

Um Modelo Fracionário Modulado por Pareto para Tráfego com Alta Variabilidade e Dependência de Longo Alcance

Ricardo J. P. de Britto Salgueiro¹, Edilayne Meneses da Silva²,
José A. Suruagy Monteiro³

¹Universidade Católica de Pernambuco – UNICAP/ DEI
Rua do Príncipe, 526 Boa Vista 50050-900 Recife, PE.

²Universidade Federal de Pernambuco – UFPE/ CIn
Av. Prof. Luis Freire s/n – Caixa Postal 7851 – 50732-970 Recife, PE.

³Universidade Salvador – UNIFACS
Rua Ponciano Oliveira, 126 Rio Vermelho 41950-275 Salvador, BA.

ricardo@unicap.br, ems@cin.ufpe.br, suruagy@unifacs.br

Abstract. *We present a Pareto modulated fractional model for the characterization of traffic with high variability and long-range dependency. In our model, the difficulties of representing traffic with high variability are overcome through a classification algorithm based on clustering. The algorithm determines the states of the modulation process associated with the first and second order statistics with sufficiently small coefficients of variation. The component of long-range dependency is captured using the Pareto distribution to define state residence times. Analysis done through the generation of series for representative video sources had shown the validity of the proposed model. The statistical multiplexing effect in a computer network, in terms of bit loss rate, was evaluated through simulation. The obtained results showed that the proposed model can be used as an efficient tool for video traffic characterization and in studies involving resource allocation in computer networks, elaboration of rules for buffer dimensioning at the network switches, predictions of QoS degradation caused by congestion, and congestion control mechanisms.*

1. Introdução

Desde a identificação da existência de Dependência de Longo Alcance (LRD) em séries temporais relacionadas a processos da natureza por Hurst [Hurst, 1951] até a atualidade, uma variedade de algoritmos para geração de séries temporais com LRD foram propostos. Dentre tais métodos, destacamos o Ruído Gaussiano Fracionário — FGN (*Fractional Gaussian Noise*) gerado pelo algoritmo RMD (*Random Midpoint Displacement*) [Lau et al., 1995], o FGN gerado pelo método da Transformada Rápida de Fourier — FFT (*Fast Fourier Transform*) [Paxson, 1995] e o modelo Auto-regressivo Integrado de Média Móvel com Diferenciação Fracionária — FARIMA (*Fractional Autoregressive Integrated Moving Average*) [Garrett and Willinger, 1994]. Porém, em casos de alta variabilidade, tais métodos tendem a gerar séries com valores negativos [de Britto Salgueiro, 2004]. Uma solução frequentemente adotada é o truncamento em zero. Todavia, as propriedades estatísticas são alteradas. Neste trabalho, apresentamos

um modelo capaz de gerar um tráfego que possua alta variabilidade e LRD, sem permitir a ocorrência de valores negativos.

2. Modelo Fracionário Modulado por Pareto

O Modelo Fracionário Modulado por Pareto — PMFM (*Pareto Modulated Fractional Model*) [de Britto Salgueiro, 2004] é um modelo fracionário duplamente estocástico, no qual um processo de Pareto modula a transição de estados que determinam os parâmetros de um modelo fracionário para geração de séries temporais. A especificação de um $\text{PMFM}(H, n, S^*, T, P)$ é dada em termos do parâmetro de Hurst (H), do número de estados (n), do conjunto de estatísticas de primeira e segunda ordem de interesse (S^*), do tempo de permanência em cada estado (T) e da matriz probabilidade de transição (P). O PMFM deve ser parametrizado de modo a possuir um conjunto de estatísticas que indiquem uma variabilidade suficientemente pequena em cada estado. Desse modo, modelos fracionários tradicionais, como o FGN e o FARIMA, podem ser utilizados sem ocorrência de valores negativos. Assim, dada uma série temporal com estatísticas de primeira e segunda ordem, definidas pelo conjunto S , indicando uma alta variabilidade, obtemos, através do algoritmo de *clustering* k^+ -Means, que definimos em [de Britto Salgueiro, 2004], um novo conjunto S^* formado por n subconjuntos com variabilidade suficientemente pequena, de modo que $S \Leftrightarrow S^* = \{S^1, S^2, \dots, S^n\}$. A especificação do processo de modulação por Pareto corresponde à definição da matriz de probabilidades de transição e do tempo médio de permanência em cada estado. A probabilidade de transição de um estado a outro do processo de modulação é determinada a partir da série temporal, após o *clustering*. Um procedimento de discretização baseada nas médias dos *clusters* auxilia a identificação das transições de estados. Definindo um processo estocástico \mathbf{P} com estados $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_n$, $n < \infty$, a probabilidade de transição de um estado \mathcal{S}_i para um estado \mathcal{S}_j é determinada empiricamente tal que $p_{ij} = N_{ij}/N_i$, onde N_{ij} representa o número de vezes em que o sistema sai do estado \mathcal{S}_i e entra no estado \mathcal{S}_j , enquanto que N_i corresponde ao número de vezes que o sistema entra no estado \mathcal{S}_i . O tempo médio de permanência em um estado é determinado calculando o número médio de valores da série que permanecem em cada um dos níveis. Desse modo, considere N_{ki} o número de valores da série que permanecem no nível relativo ao estado \mathcal{S}_i durante a k -ésima visita a esse estado e N_i como o número de vezes que o sistema entra no estado \mathcal{S}_i . O número médio de valores da série que permanecem no estado \mathcal{S}_i é determinado por $Nq_i = 1/N_i \sum_{k=1}^{N_i} N_{ki}$. O tempo médio de permanência no estado \mathcal{S}_i é dado por $T_i = t_{sc} Nq_i$, onde t_{sc} é a escala de tempo relativa ao tempo de de cada valor da série. A característica de LRD da série é resgatada pela utilização da distribuição de Pareto transladada [Brichet et al., 1996] para determinação do tempo de permanência nos estados. A distribuição de Pareto transladada corresponde à distribuição de Pareto, com parâmetro $\alpha = 3 - 2H$, onde H é o parâmetro de Hurst, acrescida de um parâmetro θ que permite o ajuste de uma média desejada sem alterar o parâmetro de Hurst. A distribuição de Pareto transladada é aplicada controlando o parâmetro θ para a obtenção de médias correspondentes aos tempos de permanência extraídos da seqüência em questão.

3. Modelagem de Tráfego com o PMFM e Análise de uma Fila PMFM/D/1/k

Para exemplificar o uso do PMFM, utilizamos a seqüência dino, estudada em [Rose, 1995], que corresponde a cerca de meia hora de codificação em MPEG-1 do filme

Jurassic Park. As estatísticas da seqüência dino ratificam a existência de alta variabilidade e LRD, exibindo uma média de 13.078 bits por quadro de imagem, um coeficiente de variação de 1,13 e um parâmetro de Hurst de 0,88. Aplicando o algoritmo k^+ Means à seqüência dino obtivemos 7 *clusters* com coeficientes de variação menores que 0,2. A série gerada pelo PMFM a partir da seqüência dino, assumindo o FGN na geração das séries, apresentou uma média de 13.542 bits por quadros, um coeficiente de variação de 1,12 e um parâmetro de Hurst de 0,72. A alta variabilidade e o efeito LRD pôde, então, ser capturado pelo PMFM. Para a análise de uma fila PMFM/D/1/k exibimos as figuras 1(a) e 1(b). Na figura 1(a), ilustramos as curvas da taxa de perdas em função do comprimento da fila tanto para a seqüência dino como para a série gerada, considerando um canal com uma capacidade disponível $W_d = 500$ kbps. Em outra instância, apresentamos na figura 1(b), o gráfico da taxa de perdas em função da capacidade do canal para os dois casos em questão, admitindo uma fila de comprimento $k = 256$ Mb. Pela figura 1(a), podemos observar que a série gerada, quando comparada com a original, apresenta taxa de perdas ligeiramente maiores para filas pequenas e taxa de perdas menores para filas longas. Esse resultado coaduna com os apresentados em [Chen et al., 1995, Chien et al., 2003], onde observaram que o tamanho médio da fila aumenta com o aumento do parâmetro de Hurst. Em nosso caso, se por um lado a série gerada apresenta uma taxa média um pouco maior que a seqüência dino, por outro lado seu parâmetro de Hurst é menor. Assim, na região de fila pequena predomina o efeito da taxa média, enquanto que na região de fila longa predomina o efeito do parâmetro de Hurst. Observando a figura 1(b), verificamos que os resultados alcançados pela série gerada foram, de modo geral, próximos ao da seqüência codificada. O fato da série gerada possuir uma taxa média um pouco maior que a seqüência codificada foi um pouco mais significativo nas regiões de baixas perdas, mas os resultados podem ser considerados satisfatórios.

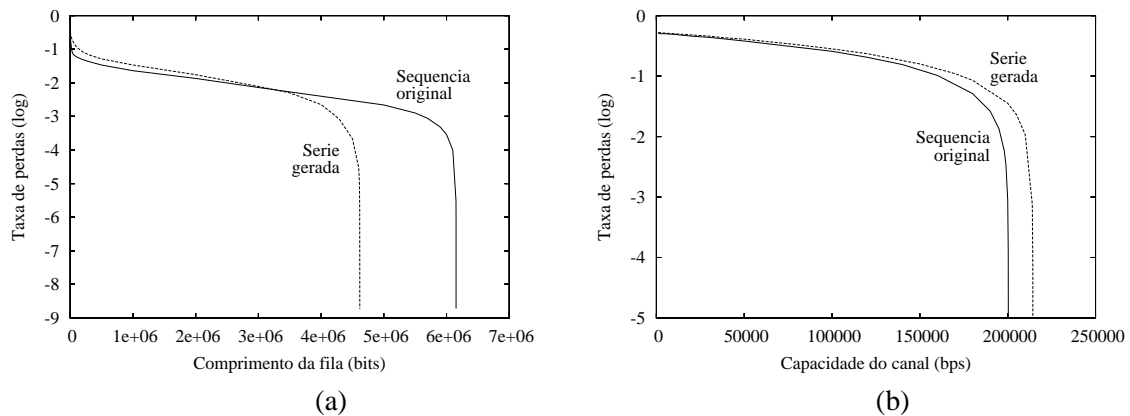


Figura 1: Taxa de perda de bits para seqüência dino e sua geração através do PMFM

4. Conclusão

Os resultados obtidos demonstram a capacidade que o PMFM possui em capturar as estatísticas de uma dada série temporal. A componente LRD pôde ser obtida com o uso da distribuição de Pareto na definição do tempo de permanência em cada estado. O algoritmo k^+ Means mostrou ser eficaz na classificação das seqüências, determinando *clusters*

com variações suficientemente pequenas para geração de séries características através do FGN, sem a ocorrência de valores negativos. Desse modo, a análise estatística dos resultados obtidos com as séries resultantes da modulação do FGN por Pareto ratificaram a adequação de nossa abordagem. Contudo, pudemos verificar também a ocorrência de divergências que podem ser previstas. Constatamos que enquanto as diferenças entre as estatísticas de primeira ordem afetam mais em situações de filas pequenas, as diferenças entre as estatísticas de segunda ordem, ou mais especificamente o parâmetro de Hurst, produzem um impacto maior em situações de filas grandes. Desse modo, é possível aferir se o resultado obtido com nossa abordagem está sendo superestimado ou subestimado. Resultados adicionais, apresentados em [de Britto Salgueiro, 2004], corroboram os aqui apresentados.

O Modelo Fracionário Modulado por Pareto pode ser utilizado, portanto, como uma ferramenta eficiente para a caracterização do tráfego com alta variabilidade e dependência de longo alcance e em estudos envolvendo alocação de recursos em redes, elaboração de regras para dimensionamento de *buffers* nos comutadores das redes, previsões de degradação de QoS causadas por congestionamento e de mecanismos de controle de congestionamento.

Referências

- Brichet, F., Roberts, J., Simonian, A., and Veitch, D. (1996). Heavy Traffic Analysis of a Storage Model with Long Range Dependent On/Off Sources. *Queueing Systems*, 23:197–215.
- Chen, Y., Deng, Z., and Williamson, C. (1995). A Model for Self-Similar Ethernet Lan Traffic: Design, Implementations, and Performance Implications. In *Proceedings of the 1995 Summer Computer Simulation Conference (SCSC'95)*, pages 831–837, Ottawa, Ontario.
- Chien, S. F., Tan, C. P., Low, A. L. Y., You, A. H., and Takahashi, K. (2003). Buffering controls for IP packets in optical packet switching. In *APCC 2003 - The 9th Asia-Pacific Conference on Communications*, volume 2, pages 473 – 476.
- de Britto Salgueiro, R. J. P. (2004). *Uma Metodologia para Caracterização de Tráfego de Vídeo Baseada nos Momentos de Primeira e Segunda Ordem*. PhD thesis, Universidade Federal de Pernambuco - Centro de Informática, Recife.
- Garrett, M. W. and Willinger, W. (1994). Analysis, Modeling and Generations of Self-Similar VBR Video Traffic. In *Proc. of SIGCOMM'94*, pages 269–280, London, UK.
- Hurst, H. E. (1951). Long-Term Storage Capacity of Reservoirs. *Trans. Amer. Soc. Civil Eng.*, 116:770–799.
- Lau, W.-C., Erramilli, A., Wang, J. L., and Willinger, W. (1995). Self-Similar Traffic Generation: The Random Midpoint Displacement Algorithm and Its Properties. In *Proc. IEEE ICC'95*.
- Paxson, V. (1995). Fast Approximation of Self-Similar Network Traffic. Technical report, Lawrence Berkeley Laboratory and EECS Division, University of California, Berkeley.
- Rose, O. (1995). Statistical Properties of MPEG Video Traffic and their impact on Traffic Modeling in ATM Systems. Technical Report 101, University of Wurzburg, Germany.