Channel-based Physical Layer Authentication

by

Chengcheng Pei

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2014

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The characteristics of the wireless physical layer can be exploited to complement and enhance traditional security. In this thesis, we study the channel-based physical layer authentication. The authentication problem is formulated as a sequence of hypothesis test problems. By exploiting the time-of-arrivals, received signal strengths, and cyclic-features of the channels, support vector machine (SVM) based authentication schemes, the linear Fisher discriminant analysis (LFDA) based authentication scheme, and the combining scheme are proposed to improve the detection probability and to reduce the false alarm probability. These schemes can reliably authenticate the sender by identifying channels from different users.

In SVM based schemes, the linear and nonlinear SVMs are used to generate classifiers to solve the hypothesis test problems. Using the real channel data measured in a regular office from Utah University, simulation is performed. Simulation results demonstrate that SVM based schemes have lower misdetection probability and false alarm probability than some existing schemes at a cost of extra time complexity and space complexity due to the training stage.

To reduce the space complexity and time complexity during the training stage, LFDA based authentication scheme is proposed. In LFDA based scheme, a linear combination of the channel features is used as the test statistic, which is compared with a threshold to perform authentication. LFDA is used to compute the weights based on some training data. Furthermore, an adaptive threshold scheme (ATS) is proposed to set and adjust the threshold. Simulation results demonstrate that the proposed LFDA based scheme performs better in terms of the sum of misdetection probability and false alarm probability, and the receiver operating characteristic curves, compared with several existing channel-based authentication schemes. Moreover, the analysis of time complexity and space complexity is provided, which shows that the LFDA based scheme is also better than SVM based schemes in terms of space complexity, time complexity, misdetection probability, and false alarm probability.

The misdetection probability and false alarm probability can be reduced greatly by two-user cooperative authentication. The combining scheme is proposed to combine the data from another legitimate user when cooperation is available. The combining scheme is proven to have the capacity to improve the performance at cost of extra communication and computation overhead. The time complexity, space complexity, and communication overhead are analyzed.

Acknowledgements

I would like to thank all the people who made this thesis possible. They are my supervisor and my colleagues in the Broadband Communications Research (BBCR) group, and my families. Without their help and encouragement, I cannot finish my study.

First of all, I gratefully acknowledge my supervisor, Professor Xuemin (Sherman) Shen. He made available his support and aid in numerous ways. He not only helps me develop the academic skills, but also guides me to strive for excellence.

At BBCR group, I would like to thank Professor Jon. W. Mark and Professor Weihua Zhuang for supporting me. I would also thank Dr. Xinshen Zhou, Dr. Xiaohui Liang, Dr. Chengze Lai, Dr. Cunhe Song, Mr. Ning Zhang, Mr. Nan Cheng, Mr. Qinghua Shen. We worked collaboratively at the BBCR group, and we had many discussions to brainstorm and collaborate on interested research topics.

I would like to thank my friends, Bo Liu, Yang Li, Han Xiao, Jiandi Yao, Junyu Lai, Shenlong Tang, Min Tan and so on.

There are many other people whose names are not mentioned here. It does not mean that I have forgotten them or their help. It is a privilege for me to work and share life with so many bright and energetic people. Their talent and friendship have made Waterloo such a great place to live for me.

In addition, grateful acknowledgements are made for Sensing and Processing Across Networks (SPAN) lab in University of Utah, who provided me the data set for simulation.

Finally, I want to thank my family, i.e., my father, my mother, my brother, and my wife. I would never get this far without their support. I thank them for always believing in me and supporting me. Their love and encouragement have been and will always be a great source of inspiration in my life. I would continuingly work hard to fullfil my career goals and never disappoint them.

Dedication

To my family. To my lost youthhood.

Table of Contents

List of Tables viii

List of Figures ix

1 Introduction 1

1.1 Background . . . 1
1.2 Literature Review . . . 3
1.3 Contributions . . . 6
1.4 Outline of the Thesis . . . 9

2 System Model and Problem Formulation 10

2.1 System Model . . . 10
2.2 Problem Formulation . . . 12
2.3 Selection of Features . . . 13
2.3.1 RSS . . . 15
2.3.2 TOA . . . 15
2.3.3 Correlation of Channel Vectors . . . 17
2.3.4 Correlation of Cyclic Feature Vectors . . . 17
2.3.5 Another Two Features (β and θ) . . . 19
2.4 Metrics . . . 20

3 Support Vector Machine Based Authentication Schemes **22**
3.1 SVMs . . . 23
3.2 Simulation of SVM Based Schemes . . . 26
3.2.1 Experimental Channel Measurements . . . 26
3.2.2 Simulation Results of SVM Based Schemes . . . 29
3.3 Complexity Analysis of SVM Based Schemes . . . 31
3.4 Summary . . . 33

4 Fisher Discriminant Analysis Based Authentication Scheme **35**
4.1 Linear Fisher Discriminant Analysis Based Authentication . . . 36
4.2 Performance Analysis and Complexity Analysis . . . 38
4.2.1 Size of Training Set . . . 38
4.2.2 Comparing LFDA Based Scheme with Other Schemes . . . 39
4.2.3 Noise's Effects . . . 40
4.2.4 Complexity Analysis of LFDA Based Scheme . . . 42
4.3 Adaptive Threshold Scheme (ATS) . . . 43
4.3.1 ATS . . . 43
4.3.2 Comparison with SVM Based Schemes . . . 47
4.4 Combining Scheme for Two-User Cooperative Authentication . . . 47
4.4.1 Combining Scheme . . . 48
4.4.2 Simulation Results of Combining Scheme . . . 49
4.4.3 Complexity Analysis of Combining Scheme . . . 49
4.5 Summary . . . 51

5 Conclusion and Future Work **52**

References **54**

List of Tables

1.1 Comparison of physical layer authentication 7

2.1 Complexity comparison of computing λ and λ' 19

3.1 Number of training locations versus $|D_{tr}|/|D|$ 30
3.2 Complexity comparison in training stage 33
3.3 Complexity comparison in testing stage 34

4.1 Comparison with CTS 40
4.2 Complexity comparison in testing stage 43
4.3 Complexity comparison in training stage 43
4.4 Complexity in testing stage 50
4.5 Complexity in training stage 50

List of Figures

1.1 Secrecy capacity 4
1.2 Key generation 5

2.1 System model 11
2.2 System model 2 14
2.3 How to classify two channel vectors 14
2.4 Plot of RSS versus Dis 16
2.5 Plot of TOA versus Dis 16
2.6 $f(\lambda'|\vec{H}_t \neq \vec{H}_A)$ and $f(\lambda'|\vec{H}_t = \vec{H}_A)$ 18
2.7 $f(\lambda|\vec{H}_t \neq \vec{H}_A)$ and $f(\lambda|\vec{H}_t = \vec{H}_A)$ 19
2.8 $f(\lambda|\vec{H}_t \neq \vec{H}_A)$ and $f(\lambda|\vec{H}_t = \vec{H}_A)$ under different Ps 20

3.1 Map of measurement locations in Utah University 27
3.2 Channel power gain in one measurement 27
3.3 Simulation working flow 30
3.4 Room layout for simulation 31
3.5 Comparison of different SVMs in terms of $P_M + P_F$ with different number of training locations 32
3.6 ROC curves of different SVMs when the number of training locations is 15 32

4.1 ROC curves of LFDA based scheme under different number of training locations 38

4.2 $P_M + P_F$ of LFDA based scheme under different number of training locations 39
4.3 ROC curves of LFDA based scheme and NPHT 40
4.4 Comparison between SVM based schemes and LFDA based scheme in terms of $P_M + P_F$ 41
4.5 ROC curves of LFDA based scheme and NPHT under different SNRs . . . 42
4.6 ROC curves of NPHT and LFDA based schemes with optimal threshold and with ATS 45
4.7 Comparison of LFDA schemes with optimal threshold and with ATS in terms of $P_F + P_M$ 46
4.8 Comparison between SVM based schemes and LFDA based scheme in terms of P_D and P_F 47
4.9 Combining scheme 49
4.10 ROC curve of combining scheme 50

Chapter 1

Introduction

In this chapter, the background about the wireless security, especially physical layer security, is reviewed. Our contribution and the outline of this thesis are presented too.

1.1 Background

In the past 20 years, there were a great number of developments in communication and networking technologies. Nowadays, wireless networks play an important role in people's life. However, due to the broadcast nature of wireless communications, the information can be easily eavesdropped or intercepted. That is why security is of paramount importance for these wireless networks. Generally speaking, security means authenticity, non-repudiation, confidentiality, integrity, and availability. These aspects of security are explained below.

- Authenticity, including entity authenticity and data origin authenticity, means mechanisms to establish identity. It is used to verify whether the information comes from legitimate users or not [1]. Another associated concept, non-repudiation means that the transmitter cannot deny having sent the information and the receiver cannot deny having received the information.
- Confidentiality means concealment of information or resources. It guarantees that the information is only disclosed to legitimate users.
- Integrity means trustworthiness of data or resources.

- Availability means the ability of legitimate users to use the information or resource which they want when they need.

Traditionally, the security of wireless networks is addressed through techniques at upper layers, such as cryptography. However, when the devices have limited computational and bandwidth resources, it is complex to implement the cryptography [1]. In reality, there are no absolutely secure cryptography protocols. There are many possible attacks in realistic environments. Attacks on wireless networks can be classified into two main categories: passive attacks and active attacks [1]. Passive attacks include traffic analysis and eavesdropping. Active attacks include denial of service (DoS), Masquerade attack, replay attack, information disclosure, message modification. Masquerade attack is always the first step to conduct other active attacks. Below are some examples of realistic attacks.

- WiFi Protected Setup (WPS) personal identification number (PIN) is susceptible to a brute force attack. A design flaw existing in the WPS specification significantly reduces the time required to brute force the entire PIN. This design flaw is that it is allowed for an attacker to know when the first half of the 8-digit PIN is correct. The lack of a proper lock out policy after a certain number of failed attempts on many wireless routers makes this brute force attack much more feasible. An attacker within range of the wireless access point (AP) may be able to brute force the WPS PIN and retrieve the password for the wireless network, change the configuration of the access point, or cause a denial of service [2].

- Three attacks on the WiFi Protected Access Temporal Key Integrity Protocol (WPA-TKIP) are described in [3]. The first attack is a DoS attack that can be executed by injecting only two frames every minute. The second attack demonstrates how fragmentation of 802.11 frames can be used to inject an arbitrary amount of packets, and it is shown that this can be used to perform a port-scan on any client [3]. The third attack enables an attacker to reset the internal state of the Michael algorithm. It is also shown that this can be used to efficiently decrypt arbitrary packets sent towards a client [3]. Implementation vulnerabilities discovered in some wireless devices are reported in [3].

As a complement to cryptography, physical layer security has drawn more and more attentions in the research community currently [4]. The main idea of physical layer security is to exploit the characteristics of the physical layer to provide security [5]. Compared with the upper layer cryptography schemes [6], the physical layer security has the following advantages: i) it does not rely on the computational hardness of certain problems, such as

the hardness of factoring numbers [7], and ii) it can protect the transmitted signal from being received or decoded in the first place. Note that the physical layer security can be combined with the upper layer security techniques to provide higher security level.

1.2 Literature Review

At a very high level, existing physical layer security works as follows. Assume that a sender and a receiver are interested in exchanging a secret message in the presence of an eavesdropper. Physical layer security exploits that the wireless channel between the sender and receiver experiences unpredictably variations that can only be measured by the sender and the receiver. Thus, one can use these unpredictable variations to communicate a secret message between the sender and the receiver even in the presence of eavesdropper. In physical layer security, various topics have been investigated, including secrecy capacity [8], channel-based key generation [7], and authentication [9], which are explained in detailed below.

The first research issue is the secrecy capacity, which means the maximum rate sent from a wireless node to its destination in the presence of eavesdroppers. More than 60 years ago, even before cryptography was introduced, Shannon introduced the concept of physical layer security which enables two nodes to communicate securely in the presence of eavesdropper without cryptography. The main idea is that if the eavesdropper channel is a degraded version of the legitimate channel, the legitimate channel can exchange secure messages at a nonzero rate. In Figure 1.1, the channel from Bob to Alice is the legitimate channel, and the channel from Bob to Eve is the eavesdropper channel. The the channel capacity from Bob to Alice is C_A. The capacity from Bob to Eve is C_E. If C_A is larger than C_E. The secrecy capacity of Bob is $C_A - C_E$. Otherwise, it is zero. Nowadays, the research in this area focuses on how to use friendly jammer to degrade the eavesdropper channel [10], how to use cooperation and beam forming to increase the secrecy capacity [11], and what is the secrecy capacity in the scenario of multiple users [12].

Another research topic is key generation. Traditional cryptographic techniques need to distribute, refresh, and revoke keys. However, key distribution needs extra bandwidth and infrastructure. For example, if you would like to use symmetrical cryptography to encrypt and decrypt the message. At first, you need to exchange the private keys, which need kind of secret channel. It may also cause key management problems in networks with high number of nodes. So some researchers proposed to make use of the channel between legitimate users to generate keys [11]. This technique is based on two assumptions. The first one is that wireless channel can be measured almost symmetrically between the legitimate

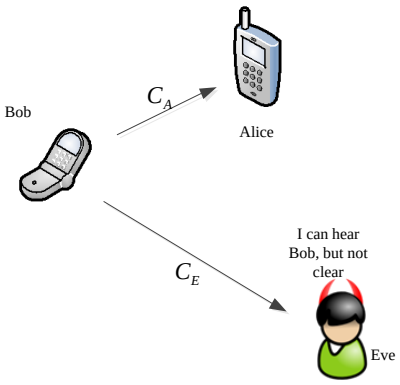

Figure 1.1: Secrecy capacity

transmitter and receiver. In Figure 1.2, Alice and Bob both know the channel between them. And it is nearly impossible for the attacker to measure the legitimate channel, which is the second assumption. So the channel can be considered as a shared secret between the legitimate users in nature. We can use this secret to generate private keys. Key generation in physical layer brings us some advantages, such as less communication overhead. Also because channels always change, the key is hard to be compromised. Nowadays, the research of secret key generation mainly focus on improving the secret bit generation rate by exploiting the characteristics of radio channels, such as temporal and spatial variations of radio channels [13], multiple antenna diversity [14], and multiple frequencies [7]. However, key generation among multiple wireless devices to ensure secure group communication remains a challenge [7].

Also, wireless networks are vulnerable to identity-based attacks, including spoofing and Sybil attacks. Faria and Cheriton (2006) pointed out that media access control (MAC) address spoofing is a problem in wireless local area networks (WLANs) [15]. Malicious users can easily launch by providing a fake MAC address or Internet Protocol (IP) address. Identity-based attacks allows for many other forms of attacks on the wireless networks [16]. Combined with other kinds of attacks, they can degrade the network performance greatly. Although the identity of the user can be verified through cryptographic authentication, cryptographic authentication is not always possible, because the cryptographic authentication requires key management and additional infrastructure overhead [17]. For example, existing 802.11 security techniques only provide authentication for data frames, but not for control frames [18]. Traditional cryptography-based authentication may not be efficient or suitable for some scenarios. It is ill suited for a less equipped distributed

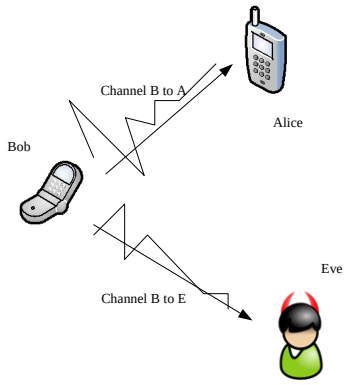

Figure 1.2: Key generation

wireless networks due to high complexity and computational requirements [18]. Moreover, traditional cryptographic methods are subject to node compromise. Physical layer authentication can help combat identity-based attacks. The main idea is to consider the nodes channel, device, and signal characteristics as their signatures. The primary users can be identified at the signal level without relying on higher layer cryptographic means.

Beside these three aspects, some code approaches, such as error correction coding, spread spectrum coding, are also considered as the physical layer security [1]. Also, research on side channel can be considered as physical layer security [19].

In this paper, our focus is physical layer authentication, which exploits the uniqueness of wireless channels to provide the entity authentication and data origin authentication. The main purpose of authentication is to verify a claim of identity. It can effectively combat the identity-based attacks. Compared with the traditional authentication, physical layer authentication is more efficient in terms of authentication speed and computation overhead to authenticate each message, which is more suitable for the scenarios where the devices have limited communication and computational capacities, such as smart grid [20] [21] and body area networks (BANs) [22].

In the literature, physical layer authentication techniques can be classified into three main categories: software-based [23], hardware-based, embedding waveforms [24], and channel-based [25][26][27] [28]. Software-based authentication are essentially based on the unique characteristics and style of the software programs or protocols running on the devices. For example, IEEE 802.11 standards have very large and complex specifications, they are usually implemented in slightly different ways by different device manufactures and driver developers. These variations in implementations can be exploited as a signature

to identify different wireless devices. For example, the probe requests sent by wireless nodes vary between manufactures. Frame sequence numbers can also be used to detect present of multiple 802.11 devices using the same MAC address. The traffic pattern (such as packet sizes and destination address) of the wireless users have be exploited to identify different users [23]. Hardware-based authentication is to exploit the characteristics of hardware, such as clock skew [29], physical unclonable function [30], and radiometric uniqueness to provide the authentication [29]. Embedded authentication is to embed the tag waveform into the transmitted physical layer waveform, which can help to reduce the bandwidth overload [24].

Different from them, channel-based authentication utilizes channel characteristics to authenticate the sender, which has more advantages, as shown in Table 1.1. Channels are estimated on most of existing systems, only the middleware should be changed slightly to implement channel-based authentication. Channel-based authentication is of low complexity, which means energy-effectiveness. Compared with hardware and software-based methods, channels are hard to mimic, which guarantees the security. Compared with embedded waveform, no extra transmit power is needed in channel-based authentication. Due to those advantages, there is a flurry of research activities in this area [9] [25] [26] [27] [28] [31]. In [9], the authors propose an authentication scheme based on Neyman-Pearson hypothesis test (NPHT). In [27], two signal-carrier binary hypothesis test approaches using time-domain training-based channel impulse responses are proposed. Also, in [27], it is assumed that the underlying complex channel is a stationary, zero-mean, Gaussian random process. In [28], two schemes are proposed by using refined multiple tone signature (RMTS) and complex temporal signature (CTS), respectively. RMTS in fact is the correlation coefficient between the frequency-domain channel vectors, while CTS takes the phase shift into the consideration. Based on the difference among noise-mitigated channel impulse responses, a new test statistic is introduced for authentication in fading environments [31]. In most of existing work, how to set the threshold is not discussed. Only ROC curves are given by varying the threshold.

1.3 Contributions

In this thesis, an overview about physical layer security, especially various kinds of non-cryptographic authentication and identification methods using physical layer properties or information in wireless networks is provided firstly. Their advantages and disadvantages are discussed. After that, we study channel-based authentication to enhance security. The authentication problem is formulated as a sequence of hypothesis test problems, where the

Table 1.1: Comparison of physical layer authentication

Main categories	Advantages	Disadvantages
Hardware-based	1, Hard to spoof the signature by using off-the-shelf devices; 2, No additional hardware is necessary; 3, Hard to mimic.	1, Vulnerable to impersonation and replay attacks, if the attacker is more powerful [18]; 2, If an attack is detected, it is hard to change the signature.
Channel-based	1, Hard to mimic; 2, Easy to implement; 3, If attack detected, easy to tune the signature; 4, No additional bandwidth; 5, No additional transmit power.	Not suitable for highly dynamic environments.
Embedded waveform	No additional bandwidth.	High energy consumption.

receiver can identify the sender when receiving frames. To better distinguish the legitimate user and the malicious one, several features including cyclic features of channels are introduced and leveraged. The idea is to compute the cyclic-feature vectors from estimated channel samples and utilize this feature to identify and authenticate different users. Based on these features, support vector machines are researched to provide some classifiers to do authentication using the training data. Simulation demonstrates that when training data is enough, low degree polynomial support vector machines can provide satisfactory performance. The time complexity and space complexity during training and testing stages are analyzed. After that, Fisher discriminant analysis is used to give these features different weights based on some training data to construct a linear classifier. To select the threshold in the authentication scheme, an adaptive threshold scheme is proposed. Based on the data from practical measurements, simulation results demonstrate that the proposed scheme performs better in terms of detection probability and false alarm probability compared with existing channel-based authentication schemes and support vector machine based schemes. Lastly, a combining scheme is presented when several legitimate users exist and one legitimate user can help the other legitimate user to do the authentication. Performance and complexity are analyzed, too.

In sum, our work have the following contributions compared with existing work.

- The sum of false alarm probability and misdetection probability can be reduced to less than 1.8% by our proposed schemes. Furthermore, this sum can be reduced to less than 0.8% by our proposed combining scheme when two-user cooperation is available. When the false alarm probability is 8%, the detection probability can reach 99.97% even without two-user cooperation.

- We address the channel-based authentication problem from the aspect of machine learning. We do not make any assumptions that the statistics characteristics of the channels are known, which are common in existing channel-based schemes. For example, it is assumed that the underlying complex channel is a stationary, zero-mean, Gaussian random process in [27]. Instead, we try to provide better performance based on some training data, which can give our schemes more capacity to be used in various kinds of environments. The kind of generality is gained at the cost of the training stage.

- We address the problem of setting the threshold in our thesis, which makes our proposed schemes more piratical.

1.4 Outline of the Thesis

The organization of the remainder of the thesis is as follows. In Chapter 1, the background about wireless security and literature review in physical layer security are presented. In Chapter 2, the system model is presented and the problem is formulated as a sequence of hypothesis tests. The measurements of indoor channels are investigated and several features used in the following chapters are introduced. Some metrics to compare different schemes are also explained. In Chapter 3, the linear support vector machine (SVM) and several nonlinear SVMs are used to get classifiers to solve the hypothesis test problems. Simulation demonstrates degree-3 polynomial SVM has good performance in terms of misdetection probability and false alarm probability. The time complexity and space complexity are analyzed. In Chapter 4, a linear combination of the channel features is used as the test statistic, which is compared with a threshold to perform authentication. Linear fisher discriminant analysis is used to compute the weights based on some training data. An adaptive threshold scheme is also proposed to adjust the threshold. Using the real channel data measured in a regular office from Utah University, simulation is performed, which demonstrates that the proposed scheme performs better in terms of the sum of misdetection probability and false alarm probability, and the receiver operating characteristic curves, compared with existing channel-based authentication schemes. The misdetection probability and false alarm probability can be reduced greatly with aid of another legitimate user. Lastly, a combining scheme is proposed to combine the data from different legitimate users, which is proven to have better performance than without combining scheme. Again, the time complexity, space complexity, and communication overhead are analyzed. Conclusion and future work is given in the last chapter.

Chapter 2

System Model and Problem Formulation

In this chapter, the system model is introduced and the channel-based authentication is formulated as a series of hypothesis test problems. In order to solve these hypothesis test problems, we need to measure the "similarity" of two channel vectors. Some features to measure the "similarity" are introduced. Lastly, some metrics are also presented to evaluate the performance of our proposed schemes and the existing schemes. These features and metrics will be used in the following chapters.

2.1 System Model

The system model, as shown in Figure 2.1, consists of three different parties: Alice, Bob, and Eve. Alice and Bob are both legitimate users, while Eve is a malicious user. Alice is transmitting frames to Bob. However, Eve tries to masquerade Alice and send frames to Bob. Bob has to authenticate the coming frames. That is, Bob has to determine whether the incoming frames are from the legitimate user, i.e., Alice, according to the channels estimated.

Channels are widely sampled and estimated in deployed wireless communication networks to demodulate the received signal. We name the vector of samples of time-domain channel impulse response (amplitude delay profile) sampled from one frame a channel vector. We assume that Bob first measures and stores the channel vector between him and Alice as the shared secret at the beginning of the authentication process . Let $\vec{h}_A$ denote

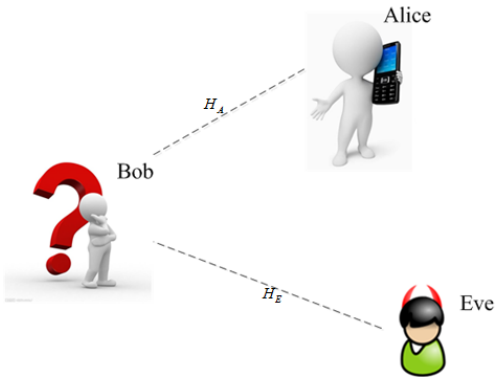

Figure 2.1: System model

the stored channel vector, which can be represented as follows:

$$\vec{h}_A = [h_{A,0}, h_{A,1}, h_{A,2}, ..., h_{A,M-1}] \tag{2.1}$$

where the subscript A denotes "the channel is from Alice", and M denotes the number of samples sampled uniformly per frame.

When one frame is received, Bob estimates the channel $\vec{h}_t(k)$ from this frame as follows:

$$\vec{h}_t(k) = [h_{t,0}(k), h_{t,1}(k), h_{t,2}(k), ..., h_{t,M-1}(k)] \tag{2.2}$$

where the subscript t denotes "transmitter to be identified and authenticated" and k denotes the frame index $k = 1, 2, \cdots$. Bob needs to determine whether the first frame is from Alice or not according to $\vec{h}_t(1)$ and $\vec{h}_A$.

If Bob determines that the first frame is from Alice, Bob should keep authenticating the second frame using the channel vector estimated in the first frame and the channel vector estimated in the second frame. That is, Bob determines whether the second frame is from Alice or not according to $\vec{h}_t(1)$ and $\vec{h}_t(2)$. Otherwise, an alarm should be given out. Also, he needs to determine whether the third frame is from Alice according to $\vec{h}_t(2)$ and $\vec{h}_t(3)$.

We have the following assumptions, which are common in the literature of physical layer authenticaion[9][25]. These assumptions are supported by the well-known Jakes uniform scattering model [32], which states that the signal rapidly de-correlates over a distance of roughly half a wavelength, and that spatial separation of one to two wavelengths is sufficient for assuming independent fading paths[25].

- Eve cannot measure or create or precisely model the channels between Alice and Bob, which means Eve cannot mimic the legitimate channels to masquerade Alice. This assumption always holds when Bob and Eve are located in spatially separated positions, especially when they are a wavelength away form the Eve in a richly scattered multi-path environment (typical of indoor wireless environments)[25].

- Alice's channel is different from Eve's channel. Take the indoor WiFi scenario for example. If Alice is more than 12.5cm away from Eve, which is larger than the WiFi wavelength, the channel from Alice to Bob and the channel from Eve to Bob are totally different.

- Channels for two successive frames from Alice are nearly the same. This assumption always holds, especially when there are no dynamic movements of Bob and Alice. For example, in the indoor WiFi scenario, where the human speed is about 1m/s, and the physical layer frame duration is about 3.5 ms. During this duration, the movement of Alice is about 3.5mm, which is less than the WiFi wavelength. Therefore, the channel between Alice and Bob almost does not change [9] [27]. Since the factual frame duration in our system model is larger than 3.5 ms, this assumption holds as well.

Based on these assumptions above, the main objective of this thesis is to distinguish the channels of Alice-Bob $\vec{h}_A$ and the channels of Eve-Bob $\vec{h}_E$.

2.2 Problem Formulation

Generally speaking, the channel-based authentication can be considered as a classification problem. All the incoming channel vectors can be classified into two categories. The first category is those from Alice. The other category is those not from Alice. When a new frame comes, we need to determine whether the new channel vector $\vec{h}_t$ and $\vec{h}_A$ belong to the same category (the same transmitter) in Figure 2.1.

We use hypothesis tests to formulate the problem as in [26]. We formulate the channel-based authentication problem as a sequence of hypothesis test problems as in [26]. The first hypothesis test problem is given by

$$\begin{aligned} \mathcal{H}_0 &: \vec{h}_t(1) \text{ is from Alice} \\ \mathcal{H}_1 &: \vec{h}_t(1) \text{ is not from Alice} \end{aligned} \tag{2.3}$$

It can be also rewritten as:

$$\begin{aligned} \mathcal{H}_0 &: \vec{h}_t(1) = \vec{h}_A \\ \mathcal{H}_1 &: \vec{h}_t(1) \neq \vec{h}_A \end{aligned} \tag{2.4}$$

The null hypothesis, $\mathcal{H}_0$, is that $\vec{h}_t(1)$ is from Alice, other than Eve. The alternative hypothesis, $\mathcal{H}_1$, is that $\vec{h}_t(1)$ is from Eve, other than Alice. To determine whether $\vec{h}_t(1)$ is same as $\vec{h}_A$ means to differentiate $\mathcal{H}_0$ and $\mathcal{H}_1$. We use the hypothesis testing to determine which hypothesis to accept. That is, we need to construct some test statistics, responding accept region, and rejection region.

Suppose that during the first hypothesis test, $\mathcal{H}_0$ is accepted. Bob continues authenticating the subsequent frames. Otherwise, an alarm will be given. The second hypothesis test is given by

$$\begin{aligned} \mathcal{H}_0 &: \vec{h}_t(2) = \vec{h}_A(1) \\ \mathcal{H}_1 &: \vec{h}_t(2) \neq \vec{h}_A(1) \end{aligned} \tag{2.5}$$

In summary, the channel-based authentication can be formulated into a general hypothesis problem: given a channel vector between legitimate users $\vec{h}_A(i-1)$, whether is the successive channel vector $\vec{h}_t(i)$ from Alice or not? It can be represented by the following hypothesis test problem:

$$\begin{aligned} \mathcal{H}_0 &: \vec{h}_t = \vec{h}_A \\ \mathcal{H}_1 &: \vec{h}_t \neq \vec{h}_A \end{aligned} \tag{2.6}$$

2.3 Selection of Features

Consider the scenario in Figure 2.2. When Alice moves from position P1 to position P2, the estimated channel vector $\vec{h}_A(1)$ is very similar to $\vec{h}_A(2)$. But $\vec{h}_E(1)$ is very different from $\vec{h}_A(1)$. In order to detect the "would-be" intruder, we need some features to measure this kind of "similarity" between any two channel vectors in Figure 2.3. These features, which reflects the "similarity" between any two channel vectors, can be used to construct the test statistic to solve (2.6). In (2.6), if the "similarity" is high, we accept $\mathcal{H}_0$. Otherwise, we reject $\mathcal{H}_0$.

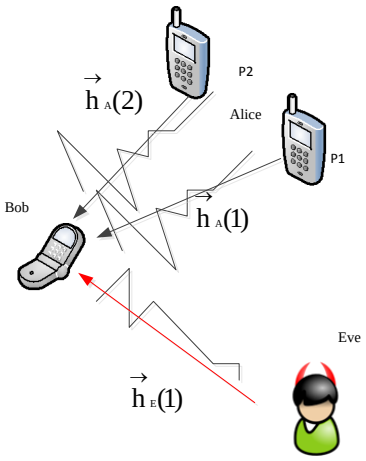

Figure 2.2: System model 2

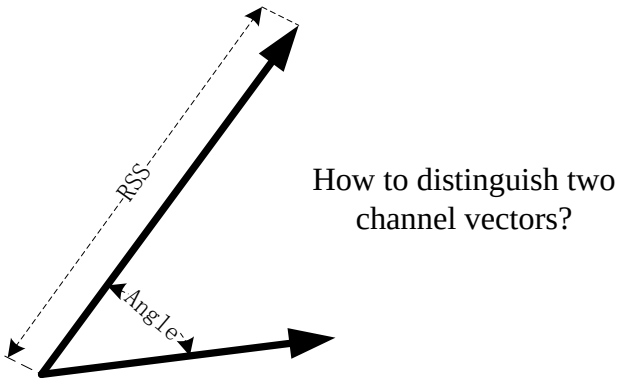

Figure 2.3: How to classify two channel vectors

2.3.1 RSS

Received signal strength (RSS) is a measurement of the power received at the receiver. It reflects the length of the channel vector in Figure 2.3. In the past, RSS is usually invisible to the users of the device, but is known to users of IEEE 802.11 protocol family now. You can get the RSS of your Android phones using the application program interfaces (APIs) provided by Google easily. RSS can be used in localization [33] and physical-layer authentication [16], due to the existing relationship between RSS and the distance. RSS is defined as ten times of the logarithm of the power of the received signal and a reference power [33]. The power dissipates from a point source as it moves further out and the relationship between power and distance is that power is inversely proportional to the square of the distance traveled [33], which is shown in (2.8) [33]. RSS can be considered as the square of the length of channel vectors.

$$\begin{aligned} RSS(\vec{H}_A) &= log(\vec{H}_A \vec{H}_A^{\mathrm{H}}) \\ &= log(\vec{h}_A \vec{h}_A^{\mathrm{H}}) \end{aligned} \tag{2.7}$$

where the superscript H denotes Hermitian Transpose, H denotes the frequency version of channel vector, and $log()$ denotes $log_{10}()$.

$$RSS = -KlogDis + constant \tag{2.8}$$

where K is the slope of the standard plot, Dis is the distance between receiver and transmitter, and $constant$ is a constant parameter.

We use the measured channel impulse response data set from Utah University [28] to analyze the relationship between distance and RSS. The measurement was done in a 14 by 13 meters regular office with many small cubes in it. The result is shown in Figure 2.4, which demonstrates that $log(RSS)$ is inversely proportional to $log(Dis)$. When Alice and Eve are in different positions, we can use RSS as a feature to reflect the "similarity".

2.3.2 TOA

Time of arrival (TOA) is the travel time of a radio signal from the transmitter to the remote receiver. TOA can be used in localization [34] and physical-layer authentication [35], due to the existing relationship between TOA and the distance. We can use TOA as a feature to do authentication, which is shown in (2.9).

$$TOA = \frac{Dis}{speed} \tag{2.9}$$

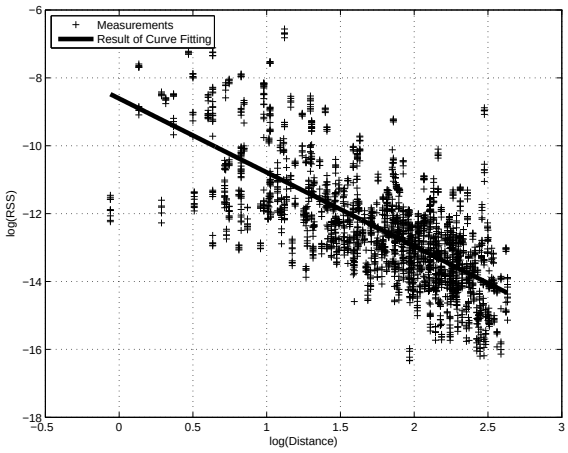

Figure 2.4: Plot of RSS versus Dis

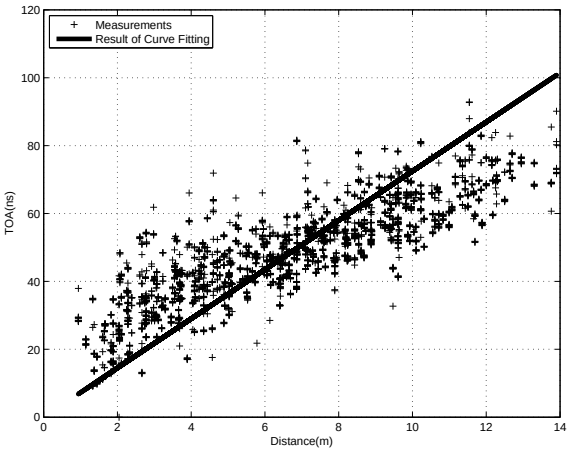

Figure 2.5: Plot of TOA versus Dis

where *speed* is the speed of signal.

We also use data set from Utah University [28] to analyze the relationship between distance and TOA. The result is shown in Figure 2.5, which demonstrates that we can use TOA as a feature to do reflect the "similarity" when Alice and Eve are in different positions.

2.3.3 Correlation of Channel Vectors

λ' is defined as the absolute value of the correlation of $\vec{H}_t$ and $\vec{H}_A$, denoted as $Corr(\vec{H}_t, \vec{H}_A)$, which is shown in (2.10). Correlation can be written as a normalized inner product, which measures how well the two vectors align linearly. Correlation can reflect the angle between $\vec{H}_t$ and $\vec{H}_A$ in Figure 2.3. The probability distribution function (PDF) of λ' given that $\vec{H}_t$ and $\vec{H}_A$ are both from Alice, as well as the PDF of λ' given that $\vec{H}_t$ is not from Alice is shown in Figure 2.6. Let $f()$ denote PDF. In Figure 2.6, it is shown that the correlation coefficients of channel vectors from the same user are close to one. The correlation coefficients of channel vectors from different users are not close to one, when they are a little far away. λ' can be used to reflect the "similarity".

$$\begin{aligned} \lambda' &= abs(Corr(\vec{H}_t, \vec{H}_A)) \\ &= abs(Corr(\vec{H}_A, \vec{H}_t)) \\ &= abs(\frac{\|\vec{H}_t \vec{H}_A^{\mathrm{H}}\|}{\|\vec{H}_t\| \|\vec{H}_A\|}) \end{aligned} \tag{2.10}$$

where $abs()$ denotes absolute value, and $\|\|$ denotes the vector norm.

2.3.4 Correlation of Cyclic Feature Vectors

λ is in fact a highly dimensional version of λ'. Let $H_A(z)$ and $H_t(z)$ denote their z-transforms of $\vec{h}_A$ and $\vec{h}_t$, respectively, which can be given as follows:

$$H_A(z) = \sum_{i=0}^{M-1} h_{A,i} z^{-i} \tag{2.11}$$

$$H_t(z) = \sum_{i=0}^{M-1} h_{t,i} z^{-i} \tag{2.12}$$

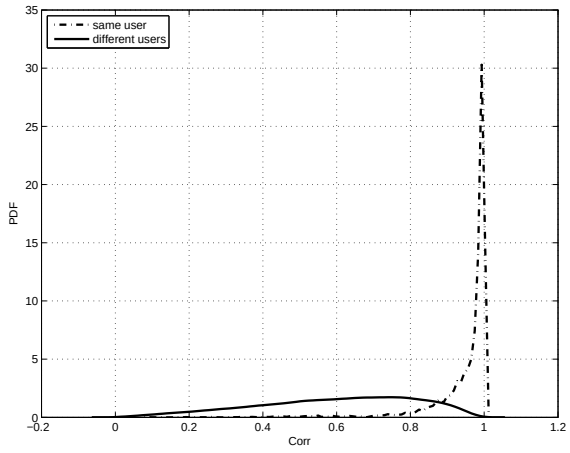

Figure 2.6: $f(\lambda'|\vec{H}_t \neq \vec{H}_A)$ and $f(\lambda'|\vec{H}_t = \vec{H}_A)$

Define the $\gamma - z$ transforms $S_A(\gamma, z)$ and $S_t(\gamma, z)$ of $\vec{h}_A$ and $\vec{h}_t$, respectively, which are similar to the cyclic spectrum of cyclostationary signals[36]. $S_A(\gamma, z)$ and $S_t(\gamma, z)$ can be given as follows:

$$S_A(\gamma, z) = H_A^{\star}(z^{\star})H_A(z^{-1}e^{j\gamma}) \tag{2.13}$$

$$S_t(\gamma, z) = H_t^{\star}(z^{\star})H_t(z^{-1}e^{j\gamma}) \tag{2.14}$$

where $\star$ denotes conjugate operation, and j denotes unit imaginary number.

Sample the $\gamma - z$ plane and gain the feature matrix of channels. For example, when $z = e^{j\frac{2\pi}{N}n}$ $(n = 0, 1, \cdots, N-1)$ and $\gamma = \frac{2\pi}{Q}q$ $(q = 0, 1, \cdots, N-1)$, we can get an $N \times Q$ matrix. To reduce the complexity, we can just use the information at several cyclic-frequencies. For example, pick up P cyclic frequencies from the set $\{\gamma = \frac{2\pi}{Q}q : q = 0, 1, \cdots, Q-1\}$ and compute an $N \times P$ matrix. We name this kind of matrixes as the cyclic feature matrix of channels and reshape them into a cyclic feature vector. Let $\vec{S}_A(q, n)$ and $\vec{S}_t(q, n)$ denote the cyclic feature vectors of $\vec{h}_A$ and $\vec{h}_t$. λ is defined as the absolute value of the correlation of the cyclic feature vectors, which is shown in Figure 2.7.

In Figure 2.8, it is shown that λ has more power than λ' to do authentication. The P in case 1 doubles the P in case 3. The P in case 3 doubles the P in case 5. When P is increased, the peak of the PDF of λ moves to the left, which makes authentication easier.

$$\begin{aligned} \lambda &= abs(CycCorr(\vec{H}_t, \vec{H}_A)) \\ &= abs(\frac{\|\vec{S}_t(q, n)\vec{S}_A(q, n)^{\mathrm{H}}\|}{\|\vec{S}_t(q, n)\|\|\vec{S}_A(q, n)\|}) \end{aligned} \tag{2.15}$$

The computation of λ and λ' involve the fast Fourier transform (FFT) of channel vectors. The N-point FFT computation involves $\frac{N}{2}log_2N$ complex multiplications and $Nlog_2N$ complex additions [37]. Let O denote O-notation. The complexity comparison of computing these features is shown in Table 2.1.

Table 2.1: Complexity comparison of computing λ and λ'

Features	Number of Complex Multiplications	Number of Complex Additions
λ	$\frac{N}{2}log_2N + NP$	$Nlog_2N + NP$
λ'	$\frac{N}{2}log_2N + N$	$Nlog_2N + N$

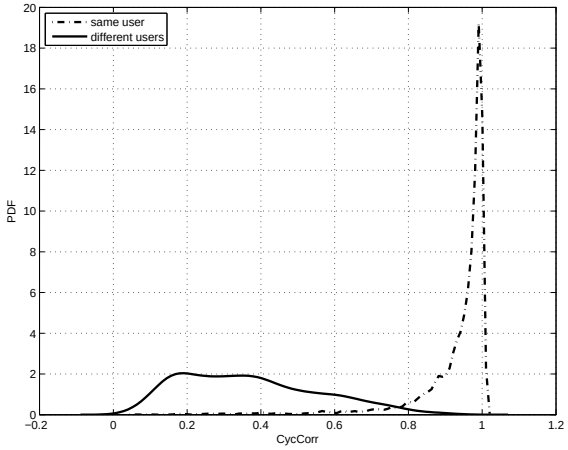

Figure 2.7: $f(\lambda|\vec{H}_t \neq \vec{H}_A)$ and $f(\lambda|\vec{H}_t = \vec{H}_A)$

2.3.5 Another Two Features (β and θ)

In order to make the setting of thresholds easier, we do some transformation on TOA and RSS. β and θ are defined as following:

$$\beta = \frac{min(RSS(\vec{h}_A), RSS(\vec{h}_t))}{max(RSS(\vec{h}_t), RSS(\vec{h}_A))} \tag{2.16}$$

$$\theta = \frac{min(TOA(\vec{h}_A), TOA(\vec{h}_t))}{max(TOA(\vec{h}_t), TOA(\vec{h}_A))} \tag{2.17}$$

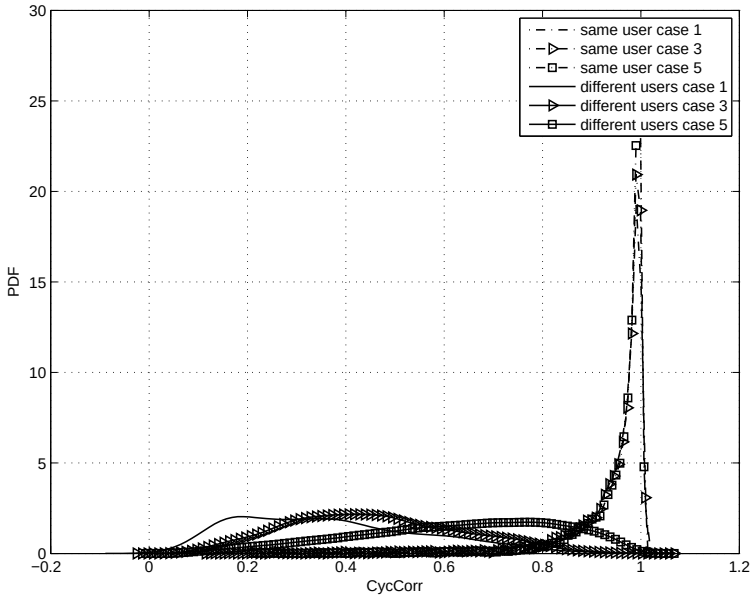

Figure 2.8: $f(\lambda|\vec{H}_t \neq \vec{H}_A)$ and $f(\lambda|\vec{H}_t = \vec{H}_A)$ under different Ps

It is shown that $\beta \in [0,1]$ and $\theta \in [0,1]$. So, the threshold should be in $[0,1]$. In the following chapters, it is shown that this kind of normalization makes the setting of threshold convenient.

In the following chapters, we use λ, β and θ to construct one vector, which is called feature vector. We use the feature vector in (2.18) to measure the "similarity". All the schemes proposed in this thesis are based on this kind of feature vectors.

$$\vec{x} = [\theta, \lambda, \beta]^{\mathrm{T}} \tag{2.18}$$

2.4 Metrics

In this section, some metrics for evaluating the performance of our proposed schemes are presented. We are interested in the statistical characterization of the attack-detection over all the possible attacks. To evaluate the performance, P_D is defined as the detection probability, which is the percentage of attack attempts that are are determined to be under attack. An successful identity-based attack will cause the hypothesis test to reject $\mathcal{H}_0$. P_F

is defined as the false alarm probability, which is the percentage of legitimate attempts that are are determined to be under attack. P_D and P_F can be given as follows:

$$P_D = Prob(\mathcal{H}_0 \text{ is rejected}|\vec{h}_t \text{ is from Eve}) \tag{2.19}$$

where $Prob()$ denotes probability.

$$P_F = Prob(\mathcal{H}_0 \text{ is rejected}|\vec{h}_t \text{ is from Alice}) \tag{2.20}$$

Let P_M denote the misdetection probability, which is defined by

$$P_M = Prob(\mathcal{H}_0 \text{ is accepted}|\vec{h}_t \text{ is from Eve}) \tag{2.21}$$

The relationship between P_D and P_M is as following:

$$P_D = 1 - P_M \tag{2.22}$$

A good scheme should have low P_F and high P_D, which means that it has a low false alarm probability and a high detection probability. Always, detection probability and false alarm probability vary under different thresholds. Receiver operating characteristic (ROC) curve can be used to study both the detection probability P_D and false alarm probability P_F. The ROC curve is a plot of detection probability compared to the false alarm probability. The ROC curve is a direct means of measuring the tradeoff between detection probability and false alarm probability. It can be gained by varying thresholds.

Also, we are interested in the sum of P_M and P_F, which reflects the accuracy of our proposed schemes. In most cases, we need to minimize this sum. So, $P_M + P_F$ is also a metric to evaluate the performance of the schemes.

The measurements and estimates of channel vectors can be considered free of measurement noise in simulation. But in reality, measurements always involves noise. As a result, we need to consider the sensitiveness of our proposed schemes and existing schemes to this measurement noise. Considering the measurement noise, we define the signal-to-noise ratio (SNR) as the channel power to the measurement noise power. ROC curves and $P_M + P_F$ should be researched under different SNRs.

In most cases, our proposed schemes need some training samples in training phase to adjust the parameters. We need to study the performance of our proposed schemes with varied sample size. In this case, we can plot the sum of the misdetection probability P_M and the false alarm probability P_F compared with sample size.

Chapter 3

Support Vector Machine Based Authentication Schemes

As explained in the last chapter, if the feature vector of two channel vectors reflects high "similarity", $\mathcal{H}_0$ is accepted. Otherwise, $\mathcal{H}_1$ is accepted. In other words, the hypothesis testing problem in (2.6) is a classification problem, too. There are two classes. If the two channel vectors are from Alice, their feature vector belongs to class -1. If the two channel vectors are from Alice and Eve, their feature vector belongs to class 1. In this thesis, we fucus on the two-class Support Vector Machines (SVMs) to solve the hypothesis testing problem in (2.6). SVMs are supervised learning models with associated learning algorithms and are used for classification and regression analysis [38]. An SVM model is a representation of the feature vectors as points in space, mapped so that the feature vectors of the two categories are divided by a clear gap that is as wide as possible. New feature vectors are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. A good separation is achieved by the hyperplane that has the largest distance to the nearest training data points of the the two classes. That is, each feature vector is a point in the space, SVM can be used to find the optimal hyper-plane to classify two classes of points. The distance is called functional margin. Always, a set of training feature vectors is necessary for an SVM training algorithm to build a model that assigns new feature vectors ($\vec{x}_i$) into one category or the other.

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

Our proposed SVM based authentication scheme: Training sets are used to train SVM

machines, which outputs the maximum-margin hyper-plane which can distinguish the feature vectors of two classes. These hyper-planes are linear or nonlinear classifiers, which can be used to solve hypothesis test problems in (2.6).

3.1 SVMs

The labeled training set D_{tr} with I_{tr} feature vectors is given by (3.1). y_i is the label of $\vec{x}_i$.

$$D_{tr} = \{\vec{x}_i | \vec{x}_i \in R^3, y_i \in \{-1, 1\}\}_{i=1}^{I_{tr}} \tag{3.1}$$

Each feature vector $\vec{x}_i$ is 3-dimensional real vector of the form (3.2).

$$\vec{x}_i = [\beta_i, \lambda_i, \theta_i]^{\mathrm{T}} \tag{3.2}$$

Our goal is to find the maximum margin hyperplane that divides these feature vectors with $y_i = +1$ from those with $y_i = -1$. Any hyperplane can be written as the set of feature vectors $\vec{x}$ satisfying

$$\vec{w}^{\mathrm{T}} \cdot \vec{x} - b = 0 \tag{3.3}$$

where $\cdot$ denotes the dot product, and $\vec{w}$ is the normal vector to the hyperplane. $b/\|\vec{w}\|$ is the offset of the hyperplane from the origin along $\vec{w}$. If $\vec{w}^{\mathrm{T}}\vec{x}_i + b > 0$, y_i is labeled as 1, otherwise y_i is labeled as -1.

$$\vec{w}^{\mathrm{T}}\vec{x}_i + b \begin{cases} < 0 & for \quad y_i = -1 \\ > 0 & for \quad y_i = +1 \end{cases} \tag{3.4}$$

Assume that the training points are linearly separable, no training points are on the hyperplane $\vec{w}^{\mathrm{T}}\vec{x} + b = 0$. Thus, in order to control separability, the following inequalities is considered.

$$\vec{w}^{\mathrm{T}}\vec{x}_i + b \begin{cases} < -1 & for \quad y_i = -1 \\ > +1 & for \quad y_i = +1 \end{cases} \tag{3.5}$$

(3.5) is equivalent to

$$y_i(\vec{w}^{\mathrm{T}}\vec{x}_i + b) \geq 1, \forall \vec{x}_i \in D_{tr} \tag{3.6}$$

The hyperplane $\vec{w}^{\mathrm{T}}\vec{x} + b = c(-1 < c < 1)$ form separating hyperplanes. If the training data D_{tr} are linearly separable, two hyper-planes can be selected so that all the feature vectors are separated by the two hyper-planes and no feature vectors are between them,

and then try to maximize their distance. The region bounded by them is called "margin". The hyperplanes with the maximum margin are so-called optimal separating hyperplanes. Our goal is to determine the optimal separating hyperplanes. The Euclidean distance from a training datum $\vec{x}$ to the separating hyperplane is $|\vec{w}^{\mathrm{T}}\vec{x}+b|/\|\vec{w}\|$ [39]. The line going through $\vec{x}$ and orthogonal to the separating hyperplane is $a\vec{w}/\|\vec{w}\|+\vec{x}$, where $|a|$ is the Euclidean distance from $\vec{x}$ to the hyperplane [39].

Then all the training data must satisfy

$$y_i(\vec{w}^{\mathrm{T}}\vec{x}_i+b)/\|\vec{w}\| \geq \delta, \forall \vec{x}_i \in D_{tr} \tag{3.7}$$

where δ is the margin. We impose the constraint in (3.8).

$$\delta\|\vec{w}\| = 1 \tag{3.8}$$

From (3.7) and (3.8), we can get the optimal separating hyperplane. The problem of finding the optimal separating hyperplane becomes (3.9).

$$\begin{aligned} &(\vec{w}, b) = arg \ \ min_{\vec{w},b} \ \ 1/2\|\vec{w}\|^2 \\ &\text{s.t.} \ \ y_i(\vec{w}^{\mathrm{T}}\vec{x}_i+b) \geq 1, \forall \vec{x}_i \in D_{tr} \end{aligned} \tag{3.9}$$

The dual problem is therefore:

$$\begin{aligned} &max \ \ \sum_{i=1}\alpha_i - 1/2 \sum_{i=1,j=1} \alpha_i\alpha_j y_i y_j \vec{x}_i^{\mathrm{T}}\vec{x}_j \\ &\text{s.t.} \ \ \alpha_i \geq 0, \sum_{i=1}\alpha_i y_i = 0 \end{aligned} \tag{3.10}$$

This is a quadratic programming (QP) problem. Because it is assumed that the training data are separable, there must be feasible solutions for $y_i(\vec{w}^{\mathrm{T}}\vec{x}_i+b) \geq 1(\forall \vec{x}_i \in D_{tr})$. The support vectors are those points satisfying $y_i(\vec{w}^{\mathrm{T}}\vec{x}_i+b) = 1(\forall \vec{x}_i \in D_{tr})$. S is called the set of the support vectors [39].

$$\vec{w} = \sum_{\vec{x}_i \in S} \alpha_i y_i \vec{x}_i \tag{3.11}$$

$$\sum_{i=1}\alpha_i y_i = 0 \tag{3.12}$$

where $\alpha_i > 0$ if $\vec{x}_i$ is a support vector, otherwise $\alpha_i = 0$.

$$b = 1/|S| \sum_{\vec{x}_i \in S}(y_i - \vec{w}^{\mathrm{T}}\vec{x}_i) \tag{3.13}$$

where || here denotes the number of elements in the set.

Always, many of α_i are zero. $\vec{w}$ is a linear combination of a small number of data points $\vec{x}_i$. For testing with new data $\vec{z}$, if $\vec{w}^{\mathrm{T}}\vec{z}+b = \sum_{\vec{x}_i \in S} \alpha_i y_i (\vec{x}_i^{\mathrm{T}}\vec{z}) + b > 0$, $\vec{z}$ belongs to class 1, otherwise, it belongs to class -1, which is given in (3.14).

$$\begin{aligned} & \vec{w}^{\mathrm{T}}\vec{z}+b \\ & = \sum_{\vec{x}_i \in S} \alpha_i y_i (\vec{x}_i^{\mathrm{T}}\vec{z}) + b \\ & \qquad\qquad\qquad\qquad \mathcal{H}_1 \\ & = (\sum_{\vec{x}_i \in S} \alpha_i y_i \vec{x}_i^{\mathrm{T}})\vec{z} + b \lessgtr 0 \\ & \qquad\qquad\qquad\qquad \mathcal{H}_0 \end{aligned} \tag{3.14}$$

In fact, the training data are always not separable. If "error" ξ_i is allowed, we need to solve the following problem:

$$\begin{aligned} & (\vec{w}, b) = arg \ \ min_{\vec{w},b} \ \ 1/2\|\vec{w}\|^2 + \mathcal{C}\sum_{i=1} \xi_i \\ & \text{s.t.} \ \ y_i(\vec{w}^{\mathrm{T}}\vec{x}_i + b) \geq 1 - \xi_i \ \ \forall \vec{x}_i \in D_{tr}, \ \ \xi_i \geq 0 \end{aligned} \tag{3.15}$$

where ξ_i is the "slack variables" in optimization [40]. $\xi_i = 0$ means that there is no error for x_i. $\mathcal{C}$ is the tradeoff parameters between error and margin.

Its dual program is

$$\begin{aligned} & max \ \ \sum_{i=1} \alpha_i - 1/2 \sum_{i=1,j=1} \alpha_i \alpha_j y_i y_j \vec{x}_i^{\mathrm{T}} \vec{x}_j \\ & \text{s.t.} \ \ \mathcal{C} \geq \alpha_i \geq 0, \sum_{i=1} \alpha_i y_i = 0 \end{aligned} \tag{3.16}$$

The dual optimization problem can be solved using QP solver [41]. $\vec{w}$ is recovered as $\vec{w} = \sum_{x_i \in S} \alpha_i y_i \vec{x}_i$. For testing with new data $\vec{z}$, if $\vec{w}^{\mathrm{T}}\vec{z}+b = \sum_{\vec{x}_i \in S} \alpha_i y_i (\vec{x}_i^{\mathrm{T}}\vec{z}) + b > 0$, $\vec{z}$ belongs to class 1, otherwise, it belongs to class -1.

Nonlinear SVMs can be used by aid of some "kernel tricks" [42]. Assume that the kernel function is defined by $\mathcal{K}(\vec{x}_i, \vec{x}_j) = \phi(\vec{x}_i)^{\mathrm{T}}\phi(\vec{x}_j)$. The problem becomes:

$$\begin{aligned} & max \ \ \sum_{i=1} \alpha_i - 1/2 \sum_{i=1,j=1} \alpha_i \alpha_j y_i y_j \mathcal{K}(\vec{x}_i, \vec{x}_j) \\ & \text{s.t.} \ \ \mathcal{C} \geq \alpha_i \geq 0, \sum_{i=1} \alpha_i y_i = 0 \end{aligned} \tag{3.17}$$

$\vec{w}$ is recovered as $\vec{w} = \sum_{x_i \in S} \alpha_i y_i \phi(\vec{x}_i)$. For testing with new data $\vec{z}$, if $\sum_{\vec{x}_i \in S} \alpha_i y_i \mathcal{K}(\vec{x}_i, \vec{z}) + b > 0$, $\vec{z}$ belongs to class 1, otherwise, it belongs to class -1.

In sum, the idea of SVM based authentication scheme is to gain the classifiers $\sum_{\vec{x}_i \in S} \alpha_i y_i (\vec{x}_i^{\mathrm{T}} \vec{z}) + b$ or $\sum_{\vec{x}_i \in S} \alpha_i y_i \mathcal{K}(\vec{x}_i, \vec{z}) + b$ using D_{tr}.

3.2 Simulation of SVM Based Schemes

In this section, we use the measured channel impulse response data set from Utah University [43] to evaluate the performance of our proposed SVM based authentication schemes. Firstly, we introduce how to use this data set to do the simulation. Then, we present the simulation results.

3.2.1 Experimental Channel Measurements

To evaluate the performance of our proposed schemes we use the measured channel impulse response data set from Utah University [43]. The measurement was done in a 14 by 13 meters regular office with many small cubes in it. There are 44 locations in this office, as shown in Figure 3.1. The small circles in Figure 3.1 are the locations used for measurements.

A 40 MHz chip rate signal generator was placed at one location and a software radio (Sigtek model ST-515) designed to receive the signal was placed at another location [43]. This is a typical channel measurement. The transmitter and receiver were closely synchronized to within 1-2 nanoseconds of each other [43]. The channels between them were estimated for several times using high-bandwidth estimate of the (complex) channel impulse response and then averaged. For example, when the transmitter is at location 1, and the receiver is at location 13, the channel measurement is shown in Figure 3.2. The data obtained from this experiment thus present the actual propagation delay of the radio frequency (RF) signal. 0 nanoseconds, in Figure 3.2, denotes the time when the transmitter starts to transmit the signal. We use the data set to verify the performance of the proposed scheme and the existing schemes.

For any two channel measurements $\vec{h}_i$ and $\vec{h}_{i-1}$ (the subscript i denotes the frame index), we compute its feature vector $\vec{x}_i$ as following:

$$x_{i,0} = \beta_i = \frac{min(RSS(\vec{h}_{i-1}), RSS(\vec{h}_i))}{max(RSS(\vec{h}_i), RSS(\vec{h}_{i-1}))} \tag{3.18}$$

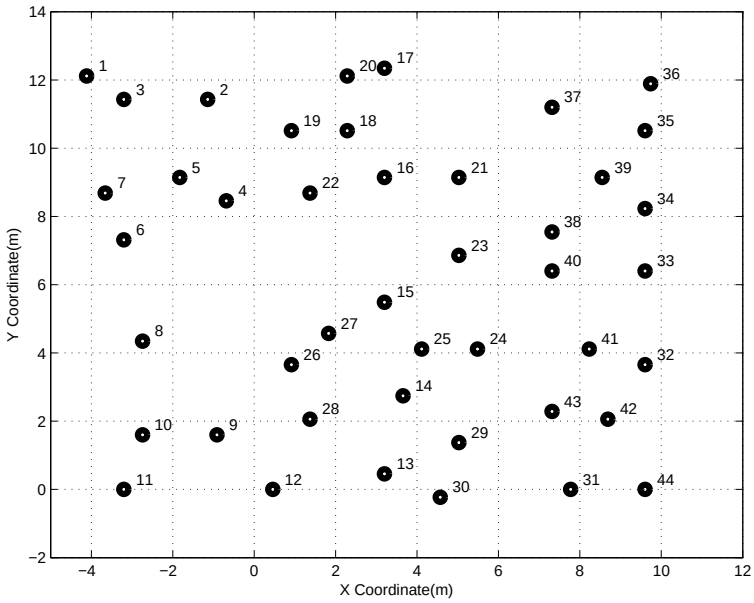

Figure 3.1: Map of measurement locations in Utah University

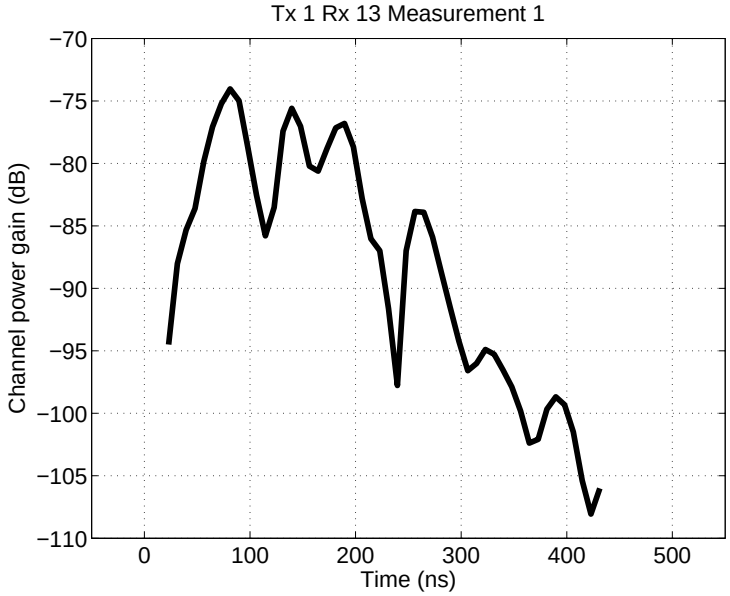

Figure 3.2: Channel power gain in one measurement

$$x_{i,1} = \theta_i = \frac{min(TOA(\vec{h}_{i-1}), TOA(\vec{h}_i))}{max(TOA(\vec{h}_i), TOA(\vec{h}_{i-1}))} \tag{3.19}$$

$$x_{i,2} = \lambda_i = abs(\frac{\|\vec{S}_i \vec{S}_{i-1}^{\mathrm{H}}\|}{\|\vec{S}_t\| \|\vec{S}_{i-1}\|}) \tag{3.20}$$

$$\vec{x}_i = [x_{i,0}, x_{i,1}, x_{i,2}]^{\mathrm{T}} = [\beta_i, \theta_i, \lambda_i]^{\mathrm{T}} \tag{3.21}$$

where $\vec{h}_0 = \vec{h}_A$.

Each channel vector has 50 elements. And we use all the 50 elements for simulation. y_i is the label of $\vec{x}_i$. In other words, y_i denotes whether the channel measurements $\vec{h}_i$ and $\vec{h}_{i-1}$ are from the same user or not. If it is, $y_i = -1$. Otherwise, $y_i = 1$. Let D denote the whole data set.

$$D = \{\vec{x}_i | \vec{x}_i \in R^3, y_i \in \{-1, 1\}\}_{i=1}^{I} \tag{3.22}$$

where I denote the total number of the feature vectors in the set D. $\vec{x}_i$ is a 3-dimensional real feature vector and y_i can be either 1 or -1, indicating which class the data point $\vec{x}_i$ belongs to.

We use the data set to evaluate the performance of our proposed schemes and the existing schemes. We use one subset of D (denoted as D_{tr}) as the training data, and use the complement set of D_{tr} as the test set, denoted as D_{test}. That is, we select one location as the AP location randomly, several other locations for training, and the remaining locations for testing. The channel vectors from the training locations to AP are used to compute D_{tr}. The channel vectors from the testing locations to AP are used to compute D_{test}. In this case, beside y_i of D_{test}, we know all the remaining information. To evaluate the performance of our proposed schemes means to predict y_i of D_{test} based on D_{tr}, y_i of D_{tr}, and D_{test}.

It is obvious that D is the union of D_{test} and D_{tr}, which is shown in (3.23).

$$D = D_{tr} \cup D_{test} \tag{3.23}$$

Let $D_{E,tr}$ denote the set of all the training feature vectors, which are computed from pairs of channel vectors from different users. Let $D_{A,tr}$ denote the set of all the training feature vectors, which are computed from pairs of channel vectors from the same user.

$$D_{E,tr} = \{\vec{x}_i | \vec{x}_i \in D_{tr}, y_i = +1\} \tag{3.24}$$

$$D_{A,tr} = \{\vec{x}_i | \vec{x}_i \in D_{tr}, y_i = -1\} \tag{3.25}$$

D_{tr} is the union of $D_{A,tr}$ and $D_{E,tr}$.

$$D_{tr} = D_{E,tr} \cup D_{A,tr} \tag{3.26}$$

Let $D_{E,test}$ denote the set of all the testing feature vectors, which are computed from pairs of channel vectors from different users. Let $D_{A,test}$ denote the set of all the testing feature vectors, which are computed from pairs of channel vectors from the same user.

$$D_{E,test} = \{\vec{x}_i | \vec{x}_i \in D_{test}, y_i = +1\} \tag{3.27}$$

$$D_{A,test} = \{\vec{x}_i | \vec{x}_i \in D_{test}, y_i = -1\} \tag{3.28}$$

D_{test} is the union of $D_{E,test}$ and $D_{A,test}$.

$$D_{test} = D_{E,test} \cup D_{A,test} \tag{3.29}$$

Let D_E denote the set of all the feature vectors, which are computed from pairs of channel vectors from different users. Let D_A denote the set of all the feature vectors, which are computed from pairs of channel vectors from the same user.

$$D_E = \{\vec{x}_i | \vec{x}_i \in D, y_i = +1\} \tag{3.30}$$

$$D_A = \{\vec{x}_i | \vec{x}_i \in D, y_i = -1\} \tag{3.31}$$

D is the union of D_A and D_E.

$$D = D_E \cup D_A \tag{3.32}$$

3.2.2 Simulation Results of SVM Based Schemes

In this section, we do the simulation according to the process in Figure 3.3. There are 44 points representing 44 measurement locations in Figure 3.4. Each round, AP is picked randomly. After that, we pick several points as the the training locations from the remaining locations in Figure 3.4. The channel vectors from these training locations to the AP are used to train the SVMs. That is, these channel vectors from these training locations to the AP are used to compute the feature vectors $\vec{x}_i$ in D_{tr}. Then, we use D_{tr} to compute the SVMs introduced above. In fact, we use MATLAB library functions to solve the QP problems in (3.10), (3.16), and (3.17) to gain the classifiers $\sum_{\vec{x}_i \in S} \alpha_i y_i (\vec{x}_i^{\mathrm{T}} \vec{z}) + b$ or $\sum_{\vec{x}_i \in S} \alpha_i y_i \mathcal{K}(\vec{x}_i, \vec{z}) + b$ [44] [45]. Besides AP and these training locations, the remaining locations are used as test locations, which means that the channel vectors from these

Table 3.1: Number of training locations versus $|D_{tr}|/|D|$

Number of Training Locations	5	10	15
$|D_{tr}|/|D|$	11.63%	23.26%	34.88%

locations to the AP are used to test the output SVM machines. That is, these channel vectors are used to compute $\vec{x}_i$ in D_{test}. Then we test the output SVM machines using D_{test}. This is the first round of simulation, which is depicted in Figure 3.4. The square point in this figure represent the AP. The circles are the training locations, and the small triangles represent test locations. Another round of simulation should be done by picking another AP randomly firstly. The simulation outputs are averaged.

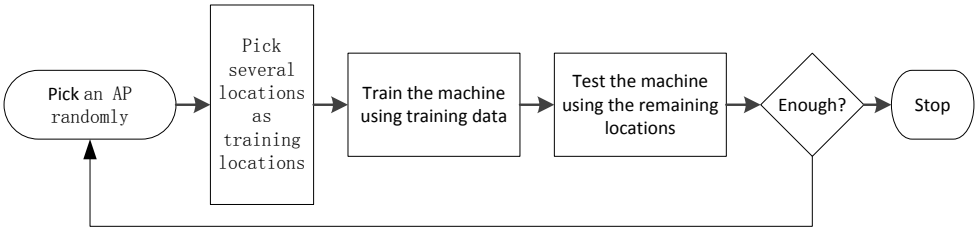

Figure 3.3: Simulation working flow

In reality, it is obvious that the training locations should be picked carefully to improve the performance of the output SVMs, because the training data should be presentive.

The number of training locations means how many locations in Figure 3.4 are used to generate D_{tr}. The relationship between the number of training locations and the size of D_{tr} is given in Table 3.1. When 5 locations in Figure 3.4 are used to generate D_{tr}, 11.63% of the total feature vectors are used for training.

Figure 3.5 and Figure 3.6 are the results. It can be seen that the performance becomes better as the size of D_{tr} increases. In Figure 3.5, the sum of misdetetion probability and false alarm probability can reach about 4%, when 34.88% of the data used for training. When different AP is selected, the detection probability and false alarm probability are plotted in Figure 3.6. Also, the average detection probability and false alarm probability are plotted too. From the two figures, it is shown that, polynomial SVMs are better than quadratic SVMs, which are better than linear SVMs. SVM based schemes have better performance than NPHT in terms of misdetection probability and false alarm probability.

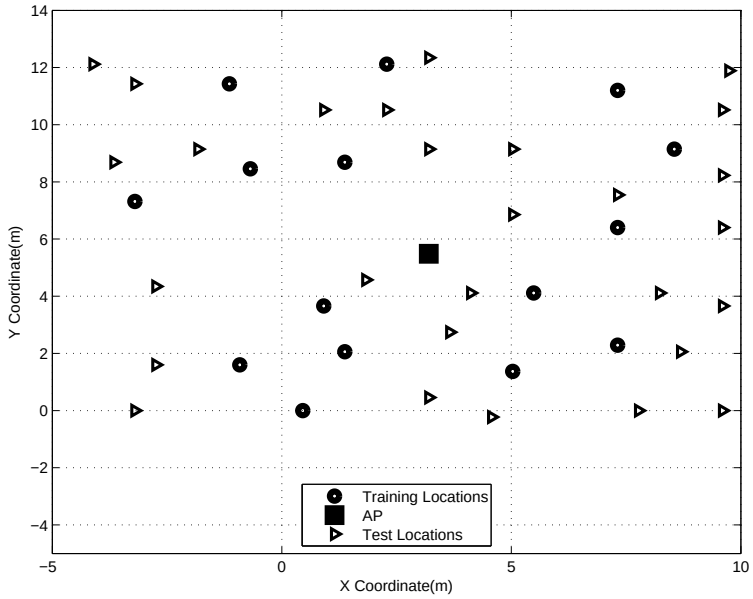

Figure 3.4: Room layout for simulation

SVMs can provide the weights and threshold at the same time, which is an advantage. While most existing schemes do not introduce the method to compute or set the threshold.

Among all the nonlinear SVMs, degree-3 polynomial SVM has the best performance in terms of ROC curves and $P_F + P_M$.

3.3 Complexity Analysis of SVM Based Schemes

Recall that $|S|$ denote the size of the set of the support vectors, $|D_{tr}|$ denotes the size of the training set, and J denote the dimension of the feature vectors. During the training stage, the complexity is related to the size of D_{tr}. There are $|D_{tr}|$ feature vectors to compute. The complexity of compute the feature vector is introduced in last chapter. Besides, the core to train a SVM is to feed to a QP Solver. The space complexity of solving QP problem is $O(J|D_{tr}|^2)$ [46]. Training time complexity of nonlinear SVMs is generally between $O(J|D_{tr}|^2)$ and $O(J|D_{tr}|^3)$ [41] [47]. The complexity comparison of these schemes during the training stage is shown in Table 3.2.

Most of the computation happens in the training phase. In the testing phase, the

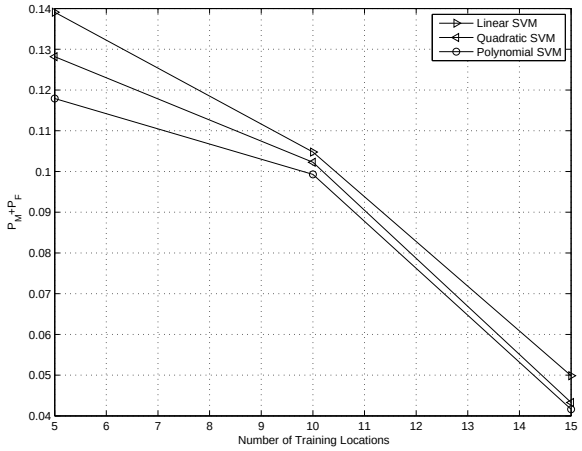

Figure 3.5: Comparison of different SVMs in terms of $P_M + P_F$ with different number of training locations

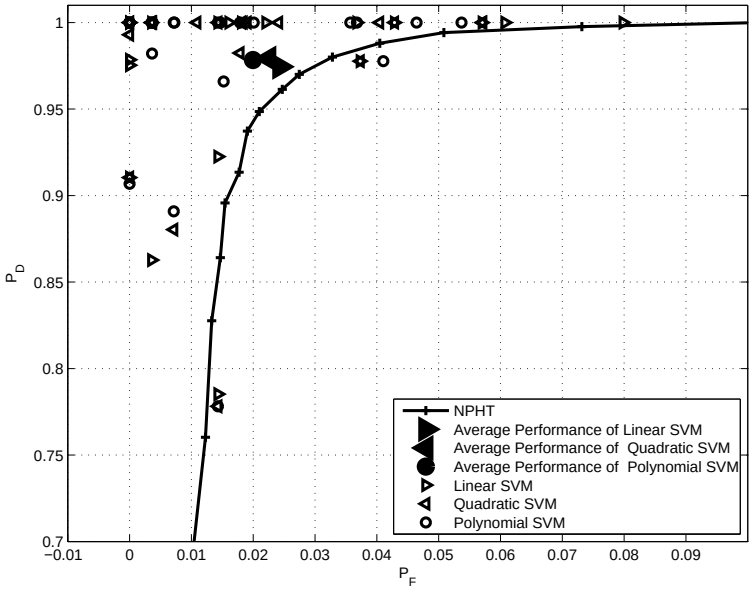

Figure 3.6: ROC curves of different SVMs when the number of training locations is 15

Table 3.2: Complexity comparison in training stage

Schemes	Computation of Feature Vectors		QP Solver	
	Number of Complex Multiplications	Number of Complex Additions	Time Complexity	Space Complexity
Linear SVM based scheme	$(\frac{N}{2}log_2N + PN)\|D_{tr}\|$	$(Nlog_2N + PN)\|D_{tr}\|$	$O(\|D_{tr}\|J)$	$O(\|D_{tr}\|^2J)$ [46]
Quadratic SVM based scheme	$(\frac{N}{2}log_2N + PN)\|D_{tr}\|$	$(Nlog_2N + PN)\|D_{tr}\|$	between $O(J\|D_{tr}\|^2)$ and $O(J\|D_{tr}\|^3)$	$O(\|D_{tr}\|^2J)$ [46]
Polynomial SVM based scheme	$(\frac{N}{2}log_2N + PN)\|D_{tr}\|$	$(Nlog_2N + PN)\|D_{tr}\|$	between $O(J\|D_{tr}\|^2)$ and $O(J\|D_{tr}\|^3)$	$O(\|D_{tr}\|^2J)$ [46]

computation is very simple. During the testing stage, in order to authenticate one frame, only the feature vector should be computed according to (3.18), (3.19), (3.20) firstly. When linear SVMs is used, the testing of new feature vector just involves the inner product of J-dimensional vectors. When nonlinear SVMs are used, the testing of new feature vector involves the kernel tricks. Assume that $\mathcal{M}$ is the number of operations required to evaluate the kernel functions. The complexity comparison is listed in Table 3.3.

3.4 Summary

In this chapter, our work is listed in the following.

- How to use SVMs to solve the problem in (2.6) is given. Our proposed SVM based scheme is to gain the classifiers $\sum_{\vec{x}_i \in S} \alpha_i y_i (\vec{x}_i^{\mathrm{T}} \vec{z}) + b$ or $\sum_{\vec{x}_i \in S} \alpha_i y_i \mathcal{K}(\vec{x}_i, \vec{z}) + b$ using $D_{E,tr}$ and $D_{A,tr}$.
- How to use the measurement data to do simulation is given.
- Simulation results show that our SVM based authentication schemes can gain satisfactory detection probability and false alarm probability.

Table 3.3: Complexity comparison in testing stage

Schemes	Time Complexity		Space Complexity
	Number of Complex Multiplications	Number of Complex Additions	
Linear SVM based scheme	$\frac{N}{2}log_2N + PN + J/4$	$Nlog_2N + PN + J/2$	$O(J)$
Quadratic SVM based scheme	$\frac{N}{2}log_2N + PN + J/4 + J\mathcal{M}$	$Nlog_2N + PN + J/2 + J\mathcal{M}$	$O(J)$
Polynomial SVM based scheme	$\frac{N}{2}log_2N + PN + J/4 + J\mathcal{M}$	$Nlog_2N + PN + J/2 + J\mathcal{M}$	$O(J)$

- Time complexity and space complexity during training stage and test stage are analyzed.

Chapter 4

Fisher Discriminant Analysis Based Authentication Scheme

In this chapter, we use the features (β, θ, and λ) to construct a test statistic to solve the problem in (2.6). If the PDFs of these random variables (β, θ, and λ) are already known, we can use the likelihood-ratio test (LRT), which is considered to be optimal according to the Neyman-Pearson criterion [48]. The LRT test statistic Λ_{LRT} and LRT testing can be formulated as (4.1).

$$\Lambda_{LRT} = \frac{f(\beta, \theta, \lambda|\mathcal{H}_1)}{f(\beta, \theta, \lambda|\mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\lessgtr}} th_{LRT} \tag{4.1}$$

where th_{LRT} is the decision threshold.

For the LRT, we need to compute the conditional PDF of (β, θ, λ), which is intractable mathematically [49]. As a result, we always adopt an linear combination of these features as the test statistics Λ. Rejection region and accept region should be determined too. If the value of Λ falls into the rejection region, Bob accepts $\mathcal{H}_1$. Otherwise, $\mathcal{H}_0$ is accepted. Λ_i in (4.2) is of common forms, which are linear combinations of some features. Λ_i is computed from $\vec{h}_i$ and $\vec{h}_{i-1}$.

$$\begin{aligned} \Lambda_i &= \vec{w}^{\mathrm{T}} \vec{x}_i \\ &= \sum_{j=0}^{J-1} w_j x_{i,j} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\lessgtr}} th \end{aligned} \tag{4.2}$$

where $x_{i,j}$ is the feature, J is the total number of features used, w_j is the weight for $x_{i,j}$, and th is the threshold. In this thesis, we just use three features, so J is 3.

In this chapter, we focus on the construction of the test statistics and the determination of the parameters (including weights and thresholds).

4.1 Linear Fisher Discriminant Analysis Based Authentication

Our proposed Fisher discriminant analysis (FDA) Based Authentication scheme is to construct a linear combination of β, θ, and λ and use linear FDA (LFDA) and D_{tr} to determine the weights. Linear FDA is always used in statistics, pattern recognition, and machine learning to find a linear combination of features, which may be used as a linear classifier. Let Λ_{LFDA} denote the test statistic. Let th_{LFDA} denote the decision threshold. Then, the hypothesis test problem in (4.2) becomes:

$$\begin{aligned} \Lambda_{i,LFDA} &= \vec{w}_{LFDA}^{\mathrm{T}} \vec{x}_i \\ &= w_0\beta_i + w_1\lambda_i + w_2\theta_i \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\lessgtr}} th_{LFDA} \end{aligned} \tag{4.3}$$

(4.4) can guarantee that the value of the test statistic Λ_{LFDA} varies between 0 and 1, which means that the threshold th_{LFDA} should be set from $[0, 1]$.

$$w_0 + w_1 + w_2 = 1 \tag{4.4}$$

Since $f(\Lambda_{LFDA}|\vec{h}_t = \vec{h}_A)$ is different from $f(\Lambda_{LFDA}|\vec{h}_t \neq \vec{h}_A)$, th_{LFDA} can be utilized to distinguish them with acceptable error. The greater the difference, the better we can distinguish them. The optimal value of the weight vector is the one that maximizes the detection probability and minimizes the false alarm probability [16]. However, a closed-form expression of the optimal linear coefficients is difficult to gain because it is hard to gain all the statistical information about these features.

The existing linear FDA algorithm can be used to find a linear coefficient vector $\vec{w}_{LFDA}$ that clearly separates different channel vectors by assigning a higher linear coefficient to a more significant feature. Let S_B and S_W denote the between- and within-class scatter

matrices respectively [50]. $\vec{w}_{LFDA}$ can be gained by maximizing the so-called Rayleigh coefficient [51] in (4.5) with respect to $\vec{w}$, which is formulated in (4.6).

$$J(\vec{w}) = \frac{\vec{w}^{\mathrm{T}} S_B \vec{w}}{\vec{w}^{\mathrm{T}} S_W \vec{w}} \tag{4.5}$$

$$\vec{w}_{LFDA} = arg \ \ max_{\vec{w}} \ \ J(\vec{w}) \tag{4.6}$$

The selection of the weights is very important. The weights reflect the importance of the features. So, the selection of the weights directly affects the performance of authentication. When the statistics properties (S_B, S_W, $\vec{m}_0$, $\vec{m}_1$) of these features are available, we can get the best weights. However, it is difficult to obtain all the information of these features in practice. Instead, only a set of channel vectors is available. For example, some labeled data can be obtained through the higher layer authentication scheme in the first place. That is how we can get D_{tr} in reality. Based on the labeled (training) data D_{tr}, the linear FDA algorithm can be performed to determine the weights in (4.3).

We use the linearize FDA to compute the weights. According to [50], $\vec{w}_{LFDA}$ can be computed according to (4.7).

$$\vec{w}_{LFDA} = S_W^{-1}(\vec{m}_1 - \vec{m}_0) \tag{4.7}$$

where $\vec{m}_0$ is the mean of all the $\vec{x}_i$ in $D_{A,tr}$, and $\vec{m}_1$ is the mean of all the $\vec{x}_i$ in $D_{E,tr}$. $\vec{m}_0$ and $\vec{m}_1$ are given in (4.8) and (4.9), respectively.

$$\vec{m}_0 = \frac{1}{|D_{A,tr}|} \sum_{\vec{x}_i \in D_{A,tr}} \vec{x}_i \tag{4.8}$$

$$\vec{m}_1 = \frac{1}{|D_{E,tr}|} \sum_{\vec{x}_i \in D_{E,tr}} \vec{x}_i \tag{4.9}$$

S_W can be estimated according to (4.10):

$$\begin{aligned} S_W = & \sum_{\vec{x}_i \in D_{A,tr}} (\vec{x}_i - \vec{m}_0) \cdot (\vec{x}_i - \vec{m}_0)^{\mathrm{T}} \\ & + \sum_{\vec{x}_i \in D_{E,tr}} (\vec{x}_i - \vec{m}_1) \cdot (\vec{x}_i - \vec{m}_1)^{\mathrm{T}} \end{aligned} \tag{4.10}$$

In summary, our proposed LFDA based authentication scheme to construct a linear combination of β_i, θ_i, and λ_i and to use linear FDA and D_{tr} to determine the weights. Specially, we use (4.7), (4.9), (4.8) and (4.10) to compute these weights.

4.2 Performance Analysis and Complexity Analysis

In this section, performance analysis and complexity analysis of our proposed LFDA based authentication are presented. Firstly, we study the effect of the training size. Then, we compare our proposed LFDA based authentication with existing schemes (CTS, RMTS, and NPHT) in terms of performance and complexity. Lastly, the complexity analysis is given.

4.2.1 Size of Training Set

The ROC curves and $P_M + P_F$ performance of LFDA based scheme with different number of training locations are shown in Figure 4.1 and Figure 4.2, respectively. These two figures are used to show the effect of the size of training set. It is obvious that the larger size the training data set, the better the performance of our proposed scheme. When 34.88% of data are used for training, our scheme can gain satisfactory performance.

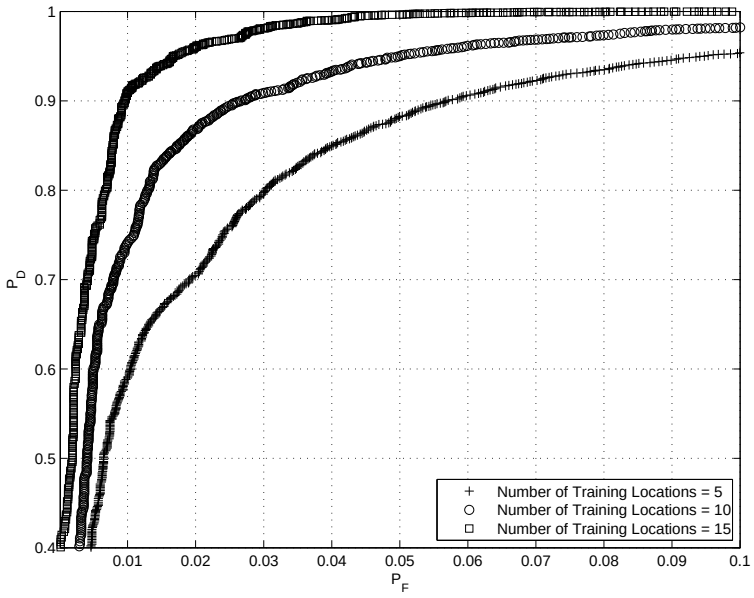

Figure 4.1: ROC curves of LFDA based scheme under different number of training locations

The optimal thresholds for different APs are different in fact. In Figure 4.2, "Optimal Threshold" means that the optimal threshold for each AP is gained by brute force in

each round of simulation. In Figure 4.1, "Optimal Threshold" is not given in the legend, which means that the detection probabilities and false alarm probabilities of each threshold are averaged on all APs and the threshold with minimum averaged sum of detection probabilities and false alarm probabilities is gained by brute force.

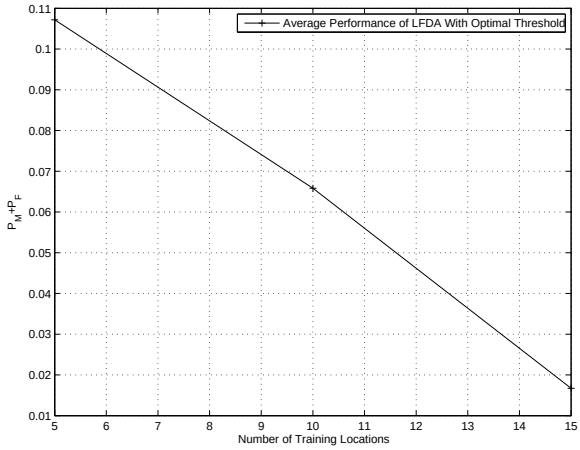

Figure 4.2: $P_M + P_F$ of LFDA based scheme under different number of training locations

4.2.2 Comparing LFDA Based Scheme with Other Schemes

We use similar methods introduced in last chapter to generate D_{tr} and D_{test}. We use D_{tr} to compute $\vec{w}_{LFDA}$ according to (4.7), (4.9), (4.8) and (4.10). Then we vary the threshold th_{LFDA} and use (4.3) to test the performance of our FDA based scheme on D_{test}. There are two cases of simulating linear FDA based scheme. The first case is that we know D before we compute the linear classifier (denoted as "LFDA global"). In this case, we use $D_{tr} \cup D_{test}$ to compute $\vec{w}$ first and then test the $\vec{w}$ on the $D_{tr} \cup D_{test}$. The second case is that we use 34.88% of the data set as D_{tr} to compute the linear classifier and then test the output linear classifier on the remaining data. This case is denoted as "LFDA training". We compare the proposed scheme with the NPHT scheme in [25]. The ROC curves are plotted in Figure 4.3, which is used to compare the ROC curves of LFDA based scheme and NPHT. In this figure, it is shown that the performance of our proposed scheme is better than NPHT. When P_F is 1%, P_D of NPHT is 70%, and P_D of the proposed scheme is extremely close to 95%. It implies that the proposed scheme can achieve higher detection probability, given the same false alarm probability.

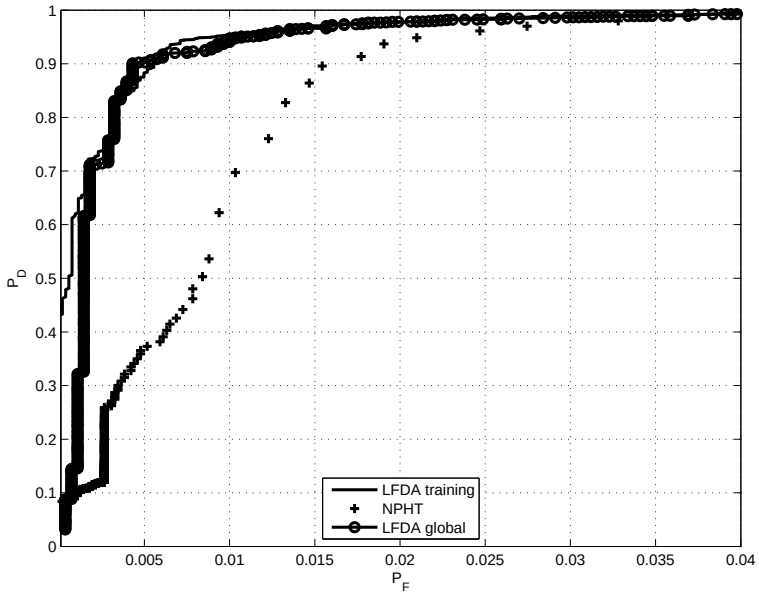

Figure 4.3: ROC curves of LFDA based scheme and NPHT

Table 4.1: Comparison with CTS

Targeted P_F	0.04	0.06	0.08
P_D of CTS	0.9467	0.9633	0.9744
P_D of LFDA based scheme	0.9898	0.9985	0.9997

We also compare the proposed scheme with CTS and RMTS schemes in [43]. When P_F is 0.1, P_D of RMTS is 96.5%, whileP_D of the proposed scheme is extremely close to 99.98%. Table 4.1 shows that the proposed scheme has higher P_D given the same P_F compared with CTS.

From Figure 4.4, the LFDA based scheme is better than SVM based schemes in terms of the sum of P_M and P_F.

4.2.3 Noise's Effects

Because in the experiments, all the pairwise links were measured for several times and averaged finally. The measurement can be considered free of measurement noise. But in

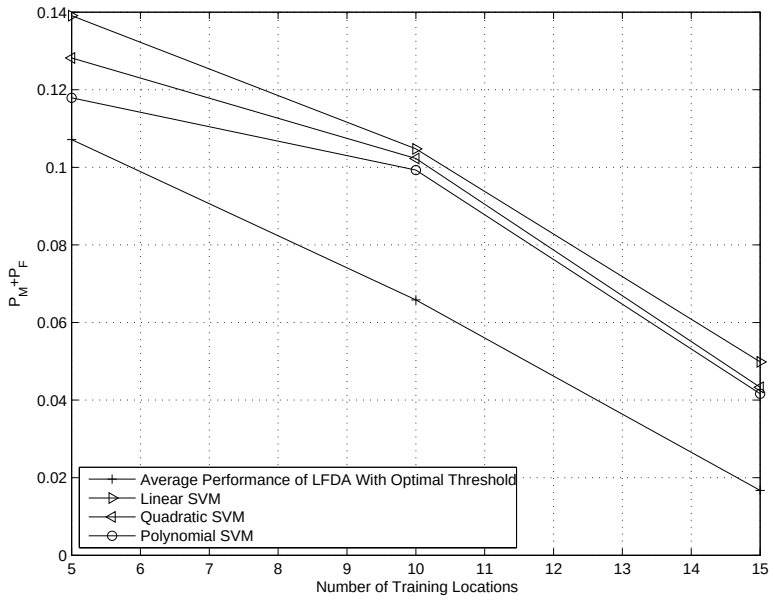

Figure 4.4: Comparison between SVM based schemes and LFDA based scheme in terms of $P_M + P_F$

reality, measurements always involves noise. As a result, we need to consider the effect of measurement noise on our proposed scheme and existing schemes. Considering the measurement noise, we define the signal-to-noise ratio (SNR) as the channel power to the measurement noise power. From Figure 4.5, it can be seen that under different SNRs, our proposed scheme always has better ROC curves than NPHT. In sum, our proposed scheme is better than NPHT in terms of ROC curves.

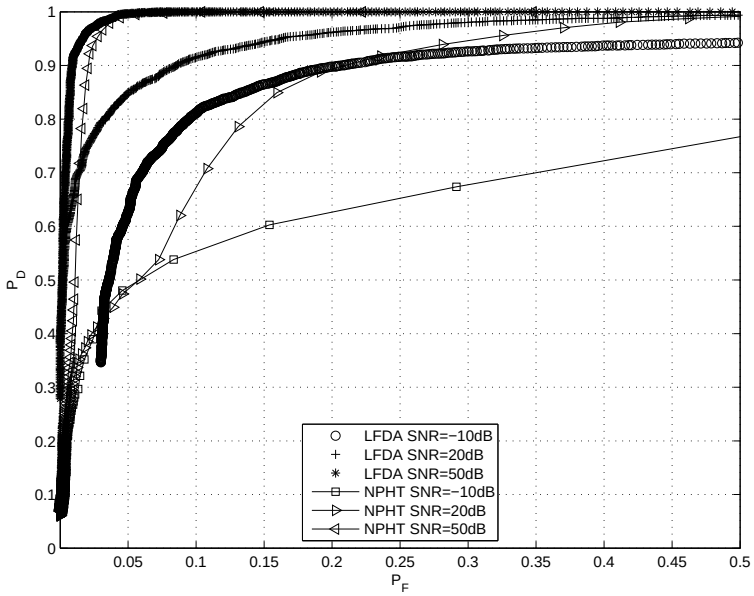

Figure 4.5: ROC curves of LFDA based scheme and NPHT under different SNRs

4.2.4 Complexity Analysis of LFDA Based Scheme

In this section, the complexity of our proposed LFDA based authentication scheme is analyzed. During the testing stage, in order to authenticate one frame, the feature vector should be computed according to (3.18), (3.19), (3.20) firstly. The complexity of the computation of the features is given in Table 2.1. After that, the inner product of two J-dimensional vectors should be computed to be compared with a threshold according to (4.3). The complexity comparison is listed in Table 4.2. In fact, NPTH involves a minimum problem, which makes NPTH more complex than just FFT computation (marked by $\divideontimes$ in Table 2.1). So we use Ω-notation.

Table 4.2: Complexity comparison in testing stage

Schemes	Time Complexity	
	Number of Complex Multiplications	Number of Complex Additions
LFDA based scheme	$\frac{N}{2}log_2N + PN + J/4$	$Nlog_2N + PN + J/2$
NPTH⁎	$\Omega(\frac{N}{2}log_2N)$	$\Omega(Nlog_2N)$
RMTS	$\frac{N}{2}log_2N$	$Nlog_2N$

During the training stage, the complexity of LFDA based scheme is related to the size of D_{tr}. NPTH and RMTS have no training stage. The complexity comparison of these schemes during the training stage is shown in Table 4.3.

Table 4.3: Complexity comparison in training stage

Schemes	Computation of Feature Vectors		Computation of $\vec{w}_{LFDA}$		
	Number of Complex Multiplications	Number of Complex Additions	Number of Real Multiplications	Number of Real Additions	Space Complexity
LFDA based scheme	$(\frac{N}{2}log_2N + NP)\|D_{tr}\|$	$(Nlog_2N + NP)\|D_{tr}\|$	$O(J^2\|D_{tr}\|)$	$O(J^2\|D_{tr}\|)$	$O(J\|D_{tr}\|)$ [49]
NPTH	0	0	0	0	0
RMTS	0	0	0	0	0

4.3 Adaptive Threshold Scheme (ATS)

4.3.1 ATS

Until now, how to set the decision threshold has not been discussed. We just researched the performance of our proposed scheme and the existing schemes by varying the threshold. The threshold defines the reject region and accept region. So, the selection of the threshold directly affects the performance of authentication. When the statistics properties of these

features are available, the optimal threshold can be determined by numerical method or derivation. However, in practice, it is difficult to obtain all the information of these features. Instead, only a set of channel vectors is available. In the section, how to set the initial threshold based on the training data, and how to adjust the threshold adaptively are discussed. The idea is that we set one initial value to the threshold at first. During the testing stage, the threshold is adjusted according to certain optimal objective. If the objective is to minimize the sum of misdetection probability and the false alarm probability, the threshold should be increased or decreased according to whether a attacker is detected and whether one legitimate user does not pass the test. If one attacker is not detected, the threshold is increased by one step. If the legitimate user is tested as the attacker, the threshold is decreased by one step. The step value affects the speed of iteration and how close the adjusted threshold approach the optimal threshold. If the step is set too big, the speed of iteration is high, but the final threshold may be too far from the optimal threshold. If the step is set too small, the speed of iteration is very slow. Another idea is to adjust the step value during the iteration process. We set the initial threshold according to (4.11), which is used to test the first feature vector. That is we compare Λ_1 with th_1. If $\Lambda_1 < th_1$, we reject $\mathcal{H}_0$ and ask the user to provide high-layer credentials. If the user can provide the right credentials, it is proven that this is a false alarm. The threshold should be updated to $th_1 + 0.5(\Lambda_1 - th_1)$ according to (4.12), where 0.5 is a factor. If the user cannot provide right credentials, it is proven the attack is detected precisely. The threshold should not be changed. Sometimes, we need to ask the user to provide the credentials randomly even if $\Lambda_i > th_i$. In this case, if the user can not provide the credentials rightly, the user is proven to be attacker. The threshold should be updated to $th_i + 0.5(\Lambda_i - th_i)$. This is the second proposed scheme: adaptive threshold scheme (ATS).

$$th_1 = 0.5(\vec{m}_0 + \vec{m}_1)^{\mathrm{T}} \cdot \vec{w}_{LFDA} \tag{4.11}$$

$$th_i \leftarrow th_{i-1} + 0.5(\Lambda_{i-1} - th_i) \tag{4.12}$$

We use simulation to verify the performance. The simulation should be done according to the following steps.

- Step 1: When the system is setup, the AP is selected.
- Step 2: Several locations are selected to generate D_{tr}. Based on D_{tr}, $\vec{w}_{LFDA}$ is worked out according to (4.7), (4.9), (4.8),and (4.10).
- Step 3: The initial threshold is computed according to (4.11).

- Step 4: D_{test} is generated. We simulate the real scenario, where the channel vectors come the AP one by one. During the process, the threshold is adjusted according to (4.12).

- Step 5: If the number of simulation rounds is big enough, stop. Otherwise, go to Step 1.

In Figure 4.6, the ROC curve of NPHT is depicted by varying the thresholds. This ROC curve is used for comparisons. We do several rounds of simulation. During each round we pick one AP randomly. We use the sequence number of AP to distinguish different APs. All the sequence numbers are depicted in Figure 3.4. When different AP is selected in each round, the false alarm probabilities and detection probabilities are depicted in Figure 4.6. The average false alarm probability and detection probability are also depicted in this figure. The sums of false alarm probability and detection probability are depicted in Figure 4.7. From these figures, it is shown that the LFDA based scheme with ATS is better than NPHT on average. It is also shown that the LFDA based scheme with ATS is worse than the LFDA based scheme with optimal threshold on average.

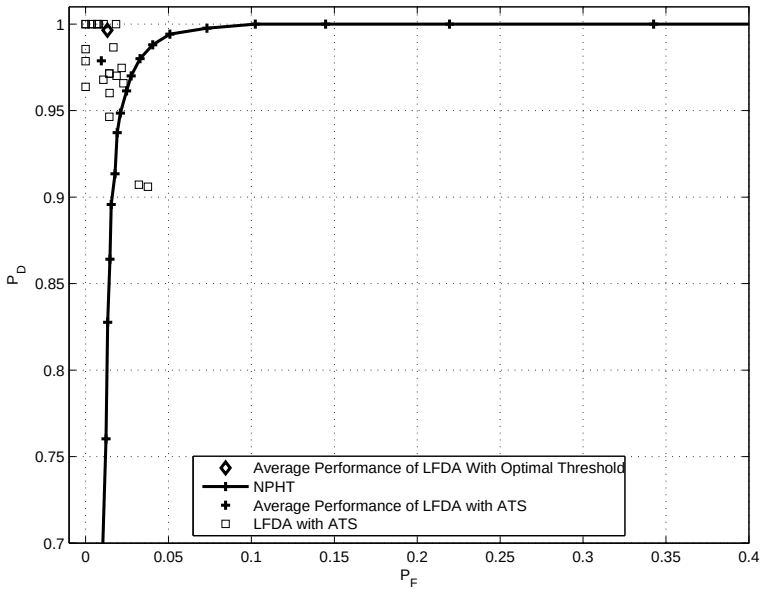

Figure 4.6: ROC curves of NPHT and LFDA based schemes with optimal threshold and with ATS

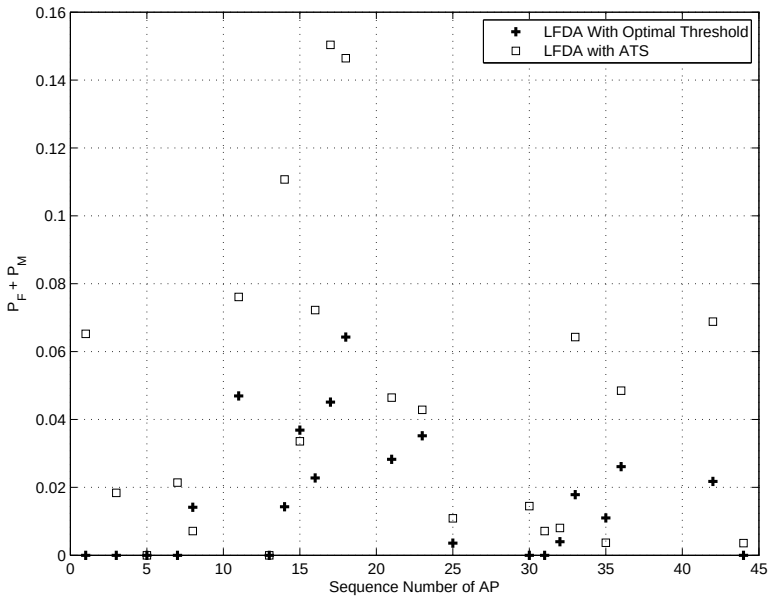

Figure 4.7: Comparison of LFDA schemes with optimal threshold and with ATS in terms of $P_F + P_M$

4.3.2 Comparison with SVM Based Schemes

From Figure 4.8, linear FDA based scheme with ATS has better performance than SVM based schemes in terms of false alarm probability and misdetection probability. SVM based schemed and LFDA based scheme with ATS are very practical, because the weights and thresholds are addressed.

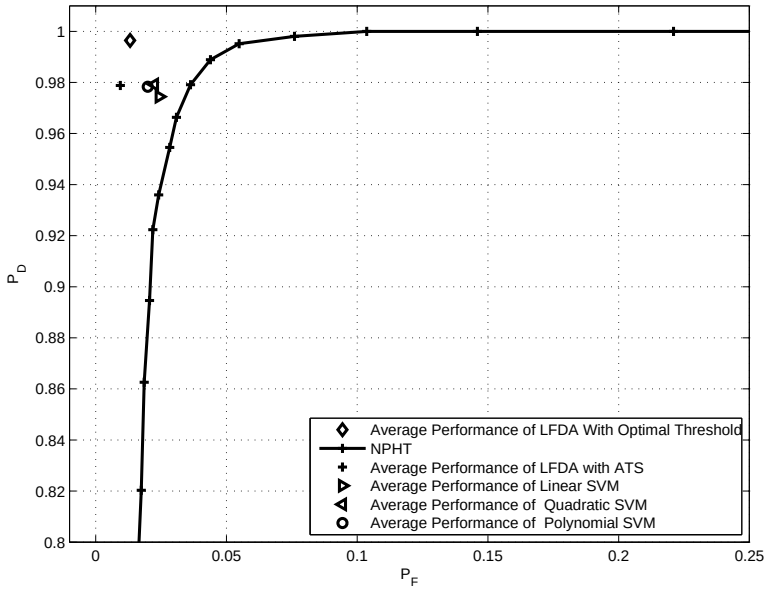

Figure 4.8: Comparison between SVM based schemes and LFDA based scheme in terms of P_D and P_F

In fact, $|D_{tr}| \gg J$ always holds. From Table 4.3, Table 4.2, Table 3.3, Table 3.2, we know that linear FDA based scheme has lower complexity than SVM based schemes.

4.4 Combining Scheme for Two-User Cooperative Authentication

In this section, we consider one special scenario, which is depicted in Figure 4.9. There are three legitimate users, Alice, Bob, and David, as well as a malicious user, Eve. Alice

is sending data to Bob. Eve tries to masquerade Alice and send frames to Bob. David can help Bob do the authentication.

The result in this subsection can be extended to multiple-antenna scenarios, where different antenna represents different legitimate user.

4.4.1 Combining Scheme

Let D_B denote the data set got by Bob. $D_{B,tr}$, used for training, is the subset of D_B.

$$D_B = \{\vec{x}_{B,i} | \vec{x}_{B,i} \in R^3, y_{B,i} \in \{-1, 1\}\}_{i=1}^{I} \tag{4.13}$$

The subscript B means "the feature vector is computed by Bob". D_B consists of training data set $D_{B,tr}$ and testing data set $D_{B,test}$.

Let D_{Da} denote the data set got by David. $D_{Da,tr}$, used for training, is the subset of D_{Da}.

$$D_{Da} = \{\vec{x}_{Da,i} | \vec{x}_{Da,i} \in R^3, y_{Da,i} \in \{-1, 1\}\}_{i=1}^{I} \tag{4.14}$$

The subscript Da means "the feature vector is computed by David". D_{Da} consists of training data set $D_{Da,tr}$ and testing data set $D_{Da,test}$.

$\vec{w}_{B,LFDA}$ and $\vec{w}_{Da,LFDA}$ are computed by Bob and David respectively, based on $D_{B,tr}$ and $D_{Da,tr}$ according to (4.7), (4.9), (4.8), and (4.10).

After $\vec{w}_{Da,LFDA}$ is computed, David sends the information $\vec{w}_{Da,LFDA}^{\mathrm{T}} \vec{x}_{Da,i} (\forall \vec{x}_{Da,i} \in D_{Da,tr})$ to Bob. These new combined feature vectors are $[w_{B,LFDA}^{\mathrm{T}} \vec{x}_{B,i}, w_{Da,LFDA}^{\mathrm{T}} \vec{x}_{Da,i}]^{\mathrm{T}}$. Bob uses these new feature vectors to compute the $\vec{w}_{BD,LFDA}$ according to (4.7), (4.9), (4.8), and (4.10).

The new test statistic is (4.15). Bob authenticate the i-th frame using the value of the test statistics according in (4.15).

$$\Lambda_{Combining} = [w_{B,LFDA}^{\mathrm{T}} \vec{x}_{B,i}, w_{Da,LFDA}^{\mathrm{T}} \vec{x}_{Da,i}] \vec{w}_{BD,LFDA} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\lessgtr}} th_{BD,LFDA} \tag{4.15}$$

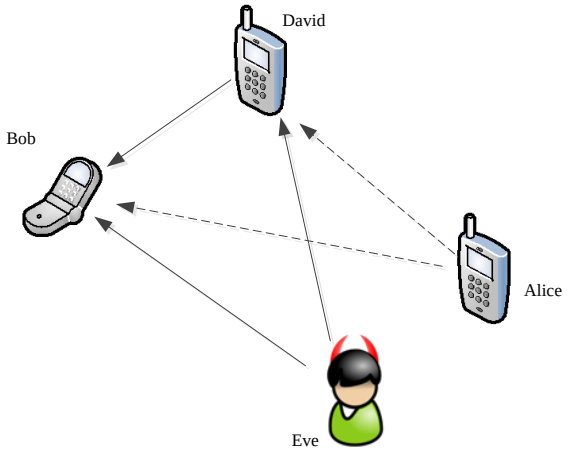

Figure 4.9: Combining scheme

4.4.2 Simulation Results of Combining Scheme

Two APs (AP1 and AP2) are selected randomly in each round of simulation. Linear FDA based scheme is simulated on AP1 and AP2, respectively. Then, the combining scheme is used on AP1, with the help from AP2. The simulation result is given in Figure 4.10. The performance of the combining scheme is better than the performance of LFDA without cooperation. It is shown that the combining scheme can improve the ROC curves at the cost of communication and computation overhead, which are analyzed next section.

4.4.3 Complexity Analysis of Combining Scheme

The complexity at David is analyzed in Table 4.2 and Table 4.3. The complexity at Bob is given in Table 4.4 and Table 4.5. Let L denote the total number of legitimate users who take part in the Bob's authentication. Here, L is 2.

As for the communication overhead, during training stage $|D_{tr}|$ numbers should be sent to Bob from David. During testing stage, only one number should be sent to Bob to authenticate each frame.

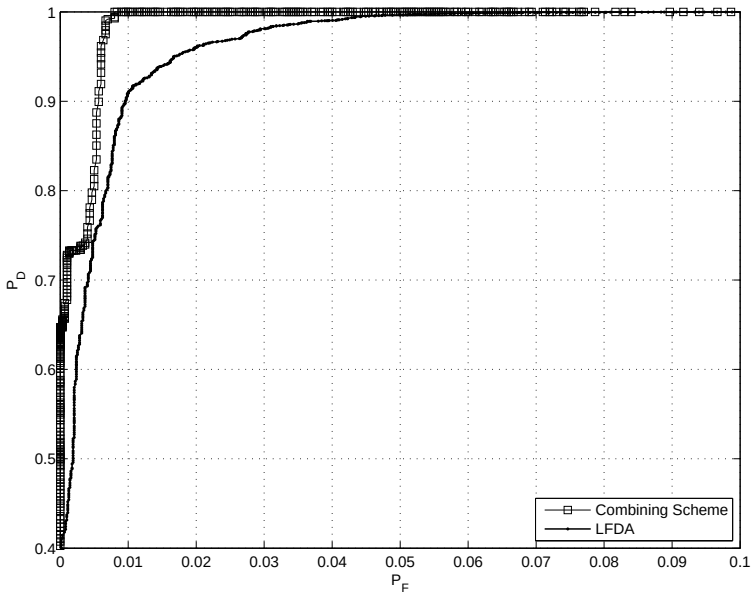

Figure 4.10: ROC curve of combining scheme

Table 4.4: Complexity in testing stage

Schemes	Time Complexity	
	Number of Complex Multiplications	Number of Complex Additions
Combining scheme	$\frac{N}{2}log_2N + PN + J/4 + L/4$	$Nlog_2N + PN + J/2 + L/2$

Table 4.5: Complexity in training stage

Schemes	Computation of Feature Vectors		Computation of $\vec{w}_{BD,LFDA}$		
	Number of Complex Multiplications	Number of Complex Additions	Number of Real Multiplications	Number of Real Additions	Space Complexity
Combining Scheme	$(\frac{N}{2}log_2N + NP)\|D_{tr}\|$	$(Nlog_2N + NP)\|D_{tr}\|$	$O((J^2 + L^2)\|D_{tr}\|)$	$O((J^2 + L^2)\|D_{tr}\|)$	$O((J + L)\|D_{tr}\|)$ [49]

4.5 Summary

In this chapter, our work is listed in the following.

- We introduce some background knowledge in FDA. How to use the linear FDA to compute the weights of the three features to generate a linear classifier is presented.
- We use the measurement data to verify that our proposed scheme have better ROC curves and better $P_F + P_M$ performance, when enough training data is provided. Also, our proposed scheme is less sensitive to noise than NPHT.
- An adaptive threshold scheme is presented in this thesis. LFDA based scheme with ATS is proven to have better performance than NPHT on average.
- Time complexity and space complexity during testing stage and training stage are analyzed.
- SVM based and LFDA based schemes are compared in terms of complexity, misdetection probability, and false alarm probability.
- Combining scheme for two-user cooperation is presented and analyzed.

Our proposed schemes:

- Linear FDA based scheme: We construct a linear combination of β, θ, and λ as the test statistic and use linear FDA to determine the weights.
- ATS: The threshold is set initially based on FDA and D_{tr}. Then, the threshold is adjusted according to certain optimal objective by variable steps during the testing stage.
- Combining scheme: When the cooperation between legitimate users is available, combining the data from other legitimate user can reduce the misdetection probability and false alarm probability at cost of extra computation and communication overhead.

Chapter 5

Conclusion and Future Work

In this thesis, we have studied the channel-based physical layer authentication. Existing literature on this topic has been reviewed firstly. Several channel-based authentication schemes, including SVM based schemes, the LFDA based scheme, and the combining scheme, have been proposed to improve the detection probability and reduce the false alarm probability. The sum of false alarm probability and misdetection probability can be reduced to less than 1.8% by our proposed LFDA based scheme. Furthermore, this sum can be reduced to less than 0.8% by our proposed combining scheme when two-user cooperation is available. When the false alarm probability is 8%, the detection probability can reach 99.97% by the LFDA based authentication scheme even without two-user cooperation.

Besides, we address the channel-based authentication problem from the aspect of machine learning. We do not make any assumptions that the statistics characteristics of the channels are known, which are common in existing channel-based schemes. For example, it is assumed that the underlying complex channel is a stationary, zero-mean, Gaussian random process in [27]. Instead, we try to provide better performance based on some training data, which can give our schemes more generality to be used in various kinds of environments. The kind of generality is gained at the cost of the training stage.

In addition, how to set the threshold is not addressed in most of existing work. We address the problem of setting the threshold in our thesis, which makes our proposed schemes more practical.

Lastly, the physical layer security is proposed to complement and enhance traditional security. Channel-based authentication should be integrated with upper layer security techniques to provide authentication. One candidate method is to use the physical layer authentication to trigger the upper layer authentication mechanisms. For example, if the

user does not pass the physical layer authentication, the user should provide upper layer credentials. In this case, the detection probability should be high enough. Our proposed scheme can guarantee the detection probability of 99.97%, which can be used in such kind of practical systems for authentication. Also, since the proposed schemes only need the channel vectors, it is easy to implement them without much modification on the existing systems.

The proposed authentication schemes are suitable for wireless network with relatively low nodes' mobility. For the future work, we will study the physical layer authentication for the scenarios with high nodes mobility. In addition, how to integrate our proposed physical layer authentication schemes with upper layer techniques should be an interesting one.

References

[1] Y.-S. Shiu, S. Y. Chang, H.-C. Wu, S.-H. Huang, and H.-H. Chen, "Physical layer security in wireless networks: a tutorial," *IEEE Wireless Communications*, vol. 18, no. 2, pp. 66–74, 2011.

[2] "Brute forcing wi-fi protected setup." [Online]. Available: http://www.coyotus.com/repo/pdf/hacking/viehboeck_wps.pdf

[3] M. Vanhoef and F. Piessens, "Practical verification of wpa-tkip vulnerabilities," in *Proceedings of ACM ASIACCS*, 2013, pp. 427–436.

[4] A. Mukherjee, S. Fakoorian, J. Huang, and A. Swindlehurst, "Principles of physical layer security in multiuser wireless networks: A survey," *IEEE Communications Surveys Tutorial*, vol. PP, no. 99, pp. 1–24, 2010.

[5] L. Dong, Z. Han, A. P. Petropulu, and H. V. Poor, "Improving wireless physical layer security via cooperating relays," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1875–1888, 2010.

[6] K. Zhang, X. Liang, X. Shen, and R. Lu, "Exploiting multimedia services in mobile social networks from security and privacy perspectives," *IEEE Communications Magazine*, vol. 52, no. 3, pp. 58–65, 2014.

[7] S. Gollakota and D. Katabi, "Physical layer wireless security made fast and channel independent," in *Proceedings of IEEE INFOCOM*, 2011, pp. 1125–1133.

[8] N. Zhang, N. Lu, N. Cheng, J. W. Mark, and X. S. Shen, "Cooperative spectrum access towards secure information transfer for crns," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 11, pp. 2453–2464, 2013.

[9] L. Xiao, L. Greenstein, N. Mandayam, and W. Trappe, "A physical-layer technique to enhance authentication for mobile terminals," in *Proceedings of IEEE ICC*, 2008, pp. 1520–1524.

[10] J. P. Vilela, M. Bloch, J. Barros, and S. W. McLaughlin, "Friendly jamming for wireless secrecy," in *Proceedings of IEEE ICC*, 2010, pp. 1–6.

[11] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector gaussian broadcast channel and uplink-downlink duality," *IEEE Transactions on Information Theory*, vol. 49, no. 8, pp. 1912–1921, 2003.

[12] F. Oggier and B. Hassibi, "The secrecy capacity of the mimo wiretap channel," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 4961–4972, 2011.

[13] J. Zhang, S. K. Kasera, and N. Patwari, "Mobility assisted secret key generation using wireless link signatures," in *Proceedings of IEEE INFOCOM*, 2010, pp. 1–5.

[14] J. W. Wallace, C. Chen, and M. A. Jensen, "Key generation exploiting mimo channel evolution: algorithms and theoretical limits," in *Proceedings of EuCAP*, 2009, pp. 1499–1503.

[15] D. B. Faria and D. R. Cheriton, "Detecting identity-based attacks in wireless networks using signalprints," in *Proceedings of ACM WiSe*, 2006, pp. 43–52.

[16] Y. Chen, J. Yang, W. Trappe, and R. Martin, "Detecting and localizing identity-based attacks in wireless and sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 5, pp. 2418–2434, 2010.

[17] Y. Chen, W. Trappe, and R. P. Martin, "Detecting and localizing wireless spoofing attacks," in *Proceedings of IEEE SECON*, 2007, pp. 193–202.

[18] K. Zeng, K. Govindan, and P. Mohapatra, "Non-cryptographic authentication and identification in wireless networks," *Elsevier Network Security*, vol. 1, p. 3, 2010.

[19] P. Luo, H. Li, G. Xu, and L. Peng, "Threat on physical layer security: Side channel vs. wiretap channel," in *Proceedings of IEEE ICCSE*, 2013, pp. 295–300.

[20] H. Wen, Y. Wang, X. Zhu, J. Li, and L. Zhou, "Physical layer assist authentication technique for smart meter system," *IET Communications*, vol. 7, no. 3, pp. 189–197, 2013.

[21] E.-K. Lee, M. Gerla, and S. Y. Oh, "Physical layer security in wireless smart grid," *IEEE Communications Magazine*, vol. 50, no. 8, pp. 46–52, 2012.

[22] L. Shi, M. Li, S. Yu, and J. Yuan, "Bana: body area network authentication exploiting channel characteristics," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 1803–1816, 2013.

[23] F. Guo and T.-c. Chiueh, "Sequence number-based mac address spoof detection," in *Recent Advances in Intrusion Detection*. Springer, 2006, pp. 309–329.

[24] P. L. Yu, J. S. Baras, and B. M. Sadler, "Physical-layer authentication," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 1, pp. 38–51, 2008.

[25] L. Xiao, L. J. Greenstein, N. B. Mandayam, and W. Trappe, "Using the physical layer for wireless authentication in time-variant channels," *IEEE Transactions on Wireless Communications*, vol. 7, no. 7, pp. 2571–2579, 2008.

[26] L. Xiao, L. Greenstein, N. Mandayam, and W. Trappe, "Fingerprints in the ether: Using the physical layer for wireless authentication," in *Proceedings of IEEE ICC*, 2007, pp. 4646–4651.

[27] J. K. Tugnait and H. Kim, "A channel-based hypothesis testing approach to enhance user authentication in wireless networks," in *Proceedings of ACM MobiCom*, 2010, pp. 1–9.

[28] N. Patwari and S. K. Kasera, "Robust location distinction using temporal link signatures," in *Proceedings of ACM MobiCom*, 2007, pp. 111–122.

[29] S. Jana and S. K. Kasera, "On fast and accurate detection of unauthorized wireless access points using clock skews," *IEEE Transactions on Mobile Computing*, vol. 9, no. 3, pp. 449–462, 2010.

[30] G. E. Suh and S. Devadas, "Physical unclonable functions for device authentication and secret key generation," in *Proceedings of ACM DAC*, 2007, pp. 9–14.

[31] F. J. Liu, X. Wang, and H. Tang, "Robust physical layer authentication using inherent properties of channel impulse response," in *Proceedings of IEEE MILCOM*, 2011, pp. 538–542.

[32] R. Clarke, "A statistical theory of mobile-radio reception," *Bell system technical journal*, vol. 47, no. 6, pp. 957–1000, 1968.

[33] A. T. Parameswaran, M. I. Husain, S. Upadhyaya *et al.*, "Is rssi a reliable parameter in sensor localization algorithms: An experimental study," in *Proceedings of FFDA*, 2009.

[34] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, "The cricket location-support system," in *Proceedings of ACM MobiCom*, 2000, pp. 32–43.

[35] P. Papadimitratos, M. Poturalski, P. Schaller, P. Lafourcade, D. Basin, S. Capkun, and J.-P. Hubaux, "Secure neighborhood discovery: a fundamental element for mobile ad hoc networking," *IEEE Communications Magazine*, vol. 46, no. 2, pp. 132–139, 2008.

[36] W. A. Gardner, "Exploitation of spectral redundancy in cyclostationary signals," *IEEE Signal Processing Magazine*, vol. 8, no. 2, pp. 14–36, 1991.

[37] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex fourier series," *Mathematics of computation*, vol. 19, no. 90, pp. 297–301, 1965.

[38] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[39] C. M. Bishop *et al.*, *Pattern recognition and machine learning*. Springer, 2006, vol. 1.

[40] A. Ben-Hur and J. Weston, "A users guide to support vector machines," in *Data mining techniques for the life sciences*. Springer, 2010, pp. 223–239.

[41] L. Bottou and C.-J. Lin, "Support vector machine solvers," *Large scale kernel machines*, pp. 301–320, 2007.

[42] J. A. Suykens, "Nonlinear modelling and support vector machines," in *Proceedings of IEEE IMTC*, vol. 1, 2001, pp. 287–294.

[43] N. Patwari, A. O. Hero, M. Perkins, N. S. Correal, and R. J. O'dea, "Relative location estimation in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 51, no. 8, pp. 2137–2148, 2003.

[44] S. Canu, Y. Grandvalet, V. Guigue, and A. Rakotomamonjy, "Svm and kernel methods matlab toolbox," *Perception Systmes et Information, INSA de Rouen, Rouen, France*, vol. 2, p. 21, 2005.

[45] "Train support vector machine classifier - matlab svmtrain." [Online]. Available: http://www.mathworks.com/help/stats/svmtrain.html

[46] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.

[47] I. W. Tsang, J. T. Kwok, and P.-M. Cheung, "Core vector machines: Fast svm training on very large data sets," *Journal of Machine Learning Research*, pp. 363–392, 2005.

[48] J. Neyman and E. S. Pearson, *On the problem of the most efficient tests of statistical hypotheses*. Springer, 1992.

[49] G. Ding, Q. Wu, Y.-D. Yao, J. Wang, and Y. Chen, "Kernel-based learning for statistical signal processing in cognitive radio networks: Theoretical foundations, example applications, and future directions," *IEEE Signal Processing Magazine*, vol. 30, no. 4, pp. 126–136, 2013.

[50] R. A. Fisher, "Contributions to mathematical statistics." 1950.

[51] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.

[52] V. Brik, S. Banerjee, M. Gruteser, and S. Oh, "Wireless device identification with radiometric signatures," in *Proceedings of ACM MobiCom*, 2008, pp. 116–127.

[53] T. Joachims, "Training linear svms in linear time," in *Proceedings of ACM SIGKDD*, 2006, pp. 217–226.