Análise de Tráfego P2P no Backbone da RNP†

Stênio F. L. Fernandes[1], Guthemberg S. Silvestre[1], Kelvin L. Dias[1], João B. Rocha Jr.[1],Djamel Sadok[1], Carlos Kamienski[1,2]

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

[2]Centro Federal de Educação Tecnológica da Paraíba - CEFET-PB
Av. 1º de Maio, 720 – Jaguaribe - 58.015-430 - João Pessoa- PB

{sflf,gss2,jbrj,kld,jamel,cak}@cin.ufpe.br

Abstract. *Nowadays, there has been a dramatic growing of peer-to-peer (P2P) traffic on the Internet due to the ever increasing usage of the popular applications such as KaZaA, particularly for sharing large audio or video files and software. As a significant part of the current Internet traffic is generated by these applications, it is of paramount importance to understand and design strategies for provisioning of ISPs (Internet Service Providers) for taking into account this new scenario. Yet, to date, few researches analyze the P2P traffic, thus leaving a bulk of Internet traffic in dark. This paper describes the first analysis of the P2P traffic in the RNP (Rede Nacional de Pesquisa).*

1. Introdução

O uso de aplicações P2P (*peer-to-peer*) vem crescendo rapidamente devido à facilidade no compartilhamento arquivos de áudio, vídeo, além de software. Em sistemas P2P puros, os computadores comunicam-se diretamente, compartilhando informações e recursos sem o uso de servidores dedicados como ocorre no paradigma cliente-servidor. Uma característica comum a estes novos sistemas é que eles constroem, no nível de aplicação, uma rede virtual (*overlay*) com seus próprios mecanismos de roteamento, provendo assim um conjunto de serviços que possibilitam a localização de um usuário ou um recurso disponível, criação de grupos de discussão, *download* e *upload* de arquivos etc. Dentre os sistemas P2P, o *Fastrack* possui a maior rede pública existente e mantém, em média, 4 milhões de usuários conectados (e.g. KaZaA [3] e Grokster [4]). A disseminação de aplicações P2P está fazendo com que o tráfego da Internet experimente uma mudança significativa no seu comportamento tradicional que era, basicamente, dominado pelo tráfego *Web*. A análise e entendimento do comportamento do tráfego Internet diante deste novo cenário é de fundamental importância para o adequado aprovisionamento de recursos pelos ISPs (*Internet Service Providers*). Além disso, a análise do comportamento do tráfego P2P torna-se estratégica para a elaboração de algoritmos eficientes, como no contexto de novas técnicas de *caching* e no roteamento otimizado na rede *overlay*.

Neste artigo desenvolvemos o primeiro estudo sobre o perfil do tráfego P2P na RNP (Rede Nacional de Pesquisa) com o objetivo de avaliar o impacto gerado pelo uso destas aplicações em seu *backbone*.

2. Metodologia de Estudo do Tráfego

Os dados analisados foram obtidos através de arquivos coletados no formato binário gerado pelo NetFlow [5] no POP de São Paulo. A filtragem tanto do tráfego P2P quanto de outras aplicações utilizadas nos estudos deste artigo, baseiam-se na filtragem da portas bem conhecidas normalmente utilizadas (e.g., HTTP (80), , DNS (53), SSH (22), HTTPS (443), TELNET (23,107,992) e FTP (20,21,69,115, etc.).

As métricas adotadas para este estudo visam descrever o comportamento básico do tráfego em termos de volume, vazão e duração dos fluxos. Os resultados gerados são utilizados para análise estatística através do software *R v1.8.0* [6].

3. Resultados

A Figura 1 mostra o perfil de tráfego relacionado ao volume de tráfego passante no POP-SP durante o período de medição. A primeira impressão deste resultado é que a contribuição de tráfego P2P não é relevante, visto que seu volume gerado é substancialmente menor do que o tráfego gerado por aplicações Web.

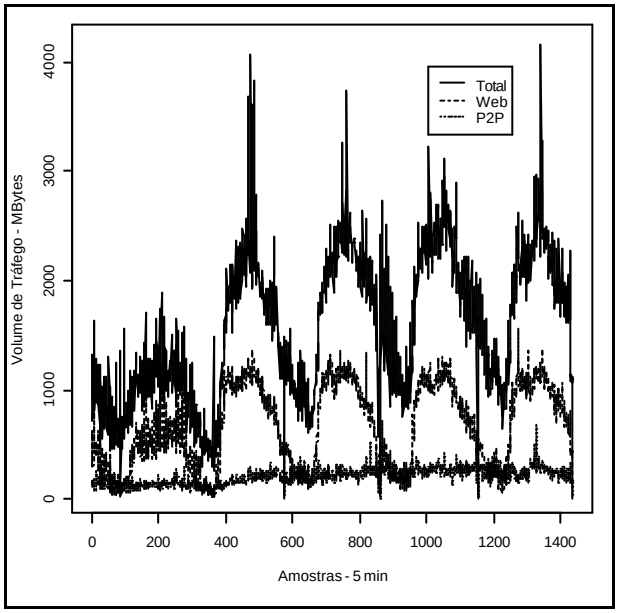

Figura 1 - Perfil de Tráfego em Volume Transferido (Mbytes) no POP-SP, 02 a 06/Nov

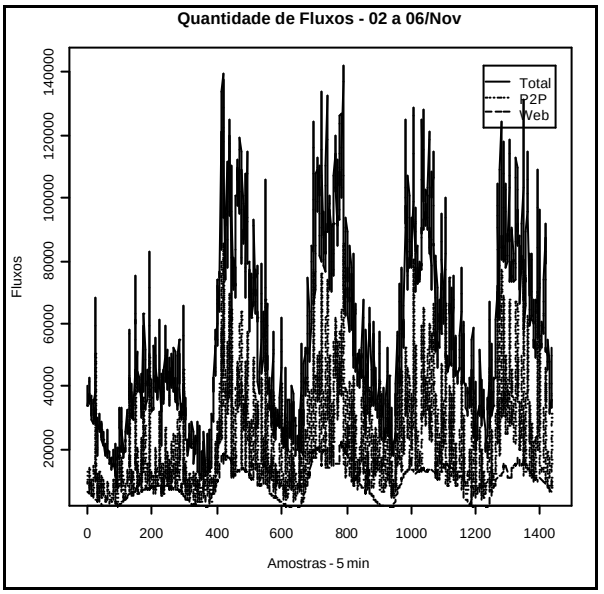

Figura 2 - Perfil de Tráfego em Quantidade de Fluxos (POP-SP), 02 a 06/Nov

Em termos de quantidade de fluxos, pode ser observado na Figura 2 seu perfil para os fluxos Web e P2P. Em contraste com o perfil do volume de tráfego, o número de fluxos P2P é substancialmente superior ao dos fluxos Web. Aparentemente, isto pode ser um indicativo de que uma quantidade significativa deste tráfego deve-se a requisições (buscas de palavras-chave) das aplicações P2P.

3.1. Distribuição do volume de tráfego por porta

Algumas aplicações P2P têm a habilidade de utilizar outras portas, diferentes de suas portas padrão, para transferência de dados. Por exemplo, através de técnicas de análise de conteúdo na camada de aplicação, foi recentemente observado que apenas uma pequena

porcentagem das sessões de downloads do Kazaa estavam utilizando a porta padrão (1214) [2]. Esta é uma tentativa de escapar de bloqueios impostos na borda da rede por regras de *firewall*. Desta forma, quaisquer outras portas podem ser utilizadas por estas aplicações P2P, independente de utilização dos protocolos normalmente associados a elas. Com isto, quaisquer anomalias na distribuição do volume de tráfego por porta de serviço mais conhecidas (e.g., 80/Web, 53/DNS), podem intuitivamente levar a conclusão que este comportamento dinâmico ocorra. Na Figura 3 é apresentada a análise da contribuição do volume de tráfego gerado por fluxos de aplicações na porta 80 (tipicamente tráfego web) de acordo com sua classificação nas faixas de 0 a 10KB, de 10KB a 100KB, de 100KB a 1MB, de 1MB a 10MB, de 10MB a 100MB e de 100MB a 1GB [1]. O perfil esperado do tráfego nesta porta é que, de acordo com [7], é de uma maior contribuição no volume de tráfego associada às duas primeiras faixas de classificação. Porém, o maior volume de tráfego gerado está associado às faixas de 100KB a 100MB. Como este comportamento é inesperado isto é um forte indicativo de que as aplicações P2P estejam também utilizando esta porta para transferência de dados. É interessante notar que estas faixas correspondem aos tamanhos típicos de arquivos de áudios codificados em MP3 (aproximadamente 3MB), de jogos, *clips* de música e software em geral (menor que 100MB).

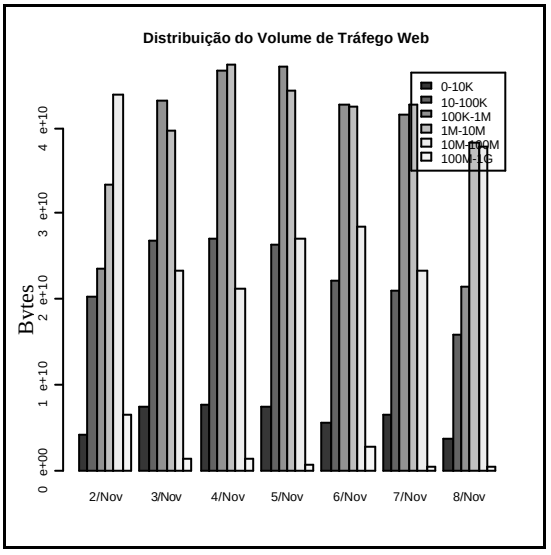

Figura 3 – Distribuição do Volume de Tráfego – Porta 80 (Web)

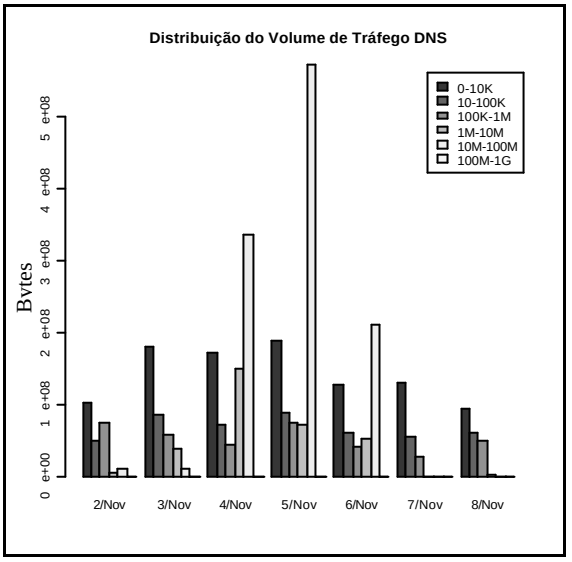

Figura 4 - Distribuição do Volume de Tráfego - Porta 53 (DNS)

Além da porta 80, esta análise foi estendida para a porta de serviço DNS (53). Os resultados da Figura 4 apresentam perfis diferentes dos usuais, pois em alguns dias aparece um grande volume de tráfego com fluxos na faixa de 10MB a 100MB, o que não é compatível com os tamanhos típicos de pacotes para estes serviços.

Uma importante métrica para caracterizar o comportamento do tráfego P2P é a inferência sobre os tempos de conexões destas aplicações. Na Figura 5 (gráfico Log-Log) esta distribuição aparenta ter as propriedades de uma distribuição Zipf (*Power Law*). A propriedade básica de uma distribuição Zipf é de um grande número de ocorrências de valores muito pequenos e a existência de cauda pesada para ocorrência de valores muito grandes. A Figura 6 mostra a caracterização do comportamento destes tempos de conexões.

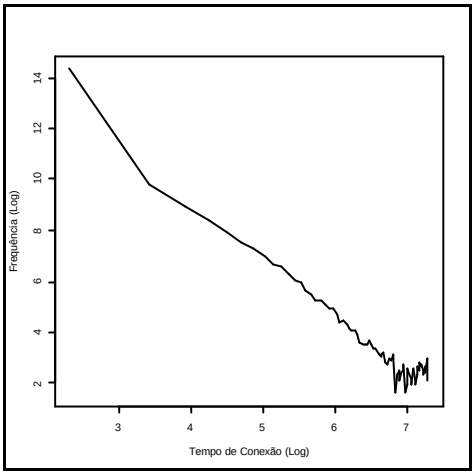

Figura 5 - Distribuição do Tempo de Conexão de Aplicações (Log-Log)

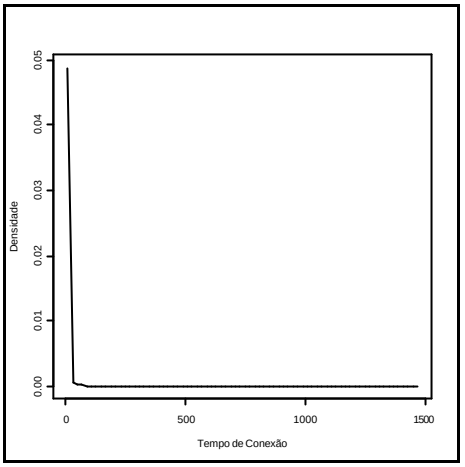

Figura 6 - Distribuição do Tempo de Conexão de Aplicações P2P

4. Conclusões e Trabalhos Futuros

Neste artigo foram realizados estudos sobre o perfil do tráfego P2P no *backbone* da RNP. Analisamos o tráfego referente a uma semana de dados coletados no POP-SP. A metodologia de trabalho consistiu no tratamento de dados coletados através da filtragem das portas utilizadas por aplicações P2P, assim como por outros tipos de tráfego. Os resultados da análise mostram indícios de um grande impacto do tráfego das aplicações P2P no tráfego global da rede. Adicionalemente, obtivemos fortes indicativos de que a maior quantidade de tráfego P2P no backbone da RNP não é transferido através das portas tradicionalmente utilizadas por essas aplicações. Ao contrário, seguindo uma tendência atual de burlar os controles de *firewalls*, as aplicações P2P (capitaneadas pelo KaZaA), utilizam qualquer porta que esteja aberta no firewall, geralmente as portas bem conhecidas, como 80/Web, 53/DNS, 22/SSH e 443/HTTPS. Verificamos isto através de anomalias nos resultados da distribuição do volume de tráfego dessas portas. Como trabalhos futuros pretende-se analisar um volume maior de informações, para obtenção de resultados mais conclusivos.

Referências

[1] Gummadi, K., *et al*. "Measurement, Modeling and Analysis of a P2P File-sharing Workload", *Proc. 11th ACM Symposium on Operating Systems*, 2003.

[2] Leibowitz, N., et al. "Deconstructing the Kazaa Network", *3rd IEEE Workshop on Internet Applications (WIAPP'03)*, June 23-24, 2003, San Jose, CA.

[3] "KaZaA", http://www.kazaa.com, acessado em Novembro 2003.

[4] "Grokster", http://www.grokster.com, acessado em Novembro 2003.

[5] NetFlow Services and Applications, Novembro 2003, www.cisco.com

[6] *R v1.8.0* http://www.r-project.org/, acessado em novembro 2003.

[7] Hyoung-Kee Choi, John O. Limb, "A Behavioral Model of Web Traffic", 7th Annual IEEE International Conference on Network Protocols (ICNP), October 1999, Canada.

† Este trabalho foi parcialmente financiado pela RNP, CAPES e CNPq.