2023 IEEE Symposium on Electromagnetic Compatibility & Signal/Power Integrity (EMC+SIPI) | 979-8-3503-0976-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/EMCSIPI50001.2023.10241712

Coupling Path Analysis for Smart Speaker Intentional Electromagnetic Interference Attacks

Tanner Fokkens[1], Shengxuan Xia[2], Aaron Harmon[3], Chulsoon Hwang[4]
EMC Laboratory
Missouri University of Science and Technology
Rolla, MO, USA
[1]trfk76, [2]sx7c3, [3]afhpq4, [4]hwangc@mst.edu

Abstract—**This paper shows an improved understanding of the coupling path for intentional electromagnetic interference (IEMI) attacks on smart speaker devices. This includes a method for finding the ideal attack angle and locating the region sensitive to the coupled EMI. In previous works, it was shown to be possible to send RF commands to a smart speaker and have these commands be interpreted as voice commands by the microphone. However, the attack still had some limited understanding in terms of the coupling path location and long-distance attack potential. Using the improved understanding of the attack, a longer attack distance is achieved (6 meters) with only 6.5 Watts of power.**

Keywords— Smart Speakers, IEMI, Hardware Security, Susceptibility

I. Introduction

A smart speaker is a device that executes commands based on a user's voice. These devices started as simply allowing for basic conversational skills, but later full-function smart home devices began allowing the smart speakers to control functions like electric plugs, locks, or thermostats. Eventually, this same technology made its way to most smart phones and computers.

In these smart-speaker devices, most utilize microphones with micro-electromechanical system (MEMS) technology [1]. MEMS microphones work like acoustic microphones, but they are active microphones that often have self-contained analog to digital converters and amplifiers. The primary benefit of these microphones is that they have lower power consumption and reduced footprints. These microphones receive an analog voltage signal from the vibrated membrane, which is then amplified and digitized all within the microphone.

This new type of microphone, while having marked benefits, also introduced unintended security implications. Ultrasounds attacks were the first to be discovered [2], followed by laser-based attacks [7]. These worked through having an ultrasound or laser signal carry a modulated audio command, which is then coupled onto the MEMS microphone circuit of the smart speaker. These two attack methods were not effective because they work through a line-of-sight attack mechanism.

Recently, it was discovered that these commands can be sent through IEMI [1]. This attack is notably more effective than the ultrasound-based attacks in that the attack can be performed through walls. The study of Buzz noise proved that an audio coupling path exists through EMI from the nearby Wi-Fi antenna to the microphone [6]. Based on the understanding of Buzz noise mechanism, the IEMI was firstly demonstrated in [1]. This attack was shown to be possible by modulating the audio range attack signal with a much higher (6 to 18 GHz) carrier signal that was radiated using a highly directional horn antenna.

Additionally, it was shown that this attack used in conjunction with machine learning techniques can be used to circumvent a common feature in smart speakers and phones that allows for the device to learn the voice of the speaker and only wake up to that person's voice [10]. In the work shown in [1], a suspected coupling path was proposed, but there was no experimental verification for the suspected most sensitive point on the device. Additionally, the attack range was more limited due to less understanding about ideal attack angles and antenna polarization relative to the smart speaker.

In this work, the IEMI attack is described in section II. Next, the previous understanding of the point on the smart speaker that is sensitive to the IEMI attack is re-evaluated and revised in section III. After the correct identification of the sensitive point, the sensitive point was validated by creating a simulated model with this sensitive point chosen as the driven port in section IV. Through simulation and measurement, the most sensitive attack angles to the attack were correctly identified, validating the simulation model. These methods were used to greatly increase the attack distance, shown in section V, further showing the necessity for mitigation against this attack. Finally, in section VI, some potential mitigation strategies against the attack are proposed based on the findings in this paper.

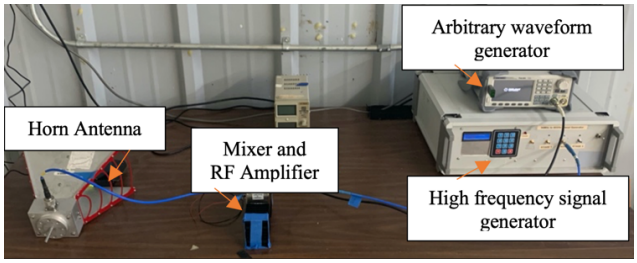
II. IEMI Attack Description and Signal Processing

The test setup for the IEMI attack is shown in Fig. 1a and represented as a block diagram in Fig. 1b. An arbitrary waveform generator was used to generate the audio range signal (an audio-range voice command), and a high frequency signal generator generates the carrier wave that is modulated by the audio signal.

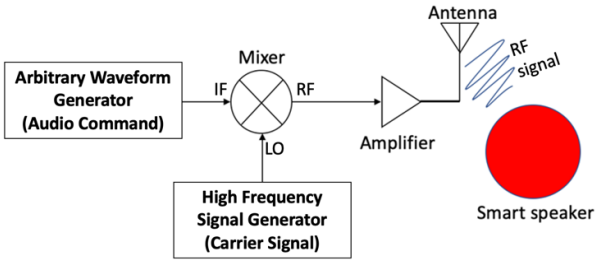
The audio commands were modulated to the high-frequency carrier by the mixer and then radiated from the attacking antenna to the microphone. The effectiveness of the attack is related to the carrier frequency. The optimal carrier signal was found by sweeping the modulating (carrier) frequency using the

This work was supported in part by the National Science Foundation under Grant No. IIP-1916535.

test setup shown in Fig. 1a with everything else held constant (including the modulated audio). Then, the recorded attack audio amplitude was retrieved from the given device online cloud and compared to the modulating frequency to find the most effective frequency for the attack. The attack carrier frequency itself was found as a transfer function of the sent attack audio amplitude versus the received command amplitude (in the device's cloud service) in [1].



(a)



(b)

Fig. 1. (a) The experimental test setup for the finding the most sensitive carrier frequency for the attack (b) Block diagram representation of the test setup. The smart speaker is placed in front of the horn antenna in the image.

The modulated signal radiates out from the aggressor antenna and then the EM wave is received at the MEMS microphone. From here, the signal is coupled back to the internal amplifier in the MEMS microphone where the non-linearity is present that causes the signal to be demodulated back to the audible range. It is then converted to digital in the analog-digital converter (ADC) of the microphone where it can be processed by the microprocessor.

Audio range amplifiers have linear amplification in the audible range when not driven to distortion. However, strong non-linearity can be observed in the non-audible range [10]. The mechanism of the non-linearity for the IEMI attack was found to be associated with the pre-amplifier that is self-contained in the MEMS microphone [4][7]. The output signal that results from amplifier nonlinearity is expressed in (1):

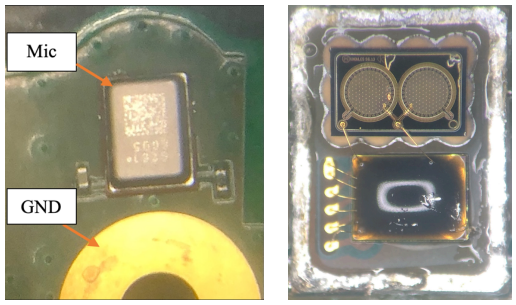
$$S_{out} = A_1 S_{in} + A_2 S^2{}_{in} + \cdots + A_4 S^4{}_{in} + A_n S^n{}_{in}. \quad (1)$$

Where S_{out} is the signal that results from the non-linearity effect, S_{in} is the input signal in the IEMI attack, and the coefficients A_n represent the amplitude of the resulting coefficients, with each subsequent one decreasing in magnitude. Coefficients n represents infinite order coefficients for this series. Previous work has shown that each subsequent S_n term decreases in magnitude strongly with each iteration [3][6], so only the S^2 term from (1) needs to be considered for this attack. The S^2 term produces both a high and low frequency component. The lower frequency component is less than the cutoff frequency of the low-pass filter of the MEMS microphone after the nonlinearity occurs within the microphone, so only the audio range signal is maintained.

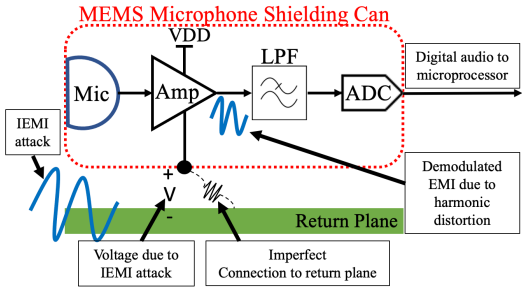
The square term of (1) necessarily produces harmonic distortion, which degrades the quality of voice injected into the system. To minimize the harmonic distortion associated with the demodulation process for the IEMI attack, an effective processing method is proposed in [1] as below (2):

$$S_{in} = \sqrt{Bf(t) + B}. \quad (2)$$

A DC offset equal to the maximum amplitude B of the audio waveform is added under the square-root to avoid imaginary part created by the square root function. This DC offset is not an issue as the internal coupling of the mixer removes this component.



(a)



(b)

Fig. 2 (a) Microphone can and PCB return plane connections for the direct injection of the IEMI attack waveform on smart speaker 1 (left), next to the decapped MEMS microphone (right), (b) Anticipated coupling mechanism and block diagram for the smart speaker IEMI attack.

A poor power supply rejection ratio (PSRR) of the internal amplifier at the attack frequencies is most likely the cause of the IEMI coupling from the outside of the microphone into the amplifier. The PSRR of an amplifier indicates how much of the change in voltage on the power rail of an amplifier will be translated onto the output. It is well known that the PSRR of an amplifier degrades with frequency [11]. Thus, if a voltage is

developing on the microphone can, which is the ground reference for the amplifier within the MEMS microphone, this variation caused by the IEMI attack will induce a voltage at the output of the internal amplifier through this poor PSRR. From here, the harmonic distortion of the internal amplifier will cause the demodulation described by equation (1). The anticipated coupling mechanism that results from the poor PSRR is shown in fig. 2b.

III. Identification of the Most Sensitive Point

Previously in [1], the sensitive point of a smart speaker (which is referred to as smart speaker 1) was identified using near field scanning. Specifically, the measurement indicated that the capacitive volume sensor was sensitive at the same frequency at which the attack was most effective. This speaker has large nearby metal structures close to the microphone in the enclosure that are thought to enhance the coupling.

To verify this finding in this work, a coaxial cable with an SMA connector was directly soldered to the capacitive switch on the PCB. After this coax-SMA was soldered to the PCB, the setup shown in Fig. 1b was used to directly inject the command known to wake the device into the smart speaker. The command was simply the 'wake word' for the device with the processing described in (2). When injecting audio to the point that was predicted to be susceptible at the 5 GHz sensitive frequency, however, no audio was heard, and the device would not interpret any direct-injected commands.

Given that the physical injection of the command should be equivalent to receiving the waveform through a radiated mechanism, this indicates that the previous understanding of the most sensitive point for the microphone could be incorrect. Thus, alternative techniques were used to find the true most sensitive point.

To start, copper tape was added to various potential coupling points on the smart speaker. The rational for this was that the added copper tape could serve as a more effective coupling structure to increase the coupling interference from the antenna to the microphone. It was observed that the audio amplitude would increase when the copper tape was added to the top of the microphone can structure, but not when it was added to the previously determined sensitive point. Thus, the direct command injection was performed once again onto the top of the microphone can as shown in Fig. 2a.

After the modification, the played back audio had a much higher amplitude during initial testing for a 5 GHz carrier signal, and the injected audio was far louder than the received amplitude purely from the normal microphone operation. This result indicated that the true sensitive point was between the top can of the microphone and the PCB return plane of the smart speaker. However, this result only showed that the sensitive point was located on the top of the microphone for this smart speaker specifically.

To see if this sensitive point location is the same for other speakers, a second smart speaker made by a separate manufacturer that does not have any obvious vulnerability to the IEMI attack was used. Additionally, the original smart speaker (smart speaker 1) also had monopoles added to the microphone can structure, while removing any nearby metal structures that enhance coupling. From here, monopoles were cut to a quarter-wavelength of the carrier frequency so that they resonate at 5 GHz (which was identified as a sensitive carrier frequency on other speakers). These monopoles were soldered to the top of the microphone can as shown in Fig. 3.

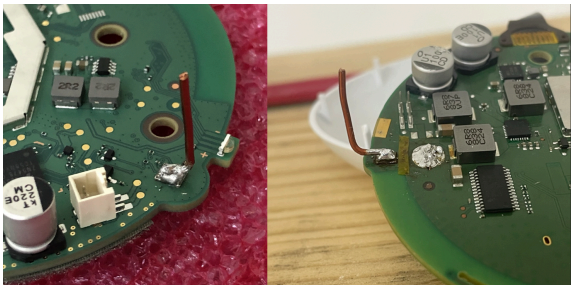


Fig 3. Adding monopole structures to the top of the second smart speaker's microphone can to try and cause the IEMI attack on a non-susceptible device.

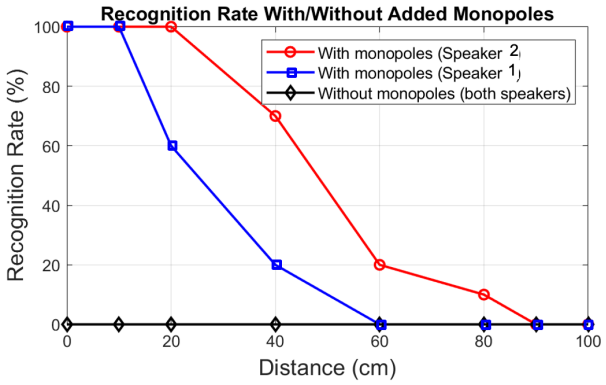


Fig. 4. Result of adding monopoles to the previously unsusceptible device.

While these devices previously showed no recognition at all (indicated by diamonds in Fig. 4), the structure with four monopoles added has command recognition out to 1 meter for the second smart speaker (indicated by circles), and 40 centimeters for the first smart speaker (indicated by squares). The videos relating to this experiment are available in [8]. From this experiment of adding monopole structures to the previously unsusceptible smart speakers, it was clear to see that the microphone can structure itself is the point of sensitivity for the IEMI attack. Additionally, this experiment and the observations seen by adding copper tape show that coupling is enhanced in the smart speaker by metal structures that are nearby to the microphone can.

IV. Ideal Attack Angle Analysis

A. Reasoning and Explanation for Attack Angle Prediciton

During testing of the smart speaker IEMI attack, it was observed experimentally that certain 'attack angles', meaning where the aggressor antenna was pointing in relation to the smart speaker, were more effective at causing the smart speaker IEMI attack. If these optimal angles for attacking could be determined through simulation techniques, it would save considerable time. The smart speaker has two microphones, so there are two susceptible points on the smart speaker based on the analysis in part III. To reduce complexity, one MEMS microphone was removed so the radiation pattern of one microphone can be measured.

The experimental setup for measuring the radiation pattern of this unintentional antenna is not straight-forward, as any augmentation of the most sensitive point, located on the can of the microphone, would change the radiation pattern of unintentional antenna. The chosen solution was to send the attack waveform with a single audio-range tone modulated to the sensitive frequency for this device, while sweeping the attack angles for this device. After radiating this attack waveform, the received audio amplitude at the smart speaker can be plotted as a function of the attack angle to determine the most effective attack angles.

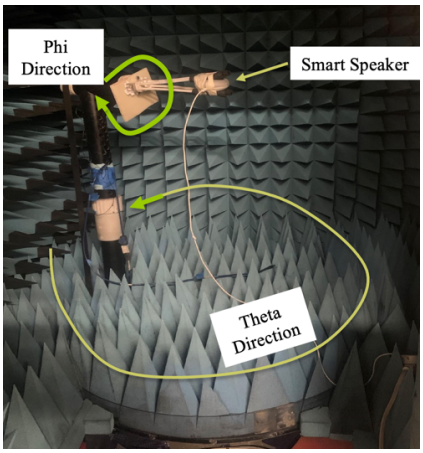
B. Measuring Smart Speaker Unintentional Antenna Radiation Pattern

The smart speaker was placed in an anechoic chamber that is equipped with a turntable and gimballed arm that can sweep both theta and phi angles 360 degrees. In this measurement, the turntable on the floor of the OTA chamber was chosen as the theta coordinate, while the part that rotates the DUT itself was chosen as the phi coordinate.

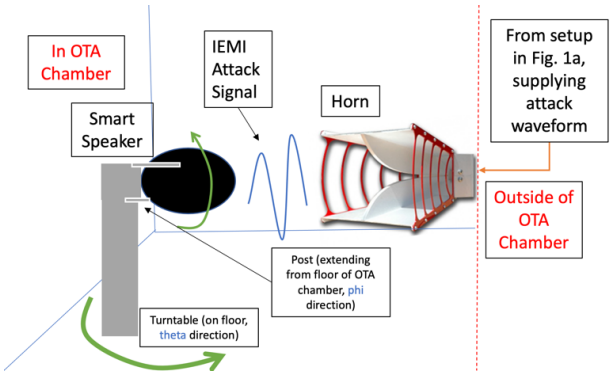
The coordinate system was developed in relation to the electric field vector position of attack antenna (polarization). For this setup, the E-field vector of the horn antenna was aligned with the theta axis. Fig. 5a shows the visual representation of this coordinate system with the arrow direction showing theta/phi spin directions for full 360-degree angle sweeping. In Fig. 5b, a diagram representing the setup is shown for improved clarity. In Fig. 5c, the coordinate system is drawn with reference to the smart speaker itself.

As in the simulation case, angle cuts for theta = 0° and 30° were measured. Data was recorded using an automated script that outputs a 1 kHz tone through the headphone connector of the computer. A 1 kHz tone was chosen because this tone is within the human speaking voice range, and command recognition was not the focus of the sensitive angle finding. The smart speaker is setup to continuously record audio during this test. The output sound of the headphone connection was sent to the IF port of the mixer for the setup shown in Fig. 1b.

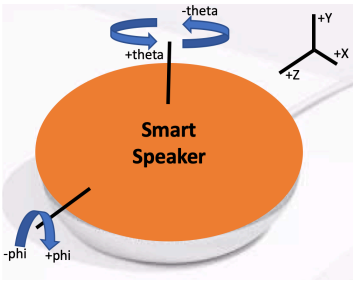
C. Modeling the Smart Speaker in Full-Wave Simulation



(a)



(b)



(c)

Fig. 5. (a) Coordinate system for this radiation pattern measurement test (b) System diagram for the measurement of the smart speaker radiation pattern for the unintentional antenna within the smart speaker allowing for the IEMI attack (c) Coordinate system relative to the smart speaker itself.

The structure was first modeled in a full-wave simulation tool. Given that the internal stack-up of the smart speaker PCB is not known, only the large metal features of the smart speaker are modeled. This includes the metal shield (which is separate from the can structure) that is placed over the microphone, the PCB ground, and PCB dielectric. Additionally, the Z11 of the loading for the sensitive point was included. This Z11 was measured using the same port connection shown in Fig. 2a. In Fig. 6, the resulting model can be seen.

To get proper comparison results between the simulation and measurement, the coordinate systems between the measurement and simulation should be matched. Additionally, the polarization of the electric field vector in the simulation was chosen so it was aligned with the theta direction, as it was in the measurement case.

As seen in Fig 6, the driven port for the simulation was placed between the driving port and PCB return plane. Ideally, this Z11 should represent the imperfect connection between the microphone can and PCB return plane at higher frequencies that causes a noise voltage. This Z11 was measured at the sensitive

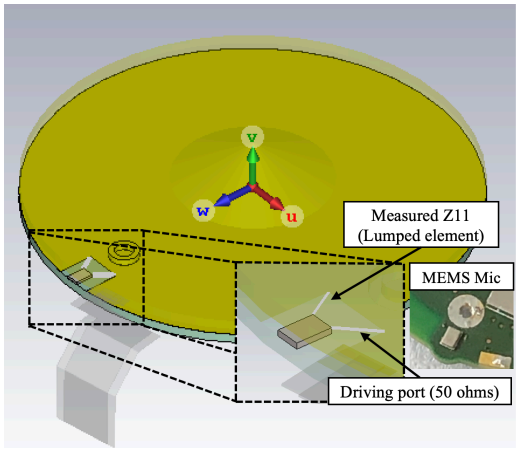


Fig. 6. Full-wave simulation Model of the internal structure of the smart speaker.

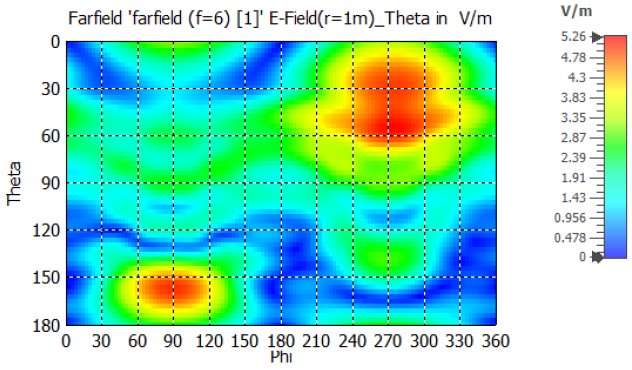


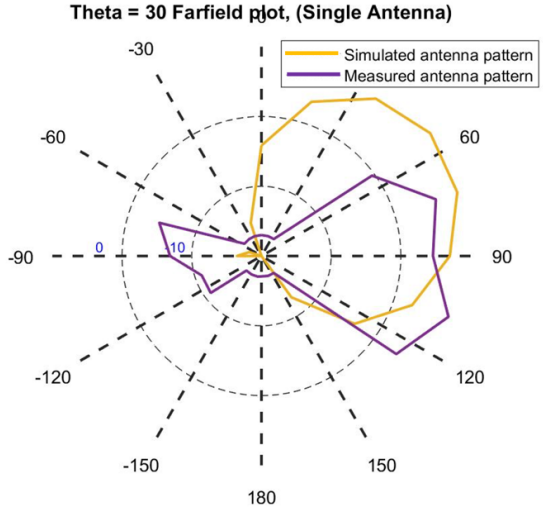
Fig. 7. Directivity of the modeled smart speaker for electric field vector oriented in the theta direction (polarization).

point shown in Fig. 2a and inserted into the model as a lumped element to add more physical characteristics to the model.

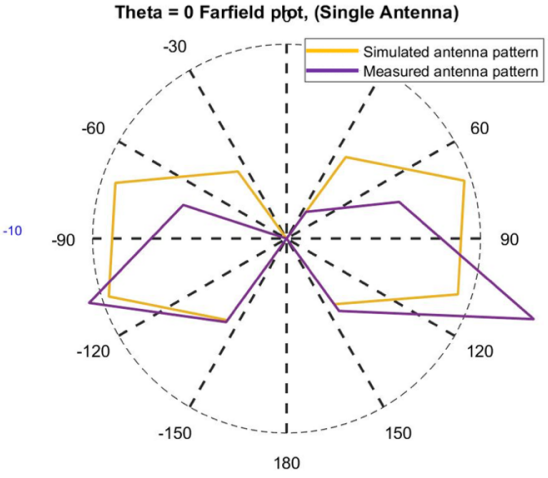
The port impedance for the driving port is selected as 50 ohms. Choosing 50-ohms as the port impedance will affect the unintentional antenna efficiency due to reflections, but not the pattern itself, which is the focus for finding the optimal attack angles.

When examining the directivity pattern in Fig. 7 (swept across theta and phi), areas of higher directivity are observed. At these combinations of theta/phi, the highest amplitude should be observed. The hotspots of high directivity are located where theta = 180 degrees/phi = 90 degrees, and theta = 30/ phi = 270 degrees.

These results are seen in Figure 8. From Figure 8a, there is a clear alignment between the simulation results and the measured results. The results were not perfectly aligned for this angle cut, but some variation is expected with this test because the measurement is carried out by simply plotting the received audio amplitudes and then normalizing them to the level of the simulation results.



(a)



(b)

Fig. 8 (a) Simulated versus Measured Directivity for the theta = 30 angle cut (in dB) (b) Simulated versus Measured Directivity for the theta = 0 angle cut (in dB)

For Figure 8b, there is better alignment. At phi=90 and phi=-90 degrees, the simulation and measurement results are relatively well aligned. These two results further support the finding that the sensitive point is located at the microphone can structure, and that the directivity for the unintentional antenna

can be obtained without knowing anything about the PCB of the smart speaker itself.

V. Long Distance Attacking

The IEMI attack was optimized to extend the range out to 6 meters by using the optimal attack angles shown in section IV of this work. Theta, in this case, was chosen to be 10 degrees. 30 degrees would be more optimal, but this was not possible with the available test space. This results in an angle that points towards the top of the smart speaker. This choice was based on the observation that increasing theta from 0 to 30 degrees resulted in increased directivity of the attack for a phi angle equal to 90 degrees. Subsequently, 90 degrees (for the shown coordinate system) was chosen for phi. This 90-degree attack angle means physically that the E-field vector of the parabolic dish must be perpendicular to the smart speaker.

This test shows the feasibility of the attack at ranges that exceed the distances (~1.5 meters) shown in the previous work [1]. A video was taken of the IEMI attack for the ~5 GHz carrier frequency case. The video of this test can be viewed at [9]. In the video, after enabling the audio-range signal, the smart speaker, which is placed 6 meters away, wakes up and says the time. This is because the audio command that was sent using the IEMI attack mechanism was "What time is it?".

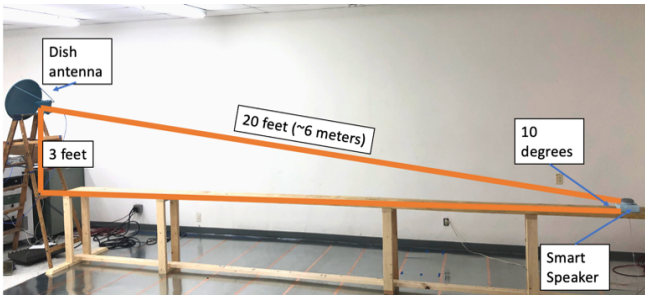


Fig. 9. Distance diagram for the 20-foot attack verification

VI. Conclusion and Mitigation Discussion

In this work, the methods for understanding the behavior of the smart speaker IEMI attack were presented and tested. A method for determining the sensitive point on the smart speaker was shown after the most sensitive point in the previous work was found to be insensitive to the IEMI attack. After the correct sensitive point was identified, the most effective attack angle was determined through simulation methods and subsequently compared with the measured most sensitive angles for alignment. Finally, a non-simulation experiment was used for a long-distance attack to show the feasibility of longer-distance attacks.

In the future, efforts should be made to mitigate this attack to prevent any security concerns associated with it. Based on the findings in section III, nearby metal structures located on or nearby the microphone structure can significantly enhance the effectiveness of the IEMI attack. To reduce this coupling, large floating metal structures, if they must be used, should be placed far away from the MEMS microphone to avoid direct capacitive coupling to the microphone.

Additionally, the high frequency connection from the microphone can structure to the PCB return plane of the smart speaker should be improved. Based on the prior measurement of the Z11 from the smart speaker to PCB return plane, the most sensitive frequencies were observed to happen where the impedance from the PCB return plane to microphone can structure was elevated. By improving this high frequency connection, the attack can be mitigated.

Finally, due to the role of the internal amplifier's non-linearities in the attack, efforts should be taken to improve the PSRR of this internal amplifier within the microphone to reduce the amount of noise that is coupled onto the microphone can structure that ends up in the output of the amplifier. This would entail characterizing how much power rail noise is coupled onto the output of this internal amplifier during the design process.

References

[1] Z. Xu, R. Hua, J. Juang, S. Xia, J. Fan and C. Hwang, "Inaudible Attack on Smart Speakers With Intentional Electromagnetic Interference," in *IEEE Transactions on Microwave Theory and Techniques*, vol. 69, no. 5, pp. 2642-2650, May 2021.

[2] Zhang, G., Yan, C., Ji, X., Zhang, T., Zhang, T., and Xu, W.,"Dolphinattack: Inaudible Voice Commands." *Proceedings of the 2017 ACM SIGSAC. ACM, 2017*

[3] Roy, N., Shen, S., Hassanieh, H., and Choudhury, R. R., "Inaudible Voice Commands: The Long-range Attack and Defense." *15th USENIX. USENIX, 2018.*

[4] J. Mao, S. Zhu, X.Dai, Q. Lin and J. Liu, "Watchdog: Detecting Ultrasonic-Based Inaudible Voice Attacks to Smart Home Systems," *IEEE Internet of Things Journal, vol. 7, no. 9, pp. 8025-8035, Sept. 2020*

[5] 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 Transactions on Electromagnetic Compatibility,* vol. 49, no. 4, pp. 849-859, Nov. 2007.

[6] Y. Zhong, Q. Huang, T. Enomoto, S. Seto, K. Araki, and C. Hwang, "Measurement Based Characterization of Buzz Noise in Wireless Devices." *IEEE Int. Symp. On Electromagnetic Compatibility*, Long Beach, CA, pp. 134-138, 2018.

[7] Takeshi Sugawara, Benjamin Cyr, Sara Rampazzi, Daniel Genkin, and Kevin Fu. 2020. "Light commands: Laser-based Audio Injection Attacks on Voice-Controllable Systems". *Proceedings of the 29th USENIX Conference on Security Symposium*. USENIX Association, USA, Article 148, 2631–2648.

[8] EMC-Laboratory. *Hardware-security/smart speaker I-EMI/videos/monopole-adding videos at main · EMC-Laboratory/Hardware-Security*. GitHub. Retrieved March 26, 2022, from https://github.com/EMC-Laboratory/Hardware-Security/tree/main/Smart%20Speaker%20I-EMI/Videos/Monopole-adding%20videos

[9] EMC-Laboratory. *Hardware-security/smart speaker I-EMI/videos/monopole-adding videos at main · EMC-Laboratory/Hardware-Security*. GitHub. Retrieved March 26, 2022, from https://github.com/EMC-Laboratory/Hardware-Security/blob/main/Smart%20Speaker%20I-EMI/Videos/6%20Meter%20attack.mp4

[10] T. Fokkens, Z. Xu, O. Hoseini Izadi and C. Hwang, "Machine Learning Voice Synthesis for Intention Electromagnetic Interference Injection in Smart Speaker Devices," *2021 IEEE International Joint EMC/SI/PI and EMC Europe Symposium*, 2021, pp. 673-677.

[11] Nirupam Roy, Haitham Hassanieh, and Romit Roy Choudhury. "BackDoor: Making Microphones Hear Inaudible Sounds", *15th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '17)*.

[12] "Op Amp Power Supply Rejection Ratio (PSRR) and Supply Voltages", Analog Devices, from https://www.analog.com/media/en/training-seminars/tutorials/mt-043.pdf