MOSQUITO: Covert Ultrasonic Transmissions between Two Air-Gapped Computers using *Speaker-to-Speaker* Communication

Mordechai Guri, Yosef Solewicz, Yuval Elovici
Ben-Gurion University of the Negev, Israel
gurim@post.bgu.ac.il; yosef.solewicz@gmail.com; elovici@bgu.ac.il
Air-gap research page: https://cyber.bgu.ac.il/advanced-cyber/airgap

Abstract—In this paper we show how two or more air-gapped computers in the same room, equipped with passive speakers, headphones, or earphones can covertly exchange data via ultrasonic waves. Microphones are not required. Our method is based on the capability of a malware to exploit a specific audio chip feature in order to reverse the connected speakers from output devices into input devices - unobtrusively rendering them microphones. We discuss the attack model and provide technical background and implementation details. We show that although the reversed speakers/headphones/earphones were not originally designed to perform as microphones, they still respond well to the near-ultrasonic range (18kHz to 24kHz). We evaluate the communication channel with different equipment, and at various distances and transmission speeds, and also discuss some practical considerations. Our results show that the speaker-to-speaker communication can be used to covertly transmit data between two air-gapped computers positioned a maximum of nine meters away from one another.

I. INTRODUCTION

Two (or more) computers in the same room are considered to be separated by an 'air-gap' if there is no physical or logical connection between them. In the context of cyber security, this measure is taken in order to ensure strict isolation between nearby computers. A common scenario involves two computers in the same room, where each computer is connected to a separate network of the organization. The air-gap separation ensures that data cannot be exchanged between the two networks, and more specifically, in a situation in which two computers have been compromised with a malware, data cannot be sent from one computer to the other and vice versa.

A. Speaker-to-Microphone Covert Channel

Despite the high degree of isolation provided by air-gapping, it doesn't provide a hermetic solution. It is known that the air-gap between two computers in the same room can be 'bridged' if the two computers are equipped with speakers and microphone [30], [10]. That is, the two computers can covertly exchange data via inaudible sound waves. In this type of communication, one computer transmits the data to the other via high frequency sound (usually at 18kHz or higher), using its loudspeaker. The receiver computer uses its microphone to receive the data. The speaker-to-microphone communication described above is mainly relevant for laptops, which have built-in speakers and microphones. Hence, previous research on this covert channel has primarily focused on laptops [30].

B. Microphone-less Environments

The speaker-to-microphone covert channel has one main drawback: in many real-life IT environments, microphones are not available to the attacker. The common cases include:

- **Desktop workstations**. Unlike laptops which have integrated microphones, desktop workstations are not always connected with an external microphone.
- **Secure environments**. In secure environments, microphones in desktop computers may be prohibited (or disconnected) to avoid the risk of eavesdropping. In secure environments, microphones may be forbidden in order to maintain an 'audio-gap' between computers. Elimination of microphones is an effective defense against the speaker-to-microphone covert channel discussed above [11].
- **Disabled/muted microphones**. A computer (desktop workstation or laptop) may be equipped with a microphone, which at some point was disabled, muted (with a physical 'off' button), or taped [3]. This typically occurs when the user wants to increase security and ensure confidentiality.

Consequently, the speaker-to-microphone covert channel limits the attacker's abilities, allowing the attacker to operate only in environments where microphones *are* present and enabled.

C. Speaker-to-Speaker Covert Channel

In this paper we show how the air-gap between two isolated computers can be bridged in 'speakers-only' environments. That is, where two computers in the same room are not equipped with microphones but are equipped with different types of output devices: (microphone-less) headphones, (microphone-less) earphones/earbuds, or passive speakers. Our method is based on the capability of a malware to transform a computer speaker from an output device into an

input device - inconspicuously changing its role from speaker to microphone and vice versa [26]. The two computers can then be used to send data (by using the speakers) and receive data (by using the transformed speakers) via inaudible sound.

The contribution of this paper is as follows:

- **Attack model.** We extensively discuss the speaker-to-speaker communication attack model of bridging the air-gap between two desktop computers. We also discuss and evaluate different types of speakers and headphones and their response to the ultrasonic range.
- **Speakers-to-headphones.** We discuss and evaluate the never discussed before threat of the speaker-to-headphones and headphones-to-headphones communication channel.
- **Evaluation.** We evaluate the speaker-to-speaker ultrasonic covert channel. In particular, we evaluate the acoustic response of passive speakers, headphones, and earphones to the near-ultrasonic range, when transformed into microphones (recall that such speakers are not designed to function as input devices).
- **Transmission protocol.** We provide a protocol stack designed for speaker-to-speaker communication. In this covert channel, the two computers must synchronize and change the speakers' roles (from speakers to microphones and vice versa) during the communication. We developed an appropriate protocol to handle this mutual communication.

The rest of this paper is organized as follows: Technical background is provided in Section II. Related work is presented in Section III. The attack is discussed in Section IV. Communication details are provided in Section V. Section VI describes the analysis and evaluation results. Countermeasures are discussed in Section VII. We conclude in Section VIII.

II. TECHNICAL BACKGROUND

In this section, we provide the technical background necessary to understand the attack itself. An essential part of the speaker-to-speaker covert channel is malware's ability to record audio signals through the speakers/headphones/earphones connected to the computer. In the following subsection, we describe this issue and discuss its limitations.

A. Speaker Reversibility

A speaker aims at amplifying audio streams out, but it can actually be viewed as a microphone working in reverse mode: a loudspeaker converts electric signals into a sound waveform, while a microphone transforms sounds into electric signals. More technically, speakers use the changing magnetic field induced by electric signals to move a diaphragm in order to produce sounds. Similarly, in microphone devices, a small diaphragm moves through a magnetic field according to a sound's air pressure, inducing a corresponding electric signal [9]. This bidirectional mechanism facilitates the use of a simple speaker as a feasible microphone simply by plugging it into a microphone jack. It should be clear that in practice, speakers were not designed to perform as microphones, and the recorded signals will be of low quality.

B. Jack Retasking

Interestingly, the audio chipsets in modern motherboards and sound cards include an option to change the function of an audio port at the software level, a type of audio port programming sometimes referred to as 'jack retasking'. This option is available on most audio chipsets (e.g., Realtek's audio chipsets) integrated into PC motherboards today. Jack retasking, although documented in the technical specifications, is not well-known [31]. For an in-depth technical discussion on malicious retasking of an audio jack, from the hardware to the operating system level, we refer the interested reader to our previous work [26].

The fact that loudspeakers, headphones, earphones, and earbuds are physically built like microphones, coupled with the fact that an audio port's role in the PC can be altered programmatically, changing it from output to input, creates a vulnerability which can be abused by attackers. A malware can stealthily reconfigure the headphone jack from a line out jack to a microphone jack. As a result, the connected output device can function as a pair of recording microphones, thereby rendering the computer a recording device - even when the computer does not have a connected microphone.

C. Passive speakers, Headphones and Earphones

The reversibility of speakers poses a limitation, in that the speaker must be passive (unpowered), without amplifier transitions. In the case of an active (externally powered) speaker, there is an amplifier between the jack and the speaker; hence, the signal will not be passed from the output to the input side [13]. Headphones, earphones, and earbuds are built from a pair of passive speakers, and hence, are always reversible. However, most PC loudspeakers today have an internal amplifier [5]. Passive speakers mainly exist in legacy and intercom systems [1].

III. RELATED WORK

Air-gap covert channels are special covert channels, which enable communication from air-gapped computers - mainly for a purpose of data exfiltration [11]. They can be classified into five main categories: electromagnetic, magnetic, acoustic, thermal, and optical. In 2014, Guri et al demonstrated AirHopper [20], [22], malware that exfiltrate data from air-gapped computers to a nearby smartphone via FM signals emitted from the screen cable. Later on Guri et al also demonstrated GSMem [19], malware that leaks data from air-gapped computers to nearby mobile-phones using cellular frequencies generated from the buses which connect the RAM and the CPU. In 2016, Guri et al showed USBee, a malware that uses the USB data buses to generate electromagnetic signals from a desktop computer [21]. In 2018, Guri et al presented ODINI [28], a malware that can

exfiltrate data from air-gapped computers via low frequency magnetic signals generated by the computer's CPU cores. The magnetic fields bypass Faraday cages and metal shields. Guri et al also demonstrated MAGNETO [16], which is a malware that leak data from air-gapped computers to nearby smartphones via magnetic signals. They used the magnetic sensor integrated in smartphones to receive covert signals. Matyunin suggested using magnetic head of hard disk drives to generate magnetic emission, which can be received by a nearby smartphone magnetic sensor [34]. Several studies have proposed the use of optical emanation from computers for covert communication. Loughry introduced the use of PC keyboard LEDs to encode binary data [32]. In 2017, Guri et al presented LED-it-GO, a covert channel that uses the hard drive indicator LED in order to exfiltrate data from air-gapped computers [29]. Guri et al also presented a method for data exfiltration from air-gapped networks via router and switch LEDs [27]. Data can also be leaked optically through fast blinking images or low contrast bitmaps projected on the LCD screen [18]. In 2017, Guri et al presented aIR-Jumper, a malware that uses the security cameras and their IR LEDs to communicate with air-gapped networks remotely [15]. In 2015, Guri et al introduced BitWhisper [23], a thermal covert channel allowing an attacker to establish bidirectional communication between two adjacent air-gapped computers via temperature changes. In acoustic covert channels, data is transmitted via inaudible, ultrasonic sound waves. Audio based communication between computers was reviewed by Madhavapeddy et al. in 2005 [33]. In 2013, Hanspach [30] used inaudible sound to establish a covert channel between air-gapped laptops equipped with speakers and microphones. Their botnet established communication between two computers alocated 19 meters apart and can achieve a bit rate of 20 bit/sec. All of the acoustic methods presented above require speakers. In 2016, Guri et al introduced Fansmitter, a malware which facilitates the exfiltration of data from an air-gapped computer via noise intentionally emitted from the PC fans [24]. In this method, the transmitting computer does not need to be equipped with audio hardware or an internal or external speaker. Guri et al also presented DiskFiltration a method that uses the acoustic signals emitted from the hard disk drive (HDD) moving arm to exfiltrate data from air-gapped computers [25].

Microphone-less computers. The attack presented in the current paper is relevant to environments in which the computers are not equipped with microphones, a common setup seen in many IT environments. Guri et al presented Speake(a)r [26] a malware that covertly turns the headphones, earphones, or simple earbuds connected to a PC into a pair of eavesdropping microphones when a standard microphone is not present, muted, taped, or turned off. They discuss technical details of this type of attack from the hardware to operating system level. However, the work of Guri et al in [26] focuses on the threat of conversation eavesdropping and did not discuss the threat of ultrasonic covert channel. As we noted, most

TABLE I
SUMMARY OF EXISTING AIR-GAP COVERT CHANNELS

Type	Method
Electromagnetic	AirHopper [20], [22] (FM radio)
	GSMem [19] (cellular frequencies)
	USBee [21] (USB bus emission)
	Funthenna [2] (GPIO emission)
Magnetic	MAGNETO [16] (CPU-generated magnetic fields)
	ODINI [28] (Faraday shields bypass)
	Hard-disk-drive [34]
Acoustic	Fansmitter [24] (computer fan noise)
	DiskFiltration [25] (hard disk noise)
	Ultrasonic [30], [10]
	MOSQUITO (speaker-to-speaker)
Thermal	BitWhisper [23] (heat emission)
Optical	LED-it-GO [29] (hard drive LED)
	VisiSploit [18] (invisible pixels)
	Keyboard LEDs [32]
	Router LEDs [27]
Optical (infrared)	aIR-Jumper [15] (security cameras & infrared)

loudspeakers connected to PCs today have an integral amplifier which prevents passing any signal from output to input.

Table I. summarizes the existing air-gap covert channels.

IV. ATTACK

In the attack scenario, two or more computers are located in the same room - separated by an air-gap (Fig. 1). That is, there is no physical or logical network connection between the two computers. The computers are not equipped with microphones but are equipped with output devices: active speakers, passive speakers, headphones, or earbuds. We distinguish between two types of communication.

- **Two computers, one-way communication.** In this case, two air-gapped computers in the same room establish unidirectional communication. This is the simplest case, where one computer is a transmitter and the other is a receiver. In this case, the transmitter is not necessarily equipped with a reversible speaker (e.g., it might be connected to an active loudspeaker).
- **Two computers, bidirectional communication.** In this case, two air-gapped computers in the same room establish bidirectional communication. In this case, each of the computers is a transmitter and a receiver. In this case, both computers are equipped with reversible speakers.

A. Malware

The communicating computers are infected with a malware. The malware has three operational components, described below.

- **Jack retasking.** Reversing the output audio jacks into input jacks, effectively turning the playing devices to microphones. This technique is described in detail in [26].
- **Synchronization.** Synchronizing between the sender and the receiver. This component is essential for a bidirectional communication. By using the synchronization, the
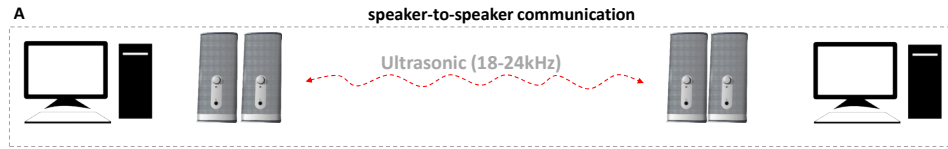
Fig. 1. The speaker-to-speaker covert communication

malware determines when the speaker should be used as a speaker and when it should be reversed to a microphone.

- **Transmission and reception.** Transmitting and receiving the data. This component performs the modulation of the data over ultrasonic wave and its demodulation back to binary data. It also includes the bit framing and the transmission protocol.

The general architecture of the malware is shown in Fig. 2. The malware consists of a user level process and a kernel driver. The user level process manages the set of malware tasks, such as communication, command and control (C&C), persistency, and so on. The kernel level driver is in charge of retasking the jacks from output to input and vice versa. Malware process runs in the user mode, reconfiguring the headphone jack into a microphone jack by sending a retasking request to the kernel driver. The kernel driver sends the configuration verb to the audio codec through the audio interface, which sends it to the audio codec via the HD audio bus.
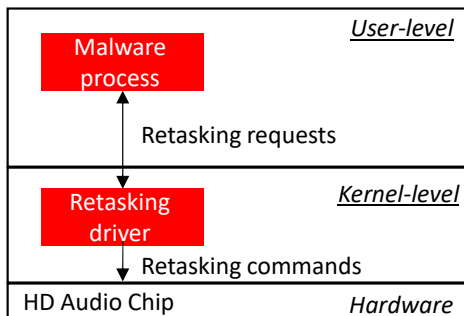

Fig. 2. The malware architecture

B. Air-Gap Communication

The attack presented in this paper allow attackers to transmit data between two computers. For example, when one computer is Internet connected and the other is an isolated, air-gapped computer. In the initial phase, the two computers must be infected with a malware. Note that it has been shown that attackers can successfully compromise air-gapped networks by using complex attack vectors, such as supply chain attacks, malicious insiders, and social engineering [36], [14], [8]. For example, in 2017 WikiLeaks published a reference to a hacking tool dubbed 'Brutal Kangaroo,' used to infiltrate air-gapped computers via USB drives [4]. When an employee of the organization inserted an infected USB drive into the air-gapped computer, a malicious code was executed.

Having established a foothold in both computers, the attacker may bridge the air-gap between the internal and external

networks using the speaker-to-speaker covert channel. The attacker can then exfiltrate information to the Internet connected computer (e.g., documents, passwords and encryption keys). Alternatively, the attacker may communicate with the isolated network by issuing commands and receiving responses.

V. COMMUNICATION

In this section, we present the design and implementation of the speaker-to-speaker communication. We discuss the detection and synchronization protocol and present the data modulation and encoding scheme. For this discussion, we assume that there are two computers (or 'nodes') denoted as **A** and **B**. We present a generic protocol which assumes that both **A** and **B** are connected with a passive speaker or headphones/earphones. At the end of this section, we discuss a case in which only one of the computers is equipped with a reversible speaker, allowing only unidirectional (rather than bidirectional) communication. Note that for simplicity we present the basic case with only two communicating peers.

A. Protocol Stack

The approach taken by other research on the ultrasonic covert channel is to use the existing implementation of protocol stacks originally designed for acoustic (non-ultrasonic) communication (e.g., [30], [37]). In this paper, we choose to implement our own light, stripped-down audio protocol stack for the evaluation of the covert channel.

B. The Near-Ultrasonic Range

Human hearing is limited to sound waves of 20kHz. In covert channel research it is acceptable to classify the range above 18kHz as practically inaudible for adults [12], [40]. In 2016, a group of researchers performed in-depth analysis on the emerging threat of ultrasonic cross-device tracking (uXDT). They found that ultrasound beacons (uBeacons) at a range of 18kHz to 20kHz are embedded into websites and TV advertisements [35]. The beacons are then picked up apps installed on nearby smartphones. Accordingly, in this paper, we consider the frequency range of 18kHz to 24kHz acceptable for the covert communication.

C. Data Modulation

In audio frequency-shift keying (AFSK) digital data is represented by changes in the frequency of an audio tone. AFSK is used to transmit binary data over radio and telephony systems. For the data transmission we implemented binary frequency-shift keying (B-FSK) modulation. In B-FSK the data is represented by a change in the frequency of a carrier

wave. In our case, two different audio frequencies f_0 and f_1 in the range of 18kHz to 24kHz represent two different symbols '0' and '1.'

D. Bit-Framing

The data packets are transmitted in small frames. Each frame consists of 46 bits and is comprised of preamble, payload, and CRC (cyclic redundancy check), as shown in Fig. 3.

Preamble. The preamble is transmitted at the beginning of every packet. It consists of a sequence of six alternating bits ('101010') which helps the receiver determine the properties of the channel, such as the carrier wave frequency and the bit period (bit rate). In addition, the preamble header allows the receiver to detect the beginning of the transmission of each packet. This is important for our covert channel, since in the case of the ultrasonic covert channel, a transmission might get interrupted, e.g., if the computer was restarted in the middle of an ongoing transmission.

Payload. The payload is the 32 bits of raw data which contains the actual packet.

CRC. For error detection, we insert eight bits of CRC code at the end of the frame. The receiver calculates the CRC for the received payload, and if it differs from the received CRC, an error is detected. In the case of error a packet retransmission request is sent (only in the case of bidirectional communication).

E. Communication Protocol

The acoustic channel is a type of shared communication channel. There are different types of multiple access protocols allowing a communication channel to be shared between many nodes (e.g., TDMA, ALOHA and CSMA [38]). Recall that in the proposed speaker-to-speaker communication a speaker can function as either a transmitter (speaker) or receiver (microphone) at a given time. Thus, each computer must know when the speaker is being used as a speaker and when to reverse it to a microphone. We used the concept of virtual 'tokens,' in which one computer acquires a transmission token. The other computer is only allowed to transmit when a transmission token is has been released. Each computer can hold the token for a maximal time slot T_{max}. When the computer has finished the transmission, it releases the token and begins to listen. The sequential flow of the communication between the two computers is illustrated in Fig. 4. At the beginning of the transmission, computer **A** acquires the transmission token, transmits n frames, and releases the token. Computer **B** then acquires the token, transmits m frames, and releases the token.

F. Discovery Broadcast Message

In order to establish the covert communication channel, the computers have to acknowledge each other's presence. To that end, each computer sends a broadcast message called a discovery beacon. The discovery beacon contains eight bits, which encode the computer identifier (ID). The identifier is a random number generated once at the beginning of a communication session. Since the computers are not connected, both computers might initially choose the same identifier. This case is handled by a simple rule: the first computer that detects the ID collision randomizes and broadcasts its ID. In order to discover the other computer, each computer is performs the discovery scheme outlined in Algorithm 1.

Algorithm 1

1: **while** (state != DISCOVERED) **do**
2:    $wait(random(5000))$
3:    $jack\_retask(SPEAKER)$
4:    $transmit(discoveryMessage(ID))$
5:    $jack\_retask(MIC)$
6:    $message \leftarrow waitForDiscoveryAck(5000)$
7:    **if** (message) **then**
8:        $set\_state(DISCOVERED)$
9:    **end if**
10: **end while**

Each computer starts by broadcasting its ID at random times every 5000ms. Note that a random time is used in order to avoid collision with a discovery message sent by the other computer, a technique which is used in communication for collision avoidance [39]. Following the ID broadcast, the computer retasks its speaker to a microphone. It waits for a discovery acknowledgment message sent from the other computer. If an acknowledgment message is received, it stops broadcasting the discovery message.

G. Type of Messages

Table II shows the main control messages in our protocol, including the DISCOVERY, ACQUIRE and RELEASE messages described earlier in this section.

TABLE II
CONTROL MESSAGES

#	Message	Description
1	DISCOVERY	The discovery broadcast message
2	ACQUIRE	Acquire the transmission token
3	RELEASE	Release the transmission token
4	ACK_OK	Frame received successfully (ack)
5	RETRANSMIT	Request to retransmit a frame
6	BITRATE_INC	Increase the current bit rate (+5%)
7	BITRATE_DEC	Decrease the current bit rate (-5%)

The ACK_OK message notifies the other computer that the message was received successfully and confirms that the CRC was correct. The RETRANSMIT message requests the other computer to retransmit a frame (e.g., if the CRC is incorrect). The BITRATE_INC and BITRATE_DEC messages enable the two computers to agree on increasing up or decreasing down the current transmission rate in 5%. In particular, the messages enable adaptive use of the channel speed to cope with environmental noise.
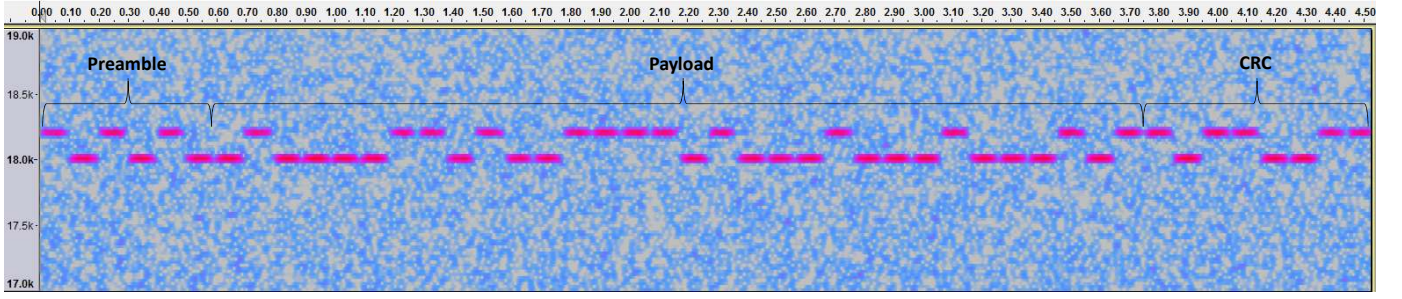
Fig. 3. The spectrogram of the 46 bit frame (preamble, payload, and CRC) transmitted at 10 bit/sec using B-FSK modualtion
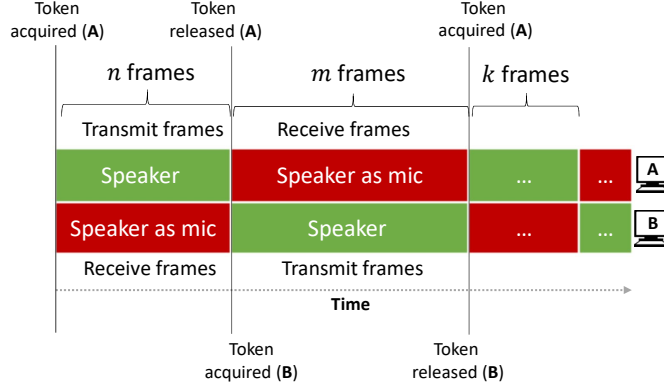

Fig. 4. The communication protocol between computers **A** and **B**

VI. ANALYSIS AND EVALUATION

Headphones, earphones, and passive speakers were not designed to perform as microphones in terms of quality and frequency range. In this section, we assess the efficacy of the speaker-to-speaker communication and present an empirical analysis of its corresponding channel capacity. We also discuss various practical considerations concerning the ultrasonic covet channel. Note that we are mainly interested in the high frequency regions that offer high channel capacity while at the same time have low auditory awareness.

A. Channel Capacity

Channel capacity (C) is a measure of the theoretical upper bound on the rate at which information can be transmitted over a communication channel. We assume that S is the power of the signal conveying the information and is corrupted by additive interfering Gaussian noise, with power N. The available communication bandwidth is B (in Hz). Given that, the channel capacity in bits per second can be calculated using the Shannon-Hartley theorem:

$$C = Blog_2(1 + \frac{S}{N}) \qquad (1)$$

Intuitively, this formula informs us that the higher the signal-to-noise ratio (SNR), and channel bandwidth, the higher the amount of information that can be conveyed.

We calculate the capacity of a communication channel formed between two loudspeakers, one of which serves as a transmitter and the other serves as a receiver. In these experiments, a sweep sinusoid of ten second in length at a range of 1Hz to 24kHz is played from the transmitter and recorded by the receiver. We use the Praat [6] tool to perform a short-time spectral analysis of the received signal.

Measurement setup. We evaluate the channel capacity for distances of one, four and eight meters between the transmitter and the receiver. To that end, we tested three off-the-shelf passive loudspeakers as receivers: (1) Logitech Z523, (2) Logitech Z213, and (3) Philips SPA5300. We also tested a pair of small Samsung earbuds for comparison. The loudspeakers were connected to a retaskable audio output jack on an Optiplex 9020 desktop PC. The sweep signal was played through a Logitech Z100 loudspeaker connected to a Gigabyte GA-H97M-D3H desktop workstation, (Intel Core i7-4790) running Ubuntu 16.04.1 kernel 4.4.0.

Calculations. The signal is analyzed in successive Gaussian windows of 200 milliseconds with 25% overlap in time. We adopt a frequency resolution of 100Hz for each band, resulting in 250 analyzed bands. The SNR is estimated for each frequency band, as the power ratio of the received signal and the measured noise in this band.

Fig. 5 presents the evaluated channel capacity for the entire frequency range. We can observe that for the 1m, 4m and 8m setups, the theoretical upper bound for the channel capacity is between 1200 bit/sec and 1800 bit/sec for the audible frequency bands (lower than 18kHz). As expected, the channel capacity is correlated with the distance between the transmitter and the receiver. The channel capacity significantly degrades in the sub-bass range (up to about 60Hz) and for high frequencies (above 18kHz). In these ranges, the theoretical upper bound is between 300 bit/sec and 600 bit/sec in most cases. The reason for that is that loudspeakers, and particularly home grade PC loudspeakers, were projected and optimized for human auditory characteristics, and therefore they are more responsive to the audible frequency ranges.
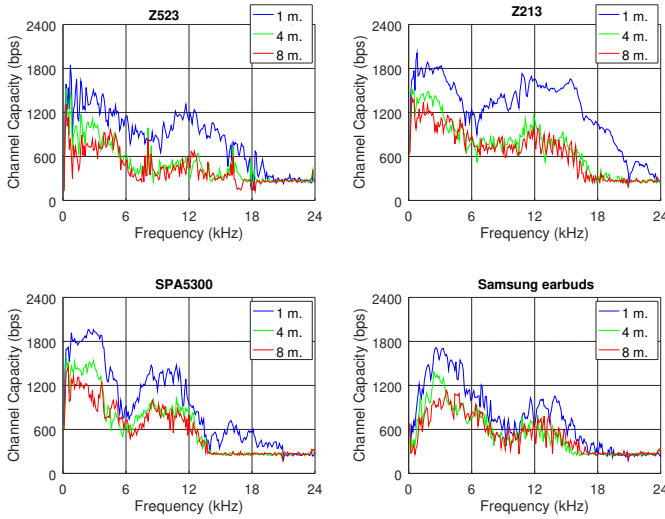
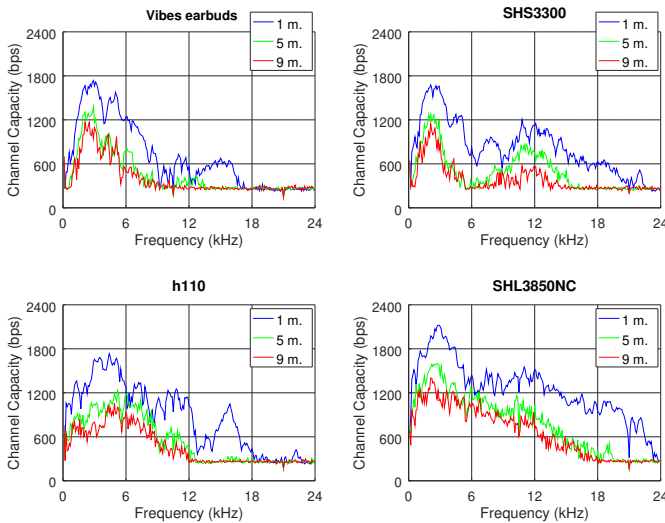Fig. 5. Channel capacity of speaker-to-speaker communication

Fig. 6. Channel capacity of speaker-to-headphones communication

1) Headphones, Earphones and Earbuds: We also calculate the capacity of a communication channel in which headphones, earphones and earbuds are used as receivers. Similar to the previous experiment, we sweep sinusoid of ten-second length in a range of 1Hz to 24kHz and record it by the headphones. We evaluate the channel capacity for one, five and eight meters. We tested four types of headphones (1) Philips vibes earbuds, (2) Philips SHS3300 earhooks, (3) Logitech h110 headphones and (4) Philips SHL3850NC headphones. The headphones were connected to a retaskable audio output jack on the desktop PC described above.

The results are shown in Fig. 6. We observed that loudspeakers do not perform significantly better as receivers than earbuds or headphones, as one could expect. In particular, for the 1m, 5m, and 8m setups, the theoretical upper bound for the channel capacity is between 300 bit/sec and 600

bit/sec in most cases.

VII. COUNTERMEASURES

In highly secure facilities it is common practice to forbid the use of any types of loudspeakers (passive or active) to create so-called audio-gap separation between computers [17]. Less restrictive policies prohibit the use of microphones but allow one-way loudspeakers. Such a policy was suggested by the NSTISSAM TEMPEST/2-95, RED/BLACK guide [1]. Accordingly, some TEMPEST certified loudspeakers are shipped with amplifiers and one-way fiber input [7]. However, the aforementioned policies and protective countermeasures are not relevant to most modern headphones, which are primarily non-powered, and built without amplifiers. A general solution for all kinds of speakers and headphones is to implement the amplifier on-board, integrating it within the audio chipset. A different approach is to mask ultrasonic transmissions in certain area by using ultrasonic jammers. These devices generate ultrasonic background noise aimed at interfering with the covert communication signals. ote that this type of solution is not trivial to deploy on a wide scale since the jamming range is limited to a radius of a few meters to a single room. The jamming efficacy also depends on the distance from the potential transmitters and receivers. It is possible to monitor the audio channel for abnormally peaks of energy, in order to detect hidden transmissions in the area. In our case, the ultrasonic frequency range above 18kHz should be scanned (continuously) and analyzed. However, if the hardware device scanning the ultrasonic spectrum is far from the transmitter this approach may not be effective. Software countermeasures include completely disabling the audio hardware in the UEFI/BIOS. This can prevent a malware from accessing the audio codec from the operating system level. However, such a configuration eliminates the use of the audio hardware (e.g., for playing audio), and hence, may not be feasible in all cases. Another option is to install a HD audio driver that prevents jack retasking or enforces a strict jack retasking policy. To provide general software-level protection, anti-malware and intrusion detection systems can employ a monitoring driver which detects unauthorized speaker-to-mic retasking operations and block them. Another approach proposed by [30] is to filter out the inaudible frequencies at the range of 18kHz and higher with a low-pass or bandpass filter.

VIII. CONCLUSION

It is known that covert communication can be established between two nearby air-gapped computers, enabling them to communicate to one another via ultrasonic waves [30]. However, the standard attack model requires the two computers to be equipped with both speakers and microphones. Consequently, this type of covert channel is not applicable in secure facilities where it is common practice to prohibit the use of microphones [11]. Also, many desktop workstations lack microphones or the microphones have been physically

muted or turned off [3]. In this work, we show how air-gapped computers without microphones can still exchange data via ultrasonic waves. The computers must be connected to passive speakers, headphones, or earphones. Our method is based on the capability of a malware to transform a PC's connected speaker from an output device to an input device, unobtrusively changing its role from a speaker to a microphone [26]. We show that although the reversed speakers are not designed to function as microphones, they are still sensitive to high frequency sound waves at a range of 18kHz to 24kHz. Transmissions in this range are practically inaudible to most adults, and hence this channel is considered covert. We evaluate the communication channel and present two attack scenarios: (1) speaker-to-speaker communication, and (2) speaker-to-headphones communication. Our results show that by using loudspeakers, data can be exchanged over an air-gap from a distance of eight meters away with effective bit rates of 10 - 166 bit/sec.

REFERENCES

[1] Nstissam tempest/2-95. https://cryptome.org/tempest-2-95.htm, 2000. (Accessed on 02/27/2018).

[2] funtenna github. https://github.com/funtenna, 2015. (Accessed on 12/03/2017).

[3] Why has mark zuckerberg taped over the webcam and microphone on his macbook? https://www.telegraph.co.uk/technology/2016/06/22/why-has-mark-zuckerberg-taped-over-the-webcam-and-microphone-on/, 06 2016. (Accessed on 02/27/2018).

[4] Wikileaks: Cia uses 'brutal kangaroo' toolkit to hack air-gapped networks. https://www.theinquirer.net/inquirer/news/3012499/-wikileaks-cia-uses-brutal-kangaroo-toolkit-to-hack-air-gapped-networks, 2017. (Accessed on 12/03/2017).

[5] Powered speakers - wikipedia. https://en.wikipedia.org/wiki/Powered_speakers, 2018. (Accessed on 02/27/2018).

[6] Praat: doing phonetics by computer. http://www.fon.hum.uva.nl/praat/, 2018. (Accessed on 02/27/2018).

[7] Tempest video solutions — amplified speaker - fiber. http://www.cissecure.com/products/tempest-amplified-speaker-fiber, 2018. (Accessed on 02/27/2018).

[8] Sherly Abraham and InduShobha Chengalur-Smith. An overview of social engineering malware: Trends, tactics, and implications. *Technology in Society*, 32(3):183–196, 2010.

[9] Glen Ballou. *Handbook for sound engineers*. Taylor & Francis, 2013.

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

[11] Josh Dean. Jumping the airgap. https://thoughtworksnc.com/2017/03/16/jumping-the-airgap/, 03 2017. (Accessed on 02/27/2018).

[12] Luke Deshotels. Inaudible sound as a covert channel in mobile devices. In *WOOT*, 2014.

[13] Ben Duncan. *High Performance Audio Power Amplifiers*. Elsevier, 1996.

[14] David Remnick Evan Osnos and Joshua Yaffa. Trump, putin, and the new cold war - the new yorker. https://www.newyorker.com/magazine/2017/03/06/trump-putin-and-the-new-cold-war, 03 2017. (Accessed on 12/03/2017).

[15] Mordechai Guri, Dima Bykhovsky, and Yuval Elovici. air-jumper: Covert air-gap exfiltration/infiltration via security cameras & infrared (ir). *arXiv preprint arXiv:1709.05742*, 2017.

[16] Mordechai Guri, Andrey Daidakulov, and Yuval Elovici. Magneto: Covert channel between air-gapped systems and nearby smartphones via cpu-generated magnetic fields. *arXiv preprint arXiv:1802.02317*, 2018.

[17] Mordechai Guri and Yuval Elovici. Bridgeware: The air-gap malware. *Communications of the ACM*, 61(4):74–82, 2018.

[18] Mordechai Guri, Ofer Hasson, Gabi Kedma, and Yuval Elovici. An optical covert-channel to leak data through an air-gap. In *Privacy, Security and Trust (PST), 2016 14th Annual Conference on*, pages 642–649. IEEE, 2016.

[19] Mordechai Guri, Assaf Kachlon, Ofer Hasson, Gabi Kedma, Yisroel Mirsky, and Yuval Elovici. Gsmem: Data exfiltration from air-gapped computers over gsm frequencies. In *USENIX Security Symposium*, pages 849–864, 2015.

[20] Mordechai Guri, Gabi Kedma, Assaf Kachlon, and Yuval Elovici. Airhopper: Bridging the air-gap between isolated networks and mobile phones using radio frequencies. In *Malicious and Unwanted Software: The Americas (MALWARE), 2014 9th International Conference on*, pages 58–67. IEEE, 2014.

[21] Mordechai Guri, Matan Monitz, and Yuval Elovici. Usbee: Air-gap covert-channel via electromagnetic emission from usb. In *Privacy, Security and Trust (PST), 2016 14th Annual Conference on*, pages 264–268. IEEE, 2016.

[22] Mordechai Guri, Matan Monitz, and Yuval Elovici. Bridging the air gap between isolated networks and mobile phones in a practical cyber-attack. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 8(4):50, 2017.

[23] Mordechai Guri, Matan Monitz, Yisroel Mirski, and Yuval Elovici. Bitwhisper: Covert signaling channel between air-gapped computers using thermal manipulations. In *Computer Security Foundations Symposium (CSF), 2015 IEEE 28th*, pages 276–289. IEEE, 2015.

[24] Mordechai Guri, Yosef Solewicz, Andrey Daidakulov, and Yuval Elovici. Fansmitter: Acoustic data exfiltration from (speakerless) air-gapped computers. *arXiv preprint arXiv:1606.05915*, 2016.

[25] Mordechai Guri, Yosef Solewicz, Andrey Daidakulov, and Yuval Elovici. Acoustic data exfiltration from speakerless air-gapped computers via covert hard-drive noise (diskfiltration). In *European Symposium on Research in Computer Security*, pages 98–115. Springer, 2017.

[26] Mordechai Guri, Yosef Solewicz, Andrey Daidakulov, and Yuval Elovici. Speake(a)r: Turn speakers to microphones for fun and profit. In *11th USENIX Workshop on Offensive Technologies (WOOT 17)*. USENIX Association, 2017.

[27] Mordechai Guri, Boris Zadov, Andrey Daidakulov, and Yuval Elovici. xled: Covert data exfiltration from air-gapped networks via router leds. *arXiv preprint arXiv:1706.01140*, 2017.

[28] Mordechai Guri, Boris Zadov, Andrey Daidakulov, and Yuval Elovici. Odini: Escaping sensitive data from faraday-caged, air-gapped computers via magnetic fields. *arXiv preprint arXiv:1802.02700*, 2018.

[29] Mordechai Guri, Boris Zadov, and Yuval Elovici. *LED-it-GO: Leaking (A Lot of) Data from Air-Gapped Computers via the (Small) Hard Drive LED*, pages 161–184. Springer International Publishing, Cham, 2017.

[30] Michael Hanspach and Michael Goetz. On covert acoustical mesh networks in air. *arXiv preprint arXiv:1406.1213*, 2014.

[31] David Henningsson. Turn your mic jack into a headphone jack! a better sounding world. http://voices.canonical.com/david.henningsson/2011/11/29/turn-your-mic-jack-into-a-headphone-jack/, 11 2011. (Accessed on 02/27/2018).

[32] Joe Loughry and David A Umphress. Information leakage from optical emanations. *ACM Transactions on Information and System Security (TISSEC)*, 5(3):262–289, 2002.

[33] Anil Madhavapeddy, Richard Sharp, David Scott, and Alastair Tse. Audio networking: the forgotten wireless technology. *IEEE Pervasive Computing*, 4(3):55–60, 2005.

[34] Nikolay Matyunin, Jakub Szefer, Sebastian Biedermann, and Stefan Katzenbeisser. Covert channels using mobile device's magnetic field sensors. In *Design Automation Conference (ASP-DAC), 2016 21st Asia and South Pacific*, pages 525–532. IEEE, 2016.

[35] Vasilios Mavroudis, Shuang Hao, Yanick Fratantonio, Federico Maggi, Christopher Kruegel, and Giovanni Vigna. On the privacy and security of the ultrasound ecosystem. *Proceedings on Privacy Enhancing Technologies*, 2017(2):95–112, 2017.

[36] Mark Maybury, Penny Chase, Brant Cheikes, Dick Brackney, Sara Matzner, Tom Hetherington, Brad Wood, Conner Sibley, Jack Marin, and Tom Longstaff. Analysis and detection of malicious insiders. Technical report, MITRE CORP BEDFORD MA, 2005.

[37] Kim McCoy, Beatrice Tomasi, and Giovanni Zappa. Janus: The genesis, propagation and use of an underwater standard. *Proc. ECUA 2010*, 2010.

[38] Raphael Rom and Moshe Sidi. *Multiple access protocols: performance and analysis*. Springer Science & Business Media, 2012.

[39] William Stallings. *Data and computer communications*. Pearson Education India, 2007.

[40] Paul Vitello. A ring tone meant to fall on deaf ears. *The New York Times*, 12, 2006.