APPLICATION OF RF-DNA FINGERPRINTING TECHNIQUES TO ICOM RADIO SATELLITE COMMUNICATION

THESIS

Patrick N. Dunkel, 2d Lt, USAF

AFIT-ENY-MS-17-M-258

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENY-MS-17-M-258

APPLICATION OF RF-DNA FINGERPRINTING TECHNIQUES TO ICOM RADIO SATELLITE COMMUNICATION

THESIS

Presented to the Faculty

Department of Aeronautical and Astronautical Engineering

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Master of Science in Astronautical Engineering

Patrick N. Dunkel, B.S.A.E.

2d Lt, USAF

March 3, 2017

AFIT-ENY-MS-17-M-258

APPLICATION OF RF-DNA FINGERPRINTING TECHNIQUES TO ICOM RADIO SATELLITE COMMUNICATION

THESIS

Patrick N. Dunkel, B.S.A.E.
2d Lt, USAF

Committee Membership:

Dr. Eric. D. Swenson, AFIT
Chair

Maj Addison Betances, AFIT
Member

Dr. Michael. A. Temple, AFIT
Member

AFIT-ENY-MS-17-M-258

Abstract

Device discrimination and identification using RF-DNA (Radio Frequency - Distinct Native Attribute) Fingerprinting techniques have been used by many researchers at AFIT (Air Force Institute of Technology), and RF-DNA techniques are well developed as an effective tool for a multitude of environments and signal types. The purpose of this research is threefold: 1) perform an analysis of the feasibility and reliability RF-DNA techniques on satellite communication through classification and identification of signals sent by satellite ground stations, 2) investigate the performance of CTSD (Combined Time and Spectral Domain) fingerprints compared to TD (Time Domain) and SD (Spectral Domain) fingerprints in an attempt to improve overall classification performance, and 3) analyze relationship between classification performance and sampling rate relative to the Nyquist sampling rate. Data was collected from six configurations of a ground station. Fingerprints for each device were generated in the spectral domain and time domain. The author had four hypothesis for the research result: 1) that there would be poor classification performance at low SNRs (Signal-to-Noise Ratio), but that performance would increase with increasing SNR; 2) Two devices would be more easily classified due to their age and use; 3) CTSD fingerprints would perform better than TD and SD fingerprints; and 4) Sampling rate would impact classification performance for TD fingerprints more than SD fingerprints. The initial device classification went against the hypothesis by having $< 90\%$ Average Correct Classification at low SNRs. The research utilizes a RndF (Random Forest) Classifier as well as MDA/ML (Multiple Discriminant Analysis/Maximum Likelihood) qualitative DRA (Dimensional Reduction Analysis) to determine the underlying cause of the higher than expected classification perfor-

mance. The cause identified is the distinguishable differences in fingerprint feature values that is dependent on signal power, leading to easy classification between devices. This research investigates several solutions to correcting power dependency, as well as makes an overall recommendation to perform total power normalization, normalizing each pulse in the spectral domain based on the total power of the region of interest. Total power normalization decreases overall performance while eliminating the power dependency. The utilization of classification techniques that do not rely on power dependency is more applicable to satellite communications. CTSD fingerprints results in improved classification performance at lower SNRs by as much as 15% when compared to TD and SD fingerprints. However, classification performance comparison between CTSD, TD, and SD fingerprints may vary based on signal characteristics, and therefore should be analyzed more in-depth for future research. The research identified the susceptibility of time domain fingerprint performance to lower sampling rates while noting the consistency of spectral domain fingerprint performance across all sampling rates considered above the Nyquist rate. RF-DNA Fingerprinting techniques are effective at device discrimination in a laboratory environment for satellite communications, but more work needs to be done to investigate performance in a real-world satellite communications environment.

Acknowledgements

I would first like to thank my Thesis Advisor, Dr. Eric Swenson, for providing such help and support during the process. He not only helped show an example of competence and expertise in fields of study, but also taught me the enthusiasm and energy a teacher can have towards his students and classes. Thank you to Dr. Michael Temple for helping me understand the techniques and methods utilized for this research, and for helping me in my limited understanding of electrical engineering. Thank you to Maj Betances for supporting me from the start of my experiment setup to the conclusion of my research. Thank you to all of the ENY students for not only helping me academically, but allowing me to have such great resources and mentors in my first Air Force assignment. Thank you especially to the members of ENY4, who taught me what it truly means to have camaraderie in an Air Force work organization. Thank you so much to my amazing wife for helping me survive graduate school, even when she was going through the same struggles I was going through. The past eighteen months would have been a lot more difficult if I did not have such an amazing wife constantly supporting me. Lastly, I would like to thank God for the wonderful blessings He has given me in life, and the wonderful opportunities I have and will experience.

Patrick N. Dunkel

Table of Contents

Page

List of Figures

List of Tables

List of Symbols

P_B Bit Error Rate

N_R Number of Subregions

$\vec{a}\,(t)$ Time Domain Amplitude Response

$\vec{\phi}\,(t)$ Time Domain Phase Response

$\vec{f}\,(t)$ Time Domain Frequency Response

μ Mean

σ^2 Variance

γ Skewness

κ Kurtosis

$S_{TD}(t)$ Processed Collected Complex Signal

$I_{TD}(t)$ In Phase component of Collected Complex Signal

$Q_{TD}(t)$ Quadrature component of Collected Complex Signal

N_x Total number of time samples in a signal

List of Abbreviations

AFIT Air Force Institute of Technology

AWGN Additive White Gaussian Noise

bps Bits per second

bps/Hz bits per second per Hertz

BW Bandwidth

dB Decibel

DFSK Differential Frequency Shift Keying

DRA Dimensional Reduction Analysis

EER Equal Error Rate

FSK Frequency Shift Keying

FVR False Verification Rate

GRLVQI Generalized Relevance Learning Vector Quantized Improved

kHz kilohertz

MATLAB® MATrix LABoratory

MDA Multiple Discriminant Analysis

MDA/ML Multiple Discriminant Analysis/Maximum Likelihood

MHz megahertz

ML Maximum Likelihood

PSD Power Spectral Density

RAR Rogue Accept Rate, interchangeable with False Verification Rate

RF-DNA Radio Frequency Distinct Native Attributes.

RndF Random Forest

ROC Receiver Operating Characteristic

ROI Region of Interest

RX Signal Reception

SATCOM Satellite Communication

SD Spectral Domain

SDR Software Defined Radio

SNR Signal-to-Noise Ratio

TD Time Domain

TNC Terminal Node Controller

CTSD Combined Time and Spectral Domain

TVR True Verification Rate

TX Signal Transmission

APPLICATION OF RF-DNA FINGERPRINTING TECHNIQUES TO ICOM RADIO SATELLITE COMMUNICATION

I. Introduction

1.1 Research Motivation

Cyber Warfare is a field of growing concern to the United States Air Force. Recently, the Cyber Resiliency Steering Group released the Air Force Cyber Campaign Plan which identifies potential threats and measures the Air Force should take in order to secure our assets from those threats [1]. Two of the focus areas are: 1) Making weapons systems capable of cyber resiliency and 2) Protecting current resources [2]. Bit level protection and encryption are a means of protecting the information and preventing unauthorized access to assets. However, encryption codes can be learned and authentication messages spoofed [3]. Having a means of providing for device discrimination and classification utilizing physical characteristics of the signal is useful for accomplishing the tasks outlined in the Air Force Cyber Campaign Plan. Radio Frequency Distinct Native Attribute (RF-DNA) Fingerprinting techniques have been used on terrestrial systems to classify and identify the source of transmissions [14, 33, 34, 35, 40]. Research has also been done on devices and components similar or identical to portable satellite hardware in an attempt to apply RF-DNA Fingerprinting techniques to satellite communication [4, 5, 6]. However, it is necessary to perform similar experiments done with previous research on a ground station for an orbiting satellite to determine if the same conclusions to the effectiveness of RF-DNA techniques can be drawn from a feasibility test of a satellite ground

station.

1.2 Research Objectives

This research will focus on addressing three objectives:

1. Can RF-DNA Fingerprinting techniques be utilized to correctly identify and classify the signals coming from six different communication ground-stations, and what is the potential effectiveness of applying the process to a larger scale satellite communication (SATCOM) system?

2. Can classification performance improve through the use of combined time and spectral domain fingerprints when compared to classifying the two domains separately?

3. What is the relationship between the proportion of oversampling and classification performance?

1.3 Research Approach

The research approach will focus on accomplishing four main steps:

1. Collect satellite communication signals from a ground station.

2. Perform pulse detection and post-processing on the collected signals.

3. Fingerprint pulses in spectral domain and time domain.

4. Perform Device Classification and Verification.

The current Air Force Institute of Technology (AFIT) RF-DNA process is summarized in Fig. 1.

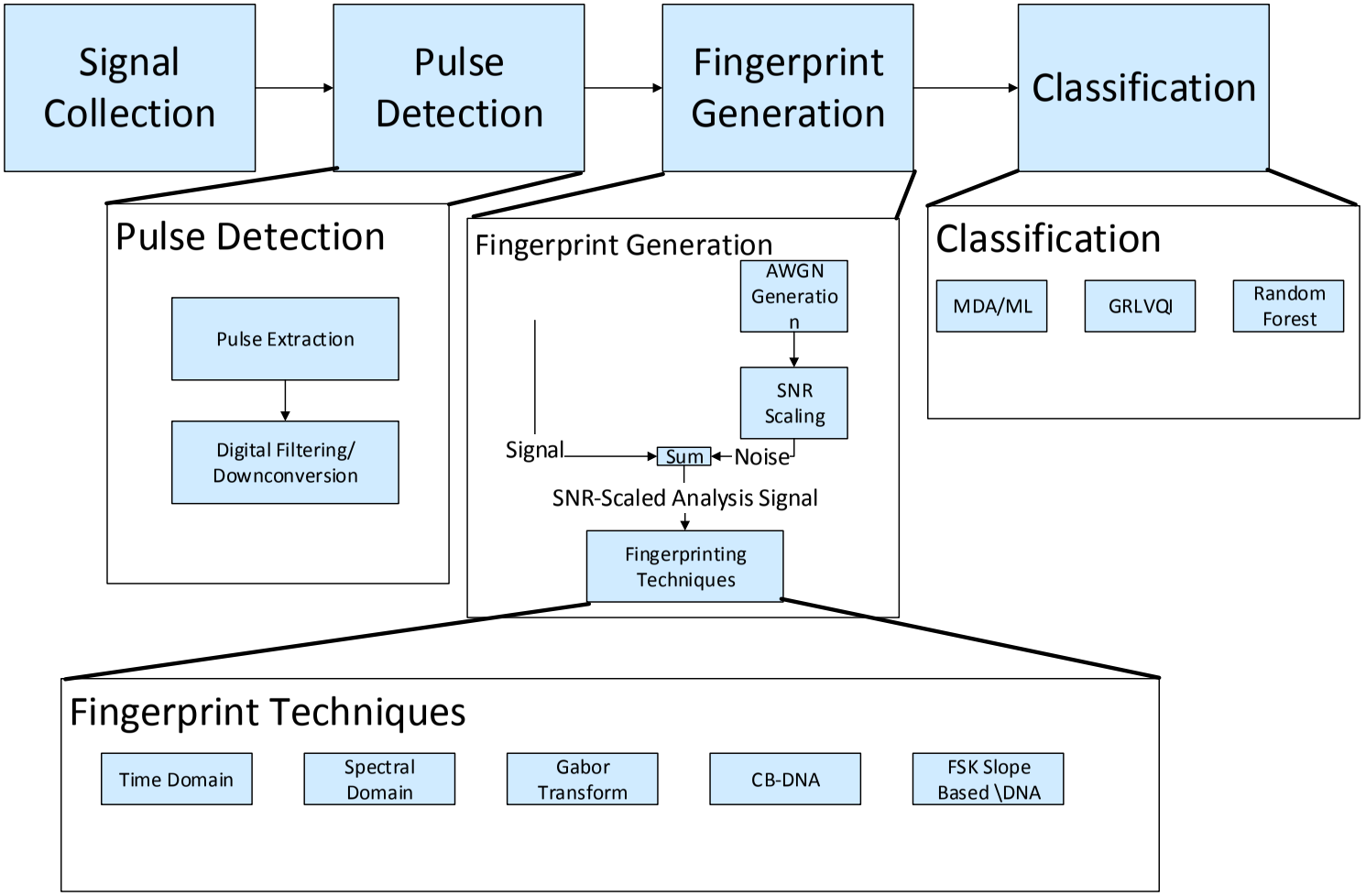

Figure 1. Current AFIT RF-DNA Process

Emission Collection and Characterization.

A valid signal from a satellite communication ground station needs to be collected as a first step in the research process. The signal can then be characterized and examined to determine what the makeup of the signal is. The objective in this section of the research is to find a source of collected data that had a large collection size for valid conclusions to be drawn from the RF-DNA process.

Signal Processing and Pulse Detection.

Once bursts of specific duration have been collected from each device, pulse detection needs to be done in order to fingerprint the signal of interest. This is accomplished by first processing the burst with filtering, down-conversion, and decimation. Then, pulse detection is done to identify and extract the individual pulses from the bursts.

Fingerprint Generation.

Once the signal is characterized, the next step will be to generate fingerprints that summarize and characterize the physical characteristics of the RF signal. Depending on the specific RF-DNA technique utilized, either a region of interest or the entire signal will be selected in order to create the fingerprints.

Device Classification and Verification.

Once the device fingerprints are generated, a classifier such as Multiple Discriminant Analysis/Maximum Likelihood (MDA/ML) or Random Forest (RndF) can be utilized to perform classification. Both methods take a set of training fingerprints and train the classifier on differences between device fingerprints. Then, a separate set of testing fingerprints is used to determine classifier accuracy. An in-depth explanation of MDA/ML and Random Forest classifiers will be provided in Chapter 2. Device ID

Verification is a process that is used after a classification model has been trained and determines how similar a "rogue" input fingerprint is to the device it is claiming to be from.

1.4 Document Organization

The document is organized as follows:

- **Chapter II - Background:** This chapter provides background information and related research on the techniques and concepts utilized, specifically signal generation and collection, fingerprint generation, and classification.
- **Chapter III - Methodology:** This chapter describes experimental methods and procedures used during the research process.
- **Chapter IV - Results And Analysis:** This chapter presents the results of the experimental process, as well as an analysis of the effectiveness of the process.
- **Chapter V - Conclusion:** This chapter summarizes the research process and results, and gives recommendations for future work.
- **Appendix A: User's Guide to RF-DNA**

1.5 Resources

Table 1 provides a summary of past research work related to RF-DNA [7]. The majority of the literature review came from the past research work, primarily from AFIT researchers as part of the RF Exploitation Lab.

Table 1. Resources

Technical Area	Previous Work
Feature Creation	
TD Features	[8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22]
SD Features	[11, 12, 18, 20, 21, 23, 24, 25]
WD Features	[16, 21]
GT Features	[10, 25, 26, 27]
CB Features	[7, 14, 28, 29, 30, 31, 32, 33]
Correlation	[34, 35]
Emission Type	
Intentional	[8, 9, 10, 11, 12, 13, 36]
Unintentional	[7, 14, 23, 24, 33, 37, 38, 39, 40, 41]
Burst	[8, 9, 10, 11, 12, 13, 14, 16, 33, 36, 40]
Continuous	[23, 24, 37, 38, 39, 42, 43]
Classification / Verification Process	
MDA/ML	[5, 6, 8, 9, 10, 12, 13, 14, 20, 21, 22, 23, 24, 25, 33, 36, 37, 44]
GRLVQI	[6, 9, 17, 20, 22, 25, 26, 41]
Random Forest	[45]
Support Vector Machine	[28, 29, 31]
k-Nearest Neighbor	[28, 29, 33]
LDA/SDA	[32]
Classification / Verification Devices	
Wireless Devices	[8, 9, 10, 12, 13, 15, 17, 21, 25, 36]
Wired Devices	[14, 33, 34, 35, 40]
Device Operations	[38, 39]
Multi-Bandwidth, Multi-User	[46]
SATCOM Devices	[4, 5, 6]
Simulated Emissions	[47]
Wired Emission Symbol Estimation	
RF SSLP	[33, 40]
CB-Based	[33]
Side Channel Analysis	
Unintentional Emissions	[14, 33, 40, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58]
Process Enhancements	
DRA	[10, 17, 20, 22, 26, 41, 59]
Constellation Point Accumulation	[28, 32]

II. Background

2.1 Overview

The purpose of this chapter is to provide background information and related research on the techniques and concepts discussed in the following chapters. It covers four main concepts:

1. Signal Modulation and Transmission.
2. Signal Detection, Collection, and Processing.
3. Fingerprint Generation.
4. Classification using MDA/ML and Random Forest Classifiers.

2.2 Kantronics KPC-9612+

The Kantronics KPC-9612+ Terminal Node Controller (TNC) is a "multi-port, multi-speed radio modem/TNC/data controller designed to fill many roles" [60]. The KPC-9612+ is part of the ground station setup used by AFIT, which will be discussed more in later chapters. The KPC-9612+ was configured into "KISS" mode which specifies a simplified communication between the TNC and ground station computer and is detailed and explained in more detail in the Kantronics User Manual [61]. The TNC acts as a modem and packet assembler/disassembler and converts the packets of a special asynchronous full duplex frame format spoken by the ground station computer into synchronous format suitable for radio transmission [62]. The packetization of the data used by the ground station and the interaction between ground station components will be discussed in the following chapters as the primary cause for the lack of an established preamble, midamble, or postamble.

The ground station is configured to use port 2 on the TNC. Port 2 utilizes Gaussian Filtered Differential Frequency Shift Keying (DFSK) with normal bandwidths (BT) of 0.3 and 0.5 [61]. BT is shorthand used to refer to the relationship between the 3 dB bandwidth for a Gaussian filter (B) and bit time (T) [63]. BT is defined as

$$BT = \frac{f_{3dB}}{R_b} \tag{1}$$

where f_{3dB} 3 dB bandwidth of the signal and R_b is the bit rate of the signal [64]. TNC Port 2 utilizes a bit rate of 9600 bits per second (bps), meaning that the 3 dB bandwidth of the signal can be configured to be either 2880 or 4800 Hz.

FSK is a method of signal modulation in which the carrier frequency of the signal is modified according to an alternating delta frequency according to the value of the desired represented bit [65]. As the bit patterns are represented by different frequencies, the signal requires more bandwidth and is less spectrally efficient (a measure of bps/Hz) than other simple modulation schemes [66]. See Fig. 2 for a visual example of FSK modulation [67].

Gaussian-filtered FSK is a variation of FSK that has one key difference: the digital data stream is passed through a premodulation baseband filter using a Gaussian-shaped frequency response. This helps avoid sharp changes in the phase slopes at the end of each bit interval [68] .

2.3 Fingerprint Generation

Analysis Signal Signal-to-Noise Ratio Scaling.

For a given collected sample of pulses, the signal-to-noise ratio (SNR) is set and based on the specific circumstances during the collection. A collected received pulse

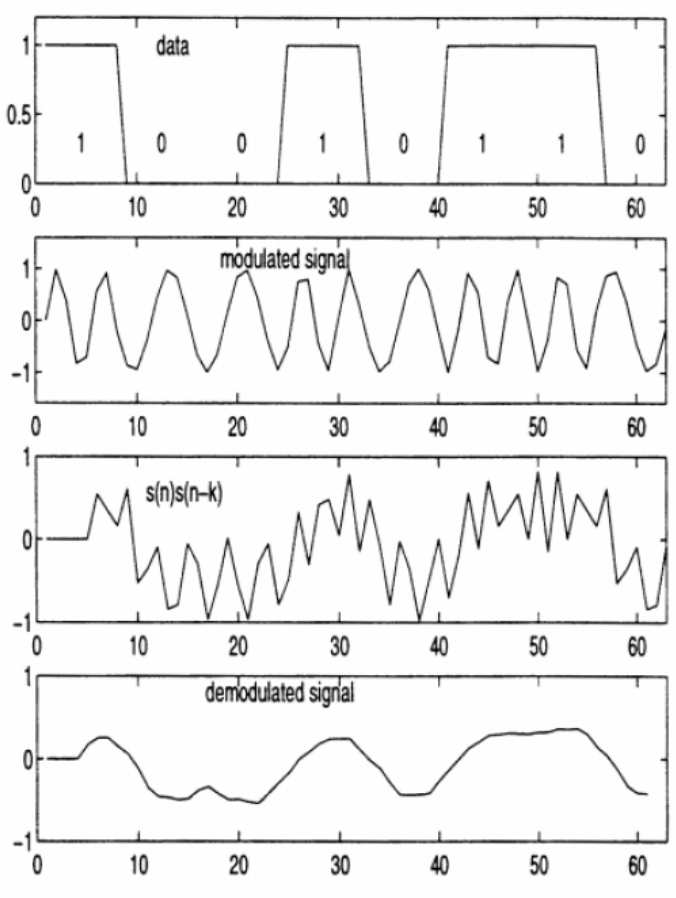

Figure 2. Example of FSK Modulation [67]

$r_c(t)$ can be expressed as

$$r_c(t) = s_b(t) + n_b(t) \tag{2}$$

where $s_b(t)$ is the collected signal and $n_b(t)$ is the collected noise [4]. The SNR can be expressed as

$$SNR = \frac{S_b}{N_b} \tag{3}$$

where S_b is the average signal power and N_b is the average noise power. A desired estimated SNR can be created with the addition of additive white Gaussian noise (AWGN), creating an $r_{c_{\text{est}}}(t)$:

$$r_{c_{\text{est}}}(t) = s_b(t) + n_b(t) + n_{AWGN}(t) \tag{4}$$

where $n_{AWGN}(t)$ is zero mean noise with a Gaussian distribution. [5] This results in an estimated SNR of

$$SNR_{est} = \frac{S_b}{N_{est}} \tag{5}$$

where N_{est} is the average power of $n_b(t) + n_{AWGN}(t)$.

An important note is that SNR's can only be estimated that are lower than the collected SNR because the addition of AWGN of smaller power than the collected noise power does not change overall noise power due to the additive property of AWGN.

For the purpose of Satellite Communications, desired SNR can be determined utilizing the bit error rate P_B. Figure 3 shows that for a bit error rate of $P_B = 10^{-2}$, the necessary E_b/N_0 is 4 dB (for BPSK modulation). A lower limit of bit error needed to be assumed. A bit error rate of $P_B = 10^{-2}$ means that one percent of all bits are incorrectly interpreted, and can be assumed to be an unacceptable error ratefor satellite communications. Therefore, $P_B = 10^{-2}$ is used as a reference P_B to

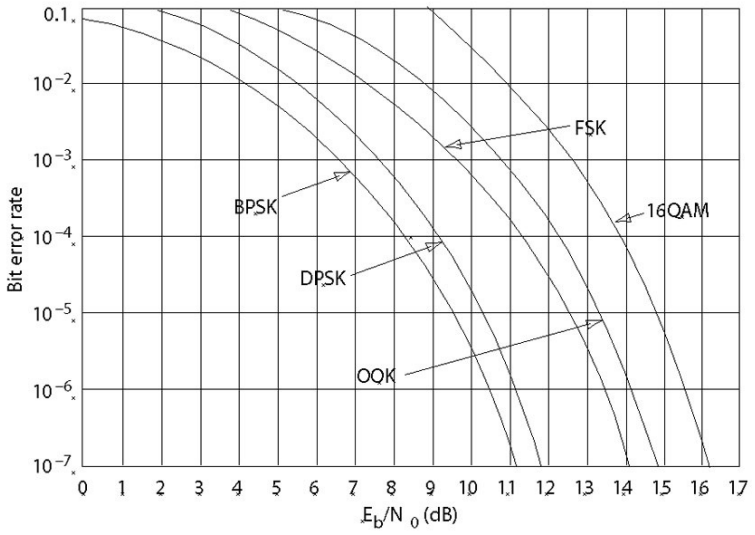

Figure 3. Bit Error Rate as a function of E_b/N_0 for different modulation schemes [69]

determine the minimum acceptable E_b/N_0. Equation (6) can be used to relate SNR with E_b/N_0 and this is dependent on the specific bandwidth B and bit rate R_b of the signal.

$$SNR = \frac{C}{N} = \frac{E_b R_b}{N_0 B} \tag{6}$$

where C is signal power and N is noise power. For the satellite signal being investigated, the SNR that would be received by the satellite would range from 5 to 35 dB.

Fingerprint Region Generation.

1D Time Domain RF-DNA Fingerprinting.

The process of 1D Time Domain (TD) RF-DNA Fingerprinting can be implemented using the following steps: [5]

1. Any final signal processing, such as decimation, down conversion, or filtering.

2. Determine TD ROI for the RF-DNA Fingerprinting process.

3. Calculate instantaneous TD responses for amplitude $\vec{a}\,(t)$, phase $\vec{\phi}\,(t)$, and frequency $\vec{f}\,(t)$.

4. Normalize each of the three TD responses separately to have a range of [-1 1].

5. Divide TD ROI into N_R subregions of equal length.

6. For each single normalized response in a single subregion, calculate the desired statistical features. For the purpose of this research, the variance σ^2, skewness γ, and kurtosis κ are considered.

7. Repeat Step 6 for every subregion and each of the three instantaneous responses.

8. Repeat Step 6-7 for all noise realizations, all SNRs, and all devices.

Steps 1 and 2 are collection specific, and the details of the process will be covered in Chapter 3. Step 3 of the process is to take the time domain signal and generate the amplitude, phase, and frequency responses. The responses can be generated with the collected complex signal $S_{TD}(t) = I_{TD}(t) + jQ_{TD}(t)$ using Eq. (7), Eq. (8), and Eq. (9) [5] [6].

The instantaneous Time Domain Amplitude Response is given by

$$\vec{a}\,(t) = \sqrt{I_{TD}(t)^2 + Q_{TD}(t)^2}. \tag{7}$$

The instantaneous Time Domain Phase Response is given by

$$\vec{\phi}\,(t) = \tan^{-1}\left[\frac{Q_{TD}(t)}{I_{TD}(t)}\right]. \tag{8}$$

The instantaneous Time Domain Frequency Response is given by

$$\vec{f}\,(t) = \frac{1}{2\pi}\left[\frac{\mathrm{d}\vec{\phi}\,(t)}{\mathrm{d}t}\right]. \tag{9}$$

The three instantaneous responses are then normalized with values from [-1 1]. Steps 5-8 will be discussed later on in the chapter in Fingerprint Subregion Generation and Fingerprint Calculation.

1D Spectral Domain RF-DNA Fingerprinting.

The process of 1D Spectral Domain (SD) RF-DNA Fingerprinting can be implemented using the following steps:

1. Any final signal processing, such as decimation, down conversion, or filtering.
2. Determine TD region of interest (ROI) for the RF-DNA Fingerprinting process.
3. Calculate power spectral density (PSD) of TD ROI.
4. Determine ROI of PSD for the RF-DNA Fingerprinting process.
5. Divide SD ROI into N_R subregions of equal length.
6. For each subregion, calculate the desired statistical features. For the purpose of this research, the variance σ^2, skewness γ, and kurtosis κ are considered.
7. Repeat Step 6 for every subregion and each of the three instantaneous responses. Combine into one vector to create a single fingerprint, as illustrated in Fig. 4.
8. Repeat step 6-7 for all noise realizations, all SNRs, and all devices.

The PSD features can be calculated using a discrete Fourier transform given by

$$X(k) = \frac{1}{N_x} \sum_{t=1}^{N_x} x(t) e^{-j\Phi(N_x,t,k)}, \tag{10}$$

where $x(t)$ is the sequence of discrete complex time domain samples and N_x is the total number of time samples in the signal [12]. Φ is given by

$$\Phi(N_x, n, k) = \left(\frac{2\pi}{N_x}\right)(n-1)(k-1). \tag{11}$$

While previous research discussed power normalization of the spectral domain [11, 12, 25], Version 10 of the AFIT RF-DNA code no longer had the capability for this process, which will be discussed later. Steps 5-8 will be discussed later on in the chapter in Fingerprint Subregion Generation and Fingerprint Calculation.

Fingerprint Subregion Generation.

Regardless of whether the ROI is in the Time Domain or Spectral Domain, the ROI is divided into N_R contiguous, equal duration subregions. If the features across the entire ROI is desired to be calculated, the total amount of regions will be (N_R+1). For each region, the statistical moments are calculated, as illustrated in Fig. 4.

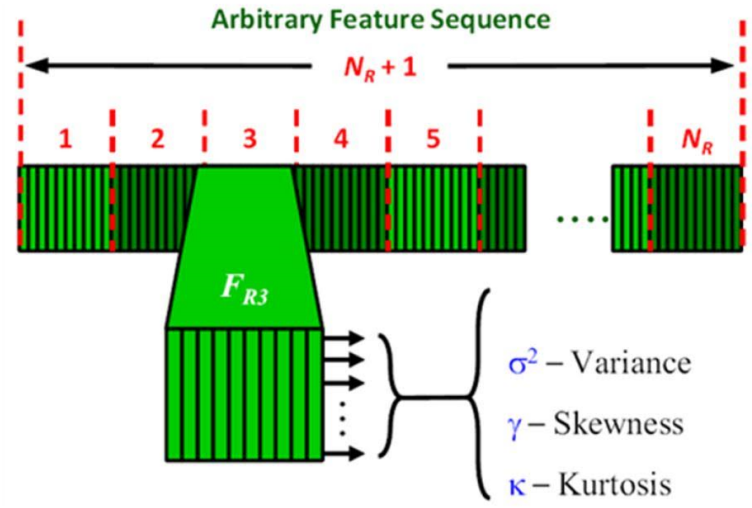

Figure 4. Depiction of Subregion Fingerprint Generation [4]

Fingerprint Feature Calculation.

The statistical moments of a signal are calculated using Eq. 12 - 17 [70]. Variance is the second statistical moment and can be calculated as

$$\sigma^2 = \frac{1}{n-1}\sum_{i-1}^{n}|x_i - \mu|^2 \tag{12}$$

where μ is defined as

$$\mu = \frac{1}{n}\sum_{i-1}^{n} x_i. \tag{13}$$

Skewness is the third statistical moment and is given by

$$\gamma_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^3}{\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2}\right)^3}. \tag{14}$$

Additionally, if bias-correction is desired, a bias-corrected skewness can be calculated as

$$\gamma_0 = \frac{\sqrt{n(n-1)}}{n-2}\gamma_1. \tag{15}$$

Kurtosis is the fourth statistical moment and is defined as

$$\kappa_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^4}{\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2}\right)^2}. \tag{16}$$

Additionally, if bias-correction is desired, a bias-corrected kurtosis can be calculated as

$$\kappa_0 = \frac{n-1}{(n-2)(n-3)}\left[(n+1)\,\kappa_1 - 3\,(n-1)\right] + 3. \tag{17}$$

2.4 Classification

The statistical fingerprints generated previously can be classified in several ways. There are many techniques and tools that have been employed by AFIT RF-DNA research; however, the two classification models that will be discussed in this research are MDA/ML and Random Forest Classifier.

Maximum Discriminant Analysis / Maximum Likelihood.

MDA/ML is actually two processes that are typically performed together. The MDA/ML process can be summarized in the following list [4]:

1. Obtain projection matrix W based on k-fold subset of training fingerprints.
2. Calculate the percent correct classification from remaining training fingerprints.
3. Repeat of Steps 1-2 for each k-fold, then select k-fold with lowest error rate.
4. Project all fingerprints into W_k, or the projection matrix corresponding to selected k-fold which had the lowest error rate.
5. Calculate means and covariances for use in distribution models later.
6. Establish trained Maximum Likelihood model boundary lines between classes.

The projection matrix W is formed using the eigenvectors of the cross-correlation and inter-correlation. Multiple subsets of fingerprints are selected to create the projection matrix W. The projection matrix W is created to maximizes the distance between the means of classes fingerprints while minimizing the variance of a single classes fingerprints [5]. k-fold cross validation is done to determine the best W matrix. The matrix with the smallest classification error is kept while all others are discarded and not used, which is done at each SNR.

Once the W model for each SNR is selected, the fingerprints are projected into the W domain and the classifier creates decision boundaries between the class spaces, as shown in Fig. 5. [5]

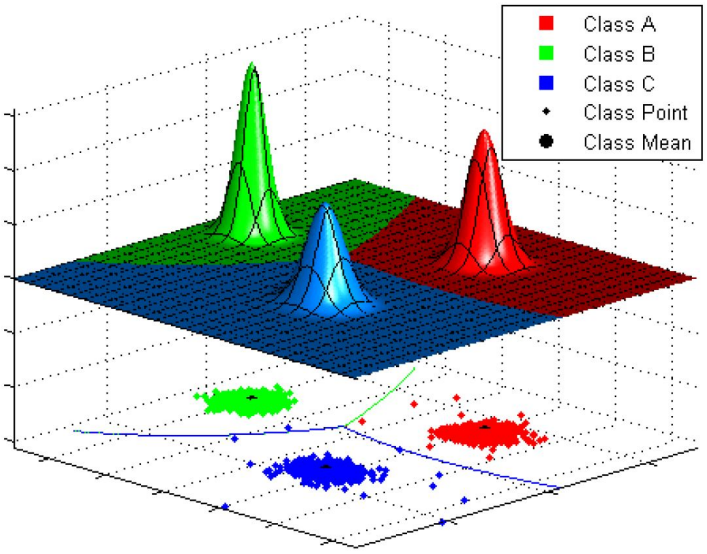

Figure 5. Representation of an RF-DNA model with decision boundaries for an arbitrary device model [5]

Device ID Verification.

Device ID Verification is a "1-to-1" comparison process that is referred to as a "Looks how much like?" assessment. It is an assessment of how much an unknown fingerprint compares to a current template of a known class of fingerprints. For a more in depth explanation of the process, see previous AFIT research [37] [24] [71] [10] [72]. For each fingerprint that is put through the verification process, there is a binary decision output of accept or reject. For a situation where there are valid device fingerprints and rogue device fingerprints together, this leads to four outcomes as shown in Table 2

A Receiver Operating Characteristic (ROC) curve is used to visualize the trade-off between accepting more valid fingerprints (True Verification) at the expense of

Table 2. Device ID Verification Outcomes

		ID Verification Result	
		Accept	Reject
Fingerprint Result	Authentic	True Verification	False Reject
	Rogue	False Verification	True Reject

additional rogue fingerprints being accepted (False Verification) [73]. Figure 6 shows an example ROC curve generated as true verification rate (TVR) vs. false verification rate (FVR) For a real world system, a desired FVR and TVR is generally decided upon and listed in communications or security requirements. The desired TVR and FVR are shown by the dotted green lines. The equal error rate (EER) is the point on the ROC curve that satisfies Eq. (18), and is typically seen as the desired operating point [4].

$$FVR = 1 - TVR. \tag{18}$$

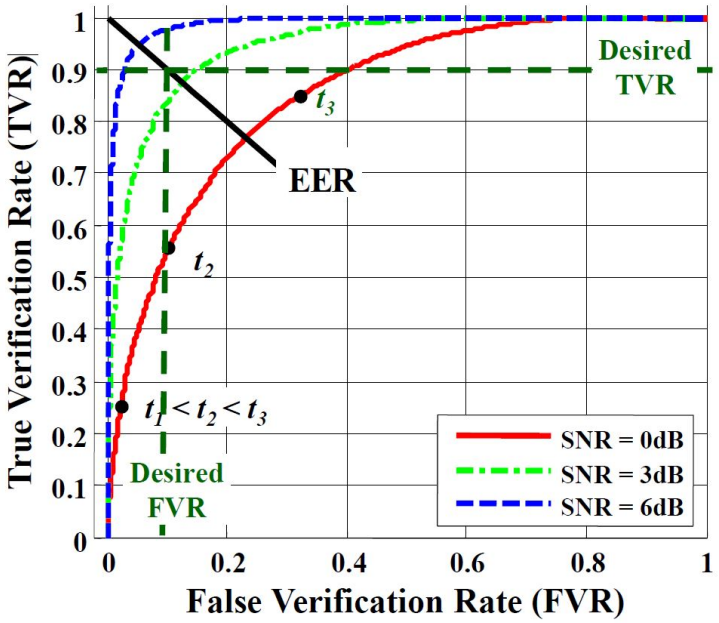

Figure 6. Representative ROC curves at three SNR values showing desired TVR and FVR performances, EER points and three arbitrary threshold t_V points [4]

Random Forest Classifier.

The Random Forest Classifier used in this research is based off of the work of Breiman [74]. A Random Forest Classifier expands on the capability of decision trees. It performs a bootstrap aggregation (bag) on ensembles of decision trees. Within the ensemble, every tree is grown on an independently drawn bootstrap replica of input data [75]. This creates a series of independent classification trees that use a subset of the data. The term "out of bag" refers to input data that is not included in the bootstrap replica for the tree.

Once the forest is "grown" Fig. 7 shows a visualization of a random forest decision. Each tree is given the unknown fingerprint and makes an independent decision, then the random forest classifier makes a decision based of the outputs of each tree.

One of the most useful outputs of the RndF Classifier is outputted in the structure with the name OOBPredictorImportance. OOBPredictorImportance is the out-of-bag estimation of predictor importance [76]. The derivation of OOBPredictorImportance is a complex process that is explained in detail in the original work of Breiman [77]. The RndF process looks at the importance of each specific variable by analyzing the change in the error when a specific variable is varied while keeping others constant [78]. Regardless of dependencies and complex interactions between variables, the value of OOBPredictorImportance is related to the increase in classification importance when the variable is selected to be used.

2.5 Summary

This chapter covered the main concepts and techniques utilized by this research. It summarized the information discovered during the research literary review, and addressed the topics of transmission and collection of data, fingerprint generation, and classification. This chapter attempted to adequately give the necessary background

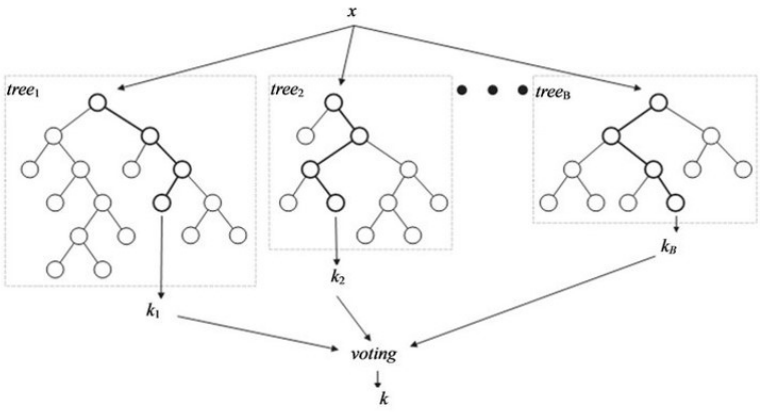

Figure 7. Random Forest Example Diagram [79]

needed to understand the concepts covered in future chapters. It is recommended to the reader that anyone wishing to continue in similar research do an in depth study of the literature contained in Table 1.

III. Methodology

3.1 Overview

The purpose of this chapter is to provide a description of the experimental methods and procedures used during this research process. Additionally, this chapter will provide clarification on the issues or concerns that were encountered along the development process.

3.2 MATLAB Setup

AFIT has a years of experience with RF-DNA, as discussed in previous chapters and Table 1. For the purpose of this research, Version 9 and 10 of the ENG RF-DNA code was combined and reorganized. Both versions of the code had similar capabilities, so the combinations of versions was done in order to increase understanding of the code, discover and correct errors with the code, and create a more user-friendly and consolidated program. The full User's Guide to RF-DNA Version 11 can be found in Appendix A.

Code Execution.

The three main portions of the code that will be discussed are

1. Pulse Detect.
2. Fingerprint Generation.
3. Classification.

3.3 Signal Collection

Experiment Components and Setup.

The experimental setup consists of the following components:

1. Signal transmission (TX) Computer with Neptune Ground Station Software.
2. Kantronics KPC-9612+ Modem.
3. ICOM IC-9100 Radios.
4. NI-2922 Software Defined Radio (SDR) Collection Receiver.
5. Signal reception (RX) Collection Computer running MATLAB®.

Figure 8 shows the interfaces between devices. For the purpose of this research, all connections were hard-wired. There were no intentional free-space collections conducted under this research.

Table 3 provides a list of the specific ICOM Radio serial numbers used for the experiment. The first two ICOM radios were purchased by AFIT several years ago, as indicated by their serial numbers, and were used operationally. The last four devices were purchased by AFIT in 2015, and have seen limited use in purely lab environments. All 4 new radios have serial numbers within 13 of each other, therefore it can be assumed they were built at the same time and with similar parts. The main hypothesis of the classification results is that since Device 1 and 2 have more wear through their use and were built at a separate time, they are expected to have the largest difference in their signal characteristics, and therefore be easiest to classify.

The first step of the experiment was to set up the TX computer. Once the Neptune ground software was started, the "per" file, or performance file, was ran. The file was set up to send 4001 pulses of the same message through the ground station. There was

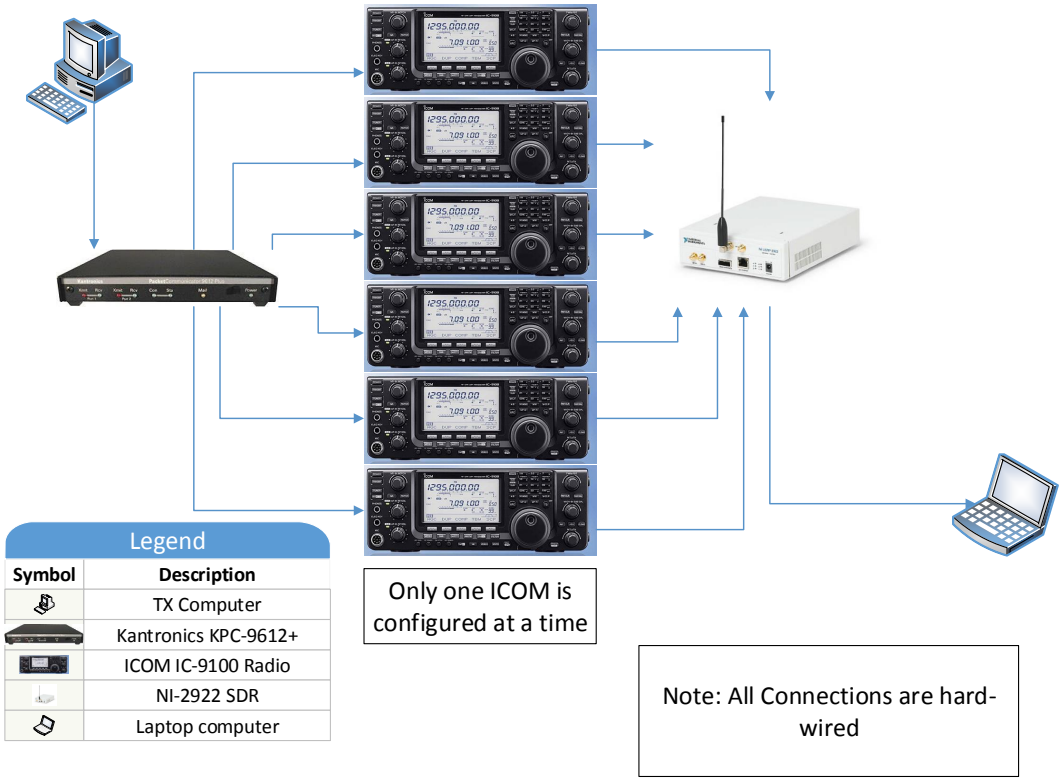

Figure 8. Hardware Experiment Setup for all Collections

Table 3. ICOM Device Serial Numbers

ICOM Dev Number	Serial Number
1	02001003
2	02001133
3	02002463
4	02002464
5	02002474
6	02002476

supposed to be a one second pause between the pulses, but the actual time between each pulse varied. This was most likely due to the TX ground station computer and the use of a wait command as opposed to a set time interval to start sending a message. The signal went through the TNC which does some data manipulation and conversion before sending it to the ICOM radios. Each time the ICOM Radios were used, they were first manually reset to ensure that all settings were the same between devices. The settings that were configured on the ICOM after being reset were:

1. Set to FM modulation.
2. Set power level.
3. Set modulation center frequency at 450 megahertz (MHz).

Next, the NI-2922 SDR was set up. Initially LabView software was used to collect and process the signal. However, the complexity of the LabView software made it difficult to determine the location of errors that might be affecting the process. Therefore, MATLAB and the USRP Software Defined Radio Toolbox was used to communicate with the NI-2922. It is recommended that future experiments also use MATLAB for collection instead of LabView. The settings configured for the NI-2922 are summarized in Table 4. One important point is that the NI-2922 collects data at a sample rate of 100 MHz. The variable "DecimationFactor" set at 32 means that the data is properly decimated in the NI-2922 by a factor of 32. This means that the sample rate of data that is sent from the NI-2922 to the laptop computer is 3.125 MHz.

A burst took approximately 60 seconds to download from the SDR to the laptop computer. Therefore, the SDR would collect for approximately 20 seconds, then send the data over to the laptop before going back to collect another burst. For this reason, there was missing data in between bursts. This is why 4000 pulses needed to be sent

Table 4. Settings for NI-2922 Collection

Setting	Value
CenterFrequency	449.95 MHz
Gain	38 dB
DecimationFactor	32
FrameLength	320,000
OverrunOutputPort	1
EnableBurstMode	1
NumFramesInBurst	50
OutputDataType	'double'

in order to collect 1600 pulses. An example burst is shown in Fig. 9, which contains 18 pulses. By analyzing the magnitude of the burst it can quickly be seen that the pulse power is above the noise power even without any filtering.

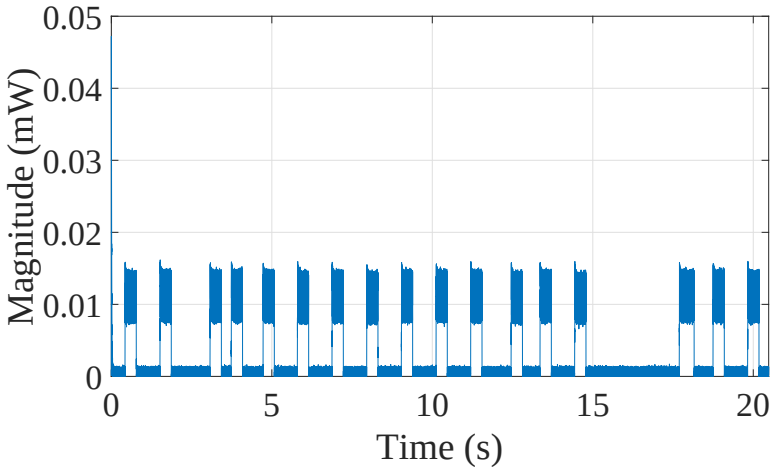

Figure 9. Representative Collection Showing Multiple Pulse Responses and DC Calibration Signal

The next step of the experiment was collection post processing. At the beginning of every burst, the NI-2922 USRP performed a DC Calibration offset that corrupted the first portion of the burst, as can be seen in Fig. 10. Each DC Calibration offset took approximately the same amount of time to complete and therefore a set amount of samples could be removed from the beginning of every burst to remove the data

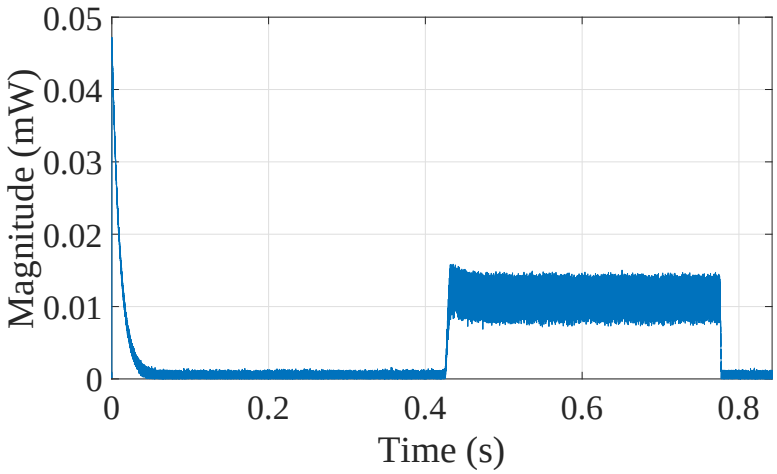

Figure 10. Expanded Collection Response from Fig. 9

corruption. The first 45,000 samples were removed from each burst in order to remove the DC calibration offset.

The collected bursts were then decimated by an additional factor of four before being saved for use in the pulse detection section of the RF-DNA process. This means that the sample rate of the data initially used for pulse detection was 781 kilohertz (kHz). This was done in order for the remaining steps of the process to be less memory intensive, as well as decreasing the computer processing time required. An initial analysis of the proper decimation factor to use resulted in the development of the third research objective, as will be discussed later in the chapter.

During post processing, the received data was analyzed to determine a possible preamble to each pulse that was the same for each signal. While each pulse contained the same command, there was no portion of the signal that was the same for each pulse. This was true when comparing magnitude, frequency, and phase as a function of time. This could be attributed to the Kantronics +9612 device that contains a bit scrambler, as well as internal packetization that occurs in the transmit computer, including sending current time, that changes with each message. It is a recommendation that future work gain a more in-depth understanding of what is being sent to

each device that is in the ground station. Therefore, it was assumed that each pulse contained a random bit stream compared to one another. The implications of this are discussed later in the chapter.

3.4 Pulse Detection

Pulse detection was accomplished using Version 11 of the RF-DNA Code. The first two section of the code are the Flag and Parameter Selections sections. A complete list of the flags and parameters selected at each step in the MATLAB RF-DNA process can be found in Appendix D.

Each pulse was bandpass filtered using an 8th order Butterworth filter centered at 54 kHz and with a bandwidth (BW) of 20 kHz to remove noise, as shown in Fig. 11. A large filter bandwidth was used for the initial filter to ensure that the desired signal was not degraded. Next, the signal was partially down-converted by 46 kHz, baseband filtered, and then decimated by a factor of four. The partial down-conversion and decimation was to decrease time of computer calculations without degrading the signal. All pulses from one device were plotted together, as shown in Fig. 12 to verify that there were no partial pulses or corrupted signals.

The first 1600 pulses for each device were then kept because the current version of the RF-DNA MATLAB code can not properly handle each device containing a different number of fingerprints. The 1600 pulses from each device were combined into one file that was then given to the Fingerprint Generation code.

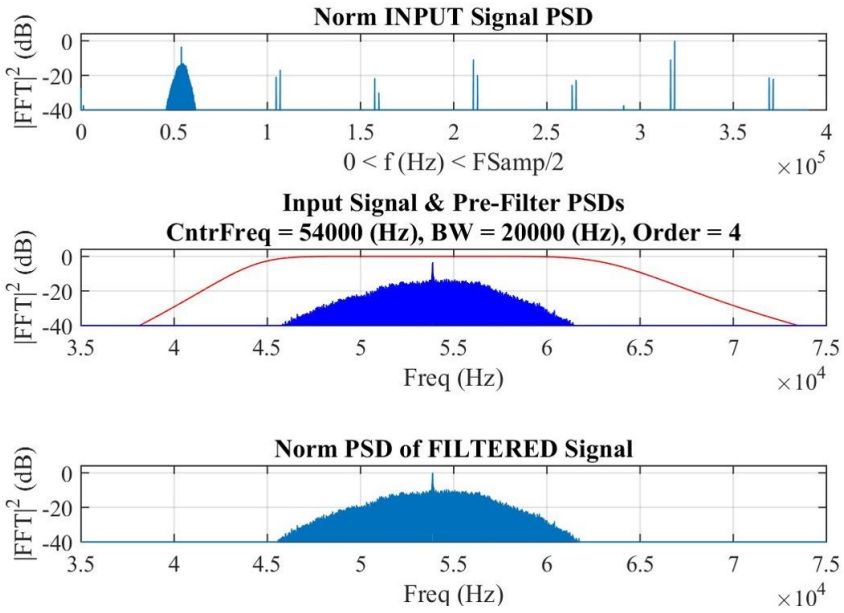

Figure 11. Pulse Detection Filter Centered at 54 kHz with a 20 kHz Bandwidth

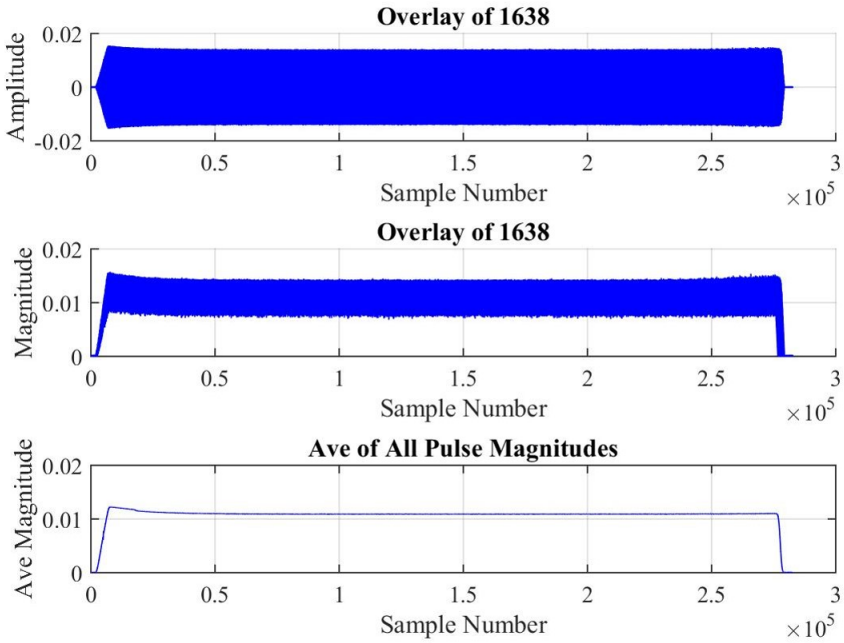

Figure 12. Pulse Overlay Demonstrating an Absence of Preamble, Midamble, or Postamble that is Common Amongst All Devices

3.5 Fingerprint Generation

Fingerprint generation was accomplished in MATLAB using Version 11 of the RF-DNA Code. The first two sections of the code are the Flag and Parameter Selections sections. A complete list of the flags and parameters selected at each step in the MATLAB RF-DNA process can be found in Appendix C.

Fingerprint generation was done in the time domain and the spectral domain. The experimental results will first be presented separately, and then combined. However, the first step of either process is the same and is to select the region of interest.

As discussed during the Signal Collection section, each pulse was assumed to have a random bit stream. There was no preamble, mid-amble, or post-amble region of interest that could be selected that was the same for every pulse. This led to the conclusion of using essentially the entire signal as the region of interest, as shown in Fig. 13.

The region of interest was then bandpass filtered, as shown in Fig. 14. Compared to the bandpass filters performed during Pulse Detection, the filter was a smaller bandwidth. Research has been done on side channel analysis as summarized in Table 1, but that was not addressed in this research. The bandwidth of the filter is most directly tied to the location of the region of interest in the spectral domain, as will be discussed later. Once the initial region of interest selection and filtering was completed, the research process diverged based on the Spectral or Time Domain fingerprinting.

Spectral Domain.

Once the signal is converted into the spectral domain, a new region of interest has to be selected. The region of interest is determined by the -3 dB limit from the bandpass filter. Twelve subregions were selected, as well as the inclusion of the total

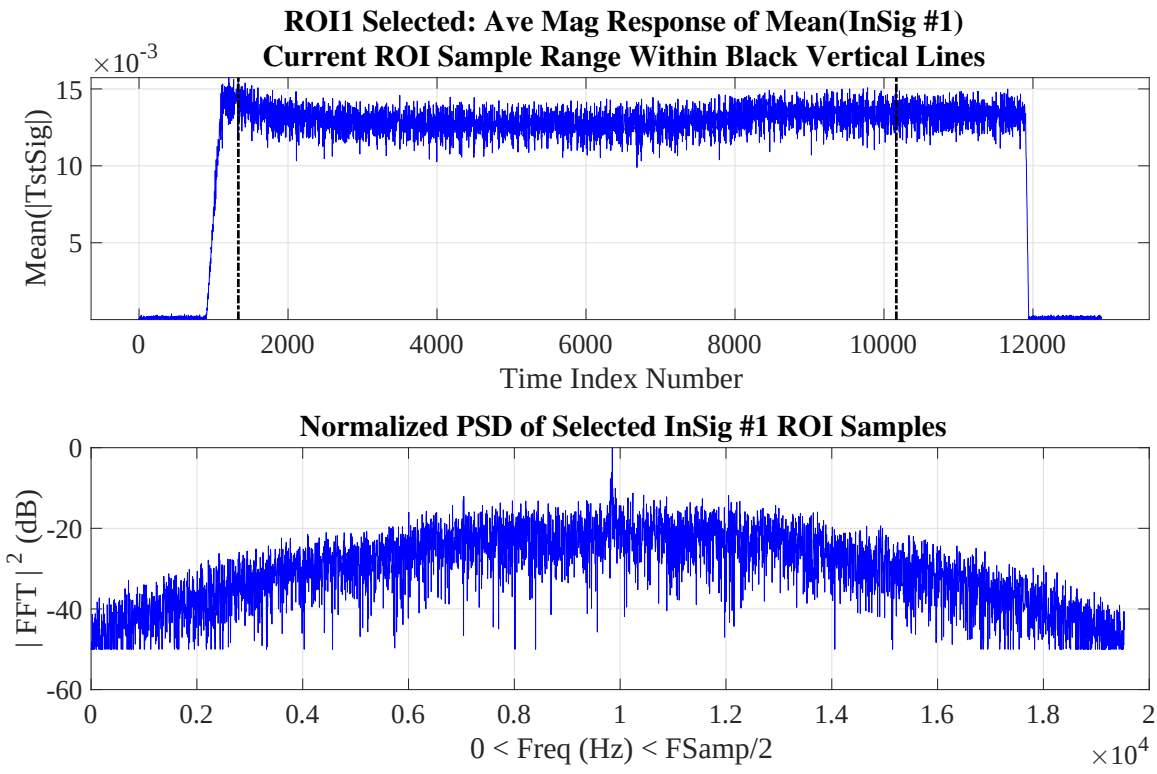

Figure 13. ROI Selection of the Majority of the Pulse Due to the Absence of an Amble

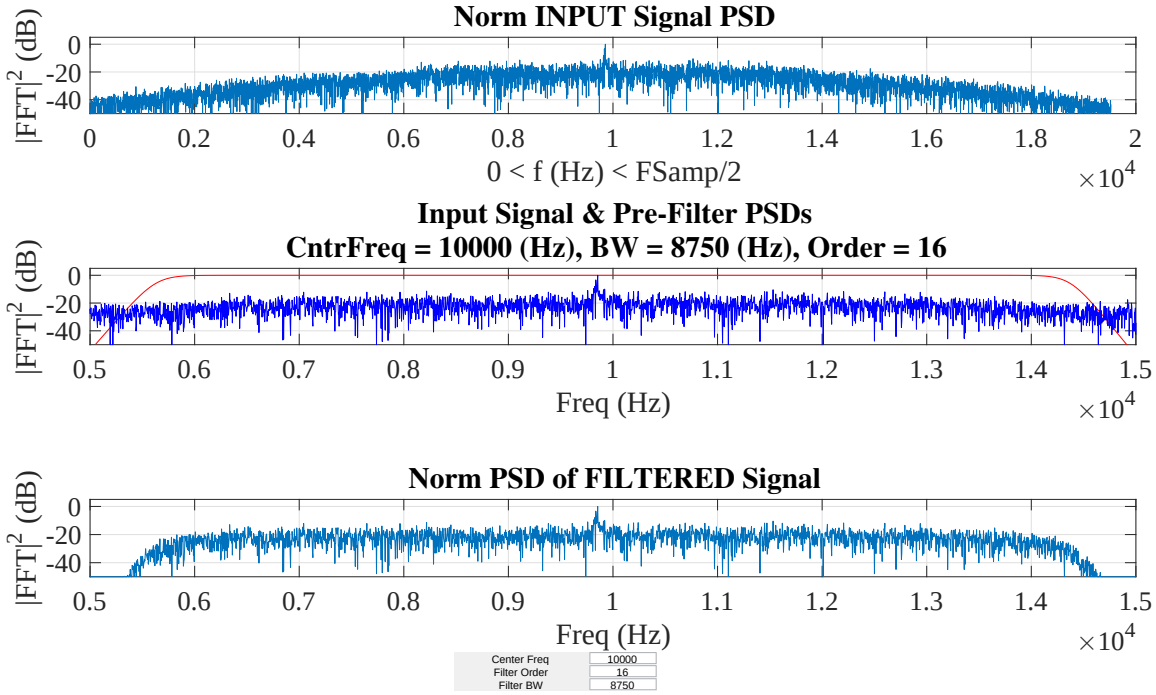

Figure 14. Fingerprint Generation Filter Centered at 10 kHz with a 8.75 kHz Bandwidth

region of interest as its own statistical region. This resulted in thirteen regions that the fingerprints are calculated for in the spectral domain. The total region of interest, as well as each of the twelve subregions, can be seen in Fig. 15. The length of each subregion is equal.

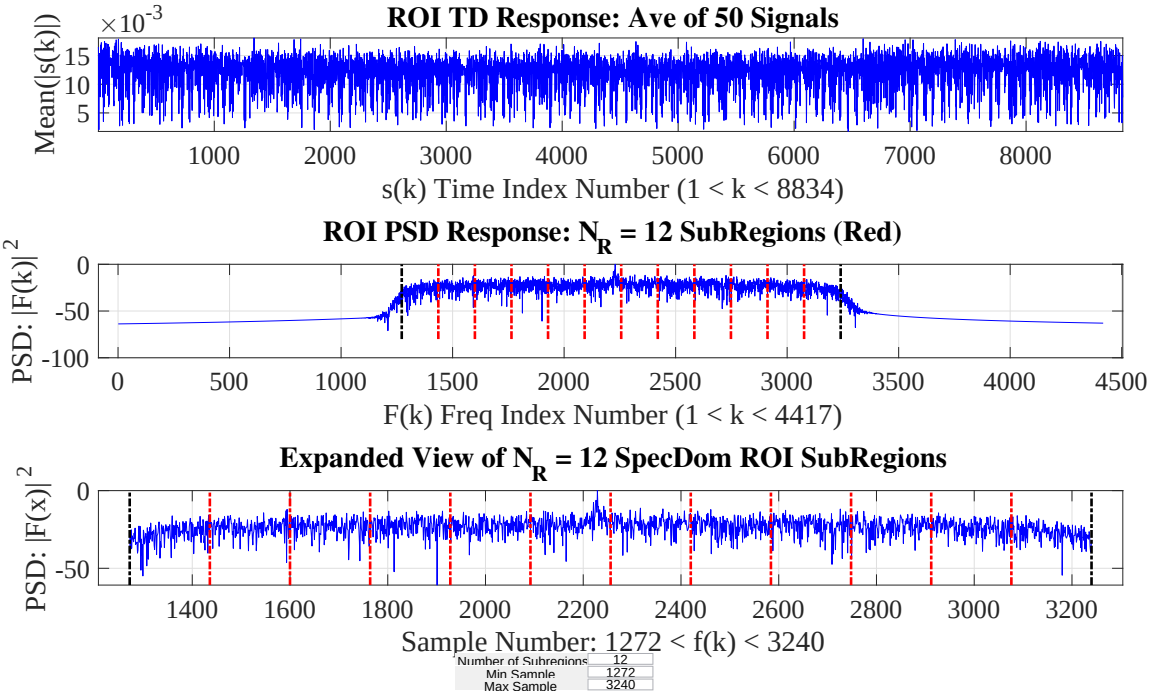

Figure 15. Fingerprint Generation Subregion Selection for SD Fingerprints, $N_R = 12$

Time Domain.

In the time domain, the features in amplitude, phase, and frequency were all calculated. Fig. 16 shows an example portion of the region of interest. The top plot is the normalized amplitude response in the time domain. The middle plot is the normalized phase response in the time domain, and the bottom plot is the normalized frequency response in the time domain. The statistical moments were calculated for each of the three signal spaces.

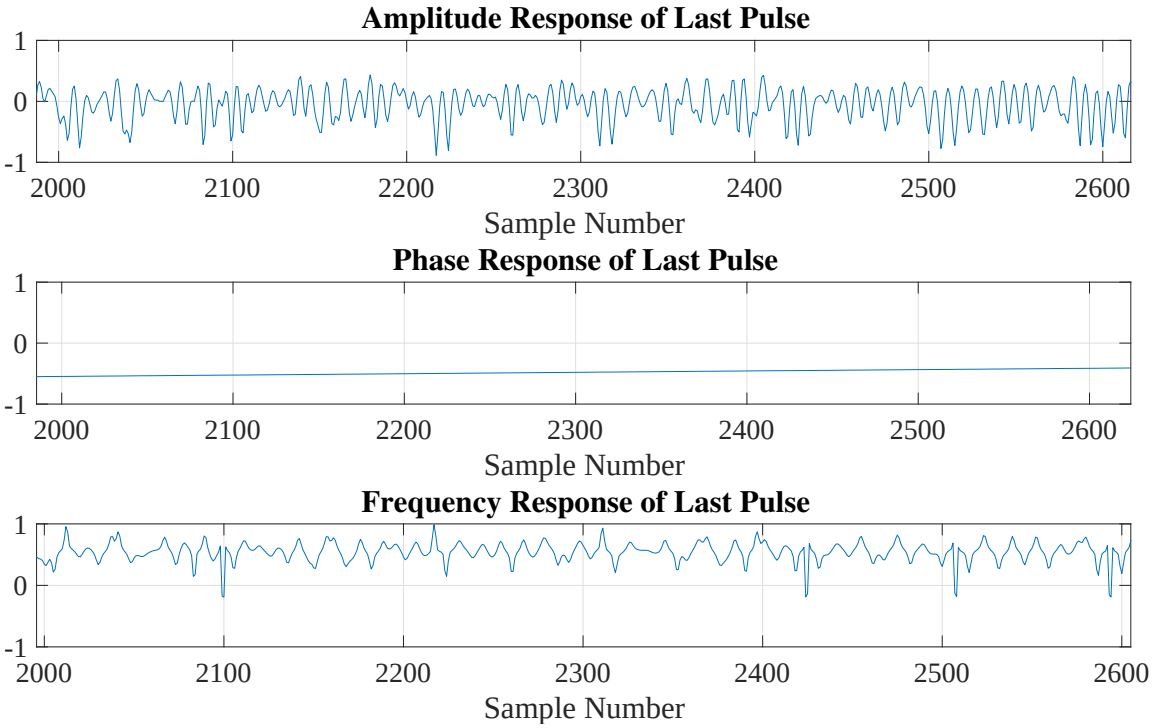

Figure 16. Time Domain Normalized Response in Amplitude, Phase, and Frequency

3.6 Initial Classification

There are two classification methods that were used in the research, MDA/ML and Random Forest.

MDA/ML Classifier.

The original fingerprint generation process was done using estimated SNR's ranging from -25 to 55 dB in order to investigate a potential issue in the classification process. When the MDA/ML classification process was done, the results were much different from the hypothesis. The average percent correct classification is shown in Fig. 17. The average percent correct classification at an SNR of -25 dB is 79%. At -25 dB, the noise power should be over 300 times the signal power, meaning that the signal is virtually indistinguishable. Therefore, the hypothesis was that there was something that the classifier was picking up on in the noise addition.

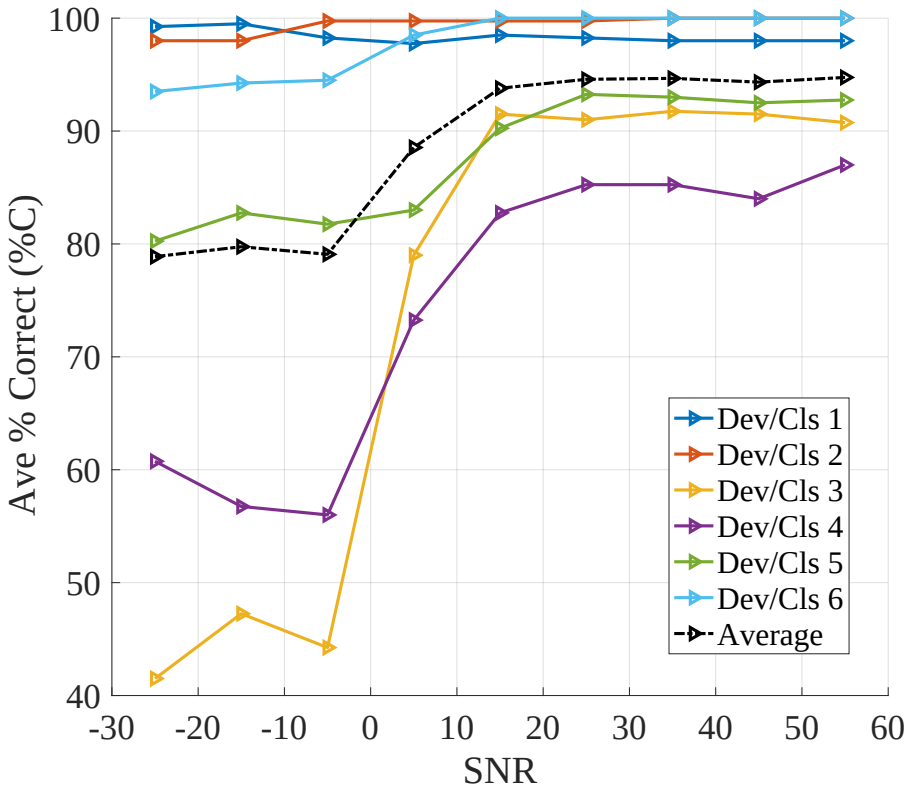

Figure 17. Initial Average Percent Correct Classification for Spectral Domain utilizing MDA/ML Classification, demonstrating higher than expected classification performance at low SNRs

Before first looking at the classifier, first the possibility of incorrect noise addition must be ruled out. Figure 18 shows the PSD of the filtered signal, filtered noise, and the addition of the signal with the noise. The desired SNR as a result of this process was -25 dB. The power of the noise being added is correct, meaning that the issue of high classification at low SNR is not the result of improper noise power addition. This leads to the necessity of gaining a better understanding of what the classifier is discriminating at low SNR's in an effort to gain some clarity. The Random Forest Classifier can be used to provide an indication of the fingerprints that MDA/ML might be using for the classification.

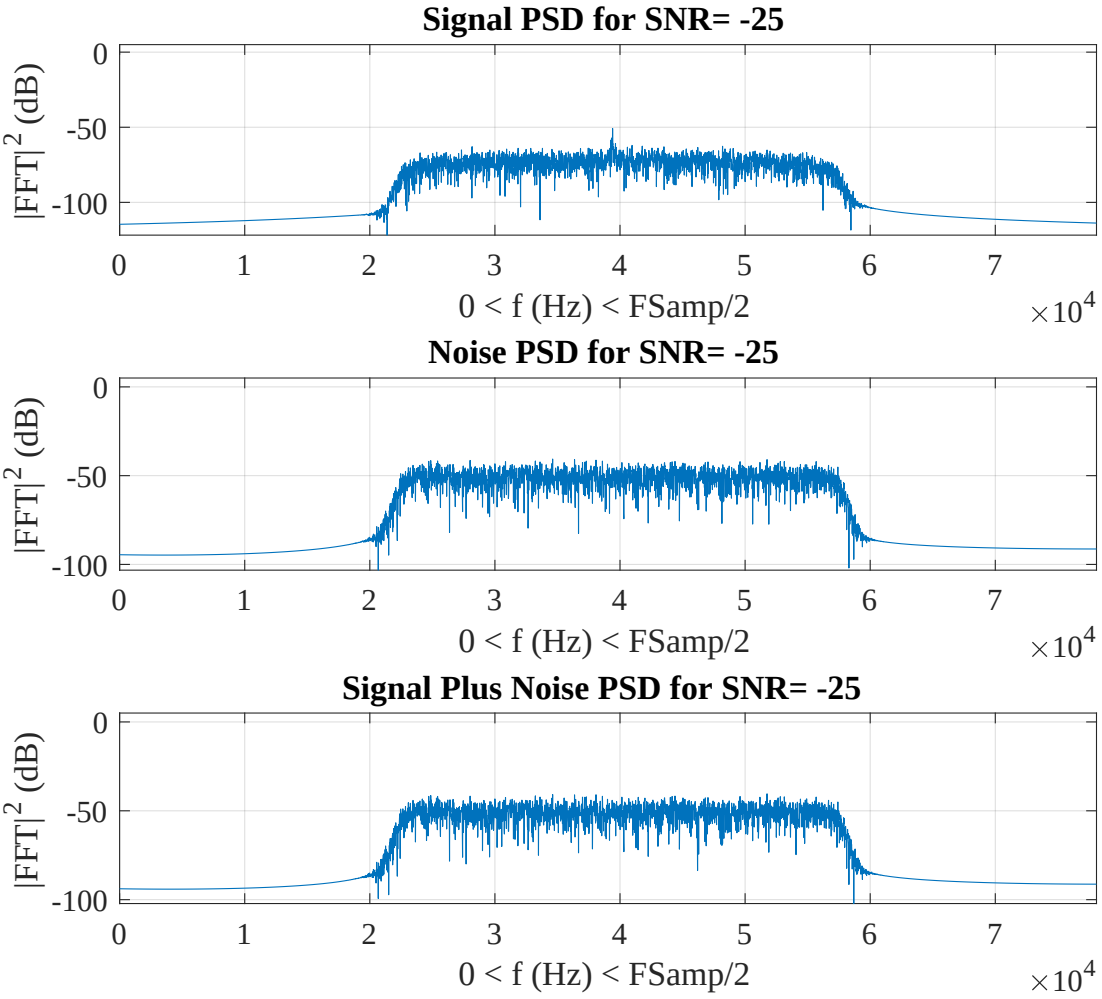

Figure 18. PSD of Signal, Noise, and Noise Plus Signal; Demonstrating the Correct Addition of Noise in the SNR Portion of the Process

Random Forest Classifier.

The Random Forest classifier shows in Fig. 19 that the fingerprint that is most important in the Random Forest classifier process is the final fingerprint in the variance subset, which is the variance of the entire region of interest. This indicates that at low SNR values, the MDA/ML classifier might be classifying based off of a difference in the variance of the region of interest. However, the variance of each subregion are also more effective at classifying than a random guess, indicating that

every variance calculation is contributing. In AFIT RF-DNA research, random forest classification does not have as much use as MDA/ML due to the speed and performance of MDA/ML. Therefore, the indications from the random forest classifier are taken to the MDA/ML classifier as a verification step.

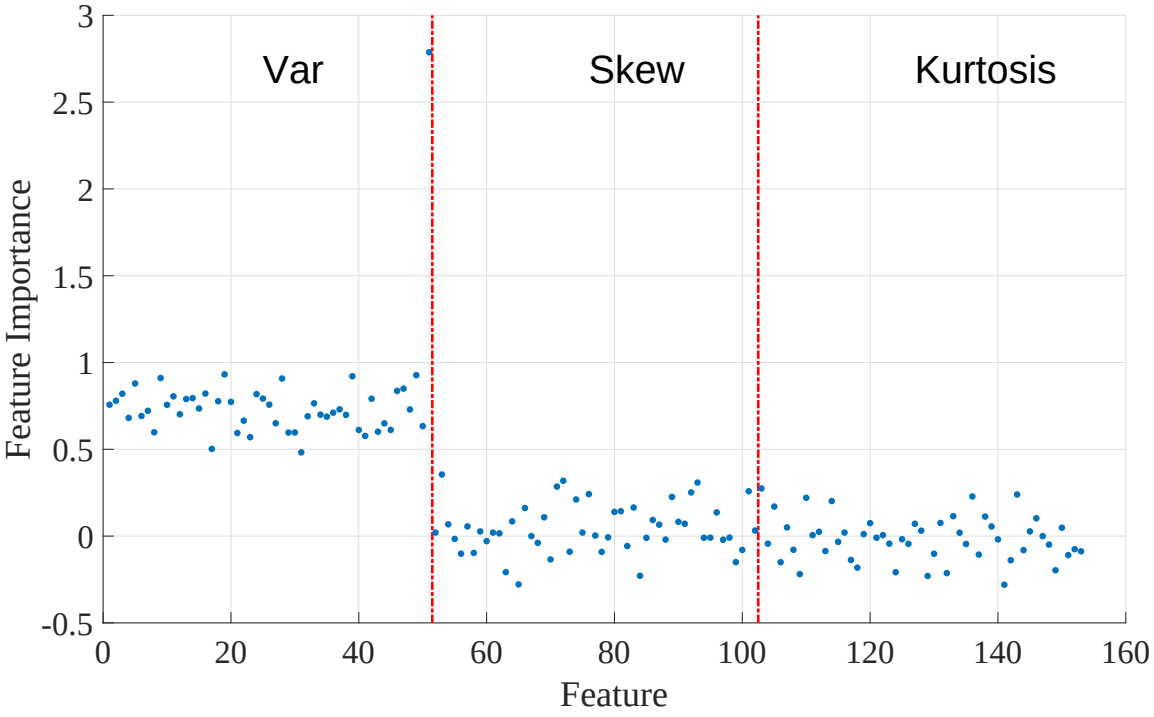

Figure 19. Random Forest Variable Feature Importance Output Divided into Features Calculated from Separate Statistical Moments

3.7 Alternate Fingerprint Features Designed to Address Power Dependency

Determination of Power Dependence.

Through Dimensional Reduction Analysis (DRA), the MDA/ML Classifier can develop a model based only on a subset of the fingerprints. Figure 20 shows classifier performance when only using variance, skewness, and kurtosis features as well using the full dimensional fingerprint classifier. The dotted lines indicate DRA utilizing the fingerprints associated with the entire region, while the solid lines utilized only the

subregions without calculating statistics on the $(N_R + 1)$ region. Figure 20 demonstrates that the variance of the noise is what the classifier is using at low SNR levels. The qualitative DRA done with MDA/ML supports the quantitative results of the Random Forest Classifier.

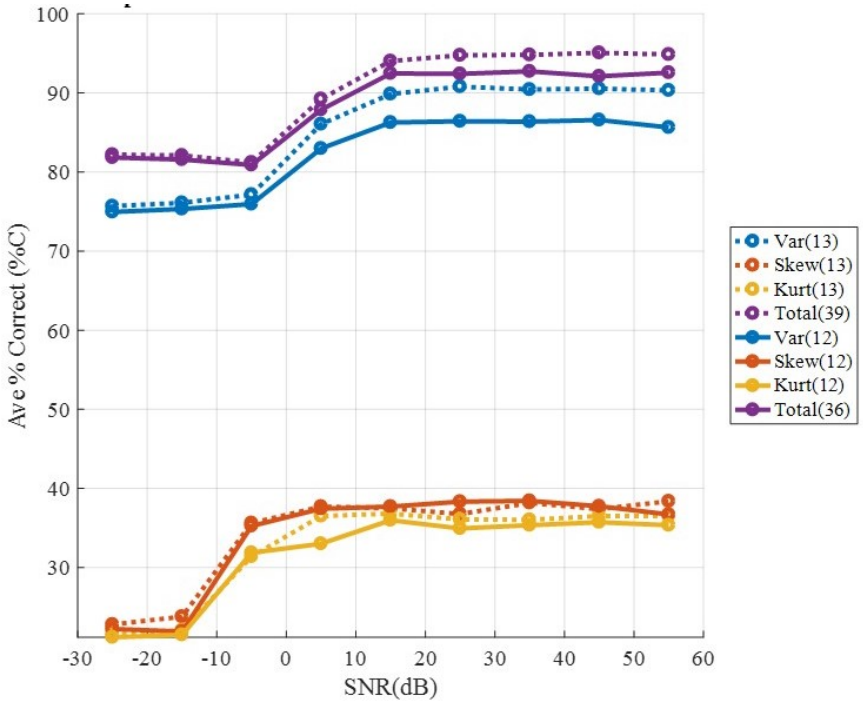

Figure 20. Original Experiment Qualitative DRA Demonstrating High Classification Permformance for Variance Features

The variance of a signal in the spectral domain is closely related to the power of the signal. Therefore, the next step is to analyze the signal power for each device. Figure 21 shows the estimated signal power of each pulse for each device. Each device is within an average of 1.8 dB from each-other for signal noise. However, the small difference between the average pulse SNR does not matter as much as the fact that the signal powers in a specific device does not vary much between pulse, and there is relatively little overlap between different devices in terms of signal power, with the notable exception of Device 3 and 4. This means that the classifier might be able to detect the subtle differences in the variance of the signal in the spectral domain. Table

5 shows the average estimated SNR for each device, and Table 6 shows the minimum and maximum variance of the spectral domain subregions. There is a correlation between the device SNR and the variance of the added noise at low estimated SNR levels.

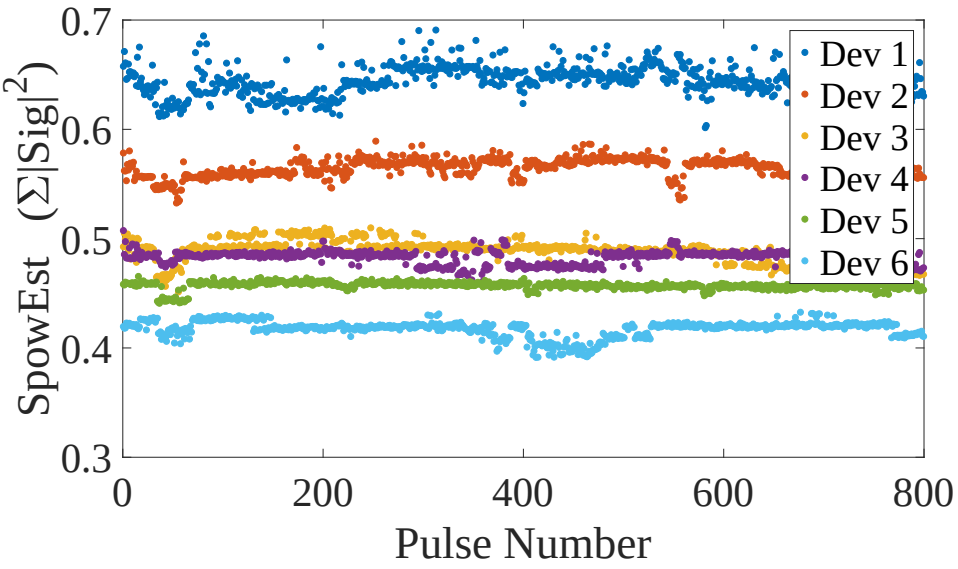

Figure 21. Estimated Collected Signal Power by Pulse for each Device

In order to generate valid classification results that are translatable to real-world satellite communications with varying SNR during uplink and downlink, a method has to be implemented for SD Fingerprinting that compensates for differences in pulse signal power.

Table 5. Device Estimated Average SNR for 800 Pulses per Device

	Ave Est SNR (dB)	SNR Ranking
Dev 1	40.1	1
Dev 2	39.2	2
Dev 3	38.4	4
Dev 4	38.5	3
Dev 5	38.3	5
Dev 6	37.9	6

Table 6. Minimum and Maximum Values of SD Subregion Variance by Device

	Min Variance	Max Variance
Dev 1	1.10E-11	8.42E-11
Dev 2	9.55E-12	5.61E-11
Dev 3	6.13E-12	4.74E-11
Dev 4	8.14E-12	4.95E-11
Dev 5	6.45E-12	3.70E-11
Dev 6	5.78E-12	3.39E-11

Methods for Addressing Power Dependence.

Four methods are proposed and described to correct for the power dependence:

1. Method 1: Normalize PSD Features by Peak Power.
2. Method 2: Normalize PSD Features by Total Power.
3. Method 3: Fingerprint Generation using SD dB Features.
4. Method 4: Adding Noise Normalized by Average Device Signal Power.

Method 1: Normalize PSD Features by Peak Power.

One possible method of correcting for a difference in signal power between devices is to do a pulse-by-pulse SD normalization based on the peak power. This is done prior to generating the fingerprints for each subregion. Equation (19) provides the mathematical formula incorporated into the MATLAB code to allow for this functionality

$$F_{\text{NORM}}(\omega) = \frac{F(\omega)}{\max\left(F(\omega)\right)}. \tag{19}$$

There is one main downside to this method, and that is frequency-specific noise dependency. The max signal power is determined in the spectral domain utilizing the maximum value for a specific frequency. Noise at the specific frequency can affect the

value of the maximum signal power. This results in a calculated max signal power that is noise-susceptible.

Method 2: Normalize PSD Features by Total Power.

The second method of correcting for a difference in signal power between devices is similar to the first method. Pulse-by-pulse normalization is still done in the Spectral Domain with the exception of the pulse being normalized based off of total power instead of peak power. Equation (20) provides the mathematical formula used in MATLAB to calculate PSD.

$$F_{\mathrm{NORM}}(\omega) = \frac{F(\omega)}{\sum [F(\omega)]}. \tag{20}$$

The main advantage of this method compared to the previous method is that, with the assumption that the noise is AWGN, the summation of noise for each frequency will average out the effects of the noise on the total signal power.

Method 3: Fingerprint Generation using SD dB Features.

This method was created through the examination and modification of the RF-DNA code, and was not expected to have the outcome it did. The mathematical comparison between calculating SD fingerprints in base-10 with calculating SD fingerprints in base-10 logarithmic is not trivial. At initial glance, this method was the most promising because it resolved low SNR incorrect classification without a loss of any information. However, further analysis revealed that this method did not correct the issue. Figure 22 highlights the comparison between SD fingerprint features in base 10 and base 10 logarithmic. The difference between the variance values are not as simple as scalar (or logarithmic) differences of each other.

An example can help to illustrate the complexity. Calculation of the first statistical

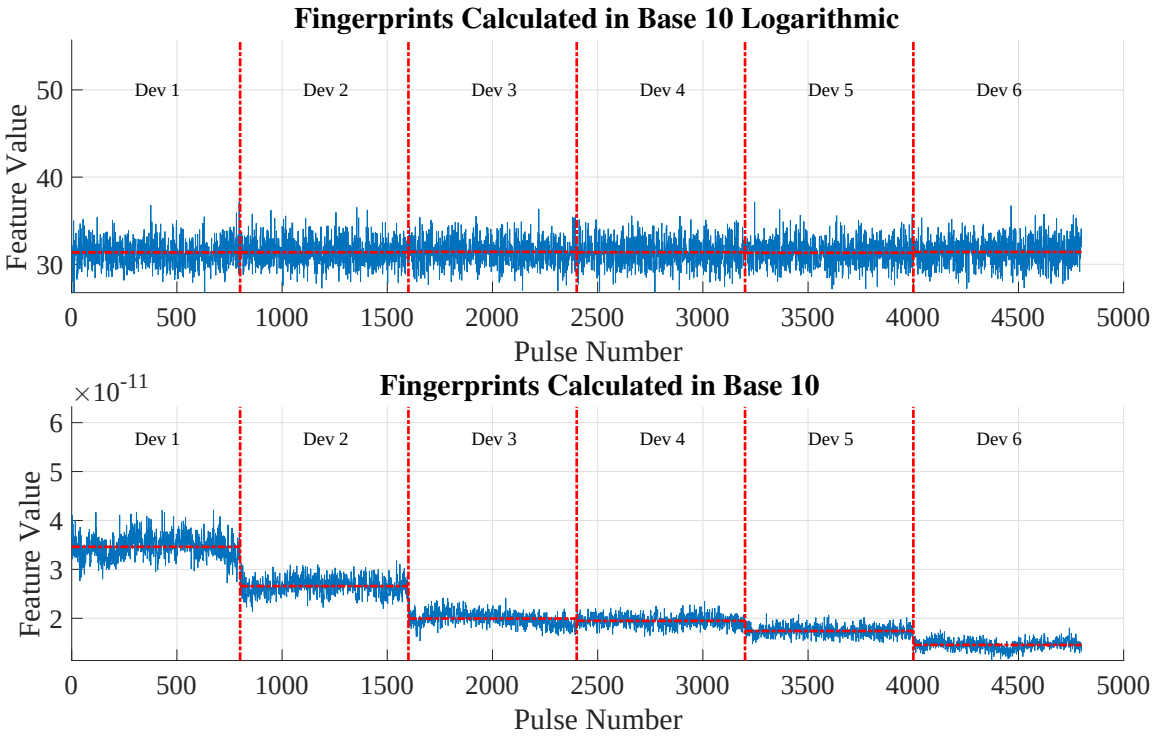

Figure 22. Feature Value for Feature=37 (Variance of entire Region of Interest in Spectral Domain)

moment of an example vector $x = [34\ 21\ 3]$ can be found with

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = 19.3\overline{3}. \tag{21}$$

The same vector converted to base 10 logarithmic is

$$x_{\text{ dB}} = 10\log_{10}(x) = [15.31\ 13.22\ 4.77]\ \text{ dB}. \tag{22}$$

The mean of this vector using Eq. (21) is 11.10 dB. Neither this value nor the base 10 conversion of 11.10 dB (12.89) equals the mean of the original vector.

The actual comparison to the mean of a base 10 logarithmic vector to a base 10 vector is given by

$$\overline{x}_{\text{ dB}} = \frac{10\log_{10}\left(\prod_{i=1}^{n} x_i\right)}{n}. \tag{23}$$

This example serves to highlight the fact that comparisons to statistical moments between base 10 and base 10 logarithmic is not a simple process. The second, third, or fourth statistical moment are more complicated and will not be explained. For this reason, a more in depth analysis of the underlying reason of the base 10 logarithmic not being power dependent was not undertaken. Instead, the process was considered to be a completely separate technique, and evaluated based on its performance.

Method 4: Adding Noise Normalized by Average Device Signal Power.

In the original RF-DNA code, the noise power N_{Pow} was determined pulse by pulse:

$$N_{\text{Pow}} = \frac{S_{\text{EstPow}}}{(N_{\text{EstPow}} \times SNR_{\text{desired}})}. \tag{24}$$

The code was modified to add noise based on the average pulse power for every device and every pulse:

$$N_{\text{Pow}} = \frac{\sum_{i=1}^{n_{\text{dev}}} \left(\sum_{j=1}^{n_{\text{pulse}}} S_{ij_{\text{EstPow}}} \right)}{n_{\text{dev}} n_{\text{pulse}} \left(N_{\text{EstPow}} \times SNR_{\text{desired}} \right)}. \tag{25}$$

Initially, this process had promising results as it led to the correction of classification performance at low SNR's while still giving a high average correct classification at higher SNR's. However, while the issue of power dependent classification is easier to identify at lower SNR's, the issue still occurs at higher SNR's. This method does not resolve different signal powers between devices once the noise power is lower than the signal.

Comparison of Methods on Performance.

Figure 23 shows the average percent correct classification for each of the methods discussed when compared with the original. Method 4 appears to have the best results, however for the reasons discussed in the previous section, it is not recommended. Method 3 had the lowest performance, and therefore is not recommended for this signal. However, it is recommended that future research look into a better understanding of the process, and determine if the method would be more useful to other signals. Method 1 and 2 have similar results, with Method 2 having slightly better performance. This is most likely due to the previously mentioned noise susceptibility of Method 1. Therefore, it is recommended that Method 2 be utilized for correcting power dependency of signals.

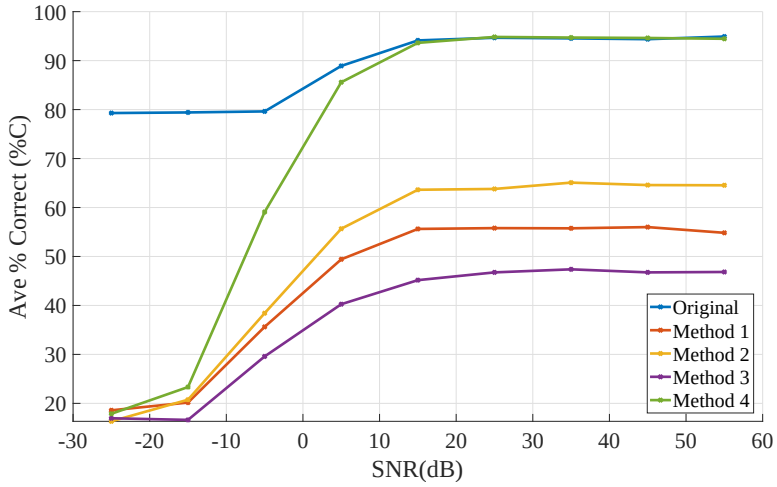

Figure 23. Average Percent Correct Classification for MDA/ML using Alternate Fingerprint Features

3.8 Sampling Rate

The RF-DNA process takes a fair amount of computational power in order to complete, meaning that depending on the signal characteristics a typical computer might struggle with the memory and computing required. Additionally, the majority

of information about the signal needs to be stored if the fingerprinting process is not completed right away. For satellite ground stations, such a capability could be added to existing hardware through the use of additional computational machines as well as storage devices. While there is a cost associated with this, it can typically be implemented. However, satellite subsystems are typically much more constrained and have less ability to add capacity during the design phase. Additionally, existing satellites can not be upgraded with new hardware as easily as their ground station counterparts. For this reason, in order for RF-DNA techniques to most easily be applied to satellite communications systems, the needed computational ability and storage space should be conserved.

An experiment was performed to determine what change in classification performance occurred when modifying the sample rate. The hypothesis is that while SD fingerprints have poorer initial correct classification, they would be better able to handle signals at lower sample rates. TD fingerprints will benefit from higher sample rates as more samples will help in averaging out noise.

Additional decimation factors of two to sixty were performed in increments of two, leading to an effective sampling rate ranging from 13 to 390 kHz. The range of sampling rates are still above the Nyquist rate of twice the signal bandwidth (4.8 kHz).

The same parameters were used for this experiment as the previous experiments, with the three notable exceptions.

1. The pulses were down-converted to baseband based on their center frequency as opposed to only being partially down-converted. The advantage of a full down-conversion is that pulses can be decimated by a higher factor while still preserving the frequencies the signal message is contained in.

2. Only four devices were used, which were Devices 2-6. These devices were chosen

because they were shown to have the highest amount of cross-class confusion, as will be discussed in Chapter 4.

3. Only 400 pulses (200 for training and 200 for testing) were used from each device in order to allow for efficient computation at higher sample rates.

All parameters and flags are constant across devices. The results are discussed in Chapter 4.

3.9 Summary

Chapter 3 discussed the methodology and experimental steps used during the research. Specifically, it addressed and summarized the initial collection, pulse detection, and fingerprint generation. Additionally, it detailed the main experimental problem encountered during the initial classification: the power dependency resulting in misleading correct classification results. Four methods were discussed to correct for the power dependency, with the method that was ultimately recommended being to normalize PSD features by total power.

IV. Results and Analysis

4.1 Overview

The purpose of this chapter is to present the results of the experimental process, as well as an analysis of the effectiveness of the process.

4.2 Classification

MDA/ML.

Figure 24 shows the 3D Fisher Plot of the six class MDA/ML classification in the Time Domain. The 3D distribution of Class 4 and 5 can be seen to intersect. The 3D distribution of Class 3 and 6 can also be seen to intersect. Device 1 and 2 are most easily classified. This supports the hypothesis of Device 1 and 2 having physical differences in the signal due to the age and use of the devices.

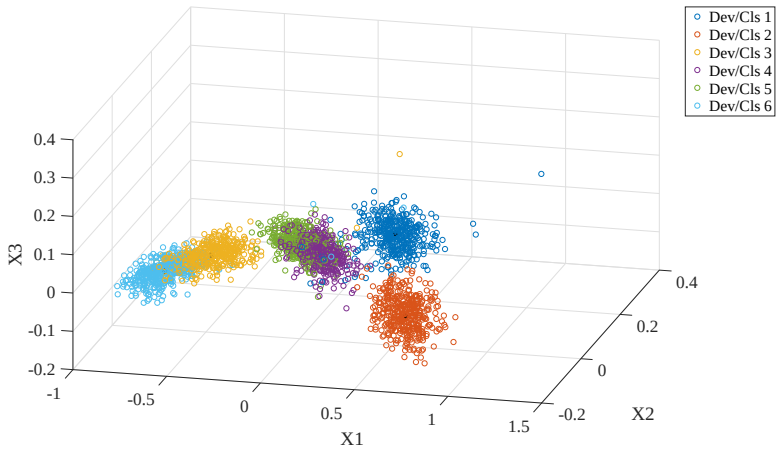

Figure 24. 3D View of Fisher Space for M=6 Class Time Domain Fingerprints, SNR = 35 dB

Figure 25 shows the 3D Fisher Plot of the six class MDA/ML classification in the Spectral Domain. Device 1, 4, and 5 have similar distributions. Device 2 is easily classifiable. The distributions of Device 3 and 6 also intersect. This finding again

supports the hypothesis that device 2 will be easily distinguishable likely due to its age, however the hypothesis was not supported for Device 1.

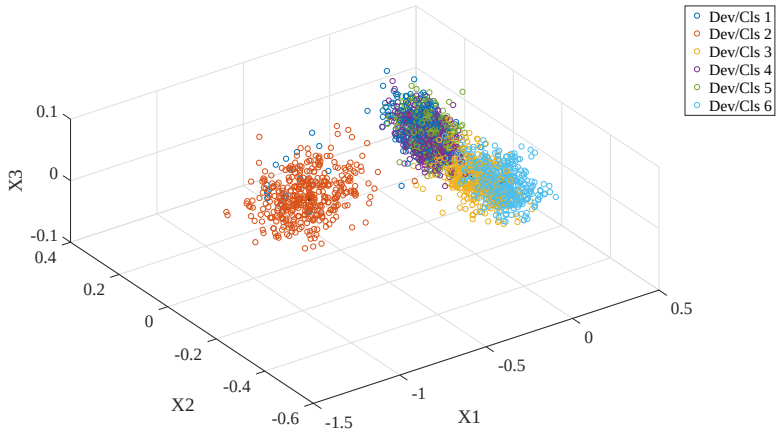

Figure 25. 3D View of Fisher Space for M=6 Class Spectral Domain Fingerprints, SNR = 35 dB

Verification of MDAML Model.

For Device ID Verification, the MDA/ML process was run for the fingerprints of every device except for Device 3. This created a five class model that was never trained on Device 3 characteristics. Then the Device ID verification process was completed using fingerprints from all six devices, meaning that Device 3 was treated as the "rogue" device. Figure 26 shows the rogue accept rate (RAR) vs TVR for the spectral domain. Varying the SNR has little effect of the results due to relatively small change in classification performance, as was demonstrated in Fig. 28.

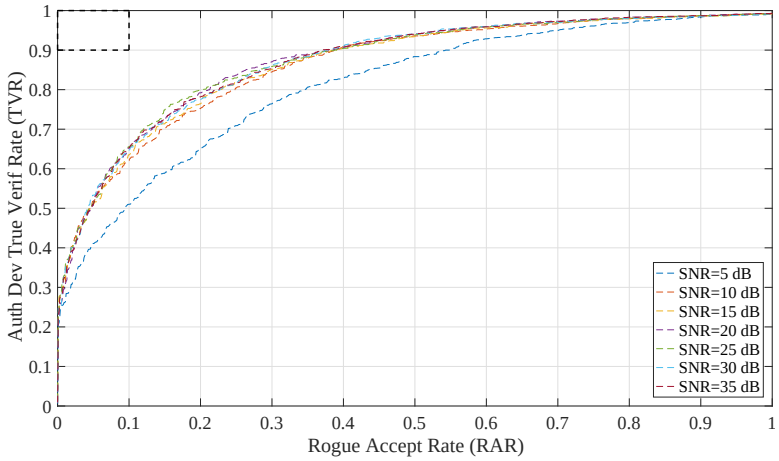

Figure 26. Rogue Accept Rate vs True Verification Rate for Spectral Domain Fingerprints, Device 3 Simulated Rogue Device

Figure 27 shows the RAR vs TVR for the Time domain. The verification process reveals good performance at high SNR's, however as SNR decreases, the performance decreases to where the EER is [.3 .7], meaning that 70% of valid fingerprints have to be rejected in order for only 30% of rogue fingerprints to accepted. In the case of the experiment, this means that 8% of all accepted fingerprints are from a rogue device.

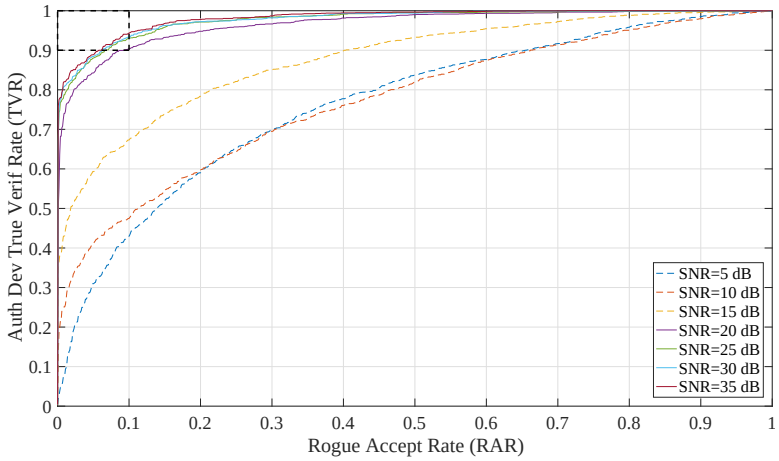

Figure 27. Rogue Accept Rate vs True Verification Rate for Time Domain Fingerprints, Device 3 Simulated Rogue Device

CTSD Fingerprint Performance.

Figure 28 shows the average percent correct classification of MDA/ML utilizing the fingerprints created in TD, SD, and CTSD. Of note is that when the fingerprints are combined together, the overall performance increases. While the SD has overall poorer performance at higher SNR's, it performs better at lower SNR's. The CTSD fingerprints results in a better average percent correct classification, with a $\approx 15\%$ increase at an SNR of 5 dB.

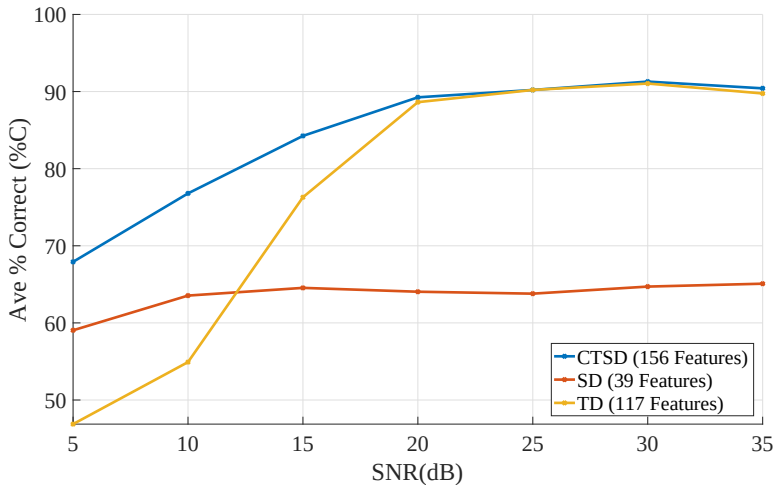

Figure 28. Average Percent Correct Classification for TD, SD, and CTSD Fingerprints

Figure 29 shows the 3D Fisher Plot of the six class MDA/ML classification when the Time and Spectral Domain fingerprints are applied. The Fisher Plot looks similar to the Time Domain plot because as the SNR increases, TD fingerprints are able to more effectively distinguish between devices.

Figure 30 shows the RAR vs TVR for the Time and Spectral Domain. The process performance is similar at SNR's to that of the Time Domain, however it performs better at low SNR than than either the Spectral or Time domain process. For example, at 10 dB, the percentage of accepted fingerprints that are from a rogue device is $\approx 4.5\%$, compared to $\approx 5\%$ for spectral domain and $\approx 8\%$ for time domain.

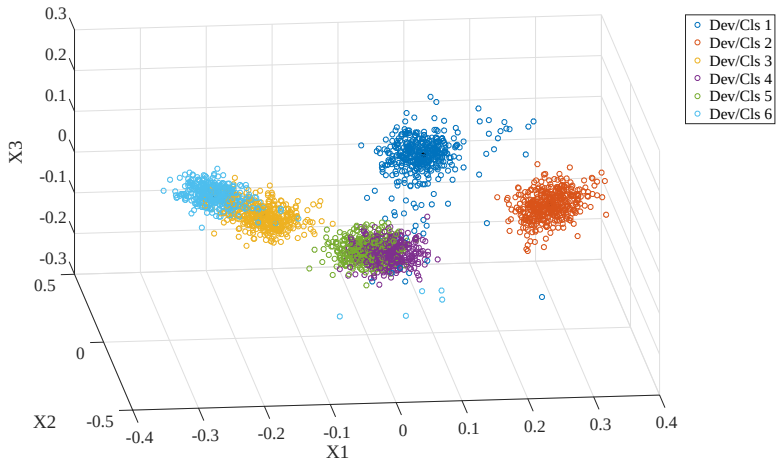

Figure 29. 3D View of Fisher Space for M=6 Class CTSD Fingerprints, SNR = 35 dB

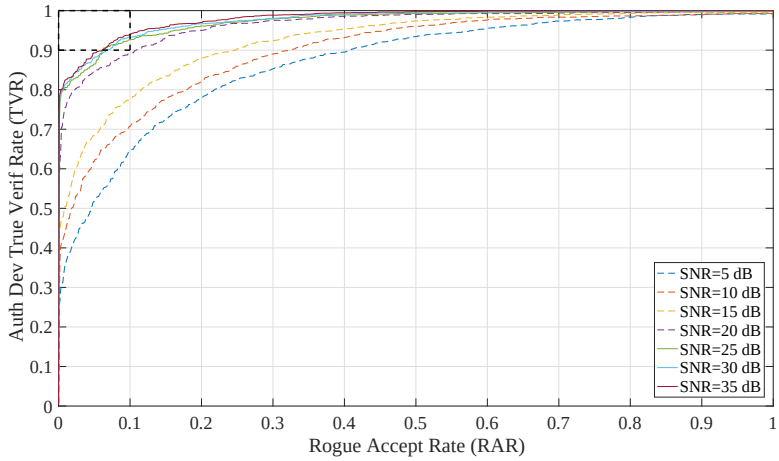

Figure 30. Rogue Accept Rate vs True Verification Rate for CTSD Fingerprints, Device 3 Simulated Rogue Device

Table 7 shows the confusion matrix for SD fingerprints, and demonstrates the highest confusion between devices. The diagonal element is highlighted for easier reference. Device 4 and 5 have the highest level of confusion between classes, followed by Device 3 and 6. The high confusion is evident by the large off-diagonal components. In the experiment, Device 2 had 100% correct classification, and Device 1 was an outlier with the poorest classification performance. Table 8 shows the confusion matrix output for the TD Fingerprinting process, and performs better than the SD Fingerprinting process. The highest confusion was still between Devices 4 and 5 as well as between Devices 3 and 6. The main difference between TD and SD Fingerprinting was the large difference in classification performance for Device 1. The classifier increased classification performance for Device 1 by over 80% when using TD fingerprints.

Table 9 shows the confusion matrix output for the CTSD Fingerprinting process. The classification performance of CTSD fingerprints was comparable to TD fingerprints with a slight increase in performance when classifying Device 4 and 5.

Table 7. Spectral Domain Confusion Matrix for SNR=35, 800 Fingerprints

		Called Class					
		1	2	3	4	5	6
Input Class	1	15.75	3.25	2.25	34	44.75	0
	2	0	100	0	0	0	0
	3	1.75	0	71.75	5	2.25	19.25
	4	6.5	0	5.25	57.5	30.75	0
	5	9	0	2.25	27.5	61.25	0
	6	0	0	15.75	0	0	84.25

Table 8. Time Domain Confusion Matrix for SNR=35, 800 Fingerprints

		Called Class					
		1	2	3	4	5	6
Input Class	1	96	0.5	0	3	0.5	0
	2	0	99.75	0	0.25	0	0
	3	0	0	88.25	0	0	11.75
	4	0.25	0.25	0	82	17.5	0
	5	0	0	0.25	18.25	81.5	0
	6	0	0.25	7.75	1	0	91

Table 9. CTSD Confusion Matrix for SNR=35, 800 Fingerprints

		Called Class					
		1	2	3	4	5	6
Input Class	1	96	0.5	0	2.75	0.75	0
	2	0	100	0	0	0	0
	3	0	0	89.25	0	0	10.75
	4	0.25	0	0	84.75	15	0
	5	0	0	0	17.75	82.25	0
	6	0	0	8.5	1.25	0	90.25

4.3 Sampling Rate

Figure 31 summarizes the results of the sampling rate experiment. It depicts the average percent correct classification across all four devices as a function of the decimation factor applied at the bursts at the beginning of the process. A decimation factor of two corresponds to a sampling frequency of 390 kHz, while the decimation factor of 60 corresponds to a sampling frequency of 13 kHz. All sampling rates are above the Nyquist rate.

SD fingerprints show no large difference in classification performance due to a change in the decimation factor. However, the performance of the TD fingerprints decrease by 20%. CTSD fingerprints decrease by $\approx$ 7%, however the performance still exceeds that of TD and SD.

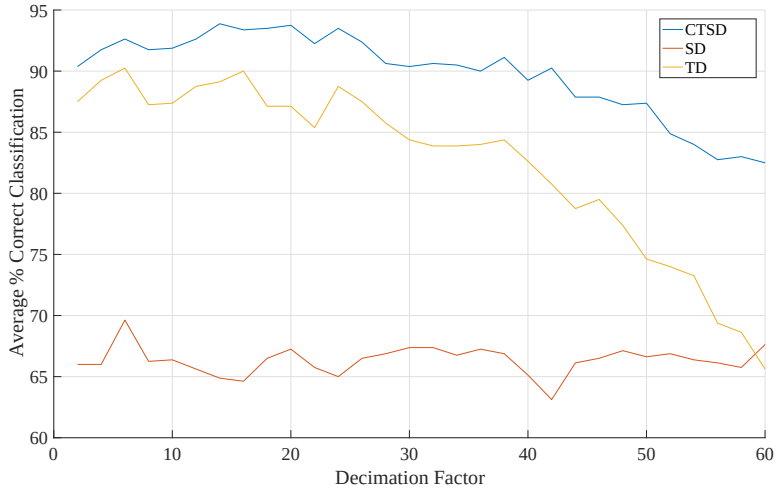

Figure 31. Proper Decimation Factor vs. Average % Correct Classification

4.4 Summary

RF-DNA Fingerprinting techniques are effective at device classification and discrimination for satellite systems, and has the potential to complement existing security measures. From the research work done, an average percent correct classification

could be achieved that adds to security without negatively impacting by a large amount the ability for authorized users to communicate with the satellite.

Time Domain and Spectral Domain Fingerprinting are both valid methods to use for classification of signals, however CTSD fingerprints allows for the classifier to have more information at its disposal in order to create the best classification model to discriminate between devices.

The research results indicate sampling rate seems to affect the performance of TD fingerprints more than SD fingerprints, and CTSD fingerprints outperforms both original methods.

V. Conclusions and Future Work

5.1 Overview

The purpose of this chapter is to summarize the research objectives, approach, and results, as well as to list recommendations for future research.

5.2 Conclusions

Research Objectives.

The research objectives fit into three categories:

1. Can RF-DNA Fingerprinting techniques be utilized to correctly identify and classify the signals coming from six different communication ground-stations, and what is the potential effectiveness of applying the process to a larger scale SATCOM communication system?

2. Can classification performance improve through the use of combined time and spectral domain fingerprints when compared to classifying the two domains separately?

3. What is the relationship between the proportion of oversampling and classification performance?

Methodology.

The research analyzed the research objectives through a process summarized in four essential steps:

1. Collect satellite communication signals from a ground station.

2. Perform pulse detection and post-processing on the collected signals.

3. Fingerprint pulses in spectral domain and time domain.

4. Perform Device Classification and Verification.

Results.

The initial device classification went against the hypothesis by having $< 90\%$ Average Correct Classification at low SNRs (Signal-to-Noise Ratio). This research investigated several solutions, however normalizing each pulse in the spectral domain based on the total power of the region of interest is ultimately recommended. While this approach decreases overall performance, it eliminates the power dependency yielding a process which is more applicable to satellite communications. RF-DNA is an effective at performing signal classification and verification for the six ground station configurations.

The combination of Time and Spectral Domain fingerprints results in improved classification performance at lower SNRs by as much as 15%, and should be analyzed more in-depth for future research.

Sampling rate affects the classification performance of the time domain more apparently than the spectral domain. The susceptibility of Time Domain fingerprint performance to lower sampling rates compared to the consistency of Spectral Domain fingerprint performance across all sampling rates above the Nyquist rate indicates that spectral domain might yield high classification performance.

5.3 Recommendations for Future Research

1. Significant modification has been performed to the RF-DNA code in order to improve usability. However, further streamlining of code should occur. Additional clarification and improvement of the code will allow for future AFIT research to focus more time on new research with RF-DNA, and not on attempting to

understand code put together from multiple thesis.

2. While lessons can be learned from lab experiments, many variables might not be properly considered until collections occur outside of a lab environments. Real world SATCOM collections, especially of satellite down link data, should occur in order to enhance understanding of RF-DNA applicability to satellite systems.

3. While this research introduced the concept of CTSD fingerprints, further analysis of the combination of the time domain and spectral domain. Specifically, an evaluation of CTSD fingerprinting effectiveness across a broad range of signals should be performed. Some applications might reveal a lower performance with CTSD than with TD or SD fingerprints.

4. Similar to further analysis needing to be done with CTSD fingerprints, the impact of classification performance due to sampling rate should be further analyzed.

VI. Appendix A: User's Guide to RF-DNA

File Organization.

The MATLAB® and data files are organized according to the folders shown in Fig 32.

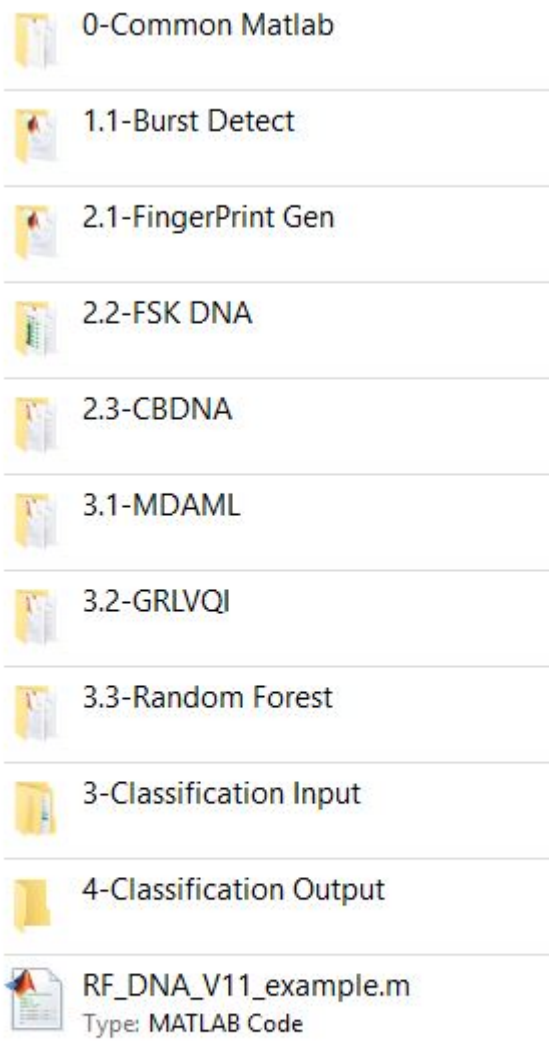

Figure 32. MATLAB File Organization

1. 0-Common Matlab contains MATLAB® .m files that are used by multiple other files in the other folders.

2. 1.1-Burst Detect contains the .m file for Burst Detection.

3. 1-Burst Detect Input contains the raw .mat files used for Burst Detection, separated into sub-folders.

4. 2.1-FingerPrint Gen contains the .m file for Fingerprint Gen using Time Domain, Spectral Domain, or Gabor Transform analysis.

5. 2.2-FSK DNA contains the .m file for Fingerprint Generation using slope-based FSK analysis.

6. 2-FingerPrint Gen Input contains the raw .mat files outputted by Burst Detection and used for Fingerprint Generation, separated into sub-folders.

7. 3.1-MDAML contains the .m file for classification using MDA/ML techniques.

8. 3.2-GRLVQI contains the .m file for classification using (Generalized Relevance Learning Vector Quantized Improved) GRLVQI. This has not been revised or updated since V9.

9. 3.3-Random Forest

10. 3-Classification Input contains the raw .mat files outputted by Fingerprint Generation and used by the Classifiers, separated into sub-folders.

11. 4-Classification Output contains the raw .mat files outputted by the Classifiers, separated into sub-folders.

Each portion of the code has three main sections within it:

1. Flag Declaration.

2. Parameter Declaration.

3. Main Code.

Version 11 of the RF-DNA is written to be used in three different ways.

1. If a user is unfamiliar with coding with MATLAB, or is getting an introduction to the RF-DNA process, the programs are designed to be run completely from GUI interfaces.

2. Once a user has become more familiar with how to use the program, the user can start to modify the first two sections of the code, Flag Declaration and Parameter Declaration. The values can be modified in the actual .m code. The main RFDNA.m file can be used to efficiently run multiple iterations of the code.

3. An experienced MATLAB programmer who spends a fair amount of time understanding the code might need to modify the main code section. This is especially applicable when adding new features or capabilities. Care should be taken to keep with the same code format when adding/modifying code.

Flags and Parameters.

The Burst Detection Flags are: The Burst Detection Parameters are:

Table 11. Burst Detect Parameters

Parameter Name	**Parameter Explanation**
LoadFileName	File name of desired load file
SaveFileName	Desired save file name
LoadExperimentSubfolder	Subfolder that load files reside in
SaveExperimentSubfolder	Subfolder that files are desired to be saved in
SaveFileName	Desired save file name
MaxDetPulses	Maximum Pulses to calculate before stopping

Parameter Name	**Parameter Explanation**
BurstMin	Minimum Number of Samples for a Single Pulse
BurstMax	Maximum Number of Samples for a Single Pulse
PulseSize	Size of First Pulse, only used if MultiplePulsesInBurst Flag is selected
AddSampLead	Amount of Samples to add to leading of pulse
AddSampTrail	Amount of Samples to add to trailing edge of pulse
NSmth	Number of samples to smooth
TrailThresh	Trailing threshhold for pulse detection
LeadThresh	Leading threshhold for pulse detection
DetFlrDb	Desired Floor of Detection in dB
PSDFlr	Floor of PSD (all lower values are not plotted)
FrqEstLowDex	Minimum Value in range of estimated frequency
FrqEstHghDex	Maximum Value in range of estimated frequency
ChSpace	Channel Seperation in Hz
EstBrstFrqHz	Estimated Burst Frequency
PreDetFiltBWHz	Bandwidth of baseband Filter in Hz
PreDetFiltOrd	Baseband Filter Order
PreDetFiltCntrFrqHz	Baseband Filter Center Frequency
PostDetFiltBWHz	Bandwidth of bandpass Filter in Hz
PostDetFiltOrd	Bandpass Filter Order
XDelta	Time between samples
DecFactVec	Vector of Desired Decimation Factors
PercentOfSignalToView	Percent of initial file to view when selecting filter characteristics
FiltImplseRespLngth	Impulse Response Length for Baseband and Bandpass Filters
FiltPercPad	Percent Padding of Filter

Table 10. Pulse Detection Flags

DisplayPlotsDuringPulseDetect	No
	Yes - Only Currently Detected Burst Plot
	Yes - Summary Plots
DisplayPlotsAtEnd	Do not plot at End
	Plot at End
SNREstimation	No SNR Estimation
	SNR Estimated
DetectBurstSelection	No Pulse Detection
	Run Pulse Detection
LeadingorTrailingEdgAlgn	Trailing Edge
	Leading Edge
TypeOfBurstFineAln	No Fine Burst Alignment
	Correlation-Based Fine Burst Alignment
	Peak-Based Fine Burst Alignment
	Correlation-Based and Peak-Based Fine Burst Alignment
PreDetectFilter	No PreDet Filter Applied
	PreDet Filter Applied, No User Validation needed
	PreDet Filter Applied, User Validation Needed
PostDetectFilter	No PostDet Filter Applied
	PostDet Filter Applied, No User Validation needed
	PostDet Filter Applied, User Validation Needed
DownConversion	No Down-Conversion
	DC EACH Burst Using an inputted Freq
	DC EACH Burst Using Its Own IntraBurstAveFrq
	D/C EACH Burst Using Channel InterBurstAveFrq
DisplayWaitBars	Do not display Wait Bars
	Display Wait Bars
MultiplePulsesInBurst	No Seperate Pulses in Burst
	Seperate Pulses in Burst
NPsdPlt	No Plotting
	Plot during SigPSD
LoadAllFilesInSubfolder	No, Input specific Files
	Yes, All Files Contained In LoadExperiment Subfolder
UserValidationofParameters	No Validation, figure will not appear
	Figure with all parameters will show up for user to change

The Fingerprint Generation Flags are:

Table 12. FingerPrint Generation Flags

Flag Name	Possible Flag Values
PrntRgn	Preamble Rgn
	SigID Rgn
BBFilt	Do NOT Apply BB Filter
	Apply BB Filter and Params
	Apply BB Filter and Params and Request User Validation
BPFilt	Do NOT Apply BP Filter
	Apply BP Filter and Params
	Apply BP Filter and Params and Request User Validation
DnConv	No Down-Conversion
	Apply Down-Conversion: All Puls by SAME Input/Set Frq - Fixed D/C
	Apply Down-Conversion: All Puls by SAME EST Frq - Est "Channel" D/C
	Apply Down-Conversion: All Puls by OWN EST Frq - Ext Puls-by-Puls D/C
ValidateROI	Do not validate ROI
	User Input, Validate ROI
ToTSig	Do Not Include Total Signal Region
	Include Total Signal Region
NPsdPlt	No Plotting
	Plot during SigPSD
UserValidationofParameters	No Validation, figure will not appear
	Figure with all parameters will show up for user to change
CalculateNoiseBasedOffAverage	Calculate Noise Based on each pulses Signal Power
	Calculate Noise Based on Average Signal Power
SetRandomStream	Do not reset random stream every SNR loop
	Reset Random Stream for experiment repeatibility. (should not be done if the data is put through code in subsets)
Std	Do not include Std Dev Metric
	Include Std Dev Metric
Var	Do not include Variance Dev Metric
	Include Variance Dev Metric
Skw	Do not include Skewness Dev Metric
	Include Skewness Dev Metric
Kur	Do not include Kurtosis Dev Metric
	Include Kurtosis Dev Metric
PntPlot	StatDNAPrints Plots off
	StatDNAPrints Plots on

Flag Name	Possible Flag Values
NormalizeFeat	Features Un-Normalized
	Features Normalized
PSDPlot	PSDFeatures Plots off
	PSDFeatures Plots on
SpecDomROI	Do not validate ROI
	User Input, Validate ROI
SDFingerprintsIndB	Calculate Fingerprints while not in dB
	Calculate Fingerprints in dB
IncludeInstAmp	Do not Include Inst Amplitude
	Include Inst Amplitude
IncludeInstPhz	Do not Include Inst Phase
	Include Inst Phase
IncludeInstFrq	Do not Include Inst Frequency
	Include Inst Frequency
FtrPlot	InstFeatures Plots off
	InstFeatures Plots on
FreqCntlrType	DO NOT REMOVE: This effectively simulates the effect of a receiver architecture that uses a FIXED local oscillator frequency for down-converting ALL collected pulses.
	REMOVE: This effectively simulates the effect of a receiver architecture that uses a VARIABLE local oscillator frequency (determined on a pulse-by-pulse basis) for down-converting EACH individual pulse.
	REMOVE: This is intended to remove the effects of CROSS-COLLECTION biases, i.e., signals collected using a different oscillator frequency.
GaborPrintFeatPlot	GaborPrintFeat Plots off
	GaborPrintFeat Plots on
GaborWaitBar	GaborPrintFeat Waitbar off
	GaborPrintFeat Waitbar on
GTNorm	No Normalization
	MAth
	Math (Reising ICC) (MISSING IN CODE?)
GaborXDisplay	GaborXForm Display Output off
	GaborXForm Display Output on

The Fingerprint Generation Parameters are:

Table 13. FingerPrint Generation Parameters

Parameter Name	Possible Parameter Explanation
LoadFileName	File name of desired load file
SaveFileName	Desired save file name
LoadExperimentSubfolder	Subfolder that load files reside in
SaveExperimentSubfolder	Subfolder that files are desired to be saved in
GTParamsFileName	File Name of Gabor Transform Parameters
PtchParamsFileName	File name where Gabor Transform Patch Parameters are saved
TstDevSel	Device desired for visualization
TstSigSel	Signal to use for plotting and verification
SNRin	Values for each desired SNR
DecFactVec	Vector of Desired Decimation Factors
MaxBurstsToCalculate	Maximum Pulses to calculate before stopping
NumParallelPools	Number of Parallel Pools to Start
TimeParallelPools	Time until Parallel Pool closes
MinRoiSampNum	Minimum Sample Value for ROI Selection
MaxRoiSampNum	Maximum Sample Value for ROI Selection
SubRgnsNum	Number of Subregions desired
MonteCarloNoiseRealizations	Number of Monte Carlo Noise Realizations
RoiLbl	Label to attach to region of interest selected
SubRgnsNum	Number of Subregions desired
MonteCarloNoiseRealizations	Number of Monte Carlo Noise Realizations
ProcAppLbl	Label to attach to the save file, typically filled in by the process
PSDFlr	Floor of PSD (all lower values are not plotted)
BPFilterBWHz	Bandwidth of bandpass Filter in Hz
BPFiltOrd	Bandpass Filter Order
BPCntrFrqHz	Baseband Filter Center Frequency
InitialDnCnvFrq	Frequency for user inputted down conversion
BBFilterBWHz	Bandwidth of baseband Filter in Hz
BBFilterOrd	Baseband Filter Order
NumImpZer	Length of Impulse Response
FiltPercPad	Percent Padding of Filter

The Classify Flags are:

Table 14. Classification Flags

Flag Name	Possible Flag Values
UserValidationofParameters	No Validation, figure will not appear
	Figure with all parameters will show up for user to change
ClfyType	Linear
	Quadratic
KeepPosteriors	Posterior probability models generated during classification for subsequent verification not kept
	Keep the posterior probability models generated during classification for subsequent verification
PrntAssgn	Contiguous Fingerprint Order
	Interleaved Fingerprint Order
	Scrambled RandPerm Fingerprint Order
TestOnlyClassification	No, divide single file into training and testing
	Yes, have all Input Fingerprints go to testing
DisplayPlots	No
	Yes - Plots Generated
FisherPlotDimensions	Do not plot Fisher Plane Projection
	Plot 2D Fisher Plane Projection
	Plot 3D Fisher Plane Projection
SelectiveModDevIndx	Use All Available Input Cls/Devs
	Select Specific Devs/Cls Indexes for Model Dev
SelectiveNz	Use all Monte Carlo Noise Realizations
	Select Monte Carlo Noise Realizations Subset
SelectiveSNR	Use all SNRs
	Select specific SNRs
NDRA	Run Full-Dimensional Fingerprints
	Run DRA Selected Fingerprints
DRASubRgn	No
	Yes - SubRegion Selection
	Yes - Load From File
SNRDependentWMatrix	Use SNR Dependent Models for TESTING
	Use a Single SNR Model to Select Proj Matrix W

The Classify Parameters are:

Table 15. Classification Parameters

Parameter Name	Possible Parameter Explanation
InputFileName	Input File Name
LoadExperimentSubfolder	Subfolder that load files reside in
PreviousResultsFileName	File name where previous results are saved
SaveFileName	Desired save file name
SaveExperimentSubfolder	Subfolder that files are desired to be saved in
ModDevIndx	Device Indexes for selective model development
TstModSel	SNR model to use for testing
KFold	Value of number of folds in the K-fold loop
NNzUsed	Number of noise realizations to use for classification
SNRDex	Specific Indexes of desired SNR's
DRAFileName	File Name if loading a DRA file
NDRASubRgns	Subregion Indexes for DRA
NDRADex	Fingerprint Indexes for DRA
NSubRgns	Number of Subregions in fingerprints
NTotRgn	Inclusion of Total Region as a subregion
NResp	Applicable only to TD fingerprints, number of responses used in fingerprint generation (Amp, Phz, Frq)
NStats	Number of statistics in fingerprints

Figure 33 shows an example of the graphical user interface for flag selection, while Fig. 34 shows the example parameter selection graphical user interface. Due to the complexity of the Fingerprint Generation process, an additional prompt needed to be added for the user determining what fingerprinting method is desired, as shown in Fig. 35.

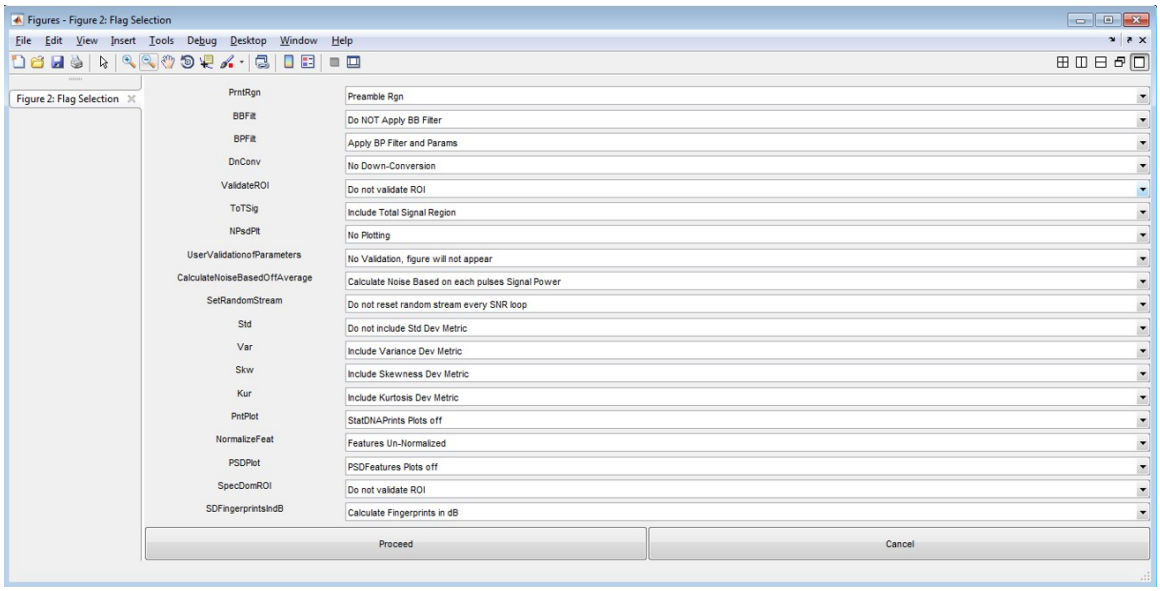

Figure 33. Flag Selection

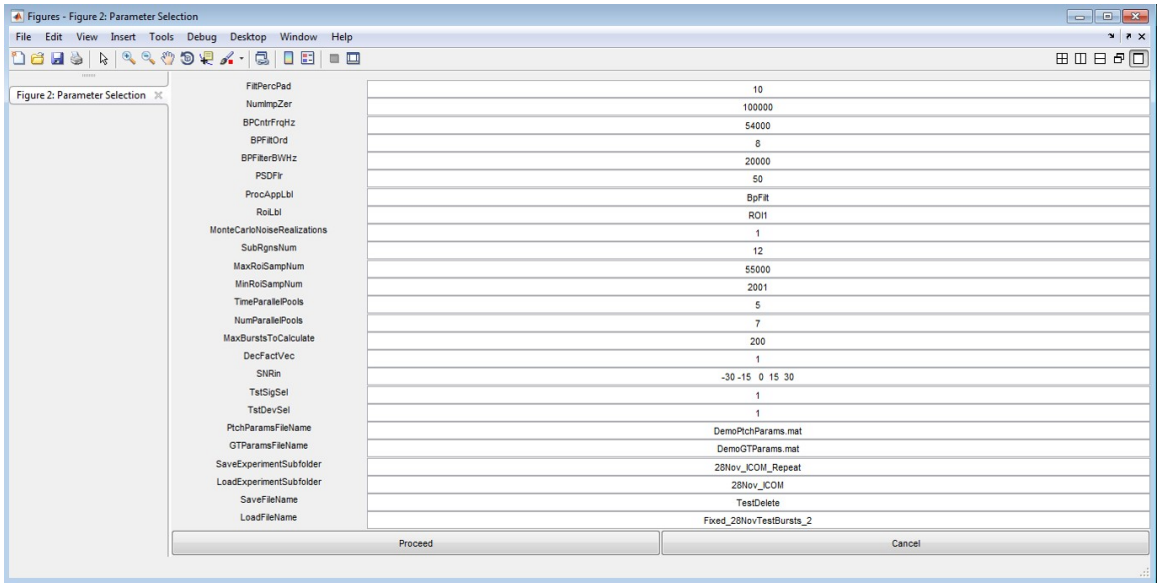

Figure 34. Parameter Selection

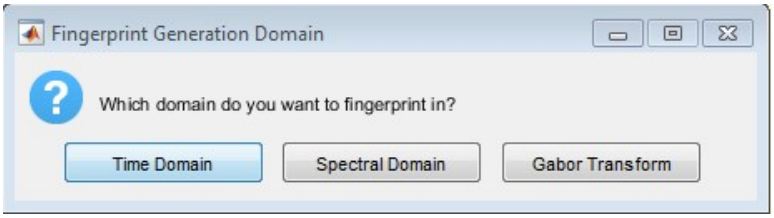

Figure 35. Method Selection

Bibliography

1. S. Nicholas, "Gen. Ellen Pawlikowski: Cyber Resiliency Steering Group Unveils Air Force Cyber Campaign Plan," *Executive Mosaic*, September 2016.

2. C. Gross, "AFMC commander says cyber threats are real, need to get ahead of them," *Air Force News Service*, September 2016.

3. J. Robertson, "Heartbleed hackers steal encryption keys in threat test," *Bloomberg Technology*, April 2014.

4. M. B. Booth, "Verification of Commercial SATCOM Device Identities Using Radio Frequency-Distinct Native Attributes (RF-DNA)," Master's thesis, AFIT-ENG-13-M-09. Air Force Institute of Technology, March 2013.

5. A. B. Tao, "Radio frequency distinct native attribute (RF-DNA) fingerprinting applied to SATCOM short burst data modems using a software-defined radio," Master's thesis, AFIT-ENG-MS-15-M-043. Air Force Institute of Technology, March 2015.

6. J. D. McGuire, "Radio frequency distict native attribute fingerprinting applied to commercial SATCOM short burst data modems," Master's thesis, AFIT-ENG-14- M-51. Air Force Institute of Technology, March 2014.

7. T. J. Carbino, *Exploitation of Unintentional Ethernet Cable Emissions Using Constellation Based-Distinct Native Attribute (CB-DNA) Fingerprints to Enhance Network Security*. PhD thesis, AFIT-ENG-DS-15-S-008. Air Force Institute of Technology, September 2015.

8. W. C. Suski II, M. A. Temple, M. J. Mendenhall, and R. F. Mills, "Radio frequency fingerprinting commercial communication devices to enhance electronic security," *Int. J. Electron. Secur. Digit. Forensic*, vol. 1, pp. 301–322, Oct. 2008.

9. R. Klein, M. A. Temple, M. J. Mendenhall, and D. R. Reising, "Sensitivity analysis of burst detection and RF fingerprinting classification performance," in *2009 IEEE International Conference on Communications*, pp. 1–5, June 2009.

10. D. R. Reising, *Exploitation of RF-DNA for Device Classification and Verification Using GRLVQI Processing*. PhD thesis, AFIT-ENG-DS-12-04. Air Force Institute of Technology, December 2012.

11. W. C. S. II, M. A. Temple, M. J. Mendenhall, and R. F. Mills, "Using spectral fingerprints to improve wireless network security," in *IEEE GLOBECOM 2008 - 2008 IEEE Global Telecommunications Conference*, pp. 1–5, Nov 2008.

12. 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*, pp. 32–39, Sept 2010.

13. 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*, pp. 1–6, Dec 2010.

14. T. J. Carbino, M. A. Temple, and J. Lopez, *A Comparison of PHY-Based Fingerprinting Methods Used to Enhance Network Access Control*, pp. 204–217. Cham: Springer International Publishing, 2015.

15. B. W. Ramsey, T. D. Stubbs, B. E. Mullins, M. A. Temple, and M. A. Buckner, "Wireless infrastructure protection using low-cost radio frequency fingerprinting receivers," *International Journal of Critical Infrastructure Protection*, vol. 8, pp. 27 – 39, 2015.

16. R. W. Klein, *Application of Dual-Tree Complex Wavelet Transforms to Burst Detection and RF Fingerprint Classification*. PhD thesis, AFIT/DEE/ENG/09-12. Air Force Institute of Technology, September 2009.

17. B. W. Ramsey, *Improved Wireless Security through Physical Layer Protocol Manipulation and Radio Frequency Fingerprinting*. PhD thesis, AFIT-ENG-DS-14-S-10. Air Force Institute of Technology, September 2014.

18. K. J. Kwak, E. van Doom, S. Stanic, and R. Levy, "NetSAID: Network Sensing, Assessment, and Intrusion Detection for Portable SDR System," tech. rep., Intelligent Automation, Inc., May 2013.

19. C. K. Dubendorfer, "Using RF-DNA Fingerprints to Discriminate ZIGBEE Devices in an Operational Environment," Master's thesis, AFIT-ENG-13-M-15. Air Force Institute of Technology, March 2013.

20. S. A. Munns, "Spectral Domain RF Fingerprinting for 802.11 Wireless Devices," Master's thesis, AFIT/GE/ENG/10-19. Air Force Institute of Technology, March 2010.

21. M. D. Williams, "Application of RF-DNA Fingerprinting to Improve WiMAX Security," Master's thesis, AFIT/GE/ENG/11-41. Air Force Institute of Technology, March 2011.

22. B. C. Wright, "PLC Hardware Discrimination using RF-DNA fingerprintin," Master's thesis, AFIT-ENG-T-14-J-12. Air Force Institute of Technology, June 2014.

23. 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*, pp. 2168–2173, Oct 2010.

24. W. E. Cobb, E. D. Laspe, R. O. Baldwin, M. A. Temple, and Y. C. Kim, "Intrinsic physical-layer authentication of integrated circuits," *IEEE Transactions on Information Forensics and Security*, vol. 7, pp. 14–24, Feb 2012.

25. 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)*, pp. 7–13, Jan 2012.

26. 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, pp. 1180–1192, June 2015.

27. F. E. Garrido, "OFDM-Based Signal Explotation Using Quadrature Mirror Filter Bank (QMFB) Processing," Master's thesis, AFIT/GE/ENG/12-16. Air Force Institute of Technology, March 2012.

28. V. Brik, S. Banerjee, M. Gruteser, and S. Oh, "Wireless device identification with radiometric signatures," in *Proceedings of the 14th ACM International Conference on Mobile Computing and Networking*, MobiCom '08, (New York, NY, USA), pp. 116–127, ACM, 2008.

29. B. Danev, H. Luecken, S. Capkun, and K. El Defrawy, "Attacks on physical-layer identification," in *Proceedings of the Third ACM Conference on Wireless Network Security*, WiSec '10, (New York, NY, USA), pp. 89–98, ACM, 2010.

30. B. Danev, D. Zanetti, and S. Capkun, "On physical-layer identification of wireless devices," *ACM Comput. Surv.*, vol. 45, pp. 6:1–6:29, Dec. 2012.

31. M. Edman and B. Yener, "Active attacks against modulation-based radio-metric identification," Tech. Rep. 09-02, Rensselaer Institute of Technology, 2009.

32. Y. Huang and H. Zheng, "Radio frequency fingerprinting based on the constellation errors," in *2012 18th Asia-Pacific Conference on Communications (APCC)*, pp. 900–905, Oct 2012.

33. T. J. Carbino, M. A. Temple, and T. J. Bihl, "Ethernet card discrimination using unintentional cable emissions and constellation-based fingerprinting," in *2015 International Conference on Computing, Networking and Communications (ICNC)*, pp. 369–373, Feb 2015.

34. R. M. Gerdes, T. E. Daniels, M. Mina, and S. F. Russell, "Device identification via analog signal fingerprinting: A matched filter approach," in *144 Proceedings of the Network and Distributed System Security Symposium (NDSS*, p. 78, 2006.

35. R. M. Gerdes, M. Mina, S. F. Russell, and T. E. Daniels, "Physical-layer identification of wired ethernet devices," *IEEE Transactions on Information Forensics and Security*, vol. 7, pp. 1339–1353, Aug 2012.

36. P. K. Harmer, M. D. Williams, and M. A. Temple, "Using DE-optimized LFS processing to enhance 4G communication security," in *2011 Proceedings of 20th International Conference on Computer Communications and Networks (ICCCN)*, pp. 1–8, July 2011.

37. W. E. Cobb, *Exploitation of Unintentional Information Leakage from Integrated Circuits*. PhD thesis, AFIT/DEE/ENG/11-06. Air Force Institute of Technology, December 2011.

38. S. J. Stone, *Radio frequency based programmable logic controller anomaly detection*. PhD thesis, AFIT-ENG-DS-13-S-05. Air Force Institute of Technology, September 2013.

39. S. J. Stone, M. A. Temple, and R. O. Baldwin, "RF-based PLC IC design verification," in *2012 DMSMS And Stand Conf*, Aug 2012.

40. T. J. Carbino and R. O. Baldwin, "Side channel analysis of ethernet network cable emissions," in *9th Intl Conf on Cyber Warfare and Security*, 2014.

41. R. J. Gonzales, "Application of Radio Frequency Distinct Native Attribute (RF-DNA) Fingerprinting to Power Substation Emissions," Master's thesis, AFIT-ENG-MS-15-M-034. Air Force Institute of Technology, March 2015.

42. D. V. Atienza, "Characterization of Noise Technology Radar (NTR) Signal Detectability Using a Non-Cooperative Receiver," Master's thesis, AFIT/GE/ENG/11-01. Air Force Institute of Technology, March 2011.

43. A. J. Zeqollari, "Ultra Wideband Radio Frequency Fingerprinting," Master's thesis, AFIT-ENG-14-M-87. Air Force Institute of Technology, March 2014.

44. W. M. Lowder, "Real-Time RF-DNA Fingerprinting of ZigBee Devices Using a Software-Defined Radio with FPGA Processing," Master's thesis, AFIT-ENG-MS-15-M-054. Air Force Institute of Technology, March 2015.

45. H. J. Patel, *Advances in SCA and RF-DNA Fingerprinting Through Enhanced Linear Regression Attacks and Application of Random Forest Classifiers*. PhD thesis, AFIT-ENG-DS-14-S-03. Air Force Institute of Technology, September 2014.

46. A. T. Alt, "Analysis of Multi User Environment Using RF-DNA," Master's thesis, AFIT-ENG-14-M-06. Air Force Institute of Technology, March 2014.

47. C. E. Fillmore, "Computational Electromagnetic Studies for Low-Frequency Compensation of the Reflector Impulse-Radiating Antenna," Master's thesis, AFIT-ENG-MS-15-M-011. Air Force Institute of Technology, March 2015.

48. Y.-L. Du, Y.-H. Lu, and J.-L. Zhang, "Novel method to detect and recover the keystrokes of ps/2 keyboard," *Progress In Electromagnetics Research C*, vol. 41, pp. 151–161, 2013.

49. Y.-L. Du, Y.-H. Lu, J.-L. Zhang, and Q. Cui, "Estimation of eavesdropping distance from conducted emission on network cable of a PC," in *2012 6th Asia-Pacific Conference on Environmental Electromagnetics (CEEM)*, pp. 347–350, Nov 2012.

50. M. Dürmuth, *Novel classes of side channels and covert channels.* PhD thesis, Saarland Univ, Postfach 151141, 66041 Saarbrcken, 2009.

51. R. Frankland, M. Keith, and R. Holloway, "Side channels, compromising emanations and surveillance: Current and future technologies side channels, compromising emanations and surveillance: Current and future technologies," 2010.

52. Z. Hongxin, H. Yuewang, W. Jianxin, L. Yinghua, and Z. Jinling, "Recognition of electro-magnetic leakage information from computer radiation with svm," *Comput. Secur.*, vol. 28, pp. 72–76, Feb. 2009.

53. M. G. Kuhn and R. J. Anderson, *Soft Tempest: Hidden Data Transmission Using Electromagnetic Emanations*, pp. 124–142. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998.

54. Z. Ling, J. Luo, Y. Zhang, M. Yang, X. Fu, and W. Yu, "A novel network delay based side-channel attack: Modeling and defense," in *2012 Proceedings IEEE INFOCOM*, pp. 2390–2398, March 2012.

55. P. Smulders, "The threat of information theft by reception of electromagnetic radiation from RS-232 cables," *Comput. Secur.*, vol. 9, pp. 53–58, Jan. 1990.

56. W. van Eck, "Electromagnetic radiation from video display units: An eavesdropping risk?," *Computers and Security*, vol. 4, pp. 269–286, Dec. 1985.

57. M. Vuagnoux and S. Pasini, "An improved technique to discover compromising electromagnetic emanations," in *2010 IEEE International Symposium on Electromagnetic Compatibility*, pp. 121–126, July 2010.

58. D. P. Montminy, *Enhancing Electromagnetic Side-Channel Analysis in an Operational Environment.* PhD thesis, AFIT-ENG-DS-13-S-01. Air Force Institute of Technology, September 2013.

59. B. W. Ramsey, M. A. Temple, and B. E. Mullins, "PHY foundation for multi-factor ZigBee node authentication," in *2012 IEEE Global Communications Conference (GLOBECOM)*, pp. 795–800, Dec 2012.

60. Kantronics, *KPC-9612 Plus Packet Communicator Product Overview*. Kantronics, 2014. Available at `http://www.kantronics.com/products/kpc9612.html`.

61. Kantronics, *KPC-9612 Plus User's Guide*. Kantronics, 2003.

62. M. Chepponis and P. Karn, *The KISS TNC: A simple Host-to-TNC communications protocol*, 1987.

63. Agilent Technologies, *GSM Design Library: Chapter 9 GSM Design Examples*. Agilent Technologies, 2003.

64. F. Kostedt and J. C. Kemerling, "Practical GMSK data transmission," *MX-COM Inc.*, 1998. Available at `http://sss-mag.com/pdf/gmsk_tut.pdf`.

65. Texas Instruments Inc., "FSK modulation and demodulation with MSP430 microcontroller," tech. rep., Texas Instruments, December 1998.

66. R. F. Mills, "AFIT EENG 571: Satellite communications lesson 4 slides." pg. 26, 1 2016.

67. G. Baudoin, F. Virolleau, O. Venard, and P. Jardin, "Teaching DSP through the practical case study of an FSK modem," in *First European DSP Education and Research Conference*, September 1996.

68. G. Maral, M. Bousquet, and Z. Sun, *Satellite Communications Systems: Systems, Techniques and Technology*. Wiley Series in Communication and Distributed Systems, Wiley, 2011.

69. R. F. Mills, "AFIT EENG 571: Satellite communications lesson 4 slides." pg. 68, 1 2016.

70. The Mathworks, Inc., Natick, Massachusetts, *MATLAB version 9.0.0.341360 (R2016a)*, 2015.

71. 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*, pp. 1–6, Oct 2012.

72. D. R. Reising and M. A. Temple, "WiMAX mobile subscriber verification using gabor-based RF-DNA fingerprints," in *2012 IEEE International Conference on Communications (ICC)*, pp. 1005–1010, June 2012.

73. J. Fogarty, R. S. Baker, and S. E. Hudson, "Case studies in the use of roc curve analysis for sensor-based estimates in human computer interaction," in *Proceedings of Graphics Interface 2005*, GI '05, (School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada), pp. 129–136, Canadian Human-Computer Communications Society, 2005.

74. L. Breimann, "Random forests." Statistics Department University of California, 2001. Available at `https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf`.

75. The Mathworks, Inc., Natick, Massachusetts, *MATLAB Help Files. "treebagger-class"*, 2015.

76. The Mathworks, Inc., Natick, Massachusetts, *MATLAB Help Files. "treebagger"*, 2015.

77. L. Breiman and A. Cutler, "Random forests: The out-of-bag (oob) error estimate." Statistics Department University of California. Available at `https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm#ooberr`.

78. A. Liaw and M. Wiener, "Classification and Regression by randomForest," *R News*, vol. 2, pp. 18–22, December 2002.

79. C. Nguyen, Y. Wang, and H. N. Nguyen, "Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic," in *Journal of Biomedical Science and Engineering*, pp. 551–560, 2013.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
23–03–2017	Master's Thesis	Sept 2015 — Mar 2017

4. TITLE AND SUBTITLE

Application of RF-DNA Fingerprinting Techniques to ICOM Radio Satellite Communication

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Dunkel, Patrick N., Second Lieutenant, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENY-MS-17-M-258

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Intentionally left blank

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A:
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This research addresses if RF-DNA techniques can be utilized to correctly identify and classify the signals coming from six different communication ground-stations. It performs this analysis through the fingerprinting of features in the Time Domain as well as in the Spectral Domain. The classification is performed utilizing MDA/ML and Random Forest Classifiers. Additionally, it introduces a Time/Spectral Domain combined fingerprint set and discusses its potential applicability. Included in the scope of the research is an analysis of the performance of the RF-DNA classification with regards to the amount of oversampling performed on the signal collection. Time/Spectral Domain fingerprints yielded a higher classification performance than the two baseline fingerprinting method. Higher classification performance can be gained through oversampling, however there is a trade-off when considering calculation time and memory. Overall, RF-DNA techniques have high applicability to satellite communications and could complement existing security methods.

15. SUBJECT TERMS

RF-DNA, Satellite Communication, Device Discrimination, MDA/ML

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. Eric D. Swenson, AFIT/ENY
U	U	U	UU	91	19b. TELEPHONE NUMBER *(include area code)* (937) 255-3636, x4555; Eric.Swenson@afit.edu