeiro termo da equação (2) representa o atraso total das mensagens de classe p chegando ao terminal i durante L_n , desde os instantes de chegadas até o término de L_n . O segundo termo é o atraso de todas as $Z_{i,n+1}^p$ mensagens devido aos intervalos de reserva e DL ($M + L_n^{DL}$). O terceiro termo é composto pelo tempo total que todas as $Z_{i,n+1}^p$ mensagens devem esperar devido às transmissões das mensagens de classe 1 até $(p-1)$, em todos os terminais (1 até M), mais as transmissões das mensagens de classe p dos terminais 1 até $(i-1)$. Por fim, o último termo representa o atraso total decorrido por todas as $Z_{i,n+1}^p$ mensagens entre os instantes em que a primeira destas mensagens começa a ser transmitida e o início das demais transmissões. Considerando que \bar{W}_i^p é o tempo médio de espera das mensagens de classe p chegando na estação i , tem-se:

$$\bar{W}_i^p = \lim_{K \rightarrow \infty} \frac{\sum_{n=1}^K W_i^p(\underline{Y}_{i,n}^p)}{\sum_{n=1}^K N_i^p(\underline{Y}_{i,n}^p)}$$

Aplicando o teorema *Markov Ratio Limit Theorem* - (MRLT) [9], obtém-se:

$$\bar{W}_i^p = \frac{E_o[W_i^p(\underline{Y}_{i,n}^p)]}{E_o[N_i^p(\underline{Y}_{i,n}^p)]}, \quad (3)$$

onde $E_o[N_i^p(\underline{Y}_{i,n}^p)]$ e $E_o[W_i^p(\underline{Y}_{i,n}^p)]$ são, respectivamente, as médias das equações (1) e (2) no estado de equilíbrio. Assim,

$$E_o[N_i^p(\underline{Y}_{p,i,n})] = E_o[Z_{i,n+1}^p] = \lambda_i^p E_o[L_n] = \lambda_i^p E[L], \quad (4)$$

com $E[L] = \lim_{n \rightarrow \infty} E[L_n]$ e $E_o[W_i^p(\underline{Y}_{i,n}^p)]$ é calculado aplicando-se a média termo a termo na equação (2). Com isso, o primeiro termo está descrito abaixo,

$$E_o\left[\sum_{k=1}^{L_n-1} N_{i,k,n}^p(L_n - k)\right] = \frac{\lambda_i^p}{2}(E[L^2] - E[L]), \quad (5)$$

onde $E[L^2] = \lim_{n \rightarrow \infty} E[L_n^2]$. No segundo termo,

$$E_o[(M + L_n^{DL})Z_{i,n+1}^p] = (M + \overline{DL})\lambda_i^p E[L], \quad (6)$$

onde $\overline{DL} = \lim_{n \rightarrow \infty} E[L_n^{DL}]$. Após algumas manipulações, têm-se as equações (7) e (8) referentes ao terceiro termo, e a equação (9) referente ao quarto termo.

$$E_o\left[\left(\sum_{j=1}^{p-1} \sum_{g=1}^M \sum_{k=1}^{Z_{g,n+1}^j} B_{g,k,n+1}^j\right)Z_{i,n+1}^p\right] = \lambda_i^p E[L^2] \sum_{j=1}^{p-1} \sum_{g=1}^M \rho_g^j, \quad (7)$$

$$E_o\left[\left(\sum_{g=1}^{i-1} \sum_{k=1}^{Z_{g,n+1}^g} B_{g,k,n+1}^g\right)Z_{i,n+1}^p\right] = \lambda_i^p E[L^2] \sum_{g=1}^{i-1} \rho_g^p \quad (8)$$

$$E_o\left[\sum_{k=1}^{Z_{i,n+1}^p-1} B_{i,k,n+1}^p(Z_{i,n+1}^p - k)\right] = \frac{\rho_i^p \lambda_i^p E[L^2]}{2} \quad (9)$$

onde $\rho_i^p = \lambda_i^p b_i^p$ nas equações (7), (8) e (9) é o tráfego no terminal i devido as mensagens de classe p . Das equações (5) à (9) obtêm-se $E_o[W_i^p(\underline{Y}_{p,i,n})]$. Usando também $E_o[N_i^p(\underline{Y}_{i,n}^p)]$, dado pela equação (4), e substituindo os termos na equação (3), tem-se o tempo médio de espera para as mensagens de classe p na estação i para a Versão I do protocolo proposto:

$$\overline{W}_i^p = M + \overline{DL} + \left(\frac{1 + \rho_i^p}{2} + \sum_{j=1}^{p-1} \sum_{g=1}^M \rho_g^j + \sum_{j=1}^{i-1} \rho_j^p\right) \frac{E[L^2]}{E[L]} - \frac{1}{2}. \quad (10)$$

A expressão para \overline{W}_i^p é dada em função de $E[L]$ e $E[L^2]$. Prosseguindo com a análise para encontrar o primeiro e o segundo momento no estado de equilíbrio para o tamanho do ciclo, tem-se a seguinte expressão recursiva para $E[L_n]$:

$$\begin{aligned} E[L_n] &= E[L_n^R + L_n^{DL} + L_n^T] \\ &= M + E[L_n^{DL}] + \sum_{p=1}^P \sum_{i=1}^M \rho_i^p E[L_{n-1}]. \end{aligned} \quad (11)$$

Assumindo que $\rho = \sum_{p=1}^P \sum_{i=1}^M \rho_i^p < 1$ e tomando os limites (com $n \rightarrow \infty$) em ambos os lados da equação (11) obtêm-se $E[L]$:

$$E[L] = \frac{M + \overline{DL}}{1 - \rho}; \quad \rho = \sum_{p=1}^P \sum_{i=1}^M \rho_i^p < 1. \quad (12)$$

De maneira similar, através de uma equação recursiva em $E[L_n^2]$ e com $\rho < 1$, tem-se a seguinte expressão para $E[L^2]$

$$\begin{aligned} E[L^2] &= \frac{1}{1 - \sum_{p=1}^P \sum_{i=1}^M (\rho_i^p)^2} \left\{ M^2 + \overline{DL}^2 + 2M\overline{DL} \right. \\ &+ \left[2(M + \overline{DL})\rho + \sum_{p=1}^P \sum_{i=1}^M \lambda_i^p b_{2,i}^p \right] E[L] \\ &+ \left[\sum_{p=1}^P \sum_{i=1}^M \sum_{j=1}^M \rho_i^p \rho_j^p + \sum_{p=1}^P \sum_{q=1}^P \sum_{i=1}^M \sum_{k=1}^M \rho_i^p \rho_k^q \right] E^2[L] \left. \right\}. \end{aligned} \quad (13)$$

Substituindo as equações (12) e (13) na equação (10), tem-se uma expressão fechada para o tempo médio de espera das mensagens de classe p na estação i com a Versão I do protocolo proposto.

B. Modelagem da Versão II

A análise para a Versão II do protocolo proposto segue de maneira direta notando-se que, de acordo com este esquema, as mensagens são transmitidas na mesma ordem que a Versão I, com as classes das mensagens trocadas pelos números das estações e vice-versa (vide Figuras 5 e 6). Então, a expressão para \overline{W}_i^p na Versão II é análoga a da Versão I, alterando-se apenas o i pelo p e o M pelo P e vice-versa:

$$\overline{W}_i^p = M + \overline{DL} + \left(\frac{1 + \rho_i^p}{2} + \sum_{j=1}^{i-1} \sum_{k=1}^P \rho_k^j + \sum_{j=1}^{p-1} \rho_j^i\right) \frac{E[L^2]}{E[L]} - \frac{1}{2}; \quad (14)$$

com $E[L]$ e $E[L^2]$ dados pelas equações (12) e (13), respectivamente.

V. RESULTADOS

Para avaliar o nível de diferenciação obtido com os protocolos descritos, serão considerados dois cenários distintos onde em cada um existe uma probabilidade diferente entre quatro classes de tráfego ($P = 4$), como mostra a Tabela I. Estas quatro classes podem ser mapeadas para os quatro tipos de serviços oferecidos pelo padrão IEEE 802.16 (vide Seção II-B). A diferença entre os cenários é que, no Cenário I existe uma maior probabilidade para as classes de maior prioridade, enquanto que, no Cenário II as classes de menor prioridade predominam sobre as de maior prioridade. Assim, pode-se comparar qual é a influência de uma alta carga dos fluxos de menor prioridade sobre os de maior prioridade e vice-versa.

Classe de Tráfego	Cenário I	Cenário II
Classe 1	40%	10%
Classe 2	30%	20%
Classe 3	20%	30%
Classe 4	10%	40%

TABELA I

CENÁRIOS DE TRÁFEGO UTILIZADOS NA MODELAGEM ANALÍTICA.

Em cada cenário existem 10 estações ($M = 10$) com o tráfego balanceado entre as mesmas, ou seja, $\lambda_i^p = \lambda^p/10$, onde λ^p representa a taxa de mensagens da classe p e λ_i^p a taxa de mensagens da classe p na estação i . Assume-se que o número de pacotes em cada mensagem de classe p na estação i segue uma distribuição geométrica com média $b_i^p = 5$ e $b_{2,i}^p = 45$, para cada $p = 1, 2, 3, 4$; e $i = 1, \dots, 10$. Com isso, o tempo médio de espera para a classe p é dado por $\overline{W}_i^p = \sum_{i=1}^{10} (\lambda_i^p / \lambda^p) \overline{W}_i^p$.

As Figuras 7 e 8 ilustram o tempo médio de espera na fila para cada classe de prioridade em relação ao tráfego oferecido no canal nos Cenários I e II respectivamente. Pelas figuras, observa-se uma diferenciação mais evidente para um alto tráfego no canal e que, com o aumento da intensidade do tráfego, cresce o tempo de espera na fila para todas as classes. Porém, esta diferenciação é menor para Versão II do protocolo proposto, como mostram as Figuras 7(b) e 8(b). Isto acontece porque, na Versão I as prioridades entre classes predominam sobre as prioridades entre estações, ocorrendo o contrário na Versão II onde as prioridades entre estações sobrepõem-se. Através das Figuras 7(a) e 8(a) observa-se que, o tempo de espera para o tráfego de alta prioridade (classe 1) é menor em relação as demais classes, até mesmo no Cenário II onde existe uma maior probabilidade dos tráfegos de baixa prioridade. Com isso, pode-se perceber que os protocolos propostos conseguem diferenciar eficientemente as classes de tráfego, garantindo menor tempo de espera na fila para as mensagens de maior prioridade.

As Figuras 9 e 10 apresentam o tempo médio de espera na fila em função dos terminais, para três valores de intensidade de tráfego ($\rho = 0, 3; 0,6$ e $0,9$), nos Cenários I e II respectivamente. Dessa

forma, o tempo médio de espera na fila para estação i é dado por $\bar{W}_i = \sum_{p=1}^4 (\lambda_i^p / \lambda_i) \bar{W}_i^p$, onde λ_i representa a taxa de mensagens na estação i . Das Figuras 9(a) e (b), observa-se que, para valores fixos de p , a diferença do tempo de espera na fila entre as estações na Versão I é menor do que na Versão II pois, como já foi dito anteriormente, na Versão I a priorização é feita primeiro por classes e não por estações. Por isso, a variação do tempo de espera entre classes é menor para a Versão II que promove uma diferenciação no acesso ao meio entre as estações de forma que, aumenta o tempo médio de espera na fila para as mensagens das estações com baixas prioridades. Com relação a mudança entre os cenários de tráfego, a Versão II do protocolo proposto comporta-se de maneira similar para o Cenário I e II, como ilustram as Figuras 9(b) e 10(b). Por outro lado, na Versão I o tempo de espera para as classes são ligeiramente menores no Cenário II, porém, preservando a diferenciação entre as classes de tráfego como mostram as Figuras 9(a) e 10(a).

Analisando o comportamento do protocolo proposto percebe-se que a atividade no sub-canal UL pode ser vista como uma fila M/G/1 com férias e prioridades, onde o UL é o único servidor do sistema. O período de férias é devido ao sub-canal DL. Durante este período, o servidor do sistema (o sub-canal UL) não serve nenhuma classe de tráfego. Para validar o modelo analítico apresentado na seção anterior, foi utilizada a ferramenta de modelagem e simulação Tangram2 [10]. Foi desenvolvido um modelo para a Versão I do protocolo proposto, com quatro fontes Poisson (uma para cada classe de tráfego) e um servidor. Através da ferramenta, é possível obter como saída o número médio de pacotes na fila para cada classe. Pelo teorema de Little e pela análise do sistema M/G/1, pode-se chegar ao tempo médio de espera na fila para cada classe de tráfego. Foram realizadas 1.000 simulações de 100.000 segundos cada, com um nível de confiança de 95%, sob o Cenário I descrito na seção anterior. A Figura 11 compara os resultados obtidos analiticamente (Figura 11(a)) e através de simulação (Figura 11(b)). Pela figura, nota-se a similaridade entre os modelos e a diferenciação entre as classes de tráfego.

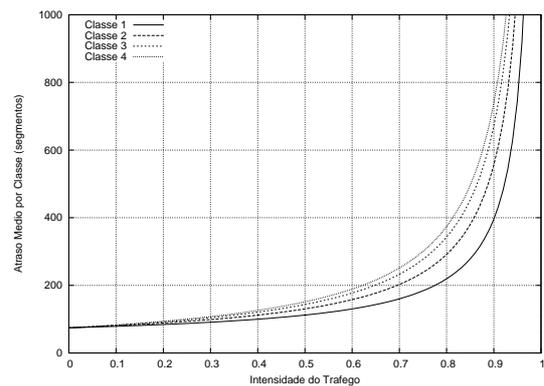
VI. CONCLUSÕES

Neste artigo foi proposto um novo protocolo de acesso ao meio para sistemas BWA que incorpora funções de escalonamento de tráfego com prioridades baseadas em mensagens e/ou em estações. Além disso, foi apresentada uma análise matemática do tempo médio de espera para as duas versões do protocolo proposto, considerando-se chegada de mensagens de acordo com uma distribuição Poisson e com o tamanho das mensagens de acordo com uma distribuição qualquer. Dos resultados obtidos, conclui-se que o protocolo proposto consegue prover, mesmo a uma alta taxa de tráfego, uma diferenciação de serviço entre as classes, reduzindo o tempo médio de espera na fila para a classe de maior prioridade. Além disso, a Versão I apresentou um maior grau de “justiça” no acesso ao meio entre as estações da rede do que a Versão II, levando em consideração uma precedência estática entre as estações.

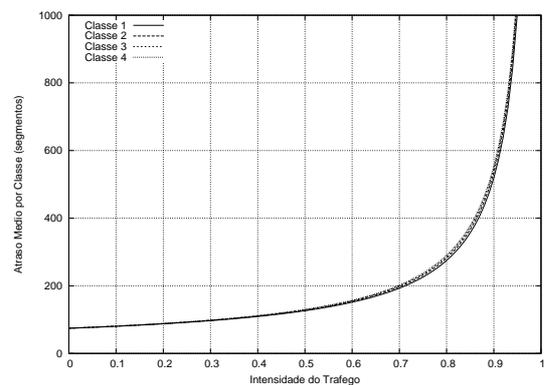
Como trabalhos futuros, pretende-se avaliar o comportamento dos protocolos propostos com a inclusão de prioridades variáveis entre as estações, onde espera-se melhorar o grau de “justiça” no acesso ao meio na Versão II. Além disso, pretende-se incluir aos protocolos um mecanismo de controle de admissão para que a sobrecarga de um determinado tipo de tráfego não afete os demais.

REFERÊNCIAS

- [1] IEEE 802.16, *IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems*, IEEE Std. 802.16, Oct. 2004.
- [2] M. Hawa and D. W. Petr, “M/G/FQ: Stochastic Analysis of Fair Queuing Systems,” in *IEEE 2nd International Conference on Networking (ICN 2002)*, Atlanta, Georgia, Aug. 2002, pp. 368–381.



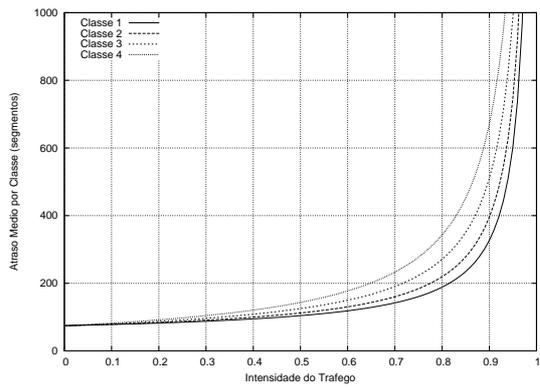
(a) Versão I



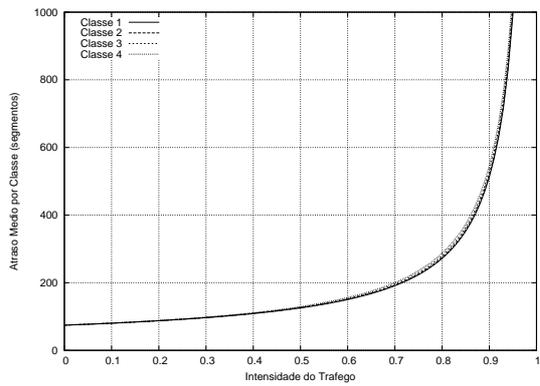
(b) Versão II

Fig. 7. \bar{W}^p no Cenário I: Versão I (a) e Versão II (b).

- [3] —, “Quality of Service Scheduling in Cable and Broadband Wireless Access Systems,” in *Tenth International Workshop on Quality of Service (IWQoS 2002)*, Miami, Florida, May 2002, pp. 247–255.
- [4] M. Hawa, “Stochastic Evaluation of Fair Scheduling with Applications to Quality-of-Service in Broadband Wireless Access Networks,” Ph.D. dissertation, Univ. of Kansas, Aug. 2003.
- [5] K. Wongthavarawat and A. Ganz, “IEEE 802.16 Based Last Mile Broadband Wireless Military Networks with Quality of Service Support,” *IEEE MILCOM'03*, vol. 2, pp. 779–784, Oct. 2003.
- [6] X. Xu, D. Liang, H. Jiang, and X. Lin, “Dynamic Bandwidth Allocation in Fixed BWA Systems,” *ICCT'03*, vol. 2, pp. 1000–1003, Apr. 2003.
- [7] L. F. M. de Moraes and A. N. L. Valverde, “Waiting-Time Analysis of a Reservation Access-Control Scheme with message-based priorities,” *IFIP International Conference on Data Communication Systems and Their Performance*, pp. 283–296, June 1987.
- [8] I. Rubin and L. F. M. de Moraes, “Message Delay Analysis for Polling and Token Multiple-Access Schemes for Local Communication Networks,” *IEEE JSAC*, vol. SAC-1, no. 5, pp. 935–947, Nov. 1983.
- [9] K. L. Chung, *Markov Chains with Stationary Transition Probabilities*. N.Y.: Springer-Verlag, 1967.
- [10] Tangram-II - LAND - PESC/COPPE/UFRJ. [Online]. Available: <http://www.land.ufrj.br/tools/tangram2/tangram2.html>

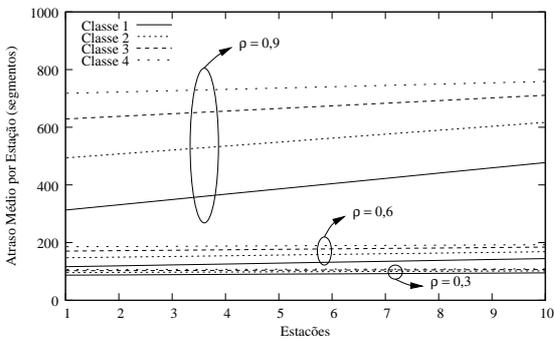


(a) Versão I

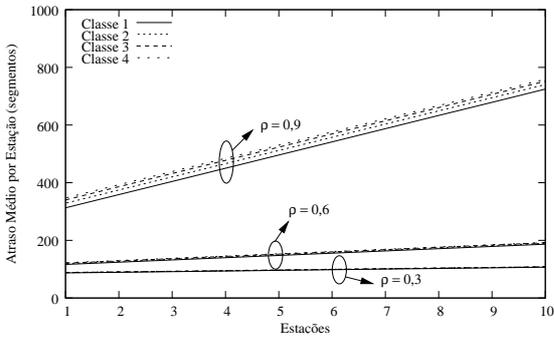


(b) Versão II

Fig. 8. \bar{W}^P no Cenário II: Versão I (a) e Versão II (b).

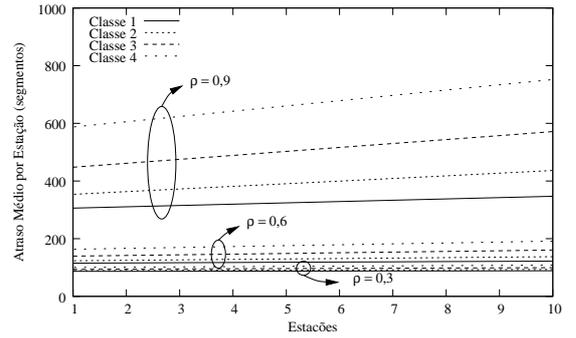


(a) Versão I

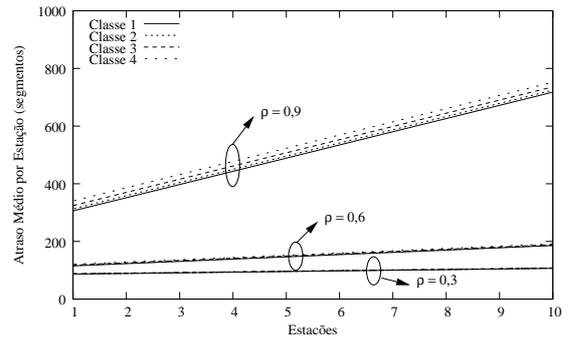


(b) Versão II

Fig. 9. \bar{W}_i no Cenário I: Versão I (a) e Versão II (b).

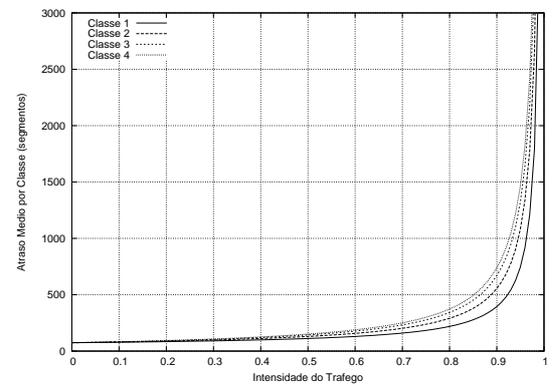


(a) Versão I

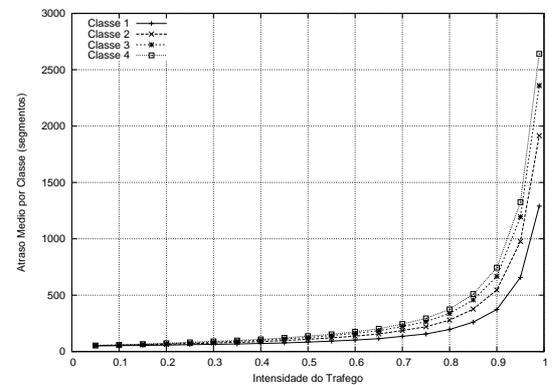


(b) Versão II

Fig. 10. \bar{W}_i no Cenário II: Versão I (a) e Versão II (b).



(a) Resultado Analítico



(b) Resultado Simulado

Fig. 11. Comparação entre os resultados analíticos e simulados.