Acoustic User Authentication with Smartphones

Neta Gevirtzer, Ofir Ben Yosef, Alon Eilam
Signal and Image Processing Laboratory (SIPL)
Andrew and Erna Viterbi Faculty of Electrical & Computer Engineering
Technion – Israel Institute of Technology
Technion City, Haifa, Israel, https://sipl.eelabs.technion.ac.il

***Abstract*—This paper describes a work on an acoustic user authentication system using smartphones. The system implements two-factor authentication for Windows workstations, where the authentication procedure, including locking and unlocking the workstation is transparent to the user. Since workstations and smartphones have built-in microphones and speakers, the system does not require additional hardware. The uniqueness of the solution is being based on acoustic signals. These signals are transmitted by the user's smartphone and received by the workstation microphone. The system is "pure play acoustic" since no wiring or radio transmission is used. The system configuration supports multiple users in the same area. Eavesdropping prevention is provided by sequentially generated random one-time keys. Acoustic communication can be applied either in the audible range or beyond the human hearing range depending on the sampling rate of the smartphone and the workstation.**

***Index Terms*—Acoustic, Android, Signal Processing, Smartphone, Two Factor Authentication, Windows**

I. Introduction

Security and privacy leakage are growing concerns for many people. Therefore, the use of authentication mechanisms is increasing in recent years. These mechanisms are based upon a single or a combination of factors that provide reliable authentication. Shah and Kanhere [1] discussed three common factors: "something you are", "something you know", and "something you have". Various applications use different approaches for authentication. The common solution is using passwords ("something you know"), yet this solution is considered unsafe. Some systems require extra steps, such as one time password (OTP), generated by a third-party application ("something you have"), as presented in [2]. Other solutions require additional hardware, e.g., a card reader ("something you have") [3] or unique bodily characteristics such as retinal or fingerprint scan ("something you are") [4]. This paper describes Windows workstation user authentication method based on "something you know" and "something you have" applied in a user-transparent form. RF based methods apply transparent operation by detecting the proximity of the users' smartphones to their workstation. As an example, Bluetooth-based authentication is proposed in [5]. A user-transparent acoustic solution is described in [6], where both workstation and smartphone record the ambient noise via their microphones. The phone compares the two recordings and determines if the computer is in the same environment. Acoustic based solutions have been applied and demonstrated in several other use cases, such as proximity detector [7], positioning systems [8], [8] QR codes replacement [10] and MAC protocol physical layer [11]. In this paper we build on the availability of the acoustic medium to provide a transparent user authentication method. As shown in Fig. 1, messages carrying the command to unlock the workstation are transmitted from the user's smartphone speaker and received by the user's workstation microphone. Since microphones and speakers are standard in smartphones and workstations, the solution does not require additional hardware. The system automatically detects the smartphone acoustic signals when the user is near the workstation, thus the process is transparent. To eliminate unauthorized unlocking by eavesdropping the transmitted messages, random keys are generated sequentially. The random keys are exclusive per user, providing for multiple users in the same room. Using acoustic communication ensures that the user is in the same room as the workstation as opposed to an RF based solutions.

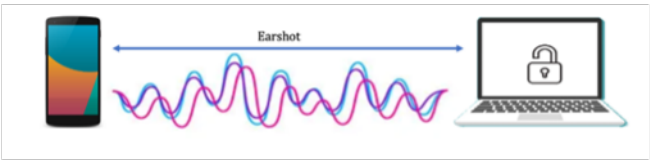

Fig. 1. System configuration

II. System Description

The solution includes two applications; the first is an application that installed on the user's smartphone, which indicates the user's presence near the workstation and sends a one-time key to open or lock it. The second is a desktop application that serves as a dynamic lock screen for the workstation. The workstation will be unlocked only when the correct password is detected. If the smartphone is within earshot (i.e., the user is in the same room, near enough to the workstation), the workstation remains unlocked. Otherwise, when the user is no longer near the workstation, it will lock automatically. Thanks to the characteristic of the audio signal, which unlike RF signals does not pass through walls, the workstation will be unlocked only when the user is in the same room as the workstation and not based solely on spatial proximity.The smartphone and the Windows application are described in the following paragraphs.

A. The smartphone application

Since the Speaker is an essential component, the smartphone application can be applied on any commercial device. In this work we have demonstrated the smartphone application using an Android smartphone. The Android application's UI was written in java, while signal processing and encryption functions were written in C++. The Android application transmits encrypted messages indicating one of three operations the workstation shall perform: lock, unlock, or auto lock. Fig. 2 shows the application's UI. Password generation: upon the first use, the application generates a random personal password, which the user shall enter into the Windows application for system initialization. The communication between the Android and Windows applications is done by transmitting acoustic signals. Each digit in the message is encoded at a symbol, which is the sum of three tones in different frequencies. Message transmission by the Android application is triggered in two manners: Either by the user, clicking the open/close button on the phone screen, or by detection of the smartphone's movement by using its accelerometers.

Fig. 2. Android application UI

B. The workstation application

The workstation application implementation is for Windows, but the proposed solution can be applied on different operation systems. The Windows application's UI was developed using Windows Forms and C#, while the signal processing and encryption functions are the same C++ code used by the Android application. As shown in Fig. 3, the lock screen interface contains an input box for password entering and an interactive messaging field for user instruction. For demonstration purpose, the received signal is plotted in the time and frequency domains using the Scottplot library [12].

A message sent by the smartphone is received by the workstation microphone and processed. Then, only if the received message is appropriately decrypted, one of the three actions is performed: lock, unlock, or auto lock. The solution includes an automatic locking mechanism based on the smartphone's accelerometers. Smartphone motion is detected by reading the accelerometers' values in three dimensions.

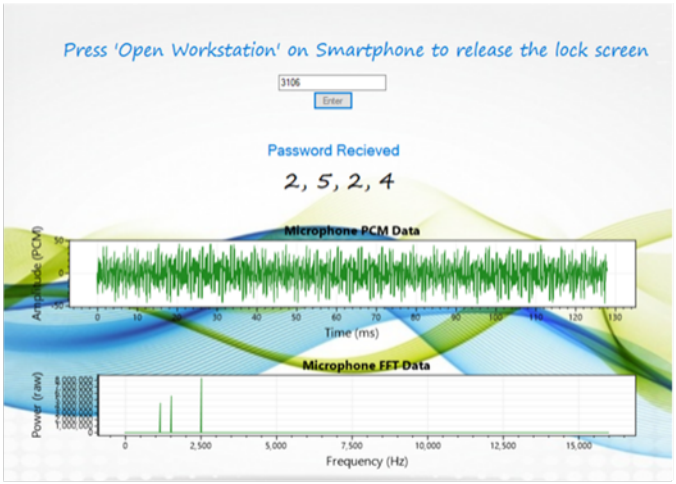

Fig. 3. Windows application UI

Fig. 4 describes the auto lock state machine. Automatic locking allows a convenient user experience by providing a user-transparent locking process that prevents data leakage due to leaving a workstation unlocked. Detection of smartphone motion is a trigger for activating the automatic locking mechanism, and a massage is transmitted accordingly. One minute later, the same message is transmitted again (duration is programmable, in our application, we used a one-minute duration). If the user is not in earshot range, the workstation will not receive the second message transmission, and the workstation will auto-lock. If the workstation receives the second message transmission, it will remain unlocked.

III. COMMUNICATION

A. Physical layer

The solution is implemented using acoustic signal encoding of the messages. The message dictionary consists of sixteen data symbols and one delimiter symbol, each symbol is composed of a combination of 3 out of 16 pre-defined frequencies. A message for locking or unlocking the workstation is a sequence of four symbols. The system is designed to function with various sampling rates and demonstrated using a 32 kHz sampling rate for both smartphone and workstation. In the demonstrated system the pre-defined frequencies ranged between 1-2 kHz, however, system operation can also be performed at ultrasonic frequencies. To optimize accurate detection probability, the three frequencies per each symbol

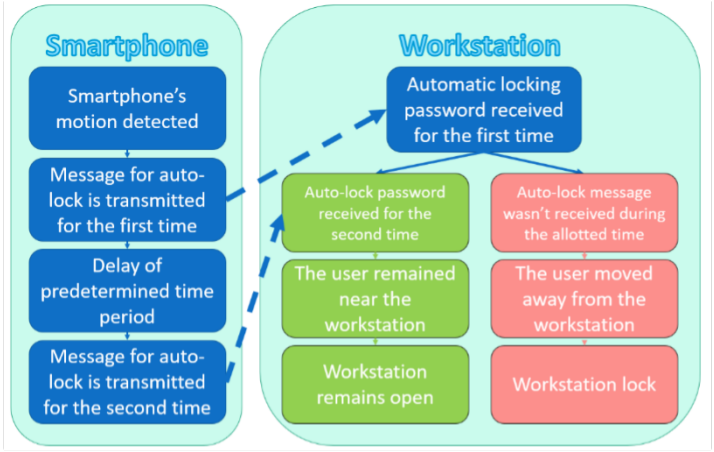

Fig. 4. Auto lock flow

were selected to provide maximum Hamming distance [13]. To optimize the Android application's memory footprint, the symbols' frequencies are defined by their indexes in the frequency domain. The symbol's samples result from IFFT to the time domain. Due to the linear characteristics of the Fourier Transform, a sum of sinusoidal waves results in the time domain as described in (1).

$$\mathcal{F}\{sin(2\pi ft)\} = -\pi j[\delta(\omega - 2\pi f) - \delta(\omega + 2\pi f)] \quad (1)$$

B. Symbol Detection algorithm

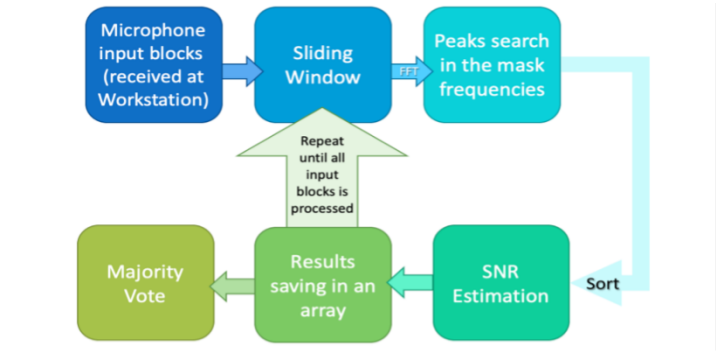

Fig. 5. Massage detection process

Fig. 5 illustrates the massage detection algorithm. As mentioned earlier, each symbol is a combination of three sinusoidal waves at different frequencies. The detection of symbols from the received audio samples block is performed by comparing the energy of 16 pre-defined frequencies (step 3 of the described algorithm). In order to avoid detection of noise when no message is transmitted, the signal-to-noise ratio (SNR) is estimated (step 7 of the described algorithm). When a data signal is received with good SNR, the FFT result will typically contain three frequencies with dominant magnitude, and the SNR : $SNR(dB) = 10 \cdot \log_{10}(P_{signal}/P_{noise})$ will be high. Therefore, the ratio between the energies of the third and fourth highest frequencies provides the SNR estimation. Fig. 6 illustrates this method. In practice, when $energy[4]/energy[3] > 1/3$, the received signal is considered as noise, "no symbol detected" will be returned.

Symbol Detection Algorithm

Given the pre-defined frequencies array F[16].

1) Get a block of 4096 samples, $input\ block\ [0:4095]$.
2) Perform FFT on the given block,
   $energy\ [0:4095] = |FFT\{input\ block[0:4095]\}|$.
3) Mask the known frequencies,
   $energy[i] = \begin{cases} energy[i] & i \in F \\ 0 & else \end{cases}$
4) Sort the amplitude of the chosen frequencies decreasingly.
5) Choose the 3 frequencies with the highest energy.
6) Check if there is a symbol
   If not $\rightarrow$ return -1 "no symbol detected". containing those 3 frequencies.
7) Check the energy of the fourth-highest frequency.
   - If $energy[3] > 3 \cdot\ energy[4] \rightarrow$ return symbol.
   - Else, return "no symbol detected".

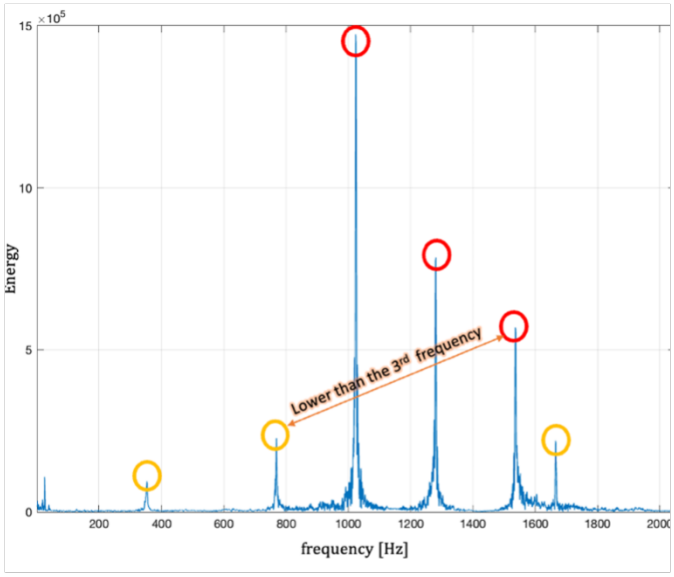

Fig. 6. FFT result of a sampled symbol

C. Sliding Window

As shown in Fig. 1, the system consists of two devices, the transmitter (smartphone) and the receiver (workstation). Each device works with its own sampling clock and thus the transmitter and receiver timings are not synchronized. The main issue with unsynchronized communication is timing the start of a message to ensure no information loss. A suggested solution to this issue is to use a Sliding Window algorithm, as described in [14] and illustrated in Fig. 7. The rationale is scanning the time domain (in the previous and subsequent blocks) to recover the transmitted symbol. The algorithm uses

a buffer of 8192 samples (2 blocks) and a resulting array with a size of 4 elements.

Sliding Window Algorithm

1) $i = 0$
2) Push the new coming block to the end of the buffer
$$\begin{cases} buffer\ [0:4095] = buffer\ [4096:8191] \\ buffer\ [4096:8191] = new\ datablock\ [0:4096] \end{cases}$$
3) Send block of 4096 samples $buffer\ [i:4095+i]$ to the Symbol Detection algorithm
$\rightarrow$ get the result.
$\rightarrow$ save in the results array.
4) Move the analysis window, $i = i + 1024$.
5) Repeat stages 2 and 3 while $i < 4096$.
6) Perform Majority Vote on the resulting array

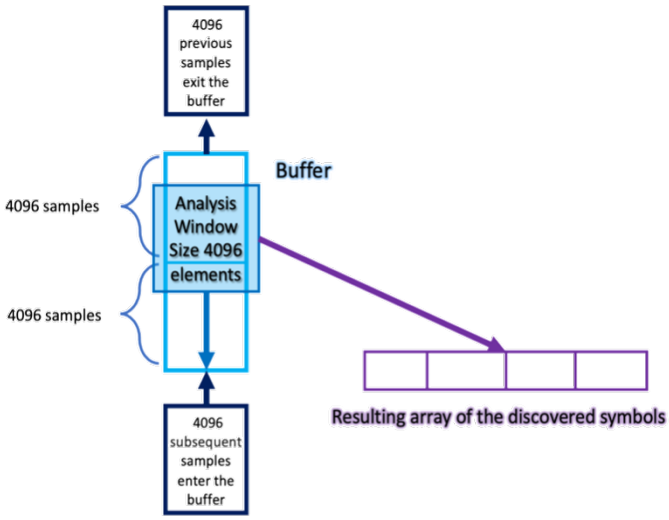

Fig. 7. Massage detection process

D. Majority Vote

The purpose of the Majority Vote is to analyse the results of the Sliding Window and return the result which was obtained most of the times. Hence only if more than half of the results indicate a specific symbol it will be saved as the identified symbol value. [15] describe audio filtering by Sliding Window approach with Majority Vote.

IV. SECURED COMMUNICATION

To eliminate unauthorized unlocking by eavesdropping the transmitted messages, random keys are generated sequentially. Upon the first use, a randomly drawn four-digit password, known to both the smartphone and workstation, is generated. Then, the smartphone's and the workstation's clocks are used for consecutive, synchronous keys generation every minute. The keys are generated using a logic function, whose arguments are the user's unique four-digit password and the current time, representing the hour and minutes in four digits. Every message from the smartphone to the workstation is encrypted using this key. Since the smartphone and the workstations clocks provide the same time, an application running on the workstation may decrypt the received messages. Each received message is decrypted with the current time, the previous minute, and the following minute; this compensates for minor time drifts between the smartphone and the workstation clocks. Unique password and sequential keys per user allow for multiple users to work in a shared space.

V. CONCLUSIONS

The motivation for this work was to design and validate the performance of an acoustic user authentication system. An essential requirement was locking and unlocking the workstation using encrypted acoustic signals from the user's smartphone. We have derived a solution that supports multiple users in the same room. The system is pure-play acoustic since no radio transmission is required. Future work may apply more elaborated message encryption. Increasing the sampling rate to 48 K samples per second shall enable working with acoustic signals beyond the human hearing range.

VI. ACKNOWLEDGMENTS

The authors are grateful to Prof. David Malah, head of the Signal and Image Processing Laboratory (SIPL) at Andrew and Erna Viterbi Faculty of Electrical and Computer Engineering in the Technion – Israel Institute of Technology, to Nimrod Peleg, SIPL's chief engineer, Yair Moshe, SIPL's Senior Engineer and the entire SIPL team for their support, advice, and useful comments.

REFERENCES

[1] Shah, Syed W., and Salil S. Kanhere. "Recent trends in user authentication–a survey." IEEE access 7 (2019): 112505-112519.J.
[2] Eldefrawy, Mohamed Hamdy, Khaled Alghathbar, and Muhammad Khurram Khan. "OTP-based two-factor authentication using mobile phones." 2011 eighth international conference on information technology: new generations. IEEE, 2011.
[3] Wang, Yongge. "Password protected smart card and memory stick authentication against off-line dictionary attacks." IFIP international information security conference. Springer, Berlin, Heidelberg, 2012.
[4] Bhattacharyya, Debnath, et al. "Biometric authentication: A review." International Journal of u-and e-Service, Science and Technology 2.3 (2009): 13-28.
[5] Saleh, Zakaria, and Ahmed Mashhour. "Double Authentication Model using Smartphones to Enhance Student on-Campus Network Access." International Journal of Advanced Computer Science and Applications 9.3 (2018).
[6] Karapanos, Nikolaos, et al. "Sound-Proof: Usable Two-Factor Authentication Based on Ambient Sound." 24th USENIX security symposium (USENIX security 15). 2015.
[7] Cohen, Z., Rodan, A., Eilam, A., & Lifshits, P. (2020, October). Acoustics Based Proximity Detector for Mobile Phones. In 2020 IEEE Workshop on Signal Processing Systems (SiPS) (pp. 1-5). IEEE.
[8] Feferman, G., Blatt, M., & Eilam, A. (2018, December). Indoor Positioning with Unsynchronized Sound Sources. In 2018 IEEE International Conference on the Science of Electrical Engineering in Israel (ICSEE) (pp. 1-4). IEEE.
[9] Dascalu, G., Movshovits, O., & Eilam, A. (2019, November). Pure Play Ultrasonic 3D Positioning System with Unsynchronized Beacons and Receivers. In 2019 IEEE International Conference on Microwaves, Antennas, Communications and Electronic Systems (COMCAS) (pp. 1-6). IEEE.

[10] I. Dagan, G. Binyamin, A. Eilam, "Delivery of QR Codes to Cellular Phones through Data Embedding in Audio", IEEE International Conference on the Science of Electrical Engineering (ICSEE) 2016
[11] Dabran, I., Eilam, A., Menhel, G., Ron, Y., & Shofen, G. (2019, November). Case study: Implementing a Personal Area Network MAC Protocol for Inaudible Sound Waves. In 2019 IEEE International Conference on Microwaves, Antennas, Communications and Electronic Systems (COMCAS) (pp. 1-5). IEEE.
[12] Harden, S., ScottPlot, 2020, https://github.com/swharden/ScottPlot, retrieved December 2021.
[13] Waggener, B., Waggener, W. N., & Waggener, W. M. (1995). Pulse code modulation techniques. Springer Science & Business Media. p.206
[14] Lee, Chang-Hung, Cheng-Ru Lin, and Ming-Syan Chen. "Sliding-window filtering: an efficient algorithm for incremental mining." Proceedings of the tenth international conference on Information and knowledge management. 2001.
[15] Geiger, Jürgen T., Björn Schuller, and Gerhard Rigoll. "Large-scale audio feature extraction and SVM for acoustic scene classification."