

Evolução do Comportamento do Usuário em Eventos de Larga Escala na Internet

Breno Machado¹, Alex B. Vieira¹, Ítalo Cunha², Artur Ziviani³

¹Departamento de Ciência da Computação – Universidade Federal de Juiz de Fora.

²Departamento de Ciência da Computação – Universidade Federal de Minas Gerais.

³Laboratório Nacional de Computação Científica (LNCC)

breno.santos@ice.ufjf.br, alex.borges@ufjf.edu.br

cunha@dcc.ufmg.br, ziviani@lncc.br

Abstract. *Video traffic accounts for 75% of Internet traffic today. In this paper, we characterize the evolution of a large video streaming service use in Brazil, comparing streaming logs from FIFA's 2014 and 2018 World Soccer Cups. We report a 166% increase in the average number of unique IP addresses and more than a 300% increase in traffic volume. Arrival rate has also changed and it's more concentrated at the beginning of the match nowadays. We also compare the impact of QoE metrics on user engagement. For instance, we note that engagement has increased for all bitrate rates and it's also higher when there are fewer bitrate adaptations. Furthermore, the high rate of rebufferings substantially decreases engagement, but only up to a certain value.*

Resumo. *Tráfego de vídeo já representa 75% do tráfego na Internet. Neste trabalho, nós caracterizamos a evolução do uso de um grande serviço de vídeo streaming no Brasil comparando registros de acesso da Copa do Mundo da FIFA de 2014 e 2018. Relatamos um crescimento de aproximadamente 176% no público médio e mais de 300% no tráfego gerado. A taxa de chegada também mudou e agora é mais concentrada no início das partidas. Comparamos também o impacto de métricas relacionadas a QoE no engajamento dos usuários. Por exemplo, notamos que o engajamento aumentou para todas as taxas de bitrate e que ele é maior quando ocorre menos adaptações de bitrate. Além disso, taxa alta de interrupções diminui o engajamento consideravelmente, mas somente até certo valor.*

1. Introdução

Distribuição de vídeo é a aplicação com maior demanda de banda na Internet, impulsionada pelo surgimento de serviços populares como Netflix, YouTube e Hulu. Para suportar essa demanda, provedores e redes de distribuição de conteúdo investem significativamente em tecnologias e infraestrutura [Schlinker et al. 2017, Open Connect 2011].

Os serviços de vídeo mais populares estão convergindo para o uso de *HTTP Adaptive Streaming (HAS)* [Nam et al. 2016]. A utilização do HTTP como transporte facilita integração com *software* e infraestrutura existente (navegadores e servidores *web*). A conectividade da rede e os dispositivos dos usuários evoluíram ao longo dos últimos anos. A

inserção digital, os padrões de acesso dos usuários, e as expectativas também mudaram. É preciso saber se essa evolução se reflete também no padrão de acesso dos usuários e em suas experiências. Essa é a principal motivação do nosso trabalho.

Neste trabalho fazemos um estudo comparativo da evolução entre as transmissões *online* ao vivo das Copas do Mundo de Futebol da FIFA de 2014 e de 2018, ambas realizadas pelo Globo.com, um dos maiores provedores de conteúdo da América do Sul (Seção 2). Realizamos uma caracterização geral da carga imposta nos servidores (Seção 3), quantificamos as principais características do comportamento dos clientes e da qualidade de serviço recebida (Seção 4) e correlacionamos métricas relacionadas à qualidade de experiência (QoE) com o engajamento do usuário (Seção 5).

Em particular, nossos resultados evidenciam:

- O tráfego total no serviço de transmissão cresceu mais de 300%. Além disso, a taxa de chegada de clientes mais do que duplicou no início das partidas. Enquanto o primeiro resultado motiva utilização de mecanismos de auto-dimensionamento (*autoscaling*), o segundo indica que sua implantação ficou mais desafiadora.
- Mostramos a evolução na qualidade do vídeo entregue pelos principais provedores de acesso à Internet (ISPs).
- Constatamos que usuários de dispositivos móveis aumentaram a taxa de adaptações de *bitrate* ao mesmo tempo que apresentaram um decréscimo na quantidade interrupções causadas por esvaziamento do *buffer*.
- A qualidade máxima do vídeo distribuído aumentou de 2564 (HD) para 3564 kbps (FULL HD).
- O público médio cresceu 176% entre 2014 e 2018. Quantidade de sessões por usuário possui comportamento similar em ambos os torneios.

Embora existam diversos trabalhos de caracterizações na literatura, eles não realizam estudos de longo prazo [Veloso et al. 2002, Ali-Eldin et al. 2015]. Esta limitação impede estudo do impacto da evolução de variáveis como população, tipo de vídeo, tecnologia de transmissão, infraestrutura e tipo evento.

Em suma, apresentamos uma nova perspectiva sobre a evolução de um serviço de distribuição de conteúdo de grande escala. Nossos resultados podem auxiliar a tomada de decisões de provedores de conteúdo relativas a modificações no *software* cliente, como no algoritmo de adaptação de taxa de codificação, e na infraestrutura de distribuição, particularmente para planejamento de capacidade e auto-dimensionamento, podendo levar a redução de custos e aumento da qualidade de experiência dos usuários.

2. Arquitetura do Sistema e Conjunto de Dados

A Copa do Mundo da FIFA é um evento esportivo de grande interesse¹. O maior provedor de conteúdo do Brasil, o Globo.com, realizou a transmissão das copas dos anos de 2014 e 2018. Neste trabalho, nós realizamos uma análise dos registros de acesso dos servidores de conteúdo desse provedor. O torneio de 2014 ocorreu entre 12 de junho e 13 de julho, enquanto o de 2018 aconteceu entre 14 de junho e 15 de julho. Em cada um desses torneios participaram 32 seleções com 64 partidas realizadas.

¹<https://www.fifa.com/worldcup/news/fans-drive-record-digital-interest-in-fifa-world-cup>

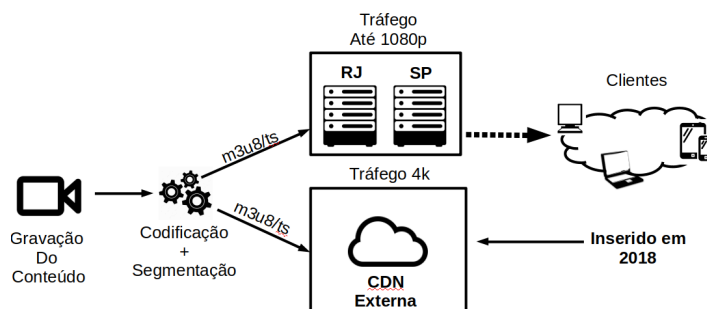


Figura 1. Infraestrutura de transmissão do Globo.com.

A infraestrutura de transmissão está representada na Figura 1. Partidas são capturadas no estádio e o vídeo é codificado em diferentes qualidades e particionado em segmentos (*chunks*) contendo alguns segundos de vídeo. Esses passos são realizados utilizando o *Apple HTTP Live Streaming (HLS)*, uma implementação de *HAS* criada pela *Apple*. No *HLS* os segmentos de vídeo são armazenados em arquivos com extensão (.ts) enquanto os arquivos de índice (chamados de *playlist*) possuem extensão (.m3u8). Os índices contêm metadados indicando parâmetros de configuração do canal como, por exemplo, a *URL* dos segmentos de vídeo para cada uma das qualidades disponíveis. O reprodutor de vídeo (*player*) utilizado foi o *Clappr*², um reprodutor de vídeo *open source* criado pela equipe do *Globo.com* com o propósito de ser utilizado na transmissão da *Copa do Mundo de 2014*³. O *player* utiliza fatores como a ocupação do *buffer*, qualidade da rede e dimensões da tela do dispositivo para escolher a qualidade apropriada.

A infraestrutura do sistema em 2014 consistia em dois centros de processamento de dados e distribuição de conteúdo localizados nas cidades de São Paulo e Rio de Janeiro. Ambos os centros são conectados a pontos de troca de tráfego (PTT) e provedores de trânsito. A arquitetura em 2018 foi estendida pela contratação de uma rede de distribuição de conteúdo (CDN) externa para entregar conteúdo com resolução 4K. Além disso a CDN externa serviria para absorver qualquer excesso de demanda além da capacidade dos centros do *Globo.com*. No entanto, um integrante da equipe do *Globo.com* nos informou que isso não ocorreu em momento algum.

Nós utilizamos os *logs* dos servidores do Rio de Janeiro e São Paulo para realizar a nossa análise. Esses *logs* consistem em respostas às requisições de vídeo HTTP realizadas periodicamente pelo *player*. Cada linha do *log* contém informações como o IP do usuário, status da requisição, *user agent*, data, hora, URL e quantidade de bytes transmitidos. Os dados de sessão avaliados não apresentam identificador único. Dessa maneira, utilizamos metodologia semelhante a trabalhos anteriores [Santos et al. 2016, Guarnieri et al. 2017] onde o endereço IP e o texto de *user agent* são usados para identificar de maneira única um cliente. Ainda assim, essa metodologia pode subestimar o número de sessões únicas. Porém, acreditamos que o número de colisões de identificação é negligenciável. De fato, quando avaliamos a metodologia de identificação proposta, em uma partida que apresentou identificadores únicos de sessão, menos de 1% das sessões identificadas tem colisão de identificadores únicos.

²<https://github.com/clappr/clappr>

³<http://blog.flavioibeiro.com/clappr-an-extensible-media-player-for-the-web/>

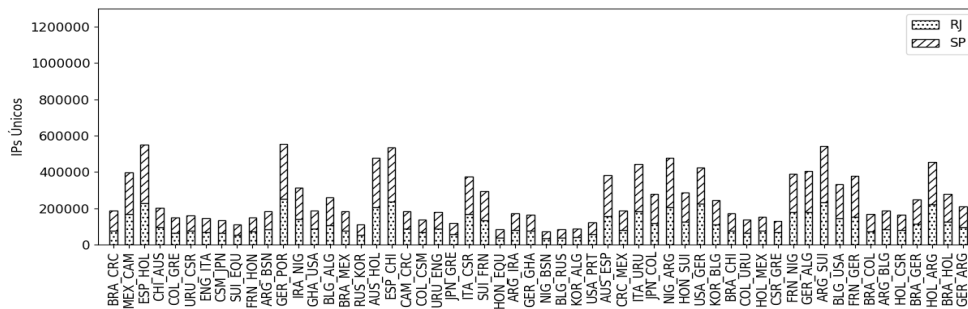


Figura 2. Endereços IP únicos por partida – 2014.

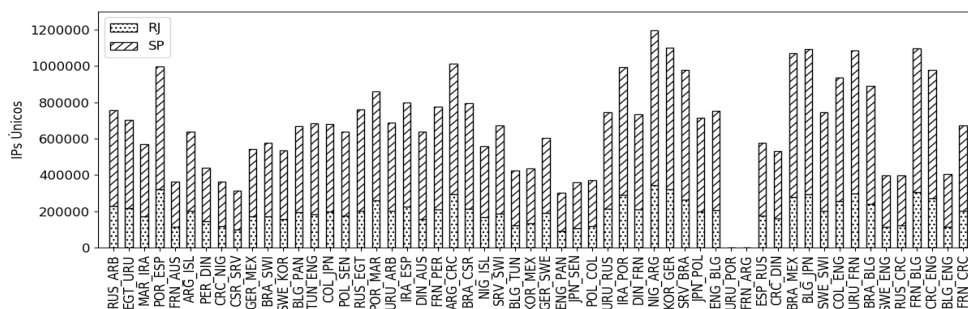


Figura 3. Endereços IP únicos por partida – 2018.

Para reconstruir uma sessão, reutilizamos método proposto em trabalho anterior [Santos et al. 2016] e consideramos que requisições de segmentos de um mesmo usuário em um período inferior a 180 s pertencem a mesma sessão. Outros limiares entre 30 e 300 s levam a resultados quantitativamente equivalentes [Guarnieri et al. 2017]. Além disso, a *string* do *user agent* nos permite separar os usuários em dois grupos de acordo seu dispositivo: computador pessoal (PC) ou dispositivo móvel (*mobile*).

3. Visão Geral da Carga

Nesta seção, caracterizamos a carga de trabalho dos servidores em função de quatro métricas: (i) quantidade de endereços IP únicos em cada jogo, (ii) quantidade de sessões, (iii) volume de tráfego gerado ao longo da transmissão e (iv) taxa de chegada de novas sessões. Essa caracterização inicial nos permite avaliar o crescimento da carga de trabalho de serviços de transmissão de vídeo ao vivo neste período de quatro anos.

As Figuras 2 e 3 mostram a quantidade de endereços IP únicos presentes em cada partida. Não conseguimos os dados para duas partidas de 2018 e, portanto, elas não estão representadas na Figura 3. A contagem por endereços IP únicos considera separadamente os centros de distribuição (RJ ou SP) do provedor de serviço.

O jogo com maior número de endereços IP únicos, observado em 2014, ocorreu entre Alemanha x Portugal. Esta partida apresentou 555.931 endereços. Em 2018, o momento de transmissão mais concorrido ocorreu durante os jogos de Nigéria x Argentina e Islândia x Croácia, que foram realizados no mesmo horário. Nesse momento, houve um total de 1.195.315 endereços IP únicos, um acréscimo de 115% quando comparado ao maior público de 2014.

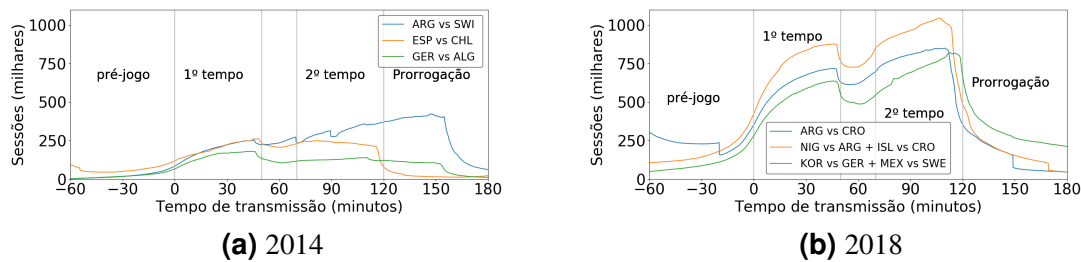


Figura 4. Sessões ao longo do tempo

O público médio, na Copa do Mundo de 2014, contou com aproximadamente 252 mil endereços IP, com 55% conectados ao centro de distribuição de São Paulo e 45% ao do Rio de Janeiro. Já em 2018 o público médio contou com aproximadamente 697 mil endereços, dos quais 29% foram servidos pelo Rio de Janeiro e 71% por São Paulo. Portanto, houve um acréscimo de 166% na quantidade de endereços IP considerando o público médio. Acreditamos que a concentração é superior em São Paulo pois há uma quantidade maior de conectividade no PTT dessa cidade se comparado com o ponto de troca de tráfego do Rio de Janeiro.

Para avaliar a quantidade de sessões e o tráfego de rede ao longo das partidas, nós consideramos três jogos representativos com: (i) alto tráfego, (ii) médio tráfego e (iii) baixo tráfego. As demais partidas apresentam padrão similar. A Figura 4 representa a quantidade de sessões ao longo do tempo da partida, com indicações sobre o período em que se encontra. Separamos esses períodos em pré-jogo, 1º tempo, 2º tempo e pós-jogo/prorrogação. Apenas duas das partidas analisadas possuíam prorrogação, sendo elas entre Argentina x Suíça e Alemanha x Argélia. Na copa de 2018, algumas partidas ocorreram concomitantemente e não foi possível distinguir o tráfego relativo a cada uma delas. Assim, essas partidas foram analisadas conjuntamente.

Há vários fatores que influenciam o interesse do espectador ao longo do jogo. Podemos citar, por exemplo, elementos como o placar da partida, horário e seleções envolvidas. Através da Figura 4a notamos que o maior número de sessões ativas para o torneio de 2014 ocorreu durante a prorrogação de Argentina x Suíça. Apresentando pouco mais de 400.000 sessões simultâneas, este jogo permaneceu empatado sem gols até os 12 minutos do segundo tempo da prorrogação.

Na Figura 4b percebemos que houve crescimento significativo de sessões, chegando até mesmo a ultrapassar 1 milhão de sessões paralelas para as partidas entre Nigéria x Argentina e Islândia x Croácia que ocorreram no mesmo horário. Vemos também que esses jogos ultrapassaram 800.000 sessões concorrentes no seus respectivos picos de usuários, que ocorreram no final do segundo tempo. É possível notar um acréscimo significativo de sessões no início dos tempos e abandono nos finais. Tal padrão de comportamento também foi notado nos demais jogos do torneio.

A Figura 5 apresenta o tráfego de rede que a aplicação de transmissão ao vivo impõe ao provedor de conteúdo. Essa figura apresenta a média do tráfego, segundo a segunda, utilizando uma janela deslizante de 30 s. O crescimento no número de usuários mostrado anteriormente é acompanhado de um aumento significativo na demanda da rede. De fato, a Figura 5 apresenta um comportamento similar ao da Figura 4 e indica um aumento de mais de 300% em 2018 se comparado com 2014, com pico de mais de 200

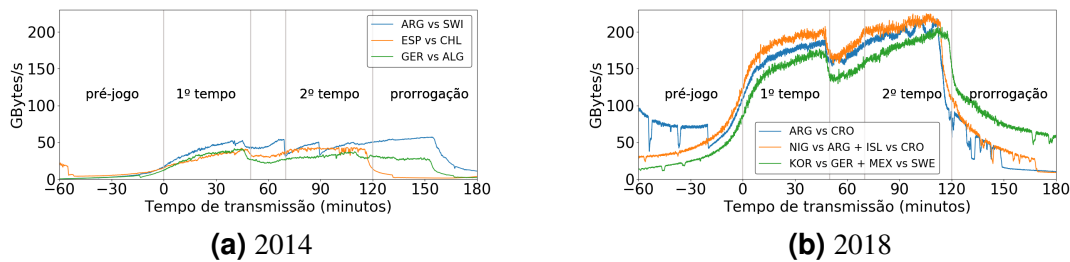


Figura 5. Tráfego ao longo do tempo.

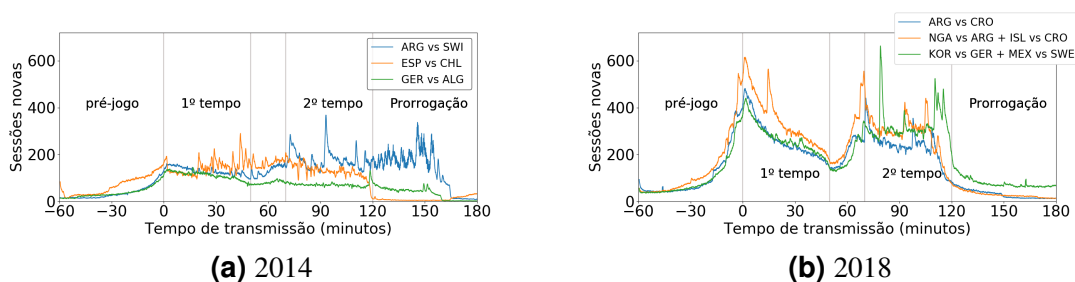


Figura 6. Sessões novas ao longo do tempo 2014 e 2018.

GBytes por segundo. Notamos que embora o tráfego tenha crescido mais 300%, a quantidade média de endereços IP únicos aumentou menos de 200%. Esse comportamento demonstra que os usuários do torneio de 2018 receberam qualidade de vídeo superior.

A Figura 6 mostra a taxa de chegada de novos usuários utilizando 30 segundos de janela deslizante. Nessa Figura percebemos que, na maior parte das vezes, os picos de chegada de clientes ocorrem no início e final do jogo. Acreditamos que o pico inicial acontece pois se trata de um evento ao vivo e com horário bem definido. Nesse caso usuários tendem a acessar o serviço no início da transmissão do evento e sua taxa de chegada ocorre em rajadas [Sripanidkulchai et al. 2004]. Já o pico nos finais do jogo nós imaginamos que ocorra devido à curiosidade e interesse no resultado final da partida e como ele influenciará no restante do torneio.

A Figura 7 apresenta a função de distribuição acumulada da qualidade do vídeo distribuído para clientes de determinados provedores de acesso à Internet (ISP). Utilizamos os dados dos 10 jogos com maior quantidade de endereços IP únicos nessa análise. Para construir esse gráfico utilizamos todas as requisições feitas no horário do jogo, pegamos o total de *bytes* de cada uma delas e convertimos o endereço IP para o ISP correspondente utilizando a biblioteca *pyisp*⁴. Utilizamos os 5 ISPs com maior número de usuários. Os provedores são numerados de forma consistente nas figuras.

Percebemos através dessa figura que a taxa máxima transferida no torneio de 2014 foi de 2564 kbps (HD), enquanto que em 2018 foi de 3475 kbps (FULL HD). Essa é uma evolução natural por conta do aumento de recursos do provedor de conteúdo e dos clientes. No entanto, lembramos aqui que a taxa máxima de 3475 kbps na copa do mundo de 2018 ocorre devido à mudanças na infraestrutura de distribuição. Taxas superiores existem e são entregues por uma CDN externa, cujos *logs* não temos acesso. Em razão disso não é possível contabilizarmos qual a magnitude desse tipo de tráfego.

⁴<https://github.com/ActivisionGameScience/pyisp>

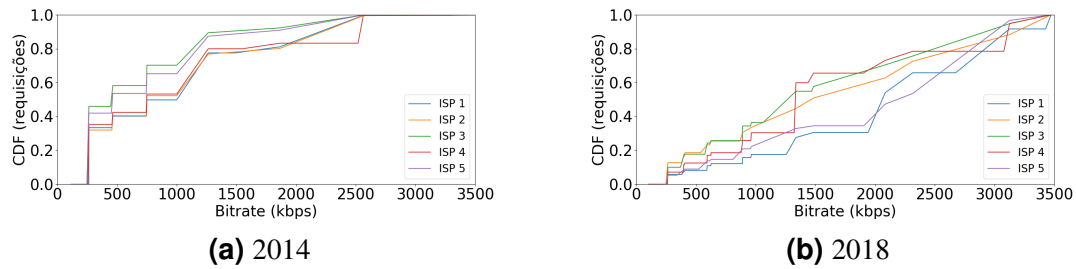


Figura 7. Taxa de transmissão nos 10 jogos com mais endereços IP Únicos.

Mais ainda, notamos que a qualidade de vídeo distribuído para clientes de todos os ISPs aumentaram consideravelmente. Usuários das operadoras 1, 2 e 4 recebiam vídeo em taxas mais altas que os demais provedores em 2014. Essas três operadoras possuíam qualidade similares até quase 2000 kbps, após essa taxa os clientes do ISP 4 recebiam vídeo de codificação mais alta. Em 2018, as curvas não se sobrepõem significativamente, como em 2014. Clientes do ISP 1 recebem vídeo com codificação superior aos demais, seguidos por usuários da operadora 5. Nesse período, a taxa entregue para clientes da operadora 5 superou a de usuários dos ISPs 2 e 4 para a maioria das codificações.

4. Comportamento dos clientes

Os clientes do sistema estudado podem ser representados pelo modelo da Figura 8. Esse modelo é o mesmo utilizado em trabalhos anteriores, que caracterizam vídeos dessa natureza [Santos et al. 2016]. Quando uma sessão de um cliente está ativa, definimos que o cliente está no estado ON. Se o usuário não contém nenhuma sessão ativa no momento é dito que ele está no estado OFF. Medimos o tempo de OFF calculando o tempo entre sessões de um cliente, enquanto o tempo de ON é a duração da sessão. Utilizamos as sessões dos 10 jogos com maior número de endereços IP únicos em nossa análise.

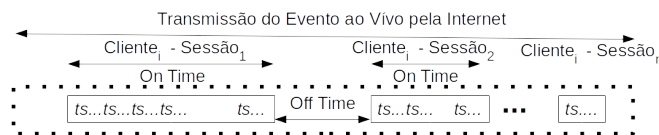


Figura 8. Modelo do comportamento dos clientes (figura adaptada de [Santos et al. 2016])

Por conta de diversos fatores como melhora na conectividade da rede e mudanças de parâmetros de configuração da infraestrutura (e.g, *buffer*, algoritmo de adaptação)

Estado	Ano	Dispositivo	Melhor Fitting	R^2	Primeiro Parâmetro	Segundo Parâmetro	Terceiro Parâmetro	Quarto Parâmetro
ON	2014	PC	Gamma	0.9910	0.6328	0.2669	46.6148	-
ON	2014	Mobile	Log Normal	0.9839	1.5396	0.1791	5.2418	-
ON	2018	PC	Mielke	0.9951	5.5022	1.0298	2.6699	1.4148
ON	2018	Mobile	Exp. Weibull	0.9943	1.7230	5.4964	2.6699	6.7985
OFF	2014	PC	Exp. Weibull	0.9893	1.2941	0.6127	2.9999	15.1016
OFF	2014	Mobile	Exp. Weibull	0.9880	1.9211	0.4507	2.9999	4.3899
OFF	2018	PC	Exp. Weibull	0.9890	2.35018	0.3860	2.9999	3.3831
OFF	2018	Mobile	Exp. Weibull	0.9951	1.9443	0.4766	2.9999	4.9710

Tabela 1. Parâmetros dos *fittings* para os tempos de ON e OFF

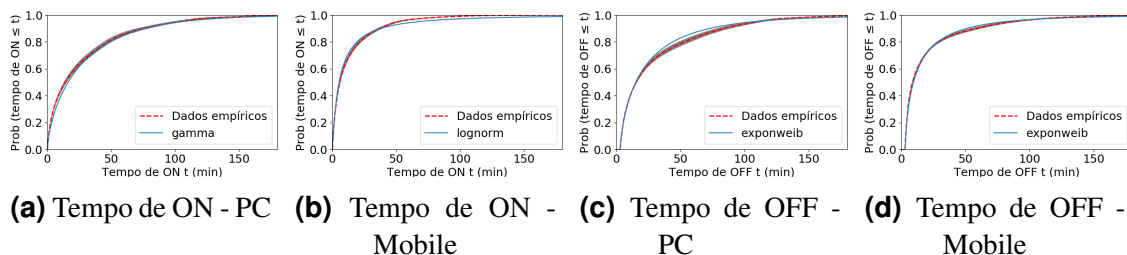


Figura 9. 2014

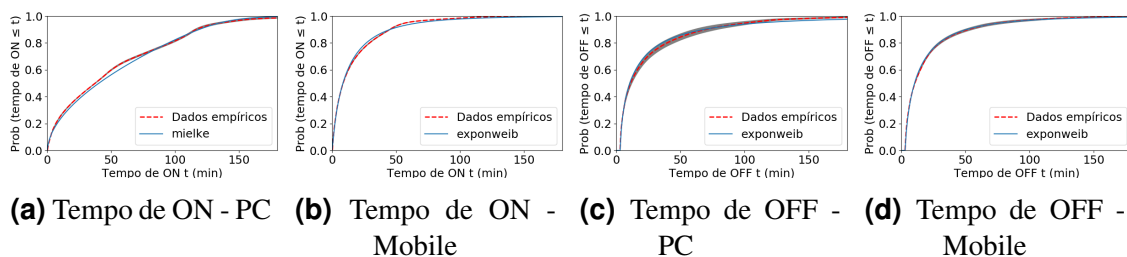


Figura 10. 2018.

imaginamos que parâmetros de modelos como o que criamos em trabalhos passados [Santos et al. 2016] precisam ser atualizados. Fazemos o *fitting* de diversas distribuições estatísticas com o objetivo de encontrar qual é a que se encaixa aos dados reais. Comparamos diferentes distribuições de duas formas: primeiro avaliamos visualmente as curvas dos dados e do *fitting*. Se houver dúvida entre mais de uma distribuição comparamos o valor do coeficiente de determinação, também chamado de R^2 , entre a função de distribuição acumulada do *fitting* e dos dados empíricos. Por fim, utilizamos a distribuição que possui o maior R^2 . Os parâmetros estão sumarizados na Tabela 1.

Podemos ver nas Figuras 9 e 10 as distribuições que mais se ajustam aos tempos de ON e OFF em cada torneio, separadas por tipo de dispositivo utilizado pelo usuário. A região cinza presente nos gráficos representa o erro com intervalo de confiança de 99% utilizando 10 amostras (10 jogos com mais endereços IP únicos). O tempo de ON aumentou consideravelmente para usuários de PC, enquanto que aumentou ligeiramente em usuários de Mobile. Tempo de ON para sessões de PC se ajustavam melhor à distribuição Gamma, mas para 2018 a Mielke é utilizada. Sessões de Mobile utilizavam a Log Normal, e passaram a ser representadas pela Weibull Exponencial. Por fim, o tempo de OFF não variou de maneira considerável e é explicado melhor pela Weibull Exponencial, independente do torneio e tipo de dispositivo.

Nós também tentamos identificar se houve mudança na quantidade de sessões de um usuário entre os períodos avaliados. Para isso construímos a função massa de probabilidade do número total de sessões para cada cliente do serviço através da Figura 11. Nessa figura, apresentamos a média e o desvio padrão da média, de todas as partidas avaliadas, além da distribuição que melhor se encaixou aos dados empíricos. Percebemos que os resultados são similares para ambos os torneios com a binomial negativa sendo a distribuição que mais se adequa aos dados.

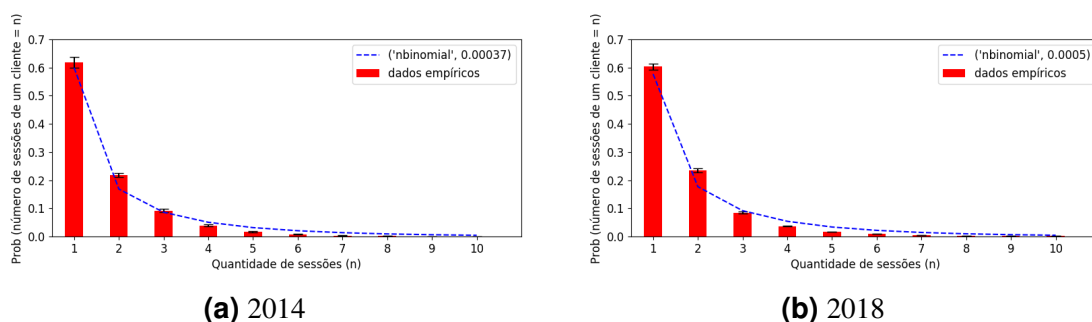


Figura 11. PMF do número de sessões por usuário para os 10 jogos com mais endereços IP Únicos.

5. Qualidade de Experiência dos Usuários (QoE)

Nos últimos anos, o que se tem utilizado para medir qualidade em serviços de *streaming* pela Internet são fatores diretamente percebidos pelos usuários, conhecidos como métricas de *QoE*. Analisamos as seguintes métricas:

- Atraso de inicialização (*Start-up Delay*): O tempo necessário para preencher o *buffer* do cliente e iniciar a reprodução.
- Taxa de reprodução média ao longo de uma sessão (*bitrate médio*).
- Taxa de interrupções (*taxa de stalls*): Frequência de eventos de esvaziamento do *buffer* (e subsequente interrupção da reprodução) durante uma sessão.
- Taxa de adaptações: Frequência de eventos de adaptação da taxa de reprodução do vídeo durante uma sessão.

Vários trabalhos [Dobrian et al. 2011, Ahmed et al. 2017, Guarnieri et al. 2017] mostram que o *bitrate* médio, *start-up delay* e *buffering ratio* são as métricas que mais impactam no engajamento do usuário. Utilizamos a duração da sessão para mensurar objetivamente o engajamento. *Buffering ratio* é representada pela divisão do tempo total em que o vídeo ficou pausado aguardando o preenchimento do *buffer* pela duração da sessão. Não utilizamos em nossa análise essa métrica porque nossos dados são provenientes dos *logs* do servidor o que dificulta reconstruir a sessão de tal forma que esses valores sejam fiéis ao que o cliente presenciou. Por conta disso utilizamos como alternativa a taxa de *stalls*. Calculamos que uma interrupção ocorre sempre que há um esvaziamento do *buffer* e simulamos o comportamento do *buffer* com base em todas as requisições recebidas em uma sessão. Acreditamos que essa seja uma boa metodologia a ser utilizada pois há uma interdependência forte entre a taxa de interrupções e o *buffering ratio* [Ahmed et al. 2017].

O comportamento do usuário varia de acordo com o dispositivo utilizado [Balachandran et al. 2013, Guarnieri et al. 2017]. Por conta disso, separamos nosso *dataset* e analisamos os usuários de computadores pessoais (PC) e de dispositivos móveis (*Mobile*) isoladamente.

Para fazer as análises a seguir filtramos todas as sessões com duração menor do que 20 segundos para eliminar os chamados *early quitters* [Balachandran et al. 2013]. Esses usuários podem sair por vários motivos, mas acreditamos que a maioria saem por falta de interesse. Como nosso objetivo não é medir o interesse e sim avaliar o impacto das métricas no engajamento, resolvemos excluir essas sessões.

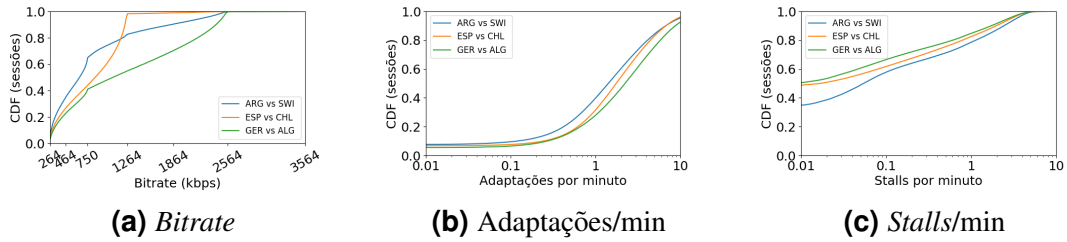


Figura 12. 2014 - PC.

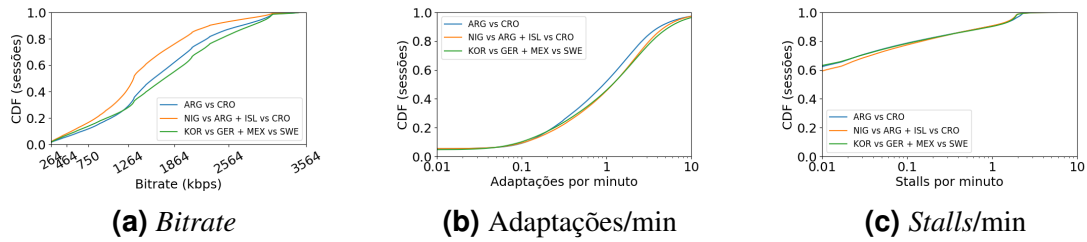


Figura 13. 2018 - PC.

A copa do mundo de 2018 foi transmitida com qualidade de até 4K, no entanto nosso *dataset* contém qualidade de até 1080p (*FULL HD*), isto é, *bitrate* até 3475 kbps (Figura 13a). O tráfego 4K era servido pela CDN externa contratada enquanto que a copa de 2014 foi transmitida com qualidade máxima de 2564 kbps (720p - *HD*) (Figura 12a). Com isso vemos um aumento significativo na qualidade do vídeo recebido. Observamos também que os usuários de dispositivos móveis possuem qualidade superior aos usuários de PC para o torneio de 2014 (14a e 12a) enquanto que para o de 2018 ela é similar (15a e 13a). Essa diferença acentuada que ocorre em 2014 se dá em parte pela limitação de qualidade que existiu em alguns jogos para usuários de PC, mas acreditamos que também seja afetada pela qualidade da rede e poder de processamento dos dispositivos que evoluíram nesse período de quatro anos.

Para medir a taxa de adaptações, nós utilizamos todas as mudanças de *bitrate* que ocorreram durante a reprodução, tanto aumento quanto diminuição de *bitrate*. Utilizamos essa abordagem porque em ambos os casos o usuário percebe de forma negativa essa mudança [Nam et al. 2016]. Para usuários de PC a taxa de adaptações caiu. Por exemplo, enquanto que em 2014 cerca de 18% dos usuários possuíam taxa de até 0.5 adaptações por minuto (Figura 12b), em 2018 esse valor subiu para um pouco mais de 30% dos usuários (Figura 13b), mostrando que menos mudanças ocorreram. Por outro lado, as sessões dos usuários de dispositivos móveis não apresentam a mesma característica. O jogo entre Espanha e Chile tem comportamento diferente dos demais jogos do primeiro torneio (Figura 14b). As qualidades para usuários de PC para essa partida foi limitada, supomos que manualmente pelo provedor de conteúdo. Essa ação gerou economia de recursos e presumimos que esses recursos “extras” permitiram melhorar as métricas analisadas para os usuários de dispositivos móveis (Figuras 14b e 14c). Os demais jogos apresentaram taxa similar entre si. Diferente dos usuários de PC, a taxa de adaptações aumentou em dispositivos móveis (Figura 14b e 15b). A princípio nos pareceu contra intuitivo esse comportamento pois imaginamos que iria ocorrer menos mudanças de *bitrate* para melhorar a experiência do usuário. Observando com mais atenção notamos que a taxa de interrupções diminuiu (Figuras 14c e 15c). Como *stalls* impactam mais negativamente

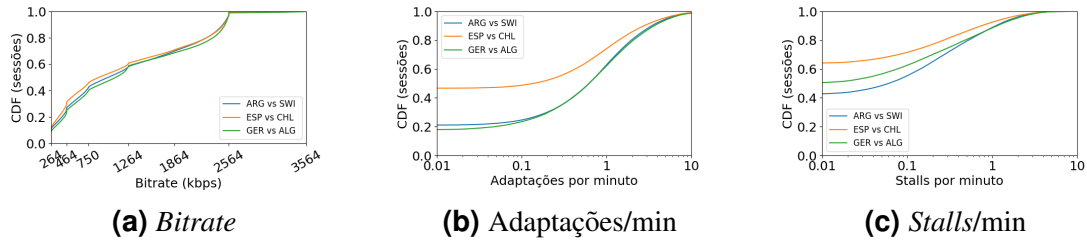


Figura 14. 2014 - Mobile.

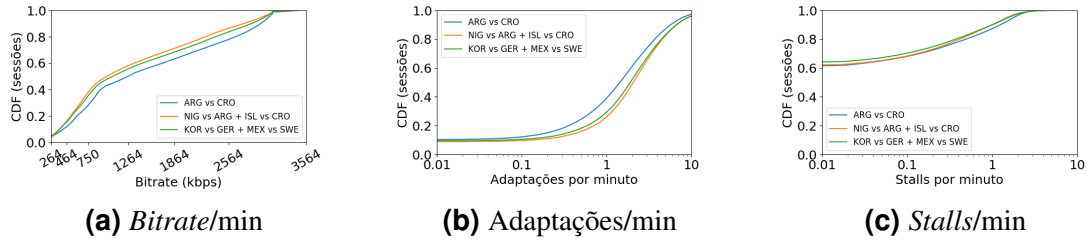


Figura 15. 2018 - Mobile.

no engajamento do que adaptações [Dobrian et al. 2011], acreditamos que a heurística de adaptação do *player* optou por alterar o *bitrate* para não ocorrer interrupções.

Avaliamos também o *start-up delay*. Observamos que dispositivos móveis em 2014 possuíam *start-up delay* superior a PCs. Já em 2018 esse valor diminuiu, não havendo diferenças significativas no atraso na inicialização, independente do dispositivo (gráficos omitidos por questões de espaço). É bem conhecido que um alto atraso para iniciar a reprodução em um serviço de vídeo influencia negativamente a experiência dos usuários. Ao diminuir o tamanho do *buffer*, o *start-up delay* também diminui. No entanto, um *buffer* muito pequeno aumenta consideravelmente o número *stalls* que influencia no engajamento dos clientes mais negativamente do que o *start-up delay* [Guarnieri et al. 2017]. Portanto, é necessário que o *buffer* tenha um tamanho que permita armazenar conteúdo suficiente para ser reproduzido em momentos de oscilações na rede. Mais ainda, por se tratar de um evento ao vivo, o *buffer* não pode crescer muito pois aumenta o atraso da transmissão. Acreditamos que por conta desses fatores o tamanho do *buffer* não variou significativamente entre os períodos avaliados.

Até agora não consideramos o engajamento do usuário em nossa análise, capturado indiretamente pela duração da sessão. Normalizamos a duração de uma sessão pelo tempo restante até o final da transmissão do jogo (incluindo prorrogação, se houver). Eliminamos de nossa análise todas as sessões que iniciaram nos 15 minutos finais da partida para evitar viés, pois elas tenderiam a ter um engajamento normalizado próximo de 1.

Discretizamos o eixo X em 50 bins nas Figuras 16 e 17 além de computarmos a média da duração normalizada para cada um deles. O engajamento aumentou para todas as taxas de *bitrate* (Figura 16), exceto para os clientes de PC do jogo entre Argentina e Suíça (Figura 16c). Acreditamos que a prorrogação nesse jogo aumentou o interesse dos usuários e causou esse comportamento.

Em relação à quantidade de adaptações, geralmente, quanto maior for a taxa, menor o engajamento (Figura 17). Limitamos o eixo X em 10 adaptações por minuto pois

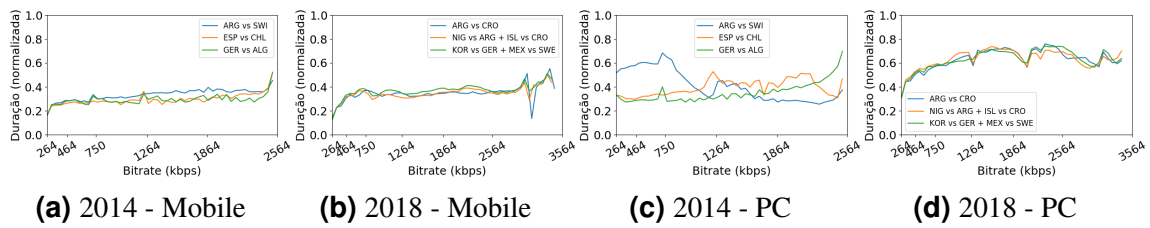


Figura 16. Bitrate x Duração normalizada.

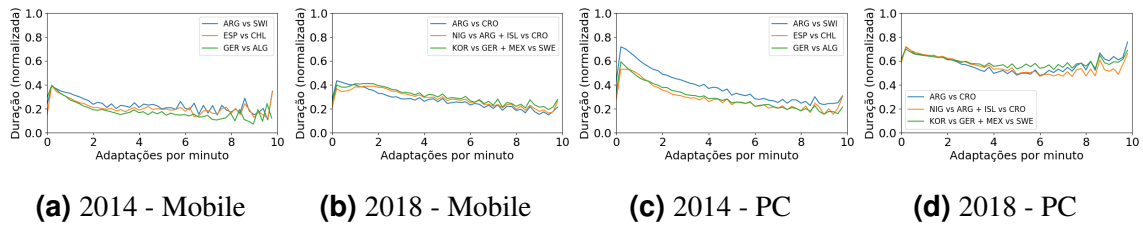


Figura 17. Adaptações x Duração normalizada.

há poucas sessões (i.e., $< 8\%$) com número superior a esse que geram viés nos dados durante o cálculo dos *bins*. Note que o impacto maior ocorre para dispositivos PC em 2014 (Figura 17c). Acima de 10 adaptações por minuto para esse gráfico temos entre 3.68% e 7.62% do total de sessões.

Interrupções na reprodução do vídeo são os eventos que mais impactam no engajamento segundo vários trabalhos existentes [Dobrian et al. 2011, Balachandran et al. 2013, Ahmed et al. 2017]. Mostramos essa influência na Figura 18. Nós calculamos a média de interrupções a cada grupo de 50 *bins* (eixo x). Quanto menor a taxa de interrupções, maior o engajamento na maioria dos casos. Essa correlação tem maior intensidade para taxas maiores. Vemos uma exceção no jogo entre Argentina e Suíça para usuários de PC. Nesse jogo ocorre um alto engajamento para sessões com alta taxa de interrupções e acreditamos que tenha relação com a prorrogação.

6. Trabalhos Relacionados

Há um conjunto de trabalhos avaliando o desempenho de serviços de *streaming* de vídeo com o objetivo de melhorar a qualidade percebida pelos usuários desse tipo de aplicação. Trabalhos como [Hands and Wilkins 1999, Mok et al. 2011] avaliaram métricas de serviço da rede do usuário (e.g., *jitter*), perda de pacotes) e seus impactos na qualidade de vídeo e experiência percebida pelos usuários. Atualmente o custo para armazenar dados históricos é pequeno gerando análises com base nesses dados. Esse tipo de estudo é chamado de *Data-Driven QoE Analysis*.

Até onde sabemos, [Dobrian et al. 2011] é o primeiro trabalho dessa natureza que ganhou bastante notoriedade. Eles analisaram um *dataset* com informações sobre a reprodução de vídeos sob-demanda e ao vivo. Essas informações foram coletadas por instrumentação no cliente. Nesse trabalho os autores definiram algumas métricas de QoE e avaliaram como cada uma delas interferiu no engajamento, além de mostrarem a magnitude da correlação entre essas métricas com o engajamento. Para contornar limitações da avaliação de correlação realizada, os autores fizeram uso da noção de ganho de informação (*Information Gain*) para quantificar o quanto cada métrica impacta no

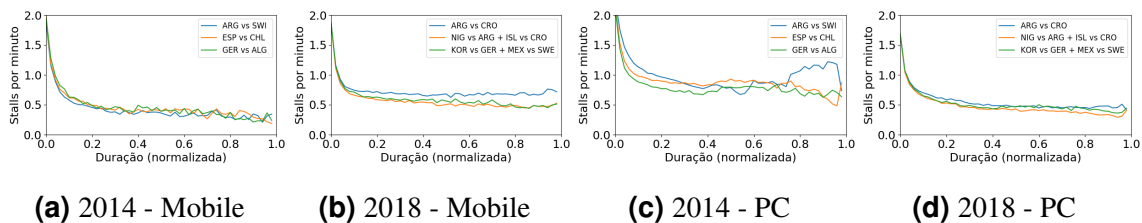


Figura 18. Stalls x Duração normalizada.

engajamento. Eles notaram que embora o *bitrate* médio possua correlação baixa com o engajamento, o ganho de informação é não trivial. Isso ocorreu pois seu relacionamento com o engajamento é não monotônico. Nós também identificamos esse relacionamento não monotônico em nosso trabalho. [Balachandran et al. 2013] utilizaram métricas semelhantes a [Dobrian et al. 2011] e incluíram alguns fatores como tipo de vídeo, dispositivo utilizado e conectividade. Com isso eles utilizaram técnicas de *machine learning* para treinar um modelo baseado em árvore de decisão com objetivo de prever o engajamento do usuário. [Ahmed et al. 2017] abordaram o problema de predição de outra maneira ao utilizar um procedimento matemático chamado de análise de componentes principais para detectar prejuízos na qualidade de experiência do usuário em tempo real.

Em nossa análise nós estendemos esses trabalhos ao realizar uma avaliação detalhada tanto do impacto nos servidores quanto do comportamento dos clientes. Além disso, fazemos um estudo comparativo da evolução ocorrida em um intervalo de quatro anos para um mesmo tipo de vídeo e conteúdo.

7. Discussão e Conclusões

Em geral, análises sobre o comportamento do usuário e qualidade percebida por eles são realizadas mais facilmente através de medição no próprio dispositivo do cliente geralmente feita por ferramentas criadas exclusivamente para monitorar algumas questões como qualidade da rede, ocupação do *buffer* e interrupções. No entanto esse tipo de informação nem sempre está disponível. Nesse caso as sessões do usuário são reconstruídas utilizando os *logs* de acesso dos servidores *web*. Essa abordagem permite realizar análises com fidelidade satisfatória e por isso também é utilizada em trabalhos da literatura (e.g. [Ahmed et al. 2017]). Utilizamos essa técnica para desenvolver nosso trabalho.

Nós ampliamos o conhecimento do impacto gerado nas redes de distribuição de conteúdo e dos fatores que influenciam no engajamento do usuário fazendo um estudo evolutivo de como a melhora na conectividade influenciou essas questões nos últimos anos em um grande provedor de conteúdo esportivo. Mostramos que houve aumento significativo na taxa de adaptações das sessões nos usuários de dispositivos móveis que pode estar ligado à melhoria na heurística de adaptação do *player* para evitar *stalls*. Relatamos também aumento de mais de 300% no tráfego gerado em determinados momentos, além de um acréscimo de 166% no público médio. Identificamos valores similares em relação à quantidade de sessões por cliente nos dois torneios. Nós também atualizamos os modelos que refletem o comportamento do usuário. Por fim mostramos a variação no engajamento em relação às métricas de QoE avaliadas. Acreditamos que nosso trabalho é um importante passo para auxiliar provedores de conteúdo a tomar decisões sobre modificações na infraestrutura e guiar alterações nos algoritmos de adaptação existentes.

Agradecimentos

Os autores agradecem o apoio de CAPES, CNPq, FAPEMIG, FAPERJ e FAPESP.

Referências

- Ahmed, A., Shafiq, Z., Bedi, H., and Khakpour, A. (2017). Suffering from buffering? detecting qoe impairments in live video streams. In *2017 IEEE 25th International Conference on Network Protocols (ICNP)*, pages 1–10. IEEE.
- Ali-Eldin, A., Kihl, M., Tordsson, J., and Elmroth, E. (2015). Analysis and characterization of a video-on-demand service workload. In *Proceedings of the 6th ACM Multimedia Systems Conference*, pages 189–200. ACM.
- Balachandran, A., Sekar, V., Akella, A., Seshan, S., Stoica, I., and Zhang, H. (2013). Developing a predictive model of quality of experience for internet video. In *ACM SIGCOMM Computer Communication Review*, volume 43, pages 339–350. ACM.
- Dobrian, F., Sekar, V., Awan, A., Stoica, I., Joseph, D., Ganjam, A., Zhan, J., and Zhang, H. (2011). Understanding the impact of video quality on user engagement. In *ACM SIGCOMM Computer Communication Review*, volume 41, pages 362–373. ACM.
- Guarnieri, T., Drago, I., Vieira, A. B., Cunha, I., and Almeida, J. (2017). Characterizing qoe in large-scale live streaming. In *IEEE GLOBECOM*, pages 1–7.
- Hands, D. and Wilkins, M. (1999). A study of the impact of network loss and burst size on video streaming quality and acceptability. In *Int. Workshop on Interactive Distributed Multimedia Systems and Telecommunication Services*, pages 45–57. Springer.
- Mok, R. K., Chan, E. W., and Chang, R. K. (2011). Measuring the quality of experience of http video streaming. In *Integrated Network Management*, pages 485–492.
- Nam, H., Kim, K.-H., and Schulzrinne, H. (2016). Qoe matters more than qos: Why people stop watching cat videos. In *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*, pages 1–9. IEEE.
- Open Connect (2011). Netflix open connect. <https://openconnect.netflix.com/en/>. Acessado em: 23/03/2019.
- Santos, B., Carnivali, G., Almeida, W., Vieira, A. B., Cunha, Í., and Almeida, J. (2016). Caracterização do comportamento dos clientes de vídeo ao vivo durante um evento de larga escala. In *Simpósio Brasileiro de Redes de Computadores*.
- Schlinder, B., Kim, H., Cui, T., Katz-Bassett, E., Madhyastha, H. V., Cunha, I., Quinn, J., Hasan, S., Lapukhov, P., and Zeng, H. (2017). Engineering egress with edge fabric: Steering oceans of content to the world. In *Proceedings of the Conference of the ACM Special Interest Group on Data Communication*, pages 418–431. ACM.
- Sripanidkulchai, K., Maggs, B., and Zhang, H. (2004). An analysis of live streaming workloads on the internet. In *Proc. of the ACM SIGCOMM conference on Internet measurement*, pages 41–54.
- Veloso, E., Almeida, V., Meira, W., Bestavros, A., and Jin, S. (2002). A hierarchical characterization of a live streaming media workload. In *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurement*, pages 117–130. ACM.