

Disponibilidade de Conteúdo Promovida por Tocadores de Playlists em Sistemas CDN-P2P

J. A. Oliveira¹, César A.V. Melo¹, Nelson L.S. da Fonseca²

¹ Instituto de Computação – Universidade Federal do Amazonas
Manaus – AM – Brasil.

{cavmelo}@icomp.ufam.edu.br

²Instituto de Computação – Universidade de Campinas
Campinas – SP – Brasil

nfonseca@ic.unicamp.br

Abstract. *In this paper, the effectiveness of playlists for improving content availability in CDN-P2P system that distributes YouTube-like videos is assessed. In such system, a playlist allows content to be accessed in a row. Using real data collected from YouTube a set of numerical evaluations was carried out to measure the importance of key parameters, such as the arrival rate and the playlist length. Results show that the improvement on content availability is about 40% for high arrival rate of playlist viewers. For scenarios with moderate arrival rate, the improvement exceeds 20%.*

Resumo. *Neste artigo, avalia-se a efetividade da presença de tocadores de listas de reprodução (playlists) na melhoria da disponibilidade de conteúdo em um sistema CDN-P2P. Nesse sistema playlists permitem a reprodução de conteúdo de diversos arquivos em sequência. A partir de dados reais coletados do YouTube avaliações numéricas foram realizadas para medir a importância de parâmetros chave, tais como a taxa de chegada e o tamanho das playlists. Os resultados do estudo numérico mostram que a melhoria na disponibilidade de conteúdo é superior a 40% quando o sistema apresenta uma alta taxa de chegada de pares tocadores de playlists. Para sistemas com taxas de chegada moderadas a melhoria é superior a 20%.*

1. Introdução

Em sistemas distribuição de conteúdo híbrido (CDN-P2P), o conteúdo armazenado nos servidores da rede de distribuição de conteúdo (CDN) é compartilhado pelos pares da rede par-a-par (P2P). Sempre que um par é o único detentor de um conteúdo e ele sai do sistema, os demais pares precisam acessar o servidor da CDN para obter esse conteúdo e torná-lo disponível para os demais pares. O atraso gerado com esse acesso degrada a qualidade de experiência do pares, e pode formar congestionamento em torno do servidor.

Quanto mais tempo um par passa conectado à rede, maior é a disponibilidade de conteúdo associado a ele. Em [Wang et al. 2008], os autores estudam a importância de pares estáveis em sistemas P2P de distribuição de fluxo de vídeo ao vivo. O ajuntamento de conteúdo popular e não-popular (*blundling*), para aumentar a disponibilidade de conteúdo, especialmente o não popular, foi estudado em [Menasche et al. 2009].

Em um sistema CDN-P2P que distribui conteúdo similar ao disponibilizado pelo YouTube, pares estáveis acessam playlists de vídeos, que são coleções de vídeos criadas e mantidas pelos seus donos. Embora à edição dessas playlists fique a cargo de seus donos, o número de vídeos que podem ser inseridos nas playlists é estabelecido pelos provedores. Por exemplo, usuários do YouTube podem criar playlists com até 200 vídeos. A maioria dos vídeos dessas playlists são sobre tópicos específicos e agregam vídeos populares e não populares, de programas de TV, grandes eventos de esporte e apresentações de artistas.

Neste artigo, investiga-se o impacto de tocadores de playlists na disponibilidade de conteúdo em redes CDN-P2P. Coleções de vídeos e playlists de diferentes tópicos foram coletadas no YouTube, e a popularidade dos vídeos dessas coleções foi caracterizada. Além disso, um sistema CDN-P2P com tocadores de playlists foi simulado considerando diversos cenários. Os resultados indicam que mais de 40% das demandas geradas pelos pares podem ser atendidas na parte P2P do sistema, quando o sistema tem uma alta taxa de chegada de tocadores de playlists. Em sistemas com baixa taxa de chegada, a disponibilidade é melhorada em pelo menos 20%, indicando que, mesmo em comunidades de pares menos participativos, tocadores de playlists podem ajudar a melhorar a disponibilidade de conteúdo em sistema CDN-P2P.

As contribuições deste artigo são **o estudo da popularidade de conteúdo multimídia para a formação de playlists e a medição do impacto dos tocadores de playlists na disponibilidade de conteúdo**. Foram usados mais de dois milhões de vídeos na análise da popularidade dos vídeos, e na forma como essa popularidade influencia a montagem de playlists. Para tal simulou-se o sistema de distribuição de conteúdo e avaliou-se a melhoria na disponibilidade do conteúdo quando tocadores de playlists estão presentes.

Os resultados do presente artigo estendem os resultados publicados em [Melo et al. 2013] pela avaliação do impacto que a interrupção das playlists causa na disponibilidade de conteúdo proporcionada por elas, tendo-se também quantificado as requisições feitas por pares comuns que são exclusivamente atendidas pelos tocadores de playlists. Além disso, os estudos realizados envolvem três novas coleções, Comédia, Filmes&Animações e Educação.

O restante do artigo está organizado da seguinte forma. Na Seção 2, os trabalhos relacionados são apresentados. Na Seção 3, as coleções de vídeos usadas são analisadas e na Seção 4, o cenário de rede considerado nos estudos numéricos é discutido. Na Seção 5, os resultados numéricos derivados dos experimentos de simulação são discutidos. Finalmente, as conclusões são apresentadas na Seção 6.

2. Trabalhos Relacionados

Nessa seção, trabalhos desenvolvidos previamente para melhorar a disponibilidade de conteúdo de sistemas de distribuição são discutidos. Os trabalhos relacionados envolvem medição [Guo et al. 2008], modelagem [Menasche et al. 2009] e implementação [Wang et al. 2008] de sistemas de distribuição. No que segue, dois trabalhos mais diretamente relacionados ao presente artigo serão discutidos.

Em [Menasche et al. 2009], os autores estudam o impacto da forma como conteúdos são agregados, chamada *bundling*, na disponibilidade desses conteúdos em um sistema de compartilhamento de arquivos baseado em *swarm*. Essa técnica ajunta vários arquivos com diferentes níveis de popularidade em um único grande *swarm*. Uma fila $M/G/\infty$ foi

proposta para modelar o sistema. Os autores mostram que um ajuntamento de K arquivos é capaz de aumentar a disponibilidade de conteúdo não popular por um fator de $e^{\Theta(K^2)}$. O presente estudo mostra que a playlist guarda similaridade com o *blundling*, dado que vídeos populares e não tão populares são ajuntados e tocados sequencialmente, se for da conveniência dos usuários. Uma playlist é um tipo especial de *bundle*, na qual usuários podem selecionar quais conteúdos irão assistir. Esse tipo de *bundle* é também chamado de *blundle* misto. Enquanto em [Menasche et al. 2009] os autores estudam apenas *bundles* puros em sistemas do tipo bitTorrent, no presente artigo, consideram-se *bundles* mistos, em sistemas do tipo YouTube.

Wang et al. [Wang et al. 2008] avaliam como pares estáveis em um sistema P2P podem melhorar uma sessão de fluxo de vídeo ao vivo, ao manterem conteúdo disponível durante parte da sessão. Os autores mostram que a QoS de sistemas P2P de fluxo de vídeo pode ser melhorada quando pares estáveis são identificados e organizados em árvores de distribuição melhorando a resiliência do sistema. Nesse sistema, a estabilidade dos pares é medida pela razão entre a participação do par em um fluxo e a duração da sessão de fluxo, classificando-se os pares como estável ou instável. Resultados numéricos mostram que a taxa de perda de dados, atraso de inicialização e atraso na reprodução são reduzidos nos sistemas P2P quando operam com os algoritmos propostos. Enquanto em [Wang et al. 2008], os autores estudam um sistema de fluxo ao vivo e a estabilidade do par é definida usando diferentes critérios, o estudo no presente artigo considera o fato do par está acessando uma playlist para caracterizá-lo como estável, e o sistema considerado é o de fluxo de vídeo sob-demanda do tipo YouTube.

3. Caracterização das Coleções de Vídeo do YouTube

Nessa seção, as coleções de vídeos usadas nos estudos numéricos são apresentadas, ou seja, as coleções de vídeos Top-Feed e as coleções de playlists. Ambas as coleções foram obtidas do site YouTube, e incluem metadados dos vídeos e playlists.

As coleções de vídeos e playlist foram analisadas para que se identificasse o papel da popularidade dos vídeos na montagem das playlists. Primeiro, o padrão de idade de vídeos e playlists foram identificados, e em seguida, esse padrão foi usado para estimar a popularidade do vídeo no instante em que as playlists eram criadas.

3.1. As Coleções de Vídeos e Playlists

Os metadados da coleção Top-Feed foram obtidos por um coletor construído a partir do método bola-de-neve (*snowball sampling*) que implementa uma busca em largura sobre as componentes conexas de um grafo G onde se encontram as sementes da busca. Para as coleções Top-Feed as sementes foram definidas pelos vídeos associados aos seguintes feeds da página principal do YouTube: Mais qualificados, Favoritos, Mais Vistos, Mais Populares, Mais Discutidos e Mais Respondidos. Esses feeds podem ser filtrados por país e por categoria de vídeo. Dado que o sistema de distribuição, descrito neste artigo, leva a localização e o contexto em consideração, os conteúdos oferecidos aos brasileiros foram coletados e filtrados por seis das dez categorias listadas no YouTube. A Tabela 1 mostra dados das coleções coletadas.

Considerando a busca em largura implementada no método bola-de-neve, uma fila foi construída e alimentada com informações dos vídeos coletados dos feeds padrões.

Para cada vídeo desenfileirado, os seguintes dados foram coletados: ID do vídeo, duração, visualizações, qualificação, autor, data de publicação, e um conjunto de 25 vídeos sugeridos pelo sistema de recomendação do YouTube. Os vídeos desse conjunto são enfileirados e usados no prosseguimento da busca em largura, desde que não tenham sido previamente visitados.

A coleção do YouTube foi explorada até um nível previamente estabelecido, permitindo que um número máximo de vídeos relacionados possam ser acessados. O limiar do nível foi estabelecido com base na duração média de uma sessão de navegação no YouTube [YouTube 2012], que é de 15 minutos. Assim, em cada coleta, definiu-se um limiar para o nível de exploração da busca em largura de tal forma que o número máximo de vídeos relacionados pudesse ser acessado em uma sessão de conexão.

Tabela 1. Estatísticas das Coleções Top-Feed e Playlist

Categoria	Coleção Top-feed				
	Princípio de Pareto	Nro de Vídeos	Duração(seg.)		Data da Coleta
			Média(μ)	σ	
Música	80% - 7%	535,687	222.6	86.5	2012-07-27
Esporte	80% - 10%	226,043	169.2	129.3	2012-08-01
Entretenimento	80% - 7%	413,524	219.3	175.2	2012-08-03
Filmes&Animações	80% - 12%	535,687	250.0	208.6	2012-11-10
Comédia	80% - 12%	448,591	235.5	188.7	2012-11-12
Educação	80% - 10%	330,843	433.0	278.3	2012-11-08
Categoria	Coleção Playlist				
	Princípio de Pareto	Nro de Playlists	Por Playlist (média)		Data da Coleta
			Duração em seg.	Nro de Vídeos	
Música	80% - 4.0%	22,550	2,607.7	9.80	2012-07-27
Esporte	80% - 2.0%	10,602	1,647.2	10.07	2012-08-01
Entretenimento	80% - 2.4%	14,969	2,190.6	12.03	2012-08-03
Filmes&Animações	80% - 3.2%	16,205	2,254.2	12.33	2012-11-15
Comédia	80% - 4.0%	19,105	2,028.1	11.20	2012-11-17
Educação	80% - 2.0%	16,788	4,762.6	12.24	2012-11-20

O outro conjunto de dado, chamado de coleção de playlists, é composto por vídeos que estão em playlists do YouTube. Toda playlist criada no YouTube tem um dono e os vídeos em uma playlist são acessados seguindo a sequência estabelecida pelo dono. Usando entradas da coleção Top-Feed, um conjunto S composto por pares (V_i, D_i) que definem o Dono D_i do Vídeo V_i foi criado. Usando S , coletou-se todas as playlists de dono D_i que possuem o vídeo V_i . As playlists foram processadas para que se tornassem únicas na coleção, uma vez que dois ou mais vídeos de um mesmo dono podem pertencer à mesma playlist. A Tabela 1 mostra as estatísticas das playlists coletadas.

3.2. Caracterizando a Popularidade dos Vídeos das Coleções Top-Feed e Playlist

A popularidade de conteúdos multimídia tem sido modelada pela distribuição Zipf bem como distribuições que são variações da Zipf, tais como Zipf com corte Exponencial e Mandelbrot-Zipf. A distribuição Exponencial Dilatada, do inglês *Stretched Exponential*,

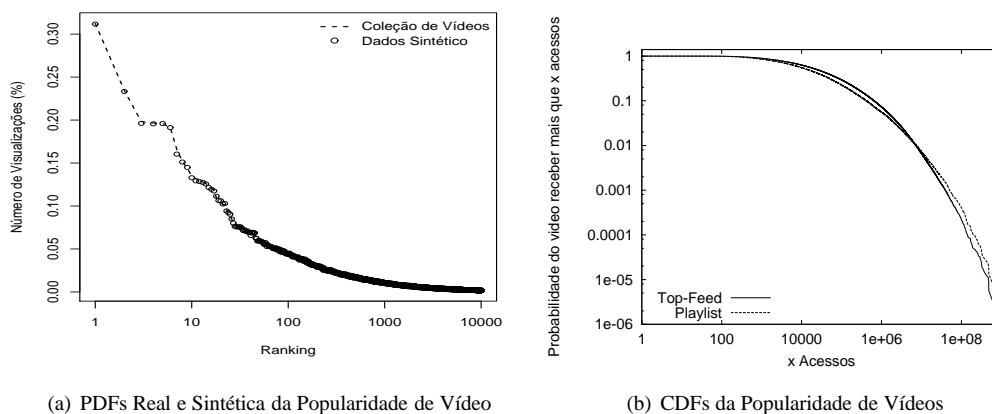


Figura 1. Modelando Popularidade dos Vídeos das Coleções Top-Feed e Playlist de Música

tem também sido considerada como apropriada para modelar a popularidade de vídeos [Guo et al. 2008]. Tais conclusões foram derivadas usando diferentes coleções de vídeos, disponibilizadas em uma variedade de sistemas de distribuição. Entretanto, o que existe de comum entre essas abordagens é que elas são todas distribuições que modelam a lei-da-potência.

O método usado para amostrar dados pode afetar significativamente os resultados obtidos. Especificamente, os autores de [Lee et al. 2005] mostraram que valores de medidas feitas usando dados que seguem a lei-de-potência são muito sensíveis à amostragem.

Reconhecendo as dificuldades de modelar popularidade em redes sociais on-line, especialmente usando as distribuições mencionadas, aplicou-se o método de Aceitação-rejeição para gerar traços sintéticos de requisições de vídeos com mais de 800 milhões de entradas. A Figura 1.a mostra a função de distribuição de probabilidade na coleção de vídeo Top-Feed Música e o seu similar sintético.

A possível similaridade entre os padrões de popularidade dos vídeos nas coleções Top-Feed e Playlist foi avaliada. A Figura 1.b mostra as coleções de Playlist e Top-feed de música classificada por visualizações. As curvas mostram que ambas as coleções compartilham o mesmo padrão de popularidade, que pode ser resumida da seguinte forma: a probabilidade de um vídeo receber um grande número de visualizações cai rapidamente a medida em que esse vídeo não está nas primeiras posições do ranking. Em outras palavras, em ambas as coleções, um subconjunto de vídeos recebe a maioria das visualizações.

Para verificar como as visualizações estão distribuídas, as coleções foram analisadas segundo o princípio de Pareto, também conhecido com regra 80-20%. Essa regra, que define um padrão geral de popularidade de objetos na Internet, estabelece que 20% dos objetos recebe 80% de todos os acessos realizados. Na playlist da coleção de música 80% das visualizações estão associadas a somente 4% de seus vídeos, enquanto na coleção Top-Feed 80% das visualizações são associadas a 7%. Para as coleções de vídeos de Entretenimento esses números são 80% - 2,4% (Playlist) e 80% - 7% (Top-Feed), e para as coleções de Esporte tem-se 80%-2% (Playlist) e 80% - 10% (Top-Feed), (ver Tabela 1 para as demais coleções). Esses resultados mostram que as popularidades dessas coleções divergem da regra geral 80-20%, especialmente as playlists. Tal divergência pode ser ex-

plicada pela dinâmica de navegação nos vídeos do YouTube, que em geral é iniciada nos vídeos da página principal, através dos feeds padrão, e então desce para vídeos de menor popularidade, seguindo as sugestões do sistema de recomendação do YouTube, e finalizando após 15 minutos de navegação em média [YouTube 2012].

Em resumo, tem-se que a curta duração das sessões de acesso e a atuação do sistema de recomendação, tornam as requisições à coleção de vídeo do Youtube mais concentradas a um subconjunto que é menor que aquele definido pelo princípio de Pareto.

3.3. A Influência da Popularidade dos Vídeos na Montagem das Playlists

Playlists podem ser criadas e mantidas inalteradas, ou podem ser atualizadas periodicamente. O YouTube não registra a data em que o vídeo foi adicionado a playlist, embora exista registro da data em que os vídeos foram disponibilizados, i.e a data de *upload*, e da data em que as playlists foram publicadas. Essa informação foi processada e verificou-se que a maioria dos vídeos são mais velhos que as playlists em que eles estão listados. Por exemplo, na coleção de música, 71% dos vídeos são mais velhos que as suas playlists, para as coleções de Entretenimento, Esporte, Educação, Comédia, Filmes&Animações essas proporções são 67%, 69%, 59%, 63% e 53%, reforçando que vídeos selecionados para playlists tinham um lastro de visualizações quando foram incluídos nas playlists.

Crane e Sornette[Crane and Sornette 2008] argumentam que a popularidade de vídeos no YouTube pode ser classificada em quatro classes. A primeira classe, composta por 90% dos vídeos, apresenta pouca flutuação em sua popularidade, e as outras três classes têm vídeos que, após um período de dormência, experimentam um aumento rápido de visualizações. Essas três classes são nomeadas como Viral, Qualidade e Lixo, devido à intensidade das atividades relacionadas a elas em um curto período de tempo. Os autores mostram que o período de dormência não dura mais que 250 dias. Nas coleções de vídeos da playlist, 76%, 67% e 77% dos vídeos mais populares nas coleções de Música, Entretenimento e Esporte, respectivamente, já tinham passado o período de dormência quando as playlists que eles estão listados foram publicadas. Para as coleções Comédia, Educação e Filmes&Animações verificou-se que 66.7%, 66.1% e 57.7% dos vídeos, respectivamente, já haviam ultrapassado aquele período de dormência.

Ao considerar as estatísticas de popularidade e idade dos vídeos e playlists, é razoável inferir que o dono da playlist tinha informações suficientes para escolher, entre os vídeos disponíveis nas coleções, os mais relevantes em relação ao tema da playlist bem como aqueles que já possuíam um número significativo de visualizações.

4. Distribuição de Conteúdo no Sistema Avaliado

Nessa seção, apresenta-se o sistema distribuição de conteúdo estudado neste artigo, bem como as diretrizes gerais da distribuição do conteúdo a partir do sistema. Os estudos realizados utilizaram simulação numérica tendo-se avaliado características do sistema que impactam a distribuição do conteúdo, especificamente a presença de pares que permanecem um maior tempo no sistema.

O sistema de distribuição de conteúdo é um sistema híbrido CDN-P2P concebido para distribuir vídeos similares aos distribuído pelo YouTube. A Figura 2 mostra a arquitetura do sistema avaliado. Nesse sistema, a requisição para um vídeo pode ser atendida por um par da porção P2P ao invés de um servidor da CDN. Para que isso aconteça,

vídeos acessados por um par da porção P2P devem ser armazenados no par tornado-o um servidor desses vídeos. Adicionalmente, tem-se que os pares são agrupados em comunidades construídas em torno de um tema, por exemplo, Música, Educação, Esporte, Entretenimento e Jogos, facilitando o compartilhamento de conteúdo. Esse agrupamento também restringe a operação de busca aos membros da comunidade que tomam conhecimento de seus vizinhos a partir de metadados disseminados pelos servidores da CDN. Embora os pares possam estar em várias comunidades simultaneamente, é muito mais provável que eles permaneçam em uma única comunidade. De fato, estatísticas mostram que sessões do YouTube duram, em média, 15 minutos, período relativamente pequeno para um par acessar dois vídeos com temas completamente diferentes. Além disso, a maioria das visualizações, algo como 60% das requisições para vídeos, vem da lista de vídeos relacionados, construída pelo sistema de recomendação que indica sua relevância [Davidson et al. 2010].

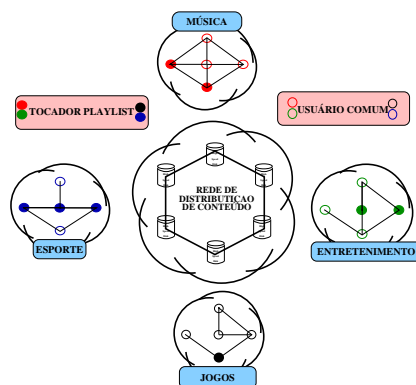


Figura 2. A Arquitetura do Sistema CDN-P2P

Nesse sistema, existem dois tipos de usuários: os que acessam os vídeos diretamente da coleção de vídeos e aqueles que acessam os vídeos a partir de uma playlist. O primeiro tipo, chamado de usuário comum, investe algum tempo considerando os vídeos recomendados ou mesmo navegando pela coleção de vídeos antes de realizar o acesso a um vídeo. O último, chamado de tocador de playlist, acessa os vídeos a partir de uma playlist, seguindo a ordem em que eles se encontram na playlist, não investindo tempo em considerações sobre que vídeo deve ser acessado.

O servidor da CDN que atende uma comunidade também atua como um tracker, com características similares ao utilizado no BitTorrent [Joung and Huang 2012]. Este tracker possui um mapa do sistema, com informações sobre os conteúdos acessados pelo pares da comunidade. Assim, quando um par chega ao sistema e realiza a requisição de um determinado conteúdo, sua requisição será dirigida ao tracker que pode então indicar uma lista de pares que já possuem o conteúdo ou indicar que tal conteúdo deve ser acessado do servidor. Sempre que um par armazena um conteúdo acessado, ou realiza o descarte de um conteúdo previamente armazenado, é sua responsabilidade informar ao tracker a ocorrência de tal evento. Neste trabalho considerou-se que o processo de busca e requisição por um conteúdo se dá a partir do vídeo, e não de partes dele. Com isso, o estudo relacionado com mecanismos de escalonamento para o acesso desse conteúdo nos pares está além do escopo desse artigo, de forma que a presença ou não do conteúdo na porção P2P foi considerada como a medida de desempenho a ser observada.

Em [Melo et al. 2011], foram avaliados diferentes processos renováveis, definidos a partir da combinação das distribuições Pareto e exponencial, na modelagem da autonomia (*churn*) do pares da porção P2P do sistema. O uso da distribuição Pareto na duração das sessões caracterizam uma participação errática dos pares, enquanto que o uso da distribuições exponencial caracteriza pares como uma participação mais próximo do comportamento médio esperado. No presente artigo, os processos renováveis definido pela distribuição exponencial foram utilizados para caracterizar a autonomia dos pares. Como destacado em [Melo et al. 2011], a qualidade das conclusões sobre a disponibilidade de conteúdo no sistema tem pouca influência do processo renovável definido pela combinação das distribuições exponencial e Pareto, e usado para caracterizar o *churn*.

5. Resultados Numéricos

Nessa seção, apresentam-se os resultados dos experimentos que mediram a melhoria na disponibilidade de conteúdo em sistemas CDN-P2P que possui pares tocadores de playlists em sua parte P2P. Nos experimentos foram usados dados reais coletados do YouTube (Seção 3), e considerou-se a disponibilidade de conteúdo gerada pela presença do vídeo requisitado na parte P2P do sistema. Avaliou-se, através de experimentos que simularam o sistema descrito na Seção 4, o impacto da taxa de chegada de usuários tocadores de playlists em um cenário em que toda a playlist é tocada e em um cenário em que a reprodução da playlist é interrompida.

No sistema estudado existem dois tipos de usuários: os tocadores de playlists e os comuns. Os usuários comuns podem sair ou entrar no sistema em qualquer instante que lhes seja conveniente. Ao entrar no sistema, o par (usuário) inicia uma sessão de conexão (sessão UP) que irá durar até a sua partida. Por outro lado, ao deixar o sistema uma sessão de desconexão (sessão DOWN) será iniciada, encerrando-se quando ocorrer o retorno do par para o sistema. Para caracterizar a dinâmica gerada pela autonomia dos usuários comuns, utilizou-se o modelo proposto em [Yao et al. 2006], que é baseado em processos renováveis independentes, para definir a duração das sessões UP e DOWN.

A autonomia dos tocadores de playlists foi baseada na ideia geral aplicada aos usuários comuns, com as sessões UP e DOWN. Entretanto, para caracterizar a duração das sessões UP considerou-se que o tocador de playlists irá acessar os vídeos listados previamente, sem realizar considerações sobre a ordem de execução ou interrupções momentâneas na execução dos vídeos na playlist. O acesso então ocorrerá de forma integral ou haverá uma interrupção definitiva na reprodução da playlist. Os usuários comuns acessam os vídeos em sessões UP que duram 1/4 de hora, em média, conforme estatística extraída de [YouTube 2012].

Em [Melo et al. 2011], estudou-se o impacto da capacidade de armazenamento e do número de pares na disponibilidade de conteúdo de uma sistema híbrido CDN-P2P. A partir desse estudo definiu-se uma comunidade com 10.000 pares, com 2.000 pares, no máximo, formando o agrupamento de tocadores de playlists e 8.000 pares, no mínimo, formando o agrupamento de usuários comuns. A capacidade de armazenamento (buffers) é reduzida, de apenas 34 vídeos, evitando os ganhos de desempenho que uma capacidade maior de armazenamento propicia, e a duração das sessões DOWN, entre 6 e 24 horas. A Tabela 2 mostra detalhadamente o conjunto de parâmetros do sistema e as valores usados nos estudos.

Tabela 2. Configuração dos agrupamentos P2P para Simulação do Sistema CDN-P2P.

Parâmetros do Sistema	Níveis do Parâmetro	Agrupamento de usuários	
		Comum	Tocador de Playlists
Duração Média da Sessão (hora)	UP	1/4	-
	DOWN	24	6, 12 e 24
Modelo de Churn	UP	exponencial	-
	DOWN	exponencial	exponencial
Tamanho do Buffer (vídeo)	Por Par	34	34

A disponibilidade de conteúdo foi à medida de desempenho usada para avaliar o sistema estudado, tendo-se considerado a taxa definida pelo número de requisições geradas na porção P2P do sistema e o número de requisições que permanecem na malha, i.e. que foram atendidas pelos próprios pares da forma a porção P2P do sistema. Essa taxa, chamada de taxa de acerto (TA), é dada por

$$TA = \frac{RA}{TRT} \quad (1)$$

RA é o número de requisições atendidas pelos pares e TRT é o número total requisições realizada pelos pares da porção P2P.

A política Greedy-Dual Size Adaptive (GDSA) [Filho et al. 2012] foi usada para gerenciar o conteúdo em cada par e o sistema foi simulado durante um ano, com tempo de aquecimento da simulação igual a seis meses, e coleta de dados semanais. Os resultados apresentados são as médias dos valores coletados após o período de aquecimento do sistema, com intervalo de confiança de 95%. Os problemas relacionados a distribuição de carga estão fora do escopo deste trabalho pois requerem a definição de mecanismos de controle de admissão e escalonamento que otimizem a utilização do canal disponível nos usuários.

A Figura 3 mostra a taxa de acerto para pares comuns e tocadores de playlists em função do número de tocadores de playlists no sistema. Nessa figura, a barra "sistema" representa a disponibilidade de conteúdo que foi calculada somando-se a disponibilidade nos dois agrupamentos de pares, tocadores de playlist e pares comuns. A medida que o número de tocadores de playlists aumenta, aumentam também suas contribuições para a disponibilidade de conteúdo. A taxa máxima de acerto foi de 45%, para sessão DOWN de seis horas e 20% de tocadores de playlists no sistema. A contribuição de pares comuns diminuiu de 22% para 1% com o aumento do número de tocadores de playlists.

A rapidez com que os tocadores de playlists tornam-se os principais responsáveis pela disponibilidade de conteúdo está relacionada com a frequência com que esses pares estão disponíveis para atender requisições de outros pares. Por exemplo, para uma sessão DOWN de 24 horas de duração, os tocadores de playlists dominam a taxa de acerto quando eles representam 15% dos pares no sistema, enquanto que, para uma sessão DOWN de seis horas, os tocadores tornam-se os maiores responsáveis pela disponibilidade de conteúdo quando são apenas 5% dos pares. A redução do tempo entre sessões UP implica em aumentar a taxa de chegada de pares tocadores de playlists ao sistema,

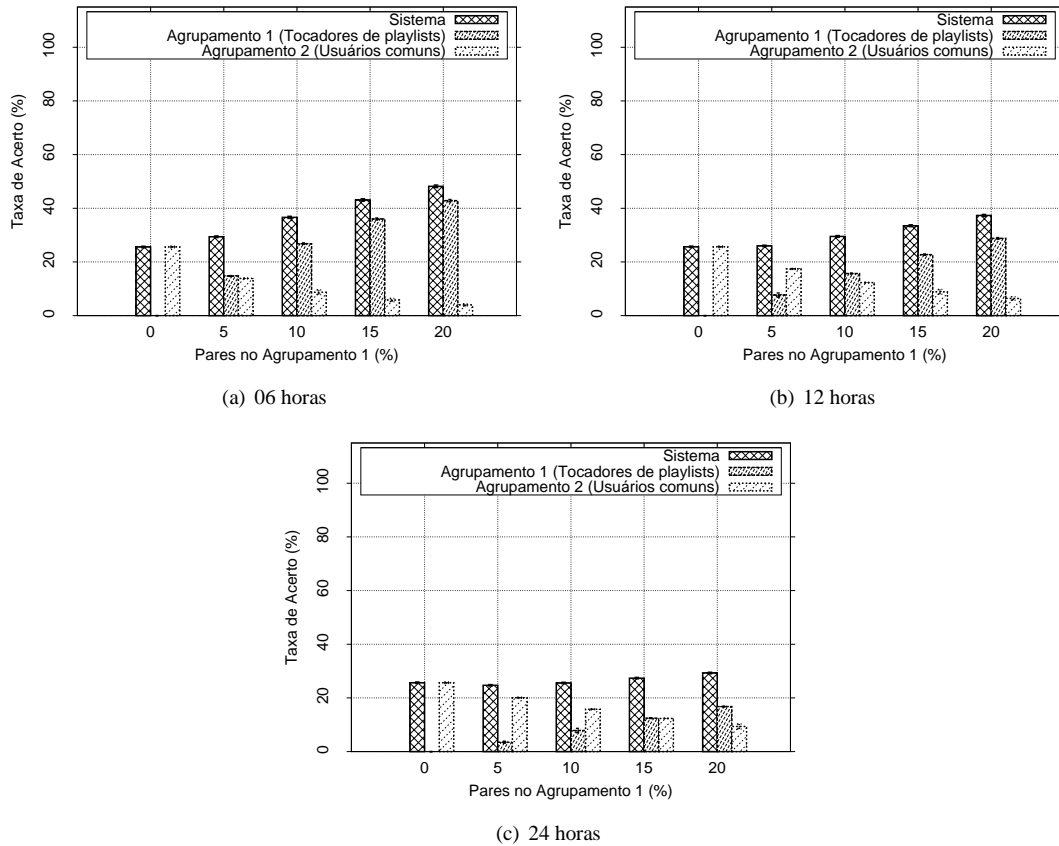


Figura 3. Taxa de Acerto de Pares Comuns e Tocadores de Playlist com três Duração de Sessão DOWN para a Coleção de Vídeos de Música.

tornando-os aptos a compartilharem os vídeos das playlists que já foram acessados. Além disso, os vídeos que estão nas playlists também são populares no contexto da comunidade formada em torno da coleção de vídeos Top-feed o que faz com que o conteúdo nas playlist esteja mais tempo disponível.

A Figura 4 mostra os resultados da avaliação realizada com as outras coleções: Esporte, Educação, Comédia e Filmes&Animações, quando as sessões DOWN duram 24 horas. As conclusões obtidas com a coleção de vídeo de Música podem ser aplicadas para as outras coleções. Ao aumentar o número de tocadores de playlists e reduzir a duração da sessão DOWN, ocorre uma melhoria na disponibilidade de conteúdo, embora haja diferença quantitativa em relação ao resultado obtido com a coleção de Música. Por exemplo, com 10% dos tocadores de playlists a contribuição dos pares comuns é maioria no atendimento das requisições na coleção de Música, porém com aquele mesmo número de tocadores de playlists, os pares comuns não mostram contribuição similar para as coleções de vídeo de Esporte.

No estudo que segue, avalia-se o impacto da saída antecipada de tocadores de playlists, i.e quando ocorre a interrupção da execução da playlist, na disponibilidade de conteúdo. Para tal, considerou-se uma configuração do sistema em que as sessões DOWN duram Seis e 24 horas, 20% dos pares na comunidade são tocadores de playlists, e as saídas prematuras acontecem com taxa $\beta = 0.25$, $\beta = 0.50$ e $\beta = 0.75$. Em outras palavras, em média um tocador de playlist irá reproduzir a lista de vídeos ocorrendo a

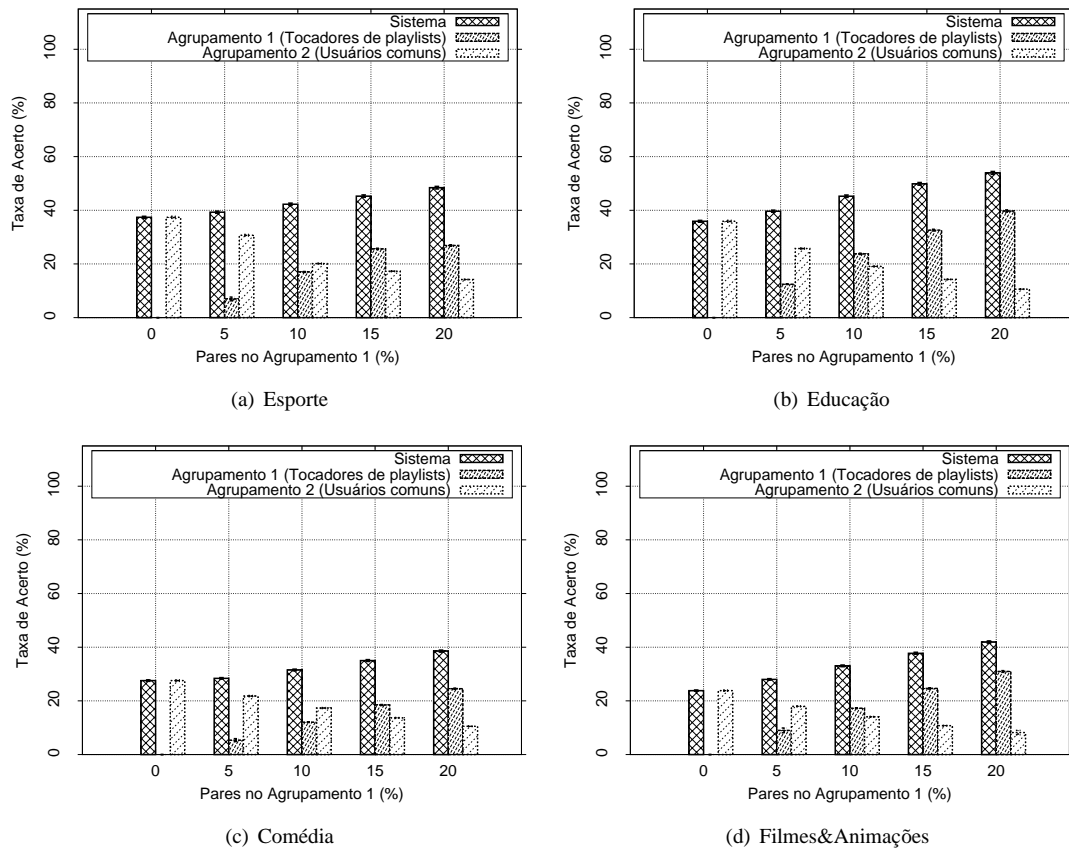


Figura 4. Taxa de Acerto em Usuários Comuns e Tocadores de Playlist com Sessão DOWN de 24 Horas

interrupção quando 25%, 50%, e 75% dos vídeos na lista forem reproduzidos. Usou-se uma distribuição geométrica truncada, configurada com os valores de β , para modelar a interrupção das playlists. Figura 5 mostra o desempenho do sistema na presença de tocadores de playlists impacientes quando a coleção de vídeo Entretenimento foi considerada.

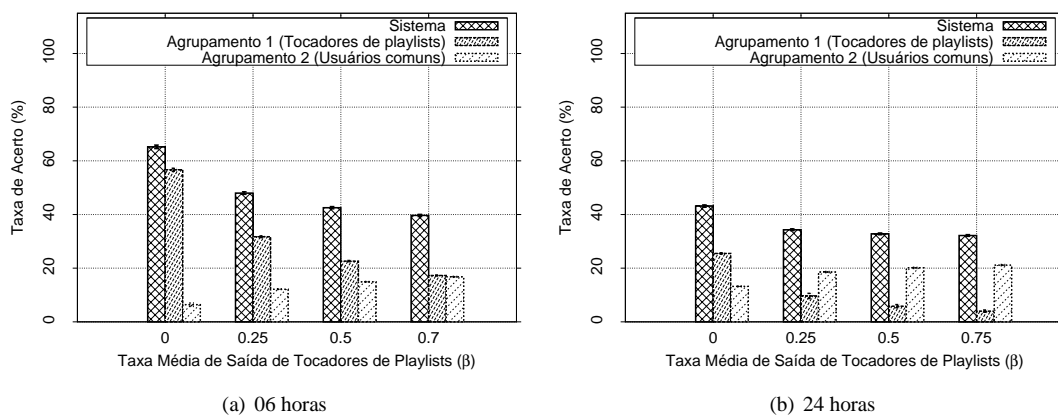


Figura 5. Impacto de Tocadores de Playlists Impacientes na Disponibilidade de Conteúdo (Coleção Entretenimento).

A conclusão geral é que à medida que as saídas acontecem mais cedo, a disponibilidade de conteúdo é afetada mais fortemente. Essa redução é mais destacada para o

período de desconexão de Seis horas, quando a disponibilidade de conteúdo diminui em 1/3. A saída prematura dos tocadores de playlists do sistema gera a indisponibilidade do conteúdo armazenado somente por eles, tendo-se observado que uma parte significativa dos vídeos mais populares da coleção estão no início das playlists. De fato, para a coleção Entretenimento verificou-se que 8% dos vídeos que respondem por 80% das visualizações na coleção Top-Feed Entretenimento encontram-se posicionados no primeiro quarto da playlist. Assim, quando o tocador de playlist interrompe a reprodução logo após reproduzir o último vídeo do primeiro quarto, todo o atendimento que poderia ser realizado a partir do armazenamento desse conteúdo popular é inviabilizado.

Os resultados apresentados até então consideraram a disponibilidade de conteúdo proporcionada pelos tocadores de playlists no sistema como um todo, o que inclui o auto-atendimento e o atendimento de requisições geradas por usuários de ambos os agrupamentos, à medida que o número de tocadores de playlists torna-se maior. Embora sejam relevantes para avaliar a sua importância para o sistema, essas medidas não revelam a intensidade dos serviços prestados pelos tocadores de playlists aos pares comuns. Para tal, mediu-se a quantidade de requisições geradas pelos usuários comuns e que é exclusivamente servida por tocadores de playlists, Figura 6, para as coleções de Entretenimento e Música.

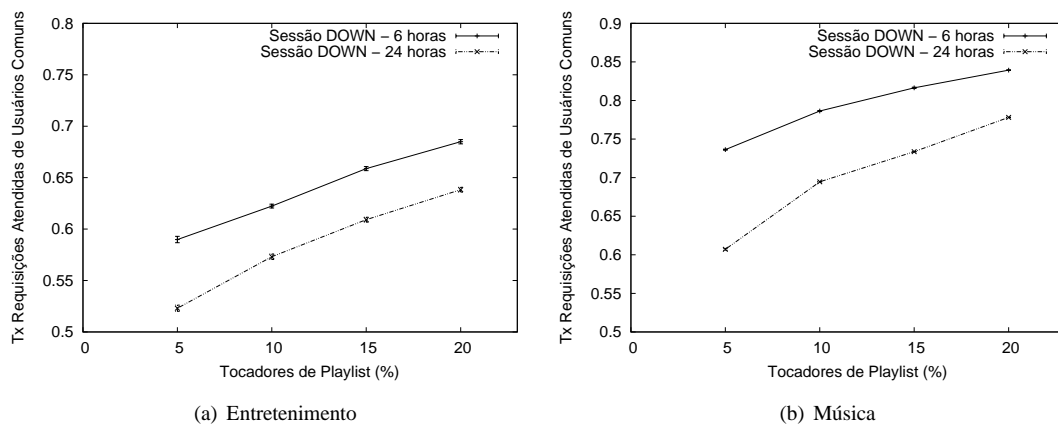


Figura 6. Impacto de Tocadores de Playlists na Disponibilidade de Conteúdo requisitado por Usuários Comuns.

Verifica-se que uma parte significativa das requisições do agrupamento de usuários comuns é atendida pelos tocadores de playlists. Por exemplo, quando os vídeos de Entretenimento são acessados esse atendimento é de 52% (63%) em um sistema que possui 5% (20%) dos seus pares compondo o agrupamento de tocadores de playlists. Esse comportamento é explicado pela intersecção do conjunto dos vídeos mais populares de ambas as coleções playlist e Top-Feed, que concentram a grande maioria dos acessos (80%), conforme análise apresentada na Tabela 1.

Outro fator que influencia o atendimento das requisições dos usuários comuns pelo agrupamento de tocadores de playlist, é o posicionamento dos vídeos mais populares na parte inicial das playlists. Por exemplo, verificou-se que 12% dos vídeos que respondem pela maioria dos acessos à coleção Top-Feed Entretenimento, 80% dos acessos, estão na primeira metade das playlists. Dessa forma, a reprodução da metade final da playlist representa a possibilidade de atendimento das requisições pelos tocadores de playlists,

principalmente aquelas feitas para os vídeos mais populares.

6. Conclusões

A disponibilidade de conteúdo em sistemas CDN-P2P é fortemente afetada pela autodeterminação dos pares sobre a sua participação no sistema. Nesses sistemas, pares que se juntam por longas sessões podem ajudar a melhorar a disponibilidade de conteúdo servindo conteúdo popular. Neste artigo, estudou-se a efetividade de pares tocadores de playlist na melhoria da disponibilidade de conteúdo em sistemas CDN-P2P que distribui vídeos de curta duração similar aos distribuídos pelo YouTube. Estudou-se os padrões de popularidade de vídeos em playlists e vídeos da primeira página do YouTube, e verificou-se que eles compartilham similaridades em seus padrões de popularidade. Além disso, avaliou-se a melhoria na disponibilidade de conteúdo em diferentes cenários de rede, i.e. quando um número grande de pares tocadores de playlists passam longas sessões conectados, a melhoria na disponibilidade de conteúdo foi de 40% para a coleção de Música. Adicionalmente, em cenários que têm todos os pares com sessão DOWN de 24 horas, a melhoria foi de 15% para a coleção de vídeos de Esporte, superou 20% para a coleção de vídeos de Filmes&Animações. Atualmente, os autores estão avaliando o papel dos pares tocadores de playlist na indexação de conteúdo baixado para a rede P2P de sobreposição de forma que pares irão precisar saber apenas como acessar o grupamento de pares que tocam playlist ao invés de saber qual par acessou um conteúdo para recuperá-lo.

Referências

- Crane, R. and Sornette, D. (2008). Robust dynamic classes revealed by measuring the response function of a social system. *The National Academy of Sciences*, 105(41):15649–15653.
- Davidson, J., Liebald, B., Liu, J., Nandy, P., Van Vleet, T., Gargi, U., Gupta, S., He, Y., Lambert, M., Livingston, B., and Sampath, D. (2010). The youtube video recommendation system. In *fourth ACM conference on Recommender systems*, pages 293–296.
- Filho, J. M. L., Melo, C. A. V., and da Fonseca, N. L. S. (2012). Políticas de gerenciamento de conteúdo gerado por usuarios em sistemas CDN-P2P. In *SBRC-WP2P, 2012*.
- Guo, L., Tan, E., Chen, S., Xiao, Z., and Zhang, X. (2008). The stretched exponential distribution of internet media access patterns. In *Twenty-seventh ACM symposium on Principles of distributed computing*, pages 283–294.
- Joung, Y.-J. and Huang, H.-L. (2012). A detailed examination of the overlay construction and maintenance mechanism in bittorrent. *Computer Communications*, 35:682–695.
- Lee, S. H., Kim, P., and Jeong, H. (2005). Statistical properties of sampled networks. *Physical Review E*, 73. updated in 2009.
- Melo, C. A. V., Oliveira, J. A., and da Fonseca, N. L. S. (2013). Promotion of content availability by playlist viewers in cdn-p2p systems. In *IEEE ICC 2013*. available at:<http://goo.gl/7EkBM>.
- Melo, C. A. V., Vieira, D., and Liborio, J. M. (2011). Impact of churn on object management policies. In *IEEE LATINCOM 2011*, pages 1–6.

- Menasche, D. S., Rocha, A. A., Li, B., Towsley, D., and Venkataramani, A. (2009). Content availability and bundling in swarming systems. *CoNEXT '09*, pages 121–132.
- Wang, F., Liu, J., and Xiong, Y. (2008). Stable peers: Existence, importance, and application in peer-to-peer live video streaming. In *IEEE INFOCOM 2008*, pages 1364–1372.
- Yao, Z., Leonard, D., Wang, X., and Loguinov, D. (2006). Modeling heterogeneous user churn and local resilience of unstructured p2p networks. In *the 2006 IEEE International Conference on Network Protocols*, pages 32–41.
- YouTube (2012). YouTube web page. último acesso, Agosto 2012.