DeHiREC: Detecting Hidden Voice Recorders via ADC Electromagnetic Radiation

Ruochen Zhou[1], Xiaoyu Ji[1†], Chen Yan[1], Yi-Chao Chen[23], Wenyuan Xu[1], Chaohao Li[1]
[1]Ubiquitous System Security Lab (USSLAB), Zhejiang University
[2]Shanghai Jiao Tong University
[3]Microsoft Research Asia
{zrccc, xji, yanchen, wyxu, lchao}@zju.edu.cn, yichao@sjtu.edu.cn

***Abstract*—Unauthorized covert voice recording brings a remarkable threat to privacy-sensitive scenarios, such as confidential meetings and private conversations. Due to the miniaturization and disguise characteristics, hidden voice recorders are difficult to be noticed in their surroundings. In this paper, we present `DeHiREC`, the first proof-of-concept system that can detect offline hidden voice recorders from their electromagnetic radiations (EMR). We first characterize the unique patterns of the emanated EMR signals and then locate the EMR source, i.e., the analog-to-digital converter (ADC) module embedded in the mixed signal system-on-chips (MSoCs). Since these unintentional EMR signals can be extremely noisy and weak, accurately detecting them can be challenging. To address this challenge, we first design an EMR Catalyzing method to stimulate the EMR signals actively and then employ an adaptive-folding algorithm to improve the signal-to-noise ratio (SNR) of the sensed EMRs. Once the sensed EMR variation corresponds to our active stimulation, we can determine that there exists a hidden voice recorder. We evaluate the performance of `DeHiREC` on 13 commercial voice recorders under various impacts, including interference from other devices. Experimental results reveal that `DeHiREC` is effective in detecting all 13 voice recorders and achieves an overall success rate of 92.17% and a recall rate of 86.14% at a distance of 0.2 m.**

***Index Terms*—hidden voice recorder, electromagnetic radiation (EMR), analog-to-digital converter (ADC).**

I. Introduction

Unauthorized voice recording has become one of the greatest threats to commercial secrets and personal privacy [1]. It is estimated that the leakage of trade secrets can cause more than billions of dollars of loss every year [2]. With the proliferation of micro-electromechanical system (MEMS) microphones, voice recorders are getting smaller in size, and therefore also become harder for human to notice especially when they are hidden, e.g., in a pocket or underneath a document. Some voice recorders are even designed in disguise of regular pens or USB sticks [3]. Therefore, in secured conference rooms such as Sensitive Compartmented Information Facility (SCIF), where smartphones and laptops are generally prohibited or strictly inspected, detecting the presence of hidden voice recorders is essential for protecting confidential meetings and private conversations.

†Corresponding author

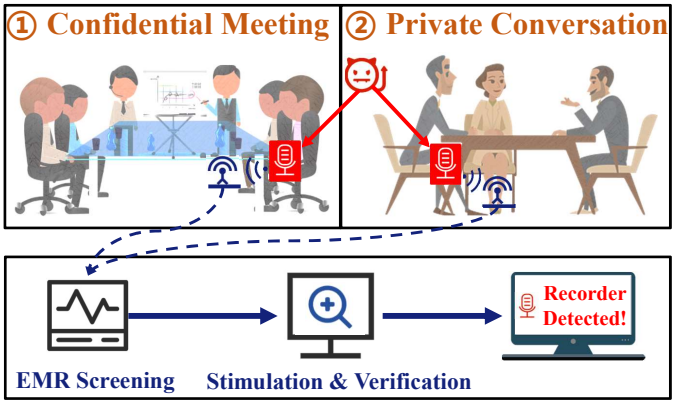


Fig. 1. Illustration of `DeHiREC`. An attacker attempts to secretly carry a hidden voice recorder and then surreptitiously record a confidential meeting or a private conversation. `DeHiREC` detects the recorder by screening and verifying the unintentional EMR signals leaked from the voice recorder.

However, detecting a hidden voice recorder is non-trivial because voice recording is essentially a *passive* process. Unlike active sensors such as radars and night-vision cameras, microphones by design do not actively emit any signals to the environment; they only passively measure the vibrations induced by sounds and convert them to electrical signals for sampling and storage. In addition, most voice recorders work offline, meaning that they do not initiate any wireless connections. They are also battery-powered and do not leak any traces in the powerline. All of the above evidence suggests that existing methods on detecting hidden cameras, e.g., by the infrared radiation [4], wireless traffic [5]–[8], or power traces [9], cannot apply to hidden voice recorders.

This paper aims to detect hidden voice recorders despite of these challenges. Our key insight is that although voice recording is passive, all electronic devices will inevitably emit electromagnetic radiations (EMR) to the environment due to the variation of currents. We believe voice recorders are no exception. Therefore, we may be able to detect a functioning hidden voice recorder if we can receive and identify its unique EMR traces over the air. To achieve this goal, we need to answer the following research questions.

- `RQ 1`: what are the characteristics of the EMR traces leaked from a voice recorder?

- RQ 2: how to verify that the EMR is indeed generated by a voice recorder when there can be interference from other devices in a similar spectrum?
- RQ 3: how to effectively measure an extremely weak EMR from a voice recorder that is typically low-powered, especially when it is covered or at a distance?

We first conduct a feasibility analysis to investigate RQ 1&2. Our spectrum analysis observes that the EMR peaks of voice recorders appear around the devices' clock frequencies with an equal interval, which we assume is a result of signal coupling inside the circuits. Additional hardware analysis suggests that the analog-to-digital converter (ADC) module integrated in the mixed signal system-on-chip (MSoC) is a primary source of such radiations, and the equal interval on the spectrum presents the ADC clock frequency. Nonetheless, signals in a similar frequency band may interfere with the detection, especially those radiated from devices that adopted a similar type of MSoC such as loudspeakers. To increase the confidence, we seek for potential EMR patterns that are *unique* to voice recorders. As the strength of EMR is closely related to the current amplitude, our key idea is to actively change the current flowing through the ADC and simultaneously measure the EMR variation for correlation. In light of prior work illustrating that ultrasound [10] and electromagnetic interference (EMI) [11] can inject signals into microphones inaudibly, we perform proof-of-concept experiments and verify the idea's feasibility with EMI injection. We name such an active stimulation method as *EMR Catalyzing*, which follows the overarching "probe-respond-detect" design principle with [12]–[15] but is the first method that can actively change the EMR strength radiated from the ADCs and derive the unique feature of the variation to identify hidden voice recorders.

Based on the feasibility analysis, we propose DeHiREC, the first proof-of-concept system that can detect hidden voice recorders from their electromagnetic radiations. In particular, to increase the signal-to-noise ratio (SNR) for weak EMRs (raised in RQ 3), we propose an enhanced signal processing algorithm, called adaptive-folding, which folds the spectrum with a phase alignment method to accumulate the EMR peaks and boost detection. Using this method, we are able to increase the detection distance from a few centimeters to a maximum of 0.5 m with our lab-grade equipment. Our evaluation in a meeting scenario on 13 consumer voice recorders of different makes and shapes under the interference of other 21 electronic devices shows that DeHiREC can effectively detect all of the voice recorders when they are hidden with an overall success rate of 92.17% and a recall rate of 86.14% at a distance of 0.2 m. As the initial work on detecting hidden voice recorders, we believe the proof-of-concept detection method can also provide new insights and inspirations to the detection of other hidden devices in different scenarios.

Our Scope. DeHiREC is designed to catch attendees that sneak hidden voice recorders into a meeting hosted in a secured conference room, where smartphones and laptops are generally prohibited or strictly inspected. As the location of people is easy to predict in the conference room, we envision that the host can use DeHiREC to protect confidential meetings and private conversations with a proper setup of the system beforehand, as illustrated in Fig. 1.

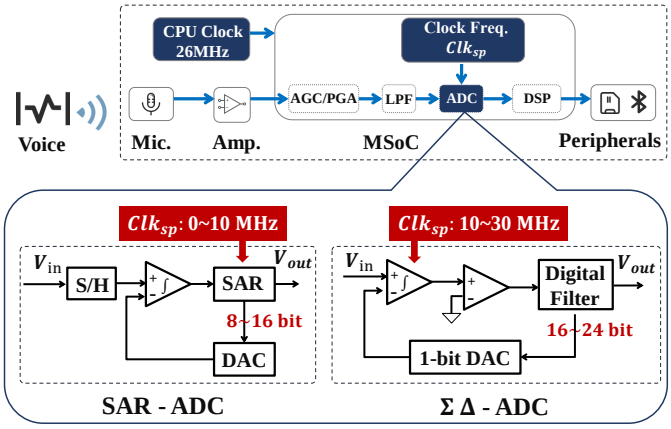


Fig. 2. The workflow of a typical voice recorder. Notably, there are two common types of ADCs, i.e., SAR-ADC and ΣΔ-ADC, and they vary in the clock frequency.

In summary, this paper makes the following contributions.

- We discover the shared EMR patterns resulting from the ADCs embedded in the MSoCs and show that they can be used for the detection of hidden voice recorders.
- We propose EMR Catalyzing, an active stimulation method that can trigger an ADC's EMR to vary in reaction of EMI. We show that it can be used to uniquely identify voice recorders and increase the confidence.
- We design DeHiREC, the first proof-of-concept system enabled with our enhanced adaptive-folding algorithm that can detect hidden voice recorders based on weak EMR signals. Our evaluation on 13 voice recorders and 21 interfering devices demonstrates its effectiveness.

II. Background

In this section, we describe the architecture and workflow of a typical voice recorder and two common types of ADCs adopted in voice recorders. Additionally, we introduce some background on electromagnetic radiation of MSoC.

A. Voice Recorder

A voice recorder is an electronic device that converts analog voice into a digital signal. Fig. 2 presents the typical architecture and workflow of a voice recorder. First, the microphone records external voice through sound vibration and converts it into an electrical signal. The amplitude of such electrical signals will increase as the voice intensity grows, generally reaching a maximum of 250 mV [16]. Second, an external amplifier is employed to increase the recorded analog signal. Third, the system on a chip (SoC) utilizes a programmable gain amplifier (PGA) to adaptively adjust the amplitude of the input signal via automatic gain control (AGC) and leverages a low-pass filter (LPF) to remove the high-frequency noise. Finally, the built-in ADC of the SoC will convert the analog

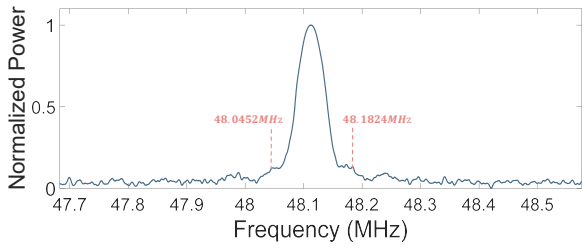


Fig. 3. The illustration of the energy distribution of a 48 MHz spread clock signal.

signal into a digital signal, which is further sent to a digital signal processor (DSP).

There are two common types of ADCs used in the voice recorder, i.e., successive-approximation ADC (SAR-ADC) and Sigma-Delta ADC ($\Sigma\Delta$-ADC). Fig. 2 presents their block diagrams. A common SAR-ADC generally consists of four sub-circuits: a sample-and-hold (S-H) circuit, an analog voltage compactor, an internal reference DAC, and a successive approximation register. When a new conversion starts, the input signal is sampled by the S-H circuit and then compared multiple times to the reference voltage. Finally, SAR-DAC stores the results and outputs the digital signals until a whole conversion finishes. Compared with SAR-ADC, $\Sigma\Delta$-ADC concentrates the high-frequency noise in the analog signals by oversampling and noise shaping, and then removes it through a digital filter. Thus, $\Sigma\Delta$-ADC can achieve a higher resolution than SAR-ADC. Specifically, the resolution of $\Sigma\Delta$-ADC can be up to 24-bits while the one of SAR-ADC generally ranges from 8-bits to 16-bits. Notably, the number of ADC bits determines the recorder's resolution and affects its clock frequency. Generally speaking, the clock frequency required for SAR-ADC is below 10 MHz, while that for $\Sigma\Delta$-ADC is in the range of 10-30 MHz.

B. EMR of MSoC

Maxwell's equation along with Lorentz force law describes how electric and magnetic fields are generated by charges, currents, and changes in the fields [17]. They also demonstrate how fluctuating electric and magnetic fields propagate in space, which is known as electromagnetic radiation. Since an MSoC is composed of digital, analog, and power circuits, the time-varying current flow inside the MSoC will always cause an EMR [18]. Existing techniques and regulations on electromagnetic interference (EMI) and electromagnetic compatibility (EMC) have made much effort to reduce the unintentional EM emanations leaked from MSoC. However, it is still inevitable that MSoC will produce EMR at the clock frequency when there are fluctuating currents.

Therefore, when we use a near-field probe close to any exposed MSoC that is running, we can get a result similar to Fig. 3, which is the EMR signal power spectrum of a 48 MHz clock MSoC. The peak shape in Fig. 3 indicates that the peaks will appear in a range when we take a long-time FFT window, around 1 second. The simplest form of a clock is a square wave of which the energy is all concentrated at the fundamental frequency or harmonics of the clock frequency. However, this leads to a high EMR intensity that may violate the regulatory requirements for EMC. To walk around this issue, modern clock generators use spread-spectrum techniques to reshape the energy distribution of the clock [19]. In other words, the oscillator on the chip will generate a clock output that varies within a certain range, generally -8% to +8% of the calibration frequency. Therefore, when we capture the RF leakage generated by clock signals, we will find that the power distribution changes among different samplings.

III. Threat Model

The attacker's goal is to surreptitiously record the contents of a confidential meeting or private conversation with a hidden voice recorder described in Fig. 1. As a detector, `DeHiREC` tries to detect the above hidden voice recorders among complex environments during the meeting.

A. Attacker Model

We make the following assumptions about the attackers.

Type of Voice Recorder. Considering that a confidential meeting may deploy signal blocking devices to disable the network connection, we assume that the malicious attacker prefers to use voice recorders without wireless transmitters, e.g., Wi-Fi or Bluetooth. Nevertheless, the attacker can use the recorders with wireless transmitters and `DeHiREC` can detect it. Moreover, we have no restrictions on the appearance and the manufacturer of the voice recorder. For example, the attacker may choose a pen-like voice recorder due to its high stealthiness.

Location of Voice Recorder. In the conference room, we assume that the voice recorder is placed in proximity to the attackers while the location of the attackers is relatively fixed. As such, the attacker may put the hidden voice recorder in the pocket, under the desk, on the chair, and so on. Thus, it is convenient for the attacker to have full control of the voice recorder, such as choosing whether and when to start or stop recording.

In addition to the type and location of voice recorder, we assume that the attacker has no prior knowledge of our detection system, i.e., he doesn't know whether the detection system exists and where it is.

B. Capability of the Detection System

Detection Equipment. The RF capturing device can capture the signal whose frequency band matches with that of the EMR signal of the voice recorders. A typical detection system of `DeHiREC` shall consist of the following devices: 1) *broadband antenna* that can emit and capture the RF signal, 2) *low noise amplifier (LNA)* that can amplify weak signals, 3) *software defined radio (SDR)* that can down-convert and digitize the signal, and 4) *PCs* that can analyze the characteristics of EMR spectrum and run detection algorithms remotely.

Location of `DeHiREC`. The `DeHiREC` system can be placed anywhere in the meeting room to be protected. To achieve a better detection performance, we can deploy multiple

detection devices near each seat in the confidential meeting room in advance. Moreover, these detection devices can be embedded into tables, chairs, or other decorations to improve their concealment.

Environment. In a scene of a confidential meeting or privacy conversation, there will inevitably be various other electronic devices, especially some legal or authorized recording devices, which will also generate EMR and interfere with the detection. However, as a detector, the system should know the positions of the above-mentioned devices and the characteristics of their EMRs to mitigate such interference.

IV. Principle of DeHiREC

To clarify why recorders can be detected by the radiated EMR, we first conduct a preliminary experiment to investigate the characteristics of the EMR signals from voice recorders. Then we locate these EMR signals and investigate their causes in order to devise a validation method to uniquely determine the EMR signal from recorders. Finally, we design an algorithm to boost the strength of the weak signals from a low-powered voice recorder in order to achieve a longer detection distance.

A. Characterizing the EMR Signals from Voice Recorders

Since every electric device generates more or less EMR, we focus on the particularity of the recording process to avoid interference from other devices during detection.

1) Experimental Setup: To show the generalizability, we select 13 voice recorders on a top online shopping website based on the ranking of sales volume. For those with similar sales volume, we try to diversify the selection by choosing recorders with different appearances or shell materials such as plastic, glass, and metal. Table II presents the model, MSoC type, ADC type, and shell material of 13 voice recorders. While the voice recorder is recording offline, i.e., not transmitting data through a wireless channel, we use a near-field probe to sense the EMR signals leaked from the voice recorder. To mitigate signal attenuation, we take off the metal shell of the voice recorders during the feasibility test. The distance between the probe and the target voice recorder is set to 3 cm. We search for the EMR components over a wideband that ranges from 20 MHz to 1 GHz. Meanwhile, the near-field probe is connected to an LNA that can amplify the weak signals with a gain of 35 dB. Next, the amplified signals are down-converted and digitized by SDR. Finally, we utilize a laptop to analyze the spectrum of received EMR signals and perform signal processing algorithms.

2) Experimental Results: The characteristics of captured EMR signals are provided in Table II, and Fig. 4 shows an example of the signal power spectrums of Sogou C1 and iFLY B1 for the follow-up analysis. We derive three observations from the preliminary test.

First, **all voice recorders have EMRs in the frequency band of about 20-150 MHz.** In addition, the characteristics of the signals become more stable with the times of capturing increase, as can be seen from the comparison between Figs. 4(a) and 4(d), which represents a single sample, with Figs. 4(b) and 4(e), which represents an average of 100 samples.

Second, **the leaked signals appear as equally spaced peaks on the power spectrum, i.e., present a stable periodicity.** The peaks in the power spectrum of EMR signals from different recorders can be various. Take Sogou C1 and iFLY B1 for example, the average peak intervals are 3.2264 MHz for the former and 24.0729 MHz for the latter according to Fig. 4(b) and Fig. 4(e).

Third, **the EMR signals are strongly related to the modes of the voice recorder.** Fig. 4(b) and Fig. 4(c) present the power spectrum of Sogou C1 under two modes, i.e., recording and standby. The average peak interval is around 26.0013 MHz when the voice recorder is on standby while the one is about 3.2264 MHz when the voice recorder starts recording. The same phenomenon also occurs on the iFLY B1 according to Fig. 4(e) and Fig. 4(f). Thus, we infer that the recording process introduces extra EMR signals.

Through the above preliminary experiment, we characterize the features of the EMR signals emanated from voice recorders, which can be further utilized to perform hidden voice recorder detection.

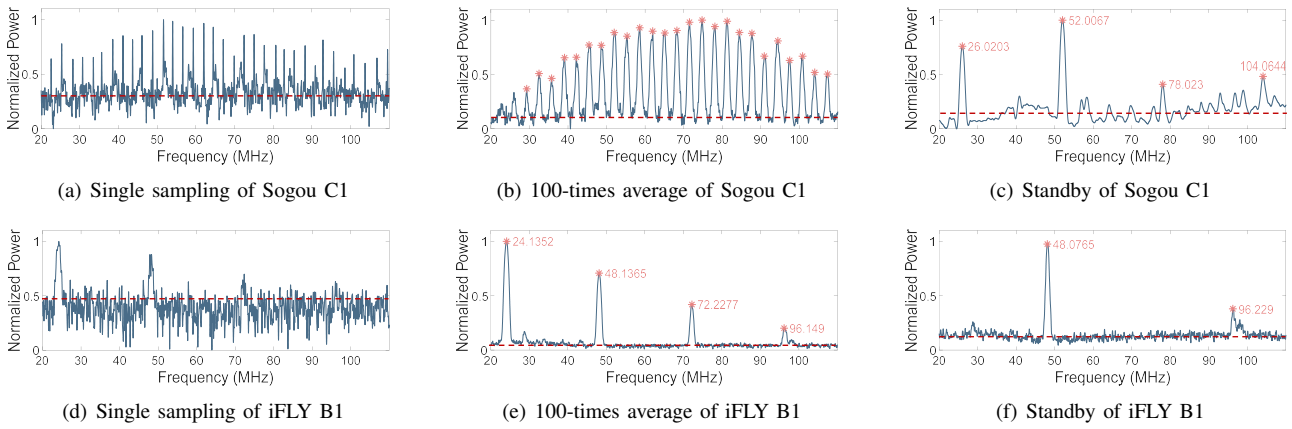
B. Why Do Voice Recorders Emit EMRs: ADC Induced

Here we dig out the EMR sources inside the voice recorder and explain the reason why the EMR signals vary under different modes. Comparing Fig. 4(b) with Fig. 4(c), as well as Fig. 4(e) and Fig. 4(f), we find that the EMR signals are similar to the modulation of two components, which we call carrier and baseband respectively.

Carrier. In Fig. 4(c), the average peak interval of Sogou C1 on standby mode is 26.0013 MHz, which is the same frequency as the system clock of MT2523S according to its datasheet [20], i.e., the MSoC used in Sogou C1. Meanwhile, the shape of these peaks matches the spectrum patterns of the clock that are reshaped by spread spectrum techniques. Thus, we infer that the carrier of the EMR signals is derived from the system clock of the MSoC, and it will exist whenever the recorder is on. iFLY B1 also has a 48 MHz clock carrier with the same characteristics according to Fig. 4(f).

Baseband. Fig. 4(b) reveals that the power spectrum of the EMR under the recording mode not only contains all peaks of the carrier but also brings extra peaks with an average peak interval of 3.2264 MHz. This number corresponds to the clock frequency of the ADC module used in the MT2523S [20]. Similarly, 24.0729 MHz in Fig. 4(e) is also the ADC's clock frequency of iFLY B1. The measured peak intervals of the other 11 voice recorders in Table II also match with their ADC clocks as well. As a result, we infer that the baseband of the EMR signals is derived from the ADC module.

In sum, we make the hypothesis that ***the EMR signal of a voice recorder is mainly derived from the system clock's EMR (i.e., the carrier) onto which the ADC's EMR (i.e., the baseband) is modulated***. In other words, the EMR signals of the system clock and the ADC within the same chip will



(a) Single sampling of Sogou C1 (b) 100-times average of Sogou C1 (c) Standby of Sogou C1

(d) Single sampling of iFLY B1 (e) 100-times average of iFLY B1 (f) Standby of iFLY B1

Fig. 4. The power spectrum of the EMR emanated from two voice recorders, Sogou C1 and iFLY B1. The first column is the results of single sampling while the second column represents 100-times average samples under recording. The third column is the EMRs when two recorders are on standby but not recording. We use the red horizontal line to distinguish the signal component from the noise. Results reveal that the peak interval of different voice recorders can be various, yet present a stable periodicity. Also, the power spectrum of the EMRs emanated from recorders is different under two modes, i.e., standby and recording.

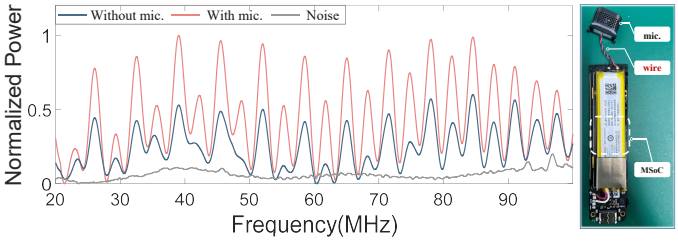


Fig. 5. The power spectrum of the EMR signals emanated from Sogou C1 with or without a microphone under recording mode. An illustration of the wire we disconnected is on the right.

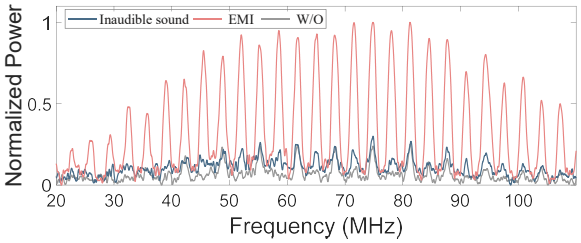


Fig. 6. The power spectrum of the EMRs under two stimulations. Results show that electromagnetic interference can significantly increase the intensity of EMR, yet inaudible sound injection presents poor performance.

be coupled and emanate to the outside. The frequency of this EMR signal can be formalized as the following.

$$f_{EMR} = k \cdot f_{clk} \pm l \cdot f_{adc} \qquad k \in \mathbb{N}_+,\ l \in \mathbb{N} \qquad (1)$$

where f_{clk} and f_{adc} denote the frequency of system clock and ADC clock, respectively.

Besides the numerical results of the above preliminary experiments matching our hypothesis, we also conduct a separation experiment to further validate our hypothesis. Specifically, we manually disassemble a voice recorder (Sogou C1) and cut off the wire connecting the microphone and the MSoC. Then we compare the EMR signals of the voice recorder with or without the microphone. Fig. 5 depicts the resulting power spectrums, showing that the strength of the EMR without a microphone is much lower than the one with a microphone. This is because the power consumption of the ADC module is related to the input of the microphone, which can be represented as an AC voltage with a DC bias.

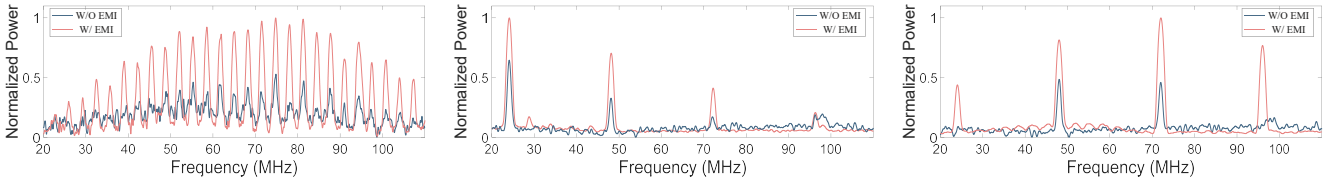
Above all, we successfully locate and validate the EMR sources inside the voice recorders, i.e., the system clock and the ADC module. However, it still leaves us a challenge that whenever we find a suspicious signal with similar characteristics, how can we validate whether it is from a hidden voice recorder?

C. Identifying the EMR from a Voice Recorder

During a real-world detection, signals coming from other devices will inevitably interfere with our judgment on the EMR of the voice recorder, especially the devices that also run the ADC module. Therefore, we need to separate or verify the component belonging to the EMR of voice recorders from a power spectrum. Recall from the finding that the strength of the emanated EMR signals is highly related to the fluctuating current that comes from the microphone. As such, the problem of signal identification can be transformed to find a method to stimulate the input current of the ADC module actively, which will then make the strength of EMR changes correspondingly.

Notably, our detection system cannot make physical contact with the hidden voice recorder directly as well as alert the malicious attacker. Thus, our proposed method shall be non-contact and covert. Motivated by previous work (i.e., DolphinAttack [10] and GhostTalk [11]), we propose two candidate active stimulation methods.

- **Increase the microphone's output**: Intuitively, the amplitude of the microphone's output will affect the input current of the ADC module. Thus, we can leverage inaudible sound injection [10] to increase the output of the microphone covertly.
- **Increase the transmission current**: GhostTalk [11] has validated that EMI can affect circuits by inducing volt-



(a) Newsmy XD01 (Metal) (b) iFLY H1 (Glass) (c) Lenovo C2 (Plastic)

Fig. 7. Comparison of the power spectrum of the EMR signals generated by three voice recorders before and after the catalyzing, the shell material of the recorder is indicated in brackets.

ages on wires. Thus, we can increase the input current of the ADC module via electromagnetic interference directly.

We follow the same device setups as the above two works, and Fig. 6 presents the performance of the two methods. Results reveal that electromagnetic interference (red line) can significantly increase the strength of EMR signals, yet inaudible sound injection (blue line) presents poor performance. This is because the increment introduced by the inaudible sound injection is limited by the maximum output of the microphone, while the EMI bypasses the microphone and couples directly into the wire.

EMR Catalyzing. To answer RQ2, we design EMR catalyzing, which emits an EMI bait signal to actively change the EMR strength of the recorders. Intuitively, the larger the intensity of EMR variation, the easier it is to identify the signal component emits from a hidden voice recorder in a spectrum. Moreover, the induced voltage caused by the EMI reaches the maximum when the frequency of the injected EMI signal matches the resonant frequency of the receiving circuit. Thus, the key to improving the intensity of EMR signals is to find the most effective EMI frequency.

We notice that the built-in microphone of a voice recorder is typically separated from the PCB mainboard, and they are connected with a wire. Meanwhile, the length of this wire can be similar (i.e., about 3 cm) among various kinds of voice recorders due to the size limitation. This observation makes it possible to find an available EMI frequency that can be applied to different voice recorders. According to Maxwell's equations, the frequency of the available EMI signal heavily relies on the electrical length of the receiving antenna, which can be formulated as $f = c/20l$, where c is the light speed and l is the length of the antenna [21].

In our case, the length of the antenna (i.e., the wire) is typically 3 cm. Thus, the frequency of our generated EMI shall be higher than 500 MHz. To search for the optimal frequency, we modulate a 1 kHz sine signal on a high-frequency carrier which ranges from 400 MHz to 1.5 GHz. Although the optimal coupling frequencies differ among each recorder, we can observe stable and various increments of the EMR produced by 13 voice recorders in Table II when the frequency of the EMI signal is set to 980.00 MHz. Fig. 7 gives an illustration of the signal gain brought by the EMR catalyzing method for three recorders, which represent three shell materials respectively.

In sum, once we find a suspicious signal with similar characteristics to the voice recorder, we will inject EMI signals to perform a secondary confirmation. If the intensity of the EMR signals increases after the active stimulation, we determine it as a voice recorder.

D. Augmenting the Strength of the Weak EMR Signals

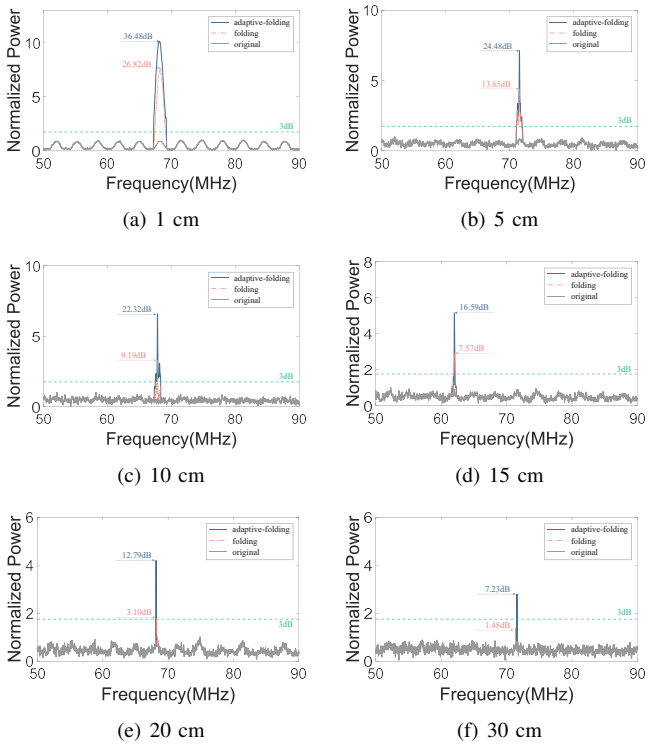
The above preliminary experiments have demonstrated the feasibility of exploiting EMR signals leaked from the MSoC for hidden voice recorder detection. However, since the original EMRs from voice recorders are generally weak, we try to improve the SNR of the received signals algorithmically.

Thus, we utilize the folding algorithm, which is generally used for amplifying periodic astronomical signals. For a known spectrum, the objective of folding algorithm is to search for a signal with a period of T, then divide the spectrum according to the time window T, and fold up all the small segments. When T is equal to the period of the signal, the energy of the signal will become stronger after folding, while the noise will be balanced out due to its randomness. In our scenario, when the window T is equal to the ADC clock frequency of the recorder, the SNR of EMR after folding will be greatly enhanced.

However, We also notice that in Fig. 4 although the peaks are distributed at intervals, the precise interval value of each adjacent peak is not fixed. Such characteristics may be due to the spread spectrum techniques of the clock, or the noise interference. In any case, we need to optimize the folding algorithm to better improve the SNR.

Adaptive-folding Algorithm. For this purpose, we design a new optimization algorithm, called adaptive-folding. The core of this algorithm is to align the phase of the peaks in each segment before folding. Specifically, we find the peaks in each segment and shift the power spectrum according to the deviation, so that the peaks can be located in the same position.

Fig. 8 shows an example of the normalized spectrums containing EMR signals of Sogou C1 which are measured from different distances. The red line represents the signal component after folding and the blue line represents that after adaptive-folding. The calculated SNRs indicate that adaptive-folding has a better effect compared to folding. In Fig. 8(f), we can hardly find the peaks of the original signal, the black line, by visual search at a distance of 30 cm. However, we can achieve a 7.23 dB gain after adaptive-folding, which is enough to be distinguished from changes in noise.



(a) 1 cm (b) 5 cm

(c) 10 cm (d) 15 cm

(e) 20 cm (f) 30 cm

Fig. 8. Effects of detecting evenly spaced peaks using folding (red) or adaptive folding (blue) algorithm. The grey line represents the original EM signal, and the threshold about 3 dB for criterion is marked by cyan line. The adaptive-folding algorithm can effectively improve SNR.

Besides, we also compare the performance of our adaptive-folding algorithm with other weak signal detection algorithms, which are folding, fast fourier transform (FFT), and variational mode decomposition (VMD). We use the original spectrum in Fig. 8 as input to evaluate the effect of the above algorithms. Table I shows the comparison between the four algorithms, and we can conclude that the effects of the four algorithms are all well when the original signal component is obvious. With the attenuation of the signal, the robustness of our adaptive-folding algorithm is revealed.

The reason why the folding algorithm has always been inferior to the adaptive-folding algorithm is due to some signal loss since the peak interval is not a fixed value in the spectrum. Meanwhile, the FFT and VMD algorithms perform poorly at a distance. Briefly, FFT fails to identify the peak interval due to its poor resolution. VMD algorithm is able to decompose the spectrum but shows an insufficient ability to detect periodic signals with severe attenuation.

Overall, by using the adaptive-folding algorithm, we are able to improve the SNR of the EMR signal, thereby increasing the detection distance.

V. DeHiREC

Based on the **EMR catalyzing** and **adaptive-folding algorithm**, we propose our detection system DeHiREC and provide its detailed design as follows. Fig. 9 presents the overview of DeHiREC, which can be mainly divided into two stages. The role of ***Stage 1*** is to determine whether there exists a suspected EMR signal with the characteristics of a voice recorder. Once there is, the system moves on to ***Stage 2***, where stimulation and verification are performed to actively distinguish whether the signal is from a voice recorder.

TABLE I
THE COMPARISON BETWEEN ADAPTIVE FOLDING, FOLDING, FFT AND VMD ALGORITHMS UNDER VARIOUS DISTANCE.

Distance	SNR			
	Adaptive-folding	Folding	FFT	VMD
1 cm	36.48 dB	26.82 dB	22.63 dB	29.14 dB
5 cm	24.48 dB	13.65 dB	8.44 dB	9.58 dB
10 cm	22.32 dB	9.19 dB	3.65 dB	4.11 dB
15 cm	16.59 dB	7.57 dB	2.82 dB	2.69 dB
20 cm	12.79 dB	3.10 dB	1.42 dB	2.03 dB
30 cm	7.23 dB	1.48 dB	0.00 dB	0.21 dB

A. *Stage 1: Preliminary Screening*

Through capturing and processing the RF signals in the surrounding environment, DeHiREC firstly analyzes whether there are suspected spectrum characteristics caused by a voice recorder. We define one screening here as the average value of continuous 100 samples of collected EMR signal, thus ensuring that the influence of noise can be reduced.

Candidate Peak Search. Once the system completes a round of screening and captures the RF signal data, it immediately starts to search for potential peaks. The interval between every two peaks is then calculated to determine if there is a dominant frequency. Recall from Section IV, we focus on the clock frequency of the ADC module in the voice recorder, which is basically between 2 MHz and 30 MHz.

Adaptive Segmentation. After a dominant frequency is obtained, the system will perform spectrum segmentation and then adaptive-folding. After that, we compare the noise with the folded peak to calculate the SNR, denoted as ξ_{signal}.

Frequency Feature Extraction. When the inequality in Eq. (2) is satisfied, we consider it to be a suspected signal and mark the frequency of peaks and intervals f_{peak} automatically to facilitate tracking.

$$Ave(\sum_{1}^{1000} \xi_{signal}) \geq \eta \tag{2}$$

After measuring 100 samples of the noise in a laboratory environment, we calculate the difference between the maximum and average value of noise power is approximately 2.6 dB. Considering a certain margin, we position the threshold η at 3 dB. We will also evaluate the impact of η on the detection accuracy in Section VI.

ADC Database Matching. To determine whether the f_{peak} might come from a recorder's EMR, we match it against the ADC database. This database is collected by ourselves, i.e., the ADC clock frequency of each MSoC in Table II. When this matching holds, the system will consider this to be a suspected EMR signal and enter the second stage.

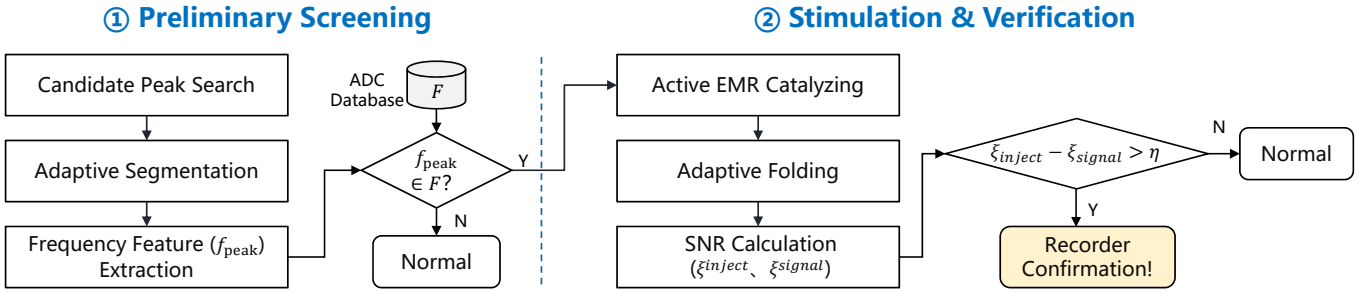


Fig. 9. Overview of DeHiREC. (1) Preliminary Screening: find the suspected spectrum by peak searching, adaptive segmentation and feature extraction. (2) Stimulation & Verification: comfirm the signal by EMR catalyzing, adaptive-folding, SNR calculation and comparison.

B. *Stage 2: Stimulation & Verification*

Once there is a suspicious EMR signal, we start to conduct EMR catalyzing to verify. The key for DeHiREC is to observe whether the tailored bait signal can cause changes in the power spectrum.

EMR Catalyzing. During frequency sweeping in Section IV, we observed that the increments of all recorders will dynamically change, and the frequencies of the highest increment are all distributed in a band of 900-1050 MHz. Therefore, in order to ensure the EMI has a better coupling effect for each recorder, we consider emitting an EMI signal that modulates a 1 kHz sine signal on a multi-frequency carrier composed of 900, 940, 980, and 1020MHz for EMR catalyzing.

DeHiREC collects the EMR signal while the EMI is emitting, define one stimulated sample here as the average value of continuous 100 samples of collected EMR signal.

Adaptive Folding & SNR Calculation. To quantify the change in EMR strength, we define the ξ_{inject} represents the SNR under EMR catalyzing. After adaptive-folding the spectrum with the marked intervals f_{peak}, the ξ_{inject} are generated.

$$Ave \sum_{1}^{1000} (\xi_{inject} - \xi_{signal}) \geq \eta \tag{3}$$

Verification. Due to the presence of noise and other interference, we apply the threshold η in the first stage again to determine the amount of change produced in the stimulation. Finally, if Eq. (3) is satisfied, we determine that there is a synchronous change on the spectrum and a hidden voice recorder is located nearby.

VI. Evaluation

In this section, we describe our experimental setup and evaluate the performance of DeHiREC. We start by examining the SNR and TPR under various distances when there are no other devices and evaluating the impact of different factors. We then consider the case when there are interference devices trying to confuse the system. Finally, we evaluate the detection accuracy in a real-world scene.

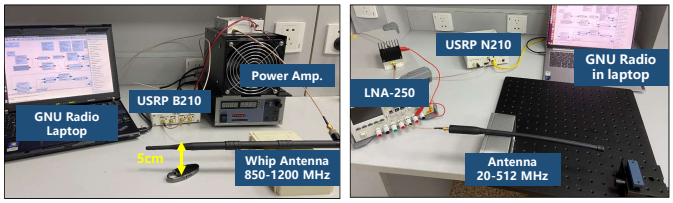


Fig. 10. The experimental setup for DeHiREC is composed of two categories. (1) **EMR receiver (right)**. We adopt a whip antenna to capture the weak EMR signals, connecting to a low noise amplifier to filter noise and amplify the signal. The EMR signal is then transmitted to the laptop by USRP and then adopt signal analysis. (2) **EMI transmitter (left)**. We modulate the baseline signal and carrier by USRP and amplify the EMI signal by the power amplifier. The generated EMI signal is emitted by a whip antenna.

Fig. 11. The appearance (left) and internal PCB circuit (right) of 13 chosen recorders, the label applies to both figures. We use the white boxes to indicate the positions of MSoCs.

A. *Experimental setup*

Equipment. The equipment utilized in our experiment can be divided into two categories: the EMR receiver and the EMI transmitter. For the EMR receiver side, we use the USRP (Universal Software Radio Peripheral) N210 [25] with two kinds of receiving antennas (i.e., a magnetic field probe NFP-3 from 30 MHz to 3 GHz [26] and a portable whip antenna GT512 from 20 MHz to 512 MHz) to sense the EMR signals leaked from the voice recorder. Then an LNA-250 low noise amplifier [27] is utilized to further augment the sensed signals and remove extra noise. For the EMI transmitter side, we adopt a USRP B210 [28] and an MPA-10-40 power amplifier [29] (typically 30 dB gain) to generate the EMI signal and use a portable whip antenna (850 MHz to 1200 MHz) to emit. The emitting EMI signal power we measured at 50 cm away from the antenna is 18 dBm, which is 64 milliwatts. Fig. 10 presents the overview of our experimental setup.

Environment. In order to minimize uncontrollable inter-

TABLE II
DETAILED INFORMATION AND OVERALL PERFORMANCE OF 13 VOICE RECORDERS. THE LAST 5 COLUMNS ARE THE RESULTS FROM OUR EXPERIMENTS.

Record Model	MSoC Type	Shell Material	ADC Type	ADC Clock Frequncy	Peak Interval		Max Distance	d = 10 cm	
				(MHz)	Min(MHz)	Max(MHz)	(cm)	SNR(dB)	TPR
Sogou C1	MT 2523S [20]	Plastic	SAR	3.25	3.2168	3.2587	42	21.67	100%
Sogou C1 pro	RTL 8722CS [22]	Plastic	SAR	5	4.8932	5.1052	30	19.56	98%
Newsmy V03	ATJ 2127 [23]	Metal	ΣΔ	24	23.8913	24.3085	15	7.64	88%
Newsmy XD01	MT 2523S	Metal	SAR	3.25	3.2212	3.2701	16	6.13	86%
Newsmy RV100	WS 200	Metal	ΣΔ	12	11.7852	12.1855	15	7.82	88%
Aigo R6811	ATJ 2127	Metal	ΣΔ	24	23.5140	24.2502	22	8.65	92%
Aigo R8822	ATJ 3315 [24]	Metal	ΣΔ	24	24.0342	24.8463	23	7.64	92%
Lenovo C2	ATJ 3315	Plastic	ΣΔ	24	23.7901	24.5103	34	15.58	98%
Shinco RV-18	JL AC6901	Plastic	SAR	6	5.8161	6.2299	33	14.11	96%
Lenovo B460	AK 2115C	Plastic	ΣΔ	24	23.7193	24.1712	40	13.24	98%
iFLY TEK H1	ATS 2837	Glass	ΣΔ	24	23.8748	24.9912	55	13.27	96%
iFLY TEK B1	ATS 2837	Plastic	ΣΔ	24	23.6139	24.1062	39	22.76	100%
Philips VTR5102	ATS 2837	Glass	ΣΔ	24	23.7035	24.3313	45	19.77	99%

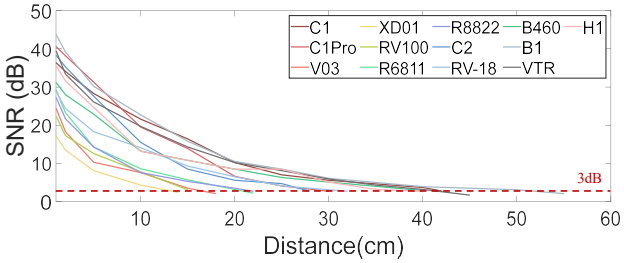


Fig. 12. The effect of the adaptive-folding algorithm. Each line represents the trend of EMR's SNR of a recorder with distance. The distance below the red line (3 dB) is considered undetectable due to noise interference.

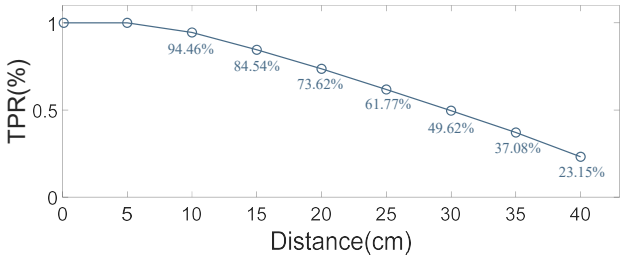


Fig. 13. The relationship between the average TPR and the detection distance, which is derived from 1300 experiments for 13 recorders.

ference, we conduct most of the experiments excluding real-world scenes in an EM shielded and silent room. For those experiments to evaluate performance under various impacts, we will explain the specific setups clearly at the beginning of each subsection.

Evaluation Metrics. We adopt the following metrics throughout the evaluation. As a premise, we consider voice recorders that are recording as the "positives" for our system, and any device else should be a "negative".

- **SNR**: characterizes the ratio of the EMR signal power to the noise power.
- **True Positive Rate (TPR)**: characterizes the probability that the system correctly detects the positive samples.
- **True Negative Rate (TNR)**: characterizes the probability that the system correctly ignores the negative samples.
- **Max Distance**: indicates the maximum distance that the system can detect the recorder with a TPR above 50%.
- **Accuracy & Recall**:

$$Accuracy = \frac{TP + TN}{\sum samples}, \quad Recall = \frac{TP}{TP + FN} \tag{4}$$

The targeted voice recorders are the 13 ones used in Section IV.

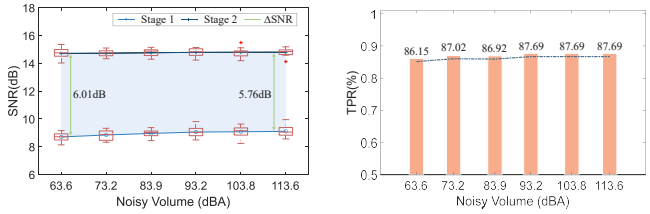
B. Performance of DeHiREC

We conduct 100 times of detection for each voice recorder at each distance that ranges from 1 cm to 50 cm with a step length of 5 cm. When DeHiREC failed to detect at a certain distance, we will reduce the distance in steps of 1 cm to get an accurate maximum distance. During the experiment, a speech video is played with a volume of about 81.5 dBA.

Table II summarizes the resulting peak interval of the sensed EMR signals for each model of voice recorders. Due to the spread spectrum techniques, we give the maximum and minimum peak interval of all samples for each voice recorder. Nevertheless, we find that the average peak interval of the sensed EMR signals always matches the clock frequency of the ADC module.

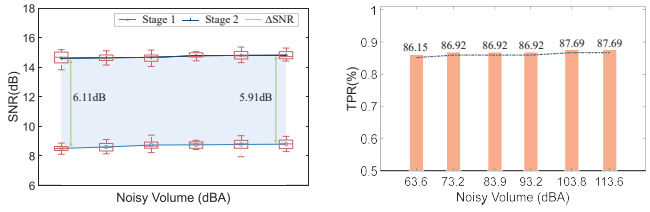
Time Overhead. With an i5-11300H&16G, the time to detect a hidden voice recorder is around 3.8 seconds in total which includes 1.5 seconds of EMR sensing and 0.4 seconds of data processing for each stage in Fig. 9.

1) Detection Distance: To illustrate the detection performance under various distances, Fig. 12 and Fig. 13 show the signal SNR of 13 recorders, and the average TPR at various distances. The maximal detection distance of all 13 voice recorders is more than 15 cm while the largest one can even reach 55 cm. We find that the recorders with metal shells generally have a lower detection distance, which is due to the



(a) (b)

Fig. 14. SNR and TPR of detection under different noise volume in a cafe.



(a) (b)

Fig. 15. SNR and TPR of detection under different noise volume in a meeting room.

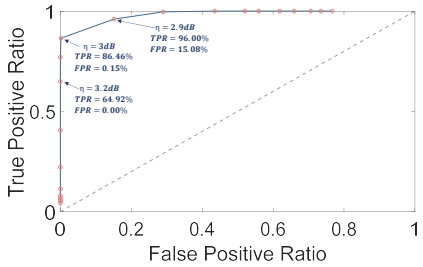


Fig. 16. The ROC curve under various thresholds between 2.0 dB and 4.0 dB with a gradient of 0.1.

TABLE III
DETECTION PERFORMANCE UNDER METAL CASING.

Material	Mic.	EMR		TPR	Audio SNR
		No. of detectable	SNR(dB)		(dB)
None	unwrapped	13	20.76	100%	76.04
faraday pouch	wrapped	0	-	0%	39.17
anti-static bag	wrapped	4	5.15	23.07%	55.24
tin foil	wrapped	0	-	0%	62.37
tin foil	unwrapped	6	8.33	43.85%	75.13

stronger attenuation by the metal than by plastic or glass. To better demonstrate the performance of our system, we show the SNR and TPR for each recorder at a distance of 10 cm in Table II.

The results show that our detection system can effectively detect recorders in the scenarios described in Section III, in which the attacker is supposed to hide the voice recorder near the table, e.g., on the seat or in the pocket during a meeting. Our receiving antenna is mounted underside the table and the distance from every possible location of a recorder is kept less than 20 cm.

2) Noisy Environment: To study the impact of background noise on the detection performance of `DeHiREC`, we simulate two environments (cafe and meeting room) by playing light music and lectures through speakers at six volume levels. The distance between the recorder and the antenna is 10 cm, and each recorder is tested 10 times in each case.

Fig. 14 and Fig. 15 show the SNR and TPR in the cafe and meeting room, respectively. The lighter blue line with dots represents the measured SNR in ***Stage 1***, and the darker blue line represents the measured SNR in ***Stage 2***. The difference between lines (shaded) indicates the EMI response of the recorders, and a larger difference means a stronger response to the EMI. We also present the experimental measurement error in the form of boxplots in the figures. The dashed lines in the right figure represent the TPR at different volume levels. The results show that a noisier background can slightly increase the EMR strength in ***Stage 1***, but it also slightly decreases the recorder's response to EMI in ***Stage 2***. Overall, the system's accuracy is almost unaffected in a volume-noisy environment.

3) Threshold Selection: 3 dB is an empirical value that we obtained through experiments to distinguish between signal and noise. To choose it, we measure the accuracy of the system under different η. We reuse the samples from *Noisy Environment* experiments, and each recorder has a total of 120 samples. We carry out the detection procedures under various thresholds ranging from 2.0 dB to 4.0 dB with a gradient of 0.1, and draw the ROC curve in Fig. 16.

The ROC curve shows that the threshold affects the trade-off between TPR and FPR. We choose 3.0 dB as the threshold because it achieves a low FPR while keeping a reasonable TPR. In practice, the system designer may choose the threshold based on the performance requirement and an adaptive measurement of the deployed noise environment.

4) Metal Casing: A rigorous attacker may wrap recorders with metal materials to shield EMR signals, which may adversely affect the detection. Thus, we evaluate the performance of `DeHiREC` under metal casing, e.g., using commercial shielding bags and handcrafted wrapping with metal foil. Specifically, we choose a faraday pouch and an anti-static bag available online which are dedicated to shielding RF signals [30], and tin foil for handcrafted wrapping. Additionally, we divide the degree of handcrafted wrapping into two types, which are 1) completely wrapped, and 2) wrapped without covering microphones (to make the recording clearer). For each case, we measure each recorder 20 times at a distance of 5 cm in an EM-shielded room and then calculate the SNR and TPR. We play music with a volume of 84.5 dBA throughout the experiment. We also analyze the audio files recorded in each case to evaluate the recording quality. We report the number of detectable recorders, the average SNR of EMRs, the TPR, and the average recording SNRs in each case in Table III.

The results suggest that 1) our system may fail to detect a hidden voice recorder when it is fully wrapped by a faraday bag or metal foil. If the microphone is left uncovered, 6 recorders can be detected at a reduced SNR and TPR. 2) The recording quality will degrade if the recorder is wrapped. Results show that the SNR of the recorded audios degraded by 20 dB when the recorders are completely wrapped. 3) 4

TABLE IV
DETAILED INFORMATION AND OVERALL PERFORMANCE OF OTHER 21 INTERFERENCE ELECTRONIC DEVICES.

Category	Type	Label	Case	MSoC Type	EMR Pattern	EMI response	TNR
Category 1: Authorized Recording Devices	Conference System	A1	Thinkplus MCP01	ATS 2836	3.1521 MHz	Y	75%
		A2	Runpu RP-M10S	N/A	2.3523 MHz	Y	70%
		A3	UGreen 30755	ATS 2831	2.4135 MHz	Y	70%
	Recording Smartphone	B1	iPhone 12	N/A	-	-	100%
		B2	OPPO Reno 3	N/A	-	-	100%
	Recording Laptop	C1	Lenovo yoga C940	N/A	-	-	100%
		C2	ThinkPad X1 Nano	N/A	-	-	100%
	Smart Speaker	D1	Xiaomi Play	BES 2300	24.1178 MHz	N	100%
		D2	XiaoDu AI Speaker 2	ES 2743	3.3795 MHz	N	100%
Category 2: Non-Recording Devices That May Emit Similar EMR	Smartphone	B1	iPhone 12	N/A	-	-	100%
		B2	OPPO Reno 3	N/A	-	-	100%
	Laptop	C1	Lenovo yoga C940	N/A	-	-	100%
		C2	ThinkPad X1 Nano	N/A	-	-	100%
	Earphone	E1	Samsung Buds 2	BES 2500	24.0413 MHz	N	100%
		E2	Xiaomi TWS	BES 2500	24.0150 MHz	N	100%
	Speaker	F1	AOMAIS A27	JL 6928	24.0138 MHz	N	100%
		F2	JBL GO2	N/A	32.6512 MHz	N	100%
	TV	G1	Samsung NUF302	N/A	-	-	100%
		G2	KONKA LED58U5	N/A	-	-	100%
Category 3: Non-Recording Devices Without DACs	Monitor	H1	AOC 27B2H	N/A	-	-	100%
		H2	DELL S2421HSX	N/A	-	-	100%
	Projector	I1	EPSON CB-FH06	N/A	-	-	100%
		I2	BenQ E520	N/A	-	-	100%
	Rounter	J1	Huawei AX3 pro	N/A	-	-	100%
		J2	TP-LINK AC1200	N/A	-	-	100%

recorders can still be detected even if they are inside the anti-static bag, which is an interesting indication that the anti-static bag's shielding capacity is not as effective as advertised. As EMR is by nature an extremely weak signal, the law of physics inherently determines that the signal will be greatly attenuated after shielding. Thus, metal casing will degrade the detection performance of all EMR-based solutions.

C. Interference from Other Devices

Since all electronic devices will radiate EMR, we evaluate the impact of 21 electronic devices which are common in a meeting room, especially the authorized recording devices which may emit EMR with similar patterns and respond to EMI. We first divide these devices into three categories as follows.

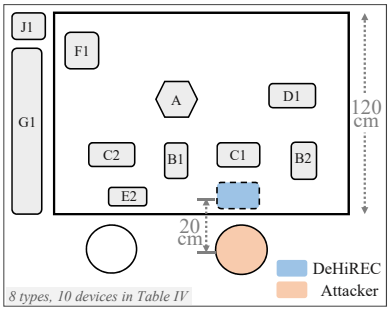
Category 1: Authorized recording devices. The meeting organizer may record the conversation with authorized devices, which are most likely to interfere with our system because the ADCs may emit similar EMR and respond to the EMI probing. We select 9 audio input devices for this category, including smartphones, laptops, and conference systems for recording and smart speakers that are always listening. Note that two selected conference systems (Think plus MCP01 and UGreen 30755) have the same series of MSoC with one of the experimented hidden voice recorders.

Category 2: Non-recording devices that may emit similar EMR. We suspect that non-recording devices with DACs may also emit similar EMRs, because audio ADCs and DACs have similar clock frequencies. Thus, these devices may interfere with the first stage of the system, but they shall be ruled out by the second stage as they do not respond to the EMI probing. We select 10 audio output devices that have audio DACs for this category, including loudspeakers, earphones, laptops, smartphones, and TVs.

Category 3: Non-recording devices without DACs. All devices radiate EMR and may interfere with our system. To narrow down our scope, we select 6 devices that are common in offices or conference rooms, e.g., monitors, projectors, and routers.

We select at least two typical devices in each type and report device's MSoC information if it uses the same MSoC architecture as the voice recorders in Table IV. For each device, We run `DeHiREC` 20 times at 10 cm away in an EM shielded room and play speech audio with a volume of about 80.3 dBA during the experiment. Table IV lists the results, e.g., the EMR pattern (peak intervals), EMI response, and TNR. Through experiment, we have the following observations for each category.

Category 1: the overall TNR for 9 devices is 89.47%. Among them, we find that 1) all 3 conference system devices not only radiate a similar EMR pattern but also respond to our EMI catalyzing, therefore resulting in a TNR of 71.67%. After we disassembled A1 and A3, we find that they adopt the same series of MSoC as that of voice recorders. 2) Although the two smart speakers radiate similar EMR patterns, their EMI responses are too weak to be detected at 10 cm. Further experiments show that the EMI responses become observable (higher than η) after decreasing the distance to 3 cm. 3) The



(a) Floor plan of real-world setup

(b) Snapshot of real-world setup

Case	Time	TP	TN	Accuracy	Recall	TNR
None	130	58	65	94.62%	89.23%	100%
A1	130	56	63	91.54%	86.15%	96.92%
A2	130	56	64	92.31%	86.15%	98.46%
A3	130	55	64	91.54%	84.62%	98.46%
M	130	56	64	92.31%	86.15%	98.46%
A1, M	130	55	63	90.77%	84.62%	96.92%
Overall	780	336	383	92.17%	86.14%	98.20%

(c) Detection performance under 6 cases

Fig. 17. The setup in a real-world scene (a) (b) shows the location of each device, including targeted recorders (peach puff) and interference devices (labeled). An enlarged snapshot of the antenna is presented in (b), RX is for EMR sensing while TX performs catalyzing. We simulate 6 cases, the first column in (c) indicates which devices are working in each case, M denotes all other interference devices excluding A.

smartphones and laptops do not show similar EMR patterns when they are recording.

Category 2: the overall TNR for 10 devices is 100%. The results show that some non-recording devices (E1, E3, F1, F2) do emit similar EMR signals, but they do not respond to the EMI we inject. Note that these devices adopt similar MSoCs that are embedded with both ADC and DAC modules. However, as the ADC module is off, the EMR caused by DAC will not respond to the EMI injected into the ADC's input.

Category 3: the overall TNR for 6 devices is 100%. We did not observe any EMR in the frequency band of 20-150 MHz.

Although the results show that some authorized recording devices do interfere with the accuracy of the detection system, we can avoid false detection with extra knowledge or human intervention. We will continue to explore this issue in real-world scenes.

D. Performance in a Real-world Scene

To investigate the practical impact of authorized recording devices, we conduct further experiments in a meeting room in our university. As a case study, we selected 6 devices (A1, A2, A3, B1, C1, D2) in category 1, 5 devices (B2, C2, E2, F1, G1) in category 2, and 1 device (J1) in category 3 from Table IV, and placed them at the locations where each device would normally appear in a meeting. For example, the conference systems were in the middle of the table, and our detection system was placed at the edge of the table near where participants would sit. Fig. 17 presents a detailed illustration. Since the 3 conference systems (A1, A2, and A3) are more likely to be detected as hidden recording devices, we evaluate DeHiREC in 6 setups as shown in Fig. 17(c) (M denotes all other negative devices excluding A).

We invite the participant (as an attacker) to bring one of the recorders in Table II into the meeting room and sit in the marked chair. We let the participants decide by themselves whether to turn on the recorder for recording or not. The only restriction is to place the recorder on the table or chair (within the detection range). In each experiment, the participant sits in the room for 2 minutes, after which he tells us if he has recorded. During the experiment, we play course lecture videos with a volume of about 86.5 dBA.

We conducted a total of 780 rounds of detection, the overall detection accuracy is 92.17% and the TNR is 98.20%. In comparison, when no interference device is turned on, the accuracy is 94.62% and the TNR is 100%. Results show that as long as the locations of DeHiREC and the authorized electronic devices are reasonably planned, it is possible to suppress the interference and improve the detection accuracy. In case that the interference from the authorized devices is inevitable even with planned device locations, we envision that the defender may collect the EMR patterns of the authorized recording devices in advance and build a whitelist to filter out their interference. Last but not the least, the defender may temporarily turn off the authorized recording devices to detect hidden voice recorders.

VII. Discussion

In this section, we discuss several considerations when deploying DeHiREC in practice.

A. Application Scenario of DeHiREC

DeHiREC can help a meeting host detect hidden voice recorders that attendees sneak into secured conference rooms, e.g., SCIF. In this scenario, since the room is under control of the host, the location of attendees, i.e., where recorders may appear, is easy to predict, and the cost of deploying multiple detection devices near attendees is easy to justify. However, due to the distance limit, DeHiREC may not be as effective in scanning hidden voice recorders inside rooms not controlled by the user, e.g., an Airbnb house, where the recorder can be stashed anywhere.

B. Precise Positioning via Active Scanning

To extend the application scenarios of DeHiREC in the future, we may improve the system to achieve active scanning. Since the EMR signal strength can be measured from the power spectrum, it may be possible to estimate the distance and orientation of the voice recorder according to the strength variation. By walking around the room and observing the signal strength, the user may scan hidden voice recorders. Achieving this idea, however, depends on a portable detecting system with higher resolution, greater detection distance,

and a redesigned detection algorithm, which requires further investigation and is left for future work.

C. Compliance with Radio Regulation

Note that when using DeHiREC, the power of EMI emission has to comply with local radio regulations. An enhanced EMR catalyzing may require increased EMI emission power in Stage 2. However, excessive transmission power may interfere with normal communication in the 900-1020 MHz frequency band, which is allocated for mobile or aeronautical communication [31]. Currently, the maximum power of the LNA we use is 1 W and the actual output power we measured is about 64 mW, while the maximum power specified by the FCC rule is 500 mW in a frequency band of 902-928 MHz [32]. We recommend complying with radio regulations when using EMI-related methods such as EMR catalyzing.

D. Detecting Other Recording Devices

Table IV reveals that devices using a similar type of MSoC, such as conference systems and smart speakers, will radiate similar EMR and respond to EMI stimulation. However, for those devices that have ADC modules (e.g., smartphones and laptops), we cannot detect similar EMR signals. We believe that there are two reasons. First, the better EMC design of these devices will reduce the EMR. Secondly, after investigating the internal circuits of smartphones, we found that the ADC module of smartphones is not integrated with other modules in the same chip as on voice recorders. Such a discrete module design leads to less signal coupling. Thus, smartphones that are recording do not cause false positives.

VIII. Related Work

In the following, we provide a summary of the existing work on EM side-channels as well as eavesdropper detection.

A. EM side-channels

Previous work has revealed that electromagnetic radiations can be exploited as a side-channel to perform information reconstruction or anomaly detection. First, researchers validate the feasibility of reconstructing display data or printed data from the EMRs of cables [33], monitors [34], [35], and printers [36], [37]. Camurati *et al.* [38] utilize the EM leakage of wireless chips to crack the encryption key running on the CPU. Sehatbakhsh *et al.* [39] exfiltrate the keystroke logging by the EMRs of the power management unit. Choi *et al.* [40] use EMRs of switching regulator and Bluetooth low energy MSoC to infer audio information. Sehatbakhshk *et al.* [41] leverage EM side-channel signals during response computation for attestation in the embedded system, i.e., Arduino UNO. In this paper, we utilize the EMRs of the ADC module to detect the presence of hidden voice recorders.

B. Device Detection

Previous work has made much effort to detect wireless hidden devices, such as wireless hidden internet-of-things (IoT) devices [6], [8], [42], [43] and wireless hidden radio frequency (RF) eavesdroppers [44]–[47]. These works heavily rely on the wireless traffic or the RF signals which are emitted, or leaked from the wireless module while the devices are working. For example, Lumos [43] sniffs and collects encrypted wireless packets over the air to detect and identify hidden IoT devices. Ghostbuster [12] leverages the leakage from the RF circuit of the wireless receiver to detect hidden eavesdroppers. Earfisher [13] detects wireless eavesdroppers by sensing memory EMR with a wireless-packet-bait technology. However, the above methods are inapplicable for common voice recorders, which generally save the real-time recorded voice locally and neither need the function of active wireless transmission (offline) nor require a wireless receiver (passive).

To detect offline and passive devices like voice recorders, nonlinear junction detection (NLJD), also known as "illumination", has been applied for counter-surveillance [14], [15] and radar sensing [48], [49]. Illumination is a general method to detect any electronic device by emitting strong EM signals and monitoring the 2nd order harmonics reflected by the nonlinear PN junctions in a device [50]. However, such a method is unable to distinguish voice recorders from other types of electronic devices. In comparison, DeHiREC manages to identify voice recorders under the interference of other electronic devices.

In summary, our work follows the overarching "probe-respond-detect" procedure for detection, which is used in related work such as Earfisher [13] and illumination [15] as well. Nevertheless, we are the first to design the EMR Catalyzing method (the "probe") that can actively change the EMR strength radiated from the ADC (the "respond") and use the unique feature embedded in EMR variation to identify hidden voice recorders (the "detect"). Specifically, our method is different from the related work in all three stages of the procedure.

IX. Conclusion

This paper presents the first attempt to detect hidden voice recorders in the surroundings. We find the shared EMR patterns resulting from the ADCs which can be used for the detection of hidden voice recorders. To uniquely identify voice recorders, we design EMR Catalyzing, an active stimulation method that can trigger an ADC's EMR to vary in reaction to EMI. Thus we design DeHiREC, the first proof-of-concept system that can detect hidden voice recorders. Our evaluation shows that DeHiREC can detect recorders under the interference of 21 electronic devices. Moreover, we envision that our method can be of profound significance for detecting other low-power and wireless hidden devices, and we will further study the potential applications of our work.

Acknowledgment

We thank the anonymous shepherd and reviewers for their valuable comments and suggestions. This work is supported by the National Natural Science Foundation of China (NSFC) Grant 62071428, 61925109, 62222114.

References

[1] D. Drab, "Economic espionage and trade secret theft: Defending against the pickpockets of the new millennium," 2003, https://www.xerox.com/downloads/wpaper/x/xgs_business_insight_economic_espionage.pdf.

[2] K. M. Robertson, D. R. Hannah, and B. A. Lautsch, "The secret to protecting trade secrets: How to create positive secrecy climates in organizations," *Business Horizons*, vol. 58, no. 6, pp. 669–677, 2015.

[3] Spyguy, "Best hidden voice recorders for 2021," 2020, https://www.spyguy.com/a/blog/best-hidden-voice-recorders.

[4] M. Roessler, *How to find hidden cameras*. March, 2002.

[5] Y. Cheng, X. Ji, T. Lu, and W. Xu, "On detecting hidden wireless cameras: A traffic pattern-based approach," *IEEE Transactions on Mobile Computing*, vol. 19, no. 4, pp. 907–921, 2019.

[6] ——, "Dewicam: Detecting hidden wireless cameras via smartphones," in *Proceedings of the 2018 on Asia Conference on Computer and Communications Security*, 2018, pp. 1–13.

[7] Y. He, Q. He, S. Fang, and Y. Liu, "Motioncompass: pinpointing wireless camera via motion-activated traffic," in *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*, 2021, pp. 215–227.

[8] K. Wu and B. Lagesse, "Do you see what i see? detecting hidden streaming cameras through similarity of simultaneous observation," in *Proceedings of 2019 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 2019, pp. 1–10.

[9] M. Jin, R. Jia, Z. Kang, I. C. Konstantakopoulos, and C. J. Spanos, "Presencesense: Zero-training algorithm for individual presence detection based on power monitoring," in *Proceedings of the 1st ACM conference on embedded systems for energy-efficient buildings*, 2014, pp. 1–10.

[10] G. Zhang, C. Yan, X. Ji, T. Zhang, T. Zhang, and W. Xu, "Dolphinattack: Inaudible voice commands," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 2017, pp. 103–117.

[11] D. F. Kune, J. Backes, S. S. Clark, D. Kramer, M. Reynolds, K. Fu, Y. Kim, and W. Xu, "Ghost talk: Mitigating emi signal injection attacks against analog sensors," in *Proceedings of the 2013 IEEE Symposium on Security and Privacy*. IEEE, 2013, pp. 145–159.

[12] A. Chaman, J. Wang, J. Sun, H. Hassanieh, and R. Roy Choudhury, "Ghostbuster: Detecting the presence of hidden eavesdroppers," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, 2018, pp. 337–351.

[13] C. Shen and J. Huang, "Earfisher: Detecting wireless eavesdroppers by stimulating and sensing memory emr," in *Proceedings of the 18th USENIX Symposium on Networked Systems Design and Implementation (NSDI 21)*, 2021, pp. 873–886.

[14] J. Raoult, A. Martorell, L. Chusseau, and C. Carel, "Intermodulation radar for rf receiver detections," in *Proceedings of the 2018 15th European Radar Conference (EuRAD)*. IEEE, 2018, pp. 273–276.

[15] H. Aniktar, D. Baran, E. Karav, E. Akkaya, Y. S. Birecik, and M. Sezgin, "Getting the bugs out: A portable harmonic radar system for electronic countersurveillance applications," *IEEE Microwave Magazine*, vol. 16, no. 10, pp. 40–52, 2015.

[16] J. Lewis, "Understanding microphone sensitivity," 2012, https://www.analog.com/media/en/analog-dialogue/volume-46/number-2/articles/understanding_microphone_sensitivity.pdf.

[17] H. D. Young, R. A. Freedman, T. Sandin, and A. L. Ford, *University physics*. Addison-Wesley Reading, MA, 1996, vol. 9.

[18] A. Fayed and M. Ismail, *Adaptive techniques for mixed signal system on chip*. Springer Science & Business Media, 2006, vol. 872.

[19] P. G. Flikkema, "Spread-spectrum techniques for wireless communication," *IEEE Signal Processing Magazine*, vol. 14, no. 3, pp. 26–36, 1997.

[20] MediaTek, "Mt2523 series datasheet," 2017, https://datasheetspdf.com/pdf-file/1343677/MediaTek/MT2523/1.

[21] J.-M. Redouté and M. Steyaert, *EMC of analog integrated circuits*. Springer Science & Business Media, 2009.

[22] REALTEK, "General description of rtl8722csm," 2019, https://www.realtek.com/en/products/communications-network-ics/item/rtl8722csm.

[23] Actions, "Atj2127 datasheet," 2012, http://www.datasheetcafe.com/atj2127-datasheet-pdf/.

[24] ——, "Atj331x datasheet," 2010, https://datasheetspdf.com/pdf-file/774998/Actions/ATJ331x/1.

[25] Ettus Research, "Usrp n200/n210 networked series," 2017, https://www.ettus.com/wp-content/uploads/2019/01/07495_Ettus_N200-210_DS_Flyer_HR_1.pdf.

[26] SCHWARZBECK, "Vulb 9162 - trilog broadband antenna," 2013, http://schwarzbeck.de/Datenblatt/K9162.pdf.

[27] RF BAY, Inc., "10 – 250mhz low noise amplifier," 2006, https://www.rfbayinc.com/upload/files/lna/lna-250.pdf.

[28] Ettus Research, "Usrp b200/b210 networked series," 2004, https://www.ettus.com/wp-content/uploads/2019/01/b200-b210_spec_sheet.pdf.

[29] RF BAY Inc., "Rf amplifier mpa-10-40," 2021, http://mpl.jp/amplifier/MPA-10-40.pdf.

[30] M. Blaze, "Testing phone-sized faraday bags," 2021, https://www.mattblaze.org/blog/faraday/.

[31] ITU, "Radio regulations," 2020, https://orarilokaljakartapusat.com/wp-content/uploads/2020/09/Radio-Regulations-2020-00013-Vol.I-EA5.pdf.

[32] Radiocommunication Sector of ITU, "Technical and operating parameters and spectrum use for short-range radiocommunication devices," 2021, https://www.itu.int/dms_pub/itu-r/opb/rep/R-REP-SM.2153-8-2021-PDF-E.pdf.

[33] M. G. Kuhn, "Electromagnetic eavesdropping risks of flat-panel displays," in *Proceedings of the International Workshop on Privacy Enhancing Technologies*. Springer, 2004, pp. 88–107.

[34] ——, "Optical time-domain eavesdropping risks of crt displays," in *Proceedings of the 2002 IEEE Symposium on Security and Privacy*. IEEE, 2002, pp. 3–18.

[35] ——, "Compromising emanations: eavesdropping risks of computer displays," Ph.D. dissertation, Citeseer, 2002.

[36] T. Tosaka, K. Taira, Y. Yamanaka, A. Nishikata, and M. Hattori, "Feasibility study for reconstruction of information from near field observations of the magnetic field of laser printer," in *Proceedings of the 2006 17th International Zurich Symposium on Electromagnetic Compatibility*. IEEE, 2006, pp. 630–633.

[37] C. Ulaş, U. Aşık, and C. Karadeniz, "Analysis and reconstruction of laser printer information leakages in the media of electromagnetic radiation, power, and signal lines," *Computers & Security*, vol. 58, pp. 250–267, 2016.

[38] G. Camurati, S. Poeplau, M. Muench, T. Hayes, and A. Francillon, "Screaming channels: When electromagnetic side channels meet radio transceivers," in *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, 2018, pp. 163–177.

[39] N. Sehatbakhsh, B. B. Yilmaz, A. Zajic, and M. Prvulovic, "A new side-channel vulnerability on modern computers by exploiting electromagnetic emanations from the power management unit," in *Proceedings of the 2020 IEEE International Symposium on High Performance Computer Architecture (HPCA)*. IEEE, 2020, pp. 123–138.

[40] J. Choi, H.-Y. Yang, and D.-H. Cho, "Tempest comeback: A realistic audio eavesdropping threat on mixed-signal socs," in *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, 2020, pp. 1085–1101.

[41] N. Sehatbakhsh, A. Nazari, H. Khan, A. Zajic, and M. Prvulovic, "Emma: Hardware/software attestation framework for embedded systems using electromagnetic signals," in *Proceedings of the 52nd Annual IEEE/ACM International Symposium on Microarchitecture*, 2019, pp. 983–995.

[42] A. D. Singh, L. Garcia, J. Noor, and M. Srivastava, "I always feel like somebody's sensing me! a framework to detect, identify, and localize clandestine wireless sensors," in *Proceedings of the 30th USENIX Security Symposium (USENIX Security 21)*, 2021, pp. 1829–1846.

[43] R. A. Sharma, E. Soltanaghaei, A. Rowe, and V. Sekar, "Lumos: Identifying and localizing diverse hidden iot devices in an unfamiliar environment," in *Proceedings of the 31st USENIX Security Symposium (USENIX Security 22)*, 2022.

[44] S. Park, L. E. Larson, and L. B. Milstein, "An rf receiver detection technique for cognitive radio coexistence," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 57, no. 8, pp. 652–656, 2010.

[45] D. Sathyamoorthy, M. J. M. Jelas, and S. Shafii, "Wireless spy devices: A review of technologies and detection methods," *Editorial Board*, p. 130, 2014.

[46] A. Mukherjee and A. L. Swindlehurst, "Detecting passive eavesdroppers in the mimo wiretap channel," in *Proceedings of the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 2809–2812.

[47] S. Park, L. E. Larson, and L. B. Milstein, "Hidden mobile terminal device discovery in a uwb environment," in *Proceedings of the 2006 IEEE International Conference on Ultra-Wideband*. IEEE, 2006, pp. 417–421.
[48] Z. Peng and C. Li, "Intermodulation fmcw (im-fmcw) radar for nonlinear wearable targets detection," in *Proceedings of the 2018 United States National Committee of URSI National Radio Science Meeting (USNC-URSI NRSM)*. IEEE, 2018, pp. 1–2.
[49] G. J. Mazzaro, "Nonlinear junction detection vs. electronics: System design and improved linearity," in *Proceedings of the 2020 IEEE International Radar Conference (RADAR)*. IEEE, 2020, pp. 654–658.
[50] Granite Island Group, "Non linear junction detector review and tutorial," 2002, http://www.tscm.com/tmdenljd.html.

X. Appendix

A. EMR Characteristics

We present the captured EMRs of all 13 voice recorders in Fig. 18, including the after-processed power spectrum with peaks found by the adaptive-folding, and the calculated intervals (i.e., ADC clock frequency). The signals are captured at a distance of 3 cm to ensure enough recognizability.

B. Performance under Other Impacts

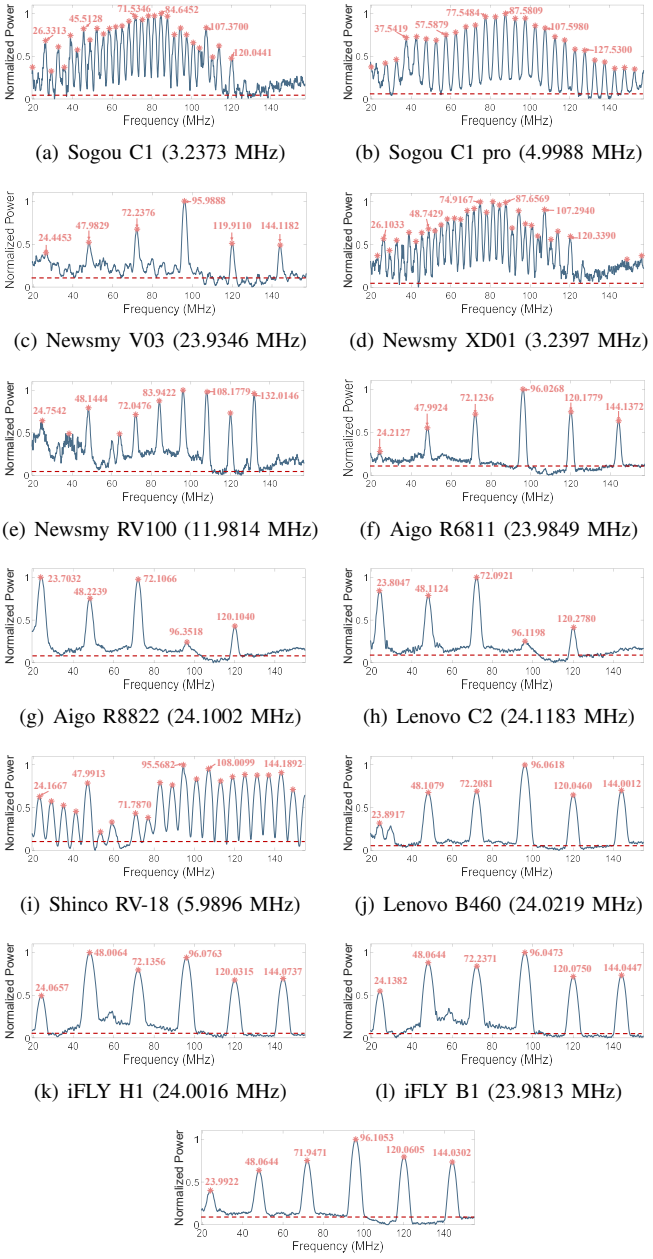
Here we evaluate the performance of DeHiREC under another two factors, including battery level and recorder's location.

1) Battery Level: Modern electronic devices will automatically adjust their performance based on the state of the battery. We believe that voice recorders have this function as well, which will result in variations in the strength of the EMR signal due to different power consumption. Therefore, we measure the SNR of the EMR signal at 10 cm away 20 times for each voice recorder when the battery of the recorder is 100%, 80%, 60%, 40%, and 20% respectively. However, there are some voice recorders that do not have the function of actively displaying the battery level, so we only evaluated 6 of them which have screens. Fig. 19 shows the correlation curve between battery levels and the corresponding SNR.

The results show that different battery levels cause changes in EMR strength, while the degree of influence varies among voice recorders. Nevertheless, we find that the SNR remains above 6 dB when the voice recorders are in a low battery state, indicating that these EM leakage signals can still be observed by our system from a distance.

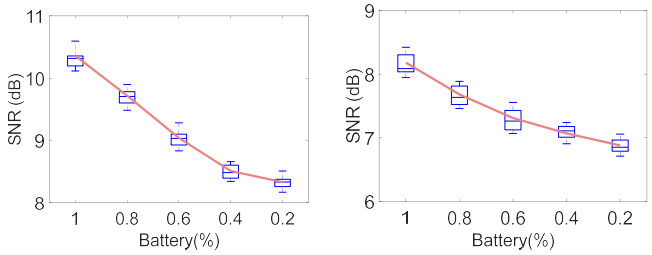
2) The Impact of Voice Recorder's Location: To evaluate the performance of DeHiREC in relation to the location of the voice recorder, we specifically design the following four scenes where the voice recorder is (a) placed directly on the table (i.e., baseline), (b) covered with a paper, (c) held in hand, or (d) put inside the pocket. For each scene, we tested DeHiREC 20 times for 13 voice recorders at a distance of 10 cm. Table V presents the TPR and SNR degradation of the voice recorders under four scenes.

The results show that, despite the fact that the presence of human body significantly reduces SNR, our system is applicable to four scenes with reasonable performance.



(a) Sogou C1 (3.2373 MHz) (b) Sogou C1 pro (4.9988 MHz)
(c) Newsmy V03 (23.9346 MHz) (d) Newsmy XD01 (3.2397 MHz)
(e) Newsmy RV100 (11.9814 MHz) (f) Aigo R6811 (23.9849 MHz)
(g) Aigo R8822 (24.1002 MHz) (h) Lenovo C2 (24.1183 MHz)
(i) Shinco RV-18 (5.9896 MHz) (j) Lenovo B460 (24.0219 MHz)
(k) iFLY H1 (24.0016 MHz) (l) iFLY B1 (23.9813 MHz)
(m) Philips VTR5102 (24.0076 MHz)

Fig. 18. The captured EMR signals of all 13 voice recorders and the peaks found by the adaptive-folding algorithm are marked in the power spectrum. We calculate the average peak interval for each recorder and present it in the sub-caption.



(a) Newsmy V03 (b) Newsmy XD01

(c) iFLY TEK H1 (d) Lenovo B460

(e) Aigo R6811 (f) Sogou C1 pro

Fig. 19. The impact of battery level. Results reveal that the SNR grows with the battery level increases.

TABLE V
TPR AND SNR DEGRADATION UNDER FOUR SCENES, WHICH ARE: (A) ON THE TABLE, (B) COVER WITH A PAPER, (C) HELD IN HAND, (D) INSIDE THE POCKET.

Model	Scene (a)	Scene (b)	Scene (c)	Scene (d)
Sogou C1	100%; 0dB	100%; -1.53dB	85%; -6.13dB	75%; -14.94dB
Sogou C1 pro	100%; 0dB	100%; -1.34dB	75%; -5.35dB	70%; -12.13dB
Newsmy V03	100%; 0dB	90%; -1.32dB	65%; -3.36dB	55%; -4.03dB
Newsmy XD01	100%; 0dB	85%; -1.63dB	65%; -3.29dB	55%; -4.17dB
Newsmy RV100	100%; 0dB	85%; -1.02dB	75%; -3.27dB	60%; -4.94dB
Aigo R6811	100%; 0dB	95%; -0.57dB	75%; -2.35dB	60%; -4.13dB
Aigo R8822	100%; 0dB	90%; -1.43dB	70%; -2.66dB	60%; -4.54dB
Lenovo C2	100%; 0dB	100%; -2.15dB	80%; -4.41dB	70%; -8.63dB
Shinco RV-18	100%; 0dB	100%; -2.38dB	75%; -6.62dB	65%; -9.23dB
Lenovo B460	100%; 0dB	100%; -1.27dB	80%; -5.14dB	70%; -8.71dB
iFLY TEK H1	100%; 0dB	100%; -1.20dB	85%; -4.27dB	70%; -9.49dB
iFLY TEK B1	100%; 0dB	100%; -2.58dB	85%; -5.55dB	75%; -10.73dB
Philips VTR5102	100%; 0dB	100%; -2.36dB	80%; -6.25dB	65%; -9.87dB