Vibrational Covert Channels using Low-Frequency Acoustic Signals

Nikolay Matyunin
matyunin@seceng.informatik.tu-darmstadt.de
TU Darmstadt, CYSEC, Germany

Yujue Wang
yujue.wang@stud.tu-darmstadt.de
TU Darmstadt, Germany

Stefan Katzenbeisser
stefan.katzenbeisser@uni-passau.de
Chair of Computer Engineering, University of Passau, Germany

ABSTRACT

In this paper, we examine how acoustic signals in sub-bass and infrasonic range can be used to establish a vibrational covert channel between speaker-equipped computers and mobile devices. We show that typical consumer speakers are capable of producing low-frequency sounds, which are not perceivable by humans. At the same time, we show that producing such sounds by the speaker's woofer inevitably generates slight vibrations of the speaker and the surface where it is located. Being unnoticeable to people, such vibrations can be captured by the accelerometer sensor of a mobile device located on the same surface. Therefore, information can be encoded into low-frequency sounds played by a speaker and received on a mobile device by analyzing the produced vibrations. Note that access to the accelerometer on modern mobile devices does not require any user permissions, making the transmission completely unnoticeable. We evaluate the presented covert channel for different speakers, apply it to several application scenarios, and give an overview of possible countermeasures.

KEYWORDS

vibrational covert channels, acoustic covert channels, infrasonic sounds, motion sensors, accelerometer, privacy, data exfiltration, cross-device tracking

ACM Reference Format:
Nikolay Matyunin, Yujue Wang, and Stefan Katzenbeisser. 2019. Vibrational Covert Channels using Low-Frequency Acoustic Signals. In *ACM Information Hiding and Multimedia Security Workshop (IH&MMSec '19), July 3–5, 2019, Paris, France.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3335203.3335712

1 INTRODUCTION

Utilizing inaudible acoustic signals as a medium for covert communication channels has been widely discussed in previous research, for a survey see [7]. Binary information can be encoded into sounds and transmitted over the air through speakers of one device. On another device, sound signals can be recorded using a microphone and decoded. Typically, ultrasonic sound (above 18kHz) is used for such a communication, as it makes transmission inaudible to humans, once the average hearing sensitivity of ≈20Hz–20kHz [15] is exceeded. Ultrasonic covert channels have been applied to a variety of attacks and applications, including data exfiltration from air-gapped systems [6, 8, 11] and sandboxed applications [5], and cross-device tracking [4, 16, 17].

To limit privacy threats emerging from acoustic covert channels, modern operating systems require explicit user permissions to access the device's microphone [1]. However, researchers have found that gyroscope sensors in smartphones can be used as zero-permission receivers of ultrasonic covert channels, as specific frequencies cause a disturbance in the sensor data due to resonance effects [5, 10, 16]. The applicability of such channels is, however, still limited, as the resonance frequency of most MEMS gyroscopes lies above 25kHz [16, 20], while sound processors in computers usually have a default sampling rate of 44.1kHz or 48kHz, and therefore cannot produce frequencies above 24kHz without aliasing.

In this work, we explore an alternative way to establish acoustic covert channels. Instead of using ultrasonic frequencies, we propose to use frequencies *below* the audible range, down to 16–24Hz. The hearing sensitivity falls sharply as the frequency goes below 50 Hz, i.e., humans can perceive such low frequencies only at a high sound pressure level (SPL) [24]. At the same time, speakers are normally designed to accurately reproduce signals only in the audible range, which often results in attenuation of the signal energy at low frequencies. As a result, at volume levels, set to conveniently perceive audible music or speech, the signal energy at low frequencies can be much lower compared to the audible range. As a result, signals deliberately produced at these frequencies remain imperceivable to human ears, but can be still captured using a microphone.

Furthermore, when a speaker produces low-frequency signals, the movements of the woofer's membrane lead to slight vibrations of the speaker enclosure, which propagate further to the surface where the speaker is located (e.g., a table). We have found that these slight vibrations, imperceptible to humans, can be successfully captured by the accelerometer sensor of a smartphone lying on the same surface. In this manner, the accelerometer can act as a zero-permission receiver of the *vibrational* covert channel, with the vibrations being deliberately produced by the speaker when playing inaudible low-frequency sounds.

We evaluated the produced vibrations for three speakers, and implemented the transmission, achieving a raw bitrate of up to 5bit/s with the bit error rate below 10%. Although bandwidth is limited in comparison to ultrasonic covert channels and transmitter and receiver must share the same surface, the proposed approach does not require high-resolution audio support for the transmitter or user permissions for the receiver, which increases the applicability of the covert channel.

IH&MMSec '19, July 3–5, 2019, Paris, France

ACM ISBN 978-1-4503-6821-6/19/07…$15.00
https://doi.org/10.1145/3335203.3335712

Contributions

Our contributions are the following:

- We show that woofer-equipped speakers can often produce sounds at frequencies as low as 16–24Hz, which can be inaudible to human ears. More importantly, when playing these sounds, woofers produce slight vibrations of the surface on which the speaker is located, which can be imperceivable to humans, but detectable by accelerometers on mobile devices.
- We propose to utilize this effect to establish a zero-permission vibrational covert channel. We identify the range of frequencies at practical SPLs which can be used for such a covert channel, and implement the transmission.
- We evaluate the proposed approach for several consumer speakers, and discuss several potential application scenarios and possible countermeasures.

2 BACKGROUND

In this section, we briefly describe how low-frequency sounds are perceived by human ears and how they are produced by speakers. Afterwards, we discuss the use of accelerometers in mobile devices.

2.1 Audibility of low-frequency sounds

It is traditionally considered that the human hearing limit lies within the so-called *audible* range of ≈20Hz–20kHz. In fact, the hearing sensitivity of human ears depends on the sound frequency and the sound pressure level (SPL): For each particular frequency, there exists a minimum SPL threshold required to perceive the sound [14]. In particular, humans can hear frequencies below 20Hz, if the SPL is sufficiently high. However, this threshold is significantly higher at the edges and outside of the audible range. For example, the average SPL threshold for the frequencies of 25 Hz and 12.5 Hz are ≈65dB and ≈95dB[1], respectively [23, 24]. For comparison, the hearing threshold for frequencies of 400Hz–4kHz is below 10dB [14].

In our work, we rely on the difference between hearing sensitivity for different frequencies: users may set the "loudness" of their speaker to a level which is convenient to hear audible frequencies (e.g., 70dB SPL for 1kHz); however, the signal at a low frequency (e.g., 25 Hz) would be inaudible as it is below the hearing threshold.

2.2 Speaker low-frequency capabilities

The primary function of a loudspeaker is to convert incoming electrical frequency signals into actual sound waves by performing physical vibrations. A detailed explanation of a speaker's design can be found in [9, 21]. In summary, a loudspeaker consists of one or multiple *drivers*, responsible to produce frequencies of a specific range. The drivers producing low frequencies are called *woofers* or *sub-woofers*, where the latter usually produces lower frequencies (we use these two terms interchangeably). The driver consists of a diaphragm connected to an electromagnet called *voice coil*, located within a permanent magnetic field. Applying a current produces electrodynamic force due to the interaction between magnetic fields, which causes the coil and diaphragm to oscillate back and forth, generating air pressure. The coil and diaphragm are connected to the speaker's enclosure through a spring called *spider*, which returns the diaphragm to its neutral position. The vibrating diaphragm, anchored to the enclosure through the spider, inevitably causes the enclosure itself to vibrate. As we observe in our experiments, these vibrations can further propagate to the surface on which the speaker is located and be measured by accelerometer sensors.

An important characteristic of the speaker is its frequency response, which describes the difference in produced SPL for the same input gain, depending on a frequency [21]. Usually, the speaker has an operational frequency range with comparably "flat" frequency response. Outside this range, however, the signal quickly attenuates. This fact further increases the range of frequencies which may be practically inaudible to human ears: For example, a speaker producing 90dB SPL at 1kHz frequency may produce only 60dB for 25Hz frequency, below the hearing sensitivity threshold.

2.3 Accelerometers

Most modern mobile devices are equipped with accelerometers, which measure the external acceleration force that is applied to the sensor along three physical axes, including the force of gravity, measured in m/s^2. The sensor data allows to identify and measure device motions, such as tilt or shaking, and is used in games, virtual reality applications, etc. In Android, iOS, and web applications, 3-axis accelerometer measurements can be retrieved by using the Sensor API [2], Core Motion [3] framework, and *DeviceMotion* event [18], respectively. The sampling rate of the measurements ranges between 60–500Hz, depending on the sensor and device.

In our work, we show that if the mobile device is located on the same surface as a loudspeaker, its accelerometer is sensitive enough to capture the slight vibrations of the reference surface itself, caused by the loudspeaker.

3 VIBRATIONAL COVERT CHANNEL

In this section, we describe the attack scenario, analyze how the vibration is presented in accelerometer data, describe the encoding scheme, transmittion details, and methods to receive the signal.

3.1 Attack scenario

We consider a generic covert channel attack model with two devices and an adversary, who is trying to extract sensitive information from one device (transmitter) by sending it over the covert channel to another device (receiver). The transmitter is assumed to be connected to a speaker equipped with a woofer. Any accelerometer-equipped mobile device (e.g., a smartphone) is considered as receiver. We further assume that transmitter and receiver are sharing a common surface, e.g., a wooden office table.

On the side of the transmitter, the attacker is able to encode the information to be exfiltrated into low-frequency sounds and play them through the speaker. We consider the case that the attacker may not be able to control the volume of the speaker, and instead has to play the signals at a level set by the device user.

On the receiver, the attacker has control over an application or a web page which records accelerometer data, captures the transmitted signals and extracts the information. Unlike traditional implementations of acoustic covert channels, which rely on access to the microphone, in our case the receiving code does not require any user permissions, except for access to the Internet, granted by default on

[1]Throughout the text, the reference sound pressure for SPL in decibels is $2 * 10^{-5}$Pa.

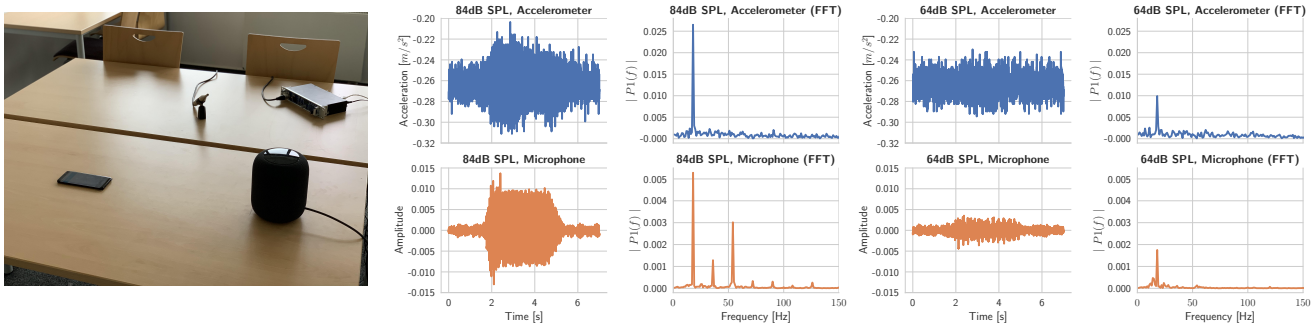

Figure 1: Examples of the experimental setup and recorded vibrations. Accelerometer measurements (blue) and sound (orange) were recorded while playing an 18Hz sine wave at 84dB and 64dB SPL in 50cm between devices. The figure shows the raw signal and its FFT representation. A clear peak at 18Hz is clearly distinguishable in the FFT values even for the low SPL.

modern operating systems. Therefore, it can be implemented as a malware, hidden in any application or a web page which the users are likely to open on their device. Finally, the decoded information is sent to the attacker over the Internet.

3.2 Analysis of the Signal Spectrum

The basic idea of our covert channel is to analyze surface vibrations produced by the speaker, by observing corresponding disturbances in accelerometer sensor data. Figure 1 shows an example of such a disturbance, as well as the experimental setup. Raw accelerometer measurements recorded on a smartphone, after playing an 18Hz sine wave on a speaker at two different SPLs. Additionally, the audio signal was recorded using a microphone located on another table at a distance of 50cm from the speaker. The figure shows raw measurements as well as their frequency domain representations after computing the Fast Fourier Transform (FFT).

The recordings were imperceivable to human ears, but the microphone was able to record the signal for both SPLs. Furthermore, when the sound was played at a high SPL, a clear disturbance in the raw accelerometer measurements can be noticed, indicating surface vibrations. The amplitude of the produced disturbance depends on the acoustic signal SPL: The woofer's diaphragm moves at larger amplitudes for the larger SPL, which results in stronger resulting vibrations of the speaker enclosure and the surface. As one can see, for the lower SPL, the raw disturbance is no longer noticeable.

For this reason, we utilize information in the frequency domain: For measurements corresponding to both SPL levels, one can see a clear peak in the signal spectrum at 18Hz, indicating that vibrations occur at the same frequency as the original sound signal. In practice, the frequency corresponding to the highest peak in the FFT can be different depending on the speaker design, surface and the relative position of the devices. For example, the speaker's diaphragm can produce vibrations for both forward and backward movements, which results in the resulting vibration frequency equal to the double value of the original sound frequency. Nevertheless, analyzing the actual frequency of produced vibrations instead of their amplitude allows us to detect even slight vibrations caused by sounds with comparably low SPL. We evaluate this approach to measure the produced vibrations in Section 4.2.

3.3 Encoding and Transmission

To encode a binary string into sound signals, we apply on-off keying (OOK) modulation. The transmitter plays a sine wave at a chosen frequency within a time interval to encode a 1, and does not play any sound within an interval to encode a 0. This encoding scheme allows us to evaluate the basic feasibility of the covert channel; we leave the evaluation of other encoding schemes for future work.

It is known that speakers can produce hearable audio clicks at the beginning and the end of each produced sine, due to abrupt changes of the signal phase and amplitude [8, 16]. To keep the transmission inaudible, we applied a Hanning window [19] at the beginning and the end of each generated sine wave. However, we found that on some speakers audible clicks could not be fully avoided on high SPLs even when the window is applied. We evaluate the audibility of the transmission in Section 4.3.

In order to help the receiver recognize the start of the transmission within a continuous recording, the transmitted string is prepended with a predefined binary sequence. We used time periods of different lengths for transmitting bits of this synchronization sequence and bits of the actual payloads, in order to avoid false positive detections when the synchronization sequence happens to appear in the payload itself. In our implementation, we used the 11-bit Barker sequence for synchronization [22], due to its low autocorrelation properties, which facilitates its accurate detection.

3.4 Receiving and Decoding

The receiver first converts the 3-axis accelerometer traces into an one-dimensional trace. We have observed that the actual direction of produced vibrations depends on the speaker design (e.g., whether the diaphragm is moving vertically or horizontally). In order to identify the direction of the vibration, we apply Principal Component Analysis [13] to the data, and choose the first component as the result, as it represents the new axis with the highest data variance.

Afterwards, the receiver identifies the start of the signal. If the actual frequency of vibrations is known, the receiver applies the Short-Time Fourier Transform (STFT) for this frequency with a short window and calculates the cross-correlation between resulting STFT amplitudes and the synchronization sequence. A time point corresponding to a high peak in the cross-correlation is considered as the start of a transmission. If the frequency of vibrations

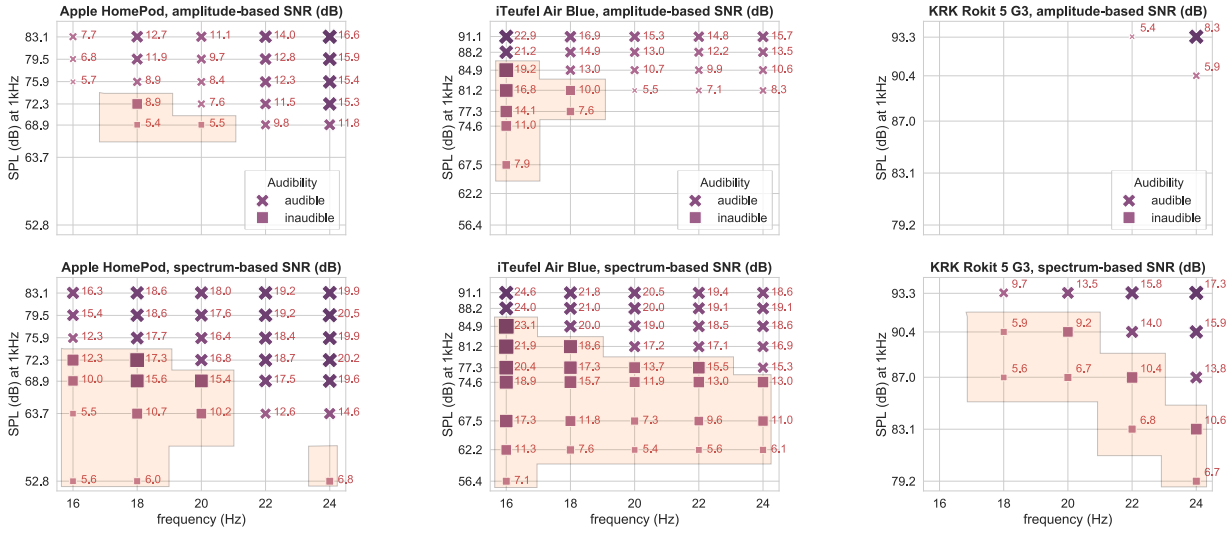

Figure 2: SNR levels of vibrations recorded on a smartphone in 50cm from the speaker, for three speakers, depending on the frequency and SPL (4.2), together with the audibility of the frequency-SPL points (4.3), inaudible: ■, audible: ×. An amplitude-based SNR (up) is compared to a spectrum-based SNR (down). Orange area highlights admissible range of frequencies/SPLs.

Speaker	16Hz	18Hz	20Hz	22Hz	24Hz
Apple HomePod	−26.3	−21.4	−17.2	−14.6	−12.2
iTeufel Air Blue	−29.7	−27.8	−21.6	−18.1	−16.8
KRK Rokit 5 G3	−35.6	−32.3	−30.8	−30.6	−30.5

Table 1: Frequency responses of the speakers, measured as difference in dB with SPL when playing 1kHz reference sine.

is not known, the receiver can perform this procedure for several candidate frequencies, e.g., checking the original sound frequency and its double value, or choosing frequencies with high average energy within the recorded interval.

Using the synchronization sequence, the receiver calculates the average FFT magnitude at the frequency of vibrations within time frames corresponding to a logical 0, and, similarly, the magnitude of all measurements corresponding to a logical 1. The mean of these two averages is used as a threshold. Finally, the transmitted payload is decoded bit by bit, by comparing this threshold with the FFT magnitude for each corresponding interval.

4 EVALUATION

In this section, we first describe the experimental setup and measure the frequency response of the speakers. Then, for each speaker, we evaluate produced vibrations and the audibility of the transmission. Finally, for the chosen setup, we evaluate the optimal bitrate.

4.1 Experimental setup

For our experiments, we chose 3 consumer speakers equipped with a woofer, listed in Table 1. We used a Galaxy S8 smartphone as the receiver, with its accelerometer sampled at the 500Hz rate. To precisely measure SPL, we used a professional calibrated Earthworks QTC30 microphone. If not stated otherwise, all experiments were performed in a typical office environment with moquette floors; speakers and smartphones were placed on a wooden office table.

As a first step, we identified the frequency response of the speakers at low frequencies. For this purpose, we played sine waves from 24 Hz down to 16 Hz, as well as a reference sine wave at 1kHz, and measured the produced SPL at 50cm. We set the volume level of the speakers so that the reference 1kHz wave is played at 80dB SPL. Afterwards, we calculated the difference between SPLs produced for each frequency within 16–24Hz and SPL at 1kHz. The average results are shown in Table 1. As one can see, all three speakers can play frequencies down to 16 Hz. However, the produced SPL at these frequencies drops significantly. This estimation shows that at practical SPLs, used to conveniently listen to the audible range sounds (e.g., 70dB for 1Khz sounds), the SPL produced when playing low-frequency sounds can fall far below the hearing threshold, making the resulting signal inaudible.

4.2 Vibration signal

Although the tested speakers can produce inaudible low-frequency sounds, they also need to conduct a sufficient amount of vibrations to the surface, to be captured by the accelerometer. In this experiment, we measured the strength of the produced vibrations.

For this purpose, a speaker and a smartphone were placed on the same table at a distance of 50cm from each other. We recorded accelerometer data when playing sine waves at each frequency within 16–24Hz range at different volume levels. Afterwards, we measured the Signal to Noise Ratio (SNR) of the recorded signals, expressed in decibels, in two different ways. First, we measured the SNR based on the raw disturbance of the vibration, i.e., the ratio of the signal power when the sine was played, to the signal power when no sound was produced. Second, we computed the spectrum-based SNR, by applying FFT to the signal and measuring SNR as the ratio between the peak value at the vibrating frequency to the median FFT value. The latter approach corresponds to the method for detecting the signal described in Section 3.

Figure 2 shows the results for 3 speakers depending on the frequency and SPL. For readability, we included only points with sufficient SNR≥5dB. We showed SPL values produced for the reference 1kHz sine at the corresponding volume level, to indicate SNR dependency on SPL for audible sounds. One can estimate the actual produced SPLs at a particular frequency using Table 1.

As one can see, all three speakers generate vibrations which can be captured by the accelerometer. Generally, the signal is higher for the higher SPL. Interestingly, the same trend does apply to frequencies: two speakers generate a higher signal for 16–18Hz signals in comparison to 20–20Hz. Even though the sound SPL gradually attenuates when lowering the frequency (see Table 1), the speaker may produce stronger vibrations at lower frequencies. More importantly, the spectrum-based approach significantly extends the range of frequencies and reduces the thresholds for SPL required for producing vibrations, which confirms its practical applicability.

4.3 Audibility

The previous experiment demonstrates the minimum SPL required for a particular frequency and a speaker to produce noticeable vibrations. In this experiment, we evaluated the audibility of the produced signals on our tested speakers. For this reason, we conducted a short user study with 5 participants, asking them if they could hear any sounds after randomly playing 2-second sine waves of 14–40Hz at different SPLs. Participants were sitting in 1m from the speaker in a typical office environment. We included intervals with no sound to verify the trustworthiness of the responses.

We included the results for different SPLs and frequencies in Figure 2, marking each pair of frequency and SPL as audible or inaudible. A pair was considered audible if it was perceived by at least 2 participants. As one can see, signals at peak SPLs are perceivable even for very low frequencies, either reaching the hearing threshold or due to audible clicks TRIM and distortions. However, when the SPL is sufficiently low for a particular frequency, the signal becomes inaudible.

Based on the results of the SNR and audibility experiments, we identified the admissible range of frequencies and SPLs, for which transmission generates a sufficient level of vibrations but remains inaudible. We highlighted this range in the Figure 2 for each speaker (orange area). We consider the frequency range of 18–20Hz with SPL levels of ≈65–75dB as the most practical, showing good performance for Apple HomePod and iTeufel Air Blue.

4.4 Transmission

In this section, we evaluated the optimal speed of transmission when using the on-off modulation scheme and decoding method described in Section 3.4. For this purpose, we chose 3 sets of frequency/SPL within the admissible ranges identified in previous experiments. Using each set, we transmitted binary messages with different bitrates, 50 messages of 20 bits in each case, and measured how the chosen bitrate affects the resulting bit error rate (BER). We considered BER of 10% as suitable, since in this case error-correcting code can be applied to ensure correct decoding with manageable overhead. Sounds were produced on the Apple HomePod speaker.

The results for five bitrates are shown in Table 2. The payloads can be successfully transmitted at bitrates up to 5bit/s. At a higher bitrate, the number of samples within a recording time frame for each bit becomes too small to correctly identify the frequency of vibrations. As a result, we consider bitrates of 2–5bit/s as practical.

Frequency	SPL, dB (at 1kHz)	BER, % 1bit/s	2bit/s	5bit/s	10bit/s
18Hz	72	0.5	4.8	5.0	20.0
18Hz	65	4.8	8.0	17.5	17.8
20Hz	65	1.8	3.2	16.0	27.0

Table 2: Transmission BER for selected frequencies & SPLs.

5 RELATED WORK

Existing works on acoustic covert channels utilize high-frequency sounds, starting from 17Khz. Ultrasonic covert channels were proposed to transmit data between isolated computers [6, 11] and mobile devices [8], and to perform cross-device tracking [4, 17]. Fortunately, operating system developers have limited the applicability of such channels by introducing a run-time permission to access the device's microphone.

However, multiple researchers demonstrated that MEMS gyroscopes can be used as zero-permission receivers for acoustic covert channels, due to their susceptibility to resonance sounds. This approach has been used to exfiltrate data from a surveillance implant [10], to break Android application sandboxing [5], and to establish a cross-device tracking [16]. In our work, we provide an alternative approach to receive covert acoustic signals without access to the microphone, by analyzing vibrations of the speaker. In comparison to ultrasonic covert channels with gyroscopes as receivers, our solution does not require high-resolution audio support and is not limited to a specific resonance frequency, but requires the transmitter and the receiver to share a common surface.

Deshotels [8] established a vibrational covert channel on mobile devices, by using the vibration motor of a smartphone as the transmitter. The idea of using accelerometers to capture vibrations from the speakers was introduced by Hasan et al. [12]. Authors hypothesized that low-frequency audio signals produced with strong sub-woofer system could be inaudible and detectable using accelerometers, but did not evaluate this idea. In our work, we show that the signal can actually be produced by comparably small consumer speakers, demonstrate how slight vibrations produced by sounds of low SPLs can be effectively decoded, and evaluate the transmission by identifying usable ranges for frequencies and SPLs.

6 APPLICATIONS

The ability to receive acoustic signals using just the accelerometer, without any user permissions, makes the channel applicable to a number of scenarios, which we briefly discuss in this section.

First, the covert channel can be used to exfiltrate information from so-called air-gapped systems, when computers are completely isolated from other devices and the Internet. In this case, covert channels can be the only way for attackers to exfiltrate data. In our case, a malicious insider can place the smartphone in close proximity to the device's speaker, or infect a smartphone of the device user with recording malware.

Second, similarly to the ultrasonic cross-device tracking technology [16], the proposed covert channel can be used to link user

profiles or activities on two devices. For example, an application or a web page on the user's laptop can encode a tracking identifier into sound signals and transmit it over the covert channel, while an application on user's smartphone can regularly record accelerometer data. If the recording application was able to receive the tracking ID, both devices must be located on a common surface, and are likely to belong to the same user.

Finally, the covert channel can be used to exfiltrate data from isolated environments within one mobile device. An example of such isolation could be the private browsing mode in web browsers, when activity traces are deleted after the user closes the session. However, an attacker may use the covert channel to transmit a tracking ID (e.g., a cookie) to another, not isolated web page.

7 COUNTERMEASURES

The most straightforward way to prevent the proposed covert channel is to apply a high-pass filter to all sounds played by the speaker, on the hardware or OS level, cutting off all frequencies below 24 Hz. In that case, transmission would be still possible using higher frequencies, but may not be inaudible. However, the attacker may still try to apply a steganographic approach, hiding hardly-perceivable signals into seemingly benign sounds, e.g., music.

Alternatively, effort can be put on reducing vibrations conducted from the speaker's enclosure to the surface, e.g., by using anti-vibration pads. However, it may be not possible to prevent the vibrations completely, as the working principle of the speaker woofer is based on physical vibrations of the diaphragm.

The high-pass filter can also be applied to the accelerometer measurements to prevent detection of the low-frequency signals in the spectrum domain, although precise sensor measurements can be required for some apps. Furthermore, mobile OS vendors can forbid access to accelerometer data without an explicit permission granted by the user, limiting the applicability of the covert channel.

8 DISCUSSION AND FUTURE WORK

In this section, we discuss several limitations of our approach and directions for future work.

First, we would like to point out that the transmission requires the transmitter and receiver to share a common surface and does not work over the air, at least at tested SPLs. We tried placing the smartphone very close to the speaker (≈5cm) on another surface, and did not observe the signal. Furthermore, the signal depends on the surface itself: We confirmed that decoding worked for 2 different office tables and 5 different relative positions of devices on the same table, with 10–160cm between devices. However, the transmission did not work on more stable surfaces (e.g., a window sill from artificial stone) or softer surfaces (e.g., a moquette floor). More extensive evaluation of the depending of the signal on the surface may be needed to confirm applicability of the approach.

Second, the low transmission rates can be improved, by investigating whether more efficient encoding schemes can be applied (e.g., multiple frequency shift keying), and more accurate signal synchronization can be implemented.

Finally, we did not thoroughly investigate the nature of vibrations produced by the speakers. Comparing speaker designs in terms of produced vibrations can be direction for future work.

9 CONCLUSION

In this work, we presented a vibrational covert channel between devices equipped with speakers and mobile devices. We showed that binary data can be encoded into inaudible low-frequency sounds. When playing these sounds, speakers produce slight vibrations of the surface, which can be captured by the accelerometer on the mobile device. We identified suitable ranges of frequencies for such transmission, and evaluated the basic transmission rate. The presented covert channel does not require any explicit permissions and can be used unnoticed to end users.

REFERENCES

[1] Android Developer Documentation. 2019. Permissions overview. Retrieved May 19, 2019 from https://developer.android.com/guide/topics/permissions/overview

[2] Android Developer Documentation. 2019. Sensors overview. Retrieved May 19, 2019 from https://developer.android.com/guide/topics/sensors/sensors_overview

[3] Apple Developer Documentation. 2019. Core Motion Framework. Retrieved May 19, 2019 from https://developer.apple.com/documentation/coremotion

[4] Daniel Arp, Erwin Quiring, Christian Wressnegger, and Konrad Rieck. 2017. Privacy threats through ultrasonic side channels on mobile devices. In *Security and Privacy (EuroS&P), 2017 IEEE European Symposium on*. IEEE, 35–47.

[5] Kenneth Block, Sashank Narain, and Guevara Noubir. 2017. An autonomic and permissionless Android covert channel. In *Proceedings of the 10th ACM Conference on Security and Privacy in Wireless and Mobile Networks*. ACM, 184–194.

[6] Brent Carrara and Carlisle Adams. 2014. On acoustic covert channels between air-gapped systems. In *International Symposium on Foundations and Practice of Security*. Springer, 3–16.

[7] Brent Carrara and Carlisle Adams. 2016. Out-of-band covert channels: a survey. *ACM Computing Surveys (CSUR)* 49, 2 (2016), 23.

[8] Luke Deshotels. 2014. Inaudible Sound as a Covert Channel in Mobile Devices. *Proceedings of the 8th USENIX Workshop on Offensive Technologies - WOOT '14* (2014), 16.

[9] Vance Dickason. 2005. Loudspeaker design cookbook. (2005).

[10] Benyamin Farshteindiker, Nir Hasidim, Asaf Grosz, and Yossi Oren. 2016. How to Phone Home with Someone Else's Phone: Information Exfiltration Using Intentional Sound Noise on Gyroscopic Sensors. In *10th USENIX Workshop on Offensive Technologies, WOOT 16, Austin, TX, USA, August 8-9, 2016.*

[11] Michael Hanspach and Michael Goetz. 2014. Recent Developments in Covert Acoustical Communications. *Sicherheit* (2014), 243.

[12] Ragib Hasan, Nitesh Saxena, Tzipora Halevi, Shams Zawoad, and Dustin Rinehart. 2013. Sensing-enabled channels for hard-to-detect command and control of mobile devices. In *8th ACM Symposium on Information, Computer and Communications Security, ASIA CCS '13, Hangzhou, China - May 08 - 10, 2013*. 469–480.

[13] H. Hotelling. 1933. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* 24, 6 (1933), 417–441.

[14] International Standard ISO 226. 2003. Normal Equal-Loudness-Level. *International Organization for Standardization, Geneva, Switzerland* (2003).

[15] Jungmee Lee, Sumitrajit Dhar, Rebekah Abel, Renee Banakis, Evan Grolley, Jungwha Lee, Steven Zecker, and Jonathan Siegel. 2012. Behavioral hearing thresholds between 0.125 and 20 kHz using depth-compensated ear simulator calibration. *Ear and hearing* 33, 3 (2012), 315.

[16] Nikolay Matyunin, Jakub Szefer, and Stefan Katzenbeisser. 2018. Zero-permission acoustic cross-device tracking. In *2018 IEEE International Symposium on Hardware Oriented Security and Trust, HOST 2018*. 25–32.

[17] Vasilios Mavroudis, Shuang Hao, Yanick Fratantonio, Federico Maggi, Christopher Kruegel, and Giovanni Vigna. 2017. On the Privacy and Security of the Ultrasound Ecosystem. *Proceedings on Privacy Enhancing Technologies* 2017, 2 (2017), 95–112.

[18] Mozilla. 2019. DeviceMotionEvent. Retrieved May 19, 2019 from https://developer.mozilla.org/en-US/docs/Web/Events/devicemotion

[19] Alan V Oppenheim. 1999. *Discrete-time signal processing*. Pearson Education India.

[20] Yunmok Son, Hocheol Shin, Dongkwan Kim, Young-Seok Park, Juhwan Noh, Kibum Choi, Jungwoo Choi, Yongdae Kim, et al. 2015. Rocking Drones with Intentional Sound Noise on Gyroscopic Sensors.. In *24th USENIX Security Symposium, USENIX Security 15, Washington, D.C., USA, August 12-14, 2015.*

[21] Michael Talbot-Smith. 2012. *Audio engineer's reference book*. Focal Press.

[22] R Turyn and J Storer. 1961. On binary sequences. *Proc. Amer. Math. Soc.* 12, 3 (1961), 394–399.

[23] Toshio Watanabe and Henrik Moller. 1990. Hearing thresholds and equal loudness contours in free field at frequencies below 1 kHz. *Journal of Low Frequency Noise, Vibration and Active Control* 9, 4 (1990), 135–148.

[24] LS Whittle, SJ Collins, and DW Robinson. 1972. The audibility of low-frequency sounds. *Journal of sound and vibration* 21, 4 (1972), 431–448.