

CARACTERIZAÇÃO DO BACKLOG E DO ATRASO EM SERVIDORES GPS SOB TRÁFEGO AUTO-SIMILAR

Flávio de M. Pereira

DECOM - FEEC - Unicamp

Caixa Postal 6101

13083-970 Campinas,SP

e-mail: flavio.pereira@ieec.org

Nelson S. Fonseca

Instituto de Computação - Unicamp

Caixa Postal 6176

13083-970 Campinas,SP

e-mail: nfonseca@ic.unicamp.br

Dalton S. Arantes

DECOM - FEEC - Unicamp

Caixa Postal 6101

13083-970 Campinas,SP

e-mail: dalton@decom.fee.unicamp.br

Resumo

Neste artigo é apresentado um algoritmo para a obtenção de valores limitantes para o *backlog* e para o atraso introduzido no tráfego de fontes auto-similares por um servidor que utiliza como disciplina de serviço o *Generalized Processor Sharing*. Este trabalho é uma extensão do estudo desenvolvido por Parekh e Gallager [9], que obtêm limitantes para fontes de tráfego reguladas pelo algoritmo do Balde Furado. Por outro lado, o presente trabalho considera que tráfego das fontes é regulado pelo algoritmo do Balde Furado Fractal proposto por Viana Neto [2] e desta forma pode ser modelado pelo processo envelope do movimento Browniano fracionário, que considera a efetiva natureza auto-similar do tráfego nas redes de serviços integrados.

1. INTRODUÇÃO

Nos últimos anos, uma significativa parte das pesquisas no segmento das redes de dados têm sido direcionada à busca de tecnologia capaz de viabilizar a convergência entre o transporte de dados e de informação multimídia por meio das denominadas Redes Digitais de Serviços Integrados de Faixa Larga (RDSI-FL) — redes extremamente flexíveis, capazes de suportar uma ampla variedade de aplicações como telefonia, vídeo-conferência, distribuição de áudio e vídeo, comércio eletrônico, ensino à distância, etc.

Uma das questões que ainda impedem a efetiva implementação da RDSI-FL é a dificuldade de se conciliar a flexibilidade da rede e a sua capacidade de fornecer garantias de desempenho aos usuários. Em particular as tecnologias hoje disponíveis mostram-se bastante limitadas neste sentido: um aumento da flexibilidade é quase sempre obtido à custa de uma redução nas garantias de desempenho.

Algumas soluções de compromisso foram obtidas com o advento da comutação por circuitos virtuais, primeiro por meio da tecnologia ATM e depois por meio de adaptações na tecnologia IP. Nos dois casos, a flexibilidade necessária é

obtida pelo uso da comutação por pacotes, empregando-se todavia mecanismos de gerência de tráfego para viabilizar as garantias de desempenho.

Tais soluções de compromisso foram consideradas satisfatórias até a constatação da natureza auto-similar do tráfego multimídia a partir dos estudos realizados por Leland *et al.* na rede do Bellcore [1]. Segundo os diversos estudos que se seguiram, a auto-similaridade tende a degradar de forma significativa o desempenho das redes, e a incapacidade dos modelos de tráfego tradicionais em representá-la sugere que os mecanismos de gerência de tráfego outrora elaborados podem tornar-se completamente ineficientes nos casos em que a auto-similaridade seja relevante.

Em linhas gerais, disciplinas de serviço são algoritmos utilizados em cada nó de uma rede comutada por pacotes para determinar a cada instante qual conexão será atendida, distribuindo assim a capacidade de transmissão entre as conexões. A escolha de uma disciplina de serviço adequada é crucial para que seja possível fornecer garantias de desempenho fim-a-fim como é exigido pelas redes de serviços integrados. Segundo Zhang [6], são basicamente quatro as características a serem consideradas: a disciplina deve ser **eficiente** para suportar um maior volume de tráfego fornecendo as garantias de desempenho exigidas por cada conexão; deve ser capaz de lhes oferecer **proteção** contra excessos de tráfego gerado por outros usuários e pelo tráfego do tipo “melhor esforço”; deve ser **flexível** para suportar aplicações diversificadas; e **simples** para que a seja possível analisá-la e estabelecer relações matemáticas entre os seus parâmetros, os parâmetros de tráfego e as exigências de desempenho de cada conexão.

Neste artigo será estudada uma disciplina com conservação de trabalho baseada em modelos fluidos denominada *Generalized processor sharing (GPS)* que reúne todas aquelas quatro características e que, por isso, mostra-se bastante adequada às redes de serviços integrados. Por basear-se em modelos fluidos, o GPS considera que o tráfego dos usuários é infinitamente divisível, o que é inviável do ponto de vista prático. Existem porém diversas aproximações eficientes que consideram a quantização do tráfego em pacotes,

Este projeto é suportado pela Fundação de Amparo à Pesquisa do Estado de São Paulo, FAPESP, processo 00/00069-0

e para as quais podem ser facilmente obtidas relações de desempenho diretamente a partir da análise do GPS. Uma descrição mais detalhada daquelas aproximações pode ser encontrada em [9], [6] e [5, Cap. 9].

A proposta deste trabalho é obter limitantes superiores para o atraso e o *backlog* em um servidor GPS alimentado por fontes de tráfego auto-similar. Mais especificamente, considera-se que o tráfego de cada conexão é regulado pelo algoritmo do Balde Furado Fractal proposto por Viana Neto em [2] e pode portanto ser modelado pelo processo envelope do movimento Browniano fracionário, a ser apresentado adiante.

Este artigo está organizado da seguinte forma : a seção 2 introduz a notação utilizada; o princípio de operação do GPS é apresentado na seção 3, que também apresenta um breve histórico de trabalhos anteriores relacionados ao problema da análise de desempenho daquela disciplina. Na seção 4 é apresentado o movimento Browniano fracionário e o seu processo envelope, este utilizado como modelo de tráfego no algoritmo proposto na seção 5 para determinar o *backlog* e o atraso introduzidos por um servidor GPS. Na seção 6 é apresentado um exemplo numérico da aplicação do algoritmo proposto, e algumas conclusões são apresentadas na seção 7.

2. NOTAÇÃO

A seguinte notação é adotada ao longo das próximas seções: para cada $1 \leq i \leq N$, \mathbf{A}_i corresponde ao processo de chegada de pacotes da fonte i , e \mathbf{S}_i corresponde ao seu processo de partida. Desta forma, $\mathbf{A}_i(t)$ representa a quantidade de tráfego da fonte i que chega ao servidor até no intervalo $[0; t]$ e $\mathbf{S}_i(t)$ aquela que dele parte no mesmo intervalo. Define-se ainda as funções $A_i(t)$ e $S_i(t)$ como as respectivas realizações. A quantidade de tráfego da fonte i que se encontra em espera no servidor no instante t (i.e. o seu *backlog*) é dada pelo processo $\mathbf{Q}_i(t) = \sup_{\tau \leq t} \{\mathbf{A}_i(\tau, t) - \mathbf{S}_i(\tau, t)\}$, sendo $Q_i(t)$ a sua realização. Finalmente, o processo $\mathbf{D}_i(t)$ descreve o atraso introduzido pelo servidor nos pacotes da fonte i que chegam no instante t , sendo a realização deste processo definida como $D_i(t)$.

3. O GENERALIZED PROCESSOR SHARING

O funcionamento do GPS pode ser compreendido da seguinte forma : seja um servidor GPS que atende o tráfego fornecido por N fontes. A cada conexão i é atribuída uma fila e um fator de ponderação ϕ_i real e positivo, sendo o conjunto $\{\phi_i\}_{1 \leq i \leq N}$ denominado parametrização do servidor GPS (*GPS assignment*). Em um determinado instante τ , a taxa de serviço fornecida a uma conexão i com *backlog* é dada por $\frac{\phi_i}{\sum_{j \in B(\tau)} \phi_j} r$, sendo $B(\tau)$ o conjunto cor-

respondente aos índices das fontes que apresentam *backlog* naquele instante. Pode-se provar que, sendo $S_i(\tau, t)$ o total de tráfego da conexão i servido no intervalo $[\tau, t]$:

$$\frac{S_i(\tau, t)}{S_j(\tau, t)} \geq \frac{\phi_i}{\phi_j}, \quad j = 1, 2, \dots, N \quad (1)$$

para qualquer conexão i que apresente *backlog* durante todo o intervalo $[\tau, t]$. Desta forma, sendo r a taxa total do servidor, cada conexão i tem garantida uma taxa de serviço mínima dada por :

$$g_i = \frac{\phi_i}{\sum_{i=1}^N \phi_j} r, \quad i = 1, 2, \dots, N \quad (2)$$

Por simplicidade, neste artigo considerar-se-á $r = 1$.

Análise de desempenho do GPS

No GPS, o serviço destinado a uma conexão não depende exclusivamente do seu tráfego e a sua relação com o tráfego das demais conexões não pode ser escrito matematicamente de forma simples. Este fato torna a análise daquela disciplina difícil e por isso a maioria dos estudos até agora desenvolvidos são baseados em modelos de tráfego mais simples; e mesmo os poucos trabalhos efetivamente direcionados à sua natureza auto-similar apresentam apenas resultados de caráter preliminar. Por exemplo, podem ser citados os estudos de :

- Dupuis e Ramanan [14], que representam a dinâmica do GPS como um caso particular do Problema de Skorokhod, e provam a existência e unicidade do mapeamento entre os processos que descrevem o tráfego gerado pelas fontes e os processos de *backlog*;
- Parekh e Gallager [9, 10] nos quais mostra-se que, se o tráfego de cada uma das conexões for regulado por um algoritmo de Balde Furado e a taxa total de chegada de pacotes for inferior à taxa de serviço, é possível obter um limitante superior para o atraso introduzido por um servidor GPS no tráfego de uma conexão e também para o seu *backlog*. Além disso, é comprovada a estabilidade das redes compostas por servidores GPS numa configuração denominada *Consistent Relative Session Treatment (CRST)*. Para um caso particular denominado *Rate Proportional Processor Sharing (RPPS)*, demonstra-se que o atraso e o *backlog* não dependem do número de nós entre a fonte e o destino do tráfego, mas apenas do estado de um único nó deste caminho — isto permite obter expressões bastante simples para caracterizar os índices de desempenho das conexões;
- Yaron e Sidi [11, 12] que introduzem a modelagem de tráfego por processos de Explosividade Exponencialmente Limitada (*Exponentially Bounded Burstiness*)

Process, EBB), e provam que se as fontes de um servidor GPS isolado fornecem tráfego EBB, o tráfego de saída do servidor também é EBB. Além disso, é demonstrado que, neste caso, o *backlog* das conexões pode ser modelado por processos de Exponencialmente Limitados (*Exponentially Bounded Process, EB*);

- Zhang *et al.* [13] que, partindo de alguns resultados obtidos por Parekh e Gallager, desenvolvem uma nova abordagem determinística para a análise do atraso e do *backlog* introduzidos por um servidor GPS no tráfego das conexões. Em seguida, esta abordagem é estendida para o caso estatístico, considerando os modelos de tráfego EBB.
- Borst *et al.* [15, 16] que mostram, sob certas condições e do ponto de vista assintótico, que o tráfego gerado por uma fonte de auto-similar é efetivamente servido pelo GPS segundo uma taxa constante. Desta forma, o serviço a ela fornecido só é influenciado pelo tráfego médio das demais conexões, sendo praticamente imune aos seus surtos. Isto comprovaria a eficiência do GPS no tratamento de tráfego auto-similar.

4. O MOVIMENTO BROWNIANO FRACIONÁRIO

O movimento Browniano fracionário (*fractional Brownian motion, fBm*) é um processo estocástico definido inicialmente por Kolmogorov; mais tarde, Mandelbrot e Van Ness [3] o definiram de modo mais formal a partir do já conhecido movimento Browniano puro. A popularização deste processo como modelo de tráfego auto-similar se deve, em grande parte, aos estudos de Norros [4].

De forma bastante simplificada, o fBm pode ser definido como um processo estocástico não-estacionário que possui a seguinte função de autocorrelação:

$$R_{\mathbf{B}_H}(s, t) = \frac{\sigma^2}{2} (|s|^{2H} + |t|^{2H} - |s - t|^{2H}) \quad (3)$$

sendo $\sigma^2 = \text{Var}[\mathbf{B}_H(1)]$. A natureza auto-similar do processo fBm se deve justamente à semelhança entre as distribuições dos processos $\mathbf{B}_H(\rho t)$ e $\rho^H \mathbf{B}_H(t)$.

4.1. Modelagem de tráfego pelo processo fBm

Norros [4] propõe o seguinte processo para a modelagem de fontes de tráfego auto-similar :

$$\mathbf{A}(t) = \rho t + \sigma \mathbf{Z}(t) \quad (4)$$

onde $\mathbf{Z}(t)$ é um processo fBm de média nula, variância $|s|^{2H}$ e com fator de Hurst H . O processo $\mathbf{A}(t)$ é então denominado tráfego Browniano fracionário e pode ser considerado um modelo parcimonioso, pois nele o tráfego é

caracterizado por apenas três parâmetros: $\rho > 0$ é a taxa média de chegada de tráfego; $0,5 \leq H < 1$ é o fator de Hurst do processo $\mathbf{Z}(t)$ e $\sigma > 0$ é um coeficiente associado à variância incremental do processo fBm :

$$\text{Var}[\mathbf{A}(t+s) - \mathbf{A}(t)] = \sigma^2 s^{2H} \quad (5)$$

Um processo importante que está associado a este modelo é o processo de armazenamento do tráfego fBm. Este processo correspondente ao máximo *backlog* de um sistema de filas de um servidor operando com taxa constante g e submetido ao tráfego descrito por (4), ou seja :

$$\delta(t) = \max_{\mathbf{A}(t), t \geq 0} \mathbf{A}(t) - gt$$

sendo que, segundo Norros [4] :

$$\Pr[\delta(t) > x] = \exp \left\{ - \frac{(g - \rho)^{2H}}{2\sigma^2 H^{2H} (1 - H)^{2-2H}} x^{2-2H} \right\} \quad (6)$$

Ou seja, a distribuição de $\delta(t)$ pode ser aproximada por uma distribuição Weibull, em particular para $x \rightarrow \infty$.

4.2. Modelagem de tráfego pelo processo envelope do fBm

Embora o processo fBm seja parcimonioso, sua tratabilidade matemática é, em muitos casos, bastante complicada. Uma alternativa à sua utilização é o emprego do processo envelope correspondente, obtido como generalização do processo envelope do movimento Browniano tradicional.

Seja $\mathbf{A}(t)$ um processo fBm com média ρ e variância incremental dada por (5), sendo H o seu fator de Hurst. Pode-se definir o processo envelope correspondente a $\mathbf{A}(t)$ como:

$$\hat{\mathbf{A}}(t) = \rho t + k\sigma t^H \quad (7)$$

A probabilidade de violação do processo envelope por $\mathbf{A}(t)$ é dada então por :

$$\Pr[\mathbf{A}(t) > \hat{\mathbf{A}}(t)] = \bar{\Phi}(k) \quad (8)$$

onde $\bar{\Phi}(k)$ é a função distribuição residual associada à distribuição gaussiana, dada por:

$$\bar{\Phi}(k) \approx \frac{1}{\sqrt{2\pi(1+k)}} \exp\left(-\frac{k^2}{2}\right) \approx \exp\left(-\frac{k^2}{2}\right) \quad (9)$$

Algumas características relacionadas ao processo envelope fBm são apresentadas por Viana Neto [2]. Também é possível obter um processo envelope para o processo de armazenamento fBm, o definindo como o *backlog* de um sistema de filas de um servidor, operando com taxa constante g e submetido ao tráfego descrito por 7 e 8. Este processo, denominado $\hat{\delta}(t)$, é dado por :

$$\hat{\delta}(t) = (\rho - g)t + k\sigma t^H \quad (10)$$

Além disso, definindo $\delta^*(t) = \max_{t \geq 0} \hat{\delta}(t)$ tem-se que [2]:

$$\delta^*(t) = \left(\frac{g - \rho}{k\sigma} \right)^{\frac{1}{H-1}} \quad (11)$$

$$\Pr[\delta(t) > \delta^*(t)] \leq \overline{\Phi}(k) \quad (12)$$

4.3. O algoritmo do Balde Furado Fractal

Após a admissão de uma conexão, são empregados mecanismos de policiamento para garantir que o tráfego gerado pelo usuário esteja de acordo com a descrição previamente negociada com a rede. Para este propósito, o mecanismo mais utilizado é o conhecido algoritmo do Balde Furado, que restringe o tráfego do usuário ao seguinte envelope :

$$\widehat{A}(t) = \rho_{BF}t + \sigma_{BF}$$

onde ρ_{BF} é a taxa de geração de permissões e σ_{BF} o tamanho do balde. Sabe-se, no entanto, que tais parâmetros são insuficientes para caracterizar fontes que transmitam dados em rajadas, como é o caso das fontes auto-similares. Neste caso, se a taxa de geração de permissões for muito próxima da taxa média de transmissão da fonte, o policiamento pode eventualmente descartar células bem-comportadas; por outro lado, se a taxa de geração for ajustada segundo a taxa de pico da fonte, reduz-se o ganho da multiplexação estatística. Além disso, o aumento do tamanho do balde como forma de absorver as rajadas não é eficiente: este parâmetro tende a assumir valores muito grandes, o que pode permitir a entrada de longas rajadas na rede.

O problema do algoritmo do Balde Furado está na implícita consideração que o tráfego acumulado é uma função linear no tempo. Como isto não é válido no caso do tráfego auto-similar, este mecanismo não é adequado ao seu policiamento. Como uma nova proposta, Viana Neto [2] apresenta um algoritmo denominado Balde Furado Fractal, que restringe o tráfego a um envelope equivalente ao do movimento Browniano fracionário, dado pela (7).

Uma descrição mais detalhada do algoritmo pode ser encontrada em [2]; cabe ressaltar no entanto sua parcimônia e eficiência no policiamento de fontes auto-similares, fatos que justificam seu uso neste trabalho.

5. ANÁLISE DO GPS A PARTIR DA CURVA DE SERVIÇO

Nesta seção é proposta uma técnica para avaliar o valor máximo do atraso e do *backlog* em um servidor GPS alimentado por fontes modeladas pelo processo envelope do fBm. Este algoritmo é inspirado no trabalho desenvolvido por Parekh e Gallager [9, 10] e aproveita o fato de que pode ser definida uma função, estritamente crescente e linear por partes, que funciona como limitante inferior para $S_i(t)$. Esta função, $\widehat{S}_i(t)$, é definida como a justaposição de segmentos

s_j^i definidos do seguinte modo : seja $t = e_1$ o instante de tempo no qual o *backlog* de uma das conexões é completamente consumido. Considerar-se-á, por simplicidade, que seja a fonte $i = 1$, ou seja, $Q_1(e_1) = 0$. Assim, no intervalo $[0; e_1]$ o serviço fornecido ao tráfego correspondente a cada conexão i é dado por :

$$c_1^i(t) = \frac{\phi_i}{\sum_{j=1}^N \phi_j} t$$

O valor de e_1 pode ser determinado utilizando a definição de $Q_1(t)$, ou seja, fazendo $Q_1(e_1) = \widehat{A}_1(e_1) - c_1^1(e_1) = 0$. Seja agora $t = e_2$ o instante no qual uma outra conexão (por simplicidade, a fonte $i = 2$) tem o seu *backlog* completamente consumido. No intervalo $[e_1; e_2]$ o serviço acumulado fornecido ao tráfego correspondente a cada conexão i é dado por :

$$c_2^i(t) = \begin{cases} \widehat{A}_1(t) & , i = 1 \\ \frac{\phi_i}{\sum_{j=2}^N \phi_j} [t - e_1 - \widehat{A}_1(t)] + c_1^i(e_1) & , i > 1 \end{cases}$$

sendo o valor de e_2 dado por $Q_2(e_2) = \widehat{A}_2(e_2) - c_2^2(e_2) = 0$. Supondo que as fontes permaneçam ordenadas em função do tempo necessário para que o servidor consuma todo o seu *backlog* (o que representa uma generalização do *feasible ordering* proposto por Parekh e Gallager em [9, 10]), pode-se concluir que, para um intervalo $[e_{j-1}; e_j]$ qualquer o serviço fornecido a cada fonte será dado por :

$$c_j^i(t) = \begin{cases} \widehat{A}_i(t) & , 1 \leq i < j \\ \frac{\phi_i}{\sum_{l=j}^N \phi_l} [t - \xi_j] + c_{j-1}^i(e_{j-1}) & , i \geq j \end{cases} \quad (13)$$

sendo $\xi_j = e_{j-1} + \sum_{l=1}^{j-1} \widehat{A}_l(t)$ e o valor de e_j dado por $Q_j(e_j) = \widehat{A}_j(e_j) - c_j^j(e_j) = 0$. Os segmentos que definem $\widehat{S}_i(t)$, denominados s_j^i são dados então por :

$$s_j^i = \left. \frac{dc_j^i(t)}{dt} \right|_{t=e_{j-1}} \quad (14)$$

A Figura 1(a) apresenta como os segmentos são justapostos para formar $\widehat{S}_i(t)$. Obviamente número de segmentos é limitado ao número de conexões que podem ter seu *backlog* efetivamente esgotado pelo servidor GPS conforme esgotasse o *backlog* das demais conexões e a capacidade ociosa é redistribuída.

Define-se, então Q_i^* como sendo o máximo *backlog* associado à conexão i e solução do seguinte problema de otimização :

$$Q_i^* = \max_t \widehat{A}_i(t) - \widehat{S}_i(t) \quad (15)$$

Para o caso do máximo atraso associado à conexão i , definido como D_i^* , o problema de otimização é dado por :

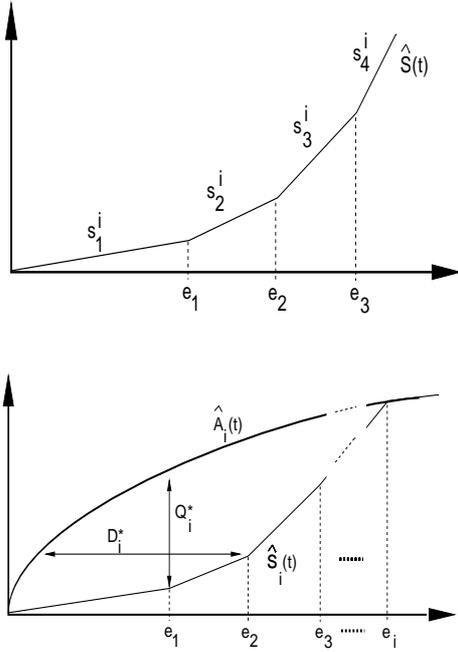


Figura 1: (a) Exemplo de construção da função limitante $\hat{S}_i(t)$. (b) Definição dos valores de Q_i^* e D_i^* .

$$D_i^* = \arg \max_{d(t)} \left\{ d(t) : \hat{A}_i(t) - \hat{S}_i(t + d(t)) = 0 \right\} \quad (16)$$

A Figura 1(b) demonstra como são definidos os valores de Q_i^* e D_i^* . É possível provar a existência e a unicidade destes problemas de otimização para uma conexão i tal que:

$$\rho_i < \frac{\phi_i}{\sum_{j=i}^N \phi_j} \left[1 - \sum_{j=1}^{i-1} \frac{d\hat{A}_j}{dt}(e_{i-1}) \right] \quad (17)$$

6. EXEMPLO NUMÉRICO

Nesta seção é apresentado um exemplo numérico para ilustrar os resultados obtidos na seção anterior. É considerado um caso simples, no qual um servidor GPS é alimentado por duas fontes de tráfego auto-similar modeladas pelo processo envelope do movimento Browniano fracionário e deseja-se obter os valores limitantes para o *backlog* das duas conexões em função do fator de Hurst de uma delas.

Os parâmetros das fontes de tráfego são apresentados na Tabela 1, que também apresenta a parametrização adotada para o servidor neste caso. Por simplicidade, assume-se que a taxa do servidor GPS é igual a 1. Os limitantes para o *backlog* fornecidos pelo algoritmo proposto e os valores obtidos por simulação, utilizando séries auto-similares sintéticas, são apresentados nas Figuras 2 e 3. Adicionalmente,

são mostrados também os valores máximos de *backlog* que seriam obtidos caso as conexões fossem atendidas com taxa constante g_i dada por (2), que corresponde à mínima taxa de serviço oferecida pelo GPS à uma conexão — constitui assim um limitante do tipo pior caso, que muitas vezes é o único disponível em função da dificuldade de análise matemática daquela disciplina de serviço. Este limitante é obtido diretamente da Eq. (11).

Fonte	ρ_i	σ_i	k_i	H_i	ϕ_i
1	0.68	0.31	4.00	-	0.80
2	0.17	0.14	4.00	0.80	0.20

Tabela 1: Parâmetros das fontes no exemplo numérico.

Analisando os resultados apresentados nas Figuras 2 e 3, é possível verificar que o algoritmo proposto fornece um limitante superior para a Fonte 1 exatamente igual ao pior caso, e outro bem melhor que este para a Fonte 2. Explica-se a coincidência entre os limitantes no caso da Fonte 1 da seguinte maneira: para $H_1 < 0,82$, a Fonte 1 tem seu *backlog* totalmente consumido antes da Fonte 2. Isto significa que até o instante e_1 o limitante inferior do serviço fornecido a esta conexão corresponde a uma taxa constante igual a g_i , justamente como ocorre no pior caso.

Após o consumo total do *backlog* da Fonte 1, uma parte da taxa a ela destinada torna-se ociosa e é direcionada ao atendimento da Fonte 2 — o que leva a uma significativa redução no limitante em relação ao valor de pior caso. À medida que o valor de H_1 se aproxima de 0,82 tal redução se torna menos significativa, justamente pelo fato dos tempos para o consumo dos *backlogs* das duas fontes se aproximarem. Quando $H_1 > 0,82$ ocorre uma inversão e a Fonte 2 passa a ter todo o seu *backlog* consumido primeiro, e por isso o limitante obtido pelo algoritmo proposto para aquela fonte e o limitante de pior caso coincidem.

7. CONCLUSÃO E TRABALHOS FUTUROS

Neste artigo foi apresentado um algoritmo para a obtenção de valores limitantes para o *backlog* e para o atraso em um servidor GPS alimentado por fontes de tráfego auto-similares reguladas pelo algoritmo do balde furado fractal proposto por Viana Neto [2] e modeladas pelo processo envelope do movimento Browniano fracionário. Um exemplo numérico foi apresentado para mostrar o bom desempenho daquele algoritmo por meio de comparação com resultados obtidos por simulação e com os limitantes do tipo pior caso obtidos por meio da Eq. (11).

Diante dos bons resultados apresentados pelo algoritmo para o caso de um servidor GPS isolado, considera-se promissora a sua extensão para o caso de uma rede de servidores; estudos neste sentido constituem as próximas etapas deste trabalho.

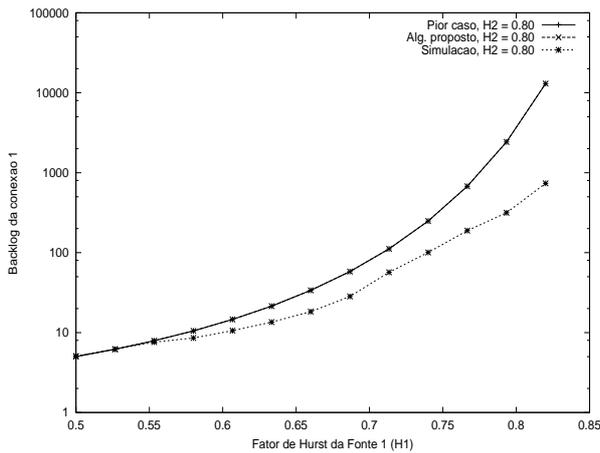


Figura 2: Curvas correspondentes ao *backlog* da conexão 1 em função do seu parâmetro de Hurst.

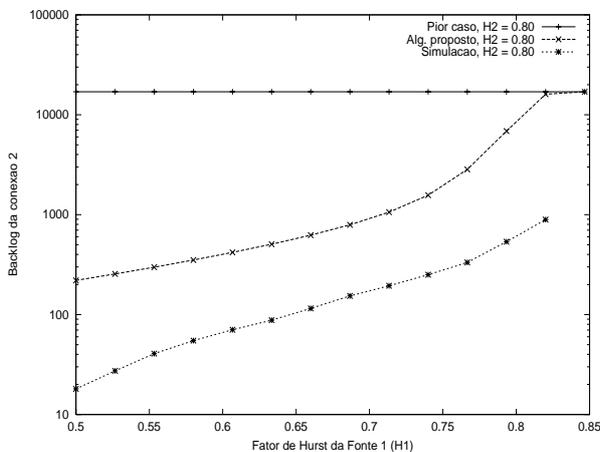


Figura 3: Curvas correspondentes ao *backlog* da conexão 2 em função do parâmetro de Hurst da conexão 1.

8. REFERÊNCIAS

- [1] W. Leland *et al.*, *On the self-similar nature of ethernet traffic (extended version)*, IEEE/ACM Trans. Networking, vol. 2, no. 1, pp. 1-15, 1994.
- [2] C. A. Viana Neto, *Multiplexação e policiamento de tráfego auto-semelhante*, Tese de mestrado, Universidade Estadual de Campinas, 1999.
- [3] B. B. Mandelbrot and J. W. Van Ness, *Fractional brownian motions, fractional noises and applications*, SIAM Review, vol. 10, pp. 422-437, 1968.
- [4] I. Norros, *On the use of fractional brownian motion in the theory of connectionless networks*, IEEE J. Select. Areas Commun., vol. 13, no. 6, pp. 953-962, 1995.
- [5] S. Keshav, *An engineering approach to computer networking : ATM networks, the Internet and the Telephone Network*, Addison-Wesley, 1997.
- [6] H. Zhang, *Service disciplines for guaranteed performance service in packet switching networks*, Proc. IEEE, vol. 83, no. 10, pp. 1374-1396, 1995.
- [7] R. L. Cruz, *A calculus for network delay, part I : Network elements in isolation*, IEEE Trans. Inform. Theory, vol. 37, no. 1, pp. 114-131, 1991.
- [8] R. L. Cruz, *A calculus for network delay, part II : Network analysis*, IEEE Trans. Inform. Theory, vol. 37, no. 1, pp. 132-141, 1991.
- [9] A.K. Parekh and R. G. Gallager, *A generalized processor sharing approach to flow control in integrated services networks : the single-node case*, IEEE/ACM Trans. Networking, vol. 1, no. 3, pp. 344-357, 1993.
- [10] A.K. Parekh and R. G. Gallager, *A generalized processor sharing approach to flow control in integrated services networks : the multiple-node case*, IEEE/ACM Trans. Networking, vol. 2, no. 2, pp. 137-150, 1994.
- [11] O. Yaron and M. Sidi, *Performance and stability of communication networks via robust exponential bounds*, IEEE/ACM Trans. Networking, vol. 1, pp. 372-385, 1993.
- [12] O. Yaron and M. Sidi, *Generalized processor sharing networks with exponentially bounded burstiness arrivals*, Proc. IEEE/ INFOCOM '94, pp. 628-634, junho de 1994.
- [13] Z. L. Zhang, D. Towsley and J. Kurose, *Statistical analysis of the generalized processor sharing scheduling discipline*, IEEE J. Select. Areas Commun., vol. 13, no. 6, pp. 1071-1080, 1995.
- [14] P. Dupuis and K. Ramanan, *A Skorokhod problem formulation and large deviation analysis of a processor sharing model*, Tech Report, Division of Applied Mathematics, Brown University, 1997.
- [15] S. Borst, O. Boxma and P. Jelenković, *Generalized processor sharing with long-tailed traffic sources*, in P. Key e D. Smith (editores), *Telletrafic engineering in a competitive world*, Proc. ITC-16, Edinburgh, UK, 345-354.
- [16] S. Borst, O. Boxma and P. Jelenković, *Asymptotic behavior of generalized processor sharing with long-tailed traffic sources*, Proc. IEEE/ INFOCOM 2000, pp. 912-921, março de 2000.