

Modelo de Comportamento dos Participantes do SopCast.

Alex Borges Vieira, Pedro Gomes, Marcus Rocha, Jussara Almeida, Sergio Campos
{borges, pcgomes, mvrocha, jussara, scampos}@dcc.ufmg.br
Computer Science Department - Federal University of Minas Gerais (UFMG)

ABSTRACT

Apesar da crescente popularidade das aplicações de transmissão de mídia contínua ao vivo em P2P, a compreensão atual de aspectos importantes do comportamento dos participantes, como o intervalo de chegada entre eles e tempo de permanência na rede, ainda é limitada. Este trabalho pretende preencher esta lacuna apresentando uma caracterização do comportamento dos participantes no SopCast, uma das aplicações mais populares de transmissão ao vivo sobre P2P. A análise apresentada inclui propriedades relativas à sessão, tais como intervalos entre chegada de novos participantes, tempo de permanência, número de sessões q e tempo de inatividade entre sessões. Também são apresentadas propriedades das parcerias entre os participantes, como número de parceiros e suas durações. Além disso, o estudo apresentado compreende diferentes canais do SopCast, destacando diferenças nos padrões de comportamento dependendo do conteúdo. As caracterizações apresentadas definem um modelo dos participantes, que pode ser usado como base para criação de novos protocolos e geração de cargas sintéticas realistas para transmissão ao vivo em P2P.

Categories and Subject Descriptors

H.1.0 [Models and Principles]: General

Keywords

P2P, Live Streaming, Measurement, Performance

1. INTRODUÇÃO

As aplicações de transmissão de mídias estão se tornando mais populares a cada dia na Internet. Atualmente, existem vários recursos que permitem a criação, edição e armazenamento de conteúdo multimídia na Internet. Essas ferramentas e sítios facilitam, por exemplo, a criação de vídeos domésticos e a sua distribuição pela rede. A existência desse ambiente fervoroso de novas ferramentas e conteúdo causam um efeito "bola-de-neve": mais conteúdo gera mais audiência,

que em consequência gera mais conteúdo [6]. Nesse cenário, o *YouTube* é a sistema mais popular para publicação e transmissão de vídeos armazenada. Alguns relatórios mostram que aproximadamente 20% de todos os usuários da Internet visitam o *YouTube* diariamente ¹.

Assim como os vídeos armazenados, as transmissões ao vivo também representam um importante segmento de aplicação e vem atraindo um crescente número de usuários. Grandes emissoras de televisão como a NBC ² e ESPN ³ por exemplo, transmitem ao vivo sua grade de programação diária pela Internet [17]. Além disso, mais e mais usuários domésticos estão transmitindo seu próprio conteúdo ao vivo para o mundo inteiro a partir de seus computadores pessoais.

Normalmente, essas aplicações de transmissão ao vivo são baseadas no modelo cliente-servidor. Apesar da popularidade deste modelo, ele apresenta pontos fracos como uma baixa escalabilidade do número de clientes e uma necessidade de alta concentração de recursos computacionais em um único ponto. Além do alto custo para manter uma arquitetura desse tipo, a concentração em um único ponto pode deixar o sistema de transmissão vulnerável a ataques.

Em resposta a estes pontos fracos e às limitações do modelo cliente-servidor, o modelo Par-a-Par (P2P) é utilizado cada vez mais para distribuir conteúdo na Internet. A crescente popularidade dos sistemas P2P para transmissão de conteúdo ao vivo e suas vantagens têm motivado o desenvolvimento de protocolos e estruturas P2P [14].

Assim, existe uma grande quantidade de trabalhos que focam no desenvolvimento de protocolos e na organização das redes P2P para difusão de mídia contínua ao vivo. Entretanto, apenas alguns estudos abordam a caracterização dos participantes da transmissão ao vivo [8]. Apesar do comportamento dos clientes poderem afetar significativamente a transmissão de mídia [12, 7], a sua compreensão ainda é superficial. Dessa forma, a caracterização de sistemas P2P de transmissão ao vivo é um tópico de interesse crescente [8].

Trabalhos anteriores sobre a caracterização de sistemas de transmissão ao vivo em P2P [10, 3, 16, 8, 9, 5, 4, 15, 18] focam seus estudos somente nas propriedades topológicas da rede P2P. Apenas alguns aspectos comportamentais dos participantes são caracterizados, como a taxa entre chegada deles e o tempo de permanência no sistema. Logo, eles não provem uma análise aprofundada do comportamento dos clientes. Além disso, estes estudos assumem um comportamento homogêneo e independente dos participantes e ig-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 200X ACM X-XXXXX-XX-X/XX/XX ...\$5.00.

¹ www.alexa.com

² www.nbc.com

³ www.espn.com

noram diferenças nos cenários de domínios distintos, como o canal transmitindo um evento de grande interesse ou um programa diário de noticiário. Assim não fica claro se os resultados se confirmam para outros domínios. Desta forma, é necessária uma caracterização mais completa do comportamento dos participantes de uma transmissão ao vivo em P2P, já que a dinâmica dos deles é apontada como um fator de forte impacto no desempenho de sistemas P2P [11].

Neste trabalho é apresentado uma análise do comportamento dos participantes do SopCast, um importante sistema de transmissão ao vivo em P2p. Nessa análise foca-se nas características do comportamento dinâmico dos usuários nessa rede P2P. O objetivo principal é prover dados para geração de cargas sintéticas mais realistas dos comportamentos dos clientes, que podem ser utilizados tanto na avaliação dos protocolos para transmissão de conteúdo contínuo ao vivo sobre P2P quanto no desenvolvimento de novas aplicações P2P para transmissões ao vivo.

A presente caracterização baseou-se em um conjunto de *logs* de tráfego sobre três canais diferentes do SopCast. Foram coletados dados em diferente datas entre novembro de 2008 e fevereiro de 2009 utilizando a rede de experimentos Planet-Lab [1]. Foram caracterizadas as cargas de trabalho desses três canais, que podem ser classificados em dois domínios diferentes: o primeiro domínio é o de um típico canal de transmissão ao vivo, que consiste de uma emissora chinesa de televisão apresentando notícias em alta qualidade de vídeo a uma audiência majoritariamente chinesa. O segundo domínio é representado por dois canais que distribuem conteúdo de entretenimento durante um grande evento de interesse, principalmente para usuários brasileiros.

Conforme em [2], foi utilizada uma abordagem hierárquica para descrever a carga de trabalho em dois níveis: um nível de sessão do sistema e outro de parcerias interativas entre os participantes. Assim, caracteriza-se uma lista dos parâmetros em cada nível dos cenários, incluindo frequências de tempo de acesso, intervalos entre a chegada de novos participantes, número de parceiros e as interações dinâmicas entre eles, também conhecidas como *churn* (entrada e saída da rede, abandono de parcerias, etc.). Este conjunto de parâmetros pode ser utilizado para construir um modelo dos usuários do sistema P2P. Uma atenção especial foi dada às variações temporais da transmissão e, as análises apresentadas são de dados selecionados de dias de carga alta do sistema (mais usuários), onde as distribuições dos parâmetros são esperadas a se manterem relativamente estáveis.

Os principais pontos observados nesse trabalho foram:

- Participantes em dias de evento exibem um comportamento interativo distinto dos clientes em dias normais.
- O tempo entre chegadas em uma sessão é aproximado por uma distribuição exponencial em todos os cenários.
- O número de parceiros de um participante é descrito por uma dist. normal com média próxima a 90 parceiros.
- Um participante tem poucas sessões. Ou seja, quando um participante deixa o sistema, ele normalmente não retorna.
- O tempo de permanência no sistema pode ser descrito por uma distribuição Gamma, porém, a distribuição Pareto descreve melhor um canal em dia de evento.
- O tempo de duração das parcerias pode ser bem representado por uma distribuição exponencial em todos os cenários.

A estrutura deste trabalho está organizada da seguinte forma. Na sessão 2 são discutidos os trabalhos relacionados. Na sessão 3 são descritos os cenários abordados e o modelo

hierárquico usado para caracterizá-los. As características gerais de tráfego nos canais são analisadas na sessão 4, e uma caracterização do comportamento dos clientes é apresentada na sessão 5. Nesta sessão também são discutidas algumas consequências desses comportamentos em sistemas P2P. Na sessão 7 é concluído o trabalho, mostrando as principais contribuições e possíveis trabalhos futuros.

2. TRABALHOS RELACIONADOS

Na literatura há vários trabalhos sobre a caracterização de cargas em transmissões de mídia contínua ao vivo [3, 16, 8, 9, 5, 4, 15]. Entretanto poucos desses trabalhos focam no comportamento dos participantes dos sistemas.

Algumas das aplicações mais populares de transmissão ao vivo como o SopCast e PPlive foram alvo destes estudos, que focaram principalmente as propriedades topológicas das redes P2P e exploraram métricas da teoria dos grafos. Modelam por exemplo, a distribuição de graus dos clientes e o coeficiente de agrupamento [16, 15]. Os autores em [3] também investigam a qualidade do conteúdo ao vivo recebido.

O padrão de tráfego, vizinhança dos clientes e tempo de vida dos participantes (definido neste trabalho como tempo de residência ou *ON time*) são explorados em [10, 8, 9, 15]. Os autores em [10, 9] analisaram o comportamento de quatro sistemas P2P de transmissão ao vivo durante um evento. Eles focaram em componentes da rede como taxa de envio, taxa de recebimento e distribuição de tráfego TCP ou UDP. No presente trabalho, é observado um comportamento da distribuição de pacotes similar a [10, 9]. O tempo de vida dos clientes em [8] é bem descrito por uma distribuição de Weibull; já em [15], o tempo de vida é modelado por uma distribuição geométrica. Os resultados aqui apresentados mostram que uma distribuição Gamma representa o tempo de vida nos canais típicos. Também foi percebido um comportamento diário de tráfego similar ao observado em [15].

Uma caracterização do padrão de chegada e saída dos clientes, a distribuição das popularidades dos canais e a localização geográfica dos participantes é apresentada em [5, 4]. Assim como ocorre nas análises apresentadas nesse trabalho, os autores observaram que os participantes têm um tempo de vida maior em canais populares. Em todos estes trabalhos, os autores não diferenciam sessão de parceria e sessão do participante.

Apesar de esses estudos anteriores cobrirem diferentes características, a maior parte deles considera apenas poucos parâmetros específicos dos cenários, focando no comportamento do sistema e nas propriedades topológicas. O presente trabalho foca em aspectos-chave para a geração de cenários sintéticos realistas, como tempo de permanência do cliente e seu tempo de parcerias, porque estes parâmetros podem caracterizar um cliente em qualquer sistema P2P de transmissão ao vivo. Além disso, dada a diversidade de canais transmitidos ao vivo, não fica claro se os resultados anteriores se confirmam em diferentes domínios, como em um canal de entretenimento ou um de notícias.

Até onde se conhece, a única tentativa de comparar características de diferentes domínios de transmissão ao vivo é a análise apresentada em [15]. Entretanto, esse trabalho fornece uma caracterização muito limitada do comportamento dos clientes, e concentra seus esforços nas propriedades topológicas do sistema.

Em comparação a trabalhos anteriores, o estudo apresentado: (1) fornece uma caracterização mais aprofundada do

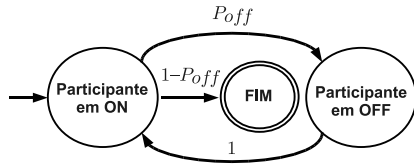


Figure 1: Modelo do Participante.

comportamento dos clientes em sistemas de transmissão ao vivo sobre P2P, (2) analisa um conjunto muito mais diversificado de canais, incluindo canais de entretenimento, notícias e uma cobertura de evento, e (3) propõe um modelo mais preciso do comportamento dinâmico dos clientes.

3. MODELO DE P2P AO VIVO

Um sistema de envio de mídia contínua ao vivo em P2P é composto por participantes (nodos/peers/usuários) que colaboram entre si para disseminar a mídia que está sendo criada por um participante especial do sistema, denominado servidor. Os participantes são organizados em uma rede virtual, sobreposta à rede física (geralmente uma rede IP). Há duas grandes vertentes de redes sobrepostas para esse fim: uma rede estruturada baseada em árvores e uma rede sem estrutura, baseado em malhas computacionais.

O servidor do sistema S gera o conteúdo ao vivo que irá ser transmitido. Ele divide a mídia em pedaços ou fatias, conhecidos como *chunks*, que irão ser transmitidos para os demais participantes do sistema P2P. Um participante guarda os pedaços de mídia em um armazenamento temporário para poder cooperar com outros participantes. Assim, um sistema P2P de transmissão ao vivo também aproveita recursos de seu conjunto de participantes P , onde $P = p_1, p_2, \dots, p_n$.

Cada participante p_i pode se unir ou abandonar o sistema a qualquer momento. Eles podem ir para um estado inoperante, o que significa que eles não interagem com nenhum outro participante. Quando os participantes saem do estado inoperante, eles retornam ao sistema e realizam novas interações. Dessa forma, a dinâmica do sistema continua. Se um participante não retorna ao sistema, interpreta-se que esse participante abandonou definitivamente a aplicação.

A Figura 1 mostra uma máquina de estado simples que captura a dinâmica dos participantes da rede P2P. Quando um participante se junta ao sistema, ele inicia suas interações com outros participantes (estado ON). Quando o participante deixa o estado ON, ele pode ficar inoperante por um momento com uma probabilidade P_{off} , ou, pode abandonar definitivamente o sistema com uma probabilidade $1 - P_{off}$.

Atualmente, as aplicações de envio ao vivo em P2P mais populares, como Sopcast (www.sopcast.com), PPlive (pplive.com) e GridMedia (www.gridmedia.com) usam uma arquitetura baseada em redes de malha, e são orientados à busca por dados (*mesh-based data-driven*). Essa rede sobreposta tende a ser resistente aos efeitos de alta dinâmica dos participantes, promovendo uma visualização da mídia sem interrupções.

Para se juntar ao sistema P2P, um cliente p_i contata um mecanismo centralizado chamado bootstrap B . Essa entidade geralmente é uma estrutura à parte da rede sobreposta. O servidor B envia ao novo participante um subconjunto dos participantes do sistema. Esse subconjunto é uma lista de potenciais parceiros para p_i , definido como LPC_i ($LPC_i \subseteq P$ e $LPC_i \neq \emptyset$).

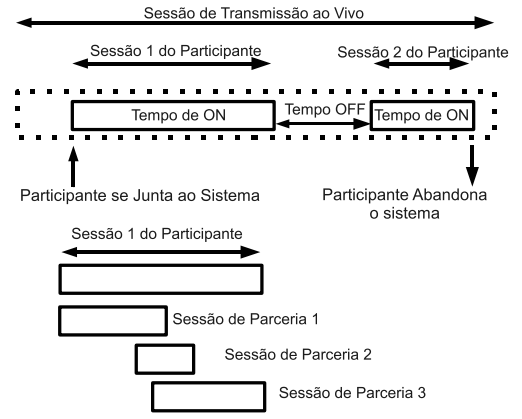


Figure 2: Modelo Hierárquico das Sessões.

O participante p_i escolhe um determinado número de participantes do conjunto LPC_i para fazer parte do conjunto de seus parceiros, denominado LP_i ($LP_i \subseteq LPC_i$). Os conjuntos LPC_i e LP_i são dinamicamente atualizados. Quando p_i detecta que um parceiro $p_j \in LP_i$ está inoperante, p_i remove p_j de seus conjuntos LPC_i e LP_i . Ele seleciona outro participante de LPC_i para realizar uma nova parceria.

Os participantes do sistema trocam dados somente com seus parceiros. Então, um participante p_i checa periodicamente quais *chunks* ele necessita e identifica quais parceiros podem servi-lo. Ele faz uma requisição dos *chunks* necessário para seus parceiros e também serve dados.

A Figura 2 ilustra o modelo proposto. Durante uma transmissão ao vivo, os participantes assistem vídeo e interagem com seus parceiros durante um período, denominado *sessão do participante*. A sessão do participante inicia quando o participante se junta ao sistema P2P e faz sua primeira requisição pela mídia ao vivo. O tempo total da sessão é conhecido como *tempo de residência* ou simplesmente tempo de ON. Se um participante retorna ao sistema após o término de uma sessão, ele inicia uma nova sessão de participante.

O participante pode ficar inoperante durante algum tempo. Esse tempo inoperante pode ser causado por uma quebra temporária em sua conexão com a rede ou congestionamento nos seus recursos. Esse tempo inoperante é denominado de tempo OFF entre sessões. O tempo OFF não necessariamente interrompe a visualização da mídia, uma vez que o cliente tem um armazenamento temporário da mídia (buffer). Por exemplo, durante a sessão de participante 1, na Figura 2, ocorreram dois tempos de inoperância. Se o participante permanecer inoperante por um tempo maior que um limiar definido no modelo, considera-se que ele abandonou a rede P2P e sua sessão foi finalizada. O valor que define o limiar do valor OFF será discutido adiante.

Dentro de uma sessão, um participante estabelece uma ou mais parcerias. Uma sessão de parceria ps_{i-j} é definida como o tempo em que p_i interage com seu parceiro p_j . Ou seja, uma sessão de parceria inicia quando dois participantes trocam dados e dura até a última interação entre eles. A Figura 2 ilustra o modelo de parcerias. São mostradas as 3 parcerias existentes na sessão de participante 1, todas elas com características distintas (tempo de início, duração, etc).

4. METODOLOGIA

Os dados utilizados nesse trabalho foram coletados do

SopCast com o tcpdump(www.tcpdump.org), que monitora a rede ao nível de pacotes. O SopCast é uma aplicação popular para transmissão ao vivo em P2P. Ele tem grande variedade de canais e suporte para vários sistemas operacionais.

4.1 Canais Analisados

Foram coletadas informações de 3 canais de transmissão do SopCast. O primeiro foi o canal de notícias da CCTV, que é um canal chinês. Esse canal é um canal popular e transmite o vídeo a uma taxa alta (600kbps). Foram medidos também dois canais brasileiros em dias distintos. Esses canais transmitiam eventos de grande interesse local (partidas de futebol). Durante esses dois eventos, esses canais transmitiram o conteúdo a uma baixa taxa (250kbps).

A Tabela 1 resume o número total de usuários, número total de sessões e o número de pontos de coleta utilizados para se obter o dado analisado.

Table 1: Características dos Canais

Canal	Taxa	# Clientes Médio	No. de Sessões	No. de Coletores	Data
CCTV	600kbps	700	4385	300	17/12/08
Jogo 1	250kbps	3,300	12233	360	30/11/08
Jogo 2	250kbps	2,800	6362	360	07/12/08

4.2 Metodologia de Coleta

Para coletar os dados do SopCast, utilizou-se o Planet-Lab [1]. Foram utilizados 421 computadores desse sistema espalhados pelo mundo. Os seguintes eventos, canais e períodos foram analisados:

Jogos de Futebol para público brasileiro: os dois eventos analisados ocorreram em 30/11/2008 e 07/12/2008. Os experimentos se iniciaram junto com a transmissão da partida de futebol e finalizaram 100 minutos depois.

Transmissão de noticiário da CCTV: Foram realizadas coletas semanais durante o período de 16/12/08 a 19/01/09; as 20:00 horas local China. Os experimentos têm duração de 100 minutos.

Como observado em [5], a coleta do comportamento dos participantes de sistemas P2P é um desafio uma vez que o acesso aos servidores de rastreamento do SopCast não está disponível. De maneira a tratar esse desafio, foi utilizado uma metodologia de coleta de informações semelhante a apresentada em [15]. Selecionou-se um conjunto de computadores do PlanetLab para se juntar ao canal de transmissão do SopCast e gravar toda sua comunicação de rede. Após a operação de coleta, reconstrói-se uma visão da rede do SopCast durante o evento capturado.

A metodologia de coleta realiza as seguintes operações:

- Uma máquina de controle dispara o processo de coleta no conjunto de máquinas do PlaneLab;
- As máquinas sincronizam o relógio local;
- As máquinas começam a captura do seu tráfego de rede nas portas específicas utilizada pelo SopCast;
- As máquinas se juntam ao canal do SopCast;
- Ao final, os dados da coleta são recuperados para análise.

Foram utilizados no mínimo 300 máquinas de coleta em cada experimento para tentar garantir a maior cobertura possível da rede analisada. Além disso, a banda de subida e descida dessas máquinas não foi limitada. Outros trabalhos com metodologia semelhantes adotam um número menor de máquinas de coleta. Por exemplo, [15] usaram 10 máquinas em seus experimentos.

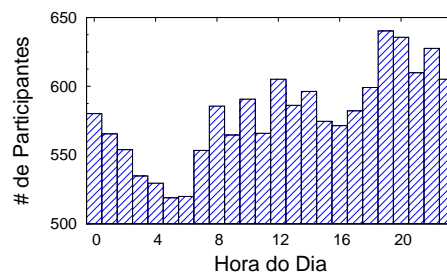


Figure 3: Utilização do Canal Durante um Dia.

5. ACESSOS DOS PARTICIPANTES

Essa seção discute o acesso dos canais do SopCast em diferentes períodos. A pretensão é discutir a atividade do usuário em um canal durante uma semana e durante um dia. Assim, as demais coletas para realizar a caracterização dos participantes do sistema podem ser mais bem direcionadas aos períodos de grande carga e estabilidade.

A figura 3 mostra a variação do número de participantes da CCTV durante um dia. Essa figura representa o período analisado do dia 16/12/2008. Durante esse dia, o número de participantes começa a crescer a partir das 7:00. O período entre 8:00 e 16:00 apresenta um número médio de usuários com pouca variação (+/- 25 usuários de variação). Esse canal alcança o seu maior número de participantes por volta das 20:00, com um período de estabilidade de cerca de 3 horas. Esse período de maior atividade coincide com o período que as pessoas estão chegando em casa após o dia de atividades.

Durante uma semana, o número de participantes em um canal também varia. Como observado em [16] e no atual trabalho, os usuários acessam mais a aplicação aos finais de semana. Isso ocorre porque durante os finais de semana as pessoas estão em casa e procuram por algum entretenimento.

6. CARACTERÍSTICAS DOS PARTICIPANTES

Esta seção analisa cada componente do modelo descrito na seção 6.1, sendo eles, intervalo entre chegadas, número de sessões, tempo de ON e OFF e número de parcerias.

6.1 Tempo Entre Sessões de Participantes

As sessões de participantes não são explicitamente informadas pelo SopCast. Assim, definiu-se que uma sessão de participante termina quando este fica inativo por um tempo maior que um valor limite. Para escolher o valor limite do tempo de inatividade (OFF), foram analisados os tempos entre requisições de um mesmo usuários.

Foram encontrados dois perfis distintos. O primeiro perfil refere-se às requisições realizadas em um período muito curto de tempo, onde mais de 70% das requisições ocorrem em um período menor que 3s. O segundo perfil, refere-se às requisições de um usuário com tempos maiores que 3s. Essas requisições geralmente são mensagens do protocolo.

A figura 4 mostra o tempo entre requisições de um mesmo usuário no segundo perfil. Por essa figura, observa-se um ponto de quebra após os 180 segundos. Por esse motivo, o limiar de OFF escolhido é 180 segundos

6.2 Processo de Chegada de Sessões

Foi analisado o tempo entre chegadas de sessões de partic-

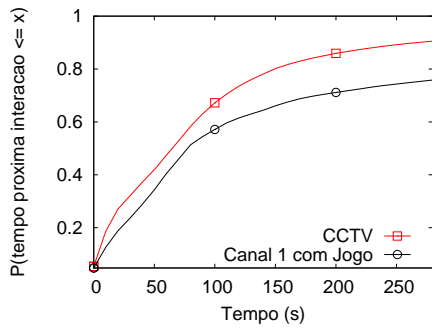


Figure 4: Distribuição do Tempo Entre Requisições.

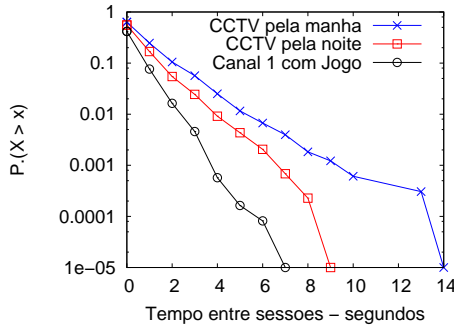


Figure 5: Tempo Entre as Chegadas de Sessões.

ipantes durante períodos de tempo que o sistema se encontrava estável. A Figura 5 resume o intervalo entre sessões para 3 comportamentos distintos.

Para achar a distribuição que melhor modela o tempo de chegada entre sessões, foram comparadas as diferenças quadráticas de um conjunto de distribuições com os dados analisados. Foram realizadas inspeções visuais para comparar os resultados de modelagem, principalmente no corpo e na cauda da curva. Foram favorecidas as distribuições que descrevem melhor o corpo da curva, uma vez que intervalos curtos entre sessões têm um impacto maior na aplicação P2P em questão (ex.: chegadas em rajadas).

A distribuição Exponencial descreve bem o tempo entre chegadas de sessões para todas as cargas observadas. Os tempos entre chegadas de sessões são exponencialmente distribuídos e são maiores em canais com menos usuários. A Tabela 2 resume as observações realizadas, provendo os parâmetros das distribuições para cada carga.

Table 2: Distribuição dos Tempos de Chegada.

Carga	Melhor Dist.	Média (segundos)	Desvio P. (segundos)
CCTV pela manhã	Exponencial	1.095908	1.304518
CCTV pela noite	Exponencial	0.819613	1.001631
Canal 1 com Jogo	Exponencial	0.514674	0.709162
Canal 2 com Jogo	Exponencial	0.392644	0.584500

Distribuição: Exponencial: $p_X(x) = \lambda e^{-\lambda x}$

6.3 Características das Sessões

O modelo apresentado na seção 6.1 possibilita a representação tanto do comportamento do participante, quanto de suas parcerias. Trabalhos anteriores como [16, 8, 13, 3], apenas caracterizam o comportamento dos usuários, e em um único ponto de coleta. Ou seja, eles não capturam muitos aspectos relevantes da caracterização de um participante.

Para determinar a possibilidade de um usuário ir a um estado de inatividade ou abandonar o sistema, foram analisadas as distribuições do número de sessões de cada participante do sistema. A Figura 6 mostra a função de distribuição de probabilidade de ocorrer a n -ésima sessão de usuário. Para todos os canais observados, a quantidade de sessões de um usuário durante os experimentos manteve o mesmo comportamento. Por exemplo, mais de 70% dos participantes abandonam o sistema após a primeira sessão.

Por regressão linear, observa-se que o parâmetro que enquadra no modelo da seção 6.1 tem valor de $P_{off}=0.40$. Nessa mesma Figura 6, além dos dados observados nos experimentos, foi traçada a curva que representa o modelo a qual descreve bem os dados observados.

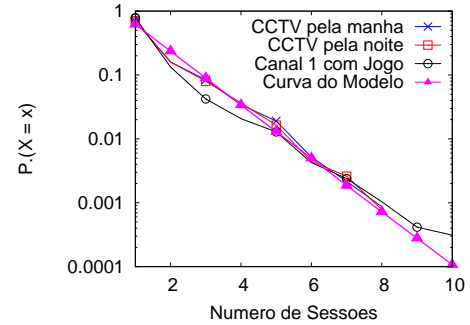


Figure 6: Número de Sessões de Participante.

6.4 Tempos de ON e de OFF

Os tempos de ON e de OFF do modelo proposto foram medidos e analisados. Os tempos de ON correspondem ao tempo de residência de um participante no sistema, ou seja, o tempo que ele busca pelo vídeo ao vivo. Enquanto isso, os tempos de OFF representam o tempo de inatividade um participante, entre duas sessões consecutivas. Nas análises apresentadas, os tempos de ON são medidas normalizadas pelo período de observação do sistema, ou seja, são apresentadas como um percentual do tempo de experimento.

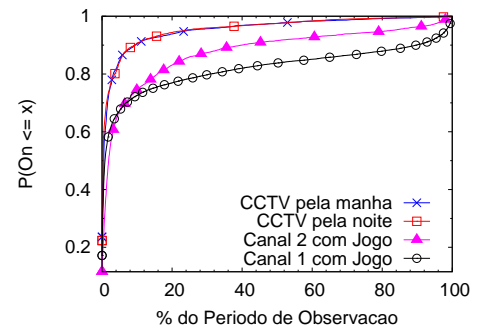


Figure 7: Tempo ON do Participante.

A distribuição que melhor descreve o tempo de ON varia entre os canais observados. Enquanto isso, os tempos de OFF não varia entre eles. Uma distribuição comum foi encontrada para descrever o tempo de OFF entre os canais observados sem uma diferença significativa. A seguir serão descritas as distribuições e as observações com relação ao tempo de ON e de OFF.

Ao contrário de trabalhos como [8] que acreditam que a partida de um participante do sistema e o mecanismo de tro-

Table 3: Distribuição dos Tempos ON.

Carga	Melhor Dist.	Média (segundos)	Desvio Padrão (segundos)	Primeiro Parâmetro	Segundo Parâmetro
CCTV	Gamma Lognormal	0.043334	0.125878	$\alpha = 0.070015$ $m = -4.261192$	$\beta = 0.618924$ $\sigma = 1.498248$
Canal 1 com Jogo	Gamma Pareto	0.179541	0.323704	$\alpha = 0.080705$ $\alpha = 0.275582$	$\beta = 2.224652$ $k = 0.000415$
Canal 1 com Jogo	Gamma Pareto	0.122206	0.235781	$\alpha = 0.115213$ $\alpha = 0.299604$	$\beta = 1.060696$ $k = 0.000187$

(Distribuição: Weibull: $p_X(x) = \alpha\beta x^{\beta-1}e^{-\alpha x^\beta} I_{(0,\infty)}(x)$, Lognormal: $p_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}$, Exponencial: $p_X(x) = \lambda e^{-\lambda x}$ and Pareto: $p_X(x) = \frac{\alpha k^\alpha}{x^{\alpha+1}}$, where $x \geq k$.)

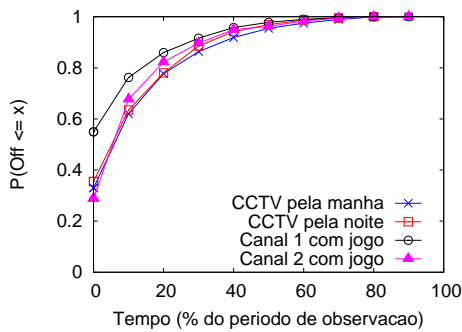
cas de parcerias de uma aplicação impactam no tempo de ON, nesse trabalho, somente o comportamento do participante é considerado como fator impactante dessa medida. Se a aplicação força a troca de parceiros, esse participante continua realizando suas atividades na mesma sessão. Somente a sua sessão de parceria é alterada por esse comportamento.

A Figura 7 mostra o tempo de ON para todas as cargas analisadas. Enquanto os tempos de ON são deslocados para a esquerda nos canais com menos participantes, no dia de evento essas curvas são deslocadas para a direita da figura. Isso indica que o tempo de ON é maior em dias de eventos, como o jogo de futebol. Em dias de evento, nota-se também que um número considerável de participantes permanece no canal por toda a duração do evento.

Os canais em dias típicos, que não estejam transmitindo um evento de interesse, a distribuição Gamma e a Lognormal representam bem os dados medidos. Para o dia de evento, a distribuição que melhor modela o tempo de ON é a Pareto ou a Gamma. A Tabela 3 resume os resultados encontrados para o tempo de ON nas cargas analisadas.

Table 4: Distribuição dos Tempos OFF.

Carga	Melhor Dist.	Média (segundos)	Desvio Padrão (segundos)
CCTV	Exponencial	0.214210	0.170650
CCTV à noite	Exponencial	0.216778	0.168359
Canal 1 com Jogo	Exponencial	0.148428	0.151016
Canal 2 com Jogo	Exponencial	0.190368	0.142840

**Figure 8: Tempo OFF dos Participantes.**

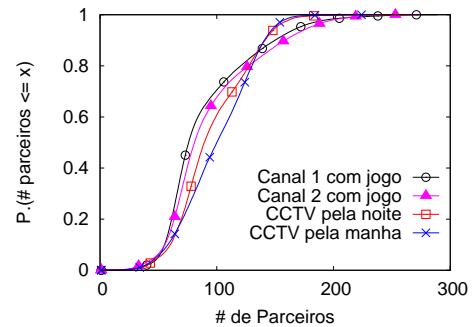
Enquanto em [8] foi encontrado uma distribuição Weibull para representa o tempo de vida dos participantes, a distribuição Gamma representa melhor a cauda dos dados dos canais típicos analisados nesse trabalho. Para o corpo da curva de dados, a distribuição Lognormal apresenta melhores resultados. Para um dia de evento, há um número considerável de usuários que permanecem no sistema durante um período grande, o que gera uma cauda mais longa para os tempos de ON. Nesse caso, a distribuição Gamma

continua representando bem a cauda, com parâmetros diferentes de um canal típico. Porém, o corpo da curva é mais bem representado por uma distribuição Pareto. Note que, nesses casos, o corpo da curva representa a maior parte dos usuários do sistema e por isso, as distribuições que representam essa parte da curva devem ter maior atenção.

A Figura 8 mostra as distribuições de tempos de OFF para as cargas observadas. Note que os comportamentos são semelhantes entre os canais. Os resultados indicam que uma distribuição Exponencial se enquadra bem aos tempos de OFF encontrados nos experimentos. Para todos os canais analisados, essa distribuição representa bem tanto a cauda quanto o corpo da curva de dados. A Tabela 4 resume os parâmetros da distribuição Exponencial para os tempos de OFF dos participantes do sistema.

6.5 Características das Parcerias

A segunda camada do modelo descrito na seção 6.1 refere-se às parcerias realizadas em uma sessão de participante. Cada participante realiza parcerias durante sua sessão e cada parceria dura por um período dentro dessa sessão. Foram

**Figure 9: Número de Parceiros.**

medidos os números de parceiro que cada participante possui. A Figura 9 mostra a distribuição do número de parceiro dos participantes do SopCast. Para todos os canais observados, o número de parceiros não se altera consideravelmente. Esse comportamento pode ser explicado pelo mecanismo intrínseco ao SopCast, que conecta a um certo número de parceiros e sempre tenta manter esse número. Para todos os canais, cerca de 40% dos participantes apresentaram pelo menos 90 parceiros. Isso pode indicar que o SopCast implementa uma política gulosa de parcerias.

Foram comparadas as diferenças quadráticas entre os dados e um conjunto de distribuições. Novamente foram realizadas inspeções visuais para verificar a melhor distribuição que descreve o número de parcerias. O número de parcerias dos participantes pode ser bem representado por uma distribuição Gamma ou Normal para canais típicos como a

Table 5: Distribuição do Número de Parceiros: Resumo.

Carga	Melhor Dist.	Média (parceiros)	Desvio Padrão (parceiros)	Primeiro Parâmetro	Segundo Parâmetro
CCTV	Gamma Normal	100.482999	31.815022	$\alpha = 8.380415$ $\mu = 100.482999$	$\beta = 11.990217$ $\sigma = 31.815022$
CCTV à noite	Gamma Normal	96.552555	31.859601	$\alpha = 8.493371$ $\mu = 96.552555$	$\beta = 11.367990$ $\sigma = 31.859601$
Canal 1 com Jogo	Gamma Lognormal Normal	90.595219	39.298512	$\alpha = 6.195481$ $m = 4.420196$ $\mu = 90.595219$	$\beta = 14.622790$ $\sigma = 0.415224$ $\sigma = 39.298512$
Canal 2 com Jogo	Gamma Lognormal Normal	94.589474	41.038147	$\alpha = 5.725874$ $m = 4.463314$ $\mu = 94.589474$	$\beta = 16.519656$ $\sigma = 0.415289$ $\sigma = 41.038147$

Table 6: Resumo dos Perfis de Participantes

Canal	Tempo entre chegadas	Número de sessões	Tempo de Residência	OFF Time	Tempo de Parceria
CCTV	Exponencial	≤ 2	Gamma Lognormal	Exponencial	Gamma Normal
Canal com Jogo	Exponencial	in 90% of cases	Gamma Pareto	Exponencial	Gamma Normal Lognormal

CCTV. Para um dia de eventos, como o jogo de futebol, além das duas citadas acima, uma distribuição Lognormal também descreve bem os dados observados.

Devido a sua simplicidade, foi escolhida a distribuição Normal para representa o número de parceiros dos canais. Além disso, observa-se que número médio de parceiros é próximo ao número encontrado em [10]. A Tabela 5 resume os resultados encontrados para o número de parceiros.

Além do número de parcerias, foram analisadas as suas durações. Cada parceria dura um tempo menor-igual ao tempo de sessão do participante. Para caracterizar o tempo de uma parceria, a análise feita tenta responder a seguinte questão: “há alguma correlação entre o tempo de parceria e o tempo de ON de um participante?”. Para responder essa questão, foram analisados os tempos de parcerias em relação ao tempo restante da sessão do participante. Ou seja, quanto tempo uma parceria ocupou do tempo restante de vida dos participantes.

A Figura 10 mostra o percentual que uma parceria durou em relação ao tempo de vida restante do participante. Observa-se que, durante o primeiro dia de evento, mais de 80% das parcerias não duram 10% do tempo restante de vida dos parceiros. Em um canal típico, na mesma situação, as parcerias duram cerca de 85% do tempo restante de vida.

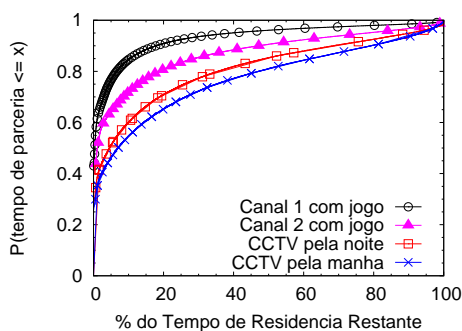


Figure 10: Duração das Parcerias

Há dois argumentos que explicam esse comportamento. Primeiro, em um dia de evento, o tempo de ON dos participantes é maior e assim, a fração da parceria pode ser menor. Segundo, em um dia típico há menos usuários no sistema o que não possibilitam muitas trocas de parcerias, e assim, elas durariam uma fração maior de tempo.

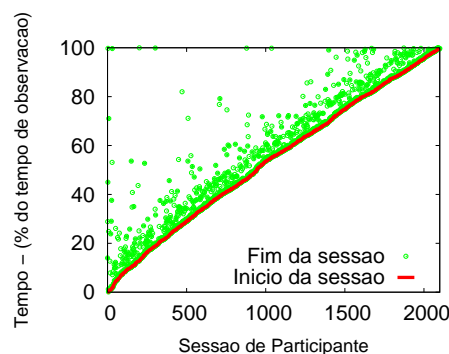


Figure 11: Participantes São Impacientes.

6.6 Resumo das Características

Esta seção apresenta os perfis de comportamento dos participantes do SopCast encontrados nesse trabalho. Os perfis apresentados são caracterizados com foco no tempo de ON, tempo de OFF e nas características das parcerias. Nas cargas de trabalho analisadas, os participantes se encaixam em um dos perfis da Tabela 6. Esse enquadramento depende principalmente do tipo de mídia transmitida pelo canal.

A principal diferença entre os canais analisados é o tempo de ON dos participantes. Os tempos de ON são menores em canais típicos do SopCast, como a CCTV, que em canais transmitindo eventos de grande interesse, como o canal de futebol. Em um dia de jogo de futebol, há um número considerável de participantes que permanecem no sistema por toda a transmissão.

Em um canal típico, uma distribuição Gamma descreve bem a cauda da curva de distribuição dos tempos de ON, uma Lognormal descreve melhor o corpo da curva. Durante a partida de futebol, uma distribuição Pareto foi a melhor descrição para o corpo da curva, capturando o comportamento da maior parte dos usuários do sistema.

Outro ponto importante é que os participantes do SopCast são impacientes. Ao contrário de aplicações de compartilhamento de arquivos, a maior parte dos participantes se junta à transmissão por alguns segundos. A Figura 11 mostra um exemplo de como os participantes assistem uma sessão de vídeo. Durante o período observado, a linha vermelha indica o momento inicial em que um participante se

junta ao sistema. Os pontos verdes mostram quando o participante abandona a sessão. Note que os pontos verdes estão próximos à linha vermelha, o que indica que os participantes entram e abandonam o sistema constantemente.

Esse tipo de comportamento dos participantes pode levar problemas ao sistema, se não tratado. Por exemplo, quando um participante se junta ao SopCast, ele se inscreve no mecanismo de *bootstrap*. Esse mecanismo guarda as informações do novo participante (como endereço) para disponibilizar aos demais participantes quando necessário (ex.: busca por novos parceiros). Porém, os mecanismos de *bootstrap* do SopCast mantêm informações sobre os participantes mesmo quando eles saem. O tempo de reação desse mecanismo é grande e assim, ele continua informando o endereço do participante que já abandonou o SopCast.

Se um grande número de participantes entra e abandona o sistema constantemente, os novos participantes receberão uma série de endereços inúteis. Isso os fará gastar tempo e recursos e parcerias que não serão efetivamente concretizadas.

7. CONCLUSÕES E TRABALHOS FUTUROS

Esse artigo provê uma caracterização do comportamento dos participantes de um sistema de transmissão ao vivo em redes P2P. Foram analisadas uma série de cargas de trabalho que se encaixam em dois domínios: um canal típico e um canal populoso por causa de um grande evento.

Em todas as cargas analisadas, os participantes da rede apresentaram um comportamento altamente dinâmico, com muitas entradas e saídas do sistema. As parcerias acompanham o comportamento geral dos participantes, sendo elas de curta duração. Em canais com um número maior de participantes, as parcerias tendem a durar menos, pois os participantes têm mais oportunidades de trocar de parceiros. Apesar do comportamento da parceria, nos canais transmitindo eventos de grande interesse foram observados tempos de residências maiores. De fato, em um dia de evento, há uma porção considerável de participantes que acompanham todo o evento.

Além desses fatos, verificou-se que o tempo de parceria pode ser expressa em termos do tempo remanescente de vida de um participante. Esse resultado facilita a geração de cargas realistas, uma vez que para todas as cargas de trabalho, a essa correlação pode ser expressa pela mesma distribuição de probabilidade. Mais ainda, apenas pela observação de algumas parcerias pode-se inferir o comportamento de um sistema completo que se encaixe nos padrões de carga de trabalho relatado nesse trabalho.

Possíveis trabalhos futuro incluem a observação e caracterização de outros sistemas P2P para transmissão ao vivo para verificar quais características são dependentes da aplicação; uma maior análise entre as correlações entre as diferentes cargas de trabalho; e propostas de novos protocolos tirando proveito do comportamento esperado dos participantes do sistema.

8. REFERENCES

- [1] B. Chun, D. Culler, T. Roscoe, A. Bavier, L. Peterson, M. Wawrzoniak, and M. Bowman. PlanetLab: An Overlay Testbed for Broad-Coverage Services. *ACM SIGCOMM Computer Communication*, 33(3), 2003.
- [2] C. P. Costa, I. S. Cunha, A. Borges, C. V. Ramos, M. M. Rocha, J. M. Almeida, and B. Ribeiro-Neto. Analyzing client interactivity in streaming media. In *WWW '04: Proceedings of the 13th international conference on WWW*, NY, USA, 2004. ACM.
- [3] B. Fallica, Y. Lu, F. Kuipers, R. Kooij, and P. V. Mieghem. On the quality of experience of sopcast. *Next Generation Mobile Applications, Services and Technologies, International Conference on*, 2008.
- [4] X. Hei, C. Liang, J. Liang, Y. Liu, and K. W. Ross. Insights into p2p: A measurement study of a large-scale p2p iptv system. In *In Proc. of IPTV Workshop, International WWW Conference*, 2006.
- [5] X. Hei, C. Liang, J. Liang, Y. Liu, and K. W. Ross. A measurement study of a large-scale p2p iptv system. *IEEE Transactions on Multimedia*, 2007.
- [6] The new york times, february 13, 2009. *The New York Times Newspaper*, 2009.
- [7] M. Rocha, M. Maia, I. Cunha, J. Almeida, and S. Campos. Scalable media streaming to interactive users. In *In Proc. of ACM Multimedia*, 2005.
- [8] T. Silverston and O. Fourmaux. Measuring p2p iptv systems. In *Proceedings of NOSSDAV'07*, June 2007.
- [9] T. Silverston and O. Fourmaux. P2p iptv measurement: a case study of tvants. In *CoNEXT '06: Proceedings of the 2006 ACM CoNEXT conference*, pages 1–2, New York, NY, USA, 2006. ACM.
- [10] T. Silverston, O. Fourmaux, A. Botta, A. Dainotti, A. Pescapé, G. Ventre, and K. Salamatian. Traffic analysis of peer-to-peer iptv communities. *Comput. Netw.*, 53(4):470–484, 2009.
- [11] D. Stutzbach and R. Rejaie. Understanding churn in peer-to-peer networks. In *IMC '06: Proceedings of the 6th ACM SIGCOMM conference on Internet measurement*, New York, NY, USA, 2006. ACM.
- [12] D. Stutzbach, R. Rejaie, and S. Sen. Characterizing unstructured overlay topologies in modern p2p file-sharing systems. *IEEE/ACM Trans. Netw.*, 2008.
- [13] Y. Tang, L. Sun, J.-G. Luo, and Y. Zhong. Characterizing user behavior to improve quality of streaming service over p2p networks. In Y. Zhuang, S. Yang, Y. Rui, and Q. He, editors, *PCM*, volume 4261 of *Lecture Notes in Computer Science*, pages 175–184. Springer, 2006.
- [14] L. Vu, I. Gupta, J. Liang, and K. Nahrstedt. Mapping the p2p network: Studying the impacts of media streaming on p2p overlays. Technical report, University of Illinois Urbana-Champaign, 2006.
- [15] L. H. Vu, I. Gupta, J. Liang, and K. Nahrstedt. Measurement and modeling of a large-scale overlay for multimedia streaming. In *QShine: International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness*, 2007.
- [16] C. Wu, B. Li, and S. Zhao. Exploring large-scale peer-to-peer live streaming topologies. *ACM Trans. Multimedia Comput. Commun. Appl.*, 4(3):1–23, 2008.
- [17] Z. Xiao and F. Ye. New insights on internet streaming and iptv. In *CIVR '08: Proceedings of the 2008 international conference on Content-based image and video retrieval*, NY, USA, 2008. ACM.
- [18] Y. Zhou, D.-M. Chiu, and J. C. S. Lui. A simple model for analyzing p2p streaming protocols. In *ICNP*, pages 226–235. IEEE, 2007.