

Caracterizando conexões P2P entre usuários do ISP: uma análise utilizando redes complexas

Agripino Gabriel M. Damasceno, Humberto T. Marques-Neto

¹Departamento de Ciência da Computação
Pontifícia Universidade Católica de Minas Gerais (PUC Minas)
30.535-901 - Belo Horizonte - Brasil

agripino.damasceno@sga.pucminas.br, humberto@pucminas.br

Resumo. *Controlar o tráfego P2P (peer-to-peer) é um desafio para os ISPs (Internet Service Providers), que precisam reduzir custos e manter o nível de serviço da rede. Um trabalho de caracterização da dinâmica das redes P2P pode auxiliar na tomada de decisões de otimização da rede pelos ISPs. Este trabalho apresenta uma caracterização do comportamento dos usuários P2P internos a rede do ISP utilizando métricas de redes complexas para entender o comportamento dos usuários. Os resultados mostram que alguns usuários internos possuem alta centralidade e maior quantidade de conexões de upload do que de download. Este resultado indica que esses usuários podem ser utilizados para promover conexões internas na rede do ISP reduzindo custos e, de certa forma, melhorando o desempenho da rede como um todo.*

Abstract. *The P2P (peer-to-peer) traffic control is a big challenge for ISPs (Internet Service Providers) since they must avoid the overload of network resources to reduce costs and maintain the level of service of their networks. Characterizing the dynamics of these networks can help the making decisions process to optimize the usage of ISPs resources. This paper presents a characterization of the behavior of internal users of P2P systems using complex network metrics for understanding the behavior of internal peers on P2P network. When they are inside the ISP network, these peers can be useful to avoid connections among peers from different ISPs, which can reduce the cross-ISP traffic. Our characterization shows that some internal peers on ISP network who have high centrality exhibit the output degree higher than input degree. It means that these peers could be used to promote the internal ISP connections, and, consequently, fostering the cost reduction of the ISP resources.*

1. Introdução

O BitTorrent é um sistemas P2P populares entre usuários de Internet no compartilhamento de arquivos [Xia and Muppala 2010]. Nesse sistema, um servidor (*tracker*) é responsável por manter a lista de usuários e os arquivos disponibilizados na rede. Para realizar o download de um arquivo o cliente precisa solicitar ao *tracker* uma lista de vizinhos que possuem o arquivo desejado. No BitTorrent original, o *tracker* pode escolher os vizinhos de forma completamente aleatória, o que pode tornar o sistema ineficiente na utilização de recursos da rede, principalmente, devido ao grande volume de dados que trafega no enlace entre ISPs [Miyoshi et al. 2012]. A grande utilização do enlace entre ISPs pode causar

o aumento dos custos de operação da rede pois estes links, normalmente, têm custos de utilização mais elevados [Hong et al. 2011, Le Blond et al. 2011].

Controlar o tráfego gerado pelos sistemas P2P é um desafio para os ISPs que precisam atingir margens de lucro satisfatórias. O controle de tráfego por parte do ISP, o controle a nível de aplicações e a cooperação entre usuários e ISP são algumas das propostas encontradas na literatura para reduzir o tráfego no enlace entre ISPs [Hong et al. 2011]. [Karagiannis et al. 2005] apresentaram um estudo que utiliza informações sobre a localidade dos *peers* na tentativa de evitar que a comunicação seja realizada sempre com *peers* externos a rede do ISP. Outras propostas podem ser encontradas na literatura como alterações na tarefa de seleção de vizinhos [Le Blond et al. 2011], a cooperação entre ISPs e sistemas P2P [Xie et al. 2008] e a interferência dos ISPs na rede P2P [Miyoshi et al. 2012].

Utilizando uma rica base de dados de um provedor de acesso brasileiro, este trabalho utiliza métricas de redes complexas para apresentar uma caracterização do comportamento de usuários de P2P, estudando suas atividades na rede interna do ISP. As transações P2P foram modeladas como um grafo onde foram aplicadas métricas de redes complexas. Diferente de outras caracterizações, esse trabalho aborda a comunicação entre os usuários internos do ISP com o objetivo de identificar propriedades que possam ser utilizadas na proposta de novas soluções de problemas conhecidos como o aumento do tráfego no enlace entre ISPs.

O restante deste artigo está organizado da seguinte maneira. A seção 2 discute trabalhos da literatura. A metodologia de caracterização, a coleta de dados e uma visão geral da base de dados utilizada são apresentados na seção 3. Os resultados da caracterização são apresentados na seção 4 e as conclusões na seção 5.

2. Trabalhos Relacionados

O sistemas P2P se tornaram populares devido a sua capacidade de compartilhamento de conteúdo e sua execução distribuída. Entretanto, os sistemas P2P utilizam seus próprios métodos de definições de rotas para os pacotes, não levando em consideração as características da rede. Este comportamento é responsável pelo grande volume de dados que trafegam no enlace entre ISPs aumentando os custos de operação para os ISPs. A localidade em sistemas P2P é uma estratégia proposta com o objetivo de tentar reduzir o volume de tráfego no enlace entre ISPs e consequentemente reduzir custos de operação da rede [Karagiannis et al. 2005].

[Le Blond et al. 2011] realizaram uma avaliação extensiva do impacto da localidade nos enlaces entre ISPs utilizam experimentos reais em um cluster de computação e avaliaram o comportamento da rede mostrando a eficiência da localidade na redução do tráfego no enlace entre ISPs. [Klachquin and Figueiredo 2011] propõe que o ISP tenha controle do servidor *tracker* ficando responsável por selecionar para a lista de vizinhos uma quantidade de *peers* X que pertençam a rede do ISP. [Hong et al. 2011] propõem uma arquitetura cooperativa entre os *peers* e ISPs onde *peers* controlados pelo ISP são inseridos na rede P2P. Estes HICs ficariam responsáveis por guiar o tráfego na rede P2P e trabalhariam de forma autônoma e transparente.

[Yu et al. 2010] modelaram a estrutura comportamental de várias aplicações como uma rede complexa. Este trabalho realiza uma análise comparativa de diferentes

tipos de aplicações baseada em métricas de redes complexas incluindo distribuição de grau dos nós, distância e a estrutura da comunidade. O trabalho apresentado nesse artigo tenta identificar características dos usuários P2P para melhorar a utilização de sistemas P2P na rede interna do ISP. [Goncalves et al. 2012] utilizam métricas de redes complexas para analisar como a estrutura da rede sobreposta do SoapCast¹. Este trabalho difere do proposto neste artigo, pois nele os autores utilizam métricas de redes complexas para caracterizar uma aplicação específica de *streaming* de vídeo.

3. Caracterização temporal do comportamento dos usuários P2P

Duas bases de dados de um provedor de acesso brasileiro foram utilizadas na caracterização dos usuários de sistemas P2P: (i) um log do DHCP (*Dynamic Host Configuration Protocol*) e (ii) um log de tráfego de dados. O primeiro log contém informações detalhadas sobre os serviços de DHCP e das sessões dos usuários. Através deste log é identificado o usuário responsável por utilizar um endereço IP em determinado período de tempo. A segunda base de dados contém informações como endereço IP, protocolo, duração, volume de dados (download e upload) e o serviço utilizado. As transações de um usuário u_i são identificadas pela sessão s_i , que fica aberta até que o usuário encerre a conexão ou fique inativo por um período superior a uma hora. A caracterização proposta utiliza apenas transações P2P, as quais representam aproximadamente 26% do total presente no log.

Uma visão geral dos dados utilizados nesta caracterização estão apresentados na Tabela 1. Os logs cobrem um período de 48 dias (27/02/2009 até 15/04/2009), durante este período um total de aproximadamente 3.200.000 de sessões de usuários foram identificadas sendo 11,71% de usuários que utilizam aplicações P2P. Entretanto, essas sessões são responsáveis por apenas 59,95% dos bytes recebidos e 86,55% dos bytes enviados. Ou seja, os usuários P2P geram a maior parte do volume de dados que trafegam na rede.

Tabela 1. Visão geral da carga de trabalho envolvendo todas as transações P2P - 27/02/2009 à 15/04/2009(48 dias)

	Total	P2P	Não-P2P
Total de sessões (%)	3.274.055	383.313 (11,71)	2.890.742 (88,29)
Total download(TB) (%)	230,83	138,38 (59,95)	92,45 (40,05)
Total upload(TB) (%)	82,77	71,64 (86,55)	11,13 (13,45)
Média (CV) serviços/sessão	2,08 (0,65)	3,59 (0,57)	1,89 (0,58)
Média (CV) download/sessão(MB)	73,92 (17,81)	378,56 (8,44)	33,53 (23,02)
Média (CV) upload/sessão(MB)	26,50 (28,91)	195,99 (8,25)	4,03 (138,85)

Para cada período com duração de uma hora (*snapshot*) as transações P2P existentes foram utilizadas na criação de um grafo direcionado $G = (V, A)$ onde transações são as arestas A e os usuários os vértices V . A cada transação entre dois usuários v_a e $v_b \in V$ existe uma aresta direcionada $(v_a, v_b) \in A$ indicando a existência dessa conexão. A caracterização do comportamento dos usuários P2P em cada *snapshot* foi realizada seguindo as métricas de centralidade como apresentado em [Goncalves et al. 2012].

4. Resultados da caracterização

Nessa seção os resultados mais relevantes da caracterização realizada são apresentados. Primeiramente é apresentada uma análise da conexão entre os usuários internos a rede do

¹<http://www.sopcast.org>

ISP, e então são apresentadas as propriedades dinâmicas dos usuários internos com alta centralidade na rede P2P interna.

Durante a análise dos logs foi identificado que a maioria dos usuários se conectam com *peers* externos a rede do ISP confirmando a necessidade da criação de estratégias para reduzir o tráfego no enlace entre ISPs. A Tabela 2 resume os dados obtidos durante a análise dos logs. Os CVs (Coeficientes de Variação) são altos para os graus máximos de entrada e saída pois existem poucos usuários com quantidade elevada de conexões. Ou seja, a maioria dos usuários se conectam com uma quantidade bem menor de *peers*.

Tabela 2. Média (CV) das métricas - 27/02/2009 a 15/04/2009 (48 dias)

Usuários internos e externos da rede P2P					
	Dia 1	Dia 2	Dia 3	Dia 4	Dia 5
Grau máximo de entrada	40,1(392,4)	37,6(442,0)	37,6(444,3)	37,6(451,1)	37,7(453,0)
Grau máximo de saída	25,8(308,2)	25,2(396,6)	25,6(408,2)	25,4(414,0)	25,4(405,5)
Caminho médio	2,4(16,6)	2,6(16,3)	2,3(10,8)	2,8(16,8)	2,8(20,3)
Diâmetro	9,7(25,8)	9,9(24,5)	8,6(24,4)	10,2(17,6)	10,7(27,2)

Na Tabela 2 estão apresentados os valores obtidos nas medidas de diâmetro e de caminho médio na rede formada a partir das transações P2P. O valor medido para o diâmetro e o caminho médio são de aproximadamente 10 e 2,7 respectivamente. Estes valores que indicam os usuários internos se comunicam com *peers* externos em comum mantendo o grafo da rede P2P mais conectado. Quanto maior o número de *peers* externos em comum maior a possibilidade de utilizar a localidade para reduzir o volume de dados que trafegam no enlace entre ISPs.

O grau médio de entrada tem valores maiores que o grau médio de saída, ou seja, na média os usuários de sistemas P2P fazem mais downloads do que uploads, confirmando o que foi apresentado na Tabela 1. Entretanto, os grandes valores de CVs mostram que existe uma grande variação nos valores. Por exemplo, um *super-peer* que, normalmente, envia dados para uma grande quantidade de usuários devido a sua maior capacidade de banda, entre outros motivos. Para diferenciar os usuários e selecionar os que possuíam altos valores de centralidade foi utilizado o valor do *betweenness*, métrica usada para tentar encontrar os pontos centrais em uma rede. Os usuários internos com maior *betweenness* têm uma quantidade maior de conexões, o que estimulou a tentativa de descobrir se eles podem ser utilizados para promover as conexões entre usuários internos.

Tabela 3. Média (CV) das métricas - 27/02/2009 a 15/04/2009 (48 dias)

Usuários internos com alto <i>betweenness</i>					
	Dia 1	Dia 2	Dia 3	Dia 4	Dia 5
Grau máximo de entrada	29,4(85,2)	70,0(123,0)	65,9(131,4)	64,0(122,5)	38,8(141,7)
Grau máximo de saída	87,2(100,0)	125,6(90,4)	118,9(85,2)	130,8(84,8)	106,2(91,5)

Para identificar os usuários com alto *betweenness* consideramos dois valores para selecionar o ponto de corte: a média e o desvio padrão. Após testar diversos valores por experimentação foi utilizado o valor de *betweenness* igual a média somada de três vezes o desvio padrão como valor de corte. Então, os usuários com alto valor de *betweenness* que têm uma quantidade maior de conexões foram identificados. Na Tabela 3 é possível observar que esses usuários internos apresentam graus de saída maiores que os graus de entrada, conseqüentemente, eles concentram uma grande quantidade de conexões com usuários externos. Esses usuários podem ser utilizados para reduzir o tráfego no enlace entre ISPs, pois, os demais usuários internos podem tirar vantagem da proximidade entre eles para o download de conteúdo.

5. Conclusões e trabalhos futuros

Neste trabalho foi apresentada uma caracterização de redes P2P utilizando métricas de redes complexas na identificação de usuários com altos valores de centralidade. Esses usuários apresentam maior quantidade de conexões e, seu comportamento pode influenciar no funcionamento da rede P2P. Os resultados mostram que as conexões P2P entre usuários internos do ISP são limitadas a uma pequena quantidade de conexões e podem ser incentivadas como uma estratégia para reduzir o tráfego no enlace entre ISP.

Usuários internos do ISP que possuem *betweenness* alto têm, normalmente, maiores taxas de upload do que de download possibilitando que outros usuários tenham a oportunidade de baixar o conteúdo compartilhado. Esses usuários podem ser utilizados na tentativa de que os usuários internos consumam maior quantidade de dados de usuários que estão na rede do ISP. A conexão entre usuários internos do ISP podem ser facilitadas por alguns métodos como esquemas de precificação ou a modificação em aplicações, evitando conexões externas que podem ser desnecessárias e onerosas.

Os resultados dessa caracterização podem ser utilizados em trabalhos futuros na proposta de precificação para usuários de P2P com o objetivo de balancear a rede do ISP, reduzir o tráfego no enlace entre ISPs e, conseqüentemente, diminuir o custos de operação da rede e do acesso a Internet. Trabalhos que utilizem essa técnica podem promover uma cobrança mais justa do acesso a Internet banda larga para todos os usuários evitar a sobrecarga de recursos na rede.

Referências

- Goncalves, K., Vieira, A., Almeida, J., da Silva, A., Marques-Neto, H., and Campos, S. (2012). Characterizing dynamic properties of the sopcast overlay network. In *Parallel, Distributed and Network-Based Processing (PDP), 2012 20th Euromicro International Conference on*, pages 319–326.
- Hong, F., Huang, D., Shan, T., and Guo, Z. (2011). Hic: A hidden cooperation strategy for cross-isp traffic control. In *Internet Technology and Applications (iTAP), 2011 International Conference on*, pages 1–4.
- Karagiannis, T., Rodriguez, P., and Papagiannaki, K. (2005). Should internet service providers fear peer-assisted content distribution? In *Proceedings of the 5th ACM SIGCOMM conference on Internet Measurement, IMC '05*, pages 6–6, Berkeley, CA, USA. USENIX Association.
- Klachquin, A. K. and Figueiredo, D. R. F. (2011). Um mecanismo orientado a isp para escolha tendenciosa de pares no bittorrent. In *XXIX Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos, WP2P 2011*, pages 71–84, Campo Grande, MS, Brasil.
- Le Blond, S., Legout, A., and Dabbous, W. (2011). Pushing bittorrent locality to the limit. *Comput. Netw.*, 55(3):541–557.
- Miyoshi, T., Shinozaki, Y., and Fourmaux, O. (2012). A p2p traffic localization method with additional delay insertion. In *Intelligent Networking and Collaborative Systems (INCoS), 2012 4th International Conference on*, pages 148–154.
- Xia, R. and Muppala, J. (2010). A survey of bittorrent performance. *Communications Surveys Tutorials, IEEE*, 12(2):140–158.
- Xie, H., Yang, Y. R., Krishnamurthy, A., Liu, Y. G., and Silberschatz, A. (2008). P4p: provider portal for applications. In *Proceedings of the ACM SIGCOMM 2008 conference on Data communication, SIGCOMM '08*, pages 351–362, New York, NY, USA. ACM.
- Yu, K., Wu, X., and Cao, Y. (2010). Comparative analysis of behavioral structure from internet traffic. In *Broadband Network and Multimedia Technology, 2010 3rd IEEE International Conference on*.