

NetBook: uma ferramenta para avaliação de desempenho de sistemas de comunicação

Marcilia Andrade Campos¹, Erick Lopes Silva¹, Diogo de Carvalho Pedrosa¹, Jorge Luiz de Castro e Silva¹, Janine Aguiar Loureiro¹, Carlos André G. Ferraz¹

¹Centro de Informática – Universidade Federal de Pernambuco (UFPE)
Caixa Postal 7851– Cidade Universitária – 50.732-970 – Recife – PE – Brasil

{mac, els2, dcp, jlcs, jal, cagf}@cin.ufpe.br

Abstract. This work talk about NetBook, a free tool, developed in Cin/UFPE with the goal of support performance analysis of communication systems, in particular, in this version, computer networks. NetBook has four modules: (i) Statistic, (ii) Generation, (iii) Transformation and (iv) Graphic. Except by the module of auto-similar traffic generation, implemented in C++, all the others NetBook modules were implemented in JavaTM.

Resumo. Este trabalho apresenta o NetBook, uma ferramenta gratuita desenvolvida no CIn/UFPE, com o objetivo de suportar análise de desempenho de sistemas de comunicação, em particular, nesta versão, redes de computadores. O NetBook contém quatro módulos: (i)-Estatística, (ii) Geração, (iii) Transformação e (iv) Gráfico. Exceto pelo módulo de geração de tráfego auto-similar, que foi implementado em C++, todos os demais módulos do NetBook foram implementados em JavaTM.

1. Introdução

Uma das mais surpreendentes conclusões sobre o tráfego *Ethernet* deveu-se ao artigo [Leland, Taqqu, Winllinger, Wilson 1993]. Nele, os autores provam que existe uma diferença estatística entre o tráfego real, medido na rede, e os modelos de tráfego tradicionais: o tráfego real exibe correlações sobre uma grande variedade de escalas de tempo, isto é, apresenta dependência de longo-alcance (*long-range*), enquanto que os modelos de tráfego tradicionais apresentam dependência de curto-alcance (*short-range*).

Todo este conhecimento novo sobre características do tráfego em redes de comunicação tanto decorreu da utilização de novos métodos quanto implica na necessidade de definir outros. A pergunta natural é então: que procedimentos usar para analisar o desempenho de redes de comunicação? Ou, o que tem sido usado na última década por pesquisadores na área para analisar o desempenho de uma rede? Os métodos mais usualmente encontrados são: estimador de Hill, distribuições de probabilidade [Willinger, Taqqu, Sherman, Wilson 1997]; séries temporais [You, Chandra 1999]; avaliação dos desvios nas caudas da distribuição de probabilidade [Crovella, Taqqu, 1999]; cd-plot, distribuições de Pareto e Lognormal, agregação de dados, regressão, parâmetro de Hurst [Crovella, Taqqu, Bestravos 1998]; modelagem de distribuições de variáveis aleatórias associadas com diferentes protocolos, estimação de parâmetros, transformação nos dados, medidas de tendência central, medidas de discrepância, intervalo de confiança, testes para comparar dados e modelos, correlação,

autocorrelação [Paxson 1994]; estatísticos descritivos e distribuições de probabilidade de variáveis na Web [Cunha, Bestravos, Crovella 1995]; processos auto-similares (*self-similar processes*) [Paxson, Floyd 1995]; agregação de processos estacionários em tempo discreto, estatística R/S, método da variância agregada [Uhlig, Bonaventure 2001]; filas [Erramilli, Narayan, Willinger 1996]; ondaletas (*wavelets*) [Feldman, Gilbert, Willinger, Kurtz 1998]; processos Markovianos, dependência de longo-alcance [Babic, Vandalore, Jain 1998]; testes estatísticos para autosimilaridade, distribuições de caudas-pesadas [Crovella, Bestravos 1997] [Willinger, Paxson, Taqqu 1999].

Este trabalho apresenta o **NetBook** (<http://www.cin.ufpe.br/~autosim/netbook>), Figura 1, uma ferramenta gratuita desenvolvida no CIn/UFPE, com o objetivo de suportar análise de desempenho de sistemas de comunicação, em particular, nesta versão, redes de computadores.



Figura 1. Tela Inicial do NetBook.

O restante deste artigo está organizado como segue. A seção 2 descreve a implementação e funcionalidades da ferramenta. Na seção 3 é apresentado um exemplo de utilização do **NetBook** para análise de redes de computadores. As conclusões do trabalho assim como os futuros desenvolvimentos são apresentados na seção 4. Ao final são mostradas as referências bibliográficas utilizadas.

2. Implementação e Funcionalidades

Uma ferramenta básica para análise de tráfego em redes de comunicação deve conter procedimentos para análise exploratória dos dados, os quais, usualmente, consistem de (i) determinar medidas de tendência central: média, mediana e moda, (ii) determinar medidas de dispersão: amplitude, coeficiente de variação, desvio padrão e variância, (iii) determinar separatrizes: quartis, decis e centis e (iv) apresentar recursos gráficos como correlograma, CD plot, gráfico de valores observados, entre outros. Os dados de entrada são as mensurações das variáveis de interesse. Supõe-se que estas mensurações foram obtidas através de *software* tais como o *MRTG* [MRTG], *Analyzer* [Analyzer], *MG-SOFT MIB Browser* [MG-SOFT], *Tivoli* [Tivoli], *Log Analyzer* [Log Analyzer], etc. ou são resultantes de algum processo de simulação.

Um aperfeiçoamento dessa ferramenta implementaria procedimentos sobre (i) inferência: estimativa, intervalo de confiança e teste de hipótese, com ênfase nos testes de bondade de ajuste [Jain 1991], [Meyer 1983], (ii) filas [Erramilli, Narayan e Willinger 1996], (iii) séries temporais: no domínio do tempo ou da freqüência [You, Chandra 1999] e (iv) processos estocásticos: cadeias de Markov e parâmetros relacionados com auto-similaridade.

Existem diversos *software* que contêm procedimentos para uma ou mais das técnicas citadas anteriormente, como, por exemplo, Maple [Maple], Minitab [Minitab], Matlab [Matlab], R [R], S-Plus [S-Plus] e TANGRAM II [TANGRAM II]. Entretanto, a existência de tráfego diferente do tráfego de voz, por exemplo, resultando na necessidade de uma diferente modelagem matemática apenas tem sido enfatizada na última década. Portanto, cálculo de estimativas para o estimador de Hill ou parâmetro de Hurst, por exemplo, não estão contemplados em alguns *software* usuais. Concluindo, mesmo que existam *software* que realizem várias das técnicas necessárias para análise de desempenho de redes de comunicação, nenhum deles, no momento, contempla todas as mencionadas anteriormente.

O **NetBook** foi projetado a fim de que o usuário fosse capaz de utilizar todas as suas funcionalidades, de maneira fácil e eficiente. Todas as funcionalidades de **NetBook** podem ser acessadas sem a necessidade de digitar qualquer linha de código, como em alguns *software* existentes. Cada procedimento contido nos módulos descritos a seguir contém parâmetros de entrada/saída e pode ser executado paralelamente/seqüencialmente dependendo do objetivo da análise. Apesar do **NetBook** apresentar uma interface gráfica própria, suas bibliotecas podem ser integradas a outros projetos, por meio da importação dos seus pacotes. Exceto pelo módulo de geração de tráfego auto-similar, que foi implementado em C++, todos os demais módulos do **NetBook** foram implementados em Java.

O **NetBook** contém quatro módulos:

- Estatística – Neste módulo estão todos os procedimentos estatísticos;
- Geração – É possível com NetBook gerar variáveis aleatórias e tráfego. Estas simulações são o conteúdo deste módulo;
- Transformação – Módulo responsável pela transformação de dados. Nesta versão, o **NetBook** apenas faz agregação de um conjunto de dados. Outras transformações serão implementadas em versões posteriores;
- Gráfico – Este módulo contém gráficos. Foi implementado usando a biblioteca gráfica JFreeChart [JFreeChart], que é gratuita e *open source*. O módulo que permite o gráfico de funções também faz uso das bibliotecas JLex [JLex] e CUP [CUP], a fim de fazer a interpretação e avaliação das funções desejadas pelo usuário.

A entrada dos dados no **NetBook** é feita por meio de arquivos de texto (com extensão txt, dat ou log) ou arquivos do Microsoft Excel (cuja extensão é xls). Os arquivos podem ser constituídos por várias colunas de dados, porém, para os cálculos de dados estatísticos, deve-se sempre selecionar uma coluna. Além disso, o arquivo só deve conter valores numéricos (inteiros ou reais). Portanto, nos procedimentos estatísticos, um arquivo de dados é um conjunto $\{x_1, x_2, \dots, x_n\}$, onde cada x_i , $i = 1, \dots, n$, é um número inteiro ou número real. No caso de x_i ser real, a precisão é a mesma apresentada na linguagem Java.

Para a leitura dos arquivos criados com o Microsoft Excel utilizou-se a biblioteca POI [POI], também com licença gratuita e *open source*.

Foi desenvolvida uma documentação on-line, por onde o usuário pode encontrar mais detalhes sobre a utilização do software.

2.1. Estatística

Este módulo contém procedimentos para (i) determinar medidas de tendência central: média, mediana e moda, (ii) determinar medidas de dispersão: amplitude, coeficiente de variação, desvio padrão e variância, (iii) determinar separatrizes: quartis, decis e centis, (iv) cálculo de autocorrelações, (v) testes de bondade de ajuste: Kolmogorov-Smirnov e Qui-quadrado, (vi) cálculo do parâmetro de Hurst e (vii) estimativa do parâmetro da Pareto.

2.2. Geradores

Com o **NetBook** é possível simular tráfego, auto-similar ou não. A geração de tráfego simulado tem, como saída, um arquivo com os dados gerados, que pode ser utilizado para análise na própria ferramenta.

Um método para gerar tráfego simulado com características auto-similares foi descrito por [Taqqu; Willinger; Sherman, 1997]. O programa “Gerador de Tráfego Auto-Similar”, invocado pela interface gráfica do **NetBook**, é uma implementação deste método e simula o tráfego real de transmissão de vídeo.

O tráfego gerado pelo programa apresenta características de auto-similaridade porque agrega duas fontes de Pareto com períodos ON e OFF distribuídos. Cada fonte gera pacotes de tamanho fixo e a cada unidade de tempo (segundo) é produzida uma nova observação. O programa simula o processo de contagem do número de pacotes que chegam em um equipamento, no intervalo de tempo $[0, t]$. O programa recebe a largura de banda da rede que se deseja simular e quantos segundos o tráfego simulado dura, Figura 2 (a); e gera (saída) um arquivo contendo o número de pacotes por segundo durante uma janela de tempo T. O número de pacotes segue uma distribuição de Pareto, enquanto que o tempo ocioso entre transmissões seguidas segue uma distribuição exponencial. O parâmetro ALFA define o grau de peso da cauda da distribuição de Pareto.

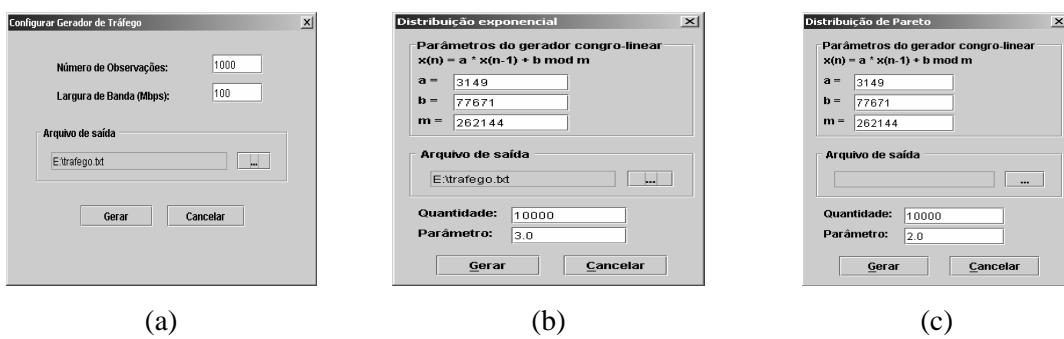


Figura 2. Telas Para Geração de Tráfego.

O **NetBook** também gera tráfego seguindo uma determinada lei de probabilidade. O processo de geração de variáveis aleatórias usado consiste de gerar uma seqüência de números aleatórios uniformemente distribuídos entre 0 e 1 através de um gerador congruo-linear misto e, em seguida, aplicar o método da transformação

inversa [Jain 1991]. A Figura 2 (b) mostra a tela de geração de tráfego com distribuição exponencial, enquanto a Figura 2 (c) mostra a geração para distribuição de Pareto.

2.3 Transformações

Nesta versão, o **NetBook** apenas faz agregação de um conjunto de dados, cuja implementação segue a fórmula

$$X_k^{(m)} = \frac{1}{m}(X_{km-m+1} + \dots + X_{km}), k = 1, 2, \dots$$

Por exemplo, a seqüência $\{X_k^{(3)}\}, k = 1, 2, \dots$, consiste de

$$\frac{X_1 + X_2 + X_3}{3}, \frac{X_4 + X_5 + X_6}{3}, \dots$$

2.4 Gráficos

O módulo gráfico do **NetBook** está dividido em (i) gráfico de dados e (ii) gráfico de funções. Os gráficos de dados, Figura 3 (a), disponíveis na ferramenta são aqueles presentes na literatura sobre tráfego em redes de computadores [Park, Kim e Crovella 1997][Winlliger et. al. 1995][Crovella, Taqqu e Bestravos 1998][Babic, Vandalore e Jain 1998][Leland et. al. 1994]: diagrama de dispersão ou gráfico de pontos, gráfico de linhas, gráfico de linhas e pontos, gráfico de freqüências, gráfico escada, correlograma, gráfico texturizado, gráfico da função de distribuição acumulada, CD-Plot. Para os gráficos de funções, Figura 3 (b), a ferramenta faz gráficos de funções lineares e quadráticas, ou de funções utilizando-se da sintaxe definida para a ferramenta. Esta sintaxe encontra-se disponível na página da ferramenta.

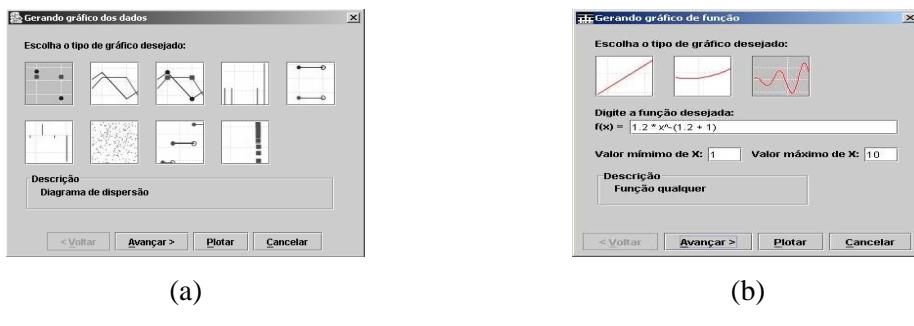


Figura 3. Tela de Gráfico de Dado (a) e de Função (b).

3. Exemplo de Utilização do NetBook

O exemplo a seguir mostra a verificação de auto-similaridade, utilizando o **NetBook**, de um arquivo de tráfego gerado pelo gerador de tráfego auto-similar da ferramenta. O tráfego, para este exemplo, foi simulado para 2.000 segundos em uma rede com 512 Kbps de largura de banda.

Para verificar a auto-similaridade desses dados, primeiramente realizamos agregações do conjunto inicial ($m = 1$) para o fator de agregação $m = 2, \dots, 10$. A Figura 4 mostra o correlograma para os dados originais (a) e para os resultados das agregações para $m = 3$ (b) e $m = 5$ (c). Um dos sinais de auto-similaridade observado foi a existência de um padrão nos correlogramas, além de as autocorrelações não tenderem a

zero. O passo seguinte foi estimar o parâmetro de Hurst para o conjunto de dados original e para os agregados. O resultado é apresentado na Tabela 1.

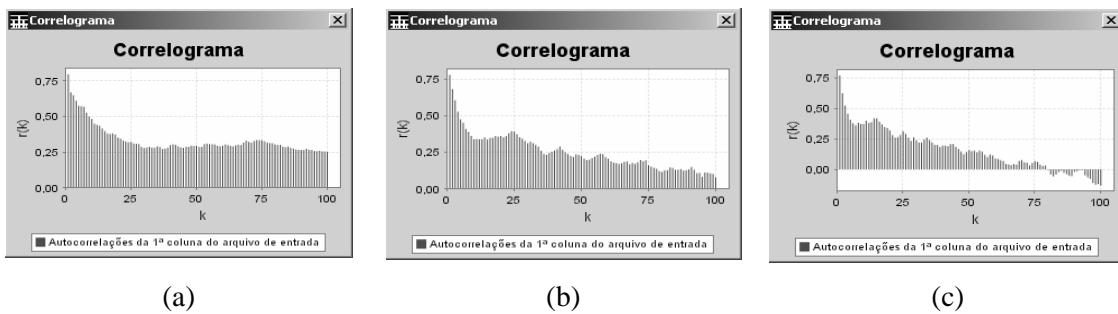


Figura 4. Correlogramas dos conjuntos de dados do exemplo.

Tabela 1. Parâmetros de Hurst para os conjuntos de dados do exemplo

Valor de m	Parâmetro de Hurst	Valor de m	Parâmetro de Hurst
1	0.8976567410748872	6	0.888698591842849
2	0.8945186937731334	7	0.8887394553179582
3	0.8930228380581087	8	0.8872240193911379
4	0.8913496055997552	9	0.8860021472269715
5	0.89007727876476	10	0.8875139921006799

Como os parâmetros de Hurst estimados estão muito próximos de 1, pode-se concluir que o conjunto de dados gerado é auto-similar.

4. Conclusões

Este trabalho apresentou uma ferramenta – **NetBook**, que pode ser usada para fazer avaliação de desempenho em sistemas de comunicação, em particular, nesta versão, redes de computadores. O **NetBook** é facilmente usada por alunos de graduação e pós-graduação, já tendo sido utilizado nos cursos “Estatística e Probabilidade para Computação” (graduação em Ciência da Computação e Engenharia da Computação / CIn / UFPE) e Avaliação de Desempenho de Sistemas (pós-graduação em Ciência da Computação / CIn / UFPE).

O **NetBook** é um dos frutos dos projetos CTRC e AUTO-SIM, 2001-2003, ambos financiados pelo CNPq, os quais envolvem, além de professores, alunos de iniciação científica, mestrado e doutorado. O **NetBook** está descrito com todos os detalhes, incluindo fórmulas e exemplos no Relatório Técnico **NetBook**, de número 01/2004 do Centro de Informática / UFPE.

Paralelamente ao contínuo e natural desenvolvimento da ferramenta, planos específicos para **NetBook** incluem (i) a implementação do módulo gerador de tráfego em Java (atualmente implementado em C++), (ii) a geração de gráfico de funções que apresentam descontinuidade, (iii) a inclusão de técnicas de avaliação de desempenho para middleware, (iv) métricas sobre filas [Erramilli, Narayan e Willinger 1996], (v) transformações nos dados (nesta versão a transformação disponível é a agregação), (vi) séries temporais, (vii) o aperfeiçoamento na obtenção da estimativa do parâmetro α da Pareto, permitindo a escolha dos pontos a serem usados no ajuste da reta e (viii) cálculo de *wavelets*.

Referências

- Analyzer, <http://analyzer.polito.it/>, último acesso em 31/01/2004.
- Babic, G., Vandalore, B. e Jain, R. (1998), "Analysis and Modeling of Traffic in Modern Data Communication Networks", <http://www.cis.ohio-state.edu/~jain/papers/trfmodel.htm>. Acessado em agosto de 2001.
- Crovella, M. E. e Bestravos, A. (1997), "Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes", <http://www.cs.bu.edu/fac/crovella/paper-archive/self-sim/journal-version.pdf>. Acessado em janeiro de 2004.
- Crovella, M. E. e Taqqu, M. S. (1999), "Estimating the Heavy Tail Index from Scaling Properties", <http://cd-www.bu.edu/faculty/crovella/paper-archive>. Acessado em setembro de 2001.
- Crovella, M. E., Taqqu, M. S. e Bestravos, A. (1998), "Heavy-Tailed Probability in the World Wide Web", A Practical Guide to Heavy Tails Statistical Techniques and Applications, Edited by R. J. Addler, R. J. Feldman and M. S. Taqqu, Birkhäuser.
- Cunha, C. R., Bestravos, A. e Crovella, M. E. (1995), "Characteristics of WWW Client-based Traces", Technical Report BU-CS-95-010, Boston University, Computer Science Department, BU-CS-95-010.
- CUP, <http://www.cs.princeton.edu/~appel/modern/java/CUP/>, último acesso em 31/01/2004.
- Erramilli, A., Narayan, O. e Willinger, W. (1996), "Experimental Queueing Analysis with Long-Range Dependent Packet Traffic", IEEE/ACM Trans. Networking, vol. 2, no. 4, p. 209-223.
- Feldmann, A., Gilbert, A. C., Willinger, W. e Kurtz, T. G. (1998), "The Changing Nature of Network Traffic: Scaling Phenomena", Computer Communication Review, vol. 28, no.2, pp. 5-29.
- Jain R., The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation and Modeling, John Wiley & Sons, 1991.
- JFreeChart, <http://www.jfree.org/jfreechart/>, último acesso em 31/01/2004.
- JLex, <http://www.cs.princeton.edu/~appel/modern/java/JLex/>, último acesso em 31/01/2004.
- Leland, W. E., Taqqu, M. S., Willinger, W. e Wilson, D. V. (1994), "On the Self-Similar Nature of Ethernet Traffic". IEEE/ACM Transactions on Networking, v.2, n.1, p.1-15.
- Log Analyzer, <http://www.netiq.com/products/log/default.asp>, último acesso em 31/01/2004.
- Maple, <http://www.maplesoft.com/>, último acesso em 31/01/2004.
- Matlab, <http://www.mathworks.com/>, último acesso em 31/01/2004.
- Meyer, P.L., Probabilidade. Aplicações à Estatística 2ª edição, Livros Técnicos Científicos, 1983.

- MG-SOFT, <http://www.mg-soft.si/>, último acesso em 31/01/2004.
- Minitab, <http://www.minitab.com>, último acesso em 31/01/2004.
- MRTG, <http://www.mrtg.org>, último acesso em 31/01/2004.
- Park, K., Kim, G. e Crovella, M. (1997), “On the Effect of Traffic Self-Similarity on the Network Performance”, Proceedings of the 1997 SPIE International Conference on Performance and Control of Network Systems, p. 296-310.
- Paxson, V. (1994), “Empirically Derived Analytic Models of Wide-Area TCP Connections”, IEEE/ACM Transactions on Networking, vol. 2, no. 4, pp 316-336.
- Paxson, V. e Floyd, S. (1995), “Wide-Area Traffic: The Failure of Poisson Modeling”, IEEE/ACM Trans. Networking, vol. 3, no. 3, p. 226-244.
- POI, <http://jakarta.apache.org/poi/>, último acesso em 31/01/2004.
- R, <http://www.r-project.org/>, último acesso em 31/01/2004.
- S-Plus, <http://www.insightful.com/>, último acesso em 31/01/2004.
- TANGRAM II, <http://www.land.ufrj.br/tools/tangram2/tangram2.html>, último acesso em 31/01/2004.
- Taqqu, M. S., Willinger e W., Sherman, R. (1997), “Proof of a fundamental result in self-similar traffic modeling”. ACM/SIGCOMM Computer Communication Review, v.27, p.5-23.
- Tivoli, <http://www.ibm.com/br/products/software/tivoli/>, último acesso em 31/01/2004.
- Uhlig, S. e Bonaventure, O. (2001), “Understanding the Long-Term Self-Similarity of Internet Traffic”, <http://citeseer.nj.nec.com/cachedpage/465608/1>. Acessado em janeiro de 2004.
- Willinger, W., Paxson, V. e Taqqu, M. S. (1998) “Self-Similarity and Heavy Tails: Strucutural Modeling of Network Traffic”, In: A Practical Guide to Heavy Tails Statistical Techniques and Applications, Edited by R. J. Addler, R. E. Feldman and M. S. Taqqu, Birkhäuser.
- Willinger, W., Taqqu, M. S., Sherman, R., e Wilson, D. V. (1995), “Self-Similarity Through High-Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level”, Proceedings of ACM SigComm'95, Cambridge, p. 100-113.
- You, C. e Chandra, K. (1999), “Time Series Models for Internet Data Traffic”, http://morse.uml.edu/~kchandra/time_series.pdf. Acessado em agosto de 2001.