CapSpeaker: Injecting Voices to Microphones via Capacitors

Xiaoyu Ji
USSLab, Zhejiang University
xji@zju.edu.cn

Juchuan Zhang
USSLab, Zhejiang University
juchuanzhang@zju.edu.cn

Shui Jiang
USSLab, Zhejiang University
3180104945@zju.edu.cn

Jishen Li
USSLab, Zhejiang University
3180105960@zju.edu.cn

Wenyuan Xu*
USSLab, Zhejiang University
xuwenyuan@zju.edu.cn

ABSTRACT

Voice assistants can be manipulated by various malicious voice commands, yet existing attacks require a nearby speaker to play the attack commands. In this paper, we show that even when no speakers are available, we can play malicious commands by utilizing the capacitors inside electronic devices, i.e., we convert capacitors into speakers and call it `CapSpeaker`. Essentially, capacitors can emit acoustic noises due to the inverse piezoelectric effect, i.e., varying the voltage across a capacitor can make it vibrate and thus emit acoustic noises. Forcing capacitors to play malicious voice commands is challenging because (1) the frequency responses of capacitors as speakers have poor performance in the range of audible voices, and (2) we have no direct control over the voltage across capacitors to manipulate their emitting sounds. To overcome the challenges, we use a PWM-based modulation scheme to embed the malicious audio onto a high-frequency carrier, e.g., above 20 kHz, and we create malware that can induce the right voltage across the capacitors such that `CapSpeaker` plays the chosen malicious commands. We conducted extensive experiments with 2 LED lamps (a modified one and a commercial one) and 5 victim devices (iPhone 4s, iPad mini 5, Huawei Nova 5i, etc.). Evaluation results demonstrate that `CapSpeaker` is feasible at a distance up to 10.5 cm, triggering a smartphone to receive voice commands, e.g., "open the door".

CCS CONCEPTS

• **Security and privacy** → *Embedded systems security*; *Malicious design modifications*; Mobile platform security.

KEYWORDS

IoT security, ASR security, malicious voice commands, capacitor sounds, voice injection

ACM Reference Format:
Xiaoyu Ji, Juchuan Zhang, Shui Jiang, Jishen Li, and Wenyuan Xu. 2021. CapSpeaker: Injecting Voices to Microphones via Capacitors. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security (CCS '21), November 15–19, 2021, Virtual Event, Republic of Korea.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3460120.3485389

*Corresponding author.

CCS '21, November 15–19, 2021, Virtual Event, Republic of Korea

ACM ISBN 978-1-4503-8454-4/21/11…$15.00
https://doi.org/10.1145/3460120.3485389

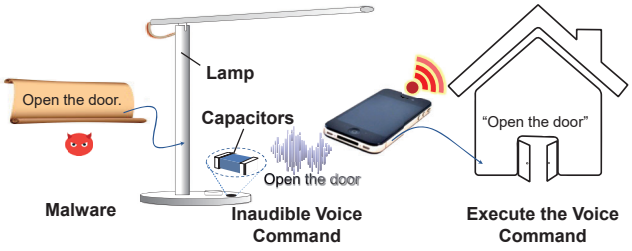


Figure 1: Attack scenario of `CapSpeaker`. The malware inside an LED lamp can manipulate the voltage across the built-in MLC capacitors of the LED lamp to produce malicious voice commands, which can further trigger a voice assistant to execute the commands such as "open the door".

1 INTRODUCTION

Capacitors are ubiquitous and indispensable components in electronic devices since they are used for voltage stabilization, filtering, etc [55]. Particularly, Multi-layer Ceramic (MLC) capacitors [46] are dominant due to their high energy density and low cost, and the number of MLC capacitors consumed per year is reported to be approximately one trillion (10^{12}) pieces [38]. MLC capacitors are known to create annoying but benign high-pitched noises, yet no one has been reported to produce voices via such capacitors, to the best of our knowledge. In this paper, we investigate the feasibility of utilizing commodity electronic devices with built-in capacitors to inject malicious voice commands into voice assistants, such as Apple Siri [11], Xiaomi Art Speaker [14], and potentially allowing an attacker to open the door sneakily.

Unlike existing work that injects malicious voice commands into voice assistants via a loudspeaker [19, 49, 50, 62, 63], we propose `CapSpeaker` that can inject voice commands by converting an electronic device (e.g., a lamp) that are not designed to produce voice into a speaker. One question is "*How can a capacitor produce voices?*" The underlying physics principle shows that capacitors can emit acoustic noises due to the inverse piezoelectric effect of ceramic materials [54], i.e., the voltage across a capacitor causes the capacitor to vibrate at the same frequency as the voltage signal. Thus, a capacitor can produce sounds in a similar way as a speaker, i.e., by converting currents into an acoustic signal.

To produce malicious voice commands, we envision that during the manufacturing phase an attacker can install the malware in an

LED lamp, as shown in Fig. 1. The program can be a modified version of a normal application, e.g., a simple brightness control application, with abilities to induce the 'right' voltage across capacitors inside the lamp circuits and to make the capacitors produce malicious voices with "open the door" embedded. As a result, CapSpeaker can opportunistically trigger the nearby voice assistants inside a smartphone to execute an unexpected command [1].

Promising yet challenging, the design of CapSpeaker faces two challenges. 1) How to play human voices? The frequency of human voices is mainly below 4 kHz [40], but the capacitors' frequency response[2] below 4 kHz is too low to produce voices. The peak frequency response of capacitors ranges between 10 kHz and 90 kHz, within which capacitors can produce a loud sound but is out of the main human voice range. 2) How to continuously control the voltage across capacitors to produce the desired voices? Commercial devices are typically well packaged and have no access control interfaces to interact with a capacitor.

To address the above challenges, we first perform experiments and simulations to analyze the factors that determine the strength of produced sounds. The experiments involve MLC capacitors connected with a switch controlled by a signal generator, whereby we have control over the frequency and the amplitude of the voltage across the capacitor. We found that the sound strength and the frequency of the MLC capacitor are directly determined by the amplitude and the frequency of voltage change across the capacitor but are almost unaffected by the capacitance or the capacitor packaging type. Our simulation of MLC capacitors soldered on a printed circuit board (PCB) with COMSOL [3] shows that the PCB sizes and materials, capacitor positions, and numbers will affect the produced sound strength as well. Thus, given an electronic device, attackers shall be able to produce desired voice with carefully crafted voltage signals.

To inject human voices by CapSpeaker despite the low frequency response below 4 kHz, we modulate the malicious voice commands on a carrier of a high-frequency band, e.g., $\geq$ 20 kHz, where the capacitors have the peak frequency response and produce a loud voice. The benefit of such a solution is that CapSpeaker can produce inaudible voice commands, but the challenge is to have the victim device demodulate the voice and extract the embedded command. By utilizing the nonlinearity of the microphone [63], we envision that the modulated commands can be demodulated and extracted. To make the attack feasible, the carrier frequency has to be carefully chosen by balancing the trade-off between the frequency response of the capacitors and the peak nonlinearity of the microphone, and we model the selection of carrier frequency as an optimization problem. Unable to directly change the voltage across capacitors, we design a program-level control mechanism to indirectly manipulate the voltages, i.e., we rely on high-level programming instructions to control the peripheral load of a device. Additionally, we found it is impossible to modulate the voltage levels continuously via amplitude modulation due to hardware constraints, we employ pulse width modulation (PWM) to control the average voltage and thus the strength of the voices continuously. Thus, we manage to produce the malicious voice commands by simply running malware on an electronic device.

To evaluate the practical performance with various factors, we implemented and evaluated CapSpeaker with extensive experiments by utilizing both a modified and a commercial LED lamp to attack various devices at multiple distances and voice commands. In total, we studied 5 victim devices and show it is feasible to inject a voice command of "Open the door" at a distance of 10.5 cm into an iPhone 4s. In short, our contributions are summarized as follows:

- We validate the feasibility of producing voices via capacitors and empirically study the factors that affect the strength and frequency of the produced voices.
- We design CapSpeaker that relies on malware to enable capacitors to produce malicious voice commands. The malicious voice commands adopt a PWM modulation mechanism to overcome the limitation imposed by the frequency response of capacitors and have the benefit of being inaudible to humans.
- We evaluate the performance of CapSpeaker by using an LED lamp against 5 victim devices including iPhone 4s, iPad mini 5, iWatch, Huawei Nova 5i Pro, and Redmi K30 Ultra. The experiment results show that CapSpeaker can fool the automatic speech recognition (ASR) of an iPhone 4s to execute a command of "Open the door" at a distance of 10.5 cm.

2 SOUND PRINCIPLE OF MLC CAPACITORS

In this section, we present the mechanical structures of a capacitor to illustrate the underlying principle of producing sounds, and discuss the factors that affect the sound strength of the capacitors with COMSOL [3] simulation and experiments. To understand command extraction at the microphones, we present the background knowledge of microphones' nonlinearity that enables the demodulation of the injected voice commands.

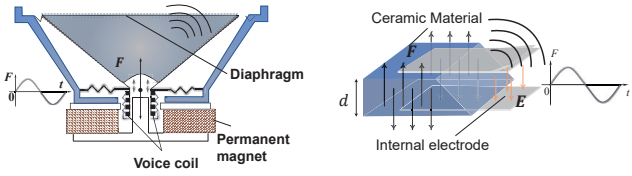
2.1 How Can Capacitors Produce Voices

The mechanical structures of an MLC capacitor and a speaker share similarities, and thus a capacitor can produce voices.

2.1.1 How Do Speakers Play Voices? Fundamentally, a speaker converts the input electrical signals into mechanical vibrations to produce voices. Most commercial devices, e.g., smartphones, laptops, and smart speakers, use electrodynamic speakers, which are mainly composed of a diaphragm, a voice coil, a permanent magnet, and a bracket, as shown in Fig. 2a. When a current passes through the voice coil, it produces a magnetic field that interacts with the magnetic field of the permanent magnet, and an Ampere force is exerted from the stable permanent magnet to the free-moving voice coil. Thus, an AC current that flows through the voice coil will drive the diaphragm to vibrate and convert the motion into sound pressure level (SPL). Suppose the magnetic induction intensity generated by the magnet is B, the current flowing in the coil is I, the number of turns of the coil is N, the circumference of the coil is L, the DC and AC components of the current are I_{DC} and I_{AC} respectively. Then, the Ampere force applied to the coil can be expressed as:

$$F = NBIL = NBL(I_{DC} + I_{AC}cos2\pi ft) \quad (1)$$

[1] Demo video is available at https://github.com/USSLab/CapSpeaker.
[2] The quantitative measure of the output in response to a given input frequency.



(a) A speaker produces sounds because the diaphragm attached to the voice coil is vibrated when a current is flowing across.

(b) An MLC capacitor produces sounds because the capacitor is vibrated under the inverse piezoelectric effect when an electric field is applied to it.

Figure 2: The principle of producing sounds for a speaker and an MLC capacitor respectively.

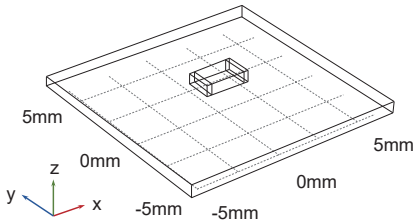


Figure 3: The COMSOL simulation layout of an MLC capacitor soldered on a PCB.

Only the AC component causes the coil to vibrate, and the vibration frequency f is the same as the frequency f of the current signal applied to the voice coil, generating sounds of frequency f.

2.1.2 How Do Capacitors Produce Sounds? Electronic devices, e.g., laptops, smartphones, LED lamps, are equipped with AC/DC or DC/DC power electronic components, resulting in inevitable high-frequency interference such as ripple signals. Capacitors play a significant role in filtering unnecessary signals and stabilizing voltage supply by alleviating voltage overshoot or undershoot [23]. With the trend of high compactness of nowadays electronic devices, MLC capacitors [46] have become dominant in electronic devices due to their low costs and small volumes [43].

MLC inverse piezoelectric. An MLC capacitor is a fixed-value capacitor using ceramic material as dielectric. As shown in Fig. 2b, an MLC capacitor consists of multiple alternating layers of ceramics and metal, acting as dielectric layers and electrodes respectively. The ceramics have the inverse piezoelectric characteristics [43], i.e., applied electric fields can induce mechanical strain force [3]. Without loss of generality, denote the mechanical strain force as $F = pE$, where p is piezoelectric constant, and E is the electric field, respectively. E can be calculated by $E = U/d$, where d is the distance between the two electrodes, and U is the voltage across the capacitor. U contains both a DC component U_{DC} and an AC component, i.e., $U_{AC}cos2\pi ft$, and therefore the deduced mechanical strain force upon an MLC capacitor can be expressed as:

$$F = \frac{pU}{d} = \frac{p}{d}(U_{DC} + U_{AC}cos2\pi ft) \quad (2)$$

[3]The piezoelectric is the electric charge that accumulates in solid materials (e.g., crystals, ceramics) in response to applied mechanical stress.

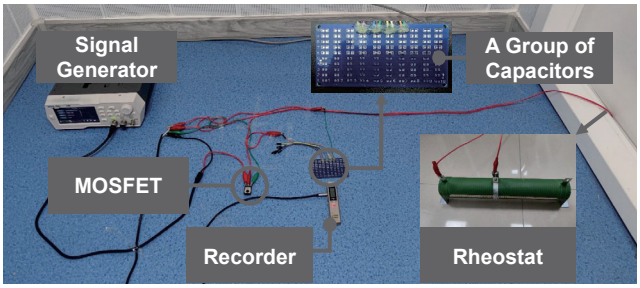


Figure 4: The real-world experimental setup to study the input signal frequency, capacitance, and package factors.

Similar to the Ampere force denoted by Equ. 1, the mechanical strain force F in an MLC capacitor contains components of frequency f of the voltage signals, which drives the MLC capacitor to vibrate with the frequency f. In addition, F is proportional to U_{AC}. Given that the voltage across the capacitor has one single frequency, we have $\Delta V = 2U_{AC}$, where ΔV is the peak-to-peak amplitude of the voltages across the capacitor, i.e., the difference between the highest and lowest voltages.

Remark. An MLC capacitor can produce sounds because of the inverse piezoelectric effect. The frequency of the sounds depends on the frequency of the voltage signal applied on the capacitor and its strength is determined by the peak-to-peak amplitude of the voltage, the piezoelectric constant p, and the printed circuit board (PCB) layout, etc. We will elaborate on those influential factors in the following section.

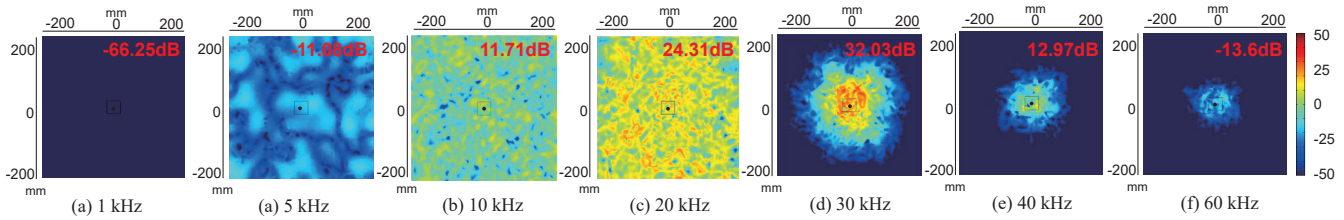
2.2 Impact Factors of MLC Capacitor Sound

In commercial products, MLC capacitors are soldered to a circuit board. In addition to the piezoelectric constant, the produced sounds can be affected by the PCB layouts, the MLC capacitors, and the applied voltage signal. To quantify those impacts, we conducted both simulation in COMSOL software [3], and real-world experiments. The validated factors include *voltage frequencies, PCB sizes, MLC capacitor positions, numbers of MLC capacitors, and PCB materials* as shown in Tab. 1.

2.2.1 COMSOL simulation. First, we study the factors that are difficult to control experimentally by utilizing COMSOL, popular simulation software for physics-based simulation tools, and the factors include the PCB sizes, capacitor locations, numbers of capacitors, and PCB materials.

Table 1: Simulation parameters of the MLC capacitor soldered on a PCB.

Impact factors	Parameters	Default	Additional
Applied voltage signal	ΔV (V)	20	4-20
Capacitor	Capacitance (nf)	1000	0.1-1000
	Package	0805	0603, 1206
PCB	Size (cm)	4*4	1*1, 10*10
	Material (MPa)	150	15, 1500
Capacitor location on PCB		(0,0)	(0, -1), (-1, -1), (-1, 0)



(a) 1 kHz (a) 5 kHz (b) 10 kHz (c) 20 kHz (d) 30 kHz (e) 40 kHz (f) 60 kHz

Figure 5: Frequency response of the MLC capacitor soldered on a PCB via COMSOL simulation. The number in red indicates the maximum sound strength for each frequency.

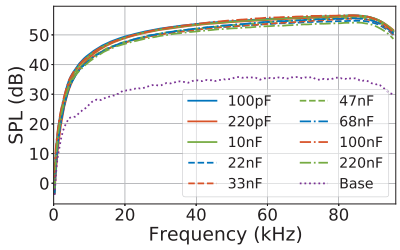


Figure 6: Frequency response vs. capacitance via real-world experiment. "Base" is the noise from other devices.

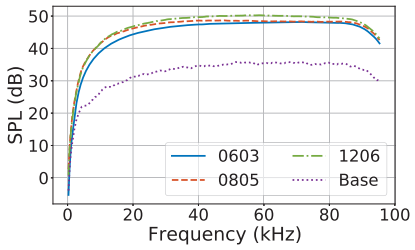


Figure 7: Frequency response vs. package via real-world experiment. "Base" is the noise from other devices.

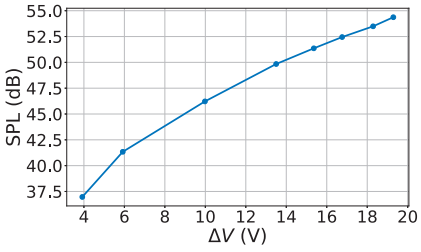


Figure 8: SPL of MLC capacitors with various load levels at 30 kHz. A higher voltage change means a higher load level.

Setup. We build a finite element model to simulate the scenarios of a capacitor welded on a PCB, as shown in Fig. 3. An MLC capacitor, which is denoted by a cube of 0805 package size [12] in the middle, is wrapped between two copper plates that are welded on the PCB using high-leaded tin alloy solder joints. All these components are surrounded by an air area, which is a cube with a side length of 40 cm. These materials and setup are supported by the COMSOL inner material library. For each simulation, we excite the electric field with a voltage, resulting in the vibration of piezoelectric ceramics. The vibration of piezoelectric ceramics drives the vibration of PCB through solder joints. These solid substances will push the surrounding air as the boundary of structural acoustics, and finally, produce sounds. We measure the sound field distribution up to 5 cm away from the board with the default setup parameters listed in Tab. 1: The capacitor electrolyte is Lead Zirconate Titanate (PZT-4), the input current frequency is 30 kHz, the size of the PCB is 4*4 cm with the capacitor located in the center of the board, and the material of PCB is glass wool board.

2.2.2 Real-world experiments. In addition to the COMSOL simulation, we conduct experiments on MLC capacitors with various capacitance values, packages, voltages, and input signal frequencies. As shown in Fig. 4, the experimental setup consists of a circuit with a 20 V power adapter, a metal oxide semiconductor field-effect transistor (MOSFET), and a sliding rheostat set to 10 Ω. The MOSFET acts as a switch to control voltages across the capacitors to generate low/high voltage levels, and a *RIGOL DG811* [5] signal generator is used to control the MOSFET with a square wave as the control signal. To validate the frequency response of MLC capacitors, we swept from 0 Hz to 96 kHz with a step size of 88 Hz and recorded the produced sounds using an *Aigo R6611* voice recorder at a distance of 5 cm. The recorded sounds are processed by an *SYBA FG-EAU02A* sound card [2]. Note that we placed the other devices that may generate noises, e.g., the sliding rheostat and the power supply, 10 m away from the recorder to reduce their influences. The default settings and instances of the studied capacitors are listed in Tab. 1.

2.2.3 Results. According to the COMSOL simulation and experiments of the aforementioned parameters, we report the results of the input voltage signal, the MLC capacitor, and the PCB, respectively.

Applied voltage signal. We measured the sounds produced by the MLC capacitor on a PCB with various voltage frequencies in COMSOL, and the results are shown in Fig. 5. We observed that the frequency response peaks at the 30 kHz band with 32.03 dB and the ones below 5 kHz are less than −11.08 dB, indicating that it is extremely difficult, if ever possible, for MLC capacitors to produce sounds in human audible frequency bands. Note that the peak frequency response of the real-world capacitors welded on a PCB is around 80 kHz, which is different from the simulation results, as shown in Fig. 6 and Fig. 7. This is because the PCB sizes and the materials of simulation and real-world experiment setup are different.

We measured the SPL by changing the peak-to-peak magnitude of the voltage signals ΔV from 4 V to 20 V. The results (shown in Fig. 8) indicate that the SPLs of the produced sounds increase almost linearly when ΔV is larger than 6 V.

Capacitor capacitance and package. Capacitors with various capacitance and packages are evaluated in real-world experiments, and the peak-to-peak amplitude of the voltage across the capacitor is $\Delta V = 20$ V. We have the following observations. 1) For capacitor capacitance, Fig. 6 indicates that the SPLs remain almost unchanged with capacitance values. The reason may be that although the MLC capacitors with higher capacitance have extra layers, the force from the inverse piezoelectric effect is mainly decided by the peak-to-peak amplitude of the applied voltages. 2) For MLC capacitor

packages, i.e., 0603, 0805, and 1206 [12] [4], Fig. 7 shows that the SPLs of capacitors with various packages remain almost unchanged, i.e., at most 2.5 dB at frequencies larger than 20 kHz.

PCB sizes and materials. To study whether the PCB sizes and materials will amplify the produced sounds by capacitors in COMSOL, we choose three typical PCB sizes (i.e., 1*1 cm, 4*4 cm, and 10*10 cm) and simulate PCB materials with various elasticities, while applying a 30 kHz voltage signal. We have the following observations. 1) Results of PCB sizes, as shown in Fig. 9, show that the 1*1 cm board produces the strongest sound while the 10*10 cm board produces the weakest sound. This is because the vibration from capacitors is amplified at a larger amplitude by a smaller PCB and vice versa. 2) To simulate PCB with various elasticities, we changed Young's modulus [59] of the PCB. A higher Young's modulus maps to a less elastic board, i.e., harder. The results in Fig. 10 indicate that the higher elastic the material is the stronger the sound produced by the PCB.

Capacitor location on PCB. In the COMSOL simulation, we place the MLC capacitor at various positions of the PCB, e.g., (0,0) cm, (0,-1) cm, (-1,-1) cm, and (-1, 0) cm, where the coordinate origin is the center of the PCB. The results in Fig. 11 show that the locations of the capacitor will affect the SPLs. Interestingly, the SPLs produced when the capacitor is located in the left-bottom and the center of the PCB are higher than the ones of middle-bottom and middle-left locations.

Remark. In summary, although the capacitance and packages of capacitors, as well as the PCB properties, of a device may affect the produced sound more or less, for a given electronic device these factors are fixed. Thus, *the sound strength produced by capacitors is fundamentally determined by the input voltage signal.* To boost the SPL of the produced sounds, the voltage signal should be carefully designed to match the frequency response of the capacitors.

2.3 Voice Assistant and Its Vulnerability

The nonlinearity of microphones. A microphone is a nonlinear component, whereby its output contains square and higher-order items of the input, and thus it can produce harmonics and cross-products. The utilization of such a hardware property is first reported by DolphinAttack [49, 50, 63]. The work utilizes an AM modulation to create malicious voice commands embedded in an ultrasonic carrier. Formally, suppose a base-band signal is $m(t) = cos(2\pi f_m t)$, the AM carrier frequency is f_c, then the modulated signal fed into the microphone can be expressed as:

$$s_{in} = m(t)cos(2\pi f_c t) + cos(2\pi f_c t) \quad (3)$$

With the help of the nonlinearity of the microphone, the output signal contains f_m and the linear terms of f_c, $f_c - f_m$, $f_c + f_m$. After the low pass filter, the output of the voice signal only contains f_m. Therefore, an attacker can successfully inject voice commands into voice assistants with inaudible ultrasonic signals.

3 THREAT MODEL

The goal of an attacker is to utilize `CapSpeaker`, i.e., the capacitors in electronic devices, to inject malicious voice commands into voice assistants on smartphones, smartwatches, and smart speakers, e.g., Siri, Google Now, Amazon Echo. Note that `CapSpeaker` may be leveraged to inject voices into other applications that involve microphones, e.g. fooling audio/video conversation, telephone calls. Nevertheless, in this paper, we focus on attacking voice assistants with the following assumptions.

Malware injection. The attacker can install malware or tamper with the firmware of the device (e.g., an LED lamp) to manipulate its load, i.e., its power consumption. The malware can be embedded inside a normal application with a hidden function, e.g. maliciously manipulate the LED brightness in the brightness control program of the LED lamp. Once the device is turned on and the program is executed, desired sounds are generated and injected into nearby devices to cause malicious consequences.

No direct access to the victim's voice assistant. We assume that the attacker has no direct access to the victim's voice assistant. She cannot install malware, change the device settings, nor physically touch it. However, she is aware of the device and the embedded microphone models of the voice assistant, and thus she can obtain a device of the same model to collect the necessary information to ensure a successful attack.

Proximity to the victim's voice assistant. Without loss of generality, we assume that the victim's voice assistant device (e.g., a smartphone) is placed close to the device with capacitors, e.g., the LED lamp.

4 DESIGN OF CAPSPEAKER

4.1 Overview

`CapSpeaker` is essentially malware preinstalled in an electronic device. For a given malicious voice command, `CapSpeaker` will induce the carefully crafted voltage signals across the capacitors on the PCB inside the device by executing high-level programming instructions. Particularly, `CapSpeaker` consists of two modules: a signal crafting module and a GPIO controlling module, whereby the signal crafting module creates an intermediate signal that contains the malicious voice command yet can drive the GPIO controlling module to induce the capacitors to produce a sound with the chosen voice command. As shown in Fig. 12, for a given command of "open the door", `CapSpeaker` inside the LED lamp will perform the following.

1) The signal crafting module chooses the desired carrier frequency that eventually can induce a loud voice from the capacitors and modulates the baseband signals of the voice command onto the carrier.
2) The voltage across the capacitors is indirectly controlled by the General Purpose Input/Output (GPIO), which is the standard interface connecting microcontrollers to the peripheral devices. Thus, a GPIO controlling module switches the GPIO that controls the LED lamp based on the modulated signals generated by the signal crafting module, inducing the desired voltage across the capacitors.

4.2 Crafting the Attack Signal

The signal crafting module addresses two issues: 1) Selecting a proper carrier frequency for modulation to make the attack signal

[4] The number indicates the physical size of the capacitor, a 0805-packaged capacitor, for example, means its length and width are 0.08 inch and 0.05 inch respectively.

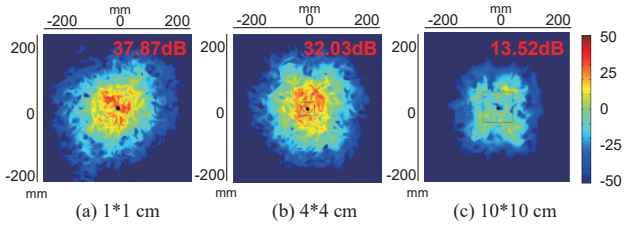


Figure 9: The MLC capacitor sound field vs. PCB size simulated with COMSOL.

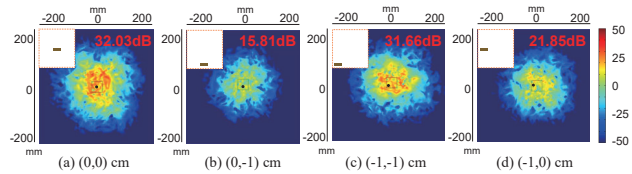


Figure 11: The MLC capacitor sound field vs. capacitor locations on PCB via COMSOL simulation.

inaudible yet match the frequency response of the capacitors. 2) Utilizing a feasible modulation scheme to modulate the voice command onto the selected carrier frequency, considering the limitation of device hardware and software.

4.2.1 Selecting the Carrier Frequency. The carrier frequency directly determines the attack performance of CapSpeaker, and it should maximize the signal-to-noise ratio (SNR) of the voice commands received by the victim's voice assistant while ensuring that the generated voice commands remain inaudible to humans. Denote the carrier frequency by f_{PWM}, and factors to be considered include the frequency response of capacitors, the nonlinearity property of the microphones, and the inaudibility of the attacks.

Frequency responses of capacitors. We define the frequency response of capacitors as the voice signal amplitude in response to a single frequency voltage signal of a peak-to-peak amplitude. Following the findings in both the COMSOL simulation (in Fig. 5) and the real-world experiments (in Fig. 6, Fig. 7), the frequency responses for capacitors are extremely weak in human audible frequency bands and a strong frequency response occur above 20 kHz. For example, consider the frequency response curve obtained by real-world experiments on a capacitor of 1uf, depicted as the orange line in Fig. 13, then the choice of a carrier frequency should be above 20 kHz and below 80 kHz.

The nonlinearity of microphones. CapSpeaker exploits the nonlinearity of the microphone to demodulate the inaudible voice command. However, the frequency response of the nonlinearity property varies with frequency. To verify the frequency response of the microphone after nonlinearity demodulation, we experimentally tested the nonlinearity property of microphones on iPhone 4s, iWatch, and Redmi K30 Ultra, respectively, using a ViFa UltraSoundGate [13]. We used a sinusoidal wave as the baseband signal and modulated it on carriers of various frequencies. For the iPhone 4s smartphone, we plot the normalized strength of the received sound after demodulation with increasing carrier frequencies in

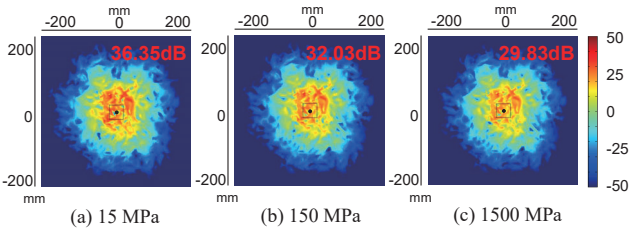


Figure 10: The MLC capacitor sound field vs. PCB material simulated with COMSOL.

Fig. 13 as a blue line. The trend is that the received signal strength decreases with the increase of the carrier frequency, and the frequency bands of [2 kHz,20 kHz], and [27 kHz,38 kHz] perform well with their normalized strength larger than 0.6. This finding is consistent with that reported in DolphinAttack [63].

Inaudibility. In addition, the selection of carrier frequency should guarantee the stealthiness of the attack. Since the frequency band of audible sound is between 20 Hz and 20 kHz, as shown in the shaded gray area of Fig. 13, the carrier frequency should be above 20 kHz to ensure inaudibility.

In summary, CapSpeaker has to simultaneously consider frequency responses of capacitors, the nonlinearity property of the victim's microphone, and the audibility to find the feasible carrier frequency. For example, according to the data of iPhone 4s plotted in Fig. 13, the feasible carrier frequency shall range from 27 kHz to 38 kHz. In reality, an attacker can obtain the information on the LED lamp and typical frequency ranges of victim devices in advance.

4.2.2 PWM-based Modulation. After selecting the proper carrier frequency, a modulation scheme is needed to modulate the baseband signal onto the carrier. Typical modulation schemes include amplitude-modulation (AM), frequency-modulation (FM), etc, where the AM modulation is popular and is used in DolphinAttack [63] to generate inaudible voice commands. However, CapSpeaker cannot use the AM modulation scheme because the victim device, i.e., the LED lamp or other IoT devices, is unable to output signals of fine-grained amplitudes. Instead, the controllable general purpose input/output (GPIO) can only output two levels, i.e., a "low" level or a "high" level, which are often 0 V and 5 V for commercial smart devices. Therefore, we resort to the pulse-width modulation (PWM) scheme which only needs two levels of output voltage. The principle of PWM modulation is illustrated in Fig. 14, where the average value of the baseband signal is controlled by switching on and off at a fast rate and thus modulated onto a carrier frequency. In our implementation, the switch is controlled by the GPIO output pin with only a low or high voltage level.

The voice command signal and the modulated signal are plotted in Fig. 12 in both time and frequency domains. After the PWM modulation, the deduced signal includes the baseband signal f_m, the carrier signal f_{PWM}, and a few harmonic frequencies such as $f_{PWM}+f_m$, $f_{PWM}-f_m$, $f_{PWM}+2f_m$, $f_{PWM}-2f_m$, etc. Note that the baseband f_m exists in this stage but has extremely low amplitude due to the poor frequency response of capacitors at low frequencies.

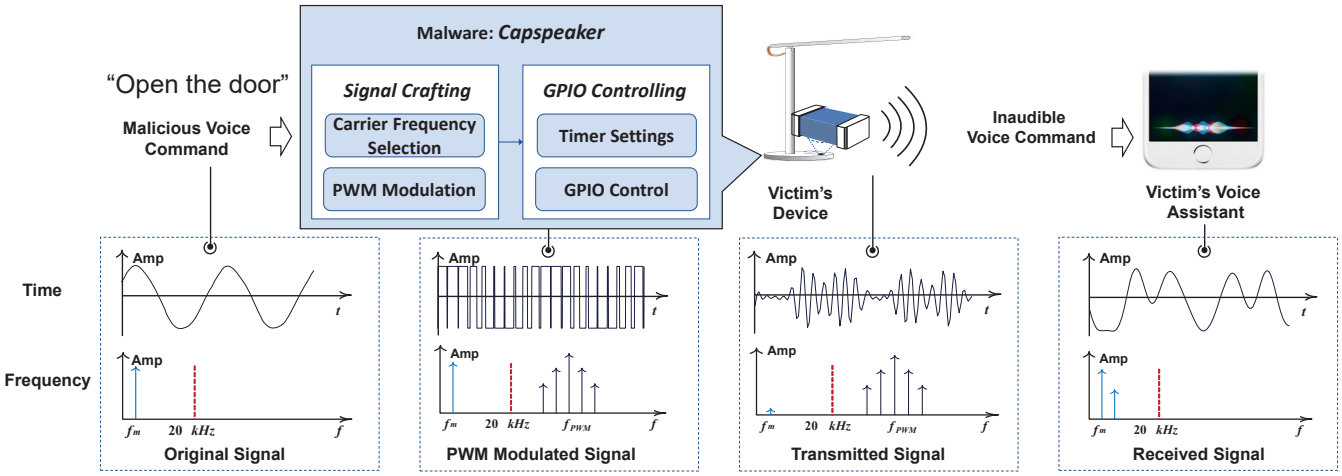


Figure 12: Workflow of CapSpeaker. To inject a malicious voice command of "open the door", CapSpeaker first selects an appropriate carrier frequency, and then PWM-modulates the baseband signal with voice commands onto the carrier. Then, CapSpeaker manipulates the GPIO pin using high-level programming instructions to generate the PWM-modulated signal, which drives the capacitors to produce the desired inaudible voice commands. Finally, the produced voice commands can be received and demodulated by the victim's voice assistant, which executes the command. We plot the signal in the time and frequency domain at various stages.

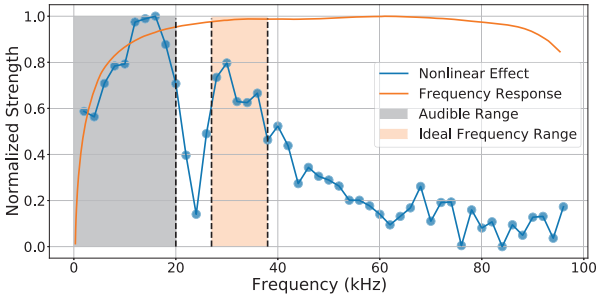


Figure 13: Carrier frequency selection. CapSpeaker has to simultaneously consider the frequency responses of capacitors, nonlinearity effect of the victim's microphone, and the audibility of the attack signal. The nonlinearity effect curve is from the measurement of the iPhone 4s smartphone.

The implementation of the PWM modulation can be referred to the Matlab Toolbox [48].

4.3 GPIO Controlling

After we derive the PWM-modulated signal, the next step is to execute the malware on the victim device to generate the attack signal. Recall that commercial devices such as the LED lamp have no accessible programming interfaces to directly control the voltages across capacitors, and we can only rely on high-level programming instructions to control the GPIO that connecting to the peripheral load. To ensure the stealthiness of the malware, the malware should not affect the normal function of the devices, e.g., a brightness control function.

Most of the microcontroller units (MCU) of IoT devices support PWM output control, and PWM is widely to realize various functions, e.g., the brightness control of monitors, the speed control of motors used in fans, the temperature control for heaters, and the volume control for loudspeakers. Thus, CapSpeaker can use PWM output control to adjust the voltages across capacitors in a similar style as controlling the brightness of LED lamps. To run the malware of CapSpeaker on the victim device, we exploit the off-the-shelf hardware PWM API on MCUs, which is usually implemented and controlled by a timer. Suppose the period of the PWM waveform is T, and the duty cycle is D. Then the PWM waveform with parameters T, D is achieved by setting the GPIO output to 1 during the active state and vice versa.

For computation-resource-constrained IoT devices, we face the challenge that the maximum carrier frequency supported by the PWM modulation is limited by the MCU clock cycles. For example, the MCU in our setup is ESP-WROOM-32D, which cannot support real-time fine-grained (i.e., 32 kHz) PWM calculation. Therefore, to strike the balance between accuracy and implementation, we increase the duty cycle to every two PWM periods to decrease the calculation overhead. Suppose that the duty cycle trace of the malicious voice command is $Duty[0]$, $Duty[1]$, $Duty[2]$, ..., we select $Duty[0]$, $Duty[2]$, $Duty[4]$, ... to set the PWM duty value. Of course, the specific PWM modulation scheme can be configured according to the calculation capability of the victim device.

How the signal is changed at different stages. After the signal is transmitted from the victim device, i.e., the LED lamp, the victim's voice assistant shall receive the attack signal and demodulate it due to the nonlinearity effect. The signal in both time and frequency domain from the original voice command signal, to the PWM-modulated one, the transmitted one from the LED lamp, and the finally received one from the voice assistant are plotted

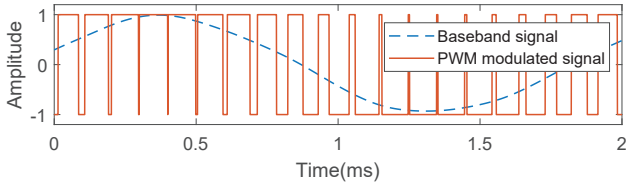


Figure 14: Pulse-width modulation (PWM). The amplitude of the baseband signal is converted to duty cycles of the PWM waveform using PWM modulation.

in Fig. 12. It is worthy to mention that even if the finally received voice command signal is not the same as the original one, i.e., the demodulated one has a $2f_m$ component, the voice assistant can still recognize it with high probability. In the evaluation section, we provide a detailed evaluation of the recognition rate of injected voice commands.

4.4 Verification of the PWM-modulation

To verify the validity of the effectiveness of the PWM-modulation scheme, we use one capacitor of 1uF. to generate the attack signal following the experimental setup in Fig. 4. The only difference is that the recorder is replaced by an iPhone 4s smartphone. We used an online text-to-speech website [1] to convert "Turn on airplane mode" into the voice command. The PWM-modulated signal is converted to executed program instructions using Alg. 1. The spectrogram of the received speech command after demodulation from the iPhone 4s microphone is plotted in Fig. 15b, which exhibits similar patterns as that of the original signal shown in Fig. 15b, indicating the effectiveness of the PWM-modulation scheme.

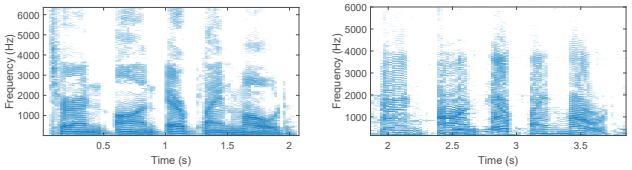
Algorithm 1: `CapSpeaker` malware.

Data: PWM duty cycle trace $Duty$, PWM Frequency f_{PWM}, PWM output pin P

```
setPwmOutputPin(P)
setTimerPeriod(1/f_PWM)
count ← 0
setDutyCycle(Duty[count])
startTimer()
Function Interrupt():
    count ← (count + 1) mod len(Duty)
    setDutyCycle(Duty[count])
return
```

Table 2: Experimental setups: The self-implemented LED lamp is for factors evaluation, and the commercial one is for feasibility validation.

Setup	Self-implemented lamp	Commercial lamp
LEDs	3 W × 11	0.5 W × 42
Driver board	700 mA PT4205 LED driver	700 mA PT4205 LED driver
MCU	ESP-WROOM-32D	ESP-WROOM-32D
Purpose	Impact factors evaluation	Feasibility validation on commercial products



(a) Original voice command signal. **(b) The demodulated attack signal.**

Figure 15: The original "Turn on airplane mode" signal (a) and the received one by the iPhone 4s from the sound emitted by a capacitor (b). The spectrograms of two signals show similar patterns, indicating the effectiveness of the PWM-modulation scheme.

5 IMPLEMENTATION AND EVALUATION

In this section, we implement and evaluate of `CapSpeaker` using the LED lamp scenario.

5.1 Experiment Setup

We utilized a self-implemented LED lamp and a commercial one to validate the performance of `CapSpeaker`. The former is used to test the performance of `CapSpeaker` at various distances, ambient noises, etc. The commercial LED lamp product is utilized to validate the feasibility of the attack against an off-the-shelf product instead of the self-implemented one, listed in Tab. 2.

The self-implemented LED lamp prototype. We implemented a prototype of `CapSpeaker` using a group of LEDs, an LED driver, and an MCU board as shown in Fig. 16. We used 11 LEDs in series, and each LED is of 3 W power. The LED driver is a commercial off-the-shelf (COTS) one with a rated current of 700 mA, and is used to drive the LEDs. The driver was powered by a 48 V DC adapter, and it has a dimming interface for adjusting the brightness of the LEDs. We used the ESP-WROOM-32D MCU, which is widely used in commercial products such as Xiaomi Lamp 1S [9], to control the driver board through the dimming interface. Due to the limited current of the I/O port of the MCU, we use a triode-based amplifier circuit to drive the LED driver board. We developed a malware that can generate a 32 kHz PWM wave and PWM-modulated 3 different voice commands. The detail of the hardware setup and the malicious voice commands are listed in Tab. 3.

The commercial smart LED lamp. To validate the effectiveness of `CapSpeaker` against commercial products, we utilized a Xiaomi Lamp 1S [9]. The LED lamp also uses the ESP-WROOM-32D MCU. There are 42 LEDs and each is of 0.5 W power. Since reverse engineering the lamp firmware requires intensive work, we replace its circuit board with an MCU of the same model (ESP-WROOM-32D) and the malware has been implanted inside the MCU in advance for convenience. The other hardware components are kept unchanged. The detailed specification of the commercial LED lamp setup can be referred to Tab. 2.

Under both setups, we placed the victim's voice assistants, e.g., the smartphones, smartwatches, and tablets close to the LED lamp at various settings, e.g., distance, ambient noises, the impact of other devices, etc., and the microphones of the voice assistants are facing the LED lamp to receive the attack signal. The detailed experimental parameters and settings of the evaluation are shown in Tab. 3.

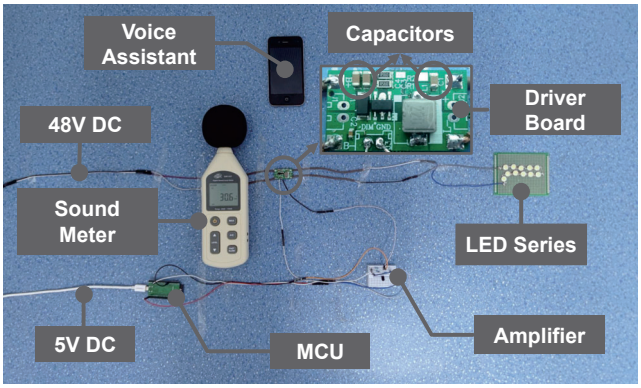


Figure 16: The implementation of the LED lamp prototype. 11 LEDs are connected in series and they are driven by a driver board which is of the same model as the one used in the commercial Xiaomi 1S smart LED lamp. (Setup #1: *Self-implemented lamp*.)

5.2 Evaluation Metrics

We define 3 metrics for the evaluation of CapSpeaker under the above setups.

- Recognition success rate $R_c = N_c/N$, where N is the number of attacks while N_c is the number of correctly recognized commands by the voice assistant.
- Responding rate $R_r = (N_c+N_w)/N$, where N_w is the number of commands that can be responded to but are incorrectly recognized as any other commands.
- The maximum attack distance D_{max}, i.e., the maximum distance that the command can be successfully recognized.

5.3 Impact of Distance and Direction

To evaluate the effectiveness of the malicious voice command generated by CapSpeaker at each location, we use the self-implemented LED prototype (in Tab. 2) and experimental settings in Tab. 3. We measure the recognition success rate and responding rate of iPhone 4s using executable speech commands including "call my wife", "open the door" or "turn on airplane mode" at various distances. We consecutively tested each command at each distance 20 times with an ambient noise of 30 dB. The remaining experimental setting parameters are by default.

Recognition rate vs. distances. Fig. 17 shows the results. The bars with complete opacity indicate the ratio of correctly recognized voice commands, while the bars with transparency indicate the ratio of incorrectly recognized voice commands (but still activated the voice assistants). The recognition success rate R_c and the responding rate R_r can be obtained from the top value of the bars with complete opacity and the bars with transparency, respectively. The results show that the recognition success rate is decreasing with the increase of distance and the maximum attack distance under the self-implemented LED lamp setup can achieve 10.5 cm and the recognition success rate for all three commands is above 80%. With larger distances, both the recognition rate and the chances to make the voice assistant responding rate gradually decrease. This is because, at large distances, the emitted voice commands from the LED lamp attenuate and shall be interfered with by the ambient noises.

Table 3: Experimental settings of the evaluation.

Settings	Default	Additional
Target device	iPhone 4s	Nova 5i Pro, iPad mini 5, K30 Ultra, iWatch S1
Voice assistant	Siri	Xiaoyi [10], iFlytek [8]
Inductors	On board	Far away
Noise (dB)	30	37, 43, 46.8, 54.5, 60.5
Distance (cm)	10.5	0-10.5
Prox. devices	None	Fan, laptop, router, monitor
Current (A)	0.41	0.35, 0.31, 0.24, 0.16
Command	Open the door	Call my wife, Turn on airplane mode

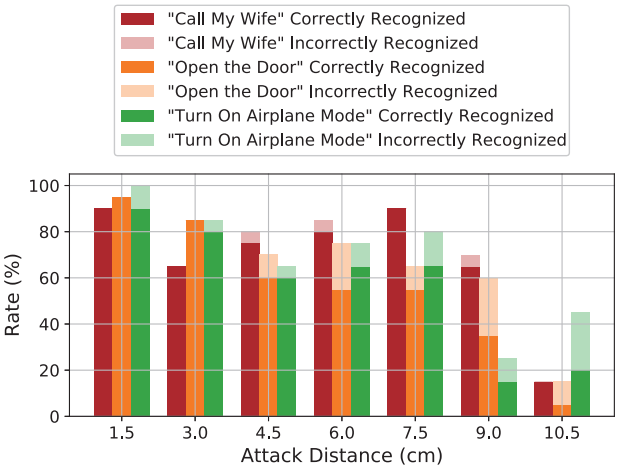


Figure 17: Recognition success rate vs. attack distances. Bars in opaque colors are the successful recognition rate, while bars of transparent colors are the rate of incorrectly recognition. The sum of them represents the responding rate.

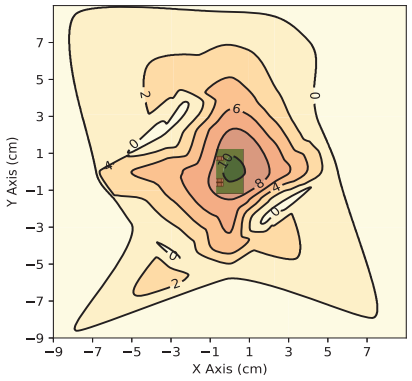


Figure 18: Recognition success rate vs. directions. The green rectangle in the middle represents the LED driver board, whose geometric center is the origin of the contour map. The number on each contour line is the maximum height (along Z axis) that the attack can succeed at the coordinates. The plot indicates that the voice commands can be emitted in all directions in the X-Y plane with various strengths.

Recognition rate vs. directions. In addition, we tested the maximum attack distance at 28 different locations from different

Table 4: Experimented devices, voice assistants of various types and brands, and the results. The carrier frequency and maximum attack distances are measured in an environment with a background noise of 30 dB SPL.

Type	Manufacturer	Model	Rel. Date	OS/Version	Voice Assistant	Recognition	$f_{pwm}(kHz)$	Max. Dist. (cm)
Smartphone	Apple	iPhone 4s	2011.10	iOS 9.3.5	Siri	Yes	32	10.5
Smartphone	Huawei	Nova 5i Pro	2019.07	EMUI 10.1.0	Xiaoyi [10]	Yes	33	9
Tablet	Apple	iPad mini 5	2019.03	iOS 13.5.1	Siri	Yes	24	3.5
Smartphone	Redmi	K30 Ultra	2020.08	MIUI 12.0.18	iFlytek [8]	Yes	28	0.5
Smartwatch	Apple	iWatch S1	2015.04	watchOS 3.1	Siri	Yes	22.3	0.2

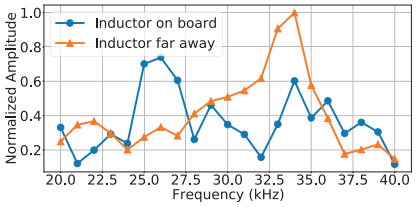


Figure 19: The normalized sound amplitude of LED driver board when the inductor is on the board and the inductor is far away.

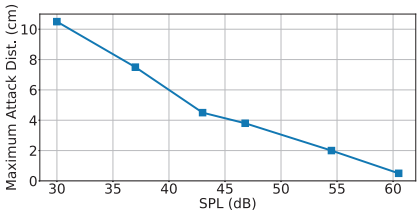


Figure 20: maximum attack distance vs. SPL of ambient noises. With 40 dB ambient noise, the attack distance is still around 6 cm.

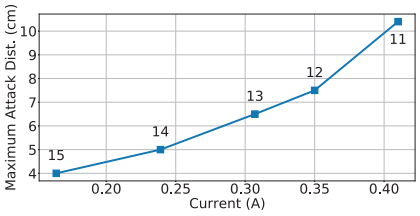


Figure 21: maximum attack distance vs. current. The attack distance can be increased by increasing the current, i.e., using fewer LEDs to get a heavier load.

directions from the LED driver board. We fit the points corresponding to the maximum attack distance of each direction to a surface, and plotted the contour map of the surface in different directions as shown in Fig. 18. The LED driver board is located at a height of 0 cm from the center of the figure. The height value at each point indicates the height of the surface at that point. The results illustrate that CapSpeaker can be successful in all directions, even the maximum attack distance varies in directions due to the uneven sound field generated, which coincides with the simulation in Fig. 5.

5.4 Impact of Voice Assistants

In addition to the iPhone 4s, we tried several other voice assistants. We tested 10 devices in total, including smartphones, smartwatches, tablets, and speakers. The experimental setup is shown in Fig. 16. The ambient noise level is 30 dB. For each device, we chose the carrier frequency with the best performance in advance and kept the microphone of the device facing towards the LED driver board. We measured the maximum attack distance of each device and recorded the voice commands used at the maximum attack distance. The results are shown in Tab 4. It shows that except iPhone 4s, Huawei Nova 5i Pro can also be successfully attacked as far as 9 cm, and iPad mini 5 can be successfully attacked as far as 3.5 cm. Note that both Huawei Nova 5i Pro and iPad Mini 5 were released in 2019 and sold 2 million and 40 million units respectively [7, 25]. In addition, iWatch and Redmi K30 Ultra can be attacked successfully at a very close range. This is because the microphones' nonlinearity of these two devices is not as significant as the other devices. We believe that more devices can be successfully attacked when the capacitors can emit stronger sounds, such as using a higher power load.

5.5 Impact of Background Noise

Controlling background noise. To demonstrate that our CapSpeaker is robust to background noise, we carried out a set of experiments under different ambient SPL settings. The main sources of the noises are an air conditioner and a server cooling fan. We controlled the sound pressure levels by putting the attack devices at various locations and used BENETECH GM1357 sound level meter to measure the SPL in the vicinity of the LED driver board. All experimental setups are the same as in Sec. 5.3 except for the SPL. Fig. 20 gives the relationship between the maximum attack distances under the corresponding SPLs of the ambient noises. The results show that the maximum attack distance decreases linearly with the increase of noise SPL. We refer to a Decibel Table [4] which presents the comparison between SPL and loudness. We found that 30 dB is equivalent to a quiet bedroom at night, 40 dB is equivalent to a quiet library, 50 dB is equivalent to an average home, and 60 dB is equivalent to a conversational speech at 1 m. Therefore, we conclude that CapSpeaker can achieve an attack distance larger than 3 cm in average in home environments. For a victim with a quiet home environment, CapSpeaker can even achieve a distance larger than 10 cm.

Interfering devices nearby. In addition to the impact of background noises, we investigated the impact of other interfering devices in proximity to the victim device in practice, as other electrical appliances have capacitors and inductors that can generate sounds. We used a cooling fan, a laptop, a router, and a monitor at 50 cm from the LED driver board respectively, and tested the maximum attack distance using the experimental setup in Sec. 5.3. All of the 4 devices were in operation and their locations with respect to the LED prototype are shown in Fig. 22 and their sound pressure levels are measured in advance shown in the caption of each figure. Tab. 5 shows the results. We can find that only the cooling fan has

a significant impact on the maximum attack distance because the noise from the fan has a frequency overlap with that of the voice command. The laptop and monitor only have a slight impact on CapSpeaker, i.e., the maximum attack is only decreased by 1 cm.

5.6 Impact of Inductors

Similar to MLC capacitors that vibrate due to voltage changes, inductors also vibrate due to current changes. Therefore, we need to confirm the sound of the LED driver board is mainly emitted by the capacitor instead of the inductor. Fig. 24 (in Appendix A) shows the LED driver board we use for this investigation. There are 3 MLC capacitors and 1 inductor on the board. We used a heat gun to remove the inductor. To make the driver board work, we kept the inductor connected using a long wire but placed it at a position 60 cm away from the LED driver board as shown in Fig. 24. The introducing resistance of the long wire is only 0.023 Ω to the circuit and can be omitted. The LED circuit is powered by a 48 V power adapter, and the load of the LED driver is the LED array used in Sec. 5.3. The voice recording device is the same as that in Sec. 2.2. We used a 20 kHz to 39 kHz swept square wave (in steps of 1 kHz) to drive the LED driver board. We recorded the sound generated by the LED driver in two cases: 1) the inductor-on-board case shown in Fig. 16; 2) the inductor-far-away case shown in Fig. 24.

The results are shown in Fig. 19. The blue and orange lines show the frequency responses for both the (1) inductor on board and the (2) inductor far away cases. Case (2) represents the frequency response of the capacitors while case (1) is the combined frequency response of both the capacitors and the inductor. The results show that after removing the inductor, the frequency response changes significantly in both amplitude and frequency. Without the inductor, the amplitude of the sound is larger, and the best frequency shifts to around 34 kHz from around 26 kHz. This indicates that the inductor can indeed produce sounds while its sound strength is weak compared to that of the capacitors. The reason why the sound becomes stronger after removing the inductor is that the PCB resonance changes after the removal of the inductor, i.e., the PCB has a lighter wight and becomes more elastic. Therefore, we conclude that the MLC capacitors are the main source of sound for the LED driver board, and the existence of inductors can interfere with the produced sound by modifying the PCB resonance frequency.

5.7 Impact of Various Loads

To verify whether the attack distance can be increased by using larger loads, i.e., larger current or higher voltage changes, we tested the maximum attack distance with 5 different loads. The experiment is based on the experimental setup in Sec. 5.3, with the input current signal of the LED driver board changed. Specifically, with a constant supply voltage, we obtained various input currents by controlling the number of LEDs in series. We used 15, 14, 13, 12, 11 LEDs in series respectively, and the generated input currents are 0.164 A, 0.239 A, 0.307 A, 0.35 A, 0.41 A, respectively. We tested the maximum attack distance for each current, and the results are shown in Fig. 21. As the current increases, the maximum attack distance increases accordingly. This gives up the opportunity to get a higher attack distance by using a higher power load.

Table 5: Other devices working in the vicinity of the LED driver board. The SPLs of the operating devices and the maximum attack distance for each case are measured.

Device	SPL (dB)	Maximum Attack Distance (cm)
Cooling Fan	49.9	4.8
Laptop with Fan On	32.1	9.5
Router	31.2	10.5
Monitor	30.9	9.5

5.8 Evaluation of Commercial Smart LED Lamp

In addition to the self-implemented LED prototype, we validated the feasibility of CapSpeaker against a commercial LED lamp. The commercial product is a Xiaomi 1S Lamp [9]. To ease the burden of reverse engineering the firmware, we directly replace its MCU board with one of the same models (ESP-WROOM-32D), and implant the malware inside the new MCU board in advance. The experiment was conducted in a 30 dB noise environment. The detailed specification of the off-the-shelf smart LED lamp can be referred to Tab. 2.

PWM modulation implementation. The only concern for the commercial LED experiment is the setting of the PWM modulation scheme. Here we used an LED PWM controller module inside the ESP-WROOM-32D MCU to generate the PWM signal. The PWM controller module is already implemented in the MCU and it is primarily designed to control the intensity of LEDs [6]. To achieve a high resolution of the PWM signal, we chose APB_CLK as the internal clock source as it has a higher clock frequency (80MHz) than that of REF_TICK (1MHz).

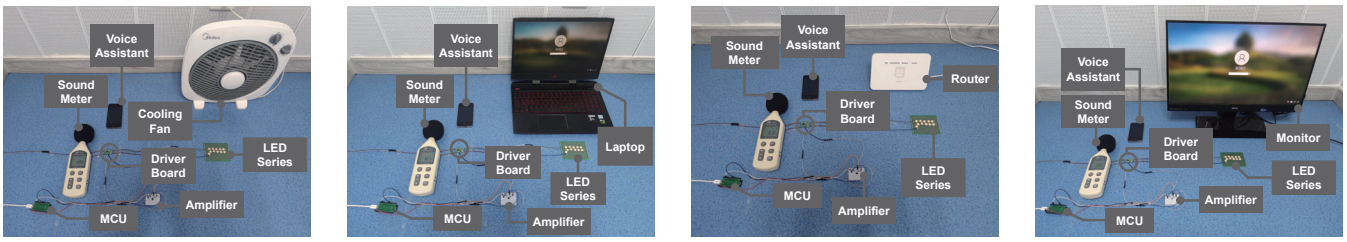
Results. As Fig. 23 shows, the commercial LED lamp can successfully respond and make the victim's voice assistant, i.e., iPhone 4s Siri recognizes the "open the door" voice command. The maximum attack distance can be as far as 3.2 cm, i.e., the distance from the lamp's outer case to the smartphone. In addition, the commercial lamp can attack other voice assistants. For Huawei Nova 5i Pro, the maximum distance is 2.9 cm, and for iPad mini 5 it is 0.5 cm respectively.

6 DISCUSSION

In this section, we discuss the defense strategies against CapSpeaker attacks and introduce components that can be used to produce sounds, e.g., inductors and rheostats.

6.1 Countermeasures

At the hardware level, a naive method to defend CapSpeaker is to avoid using MLC capacitors, but use electrolytic capacitors or tantalum capacitors instead. However, electrolytic capacitors are too large. With the miniaturization trend of electronic devices, devices do not always have enough space to use electrolytic capacitors. Although tantalum capacitors have small sizes, they are expensive and have risks of failure [39]. Therefore, MLC capacitors are irreplaceable. To reduce the sound emitted by MLC capacitors, one feasible way is to mount the MLC capacitor on an interposer to decrease the amount of vibration that will be transmitted to the PCB substrate [54]. The other way is to fabricate the MLC capacitor



(a) Fan (49.9 dB) (b) Laptop (32.1 dB) (c) Router (31.2 dB) (d) Monitor (30.9 dB)

Figure 22: Impact of other electrical appliances. We put 4 running devices, i.e., a cooling fan, a laptop, a router and a monitor 20 cm away from the LED driver board to test their interference with the voice commands.

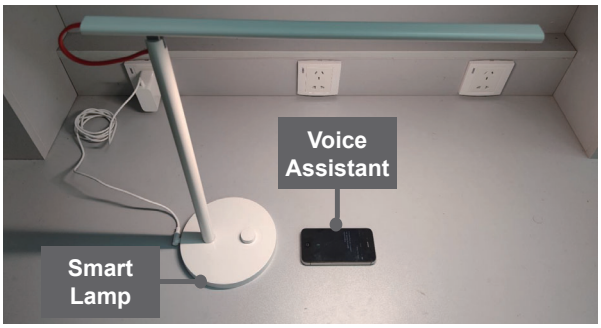


Figure 23: Evaluation on a commercial LED lamp. The Xiaomi 1S Lamp is used to inject an "open the door" voice command. Voice assistants including Siri on iPhone 4s can be successfully attacked with a maximum distance of 3.2 cm. (Setup #2: *Commercial LED lamp.*)

by adding two metal "legs" made of ferric alloy with Ni/Sn to isolate the capacitor vibration from the board [54].

At the software level, detection mechanisms can identify the possibility of malicious content in benign PWM control programs. For example, in brightness control of LEDs and temperature control of heaters, the duty cycle of the PWM wave does not change drastically in normal scenarios, while `CapSpeaker` continuously changes the duty cycle to be up and down rapidly. Thus, `CapSpeaker` can be detected by monitoring the frequency and amplitude of duty cycle changes.

6.2 Alternative Potential 'Speakers'

In addition to capacitors, inductors and rheostats are widely used in commercial electronic devices and can produce sounds, too.

Inductors. An inductor typically consists of an insulated wire wound onto a coil [58], and it is a passive two-terminal electrical component that stores energy in a magnetic field when electric current flows through it [58]. The structure of an inductor is similar to a speaker, and thus the coil can vibrate according to the input currents. To investigate the feasibility of using an inductor to attack the voice assistant, we produced the attack voice command using a single inductor and a DC-DC converter with the same type of inductor, following the experimental setup shown in Fig. 25 (in Appendix A). The frequency of the PWM wave is set to 50 kHz. Both settings can successfully cause the iPhone 4s to recognize the attack command "Turn on airplane mode". However, in reality, most of the current electronic devices use inductors of metallic integral molding types, whose produced acoustic signal is mitigated to 1/10 of the original sounds [56].

Rheostats. Similar to inductors, a rheostat is made of a coil and can generate sounds with varying currents, due to the repelling force from the magnetic field. To validate, we carried out an individual experiment on the rheostat, with no MLC capacitors or inductors in the circuit. We sweep the frequency to obtain the frequency response of sound produced by the rheostat, and the frequency response is approximately 5 dB, lower than that of capacitors. Note that in our experiments of MLC capacitors, we placed the rheostat outside of the room to avoid the interference from rheostats.

6.3 Limitations

Admitted that the attack distance of `CapSpeaker` is 10.5 cm and requires the targeted devices to be nearby, the attack distance can be increase with a higher voltage signal, possibly in electronic devices with a larger power, e.g., LED lamps with a higher power, rice cookers. Nevertheless, the goal of the work is to raise awareness of such an attack, since people tend to put their smartphones on the desk and in the vicinity of other devices such as a lamp.

7 RELATED WORK

7.1 Implicit Voice Commands

Using implicit voice commands to attack speech recognition systems is well known and studied. Zhang et al. [63] achieve inaudible voice commands by modulating low-frequency voice commands on ultrasonic carriers. The modulated low-frequency voice commands can be demodulated by the nonlinearity of microphones. Such an attack can attack most of the off-the-shelf smartphones, smart speakers, and laptops. Roy et al. [50] extend the attack range and develop a defense against this class of voice attacks that exploit nonlinearity.

In addition to inaudible voice commands, commands that are indistinguishable can also be implicit, although they are audible. The attacker can generate implicit voice commands by removing the audio features that are not used in the speech recognition system but a human listener might use them for comprehension [19]. In detail, they first calculate the Mel-frequency cepstral coefficients

(MFCC) of the original speech. Then, they use the MFCCs to reconstruct the original speech. The reconstructed speech loses the phase information of the speech, and thus cannot be comprehended. Yuan et al. [62] find that the voice commands can be stealthily embedded into songs. When the song is played, the voice command can be recognized by the speech recognition system while not being noticed by humans. Schönherr et al. [51] introduce a new type of adversarial examples based on psychoacoustic hiding. Du et al. [24] propose a method to generate adversarial audios which can attack both black box and white box. Abdullah et al. [15] make hidden command attacks more practical through black-box attacks. Specifically, they perturb the original audio by time domain inversion, random phase generation, high frequency addition, and time scaling.

7.2 Privacy Leakage by Devices

The privacy leakage through devices can be divided into side-channel attacks and covert channel ones.

Side-channel attacks exploit unintended information leakage of computing devices or implementations to infer sensitive information. When a device is computing, it consumes power which dissipates in various forms of physical signals including sound, light, electromagnetism, force, and heat. Those signals contain information related to the computation process and thus can be used to extract sensitive data. Existing physical side-channel attacks can be categorized as acoustic ones [26], electromagnetic ones [16, 17, 22, 64, 67], motion ones [60], optical ones [20] and thermal ones [42]. In addition, there are side-channel attacks in the digital domain, including cache-based side-channel attacks [47], timing-based side-channels [53] and encrypted-traffic-based side-channels [21, 66]. Multiple attacks can be realized through a side-channel. For example, Genkin et al. [28] use the ground electric potential of computers to infer RSA keys. They also achieve screen contents inference via the acoustic noise generated by the screen circuits [26].

Covert channel is defined as the channel that is not intended for information transfer at all but leaks sensitive data [45]. As the community becomes more and more aware of network security, air-gap networks are widely used in security-aware organizations such as power grids and military bases. However, covert channel can break the air gap because devices inevitably generate physical signals during operations. If an attacker can control the operation of the devices, she can leak sensitive information from the air-gapped networks. Covert channels can be classified according to the type of physical signals they use, such as voltage signals on power lines [35, 41, 52], electromagnetic signals emanated from memory reading and writing [30], USB cable [31], magnetic signals produced by CPU [37, 65], acoustic signals generated by hard-drive [33] and ultrasonic communications [34]. In addition, optical signals [36] and thermal signals [32] can be used to leak sensitive data.

7.3 Utilization of Non-Speaker Devices and Non-Microphone Devices

There have been some works that focus on utilizing the sound generated by non-speaker devices. Guri et al. [29] use the sound generated by power supplies to create a covert channel. Specifically, they suppose that malware can be implanted in a computer in advance. Then, the malware controls the CPU cores to be busy and idle, causing the current drained from the power supply to change periodically. Then, the power supply can generate a sound of the corresponding frequency. Thus, the malware embeds information into sound, and the sound can be received by a general microphone. Genkin et al. [27] show that acoustic noise of the screen circuits can be used to infer screen content. The principle is that the momentary power draw induced by the monitor's digital circuits varies as a function of the screen content being processed in raster order. This process affects the electrical load on the power supply components. As the power supply components have capacitors and inductors, the ripple current and voltage will cause vibrations and thus generates sounds. Yang et al. [61] use sound from power supply to fingerprint appliance. As switching-mode power supply (SMPS) is widely used in off-the-shelf electronic devices, the differences of SMPS working modes, frequencies, topologies can be used to fingerprint an appliance. Compared with traditional device fingerprint methods such as electromagnetic-based methods, their method is low-cost and easy to deploy.

In addition, non-microphone devices can be used to sense sound. Trippel et al. [57] found that although a MEMS accelerometer is not a microphone, it can still be interfered with by sound waves and even controlled by sound waves. Likewise, the shock sensor in the hard drive can also be interfered with by sound waves [18], and the This interference could cause major data centers to go down. In addition, the head of the hard drive can be used as a microphone to record sound by recording the Position Error Signal (PES) of the head [44].

8 CONCLUSION

In this paper, we demonstrated that acoustic noises produced by MLC capacitors may appear to be benign, yet they can be manipulated to inject malicious voice commands to nearby voice assistant, which we call `CapSpeaker` attacks. After studying the factors that influence the strength of the sounds produced by the MLC capacitors, we found that MLC capacitors can produce sounds in the inaudible range much better than audible ranges. Thus, `CapSpeaker` modulates the malicious voice commands to an inaudible frequency carrier to take advantage of the high frequency response, and exploits the nonlinearity property of the microphone to automatically demodulate the PWM-modulated signals embedded with malicious voice commands. We implement the `CapSpeaker` attack on a smart LED lamp, and evaluate the performance of `CapSpeaker` attacks with various voice commands, receiving devices, ambient noise levels, and distances between the capacitor to the receiver. The results show that malicious voice commands can be successfully recognized by an iPhone 4s at a distance of 10.5 cm. The work aims to raise awareness of such vulnerabilities in light of the growing trends of voice assistants.

9 ACKNOWLEDGMENTS

We thank all anonymous reviewers for their insightful comments on this paper. We thank You Wu for conducting experiments on MLC capacitors with various impact factors. We thank Shuai Chen for conducting simulations and experiments on verification of the PWM-modulation. This work is supported by China NSFC Grant 62071428, 61925109, and 61941120.

REFERENCES

[1] 2020. From Text To Speech - Free Online TTS Service. http://www.fromtexttospeech.com/.

[2] 2021. 192/24 PCI-E 8-Channel sound card. Retrieved 20-August-2021 from http://www.syba.cc/e/wap/show.php?classid=24&id=418&style=0&cpage=4&cid=3&bclassid=1

[3] 2021. COMSOL - Software for Multiphysics Simulation. Retrieved 20-August-2021 from https://www.comsol.com/

[4] 2021. Decibel Table - SPL - Loudness Comparison Chart. Retrieved 20-August-2021 from http://www.sengpielaudio.com/TableOfSoundPressureLevels.htm

[5] 2021. DG800 series RIGOL waveform generators. https://www.rigolna.com/products/waveform-generators/dg800/dg811/.

[6] 2021. ESP32 Technical Reference Manual. Retrieved 20-August-2021 from https://www.espressif.com/sites/default/files/documentation/esp32_technical_reference_manual_en.pdf#ledpwm

[7] 2021. Huawei Nova 5 Series Sales Cross 2 Million Units in a Month. Retrieved 20-August-2021 from https://gadgets.ndtv.com/mobiles/news/huawei-nova-5-pro-5i-pro-2-million-sales-china-2077454

[8] 2021. iFLYTEK - Empower The World With A.I. Retrieved 20-August-2021 from https://www.iflytek.com/en/

[9] 2021. mi-led-desk-lamp-1s. Retrieved 20-August-2021 from https://www.mi.com/global/mi-led-desk-lamp-1s/overview/

[10] 2021. [NEWS] HUAWEI UNVEILS ITS OWN VOICE ASSISTANT. Retrieved 20-August-2021 from https://consumer.huawei.com/ae-en/community/details/NEWS-HUAWEI-UNVEILS-ITS-OWN-VOICE-ASSISTANT-CELIA/topicId_82910/

[11] 2021. Siri. https://www.apple.com/siri/.

[12] 2021. Surface mount component packages. Retrieved 20-August-2021 from https://www.surfacemountprocess.com/smd-component-packages.html

[13] 2021. UltraSoundGate. Retrieved 07-May-2021 from http://www.avisoft.com/ultrasoundgate/

[14] 2021. Xiaomi XiaoAI Art Speaker. https://xiaomi-mi.com/portable-speakers/xiaomi-xiaoai-art-speaker/.

[15] Hadi Abdullah, Washington Garcia, Christian Peeters, Patrick Traynor, Kevin R. B. Butler, and Joseph Wilson. 2019. Practical Hidden Voice Attacks against Speech and Speaker Recognition Systems. In *Proceedings of 26th Annual Network and Distributed System Security Symposium, NDSS 2019*. The Internet Society.

[16] Lejla Batina, Shivam Bhasin, Dirmanto Jap, and Stjepan Picek. 2019. CSINN: Reverse Engineering of Neural Network Architectures Through Electromagnetic Side Channel. In *Proceedings of the 28th USENIX Security Symposium (USENIX Security 19)*. 515–532.

[17] Alexandru Boitan, Simona Halunga, Valerică Bîndar, and Octavian Fratu. 2020. Compromising Electromagnetic Emanations of USB Mass Storage Devices. *Wireless Personal Communications* (April 2020).

[18] Connor Bolton, Sara Rampazzi, Chaohao Li, Andrew Kwong, Wenyuan Xu, and Kevin Fu. 2018. Blue note: How intentional acoustic interference damages availability and integrity in hard disk drives and operating systems. In *Proceedings of 2018 IEEE Symposium on Security and Privacy (SP)*. IEEE, 1048–1062.

[19] Nicholas Carlini, Pratyush Mishra, Tavish Vaidya, Yuankai Zhang, Micah Sherr, Clay Shields, David Wagner, and Wenchao Zhou. 2016. Hidden voice commands. In *Proceedings of 25th USENIX Security Symposium (USENIX Security 16)*. 513–530.

[20] S. Chakraborty, W. Ouyang, and M. Srivastava. 2017. LightSpy: Optical Eavesdropping on Displays Using Light Sensors on Mobile Devices. In *Proceedings of the 2017 IEEE International Conference on Big Data (Big Data)*. 2980–2989.

[21] Yushi Cheng, Xiaoyu Ji, Tianyang Lu, and Wenyuan Xu. 2018. DeWiCam: Detecting Hidden Wireless Cameras via Smartphones. In *Proceedings of the 2018 on Asia Conference on Computer and Communications Security (ASIACCS '18)*. ACM, 1–13.

[22] Yushi Cheng, Xiaoyu Ji, Wenyuan Xu, Hao Pan, Zhuangdi Zhu, Chuang-Wen You, Yi-Chao Chen, and Lili Qiu. 2019. MagAttack: Guessing Application Launching and Operation via Smartphone. In *Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security - Asia CCS '19*. ACM, 283–294.

[23] S. Knudtsen G. Shirn D. Burks, R. Hofmaier. 1989. A ceramic capacitor for AC applications. (1989), 194–201.

[24] Tianyu Du, Shouling Ji, Jinfeng Li, Qinchen Gu, Ting Wang, and Raheem Beyah. 2020. Sirenattack: Generating adversarial audio for end-to-end acoustic systems. In *Proceedings of the 15th ACM Asia Conference on Computer and Communications Security*. 357–369.

[25] Facebook, Twitter, and LinkedIn. 2021. Did You Know This Many iPads Had Been Sold? https://www.lifewire.com/how-many-ipads-sold-1994296 Section: Lifewire.

[26] Daniel Genkin, Mihir Pattani, Roei Schuster, and Eran Tromer. 2019. Synesthesia: Detecting Screen Content via Remote Acoustic Side Channels. In *Proceedings of the 2019 IEEE Symposium on Security and Privacy (SP)*. IEEE, 853–869.

[27] Daniel Genkin, Mihir Pattani, Roei Schuster, and Eran Tromer. 2019. Synesthesia: Detecting screen content via remote acoustic side channels. In *Proceedings of 2019 IEEE Symposium on Security and Privacy (SP)*. IEEE, 853–869.

[28] Daniel Genkin, Itamar Pipman, and Eran Tromer. 2015. Get Your Hands off My Laptop: Physical Side-Channel Key-Extraction Attacks on PCs. *Journal of Cryptographic Engineering* 5, 2 (2015), 95–112.

[29] Mordechai Guri. 2020. POWER-SUPPLaY: Leaking Data from Air-Gapped Systems by Turning the Power-Supplies Into Speakers. *arXiv:2005.00395 [cs]* (May 2020). arXiv:2005.00395

[30] Mordechai Guri, Assaf Kachlon, Ofer Hasson, Gabi Kedma, Yisroel Mirsky, and Yuval Elovici. 2015. GSMem: Data Exfiltration from Air-Gapped Computers over GSM Frequencies. In *Proceedings of 24th USENIX Security Symposium (USENIX Security 15)*.

[31] Mordechai Guri, Matan Monitz, and Yuval Elovici. 2016. USBee: Air-Gap Covert-Channel via Electromagnetic Emission from USB. In *Proceedings of the 14th Annual Conference on Privacy, Security and Trust (PST)*. 264–268.

[32] Mordechai Guri, Matan Monitz, Yisroel Mirski, and Yuval Elovici. 2015. BitWhisper: Covert Signaling Channel between Air-Gapped Computers Using Thermal Manipulations. In *Proceedings of the 28th IEEE Computer Security Foundations Symposium*. 276–289.

[33] Mordechai Guri, Yosef Solewicz, Andrey Daidakulov, and Yuval Elovici. 2017. Acoustic Data Exfiltration from Speakerless Air-Gapped Computers via Covert Hard-Drive Noise ('DiskFiltration'). In *Proceedings of European Symposium on Research in Computer Security*. Springer, 98–115.

[34] Mordechai Guri, Yosef Solewicz, and Yuval Elovici. 2018. MOSQUITO: Covert Ultrasonic Transmissions Between Two Air-Gapped Computers Using Speaker-to-Speaker Communication. In *Proceedings of 2018 IEEE Conference on Dependable and Secure Computing (DSC)*. 1–8.

[35] Mordechai Guri, Boris Zadov, Dima Bykhovsky, and Yuval Elovici. 2019. PowerHammer: Exfiltrating Data from Air-Gapped Computers through Power Lines. *IEEE Transactions on Information Forensics and Security* (2019), 1–1.

[36] Mordechai Guri, Boris Zadov, and Yuval Elovici. 2017. LED-It-GO: Leaking (A Lot of) Data from Air-Gapped Computers via the (Small) Hard Drive LED. In *Proceedings of Detection of Intrusions and Malware, and Vulnerability Assessment*. Springer, 161–184.

[37] Mordechai Guri, Boris Zadov, and Yuval Elovici. 2020. ODINI: Escaping Sensitive Data From Faraday-Caged, Air-Gapped Computers via Magnetic Fields. *IEEE Transactions on Information Forensics and Security* 15 (2020), 1190–1203.

[38] J. Ho, T. R. Jow, and S. Boggs. 2010. Historical Introduction to Capacitor Technology. 26, 1 (Jan. 2010), 20–25.

[39] Jiaoying Huang, Yongkang Wan, Cheng Gao, and Yuanyuan Xiong. 2015. Discussion on multilayer ceramic replacements for tantalum capacitors. In *Proceedings of 2015 Prognostics and System Health Management Conference (PHM)*. IEEE, 1–5.

[40] Larry E Humes. 1996. Speech understanding in the elderly. *Journal-American Academy of Audiology* 7 (1996), 161–167.

[41] Mohammad A. Islam and Shaolei Ren. 2018. Ohm's Law in Data Centers: A Voltage Side Channel for Timing Power Attacks. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 146–162.

[42] Mohammad A. Islam, Shaolei Ren, and Adam Wierman. 2017. Exploiting a Thermal Side Channel for Power Attacks in Multi-Tenant Data Centers. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 1079–1094.

[43] Dongjoon Kim, Byung-Han Ko, Sanggeuk Jeong, No-Cheol Park, and Young-Pil Park. 2015. Vibration reduction of MLCC considering piezoelectric and electrostriction effect. In *Proceedings of 2015 Joint IEEE International Symposium on the Applications of Ferroelectric (ISAF), International Symposium on Integrated Functionalities (ISIF), and Piezoelectric Force Microscopy Workshop (PFM)*. IEEE, 186–189.

[44] Andrew Kwong, Wenyuan Xu, and Kevin Fu. 2019. Hard Drive of Hearing: Disks that Eavesdrop with a Synthesized Microphone. In *Proceedings of 2019 IEEE Symposium on Security and Privacy (SP)*. IEEE, 905–919.

[45] Butler W Lampson. 1973. A note on the confinement problem. *Commun. ACM* 16, 10 (1973), 613–615.

[46] Stratistics Market Research Consulting Pvt Ltd. 2020. Multi-Layer Ceramic Capacitor - Global Market Outlook (2019-2027). (2020).

[47] Yangdi Lyu and Prabhat Mishra. 2018. A Survey of Side-Channel Attacks on Caches and Countermeasures. *Journal of Hardware and Systems Security* 2, 1 (March 2018), 33–50.

[48] MathWorks. 2021. Pulse Width Modulation - MATLAB & Simulink - MathWorks. https://www.mathworks.com/help/physmod/sps/pulse-width-modulation.html

[49] Nirupam Roy, Haitham Hassanieh, and Romit Roy Choudhury. 2017. Backdoor: Making microphones hear inaudible sounds. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*. 2–14.

[50] Nirupam Roy, Sheng Shen, Haitham Hassanieh, and Romit Roy Choudhury. 2018. Inaudible voice commands: The long-range attack and defense. In *Proceedings of 15th USENIX Symposium on Networked Systems Design and Implementation (NSDI 18)*. 547–560.

[51] Lea Schonherr, Katharina Kohls, Steffen Zeiler, Thorsten Holz, and Dorothea Kolossa. 2019. Adversarial Attacks Against Automatic Speech Recognition Systems via Psychoacoustic Hiding. In *Proceedings 2019 Network and Distributed System Security Symposium*. Internet Society, San Diego, CA.

[52] Zhihui Shao, Mohammad A. Islam, and Shaolei Ren. 2020. Your Noise, My Signal: Exploiting Switching Noise for Stealthy Data Exfiltration from Desktop

Computers. In *Proceedings of the ACM on Measurement and Analysis of Computing Systems*. ACM, 1–39.
[53] Laurent Simon, Wenduan Xu, and Ross Anderson. 2016. Don't Interrupt Me While I Type: Inferring Text Entered Through Gesture Typing on Android Keyboards. In *Proceedings on Privacy Enhancing Technologies*. 136–154.
[54] Yin Sun, Jianmin Zhang, Zhiping Yang, Chulsoon Hwang, and Songping Wu. 2019. Measurement Investigation on Acoustic Noise Caused by "Singing" Capacitors on Mobile Devices. In *in Proceedings of the 2019 IEEE International Symposium on Electromagnetic Compatibility, Signal & Power Integrity (EMC+ SIPI)*. IEEE, 505–510.
[55] James A. Svoboda and Richard C. Dorf. 2013. *Introduction to Electric Circuits*. John Wiley & Sons.
[56] TDK. 2021. Measures Against Acoustic Noise in Power Inductors. https://product.tdk.com/en/techlibrary/solutionguide/acoustic-noise.html
[57] Timothy Trippel, Ofir Weisse, Wenyuan Xu, Peter Honeyman, and Kevin Fu. 2017. WALNUT: Waging doubt on the integrity of MEMS accelerometers with acoustic injection attacks. In *Proceedings of 2017 IEEE European symposium on security and privacy (EuroS&P)*. IEEE, 3–18.
[58] Wikipedia contributors. 2021. Inductor. https://en.wikipedia.org/wiki/Inductor.
[59] Wikipedia contributors. 2021. Young's modulus — Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Young%27s_modulus&oldid=1039311706
[60] Zhi Xu, Kun Bai, and Sencun Zhu. 2012. TapLogger: Inferring User Inputs on Smartphone Touchscreens Using on-Board Motion Sensors. In *Proceedings of the Fifth ACM Conference on Security and Privacy in Wireless and Mobile Networks*. ACM, 113–124.
[61] Lanqing Yang, Honglu Li, Zhaoxi Chen, Xiaoyu Ji, Yi-Chao Chen, Guangtao Xue, and Chuang-Wen You. 2020. Appliance fingerprinting using sound from power supply. (2020), 160–163.
[62] Xuejing Yuan, Yuxuan Chen, Yue Zhao, Yunhui Long, Xiaokang Liu, Kai Chen, Shengzhi Zhang, Heqing Huang, XiaoFeng Wang, and Carl A Gunter. 2018. Commandersong: A systematic approach for practical adversarial voice recognition. In *27th USENIX Security Symposium (USENIX Security 18)*. 49–64.
[63] Guoming Zhang, Chen Yan, Xiaoyu Ji, Tianchen Zhang, Taimin Zhang, and Wenyuan Xu. 2017. Dolphinattack: Inaudible voice commands. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. 103–117.
[64] Juchuan Zhang, Xiaoyu Ji, Yuehan Chi, Yi-chao Chen, Bin Wang, and Wenyuan Xu. 2021. OutletSpy: Cross-Outlet Application Inference via Power Factor Correction Signal. In *Proceedings of the 14th ACM Conference on Security and Privacy in Wireless and Mobile Networks (WiSec '21)*. 181–191.
[65] Juchuan Zhang, Xiaoyu Ji, Wenyuan Xu, Yi-Chao Chen, Yuting Tang, and Gang Qu. 2020. MagView: A Distributed Magnetic Covert Channel via Video Encoding and Decoding. In *Proceedings of IEEE International Conference on Computer Communications*. IEEE.
[66] Xuan Zhao, Md Zakirul Alam Bhuiyan, Lianyong Qi, Hongli Nie, Wajid Rafique, and Wanchun Dou. 2018. TrCMP: An App Usage Inference Method for Mobile Service Enhancement. In *Proceedings of Security, Privacy, and Anonymity in Computation, Communication, and Storage*. Springer, 229–239.
[67] Zhou Zhuang, Xiaoyu Ji, Taimin Zhang, Juchuan Zhang, Wenyuan Xu, Zhenhua Li, and Yunhao Liu. 2018. FBSleuth: Fake Base Station Forensics via Radio Frequency Fingerprinting. In *Proceedings of the 2018 on Asia Conference on Computer and Communications Security (ASIACCS '18)*. ACM, 261–272.

A EXPERIMENT SETUP OF INDUCTORS

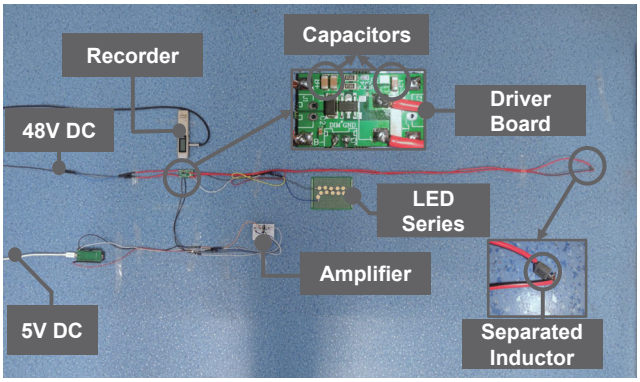
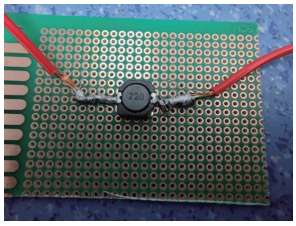
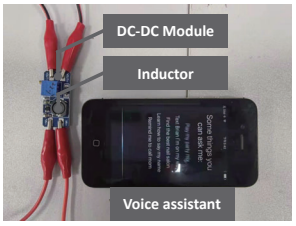


Figure 24: Experimental setup of inductor factor validation. The inductor is taken down from the board and put far away yet connected to the driver board with long wires.



(a) Inductor.



(b) Inductor on a DC-DC module.

Figure 25: Generating the malicious voice command using an inductor in a DC/DC module.