

Uma Avaliação de políticas de adaptação de Taxa de bits para sessões de Vídeo 3D

Introdução

A técnica de adaptação de taxa de bits a partir dos clientes é um dos conceitos responsáveis pelo sucesso das aplicações de streaming de vídeo. Existe um corpo grande de trabalhos nessa temática que abordam o problema da adaptação da taxa sob diferentes perspectivas. Do ponto de vista dos dados usados para tomada de decisão as diversas abordagens podem ser classificadas em dois grandes grupos: i) sessão em andamento e ii) sessão passada. A primeira usa apenas os dados da sessão em andamento para tomar decisão sobre como será a adaptação de sua taxa de bits, sendo as técnicas baseadas em médias as que mais se destacam. A segunda considera o histórico de sessões anteriores para tomar a decisão sobre a adaptação da taxa, tendo-se utilizado técnicas de aprendizagem de máquina, e.g., aprendizagem por reforço, para definir as taxas de bits.

Em ambientes virtuais imersivos, i.e., vídeos em 3D, uma grande quantidade de dados e a existência de características próprias desses ambientes requer a definição de novas técnicas ou o aperfeiçoamento das técnicas de adaptação de taxa de bits já consagradas. Um exemplo de característica particular dos ambientes imersivos é o conceito de campo de visão que define onde está o interesse da audiência que consumirá o conteúdo. A técnica de codificação de vídeo baseada em esquadrilhas, do inglês *tile streaming*, permite que as decisões tomadas possam considerar o conteúdo em escala diferentes. A região onde está o campo de visão tem maior importância e portanto precisa ser priorizada. À medida que o conteúdo se afasta do campo de visão a sua importância para definição da qualidade da sessão diminui.

As técnicas de adaptação de taxa de bits até então desenvolvidas consideram a escala de tempo definida pela duração dos segmentos criados durante a codificação do vídeo. Em outras palavras, um conjunto de cenas em um segmento terá a mesma qualidade, fato que não tem correspondência em uma sessão de vídeo em 3D, que tem no campo de visão da audiência um fator importante mas não o único para definição da qualidade da sessão. A investigação da adequação das técnicas de adaptação de taxas de bits, desenvolvidas originalmente para operar na escala definida pela duração dos segmentos, é oportuna.

As avaliações devem ocorrer em um cenário de conectividade caracterizado pelo uso do protocolo HTTP3 e QUIC, que permitem a multiplexação de vários fluxos em uma única conexão. A possibilidade de priorização desses fluxos, que correspondem às diversas regiões, que formam as cenas de um segmento de vídeo 3D, sugere que a decisão sobre a taxa de bits em cada fluxo seja também orientada por sua prioridade, e considere critérios mais alinhados à prioridade.

Objetivo

Implementar e avaliar abordagens clássicas para adaptação de fluxos de vídeos, construídos usando técnicas de HTTP streaming, no contexto de ambientes imersivos, medindo a eficiência dessas abordagens por meio de métricas qualitativas.

Especificamente, esse objetivo será realizado da seguinte forma:

- Implementar e avaliar política baseado em ocupação de buffer;
- Implementar e avaliar política baseada em vazão do canal;
- Implementar e avaliar política baseada em vazão do canal e ocupação

Metodologia

As técnicas de adaptação de fluxos de vídeos em ambientes convencionais (2D) possuem uma extensa lista de resultados na literatura da área. Os esforços despendidos para organizar esses resultados classificam essas técnicas a partir de diferentes perspectivas. Neste trabalho, utilizaremos como critério de caracterização das técnicas as variáveis envolvidas na tomada de decisão sobre a qualidade dos próximos segmentos a serem baixados. Essas variáveis são: vazão do canal e a ocupação do buffer.

As técnicas escolhidas devem ser implementadas em um cliente HAS que seja capaz de simular a reprodução de conteúdo imersivo, e.g., 3D. A decisão sobre a qualidade do fluxo deve considerar o campo de visão da audiência, o que torna a transmissão de vídeo 3D única. Dessa forma, será preciso a tomada de decisão considerando que no entorno do campo de visão existem regiões que podem ser tratadas de forma diferentes, conforme ilustrado na figura abaixo.



A região identificada por "visão" é onde se encontra o campo de visão da audiência, ou seja, a região de maior prioridade na oferta de qualidade. À medida que nos afastamos do campo de visão, a prioridade deve ser reduzida com relação à qualidade ofertada na sessão.

Ao medir a qualidade de uma sessão de vídeo é preciso considerar: continuidade, imagens apresentadas e a vivacidade. A dimensão da continuidade captura o nível de interrupção no fluxo durante a sessão. Frequência e duração das interrupções devem ser utilizadas para avaliar como essas interrupções afetaram a percepção da audiência. A dimensão de imagens apresentadas captura quão distante essas imagens estão da melhor qualidade relativa, i.e., que poderia ser apresentado no dispositivo da audiência, disponível para acesso. Finalmente, a dimensão da vivacidade captura quão distante está a duração da sessão e do vídeo apresentado. Nessa dimensão, mede-se o impacto da bufferização inicial, que ocorre antes da apresentação do primeiro segmento, e a bufferização gerada por interações da audiência, e.g. avanços e retomadas.

Tarefas

Nesta seção listam-se as tarefas necessárias para a execução do projeto.

Implementar abordagens de adaptação de taxa de fluxos de vídeo. Deve-se implementar uma técnica de referência em cada uma das abordagens relacionadas a adaptação do fluxo em sessões de vídeo streaming. Cada técnica de referência deverá ser implementada e validada em simulador de consumo de vídeo 3D (Cliente). Existe um cliente de referência disponível em [1].

As técnicas que serão consideradas neste estudo são estas:

1. Ocupação do buffer
2. Vazão do canal
3. BOLA

O algoritmo de **ocupação do buffer** usa a seguinte regra para definir a qualidade(bitrate) do próximo segmento a ser baixado, considerando que

Throughput: vazão média de um canal (e.g. EWMA conforme descrito abaixo)
bufferLevel: Ocupação atual do buffer (unidade de tempo, e.g. segundos)
fragmentDuration: Duração do segmento (definido em tempo de codificação do vídeo)

INSUFFICIENT_BUFFER_SAFETY_FACTOR: Parâmetro de controle da agressividade da vazão medida no canal, e.g. 0.5 foi usado na implementação do algoritmo.

O bitrate é definido como segue:

```
bitrate = throughput * (bufferLevel / fragmentDuration) * INSUFFICIENT_BUFFER_SAFETY_FACTOR;
```

É preciso mapear o valor do `bitrate` para uma das taxas disponíveis no conjunto de versões geradas para um determinado vídeo. Ao final de cada evento de download o `bitrate` precisa ser reavaliado.

O Algoritmo baseado em **Vazão do canal** usa uma média móvel exponencialmente ponderada. As considerações feitas para a implementação do algoritmo são estas:

1. O número de segmentos considerado na média é quatro;
2. A ponderação da última vazão medida e do histórico de medidas considera o tempo de download do segmento
 - a. Quanto maior o tempo de download mais relevante é a última medida da vazão;
 - b. Quanto menor o tempo de download mais relevante é o histórico de medidas da vazão.
3. O cálculo é realizado em dois modos. O primeiro considera a ponderação completa, e é chamado de *fast*, e o segundo considera apenas a metade da ponderação, e é chamado de *slow*. Dessa forma, a ponderação é calculada como segue:
 - $\alpha_{fast} = (1/2)^w$, w é o tempo de download, em segundos, do último segmento baixado
 - $\alpha_{slow} = (1/2)^{w/2}$, w é definido da mesma forma que no modo *fast*.

A atualização da vazão média (μ) é então calculada para os dois modos, conforme segue, e o menor valor obtido no cálculo é usado para definir a taxa de bits do próximo segmento :

- $\mu_{fast} = (1 - \alpha_{fast}) * \delta + \alpha_{fast} * \mu_{fast}$, δ é a vazão instantânea medida pelo acesso ao último segmento
- $\mu_{slow} = (1 - \alpha_{slow}) * \delta + \alpha_{slow} * \mu_{slow}$, δ é a vazão instantânea medida pelo acesso ao último segmento
- $bitrate = \min(\mu_{fast}, \mu_{slow})$

É preciso mapear o valor do `bitrate` para uma das taxas disponíveis no conjunto de versões geradas para um determinado vídeo. Ao final do download de cada segmento o `bitrate` precisa ser reavaliado.

O Algoritmo **BOLA** possui uma lógica mais complexa e está descrito em [2].

Para todas as técnicas implementadas será necessário considerar que para o vídeo 360, codificado usando tile, a escala de decisão é diferente que as usadas com vídeo convencional (2D) , que é dada pelo segmento. Uma codificação de tile 10x12 indica que 120 fragmentos precisam ser baixados para efetuarmos a reprodução do segmento. Esse cenário exige que as técnicas descritas anteriormente sejam adaptadas. Em outras palavras, é preciso considerar a escala do segmento para a tomada de decisão.

O conteúdo deve ser provido por um servidor de vídeo, que responde às requisições usando o protocolo QUIC. Existe um servidor de referência disponível em [1].

Para a realização da validação das implementações, o proponente deverá criar uma instância de ambiente de simulação controlado, usando mininet. Nesse ambiente, cenários de conectividade de rede devem ser criados simulando o provimento de conteúdo na última milha. Um canal opera com uma taxa constante (bandwidth) mas as instâncias da aplicação de vídeo devem dividir esse canal com outras aplicações, gerando um tráfego background que consuma de 10 a 30% da banda passante. O canal vai funcionar com as seguintes bandas: 10 e 8 Mbps. A rede conecta o servidor aos clientes com um certo atraso de propagação de 100ms(instantânea), 1000(machine is working), e 2000ms(mental context switch). Outro parâmetro da rede é a taxa de perda, que é igual a 2%.

Ordem de prioridade para construção dos cenários:

- 1) Cenário #01: 10%(load), 10 Mbps (Banda), 100ms(atraso)
- 2) Cenário #02: 30%(load), 8 Mbps (Banda), 2000ms(atraso)
- 3) Cenário #03: 10%(load), 10 Mbps (Banda), 2000ms(atraso)
- 4) Cenário #04: 30%(load), 10 Mbps (Banda), 1000ms(atraso)
- 5) Cenário #05: 10%(load), 8 Mbps (Banda), 2000ms(atraso)
- 6) Cenário #06: 30%(load), 10 Mbps (Banda), 2000ms(atraso)
- 7) Cenário #07: 10%(load), 10 Mbps (Banda), 1000ms(atraso)
- 8) Cenário #08: 30%(load), 8 Mbps (Banda), 1000ms(atraso)
- 9) Cenário #09: 10%(load), 8 Mbps (Banda), 100ms(atraso)
- 10) Cenário #10: 30%(load), 10 Mbps (Banda), 100ms(atraso)
- 11) Cenário #11: 10%(load), 8 Mbps (Banda), 1000ms(atraso)
- 12) Cenário #12: 30%(load), 8 Mbps (Banda), 100ms(atraso)

Referências

[1] Fernandez, G. 360 Video streaming over QUIC.

<https://github.com/gufernandez/aioquic-360-video-streaming>

[2] K. Spiteri, R. Ugaonkar and R. K. Sitaraman, "BOLA: Near-Optimal Bitrate Adaptation for Online Videos," in *IEEE/ACM Transactions on Networking*, vol. 28, no. 4, pp. 1698-1711, Aug. 2020, doi: 10.1109/TNET.2020.2996964.