

Provisionamento de QoS baseado em Reserva Adaptativa de Banda em Redes Móveis

Kelvin Lopes Dias¹, Djamel F. H. Sadok²

¹CCET - Universidade da Amazônia - UNAMA
Campus Alcindo Cacela- 66060-902 – Belém - PA - Brasil

kld@bcc.unama.br

²Centro de Informática – Universidade Federal de Pernambuco (UFPE)
Caixa Postal 7851 – 50732-930 – Recife – PE – Brasil

jamel@cin.ufpe.br

Resumo. *Este artigo propõe um arcabouço adaptativo para a gerência de banda em redes móveis 4G. A proposta baseia-se em previsão de mobilidade e considera a banda passante requerida por chamadas de handoff para a reserva de recursos durante uma janela de tempo. Nossa proposta é comparada, através de simulações, com o esquema fixo de canal de guarda e outras propostas otimizadas baseadas em reserva dinâmicas de recursos presentes na literatura. Os resultados de simulação mostraram a superioridade de nossa proposta em termos de utilização, garantias de níveis desejáveis para a probabilidade de descarte de handoffs e redução da sinalização.*

Abstract. *This paper proposes an adaptive bandwidth management scheme that avoids per-user reservation signalling overhead and meets the scalability requirement advocated for future 4G networks. The proposal uses mobility prediction based on mobile positioning system and takes into account the expected bandwidth to be used by calls handed off from neighboring cells within an estimation time window. Our scheme is compared through simulations with the fixed guard channel and other optimized dynamic reservation-based proposals present in the literature. The simulation results show that the proposed scheme is superior to the other schemes in terms of handoff dropping probability, utilization and signalling overhead.*

1. Introdução

O advento da Quarta Geração (4G) das comunicações móveis [Hui 2003], implica na necessidade de novos algoritmos e mecanismos para tratar a *mobilidade* do usuário a fim de viabilizar uma comunicação transparente, sem interrupções e com garantias de qualidade de serviço (QoS – *Quality of Service*). Em redes móveis celulares, o *handoff* é o procedimento que transfere uma chamada em andamento de uma célula para outra à medida que uma estação móvel desloca-se através da área de cobertura da rede. Se a célula alvo para a qual a estação móvel migra não tem recursos suficientes de banda passante, a chamada será descartada. Da perspectiva do usuário, o descarte de chamadas em andamento durante o *handoff* é menos desejável que o bloqueio de novas chamadas. Um mecanismo essencial para a provisão de QoS no nível

de chamadas é o controle de admissão de chamadas (CAC) e reserva de recursos. O CAC deve assegurar não apenas os requisitos de QoS das novas chamadas que chegam ao sistema, se aceitas, mas também garantir que os requisitos das chamadas existentes não serão deteriorados com a admissão de mais usuários.

A reserva de recursos tem efeitos mais negativos em redes móveis sem fio do que nas redes fixas de alta velocidade devido à escassez de largura de banda dos enlaces sem fio. Uma arquitetura escalável para as futuras redes móveis 4G, deve evitar o uso excessivo de sinalização por usuário para efetuar reservas de recursos de rádio. Neste artigo propõe-se um novo esquema para provisionamento de QoS em redes móveis que evita a tradicional reserva de recursos por usuário e reduz o *overhead* de sinalização associado. As características chave do esquema proposto são:

- ✓ Utiliza predição de mobilidade e estima a carga agregada que migrará para a célula de admissão¹, dentro de uma janela de tempo. O objetivo é evitar que a admissão deste novo usuário provoque o descarte de chamadas em andamento, caso estas efetuem *handoffs* para a célula de admissão, e;
- ✓ Introduce uma nova técnica denominada *Reserva Virtual (RV)*, através da qual evita-se que usuário sinalize e efetue reservas individuais de banda na próxima célula a ser visitada, aumentando a escalabilidade da rede.

Nossa proposta é comparada por meio de simulações com esquemas baseados em reservas fixas e dinâmicas de banda. Os resultados das simulações demonstram que nossa proposta garante melhores resultados em termos de probabilidade de descarte em casos de *handoffs* mal sucedidos e probabilidade de bloqueio de novas chamadas, melhor utilização dos recursos dos enlaces sem fio, além de um *overhead* de sinalização reduzido. O restante deste artigo é organizado da seguinte forma. Na seção 2, descrevemos os trabalhos relacionados. Em seguida, apresentamos a proposta do novo CAC na seção 3. A seção 4 descreve o modelo de simulação e métricas. Os resultados de desempenho são apresentados na seção 5. Apresentamos as conclusões e trabalhos futuros na seção 6.

2. Trabalhos relacionados

De uma maneira geral, a idéia básica das estratégias para controle de admissão e gerencia de banda passante na literatura é reservar recursos *a priori* nas células do sistema para reduzir a probabilidade de descarte no caso do móvel efetuar um procedimento de *handoff* [Oliveira 1998] [Jiang 2001] propostas existentes na literatura para CAC em redes móveis sem fio podem ser divididas, basicamente, em duas categorias: fixas e dinâmicas. Estratégias fixas, como o *canal de guarda (GC - Guard Channel)* 0, proporcionam tratamento diferenciado para chamadas de *handoffs*, reservando um número fixo de canais exclusivamente para tais chamadas. Este esquema reduz a probabilidade de descarte devido ao *handoff*, entretanto, aumenta a probabilidade de bloqueio de novas chamadas. A vantagem desta estratégia é sua simplicidade porque não há a necessidade de troca de informações entre as estações

¹ Utilizaremos o termo *célula de admissão* para indicar a célula onde o usuário está requisitando a admissão.

rádio-base. Por outro lado, não é flexível para tratar cargas variáveis, uma vez que não utiliza informações sobre a carga da célula atual e das vizinhas.

Propostas de esquemas dinâmicos baseados em reservas estendem o esquema básico da estratégia fixa de acordo com estimativas da taxa de handoffs derivadas do número de chamadas na vizinhança e do padrão de mobilidade destas chamadas para reservar banda na próxima célula a ser visitada pelo usuário. De uma maneira geral, o usuário será admitido se existir banda suficiente na célula de admissão e se for possível efetuar reservas na vizinhança desta célula.

O controle de admissão proposto em [Jiang 2001] leva em consideração o número de chamadas nas células adjacentes, além do número de chamadas existentes na célula onde o usuário requisita a admissão. Os autores desenvolveram um modelo teórico para computar os requisitos dos pedidos de *handoff* com o objetivo de manter uma determinada probabilidade de descarte de *handoff*. O modelo proposto assume que todas as chamadas são idênticas, o que deixa de ser uma suposição válida em ambientes com aplicações multimídia com requisitos de banda variáveis. Este esquema não utiliza qualquer informação de mobilidade para prever a próxima célula na trajetória do usuário.

Recentemente, o GPS (*Global Positioning System*) tem sido adotado para auxiliar as propostas de controle de admissão de novas chamadas, uma vez que pode ser usado para prever a próxima célula a ser visitada pelo móvel de forma mais precisa [Chiu 2000][Xu 2002][Soh 2003]. Além disso, esta abordagem pode evitar a degradação da utilização da capacidade da rede devido às reservas de recursos desnecessárias. A idéia básica desta proposta é iniciar o pedido de reserva de recursos na próxima célula predita para o usuário, apenas quando este alcançar um certo limiar que indique a iminência da ocorrência de um *handoff*. Estes esquemas exigem uma quantidade excessiva de mensagens de sinalização para reserva por usuário e cancelamento de reservas falsas. Além disso, reservas podem reduzir a probabilidade de descarte de *handoffs* com o efeito colateral de aumentar a probabilidade de bloqueios, o que pode acarretar utilização ineficiente dos recursos da rede.

3. Descrição da Proposta

A proposta utiliza duas fases. Inicialmente, considera-se o comportamento do agregado de tráfego de *handoff*, proveniente de células adjacentes, para prover uma expectativa de carga para a célula de admissão. Posteriormente, verifica-se a disponibilidade de recursos suficientes na próxima célula predita na trajetória do móvel utilizando uma técnica que denominamos de Reserva Virtual (RV).

Seja B_i a taxa de transmissão requerida por cada um dos usuários pertencente ao conjunto A de usuários com chamadas em andamento na célula de admissão; B_k o requisito de taxa de transmissão do novo usuário e , considerando C como a capacidade do enlace sem fio da célula de admissão, uma primeira condição que deve ser satisfeita na fase inicial do algoritmo é dada por:

$$\sum_{i \in A} B_i + B_k \leq C \quad (1)$$

Com o intuito de evitar o descarte súbito de chamadas em andamento dos usuários presentes em células vizinhas, caso estes realizem *handoffs* para a célula de

admissão, uma segunda condição pertencente à primeira fase deve ser satisfeita. Denominamos este último teste da primeira fase de *teste da carga agregada* das vizinhas com relação à célula de admissão. A idéia é obter uma estimativa das necessidades de banda da célula de admissão que será requerida dentro de uma janela de tempo. Esta estimativa corresponde ao somatório da capacidade que será utilizada pelas chamadas em andamento na célula de admissão, bem como daquelas que migrarão para esta célula dentro da janela de tempo.

Em nosso sistema, cada usuário admitido, s , está associado com uma célula vizinha dependendo de sua trajetória. Um grupo de usuários registrado com a célula i , cuja vizinha (célula prevista para o próximo *handoff*) seja j , é denotado por S_{cij} e corresponde à *base de dados direcional*² da célula i para a célula j .

Utilizamos o GPS também para a verificação de limiares de tempo que indiquem a iminência de *handoff* de um determinado usuário. Isto é feito com o intuito de otimizar o cálculo da carga direcional de uma célula para cada uma de suas vizinhas, considerando a localização e velocidade do usuário. A idéia é utilizar o requisito total de taxa do usuário para atualizar a carga direcional para a célula vizinha correspondente, quando o tempo restante para o usuário efetuar o *handoff* para tal célula for inferior a um determinado limiar de tempo que denominamos de LTH (limiar de tempo para o *handoff*). Caso contrário, se as medições atuais do GPS indicarem que o usuário demorará mais do que o LTH para efetuar o *handoff* (em função de sua localização e velocidade), utiliza-se uma carga proporcional, cujo valor depende do tempo que resta para o usuário efetuar o *handoff*. Denominamos este tempo de TR (tempo restante para o *handoff*). Todas estas atualizações são realizadas localmente por cada estação rádio-base. O objetivo desta abordagem é aumentar a quantidade de reservas quando um grande número de usuários de alta mobilidade estiver presente na rede, ou de outra forma, minimizar o efeito de bloqueios desnecessários de novas chamadas devido a reservas excessivas, quando a rede possuir usuários de baixa velocidade e/ou estacionários.

Seja W_s um fator que indica a iminência de *handoff* e B_s a taxa de transmissão requerida por cada usuário pertencente à S_{cij} e, considerando M conexões ativas, a carga direcional de i para a célula alvo j é dada por:

$$\Omega_{cij} = \sum_{s \in S_{cij}} (W_s \times B_s) \quad (2)$$

onde $0 < W_s \leq 1$.

Considerando LTH como um limiar de tempo configurado pelo administrador da rede e TR o tempo que resta para o usuário efetuar o *handoff*, se $TR \leq LTH$ então:

$$W_s = 1, \text{ caso contrário, } W_s = \frac{LTH}{TR}.$$

² Contém informações sobre a carga direcional de uma célula para as suas vizinhas. A carga direcional da célula i para a j corresponde ao somatório dos requisitos de banda dos usuários previstos para migrar de i para j .

Além da carga proveniente de usuários que migrarão para a célula em questão, deve-se levar em consideração a carga remanescente devido a usuários estacionários, sobretudo em células onde a carga atual (U) esteja próxima à capacidade máxima do enlace (C). Dessa forma, se na célula de admissão, N usuários estão com chamadas em andamento, podemos ter um componente de *prevenção de congestionamento* que evita a admissão excessiva na célula em questão:

$$\Omega_{Cij} = \sum_{i=1}^N (L \times Bi) \quad (3)$$

onde $L = \frac{U}{C} \times MaxL$,

e $0 < MaxL \leq 1$. Assim, $MaxL$ representa o valor máximo assumido por L . Um valor baixo para $MaxL$ beneficiará a utilização global da rede a um custo de aumentar a probabilidade de descarte de *handoff*. Valores altos para $MaxL$ devem gerar uma baixa utilização da rede, contudo, a probabilidade de descarte de *handoffs* é minimizada.

Seja Ω_{ij} a carga referente aos móveis presentes na célula i que migrarão para a célula j e Ω_{jj} a carga referente àqueles que permanecerão na célula j . Considerando, ainda, W_k o peso a ser multiplicado pelo requisito de banda B_k do novo usuário k que tem chances de migrar para a célula C_j (adjacente à célula de admissão), a segunda condição (*teste de carga agregada*) pertencente à fase inicial é dada por:

$$\Omega_{Cij} + \Omega_{Cjj} + W_k \times B_k \leq C_j \quad (4)$$

A cada período τ , os valores de Ω_{Cij} e Ω_{Cjj} são calculados por cada estação rádio-base e são enviados a um servidor de QoS existente na rede. Diferente das propostas apresentadas em [Jiang 2001], onde cada estação rádio-base é diretamente responsável pelo envio e recepção de mensagens de suas células vizinhas, utilizamos uma estrutura similar àquela de gerenciamento escalável de recursos usando *bandwidth brokers* [Zhang 2001]. Um servidor denominado SQ (Servidor de QoS) é responsável por receber as informações enviadas pelas estações rádio-base que pertencem a um domínio administrativo e enviar de volta as informações sumarizadas para a realização local (nas estações rádio-base) do controle e gerenciamento de banda. Dessa forma, a informação sobre carga predita para a provável célula de *handoff* não é trocada diretamente entre as estações rádio-base a cada nova requisição de admissão ou *handoff*. Esta informação é trocada periodicamente através da rede de *backbone*.

A *última fase* de nossa proposta utiliza informações da carga direcional das células vizinhas, recebidas do SQ no início da janela de tempo, para implementar o que denominamos de *Reserva Virtual*. Este é um esquema virtual porque não há reservas explícitas por usuário. A estação rádio-base da célula, onde o usuário requisita admissão, consulta valores aproximados sobre a disponibilidade de recursos na célula predita para um futuro *handoff* do usuário baseando-se nos valores recebidos do SQ. O objetivo é evitar que a nova chamada seja admitida quando estima-se que a próxima célula no caminho do usuário estará sem recursos suficientes. Desta forma, minimiza-se o descarte de chamadas devido à ausência de recursos na célula para onde o usuário efetuará o *handoff*. Para que a segunda fase da proposta seja bem sucedida, a condição abaixo deve ser satisfeita:

$$\Omega_{C_{jh}} + \Omega_{C_{jr}} \leq C_j \quad (5)$$

onde $\Omega_{C_{jh}}$ representa o somatório das carga direcionais de todas as vizinhas da célula j predita para o *handoff*. $\Omega_{C_{jr}}$ representa uma fração da carga presente em j , calculada no início do período, como em (3). C_j representa a capacidade da célula predita para o *handoff*. Esta fase também se beneficia do *backbone* da rede de acesso e do SQ para troca e processamento das informações sobre carga direcional enviadas pelas estações rádio-base.

Em suma, ao final de uma janela de tempo de τ segundos, as estações rádio-base enviam e recebem as informações necessárias para a realização das duas fases da proposta. Esta troca se dá uma única vez a cada τ segundos, isto é, cada estação rádio-base envia/recebe uma mensagem para/do SQ. As estações rádio-base realizam localmente os testes de admissão com informações sobre carga direcional de cada uma de suas vizinhas enviada pelo SQ.

3.1 Controle Adaptativo do Valor de LTH em cada Célula

Um aspecto fundamental para o funcionamento adequado de nossa proposta é a seleção do valor de LTH. O valor assumido pelo LTH governa a atualização dos valores armazenados na base de dados de carga direcional da célula em questão para suas vizinhas. Em cenários com células congestionadas, espera-se que o LTH seja alto, permitindo que as células vizinhas sejam informadas, durante a próxima troca de mensagens, da necessidade de aumentarem suas reservas para evitar que mais chamadas de *handoff* sejam descartadas na célula congestionada. A escolha de um valor ótimo para o LTH depende das características do tráfego e mobilidade dos usuários. Para assistir à seleção do valor do LTH, propomos um esquema adaptativo que tem como critério o valor medido da taxa de descarte de *handoff* (Pd) em cada célula, a cada intervalo entre troca de mensagens sobre carga direcional. A Figura 1 mostra o algoritmo utilizado para o ajuste do valor de LTH em cada célula.

Max_Pd - Valor de Pd a partir do qual deve-se diminuir o valor de LTH.

Min_Pd - Valor de Pd a partir do qual pode-se aumentar o valor de LTH

Max_LTH - Valor máximo para o LTH

Min_LTH - Valor mínimo para o LTH

1. Ao final de cada intervalo entre troca de mensagens
2. Para cada célula
3. Se ($Pd_Medido \geq Max_Pd \ \&\& \ LTH < Max_LTH$)
4. Incrementar LTH;
5. Se ($Pd_Medido < Min_Pd \ \&\& \ LTH > Min_LTH$)
6. Decrementar LTH;

Figura 1 – Algoritmo para o ajuste dinâmico do LTH da proposta.

O algoritmo é executado a cada intervalo entre troca de mensagens de carga direcional. Cada célula verifica se a probabilidade de descarte medida na célula durante

o intervalo anterior é superior ou igual ao limite pré-estabelecido para esta métrica (linha 3). Neste caso, se o valor máximo de LTH ainda não tiver sido alcançado, este é incrementado. Por outro lado, caso a probabilidade de descarte medida esteja abaixo de um limite inferior e o LTH atual esteja acima de um valor mínimo (linha 5), diminui-se o valor de LTH de forma que a utilização média seja otimizada com uma menor quantidade de reservas realizada pelas estimativas existentes de carga direcional.

De acordo com simulações realizadas variando um parâmetro e deixando os demais fixos, selecionamos os seguintes valores para o algoritmo adaptativo: $Max_Pd = 1\%$, $Min_Pd = 0,1\%$, $Max_LTH = 60s$, $Min_LTH = 10s$, $LTH_inicial = 30s$, incremento do LTH = 5 e decremento do LTH = 1. Algumas destas simulações são apresentadas neste artigo, nas seções iniciais que descrevem os resultados. O valor de Max_Pd igual a 1% refere-se ao limite superior que deve ser garantido, isto é, não violado 000. O intervalo entre execuções do algoritmo é igual a janela de tempo $\tau = 30s$, que define o intervalo entre trocas de mensagens de cargas direcionais. É importante ressaltar que a escolha dos parâmetros relacionados ao LTH (Max, Min e inicial) e à janela podem influenciar nos resultados da simulação. Optamos por escolher valores próximos ao tempo médio de residência utilizado em nossas simulações que foi de 60s. Note que em um ambiente real poderíamos avaliar o tempo médio de residência e velocidade dos móveis para utilizar valores condizentes com o cenário sendo estudado.

4. Modelo de Simulação e Métricas

Para a avaliação de nossa proposta, foi desenvolvido um simulador escrito em linguagem C. Além do simulador desenvolvido para nossa proposta, desenvolvemos também três simuladores para esquemas de controle de admissão baseados em reservas fixas (Canal de Guarda - CG) [Hong 1986], dinâmicos com reserva antecipada (RA – Reservas Antecipadas) representando uma classe geral de propostas que reservam banda imediatamente ao requisitar uma conexão [Oliveira 1998] e o ACR (*Adaptive Channel Reservation*) que otimiza o pedido de reserva dependendo de informações de posicionamento e limiares de tempo para a efetivação do *handoff* [Xu 2002]. Este último representa uma classe de algoritmos que realiza a reserva dependendo de limiares de distância ou tempo para a ocorrência de uma *handoff* [Chiu 2000][Soh 2003].

A Figura 2 apresenta a topologia celular simulada que é composta de 19 células, utilizando hexágonos para representar as relações de adjacência. Cada célula tem a capacidade de 100 BUs (*bandwidth units*) [Xu 2002][Soh 2003], onde cada unidade representa a largura de banda requerida por uma chamada de voz. Para evitar o efeito de borda utilizamos a mesma abordagem adotada na maioria dos trabalhos relacionados da literatura. Ou seja, quando um usuário move-se para fora da topologia deverá entrar novamente na rede pelo outro lado. Por exemplo, usuários nas células de borda 8 e 12 que estejam deslocando-se para fora da topologia deverão entrar novamente no sistema através das células 14 e 18, respectivamente. O modelo de tráfego é similar ao utilizado em [Choi 2002][Soh 2003]. Chamadas são geradas de acordo com um processo de Poisson com taxa de λ (chamada/célula/segundo) em cada célula. Os modelos de tráfego simulado consistem em usuários com requisitos de 1 BU (voz) e 4 BUs (vídeo) com probabilidades Rvo and $1-Rvo$, respectivamente, onde Rvo também é denominada de proporção de voz (*voice ratio*) como em [Soh 2003]. Em nossas simulações Rvo é configurada com o valor 0.7, isto é, 70% do tráfego simulado é de voz e 30% de vídeo.

A duração de cada chamada (T_c) é exponencialmente distribuída com média 180s [Chiu 2000] [Xu 2002][Soh 2003].

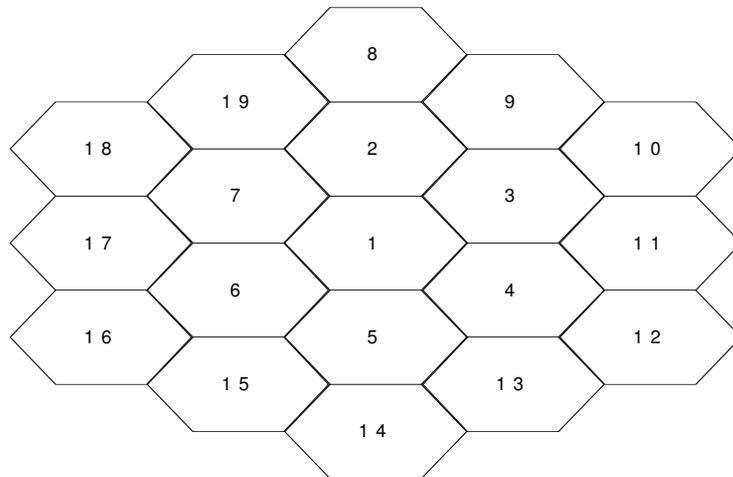


Figura 2 - Topologia celular adotada nas simulações.

Os modelos de mobilidade utilizados nas simulações são similares aos adotados em [Chiu 2000][Xu 2002]. No modelo 1 (*Mob1*), quando uma nova chamada é gerada, inicialmente o móvel escolhe um tempo de residência exponencialmente distribuído com média 60s. A direção do movimento é uniformemente distribuída sobre as seis células vizinhas. Em cada período de tempo de duração variável T_u (exponencialmente distribuída com média $E[T_c]/2$), o móvel desloca-se em linha reta sem mudança de direção a cada *handoff* e com velocidade constante. Após este período T_u , o móvel pode parar com probabilidade $P_{parada} = 0,1$, por um tempo $T_p = 20s$ ou continuar a movimentar-se (com probabilidade $P_{cont} = 1 - P_{parada}$) na mesma direção. O modelo 2 (*Mob2*) difere do primeiro por permitir um grau de incerteza na escolha da próxima célula no trajeto do móvel. O móvel pode realizar mudanças de direção enquanto cruza a célula. Dessa forma, o móvel seguirá a mesma direção com probabilidade 0,5. Caso o móvel mude de direção, deverá selecionar uniformemente uma das outras cinco células vizinhas como a próxima célula para a qual efetuará o *handoff*.

As métricas avaliadas nos modelos simulados são: Probabilidade de bloqueio de novas chamadas (P_b): razão do número de novas chamadas admitidas sobre o número total de requisições de novas chamadas ao sistema; Probabilidade de descarte de *handoff* (P_d): razão do número descartes sobre o total de tentativas de *handoff*; e utilização média: razão da quantidade de recursos de banda passante em uso no sistema sobre a capacidade total. Os gráficos obtidos das simulações apresentam as métricas de interesse em função da carga oferecida por célula [Choi 2002], definida como: *taxa média de geração de chamadas x requisito de banda da chamada x tempo médio da duração da chamada*, isto é: $\Omega = [1 \times R_{vo} + 4 \times (1 - R_{vo})] \times \lambda \times 180$. O significado físico da carga oferecida por célula é a largura de banda requerida, em média, para suportar todas as conexões existentes em uma célula. Para as simulações apresentadas neste artigo, consideramos uma faixa de carga oferecida variando de 30 a 300 BUs. Em geral, é desejável que esta carga esteja abaixo, ou seja, igual à capacidade do enlace de cada célula, isto é, 100 BU. Entretanto, a carga oferecida pode aumentar rapidamente quando usuários são bloqueados e tentam, repetidamente, estabelecer a conexão na mesma célula [Choi 2002]. Nós consideramos valores até 300 BUs, uma vez que mesmo para cargas oferecidas bastante altas, um P_d máximo deve ser garantido.

5. Resultados das Simulações

Esta seção apresenta os resultados da avaliação de nossa proposta. Os resultados das simulações foram obtidos através da técnica de médias em lotes (*batch means*), removendo-se estados transientes do início da simulação, computando a média dos lotes e calculando a média final a partir das médias obtidas destes lotes. As barras verticais para os valores médios apresentados nos gráficos foram calculadas considerando um intervalo de confiança de 95%.

5.1 Estudo da Variação do LTH da Proposta

Inicialmente, verificamos o impacto da seleção de diferentes valores do *LTH* (limiar de tempo para *handoff*) sem a adaptação. O valor de *LTH* dita se o requisito de banda de um móvel deve ser considerado no cálculo da carga direcional trocada entre as células. A Tabela 1 mostra os valores obtidos pelas métricas avaliadas para as cargas oferecidas de 100 BUs (operação normal da rede) e 300 BUs (sobrecarga da rede) considerando *LTHs* iguais a 10s, 30s e 60s. Os demais parâmetros de nossa proposta foram configurados com os mesmos valores para todos os cenários de *LTHs* analisados. Assim, *MaxL* e o tempo entre trocas de cargas direcionais foram configurados com 1 e 30s, respectivamente.

Podemos observar que o bloqueio de novas chamadas aumenta com a utilização de valores mais altos para o *LTH*, isto é, quanto mais antecipada for a inclusão dos requisitos de banda do usuário na base de dados de carga direcional, maior será a quantidade de banda reservada, evitando que mais usuários tenham acesso à rede. O efeito de uma maior quantidade de reserva de banda passante é a redução do descarte de chamadas de *handoff*. A utilização média também é afetada pelos diferentes valores de *LTH*. Note que quanto maior o valor do *LTH*, menor a utilização. Este comportamento da utilização é uma consequência do maior número de bloqueios de usuários gerados pela maior quantidade de reservas devido ao aumento do *LTH*. Mantendo-se fixo o valor do parâmetro *MaxL*, limites máximos de *Pd* não são extrapolados com a escolha apropriada do *LTH*. Por exemplo, para um requisito de 1%, a configuração com *LTH* = 30s seria a mais indicada.

Tabela 1 – Resultados para a carga de 100 e 300 BUs, variando o *LTH* (Mob1).

Configuração	Carga Oferecida	Bloqueio (<i>Pb</i>)	Descarte (<i>Pd</i>)	Utilização
<i>LTH</i> = 10s	100	17%	3,2%	81,4%
<i>LTH</i> = 30s	100	31%	0,39%	72,9%
<i>LTH</i> = 60s	100	37,9%	0,004%	65,3%
<i>LTH</i> = 10s	300	65%	7,2%	88,3%
<i>LTH</i> = 30s	300	73,5%	1%	78,9%
<i>LTH</i> = 60s	300	76,7%	0,17%	70,8%

5.2 Estudo da Variação do *MaxL*

A Tabela 2 mostra os resultados obtidos da avaliação da variação do parâmetro *MaxL*, utilizado como o limite superior no cálculo da porcentagem da carga atual da célula a ser acumulada no cômputo da carga direcional. Foram simulados cenários com

valores iguais a 0,5, 0,7 e 1. Os valores para o tempo entre troca de cargas direcionais e LTH foram ambos iguais a 30s, para todos os três cenários simulados.

Como esperado, o aumento do *MaxL* ocasiona um acréscimo no bloqueio de novas chamadas. A utilização do valor integral (*MaxL* = 1) do requisito de banda do usuário proporcionou o maior bloqueio, como indicado pelos valores da métricas apresentadas na Tabela 2. O efeito colateral do maior bloqueio na configuração do parâmetro *MaxL* com o valor unitário é a baixa utilização média da banda passante da rede que foi 10,4% inferior à utilização do cenário com a configuração do *MaxL* = 0,7 na carga de 100 BUs, bem como o descarte excessivamente baixo. O valor desejável como o limite máximo a ser garantido para o descarte de chamadas de *handoffs* é de 1% [Choi 2002][Wu 2002][Soh 2003]. A configuração da proposta que utiliza *MaxL* com o valor igual a 0,5 obteve os melhores resultados de bloqueio e utilização. Entretanto, o menor número de usuários bloqueados e a maior utilização média para esta configuração da proposta causam um descarte muito alto e inaceitável de 25,26% para a carga oferecida de 300 BUs.

Tabela 2 - Resultados para cargas de 100 e 300 BUs, variando o *MaxL*.

<i>MaxL</i>	Carga Oferecida	Bloqueio	Descarte	Utilização
0,5	100	7,8%	5,9%	84,1%
0,7	100	17,2%	2,5%	81,4%
1	100	31%	0,39%	72,9%
0,5	300	46,8%	25,26%	94,8%
0,7	300	65%	72,3%	88,3%
1	300	73,5%	1,02%	78,9%

5.3 Comparação com o Esquema de Canal de Guarda

A Figura 3 apresenta os resultados (em escala logarítmica, exceto para a utilização) da comparação de nossa proposta com o esquema fixo de canal de guarda (CG), que aloca parte dos recursos da célula exclusivamente para tráfego de *handoff*. Os parâmetros de nossa proposta são configurados com os valores descritos na Seção 3.1. Para o CG, simulamos três alocações: 5%, 15% e 20% da capacidade das células, exclusivamente para tráfego de *handoff*. Dos três esquemas de CG simulados, apenas a configuração com 20% de reserva consegue manter o *Pd* em 1%. Nossa proposta mantém o *Pd* sempre abaixo de 1% ao longo de todo o espectro de cargas simuladas e bloqueia uma quantidade menor de usuários para cargas oferecidas abaixo de 100BUs, com exceção do CG com reserva de 5% de banda passante que obteve um menor bloqueio a partir de 80BUs de carga oferecida.

A utilização média de banda obtida por nossa proposta só foi inferior que os esquemas de canal de guarda com 5% e 15% de reserva para cargas acima de 80 BUs e 130 BUs, respectivamente. Note que o *Pd* obtido por estas configurações do CG está acima do máximo desejável de 1% para estas cargas. Para cargas acima de 100BUs, nossa proposta obteve uma utilização similar a do CG com reserva de 20%, que foi o único esquema com reserva fixa que não extrapolou o requisito máximo em termos de descarte de *handoffs*. O ganho de nossa proposta com relação aos esquemas CG-15 e CG-20 foi de 8% e 12,8% na carga oferecida de 80 BUs, respectivamente. Apesar da

melhor utilização obtida pelo esquema CG-5 para a carga igual a 300 BUs, esta configuração atingiu um descarte de 10,7%, enquanto nossa proposta obteve o valor de 0,92%. Dessa forma, o RV otimiza a utilização da rede para cargas baixas, evitando o bloqueio excessivo gerado pela proposta de canal de guarda fixo, bem como garante um P_d abaixo de 1% para toda a faixa de cargas simuladas.

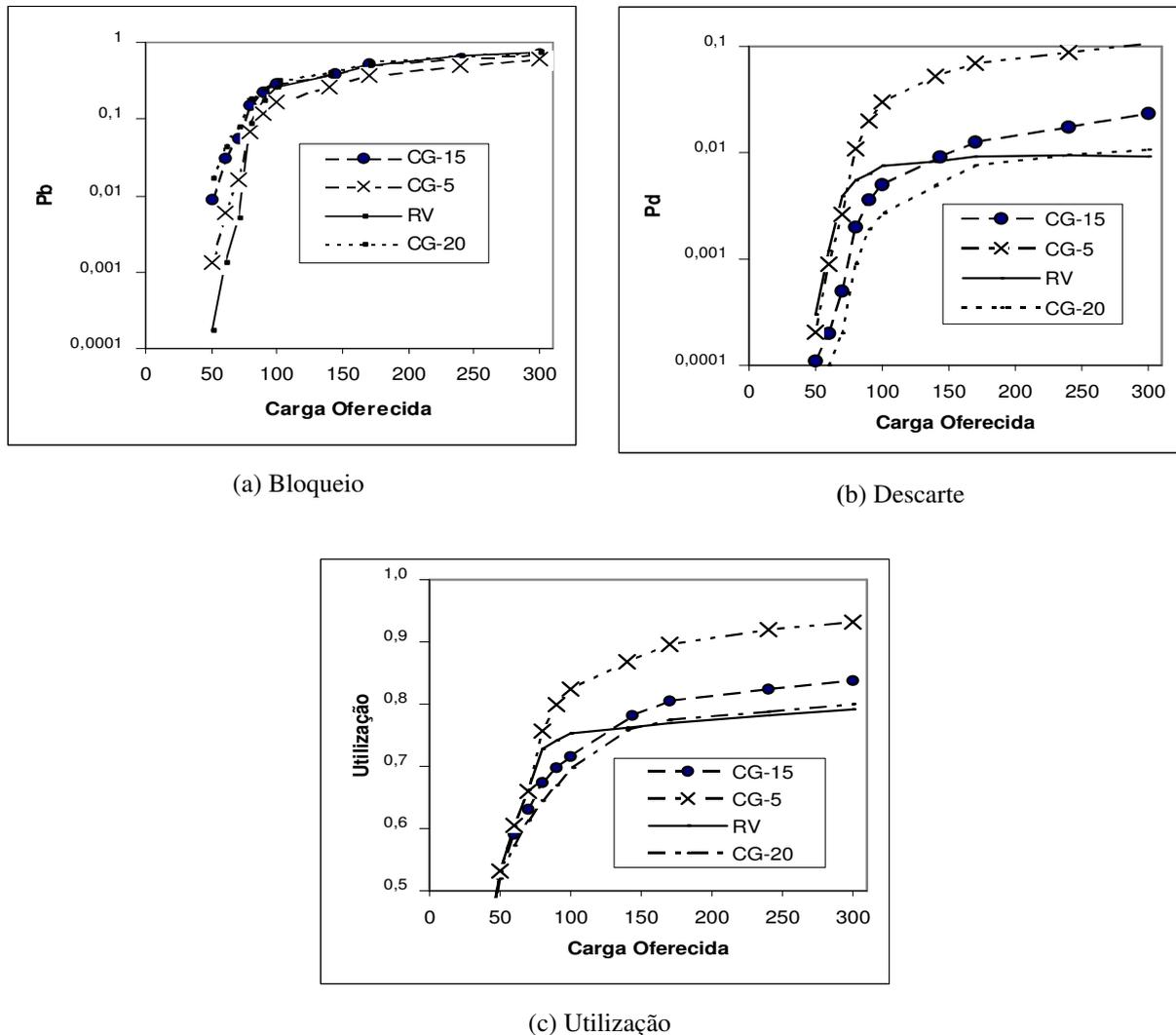


Figura 3 – Comparação com o esquema alocação fixa (CG).

5.4 Comparação com Esquemas de Reservas Antecipadas por Usuário

Esta seção apresenta o estudo comparativo com esquemas que reservam de forma antecipada por usuário em células vizinhas. Inicialmente, avaliamos a reserva antecipada imediata que efetua a reserva na próxima célula (RA-Simples), bem como reservando nas duas próximas células (RA-Dupla) na trajetória do usuário assim que este requisita admissão ou realiza *handoffs*. A Tabela 3 resume os resultados obtidos da comparação realizada entre nossa proposta e os esquemas de reservas antecipadas simples e dupla. São apresentados os resultados referentes aos cenários utilizando o modelo de mobilidade determinístico (Mob1) e o aleatório (Mob2). De uma maneira geral, nossa proposta consegue manter níveis similares de desempenho em ambos os cenários de mobilidade. O limite superior de 1% para a probabilidade de descarte é

garantido, isto é, não é ultrapassado, por nossa proposta. O esquema que realiza reserva antecipada simples não evita a violação do Pd . Com relação à métrica de utilização média, os melhores resultados foram obtidos por nossa proposta nos dois cenários de mobilidade.

Dependendo da carga oferecida e da quantidade de *handoffs* realizadas pelos usuários, a garantia de Pd inferior ou igual a 1% pode exigir a reserva em mais de uma célula [Wu 2002]. Considerando a carga oferecida de 300 BUs e o modelo de mobilidade 2, as probabilidades de descarte de *handoffs* foram iguais a 5,2% e 0,9% para os esquemas de reserva simples e dupla, respectivamente.

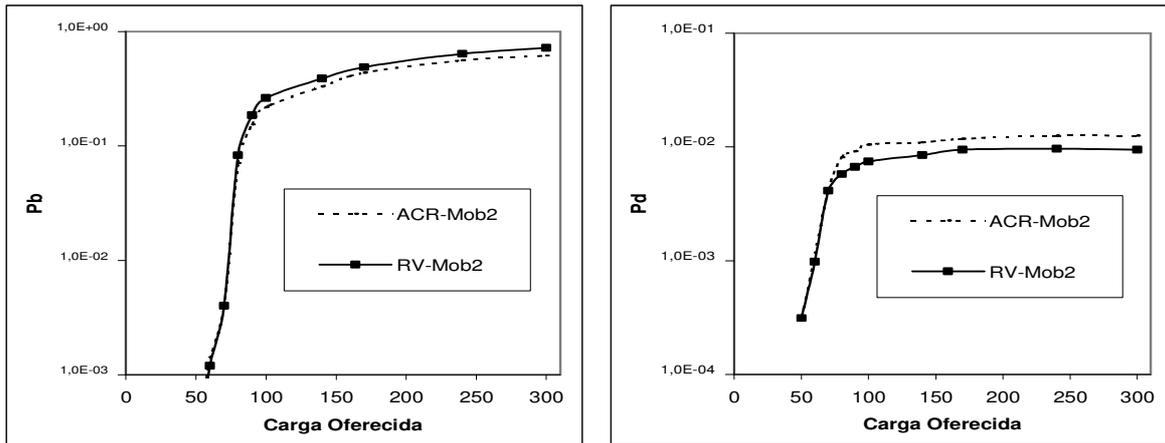
Tabela 3 – Reservas Antecipadas Imediatas vs. RV.

Esquema	Pd em 300 BUs		Utilização média em 100 BUs	
	Mob1	Mob2	Mob1	Mob2
RV (Proposta)	0,92%	0,95%	75%	76%
RA-simples	4,2%	5,2%	70%	74%
RA-dupla	0,04%	0,9%	59%	63%

5.5 Comparação entre as Propostas RV e ACR

Para realizar uma comparação justa com nossa proposta, não utilizamos um limiar fixo para o ACR [Xu 2002]. Ao invés disso, o valor do limiar de tempo do ACR é adaptado da mesma forma como realizado na nossa proposta, isto é, em função do valor medido do Pd na célula atual. Nos experimentos, a faixa de valores usada para o ACR corresponde à mesma utilizada por nossa proposta (valores de Max_LTH , Min_LTH , LTH inicial, Max_Pd e Min_Pd , incremento e decremento do LTH). Caso o valor máximo para o limiar do ACR seja alcançado, e o descarte medido estiver acima de Max_Pd , o ACR realiza reserva dupla para minimizar o descarte em situações de sobrecarga da rede [Dias 2003].

A Figura 4 apresenta os gráficos referentes à comparação de nossa proposta com o ACR modificado, utilizando o ajuste dinâmico do LTH. O valor do intervalo entre atualizações do LTH do ACR é igual ao utilizado por nossa proposta, isto é, 30s. O modelo de mobilidade 2 foi utilizado nas simulações. Para as simulações utilizando o modelo de mobilidade 1 não houve diferença significativa entre as propostas. Como apresentado na Figura 4 (a), à medida que a carga oferecida aumenta, o valor do Pb torna-se maior para o RV do que para o ACR. O RV consegue um controle adequado do nível de congestionamento nas células com a escolha adequada do valor do parâmetro $MaxL$. A consequência de uma menor quantidade de usuários bloqueados pelo ACR é apresentada na Figura 4 (b), onde o seu Pd é maior que o obtido pelo RV. Enquanto o RV garante um valor abaixo de 1%, isto é, com um $Pd = 0,95%$, o ACR extrapola ligeiramente com o valor igual a 1,2%. O movimento aleatório dos usuários prejudicou o ACR em situações com cargas oferecidas excessivamente altas (e.g., 300 BUs).



(a) Bloqueio

(b) Descarte

Figura 4 – Comparação dos esquemas RV e ACR.

A Figura 5 apresenta uma comparação em termos de mensagens de sinalização. O gráfico apresenta os valores normalizados do número de mensagens gerados pelo ACR em relação ao número de mensagens de nossa proposta. A carga de sinalização de nossa proposta refere-se ao número de mensagens trocadas periodicamente a cada 30s entre o servidor de QoS da rede e as estações rádio-base a fim de distribuir informações sobre as cargas direcionais. As mensagens consideradas para o ACR correspondem ao pedido de reserva quando o LTH for alcançado e ao cancelamento de reservas devido à mudança de trajetória e finalização de chamadas. Os parâmetros dos esquemas são os mesmo utilizados pelos cenários de mobilidade determinística (Mob1) e aleatória (Mob2) utilizados nas simulações anteriores.

Nossa proposta reduz consideravelmente a quantidade de mensagens de sinalização quando comparada ao ACR. O ACR tem um sobrecarga de sinalização que chega a ser 18,7 vezes maior que a gerada pela troca periódica da proposta RV no cenário utilizando o modelo *Mob1*. A sobrecarga de sinalização do ACR é 39,6 vezes maior que a do RV quando o cenário utiliza o modelo de mobilidade 2. Dessa forma, considerando os resultados de simulações discutidas anteriormente, podemos concluir que nossa proposta obtém níveis similares de desempenho em termos das métricas P_d , P_b , utilização média sem a sobrecarga excessiva de sinalização gerada pelo ACR, que realiza reservas antecipadas por usuário.

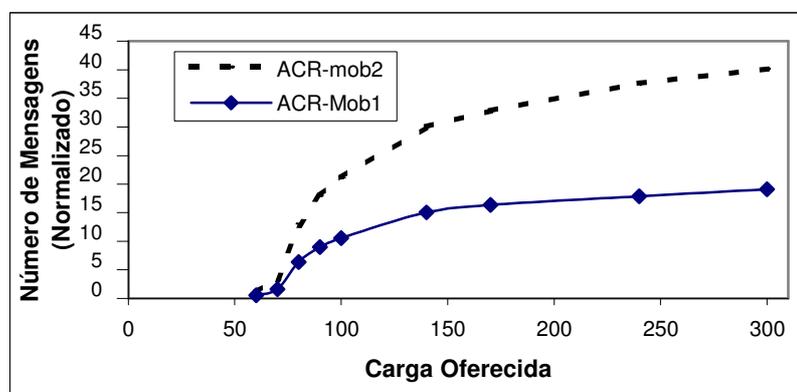


Figura 5 – Número de mensagens (#ACR/#RV)

6. Conclusão e Trabalhos Futuros

Neste artigo, propusemos um esquema para provisionamento de QoS em redes móveis. Nossa proposta é compatível com os atuais arcabouços escaláveis de QoS para a Internet, pois reduz o *overhead* de sinalização de reservas antecipadas por usuário e os bloqueios desnecessários devido a reservas não utilizadas.

Os resultados de simulações mostraram que nossa proposta obteve o melhor compromisso em termos das métricas analisadas na comparação com os resultados obtidos por esquemas fixos e dinâmicos baseados em reservas antecipadas por usuário. Além disso, a proposta consegue manter os níveis de probabilidade de descarte de chamadas de *handoff* abaixo de limites especificados, independente da mobilidade do usuário e sem degradar a utilização da rede com bloqueios excessivos de novas chamadas. Nossa proposta também reduz consideravelmente a quantidade de mensagens de sinalização quando comparada aos esquemas dinâmicos baseados em reservas antecipadas.

Como trabalhos futuros, pretendemos estender este estudo para o nível de pacotes IP, avaliando outras métricas de QoS como *atraso* e *jitter*. Além disso, estamos elaborando uma arquitetura onde o mecanismo proposto neste artigo deve ser um dos componentes.

Referências

- Hui, S. Y., Yeung, K. H., "Challenges in the Migration to 4G Mobile Systems", *IEEE Communications Magazine*, pp. 54-59, December 2003.
- Hong, D. and Rappaport, S., "Traffic Model and Performance Analysis for Cellular Mobile Radio Telephone Systems with Prioritised and Nonprioritised Handoff Procedures," *IEEE Transactions on Vehicular Technology*, 35(3), pp. 77-92, Aug. 1986.
- Choi, S. and Shin, K. G., "Adaptive Bandwidth Reservation and Admission Control in QoS-Sensitive Cellular Networks," *IEEE Transactions on Parallel and Distributed Systems* 13(9), pp. 882-897, Sep. 2002.
- Oliveira, C., Kim, J. B., and Suda, T., "An Adaptive Bandwidth Reservation for High-Speed Multimedia Wireless Networks," *IEEE J. on Selected Areas in Communications*, 16(6), pp. 858-874, Aug. 1998.
- Wu, S., Wong, K. Y. M., and Li, B., "A Dynamic Call Admission Policy With Precision QoS Guarantee Using Stochastic Control for Mobile Wireless Networks," *ACM/IEEE Transactions on Network*, 10(2) pp. 257-271, April 2002.
- Jiang, S., B. et al., "A Modified Distributed Call Admission Control Scheme and Its Performance," *Journal of Wireless Networks*, 7(2), pp. 127-138, 2001.
- Chiu, M.H., and Bassiouni, M. A., "Predictive Schemes for Handoff Prioritization in Cellular Networks based on Mobile Positioning," *IEEE JSAC*, 18(3), pp. 510-522, Mar. 2000.
- Xu, Z. et al. "A New adaptive channel reservation scheme for handoff calls in wireless cellular networks," *Proc. of IFIP Networking2002*. pp 672-684.
- Soh, W.-S., and Kim, H. S., "QoS Provisioning in Cellular Networks Based on Mobility Prediction Techniques," *IEEE Comm. Mag.*, pp 86-92, Jan. 2003.
- Dias, K. L. Kelner, J., and Sadok, D., "A Hierarchical Resource Management Approach for IP-based Wireless and Mobile Networks", *IEEE Vehicular Technology Conference (VTC)*, Orlando, Florida –USA, October 2003.
- Zhang, Z-L. et al., "On Scalable Design of Bandwidth Brokers.," *IEICE Transactions on Communications*, E84-B(8), Aug. 2001.