THE IMPACT OF RAYLEIGH FADING CHANNEL EFFECTS ON THE RF-DNA FINGERPRINTING PROCESS

By

Mohamed Fadul

Donald R. Reising Assistant Professor (Committee Chair)	Abdul R. Ofoli UC Foundation Associate Professor (Committee Member)
Thomas D. Loveless UC Foundation Assistant Professor (Committee Member)	

THE IMPACT OF RAYLEIGH FADING CHANNEL EFFECTS

ON THE RF-DNA FINGERPRINTING PROCESS

By

Mohamed Fadul

A Thesis Submitted to the Faculty of the University of
Tennessee at Chattanooga in Partial
Fulfillment of the Requirements of the Degree
of Master of Science: Engineering

The University of Tennessee at Chattanooga
Chattanooga, Tennessee

August 2018

ABSTRACT

The Internet of Things (IoT) consists of many electronic and electromechanical devices connected to the Internet. It is estimated that the number of IoT-connected devices will be between 20 and 50 billion by the year 2020. The need for mechanisms to secure IoT networks will increase dramatically as 70% of the edge devices have no encryption. Previous research has proposed RF-DNA fingerprinting to provide wireless network access security through the exploitation of PHY layer features. RF-DNA fingerprinting takes advantage of unique and distinct characteristics that unintentionally occur within a given radio's transmit chain during waveform generation. In this work, the application of RF-DNA fingerprinting is extended by developing a Nelder-Mead-based algorithm that estimates the coefficients of an indoor Rayleigh fading channel. The performance of the Nelder-Mead estimator is compared to the Least Square estimator and is assessed with degrading signal-to-noise ratio. The Rayleigh channel coefficients set estimated by the Nelder-Mead estimator is used to remove the multipath channel effects from the radio signal. The resulting channel-compensated signal is the region where the RF-DNA fingerprints are generated and classified. For a signal-to-noise ratio greater than 21 decibels, an average percent correct classification of more than 95% was achieved in a two-reflector channel.

ACKNOWLEDGEMENTS

First, I would like to express my sincere gratitude to my advisor, Dr. Reising, for the continuous support of my thesis study and related research, and for his patience, motivation, and immense knowledge. His guidance helped me over the whole course of the research and writing for this thesis.

Besides my advisor, I would like to also thank the rest of my thesis committee: Dr. Loveless, and Dr. Ofoli, for their insightful comments and encouragement. I would also like to thank the Department of Electrical Engineering of The University of Tennessee at Chattanooga for all its support and the great professors it provided.

TABLE OF CONTENTS

CHAPTER

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AWGN, Additive White Gaussian Noise

CFO, Carrier Frequency Offset

DFT, Discrete Fourier Transform

DGT, Discrete Gabor Transform

GI, Guard Interval

GT, Gabor Transform

IoT, Internet of Things

LS, Least Square

LTE, Long Term Evolution

LTS, Long Training Symbol

MAC, Medium Access Control

MDA/ML, Multiple Discrimination Analysis/Maximum Likelihood

MMSE, Minimum Mean Square Error

N-M, Nelder-Mead

OFDM, Orthogonal Frequency Division Multiplexing

OSI, Open System Interconnect

PAM, Pulse Amplitude Modulation

PHY, Physical Layer

QAM, Quadrature Amplitude Modulation

RF-DNA, Radio Frequency Distinct Native Attribute

RFSICS, RF Signal Intercept and Collection System

SEI, Specific Emitter Identification

SNR, Signal-to-Noise Ratio

STS, Short Training Symbol

TDL, Tap Delay Line

Wi-Fi, Wireless Fidelity

WiMAX, Worldwide Interoperability for Microwave Access

WLANs, Wireless Local Area Networks

CHAPTER 1

INTRODUCTION

1.1 Overview

Specific Emitter Identification is a physical (PHY) layer approach focused on the discrimination of radios (RF Emitters) for the purpose of augmenting traditional digital-level network security techniques. One of the common Specific Emitter Identification techniques used in wireless networks is the Radio Frequency Distinct Native Attribute (RF-DNA). RF-DNA fingerprinting takes advantage of unique and distinct characteristics, behaviors, and interactions that unintentionally occur within a given radio's transmit chain during waveform generation.

Prior work has shown that RF-DNA fingerprinting is impacted by the channel through which the transmitted waveform propagates [1-12]. These works focused solely on Additive White Gaussian (AWGN) channel models. However, a major concern within wireless communication channels is multipath fading and it remains a minimally researched topic within RF fingerprinting as a whole. Multipath occurs when the transmitted signal interacts with objects along the transmission path. These objects lead to the transmitted waveform being reflected, scattered, or a combination thereof [13]. The result is two or more instances of the transmitted waveform reaching the receiving radio antenna; thus, the received signal is a combination of copies of the originally transmitted signal in which each copy experiences its own path delay, attenuation, and phase shift [13].

In the case of RF-DNA fingerprinting, multipath channel effects can distort, change, mask, and even eliminate the unique and distinct waveform characteristics exploited by the fingerprinting process. The goal is to mitigate or even eliminate multipath channel effects while preserving the unique and distinct characteristics, behaviors, and interactions that are exploited by the RF-DNA fingerprinting process.

1.2 Motivation

The Internet of Things (IoT) consists of many electronic and electromechanical devices connected to the Internet. It is estimated that the number of connected IoT devices will be between 20 and 50 billion by the year 2020 [14-16]. The need for mechanisms to secure IoT networks will increase dramatically as 70% of the edge devices employ weak or no encryption [17].

Wireless networks are governed by the Open System Interconnect (OSI) model described by Figure 1.1, which defines the services provided and data units produced at each of the seven layers. Traditionally, detection and security of unauthorized network access is provided using digital techniques within the higher layers of the OSI model, e.g., Network and Datalink layers. By default, the security mechanisms provided at the higher layers are independent and they ignore the Physical (PHY) layer, which is the first layer exposed to major malicious activities [5].

In an effort to enhance more traditional digital security mechanisms, RF-DNA fingerprinting has been proposed as a security approach by which to detect and prevent unauthorized wireless network access through the exploitation of PHY layer features [5]. There has been a significant amount of research in RF-DNA fingerprinting and the results have demonstrated success in the discrimination of wireless radios for the case of an AWGN channel model [2, 4-8, 10-12, 18].

Despite the amount of work within RF-DNA fingerprinting, the investigation of radio discrimination when the transmitted waveforms have been subjected to a multipath channel remains largely unaddressed. There has been prior research into wireless device discrimination for the case of waveforms that have propagated through a multipath fading channel [19-21]. The contributions of these works as well as their limitations will be described in further detail within the next section and Chapter 2.

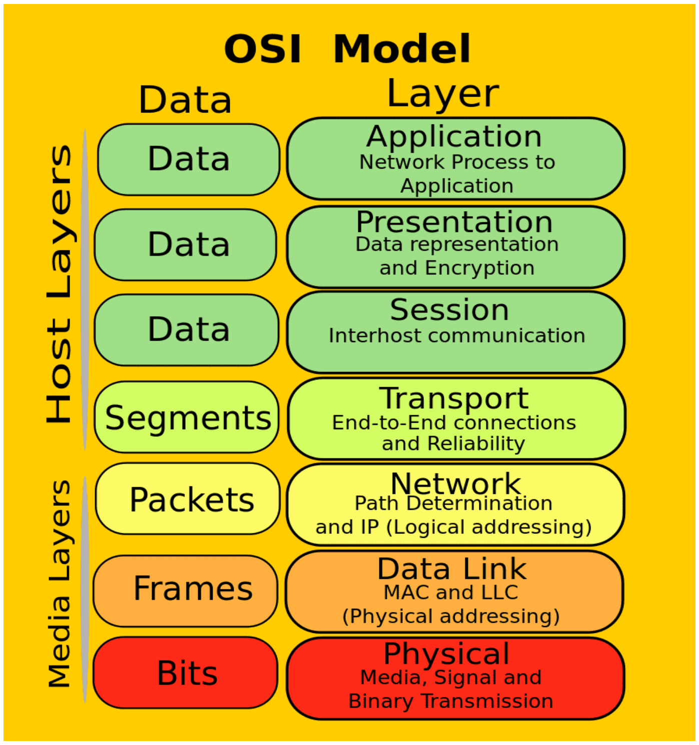

OSI Model
Data
Layer
Host Layers
Media Layers
Data
Application
Network Process to Application
Data
Presentation
Data representation and Encryption
Data
Session
Interhost communication
Segments
Transport
End-to-End connections and Reliability
Packets
Network
Path Determination and IP (Logical addressing)
Frames
Data Link
MAC and LLC (Physical addressing)
Bits
Physical
Media, Signal and Binary Transmission

Figure 1.1 Open Systems Interconnect (OSI) network model [1]

1.3 Problem Statement

Discrimination of wireless devices using waveforms that have been exposed to multipath environments is not new; however, the research has been limited to just a few efforts [19-21]. The work in [19] used an iterative approach to perform joint channel estimation and SEI within multipath environments. This approach estimates the path delays, associated attenuation coefficients, and discriminates between the wireless devices by computing the residual power between the received waveform and each waveform contained within a set of candidate waveforms associated with each of the 32 radios that are to be discriminated. All of the candidate waveforms were collected within an anechoic chamber to prevent interference from other transmitters as well as to remove multipath channel effects. The work in [19] presents the first case of waveform-based SEI under multipath channel conditions. Two key oversights in [19] are:

1) The details, e.g., distribution type, for the multipath channel were not provided. The multipath channel is simply described as, “an office multipath environment”.
2) SEI performance was not assessed under degrading Signal-to-Noise Ratio (SNR) conditions. The SNR of the assessed case was not presented at all; thus, limiting the scope of the work.

The work in [20, 21] implemented a Specific Emitter Identification (SEI) technique, by leveraging the features associated with the RF transmitter’s amplifier. Specifically, these works performed SEI using a constellation-based approach, which requires the demodulation of the transmitted waveform to facilitate discrimination via the complex symbols. This approach is distinctly different from RF-DNA fingerprinting, because the latter approach extracts the discriminating features from the waveform itself. Additionally, the work in [20, 21] utilized an RF front-end component that was modeled with fixed features for a given simulated transmission and across transmissions. The use of fixed features may unduly bias the discrimination results, because RF-DNA fingerprinting has demonstrated that a given feature can vary across a given radio’s transmissions [10]. To assess the multipath impact on SEI, this work simulated the channel with fixed coefficients instead of time varying channel coefficients.

Therefore, to advance the current state-of-the-art in SEI, the assessment of wireless radio discrimination using 1) a specific multipath model based upon a time-varying random process, and 2) degrading signal-to-noise ratio (SNR) conditions must be addressed. It is important to note that this work represents the first case of RF-DNA fingerprinting of wireless devices under multipath channel conditions.

1.4 Objectives

The objective of this research is to quantify RF-DNA fingerprinting in IEEE 802.11a Wireless Fidelity (Wi-Fi) indoor Multipath environments.

The work is divided into two parts, the first part objective is to estimate the channel impulse response from IEEE 802.11a signals that have been corrupted by the simulated Rayleigh fading channel. The objective of the second part is to use the estimated channel impulse response to compensate for the channel effects and prepare the signals for subsequent RF-DNA fingerprinting processing. In this work, the RF-DNA fingerprinting uses signals collected from 4 IEEE 802.11a radios

1.5 Research Contributions

This section provides a relational mapping between previous SEI in multipath environments research and the contributions of the current work presented in this thesis.

- This work presents a waveform based approach at the PHY layer, which means that the signal processing and analysis including fingerprints generation is done directly without demodulating the received signal. The work in [19] has presented an approach that is based on waveforms, while [20, 21] used the demodulated signals at higher layers for analysis and discrimination between devices.

- In this work, fingerprints are generated from 2-D, joint time-frequency (T-F) responses of signals using Discrete Gabor Transform (DGT), while the work in [20, 21] uses an approach that extracts the features from nonlinearities in transmitter’s amplifier. The approach presented in [19] does the fingerprinting and classification by computing the residual power between the received signal and a training database of signals.
- This work assesses the performance of the SEI in multipath environments under degrading signal-to-noise ratio (SNR), while the work in [19, 20, 21] do not.
- While performing channel estimation, the work in [19] didn’t specify the multipath channel model, while that in [20, 21] used a generic Tap Delay Line (TDL) channel with fixed coefficients. In this work, a Rayleigh fading channel is used to model a time varying indoor multipath environment with random coefficients.
- The work in [19] used an iterative approach that jointly estimate the channel impulse response and classify the received signals, while that in [20, 21] used a linear approximation approach for channel estimation and then proposed an iterative approach to improve the channel coefficients and transmitter’s nonlinearity estimation. This work uses Least Square (LS) technique to provide an initial channel estimate, then applies a Nelder-Mead (N-M) based approach to further improve the coefficients estimation. The use on N-M in this work represents the first application of N-M algorithm in channel coefficients estimation.

1.6 Thesis Outline

The remaining chapters of this thesis are organized as follows:

- Chapter 2: This chapter provides an overview of the literature on: prior RF fingerprinting involving multipath channels, the IEEE 802.11a Wi-Fi signal structure, Rayleigh fading channel model, Gabor Transform (GT) based fingerprint generation, and Multiple Discrimination Analysis/Maximum Likelihood (MDA/ML) classifier.
- Chapter 3: This chapter presents the channel estimation and removal process using the least square method for coarse multipath channel estimation and Nelder-Mead algorithm for fine multipath channel estimation. Lastly, an overview of the RF-DNA fingerprint generation and classification process is provided.
- Chapter 4: This chapter provides the simulation results and discussions on the performance of RF-DNA fingerprinting of radios operating within Rayleigh fading multipath channels.
- Chapter 5: This chapter concludes the findings and contributions of this work as well as proposes topics and challenges for follow-on research.

CHAPTER 2

BACKGROUND

This chapter presents the background needed for conducting the performance assessment of the RF-DNA fingerprinting process that is used to discriminate between wireless devices located in an indoor multipath environment. The chapter starts by introducing the signal of interest which is the IEEE 802.11a signal, then explains the concept of the multipath propagation and modeling using Rayleigh distribution. This chapter also presents the two channel estimation techniques that were used to estimate the multipath channel coefficients and delays in this work. The end of this chapter introduces the technique by which the RF-DNA fingerprints are generated and classified using the Multiple Discriminant Analysis, Maximum Likelihood (MDA/ML) classifier.

2.1 Signal of Interest

This work makes use of wireless radios which communicate using the IEEE 802.11a Wi-Fi standard [22]. The selection of this communication standard was made due to the significant amount of SEI research using 802.11a signals [3, 6, 9, 10, 20, 21, 23-26] as well as the use of the Orthogonal Frequency Division Multiplexing (OFDM) scheme in current and future wireless communication systems, e.g., 802.11ac, 802.11ad, 802.11ax, Long Term Evolution (LTE), Worldwide Interoperability for Microwave Access (WiMAX) [27]. The IEEE 802.11a standard is a wireless network protocol used in local and metropolitan area networks. The OSI model associates IEEE 802.11a to the PHY and Data-link layers [22]. The standard utilizes the a OFDM scheme to provide communication capabilities of 6, 9, 12, 18, 24, 36, 48, and 54 Mbit/s [22]. In OFDM systems, the high data-rate stream is broken down into many low-rate streams that are transmitted in parallel, which increases the symbol duration while reducing the ISI [28].

The IEEE 802.11a standard is an OFDM based system designed to efficiently utilize the allocated 300 MHz spectrum within the unlicensed national information infrastructure band from 5.15 to 5.725 GHz [22]. OFDM modulates information separately on to each of the 52 parallel subcarriers, and those subcarriers are equally spaced and orthogonal to each other [29]. Figure 2.1 shows the structure of the OFDM system adopted by the IEEE 802.11a in which the upper portion describes the transmitter and the lower portion describes the receiver. The data to be transmitted is first encoded using convolutional encoding techniques, then block interleaved to improve the symbol per bit error rate.

At the transmitter, the input data stream is mapped into N complex symbols, i.e., code words, in the frequency domain, including null data symbols for virtual subcarriers. These N complex code words are modulated onto N subcarriers of one OFDM symbol through the use of the Inverse Discrete Fourier Transform, which results in the time domain OFDM symbol given by,

$$x(m) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X(k) e^{\frac{j2\pi km}{N}} \tag{2.1}$$

where $m = 0, 1, 2, \ldots, N-1$, $X(k)$ denotes the data symbol in subcarrier k, $x(\mathrm{m})$ is the m^{th} sample of the OFDM symbol [30]. In IEEE 802.11a the number of data subcarriers is 52 while the total number of the parallel subcarriers including the virtual pilots is $N = 64$. A Cyclic extension often called the Guard Interval (GI) is added between adjacent OFDM symbols to reduce the ISI [31].

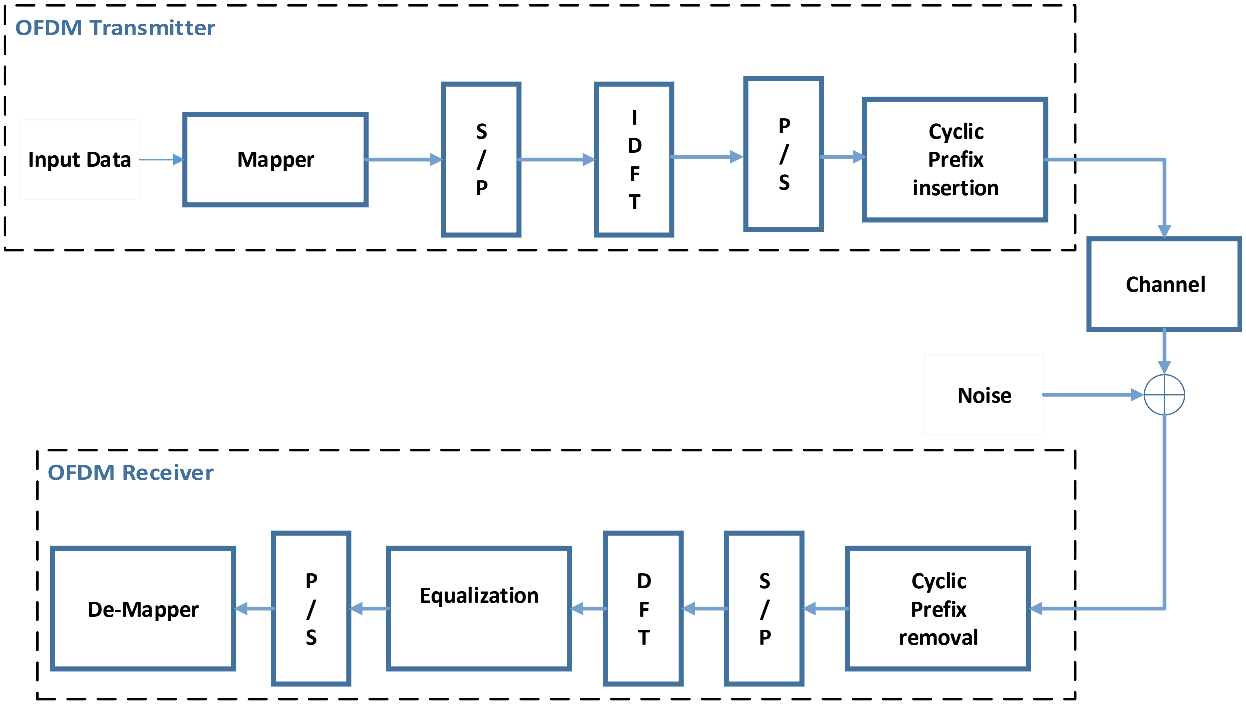


Figure 2.1 OFDM system model

Before transmission, the IEEE 802.11a OFDM frame is prepended with a known fixed symbol sequence known as a preamble. The preamble structure is illustrated in Figure 2.2 and is comprised of ten Short Training Symbols (STS) denoted by t1 through t10 and two Long Training Symbols (LTS) T1 and T2. Each STS and LTS are 0.8 µs and 3.2 µs in duration, respectively [10].

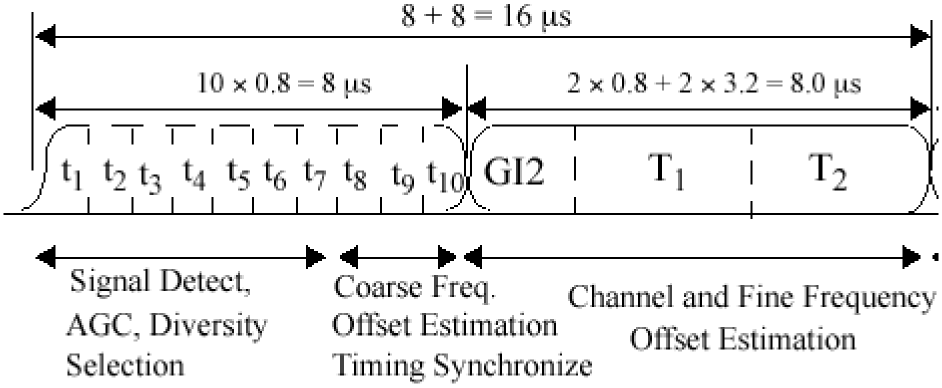

8 + 8 = 16 µs
10 × 0.8 = 8 µs
2 × 0.8 + 2 × 3.2 = 8.0 µs
t_1 t_2 t_3 t_4 t_5 t_6 t_7 t_8 t_9 t_{10}
GI2
T_1
T_2
Signal Detect, AGC, Diversity Selection
Coarse Freq. Offset Estimation Timing Synchronize
Channel and Fine Frequency Offset Estimation

Figure 2.2 Preamble structure [22]

IEEE 802.11a compliant radios use the preamble in the estimation of time and frequency offsets that can exist between the transmitter and receiver (time/frequency synchronization), as well as channel estimation. The preamble remains the same for all IEEE 802.11a Medium Access Control (MAC) frames and is always transmitted at 6 Mbps regardless of the bitrates used by the rest of the frame [22]. In this work, the preamble is used to estimate the multipath channel impulse response as well as the portion of the IEEE 802.11a signal from which RF-DNA fingerprints are extracted.

2.2 Channel Modeling

2.2.1 Overview

Multipath channel fading affects propagating signals over long and short distances, and is one of the severe challenges that must be overcome to maintain reliable high speed data communications within Wireless Local Area Networks (WLANs). Reflections and scattering of the transmitted signal caused by objects within the propagation environment, including human bodies, generate time and phase shifted versions of the transmitted signal which combine and interfere at the receiver [32]. Multipath channel fading effects can be summarized in three points as follows:

1) Changing the signal strength over small distances or during short time intervals.
2) Shifting the carrier frequency randomly as a result of transmitter or receiver motion.
3) Echoes caused by variable delays of multipath components [33].

Figure 2.3 shows a representative multipath environment in which the original transmitted signal is reflected by four objects that constitute 3 reflected paths, i.e., multipath components, in addition to the direct line of sight path.

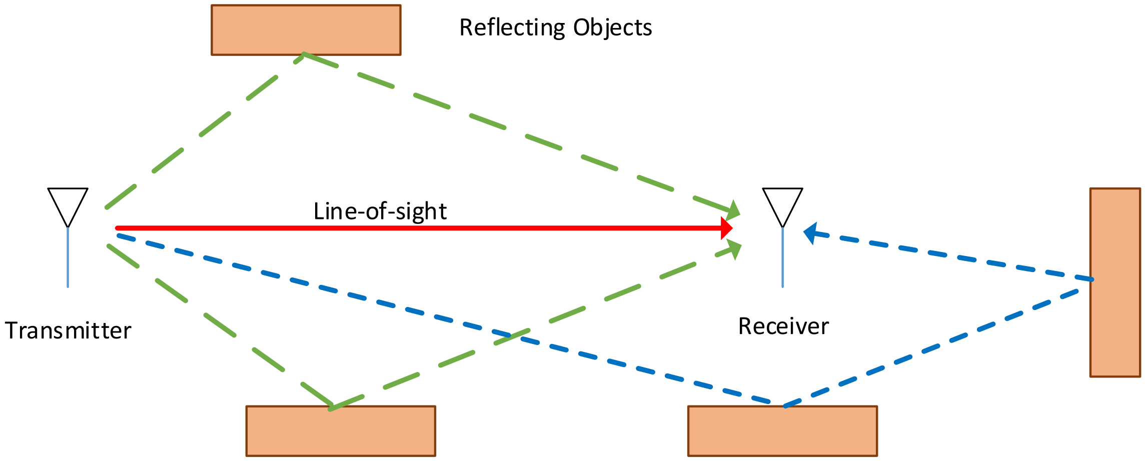


Figure 2.3 Multipath propagation scenario with a line-of-sight component and three multipath components [13]

If the transmitter transmits a single pulse through the channel, shown in Figure 2.3, it will traverse multiple paths from the transmitter to the receiver resulting in a final pulse train signal. The first pulse within the received pulse train will correspond to the line of sight component and the last one will correspond to the longest reflected path [13].

2.2.2 Indoor Multipath Channel Modeling

In general, there are two approaches by which to model multipath channel propagations: 1) deterministic modeling and 2) empirical, a.k.a., statistical, modeling. Deterministic models are based on the numerical calculations of the Maxwell's equations for certain environment conditions. These model are site specific and cannot be generalized as they take into account details of a specific environment they were developed for. Moreover, they are much more complicated than the empirical models as solving Maxwell's equations for certain conditions requires the availability of a database of all the environment details [34]. Empirical models are based on statistical data collected for various environments under different conditions. Empirical models take into consideration all of the environmental influences and conditions regardless of whether they can be separately recognized. Empirical models are more practical and greatly reduce the computational complexity [34]. One of the most common empirical models used in the representation of IEEE 802.11a indoor, multipath environments is the Rayleigh Channel Model.

2.2.3 Rayleigh Channel Model

Multipath is one of the major concerns for indoor environments, in which 802.11a Wi-Fi transceivers operate, because it negatively impacts their performance (destructive interference and random frequency shift) as explained in section 2.2.1 [35]. The Rayleigh distribution has been selected by the IEEE 802.11 working group to assess the performance of the modulation used in IEEE 802.11a and IEEE 802.11b wireless multipath environments [35]. The Rayleigh distribution is used in the creation of a channel model to exhibit a statistically time varying nature of the multipath environment in which one or more reflectors are present and the line-of-sight/direct path does not. The delay of the reflected path is one of the parameters used to characterize multipath and is known as the delay spread, τ [35]. The delay spread of the multipath channel is a random process that depends on the number of reflections, path length between the transmitter and receiver, the reflector materials, and transmitter/receiver motion. The delay spread will vary based upon the type of indoor multipath environment being modeled. The delay spread will be below 50 ns for home environments, and around 100 ns for office environments [35]. The multipath channel is modeled using a tap delay line in which each multipath component is represented by a single tap and associated delay as given by Figure 2.4 [28].

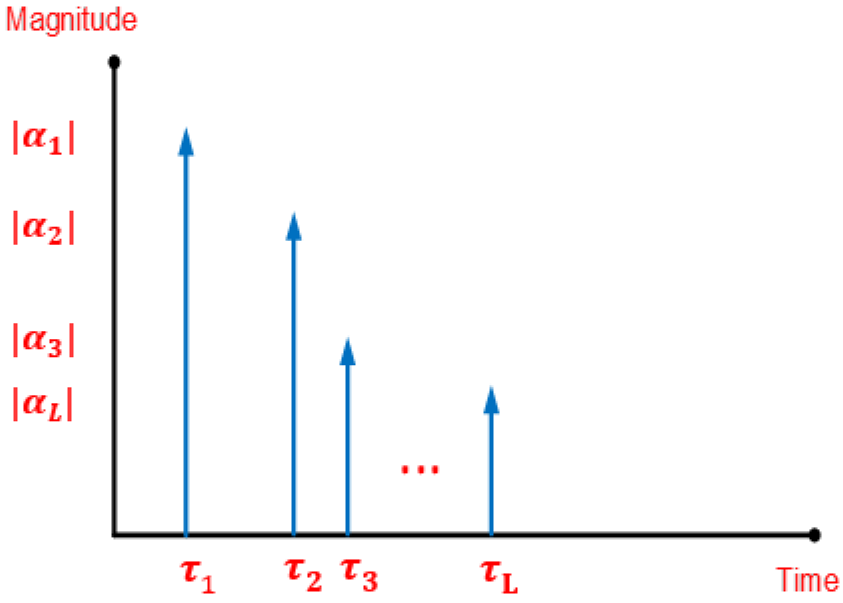


Figure 2.4 TDL representation of the Multipath channel

For Rayleigh fading channels, each tap coefficient can be modeled as circularly complex Gaussian random variable given by,

$$\alpha_k = A + jB \tag{2.2}$$

where k is the path index, and A, B are zero mean independent, identically distributed (iid) Gaussian random variables with variance σ^2. The variance σ^2 of the iid Gaussian random variable A, B is given by,

$$\sigma^2 = \frac{\sigma_k^2}{2} \tag{2.3}$$

where

$$\sigma_k^2 = \sigma_0^2 e^{\frac{-kT_s}{T_{RMS}}}, \tag{2.4}$$

T_s is the sampling period, T_{RMS} is the Root-Mean-Squared (RMS) delay spread of the channel, and σ_0^2 is the variance of the first multipath component given by,

$$\sigma_0^2 = 1 - e^{\frac{-T_s}{T_{RMS}}}. \tag{2.5}$$

Each path coefficient, α_k has statistics specified by the variance, σ_k^2, which should satisfy:

$$\sum \sigma_k^2 = 1,$$

and a magnitude drawn from a Rayleigh distribution. The tap delay line channel with L total paths is characterized by [28],

$$h(t,\tau) = \sum_{k=1}^{L} \alpha_k(t)\delta(t - \tau_k T_s), \tag{2.6}$$

where τ_k is the delay of the k^{th} path normalized by T_s [7]. If the transmitted IEEE 802.11a signal is $x(t)$, the received signal $y(t)$, filtered by a noisy multipath channel is,

$$y(t) = x(t) * h(t,\tau) + n(t) \tag{2.7}$$

where '$*$' denotes convolution, and $n(t)$ is complex Additive White Gaussian Noise (AWGN) with variance σ_n^2 [28].

2.3 Time Offset Estimation

Time offset estimation is a necessary step for enabling symbol timing synchronization in IEEE 802.11a WLAN systems [28]. Symbol timing synchronization facilitates the detection of whether a transmission is present within the wireless frame as well as determination of the start of the payload portion of the transmission [29]. OFDM systems are known to be sensitive to synchronization errors; therefore, improving the accuracy of timing offset estimation is essential to improving the overall system performance, i.e., recovery of the transmitted data. In IEEE 802.11a WLAN systems, the time offset is estimated upon reception of each transmitted Wi-Fi signal, and prior to estimation of the carrier frequency offset and channel impulse response. The estimated time offset is directly used to determine the delay of the first tap with respect to the receiver's point-of-view, and all the remaining calculated path delays are relative to that offset. The time offset estimation approach detailed here is based on the work presented in [30], it leverages the repeating 10 STSs and calculates the normalized autocorrelation function using the IEEE 802.11a preamble as an input.

A received OFDM signal transmitted through a discrete time multipath channel is expressed by,

$$r(m) = \sum_{k=0}^{L-1} x(m - \theta - \tau_k)h(k)e^{\frac{j2\pi\varepsilon m}{N_{sc}}} + n(m), \tag{2.8}$$

where m is the time index, $h(k)$ is the sampled complex channel impulse response, ε is the carrier frequency offset, $n(m)$ is the sampled complex AWGN, N_{sc} is the number of subcarriers used to each symbol, and θ is the time offset to be estimated [30]. The approach in [30] and [29], utilizes the autocorrelation of the ten STSs extracted from the received IEEE 802.11a preamble to estimate the time offset. The time offset is determined through the calculation of two normalized autocorrelation timing metrics $M_1(\theta)$ and $M_2(\theta)$. The first timing metric $M_1(\theta)$, is the normalized autocorrelation of the received preamble with a delayed copy of itself. The amount of delay for the first timing metric is exactly the duration of one STS, i.e., 0.8 □s. The result is a plateau that is nine STSs in duration, i.e., 7.2 μs, which begins at the start of the first STS. The second timing metric, $M_2(\theta)$, is the normalized autocorrelation of the received signal with itself delayed by two STSs. The result is a plateau that is eight STSs in duration, i.e., 6.4 μs [9]. The two timing metrics are given by,

$$M_1(\theta) = \frac{\sum_{m=0}^{N_s-1} r(\theta+m) r^*(\theta+m+N_s)}{\sum_{m=0}^{N_s-1} |r(\theta+m)|^2}, \tag{2.9}$$

and

$$M_2(\theta) = \frac{\sum_{m=0}^{N_s-1} r(\theta+m) r^*(\theta+m+2N_s)}{\sum_{m=0}^{N_s-1} |r(\theta+m)|^2}, \tag{2.10}$$

where N_s is the length of one STS. The start of the received preamble's ninth STS is found by computing the maximum of the difference of the two timing metrics as given by,

$$\hat{\theta} = a\, rg \underbrace{max}_{\theta} \left(M_1(\theta) - M_2(\theta)\right) \tag{2.11}$$

If $\hat{\theta}$ is earlier than the true time, part of the cyclic prefix of the current symbol is taken as data; thus, causing no interference. If $\hat{\theta}$ is later than the true time, part of the cyclic prefix of next symbol is taken as data, which results in ISI. Figure 2.5 shows the structure of the IEEE 802.11a MAC frame, where each Data symbol is prepended with a cyclic prefix denoted by GI. Figure 2.6, and Figure 2.7 show the case when the estimated time offset is earlier, and later than the true time respectively.

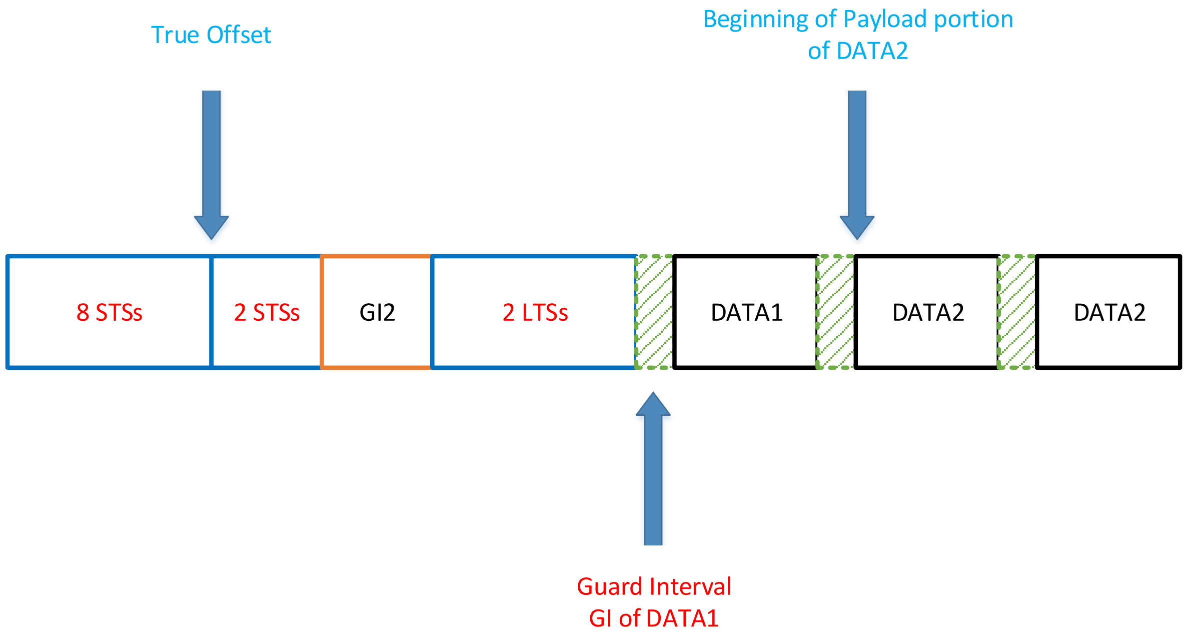

True Offset
Beginning of Payload portion of DATA2
8 STSs
2 STSs
GI2
2 LTSs
DATA1
DATA2
DATA2
Guard Interval GI of DATA1

Figure 2.5 True time offset and beginning of the 802.11a payload

2.4 Carrier Frequency Offset Estimation

Carrier Frequency Offset (CFO) is due to the frequency difference that can occur between a transmitter’s Local Oscillator (LO) and the LO within the receiver [10]. All OFDM systems are sensitive to the presence of CFO, and can only tolerate CFO values which are fractions of the spacing between subcarriers, 312 KHz [36].

The work in [36] presents a novel approach in the estimation of CFO. The approach uses the correlation of the frequency responses of two identical and consecutive training sequences to compute the phase difference between them. This phase difference is then used to calculate an estimated CFO value.

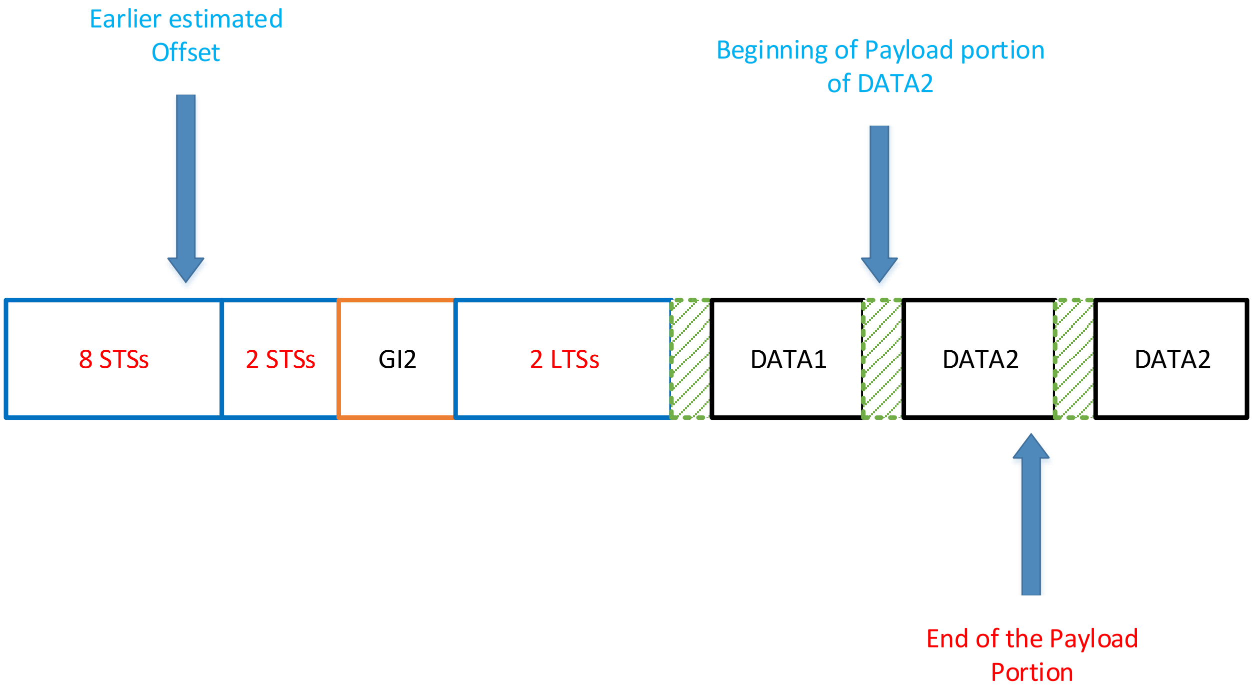

Earlier estimated Offset
Beginning of Payload portion of DATA2
8 STSs
2 STSs
GI2
2 LTSs
DATA1
DATA2
DATA2
End of the Payload Portion

Figure 2.6 Estimated time offset is earlier than the true time

If S(m) is the frequency response of the training sequence samples, the phase difference between frequency responses of two consecutive identical training sequences is given by [10],

$$\widehat{\emptyset} = angle\left\{\sum_{m=0}^{L-1} S(d+m)^{*} S(d+m+L)\right\}, \qquad (2.12)$$

where $*$ denotes the complex conjugate, d is the time index corresponding to the first sample in a window of $2L$ samples, and L is the length of one training sequence. The estimated carrier frequency offset $\Delta\hat{f}$ is calculated from the phase difference by,

$$\Delta\hat{f} = \frac{\widehat{\emptyset}}{\pi T}, \qquad (2.13)$$

where T is the time between the two sequences. The CFO given by equation (2.13), causes a rotation of the constellation as well as a random spread of the constellation points if uncorrected [36]. The estimated CFO can be removed from the complex In-Phase (I) and Quadrature (Q) samples of the time domain signal $s(n)$ using,

$$s'(n) = s(n)e^{-2\pi j \Delta \hat{f} t}, \tag{2.14}$$

where t is time. The approach presented in [36], estimates and removes the CFO in two steps. The first step is called "course" CFO estimation, where (2.13) and (2.14) are used with the two consecutive STSs t_8 and t_9. In the second step which is called "fine" CFO estimation and removal, the two LTSs T_1 and T_2, are used to remove any remaining CFO left after the "course" CFO estimation.

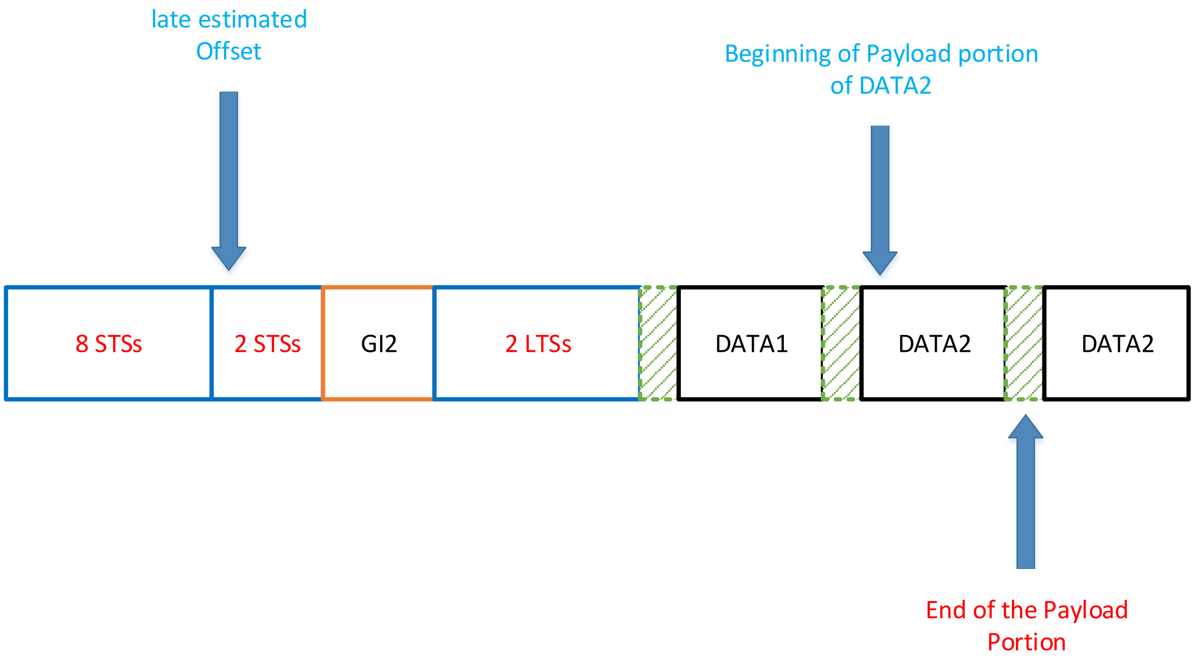


Figure 2.7 Estimated time offset is later than the true time

2.5 Channel Estimation and Equalization

In addition to the multipath channel’s path delays, each reflected path is associated with an attenuation amplitude value, which corresponds to the amplitude attenuation applied to the original transmitted signal. One technique used to compensate for these multipath effects is known as equalization. Equalization is implemented within the receiver as a filter in which the non-zero filter coefficient locations and values correspond to the path delays and inverses of the estimated attenuation values associated with each reflected path, respectively. The path delays and attenuation values form the multipath channel impulse response. Therefore, the goal within every wireless receiver is to estimate this impulse response to enable mitigation of the multipath effects on the received signal.

For OFDM systems, there are two general approaches to estimation of the channel’s impulse response. The first approach is based on using training data transmitted on each subcarrier to estimate the channel, while the second uses training information transmitted on a subset of the subcarriers [37]. In this work, two channel estimation techniques, based on the first approach, are presented. The first channel estimation technique is based on the work in [37], [38], [39] and is known as Least Square (LS) channel estimation. A brief explanation of the LS estimation process is presented in Section 2.5.1. The second channel estimation technique was developed within this effort and is built upon the Nelder-Mead direct search method presented in [40], [41]. The Nelder-Mead (N-M) approach estimates the channel coefficients through the minimization of an error function. The N-M direct search method will be explained in Section 2.5.2 and its application to channel estimation presented in Chapter 3.

2.5.1 Least Square Estimator

The two, known LTS are used to estimate the channel impulse response based on the work presented in [9], [4], [37]. The frequency response of the transmitted and received LTSs are denoted as X and Y, respectively. The frequency response of the channel impulse response is denoted as H. The relationship between X, Y, and H is given by,

$$Y = \begin{bmatrix} H_0 \\ H_1 \\ \vdots \\ H_{N_{sc}-1} \end{bmatrix} \begin{bmatrix} X_0 & 0 & \dots & 0 \\ 0 & X_1 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & X_{N_{sc}-1} \end{bmatrix} + \begin{bmatrix} N_0 \\ N_1 \\ \vdots \\ N_{N_{sc}-1} \end{bmatrix}, \qquad (2.15)$$

where N is the frequency response of the complex AWGN. The goal of the LS estimator is to minimize the cost function $J(\widehat{H})$ given by,

$$J(\widehat{H}) = (Y - HX)^H (Y - HX), \qquad (2.16)$$

where $\widehat{\mathrm{H}}$ is the estimate of the channel frequency response, and $(\square)^{\mathrm{H}}$ represents the conjugate transpose of matrix [10]. LS algorithm solves equation (2.16) and estimates the channel frequency response by using,

$$\widehat{H}_{ls} = X^{-1} Y. \qquad (2.17)$$

The use of both LTSs in (2.17) will result in the reduction of the variance of the noise and the square error of the channel estimate by one-half. This LS approach to channel estimation is given by,

$$\widehat{H}_{ls} = \frac{1}{2} . X^{-1} . (Y_1 + Y_2), \qquad (2.18)$$

where Y_1, and Y_2 are the Discrete Fourier Transforms of the first and second received LTSs, respectively.

2.5.2 Nelder-Mead Estimator

The Nelder-Mead estimator, developed within this work, is based upon the N-M simplex algorithm. N-M simplex algorithm is a direct search method used in optimization problems with the goal of minimizing a specific function. The N-M method is widely used due its robustness and computational efficiency [42]. It attempts to minimize a n-variable nonlinear function through an iterative process using only the function values, i.e., without the need for calculation of first nor second order derivatives [28]. The problem to be solved by the N-M algorithm can be defined by the formula:

$$\underbrace{minmize}_{x\epsilon R^n}\ f(x)$$

The N-M method uses an iterative approach to find the problem's solution. Each iteration begins with a simplex determined by $n+1$ vertices where n is the number of the scalar variables in that problem. For the case when $n=2$, the simplex is a triangle. The k^{th} iteration where $k \geq 0$, is terminated when the function value at the vertices of the simplex of that iteration satisfies a certain condition. This condition is presented later within this section. Complete definition of the N-M algorithm requires the specification of four parameters: reflection coefficient (ρ), contraction (γ), expansion (χ), and shrinkage (φ). These four parameters should satisfy the following constrains [37]:

$$\rho > 0 \quad \chi > 1 \quad \chi > \rho \quad 0 < \gamma < 1 \qquad 0 < \varphi < 1$$

Trial steps within each iteration are generated by reflection, expansion, contraction, and shrinkage operations. At each interaction k, the following steps and function evaluations are performed:

- Each iteration starts with a simplex defined by the points $x_i \epsilon R^n$, where $1 \leq i \leq n+1$. The simplex points should be ordered to satisfy $f(x_1) \leq f(x_2) \leq \cdots \leq f(x_{n+1})$,
- Compute the reflected vertex using,

$$x_r = (1+\rho)\overline{x} - \rho x_{n+1}, \tag{2.19}$$

where

$$\overline{x} = \sum_{i=1}^{n} \frac{x_i}{n}, \tag{2.20}$$

is the center of the n best points. It is important to note that the worst point, $x_{\mathrm{n+1}}$ is not included. The value of the function at x_r determines which operation is done next and whether the trial is accepted or rejected. If the function value at x_r is anywhere between the best point x_1 and the second worst point x_n, then the iteration is ended and the reflected point replaces the worst point x_{n+1} within the simplex of the next iteration. For the case when the function value, corresponding to the reflected point $f(x_r)$, is less than its value at $x_1$or greater than its value at x_n, one or more of the operations mentioned earlier needs to be done. If $f(x_r) \leq f(x_1)$, then the iteration is expanded and the expansion point is calculated by,

$$x_e = (1+\rho\chi)\overline{x} - \rho\chi x_{n+1}. \tag{2.21}$$

The iteration is ended and the expansion point x_e is accepted when $f(x_e) < f(x_r)$. If $f(x_r) \geq f(x_n)$, a contraction operation is performed and it has two cases, the first case is when $f(x_n) \leq f(x_r) \leq f(x_{n+1})$; thus, the outside contraction is performed using,

$$x_c = (1+\rho\gamma)\overline{x} - \rho\gamma x_{n+1}. \tag{2.22}$$

The function is then evaluated at x_c, and if $f(x_c) \leq f(x_r)$, then the iteration is ended and x_c is accepted to replace the worst point in the simplex for the next iteration. Otherwise, the shrink operation is performed. The second case is when $f(x_r) \geq f(x_{n+1})$, which results in an inside contraction being performed using,

$$x_{cc} = (1-\gamma)\overline{x} + \gamma x_{n+1}. \tag{2.23}$$

If $f(x_{cc}) \leq f(x_{n+1})$ the iteration is ended and x_{cc} is accepted; otherwise, the shrink operation is performed. The shrink operation is performed by calculating a new set of vertices v_i given by,

$$v_i = x_1 + \varphi(x_i - x_1), \tag{2.24}$$

where $2 \leq i \leq n+1$ and the new simplex for the next iteration is $(x_1, v_2, \dots, v_{n+1})$ [41]. In this work, two stopping criteria were adopted to terminate the iterations and end the search. A termination condition based on the function values given by,

$$\frac{1}{n}\sum_{i=1}^{n+1}(f(x_i) - \overline{f})^2 < \epsilon_1, \tag{2.25}$$

where $\overline{f}$ is the mean of the function and ϵ_1 is a tolerance based on function values. The second stopping criterion is based on the points x_i and is given by [41],

$$\frac{1}{n}\sum_{i=1}^{n}\left\|x_i^k - x_i^{k+1}\right\|^2 < \epsilon_2, \tag{2.26}$$

where $\|\blacksquare\|$ is the l_2 norm. If the l_2 norm of the point x_i between two successive iteration is less than the tolerance ϵ_2, then the algorithm is terminated [42].

In this work, the function to be minimized is defined as follows: If the received signal defined by (2.8) is corrected for the time and carrier frequency offset, then the function of interest $f(h)$ can be defined as a square error function given by,

$$f(h) = \sum_{k \in m} \left| r(m) - \sum_{k=0}^{L-1} x(m - \tau_k) h(k) \right|^2 , \tag{2.27}$$

where $r(m)$ is the received signal, $x(m)$ is the IEEE 802.11a transmitted preamble, and $h(k)$ is the multipath coefficient associated with the kth path.

2.6 RF-DNA Fingerprinting

This work assesses the impact of indoor multipath channel removal has on RF-DNA fingerprint based wireless radio discrimination performance. The RF-DNA fingerprints are extracted from two-dimensional, joint Time-Frequency (T-F) responses. Based on the work in [5], the features were extracted using the Discrete Gabor Transform (DGT) in which the complex Gabor coefficients are calculated as follows,

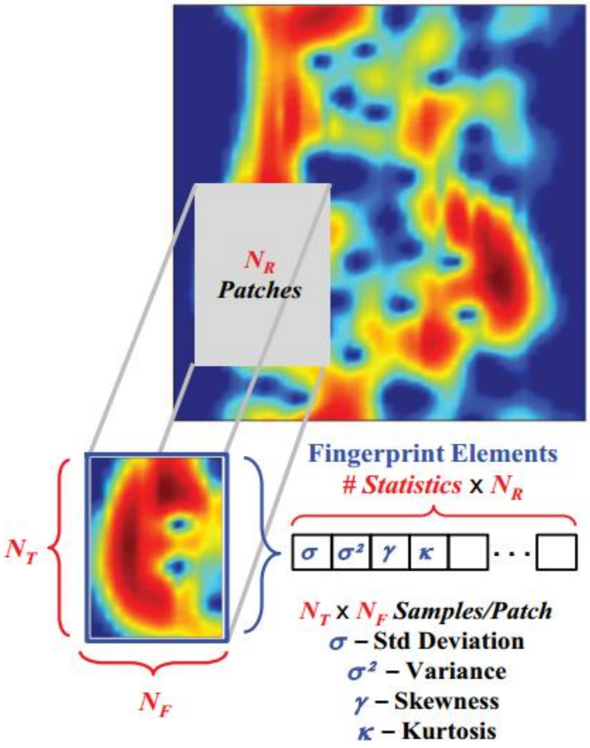
$$G_{mk} = \sum_{n=1}^{MN_\Delta} s(n) W^*(n - mN_\Delta) e^{-j\gamma}, \tag{2.28}$$

where G_{mk} are the Gabor coefficients, $s(n) = s(n + lMN_\Delta)$ is the periodic input signal, $W(n) = W(n + lMN_\Delta)$ is the periodic analysis window, $\gamma = 2\pi kn/K_G$, N_Δ is the total number of shifted samples, $m = 1, 2, \ldots, M$ for M total number of shifts, $k = 1, 2, \ldots, K_G - 1$ for $K_G \geq N_\Delta$ and $mod(MN_\Delta, K_G) = 0$ satisfied [10]. In calculation of the Gabor coefficients the signal $s(n)$ and window $W(n)$ should have finite support. The window function is usually required to have finite support, while the signal is allowed to have a very large or infinite support. If the signal is very large, overlap-add techniques are applied by splitting up the signal into smaller pieces, and each piece is treated separately [43]. The Gabor transformation is oversampled for the case of $K_G > N_\Delta$ and critical sampling occurs when $K_G = N_\Delta$. Oversampling is preferred when processing noisy data, because it adds more reliability to the analysis of signals under varying SNR conditions [5]. Thus, K_G and $N_\Delta$were selected to achieve the oversampling condition.

In Gabor based fingerprinting, the RF-DNA fingerprints are generated from the normalized magnitude-squared Gabor coefficients $|G_{mk}|^2$ which are given by [5],

$$\overline{|G_{mk}|^2} = \frac{|G_{mk}|^2 - min\{|G_{mk}|^2\}}{max\{|G_{mk}|^2 - min\{|G_{mk}|^2\}\}}. \tag{2.29}$$

Figure 2.7, shows a representative normalized magnitude-squared T-F surface generated from the complex Gabor coefficients. The surface is divided up into N_R two-dimensional sub-regions, which are also known as patches, in which each sub-region contains a total of $N_T \times N_F$ values. The variables N_T and N_F correspond to length of the sub-region along the time and frequency dimensions, respectively. The value of N_T and N_F are chosen to ensure that a minimum number of $N_{TF} = 120$ values are contained by each sub-region. This criterion was set to facilitate a sufficient number of values are present within a given sub-region to satisfy statistical constraints. The statistics: standard deviation, variance, skewness, and kurtosis, are calculated for each individual sub-region as well as over the entire T-F surface, i.e., the $N_R + 1$ sub-region [8].



$$S_b = \sum_{i=1}^{C} P_i \Sigma_i \tag{2.30}$$

Figure 2.8 representation of a 2-D response generated using Gabor transform and normalized magnitude-squared GT coefficients [8]

2.7 Device Classification

This work uses the same Multiple Discriminant Analysis/Maximum Likelihood (MDA/ML) classification approach for feature selection and device discrimination as presented in [5, 10]. The process begins with MDA, which is an extension of Fisher's Linear Discriminant Analysis from a two-class to the NC-class, where NC is the total number of classes, a.k.a., devices. MDA is used to project the RF-DNA fingerprints into a NC-1 dimensional subspace while maintaining class separability [5]. It reduces feature dimensionality while improving separation between classes by maximizing the between-class separation and minimizing the within-class variance [10]. Figure 2.8 provides a $N_D = 3$ class problem where the ML boundaries associated with each of the Gaussian class likelihood functions are projected into a (NC – 1 = 2)-dimensional subspace. The inter-class S_b and intra-class S_w scatter matrices are given by [44], and,

$$S_w = \sum_{i=1}^{C} P_i(\mu_i - \mu_0)(\mu_i - \mu_0)^T, \tag{2.31}$$

where $\Sigma_i$the covariance matrix and P_i is is the prior probability of class ci. The generated RF-DNA fingerprints are projected into the (NC-1)-dimensional subspace by,

$$f_i^w = W^T f, \tag{2.32}$$

where W is the projection matrix formed from the eigenvectors of $S_w^{-1}S_b$ [5]. The MDA tries to find the best projection matrix that provides the optimal ratio between the inter-class distances and intra-class variances [5]. Figure 2.6 provides an example for two MDA projection matrices where projection matrix W_1 maximizes the inter-class distances and so it provides better class separation performance [5].

During the MDA training process, the training RF-DNA fingerprints generated from each device are projected to NC-1 subspace to comprise the projected training matrix f^w. The mean vector and covariance matrix are estimated for projected training fingerprints of the individual classes. A multi-variate Gaussian distribution is computed, and it’s mean vector is fitted to the projected training samples of each class to develop the reference model as shown by Figure 2.9 [5].

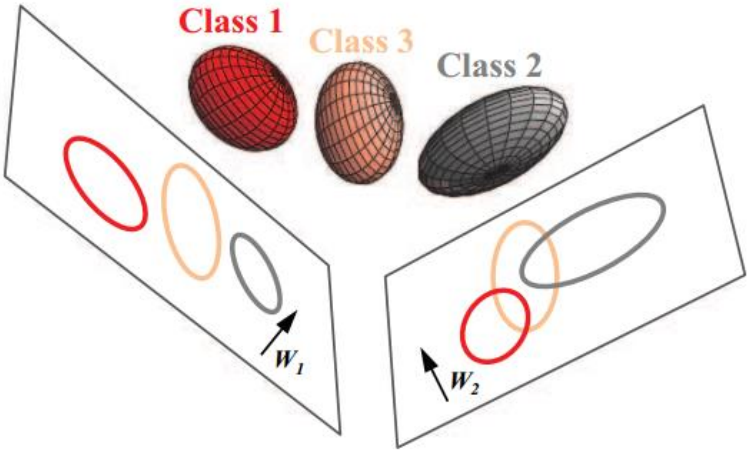


Figure 2.9 MDA projection of fingerprints from NC = 3 class space to 2-D subspaces [22]

The identity of the originating device is determined by comparing its RF-DNA fingerprints to the developed reference models that were fit to the individual projected training sets. The classifier makes decides based on calculation of similarity measure between the unknown RFDNA fingerprint and each of the known reference models. It assigns the unknown fingerprint to the class that results in the best match [5]. As in [5, 10], this work uses the Bayesian posterior probability assuming uniform costs and equal prior probabilities.

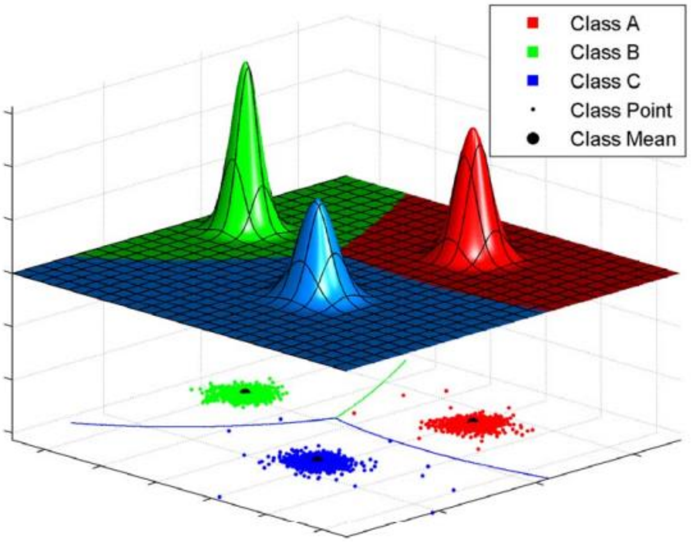


Figure 2.10 MDA/ML Classification subspace for NC = 3; multivariate Gaussian distributions are fitted to projected fingerprints; ML decision boundaries [23]

2.8 Relevant Work

2.8.1 Joint Channel Estimation and Classification

The work in [19] presents an approach that performs wireless device discrimination through the use of RF fingerprints extracted from signals that were transmitted through a wireless multipath channel. This approach was tested on signals transmitted by Universal Mobile Telephone System (UMTS) user equipment, but it can be applied to any system with a known repeated symbol sequence such as the preamble of the IEEE 802.11a MAC frame. It is motivated by an authentication problem that can exist within the femtocell base stations used in cellular wireless networks, like UMTS. This problem is prevalent for cases where the core network signaling, due to the location management, is huge especially when the cell sizes are small. It uses RF fingerprinting to identify the cellular devices at the first contact with the femtocell base station. Authentication at the PHY layer reduces the signaling traffic at higher layers of the cellular core network.

The work in [19] assumes the availability of N-record training database of UMTS preambles $s_n(t)$ originating from different devices. One of the N-record database preambles is input to a multipath fading channel plus noise to generate output $y(t)$, which is the received signal. The original transmitted signal is needed to suppress the effects of the channel. However, the identity of the originating device is unknown prior to classification. In an effort to overcome this challenge, the work in [19] performed joint channel estimation and device classification. The joint channel estimation and device classification was performed using the linear adaptive compensator illustrated by the block diagram in Figure 2.10. The linear adaptive compensator in Figure 2.10 can be summarized by the following equations:

$$n_0 = arg_{n=1,\ldots,N}\, min\, G(n), \tag{2.33}$$

$$G(n) = min_{h_l,\delta_l,w} \sum_{t\in T} |y(t) - \sum_{l=1}^{L} h_l e^{-wn} s_n(t-\delta_l)|^2, \tag{2.34}$$

where $G(n)$ is the residual power at the output of the adaptive compensator for the nth signal candidate, L is the length of the channel, T is the duration of the samples, w, δ_l, and h_l are the frequency offset, delay of the lth tap, and the corresponding tap coefficient. The algorithm first compensates for the frequency offset, then calculates the residual power by subtracting the channel modified version of the candidate signal from the received signal. The nth channel modified signal is the combination of delayed and weighted copies of the nth candidate. The candidate signal that results in the minimum residual power is selected as the closest match to the received signal. The received signal is determined to have originated from the device associated with the selected candidate signal.

2.8.2 Nonlinearity Estimation for Specific Emitter Identification

The work in [20, 21] presents a constellation based method for SEI in an empirical indoor multipath channel model. This method is based on estimating the nonlinearity in the RF front-ends, such as the amplifier, to identify a given radio. Reliable nonlinearity estimation and radio identification requires estimation and removal of the multipath channel to suppress ISI. The algorithm begins with the differentiation of the nonlinearity in each symbol, where the symbols with lower amplitudes are ignored as they are less affected by RF front-end nonlinearity distortion. After initial estimation of the RF front-end nonlinearities, the accuracy of the estimation is improved through the use of an iterative estimation approach. This iterative approach estimates the transmitted symbols and the channel coefficients to achieve an asymptotically unbiased estimation. The algorithm is applicable on multiple amplitude level modulation schemes such as Quadrature Amplitude Modulation (QAM), OFDM, and Pulse Amplitude Modulation (PAM).

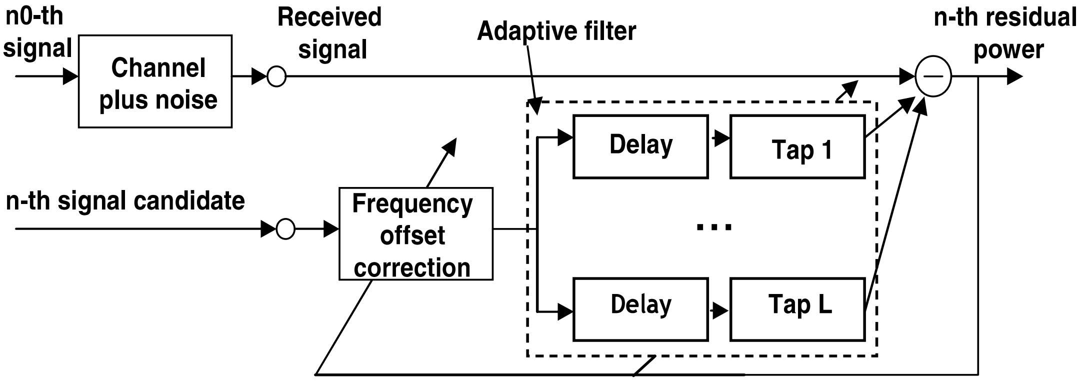


Figure 2.11 Linear adaptive compensator

The work in [20] uses the linear approximation approach to estimate the channel coefficients , This linear approximation based estimate is achieved through the use of a known sequence of waveform symbols, e.g., the preamble of the IEEE 802.11a MAC frame. The estimated channel impulse response is then used to recover the transmitted signal, i.e., the channel input. The estimated transmitted signal can then be used to derive the nonlinearity coefficients. This initial estimate of the channel impulse response, transmitted signal, and the nonlinearity coefficients, can be used to further improve the identification performance through the use of more iterations. The initial estimate of the transmitted signal can be used to re-estimate the channel impulse response, which is used to update the estimation of the transmitted signal and then the nonlinearity coefficients. The estimated nonlinearities can then be compared with the training RF front-end nonlinearity models to identify the transmitting device.

CHAPTER 3
METHODOLOGY

3.1 Introduction

In this chapter, the process developed for assessing RF-DNA fingerprinting performance in wireless multipath channels is described. The methodology presented here uses the same IEEE 802.11a signal set in [10]; thus, the carrier frequency offset value for each signal has been estimated and corrected prior to the methodology presented in the remainder of this chapter. The collected signals are filtered by Rayleigh multipath channels. Following Rayleigh channel filtering, the received signals then undergo time synchronization through estimation and correction of the time offset. The resultant signals are used with the known short and long training symbol sequences to estimate the channel impulse response. The channel impulse response is estimated using the Least Square (LS) and Nelder-Mead (N-M) estimators presented in Chapter 2. The LS estimator provides an accurate estimation of the path delays, while the estimation of the coefficients is further improved by the N-M estimator. A comparison between the LS and the N-M estimator's performance is presented in Chapter 4. The estimated channel impulse response is used to recover the original transmitted preamble, which input into the RF-DNA fingerprinting process. The statistical fingerprints are generated from the recovered preambles using the Gabor Transform (GT). The generated fingerprints are then classified using the MDA/ML classifier presented in Chapter 2. Figure 3.1 shows the processing blocks at each stage of this work.

3.2 Signal Detection and Collection

Figure 3.1 shows the collection and post-collection processes adopted from [6] and adapted for this work. In particular, this work adds the Rayleigh fading channel, time offset correction, channel estimation, and channel equalization blocks. The IEEE 802.11a Wi-Fi signals used in this work are the same as those used to generate the results presented in [6, 10]. These signals were collected from a Cisco AIR-CB21G-A-K9 with a total of $N_D = 4$ Wi-Fi cards in an office environment using Agilent E3238S-based RF Signal Intercept and Collection System (RFSICS). The RFSICS has a tunable frequency range from 0.02 to 6 GHz, and an RF filter with a bandwidth W_{RF} = 36.0 MHz. In the collection process, the captured frequency spectrum is down converted to an intermediate frequency f_{IF} = 70 MHz before going through the analog-to-digital converter (ADC). The ADC used in RFSICS has 12-bit resolution with a sampling rate of f_s = 95 mega-samples-per-second. After analog-to-digital conversion, the digital signal is down converted and filtered with a W_{BB} = 9.28 MHz digital filter and stored as complex In-phase and Quadrature components (i.e., samples). A total of $N_B = 2000$ individual 802.11a signals were extracted from each device's collection record. The amplitude-based variance trajectory technique in [4] was used to detect the individual Wi-Fi transmissions within the overall collection record. The Carrier Frequency Offset (CFO) values were then estimated and removed from each detected signal using the approach in [10]. Following CFO removal, each signal was resampled from 23.75 MHz to 20 MHz to improve the accuracy of the time synchronization process. The time synchronization process, as described in Chapter 2, uses the sampled version of the 16 μs long preamble.

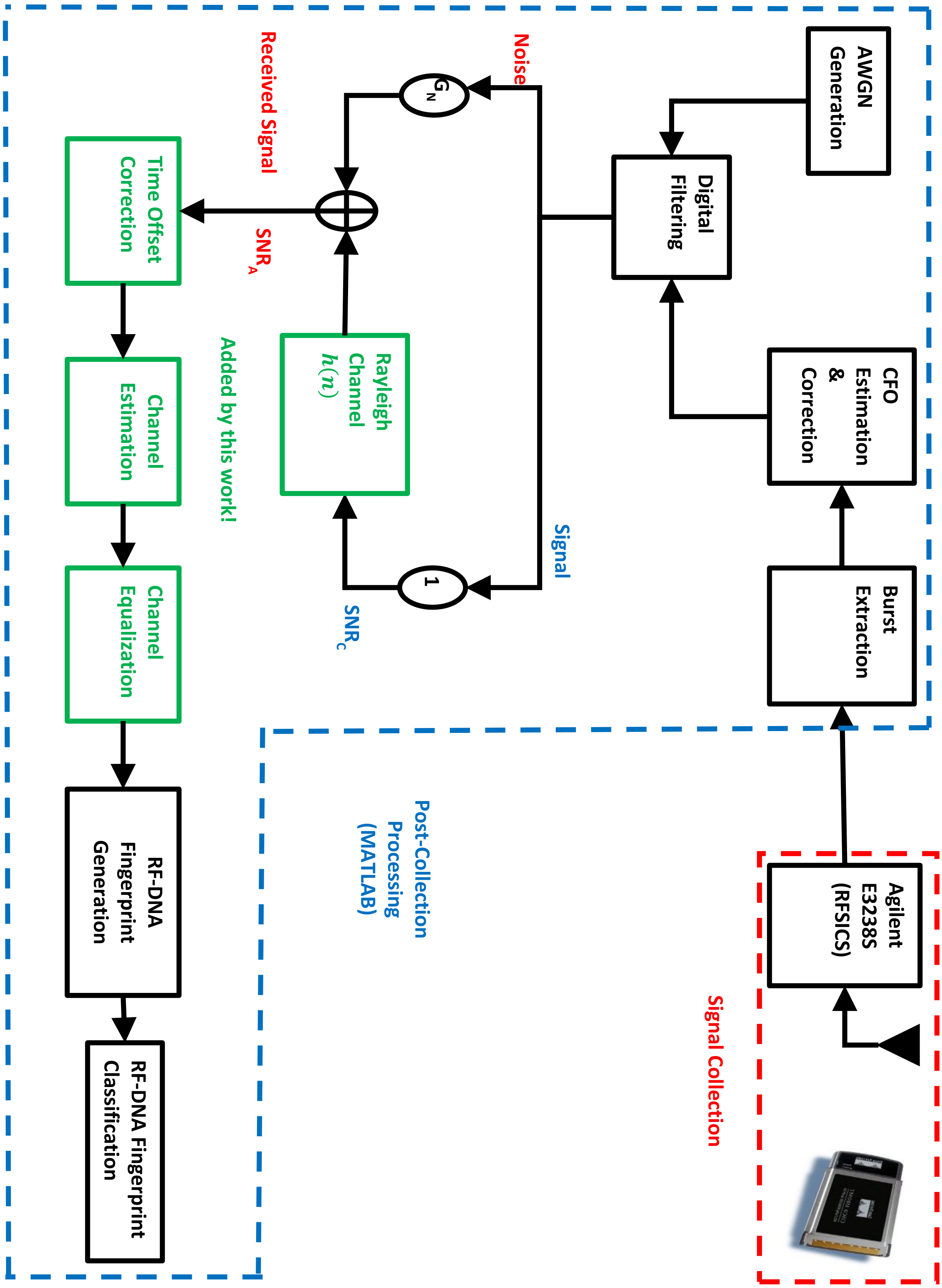


Figure 3.1 Signal collection and post-collection processing [5]

For the 20 MHz sampling rate, the number of samples for every STS and LTS is 16 and 64, respectively. However, when sampled at 23.75 MHz the number of samples per STS and LTS is no longer integer, e.g., STS has 19.001 samples. The algorithm presented in section 2.3 requires the number of samples to be integer for accurate time offset estimation because the offset is estimated by correlating the received signal with itself shifted by 1 or 2 STSs periods (in samples).

3.3 Noisy Multipath Signal Generation

Prior to application of the Rayleigh fading channel and addition of noise, 1,000 signals, from each Wi-Fi device, are randomly selected. This set of 4,000 signals, 1,000 per device, is designated as the “test” set and will be the only signals exposed to Rayleigh fading. The remaining 4,000, the 1,000 not selected for each of the four devices, is designated as the “training” set. The training set is only passed through an Additive White Gaussian Noise (AWGN) channel prior to RF-DNA fingerprint generation; thus, this set of signals does not contain multipath channel effects. The RF-DNA fingerprints generated from the training set of signals is used for model development within the MDA/ML classifier.

Figure 3.2 provides a more detailed representation of the Rayleigh channel block in Figure 3.1. The input signal, x(n), is filtered by the Rayleigh channel, detailed in Section 2.2.3, given by,

$$h(n) = (A_1 + jB_1)\delta(n - \tau_1) + \cdots + (A_L + jB_L)\delta(n - \tau_L), \qquad (3.1)$$

where A_k, and B_k are zero mean independently identically distributed (iid) Gaussian random variables with variance σ_k^2 given by (2.3), (2.4), and (2.5) in section 2.2.3, and τ_k is the delay of the kth path. Each channel coefficient $(A_k + jB_k)$ is generated from a Gaussian random variable with zero mean and unit variance as follows,

$$(A_k + jB_k) = \frac{\sigma_k}{\sqrt{2}}[\mathcal{N}(0,1) + j\mathcal{N}(0,1)], \tag{3.2}$$

where $\mathcal{N}(0,1)$ is a zero mean, unit variance Gaussian random variable. The signal $M(n)$, in Figure 3.2, is generated by the convolution of x(n) with the channel impulse response in (3.1).

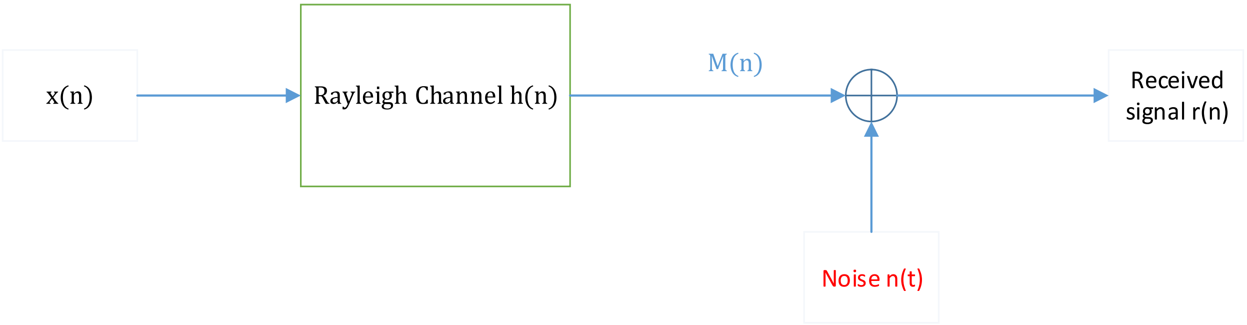


Figure 3.2 Multipath signal generation

The noisy channel output signal, r(n), in Figure 3.2 is generated by adding scaled and like-filtered noise to M(n). Prior to filtering, the added noise is complex with a variance $\sigma_n{}^2$. The variance $\sigma_n{}^2$ is set to generate a received signal r(n) with signal-to-noise (SNR) ranging from 9 dB to 30 dB in 3 dB steps. The like-filtered AWGN noise is simulated by first generating a zero-mean, and unit-variance Gaussian random sequence with length L_n given by,

$$L_n = L_x + L - 1, \tag{3.3}$$

where L_x, is the length of the transmitted signal x(n), and L is the length of the channel. This sequence is filtered using the same parameters that were used to filter the collected signal. The like filtered Gaussian noise is then scaled to achieve analysis $SNR_A^{db} \in [9, 30]$ dB in 3 dB steps. The SNR_A^{db} is given by,

$$SNR_A^{db} = 10 \times log_{10}\left(\frac{S_t}{P_G}\right), \tag{3.4}$$

where S_t is the power of the transmitted signal, and P_G is the power of the scaled, like-filtered noise. The noise scale factor R_n is given by,

$$R_n = \sqrt{10^{\frac{SNR_A^{db}}{10}} \times S_t}. \tag{3.5}$$

The power of the transmitted signal S_t is given by,

$$S_t = \frac{1}{L_x} \sum_{\mathrm{m}=1}^{L_x} \mathrm{x(m)\ x^*(m)}. \tag{3.6}$$

3.4 Time Synchronization

Carrier frequency offset correction and time synchronization must be performed prior to channel estimation and correction. The Wi-Fi signals used in this work are the same as those used in [10]; thus, the carrier frequency offset has been corrected and is neglected here. As explained in Section 2.3, the timing metrics $M_1(\theta)$, and $M_1(\theta)$ are first computed by calculating the normalized autocorrelation of the received signal with itself delayed by one and two STS durations, respectively. For a sampling frequency of 20 MHz, each subcarrier in one OFDM symbol is represented by one sample in time, so one STS and LTS are 16 and 64 samples in length, respectively. The Guard Interval (GI) portion of the preamble will be 32 samples in length.

In effort to improve clarity of the time synchronization process, an example of time synchronization is presented here using an ideal 802.11a preamble, i.e., one void of any device coloration and generated using modeling and simulation software. In this example, a delay of 1,500 ns, i.e., 30 samples, is applied to the ideal preamble. Calculation of the first metric $M_1(\theta)$ is computed using (2.9) and the resulting magnitude shown in Figure 3.3. In Figure 3.3, the magnitude of $M_1(\theta)$ reaches a maximum value of 1 at index 31, which corresponds to the beginning of the first STS of the delayed signal, and maintains this maximum through index 159. Figure 3.4 shows the magnitude of the second metric, $M_2(\theta)$, calculated using (2.10). The magnitude of metric $M_2(\theta)$ reaches it maximum value of 1 at the same index as that of $M_1(\theta)$, but only maintains this maximum magnitude value through index 143.

Estimation of the time offset within the received signal is determined through calculation of (11) in Section 2.3, which is the magnitude of the difference between the two metrics $M_1(\theta)$, and $M_1(\theta)$. For this example, the result of (11) is a single peak corresponding to index 159 as illustrated in Figure 3.5. Thus, the estimated time offset $\hat{\theta}$ would be 159, which corresponds to the start of the ninth STS. The estimated offset $\hat{\theta}$ is essential to the estimation of the channel impulse response, because it provides a point of reference that is used to locate the start and end of the LTSs, which are used by the LS and N-M channel estimators. The process detailed in this section is applied to all the received signals generated in Section 3.3.

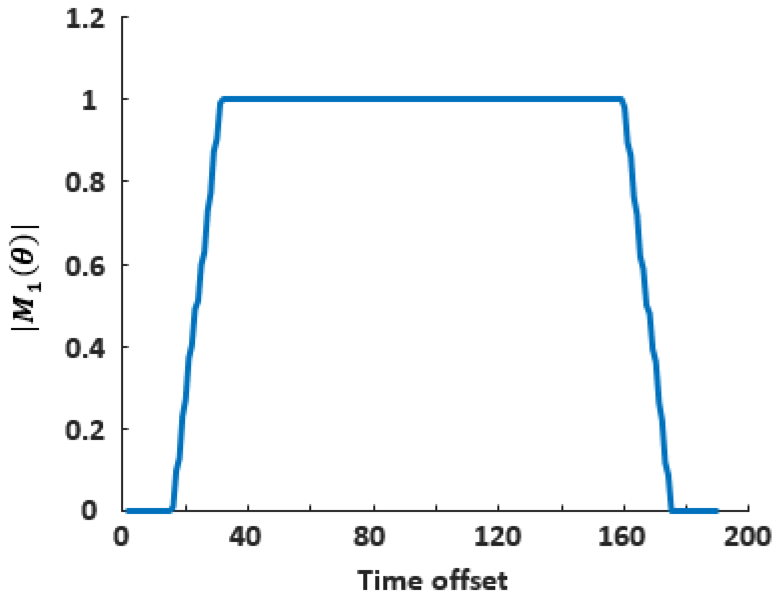


Figure 3.3 Normalized timing Metric $M_1(\theta)$ for a delay of 30 samples

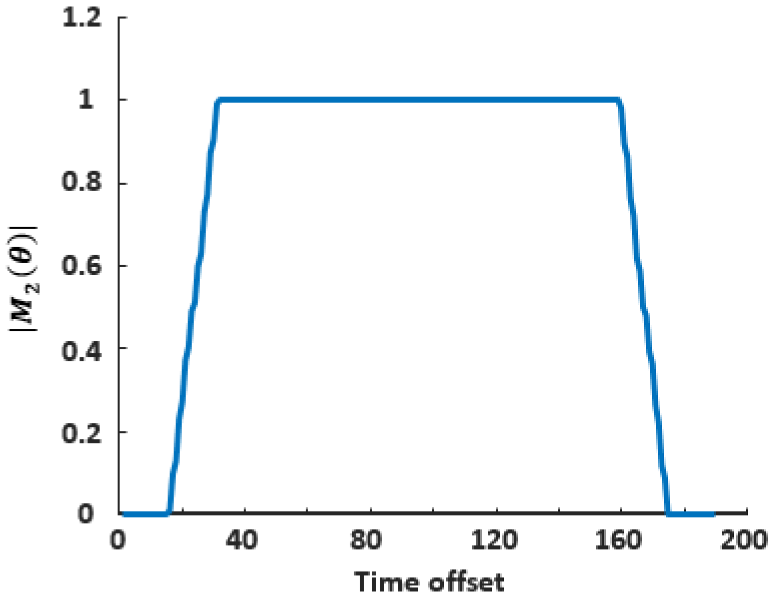


Figure 3.4 Normalized timing Metric $M_2(\theta)$ for a delay of 30 samples

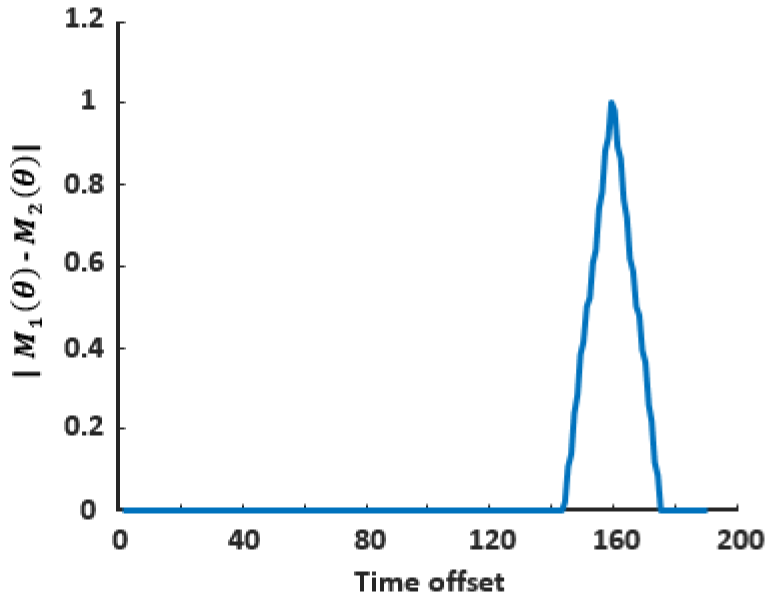


Figure 3.5 Metric difference | $M_1(\theta)$ - $M_2(\theta)$| as a function of offset θ

3.5 Least Square Estimator

As explained in Section 2.5.1, the LS estimator is used to estimate the channel impulse response using (2.18). The resulting estimate serves as the first estimate of the channel's frequency response. A key assumption is that the overall length of the channel is less than that of the GI, i.e., $L < 32$ samples when $f_s = 20$ MHz [45]. This assumption is applied in both the LS and N-M estimators. The remainder of this section explains the solution to (2.18) and determination of a course estimate of the channel's impulse response. In (2.18), X is the Discrete Fourier Transform (DFT) of the ideal long training symbol and is comprised of 52 nonzero elements corresponding to the data subcarriers as given by:

$$X(1:26) = \{1,-1,-1,1,1,-1,1,-1,1,-1,-1,-1,\\ -1,-1,1,1,-1,-1,1,-1,1,-1,1,1,1,1\},$$

$$X(38:63) = \{1,1,-1,-1,1,1,-1,1,-1,1,1,1,1,1,1,\\ -1,-1,1,1,-1,1,-1,1,1,1,1\},$$

where $X(k) = 0$ for $26 < k < 38$ corresponds to the virtual subcarriers [42]. In the discrete time domain, let $P(n)$ be the portion of the IEEE 802.11a ideal preamble that contains the GI and the two LTSs as shown in Figure 3.6. Let the received signal portion corresponding to $P(n)$ be designated as $C(n) = P(n) * h(n)$, which is the convolution of the length L channel with that of $P(n)$. Due to the channel $h(n)$, the first $L - 1$ elements of $C(n)$ correspond to $P(n)$ and the ten STS of the transmitted preamble. The next $5N/2 - L + 1$ elements only depend upon the channel and $P(n)$ [45]; thus, the focus is placed on this portion of the received signal to facilitate estimation of the channel's impulse response. These elements are designated as,

$$v(n) = C(n + L - 1), \qquad n = 0, \dots, \frac{5N}{2} - L.$$

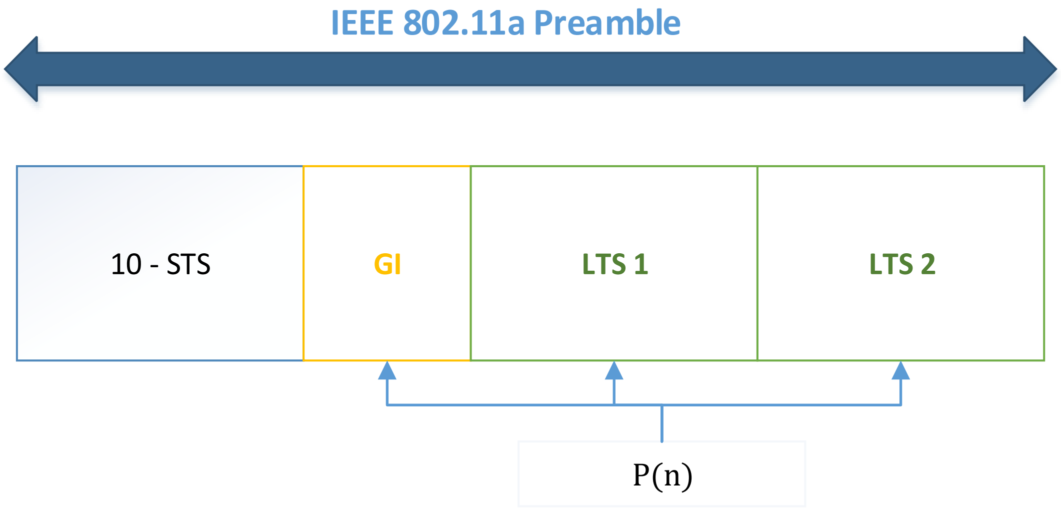


Figure 3.6 $P(n)$ signal portion of the IEEE 802.11a Preamble

After determination of the time offset $\hat{\theta}$, as described in Section 3.4, the first index $n_{v,1}$ of $v(n)$ is determined to be,

$$n_{v,1} = \hat{\theta} + 2 \text{ x } L_{STS} - 1 + L - 1, \tag{3.7}$$

and its last index $n_{v,2}$ is given by,

$$n_{v,2} = \hat{\theta} + 2 \text{ x } L_{STS} - 1 + L_{GI} + 2 \text{ x } L_{LTS}, \tag{3.8}$$

where L_{STS} , L_{GI}, and L_{LTS} are the number of samples associated with one STS, the GI, and one LTS ,respectively.

Following the extraction of $v(n)$, the first estimate of the channel's frequency response $\widehat{\mathrm{H}}_{\mathrm{ls}}$can be calculated using (2.17) and (2.18). In (2.18), Y_2 is the DFT of the last N elements of $v(n)$ and Y_1 corresponds to the N elements preceding Y_2. For the case of (2.17), Y can be either Y_1 or Y_2, because both are the result of filtering the same LTS by the channel $h(n)$. For the results presented in Chapter 4, the channel's frequency response is estimated using (2.18). The use of two LTSs reduces the variance of the noise, which results in the reduction of the square error by one half when compared to the use of only one LTS as in (2.17) [42]. A preliminary normalized estimate of the channel impulse response is calculated by,

$$\bar{h}(k)_{pre} = \frac{h(k)_{pre}}{max\left|h(k)_{pre}\right|} \tag{3.9}$$

where $h(k)_{pre}$ is given by,

$$h(k)_{pre} = IDFT\{\widehat{H}_{ls}\}. \tag{3.10}$$

Using the estimate given by equation (3.9), each multipath coefficient's position which represents the delay associated with that component can be determined by applying a threshold rule given by,

$$\hat{h}(k) = \begin{cases} \bar{h}(k)_{pre} & when\ \left|\bar{h}(k)_{pre}\right| > \Gamma \\ 0 & otherwise, \end{cases} \tag{3.11}$$

where Γ is a threshold that satisfies $\Gamma < 1$. $\hat{h}(k)$ Provides a coarse estimate for the channel coefficients, which can be further improved by feeding delays associated with the channel coefficients to the N-M estimator as will be explained in the next section.

3.6 Nelder-Mead Estimator

This work presents a novel application of the Nelder-Mead (N-M) direct search algorithm, described in Section 2.5.2, in the development of a multipath channel estimator. This application is designated here as the N-M estimator and is used to provide a fine estimate of the channel coefficients prior to equalization. As previously stated, the N-M algorithm is applied in the minimization of the square error function $f(h)$ given by (29). One restriction of the N-M algorithm is that it solves only real-valued functions; however, the function $f(h)$ to be minimized is complex [36, 37]. This is due to the fact that both the transmitted signal and channel coefficients are complex. In an effort to overcome the N-M algorithm's limitation, the function $f(h)$ is expanded into real and imaginary parts as follows:

$$C_1 = \sum_{\mathrm{m} \in T} \left| Re\{r(m)\} - \left(\sum_{k=1}^{L} h_{r,1}(k) \times Re\{x(m-\tau_k)\}\right) - h_{i,1}(k) \times Im\{x(m-\tau_k)\}\}\right) \right|^2 \qquad (3.12)$$

$$C_2 = \sum_{\mathrm{m} \in T} \left| Im\{r(m)\} - \left(\sum_{k=1}^{L} h_{r,2}(k) \times Im\{x(m-\tau_k)\}\right) + h_{i,2}(k) \times Re\{x(m-\tau_k)\}\}\right) \right|^2, \qquad (3.13)$$

where C_1 and C_2 are the real and imaginary parts of the function $f(h)$, respectively, $h_{r,1}(k)$ and $h_{r,2}(k)$ are the real parts of the kth estimated channel coefficient, $h_{i,1}(k)$ and $h_{i,2}(k)$ are the imaginary parts of the kth estimated channel coefficient, $Re\{\}$ is the real part of a function, $Im\{\}$ is the imaginary part of the function, and τ_k is the path delay of the kth coefficient. The path delay

τ_k represents the delay of the kth non-zero coefficient obtained through the use of the LS estimator as described in Section 3.5.

Splitting $f(h)$ into its real and imaginary components allows for estimation of the real and imaginary parts of the channel impulse response separately and without violating the real-valued function limitation of the N-M algorithm. In an effort to minimize the estimation error, the coefficients estimated by (2.12) and (2.13) are averaged together to provide the final estimate of the real and imaginary components of the channel, respectively. The real and imaginary components for the final estimated channel are given by,

$$h_r(n) = \begin{cases} \dfrac{h_{r,1}(n) + h_{r,2}(n)}{2} & , \quad when\ n \in [\tau_1, \tau_L] \\ 0 & , \quad otherwise \end{cases}, \tag{3.14}$$

$$h_i(n) = \begin{cases} \dfrac{h_{i,1}(n) + h_{i,2}(n)}{2} & , \quad when\ n \in [\tau_1, \tau_L] \\ 0 & , \quad otherwise \end{cases}, \tag{3.15}$$

where $h_r(n)$ is the real part, and $h_i(n)$ is the imaginary part of the estimated channel impulse response. The estimated channel impulse response can then be obtained by combining the real and imaginary parts as follows,

$$h(n) = h_r(n) + jh_i(n). \tag{3.16}$$

Depending on the transmitted signal $x(n)$, two different cases are considered in the estimation of the channel. In the first case, $x(n)$ is an ideal preamble, while the received signal is constructed using the collected signals, from the “test” set, convolved with the generated channel impulse response followed by addition of appropriately scaled noise to achieve the desired SNR. Thus, for this case, the coefficients used in construction of the equalization filter are based upon a comparison of a received signal in which device specific waveform coloration is present to that of a transmitted signal that is devoid of such coloration. In the second case, a collected “candidate” preamble is used as $x(n)$ instead of the ideal preamble. In this work, a total of five candidate preambles were randomly selected from the training set of each of the $N_D = 4$ devices; thus, a total of $N_P = 20$ candidate preambles are used in obtaining the fine estimate of the channel coefficients. For every received signal, there is one set of channel coefficients corresponding to each of the N_P candidate preambles for which (2.27) is minimized.

The “best” estimate of the channel is selected based upon the candidate preamble that results in the smallest residual power given by,

$$\hat{h}(m) = \text{argmin}_{\ c} \left\{ \sum_m \left| r(m) - \hat{h}_c(m) * x_c(m) \right|^2 \right\}, \tag{3.17}$$

where $1 \leq c \leq N_P$, and $\hat{h}_c(m)$ is the estimated channel associated with candidate preamble $x_c(m)$.

3.7 Channel Equalization

Removal of the channel effects from the received signal is achieved through the use of an equalization filter. The frequency response of the equalization filter is generated from the channel estimate obtained from the N-M estimator and is given by,

$$G(k) = \frac{1}{\widehat{H}(k)}, \tag{3.18}$$

where $\widehat{H}(k)$ is an L_m point DFT of the "best" channel estimate given by (48), $L_m = L_x + L - 1$ is the length of the received signal, L_x is the length of the transmitted signal, and L is the length of the channel. The equalization filter $G(k)$ is used to remove the channel as follows:

$$\hat{x}(m) = \frac{1}{N_{DFT}} \sum_{k=0}^{N_{DFT}-1} R(k).G(k)e^{\frac{j2\pi km}{N_{DFT}}}, \tag{3.19}$$

where $R(k)$ is the DFT of the received signal, and $0 \leq m < L_m$. The channel coefficients estimated by the two cases, presented in Section 3.6, are used in generation of the equalization filter. This results in two sets of channel corrected preambles from which RF-DNA fingerprints are generated prior to MDA/ML classification. Using these two sets of RF-DNA fingerprints, comparative assessment and analysis of percent correct classification performance is presented in Chapter 4.

3.8 RF-DNA Fingerprint Generation

The equalized signal $\hat{x}(m)$, given by (3.19), serves as the region of interest from which RF-DNA fingerprints are generated. The RF-DNA fingerprints can be extracted based on amplitude, phase, and/or frequency characteristics [12]. In this work the RF-DNA fingerprints are generated using the same approach presented in [12], and briefly explained in Section 2.6. This approach leverages the Discrete Gabor Transform (DGT) to jointly capture the momentary T-F variations that occur within a signal [6]. The DGT was calculated using (2.28), a Gaussian synthesis window $W(n)$, and the variables defined in Section 2.6 [6]. The normalized magnitude response of the resulting 2-D T-F plane is calculated using (2.29) and subsequently divided into N_R patches. Each patch is associated with a total of $N_T \times N_F$ Gabor coefficients and is reshaped into a N_{TF} length vector. The statistics: standard deviation (σ), variance (σ^2), skewness (γ), and kurtosis (□) are calculated from this vector and used to form the RF-DNA fingerprint corresponding to $\hat{x}(m)$. For the classification results presented in Chapter 4, all RF-DNA fingerprints are generated using: $M = 186$, $K_G = 186$, $N_\Delta = 1$, $N_{TF} = 120$, $N_T = 12$, and $N_F = 10$, at SNR $\in$ [9,30] dB in 3 dB steps.

3.9 Device Classification

The RF-DNA fingerprints, associated with the “training” set and generated in accordance with Section 3.8, serve as the input to the MDA/ML classifier as described in Section 2.7. MDA serves as the feature selection process by projecting the “training” RF-DNA fingerprints for each of the $N_D = 4$ devices into a $(N_D - 1)$ = 3-dimesional subspace. This subspace is associated with the projection that results in the maximum distances between classes and the minimum within class spread, a.k.a., variance [12].

The MDA process is followed by ML classification, which uses the Bayesian Decision rule to assign each of the projected RF-DNA fingerprints to one of the N_D classes. Each class assignment is based upon the reference model that returns to the largest likelihood value [10]. The MDA/ML classification performance results are presented in Chapter 4 and are measured using percent correct classification. Percent correct classification is calculated through the tracking of the number of times the classifier correctly assigns an RF-DNA fingerprint to its class over all trials [10].

CHAPTER 4

RESULTS AND DISCUSSION

This chapter presents the results and analysis for: 1) multipath channel estimation using the LS (Sect. 2.5.1) and N-M (Sect. 2.5.2) estimators, and 2) IEEE 802.11a Wi-Fi device classification performance associated with the MDA/ML classifier, Section 2.2.3. Device classification performance is presented for four devices and multipath channel lengths of: L = 2 and L = 5 reflectors. The multipath channel was implemented as a TDL in which each tap is characterized by a variance $\sigma_k^2/2$ and a delay spread τ_k. The LS and N-M based estimator performance is presented in Section 4.1.

The RF-DNA fingerprints were generated from the 2-D, joint T-F signal responses generated using the Discrete Fourier Transform (DGT) described in Section 2.6. The RF-DNA fingerprints were generated from $N_B = 2{,}000$ collected signals for each device. The device classification results were generated based on four Wi-Fi devices from the same manufacturer, but with different serial numbers, which represent serial number discrimination [5]. Serial number discrimination represents the most challenging SEI case. In this work the RF-DNA fingerprints were divided into two sets:

1. *Training Set of Wi-Fi Signals*: This set is used in the training of the MDA/ML classifier in the development of the reference model. The RF-DNA fingerprints comprising each device's training set, were generated using 1,000 signals that were randomly drawn from

the collected set of $N_B = 2{,}000$ signals. The training set of signals was subjected to an AWGN channel only, i.e., they are not used for multipath channel estimation.

2. *<u>Test Set of Wi-Fi Signals:</u>* The test set is generated from the remaining 1,000 collected signals that were not selected for inclusion within training set. Unlike the training set signals, these signals were subjected to Rayleigh fading channel simulation in addition to the AWGN channel. The RF-DNA fingerprints associated with this set of signals serve as the "blind" test of the MDA/ML classifier model developed using the training set's RF-DNA fingerprints [5].

Table 4.1 Reflector parameters for L = 2 Rayleigh channel

	K = 1	**K = 2**
Variance $\frac{\boldsymbol{\sigma_k^2}}{\mathbf{2}}$	0.8	0.2
uDelay spread $\boldsymbol{\tau_k}$	50 ns	200 ns

4.1 LS and N-M Estimator

In this study, the LS and N-M channel estimators were used to estimate the channel coefficients, which were then stored to facilitate a comparative assessment via squared error. For the results in this section, a unique L = 2 Rayleigh fading channel, with variances and associated delay spreads given in Table 4.1, is generated using (3.1) and (3.2) and applied to each signal within the test set for each of the four devices. The selection of σ_k was limited such that $\sum \sigma_k^2 = 1$. This limitation ensures that the average received power remained the same across all the signals at each SNR.

Following generation and application of the Rayleigh fading channel, each of the test set signals has scaled, like-filtered AWGN added to it for SNR ∊ [0, 30] dB in 3 dB steps and ten noise realizations. Afterward, channel estimation is conducted using both the LS and N-M estimators. The result is a set of estimated coefficients for each of the signals associated with each device and every SNR, i.e., there is a total of 88,0000 channel estimates performed by each estimator. The performance of each estimator is assessed using square error given by,

$$\epsilon = \sum_{n \in L} \left| h(m) - \hat{h}(m) \right|^2, \tag{4.1}$$

where $h(m)$ is the true channel and $\hat{h}(m)$ is the estimated channel. Figure 4.1 shows the average of the square error ϵ given by equation (51) across all the $N_B = 2000$ bursts and all the $N_D = 4$ devices at each SNR. For SNR ≥ 6 dB, the N-M based estimator results in the lowest square error in estimation of the channel coefficients; however, the performance of the LS estimator equals or outperforms the N-M estimator for SNR ≤ 3 dB. Due to the performance of the N-M estimator for SNR ≥ 6 dB, it is selected for use in the estimation of the Rayleigh fading channel coefficients for all of the results presented in Section 4.2.

4.2 Device Classification

The MDA/ML classifier uses a “best” match criterion to make the final device/class assignment. The classifier compares an “unknown”, i.e., one not seen during training, fingerprint and computes the likelihood values for each of the normal distribution functions that were “fit” to the projected, training RF-DNA fingerprints. The unknown fingerprint is then assigned to the device/class associated with the function that returns the largest likelihood value.

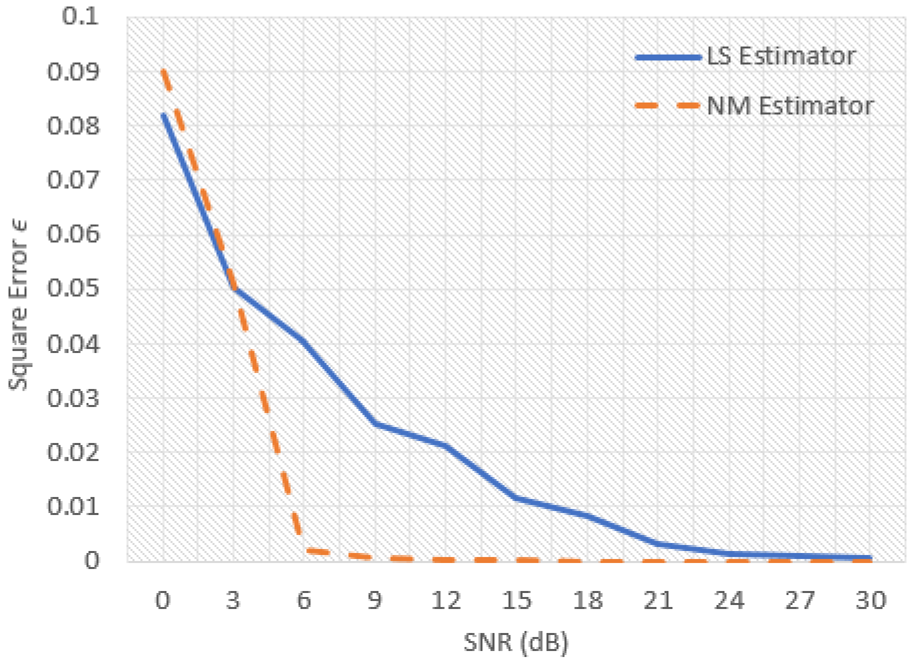


Figure 4.1 The average square error of the LS and NM estimators for SNR∈ [0, 30] dB.

The reliability of the developed model has been assured through the use of: 1) Monte Carlo simulation in which $N_z = 10$ independent, like-filtered AWGN noise realizations are generated for each SNR ∊ [9,30] dB in 3 dB steps, and 2) k-fold cross validation with k = 5. k-fold cross validation is performed by randomly assigning an equal number of training RF-DNA fingerprints to one of the five subsets. In this work, each subset is comprised of 200 RF-DNA fingerprints for each device. The implementation of k-fold cross validation was done at every noise realization and SNR; thus, for a given SNR and noise realization the random assignments were made and remained unchanged during the development and validation each “folds” corresponding model. In k-fold validation four of the five subsets, e.g., 1, 3, 4, and 5, are used in developing the classifier model and the fifth subset “held out” for model validation. The process is then repeated five additional times with a new and previously unselected subset being “held out”. This ensures that every RF-DNA fingerprint, within the training set, was used for model validation at least once. At each SNR, the average percent classification error is calculated and tracked across all folds and noise realizations. The model that resulted in the smallest average percent classification error is designated as the “best” model and used for “blind” classification of the test set of RF-DNA fingerprints. The term “blind” refers to RF-DNA fingerprints that have never been used by the classifier for model development nor validation. RF-DNA fingerprint classification performance results are presented for L = 2 (Sect. 4.2.1) and L = 5 (Sect. 4.2.2) length Rayleigh fading channels used to corrupt the test set of Wi-Fi signals. All RF-DNA fingerprints are generated using the DGT given by (2.28) and (2.29) using parameters: $M = 186$, $K_G = 186$, $N_\Delta = 1$, $N_T = 12$, and $N_F = 10$.

4.2.1 Two-Reflector Channel

For the results presented here, a given SNR represents a total of $N_D \times 1{,}000 = 4{,}000$ unique Rayleigh fading channels [23]. The N-M estimator was used to estimate the channel impulse response, which is then used for channel equalization as described in Section 3.7. In this section, MDA/ML classification performance is presented for the two N-M based channel estimation approaches presented in Section 3.6. These two cases are: 1) an ideal preamble and 2) a set of candidate preambles.

Figure 4.2 shows the percent correct classification performance results when N-M channel estimator is used to estimate the coefficients for the L = 2 reflector Rayleigh fading channel. The percent correct classification results associated with the N-M estimator’s use of the ideal preamble is shown in Figure 4.2(a). In Figure 4.2 (a), Device #1, #2, and #4 show percent correct classification performance of more than 90% for SNR ≥ 24, with device #4 achieving 100% at SNR ≥ 30 dB. At an SNR = 18 dB, Device #3 has the worst percent correct classification performance of the 4 devices at 69%. The remaining three devices are classified correctly more than 83% at the same SNR. For SNRs below 15 dB, the percent correct classification performance was dramatically degraded for Device #1 and Device #4. When the candidate preambles are used in channel estimation, the percent correct classification performance is greater than 95% for all four devices at SNR ≥ 21, Figure 4.2(b). The percent correct classification performance of Device #4 remains above 93% for SNR ≥ 15 dB. It is important to note that the percent correct classification associated with the validation of the developed classifier model is included in Figure 4.2(a) and Figure 4.2(b). These results represent classification of RF-DNA fingerprints generated from signals that have only an AWGN channel applied to them.

Figure 4.2(c) shows the overlay of the percent correct classification for the ideal and candidate preamble cases. It shows that using candidate preambles improves the device classification performance for all SNR ∈ [9, 30] dB. When candidate preambles were used, the average percent correct classification was greater than 95% at SNR ≥ 21 dB. The ideal preamble case achieved this same performance at SNR ≥ 27 dB; therefore, using candidate preambles provides a 6 dB improvement over that of the ideal preamble case.

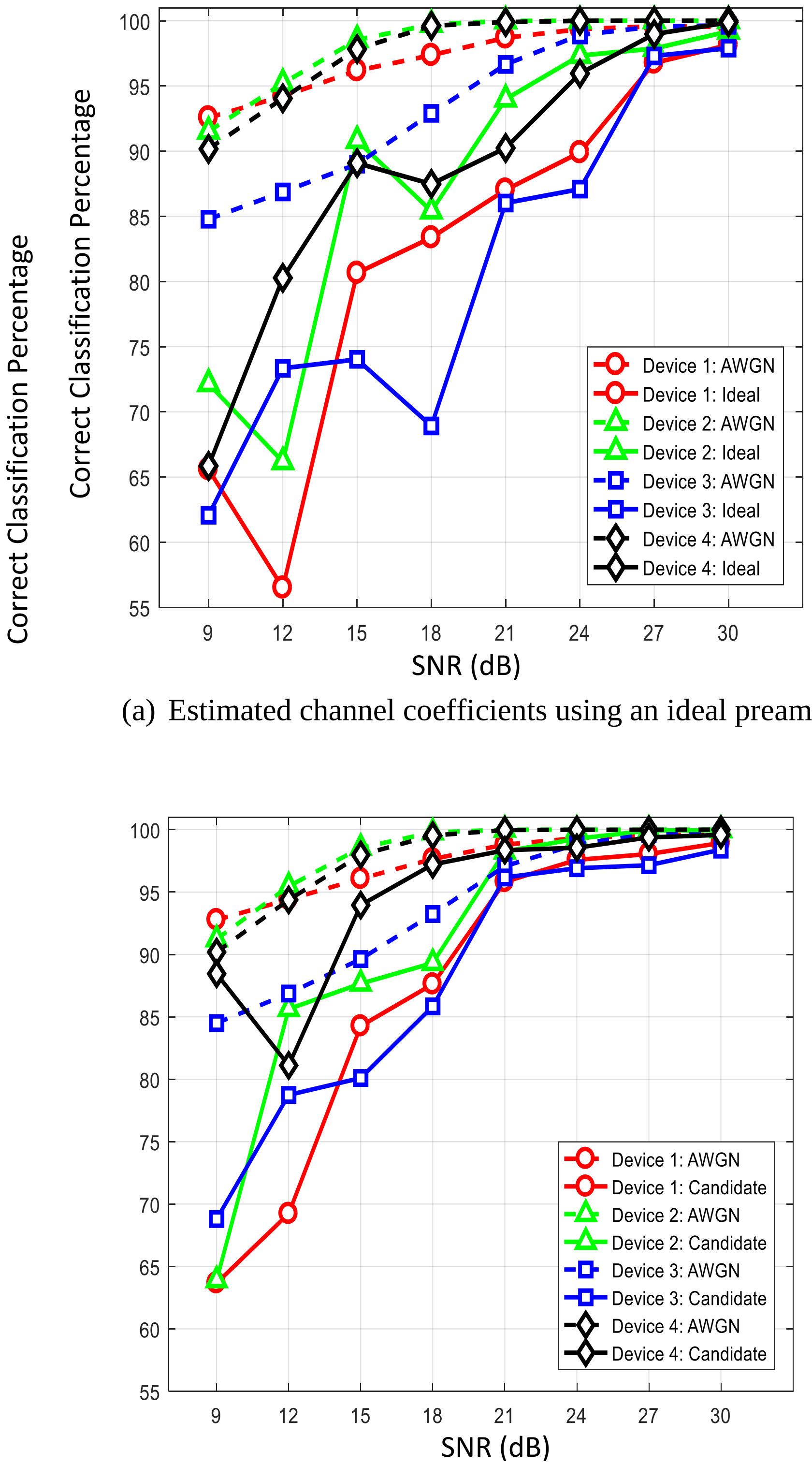
4.2.2 Five-Reflector Channel

This section presents the classification results for the L = 5 reflector fading channel in which the reflected path variances and delay spreads are given in Table 4.2. The results were generated using the same process used in the L = 2 case using candidate preambles by the N-M based estimator.

Table 4.2 Reflector Parameters for L = 5 Rayleigh Channel

	K = 1	K = 2	K = 3	K = 4	K = 5
Variance $\frac{\boldsymbol{\sigma_k^2}}{\mathbf{2}}$	0.865	0.117	0.016	0.002	0.0003
Delay spread $\boldsymbol{\tau_k}$	50 ns	100 ns	150 ns	200 ns	250 ns

Figure 4.2.2 shows that a correct classification percentage of greater than 90% was achieved for all of the 4 devices, for SNR ≥ 27 dB. Device 4 kept a correct classification of greater than 90% for SNR ≥ 21 dB. The overall classification performance was degraded when the number of multipath reflectors was increased from two to five.



(a) Estimated channel coefficients using an ideal preamble.

(b) Estimated channel coefficients using candidate preambles.

Figure 4.2 Two teflectors: RF-DNA fingerprint percent correct classification performance using ideal and candidate preambles

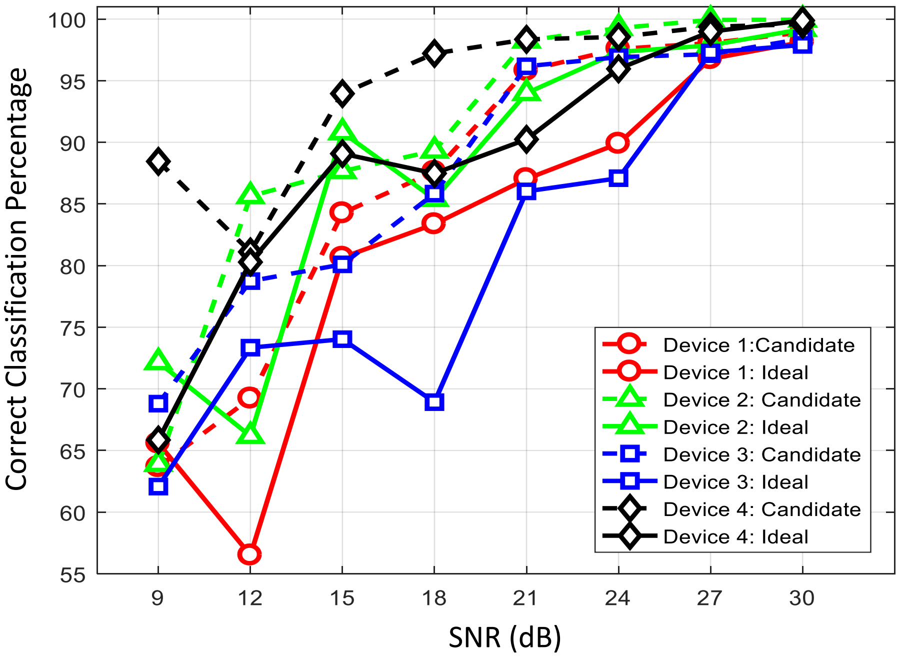


Figure 4.3 Two reflectors: ideal versus candidate preamble

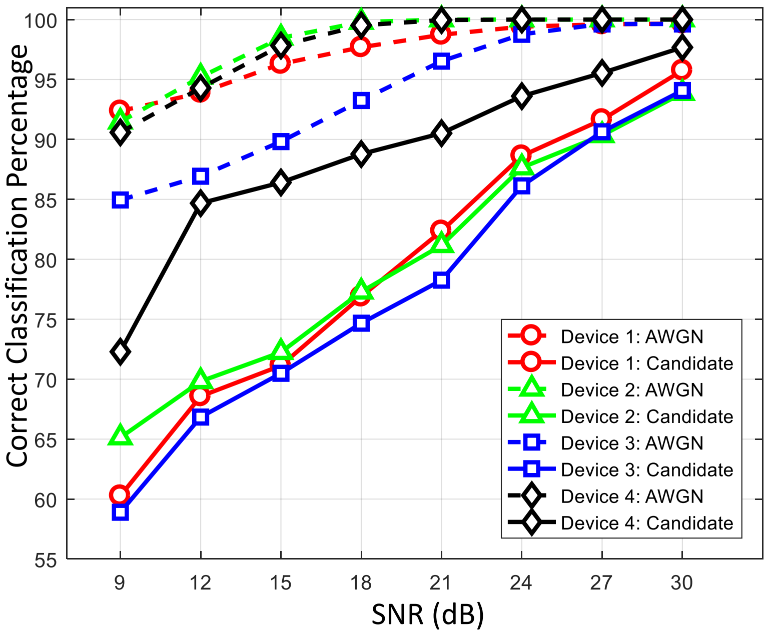


Figure 4.4 Five reflectors: IEEE 802.11a Wi-Fi RF-DNA fingerprint percent correct classification performance

CHAPTER 5

CONCLUSION

This work presents the first investigation of RF-DNA fingerprint classification performance of four IEEE 802.11a Wi-Fi devices using collected signals under an indoor multipath environment. The indoor multipath environment is simulated using a Rayleigh fading channel and degrading SNR. Additionally, an assessment of two multipath channel estimation approaches, Least Square (LS) and Nelder-Mead (N-M), is presented. This is the first known application of the N-M search algorithm in the estimation of multipath channel coefficients.

When compared with the LS estimator, the N-M based estimator resulted in superior channel coefficient estimation performance for SNR $\geq$ 6 dB. This performance was assessed through the calculation of square error. Based upon these results, the N-M based estimator was chosen as the channel estimation process of choice prior to RF-DNA fingerprint classification and MDA/ML classification. The classification performance results were generated from signals that had been convolved with a Rayleigh fading channel consisting of either two or five reflectors. This work investigated channel estimation in which the N-M estimator used either an ideal preamble or a set of candidate preambles to determine the channel coefficients. The use of candidate preambles corresponded with superior percent correct classification performance for a two reflector channel.

For this case, the percent correct classification of 95% was achieved for all four Wi-Fi devices for SNR ≥ 21 dB. Percent correct classification for the five reflector case was poorer than that of the two reflector case and that was expected because the coefficients estimation errors combine and affect the final recovered signal, so with five reflectors, the channel estimation error increases, and that affects the classification result. Percent correct classification was greater than 90% for all of the 4 devices at SNR ≥ 27.

5.1 Future Work

A better "apples-to-apples" comparison of the N-M approach's performance is to use the signals that have gone through the channel estimation, and equalization in constructing the "training" set used in MDA/ML classifier model development.

Employ an alternative equalization technique better suited to degrading SNR. One such example is the Minimum Mean Square Error (MMSE) technique. MMSE is designed for demodulated data, but the same technique can be modified to perform equalization at the waveform i.e., PHY level. MMSE is capable of accounting for environmental statistics such as noise power, and the signal's power spectral density. Thus, it may provide better performance as SNR degrades.

REFERENCES

[1] W. E. Cobb, E. W. Garcia, M. A. Temple, R. O. Baldwin, and Y. C. Kim, "Physical layer identification of embedded devices using RF-DNA fingerprinting," in *2010 - MILCOM 2010 Military Communications Conference*, 2010, pp. 2168-2173.

[2] C. K. Dubendorfer, B. W. Ramsey, and M. A. Temple, "An RF-DNA verification process for ZigBee networks," in *MILCOM 2012 - 2012 IEEE Military Communications Conference*, 2012, pp. 1-6.

[3] P. K. Harmer, D. R. Reising, and M. A. Temple, "Classifier selection for physical layer security augmentation in Cognitive Radio networks," in *2013 IEEE International Conference on Communications (ICC)*, 2013, pp. 2846-2851.

[4] R. Klein, M. Temple, M. Mendenhall, and D. Reising, "Sensitivity Analysis of Burst Detection and RF Fingerprinting Classification," in *IEEE International Conference on Communications (ICC09)*, Dresden, Germany, 2009, pp. 1-5.

[5] D. R. Reising, "Exploitation of RF-DNA for Device Classification and Verification Using GRLVQI Processing," Air Force Institute of Technology, 2012. Available: http://www.dtic.mil/dtic/tr/fulltext/u2/a572506.pdf

[6] D. R. Reising, M. A. Temple, and J. A. Jackson, "Authorized and Rogue Device Discrimination Using Dimensionally Reduced RF-DNA Fingerprints," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 6, pp. 1180-1192, 2015.

[7] D. R. Reising, M. A. Temple, and M. J. Mendenhall, "Improving Intra-Cellular Security Using Air Monitoring with RF Fingerprints," in *2010 IEEE Wireless Communication and Networking Conference*, 2010, pp. 1-6.

[8] D. R. Reising, M. A. Temple, and M. E. Oxley, "Gabor-based RF-DNA fingerprinting for classifying 802.16e WiMAX Mobile Subscribers," in *2012 International Conference on Computing, Networking and Communications (ICNC)*, 2012, pp. 7-13.

[9] W. Suski, M. Temple, M. Mendenhall, and R. Mills, "Radio frequency fingerprinting commercial communication devices to enhance electronic security," *International Journal of Electronic Security and Digital Forensics* vol. 1, no. 3, 2008, pp. 301-322.

[10] C. G. Wheeler and D. R. Reising, "Assessment of the impact of CFO on RF-DNA fingerprint classification performance," in *2017 International Conference on Computing, Networking and Communications (ICNC)*, 2017, pp. 110-114.

[11] M. D. Williams, S. A. Munns, M. A. Temple, and M. J. Mendenhall, "RF-DNA Fingerprinting for Airport WiMax Communications Security," in *2010 Fourth International Conference on Network and System Security*, 2010, pp. 32-39.

[12] M. D. Williams, M. A. Temple, and D. R. Reising, "Augmenting Bit-Level Network Security Using Physical Layer RF-DNA Fingerprinting," in *2010 IEEE Global Telecommunications Conference GLOBECOM 2010*, 2010, pp. 1-6.

[13] M. Jakobsen, "Modeling and Estimation of Wireless Multipath Channels - An Application within Pilot-assisted Channel Estimation for Downlink OFDM," MSc, Aalborg University, Denmark, 2009.

[14] D. Evans, "The Internet of Things: How the Next Evolution of the Internet is Changing Everything," vol. 1, Jan 2011, 1-11.

[15] G. Research, "Gartner Says 6.4 Billion Connected 'Things' Will Be in Use in 2016, Up 30 Percent from 2015," Nov 2015.

[16] S. Smith, "Internet of Things Connected Devices to Tripple by 2021, Reaching Over 46 Billion Units," Juniper Research, 2016. Available from: https://www.juniperresearch.com/press/press-releases/%E2%80%98internet-of-things%E2%80%99-connected-devices-to-triple-b

[17] W. Trappe, R. Howard, and R. Moore, "Low-Energy Security: Limits and Opportunities in the Internet of Things," presented at the IEEE Security Privacy, Jan 2015. Available from: https://ieeexplore.ieee.org/document/7031838/

[18] W. C. S. II, M. A. Temple, M. J. Mendenhall, and R. F. Mills, "Radio frequency fingerprinting commercial communication devices to enhance electronic security," *International Journal of Electronic Security and Digital Forensics*, vol. 1, no. 3, pp. 301-322, 2008.

[19] I. O. Kennedy and A. M. Kuzminskiy, "RF Fingerprint detection in a wireless multipath channel," in *2010 7th International Symposium on Wireless Communication Systems*, 2010, pp. 820-823.

[20] M. W. Liu and J. F. Doherty, "Nonlinearity estimation for Specific Emitter Identification in multipath environment," in *2009 IEEE Sarnoff Symposium*, 2009, pp. 1-5.

[21] M. W. Liu and J. F. Doherty, "Nonlinearity Estimation for Specific Emitter Identification in Multipath Channels," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 1076-1085, 2011.

[22] *IEEE Std 802.11-2007, Local and Metropolitan Area Networks, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, 2007.

[23] P. K. Harmer, T. Michael, M. Buckner, and E. Farquhar, "4G Security Using Physical Layer RF-DNA with DE-Optimized LFS Classification," *Journal of Communications,* vol. 6, no. 9, pp. 671-681, Dec 2011.

[24] P. Jeffrey, G. Ben, G. Ramakrishna, S. Srinivasan, and W. David, "802.11 user fingerprinting," in *MobiCom '07: Proceedings of the 13th Annual ACM International Conference on Mobile Computing and Networking*, Montréal, Québec, Canada, pp. 99-110, 2007.

[25] R. W. Klein, M. A. Temple, and M. J. Mendenhall, "Application of wavelet-based RF fingerprinting to enhance wireless network security," *Journal of Communications and Networks,* vol. 11, no. 6, pp. 544-555, 2009.

[26] D. Takahashi, Y. Xiaoa, Y. Zhang, P. Chatzimisios, and H.-H. Chend, "IEEE 802.11 User Fingerprinting and its App for Intrusion Detection," presented at the Computers and Mathematics with Applications, 2010.

[27] H. Lajos, A. Yosef, W. Li, and J. Ming, *MIMO-OFDM for LTE, Wi-Fi, and WiMAX*. Hoboken, NJ: John Wiley and Sons, Ltd, 2011.

[28] M. Fadul and D. Reising, "RF-DNA Fingerprinting of 802.11a Wi-Fi Signals in an Indoor Rayleigh Fading Channel," May 2018.

[29] M. Gadhiok, "Symbol Timing Synchronization for OFDM-based WLAN Systems," Master’s thesis, Rice University, 2007.

[30] M. Faulkner, J. Singh, I. Tolochko, and K. Wang, "Timing Synchronization for 802.11a WLANs under Multipath Channels," presented at the Australian Telecommunications, Networks and Applications Conference (ATNAC), 2003.

[31] M. R. Raghavendra and K. Giridhar, "Improving channel estimation in OFDM systems for sparse multipath channels," *IEEE Signal Processing Letters*, vol. 12, no. 1, pp. 52-55, 2005.

[32] A. Goldsmith, *Wireless Communications*. Cambridge, UK: Cambridge University Press, 2005.

[33] A. Mitra, "A Curriculum Development Cell Project Under QIP, IIT Guwahati," Indian Institute of Technology Guwahati.

[34] K. Anusuya, S. Bharadhwaj, and S. Rani, "Wireless Channel Models for Indoor Environments," *Defence Science Journal,* vol. 58, no. 6, pp. 771-777, November 2008.

[35] B. O'Hara and A. Petrick, *IEEE 802.11 Handbook: A Designer's Companion*. Standard Information Network, IEEE Press, 2005.

[36] T. M. Schmidl and D. C. Cox, "Robust frequency and timing synchronization for OFDM," *Communications, IEEE Transactions on,* vol. 45, no. 12, pp. 1613-1621, 1997.

[37] M. Serra, P. Marti, and J. Carrabina, "Implementation of a channel equalizer for OFDM wireless LANs," USA: IEEE, 2004, pp. 232-238.

[38] Y. Huimei, L. Yingzhuan, S. Hao, and C. Wen, "Research on channel estimation for OFDM receiver based on IEEE 802.11a," in *2008 6th IEEE International Conference on Industrial Informatics,* 2008, pp. 35-39.

[39] P. H. Moose, "A technique for orthogonal frequency division multiplexing frequency offset correction," *IEEE Transactions on Communications,* vol. 42, no. 10, pp. 2908-2914, 1994.

[40] J. A. Nelder and R. Mead, "A Simplex Method for Function Minimization," *The Computer Journal,* vol. 7, no. 4, pp. 308-313, 1965.

[41] J. Lagarias, J. Reeds, M. Wright, and P. Wright, "Convergence Properties of the Nelder--Mead Simplex Method in Low Dimensions," *SIAM Journal on Optimization* vol. 9, no. 1, pp. 112-147, 2006.

[42] A. Wouk, *New Computing Environments: Microcomputers in Large-Scale Computing*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1987.

[43] M. Bastiaans, "Discrete Gabor Transform and Discrete Zak Transform," in *IEEE Int'l Conf on Signal and Image Processing Applications (ICSIPA96),* 1996, pp. 122–126.

[44] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4th ed. Burlington, MA: Academic Press, 2009.

[45] M. G. Troulis, "Fixed Point Channel Estimation for 802.11a Wireless Receivers," in *12th European Wireless Conference 2006 - Enabling Technologies for Wireless Multimedia Communications*, 2006, pp. 1-6.

VITA

Mohamed Fadul was born in Qaseem, Saudia Arabia, to his parents Khalid and Aziza. He is the third of four children, with two older sisters and one younger sister. He received his Bachelor of Science degree in Electrical and Electronics Engineering (Communications Concentration) in 2012 from University of Khartoum in Khartoum, Sudan. After graduation, he joined MTN Group LTD as a telecom engineer. He worked there for three years before accepting a graduate assistantship offer from The University of Tennessee at Chattanooga to pursue a Master of Science degree in Electrical Engineering.