RIStealth: Practical and Covert Physical-Layer Attack against WiFi-based Intrusion Detection via Reconfigurable Intelligent Surface

Yuxuan Zhou[1], Chenggao Li[1], Huangxun Chen[2,3], Qian Zhang[1]
[1]The Hong Kong University of Science and Technology, Hong Kong SAR, China
[2]The Hong Kong University of Science and Technology (Guangzhou), Guangdong, China
[3] Huawei, Hong Kong SAR, China
yzhoudo@connect.ust.hk,chenggao.li@connect.ust.hk,huangxunchen@hkust-gz.edu.cnk,qianzh@cse.ust.hk

ABSTRACT

The emerging reconfigurable intelligent surface (RIS) technique introduces novel threats to wireless sensing owing to its channel customization ability. Unlike active radios, the RIS's interference behaves akin to natural reflections, exhibiting a higher level of stealthiness and difficulty in detection. However, the majority of current RIS-based attacks lack generalizability to real-world scenarios, as they assume complete coverage of the RIS over objects and develop their techniques within electromagnetic-controlled environments such as an anechoic chamber. To bridge this gap, we present RIStealth, a practical and covert attack that leverages RIS technology to render a moving individual undetectable by WiFi-based intrusion detection systems in real-life scenarios. RIStealth integrates the strengths of both motion reduction and threshold lifting strategies to address challenges of limited RIS affordability, constrained cooperation in adversary settings, and complex and unpredictable environments. Through real-world evaluations conducted with our RIS prototype, we demonstrate that RIStealth effectively reduces the victim's intrusion detection rate from 95.1% to 16.4%. Our findings shed light on the practical threats posed by RIS, thereby encouraging further countermeasure development.

CCS CONCEPTS

• **Security and privacy → Domain-specific security and privacy architectures**.

KEYWORDS

Intrusion detection, Wireless sensing, Physical-layer attack, Reconfigurable intelligent surface

ACM Reference Format:
Yuxuan Zhou[1], Chenggao Li[1], Huangxun Chen[2,3], Qian Zhang[1]. 2023. RIStealth: Practical and Covert Physical-Layer Attack against WiFi-based Intrusion Detection via Reconfigurable Intelligent Surface . In *The 21st ACM Conference on Embedded Networked Sensor Systems (SenSys '23), November 12–17, 2023, Istanbul, Turkiye.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3625687.3625790

Huangxun Chen participated in this research when she was a Researcher at Huawei. She now joins HKUST (GZ) as an Assistant Professor.

SenSys '23, November 12–17, 2023, Istanbul, Turkiye

ACM ISBN 979-8-4007-0414-7/23/11…$15.00
https://doi.org/10.1145/3625687.3625790

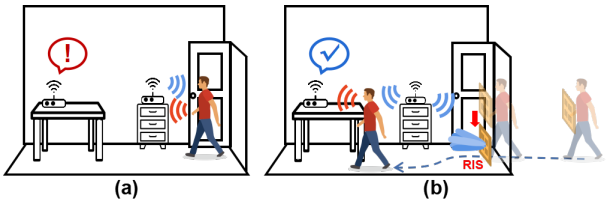

Figure 1: The RIS may bring covert threats to the wireless sensing foundation.

1 INTRODUCTION

WiFi-based indoor intrusion detection [13, 24, 25, 28, 49, 52] has been extensively studied as one of the important applications in wireless sensing as depicted in Figure 1(a). Technically, WiFi sensing extracts valuable information about the state of target objects through channel state information (CSI), which encompasses the reflection, scattering, and attenuation caused by these objects. Compared with camera-based solutions, WiFi-based detection alleviates privacy concerns to a greater extent and operates consistently under diverse lighting conditions, including both day and night. This technology has now become relatively mature, and some commercial products, such as Hex Home[35] and Linksys Aware [29], have been manufactured and deployed in real-world scenarios.

Along with the technical advancements in wireless sensing, relevant attack and defense studies also alternately emerge and revolve around the open nature of the wireless medium. Jamming attacks [22, 37, 47, 50], the most intuitive ones, actively generate signals to obstruct the sensing functionalities. However, their defense schemes are also straightforward via detecting abnormal energy fluctuation and power analysis [2, 21]. Replay attacks [16, 46] are more advanced to alter the state, *i.e.*, amplitude and phase of legitimate transmitted signals to deceive the sensing system. However, randomized pause-and-detect methods could detect and mitigate replay attacks effectively.

Despite previous defense endeavors, we have observed that the emergence of reconfigurable intelligent surface (RIS) technology [7, 9, 10, 12, 15, 36] introduces a new breed of covert threats. RIS features massive low-cost, passive, and reconfigurable electromagnetic reflection units. From an adversarial standpoint, RIS could

manipulate the channel to compromise sensing. More importantly, passive RIS is less detectable compared with active radios because it can emulate the behavior of benign natural reflectors. Previous defense schemes, such as power analysis [2, 21] and pause-and-detect methods, cannot distinguish RIS reflections. While radios equipped with full-duplex antenna arrays [55] can emulate the effects of RIS, they typically incur higher costs.

It is noted that certain recent studies have explored the RIS to render objects invisible to radar systems, but most existing works assume relatively ideal working settings. Cloaking-based techniques [39, 53, 54] are studied in anechoic chambers and assume a single incident signal with clear direction. However, massive, unpredictable reflections in real indoor scenarios will make them collapse in practice. Radar cross-section (RCS) reduction methods [11, 17, 51] focus on relatively static scenarios, *i.e.*, static RIS employs specific configuration to minimize visibility under radar detection. However, the feasibility of maintaining invisibility while in motion remains largely unexplored, posing a potentially more significant threat in practical scenarios. Consequently, the practical feasibility of RIS-aided attacks against wireless sensing urges more research endeavors.

To fill this gap, we investigate the threat model of leveraging RIS to render a moving intruder undetectable by WiFi-based intrusion detection systems in real-world scenarios, as illustrated in Figure 1(b). Specifically, we design RIStealth, a RIS-aided practical and covert attack to achieve our attack goal, and mainly address three practical challenges (C1-C3).

Limited Affordability of Practical RIS (C1): Affordability mainly refers to the size and capacity of RIS. In practical attacks, RIS should be handbag-sized to maintain stealthiness and mobility. This constraint naturally excludes those methods requiring fully covering objects with RIS [17, 39, 53, 54]. Additionally, the limited size also restricts the available manipulation space, *e.g.*, the amount of manipulated energy is highly correlated to the RIS size.

Restrained Cooperation in Adversary Setting (C2): Configuring RIS in benign settings generally has assistance from legitimate transceivers. For example, for boosting communication, RIS is proactively configured based on receivers' CSI to increase signal quality and throughput [1, 14]. For defending malicious motion sensing, [45] deploys RIS along with the benign transmitter to control more energy for protection. However, benign transceivers will not assist RIS in our attack setting, The attacker's RIS will be initially far away from the victim system and should rely on himself to derive deployment details, *e.g.*, the location of transceivers if needed. It may have estimation errors due to the non-cooperation of benign transceivers, *e.g.*, sparse traffic from the receiver makes localization harder.

Complex and Unpredictable Environment (C3): The area being intruded into may be a complicated environment full of unknown reflectors, *e.g.*, furniture and metal devices. Configuring RIS properly to render a moving intruder invisible involves not only disturbing channels but also effectively concealing natural disturbances caused by human motion. Thus, the attacker should consider the complex interaction between the RIS device, legitimate transceivers, surrounding objects, and the moving person, and also their dynamics with their changing relative locations. This prevents direct use of [39] requiring precise incident angles for configuring RIS. In addition, channel obfuscation methods [45] do not fit because obfuscation without careful design may either be filtered or accidentally trigger alarms of sensing system, failing to achieve our covert attack objective.

To address these practical challenges, RIStealth incorporates a delicate attacking scheme that takes these factors into account. Our key insight is rooted in the underlying mechanism of state-of-the-art WiFi-based intrusion detection [13, 49, 52]. It is noted that robustness is an essential consideration in practical WiFi sensing systems. Intrusion detection aims to achieve both a high detection rate and a low false alarm rate. Thus, these systems typically employ adaptive schemes that can adjust to varying levels of interference caused by ambient noise, transceiver noise, and disturbances from irrelevant objects in a real-world deployment. Unfortunately, our research reveals that the RIS can strategically manipulate the environmental interference to desensitize the victim sensing system and alter its detection outcomes. Highlights of our original contributions are as follows:

- Firstly, we fully consider the RIS's different channel manipulation capacities under various distances to benign transceivers (**C1**) and proactively schedule the attack in two phases accordingly. In the distant field, our strategy aims to keep the motion-induced disturbance below the alert threshold (Sneaking Phase), while in the near field, our strategy aims to covertly lift the alert threshold so that the intruder's afterward movement will not trigger the alarm (Radio Blast Phase). The former strategy helps bring RIS into the effective range of the latter one. They work together to enable a physically achievable attack.
- Secondly, to realize the goal of Sneaking Phase, we develop RIS-aided beamforming to deflect the reflection energy away from the victim receiver. Specifically, we direct the beam towards the ground to mitigate unexpected multipath reflection hitting at the receiver (**C3**). We also design a beam broadening method to tolerate the localization error of the transmitter caused by the intruder's walking movement and device imperfection (**C2**).
- Thirdly, to realize the goal of Radio Blast Phase, we integrate the RIS's phase shift reconfigurability with RIS-aided beamforming to covertly inject disturbance to the receiver. Specifically, we design a sub-array-based spiral manner to activate RIS elements progressively, imitating the increase of normal noise to covertly lift the alert threshold (**C1**). We also design RIS-aided AoA estimation to help increase the error tolerance of receiver localization (**C2**).

We implement a RIS prototype and conduct extensive experiments to evaluate RIStealth against the state-of-the-art WiFi-based intrusion detection system. RIStealth lowers the intrusion detection rate from 95.1% to about 16.4% in practical settings. Detailed evaluations also demonstrate the effectiveness and robustness of RIStealth.

2 PRELIMINARIES

2.1 Wi-Fi Sensing

In typical Wi-Fi sensing, object movements could manifest their effects on the propagation channel between a pair of transceivers. Thus, we could exploit channel state information (CSI) to infer surrounding dynamics. CSI represents the amplitude attenuation and phase change of Wi-Fi signals caused by the environment. The

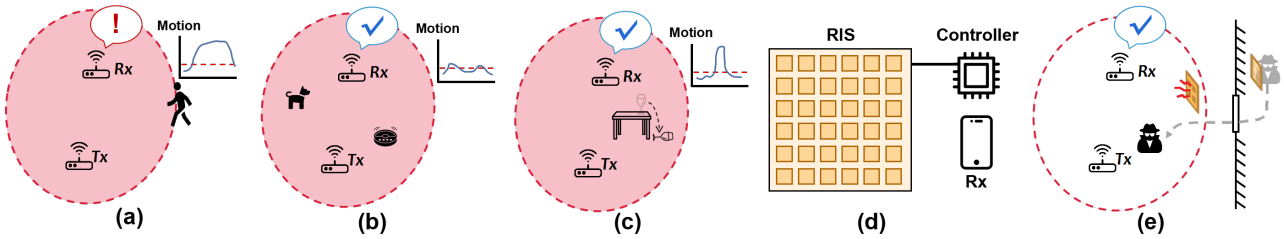

Figure 2: Threat model of RIStealth. (a) shows that the system detects the intruder as soon as he enters the monitoring area (high detection rate); (b)-(c) shows that the system ignores both the transitory movement like a dropping object and the movement of small objects like pets (low false alarm rate). (d) shows the attacker's equipment; (e) shows the attack goal.

CSI arriving at time t and at frequency f can be represented as follows:

$$H(f,t) = \sum_{i=1}^{L} \alpha_l(f,t) e^{-j2\pi f \tau_l(f,t)} \tag{1}$$

where L is the number of multi-paths, α_l and τ_l are the complex attenuation and propagation delay of the l-th path. The CSI could be divided into two components as follows:

$$H(f,t) = H_s(f,t) + H_d(f,t) \tag{2}$$

The static component $H_s(f,t)$ usually represents the signal reflected from the static environment or propagating through the direct path, which often dominates the CSI, while the dynamic component $H_d(f,t)$ represents the distortion induced by surrounding moving objects. Many Wi-Fi sensing applications, *e.g.*, intrusion detection, extract the dynamic components $H_d(f,t)$ to distill motion-relevant features, *e.g.*, the Doppler frequency shift, signal standard deviation, *etc*, to detect the target object.

2.2 Reconfigurable Intelligent Surface

Reconfigurable Intelligent Surface (RIS). A RIS device consists of an array of reflection units, each of which can independently alter the incident signal in terms of phase, amplitude, frequency, or even polarization [4]. The RIS considered in our work only alters the phase of incident signals hitting the surface, *i.e.*, a low-power passive mode without a power amplifier. Under the impact of RIS, the CSI measured by the receiver can be re-written as follows:

$$H'(f,t) = H(f,t) + \sum_{m=1}^{M} H_{RIS,m}(f,t) e^{-j\phi_m} \tag{3}$$

where $H(f,t)$ denotes the channel state without RIS impact as in Equation 2, M denotes the number of RIS elements, $H_{RIS,m}$ denotes the native reflection from the m-th element, and ϕ_m denotes the additional phase shift deliberately imposed by the m-th element.

RIS-aided Beamforming. The passive reflection mode and the array layout of RIS make it intrinsically fit for beamforming. Similar to directional antennas [31, 32, 48], beamforming is to coordinate additional phase shifts of adjacent RIS elements to superpose constructively at the target direction. Both directions of the transmitter and receiver relative to the RIS device are required to enable a precise beamforming effect. The former information is used to compensate for the phase difference of incident signals arriving at multiple RIS elements, *i.e.*, enabling the transmitter-side beam. The latter information is used to calculate the required phase shifts of each RIS element to ensure the constructive signal superposition at the receiver, *i.e.*, enabling the receiver-side beam. For example, if we consider only two elements in a 1-D array layout, the additional phase shift of the second element relative to the first is as follows:

$$\begin{aligned} \Delta\phi =& \phi_{txdir} + \phi_{rxdir} \\ =& -2\pi \frac{d_{tx,ele1} - d_{tx,ele2}}{\lambda} + 2\pi \frac{d_{rx,ele1} - d_{rx,ele2}}{\lambda} \end{aligned} \tag{4}$$

where ϕ_{txdir} and ϕ_{rxdir} denote the phase difference because of the transmitter direction and the receiver direction, respectively, and $d_{x,y}$ denotes the distance between x and y. For a digital RIS device, the phase shift actually applied will be rounded to a valid value within the phase shift space. A 1-D array of elements could only enable 2-D plane beamforming, while a 2-D array could enable beamforming in 3-D space.

3 THREAT MODEL

Here we illustrate the overall setup of the attack scenario targeted by RIStealth and the capability of the attacker.

Attack Scenario. RIStealth aims to render a moving person invisible against a WiFi-based intrusion detection system. A pair of transceivers are deployed to detect any intruder stepping into the protected area. In normal functioning mode, the intrusion detection system will report any human presence as shown in Figure 2(a) (high detection rate), but ignore the motion of the pet, robot, curtain, and *etc*, as shown in Figure 2(b)-(c) (low false alarm rate). The goal of the attacker is to intrude into the monitoring area without being detected (Figure 2(e)). Once the intruder can get inside the protected area, he can cut off the system power cable, steal the valuables, *etc.*

Attacker Setup. As shown in Figure 2(e), the attacker has no prior knowledge about the deployment of the intrusion detection system, including the precise location of the transmitter and the receiver, which means he has to manage to obtain the necessary information by himself. The attack does not get the detailed parameters (*e.g.*threshold) of the victim systems. The attacker is equipped with a passive RIS and an auxiliary off-the-shelf receiver (*e.g.*, a smartphone) that can acquire the CSI information, as in Figure 2(d). The RIS has 16×16 elements and is able to impose an extra phase shift on the reflected signal. The phase shift function supports 4 phase shifts (0, $\frac{\pi}{2}$, π, $\frac{3\pi}{2}$), configured by a controller, *e.g.*, an FPGA board or a laptop.

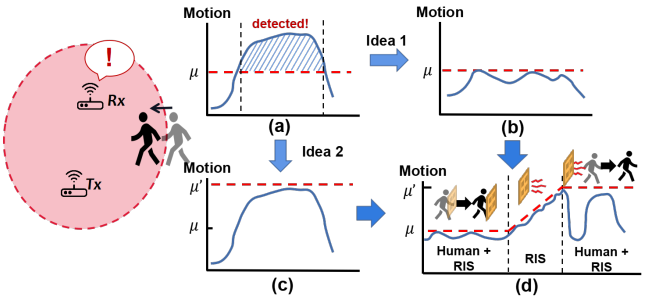

Figure 3: Key insight for RIStealth. RIStealth integrates the strengths of both motion reduction and threshold lifting strategies to launch the attack.

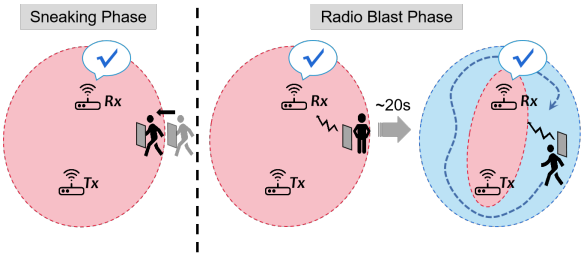

Figure 4: Overview of RIStealth. Sneaking Phase is to make the intruder holding the RIS approach the protected area sneakily to bring the RIS closer to the victim transceiver to facilitate the Radio Blast Phase. In Radio Blast Phase, the intruder puts down the RIS and then uses it to raise the victim system's motion detection threshold so that the intruder's afterward movement will not trigger the alarm.

4 KEY INSIGHT OF RISTEALTH.

In this section, we elaborate on RIStealth's key insight.

WiFi-based Intrusion Detection. The basic principle of WiFi-based intrusion detection is illustrated in Figure 3(a). The intrusion warning will be triggered when the human motion signal exceeds the threshold value of the detection system, *i.e.*, the blue line exceeds the red dotted line.

Adaptive-threshold Scheme. The quality of motion signal will vary across different environments and even different times of the same environment with the same device [5]. Thus, to combat such fluctuation, researchers have designed adaptive-threshold schemes [25, 28, 52, 57] to maintain a high detection rate and low false alarm rate in practice (Figure 2(a)-(c)). Without loss of generality, the detection criterion of the adaptive-threshold scheme is described as follows:

$$\sigma_t > \delta'_{mov} = \delta_{mov} \times f(\sigma^{st}_{t'})(t' < t) \tag{5}$$

σ_t denotes motion feature signal at t-th processing window. δ'_{mov} denotes the adaptive threshold, which is equal to a basic threshold δ_{mov} multiplying a function of the signal of the previous static period $f(\sigma^{st}_{t'})$. In this way, the threshold adapts itself to the varying noise level for robustness.

Attack Basis of RIStealth. In order to make the intruder invisible, the basic idea is to keep the blue line below the red dotted line as shown in Figure 3(b)-(c). Figure 3(b) is to suppress the blue line, *i.e.*, reduce the reflection of the intruder to prevent it from exceeding the threshold (motion reduction strategy). While Figure 3(c) is to lift the red dotted line, *i.e.*, raise the threshold to clear the alarm (threshold lifting strategy).

The threshold-lifting strategy seems like a once-and-for-all choice for hiding the intruder. As shown in Equation 5, the coefficient multiplier of the motion threshold δ'_{mov} is determined by a function of the signal of the latest 'no motion' duration. Given the channel customization capability of RIS, it is possible to strategically lift the next δ'_{mov} by increasing σ gradually yet remaining below the current δ'_{mov}. However, through our experiment, it is found that RIS should be close to the transceivers and control relatively high energy towards the receiver, *e.g.*, by beamforming, to realize the threshold lifting strategy successfully.

Therefore, we further introduce a motion reduction strategy to help bring the RIS into the effective working range of the threshold lifting strategy. When the intruder is relatively away from the transceivers, the reflection energy controlled by a handbag-sized RIS could be comparable with that induced by uncovered body parts. Thus, RIS could steer its energy away from the receiver to realize the motion reduction strategy successfully. As shown in Figure 3(d) and Figure 4, RIStealth integrates the best of both strategies into a dual-phase attack scheme to make the intruder invisible.

5 RISTEALTH DESIGN

5.1 Attack Scheme Overview

As illustrated in Figure 4, RIStealth consists of two major phases, Sneaking Phase and Radio Blast Phase.

The goal of the Sneaking Phase is to make the intruder (holding the RIS) approach the protected area sneakily and reach a pre-specified position close to the victim transceiver to facilitate the Radio Blast Phase, and the goal of the Radio Blast Phase is to lift the threshold of the victim system to make himself invisible. The process would be as follows:

(1) Before the attack, the attacker must determine the RIS's effective working range while planning the scheme.

(2) To launch the attack, the attacker arrives near the victim's home and estimates the victim system's location. By sniffing signals from the transmitter (which continuously emits signals) and receiver (which occasionally emits signals to report the detection results) from different locations, the attacker can approximate the transmitter and receiver's positions using some simple RSS-based methods [57].

(3) Knowing the transceivers' rough location, the house entry, and the RIS's effective range, the attacker estimates the switching point from the Sneaking Phase to the Radio Blast Phase. For example, if the entry is 8 meters from the transceivers and the RIS's range is 6 meters, the attacker plans to intrude at least 2 meters and use simple self-localization methods like the inertial measurement unit on the smartphone to sense the intrusion distance.

(4) The attacker initiates the Sneaking Phase to avoid detection by redirecting energy towards the ground. Upon reaching the

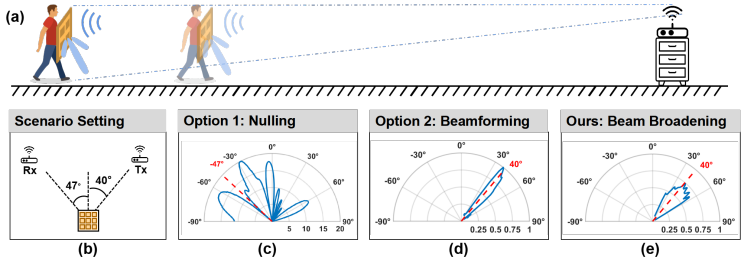

Figure 5: RIS-aided motion reduction strategy. (a) The attack method of Sneaking Phase; (b) The attack setting for simulation; (c) The results of the nulling strategy; (d) The results of the strategy; (e) The results of our strategy.

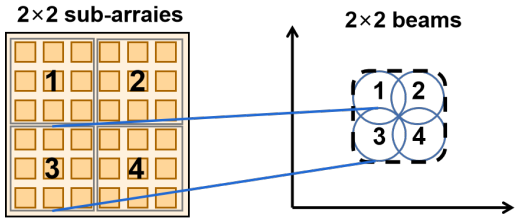

Figure 6: RIS-aided beam-broadening beamforming. The beam broadening operation could be applied to enable either transmitter-side wide beam or receiver-side wide beam, but not both simultaneously.

switching point, the attacker carefully places the RIS (to avoid being detected) and starts the Radio Blast Phase.

(5) In the Radio Blast Phase, the attacker localizes the victim system's transmitter accurately using RIS-aided AoA estimation. The transmitter is localized since its continuous emitting enables our scanning method. Accurate localization is required since two-way beam broadening isn't possible for our scheme.

(6) Finally, the attacker launches the Radio Blast Phase, using the RIS to focus energy and create artificial noise to increase the detection threshold. With a high enough threshold, the attacker can move freely in the monitored area and commit crimes.

In the following, we elaborate on the detailed design in two phases for addressing their technical challenges respectively.

5.2 Sneaking Phase Design

As shown in Figure 5(a), the intruder is initially away from the receiver. Thus, RIS could dominate the reflection from the intruder to the receiver. Two candidate strategies come into our mind: nulling and beamforming.

5.2.1 Option 1: RIS-aided Nulling Strategy. Nulling aims to enforce specific RIS configurations so that the RIS signals can destructively cancel out each other at the receiver. It can be implemented by making reflection signals of a group of adjacent elements adding up to zero. For the setting shown in Figure 5(b), we emulate the nulling strategy to reduce the reflection energy towards the receiver. It seems feasible in principle. However, we find that the directions adjacent to the receiver direction have high energy (Figure 5(c)). It implies that our estimation for the receiver direction should be extremely accurate, and the intruder should maintain the pointing direction stably on the move to avoid high-energy sidelobe accidentally hitting the receiver. Thus, the nulling strategy will collapse in practice.

5.2.2 Option 2: RIS-aided Beamforming Strategy. The basic idea of beamforming is to direct most energy towards a direction other than the receiver direction, *i.e.*, equivalently reducing the number of reflection signals at the receiver. Compared with relatively scattered energy in the nulling strategy, the beamforming strategy can make the energy more controllable towards a specific direction (Figure 5(d)) to avoid unanticipated yet powerful reflections in a complicated environment full of strong reflectors. In our scenario, it is appropriate to enforce beamforming towards the ground as shown in Figure 5(a) because it is normally unoccupied and predictable to be less likely to produce a significant reflection toward the receiver. In addition, some intrusion detection systems [28] leverage height information to differentiate the adult intruder, the pet, and the infant. Thus, the energy beamformed towards the ground, together with the normal reflection from the intruder's legs, could be misidentified as the pet and ignored.

Revisiting the RIS-aided beamforming principle shown in Equation 4, we conclude that it requires an accurate estimation of both the incident signal direction and the target direction to enable precise beamforming. Otherwise, unpredictable scattered energy may appear to cause a strong reflection and trigger the alarm. In this phase, we choose the ground as the target direction to greatly relax the requirement on its direction estimation. For estimating incident signal direction, *i.e.*, the transmitter's direction, the intruder could apply RSS-based (received signal strength) localization approach [57] with its auxiliary receiver, *e.g.*, a smartphone to collect the channel information of transmitter's packets to derive its direction. However, being relatively away from the victim transmitter brings non-negligible estimation error. Based on Equation 4, the width of the transmitter-side beam is the same as the receiver-side beam. As the narrow beam is shown in Figure 5(d), a little estimation error in the transmitter's direction could break the desired beamforming effect, especially for acquiring and updating the direction during the process of approaching victim transceivers.

5.2.3 Ours: RIS-aided Beam-Broadening Beamforming. Based on the above analysis, a wider transmitter-side beam could be more favorable for the attacker. Because it can increase the error tolerance on estimating the transmitter's direction relative to the RIS and also flatten the potential peak energy as shown in Figure 5(e). Inspired by the existing beam-broadening methods on phased antenna array [23, 41], we design a beam-broadening technique to enhance the original RIS-aided beamforming strategy. Specifically, we could divide the RIS elements into equal-sized sub-arrays and then configure these sub-arrays to concentrate on the adjacent directions to synthesize a wider beam (to either transmitter-side or receiver-side) as illustrated in Figure 6.

Technically, all RIS reflecting elements are grouped into $N_s \times N_s$ sub-arrays. Each sub-array will be assigned and configured with

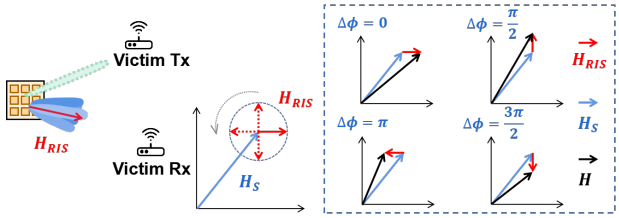

Figure 7: Beamforming-enabled variation increasing. The additional phases are imposed upon the signals periodically to introduce more signal variation at the receiver.

a corresponding beamforming angle towards the same selected direction (either the transmitter's or receiver's direction). The RIS center is associated with the center direction of the synthesized wide beam, the elevation and azimuth angles of which are denoted as (Θ_0, Φ_0). The beamforming angle of each sub-array is determined as follows:

$$\begin{cases} \Theta_{i,j} = \Theta_0 + \frac{i-(N_s-1)}{2}\Delta_{angle}, i = 0, 1, ..., N_s \\ \Phi_{i,j} = \Phi_0 + \frac{j-(N_s-1)}{2}\Delta_{angle}, j = 0, 1, ..., N_s \end{cases} \quad (6)$$

where i and j denote the row index and column index of the sub-array, respectively, Δ_{angle} denotes the pre-defined interval between adjacent beams. Figure 5(d) and Figure 5(e) share the same setting except that the former uses RIS-aided beamforming and the latter uses the beam-broadening version with $N_s = 4$ and $\Delta_{angle} = 4°$. Our approach broadens the beam effectively to strengthen the resilience towards the estimation error of the transmitter's direction. A 30° beam in Figure 5(e) can tolerate around ± 1.2 meters localization error at 4 meters far way.

5.3 Radio Blast Phase Design

In Sneaking Phase, we direct reflection energy away from the receiver, while in Radio Blast Phase, we should bring reflection energy towards the receiver and further introduce some disturbance to strategically influence the signal variation for threshold lifting. Sneaking Phase helps bring RIS closer to the receiver, but two problems should be addressed to achieve the attack goal of Radio Blast Phase: (1) what is the effective and physically-realizable approach to increase signal variation, *i.e.*, the basics of threshold lifting; (2) how to increase signal variation covertly, *i.e.*, not be detected as a motion-induced variation so as to lift the threshold in Equation 5 successfully. In the following, we introduce our design considerations for these two issues respectively.

5.3.1 Beamforming-enabled Variation Increasing. To increase the signal variation at the receiver, we make full use of two capabilities of the RIS: the energy-focusing capability of the RIS-aided beamforming and the phase shift reconfigurability of the RIS, $(0, \frac{\pi}{2}, \pi, \frac{3\pi}{2})$ for a 2-bit RIS. As shown in Figure 7, we can first assign the RIS with a beamforming configuration to focus reflection energy on the receiver. Then, we further impose an extra phase shift upon the basic beamforming configuration to introduce disturbance, *i.e.*, more signal variation at the receiver. Specifically, we simultaneously increment the phases of all RIS elements with the same extra shift in the circle of $(0, \frac{\pi}{2}, \pi, \frac{3\pi}{2})$ as shown in Figure 7 so as to satisfy both variations increasing and designated beamforming direction preservation.

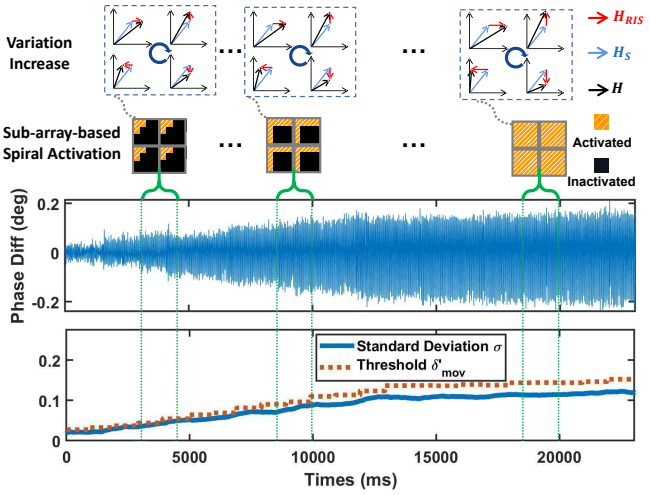

Figure 8: Covert threshold lifting. For a certain intermediate time period denoted between green dotted lines, a group of phase shifts is added circularly to increase signal variation. Over the whole period of time, the number of activated RIS elements gradually increases to covertly lift the threshold.

5.3.2 Covert Threshold Lifting Strategy. Signal variation increasing itself is not sufficient to trigger threshold lifting. As shown in Equation 5, the motion detection threshold of the victim system is only updated by the variation of the last 'no motion' duration/processing window. Thus, if we increase the variation sharply, the extra variation will be detected as motion-induced variation, failing to trigger threshold adapting in Equation 5. The attacker should increase the signal variation steadily and covertly to lift the threshold gradually. To this end, instead of exploiting the whole RIS to increase signal variation, we choose to amortize the amount of variation brought by all elements over a period of time to keep the variation-increasing procedure covert and undetected.

RIS Configuration Strategy. The amount of variation is highly correlated with the number of involved RIS elements. Therefore, we design a progressive configuration strategy to activate RIS elements in batches over a period of time. For a certain intermediate time period in Figure 8 (assuming a RIS with $N_s \times N_s$ sub-arrays and $N_s = 2$), only partial RIS elements, *i.e.*, orange ones with diagonal stripe, will be activated and assigned the basic phases for beamforming and an extra phase shift circulating within $(0, \frac{\pi}{2}, \pi, \frac{3\pi}{2})$. The cycle time is at a millisecond level. In other words, a RIS with partially activated elements focuses a part of upper bound beamforming energy towards the receiver and introduces extra disturbance by rapidly rotating the RIS-induced channel response H_{RIS} to covertly lift the threshold. When proceeding to the next duration, we increase the number of activated RIS elements to involve more reflection energy. The handover interval between different activation configurations in the second level should be longer than the processing window size of intrusion detection. In this way, we gradually enhance the upper bound capability of signal variation increasing in order to lift the threshold progressively.

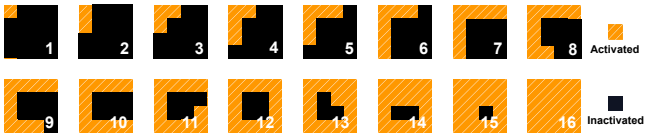

Figure 9: Illustration of sub-array-based spiral activation strategy with $M_s = 4$. The elements are activated in a spiral manner.

RIS Element Activation Sequence. The activation sequence of RIS elements is very critical to enable covert threshold lifting. An intuitive manner is to activate the RIS elements in a square manner, *i.e.*, activating squares of elements in batches as shown in Figure 14(c). However, these schemes have two drawbacks. Firstly, they may result in abrupt variation change once some elements that contribute significantly to beamforming are turned on at one moment [26]. Secondly, they can not maintain a wide beam for most moments to guarantee a lasting effect on the receiver, considering we only have a coarse receiver direction. Thus, we propose a sub-array-based spiral activation manner. Similar to Sneaking Phase, all RIS elements are divided into $N_s \times N_s$ sub-arrays to enable beam broadening. For each sub-array with the size of $M_s \times M_s$, we activate its elements in the spiral manner as shown in Figure 9. Specifically, we apply a mask on every sub-array. The mask is a $M_s \times M_s$ matrix consisting of '0' and '1'. The elements corresponding to the '0' will remain idle, while the elements corresponding to the '1' will be activated. Figure 9 depicts the two stages of the sub-array-based spiral activation scheme.

***Stage 1*:** From the first to the seventh pattern, a whole row and a column of elements are gradually activated. In principle, a row and a column of elements enable 3-D beamforming. Thus, we could activate the basic beamforming functionality with the whole row and column of elements. The rationale behind this is that the wide beam is synthesized by many small beams, and every sub-array is in charge of a specific beamforming direction as in Figure 6 for beam-broadening. The wide beam is successfully produced as soon as all small beams are configured as expected, and a row and a column of RIS elements in sub-arrays are sufficient to support 3D beamforming. Thus, compared with the activation scheme shown in Figure 14(c), it is a safer strategy to synthesize the wide beam sooner to ensure the influence coverage at the receiver.

***Stage 2*:** From the 8th to the 15th pattern, we use a spiral manner to gradually increase the number of activated elements within sub-arrays. In this way, we avoid creating mirror symmetry to cause potential dramatic changes when beamforming is insufficient (*i.e.*, from the 8th to the 11th pattern) [51] and accomplish gradual growth when the threshold is lifted high enough (*i.e.*, from the 12th to the 16th pattern). Thus, compared with the activation scheme shown in Figure 14(c), our scheme uniformly activates the elements across the area of the whole RIS to reduce the risk of abrupt variation increasing.

As the experimental data is shown in Figure 8, our RIS configuration strategy steadily increases the motion feature signal σ so as to lift the threshold δ'_{mov} covertly.

5.3.3 Practical Issue. In Figure 6, all RIS elements should at least achieve phase alignment on one side, *i.e.*, either the transmitter or receiver side, then we can configure multiple sub-arrays with sub-beams synthesizing a wide beam at the other side. Thus, before applying the RIS-aided beam-broadening beamforming in the variation-increasing stage shown in Figure 7, we need to estimate the precise direction of one transceiver since it is infeasible to synthesize wide beams for both directions.

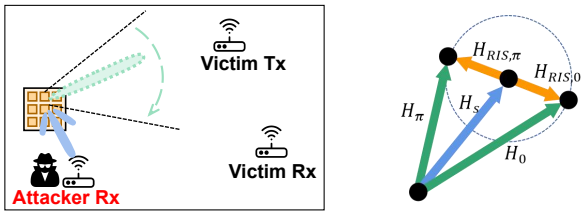

Figure 10: Sweeping-based RIS-aided AoA estimation. The beam for the transmitter side sweeps the potential transmitter's direction. For each direction, the corresponding configuration and the same one with an additional 180-degree phase shift are enforced successively to measure the pure influence of the RIS via $|H_{RIS,\pi} - H_{RIS,0}|$.

The transmitter generally emits the signal continually for intrusion detection in our scenario. Therefore, we design a RIS-aided Angle-of-Arrival (AoA) estimation scheme to estimate its location. As shown in Figure 10, we have a static RIS, and a collocated auxiliary receiver (e.g., a smartphone) with one or two antennas, *i.e.*, their relative direction is known. We then enforce the RIS to sweep the transmitter-side beam across the front area. For each candidate transmitter direction, we first configure the RIS accordingly and use the auxiliary receiver to collect the channel response data H_0, then further assign the RIS with the sample configuration plus an additional 180-degree phase shift and collect the channel response data H_π. H_0 is composed of the channel response without RIS impact H_s and the extra channel response introduced by RIS $H_{RIS,0}$, *i.e.*, $H_0 = H_s + H_{RIS,0}$. Similarly, $H_\pi = H_s + H_{RIS,\pi}$. As the only difference between $H_{RIS,0}$ and $H_{RIS,\pi}$ is a 180-degree phase difference, we can derive the following formula:

$$\begin{aligned} |H_\pi - H_0| &= |H_{RIS,\pi} - H_{RIS,0}| \\ &= 2|H_{RIS,0}| = 2|H_{RIS,\pi}| = 2|H_{RIS}| \end{aligned} \tag{7}$$

For the candidate case where its direction matches the true direction of the transmitter, the RIS will beamform the most energy to the auxiliary receiver, *i.e.*, observing the maximal $|H_\pi - H_0|$. In this way, we can infer the precise location of the transmitter relative to the RIS. Estimation accuracy could be improved by taking multiple measurements at various locations. Our preliminary experiment shows that this method achieves a mean error of 2.17 degrees and a median error of 1.66 degrees with single measurements. It is noted that this method may improve the localization of the transmitter in Sneaking Phase, but it cannot localize the receiver as the sparse traffic (e.g., reporting packets) of the receiver does not adequately support the scanning-based method.

6 EVALUATION

In this section, we experimentally investigate the effectiveness of RIStealth. First, we introduce the reproduction of the victim

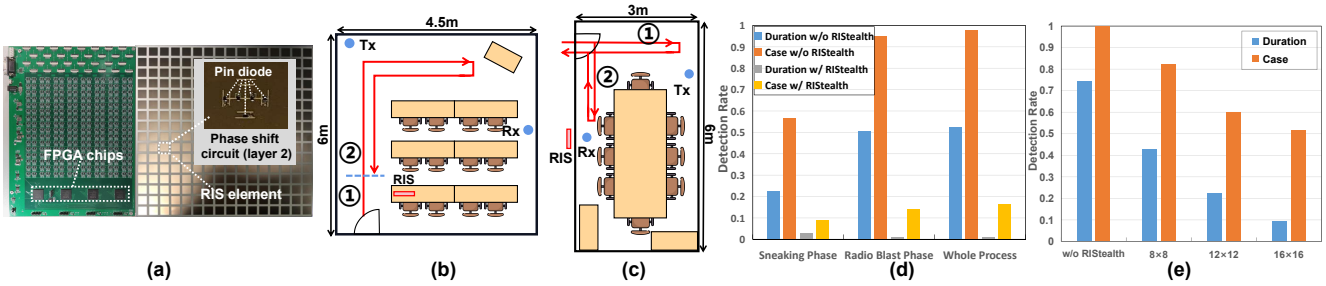

Figure 11: (a) RIS hardware prototype. Left (Ground layer): Control circuitry. Right (Patch layer and slot-loaded layer): RIS elements and phase shift circuits; (b) The LOS evaluation site layout and the trial routes; (c) The NLOS evaluation site layout and the trial routes; (d) The LOS evaluation results; (e) The NLOS evaluation results.

intrusion detection system and the implementation of the RIStealth hardware prototype. Then, we conduct the end-to-end evaluation of RIStealth in two typical indoor intrusion detection scenarios. Finally, we assess two major phases of the RIStealth design separately and thoroughly. All experiments are approved by our IRB.

6.1 Implementation

6.1.1 Victim Intrusion Detection System Reproduction. We build a victim intrusion detection system on two off-the-shelf mini-desktops equipped with Intel 5300 wireless NICs. The transmitter employs one antenna, while the receiver uses three antennas. The devices are set to monitor mode on channel 140 at 5.7 GHz. The CSI data for detection is collected with Linux CSI Tools [18]. Both the transmitter and the receiver are set at 120 cm height so that the motion of intruders with different heights will be clearly detected. Typically, the distance between the transmitter and the receiver is set to around 4m to cover a room.

We reproduce the state-of-the-art practical algorithm, AR-Alarm [25]. To our best knowledge, with some practicality improvement modules involved, AR-alarm is robust to more practical scenarios than other state-of-the-art works; thus, it is a stronger victim candidate for attack evaluation. It utilizes the widely-adopted standard deviation waveform as the motion feature [13, 49], which is derived by calculating the standard deviation of the CSI phase difference of adjacent antennas in a sliding window manner, and adopts $f(\cdot) = max(\cdot)$ in Equation 5. We select the best antenna pair and use the mean value of five successive subcarriers to calculate the phase difference. The initial threshold and the hyper-parameters of the duration-based and magnitude-based filter for false alarm reduction as in Figure 2(b)-(c) are well-tuned by collecting around 5 min reference measurement in 'no motion' status before experiments. We invite volunteers to walk inside the monitoring area uniformly. The detection rate, *i.e.*, the proportion of the detected trials overall, is about 95.1%, and the false alarm rate is around 1.13%. It validates the effectiveness of our reproduction.

6.1.2 RIStealth Hardware Prototype Implementation. We implement our 16×16 RIS prototype working at WiFi 5 GHz band based on the principle and RIS structure in [12]. Our prototype consists of a three-layer structure: patch layer, slot-loaded layer, and ground layer printed on the 31×31 cm standard Rogers 4350B substrates. As shown in Figure 11(a), for one RIS element, the phase shift circuit on the second layer has five pin diodes to enable four different phase configurations ($0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}$). These configurations can be switched by setting different bias voltages on two independent DC control lines. Thus, we design and implement an FPGA-based control circuit (see the left in Figure 11(a)) by integrating 512 independent control interfaces for standard serial port communication. In this way, we have 4^{256} different configurations. We assign the configuration schemes for the Sneaking Phase, Radio Blast Phase, and AoA estimation initiation to different push buttons on the FPGA board. Thus, the attacker can switch modes and advance the attack process. The result of AoA estimation can be obtained from the smartphone via serial port. It takes only microsecond-level time for the control circuit to flip the configurations of all 256 elements. It guarantees sufficient processing speed to support our signal variation increasing and threshold lifting design. For the power consumption, the static power of FPGA is about 0.65 W. The dynamic power will not exceed 1.65 W, and a normal 5V/0.5A USB port on the labtop can supply the system.

6.2 Overall Evaluation

We evaluate the end-to-end implementation of RIStealth, *i.e.*, including two consecutive phases, in two typical indoor settings: Line-of-Sight (LOS) scenario in Figure 11(b) and Non-Line-of-sight (NLOS) scenario in Figure 11(c).

Evaluation metrics: We adopt two metrics as follows:

$$\begin{aligned} \text{duration\_detection\_rate} &= T_{\text{detected}}/T_{\text{all}} \\ \text{case\_detection\_rate} &= N_{\text{detected}}/N_{\text{all}} \end{aligned} \tag{8}$$

where T_{detected} denotes the detected duration within the data duration, T_{all} denotes the whole data duration. N_{detected} denotes the number of trials that the intruder is detected, and N_{all} denotes the number of all trials. The case detection rate oppositely implies the success rate of all intrusions, while the duration detection rate inversely indicates the stealthiness within a single intrusion.

6.2.1 LoS Scenario Evaluation. We deploy the victim system in a classroom with the layout shown in Figure 11(b). The intruder first enters the room holding the RIS to apply Sneaking Phase (route 1 in Figure 11(b)). After reaching the specified location, he places down the RIS and applies Radio Blast Phase to lift the threshold to a certain level so that he can walk inside the protected area later

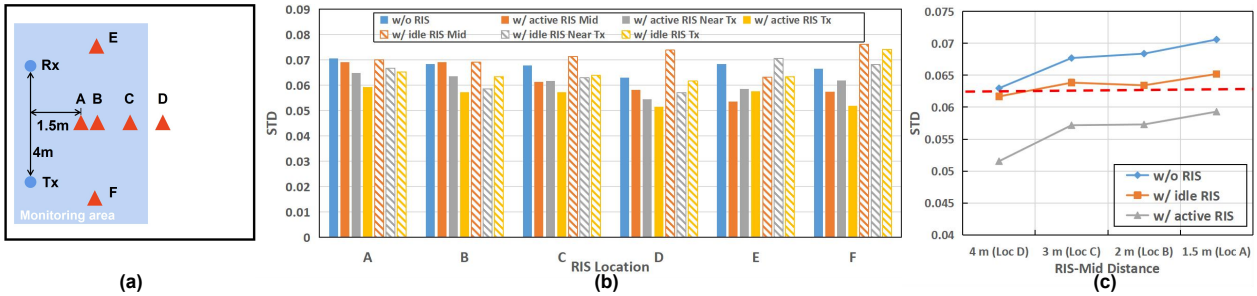

Figure 12: Sneaking Phase evaluation: (a) The evaluation site layout: The red triangles denote different positions of the intruder holding the RIS; (b) Results across various settings of RIS orientation (Tx, Neart Tx, Mid), location (A-F) and status (active/idle/none); (c) Results across various RIS-TransceiverMid distances. The red dotted line denotes the uplifted detection threshold in Sneaking Phase.

without being detected (route 2 in Figure 11(b)). We invite four volunteers, and each volunteer repeats the above process about 30 times. We evaluate two phases in terms of two metrics independently, as shown in Figure 11(d). For trials without RIStealth, the case detection rates are around 55% and 95% for route 1 and route 2, respectively. However, the case detection rate of route 1 drops to about 10% with the Sneaking Phase design, and that of route 2 drops to about 15% with the Radio Blast Phase design. For the whole two-phase process, the case detection rate drops to about 16.4%. Our evaluation of RIStealth indicates a significant reduction in both duration detection rate and case detection rate, *i.e.*, successfully exploiting RIS to spoof WiFi sensing. Though our attack did not reduce the detection rate to 0%, This finding demonstrates that the WiFi intrusion system is no longer entirely reliable when facing the threat of RIS-aided attacks.

Regarding attack time consumption, the time cost of Sneaking Phase relies on walking speed and distance to the switching point. In a LoS scenario, it takes approximately 2 seconds from entry to the switching point. In contrast, Radio Blast Phase takes roughly 20 seconds, primarily due to configuration switching speed, to increase the threshold and enable the attacker to move freely.

6.2.2 NLoS Scenario Evaluation. Sometimes, the intrusion detection system will be deployed near the door and the wall to monitor the entrance. Thus, we set up a site shown in Figure 11(c) to evaluate whether RIStealth still works. The intruder directly launches the Radio Blast Phase outside the wall and then tries to intrude into the protected area through route 1 and 2. In this part, we evaluate settings with different numbers of RIS-activated elements: 8×8, 12×12, and 16×16. All inactivated elements keep the same status and can be considered idle. For each test trial, we ask the volunteer to repeat the intrusion about 25 times. As presented in Figure 11(e), with the increasing of the RIS size, the duration detection rate decreases. However, it is notable that the case detection rate is not as low as the duration detection rate. This implies that although the capability of RIS on variation increasing can be enhanced by more activated elements to lift the threshold more effectively, the wall has dramatically attenuated the RIS-induced disturbance towards the receiver so that the threshold is not lifted sufficiently to prevent the most prominent motion signals from triggering the alert. We suggest that the number of elements of RIS should be at least 16×16 for the LOS scenario and even larger for the NLOS scenario.

6.3 Sneaking Phase Evaluation

To evaluate the effectiveness of Sneaking Phase on motion variation reduction, we set up an evaluation site as shown in Figure 12(a). We ask the volunteer to hold the configured RIS and perform in-situ movement repeatedly in position A-F. In each position, we repeat the experiment with different orientations, including facing the middle of the transmitter-receiver link (denoted as Mid), facing around the transmitter (denoted as Near Tx), and facing the transmitter (denoted as Tx). In addition, We repeat the experiment with idle RIS for reference and without RIS for comparison. We collect about 1-minute CSI data for each trial to quantify the average standard deviation of the phase difference (denoted as STD).

6.3.1 Impact of RIS Orientation/Location/Status. As the results are shown in Figure 12(b), it is found that the variation reduces the most when the RIS faces the transmitter, *i.e.*, the 'w/ active RIS Tx' case. The reason behind this is that making the RIS surface face the transmitter could redirect more energy toward the direction other than the receiver direction. It is worth mentioning that when the RIS is facing the middle of the link, *e.g.*, 'w/ idle RIS Mid' case, the receiver observes a quite high signal variation because partial signals experience the specular reflection due to the generalized Snell's Law to result in more disturbance at the receiver. The exceptions are position E and F, where the transmitter and the receiver are in adjacent orientations relative to RIS. Thus, the intruder has better make RIS surface face the transmitter in a practical attack.

In terms of location impact, we find that the signal variations are similar for position C, E, and F because these positions have the same length of the signal reflection path, *i.e.*, the total path length of Tx-[Loc]-Rx. Their only difference is whether the intruder is close to the transmitter or the receiver, which has not much impact on the results.

In terms of RIS status, the RIS is similar to a metal reflector in 'w/ idle RIS' cases. However, the natural reflection energy may hit the receiver due to the uncontrollable multipath reflection in a complicated indoor scenario. To make it worse, once triggering

specular reflection, the variation at the receiver will be much higher than 'w/o RIS' cases.

6.3.2 Impact of RIS-TransceiverMid Distance. The impact of the distance of RIS relative to the middle of the transceiver could be demonstrated by experiments on position A-D. We choose the cases of 'w/o RIS', 'w/ active RIS Tx', and 'w/ idle RIS Tx' in Figure 12(b) and summarize their results in Figure 12(c). It shows that the Sneaking Phase design can significantly reduce the variation compared to no RIS and idle RIS cases. However, the reduction effect will be weakened when the intruder gets closer to the transceiver because it becomes more difficult to neutralize reflection from the uncovered body of the intruder.

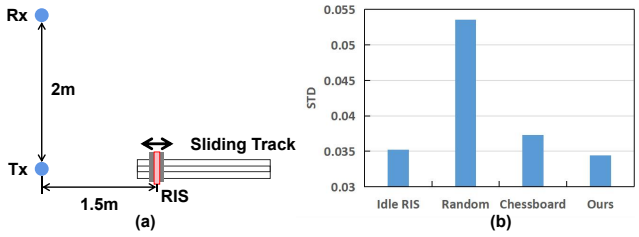

Figure 13: RIS configuration strategy comparison. (a) Experiment setup; (b) Variation results of different RIS configurations for motion reduction.

6.3.3 RIS Configuration Strategy Comparison. We conduct experiments to compare the RIS configuration of Sneaking Phase with others, including idle RIS, random RIS configuration, and chessboard RIS configuration [30]. Figure 13(a) shows our experimental setup where the RIS is fixed on a sliding track to move slightly and repeatedly, excluding inconsistent human factors for a fair comparison. We also prevent the specular reflection for effective comparison.

As in Figure 13(b), the random one performs the worst as the energy will be diffused to all sides. Our beam broadening configuration achieves the lowest average variation, close to the chessboard and idle RIS status. However, both of them cannot control the direction of the energy to prevent unexpected strong multipath reflection towards the receiver, *e.g.*, from some metal electronic devices and furniture in a complicated indoor environment. In contrast, our method keeps directing the energy to the ground to be more likely to avoid such cases. What's more, some intrusion detection systems [28] may wrongly recognize the redirected energy from the ground and the intruder's lower limbs as pets.

6.4 Radio Blast Phase Evaluation

We first evaluate the false alarm rate during the process of threshold lifting and then evaluate the duration detection rate once the threshold has been saturated.

6.4.1 False Alarm Rate during Threshold Lifting. Avoiding being detected during the threshold lifting means keeping a low false alarm rate in a static environment.

$$\text{false\_alarm\_rate} = \text{T}_{\text{detected}}/\text{T}_{\text{static}} \tag{9}$$

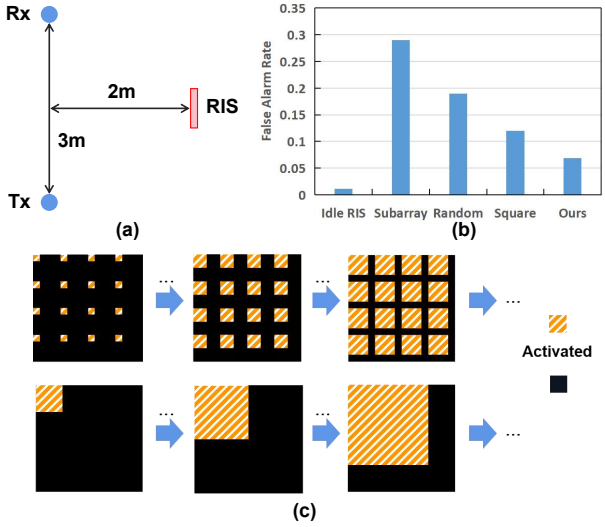

Figure 14: Different threshold lifting strategies comparison. (a) Experiment setup; (b) Comparison results; (c) sub-array square strategy and square strategy.

where T_{detected} denotes the detected in a 'no motion/static' environment and T_{static} denotes static duration. We set up an evaluation site shown in Figure 14(a) and compare our covert sub-array spiral activation strategy with other baselines: random, square, and sub-array square activation strategies shown in Figure 14(b)-(c). For the square method, the activated elements gradually grow to form a large square. The sub-array square method is to apply the square strategy to the sub-array granularity.

As shown in Figure 14(b), the sub-array square method performs the worst. The reason behind this is that each sub-array is responsible for a direction in the beam broadening method (Figure 6). Therefore, once we activate the sub-array covering the receiver direction with a great number of activated RIS elements, the influence will be increased rapidly to trigger the alarm. Our method achieves the lowest false alarm rate, while the random one and the square one perform around 2x~3x worse.

6.4.2 Intrusion Detection Rate after Threshold Lifting. To quantify the final effect of threshold lifting in Radio Blast Phase, we evaluate the duration detection rate across the monitoring area after the threshold reaches saturation. Specifically, we divide the monitoring area into 4×7 grids, each with the size of 1m×1m, as shown in Figure 15(a). We place the RIS at location A-F respectively to launch Radio Blast Phase strategy. For each attack location, after the threshold lifting process, we ask the subject to walk randomly inside each grid for about 30 seconds, except for the grid to place the transmitter, the receiver, and the RIS. Figure 15(b)-(h) show the spatial distribution of duration detection rates of the victim system without attack and under attacks with RIS at location A-F. It shows that the duration detection rates decrease significantly after RIS-aided threshold lifting. We further derive the metric as follows to facilitate comparison:

$$\text{blind\_zone\_rate} = \text{N}_{\text{accessible\_grids}}/\text{N}_{\text{all\_grids}} \tag{10}$$

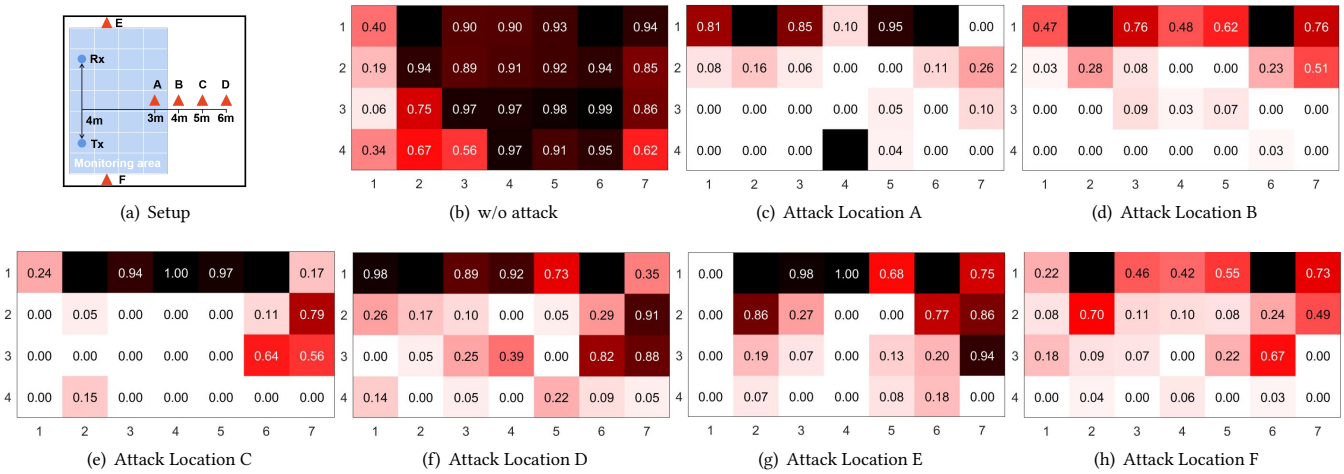

Figure 15: Evaluation on intrusion detection rate after threshold lifting. (a) Experiment setup. The red triangle denotes the attack position of RIS; (b)-(h) show the spatial distribution of duration detection rates of the victim system without attack and under attacks with RIS at location A-F, respectively.

The accessible grid is defined as the grid with a duration detection rate of less than 20%, which represents that only $\frac{1}{5}$ of the intrusion behavior in the grid will be detected. In addition, the accessible grid should be adjacent to other accessible grids or the boundary, *i.e.*, not be completely surrounded by inaccessible grids. In other words, these grids indicate the area where the intruder can access with a rather low probability of being detected.

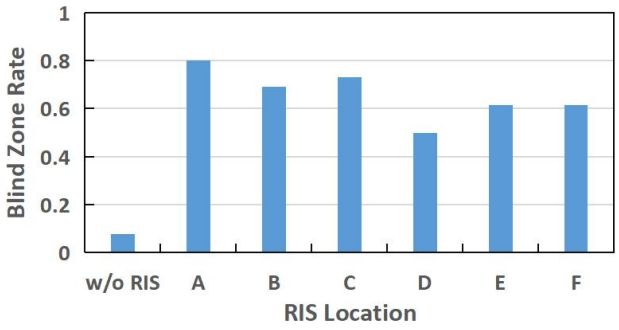

Figure 16: The blind zone rate across different RIS attack locations.

Figure 16 presents blind zone rates for different attack locations. The blind zone rate of the victim system could be increased to at least 50% from less than 10% under attacks. Noticeably, this result is aligned with the result of Section 6.2.1, where the intrusion behavior is hardly detected in blind zones, and most detected intrusion cases fall in the non-blind zone.

The effectiveness of Radio Blast Phase could be enhanced with a smaller distance of RIS relative to the transceiver based on the results on location A~D. Based on the results in Figure 15 (d), (g)-(h), it is noticed that the distance between the RIS and the transmitter matters more than that between the RIS and the receiver.

7 DISCUSSION

In this section, we discuss the generalization and enhancement of RIStealth and potential defense methods.

Target of RIStealth. Two categories of detection methods exist: learning-based and statistics-based. Learning-based methods [20, 33, 38, 42, 56] require prior training and recalibration over time to adapt to environmental changes, making them impractical [25, 52]. Practical statistics methods [25, 28, 57], including commercial products [29, 35], widely use adaptive threshold mechanisms. Thus, we choose this mechanism as the basis for attack since it is common in most practical Wi-Fi-based intrusion detection systems. For learning-based methods, the disturbed CSI signals caused by RIS resemble adversarial examples, potentially leading to performance degradation. We plan to consider RIStealth for learning-based methods in future work.

Generalization of RIStealth. (i) The design of RIStealth is based on physical-layer channel characteristics rather than WiFi-specific ones, ensuring its effectiveness as long as the operating frequency of RIS and auxiliary receiver match the victim system. (ii) RIStealth is not limited by specific motion features or threshold-changing algorithms used by victim systems. The essence of intrusion detection is to distill motion-relevant features from the dynamic channel component in Equation 2. Our evaluation uses the standard deviation of CSI phase difference, a common motion-relevant feature, but other features like CSI amplitude in Figure 10 also experience disturbance under RIS-aided attacks. Furthermore, Radio Blast Phase is based on the adaptive threshold mechanism, where the threshold is dynamically adjusted to account for varying environmental channel variance and minimize intrusion detection false alarms. While implementation details may affect the changing speed, the adaptive nature remains unchanged. In essence, RIStealth can pose universal threats to relevant wireless sensing applications by increasing motion-irrelevant disturbance and impairing the victim system's signal quality and sensing capacity.

Auxiliary Information Acquisition. The lack of prior knowledge of the victim transceiver's positions drives many design trade-offs of RIStealth. The attacker could infer auxiliary information in advance to facilitate attacks. For instance, receivers are typically located in the corner or next to the wall to increase sensing coverage and minimize disruption to daily activities. Thus, the attacker can estimate the transceiver's locations more precisely with the floor map of the target area. Then, the attacker could set a narrow beam safely to exchange an energy-concentrated beam for more threshold lifting space.

Extended Monitoring Area Intrusion. The intruder may fail when the deepest location he can intrude into with one-round Sneaking Phase is still beyond Radio Blast Phase's effective zone for sufficient threshold lifting. In this case, the attacker can carry two RISs and take turns executing two phases. He may initiate Sneaking Phase and move forward to a location, then initiate Radio Blast Phase to lift the threshold as much as possible. After that, he further leverages the second RIS to execute the Sneaking Phase and Radio Blast Phase respectively, before collecting the first RIS for the subsequent round. In this way, the intruder can finally intrude into the desired area.

Multiple Intrusion Detection Systems. This work represents the initial step in addressing practical RIS-related sensing security issues by focusing on one intrusion detection system with a single pair of transceivers, as recommended by commercial off-the-shelf products. Our aim is to raise awareness that a single WiFi intrusion detection system may not be entirely secure. In future work, we plan to explore handling multiple pairs of transceivers or multiple systems.

Potential Defense. The RIS-induced disturbance is almost indistinguishable from the natural variation of multipath reflection. Thus, it is hard to detect its existence via existing techniques, such as packet power analysis methods [2, 21] for jamming attacks and pause-and-detect approaches for a replay attack. One potential defense is to detect the threshold lifting pattern as Figure 8. However, since adaptive-threshold mechanisms are initially for enhancing robustness, merely detecting threshold lifting for defense purposes may also disrupt the original functionalities of threshold-changing algorithms and lead to increased false alarms. Defense development requires careful consideration of this dilemma. Furthermore, attackers can adapt their strategies by introducing more natural and indistinguishable patterns to avoid detection, e.g., random threshold lifting/dropping and phase rotating. Another potential defense is to equip the receiver with a high-sensitivity antenna array to identify the subtle beamforming energy from the malicious RIS, but relevant algorithms are required to differentiate malicious RIS reflections from benign ones.

8 RELATED WORKS

WiFi-based Intrusion Detection. There exist two categories of detection methods: learning-based and statistics-based ones. Learning-based methods [20, 33, 38, 42, 56] leverage machine learning models such as SVM, LSTM, KNN, and CNN to detect the intruder with very high accuracy. However, these works require prior training to capture scenario-tailored parameters as well as recalibration over time to adapt to environmental dynamics during long-term deployment [25, 52], which makes them unpractical. For statistics-based methods, some works extract RSS [24], while others leverage more fine-grained CSI to detect intrusion behaviors [13, 49, 52] with statistics features such as the standard deviation. To improve robustness, AR-Alarm [25] designs duration-based and magnitude-based filters to exclude the dropping objects and small objects, while PetFree [28] infers the height information to exclude the pets. We select AR-Alarm [25] as our evaluation target due to its practicability.

RIS-based Attacks against Sensing. Existing works on cloaking technique [39, 53, 54] and radar cross-section (RCS) reduction [11, 17, 34, 51] have utilized RIS to make objects invisible against radar systems. However, most works are limited to ideal lab settings and cannot be applied in real-world scenarios. They either overlook the unpredictable and extensive multi-path reflections or assume static target scenarios. In contrast, our work takes into account various conditions in real-world scenarios and proposes a practical attack scheme.

Wireless Physical Layer Security. The wireless physical layer attack has been widely discussed in wireless communication, and some recent works have focused on wireless sensing security. Some works develop geofencing methods for privacy protection [3, 19, 44]. There have also been attempts to tamper channel properties to spoof wireless localization [6, 27]. WiAdv [55] uses full-duplex radio to craft robust adversarial examples to cheat WiFi-based gesture recognition. In contrast, some works defend against adversarial wireless sensing [40, 43, 45]. Shenoy et al. [43] craft fake trajectories to obfuscate malicious mmWave radars, while others design full-duplex-based methods to prevent sensing information leakage [8, 40]. IRShield [45] also works on RIS-related sensing security. However, it aims to prevent adversarial motion sensing. Thus, it can deploy RIS close to the benign transmitter to obtain more energy for channel customization. Nevertheless, RIStealth exploits RIS to cover a moving intruder, and RIS is initially far away from the benign transmitter. Thus, we design beamforming techniques to boost the capability of RIS for channel customization.

9 CONCLUSION

Our work takes the first step to realizing a practical and covert RIS-aided attack in real-world scenarios. Through RIStealth design and extensive evaluation, we demonstrate how to strategically configure a handbag-sized passive RIS to precisely control channel disturbance and successfully render a moving person invisible to WiFi-based intrusion detection. RIStealth intriguingly shows that adaptive-threshold mechanisms, initially for enhancing robustness, inadvertently provide attack opportunities. We hope our RIS techniques and practices can promote advancements in RIS utilization and more research on RIS-relevant security issues.

10 ACKNOWLEDGMENTS

This research is supported in part by RGC under Contract CERG 16204820, 16206122, AoE/E-601/22-R, Contract R8015, and 3030_006. The authors would like to express their gratitude to the anonymous editors and reviewers for their valuable comments and suggestions.

REFERENCES

[1] Venkat Arun and Hari Balakrishnan. 2020. RFocus: Beamforming Using Thousands of Passive Antennas.. In *NSDI*. 1047–1061.

[2] Saeed Bagherinejad and S Mohammad Razavizadeh. 2021. Direction-based jamming detection and suppression in mmWave massive MIMO networks. *IET Communications* 15, 14 (2021), 1780–1790.

[3] Justin Chan, Changxi Zheng, and Xia Zhou. 2015. 3D Printing Your Wireless Coverage. In *Proceedings of the 2nd International Workshop on Hot Topics in Wireless* (Paris, France) *(HotWireless '15)*. Association for Computing Machinery, New York, NY, USA, 1–5. https://doi.org/10.1145/2799650.2799653

[4] Lili Chen, Wenjun Hu, Kyle Jamieson, Xiaojiang Chen, Dingyi Fang, and Jeremy Gummeson. 2021. Pushing the Physical Limits of IoT Devices with Programmable Metasurfaces.. In *NSDI*. 425–438.

[5] Xi Chen, Chen Ma, Michel Allegue, and Xue Liu. 2017. Taming the inconsistency of Wi-Fi fingerprints for device-free passive indoor localization. In *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE, 1–9.

[6] Yingying Chen, Wade Trappe, and Richard P Martin. 2007. Attack detection in wireless localization. In *IEEE INFOCOM 2007-26th IEEE International Conference on Computer Communications*. IEEE, 1964–1972.

[7] Qiang Cheng, Lei Zhang, Jun Yan Dai, Wankai Tang, Jun Chen Ke, Shuo Liu, Jing Cheng Liang, Shi Jin, and Tie Jun Cui. 2022. Reconfigurable Intelligent Surfaces: Simplified-Architecture Transmitters—From Theory to Implementations. *Proc. IEEE* 110, 9 (2022), 1266–1289. https://doi.org/10.1109/JPROC.2022.3170498

[8] Marco Cominelli, Francesco Gringoli, and Renato Lo Cigno. 2022. AntiSense: Standard-compliant CSI obfuscation against unauthorized Wi-Fi sensing. *Computer Communications* 185 (2022), 92–103.

[9] Tie Jun Cui, Shuo Liu, and Lei Zhang. 2017. Information metamaterials and metasurfaces. *Journal of materials chemistry C* 5, 15 (2017), 3644–3668.

[10] Tie Jun Cui, Mei Qing Qi, Xiang Wan, Jie Zhao, and Qiang Cheng. 2014. Coding metamaterials, digital metamaterials and programmable metamaterials. *Light: science & applications* 3, 10 (2014), e218–e218.

[11] Huijuan Dai, Yongjiu Zhao, Huangyan Li, Jiaqing Chen, Zheng He, and Wenjun Qi. 2019. An ultra-wide band polarization-independent random coding Metasurface for RCS reduction. *Electronics* 8, 10 (2019), 1104.

[12] Linglong Dai, Bichai Wang, Min Wang, Xue Yang, Jingbo Tan, Shuangkaisheng Bi, Shenheng Xu, Fan Yang, Zhi Chen, Marco Di Renzo, et al. 2020. Reconfigurable intelligent surface-based wireless communications: Antenna design, prototyping, and experimental results. *IEEE access* 8 (2020), 45913–45923.

[13] Enjie Ding, Xiansheng Li, Tong Zhao, Lei Zhang, and Yanjun Hu. 2018. A robust passive intrusion detection system with commodity WiFi devices. *Journal of Sensors* 2018 (2018).

[14] Manideep Dunna, Chi Zhang, Daniel Sievenpiper, and Dinesh Bharadia. 2020. ScatterMIMO: Enabling virtual MIMO with smart surfaces. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–14.

[15] Chao Feng, Xinyi Li, Yangfan Zhang, Xiaojing Wang, Liqiong Chang, Fuwei Wang, Xinyu Zhang, and Xiaojiang Chen. 2021. RFlens: Metasurface-Enabled Beamforming for IoT Communication and Sensing. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking* (New Orleans, Louisiana) *(MobiCom '21)*. Association for Computing Machinery, New York, NY, USA, 587–600. https://doi.org/10.1145/3447993.3483238

[16] Zi Feng, Jianxia Ning, Ioannis Broustis, Konstantinos Pelechrinis, Srikanth V. Krishnamurthy, and Michalis Faloutsos. 2011. Coping with packet replay attacks in wireless networks. In *2011 8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*. 368–376. https://doi.org/10.1109/SAHCN.2011.5984919

[17] Mohammad-Javad Haji-Ahmadi, Vahid Nayyeri, Mohammad Soleimani, and Omar M Ramahi. 2017. Pixelated checkerboard metasurface for ultra-wideband radar cross section reduction. *Scientific Reports* 7, 1 (2017), 1–12.

[18] Daniel Halperin, Wenjun Hu, Anmol Sheth, and David Wetherall. 2011. Tool release: Gathering 802.11 n traces with channel state information. *ACM SIGCOMM computer communication review* 41, 1 (2011), 53–53.

[19] Jiang Haofeng and Gong Xiaorui. 2019. Wi-Fi Secure Access Control System Based on Geo-fence. In *2019 IEEE Symposium on Computers and Communications (ISCC)*. 1–6. https://doi.org/10.1109/ISCC47284.2019.8969707

[20] Yuqian Hu, Muhammed Zahid Ozturk, Beibei Wang, Chenshu Wu, Feng Zhang, and KJ Ray Liu. 2022. Robust Passive Proximity Detection Using Wi-Fi. *IEEE Internet of Things Journal* 10, 7 (2022), 6221–6234.

[21] Sunakshi Jaitly, Harshit Malhotra, and Bharat Bhushan. 2017. Security vulnerabilities and countermeasures against jamming attacks in Wireless Sensor Networks: A survey. In *2017 International Conference on Computer, Communications and Electronics (Comptelix)*. IEEE, 559–564.

[22] T. Karhima, A. Silvennoinen, M. Hall, and S.-G. Haggman. 2004. IEEE 802.11b/g WLAN tolerance to jamming. In *IEEE MILCOM 2004. Military Communications Conference, 2004.*, Vol. 3. 1364–1370 Vol. 3. https://doi.org/10.1109/MILCOM.2004.1495141

[23] J Clayton Kerce, George C Brown, and Mark A Mitchell. 2007. Phase-only transmit beam broadening for improved radar search performance. In *2007 IEEE Radar Conference*. IEEE, 451–456.

[24] Ahmed E Kosba, Ahmed Saeed, and Moustafa Youssef. 2012. RASID: A robust WLAN device-free passive motion detection system. In *2012 IEEE International Conference on Pervasive Computing and Communications*. IEEE, 180–189.

[25] Shengjie Li, Xiang Li, Kai Niu, Hao Wang, Yue Zhang, and Daqing Zhang. 2017. Ar-alarm: An adaptive and robust intrusion detection system leveraging csi from commodity wi-fi. In *Enhanced Quality of Life and Smart Living: 15th International Conference, ICOST 2017, Paris, France, August 29-31, 2017, Proceedings 15*. Springer, 211–223.

[26] 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.

[27] Zang Li, Wade Trappe, Yanyong Zhang, and Badri Nath. 2005. Robust statistical methods for securing wireless localization in sensor networks. In *IPSN 2005. Fourth International Symposium on Information Processing in Sensor Networks, 2005.* IEEE, 91–98.

[28] Yuxiang Lin, Yi Gao, Bingji Li, and Wei Dong. 2020. Revisiting indoor intrusion detection with WiFi signals: do not panic over a pet! *IEEE Internet of Things Journal* 7, 10 (2020), 10437–10449.

[29] Linksys. 2019. Linksys Aware. https://www.linksys.com/for-home/software-and-services/linksys-aware/

[30] Xiao Liu, Jun Gao, Liming Xu, Xiangyu Cao, Yi Zhao, and Sijia Li. 2016. A coding diffuse metasurface for RCS reduction. *IEEE Antennas and wireless propagation letters* 16 (2016), 724–727.

[31] Xi Liu, Anmol Sheth, Michael Kaminsky, Konstantina Papagiannaki, Srinivasan Seshan, and Peter Steenkiste. 2009. DIRC: Increasing Indoor Wireless Capacity Using Directional Antennas. In *Proceedings of the ACM SIGCOMM 2009 Conference on Data Communication* (Barcelona, Spain) *(SIGCOMM '09)*. Association for Computing Machinery, New York, NY, USA, 171–182. https://doi.org/10.1145/1592568.1592589

[32] Xi Liu, Anmol Sheth, Michael Kaminsky, Konstantina Papagiannaki, Srinivasan Seshan, and Peter Steenkiste. 2010. Pushing the Envelope of Indoor Wireless Spatial Reuse Using Directional Access Points and Clients. In *Proceedings of the Sixteenth Annual International Conference on Mobile Computing and Networking* (Chicago, Illinois, USA) *(MobiCom '10)*. Association for Computing Machinery, New York, NY, USA, 209–220. https://doi.org/10.1145/1859995.1860020

[33] Jiguang Lv, Dapeng Man, Wu Yang, Liangyi Gong, Xiaojiang Du, and Miao Yu. 2019. Robust device-free intrusion detection using physical layer information of WiFi signals. *Applied Sciences* 9, 1 (2019), 175.

[34] Akila Murugesan, Krishnasamy T Selvan, Ashwin Iyer, Kumar Vaibhav Srivastava, and Arokiaswami Alphones. 2021. A review of metasurface-assisted RCS reduction techniques. *Progress In Electromagnetics Research B* 94 (2021), 75–103.

[35] Inc. Origin Wireless. 2021. Hex Home Smart Home Security System. https://myhexhome.com/

[36] Xilong Pei, Haifan Yin, Li Tan, Lin Cao, Zhanpeng Li, Kai Wang, Kun Zhang, and Emil Björnson. 2021. RIS-Aided Wireless Communications: Prototyping, Adaptive Beamforming, and Indoor/Outdoor Field Trials. *IEEE Transactions on Communications* 69, 12 (2021), 8627–8640. https://doi.org/10.1109/TCOMM.2021.3116151

[37] 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. https://doi.org/10.1109/COMST.2022.3159185

[38] Fang Qi, Yingkai Zhao, Md Zakirul Alam Bhuiyan, Hai Tao, Weifeng Yan, and Zhe Tang. 2022. Artificial intelligence driven Wi-Fi CSI data mining: Focusing on the intrusion detection applications. *International Journal of Communication Systems* (2022), e5338.

[39] Chao Qian, Bin Zheng, Yichen Shen, Li Jing, Erping Li, Lian Shen, and Hongsheng Chen. 2020. Deep-learning-enabled self-adaptive microwave cloak without human intervention. *Nature photonics* 14, 6 (2020), 383–390.

[40] Yue Qiao, Ouyang Zhang, Wenjie Zhou, Kannan Srinivasan, and Anish Arora. 2016. PhyCloak: Obfuscating sensing from communication signals. In *13th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 16)*. 685–699.

[41] Sridhar Rajagopal. 2012. Beam broadening for phased antenna arrays using multi-beam subarrays. In *2012 IEEE International Conference on Communications (ICC)*. IEEE, 3637–3642.

[42] Jatin Sadhwani and M Sabarimalai Manikandan. 2021. Non-collaborative human presence detection using channel state information of Wi-Fi signal and long-short term memory neural network. In *2021 13th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*. IEEE, 1–6.

[43] Jayanth Shenoy, Zikun Liu, Bill Tao, Zachary Kabelac, and Deepak Vasisht. 2022. RF-protect: privacy against device-free human tracking. In *Proceedings of the ACM SIGCOMM 2022 Conference*. 588–600.

[44] Anmol Sheth, Srinivasan Seshan, and David Wetherall. 2009. Geo-fencing: Confining Wi-Fi Coverage to Physical Boundaries.. In *Pervasive*, Vol. 5538. 274–290.

[45] 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.

[46] Zhi Sun, Sarankumar Balakrishnan, Lu Su, Arupjyoti Bhuyan, Pu Wang, and Chunming Qiao. 2021. Who Is in Control? Practical Physical Layer Attack and Defense for mmWave-Based Sensing in Autonomous Vehicles. *IEEE Transactions on Information Forensics and Security* 16 (2021), 3199–3214. https://doi.org/10.1109/TIFS.2021.3076287

[47] Mathy Vanhoef and Frank Piessens. 2014. Advanced Wi-Fi Attacks Using Commodity Hardware. In *Proceedings of the 30th Annual Computer Security Applications Conference* (New Orleans, Louisiana, USA) *(ACSAC '14)*. Association for Computing Machinery, New York, NY, USA, 256–265. https://doi.org/10.1145/2664243.2664260

[48] Ambuj Varshney, Luca Mottola, Mats Carlsson, and Thiemo Voigt. 2015. Directional Transmissions and Receptions for High-Throughput Bulk Forwarding in Wireless Sensor Networks. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems* (Seoul, South Korea) *(SenSys '15)*. Association for Computing Machinery, New York, NY, USA, 351–364. https://doi.org/10.1145/2809695.2809720

[49] Tong Xin, Bin Guo, Zhu Wang, Pei Wang, Jacqueline Chi Kei Lam, Victor Li, and Zhiwen Yu. 2018. FreeSense: A robust approach for indoor human detection using Wi-Fi signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–23.

[50] Chen Yan, Wenyuan Xu, and Jianhao Liu. 2016. Can you trust autonomous vehicles: Contactless attacks against sensors of self-driving vehicle. *Def Con* 24, 8 (2016), 109.

[51] Fang Yuan, Guang-Ming Wang, He-Xiu Xu, Tong Cai, Xiao-Jun Zou, and Ze-Hao Pang. 2017. Broadband RCS reduction based on spiral-coded metasurface. *IEEE Antennas and wireless propagation letters* 16 (2017), 3188–3191.

[52] Feng Zhang, Chenshu Wu, Beibei Wang, Hung-Quoc Lai, Yi Han, and KJ Ray Liu. 2019. WiDetect: Robust motion detection with a statistical electromagnetic model. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–24.

[53] Xin Ge Zhang, Ya Lun Sun, Qian Yu, Qiang Cheng, Wei Xiang Jiang, Cheng-Wei Qiu, and Tie Jun Cui. 2021. Smart Doppler cloak operating in broad band and full polarizations. *Advanced Materials* 33, 17 (2021), 2007966.

[54] Zheng Zhen, Chao Qian, Yuetian Jia, Zhixiang Fan, Ran Hao, Tong Cai, Bin Zheng, Hongsheng Chen, and Erping Li. 2021. Realizing transmitted metasurface cloak by a tandem neural network. *Photonics Research* 9, 5 (2021), B229–B235.

[55] 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.

[56] Guozhen Zhu, Chenshu Wu, Xiaolu Zeng, Beibei Wang, and KJ Ray Liu. 2022. Who Moved My Cheese? Human and Non-human Motion Recognition with WiFi. In *2022 IEEE 19th International Conference on Mobile Ad Hoc and Smart Systems (MASS)*. IEEE, 476–484.

[57] Yanzi Zhu, Zhujun Xiao, Yuxin Chen, Zhijing Li, Max Liu, Ben Y Zhao, and Haitao Zheng. 2020. Et tu alexa? when commodity wifi devices turn into adversarial motion sensors. In *Network and Distributed Systems Security (NDSS) Symposium*.