Sensing and Anti-sensing with Wireless Communication Signal

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree
Doctor of Philosophy in the Graduate School of The Ohio State
University

By

Ouyang Zhang, B.S., M.S.

Graduate Program in Computer Science and Engineering

The Ohio State University

2019

Dissertation Committee:

Dr. Kannan Athreya, Advisor

Dr. Srinivasan Parthasarathy

Dr. Yinqian Zhang

Dr. Wei Zhang

Abstract

Wireless sensing is the core technology that enables a countless number of applications in various areas, such as human-computer interaction, home-security monitor and indoor navigation. Compared with various existing techniques, such as audio-based, camera-based, and radar-based systems, wireless signal based techniques show advantages in low-cost, preserving users' privacy, free from limitations such as the light condition, the line-of-sight link and so on. The ubiquity and drastically growth of communication devices (e.g., WiFi) in the world today represents a huge opportunity in delivering useful services to the society.

However, in designing wireless sensing system, there exist several technical challenges, such as significant noise in the wireless channel, random phase shift due to lack of clock synchronization, continuously changing and dynamic wireless channel and heterogeneous links' qualities. These factors prevent the reliability and robustness of existing wireless sensing systems, which utilizes commodity communication signal. On the other side, given enough time on efforts in the research community and the industry, these sensing systems will in due course be realized robustly and commoditized for broad use. Thus, a great concern comes onto the table as a new form of threat has emerged recently that leaks private information about the whereabouts and activities of physical targets merely by observing the ongoing wireless communications in the scene.

Therefore, to exploit the potential of prevalent existing communication infrastructures for the well-being of human, we as researchers need to not only propose innovative solutions on providing reliable and robust wireless sensing services but also put a significant priority on endeavors to protect private information of the innocent users from leaking. In this dissertation, we study the problems of sensing and anti-sensing with wireless communication signal, e.g., WiFi.

First, we study user-friendly fine-grain finger-gesture recognition. We propose Mudra, a novel wireless sensing technique that takes the first step to recognize finger-gestures with commodity WiFi signal. Mudra provides consistent services against the dynamics in the wireless environment. Second, we study whole-home human activities recognition. To boost the performance over the existing activity recognition systems, we design TifWiFi, which utilizes a two-profile integration approach to complement the drawback on each individual profile. Lastly, we work on protecting users' privacy against general wireless sensing techniques. We propose PhyCloak, which is the first work of its kind to disable eavesdropping on users' private information against any wireless sensing technique with one-antenna setting.

In the above, the two wireless sensing designs cover different application scenarios, i.e., near-field fine-grain small-scale motion detection and far-field course-grain large-scale activity recognition. Thus, they can represent distinct challenges and issues with regards to each scenario. In the protection system, instead of naive jamming method, we design a novel signal processing approach to obfuscate the eavesdropper without its awareness while preserving the legitimate sensing capability. We have implemented the above three systems and conducted extensive experiments in various real-world conditions. Evaluation results show the effectiveness of our systems.

This is dedicated to My Family.

Acknowledgments

In the long journey of the PhD career, I have received numerous help, support and advice from many people around me which are indispensable to make it fruitful. Otherwise, this dissertation would not be possible. I would like to express the deepest appreciation to all of them.

First and foremost, I would like to thank my advisor, Dr. Kannan Athreya, who has been supportive of my career goals. His expertise was invaluable in the formulation of the research topic and methodology in particular. His guidance helps me to develop as a good researcher. Throughout time at The Ohio State University, a large part of my skills for research are developed from the discussion during the project meeting, such as how to seek and define a problem with potential significance, how to solve a problem with solid contribution and how to evaluate a work in a critical manner. He believes in me like nobody else and gives personal advice in daily life just like a close friend.

I would like to thank my dissertation committee members for their constructive advice. Their remarks has helped to improve the writing and shape my final dissertation. I am also thankful to Dr. Nandi, Dr. Parthasarathy and Dr. Zhang for their incisive comments during my candidacy exam to guide my following research works. I am also grateful to lots of faculty members at the The Ohio State University. They have provided me with an excellent education. In particular, I would like to thank Dr.

Ness Shroff who has given me plenty of valuable advice during our collaboration. His knowledge of techniques and attitude to research deeply impress me. I also extend my appreciation to Dr. Jihun Hamm, Dr. Mikhail Belkin, Dr. Can Emre Koksal, Dr. Raef Bassily, Dr. Srinivasan Parthasarathy and Dr. Anish Arora, who taught me knowledge on machine learning, artificial intelligence, database and networking.

In addition, I would like to express sincerest thanks to my collaborators, especially, Dr. Wenjie Zhou, Dr. Avishek Banerjee, Dr. Yue Qiao, Dr. Fei Wu, Dr. Zhenzhi Qian, Dr. Yifan Mao, Dr. Yang Yang, Dr. Bo Chen, Dr. Xingya Zhao and Dr. Arjun Bakshi. Without their continued supports and contributions, it is not possible for me to accomplish my research works. Beyond the research, they make my experience in the research group warm and fun in daily life. I am also grateful to my industrial collaborators. I spent one summer in the wireless connectivity group at Facebook. I would like to thank Javier Cardona (my mentor at Facebook) for his great mentorship and guidance.

Last but not least, I am grateful to my family and friends. My parents are always there with support and encouragement in whatever I pursue. I love you. Finally, I would like to thank my friends, who are of great help and shared valuable happy moments with me in daily life.

Vita

July 11, 1994	Born - Fuzhou, China
2014	B.S. Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, China
2019	M.S. Computer Science and Engineering, The Ohio State University
2014-present	Graduate Research Associate / Graduate Teaching Associate, The Ohio State University

Publications

Research Publications

Ouyang Zhang, Zhenzhi Qian, Yifan Mao, Kannan Srinivasan, and Ness B. Shroff "ERSCC: Enable Efficient and Reliable Screen-camera Communication". *in Proc. of the Nineteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, July 2019.

Ouyang Zhang and Kannan Srinivasan "Mudra: User-friendly Fine-grained Gesture Recognition using WiFi signals". *in Proc. of the 12th International on Conference on emerging Networking EXperiments and Technologies*, December 2016.

Yue Qiao, Ouyang Zhang, Wenjie Zhou, Kannan Srinivasan, and Anish Arora "PhyCloak: Obfuscating Sensing from Communication Signals". *13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16)*, March 2016.

Fei Wu, Yang Yang, Ouyang Zhang, Kannan Srinivasan, and Ness B. Shroff "Anonymous-Query based Rate Control for Wireless Multicast". *in Proc. of the 17th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, July 2016.

Bo Chen, Yue Qiao, Ouyang Zhang, and Kannan Srinivasan "Airexpress: Enabling in-band Wireless Cut-through Transmission". *in Proc. of the 21st Annual International Conference on Mobile Computing and Networking*, September 2015.

Anran Wang, Chunyi Peng, Ouyang Zhang, Guobin Shen, and Bing Zeng "InFrame: Multiflexing Full-Frame Visible Communication Channel for Humans and Devices". *in Proc. of the 13th ACM Workshop on Hot Topics in Networks*, Oct 2014.

Fields of Study

Major Field: Computer Science and Engineering

Studies in:

Computer Networking	Prof. Kannan Athreya
	Prof. Can Emre Koksal
	Prof. Anish Arora
Artificial Intelligence	Prof. Mikhail Belkin
	Prof. Raef Bassily
	Prof. Jihun Hamm
Database	Prof. Srinivasan Parthasarathy

Table of Contents

Page

Abstract ii

Dedication iv

Acknowledgments v

Vita vii

List of Tables xii

List of Figures xiii

1. Introduction 1

   1.1 User-friendly Fine-grain Finger-gesture Recognition with WiFi signal 3
   1.2 Two-profile Integration Framework for Device-free Activities Recognition with Communication Signal 5
   1.3 Obfuscating Sensing from Communication Signal 7

2. Related Work 10

   2.1 Fine-grain Gesture Recognition with Wireless Signal 10
   2.2 Two-profile Integration Framework for Device-free Activities Recognition with Communication Signal 13
   2.3 Obfuscating Sensing from Communication Signal 16

3. Mudra: User-friendly Fine-grain Finger-gesture Recognition with WiFi signal 19

   3.1 Overview 19
   3.2 Mudra Design 24

3.2.1 Preliminary: System Scenario and Setup 24
3.2.2 Highly-sensitive Motion Indication 25
3.2.3 Resource-friendly Signal Processing. 28
3.2.4 Environment-agnostic Gesture Pattern. 33
3.2.5 System Model 36
3.3 Evaluation 38
3.3.1 Implementation 38
3.3.2 Micro-Benchmarks 40
3.3.3 System Evaluation 48
3.4 Discussion 55
3.5 Summary 57

4. TifWiFi: Two-profile Integration Framework for Device-free Activities Recognition with Communication Signal 58

4.1 Overview 58
4.2 TifWiFi Design 62
4.2.1 Insufficiency of Single Profile 62
4.2.2 Multipath-Profile Analysis 64
4.2.3 Motion-Profile Analysis 70
4.2.4 Profile Integration Mechanism 74
4.2.5 Activity Detection and Segmentation 76
4.2.6 Integration of multiple links 77
4.3 Implementation and Evaluation 78
4.3.1 Testbed 78
4.3.2 Infrastructure Setup 79
4.3.3 Data Collection 80
4.3.4 Activity Detection 81
4.3.5 Multipath-Profile Analysis 82
4.3.6 Motion-profile Analysis 85
4.3.7 Holistic System Performance 86
4.4 Summary 89

5. PhyCloak: Obfuscating Sensing from Communication Signal 90

5.1 Overview 90
5.2 System Model and Preliminary 94
5.2.1 Threat Model 94
5.2.2 System and Goals 94
5.2.3 Three Degrees of Freedom 95
5.3 PhyCloak Design 98
5.3.1 Online Maintenance of Self-Channel Estimates 98

5.3.2 Obfuscation of Patterns in 3 DoFs 105
5.3.3 Spoofing 109
5.3.4 PhyCloak 110
5.4 Implementation and Evaluation 111
5.4.1 Experimental Setup 111
5.4.2 Self-Interference Cancellation 112
5.4.3 Obfuscation Performance 114
5.4.4 Throughput Performance 122
5.5 Discussion 123
5.6 Summary 123

6. Final Remarks 125

Bibliography 127

List of Tables

Table **Page**

2.1 Summary of recent SISO sensing systems 16

3.1 Consistency study of gesture waveforms with various signal source locations. 35

3.2 Various computation costs w.r.t two approaches. 54

4.1 Codes and Locations for Tested Activities 79

List of Figures

Figure **Page**

1.1 Raw traces of CSI value on one subcarrier. 7

2.1 **Phase/Amplitude variation in channel.** When fingers repeat the 'Come' gesture ten times, the corresponding CSI amplitude/phase change is buried within large noise, making it infeasible for finger-gesture recognition. This is the same with other gestures while we just show 'Come' to save space, same with Fig. 2.2. 12

2.2 **Doppler profile in frequency domain.** Large gestures such as Limb Push produce noticeable Doppler shift (the above one) while figure gesture Come can not induce discernible shift, which is the same as no motion. 13

3.1 The system scenario of Mudra. 21

3.2 Gestures considered in Mudra. 21

3.3 Direction Demonstration 24

3.4 **Cancellation Performance Comparison.** We test in various scenarios. 'Full CSI' estimates on the whole band while 'Partial CSI' just calculates overall amplitude/phase difference. 26

3.5 WiFi spectrum 28

3.6 Canceled signal spectrum before and after smoothing over 200 neighbors. The left plot is the power spectrum across 20MHz generated by original samples and the right one is a smoothed version by averaging over adjacent 200 values. The trough in each of them is magnified in a subplot on the top. 32

3.7 System SNR with and without iron plate. When signal source is very close (2cm), we test waveform-to-noise-amplitude ratio using 'Come' gesture in a circle in eight equi-angular directions. 33

3.8 Waveform of gestures considered in Mudra. 35

3.9 **System model** 37

3.10 Floor plan 39

3.11 Noise Variation: 'wom' is short for "without body movement" and 'wm' is short for "with body movement" between the source and receiver. . 40

3.12 Testbed 40

3.13 Sensitivity Study. **Takeaway**: Sensitivity of Mudra changes with antenna distance while keeping stable with source distance. 42

3.14 **Orientation Effect.** Upper four plots study moving direction; bottom ones study hand position. 42

3.15 Impact of packet interval. 45

3.16 Impact of packet length. 46

3.17 False detection rate during a whole day. 49

3.18 Confusion Matrix with Local Source. 50

3.19 Confusion Matrix with Remote Source. 50

3.20 Variation range of gesture duration across users. *The box represents the quantiles while the whisker means the standard deviation. The red line is the mean while the red plus is the median.* 52

3.21 Performance of gesture recognition with CNN-RNN model in various environment settings. 54

4.1 Experiment scenario for preliminary study. 62

4.2 Drink. 63

4.3 Spine Stretch. 63

4.4 Draw. 63

4.5 Bend-over. 63

4.6 Drink. 64

4.7 Spine Stretch. 64

4.8 Draw. 64

4.9 Bend-over. 64

4.10 Normalized CSI amplitude in 4 links of 2x2 MIMO. 65

4.11 Bi-directional Recurrent Neural Network Architecture. 67

4.12 Diagram of training sample extraction. 69

4.13 Effect of PCA-based de-noising. 73

4.14 Motion intensities of three activities. 75

4.15 Testbed. The left is an office room and the right is a two-bedroom apartment. 78

4.16 Detection range of TifWiFi 81

4.17 Confusion matrix of TifWiFi with the multipath profile. 82

4.18 Confusion matrix of 'amp-dist' with the multipath profile. 83

4.19 Accuracy of three approaches with the multipath profile. 83

4.20 Confusion matrix of TifWiFi with motion profile. 85

4.21 Confusion matrix 86

4.22 Detection rate of all activities with two links and integration. 87

4.23 Recognition accuracy of all activities with two links and integration. . 88

4.24 Recognition accuracy of all activities with three approaches. 89

5.1 4 single-input-single-output (SISO) nodes exist in the system: Alice, Bob, Carol and Eve: Alice and Bob perform data transmission and reception; Eve performs illegitimate sensing by exploiting Alice's transmission; Carol also performs sensing, but her obfuscator module forwards the received signal in a way that distorts physical information but preserves logical information 95

5.2 Expected Doppler shift and phases are generated at a forwarder . . . 97

5.3 High-level block diagram of PhyCloak 98

5.4 Training sequence 102

5.5 Channel coefficients measured at different sampling rates 102

5.6 The granularity of the spectral decreases as the Doppler shifts change from 1s to 0.05s 103

5.7 The pattern that a WiSee sensor sees in Figures 5.2(a) is hidden by an obfuscated signal where Doppler shift changes every 0.1 second . . . 107

5.8 The constant angle of human motion (starting from 9th second) that a Wi-Vi style sensor sees in (a) is hidden by an obfuscated signal where phase changes randomly every 0.1 second 107

5.9 Spoofing 109

5.10 Self-interference cancellation performance 112

5.11 Obfuscation level of each of the three features decreases as SOR (original signal over obfuscation signal) increases 115

5.12 Eve's classification accuracy vs d_{EO} 117

5.13 Eve's classification accuracy vs d_{TO} 118

5.14 Eve's classification accuracy under different Tx-Ox and Ox-Eve channel conditions 119

5.15 False positives with a spoofing Ox 120

5.16 Square wave based training preserves legitimate sensing 121

5.17 Throughput 122

Chapter 1: Introduction

Wireless sensing is the core technology that enables a countless number of applications in various areas, such as human-computer interaction, home-security monitor and indoor navigation. Compared with various existing techniques, such as audio-based ([42, 68]), camera-based ([28, 62, 78]), and radar-based ([2]) systems, wireless signal based techniques show advantages in low-cost and preserving users' privacy. Besides, it is free from limitations such as interference of the background sound, the requirement of the light condition, and the line-of-sight (LoS) path, encountered by the others. The ubiquity and drastically growth of Wi-Fi devices in the world today represents a huge opportunity in delivering useful services to the society. Up to now, there are 23.1 billion connected IoT devices worldwide and it is predicted that by 2022 the number will be boosted to 42.6 billion ([58]). However, to exploit the potential, there exist several technical challenges which prevent the reliability and robustness of wireless sensing applications based on commodity Wi-Fi signal.

First of all, due to the environmental variation and device imperfection, the wireless channel is noisy. As such, the collected channel state information (CSI) deviates from the true channel coefficients. In some cases, the noise is so large that could cause a significant error in the sensing applications. What's more, without clock synchronization between the transmitter and the receiver, random phase shift occurs in each received packet. Therefore, the useful channel phase information is overwhelmed by the random shift and thus loses its function in sensing. Secondly, the wireless environment is dynamic and unpredictable. There are three main factors that cause dynamics in wireless channel - moving objects (human beings and vehicles), moving settings (furnitures), and moving streaming endpoints (transmitter and receiver). The dynamics of the channel makes it hard for the sensing technology to be reliable. Last but not least, link conditions in the wireless environment are quite heterogeneous. That is, different wireless links in the space would exhibit large discrepancy in the channel quality and thus the performance as well. Thus, it remains an issue to select proper locations for installation and combine the results from different wireless links.

On the other side, given enough time on efforts in the research community and the industry, these sensing systems will in due course be realized robustly and commoditized for broad use. Thus, a great concern comes onto the table as a new form of threat has emerged recently that leaks private information about the whereabouts and activities of physical targets merely by observing the ongoing wireless communications in the scene. Broadly speaking, as a wireless signal gets reflected off of people and other objects in the scene, information about them is leaked to eavesdroppers by computational analysis of the signal distortions. Increasingly, researchers have been demonstrating proof of concepts where not only people presence but also fine-grain information about their locations and even breathing, lip movement or keystrokes is leaked [8, 39, 50, 65, 70]—all from observing communication signals that are widely prevalent in our homes.

In this dissertation, we present three pieces of work regarding the problem of sensing and anti-sensing with wireless communication signal. In the following, we briefly introduce each part.

1.1 User-friendly Fine-grain Finger-gesture Recognition with WiFi signal

In modern society, wireless signals have been prevalent and ubiquitous in coverage in our daily life. Although these signals are originally intended for communication purpose to connect different computing endpoints wirelessly, researchers both in the academia and in the industry have explored and demonstrated their potential in sensing the context in the surrounding environment. Recently, with the advances in wireless sensing, researchers have extended their studies from macro-gesture recognition to micro-gesture recognition. In this work, we take the first step to recognize finger-gestures, which are fine-grain motion with centimeter-range, with wireless communication signal. We envision such a technology could deliver a new set of service to the public over their portable and mobile devices.

One potential application especially supported by this technique is the human-machine interaction with wearables that have limited screen space. With nine gestures, users can easily input instructions into the device as controlling command on the application which is active on the screen, such as scrolling through one page, clicking button, and switching among applications. Let's imagine a reader is reading an e-book on apple watch. He can move to next page with a 'tap' gesture and return back with 'double tap' gesture, or circling his finger to get to the bottom. It can also enable complex and in-the-air interactive games on portable devices. Users can control the action of figures with finger gestures in the air, such as going ahead, returning back, jumping and punching.

Enabling a fine-grain finger gesture recognition system like this needs millimeter-level sensitivity on the system over the motion of the object. In this paper, we develop a novel finger-gesture recognition system, Mudra utilizing WiFi signals. Mudra just needs regular WiFi data transmissions either from local or remote sources. In this way, Mudra doesn't need to generate special signals (like Soli), thus, saving power for power-hungry devices [1,49] and not hurting communication opportunities. A key feature of Mudra is its training-free nature. This makes our system more promising in reality compared with other micro-motion technologies [17,30,64–66], all of which require training for specific location and user.

However, it is non-trivial to meet the above design goals to make the system user-friendly. Finger gestures are micro-motion which is hard to detect because they result in subtle changes in the wireless channel. The mobile and dynamic application scenario of the system poses challenges in delivering training-free services to the users. To overcome these difficulties, we design three main components in our system. The first one is *Highly-sensitive Motion Indication*, which makes our system robust against the large noise in the wireless channel. The second one is *Resource-friendly Signal Processing* which enables the system functionality with regular WiFi signals. The last one is *Environment-agnostic Gesture Pattern* , which generates consistent and environment-agnostic patterns, making our system free from specific training data from users per surrounding environment.

We implement a prototype with software defined radio and COTS WiFi transmission. The receive infrastructure of our system just includes two antennas without special structure requirements. With 802.11n protocol [75], MIMO devices with multiple antennas are very common. Thus, Mudra can be implemented on COTS receiver if manufacturers open access to the samples. Mudra tracks finger-motion by deriving motion indication with cancellation between two received signals. Stretch Limited DTW algorithm is proposed to classify gestures with variation in duration and shape.

1.2 Two-profile Integration Framework for Device-free Activities Recognition with Communication Signal

Human activity recognition serves as the crucial part of numerous human-centered computing services, such as smart home, elderly care, assisted living, and etc. WiFi signal based passive activity recognition is to capture the on-going wireless communication signal in the air and infer the context information about the human activity in the target area with proper signal processing approach. Similar to the gesture recognition in the previous section, this topic also falls in the area of wireless sensing based on communication signal. However, activity recognition targets at a totally different application scenario in the aspect of the scale of the motion and the size of the object. Additionally, the nature of the available infrastructure and the environment dynamics is also different. Thus, in the problem of activity recognition, we need to address a distinct set of challenges for high-performance sensing.

In the human activity recognition system, the users perform normal activities such as cooking, washing and eating. By inferring the type of activities that he/she is performing, such a system could assist to deliver a context-aware home experience. For example, light up different bulbs with different color to each distinct activity. Apart from in-position activities, the users could also move around across multiple rooms with their trajectory recorded. With regards to such a smart home system, it could be easily seen that the target moving object is the whole human body. In the meantime, the range of the motion is now in the scale of hundreds of square meters. Does it mean activity recognition could be much more easier than gesture recognition? No exactly.

Without attached sensors on human body, device-free activity recognition provides a more favorable user experience by analyzing signal received from on-going transmission. In this setting, the network infrastructures are the machines that could be far from the human object. In the wireless channel, when the object is far from the communication signal link in the air, then its motion has unnoticeable distortion on the receive signal. Besides, in this work, we target at enabling the application on COTS WiFi devices by utilizing the protocol-support channel information of CSI reported by the network chipsets. However, the raw CSI traces are hard to be used directly for pattern recognition due to the large noise. This can be seen from the example in the Fig.1.1, where the variation related to a person's walking is buried in the noise.

In this work, we propose TifWiFi, a two-profile integration framework for device-free human activity recognition. TifWiFi is built upon existing WiFi devices and thus removes the cost and burden of deployment. It is also a passive detection system with no privacy concern for users. The basic idea behind TifWiFi is to utilize status and motion of human activity. Specifically, the status of the activity is captured by a multipath profile, which represents the multipath propagation condition. The profile for the motion part is represented by a frequency domain speed model. As far as we know, TifWiFi is the first to utilize both profiles in the literature. It is worth noting

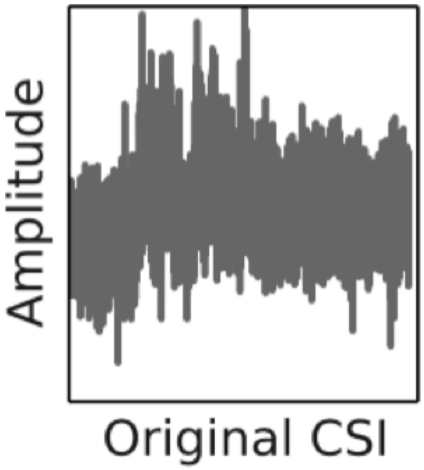


Figure 1.1: Raw traces of CSI value on one subcarrier.

that TifWiFi doesn't require any extra information source to construct two profiles. The reason is that they are constructed with different processing on the CSI data.

We build our system on commodity WiFi devices and conduct experiments in real home and office environment to validate our design and evaluate the system performance. The results show that our system achieves a substantial improvement over the conventional approaches.

1.3 Obfuscating Sensing from Communication Signal

Recognition of human activities and gestures using pre-existing WiFi signals has been shown to be feasible in recent studies. Given the pervasiveness of WiFi signals, this emerging sort of sensing poses a serious privacy threat. This new form of threat has emerged recently that leaks private information about the whereabouts and activities of physical targets merely by observing the ongoing wireless communications in the scene. Broadly speaking, as a wireless signal gets reflected off of people and other objects in the scene, information about them is leaked to eavesdroppers by computational analysis of the signal distortions. Increasingly, researchers have been demonstrating proof of concepts where not only people presence but also fine-grain information about their locations and even breathing, lip movement or keystrokes is leaked [8, 39, 50, 65, 70]—all from observing communication signals that are widely prevalent in our homes.

While the upside is that legitimate users can detect these physical "signatures" simply using existing signals, a burglar can also detect that there are no people in a house, a passerby can decipher key presses without leaving a trace [17], and a neighbor can snoop on the activities in our homes [70]. There is little doubt that several of these privacy exploits will in due course be realized robustly and commoditized for broad use. And, given the pervasive nature of wireless communications, the privacy implications of such attacks will undoubtedly be of major social importance.

It is thus timely and important to develop suitable counter-measures for this type of privacy leakage. This paper is the first to counter the threat of unwanted or even malicious communication based sensing. To avoid any modification of existing receivers, we need to build an obfuscator (Ox) that works independently from a receiver (Rx) and can yet deter privacy leakage. At the same time, Ox should not hurt the ongoing reception at the intended receiver. In addition, given the diversity of the design of RF based sensors and invisibility of eavesdroppers, it is not reasonable to assume Ox that uses a specific obfuscation approach against a specific Eve. Thus, our goal is to build a black-box solution which distorts only the privacy sensitive information while not affecting the logical information.

We propose PhyCloak to protect privacy information from unwanted or even malicious sensing with no modification to existing wireless infrastructures. The effectiveness of the design is validated via a prototype implementation on an SDR platform.

Chapter 2: Related Work

2.1 Fine-grain Gesture Recognition with Wireless Signal

Human tracking and motion detection have been broadly studied in the literature. We discuss existing techniques and their shortcomings in our setting.

Non-RF based micro-gesture recognition: Imaging-based systems (e.g. Xbox Kinect [4], leap motion [3] and Maestro [21]) use monochromatic infrared cameras and LEDs to build body-depth imaging. Both Radar and infrared systems require line-of-sight (LOS) operation. Sonar-based system [47] requires specially modulated sound wave and exclusive devices. Mudra, on the other hand, utilizes regular WiFi signals in the ISM band, operating in parallel with WiFi communication. Furthermore, Mudra supports both line-of-sight (LOS) and non-line-of-sight (NLOS) scenarios.

RF based micro-motion tracking: Another set of works exploit RF signals to track human's minor movements. Radar based systems (e.g. Google Soli [2]) show the ability to track minor finger movement by constructing Doppler profile using 60GHz radar signals. Such systems, however, require embedded chip to generate/capture and process radar signals. Besides, it is limited to line-of-sight (LOS) scenario due to the extra-directional feature of radar signals.

[17,30,64,66] look into the keystroke detection problem. Mudra differs from them fundamentally in that Mudra is targeted at identifying finger gestures, not keystroke action. More importantly, Mudra is a user-friendly system without the need for training. Whereas, the others require location-specific or user-specific training for classification. One recent work detects minor keystroke motion by canceling two signals from two receiving antennas [17]. This system can not be applied to WiFi signals since (i) it assumed continuous signal transmissions and (ii) assumed the frequency response to be flat across the whole band. Therefore, they generate continuous wireless signal from SDR platform and only estimate the overall phase and amplitude of the channel. WiFi transmissions are not continuous and do not have a flat frequency response due to a small number of subcarriers spanning the whole band. The second issue affects cancellation quality, which is what Mudra relies on to identify finger-gestures. Mudra enables micro-gesture recognition with WiFi signals after solving those problems by implementing smart packet capture and estimating full channel state information (CSI). Mudra also resolves additional issues for WiFi signals such as variable length of packets.

[64, 66] make use of sound and electromagnetic signals emanating from keyboards, a feature unavailable with finger motion. [30] mainly use Principle Component Analysis (PCA) to remove noise based on correlated variation of CSI with different keystrokes. It is highly unlikely that such correlation still works with various finger gestures since finger movements cause subtle changes to CSI: Fig. 2.1 shows the amplitude and phase variation over time when the user performed the same gesture 10 times. The plot shows no periodic variations. Another line of approaches (e.g. [65]) use MIMO beamforming, using multiple antennas and step motors to project signal to

target. Target locating needs 6s and 85% of accuracy greatly affects overall accuracy. Mudra, on the other hand, doesn't require such a luxurious infrastructure.

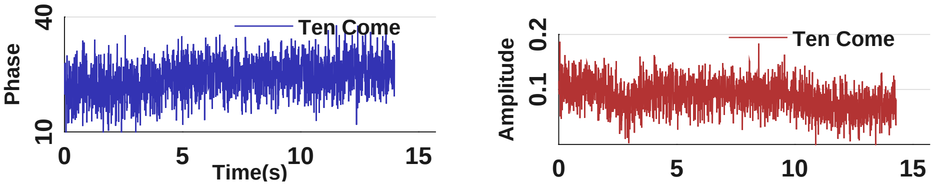


Figure 2.1: Phase/Amplitude variation in channel. When fingers repeat the 'Come' gesture ten times, the corresponding CSI amplitude/phase change is buried within large noise, making it infeasible for finger-gesture recognition. This is the same with other gestures while we just show 'Come' to save space, same with Fig. 2.2.

Coarse-grained gesture/motion detection: A large body of works track and identify larger-scale gesture motion (e.g. human limbs), which typically spans over range of several decimeters or target at the whole body, using characteristics of wireless signal. [50, 59] enables gesture recognition using Doppler profile generated from multi-second FFT to get sub-Hertz granularity. However, typical finger gestures are much shorter (e.g. tap just spans 0.5s), making this approach infeasible. To see this, we implement Wisee [50] on PXIe-1082 platform and see no observable Doppler shift with finger motion in Fig. 2.2. Frequency modulation was used with sound wave in [46] and carrier wave in [9] to track human breath. Mudra differs in that we don't change the regular source to generate a delicately-designed signal, which can only serve recognition purpose.

Channel CSI / RSSI-based systems [32, 70] use coarse CSI feature to extract motion information, which is not feasible in detecting slight finger motion (Fig. 2.1). As to antenna array based technologies (e.g. [7] and [29]), WiDeo [29] spans four-antenna

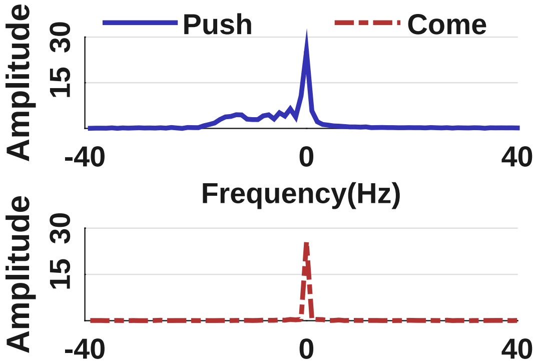


Figure 2.2: Doppler profile in frequency domain. Large gestures such as Limb Push produce noticeable Doppler shift (the above one) while figure gesture Come can not induce discernible shift, which is the same as no motion.

infrastructure covering a length of 18cm, which is not available in portable devices. Wi-Vi [7] assumes a constant motion speed (1m/s) simply to identify whether the object is walking towards or away from a receiver. Another line of works get high positioning accuracy by attaching object with specific devices([37, 79, 81]). Tagoram [79] locates at a centimeter level using phase information of RFID backscatter signal. [81] tracks smartphone as a mouse by active sound wave transmission. Mudra is a user-friendly system which doesn't require such intrusive body instrumentation.

2.2 Two-profile Integration Framework for Device-free Activities Recognition with Communication Signal

Human activity recognition has been a hot research area in the past decades. Existing approaches could be broadly separated into two categories. One category of methods make utilization of ubiquitously deployed WiFi infrastructure and devices.

While, the other set of approaches exploit non-WiFi signal by installing extra devices in the target area.

Non-WiFi based approaches. Non-RF signal based approaches such as vision-camera based [3, 4, 10], and radar-based [2], has been proposed in the literature. Compared with WiFi signal based methods, they are all limited to the line-of-sight (LoS) scenario. Apart from installing extra hardware in the environment, they also suffer disturbance in the environment condition. For instance, vision-based systems have implicit requirements of proper background light condition. Besides, user's privacy is also a big concern with these techniques due to the abundant information leaked by images. Similarly, radar-based methods suffer high directionality with 60 GHz signal and low-cost radar system has limited range (tens of centimeters).

By attaching devices on the user's body, researchers can infer the activities he/she engaged in by analyzing data from various sensors. Philips [5] and Pbn [31] attached accelerometer on human body to infer the activity. Body-Scope [80] deploy acoustic sensors to recognize daily activities like eating or coughing. However, attaching these sensors either cause a burden on the users or extra deployment cost and thus are not preferred and available in most applications.

WiFi-based approaches. Due to the ubiquity of WiFi infrastructures and devices, WiFi signal has attracted lots of interests in the research community.

One group of approaches utilize the statistics of CSI values collected during the activity for recognition. E-eyes [70] explored the channel state information from commodity devices to build a database of location-activity profiles. As different activities cause different CSI distributions, this system identified the activity by comparing its amplitude histograms with the pre-stored profile database. In contrast, another group of methods look into the variation pattern in CSI training data for pattern recognition. For instance, WiSee [50] infers the human gestures by looking into the Doppler shift caused by moving human body parts. This system relied on a special hardware design with USRP to obtain the fine-grain frequency resolution in OFDM signal. CARM [67] built a CSI-activity model by constructing a mapping from CSI to human speed. WiFinger [60] extracted and identified the CSI variation patterns caused by human finger gestures.

The fundamental limitation of the above approaches is that they just utilize a single facet of the human activity, i.e., either the status or the motion, which is explained in Sec.4.1. In this work, we design a two-profile integration framework which can utilize both facets of the activity to improve the activity recognition system. Unlike current work [70] which builds the multipath profile with just single-subcarrier amplitude distribution, we exploit the statics in the whole CSI information to make use of the diversity embedded among multiple subcarriers. Part of the idea is inspired by CSI-fingerprinting based works [55, 69, 74, 82]. Moreover, we design a statistical analysis scheme on the motion profile for loosely-defined human activities. To optimally integrate both analysis, we design the PBD algorithm to make use of the inherent priority.

Another line of related works [38,41,57,77] use signal processing based approaches which derives the angle-of-arrival (AoA) and time-of-flight (ToF) information. Those approaches rely on at least three antennas on the receivers which may not be satisfied by a large portion of commodity WiFi devices. Besides, they need dedicate calibration and deployment to have accurate location reference from the static WiFi infrastructures and more than one receivers to enable intersection for pointing to the

Existing Work	Feature Basis	Device	Sensing Task
WiSEE: Pu et al. [50]	Doppler Shift	USRP-N210	Gesture recognition
Wi-Vi: Adib and Katabi [8]	Phase	USRP-N210	Gesture based communication,tracking
E-eyes: Wang et al. [70]	RSSI, CSI	COTS 802.11n devices	Activity classification
Gonzalez-Ruiz et al. [25]	RSSI	IEEE 802.11g wireless card	Obstacle mapping
Wang et al. [67]	Phase, CSI	COTS 802.11ac devices	Activity classification
WiKey: Ali et al. [11]	CSI	COTS 802.11n devices	Key recognition
RSA: Zhu et al. [85]	RSS	HXI Gigalink 6451 60GHz radios	Object imaging

Table 2.1: Summary of recent SISO sensing systems

target. Moreover, the AoA information is only useful in the direct link which implies those approaches fail in non-LoS scenario and area with strong obstructions. In contrast, TifWiFi are not limited to LoS. Besides, TifWiFi has a minimum hardware requirement and doesn't require special deployment and location calibration.

2.3 Obfuscating Sensing from Communication Signal

RF sensing from communications has been of great interest in the last few years, as it allows data signals to be exploited to infer remarkable details about the physical world. Although the primary purpose of the communication signals is to carry logical information, concepts of radar analysis [14,20,26,27,33,34,36,40,48,52,56,61] are adapted to extract these details. There are however several challenges in the adaptation since communication signal is defined particularly for carrying data. For example, radar systems control their resolution by specially encoding their transmitting signals, say in the form of Frequency-Modulated Carrier Waves (FMCW) for spectrum sweeping, but when sensing from RF communication a similar sort of transmitter cooperation typically cannot be leveraged. As another example, sophisticated radar signal processing techniques, say creating a synthetic aperture using a *large* number of antennas, cannot be implemented directly in communication systems due to resource limitations.

Many techniques have been developed and demonstrated to address the above mentioned challenges for diverse sensing tasks including motion tracking [8], activity/gesture recognition [50,67,70], and obstacle/object mapping/imaging [25,85], and even minor motions like keystrokes recognition [11,17] and lip reading [65]. One idea is to use one antenna to emulate an antenna array in the presence of human movement. By tracking the angle of the reflected signal from the target (human) [8], the system is able to track the motion of the target as a form of inverse synthetic aperture radar (ISAR). Ubicarse [39] exploits the idea of circular synthetic aperture radar (SAR), in which the system rotates a single antenna so as to emulate a circular antenna array. As SAR does not require the target to be in motion, unlike the case of ISAR, Ubicarse proposes a method of using a handheld device to create circular antenna array to perform localization. To overcome any imprecision in the circle created by the rotation, it refines the formulation of SAR by using the relative trajectory between two receive antennas. Some other techniques characterize signatures corresponding to the channel variation caused by human activities. E-eyes [70] shows that temporal RSS and CSI features, which are available in COTS devices, can be used in activity classification, albeit this requires relatively heavy training. WiSee [50] proposes a method to extract Doppler shifts from OFDM symbols by applying a large FFT over repeated symbols, and gesture recognition is then shown to be possible from the extracted Doppler shifts. Another interesting technique used by communication based sensors maps obstacles/objects [25, 44]. The Tx-Rx pairs detect the presence of obstacles via wireless measurements and thereby co-operatively draw the indoor obstacle map.

As our protection system is single-input-single-output (SISO), we focus on breaking any SISO illegitimate sensing system in this work. Although SISO sensing systems use diverse techniques exemplified in Table 2.1, they all leverage a subset of the 3 DoFs discussed in Section 5.1. Since PhyCloak provides a generic tool to obfuscate in all these three dimensions, it can protect against any SISO sensor.

In contrast, for a multi-antenna sensing system, there is an additional DoF—the relative placement of antennas—that yields other types of information like angle of arrival (AoA) and time difference of arrival (TDoA). Nevertheless, by rotating PhyCloak's transmit antenna or extending our framework to a multi-antenna protection system, we would have the freedom to also obfuscate the fourth dimension provided by a multi-antenna sensing system.

Chapter 3: Mudra: User-friendly Fine-grain Finger-gesture Recognition with WiFi signal

In this chapter, we study the challenge of detecting micro-gestures of human and propose the design of a reliable and robust fine-grain finger-gesture recognition system, Mudra with the commodity communication signal, e.g., WiFi signal. In Mudra, we propose a highly-sensitive motion indication which enables the system to track fine-grain finger motions precisely. Apart from the high sensitivity, we make Mudra user-friendly and propose a delicate gesture pattern design which is consistent in the dynamic wireless environment. What's more, Mudra is also resource-friendly with parallel operating with the commodity communication signal using novel signal processing techniques.

3.1 Overview

We propose Mudra, a system that is able to precisely track and recognize human finger gestures using only the commodity communication signal. Mudra falls in the area of wireless sensing, which generally utilizes wireless signal, e.g., radio-frequency signal, to detect and/or recognize object's motion in the target area. Currently, wireless sensing technologies have tried to localize human body or detect human motion in the range of a large building or whole-home environment. As illustrated below, we believe that sensing and recognizing fine-grain gestures could enable a set of new applications to reveal a whole-new prospect of human-machine interaction.

As like other wireless sensing technologies, Mudra utilizes the changes in the received wireless signal to derive the motion of the objects in the space. However, Mudra takes the first step to look into finger-gesture recognition using conventional ISM-band signals. Specifically, when the user performs micro finger-gestures in the vicinity of a mobile device, the received communication signal is influenced due to the changes in the multipath propagation channel. The principle of our system is to derive the finger motion based on such changes in the received signal. In this way, Mudra provides a service to recognize different gestures performed by the target user.

The impetus behind designing Mudra system is the ubiquity and drastic growth of Wi-Fi devices in the world today. Up to now, there are 23.1 billion connected IoT devices worldwide and it is predicted that by 2022 the number will be boosted to 42.6 billion. Therefore, a gesture recognition system based on conventional communication signal represents a potential in delivering broadly-available service to the society. What's more, the communication radios are no longer standalone communication processing circuits. On the one hand, the extra processing capability from software defined radio or customized firmware enables delicate signal processing on received signals. On the other hand, with the support from hardware protocol to report meta-data from driver to the user space, the powerful processing platform of the smartphone or PC where the wireless cards are attached with can be utilized to do computation-heavy analysis and thus enable innovative applications.

There are recent advances in pushing gesture recognition techniques. Google Soli [2] makes a huge breakthrough pushing motion sensing to millimeter-level accuracy,

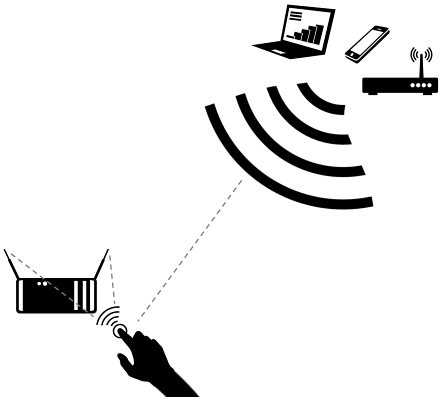


Figure 3.1: The system scenario of Mudra.

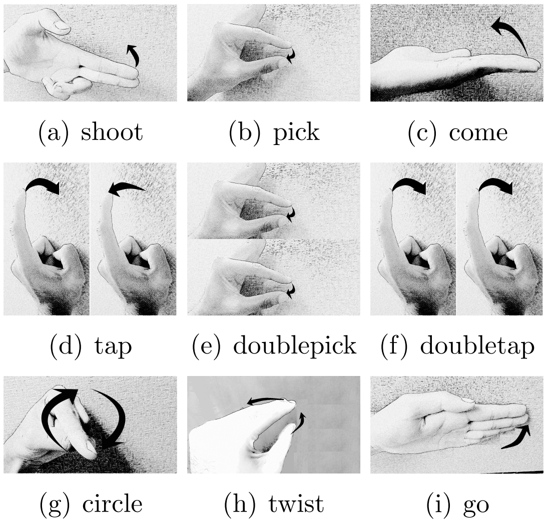


Figure 3.2: Gestures considered in Mudra.

enabling micro-gesture recognition. This technology enables finger control application in portable and wearable devices. Soli requires a dedicated RADAR-like chipset that needs to be integrated with the existing hardware. This motivated us to explore the following question: Can we detect fine gestures with conventional wireless signals, e.g.

WiFi? If this is possible then we won't need separate hardware for gesture recognition from communication. Shown in Fig. 3.1, we expect users to control their portable device with finger motion nearby.

Since Mudra utilizing WiFi signals, it just needs regular WiFi data transmissions either from local or remote sources. In this way, Mudra doesn't need to generate special signals (like Soli), thus, saving power for power-hungry devices [1, 49] and not hurting communication opportunities. Service of Mudra is always available as long as there are active transmissions in vicinity: Mudra can use a nearby WiFi-enabled desktop or laptop's transmissions to its access point (AP) to detect gestures. Furthermore, Mudra can use transmissions from multiple sources over time.

A key feature of Mudra is its training-free nature. This makes our system more promising in reality compared with other micro-motion technologies [17, 30, 64–66], all of which require training for specific location and user. One advantage is Mudra is not constrained by stationary scenario. That means, even if the user moves to another place, Mudra is still able to provide gesture recognition service using a WiFi transmitter at a new location without reconfiguration.

However, to meet these design goals, there are the following **challenges** in building Mudra:

- Compared with human body motion, finger gestures are micro-motion that involves much smaller objects (fingers), slow speed and short distance. Thus, finger gestures are hard to detect because they result in subtle changes in the wireless channel.

- To utilize regular WiFi signals, we confront with lots of critical issues. Firstly, WiFi packets are typically not continuous. We need to implement smart packet detect/capture module in Mudra. Second, short subcarrier length and variable packet lengths in WiFi limit the tracking precision.

- User's movement changes the channel between the source and receiver. This makes finger-gesture recognition hard to be reliable and training for every location is not favorable for a user-friendly system.

To solve the above-mentioned challenges, we propose three main design components in Mudra system:

—*Highly-sensitive Motion Indication.* To make our system robust against the small signal-to-noise-ratio (SNR), we propose a highly-sensitive motion indication.

—*Resource-friendly Signal Processing.* To enable the system functionality with regular WiFi signals, we design novel signal processing approach to solve the challenges. In this way, our system doesn't generate any dedicated wireless signal or take up extra spectrum resources.

—*Environment-agnostic Gesture Pattern.* To support the user-friendly feature, our system needs to require no specific training data from users per surrounding environment. To this goal, we design and validate a set of gestures which generate consistent and environment-agnostic patterns.

We build a prototype on NI based SDR platform. Our evaluation uses signals from COTS WiFi sources. We evaluate our system in various environmental conditions and system settings across multiple users. It shows that our system can achieve an average accuracy of 98% with a local source (2 cm from Mudra) and 96% with remote source(s) (0.5 - 7 m from Mudra). In the following sections, we will introduce the design of each component in detail.

3.2 Mudra Design

3.2.1 Preliminary: System Scenario and Setup

Before we go to the detail of each design component, in this section let us first understand the application scenario considered in this work. With Mudra, we are designing a system that enables micro-gesture recognition in vicinity. As Fig. 3.1 shows, the hand moves roughly 7cm to the receiving antennas over 2-4cm with gestures in Fig. 3.2. This distance is measured as the minimum distance between the hand and the two-antenna segment. On the target device, there are two antennas capturing incoming WiFi signals. As to the distance between them, it is a critical design choice and will be studied in Sec. 3.3.2, which gives 10cm as a default setting for the best sensitivity while noting that Mudra can also work with shorter distance (Sec. 3.3).

Mudra can work with a remote source such as desktop, tablet and AP, and also a local source, i.e. an extra transmit antenna on the target device itself. Thus, distance between signal source and Mudra receiver could be ranging from several centimeters to a dozen meters. The motivation behind utilizing existing signals is two folds. On the one hand, it eliminates the need for special transmissions for gesture recognition. On the other hand, including distant sources allows Mudra to utilize different sources over time and provide gesture recognition service throughout.

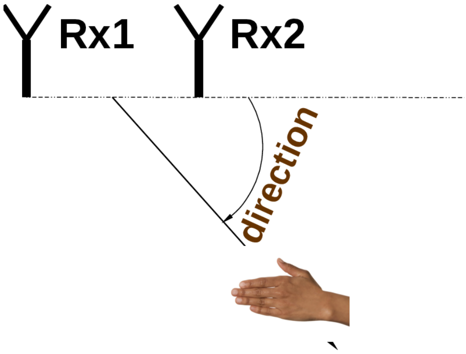


Figure 3.3: Direction Demonstration

Two directions relate to finger motion in this work. One is the hand direction which is measured as the angle relative to the line formed by two antennas, increasing clockwise (Fig. 3.3). The other is the finger moving direction. User experiences show that finger moving along forearm with gestures in Fig. 3.2 is comfortable and adopted naturally. Note that "moving along forearm" is a 2-D description as the 'shoot' rotates the wrist while making the fingers move along forearm in horizontal plane. Thus, we measure the forearm direction relative to the two-antenna-line as moving direction while telling the users to move along forearm. In Fig. 3.3, moving direction is 0 degree with forearm being parallel with the two-antenna-line.

3.2.2 Highly-sensitive Motion Indication

In order to track minor finger motion, what indicator does Mudra use that is extra-sensitive to finger position?

Channel Estimation and Signal Equalization

Mudra is equipped with two receiving antennas as shown in Fig. 3.1. Each of the antennas captures incoming signal samples, denoted as $s_1(t)$ and $s_2(t)$ respectively. The first thing is to equalize $s_1(t)$ and $s_2(t)$ for cancellation purpose. The solution here is to estimate the relative channel response between them.

To explain this, let two channel frequency responses be $H_1(w)$ ('w' is angular frequency) and $H_2(w)$ respectively, transmitted signal be $T(w)$ and two received signals be $S_1(w)$ and $S_2(w)$. Then, $S_1(w) = T(w) * H_1(w) = T(w) * H_2(w) * \frac{H_1(w)}{H_2(w)} = S_2(w) * \frac{H_1(w)}{H_2(w)}$. As it shows, signal S_2 is equalized to S_1 by compensating for the relative channel response $\frac{H_1(w)}{H_2(w)}$. We denote the time domain signal as $\overline{s}_2(t)$ after this equalization. In Mudra, $\frac{H_1(w)}{H_2(w)}$ is transformed to time domain impulse response then put into FIR filter as coefficients to equalize S_2.

In Mudra, since WiFi signals exhibit non-flat relative channel response for two receivers, we need to estimate $\frac{H_2(w)}{H_1(w)}$ across the whole band. This is totally different from [17] which estimates just amplitude/phase difference with the flat-response assumption. In Fig. 3.4, the result validates the effectiveness and necessity of full channel estimation.

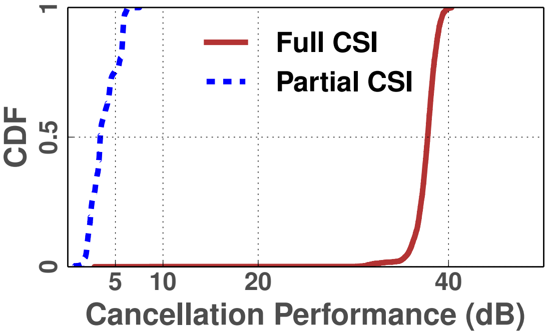


Figure 3.4: Cancellation Performance Comparison. We test in various scenarios. ‘Full CSI’ estimates on the whole band while ‘Partial CSI’ just calculates overall amplitude/phase difference.

Motion Indication with Signal Cancellation

After we obtain $\overline{s}_2(t)$ from equalization, $s_1(t)$ can be canceled from $\overline{s}_2(t)$. However, we want an indication which changes with finger moving. To this end, inspired by previous work [17], manual delay Δt and manual phase ϕ are introduced to $\overline{s}_2(t)$. Now, when finger moves, an extra delay $\delta\tau$ is introduced (Eq. 3.1a). Then, their frequency domain components ($S_1(w)$ and $\overline{S}_2(w)$) are related as shown in Eq. 3.1b.

$$s_1(t) = \overline{s}_2(t - \Delta t - \delta\tau)e^{-j\phi} \tag{3.1a}$$

$$S_1(w) = \overline{S}_2(w)e^{-jw(\Delta t+\delta\tau)}e^{-j\phi} \tag{3.1b}$$

Thus, as s_1 is subtracted from $\overline{s}_2$, the cancellation is perfect (i.e. zero value) only at specific frequency points where the phase shift is an integer multiple of 2π.

$$w^* = \frac{2\pi k - \phi}{\Delta t + \delta\tau}, \; k \in \mathbb{Z} \tag{3.2}$$

w^* is a set of frequencies where the cancellation will be perfect.

Since at frequencies farther away from w^* cancellation becomes worse (due to larger phase mismatch), we would get a trough at w^* shown in Fig. 3.6(a). Now, as the finger moves, $\delta\tau$ varies accordingly causing trough location to change. Taking 'Come' (Fig. 3.2(c)) for instance, when fingers move from Rx1 to Rx2 and then back from Rx2 to Rx1, $\delta\tau$ decreases first then increases to origin, which translates to increasing and decreasing of w^*. Then, we get waveform as shown in Fig. 3.8(c).

One could expect that a sharp trough gives us precise tracking, enabling gesture recognition with good performance. This relies both on: *perfect cancellation performance* (trough depth), which is guaranteed by our full CSI estimation, and *high sensitivity* of cancellation performance on phase mismatch (trough slope). We refer readers to [17] which has a study on sensitivity of cancellation degradation with phase mismatch.

How to select manual delay Δt and manual phase ϕ? - based on WiFi frequency structure.

In the OFDM physical layer of 802.11 protocol, there are 52 subcarriers over 20MHz band and no data at the center and edges (Fig. 3.5). Thus, if trough moves through these non-signal frequencies, we lose the track of finger motion. Mudra deals with this problem in two steps.

First, from Eq. 3.2, we can see the trough varying rate with motion drops with larger Δt. Thus, by setting a large enough Δt, we limit the varying range of trough.

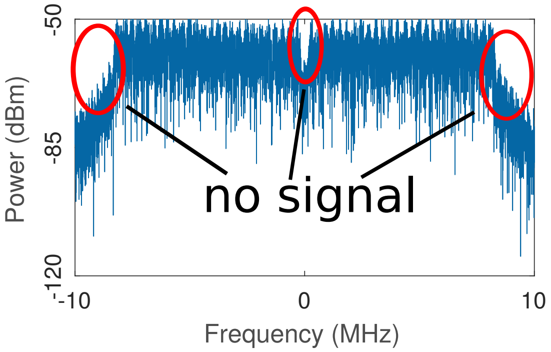


Figure 3.5: WiFi spectrum

Second, we naturally want to make use of the maximum continuous left or right half band. Thus, manual phase ϕ is chosen as to tune the initial trough location to the center of left or right half band.

3.2.3 Resource-friendly Signal Processing.

In order to utilize the prevalent commodity communication signal, e.g., WiFi, Mudra needs to answer various questions and address several issues regarding to the realistic WiFi signals.

Fewer subcarriers in WiFi and cyclic prefix (CP)

In 802.11 a/g/n protocol, an OFDM symbol has just 64 subcarriers in 20/40MHz band. This is a critical issue which is not discussed in [17] as they send special OFDM symbols with 8192 subcarriers continuously. If we just use 64 subcarriers, then the frequency resolution is too coarse to perform gesture recognition. Here, we note that WiFi signals have power over all frequency points in its band because OFDM symbols are different from each other (shown in Fig. 3.5). So we are safe to do a large size

FFT by simply connecting multiple OFDM symbols. However, do we need to avoid CP when combining multiple symbols together? The answer is no. In Mudra, signal copies on the receive antennas are for cancellation, not decoding. Thus, CP would not affect cancellation result as it also complies to channel response. Compared with WiSee [50], which equalized and combined OFDM symbols to enable fine frequency granularity, Mudra eliminates equalization because it relies on heterogeneity among those symbols.

Various packet lengths in WiFi signals

Packet length decides how large the FFT could be: Larger this window better the frequency granularity. WiFi transmissions could have variable packet lengths causing the granularity to change from packet to packet. In Mudra, we utilize two signal processing properties for this issue, Zero-fill Invariance and Connect Invariance.

- **Zero-filling Invariance**

When packet length is a little less than the expected FFT length, we want to maintain the same granularity. The approach here is to use zero-filling at the end. In Eq. 3.3, we show the zero-filling invariance property, which means zero-filling doesn't change the value on the same frequency. **DTFT:**

$$S(w) = \sum_{n=0}^{N-1} s[n]e^{-jwn} \tag{3.3}$$

This is Discrete Time Fourier Transform (DTFT). $s[n]$ can be seen as the original signal samples with any length of zero-filling at the end. $S(w)$ doesn't change with those filled zeros.

- **Connect Invariance**

Zero-filling just increases samples in frequency domain while not really contributing to signal spectrum. That means, although we get more values in frequency domain with zero-filling, the precision doesn't change because the trough is smoothed at the bottom. Thus, we have to drop those packets which may frequently appear in transmission, greatly sacrificing trough tracking opportunities.

However, we note the connect invariance property, which means trough location doesn't change if we combine residual signals of two neighboring packets. In Eq. 3.4, $r1, r2$ are cancellation residuals from adjacent packets, which are connected to get $r[n]$. **DTFT:**

$$\begin{aligned} R(w) &= \sum_{n=0}^{N+M-1} r[n]e^{-jwn} \\ &= \sum_{n=0}^{M-1} r1[n]e^{-jwn} + \sum_{n=0}^{N-1} r2[n]e^{-jw(n+M)} \\ &= R1(w) + e^{-jwM}R2(w) \end{aligned} \tag{3.4}$$

If $R1(w)$ and $R2(w)$ have trough at the same frequency, then $R(w)$ will also have trough at that frequency. Thus, by connecting neighboring packets after cancellation, we get the desired precision. Since motion is continuous in space, troughs within adjacent packets would share adjacent positions.

Noncontinuous transmissions of WiFi signals

Transmissions in WiFi are not continuous. For this, we implement a smart packet capture module with power threshold. Inter-packet interval would affect timing resolution of motion indications. When packet's interval is smaller, we get finer motion indications and thus, the recognition performance would be better. Fig. 3.15 shows experimental results on the effect of inter-packet interval on recognition performance.

Frequency/timing Offset, Multi-path Effect and Irritable Neighbor as well as Local Source

1. How does Mudra deal with frequency/timing offset with remote source?

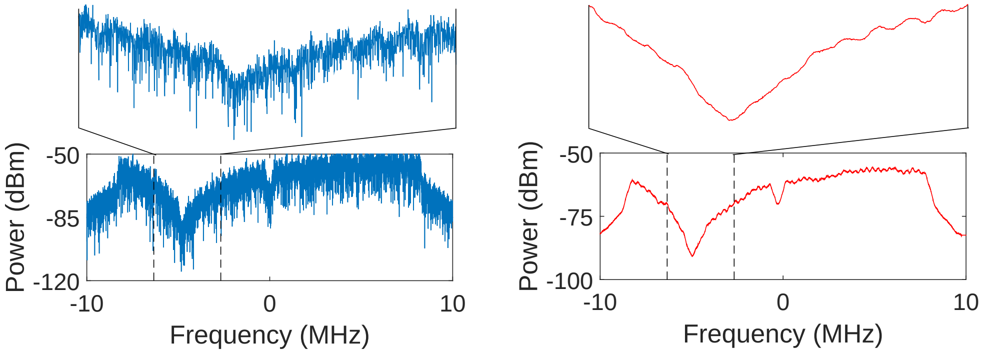
With remote source, frequency and timing offsets between transmitter and receiver ([15, 24, 51]) would impose a varying phase shift in the received signals, which can disable our system. In Mudra, however, two receiving antennas are synchronized with the same RF/sampling clock and timing offset as they are on the same device. Since our scheme obtains motion indication from cancellation of these two, we then get out of the phase shift problem.

2. Multi-path effect in received signals

In experiment, we find that multi-path effect would incur signal destruction at some frequency, causing a trough similar to the one from cancellation (shown in Fig. 3.6(a)). This trough will remain in the cancellation result, thus, confounding with the desired one. In Mudra, adaptive strategy is proposed to move desired trough out of confounding range. Specifically, we can detect multi-path trough from original received signals, the one without manual delay/phase. If it is in the left half band, we let the desired trough from cancellation be in the right half band by adjusting the manual phase (described earlier).

3. Irritable neighbor around trough

Fig. 3.6(a) shows that the spectrum after cancellation has irritable neighboring value around trough location, which are spikes in the magnified figure, seriously deteriorating the accuracy of trough location. We use neighboring average to remedy the problem and get a smoothed curve, shown in Fig. 3.6(b). The effectiveness and length choice of smoothing is discussed in the evaluation part, shown in Fig. 3.11(a).



(a) Original Spectrum

(b) Smoothed Spectrum

Figure 3.6: Canceled signal spectrum before and after smoothing over 200 neighbors. The left plot is the power spectrum across 20MHz generated by original samples and the right one is a smoothed version by averaging over adjacent 200 values. The trough in each of them is magnified in a subplot on the top.

Mudra with a local source With a signal source on the same device, transmitting and receiving antennas are very close. Thus, the signal power over direct LOS path is much larger than reflection from hands, which means finger motion has a minor effect on received signals. Besides, due to near-field effect [71], the received signals see much noisier variation in trough recording.

Solution of Mudra: In Mudra, we deal with this problem by adding a metal plate as an electromagnetic partition between sender and receiver antennas so as to cut off the direct-path signals. Similar approach is adopted in [22] to implement absorptive shielding.

We test the effectiveness of this approach by comparing signal-to-noise ratio (SNR) in different directions with and without metal partition. We define SNR as the ratio between gesture-waveform and noisy variation amplitude. Fig. 3.7 validates that our

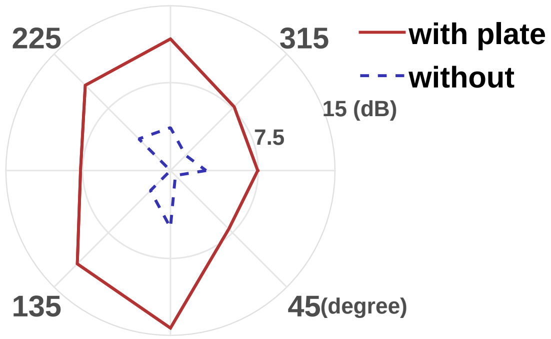


Figure 3.7: System SNR with and without iron plate. When signal source is very close (2cm), we test waveform-to-noise-amplitude ratio using 'Come' gesture in a circle in eight equi-angular directions.

method can increase SNR by 6 times on average. For fair comparison, finger moving direction is 0 for all positions.

3.2.4 Environment-agnostic Gesture Pattern.

One of the goals of Mudra is to perform gesture recognition without any training and even when the user has moved to another environment. In this way, our system is user-friendly in the aspect that it doesn't require per-user-per-environment training. There are two aspects in Mudra that enables it to support this feature:

- *Environment-agnostic.* The system is environment-agnostic means that the derived motion indication is not affected by the surrounding environments. That is, when the user moves from one room to another room, the trough location remains the same with the same position of the finger. This characteristic is realized with the relative-channel design of the motion indication, i.e., the motion indication is only affected by the relative channel between the hand the two receive antennas.

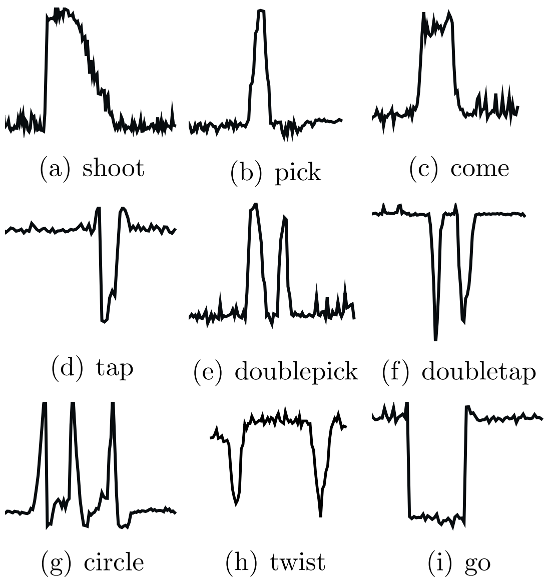
- *Reliable consistency.* However, maintaining the consistent motion indication doesn't ensure the consistency of gesture pattern from motion indication time series. The essential reason comes from the practical issues of the operating subject. On the one hand, different people perform the same gesture differently in the duration, speed and physical trajectory. One the other hand, human hand is not a point object but can be viewed as a collection of multiple point objects which reflect the signal altogether to the receiver. Thus, to ensure a consistent pattern, we need to make sure the moving trajectory of all point objects on the hand appears the same with regards to the two receive antennas.

In this subsection, we study how the trough variation over time can be used to identify different gestures and study how reliable and consistent the pattern is to the various settings.

Fig. 3.8 shows the trough location (in MHz) when plotted over time for different gestures (The axes are not shown due to space issues). Do these patterns remain the same across different environments? Do they change with different users?

When user moves, the environment around him would change. We care about different directions and distances of signal source in both LOS and nLOS scenarios. Thus, we choose eight equi-angular LOS positions with 2m distance and nLOS positions 5m away. For hand positions, we put hand at 0, 45 and 90 degree respectively. Note that we can ask the users to keep their hands at an angle that gives the best accuracy for all the finger-gestures. We use this best angle in our evaluations in Sec. 3.3.2.

Five users were asked to perform nine gestures at every source and hand position in a quiet office with no other people. We combine the results of all users in Table 3.1.



(a) shoot (b) pick (c) come

(d) tap (e) doublepick (f) doubletap

(g) circle (h) twist (i) go

Figure 3.8: Waveform of gestures considered in Mudra.

Position	**P**1	**P**2	**P**3	**P**4	**P**5	**P**6	**P**7	**P**8
0degree	+	⋆,-	+,⋆,-	+,⋆,-	⋆,-	+,⋆,-	+,⋆	+,⋆,-
45degree	+	+,⋆,-	+,⋆	+,⋆,-	+,⋆	+,⋆,-	⋆,-	⋆,-
90degree	+	+	+	+	+	+	+	+
Position	**P**9	**P**10	**P**11	**P**12	**P**13	**P**14	**P**15	**P**16
0degree	+,⋆	+,-	+,⋆,-	+,⋆	+,⋆	⋆,-	+,-	+,⋆,-
45degree	+,⋆,-	⋆,-	+,-	+,⋆,-	+,⋆,-	+,⋆	+	⋆,-
90degree	+	+	+	+	+	+	+	+

Table 3.1: Consistency study of gesture waveforms with various signal source locations.

P1-**P**16 denotes sixteen distinct source positions with direction increasing starting from 0 degree: **P**1-**P**8 are LOS while **P**9-**P**16 are nLOS. '+' means the waveform shape is same or similar with the same gesture in first position **P**1 while '-' represents direction flip of waveform shape. '⋆' is to say the corresponding waveform shape has no similarity with **P**1. In other words, '+' or '-' is good since the shape of the pattern is retained. However, '$\star$' indicates that the pattern has changed and denotes a case against Mudra.

The results in Table 3.1 show that hand position in 90 degree could ensure consistency in gesture waveforms (Fig. 3.8) across all scenarios. While this is not surprising, as we know, there exists a linear relationship between signal phase and delay, and the reflection from fingers is a combination from all point objects. In later experiments, we will adopt this setting as default. In addition, clustering approach in data mining [**?**] might also be useful to automatically group gesture segments without explicit waveform design, potentially increasing gesture classes.

3.2.5 System Model

Mudra has five main components (Fig. 3.9):

A. Packet detect/connect: This component captures packets from WiFi on two antennas. Since power level is not equal over samples, envelop of power trace is used for finding the beginning and end of continuous packet with a threshold. Mudra will discard packets shorter than 1000 samples to ensure a reasonable trough locating confidence. After that, neighboring packets will be connected together if packet length is less than 8192.

B. Channel estimator/signal equilization: As shown in Fig. 3.9 block **B**, channel estimator gets relative channel response between the two receiving antennas. Then, this coefficient is sent to an FIR filter, which equalizes the samples on the second antenna.

C. Manual delay/phase injector: As mentioned in Sec. 3.2.3 and 3.2.3, manual phase/delay is used to address WiFi frequency structure and multi-path issues. An

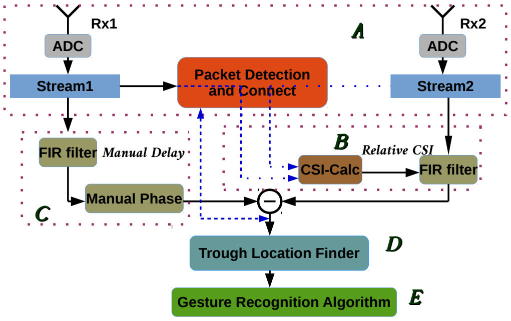


Figure 3.9: System model

FIR filter is used to inject manual delay to the captured packets followed by a manual phase addition.

D. Trough location finder: After cancellation, residual signal will be searched in the selected band and the minimum-value (frequency) position is the trough location. Then, timing series of trough location would be sent to gesture detection/recognition module.

E. Gesture recognition algorithm: Gesture recognition module comprises of preprocessing, segmentation and classification. We implements online recognition with four gestures using simplified classification logic. To classify all nine gestures, we conduct gesture recognition off-line.

Design detail

Here, we talk about how to choose design parameters.

1. Length of FFT

Length of FFT decides the granularity we can look into in frequency domain. In 802.11 a/g/n protocol, the maximum size of PLCP Protocol Data Unit (PPDU) in physical layer is larger than 4000 bytes. Considering that one byte has 8 bits and

64-QAM modulation has 6 bits in one sample, we choose 8192 as the number of frequency points we look into, which is the length of FFT operation.

2. Manual delay $\Delta\tau$ and manual phase ϕ

As mentioned in Sec. 3.2.3, we tune the manual delay $\Delta\tau$ to avoid trough running out of the selected range. After testing all target gestures by five testers, we choose 1, which means skipping one sample, as $\Delta\tau$ to achieve largest sensitivity while satisfying this restriction, ensuring that it is feasible in almost all situations. After that, we move the trough location to the center of the range by choosing a value for ϕ. Note that ϕ is fixed after $\Delta\tau$ is decided.

Authorization Key for System Access

To avoid target device being miscontrolled by surrounding moving objects, we designed an authorization key containing multiple 'Tap' gestures as preample for users to access this system. We have test in Sec. 3.3.3 to evaluate the effectiveness of different gesture repetitions.

3.3 Evaluation

3.3.1 Implementation

We implement Mudra receiver system on the NI-based SDR platform. NI PXIe-1082 chassis is equipped with PXIe-8133 Express Controller and two NI-5791 FlexRIO adapters. Each adapter is connected with a VERT2450 3dBi gain antenna. In Virtex-5 based FPGA, DSP decimation after ADC with resolution of 14 bits generates 20MHz baseband samples. Direct Memory Access is built to transfer baseband samples from Rx1 to Rx2. Central controller is built on RTOS based PXIe-8133. We implement packet detect/connect and channel estimation on central controller.

In the FPGA of Rx2, we implement manual delay/phase injector on stream from Rx1, shown in system model Fig. 3.9. To equalize signal, samples from Rx2 are thrown into FIR filter with relative channel timing response as coefficients. On the central controller, two signal streams and cancellation stream are fetched from host FIFOs. Then, we use the captured packets to calculate relative channel response, sent to FIR filter in second stream pipeline with FPGA Module register map. Motion indication trace is generated online and stored to file along with timestamp indicating packet arrival time.

We design parallel architecture on central controller to solve computation challenge. First of all, we use multiple single-time ***while*** loops to distribute computation on different CPU cores. To avoid overflow of data transmission from FPGA to central controller, Mudra has a special loop assigned with highest priority dedicated for fetching data from host FIFO. Further, to maintain packet order, queue structure is used to pass data from one loop to next, forming producer-consumer structure.

We use TP-LINK TL-WN722N adapter as signal source, with output power peak at 17.8 dBm [6], and NETGEAR WNDA3100 adapter as WiFi receiver. We build an

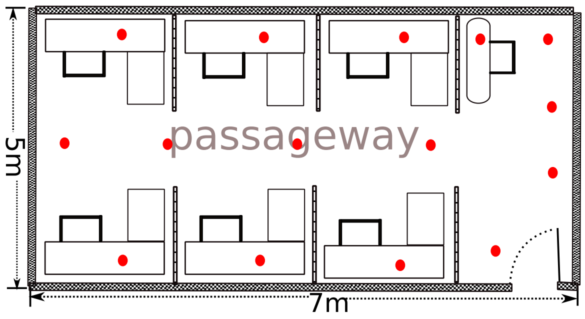
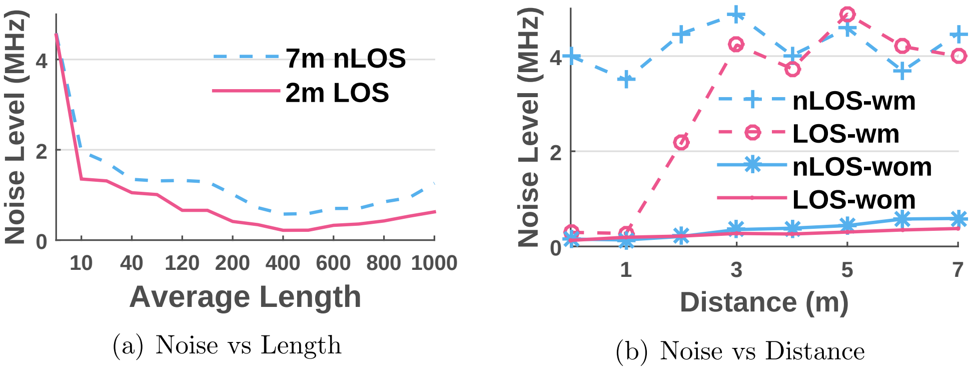


Figure 3.10: Floor plan



(a) Noise vs Length

(b) Noise vs Distance

Figure 3.11: Noise Variation: 'wom' is short for "without body movement" and 'wm' is short for "with body movement" between the source and receiver.

adhoc connection between these two and enable UDP stream on Windows7 system. Fig. 3.12 shows our testbed.

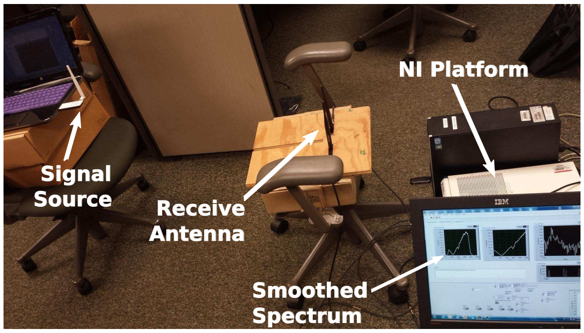


Figure 3.12: Testbed

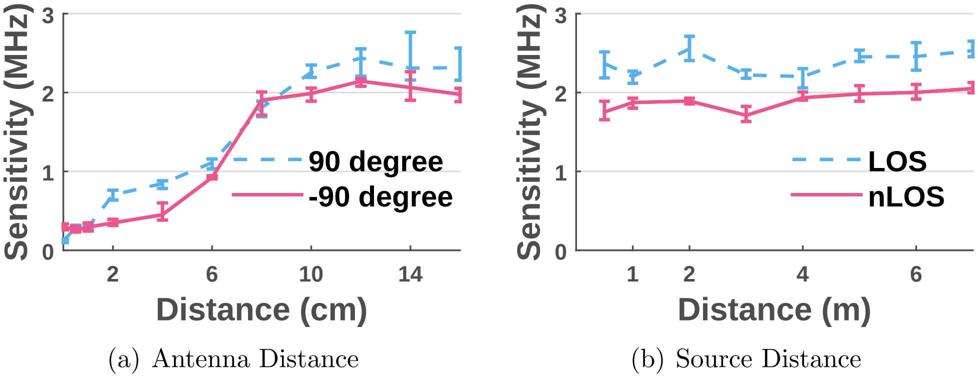
3.3.2 Micro-Benchmarks

We set out to look into various performance-affecting factors. First, what is the noise variation level and how it varies with changing environment. Apart from this, we also care about system sensitivity, distance between the source and the receiver and orientation. After that, impact of packet length/interval will be studied.

Noise variation

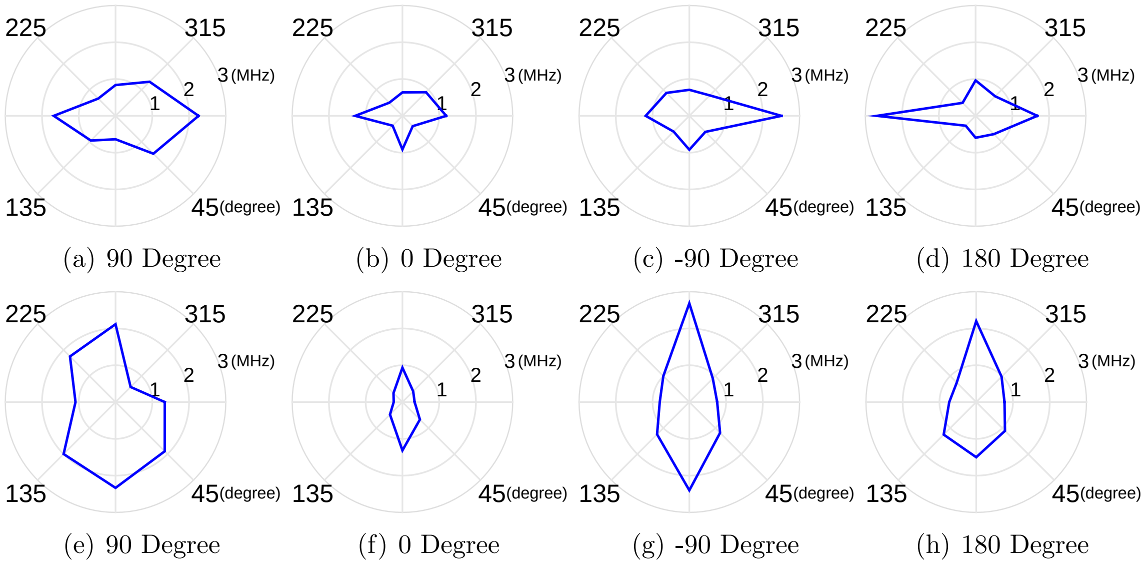
The irritable neighbor problem introduces severe noise in trough tracking, as mentioned in Sec. 3.2.3. To study this effect, we vary the distance between two antennas in Mudra. In a quiet office with door closed (floor plan shown in Fig. 3.10), we put source 2m away in LOS then 7m away in nLOS at 90 degree to receiver. For fair comparison, we just let one user to be around receiver statically. We measure indication varying range shown in Fig. 3.11(a). We observe that the average length (for smoothing) larger than 400 samples doesn't reduce noise further. Thus, we select 400 as the default setting. We also find that noise gets even worse with more than 500. This is expected since averaging also makes the trough smoother.

Noise and environmental variation (e.g. human breath) would also induce noise variation. For this test, we choose different distances between signal source and receiver. To study the effect of moving body, another tester will move 3m away from receiver, blocking direct-path signals at some points. As indicated in Fig. 3.11(b), the noise increases with distance with no moving body. While, with moving body, noise gets extremely serious, disabling our system: We show how such scenarios can be detected and the corresponding samples can be avoided while performing gesture recognition in Mudra later. Another observation is that, with direct-path signal, moving body keeping 1m away from LOS generates a tolerable noise, while nLOS scenario doesn't share this feature.



(a) Antenna Distance (b) Source Distance

Figure 3.13: Sensitivity Study. **Takeaway**: Sensitivity of Mudra changes with antenna distance while keeping stable with source distance.



(a) 90 Degree (b) 0 Degree (c) -90 Degree (d) 180 Degree

(e) 90 Degree (f) 0 Degree (g) -90 Degree (h) 180 Degree

Figure 3.14: Orientation Effect. Upper four plots study moving direction; bottom ones study hand position.

Sensitivity

We evaluate sensitivity of our system using WiFi signal in different scenarios. First, we study the factor of distance, including antenna distance and source distance.

Then, we explore the impact of orientation i.e. hand/source direction and finger moving direction. We ask the users to perform ‘Come’ with 6 rounds. For each round, we obtain sensitivity as the average amplitude of (trough location) waveforms.

(a) Versus distance

To study antenna distance, we put the source 2m away at 90 degree and -90 degree in LOS. The user is asked to perform 'Come' moving in 0 degree i.e. along two antennas with left hand position of 90 degree. Then, we gradually change distance between two antennas for each test. Fig. 3.13(a) indicates that sensitivity increases with antenna distance and stops at about 10cm. Considering that 10cm is a favorable scale for most portable devices, we select 10cm as the default setting throughout this work.

After that, we continue to look into source distance. This time, we let the source be at 90 degree. We adjust the distance between source and receiver both in LOS and nLOS. To fairly compare sensitivity, we let user vary hand position and moving direction in eight equi-angular orientations, choosing the largest one. Fig. 3.13(b) shows that sensitivity with LOS is greater than nLOS. To our surprise, sensitivity almost doesn't change with distance in both scenarios. We believe the reason is that sensitivity is mainly affected by distance between hand and receiver. Those results, combined with study on noise variation (Sec. 3.3.2), also validates the feasibility of finger-gesture recognition using WiFi signals.

(b) Versus orientation

We first study the effect of finger moving direction. To this end, the user moves his fingers in eight equi-angular directions. We test with four equi-angular source positions with a distance of 2m away from receiver. For example, Fig. 3.14(a) represents a source at 90 degree. Hand is at 90 degree. As stated before, all directions are relative to two antennas in clockwise mode. In Fig. 3.14(a)-(d), we observe that moving in 0 degree guarantees at least 1MHz sensitivity for all source positions. Besides, moving in 180 degree provides comparable overall sensitivity as in 0 degree. This is expected since both directions are along two antennas. Then, we seek to evaluate the effect of hand position. We put hand in eight equi-angular positions. This time, however, the user tries to move in all directions and selects the largest one as sensitivity measurement. For example, in Fig. 3.14(e), with hand position at 90 degree, moving in 0 degree generates largest waveform amplitude. Fig. 3.14(e)-(h) show that putting hand vertical to antenna, in all eight positions, has the best overall sensitivity. We have roughly the same observations across other gestures, thus, pushing us to select 90 degree for hand position and 0 degree for moving direction, which means, fingers move along two antennas with hand at vertical position. With such a default setting, combing with waveform consistency study in Sec. 3.2.4, we desire to make our system feasible with best recognition performance.

Packet length/interval

We run experiments to understand the effects of these parameters on recognition performance. Specifically, we setup an ad hoc connection between two wireless adapters and enable UDP streaming on sender side. With socket programming, we

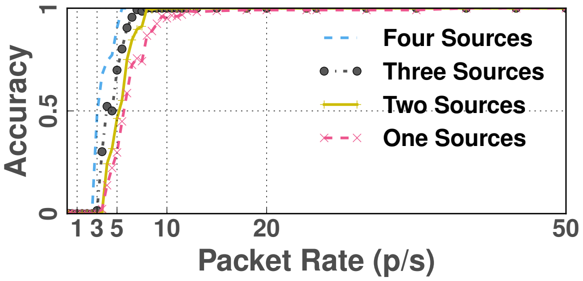


Figure 3.15: Impact of packet interval.

can adjust the length of character string which is sent as a bunch. On both the SDR platform and Wireshark packet capturer, we validate the packet size. Inspired by this, we adjusted packet rate by tuning the while loop interval. Unfortunately, the maximum frequency we get is around 100 packets/second. If we set a lower delay, packet rate increased drastically in a burst mode, which means large portion of duration has no transmission. While, a good thing is that our system already has favorable classification accuracy with 50 packets/second. Thus, such a limitation doesn't affect our evaluation. Signal source is put 2m away at 90 degree position. We have five users perform four gestures - 'Come, Go, Tap, Pick,' each with 50 repetitions. We implement a simplified algorithm on SDR platform to recognize these gestures in real time. Specifically, a threshold is applied to derivation of the indication trace after reducing noise, where boundaries of gesture is detected. Then, we get gesture duration using the attached timestamp. Since 'Come, Go' last much longer than 'Tap' and 'Pick' (2s vs 0.5s), we can distinguish between them. Further, by utilizing the waveform direction, we come to the final decision by judging whether the first derivation is positive or negative.

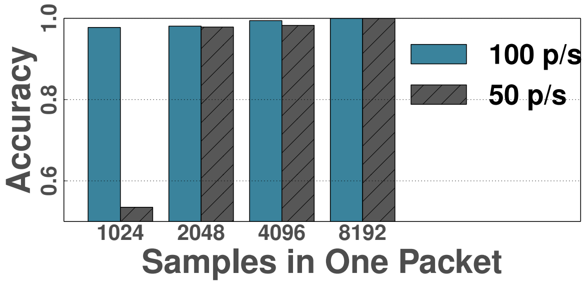


Figure 3.16: Impact of packet length.

To evaluate interval effect, we transmit packets at different rates. Fig. 3.15 shows how classification accuracy changes with packet rate. Let's first look at the resulting line with just one signal source. Here, accuracy is as high as 0.95 with just 10 packets/second - typical WiFi beacon rate, occupying 0.4% of the whole duration. This result means that our system just needs very conservative WiFi signals. We note that no gesture can be detected with rate less than 3 packets/second. This is because such low rate would miss gesture waveform. Then, we apply multiple signal sources and combine results. There are four sources located in four directions. The results indicate that with more signal sources we can get better performance. One reason is that with more sources we reduce the rate of missing detection. Another reason is that with majority voting, large errors with specific source could be avoided. Typically, with four sources, accuracy could be 0.95 same as one source while with just 5 packets/second per source.

As to the packet length, we evaluate recognition performance with four different choices. Packet rate is set to 100 packets/second and 50 packets/second. In Fig. 3.16, the results indicate that by connecting short packets, Mudra could be effective in gesture recognition. We note, compared with Fig. 3.15, that the performance is a little worse. This is because connecting packets not only reduces indication frequency but also degrades indication precision due to averaging over multiple packets. The drastic drop with 1024 at 50 p/s indicates the minimum indication length needed for good performance.

3.3.3 System Evaluation

False Positive

To avoid mis-detection with surrounding moving object, Mudra use multiple 'Tap' as authorization key. We evaluate its effectiveness in this section.

In our experiment, we put two antennas on one desk so that it can be affected by human regular activities. All members are free to do anything, walking, speaking or eating, in the office. We have four position settings which put source 5m away from receiver. Recording last 24 hours over a whole weekday from 8am to 8am, we get the result of false detection number per hour in Fig. 3.17.

The results show that when one 'Tap' is used as access key, the false detection rate over 24-hour period is 6.92 per hour. We note that this is very low. This is because 'Tap' lasts just about 0.5s, while typical human movements such as walking, eating and typing, last much longer. Besides, those large-scale motion usually push trough out of legal range. Lastly, 'Tap' waveform is special which is not easy for random motion and environmental variation to generate. We also find that with multiple repetitive gestures as the key, the false rate decreases greatly. It is reasonable since we expect gestures in one key would be in a certain range. In this way, for two repetitive 'Tap' false rate is 1.37 per hour and using three 'Tap' makes it as low as 0.08.

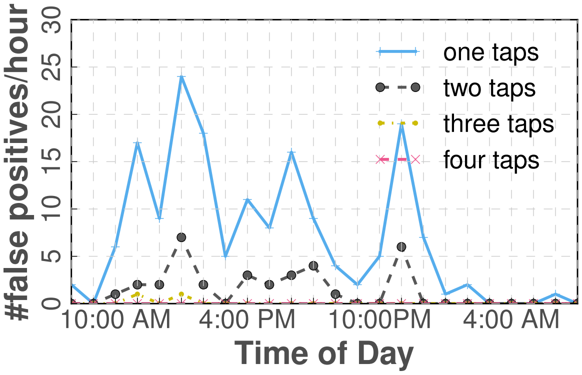
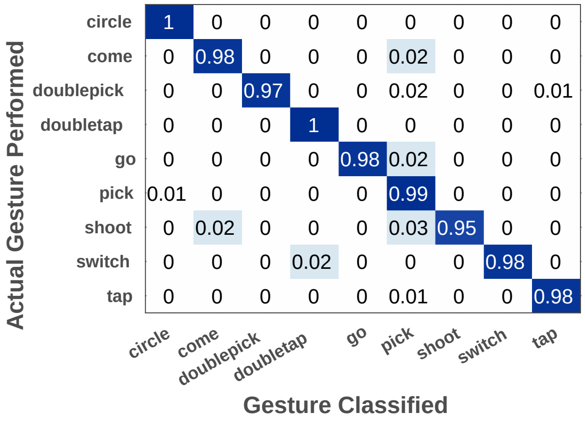


Figure 3.17: False detection rate during a whole day.

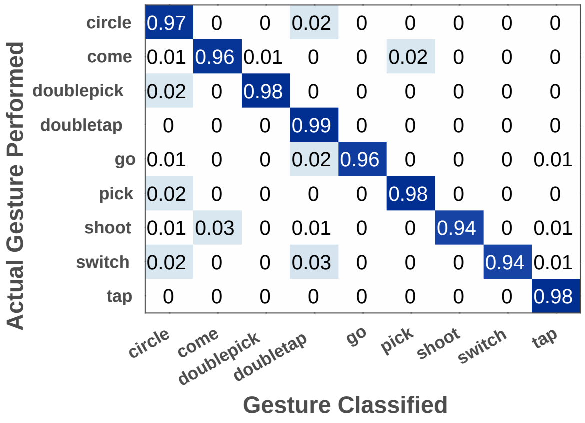
Mudra Recognition Performance

(a) Evaluation with Local Source: Mudra can utilize signals from the target device with solution in Sec. 3.2.3. We put source (WiFi adapter) 2cm away from the first antenna along the line with iron plate to block the direct path. Receiver is put on desks with surrounding strong reflective objects and empty space near the door shown in Fig. 3.10. Five users, who don't know how Mudra works, are shown how to perform each gesture. Each gesture in Fig. 3.2 is performed for 40 times by each user, following regulation of position and direction. During test, users sit in a chair. As to WiFi signals, we adjust packet rate to 50 per second with length of 4096, which means, indication after packet-connect would be 25 per second. We measure recognition accuracy combining detection and classification across all positions and users. Fig. 3.18 shows the average recognition accuracy is 98%. In this matrix, we observe that recognition has the largest error with 'shoot.' Checking the intermediate result, we find that waveform of 'shoot' is the most variable one across users.



	circle	come	doublepick	doubletap	go	pick	shoot	switch	tap
circle	1	0	0	0	0	0	0	0	0
come	0	0.98	0	0	0	0.02	0	0	0
doublepick	0	0	0.97	0	0	0.02	0	0	0.01
doubletap	0	0	0	1	0	0	0	0	0
go	0	0	0	0	0.98	0.02	0	0	0
pick	0.01	0	0	0	0	0.99	0	0	0
shoot	0	0.02	0	0	0	0.03	0.95	0	0
switch	0	0	0	0.02	0	0	0	0.98	0
tap	0	0	0	0	0	0.01	0	0	0.98

Figure 3.18: Confusion Matrix with Local Source.



	circle	come	doublepick	doubletap	go	pick	shoot	switch	tap
circle	0.97	0	0	0.02	0	0	0	0	0
come	0.01	0.96	0.01	0	0	0.02	0	0	0
doublepick	0.02	0	0.98	0	0	0	0	0	0
doubletap	0	0	0	0.99	0	0	0	0	0
go	0.01	0	0	0.02	0.96	0	0	0	0.01
pick	0.02	0	0	0	0	0.98	0	0	0
shoot	0.01	0.03	0	0.01	0	0	0.94	0	0.01
switch	0.02	0	0	0.03	0	0	0	0.94	0.01
tap	0	0	0	0	0	0	0	0	0.98

Figure 3.19: Confusion Matrix with Remote Source.

(b) Evaluation with Remote Source: To evaluate system performance with signals from remote device, we put transmitting WiFi adapter across positions in the office (dots in Fig. 3.10). With receiver on the walkway and on desk, we include scenarios with distances ranging from 0.5m to 7m both in LOS and nLOS. Each gesture is performed a total of 40 times. As Fig. 3.19 shows, the average accuracy is 96%. **Takeaway:** Mudra can deliver high accuracy with sources less than 7m away under quiet environment. The performance is worse than local source since larger distance can introduce more noisy variation. After that, we combine results of local and remote sources to look into performance with individual user (shown in Fig. ??): The accuracy is consistent across users showing that Mudra is generic and doesn't need training.

- ***Hand-held Device and Multi-source Combining:***

For portable devices, e.g. smartphone and tablet, people usually hold devices in hand when using them. It is a good utility if our system can support gesture recognition in this scenario. For this exercise, we bundle two antennas with a stick at a distance of 10cm. Users were asked to hold the stick with right hand statically and perform gestures with left hand. To test in the worst case, we put source at distances of 6m and 7m in four nLOS locations. Observation from this test is larger noisy variations due to hand-holding. However, the result shows that average recognition accuracy reaches 90% since noise is still tolerable with users sitting in a chair.

Further, to explore how multiple sources could help in such scenario, we combine recognition with traces from those sources. As a result, the overall accuracy is improved to 96%. It validates the effectiveness of utilizing multiple sources.

- ***Operating with Other Users:***

To test whether multiple users can perform gestures at the same time, we let one user operate on our system with other people nearby, also performing finger gestures. The interfering users are asked to be 2m away while facing to receiver. Also, we put the source across various locations. With three users nearby performing gestures, recognition accuracy is 94%. Thus, we are glad to show multiple users can operate simultaneously in the same room.

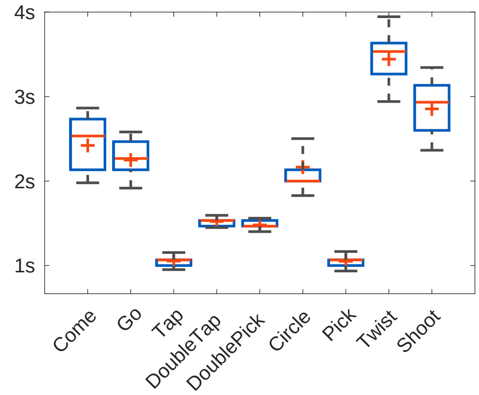


Figure 3.20: Variation range of gesture duration across users. *The box represents the quantiles while the whisker means the standard deviation. The red line is the mean while the red plus is the median.*

- ***Experiment with a larger diversity of users:***

To examine the performance with a larger group of users, we conduct extended experiments with six more users, including male and female. We use local source for the experiments. Each users are asked to conduct each gestures for about 10 rounds. Totally, we have 500 more test samples of gesture segments. As expected, we observe a larger diversity in the gesture waveform. In Fig.3.20, the plots show the variation range of the duration of each gesture across these users. In addition, with the larger diversity, we obtain a decrease in the recognition performance which is 91.3%.

However, we observe that the variation of the gesture waveform of one specific user is smaller than the variation across all users. It inspires us to learn the waveform pattern for each user individually to improve the performance. Specifically, we take the average duration as the metric to learn for each user and adapt our waveform template to the specific duration for each user. As such, the overall performance is enhanced to 94.9%.

- ***Compare with computer-vision technique:***

Computer-vision technologies have been popular in information discovery in the scene. However, to obtain a good performance, such approaches highly depend on high quality video capture. Thus, they are limited to light condition, background scene and target appearance. We conduct experiment to test the computer-vision approaches on our finger gesture recognition under various conditions. Specifically, we compare the results of pure-color and natural backgrounds, bright and dark lighting conditions, and glove and non-glove settings. Glove setting means the hand wears a pure-color glove and thus is easier for the video processing to distinguish and segment from the background.

We adopt a existing workflow [43] with open-source libraries to build the video processing model. The program uses tflearn of tensorflow to build two layers of recognition system. In the first layer, an inception model is trained as CNN network to extract spatial features from each frame of the video. In the second layer, an RNN model is trained to extract temporal features and learn dependencies between current and previous features. The video is captured at 30 FPS and typically last for 2 to 4 seconds for each gesture. We separate the training and testing dataset with a rate of 4:1 and obtain the performance with cross validation. The results are shown in Fig.3.21. From the figure, we can see that those factors affect the performance of the gesture recognition greatly. Compared with pure background, a nature background setting decreases the overall recognition accuracy by 27%. Besides, with bad lighting condition, the dark environment essentially disables the recognition as the accuracy is almost equal to random guess ($1/9 = 11.1\%$). Lastly, without a pure appearance of the object, the non-glove setting degrades the performance by 53 percent. Fortunately, the WiFi signal based approach is free from those limitations as it doesn't rely on visible light.

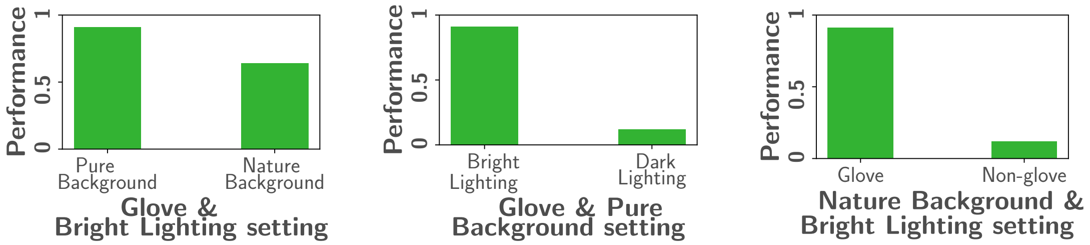


Figure 3.21: Performance of gesture recognition with CNN-RNN model in various environment settings.

Apart from recognition performance, we also compare the two approaches on the computation cost. For fair comparison, we run two programs in the same laptop, i.e., HP EliteBook with 7.7 GB memory and 2.9 GHz i7 cores, and record metrics such as disk storage, running time and memory cost.

Metric	Mudra	CNN-RNN model
Storage (KB/per gesture)	0.48	140 – 300
Time (s/per gesture)	0.085	train 82.4 test 44.9
Memory (MB)	0.4	394.2

Table 3.2: Various computation costs w.r.t two approaches.

In Fig.3.2, we show the metric values when running these two algorithms. For each gesture segment, the longest segment has around 60 double-value points which take storage of 0.48 KB with 8 Bytes per number. While, with gesture duration ranges from 2s to 4s, the video size is about 140 KB – 300 KB. To evaluate the time cost, we average over all gestures with the total running time for pattern recognition processing. As to video-processing model, we calculate the training time as the summation over training CNN model, extracting spatial feature and training RNN model. The testing time is the summation over extracting spatial feature with inception model and testing RNN model. Lastly, the memory cost is the peak memory used during the program running.

As we can see from the results (Table 3.2), the computer vision based approach takes up much larger storage due to the large size of video (we actually use the lowest video capture resolution 320x240). Besides, due to the large size of the model in video processing, each stage of running the model also takes up much more memory and time for each gesture both in off-line training and on-line testing.

3.4 Discussion

(a) Human interference

Surrounding human motion will incur great variation as shown in noise study Sec. 3.3.2. However, we also find that when moving object is far from direct path between source and receiver, noise level could be tolerant for gesture recognition. In personal spaces, such as office and living room, such requirements can be satisfied, i.e. people are separated by individual rooms with laptop/desktop nearby.

In the open space with nearby moving objects, the system cannot work well since body motion causes much stronger variation than finger motion in the received signal. In this case, a directional-antenna design or beam-forming could possibly control the path of the signal thus limiting the impact of surrounding interference source.

(b) Limited diversity of users

We also note that participants in the experiment have limited diversity. The ages of users in the experiments range from 16 to 25. Generally, the young children and elderly exhibits a larger difficulty to follow instructions to perform the gestures. Therefore, our system may see a degradation of performance across people out of the above age range. We expect to extend our experiments in the future and improve the system across a larger diversity of users by understanding the challenges with the young children and elderly.

(c) Constraints on external traffic

The resource-friendly processing in Mudra is not perfect. Although it solves the issues of various packet lengths, in practice we drops packets with too short lengths to preserve the system's accuracy. That is because too short packets need to be connected with a large number to infer the correct frequency resolution. Therefore, the resulting trough location is the average across much longer duration with very-low precision. In practice, we drops packets with length less than 1024 samples.

(d) Limited number of gestures

In this work, we design nine gestures while more gestures are possible. However, since we utilize the waveform shape pattern which is derived from the relative delay, more gestures would require a design of a larger number of delay variation patterns. In this case, it is expected that the distance between gesture waveforms is reduced and larger error will be in the recognition system.

(e) User's privacy

There may be a concern on user's privacy with a fine-grained gesture recognition system. However, our system operates in near-human-to-machine scenario which requires the finger motion to be close to the receivers. Therefore, the sensing technology developed in the work is not usable by a remote malicious eavesdropper with just passively receive antennas.To hack the information of the gesture recognition results, the hacker first needs to install malware on the users' phone to sniff the packet's information. As our gesture system performs like in-air interaction with the machine, we could adopt a similar security mechanism as the keyboard to protect the input from sniffing.

3.5 Summary

In this work, we take the first step towards designing a finger-gesture recognition system with WiFi signals and validate our design with COTS WiFi sources. Our system is user-friendly with no requirement of training. We believe such a technology would stimulate new applications, such as in-air interaction, untouchable-device control and disabled-friendly design. With both on-target local source and remote source, Mudra could support high recognition accuracy. While, our system could not support scenario with moving body nearby, we showed how such scenarios can be detected and avoided during gesture recognition.

Chapter 4: TifWiFi: Two-profile Integration Framework for Device-free Activities Recognition with Communication Signal

4.1 Overview

In the past decades, researchers have explored various techniques to achieve human activity recognition, such as camera-based [10], radar-based [2] and electronic wearable devices [5, 31, 80]. Camera-based approaches are restricted to line-of-sight (LoS) areas and require a good light condition. Also, the abundant image information will potentially threat users' privacy. Low-cost radar system also suffers the high directionality and a limited coverage (tens of centimeters). By attaching devices on the user's body, researchers can infer the activity he/she engages in by analyzing data from sensors like accelerator or gyroscope. However, attached sensors are neither desirable nor available in most applications. In contrast, WiFi devices provide the opportunity to achieve a low-cost system as well as get rid of the above limitations and security concerns. Without a doubt, WiFi signal has attracted researchers for designing the activity recognition system.

Limitations of prior works. The principle of activity recognition using WiFi signal is that different human activities would introduce different multipath distortions on wireless propagation. WiSee [50] infers the human gestures by looking into the Doppler shift and extracts the direction of body motion. This system relies on a special hardware design with USRP [73] to obtain the fine-grained resolution in the frequency domain. Like data-drive prediction [83, 84], E-eye [70] uses the channel state information (CSI) to build location-activity profiles. As different activities cause different CSI distributions, this system identifies the activity by comparing the amplitude histogram on each subcarrier. In contrast, another line of approaches look into the variation in CSI values caused by human motion. CARM [67] extracts the speed information from CSI-speed model. WiFinger [60] identifies the CSI variation patterns caused by close gestures.

Although the above techniques have demonstrated the feasibility, we note that all those approaches fall into the limitation of utilizing only one facet of the human activity, i.e., either the status or the motion. To be specific, the status here means the condition of the target such as position, orientation, and posture which introduce a certain CSI. Then, the collection of the status during the activities makes up the statistics of CSI values. While the motion means how that person moves his/her body, including body part, speed, and direction. With two observations, we find that both facets are critical in identifying the activity. On the one hand, two activities performed with the same location and orientation may introduce similar statistics of CSI values. In this case, the motion facet of the activity is the key to distinguish between them. On the other hand, two activities performed at different places may also have similar motions. Then, the status turns to be important. We will show later how these two cases make the current approaches fail. Thus, by analyzing both the status and the motion facets, the recognition system can achieve better performance.

Proposed approach. In this work, we propose TifWiFi , a two-profile integration framework for device-free human activity recognition. TifWiFi is built upon existing WiFi devices and thus removes the cost and burden of deployment. It is also a passive detection system with no privacy concern for users. The basic idea behind TifWiFi is to utilize status and motion of human activity. Specifically, the status of the activity is captured by a multipath profile, which represents the multipath propagation condition. The profile for the motion part is represented by a frequency domain speed model. As far as we know, TifWiFi is the first to utilize both profiles in the literature. It is worth noting that TifWiFi doesn't require any extra information source to utilize construct two profiles. The reason is that they are constructed with different processing on the CSI data.

Technical challenges. In designing TifWiFi, we first solve two challenges in the multipath-profile analysis and motion-profile analysis. In multipath-profile analysis, the conventional subcarrier-level distribution approach (E-eye [70]) causes a large error between different activities with close human status, i.e., position and orientation, which is validated in preliminary study Sec.4.2.1 and evaluation Sec.4.3.5. To combat this error, our solution is to explore the distribution of the CSI as a whole vector during the activity. The insight here is that one subcarrier contains much less useful channel information compared with the whole CSI vector and the diversity across CSI subcarriers also embeds important information about the multipath channel. However, it is also challenging to analyze high-dimensional CSI statistics. We will present our dual-statistics analysis model in Sec.4.2.2.

In the motion-profile analysis, current works only focus on well-defined human gestures such as punch, falling and push. The intrinsic pattern matching over time in these systems implicitly requires the motion to be in a specific sequence of speeds. Therefore, it is necessary to design new analyzing approach on motion profile for loosely-defined human activities. Lastly, to integrate both profiles, a simple strategy could only achieve the average performance. In this work, we make use of the insight on individual strength of each profile to design the integration strategy to enhance the overall performance.

There are three main components in the design of our system:

—For the multipath-profile analysis, we propose a dual-statistics scheme to enhance the recognition performance. To achieve the dual-statistics analysis on high-dimensional CSI data, we use a bi-directional RNN model to handle the challenge.

—As for the motion-profile analysis, we propose a statistical analysis on motion profile. After comparing with alternative approaches, we utilize the motion intensity as the metric to differentiate different activities.

—We propose the whole TifWiFi system design with the integration on two profile analysis.

To evaluate our system, we implement TifWiFi on commodity WiFi cards and personal computers. With extensive experiments in two typical environments, we shows both various benchmarks and holistic performance of TifWiFi and compare with alternative approaches.

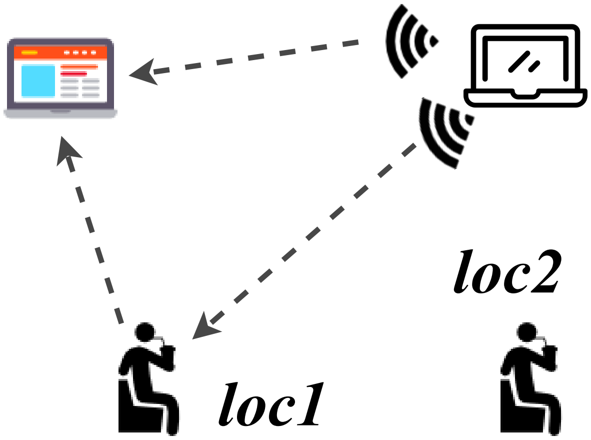


Figure 4.1: Experiment scenario for preliminary study.

4.2 TifWiFi Design

4.2.1 Insufficiency of Single Profile

In last section, the hypothesis is that both status and motion facets of the activity are critical for identifying the activity. To validate this, we conduct the experiment in two points of view. First, different activities performed with the same position and orientation have similar distributions of CSI values. Next, different activities performed at different places may also have similar motions, i.e., moving speed, of body parts.

In a typical office, we put two laptops equipped with WiFi NICs, with one on the desk and another on the ground, separated by 3 meters. While one laptop is the transmitter, the other runs as the receiver and records the CSI values during the experiment. The participant is guided to sit on one chair and conducts multiple times of two activities; one is drinking water and the other is spine-stretch (i.e., an exercise to relieve the muscle tension). Then, he switched to another chair and did

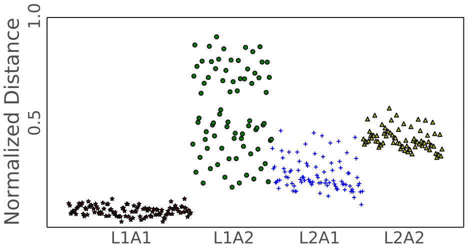


Figure 4.2: Drink.

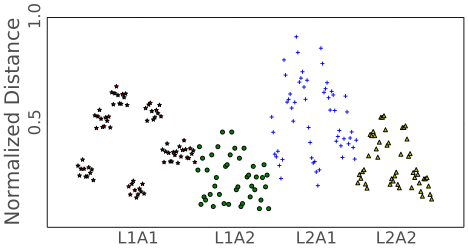


Figure 4.3: Spine Stretch.

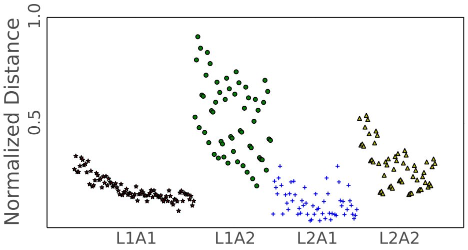


Figure 4.4: Draw.

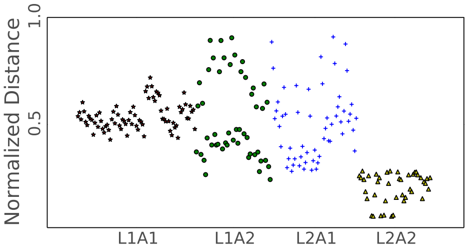


Figure 4.5: Bend-over.

two different activities; one is drawing and the other is bend-over (i.e., also an exercise with body leaning forward). The two chairs are 1-2 m away from the laptops. Figure 4.1 shows the scenario.

The WiFi transmission speed is 1250 packets per second (the choice of speed is explained in Sec.4.2.3). Each activity is repeated for 15 - 20 times with an interval of about four seconds. Then, we segment the CSI time series for each movement and compare their multipath profiles and motion profiles. Fig.4.2 - 4.5 shows the earth mover's distance (EMD) among the motion profiles (we will introduce the construction of motion profile in Sec. 4.2.3). We use short name 'L1A1' to represent first activity at location one, i.e., drinking water, and name the other activities with the same logic. As we can see, it is difficult to distinguish 'L1A2' from 'L2A2'. Also, 'L2A1' is hard to be differentiated from 'L1A1'. This validates that when activities have

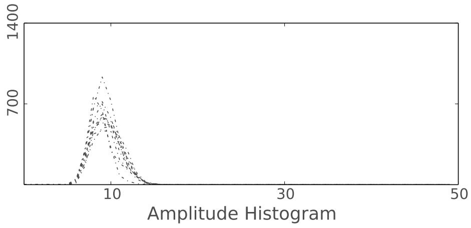


Figure 4.6: Drink.

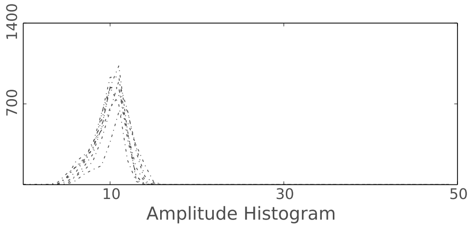


Figure 4.7: Spine Stretch.

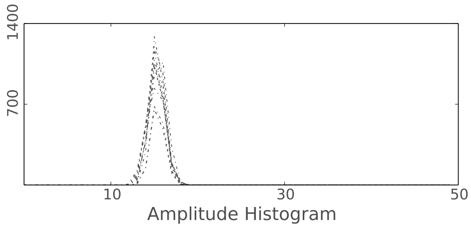


Figure 4.8: Draw.

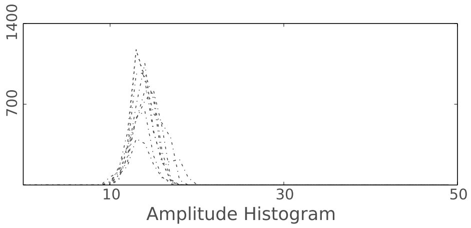


Figure 4.9: Bend-over.

similar motions then have similar motion profiles. On the other hand, Fig.4.6 - 4.9 shows the histograms of CSI amplitude on subcarrier 15 where each dashed line represents one instance. The observation is that despite the different motions, 'drink' and 'spine stretch' have similar amplitude histogram and the same applies to 'draw' and 'bend-over'. It also applies to other subcarriers (not shown due to limited space). Thus, the multipath profile which is based on the CSI statistical distribution is also not sufficient. This preliminary study gives us the motivation to combine those two profiles for a higher capability in recognizing activities.

4.2.2 Multipath-Profile Analysis

CSI to Robust Multipath Profile. In the wireless environment, each status of the human activity, i.e., the person's position and orientation, will cause a unique multipath propagation condition. As a result, the distribution of channel properties

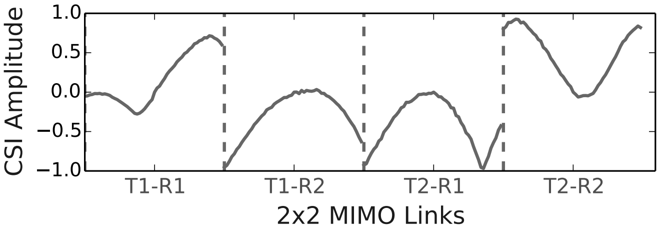


Figure 4.10: Normalized CSI amplitude in 4 links of 2x2 MIMO.

during one activity can be utilized to match this activity. In the channel state information (CSI), the value on each subcarrier represents the channel property on the corresponding frequency. In this work, we collect all CSI data during one activity to construct the multipath profile.

Traditional approach [70] just utilizes the distribution of amplitudes on each subcarrier, which substantially under-utilizes the full CSI information across the whole band. As explored by existing works [55, 69, 74, 82] in the literature, the full CSI vector can offer more useful information about the channel condition due to the diversity (Fig.4.10) across different frequencies in the 20 MHz WiFi band as mentioned in Sec.4.2.1.

Different from the localization works [55, 69, 74, 82], the object is moving at each location doing some activity in our case. Thus, the CSI information during the activity varies. Although one CSI vector for one feature may be feasible as different gestures differ in CSI with the same location, we believe that the statistics of CSI values during the activity could better distinguish the activity. In the localization works, each CSI vector is the feature for each location. Thus, they just need to train the model to learn the statistics of CSI values across different instances. However, in our case, we need to get the statistics of CSI values during each instance as the feature and learn the other statistics which represents the variation across different instances, i.e., dual-statistics problem. Interestingly, this idea also relates to the approach of applying distribution histogram of CSI amplitude on each subcarrier (E-eye [70]). The difference is that here we exploit the statistics of the whole CSI vector as the feature which is more challenging due to the high-dimensional vector, i.e., a 56-dimensional vector with Atheros chipset.

Multipath Profile Matrix. Although each CSI value has amplitude and phase, the phase information from commodity WiFi is unreliable due to the unsynchronized clock between the sender and the receiver [67]. To construct the multipath profile, we collect the amplitudes of CSI values during the activity. Specifically, the multipath profile is as the following matrix:

$$F = [f_1, f_2, ..., f_t, ..., f_T] \tag{4.1}$$

where, T is the total number of CSI values during the activity and $f_t = [|h_1|, |h_2|, ..., |h_{56}|]^T$ is the amplitude vector of CSI collected at timestamp t. With Atheros WiFi card, one CSI vector would be reported to the user space from the kernel with one received packet. Due to the uneven transmission of the WiFi packets, we use linear interpolation to achieve a consistent interval.

Dual-Statistics Analysis Approach. In this part, we introduce our model to deal with the challenging dual-statistics problem. First, let's review the current approach in E-eye [70]. This work calculates the distribution histogram on each subcarrier. Then, it obtains the distance between two distribution with earth mover's

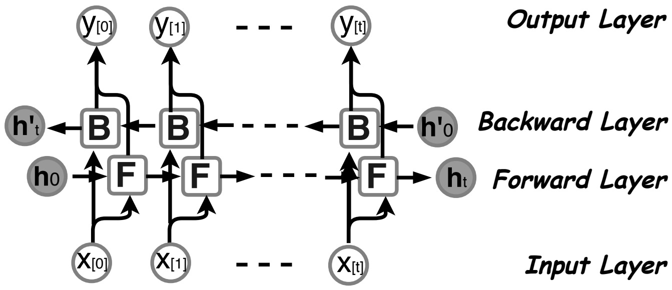


Figure 4.11: Bi-directional Recurrent Neural Network Architecture.

distance (EMD) algorithm. Although the scheme - get statistics and compare statistics, is simple. It doesn't apply to our problem. The essential reason is that high-dimensional distribution is hard to construct and compare. In the following, we introduce an end-to-end model to deal with the dual statistics.

First of all, we consider the neural network to learn the statistics of the high-dimensional CSI vectors. This approach has been validated in CSI-fingerprint based localization [55, 69, 74, 82]. However, our problem is different in that the feature is now a collection of CSI values during the activity instance. Thus, apart from the variation across different activity instances (i.e., each time the participant does the activity slightly differently), the model needs to extract the statistics of CSI values in the collection of one instance. Therefore, our second consideration is recurrent neural network (RNN) model which learns the feature across a collection of time series data.

Nevertheless, the RNN model only associates the information in one direction, i.e., pass the information from the past to the future timestamp. According to the nature of our problem, the loosely-defined activity doesn't have a strict order of body movement. Thus, a model that can associate the information in both directions would be more suitable. Based on this understanding, our solution is to use the bidirectional recurrent neural network (BiRNN) as the dual-statistics analysis model. BiRNN [54] model is capable to also associate with the future state and thus is more suitable for the activity with loosely-defined motions. In Sec.4.3.5, we validate the effectiveness of our model in dealing with dual statistics and also prove that other models are 36% worse in accuracy.

Fig.4.11 shows the bidirectional RNN model architecture. It includes two hidden layers connected to the output layer with the input sequence fed in normal time order for one layer, and in reverse time order for another. With this architecture, the output layer can get information from past and future states simultaneously. With the gradient descent method, every parameter in the neural network is updated in the direction of the deepest descent. For brevity, we refer interesting readers to the literature [54] for more details on bidirectional RNN. Specifically, our model uses gated recurrent unit cell (GRU) [18], which has a similar ability with LSTM cell [53] but fewer parameters. The input dimension is equal to the length of CSI vectors, which is 56*4 with 2x2 MIMO. The dimension of the internal state is set to 120. The batch size is set to 10. To enhance generality, we use a dropout wrapper with dropout rate as 0.5. Adam optimizer [35] is used to adaptively change the learning rate to precisely achieve the minimum cost.

Issue of Noisy CSI Values. As of the imperfection of WiFi chipset, the obtained CSI amplitude values is noisy. To reduce the noise, we average over 5 consecutive CSI values. As the transmission rate is 1250 p/s, 5 consecutive CSI values span the period of 4 ms, which is the time resolution of our CSI data.

Issue of Various Activity Durations. The detected and extracted activity segments would have various lengths. Therefore, it is not proper to directly put the

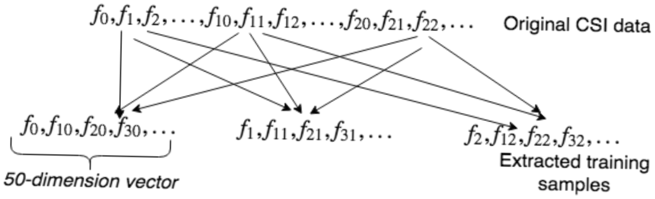

$f_0, f_1, f_2, \ldots, f_{10}, f_{11}, f_{12}, \ldots, f_{20}, f_{21}, f_{22}, \ldots$
Original CSI data
$f_0, f_{10}, f_{20}, f_{30}, \ldots$
$f_1, f_{11}, f_{21}, f_{31}, \ldots$
$f_2, f_{12}, f_{22}, f_{32}, \ldots$
50-dimension vector
Extracted training samples

Figure 4.12: Diagram of training sample extraction.

original data into the model. We note that since statistics is the useful information, it is feasible to evenly sample the original data. First, we assume that each activity segment is typically larger than 2 s. Then, the total number of CSI values after averaging is at least 500. By real tests, we select 50 as the reference length which gives us sufficient time resolution, i.e., 0.04 s. Besides, this configuration keeps the model run efficiently with Intel i5 processor and 7.7 GB memory.

Trick of Increasing Training Samples. In the machine learning model, more number of samples can increase the generality of the model and increase the accuracy. However, it is against user experience to obtain more training samples. Here, we utilize the high packet transmission rate to obtain multiple training samples for one activity instance. Recall that in the above, we have more than 500 CSI values for one activity instance. While we only use 50 for the model with even sampling to preserve the statistics. Therefore, by shifting the sampling start point, we can actually obtain 10 more training samples from just one activity instance. Fig.4.12 shows the diagram to demonstrate the idea.

4.2.3 Motion-Profile Analysis

CSI to Motion Profile In the literature, several existing works extract the motion profile from channel state information in different aspects. WiSee [50] looks into the frequency shift to infer the motion direction according to Doppler effect theorem [72]. WiFinger [60] extracts the variation pattern in CSI traces caused by finger motion. These approaches target at well-defined gestures and thus are not suitable for the loosely-defined human activity. To construct the motion profile, we take the idea in recent works [63, 67] that builds the relation between CSI values and object's speed.

Apart from the static signal paths, suppose there are K dynamic signal paths between the transmitter and the receiver, whose length is changing due to the human motion. For each path $k(k \in 1...K)$ and frequency f, the path length is $l_k(t)$ and the complex attenuation is $a_k(f, t)$. Then, with the signal phase shifting in the traveling and transmitter-receiver frequency offset Δf, the channel property $h(f, t)$ can be represented as:

$$h(f,t) = e^{-j2\pi\Delta ft}(h_s(f,t) + \sum_{k=1}^{K} a_k(f,t)e^{-j2\pi l_k(t)/\lambda})$$

where $h_s(f)$ is the contribution of all static signal paths. As the dynamic path changes in length, the power of channel property $|h(f,t)|^2$ changes. Within a small period time t, the object moves a short distance with speed v_k on dynamic path k. Then, we have the instantaneous path length $l_k(t) = l_k(0) + v_k * t$. In this way, we can derive the power $|h(f,t)|^2$ in Eq.4.2.

The above result reveals that the object's moving speeds are directly related to the frequencies in sinusoid components of $|h(f,t)|^2$. Therefore, by transforming the

CSI amplitude time series to the frequency domain, we can get the speed condition of the target.

[t!]

$$\begin{aligned}|h(f,t)|^2 = &\sum_{k=1}^{K} 2|h_s(f)a_k(f,t)|cos(\frac{2\pi v_k t}{\lambda} + \frac{2\pi l_k(0)}{\lambda} + \phi_{sk}) \\ &+ \sum_{\substack{k,l=1 \\ k\neq l}}^{K} 2|a_k(f,t)a_l(f,t)|cos(\frac{2\pi(v_k - v_l)t}{\lambda} + \frac{2\pi(l_k(0) - l_l(0))}{\lambda} \\ &+ \phi_{kl}) + \sum_{k=1}^{K} |a_k(f,t)|^2 + |h_s(f)|^2\end{aligned} \tag{4.2}$$

Transmission Rate and PCA-based Denoising. The packet transmission rate decides the time resolution of CSI values because the WiFi card reports one CSI value for each packet. As per the Nyquist theorem, the CSI sampling frequency should be at least two times of the variation frequency of CSI amplitude over time. As for 2.4 GHz WiFi signal, the variation frequency of CSI amplitudes is equal to the number of wavelengths in the distance that the target moves in one second, i.e., according to Eq.4.2. Therefore, 150 Hz is the upper bound of variation frequency caused by human speed less than 8 m/s. To have a good de-noising effect, we choose 1250 as the packet rate as suggested in CARM [67].

To combat the noisy CSI data from commodity WiFi cards, we use the PCA-based de-noising technique. It successfully extracts the correlation of variation across CSI subcarriers. Also, according to CARM [67], the major noise source is due to the internal WiFi card state transition. This transition causes a similar effect on all subcarrier and hence PCA-analysis can remove the correlated noise. However, different from CARM [67], the noise typically exists in component four and above in our data, which may be due to the different WiFi cards. Thus, our strategy is to keep the first three principal components. Fig.4.13 shows the original CSI value and the first principle component.

Motion Profile Construction. We convert the time-domain power $|h(f,t)|^2$ to the frequency domain so as to extract the motion profile. Compared with short time frequency transform (STFT), discrete wavelet transform (DWT) has advantages in obtaining high-frequency value with high time resolution and low-frequency value with high frequency resolution. We conduct DWT analysis on each CSI segment of 300 ms and extract 10 levels in the frequency domain. Our system applies DWT to decompose the principal components into 10 levels and average them to obtain the mean power value in each level. 10 is the maximum level we can get from the data and higher level would be zero-value due to the boundary effect. Each level represents one the frequency range which is exponentially decreased. For example, if level 1 represents a frequency range of $150 \sim 300$ Hz, then level 2 represents a frequency range of $75 \sim 150$ Hz. For brevity, we refer interesting readers to CARM [67] for detail of DWT processing.

We store the vector of power value in DWT levels for each CSI segment. To achieve smooth value over time, we move the segment window with a step of 100 ms. Thus, as to an activity instance about 2 s, the overall motion profile is a collection of 25 vectors with 10 dimensions. Fig.4.13 (right) shows the DWT power distribution over time.

Motion-Profile Analysis. Previous works [63, 67] match the DWT power vectors over time sequentially during the activity. This approach implicitly requires the motion to be a well-defined gesture with a certain order. As to general human activity, the motion is not well ordered over time. For instance, eating includes fetching food

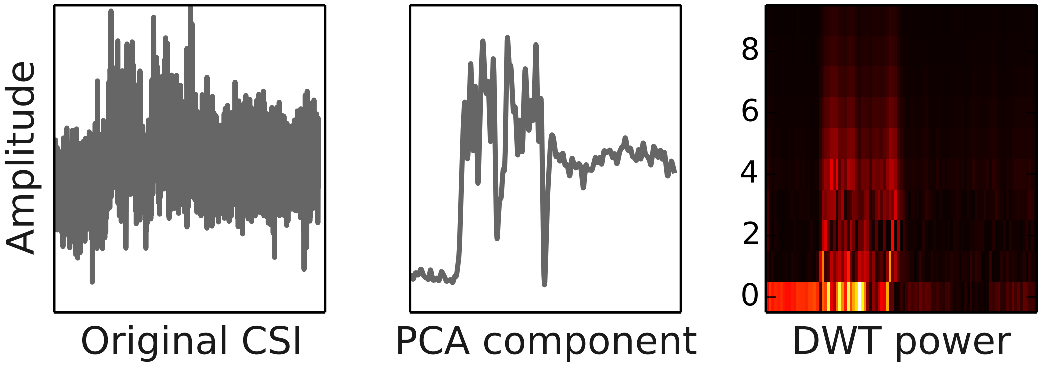


Figure 4.13: Effect of PCA-based de-noising.

to your plate, sending food to the mouth and chewing, which are not always in the same sequence. Thus, a new scheme of analysis is needed. Inspired by the previous work E-eyes [70], which use the distribution of CSI to identify the activity, our idea is to use statistical features of the motion profile to match the loosely-defined human activity.

The first approach we tried is to build the distribution on each DWT level and compare distributions of two activities to get the distance. Fig.4.2-4.5 show the distances between those instances. We use earth mover's distance (EMD) to calculate the distance between two distributions. As we can see, this approach roughly differentiates different activities with the distance metric but with poor performance. In Fig.4.3, the distances of 'L1A2', i.e., Spine-stretch, with other activities are not clearly separated with the distances among its own instances. This means that distance-based clustering could not robustly identify this activity. Besides, we note that 'L1A2' and 'L1A1' are quite different in motion, i.e., Spine-stretch versus Drink. Note that even this approach is sub-optimal it doesn't affect the conclusion in Sec.4.2.1 that with similar motion different activities have similar statistics in motion profile.

Motion Intensity. The second approach we tried is to calculate the total power across the activity to indicate the motion intensity. Other than directly adding coefficients in DWT levels, we use DWT level as a weight to the values. Specifically, assuming the coefficient in DWT level K is P, then we add $K * P$ to the sum as the overall motion intensity. The essential reason is that high DWT level corresponds to high frequency, which indicates a high speed of human motion (refer to Eq.4.2). As such, we calculate the motion intensity $\mathcal{I}$ for each activity instance. To show the performance in challenging case, we show the $\mathcal{I}$ with activities 'h' - 'j' which can be effectively detected in the first link in the experiment Sec.4.3 (the activity codes are defined in Table.4.1). This case is more challenging since these three activities (i.e., 'h', 'i' and 'j') are in the same place. Fig.4.14 shows the values of $\mathcal{I}$. From the results, we can see that by setting proper thresholds the accuracy of distinguishing these three same-location activities could be 93%. With the training samples, we set the thresholds to achieve the largest soft margins across activities with the same location. The reason that we focus on distinguishing activities with the same location using motion profile is that multipath profile analysis achieves high accuracy in differentiating activities in different locations, which is shown in Fig.4.17.

4.2.4 Profile Integration Mechanism

In this section, we discuss the integration strategy on the analysis results from the above two profiles.

Problem description. With the multipath-profile analysis, the output of the bi-RNN model is a probability vector $P = [p_{mul}(1), p_{mul}(2), ..., p_{mul}(n), ..., p_{mul}(N)]$ where $p_{mul}(n)$ is the probability of activity n. Each activity n in this model is

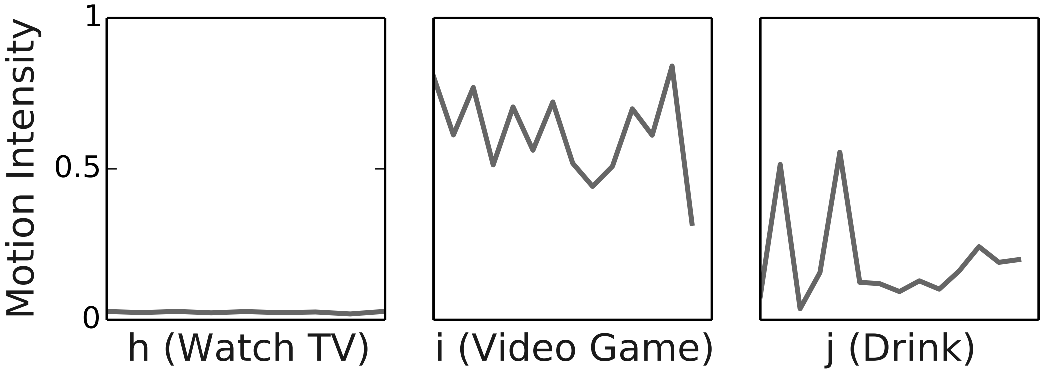


Figure 4.14: Motion intensities of three activities.

associated with a location $m = Loc(n), m \in 1...M$. Accordingly, for the motion-profile based approach, the output is the identified activity vector on each location $A = [a_{mot}(1), a_{mot}(2), ..., a_{mot}(m), ..., a_{mot}(M)]$ based on the motion intensity and the location-based threshold. The problem here is to achieve best recognition accuracy based on P and A.

Integration based on the priority. We first introduce a baseline approach. This approach is maximum probability based decision (MPD). MPD selects the decision between multipath-profile analysis and motion-profile analysis with the larger probability. The probability of the decision motion-profile analysis is based on the number of candidate activities. Note that vector A would have none value in some location if the motion intensity is out of range of all possible activities at that location.

The second approach is called priority based decision (PBD). We design PBD based on the observation from the experiment. That is, multipath-profile based analysis demonstrates high accuracy (~ 1) in distinguishing activities with different locations, shown in Fig.4.17. Thus, we put higher priority on the results from multipath-profile based analysis to decide the location of the activity. Specifically, we bundle the probability of same-location activities in vector P to obtain the probability of each location. Then, we only consider locations with probability larger than 30% (no more than two). What's more, if one location has a much higher probability (10% more) than the other, then we only look at one location. And, the activity with that location in vector A is the final decision. If the other location has close probability (within 10%), we then look into the motion intensity value and select the activity whose $\mathcal{I}$ value is farther from the threshold, i.e., less error with a larger margin.

With MPD, the accuracy of integration can only reach the best between multipath-profile analysis and motion-profile analysis. In contrast, due to the utilization of the priority, PBD approach can combine the advantages of these two profiles and boost the recognition accuracy, which will be demonstrated in evaluation Sec.4.3.7.

4.2.5 Activity Detection and Segmentation

In this section, we introduce the activity detection of TifWiFi system. In TifWiFi , we consider both static and motion activities. Static activity is the activity with no or little human movement, such as sleeping, reading, etc. Motion activity is thus the activity with human body movements, such as drinking, eating, etc. A successful activity detection is not only detecting the activity but also needs to correctly segment the activity in CSI data for recognition purpose. To achieve this with both static and motion activities, the activity detection of TifWiFi system utilizes both multipath profile and motion profile.

The strategy of activity detections in TifWiFi goes in two steps. First, we utilize multipath profile to detect the activity. After that, we use motion profile to distinguish motion activity from static activity and extract the segment of motion activity from

CSI data. Specifically, in the first step, we first obtain the reference profile with 'Empty' activity (code 'a' in Table.4.1), which means no human in the test area. Then, with the multipath profile at the current time, the activity is detected with a large deviation from the reference profile by calculating the Euclidean distance. To avoid random error, the activity is detected only when the large deviation lasts longer than 1 second. The threshold is set to three times of the maximum Euclidean distance among the instances of reference profiles.

As to the second step, we detect the motion activity if the motion intensity $\mathcal{I}$ (Sec.4.2.3) reaches 2 times of the maximum value during static period. Similarly, the minimum duration is 1 second to avoid random abrupt error. The start of the activity is when the motion intensity $\mathcal{I}$ goes up while the end is the time point when $\mathcal{I}$ starts 1 second duration of being a low value, i.e., less than 1.5 times of that in the static period. In Fig.4.16 of the evaluation section, we show the performance of activity detection of TifWiFi .

4.2.6 Integration of multiple links

As demonstrated in Sec.4.3.4, different links have different performances in recognizing activities in a certain area. This is because different links have different signal powers that go through that area, which causes different impacts on the received signal. Thus, to effectively recognize activity in the whole area, we need to integrate the results from multiple WiFi links.

The first part is the integration of activity detection. From our evaluation of the false alarm rate (FAR), we find that the FAR is quite low across a long duration (12 hours). Thus, the integration on activity detection is that as long as there is detection on one link then the activity is positive.

The second part is the integration of activity recognition. Since each activity has different impacts on different links, we should allocate different priorities to the recognition results. In this work, we use the motion intensity $\mathcal{I}$ to decide the priority. Specifically, if two WiFi links report different results of recognition, the one with the larger motion intensity $\mathcal{I}$ during the activity is the final decision. In Fig.4.23 of the evaluation section, we show the performance with individual links and the integration of two links.

4.3 Implementation and Evaluation

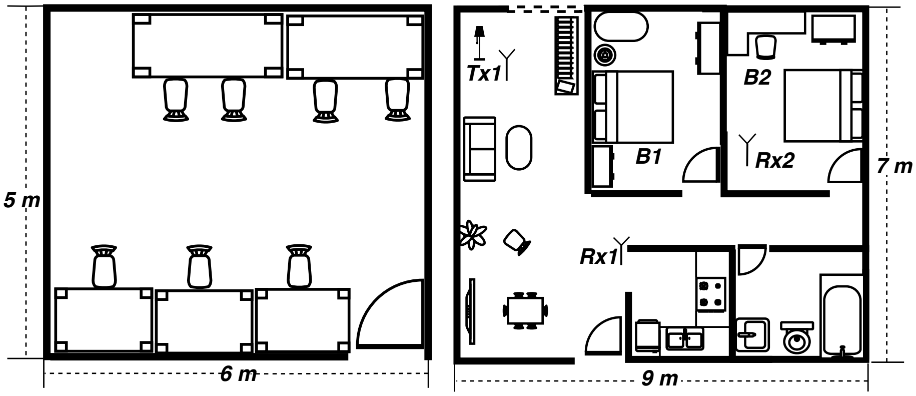


Figure 4.15: Testbed. The left is an office room and the right is a two-bedroom apartment.

4.3.1 Testbed

We conduct extensive experiments in two environments, shown in Figure 4.15. One is a typical office environment with multiple desks and chairs. The other is a two-bedroom home environment. Both of these occasions have abundant multipath propagation. The apartment setup provides NLoS scenario for our testing purpose.

4.3.2 Infrastructure Setup

We install Qualcomm Atheros chipsets (i.e., Atheros AR9382 and Atheros AR9462) on HP laptops as the transmitter and receiver, with one of them serving as access point. Each Wi-Fi PCIe card can support two antennas. Thus, with 802.11n Wi-Fi protocol, the signal transmission can support 2x2 MIMO stream. As such, for each correctly received packet, the Wi-Fi card will report 4 spatial CSI sets. To enable the CSI calculation and reporting functionality of the Wi-Fi card, 802.11 Wi-Fi standard

Table 4.1: Codes and Locations for Tested Activities

Codes	Activities	Location
a	Empty	
b	Sleep (bed)	Bedroom 2
c	Read (bed)	Bedroom 2
d	Phone call (bed)	Bedroom 2
e	Read (chair)	Bedroom 2
f	Bend-over (chair)	Bedroom 2
g	Type (chair)	Bedroom 2
h	Watch TV (sofa)	Living Room
i	Video Game (sofa)	Living Room
j	Drink (sofa)	Living Room
k	Wash Dishes (sink)	Kitchen
l	Eat (table)	Kitchen
m	Cook (stove)	Kitchen

specifies the requirement of setting the sounding flag. For this purpose, the linux kernal is modified with Atheros CSI tool [76] in Ubuntu 14.04 LTS environment, which supports up to 9 spatial CSI sets while only 4 of them are valid in our setting. In the 20 MHz Wi-Fi band, there are 56 subcarriers. Thus, each CSI set is a vector of 56 complex values, each of which has a resolution of 10 bits in both real and imaginary parts. Apart from CSI values, the timestamp is also recorded for later processing to account for the uneven arrival of the packets.

4.3.3 Data Collection

The carrier frequency is 2.462 GHz (WLAN channel 11) and the transmission bandwidth is 20 MHz. With each received packet, the WiFi chipset calculates CSI and reports it from the kernel space to the user space. However, due to the burst transmission, packet error and congestion, the actual packet transmissions may not be evenly distributed over time. Therefore, we utilize the timestamp information to interpolate between CSI values.

We have five participants in the experiments. They are college students, including male and female. In the two environments in Fig.4.15, we conduct different sets of activities. In the office, a simple set of four activities are tested, including drinking, drawing, bend-over and spine-stretch. While, in the apartment, there are 13 activities tested, shown in Table 4.1. During the experiment, each participant is guided to repeat each activity for 15-30 times with an interval of 4 seconds. The authenticate time for each motion is recorded for evaluation purpose.

In this section, we first present the benchmark performances. Then, we show the holistic performance of the system in two environments.

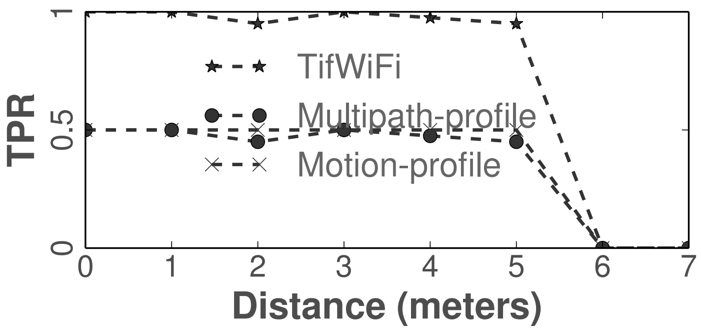
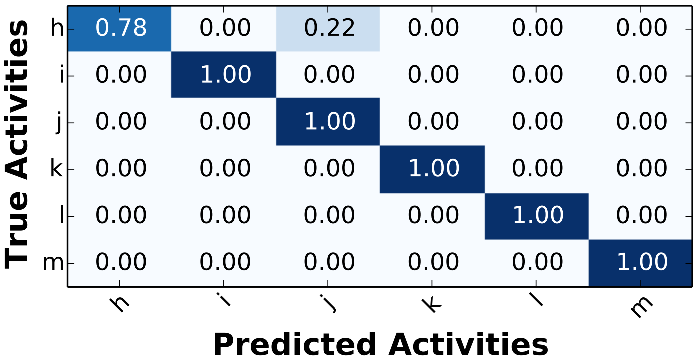


Figure 4.16: Detection range of TifWiFi .

4.3.4 Activity Detection

Here, we show the accuracy of TifWiFi in detecting the presence of an activity. True positive rate (TPR) and false alarm rate (FAR) are used as two metrics. In this test, we consider both motion activity (i.e., bend-over) and static activity (i.e., read). We compare the detection performance of our approach with multipath-profile based and motion-profile based approach. Fig.4.16 shows the TPR of TifWiFi for two different activities in the apartment, shown in Fig.4.15. We put the sender in the living room and the receiver in the kitchen with the non-line-of-sight (nLoS) link and a distance of 7 meters between them. For calculating each TPR, the participant repeated the activity for 20 times at one location with some distance to the receiver and another location with the same distance to the transmitter. From the results, we can see that TifWiFi detects both static and motion activities with a TPR larger than 95% at a distance of up to 5 meters. In comparison, multipath-profile based approach can only achieve TPR of 45% at the same distance while motion-profile based approach also only achieves TPR of 50%. The reason is that multipath profile cannot detect the motion and thus it fails in extracting motion activity segment in



	h	i	j	k	l	m
h	0.78	0.00	0.22	0.00	0.00	0.00
i	0.00	1.00	0.00	0.00	0.00	0.00
j	0.00	0.00	1.00	0.00	0.00	0.00
k	0.00	0.00	0.00	1.00	0.00	0.00
l	0.00	0.00	0.00	0.00	1.00	0.00
m	0.00	0.00	0.00	0.00	0.00	1.00

Figure 4.17: Confusion matrix of TifWiFi with the multipath profile.

CSI data. On the other hand, the motion profile is insensitive to the static activity. We also note the limitation that it is hard to detect the activity with nLoS locations, especially when the human is two-room away, e.g., the second bedroom in the Fig.4.15 which is blocked by two walls in the middle (cases of 6 m and 7 m in this test).

To measure the false alarm rate, we keep the above setting and collect the data during normal daytime from 8:00am to 8:00pm. To achieve real-time processing, we construct a producer-consumer virtual pipe with data collection script and data processing script. The virtual pipe is storing each 10 s data as a file in a folder with a number ranging from 1 - 1000. Then, the data processing script sequentially process and delete these files after finish. During 12 hours with no one in the apartment, the total number of false alarms is 16. Those infrequent false alarms are mostly caused by passengers outside of the apartment.

4.3.5 Multipath-Profile Analysis

We now present the performance of our dual-statistics based approach (denoted as ‘dual-stat’) in multipath-profile analysis. For comparison, we also implement two

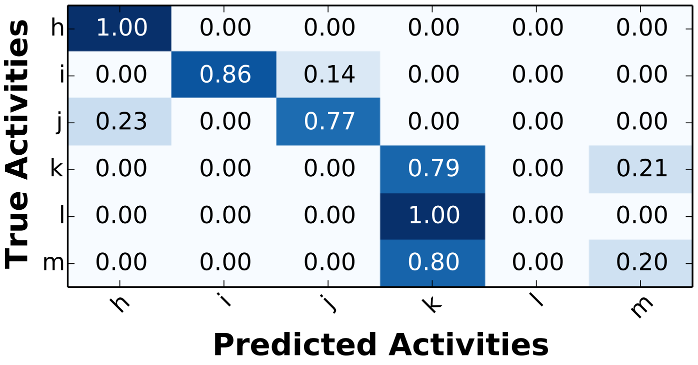


Figure 4.18: Confusion matrix of ‘amp-dist’ with the multipath profile.

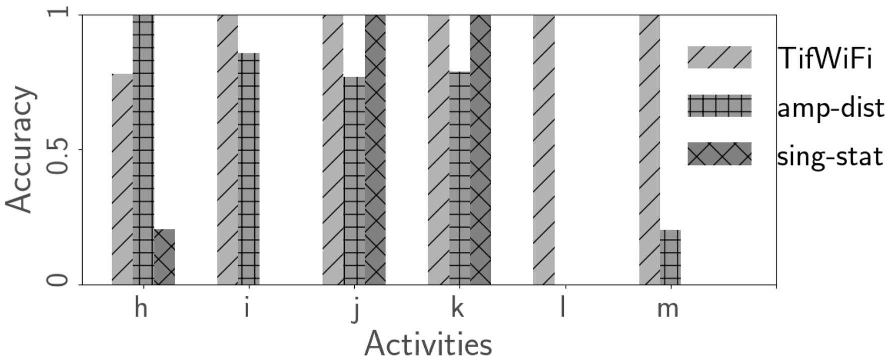


Figure 4.19: Accuracy of three approaches with the multipath profile.

other alternative approaches. One is the amplitude distribution per subcarrier used in E-eyes [70] (denoted as ‘amp-dist’) and the other is single-statistics based approach (denoted as ‘sing-stat’). The ‘sing-stat’ approach takes each CSI vector collected during the activity as the feature and trains the model with fully-connected neural network. TifWiFi can achieve an average cross-validation accuracy of 96% in recognizing the activities. Fig.4.17 shows the confusion matrix. For ease of analysis, we just show the results from the Tx1-Rx1 link. As such, six activities (i.e., ‘h’-‘m’) can be effectively detected with TPR larger than 96.3%. Then, the accuracy is evaluated with the detected and extracted activity segments in CSI data. Later, we will combine results from all links in the holistic system evaluation. In comparison, the 'amp-dist' approach only gets an accuracy of 60.3% while 'sing-stat' gets an accuracy of 36.7% (Fig.4.19). Within the confusion matrix of 'amp-dist' approach (Fig.4.18), we can see that activities with the same and close locations are prone to be confounded. That is because 'amp-dist' approach doesn't utilize the diversity information across different frequencies across the WiFi band.

Therefore, the results validate the hypothesis that the statistics of CSI vectors is more reliable than amplitude distribution per subcarrier in activity recognition. Also, it demonstrates that our bi-directional RNN model successfully handles the dual-statistics analysis task. Besides, the low accuracy of 'sing-stat' also demonstrates that statical distribution of CSI vectors during the activity is more reliable than individual CSI snapshots. We also note that TifWiFi loses accuracy in recognizing activity 'h' with wrong detection of activity 'j'. However, 'h' and 'j' share the same location, i.e., sofa. The results in link 1 show that TifWiFi has an accuracy of 100% in recognizing activities in different locations. We also verify the high accuracy in location differentiation with link 2, i.e., Tx1-Rx2 link.

Different numbers of histogram bins. In E-eyes [70], it is reported that with more bins in the distribution histogram, the classification could achieve higher accuracy. Thus, we tested the 'amp-dist' approach with various numbers of histogram bins ranging from 10 to 60. In comparison with TifWiFi , we use the best performance of the 'amp-dist' among those cases.

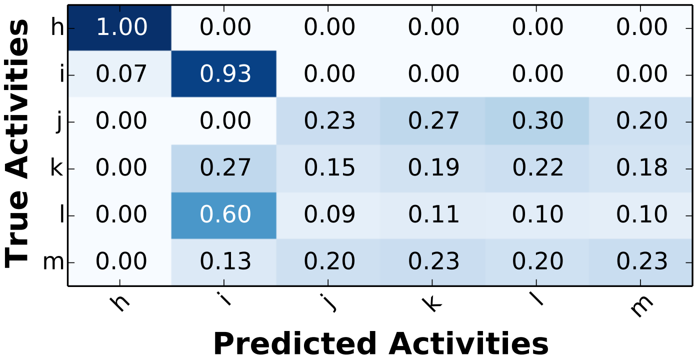


Figure 4.20: Confusion matrix of TifWiFi with motion profile.

4.3.6 Motion-profile Analysis

In the motion-profile analysis, we show the performance of recognizing different activities. Similarly, we just look into the data collected from link Tx1-Rx1 to facilitate the analysis.

Fig.4.20 shows the confusion matrix with the six effectively detected activities, i.e., ‘h’-‘m’. As we can see, the overall average cross-validation accuracy of motion-profile analysis is about 44.7%, which is worse than multipath-profile analysis. The reason is that several sets of activities share a similar motion pattern, i.e., body parts and speed. For example, both ‘wash’ and ‘cook’ includes the shaking of the whole upper body. Thus, with the motion intensity as the metric, the activities with the similar motion pattern are prone to be confounded. As shown in Fig.4.20, the even distribution of the error across activities ‘j’ - ‘m’ is because they share a similar motion intensity threshold. Despite the poor overall performance, we note that motion-profile analysis performs well in distinguishing activities in the same location with different motion intensity, i.e., larger than 93% across ‘h’, ‘i’ and ‘j’. With no error in recognizing ‘h’, the motion-profile analysis can enhance the accuracy of multipath-profile analysis in Fig.4.19 to 100%. In the following section, we will show that multipath profile and motion profile could also improve the holistic accuracy with the integration strategy proposed in Sec.4.2.4.

4.3.7 Holistic System Performance

In this section, we present the holistic evaluation on TifWiFi in two typical indoor environments. In calculating the recognition accuracy, we take into account the missing detection. The missing detection would be regarded as 'Empty' activity (i.e., 'a' in Table.4.1).

Office Environment

The office area is an open space and provides a line-of-sight (LoS) scenario. The space is 5m x 7m as shown in Fig.4.15. As mentioned in Sec.4.3.3, we tested four activities, which are drinking, drawing, bend-over and spine-stretch. As the WiFi infrastructure, the sender is put on the desk and the receiver is put on the ground with a distance of 3 m. Two chairs are put in a range of 1-2 m from the sender and the receiver respectively. The participants are guided to conduct two activities on one chair and the other two on the other chair, as explained in Sec.4.2.1. Each activity was repeated 15 times with an interval of ~4s.

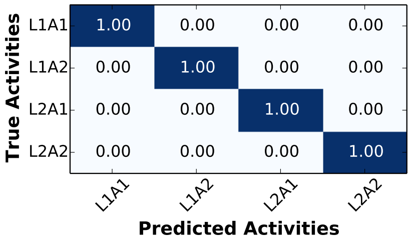


Figure 4.21: Confusion matrix

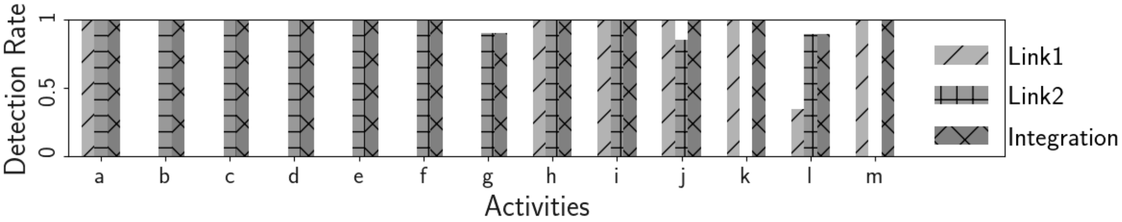


Figure 4.22: Detection rate of all activities with two links and integration.

Fig.4.21 shows the confusion matrix. We use cross-validation to obtain the recognition accuracy. The results show that TifWiFi can reliably detect and recognize all four activities in an open area with LoS links.

Apartment Environment

As shown in Fig.4.15, the apartment area is 7m x 9m with two bedrooms. To cover the whole area, we set up two WiFi links with one transmitter and two receivers. The transmitter Tx1 is put in the living room. One receiver is put in kitchen room and another is put in the second bedroom. The participants are guided to perform the activities in Table.4.1. Each activity is repeated for 20 times. ‘Empty’ activity means there is no human in the apartment. In the experiment, we trigger the transmission with wireless mouse and ubuntu Onboard when the person is outside. As mentioned in Sec.4.2.5, both activities have some blind areas in the apartment which is far from the sender and the transmitter. In Fig.4.22, we show the detection rate (TPR) across all activities in three cases, i.e., Tx1-Rx1 link, Tx1-Rx2 link and two-link integration. As we can see, the overall average detection accuracy is improved from 65.4% with one link to 98.4% with two-link integration. With link 1 (Tx1-Rx1), the missing

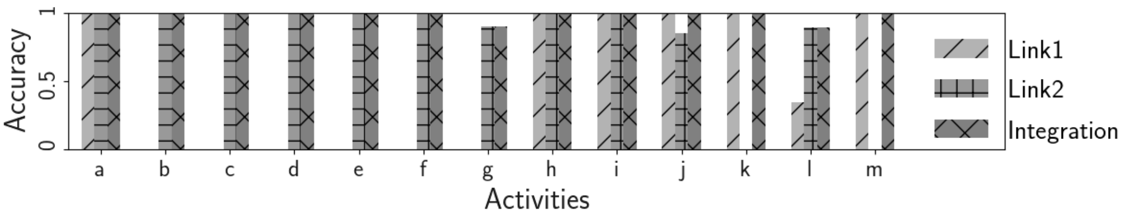


Figure 4.23: Recognition accuracy of all activities with two links and integration.

detection comes from activities 'k' and 'm' in the kitchen. As to link 2 (Tx1-Rx2), the missing detection comes from activities 'b' to 'g' which are in Bedroom 2.

The next is the evaluation on the activity recognition. *TifWiFi achieves an average cross-validation accuracy of 98% across all activities.* We show the accuracies of individual links and two-link integration in Fig.4.23. The low accuracy of individual links is due to the large missing detection rate, i.e., missing detection case is misclassified as 'Empty'.

Two-profile integration. To analyze the effect of two-profile integration in the holistic evaluation, we take the recognition results from multipath profile and motion profile independently in the integration of two links. In this way, we can compare the performance of individual profile with the two-profile integration. Fig.4.24 shows the accuracy across all activities. Here, we keep using two profiles in activity detection and only separate them in activity recognition.

As we can see, the two-profile integration strategy improves the accuracy over both multipath profile and motion profile. The multipath-profile analysis with dual-statistics scheme achieves overall average accuracy of 92.8% while motion-profile analysis achieves accuracy of 23.08%. In comparison, the two-profile integration achieves

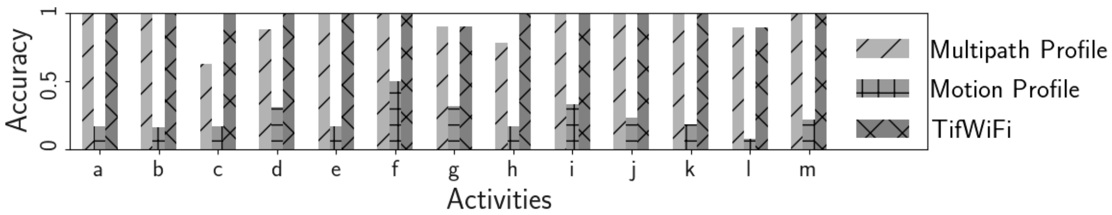


Figure 4.24: Recognition accuracy of all activities with three approaches.

a higher accuracy of 98%. In Fig.4.24, the multipath-profile analysis loses accuracy with confounding between ‘c’ and ‘d’, whose accuracies are 62.5% and 87.5% respectively. Although motion-profile also performs poorly on ‘c’ and ‘d’ (i.e., 17% and 30.6%), it obtains a high recognition ability when it knows that ‘c’ and ‘d’ are in the same location from multipath-profile analysis. As such, the accuracies of both ‘c’ and ‘d’ are enhanced to 100% with two-profile integration. Thus, it’s validated that *two-profile integration strategy is better than individual profile analysis.* Besides, it also proves that our dual-statistics scheme achieves high accuracy even in recognizing activities with the same location.

4.4 Summary

In this work, we propose the design of a two-profile integration framework TifWiFi . With dedicate design on both multipath-profile analysis and motion-profile analysis, TifWiFi provides a better performance over existing approaches. In the whole home environment, TifWiFi have an accuracy of 98% across 13 activities. With TifWiFi , we expect to provide a better activity recognition service to the public.

Chapter 5: PhyCloak: Obfuscating Sensing from Communication Signal

5.1 Overview

At first glance, it might appear that an obvious way to prevent or deter the privacy leakage is to simply jam the signals [23, 45]. However, jamming is an overkill for this problem, as the protection we wish lies in physical and not in the logical (data) layer. Jamming distorts the information of both layers, therefore it hurts the channel capacity of the network. In contrast to jamming, our approach is to distort the physical information that is environmentally superimposed on the signal as opposed to the data itself. *To make clear the distinction between these two forms of signal distortion, we refer to the latter as signal obfuscation.*

To avoid any modification of existing receivers, we need to build an obfuscator (Ox) that works independently from a receiver (Rx) and can yet deter privacy leakage. At the same time, Ox should not hurt the ongoing reception at the intended receiver. In addition, given the diversity of the design of RF based sensors and invisibility of eavesdroppers, it is not reasonable to assume Ox that uses a specific obfuscation approach against a specific Eve. Thus, our goal is to build a black-box solution which distorts only the privacy sensitive information while not affecting the logical information. We design Ox by answering the two questions below:

1. How to distort physical information regardless of the RF-sensing mechanism? To answer this question, let us first examine what kind of physical information is contained in RF signals. Assume the received signal at a reflector is $s(t)$, then the received signal $r(t)$ reflected by the reflector can be expressed as follow: $r(t) = a \times s(t) \times e^{j2\pi(f_c+\Delta f)(t+\Delta t)}$, where a is the amplitude gain, f_c is the carrier frequency, Δf is the Doppler shift caused by a reflector that moves at a constant speed relative to the receiver, and Δt is the delay due to transmission over the path. Here, we can see that the reflector modifies the reflected copies by controlling three orthogonal components: amplitude gain a, delay Δt and Doppler shift Δf. All the features exploited by single-antenna RF based sensors are created by these three degrees of freedom (DoFs). Hence, if an Ox distorts the three orthogonal bases respectively, any features that reveal physical information are distorted too.

2. How to preserve logical information (data communication)? As the previous observation suggests, Ox needs to change the 3 degrees of freedom (DoFs) of a signal in order to deter eavesdropping of physically sensed features. Note that in a wireless environment, signals traverse through many paths and experience Doppler shifts: These effects are similar to dynamic multipath reflections. Thus, Ox can be a relay node that introduces dynamically changing multipath components of the communication signal. In other words, Ox receives the incoming communication signal, manipulates the signals and forwards them back to the environment. To a legitimate receiver, this forwarded signal will simply look like a multipath component of the signal from the legitimate transmitter (Tx). Commercial off-the-shelf (COTS) Rx are capable of tolerating and even exploiting multipath reflections to decode data. Thus, a carefully designed Ox can distort sensing and still preserve communication.

Challenges: PhyCloak works as a full-duplex amplify-and-forward (A&F) relay at logic layer, and an Ox at physical layer by distorting the 3 DoFs. While the solution may appear at first blush to be a simple instance of full-duplex A&F forwarder [12,15], there are key challenges that arise from this design that need to be resolved.

1. *Online self-channel estimation with an ongoing external transmission*: Traditional self-interference cancellation of full duplex assumes that the self-channel –the channel between the relay's transmit and receive antennas– is stable for a relatively long time and so, estimates self-channel infrequently. An Ox, however, works in an environment where the channel is varying as a result of target movement, gestures and activities. Self-channel estimation in the presence of this variation is particularly relevant when we combine the Ox module with a legitimate sensor, so as to preserve legitimate sensing while simultaneously obfuscating illegitimate sensing. Therefore, Ox needs to transmit training signals frequently to perform self-channel estimation so as to achieve sufficient and stable cancellation. To make matters worse, Ox needs to do channel estimation during an ongoing transmission. Therefore, it has to tolerate external transmission during training. A straightforward way to overcome this problem is to adopt medium access control (MAC), however, that would introduce contention and hurt throughput of legitimate data transmission.

2. *Effectiveness of obfuscating physical information*: No work has been done in validating a full-duplex A&F forwarder's capability of controlling physical information contained in the forwarded copy. In addition, the effectiveness of superposing an Ox's

distorted signal and a target's reflected signal in obfuscating an eavesdropping sensor has yet to be shown.

Contributions: We propose PhyCloak to protect privacy information from unwanted or even malicious sensing with no modification to existing wireless infrastructures. In this work, we make the following contributions:

1. To our knowledge, we are the first to address the potential threats due to the recent development of communication-based sensing.
2. We propose PhyCloak, the first full-duplex forwarder-based solution that hides physical information superimposed by the channel via adding interference in a 3-dimensional orthogonal bases so that illegitimate sensing is disabled and meanwhile data transmission is not affected (and even improved). We go further and add the capability to spoof human gestures to further confuse illegitimate sensors.
3. We propose an alternative online self-channel estimation scheme that is contention-free and operates in the presence of an ongoing transmission. By doing so we also allow for legitimate sensing by integrating the sensor with our obfuscator.
4. We build a prototype PhyCloak on PXIe-1082, an SDR platform. Experimental results (Section 5.4.3) on a state-of-the-art sensor show that PhyCloak successfully obfuscates illegitimate sensing, enables legitimate sensing and improves overall throughput of data transmission. Gesture spoofing to the same type of sensor is also proved to be feasible.

5.2 System Model and Preliminary

5.2.1 Threat Model

Assume there is an adversary who is interested in inferring physical information from a SISO wireless communication channel. The adversary may be active or passive, i.e., it can transmit itself or just exploit ongoing wireless transmissions. In both cases, we assume that the adversary uses a single-antenna receiver to sniff the wireless transmission. In general, the design and implementation of adversarial sensing is unknown to the protection system designer.

Note that some types of sensing require a training phase to tune recognition patterns with respect to the environment of interest. To protect against stronger adversaries, we assume that the adversary is well trained for the environment at hand. The details of this training, whether it occurs concurrently with the training of a legitimate sensor or is based on some historical knowledge, are outside the scope of our interest here.

5.2.2 System and Goals

Our protection system comprises 4 SISO nodes as shown in Figure 5.1: Alice (data transmitter), Bob (data receiver), Carol (legitimate sensor) and Eve (illegitimate sensor). Both Alice and Bob can be controlled by Eve, thus Carol does not assume that Alice and Bob are honest.

Goals: 3 tasks co-exist in the network: data transmission between Alice and Bob, illegitimate sensing at Eve and legitimate sensing at Carol. By adding Ox to Carol with no cooperation from any of the other nodes, the protection system must satisfy the following three goals:

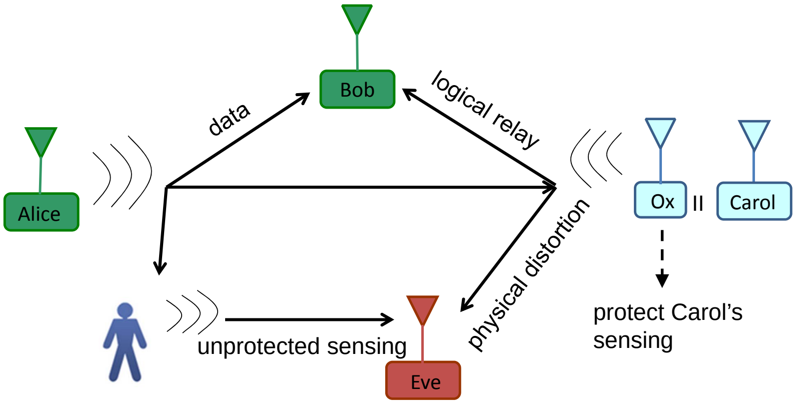


Figure 5.1: 4 single-input-single-output (SISO) nodes exist in the system: Alice, Bob, Carol and Eve: Alice and Bob perform data transmission and reception; Eve performs illegitimate sensing by exploiting Alice's transmission; Carol also performs sensing, but her obfuscator module forwards the received signal in a way that distorts physical information but preserves logical information

1. Obfuscate Eve's sensing.
2. Preserve Carol's sensing.
3. Not degrade the throughput of the link between Alice and Bob, nor introduce extra computation at Alice and Bob; i.e., Alice's and Bob's behaviors stay unaltered when Ox operates.

5.2.3 Three Degrees of Freedom

Usually a forwarder relays the signal directly, but in the context of an Ox a forwarder can do far more. In fact, a forwarder can be viewed as a special type of reflector; in theory, whatever change a natural reflector can induce on a signal, a forwarder can induce likewise. We begin by examining how a reflector changes the signal.

Letting the received signal at a reflector be $s(t)$, the received signal $r(t)$ that it reflects can be expressed as

$$r(t) = a \times s(t) \times e^{j2\pi(f_c+\Delta f)(t+\Delta t)} \tag{5.1}$$

where a is the amplitude gain due to reflection and propagation, f_c is the carrier frequency, Δf is the Doppler shift caused by a reflector that moves at a constant speed relative to the receiver, and Δt is the delay due to propagation over the path. We see that a reflector modifies signals by changing three components: a, Δf and Δt. Namely reflectors enjoy three DoFs when modifying signals.

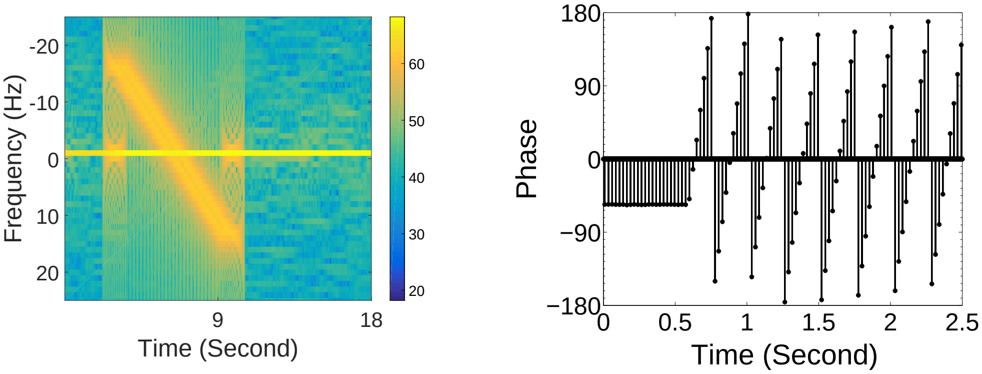
We examine what kind of signal processing is needed at the Ox to effect similar changes in the signal being forwarded. Rewrite Equation 5.1 into the following form:

$$r(t) = a \times s(t) \times e^{j2\pi\Delta ft} \times e^{j2\pi(f_c+\Delta f)\Delta t} \times e^{j2\pi f_c t} \tag{5.2}$$

Amplitude gain a: It is clear that if a forwarder receives $s(t)$ from the source, then by amplifying the samples with different levels, a can be easily changed.

Doppler shift Δf: To emulate a Doppler shift of Δf, a forwarder can rotate the nth received sample by $2\pi n\Delta f\overline{\Delta t}$, where $\overline{\Delta t} =$ sampling interval.

Delay Δt: A delay of Δt can be introduced by simply delaying the to-be-forwarded signals in either the digital domain or the analog domain at the forwarder. A problem with delaying signals in the digital domain is that digital delays are discrete and do not match the speed of human movement. For example, if an ADC works with a sampling rate 100MHz, then the minimum delay that can be introduced in digital domain is 10ns, which corresponds to a distance of 3m. Controlling analog delay while feasible, however requires effort in modifying existing SDR platforms. Our solution then is to rotate the to-be-forwarded samples by a fixed phase $2\pi(f_c + \Delta f)\Delta t$ in the digital domain, which matches the expected delay of Δt. In our NI PXIe platform, this calculation can be made in two clock cycles ($\frac{1}{\text{ADC sampling rate}}$).



(a) By multiplying the nth to-be-forwarded sample with $2\pi n\Delta f\overline{\Delta t}$, and changing Δf from 20Hz to -20Hz, the Doppler shift profile at the receiver is as expected

(b) By rotating the to-be-forwarded signals with a certain phase which changes by 36° every 30ms at the forwarder, the phase of the signal changes $\sim 36°$ every 30ms

Figure 5.2: Expected Doppler shift and phases are generated at a forwarder

Figure 5.2(a) depicts the Doppler shift profile of the received signals that are sent by a forwarder who keeps changing the to-be-forwarded samples' Doppler shift from 20Hz to -20Hz according to the above algorithm. Similarly, from Figure 5.2(b) we can see that by multiplying the to-be-forwarded sample with a phase φ which increases 0.2π every 30ms at the forwarder, the phase of the received samples changes by $\sim 0.2\pi$ every 30ms. These results show that a forwarder can predictably control Doppler shift and phase.

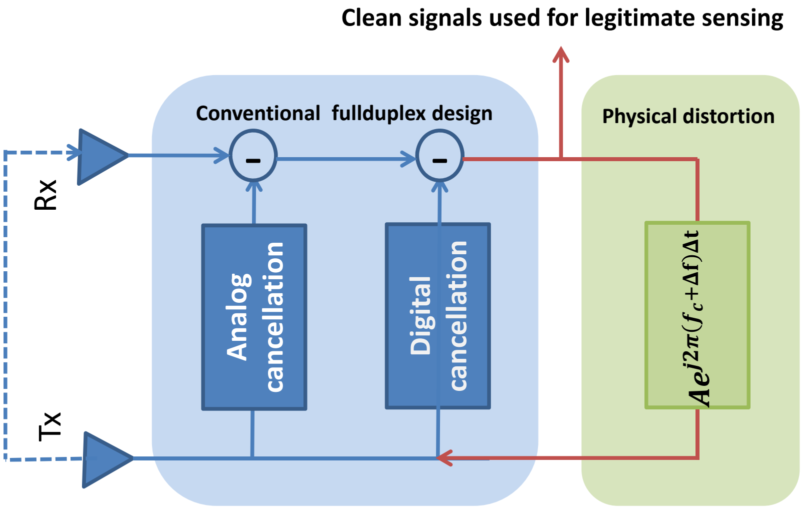

Clean signals used for legitimate sensing
Conventional fullduplex design
Physical distortion
Rx
Tx
–
–
Analog cancellation
Digital cancellation
$Ae^{j2\pi(f_c+\Delta f)\Delta t}$

Figure 5.3: High-level block diagram of PhyCloak

5.3 PhyCloak Design

Figure 5.3 shows a simplified block diagram of our system PhyCloak. The physical distortion is introduced after self-interference cancellation, the distorted signal is then forwarded to the transmit antenna.

5.3.1 Online Maintenance of Self-Channel Estimates

As mentioned earlier, PhyCloak is a full-duplex system that needs to cancel self-interference to operate. However, human movements change the self-channel and affect cancellation. To maintain reliable self-interference cancellation, PhyCloak needs to re-train itself online to update its self-channel estimate periodically and frequently. More precisely, updates need to be performed every coherence interval, which is typically in the range of tens to hundreds of milliseconds.

A complication arises when an update is attempted during an ongoing external transmission: the external transmission may distort self-channel estimation while the transmission that helps with self-channel estimation may interfere with external data reception. There are two straightforward solutions to this problem: 1) using MAC; 2) exploiting the silent period defined by wireless protocols, like short inter-frame space (SIFS) in WiFi. The former hurts the throughput of data transmission and moreover interrupted external transmission degrades coupling legitimate sensors with the Ox. And in addition, both of the solutions require a big effort to design careful adaptation to various wireless communication protocols.

We therefore propose a self-channel estimation algorithm for PhyCloak that addresses this complication. It uses two main elements: 1) oversampling and differential to get rid of any ongoing external transmission, and 2) a special training sequence that yields minimum interference to external transmissions.

Self-channel estimation with and without external interference

Before we describe our self-channel estimation algorithm, let us first see the impact of training with and without external interference. Assume $A = \{a_{-m}, a_{-m+1}, \ldots, a_m\}$ is the transmitted training sequence, $B = \{b_0, b_1, \ldots, b_m\}$ is the received sample sequence, and $H = \{h_0, h_1, \ldots, h_m\}$ is the channel coefficient vector in time domain with $m + 1$ taps. Therefore, we have

$$\left\{\begin{matrix} b_0 \\ b_1 \\ \ldots \\ b_m \end{matrix}\right\} = \left\{\begin{matrix} a_0 & \ldots & a_{-m} \\ a_1 & \ldots & a_{-m+1} \\ \ldots & \ldots & \ldots \\ a_m & \ldots & a_0 \end{matrix}\right\} \times \left\{\begin{matrix} h_0 \\ h_1 \\ \ldots \\ h_m \end{matrix}\right\} \tag{5.3}$$

In the presence of external transmission, B becomes:

$$\left\{\begin{matrix} b_0 \\ b_1 \\ \dots \\ b_m \end{matrix}\right\} = \left\{\begin{matrix} a_0 & \dots & a_{-m} \\ a_1 & \dots & a_{-m+1} \\ \dots & \dots & \dots \\ a_m & \dots & a_0 \end{matrix}\right\} \times \left\{\begin{matrix} h_0 \\ h_1 \\ \dots \\ h_m \end{matrix}\right\} +$$

$$\left\{\begin{matrix} s_0 & \dots & s_{-m} \\ s_1 & \dots & s_{-m+1} \\ \dots & \dots & \dots \\ s_m & \dots & s_0 \end{matrix}\right\} \times \left\{\begin{matrix} h_0^{'} \\ h_1^{'} \\ \dots \\ h_m^{'} \end{matrix}\right\} \tag{5.4}$$

where $S = \{s_{-m}, s_{-m+1}, \dots, s_i, \dots, s_m\}$ is the external transmitted sample sequence, and $H^{'} = \{h_0^{'}, h_1^{'}, \dots, h_m^{'}\}$ is the channel coefficient vector which corresponds to the channel between the transmit antenna of the external device and the receive antenna of the Ox.

Oversampling and differential to get rid of external interference

To overcome the external interference in Equation 5.4, which is unknown to PhyCloak, we exploit oversampling. Say PhyCloak samples at a rate $2m$ times higher than the sampling rate of the external transmitter, it follows that approximately $s_{-m} = \dots = s_m$. So

$$\left\{\begin{matrix} s_0 & \dots & s_{-m} \\ s_1 & \dots & s_{-m+1} \\ \dots & \dots & \dots \\ s_m & \dots & s_0 \end{matrix}\right\} \times \left\{\begin{matrix} h_0^{'} \\ h_1^{'} \\ \dots \\ h_m^{'} \end{matrix}\right\} = \left\{\begin{matrix} s_0 \times (h_0^{'} + \dots + h_m^{'}) \\ s_0 \times (h_0^{'} + \dots + h_m^{'}) \\ \dots \\ s_0 \times (h_0^{'} + \dots + h_m^{'}) \end{matrix}\right\} \tag{5.5}$$

Therefore, by differential we have

$$\left\{\begin{matrix} b_1 - b_0 \\ b_2 - b_1 \\ \dots \\ b_m - b_{m-1} \end{matrix}\right\} = \left\{\begin{matrix} a_1 - a_0 & \dots & a_{-m+1} - a_{-m} \\ a_2 - a_1 & \dots & a_{-m+2} - a_{-m+1} \\ \dots & \dots & \dots \\ a_m - a_m & \dots & a_1 - a_0 \end{matrix}\right\} \times \left\{\begin{matrix} h_0 \\ h_1 \\ \dots \\ h_m \end{matrix}\right\} \tag{5.6}$$

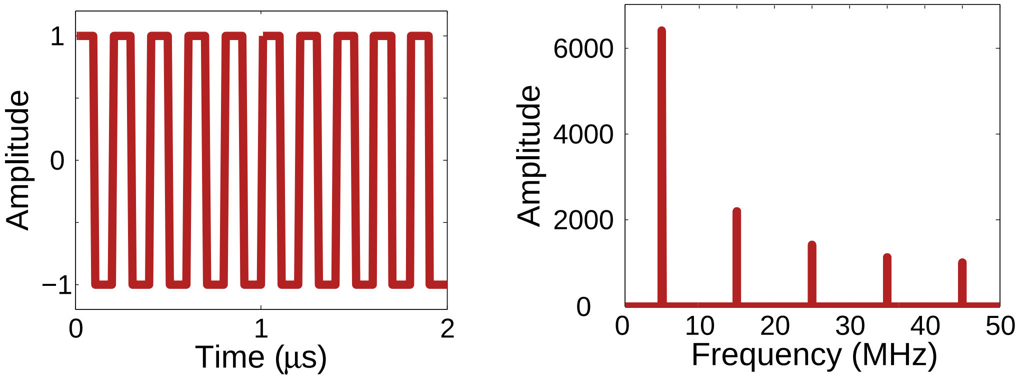
It may appear that we have already been able to get rid of external interference, however, $\mathbf{A}$ is an $m \times (m+1)$ matrix, so the rank of A is less than $m+1$. This means that we can get only a unique solution for at most m of the $m+1$ unknowns contained in $\mathbf{H}$, where $\mathbf{H} = \{h_0, h_1, \ldots, h_m\}^T$ and

$$\mathbf{A} = \left\{ \begin{matrix} a_1 - a_0 & \ldots & a_{-m+1} - a_{-m} \\ a_2 - a_1 & \ldots & a_{-m+2} - a_{-m+1} \\ \ldots & \ldots & \ldots \\ a_m - a_{m-1} & \ldots & a_1 - a_0 \end{matrix} \right\} \tag{5.7}$$

A special training sequence

To ensure that Equation 5.6 has a unique solution for $\{h_0, h_1, \ldots, h_{m-1}\}^T$, we leverage a special training sequence, namely a square wave, which is shown in Figure 5.4(a). As shown in Figure 5.4(b), the fundamental frequency of the square wave is the square wave frequency, and its odd harmonics are decreasing in size. To be more specific, for a square wave over a period consisting of N samples with B MHz sample rate, the frequency components are at $1f$, $3f$,..., $(2i+1)f$, ... with decreasing amplitude, where $f = \frac{B}{N}$MHz.

The rationale for using this training sequence is two-fold: First, the square wave has a unique solution to $\{h_0, h_1, \ldots,\ h_{m-1}\}^T$ as long as $a_{-m} = a_{-m+1} = \ldots = a_0 = a_1 + c = \ldots = a_m + c$, where c is a non-zero constant. And second, the spikes it produces in the frequency domain are sparse. For example, with $B = 100$MHz and $N = 16$, the space between neighboring spikes is 12.5MHz. Such sparse spikes are tolerable in wireless systems. For example, in a 20MHz WiFi band using OFDM, as claimed by Flashback [19], existing WiFi systems have a relatively large SNR margin. And because the interference of any such spike is constrained to at most one



(a) Training sequence in time domain

(b) Training sequence in frequency domain

Figure 5.4: Training sequence

subcarrier, the loss of a few bits does not significantly affect decoding, as successful packet transmissions always respect SNR margins.

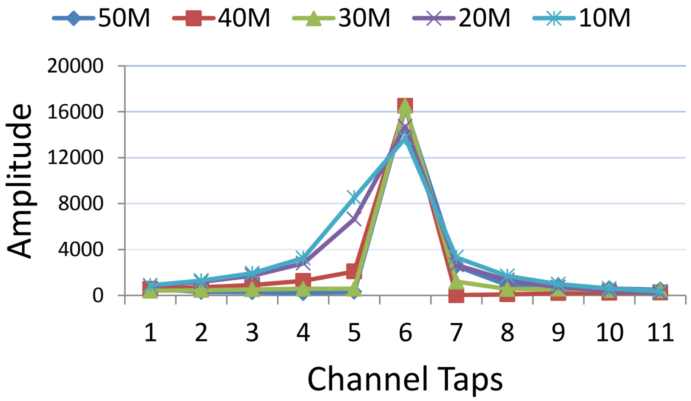
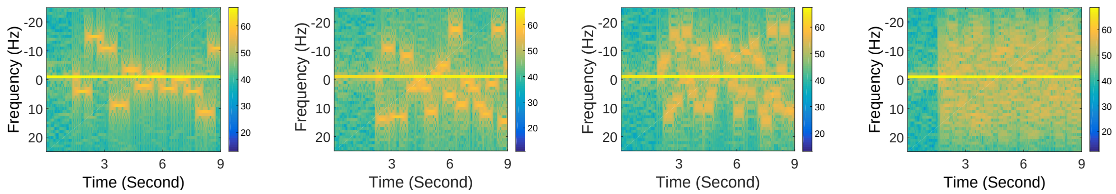


Figure 5.5: Channel coefficients measured at different sampling rates



(a) 1s (b) 0.5s (c) 0.125s (d) 0.05s

Figure 5.6: The granularity of the spectral decreases as the Doppler shifts change from 1s to 0.05s

The training procedure

Training is performed as follows: PhyCloak samples at a rate n times higher than that of external transmission. A training sequence which is the concatenation of consecutive 1s and -1s is sent during training. The received samples corresponding to the transition points (1 to -1 or vice versa) are used to calculate the channel coefficients. More specifically, the received sample b_0 which corresponds to the point right before the transition occurs is equal to $h_0 + \cdots + h_m$, and the next received sample b_1 is equal to $-h_0 + \cdots + h_m$. Thus, we can compute $h_0 = (b_0 - b1)/2$. The rest of the channel coefficients are calculated in a similar way. One concern is whether the desired oversampling rate can be supported. Take 802.11g as an instance, which has the smallest bandwidth (20MHz) among WiFi standards. If training needs a 20X oversampling rate, we would need a platform that supports 400MHz sampling rate. We figure out that, however, a 4X oversampling rate is sufficient to eliminate the effect of an external transmission of 802.11g.

Figure 5.5 plots for different sampling rates the measured channel taps of the self-channel in the same environment. We see that under different sampling rates, the delay spread alway expands across 3 taps. This phenomena follows from the fact: delay spread of non-ultra-wideband is dominated by sampling offset not multipath propagation. The reason is that a multi-path component affects self channel non-negligibly if and only if its propagation path is not too long compared to the direct path between the self transmit and receive antennas. Thus the delay caused by multi-path propagation is small. The reason that the delay spread caused by sampling offset is always 3 is that sampling offset leads to inter-symbol-interference (ISI). Due to ISI, each received sample is affected by not only the intended sample, but also its two neighbors. Therefore the delay spread expands across 3 taps. So as long as we can accurately estimate the three dominant taps in non-ultra-wideband, we can achieve good cancellation performance. That implies we need the external interference to be stable during the reception of at least four consecutive samples at the transition point of the training sequence so as to get the three main taps by differential. Namely 4X oversampling is required.

Note that 4X oversampling does not guarantee the reception of the desired 4 samples happen in the duration of one external interference sample. But we can leverage the interference reduction provided by averaging over multiple transition points, and partially accurate estimation of the channel taps, and still achieve good performance. Even lower oversampling rate (2X/3X) also performs well according to the experiment (see Section 5.4.2).

5.3.2 Obfuscation of Patterns in 3 DoFs

To motivate how we obfuscate patterns in the three DoFs, let us first examine the result of superposing a signal via one path with an obfuscated version via another path.

Assume we have two paths: one with $\{a_1, \Delta f_1, \Delta t_1\}$, and the other via the Ox with $\{a_2, \Delta f_2, \Delta t_2\}$. The superposition of the signals through these two paths is given by the following formula:

$$\begin{aligned} \hat{r}(t) =& a_1 \times s(t) \times e^{j2\pi(f_c+\Delta f_1)(t+\Delta t_1)} \\ &+ a_2 \times s(t) \times e^{j2\pi(f_c+\Delta f_2)(t+\Delta t_2)} \end{aligned} \tag{5.8}$$

Now, is superposing an obfuscated signal sufficient for hiding the original triplet $\{a_1, \Delta f_1, \Delta t_1\}$? The answer is partially yes: The amplitudes and delays are not separable in the superposed signals, but the respective Doppler shifts remain distinguishable after superposition. So, a and Δt can be hidden instantly by randomly changing amplitude and delay of the signal by the Ox.[1] To see why Doppler shifts are distinct even after superposition, consider the frequency response of the received signals:

[1] In theory for a high sampling rate receiver, delays might be separable in the brief prefix that arrives before the obsfuscated signal arrives, but how much information a sensor can accurately extract from the brief clean prefix is questionable.

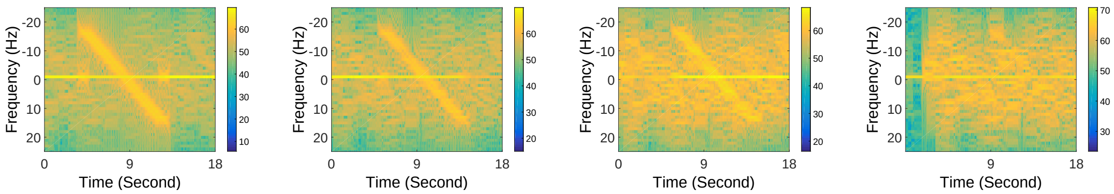
$$\begin{aligned} R(f) =& \int \hat{r}(t) e^{-2\pi j f t} dt \\ =& \int (a_1 \times s(t) \times e^{j2\pi(f_c+\Delta f_1)(t+\Delta t_1)}) e^{-j2\pi f t} dt \\ &+ \int (a_2 \times s(t) \times e^{j2\pi(f_c+\Delta f_2)(t+\Delta t_2)}) e^{-j2\pi f t} dt \\ =& a_1 e^{j2\pi(f_c+\Delta f_1)\Delta t_1} S(f - fc - \Delta f_1) \\ &+ a_2 e^{j2\pi(f_c+\Delta f_2)\Delta t_2} S(f - fc - \Delta f_2) \end{aligned} \tag{5.9}$$

where $S(f)$ is the frequency response of $s(t)$. In an OFDM system, we can see two frequency components that are shifted by Δf_1 and Δf_2 around the subcarrier f.

Doppler shift obfuscation

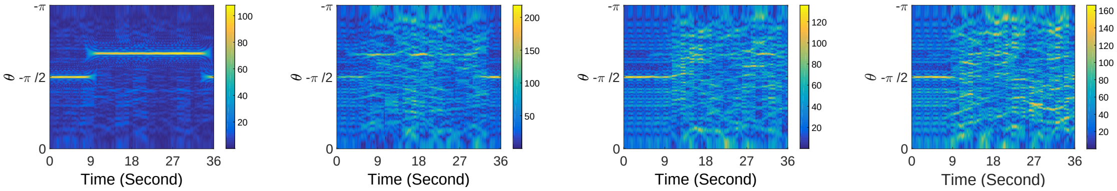
As amplitude and delay can be instantly changed by superposition with an obfuscated signal, patterns that rely only on amplitude and delay can be hidden by Ox, by randomly changing them on a per packet basis. At first glance, it may appear that this scheme cannot be made to work for patterns that rely on Doppler shift, but it turns out the scheme can be made to work for Doppler shift, assuming the moments of change are carefully chosen.

The rationale for choosing the moments of change is based on the fact that a t-second observation in the time domain leads to $1/t$ Hz granularity in the frequency domain. To choose the appropriate Δf at $1/t$ Hz granularity, there is an implicit requirement that the Δf needs to last for at least t seconds. Therefore, if the forwarder changes its Δf every t seconds while the other copy's Δf does not change, an observer would still only see $1/t$ Hz granularity. Since human movements typically result in -20Hz to 20Hz Doppler shifts in the 2.4GHz band, a Doppler shift of the forwarded copy that changes every 0.1s creates sufficient confusion at an observer. Figure 5.6



(a) 0dB (b) -3dB (c) -6dB (d) -9dB

Figure 5.7: The pattern that a WiSee sensor sees in Figures 5.2(a) is hidden by an obfuscated signal where Doppler shift changes every 0.1 second



(a) Motion towards a Wi-Vi style sensor with constant angle (b) 0dB (c) -3dB (d) -6dB

Figure 5.8: The constant angle of human motion (starting from 9th second) that a Wi-Vi style sensor sees in (a) is hidden by an obfuscated signal where phase changes randomly every 0.1 second

shows that when the Doppler shifts of the transmitted signals are varied from every 1s to every 0.05s, the spectral seen by an observer with 1s observation interval have progressively finer granularity, to the point where a time-frequency pattern gets hidden.

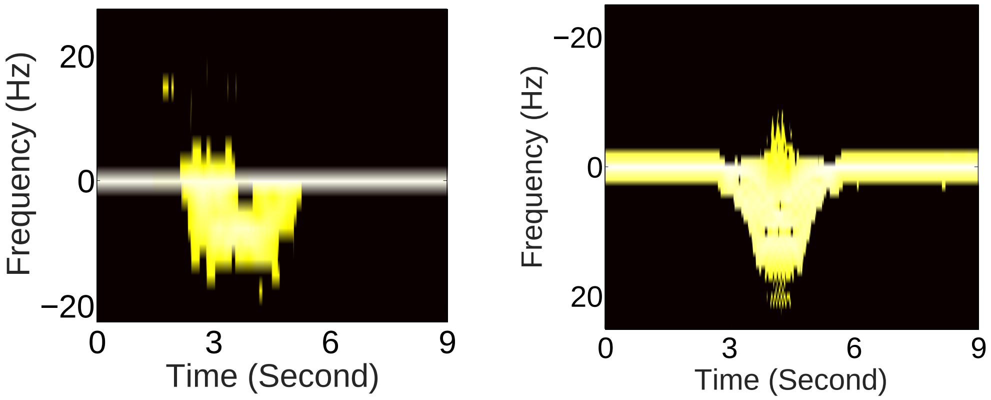
Effect of superposing with randomly changing obfuscated signals

The basic idea of PhyCloak then is to superpose signals from the target with naturally changing $\{a, \Delta f, \Delta\phi\}$ with the obfuscated signals with randomly changing

$\{a, \Delta f, \Delta\phi\}$. More specifically, as analyzed above, PhyCloak changes the value of the triple every 0.1s. We illustrate the blackbox effect of obfuscation experimentally using two state-of-the-art sensors, WiSee [50] and Wi-Vi [8], which we implemented. WiSee performs gesture recognition by extracting Doppler shifts from OFDM symbols, whereas Wi-Vi uses ISAR to track the angle of human motion w.r.p.t. the receive antenna of the sensor.

For the case of obfuscating Doppler shift patterns, Figure 5.7 shows the superposition of a signal with the synthetically generated Doppler shift pattern described in Figure 5.2(a) and an obfuscated copy of the pattern where Doppler shift changes randomly every 0.1s. We see that pattern of Figure 5.2(a) is covered by the “noise map” created by the randomly changing copy. As the strength ratio of the former relative to the latter, which we define as signal to obfuscation ratio (SOR), decreases from 0dB to -9dB, the visibility of the artificial pattern at the WiSee sensor decreases. As Doppler shifts generated by human reflections are small compared to that generated by an Ox that has its own power supply, -9dB higher Doppler shifts would be easily achieved in practice.

For the case of obfuscating phase-based patterns, we synthetically emulated a human moving towards the receive antenna of our Wi-Vi style sensor at a constant angle, as shown in Figure 5.8(a), and then superposed the signal with a randomly obfuscated copy where phase changes every 0.1s. Figure 5.8 shows that as SOR decreases from 0dB to -6dB, the pattern shown in Figure 5.8(a) becomes progressively invisible at the Wi-Vi style sensor.



(a) Doppler shifts created by a human gesture (pull)

(b) Doppler shifts emulated by an Ox

Figure 5.9: Spoofing

5.3.3 Spoofing

According to the above discussion, our design succeeds in obfuscating any RF-based single-antenna sensors by creating false negative results. But an Ox can achieve more than that: it can create false positives also by spoofing changes in the 3 DoFs that are similar to the changes created by a target. By splitting the to-be-forwarded samples into multiple streams, applying different instantiations of the triple $\{a, \Delta f, \Delta t\}$ to them, and forwarding the combination of the processed streams as one stream, an Ox can emulate multiple reflectors corresponding to different parts of the target (say a human body). But unlike the case of false negatives, the effectiveness of creating false positives at a sensor grows as the Ox knows more about the features and algorithms used by the sensor. For example, if an Ox knows a sensor uses the WiSee algorithm [50], it can create a Doppler shift profile accordingly without making an effort to model accurate human movement. Figure 5.9 depicts the extracted Doppler profile of a human gesture (pull) and that spoofed by an Ox. WiSee segments a Doppler profile into positive and negative parts according to its power distribution and encodes them into 1s and -1s respectively. Since both of the profiles contain positive Doppler shifts of negligible power, they will be encoded as -1s and mapped to the same target by a WiSee sensor.

5.3.4 PhyCloak

By obfuscating using random physical distortion, an Ox is able to confuse Eve, and by online maintenance of self-channel estimates, Ox is able to output interference-free signals to Carol for legitimate sensing. However, one critical requirement is still not met: preserving the communication throughput in the presence of Ox.

Although PhyCloak works as a relay at logical layer which improves the throughput, it is not clear that obfuscation would not hurt the decoding process. We find that, however, as long as the change of the triplet $\{a, \Delta f, \Delta\phi\}$ does not happen in the middle of packet transmission, obfuscation is safe with respect to data communication. The reason for this is that from the perspective of a data receiver, the Ox effectively just adds variability to the channel. Since data receivers usually perform channel estimation at the beginning of the received packet, as long as the channel is stable during the reception of the packet, decoding can be successful. We, therefore, refine the design of PhyCloak as follows: PhyCloak switches between two transmitting modes: training and forwarding. In the training phase, the PhyCloak sends the above mentioned training sequence and computes its self-channel estimate according to Section 5.3.1; in the forwarding phase, PhyCloak then performs self-interference cancellation, applies the physical distortion $\{a, \Delta f, \Delta\phi\}$ to the interference-free signal and forwards the distorted signal via the transmit antenna. The PhyCloak randomly chooses an instance of $\{a, \Delta f, \Delta\phi\}$ in the predefined pool and updates the current value when the channel is free and the last update happened more than 0.1s ago. In this way, PhyCloak avoids interfering with the transmission. And in theory, there is still a chance that due to the delay caused by free-channel detection, PhyCloak changes the channel after several samples of a packet has been transmitted, but that chance is quite low. Even if it happens, because PhyCloak only affects a few samples at the beginning, the packet might be still decodable.

5.4 Implementation and Evaluation

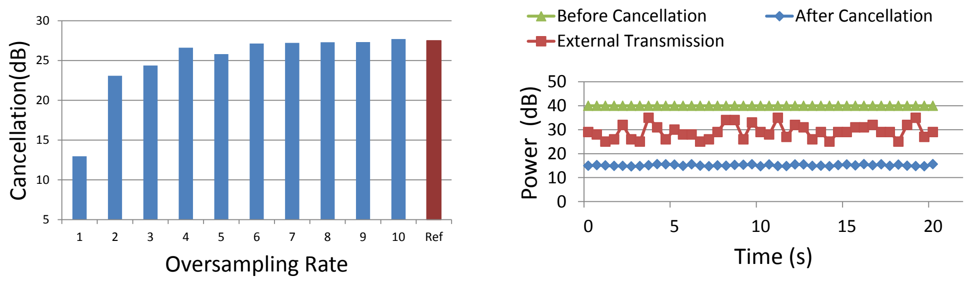
We now describe a prototype of PhyCloak that we have built, and our experiments to validate its performance.

5.4.1 Experimental Setup

Our prototype is based on PXIe 1082 SDR platform. We built the transmitter, receiver, eavesdropping sensor and legitimate sensor on the same platform, which all follow the 802.11g standard, i.e., working at 2.4GHz with a 20MHz band. PhyCloak works at the same center frequency but with a 50MHz sampling rate, about 3 times the rate of an external data transmission, which gives it a reasonable margin to perform self-channel estimation with an ongoing external transmission (see Section 5.4.2).

PhyCloak contains two RF chains, one for transmitting and one for receiving. Each of the RF chains contains an NI-5791 (FlexRIO RF transceiver equipped with one antenna) for transmitting or receiving and an NI PXIe-7965R (a Xilinx Virtex-5 FPGA) for digital processing. Analog cancellation is implemented according to our earlier design [13, 16]. The self-channel estimation, digital cancellation and physical layer distortion are implemented on the FPGA. The distortion processing introduces a latency of about 100ns. Our experiments were conducted in a 5m×7m lab.

5.4.2 Self-Interference Cancellation



(a) Cancellation performance of square wave based training increases when oversampling rate increases from 1 to 4

(b) Insensitivity of square wave based training to external transmission power variation, which is necessary for preserving legitimate amplitude based sensing

Figure 5.10: Self-interference cancellation performance

We begin with the performance of the digital cancellation of our self-channel estimation algorithm. As discussed in Section 5.3.1, Ox tolerates external interference during self-channel estimation using oversampling. So, we first examine the oversampling rate needed to achieve reasonably accurate self-channel estimates in the presence of external transmission. We let a full-duplex transceiver operate at 50MHz with a 10-tap filter for self-interference (digital) cancellation. Self-channel estimation is obtained by averaging over 128 training rounds, which altogether takes about 20μs.

Figure 5.10(a) plots the self-interference cancellation performance of our square wave-based training. In the figure, as we fixed the sampling rate of the full-duplex radio (50MHz), different oversampling rates correspond to different external transmission rates with the received power of the external transmissions being the same as that of self-interference signal at Ox's receive antenna[2]. 1X oversampling rate corresponds to the case when the training and data communication use the same sampling rate, in which case square wave-based training and traditional pilot based training would achieve similar performance. We see that the performance of self-interference cancellation of square wave based training increases as the oversampling rate increases from 1 to 4, but it stops increasing after 4, and achieves similar performance as that in the case when there is no external transmission going on (indicated by the red bar). It shows that Ox can reliably estimate and cancel self-interference even in the presence of strong external transmission when the oversampling parameter is 4X as supported by our observation in Section 5.3.1. In addition, 2X and 3X oversampling rates also produce high cancellation as they benefit from two factors: 1) accurate estimation of part of the channel taps, and 2) averaging over multiple transition points. **Takeaway:** *Our oversampling technique makes self-interference cancellation reliable at modest oversampling rates even in the presence of strong ongoing external transmission.*

The analysis above considers external interference sent at a fixed power. To enable legitimate sensing, self-interference cancellation performance needs to be stable even when the received power from external transmission is varying. For example, an unstable self-interference canceler can render an amplitude-based sensor useless since the (varying) residual self-interference will affect the received signal amplitude. Figure 5.10(b) plots the full-duplex radio's cancellation performance with 3X oversampling rate over time during which the received power from the external transmitter

[2] Note that this is a very strong external interference and we choose this setting to show oversampling strategy's performance even under strong external interference.

fluctuates. We see that the self-interference cancellation performance of square wave-based training is insensitive to the variation of external interference. **Takeaway:** *Our oversampling technique results in a stable cancellation performance at modest oversampling rates even when the received signal from external transmitter is varying.*

5.4.3 Obfuscation Performance

Obfuscation vs. SOR in 3 DoFs

We first measure the different levels of obfuscation created by PhyCloak by comparing the correlation of the amplitude, phase and Doppler shift with and without the presence of PhyCloak. The transmitter is programmed to send continuous OFDM symbols with QPSK modulation; the amplitude and phase thus stay fixed during reception when PhyCloak is absent. An artificial Doppler shift of 10Hz is also added at the transmitter. PhyCloak performs obfuscation by randomly changing the amplitude, phase and Doppler shifts every 0.1s.

Figure 5.11 depicts the correlation between the pairs of amplitude, phase, and Doppler shifts at different SORs: Again, SOR is the signal strength ratio of original signal over obfuscation signal (see Section 5.3.2). We see that as SOR increases, the correlation of each pair of the three features increases, i.e., the obfuscation degree decreases. Amplitude sequence pair and phase sequence pair see lower correlation than Doppler shift pair when SOR is high. This is because amplitude and phase are instantaneous quantities, while Doppler is a statistical quantity that is derived from multiple instantaneous samples. But, even for Doppler shift, a 10dB SOR is low enough to hide the patterns contained in signals reflected by targets. It's worth noting that in practice, as PhyCloak is independently powered while the target only

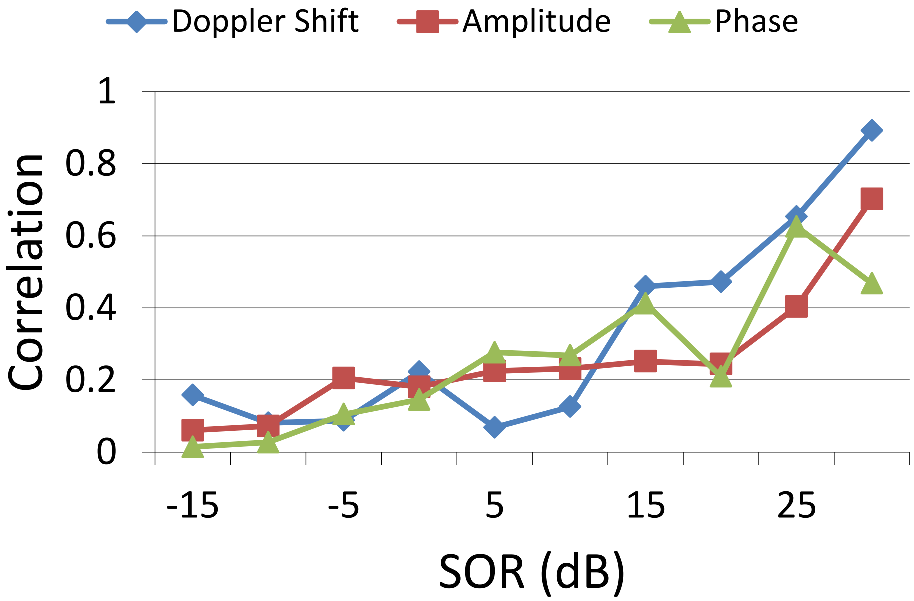


Figure 5.11: Obfuscation level of each of the three features decreases as SOR (original signal over obfuscation signal) increases

passively reflects signals, the desired SOR to successfully obfuscate is readily achieved.

Takeaway: *PhyCloak effectively obfuscates sensing even at a relatively high SOR.*

As different sensors differ in their robustness to noise, PhyCloak 's effectiveness is sensor dependent. While we are unaware of any research on the robustness of the communication-based sensors, we may infer from Figures 5.7, 5.8 and 5.11 that less obfuscation power is needed to confuse a phase or amplitude based sensor as compared to a Doppler shift based sensor. Therefore, we choose to validate the PhyCloak 's capability of confusing illegitimate sensing and preserving legitimate sensing in the context of WiSee, which is the state-of-the-art Doppler shift-based sensor.

Degradation of illegitimate sensing

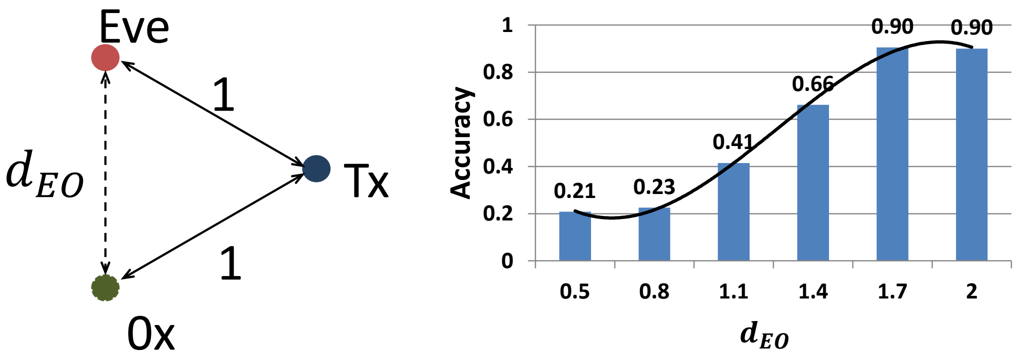
We built a Doppler-based sensor in our platform per the method proposed by WiSee [50]. The method consists of two parts: 1) extraction of Doppler shifts from repeated OFDM symbols by applying a large size FFT; and 2) using sequence matching to classify gestures. We note since we could not get to the original WiSee code and some of the details are missing, we implemented WiSee with a few adaptations. For example, we randomly map the sequence to the predefined classes with uniform distribution in case the sequence does not match any of the predefined sequence. Our implementation shows a classification accuracy of 93% across 5 gestures in none-line-of-sight (NLoS) setting with the human target 5 feet away from the WiSee sensor, while WiSee reports 94% across 9 gestures. While there is this small discrepancy in replication, the core algorithm is the same and our main goal is to study obfuscation performance.

We examine the performance of an illegitimate WiSee sensor with obfuscation from a PhyCloak. We conduct two sets of experiments to validate PhyCloak 's coverage range and its overall effectiveness under different channel conditions respectively.

Obfuscation coverage: First, we randomly choose 10 pairs of locations to place Tx and Eve, and then place Ox in locations such that the distance d_{TE} between Tx and Eve is equal to the distance d_{TO} between Tx and Ox as shown in Figure 5.12(a), but the distance d_{EO} between Eve and Ox varies from $0.5d_{TE}$ to $2d_{TE}$. The channels between any two of the three parties are line-of-sight (LoS).[3] A human target performs

[3] WiSEE sensors have a slightly worse performance in LoS ($\approx 90\%$) than NLoS ($\approx 93\%$) as strong direct power from the transmitter hides the information provided by target's reflection. For the next two experiments, we choose LoS instead of NLoS because it makes the placement easier to make sure d_{EO} is the only variable which would change the power ratio of the obfuscation and human reflection.

five gestures drag, push, pull, circle and dodge close to Eve. With no obfuscation, Eve's classification accuracy in this placement is about 90% across the five gestures.

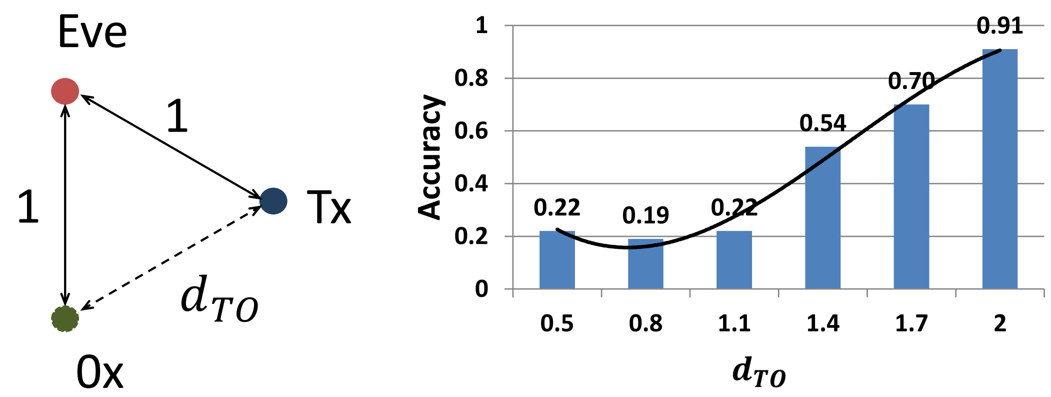


(a) Placement of Tx, Ox and Eve with all three channels LoS

(b) Classification accuracy of Eve in the presence of PhyCloak increases as d_{EO} increases

Figure 5.12: Eve's classification accuracy vs d_{EO}

For simplicity, we normalize d_{EO} by d_{TO} (d_{TE}), and plot the classification accuracy against the normalized d_{EO} in Figure 5.12(b). As we know, the received obfuscation power at Eve from Ox is a function of d_{TO} and d_{OE}, therefore as d_{EO} increases the power ratio of obfuscation over human reflection decreases. From the figure we see that classification accuracy of Eve increases as d_{EO} increases as expected. Note that since we have 5 classes, a classification accuracy of 0.2 means a random guess. PhyCloak can obfuscate Eve near perfectly when d_{EO} is smaller than 0.8, and it totally fails when it is larger than 1.7. **Takeaway:** *The closer Ox is to Eve, the better the achieved obfuscation.*



(a) Placement of Tx, Ox and Eve with all three channels LoS.

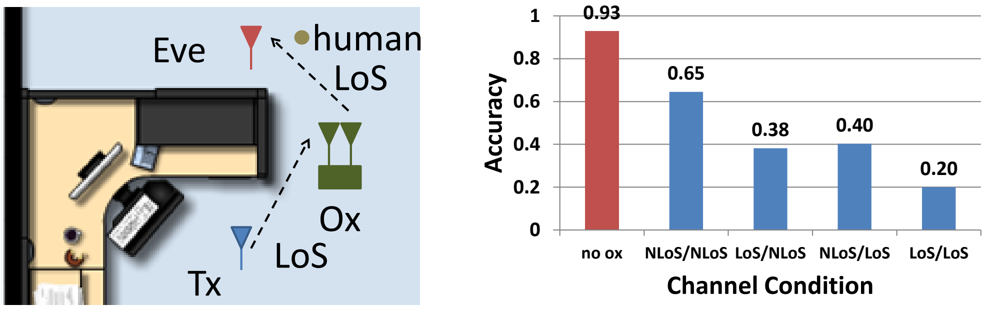
(b) Classification accuracy of Eve in the presence of Phy-Cloak increases as d_{TO} increases

Figure 5.13: Eve's classification accuracy vs d_{TO}

In the second experiment, we make $d_{TE} = d_{OE}$, and vary d_{TO} as shown in Figure 5.13(a). And again in Figure 5.13(b), we see that as d_{TO} increases, Eve's classification accuracy increases. **Takeaway:** *the closer Ox is to Tx, the better teh achieved obfuscation.*

In other experiments we vary either the human-Eve or human-Ox distance while keeping the power recevied by Ox and human from Tx stay constant. As these distances respectively reduced, the effectiveness of the sensing and obfuscation respectively increased.

Obfuscation effectiveness under different channel conditions: In addition to the coverage range in LoS setting, we also measure Eve's classification accuracy when channels between the transmitter and obfuscator and the channel between the obfuscator and Eve are under different LoS and NLoS combinations. Intuitively, when both channels are NLoS, Eve receives the least power forwarded by the obfuscator, and therefore, she achieves the best performance. We care about these channel conditions



(a) Example of LoS/LoS placement in the lab

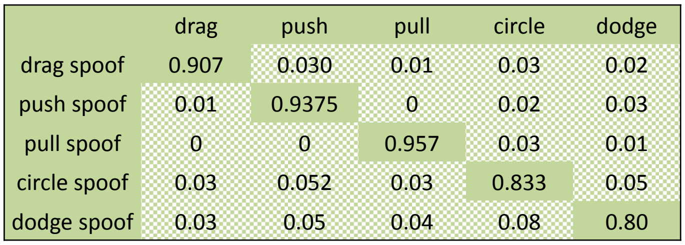
(b) Obfuscation degrades somewhat if Tx-Ox or Ox-Eve channels are NLoS

Figure 5.14: Eve's classification accuracy under different Tx-Ox and Ox-Eve channel conditions

because in some scenarios the transmitter is under control of the adversary, and therefore the adversary may enjoy the freedom to create "good" channels to mitigate PhyCloak 's obfuscation.

In the experiment, we make the channel between Tx and Eve NLoS, and the channel between Tx and the human and that between human and Eve LOS, so as to make sure Eve sees high classification accuracy when no obfuscation is going on. The channel between Tx and Ox and the channel between Ox and Eve have four possible channel condition combinations. A human target performs 500 times of the 5 predefined gestures near Eve in each of the four combinations. Figure 5.14(a) is an example of how we create a channel combination of Los/Los in the lab, where the first LoS refers to the channel condition of the channel between Tx and Ox, while the second refers to that of the channel between Ox and Eve. NLoS channels are created by placing obstacles in the direct propagation paths.

Figure 5.14(b) depicts Eve's classification accuracy without obfuscator and with obfuscator in four channel combinations. We can see that as expected, Eve sees the highest classification accuracy (65%) in NLoS/NLoS setting among the four channel conditions, but it is still smaller than the case when no obfuscation is happening (93%). Eve sees similar performance in Los/NLoS and NLoS/LoS scenarios as power forwarded by obfuscator in both the settings is similar. **Takeaway:** *although NLoS channel degrades the received power at Eve from Ox, the degradation is not dramatic since there is rich multipath propagation in indoor environment.*



	drag	push	pull	circle	dodge
drag spoof	0.907	0.030	0.01	0.03	0.02
push spoof	0.01	0.9375	0	0.02	0.03
pull spoof	0	0	0.957	0.03	0.01
circle spoof	0.03	0.052	0.03	0.833	0.05
dodge spoof	0.03	0.05	0.04	0.08	0.80

Figure 5.15: False positives with a spoofing Ox

Feasibility of spoofing

We built a spoofing obfuscator by reverse engineering the five predefined sequences corresponding to the five gesture types that our WiSee sensor recognizes. The basic difference between this spoofing obfuscator and PhyCloak is that the former changes Doppler shift according to the five well-defined gestures, while the latter changes Doppler shift randomly. The result is shown in Figure 5.15. **Takeaway:** *the spoofing obfuscator fools a WiSee sensor with a high success rate, averaging 88.69% across the 5 gestures, in the absence of human gesturing.*

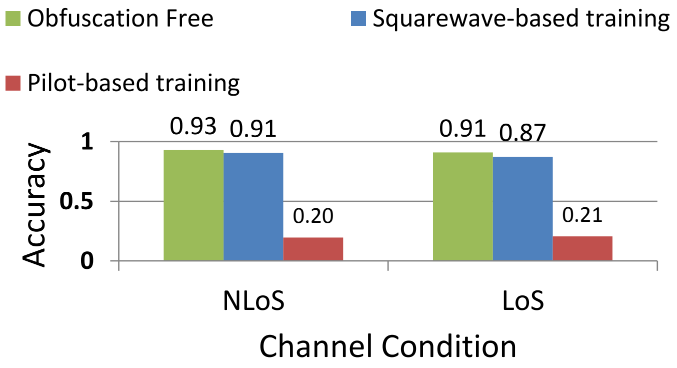


Figure 5.16: Square wave based training preserves legitimate sensing

Preservation of legitimate sensing

Next, we examine PhyCloak 's capability of supporting coupled legitimate sensing. That is, we evaluate whether our self-channel estimation method produces consistent and sufficient self-interference cancellation in a changing environment to preserve legitimate sensing. Figure 5.16 depicts the legitimate sensor's classification accuracy for three different sensing modes: 1) obfuscation free sensing; 2) legitimate WiSee sensing coupled with a PhyCloak module that uses the proposed square waved based self-channel estimation; 3) legitimate WiSee sensing coupled with a PhyCloak module that uses traditional pilot based self-channel estimation. We also vary the channel between Tx and the legitimate sensor by placing and removing obstacles. From the figure we see that the WiSee sensor equipped with PhyCloak module that uses square wave based training achieves comparable performance as obfuscation-free sensing in both LoS and NLoS, while the WiSee sensor equipped with PhyCloak module that uses traditional training fails dramatically. This is because not enough self-interference cancellation is achieved in the presence of external transmissions using extant self-channel estimation techniques. **Takeaway:** *Square wave based training provides sufficient self-interference cancellation to preserve legitimate sensing with external transmission going on.*

5.4.4 Throughput Performance

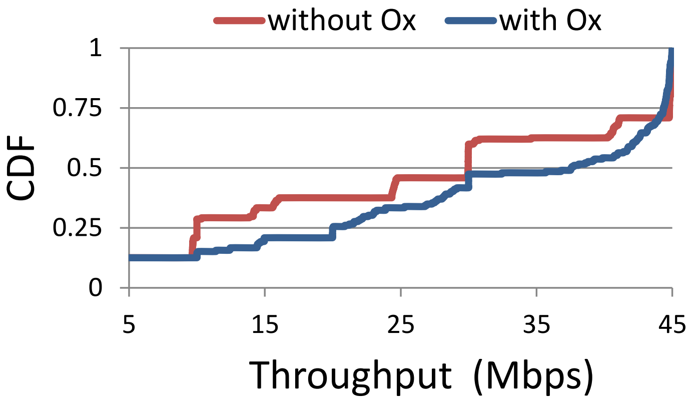


Figure 5.17: Throughput

As discussed in Section 5.3.4, PhyCloak would not hurt the average throughput by virtue of being a relay as long as it avoids parameter changes in the middle of packet transmissions. And, its online training would introduce some interference albeit of small measure. To validate that the net throughput benefit that a data receiver obtains from PhyCloak is not affected but even improved, we measured the throughput performance of a data link with and without PhyCloak in our testbed. 20 location triples were randomly picked for a data transmitter, a data receiver, and PhyCloak. The data transmitter transmits back-to-back packets continuously, and we can thus see the throughput performance in the worst case where PhyCloak performs parameter updates in the middle of some packets. Figure 5.17 plots the CDF of the throughput with and without the PhyCloak. **Takeaway:** *The average throughput increases with the help of PhyCloak.*

5.5 Discussion

(a) Limitations of PhyCloak

While the design of PhyCloak is effective across the whole three dimensions of physical information in the signal, we note several limitations. First, PhyCloak design doesn't taken into account the exploitation of angle-of-arrival (AoA) information which can be inferred by multi-antenna eavesdropper. Second, the current design is for a single node which has limited power supply both due to the device and the cancellation performance. Thus, to extend the coverage to the large target area, it is necessary to consider the collaborative network of multiple PhyCloak nodes. Besides, although the support for legal sensing service is attractive, the service is restricted to the same obfuscation node. The co-located constraint limits the availability and flexibility of the service.

5.6 Summary

We have shown that the threat created by recent developments in communication based sensing can be countered in a black-box fashion. PhyCloak obfuscates multi-dimensional physical signatures of human targets. We have empirically validated this for certain state-of-the-art sensors. We have also shown that when white box details of particular sensors can be obtained, PhyCloak can be refined to spoof those sensors. Notably, the methodology not only preserves but in fact improves the link throughput of the ongoing data transmissions, and supports co-existence of legitimate sensors while obfuscating illegitimate sensors.

Looking beyond the scope of the present work, we find that the methodology is readily generalized to protect against sensing of other types of physical targets and their properties, and allows for a network of PhyCloak devices to collaboratively cover a large region, the details of which are topics for future studies. In addition, when we extend our current single-antenna PhyCloak to a multiple-antenna system, how to fully exploit the space diversity provided by the multiple antennas is worth studying.

Chapter 6: Final Remarks

In this dissertation, we study the problem of sensing and anti-sensing with wireless communication signal. In particular, we consider two different application scenarios in wireless sensing and a generally-effective privacy protection strategy for one-antenna eavesdropper.

First, we propose Mudra to enable fine-grain finger-gesture recognition with WiFi signal. To detect the fine-grain motion of human fingers, we design a highly-sensitive motion indication utilizing the sensitivity of cancellation performance. Second, we enable a human activity recognition system, TifWiFi with an integration framework on two profiles. TifWiFi enhances the recognition performance over conventional approach by overcoming the drawback of each individual profile and utilizing the complementation of them. Lastly, we propose PhyCloak to combat against any kind of one-antenna eavesdropper over the information in three degrees of freedom. PhyCloak goes beyond the simple jamming strategy by only obfuscating the physical-layer information while preserving the logical-layer information.

For each of the above systems, we implement a prototype and conduct extensive experiment to evaluate the performance over various conditions. The results validate the effectiveness of our proposed designs. We also note that for each system there exist limitations. For example, the gesture recognition receiver can be confused if there is surrounding moving body, which is much intenser than finger-motion. As for the PhyCloak system, a multi-antenna eavesdropper can work with the angle-of-arrival information, which is out of the three degrees of freedom. These more challenging topics would be interesting for us to study in the future.

Bibliography

[1] Findlay Shearer. Power management in mobile devices, chapter Hierarchical View of Energy Conservation, pages 32-75. Newnes, 2008.

[2] Google project soli. https://www.youtube.com/watch?v=0qnizfsspc0.

[3] Leap motion. https://www.leapmotion.com.

[4] Microsoft. x-box kinect. http://www.xbox.com.

[5] Philips lifeline. http://www.lifelinesys.com/content/.

[6] TPLINK. TL-WN722N. https://wikidevi.com/wiki/TP-LINK_TL-WN722N.

[7] Fadel Adib and Din Katabi. Wi-Vi: See Through Walls with WiFi! In *ACM SIGCOMM*. 2013.

[8] Fadel Adib and Dina Katabi. *See through walls with WiFi!*, volume 43. ACM, 2013.

[9] Fadel Adib, Hongzi Mao, Zachary Kabelac, Dina Katabi, and Robert C Miller. Smart Homes that Monitor Breathing and Heart Rate. In *ACM CHI*. 2015.

[10] Jake K Aggarwal and Michael S Ryoo. Human activity analysis: A review. *ACM Computing Surveys (CSUR)*, 43(3):16, 2011.

[11] Kamran Ali, Alex Xiao 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. ACM, 2015.

[12] Dinesh Bharadia and Sachin Katti. Fastforward: fast and constructive full duplex relays. In *Proceedings of the 2014 ACM conference on SIGCOMM*, pages 199–210. ACM, 2014.

[13] Dinesh Bharadia, Emily McMilin, and Sachin Katti. Full duplex radios. In *ACM SIGCOMM Computer Communication Review*, volume 43, pages 375–386. ACM, 2013.

[14] Igal Bilik and Joseph Tabrikian. Radar target classification using doppler signatures of human locomotion models. *IEEE Transactions on Aerospace and Electronic Systems*, 43(4):1510–1522, 2007.

[15] Bo Chen, Yue Qiao, Ouyang Zhang, and Kannan Srinivasan. Airexpress: Enabling seamless in-band wireless multi-hop transmission. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, pages 566–577. ACM, 2015.

[16] Bo Chen, Vivek Yenamandra, and Kannan Srinivasan. Flexradio: Fully flexible radios. *The Ohio State University, Tech. Rep*, 2013.

[17] Bo Chen, Vivek Yenamandra, and Kannan Srinivasan. Tracking keystrokes using WiFi. In *Proceedings of ACM MobiSys*, 2015.

[18] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

[19] Asaf Cidon, Kanthi Nagaraj, Sachin Katti, and Pramod Viswanath. Flashback: decoupled lightweight wireless control. *ACM SIGCOMM Computer Communication Review*, 42(4):223–234, 2012.

[20] Theodoros Damoulas, Jin He, Rich Bernstein, Carla P Gomes, and Anish Arora. String kernels for complex time-series: Counting targets from sensed movement. In *2014 22nd International Conference on Pattern Recognition (ICPR)*, pages 4429–4434. IEEE, 2014.

[21] Nicola Dell, Krittika D'Silva, and Gaetano Borriello. Mobile Touch-Free Interaction for Global Health. In *ACM HotMobile*. 2015.

[22] Evan Everett, Anant Sahai, and Ashutosh Sabharwal. Passive self-interference suppression for full-duplex infrastructure nodes. *IEEE Transactions on Wireless Communications*, 2014.

[23] Shyamnath Gollakota, Haitham Hassanieh, Benjamin Ransford, Dina Katabi, and Kevin Fu. They can hear your heartbeats: non-invasive security for implantable medical devices. *ACM SIGCOMM Computer Communication Review*, 41(4):2–13, 2011.

[24] Shyamnath Gollakota and Dina Katabi. Zigzag decoding: combating hidden terminals in wireless networks. In *ACM SIGCOMM*. 2008.

[25] Alejandro Gonzalez-Ruiz, Alireza Ghaffarkhah, and Yasamin Mostofi. An integrated framework for obstacle mapping with see-through capabilities using laser and wireless channel measurements. *Sensors Journal, IEEE*, 14(1):25–38, 2014.

[26] Jin He and Anish Arora. A regression-based radar-mote system for people counting. In *International Conference on Pervasive Computing and Communications (PerCom), 2014 IEEE*, pages 95–102, March 2014.

[27] Chih-Wei Huang and Kun-Chou Lee. Application of ica technique to pca based radar target recognition. *Progress In Electromagnetics Research*, 105:157–170, 2010.

[28] Kinjal A Joshi and Darshak G Thakore. A survey on moving object detection and tracking in video surveillance system. *International Journal of Soft Computing and Engineering*, 2(3):44–48, 2012.

[29] Kiran Joshi, Dinesh Bharadia, Manikanta Kotaru, and Sachin Katti. Wideo: Fine-grained device-free motion tracing using RF backscatter. In *USENIX NSDI*. 2015.

[30] Wei Wang Kamran Ali, Alex X. Liu and Muhammad Shahzad. Keystroke Recognition Using WiFi Signals. In *ACM MobiCom*. 2015.

[31] Matthew Keally, Gang Zhou, Guoliang Xing, Jianxin Wu, and Andrew Pyles. Pbn: towards practical activity recognition using smartphone-based body sensor networks. In *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems*, pages 246–259. ACM, 2011.

[32] Bryce Kellogg, Vamsi Talla, and Shyamnath Gollakota. Bringing gesture recognition to all devices. In *USENIX NSDI*. 2014.

[33] Youngwook Kim and Hao Ling. Human activity classification based on micro-doppler signatures using an artificial neural network. In *Antennas and Propagation Society International Symposium, 2008. AP-S 2008. IEEE*, pages 1–4. IEEE, 2008.

[34] Youngwook Kim and Hao Ling. Human activity classification based on micro-doppler signatures using a support vector machine. *IEEE Transactions on Geoscience and Remote Sensing*, 47(5):1328–1337, 2009.

[35] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[36] Vinit Kizhakkel, Rajiv Ramnath, and at el. Pulsed doppler radar target recognition based on micro-doppler signatures using wavelet analysis. In *IEEE High Performance Extreme Computing Conference (HPEC)*, September 2014.

[37] Manikanta Kotaru, Kiran Joshi, Dinesh Bharadia, and Sachin Katti. Spoton: Indoor localization using commercial off-the-shelf WiFi nics. In *IPSN*. 2015.

[38] Manikanta Kotaru, Kiran Joshi, Dinesh Bharadia, and Sachin Katti. Spotfi: Decimeter level localization using wifi. In *ACM SIGCOMM Computer Communication Review*, volume 45, pages 269–282. ACM, 2015.

[39] Swarun Kumar, Stephanie Gil, Dina Katabi, and Daniela Rus. Accurate indoor localization with zero start-up cost. In *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*, pages 483–494. ACM, 2014.

[40] Kun-Chou Lee, Jhih-Sian Ou, and Ming-Chung Fang. Application of svd - reduction technique to pca based radar target recognition. *Progress In Electromagnetics Research*, 81:447–459, 2008.

[41] Xiang Li, Shengjie Li, Daqing Zhang, Jie Xiong, Yasha Wang, and Hong Mei. Dynamic-music: accurate device-free indoor localization. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 196–207. ACM, 2016.

[42] Wenguang Mao, Jian He, and Lili Qiu. Cat: high-precision acoustic motion tracking. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, pages 69–81. ACM, 2016.

[43] Sarfaraz Masood, Adhyan Srivastava, Harish Chandra Thuwal, and Musheer Ahmad. Real-time sign language gesture (word) recognition from video sequences using cnn and rnn. In *Intelligent Engineering Informatics*, pages 623–632. Springer, 2018.

[44] Yasamin Mostofi. Cooperative wireless-based obstacle/object mapping and see-through capabilities in robotic networks. *IEEE Transactions on Mobile Computing*, 12(5):817–829, 2013.

[45] Rajalakshmi Nandakumar, Krishna Kant Chintalapudi, Venkat Padmanabhan, and Ramarathnam Venkatesan. Dhwani: secure peer-to-peer acoustic NFC. In *ACM SIGCOMM Computer Communication Review*, volume 43, pages 63–74. ACM, 2013.

[46] Rajalakshmi Nandakumar, Shyamnath Gollakota, and Nathaniel Watson. Contactless Sleep Apnea Detection on Smartphones. In *ACM MobiSys*. 2015.

[47] Rajalakshmi Nandakumar, Vikram Iyer, Desney Tan, and Shyamnath Gollakota. FingerIO: Using Active Sonar for Fine-Grained Finger Tracking. In *ACM CHI*. 2016.

[48] Jeffrey A Nanzer and Robert L Rogers. Bayesian classification of humans and vehicles using micro-doppler signals from a scanning-beam radar. *Microwave and Wireless Components Letters, IEEE*, 19(5):338–340, 2009.

[49] Gian Paolo Perrucci, Frank HP Fitzek, and Jörg Widmer. Survey on energy consumption entities on the smartphone platform. In *IEEE Vehicular Technology Conference*. 2011.

[50] Qifan Pu, Sidhant Gupta, Shyamnath Gollakota, and Shwetak Patel. Whole-home gesture recognition using wireless signals. In *Proceedings of the 19th annual international conference on Mobile computing & networking*, pages 27–38. ACM, 2013.

[51] Hariharan Shankar Rahul, Swarun Kumar, and Dina Katabi. JMB: scaling wireless capacity with user demands. In *ACM SIGCOMM*. 2012.

[52] RG Raj, VC Chen, and R Lipps. Analysis of radar human gait signatures. *Signal Processing, IET*, 4(3):234–244, 2010.

[53] Haşim Sak, Andrew Senior, and Françoise Beaufays. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In *Fifteenth annual conference of the international speech communication association*, 2014.

[54] Mike Schuster and Kuldip K Paliwal. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11):2673–2681, 1997.

[55] Moustafa Seifeldin, Ahmed Saeed, Ahmed E Kosba, Amr El-Keyi, and Moustafa Youssef. Nuzzer: A large-scale device-free passive localization system for wireless environments. *IEEE Transactions on Mobile Computing*, 12(7):1321–1334, 2013.

[56] Graeme E Smith, Karl Woodbridge, and Chris J Baker. Radar micro-doppler signature classification using dynamic time warping. *IEEE Transactions on Aerospace and Electronic Systems*, 46(3):1078–1096, 2010.

[57] Elahe Soltanaghaei, Avinash Kalyanaraman, and Kamin Whitehouse. Multipath triangulation: Decimeter-level wifi localization and orientation with a single unaided receiver. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, pages 376–388. ACM, 2018.

[58] Statista. Iot connected devices. https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/. 2018.

[59] Bo Tan, Karl Woodbridge, and Kevin Chetty. A real-time high resolution passive WiFi Doppler-radar and its applications. In *International Radar Conference*. 2014.

[60] Sheng Tan and Jie Yang. Wifinger: leveraging commodity wifi for fine-grained finger gesture recognition. In *Proceedings of the 17th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pages 201–210. ACM, 2016.

[61] Thayananthan Thayaparan, Sumeet Abrol, Edwin Riseborough, LJ Stankovic, Denis Lamothe, and Grant Duff. Analysis of radar micro-doppler signatures from experimental helicopter and human data. *IET Radar, Sonar & Navigation*, 1(4):289–299, 2007.

[62] Sonali Vaidya and Kamal Shah. Real time video surveillance system. *International Journal of Computer Applications*, 86(14), 2014.

[63] Raghav H Venkatnarayan, Griffin Page, and Muhammad Shahzad. Multi-user gesture recognition using wifi. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, pages 401–413. ACM, 2018.

[64] Martin Vuagnoux and Sylvain Pasini. Compromising Electromagnetic Emanations of Wired and Wireless Keyboards. In *USENIX SSYM*. 2009.

[65] 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*, pages 593–604. ACM, 2014.

[66] Junjue Wang, Kaichen Zhao, Xinyu Zhang, and Chunyi Peng. Ubiquitous keyboard for small mobile devices: harnessing multipath fading for fine-grained keystroke localization. In *ACM MobiSys*. 2014.

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

[68] Wei Wang, Alex X Liu, and Ke Sun. Device-free gesture tracking using acoustic signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, pages 82–94. ACM, 2016.

[69] Xuyu Wang, Lingjun Gao, Shiwen Mao, and Santosh Pandey. Csi-based fingerprinting for indoor localization: A deep learning approach. *IEEE Transactions on Vehicular Technology*, 66(1):763–776, 2017.

[70] 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*, pages 617–628. ACM, 2014.

[71] Wikipedia. Near and far field — wikipedia, the free encyclopedia, 2015. [Online; accessed 20-October -2015].

[72] Wikipedia contributors. Doppler effect — Wikipedia, the free encyclopedia, 2018. [Online; accessed 18-November-2018].

[73] Wikipedia contributors. Universal software radio peripheral — Wikipedia, the free encyclopedia, 2018. [Online; accessed 18-November-2018].

[74] Jiang Xiao, Kaishun Wu, Youwen Yi, Lu Wang, and Lionel M Ni. Pilot: Passive device-free indoor localization using channel state information. In *Distributed computing systems (ICDCS), 2013 IEEE 33rd international conference on*, pages 236–245. IEEE, 2013.

[75] Yang Xiao. IEEE 802.11 n: enhancements for higher throughput in wireless LANs. *IEEE Wireless Communications*, 2005.

[76] Yaxiong Xie, Zhenjiang Li, and Mo Li. Precise power delay profiling with commodity wifi. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, MobiCom '15, page 53–64, New York, NY, USA, 2015. ACM.

[77] Jie Xiong and Kyle Jamieson. Arraytrack: a fine-grained indoor location system. Usenix, 2013.

[78] Zhengya Xu and Hong Ren Wu. Smart video surveillance system. In *Industrial Technology (ICIT), 2010 IEEE International Conference on*, pages 285–290. IEEE, 2010.

[79] Lei Yang, Yekui Chen, Xiang-Yang Li, Chaowei Xiao, Mo Li, and Yunhao Liu. Tagoram: real-time tracking of mobile rfid tags to high precision using cots devices. In *Proceedings of the 20th annual international conference on Mobile computing and networking*, pages 237–248. ACM, 2014.

[80] Koji Yatani and Khai N Truong. Bodyscope: a wearable acoustic sensor for activity recognition. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 341–350. ACM, 2012.

[81] Sangki Yun, Yi-Chao Chen, and Lili Qiu. Turning a Mobile Device into a Mouse in the Air. In *ACM MobiSys*. 2015.

[82] Rui Zhou, Jiesong Chen, Xiang Lu, and Jia Wu. Csi fingerprinting with svm regression to achieve device-free passive localization. In *A World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2017 IEEE 18th International Symposium on*, pages 1–9. IEEE, 2017.

[83] Gangyi Zhu and Gagan Agrawal. A performance prediction framework for irregular applications. In *2018 IEEE 25th International Conference on High Performance Computing (HiPC)*, pages 304–313. IEEE, 2018.

[84] Gangyi Zhu, Peng Jiang, and Gagan Agrawal. A methodology for characterizing sparse datasets and its application to simd performance prediction. In *Proceedings of the 28th International Conference on Parallel Architectures and Compilation Techniques*. ACM, 2019.

[85] Yanzi Zhu, Yibo Zhu, Ben Y Zhao, and Haitao Zheng. Reusing 60ghz radios for mobile radar imaging. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, pages 103–116. ACM, 2015.