Transmissão de Áudio em Tempo Real sobre Redes Wireless IEEE 802.11 com Foco em Acessibilidade para o Cinema Digital

Caio Marcelo Campoy Guedes, Danilo Assis Nobre dos Santos Silva, Guido Lemos de Souza Filho

Programa de Pós-Graduação em Informática
Centro de Informática - Universidade Federal da Paraíba
{caio.marcelo, danilo, guido}@lavid.ufpb.br

Resumo: A transmissão de reforço de áudio acessível em salas de cinema é um exemplo em que o envio de conteúdo em tempo real é um fator limitante. Nessa aplicação, é desejável que o retardo entre a captura e exibição do áudio mantenha-se abaixo de um limiar de 40 ms. Este estudo tem como objetivo desenvolver uma solução capaz de transmitir áudio em tempo real sobre redes Wi-Fi entre um servidor e múltiplos clientes. Almejando o máximo de resiliência, a solução faz uso de *Forward Error Correction* e redundância temporal para mitigar os efeitos indesejáveis. Por fim, ainda é proposto um novo algoritmo adaptativo que busca ajustar o *buffer* de *playback* com o propósito de atenuar o *jitter* da rede.

Palavras-chave: *transmissão de áudio; áudio; FEC; algoritmo adaptativo; redundância;* smartphones*;* wireless*;* jitter.

1. Introdução

Com a constante luta para inclusão de deficientes visuais e auditivos em apresentações artísticas, observa-se o surgimento de diversas tecnologias desenvolvidas cujo propósito é auxiliar esses indivíduos a terem melhor compreensão sobre determinada apresentação, seja ela uma peça teatral ou uma exibição em salas de cinema. As técnicas presentes no mercado variam desde o uso de painéis textuais em que é realizada estenografia, ao uso de transmissores de áudio e a presença de intérpretes de línguas de sinal.

O reforço de áudio individual é um recurso bastante delicado, uma vez que variações no envio do conteúdo, seja por perda de pacotes ou latência, afetam diretamente a experiência do usuário. Dentre as tecnologias desenvolvidas para transmitir reforço de áudio em eventos de acessibilidade, destaca- se o uso de transmissão por FM (*Frequency Modulation*) e o uso de tecnologia IR (*infrared*). Embora ambas as soluções sejam maduras e possuam excelente desempenho no processamento e envio dos dados, existem problemas de alto custo e ajustes legais.

A alocação de frequências FM deve ser efetuada junto a um órgão governamental especializado, sendo necessário realizar um trâmite extenso em cada local onde se deseja implantar a solução. Ressalta-se também que a regulamentação depende dos órgãos nacionais especializados, que pode variar a depender do país em questão. Segundo Lan [1], o principal problema no acesso à tecnologia de radiofrequência está intrinsecamente relacionado à defasada regulamentação e à baixa efetividade nos padrões de alocação, o que limita o acesso aos recursos e vantagens oferecidas pela tecnologia. Outro fator relevante é o investimento necessário para a compra de equipamentos dedicados e aquisição da concessão pública para uso da frequência desejada que difere entre países e regiões.

Já as soluções que utilizam a tecnologia IR, lidam com um alcance muito menor quando comparada a solução que utiliza FM. Entretanto, há o benefício de não necessitar de homologação junto aos órgãos governamentais para atuação, o que a torna mais atrativa. Todavia, ainda há a utilização de um hardware dedicado, o que encarece a solução.

Dentre as alternativas baseadas em transmissão não guiada, a comunicação através de redes *wireless* tem-se difundido rapidamente por conta da sua praticidade e eficiência, beneficiando diferentes aplicações como, por exemplo, a transmissão de conteúdos audiovisuais, jogos e centros de entretenimento multimídia [2]. A transmissão sem fio também destaca-se pela flexibilidade no conjunto de soluções tecnológicas como as propostas nos padrões de comunicação elaborados pelo IEEE (*Institute of Electrical and Electronics Engineers*), que permitem o uso de diferentes frequências e larguras de banda [3]. Com a facilidade de implantação e baixo custo na instalação de bases *wireless*, serviços que utilizam transmissão de dados para múltiplos pontos podem ainda se beneficiar de protocolos que otimizam a transmissão de fluxos de áudio para grupos (*multicast*). Todavia, ao utilizar redes *wireless*, observa-se grande instabilidade quando realiza-se uma transmissão de baixa latência, o que é, em geral, proveniente de atraso no envio do conteúdo e a colisão entre os pacotes [4].

Embora a transmissão por modulação em frequência seja mais eficaz em transmitir o conteúdo, por possuir menor latência, o *tradeoff* entre alto custo e *delay* é extremamente crucial neste tipo de aplicação. Assim, diante dos resultados das análises realizadas sobre as tecnologias disponíveis, optou-se pela utilização de bases *wireless* IEEE 802.11.

Além do uso das bases *wireless* é importante definir o dispositivo que será utilizado para reproduzir os conteúdos transmitidos. Como um dos focos da solução é prover mobilidade ao usuário, já que a mesma será utilizada primordialmente em cinemas, decidiu-se utilizar *smartphones* como *hardware* de *playback*. *smartphones* modernos possuem alto poder computacional, permitindo flexibilidade durante o desenvolvimento do *software* de reprodução, assim como fácil substituição caso haja a necessidade de troca

do dispositivo ou caso se faça necessário um *upgrade* em algum dos aparelhos.

O grande desafio da escolha da tecnologia a ser utilizada na resolução do problema proposto é a latência, caracterizada como o intervalo entre o que está sendo emitido pelo locutor e o que está sendo escutado pelo cliente, pois não é desejado que haja a ocorrência de eco ou perda de sincronização labial, o que ocasiona uma experiência desagradável ao usuário. Eventos audiovisuais que necessitam de reforço de áudio são, em geral, limitados por um fator de tempo bem definido. Na literatura, um limiar onde não se percebe a divergência na apresentação entre as trilhas de áudio e vídeo, em televisores, deve ser um delta igual ou inferior a 40 ms entre o momento em que a imagem é exibida e o áudio é reproduzido [5][6]. Logo, se a transmissão do conteúdo não conseguir suprir a latência em valor máximo de 40 ms, ocorrerá um grande desconforto para o usuário, já que para deficientes auditivos o filme poderá ser exibido na frente da trilha de áudio, e para deficientes visuais a audiodescrição poderá sobrepor diálogos do filme.

Dessa forma, o escopo do trabalho aqui proposto é a transmissão de áudio com valores de latência da ordem de 40 ms, que é baixa o suficiente para atender requisitos de sincronização labial e transmissão de reforço de áudio em serviços de acessibilidade em salas de cinema.

2. Metodologia

Este trabalho busca desenvolver uma solução cliente-servidor, denominada FAST (*Fast Audio Streamer*), que gerencia a transmissão e reprodução de fluxos de áudio em tempo real com baixo retardo através do uso de redes Wi-Fi para múltiplos destinatários, cujo objetivo é melhorar a experiência de usuários em salas de cinema. Buscando resumidamente responder a seguinte questão de pesquisa: Como transmitir áudio para grupos de usuários utilizando dispositivos móveis como receptores com latência entre captura e exibição igual ou inferior a 40 ms?

Sendo assim, o objetivo geral deste trabalho é desenvolver uma solução de transmissão e recepção de áudio em tempo real de baixo retardo, a qual subdivide-se em dois componentes: o servidor, responsável pelo envio do conteúdo, e a aplicação cliente para dispositivos móveis, que reproduzirá a mídia transmitida. A comunicação entre módulos será realizada através do protocolo IP (*Internet Protocol*) e os dados serão transmitidos através de uma rede sem fio Wi-Fi.

Essa solução buscará enviar os dados da forma mais ágil possível sem comprometer a experiência do usuário, ou seja, o FAST atuará realizando uma transmissão de áudio com latência igual ou inferior a 40 ms almejando o mínimo de perda de pacotes. Dessa forma, uma revisão sistemática da literatura foi realizada e três estratégias foram selecionadas, são elas: *Forward Error Correction* [7], redundância temporal e o desenvolvimento de um algoritmo adaptativo que manipula a frequência das amostras. Essas estratégias foram selecionadas com base no uso das tecnologias em mercado e na sua presença em artigos científicos, cujos resultados foram os mais promissores.

A Figura 1 apresenta uma visão macro da arquitetura do FAST e nela é possível visualizar a comunicação entre a fonte e o pipeline de processamento do servidor.

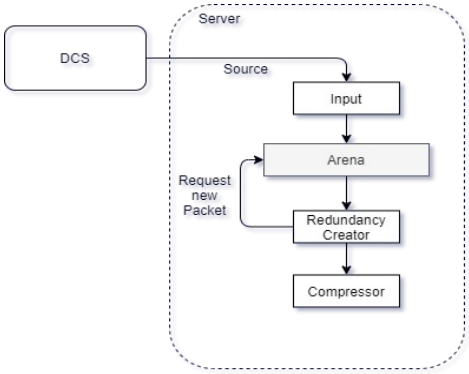

Figura 1. Arquitetura do FAST, visão do servidor.

Inicialmente, uma fonte de áudio, que neste trabalho é definida como DCS (*Digital Cinema Server*), transmite um fluxo de áudio em paralelo com a reprodução do vídeo referente a determinado filme. Em seguida, o módulo de captura do servidor, denominado *Input*, captura o áudio estéreo a uma taxa constante, para este trabalho, em que cada *sample* é definido como uma porção de áudio de 4 milissegundos. Os dados são então enviados a uma arena, que é uma região contínua de memória alocada junto ao *kernel* do sistema operacional. Essa região possui acesso privilegiado e é garantido que o sistema operacional não realizará paginação com essa porção. Em seguida, o módulo *Redundancy Creator* agrupa as amostras de áudio e procede com a criação de um pacote com 5 samples, das quais 4 são redundâncias, ou seja, 16 ms de conteúdo replicado. Esse é então enviado ao *Compressor*, módulo responsável por comprimir os dados e adicionar metadados para reconstrução das amostras através de FEC no lado do cliente. Assim que o processo de compressão é concluído, os dados são transmitidos através do protocolo UDP (*Multicast*) para os clientes conectados no grupo de recepção.

A Figura 2 exemplifica o funcionamento do FAST no lado do dispositivo móvel, neste trabalho, definido como um *smartphone*. Os dados que são recebidos na antena *wireless* do *smartphone* são enviados diretamente ao *Extractor*, módulo responsável pela decodificação dos pacotes, aplicação do FEC e pela inserção apropriada dos pacotes em uma nova *Arena*. É importante destacar que essa implementação da *Arena* não se beneficia da otimização do *kernel* e está sujeita a paginação. O *Extractor* tem visão direta sobre a região de memória onde os pacotes extraídos estão localizados, e, dessa forma, ele consegue identificar se é necessário utilizar alguma das redundâncias para corrigir uma falha na rede, que pode ser uma perda completa de um pacote ou um atraso na recepção. Assim, o módulo pode

corrigir até 4 pacotes perdidos na rede e em última hipótese ele consegue criar uma amostra de áudio, de pior qualidade, a partir dos metadados criados pelo algoritmo de FEC.

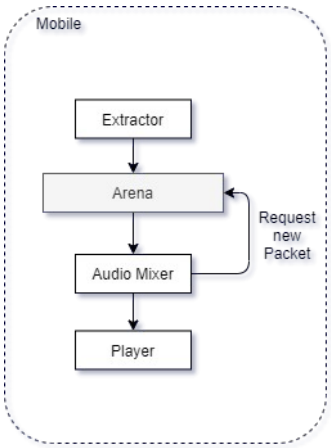

Figura 2. Arquitetura do FAST, visão do cliente.

Após a inserção na *Arena*, o *Audio Mixer* encarrega-se de preparar a trilha de áudio para ser executada para o ouvinte. Dentre as manipulações que o mixer pode realizar sobre o áudio original estão: amplificação do áudio, replicação de um dos canais e reamostragem, que é o algoritmo adaptativo proposto nesse trabalho. Ressalta-se que cada uma dessas opções pode ser desativada pelo usuário da aplicação.

A amplificação do áudio é o processo em que se aumenta o sinal referente ao áudio de entrada de forma uniforme. Esse processo é feito de forma linear onde um fator numérico é multiplicado a cada sample do áudio original. Porém, é importante definir um valor máximo para o processo de ganho, uma vez que existe a possibilidade de *clipping*, visto que o valor resultante da multiplicação pode exceder a precisão das amostras (*bit depth*) definido para o áudio em questão. A replicação dos canais é um requisito do uso da solução em salas de cinema, visto que no áudio estéreo que é transmitido aos espectadores são transmitidos o áudio amplificado da sala, lado esquerdo, que é utilizado por deficientes auditivos e a audiodescrição, lado direito, que é utilizada por deficientes visuais. O algoritmo de replicação de áudio tem complexidade linear, ou seja, dada uma entrada de áudio, é necessário a iteração sobre metade da informação para replicar os dados, e, dessa forma, o algoritmo não é impactante para a latência final da proposta. Além disso, o processo de replicação do conteúdo e a aplicação do ganho podem ser feitos concomitantemente, e, dessa forma, apenas uma iteração é necessária para aplicar ambos os algoritmos.

O *mixer* tem o papel de amenizar os efeitos de variação de retardo causado pela rede através do uso de um algoritmo adaptativo, que é responsável por expandir e contrair as amostras de áudio, assim como o funcionamento de um *buffer* elástico (uma espécie de sanfona), reproduzindo o áudio de forma mais lenta quando o *buffer* está se esvaziando e de forma mais rápida quando ele retorna a um nível aceitável. O áudio é manipulado através da adição ou remoção de dados de amostras do áudio, e, para isso, a ferramenta proposta utiliza a *libsamplerate*, um projeto *open source* para *upsampling* e *downsampling* de dados PCM (*Pulse Code Modulation*). A API (*Application Programming Interface*) fornece 5 mecanismos de *resampling*, variando entre algoritmos lineares e algoritmos por interpolação, e, dessa forma, possibilita a flexibilidade no ajuste da qualidade do algoritmo de *resampling* sem que se faça necessário a troca de bibliotecas. Como o FAST lida com oscilações na transmissão de dados muito rápidas, a correção efetuada pelo algoritmo precisa ocorrer instantaneamente, trazendo benefícios para a qualidade do serviço, uma vez que ao invés do espectador escutar um intervalo de silêncio que, em geral, é percebido como um estalo, o usuário escuta o áudio com uma pequena mudança de tonalidade.

Por fim, os dados são enviados ao *player* para que a informação seja reproduzida para o usuário.

3. Resultados

Para avaliar a solução, foram mapeados oito experimentos distintos utilizando a estratégia 2^kfatorial com o propósito de simular diferentes situações que podem ser encontradas em ambiente de cinema. Para estes experimentos foram mapeadas três variáveis independentes que são variadas com o propósito de avaliar se a aplicação consegue manter o *playback* em 40 ms e se o *buffer* da mesma permanece com ao menos uma amostra. Dessa forma, as variáveis independentes são:

1. Distância entre roteador e dispositivo (3.5 metros ou 7.3 metros);
2. Quantidade de redes na mesma faixa de canais (1 ou 3 roteadores);
3. Quantidade de dispositivos conectados na rede (1 ou 2 dispositivos).

Para cada execução do experimento, foram coletados a ocupação do *buffer*, os áudios da fonte e do dispositivo móvel, com o intuito de comparar a discrepância entre entrada e saída.

Dentre os resultados obtidos nos testes realizados, são apresentados a seguir os resultados para o primeiro experimento, roteador a 3.5 metros, 1 roteador na mesma faixa de canal e 1 dispositivo móvel. A partir de um histograma da ocupação do *buffer* foi possível identificar que a solução consegue manter uma média de alocação em *buffer* de 35 ms de áudio, sendo assim, a solução conseguiu manter-se abaixo do limite estipulado. Também foi possível notar que a solução oscilou a ocupação do *buffer* entre 4 ms de conteúdo e 40 ms, indicando que ela evitou a drenagem completa do *buffer* de *playback*. Ainda foi possível analisar a partir de um gráfico de *timeline* que o *resampling* realizado pelo algoritmo adaptativo auxilia a solução a manter-se em um ponto médio definido em 35 ms. No gráfico, foi visualizado um degrau onde amostras são

perdidas, esticadas por *upsampling*, até atingir um limiar estipulado e contraídas por *downsampling* para compensar a inserção de tempo.

Os demais experimentos realizados indicaram que a solução mantém valores muito similares com relação à ocupação de pacotes em *buffer* independentemente da variação das variáveis independentes, e, portanto, foram omitidos nesse trabalho.

4. Discussão

A solução deste trabalho destaca um serviço de transmissão de reforço de áudio acessível em salas de cinema. Todavia, esse não é o único cenário de uso para a solução. É possível utilizar o FAST em outras situações, como, por exemplo, visitas guiadas a museus, apresentações teatrais, partidas de futebol, dentre outros, trazendo tanto inclusão para os deficientes visuais e auditivos quanto novas formas de consumo de informação aos demais usuários.

5. Conclusões

Neste trabalho foi apresentada uma solução cliente-servidor denominada FAST, cujo foco é a transmissão de áudio em baixa latência sobre redes Wi-Fi. Ela trabalha com três estratégias distintas, que almejam baixa latência através do alívio de perda de pacotes e perturbações causadas sobre os sinais *wireless* pelo meio de transmissão. Desse modo, o uso de FEC, redundância temporal e controle adaptativo de consumo de amostras foram propostos como soluções, que melhoram significativamente a transmissão de fluxos de áudio. Ainda é possível afirmar que a solução apresenta diversos benefícios quando comparada às demais soluções do mercado e a outros trabalhos da literatura, que atingem latência média de 200 ms [7], como a baixíssima latência e o baixo custo de implantação. Esta solução é um dos poucos trabalhos que lida com reforço de áudio entre redes *wireless* e dispositivos móveis com latência inferior a 40 ms.

Bibliografia

[1] Lan, K. and Li, M. (2010) Feasibility study of using FM radio for data transmission in a vehicular network. International Computer Symposium (ICS2010) pp. 55-60. DOI: 10.1109/COMPSYM.2010.5685448.

[2] Kovacevic J.; Samardzija D.; Temerinac M. (2009) Joint coding rate control for audio streaming in short range wireless networks. *IEEE Transactions on Consumer Electronics* 55(2): 486-491. DOI: 10.1109/TCE.2009.5174411.

[3] Mangold, S.; Choi, S., Hiertz, G., Klein, O. and Walke, B. (2003) Analysis of IEEE 802.11e for QoS support in wireless LANs. *IEEE Wireless Communications* 10(6): 40-50. DOI: 10.1109/MWC.2003.1265851.

[4] Hardman V. et al. (1997) Reliable Audio for Use over the Internet. Proc. IINET'95 Conference pp. 171-178.

[5] Ravindran, K.; Bansal, V. (1993) Delay compensation protocols for synchronization of multimedia data streams. *IEEE Transactions on Knowledge and Data Engineering* 5(4): 574-589. DOI: 10.1109/69.234770.

[6] EBU Recommendation R37-2007 The relative timing of the sound and vision components of a television signal https://tech.ebu.ch/docs/r/r037.pdf. Acesso em 05/01/2020.

[7] McKinley, P. and Gaurav, S. (2000) Experimental evaluation of forward error correction on multicast audio streams in wireless LANs. In: Proc. eighth ACM International Conference on Multimedia (MULTIMEDIA '00). Association for Computing Machinery, New York, NY, USA, 416-418. DOI: 10.1145/354384.376304.

[8] Raju, G. et al. (2006) On Supporting Real-time Speech over Ad hoc Wireless Networks. 14th IEEE International Conference on Networks, pp. 1-6, DOI: 10.1109/ICON.2006.302667.