Revealing Hidden IoT Devices through Passive Detection, Fingerprinting, and Localization

Wei Sun
UC San Diego
redsunwit@gmail.com

Hadi Givehchian
UC San Diego
hgivehch@ucsd.edu

Dinesh Bharadia
UC San Diego
dineshb@ucsd.edu

ABSTRACT

Internet-of-things (IoT) devices (e.g., micro camera and microphone) are usually small form factor, low-cost, and low-power, which makes them easy to conceal and deploy in the indoor environment to spy on people for human private information such as location and indoor activities. As a result, these IoT devices introduce a great privacy and ethical threat. Therefore, it is important to reveal these concealed IoT devices in the indoor environment for human privacy protection.

This paper presents *RFScan* [1], a system that can passively detect, fingerprint, and localize diverse concealed IoT devices in the indoor environment by sensing their unintentional electromagnetic emanations. However, sensing these emanations is challenging due to the weak emanation strength and the interference from the ambient wireless communication signals. To this end, we boost the emanation strength through the non-coherent averaging based on the emanation signal's characteristics and design a novel suppression algorithm to mitigate interference from the wireless communication signals. We further profile emanations across frequency and time that act as the emanation source's unique signature and customize a deep neural network architecture to fingerprint the emanation sources. Furthermore, we can localize the emanation source with an angle-of-arrival (AoA) based triangulation approach. Our experimental results demonstrate the efficiency of the IoT devices' detection, fingerprinting, and localization across different indoor environments.

KEYWORDS

Concealed IoT Devices, Detection, Localization, Fingerprinting, Privacy Enhancement, Wireless Sensing

1 INTRODUCTION

Affordable and compact IoT devices [2, 6, 8, 11, 18, 19] are increasingly portable, accessible, and easily concealed, making them ideal tools for covert monitoring. Malicious actors can exploit these devices to deploy hidden video cameras, audio recorders, or even radio frequency (RF) receivers to observe private activities and conversations. For instance, a hidden camera could be placed in a victim's home or office to discreetly watch their daily activities [25], or a small audio recorder could be concealed in a pocket to intercept private conversations [49]. Therefore, it is essential to develop and deploy sensing technologies and detection methods [50] to identify and counteract these covert IoT threats.

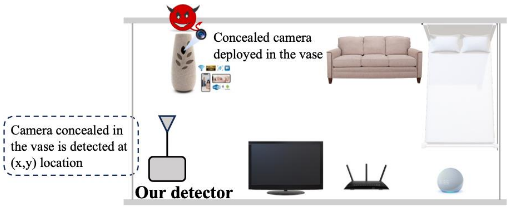


Figure 1: An attacker can stealthily use an IoT device concealed in the vase to spy on people in an indoor environment. Our detector is able to detect, fingerprint, and localize the concealed IoT device.

Ideally, such a detector must detect a variety of devices, localize the hidden device, and perform well in the presence of other electronic devices in the environment. To the best of our knowledge, none of the existing approaches satisfy all these requirements. Commercial off-the-shelf (COTS) detectors [38, 40] require the user to turn surrounding electronic or wireless communication devices off. The dedicated sensor-based detectors (e.g., LAPD [43]) which rely on the light indicator of the camera, only detect hidden cameras and further, do not work in non-line-of-sight scenarios. The traffic-based detectors [12, 24, 48] assume that IoT devices have wireless transceivers for data streaming, while most IoT devices only quietly record and save data to local memory (e.g., SD card) to reduce power consumption and maintain a small form factor [2, 11]. Recent work [34, 57] takes advantage of the unintentional electromagnetic (EM) emanations from the IoT device as most IoT devices leak such emanations. However, these EM emanations are usually weak and not detectable in the presence of other electronic devices. Therefore, prior work [34, 57] employs the excitation-probe framework to magnify and detect the emanated signals. However, such active sensing approaches can be easily detected and defended against by the attacker, they are hardware-dependent (e.g., Memscope [46]), and further confuse the detector by stimulating other legitimate electronic devices.

In this paper, we propose *RFScan*, a system that can passively detect, fingerprint, and localize concealed IoT devices by sniffing their emanations as shown in Fig. 1. Our approach builds on the idea of using EM emanations to detect hidden devices. However, it stands out from previous work as it offers passive detection, fingerprinting, and localization capabilities for diverse IoT devices. Unlike existing methods, our approach doesn't rely on network access or external excitation to stimulate the IoT devices and is independent of the hardware architecture of the diverse IoT devices. Therefore, *RFScan* satisfies the aforementioned requirements, making it a versatile and

[1] *RFScan* represents Radio Frequency-based Scanning to sense the concealed IoT devices. The GitHub repo for RFScan is https://github.com/ucsdwcsng/RFScan_emanations.git

Proceedings on Privacy Enhancing Technologies YYYY(X), 1–14

https://doi.org/XXXXXXX.XXXXXXX

	No network access	**No external stimulation**	**Localization ability**	**Diverse devices**
COTS detectors, e.g., [38]	✓	✓	✗	✗
Sensor-based detection, e.g., [10]	✓	✓	✗	✗
Traffic-based detection, e.g., [24]	✗	✗	✓	✓
Active sensing, e.g., [45, 57]	✓	✗	✗	✗
Memscope [46]	✓	✓	✗	✗
***RFScan* (our approach)**	✓	✓	✓	✓

Table 1: Comparing the existing approaches vs. *RFScan*

effective detector. A comparison of *RFScan* with previous detectors is presented in Table 1.

However, there are several challenges that *RFScan* has to overcome to passively detect the hidden IoT devices. First, the emanations received by radio have a weak signal strength as they are amplitude-modulated clock signals, which can quickly dissipate when transmitted over the air. To this end, we employ a non-coherent averaging technique to enhance the strength of the emanation signals. Second, the emanations of a single device are often dispersed across a wide frequency range as multiple periodic and extremely narrowband signals or tones. Therefore, we employ a frequency hopping technique to scan a wide range of frequencies comprehensively and use a median filter to estimate and smooth the noise floor in each band as the noise floor varies across bands so that we can detect emanation spikes based on their power spectral densities with greater accuracy and reliability. Further, we develop an algorithm that uses a weighted pair-wised distance between the emanation spikes to detect the dispersed tones of a single emanation. The third challenge arises from the presence of ambient emanations generated by legitimate IoT devices, such as Amazon Echo Dot and WiFi access points that are deployed by homeowners for smart homes, as well as wireless communication signals like cellular signals. *RFScan* must be able to distinguish the emanation of covert IoT devices from such background signals in the same frequency range. To address this challenge, we profile the background signals, categorize them as "baseline", and distinguish the signals introduced by the hidden IoT devices during the "test" stage by eliminating the baseline ambient emanations and wireless communication signals.

Once the emanations from hidden IoT devices are detected, *RFScan* further characterizes the devices. First, we must know if the observed emanations belong to one or more devices. To this end, *RFScan* profiles the emanations across the frequency and time that can uniquely characterize the concealed IoT devices. We further customize an attention-based deep neural network that takes the emanation profile as input to fingerprint the concealed IoT devices. To further refine our detection capabilities, we equip our detector with a directional antenna. This antenna enables us to scan the wireless environment, providing us with the angle of arrival (AoA) of the emanation source. By triangulating these AoAs, we can precisely pinpoint the location of the emanation source and remove the hidden IoT devices from the indoor environment.

We summarize our contributions as follows:

- To the best of our knowledge, this is the first work of passively detecting, fingerprinting, and localizing concealed IoT devices based on their unintentional electromagnetic emanations in the indoor environment.
- We introduce a non-coherent averaging technique to enhance emanation signals while effectively suppressing wireless communication signals and ambient emanations through a two-step subtraction process.
- We customize an attention-based deep neural network using emanation profile across the frequency and time to uniquely fingerprint IoT devices and further determine their locations through AoA-based triangulation.
- Our experimental results show an average detection accuracy of around 0.95, an average fingerprinting accuracy of around 0.96, and a localization error of about 2.62 feet with a maximum detection range of 9.8 feet across different IoT devices in different indoor environments.

2 THREAT MODEL

Attack scenario. The attacker wants to spy on private activities or conversations of the subjects of interest. As shown in Fig. 1, the attacker either conceals the IoT device in a specific location such as the victim's house or office, or carries the device in their pockets or bags during a confidential or private conversation. In either case, we assume there is a baseline state in which the IoT device is not in the environment, for example, before the attacker places their device in the house or before the attacker shows up in the meeting. **Attacker's capability.** The attacker may use different types of IoT devices with different hardware architectures. We consider three types of the most popular devices, namely, acoustic-based devices, vision-based devices, and radio frequency-based devices. The IoT device may also have different properties or capabilities. For example, it may use wireless communications to leak out the information or use internal storage (e.g., SD card) to record the data, or it might be capable of detecting excitation signals activated by the victim's detector.

To this end, the detector should be able to detect the concealed IoT devices passively by scanning the radio frequency signals, regardless of what device the attacker uses. We can simply deploy the detector in the indoor environment to continuously detect, fingerprint, and localize the newly added concealed IoT devices. As such, we can kick out these concealed IoT devices from the indoor environment for privacy protection.

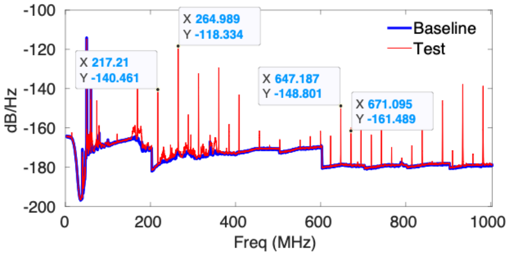


Figure 2: Power spectral density in the anechoic chamber room without spy camera deployment (i.e., baseline) and with the spy camera deployment (i.e., test).

3 PRIMER ON EMANATIONS

Physical principle. Electromagnetic waves are usually unintentionally leaked from IoT devices either through the electronic components on the IoT devices or through the transceiver's antennas of the IoT devices. These leaked electromagnetic waves are termed emanations. The physical principle of IoT devices' emanations is that the amplitude-modulated clock signals have leaked from these IoT devices due to the computation activities performed on these IoT devices (e.g., CPU or memory access). Since every IoT device has a clock for synchronization purposes during computation, the clock signals are going to be leaked whenever the IoT devices are powered up. Specifically, these clock signals can be modulated by the computation activities (e.g., CPU or memory access) of the IoT devices, resulting in the amplitude-modulated squared wave signals that are emanations.

The emanations exhibit the squared waves in the time domain and periodical spikes (i.e., harmonics) in the frequency domain. The periodicity of the emanation spikes indicates the clock frequency of the emanation source. Let's formally model the IoT device's emanations. The ideal squared wave using Fourier expansion with a cycle frequency of f over time t can be represented as follows:

$$x(t) = \frac{4}{\pi}\sum_{k=1}^{\infty}\frac{sin(2\pi(2k-1)ft)}{2k-1}.$$

Its frequency-domain representation is shown in the following.

$$x(f) = \sum_{k=-\infty}^{\infty}\frac{2sin(2\pi k f_0 T)}{k}\delta(f-kf_0),$$

where $f_0 = \frac{1}{T}$ is the frequency of the fundamental harmonic, and $\delta(f-kf_0)$ indicates the harmonic component at frequency of kf_0 with amplitude of $\frac{2sin(2\pi k f_0 T)}{k}$. As we can see, the squared wave consists of an infinite number of sine wave components. Moreover, since the emanations are amplitude-modulated, they can spread over the spectrum. So, each sine wave component acts as a different carrier for the modulation signals. However, in reality, suppose the frequency of the modulated signals due to computation activities is f_l, which can be further mixed with the clock signals with the frequency of f_c. If the IoT device has a transceiver with the carrier frequency of $f_{carrier}$, these amplitude-modulated signals can be converted to other frequency bands. As a result, the frequency of emanation signals received over the air can be represented as follows.

$$f_r = p \cdot f_{carrier} + q \cdot f_c + r \cdot f_l,$$

where p, q, and r are integers due to the mixing and amplification of the emanations within the IoT devices. Note that not all of these multiples are applied, which is highly dependent on the hardware architecture and components (e.g., filters) of the circuit. Some IoT devices may not even have RF transceivers, thereby the emanations are emitted through the data lines (i.e., acting as antennas) on the IoT device.

Showcasing. To demonstrate the squared waves of emanations from the IoT devices, we only put a spy camera (i.e., OV5640 [11]) in the anechoic chamber room where the RF signals are isolated. We use a signal hound [47] as a spectrum analyzer that connects the receive antenna [7] inside the anechoic chamber room. Fig. 2 shows the power spectral density (PSD) of the received signals within the frequency band of 3MHz-1GHz. As shown in the blue line, when there is no spy camera deployed in the anechoic chamber room (i.e., baseline), the PSD is quite clean without any outstanding spikes. However, there are periodical spikes in PSD that can indicate the emanations from the spy camera as shown in the red line (i.e., test). Note that the distances between two adjacent emanation spikes in PSD indicate the clock frequencies of the spy camera, which are 24MHz and 48MHz.

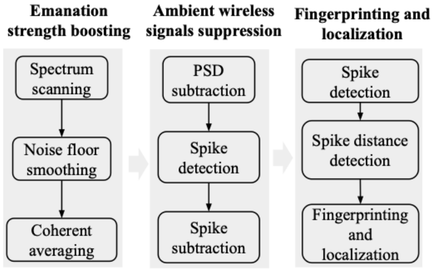


Figure 3: Workflow of *RFScan* consists of emanation strength boosting module, wireless communication signal suppression module, and emanation source fingerprinting and localization module.

4 OVERVIEW

Fig. 3 shows the workflow of *RFScan*, consisting of the emanation strength boosting module, ambient wireless communication signals suppression module, and emanation source fingerprinting and localization module. For example, in a typical indoor environment (e.g., a home), our detector detects the new IoT device introduced to the environment and localizes it. Then, the homeowner could find out if this new IoT device is unrecognizable which should be removed from the environment. Then, we can fingerprint this new IoT device. We illustrate *RFScan*'s workflow in detail as follows.

Boosting emanation strength. Our detector obtains the power spectral density (PSD) of the received signals through spectrum scanning, which consists of the emanations and ambient wireless communication signals. We first need to smooth the noise floor of the derived power spectral density of the received signals for accurate emanation detection. Then, we boost the emanation strength through non-coherent averaging.

Suppressing wireless signals. To remove the ambient wireless communication signals (e.g., WiFi or cellular signals) and the emanations from legitimate IoT devices, we first apply the subtraction to the power spectral density of the received signals from the baseline and test. Then, we detect the emanation spikes on the subtracted PSD. To further eliminate the artifacts introduced by the wireless communication signals, we propose the subtraction between the spikes detected in the PSD from the baseline and test.

IoT device fingerprinting and localization. *(1) Fingerprinting IoT.* We first profile the emanations across the frequency and time, which can be used as the input of the customized deep neural network for IoT device fingerprinting. As such, we can remove the IoT devices that are unrecognizable. Specifically, our customized deep neural network is trained on the common IoT devices that we can purchase on the market. Therefore, we can have one extra class with a small output probability to indicate the new IoT device that has not been seen by the well-trained deep neural network. As such, the softmax threshold is chosen based on the smallest softmax output probability from the training set. During the fingerprinting, if the new device is introduced to the environment, our deep neural network flags it as the new class due to the small output probability. Then, we can add this new device's emanation profiles to the current training dataset to retrain our neural network model. After the IoT devices are fingerprinted, we can further localize them *(2) Localizing IoT.* After we detect the new IoT device introduced to the environment, we can further localize this new IoT device by pinpointing the emanation source with the triangulation of the emanation source's angle of arrivals (AoAs). Specifically, we can use a directional antenna instrumented on the detector to derive the AoA information for localization. After we know the exact location of this new IoT device, we can further check if this new IoT device is malicious or not. We can look for this new IoT device using AoA information and even localize it.

5 SYSTEM DESIGN

5.1 Boosting the emanation strength

Noise floor smoothing. The power spectral density (PSD) of the received wireless signals across the frequency bands has different noise floor levels, which can affect our emanation spikes detection. As we discussed in Section 3, the emanations exhibit periodical spikes in the frequency domain, which we need to extract for emanation detection. Therefore, we first need to smooth the noise floor across the frequency bands to mitigate the emanation spike detection error with the move median filter. Basically, we use a window sliding over the PSD. Then, we calculate the median value within the sliding window, which can be subtracted by the PSD to smooth the noise floor across the frequency bands. To demonstrate the effectiveness of our noise floor smoothing with the move median approach, we scan the wireless spectrum from 3MHz to 1GHz in a typical office room. This is because the emanations from the hidden IoT devices are usually within this frequency band. Fig. 4 shows the power spectral density of the spectrum within a 3MHz-1GHz frequency band in the office room. As we can see, the noise floor across the frequency bands is quite varying. After we use our move median approach for noise floor smoothing, as shown in Fig. 5, the noise floor across the frequency bands becomes flat and smooth, which is beneficial to accurate emanation spikes detection.

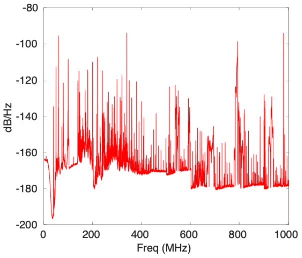


Figure 4: PSD of the received signals without noise floor smoothing.

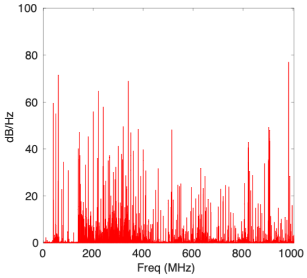


Figure 5: PSD of the received signals with noise floor smoothing.

Non-coherent averaging. Since the IoT device's emanations are amplitude-modulated clock signals, it spreads across a wide spectrum band. It is impossible to scan the GHz spectrum band [22]. Therefore, we can scan a MHz spectrum band with frequency hopping for GHz spectrum band sensing. However, this frequency-hopping approach may introduce different levels of noise floor across multiple frequency hops, which can confuse our emanation spikes detection. This is because the emanations are amplitude-modulated clock signals, which can spread across a wide frequency band. To further boost the emanation strength, we can do non-coherent averaging of the received wireless signals over multiple time sweeps. This is because emanations exhibit the same pattern (i.e., amplitude-modulated clock signals) across the time sweeps, while the noise can be averaged out.

5.2 Suppressing Ambient Wireless Signals

Subtraction on PSD. To eliminate the ambient artifacts of wireless communication signals, we first collect the received wireless signals in the ambient wireless environment, which can be regarded as the baseline dataset. As time goes on, new IoT devices may be introduced to the environment. We continuously collect the received wireless signals, which can be regarded as the test dataset. Then, we can obtain the power spectral density of the baseline and test datasets, which is averaging over time to strengthen the emanations. Since the goal of our emanation detection is to extract the intermediate frequency of these emanations in PSD, we need to suppress the power spectral density of the artifacts. Therefore, the subtraction of PSD from the baseline and test dataset can remove the ambient wireless communication signals and emanations in the environment.

Subtraction on detected spikes. To further eliminate the effect from the ambient artifacts, we first detect the spikes in the PSD from the baseline dataset. The frequencies of those detected spikes are supposed to indicate the ambient emanations and wireless communication signals. We use F_{bs} to represent the frequencies of those detected spikes. Furthermore, we detect the spikes in the subtracted PSD, where the frequencies of the detected spikes mainly represent the emanations from the IoT devices. We use F_{ps} to represent the frequencies of the detected spikes in the subtracted PSD. Then, we can do the subtraction between the frequencies of the detected

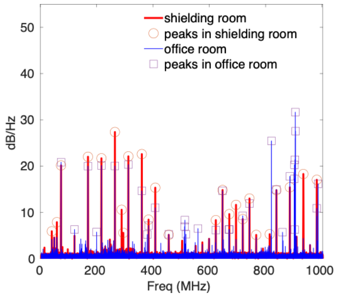


Figure 6: Detected emanation spikes in the office room and anechoic chamber room after suppressing the wireless communication signals.

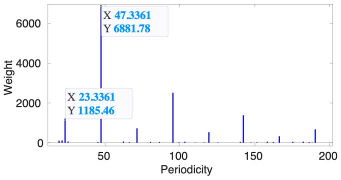


Figure 7: Periodicity detection of the spy camera's emanations.

spikes in baseline and subtracted PSD to mitigate the artifacts as follows:

$$\hat{F_e} = \{f_1 \mid |f_1 - f_2| < Th, \forall f_1 \in F_{bs}, f_2 \in F_{ps}\}$$

$$F_e = F_{ps} \setminus \hat{F_e}$$

where F_e represents the frequencies of the detected emanations from the object of interest (e.g., spy camera) and Th indicates the threshold for the same emanation spike identification that is configured as 1 by default in our experiments.

To further demonstrate the effectiveness of our two-step subtraction approach for artifact removal, we collect the baseline dataset in a typical office room without spy camera deployment. Then, we collect the test dataset in this office room with the hidden camera being deployed. We conduct our two-step subtraction for the camera's emanation detection. As shown in Fig. 6, the detected spikes after two-step subtraction match with the emanation spikes obtained from the experiments in the anechoic chamber room (i.e., the ground-truth emanations). This is because the wireless communication signals are usually stable over time and the ambient emanations are also stable as long as there are no new electronic devices introduced to the environment. The mismatched emanation spikes come from the noise and imperfect signal cancellation. More importantly, our emanation spike detection is designed to be resilient to noisy spikes by leveraging the periodicity of the emanation spikes. Moreover, we can also frequently update our baseline dataset over time to ensure that our subtraction can mitigate the artifacts introduced by the ambient wireless communication signals and emanations.

5.3 Emanation Source Detection Using Clock Frequency

Pair-wised distance extraction through weighting. To detect the emanations, we propose to compute the pair-wised distances of the emanation spikes detected in two-step subtracted PSD. Obviously, one of these pair-wised distances can indicate the clock frequency or fundamental periodicity of these emanation spikes. Then, our problem becomes how to make the clock frequency or fundamental periodicity outstanding among all those pair-wised distances. To do so, we propose to use the power of the detected spikes contributed to the pair-wised distance (i.e., periodicity) as the weight. The intuition is that the emanation spikes can have high power in comparison to the other noisy spikes due to false positive spike detection. However, since we use the pair-wised distances between the detected emanation spikes, we can see pair-wised distances that are multiple times to each other have a similar weight, which can make it difficult for us to identify the clock frequency or fundamental periodicity.

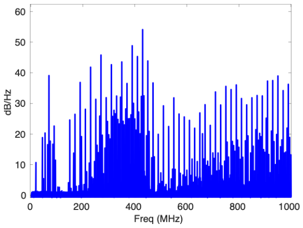


Figure 8: PSD of camera's emanations.

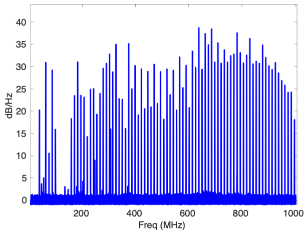


Figure 9: PSD of microphone's emanations.

Re-weighting the pair-wised distance through folding. To make the pair-wised distance that can indicate the clock frequency more outstanding, we can further fold the pair-wised distances that are multiple times to each other. Specifically, we add the weight of the pair-wised distance d to the weight of the pair-wised distances that are multiple times d. In this way, we can make the pair-wised distance indicating the clock frequency to have more weights, which can be outstanding among all the pair-wised distances. Fig. 7 shows the histogram of the pair-wised distance (i.e., periodicity) for the camera's emanation detection with our two-step artifact removal and clock frequency detection. As we can see, the highest bar in the histogram indicates the periodicity of 48MHz which is the camera's clock frequency. The second-highest bar in the histogram indicates a periodicity of 24MHz which is another clock frequency of the camera. This demonstrates the feasibility of our clock frequency detection approach.

5.4 IoT Device Fingerprinting

To demonstrate the feasibility of fingerprinting IoT devices using their emanations, Fig. 8 and Fig. 9 show the emanation pattern of common IoT devices (i.e., cameras and microphones). As we can see, the intermediate frequencies of the emanations from cameras and microphones are different due to their different circuit architectures and components. Since these emanation spikes are periodical, the straightforward idea is to use periodicity (i.e., clock frequency) to characterize these emanation spikes. However, different IoT devices may have the same CPU clock. As a result, we cannot simply use the clock frequency extracted from the IoT devices to fingerprint them. So, we propose to extract the features from IoT devices' emanation profiles. Then, we design a deep neural network architecture for IoT device fingerprinting based on the extracted features.

5.4.1 *Feature engineering.*

The emanation signals exhibit periodic properties across frequencies, which are highly dependent on the IoT devices' hardware architectures. Therefore, we mainly extract the emanation frequency spikes as the features for IoT devices' fingerprinting. To do so, we need to identify the emanation spikes

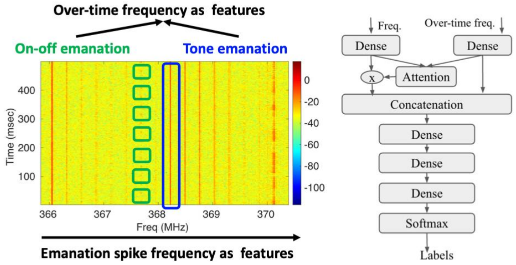


Figure 10: The left figure shows frequency features and over-time frequency features extracted from the spectrogram. The right figure shows the overall framework of *RFScan*'s attention-based multimodal deep neural network architecture consisting of multiple dense layers, a softmax layer, and an attention layer.

in the power spectral density of the emanation profile. Fig. 6 shows the emanation spike detection on the power spectral density of the emanation signals collected from the camera. The frequencies of the identified emanation spikes can formulate a frequency series, which can be used as the input of the deep neural network for IoT fingerprinting. Since different IoT devices exhibit emanations at different frequencies, the size of the extracted frequency series is not consistent across different IoT devices. Therefore, we need to equalize the length of the extracted frequency series through interpolation.

To further enhance our fingerprinting, we leverage the over-time frequency features. Specifically, we slice out the emanation spike in the spectrogram and further infer its emanation pattern. Each emanation spike either exhibits a squared wave or tone-like property. So, we create a series of 0-1 vectors to represent this over-time frequency feature. For example, let's assume there are two emanation spikes in the emanation profile. One emanation spike exhibits a squared wave over time and another one exhibits a tone-like signal. We can have a vector of two elements, where element 0 represents a tone-like signal and element 1 represents a squared wave. At last, we have a feature of two vectors $\mathbf{f}$ and $\mathbf{s}$, where f_i represents $i-$th emanation at the frequency of f_i and s_i is the corresponding over-time frequency feature. We process the over-time frequency features with spline interpolation, which should give us the features with the same length and further be used for IoT fingerprinting.

We propose to combine these two features for IoT device fingerprinting as shown in Fig. 10. For the sake of simplicity and visualization, we just concatenate the frequency and over-time frequency features and further use these combined features to see if they can be used to differentiate the IoT devices.

5.4.2 Deep neural network architecture. To fingerprint the IoT devices, we mainly leverage the frequency and over-time frequency features that we extract from the frequency-domain emanation signals. So, we propose the multimodal deep neural network architecture to model the frequency and over-time frequency features separately, which can be fused to fingerprint IoT devices. Attention mechanism [37] has been exploited to enable the deep neural network to focus on the essential features by adjusting the importance weights. To exploit the role of these two features, we also introduce the attention layer in our deep neural network architecture. Since the over-time frequency features indicate the over-time characteristics of the emanation spikes in the frequency domain, these over-time frequency features are highly related to the frequency features. So, we mainly rely on the frequency features and exploit the over-time frequency features to highlight the frequency features that are important for IoT fingerprinting. Specifically, the attention mechanism learns a mapping from the frequency and over-time frequency features to the weights of the frequency features, which can highlight the features that are important for IoT device fingerprinting. We illustrate this process with the following equations.

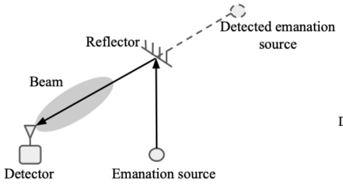


Figure 11: Multipath reflection results in the wrong AoA estimation.

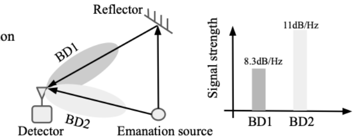


Figure 12: When the emanation source is within the directional antenna's beam, the received emanation signal strength becomes larger.

$$\mathbf{w} = \mu(\mathbf{a^f}, \mathbf{b^t}) \tag{1}$$

$$\mathbf{c^f} = \sum_{i=1}^{N} (\mathbf{w_i} \times \mathbf{a_i^f}) \tag{2}$$

where μ represents the attention layer that takes the over-time frequency features $\mathbf{b^t}$ and frequency features $\mathbf{a^f}$ as the input to generate the weight that can characterize the importance of the frequency features. $\mathbf{c^f}$ indicates the weighted frequency features that can be used to fuse with the over-time frequency features for IoT fingerprinting.

Our deep neural network architecture is shown in Fig. 10, which consists of five layers. The frequency and over-time frequency features can first be used as the input of the dense layer with 512 neurons separately. Then, the attention layer is used to highlight the important frequency features and generate the weighted frequency features, which can be concatenated with the over-time frequency features. Then, three dense layers are used with the number of neurons of 256, 128, and 64 respectively, which can generate the features for classification with the softmax layer. We train our deep neural network with the categorical cross-entropy loss function and Adam optimizer.

5.5 Localizing Hidden IoT Device

After we have detected the existence of IoT devices, we need to localize these IoT devices in the environment so that we can remove them out to protect our privacy. The straightforward idea is to leverage the widely used wireless localization algorithms (e.g., MUSIC [17] algorithm) to localize the IoT devices based on their emanations, while this is not feasible. Because these algorithms require wireless channel measurements across receive antennas on the array, which is not available for emanations as we do not know what the emanation leakage looks like. Moreover, these algorithms

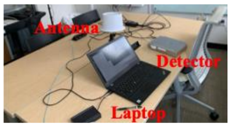


Figure 13: Experimental setup in the typical office room.

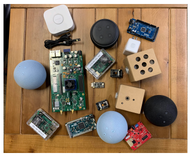


Figure 14: IoT devices used in our experimental evaluation.

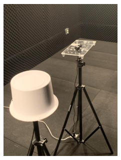


Figure 15: Experimental setup in an anechoic chamber room.

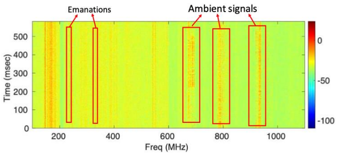


Figure 16: Spectrogram presents the ambient emanations and wireless communication signals in the typical office room.

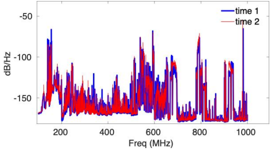


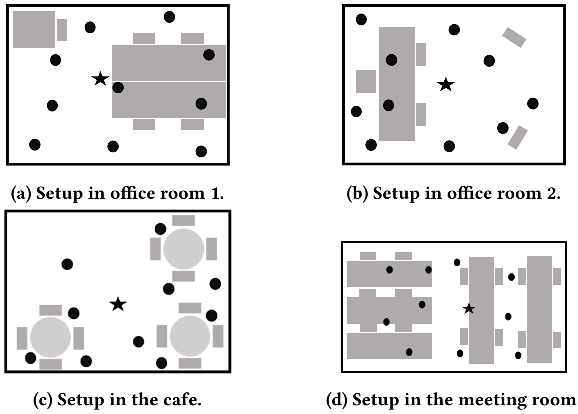
Figure 17: PSD of the ambient wireless signals collected at different times in the hallway of a typical office building.

require an antenna array and since emanations usually occur in sub-GHz down to MHz bands, the antenna array must be very large and impractical.

To do so, we can use a directional antenna instrumented on the detector and automatically scan over the physical environment. In the typical multipath-rich indoor environment, the emanation signals may be reflected off different objects (e.g., walls, chairs, desks, etc.) in the indoor environment. As such, the detector instrumented with the directional antenna may detect the emanation source in the wrong direction as shown in Fig. 11. Therefore, we need to rotate our detector's directional antenna over 360 degrees to scan the environment. As a result, when the emanation source is within the directional antenna's beam direction, the received emanation signal strength becomes larger (i.e., around 3dB increase) as shown in Fig. 12. Therefore, we can find the correct angle of arrival (AoA) of the emanation source after scanning the 360-degree direction. However, this can only allow us to know the coarse-grained AoA of the emanation source. To further localize the emanation source, we can put our detector at another different location to scan the wireless environment, which can give us another AoA of the emanation source. Then, the crossing point of the AoAs can help us to pinpoint or triangulate the location of the emanation source.

6 IMPLEMENTATION AND EVALUATION

***RFScan*'s detector.** We built our *RFScan* prototype with signal hound [47] as the detector, which is instrumented with the omnidirectional antenna [53] for IoT device detection as shown in Fig. 13. For IoT device localization, we use a directional antenna [1, 9] to automatically scan the wireless environment, which is mounted on the ComXim Turntable [14] with an angle resolution of 1 degree. We scan the frequency band from 100MHz to 1GHz, where the emanations are. The detector scans the spectrum with a frequency hop of 100MHz and a sampling rate of 200MHz. Within each frequency hop of the 100MHz spectrum band, the detector collects 32768 data samples. Therefore, we sweep 9 frequency bands with a bandwidth of 100MHz. To strengthen the emanations, the detector sweeps the spectrum 500 times (i.e., an experimental parameter) to average out the noise, resulting in the power spectral density across the frequency band from 100MHz to 1GHz. As a result, it takes about 0.8s to collect all the IQ samples for detection. To improve the emanation spike detection accuracy, we can use more time sweeps, while it introduces the processing delay and a large amount of IQ samples. Therefore, we find that a 500-time sweep is a good balance. Then, we employ the workflow elaborated in Section 5 for emanation source detection, identification, and localization, which are processed in MATLAB on a ThinkPad P1 laptop connected with the signal hound through a Thunderbolt Ethernet adapter.



(a) Setup in office room 1.

(b) Setup in office room 2.

(c) Setup in the cafe.

(d) Setup in the meeting room.

Figure 18: *RFScan*'s set up in different indoor environments, where the black dots indicate the spy IoT device's position and the black star indicates the detector's position.

IoT devices. We tested 14 IoT devices as shown in Fig. 14, consisting of video recorders (i.e., cameras), voice recorders (i.e., microphones), and RF-based sensing radios. Those IoT devices are chosen that can cover the major types of IoT devices on the market such as Amazon, including cameras (e.g., ov5640 [11], ov2640 [6], HZVS4730 [3], and AIY-camera [18]) and microphones (e.g., google home [20], AIY-microphone [19]) that can collect private activity, facial, and speech information about the human beings in the environment.

Anechoic chamber room. The ground-truth emanations of IoT devices are measured in the anechoic chamber room as shown in Fig. 15. We only deploy the detector's receive antenna in the chamber room and the detector outside the chamber room to avoid signal interference from the detector itself. To obtain the ground-truth emanations from the IoT devices, the distance between the detector's antenna and the IoT device is around a half meter. This ground-truth emanations from the IoT devices can be provided by the IoT device manufacturers.

Open and real-world indoor environments. To evaluate the performance of *RFScan*, we mainly conduct the experiments in two office rooms, one meeting room, and one cafe as shown in Fig. 18. Each office room is about 3x4 meters, the meeting room is about 5x6 meters, and the cafe is about 4x5 meters. Note, these are typical rooms in the enterprise building, thereby there are diverse IoT devices such as computers, printers, TV monitors, cameras,

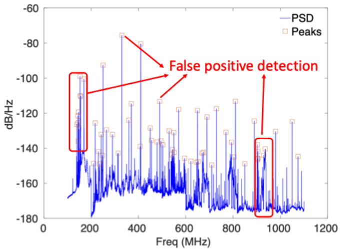


Figure 19: Emanation spikes detection without stimulating the emanation source, resulting in false positive emanation spikes detection.

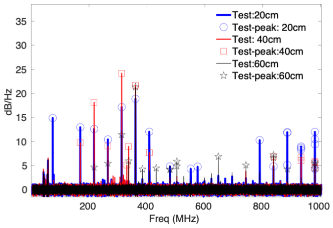


Figure 20: Emanations from the hidden camera with the distance of 20cm, 40cm, and 60cm away from the receive antenna of the detector.

TV controllers, phones, etc. in the environment. As such, there are different kinds of ambient wireless communication signals and electromagnetic emanation signals in these room environments as shown in Fig. 16. Moreover, since they are typical rooms in office buildings, people may move around during the experiments and introduce new IoT devices (e.g., 3-5 people carrying smartphones or laptops) to the environment.

Experimental settings. The signal cancellation-based approach requires that the ambient wireless communication signals be consistent over time. To demonstrate this, we deploy our receiver in a typical office building and receive the ambient wireless signals at different times in one day over the frequency band between 100MHz and 1GHz. Specifically, we first collect the wireless signals at 10:06 am which is indicated as time 1, and re-collect the wireless signals at 11:10 am which is indicated as time 2 in the hallway of a typical office building. As shown in Fig. 17, the power spectral density of the ambient wireless signals is quite consistent over hours. We note that the power density of wireless signals at the specific frequency band is slightly changing over time due to the multipath effect in the dynamic wireless environment, which can not affect our signal cancellation over frequencies. This is because the wireless signals are consistent over the wide frequency band. Even though there are some abrupt wireless communication signals, we can mitigate them by averaging over a long time duration of data collection. Therefore, the emanations from any new IoT devices introduced to the environment can be profiled for detection, while the communication signals cannot affect our system performance. This is because the background wireless signals are consistent over time. The communication signals from the new IoT device may occupy the same frequency band as the existing ones, resulting in efficient interference cancellations.

In the office rooms and cafe, we mainly conduct experiments for IoT device detection and fingerprinting by instrumenting the omnidirectional antenna [53] at the detector. Specifically, we deploy our detector's receive antenna at the center of the room to scan the existence of the IoT devices in the room, such that we can detect the emanations in all directions. In the meeting room, we mainly conduct experiments for IoT device localization. Specifically, we instrument the directional antenna [1, 9] at the detector to scan the wireless environment for the emanation source's AoA estimation with ComXim Turntable [14]. We randomly select 10 spots to deploy the IoT devices and evaluate *RFScan*'s detection accuracy as shown in Fig. 18. As such, we can cover almost all the areas in the indoor environment. At each location, we collect the emanations from each IoT device as one experimental trial. The average distance between the detector's antenna and 10 spots is around 2.5 meters.

During the end-to-end system evaluation, we can first profile the emanations of the common IoT devices purchased on the market, which can be used to train our deep neural network for IoT fingerprinting. Our detector should be continuously monitoring the wireless environment. Whenever a new IoT device is detected, we can further localize it through triangulation using AoA information. Then, we can further fingerprint this new IoT device. If it does not exist in the current training dataset, we can regard it as a new class of IoT device and add its emanation profiles to the training dataset.

To demonstrate the performance of our IoT device fingerprinting, we collect emanations from multiple devices with the same type or brand (e.g., Raspberry Pi and mmWave radar devices). Specifically, we repeat experiments for emanation signal collection 25 times in each indoor environment. As a result, we are able to have 50 frequency feature vectors and 50 over-time frequency feature vectors for each IoT device. We repeat the above experiments in two different indoor environments. In total, we collect 675 measurements. Then, we conduct the feature extraction to obtain the features for IoT device fingerprinting with the deep neural network. In comparison, the conventional machine learning models (e.g., SVC, logistic regression, KNN, etc.) for IoT fingerprinting are developed with the sklearn [15] library. We also leverage the XGBoost classifier in Python to fingerprint the IoT devices. We further split the collected dataset into 80% for training and 20% for testing.

7 EXPERIMENTAL RESULTS

To evaluate the *RFScan*'s system performance, We first conduct system-level evaluations in the practical indoor environment for IoT device detection, fingerprinting, and localization. Then, we conduct the microbenchmarks to understand and show the impact of different factors on IoT device detection.

7.1 End-to-End System Evaluation

7.1.1 Detection Performance. To measure the IoT device detection accuracy with *RFScan*, we do the experiments in three indoor places such as office rooms and cafes. The detection accuracy is defined as the ratio of the correct detection trials to the total trials. The correct IoT device detection is made if we can correctly detect its clock frequency. Since they are typical indoor environments, there are some legitimate IoT devices (e.g., phones, TV monitors, computers, etc.) deployed in the rooms. Specifically, the detector is deployed at the center of the room and we randomly select ten spots to deploy the IoT devices (e.g., spy camera HZVS4730).

Result. Fig. 21 shows the IoT device (e.g., spy camera HZVS4730) detection accuracy in three indoor areas. As we can see, the detection accuracy in office room 1, office room 2, and the cafe is around 0.97, 0.95, and 0.98 respectively, which indicates the good performance of our *RFScan* on IoT device detection. Moreover, the detection accuracy across different indoor areas is close to each other, which indicates that our detection approach is resilient to different physical environments. Even though the different physical

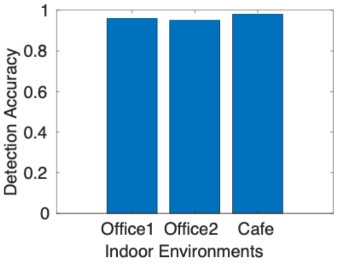


Figure 21: IoT devices detection accuracy in office rooms and cafes.

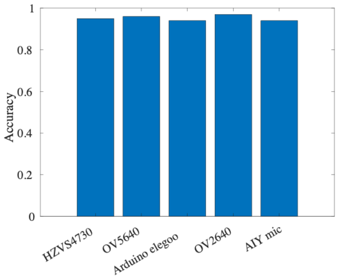


Figure 22: Showcasing different IoT devices' detection accuracy.

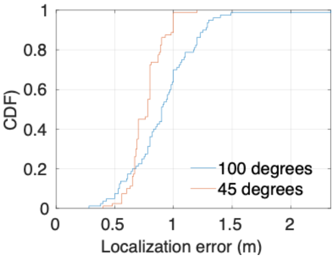


Figure 25: CDF of the IoT device localization error using directional antennas with a beamwidth of 100 degrees or 45 degrees.

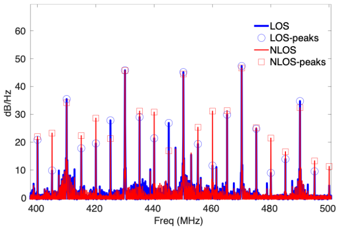


Figure 26: Emanations from the camera in LOS and NLOS scenarios.

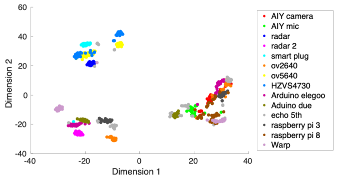


Figure 23: Combined feature-based clustering for the IoT devices in the hallway and office room environments using TSNE algorithm.

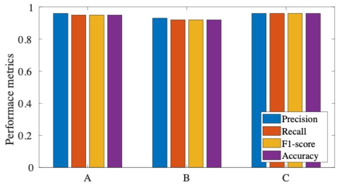


Figure 24: Performance of the IoT fingerprinting when Arduino, echo dot, or spy camera is missing in the training dataset indicated as A, B, or C respectively.

	Precision	Recall	F1-score	Accuracy
XGB Classifier	0.8	0.77	0.77	0.77
Logistic regression	0.46	0.46	0.44	0.46
SVC	0.73	0.72	0.71	0.72
KNN	0.84	0.81	0.8	0.81
DTC	0.85	0.83	0.82	0.83
RFC	0.88	0.85	0.85	0.85
RFScan	**0.96**	**0.95**	**0.95**	**0.95**

Table 2: Precision, recall, F1-score, and accuracy of fingerprinting IoT devices with conventional machine learning models and attention-based multimodal deep neural network using frequency and over-time frequency features.

environments exhibit different multipath effects, it can not affect our IoT device detection. This is because this multipath effect can only affect the received signal strength at our detector due to the destructive or constructive signal addition at our detector's receive antenna. However, it can not affect our detector's detectability as long as our detector can receive the emanation signals. Fig. 22 shows the detection accuracy for five different IoT devices (i.e., HZVS4730, OV5640, Arduino elegoo, OV2640, and AIY mic) in the office room. We chose them because these IoT sensors represent different spy camera sensors, microcontrollers, and assembled sensors. As we can see, the detection accuracy across different IoT devices is similar with an accuracy of around 0.97, which demonstrates the efficiency of our hidden IoT device detection approach. This is because our *RFScan* is resilient to the diversity of IoT devices.

7.1.2 Fingerprinting Performance. To demonstrate the efficiency of *RFScan* on IoT fingerprinting, we compare its performance with the conventional machine learning models. For the conventional machine learning models, we simply concatenate the frequency and time features and then we conduct the TSNE algorithm on the concatenated features. As such, two important features are extracted and used for fingerprinting. However, *RFScan* uses the multimodal deep neural network with an attention layer to model the frequency and over-time frequency features for IoT fingerprinting.

Result. Fig. 23 shows the scatter plot using the combined features, where dimensions 1 and 2 indicate the two important feature dimensions extracted by the TSNE algorithm. As we can see, different IoT devices can be differentiated through the combined features extracted from their emanation signals. However, some IoT devices cannot be accurately differentiated, as their combined features are overlapped. This indicates that we need to use a deep neural network to exploit the frequency and over-time frequency features for fingerprinting.

Fig. 24 shows the performance of IoT fingerprinting on 14 devices, while the deep neural network is trained on 13 devices. We consider there is only one class of IoT devices missing. This is because we add one new IoT device to the training dataset as time goes on. As such, the deep neural network outputs a lower probability on the new device that is not seen during the training process. As we can see, the precision, recall, F1-score, and accuracy are pretty high, even though the deep neural network is trained only on 13 devices. This is because the deep neural network can model 13 devices' emanations very well and regard the unseen device as the new device. Moreover, for different missing devices in the training process, the deep neural network presents a similar performance, demonstrating the resilience of our deep neural network model.

Table 2 shows the performance of IoT fingerprinting with attention-based multimodal deep neural networks and conventional machine learning models. Our attention-based multimodal deep neural network achieves a precision of 0.96, recall of 0.95, F1-score of 0.95, and accuracy of 0.95. As we can see, *RFScan* with the attention-based multimodal deep neural network can achieve high classification accuracy in comparison to the other conventional machine learning models. This is because the deep neural network exploits the feature importance and relations for IoT fingerprinting.

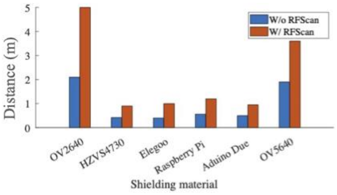


Figure 27: Maximum detection distance for different IoT devices.

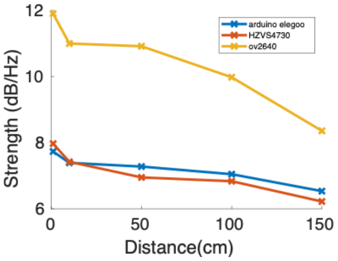


Figure 28: Strength of emanations over IoT device-receiver distance.

7.1.3 Localization Performance. To demonstrate the performance of IoT device localization, we do experiments in a meeting room. The detector's directional antenna is mounted on the turntable which can have an angle resolution of 1 degree. Since the beamwidth of the directional antenna can affect the AoA estimation accuracy, we conduct the experiments with two different directional antennas with a beamwidth of 100° [1] and 45° [9]. The IoT device is deployed in this meeting room and the detector can scan the physical environment by rotating the directional antenna's direction with the turntable. The emanation's AoA is obtained during the directional antenna's rotation when the detector can accurately detect the clock frequency of the IoT device. To localize this IoT device, we can move the directional antenna to another spot and re-scan the environment to detect the IoT device's clock frequency. Then, the location of the IoT device is pinpointed by the triangulation of emanation's AoAs.

Result. Fig. 25 shows the CDF of the localization error for the IoT devices. As we can see, the median localization error is around 0.85m across different IoT devices using a directional antenna with a beamwidth of 100 degrees and a directional antenna with a beamwidth of 45 degrees. The localization error using the directional antenna with a beamwidth of 100 degrees is larger than the directional antenna with a beam width of 45 degrees. This is because the directional antenna with a narrower beamwidth can provide us with a more accurate AoA estimation, resulting in smaller localization errors. We can further use a directional antenna with an even narrower beamwidth to increase our localization accuracy.

7.2 Microbenchmarks

7.2.1 Impact of the Ambient Wireless Signals. To demonstrate this, we put a camera in the office room, which is deployed around 1m away from the detector's receive antenna. We choose the distance of 1m as an example, as we only want to measure the impact of the ambient wireless signals. We try to exhibit the spectrogram of the received signals in the spectrum band between 100MHz and 1GHz. These ambient artifacts can interfere with the emanation source detection, as the emanations are spreading across a wide frequency band.

Result. Fig. 16 shows the spectrogram in the typical office room. As we can see, the cellular communication signals are shown in the spectrogram within the frequency band between 600MHz and 900MHz. Since the amplitude-modulated clock signals are spread over a wide frequency band, the normal wireless communication signals can affect emanation detection when the emanation spikes are overlayed by these wireless communication signals. Moreover, we can see that the ambient emanations from legitimate IoT devices (e.g., TV monitors, computers, phones, etc.) can be mixed up with the spy camera's emanations within the frequency band between 200MHz and 400MHz, which can significantly affect the camera's emanation detection. As shown in Fig. 19, the emanation spike detection in the ambient wireless environment can be very challenging, as the ambient wireless communication signals and background emanations from the legitimate IoT devices may occupy the same frequency band as the emanations from the IoT device of interest due to the frequency spread of the amplitude-modulated clock signals. Therefore, this can motivate us to leverage our core technique of signal cancellation to remove interference from the normal wireless communication signals and emanation signals in the wireless environment. Incorporating our findings of stable ambient wireless signals in Fig. 17, our signal cancellation approach should accurately eliminate the interferences for emanation detection.

7.2.2 Effect on the Detection Range. To measure the effect on emanation detection range, we did experiments in the anechoic chamber room and typical office room with diverse IoT devices. Basically, we measure the power spectral density of the emanation spikes in the anechoic chamber room when the IoT device is deployed 20cm, 40cm, and 60cm away from the detector's receive antenna. We also did experiments in the typical office room to measure the detection range and emanation strength of different IoT devices.

Result. Fig. 20 shows the emanations from the camera which is 20cm, 40cm, and 60cm away from the detector's receive antenna. We can see that the detected emanation spikes are not linearly related to the distance between the receive antenna and the camera. This is because of the multipath effect in the anechoic chamber room that can cause the constructive or destructive emanation signals addition at the detector. This also affects the maximum emanation detection distance. For example, in the anechoic chamber room, we find that a camera (i.e., ov5640) can be detected at 5 meters away from the detector's receive antenna. However, in the office environment, we can detect it at most 3 meters away from the detector's receive antenna due to the multipath effect and the inference of the wireless communication signals that can not be fully eliminated. Moreover, we find that different IoT devices have different maximum detection ranges due to different circuit architectures and components. For example, the microphone (i.e., Fillman microphone) can be detected at most 50cm away from the detector's receive antenna in the office room.

Fig. 27 shows the maximum emanation detection distance for six different IoT devices with and without *RFScan*. As we can see, different IoT devices have different maximum emanation detection distances due to their heterogeneous hardware architectures. For some spy camera devices (e.g., ov2640), we can even detect its emanations 5m away using *RFScan* for noncoherent averaging to boost the emanation strength, while its emanations can only be detected within 2.1m without *RFScan*. We can see that the detection range is increased by 2× with *RFScan*. Fig. 28 shows the strength of the

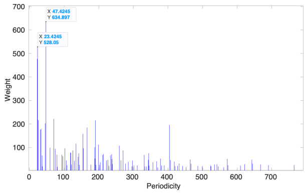


Figure 29: The clock frequency detection result for the camera with a one-hour time gap for baseline and test dataset collection.

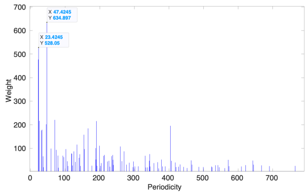


Figure 30: The clock frequency detection result for the camera with baseline and test dataset collected on different days.

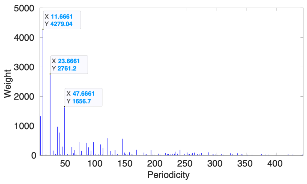


Figure 31: The clock frequency detection result for camera and microphone deployed in the office room.

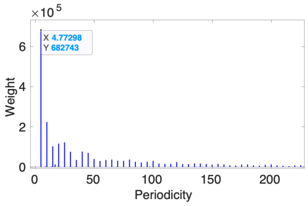


Figure 32: The clock frequency detection result for two different cameras deployed in the office room.

emanation signals decreases as the IoT device-receiver distance increases. This is because the emanation signals have been attenuated over the air. As we can see, the strength of the emitted emanation signals from the ov2640 is much higher than the other IoT devices (e.g., Arduino elegoo and HZVS54730), which are dependent on the strength of the clock signals generated by these IoT devices.

7.2.3 Impact of Non-line-of-sight Emanations. To demonstrate the impact of the NLOS emanation propagation, we do experiments in the anechoic chamber room with a camera deployed 50cm away from the detector's receive antenna. We chose the distance of 50cm as an example, as we want to measure the impact of non-line-of-sight emanations. We first measure the emanation detection when there is only a line-of-sight path between the camera and our detector's receive antenna. Then, we deploy some paper boxes (i.e., $45cm \times 45cm \times 25cm$) around our setup to create the NLOS environment, resulting in constructive or destructive signal addition at the detector. Using a paper box to create the non-line-of-sight environment for the performance evaluation is also employed by the prior work [54]. Note that we use this paper box as the reflector. The other reflectors such as metal boxes could also work to create a multipath wireless environment.

Result. Fig. 26 shows the detected emanations within the frequency band of 400MHz and 500MHz in LOS and NLOS scenarios in the anechoic chamber room. As we can see, the emanation spikes from the camera in the LOS and NLOS scenarios are matched very well, as we use the same emanation source in both scenarios. We also find that some emanation spikes in the NLOS scenario have larger power densities than the emanation spikes in the LOS scenario due to constructive signal addition, and some emanation spikes in the NLOS scenario have smaller power densities than the emanation spikes in the LOS scenario due to the destructive signal addition. This is because of the multipath effect which can affect our emanation spike detection.

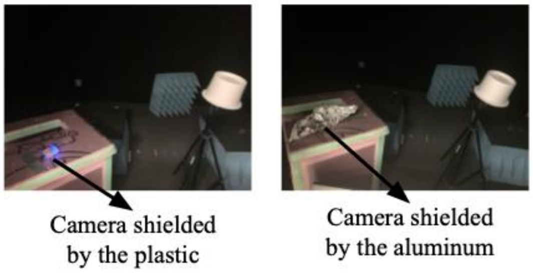


Figure 33: The Spy camera is shielded by plastic and aluminum in the anechoic chamber room for emanation detection.

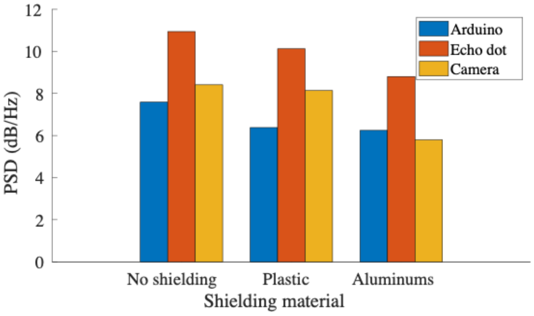


Figure 34: Power spectral density (PSD) of the emanation spikes from three different IoT devices, when they are shielded by different materials.

7.2.4 Effectiveness of Suppressing Wireless Signals. To measure the effectiveness of our two-step subtraction, we collect the baseline and test datasets with a time gap. We put the camera at a distance of 80cm away from the detector's receive antenna. We chose the distance of 80cm as an example as we want to measure the impact of the wireless signals. Our findings could be easily generated to the other distance choices. We expect that the artifacts in the wireless environment should be consistent even with a large time gap in data collection, which can be canceled out with our artifacts removal approach.

Result. Fig. 29 shows the histogram of clock frequency detection for the camera in the office room, when we collect the baseline and test dataset with a time gap of an hour. The two highest spikes in the histogram indicate the detected clock frequencies. As we can see, our algorithm can accurately detect two clock frequencies of the camera which are 24MHz and 48MHz. Since we collected the baseline and test dataset with a time gap of an hour, this result could demonstrate the effectiveness of our artifacts removal approach.

To further demonstrate the efficiency of artifact removal with our two-step subtraction approach, we collect the baseline and test dataset with a time gap of tens of hours. Specifically, we collect the baseline dataset on one day and test the dataset on another day. Fig. 30 shows the emanation source's clock frequency detection result with camera ov5640. As we can see, the two highest bars in the histogram indicate the clock frequency of 24MHz and 48MHz, which matches the camera's clock frequencies that we used during our experiments. We find that we can still correctly identify the clock frequencies of the camera in the office room. This is because the wireless signals in the environment are usually quite consistent. Even though there are some bursty wireless communication signals, they can be captured through baseline dataset collection with a longer time duration. More importantly, we can regularly update our baseline dataset for artifact removal, when we find that there are no periodical emanation spikes in the result of two-step PSD subtraction.

7.2.5 Impact of Multiple IoT Devices. To demonstrate the performance of *RFScan* on detecting multiple IoT devices. We first deploy one camera and one microphone in the office room to see if *RFScan* can accurately detect the IoT device by predicting their clock frequencies. Furthermore, we deploy two different cameras in the office room to see if we can still discover them.

Result. Fig. 31 shows the clock frequency detection result with a histogram for the camera (i.e., ov5640) and microphone (i.e., Fillman mic). As we can see, the highest spikes in this histogram have the periodicity values of 11.6MHz, 23.6MHz, and 47.6MHz, which is close to the ground-truth clock frequencies of the camera and microphone are 24MHz and 48MHz for the camera and 12MHz for the microphone. Fig. 32 shows the clock frequency detection result with a histogram for two different cameras (i.e., ov5640 and ov2640) deployed in the office room. Since the clock frequency of the ov2640 camera is 5MHz and the clock frequency of ov5640 is 24MHz and 48MHz, their clock frequencies are almost integer multiples of one another. As a result, we can see that the highest spike in the histogram has a periodicity value of 4.7MHz which is close to 5MHz. Since the clock frequencies of the ov5640 camera are almost integer multiples of the ov2640 camera, we can only see one outstanding peak in the histogram. However, *RFScan* can still detect IoT devices in the environment due to the periodical emanation spikes in the PSD profile, even though we can not distinguish them based on their clock frequencies. Moreover, we can detect one of the cameras and localize it to kick it out of the environment. Then, we can detect and localize another one. In this way, we can expose all the possible IoT devices deployed in the environment, even though they share similar clock frequencies.

We discuss the impact of the multiple IoT devices on detection, fingerprinting, and localization. As long as the IoT devices have different clock frequencies, they can be detected and separated as different devices, no matter if they appear at the same time or not. This is because we can detect emanation spikes and further identify the clock frequencies to differentiate them. However, if the devices have similar clock frequencies, RFScan still detects the presence of a device but does not separate the devices with similar clock frequencies (i.e., fingerprinting fails). If more than one new device is introduced, *RFScan* cannot simultaneously localize all of them. However, we can localize and remove them one by one as explained above. Therefore, eventually, it can be able to localize all new devices if they have different clock frequencies. Please note that this can not be a problem for existing trusted devices in the environment as they have been removed in the prior baseline stage.

7.2.6 Impact of the IoT device shielding. To evaluate the impact of the IoT device shielding on emanation detection, we mainly measure the power spectral density of the emanation spikes when the IoT devices are not shielded or shielded by plastic and aluminum. This is because the existing IoT devices are usually shielded by plastic or metal and some commercial-of-the-shelf IoT devices are not even shielded at all as shown in Fig. 33. To do so, we use three IoT devices (i.e., Amazon Echo Dot, Arduino, and spy camera HZVS4730) deployed in the anechoic chamber room and cover them with plastic or alumni for power spectral density measurements of the emanations spikes.

Result. Fig. 34 shows the power spectral density (PSD) of the detected emanation spikes from three different IoT devices, when they are not shielded or covered by plastic and aluminum in the anechoic chamber room. Specifically, since there are multiple emanation spikes, we mainly measure the average PSD of all the emanation spikes. As we can see, the different IoT devices without shielding show different average PSD. The strength of the emanation spikes becomes weaker as the IoT devices are covered by plastic and aluminum. This is because the emanation signals are attenuated by the plastic and aluminum. However, since we cannot fully cover or shield the IoT device, the emanations can still be received by the detector afar.

8 RELATED WORK

Dedicated detectors. The dedicated detectors [4, 5, 38, 40] have been widely used for camera and microphone detection, which requires us to thoroughly scan the wireless environment and shut down all the legitimate IoT devices in the background. Moreover, the users need to hold the detector and walk around in the indoor area to detect potential IoT devices. As a result, it needs heavy human workloads and can interrupt the existing IoT device's wireless communication, which disables it from being widely used. Moreover, these dedicated detectors detect spy cameras based on the signal's strength, which cannot distinguish between multiple devices. However, our *RFScan* can passively and automatically scan the wireless environment to detect all kinds of IoT devices.

Sensor-based detection. Prior works also use the camera to detect the hidden camera in the environment [10, 21, 23, 27, 30–32, 39, 41, 43, 51, 52, 56]. The basic idea is that the hidden camera usually has a light indicator when it is turned on, which can be detected by our camera or dedicated sensor such as laser sensor [43]. These camera-based detection approaches usually leverage the hand-held smartphone and use its camera to detect the light indicator on the camera. However, these camera-based detection approaches heavily rely on the light indicator of the hidden camera, which can be easily hidden by the attacker at the software level. For example, the attacker can turn off this light indicator to avoid being detected. More importantly, we still need to hold the camera and walk around in the environment to accurately detect the hidden camera due to the camera's line-of-sight restriction and weak light signals.

Traffic-based detection. The traffic-based detection approaches target wireless cameras, which can transmit wireless signals over the air [12, 13, 24, 26, 28, 33, 35, 36, 42, 44, 48, 55]. As a result, we can sense the variation of the wireless traffic pattern to detect if the wireless camera is recording. Specifically, these works require the detector to make movements (i.e., motion triggering) or use light signals (i.e., light triggering) to trigger the wireless camera's recording and then detect the variation of the wireless traffic pattern. For example, MotionCompass [24] triggers the motion sensor of the wireless camera, which can introduce a sudden variation of the wireless traffic generated by the camera. This sudden variation of the wireless traffic can be captured to detect if there is a camera around and further localize it by triggering the camera in different indoor places. Unfortunately, these traffic-based detection approaches cannot detect the hidden camera without a wireless transceiver, which is widely used to spy on people.

Excitation-probe detection. Recently, some works have tried to excite spy cameras with light or radio frequency signals to sense the electromagnetic wave leakage from them [45]. For example, Digitus [16] mainly leverages the emanations to fingerprint the IoT devices (e.g., Arduino Unos and STM 32s) with deep learning, which requires to design well-trained machine learning models without accurately extracting the clock frequency features from these IoT devices. Moreover, it suffers from the impact of the ambient artifacts. CamRadar [34] uses light signals to scan the environment to excite the cameras and sense the emanations from them. The underlying assumption is that the emanations become stronger as the camera captures the bright scenes. E-eye [29] uses the mmWave signals as the excitation source for IoT device detection, as the electronic components on the IoT device can introduce the nonlinear harmonics due to the interaction with the impinged mmWave signals. Similar to CamRadar and E-eye, DeHiREC [57] exploits electromagnetic interference or ultrasound to excite the hidden microphones and further detect them based on their strengthened emanations. Even though the excitation source can stimulate the hidden cameras or microphones, they can also excite the other legitimate IoT devices in the environment [29] which can aggravate the difficulty of distinguishing between the ambient emanations from the legitimate IoT devices and emanations from the IoT devices. Memscope [46] exploits the emanations from the device's memory to fingerprint the devices. Most importantly, the typical wireless communication signals can significantly interfere with the emanations from the cameras and microphones. However, all of these works have not discussed how to de-clutter the ambient noise and inference in the wireless environment for camera or microphone detection without excitation. As a result, we cannot directly apply them to detect the diverse IoT devices, as we cannot use one excitation source to stimulate all the IoT devices. More importantly, the excitation signals can be easily detected by the attackers to prevent themselves from being detected, while our *RFScan* is fully passive for IoT devices detection which is agnostic to the attackers.

9 CONCLUSION, DISCUSSION, AND FUTURE WORK

In this paper, we present *RFScan*, a system that can detect, identify, and localize hidden IoT devices using unintentional electromagnetic emanations. Specifically, we propose to use non-coherent averaging to boost the emanation strength and design a novel algorithm to eliminate the ambient wireless communication signals, which can be used to sense the hidden IoT devices in the cluttered indoor environment. Further, we design a novel clock frequency detection algorithm for IoT device fingerprinting and localize the IoT devices through AoA-based triangulation. Below, we discuss some limitations of our *RFScan*'s design and propose future potential development opportunities.

Long-range emanation sensing. One great challenge of emanation source detection, identification, and localization is the weakness of emanation signals, which can restrict our emanation detection range. The existing work of using emanations to achieve long-range detection either through exciting the emanation source or relying on the transceiver architecture of the device [34, 45, 46, 57], which can be either easily defended or hardware dependent. Since the wireless communication signals may also cover the emanation signals, we need to further de-cluttering the impact of the normal wireless communication signals. Therefore, it is important to suppress these wireless communication signals and strengthen the emanation signals based on their signal patterns with machine learning models or advanced signal filtering techniques, which can further boost emanations for detection, identification, and localization. Note that our *RFScan* takes the first step to achieve passive emanation sensing for diverse IoT devices.

Fine-grained emanation source localization. To localize the IoT device in the cluttered wireless environment, *RFScan* mainly leverages the directional antenna for the emanation source's AoA estimation. However, *RFScan*'s directional antenna-based AoA estimation is coarse-grained, which is highly dependent on its beam width. Since antenna array-based AoA estimation has been extensively exploited in the wireless localization domain, we can use an antenna array for fine-grained AoA-based emanation source localization. However, the localization accuracy is compromised by the length of the antenna array, especially considering the emanations in low-frequency bands.

ACKNOWLEDGMENTS

This research does not use any AI-based tools for writing. We are thankful to the WCSNG lab members at UC San Diego for insightful feedback. This paper is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via [2021-2106240007] and in part funded by NSF #2232481. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government.

REFERENCES

[1] Alien. 2023. Alien ALR-8698 RFID Antenna. https://rfid.atlasrfidstore.com/hubfs/Tech_Spec_Sheets/Alien/ATLAS_Alien_8697_and_8698.pdf?t=1441381228634.
[2] amazon. 2023. fullfillman speech recoder. https://a.co/d/gGmfkvn.
[3] amazon. 2023. HZVS4730 spy camera. https://a.co/d/hCMR3sX.
[4] Amazon. 2023. JMDHKK Anti Spy Detector, Bug Detector, Hidden Camera Detectors, GPS Detector, RF Signal Scanner Device Detector for GPS Tracker Listening Device Camera Finder. https://a.co/d/17o2DyG.
[5] Amazon. 2023. Knight 5-in-1 Hidden Devices Detector Hidden Camera Detectors GPS, RF Detector Bug Detector Anti Spy Detector Hidden Camera Finder GPS Radio Frequency Detector de camaras y microfonos ocultos. https://a.co/d/3onwqGh.
[6] amazon. 2023. ov2640 spy camera. https://a.co/d/7t86z8O.
[7] Amazon. 2023. Portable Radio Signal Antenna. https://a.co/d/1Tsmgzv.
[8] amazon. 2023. v11 spy microphone. https://a.co/d/j4RKmAt.

[9] Amazon. 2023. ZDTECH Directional Antenna. https://a.co/d/4ZI51Ix.

[10] apple store. 2023. Hidden camera detector. https://apps.apple.com/us/app/hidden-camera-detector/id532882360.

[11] Bubcos. 2023. OV5640 hidden spy camera. https://a.co/d/7a9Lz6M.

[12] Yushi Cheng, Xiaoyu Ji, Tianyang Lu, and Wenyuan Xu. 2018. Dewicam: Detecting hidden wireless cameras via smartphones. In *Proceedings of the 2018 on Asia Conference on Computer and Communications Security*. ACM, none, 1–13.

[13] Yushi Cheng, Xiaoyu Ji, Tianyang Lu, and Wenyuan Xu. 2019. On detecting hidden wireless cameras: A traffic pattern-based approach. *IEEE Transactions on Mobile Computing* 19, 4 (2019), 907–921.

[14] ComXim. 2023. Ceiling-Mount Antennas. https://a.co/d/3MZAeFI.

[15] David Cournapeau. 2023. scikit-learn Machine Learning in Python. https://scikit-learn.org/stable/ none.

[16] Justin Feng, Tianyi Zhao, Shamik Sarkar, Dominic Konrad, Timothy Jacques, Danijela Cabric, and Nader Sehatbakhsh. 2023. Fingerprinting IoT Devices Using Latent Physical Side-Channels. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 2 (2023), 1–26.

[17] Benjamin Friedlander. 1990. A sensitivity analysis of the MUSIC algorithm. *IEEE Transactions on acoustics, speech, and signal processing* 38, 10 (1990), 1740–1751.

[18] google. 2023. AIY camera. https://aiyprojects.withgoogle.com/vision.

[19] google. 2023. AIY microphone. https://aiyprojects.withgoogle.com/voice/.

[20] google. 2023. google home. https://assistant.google.com/platforms/speakers/.

[21] google play. 2023. Glint Finder. https://play.google.com/store/apps/details?id=com.workshop512.glintfinder.

[22] Yeswanth Guddeti, Raghav Subbaraman, Moein Khazraee, Aaron Schulman, and Dinesh Bharadia. 2019. {SweepSense}: Sensing 5 {GHz} in 5 Milliseconds with Low-cost Radios. In *16th USENIX Symposium on Networked Systems Design and Implementation (NSDI 19)*. USENIX, none, 317–330.

[23] Sifeng He, Yuan Meng, and Mali Gong. 2018. Active laser detection system for recognizing surveillance devices. *Optics Communications* 426 (2018), 313–324.

[24] Yan He, Qiuye He, Song Fang, and Yao Liu. 2021. MotionCompass: pinpointing wireless camera via motion-activated traffic. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, none, 215–227.

[25] Zoe Christen Jones. 2022. "American Idol" winner Laine Hardy arrested in Louisiana for allegedly placing recording device in ex-girlfriend's room. https://www.cbsnews.com/news/laine-hardy-arrest-louisiana-recording-device-american-idol/.

[26] Brent Lagesse, Kevin Wu, Jaynie Shorb, and Zealous Zhu. 2018. Automated Hidden Sensor Detection in Sensor-Rich Spaces. In *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, none, none, 433–435.

[27] Li Li, Jianlin Ren, and Xingbin Wang. 2015. Fast cat-eye effect target recognition based on saliency extraction. *Optics Communications* 350 (2015), 33–39.

[28] Zhijing Li, Zhujun Xiao, Yanzi Zhu, Irene Pattarachanyakul, Ben Y Zhao, and Haitao Zheng. 2018. Adversarial localization against wireless cameras. In *Proceedings of the 19th International Workshop on Mobile Computing Systems & Applications*. none, none, 87–92.

[29] Zhengxiong Li, Zhuolin Yang, Chen Song, Changzhi Li, Zhengyu Peng, and Wenyao Xu. 2018. E-eye: Hidden electronics recognition through mmwave non-linear effects. In *Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems*. none, none, 68–81.

[30] Chun Liu, Changming Zhao, Haiyang Zhang, Zilong Zhang, Zitao Cai, and Zhipeng Li. 2019. Spectrum classification using convolutional neural networks for a mini-camera detection system. *Applied optics* 58, 33 (2019), 9230–9239.

[31] Chun Liu, Changming Zhao, Haiyang Zhang, Zilong Zhang, Shuyuan Gao, and Yunshi Wang. 2019. Analysis of mini-camera's cat-eye retro-reflection for characterization of diffraction rings and arrayed spots. *IEEE Photonics Journal* 11, 4 (2019), 1–12.

[32] Chun Liu, Changming Zhao, Haiyang Zhang, Zilong Zhang, Yanwang Zhai, and Yali Zhang. 2019. Design of an active laser mini-camera detection system using cnn. *IEEE Photonics Journal* 11, 6 (2019), 1–12.

[33] Tian Liu, Ziyu Liu, Jun Huang, Rui Tan, and Zhen Tan. 2018. Detecting wireless spy cameras via stimulating and probing. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*. none, none, 243–255.

[34] Ziwei Liu, Feng Lin, Chao Wang, Yijie Shen, Zhongjie Ba, Li Lu, Wenyao Xu, and Kui Ren. 2023. CamRadar: Hidden Camera Detection Leveraging Amplitude-modulated Sensor Images Embedded in Electromagnetic Emanations. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 4 (2023), 1–25.

[35] Markus Miettinen, Samuel Marchal, Ibbad Hafeez, N Asokan, Ahmad-Reza Sadeghi, and Sasu Tarkoma. 2017. Iot sentinel: Automated device-type identification for security enforcement in iot. In *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, none, none, 2177–2184.

[36] Richard Mitev, Anna Pazii, Markus Miettinen, William Enck, and Ahmad-Reza Sadeghi. 2020. Leakypick: Iot audio spy detector. In *Annual Computer Security Applications Conference*. none, none, 694–705.

[37] Volodymyr Mnih, Nicolas Heess, Alex Graves, et al. 2014. Recurrent models of visual attention. *Advances in neural information processing systems* 27 (2014), none.

[38] online source. 2023. Dedicated detectors. https://www.spygadgets.com/collections/counter-surveillance.

[39] Feng Qian, Bao Zhang, Chuanli Yin, Mingyu Yang, and Xiantao Li. 2015. Recognition of interior photoelectric devices by using dual criteria of shape and local texture. *Optical Engineering* 54, 12 (2015), 123110–123110.

[40] REI. 2023. Dedicated detectors. https://reiusa.net/nljd/orion-hx-deluxe-nljd/.

[41] Laurel Sadler and Troy A Alexander. 2010. Mobile optical detection system for counter-surveillance. In *Ground/Air Multi-Sensor Interoperability, Integration, and Networking for Persistent ISR*, Vol. 7694. SPIE, none, none, 227–234.

[42] Muhammad Salman, Nguyen Dao, Uichin Lee, and Youngtae Noh. 2022. CSI: DeSpy: Enabling Effortless Spy Camera Detection via Passive Sensing of User Activities and Bitrate Variations. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (2022), 1–27.

[43] Sriram Sami, Sean Rui Xiang Tan, Bangjie Sun, and Jun Han. 2021. LAPD: Hidden Spy Camera Detection using Smartphone Time-of-Flight Sensors. In *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*. none, none, 288–301.

[44] Rahul Anand Sharma, Elahe Soltanaghaei, Anthony Rowe, and Vyas Sekar. 2022. Lumos: Identifying and localizing diverse hidden {IoT} devices in an unfamiliar environment. In *31st USENIX Security Symposium (USENIX Security 22)*. none, none, 1095–1112.

[45] Cheng Shen and Jun Huang. 2021. {EarFisher}: Detecting Wireless Eavesdroppers by Stimulating and Sensing Memory {EMR}. In *18th USENIX Symposium on Networked Systems Design and Implementation (NSDI 21)*. none, none, 873–886.

[46] Cheng Shen, Jun Huang, Guangyu Sun, and Jingshu Chen. 2022. Electromagnetic Fingerprinting of Memory Heartbeats: System and Applications. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (2022), 1–23.

[47] signal hound. 2023. Signal hound. https://signalhound.com/products/sm200c-20-ghz-real-time-spectrum-analyzer-with-10gbe/.

[48] Akash Deep Singh, Luis Garcia, Joseph Noor, and Mani B Srivastava. 2021. I Always Feel Like Somebody's Sensing Me! A Framework to Detect, Identify, and Localize Clandestine Wireless Sensors.. In *USENIX Security Symposium*. none, none, 1829–1846.

[49] Ritik Singh. 2020. Use Phone Mic To Spy And Record Audio and Listen In Real Time. https://gadgetstouse.com/blog/2020/05/27/use-phone-mic-to-spy-and-record-audio/.

[50] Wei Sun, Tingjun Chen, and Neil Gong. 2024. SoK: Secure Human-centered Wireless Sensing. *Proceedings on Privacy Enhancing Technologies* 2024 (2024), 313–329.

[51] Daniel Svedbrand, Lars Allard, Magnus Pettersson, Pontus Köhler, Markus Henriksson, and Lars Sjökvist. 2019. Optics detection using an avalanche photo diode array and the scanning-slit-method. In *Technologies for Optical Countermeasures XVI*, Vol. 11161. SPIE, none, none, 167–177.

[52] Jakobi Teknik. 2023. Spy hidden camera Detector. https://apps.apple.com/us/app/spy-hidden-camera-detector/id925967783?mt=8..

[53] transformational security. 2023. Ceiling-Mount Antennas. https://www.powerfulsecurity.com/Products?id=cma.

[54] Jue Wang and Dina Katabi. 2013. Dude, where's my card? RFID positioning that works with multipath and non-line of sight. In *Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM*. none, none, 51–62.

[55] Kevin Wu and Brent Lagesse. 2019. Do you see what i see?< subtitle> detecting hidden streaming cameras through similarity of simultaneous observation. In *2019 IEEE International Conference on Pervasive Computing and Communications (PerCom*. IEEE, none, none, 1–10.

[56] X Zhang and Feng Qian. 2017. A Fast Recognition Algorithm for Photoelectric Peeping Equipment. *DEStech Transactions on Computer Science and Engineering* none (2017), none.

[57] Ruochen Zhou, Xiaoyu Ji, Chen Yan, Yi-Chao Chen, Wenyuan Xu, and Chaohao Li. 2022. DeHiREC: Detecting Hidden Voice Recorders via ADC Electromagnetic Radiation. In *2023 IEEE Symposium on Security and Privacy (SP)*. IEEE Computer Society, none, none, 658–673.