

Caracterização de tráfego utilizando fluxos de comunicação *

Luís Felipe M. de Moraes e Guilherme S. Vilela

¹Laboratório de Redes de Alta Velocidade – RAVEL
Programa de Engenharia de Sistemas e Computação – COPPE/UFRJ
Caixa Postal: 68.511 – 21941-972 – Rio de Janeiro, RJ

moraes,vilela@ravel.ufrj.br

Abstract. *Traffic characterization is an important tool for the planning and management of computer networks. This article presents a characterization methodology for the network traffic using communication flows. A case study using the proposed approach was performed with Rede Rio[1], from which a few important results could be drawn through statistical analysis of the traffic.*

Resumo. *A caracterização de tráfego é hoje um importante instrumento para o planejamento e gerenciamento das redes de computadores. Este artigo apresenta uma metodologia para caracterização do tráfego de rede utilizando fluxos de comunicação. Um estudo de caso utilizando a metodologia proposta foi realizado com a Rede Rio[1], através do qual alguns resultados importantes puderam ser obtidos através da análise estatística do tráfego.*

1. Introdução

No decorrer dos últimos anos, a engenharia de tráfego passou a desempenhar um importante papel no planejamento e gerenciamento de redes de computadores. Uma importante técnica utilizada para a caracterização do tráfego é a medição dos fluxos de comunicação. Apesar de ser uma preocupação recente, o estudo da caracterização dos fluxos vem crescendo em importância, no entanto, quase não há estudos sobre como seus tamanhos, durações e taxas estão relacionados entre si. Com a finalidade de ajudar a responder essas questões é proposta neste trabalho uma metodologia para análise de tráfego. No intuito de validar a metodologia proposta, utilizou-se um estudo de caso com a rede de pesquisa do Estado do Rio de Janeiro, Rede Rio [1, 2].

2. Metodologia

Com o objetivo de identificar diferenças entre os fluxos, algumas definições foram propostas na literatura [3, 4], visando classificá-los quanto aos seus tamanhos, durações e taxas. Tendo em vista a diversidade das redes e do tráfego, a média e o desvio padrão dos fluxos devem ser considerados ao se propor tal classificação. A forma utilizada neste trabalho é descrita a seguir.

Tamanho: Fluxos de grande tamanho (“elephant” ou elefante) são aqueles com tamanho maior que x KB, sendo $x = \mu + 3\sigma$, onde μ é a média e σ é o desvio padrão da variável medida.

Duração: Fluxos de longa duração (“tortoise” ou tartaruga) são aqueles com duração maior que y minutos, onde $y = \mu + 3\sigma$.

Taxa: Fluxos de alta taxa (“cheetahs” ou chitá) são aqueles com taxa maior que z KB/s, onde $z = \mu + 3\sigma$.

*Este trabalho conta com suporte da FAPERJ (Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro).

2.1. Metodologia Utilizada

A metodologia utilizada para a análise dos dados coletados tem como objetivo verificar as distribuições cumulativas, classificar os fluxos e saber como os mesmos estão relacionados.

Para a classificação dos fluxos é necessário o cálculo da média e do desvio padrão da variável a ser analisada (tamanho, duração e taxa). Através destas medidas é possível fazer a classificação dos fluxos. Para tal, deve-se analisar e verificar se o valor da variável excede $\mu + 3\sigma$. Caso exceda, esse fluxo deve ser classificado em uma categoria (ex: se a variável medida for o tamanho, o fluxo será classificado como elefante).

Para relacionar os fluxos três análises foram feitas: a correlação das variáveis tamanho, taxa e duração, correlação destas variáveis dado que o fluxo pertence a uma determinada categoria e correlação das diferentes categorias de fluxos.

3. Resultados

Os resultados apresentados a seguir foram obtidos através dos dados coletados do roteador de borda da Rede Rio [1] durante as 24 horas do dia 17 de novembro de 2004 (quarta-feira).

As Figuras 1(a) a 1(c) mostram as distribuições dos fluxos de acordo com o tamanho, taxa e duração, respectivamente. A distribuição do tamanho dos fluxos, mostrada na Figura 1(a), indica que aproximadamente 95% dos fluxos possuem tamanho de até 10KBytes. A partir deste resultado é possível concluir que o tráfego é predominantemente composto por fluxos de tamanhos pequenos. A distribuição encontrada foi próxima daquela apresentada em [4, 5].

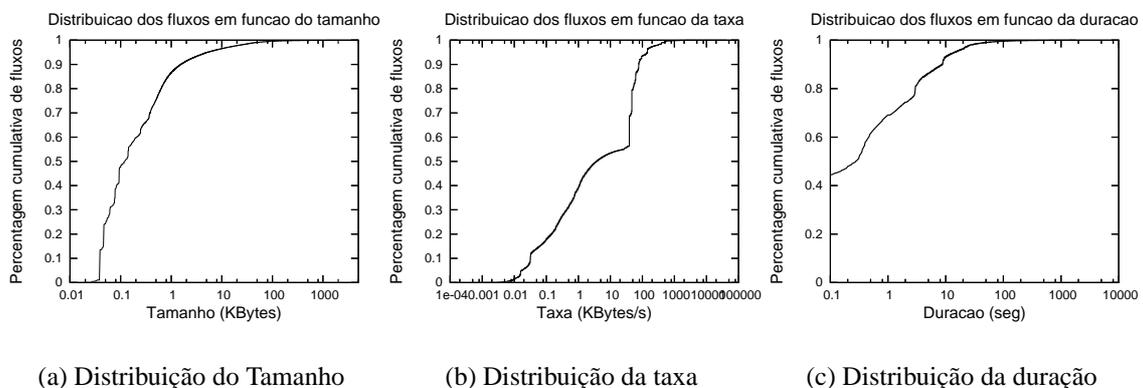


Figura 1: Distribuição cumulativa das diferentes variáveis dos fluxos

A Figura 1(b) indica que a grande maioria dos fluxos possui taxas de transmissão relativamente altas. Essa alta taxa pode ser relacionada com as distribuições das Figuras 1(a) e 1(c), que mostram que uma grande parte dos fluxos possui duração e tamanho pequenos. Em [4] a distribuição da taxa se mostrou inferior. Cerca de 85% das taxas eram iguais ou inferiores a 10Kbytes/s.

A Figura 1(c) mostra que, em sua maioria, os fluxos são de pequena duração, assim como observado em [3, 4, 5].

3.1. Média e desvio padrão medidos

Utilizando a metodologia descrita, foram encontrados os valores para a média e variância do tamanho, duração e taxa dos fluxos. Os resultados obtidos são mostrados na Tabela 1.

Na Tabela 2 pode-se verificar que uma pequena quantidade de fluxos é responsável por um grande volume de tráfego. Resultados semelhantes foram verificados em [3]. Esse

Tabela 1: Média e desvio padrão dos Fluxos

	Média	Desvio Padrão	Média + 3*desvio padrão
Tamanho (KBytes)	6.53	402.41	1213.76
Duração (seg)	4.30	38.22	118.95
Taxa (KBps)	39.53	119.42	397.80

Tabela 2: Percentagem do tráfego de cada categoria

	Elefante	Tartaruga	Chitá
Fluxos	0.071%	0.352%	1.057%
Bytes	62.261%	63.075%	2.143%

Tabela 3: Correlação dos fluxos

	Coefficiente de Correlação
Tamanho e Duração	0.29
Tamanho e Taxa	0.00
Taxa e Duração	-0.04

comportamento tem um grande impacto na rede. As infra-estruturas de rede e protocolos atuais são baseados em tráfego por rajada. O aumento da duração e do tamanho dos fluxos acarreta uma utilização maior do enlace de rede por um grande período de tempo, mudando o perfil do tráfego. Em relação aos fluxos chitá, é observado que os percentuais de fluxos e bytes são próximos, mostrando que os mesmos têm pouco impacto no tráfego de rede.

3.2. Correlação

Para fazer a caracterização do tráfego é necessário colher o maior número possível de informações. Um importante fator a ser estudado é a correlação entre as variáveis dos fluxos. Tal estudo torna possível o conhecimento de como elas estão relacionadas entre si, permitindo estudar mecanismos de tarifação de redes, distinguir um tráfego padrão de um malicioso e obter importantes informações para os projetistas de rede e de equipamentos [4].

A Tabela 3 mostra a correlação entre os fluxos indicados. Pode-se observar que há uma correlação entre duração e tamanho. Este comportamento pode ser explicado pelo fato de que quanto maior um fluxo, maior será a duração para uma taxa de transmissão constante. Ao se analisar o coeficiente de correlação entre taxa e tamanho dos fluxos, observa-se um valor nulo. Isto pode ser explicado pela diversidade do tráfego (tráfegos longos e de rajada). O resultado da correlação entre taxa e duração pode ser explicado pelo fato de longas durações ocorrerem devido a pequenas taxas de transmissão.

Nas Tabelas 4 e 5 foram analisadas as correlações dos fluxos em relação às suas categorias. A Tabela 4 mostra uma pequena correlação entre os fluxos de grande tamanho e longa duração. Os resultados obtidos entre os fluxos de alta taxa e longa duração têm comportamento esperado se considerarmos o tamanho dos fluxos constante. O pequeno valor encontrado no coeficiente de correlação indica uma grande variação nos tamanhos dos fluxos. O coeficiente de correlação entre os tráfegos chitá e elefante apresenta um percentual próximo de zero.

A Tabela 5 indica que nos fluxos chitá o aumento de tamanho implica aumento de duração. Este fato também explica o coeficiente de correlação nulo entre tamanho e taxa e entre duração e taxa nos fluxos chitá. Os fluxos de longa duração apresentam alta correlação entre tamanho e taxa. Pode-se também observar um coeficiente de correlação ligeiramente negativo entre duração e taxa nos fluxos de grande tamanho.

A Figura 2(a) mostra a distribuição dos fluxos de cada categoria em função do tamanho. É possível observar que a grande maioria dos fluxos de longa duração são fluxos de pequeno tamanho. Esta informação indica que fluxos tartaruga ocorrem devido

Tabela 4: Correlação entre as categorias

	Coef. de Correlação
Elefante e Tartaruga	0.19
Tartaruga e Chitá	-0.09
Chitá e Elefante	-0.01

Tabela 5: Correlação entre os fluxos

	Coeficiente de Correlação Dado		
	Elefante	Tartaruga	chitá
Tamanho e Taxa	0.14	0.74	0.00
Tamanho e Duração	0.21	0.24	0.95
Duração e Taxa	-0.26	0.074	0.00

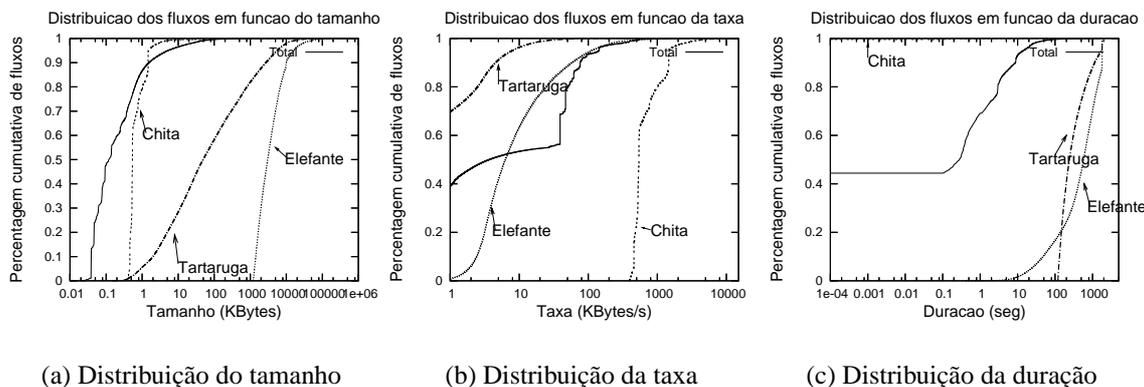


Figura 2: Distribuição cumulativa das diferentes categorias dos fluxos

a enlaces de baixa taxa ou devido ao comportamento dos protocolos utilizados. O gráfico mostra também que cerca de 90% dos fluxos de alta taxa têm até 1KB de tamanho.

A Figura 2(b) mostra que cerca de 80% dos fluxos de grande tamanho têm taxa de até 25KB/s, indicando que a maioria dos fluxos elefantes têm pequenas taxas de transmissão. Os fluxos tartaruga apresentaram as menores taxas, com cerca de 90% dos fluxos possuindo taxas até 4.5KB/s.

Finalmente, na Figura 2(c), observa-se que cerca de 50% dos fluxos elefante duram até 10 minutos e cerca de 80% duram 22 minutos. Mais de 95% dos fluxos de alta taxa duram menos de 1 segundo. Além disso, a figura indica que a maioria dos fluxos elefante possuem longa duração.

4. Conclusões

A engenharia de tráfego vem se tornando fundamental para as redes de computadores. Para se projetar e gerenciar uma rede é imprescindível que o tráfego na mesma seja conhecido. Este trabalho contribuiu para o melhor entendimento do tráfego da rede, mostrando a distribuição dos fluxos em função de seus tamanhos, durações e taxas. O trabalho mostrou ainda como as categorias dos fluxos estão relacionadas. Como trabalhos futuros pretende-se estudar outras formas de classificação dos fluxos, dividindo-os em um número maior de conjuntos.

Referências

- [1] “Rede rio.” <http://www.rederio.br/>, último acesso em 25/03/2005.
- [2] “Iptraf.” <http://iptraf.ravel.br/>, último acesso em 25/03/2005.
- [3] K. C. N. Brownlee, “Understanding internet traffic streams: Dragonies and tortoises,” *IEEE Communications*, vol. 40, pp. 110–117, Out 2002.
- [4] J. K. Lan, “On the correlation of internet flow characteristics,” 2003. <http://www.isi.edu/trpublic/pubs/au-kclan.html>, último acesso em 25/03/2005.
- [5] V. P. Yin Zhang, Lee Breslau and S. Shenker, “On the characteristics and origins of internet flow rates,” *Proceedings of ACM SIGCOMM*, Ago 2002.