

Como influenciar seus pares?

Sistemas de recomendação de conteúdo para redes P2P

Diogo M. Vieira¹, Carla A. D. M. Delgado¹, Daniel S. Menasché¹

¹Departamento de Ciência da Computação – Universidade Federal do Rio de Janeiro (UFRJ)
Rio de Janeiro – RJ – Brasil

diogo_munaro@ppgi.ufrj.br, {carla,sadoc}@dcc.ufrj.br

Abstract. *A recommendation system is one of the fundamental components in most distributed systems for content delivery. In essence, a recommendation system matches peers to content, based on their interest. Although these matches are one of the key aspects of distributed systems for content delivery, it has been almost independently analyzed by the content recommendation community. The main goal of this position paper is to show the potential benefits, to user and providers, of taking an holistic perspective on content recommendation where both the networking (QoS) and by the content recommendation aspects are taken into account.*

Resumo. *A humanidade nunca sabe exatamente o que quer. Temos preferências, mas dentro delas fica difícil escolher o que seria melhor para nós. Esses aspectos sociais podem ser usados como novos conceitos para novas tecnologias. Com o uso das tecnologias AVOD (Audio and Video on demand) via P2P (Peer-to-Peer), abrem-se novas perspectivas para implementação dos sistemas de distribuição de conteúdo. E se o conteúdo com melhor QoS (Quality of Service) pudesse ser recomendado com maior probabilidade? Nesse artigo é proposta a utilização de métodos de recomendação para sugerir o melhor conteúdo aos pares (peers) - dentro das suas preferências - levando em consideração o QoS de cada conteúdo para aquele par.*

1. Introdução

Redes P2P tradicionalmente permitem a troca de arquivos com maior escalabilidade e rapidez do que o modelo cliente-servidor, já que podem contar com recursos, como a capacidade de *upload* de vários *peers* [Kurose and Ross 2010]. Um problema enfrentado comumente por essas redes é a falta de *peers* para disseminar algum conteúdo que ficou esquecido, ou até a falta de alguma parte de um conteúdo.

A tecnologia P2P é aplicada em vários ambientes, desde AVOD [Wu and Lui 2012, Zhou et al. 2012, Ma et al. 2012] e ambientes de colaboração [Saddik et al. 2007] até um dos comunicadores mais utilizados do mundo: o Skype [Kurose and Ross 2010]. Para todos esses ramos da tecnologia P2P já existe uma série de algoritmos de recomendação visando maior comodidade do usuário em encontrar um melhor conteúdo de acordo com seu perfil [Hecht et al. 2012]. Mesmo com a comprovação de que os mecanismos de recomendação possuem grande influência na decisão de consumo - principalmente multimídia (em média 30% das visualizações de cada vídeo do YouTube é oriunda de recomendação [Zhou et al. 2010]) - nenhum desses

algoritmos de recomendação foi até então usado para influenciar a QoS fornecida pela rede [Sitaraman 2013].

Nesse trabalho propomos a utilização de um algoritmo de recomendação que, além de priorizar os aspectos comuns de *precision and recall* para manter a qualidade das recomendações, melhora o QoS da rede P2P; para tal, avaliamos como o sistema de recomendação (SR) pode fazer uso de métricas de QoS para sugerir conteúdo visando maximizar a satisfação do usuário e, reciprocamente, como o SR pode ser usado para aumentar a QoS experimentada por usuários que requerem conteúdo que, *a priori*, seria pouco replicado.

O artigo inicia com uma seção mostrando alguns trabalhos que nos inspiraram na seção 2, seguindo para uma solução que propomos para alcançar nossos objetivos na seção 3. A partir dessa solução já temos alguns resultados parciais na seção 4 e discutimos sua relevância com perspectivas futuras na seção 5.

2. Trabalhos Correlatos

Os pilares do presente trabalho são redes P2P e SRs de conteúdo. Embora a literatura nessas duas áreas seja vasta, não é de nosso conhecimento nenhum trabalho sobre os possíveis benefícios de uma abordagem integrada para estudo desses dois tópicos.

No escopo P2P, existem alguns trabalhos utilizando AVOD P2P [Zhou et al. 2012, Wu and Lui 2012]. Dentro desse escopo, nos deparamos com uma maior atenção a estudos de AVOD utilizando o protocolo BitTorrent [Ma et al. 2012], que tradicionalmente é um dos protocolos mais usados para transferências de arquivos, principalmente arquivos multimídia [Mendonça and Leão 2012].

Arquivos multimídia estão em foco também na área de recomendação [Mei et al. 2007, Chakoo et al. 2008], principalmente depois dos prêmios oferecidos pelo Netflix para melhorar a recomendação do seu sistema AVOD [Netflix 2013]. Um exemplo recente de recomendação multimídia em sistemas P2P é o Radiomender [Hecht et al. 2012], que propôs um SR de músicas distribuído em que o *peer* possui sua rádio personalizada. Mesmo com toda essa inovação, nenhum desses trabalhos tentou melhorar a QoS da rede utilizando recomendação, sendo assim, a recomendação fica voltada somente para os usuários e o conteúdo, não levando em conta a rede.

3. Solução Proposta

Tendo em vista a oportunidade de poder melhorar o QoS da rede sem necessidade de ampliar os recursos para disseminação de conteúdo - como contratar CDNs ou melhorar servidores - ou realizar qualquer alteração do ponto de vista do usuário [Menasche et al. 2013], *a recomendação de conteúdo multimídia se mostra interessante tanto para usuários como para administradores de rede.*

Propomos aqui estratégias adaptativas para satisfazer o desempenho dos sistemas e dos usuários, em que se torna possível regular o quanto o QoS da rede será prevalecido em detrimento da melhor recomendação que deve ser sugerida para o usuário, ou decidir em favor da recomendação. Para isso é necessário que seja levado em conta durante a recomendação de um conteúdo, a QoS com que ele poderá ser servido. Isso tanto melhora a qualidade da rede, quanto a satisfação dos usuários - principalmente se tratando de AVOD - em que o serviço requer melhor QoS [Zhou et al. 2012, Wu and Lui 2012].

A ideia geral é apresentada na Figura 1, onde ilustramos o estado atual dos SRs e nossa proposta, que coloca a rede e o QoS dela como beneficiários do sistema também. Assim o usuário não necessariamente recebe o conteúdo que ele pode gostar mais, mas sim um conteúdo que ele gostará e terá melhor QoS - além de poder aumentar a QoS de outros usuários.

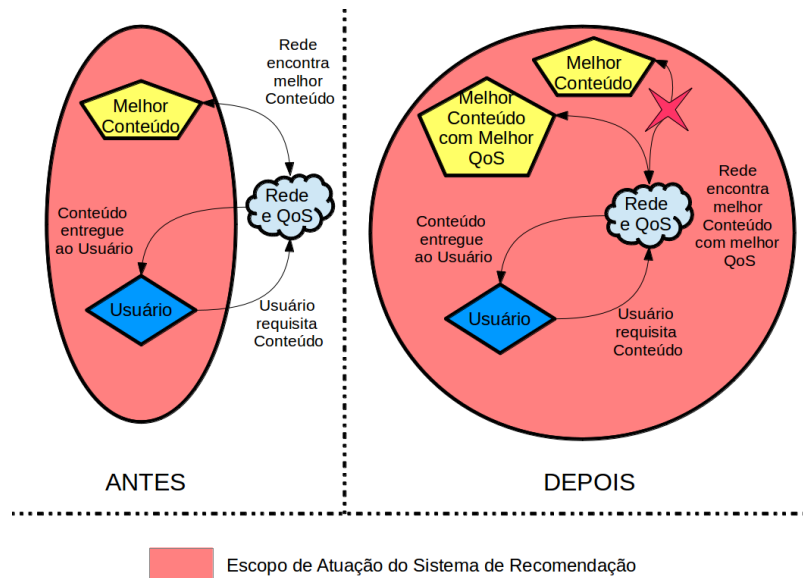


Figura 1. Mudança de atuação do sistema de recomendação (SR) que além de pensar na satisfação do usuário leva em conta a QoS (Qualidade do Serviço) que ele pode conseguir.

4. Resultados Preliminares

A seguir, usamos simulações para avaliar os potenciais benefícios de uma visão integrada de redes P2P e SRs. Nossos objetivos são: indicar os potenciais problemas encontrados em sistemas nos quais não haja flexibilidade alguma para recomendação de conteúdo e; ilustrar como que um sistema no qual os usuários têm flexibilidade sobre como farão o *download* e consumo dos conteúdos pode beneficiar-se de recomendações. Assim, na seção 4.1 descrevemos o simulador implementado, e na seção 4.2, apresentamos resultados preliminares.

4.1. Cenário Experimental

Para análise da rede P2P foi implementado em Python um simulador com diversas configurações para verificar *downloads* de arquivo P2P e consumo desses arquivos, sendo que cada *peer* participante do sistema possui uma lista de arquivos que já traz consigo e outra lista de arquivos que irá consumir (*playlist*). Assumimos que a interseção dos conteúdos que o usuário traz consigo ao sistema e de sua *playlist* é vazia. Como parâmetros da simulação foram utilizados 50 *peers*, sendo que cada um deles possuía uma *playlist* e uma lista de arquivos com 120 arquivos cada. Essas duas listas derivavam sempre do total de 5000 arquivos do sistema de forma aleatória.

O tempo foi dividido em *timeslots* e os *peers* interagiam a cada *timeslot* pareando aleatoriamente e verificando se estavam interessados em algum arquivo que o *peer* com

quem parearam possuía em sua lista de arquivos para realizar um *download*. Na simulação foram mantidos alguns parâmetros fixos e como variáveis a taxa de *download*, o tipo de *download* e o tipo de consumo.

O *download* ocorre quando um *peer* pega um arquivo da lista de arquivos do outro *peer* para sua lista de arquivos de acordo com sua taxa de *download*. As taxas de *download* utilizadas foram 10%, 30%, 50%, 70% e 100% e são definidas como a porcentagem do arquivo a ser feito o *download* a cada *timeslot*. O tipo de *download* pode ser sequencial em que os *peers* realizam *download* dos arquivos da *playlist* sequencialmente; ou aleatório em que o *peer* efetua *download* de qualquer arquivo de sua *playlist* que esteja na lista de arquivos do *peer* com quem pareou. O *download* aleatório simula um SR que indica um melhor conteúdo a ser feito *download* dentro das preferências do usuário (*playlist*). Cada *download* era feito totalmente ou parcialmente dependendo da taxa de *download* estipulada.

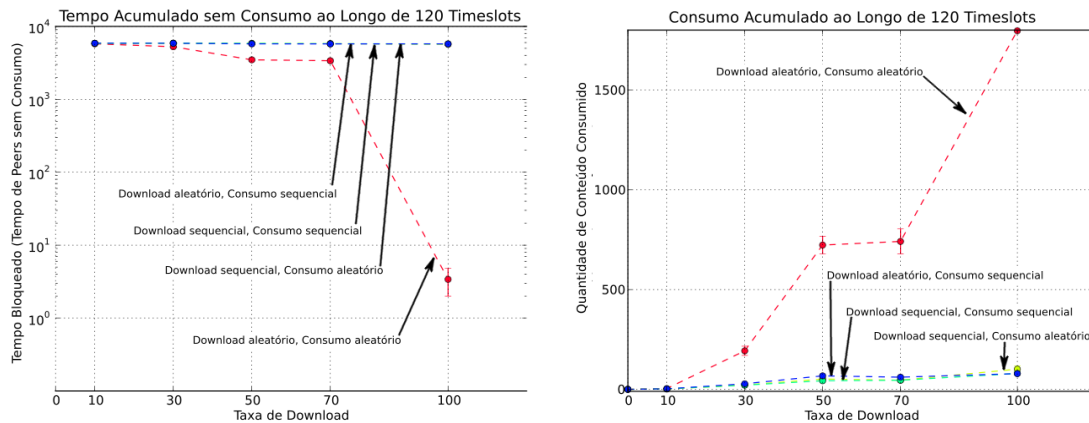
Depois de realizar um *download*, os *peers* consomem os arquivos da *playlist* que sofreram *download* e o *timeslot* se encerra. O consumo ocorre segundo uma taxa de consumo. A taxa de consumo utilizada foi fixada em 30% e é a porcentagem do arquivo a ser consumido a cada *timeslot*. Esse consumo é o equivalente a ouvir um áudio ou assistir um vídeo em um sistema AVOD, então assumimos que o *peer* só começa a consumir outro arquivo quando termina o corrente. O tipo de consumo pode ser sequencial em que os *peers* realizam consumo dos arquivos da *playlist* sequencialmente; ou aleatório em que a ordem de consumo é escolhida de forma aleatória entre arquivos com *download* concluídos ou parcialmente concluídos, novamente possibilitando uma segunda camada de abstração para o SR.

Em cada simulação, todos os *peers* possuem a taxas de *download* e consumo fixas em um dos valores pré-definidos. Cada rodada de simulação dura 120 *timeslots* e foram repetidas 10 rodadas para cada grupo de parâmetros.

4.2. Desafios e Soluções

Nessa seção iniciamos apresentando resultados referentes a cenários em que os *peers* não têm flexibilidade para ajustar seus *downloads* e consumos de acordo com recomendações do sistema. Nesses casos, os dois principais desafios encontrados em nossas simulações foram: 1) eventualmente *peers* ficam bloqueados sem ter conteúdo para consumir e 2) ao final do tempo de simulação muitos *peers* não conseguem obter grande parte do conteúdo desejado. Indicaremos então que, se assumirmos que os *peers* têm flexibilidade para ajustar seus *downloads* e consumos, a situação muda radicalmente.

Nas Figuras 2(a) e 2(b) apresentamos o tempo acumulado sem consumo de todos os *peers* em *timeslots* e o consumo acumulado de arquivos de todos os *peers*, respectivamente, em função da taxa de *download*. Essas duas figuras servem para ilustrar os problemas 1 e 2 descritos acima, respectivamente. Na Figura 2(a) percebemos que as simulações com *consumo sequencial*, independente da taxa de *download* ou do tipo de *download* (sequencial ou aleatório), resultam em muito tempo com os *peers* bloqueados. O mesmo se repete com as simulações com *download sequencial - consumo aleatório*, já que o *download* força o consumo a ser sequencial também, pois não são feitos *downloads* fora da ordem da *playlist*. A Figura 2(b) mostra outro problema dos sistemas com *download sequencial - consumo aleatório*, *download sequencial - consumo sequencial* e



(a) Tempo que *peers* ficam bloqueados. Somente com download e consumo aleatórios foi possível consumo contínuo.

(b) Quantidade de arquivos completamente consumidos pelos *peers*. Observamos que os ganhos obtidos quando se combina flexibilidade no consumo com flexibilidade no *download* são grandes.

Figura 2. Resultado do consumo dos *peers* para simulações realizadas

download aleatório - consumo sequencial. Nesses sistemas, os *peers* não conseguem fazer *download* de todo o conteúdo desejado. A quantidade de arquivos consumidos cresce sutilmente na medida em que a taxa de *download* aumenta, mas não passou de 90 arquivos ao longo dos 120 *timeslots*, em nenhuma das rodadas de simulação consideradas.

As Figuras 2(a) e 2(b) mostram que o comportamento do sistema muda radicalmente quando consideramos que os *peers* têm flexibilidade para decidir a ordem de *download* e consumo do conteúdo (curvas vermelhas). Nesse caso (*download* e consumo aleatórios), isso ocorre devido ao aumento da probabilidade de *download* de um arquivo durante o pareamento entre dois *peers* por pelo menos um desses *peers*, pois a possibilidade de busca aumenta. Além disso, o consumo aleatório permite que qualquer arquivo da *playlist*, que tenha seu *download* em andamento ou concluído, seja consumido.

Note que é essencial a flexibilidade tanto do *download* quanto do consumo. Se assumirmos que apenas o consumo é flexível (curva *download sequencial - consumo aleatório*), o desempenho do sistema não é satisfatório. Dessa forma, é fundamental que qualquer sistema de recomendação que vise aumentar o QoS da rede afete não apenas o consumo, mas também a ordem em que os conteúdos são baixados.

5. Conclusão

Neste trabalho discutimos alguns dos potenciais benefícios de se estabelecer uma visão unificada sobre redes P2P e recomendação de conteúdo. Nós acreditamos que as ideias aqui apresentadas podem ter importante impacto em outras redes, como em ICNs (*Information Centric Networks*) para armazenamento do melhor conteúdo [Ghods et al. 2011, Perino and Varvello 2011], ou até mesmo na determinação do melhor conteúdo a ser guardado no *cache* de servidores *proxy* para diminuir a quantidade de *cache misses* em uma rede privada.

Essa visão unificada pôde ser verificada no simulador construído que, por sua vez, possibilita o pareamento de *peers* com arquivos e preferências semelhantes mais

facilmente, permitindo testar mudanças na recomendação e no QoS.¹

Referências

- [Chakoo et al. 2008] Chakoo, N., Gupta, R., and Hiremath, J. (2008). Towards Better Content Visibility in Video Recommender Systems. *2008 Japan-China Joint Workshop on Frontier of Computer Science and Technology*, 201305:181–185.
- [Ghodsi et al. 2011] Ghodsi, A., Shenker, S., Kooponen, T., Singla, A., Raghavan, B., and Wilcox, J. (2011). Information-centric networking: seeing the forest for the trees. In *HotNets-X*, pages 1:1–1:6, New York, NY, USA. ACM.
- [Hecht et al. 2012] Hecht, F., Bocek, T., Bar, N., Erdin, R., Kuster, B., Zeeshan, M., and Stiller, B. (2012). Radiommender: P2p on-line radio with a distributed recommender system. In *P2P Computing, 2012 IEEE 12th*, pages 73–74.
- [Kurose and Ross 2010] Kurose, J. F. and Ross, K. W. (2010). *Computer Networking: A Top-Down Approach (5th Edition)*. Addison Wesley, 5 edition.
- [Ma et al. 2012] Ma, Z., Xu, K., Liu, J., and Wang, H. (2012). Measurement, modeling and enhancement of BitTorrent-based VoD system. *Computer Networks*, 56(3):1103–1117.
- [Mei et al. 2007] Mei, T., Yang, B., Hua, X., and Yang, L. (2007). VideoReach: an online video recommendation system. *SIGIR '07. ACM*, pages 767–768.
- [Menasche et al. 2013] Menasche, D. S., Massoulie, L., and Towsley, D. (2013). Reciprocity and Barter in Peer-to-Peer Systems. In *ICC'2013*.
- [Mendonça and Leão 2012] Mendonça, G. G. and Leão, R. M. M. (2012). BTStream — Um ambiente para desenvolvimento e teste de aplicações de streaming P2P. In *XXX SBRC, Salão de Ferramentas*, Ouro Preto, MG, Brazil.
- [Netflix 2013] Netflix (2013). Wikipedia Netflix Prize. http://en.wikipedia.org/wiki/Netflix_Prize. [Online; accessed 11-March-2013].
- [Perino and Varvello 2011] Perino, D. and Varvello, M. (2011). A reality check for content centric networking. In *ACM SIGCOMM workshop on ICN, ICN '11*, pages 44–49, New York, NY, USA. ACM.
- [Saddik et al. 2007] Saddik, A., Rahman, A., Abdala, S., and Solomon, B. (2007). PE-COLE: P2P multimedia collaborative environment. *Multimedia Tools and Applications*, 39(3):353–377.
- [Sitaraman 2013] Sitaraman, R. K. (2013). Network performance: Does it really matter to users and by how much? pages 1–10.
- [Wu and Lui 2012] Wu, W. and Lui, J.-S. (2012). Exploring the optimal replication strategy in p2p-vod systems: Characterization and evaluation. In *INFOCOM, 2011 Proceedings IEEE*, pages 1206–1214.
- [Zhou et al. 2010] Zhou, R., Khemmarat, S., and Gao, L. (2010). The impact of YouTube recommendation system on video views. *IMC '10*, page 404.
- [Zhou et al. 2012] Zhou, Y., Fu, T., and Chiu, D. (2012). A unifying model and analysis of p2p vod replication and scheduling. *INFOCOM, 2012 Proceedings IEEE*.

¹Mais detalhes e o código fonte do simulador estão em: <http://code.google.com/p/simuqos>