Inaudible Attack on Smart Speakers With Intentional Electromagnetic Interference

Zhifei Xu, *Member, IEEE*, Runbing Hua, *Graduate Student Member, IEEE*, Jack Juang, Shengxuan Xia, Jun Fan, *Fellow, IEEE*, and Chulsoon Hwang, *Senior Member, IEEE*

***Abstract*— This article demonstrates an inaudible attack on smart speakers using electromagnetic interference (EMI). The EMI induces voltages on the order of a few millivolts on conductors, which are then converted into baseband signals by exploiting the inherent nonlinearity of microphones. The EMI signal is specially preprocessed to minimize the useless harmonics generation at the microphone output signals, which significantly improves the recognition rate as well as nullify the previous countermeasures based on the harmonics detection. The sensitive carrier frequency found by our proposed method can improve the attack distance as well. A measurement-based methodology is applied to locate the sensitive regions for noise coupling without knowing the layout of the printed circuit board (PCB), and the transfer function is also obtained to insure the main coupling location. Our experiments show that in open space, intentional EMI under 2.5 W can inject commands at distances up to 2.5 m on smart speakers.**

***Index Terms*— Hardware security, inaudible attack, intentional electromagnetic interference (IEMI), smart speaker.**

I. Introduction

A SMART speaker nowadays is not a just music player. With more and more devices connected, smart speakers such as Google Home and Amazon Echo can serve as a "home assistant" that can provide control of common household tasks, such as environmental control (thermostat), lighting, door locks, security monitoring, and more.

The security of these smart speakers can have substantial effects, such as compromised home security and information leakage. Since smart speakers are having Wi-Fi or Bluetooth connections, various attacks can be performed through the apps and networks [1]–[3]. Researchers have implemented two application-level attacks, voice squatting and voice masquerading, which impersonate the smart speakers to steal and eavesdrop the conversations [4], [5]. A security researcher from MWR Info Security has demonstrated an attack on Amazon Echo speakers by placing the malware which enables the adversary to have access to control the smart speaker [6], [7]. However, these attacks cannot be executed remotely.

Manuscript received October 8, 2020; revised December 22, 2020; accepted January 6, 2021. Date of publication March 1, 2021; date of current version May 5, 2021. This work was supported by the National Science Foundation under Grant IIP-1916535. *(Corresponding author: Zhifei Xu.)*

The authors are with the Electromagnetic Compatibility Laboratory, Missouri University of Science and Technology, Rolla, MO 65401 USA (e-mail: zxfdc@mst.edu; rhzn2@mst.edu; jjryb@mst.edu; sx7c3@mst.edu; jfan@mst.edu; hwangc@mst.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TMTT.2021.3058585.

Digital Object Identifier 10.1109/TMTT.2021.3058585

On the other hand, as the system trusts the microphone readings, a physical layer attack can readily bypass conventional security algorithm providing an unchecked entry point to the system. The application layer with software running on the smart speaker systems makes critical decisions of the input data acquired by the microphone circuit. Recently, several publications have demonstrated inaudible voice commands injection on the physical layer of the smart speakers by exploiting the nonlinearity of the microphone [8]–[17].

The dolphin attack or ultrasound attack [9]–[15] has demonstrated that a voice-enabled device can respond to inaudible ultrasound voice commands. More recently, laser pointers have been demonstrated as another tool for attacking microphone-based devices [8]. Some defense methods regarding the ultrasound attack have been investigated recently [9], [15] such as the voice signal processing method proposed in [15]. However, these attack types are limited by obstacles such as windows. The ultrasound is mechanical waves which need strong power to propagate through the window [27], and the laser pointer attack requires the device insight because the microphone of the smart speaker needs to be pointed while attacking.

In contrast to other types of attacks such as ultrasound and light command, EMI-based attack can penetrate windows with relatively low loss and does not need to have the target in sight. The high-power intentional EMI can stop electrical network such as electric cars, trains, and transformers [20]–[27], and radio communicating devices such as cellphone, computer, and other electronics will be impacted as well [20], [23]; the required high power can be deduced by [22] for a long-distance attack. The IEMI can also be applied to inject information into the analog devices which operate in the order of a few millivolts [16]–[19]. This attack with circuit easily affected is known as "back-door" interfering [17]. Since the acquisition process requires much lower energy, microphone circuit with cables or copper PCB interconnects is vulnerable to interference [21] and allows the information injection [17]. The intentional EMI has been employed to attack the headset cable of smartphones [16], and the audio signal has been injected through the electromagnetic coupling on the cable of the headset because the cable can act as an antenna which can receive the electromagnetic interference (EMI). The intentional EMI has also been employed to attack the analog sensor of the microphone of cardiac electronic devices [17]. However, in their application, the attack setup needs to be placed very close to the cardiac electronic device.

This article is the first article demonstrating the IEMI attack on the smart speakers and attempted to increase the IEMI attack distance for the smart speakers and cellphones. Different from previous cellphone attack work, we targeted at the microphones of the devices not the headset cable. The attack principles including the electromagnetic (EM) coupling and microphone nonlinearity are presented in Section II. Then, the attack signal is optimized by exploring the nonlinearity performance of the digital microphone of the smart speakers to achieve longer distance in Section III. In Section IV, since the coupling efficiency to the smart speaker will differ depending on the frequencies of the attack signal, the sensitive frequency of the attack EM wave is explored based on measurement. A measurement-based methodology is applied to investigate the sensitive location without knowing the exact layout of the board. The maximum real voice attack distance is discussed, and the required electrical field intensity to attack different devices is presented. Finally, conclusions and countermeasures are proposed in Section V.

II. Attack Overview

A. Threat Model

We assumed that the adversary could obtain a smart speaker having the same model as the targeting smart speaker being widespread in the market. The investigation and attack experiments can be performed for the obtained smart speaker within a private home or a laboratory. In addition, the same smart speaker model can be attacked using the same setup developed in the laboratory. As the system of the smart speaker trusts the microphone readings from the physical circuit, the physical attack can bypass conventional security algorithm. The adversary can manipulate the microphone's reading by injecting voice command signals to the analog circuit of the microphone in the smart speaker; then, the injected voice command can pass the application layer algorithm of the system and be recognized by the smart speaker. Finally, the intended voice command from the adversary will be executed by the smart speaker. Meanwhile, the adversary can make the voice command signal inaudible to human but audible to the smart speaker. Generally, human ear can receive audio signals with frequency between 20 Hz and 20 kHz and, therefore, the microphones of the smart speakers are also designed to receive the audio signals in this frequency range. Thus, no one can realize that the target device is being attacked. This is a critical security issue for the smart speakers placed at homes.

Although some smart speakers can be set to recognize only the owner's voice, the owner's voice record or speeches can be found on the internet or somewhere; then, the recompose of the voice command word by word can be done through the software programming [37]–[39]. In addition, the owner's voice can also be constructed through deep learning [40], [41].

B. IEMI Coupling Mechanism

The working smart speaker with an electrical circuit can be taken as a receiving antenna; EM waves can be coupled to conductors on the printed circuit board (PCB) [32], [33], as shown by the dashed line in Fig. 1. When the signal is coupled onto the power/ground net and reached the amplifier, the induced nonlinearity can be modeled by developing the output signal equations of a simple amplifier [34], [35]. The injection path of the EMI attack is different from the previous attacks such as ultrasound commands where the commands are injected through the membrane [9]–[12]. However, the IEMI attacks are performed by injecting the signal to the electrical circuit.

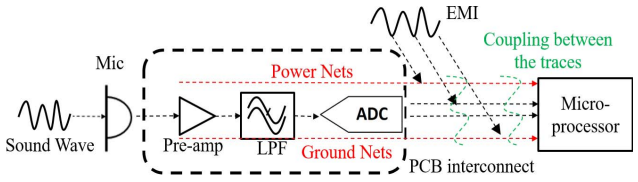


Fig. 1. Microphone circuit diagram and the anticipated coupling path.

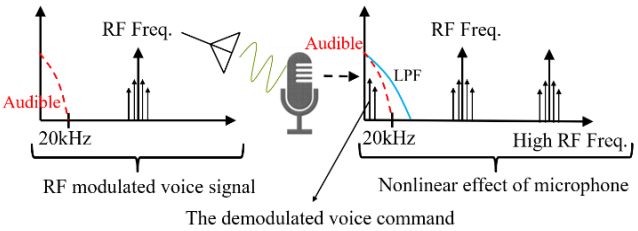


Fig. 2. Demodulation due to the inherent nonlinearity of microphones.

All the circuits are having components that can couple the EM signal efficiently from MHz to GHz depends on the resonant frequency of the receiving circuit. Once the EM signal is coupled to the PCB, the traces on the PCB will deliver the signal to the microphone module.

C. Nonlinearity of Microphones

A typical microphone system in a smart speaker consists of four primary blocks, as shown in Fig. 2 [9]–[12]: a microelectricalmechanical system (MEMS)-based microphone sensor [28]–[31], an amplifier, a low-pass filter (LPF), and an analog-to-digital converter (ADC). The acoustic waves passing through the microphone sensor induce vibrations in the membrane and are processed by the rest of the circuit. Most microphones are designed to only capture voice commands below 24 kHz. An amplifier is used because the captured voice command is too low in amplitude to be processed by the ADC. The ADC quantifies the signal levels, usually with a sampling rate of twice the maximum voice signal frequency. As a result, audio signals with frequencies greater than 24 kHz will be removed by the LPF.

As previously reported, nonlinearity is induced in the microphone circuit. The nonlinearity can be expressed by the following equation:

$$S_{\text{out}} = aS_{\text{in}} + bS_{\text{in}}^2 + \cdots dS_{\text{in}}^4 + mS_{\text{in}}^n. \tag{1}$$

In general, the coefficients of the higher order terms decrease dramatically, with the coefficients $m \ll c \ll b$; hence, only the second-order coefficient needs to be considered for the nonlinearity [10], [28].The attack signal, $A\cos_i t$, is multiplied with the carrier signal, $B\cos_r t$, to generate the

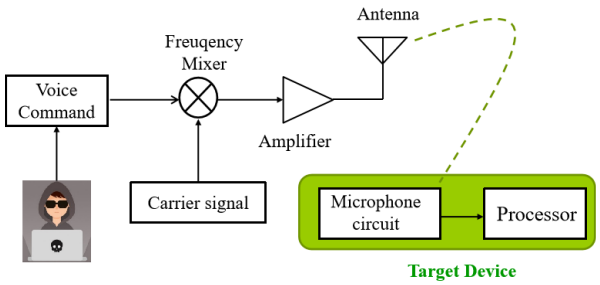


Fig. 3. Intentional EMI attack schematic.

amplitude-modulated signal

$$A\cos\omega_i t \times B\cos\omega_r t = \frac{AB}{2}[\cos(\omega_r - \omega_i)t + \cos(\omega_r + \omega_i)t] \tag{2}$$

where A and B are the amplitude of the signals, $\omega_i = 2.\pi f_i$ and $\omega_r = 2.\pi f_r$ represent the angular frequency of the attack and carrier signals, and f_i and f_r are the frequencies of the attack signal and carrier signal with relation $f_i \ll f_r$. Due to the second-order term of (1), S_{in}^2, the manipulated voice command will be shifted to the audible range as shown in (3). Since the carrier signal is a high-frequency signal, both low and high frequency components will present after the square function. The high frequency will be removed by the LPF in the microphone circuit, therefore, only low-frequency components presented in the following equation:

$$(A\cos\omega_i t \times B\cos\omega_r t)^2 \rightarrow \frac{A^2B^2}{4}\cos 2\omega_i t. \tag{3}$$

We assume f_i is the voice command below 10 kHz in the audible range. After the nonlinear operation of the microphone, low-frequency audible components up to 20 kHz containing the information of the voice command are generated. Since the spectrum of the audible output is doubled compared to the voice command, the voice command signal needs to be preprocessed before it is modulated into the attack signal so that the exact voice command can be recovered after this nonlinearity of the smart speaker.

D. Attack Setup

Fig. 3 shows a general setup of the attack. A computer or audio signal generator is used to generate the attack voice signal that is sent to a mixer. Different mixers can be applied to mix/modulate the attack signal to the carrier signal depending on the sweeping frequency band. A frequency synthesizer, vector network analyzer (VNA), or signal generator can be employed to generate the carrier signal. A power amplifier is used for amplifying the modulated signal. Since the unidirectional antennas radiate to all directions, a directional antenna is used to radiate more power in the dedicated direction of the modulated signal. The real attack setup is shown in Fig. 4.

III. Amplitude Modulated Attack Signal Manipulation

Assuming that we modulate a 2 kHz single tone signal to the carrier signal directly without pre-processing and send

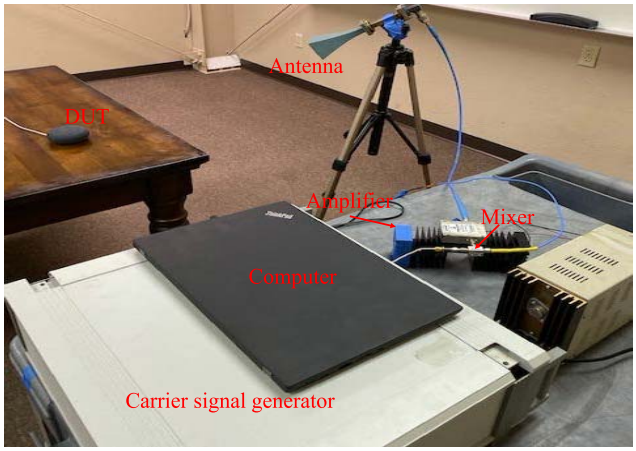


Fig. 4. Real attack setup.

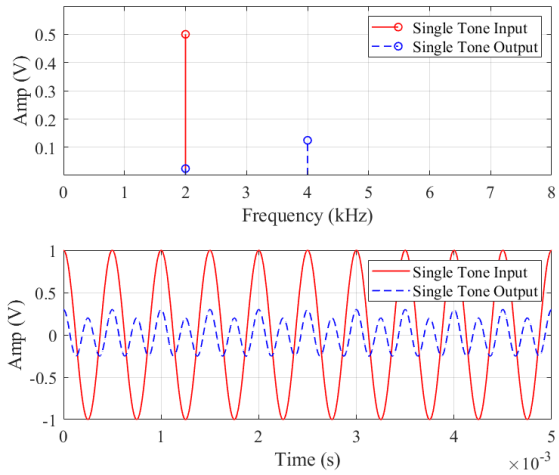


Fig. 5. Original single-tone input and its model output.

out to the microphone circuit, a square function exhibiting nonlinear behavior of the microphone circuit will be applied to the modulated signal. The resulting signal passes through the LPF in the microphone circuit, and only the low-frequency components remain. Through the mathematical derivation, the low-frequency component $\cos(\omega_i t)$ with $f_i = 2$ kHz and $\cos(2\omega_i t)$ with $2f_i = 4$ kHz are found after the LPF as shown in the equation below

$$(A\cos\omega_i t \times B\cos\omega_r t + F\cos\omega_r t)^2 \rightarrow \frac{A^2B^2}{4}\cos 2\omega_i t + ABF\cos\omega_i t \tag{4}$$

where $\cos(\omega_r t)$ is the feed-through component generated by the mixer due to the limited isolation of the mixer. The measurement of the modulated signal through the mixer exposed this feed through component. And this component has been applied in all the computations in the following sections. As shown in Fig. 5, the generated 4 kHz is much stronger than the 2 kHz signal. Since we want to recover the attack signal of 2 kHz after the microphone's nonlinearity, the preprocessing of the attack signal needs to be performed.

A. Dc Added Attack Signal Preprocessing

By adding the dc component to the attack signal, still using a 2-kHz signal as an example, the model output will change.

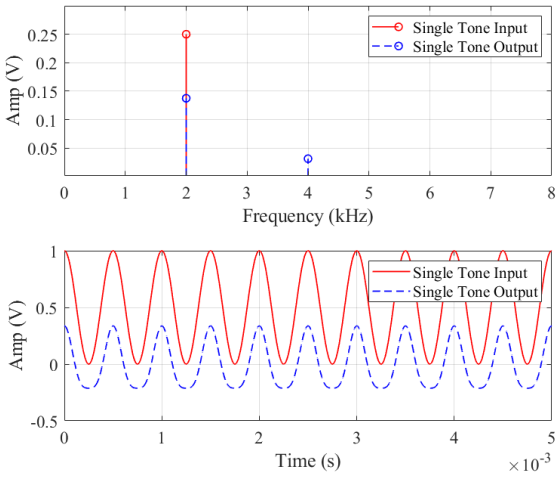


Fig. 6. Dc-added single-tone input and its model output.

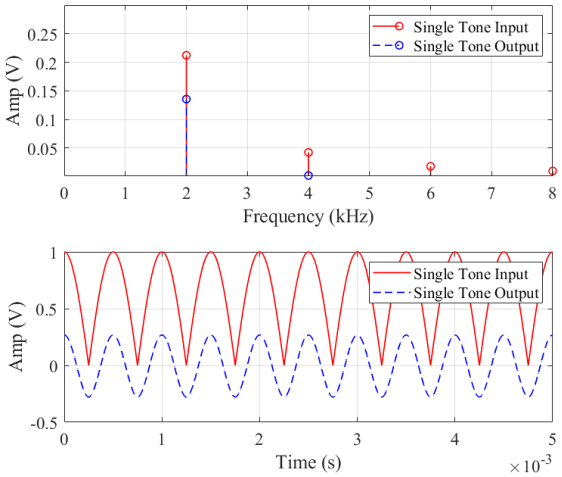


Fig. 7. Square-rooted single-tone input and its model output.

As shown in (5) below, where C is the amplitude of the dc component, after the LPF, we still have both $\cos\omega_i t$ and $\cos 2\omega_i t$, in other words, the 2- and 4-kHz signals, as shown in Fig. 6. But now the 2-kHz signal has a higher amplitude compared to the previous case. Notably, the time domain output waveform is deformed compared to the original signal waveform shown as the red solid curve in Fig. 5

$$((A\cos\omega_i t + C)\times B\cos\omega_r t + F\cos\omega_r t)^2 \rightarrow \frac{A^2B^2}{4}\cos 2\omega_i t + (ACB^2 + ABF)\cos\omega_i t. \quad (5)$$

To ensuring that the coefficient of the $cos2\omega_i t$ component is much smaller than the coefficient of the $cos\omega_i t$ component, as shown in (5), the relation in (6) can be developed

$$\frac{A^2B^2}{4} \ll (ACB^2 + ABF) \Rightarrow AB \ll 4CB + 4F. \quad (6)$$

$4(CB + F)/AB \gg 1$ should be the condition to minimize the $\cos 2\omega_i t$ component.

B. Square-Rooted Attack Signal Preprocessing

Since the nonlinearity is represented as the square term as shown in (1), a square root of the signal can be first performed. Therefore, after the square function of the signal, the original signal can be recovered. Since the computer can only output the real number of the signals, the dc value has to be added first before square root to avoid generating complex values. Continuing to preprocess the attack signal, the operation shown in the following equation can be performed:

$$\left(\sqrt{(A\cos\omega_i t + C)}\times B\cos\omega_r t + F\cos\omega_r t\right)^2 \rightarrow \frac{A^2B^2}{2}\cos\omega_i t + BF\sqrt{(A\cos\omega_i t + C)}. \quad (7)$$

As shown in Fig. 7, by applying this operation to the attack signal, we still have the $\cos 2\omega_i t$ signal (4 kHz), but it is much lower in amplitude and has less effect on the original signal, $\cos 2\omega_i t$. Moreover, the shape of the time domain output curve is well recovered compared with the dc added case. Therefore, the square-rooted injection signal is a better attack signal recovered in the smart speaker.

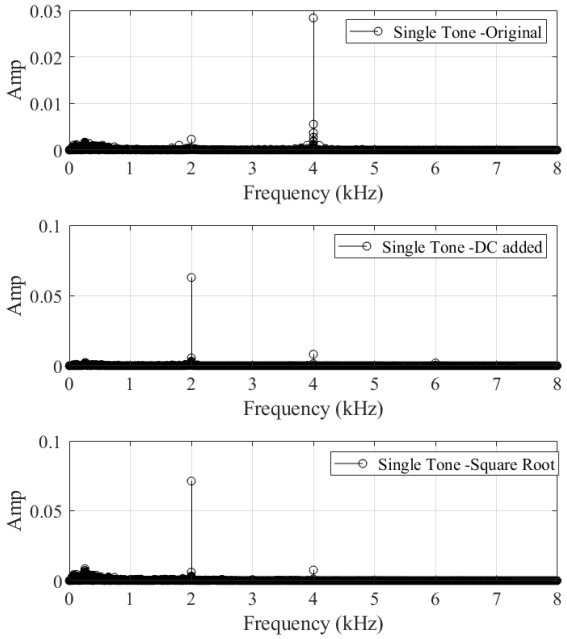


Fig. 8. Single-tone injection signal measurement.

C. Verification of the Preprocessed Attack Signal

To verify the preprocessing procedure, three types of injections are applied to a target device by applying the setup shown in Fig. 8. All the attacks in this section are performed at the sensitive frequency of the smart speaker, which will be explained in Section IV-A. After the frequency analysis of the recorded files, the results shown in Fig. 8 are obtained. The results from Fig. 8 prove that the square-rooted attack signal has the best performance when injected into the device.

D. Real Voice Command Attack Analysis

To be more confident on the attack signal preprocessing, some real voice command attacks were performed. One attack voice command was "What time is it?," and the target device responded with the current time. The command was sending continuously. Fig. 9 shows that the recorded voice signal matches well with the original signal. The square-root function of the original signal was applied to form the attack signal,

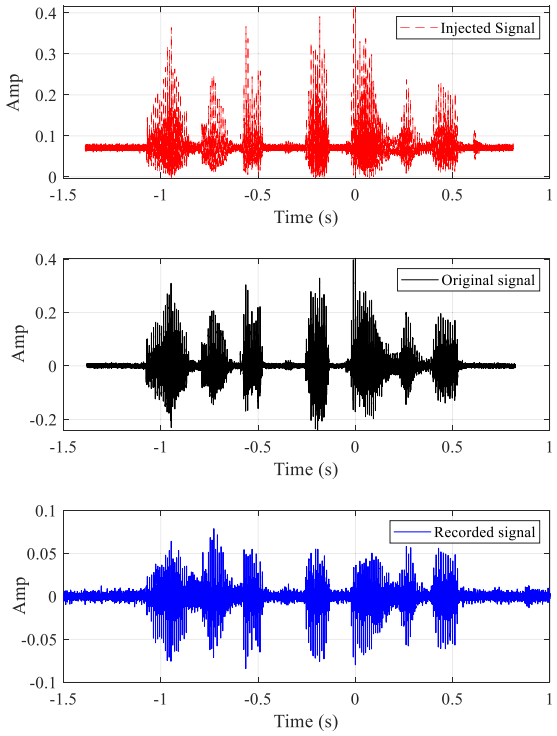


Fig. 9. Real voice command injection measurement.

TABLE I
MAXIMUM ATTACK DISTANCE BASED ON THE CURRENT SETUP WITH DIFFERENT ANTENNAS

Product	*Smart Speaker 1*	*Smart Speaker 2*	*Smart Speaker 3*	*Cellphone 1*
Maximum attack distance	2.5 m	40 cm	40 cm	20 cm
Minimum attack power density	2.94 Watts/m	39.3 Watts/m	39.3 Watts/m	157.2 Watts/m
Minimum attack electrical field intensity	150 dBuV/m	161.7 dBuV/m	161.7 dBuV/m	167.7 dBuV/m

and the resulting signal was then injected to the target device to ensure better signal recovery in the recorded file. However, without the preprocessing, the target device could not understand the voice command because the frequency of the signal changed due to the nonlinear effect.

At the maximum attack distance shown in Table I, the target devices can barely recognize the voice command. Therefore, we can analyze the efficiency of the different preprocessed attack signals with the peak-to-peak value normalized to 1. The recognition rates of the various preprocessed attack signals for different products are compared in Fig. 10, indicating that the square-rooted input has the best attack performance. The recognition rates are determined from the execution times of the target devices over ten attacks for each preprocessed attack signal.

IV. EXPERIMENTAL FEASIBILITY ANALYSIS

In this section, the sensitive carrier frequency is analyzed first; then, the sensitive location for EM coupling is studied based on the measurement without knowing the PCB layout.

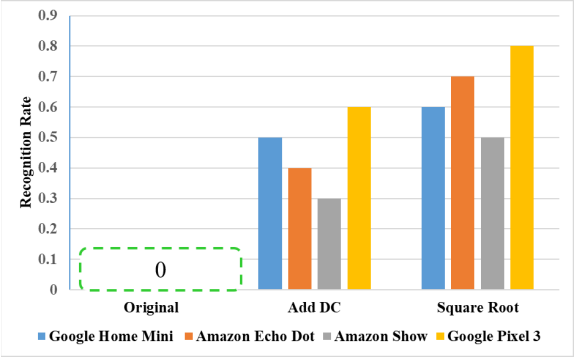


Fig. 10. Recognition rate for different preprocessed input signals and products.

The maximum attack distances are investigated for different smart speakers. And then the required electrical field densities are analyzed for different device under tests (DUTs).

A. Sensitive Carrier Frequency Analysis

The most sensitive frequency of the carrier signal needs to be identified to have efficient energy coupled to the smart speaker. In addition, attacking at the sensitive frequency can increase both the attack distance and the success rate. The following process can be applied to find the most sensitive frequency of the carrier signal for implementing an attack on a smart speaker.

1) A single-tone audible signal (2 kHz was chosen in this article; other single-tone signals within the audible frequency band can be chosen as well) modulated to the carrier signal is applied for the attack for simplicity, because a real voice command is a signal with multiple tones, which is difficult to define the amplitude because there might be some other noise in the recorded file.
2) Then, sweep the frequency of the carrier signal with the attack setup and send the modulated signal to the smart speaker with an activation voice command to wake up the device. Alternatively, let the device make a voice call to a phone which can record before sending the modulated signal.
3) The record file can be downloaded from the cloud because most smart speakers upload the voice command to the cloud automatically. Alternatively, the recorded file on the phone can be transferred to the computer for analysis.
4) Finally, the recorded file can be analyzed through Fast Fourier Transform (FFT) to determine whether the frequency harmonic at 2 kHz is present. By comparing the amplitude of the harmonic at 2 kHz, the sensitive frequency can be determined.

The frequency of the carrier signal was swept from 1 to 18 GHz with 1-GHz frequency step using the setup shown in Fig. 11. When the setup is fixed, the sweeping process was automated by programming the signal generator.

Fig. 12 shows the ratio of the power of the recorded 2-kHz component to the power of the attack signal at the antenna output for two different products, and the ratio is representing the transfer function from the antenna output to the record

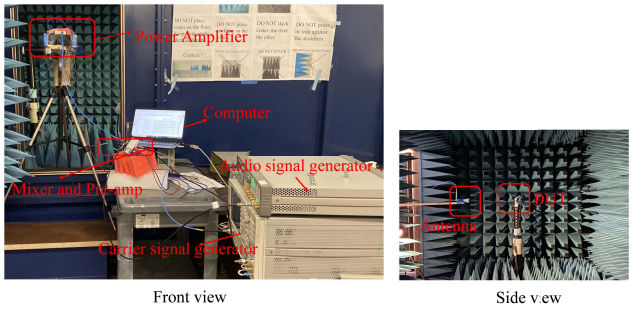


Fig. 11. Sensitive frequency measurement setup.

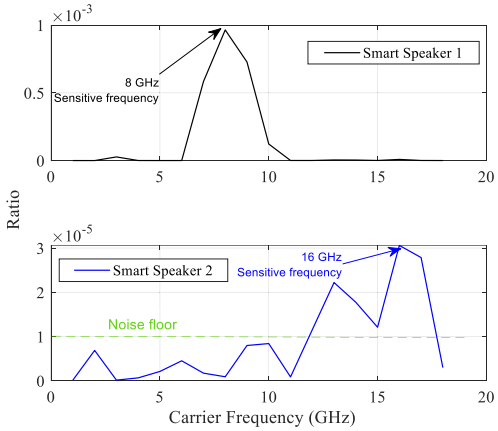


Fig. 12. Sensitive carrier signal frequency analysis for two different types of smart speakers.

file output. Four main propagation paths are included in this ratio: air propagation, coupling path, demodulation process, and record file. The same distances, 50 cm for *Smart Speaker 1*, 20 cm for *Smart Speaker* 2, are maintained for the different frequencies of the carrier signal. The sensitive frequencies of these two products are obtained at 8 and 16 GHz, respectively. From the amplitude ratios of the two products, we observe that the *Smart Speaker* 1 can easily be coupled at 8 GHz. Since the environmental noise may contain the audible signals that can be recorded by the devices, this may impact the final obtained results. Thus, the experiments need to be performed in a quiet room to have reliable results. The sensitive carrier signal frequency is found at 16 GHz for the *Smart Speaker* 2. Although the ratio is very low, the attack still succeeded because the application layer of different smart speakers has different decisions on the input signal level.

B. Measurement-Based Sensitive Location Analysis

To explore how the EM wave was coupled to the smart speakers, a measurement-based methodology is proposed. The *Smart Speaker* 1 was used as the DUT to present our methodology. A near field injection technique was applied, as shown in Fig. 13. The field probe was used to inject the modulated EM signal which is different from the normal near-field scan that measures the electromagnetic field component at a scanning location. The rest of the setup is the same as in Figs. 4 or 11; only the antenna was replaced with a high-frequency field probe which was used to inject the modulated EM signal.

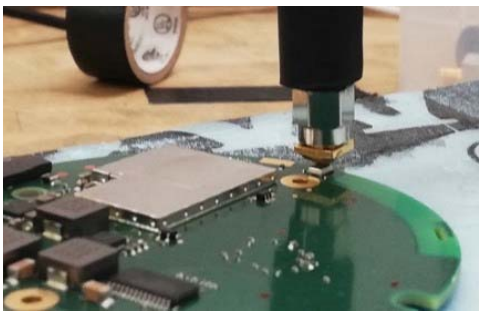


Fig. 13. Near-field injection setup.

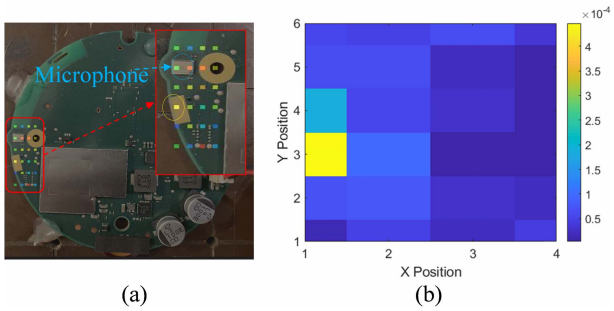


Fig. 14. Near-field injection region and results.

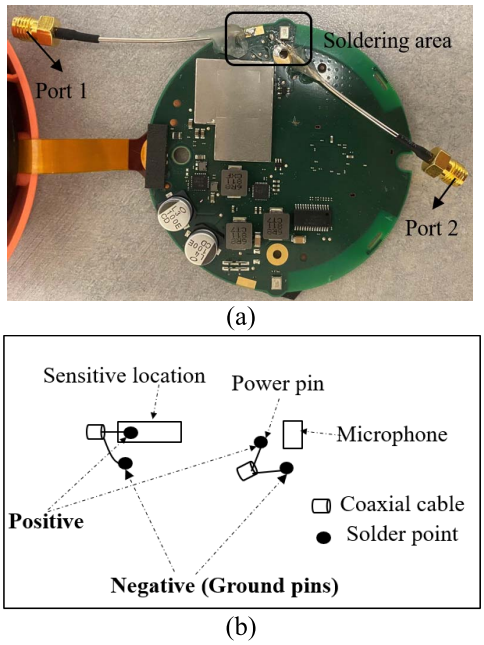


Fig. 15. (a) S-parameter measurement setup for the *Smart Speaker 1*. (b) Soldering area description.

At each position, the injection was performed and the recorded file was analyzed to monitor the strength of the 2 kHz signal. The injection area is the red-colored region shown in Fig. 14(a), where the microphone is located. Fig. 14(b) represents the 2-kHz magnitudes received in the recorded file at different locations. The results indicate that the yellow circled region is the most sensitive location.

To support that the sensitive location results in the highest noise level coupled to the microphone, the coupling path transfer function was obtained between the power pin of the microphone and the sensitive location. Fig. 15(a) shows

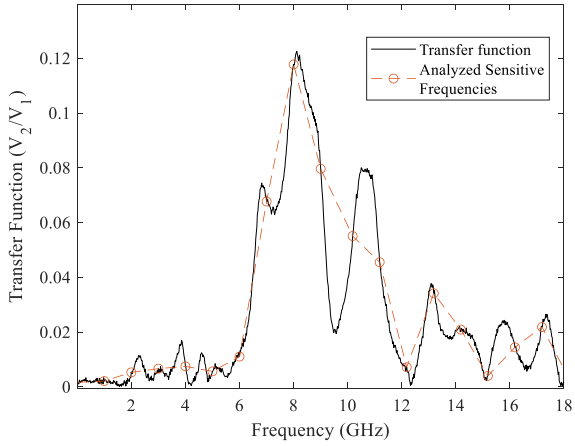


Fig. 16. Transfer function of the sensitive location.

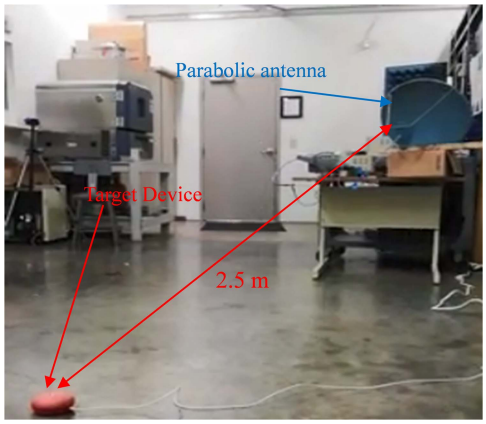


Fig. 17. Voice command attack on a *Smart Speaker 1* with a parabolic antenna.

the 2-port S-parameter measurement setup of the DUT. The positive terminals of the two identical coaxial cables were soldered on the sensitive location and the power pin of the microphone; the negative terminals are soldered on the adjacent ground pins, as shown in Fig. 15(b). The measured 2-port S-parameter data were transformed into the ABCD matrix to obtain the transfer function, as shown in Fig. 16. The plot in circles in Fig. 16 represents the analyzed sensitive frequencies in Fig. 12. It can be seen that the strongest coupling happens at around 8 GHz which meets our expectation based on the previous results shown in Fig. 12.

C. Maximum Attack Distance for Real Voice

The maximum attack distances for different target devices were achieved with a square-rooted attack voice command. The maximum distance reached for *Smart Speaker* 1 is 2.5 m with a parabolic antenna, as shown in Fig. 17. The antenna is giant, but the adversary can drive a truck on the street and sent the attack messages to the smart rooms, and we can also change the antenna to smaller ones. For different products, we obtained varying maximum attack distances based on the current setup with different antennas, as shown in Table I. The maximum attack distance varied from 20 cm to 2.5 m for different target devices with an output power of only 2.5 W, and the antenna gain varies from 15 to 22 dBi. Despite the attack distances are small for some products compared to previous works [8], the attack distance can be increased by employing a high power amplifier.

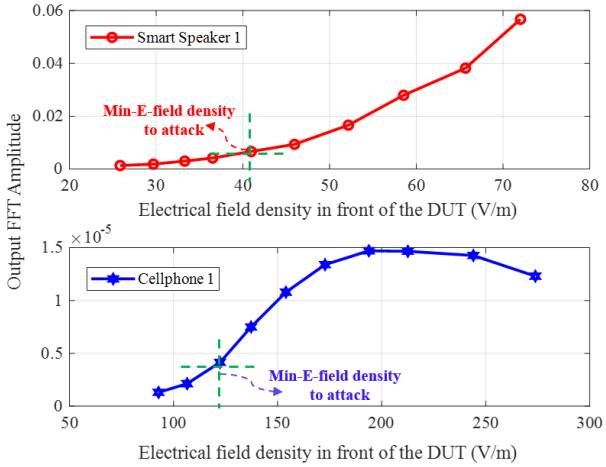


Fig. 18. Relation between input and output power.

D. Attack Electrical Field Density Analysis

In this analysis, the attack distance is fixed, and different attack powers are applied to generate different electrical field densities in front of the DUTs.

The power density in front of the smart speakers can be derived from the Friis transmission equation, as shown in the following equation:

$$P_D = \frac{P_t G_t}{4\pi d^2}. \tag{8}$$

The electric field strength at a given location can be obtained as follows [42]:

$$E = \sqrt{P_D Z_0} = \sqrt{120\pi P_D} \tag{9}$$

where P_t is the transmitter power (either the peak or average power), G_t is the gain of the antenna, d is the distance, and Z_0 is the air impedance. In this case, the electric field strength in front of the device can be characterized. The minimum required power density and electrical field intensity in front of the smart speakers are listed in Table I.

The gain of the antenna is 18 dBi at 8 GHz for the *Smart Speaker* 1 attack and 22 dBi at 18 GHz for the *Cellphone* 1 attack. The single-tone audible output spectrum is obtained in the recorded files. The relation between the E-field density in front of the DUTs and the obtained single-tone audible output is shown in Fig. 18. The green lines indicate the minimum E-field densities needed for the different target devices to recognize a real voice command. The different target devices exhibit varying limits and coupling strengths; for example, to attack *Smart Speaker 1*, the required minimum E-field density in front of the device is around 40 V/m, with a distance of 20 cm. However, for *Cellphone 1*, the requirement is around 125 V/m. In addition, the recognition level varies due to the noise cancellation technique applied by the *Cellphone 1*. The coupling efficiency which is the ratio between the input and output power can be obtained by calculating the slope of the curve.

V. Conclusion

This article presents an optimized EM attack process and the sensitivity analysis. The mechanism of the nonlinearity in the microphone circuit was described. The attack signal was preprocessed to increase the probability of a successful attack based on the nonlinearity characteristics, and measurements were performed for the single-tone signal attack to illustrate the effectiveness of the attack signal preprocessing. In addition, a methodology for sensitivity frequency analysis was discussed in order to find the most sensitive carrier frequency of a given product. The coupling sensitivity is studied based on near-field injection technique, and the transfer function from the sensitive location to the microphone under test is measured. The real voice commands were also successfully injected and executed by the target devices. Different maximum distances have been reached for different target devices. Generally, the maximum distance is depending on the output power of the antenna and types of DUT. A model can be built to estimate the required attack power (output power from the antenna or the power density in front of the device). Thus, the designer can optimize their device based on their standards regarding attackable distance and power. To eliminate the attack possibility, some countermeasures are proposed below for future studies:

A. Layout Optimization

Most EM threats arise due to an unintentional antenna structure [33], [34], [36] associated with the PCB layout design. Additional efforts to minimize exposed traces in the outer layers can reduce EM coupling. Moreover, the unintentional antenna structure near the microphone can act as an antenna to receive the I-EMI signal and conduct it to the microphone, allowing the microphone to demodulate the voice command.

B. Shielding Technique

Because the EM field must travel to the microphone circuit, a full structure shielding technique can be integrated into the device by exposing only the necessary parts, for example, by including a small hole for the microphone. An outer metal shield will prevent the field from coupling to the interconnects of the microphone circuit. Although the cost will increase, security risks can be minimized.

C. Inaudible Voice Command Detection

RF modulated signals operate at high frequencies; thus, another circuit can be added to detect the high-frequency component, in parallel to the microphone circuit. If modulated RF signals are detected, the circuit can give a signal to the microphone to stop listening. Thus, the smart device will not execute the attack command.

The future investigation will conduct in the protection technique development.

References

[1] D. Brannon. (Jun. 2018). *Attacking Private Networks from the Internet with DNS Rebinding*. [Online]. Available: https://medium.com/@brannondorsey/attacking-private-networks-from-the-internet-with-dns-rebinding-ea7098a2d325

[2] X. Xiao and S. Kim, "A study on the user experience of smart speaker in China-focused on Tmall Genie and Mi AI speaker," *J. Digit. Converg.*, vol. 16, no. 10, pp. 409–414, 2018.

[3] W. Diao, X. Liu, Z. Zhou, and K. Zhang, "Your voice assistant is mine: How to abuse speakers to steal information and control your phone," in *Proc. 4th ACM Workshop Secur. Privacy*, Nov. 2014, pp. 63–74.

[4] D. Kumar *et al.*, "Skill squatting attacks on Amazon Alexa," in *Proc. 27th USENIX*, 2018, pp. 33–47.

[5] N. Zhang, X. Mi, X. Feng, X. Wang, Y. Tian, and F. Qian, "Dangerous skills: Understanding and mitigating security risks of voice-controlled third-party functions on virtual personal assistant systems," in *Proc. IEEE Symp. Secur. Privacy (SP)*, San Francisco, CA, USA, May 2019, pp. 1381–1396.

[6] I. Clinton, L. Cook, and S. Banik, "A survey of various methods for analyzing the Amazon echo," Citadel, Military College South Carolina, Charleston, SC, USA, Tech. Rep., 2016. [Online]. Available: https://vanderpot.com/2016/06/amazon-echo-rooting-part-1/ and https://vanderpot.com/Clinton_Cook_Paper.pdf

[7] J. Lau, Z. Benjamin, and S. Florian, "Alexa, are you listening: Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers," *Proc. ACM Hum.-Comput. Interact.*, vol. 2018, no. 2, pp. 1–31, 2018.

[8] T. Sugawara, B. Cyr, S. Rampazzi, D. Genkin, and K. Fu, "Light commands: Laser-based audio injection attacks on voice-controllable systems," in *Proc. 29th USENIX Secur. Symp.*, 2020, pp. 2631–2648.

[9] G. Zhang, C. Yan, X. Ji, T. Zhang, and W. Xu, "Dolphinattack: Inaudible voice commands," in *Proc. ACM SIGSAC.*, 2017, pp. 103–117.

[10] N. Roy, S. Shen, H. Hassanieh, and R. R. Choudhury, "Inaudible voice commands: The long-range attack and defense," in *Proc. 15th USENIX Symp. NSDI.*, 2018, pp. 547–560.

[11] L. Song and M. Prateek, "Poster: Inaudible voice commands," in *Proc. ACM SIGSAC*, 2017, pp. 2583–2585.

[12] Q. Wang, K. Ren, M. Zhou, T. Lei, D. Koutsonikolas, and L. Su, "Messages behind the sound: Real-time hidden acoustic signal capture with smartphones," in *Proc. 22nd Annu. Int. Conf. Mobile Comput. Netw.*, Oct. 2016, pp. 29–41.

[13] R. Iijima *et al.*, "Audio hotspot attack: An attack on voice assistance systems using directional sound beams and its feasibility," *IEEE Trans. Emerg. Topics Comput.*, early access, Nov. 19, 2019, doi: 10.1109/TETC.2019.2953041.

[14] C. Yan, G. Zhang, X. Ji, T. Zhang, T. Zhang, and W. Xu, "The feasibility of injecting inaudible voice commands to voice assistants," *IEEE Trans. Dependable Secure Comput.*, early access, Mar. 19, 2019, doi: 10.1109/TDSC.2019.2906165.

[15] J. Mao, S. Zhu, X. Dai, Q. Lin, and J. Liu, "Watchdog: Detecting ultrasonic-based inaudible voice attacks to smart home systems," *IEEE Internet Things J.*, vol. 7, no. 9, pp. 8025–8035, Sep. 2020.

[16] C. Kasmi and J. L. Esteves, "IEMI threats for information security: Remote command injection on modern smartphones," *IEEE Trans. Electromagn. Compat.*, vol. 57, no. 6, pp. 1752–1755, Dec. 2015.

[17] D. F. Kune *et al.*, "Ghost talk: Mitigating EMI signal injection attacks against analog sensors," in *Proc. IEEE Symp. Secur. Privacy*, Berkeley, CA, USA, May 2013, pp. 145–159.

[18] J. Selvaraj, "Intentional electromagnetic interference attack on sensors and actuators," Ph.D. dissertation, Dept. Elect. Eng., Iowa State Univ., Ames, IA, USA, 2018.

[19] C. Kasmi and J. L. Esteves, "Electromagnetic threats for information security: Ways to chaos in digital and analogue electronics," in *Proc. 34c3 Chaos Commun. Congr.*, Dec. 2017.

[20] D. You, R. R. Jones, P. H. Bucksbaum, and D. R. Dykaar, "Generation of high-power sub-single-cycle 500- fs electromagnetic pulses," *Opt. Lett.*, vol. 18, no. 4, pp. 290–292, 1993.

[21] R. W. P. King, "Lateral electromagnetic waves and pulses on open microstrip," *IEEE Trans. Microw. Theory Techn.*, vol. 38, no. 1, pp. 38–47, Jan. 1990.

[22] M. Friedman and F. Fernsler, "Low-loss RF transport over long distances," *IEEE Trans. Microw. Theory Techn.*, vol. 49, no. 2, pp. 341–348, Feb. 2001.

[23] M. Okoniewski and M. A. Stuchly, "Modeling of interaction of electromagnetic fields from a cellular telephone with hearing aids," *IEEE Trans. Microw. Theory Techn.*, vol. 46, no. 11, pp. 1686–1693, Nov. 1998.

[24] D. L. Birx, "High power electromagnetic pulse driver using an electromagnetic shock line," U.S. Patent 5 319 665, Jun. 7, 1994.

[25] J. M. Snow and T. M. Snow, "Electromagnetic weapon," U.S. Patent 7 051 636, May 30, 2006.

[26] D. V. Giri, *High-Power Electromagnetic Radiators: Nonlethal Weapons and Other Applications*. Cambridge, MA, USA: Harvard Univ. Press, 2004.

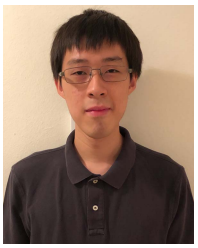
[27] T. D. Mast, L. M. Hinkelman, M. J. Orr, V. W. Sparrow, and R. C. Waag, "Simulation of ultrasonic pulse propagation through the abdominal wall," *J. Acoust. Soc. Amer.*, vol. 102, no. 2, pp. 1177–1190, Aug. 1997.
[28] Y. Zhong, Q. Huang, T. Enomoto, S. Seto, K. Araki, and C. Hwang, "Measurement based characterization of buzz noise in wireless devices," in *Proc. IEEE Int. Symp. Electromagn. Compat.*, Long Beach, CA, USA, Jul. 2018, pp. 134–138.
[29] J. W. Weigold, T. J. Brosnihan, J. Bergeron, and X. Zhang, "A MEMS condenser microphone for consumer applications," in *Proc. 19th IEEE Int. Conf. Micro Electro Mech. Syst.*, Jan. 2006, pp. 86–89.
[30] S. Castro, R. Dean, G. Roth, G. T. Flowers, and B. Grantham, "Influence of acoustic noise on the dynamic performance of MEMS gyroscopes," in *Proc. ASME Int. Mech. Eng. Congr. Expo.*, 2007, pp. 1825–1831.
[31] R. N. Dean *et al.*, "A characterization of the performance of a MEMS gyroscope in acoustically harsh environments," *IEEE Trans. Ind. Electron.*, vol. 58, no. 7, pp. 2591–2596, Jul. 2011.
[32] S. Shahparnia and O. M. Ramahi, "Electromagnetic interference (EMI) reduction from printed circuit boards (PCB) using electromagnetic bandgap structures," *IEEE Trans. Electromagn. Compat.*, vol. 46, no. 4, pp. 580–587, Nov. 2004.
[33] C. Hwang *et al.*, "Noise coupling path analysis for RF interference caused by LCD noise modulation," in *Proc. IEEE Int. Symp. Electromagn. Compat. (EMC)*, Jul. 2016, pp. 348–352.
[34] J. Gago, J. Balcells, D. GonzÁlez, M. Lamich, J. Mon, and A. Santolaria, "EMI susceptibility model of signal conditioning circuits based on operational amplifiers," *IEEE Trans. Electromagn. Compat.*, vol. 49, no. 4, pp. 849–859, Nov. 2007.
[35] M. T. Abuelma'atti, "Analysis of the effect of radio frequency interference on the DC performance of bipolar operational amplifiers," *IEEE Trans. Electromagn. Compat.*, vol. 45, no. 2, pp. 453–458, May 2003.
[36] A. R. Rofougaran, "RF diversity antenna coupling structure," U.S. Patent 7 170 465, Jan. 30, 2007.
[37] J. Vincent, "Lyrebird claims it can recreate any voice using just one minute of sample audio," The Verge, 2017, vol. 24.
[38] P. H. Huart and L. K. Surazski, "Method and apparatus for reconstructing voice information," U.S. Patent 7 013 267, Mar. 14, 2006.
[39] B. W. Wah and D. Lin, "Transformation-based reconstruction for audio transmissions over the Internet," in *Proc. 17th IEEE Symp. Reliable Distrib. Syst.*, Oct. 1998, pp. 211–217.
[40] S. George. (Jul. 2019). *You Can Now Speak Using Someone Else's Voice With Deep Learning*. Accessed: Mar. 27, 2020. [Online]. Available: https://towardsdatascience.com/you-can-now-speakusing-someone-elses-voice-with-deep-learning-8be24368fa2b
[41] M. Kayla. *What You Need to Know About Voice Hacking*. Accessed: Mar. 27, 2020. [Online]. Available: https://productivitybytes.com/needknow-voice-hacking/
[42] R. J. Burkholder and P. H. Pathak, "Analysis of EM penetration into and scattering by electrically large open waveguide cavities using Gaussian beam shooting," *Proc. IEEE*, vol. 79, no. 10, pp. 1401–1412, Oct. 1991.

Zhifei Xu (Member, IEEE) received the master's degree in microelectronics from École Supérieure d'Ingénieurs en Électrotechnique et Électronique, Paris, in 2016. He received the Ph.D. degree in electrical engineering from the Université de Rouen Normandie, Mont-Saint-Aignan, France, in 2019.

He is currently a postdoctoral researcher with the Electromagnetic Compatibility Laboratory, Missouri University of Science and Technology, Rolla, MO, USA. He is also one of the contributors on the applications of Kron's model. His research area covers hardware security, signal/power integrity, and electromagnetic compatibility (EMC) analysis in high-speed interconnect systems.

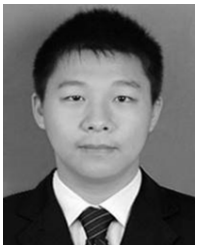
Runbing Hua (Graduate Student Member, IEEE) received the B.S. degree in electrical engineering from the Electromagnetic Compatibility Laboratory, Missouri University of Science and Technology, Rolla, MO, USA, in 2018, where she is currently pursuing the M.S. degree in electrical engineering.

Her research interests include intentional electromagnetic interference and hardware security, signal integrity, and system-level ESD modeling and testing.



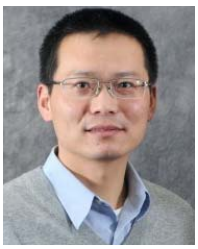
Jack Juang received the B.S. degree in electrical engineering from the Electromagnetic Compatibility Laboratory, Missouri University of Science and Technology (S&T), in 2020, where he is currently pursuing the M.S. degree.

His research interests include power distribution network modeling and optimization. He has been involved in projects related to the optimization of decoupling capacitor placement and RF susceptibility of smart devices.



Shengxuan Xia received the B.E. degree in electrical and computer engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2018. He is currently pursuing the Ph.D. degree in electrical engineering at the Electromagnetic Compatibility Laboratory, Missouri University of Science and Technology, Rolla, MO, USA.

His research interests include radio frequency interference, desense, radiated emission and susceptibility modeling, and hardware security.



Jun Fan (Fellow, IEEE) received B.S. and M.S. degrees in electronic engineering from Tsinghua University, Beijing, China, in 1994 and 1997, respectively. He received the Ph.D. degree in electrical engineering from the University of Missouri-Rolla, Rolla, MO, USA, in 2000.

From 2000 to 2007, he worked with NCR Corporation, San Diego, CA, USA, as a Consultant Engineer. In July 2007, he joined the Missouri University of Science and Technology (formerly the University of Missouri-Rolla) and became a Tenured Professor in 2016. He was also a Senior Investigator with the Missouri S&T Material Research Center. From 2013, he has served as the Director for the Missouri S&T EMC Laboratory and the Director for the National Science Foundation (NSF) Industry/University Cooperative Research Center (I/UCRC) for Electromagnetic Compatibility (EMC). From October 2018, he has been the Cynthia Tang Missouri Distinguished Professor in Computer Engineering. His research interests include hardware design and fundamental research for electromagnetic compatibility (including signal and power integrity) at the levels of integrated circuit, package, PCB, and system, and the development of specialized design tools and innovative measurement technologies.

Dr. Fan served as a member of the Board of Directors, the Chair of the Technical Advisory Committee, the Chair of the TC-9 Computational Electromagnetics Committee, and a Distinguished Lecturer. He currently is an Associate Editor for the IEEE TRANSACTIONS ON ELECTROMAGNETIC COMPATIBILITY and IEEE ELECTROMAGNETIC COMPATIBILITY MAGAZINE. He was the Technical Paper Chair and Technical Program Chair for a few IEEE INTERNATIONAL SYMPOSIA ON ELECTROMAGNETIC COMPATIBILITY, the General Chair for IEEE International Conference on Signal and Power Integrity, the Founding Chair for the SC-4 EMC for Emerging Wireless Technologies Special Committee, and so on. He received an IEEE TRANSACTIONS ON ELECTROMAGNETIC COMPATIBILITY Society Technical Achievement Award in August 2009.

Chulsoon Hwang (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2007, 2009, and 2012, respectively.

From 2012 to 2015, he was a Senior Engineer with Samsung Electronics, Suwon, South Korea. In July 2015, he joined the Missouri University of Science and Technology (formerly the University of Missouri-Rolla), Rolla, MO, USA, where he is currently an Assistant Professor. His research interests include RF desense, signal/power integrity in high-speed digital systems, EMI/EMC, hardware security, and machine learning.

Dr. Hwang was a recipient of the AP-EMC Young Scientist Award, the Google Faculty Research Award, and Missouri S&T's Faculty Research Award. He was a co-recipient of the IEEE TRANSACTIONS ON ELECTROMAGNETIC COMPATIBILITY Best Paper Award, the AP-EMC Best Paper Award, and a two-time co-recipient of the DesignCon Best Paper Award.