

O papel das redes sociais na disseminação de arquivos em redes peer-to-peer

Thiago Amaral Guarnieri¹, Alex Borges Vieira¹, Ana Paula Couto Silva¹

¹Departamento de Ciência da Computação - Instituto de Ciências Exatas
Universidade Federal de Juiz de Fora - Juiz de Fora - MG - Brazil

thiago.guarnieri@ice.ufjf.br, {ana.coutosilva,alex.borges}@ufjf.edu.br

Resumo. *Sistemas de compartilhamento peer-to-peer (P2P), mais especificamente o BitTorrent, tem alcançado um grande sucesso na internet, pois elevam à um novo nível a escalabilidade de sistemas de difusão de conteúdo. A vazão de upload dos clientes, antes negligenciada, passa a ser usada, aliviando a carga dos servidores, além de eliminar a vulnerabilidade do paradigma cliente-servidor, cuja arquitetura possui somente um ponto de falha.*

Sistemas P2P têm sua eficiência atrelada à cooperação entre seus peers. É desejável que um sistema iniba o aparecimento de freeriders, que são peers que não contribuem com a escalabilidade do sistema. O BitTorrent possui mecanismos como o tit-for-tat que eficientemente reduz o surgimento de freeriders, mas não é uma solução definitiva. O presente trabalho procura mostrar novas abordagens de associação entre redes sociais e redes P2P, que visam aumentar a confiabilidade e qualidade de difusão de conteúdo.

- **Área de pesquisa:** Redes de Computadores
- **Ingresso:** Março de 2012
- **Previsão de conclusão:** Março de 2014

Palavras-chave: peer-to-peer, redes sociais, comunidades temporais, redes

1. Introdução

Muitos sistemas *peer-to-peer* são caracterizados por serem abertos. Qualquer *peer* pode ingressar na rede e potencialmente comunicar-se com outro no sistema, de forma que a confiabilidade reside na cooperação entre eles. De maneira geral, um *peer* não tem conhecimento prévio sobre os outros participantes, não podendo assumir se todos irão colaborar para a difusão de informação.

Muitos mecanismos têm sido propostos para evitar o surgimento de *freeriders*, que são os *peers* que não contribuem para a escalabilidade do sistema, entretanto esta não é uma tarefa trivial [Piatek et al. 2007]. Mais recentemente, o protocolo *BitTorrent* [Cohen 2003] definiu a estratégia *tit-for-tat* para combater *freeriders*, que se mostrou bastante eficiente, mas não resolveu completamente o problema: [Locher et al. 2006] demonstra que é possível fazer download de conteúdo sem precisar contribuir com nenhum dado útil, bastando que o cliente burle o mecanismo de *optimistic unchoking* do protocolo.

O relacionamento social entre os *peers* como um fator de confiabilidade tem sido foco de alguns estudos atualmente. [Wang et al. 2011] demonstra que certos *peers* possuem um comportamento mais estável que outros, ingressando em *swarms* em intervalos de tempo diários similares e mais longos, podendo contribuir melhor com a disseminação de informação. Além disso, [Wang et al. 2012] demonstrou que essa estabilidade pode ser melhor explorada em *torrents* compartilhados em redes sociais, particularmente o *twitter*. Mais recentemente as redes sociais também têm sido associadas a comunidades temporais [Pietiläinen and Diot 2012], em que *clusters* reaparecem ao longo do tempo, o que pode levar à uma possível implicação de que não somente nós individuais, mas grupos de nós podem ser explorados para acelerar compartilhamento, embora ainda não se saiba a taxa de incidência desses grupos em redes *peer-to-peer*.

2. Caracterização do Problema

O foco deste projeto é caracterizar o papel das redes sociais, particularmente o *facebook*, na disseminação de conteúdo em redes *peer-to-peer*. Mais recentemente esta ferramenta implementou a adição de grupos, muitos dos quais dedicados à distribuição de *torrents*, desta forma, é relevante a investigação da ocorrência de fenômenos já observados em outras redes sociais por [Wang et al. 2012] que melhoram a disseminação de informação, como nós estáveis e maior correlação temporal entre *peers*.

As áreas de particular interesse são o impacto da disseminação social de *swarms* na proliferação de *freeriders*, redução de tráfego entre sistemas autônomos e em métricas de desempenho como velocidade de *upload e download*, latência, tempo de vida do *swarm*, coeficiente de clusterização, número de saltos, entre outros. A incidência e o benefício de comunidades temporais induzidas por tal disseminação também será explorada.

3. Fundamentação Teórica

3.1. O Protocolo BitTorrent

O protocolo *BitTorrent* é focado na transmissão de dados em massa. Todos os usuários num *swarm* em particular estão interessados em obter um mesmo arquivo ou conjunto de

arquivos. Com o objetivo de ingressar em um *swarm*, o *peer* baixa um arquivo de meta-dados chamado *torrent*, de um provedor de conteúdo, usualmente via requisição HTTP simples. O arquivo de meta-dados especifica o nome e o tamanho do arquivo a ser baixado, bem como as assinaturas de integridade SHA-1 dos blocos de dados (tipicamente de 64 a 512KB). O arquivo também contém o endereço de um servidor *tracker*, que coordena a interação entre os *peers* do *swarm*. Cada *peer* contacta o *tracker* periodicamente, geralmente com a frequência de 15 minutos. O *tracker* mantém uma lista dos *peers* ativos e entrega um subconjunto aleatório deles quando requisitado.

Os usuários que possuem o arquivo completo, conhecidos como *seeds*, redistribuem pequenos blocos para os outros participantes do *swarm*. Os *peers* trocam blocos de informação e dados de controle com o conjunto de *peers* diretamente conectados, conhecido como vizinhança local. Este conjunto de *peers* obtido do *tracker* não possui nenhum tipo de hierarquia e é aleatório, eliminando a necessidade de algum tipo especial de operação de junção ou recuperação quando algum *peer* chega ou sai da vizinhança. A adoção desta abordagem traz uma série de implicações. Como os *peers* são escolhidos aleatoriamente, eles podem estar distantes geograficamente, encarecendo o tráfego de dados. Além disso, nenhum tipo de relação social e temporal entre eles pode ser explorada, o que inviabiliza relações de longa duração.

O conjunto de *peers* para o qual o cliente está enviando dados é denominado conjunto ativo. O protocolo usa a estratégia *tit-for-tat* baseada em taxas para determinar quais *peers* incluir no conjunto ativo. A cada rodada, um *peer* envia blocos para os *peers* em um estado conhecido como *unchoked*, que são aqueles dos quais ele recebeu dados mais rapidamente num passado recente. Esta estratégia tem como objetivo prover incentivos à contribuição no sistema e inibir *freeriders*. Ocasionalmente, os clientes também enviam dados para um pequeno número de *peers* aleatoriamente escolhidos que ainda não adquiriram o status de recepção. Este mecanismo é chamado de *optimistic unchoking* e serve para que *peers* novos ingressem no *swarm* e possam começar a participar do esquema de *tit-for-tat* além de facilitar a descoberta de novas e melhores fontes de dados. Um *peer* pode dissimuladamente fingir ser um novo *peer* por um tempo maior, como demonstrado em [Locher et al. 2006], fazendo com que ele não precise contribuir com o *swarm* e ainda sim receber pacotes de outros *peers* através do *optimistic unchoking*. *Peers* que não enviam dados rápido o suficiente, são removidos do conjunto ativo durante a rodada e ficam no estado de *choked*.

A taxa de *upload* compartilhada de cada *peer* é proporcional ao número de *peers* do conjunto ativo e é dividida igualmente entre eles. Na implementação oficial do *bittorrent* o conjunto ativo é proporcional à $\sqrt{\text{uploadrate}}$ do *peer*, embora em alguns clientes, o tamanho do conjunto seja estático.

A característica centralizadora dos *trackers* também permite a implementação de mecanismos de incentivo e reputação tais como o *sharing-ratio*, que é a obrigação de o *peer* manter uma certa proporção entre seu montante de dados de *upload* e *download*. Entretanto esse mecanismo pode ser burlado, devido ao fato de o *peer* poder anunciar *peers* falsos ao *tracker*, fazendo com que este pense que aquele está compartilhando quando de fato não está. Portanto é necessária a análise individual dos *sharing-ratios*, para que haja a exposição dos *freeriders*.

3.2. Comportamento social e comunidades temporais

A formação de relações de longa duração entre *peers* pode ser utilizada para melhorar a eficiência de distribuição de conteúdo como demonstrado em [Wang et al. 2012], porém o padrão temporal dos *peers online* é altamente diverso: nas medições efetuadas apenas 5% deles puderam se reencontrar, sendo que seus intervalos de permanência na rede não são bem sobrepostos e nem podem ser previstos. Entretanto, um pequeno conjunto de *peers* apresentou um padrão estável, periódico e autossimilar de chegada, que podia ser previsto através de seu comportamento histórico. Tais *peers*, chamados de *peers estáveis*, se mostraram mais frequentes em *swarms* disseminados via *twitter*, tendo um total de mais de 35% de reencontros.

[Pietiläinen and Diot 2012] traça também uma relação entre redes oportunistas e redes sociais. Segundo os autores ambas possuem semelhanças em alguns fatores chave: são altamente clusterizadas, além de os nós serem alcançáveis por um pequeno número de passos, o que é conhecido como fenômeno *small-world*. Também observou-se o aparecimento de *clusters* chamados de comunidades temporais: conjunto de nós que se encontram mais de uma vez durante a medição. Foi constatada estrutura social representativa nas comunidades temporais, sendo que em alguns casos até 80% das comunidades temporais compartilhavam mais de 50% de membros com as redes sociais.

3.3. Agregação multi-torrent

Existem muitos trabalhos com foco em agregação de *torrents* e *peers*, particularmente seguindo a abordagem de [Guo et al. 2005], que mostra que mais de 85% dos *peers* participam de múltiplos *torrents*. [Dan and Carlsson 2009] mostra que muitos *torrents* são compostos de múltiplos *swarms*, que podem ser agregados para uma melhor distribuição de *peers* e aumento de vazão. [Piatek et al. 2008] expõe os problemas de performance decorrentes da publicação de um pacote de arquivos relacionados entre si num único *swarm*. De maneira geral estes trabalhos argumentam que os mecanismos de incentivos são insuficientes porque não têm o escopo duradouro, ou seja, as vantagens recebidas ao semear em um *swarm*, não são passadas para as próximas transferências.

4. Metodologia

4.1. Descrição das Coletas

Para se ter acesso aos anúncios de *peers* ativos em um *swarm* é necessário o ingresso nos mesmos. Baseado em trabalhos anteriores [Wang et al. 2012][Piatek et al. 2008], as coletas deverão conter entre 60000 e 100000 *swarms*. O processamento se dará através de um conjunto de clientes modificados rodando num ambiente distribuído, para obtenção, principalmente, dos IP's dos *peers*, dos quais se pode extrair localização geográfica, sistema autônomo, entre outras informações. Apenas a coleta das informações será feita e não haverá troca real de dados entre os *peers*.

4.2. Métricas

Embora em estagio inicial, é possível extrair as seguintes métricas:

- Popularidade do *swarm* - número obtido da proporção entre *seeders* e *leechers*. Outros fatores como tamanho e idade do *swarm* podem eventualmente fazer parte da medida.

- Disponibilidade de *peers* - proporção entre *peers* alcançáveis pelo IP e *peers* inalcançáveis. É um indicativo para definir se algum *peer* está anunciando propositalmente *peers* falsos para burlar mecanismo de *sharing-ratio*.
- Coeficiente de clusterização - define em qual nível está a organização dos nós em grupos. A divisão em *clusters* é o ponto de partida para a descoberta de comunidades temporais e relações de longa duração.
- Índice de similaridade - define o quão similar é um *peer* em relação aos demais. *Peers* com alto índice de similaridade tem um *overlap* de permanência *online* que abrange uma quantidade maior de outros *peers*, podendo ajudar na disseminação de conteúdo.
- Taxa de contato - Define o número de contatos ou *unchokes* à um nó por unidade de tempo. É uma métrica utilizada para definir a centralidade de um nó. Trabalhos anteriores [Hui et al. 2008] estabeleceram que nós centrais contribuem mais na disseminação de conteúdo. Entretanto, de acordo com [Pietiläinen and Diot 2012], os nós de alta taxa de contato sociais contribuem menos do que os não-sociais.

5. Estado Atual do Trabalho e Próximas Etapas

Até o presente momento estão sendo estudadas as metodologias de coleta e análise expostas na literatura. Também está sendo feito o estudo das métricas mais importantes para caracterização. Coletas preliminares já foram realizadas para demonstrar a viabilidade de obtenção de dados de cada *peer*, como sua localização geográfica e endereço IP. A metodologia de coleta está evidenciada na seção 4.1.

Após a coleta, a caracterização se dará avaliando a incidência de comunidades temporais, relações de longa duração, número de repetições de arquivos e *peers* em *swarms* diferentes.

Em um segundo momento um cliente, ou uma série de clientes serão modificados para que o *unchoking* use como critério a escolha de nós sociais ou de comunidades temporais, com o objetivo de mostrar se os fenômenos encontrados servem para a melhoria da disseminação de conteúdo.

References

- Cohen, B. (2003). Incentives build robustness in bittorrent. *Proc. of IPTPS*.
- Dan, G. and Carlsson, N. (2009). Dynamic swarm management for improved bittorrent performance. *Proc. USENIX IPTPS*.
- Guo, L., Chen, S., Xiao, Z., E.Tan, Ding, X., and Zhang, X. (2005). Measurements, analysis, and modeling of bittorrent-like systems. *Proc. ACM/USENIX IMC*.
- Hui, P., Crowcroft, J., and Yoneki, E. (2008). Bubble rap: Social based forwarding in delay tolerant networks. *Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing*.
- Locher, T., Moor, P., Schmi, S., and Wattenhofer, R. (2006). Free riding in bittorrent is cheap. *Fifth Workshop on Hot Topics in Networks*.
- Piatek, M., Isdal, P., Anderson, T., Krishnamurthy, A., and Venkataramani, A. (2007). Do incentives build robustness in bittorrent? *NSDI*.

- Piatek, M., Isdal, T., Krishnamurth, A., and Anderson, T. (2008). One hop reputations for peer to peer file sharing workloads. *Proc. NSDI*.
- Pietiläinen, A. and Diot, C. (2012). Dissemination in opportunistic social networks: the role of temporal communities. *Proceedings of the thirteenth ACM international symposium on Mobile Ad Hoc Networking and Computing*, pages 165–174.
- Wang, F., Liu, J., Xu, K. ., and Wu, D. (2012). Torrents on twitter: Explore long-term social relationships in peer-to-peer systems. *IEEE Transactions on Network and Service Management*, pages 1–10.
- Wang, H., Wang, F., and Liu, J. (2011). On long-term social relationships in peer-to-peer systems. *IEEE 19th International Workshop on Quality of Service*, pages 1–8.