RISiren: Wireless Sensing System Attacks via Metasurface

Chenghan Jiang[†]
Northwest University
Xi'an, Shaanxi, China
jiangchenghan@stumail.nwu.edu.cn

Jinjiang Yang
Northwest University
Xi'an, Shaanxi, China
yangjinjiang@stumail.nwu.edu.cn

Xinyi Li
Tsinghua University
Beijing, China
xinyili@tsinghua.edu.cn

Qi Li
Tsinghua University
Beijing, China
Zhongguancun Laboratory, Beijing
China
qli01@tsinghua.edu.cn

Xinyu Zhang
University of California San Diego
San Diego, California, USA
xyzhang@ucsd.edu

Ju Ren[*]
Tsinghua University
Beijing, China
Zhongguancun Laboratory, Beijing
China
renju@tsinghua.edu.cn

Abstract

After over a decade of intensive research, wireless sensing technology is nearing commercialization. However, the inherent openness of the wireless medium exposes this technology to security flaws and vulnerabilities. In this paper, we introduce RISiren to reveal the risk. RISiren is a pioneering end-to-end black-box attack system leveraging programmable metasurface with a high level of stealthiness. The key insight of RISiren lies in its ability to generate malicious multipath using metasurface, thereby disrupting wireless channel metrics influenced by genuine human activities and facilitating malicious attacks. To ensure the effectiveness of RISiren, we propose a novel metasurface configuration strategy aiming at creating human-like activities that stem from a comprehensive analysis of how human activities impact wireless signal propagation. We have implemented and validated RISiren using commercial Wi-Fi devices. Our evaluation involved testing our attack strategies against five state-of-the-art systems (including five different types of recognition frameworks) representative of the current landscape. The experimental results show that the adversarial wireless signals generated by RISiren achieve over 90% attack success rate on average, and remain robust and effective across different environments and deployment setups, including through wall attack scenarios.

CCS Concepts

- **Security and privacy** → *Mobile and wireless security*.

Keywords

Metasurface, Adversarial Attack, IoT Security, Security Wireless Sensing, Sensing, Physical Layer Security

ACM Reference Format:
Chenghan Jiang[†], Jinjiang Yang, Xinyi Li, Qi Li, Xinyu Zhang, and Ju Ren[*]. 2024. RISiren: Wireless Sensing System Attacks via Metasurface. In *Proceedings of the 2024 ACM SIGSAC Conference on Computer and Communications Security (CCS '24), October 14–18, 2024, Salt Lake City, UT, USA.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3658644.3690186

1 Introduction

Wireless sensing technology has garnered significant interest from both academia and industry. Its advantage cost-effectiveness, non-intrusive nature, and ability to operate continuously under any lighting conditions make it promising many exciting applications, such as smart homes [11, 12, 14, 15, 21, 31, 33, 34, 49, 57, 61], health monitoring [9, 20, 30, 36, 37, 45, 48, 52, 58, 59], and security authentication [22, 23, 28, 42]. After over a decade of intensive research, this technology is on the cusp of commercialization, exemplified by the forthcoming launch of Zoe Care [7] and Origin [5].

Despite the advancements in enhancing the sensing capabilities and reliabilities of state-of-the-art wireless sensing systems [12, 21, 31, 55, 60], the inherent security vulnerabilities remain significantly overlooked. This hole introduces substantial risks to forthcoming commercial wireless sensing applications. For instance, an attacker can extort a sum in commercial insurance from a nursing home by hiding the fall detection and delaying emergency response times [1]. The reason for risk hole is that the inherently open nature of the wireless medium, there exists the potential for malicious users to tamper with or interfere with the sensing signals in free space.

To emphasize the holes within wireless sensing systems, several pioneering studies have explored the vulnerabilities of these systems by deliberately interfering with wireless sensing channels [19, 29, 39, 47, 62]. Although these works have demonstrated commendable attack performance, they often fall short in terms of practical applicability for two primary reasons. Firstly, the use of additional active devices to execute attacks makes them more detectable by the victim systems. Secondly, some strategies require a profound understanding of the victim system's framework and operational parameters, a requirement that significantly reduces their feasibility in real-world scenarios where such information might not be readily accessible. These limitations render existing approaches less effective in investigating the potential risk of wireless sensing.

[†] This work was done during his visit at Tsinghua University.
[*] Corresponding author.

CCS '24, October 14–18, 2024, Salt Lake City, UT, USA
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0636-3/24/10
https://doi.org/10.1145/3658644.3690186

Consequently, is there an adversarial attack strategy capable of enabling attackers to launch attacks without prior knowledge of the victim system while maintaining a high level of stealth to evade detection? If such a strategy exists, it exposes a critical vulnerability in ongoing commercial wireless sensing systems. This revelation would, in turn, push the urgent security need for commercial wireless sensing systems to evolve.

In this paper, we propose an affirmative answer through *RISiren*[1], an end-to-end black-box attack system with high stealthiness to reveal the risk. RISiren injects attack by programming radio frequency (RF) environment based on metasurface as illustrated in Figure 1. Inspired by wireless identification accuracy is easily affected by spatial multipath noise [41, 51, 54, 63], our key insight is to generate malicious multipath by leveraging metasurface, thereby tampering with the RF link affected by human activities to enable malicious attacks. The stealth capabilities of RISiren stem from the following factors: (i) The metasurface, as an innovative device, does not generate signals by itself but reflects signals from the transmitter. It can prompt RISiren evading from extra source detection; (ii) Through careful design of the metasurface configuration, we can manipulate the phases and frequency of the signal, allowing for the injection of pseudo activities.

Though the basic idea sounds straightforward, it is non-trivial to construct RISiren due to the following challenges:

(C1) How to determine an efficient metasurface configuration strategy for generating disruptive multipath and interfering with human activity characteristics within the sensing signal? The configuration strategy of metasurface determines the combination of the signals reflected from human activities and attack multipath at the receiver. An effective configuration strategy can disrupt the activity pattern, conversely, an ineffective configuration strategy will fail to mask the activity pattern, leading to an unsuccessful attack. This is because: 1) Human activity pattern changes dynamically in the time and frequency domains, which means the time-invariant static multipath introduced by the metasurface is inadequate for interfering with activity recognition; 2) different human activities have different activity magnitude distributions, which makes it difficult to achieve a one-size-fits-all metasurface configuration strategy.

(C2) How to build a black-box attack strategy without knowing the framework and parameters of the victim system? Existing wireless sensing solutions include multiple models in recognition, such as machine learning (ML) models and deep neural network (DNN) models. The attacker doesn't have access to the details of the victim system classifier. Therefore, implementing a black-box attack strategy to cross-recognition systems is an important challenge.

To answer the first challenge (**C1**), our key insight is that switching different metasurface configurations can generate time-variant interference to the wireless channel. In addition, increasing the energy of interference can enhance the submergence of human activity characteristics. Therefore, RISiren aims to identify configurations that maximize interference differences. To this end, we design a configuration optimization algorithm to maximize interference signals. Specifically, through the combination of beamforming and nulling beam, RISiren ensures that the metasurface can generate maximum perturbations in all directions, thus maximizing the effectiveness of the attack.

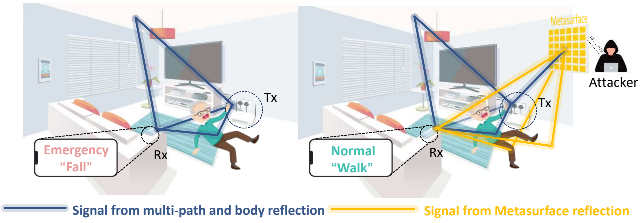

Emergency
"Fall"
Rx
Tx
Normal
"Walk"
Rx
Tx
Metasurface
Attacker
Signal from multi-path and body reflection
Signal from Metasurface reflection

Figure 1: Illustration of RISiren attack cases.

To deal with the second research question **(C2)**, our key observation is that the accuracy of wireless sensing recognition systems is sensitive to the feature representation [26]. In other words, the ambiguity of the extracted features blurs the intersectional boundaries within a classifier, leading to considerable variances in accuracy. Hence, RISiren can choose some activities (such as walking) with obvious and highly independent features in all recognition systems as camouflage activity, thereby enabling robust attack across recognition systems. However, the key question is: how does the metasurface generate high-quality human-like camouflage activities by reflecting the wireless signals? A straightforward solution is to directly change the switching rate of metasurface configurations based on the characteristics of the target activity. Yet its disguised result is coarse-grained and hardly achieves high accuracy attack performance. Hence, RISiren firstly uses an approximation algorithm to obtain the time-frequency profile of human activity, and then we utilize a genetic algorithm to optimize the switching rate of configurations at different time slices, to create highly fidelity signal patterns corresponding to the camouflage activity.

We implement RISiren on a metasurface consisting of 16×16 meta-atoms and spanning an area of $0.35 \times 0.35 m^2$. Its compact size enables discreet integration into everyday environments like walls, furniture, and wall art, facilitating inconspicuous attacks. We conducted extensive experiments on 5 representative systems to validate the performance in stealthy, robust, and destructive of RISiren under different impact factors and environments. We achieved an average Attack Success Rate of up to 90% and a maximum range of 10 meters. Besides, RISiren can even maintain high attack performance when attacking across different obstacles. Importantly, RISiren allows for attacks without disrupting the victim's normal communications on the link. In summary, our contributions include the points below:

- To the best of our knowledge, RISiren is the first black-box attack system to conduct adversarial tampering attacks by metasurface without providing additional sources, which only regulates the existing electromagnetic wave environment in the environment.
- We build a novel attack strategy to maximize the interference differences and generate human-like activity from the metasurface by carefully designing the approximation and optimization algorithm.
- RISiren reveals serious overlooked vulnerabilities in wireless sensing systems, prompting the industry to think about the security aspects of this emerging technology. Furthermore,

[1] "RISiren" derived from the sea-nymphs "Siren" who lured sailors to their death with a bewitching song in ancient Greek mythology.

we discuss and propose potentially feasible countermeasures to protect the availability of wireless sensing systems.

2 Preliminary

In this section, we introduce a background of programmable metasurfaces, followed by a description of the traditional wireless sensing model.

2.1 Programmable Metasurface

Programmable metasurface is a key technology in new-generation wireless networks as it enables dynamic control of the radio environment. A metasurface is a synthetic surface with digitally reconfigurable radio wave reflection properties. This emerging concept originated from the physics of metamaterials and a metasurfaces, and has recently gained substantial traction along with the emergence of "intelligent reconfigurable surfaces" for 6G wireless networks. A programmable metasurface is commonly implemented using cost-effective 2D arrays of electronically adjustable meta-atom reflectors, controlled by microcontrollers or FPGAs. Many meta-atoms can modify the amplitude and/or phase of the reflecting signals in a coherent manner to achieve beamforming or beam steering. More specifically, a programmable metasurface can consist of M columns and N rows of reflecting meta-atoms, and each meta-atom functions as a n-bit phase shifter. For example, a 2-bit phase shifter can provide 4 phase states (i.e., 0, $\pi/2$, π, and $3\pi/2$). A phase value of each meta-atoms is referred to as a *coding parameter*, and the set of coding parameters for all the meta-atoms is also called a configuration. Given that only ambient signals are reflected, the metasurface inherently operates with high energy efficiency and does not necessitate active RF chains. Consequently, the metasurface entails low hardware complexity.

2.2 Traditional Wireless Sensing Model

Typically, RF signals from a transmitter bounce off multiple objects (i.e., walls and human bodies) and eventually combine at the receiver. Suppose $X(f,t)$ and $Y(f,t)$ represent the transmitted and received signals on the sub-carrier with frequency f at time t. $Y(f,t)$ can be expressed as:

$$Y(f,t) = H(f,t) \times X(f,t) \tag{1}$$

where $H(f,t)$ refers to the channel frequency response (CFR), which encapsulates the multipath of the surrounding environment and human activities. $H(f,t)$ can be written as:

$$\begin{aligned} H(f,t) &= H_s(f,t) + H_d(f,t) \\ &= H_s(f,t) + \sum_{p=1}^{P} a_p(f,t)\, e^{-j2\pi f \tau_p(t)}, \end{aligned} \tag{2}$$

where $H_s(f,t)$ is the static multipath component, including the Los path and the reflected path from the environment; $H_d(f,t)$ is the dynamic path reflected by human activity; $a_p(f,t)$ represents the signal attenuation of p^{th} path; $\tau_p(t)$ denotes the time delay resulting from the p^{th} path length.

3 Threat Model

Attack Scenarios: Suppose a pair of transmitter and receiver, denoted as Alice and Bob, are used for wireless sensing (Figure 1). These devices can play various roles such as Wi-Fi access points, mobile devices, or laptops. Eve acts as an adversary who can manipulate a metasurface to generate a malicious multipath for disrupting Bob's sensing capabilities. Eve'sRISiren metasurface serves as a sophisticated physical layer attacker, manipulating the signal patterns received by Bob, to create an illusion of human activities. The metasurface launches stealthy attacks from behind walls, without requiring a direct line-of-sight to Bob and Alice. Notably, no synchronization or real-time coordination is needed between the RISiren metasurface and Alice/Bob. Furthermore, Bob remains stationary and the RISiren metasurface is aware of Bob's relative direction. The RISiren metasurface can be equipped with a Wi-Fi packet sniffer [8, 56] for monitoring human activity and triggering the RISiren attack.

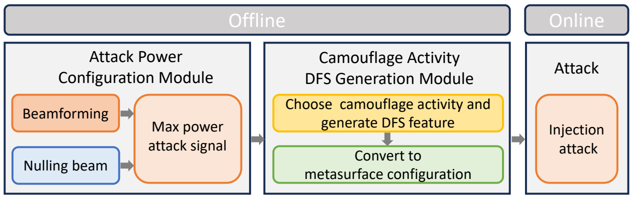

Offline
Online
Attack Power Configuration Module
Beamforming
Nulling beam
Max power attack signal
Camouflage Activity DFS Generation Module
Choose camouflage activity and generate DFS feature
Convert to metasurface configuration
Attack
Injection attack

Figure 2: The attack overview of RISiren.

A typical attack scenario can be an intelligent medical center equipped with advanced commercial wireless recognition system equipment. The elder's dangerous activities (i.e., falling) can be monitored in real-time through wireless sensing devices deployed inside the room or in the corridor. Owing to a thin form factor, metasurface can be hidden in the facades of the ambient environment, such as murals or outside walls. The attacker uses the disguised activity strategy to trigger the victim system to misjudge the user's activities, such as identifying a fall as walking, which is ignored by medical staff, causing security risks. Such motivations may stem from defrauding high medical insurance or harming the life safety of guardians.

Attack mode: We focus on two attack modes:

(1) The first attack goal is play-in-plug malicious attacks. The metasurface continuously injects malicious multipath to generate pseudo activity for deceiving the sensing system recognition even when no one is active.

(2) The second one is the smart trigger attack, which focuses on the long-term latent attack. The metasurface stays dormant and only activated when the wireless signals are disturbed by human activities such as falls. In such scenarios, the metasurface needs to integrate auxiliary devices such as mmWave [3] or Wi-Fi sniffers to capture the occurrence of activities. Then, RISiren immediately triggers the generation of a spoofing activity. Note that sniffers are only used to monitor human activity and trigger the attack mode of metasurface. Sniffers don't affect the stealthy of RISiren. For example, the size of commercial millimeter-wave (mmWave) sniffer HLK-LD2420 is only $20mm \times 20mm$[3], thus it can be embedded in the metasurface easily.

4 System Overview

4.1 System Workflow

The system workflow is shown in Figure 2. Before the attack, the metasurface generates attack configurations offline (Section 5.2). These configurations are essentially a sequence of the meta-atoms' states, which distort the impinging signal patterns to fake human activities (Section 6.2). Ultimately, under RISiren' adversarial attack, the victim's activity recognition system will be tampered with the activity results expected by the attacker.

4.2 Design Choice

RISiren is an end-to-end black-box stealth attack system against wireless sensing. Actually, it reveals universal risks inherent in wireless sensing systems, irrespective of the frequency band and signal protocol. To comprehensively illustrate the working principles of RISiren, we take a Wi-Fi-based fall monitoring system as an attack example for the following reasons:

(i) *Real-world significance and practicality.* Fall detection has become one of the main causes of death for the elderly [2]. Medical reports indicate that the golden rescue time is only one hour for severe falls [53]. Fall detection represents of the very few practical wireless sensing use cases, which has been commercialized [5] and recently deployed by Verizon [16]. Therefore, the consequences of attacks on such fall detection systems are undoubtedly fatal, raising significant concerns about the safety of wireless sensing.

(ii) *Commonality of sensing principles.* Human fall detection shares the same sensing model as other activity recognition systems, such as gesture recognition. These systems identify various activities by using signal processing or deep learning models to analyze the time/frequency domain features within the RF data. Therefore, while RISiren primarily focuses on attacking the human fall scenario, its fundamental concept can easily translate to other sensing applications.

(iii) *Fall detection is the toughest obstacle to overcome in attacks.* Human fall events are abrupt and often cause highly disturbing features on wireless sensing signals. Hence, fall detection is one of the most reliable classes of activities that off-the-shelf wireless sensing devices can easily detect. By attacking this challenging baseline, we can essentially push the limit of RISiren and verify its potential against other wireless sensing activities with less dramatic features.

5 Metasurface Configuration Strategy

In this section, we first describe the metasurface-based sensing model and then detail our proposed schemes for generating disruptive multipath signal patterns by using an efficient metasurface configuration strategy.

5.1 Metasurface-based Sensing Model

We now extend the basic wireless sensing model (Eq. 2) to accommodate the case when a metasurface (MTS) is involved. As illustrated in Figure 1, the metasurface introduces a new controllable path between the Tx and Rx (i.e., Tx → metasurface → Rx). Thu the wireless channel can be written as:

$$H\left(f,t\right) = H_s\left(f,t\right) + H_d\left(f,t\right) + H_m\left(f,t\right) \tag{3}$$

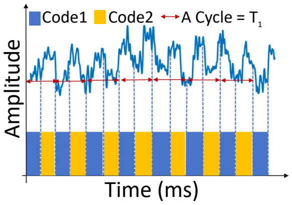


Figure 3: The CSI under different codes.

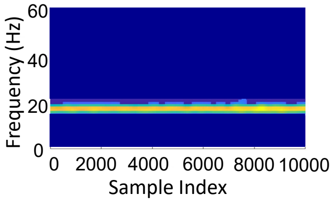
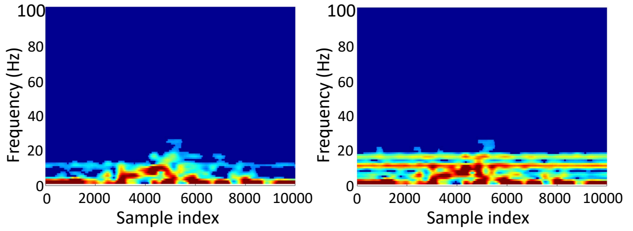


Figure 4: Frequency of 50ms switching rate.



(a) The result in low-intensity attack (b) The result in high-intensity attack

Figure 5: The result of frequency spectrum under different CSI amplitude change.

where $H_m(f,t)$ is the metasurface path.

When changing the metasurface configurations, the wireless channel is distorted. This is because the configuration determines how a metasurface interacts with the electromagnetic waves (i.e., redirection or reshaping). Therefore, the Eq. 2 can be rewritten as:

$$H\left(f,t\right) = H_s\left(f,t\right) + H_d\left(f,t\right) + \alpha_{code_i} e^{-j\phi_{code_i}} \tag{4}$$

where α_{c_i} and ϕ_{c_i} are the amplitude and phase provided by i^{th} configurations.

Finally, to extract the Doppler frequency shift (DFS) spectrogram from $H\left(f,t\right)$, we adopt the short-time Fourier transform (STFT) to obtain Doppler shifts spectrogram. It can be represented by:

$$STFT(H\left(f,t\right)) = f_s\left(t\right) + f_d\left(t\right) + f_m\left(t\right) \tag{5}$$

where $f_s\left(t\right)$ represents the frequency resulting from the static path, typically considered as 0 since the static path is generally not subject to time variations. $f_d\left(t\right)$ is the DFS caused by human activity. $f_m\left(t\right)$ is the frequency brought by the metasurface link.

5.2 Metasurface configuration Strategy

In this section, we mainly focus on introducing how to determine effective metasurface configuration strategies to generate destructive multipath, including the attack frequency 5.2.1 and power 5.2.2.

5.2.1 *How to generate frequency offset patterns?*

To ensure the effectiveness of attack performance, RISiren should generate attack frequency within the frequency range introduced by human activities (i.e., 0 ∼ 80 Hz [46]). To do so, the key insight of RISiren is that the configuration of the metasurface needs to be periodically switched, to distort the impinging signals and interfere with those signals reflected by human activity.

For ease of explanation, we establish a simple example only focusing on the metasurface path $H_m(f,t)$. We set T_1 as a period to switch the meta-atoms' configurations. Without considering

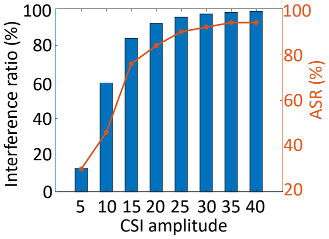
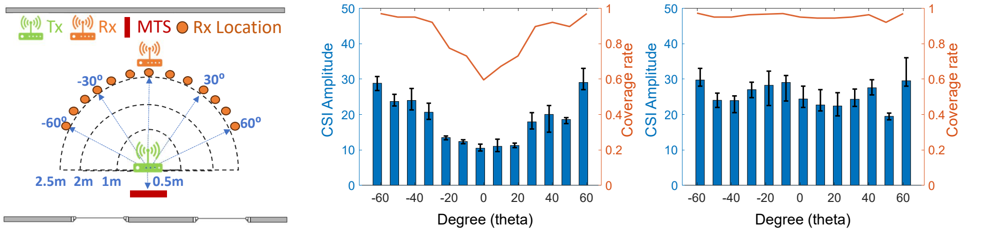


Figure 6: Different ratio of different CSI amplitudes.



(a) The experiment layout (b) Beamforming and metasurface "OFF" (c) Beamforming and nullforming

Figure 7: Different solutions correspond to the different attack powers.

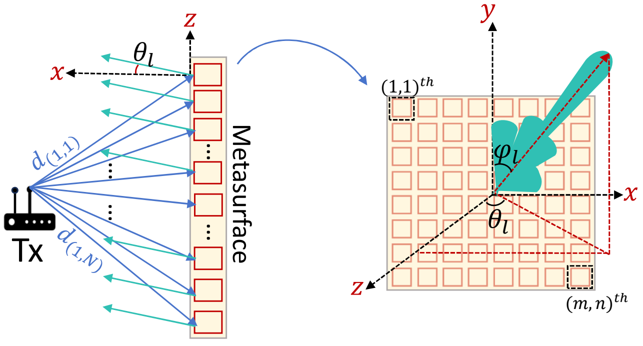


Figure 8: The geometry of the metasurface.

external interference, the amplitude of the Tx-MTS-Rx channel only depends on the configurations of the metasurface. So we can simplify $H_m(f,t)$ as a square wave signal with T_1 period. In other words, the periodic variation of the configurations causes a frequency offset of $1/T_1$ Hz. Thus, we can intentionally design the switching period T_1 to impose any frequency offset we want.

We conduct experiments to verify the feasibility of introducing dynamic path from metasurface to interfere with human activities. Specifically, the transmitter and receiver are Wi-Fi devices, separated by 3 m. The metasurface is deployed along the 30° angle relative to transmitter with a 1 m distance. All devices are placed at a height of 1.1 m. We set T_1 to 50 ms. Although the CSI data is inevitably subject to external interference in real scenarios, we can still clearly see periodic CSI amplitude changes in Figure 3, which is consistent with the configuration switching period. As shown in Figure 4, when the interval between two metasurface coding configurations is 50 ms, a frequency offset of 20 Hz (1/50 ms= 20 Hz) can be observed in the spectrum.

5.2.2 *How to guarantee the attack power?*

Insufficient attack power (i.e., the intensity of the adversarial frequency) can impede the successful execution of an attack. For instance, in Figure 5(a), when the metasurface generates a lower attack power at 15 Hz, it's apparent that the original human activity feature remains largely unaffected.

To investigate the impact of the limited attack power, we first collect a dataset of typical in-home human activities - including sitting, standing up, falling, bending, and walking - by using commercial Wi-Fi devices. We implemented a CNN-based wireless sensing model following [36] as a simple 5-class activity recognition system. Subsequently, we artificially superimpose different adversarial frequency intensities on the data of the original activity data through changing the amplitude of adversarial frequency. The amplitude of adversarial frequency changes from 0 to 40. The results are shown in Figure 6. As the frequency amplitude increases, there is a noticeable increase in the attack success rate. This is because the higher the adversarial frequency intensity, the interference towards the human activities (Figure 6. Consequently, the classifier is more inclined to extract features from the adversarial frequency, resulting in misidentification.

Therefore, RISiren should create a high-intensity attack signal. Yet this is challenging in practice, given the metasurface's limited size and lack of power amplifiers. A straightforward solution is to switch the metasurface between the "OFF" state and beamforming state. Unfortunately, when in the "OFF" state, the metasurface acts like a mirror for the impinging signals, leading to only minor difference in reflected signal intensity between the two states. We conduct an experiment to demonstrate the phenomenon. The deployment layoutt and results are shwon in Figure 7(a) and Figure 7(b), respectively. In Figure 7(b), we can see that when the receiver is located in the range of −20° to −20°, the metasurface can not maintain a high intensity, implying that RISiren cannot achieve high-intensity attacks for the majority of the angles.

To overcome this challenge, we propose switching the metasurface between beam "nulling" and beamforming states. Specifically, beamforming configuration aims to focus its energy towards a specific direction, whereas beam nulling diminishes or nullifies the signal in that direction. By combining both, we can effectively increase the contrast between different states of the metasurface, leading to an adversarial frequency offset pattern with high intensity. In what follows, we elaborate on the beamforming and nulling design.

Beamforming algorithm: The metasurface has $M \times N$ meta-atoms ($M = N = 16$ in our prototype). As shown in Figure 8, the propagation distance experienced by the incident electromagnetic wave before impinging on the $(m,n)^{th}$ meta-atom is $d_{(m,n)}$, which leads to an initial phase shift of $\phi^{I}_{(m,n)} = -kd_{(m,n)}$, where $k = 2\pi/\lambda$; λ is the wavelength of signal; $m \in [1,M]$ and $n \in [1,N]$. Suppose the direction of the legitimate receiver is (θ_l, φ_l), where θ_l and φ_l are the elevation and azimuth angles, respectively. Therefore, to beamform the signal towards the direction (θ_l, φ_l), the theoretical phase distribution for the meta-atoms is:

$$\phi^{T}_{(m,n)} = -k\left(x_m \sin\theta_l \cos\varphi_l + y_n \sin\theta_l \sin\varphi_l\right), \tag{6}$$

where x_m and y_n are the X-axis and Y-axis distances of the $(m,n)^{th}$ meta-atom relative to the origin of coordinate. Thus, the *ideal phase compensation* generated from each meta-atom should be the difference of $\phi^{I}_{(m,n)}$ and $\phi^{T}_{(m,n)}$:

$$\phi^{C}_{(m,n)} = \phi^{T}_{(m,n)} - \phi^{I}_{(m,n)}. \tag{7}$$

In our prototype, each meta-atom is a 2-bit phase shifter with 4 possible states: 0, $\pi/2$, π, $3\pi/2$. Phase shifters approximate the desired phase shift through a quantization rule:

$$\mathrm{Q}\left(\phi^{C}_{(m,n)}\,|_{2-bit}\right) = \begin{cases} 0, & otherwise \\ \frac{\pi}{2}, & if\ \frac{\pi}{4} \le \phi^{C}_{(m,n)} < \frac{3\pi}{4} \\ \pi, & if\ \frac{3\pi}{4} \le \phi^{C}_{(m,n)} < \frac{5\pi}{4} \\ \frac{3\pi}{2}, & if\ \frac{5\pi}{4} \le \phi^{C}_{(m,n)} < \frac{7\pi}{4} \end{cases} \tag{8}$$

Beam nulling algorithm: Like beamforming, one straightforward method to achieve beam-nulling configuration is to adjust the phase shifts across meta-atoms to counteract signals in a specific direction. However, this method can only create the null-line rather than null-beam with any significant width. It leads to the attacker having extremely precise information about the victim receiver's location, which is hardly obtained in reality. Additionally, even when a receiver is positioned within the null-line, it remains vulnerable to interference from sidelobes in adjacent directions, degrading the effectiveness of the attack.

To overcome this issue, we use particle swarm optimization (PSO) to optimize the configuration. Note that other optimization methods can also be used [10, 40]. We define the objective function of optimization as follows:

$$\begin{aligned} &\mathcal{J} \in min\sqrt{(l_1)^2 + (l_2)^2 + (l_3)^2} \\ &S.t. \\ &\theta_l \in [\theta_l - BW_1/2, \theta_l + BW_1/2] \\ &\varphi_l \in [\varphi_l - BW_2/2, \varphi_l + BW_2/2] \\ &\gamma \in C_u\left(\theta_l, \varphi_l\right) \end{aligned} \tag{9}$$

where $l_1 = \frac{1}{|Gain_{(\theta_l,\varphi_l)} - BFGain_{(\theta_l,\varphi_l)}|}$ aims to minimize the gain of the desired beam; $l_2 = Var(Gain_{(\theta_l,\varphi_l)})$ targets beam flatness within the desired direction, reducing variance in the desired area, enhancing uniformity; $l_3 = \max(Gain_\gamma) - \min(Gain_\gamma)$ intends to eliminate high-gain mainlobes in directions other than the specified victim receiver. (θ_l, φ_l) represents the desired beam nulling direction. $Gain(\cdot)$ refers to the null-gain function within the beam pattern. $Var(.)$ is the variance function. BW_1 is the beam width around the θ_l in elevation angle, and BW_2 is the beam width around the φ_l in azimuth angle. γ is the collection of directions other than the target direction (θ_l, φ_l).

We conduct an experiment to demonstrate the effectiveness of the beamforming and nulling design. The deployment is shown in Figure 7(a). Compared to Figure 7(b) and Figure 7(c), when the beamforming gain remains constant, the beamforming and nulling beam can obtain larger amplitude changes in all directions. This is because the nulling beam configuration can achieve constant low levels of gain in all directions. The results of the frequency spectrum are as shown in Figure 5(b), it can cover the original sample seriously when using beamforming and nulling beam configurations.

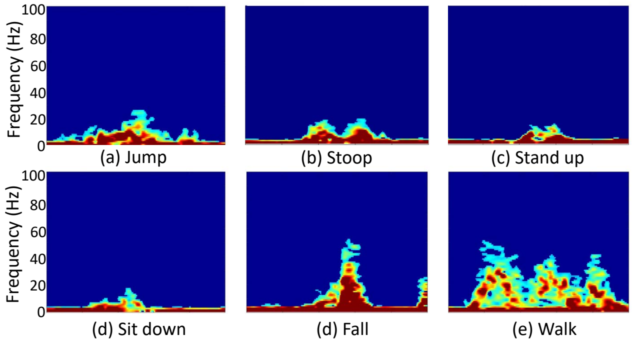


Figure 9: The different DFS samples of different activities.

6 RISiren Attack Strategy

In this section, we devise a universal attack strategy that ensures targeted attacks while maintaining the stealthiness of RISiren.

6.1 How to Choose Camouflage Activity?

Existing wireless sensing methods for human activity recognition have commonly utilized classifiers to discern and categorize various activities:

$$y = F(x; \omega) \tag{10}$$

where y is the ground-truth label of human activity and x is the input data, which can be extracted features or original data. $F(\cdot)$ is the representation of a classifier, such as Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest (RF), CNN, and so on. ω is the set of all parameters in the classifier.

Traditional attack systems [19, 29] aim to optimize attack data x_{adv} and assume prior knowledge of the classifier model (i.e., $F(\cdot)$) is known. They then inject x_{adv} to push the original data across the decision boundary, leading to the misidentification of activities. The process can be formulated as:

$$y_{fake} = F(x_{ori} + x_{adv}; \omega) \tag{11}$$

where x_{adv} and x_{ori} represent the attack data and the original data of activity, respectively. y_{fake} is the desired activity label of the attacker.

Although such systems have shown promising results in terms of attack performance, the classifier model is hard to obtain for attackers in reality. In addition, these methods lack generalization, as the impact of the adversarial attack data x_{adv} heavily depends on a specific classifier [62]. Any alterations in the recognition system tend to reduce the attack accuracy. This is because the accuracy of wireless sensing recognition systems is sensitive to the feature representation from the original data.

To overcome this challenge, we propose a feature camouflage scheme to achieve an attack without knowing the classifier model. Specifically, we choose one normal daily activity (e.g., walking) as the camouflage activity. Then, we design the metasurface to generate high-quality human-like camouflage activities by reflecting

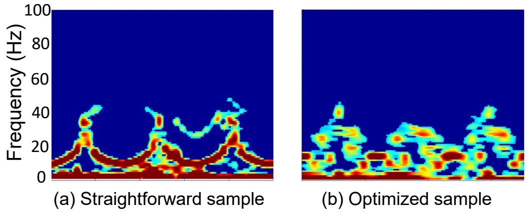


Figure 10: Results of generating camouflage sample.

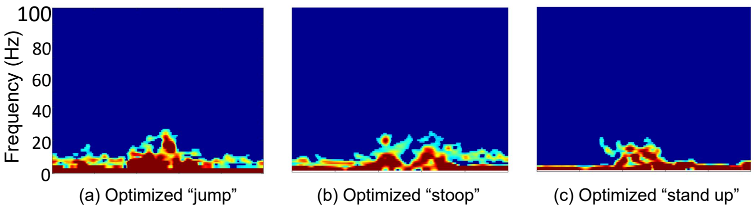


Figure 11: The RISiren metasurface can create different human-like activities.

the wireless signals. By doing so, we can cover the features of the original activity, so that the injected adversarial activity features will overwhelm the normal legitimate features. With this measure, our attack strategy can be characterized as:

$$X_{ori} + X_{adv} \longrightarrow X_{adv} \tag{12}$$

One immediate question here is: what kind of activities should we select as X_{adv}? Since the wireless signals usually carry adverse domain[2] information unrelated to human activities, leading to extracting features that may not accurately represent a unique activity. Choosing such activities as part of the camouflage for RISiren's attacks could lead to unstable attack performance. For instance, if "push" is selected as the camouflage activity to target the recognition system, the ambiguous features might yield an attack result of "pull", which deviates from the intended outcome.

We take daily activities as an example, we collect several activity samples, and the DFS are as shown in Figure 9. Among these activities, "walking" exhibits distinct feature differences owing to its duration and periodicity. Therefore, our primary approach involves selecting these types of activities, such as walking, as target camouflage activities. It is important to note that while walking is a prime example, various other activities are exhibiting robust and distinct features that can also serve as suitable candidates for camouflage activities. RISiren can camouflage target activity into the original activity sample due to the independent and intense features exhibited by the adversarial activity. Consequently, the recognition system outputs the incorrect result that RISiren specifically devises.

6.2 How to create the camouflage activity?

The next question is: how does the metasurface generate high-quality human-like camouflage activities by reflecting the wireless signals? A straightforward method involves altering the metasurface switching sequence according to feature fluctuations. However, this manually crafted pattern, as depicted in Figure 10(a), results in rough features that noticeably differ from normal features, losing its stealthiness.

To address this limitation, we propose an approximation algorithm that derives the time-frequency profile of human activity. We then employ a genetic algorithm to optimize the configuration switching rates across various time intervals. This optimized approach aims to create highly realistic signal patterns corresponding to the chosen camouflage activity, ensuring higher fidelity in mimicking human activity.

Specifically, we simulate the relationship between the switching sequence S_n of metasurface and the DFS profile of human activities. Note that the DFS profile indicates the frequency of human activity at each time slice or the frequency of switching metasurface configuration. We extract the DFS profile L_{act} and L_{S_n} by:

$$\begin{aligned} L_{S_n} &= Max(DFS_{S_n}) \\ L_{act} &= Max(DFS_{act}) \end{aligned} \tag{13}$$

where DFS_{S_n} is the DFS generated by S_n, and DFS_{act} represents the truth activity DFS.

To optimize the similarity between L_{S_n} and the true DFS profile L_{act}, we use the Pearson correlation coefficient [13] as the fitness measure. Pearson correlation coefficient is used in statistics to measure the correlation of two vectors and can be written as:

$$\rho_{S_n,L_n} = \frac{cov(S_n, L_n)}{\sigma_{S_n}\sigma_{L_n}} \tag{14}$$

Where L_1 and L_2 are two independent vectors. $cov(L_1, L_2)$, σ_{L_1}, and σ_{L_2} are the covariance and standard deviation of L_1 and L_2. The value range is $[-1, 1]$. The result is closer to 1, L_1 and L_2 are more similar. We formulate the fitness score Fs as :

$$Fs = \frac{1}{\rho_{(L_{S_n},L_{act})} + 1} \tag{15}$$

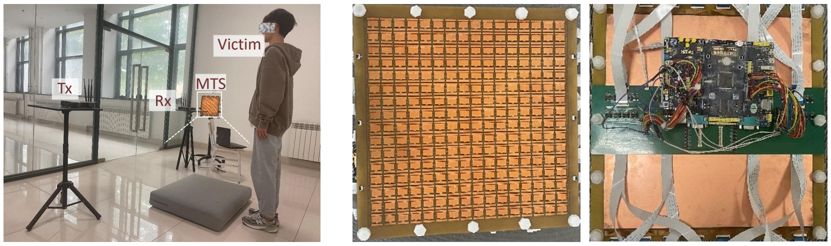
which reflects the correlation between the L_{S_n} and the real activity ones. The lower the fitness score is, the closer they are. As shown in Figure 10(b), the final optimized sequence exhibits high similarity to the ground truth data. We also use the optimization we proposed to generate other activities' camouflage samples like jump, stoop, and stand. The results are shown in Figure 11. We can find they are similar to the original activities. Thus, through the optimization of the switching sequence, RISiren exhibits its ability to generate diverse human-like activities by harnessing metasurface technology.

7 Implemention

Metasurface Prototype. RISiren metasurface is designed by assembling multiple optimized meta-atoms. We build a prototype of RISiren metasurface that consists of 16 × 16 meta-atoms. All the meta-atoms are evenly distributed inside an area of $0.35 \times 0.35\ m^2$, with a distance of 19.5 mm between adjacent meta-atoms, as shown in Figure 12(b). To reconfigure the PIN diode states of each meta-atom, we embed a bias layer to transmit DC bias voltage to each PIN diode (SMP1340-040LF PIN diodes [6]).

Hardware control. To configure the whole metasurface, we design a control circuit module consisting of a microcontroller

[2] In this work, a domain is a deployment setup including factors like users, deployment environment, device setup, etc.



(a) Experiment scenario

(b) Top and bottom view of metasurface

Figure 12: The experiment layout and the details of metasurface.

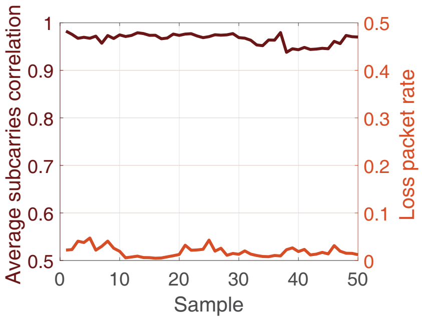


Figure 13: The stealth verification.

(i.e., STM32H743IIT6, STMicroelectronics) and 64 SN74LV595 shift registers to provide different DC voltages (0 V or 5 V) for each meta-atom. Specifically, we divide the entire metasurface board into 4 zones. For each zone, we use two channels in the microcontroller to transmit a data stream with 128 bits to control 128 PIN diodes, as shown in Figure 12(b). Via the above set up, the controller is now able to independently configure the state of each meta-atom's PIN diode.

8 Evaluation

8.1 Experiment setup

In order to verify the performance of RISiren under different classifier architectures, we reproduced different systems, including 5 different classifiers model. Specifically: (1) Falldefi [37], splits activity data through power burst curves, and uses SVM to distinguish falls from other daily events. The highest reference recognition accuracy in Falldefi is 93%, and the recognition accuracy we reproduce is 88.30% (2) The work [36] analyzes the limitations of existing segmentation based on signal energy, chooses to use a sliding window to segment the spectral features of the data, and then combines and inputs into the CNN network for identification. The reference recognition accuracy in this work was 92%, and the recognition accuracy we reproduced is 97.4%. For ease of reference, we use its abbreviation WFDUSI instead in the experiments. (3) Actrec [11] identifies by extracting multiple feature combinations such as mean, variance, maximum, and minimum in mean Doppler shift (MDS) from the spectrum. Naive Bayes, KNN, and decision tree models are selected for identification. The recognition accuracy of the three classification models in Actrec is 94.6%, 96.2%, and 98.9% respectively. The recognition accuracy we reproduced is 96.61%, 93.42%, and 90.77% respectively. For ease of reference, we use the abbreviations ActrecBayes, ActrecKNN, and ActrecTree instead.

We experiment with RISiren in three different environments, including a single apartment, a corridor, and a bedroom. The experiment scenario is shown in Figure 12(a). The victim transmitter and receiver are equipped with an Intel 5300 NIC and three antennas. The transmission rate of systems is 1000 packets per second. All transceivers are put in 110 cm height so that the motion of users with different heights can be detected. All the experiments have been approved by our Institutional Review Board (IRB).

Data size[3] We invite 7 volunteers (5 males and 2 females) to simulate the victims' falling, and collect five daily activities (including falling, walking, sitting, stooping, and standing) and four gestures (pushing, circling, zigzagging, and sliding) as dataset to reproduce the five recognition systems. We totally collect 1,670 data samples in 3 environments and 13 evaluation scenarios experiments. We conduct 10 tests for each variable changed in different experiments.

Metric. We define two metrics to quantify the effectiveness of attack: *Attack success rate (ASR)*, which is the ratio between the number of successful attacks and the total number of attacks. A successful attack means that when sending a confrontational signal, the victim system misjudges it as the type of target gesture the attacker needs. Can be accessed by: $ASR = \frac{N^{Success}_{A\longrightarrow target}}{N^{Aall}_{A\longrightarrow target}}$, where the $N^{Success}_{A\longrightarrow target}$ is the number of activity A was successfully misjudged as the target activity, and the $N^{Aall}_{A\longrightarrow target}$ is total number of attacks launched. *Recognition success rate (RSR)*, which is defined to evaluate the attack performance in the victim's recognition system. Specifically, the *RSR* is defined as the ratio of the number of samples correctly classified by the recognition classifier to the total number of samples. In our system, the lower *RSR* is, the better attack performance is.

8.2 Micro-benchmark

Stealthiness of RISiren. In order to verify the stealth of RISiren, we reproduce the existing Wi-Fi aware co-channel interference detection method. WiAnti [17] and Phaseanti [18] proposed subcarrier correlation judgment during each subcarrier and packet loss rate per time to check whether there is co-channel interference, respectively. The lower the subcarrier correlation or the higher the packet loss rate, the more severe the environmental channel interference. And they respectively defined threshold indicators for judging whether there is interference. We collect CSI samples under the RISiren attack and calculate the average of the two metrics, as shown in Figure 13. The numerical distribution of the two metrics during the attack is consistent with the reference normal. The average subcarrier correlations are all above 0.9, higher than the recommended threshold of 0.86, and the average packet loss rate is also lower than the threshold level of 0.07. And RISiren will not affect the communication performance of the victim system.

[3] Dataset is available at: https://github.com/HappyChenghan/RISiren.

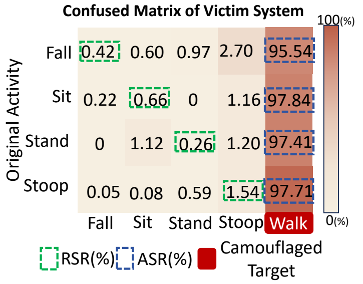
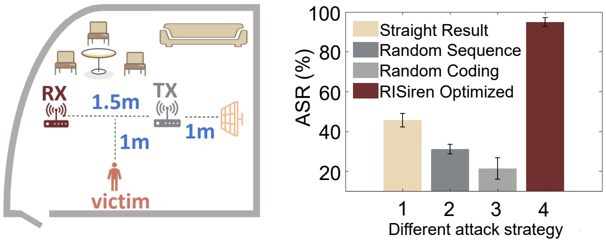


Figure 14: The attack results under different activities.



(a) Experiment layout
(b) ASR for different strategy

Figure 15: Verification of the camouflage algorithm

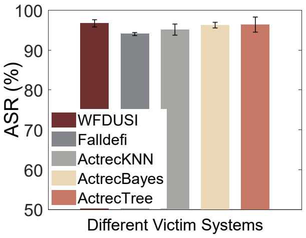
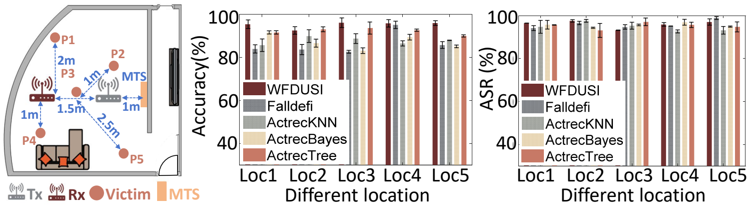


Figure 16: The overall performance of RISiren.

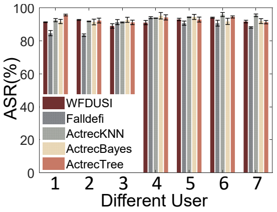


(a) Experiment layout
(b) The result w/o attack
(c) The result under attack

Figure 17: The performance of different victims' locations.

User ID	Gender	Height (cm)	Weight (Kg)	Age
1	Male	170	65	23
2	Female	158	48	19
3	Male	179	75	22
4	Male	183	86	31
5	Female	164	52	46
6	Male	176	69	20
7	Male	170	55	28

(a) Different volunteers' information



(b) The ASR of different Users

Figure 18: The performance of different users.

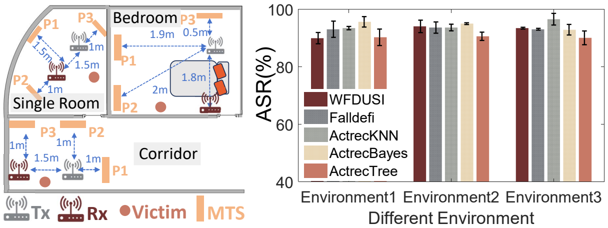
Therefore, we demonstrate that the RISiren system remains stealthy and difficult to detect during attack.

Verification of attacking different activities. We choose four daily activities including falling, standing, sitting, and stooping to evaluate the attack performance under different activities. We set the metasurface behind the transmitter at 1 m, and keep the distance between the transceiver at 1.5 m. The attack results (i.e., *ASR* and *RSR*) are shown in Figure 14. We observe that *ASR* exceeds 95% for all activities. Additionally, the *ASR* for other daily activities is higher than for falling. This is because these activities involve slower speeds, resulting in lower Doppler frequencies and smaller CSI amplitude differences (Figure 9), which indicates that they are easier to tamper with. Furthermore, the *RSR* for different activities is less than 1.6%, meaning the recognition system struggles to accurately identify human activities. Note that while both *ASR* and *RSR* measure the attack performance, *ASR* is a more rigorous metric than *RSR*. *ASR* indicates a successful attack only if the recognition result matches the target disguised activity. In contrast, *RSR* considers an attack successful if the recognition result differs from the ground-truth label, allowing for ambiguous recognition to be considered successful attacks. Therefore, to accurately represent the attack performance of RISiren, we use *ASR* in subsequent experiments.

Compared the attack performance of RISiren with different baselines. In this experiment, we aim to verify the effectiveness of the configuration strategy and camouflaged activity strategy of RISiren by comparing the following three baselines, the details are as follows: We use two cases to verify the camouflaged activity strategy's effectiveness. The first one is a naive approach: switching optimized beamforming-nullforming coding configurations by using a random switching sequence. More advanced, the second one is switching the optimized coding configurations using a rough manually human-like configuration switching sequence without optimizing. We represent the two baselines as "Straight Result" and "Random Sequence". To verify the effectiveness of the configuration strategy, we use switching random coding configurations by using the switching sequence optimized by our algorithm. We represent the baseline as "Random Coding". We set the metasurface behind the transmitter at 1m, and keep the distance between the transceiver at 1.5 m. We evaluate the average *ASR* of RISiren and three baselines. The results are shown in Figure 15, we can clearly see the average ASR of RISiren is 95.04%, and the *ASR* of the "Random Coding" is only 22.45%. It is due to the random coding configuration can hardly provide the powerful adversarial feature to mislead the recognition system. Besides, the *ASR* results of "Random Sequence" and "Straight Result" are both lower than 50%. Although the "Straight Result" achieve higher attack performance compared with "Random Sequence", it still can not camouflage the adversarial activity feature well.

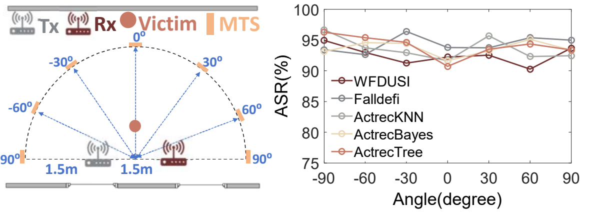
8.3 Overall performance of RISiren

We deploy the actual attack scenario in a fan-shaped single room to verify the overall effect of the attack. As shown in Figure 15(a), we set the same layout that keep transceiver spacing to 1.5 m, and conduct the volunteer fall activity at 1 m in the direction of the center line. The metasurface is deployed 1 m away from the transmitter, surrounded by various furniture to mimic the complex multipath in the real room. As shown in Figure 16, the ASR of five different recognition systems are all above 94%. It can be seen that RISiren has robust attack generalization for different recognition systems.



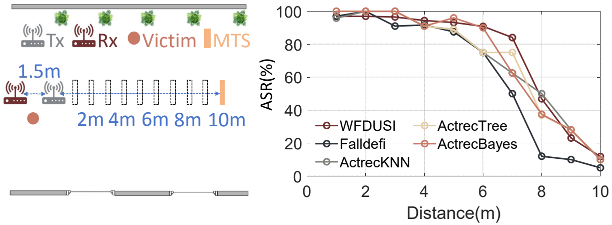
(a) The Scenario layout (b) The ASR of different environments

Figure 19: The performance of different environments.



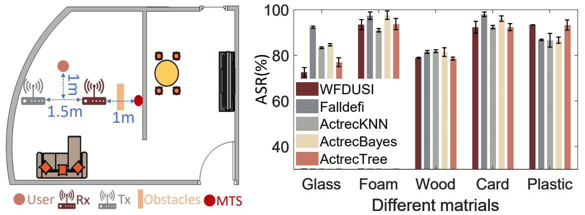
(a) Experiment layout (b) The ASR of different orientation

Figure 20: The performance of different orientations between metasurface and victims' system.



(a) The experiment layout (b) The ASR of different distances

Figure 21: The performance of different distances between Tx-metasurface.



(a) The experiment layout (b) The ASR under different obstacles

Figure 22: The performance of traversing obstacles.

8.4 Evaluation of extrinsic impact factors

We investigate major external factors that may affect the attack performance: the impact of victims' falling locations, impacts of personal differences, and different environments, In each experiment, we only change a single variable while maintaining others.

Impact of victim's location. In order to evaluate the impact of the victim falling at different locations on the attack, we randomly select five locations (P1-P5) in a single room to conduct the victim falling test. The actual deployment is shown in Figure 17(a). We separate the distance of the transmitter (Tx) and receiver (Rx) by 1.5 m and deploy the metasurface at a distance of 1 m from Tx. We also verify the ASR when the metasurface attack is working and the recognition accuracy of five different systems without attacking. The results are shown in Figure 17(b). We can see that the victim recognition system we reproduce is robust. The average attack success rates are shown in Figure 17(c). We find that the victim falling down at different locations has little impact on our attack, because RISiren directly injects the adversarial perturbation into the Rx end through the newly created malicious multipath provided by metasurface, and is almost not affected by the victim's location.

Impact of different users. In order to explore the impact of different people on attacking, we recruit 7 volunteers, with heights ranging from 158 cm to 183 cm and weight ranges from 48 kg to 86 kg. The specific information is as shown in Figure 18(a). We deploy the metasurface 1 m behind the Tx, and the transceivers are 1.5 m apart. Each volunteer performs the falling action individually. The experimental results are shown in the Figure 18(b). The average attack success rates for each volunteer are all above 90%. Therefore, we demonstrate the robustness of our attack system against different users.

Impact of different environments. We further conduct experiments in 3 different environments: a single apartment, a bedroom, and a corridor. The bedroom and hall represent the multipath complex environment and the multipath simple environment respectively. In each environment, we select three different attack positions to deploy the metasurface, as shown in Figure 19(a). The experimental results are shown in Figure 19(b). In three different environments, our average attack success rates are 92.4%, 93.60%, and 93.4% respectively. This shows that our attack is extremely robust to the environment.

8.5 Evaluation of intrinsic impact factors

Impact of different orientations between metasurface and victims' system. In order to explore the impact of the deployment position of the metasurface on the attack performance, our deployment is shown in Figure 20(a). We use a radius of 2.25 m from the center of Tx-Rx and evaluate at every position between $-90°$ and $90°$ in 30° intervals. We verify the performance of our attack under four recognition systems. The results are shown in the Figure 20(b). In each direction, RISiren can maintain over 90% ASR.

Impact of different distance between metasurface and victims' system. In order to verify the impact of the distance between the metasurface and the victim system, our deployment is shown in Figure 21(a). The distance between the Tx and Rx is 1.5 m. The distance between the metasurface and Tx varied from 1 m to 10 m at 1m intervals. The results are as shown in Figure 21(b), we can see RISiren can maintain 10.08% ASR effective until 10 m. As the distance between metasurface and victims' system increases, the average ASR decreases. RISiren can still maintain a 66.71% ASR in average at 7 m, ensures the attacker's ability to launch attacks from long distance.

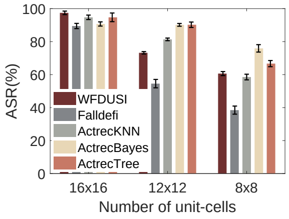


Figure 23: Result of different number of meta-atoms.

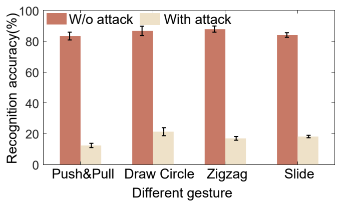


Figure 24: Attacks on gesture recognition systems.

Impact of metasurface traversing different obstacles. In order to consider the concealment problem of metasurface and the existence of obstacles in real deployment. We test the impact of the metasurface when blocked by different materials in the environment. The materials we explore are glass, foam, wood, cardboard,and plastic. The specific experimental deployment is shown in the Figure 22(a). We place the metasurface 1 m away from the Tx end and place plates of different materials in the middle. The experimental results are shown in the Figure 22(b). We found that under these types of obstacles, the average attack accuracy is higher than 80.45%. Besides, compared to drywall which is typically found in hospitals or nursing homes[4], the materials we evaluated like wood and glass exhibit higher signal path loss[44]. Therefore, RISiren can penetrate daily obstacles and achieve outside wall attack.

Impact of different number of meta-atoms. Here we verify the impact of different numbers of meta-atoms on attack performance. We keep the deployment as same as Figure 15(a) shows. For each test, we set the two outermost meta-atoms of the metasurface to a state of 0. When the element is set to 0, It can be understood that the meta-atoms do not work. This procedure is consistently repeated in each iteration, gradually reducing the count of active meta-atoms. That is, the number of meta-atoms each time is: 16×16, 12×12, 8×8. The experimental results are shown in the Figure 23. As the number of meta-atoms decreases, the attack success rate gradually decreases. This is because the intensity of the injected adversarial frequencies decreases as the number of meta-atoms decreases. It is worth noting that the attack effect of Falldefi is the most affected. This is because a frequency shift intensity threshold is set in its preprocessing stage to divide action occurrences. When the attack intensity is lower than the threshold, the activity cannot be triggered. In addition, as the number of meta-atoms decreases, the dimensions of the metasurface are: 21.6 cm×21.6 cm and 14.4 cm×14.4 cm respectively. Therefore, the ease of deployment of the metasurface will be further improved. In order to balance this trade-off, the attacker can customize the number of metasurface units according to the actual attack scenario.

8.6 Attacking other sensing systems

In this section, we evaluate the compatibility of our system against threats to sensing systems outside of the above evaluation cases. Specifically, we analyze the generality of the RISiren attack method to systems with other pre-processing methods; and use a gesture recognition system as a case to demonstrate the generalization performance of the RISiren attack scheme.

In order to verify the attack performance in other gesture perception aspects, We reproduced Widar3.0 [60] recognition system. We invite 10 volunteers to collect a set of data on four gestures (Push&Pull, Draw Circle, Draw Zigzag, and Slide), and use the CNN model to implement a gesture recognition system as our attack target, in which the CSI amplitude is used as input data. The result of RISiren attack is shown in Figure 24, the average recognition accuracy under attack is lower than 25%. It proves the potential threat of RISiren to gesture recognition applications.

8.7 Evaluation under harsh attacking conditions

In this section, we evaluate the performance of RISiren under harsh conditions, including attack multi-receivers and attack without exactly location of transceiver.

Performance without transceiver location information. In the RISiren attack, we assume that the victim transceiver is known, but at the same time in actual scenarios, this location information may be not easy to collect. So, we verified the angle error tolerance performance of RISiren for the transceiver. The experimental deployment and results are shown in the figure 25(a)(c). We fixed the metasurface beam pointing at 0° and moved the transceiver within plus or ±30° at intervals of 10°. The experimental results are shown in the figure 25(b)(d). RISiren attack success rate can still reach average of 78.1% within a receiver offset of ±10°, and the attack effect still exists within ±30°. For transmitter offset, RISiren can maintain an average attack effect of over 80.4% within ±30°.

Performance under multiple receivers. In order to verify the attack performance under multiple receivers, we consider a harsh attack scenario. We assume victim systems use multiple receivers for fall detection recognition. We use the feature combination process[50] and the experiment setup is shown in Figure 26(a), We enable the receivers in order until all five receivers are enabled. The result is shown in Figure 26(b), we found that the attack success rate decreases as the number of receivers increases, because the beam of metasurface can only cover limited receivers. When the victim system has three receivers, RISiren can still maintain an average attack success rate of over 50%. And up to five receivers, it still has the attack effect. This proves that we can maintain good attack performance under multiple receivers.

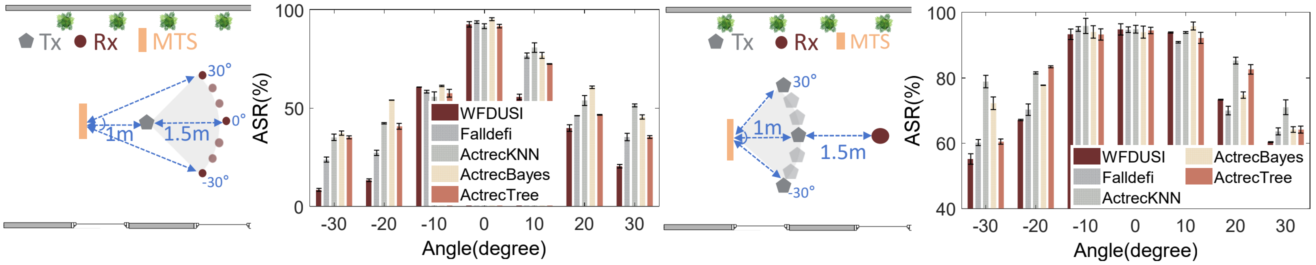
9 Discussion

9.1 Attack generality and Future work

In this section, we aim to discuss the attack generality of RISiren from potential attack scenarios:

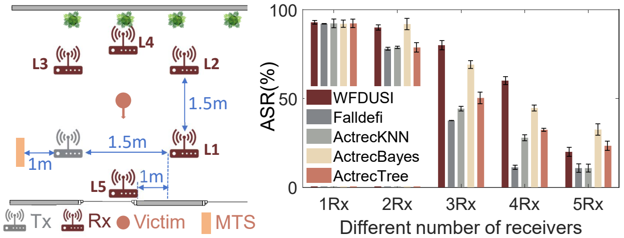
Attack under 3D scenarios. RISiren can attack when the transceiver and metasurface are positioned on different planes. This is because RISiren's metasurface can manipulate electromagnetic waves in a 3D scenario, ranging from −60° to 60° in both azimuth and elevation angles [25]. This region is the metasurfaces' field-of-view (FoV). When the receiver's position falls in the region of FoV, RISiren can still attack effectively. However, the *ASR* will diminish as the boundary is approached since the limitation of FoV.

Attack under multi-activities detection systems. In general, if the operating frequency band of the metasurface is same as that of the recognition system. RISiren can achieve effective attack because metasurface can reshape the channel characteristics of electromagnetic waves. In Section 8.2, we have evaluated RISiren's attack performance under different daily activities. Beyond these



(a) The layout of Rx error (b) The result of Rx error (c) The layout of Tx error (d) The result of Tx error

Figure 25: The performance of different location direction errors from transmitter and receiver.



(a) Experiment layout (b) The ASR of different number receivers

Figure 26: The performance of multiple receivers.

evaluated activities, RISiren's attack strategy can expand to attack other daily activities with lower frequency and strength than target camouflage activities like walking. RISiren's attack performance under gesture recognition is evaluated in Section 8.6. Additionally, the attack target of the RISiren is the receiver rather than the activities themselves. RISiren's key attack strategy is injecting powerful camouflaged activity features to the receiver. If the location of the receiver changes, RISiren only needs to reshape the electromagnetic signal towards the new receiver's location. Therefore we can apply this to the scenarios of multi-activity detection.

Differentiated Attack. RISiren aims to indiscriminately (no matter what the victim's daily activities) tamper with the victim's activity to resemble a camouflaged activity (e.g., walking) while maintaining the passive and low power consumption characteristics of the metasurface. To achieve differentiated attacks, a potential way is embedded an activity recognition algorithm and synchronization scheme on the metasurface to handle transient activities effectively for differentiated attack. Although this approach may increase the power consumption of the metasurface and make it harder for the attacker to succeed, it is an interesting topic for the future work.

Attack under networked metasurface. Networked metasurface can expand the attack coverage region of RISiren when multiple metasurfaces cooperate with each other. Prior work [27] has proposed mechanisms to coordinate multiple metasurfaces for communication coverage improvement. A potential approach is to apply RISiren attack strategies in the N^{th} metasurface and maintain the original coordination algorithm in the $1^{th} \sim (N-1)^{th}$ metasurfaces. It is an interesting topic for our future work."

9.2 Limitations

Inherent limitations of metasurfaces. Due to the passive nature of the metasurface, it can only reach electromagnetic waves on its surface by reshaping Tx. The evaluation results show that its attack distance is limited compared to attack devices that provide additional sources. However, we can achieve a more stealth attack. It is worth noting that for activity-aware attacks in indoor scenes, our requirements for distance and a metasurface deployment location are relaxed, and attackers can deploy metasurfaces at nearby yet stealthy locations (e.g., across wall). RISiren can expand the attack methods against wireless sensors and complement existing active attacks. Besides, since the limitation of the beam width, RISiren can not maintain the high performance without the knowledge of victim's transceiver. We have evaluated the angle tolerance for transceiver in Section 8.7. Fortunately, the result shows that RISiren only need the approximate location. And there are currently some methods that can assist metasurface to estimate the receiver and transmitter position, such as beam scanning, which will greatly reduce the attacker's attack limit, which will be the future work.

9.3 Potential defense

The adversarial interference caused by metasurface is almost indistinguishable from the natural variations of multipath reflections. We verified RISiren's resistance to channel interference detection, and the result was that it was unable to identify the presence of interference in the environment. Therefore, its presence is difficult to detect through existing techniques, such as packet loss analysis method [17] and subcarrier correlation detection method [18] for interference attacks. At the same time, the perceptual encryption methods proposed by some existing work [32] can only prevent malicious attackers from stealing CSI information, but they lack resistance to tampering attacks.

A potential defense measure is to detect the intensity threshold of the feature during feature extraction for activity segmentation, as in the system we use to attack [37]. However, it should be noted that raising the threshold solely for defense purposes will make the user's normal activities impossible. Energy also fails to reach the threshold and is misjudged as no activity, thus affecting the accuracy of normal perception and leading to an increase in error rates. Defense development needs to carefully consider this dilemma. Another potential defense measure is to equip the receiver with a high-sensitivity antenna array to separate benign reflection links from malicious metasurface paths [38], but this goes against the original intention of low-cost and low-power consumption of current Wi-Fi sensing devices. Therefore, a low-cost and efficient wireless sensing protection system deserves further study to resist malicious attacks.

10 Related works

Wireless sensing systems: With the development of wireless sensing technology, there have been some works dedicated to using RFID and mmWave for exciting applications such as human activity sensing, fall monitoring, and vital sign detection. Due to WiFi's excellent popularity and low cost, WiFi-based recognition systems have been extensively explored in the field of human-computer interaction. In recent years, it has been widely proposed for applications such as activity perception and health monitoring. The identification of healthcare applications mainly includes hazardous movements and vital sign signal monitoring. Hazardous action recognition [11, 20, 36, 37, 45, 48, 52] typically utilizes wireless transceivers deployed in the home to monitor the occurrence of fall activity and be able to accurately alert potential risks. Vital signs include heartbeat, breathing, etc. [9, 30, 58, 59], and the relevant displacement of the human body surface is detected through wireless signal characteristics. Gesture recognition through wireless sensing has also been well-studied. In [31, 61], for example, the authors perform gesture perception through commercial off-the-shelf (COTS) WiFi devices and establish a completely open data set for future work. However, the consequences of successful attacks on these important wireless sensing applications are serious.

Attack to wireless sensing system: Previous work has warned us about the vulnerability of wireless sensing systems. Physical-World [29] uses the extra WiFi signal to attack the human behavior recognition system. The extra WiFi signal can cause packet collision by the CSMA/CA protocol [35], which results in the loss of CSI packets. Is-Wars [19] uses cross-technology interference (CTI) to attack, the attacker transmits the noise into the victim CSI data on overlapped frequency bands by ZigBee device. However, this extra active source will be detectable and may lose stealthy. Wiadv [62] uses state-of-the-art full-duplex devices tactfully forward to WiFi signals to mimic dynamic multipaths. It can cause an adversarial Doppler frequency shift in the victim spectrum. However, the high cost of full-duplex antennas creates a burden for attackers

Smart metasurface applications: Intelligent metasurfaces have emerged as a key theme in next-generation wireless communication systems. Substantial research has investigated how a metasurface modulates electromagnetic waves in the environment. In the field of communication, RF-Bouncer [25] proposed a dual-band frequency ($5GHz$ and $2.4GHz$) metasurface to concentrate on the signal and enhance the coverage area; Protego [24] proposed a transmission metasurface provides sidelobe obfuscation communication security and enhances the signal strength in mainlobe. In addition, the application of metasurface in wireless recognition has also garnered attention. IRShield [43] diminished the chance of eavesdroppers eavesdropping on leaked CSI by utilizing metasurface obfuscations. To the best of our knowledge, RISiren is the first work to apply metasurface to attack against wireless recognition.

11 Conclusion

This paper introduces RISiren, a metasurface-assisted end-to-end black-box attack system with high stealthiness. By invisibly injecting adversarial activity features through a generalized camouflage generation strategy, our experimental results show that RISiren attacks achieved the attack success rate over 90% on average, and it maintains robustness under different physical settings. Furthermore, our proposed attack method can be easily generalized to other wireless sensing recognition applications. This work explains potential vulnerabilities of wireless sensing systems and provides insights for future secure and tamper-resistant system designs. Furthermore, RISiren prompts a new contemplation on insidious threats by manipulating the RF environment using metasurface.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant No. 62341201, 62122095, 62432004 and 62072472, and by a grant from the Guoqiang Institute, Tsinghua University.

References

[1] [n. d.]. CareClaim. https://www.accidentclaimsadvice.org.uk/care-home-negligence-claims/.

[2] [n. d.]. Get the facts on falls prevention. https://www.ncoa.org/article/get-the-facts-on-falls-prevention.

[3] [n. d.]. HLK-LD2420. https://www.hlktech.net/index.php?id=1150.

[4] [n. d.]. Nursing home standart. https://www2.gnb.ca/content/dam/gnb/Departments/sd-ds/pdf/NursingHomes/NursingHomeDesignStandards-e.pdf.

[5] [n. d.]. Origin Wireless. https://www.originwirelessai.com/.

[6] [n. d.]. SMP1340. https://www.skyworksinc.com/Products/Diodes/SMP1340-Series.

[7] [n. d.]. Zoe Care. https://techcrunch.com/2024/01/10/zoe-care-zoe-fall/.

[8] Abbas Acar, Hossein Fereidooni, Tigist Abera, Amit Kumar Sikder, Markus Miettinen, Hidayet Aksu, Mauro Conti, Ahmad-Reza Sadeghi, and Selcuk Uluagac. 2020. Peek-a-boo: I see your smart home activities, even encrypted!. In *Proceedings of the 13th ACM Conference on Security and Privacy in Wireless and Mobile Networks*. 207–218.

[9] Fadel Adib, Hongzi Mao, Zachary Kabelac, Dina Katabi, and Robert C Miller. 2015. Smart homes that monitor breathing and heart rate. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 837–846.

[10] Mohammed A Almagboul, Feng Shu, Yuwen Qian, Xiaobo Zhou, Jin Wang, and Jinsong Hu. 2019. Atom search optimization algorithm based hybrid antenna array receive beamforming to control sidelobe level and steering the null. *AEU-International Journal of Electronics and Communications* 111 (2019), 152854.

[11] Ali Chelli, Muhammad Muaaz, and Matthias Pätzold. 2020. ActRec: A Wi-Fi-Based Human Activity Recognition System. In *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 1–6.

[12] Zhenghua Chen, Le Zhang, Chaoyang Jiang, Zhiguang Cao, and Wei Cui. 2018. WiFi CSI based passive human activity recognition using attention based BLSTM. *IEEE Transactions on Mobile Computing* 18, 11 (2018), 2714–2724.

[13] Israel Cohen, Yiteng Huang, Jingdong Chen, Jacob Benesty, Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. 2009. Pearson correlation coefficient. *Noise reduction in speech processing* (2009), 1–4.

[14] Qinhua Gao, Jie Wang, Xiaorui Ma, Xueyan Feng, and Hongyu Wang. 2017. CSI-based device-free wireless localization and activity recognition using radio image features. *IEEE Transactions on Vehicular Technology* 66, 11 (2017), 10346–10356.

[15] Linlin Guo, Lei Wang, Chuang Lin, Jialin Liu, Bingxian Lu, Jian Fang, Zhonghao Liu, Zeyang Shan, Jingwen Yang, and Silu Guo. 2019. Wiar: A public dataset for wifi-based activity recognition. *IEEE Access* 7 (2019), 154935–154945.

[16] Claus Hetting. 2022. Verizon Fios launches Wi-Fi sensing service powered by Origin. https://wifinowglobal.com/news-and-blog/verizon-fios-launches-wi-fi-sensing-service-powered-by-origin/.

[17] Jinyang Huang, Bin Liu, Hongxin Jin, and Zhiqiang Liu. 2018. WiAnti: An anti-interference activity recognition system based on WiFi CSI. In *2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*. IEEE, 58–65.

[18] Jinyang Huang, Bin Liu, Chenglin Miao, Yan Lu, Qijia Zheng, Yu Wu, Jianchun Liu, Lu Su, and Chang Wen Chen. 2021. Phaseanti: an anti-interference wifi-based activity recognition system using interference-independent phase component. *IEEE Transactions on Mobile Computing* (2021).

[19] Pei Huang, Xiaonan Zhang, Sihan Yu, and Linke Guo. 2021. Is-wars: Intelligent and stealthy adversarial attack to wi-fi-based human activity recognition systems. *IEEE Transactions on Dependable and Secure Computing* 19, 6 (2021), 3899–3912.

[20] Sijie Ji, Yaxiong Xie, and Mo Li. 2022. SiFall: Practical Online Fall Detection with RF Sensing. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*. 563–577.

[21] Wenjun Jiang, Chenglin Miao, Fenglong Ma, Shuochao Yao, Yaqing Wang, Ye Yuan, Hongfei Xue, Chen Song, Xin Ma, Dimitrios Koutsonikolas, et al. 2018.

Towards environment independent device free human activity recognition. In *Proceedings of the 24th annual international conference on mobile computing and networking*. 289–304.

[22] Hao Kong, Li Lu, Jiadi Yu, Yingying Chen, and Feilong Tang. 2020. Continuous authentication through finger gesture interaction for smart homes using WiFi. *IEEE Transactions on Mobile Computing* 20, 11 (2020), 3148–3162.

[23] Mengyuan Li, Yan Meng, Junyi Liu, Haojin Zhu, Xiaohui Liang, Yao Liu, and Na Ruan. 2016. When CSI meets public WiFi: inferring your mobile phone password via WiFi signals. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*. 1068–1079.

[24] Xinyi Li, Chao Feng, Fengyi Song, Chenghan Jiang, Yangfan Zhang, Ke Li, Xinyu Zhang, and Xiaojiang Chen. 2022. Protego: securing wireless communication via programmable metasurface. In *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*. 55–68.

[25] Xinyi Li, Chao Feng, Xiaojing Wang, Yangfan Zhang, Yaxiong Xie, and Xiaojiang Chen. 2023. RF-Bouncer: A Programmable Dual-band Metasurface for Sub-6 Wireless Networks. In *20th USENIX Symposium on Networked Systems Design and Implementation (NSDI 23)*. 389–404.

[26] Xinyi Li, Fengyi Song, Mina Luo, Kang Li, Liqiong Chang, Xiaojiang Chen, and Zheng Wang. 2023. Caring: Towards Collaborative and Cross-domain Wi-Fi Sensing. *IEEE Transactions on Mobile Computing* (2023).

[27] Xinyi Li, Gaoteng Zhao, Ling Chen, Xinyu Zhang, and Ju Ren. 2024. RFMagus: Programming the Radio Environment With Networked Metasurfaces. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*. 16–30.

[28] Hongbo Liu, Yan Wang, Jian Liu, Jie Yang, and Yingying Chen. 2014. Practical user authentication leveraging channel state information (CSI). In *Proceedings of the 9th ACM symposium on Information, computer and communications security*. 389–400.

[29] Jianwei Liu, Yinghui He, Chaowei Xiao, Jinsong Han, Le Cheng, and Kui Ren. 2022. Physical-World Attack towards WiFi-based Behavior Recognition. In *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*. IEEE, 400–409.

[30] Jian Liu, Yan Wang, Yingying Chen, Jie Yang, Xu Chen, and Jerry Cheng. 2015. Tracking vital signs during sleep leveraging off-the-shelf wifi. In *Proceedings of the 16th ACM international symposium on mobile ad hoc networking and computing*. 267–276.

[31] Yong Lu, Shaohe Lv, and Xiaodong Wang. 2019. Towards location independent gesture recognition with commodity wifi devices. *Electronics* 8, 10 (2019), 1069.

[32] Jun Luo, Hangcheng Cao, Hongbo Jiang, Yanbing Yang, and Zhe Chen. 2023. mimoCrypt: Multi-User Privacy-Preserving Wi-Fi Sensing via MIMO Encryption. In *2024 IEEE Symposium on Security and Privacy (SP)*. IEEE Computer Society, 25–25.

[33] Junyi Ma, Hao Wang, Daqing Zhang, Yasha Wang, and Yuxiang Wang. 2016. A survey on wi-fi based contactless activity recognition. In *2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld)*. IEEE, 1086–1091.

[34] Yongsen Ma, Gang Zhou, Shuangquan Wang, Hongyang Zhao, and Woosub Jung. 2018. SignFi: Sign language recognition using WiFi. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–21.

[35] Vuong V Mai, Truong C Thang, and Anh T Pham. 2017. CSMA/CA-based uplink MAC protocol design and analysis for hybrid VLC/Wifi networks. In *2017 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 457–462.

[36] Takashi Nakamura, Mondher Bouazizi, Kohei Yamamoto, and Tomoaki Ohtsuki. 2022. Wi-Fi-based fall detection using spectrogram image of channel state information. *IEEE Internet of Things Journal* 9, 18 (2022), 17220–17234.

[37] Sameera Palipana, David Rojas, Piyush Agrawal, and Dirk Pesch. 2018. FallDeFi: Ubiquitous fall detection using commodity Wi-Fi devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–25.

[38] Jeong-Ki Park, Jae-Hyun Park, and Kyung-Tae Kim. 2023. Multipath signal mitigation for indoor localization based on mimo fmcw radar system. *IEEE Internet of Things Journal* (2023).

[39] Hossein Pirayesh and Huacheng Zeng. 2022. Jamming attacks and anti-jamming strategies in wireless networks: A comprehensive survey. *IEEE communications surveys & tutorials* 24, 2 (2022), 767–809.

[40] Prerna Saxena and Ashwin Kothari. 2016. Ant lion optimization algorithm to control side lobe level and null depths in linear antenna arrays. *AEU-International Journal of Electronics and Communications* 70, 9 (2016), 1339–1349.

[41] Souvik Sen, Jeongkeun Lee, Kyu-Han Kim, and Paul Congdon. 2013. Avoiding multipath to revive inbuilding WiFi localization. In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*. 249–262.

[42] Cong Shi, Jian Liu, Hongbo Liu, and Yingying Chen. 2017. Smart user authentication through actuation of daily activities leveraging WiFi-enabled IoT. In *Proceedings of the 18th ACM international symposium on mobile ad hoc networking and computing*. 1–10.

[43] Paul Staat, Simon Mulzer, Stefan Roth, Veelasha Moonsamy, Markus Heinrichs, Rainer Kronberger, Aydin Sezgin, and Christof Paar. 2022. IRShield: A countermeasure against adversarial physical-layer wireless sensing. In *2022 IEEE Symposium on Security and Privacy (SP)*. IEEE, 1705–1721.

[44] William C Stone. 1997. Electromagnetic signal attenuation in construction materials. (1997).

[45] Yonglong Tian, Guang-He Lee, Hao He, Chen-Yu Hsu, and Dina Katabi. 2018. RF-based fall monitoring using convolutional neural networks. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–24.

[46] Philip van Dorp and FCA Groen. 2008. Feature-based human motion parameter estimation with radar. *IET Radar, Sonar & Navigation* 2, 2 (2008), 135–145.

[47] Mathy Vanhoef and Frank Piessens. 2014. Advanced Wi-Fi attacks using commodity hardware. In *Proceedings of the 30th Annual Computer Security Applications Conference*. 256–265.

[48] Hao Wang, Daqing Zhang, Yasha Wang, Junyi Ma, Yuxiang Wang, and Shengjie Li. 2016. RT-Fall: A real-time and contactless fall detection system with commodity WiFi devices. *IEEE Transactions on Mobile Computing* 16, 2 (2016), 511–526.

[49] Wei Wang, Alex X Liu, Muhammad Shahzad, Kang Ling, and Sanglu Lu. 2015. Understanding and modeling of wifi signal based human activity recognition. In *Proceedings of the 21st annual international conference on mobile computing and networking*. 65–76.

[50] Wei Wang, Alex X Liu, Muhammad Shahzad, Kang Ling, and Sanglu Lu. 2017. Device-free human activity recognition using commercial WiFi devices. *IEEE Journal on Selected Areas in Communications* 35, 5 (2017), 1118–1131.

[51] Xuanzhi Wang, Kai Niu, Jie Xiong, Bochong Qian, Zhiyun Yao, Tairong Lou, and Daqing Zhang. 2022. Placement matters: Understanding the effects of device placement for WiFi sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 1 (2022), 1–25.

[52] Yuxi Wang, Kaishun Wu, and Lionel M Ni. 2016. Wifall: Device-free fall detection by wireless networks. *IEEE Transactions on Mobile Computing* 16, 2 (2016), 581–594.

[53] Deidre Wild, US Nayak, and B Isaacs. 1981. How dangerous are falls in old people at home? *Br Med J (Clin Res Ed)* 282, 6260 (1981), 266–268.

[54] Tuo Xie, Hanjun Jiang, Xijin Zhao, and Chun Zhang. 2019. A Wi-Fi-Based wireless indoor position sensing system with multipath interference mitigation. *Sensors* 19, 18 (2019), 3983.

[55] Jianfei Yang, Xinyan Chen, Han Zou, Dazhuo Wang, Qianwen Xu, and Lihua Xie. 2022. EfficientFi: Toward Large-Scale Lightweight WiFi Sensing via CSI Compression. *IEEE Internet of Things Journal* 9, 15 (Aug. 2022), 13086–13095. https://doi.org/10.1109/jiot.2021.3139958

[56] Abdulsalam Yassine, Shailendra Singh, and Atif Alamri. 2017. Mining human activity patterns from smart home big data for health care applications. *IEEE Access* 5 (2017), 13131–13141.

[57] Siamak Yousefi, Hirokazu Narui, Sankalp Dayal, Stefano Ermon, and Shahrokh Valaee. 2017. A survey on behavior recognition using WiFi channel state information. *IEEE Communications Magazine* 55, 10 (2017), 98–104.

[58] Youwei Zeng, Dan Wu, Jie Xiong, Jinyi Liu, Zhaopeng Liu, and Daqing Zhang. 2020. MultiSense: Enabling multi-person respiration sensing with commodity wifi. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3 (2020), 1–29.

[59] Youwei Zeng, Dan Wu, Jie Xiong, Enze Yi, Ruiyang Gao, and Daqing Zhang. 2019. FarSense: Pushing the range limit of WiFi-based respiration sensing with CSI ratio of two antennas. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–26.

[60] Yi Zhang, Yue Zheng, Kun Qian, Guidong Zhang, Yunhao Liu, Chenshu Wu, and Zheng Yang. 2021. Widar3. 0: Zero-effort cross-domain gesture recognition with wi-fi. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 11 (2021), 8671–8688.

[61] Yue Zheng, Yi Zhang, Kun Qian, Guidong Zhang, Yunhao Liu, Chenshu Wu, and Zheng Yang. 2019. Zero-effort cross-domain gesture recognition with Wi-Fi. In *Proceedings of the 17th annual international conference on mobile systems, applications, and services*. 313–325.

[62] Yuxuan Zhou, Huangxun Chen, Chenyu Huang, and Qian Zhang. 2022. WiADv: Practical and robust adversarial attack against WiFi-based gesture recognition system. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (2022), 1–25.

[63] Zimu Zhou, Chenshu Wu, Zheng Yang, and Yunhao Liu. 2015. Sensorless sensing with WiFi. *Tsinghua Science and Technology* 20, 1 (2015), 1–6.