ProtectMe: An Approach to Provide Audio Privacy in Real-time

Nishat Tasnim Niloy
MSSE0706
Email: bsse0723@iit.du.ac.bd

Abu Rafe Md Jamil
MSSE0712
Email: bsse0722@iit.du.ac.bd

Asadullah Hill Galib
MSSE0718
Email: bsse0712@iit.du.ac.bd

Naushin Nower
Associate Professor
Institute of Information Technology
University of Dhaka

Mohammad Shoyaib
Professor
Institute of Information Technology
University of Dhaka

***Abstract*—Audio communication has become very much widely prevalent in this modern era. With the increase of its availability and usability, the privacy concern has become an important issue to be addressed. The participants of an audio conversation like to secure their privacy without any hindrance in the regular activities. Especially, such type of technology is expected in real time. In the past, some notable solution has been provided to resolve privacy-related issues, however, that solution was not that much user-friendly. In this paper, an approach has been proposed to provide the participants of an audio communication full audio privacy in real time. It will completely hide the vocal identity without compromising any information of the audio stream. This approach does not involve any extra hardware mechanism or data analysis to provide the service. In either way, it is very much user-friendly and inexpensive solution for the users to preserve their security in audio conversation.**

***Index Terms*—Audio Privacy, Stream, Audio Modification, Pitch Shifting, Frequency**

I. INTRODUCTION

Audio Privacy is a prevalent issue nowadays [1]. In the current digital communication era, audio communication is ubiquitous. Thousands of million audio signals are transmitted constantly over the internet. But privacy issues regarding that audio communication [2] are often ignored. Lack of audio privacy [3] engenders real risk[4], especially when the speaker tends not to reveal their identity in audio communication. Uncovering the speaker's identity is quite annoying to the speaker. Besides, it breaches the privacy of the speaker. Hence exposing audio privacy should be ceased and performing audio privacy in real time should be done for effectiveness.

In resolving this issue, different approaches are performed in the previous works. Altering the pitch of the sound using predictive filters and Fourier transforms [5] is quite a creditable approach though it lacks performance due to its heavy computation. Modifying the pitch of the sound is also performed in another work [7], but its efficacy in audio privacy is arguable as it needs physical involvement of the speaker. RTFace [6] ensures privacy in video streams using a privacy-aware architecture, though it is not applicable in audio streams. Replacing vocalics using prerecorded LPC coefficients [8] is another sound approach. But its additional requirements of prerecorded LPC coefficients halts its viability in real circumstances. The flaws in these works are to be settled as none of these work can safeguard actual audio privacy in real time.

In this research, we overcome the inadequacy of the existing works. Primarily resampling of the original audio [9] is used to conceal the identity of the speaker. Besides, proper audio privacy is secured by introducing a trustworthy medium [10] and encrypting the data. Moreover, real-time experience is preserved by keeping it a simple, structured and continuous process.

This work is a noteworthy contribution to the field of audio privacy since it protects audio privacy precisely. Furthermore, the real-time processing of audio assures the applicability and effectiveness of the work.

II. RELATED WORKS

Many previous works have been done to modify the audio information in the past. There different kinds of techniques are proposed which are good for certain situations.

An algorithm is proposed by T. Drugman et el [5], which is responsible for altering the pitch of the sound sample. They have used linear predictive filters and Fourier transforms to do the work. But it requires a large amount of computation to retain the formant shape.

RTFace is a mechanism proposed by J. Wang et el [6] which can denature video streams and afterward blurs some selective faces. It follows some predefined policies before providing privacy to certain users. Besides, this approach maintains a scalable and privacy-aware architecture. However, they have not considered audio privacy in their system.

V.Venkatesh et el [7] has come up with a solution to provide an audio-visual privacy solution. There the audio

privacy was confirmed by modifying the pitch of the participants. But this requires the participants to wear a Bluetooth microphone so that the system can identify her.

Chen, Adcock, and Krishnagiri [8] presented a technique which replaces vocalics to preserve the speech along with its intonation automatically. This method promises to provide audio privacy by obfuscating speech while preserving the identifiability of environmental sounds and prosodic information. At first, it detects a vocalic, select that vocalic and finally substitutes that vocalic. In order to do so, it needs a prerecorded LPC coefficient to identify the vocal syllables and then changes the LPC coefficient, where our method does not need any prerecorded information to serve the purpose.

To mitigate these shortcomings, a solution will be provided in this paper, where resampling will be used to bring a modification to the audio to the listener. In this way, the vocal information will remain the same and one can intelligently understand the audio but the speaker cannot be identified through the voice.

III. METHODOLOGY

Audio Privacy is a technique to modify the audio in such a way that one can understand the audio but can not recognize the voices of the speakers. This goal can be achieved by audio time stretching, pitch scaling or audio resampling. In the proposed system the audio resampling technique is used.

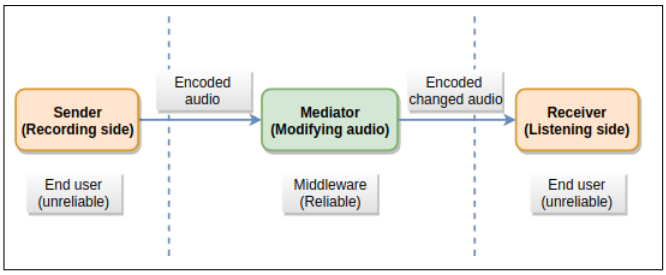

Fig. 1: Architecture of Protect Me software

A. Proposed System

Our proposed system architecture is segmented into three parts:

- Sender
- Mediator
- Receiver

1) Sender: The sender records the audio and sends it to the mediator. It does not save the audio because it is not a reliable part of the architecture.

2) Mediator: The mediator receives the original audio and sends it to the receiver after changing the voices with audio resampling technique. It saves both the original and modified audio because it is the only trusted part of the system.

3) Receiver: The receiver receives the modified audio from the mediator and plays the audio to the listener.

Every communication happened via socket and the exchanged data is always encrypted. It is a continuous process that makes the system real-time audio streaming with privacy. The fig. 1 demonstrate the whole architecture of our proposed system named ProtectMe.

B. Algorithm for Audio Modification

The audio modification algorithm has been implemented in the mediator section. It requires an audio stream as an input while it produces the modified audio streams. The procedure will be accessed every time in the callback which is responsible for the continuous incoming of the input stream from the sender.

Before modification it decodes the encoded audio inputs sent from the sender and after modification again encodes before sending it to the receiver. The audio modification technique is described in algorithm 1.

Algorithm 1 Audio Modification

```
function ToModifyAudio
    input ← encoded audio stream
    if input = null then
        return
    decoded input ← decode(input)
    initialize output[size of decoded input[0] * 2]
                   [size of decoded input[1]]
    i,j ← 0
    loop:
    for decoded input do
        output[j] ← decoded input[i]
        output[j+1] ← decoded input[i]
        i ← i+1
        j ← j+2
    output ← output/max(abs(output))
    encoded output ← encode(output)
```

IV. EXPERIMENT

The solution is performed upon 10 speakers to identify if they are recognizable. The listeners could identify the presence of different speakers, however, the speakers cannot be identified. In the study, the speaker used to send the audio streams to the server from the remote end node. The server propagates the audio streams to the mediator where the modifications occur. After that, the mediator sends the modified streams which are sent to the listening end immediately.

Among all the audio streams, A sample has taken and saved it as a .wav file. A frequency graph is generated using the original audio file and modified audio file to analyzed the characteristics of the audio streams. The graphs are shown in

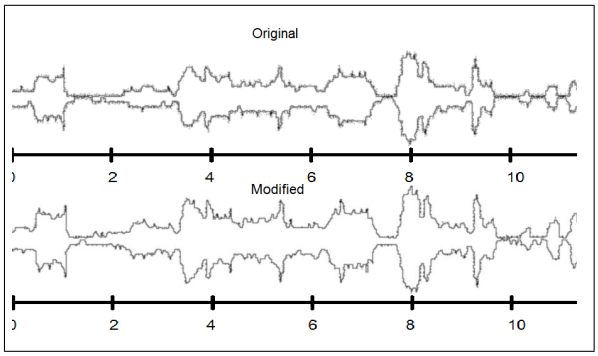

Fig. 2: Graphical representation of the original audio and modified audio streams

figure 2. The fluctuations of the frequencies are visible in the graphs.

V. CONCLUSIONS

As a straightforward and viable solution of audio privacy, this work uses a sound technique - resampling for modifying the audio signal. This technique is sound in a sense that, it is widely used for altering audio signal effectively. Applying this simple approach for audio privacy is a smart solution indeed. Apart from that, providing a real-time solution enhances its feasibility and usefulness. Future work concerns improving the current algorithm and architecture for large scale uses of audio privacy, incorporating the work to provide user-specific user privacy in real time and introducing a privacy-aware policy to the system.

REFERENCES

[1] Philosophy247.org. (2019). Why Care about Privacy?. [online] Available at: http://www.philosophy247.org/podcasts/privacy/ [Accessed 19 May 2019].

[2] Tech, P. (2019). Are smart speakers spying on you? Privacy and security fears rise as devices popularity grows. [online] Financial Post. Available at: https://business.financialpost.com/technology/personal-tech/are-smart-speakers-spying-on-you-privacy-security-fears-grow-as-devices-get-more-popular [Accessed 19 May 2019].

[3] Forbes.com. (2019). Is Your Smart Security Camera Protecting Your Home Or Spying On You?. [online] Available at: https://www.forbes.com/sites/marcwebertobias/2016/08/22/is-your-smart-security-camera-protecting-your-home-or-spying-on-you/#a980f8056dd0 [Accessed 19 May 2019].

[4] Louroe Electronics. (2019). Security: Why Audio? Why Not? - Louroe Electronics. [online] Available at: https://www.louroe.com/security-why-audio-why-not/ [Accessed 19 May 2019].

[5] Drugman, T. and Dutoit, T., 2010, August. A comparative evaluation of pitch modification techniques. In 2010 18th European Signal Processing Conference (pp. 756-760). IEEE.

[6] Wang, J., Amos, B., Das, A., Pillai, P., Sadeh, N. and Satyanarayanan, M., 2017, June. A scalable and privacy-aware IoT service for live video analytics. In Proceedings of the 8th ACM on Multimedia Systems Conference (pp. 38-49). ACM.

[7] Venkatesh, M.V., Zhao, J., Profitt, L. and Sen-ching, S.C., 2009, June. Audio-visual privacy protection for video conference. In 2009 IEEE International Conference on Multimedia and Expo (pp. 1574-1575). IEEE.

[8] Chen, F., Adcock, J. and Krishnagiri, S., 2008, October. Audio privacy: reducing speech intelligibility while preserving environmental sounds. In Proceedings of the 16th ACM international conference on Multimedia (pp. 733-736). ACM.

[9] Ccrma.stanford.edu. (2019). Digital Audio Resampling Home Page. [online] Available at: https://ccrma.stanford.edu/ jos/resample/ [Accessed 19 May 2019].

[10] Ford, J. (2019). Key Communication Skills for the Mediator. [online] Workplace, Team Mediation Conflict Resolution — resologics. Available at: https://www.resologics.com/resologics-blog/2016/7/26/key-communication-skills [Accessed 19 May 2019].