UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

DECEPTION IN WIRELESS COMMUNICATIONS: BENEFITS AND PERILS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

BY

QIUYE HE
Norman, Oklahoma
2024

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/340391

DECEPTION IN WIRELESS COMMUNICATIONS: BENEFITS AND PERILS

A DISSERTATION APPROVED FOR THE
SCHOOL OF COMPUTER SCIENCE

BY THE COMMITTEE CONSISTING OF

Dr. Song Fang, Chair

Dr. Qi Cheng

Dr. Anindya Maiti

Dr. Choon Yik Tang

© Copyright by Qiuye He 2024
All Rights Reserved.

Acknowledgements

First and foremost, I would like to thank my advisor, Dr. Song Fang, for his invaluable guidance, inspiration, patience, feedback, and continuous support throughout my Ph.D. journey. He is very nice and always there to help. I consider myself incredibly lucky to have him as my advisor.

I would also like to thank all my committee members, Dr. Qi Cheng, Dr. Anindya Maiti, and Dr. Choon Yik Tang, for dedicating their precious time to reviewing my dissertation and providing insightful comments and suggestions on my research. Additionally, I am grateful to the National Science Foundation (NSF) for its funding support.

To my labmates in the Cybersecurity research lab, thank you for all the interesting and inspiring discussions. They are Edwin Yang, Yan He, Guanchong Huang, and Yan Zhou.

I would like to extend my appreciation to the University of Oklahoma and the School of Computer Science for providing me with the resources, facilities, and funding that have enabled me to conduct my research.

Last but not least, I owe a debt of gratitude to my family and friends. Their unwavering love, support, and understanding throughout my academic journey have been a constant source of motivation and inspiration.

Thank you all from the bottom of my heart.

This work incorporates material that has previously been published in peer-reviewed articles authored by me. Full citations of these publications are provided at the beginning of each respective chapter where the material appears.

Abstract

Wireless signals, ubiquitous and capable of penetrating obstacles, are increasingly exploited to infer personal information such as vital signs, human activities, and crowd counts. This surge in wireless sensing techniques benefits society but also poses significant security threats. Adversaries could potentially use these signals for wireless inference attacks, determining a target's activities with little regard for privacy or security. This dissertation mainly includes two studies on deception in existing wireless networks and systems. On one hand, it proposes proactive measures to send deceptive signals to mislead eavesdroppers, thus convincing them their attempts are successful when, in fact, they are failing. On the other hand, it identifies vulnerabilities in wireless liveness detection systems, where an attacker can use such wireless deceptive techniques to compromise these systems.

The first study proposes proactive defenses against all existing Channel State Information (CSI)-based vital signs and crowd counting inference attacks by establishing ambush locations and then transmitting deceptive signals designed to mimic vital signs or person count, albeit falsely, thereby protecting the user's true vital signs or person count data. Experimental results on software-defined radio (SDR) platforms demonstrate that these defensive strategies can effectively mislead eavesdroppers, who then receive falsified breathing rates or person counts.

Conversely, the second study demonstrates how attackers exploit these wire-

less deceptive techniques to compromise wireless liveness detection systems. Malicious attackers are capable of crafting fake wireless signals synchronized with spoofed video or audio streams to deceive systems into accepting fraudulent activities as legitimate. We refer to such attacks as *phantom-CSI attacks*. Experimental results on SDR platforms verify that these *phantom-CSI attacks* can significantly reduce the accuracy of spoof detection in wireless liveness systems.

Contents

Acknowledgements **iii**

Abstract **vi**

List of Figures **x**

List of Tables **xii**

1 Introduction **1**
- 1.1 Counteractive Strategies for Wireless Breath and Crowd Inference Attacks . . . 3
- 1.2 Vulnerabilities in Wireless Liveness Detection Systems . . . 8
- 1.3 Summary of Contributions . . . 12

2 Related Work **14**
- 2.1 Wireless Breathing Rate Inference Techniques . . . 14
- 2.2 Wireless Crowd Counting Techniques . . . 16
- 2.3 Wireless Human Activity Detection . . . 17
- 2.4 Liveness Detection . . . 18

3 Proactive Ambush Tactic Against Wireless Breath and Crowd Inference **20**
- 3.1 Preliminaries . . . 20
  - 3.1.1 Fresnel Zone . . . 21
  - 3.1.2 CSI-based Breathing Rate Inference . . . 22
  - 3.1.3 CSI-based Crowd Counting Inference . . . 22
- 3.2 Attack Model and Assumptions . . . 23
- 3.3 Ambush Design for Breath Inference Attacks . . . 24
  - 3.3.1 Overview . . . 24
  - 3.3.2 Planning Phase . . . 25
  - 3.3.3 Disturbance Manipulation . . . 27
  - 3.3.4 Breathing Rate Retrieval . . . 29
  - 3.3.5 From Point Ambush to Area Ambush . . . 33

3.3.6 Security Analysis . . . 37
3.4 Ambush Design for Crowd Inference Attacks . . . 39
3.4.1 Overview . . . 40
3.4.2 Planning Phase . . . 41
3.4.3 Disturbance Manipulation . . . 42
3.4.4 Person Number Estimation . . . 42
3.5 Experimental Evaluation . . . 44
3.5.1 Evaluation Setup . . . 44
3.5.2 Breathing Rate Inference Attacks . . . 46
3.5.3 Example Defenses . . . 48
3.5.4 Overall Defense Impact . . . 52
3.5.5 Two-user Scenario . . . 54
3.5.6 Trap Area Evaluation . . . 55
3.5.7 Ghost Defense against Crowd Inference Attacks . . . 58

4 Phantom-CSI Attacks against Wireless Liveness Detection 62
4.1 Preliminaries . . . 62
4.1.1 CSI Estimation . . . 63
4.1.2 CSI-aided Liveness Detection . . . 63
4.2 Adversary Model . . . 64
4.3 System Design . . . 65
4.3.1 Attack Overview . . . 65
4.3.2 Video-based Pipeline . . . 66
4.3.3 Artificial CSI Generation . . . 68
4.3.4 Transmission Manipulation . . . 69
4.3.5 CSI-aided Liveness Detection . . . 72
4.4 Experimental Results . . . 78
4.4.1 Evaluation Setup . . . 79
4.4.2 Effectiveness of Channel Manipulation . . . 80
4.4.3 Two Attack Cases . . . 83
4.4.4 Overall Attack Impact . . . 85
4.4.5 User Study . . . 90
4.5 Attack Against Wireless Voice Liveness Detection . . . 92
4.5.1 Implementation Setup . . . 93
4.5.2 Case Study . . . 93
4.5.3 Overall Performance . . . 95
4.5.4 User Study . . . 98
4.6 Discussions . . . 99
4.6.1 Limitations . . . 99
4.6.2 Countermeasures . . . 100

5 Future Work **102**
5.1 Challenges with IoT Devices 103
5.2 ML for Wireless HPI Inference 106
5.3 HPI Inference with mmWave 109

6 Conclusion **112**

Bibliography **114**

List of Figures

1.1 Creating a fake (sensitive or insensitive) CSI. 6
1.2 Crafting wireless signal affected by human activity. 11
1.3 General structure of a wireless liveness detection system. 12

3.1 Demonstration of Fresnel Zones. 21
3.2 Flow chart of the proposed ambush tactic. 25
3.3 Selecting an ambush location. 25
3.4 An MAC process. 28
3.5 CSI pre-processing. 30
3.6 Subcarrier sensitivity. 31
3.7 Local peaks. 32
3.8 Peaks in PSD. 33
3.10 Flow chart of the proposed *Ghost* defense. 41
3.11 Layout of the experimental environment. 45
3.12 Setup for deploying an ambush area. 46
3.13 Values of ϵ and ϵ at Eve when no defense is enforced. 47
3.14 Enabling Eve to obtain no breathing activity. 49
3.15 Fabricating normal breath. 50
3.16 Making Eve obtain abnormal breath. 51
3.17 CDFs of $P(\epsilon \leq x)$ for **D1**. 52
3.18 CDFs of $P(\epsilon \leq x)$ and $P(\eta \leq x)$ for **D2**. 53
3.19 CDFs of $P(\epsilon \leq x)$ and $P(\eta \leq x)$ for **D3**. 54
3.20 Mean absolute estimation errors (AEE). 55
3.21 Extending defenses in two-user scenario. 56
3.22 Fabricating normal breath for a trap area. 57
3.23 Confusion matrix for crowd inference attack. 59
3.24 Estimated CSI at Eve in an empty room when our defense is enforced. 60

4.1 Flow chart of the phantom-CSI attack. 64
4.2 Body keypoints extracted by *OpenPose*. 67
4.3 Subcarrier-level CSI wave morphing. 70
4.4 Procedures of CSI data preprocessing. 73

4.5 Layout of the experimental environment. 79
4.6 Three daily events. 79
4.7 Channel manipulation in a static environment. 81
4.8 Channel manipulation in a dynamic environment. 82
4.9 Video and the CSI signals when fabricating events. 83
4.10 Video and CSI signal comparison when hiding events. 85
4.11 CDF of the extracted features in a normal situation and when a video spoofing only attack happens. 87
4.12 CDF of the extracted features with our attack. 88
4.13 Event start time discrepancies. 89
4.14 Event end time discrepancies. 90
4.15 Mean frequency discrepancies. 91
4.16 An example of a wireless-based voice liveness detection. 93
4.17 CDFs of start/end time for normal and voice spoofing attack only cases. 94
4.18 CDFs of word count for normal and voice spoofing attack only cases. 94
4.19 CDFs of start/end time when the proposed attack is launched. . . 95
4.20 Speaking start time differences. 96
4.21 Speaking end time differences. 97
4.22 Mean word count differences. 98

5.1 Extensive applications of IoT. 103

List of Tables

4.1 Different human activity combinations. 84
4.2 Wireless video liveness detection vs. feature count. 89
4.3 Impact of different event types. 90
4.4 The list of voice commands we test. 91
4.5 Wireless voice liveness detection vs. feature count. 97
4.6 Wireless voice liveness detection vs. word count. 98

Chapter 1

Introduction

In recent years, the utilization of wireless signals for inferring various personal information, such as vital signs [68], postures [66], and keystrokes [178], has garnered increasing interest. While the proliferation of wireless sensing techniques has offered significant societal benefits, it also introduces substantial security concerns. Specifically, adversaries may exploit these techniques to determine a target user's activities by collecting corresponding wireless signals and conducting inference attacks. For instance, [98] demonstrates a technique by which an attacker could infer the vital signs of the target user, such as breathing rate and heartbeat, potentially resulting in privacy leakage.

However, there is little attention paid to the security and privacy issues in wireless sensing. This gap is particularly alarming given the pervasive reliance on wireless technologies in various aspects of daily life and industry. To mitigate the misuse of wireless techniques, we propose proactive measures against wireless eavesdropping attacks. These measures deceive eavesdroppers with fake but seemingly meaningful wireless signals, obscuring the fact that their eavesdrop-

ping attempts have failed. Consequently, this may prompt the eavesdropper to exert further efforts to break into the wireless communication [51, 67].

Moreover, due to the fact that wireless signals are ubiquitous, invisible, and able to penetrate through obstacles, wireless signals are widely used as second-factor authentication in wireless liveness detection systems to enhance the owner's privacy. Such systems, which monitor human behavior in real-time, are inherently susceptible to spoofing attacks. However, the nature of wireless signals that facilitates their use in authentication also renders them vulnerable to spoofing. Adversaries can generate counterfeit wireless signals synchronized with spoofed video or audio streams, significantly complicating the task of distinguishing genuine human activity from fraudulent activities.

This dissertation proposes a deception strategy in wireless communication, which serves as a double-edged sword: it can be employed by defenders to mislead attackers and protect personal health information from being leaked; it can also be utilized by attackers targeting wireless liveness detection systems to bypass them successfully. The dissertation mainly introduces two studies. In the first study, new proactive countermeasures against all existing Channel State Information (CSI)-based vital signs and crowd counting inference methods are introduced. Specifically, we establish ambush locations with carefully designed wireless signals, enabling eavesdroppers to deduce a false breathing rate and person count as specified by the transmitter, thereby safeguarding the breathing rates and true person counts. In the second study, we demonstrate how an adversary can easily generate phantom wireless signals and synchronize them with spoofed video/voice signals, making it challenging for legitimate users to differentiate between real and fake human activity.

Chapter 2 summarizes the related work. The details of my two studies are

presented in Chapters 3 and 4, respectively. Chapter 5 discusses future work. In the following, I provide the motivations and background for each of the two studies.

1.1 Counteractive Strategies for Wireless Breath and Crowd Inference Attacks

Crowd counting and vital signs inference via wireless signals has drawn increasing attention due to the widespread availability of wireless infrastructures and the lack of need for direct contact with devices [4, 24, 62, 64, 73, 81, 98, 99, 102, 113, 119, 154, 160, 172, 173, 179, 184, 185]. With such a technique, an eavesdropper can stealthily set up a wireless receiver on one side of the user to passively collect the signals emitted by a wireless Access Point (AP) which is on the other side of the user. The presence of multiple moving people leads to a larger variation in the channel state information (CSI) over time, which can be used to estimate the number of people in a room for crowd counting. Similarly, the fluctuations in the received signals caused by respiration-induced chest and stomach movements can reveal sensitive vital signs, which can be analyzed by the eavesdropper for vital signs inference.

The popularity of such techniques also brings privacy concerns. In detail, attackers can extract personal information such as the number of individuals present, their location, and even their vital signs, which can be exploited for malicious purposes like stalking or theft. For example, an eavesdropper can track occupancy in a home using a crowd inference attack, and then break in once the target homes are vacant to reduce the chance of getting caught [21]. Fur-

thermore, there are also extensive research efforts that detect breathing for user presence identification [105, 123, 152, 169], which can result in serious security issues. On the other hand, vital signs often contain sensitive information related to the state of personal essential body function [1, 53, 98, 102, 168]. Generally, the normal breathing rate for an adult at rest is 12 to 20 breaths per minute (bpm). Abnormal breathing may be a symptom of diseases, such as pulmonary diseases [30], hypertension or hyperthyroidism [14], heart problems or drug overdose [44], asthma or pneumonia [147], and cardiovascular diseases like stroke [53]. The disclosure of such health information can cause serious consequences such as employment discrimination based on health status [91], and a company's stock plummeting due to its CEO's health concerns [37, 39].

Though research is booming in crowd and vital signs inference through wireless signals, there are few research efforts discussing corresponding countermeasures. Traditional anti-eavesdropping methods usually take the following two defenses: (1) *Cryptographic key based:* by encrypting transmitted messages between legitimate parties [48, 135], an eavesdropper without the secret key cannot successfully decode the received message; and (2) *Friendly jamming based:* an ally jammer actively sends jamming signals (e.g., [59, 132]) which interrupt the eavesdropping while the receiver can decode messages by canceling the impact of the inference signals [157]. With either mechanism, the eavesdropper would capture encrypted or disrupted signals, which are often random and meaningless. Though the eavesdropper may not get the correct wireless signals, the unintelligibility of those signals indicates to her that her eavesdropping fails. She may thus make further efforts to break the wireless communication. For example, an eavesdropper may attempt to steal the secret key via social engineering methods (e.g., [89]) or side-channel attacks (e.g., [56]). Also, it has been shown that an

attacker equipped with multiple antennas is able to separate the message from the jamming signals [140]. Due to the importance of personal privacy, a more effective defense strategy is thus much-needed to prevent wireless crowd and vital signs eavesdropping.

Orthogonal frequency-division multiplexing (OFDM) is widely used in modern wireless communication systems (e.g., 802.11a/g/n/ac/ad) with multiple subcarrier frequencies to encode a packet. The minute wireless signal disturbance caused by human motion can be captured by *received signal strength* (RSS) or *channel state information* (CSI). RSS only provides the average power in a received radio signal over the whole channel bandwidth, while CSI represents how the wireless channel impacts the radio signal that propagates through it (e.g., amplitude attenuation and phase shift). CSI offers fine-grained channel information, consisting of subcarrier-level information. As a result, CSI is more sensitive to human activity and has shown the best performance in inferring human activity compared with other wireless techniques [73].

What if we actively feed the eavesdropper with a meaningful but bogus person count or breathing rate? When the eavesdropper is misled by the fake person number or breathing rate, she would not take further methods to compromise the true one. In this paper, we thus develop a novel scheme against CSI-based crowd counting and vital signs inference techniques. Specifically, we set up an *ambush location*, choose a fake person number or breathing rate, and convert it into a fake CSI. The transmitter then delivers the specified CSI to the ambush location by manipulating the transmitted wireless signals. As a result, the eavesdropper at the ambush location would infer the fake person count or breathing rate with the estimated CSI.

We first take the breathing inference system as an example, where the user

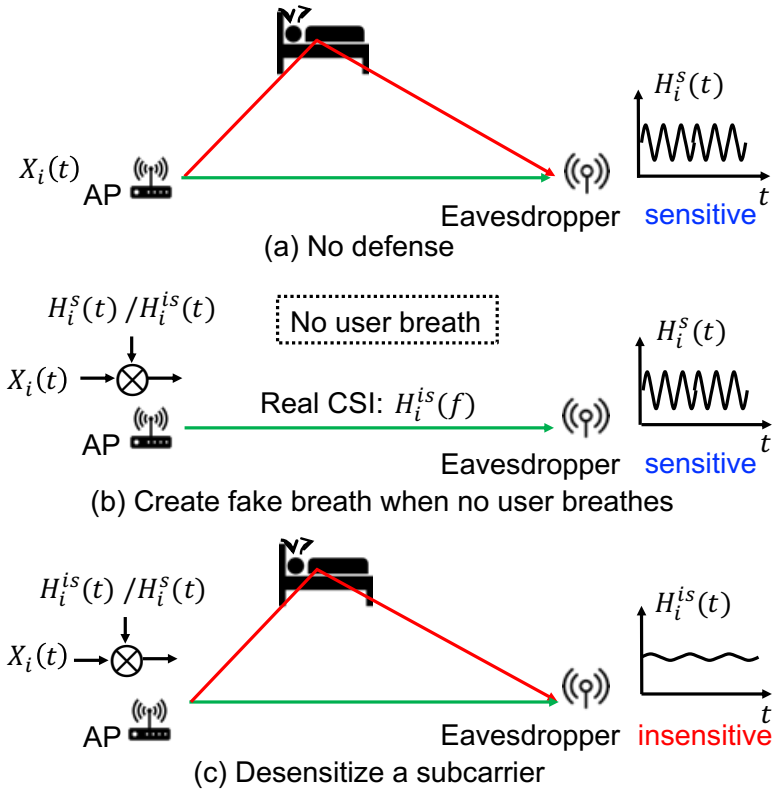


Figure 1.1: Creating a fake (sensitive or insensitive) CSI.

remains static. We observe that various subcarriers exhibit varying degrees of variance in the breathing rate. This variance is attributed to the effects of the Fresnel Zone [154], a region of space between the transmitter and receiver where the radio waves propagate. Generally, as the reflected and line-of-sight (LOS) signals interfere constructively or destructively, a receiver may observe enhanced or weakened signals. Such effects may vary for different subcarriers, which can be categorized into two groups: sensitive and insensitive. With respiration-induced body movements, sensitive subcarriers enable the receiver to observe large amplitudes (or variances) of CSI measurements, while insensitive subcarriers rarely show correlated fluctuations. Thus, the breathing rate can be determined via observations of sensitive subcarriers.

We give an example to illustrate our idea. Without loss of generality, we utilize a single subcarrier for discussion. For OFDM systems, a transmitter sends a publicly known pseudo-noise sequence $X_i(t)$, and the receiver estimates the channel frequency response $H_i(t)$ (i.e., subcarrier CSI) from the received, distorted copy $Y_i(t)$, i.e., $H_i(t) = \frac{Y_i(t)}{X_i(t)}$ [31, 45, 47, 49, 58]. If no defense strategy is enforced, as shown in Figure 1.2a, the eavesdropper (malicious receiver) can obtain the real CSI for the sensitive i^{th} subcarrier between itself and the AP, denoted with $H_i^s(t)$, which enables her to derive the breathing rate of the target user.

If there is no breathing activity, as shown in Figure 1.2b, the ith subcarrier should be insensitive and the true CSI is denoted with $H_i^{is}(t)$. However, the AP multiples the signal $X_i(t)$ with a coefficient $H_i^s(t)/H_i^{is}(t)$, and sends the resultant signal, which also goes through the real wireless channel. Consequently, the received signal becomes $X_i(t) \cdot H_i^s(t)/H_i^{is}(t) \cdot H_i^{is}(t) = X_i(t)H_i^s(t)$, and thus the eavesdropper obtains an estimated subcarrier CSI $H_i^s(t)$ (sensitive), with which the breath rate specified by the transmitter can be extracted.

Now consider the scenario in Figure 1.2c: the transmitter aims to hide the user's true breathing rate. Therefore, it multiples the signal $X_i(t)$ with a coefficient $H_i^{is}(t)/H_i^s(t)$. As a result, the eavesdropper obtains $X_i(t) \cdot H_i^{is}(t)/H_i^s(t) \cdot H_i^s(t) = X_i(t)H_i^{is}(t)$. The calculated subcarrier CSI then becomes $H_i^{is}(t)$, which means that such subcarrier is insensitive, causing failure of inferring the true breathing rate.

Unlike the breathing inference system where the user is stationary, people move randomly in the crowd counting system. Based on the observation that all subcarriers exhibit similar fluctuations due to the continuous movement and changing positions of people, all subcarriers can be considered sensitive. Therefore, we propose a defense scheme that manipulates the CSIs across all subcar-

riers, as shown in Figure 1.2b. The specific CSI is extracted from a pre-built profile consisting of collected CSI data for different numbers of moving people. As a result, the transmitter provides false information to the attacker, leading them to estimate incorrect crowd counts.

Our real-world experimental results show the proposed defenses can fool an eavesdropper into believing any desired breathing rate with an error of less than 1.2 bpm when the user lies on a bed in a bedroom and 0.9 bpm when the user sits in a chair in an office room. Furthermore, our proposed defense mechanisms can deceive an attacker into believing that there are moving individuals in an empty room with a probability of 100% and 100% for Support Vector Machine (SVM) and Decision Tree (DT) classifiers, respectively.

1.2 Vulnerabilities in Wireless Liveness Detection Systems

Liveness detection using wireless signals aims to detect whether human activity is real (from a live person present at the point of capture) or fake (from a spoof artifact or lifeless body part) by exploring the correlation between feeds of a sensor capturing human motion and co-existing wireless signals. Wireless liveness detection has proven successful in securing various practical systems [77, 78, 92, 110, 123], such as

- **Video liveness detection:** By launching a video spoofing attack (e.g., [12]), an adversary can hijack the camera feed to replay benign footage while stealing valuables (e.g., contents of a vault) without getting caught. A security guard can detect such attacks by observing mismatches between the live

video feeds and the captured wireless signals [92].

- **Voice liveness detection:** Voice controllable systems are especially vulnerable to spoofing attacks (e.g., with pre-recorded voice [34]) due to the inherent broadcast nature of voice transmissions. It can tell whether the voice command is generated by a live user by comparing the features extracted from both voice and wireless signals [110].

- **Human presence detection:** Wireless signals can be utilized to detect human presence by human breathing [98, 154, 169]. Wireless liveness detection can thus associate the detection of breathing with the user presence to combat replay attacks against voice assistants [123].

Human activity usually causes subtle environmental impacts unique to that human activity pattern, which can be observed by analyzing collected nearby wireless signals. As a result, wireless signals can be utilized to detect human activities and thus verify the authenticity of the captured data of another co-existing sensor such as video or microphone.

Mainstream WiFi systems are based on the Orthogonal frequency-division multiplexing (OFDM) technique, which utilizes multiple parallel narrowband subcarriers to encode a packet. Disturbances in wireless signals can be quantified by the *channel state information (CSI)* measurement [58], which describes how the wireless channel impacts the radio signal that propagates through the channel (e.g., amplitude attenuation and phase shift). CSI can be considered as an aptly initialed wireless analog to traditional "Crime Scene Investigation", measuring what has happened on a wireless channel [60]. Specifically, the variation of CSI time series has been widely utilized to identify the motion changes of a target user between a wireless transmitter and receiver pair.

In this work, however, we design a new *phantom-CSI* attack against all existing liveness detection built on the correlation between recorded human activity and co-existing CSI measurements. This attack accompanies traditional spoofing of video or microphone recorders by creating measurable CSI which exhibits corresponding spoofed human activity, bypassing the enforced wireless liveness detection system.

To understand the phantom-CSI attack, we first explain the impact of human activity on wireless signals. Generally, the presence of human and related body motion will result in significant changes in both amplitude and phase of the received wireless signals [97]. Accordingly, the received wireless signal (or CSI) at the receiver can thus capture the timing information (e.g., start or end time) and prominent frequency of occurrence of activities [92], and will exhibit a unique pattern corresponding to each activity [110]. For example, the repetitive (rhythmic) patterns of human breathing induce wave-like (sinusoidal-like) periodic change patterns over time in the CSI amplitudes at subcarrier level [98,154,169]. To fool a receiver to believe that an event occurs, the attacker needs to create a "virtual channel" that can exhibit a pattern similar to the real wireless channel affected by the event.

Figure 1.2 presents an example at the OFDM subcarrier level to illustrate how the attacker can build such a channel. Figure 1.2a shows a real scenario without an attack, where the transmitter sends a wireless signal and a human activity (e.g., walking) occurs between the transmitter and the receiver during the period from time t_1 to t_2. As a result, the received signal at the receiver would reflect the corresponding interference during the activity period $[t_1, t_2]$. Figure 1.2b shows an attack scenario, where there is no human activity happening between the attacker (i.e., a compromised transmitter) and the receiver, but the attacker

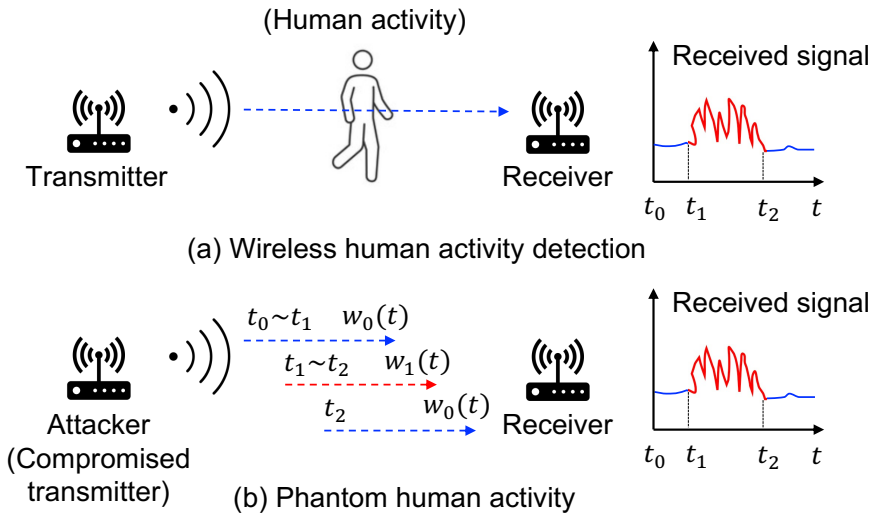


Figure 1.2: Crafting wireless signal affected by human activity.

aims to make the receiver detect some activities similar to that in Figure 1.2a. For each transmitted signal at time t, the attacker multiplies it with a corresponding coefficient, i.e., $w_0(t)$ when $t \in [t_0, t_1)$ or $t > t_2$, or $w_1(t)$ when $t \in [t_1, t_2]$, to mimic the distortion effect of the real subchannel in Figure 1.2a. Consequently, the receiver observes a distinguishable time series in period $[t_1, t_2]$ and incorrectly deduces that it is caused by the activity performed in Figure 1.2a.

Beyond this example of spoofing human activity in its absence, an attacker may have other goals, such as obscuring a particular human activity or portraying a different fake activity. Performing this general attack requires two technical solutions. First, the phantom motion must be encoded in the form of CSI for the receiver to estimate and map to the intended motion. Accordingly, we design a custom technique to convert an event into manipulated CSI of a wireless channel. Second, the transmitted signal crafted by the adversary is affected by the real wireless channel between herself and the receiver. Thus, the attacker requires a method to cancel the effect of the real channel, so that the receiver only observes

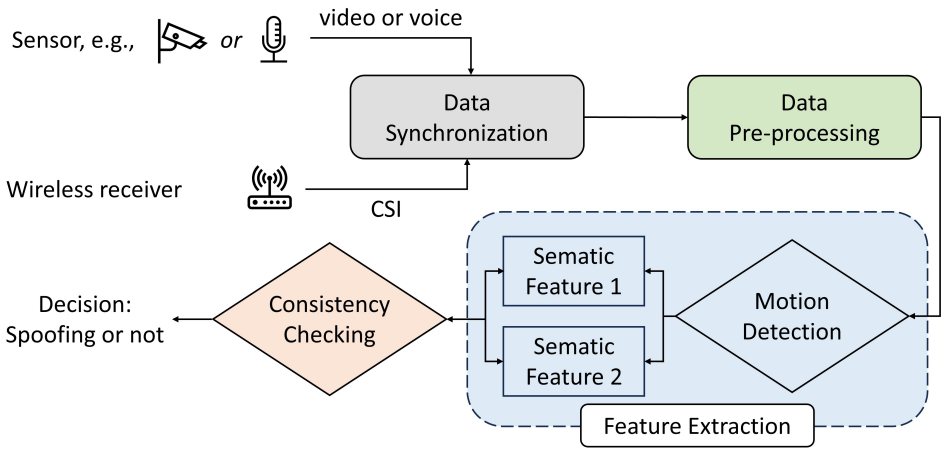


Figure 1.3: General structure of a wireless liveness detection system.

the phantom channel corresponding to spoofed activity. We address this challenge by reverse-engineering existing channel estimation algorithms for OFDM systems and pre-coding the original signal.

The discovered attack reveals that an attacker can create fake CSI data corresponding to spoofed voice or video signals. We conduct real-world experimental evaluations on top of Universal Software Radio Peripheral (USRP) X300 platforms. The experimental results show that an attacker camouflaged via our phantom CSI can inject spoofed video and voice to successfully bypass wireless liveness detection systems with a probability of 95.6% and 100%.

1.3 Summary of Contributions

The contributions of this dissertation are summarized below:

- Counteractive strategies for wireless breath and crowd inference attacks: To the best of our knowledge, we are the first to propose a deceptive strategy to defend against wireless vital signs and crowd inference attacks. By

reverse engineering existing techniques based on Channel State Information (CSI) for breathing rate and crowd inference, we have designed a customized scheme that transforms a selected breathing rate or crowd count into a fabricated CSI. Additionally, we have developed methods that allow an eavesdropper to estimate this fake CSI, leading them to derive the manipulated breathing rate or crowd count. We have implemented real-world prototypes for both the existing CSI-based inference techniques and our proposed defense mechanisms. Through experimentation with these prototypes, we assess the effectiveness of our defenses.

- Phantom-CSI attacks and corresponding defenses: We are the first to identify the vulnerability in wireless liveness detection systems through phantom-CSI attacks, which cause wireless signals and spoofed video/voice data to present fake human semantic information that appears genuine. We have developed a technique capable of crafting fake CSI based on human activities, which is then transmitted to the receiver via a realistic wireless channel. We have implemented real-world prototypes for both existing wireless video/voice liveness detection systems and our proposed attack techniques, validating the efficacy of the latter in compromising the former.

Chapter 2

Related Work

In this chapter, we review all existing techniques related to wireless breathing rate inference, crowd counting, human activity detection, and liveness detection.

2.1 Wireless Breathing Rate Inference Techniques

Generally, existing wireless breathing rate inference techniques fall into the following categories:

Ultra-wideband (UWB) radar based: The expansion and contraction of the chest cavity may create changes in the multipath profile of the transmitting signal, which can be captured with UWB impulse responses for breathing rate estimation [73,127,143]. UWB transmissions, however, spread over a large frequency bandwidth [52]. Also, the receiver structure for UWB is highly complex [93].

Doppler radar based: Doppler radar systems have been proposed to achieve breathing detection [11,36,94,95,125]. According to the Doppler theory, a target

with time-varying movement but zero net velocity will reflect the signal, whose phase is modulated in proportion to the displacement of the target [17]. A stationary person's chest and stomach can be thus regarded as a target. However, such Doppler radar based techniques suffer from the null point problem, which significantly degrades the measurement accuracy [61, 95, 181].

Frequency Modulated Continuous Wave (FMCW) radar based: An FMCW radar has also been utilized for breathing rate inference [4,10,142]. The breathing-induced body movement changes the signal reflection time. By analyzing such changes, the breathing rate can be extracted. However, high resolution (i.e., the minimum measurable change) requires a large swept bandwidth B as the resolution equals $\frac{C}{2B}$ [3], where C is the speed of light.

RSS-based: The changes in received signal strength (RSS) on wireless links have been successful in estimating breathing rate [1, 86, 119, 120]. For example, [1] puts a mobile device on the chest to collect RSS for inferring breathing rates. However, those methods are workable only when the target user stays close to the receiver. As an eavesdropper usually has a preference to be located far away to avoid being discovered, such RSS-based methods are not optimal.

*CSI-based:*RSS represents coarse channel information while CSI represents fine-grained channel information, consisting of subcarrier-level information. As a result, CSI is more sensitive to detecting breathing activity and the CSI-based approaches are able to capture breathing from a distance. Accordingly, CSI-based breathing rate inference has drawn increasing attention [98, 101, 102, 154, 161, 162, 187]. In particular, a recent empirical study [73] reveals CSI provides the most robust estimates of breathing rate compared with UWB radar or RSS.

2.2 Wireless Crowd Counting Techniques

Existing studies on crowd counting can be broadly categorized into the following groups:

RSS-based: It observed that the RSS value will be stable if there is no person present between a pair of transmitter and receiver. However, the RSS value exhibits a larger variance when a person crosses the wireless link, with this variance increasing as the number of people increases [113, 173, 184]. Correspondingly, several studies [113, 173, 184] have explored this relationship further. In detail, the study [113] derived a linear approximation formula to estimate crowd density based on the relationship between the number of people and the average/variance of RSS. SCPL [173] was the first to perform multi-subject counting and localization based on coarse-grained RSS. Moreover, [184] introduced a novel Wireless Sensor Network (WSN) application that uses RSS to estimate different crowd densities in subareas, utilizing the K-means algorithm for precision. However, these approaches require extensive deployment of sensor nodes and the construction of a fingerprint database, resulting in significantly high costs and substantial training efforts.

CSI-based: CSI-based approaches are motivated by the observation that CSI is highly sensitive to environmental variations. Therefore, a larger number of moving people will result in a greater CSI variance in the target area [62, 64, 99, 172]. For example, FCC [172] theoretically found a stable monotonic function to characterize the relationship between the number of moving people and the variation in the wireless channel. To improve accuracy, WiSpy [64] analyzes the CSI variation caused by human motion and identifies the number of moving targets by using machine learning algorithms, including KNN, stochastic gradient

descent (SGD), SVM, naive Bayes (NB), and decision tree (DT). Furthermore, [62] provides a human dynamics monitoring system by estimating the number of participants. It uses a semi-supervised learning approach based on non-linear regression to reduce training efforts. Another study [99] proposes a deep learning based approach to crowd counting using both CSI amplitude and phase.

2.3 Wireless Human Activity Detection

Due to the pervasive, low-cost, and non-intrusive sensing nature, wireless human activity sensing has drawn increasing attention [97]. The received signal strength (RSS) or channel state information (CSI) obtained at the receiver may vary with environmental human activity. RSS represents the average power in a received wireless signal over the whole power bandwidth. Different from RSS, which uses synthetic values, CSI offers fine-grained channel information by decomposing the entire channel measurement into subcarriers and obtains better human activity detection performance than other metrics (e.g., received signal strength) [73]. CSI contains both subcarrier-level amplitude and phase information. Extensive research efforts show that CSI amplitudes can capture various human activities, such as walking [158, 164, 186], breathing [98], gestures [145], and keystrokes [8,50,178]. Also, the work [163] exploits CSI phase difference data to monitor vital signs. Moreover, CSI amplitude and phase information can be employed together to achieve human activity detection [123, 124, 133, 185]. For example, the study [185] points out that human respiration cannot be detectable in all the locations when CSI amplitude or phase is used individually, and then proposes to use both phase and amplitude that are complementary to remove blind spots (where respiration detection experiences poor performance). Another

study [123] presents that compared with using CSI amplitude alone, leveraging CSI amplitude along with CSI phase improves the accuracy of breathing rate estimation.

2.4 Liveness Detection

With the rapid advance in speech synthesis and video editing methods, it becomes increasingly popular to replay tampered voices/videos [85, 88, 156]. Specifically, in an audio replay attack, a recording of a target speaker's voice is replayed to a voice recognition system in place of genuine speech [88]; in a video spoofing attack, an attack can play back a clip of footage to cover up a crime [85]. With such spoofing techniques, attackers may bypass voice authentication or video monitoring, and even stealthily inject illegal voice commands or conduct malicious activities. To deal with these spoofing attacks, liveness detection is widely applied to differentiate the alive and present data (originating from live users) from forged data that are pre-recorded, concatenated, or synthesized by the attacker. Liveness detection against those spoofing attacks mainly includes the following three categories.

Intrinsic feature-based: Non-live representations often miss some intrinsic features in the corresponding live source. For example, a smartphone's loudspeaker usually presents strongly attenuated frequency responses in the low part of the spectrum [144], but it often has a high false acceptance rate to use this observation for liveness detection. Also, [191] uses the unique time-difference-of-arrival (TDoA) dynamic (i.e., the TDoA changes in a sequence of phoneme sounds to the phone's two microphones) for liveness detection, as it does not exist under replay attacks. Nevertheless, this method is not applicable to a device with only

one microphone.

Another sensor-assisted: Liveness detection can also be achieved by combining a microphone/camera with other co-existing sensors [26, 76, 131, 190]. For example, [76] correlates sound and breathing-induced chest motion (obtained via a gyroscope) to build a liveness detection system; [131] uses earbuds to measure the air pressure in the ear canal for voice liveness detection. These two methods, however, require the user to wear a chest-mounted gyroscope or earbuds. [190] leverages a speaker to emit inaudible signals, and exerts a microphone to record the reverberant signals to distinguish bone-conducted vibrations from air-conducted voices for liveness detection. Unfortunately, not all loudspeakers can emit ultrasound, which limits its practicality.

Wireless-based: There are emerging research efforts (e.g., [78, 92, 109, 110, 116, 123, 141, 193]) performing liveness detection leveraging wireless sensing due to its non-invasive and device-free nature, as well as the ubiquitous deployment of wireless infrastructures. In particular, [116] uses the ratio of the energy in motion affected bands (35-60 Hz) over the entire mmWave radar spectrogram as an indicator for liveness; [78, 92] develops techniques to detect video replay or forgery attacks using CSI extracted from wireless signals near the camera spot; [109] utilizes CSI to capture mouth motions, which can help distinguish authentic voice command from a spoofed one; [123] exploits the synchronized changes in voice and breathing to detect voice replay attacks. Our attack can make the CSI convey the same event semantic information with the spoofed video or voice signals, compromising those wireless liveness detection systems.

Chapter 3

Proactive Ambush Tactic Against Wireless Breath and Crowd Inference

This chapter [1] introduces the existing attacks on wireless breathing rate inference and crowd inference. Aiming at these attacks, this chapter also proposes corresponding proactive ambush tactics.

3.1 Preliminaries

In this section, we impart preliminary knowledge about the Fresnel Zone model and the general method used by existing work using CSI to infer breathing rates.

[1] This chapter was published in He, Q., Yang, E., Fang, S., Zhao, S. (2023). HoneyBreath: An Ambush Tactic Against Wireless Breath Inference. In: Longfei, S., Bodhi, P. (eds) Mobile and Ubiquitous Systems: Computing, Networking and Services. MobiQuitous 2022. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 492. Springer, Cham. https://doi.org/10.1007/978-3-031-34776-4_12. Used with permission.

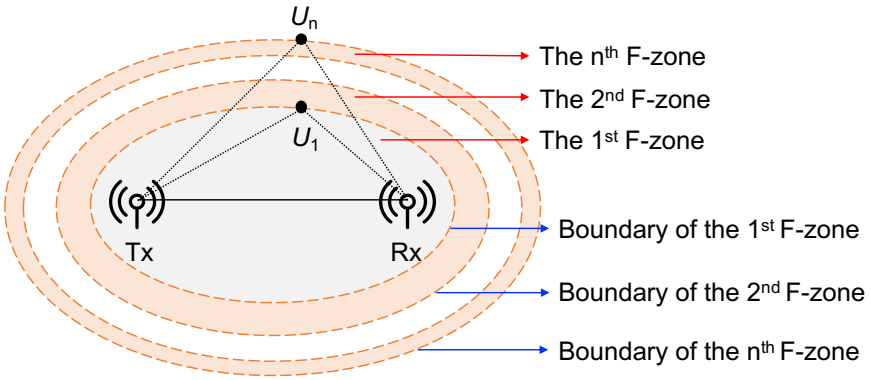


Figure 3.1: Demonstration of Fresnel Zones.

3.1.1 Fresnel Zone

In the context of wireless signal propagation, Fresnel Zones refer to concentric ellipses with the transmitter (Tx) and receiver (Rx) at two focal points, and denote regions of different wireless signal propagation strengths between the pair of communicators, as shown in Figure 3.1. For a given radio wavelength λ, each ellipse can be constructed by ensuring

$$|\mathrm{Tx}, U_n| + |\mathrm{Rx}, U_n| - |\mathrm{Tx}, \mathrm{Rx}| = n\lambda/2, \tag{3.1}$$

where U_n is a point in the nth ellipse, and $|u, v|$ denotes the Euclidean distance between two points u and v. The innermost ellipse is the first Fresnel Zone, representing the region where the LOS signals can pass through. The nth (when $n \geq 2$) Fresnel Zone is the region between the $(n-1)^{\text{th}}$ and n^{th} ellipses.

The received signal at Rx is a linear combination of reflected and LOS signals. The distance difference ΔD (i.e., $n\lambda/2$) between the two paths generates a phase difference of $\frac{\Delta D}{\lambda} \cdot 2\pi = n\pi$ between the two signals. As the phase shift introduced by the reflection is π [154], the total phase difference $\Delta\phi$ between reflected and LOS signals equals $(n+1)\pi$. Thus, if n is even, we obtain $\Delta\phi \bmod 2\pi = \pi$,

causing the two signals to arrive at Rx to have opposite phases and destructively interfere with each other. In contrast, we have $\Delta\phi \bmod 2\pi = 0$ if n is odd, i.e., both signals have the same phase and constructively interfere with each other to form a boosted signal. The Fresnel Zone model can thus help reveal the signal change pattern (i.e., sensitive or insensitive) in each subcarrier (with different waveforms) caused by respiration-induced body movement [154].

3.1.2 CSI-based Breathing Rate Inference

Existing CSI-based breathing rate inference schemes [73, 98, 154] usually utilize three steps to infer breathing rates, namely, CSI pre-processing, subcarrier selection, and breathing cycle extraction. The first phase removes outliers and noise from the CSI to improve its reliability. As discussed earlier, each subcarrier may be sensitive or insensitive to respiration due to the constructive or destructive interference effect of LOS and reflected signals. The second phase picks up sensitive subcarriers for breathing rate inference. A sensitive subcarrier often exhibits a sinusoidal-like periodic change pattern over time in the CSI amplitudes, which corresponds to periodic breathing. In the third phase, the peak-to-peak time interval of sinusoidal CSI amplitudes can be then extracted as the breathing cycle, with which, the breathing rate can be calculated.

3.1.3 CSI-based Crowd Counting Inference

Existing studies on CSI-based crowd counting approaches [27, 87, 172] utilize existing WiFi infrastructure for crowd classification in indoor scenarios. The key idea is that an increase in the number of moving people introduces larger multipath variations, resulting in greater CSI variation over time. In this way, based on

the measurement of how CSI varies over time, the number of moving people can be estimated. In general, there are three key phases: CSI pre-processing, feature extraction, and crowd classification. To obtain the CSI measurements caused by moving people, CSI pre-processing phase removes redundant components, such as noise, from the CSI data. Based on such processed data, distinct features (e.g., mean, standard deviation, maximum, and minimum of CSI amplitude) are extracted and then fed into a classifier (e.g., SVM, DT) to output the estimated number of moving people.

3.2 Attack Model and Assumptions

We consider a general scenario, where an attacker only uses a wireless receiver to launch a breathing rate inference attack or crowd inference attack, as she has a preference to take advantage of an existing wireless transmitter to make the attack stealthier [98, 172]. The transmitter (i.e., defender) is benign and aims to hide true breathing rates or person count and inject fake ones into the eavesdropper.

We assume that the receiver (i.e., attacker) attempts to find a position that enables her to eavesdrop on the breathing rate or person count, which is a common strategy [9]. We borrow the idea from a long-established military tactic – ambush: set up one or multiple ambush locations where an attacker may appear and be trapped. We further assume that the transmitter is able to obtain actual CSI measurements between itself and an ambush location. This can be achieved by estimating the CSI measurements from wireless signals emitted by a helper node placed at the ambush location.

3.3 Ambush Design for Breath Inference Attacks

3.3.1 Overview

To lay an ambush, the transmitter first selects an ambush location and arbitrarily specifies a fake breathing rate to fool the attacker entering the ambush. The locations where an eavesdropper may appear with the highest probabilities can be determined via eavesdropper tracking techniques (e.g., [23]) and ambush locations can be then deployed along the eavesdropper's possible route.

The transmitter then enters the *planning* phase, which consists of two parallel tasks: (1) determining sensitive subcarriers; and (2) converting a specified breathing rate into an artificial CSI. We utilize a binary decision variable α_i to indicate the sensitivity of the ith subcarrier, with 1 denoting sensitive while 0 showing insensitive. The sensitivities of all N subcarriers can be represented by a vector $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_N]^T$. Since insensitive subcarriers do not contribute to the breathing rate inference, there is no need to manipulate their CSIs.

The next phase is *disturbance manipulation*. For signals on sensitive subcarriers, the transmitter aims to make the attacker estimate the converted CSI. As any transmitting signal has to go through the real wireless channel, the transmitter then applies a module of desensitizing subcarriers to remove the real impact of corresponding wireless sub-channels, and also crafts the artificial disturbance on these originally sensitive subcarriers for the attacker to observe. Finally, the transmitter combines the crafted signals on sensitive subcarriers with unchanged signals on insensitive subcarriers and transmits the aggregated signal out.

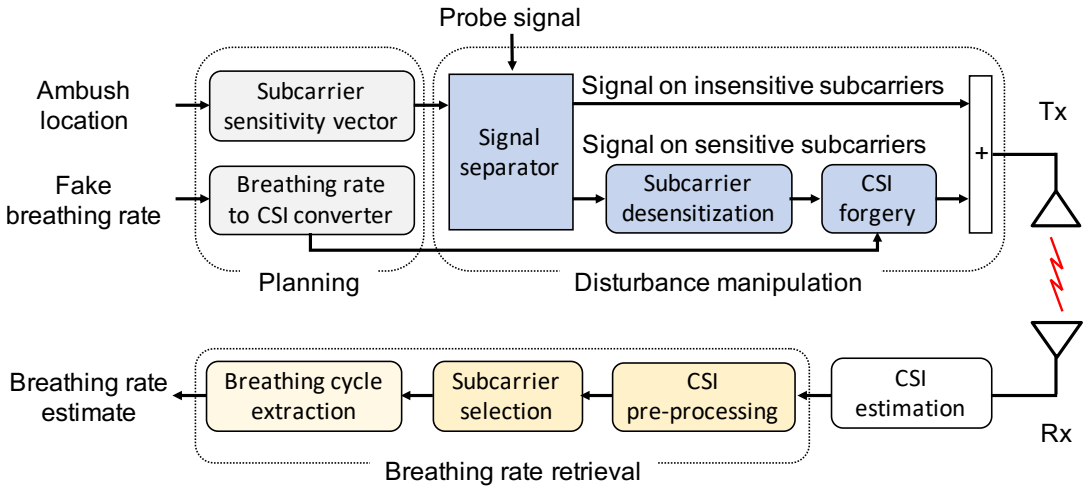


Figure 3.2: Flow chart of the proposed ambush tactic.

Consequently, the attacker infers breathing rate with estimated CSI by performing the general *breathing rate retrieval* process. Figure 3.2 shows the flow chart of the proposed ambush tactic.

3.3.2 Planning Phase

3.3.2.1 Obtaining Subcarrier Sensitivity

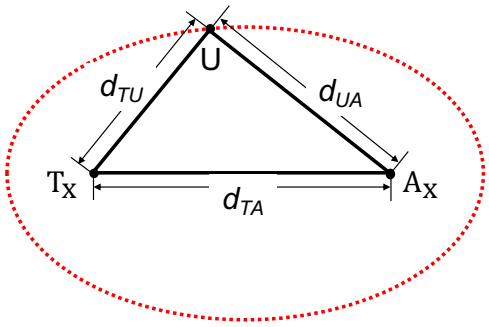


Figure 3.3: Selecting an ambush location.

As shown in Figure 3.3, T_x, U, and A_x denote the transmitter, the user, and an ambush location, respectively. A wireless signal sent by T_x travels on two

paths, the LOS path and the reflection one. The distance difference Δd between the two paths is $\Delta d = d_{TU} + d_{UA} - d_{TA}$.

Let λ_i denote the wavelength of the ith subcarrier with frequency f_i, i.e., $\lambda_i = c/f_i$, where c is the speed of light. Correspondingly, the phase difference $\Delta\theta_i$ (between signals arrived at $\mathrm{A_x}$ through the two paths) equals the sum of the respective phase shifts caused by Δd and the reflection phenomenon, i.e., $\Delta\theta_i = \frac{2\pi\Delta d}{\lambda_i} + \pi$. We perform a modulus 2π operation on $\Delta\theta_i$ and obtain a phase difference $\Delta\theta_i'$ within the range of $[0, 2\pi)$, i.e., $\Delta\theta_i' = \Delta\theta_i \pmod{2\pi}$.

Based on the Fresnel Zone theory [154], if $\Delta\theta_i'$ is close to 0 or 2π, the ith subcarrier is sensitive, i.e., when $\Delta\theta_i' \in [0, \pi/2) \cup (3\pi/2, 2\pi)$, we obtain the binary decision variable $\alpha_i = 1$. On the other hand, if $\Delta\theta_i'$ approaches to π, this subcarrier becomes insensitive, i.e., $\alpha_i = 0$ for $\Delta\theta_i' \in [\pi/2, 3\pi/2]$. The relationship between α_i and $\Delta\theta_i'$ can be then denoted as $\alpha_i = \frac{|\Delta\theta_i' - \pi|}{\pi/2}$, where x denotes the floor function, representing the largest integer less than or equal to x.

3.3.2.2 Converting Breathing Rate to CSI

Breathing rate to CSI conversion is the process of translating a selected breathing rate into a subcarrier CSI. It has been observed that periodic chest and stomach movement caused by respiration would make the amplitude of CSI on a sensitive subcarrier present a sinusoidal-like pattern over time [98, 102, 154]. We thus model the respiration-induced CSI amplitude stream on a sensitive subcarrier as a sinusoidal wave.

Let f_b denote the specified respiration frequency (Hz), so the corresponding breathing rate equals $60 \cdot f_b$ (bpm). We then convert it into a subcarrier CSI $W_b(t)$, which can be then denoted with $|W_b(t)|e^{j\varphi(t)}$, where $|W_b(t)|$ and $\varphi(t)$ represent amplitude and phase, respectively. Since the phase could be distorted due to an

unknown time lag caused by the non-synchronized transmitter and receiver [129], most studies only use the amplitude to characterize the wireless channel [149] and extract breathing rate [98, 102, 154]. We also explore CSI amplitude and refer to it as just "CSI" in the following. In terms of $\varphi(t)$, it has no impact on breathing rate inference and we omit it for the sake of simplicity.

With the sinusoidal model, the CSI envelope at time t can be denoted by

$$|W_b(t)| = a \cdot sin(2\pi f_b t + \beta) + m + \mathcal{N}_0, \tag{3.2}$$

where a, β, m, and $\mathcal{N}_0$ are the amplitude, initial phase, constant shift (which defines a mean level) of the sinusoidal wave, and the additive noise. In turn, with such a CSI envelope, the attacker can infer the breathing rate as $60 \cdot f_b$.

Formation of the Specified OFDM CSI: The specified CSI for an OFDM system with N subcarriers can be denoted with $\mathbf{W}(t) = [W_1(t), W_2(t), \cdots, W_N(t)]$. Let $\mathcal{S} = \{s_1, s_2, \ldots, s_K\}$ and $\bar{\mathcal{S}} = \{p_1, p_2, \ldots, p_{K'}\}$ denote the sets formed by the indexes of the sensitive and insensitive subcarriers, where $K + K' = N$. For $i \in \mathcal{S}$, we enable $W_i(t) = W_b(t)$; for $i \in \bar{\mathcal{S}}$, we have $W_i(t) = H_i(t)$ (i.e., no manipulation is required), where $H_i(t)$ is the original CSI of the ith subcarrier.

3.3.3 Disturbance Manipulation

The transmitter can utilize a multiply-accumulate (MAC) process to generate desired artificial disturbance, as shown in Figure 3.4. Specifically, the public training sequence $\mathbf{X}(t)$ is encoded into N subcarrier signals by a serial-to-parallel (S/P) converter module, represented with $[X_1(t), X_2(t), \cdots, X_N(t)]^T$. We use $\mathbf{J}$ to represent an $N \times 1$ vector of all 1's. Thus, after the signal separator, the original N subcarrier signals will be divided into two groups: $\mathbf{S}(t) = \text{diag}(\boldsymbol{\alpha}) \cdot \mathbf{X}(t)$ and

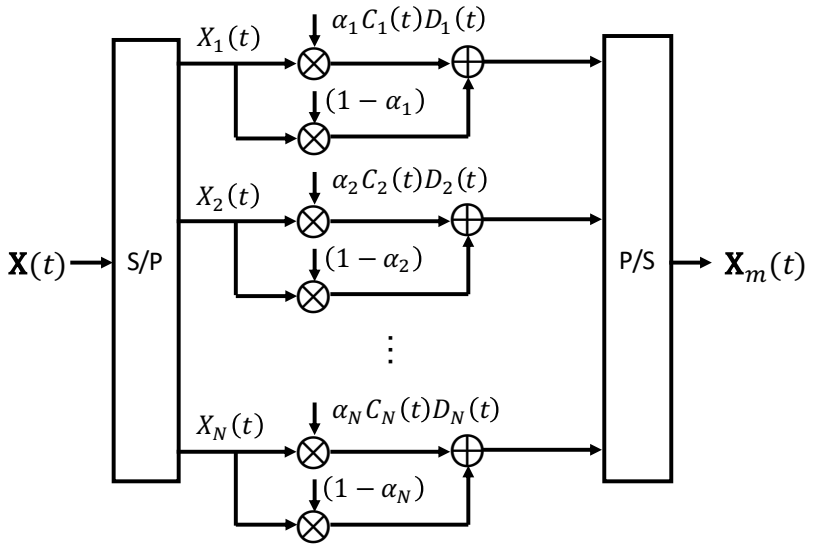


Figure 3.4: An MAC process.

$\mathbf{IS}(t) = \text{diag}(\mathbf{J}-\boldsymbol{\alpha})\cdot\mathbf{X}(t)$, denoting signals on sensitive and insensitive subcarriers, respectively, where $\text{diag}(\mathbf{V})$ denotes a square diagonal matrix with the elements of vector $\mathbf{V}$ on the main diagonal.

Signals on sensitive subcarriers would then go through two modules: subcarrier desensitization and CSI forgery. The former module with the coefficient vector $\mathbf{C}(t) = [C_1(t), C_2(t), \cdots, C_N(t)]$ aims to cancel the original channel impact, so that the real respiration-induced channel disturbance (i.e., the real breathing rate) can be hidden for the attacker. Accordingly, we have $C_i(t) = H_i^{-1}(t)$ if the ith subcarrier is sensitive, i.e., $i \in \mathcal{S}$, and set $C_i(t) = 0$ for $i \in \bar{\mathcal{S}}$. The latter module with a coefficient vector $\mathbf{D}(t) = [D_1(t), D_2(t), \cdots, D_N(t)]$ would add the effect of the artificial CSI for the attacker to estimate, where the forged subcarrier CSI $D_i(t) = W_i(t)$ if $i \in \mathcal{S}$ and we set $D_i(t) = 0$ for $i \in \bar{\mathcal{S}}$.

Finally, signals on originally sensitive and insensitive subcarriers are concatenated through a parallel-to-serial (P/S) converter module to form OFDM symbols to send via the realistic wireless channel. The resulting transmitting signal $\mathbf{X}_m(t)$

can be represented by

$$\mathbf{X}_m(t) = \text{diag}(\mathbf{D}(t)) \cdot \text{diag}(\mathbf{C}(t)) \cdot \mathbf{S}(t) + \mathbf{IS}(t). \tag{3.3}$$

Let $\mathbf{H}(t) = [H_1(t), \cdots, H_N(t)]^T$ denote the true OFDM CSI. The received signal at the attacker thus becomes $\mathbf{R}_m(t) = \text{diag}(\mathbf{X}_m(t)) \cdot \mathbf{H}(t)$, where we omit the noise term for the sake of simplicity. The attacker estimates CSI with the received signal and the public training sequence, i.e., $\mathbf{R}_m(t) = \text{diag}(\mathbf{X}(t)) \cdot \hat{\mathbf{H}}(t)$, where $\hat{\mathbf{H}}(t) = [\hat{H}_1(t), \cdots, \hat{H}_N(t)]^T$ represents the estimated CSI. Consequently, we have

$$\begin{aligned} \hat{H}_i(t) &= \alpha_i \cdot \frac{X_i(t) C_i(t) D_i(t)}{X_i(t)} \cdot H_i(t) + (1 - \alpha_i) \cdot H_i(t) \\ &= \alpha_i \cdot D_i(t) + (1 - \alpha_i) \cdot H_i(t) = W_i(t). \end{aligned} \tag{3.4}$$

This demonstrates that with the disturbance manipulation, when the ith subcarrier is sensitive, the transmitter is able to make the attacker obtain a fake subcarrier CSI $W_i(t)$ specified by itself in the planning phase. Meanwhile, if the ith subcarrier is insensitive, it is still observed as insensitive, i.e., the corresponding estimated subcarrier CSI equals the real value $H_i(t)$. This is because the transmitter does not manipulate signals on insensitive subcarriers.

3.3.4 Breathing Rate Retrieval

3.3.4.1 CSI Pre-processing

CSI pre-processing, consisting of outlier removal and noise reduction, aims to make the collected CSI reliable. The imperfect CSI can be caused by non-respiratory environmental change or hardware imperfections.

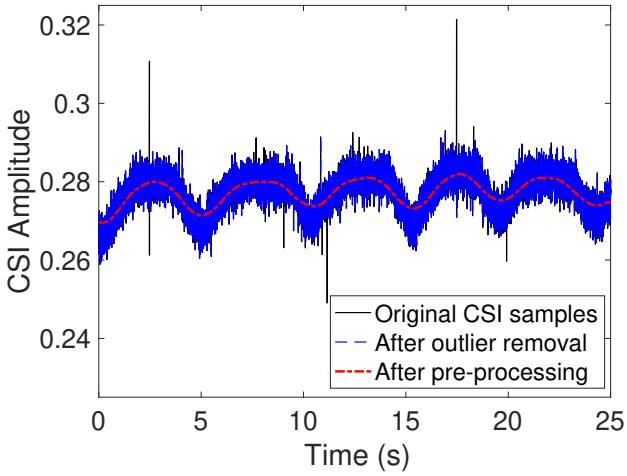


Figure 3.5: CSI pre-processing.

Hampel filter is a classical technique to remove outliers (i.e., samples that significantly differ from neighboring ones) in a given series [33, 102]. As the collected CSI may have abrupt changes that are not caused by respiration, a Hampel filter is enforced to remove those outliers. It is observed that the CSI variations caused by the chest and stomach movement usually lie at the low end of the spectrum. Thus, we further adopt the moving average filter, which is optimal for reducing high-frequency noise while retaining a sharp step response [137]. Figure 3.5 illustrates an example of CSI pre-processing. It can be seen that the outliers and high-frequency noise are effectively removed.

3.3.4.2 Subcarrier Selection

Empirically, the CSI variance of a sensitive subcarrier is usually more than one order of magnitude larger than that of an insensitive subcarrier. This observation implies a threshold-based approach to distinguish the two types of subcarriers. Specifically, when there is no breathing activity, the average CSI variance σ^2 across all subcarriers can be measured, called *reference variance*, which will be

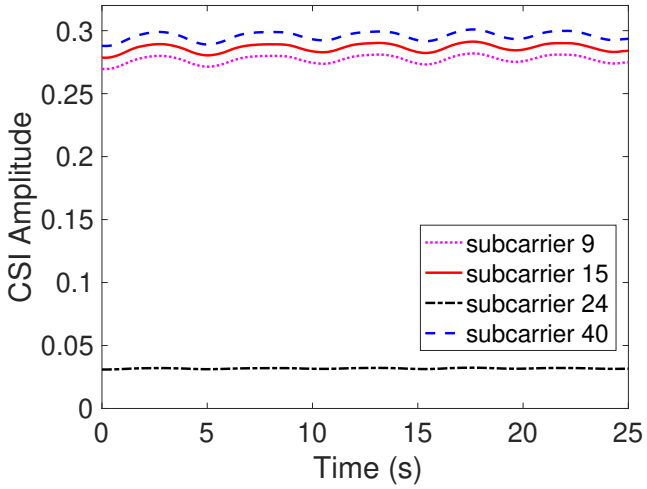


Figure 3.6: Subcarrier sensitivity.

then utilized as the threshold to determine the sensitivity of each subcarrier. Let v_i^2 denote the CSI variance for the ith subcarrier. If $\log_{10}(v_i^2/\sigma^2) < 1$ holds, we regard that the variance is caused by noise and the subcarrier is insensitive; otherwise, this subcarrier is sensitive. If CSI variances on all subcarriers have the same order as the reference variance, all subcarriers are insensitive (i.e., no breathing activity is detected).

Figure 3.6 plots the CSIs observed on 4 different subcarriers. In this example, we can see that subcarrier 24 has a quite flat CSI which rarely discloses any useful information about the breathing activity, while the CSIs of the remaining subcarriers show evident periodical fluctuations. Accordingly, we can determine that subcarriers 9, 15, and 40 are sensitive, while subcarrier 24 is insensitive.

3.3.4.3 Breathing Cycle Identification

The CSI on a sensitive subcarrier often shows a sinusoidal pattern correlated with breathing activities. To obtain a breathing cycle, we can thus compute the inter-peak interval (i.e., the time between successive peaks) of the sinusoidal CSI.

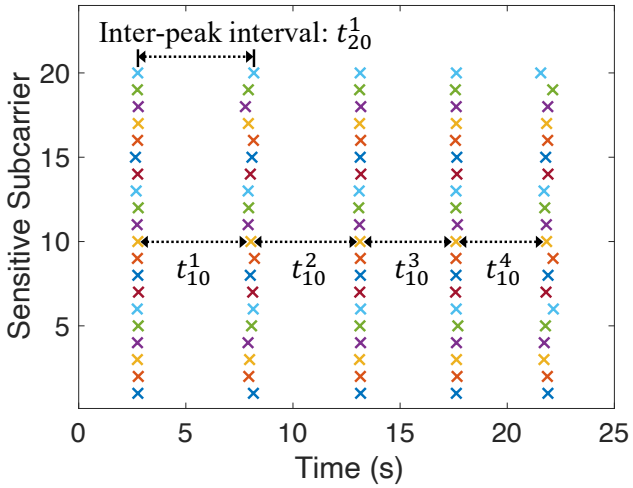


Figure 3.7: Local peaks.

Intuitively, the first derivative of a peak switches from positive to negative at the peak maximum, which can be used to localize the occurrence time of each peak. However, there may exist fake peaks caused by noise and consequently false zero-crossings. Motivated by the fact that a person usually cannot breathe beyond a certain frequency, a fake peak removal algorithm can be developed. Specifically, if the calculated interval between the current peak with the previous one is less than $60/R_{max}$ (seconds), where R_{max} (bpm) denotes the maximum possible breathing rate, this peak will be labeled as a fake one and then removed.

Figure 3.7 shows all detected local peaks on 20 sensitive subcarriers during 25 seconds. The breathing rate is calculated as 12.7 bpm for this example.

3.3.4.4 Inferring Multi-user Breathing Rates

For the multi-user scenario, we use the power spectral density (PSD) [98] to identify the frequencies with strong signal power in the frequency domain. Normally, each breathing signal from one person contributes to one evident peak in the obtained PSD [152]. The PSD on the ith sensitive subcarrier with L samples can be

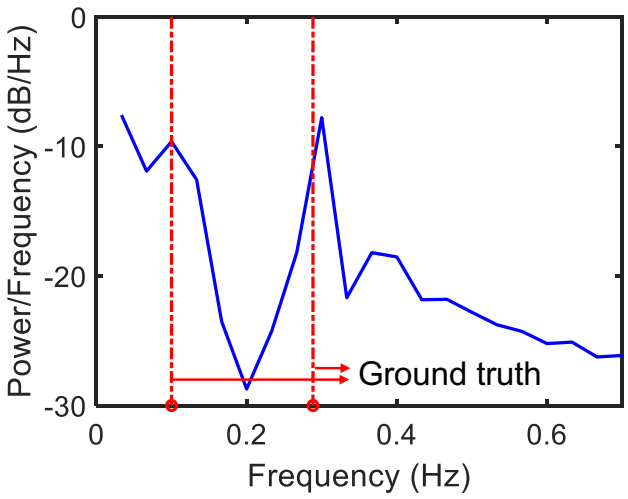


Figure 3.8: Peaks in PSD.

obtained by $PSD_i = 10\log_{10}\frac{|FFT(H_i)|^2}{L}$, where H_i is the vector of CSI amplitude on the ith subcarrier.

When there are two users, the two strongest peaks in the PSD would indicate their breathing rates, as in an example shown in Figure 3.8. The ground truths of two users' breathing rates are 6.0 and 17.3 bpm (corresponding to 0.10 ad 0.29 Hz); the estimated breathing rates based on the first two strongest peaks are 6.0 and 18.0 bpm (i.e., 0.10 and 0.30 Hz), showing that the estimation of two-user breathing rates is accurate.

3.3.5 From Point Ambush to Area Ambush

With more deployed ambush locations, the probability that an eavesdropper happens to be at any of them would be higher. Meanwhile, it helps to defend against multiple collaborative attackers, each of which searches for opportune eavesdropping locations.

3.3.5.1 Setting up Two Ambush Locations

The transmitter with two antennas can set up two ambush locations. Let $\mathbf{H}_{sr}(t)$ $(s,\ r \in \{1,2\})$ denote the overall CSI between the sth transmit antenna and the rth ambush location. The corresponding subcarrier sensitivity vector is represented by $\boldsymbol{\alpha}_{sr} = [\alpha_{sr}^1, \cdots, \alpha_{sr}^N]$, which can be pre-obtained with the method proposed in Section 3.3.2.1. At each ambush location, the received signal is the superposition of two signals, each from a different transmit antenna. If at least one of the two subcarriers between the respective transmit antenna and the rth ambush location is sensitive, we regard that this overall subcarrier between the transmitter and the rth ambush location is sensitive. Mathematically, let $\boldsymbol{\alpha}_r = [\alpha_r^1, \cdots, \alpha_r^N]$ denote the resultant subcarrier sensitivity vector of the transmitter for the rth ambush location, and $\alpha_r^i = \alpha_{1r}^i \vee \alpha_{2r}^i$. On the other hand, it may arouse suspicion of two colluding eavesdroppers if the breathing rates they infer separately are different. Thus, the transmitter should enable both ambush locations to observe the same breathing rate, i.e., the manipulated CSIs at corresponding sensitive subcarriers should be equal. If a subcarrier at either ambush location is sensitive, we then regard that the overall subcarrier between the transmitter and the two ambush locations is sensitive. Similarly, let $\boldsymbol{\alpha} = [\alpha^1, \cdots, \alpha^N]$ denote the subcarrier sensitivity vector of the transmitter for the two ambush locations, and $\alpha^i = \alpha_1^i \vee \alpha_2^i$.

Let $W(t)$ denote the fake CSI which is converted with a specified breathing rate. The transmitter aims to make the estimated CSI on sensitive subcarriers at each eavesdropper to be equal to $W(t)$.

As discussed in Section 3.3.3, the transmitting signals on sensitive subcarriers will be first desensitized and then multiplied with the forged CSI before being

sent out. In this scenario, let $H_{sr}^i(t)$ denote the CSI on i^{th} subcarrier between the sth transmit antenna and the rth ambush location. Thus, in terms of the coefficient vector $\mathbf{C}_s(t) = [C_s^1(t), \cdots, C_s^N(t)]$ for subcarrier desensitization at the sth transmit antenna, if $\alpha^i = 0$ (i.e., the ith subcarrier between the transmitter and the two ambush locations is insensitive), we set $C_s^i(t) = 0$, otherwise, we have $C_1^i(t) = \frac{H_{21}^i(t) - H_{22}^i(t)}{\zeta^i}$ and $C_2^i(t) = \frac{H_{12}^i(t) - H_{11}^i(t)}{\zeta^i}$, where $\zeta^i = H_{21}^i(t)H_{12}^i(t) - H_{22}^i(t)H_{11}^i(t)$. Also, the coefficient vector for the CSI forgery module at each transmit antenna is $\mathbf{D}(t) = [D_1(t), \cdots, D_N(t)]$, where we set $D_i(t) = 0$ if $\alpha^i = 0$ and have $D_i(t) = W(t)$ if $\alpha^i = 1$.

We rewrite Equation 3.3 and the transmitting signal $\mathbf{X}_m(t) = [\mathbf{X}_1(t), \mathbf{X}_2(t)]^T$ after manipulation becomes

$$\mathbf{X}_m(t) = \begin{bmatrix} \text{diag}(\mathbf{D}(t)) \cdot \text{diag}(\mathbf{C}_1(t)) \cdot \mathbf{S}(t) + \mathbf{IS}(t) \\ \text{diag}(\mathbf{D}(t)) \cdot \text{diag}(\mathbf{C}_2(t)) \cdot \mathbf{S}(t) + \mathbf{IS}(t) \end{bmatrix}. \tag{3.5}$$

The transmitting signal $\mathbf{X}_m(t)$ would go through the realistic wireless channel. At the ambush location side, the received signal and the public training sequence will be then utilized to estimate CSI. Let $\hat{\mathbf{W}}_1(t)$ and $\hat{\mathbf{W}}_2(t)$ denote the estimated CSIs at the two ambush locations. We thus obtain

$$\hat{W}_r^i(t) = \alpha^i \cdot W(t) + (1 - \alpha^i) \cdot (H_{1r}^i(t) + H_{2r}^i(t)). \tag{3.6}$$

This implies the success of setting up two ambush locations simultaneously.

3.3.5.2 General Scheme for Area Ambush

The transmitter can deploy κ ambush locations with κ antennas. We consider colluding eavesdroppers and need to guarantee the breathing rate inferred by

each eavesdropper at any ambush location stays the same.

The sensitivity of the ith subcarrier between the sth transmit antenna and the rth ambush location can be represented by α_{sr}^{i} $(s,\ r \in \{1, 2, \cdots, \kappa\})$. Meanwhile, let α_{r}^{i} denote the overall sensitivity of the ith subcarrier between the transmitter and the rth ambush location, i.e., $\alpha_{r}^{i} = \alpha_{1r}^{i} \vee \alpha_{2r}^{i} \cdots \vee \alpha_{\kappa r}^{i}$. Thus, in terms of the subcarrier sensitivity vector $\boldsymbol{\alpha}$ of the transmitter for all κ ambush locations, we have $\alpha^{i} = \alpha_{1}^{i} \vee \alpha_{2}^{i} \cdots \vee \alpha_{\kappa}^{i}$. Let $\mathbf{X}(t) = [\mathbf{X}_1(t), \cdots, \mathbf{X}_\kappa(t)]^T$ denote the manipulated signal sent by κ transmit antennas. The transmitter aims to make the estimated CSI at each ambush location be equal to the specified fake CSI, i.e., $\hat{\mathbf{W}}_r(t) = \mathbf{W}(t)$. Similarly, each transmit antenna utilizes the same coefficient vector $\mathbf{D}(t)$ for the CSI forgery module.

Accordingly, we can then solve the manipulated signal $\mathbf{X}_m(t)$, and rewrite Equation 3.5 as

$$\mathbf{X}_m(t) = \begin{bmatrix} \text{diag}(\mathbf{D}(t)) \cdot \text{diag}(\mathbf{C}_1(t)) \cdot \mathbf{S}(t) + \mathbf{IS}(t) \\ \vdots \\ \text{diag}(\mathbf{D}(t)) \cdot \text{diag}(\mathbf{C}_\kappa(t)) \cdot \mathbf{S}(t) + \mathbf{IS}(t) \end{bmatrix}, \tag{3.7}$$

where $\mathbf{C}_s(t)$ is the coefficient vector for the subcarrier desensitization module at the sth transmit antenna.

Equation 3.7 has κ unknowns ($\mathbf{C}_1(t)$ to $\mathbf{C}_\kappa(t)$). As the number of transmit antennas equals the number of unknowns, the linear system formed by Equation 3.7 has a unique solution. It demonstrates when the transmitter is able to set the coefficient vector for the subcarrier desensitization module at the sth transmit antenna with the computed $\mathbf{C}_s(t)$, the goal of deploying κ simultaneous ambush locations can be achieved.

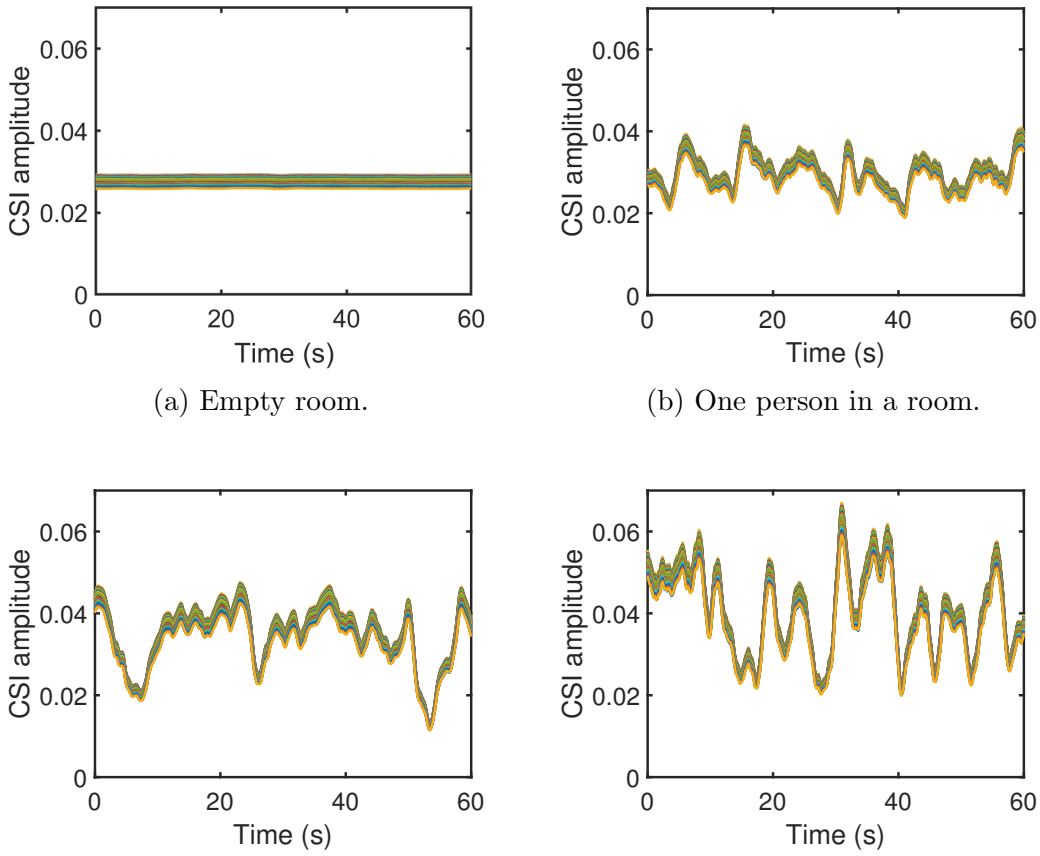
3.3.6 Security Analysis

The proposed scheme is known by the eavesdropper. One concern is whether the eavesdropper can distinguish ambush locations or even indirectly compute the real CSI of sensitive subcarriers (to infer the true breathing rate).

Ambush Indistinguishability: With the Fresnel Zone principle, CSI-based breathing rate inference works at certain locations, while its performance may deteriorate greatly at other locations [24]. Thus, when the eavesdropper moves out of the ambush location, though she cannot detect the breathing rate as when she is at the ambush location, she is still unable to distinguish this case from the normal one when the ambush scheme is not enforced. Such ambush indistinguishability leaves the eavesdropper in a dilemma: if she believes the inferred breathing rate, she will be deceived; instead, if she does not trust any inferred breathing rate, her ability to eavesdropping breathing rate is lost.

Indirect Calculation: To calculate the real CSI, an eavesdropper must compromise the phase of distribution manipulation. As shown in Section 3.3.3, suppose that the ith subcarrier is sensitive, the transmitting signal on this subcarrier can be represented as $X_i^m(t) = \alpha_i C_i(t) D_i(t) X_i(t) + (1 - \alpha_i) X_i(t)$. We utilize $M_i(t) = C_i(t) D_i(t)$ to denote the total impact of disturbance manipulation. Let R_i^e denote the signal received by the eavesdropper on the ith subcarrier, and $H_i^e(t)$ denote the corresponding real subcarrier CSI between the transmitter and eavesdropper. Thus, we have $R_i^e = X_i^m(t) H_i^e(t) = a_i M_i(t) X_i(t) H_i^e(t) + (1 - a_i) X_i(t) H_i^e(t)$.

To learn $M_i(t)$, the eavesdropper must learn both a_i and $H_i^e(t)$. However, this imposes a strong requirement for the eavesdropper. On one hand, without the knowledge of the accurate positions of the target user and the transmitter, the



(a) Empty room.

(b) One person in a room.

(c) Two persons in a room.

(d) Three persons in a room.

Figure 3.9: CSI amplitudes at all subcarriers in different situations.

eavesdropper can hardly determine the subcarrier sensitivity except by guessing. On the other hand, the transmitter can always hide its real CSI between itself and the eavesdropper. Thus, $H_i^e(t)$ is not available. Consequently, the eavesdropper would fail to obtain $M_i(t)$ and cannot calculate the real CSIs of sensitive subcarriers for inferring the true breathing rate.

3.4 Ambush Design for Crowd Inference Attacks

A crowd inference attack using CSI is a technique that leverages wireless signals to estimate the number of people in an area of interest. The key idea of a crowd inference attack is that as more people enter the given area, the wireless signals will reflect off their bodies and change the channel characteristics. After extracting the CSI features, machine learning algorithms can be used to analyze the CSI data and estimate the number of people in the area accurately. To counteract such a CSI-based crowd inference attack, we propose the corresponding defenses called *Ghost* in the following.

We first explore the relationship between the CSI amplitude and the number of moving people in the given area. Thus, we investigate the amplitude of CSI measurements at each subcarrier within a certain time interval in several different situations (i.e., empty, one person in a room, two persons in a room, three persons in a room). As shown in Figure 3.9, we have the following observations: (1) different subcarriers show similar fluctuations, which demonstrates they have similar responses towards movement, and all subcarriers are sensitive due to the random changes of position; (2) the CSI amplitude is not periodic or predictable due to the random walking; (3) when the room is empty (i.e., no person is present), the CSI amplitude is almost flat and stable; (4) the variation of CSI amplitude becomes larger when the number of moving people increases. Based on these observations, we can subsequently design our *Ghost* defense against the crowd inference attack.

3.4.1 Overview

To defend against the crowd inference attack, the transmitter first arbitrarily specifies a fake person number to fool the attacker into entering the ambush. Due to the fact that all subcarriers are sensitive to the random movements of people in the given area, the ambush locations can be selected in hidden or concealed areas. Additionally, the ambush locations can be determined based on the locations where the eavesdropper is most likely to appear, and then deployed along the eavesdropper's possible route.

Different from *HoneyBreath* in Section 3.3.1, *Ghost* targets an empty room and fools the receiver into estimating the wrong person number in the room. The transmitter first proceeds to the *planning* phase, which consists of two tasks: CSI profile construction and CSI retrieval. Based on observation (2), the CSI amplitude is not periodic or predictable. Therefore, to map the specified person number to the CSI measurements, the transmitter can construct the CSI profile by collecting the CSI measurements corresponding to the different numbers of individuals. Later, according to the fake person number selected by the transmitter, the corresponding CSI stream can be retrieved from the built library and then fed into the next phase, *disturbance manipulation*. Due to observation (1), as all subcarriers are sensitive to movements, the transmitter is able to perform the same operation on each subcarrier. Then, all crafted signals are aggregated and sent out. Consequently, the attacker estimates the person number with the estimated CSI by performing the general *person number estimation* process. Figure 3.10 illustrates the flow chart of the proposed *Ghost* defense.

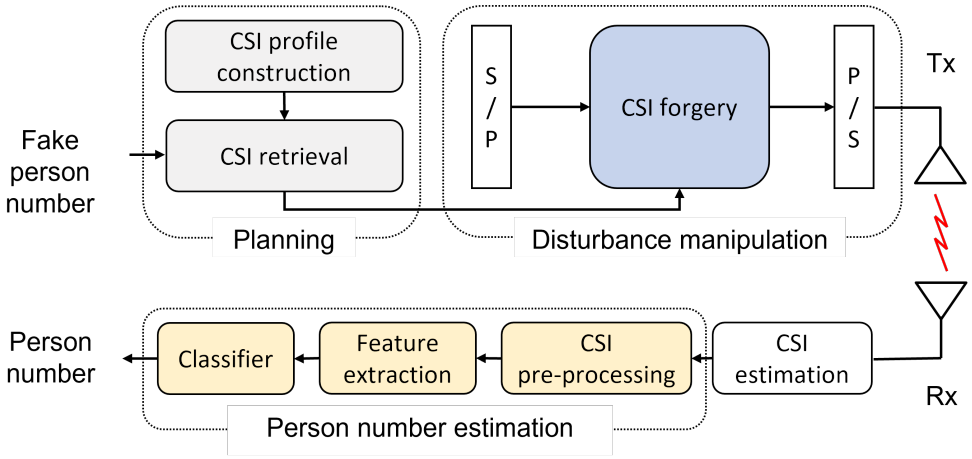


Figure 3.10: Flow chart of the proposed *Ghost* defense.

3.4.2 Planning Phase

In this phase, the following two steps are performed.

CSI Profile Construction: Different from the breathing inference attack where the victim is static, the victims in the crowd counting system move randomly. In this way, it is not possible to predict how the CSI varies with time. To address this challenge, the CSI profile can be constructed by collecting several CSI measurements when there are different numbers of moving people. Let $\mathbf{S}^p(t) = \{\mathbf{S}_1^p(t), \mathbf{S}_2^p(t), \cdots, \mathbf{S}_k^p(t)\}$ denote the CSI profile, where $p = 0, 1, \cdots, P$ represents the person number and $k = 0, 1, \cdots, K$ means the trial number. For each scenario with a different number of individuals, CSI measurements are collected in multiple trials at different times.

CSI Retrieval: The transmitter plans to send specified CSI to the receiver, from which the person number can be estimated. Thus, for the given person number p that is used to fool the attacker, the corresponding CSI $\mathbf{S}_k^p(t)$ can be extracted from the library as the specified CSI.

3.4.3 Disturbance Manipulation

Different from Section 3.3.3, it is not required to perform different operations on sensitive subcarriers and insensitive subcarriers, respectively. Since all subcarriers are sensitive to human movements during the walking period, CSI forgery needs to be performed on each subcarrier. To achieve it, the multiply-accumulate (MAC) process is exploited by the transmitter to generate the desired artificial disturbance.

3.4.4 Person Number Estimation

In the person number estimation phase, a general scheme has the following three steps:

CSI Pre-processing: After gathering the CSI measurements from the transmitter, the receiver first performs the CSI estimation based on the original training sequence and the received data. Since the CSI data is considerably noisy due to various factors such as interference, multipath fading, and hardware imperfections, it is necessary to remove the redundant components from the calculated amplitude values. For this smoothing process, we apply two kinds of filters, one is the Hampel filter for eliminating impulse noises, and the other is the moving average filter for removing high-frequency noise while preserving the low-frequency components.

Feature Extraction: By leveraging the CSI amplitude from each subcarrier, various statistical features [27,189] can be extracted for crowd counting as follows:

- *Mean:* the average value of amplitude.
- *Standard deviation (STD):* shows how individual amplitude values deviate

from the mean amplitude value.

- *Median Absolute Difference (MAD):* a robust measure of dispersion that is not affected by outliers.
- *Maximum:* the highest amplitude value.
- *Minimum:* the lowest amplitude value.
- *Skewness:* encompasses the asymmetric shape of the CSI subcarrier profile and indicates a stronger or weaker signal on the left or right.
- *Kurtosis:* represents how tail-heavy the shape is compared to a normal distribution (meaning more extreme values).
- *Entropy:* measures the amount of signal information.

After that, the CSI at each subcarrier corresponds to a 1×8 feature vector, which can be combined into a $N \times 8$ feature matrix, where N represents the number of subcarriers. Since these CSI amplitudes present similar variations and describe the same human movement, the average value of each feature across all subcarriers is calculated and regenerates the final 1×8 feature vector, which is fed into the classifier later.

Classifier: Accordingly, based on these common statistical features, a feature set can be created for each training sample to form a labeled dataset. Two widely used classifiers are trained to divide inputs into different predefined classes and then make decisions as follows:

- *Support Vector Machine (SVM):* used with one-versus-one (OvO) strategies, finds the hyperplane that maximally separates the data points of different classes.

- *Decision Tree (DT):* used with OvO strategies, recursively partitions the data and selects the optimal boundaries that best separate the data points of different classes.

Based on the trained classifier, the number of moving people can be estimated from the collected CSI measurements.

3.5 Experimental Evaluation

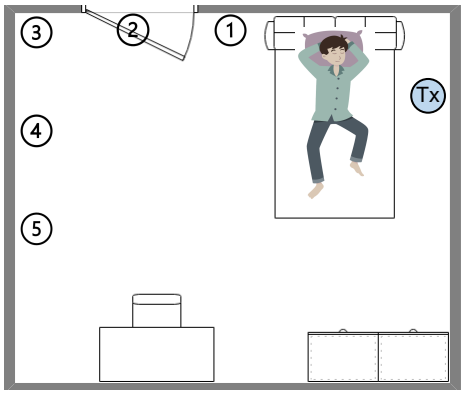
We implement CSI-based breathing rate inference and our proposed ambush schemes on top of Universal Software Radio Peripheral (USRP) X310s [43], which are equipped with SBX-120 daughterboards [42] and run GNU Radio [57] – an open-source software toolkit.

3.5.1 Evaluation Setup

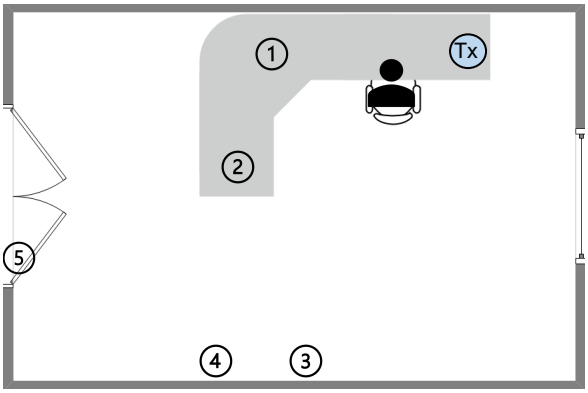
The prototype system includes a transmitter Tx and an eavesdropper Eve (i.e., malicious receiver). Each node is a USRP X310. We recruited 5 participants and asked each to act as the target user of the inference attacks over three months.[2] Also, each wore a Masimo MightySat Fingertip Pulse Oximeter [107] with hospital-grade technology to obtain ground-truth breathing rate.

Testing Scenarios: We test two typical scenarios: (1) a bedroom, where the user lies on a bed; and (2) an office room, where the user sits in a chair. Figure 3.11 shows the ambush locations and the position of the transmitter. For each scenario, we place Eve at 5 different ambush locations to infer the user's breathing rate, and the transmitter launches the proposed ambush scheme.

[2]The study has been approved by our institution's IRB.



(a) Bedroom (User lies down).



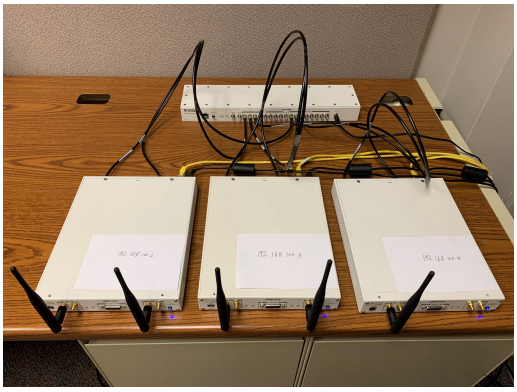
(b) Office (User sits).

Figure 3.11: Layout of the experimental environment.

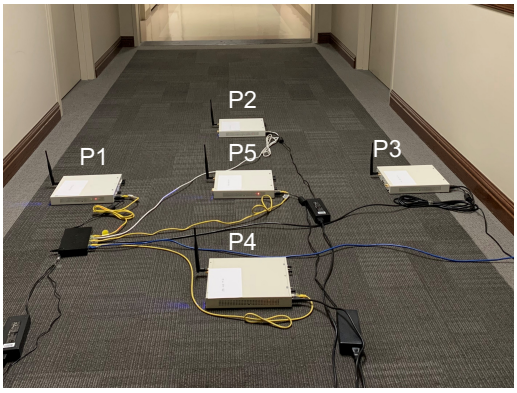
To deploy a trap area, as shown in Figure 3.12a, we use a 5-antenna transmitter, consisting of three USRP X310s, which are connected with a host computer through an Ethernet switch and synchronized with OctoClock-G [40]. As shown in Figure 3.12b, five collaborative eavesdroppers are placed at 5 specified ambush points on the corridor outside of the office room: one in the center and the other four in the circle with a radius (i.e., antenna-antenna distance) of 0.75 m.

Metrics: Let $\hat{r}$ denote the estimated rate.We apply the following two metrics.

- *Absolute estimation error* ϵ*:* the difference between true and estimated breathing rates, i.e., $|r_{gt} - \hat{r}|$, where r_{gt} is the ground truth.



(a) Five-antenna transmitter with USRPs.



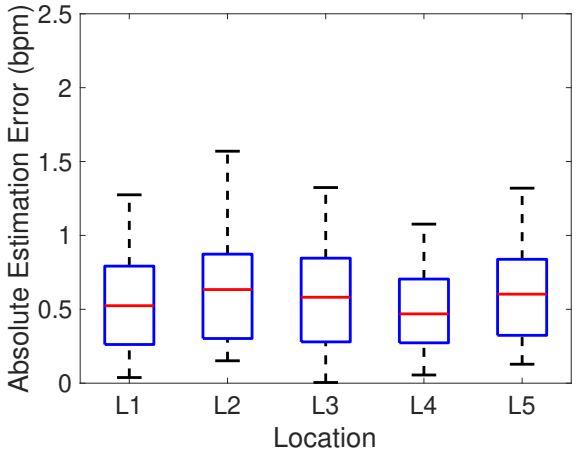
(b) Ambush area.

Figure 3.12: Setup for deploying an ambush area.

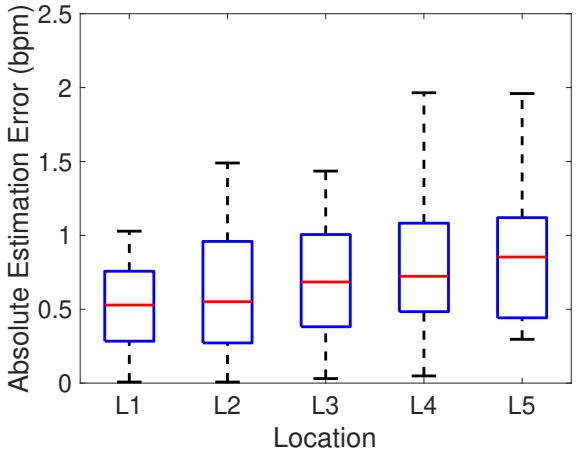
- *Absolute ambush error η:* the difference between estimated and specified breathing rates, i.e., $|r_a - \hat{r}|$, where r_a is the one specified by the transmitter.

3.5.2 Breathing Rate Inference Attacks

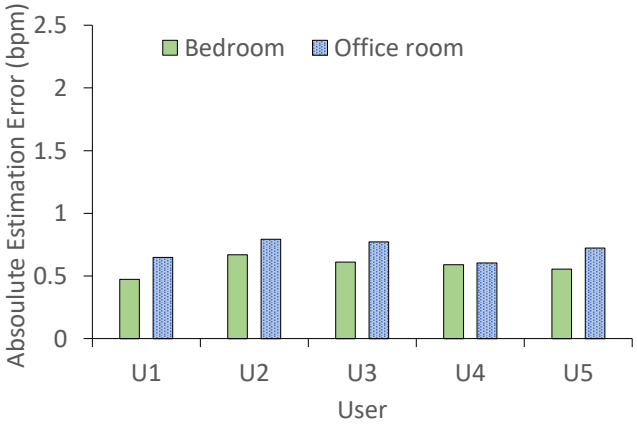
We first verify the effectiveness of using CSI to infer breathing rates. As shown in Figure 3.11, Eve is put at each ambush location in both of the two scenarios to estimate each participant's breathing rate, with 100 trials performed for every estimate. Figure 3.13 shows the obtained absolute estimation error when the proposed ambush scheme is not launched.



(a) In the bedroom.



(b) In the office room.



(c) Mean value of ϵ.

Figure 3.13: Values of ϵ and ϵ at Eve when no defense is enforced.

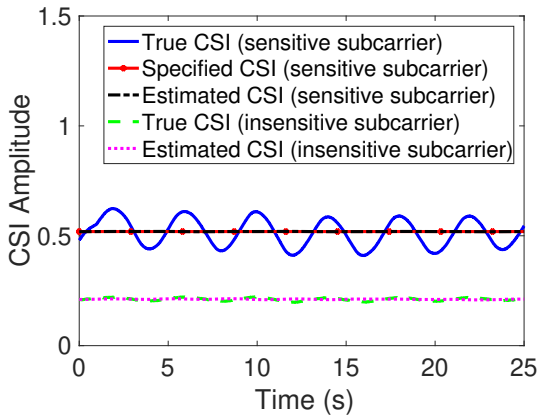
Figure 3.13a shows that the inference technique always achieves high accuracy with less than 1.6 bpm of error at all locations in the bedroom. The median absolute estimation error ranges from 0.4 to 0.6 bpm across all locations. Meanwhile, we see the value of ϵ on average is slightly larger at Location 2 than at other locations. This is because Location 2 is not in the LOS of the user and the resultant signal fading degrades the inference performance. We have similar observations from Figure 3.13b. Figure 3.13c depicts the mean absolute estimation errors for different users (referred to as U1$\sim$U5). We can observe that the mean absolute estimation error is consistently low (i.e., below 0.8 bpm) across all users in both environments. Also, the average absolute estimation error for each user in the office room is larger than that in the bedroom. It can be explained by the fact that the user has less body movement irrelevant to breathing activity when lying on the bed than when sitting in a chair. These results demonstrate convincingly that an eavesdropper could utilize passively collected CSI to accurately infer a person's breathing rate in different scenarios.

3.5.3 Example Defenses

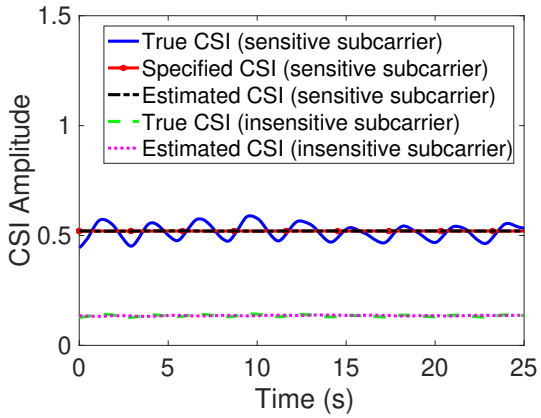
We examine three example defenses, in which we deploy the ambush location at Location 1 shown in Figure 3.11a and Location 3 shown in Figure 3.11b.

Example 1 - Making Breath Unobservable: We first show a defense method by hiding breathing rates, i.e., when Eve appears at the ambush location, she would obtain a breathing rate of 0 (i.e., no breathing activity is detected).

Figure 3.14 plots the real CSIs between the transmitter and the ambush location, the estimated CSIs at the ambush location, as well as the subcarrier CSI specified by the transmitter. In both environments, the transmitter can make



(a) In the bedroom.

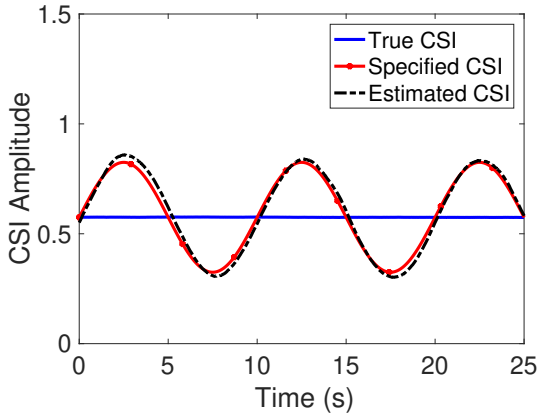


(b) In the office room.

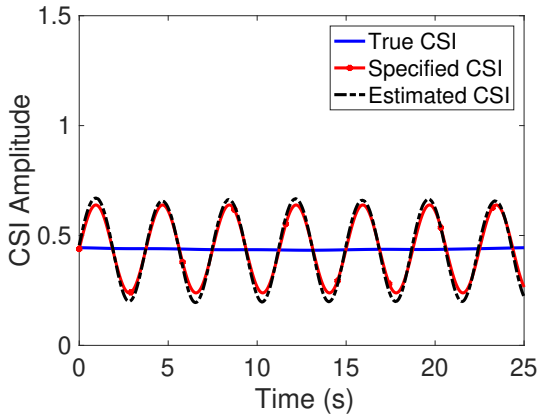
Figure 3.14: Enabling Eve to obtain no breathing activity.

Eve observe a CSI on a sensitive subcarrier significantly near to the specified one while both greatly deviate from the true one; with the estimated CSI, Eve obtains a breathing rate of 0 though the respective true breathing rates are 15.1 and 20.8 bpm. The absolute estimation errors in the bedroom and the office room are thus 15.1 and 20.8 bpm, while the corresponding absolute ambush errors are both 0. Besides, the CSI of the insensitive subcarrier keeps insensitive with the defense (we thus only focus on sensitive subcarriers in the later evaluation).

Example 2 - Fabricating Nonexistent Breath: We aim to make Eve obtain a fake breathing rate while there is no breathing activity in both scenarios.



(a) In the bedroom.



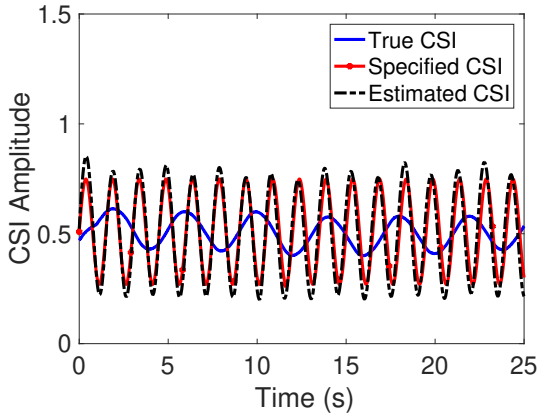
(b) In the office room.

Figure 3.15: Fabricating normal breath.

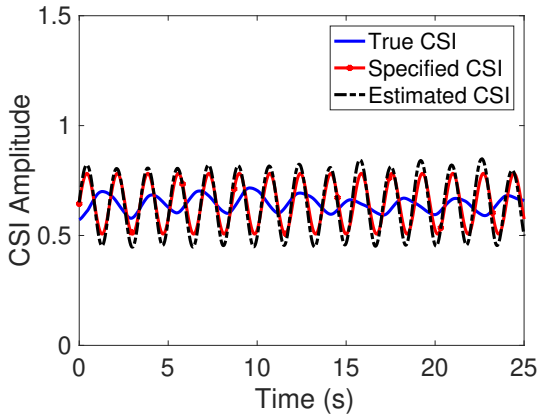
We specify a fake breathing rate of 6 (16) bpm for the bedroom (office) room.

As shown in Figure 3.15, we see the true CSI is almost flat, as there is in fact no breathing activity, and the estimated CSI is quite consistent with the CSI specified by the transmitter. With the estimated CSI, Eve obtains a breathing rate of 6.4 bpm in the bedroom and 16.1 bpm in the office room. The absolute estimation errors in the two scenarios become 6.4 and 16.1 bpm, respectively; the respective absolute ambush errors are as small as 0.4 bpm and 0.1 bpm.

Example 3 - Falsifying Breath: We aim to hide a normal breathing rate by making Eve observe an abnormal one. We randomly specify an abnormal



(a) In the bedroom.



(b) In the office room.

Figure 3.16: Making Eve obtain abnormal breath.

breathing rate of 40 bpm for the bedroom and 35 bpm for the office room.

Similar to the above examples, we observe from Figure 3.16 that the estimated CSI is quite close to the specified CSI while it greatly differs from the true CSI in both environments. The estimated breathing rate of Eve in the bedroom becomes 40.2 bpm, instead of the true one (i.e., 19.9 bpm) derived from the Masimo Oximeter. In the office room, Eve obtains a breathing rate of 35.2 bpm, instead of the ground truth (i.e., 17.0 bpm). Therefore, the absolute estimation errors for the bedroom and the office room are 20.3 bpm and 18.2 bpm, respectively, while the absolute ambush errors in these two scenarios are both just 0.2 bpm.

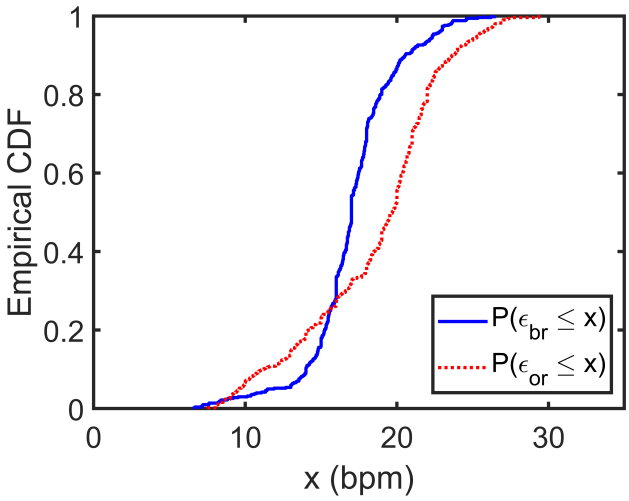


Figure 3.17: CDFs of $P(\epsilon \leq x)$ for **D1**.

3.5.4 Overall Defense Impact

We examine the overall impact of the three defenses (numbered according to their respective cases): (1) a user is breathing while we aim to make Eve obtain no breathing activity; (2) no breathing activity occurs while we aim to make Eve obtain a fake breathing rate; (3) a user is breathing while we aim to make Eve obtain a different non-zero breathing rate. Eve estimates the breathing rate at every ambush location. For each estimate, we perform 100 trials.

D1: We test when the user has different breathing rates in the range of 6-27 bpm. For all trials, we find that Eve always obtains an estimated breathing rate of 0, indicating the consistent success of the defense. Let $P(\epsilon_{br} \leq x)$ and $P(\epsilon_{or} \leq x)$ denote the empirical cumulative distribution functions (CDFs) of the absolute estimation error ϵ_{br} for the bedroom and ϵ_{or} for the office room. Figure 3.17 shows that ϵ_{br} and ϵ_{or} lie in the ranges of [6.6, 26.5] and [7.5, 29.6] with probability 100%. Both demonstrate that Eve always has a significant error in the breathing rate estimation with the proposed defense.

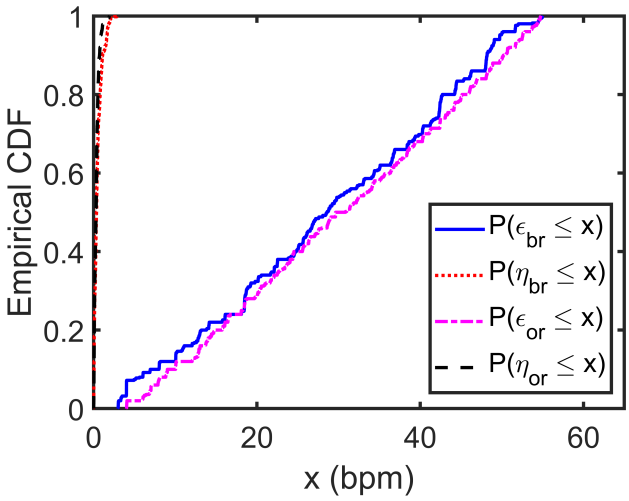


Figure 3.18: CDFs of $P(\epsilon \leq x)$ and $P(\eta \leq x)$ for **D2**.

D2: We randomly specify a fake breathing rate within the range of 3-55 bpm in each trial. Let $P(\eta_{br} \leq x)$ and $P(\eta_{or} \leq x)$ denote the CDFs of the absolute ambush errors η_{br} for the bedroom and η_{or} for the office room. As shown in Figure 3.18, we observe a small η and a high ϵ for both environments. For example, η_{br} is less than 1.5 bpm with a probability of 95.0%, while ϵ_{br} ranges from 3.0 to 54.8 bpm and is larger than 3.1 with a probability of 98.2%.

D3: Each participant has a normal breathing rate, and the transmitter chooses a bogus breathing rate randomly in an abnormal range (31-56 bpm). Figure 3.19 shows the CDFs of the corresponding ϵ and η. We can see that ϵ_{br} and ϵ_{or} are larger than 11 bpm with probabilities of 96.2% and 99.0%, respectively. Meanwhile, η_{br} is always less than 1.2 bpm, and η_{or} is always less than 0.9 bpm.

Figures 3.20a and 3.20b show the mean value of ϵ across all locations in both environments when the proposed defenses are employed. We observe that ϵ stays consistently high at all ambush locations for both environments. Compared with no defense, all defenses can significantly increase ϵ at Eve.

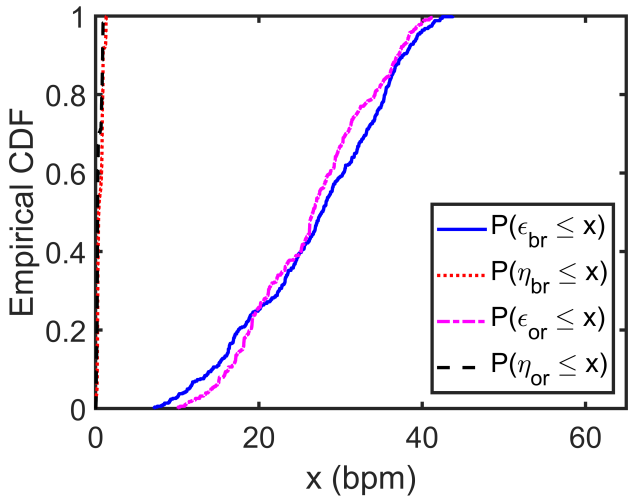
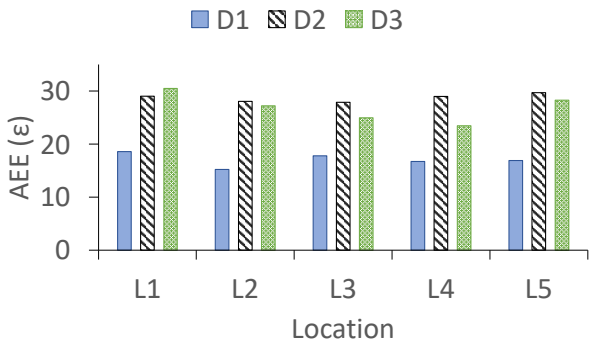


Figure 3.19: CDFs of $P(\epsilon \leq x)$ and $P(\eta \leq x)$ for **D3**.

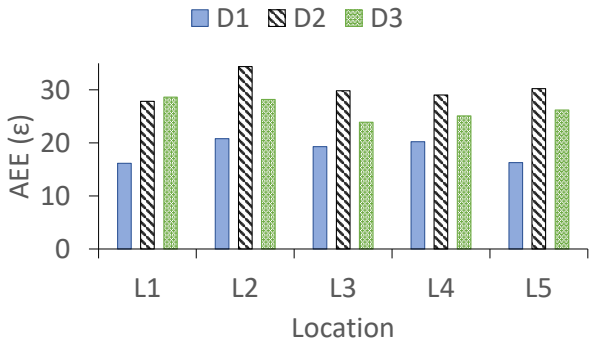
3.5.5 Two-user Scenario

First, we aim to make two persons' breathing unobservable (referred to as **D1**). We consider the scenario when two participants are in the office room simultaneously. As shown in Figure 3.21a, the estimated CSI is quite close to the specified one while both deviate from the true CSI. Consequently, Eve obtains a breathing rate of 0 though the true breathing rates of the two users are 6.0 and 10.0 bpm, respectively. Second, we aim to make Eve observe two specified breathing rates (16 and 22 bpm) when there is no breathing activity (referred to as **D2**). As shown in Figure 3.21b, though the true CSI is almost flat, indicating no person in the room, the estimated CSI and the specified one are alike, leading Eve to obtain two-person breathing rates of 16.0 and 22.1 bpm.

We repeat the above two experiments 40 times. For comparison, we also perform 40 attempts of inferring two-person breathing rates when no defense is applied (this case is denoted with **ND**). Figure 3.21c presents the absolute estimation errors (ϵ) for the cases with two real or fake users (U1 and U2).



(a) In the bedroom.



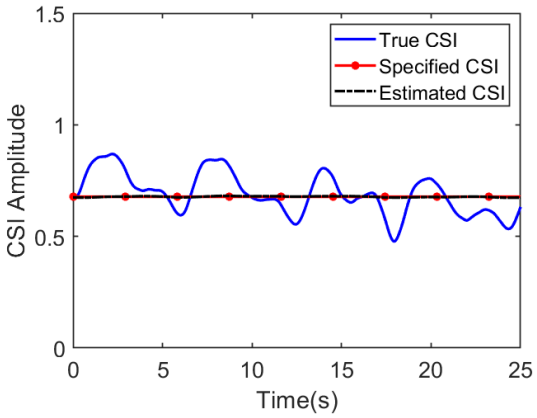
(b) In the office room.

Figure 3.20: Mean absolute estimation errors (AEE).

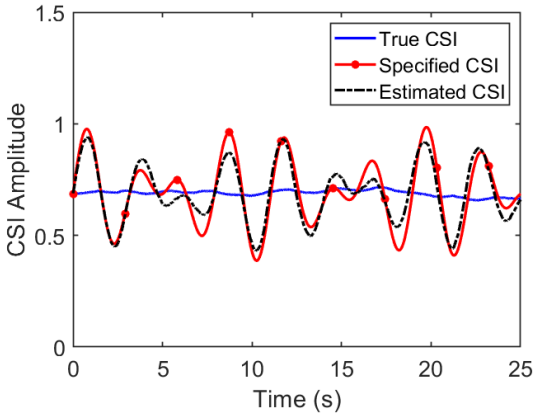
Without the defenses, the mean value of ϵ is quite small (around 0.8 bpm); while it is significantly increased (within the range of 12.6-17.2 bpm) with the proposed defenses (D1 and D2). Also, for D1, the mean values of the absolute ambush error η for the two users both equal about 0, while for D2, they are 0.1 and 0.4 bpm. These results convincingly show the proposed scheme can successfully mislead Eve with specified breathing rates for the two-user scenario.

3.5.6 Trap Area Evaluation

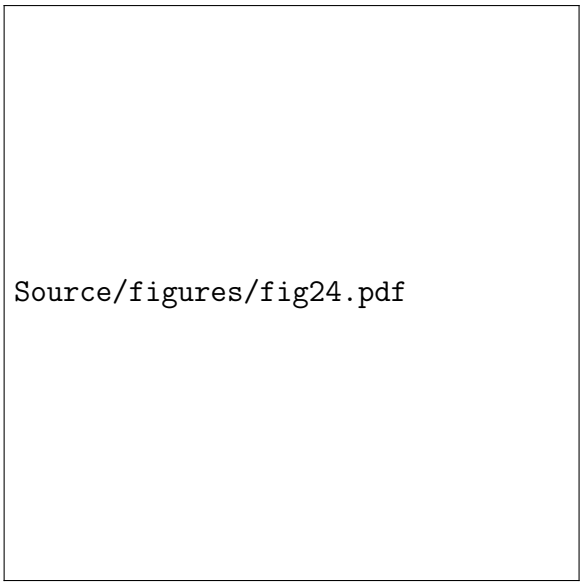
We aim to generate fake breath rates in the trap area consisting of five ambush points (referred to as P1$\sim$P5), as shown in Figure 3.12b. We choose a breathing



(a) Canceling two-user breath.

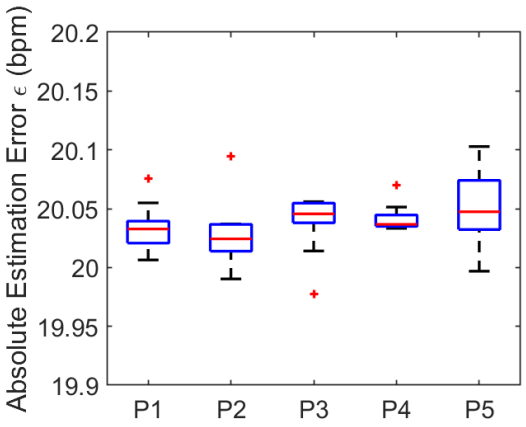


(b) Creating two-user breath.

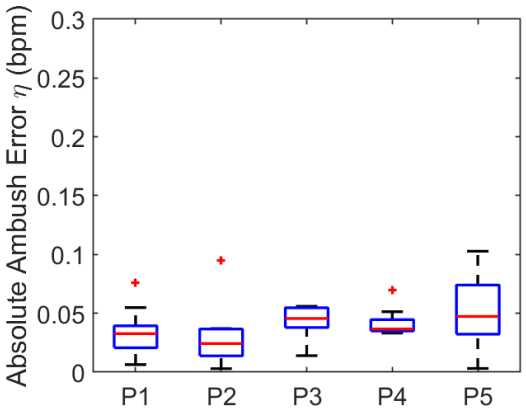


(c) Comparison of ϵ.

Figure 3.21: Extending defenses in two-user scenario.



(a) Absolute estimation error.



(b) Absolute ambush error.

Figure 3.22: Fabricating normal breath for a trap area.

rate of 20 bpm when the target room has no breathing activity. We perform 10 trials of deploying a trap area.

Figure 3.22a shows that the absolute estimation errors at all ambush points are consistently large (close to 20 bpm). Figure 3.22b demonstrates that the absolute ambush errors at all ambush points are quite small, with the mean value ranging from 0.03 to 0.05 bpm across all ambush points. These results demonstrate that the proposed scheme can simultaneously deploy multiple ambush points to mislead collaborative eavesdroppers (or simply increase the probability of trapping a single eavesdropper) with fake breathing rates.

3.5.7 Ghost Defense against Crowd Inference Attacks

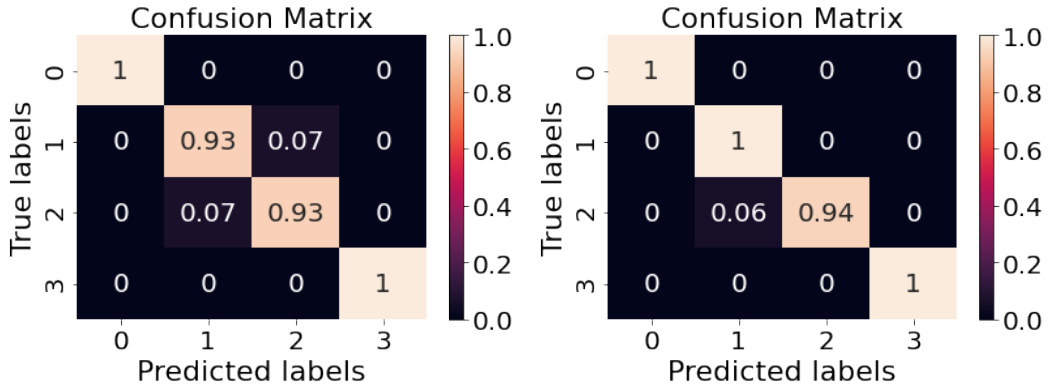
3.5.7.1 Experimental Setup

Similarly, we implemented CSI-based crowd inference and our proposed defense schemes using USRP X310s [43], which were used as a transmitter (Tx) and an eavesdropper (Eve, i.e., malicious receiver).

We asked 3 participants to randomly walk into a room. To build the CSI profile, we collected CSI data in four scenarios: empty room, one person in a room, two persons in a room, and three persons in a room. We performed 50 estimations for each scenario, resulting in a total of $50 \times 4 = 200$ estimations. Each estimation lasted for one minute. We then split the dataset into training (70%) and testing sets (30%). The training set was used to train the classifier, while the testing set was used to evaluate its performance.

For the defense scheme, we first performed 20 trials in the empty room. After that, we randomly selected one pre-collected CSI for each of the three defense strategies (i.e., fabricating the presence of one person, two persons, or three persons in an empty room). We conducted 20 trials for each defense strategy, resulting in $20 \times (1 + 3) = 80$ trials in total. We used the estimated CSI data to test the trained classifiers and evaluate our defense scheme.

To evaluate the crowd inference attack and the proposed defense, we used a confusion matrix to visualize the results. The accuracy was then calculated to indicate the proportion of correct predictions out of the total number of predictions.



(a) SVM classifier. (b) DT classifier.

Figure 3.23: Confusion matrix for crowd inference attack.

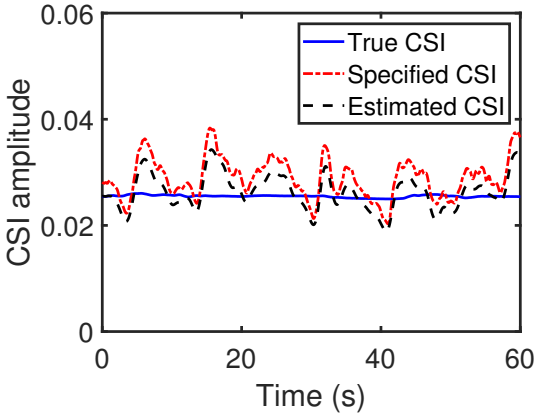
3.5.7.2 Crowd Inference Attacks

We first verified the effectiveness of using CSI to estimate the number of people present. Figure 3.23 shows the confusion matrix for two classifiers when the proposed defense scheme is not employed. As observed, the SVM classifier achieves an accuracy of 96.5%, while the DT classifier achieves an accuracy of 98.5%.

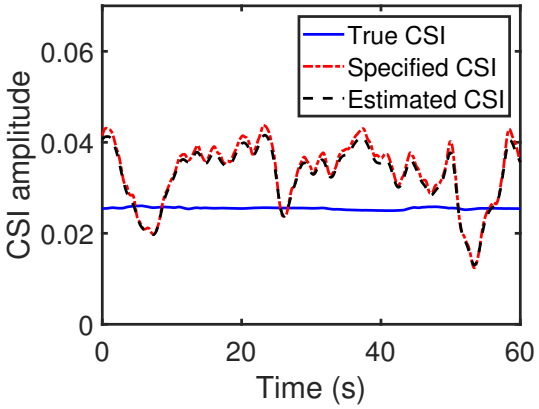
3.5.7.3 Example Defenses

We examine three example defenses, in which we deploy the proposed defense in the empty room to fool the attacker to consider it as an occupied room.

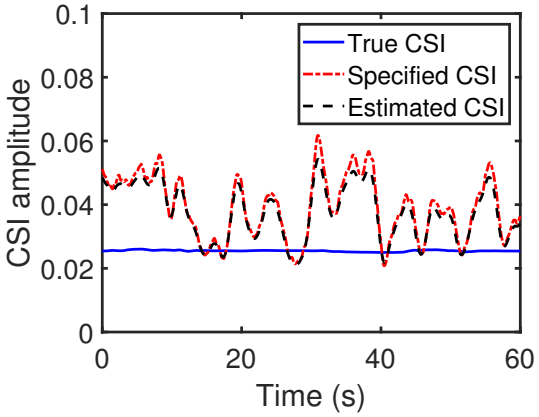
We demonstrate a defense method that involves fabricating one, two, or three moving people in an empty room when Eve attempts to launch a crowd inference attack. Figure 3.24 plots the real CSIs between the transmitter and the ambush location, the estimated CSIs at the ambush location, and the subcarrier CSI specified by the transmitter. In an empty room, the transmitter can cause Eve to observe a CSI on each subcarrier that is significantly close to the specified one, while both are greatly different from the true one (since the CSI is almost flat



(a) One person.



(b) Two persons.



(c) Three persons.

Figure 3.24: Estimated CSI at Eve in an empty room when our defense is enforced.

and stable in an empty room). It is noted that the specified CSI is extracted from the built CSI profile, which consists of various collected true CSIs. In this way, by obtaining these CSIs in Figures 3.24a, 3.24b, and 3.24c, respectively, the attacker will estimate the corresponding person numbers as 1, 2, and 3, after inputting such estimated CSI into the classifier.

3.5.7.4 Overall Defense Impact

We examined the overall impact of our defense strategies. Specifically, we consider a scenario where the actual number of people in a room is zero (i.e., the true label is 0). If Eve launches a crowd inference attack in this empty room, both the trained SVM and DT classifiers can initially identify the empty room with 100% accuracy. However, after each defense strategy is implemented, both classifiers misclassify the empty room as containing one, two, or three people, also with a 100% probability. Consequently, the accuracy of both classifiers dropped to 0%. These results demonstrate that Eve consistently makes significant errors in estimating the crowd count when our defense strategies are applied, further confirming the effectiveness of our defense scheme.

Chapter 4

Phantom-CSI Attacks against Wireless Liveness Detection

In this chapter [1], we identify vulnerabilities in existing wireless liveness detection systems and propose corresponding countermeasures to defend against such Phantom-CSI attacks.

4.1 Preliminaries

In this section, we introduce the prevalent algorithm used to estimate CSI for OFDM and the general method used by existing work employing CSI to achieve liveness detection.

[1] This chapter was published in Qiuye He and Song Fang. 2023. Phantom-CSI Attacks against Wireless Liveness Detection. In Proceedings of the 26th International Symposium on Research in Attacks, Intrusions and Defenses (RAID '23). Association for Computing Machinery, New York, NY, USA, 440-454. https://doi.org/10.1145/3607199.3607245. Used with permission.

4.1.1 CSI Estimation

As discussed earlier, the occurrence of human activities can induce disturbances in the surrounding wireless signal and thus variation in the observed CSI at the receiver.

The OFDM technique has been widely used in modern wireless communication systems, e.g., 802.11 a/g/n/ac/ad. The *channel frequency responses* measured from all subcarriers form the CSI of OFDM. Let $H(f,t)$ denote the channel frequency response at time t for a particular subcarrier with a frequency f. It is usually estimated by using a pseudo-noise sequence that is publicly known [50, 58, 176]. Specifically, a transmitter sends a pseudo-noise sequence, denoted with $X(f,t)$, over the wireless channel, and the receiver estimates $H(f,t)$ from $X(f,t)$ and the received, distorted copy, denoted with $Y(f,t)$, i.e., $H(f,t) = \frac{Y(f,t)}{X(f,t)}$.

4.1.2 CSI-aided Liveness Detection

A myriad of recent studies have shown the success of using CSI to recognize subtle human movements, including walking [158, 164], falling [118], breathing [98], mouth movements [153], and activities of daily living [126]. Existing CSI-based liveness detection techniques discover that CSI from widely available wireless signals is able to perceive human existence or activities in the place of interest in addition to surveillance cameras [78, 92] or a microphone [110, 123], and thus spoofing attacks can be detected by catching dissimilarities between CSI and video/voice signals.

These techniques normally use four steps to verify live users and detect spoofing attacks, namely, data synchronization, data pre-processing, feature extrac-

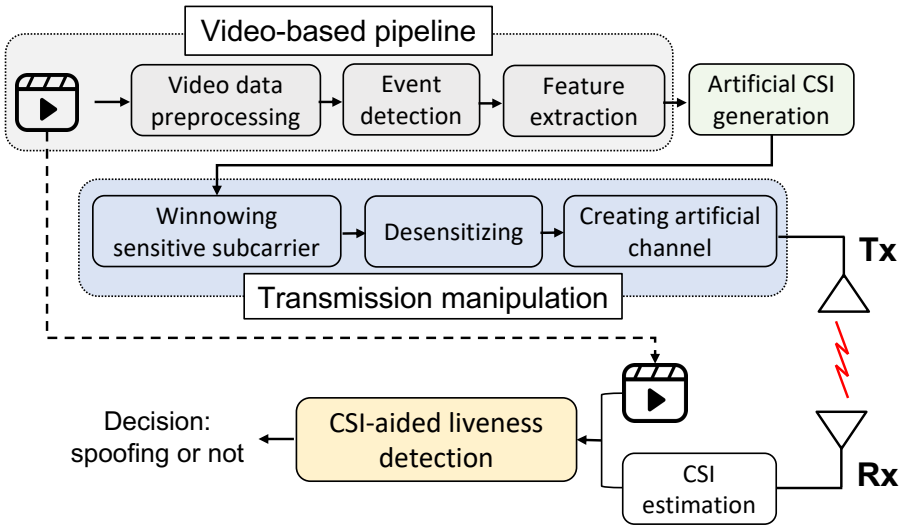


Figure 4.1: Flow chart of the phantom-CSI attack.

tion, and consistency checking. The first phase synchronizes signals in both modalities. The following phase pretreats video/voice feeds for activity detection and removes noise from the CSI. Next, specific features are extracted from both CSI and video/voice signals. They are then correlated and exploited to decide whether a spoofing attack happens in the final phase. Figure 1.3 illustrates a general flowchart of the CSI-aided liveness detection system.

4.2 Adversary Model

A general wireless liveness detection system utilizes wireless signals as a second-factor authentication for human activity, which is detected via another co-existing sensor. Without loss of generality, we consider a common surveillance scenario, where a camera is used to monitor an open area, and a transmitter and receiver pair is utilized to verify the authenticity of the video captured by the camera. Specifically, the public transmitter constantly transmits the wireless signal; the receiver estimates the CSI based on the received signal. We point out that such

a public transmitter can be unreliable and can be exploited for launching the proposed attack. If the detected human activities from wireless signals and the camera match with each other, the video is authentic, otherwise the video spoofing attack is detected.

To demonstrate the impact of our attack, we consider an attacker who can craft a fake video and feed it to the camera (e.g., [12, 72]). This aligns with existing liveness detection studies (e.g., [77, 78, 83, 92]). The attacker aims to make the target system unable to detect the fake video. She may use the public transmitter as an accomplice. Alternatively, if the defender secures the public transmitter, the attacker can set up a hidden transmitter nearby. Similar to other wireless attacks such as GPS spoofing [140], the attacker's transmitter then employs wireless jamming or spoofing techniques [182] to cancel the real signals and let the receiver take the fake signals from the attacker as the real ones. Toward the goal, the malicious transmitter attempts to mislead the receiver by generating a phantom CSI that matches the forged video.

4.3 System Design

4.3.1 Attack Overview

Existing wireless liveness detection systems rely on wireless environmental fluctuations to detect video- or voice-spoofing attacks. Our key idea of the proposed attack is to manipulate the wireless environmental fluctuations so that both the coexisting video/voice and CSI data have a consistent observation of human activities. Wireless liveness detection systems would thus be unaware of the spoofing attacks. Without loss of generality, we assume that the attacker aims to launch

video spoofing attacks.

In a typical video spoofing attack, the attacker replaces the live video frames with fake ones (e.g., what are previously recorded) so that she can perform activities in the area monitored by the camera without being recorded. With a stream of video frames, the *data pre-processing* phase first identifies body keypoints in each video frame. Such keypoints input to the *event detection* phase, which determines the ongoing event. After that, the *feature extraction* phase generates semantic features from the processed video data, which are compared with that extracted from the CSI to determine the authenticity of the captured video.

To make the receiver observe fake CSI, whose semantic features are consistent with that extracted from the video data, the attacker first specifies such artificial CSI, and then delivers it to the receiver by manipulating the transmitted signal. Since the transmitted signal has to experience the distortion effect applied by the real wireless channel, the attacker compensates for such distortion effect at the transmitter side. Consequently, the receiver extracts the semantic features of the ongoing event with estimated CSI. Figure 4.1 depicts the flow chart of the proposed attack.

4.3.2 Video-based Pipeline

Traditional video-based monitoring system usually involves three steps, data preprocessing, event detection, and feature (i.e., event parameter) extraction.

Data Pre-processing: *OpenPose* is the first open-source real-time video processing tool for 2D pose detection, including tracking body, foot, hand, and facial keypoints [20]. It is also widely used in existing wireless liveness studies (e.g., [78, 92]). We also utilize OpenPose to process video frames, each of which

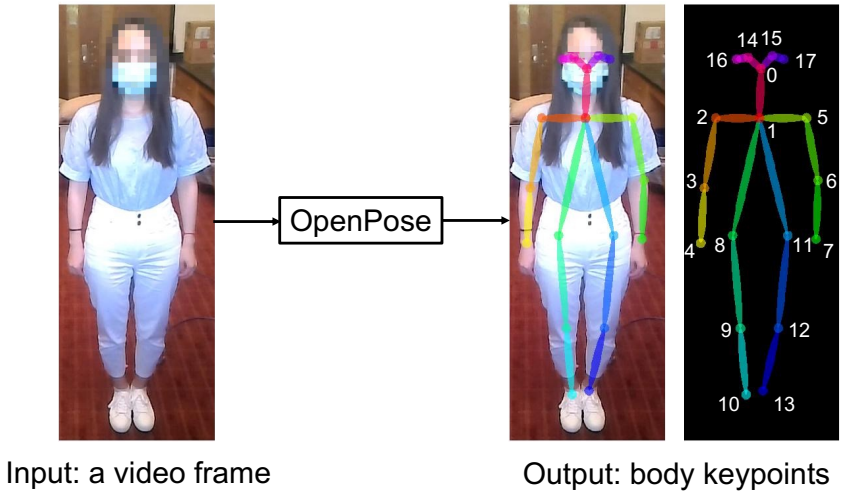


Figure 4.2: Body keypoints extracted by *OpenPose*.

then generates X-Y coordinates of the 18 body keypoints. Figure 4.2 shows an example of the body keypoints extracted from a video frame using OpenPose. We see that there are 18 keypoints (labeled with 0-17) of the target person. The displacement of those keypoints over time can then help infer occurrent events (e.g., human activities).

Event Detection: The input of this step is the X-Y coordinates of the 18 body keypoints extracted from each video frame. Let P_m^i denote the ith point in the mth video frame, where $i \in \{1, 2, \cdots, 18\}$ and $m \in \{1, 2, \cdots, M\}$, where M denotes the total amount of video frames. The Euclidean distance of each point between the mth frame and the $(m+1)^{th}$ can be denoted as $L_m^i = \left|P_m^i - P_{m+1}^i\right|$. We then add up all these Euclidean distances and obtain the sum $D_m = \sum_{i=1}^{18} L_m^i$. If D_m is larger than the predefined threshold D_0, we regard that the motion is detected; otherwise, there is no motion detected if $D_m \leq D_0$. We iterate over all neighboring video frames with this scheme, separating dynamic scenes (with motion) from static ones.

Feature Extraction: We need to select a set of distinctive semantic features

of motion, so that we can use them to design corresponding phantom-CSI flows. The start time and the end time of motion are often chosen as such features. If the motion occurring in the video is periodic, the motion frequency is also recorded as another. Particularly, to determine the frequency, we apply a metric referred to as motion energy which captures the energy in the different frequency bands of the body keypoints. With the FFT profile of the body keypoints, a single frequency component that exhibits the maximum signal magnitude can be extracted.

4.3.3 Artificial CSI Generation

The attacker would deliver specified CSI to the receiver, which matches events occurring in the injected fake video. Let $\mathbf{h}_T(t) = [h_{T_1}(t), h_{T_2}(t), \cdots, h_{T_N}(t)]$ denote the target CSI for N subcarriers. Intuitively, we may pre-record the CSI corresponding to the events in the video as $\mathbf{h}_T(t)$. However, this profiling process of collecting CSI is laborsome and may place an extra burden on the attacker. Instead, we propose a method that enables the attacker to generate such artificial CSI.

In general, to craft $\mathbf{h}_T(t)$, there are the following two cases: 1) the video just contains static images and has no human activity in the video; 2) the video contains human activity. For the first case, the target CSI $\mathbf{h}_T(t)$ can be easily crafted, denoting the random noise in the environment. For the latter case, we then need to convert the human activities into $\mathbf{h}_T(t)$.

Different human activities may cause different impacts on the environmental CSI. Specifically, the CSI amplitude on a sensitive subcarrier often shows a strong correlation with human activities. As a non-synchronized transmitter and

receiver pair may bring an unknown phase lag [129], the CSI amplitude is often only chosen to characterize the wireless channel for human activity detection. Correspondingly, this paper also focuses on wireless liveness detection using CSI amplitudes.

It is widely observed that periodic movement usually makes the CSI amplitude on a sensitive subcarrier present a sinusoidal-like pattern over time [154]. Let f_a denote the frequency (Hz) of the occurred event. We then convert the event into a subcarrier CSI $h_{T_i}(t) = |h_{T_i}(t)| \cdot e^{j\theta(t)} + N_i(t)$, where $|h_{T_i}(t)|$ represents amplitude. We model the CSI envelope on a sensitive subcarrier as a sinusoidal-like wave, i.e.,

$$|h_{T_i}(t)| = a \cdot sin(2\pi f_a t + \beta) + N_i, \text{when } t \in [\tau_s, \tau_e], \tag{4.1}$$

where a, β, and N_i are the amplitude, initial phase, and additive noise. When $t \notin [\tau_s, \tau_e]$ (i.e., outside of the activity period), there is no need to craft specific CSI and we then have $|h_{T_i}(t)| = 0$. In turn, with such a CSI envelope, the receiver can infer the start and end times of the activity, as well as the event frequency.

4.3.4 Transmission Manipulation

To invalidate wireless liveness detection, the transmitter (i.e., attacker) needs to make the receiver believe the target CSI $h_{T_i}(t)$ on sensitive subcarriers. To achieve this goal, the following three steps are required to craft the transmitted signal.

4.3.4.1 Winnowing Sensitive Subcarrier

Due to the multipath effect, signals usually arrive at the receiver via different paths, e.g., line-of-sight (LOS) and non-line-of-sight (NLOS). These signals may

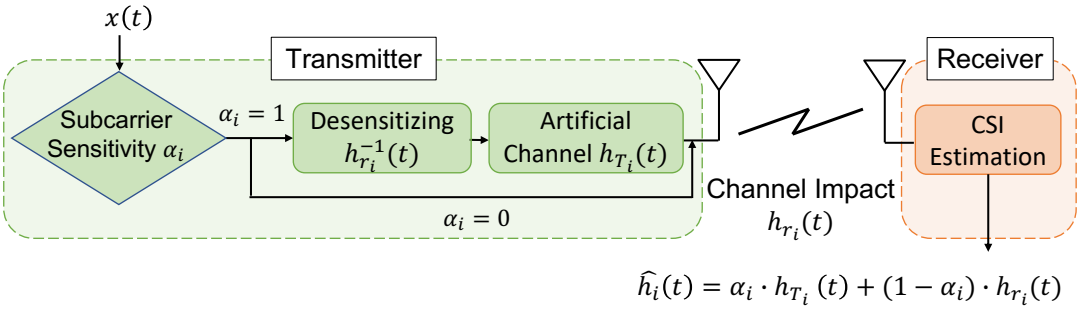


Figure 4.3: Subcarrier-level CSI wave morphing.

interfere constructively or destructively, leading the receiver to observe enhanced or weakened signals. This phenomenon may vary for different subcarriers as they have varying wavelengths. Consequently, all subcarriers can be divided into two groups: sensitive and insensitive. Sensitive subcarriers show large amplitudes (or variances), while insensitive subcarriers have imperceptible signal fluctuations. Thus, observations on sensitive subcarriers are utilized to detect human activities.

We utilize a binary decision variable α_i to indicate the subcarrier sensitivity, with 1 denoting sensitive while 0 showing insensitive. Since insensitive subcarriers are not involved in wireless liveness detection decisions, we only exploit sensitive subcarriers for achieving CSI manipulation.

4.3.4.2 Desensitizing

Since the transmitted signal has to experience the real wireless channel, the transmitter needs to cancel the actual distortion effect of the real channel. We call this process *desensitizing*. Let $h_{r_i}(t)$ denote the real CSI of the ith sensitive subcarrier, and $d_i(t)$ represent the corresponding coefficient of the desensitizing module. $d_i(t)$ would be the inverse of $h_{r_i}(t)$ to eliminate the impact of the real channel on the transmitted signal $x(t)$. We then have $d_i(t) \cdot h_{r_i}(t)$=1, i.e., $d_i(t) = h_{r_i}^{-1}(t)$.

Activity Removal in Dynamic Scenarios: Generally, to obtain the real

CSI in environments with human motion, an attacker can utilize a CSI profiling process. Particularly, rhythmic human activities (e.g., breathing) periodically affect the CSI waveforms, and the resultant CSI often presents a sinusoidal-like pattern, which can be then modeled by the attacker, as illustrated in Section 4.3.3.

Signal Annihilation in Realistic Settings: To cancel the real channel effect, the attacker needs to know the real CSI via CSI profiling or modeling ahead. In certain cases, human activity is complex and the real CSI is not available. However, the attack impact still exists. Although the attacker cannot control the CSI obtained at the receiver, she can then utilize a random coefficient of the desensitizing model. This may not successfully cancel the real channel effect, but it can make the target wireless liveness detection system obtain random and incorrect decisions. In the following, we focus on the scenarios where the attacker has knowledge of the real CSI due to the higher manipulability and more misleading nature of such attacks.

4.3.4.3 Creating Artificial Channel

After canceling the real channel effect, the attacker also needs to create an artificial channel to make the receiver obtain the target CSI, crafted during the phase of event-CSI conversion, as demonstrated in Section 4.3.3. Let $h_{a_i}(t)$ denote the specified CSI of the artificial i^{th} subchannel, and we thus obtain $h_{a_i}(t) = h_{T_i}(t)$.

Figure 4.3 illustrates subcarrier-level transmission signal manipulation. We use $x_{a_i}(t)$ to show the actual transmitted signal on the ith subchannel. After the original signal $x(t)$ goes through the two steps of desensitizing and artificial channel, we have $x_{a_i}(t) = (1-\alpha)\cdot x(t) + \alpha \cdot x(t) \cdot h_{r_i}^{-1}(t) \cdot h_{a_i}(t)$. The received signal at the receiver then becomes $y_{a_i}(t) = x_{a_i}(t) \cdot h_{r_i}(t)$ (where we omit the noise term for the sake of simplicity). With $y_{a_i}(t)$ and the publicly known training sequence,

the receiver can estimate the subcarrier CSI $\hat{h_i}(t)$, i.e., $y_{a_i}(t) = x(t) \cdot \hat{h_i}(t)$. As a result, we have

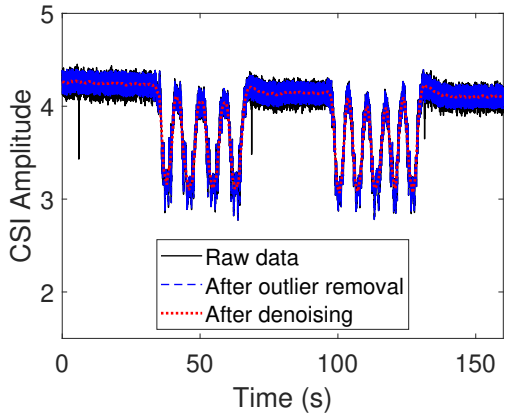
$$\hat{h_i}(t) = \alpha \cdot h_{T_i}(t) + (1 - \alpha) \cdot h_{r_i}(t). \tag{4.2}$$

Consequently, for insensitive subcarriers ($\alpha = 0$), we obtain $\hat{h_i}(t) = h_{r_i}(t)$, i.e., no manipulation is applied; while for sensitive subcarriers ($\alpha = 1$), we have $\hat{h_i}(t) = h_{T_i}(t)$, demonstrating that the proposed method is able to make the receiver estimate the specified CSI via creating an artificial channel.

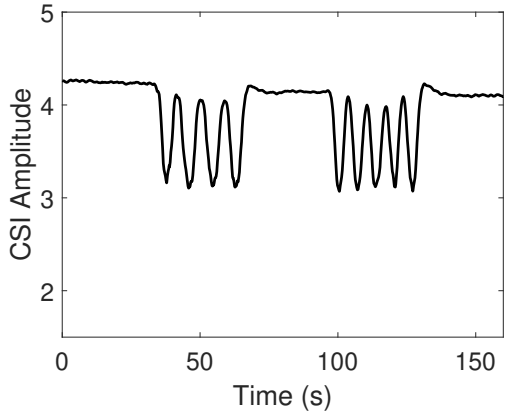
Synchronization for Real CSI Cancellation: CSI patterns (e.g., peaks and valleys in sinusoidal waves) change with human motion, and CSI during the motion period shows a larger variance than those happening out of the period. We can thus utilize human motion and the corresponding CSI feature to achieve synchronization, so that the real channel effect can be compensated.

4.3.5 CSI-aided Liveness Detection

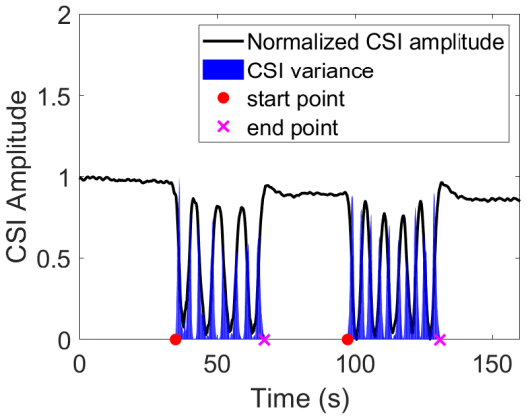
With both the video and CSI signals, as discussed in Section 4.1.2, we apply the general wireless liveness detection process in existing studies (e.g., [92]). Particularly, we first synchronize both signals and then process each. The video data processing follows the procedures described in Section 4.3.2, while the CSI-based monitoring pipeline is an inverse process of event-CSI conversion, including CSI data preprocessing, event detection, and feature extraction. Finally, we cross-check features extracted from the two sources to determine whether a spoofing attack happens.



(a) Outlier removal and noise reduction.



(b) Waveform after PCA.



(c) Variance scan and event extraction.

Figure 4.4: Procedures of CSI data preprocessing.

4.3.5.1 CSI and Video Data Synchronization

Spoofing detection relies on the concurrent camera and wireless signals, thus it is crucial to synchronize both. The out-of-sync data may result in different semantic features, causing a high false alarming rate when they are used for spoofing detection [78].

Suppose that f_v denotes the frame per second (FPS) or frame rate of the camera, and Δ_v represents frame interval, i.e., the interval between two consecutive frames. The frame interval is normally constant and mathematically, we have $\Delta_v = 1/f_v$. The common frame rates for video are 24 FPS (standard), 30 FPS (close-second standard), and 60 FPS (for slow motion) [139]. Thus, the corresponding frame intervals are 42 ms, 33 ms, and 17 ms. Meanwhile, let f_w represent the CSI sampling rate at the receiver, which is much larger than f_v. We use N_c to denote the number of CSI measurements that a frame interval corresponds to. Note that if there is no packet loss, N_c is constant and equals $\frac{f_w}{f_v}$. Due to packet loss, unlike video frames, CSI measurements may have variable time intervals between them. As a result, each frame interval corresponds to a varying number of CSI measurements, i.e., N_c varies.

To address the issue, we apply linear interpolation to resample CSI measurements with a constant interval $\Delta_c = \frac{\Delta_v}{N_c}$, so that each video frame corresponds to a fixed amount of resampled CSI measurements.

4.3.5.2 CSI Data Preprocessing

The imperfect CSI can be caused by environmental noise, radio signal interference, and hardware imperfection. CSI data preprocessing includes (1) outlier removal and noise reduction, making CSI more accurately reflect the impact of

human activities; (2) Principle Component Analysis (PCA) [136], reducing the dimensionality of feature vectors to facilitate data analysis.

Outlier Removal and Noise Reduction: The collected CSI series may have some abrupt changes that are not caused by human activities, and such abnormal values should be corrected. Hampel filter is generally applied to identify and replace outliers (which differ significantly from other samples) in a given series [33, 121]. It uses a sliding window of configurable width to go over the input data. For each window, the median η and the median absolute deviation (MAD) λ can be calculated. The sample of the input is regarded as an outlier if it lies outside of the range of $[\eta - \gamma \cdot \lambda, \eta + \gamma \cdot \lambda]$, where γ is a pre-determined scalar threshold. In this way, the Hampel filter is able to identify all outliers in the CSI series and then replace them with the corresponding median.

Besides, CSI variations caused by human activities may occur at the low end of the frequency range. We thus utilize the moving average filter [137] to smooth the CSI series. This filter is simple to use and is optimal for retaining a sharp step response [112]. It computes the arithmetic mean of M input points at a time to produce each point of the output stream, where M is the pre-defined number of points. Thus, the high-frequency noise in the raw CSI measurements can be eliminated.

Figure 4.4a shows an example of applying outlier removal and noise reduction, where we effectively reduce outlier peaks and the strong high-frequency noise.

Dimension Reduction: We apply the PCA technique to decrease computational complexity by converting the received CSI into a set of orthogonal components (i.e., the most representative or principal components), which are influenced by human activity. Meanwhile, PCA also facilitates removing the uncorrelated noisy components. Figure 4.4b shows the CSI waveform after PCA,

and we can clearly observe CSI fluctuations that correspond to human activity and smooth waveform, indicating static periods within which there is no human activity.

4.3.5.3 Event Detection

Generally, when there is no movement in the monitored area, the CSI fluctuation is small and maintains stability in the time domain [170], while human activity would bring distinguishable CSI fluctuations [165]. To segment CSI waveforms corresponding to human activities, we need to determine the start and end points of the CSI time series, which covers as much of the activity-disturbed waveform as possible while minimizing the coverage of the non-activity portion.

We then calculate the moving variance σ^2 of each window $\mathbf{h} = \{h_1, h_2, \cdots, h_J\}$, where J is the pre-defined size of the window and h_j is the jth CSI value in this window. Mathematically, we have $\sigma^2 = \frac{\sum_{j=1}^{J}(h_j - \mu)^2}{J-1}$, where μ is the mean CSI value of the window $\mathbf{h}$. Empirically, the CSI segments during the human motion period show a much larger variance than those happening out of the period. Thus, we are only interested in the CSI segments with a variance larger than a predetermined threshold while ignoring the segments with a variance under this threshold. Later, those segments containing information about human activities will be further processed to extract semantic features about human activities. As shown in Figure 4.4c, by scanning the CSI variances, we can determine the start and end points for each event (two are detected, occurring during [34.9 s, 67.1 s] and [97.3 s, 130.7 s], respectively).

4.3.5.4 Feature Extraction

With CSI segments during human activities, a set of distinctive semantic features would be extracted and compared with those obtained from the video streams. The time period of human activities intercepted by CSI waveforms and video frames would usually match. Thus, the start and end times of each CSI segment, corresponding to that of human activity, will be recorded as the features. The frequency of CSI variations denotes the frequency of the event, which the video frames can also generate. Accordingly, we use the inter-peak intervals (i.e., the time period between successive peaks) to compute the frequency of occurred events.

As the first derivative of a peak switches from positive to negative at the peak maximum, it can be used to localize the occurrence time of each peak. However, noise may occasionally bring fake peaks and consequently false zero-crossings. Generally, the event usually cannot occur beyond a certain frequency. This observation enables us to develop a threshold-based fake peak removal algorithm. Specifically, if the calculated interval between the current peak with the previous one is less than $1/f_{max}$ (seconds), where f_{max} (Hz) denotes the maximum possible event frequency, this peak will be labeled as a fake one and thus discarded.

Let p_i denote the number of true peaks detected via an event-associated CSI segment, and $[t_1, t_2, \cdots, t_{p_i-1}]$ denote the corresponding sequence of inter-peak intervals. The event frequency f can be then estimated using the mean inter-peak interval, i.e., $f = \frac{p_i-1}{\sum_{j=1}^{p_i-1} t_j}$.

4.3.5.5 Consistency Checking

Given two tuples of features $\mathbf{f}^v = [f_1^v, \cdots, f_n^v]$ (from video) and $\mathbf{f}^c = [f_1^c, \cdots, f_n^c]$ (from CSI), where n is the number of extracted features, the multi-feature similarity score S can be calculated by comparing the similarity of each corresponding feature.

If the difference between the two features, each extracted from one of the two sources, is within a predefined threshold, we regard that both sources show the same feature. Mathematically, let s_j denote the single-feature similarity score and it can be obtained through

$$s_j = \begin{cases} 1 & \text{if } |f_j^v - f_j^c| \leq D_j \\ 0 & \text{otherwise} \end{cases}, j \in [1, \cdots, n]. \tag{4.3}$$

D_j is chosen empirically to achieve a high detection accuracy with a low false positive rate. We set the optimal thresholds for both the start and end times as 1.5 seconds, and that for the event frequency as 0.08 Hz. As a result, we have $S(i) = \sum_{j=1}^{n} s_j$. If all features extracted from both sources are consistent, i.e., $S(i) = n$, we determine that there is no spoofing attack present; otherwise, the video spoofing attack is detected.

4.4 Experimental Results

We implement an existing wireless liveness detection (e.g., [92, 92]) and our proposed attack on top of a typical surveillance camera (CODi HD 1080p [28]) and two USRP X300s [41], each equipped with an SBX-120 daughterboard [42].

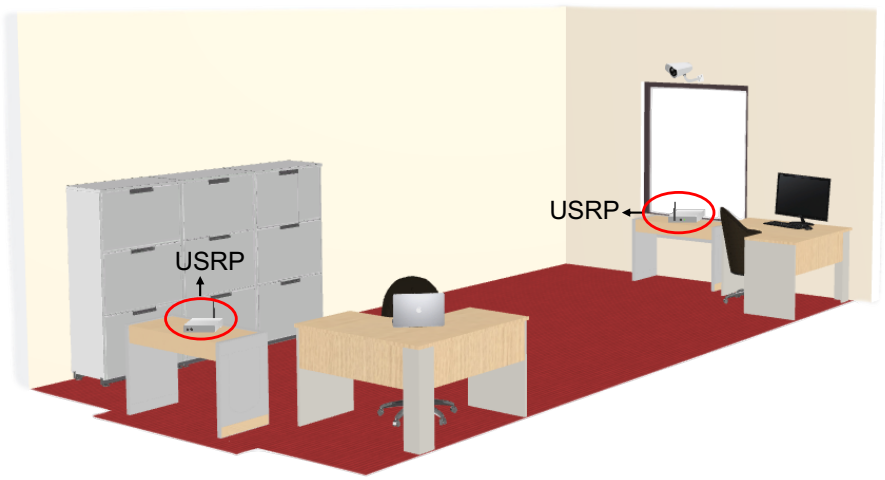


Figure 4.5: Layout of the experimental environment.



Figure 4.6: Three daily events.

4.4.1 Evaluation Setup

We perform the experiment in a laboratory office. For a good field of view, the camera is mounted on a wall 2.2 meters above the floor to monitor human activities in the office. It creates 1280×720 RGB images at 30 frames per second (FPS). Meanwhile, a wireless transmitter and receiver pair is utilized to verify the authenticity of the recorded video. Each node is a USRP X300.

The channel estimation algorithm runs at the receiver to extract the CSI for liveness detection. The attacker launches the phantom-CSI attack by replacing the original real-time video frames with pre-recorded fake ones (e.g., [12,72]) and

simultaneously manipulating the transmitted signal, aiming to make both the recorded video and the measured CSI at the receiver consistently show the same human activities. Figure 4.5 shows the positions of the camera, the transmitter, and the receiver.

We ask the user to perform the following three daily activities, as shown in Figure 4.6, including E_1: walking on the floor; E_2: sitting on a chair and then standing on the floor; E_3: moving the arm up and down. We consider two typical attack scenarios based on the goal of the attacker.

- *Fabricating Event*: when no event occurs in the monitored area, the attacker feeds a video with a motion to the camera and synchronously makes the CSI detect the same motion.
- *Hiding Event*: when motion appears in the area, the attacker feeds a static shot to the camera and meanwhile makes CSI exhibit no motion.

Metrics: We use the following two evaluation metrics.

- *True Positive Rate*: this is the percentage of actual spoofing incidents that are correctly detected, denoting the accuracy of the spoofing detection.
- *False Positive Rate*: this is the proportion of all negatives (i.e., when no spoofing occurs) that are wrongly categorized as cases with spoofing.

4.4.2 Effectiveness of Channel Manipulation

In the section, we utilize examples to demonstrate the effectiveness of channel manipulation in different environments, which aims to make the receiver obtain the channel specified by the attacker.

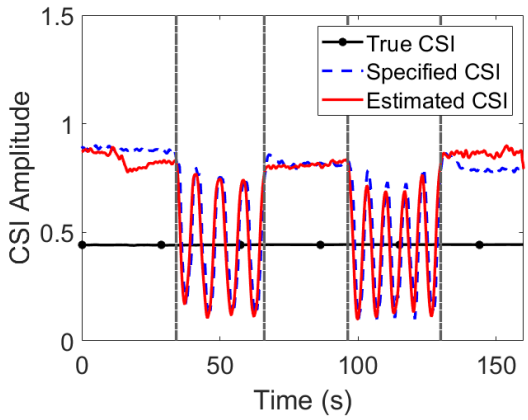
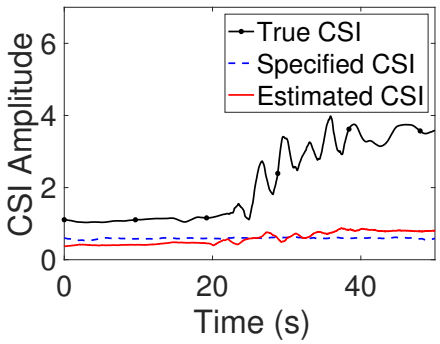


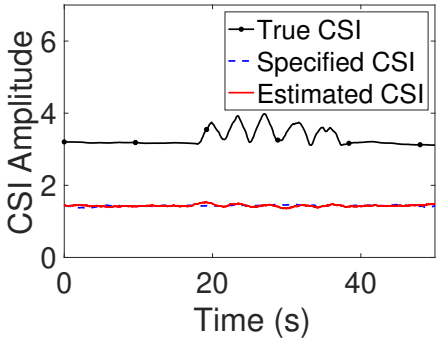
Figure 4.7: Channel manipulation in a static environment.

Static Environment: Figure 4.7 presents the true CSI between the transmitter and the receiver, the CSI specified by the attacker, and the estimated CSI at the receiver in a static environment (with no human activity). We can observe that the estimated CSI is greatly similar to the specified one, while both significantly deviate from the true CSI. The estimated CSI further causes the receiver to believe that there are human activities during the periods from 34.2 s to 66.1 s, and from 96.3 s to 130.0 s. The activity repeats four and five times in the two periods, respectively. When the attacker injects a fake video with such events (e.g., waving arms) into the camera, the system would alert as the true CSI and the video detect inconsistent results without our attack, whereas our attack can successfully bypass the CSI-aided liveness detection system.

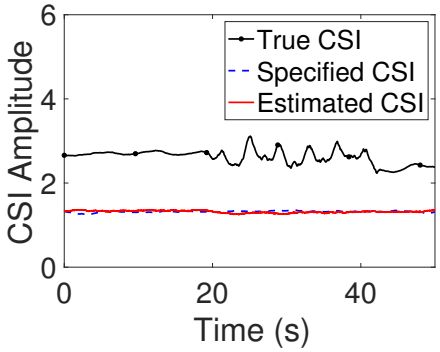
Dynamic Environment: Figure 4.8 presents the true CSI between the transmitter and the receiver, the CSI specified by the attacker, and the estimated CSI at the receiver in an environment with human activities present. Human activities bring fluctuations in the CSI waveforms. Specifically, a walking activity involves significant body movements and location changes. Thus, it causes significant CSI changes over time. However, an in-place activity, i.e., sitting/standing



(a) Walk.



(b) Sit/stand.



(c) Wave arms.

Figure 4.8: Channel manipulation in a dynamic environment.

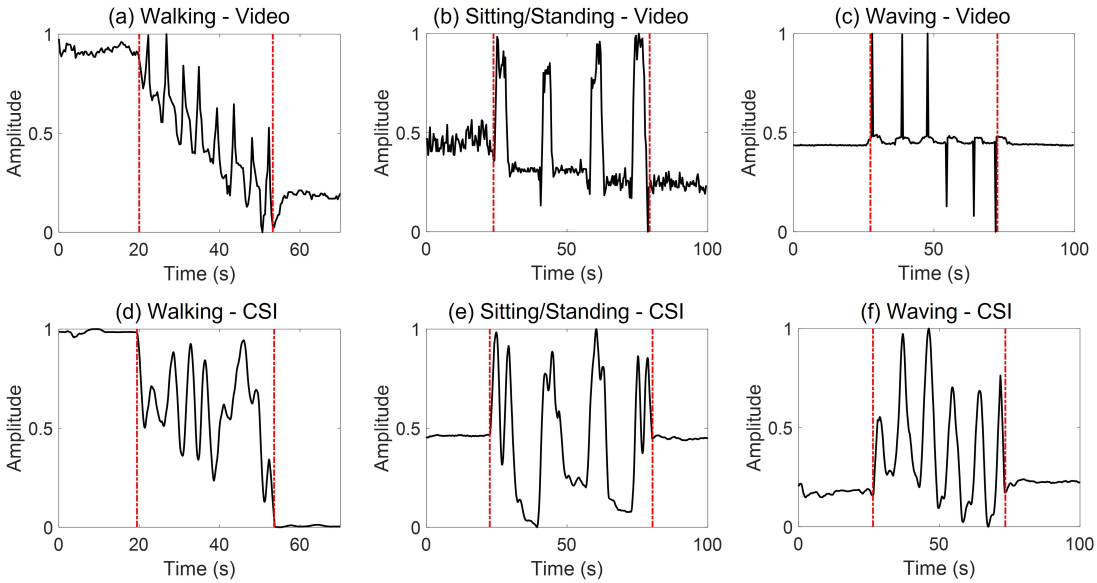


Figure 4.9: Video and the CSI signals when fabricating events.

and waving arms, only involves relatively smaller body movements and does not cause significant CSI changes. Also, channel manipulation enables the receiver to obtain an estimated CSI that is almost flat and close to the CSI specified by the attacker, causing the receiver to believe that no event happened. Thus, when the attacker injects a fake static video into the camera and meanwhile human activities occur in the monitored area, the system may alert without our attack due to the inconsistent detection results from the video and CSI, whereas our attack can make the CSI present no event and succeed to defraud the CSI-aided liveness detection system.

4.4.3 Two Attack Cases

Case I - Fabricating Nonexistent Events: The attacker makes the estimated CSI at the receiver side change with the injected fake video containing scenes of human activities, where the environment is in fact static.

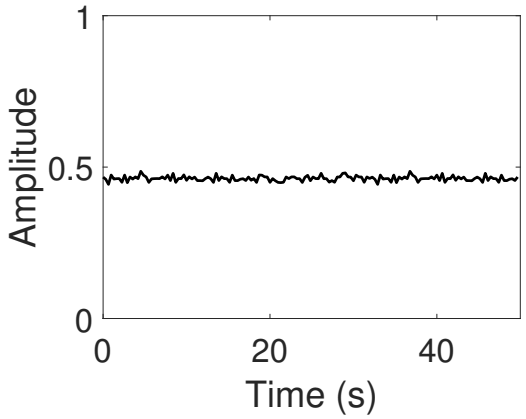
Table 4.1: Different human activity combinations.

Number of events	Human activity combination
1	E1 only; E2 only; E3 only
2	E1+ E2; E2 + E3; E1 + E3
3	E1+E2+E3

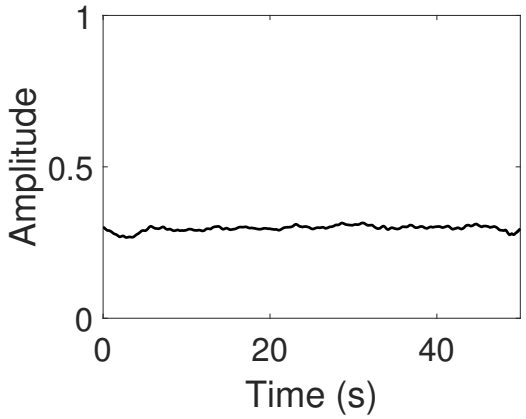
Figure 4.9 compares the time series of the video and CSI when the fake video contains different activities. As shown in Figures 4.9a and 4.9d, with the video signal, the extracted feature tuple (including start time, end time, and frequency) for walking equals $(20.0\ s, 53.2\ s, 0.15\ Hz)$; with the CSI data stream, the corresponding tuple is $(19.5\ s, 53.5\ s, 0.15\ Hz)$. The absolute errors between features from the two sources are thus 0.5 s, 0.3 s, and 0. As the optimal thresholds for start time, end time, and event frequency are 1.5 s, 1.5 s, and 0.08 Hz, the similarity score equals 3. We have similar observations for the cases of sitting/-standing (Figures 4.9b and 4.9e) and waving arms (Figures 4.9c and 4.9f). In all cases, our attack successfully bypasses wireless video liveness detection.

Case II - Hiding True Events: The attacker aims to make the CSI disclose no human activities when feeding a fake video containing only static scenes, though the user performs activities in the monitored area.

When the spoofed video contains no person, *OpenPose* extracts no keypoints from it and thus shows the empty output. When the spoofed video of a static scene contains a still user, the extracted keypoints have no movement, as shown in Figure 4.10a. Figure 4.10b plots the corresponding CSI time series obtained at the receiver side when the user performs events (e.g., walking). From the video and CSI signals, the respective extracted features are consistent. Thus, the wireless liveness detection system generates no alarm of spoofing detection, verifying the success of the proposed attack.



(a) Video stream.



(b) CSI stream.

Figure 4.10: Video and CSI signal comparison when hiding events.

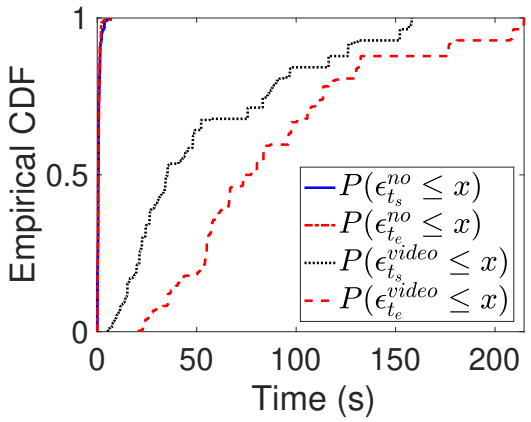
4.4.4 Overall Attack Impact

We test both static and dynamic environments. Each has two scenarios: (i) the attacker launches a video spoofing only attack; (ii) the attacker launches the proposed attack. For comparison, we also test the performance of the wireless liveness detection system when there is no attack. The above three scenarios are referred to as "video", "csi", and "no", respectively. We consider the number of actual or spoofed events ranging from 1 to 3, and test 7 different combinations of the three daily events (E_1, E_2, and E_3), as shown in Table 4.1, where "$E_i + E_j$"

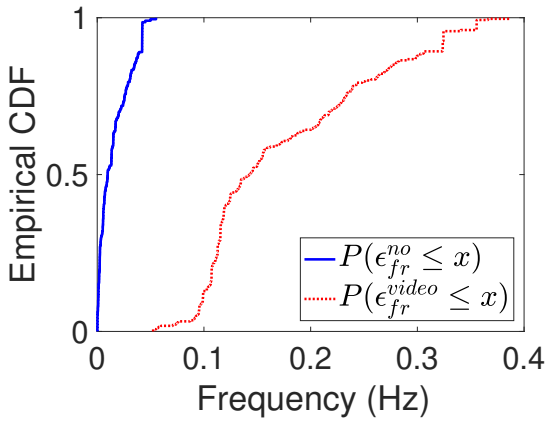
$(i, j \in \{1, 2, 3\})$ denotes that events E_i and E_j occur sequentially. For every combination under each case, we perform 10 trials. Thus, in total, we perform $(2 \times 2 \times 7 + 7 + 1) \times 10 = 360$ attempts.

Event Feature Matching: Let $\epsilon_{t_s}^{sce}$, $\epsilon_{t_e}^{sce}$, and ϵ_{fr}^{sce} denote the measured absolute estimation errors for start time, end time, and event frequency, in scenario sce ($sce \in no, video, csi$). We show the empirical cumulative distribution functions (CDFs) of $\epsilon_{t_s}^{no}$, $\epsilon_{t_e}^{no}$, $\epsilon_{t_s}^{video}$, and $\epsilon_{t_e}^{video}$ in Figure 4.11a. Also, Figure 4.11b shows the CDFs of ϵ_{fr}^{no} and ϵ_{fr}^{video}. We see that the absolute errors for all three features are always small with no attack. Specifically, $\epsilon_{t_s}^{no}$ and $\epsilon_{t_e}^{no}$ are less than 2.0 s with probabilities 92.9% and 98.6%, respectively; ϵ_{fr}^{no} is always less than 0.045 Hz. Such results clearly show that without any attacks, the co-existing video and CSI data are highly consistent, i.e., the false positive rate of wireless liveness detection is low. On the other hand, for a video spoofing only attack, the features extracted from the two sources show an apparent mismatch. We observe that $\epsilon_{t_s}^{video}$ and $\epsilon_{t_e}^{video}$ are larger than 7.8 s and 23.8 s with probability 97.6%, respectively. Also, ϵ_{fr}^{video} ranges from 0.05 to 0.39 Hz, and is larger than 0.07 Hz with a probability of 97.6%. These results convincingly demonstrate that the wireless liveness detection system can effectively detect video spoofing only attacks.

Figure 4.12 presents CDFs of $\epsilon_{t_s}^{csi}$, $\epsilon_{t_e}^{csi}$, and ϵ_{fr}^{csi}. We observe that the absolute estimation errors for all three features become consistently small. Particularly, $\epsilon_{t_s}^{csi}$ and $\epsilon_{t_e}^{csi}$ are less than 1.5 s with probabilities 93.8% and 95.3%; ϵ_{fr}^{csi} is less than 0.042 Hz with probability 98.6%. These results show that our attack can successfully synchronize the CSI and video signals observed at the receiver. With consistent CSI and video data streams, the wireless liveness detection system would fail to send out an alarm when video spoofing attacks happen.



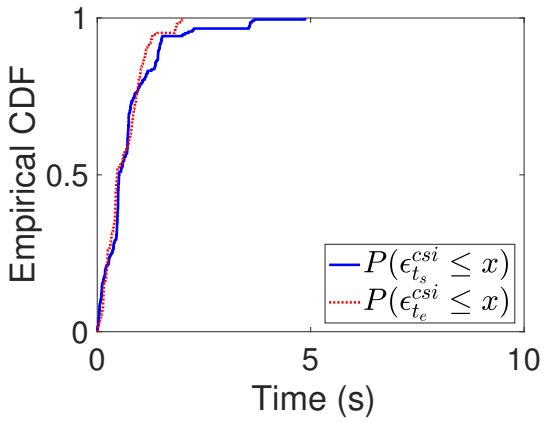
(a) Start/end time.



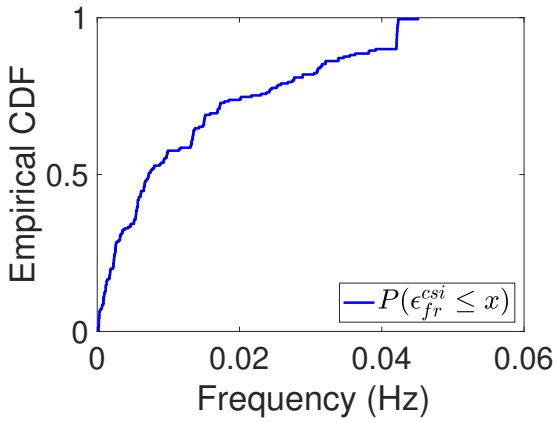
(b) Event frequency.

Figure 4.11: CDF of the extracted features in a normal situation and when a video spoofing only attack happens.

Impact of Feature Count: By comparing extracted features from both sources, it can determine whether the recorded video is spoofed or not. Table 4.2 presents TPRs and FPRs of the liveness detection system when the video spoofing only attack happens and when our attack initiates. We see that if using two features (start and end time), the overall TPR can be up to 1 when there is a video spoofing only attack, while it is decreased to as small as 3.1% when the proposed attack is launched. This implies that the CSI-aided liveness detection system can reliably detect traditional video spoofing attacks, but becomes ineffective with



(a) Start/end time.



(b) Event frequency.

Figure 4.12: CDF of the extracted features with our attack.

our attack (with just 9.1% accuracy). Besides, we observe that the proposed attack rarely has an impact on FPR, which maintains a relatively low value. Moreover, when using three features (start time, end time, and event frequency) for event detection, we have similar observations. Specifically, compared with the video spoofing only attack, the TPR of our attack is slightly increased but still below 4.5%, again indicating the attack effectiveness against the wireless liveness detection scheme.

Impact of Event Type: For different types of events in the spoofed video, we construct respective phantom CSI to launch our attack. As shown in Table 4.3,

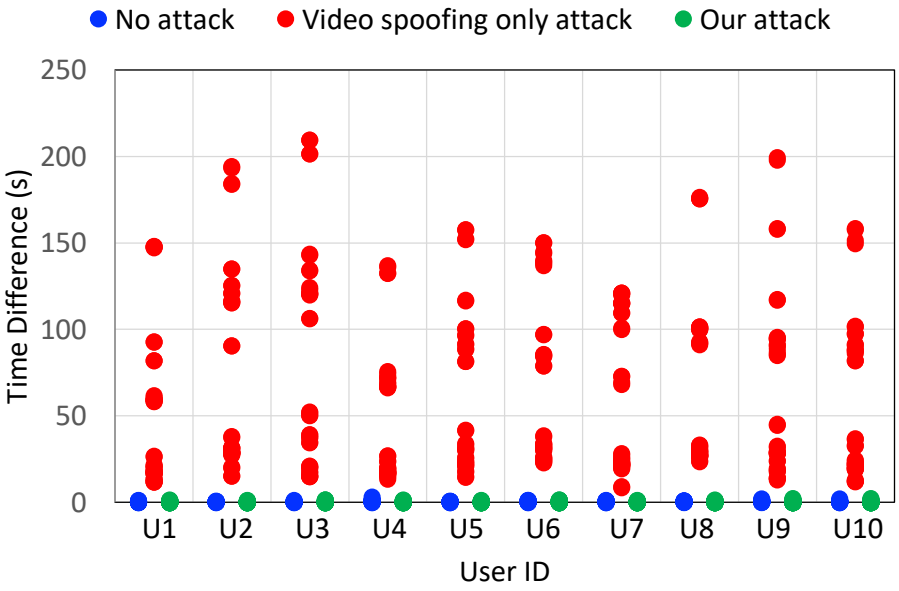


Figure 4.13: Event start time discrepancies.

Table 4.2: Wireless video liveness detection vs. feature count.

Count	Two		Three	
Case	*video spoofing only attack*	*our attack*	*video spoofing only attack*	*our attack*
TPR	1	3.1%	1	4.4%
FPR	4.4%	4.4%	13.3%	13.3%

the TPR of the liveness detection system is always 100% under video spoofing only attacks regardless of event type, while it drops dramatically to 5.0%. 6.0%, and 4.3% for E_1, E_2, and E_3, respectively. Also, the FPRs across all event types under both scenarios are no larger than 10%. These results demonstrate our attack is robust against event type.

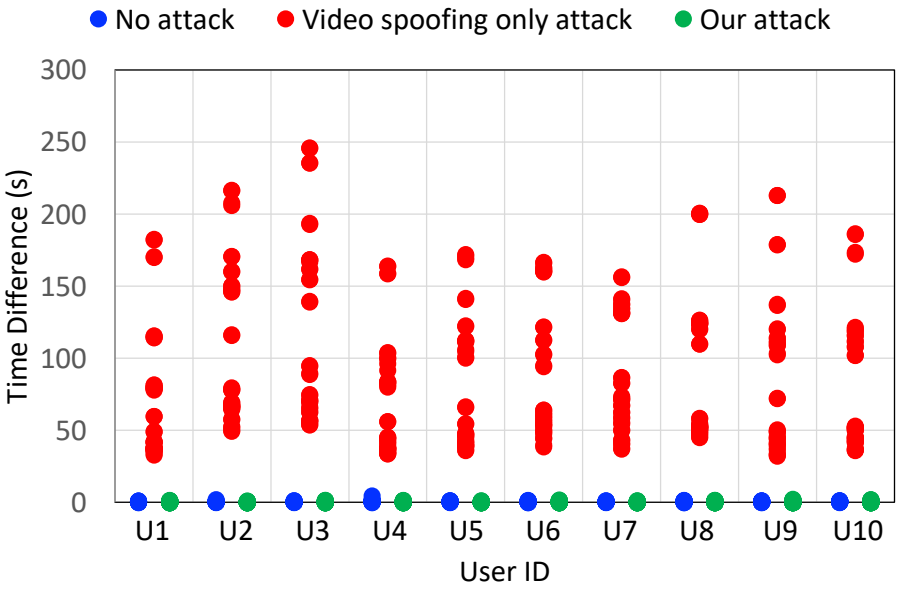


Figure 4.14: Event end time discrepancies.

Table 4.3: Impact of different event types.

	E_1		E_2		E_3	
Case	*Video**	*our attack*	*Video*	*our attack*	*Video*	*our attack*
TPR	1	5.0%	1	6.0%	1	4.3%
FPR	10.0%	10.0%	10.0%	10.0%	7.1%	7.1%

* *Video*: *video spoofing only attack.*

4.4.5 User Study

We recruited 10 volunteers (aged 18-35 years old; 5 self-identified as females and the rest as males).[2] Every participant was asked to perform each motion event in Table 4.1 twice in a normal scenario (i.e., without any attacks). We also recorded the corresponding videos and replayed them in the other two cases, i.e., the video spoofing only attack and the proposed attack. For each case, we test

[2]Our study has been approved by our institution's IRB.

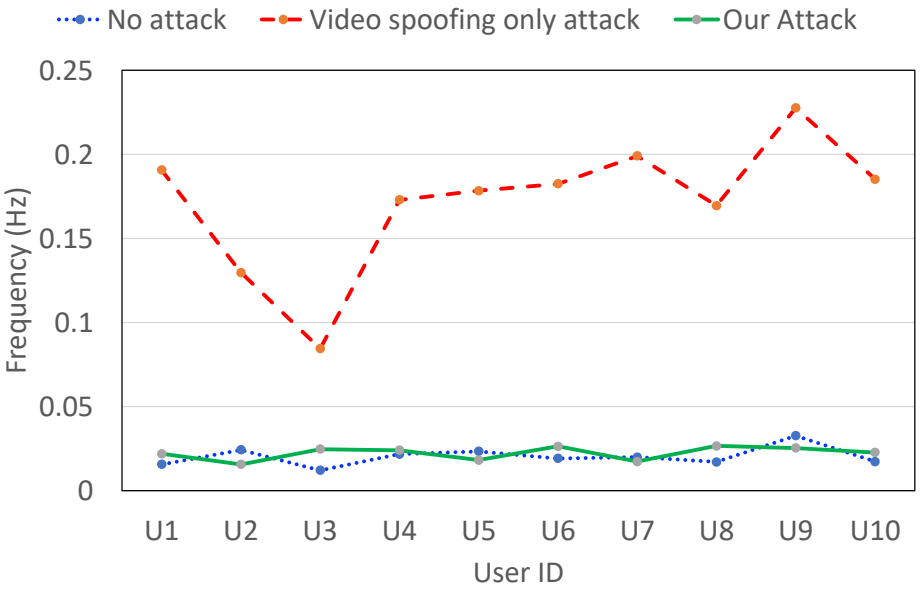


Figure 4.15: Mean frequency discrepancies.

Table 4.4: The list of voice commands we test.

ID	Command	Word #
C1	Please call 911	3
C2	Please play music	3
C3	Please open the door	4
C4	Please turn on the TV	5
C5	Please open the notification center	5

the performance of wireless video liveness detection for $(3+4+1)\times 2 = 16$ trials per participant.

Figures 4.13, 4.14, and 4.15 illustrate respective feature differences. We see that all feature differences are consistently low with no attack. Specifically, for the start/end time, the feature difference is less than 1.5 s while it is less than 0.03 for the frequency. With the video spoofing only attack, each feature discrepancy of all users increases greatly, which becomes an effective indicator of the existence of video spoofing. However, when the proposed attack is launched, all feature differences become consistently small again, similar to that in the scenario

of no attack. These results convincingly demonstrate that an attacker can effectively bypass the wireless video liveness detection system with spoofed videos by launching the phantom-CSI attack.

4.5 Attack Against Wireless Voice Liveness Detection

Voice assistants, such as Amazon Alexa and Google Assistant, have been embedded in a slew of digital devices (e.g., smartphones and smart TVs). Due to the open nature of voice assistants' input channels, a malicious attacker could easily record people's use of voice commands [5, 54], and even build a model to synthesize a victim's voice [111]. The attacker plays pre-recorded or synthesized voice commands, which may spoof voice assistants, causing these devices to perform operations against the desires of their owners [15,192]. Wireless voice liveness detection cross-checks the consistency between simultaneously obtained audio and wireless signals. Specifically, we preprocess audio signals using the spectral subtraction technique [16] to remove the background noise, where the average noise spectrum is first estimated and then subtracted from the noisy speech spectrum. By extracting semantic features (e.g., start time, end time, and word count) from the audio and wireless signals, spoofing attacks via pre-recorded or synthesized voice can be then detected [109, 110, 123, 141, 193]. Our attack can generate fake CSI and make it synchronized with the voice signal played by a speaker.

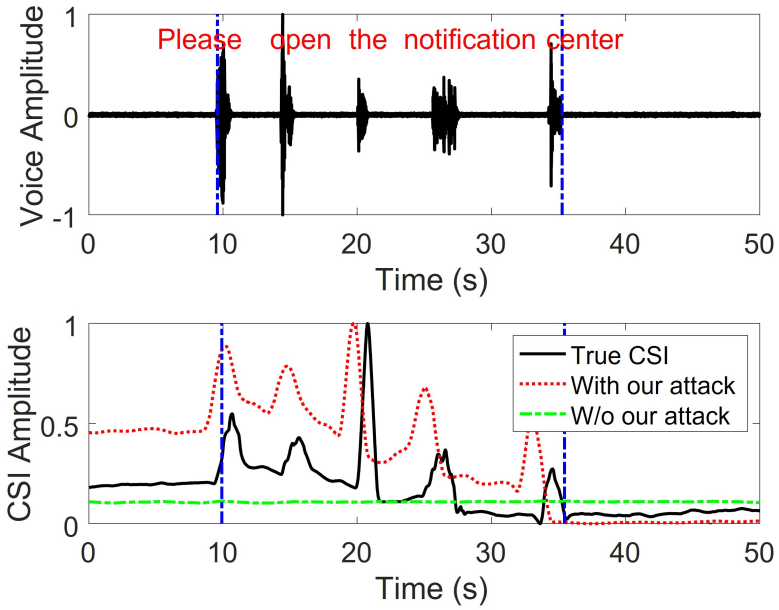


Figure 4.16: An example of a wireless-based voice liveness detection.

4.5.1 Implementation Setup

We implement wireless voice liveness detection and our attack in real-world environments. We utilize USRP X300 as a transceiver to collect CSI, and a microphone to collect voice signals. The transmitter and the receiver are placed at opposite positions relative to the target speaker. We randomly select 5 commands (C1-C5) from a list of the best Siri voice commands for a variety of daily tasks [22], as shown in Table 4.4. The evaluation metrics are the same as those for assessing the attack against wireless video liveness detection.

4.5.2 Case Study

We compare the following cases: (1) *Normal Case*: the user speaks command C5 in Table 4.4; (2) *Voice Replay Only*: a speaker plays C5; (3) *Our Attack*. Figure 4.16 plots corresponding voice and CSI signals.

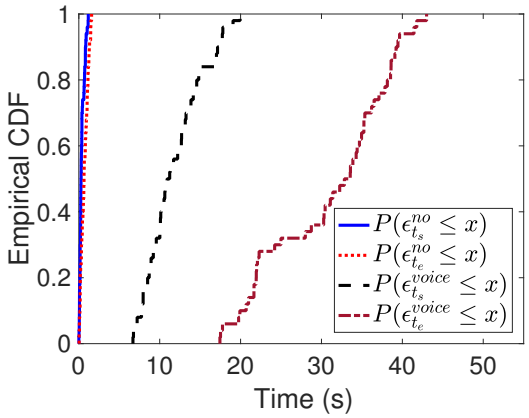


Figure 4.17: CDFs of start/end time for normal and voice spoofing attack only cases.

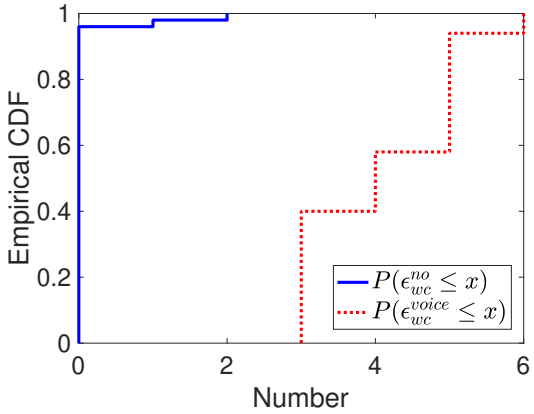


Figure 4.18: CDFs of word count for normal and voice spoofing attack only cases.

Normal Case: From the voice signal, the speaking interval is [9.6 s, 35.2 s] and there are 5 separate segments full of fluctuations, corresponding to 5 words. Meanwhile, the fluctuations of the CSI time series (referred to as "true CSI" in Figure 4.16) happen with the occurrence of the command; accordingly, we get the speaking interval [9.9 s, 35.3 s] and the word count 5 (as the sharp and rise pattern appears 5 times, each caused by speaking a word). Thus, the errors between corresponding features extracted from the voice and CSI signals are all small, indicating that both signals are consistent.

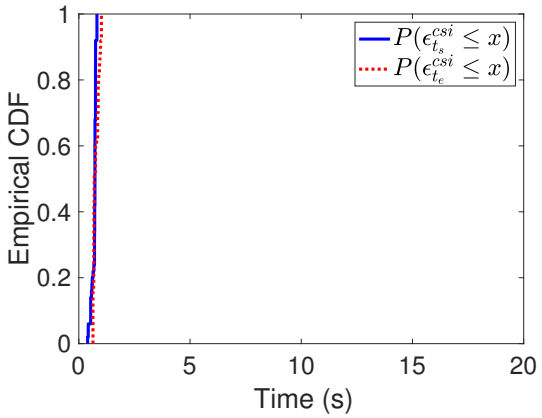


Figure 4.19: CDFs of start/end time when the proposed attack is launched.

Voice Replay Only: When an attacker launches a voice spoofing only attack (with no mouth motion), the voice signal that the microphone captures maintain almost unchanged. However, the CSI waveform (referred to "W/o our attack' in Figure 4.16) becomes flat, demonstrating that the CSI would detect no event. The inconsistency of event detection via voice and CSI data facilitates the detection of the voice spoofing attack.

Our Attack: The waveform of the estimated CSI is highly similar to the true one. The correspondingly extracted features are 9.0 s, 34.4 s, and 5. By comparing them with the features extracted from the voice signal, we obtain the absolute errors as 0.6 s, 0.8 s, and 0, each of which is smaller than the respective threshold, indicating the failure of the liveness detection.

4.5.3 Overall Performance

For each command in Table 4.4, we perform the proposed attack 10 times. We synchronize the CSI and spoofed voice signals each time to bypass the wireless-based liveness detection system. For comparison, we also record the performance of the normal case with no attack, and the voice spoofing only attack. We refer

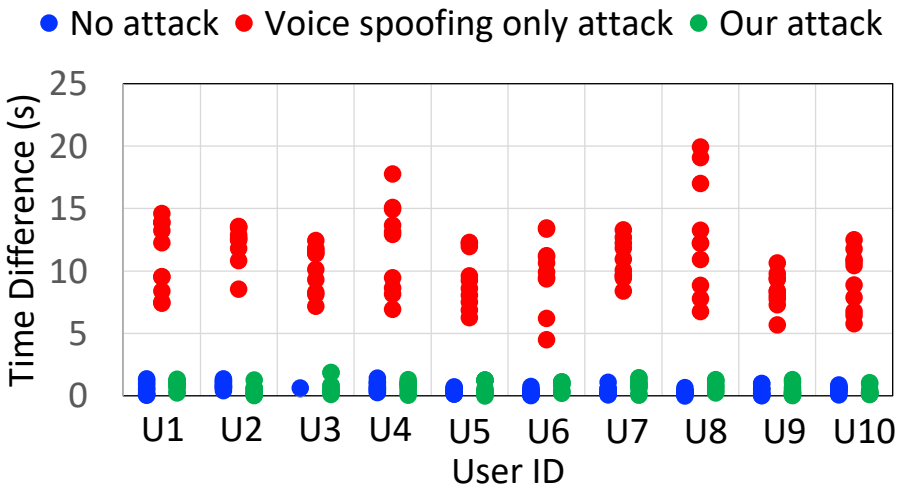


Figure 4.20: Speaking start time differences.

to the above three scenarios as "*csi*", "*no*", and "*voice*", respectively.

Speaking Activity Detection: Let $\epsilon_{t_s}^{sce}$, $\epsilon_{t_e}^{sce}$, and ϵ_{wc}^{sce} denote the absolute estimation errors of start time, end time, and word count in scenario *sce*, where $sce \in \{no, voice, csi\}$. Figure 4.17 shows CDFs of $\epsilon_{t_s}^{no}$, $\epsilon_{t_e}^{no}$, $\epsilon_{t_s}^{voice}$ and $\epsilon_{t_e}^{voice}$. We see that $\epsilon_{t_s}^{no}$ is always less than 1.2 s and $\epsilon_{t_e}^{no}$ is less than 1.5 s with probability 98.0%, while $\epsilon_{t_s}^{voice}$ and $\epsilon_{t_e}^{no}$ are apparently larger. Meanwhile, ϵ_{wc}^{no} equals 0 with probability of 96.0%, whereas ϵ_{wc}^{voice} ranges from 3 to 6, as shown in Figure 4.18. These results convincingly imply that the wireless liveness detection system can effectively recognize voice spoofing attacks via feature differences. Figure 4.19 presents CDFs of $\epsilon_{t_s}^{csi}$ and $\epsilon_{t_e}^{csi}$. We see that $\epsilon_{t_s}^{csi}$ and $\epsilon_{t_e}^{csi}$ are always less than 0.8 s and 1.1 s, respectively. Also, ϵ_{wc}^{csi} is always 0. Evidently, with our attack, the extracted features from both voice and CSI signals are highly consistent, leading to the failure of the liveness detection system.

Impact of Feature Count: Table 4.5 compares TPR and FPR for different cases when utilizing two features (start and end time) or three features (start

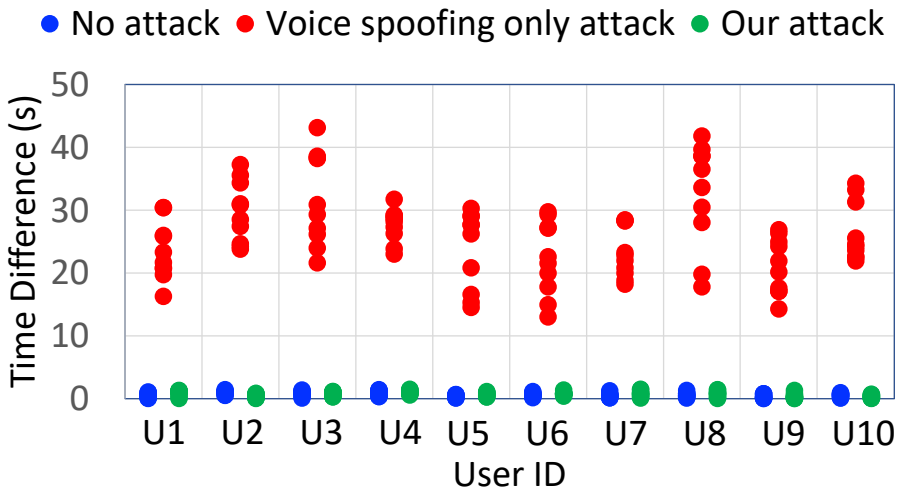


Figure 4.21: Speaking end time differences.

Table 4.5: Wireless voice liveness detection vs. feature count.

	Two			Three		
Case	no	voice	csi	no	voice	csi
TPR	N/A	1	0	N/A	1	0
FPR	6.0%	6.0%	6.0%	8.0%	8.0%	8.0%

time, end time, and word count) to detect spoofing attacks. We observe that regardless of the feature count, the wireless voice liveness detection system can achieve a TPR of 100% to recognize voice spoofing only attacks, while the TPR plummets to 0 with the proposed attack, implying that a voice replay attack is no longer to be correctly recognized. Meanwhile, we see that the FPR maintains small and consistent in different cases, demonstrating that our attack does not raise extra false alarms.

Impact of Number of Spoken Words: Aligned with existing work [110, 155, 171, 191], we also investigate the impact of the count of spoken words. As shown in Table 4.6, for word count ranging from 3 to 5, the FPR of the liveness detection system is always 100% without considering our attack, while it drops 0

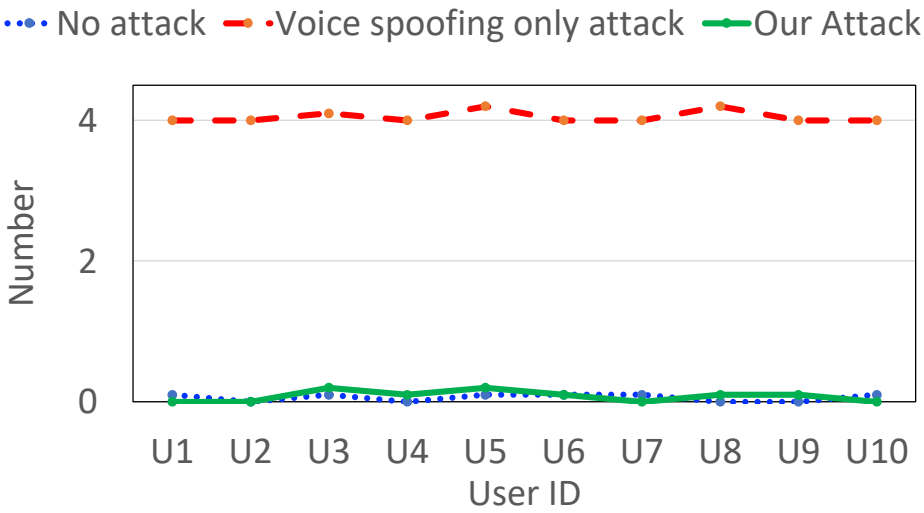


Figure 4.22: Mean word count differences.

Table 4.6: Wireless voice liveness detection vs. word count.

	3		4		5	
Case	voice	csi	voice	csi	voice	csi
TPR	1	0	1	0	1	0
FPR	10.0%	10.0%	10.0%	10.0%	5.0%	5.0%

under our attack. This verifies the robustness of our attack against word count. Also, the FPRs across all word counts for two cases are no larger than 10%, and the small fluctuation in FPR appears due to the minute changes in the environment.

4.5.4 User Study

The 10 volunteers (as described in Section 4.4.5) were asked to speak each command in Table 4.4 twice in a normal scenario. We also recorded the voices and replayed them in the other two cases with the voice spoofing only attack and our proposed attack, respectively. Figures 4.20, 4.21, and 4.22 illustrate respective feature discrepancies. We have the following observations. With no attack, the

differences in both start time and end time are consistently low (less than 1.5 s) across all users. Also, the mean difference in word count for each user is always small (less than 0.1). However, for a voice spoofing only attack, the discrepancies in all features for all users jump sharply. These results convincingly show that the wireless liveness detection system can robustly detect voice spoofing only attacks. With our attack, however, those feature discrepancies decrease to small values, similar to that in the scenario of no attack, indicating that spoofed voice can successfully bypass the wireless voice liveness detection system.

4.6 Discussions

4.6.1 Limitations

Cross-modality Sensing: Currently, the proposed attack targets compromising wireless video/voice liveness detection systems. Thus, except for generating fake CSI time series, it should also perform a video spoofing or voice replay attack simultaneously. In general, phantom-CSI can be utilized alone to confuse any CSI-based applications, such as keystroke recognition techniques [7, 50] or vital signs inference methods [81, 98].

Complex Human Activities: Our work currently just considers three popular daily activities (i.e., walking, sitting/standing, and waving arms), while a person may perform more complex activities (e.g., playing games). It may thus become difficult to construct phantom CSI associated with these activities. Accordingly, we expect that if the adversary could pre-collect CSI traces from such activities, she can feed them to the wireless liveness detection system to launch the proposed attack.

Scenarios Where Real CSI is Unknown: The proposed work may fail to make the receiver obtain the specific CSI in scenarios where real CSI is unavailable or cannot be correctly predicted. Machine learning-based approaches have demonstrated success in achieving accurate CSI prediction (e.g., [104, 183]). They thus can be added to our technique to improve the attack effectiveness, and we leave such integration to our future work.

Channels with Noise and Interference: Normally, if the real channel has noise and interference, existing wireless liveness detection may not work, and thus in this case, it is unnecessary to explore the feasibility of the proposed attack. The directional antenna can be adopted to eliminate CSI noises and other interferences.

4.6.2 Countermeasures

The proposed attack needs to compromise the transmitter and cancel the real channel effect before injecting phantom CSI to mislead the target system. Intuitively, to defend against such attacks, we can *utilize a trustful transmitter or a protected frequency* (on which the attacker is not allowed to inject signals). Such methods, however, would incur extra costs. Alternatively, we can also directly stop the attacker from obtaining the true wireless channel information by leveraging *friendly jamming* [46]. Specifically, an ally jamming sends out intentional radio interference signals, i.e., jamming signals, to the wireless channel to prevent the attacker from measuring the real CSI, while the receiver itself can eliminate the impact of interference signals to guarantee that the wireless liveness detection system still works when the proposed attack is not launched. Similarly, this defense brings additional overhand for jamming hardware.

To validate the liveness detection result, another viable defense strategy is to integrate extra sensors. For example, the work [138] uses thermal infrared (IR) images to detect live signals; motion sensors can be employed to detect the presence of humans from the radiation of their body heat [69–71]; by exploiting the circular microphone array of the smart speaker, voice spoofing attacks can be thwarted [108]. However, these extra sensors are not always available, and the deployment of additional infrastructure requires authentication of the new sensor data that may potentially introduce a new attack surface [92].

Chapter 5

Future Work

This chapter [1] discusses three future trends for wireless human profile information (HPI) inference.

To utilize existing wireless HPI inference techniques, we often need to first establish a wireless environment. This involves setting up wireless transceivers or commercial off-the-shelf WiFi devices or software-defined radio (SDR) systems-around the target user. As IoT devices with wireless connectivity become increasingly pervasive and crucial in various applications, they may handle sensitive HPI that needs protection. Furthermore, the advancement of machine learning techniques, complementing traditional signal processing methods, offers new possibilities for wireless HPI inference. Machine learning-based techniques can learn from and adapt to the environment through experience [25]. Additionally, the use of mmWave communication technology is growing in emerging wireless applications, such as virtual reality (VR) [38]. mmWave radar signals, with their shorter

[1] This chapter was published in Q. He, E. Yang and S. Fang, "A Survey on Human Profile Information Inference via Wireless Signals," in IEEE Communications Surveys & Tutorials, doi: 10.1109/COMST.2024.3373397. Used with Permission.

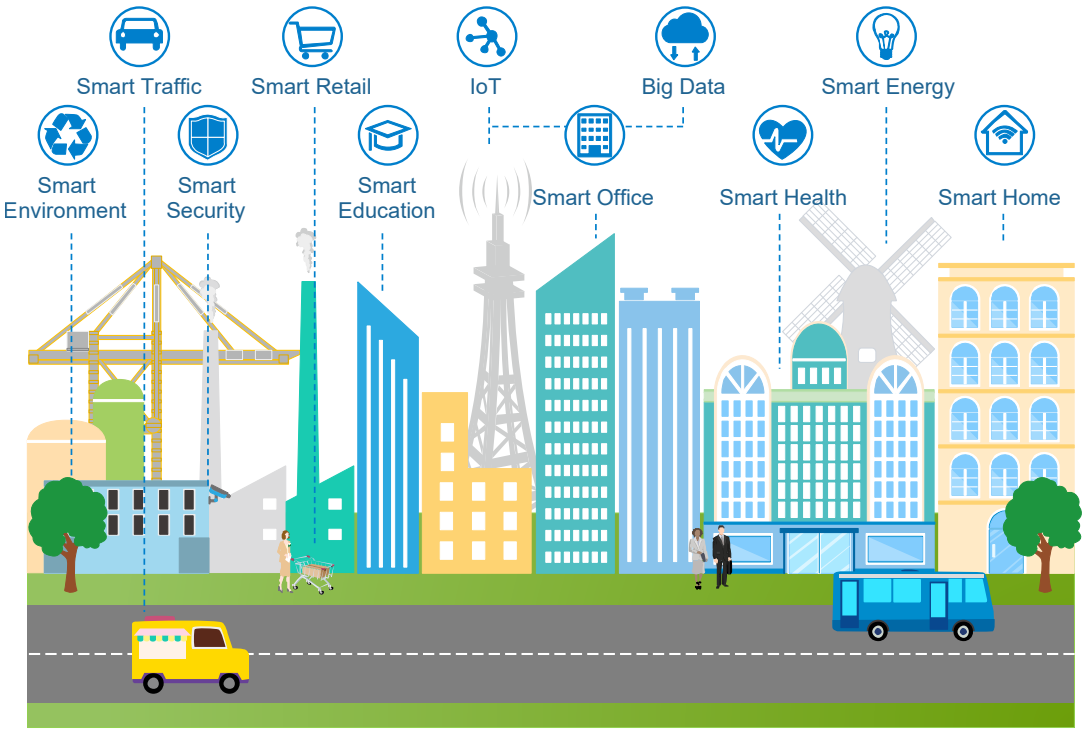


Figure 5.1: Extensive applications of IoT.

wavelengths compared to WiFi signals, can be used to detect subtle movements. In the following sections, we will discuss the future trends and challenges in wireless HPI inference. These include the popularity of Internet-of-Things (IoT) devices, integration with machine learning techniques, and increasing adoption of millimeter wave (mmWave) communications.

5.1 Challenges with IoT Devices

IoT is a system of interrelated computing devices connected to a network and/or to one another, exchanging data without necessarily requiring human-to-machine interaction [29]. Due to their low-cost and low-power characteristics, IoT devices

have been extensively deployed in various domains, including smart homes, transportation, health care, and manufacturing, as shown in Figure 5.1. According to a recent study about the global IoT market, it is expected that by 2025, there will be approximately 27 billion connected IoT devices worldwide [65]. However, the proliferation of IoT devices brings several key challenges:

Challenge 1: Ensuring the security and privacy of sensed data transmitted wirelessly. IoT devices may monitor users' private activities, and the data they collect often carry a great potential for privacy risks regarding the use of the data and its access [122, 175, 177]. Intuitively, the behaviors of various smart devices, such as smart door locks, lighting control systems, and wireless security cameras can be easily affected by human activity. For instance, in a smart home, the lighting condition can be adjusted automatically according to whether the user enters or walks out of the room, and the door will open if the identity is verified for the user who wants to enter the room [84]. As wireless signals may carry important information about these devices' behavior, they can be captured and analyzed to infer HPI.

Challenge 2: Keeping multiple IoT devices to be well-calibrated and synchronized. A vast array of devices with their own set of capabilities are sourced from different manufacturers. Therefore, inter-device data exchange for IoT devices is challenging. First, IoT devices are often portable, and their high mobility may introduce noise and interference, decreasing communication performance. Second, synchronization among diverse IoT devices is crucial when data from multiple devices need to be combined or compared, especially for multi-modal systems that collect data from different sensors, such as capturing audio, video, temperature, motion, and more. For instance, if a surveillance system captures both video and audio [92], a sound should correspond accurately with the visual event causing

it; otherwise, it may cause false alarms to make the security system unreliable.

Challenge 3: Addressing the heterogeneity among various IoT devices. On the one hand, different IoT devices may support incompatible communication standards (e.g., WiFi, Bluetooth, and ZigBee) and have varying sensing modalities. For example, a Zigbee smart bulb might struggle to relay its readings to a smart home hub that exclusively supports WiFi or Bluetooth. On the other hand, data quality may differ in different IoT devices. For example, a 4K security camera might capture 4K video while a normal one may only capture 1080p. Such disparities can lead to inconsistencies in data quality, which may deteriorate data aggregation or data fusion.

Challenge 4: Designing new authentication method in emerging IoT devices. Most emerging IoT devices lack a user interface (e.g., a touchscreen or keypad), and traditional authentication methods using direct text entry become inapplicable. New secure and robust mechanisms are thus required to enable wireless communication among IoT devices [63]. For example, the study [55] develops a robust communicating system for a mobilizable IoT network. It exploits ultrasonic signals at a frequency corresponding to the target receiver, forcing the inertial sensors to resonate, so as to convey information. Also, to authenticate users of IoT devices, [96] presents a virtual sensing technique that allows IoT devices to virtually sense user touches on the devices.

Challenge 5: Achieving real-time sensing while improving energy efficiency. Furthermore, real-time processing and analysis of HPI within IoT devices have become essential requirements for various applications, especially for health monitoring, intruder detection, and fall detection systems. Responding promptly to the new arrival of data and analyzing it without delay is crucial in these contexts. This presents several challenges, including resource constraints of IoT de-

vices, data quality, environmental noise, and the need for robust and adaptable models. Another critical aspect to consider is that devices continuously monitoring and transmitting data rapidly consume energy. To address these issues, future research should focus on developing lightweight machine learning models, efficient data pre-processing techniques, and adaptive learning mechanisms that can operate within IoT device constraints while ensuring accurate and real-time analysis.

5.2 ML for Wireless HPI Inference

Machine Learning (ML) plays an essential role in wireless HPI inference systems. ML approaches leverage wireless signals to sense our environment, detect and monitor our activities, and localize and track the users. For example, the work [159] proposes a Hidden Markov Model (HMM) for human activity recognition; the study [7] trains k-Nearest Neighbour (kNN) classifiers for recognizing keystrokes; the Support Vector Machine (SVM) model can be built to perfectly classify the gestures [130]. Also, Convolutional Neural Network (CNN) classifiers can be leveraged for sign language recognition [106], and another work [133] exploits a three-layer Deep Neural Network (DNN) for user authentication. These ML approaches are widely used in past studies, providing high-accuracy performance. However, the applications of ML in wireless HPI inference present a lot of challenges.

Challenge 1: Ensuring Reliability of ML Approaches in HPI Inference. Despite the extensive use of ML techniques in wireless HPI inference and their high-accuracy performance, ensuring consistent reliability across diverse scenarios remains a challenge. Different activities and environments introduce significant

variations in signal patterns, which may cause a single ML model to underperform in certain situations. Ensemble learning methods offer a solution by combining predictions from multiple models, thereby capturing a wider spectrum of data patterns. For instance, WiARes [32] leverages ensemble learning, fusing predictions from a multiple layer perceptron (MLP), a random forest (RF), and a support vector machine (SVM) to enhance human activity recognition accuracy. Similarly, a recent study [174] presents an ensemble approach for cross-person activity recognition, demonstrating increased reliability and more robust predictions compared to standalone models.

Challenge 2: Integrating machine learning with multi-modal sensing. Integrating machine learning techniques with wireless sensing technologies offers vast opportunities for a range of applications, with most of these methods focused on a single type of wireless measurement. However, an approach that combines different types of signals (e.g., acoustic, CSI, mmWave, infrared, altrasound) can provide more comprehensive and diverse information for enhanced HPI inference [35, 100, 103]. For instance, [35] fuses ultrasound (which is immune to ambient noise and provides additional information about the speaker) with acoustic signals (which offer rich auditory data and are less susceptible to airflow) for speech enhancement. If machine learning networks can learn from the combination of these wireless measurements, it can result in more robust HPI inference techniques. Such techniques would be particularly beneficial in ubiquitous deployments, especially in future smart homes with many IoT devices.

Challenge 3: Achieving scalability and adaptability. Existing WiFi sensing techniques based on ML (deep learning) require a labor-intensive and time-consuming process of collecting training data or fingerprints. The training data need to be collected for each target subject or activity across diverse environ-

ments. While feasible for a subset of users and typical environments, it becomes impractical when expanding to new users or environments. These constraints limit the applicability of such techniques in larger, more complex settings. Therefore, there is an urgent need for innovative solutions that can reduce the extensive data collection requirement, enabling more scalable and flexible WiFi sensing applications. To address it, [188] applies transfer learning to effectively reuse knowledge across different sites and tasks. In addition, [150] utilizes domain adaptation, allowing the trained model to be applied to untrained domains (e.g., new cars, new drivers) for in-car activity recognition. Despite promising progress in these areas, several challenges, such as ensuring model robustness and reliability in new environments, still need to be addressed. Nonetheless, the potential of these techniques to significantly enhance the scalability and usability of deep learning models makes this an exciting area for future research.

Challenge 4: Mitigating security concerns in wireless ML. The usage of machine learning algorithms in the wireless domain also brings security concerns. Specifically, adversarial machine learning is receiving increasing attention nowadays, which can effectively disrupt wireless communications [2]. It studies vulnerabilities of machine learning approaches in adversarial settings and develops systems to make learning robust to adversarial manipulation [146]. An adversary can carefully design inputs, then feed them to machine learning models in the test or training phase to manipulate the behavior of a legitimate system by launching adversarial attacks [90]. For example, [134] trains a generative adversarial network (GAN) to spoof wireless signals. Hence, adversarial attacks and countermeasures should be considered when applying machine learning tools in achieving wireless HPI inference.

5.3 HPI Inference with mmWave

Millimeter wave (mmWave) communication has been witnessed as a promising technology for next-generation wireless systems. Millimeter wave frequencies range from 30 GHz to 300 GHz, which are much higher than those used by traditional wireless technologies (e.g., WiFi). As the wavelength of a signal is inversely proportional to its frequency, the wavelength at mmWave frequencies is much shorter than at lower frequencies. Thus, the size of the electronic components designed for transmitting and receiving these signals can be reduced [79], and it is possible to design smaller, more compact, and more portable mmWave-supported devices.

Nowadays, mmWave-supported devices are increasingly popular in everyday life. For example, 5G smartphones are equipped with mmWave technology, which allows them to connect to 5G networks and take advantage of the high speeds and low latency [74]; some wireless routers use mmWave technology to provide high-speed wireless Internet connectivity to devices at home or in an office [115]; autonomous driving systems consisting of mmWave radars provide high-resolution radar images for obstacle detection and avoidance [167]; mmWave frequencies are also used in medical imaging devices, such as CT scanners and MRI machines, to produce detailed images of the human body [6].

Besides, mmWave operates across a wide bandwidth, which results in greater sensing resolution. In detail, the resolution can be computed as $R = \frac{C}{2B}$, where C is the speed of light and B is the sweeping bandwidth. Thus, mmWave technology with a chirp bandwidth of a few GHz will have a range resolution in the order of centimeters (e.g., a chirp bandwidth of 4 GHz translates to a range resolution of 3.75 cm) [79].

The high resolution of mmWave enables it to sense minute human motion. Recent studies show that mmWave systems have improved performance compared with traditional wireless systems in terms of achieving various applications, such as user localization [117], vital signs monitoring [151, 180], activity recognition [82, 166], occupancy detection [128], user identification [80, 194], and speech acquisition [13, 75, 148]. Those studies provide the initial foray into HPI inference using mmWave, and we expect more such schemes will be designed targeting a broader category of HPI with the increased adoption of mmWave techniques.

Challenge 1: Extending the effective range of mmWave sensing. Indeed, millimeter-wave (mmWave) technology demonstrates significant potential for high-precision, non-intrusive HPI inference applications. However, there are inherent challenges that need to be addressed, including occlusion and signal attenuation. mmWave signals are highly susceptible to obstruction by obstacles and suffer from significant signal attenuation over long distances. It makes reliable HPI inference in diverse environments challenging. Because signal attenuation tends to increase with frequency, mmWave radar operating at a higher frequency may have a shorter effective range. To address this problem, an intuitive approach is to simply increase the transmitter power, but this solution is not energy-efficient and may pose additional issues such as interference with other systems and potential health concerns. Therefore, it is an important direction for the future to investigate intelligent reflecting surfaces (IRS) [19] and reconfigurable intelligent surfaces (RIS) [114], which are employed in the communications domain for signal propagation and beam steering for a larger coverage area.

Challenge 2: Designing advanced integrated circuits and systems. High carrier frequencies and bandwidths introduce design challenges for mmWave communication circuit components and antennas. The high transmit power and large

bandwidth can cause nonlinear distortion in power amplifiers. RF integrated circuits also face issues related with phase noise and IQ imbalance. On the other hand, the implementing mmWave technology requires high-frequency and high-speed components, demanding advanced system design and precise manufacturing techniques to produce these energy-efficient, compact, and cost-effective components [79].

Challenge 3: Adopting mmWave techniques in multimodal sensing. Another significant trend in the future is multimodal sensing. Combining mmWave sensing with other sensing modalities (e.g., acoustic, infrared) could enhance the accuracy and robustness of HPI inference systems. For example, the study [100] integrates mmWave and acoustic signals from a microphone, thereby facilitating a noise-resistant, long-distance speech recognition application. Similarly, the work [18] jointly analyzes mmWave and thermal camera signals, achieving privacy-preserving temperature screening and human tracking. These studies provide exciting opportunities for innovative interaction techniques, applications, and use cases.

In summary, wireless HPI inference is a promising field with significant challenges, including privacy and security concerns, robustness in various environments, and scalability for large-scale deployments. As for future trends, multimodal sensing is expected to gain prominence, as systems integrate different sensing modalities to gather comprehensive data for more accurate and detailed inference. The rise of edge computing and AI, coupled with the growth of the IoT ecosystem, paves the way for real-time data processing and broader integration of wireless HPI sensing. This expansion opens up potential applications in areas such as healthcare, retail, and smart homes. Considering these trends and challenges, wireless HPI sensing is a promising area for future research.

Chapter 6

Conclusion

This dissertation includes two studies about deception strategies in existing wireless networks and systems.

In the first work, wireless signal has demonstrated exceptional capability to detect breathing activity and estimate person count, which introduces a new threat to the security of personal information. To address this issue, we design an ambush-based strategy by actively deploying ambush locations and feeding eavesdroppers who move to those ambush locations with fake breathing rates or person count. This scheme enables the transmitter to encode the specified fake breathing rate or person count into CSI, and then utilize disturbance manipulation to deliver it to the eavesdropper. We conduct an extensive real-world evaluation on the USRP X310 platform. Experimental results in different scenarios consistently demonstrate the effectiveness of the proposed defenses.

In the second work, we have identified a new attack against liveness detection systems that use CSI to authenticate environmental human activities. Our *phantom-CSI* attack can manipulate CSI to exhibit the same semantic informa-

tion as that measured by a co-existing camera or microphone, allowing spoofed video or voice signals to bypass the CSI-based liveness detection system. Our attack implementation on USRPs running GNURadio validates the effectiveness and robustness of the proposed attack, with experimental results showing that the proposed attack drastically lowers the true positive rates (TPRs) of the wireless liveness detection system from 100% to just 4.4% and 0% for detecting spoofed video and voice, respectively.

Bibliography

[1] Heba Abdelnasser, Khaled A. Harras, and Moustafa Youssef. UbiBreathe: A Ubiquitous Non-Invasive WiFi-Based Breathing Estimator. In *Proc. of ACM International Symp. on Mobile Ad Hoc Networking and Computing (MobiHoc)*, pages 277–286, 2015.

[2] Damilola Adesina, Chung-Chu Hsieh, Yalin E. Sagduyu, and Lijun Qian. Adversarial Machine Learning in Wireless Communications using RF Data: A Review. *IEEE Communications Surveys & Tutorials*, 2022.

[3] Fadel Adib, Zach Kabelac, Dina Katabi, and Robert C. Miller. 3D Tracking via Body Radio Reflections. In *11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14)*, pages 317–329, Seattle, WA, 2014.

[4] Fadel Adib, Hongzi Mao, Zachary Kabelac, Dina Katabi, and Robert C. Miller. Smart Homes That Monitor Breathing and Heart Rate. In *Proc. of ACM Conference on Human Factors in Computing Systems (CHI)*, pages 837–846, 2015.

[5] Muhammad Ejaz Ahmed, Il-Youp Kwak, Jun Ho Huh, Iljoo Kim, Taekkyung Oh, and Hyoungshick Kim. Void: A Fast and Light Voice Liveness Detection System. In *29th USENIX Security Symposium (USENIX Security 20)*, pages 2685–2702, 2020.

[6] Sherif Sayed Ahmed, Andreas Schiessl, Frank Gumbmann, Marc Tiebout, Sebastian Methfessel, and Lorenz-Peter Schmidt. Advanced Microwave Imaging. *IEEE Microwave Magazine*, 13(6):26–43, 2012.

[7] Kamran Ali, Alex X Liu, Wei Wang, and Muhammad Shahzad. Keystroke Recognition Using WiFi Signals. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, pages 90–102, 2015.

[8] Kamran Ali, Alex X Liu, Wei Wang, and Muhammad Shahzad. Recognizing Keystrokes Using WiFi Devices. *IEEE Journal on Selected Areas in Communications*, 35(5):1175–1190, 2017.

[9] Narendra Anand, Sung-Ju Lee, and Edward W. Knightly. STROBE: Actively Securing Wireless Communications Using Zero-Forcing Beamforming. In *2012 Proc. IEEE INFOCOM*, pages 720–728, 2012.

[10] Laura Anitori, Ardjan de Jong, and Frans Nennie. FMCW Radar for Life-sign Detection. In *2009 IEEE Radar Conference*, pages 1–6, 2009.

[11] Marcello Ascione, Aniello Buonanno, Michele D'Urso, Leopoldo Angrisani, and Rosario Schiano Lo Moriello. A New Measurement Method Based on Music Algorithm for Through-the-Wall Detection of Life Signs. *IEEE Transactions on Instrumentation and Measurement*, 62(1):13–26, 2013.

[12] Zach Banks and Eric Van Albert. Looping Surveillance Cameras through Live Editing of Network Streams. https://infocondb.org/con/def-con/def-con-23/looping-surveillance-cameras-through-live-editing-of-network-streams, 2015.

[13] Suryoday Basak and Mahanth Gowda. mmspy: Spying Phone Calls using mmWave Radars. In *IEEE Symposium on Security and Privacy (SP)*, pages 1211–1228, 2022.

[14] Clark BJ and 3rd. Treatment of Heart Failure in Infants and Children. *Heart Disease*, 2(5):354–361, 2000.

[15] Logan Blue, Luis Vargas, and Patrick Traynor. Hello, Is It Me You're Looking for? Differentiating Between Human and Electronic Speakers for Voice Interface Security. In *Proceedings of the 11th ACM Conference on Security & Privacy in Wireless and Mobile Networks*, pages 123–133, 2018.

[16] Steven Boll. Suppression of Acoustic Noise in Speech Using Spectral Subtraction. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 27(2):113–120, 1979.

[17] Olga Boric-Lubecke, Victor M Lubecke, Amy D Droitcour, Byung-Kwon Park, and Aditya Singh. *Doppler Radar Physiological Sensing*. Wiley Online Library, 2016.

[18] Marco Canil, Jacopo Pegoraro, and Michele Rossi. milliTRACE-IR: Contact Tracing and Temperature Screening via mmWave and Infrared Sensing. *IEEE Journal of Selected Topics in Signal Processing*, 16(2):208–223, 2022.

[19] Yashuai Cao, Tiejun Lv, and Wei Ni. Intelligent Reflecting Surface Aided Multi-User mmWave Communications for Coverage Enhancement. In *2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, pages 1–6, 2020.

[20] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. OpenPose: Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(1):172–186, 2019.

[21] Carl Willis. Vacant Homes Becoming Latest Targets for Burglars, 2017. https://www.wsbtv.com/news/local/vacant-homes-becoming-latest-targets-for-burglars/579935438/.

[22] Edgar Cervantes. The Best Siri Commands for Productivity, Information, Laughter, and More. https://www.androidauthority.com/best-siri-commands-1094484/, 2021.

[23] Anadi Chaman, Jiaming Wang, Jiachen Sun, Haitham Hassanieh, and Romit Roy Choudhury. Ghostbuster: Detecting the Presence of Hidden Eavesdroppers. In *Proc. of the 24th Annual International Conference on Mobile Computing and Networking*, MobiCom '18, pages 337–351, New York, NY, USA, 2018.

[24] Lili Chen, Jie Xiong, Xiaojiang Chen, Sunghoon Ivan Lee, Daqing Zhang, Tao Yan, and Dingyi Fang. LungTrack: Towards Contactless and Zero Dead-Zone Respiration Monitoring with Commodity RFIDs. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 3(3):79:1–79:22, 2019.

[25] Mingzhe Chen, Ursula Challita, Walid Saad, Changchuan Yin, and Mérouane Debbah. Artificial Neural Networks-Based Machine Learning for Wireless Networks: A Tutorial. *IEEE Communications Surveys & Tutorials*, 21(4):3039–3071, 2019.

[26] Shaxun Chen, Amit Pande, and Prasant Mohapatra. Sensor-assisted Facial Recognition: An Enhanced Biometric Authentication System for Smartphones. In *Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services*, pages 109–122, 2014.

[27] Hyuckjin Choi, Manato Fujimoto, Tomokazu Matsui, Shinya Misaki, and Keiichi Yasumoto. Wi-Cal: WiFi Sensing and Machine Learning Based Device-Free Crowd Counting and Localization. *IEEE Access*, 10:24395–24410, 2022.

[28] CODi. FALCO HD 1080P Auto Focus Webcam. https://www.codiworldwide.com/mobile-accessories/falco-hd-1080p-webcam/, 2021.

[29] Congressional Research Service. The Internet of Things (IoT): An Overview. https://crsreports.congress.gov/product/pdf/IF/IF11239, 2020.

[30] Rafael A. Cox and Carlos Zayas Torres. Acute Heart Failure in Adults. *Puerto Rico Health Sciences Journal*, 23(4):265–271, 2004.

[31] Riccardo Crepaldi, Jeongkeun Lee, Raul Etkin, Sung-Ju Lee, and Robin Kravets. CSI-SF: Estimating Wireless Channel State Using CSI Sampling Fusion. In *2012 Proc. IEEE INFOCOM*, pages 154–162, 2012.

[32] Wei Cui, Bing Li, Le Zhang, and Zhenghua Chen. Device-free Single-user Activity Recognition Using Diversified Deep Ensemble Learning. *Applied Soft Computing*, 102:107066, 2021.

[33] Laurie Davies and Ursula Gather. The Identification of Multiple Outliers. *Journal of the American Statistical Association*, 88(423):782–792, 1993.

[34] Wenrui Diao, Xiangyu Liu, Zhe Zhou, and Kehuan Zhang. Your Voice Assistant Is Mine: How to Abuse Speakers to Steal Information and Control Your Phone. In *Proceedings of the 4th ACM Workshop on Security and Privacy in Smartphones & Mobile Devices*, pages 63–74, 2014.

[35] Han Ding, Yizhan Wang, Hao Li, Cui Zhao, Ge Wang, Wei Xi, and Jizhong Zhao. UltraSpeech: Speech Enhancement by Interaction between Ultrasound and Speech. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 6(3), 2022.

[36] Amy D. Droitcour, Olga Boric-Lubecke, and Gregory T. A. Kovacs. Signal-to-Noise Ratio in Doppler Radar System for Heart and Respiratory Rate Measurements. *IEEE Transactions on Microwave Theory and Techniques*, 57(10):2498–2507, 2009.

[37] Wayne Duggan. CSX Stock Plummets On CEO's Health Concerns. *US News*, 2017. https://money.usnews.com/investing/stock-market-news/articles/2017-12-15/csx-corporation-stock-plummets-on-ceos-health-concerns.

[38] Mohammed S. Elbamby, Cristina Perfecto, Mehdi Bennis, and Klaus Doppler. Edge Computing Meets Millimeter-wave Enabled VR: Paving the Way to Cutting the Cord. In *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6, 2018.

[39] Holly Ellyatt. How CEO health can affect your wealth. *CNBC*, 2012. https://www.cnbc.com/id/49115208.

[40] Matt Ettus. Usrp Users and Developers Guide. www.olifantasia.com/gnuradio/usrp/files/usrp_guide.pdf, 2005.

[41] Ettus Research. USRP X300. https://www.ettus.com/all-products/x300-kit/, 2021.

[42] Ettus Research. SBX 400-4400 MHz RX/TX, 2022. https://www.ettus.com/all-products/sbx120/.

[43] Ettus Research. USRP X310, 2022. https://www.ettus.com/all-products/x310-kit/.

[44] Dou Fan, Aifeng Ren, Nan Zhao, Xiaodong Yang, Zhiya Zhang, Syed A. Shah, Fangming Hu, and Qammer H. Abbasi. Breathing Rhythm Analysis in Body Centric Networks. *IEEE Access*, 6:32507–32513, 2018.

[45] S. Fang, Y. Liu, and P. Ning. Mimicry Attacks Against Wireless Link Signature and New Defense Using Time-Synched Link Signature. *IEEE Transactions on Information Forensics and Security*, 11(7):1515–1527, 2016.

[46] S. Fang, Y. Liu, and P. Ning. Wireless Communications under Broadband Reactive Jamming Attacks. *IEEE Transactions on Dependable and Secure Computing*, 13(3):394–408, 2016.

[47] S. Fang, Y. Liu, W. Shen, H. Zhu, and T. Wang. Virtual Multipath Attack and Defense for Location Distinction in Wireless Networks. *IEEE Transactions on Mobile Computing*, 16(2):566–580, 2017.

[48] S. Fang, I. Markwood, and Y. Liu. Manipulatable Wireless Key Establishment. In *2017 IEEE Conference on Communications and Network Security (CNS)*, pages 1–9, 2017.

[49] Song Fang, Yao Liu, Wenbo Shen, and Haojin Zhu. Where Are You from? Confusing Location Distinction Using Virtual Multipath Camouflage. In *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*, MobiCom '14, pages 225–236, Maui, Hawaii, USA, 2014.

[50] Song Fang, Ian Markwood, Yao Liu, Shangqing Zhao, Zhuo Lu, and Haojin Zhu. No Training Hurdles: Fast Training-Agnostic Attacks to Infer Your Typing. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, CCS '18, pages 1747–1760, New York, NY, USA, 2018.

[51] Song Fang, Tao Wang, Yao Liu, Shangqing Zhao, and Zhuo Lu. Entrapment for Wireless Eavesdroppers. In *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, pages 2530–2538, 2019.

[52] Federal Communications Commission and others. Revision of Part 15 of the Commission's Rules Regarding Ultra-wideband Transmission Systems. *First Report and Order, FCC 02-48*, 2002.

[53] Atena Roshan Fekr, Majid Janidarmian, Katarzyna Radecka, and Zeljko Zilic. Respiration Disorders Classification With Informative Features for m-Health Applications. *IEEE Journal of Biomedical and Health Informatics*, 20(3):733–747, 2016.

[54] Huan Feng, Kassem Fawaz, and Kang G. Shin. Continuous Authentication for Voice Assistants. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, MobiCom '17, pages 343–355, New York, NY, USA, 2017.

[55] Ming Gao, Feng Lin, Weiye Xu, Muertikepu Nuermaimaiti, Jinsong Han, Wenyao Xu, and Kui Ren. Deaf-Aid: Mobile IoT Communication Exploiting Stealthy Speaker-to-Gyroscope Channel. In *Proc. of the 26th Annual International Conference on Mobile Computing and Networking*, MobiCom '20, 2020.

[56] Daniel Genkin, Lev Pachmanov, Itamar Pipman, and Eran Tromer. Stealing Keys from PCs Using a Radio: Cheap Electromagnetic Attacks on Windowed Exponentiation. In Tim Güneysu and Helena Handschuh, editors, *Cryptographic Hardware and Embedded Systems - CHES 2015*, pages 207–228, Berlin, Heidelberg, 2015.

[57] GNU Radio project. GNU Radio - The Free & Open Source Radio Ecosystem, 2022. https://www.gnuradio.org.

[58] Andrea Goldsmith. *Wireless Communications*. New York, NY, USA, 2005.

[59] Shyamnath Gollakota and Dina Katabi. Physical Layer Wireless Security Made Fast and Channel Independent. In *2011 Proc. IEEE INFOCOM*, pages 1125–1133, 2011.

[60] Francesco Gringoli, Matthias Schulz, Jakob Link, and Matthias Hollick. Free Your CSI: A Channel State Information Extraction Platform for Modern Wi-Fi Chipsets. In *Proceedings of the 13th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization*, pages 21–28, 2019.

[61] Changzhan Gu. Short-range Noncontact Sensors for Healthcare and Other Emerging Applications: A Review. *Sensors*, 16(8):1169, 2016.

[62] Xiaonan Guo, Bo Liu, Cong Shi, Hongbo Liu, Yingying Chen, and Mooi Choo Chuah. WiFi-Enabled Smart Human Dynamics Monitoring. In *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*, SenSys '17, New York, NY, USA, 2017.

[63] Jun Han, Albert Jin Chung, Manal Kumar Sinha, Madhumitha Harishankar, Shijia Pan, Hae Young Noh, Pei Zhang, and Patrick Tague. Do You Feel What I Hear? Enabling Autonomous IoT Device Pairing Using Different Sensor Types. In *IEEE Symposium on Security and Privacy (SP)*, pages 836–852, 2018.

[64] Asif Hanif, Mazher Iqbal, and Farasat Munir. WiSpy: Through-Wall Movement Sensing and Person Counting Using Commodity WiFi Signals. In *2018 IEEE Sensors*, pages 1–4, 2018.

[65] Mohammad Hasan. State of IoT 2022: Number of Connected IoT Devices Growing 18% to 14.4 Billion Globally. https://iot-analytics.com/number-connected-iot-devices/#:~:text=The%20forecast%20for%20the%20total, impacted%20both%20demand%20and%20supply., 2022.

[66] Qiuye He and Song Fang. Phantom-CSI Attacks against Wireless Liveness Detection. In *Proceedings of the 26th International Symposium on Research in Attacks, Intrusions and Defenses*, RAID '23, pages 440–454, New York, NY, USA, 2023.

[67] Qiuye He, Song Fang, Tao Wang, Yao Liu, Shangqing Zhao, and Zhuo Lu. Proactive Anti-Eavesdropping With Trap Deployment in Wireless Networks. *IEEE Transactions on Dependable and Secure Computing*, 20(1):637–649, 2023.

[68] Qiuye He, Edwin Yang, Song Fang, and Shangqing Zhao. HoneyBreath: An Ambush Tactic Against Wireless Breath Inference. In *Mobile and Ubiquitous Systems: Computing, Networking and Service*, pages 203–226, Cham, 2023.

[69] Y. He, Q. He, S. Fang, and Y. Liu. Precise Wireless Camera Localization Leveraging Traffic-aided Spatial Analysis. *IEEE Transactions on Mobile Computing*, (01):1–13, 2023.

[70] Yan He, Qiuye He, Song Fang, and Yao Liu. MotionCompass: Pinpointing Wireless Camera via Motion-Activated Traffic. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys)*, pages 215–227, New York, NY, USA, 2021.

[71] Yan He, Qiuye He, Song Fang, and Yao Liu. When Free Tier Becomes Free to Enter: A Non-Intrusive Way to Identify Security Cameras with no Cloud Subscription. In *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security*, CCS '23, pages 651–665, New York, NY, USA, 2023.

[72] Craig Heffners. Exploiting Network Surveillance Cameras Like a Hollywood Hacker. https://www.youtube.com/watch?v=B8DjTcANBx0, 2013.

[73] Peter Hillyard, Anh Luong, Alemayehu Solomon Abrar, Neal Patwari, Krishna Sundar, Robert Farney, Jason Burch, Christina Porucznik, and Sarah Hatch Pollard. Experience: Cross-Technology Radio Respiratory Monitoring Performance Study. In *Proc. of the 24th Annual International Conference on Mobile Computing and Networking*, MobiCom '18, pages 487–496, New York, NY, USA, 2018.

[74] Wonbin Hong, Kwang-Hyun Baek, and Seungtae Ko. Millimeter-wave 5G Antennas for Smartphones: Overview and Experimental Demonstration. *IEEE Transactions on Antennas and Propagation*, 65(12):6250–6261, 2017.

[75] Pengfei Hu, Wenhao Li, Riccardo Spolaor, and Xiuzhen Cheng. mmEcho: A mmWave-based Acoustic Eavesdropping Method. In *IEEE Symposium on Security and Privacy (SP)*, pages 836–852, 2022.

[76] Chenyu Huang, Huangxun Chen, Lin Yang, and Qian Zhang. BreathLive: Liveness Detection for Heart Sound Authentication with Deep Breathing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(1):1–25, 2018.

[77] Yong Huang, Xiang Li, Wei Wang, Tao Jiang, and Qian Zhang. Forgery Attack Detection in Surveillance Video Streams Using Wi-Fi Channel State Information. *IEEE Transactions on Wireless Communications*, 21(6):4340–4349, 2021.

[78] Yong Huang, Xiang Li, Wei Wang, Tao Jiang, and Qian Zhang. Towards Cross-Modal Forgery Detection and Localization on Live Surveillance Videos. In *Proceedings of the IEEE International Conference on Computer Communications*, INFOCOM '21, 2021.

[79] Cesar Iovescu and Sandeep Rao. The Fundamentals of Millimeter Wave Radar Sensors, 2020. https://www.ti.com/lit/wp/spyy005a/spyy005a.pdf?ts=1673037835044&ref_url=https%253A%252F%252Fwww.google.com%252F.

[80] Prabhu Janakaraj, Kalvik Jakkala, Arupjyoti Bhuyan, Zhi Sun, Pu Wang, and Minwoo Lee. STAR: Simultaneous Tracking and Recognition through Millimeter Waves and Deep Learning. In *2019 12th IFIP Wireless and Mobile Networking Conference (WMNC)*, pages 211–218, 2019.

[81] Weijia Jia, Hongjian Peng, Na Ruan, Zhiqing Tang, and Wei Zhao. WiFind: Driver Fatigue Detection with Fine-Grained Wi-Fi Signal Features. *IEEE Transactions on Big Data*, 6(2):269–282, 2020.

[82] Wenjun Jiang, Chenglin Miao, Fenglong Ma, Shuochao Yao, Yaqing Wang, Ye Yuan, Hongfei Xue, Chen Song, Xin Ma, Dimitrios Koutsonikolas, et al. Towards Environment Independent Device Free Human Activity Recognition. In *Proc. of the 24th Annual International Conference on Mobile Computing and Networking*, pages 289–304, 2018.

[83] Jesse S Jin, Changsheng Xu, Min Xu, Dai-Kyung Hyun, Min-Jeong Lee, Seung-Jin Ryu, Hae-Yeoun Lee, and Heung-Kyu Lee. Forgery Detection for Surveillance Video. In *The Era of Interactive Media*, pages 25–36, 2013.

[84] Arun Cyril Jose and Reza Malekian. Improving Smart Home Security: Integrating Logical Sensing Into Smart Home. *IEEE Sensors Journal*, 17(13):4269–4286, 2017.

[85] Naor Kalbo, Yisroel Mirsky, Asaf Shabtai, and Yuval Elovici. The Security of IP-based Video Surveillance Systems. *Sensors*, 20(17):4806, 2020.

[86] Ossi Kaltiokallio, Hüseyin Yiğitler, Riku Jäntti, and Neal Patwari. Non-invasive Respiration Rate Monitoring Using a Single COTS TX-RX Pair. In *Proc. of the 13th International Symp. on Information Processing in Sensor Networks (IPSN)*, pages 59–69, 2014.

[87] Danista Khan and Ivan Wang-Hei Ho. CrossCount: Efficient Device-Free Crowd Counting by Leveraging Transfer Learning. *IEEE Internet of Things Journal*, 10(5):4049–4058, 2023.

[88] Tomi Kinnunen, Md Sahidullah, Héctor Delgado, Massimiliano Todisco, Nicholas Evans, Junichi Yamagishi, and Kong Aik Lee. The ASVspoof 2017 Challenge: Assessing the Limits of Replay Spoofing Attack Detection. 2017.

[89] Katharina Krombholz, Heidelinde Hobel, Markus Huber, and Edgar Weippl. Advanced social engineering attacks. *Journal of Information Security and Applications*, 22:113 – 122, 2015.

[90] Alexey Kurakin, Ian J. Goodfellow, and Samy Bengio. Adversarial Machine Learning at Scale. In *International Conference on Learning Representations*, 2017.

[91] Deborah Beranek Lafky and Thomas A. Horan. Personal Health Records: Consumer Attitudes Toward Privacy and Security of Their Personal Health Information. *Health Informatics Journal*, 17(1):63–71, 2011.

[92] Nitya Lakshmanan, Inkyu Bang, Min Suk Kang, Jun Han, and Jong Taek Lee. SurFi: Detecting Surveillance Camera Looping Attacks with Wi-Fi Channel State Information. In *Proceedings of the 12th Conference on Security and Privacy in Wireless and Mobile Networks*, WiSec '19, 2019.

[93] Lutz Lampe and Klaus Witrisal. Challenges and Recent Advances in IR-UWB System Design. In *Proc. of 2010 IEEE International Symposium on Circuits and Systems*, pages 3288–3291, 2010.

[94] Changzhi Li, Jun Ling, Jian Li, and Jenshan Lin. Accurate Doppler Radar Noncontact Vital Sign Detection Using the RELAX Algorithm. *IEEE Transactions on Instrumentation and Measurement*, 59(3):687–695, 2010.

[95] Changzhi Li, Victor M. Lubecke, Olga Boric-Lubecke, and Jenshan Lin. A Review on Recent Advances in Doppler Radar Sensors for Noncontact Healthcare Monitoring. *IEEE Transactions on Microwave Theory and Techniques*, 61(5):2046–2060, 2013.

[96] Xiaopeng Li, Fengyao Yan, Fei Zuo, Qiang Zeng, and Lannan Luo. Touch Well Before Use: Intuitive and Secure Authentication for IoT Devices. In *The 25th Annual International Conference on Mobile Computing and Networking*, MobiCom '19, 2019.

[97] J. Liu, H. Liu, Y. Chen, Y. Wang, and C. Wang. Wireless Sensing for Human Activity: A Survey. *IEEE Communications Surveys & Tutorials*, 22(3):1629–1645, 2020.

[98] Jian Liu, Yan Wang, Yingying Chen, Jie Yang, Xu Chen, and Jerry Cheng. Tracking Vital Signs During Sleep Leveraging Off-the-shelf WiFi. In *Proc. of ACM International Symp. on Mobile Ad Hoc Networking and Computing (MobiHoc)*, pages 267–276, 2015.

[99] Shangqing Liu, Yanchao Zhao, and Bing Chen. WiCount: A Deep Learning Approach for Crowd Counting Using WiFi Signals. In *2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC)*, pages 967–974, 2017.

[100] Tiantian Liu, Ming Gao, Feng Lin, Chao Wang, Zhongjie Ba, Jinsong Han, Wenyao Xu, and Kui Ren. Wavoice: A Noise-Resistant Multi-Modal Speech Recognition System Fusing mmWave and Audio Signals. In *Proc. of the 19th ACM Conference on Embedded Networked Sensor Systems*, SenSys '21, pages 97–110, 2021.

[101] Xuefeng Liu, Jiannong Cao, Shaojie Tang, and Jiaqi Wen. Wi-Sleep: Contactless Sleep Monitoring via WiFi Signals. In *2014 IEEE Real-Time Systems Symposium*, pages 346–355, 2014.

[102] Xuefeng Liu, Jiannong Cao, Shaojie Tang, Jiaqi Wen, and Peng Guo. Contactless Respiration Monitoring Via Off-the-Shelf WiFi Devices. *IEEE Transactions on Mobile Computing*, 15(10):2466–2479, 2016.

[103] Xinxin Lu, Lei Wang, Chi Lin, Xin Fan, Bin Han, Xin Han, and Zhenquan Qin. AutoDLAR: A Semi-Supervised Cross-Modal Contact-Free Human Activity Recognition System. *ACM Trans. Sen. Netw.*, 2023.

[104] Changqing Luo, Jinlong Ji, Qianlong Wang, Xuhui Chen, and Pan Li. Channel State Information Prediction for 5G Wireless Communications: A Deep Learning Approach. *IEEE Transactions on Network Science and Engineering*, 7(1):227–236, 2020.

[105] Yongsen Ma, Gang Zhou, and Shuangquan Wang. WiFi Sensing with Channel State Information: A Survey. *ACM Comput. Surv.*, 52(3), 2019.

[106] Yongsen Ma, Gang Zhou, Shuangquan Wang, Hongyang Zhao, and Woosub Jung. SignFi: Sign Language Recognition Using WiFi. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(1), 2018.

[107] Masimo. MightySat Fingertip Pulse Oximeter with Bluetooth LE, RRp, & PVi, 2021. https://www.masimopersonalhealth.com/products/mightysat-fingertip-pulse-oximeter-with-bluetooth-le-rrp-pvi.

[108] Yan Meng, Jiachun Li, Matthew Pillari, Arjun Deopujari, Liam Brennan, Hafsah Shamsie, Haojin Zhu, and Yuan Tian. Your Microphone Array Retains Your Identity: A Robust Voice Liveness Detection System for Smart Speaker. In *USENIX Security*, 2022.

[109] Yan Meng, Zichang Wang, Wei Zhang, Peilin Wu, Haojin Zhu, Xiaohui Liang, and Yao Liu. WiVo: Enhancing the Security of Voice Control System via Wireless Signal in IoT Environment. Mobihoc '18, pages 81–90, New York, NY, USA, 2018.

[110] Yan Meng, Haojin Zhu, Jinlei Li, Jin Li, and Yao Liu. Liveness Detection for Voice User Interface via Wireless Signals in IoT Environment. *IEEE Transactions on Dependable and Secure Computing*, 2020.

[111] Dibya Mukhopadhyay, Maliheh Shirvanian, and Nitesh Saxena. All Your Voices are Belong to Us: Stealing Voices to Fool Humans and Machines. In *European Symposium on Research in Computer Security*, pages 599–621, 2015.

[112] Eduardo F. Nakamura and Antonio A. F. Loureiro. Information Fusion in Wireless Sensor Networks. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, SIGMOD '08, pages 1365–1372, New York, NY, USA, 2008.

[113] Masayuki Nakatsuka, H Iwatani, and Jiro Katto. A Study on Passive Crowd Density Estimation Using Wireless Sensors. In *The 4th Intl. Conf. on Mobile Computing and Ubiquitous Networking (ICMU 2008)*, 2008.

[114] Mahyar Nemati, Jihong Park, and Jinho Choi. RIS-Assisted Coverage Enhancement in Millimeter-Wave Cellular Networks. *IEEE Access*, 8:188171–188185, 2020.

[115] novotech. 5G Routers, Gateways and Antennas (updated 2021), 2022. https://novotech.com/5g-routers-gateways-antennas/.

[116] Muhammed Zahid Ozturk, Chenshu Wu, Beibei Wang, and KJ Liu. RadioMic: Sound Sensing via mmWave Signals. *arXiv preprint arXiv:2108.03164*, 2021.

[117] Joan Palacios, Guillermo Bielsa, Paolo Casari, and Joerg Widmer. Communication-driven Localization and Mapping for Millimeter Wave Networks. In *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*, pages 2402–2410, 2018.

[118] Sameera Palipana, David Rojas, Piyush Agrawal, and Dirk Pesch. FallDeFi: Ubiquitous Fall Detection Using Commodity Wi-Fi Devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(4), 2018.

[119] Neal Patwari, Lara Brewer, Quinn Tate, Ossi Kaltiokallio, and Maurizio Bocca. Breathfinding: A Wireless Network That Monitors and Locates Breathing in a Home. *IEEE Journal of Selected Topics in Signal Processing*, 8(1):30–42, 2014.

[120] Neal Patwari, Joey Wilson, Sai Ananthanarayanan, Sneha K. Kasera, and Dwayne R. Westenskow. Monitoring Breathing via Signal Strength in Wireless Networks. *IEEE Transactions on Mobile Computing*, 13(8):1774–1786, 2014.

[121] R.K. Pearson. Outliers in Process Modeling and Identification. *IEEE Transactions on Control Systems Technology*, 10(1):55–63, 2002.

[122] Timothy J. Pierson, Travis Peters, Ronald Peterson, and David Kotz. Proximity Detection with Single-Antenna IoT Devices. In *The 25th Annual International Conference on Mobile Computing and Networking*, MobiCom '19, 2019.

[123] Swadhin Pradhan, Wei Sun, Ghufran Baig, and Lili Qiu. Combating Replay Attacks Against Voice Assistants. *Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(3):1–26, 2019.

[124] Kun Qian, Chenshu Wu, Zheng Yang, Yunhao Liu, Fugui He, and Tianzhang Xing. Enabling Contactless Detection of Moving Humans with Dynamic Speeds Using CSI. *ACM Transactions on Embedded Computing Systems (TECS)*, 17(2):1–18, 2018.

[125] Tauhidur Rahman, Alexander T. Adams, Ruth Vinisha Ravichandran, Mi Zhang, Shwetak N. Patel, Julie A. Kientz, and Tanzeem Choudhury. DoppleSleep: A Contactless Unobtrusive Sleep Sensing System Using Short-Range Doppler Radar. In *Proc. of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '15, pages 39–50, New York, NY, USA, 2015.

[126] Muhammad Salman, Nguyen Dao, Uichin Lee, and Youngtae Noh. CSI:DeSpy: Enabling Effortless Spy Camera Detection via Passive Sensing of User Activities and Bitrate Variations. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 6(2), 2022.

[127] Jussi Salmi and Andreas F. Molisch. Propagation Parameter Estimation, Modeling and Measurements for Ultrawideband MIMO Radar. *IEEE Transactions on Antennas and Propagation*, 59(11):4257–4267, 2011.

[128] Avik Santra, Raghavendran Vagarappan Ulaganathan, and Thomas Finke. Short-Range Millimetric-Wave Radar System for Occupancy Sensing Application. *IEEE Sensors Letters*, 2(3):1–4, 2018.

[129] Souvik Sen, Božidar Radunovic, Romit Roy Choudhury, and Tom Minka. You Are Facing the Mona Lisa: Spot Localization Using PHY Layer Information. In *Proc. of the 10th International Conference on Mobile Systems, Applications, and Services*, MobiSys '12, pages 183–196, New York, NY, USA, 2012.

[130] Jiacheng Shang and Jie Wu. A Robust Sign Language Recognition System with Multiple Wi-Fi Devices. In *Proc. of the Workshop on Mobility in the Evolving Internet Architecture*, MobiArch '17, pages 19–24, 2017.

[131] Jiacheng Shang and Jie Wu. Voice Liveness Detection for Voice Assistants through Ear Canal Pressure Monitoring. *IEEE Transactions on Network Science and Engineering*, 2022.

[132] Wenbo Shen, Peng Ning, Xiaofan He, and Huaiyu Dai. Ally Friendly Jamming: How to Jam Your Enemy and Maintain Your Own Wireless Connectivity at the Same Time. In *2013 IEEE Symposium on Security and Privacy*, pages 174–188, 2013.

[133] Cong Shi, Jian Liu, Hongbo Liu, and Yingying Chen. 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*, pages 1–10, 2017.

[134] Yi Shi, Kemal Davaslioglu, and Yalin E. Sagduyu. Generative Adversarial Network for Wireless Signal Spoofing. In *Proc. of the ACM Workshop on Wireless Security and Machine Learning*, WiseML 2019, pages 55–60, 2019.

[135] Yi-Sheng Shiu, Shih Yu Chang, Hsiao-Chun Wu, Scott C.-H. Huang, and Hsiao-Hwa Chen. Physical Layer Security in Wireless Networks: A Tutorial. *IEEE Wireless Communications*, 18(2):66–74, 2011.

[136] Jonathon Shlens. A Tutorial on Principal Component Analysis. *arXiv preprint arXiv:1404.1100*, 2014.

[137] Steven W Smith. *The Scientist and Engineer's Guide to Digital Signal Processing, Second Edition*. California Technical Pub. San Diego, 1999.

[138] Lin Sun, WaiBin Huang, and MingHui Wu. TIR/VIS Correlation for Liveness Detection in Face Recognition. In *International Conference on Computer Analysis of Images and Patterns*, pages 114–121, 2011.

[139] Benjamin Tag, Junichi Shimizu, Chi Zhang, Kai Kunze, Naohisa Ohta, and Kazunori Sugiura. In the Eye of the Beholder: The Impact of Frame Rate on Human Eye Blink. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, pages 2321–2327, New York, NY, USA, 2016.

[140] Nils Ole Tippenhauer, Luka Malisa, Aanjhan Ranganathan, and Srdjan Capkun. On Limitations of Friendly Jamming for Confidentiality. In *Proc. of the 2013 IEEE Symposium on Security and Privacy*, SP '13, pages 160–173, USA, 2013.

[141] Bang Tran, Shenhui Pan, Xiaohui Liang, and Honggang Zhang. Exploiting Physical Presence Sensing to Secure Voice Assistant Systems. In *ICC 2021 - IEEE International Conference on Communications*, pages 1–6, 2021.

[142] K Van Loon, MJM Breteler, L Van Wolfwinkel, AT Rheineck Leyssius, S Kossen, CJ Kalkman, B van Zaane, and LM Peelen. Wireless Non-invasive Continuous Respiratory Monitoring with FMCW Radar: A Clinical Validation Study. *Journal of Clinical Monitoring and Computing*, 30(6):797–805, 2016.

[143] Swaroop Venkatesh, Christopher R. Anderson, Natalia V. Rivera, and R. Michael Buehrer. Implementation and Analysis of Respiration-rate Estimation Using Impulse-based UWB. In *2005 IEEE Military Communications Conference (MILCOM)*, pages 3314–3320 Vol. 5, 2005.

[144] Jesus Villalba and Eduardo Lleida. Preventing Replay Attacks on Speaker Verification Systems. In *2011 Carnahan Conference on Security Technology*, pages 1–8, 2011.

[145] Aditya Virmani and Muhammad Shahzad. Position and Orientation Agnostic Gesture Recognition Using WiFi. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, pages 252–264, 2017.

[146] Yevgeniy Vorobeychik and Murat Kantarcioglu. *Adversarial Machine Learning*. Morgan & Claypool Publishers, 2018.

[147] Steven A Wahls. Causes and Evaluation of Chronic Dyspnea. *American family physician*, 86 2:173–82, 2012.

[148] Chao Wang, Feng Lin, Zhongjie Ba, Fan Zhang, Wenyao Xu, and Kui Ren. Wavesdropper: Through-wall Word Detection of Human Speech via Commercial mmWave Devices. *Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(2):1–26, 2022.

[149] Chen Wang, Jian Liu, Yingying Chen, Hongbo Liu, and Yan Wang. Towards In-baggage Suspicious Object Detection Using Commodity WiFi. In *2018 IEEE Conference on Communications and Network Security (CNS)*, pages 1–9, 2018.

[150] Fangxin Wang, Jiangchuan Liu, and Wei Gong. WiCAR: Wifi-Based in-Car Activity Recognition with Multi-Adversarial Domain Adaptation. In *Proc. of the International Symposium on Quality of Service*, IWQoS '19, 2019.

[151] Fengyu Wang, Xiaolu Zeng, Chenshu Wu, Beibei Wang, and KJ Ray Liu. Driver Vital Signs Monitoring Using Millimeter Wave Radio. *IEEE Internet of Things Journal*, 2021.

[152] Fengyu Wang, Feng Zhang, Chenshu Wu, Beibei Wang, and K. J. Ray Liu. Respiration Tracking for People Counting and Recognition. *IEEE Internet of Things Journal*, 7(6):5233–5245, 2020.

[153] Guanhua Wang, Yongpan Zou, Zimu Zhou, Kaishun Wu, and Lionel M. Ni. We Can Hear You with Wi-Fi! In *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*, MobiCom '14, pages 593–604, New York, NY, USA, 2014.

[154] Hao Wang, Daqing Zhang, Junyi Ma, Yasha Wang, Yuxiang Wang, Dan Wu, Tao Gu, and Bing Xie. Human Respiration Detection with Commodity WiFi Devices: Do User Location and Body Orientation Matter? In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '16, page 25â36, New York, NY, USA, 2016.

[155] Qian Wang, Xiu Lin, Man Zhou, Yanjiao Chen, Cong Wang, Qi Li, and Xiangyang Luo. VoicePop: A Pop Noise based Anti-spoofing System for Voice Authentication on Smartphones. In *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, pages 2062–2070, 2019.

[156] Shu Wang, Jiahao Cao, Xu He, Kun Sun, and Qi Li. When the Differences in Frequency Domain Are Compensated: Understanding and Defeating Modulated Replay Attacks on Automatic Speech Recognition. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, pages 1103–1119, 2020.

[157] T. Wang, Y. Liu, T. Hou, Q. Pei, and S. Fang. Signal Entanglement Based Pinpoint Waveforming for Location-Restricted Service Access Control. *IEEE Transactions on Dependable and Secure Computing*, 15(5):853–867, 2018.

[158] Wei Wang, Alex X. Liu, and Muhammad Shahzad. Gait Recognition Using Wifi Signals. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '16, pages 363–373, New York, NY, USA, 2016.

[159] Wei Wang, Alex X. Liu, Muhammad Shahzad, Kang Ling, and Sanglu Lu. Understanding and Modeling of WiFi Signal Based Human Activity Recognition. In *Proc. of the 21st Annual International Conference on Mobile Computing and Networking*, MobiCom '15, pages 65–76, 2015.

[160] Xuanzhi Wang, Kai Niu, Jie Xiong, Bochong Qian, Zhiyun Yao, Tairong Lou, and Daqing Zhang. Placement Matters: Understanding the Effects of

Device Placement for WiFi Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 6(1), 2022.

[161] Xuyu Wang, Chao Yang, and Shiwen Mao. PhaseBeat: Exploiting CSI Phase Data for Vital Sign Monitoring with Commodity WiFi Devices. In *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*, pages 1230–1239, 2017.

[162] Xuyu Wang, Chao Yang, and Shiwen Mao. TensorBeat: Tensor Decomposition for Monitoring Multiperson Breathing Beats with Commodity WiFi. *ACM Trans. Intell. Syst. Technol.*, 9(1), 2017.

[163] Xuyu Wang, Chao Yang, and Shiwen Mao. On CSI-based Vital Sign Monitoring Using Commodity WiFi. *ACM Transactions on Computing for Healthcare*, 1(3):1–27, 2020.

[164] Yan Wang, Jian Liu, Yingying Chen, Marco Gruteser, Jie Yang, and Hongbo Liu. E-Eyes: Device-Free Location-Oriented Activity Identification Using Fine-Grained WiFi Signatures. In *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*, MobiCom '14, pages 617–628, New York, NY, USA, 2014.

[165] Bo Wei, Wen Hu, Mingrui Yang, and Chun Tung Chou. From Real to Complex: Enhancing Radio-based Activity Recognition Using Complex-valued CSI. *ACM Transactions on Sensor Networks (TOSN)*, 15(3):1–32, 2019.

[166] Teng Wei and Xinyu Zhang. MTrack: High-Precision Passive Tracking Using Millimeter Wave Radios. In *ACM Conference on Mobile Computing and Networking (MobiCom)*, pages 117–129, 2015.

[167] Zhiqing Wei, Fengkai Zhang, Shuo Chang, Yangyang Liu, Huici Wu, and Zhiyong Feng. mmWave Radar and Vision Fusion for Object Detection in Autonomous Driving: A Review. *Sensors*, 22(7):2542, 2022.

[168] Lacey Whited and Derrel D. Graham. Abnormal Respirations. Treasure Island, FL, USA, 2019. https://www.ncbi.nlm.nih.gov/books/NBK470309/.

[169] Chenshu Wu, Zheng Yang, Zimu Zhou, Xuefeng Liu, Yunhao Liu, and Jiannong Cao. Non-Invasive Detection of Moving and Stationary Human With WiFi. *IEEE Journal on Selected Areas in Communications*, 33(11):2329–2342, 2015.

[170] Kaishun Wu, Jiang Xiao, Youwen Yi, Dihu Chen, Xiaonan Luo, and Lionel M. Ni. CSI-Based Indoor Localization. *IEEE Transactions on Parallel and Distributed Systems*, 24(7):1300–1309, 2013.

[171] Libing Wu, Jingxiao Yang, Man Zhou, Yanjiao Chen, and Qian Wang. LVID: A Multimodal Biometrics Authentication System on Smartphones. *IEEE Transactions on Information Forensics and Security*, 15:1572–1585, 2020.

[172] Wei Xi, Jizhong Zhao, Xiang-Yang Li, Kun Zhao, Shaojie Tang, Xue Liu, and Zhiping Jiang. Electronic Frog Eye: Counting Crowd Using WiFi. In *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, pages 361–369, 2014.

[173] Chenren Xu, Bernhard Firner, Robert S. Moore, Yanyong Zhang, Wade Trappe, Richard Howard, Feixiong Zhang, and Ning An. SCPL: Indoor Device-Free Multi-Subject Counting and Localization Using Radio Signal Strength. In *Proceedings of the 12th International Conference on Information Processing in Sensor Networks*, IPSN '13, pages 79–90, New York, NY, USA, 2013.

[174] Siyuan Xu, Zhengran He, Wenjuan Shi, Yu Wang, Tomoaki Ohtsuki, and Guan Guiy. Cross-Person Activity Recognition Method Using Snapshot Ensemble Learning. In *2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall)*, pages 1–5, 2022.

[175] Edwin Yang and Song Fang. GPSKey: GPS-based Secret Key Establishment for Intra-vehicle Environment. In *Workshop on Automotive and Autonomous Vehicle Security (AutoSec) 2022*, 2022.

[176] Edwin Yang, Song Fang, Ian Markwood, Yao Liu, Shangqing Zhao, Zhuo Lu, and Haojin Zhu. Wireless Training-Free Keystroke Inference Attack and Defense. *IEEE/ACM Transactions on Networking*, 30(4):1733–1748, 2022.

[177] Edwin Yang, Song Fang, and Dakun Shen. DASK: Driving-Assisted Secret Key Establishment. In *2022 IEEE Conference on Communications and Network Security (CNS)*, pages 73–81, 2022.

[178] Edwin Yang, Qiuye He, and Song Fang. WINK: Wireless Inference of Numerical Keystrokes via Zero-Training Spatiotemporal Analysis. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*, CCS '22, pages 3033–3047, New York, NY, USA, 2022.

[179] Yanni Yang, Jiannong Cao, Xuefeng Liu, and Kai Xing. Multi-person Sleeping Respiration Monitoring with COTS WiFi Devices. In *2018 IEEE 15th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*, pages 37–45, 2018.

[180] Zhicheng Yang, Parth H Pathak, Yunze Zeng, Xixi Liran, and Prasant Mohapatra. Monitoring Vital Signs Using Millimeter Wave. In *Proc. of the 17th ACM international symposium on mobile ad hoc networking and computing*, pages 211–220, 2016.

[181] Yanming Xiao, Jenshan Lin, Olga Boric-Lubecke, and Victor M. Lubecke. Frequency-tuning Technique for Remote Detection of Heartbeat and Respiration Using Low-power Double-sideband Transmission in the Ka-band. *IEEE Transactions on Microwave Theory and Techniques*, 54(5):2023–2032, 2006.

[182] Mustafa Harun Yılmaz and Hüseyin Arslan. A survey: Spoofing Attacks in Physical Layer Security. In *2015 IEEE 40th Local Computer Networks Conference Workshops (LCN Workshops)*, pages 812–817, 2015.

[183] Jide Yuan, Hien Quoc Ngo, and Michail Matthaiou. Machine Learning-Based Channel Prediction in Massive MIMO With Channel Aging. *IEEE Transactions on Wireless Communications*, 19(5):2960–2973, 2020.

[184] Y. Yuan, J. Zhao, C. Qiu, and W. Xi. Estimating Crowd Density in an RF-Based Dynamic Environment. *IEEE Sensors Journal*, 13(10):3837–3845, 2013.

[185] Youwei Zeng, Dan Wu, Ruiyang Gao, Tao Gu, and Daqing Zhang. FullBreathe: Full Human Respiration Detection Exploiting Complementarity of CSI Phase and Amplitude of WiFi Signals. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(3), 2018.

[186] Yunze Zeng, Parth H. Pathak, and Prasant Mohapatra. WiWho: WiFi-Based Person Identification in Smart Spaces. In *2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, pages 1–12, 2016.

[187] Fusang Zhang, Daqing Zhang, Jie Xiong, Hao Wang, Kai Niu, Beihong Jin, and Yuxiang Wang. From Fresnel Diffraction Model to Fine-Grained Human Respiration Sensing with Commodity Wi-Fi Devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(1), 2018.

[188] Jie Zhang, Zhanyong Tang, Meng Li, Dingyi Fang, Petteri Nurmi, and Zheng Wang. CrossSense: Towards Cross-Site and Large-Scale WiFi Sensing. In *Proc. of the 24th Annual International Conference on Mobile Computing and Networking*, MobiCom '18, pages 305–320, 2018.

[189] Jin Zhang, Weitao Xu, Wen Hu, and Salil S. Kanhere. WiCare: Towards In-Situ Breath Monitoring. In *Proc. of the 14th EAI International Conference*

on Mobile and Ubiquitous Systems: Computing, Networking and Services, MobiQuitous 2017, pages 126–135, New York, NY, USA, 2017.

[190] Linghan Zhang, Sheng Tan, Zi Wang, Yili Ren, Zhi Wang, and Jie Yang. VibLive: A Continuous Liveness Detection for Secure Voice User Interface in IoT Environment. In *Annual Computer Security Applications Conference*, pages 884–896, 2020.

[191] Linghan Zhang, Sheng Tan, Jie Yang, and Yingying Chen. Voicelive: A Phoneme Localization based Liveness Detection for Voice Authentication on Smartphones. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 1080–1091, 2016.

[192] Zhaohe (John) Zhang, Edwin Yang, and Song Fang. CommanderGabble: A Universal Attack Against ASR Systems Leveraging Fast Speech. In *Annual Computer Security Applications Conference*, ACSAC '21, pages 720–731, New York, NY, USA, 2021.

[193] Cui Zhao, Zhenjiang Li, Han Ding, Wei Xi, Ge Wang, and Jizhong Zhao. Anti-Spoofing Voice Commands: A Generic Wireless Assisted Design. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 5(3), 2021.

[194] Peijun Zhao, Chris Xiaoxuan Lu, Jianan Wang, Changhao Chen, Wei Wang, Niki Trigoni, and Andrew Markham. mID: Tracking and Identifying People with Millimeter Wave Radar. In *2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, pages 33–40, 2019.