

Identificação de Pares Pouco Cooperativos em Sistemas Par-a-Par de Distribuição de Vídeos ao Vivo

Glauber D. Gonçalves
DCC - UFMG - Brasil
ggoncalves@dcc.ufmg.br

Alex Vieira Borges
DCC - UFJF - Brasil
alex.borges@ufjf.edu.br

Jussara M. de Almeida
DCC - UFMG - Brasil
jussara@dcc.ufmg.br

ABSTRACT

Este artigo descreve um projeto de dissertação de mestrado em andamento que visa desenvolver estratégias para identificar participantes pouco cooperativos em sistemas de distribuição de vídeos ao vivo baseados em arquiteturas par-a-par. Para tanto, investigou-se a correlação entre a razão das taxas de *upload* por *download* dos pares e métricas da estrutura topológica da rede sobreposta formada pelas parcerias entre eles. Os resultados preliminares mostram que métricas de centralidade, como o grau e o *betweenness*, provêm boas evidências sobre o nível de cooperação do par na rede, isto é, sobre a razão entre a taxa de *upload* e *download* do par. Propõe-se então desenvolver um modelo de regressão capaz de prever o nível de cooperação de um par em um dado instante no tempo, tendo como entrada valores da sua centralidade no passado recente. Tal modelo pode ser utilizado para detecção de pares pouco cooperativos (*free-riders*) assim como prover subsídios ao projeto futuro de mecanismos de incentivo a cooperação.

Categories and Subject Descriptors

C.2.0 [Computer-Communication Networks]:
General-Data communications

General Terms

Performance, Design, Measurement

Keywords

P2P, free riding, Topology

1. CONTEXTO TEÓRICO

Este artigo descreve um projeto de dissertação de mestrado em andamento, sendo desenvolvido na Universidade Federal de Minas Gerais. O projeto foi iniciado em março de 2010 e a sua conclusão está prevista para março de 2012.

WebMedia'11: Proceedings of the 17th Brazilian Symposium on Multimedia and the Web. XI Workshop on Ongoing Thesis and Dissertations.

October 3 -6, 2011, Florianópolis, SC, Brazil.

ISSN 2175-9650.

SBC - Brazilian Computer Society

O uso da Internet banda larga tem contribuído para a popularização de aplicações de distribuição de mídia contínua. Isso pode ser observado com o crescente número de canais profissionais ou amadores que abraçam a oportunidade de alcançarem uma audiência além das suas regiões de origem. Assim, a maior disponibilidade de conteúdo multimídia na Internet desperta o interesse de um número maior de internautas tornando esse serviço cada vez mais popular.

A arquitetura tradicional de serviços de distribuição de vídeo ao vivo - arquitetura cliente-servidor - tem se mostrado não escalável o suficiente para atender a demanda desse serviço pela Internet. Isso se deve, principalmente, ao consumo excessivo de banda no servidor de vídeo. Esse consumo torna-se mais crítico na distribuição ao vivo devido às condições estritas de entrega de conteúdo em tempo real. Logo, se um número grande de usuários têm interesse em obter serviço simultaneamente, a rede pode se tornar um ponto de contenção.

Nesse contexto, a maior vantagem da arquitetura P2P sobre o modelo cliente-servidor diz respeito à melhor escalabilidade quanto à quantidade de clientes atendidos simultaneamente. No modelo P2P, na medida em que os participantes conectados à rede obtêm a visualização da mídia, eles são habilitados a compartilhar os segmentos (*chunks*) da mídia mais atuais com os demais participantes do sistema. Logo, quem participa da rede contribui de modo significativo para aliviar a carga do servidor, assim como da rede em geral.

A formação de uma rede P2P normalmente ocorre de modo dinâmico, isto é, os participantes têm a liberdade de conectar ou sair da rede a qualquer momento. Quando um novo participante ingressa no sistema, ele interage com um subconjunto de participantes atualmente ativos, estabelecendo assim parcerias. Desse modo, nós parceiros formam uma rede lógica comumente chamada de rede sobreposta. Dependendo da estrutura da rede sobreposta, os sistemas par-a-par podem ser categorizados em dois grupos: árvore ou malha. A última estrutura é a mais utilizada, devido à sua maior robustez quanto à dinâmica dos parceiros [11]. De fato, alguns sistemas P2P de distribuição de vídeos ao vivo usam organização dos participantes baseados em malha e têm-se mostrado escaláveis para atender milhares de participantes com uma qualidade de serviço aceitável [7].

2. IDENTIFICAÇÃO DO PROBLEMA

Os projetistas de sistemas de distribuição de vídeo têm como foco principal atingir uma boa qualidade de serviço. A principal premissa para que um sistema P2P proveja uma

boa qualidade de serviço a todos os participantes é que os mesmos atuem de modo altruísta e compartilhem um volume de dados proporcional ao recebido de seus parceiros. Todavia, isso nem sempre acontece em um ambiente aberto e de difícil controle como a Internet.

Segundo [10], em alguns tipos de redes P2P, como redes de compartilhamento de arquivos, uma quantidade pequena de pares altruístas podem fornecer os serviços requisitados e colaborar para a sobrevivência do sistema. No entanto, as redes P2P de distribuição de vídeos ao vivo precisam eliminar ou reduzir o impacto causado por pares pouco cooperativos para cumprir com as restrições de tempo real e a qualidade de serviço exigidos na distribuição de vídeo. Um dos impactos causado por esses pares diz respeito a sobrecarga na taxa de *upload* dos pares altruístas que podem se ver desestimulados a participarem do sistema.

As principais técnicas de identificação de pares pouco cooperativos em redes P2P de vídeos ao vivo [2, 6] exploram apenas uma propriedade dos pares, a saber, a razão entre as taxas de *upload* e *download* de dados. Isso se deve ao fato de que um par pouco cooperativo tende a receber (*download*) um valor muito maior de dados do que ele compartilha (*upload*) com os demais. No entanto essas taxas podem ser falsificadas por pares mentirosos [9] para enganar os relatórios de monitoramento em sistemas que detectam comportamentos suspeitos. Faz-se então necessário investigar novas propriedades dos pares que podem ser exploradas para a detecção eficaz de usuários pouco cooperativos.

3. OBJETIVO

Esse trabalho tem por objetivo encontrar propriedades que permitam identificar pares pouco cooperativos em sistemas P2P de distribuição de vídeos ao vivo. O ponto de partida é a observação da razão entre a taxa de *upload* por *download* dos pares. A seguir, essa razão é relacionada com outras informações extraídas desses pares, preferencialmente propriedades centralidade. Nesse propósito, considerando a constante evolução da rede, busca-se responder a seguinte pergunta: propriedades de centralidade do par na rede P2P sobreposta poderiam prever o seu nível de cooperação, estimado pela razão entre o volume de *upload* por *download*?

4. CONTRIBUIÇÕES ESPERADAS

As principais contribuições desse trabalho podem ser delineadas nos seguintes pontos:

- um estudo sobre as correlações entre o nível de cooperação de um par, definido como a razão entre a taxa de *upload* por *download*, e propriedades do par na rede sobreposta, particularmente métricas da centralidade.
- um modelo de regressão para prever o nível de cooperação de um par no futuro próximo dado suas métricas de centralidade no passado recente

O estudo e o modelo propostos servem para que os administradores de sistemas p2p de distribuição de vídeo ao vivo detectem pares pouco cooperativos e possam tomar ações adequadas com suas políticas internas. O modelo de predição também pode fomentar o desenvolvimento futuro de mecanismos de incentivos e de reputação.

5. METODOLOGIA

A metodologia geral consiste de 3 grandes passos: (1) coleta de dados referentes às propriedades topológicas de redes P2P de distribuição de vídeos ao vivo reais assim como às trocas de dados entre parceiros; (2) caracterização das propriedades topológicas dos participantes e da correlação entre essas propriedades e o nível de cooperação dos mesmos e (3) desenvolvimento e validação de um modelo de predição do nível de cooperação de um par.

No primeiro passo, foram realizados experimentos utilizando o sistema P2P de distribuição de vídeos ao vivo SopCast¹, um dos mais populares atualmente, e o PlanetLab [3]. Assim, foi possível formar redes P2P em escala global e posteriormente analisar toda a estrutura sobreposta formada, semelhante a metodologia descrita em Tang et. al.[13]. Foram criados dois tipos de pares no experimento: o par provedor do vídeo, que é um computador situado no laboratório do campus, onde foi configurado um canal dedicado para o experimento exibindo um vídeo a 280 kbps; e os computadores do PlanetLab que foram sincronizados para conectar ao canal no mesmo horário e permanecerem conectados durante 60 minutos. Diferente do experimento realizado por [13] que contava com apenas 51 pares do PlanetLab, esse trabalho utilizou 450 pares. Cada par foi configurado com um cliente SOPCast e Wireshark² para coletar e armazenar dados referentes às parcerias estabelecidas por estes clientes coletores e seus parceiros.

Foram realizados um total de 6 experimentos entre os dias 22 e 24 de fevereiro de 2011. A seguir, os dados de cada par foram processados para serem analisados apenas os 30 minutos intermediários da transmissão, evitando a instabilidade inicial da rede e a perda de pares que desconectaram inesperadamente próximo ao fim do experimento. Também foram descartadas todas as mensagens menores que 1300 bytes uma vez que essas mensagens podem se referir a pacotes de controle. O nosso foco primordial foi na troca de mensagens de *dados* entre participantes.

Conforme discutido na Seção 1, a rede sobreposta é dinâmica, dado que parcerias podem ser iniciadas e finalizadas a todo instante, durante uma transmissão. Assim, no segundo passo da metodologia foram analisados o volume de dados trocados entre parceiros bem como as propriedades topológicas dos pares em fotografias consecutivas do sistema, cada uma representando 60 segundos, ao longo de cada transmissão. Cada fotografia do sistema foi modelada como um grafo direcionado ponderado, onde os vértices representam os pares e uma aresta direcionada entre pares i e j com peso $p_{i,j}$ indica que i fez *upload* de $p_{i,j}$ bytes para seu parceiro j durante os 60 segundos correspondentes.

As métricas de centralidade *proximidade*, *grau* e *betweenness* foram calculadas para cada nó do grafo, considerando a sua versão não direcionada e as equações 1 de acordo [5]. O *grau* é a quantidade de parcerias, ou seja, arestas incidentes ao par v . A *proximidade* $P(v)$ de um par v é dada pelo inverso do tamanho médio dos caminhos mínimos desse par para todos os outros pares no grafo conforme a função $P(v)$. O *betweenness* é a quantidade de caminhos mínimos que cruzam um par v , dado pela função $B(v)$, onde $\sigma_{ij}(v)$ é a quantidade de caminhos mínimos entre os pares i e j passando por v , e σ_{ij} é a quantidade total de

¹www.sopcast.org

²www.wireshark.org

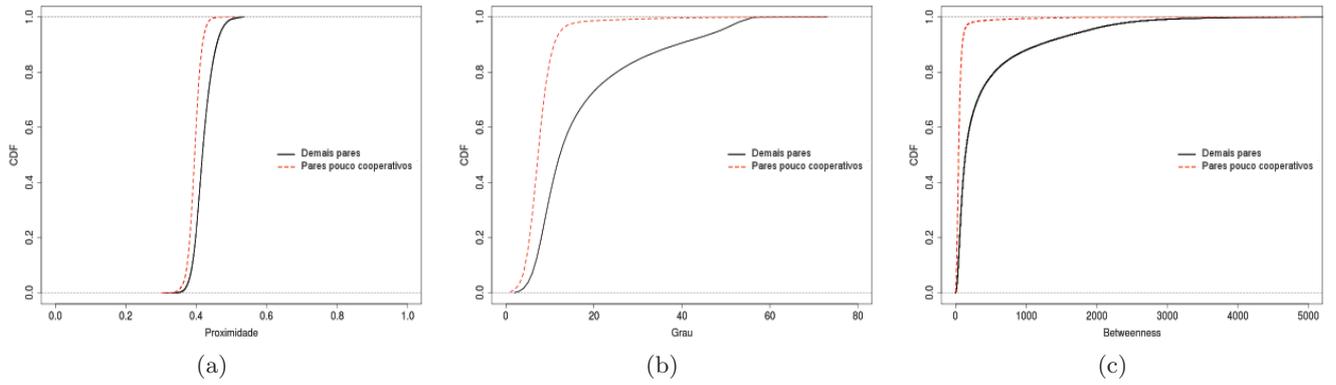


Figure 1: CDF de métricas de centralidade para pares pouco cooperativos e demais pares da rede.

caminhos mínimos entre os pares i e j .

$$P(v) = \frac{|v| - 1}{\sum_{i \neq j} d(v, i)} \quad ; \quad B(v) = \sum_{i \neq v, i \neq j, j \neq v} \frac{\sigma_{ij}(v)}{\sigma_{ij}} \quad (1)$$

6. ESTÁGIO ATUAL DO TRABALHO

Conforme a metodologia apresentada, foram realizados o passo 1, o passo 2 e o estágio atual do trabalho refere-se ao passo 3 que compreende o desenvolvimento e validação de um modelo para predição do nível de cooperação de um par.

A análise preliminar dos experimentos mostrou que cerca de 20% dos pares apresentam razão entre taxas de *upload* e *download* zero, considerando mensagens de dados de vídeo maiores que 1300 bytes. Por outro lado, cerca de 3% dos pares têm um volume de *upload* de 10 a 50 vezes maior que o seu volume de *download*, mostrando que o protocolo SOPCast não aplica mecanismos para balancear a carga entre os pares.

Visto a quantidade significativa de pares com a razão entre *upload* e *download* igual a zero, denominado nesse trabalho como pares pouco cooperativos, a figura 1 mostra a função de probabilidade cumulativa (CDF) para as métricas *proximidade*, *grau* e *betweenness* em curvas separadas para pares pouco cooperativos e demais pares. As figuras 1(b) e 1(c) mostram que pares pouco cooperativos exibem distribuições de grau e *betweenness* bem distintas, quando comparadas às mesmas distribuições dos demais pares: pares pouco cooperativos tendem a ter grau e *betweenness* menores. Em contrapartida, a métrica *proximidade* se mostrou pouco eficaz para discriminar pares pouco cooperativos dos demais, conforme mostrado na figura 1(a).

A diferença entre as curvas de pares pouco cooperativos e demais pares para grau e *betweenness*, mostram indícios que essas métricas estão correlacionadas com a cooperação do par estimada pela razão entre volume de *upload* por *download*. Em consequência disso, foram propostos modelos de regressão [8] para prever o nível dessa cooperação em função do grau e *betweenness* na rede sobreposta. Foram construídos modelos com diferentes parametrizações envolvendo as duas métricas em equações com crescimento linear, quadrático e exponencial, e o modelo quadrático mostrado na equação

$$c(b, d) = \alpha \times b^2 + \beta \times d^2 + \gamma \times (d * b)^2 \quad (2)$$

onde c é a estimativa de cooperação do par, d é o grau e b é o *betweenness*, foi o que apresentou melhores resultados, conforme será discutido abaixo.

As fotografias da rede obtidas nos experimentos foram utilizadas para avaliar o modelo de regressão da seguinte forma. Os coeficientes α, β, γ foram estimados usando as métricas grau e *betweenness* e as taxas de cooperação (*upload/download*) para cada nó, observadas na fotografia i . A seguir, o modelo da equação 2 foi aplicado pra prever o nível de cooperação dos pares na fotografia $i + 1$. Em situações práticas, a eficiência do modelo será observada em sua capacidade em prever a cooperação do par dado o dinamismo da rede P2P. Desse modo, a qualidade das predições foram avaliadas usando o coeficiente de determinação R^2 do modelo [8] e os valores absolutos do erro residual, definido pela diferença entre valores previstos e valores reais na fotografia $i + 1$, como mostra a figura 2.

Observa-se na figura 2(a) que o modelo descreve bem a variação dos dados pois apresenta R^2 entre 0.75 a 0.95 em mais de 80% das fotografias da rede. Todavia, verifica-se ainda pares cujo comportamento foge à curva do modelo, visto na figura 1 onde há uma minoria de pares pouco cooperativos com grau e *betweenness* alto, também na figura 2(c) onde o erro é maior que 1 para cerca de 10% das predições.

O próximo passo desse trabalho consiste em melhorar a predição buscando modelos mais precisos e também realizar uma avaliação mais extensa da solução proposta. Por conseguinte, pretende-se investigar com que frequência o modelo precisa ser ajustado, o que por sua vez depende de quão frequentemente as características dos pares observadas mudam.

7. TRABALHOS RELACIONADOS

Os pesquisadores Eytan Adar e Bernardo A. Huberman [1] foram os primeiros que estudaram o problema de pouca cooperação em sistemas P2P e advertiram quanto ao risco que ele representa ao desempenho desses sistemas. Eles rastream o sistema de compartilhamento de arquivos Gnutella e observaram que 70% dos usuários não compartilhavam arquivos e que apenas 25% dos usuários atendiam 99% de todas as consultas no sistema. A partir desse trabalho o problema de *free riding* em P2P ganhou notoriedade e desde então tem surgido muitas contribuições para tentar resolvê-lo ou ao menos amenizá-lo.

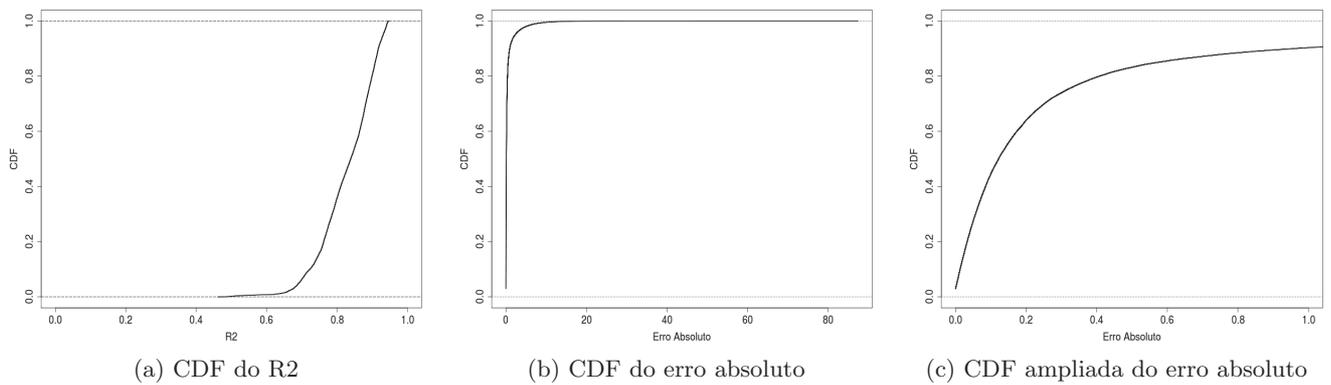


Figure 2: Avaliação do modelo de regressão em todas as fotografias da rede obtidas nos experimentos

Recentemente, os mecanismos de incentivo que usam contribuição bilateral entre os pares como *tit-for-tat* (TFT) [4] são amplamente aplicados em sistemas P2P de compartilhamento de arquivos. Contudo, Silverston et al. [12] adverte que TFT pode não ser eficaz em ambientes P2P de distribuição de vídeo ao vivo devido as restrições de tempo. Isso porque os segmentos de vídeo tornam-se obsoletos ao longo da transmissão, impossibilitando a alguns pares compartilharem todos os segmentos recebidos. Nesse contexto eles propõem um mecanismo de incentivo onde os pares incapacitados de servir dados aos seus parceiros provedores podem retribuí-los indicando parceiros que sejam potenciais fontes de dados. Embora essa estratégia seja promissora, os autores ainda não conseguiram avaliar eficientemente o quanto ela melhora a cooperação entre os pares.

Azzedin [2] propõe um mecanismo para filtrar e prevenir pares pouco cooperativos baseado em monitoramento e reputação. Nesse mecanismo, cada par mantém uma *lista negra* para armazenar os parceiros suspeitos e as suas informações sobre contribuição são enviadas para determinados nós da rede responsáveis pela manutenção do sistema. Guerraoui et. al. [6] propõe LiFTing, um protocolo para detectar pares pouco cooperativos que consiste em verificações diretas e indiretas de cooperação realizadas pelos pares. A primeira verifica se as requisições de dados foram atendidas pelos parceiros diretos, a segunda verifica se os parceiros diretos retransmitiram os dados recebidos aos seus respectivos parceiros (parceria indireta).

Os dois mecanismos apresentados acima detectam uma proporção razoável de pares pouco cooperativos, de acordo a avaliação proposta nos respectivos trabalhos. Contudo, eles utilizam apenas a taxa de *upload/download*, uma propriedade do nó já utilizada por outros métodos que pode ser manipulada pelos usuários maliciosos [9]. Além disso, eles aumentam a sobrecarga de mensagens de sinalização já existente nos atuais sistemas.

8. CONSIDERAÇÕES FINAIS

Este trabalho propõe uma nova estratégia para identificar participantes pouco cooperativos em sistemas P2P de distribuição de vídeos ao vivo. Foram investigadas a correlação entre a razão das taxas de *upload* por *download* dos pares e as métricas de centralidade na rede sobreposta formada pelas parcerias entre eles. Os resultados mostraram

que as métricas de centralidade grau e *betweenness* provêm boas evidências sobre o nível de cooperação do par na rede. Foi proposto e avaliado um modelo de regressão para prever o nível de cooperação de um par em um dado instante no tempo, tendo como entrada valores da sua centralidade no passado recente. O próximo passo desse trabalho consiste em melhorar a previsão e exatidão desse modelo.

9. REFERENCES

- [1] E. Adar and B. Huberman. Free riding on gnutella. In *First Monday: Peer-Reviewed J. Internet*, 2000.
- [2] F. Azzedin. Trust-based taxonomy for free riders in distributed multimedia systems. In *HPCS*, 2010.
- [3] B. Chun, D. Culler, T. Roscoe, A. Bavier, L. Peterson, M. Wawrzoniak, and M. Bowman. Planetlab: an overlay testbed for broad-coverage services. *SIGCOMM*, 33, 2003.
- [4] B. Cohen. Incentives build robustness in bittorrent. In *Workshop on Economics of Peer-to-Peer Syst.*, 2003.
- [5] L. Freeman. Centrality in social networks: I. conceptual clarification. *Soc. Net.*, 1, 1979.
- [6] R. Guerraoui, K. Huguenin, A.-M. Kermarrec, and M. Monod. LiFTinG: Lightweight Freerider-Tracking Protocol in Gossip. In *11th MIDDLEWARE*, 2010.
- [7] X. Hei, C. Liang, J. Liang, Y. Liu, and K. Ross. A measurement study of a large-scale p2p iptv system. *Multimedia, IEEE Transactions on*, 9(8), dec. 2007.
- [8] R. Jain. *The art of computer systems performance analysis*. Wiley, 1991.
- [9] X. Jin, S. Chan, W. Yiu, Y. Xiong, and Q. Zhang. Detecting malicious hosts in the presence of lying hosts in peer-to-peer streaming. In *Multim. and Expo*, 2006.
- [10] M. Karakaya, I. Korpeoglu, and O. Ulusoy. Free riding in peer-to-peer networks. *Internet Comp.*, 13(2), 2009.
- [11] N. Magharei, R. Rejaie, and Y. Guo. Mesh or multiple-tree: A comparative study of live p2p streaming approaches. In *INFOCOM 2007*, 2007.
- [12] T. Silverston, O. Fourmaux, and J. Crowcroft. Towards an incentive mechanism for peer-to-peer multimedia live streaming systems. In *P2P '08*, 2008.
- [13] S. Tang, Y. Lu, J. M. Hern'andez, F. A. Kuipers, and P. V. Mieghem. Topology dynamics in a p2ptv network. In *Networking 2009*. Springer, 2009.