

# Controle de Admissão e Reserva de Recursos em Redes Móveis utilizando Previsão de Séries Temporais<sup>†</sup>

Kelvin L. Dias, Stênio F. L. Fernandes, Djamel F. H. Sadok

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

{kld, sflf, jamel}@cin.ufpe.br

**Abstract.** *This paper proposes a novel call admission control (CAC) and reservation scheme for wireless mobile networks that avoids per-user reservation signalling. To this end, we propose the use of time series-based models for predicting handoff load. These methods are executed locally by each base-station or access router and forecast how much bandwidth should be reserved on a periodic time window basis. In addition, depending on the schemes settings, the prediction methods can grant an upper bound on handoff dropping probability even under very high load scenarios. The proposal is also improved with an adaptive approach to achieve a better bandwidth utilization.*

**Resumo.** *Este artigo propõe um novo esquema para controle de admissão de chamadas (CAC) e reserva de recursos em redes móveis sem fio que evita a sinalização de reservas antecipadas. Utilizamos previsão local do tráfego de handoff com base no uso de métodos de previsão para séries temporais. O CAC é realizado localmente em cada célula e a aceitação ou não de um novo usuário depende da disponibilidade de recursos não reservados na célula em questão. Dependendo das configurações do esquema, pode-se garantir que limites máximos para a probabilidade de descarte de handoffs não sejam violados. Para melhorar a utilização dos recursos da rede propomos um esquema adaptativo para diminuir o bloqueio de novas chamadas para cargas moderadas.*

## 1. Introdução

Pesquisas sobre as futuras redes móveis sem fio de quarta geração (4G) são especialmente desafiadoras quando o tópico envolvido é a provisão de qualidade de serviço (QoS – *Quality of Service*). As aplicações dos usuários móveis podem sofrer degradações de desempenho devido às propriedades dos enlaces sem fio e devido ao processo de *handoff* [3][4]. *Handoff* é o mecanismo que transfere uma chamada em andamento de uma célula para outra à medida que o usuário se movimenta através da área de cobertura do sistema. Caso a nova célula para onde o usuário acabara de migrar não possua capacidade suficiente para acomodar os requisitos de banda da aplicação, a chamada será descartada. Da perspectiva do usuário, descartes de chamadas (*dropped calls*) em andamento são menos desejáveis do que o bloqueio de novas chamadas. Dessa forma, a literatura considera a probabilidade de descarte para chamadas de *handoff* (HDP – *Handoff Dropping*

*Probability*) como uma métrica de QoS no nível de chamada que deve ser garantida [3][4][8].

O controle de admissão de novas chamadas (*CAC – Call Admission Control*) e reserva de recursos são fundamentais para garantir um limite superior no HDP. O CAC deve ser utilizado para atacar o aspecto de mobilidade, decidindo se um novo usuário deve ser admitido ou rejeitado. Para que o usuário seja admitido, tanto o novo usuário deverá receber garantias de manutenção dos requisitos de QoS para sua aplicação, bem como sua admissão não deverá degradar o desempenho das aplicações já presentes na rede. Vários esquemas para CAC e reserva de recursos foram propostos na literatura para tratar o efeito da mobilidade do usuário em redes sem fio [2][3][4][5][6][7][8][9][10]. Em geral, as propostas baseiam-se na idéia de priorizar chamadas de *handoffs* por meio de reserva antecipada nas células vizinhas àquela onde o usuário encontra-se atualmente.

Um outro aspecto importante que deve ser levado em consideração no projeto destas redes é a escalabilidade na provisão de QoS em termos de estados mantidos nos elementos da rede e sinalização excessiva. Em redes fixas, defendem-se soluções escaláveis que evitem a reserva antecipada por usuário [1]. No contexto das redes móveis, a reserva de recursos por usuário tem efeitos ainda mais negativos devido à escassez de largura de banda de seus enlaces sem fio e bloqueios desnecessários de novas chamadas gerados por reservas antecipadas e não utilizadas. Além disso, as propostas de CAC em redes móveis são, em geral, distribuídas, com a troca excessiva de mensagens entre os pontos de acesso a fim de coletar informações sobre carga da vizinhança [3][5][8][9].

Este artigo descreve uma proposta para CAC e reserva de recursos que evita totalmente a troca de mensagens entre os elementos da rede e a reserva de recursos por usuário. Esta proposta baseia-se na previsão da carga de *handoffs* através da coleta local de informações sobre o tráfego em intervalos regulares. Para isto, é proposto o uso de modelos para previsão de séries temporais para prever a quantidade necessária de recursos de banda para suportar a demanda de *handoffs* em cada célula. Utilizamos a metodologia de alisamento exponencial adaptativo, também denominado *Trigg and Leach*, e modelos ARIMA para estimar periodicamente a quantidade de banda necessária para suportar a carga de *handoffs* para uma janela de tempo futura. O CAC é realizado localmente, simplesmente, verificando-se a quantidade de recursos que não foi reservada para decidir pela admissão de um novo usuário.

Este artigo está organizado da seguinte forma: a Seção 2 apresenta os trabalhos relacionados. Na Seções 3 e 4 descrevemos a metodologia de alisamento exponencial adaptativo (*Trigg and Leach*), modelos ARIMA e a metodologia *Box-Jenkins*. O algoritmo para controle de admissão e reserva de recursos através da previsão de séries temporais é apresentado na Seções 5 e 6. A Seção 7 apresenta as avaliações baseadas em simulações do uso das duas metodologias. Finalmente, na Seção 8 apresentamos as considerações finais.

## **2. Trabalhos Relacionados**

De uma maneira geral, a idéia básica das estratégias para controle de admissão existentes 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* [2][3][6][9][11].

Estratégias fixas, como o *canal de guarda* (GC - *Guard Channel*) [2], 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 rádio-base. Por outro lado, a abordagem não é flexível para tratar cargas variáveis de tráfego, uma vez que não utiliza informações sobre a carga da célula atual e das vizinhas.

As propostas em [5][8] levam 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. Reservas em grupos de células são considerados em [9] e [10], mas a excessiva sinalização tornam-nos proibitivos, além de serem propostas de difícil implementação em ambientes reais.

Recentemente, o uso do GPS (*Global Positioning System*) tem sido defendido como uma abordagem promissora 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. Além disso, esta abordagem pode evitar a degradação da utilização da capacidade da rede devido às reservas de recursos desnecessárias [4][6][7]. A idéia básica 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*. Dessa forma, mais usuários poderão ser admitidos na rede, aumentando a sua utilização. Apesar de proporcionarem uma boa utilização dos recursos, estas propostas sofrem com altos HDPs quando empregadas em redes congestionadas [14][15].

Na proposta descrita em [12], os autores propõem um método local de reservas preditivas para *handoffs* baseada no processo de *Wiener*, que corresponde a um processo Markov onde apenas os valores atuais são relevantes para a previsão um passo a frente. Além disso, propõem uma metodologia para garantir um limite superior no HDP, que consiste em utilizar o limite superior do intervalo de confiança para o valor predito, como a quantidade a ser reservada para futuros *handoffs* durante um determinado intervalo de tempo. Por exemplo, se o máximo HDP permitido é 5%, a reserva deve corresponder ao limite superior do intervalo de confiança para o valor predito considerando um nível de confiança de 95%. Os autores, também, sugerem um modelo ARIMA (*Autoregressive Integrated Moving Average*) e mostram que o ARIMA proporciona uma previsão ligeiramente melhor do que o processo de *Wiener*, através da comparação de dados coletados a partir de simulações de tráfego de *handoff* em um cenário de uma única célula. Entretanto, os resultados preliminares do artigo, com relação às métricas de probabilidade de bloqueio de chamadas e probabilidade de descartes para *handoffs*, são apresentados apenas para o método de previsão baseado no processo de *Wiener*. Em [13], mostra-se que esta metodologia pode causar reservas desnecessárias e bloqueios excessivos de novas chamadas. Em outras

palavras, apesar de um dos métodos de previsão ser aparentemente mais preciso, não foi demonstrada sua eficácia na melhoria das métricas de probabilidade de bloqueio de chamadas e probabilidade de descartes para *handoffs*.

### 3. Alisamento Exponencial e suas Variantes

A técnica de alisamento exponencial tem sido por muito tempo o método escolhido para a previsão univariada devido sua precisão e facilidade de uso, sendo bastante atrativa devido ao seu baixo esforço computacional. A técnica é recomendada para previsão de curto prazo. Destacamos a seguir, sua forma mais simples, o alisamento exponencial e, em seguida, o procedimento chamado *Trigg & Leach* que utiliza uma abordagem adaptativa do alisamento exponencial simples.

#### 3.1. Alisamento Exponencial Simples

Seja  $Y_t$  uma série temporal univariada. O alisamento exponencial simples assume que a previsão  $\hat{Y}$  para o período  $t+h$  é dada pelo nível da variável  $\hat{a}$  no período  $t$

$$Y_{t+h} = \hat{a}_t,$$

que é recursivamente estimada pela média ponderada dos valores observados e preditos de  $Y_t$ .

$$\hat{a}_t = \mathbf{a}Y_t + (1-\mathbf{a})\hat{Y}_t,$$

$$\hat{a}_t = \mathbf{a}Y_t + (1-\mathbf{a})\hat{a}_{t-1}$$

onde  $0 < \mathbf{a} < 1$  é conhecido como parâmetro ou constante de alisamento. A principal desvantagem desta técnica é a escolha do parâmetro de alisamento, pois o configurando próximo de 1 pode-se proporcionar modelos altamente reativos. Por outro lado, a escolha da constante de alisamento com valor próximo a 0 pode gerar um modelo menos sensível a mudanças no padrão da série.

#### 3.2. Alisamento Exponencial Adaptativo: Trigg & Leach

Para assistir a seleção do parâmetro  $\mathbf{a}$  e melhorar a sensibilidade do preditor com relação às variações no padrão da série, métodos adaptativos têm sido recomendados na literatura [20]. Um dos mais representativos e amplamente difundido é o método de alisamento exponencial adaptativo chamado *Trigg & Leach* (TL) [19]. Sua principal vantagem baseia-se no fato de que não há necessidade de uma especificação prévia do parâmetro de alisamento. O procedimento TL pode regular o parâmetro de alisamento  $\mathbf{a}$  sempre que uma mudança ocorrer na estrutura básica da série temporal. Seja  $\mathbf{a}_{t+1}$  o parâmetro de alisamento um passo à frente, então, a previsão em  $t+1$  para o nível da série é dada por:

$$\hat{a}_{t+1} = \mathbf{a}_t Y_t + (1-\mathbf{a}_t) \hat{Y}_t \text{ e } \mathbf{a}_{t+1} = \left| \frac{E_t}{M_t} \right|$$

$$\text{onde } E_t = \mathbf{b}e_t + (1-\mathbf{b})E_{t-1}, M_t = \mathbf{b}|e_t| + (1-\mathbf{b})M_{t-1} \text{ e } e_t = Y_t - \hat{Y}_t.$$

#### 4. Modelos ARIMA e a Metodologia Box & Jenkins

Existem algumas abordagens clássicas para modelagem de séries temporais estacionárias, entre elas os modelos autorregressivo (AR – *Auto-regressive*), Médias Móveis (MA – *Moving Average*) e Autorregressivo de Médias Móveis (ARMA – *Autoregressive Moving Average*). Particularmente, processos que após a aplicação de um número finito  $d$  de diferenças reduzem-se a modelos ARMA são denominados de modelos ARIMA( $p, d, q$ ). A aplicação da diferença em séries temporais é um método para transformar uma série temporal não estacionária em outra que seja estacionária. Um modelo ARIMA( $p, d, q$ ) pode ser representado por:

$$\Delta^d X_t = \mathbf{m} + \mathbf{f}_p \Delta^d X_{t-p} + \mathbf{e}_t + \mathbf{q}_q \mathbf{e}_{t-q}, \quad (1)$$

onde  $\Delta^d$  indica que a ordem da diferenciação é  $d$ .

Observando mais detalhadamente a equação 1, verifica-se a necessidade de se encontrar uma forma de estimar os valores de  $\mathbf{q} \equiv (c, \mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_p, \mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_q)$ , com base nas observações em  $\{X_t\}$ . Uma técnica de inferência usual para a estimação é a máxima verossimilhança (*Maximum Likelihood*- ML). Dada uma amostra de tamanho T, o primeiro passo seria calcular a função de verossimilhança (*likelihood function* - LF),  $L(\mathbf{q}; x)$ . Esta função pode ser encontrada através do cálculo da densidade de probabilidade  $f_{x_T, x_{T-1}, \dots, x_1}(x_T, x_{T-1}, \dots, x_1; \mathbf{q})$  que está fortemente relacionada com a suposição de que a distribuição particular para o processo de ruído branco  $\mathbf{e}_t$  assume uma forma de ruído branco Gaussiano, i.e.,  $\mathbf{e}_t \sim i.i.dN(0, \mathbf{S}^2)$ . Comumente, utiliza-se o log-verossimilhança condicional ou reduzida  $l(\mathbf{q}; x) \propto \ln L(\mathbf{q}; x)$ , onde a LF tem a forma

$$L(\mathbf{q}; x) = f_{x_1}(x_1; \mathbf{q}) \prod_{t=1}^T f_{x_t | x_{t-1}}(x_t | x_{t-1}; \mathbf{q}).$$

O sistema de equações dado por  $\nabla l(\hat{\mathbf{q}}) = 0$  é usualmente referida como equações de verossimilhança. Como normalmente não há uma solução explícita um algoritmo de maximização numérica deve ser empregado. Neste trabalho foi utilizado o algoritmo *Broyden-Fletcher-Goldfarb-Shanon* (BFGS).

A metodologia Box & Jenkins tenta prover um procedimento flexível de forma a se obter modelos ARIMA adequados e de alta qualidade. A metodologia consiste de três estágios básicos: identificação, estimação de parâmetro e verificação de diagnóstico. No primeiro estágio, um modelo de temporário é normalmente selecionada com base na função de autocorrelação ou correlograma da amostra, que tenta identificar as ordens de  $p$  e  $q$  do processo ARIMA. Dada uma série temporal, o primeiro estágio pode recomendar várias especificações (i.e., ordens  $p$  e  $q$ ), cada uma satisfazendo algumas verificações de diagnóstico. Por esta razão, algum tipo de medida de *qualidade do ajuste* é necessária para decidir sobre os melhores modelos apresentados. Neste artigo o critério adotado foi o AIC (*Akaike Information Criterion*). A estimação de máxima verossimilhança explicada anteriormente é realizada durante o segundo estágio da metodologia.

Neste artigo, automatizamos a metodologia Box & Jenkins para identificar, estimar e realizar a verificação de diagnóstico para a carga dos *handoffs* em cada célula na topologia adotada em nossas simulações. Utilizamos um intervalo de tempo para amostras de 30 ou 60s e coletamos as primeiras 30 amostras (denominado de período de treinamento do modelo) antes de iniciar o procedimento automatizado Box & Jenkins. As 30 amostras de treinamento representam uma quantidade suficiente para alcançar convergência na estimativa ML. Depois do período de treinamento, para cada nova carga de *handoff* medida em cada célula durante um intervalo de medição, realizamos todo o procedimento Box & Jenkins novamente. Neste trabalho utilizamos as APIs (*Application Program Interfaces*) do software estatístico *R* v1.8.0 [21] foram utilizadas em conjunto com o simulador de nossa proposta escrito em linguagem C [16].

## 5. CAC e Reserva de Recursos Propostos

Nossa proposta para reserva local de recursos estima a quantidade de largura de banda requerida devido à demanda de futuros *handoffs* baseando-se no TL e ARIMA. O processo de previsão é local, isto é, utiliza apenas a informação coletada localmente por cada estação rádio-base que serve como entrada para o método de previsão sem a necessidade de troca de mensagens entre células vizinhas.

Suponha que uma estação rádio-base conheça a quantidade  $\Omega$  de recursos necessários em termos de largura de banda para chamadas de *handoff* no tempo atual  $t$ . A quantidade de recursos requerida para *handoffs*  $E(\Omega)$  em um tempo futuro  $t + \Delta t$ , pode ser predita com base no  $\Omega$  atual e o valor predito do intervalo de tempo anterior  $t - \Delta t$ . O novo CAC deve determinar se a célula de admissão possui largura de banda disponível suficiente para suporta os requisitos dos usuários e, para isto, considera a carga predita de *handoff* para esta célula. Seja  $y$  a quantidade a ser reservada localmente por uma célula para o próximo intervalo de previsão, a seguinte condição deve ser satisfeita para que o usuário seja admitido:

$$\sum_{i \in A} B_i + B + y \leq C$$

onde  $A$  representa o conjunto de usuários com chamadas em andamento na célula de admissão,  $C$  é a capacidade do enlace sem fio e  $B_i$  representa a largura de banda sendo utilizada pela  $i$ -ésima chamada nesta célula.  $B$  é a largura de banda requerida pelo novo pedido de admissão. No início de cada intervalo de tempo, um novo  $y$  é usado para controlar o processo de admissão. A cada novo chegada de *handoff*, durante um intervalo de previsão, o  $y$  atual é decrementado até o valor 0 ou até que um novo intervalo de previsão seja iniciado, o que ocorrer primeiro. É importante ressaltar que o intervalo de previsão está intrinsecamente associado à quantidade de bloqueios de novas chamadas e garantias de HDP. Dessa forma, intervalos pequenos podem gerar HDPs altos, caso a carga na rede esteja alta. Por outro lado, intervalos grandes contabilizarão mais *handoffs*, capturando a variabilidade do tráfego, e poderão gerar bloqueios excessivos de novas chamadas quando a carga na rede for baixa.

## 6. Adaptação do Intervalo de Medição e Previsão

Estudos sobre previsão de tráfego em redes fixas demonstraram que a escolha adequada do intervalo de medição e previsão para fins de implementação de mecanismos de controle da rede, depende de vários fatores, como precisão desejada, requisito de utilização da rede [23]. Quanto maior o intervalo de previsão, menor é a precisão na previsão do tráfego e, conseqüentemente, o desempenho do esquema de controle pode ficar comprometido. A baixa precisão pode ainda degradar a utilização dos recursos da rede: super-previsão (previsão acima do valor real) para o próximo intervalo de controle pode gerar uma baixa utilização, enquanto a sub-previsão (previsão abaixo do valor real) pode gerar uma boa utilização com um controle menos rígido e, talvez, ineficaz.

No caso das redes móveis, um valor ótimo para o intervalo de medição e previsão depende das características do tráfego e mobilidade dos usuários. A escolha de um valor fixo para este intervalo pode prejudicar a utilização global da rede, uma vez que a quantidade de carga prevista e, conseqüentemente, a ser reservada na célula, é proporcional ao tamanho do intervalo selecionado. Para assistir à seleção deste intervalo, propomos a adoção de um esquema adaptativo que tem como critério o valor medido do HDP em cada célula. O algoritmo abaixo descreve a adaptação do intervalo de medição e previsão.

*Max\_HDP -> Valor máximo para o HDP a partir do qual deve-se aumentar a janela.*

*Min\_HDP -> Valor mínimo do HDP a partir do qual pode-se diminuir a janela*

*Max\_Intervalo -> Valor máximo para o intervalo de previsão*

*Min\_Intervalo -> Valor mínimo para o intervalo de previsão*

**Ao final de cada Intervalo de Previsão da célula**

*Se (HDP\_Medido  $\geq$  Max\_HDP && Intervalo\_Previsão  $<$  Max\_Intervalo)*

*Incrementar Intervalo\_Previsão;*

*Se (HDP\_Medido  $<$  Min\_HDP && Intervalo\_Previsão  $>$  Min\_Intervalo)*

*Decrementar Intervalo\_Previsão;*

A idéia por trás da adaptação é que, caso o HDP da célula sob o controle do algoritmo esteja acima do valor máximo permitido (*Max\_HDP*), o intervalo de medição e previsão deve ser aumentado gradativamente até um limite (*Max\_Intervalo*) para considerar a variabilidade do tráfego, aumentando as reservas e, conseqüentemente, causando um maior bloqueio de novos usuários. Por outro lado, para a célula onde o valor medido do HDP esteja abaixo de *Min\_HDP* (usamos um limite inferior para gerar uma região de estabilidade), reduzimos o intervalo de medição até um limite inferior denominado *Min\_Intervalo*, para que haja uma previsão mais próxima do valor real do tráfego, causando assim, menos bloqueios.

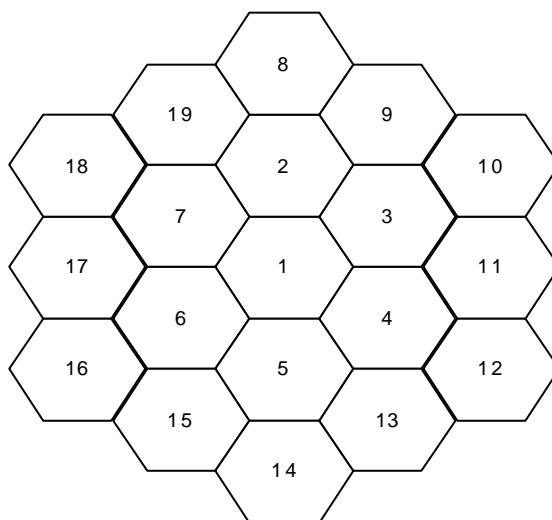
## 7. Avaliação da Proposta

Para representar os requisitos das aplicações utilizamos a mesma nomenclatura dos trabalhos relacionados em unidades de largura de banda (BU - *bandwidth unit*), onde cada unidade representa a largura de banda requerida por uma chamada de voz [3]. O modelo de

tráfego é similar ao utilizado em [3][4]. Chamadas são geradas de acordo com a distribuição de *Poisson* com taxa de  $\lambda$  (chamada/célula/segundo) em cada célula. Os valores simulados para a taxa de chegada de chamadas variam desde uma carga muito baixa (0,05), até situações de congestionamento com uma alta taxa de chegada (1,7), com o intuito de verificar se um HDP máximo não é ultrapassado.

Os modelos de tráfego simulado consistem em usuários com requisitos de 1 BU (voz) e 4 BUs (vídeo) com probabilidades  $R_{vo}$  and  $1-R_{vo}$ , respectivamente, onde  $R_{vo}$  também é denominada de proporção de voz (*voice ratio*) como em [3][4]. Em nossas simulações  $R_{vo}$  é configurada com o valor 0,6, isto é, 60% do tráfego simulado de voz e 40% de vídeo. A duração de cada chamada é exponencialmente distribuída com média 180s [3][4][6].

A Figura 1 apresenta a topologia simulada que consiste de 19 células cada um com capacidade de 100 BUs, utilizando hexágonos para representar as relações de adjacência entre as células como feito em [8]. Para evitar o efeito de borda utilizamos a mesma abordagem adotada na maioria dos trabalhos relacionados da literatura, ou seja, quando um usuário mover-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.



**Figura 1- Topologia Celular.**

O modelo de mobilidade segue esquemas previamente adotados em [6][7]. A cada nova chamada ou *handoff*, o usuário seleciona aleatoriamente uma direção para movimento entre as seis células vizinhas a sua célula atual utilizando uma distribuição uniforme. O tempo de residência pode ser exponencialmente distribuído, como assumido nos trabalhos relacionados, ou seguir uma distribuição de cauda pesada (e.g. Pareto) como apontada nas medições em ambientes de micro-células apresentadas em [22]. Ambas as distribuições para o tempo de residência têm média de 60s. A qualquer momento, enquanto cruza a célula, a MS pode mudar sua direção com probabilidade de 0,5. Se a MS muda sua direção (próxima



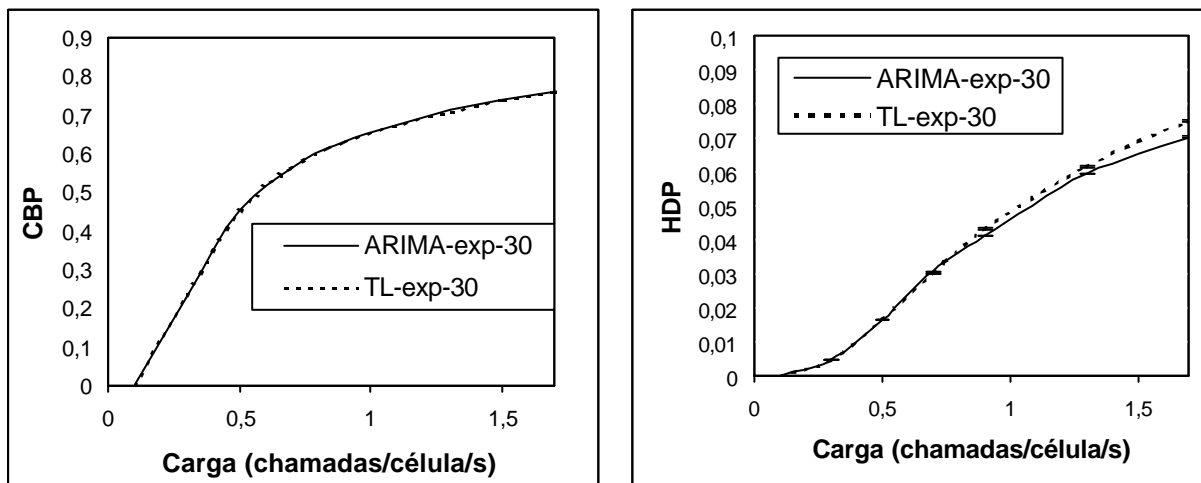
célula a ser visitada), uma nova célula alvo é selecionada, assim como um novo tempo de residência.

### 7.1. Resultados de Simulação

As métricas de interesse neste estudo são: (1) probabilidade de descarte para *handoff* (HDP – *Handoff Dropping Probability*) definido com a razão entre o número de *handoffs* sem sucesso sobre o total de tentativas de *handoffs*, (2) probabilidade de bloqueio de novas chamadas (CBP - *Call Blocking Probability*), isto é, a razão entre o número de chamadas bloqueadas por falta de recursos no momento da admissão pelo número total de pedidos e (3) utilização que corresponde à razão da quantidade de recursos de banda passante em uso no sistema sobre a capacidade total do sistema. Os modelos são rotulados ‘*M-B-T*’ nos gráficos, onde *M* representa o modelo adotado para a previsão (TL ou ARIMA), *B* a distribuição de probabilidade para os tempos de residência na célula (Exponencial ou Pareto). *T* é o intervalo para o qual a previsão é realizada (30 ou 60s).

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. O tempo de simulação foi de 10<sup>6</sup>s. As barras verticais para os valores médios apresentados nos gráficos foram calculadas para representar um intervalo de confiança de 95%.

A Figura 2 mostra comparações para o CBP e HDP utilizando TL e ARIMA e considerando um intervalo de previsão de 30s. Ambos os modelos obtiveram níveis similares de CBP. A comparação com relação à métrica HDP mostra que o TL obteve um HDP ligeiramente maior que o ARIMA para cargas altas. Este cenário indica que o modelo TL pode alcançar previsões satisfatórias comparáveis aos modelos mais sofisticados (ARIMA).



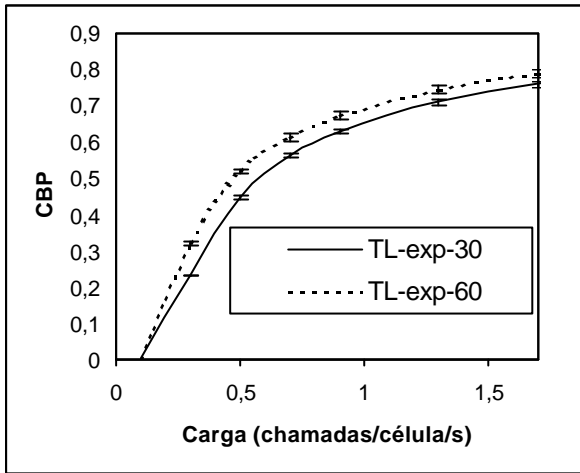
(a) Probabilidade de Bloqueio de Novas Chamadas.

(b) Probabilidade de Descarte para Handoffs.

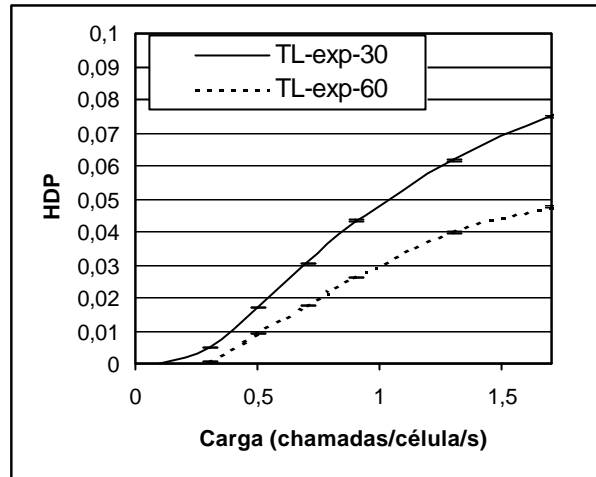
Figura 2- Intervalo de Previsão igual a 30s.

A Figura 3 e Figura 4 mostram os resultados obtidos para os métodos TL e ARIMA, utilizando intervalos de previsão de 30s e 60s. Quando os métodos utilizam 60s, ambos

conseguem manter o HDP abaixo do valor alvo de 5%. O modelo ARIMA alcança o menor HDP em ambos os cenários (30s e 60s). Novamente, a escolha apropriada do intervalo de previsão é um compromisso entre o limite máximo desejável para o HDP e a utilização eficiente dos recursos de banda passante da rede.

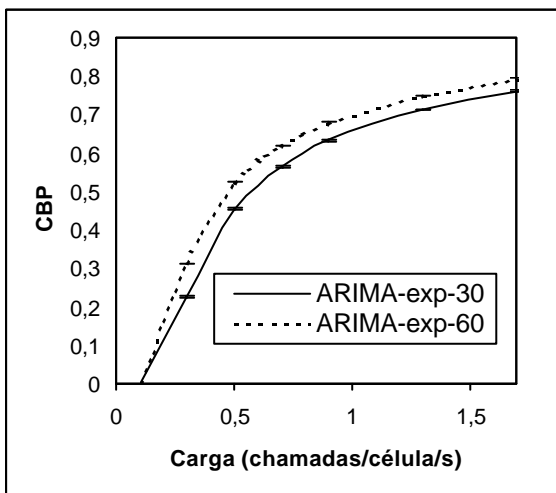


(a) Prob. De Bloqueio de Novas Chamadas.

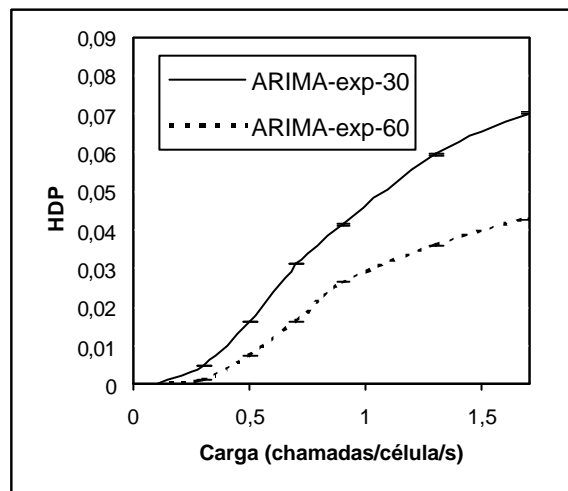


(b) Prob. De Descarte para Handoffs.

Figura 3– Trigg & Leach – Intervalos de previsão iguais a 30s e 60s.



(a) Prob. de Bloqueio de Novas Chamadas.



(b) Prob. de Descarte para Handoffs.

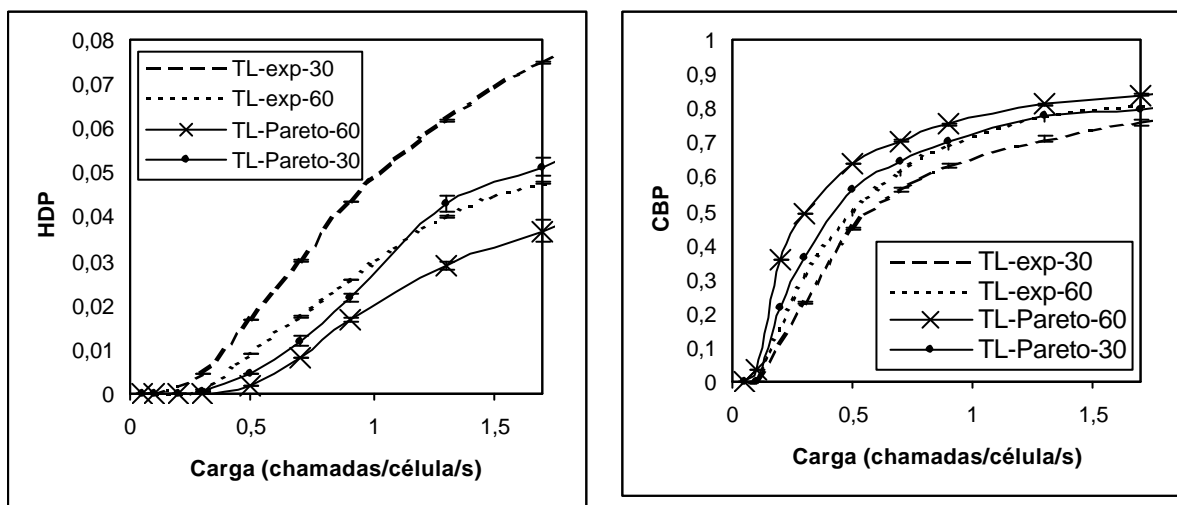
Figura 4–ARIMA – Intervalos de Previsão iguais a 30 e 60s.

Com base nos resultados discutidos anteriormente, uma opção viável para utilização prática pelas operadoras seria a previsão baseada no modelo TL, uma vez que este é computacionalmente simples e pode garantir limites máximos do HDP similares aos obtidos pelo modelo ARIMA. Modelos ARIMA são interessantes para previsões de longo prazo, como para planejamento de capacidades em redes fixas de alta velocidade onde a escala de tempo pode compreender dias [18]. No caso de previsão de *handoffs*, onde devem ser consideradas escalas de tempo menores (segundos ou minutos), o modelo ARIMA pode não

ser útil para a realização de previsões de carga, devido à complexidade e tempos necessários para sua estimação, além da necessidade em se ter uma quantidade maior de dados armazenados para a fase de treinamento.

A seguir, utilizamos apenas o modelo TL para avaliar o impacto do uso distribuição Pareto na geração dos tempos de residência para diferentes intervalos de previsão. Trabalhos futuros devem considerar o estudo do impacto deste novo cenário quando o modelo ARIMA é utilizado.

A Figura 5 (a) e (b) apresentam resultados da comparação do HDP e CBP para dois intervalos de previsão considerando o tempo de residência com distribuições *Pareto* e Exponencial. Pode-se perceber que os valores de HDP obtidos utilizando *Pareto* são menores que os do cenário utilizando a distribuição exponencial. Isto ocorre devido à característica da distribuição *Pareto* que gera *handoffs* mais freqüentes que a exponencial. Os tempos médios de residência considerando a distribuição exponencial oscilam em torno da média, enquanto que a distribuição *Pareto* gera muitos valores abaixo da média e alguns poucos muito acima, uma vez que, dependendo de sua parametrização, sua variância tende ao infinito. Dessa forma, com tempos de residência pequenos, a quantidade de *handoffs* prevista tende a aumentar porque usuários estão efetuando mais *handoffs* o que gera mais bloqueios de novas chamadas pela maior quantidade de reserva para estes *handoffs* mais freqüentes durante o intervalo de tempo escolhido.



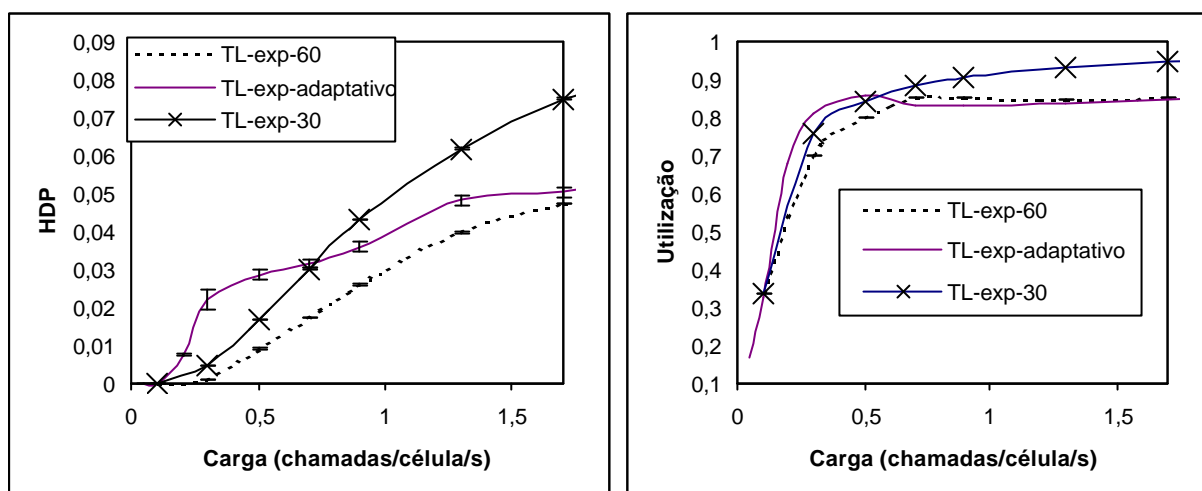
(a) Prob. de Descarte para *Handoffs*.

(b) Prob. de Bloqueio de Novas Chamadas.

Figura 5. Tempo de residência exponencial vs. *Pareto*

A Figura 6 apresenta o emprego do esquema adaptativo considerando o tempo médio de residência exponencial que, como mostrado nos gráficos anteriores, gerou HDPs maiores que os do cenário que utiliza a distribuição *Pareto*. O esquema adaptativo é comparado com os resultados de simulações que usam intervalos fixos de 30s e 60s. Os valores de *Max\_HDP* e *Min\_HDP* foram iguais a 5% e 2%, respectivamente. Estes valores devem ser escolhidos para gerar uma região de operação estável entre *Max\_HDP* e *Min\_HDP*.

*Max\_Intervalo* e *Min\_Intervalo* foram configurados com valores iguais a 60s e 10s, respectivamente. O valor inicial do intervalo de previsão para todas as células é de 30s. Os valores dos incrementos e decrementos para os intervalos de previsão foram 5 e 1, respectivamente. O valor de incremento igual a 5 foi utilizado para que, caso o valor máximo permitido para o HDP seja ultrapassado, o intervalo de previsão cresça rapidamente evitando que mais usuários entrem na célula já congestionada, por outro lado, o valor unitário para o decremento proporciona um ajuste cauteloso na permissão de mais usuários na célula.



(a) Prob. de Descarte para *Handoffs*.

(b) Utilização.

Figura 6 –Esquema Adaptativo.

Note que a utilização para as cargas abaixo de 0,5 chamadas/célula/s, é sempre superior para o esquema adaptativo. Considerando a carga de 0,29 chamada/célula/s, temos uma utilização média igual a 80,9% para o esquema adaptativo, 75,6% para o esquema utilizando um intervalo de 30s e 69% para o que utiliza 60s. Ou seja, o esquema adaptativo, supera em 7% o esquema que utiliza intervalo de 30s e em 17% o esquema que utiliza intervalo de 60s.

Para taxas de chegadas mais altas, acima de 0,5 chamadas/s/célula, o esquema fixo com 30s ultrapassa a utilização dos demais, no entanto, seu HDP extrapola o valor limite de 5% como pode ser verificado na Figura 6(a). O esquema adaptativo tem o maior HDP para cargas baixas, mas manteve-se dentro do valor máximo desejável de 5%, assim como o esquema fixo de 60s.

## 8. Conclusão

Neste artigo propomos um novo esquema para controle de admissão local com reserva de recursos baseado em previsão de séries temporais para redes móveis sem fio. A proposta evita a troca de mensagens entre os elementos da rede características de propostas distribuídas, bem como evita o *overhead* da sinalização por usuário para a requisição e efetivação de reservas antecipadas.

Utilizamos a técnica de alisamento exponencial adaptativo (*Trigg and Leach*- TL) e a metodologia *Box-Jenkins* na determinação de modelos ARIMA para a previsão da quantidade de banda passante necessária para o tráfego de *handoff*. Apesar da simplicidade e do baixo esforço computacional requerido pelo esquema baseado no *Trigg and Leach*, foi possível obter resultados bem próximos aos obtidos pelo esquema baseado em modelos ARIMA. Modelos ARIMA requerem um maior esforço computacional para realização da seleção do modelo e previsão da demanda futura de *handoffs*. Assim, em se tratando de previsões de curto prazo, isto é, intervalo de minutos ou segundos, modelos ARIMA não são tão vantajosos quanto modelos mais simples baseados no método *Trigg and Leach*. Neste artigo, também propomos um esquema adaptativo que varia o intervalo de previsão em função do HDP medido em cada célula. Isto permite que a utilização global da rede seja otimizada e, ao mesmo tempo, que os limites superiores para o HDP sejam mantidos.

Apesar dos trabalhos relacionados considerarem a distribuição exponencial [3][12][5][9] em seus estudos, pesquisas recentes de medição sugerem que outras distribuições, como a Pareto, caracterizam melhor o tempo de residência em ambientes de micro-células [22]. Neste trabalho, avaliamos o método TL com diferentes distribuições para o tempo de residência. Verificou-se que a distribuição Pareto gera mais bloqueios de novas chamadas (CBP) do que a distribuição exponencial para nossa proposta. Em contrapartida, o descarte de *handoffs* (HDP) foi menor para o cenário com tempo de residência Pareto. Isto ocorre devido às características da distribuição Pareto que gera uma grande quantidade de tempos de residência muito abaixo da média e alguns poucos bastante altos, bem acima da média. Já os valores gerados pela distribuição exponencial têm um comportamento menos disperso em torno do valor médio. Dessa forma, com tempos de residência pequenos, a quantidade de *handoffs* prevista tende a aumentar porque usuários estão efetuando mais *handoffs*. Isto gera mais bloqueios de novas chamadas devido a maior quantidade de reserva para estes *handoffs* mais frequentes durante o intervalo de tempo escolhido, segundo nossa proposta.

Como trabalhos futuros pretendemos avaliar novos cenários de tráfego e mobilidade, além de estender a proposta para avaliá-la no nível de pacotes, considerando as métricas de atraso, *jitter* e perda de pacotes IP.

## Referências

- [1] S. Blake *et al.*. An architecture for differentiated services. IETF RFC 2475, Dec. 1998.
- [2] D. Hong and S. Rappaport. Traffic model and performance analysis for cellular mobile radio telephone systems with prioritised and nonprioritised handoff procedures. *IEEE Trans. On Veh. Teh.*, 35(3):77-92 August 1986.
- [3] S. Choi and K. G. Shin, "Adaptive Bandwidth Reservation and Admission Control in QoS-Sensitive Cellular Networks," *IEEE Transactions on Parallel and Distributed Systems* 13( 9), pp. 882-897, September, 2002.
- [4] W.-S. Soh and H. S. Kim, "QoS Provisioning in Cellular Networks Based on Mobility Prediction Techniques," *IEEE Comm. Mag.*, Jan. 2003, pp 86-92.
- [5] M. Naghshineh and M. Schwartz. Distributed call admission control in mobile/wireless networks. *IEEE J. Selected Areas in Commun.*, 14(4):711-717, May 1996.

- [6] M.H. Chiu and M. A. Bassiouni. Predictive schemes for handoff prioritization in cellular networks based on mobile positioning. *IEEE J. Select. Areas Commun.* 18(3), March 2000.
- [7] Z. Xu *et al.* “ A New adaptive channel reservation scheme for handoff calls in wireless cellular networks,” *Proc. of IFIP Networking2002*. pp 672-684, 2002.
- [8] S. Wu, K. Y. M. Wong and Bo Li, “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
- [9] D. A. Levine, I. F. Akyldiz and M. Naghshineh. A resource estimation and call admission control algorithm for wireless multimedia networks using the shadow cluster concept. *IEEE/ACM Trans. Networking*, vol. 5, pp. 1-12, Feb. 1997.
- [10] B. Sadeghi, E. W. Knightly, “Architecture and Algorithms for Scalable Mobile QoS,” *Journal of Wireless Networks*, 9(1). pp. 7-20, 2003
- [11] C. Oliveira, J. Kim, T. Suda. An adaptive bandwidth reservation scheme for high speed multimedia wireless networks. *IEEE J. Select. Areas in Commun.*, 16(6), pp 858-874, 1998.
- [12] T. Zhang *et al.* “Local Predictive Resource Reservation for Handoff in Multimedia Wireless IP Networks,” *IEEE Journal on Selected Areas in Communications*, 19(10), Oct. 2001.
- [13] K L. Dias, S. F. L. Fernandes, D. H. F. Sadok , “Predictive Call Admission Control for All-IP Wireless and Mobile Networks”, In Proceedings of IFIP/ACM Latin America Network Conference, ISBN 1-58113-789-3 p. 131-139, La Paz, Bolívia, Oct. 2003.
- [14] K. L. Dias, J. Kelner and D. F. H. Sadok, “A Hierarchical Resource Management Approach for IP-based Wireless and Mobile Networks”, *IEEE Vehicular Technology Conference (VTC)*, Orlando, Florida – USA, October 2003.
- [15] K. L. Dias and D. F. H. Sadok. A scalable call admission control for QoS provisioning in all-IP Networks. *International Conference on Third Generation Wireless and Beyond (3Gwireless'2003)*. San Francisco, USA, May 27-30, 2003.
- [16] M. H McDougall. *Simulating computer systems: techniques and tools*. The MIT Press. 1987
- [17] W. A. Fuller, “Introduction to statistical time series,” New York: John Wiley & Sons, Inc., 1996.
- [18] D. Papagiannaki, N. Taft, Z.-L. Zhang, C. Diot. "Long-Term Forecasting of Internet Backbone Traffic: Observations and Initial Models". *IEEE Infocom*. San Francisco, March 2003.
- [19] D. W. Trigg, D. H. Leach, “Exponential Smoothing with an Adaptive Response Rate,” *Operational Research Quarterly*, vol. 18, 1967, pp. 53-59.
- [20] D. C Montgomery, L. A. Johnson, and J. S. Gardiner, “Forecasting & time series analysis,” MacGraw-Hill International Editions, 2nd Ed., 1990.
- [21] *R v1.8.0* <http://www.r-project.org/>
- [22] S. Thajchayapong and J. M. Peha, “Mobility Patterns in Microcellular Wireless Networks”, *IEEE Wireless Communications and Networking Conference*, March 2003.
- [23] A. Sang and A. Li, “Predictability Analysis of Network Traffic”, *IEEE INFOCOM'2000*, March 2000.

---

<sup>†</sup> Este trabalho foi parcialmente financiado pela CAPES e CNPq.