Coresident Evil: Noisy Vibrational Pairing in the Face of Co-located Acoustic Eavesdropping

S Abhishek Anand
University of Alabama at Birmingham
anandab@uab.edu

Nitesh Saxena
University of Alabama at Birmingham
saxena@uab.edu

ABSTRACT

An interesting approach to pairing devices involves the use of a vibrational channel, over which the keying material (e.g., a short PIN) is sent. This approach is efficient (only a unidirectional transfer of PIN is needed) and simple (the sending device requires a vibration motor and receiving device requires an accelerometer). However, it has been shown to be susceptible to *acoustic emanations* usually produced by the vibration motor. Recent research introduced a mechanism to defeat these attacks by attempting to *mask* the acoustic leakage with deliberate acoustic noises. In this paper, we pursue a systematic investigation of the security of such a "noisy vibrational pairing" mechanism in a *strong yet realistic* adversarial model where the eavesdropper is *co-located* with the victim device(s).

Our contributions are two-fold. *First*, we show that existing noisy vibrational pairing mechanisms – based on *white noise* as the masking signal – are vulnerable against a co-located eavesdropper (although they may defeat a distant eavesdropper). We build our attack based on standard signal processing and noise filtering techniques, and show that it can result in a complete compromise of pairing security. *Second*, we propose a defense that bolsters the masking signal with *low-frequency audio tones*. We present and address the challenges associated with producing such low-frequency sounds with current commodity hardware. We show that our defensive approach can not only resist our above attack but is also robust to more sophisticated, noise filtering and source separation methods when applicable. We also establish that the insertion of low-frequency sounds does not affect the receiving device's capability to sense the vibrations generated by the sending device. The suggested defense may therefore be used to enhance the security of noisy vibrational pairing without affecting its performance on a wide variety of devices.

ACM Reference format:
S Abhishek Anand and Nitesh Saxena. 2017. Coresident Evil: Noisy Vibrational Pairing in the Face of Co-located Acoustic Eavesdropping. In *Proceedings of WiSec '17 , Boston, MA, USA, July 18-20, 2017,* 11 pages.
DOI: 10.1145/3098243.3098256

1 INTRODUCTION

The wireless radio communication, e.g., Bluetooth, WiFi or RFID/NFC, is easy to eavesdrop and manipulate, and therefore a fundamental security goal is to secure this communication channel. "Pairing" refers to the operation of establishing secure communication between two wireless devices, while resisting against eavesdropping and man-in-the-middle attacks. Pairing is challenging to achieve in practice due to the lack of a global infrastructure enabling devices to share an on- or off-line trusted third party, a certification authority, a PKI or any pre-configured secrets.

A well-studied pairing methodology [13] makes use of an auxiliary channel governed by the users operating the devices. Such auxiliary channels include audio, visual, and vibrational channels. Unlike the radio communication channels, auxiliary channels are "human-perceptible", i.e., the underlying transmission/reception can be perceived by one or more of human senses.

An interesting approach to pairing devices involves the use of a vibrational channel, over which the keying material (e.g., a short PIN) is sent. This PIN may then used as an input to a password-based authenticated key agreement (PAKA) protocol [3] to achieve pairing. This pairing approach is efficient, since only a unidirectional transfer of PIN is needed, and simple, since the sending device requires a vibration motor and receiving device requires a vibration sensor or an accelerometer. It is therefore suitable for many pairing contexts, such as pairing between a phone and an RFID card (e.g., asset tracking tags [8, 19]), two phones, and a phone and a point-of-sale terminal, to name a few.

A notable instance of such vibrational pairing is a system called PIN-Vibra [18]. It uses an automated vibrational channel to pair a personal RFID tag with a mobile phone. The phone generates a PIN and transmits it to (an accelerometer-equipped) tag through its vibrations, while the user presses the phone against the tag. The same channel is later used by the phone to authenticate to, or activate, the RFID tag and prevent unauthorized reading of the tag's information [18]. A similar general approach can be used on a multitude of other pairing/authentication settings as long as the sending device has a vibration-generation capability and the receiving device has a vibration-sensing capability.

However, vibrational pairing in general, and PIN-Vibra in particular, have been subject to acoustic eavesdropping attacks and shown to be vulnerable [9]. Specifically, [9] demonstrated highly accurate eavesdropping attacks on PIN-Vibra, from several centimeters to few meters away, in which the acoustic signals are a by-product of the vibration as a side channel. These attacks thus serve to call the security of vibrational pairing into question.

Subsequently, recent research [14, 17] introduced a novel strategy to defeat these attacks by attempting to *mask the acoustic leakage* associated with vibration with the use of deliberate acoustic

WiSec '17 , Boston, MA, USA
© 2017 ACM. 978-1-4503-5084-6/17/07…$15.00
DOI: 10.1145/3098243.3098256

noises.[1] The key principle is to selectively jam the acoustic channel leakage by having the sending device produce deliberate sounds that would mask the sounds produced by the vibration motor. An important advantage of such a jamming approach is that the receiving device remains *oblivious* to the inserted acoustic noises as it only needs to sense the vibrations (through accelerometer), *not the sounds*, to decode the transmitted PIN. This is in contrast to selective jamming of key material or arbitrary data transmission over the audio channel [23] itself, where the jamming signal may also make it error-prone for the receiving device to decode the key or data signal.

1.1 Our Contributions

In this paper, we pursue a systematic investigation of the security of the above-mentioned *noisy vibrational pairing* methodology [14, 17] under a *strong yet realistic* adversarial model, where the eavesdropper is *co-located* with the victim device(s) (i.e., at a distance of essentially 0 centimeters from the victimized pairing device(s)). A compromised microphone of the victim device is a prime example of such a co-located acoustic eavesdropper. We show that noisy vibrational pairing schemes, resistant to close distance attackers, are completely insecure against co-located attackers. Henceforth, we propose a new scheme that is resistant against powerful adversaries such as co-located attackers. Our main contributions underlying this work are two-fold:

(1) ***Attacks Against Noisy Vibrational Pairing***: We show that existing noisy vibrational pairing mechanisms [14, 17] — based on *white noise* as the masking signal — are vulnerable against a co-located eavesdropper (although they may defeat a distant eavesdropper). We build our attack based on standard signal processing and noise filtering techniques, and show that it can result in a complete compromise of pairing security. The key insight in the vulnerability of white noise based vibrational pairing against a co-located eavesdropping is that the vibrational motor creates *low-frequency sounds* in the frequency band 50Hz-250Hz that the white noise can not cloak, thereby exposing the keying material to the attacker residing right at the source of these sounds.

(2) ***New Defense based on Bolstered Noises***: We introduce a defense that carefully bolsters the white noise based masking signal with *low-frequency (50Hz-250Hz) audio tones*. We present and address the challenges associated with producing such low-frequency sounds with current commodity hardware. We show that our defensive approach can not only resist our attack above but is also robust to more sophisticated, noise filtering and source separation methods when applicable. We also establish that the insertion of low-frequency sounds does not affect the receiving device's capability to sense the vibrations generated by the sending device. The suggested defense may therefore be used to enhance the security of noisy vibrational pairing without affecting its overall performance on a wide variety of devices.

1.2 Outline of the Paper

The rest of this paper is organized as follows. Section 2 details the background for the work done in this paper and introduce the co-located eavesdropping threat model. Section 3 demonstrates the vulnerability of the noisy vibrational pairing mechanisms in the presence of co-located adversary. Next, Section 4 proposes a novel defense mechanism that improves upon the security of noisy vibration pairing even under co-located eavesdroppers. Section 5 follows it up with the summary of the results and possible future work to be done. Finally, Section 6 provides the take home message of our work.

2 BACKGROUND AND THREAT MODEL

Prior Work on Vibrational Pairing: The use of auxiliary channels for establishing secure communications has been well-studied over recent years with the increase in the use of smartphones and constrained devices, such as RFID cards and POS terminals. The auxiliary channels, often referred to as Out-of-Band (OOB) channels [13], have the promising characteristic of providing a novel decentralized mechanism of security unlike the traditional security approaches based on centrialized entities (e.g., Public-Key Infrastructure or Key Distribution Centers). As a result, there has been a rise in a number of communication, pairing and authentication protocols that choose an auxiliary channel for securing the communicated information.

PIN-Vibra [18] proposed a pairing protocol that used vibrations of a sending device (a phone) to communicate a short PIN to the receiving device (an RFID card) for pairing. This PIN is later to be used for unlocking the card using phone, again over the same vibrational channel. The vibrational communication was based on a simple "ON-OFF" encoding where the set bits were represented as vibrations of a predetermined time period and the unset bits were represented as silence. However, this scheme was shown to be vulnerable against an acoustic eavesdropping attack in [9]. The vulnerability of the scheme resulted from acoustic side channel leakage due to the vibrations of the PIN-transmitting device.

To address such side channel attacks, the work presented in [14] proposed a "noisy" vibrational key exchange protocol that utilized a band-limited Gaussian white noise to mask the acoustic side channel leakage during the key exchange procedure. They tested the scheme against an acoustic eavesdropping adversary at a distance of 30cm and against independent component analysis (ICA) [4] from a distance of 1m, and showed it to be robust to such an adversary.

Further, in a system called Ripple [17], authors introduced a vibration based communication scheme that created an *anti-noise* signal to cancel out the acoustic side channel leakage. The canceling signal was boosted with *pseudorandom* white noise to help mask the residual data bits that were susceptible to an acoustic side channel attack while the *anti-noise* signal was being generated. They tested the canceling signal's effectiveness with varying distance by measuring the residual signal's power. They also evaluated the jamming signal (white noise) at a very close distance (few centimeters away) by calculating the correlation coefficient of the actual signal with the jammed signal that indicated a decrease in correlation with increase in the power of the jamming signal.

[1] The work of [17] focuses on arbitrary yet potentially sensitive communications using the vibrational channel, not pairing.

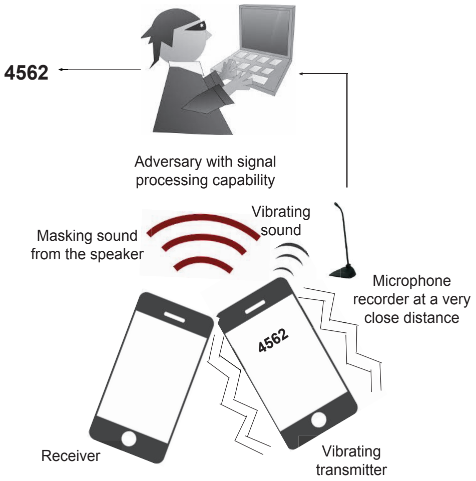

4562
Adversary with signal processing capability
Vibrating sound
Masking sound from the speaker
Microphone recorder at a very close distance
4562
Receiver
Vibrating transmitter

Figure 1: Attack setup

Existing and Proposed Threat Models: Prior work on non-noisy vibrational pairing [18] and noisy vibrational pairing ([2, 14, 17]) follow a threat model that placed the eavesdropping adversary nearby but at some distance from the victim device (i.e., to the device(s) being paired). In this paper, we strengthen this threat model. In particular, we assume the adversary to be co-located with the transmitter (at a distance 0cm, without touching). We further extend and strengthen the model to consider multiple co-located adversaries that may even be co-resident with the victim device.

Another advantage offered by a co-located adversary is more accurate eavesdropping as vibration sounds are susceptible to inbuilt noise cancellation mechanisms by off the shelf PC microphones (that have been used in prior attacks [14, 17, 18] as well as in our analysis in this paper) when the eavesdropping is done at a distance from the vibrating device. Such a strong threat model could be realized by an adversary by creating a communication setup where the device transmitting the vibrations is equipped with a microphone (like a smartphone) that can provide on-board recording capability to the adversary. In addition, the adversary could have implanted a tiny listening bug in either of the devices for eavesdropping. This would require one-time access to the compromised device by the adversary constituting a *lunch-time* attack.

Although strong, the model is mostly realistic. For instance, it could be realized when the victim is trying to pair the device while on a phone or VoIP call, in which case the entity at the other end of the call may surreptitiously eavesdrop over the vibrational sounds. Another scenario for this threat model involves a malicious application trying to access the microphone of the involved devices for the purpose of eavesdropping. Since both the transmitter and the receiver need to be in physical contact with each other from transmitting data through vibrations, unauthorized access to either device's microphone would qualify as the threat model for a co-located eavesdropping adversary. Compromising the devices by attaching a small listening device can also be achieved through products available in the market [11]. We emphasize that it is important to consider stronger adversarial models and design security mechanisms that can abide by such models (a primary goal of our work in this paper).

An important point to note is that while co-located vibrational eavesdropping (e.g., through accelerometer readings) is another form of a viable co-located attack, in this work, our focus is on co-located acoustic eavesdropping since the audio channel opens up many different avenues for such eavesdropping (e.g., via an on-board malware with access to microphone or even a remote eavesdropper over a voice-based call, as mentioned above).

Finally, in line with the threat model of prior vibrational pairing schemes [14, 17, 18], we also assume that the attacker does not need to decode the vibrations online. The eavesdropped vibration sounds can be processed offline using signal processing algorithms. Also, the recording is assumed to be done in a noise-free environment except for the sounds emanating from the communicating devices.

3 ATTACKING EXISTING VIBRATIONAL PAIRING SCHEMES

In this section, we will demonstrate an attack on noisy defense mechanisms under the threat model with a co-located adversary as described in Section 2. We borrow the idea of using white noise as the masking signal that is used to mitigate acoustic eavesdropping attacks as proposed in [14] and [17]. We will also reuse the basic principles from the acoustic eavesdropping attack on the PIN-Vibra scheme as detailed by Halevi et al.[9].

3.1 Overview of the Attack

An acoustic eavesdropping attack in our studies involves an adversary eavesdropping on the audio leakage from the two devices in communication with each other. The audio leakage from the communication refers to the sound emanating from the vibration of the device acting as a transmitter.

The vibrations in a phone is caused by a tiny electric motor residing inside the device. The motor has a weight mounted off-center on a shaft and when the motor spins, the off-center mounted weight produces the vibrations. When the transmitter in the PIN-Vibra scheme produces vibrations in a unique pattern specified by the scheme, the resulting sound is recorded by the adversary.

The recorded sound is processed by the adversary in the frequency domain using signal processing algorithms. Various features exist that can characterize unique patterns in an audio signal like Fast Fourier Transformation (FFT) and Mel-Frequency Cepstral Coefficient (MFCC). These features are used to identify the timings of the vibrations thereby revealing the pattern of the vibrations that lead to the data being transferred. We will begin with the recreation of an attack on a vibration based pairing scheme such as PIN-Vibra and then evaluate the effectiveness of the attack against the proposed defense measures in [14] and [17].

3.2 Attack Experiment Setup

In our experiments, we implement the PIN encoding algorithm as detailed in [18]. The algorithm converts a 4-digit PIN to its equivalent 14-bit binary string and adds a preamble "110" to the binary string. For example, "4562" would be converted to "11001000111010010" after the addition of the preamble (each digit is converted to its four bit binary equivalent). Each bit that is set

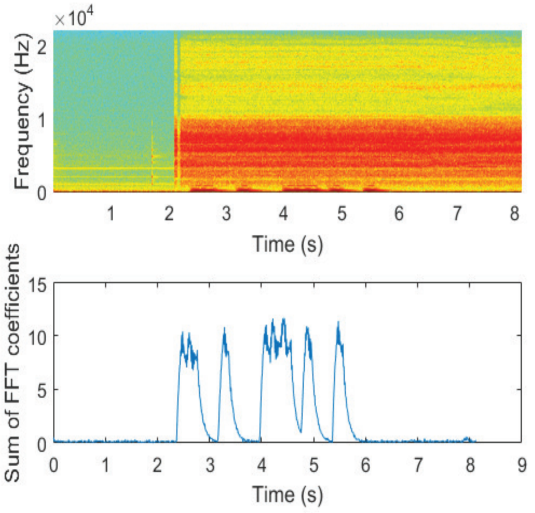


Figure 2: Frequency spectrogram for audio recorded at a distance 0cm. Intensity in the top graph is proportional to energy in the frequency band. Sum of the FFT coefficients indicates the estimated energy at the time instant.

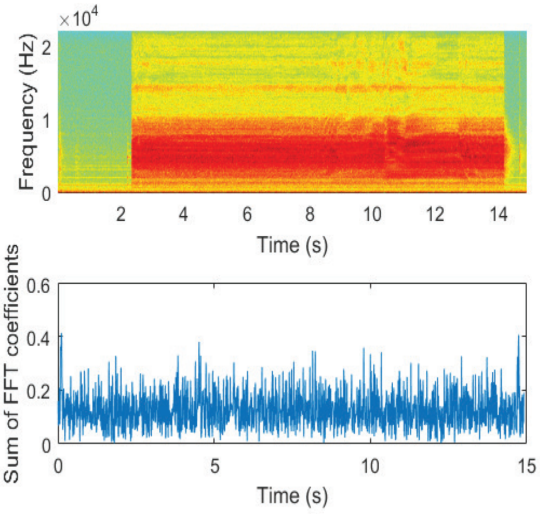


Figure 3: Frequency spectrogram for audio recorded at a distance 10cm. Intensity in the top graph is proportional to energy in the frequency band. Sum of the FFT coefficients indicates the estimated energy at the time instant.

(i.e. "1") is encoded as a vibration lasting 200 ms while an unset bit (i.e. "0") is encoded a silence lasting 200 ms. The total duration for the PIN transmission is (14 + 3) × 200 ms = 3.4 seconds.

In our attack setup (Figure 1), we used a set of Motorola Droid™ X2 smartphones as the communicating devices, with one acting as the transmitter and the other as the receiver. To record the audio generated from the vibrations, we used a Dynex PC microphone and Matlab's signal processing tools for processing of the recorded audio. The microphone is placed almost touching to the vibrating device in order to record the vibration sounds at the closest possible distance for emulating a co-located adversary. The on-board microphone can also be used for this purpose as per detailed in our threat model described in Section 2.

To complete the noisy vibrational pairing setup, we also implemented the defense measures as proposed in [14]. The defense measure in [14] utilized band limited Gaussian white noise that lies in the same frequency range as the audio leakage from the vibrations. It is used as the masking sound to hide the acoustic leakage from the vibrations. The masking sound was generated as the transmitting device starts the communication with the receiving device for pairing.

On the side of the attacker that has the capability to process the eavesdropped audio signal offline, we also used the "Noise Reduction" feature of the audio processing tool *Audacity* to filter out the noise and reveal part or whole of the audio leakage. This feature allows the selection of a small portion of the audio signal consisting of the noise only to build a noise profile that is then filtered from the whole audio signal.

3.3 Effectiveness of the Attack

As per the threat model detailed in Section 2, we recorded vibration sounds superimposed by the masking sound at a distance of 0cm. We also recorded vibration sounds at a distance of 10cm for comparing it with the co-located adversary scenario. Figure 2 and Figure 3 represent the frequency spectrum of the eavesdropped signal at distances 0cm (as per our threat model) and 10cm (similar to [14]). The frequency spectrum revealed that masking sound may be able to hide the audio leakage due to the vibrations from an adversary eavesdropping at a distance. However, for a co-located adversary, the masking sound was unable to hide the audio leakage resulting from the vibrations at the lower frequency range of 50Hz-250Hz.

In the lower graph presented in Figure 2 depicting the sum of FFT coefficients in the frequency range 50Hz-250Hz plotted against time, a suitable threshold was chosen to identify the beginning of a vibration. Then we calculated the mean sum of FFT coefficients over time intervals of 0.2 seconds. The mean sum of FFT coefficients is used as a representative of the energy contained in that time interval. If the energy was above the threshold, the interval was believed to have contained a vibration sound and is therefore labeled as "1" else it was deemed to be a silence period and labeled "0" accordingly. This approach led us to to the decoded binary string "11001000111010010" i.e. "4562" for the audio captured in Figure 2.

Roy et al. in their work Ripple [17] used an *anti noise* signal to cancel out the vibration sounds for minimizing the audio leakage. The *anti noise* signal initially started with a sampling frequency that put the fundamental frequency of the *anti noise* signal higher than the frequency of the vibration sounds. However, there existed a phase difference between the *anti noise* signal and the vibration sounds which led to the existence of some residual vibration sounds till the time the phase of the *anti noise* signal matched with the vibration sounds for canceling out the vibration sounds.

While the *anti noise* signal may be able to cancel out the vibration sounds, some residual vibration sounds still existed at the beginning of the transmission. These residual vibration sounds have the potential of leaking out bits of information to an eavesdropping adversary. For this purpose, Ripple [17] used *pseudorandom noise* as a jamming signal to mask the residual vibration sounds. Yet, as we have demonstrated in our experiments, white noise (same as

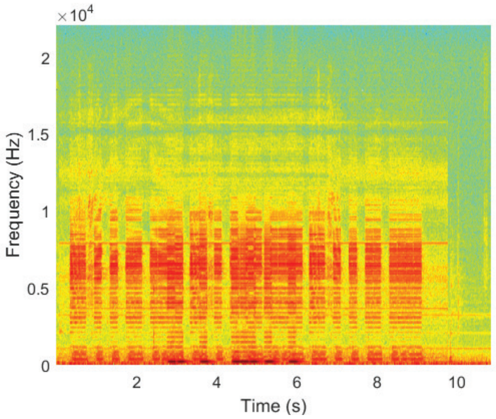


Figure 4: Frequency spectrogram of real vibration sounds mixed with fake vibrations sounds. Intensity in the graph is proportional to energy in the frequency band.

pseudorandom noise) alone was inadequate to mask the vibration sounds. Hence, we needed to explore further options to bolster the white noise masking signal in order to mitigate an eavesdropping attack from a co-located adversary.

4 NOVEL DEFENSE BASED ON LOW-FREQUENCY NOISE

In this section, we would try to strengthen the white noise to mask the vibration sounds at the lower frequency band of 50Hz-250Hz against a co-located adversary. We would also detail the challenges faced in the implementation of the proposed masking signal with the attack setup described in Section 3. Lastly, we would evaluate the efficiency of the proposed masking signal against sophisticated attacks and its effect on vibrational sensors of the receiving device.

4.1 Masking with White Noise and Low-Frequency Tones

As observed in the previous section, the white noise alone proved ineffective at masking low-frequency vibration sounds. In order to overcome this shortcoming, we considered other signals that could prove effective at masking audio leakage at low frequencies. We tried to add acoustically similar vibration sounds to the white noise to confuse the adversary between the real vibration sounds and the pre-recorded vibration sounds that were played along with the white noise.

We recorded vibrations of the Droid X2 phone from our setup with the inbuilt microphone, with the phone placed on a glass surface. Ripple [17] indicated a glass plate as producing the strongest side channel leakage when the vibrating device is placed on it. This motivated us to record the vibrations on a glass surface (henceforth referred as fake vibration sounds) as stronger the vibration sounds, stronger would be their recording, producing similar effect as actual vibration sounds when played back during the pairing process.

After recording fake vibration sounds, we played it on the device during the pairing process to gauze the similarity of the fake vibration sounds with the audio leakage of real vibration on the frequency spectrum. Figure 4 shows the resulting frequency spectrum. As it is clear from the spectrum that while fake vibration sounds matched the audio leakage from the real vibration to most extent, they lacked the required low-frequency components contained in the audio leakage. Hence, they offered no better protection over the white noise masking signal and neither did the combination of both the white noise and fake vibration sounds (due to their inefficiency at lower frequency range).

The lack of low-frequency components in the fake vibration signal implored us to explore software based solutions for improving the quality of the audio playback to recover the desired low-frequency response. Since, our device ran on Android platform, we utilized audio effects and controls offered by the platform via AudioTrack API (Application Programming Interface) but no improvements were observed and boosting the signal only resulted in clipping of the audio signal, a phenomenon explained below:

Non Linearity: This phenomenon is widely encountered in electrical circuits e.g. an amplifier, where the generated output signal strength is not directly proportional to the input signal strength. The ratio of the output signal strength to input signal strength, also called "gain" depends upon the input signal strength.

Clipping: This phenomenon occurs due to distortion of the waveform when an amplifier is over-driven by trying to produce an output signal, the strength of which is beyond the specified limits of the amplifier. This causes the signal to be clipped at the limits resulting in a distorted wave. A side effect of clipping is the introduction of harmonics of the signal at higher frequencies.

The next choice in our experiments was to generate tones in the desired frequency range and add them to the white noise to obfuscate the audio leakage from the vibration sounds. For this purpose, we used the Tone Generator function in Audacity™ along with the Noise generator and used "mix and render" functionality to produce the combined signal that is a mixture of white noise and a sinusoidal tone of 150Hz. The resulting observations are shown in Figure 5. As Figure 5 shows, there was no masking at lower frequency band despite the introduction of a low-frequency (150Hz) tone. In particular, there was no presence of the tone at the intended frequency level. This behavior was similar to that of fake vibration sounds which also lacked the low-frequency components present in the audio leakage.

Since, software based solutions offered no improvement, the reasonable explanation for this behavior was the inefficiency of speaker to correctly reproduce sounds at low frequencies. We further investigated the issue by trying to reproduce various low-frequency sounds on two devices: Motorola Droid™ X2 and LG™ G4 smartphones. Droid X2 is an old smartphone, first released in 2011 whereas G4 is one of the latest devices announced in 2015.

During our attempts to reproduce low-frequency tones while testing the speakers of both old (Droid X2) and new (LG G4) devices, we re-encountered the non-linear behavior of the speaker response. The output audio signal for low-frequency tones was very low, barely registering on the microphone. Any attempts to increase the gain would inadvertently result in clipping of the signal producing unwanted harmonics at higher frequency levels with no improvement at the intended low frequency. We expected better performance from LG G4 smartphone featuring an improved speaker but the results were only slightly better (Figure 6). The speaker was barely an improvement over the Droid X2 speaker

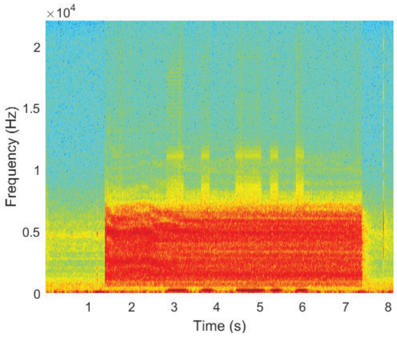


Figure 5: Frequency spectrogram for the vibrations recorded by Droid X2. Intensity in the graph is proportional to energy in the frequency band.

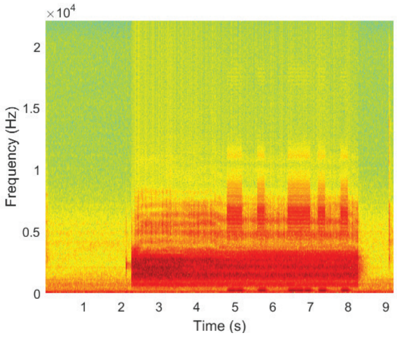


Figure 6: Frequency spectrogram for the vibrations recorded by LG G4. Intensity in the graph is proportional to energy in the frequency band.

suffering from the same drawbacks of non linearity and clipping aspects of the system. Since, the inbuilt speakers of the smartphones did not fulfill our purpose, we turned to other setups where we could obtain better speaker quality for improved sound reproduction.

4.2 Modified Setting

While smartphone speakers may be limited in their capacity to reproduce low frequency sounds (sub 300Hz), we can try to boost their capabilities by complementing them with better hardware. Such an approach has been used in the industry by introducing a case like system with built-in speakers and/or a separate audio engine to boost the quality of smartphone's speaker [1, 16, 22]. While [1] and [22] are geared towards iPhones, [16] is offered as an accessory for Moto Z™ family of phones. These accessories can be put on as a case on the phone (Figure 18).

We simulated the concept by taping a small portable speaker to our device and playing the sound through it. This setup also emulated the scenario where the receiving device could have an inbuilt powerful speaker like a payment terminal or high end media devices e.g. a smart television.

For our experiments, we used three different portable speakers Altec™[15], Sony SRS-XB2™[20] and JBL™[12]. The frequency specifications for the tested speakers are presented in Table 1. In order to test the effectiveness of speakers in producing low frequency sounds, we played a 150Hz sinusoidal tone through each speaker and observed the recorded signal in frequency domain (Figure 7). The tone for each speaker is denoted by a yellow band present at the lower end of the spectrum. The data cursors for each of the speakers demonstrate that a tone of 150HZ produces response in the range 86Hz-344Hz. Since a tone below 150Hz distorted the response from Altec speaker, we used 150Hz tone in our next stage of experiment.

In our experiment, the speaker was connected (via an audio cable or bluetooth) and attached to the smartphone. Rest of the experimental setup was similar to our previous attack experiment (Section 3) under similar threat model (Section 2). The masking sound that was used to obfuscate audio leakage was a mix of white noise and a low-frequency tone (150Hz). We generated two separate tracks containing white noise (generated using noise generator functionality in Audacity) and 150 Hz tone (generated using tone generator functionality in Audacity) which were then mixed and rendered to form a new track. The low-frequency tone helped in masking the low frequencies of the audio leakage while the white noise spread across rest of the frequency spectrum masked the audio leakage at higher frequencies.

During our experiments, we observed the effectiveness of masking signal against co-located adversary. We also observed the effect of sound level of the masking signal in the event of clipping. This was of particular importance as we were operating around the lowest frequency response for some of the tested speakers.

Table 1: Frequency response for tested speakers

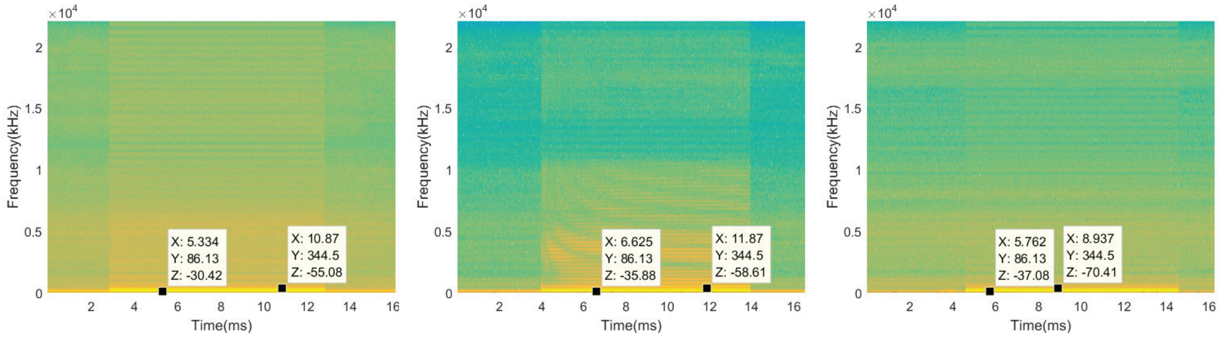
Speaker	Frequency Response (in Hz)
Altec Lansing Mini H2O Speaker	Not specified
JBL Clip Portable Bluetooth Speaker	160-20,000
Sony SRS-XB2 Speaker	20-20,000

The results for the portable speaker are shown in Figure 8, 9 and 12. The frequency spectrum did not show the presence of audio leakage resulting from vibrations, particularly at low frequencies (50Hz-250Hz). The graph of the sum of FFT coefficients vs time showed that the quality of audio leakage degraded to an extent that it became very hard to choose a suitable threshold to determine a constant period of vibration. While the spikes in the graph may indicate towards presence of vibration, the resulting pattern could not be decoded into a valid PIN making the detection infeasible.

This observation showed that external portable speaker had the required sound reproduction quality that was found lacking on the inbuilt smartphone speakers. We also measured the sound level of the masking signals via a sound level measurement application for Android phone and recorded the sound level at a distance of 10cm. We observed that the optimal sound level for producing low-frequency sound of an amplitude sufficient to mask the audio leakage from vibration sounds was around 58 decibels. This sound level is approximately equal that of conversational speech and thus not considered harmful to the human ear.

4.3 Security under Sophisticated Attacks

In this section, we will evaluate the masking effectiveness of the white noise boosted with low-frequency tones against some attack vectors that may be used by the adversary for a more sophisticated analysis of the eavesdropped signal. The attack techniques that we



(a) **JBL speaker**

(b) **Altec speaker**

(c) **Sony SRS-XB2 speaker**

Figure 7: Frequency response for tested speakers for 150Hz tone (data cursor indicates X as time instant, Y as frequency in Hz and Z as power/energy estimate)

will discuss here would involve noise filtering and source separation techniques.

4.3.1 Noise Filtering: The defense mechanisms that we studied till now, relied on deliberate injection of a masking signal in the environment for obfuscating the audio leakage during vibrational pairing. From the adversary's point of view, the masking signal was the noise accompanying the audio leakage (that was to be acquired and decoded). Hence, the adversary could try to remove or suppress the noise using noise removal algorithms.

As per our threat model in Section 2, a co-located adversary had the ability to process the eavesdropped signal offline and could try to recover the information from the audio leakage. We repeated the attack experiment (Section 3) according to our threat model (Section 2) with the masking signal comprising of white noise with a low-frequency tone of 150Hz that was capable of masking the audio leakage from the vibrations at the low-frequency bands as detailed previously.

To evaluate the efficiency of our masking signal against noise filtering, we applied the noise reduction technique called "spectral noise gating" to the eavesdropped signal. This technique is used in most of the audio processing software tools like Audacity™. We chose a short sample from the eavesdropped signal as the noise profile and applied it to the signal to be removed as noise. This process could be repeated multiple times until satisfactory results were obtained. The results are shown in Figure 10, Figure 11 and Figure 13. Figure 11 and Figure 13 showed no indication of the audio leakage from the vibrations in part of the frequency spectrum. This affirmed the effectiveness of masking signal at hiding the audio leakage from the vibrations. Figure 10 showed some residual leakage after noise filtering that we were unable to confirm as a part of the transmitted PIN. Since we had no knowledge about the frequency response for Altec speakers, we attribute the residual leakage result to audio distortion at lower frequency level in this speaker. Both JBL speaker and Sony speakers are advertised as having a good bass performance and the Sony speaker has an additional functionality to boost bass performance that helps in preventing audio distortion at lower frequency levels. This analysis serves to highlight the requirement of the proposed defense method to possess the ability of producing bass rich (low frequency) sounds for effectively masking vibration sounds.

4.3.2 Source Separation Analysis: In our modified defense design, we introduced an external speaker with the existing setup to produce the masking sound. In case of an on-board speaker existing on one of the devices, we assumed the location of the speaker to be close enough to the vibration device (as the devices are touching each other) to be of any significance against source separation attacks. However, with an external speaker (even at 0cm), there exists a chance that an adversary may be able to separate the masking signal from the audio leakage from the vibrations as the two sounds were generated from two different sources.

There exist two statistical techniques that allow us to separate multiple unknown signal sources by analyzing several recordings of the mixed signal taken at the same time. The process to separate multiple unknown sources from a set of mixture of the source signals is referred as Blind Source Separation (BSS) [5]. Two common methods that are used to implement BSS are Principal Component Analysis (PCA) and Independent Component Analysis (ICA) [4]. PCA is helpful when the data consists of Gaussian variables, however ICA is useful for non Gaussian data and utilizes higher order statistics for source separation. In our setup, the masking signal makes the raw data very noisy hence ICA being the more powerful tool is applied for source separation.

We used the same attack setup with the modified defense as in Section 4.2 and added another microphone approximately equidistant (but still at a distance of 0cm) from the setup as the original eavesdropping microphone to simulate multiple colocated adversary. We used the FastICA algorithm [10] for source separation and plotted the resulting sources on the frequency spectrum. As Figure 14 and Figure 15 show, the two sources as per approximated by the FastICA algorithm did not seem to contain the vibrations in the low-frequency range of 50Hz-250Hz. Both the sound source seemed to be variations of white noise which indicated that the audio leakage from the vibrations of the transmitting device was completely masked and could not be detected by FastICA. Thus our defense mechanism seems resistant against an adversary equipped with BSS techniques.

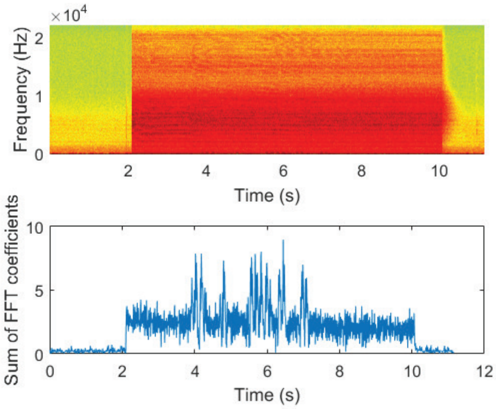


Figure 8: Frequency spectrogram of the audio signal in presence of masking sound from Altec speaker. Intensity in the graph is proportional to energy in the frequency band. Sum of the FFT coefficients indicates the estimated energy at the time instant.

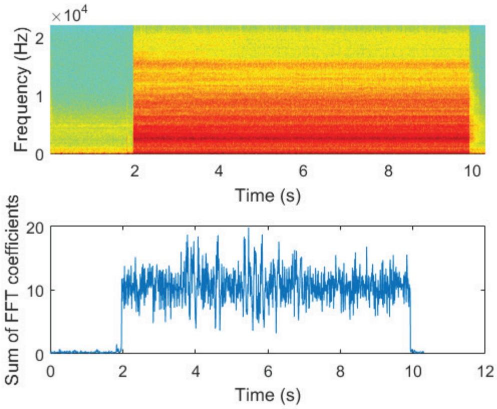


Figure 9: Frequency spectrogram of the audio signal in presence of masking sound from JBL speaker. Intensity in the graph is proportional to energy in the frequency band. Sum of the FFT coefficients indicates the estimated energy at the time instant.

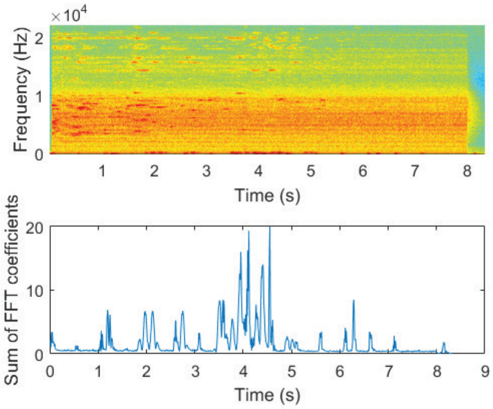


Figure 10: Frequency spectrogram of the audio signal in presence of masking sound from Altec speaker after noise filtering. Intensity in the graph is proportional to energy in the frequency band. Sum of the FFT coefficients indicates the estimated energy at the time instant.

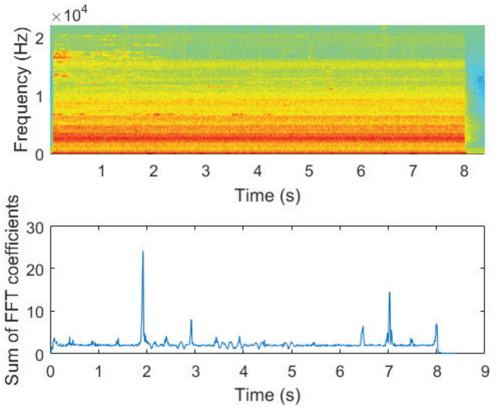


Figure 11: Frequency spectrogram of the audio signal in presence of masking sound from JBL speaker after noise filtering. Intensity in the graph is proportional to energy in the frequency band. Sum of the FFT coefficients indicates the estimated energy at the time instant.

4.4 Effect on Vibrational Sensing

In a pairing mechanism based on vibrations like PIN-Vibra [18], the receiving device uses its accelerometer to read the vibrations and then decode it based on the protocol. The masking signal, proposed in this work, comprised of a low-frequency tone along with the white noise. The bass effect of the low-frequency tone has a tendency to produce deep rumbling sounds that have the capability of producing faint vibrations in the speaker. This effect may negatively affect the accelerometer readings of the receiving device that could have an negative impact on the accuracy of the vibrational decoding and thereby the success of the pairing process.

In order to test the impact of the masking signal on the ability of the receiving device to decode the vibrations correctly, we collected accelerometer readings in the background on the receiving device during the vibrational pairing in the presence of masking signal (as proposed in Section 4.2). We plotted the accelerometer readings (a scalar component derived from the three axes of the accelerometer) against time and the results are shown in Figure 16 and Figure 17. If we compare the two figures, we do not see any noticeable effect of the masking signal on the accelerometer readings. The only evidence of the masking signal is the slight wiggle of the baseline accelerometer reading during the 8 seconds of time including the 3.4 seconds of the PIN transmission via vibrations. Thus, we conclude that the masking signal does not deteriorate the decoding accuracy of the receiving device while enhancing the security by obfuscating the audio leakage resulting from the vibrations at the same time.

5 DISCUSSION AND FUTURE WORK

Summary of Results: We showed the vulnerability of the noisy vibrational pairing mechanism against a co-located adversary that can accurately determine the pairing key by eavesdropping on the audio leakage emanating from the vibrating device. In particular, we determined that white noise alone is insufficient in masking the audio leakage of the vibrations from a co-located eavesdropping adversary at lower frequency range of 50Hz-250Hz.

We then showed that the inadequacy of white noise in masking the audio leakage at low frequency range could be remedied by

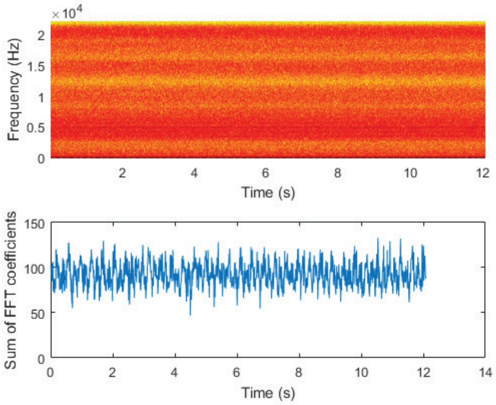


Figure 12: Frequency spectrogram of the audio signal in presence of masking sound from Sony SRS-XB2 speaker. Intensity in the graph is proportional to energy in the frequency band. Sum of the FFT coefficients indicates the estimated energy at the time instant.

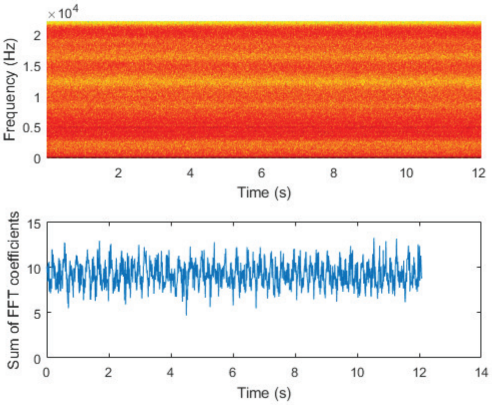


Figure 13: Frequency spectrogram of the audio signal in presence of masking sound from Sony SRS-XB2 speaker after noise filtering. Intensity in the graph is proportional to energy in the frequency band. Sum of the FFT coefficients indicates the estimated energy at the time instant.

adding a low-frequency sound like a sinusoidal tone of appropriate frequency (150Hz in our work). We also found out that it is not possible to generate the required low frequency sounds in all types of devices where the frequency response of the inbuilt speaker may not be enough for our requirements. We offered a solution by allowing a more powerful, yet low-cost, external speaker co-located with the device in our setup that would generate the required masking signal. We further showed that this defense design had no side effects on the decoding ability of the receiving device (if the speaker is co-located with the receiver) and the defense mechanism adequately serves its purpose.

We also tested our proposed defense mechanism against noise filtering and blind source separation attacks. The results suggested that noise filtering is ineffective in removing the masking signal from the audio leakage of vibrations. Similar results were obtained against blind source separation attacks (ICA), further establishing the extent of security offered by the proposed defense mechanism. Based on our tested speakers and the speakers inbuilt in current generation of smartphones, we find that there exists a significant gap in performance of an external speaker and speakers generally found in smartphones. Smartphone speakers are not built to deliver a bass enriched sound as this may require additional hardware making the phone bulky. Solutions like [1], [22], and [16] are designed to overcome this deficiency by adding an external speaker designed as a cover for the smartphones. JBL Soundboost speaker, as a part of moto mods [16], boasts of a frequency response range of 200Hz-20kHz. While this specification still falls short of the frequency response range for the JBL Clip speaker [12] and Sony SRS-XB2 speaker [20] as shown in Table 1, we believe that it is not far-fetched to assume that such a speaker case can be designed that is free from audio distortion and clipping at low frequency ranges thereby presenting a practical defense against coresident adversary.

Pairing Application Settings: Since our proposed defense system relies upon a good quality speaker with frequency response in the low frequency range (50Hz-250Hz), the most suitable settings where it would be applicable is in pairing scenarios where the receiving device is equipped with a good quality speaker. Such pairing scenarios could involve a phone and a bluetooth speaker, a phone and a POS terminal, or an upgraded phone (with a case having powerful inbuilt speakers) and another device.

Other Masking Sounds: While the masking signal in the proposed work used low frequency sinusoidal tones for masking the audio leakage at lower frequency range, there may be other possible sounds that can be used for achieving similar effect. Human voice, drum beats and bass guitar chords are just some of the examples that should be tested in future work for their effectiveness at masking. A probable drawback for these sounds would be their discrete nature, potentially making it easier for an attacker to identify and filter them out.

Multiple Co-located Eavesdroppers: A possible avenue for compromising the proposed defense system is the triangulation attack [7] that classifies each sound source based on the difference in the time of arrival at multiple equidistant adversaries. However, there are two major limitations to this attack when applied to the pairing design: 1) triangulation attack lacks sufficient granularity to distinguish sound sources that may be located very close to each other. In the pairing setup, the transmitting and the receiving device are in contact with each other. In addition, the vibration sound does not have a single source, rather the transmitting device vibrates while in contact with the receiving device. 2) To increase the effectiveness of the triangulation attack, multiple co-located adversaries would be needed. In a pairing process, the devices would most likely have at the most two microphones (one on each device) for a co-located adversary (such as a resident malware) to exploit. This limitation reduces the effectiveness of the triangulation attack against the proposed defense mechanism.

Arbitrary Vibrational Communications: Our work focused on the security of the vibrational pairing schemes like PIN-Vibra [18].

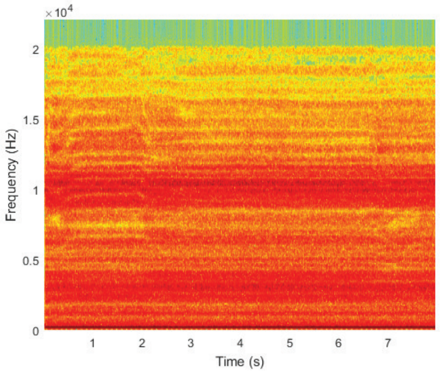


Figure 14: Frequency spectrogram for Source A after application of ICA. Intensity in the graph is proportional to energy in the frequency band.

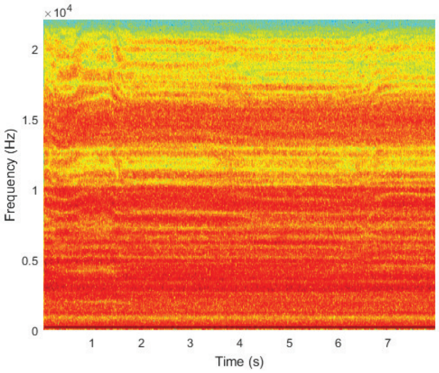


Figure 15: Frequency spectrogram for Source B after application of ICA. Intensity in the graph is proportional to energy in the frequency band.

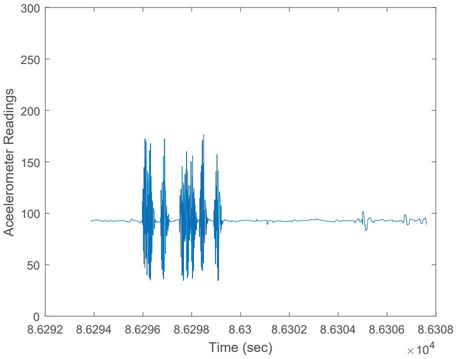


Figure 16: Peaks in accelerometer reading (without masking sound)

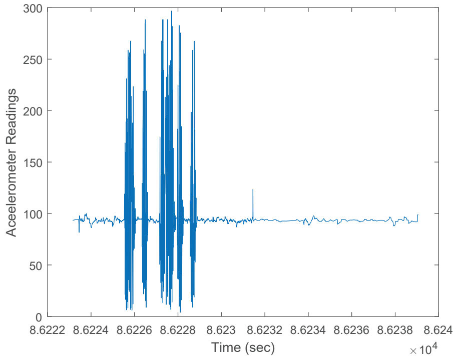


Figure 17: Peaks in accelerometer reading (after addition of masking sound via external speaker)

However, vibrations could also be used for arbitrary communications (Ripple [17]) that reach beyond the scope of the short duration pairing application. The masking signal that we proposed in this paper is suited for short transmissions like pairing devices. With arbitrary communications, the time taken for the communication to last could be much longer, affecting the usability of the system. It would also be worthwhile to explore other options, like the *anti-noise* signal proposed in Ripple [17]. We excluded it from the pairing scenario because creating an *anti-noise* signal is a computationally exhaustive task on constrained devices and might take more time than the actual time taken for the pairing process to complete. If time is not a limiting factor, an *anti-noise* signal might prove to be a better suited alternative than the proposed masking signal. Another potential for arbitrary secure vibrational communication may involve the use of masking sounds that would not distract the user. A comprehensive future investigation is necessary to extend our work to the arbitrary communication contexts.

Other Security Applications: The proposed defense mechanism in this work can be used to bolster the security of other authentication schemes, such as VibraPass [6], that use vibrations from the user's device to resist observation and shoulder-surfing attacks during password or PIN entry. However, a threat model similar to ours that involves co-located acoustic eavesdropping may determine the pattern of the vibrations thereby invalidating the security of the scheme. Further work is need to evaluate these scenarios in such authentication schemes and analyze the applicability of our work (both attacks and defenses).

6 CONCLUSIONS

In this paper, we emphasized the need for appropriately chosen acoustic noises to mask the sounds of vibration in vibrational pairing schemes. We considered acoustic eavesdroppers co-located with the device(s) being paired, and demonstrated that white noise alone as a masking signal, proposed in prior literature, is insufficient to cloak the acoustic emanations against such strong-yet-realistic attackers. On the positive side, we further showed that carefully incorporating low-frequency noises to the white noise signal serves as a viable defense even against co-located eavesdropping attacks. While current breed of smartphones may not be capable of producing such low-frequency sounds, the use of emerging phone cases that are equipped with better speakers, good-quality speakers present on the other pairing device, low-cost external speakers or even next generation smartphones would address this limitation. The proposed defense can help secure vibrational pairing against acoustic side channel eavesdropping, near or far, without undermining the overall performance of the vibrational decoding or the system efficiency and usability.

REFERENCES

[1] AmpAudio. 2016. AmpAudio. https://www.ampaudio.com/. (4 2016).

[2] S Abhishek Anand and Nitesh Saxena. 2016. Vibreaker: Securing Vibrational Pairing with Deliberate Acoustic Noise. In *9th ACM Conference on Security & Privacy in Wireless and Mobile Networks (WiSec '16)*.

[3] V. Boyko, P. MacKenzie, and S. Patel. 2000. Provably Secure Password-Authenticated Key Exchange Using Diffie-Hellman. In *Eurocrypt*.

[4] J Cardoso. 1998. Blind source separations: statistical principles. *Proc. IEEE* 9, 10 (1998), 2009–2025.

[5] Andrzej Cichocki, Juha Karhunen, Wlodzimierz Kasprzak, and Ricardo Vigario. 1999. Neural networks for blind separation with unknown number of sources. *NeuroComputing* 24, 1 (1999), 5593.

[6] Alexander De Luca, Emanuel von Zezschwitz, Vijay Raghunathan, and Heinrich Humann. 2009. VibraPass-Secure Authentication Based on Shared Lies. In *International Conference for Human-Computer Interaction (CHI) (CHI '09)*.

[7] A.H.Y. Fiona. 2006. Keyboard Acoustic Triangulation Attack. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.100.3156&rep=rep1&type=pdf. (2006). Final Year Project.

[8] guardRFID. 2017. GuardRFID announces Ultra Low Profile Active RFID Tag which includes Motion and Temperature Sensing. http://www.guardrfid.com/news/guardrfid-announces-ultra-low-profile-active-rfid-tag-which-includes-motion-and-temperature. (3 2017).

[9] Tzipora Halevi and Nitesh Saxena. 2013. Acoustic Eavesdropping Attacks on Constrained Wireless Device Pairing. In *IEEE Transactions on Information Forensics and Security (TIFS)*.

[10] A. Hyvarinen. 1999. Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. *IEEE Transactions on Neural Networks* 10, 3 (1999), 626–634.

[11] Countryman Associates Inc. 2017. B6 Omnidirectional Lavalier. http://www.countryman.com/b6-omnidirectional-lavalier-microphone. (3 2017).

[12] JBL. 2016. JBL - Clip Portable Bluetooth Speaker. http://www.bestbuy.com/site/jbl-clip-portable-bluetooth-speaker-purple/6050039.p?id=1219696711027&skuId=6050039. (4 2016).

[13] Ronald Kainda, Ivan Flechais, and A. W. Roscoe. 2009. Usability and Security of Out-of-band Channels in Secure Device Pairing Protocols. In *SOUPS*.

[14] Younghyun Kim, Woo Suk Lee, Vijay Raghunathan, Niraj K. Jha, and Anand Raghunathan. 2015. Vibration-based Secure Side Channel for Medical Devices. In *Proceedings of the 52Nd Annual Design Automation Conference (DAC '15)*.

[15] Altec Lansing. 2016. Altec Lansing - Mini H2O Bluetooth Speaker . http://www.alteclansing.com/en/al-products/mini-h20-speaker/. (4 2016).

[16] Motorola. 2017. JBL Soundboost Speaker. https://www.motorola.com/us/products/moto-mods/jbl-soundboost-speaker. (3 2017).

[17] Nirupam Roy, Mahanth Gowda, and Romit Roy Choudhury. 2015. Ripple: Communicating Through Physical Vibration. In *12th USENIX Symposium on Network Systems Design and Implementation (NSDI '15)*.

[18] Nitesh Saxena, Md. Borhan Uddin, Jonathan Voris, and N. Asokan. 2011. Vibrate-to-unlock: Mobile phone assisted user authentication to multiple personal RFID tags. In *IEEE International Conference on Pervasive Computing and Communications (Percom '11)*.

[19] SensMaster. 2017. SensMaster's Samui and Boyard available now with cost-effective motion detector. http://www.veryfields.net/active-rfid-tags-with-motion-detector-sensmaster. (3 2017).

[20] Sony. 2017. Sony SRS-XB2. http://www.sony.com/electronics/wireless-speakers/srs-xb2. (3 2017).

[21] Utterly Random Techie. 2017. LET'S SAY HELLO MOTO, CEBU! http://www.utterlyrandomtechie.com/lets-say-hello-moto-cebu/. (3 2017).

[22] Zagg. 2016. Zagg Speaker Case. http://www.zagg.com/us/en_us/cases/iphone-6-case/speaker-case. (4 2016).

[23] Bingsheng Zhang, Qin Zhan, Si Chen, Muyuan Li, Kui Ren, Cong Wang, and Di Ma. 2014. PriWhisper: Enabling Keyless Secure Acoustic Communication for Smartphones. *IEEE Internet of Things Journal* 1, 1 (2014), 33–45.

A EXAMPLE OF SPEAKER CASE FOR SMARTPHONES

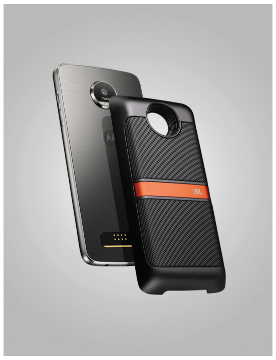


Figure 18: JBL® SoundBoost Speaker (image courtsey [21])