COMPARISON OF RADIO FREQUENCY DISTINCT NATIVE ATTRIBUTE AND MATCHED FILTERING TECHNIQUES FOR DEVICE DISCRIMINATION AND OPERATION IDENTIFICATION

THESIS

Barron D. Stone, 1st Lt, USAF

AFIT-ENG-MS-16-M-048

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENG-MS-16-M-048

COMPARISON OF RADIO FREQUENCY DISTINCT NATIVE ATTRIBUTE AND MATCHED FILTERING TECHNIQUES FOR DEVICE DISCRIMINATION AND OPERATION IDENTIFICATION

THESIS

Presented to the Faculty

Department of Electrical and Computer 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 Electrical Engineering

Barron D. Stone, B.S.E.E.

1st Lt, USAF

March 2016

AFIT-ENG-MS-16-M-048

COMPARISON OF RADIO FREQUENCY DISTINCT NATIVE ATTRIBUTE AND MATCHED FILTERING TECHNIQUES FOR DEVICE DISCRIMINATION AND OPERATION IDENTIFICATION

THESIS

Barron D. Stone, B.S.E.E.
1st Lt, USAF

Committee Membership:

Maj Samuel J. Stone, PhD
Chair

Capt Timothy J. Carbino, PhD
Member

Dr. Michael A. Temple
Member

AFIT-ENG-MS-16-M-048

Abstract

The research presented here provides a comparison of *classification*, *verification*, and *computational time* for three techniques used to analyze Unintentional Radio-Frequency (RF) Emissions (URE) from semiconductor devices for the purposes of *device discrimination* and *operation identification*. URE from ten MSP430F5529 16-bit microcontrollers were analyzed using: 1) RF Distinct Native Attribute (RF-DNA) fingerprints paired with Multiple Discriminant Analysis/Maximum Likelihood (MDA/ML) classification, 2) RF-DNA fingerprints paired with Generalized Relevance Learning Vector Quantized-Improved (GRLVQI) classification, and 3) Time Domain (TD) signals paired with matched filtering. These techniques were considered for potential applications to detect counterfeit/Trojan hardware infiltrating supply chains and to defend against cyber attacks by monitoring executed operations of embedded systems in critical Supervisory Control And Data Acquisition (SCADA) networks.

RF-DNA with MDA/ML achieved the best *device classification* performance at high analysis Signal-to-Noise Ratio (SNR_A) among $N_{DevAuth} = 8$ "authorized" devices with average percent correct classification of $\%C_{Ave} \geq 75.38\%$ achieved for $SNR_A \geq 20$ dB. RF-DNA with GRLVQI performed second best with $\%C_{Ave} \geq 60.85\%$ achieved for $SNR_A \geq 20$ dB. Matched filtering yielded the lowest performance of $\%C_{Ave} \geq 52.38\%$ achieved for $SNR_A \geq 20$ dB. The performance for *operation classification* among $N_{Op} = 12$ register-to-register operations was lower than the *device classification* performance and included 1) RF-DNA with MDA/ML achieving $\%C_{Ave} \geq 10.73\%$ for $SNR_A \geq 20$ dB, 2) RF-DNA with GRLVQI achieving $\%C_{Ave} \geq 8.62\%$ for $SNR_A \geq 20$ dB, and 3) matched filtering achieving $\%C_{Ave} \geq 11.09\%$ for $SNR_A \geq 20$ dB.

For *device ID verification* with $N_{DevAuth} = 8$ "authentic" and $N_{DevRogue} = 2$ "rogue" devices, both RF-DNA with MDA/ML and matched filtering techniques successfully verified/rejected at least one authentic/rogue device using an Equal Error Rate (EER) benchmark of $EER \leq 10\%$ at $SNR_A = 30$ dB. RF-DNA with GRLVQI did not meet the benchmark in any evaluated *verification* scenarios.

When comparing the *computational time* required to process and classify collected emissions, both RF-DNA techniques took roughly 3 times longer for *device classification* and 42 times longer for *operation identification* compared to matched filtering.

Acknowledgements

I would like to thank my advisor and committee members for their guidance and feedback during this research adventure, my parents for supporting my interest in electronics at a young age, my wife for her love and patience during the late nights I spent writing this thesis, and God for blessing my life with all of the above people.

Barron D. Stone

Table of Contents

Page

Page

List of Figures

List of Tables

List of Abbreviations

Abbreviation		Page
RF	Radio-Frequency	1
RF-DNA	RF Distinct Native Attribute	1
USAF	United States Air Force	1
DOD	Department Of Defense	1
IT	Information Technology	1
IC	Integrated Circuit	1
CAC	Common Access Card	1
AFCERT	Air Force Computer Emergency Response Team	2
US-CERT	United States Computer Emergency Readiness Team	2
SCADA	Supervisory Control And Data Acquisition	2
DARPA	Defense Advanced Research Projects Agency	4
URE	Unintentional RF Emissions	4
MCU	MicroController Unit	4
PLC	Programmable Logic Controller	5
COTS	Commercial Off The Shelf	5
FBI	Federal Bureau of Investigation	5
ICS-CERT	Industrial Control Systems Cyber Emergency Readiness Team	5
OSI	Open Systems Interconnect	6
EM	Electro-Magnetic	7
SCA	Side Channel Analysis	7
DPA	Differential Power Analysis	8

Abbreviation		Page
DUT	Device Under Test	8
MDA/ML	Multiple Discriminant Analysis/Maximum Likelihood	9
GRLVQI	Generalized Relevance Learning Vector Quantized-Improved	9
AFIT	Air Force Institute of Technology	10
CBAD	Correlation-Based Anomaly Detection	10
RFINT	RF INTelligence	11
TD	Time Domain	11
SD	Spectral Domain	11
CD	Correlation Domain	11
IRE	Intentional RF Emissions	11
LFS	Learning From Signals	11
MF	Matched Filtering	11
GND	GrouND	17
SRC	SouRCe	18
DST	DeSTination	18
CHASE	Center for Hardware Assurance, Security, and Engineering	21
ADEC	Advanced Detection of Electronic Counterfeits	21
PFP	Power FingerPrinting	22
EER	Equal Error Rate	22
AES	Advanced Encryption Standard	24
SCARE	SCA for Reverse Engineering	25
SIM	Subscriber Identity Module	25

Abbreviation		Page
GSM	Global System for Mobile	25
DES	Data Encryption Standard	25
PCA	Principal Component Analysis	26
LDA	Linear Discriminant Analysis	26
SCANDALee	Side-ChANnel-based DisAssembLer using Local Electromagnetic Emanations	26
M-LDA	Multiclass Fisher's LDA	26
P-LDA	Polychotomous LDA	26
RISC	Reduced Instruction Set Computing	27
MDA	Multiple Discriminant Analysis	30
ML	Maximum Likelihood	31
ANN	Artificial Neural Network	32
DRA	Dimensional Reduction Analysis	32
DSP	Digital Signal Processing	35
ADC	Analog-to-Digital Converter	36
PMM	Power Management Module	36
UCS	Unified Clock System	36
USB	Universal Serial Bus	36
BSL	BootStrap Loader	36
LCD	Liquid Crystal Display	37
MCLK	Master CLocK	38
REFO	REFerence Oscillator	38
FLL	Frequency Locked Loop	38
DCO	Digitally Controlled Oscillator	38

Abbreviation		Page
IR	InfraRed	38
GPIO	General Purpose Input/Output	41
LPF	Low Pass Filter	43
BPF	Band Pass Filter	47
DDC	Digital Down Conversion	47
HT	Hilbert Transform	47
SNR	Signal-to-Noise Ratio	47
AWGN	Additive White Gaussian Noise	48
ROI	Region Of Interest	49
I	In-phase	51
Q	Quadrature-phase	51
kF-CV	k-Fold Cross-Validation	53
TVR	True Verification Rate	66
FRR	False Rejection Rate	66
FVR	False Verification Rate	66
TRR	True Rejection Rate	66
RAR	Rogue Accept Rate	66
RRR	Rogue Rejection Rate	66
ROC	Receiver Operating Characteristics	67
PMF	Probability Mass Function	67
OS	Operating System	70
ZIF	Zero Insertion Force	100
CB-DNA	Constellation-Based Distinct Native Attribute	101
FPGA	Field-Programmable Gate Array	103

COMPARISON OF RADIO FREQUENCY DISTINCT NATIVE ATTRIBUTE AND MATCHED FILTERING TECHNIQUES FOR DEVICE DISCRIMINATION AND OPERATION IDENTIFICATION

I. Introduction

This chapter introduces the research topic and outlines the motivation for investigating Radio-Frequency (RF) Distinct Native Attribute (RF-DNA) and matched filtering techniques for *device discrimination* and *operation identification*. Section 1.1 provides a brief overview of the operational motivation for this research and is divided into two subsections: 1) Section 1.1.1 describes *hardware*-based security concerns and 2) Section 1.1.2 describes the threat of *software*-based vulnerabilities. Section 1.2 gives a brief overview of existing research and technologies being leveraged by the current research effort. Section 1.3 summarizes how this research effort relates and contributes to the existing research and technologies.

1.1 Operational Motivation

The proliferation of computing technology in recent decades has led to an increased vulnerability to cyber attack in both the private and government sectors. The United States Air Force (USAF) and Department Of Defense (DOD) have become critically dependent on computer hardware and software to carry out nearly every aspect of daily operations. Their digital systems range in size from globally networked Information Technology (IT) resources that manage email and databases to the tiny Integrated Circuit (IC) chips embedded in the Common Access Card (CAC) issued to every DOD employee [51]. While there are many advantages to

using these digital systems, they come with the cost of an increased vulnerability to malicious cyber attacks. The USAF initially responded to the increase in cyber vulnerability by establishing the Air Force Computer Emergency Response Team (AFCERT) as the primary agency responsible for protecting USAF network assets from attack. The monitoring efforts of AFCERT highlighted the magnitude of the cyber threats facing military IT networks, reporting over 150 verified incidents of "hackers" gaining access to USAF information systems in 2011 and nearly 2 million weekly alerts indicating potential cyber attacks against USAF bases [122].

These cyber attacks are not limited to just military and DOD organizations. From 2006 to 2012, the number of cyber incidents reported by federal agencies to the United States Computer Emergency Readiness Team (US-CERT) increased by 782 percent with improper usage, malicious code, and unauthorized access being the most widely reported types of incidents across the federal government [126]. With these findings came increased concern to protect national infrastructures which have become increasingly reliant on computerized systems and electronic data for their operation. A significant challenge to defending these infrastructures is that they are largely owned by private sector organizations and many of the systems performing key functions are proprietary in nature [126]. President Obama responded to the concern in 2013 by issuing Executive Order 13636 with the goal of "improving critical infrastructure cybersecruity" [84].

The following sections describe the types of the *hardware* and *software* vulnerabilities that threaten the computerized Supervisory Control And Data Acquisition (SCADA) systems that are at the heart of critical infrastructures and military applications.

1.1.1 Hardware Vulnerabilities.

The desire for cheaper electronic components has driven most IC manufacturers to outsource the fabrication of semiconductor devices to countries with a lower cost of labor such as China and Taiwan [1]. U.S. companies first moved their assembly, testing, and packing operation to Asia in the 1960s. Later, in the 1980s, they began shifting their fabrication abroad as well to produce semiconductor wafers from designs created in the United States [130]. This growing reliance on foreign suppliers has raised concern regarding the authenticity of hardware devices used in critical infrastructure applications as well as military defense systems [106]. Counterfeit hardware has been discovered in systems ranging from common network routers to high-altitude missile computers [40]. The incidents of counterfeit electronic parts encountered by original component manufacturers more than doubled between 2005 and 2008 [17] and reports of counterfeit parts saw a four fold increase from 2009 to 2011 [54]. The presence of counterfeit hardware in the supply chain is a serious threat to the reliability of systems performing critical functions [106].

In addition to propagating unreliable counterfeit electronics, the outsourcing of microchip production to foreign countries also presents the opportunity for malicious hardware Trojans to be implanted in devices [85]. Complex supply chains like that of the aerospace sector depend on an extensive network of purchasers, subcontractors, suppliers, and partners which create many potential entry points for overseas fabrication facilities to embed unauthorized designs into devices at early stages in the supply chain [130]. Current IC verification practices focus on testing chips to meet a functional specification. While this approach may detect if certain functionality was removed from the design it will likely fail to detect additional covert functionality inserted by an adversary [7].

The ability to identify and authenticate semiconductor devices is important

to prevent counterfeit components and hardware Trojans from entering the supply chain for critical systems. The Defense Advanced Research Projects Agency (DARPA) recently initiated the "TRUST in Integrated Circuits" program to develop technologies capable of verifying the contents of semiconductor components used in military systems that were designed and fabricated under untrusted condition. The TRUST program is pursuing a metrics based approach, formulated in terms of *probability of detection* versus *probability of false alarm*, to identify maliciously altered ICs [9].

This research effort focused on microprocessor ICs which were the second most counterfeited semiconductor device in 2011 as shown in Table 1, accounting for 13.4% of reported incidents [55]. IC manufacturing processes operate within tolerances which allow for small physical variations between individual devices so that no two devices are exactly the same. Those variations can impact the RF energy emitted by devices during operation [14]. The *device discrimination* aspect of this research leveraged those variations in Unintentional RF Emissions (URE) to provide a means of discriminating between individual MicroController Unit (MCU) devices that could be applied to detect the presence of counterfeit or unauthorized Trojan hardware.

Table 1. Top-5 most counterfeited semiconductors in 2011 (percentage of counterfeit part reports) [55].

Rank	Commodity Type	% of Reported Incidents
#1	Analog IC	25.2%
#2	Microprocessor IC	13.4%
#3	Memory IC	13.1%
#4	Programmable Logic IC	8.3%
#5	Transistor	7.6%

1.1.2 Software Vulnerabilities.

Programmable Logic Controllers (PLCs) are at the core of the SCADA systems that control critical functions in military and civilian infrastructures. They are purpose-built machines which are often proprietary in nature and commercially available as Commercial Off The Shelf (COTS) hardware. In recent years, as SCADA networks have grown in complexity and become connected to the open Internet, it has raised security concerns regarding their increased vulnerability to remote cyber attack. In 2011, the Federal Bureau of Investigation (FBI) Cyber Security division revealed that hackers had successfully accessed SCADA systems controlling the infrastructure of three unnamed cities [50]. With that sort of access, malicious hackers could cause significant damage to sewage, power, and water facilities.

One of the most publicized cyber attacks on a SCADA system was the Stuxnet virus which infected the PLCs used to control Iranian nuclear centrifuges [44]. Stuxnet was significant because it was the first cyberweapon created and deployed to cause physical damage with the intent of degrading, disrupting, and destroying a specific information system [35]; however, several other Stuxnet related viruses have since been discovered including Duqu [18, 111], Flame [19], and Shamoon [76]. In January 2015, hackers manipulated the control systems at a German steel mill to cause massive damage when a blast furnace could not be properly shut down - the second cyber attack to cause physical damage [132]. Later that year, a cyber attack using the BlackEnergy Trojan targeted the Ukrainian electric power industry leaving around 700,000 homes in the Ivano-Frankivsk region without power for several hours [75]. The Industrial Control Systems Cyber Emergency Readiness Team (ICS-CERT), which operates side-by-side with US-CERT, was aware of the BlackEnergy Trojan's capability to infect SCADA systems as early as 2014 [53].

Modern worms and viruses that infect SCADA systems, including Stuxnet, may hide their existence by reporting false status to the operator while performing nefarious deeds [131]. Stuxnet also took advantage of several previously unknown "zero-day" vulnerabilities which would likely not have been detected by traditional antivirus techniques which depend on previously known "virus definitions" [112].

The 7-layer Open Systems Interconnect (OSI) model shown in Figure 1 is used to describe the different levels of a networked computer systems [113]. Modern methods for detecting unauthorized activity on informations systems focus on analyzing data within the Application (level 1) and Network (level 5) layers of the OSI model for anomalous program behavior. The execution of that analysis requires additional processing resources beyond normal operation. The embedded MCUs used in PLCs tend to have limited processing resources which prevent them from running onboard cybersecurity processes such as antivirus or intrusion detection [102]. Additionally, SCADA systems are often out-dated by IT standards, being deployed for decades in easily accessible locations with poor physical security where they can be easily tampered with and altered [32].

Attackers are constantly finding new ways to conceal their malicious code from traditional detection methods which typically operate at the Network (level 5) layer and above on the OSI model [97]. Therefore, it is necessary to develop efficient and effective methods of detecting anomalous system behavior to prevent compromised programming or zero-day attacks from entering critical infrastructure or military combat systems [7]. The *operation identification* aspect of this research aims to reverse-engineer the instructions executed on a MCU as a method of passively monitoring a device to validate that it is functioning as expected. These types of monitoring systems are often implemented using low-cost hardware with limited processing power which necessitate a simple algorithm with a low computational cost.

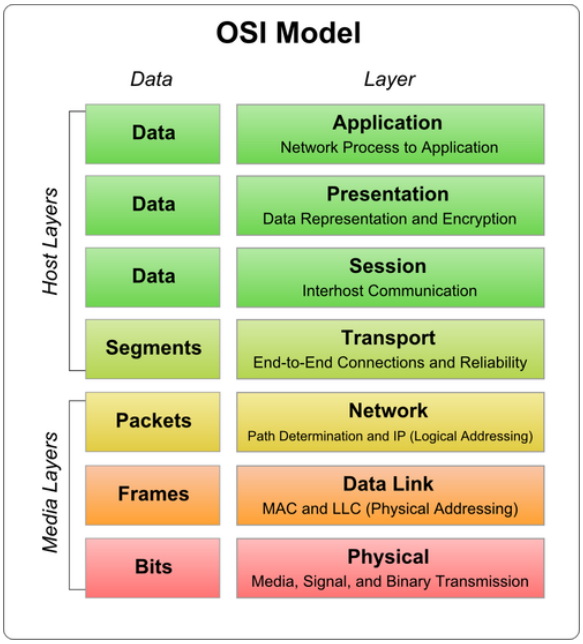

Figure 1. OSI 7-layer network model [113].

1.2 Technical Motivation

This section provides a brief overview of the technical motivations and methods utilized in this research effort.

1.2.1 RF Emission Exploitation.

For decades, it has been known that Electro-Magnetic (EM) emissions can contain exploitable information about system operations. Declassified TEMPEST documents revealed that in the 1960s the United States was aware that cryptographic systems were vulnerable to EM analysis and the Soviets had guidelines regarding RF interference indicating that they also recognized the threat of exploitation [10]. The field of Side Channel Analysis (SCA) aims to extract information from unintended sources to gain insight into the operation and nature of a device. In 1996, Paul C. Kocher presented the first SCA attack which used information about the execution time of a MCU-based RSA algorithm implementation to extract the pri-

vate cryptographic key [69]. A few years later, Kocher presented a method of Differential Power Analysis (DPA) which analyzed power consumption measurements to extract secret keys from "tamper resistant" devices [70]. These groundbreaking publications paved the way for many new exploitation techniques ranging from visual analysis to statistical models.

This research effort exploited the URE that are produced by all electronic devices during their operation. One benefit of using URE attributes to classify devices is that they based on physical characteristics of a device and are therefore difficult to spoof. A large body of prior research has successfully demonstrated the ability to detect and classify devices [11, 14–16, 20–22, 24, 27, 29–31, 34, 46, 49, 91, 93, 101, 102, 125, 127] and device operations [100–105, 127–129] based on RF attributes using a variety of methods. This research utilized a non-contact, near-field RF probe to acquire URE from a Device Under Test (DUT) during its operation, providing a *non-destructive* collection method that does not require the IC to be removed and limits interference from other system components. For each of the three considered classification methods, classifier models were first developed using a set of "training" emissions corresponding to each of the authorized device or operation classes. Afterwards, the performance of each classifier model was evaluated using a second, independent set of "testing" emissions.

1.2.2 RF-DNA Fingerprints and Classifiers.

The RF-DNA process captures, analyzes, and quantifies the variance in RF emissions related to physical variances in manufactured semiconductor devices [14] and/or different device operations [102]. Statistical features are extracted from an observed RF emission to construct a collection of values called the RF "fingerprint" which is associated with a specific device or operation. RF fingerprints can be used

with a variety of classification methods to identify the source device or operation corresponding to an observed RF emission. This research considered two classification methods that have been utilized in previous RF-DNA research: Multiple Discriminant Analysis/Maximum Likelihood (MDA/ML) [14–16,26,29,67,68,91–93,109, 110, 124, 125, 127] and Generalized Relevance Learning Vector Quantized-Improved (GRLVQI) [26, 67, 68, 90–93, 102, 127]. Both methods have been successfully used for device classification and verification; however, this research effort extends their use to also classify executed operations.

1.2.3 Correlation and Matched Filtering.

The implementation of MDA/ML and GRLVQI classification methods used in previous RF-DNA research, while effective, can become computationally expensive for a large number of classes and/or RF-DNA fingerprint characteristics. The high complexity of classification can prevent these processes from being implemented on information systems with limited processing capabilities or power constraints, such as mobile or embedded platforms [102]. While work continues to quantify and reduce the computational complexity of classification processes [6, 59], this research effort evaluated the viability of using relatively simple correlation-based matched filtering for classification as an alternative to computationally complex algorithms like MDA/ML and GRLVQI. Correlation is well suited to implementation in real-time, inexpensive hardware because it is a conceptually straight forward function with a well-defined computational complexity [102].

Matched filtering implemented with correlation is commonly used for symbol estimation in digital communication systems due to its optimal performance as a receiver in the presence of noise [89, 99]. Correlation is also used in other fields such as image processing which required signals to be identified in noisy environ-

ments [28]. Prior research at the Air Force Institute of Technology (AFIT) has analyzed URE with Correlation-Based Anomaly Detection (CBAD) techniques to detect anomalous software execution in PLCs [102, 103, 105] and MCUs [128, 129]. This research effort continued the investigation of correlation-based techniques and matched filtering by aiming to identify the individual operations that are executed by a MCU as well as discriminate between multiple, similar MCU devices.

1.2.4 Device and Operation Classification.

Classification performs a one-to-many comparison to classify an observed RF emission as belonging to one of N_C known classes. In the case of *device discrimination* the known classes correspond to authorized devices. For *operation identification* the classes represent possible executed operations. This research evaluated the classification performance of three methods to classify RF emissions as one of $N_C = N_{DevAuth} = 8$ authorized devices or $N_C = N_{Op} = 12$ possible operations.

1.2.5 Device ID Verification.

Verification performs a one-to-one comparison of an unknown RF emission to a specific known device with the goal of determining if the unknown emission originated from the claimed device. The verification process can be used to verify a device's bit-level identity to grant it system access. This process parallels procedures for biometric human ID verification, such as using a photo ID card to verify a person's identity before granting them access [102]. This research evaluated the verification performance when presenting each of $N_{DevAuth} = 8$ authorized devices and $N_{DevRogue} = 2$ unauthorized "rogue" devices for verification claiming one of the authorized identities. A verification assessment for "rogue operations" was not performed because a MCU can only execute operations from a known, finite set.

1.3 Research Contributions

The research goal involved comparing the effectiveness and computational costs of three different techniques to classify URE: 1) MDA/ML with RF fingerprints, 2) GRLVQI with RF fingerprints, and 3) matched filtering with time domain emissions. The three techniques were evaluated for both the task of discriminating between multiple MCU devices and identifying the operations executed on a device. As summarized in Table 2, AFIT research contributions in the RF INTelligence (RFINT) field have been made in several technical areas. Previously undefined acronyms that are used in the table include: Time Domain (TD), Spectral Domain (SD), Correlation Domain (CD), Intentional RF Emissions (IRE), Learning From Signals (LFS), and Matched Filtering (MF).

Table 2. Relational mapping between RFINT *Technical Areas* in *Previous* related work and *Current* AFIT research contributions. The × symbol denotes specific areas addressed.

Technical Area	Previous Work		Current Research	
	Addressed	Ref #	Addressed	Ref #
TD Features	×	[67, 68, 90–92, 102–105] [109, 110, 124, 125, 127–129]	×	[100, 101]
SD Features	×	[15, 16, 93, 125]		
CD Features	×	[102–105, 109, 110, 128, 129]	×	[100, 101]
Emission Type				
Intentional (IRE)	×	[29, 46, 47, 49, 67, 68, 90] [92, 93, 109, 110, 124, 125]		
Unintentional (URE)	×	[14–16, 25, 26] [102–105, 127–129]	×	[100, 101]
Burst	×	[29, 46, 47, 49, 67, 68, 90] [92, 93, 109, 110, 124, 125]		
Continuous	×	[14–16, 102–105, 127–129]	×	[100, 101]
High SNR	×	[29, 46, 47, 49, 67, 68, 90] [92, 93, 109, 110, 124, 125]		
Low SNR	×	[14–16, 102–105, 127–129]	×	[100, 101]
Classification/Verification Processes				
MDA/ML	×	[14–16, 26, 29, 67, 68, 91–93] [109, 110, 124, 125, 127]	×	[101]
GRLVQI	×	[26, 67, 68, 90–93, 102, 127]	×	
LFS	×	[46–49]		
MF			×	[100, 101]
Dimensional Reduction Analysis (DRA)				
MDA/ML	×	[47, 67, 68, 91–93, 127]		
GRLVQI	×	[66, 90–93, 102, 127]		
LFS	×	[46–49]		
Verification				
Electronic Components	×	[14–16, 102, 127]	×	[101]
Wireless Devices	×	[29, 91–93]		
Device Operations	×	[102–105, 127–129]	×	[100, 101]

1.4 Document Organization

The remaining chapters are organized as follows. Chapter II provides background information regarding sources of RF emissions, detecting counterfeit and Trojan devices, reverse engineering and validating device operations, the MSP430 DUT, MDA/ML, GRLVQI, and correlation-based classification methods. Chapter III provides details on the methodology used for this research effort including the DUT operating conditions, RF signal collection, post collection processing, classifier model development, and the procedures for evaluating classification, verification, and computational time. Chapter IV presents the results of the methodologies from Chapter III including device classification, operation classification, device ID verification, and computational time. Chapter V provides a summary of the research results and recommendations for potential future research efforts.

II. Background

This chapter provides background information on topics associated with this research effort. Section 2.1 provides information about sources of Radio-Frequency (RF) emissions and factors that can cause variations in emissions between devices and executed operations. Section 2.2 covers the applications and techniques pursued by prior related research efforts evaluating unintentional device emissions. Details about the MSP430 MicroController Unit (MCU), which was used as the Device Under Test (DUT) for this research effort, are cover in Section 2.3. Background information and justification for the signal collection techniques used in this research are provided in Section 2.4. Section 2.5 provides background on methods used to classify RF Distinct Native Attribute (RF-DNA) fingerprints and Section 2.6 describes the correlation-based classification process used with Time Domain (TD) signals.

2.1 RF Emissions

The field of Side Channel Analysis (SCA) aims to extract information about a device from *unintended* observable phenomena that are correlated with its internal state and operation. Commonly exploited side channels include variations in computational time, power consumption, acoustic, RF, and optical emissions [80]. This research focused specifically on exploiting the RF energy that is radiated by all electronic devices.

2.1.1 Intentional and Unintentional Emissions.

The Electro-Magnetic (EM) emissions produced by electronic devices can be separated into two broad categories: *intentional* and *unintentional* [2]. The term

Intentional RF Emissions (IRE) describes RF energy that is *intentionally* broadcast by devices such as wireless radios, cellular phones, and IEEE 802.15 Bluetooth devices as part of their primary function [102]. In addition to serving their purpose to transmit wireless information, IRE can also leak secondary information about the identity and operating conditions of the device. Prior RF-DNA research efforts analyzed IRE to improve network security for several wireless communication standards including IEEE 802.11 WiFi [46, 47, 67, 68, 109, 110], IEEE 802.15 ZigBee [29], and IEEE 802.16 WiMAX [49, 90, 93, 124, 125].

In addition to the RF energy that wireless communication devices *intentionally* broadcast, all modern semiconductor devices also *unintentionally* broadcast RF energy due to electronic and EM field coupling between components. Operations such as transistors switching on and off cause current fluctuations within semiconductor devices which produce Unintentional RF Emissions (URE) through three types of coupling: conductive, inductive, and radiative [80]. When a physical conductive path carries a signal through a system, conductive coupling produces faint currents on all conductive surfaces or lines attached to the system. These conductive emanations often ride on top of strong, intentional currents within the same conductors [2] and can be observed in the power supply, ground line, and cables attached to the device [10]. Inductive coupling occurs when two conductive paths are separated by less than a wavelength and the current flowing through one wire induces a voltage in the neighboring wire. When two circuit components are separated by more than a wavelength, radiative coupling can cause parts of the circuit to act as an antenna that transmits undesired RF waves which interfere with the other circuit components [86]. The URE that escape the device package can be observed externally using a near-field RF probe as illustrated by Figure 2 which shows the normalized RF spectral intensity observed across the surface of a MSP430 chip.

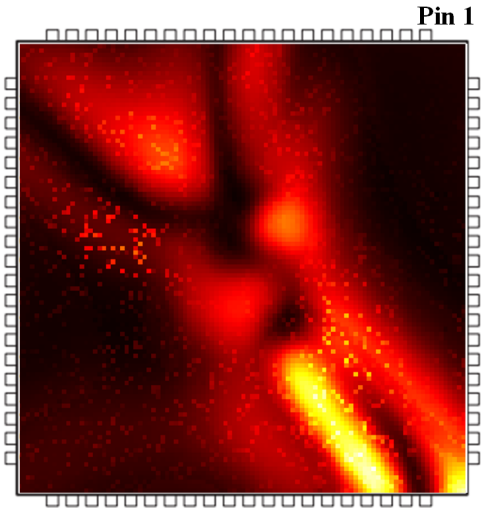

Figure 2. Normalized RF spectral intensity of URE measured at 100×100 locations spanning the surface of an MSP430F5529 chip package.

IRE and URE are both targets for SCA because they can carry potentially compromising information; however, the differences between URE and IRE require differing approaches for exploitation. IRE are based on wireless broadcast standards that have a well defined signal structure and are typically transmitted at a high signal power which can be received by a "distant" antenna. URE, on the other hand, do not have a specifically designed signal structure and are emitted at a significantly lower signal power than IRE signals. The acquisition of URE requires an RF probe and collection details, such as bandwidth and target frequency, are typically developed based on knowledge of device characteristics (clock rate, feature size, etc...) and observation of captured RF signals [102]. Prior Air Force Institute of Technology (AFIT) research efforts have analyzed URE from a variety of devices such as Programmable Logic Controllers (PLCs) [25, 26, 102–105, 127] and MCUs [14–16, 100, 101, 128, 129].

2.1.2 Variations Between Devices.

The unintentional RF energy that all electronic devices radiate as part of their normal operation is characteristic of their function, design, and construction. Additionally, the Integrated Circuit (IC) fabrication process introduces very small scale (deep sub-micron in modern IC technology) variations in the final structure of devices. As long as those variations are within acceptable tolerances the device will be functionally correct; however, like snowflakes or human fingerprints, no two chips are *exactly* alike. The tiny structural variants that make integrated circuits unique also color the inherent URE in a manner that can be exploited to identify the source device [15]. As demonstrated by previous RF-DNA research, even two "same model" devices can be differentiated from each other based on their URE with a high degree of accuracy [14–16, 102].

2.1.3 Variations Due to Operation.

In addition to the impact of physical and electrical device characteristics on URE, the operations and data being processed by digital hardware will also influence emissions [2]. Prior research to perform side-channel based reverse engineering of PIC16F687 MCUs evaluated the effects that clock rate, working register, fetching process, bus data, processed data, opcode, instruction register, and noise had on device power consumption as measured using a shunt resistor between the chip's GrouND (GND) and power supply. The research considered two simplistic models to relate the device's power consumption to the opcode and data being processed: *Hamming-Weight* and *Hamming-Distance* [39].

The *Hamming-Weight* model assumes that power consumption is proportional to the number of bits set in a chosen intermediate value, such as the opcode or operand data. The observed power consumption can then be used to estimate the

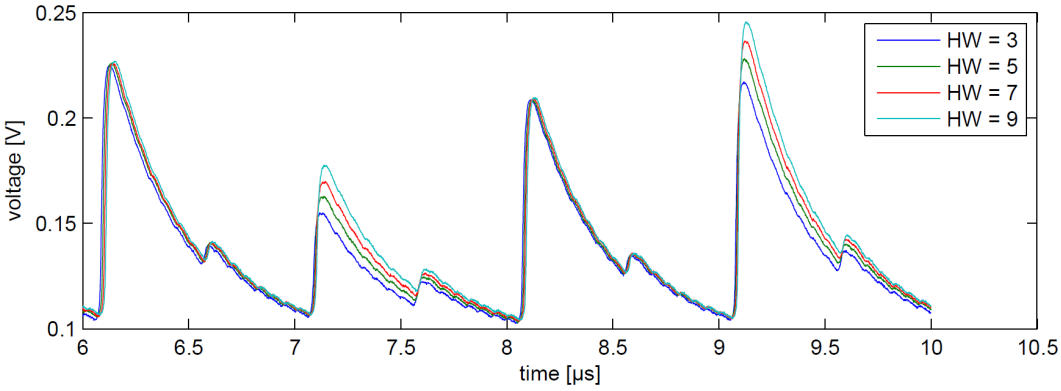

Figure 3. Averaged traces demonstrating the effects on power consumption caused by Hamming weight of the current opcode in a PIC16F687 MCU [39].

Hamming weight of the intermediate value. Figure 3 shows that the Hamming weight of the opcode has a characteristic impact on the power consumption in the second and fourth clock cycles of PIC16F687 operation execution [39].

The *Hamming-Distance* model takes into account the number of altering opcode and/or data bits between consecutive operations. This model is generally more accurate than Hamming-Weight; however, it is more complicated to apply in practice because both the previous and current values need to be known. Figure 3 shows that the Hamming distance between consecutive opcodes has a characteristic impact on the power consumption in the fourth clock cycles of PIC16F687 operation execution [39].

The current research effort built on the previous power analysis based efforts to determine if factors such as Hamming weight and Hamming distance of data and opcodes also influenced URE in a useful way that could enable operation classification. Initial experiments analyzing URE from the MSP430F5529 MCU showed that there were visibly discernible differences in the TD waveforms for certain operations related to the Hamming weight of the input operands and result value. One such experiment executed one of three operations (`AND.B`, `MOV.B`, and `XOR.B`) after loading the SouRCe (SRC) and DeSTination (DST) registers with one of seven values

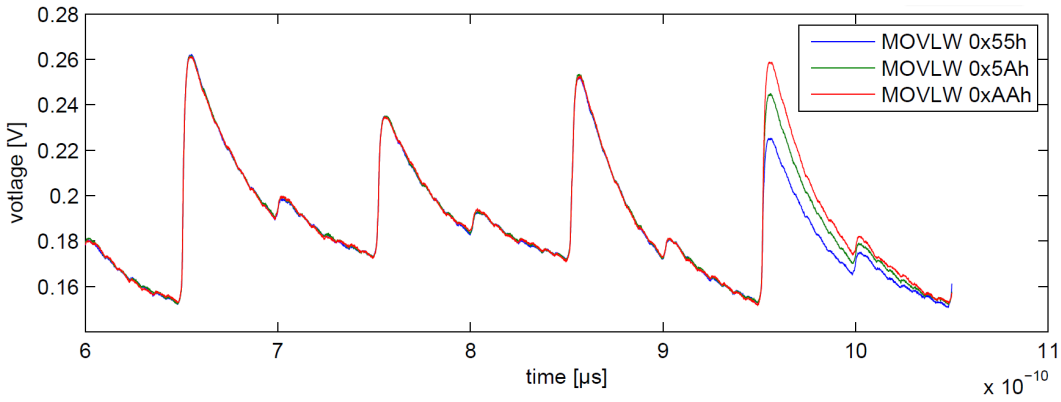

Figure 4. Averaged traces demonstrating the effects on power consumption caused by the Hamming distance of the current and previous opcode in a PIC16F687 MCU. Each trace was generated by first executing a `MOVLW 0x55h` instruction followed by another `MOVLW` with the operand either being `0x55h`, `0x5Ah`, or `0xAAh` [39].

having various Hamming weights. Since the SRC and DST registers were initialized to the same value in each case, the result values stored back into the DST register for the `AND.B` and `MOV.B` operation would be the same as the initial value that was already in the DST register. Figure 5 shows that the averaged TD waveform peaks of observed URE corresponding to the `AND.B` and `MOV.B` operations were closely grouped and intermixed in amplitude with negligible impact from the differing initialization values [100].

Since `XOR.B` stores the XOR'd value of the SRC and DST operands into the DST register (SRC $\oplus$ DST $\rightarrow$ DST) the result value being stored back into DST would be different from the initialization value in each case. Figure 5 shows that the averaged TD waveform peaks of observed URE corresponding to the `XOR.B` operation were visibly separated in amplitude and grouped according to operand Hamming weight. Additional early experiments also showed that it was possible to estimate the executed instruction or to estimate the SRC and DST registers from URE with better accuracy than "random guess" [100].

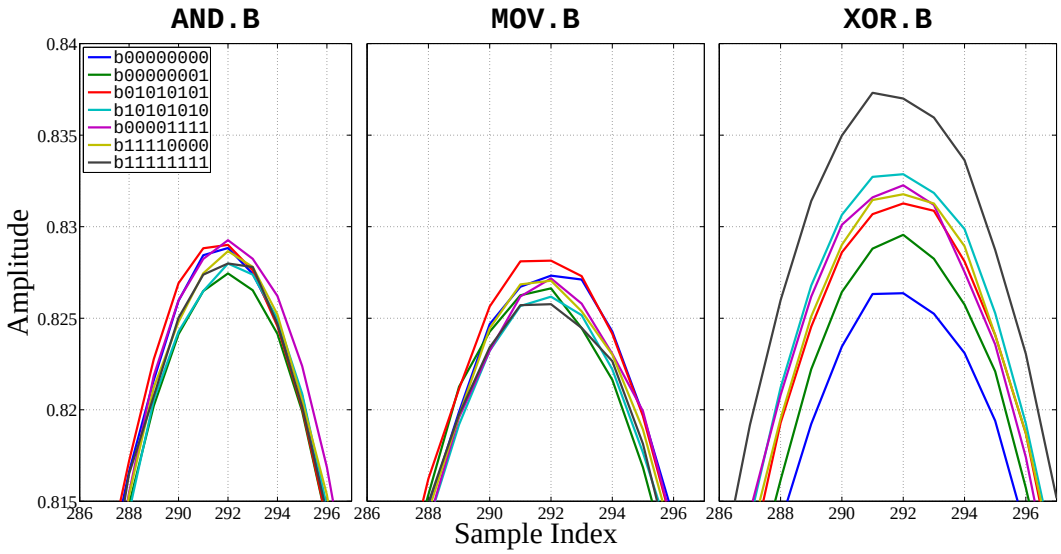

Figure 5. Expanded view of averaged TD waveform peaks of URE collected from an MSP430F5529 executing one of three operations. SRC (R4) and DST (R5) registers were loaded with the listed values prior to execution. Traces for `AND.B` and `MOV.B` are closely grouped because the input/result values are the same; however, traces for `XOR.B` are visibly separated according to input/result Hamming distance [100].

2.2 Applications

The following sections provide an overview of prior efforts focused on device authentication to detect counterfeit/Trojan hardware and operation validation through the reverse engineering of executed instructions.

2.2.1 Counterfeit and Trojan Hardware Detection.

The term "counterfeit hardware" can be applied to a wide range devices including cloned parts, mismarked parts, recycled parts from aged/reclaimed e-waste, and parts that have been intentionally modified for malicious purposes (hacking, theft, espionage). Devices that have been modified to contain malicious "Trojan" circuitry are especially concerning because they can act as backdoors to leak information or kill switches to disrupt operations.

The proliferation of counterfeit components has made it necessary for manufac-

turers and distributors to inspect incoming electronic components for authenticity. Physical evaluation processes range in complexity from visual inspection of the exterior component packaging using a low power microscope to internal analysis of a component's chemical composition using techniques such as X-ray fluorescence and Fourier transform infrared spectroscopy. Electrical inspection techniques are also often employed including parametric, functional, burn-in, and structural tests [43].

In addition to trying to detect existing counterfeit components, new anti-counterfeit mechanisms have also been developed which can be put in place during the design of new chips so they can be uniquely identified [43]. Prior research efforts have investigated the effectiveness of injecting specific compounds during the component manufacturing process to improve the RF fingerprinting performance [65] and intentionally inserting information into hidden device side-channels to act as a watermark [8].

One way that the Department Of Defense (DOD) is responding to the threat of counterfeit and Trojan hardware is by investing in the research groups at academic institutions. One such group, the University of Connecticut's Center for Hardware Assurance, Security, and Engineering (CHASE), received a $7.5 million grant in 2014 from the DOD to study security for nanoscale devices [98]. Commercial products to screen for counterfeit devices have also become available such as the Advanced Detection of Electronic Counterfeits (ADEC) system sold by Nokomis Inc. which analyzes URE using non-contact and non-invasive methods to detect counterfeits. Since ADEC is a proprietary technology, many of the implementation details are protected; however, patent filings suggest that the system may use a signal generator to stimulate a DUT and then collect emitted RF energy using an antenna array [63, 64]. Marketing literature from Nokomis Inc. claims that the ADEC system achieved 100% accuracy in two blind pilot tests in which 7 ADUM5241ARZ digital

isolator chips and 40 Atmel AT89S52 8-bit MCUs were correctly identified as being counterfeit or authentic [62]. Another company, Power FingerPrinting (PFP) Cybersecurity Inc., has developed their own technology for detecting counterfeit hardware based on observing side channel information such as power consumption. A test of the PFP solution examining 9 different samples of Intel's TB28F400B5-T80 Flash memory correctly identified 100% of the counterfeit parts [88].

Prior research efforts have demonstrated the ability to extract information from a variety of side-channels to detect counterfeit/Trojan devices including path delay [58], thermal profiling [72], and power consumption [32]. The two RF-DNA based techniques evaluated in this research effort build on prior AFIT work that successfully demonstrated *device discrimination* using the Multiple Discriminant Analysis/Maximum Likelihood (MDA/ML) and Generalized Relevance Learning Vector Quantized-Improved (GRLVQI) classification techniques with RF-DNA fingerprints generated from URE. Using RF-DNA with MDA/ML, [14] evaluated URE from 40 PIC24 16-bit MCUs and achieved a percent correct *classification* ($\%C$) of $\%C > 90\%$ for $SNR_A \geq 15$ dB and an average Equal Error Rate (EER) for *verification* of $EER_{Ave} < 0.05\%$. Later, [102] evaluated the verification performance of ten Allen Bradley PLCs using GRLVQI with RF-DNA fingerprints and achieved $EER_{Ave} \leq 1.6\%$ at $SNR = 15$ dB.

A process that is similar to the matched filtering technique used in this research has been applied to device identification before by analyzing the IRE of Ethernet packets to discriminate between Ethernet cards [38]. To this researcher's knowledge, the current effort is the first to apply matched filtering for *device discrimination* based on URE.

2.2.2 Operation Validation and Reverse Engineering.

Supervisory Control And Data Acquisition (SCADA) networks have become increasingly vulnerable to cyber attack in recent years as they have grown in complexity and become connected to the open Internet. Malicious hackers have recently exploited SCADA vulnerabilities to take control of critical systems, causing them to execute unauthorized operations that cause physical damage [131, 132]. Since the embedded hardware used in SCADA systems is often unable to defend itself by running anti-virus software due to limited processing capabilities, it is necessary to develop external monitoring methods that can verify critical system operation. Analyzing side-channels such as power consumption [13, 41], URE [102, 129], and Modbus/TCP register values [36] can provide vital insight into the actual operation of embedded devices to detect deviations from the authorized program.

Embedded devices commonly run a single application that performs a small number of repetitive actions such as actuating an electrical relay, controlling a pump, or collecting sensor readings which results in a small externally visible state space [13]. One approach to detect anomalous behavior in embedded systems is to monitor side-channel information for deviations from normal operation. Prior research analyzing the dynamic power consumption of a PIC18 8-bit MCU on a transceiver board generated a reference template representing normal operation by averaging over numerous training collections. Test signals were then collected and subdivided into small regions which were compared to the corresponding sections of the reference signal using correlation. The technique was able to discriminate between execution of the original and modified code to sending encrypted and unencrypted unicast transmissions with 100% accuracy [41].

Correlation-Based Anomaly Detection (CBAD) is another correlation based technique that has been demonstrated to successfully detect anomalous program

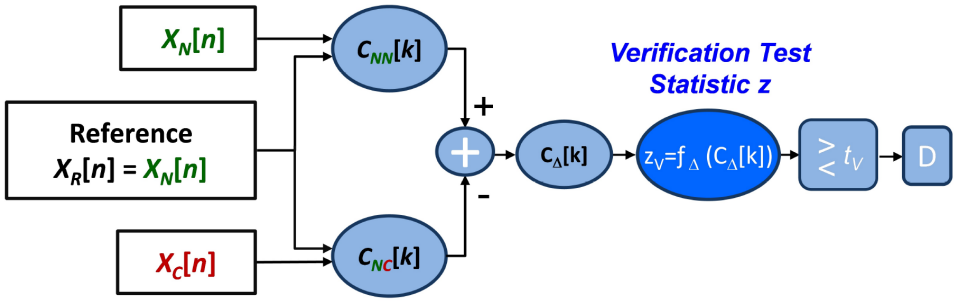

Figure 6. CBAD process comparing the *current unknown* sequence, $X_C[n]$, to the *reference* sequence representing *normal* operating conditions, $X_N[n]$. A binary decision is made to declare $X_C[n]$ the result of either a) *normal* operating conditions ($z_V \leq t_V$) or b) *anomalous* operation conditions ($z_V > t_V$) [102].

execution based on observed URE. The CBAD process calculates the difference between the auto-correlation of a *reference* signal representing *normal* operating conditions, $X_N[n]$, and the cross-correlation of that reference with the *current unknown* signal, $X_C[n]$, as shown in Figure 6. A difference function, $f_\Delta(C_\Delta[k])$, produces a test statistic value, z_V, which is compared to a threshold, t_V, to make a binary declaration that $X_C[n]$ resulted from *normal* ($z_V \leq t_V$) or *anomalous* ($z_V > t_V$) operating conditions [102].

When used to detect anomalous execution in ten Allen Bradley PLCs, a version of the CBAD technique using Hilbert transformed sequences achieved *verification* performance of $EER \leq 10\%$ for all ten PLC devices at $SNR = 0$ dB [102]. A similar CBAD process using Hilbert transforms was able to detect a modified version of the Advanced Encryption Standard (AES) algorithm running on a MSP430 16-bit MCU with $EER \leq 10\%$ at $SNR = 10$ dB [129]. While the CBAD approach has proven itself effective for detecting anomalous behavior, it only produces a binary decision of *normal* or *anomalous* and does not provide any additional information about the type of anomalous behavior.

Rather than using deviations from normal behavior to detect anomalous program execution, this research effort investigated the ability to estimate the *individ-*

ual instructions executed by a MCU. A sequence of extracted instructions can be compared to the authorized program to verify the device's operation. In addition to identifying anomalous execution, this approach has the added benefit that it can provide information about the unauthorized sequence of instructions which may help to determine the source of an anomaly.

A large portion of the existing body of SCA research targets encryption algorithms using methods such as power analysis [4, 61, 78] or template attacks [3, 5, 37, 42]. Those cryptographic focused efforts often target a specific encryption implementation, the details of which are known to the attacker, with the goal being to extract the secret key.

The emerging field of SCA for Reverse Engineering (SCARE) began exploiting side-channel leakage to recover non-trivial details about the way cryptographic functions were implemented. One of the first proposed SCARE efforts targeted substitution block lookup tables to attack a secret authentication and session key generation algorithm on Subscriber Identity Module (SIM) cards in Global System for Mobile (GSM) communication networks [83]. A later SCARE attack targeting block cipher Data Encryption Standard (DES) managed to retrieve details about some of the constants used by the algorithm for permutation tables and key scheduling as well as implementation choices such as the register where subkeys were loaded [23].

More recently, SCARE efforts have focused on extracting information about the executed code and processed data in simple microprocessor devices. The Java Card Virtual Machine was one of the first SCARE targets which was attacked to extract executed bytecodes [60, 123]. The research presented in [39] examined the types of information that could be gathered from single side-channel observations and laid the groundwork for later efforts targeting the executed operations in 8-bit MCUs.

A SCARE effort analyzing power signatures from an 8-bit PIC16 MCU correctly identified executed instructions with recognition rates around 66% using Principal Component Analysis (PCA) and 70% using Linear Discriminant Analysis (LDA). When hidden Markov Chains were also applied to help identify the most likely instructions to have occurred it increased the recognition rate by 17% [33]. A later research effort analyzing power consumption of an 8-bit ATMega163 MCU running the RSA signature screening algorithm claimed to have achieved 100% correct classification using PCA to identify 39 commonly used instructions [81, 82]; however, other sources were unable to reproduce those results [108].

The Side-ChANnel-based DisAssembLer using Local Electromagnetic Emanations (SCANDALee) project extracted executed code from an 8-bit PICF687 MCU based on URE and is the prior effort most closely related to the current research. The SCANDALee collection process is destructive and requires decapsulation of the DUT. A localized RF probe is place at 20 different locations in a 4×5 grid over the DUT to capture emissions. To produce training signals for each instruction, the registers were loaded with random data and the instruction to be learned was repeated with other random instructions executed before and after it. The SCANDALee research evaluated Multiclass Fisher's LDA (M-LDA) and Polychotomous LDA (P-LDA) techniques for dimensionality reduction. The template for each instruction class was formed from the mean and covariance matrix of the training collections [107, 108]. Table 3 provides a comparison showing how this research effort differed from the approach taken with SCANDALee including the choice of DUT, collection methods, and classification processes. SCANDALee achieved an instruction recognition rate of 96.24% using test patterns and 87.69% using a real code implementation of the AES [108].

Table 3. Comparison showing differences between SCANDALee [108] and the current research effort.

	SCANDALee	Current Research
DUT	PIC16F687 8-bit MCU	MSP430F5529 16-bit MCU
Evaluated Instructions	36 instructions (4 clock cycles each)	12 instructions (1 clock cycle each)
URE Collection Methods	RF probe @ 20 locations (decapsulated DUT)	RF probe @ 1 location (non-destructive, non-contact)
Analysis Methods	M-LDA P-LDA	MDA/ML (w/ RF-DNA) GRLVQI (w/ RF-DNA) Matched Filtering (w/ TD)

2.3 MSP430 Device Description

Prior research efforts focused on reverse-engineering MCU operations have analyzed simple 8-bit MCUs like the PIC16 [33, 39, 107, 108] and ATMega163 [81, 82]. In 2011, 16-bit MCUs became the largest market segment, surpassing 4 and 8-bit MCUs as being the most commonly shipped devices [52]. To maintain relevance with current trends, this research effort selected the MSP430F5529 16-bit MCU manufactured by Texas Instruments [117] as the DUT because it is widely used and representative of modern MCU architecture and semiconductor manufacturing processes. Although the exact fabrication details are proprietary, the MSP430 was most likely fabricated using a 65 nm process [119].

The MSP430 utilizes a Reduced Instruction Set Computing (RISC) architecture with 27 instructions having zero, one, or two operands with each operand using one of seven possible addressing modes. Information about the addressing modes is included in Appendix A. The execution time of a each operation ranges from one to six clock cycles depending on the instruction and addressing mode used as demonstrated by Table 4 which lists several instruction/addressing mode combinations and their corresponding execution duration [118]. This variable execution time

Table 4. Example MSP430 instruction and addressing mode combinations with different execution durations [118].

Instruction Mnemonic	SRC Addressing Mode	DST Addressing Mode	Execution Duration
`MOV.B`	Register	Register	1 Cycle
`MOV.B`	Register	Indexed	3 Cycles
`MOV.B`	Indirect Register	Indexed	4 Cycles
`MOV.B`	Indexed	Indexed	5 Cycles
`XOR.B`	Register	Register	1 Cycle
`XOR.B`	Register	Indexed	4 Cycles
`XOR.B`	Indirect Register	Indexed	5 Cycles
`XOR.B`	Indexed	Indexed	6 Cycles

complicates the process of reverse-engineering executed operations from observed URE because it is difficult to determine when one operation ends execution and the next one begins in a long sequence of unknown operations. Prior research efforts analyzing the simpler 8-bit PIC16 MCU did not have to address this issue because each instruction cycle of the PIC lasts exactly four clock cycles [33].

One possible approach for dealing with instructions that span multiple clock cycles is to treat each clock cycle as consecutive, individual instructions and create classification templates for each one [81]. Given the number of valid MSP430 instruction/addressing mode combinations and their respective durations, such an approach would result in 1980 single cycle "instruction" templates which is beyond a reasonable scope for this effort. Therefore, for the purpose of *operation identification*, the scope of this research was limited to only consider the Format I instructions listed in Table 5 using the register-to-register addressing mode. The input/output relationship of each operation and the affected status bits is not directly relevant to this research effort; however, it has been included in Table 5 for completeness. The MSP430 Format I operations require two operands, SRC and DST, and have an effective duration of one clock cycle when using the register-to-register addressing mode [118].

Table 5. MSP430 Double-Operand (Format I) Instructions [118].

Mnemonic	Operation	Status Bits			
		V	N	Z	C
MOV.B	src → dst	-	-	-	-
ADD.B	src + dst → dst	*	*	*	*
ADDC.B	src + dst + C → dst	*	*	*	*
SUB.B	dst + .not.src + 1 → dst	*	*	*	*
SUBC.C	dst + .not.src + C → dst	*	*	*	*
CMP.B	dst - src	*	*	*	*
DADD.B	src + dst + C → dst (*decimally*)	*	*	*	*
BIT.B	src .and. dst	0	*	*	$\bar{Z}$
BIC.B	.not.src .and. dst → dst	-	-	-	-
BIS.B	src .or. dst → dst	-	-	-	-
XOR.B	src .xor. dst → dst	*	*	*	$\bar{Z}$
AND.B	src .and. dst → dst	0	*	*	$\bar{Z}$

This subset of instructions was selected for evaluation because register-to-register operations are commonly used in compiled code. For example, a disassembly analysis of the AES-128 encryption/decryption function [114] compiled for the MSP430F5529 revealed that approximated 29% of the Format I instructions in the routine used the register-to-register addressing mode. Due to the proprietary nature of the MSP430, there is very little official documentation that reveals implementation details about its architecture. According to a thread on the Texas Instrument E2E forum [116], although the MSP430 process is not pipelined, it does have the ability to pre-fetch the next instruction if the destination address of the current instruction is a register. For the purposes of this research the existence of a pre-fetch stage is ignored and, per the User's Guide, all of the register-to-register operations are treated as having an execution duration of one clock cycle [118].

2.4 Signal Collection

One common avenue for SCA of MCUs is to analyze the power consumption of the DUT. This is typically accomplished by measuring the voltage across a shunt

resistor which is connected between the device GND and power supply. While power analysis has provided successful results in other research efforts [33, 39, 81, 82], it requires a contact probe to be physically connected to the DUT which could potentially interfere with device operations. Other prior research focused on reverse-engineering MCU operations from URE used a non-contact EM probe to collect emissions; however, the techniques used there required the DUT to be decapsulated and thus left physical traces [107, 108]. Techniques that require chip decapsulation may be useful in a laboratory test setting, but they are impractical for a deployed monitoring implementation because they are destructive to the DUT.

This research effort used a RF near-field probe similar to other AFIT research efforts to collect URE from above the surface of the chip package [14–16, 100–105, 127–129]. Using the near-field probe provides a non-destructive RF signal collection method which will not interfere with the DUT's normal operation and will support future in situ analysis of legacy systems.

2.5 RF-DNA-Based Classification

"RF-DNA" is itself not a classification method and only refers to the process used to extract statical values from RF emissions to generate "RF fingerprints". The resulting fingerprints can then be evaluated using a variety classification methods. The following two sections describe the classification methods that were used in this research to classify RF-DNA fingerprints: MDA/ML and GRLVQI.

2.5.1 MDA/ML Classifier.

Multiple Discriminant Analysis (MDA) is a multi-class form of Fisher's (two-class) LDA extended to evaluate N_C classes. The MDA projection matrix, $\mathbb{W}$, is used to project an N_f-dimensional RF-DNA fingerprint vector, $\mathbf{F}$, in $N_C - 1$ space

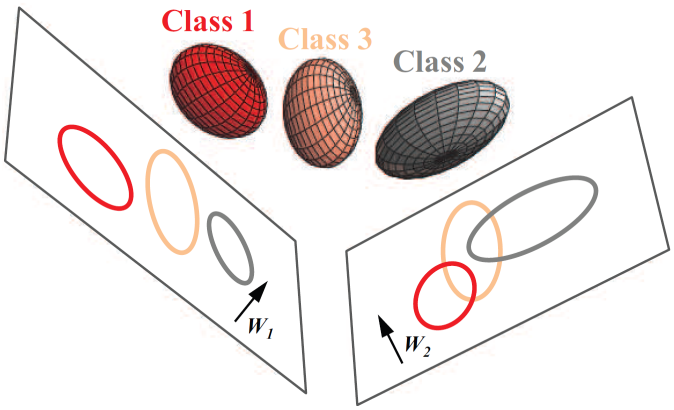

Figure 7. Representative MDA projection of N_C=3 class inputs onto two possible N_C-1=2-dimensional subspaces [90].

[74] as illustrated in Figure 7 for $N_C = 3$ classes. The set of N_{Tng} fingerprints for each of the N_C classes are use to generate a projection matrix $\mathbb{W}$ which optimally maximizes the *inter-class* distance and minimizes the *intra-class* distance. The MDA process assumes a Gaussian distribution of input data and the transformed features remain Gaussian distributed. The projected training fingerprints are used to determine the mean location for each class in $N_C - 1$ space.

To classify an *unknown* test input, test statistics, such as the Euclidean distance to the mean class locations, are calculated from the projected input fingerprint for each of the N_C classes. A Maximum Likelihood (ML) decision is made by choosing the class with the smallest distance from the projected fingerprint and assigning it as the estimated class for the unknown test input. The MDA model development and classification processes used in this research were conducted in accordance with prior AFIT RF-DNA research [14–16, 26, 29, 47, 67, 68, 91–93, 109, 110, 124, 125, 127] and are described in Section 3.4.2 and Section 3.5.1.

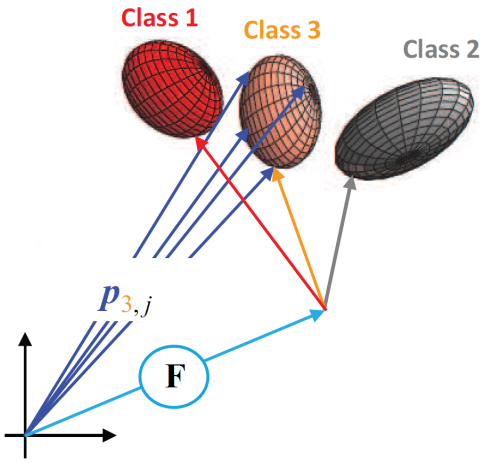

Figure 8. Representative GRLVQI *classification* assigning *unknown* fingerprint, F, to class C_i based upon the minimum Euclidean distance to prototype vectors [90].

2.5.2 GRLVQI Classifier.

GRLVQI is an Artificial Neural Network (ANN)-based classification process that uses iterative learning to generate a set of N_P "prototype vectors" that describe each of the N_C classes [45, 77] as illustrated in Figure 8 for $N_C = 3$ classes. The learning process supports Dimensional Reduction Analysis (DRA) to define the prototype vectors using a subset of the total N_f RF-DNA features by selecting the most relevant features; however, this research effort only considered full-dimensional fingerprints. GRLVQI was adopted for AFIT RF-DNA research [26, 66–68, 90–93, 102, 127] because it overcomes some of the drawbacks of MDA/ML, including [45, 77]:

- Each RF-DNA fingerprint feature is assigned a relevance rank measuring its significance to the overall *classification* decision.
- Feature selection is performed in conjunction with *classification*.
- There are no assumptions made or knowledge required regarding the input data distribution.

- Processing is well-suited for noisy or inconsistent input data.

Details about the GRLVQI model development and classification processes used in this research are provided in Section 3.4.3 and Section 3.5.2, respectively.

2.6 Correlation-Based Classification

Correlation is a very useful tool for detecting signals that are corrupted by additive random noises [73]. The mathematical operation of *correlation* used to implement a "correlator" takes a signal and correlates it with a replica of itself. A "matched filter" uses the related mathematical operation of *convolution* to convolve a signal with the mirror image of the signal, delayed by its time duration. For this document, the term "matched filter" will be used synonymously with "correlator" because the matched filter convolution of a signal and its time-reversed-and-delayed replica produces the same output as correlation between the signal and its replica at the end of the time duration [99].

The correlation processing used in this research is similar to that used for symbol estimation in a traditional digital communication system [89, 99] whereas the CBAD methods used in prior research [102, 103, 105, 128, 129] were more consistent with correlation techniques commonly used in other fields like image processing that require signal identification in noisy environments [28]. The goal of the CBAD process was to make the binary decision whether a device was executing *normal* or *anomalous* operations from observed URE. CBAD declared an unknown input emission as *normal* or *anomalous* depending on how closely it was correlated to a reference signal representing the expected *normal* behavior [102, 103, 105, 128, 129]. This research differs from the CBAD approach because it aims to classify an unknown input emission as corresponding to one of several possible devices or MCU operations. Rather than comparing an input emission to a single "normal" refer-

ence signal, the input emission is evaluated using a bank a parallel matched filters corresponding to each of the possible devices or MCU operations.

The matched filters used discrete correlation given by

$$y[n] = \sum_{k=-\infty}^{\infty} h[n-k]x[k] \tag{1}$$

to compare the unknown input emission, x_{Unk}, with a reference signal, x_{Ref}, corresponding on of the possible devices or MCU operations. The output is a single test statistic value, z_V, which is equivalent to the cross-correlation of x_{Ref} and x_{Unk} at $k = 0$ time lag, calculated as

$$z_V = y[0] = \sum_{k=0}^{N} x_{Ref}[k]x_{Unk}[k] \tag{2}$$

with a higher value of z_V represents a closer match between x_{Unk} and x_{Ref}. From an a-posterior probability perspective, the raw z_V test statistics can be used to independently implement the *classification* and *verification* processes [16].

As discussed in Chapter I, the motivation for using correlation is that it a relatively simple function that is well suited for implementation in systems with limited computing capability. The execution cost of other classification processes can vary greatly, but the computational cost of correlation is predictable and well bounded. For two discrete sequences of length N, the computational time complexity is computable and analytically bounded by $\mathcal{O}(N^2)$ [102].

III. Methodology

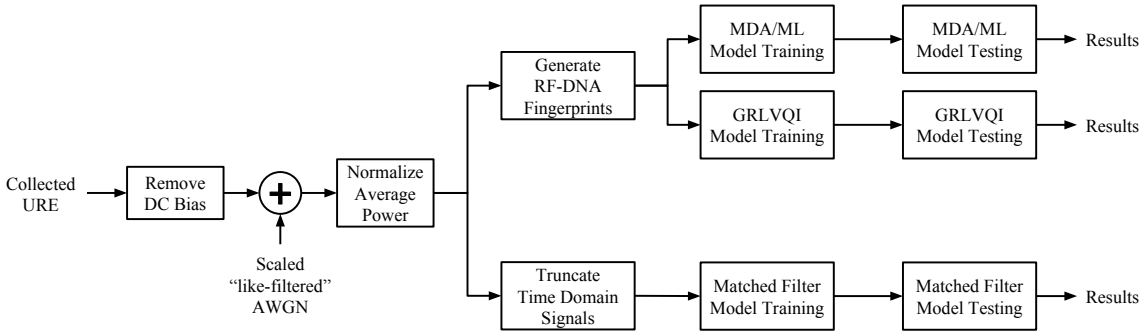

Figure 9. Block diagram outlining the methodology described in Chapter III.

This chapter provides details on the research methodology and process stages outlined in Figure 9 which were used to generate the results presented in Chapter IV. First, the operating conditions of the MSP430 Device Under Test (DUT) during signal collection for device discrimination and operation identification are discussed in Section 3.1. Details about the configuration of the acquisition hardware are described in Section 3.2. Section 3.3 describes the series of Digital Signal Processing (DSP) steps applied to all collected Unintentional Radio-Frequency (RF) Emissions (URE) to generate RF Distinct Native Attribute (RF-DNA) fingerprints and truncated waveforms for matched filtering. Details about the generation of the Multiple Discriminant Analysis/Maximum Likelihood (MDA/ML), Generalized Relevance Learning Vector Quantized-Improved (GRLVQI), and matched filtering classification models are discussed in Section 3.4. The methods used to evaluate and present results for classification and verification performance are described in Sections 3.5 and 3.6, respectively. Finally, Section 3.7 describes the process used to evaluate the computational cost associated with each of the three classification methods.

3.1 MSP430 Operating Conditions

This section provides details about the $N_{Dev} = 10$ MSP430F5529 MicroController Units (MCUs) which were evaluated for this research and their physical operating conditions during URE acquisition. The purchased devices were mounted on the MSP-EXP430F5529 Experimenter Board [115] which was used to control the DUT during collections. Assembly code was used to write different software routines to execute during acquisition for the purposes of *device discrimination* or *operation identification*.

3.1.1 Evaluated Devices.

Table 6 shows the MSP430F5529 revision and MSP-EXP430F5529 board manufacturing lot information for each of the $N_{Dev} = 10$ devices. $N_{DevAuth} = 8$ of the DUTs were treated as "authorized" devices and evaluated for *device classification*. The remaining $N_{DevRogue} = 2$ DUTs were treated as unauthorized "rogue" devices and considered for *device verification*. However, all evaluated devices were actually new, authentic MSP-EXP430F5529 Experimenter Boards.

The MSP-EXP430F5529 boards selected to represent "authentic" devices all contained Rev. F MSP430F5529 MCUs whereas the boards representing "rogue" devices used the older Rev. E MSP430F5529. Although both versions of the MCU are functionally similar, the minor differences between revisions may impact URE. The Rev. F and Rev. E devices were grouped as "authentic" and "rogue" to simulate the differences that may exist between real and counterfeit/altered devices. In reality, Rev. F of the MSP430F5529 fixed several bugs from Rev. E related to the Analog-to-Digital Converter (ADC), flash memory, Power Management Module (PMM), system module, Unified Clock System (UCS), and Universal Serial Bus (USB) as well as introduced a new error in the BootStrap Loader (BSL) [120].

Table 6. DUT Names Mapped to N_{Dev}=10 MSP-EXP430F5529 Boards

DUT Name	MSP-EXP430F5529 Manufacturing Lot	MSP430F5529 Revision
Auth1	01182013GM	F
Auth2	01182013GM	F
Auth3	01182013GM	F
Auth4	01182013GM	F
Auth5	01182013GM	F
Auth6	01182013GM	F
Auth7	01182013GM	F
Auth8	01182013GM	F
RogueA	09052012S2	E
RogueB	11202012	E

The DUT named "RogueA" was arbitrarily selected as the device to use for evaluating *operation identification* and all training and testing emissions for that purpose were collected from it. This research effort did not evaluate the effectiveness of using training emissions collected from one device to estimate operations using testing emissions collected from a different device.

3.1.2 DUT Configuration.

The eZUSB connector of the MSP-EXP430F5529 Experimenter Board was connected to a host desktop computer which provided power to the DUT and was used to load the software routines to execute during URE collection. The MSP-EXP430F5529 is a development board which contains a variety of peripherals including an accelerometer, potentiometer, Liquid Crystal Display (LCD) screen, and capacitive touchpad buttons [115]. To reduce the impact that these peripherals might have on URE collected above the MSP430 chip, all non-critical board elements were powered-off and disconnected by removing jumper pins. The only jumpers that remained were the *MSP430* jumper, which provided power to the MCU, and the *DVCC*, *TXD*, *RXD*, *RST*, and *TEST* jumpers.

The DUT was configured to operate at a Master CLocK (MCLK) frequency of $f_{MCLK} = 1.049MHz$ which was derived from an internal trimmed low-frequency REFerence Oscillator (REFO) with a typical frequency of $f_{REFOCLK} = 32.768$ kHz. A Frequency Locked Loop (FLL) used the REFO with a frequency integrator to drive a Digitally Controlled Oscillator (DCO) which provided the MCLK signal [118].

After a DUT is initially powered, the URE may exhibit a temperature-dependency as the device warms up to its normal operating temperature [15]. To compensate for these effects, each DUT was allowed to run for fives minutes prior to signal collection to allow the temperature to stabilize. A hand-held InfraRed (IR) thermometer was used to measure and record the surface temperature at the center of the MSP430F5529 after the five minute warm up period. Table 7 shows that all DUTs had a surface temperature between 73.5 °F and 75.0 °F prior to acquisition.

Table 7. Surface Temperature of DUT in Fahrenheit After Five Minute Warm-Up Period Prior to First Acquisition

DUT Name	**Auth**								**Rogue**	
	1	2	3	4	5	6	7	8	A	B
Temperature (°F)	74.0	74.5	74.5	74.5	75.0	75.0	75.0	74.5	75.0	73.5

3.1.3 Program Sequence for Device Discrimination.

For the purpose of *device discrimination*, a single software routine was created which executed the sequence of nine operations shown in Table 8 which corresponds to a subsection of the MSP430 AES-128 encryption function distributed by Texas Instruments [114]. The specific subsection was selected because it contains a variety of Format I operations using different addressing modes. The sequence has a total duration of 16 clock cycles with individual instructions having durations of one, two, or three cycles.

Table 8. Subsection of AES-128 Encryption Function Used as Program Sequence for Device Discrimination with Clock Cycle Durations Listed for Each Instruction

```
XOR.B  R8,     R12     ; 1 clock cycle
XOR.B  R9,     R12     ; 1 clock cycle
MOV.B  R12,    0(R10)  ; 3 clock cycles
MOV.B  4(R1),  R12     ; 3 clock cycles
ADD.B  #1,     R12     ; 1 clock cycle
MOV.B  R12,    R13     ; 1 clock cycle
MOV.W  R13,    4(R1)   ; 3 clock cycles
CMP.B  #4,     R12     ; 1 clock cycle
MOV.W  R13,    R13     ; 2 clock cycles
```

For each of the $N_{Dev} = 10$ DUTs, the 16 cycle sequence was repeated enough times to collect $N_B = 5000$ emissions per DUT. The $N_B = 5000$ emissions were then separated into two sets: $N_{Tng} = 1000$ training emissions to develop the classification model and $N_{Tst} = 4000$ testing emissions to evaluate its performance.

Early experiments showed that the act of physically swapping out devices in the acquisition fixture produced variations in the observed URE that could potentially bias the *device classification* process. Initially, emissions were collected from each DUT over $N_{Acq} = 2$ separate acquisition sessions with the DUTs being removed and replaced in the acquisition fixture between sessions. Matched filter templates were then generated using an equal number of training emissions from each of the $N_{Acq} = 2$ sessions per DUT. When the testing emissions for each DUT were evaluated using the corresponding matched filter template for that DUT, the resulting test statistics were distributed such that it was possible to distinguish many of the emissions as having originated from one of the $N_{Acq} = 2$ acquisition session.

Efforts were made to maintain consistent DUT and probe placement between acquisition sessions by using the custom built device mount described in Section 3.2.1 and keeping the horizontal XY position of the probe constant between sessions. The probe was only raised and lowered vertically as required to exchange DUTs and it was always returned to the same resting position on the surface of the chip

package. The post-collection processing steps described in Section 3.3.1 were also applied to reduce the impact of collection bias. Despite these efforts, minor differences in URE persisted which often made it possible to distinguish individual collection sessions from each other.

As the number of acquisition sessions per DUT was increased, the impact of the differences between collection sessions decreased and the resulting distribution of test statistics approached a Gaussian. After additional experimentation, it was determined that increasing the number of acquisition sessions to $N_{Acq} = 5$ provided a distribution of test statistics that reasonably approximated a Gaussian as illustrated in Figure 10. Therefore, to reduce the impact of collection bias, the $N_B = 5000$ emissions for each DUT were collected over the course of $N_{Acq} = 5$ acquisition sessions per DUT with the DUT being physically removed and then repositioned in the test fixture between each session. The groups of $N_{Tng} = 1000$ training and $N_{Tst} = 4000$ testing emissions per DUT were formed such that they contained an equal number of emissions from each of the $N_{Acq} = 5$ sessions.

3.1.4 Program Sequence for Operation Identification.

For the purpose of *operation identification*, the scope of this research was limited to only consider the following $N_{Op} = 12$ Format I operations using the register-to-register addressing mode: `ADD.B`, `ADDC.B`, `AND.B`, `BIC.B`, `BIS.B`, `BIT.B`, `CMP.B`, `DADD.B`, `MOV.B`, `SUB.B`, `SUBC.B`, and `XOR.B`. All instructions were configured to use registers `R4` and `R5` as the SouRCe (SRC) and DeSTination (DST) registers, respectively. To eliminate the impact that changes in the Hamming weight of operand and results values can have on URE [100], all registers were initialized to contain the value zero prior to execution which prevented the result values stored into `R5` from changing as the operations were repeated.

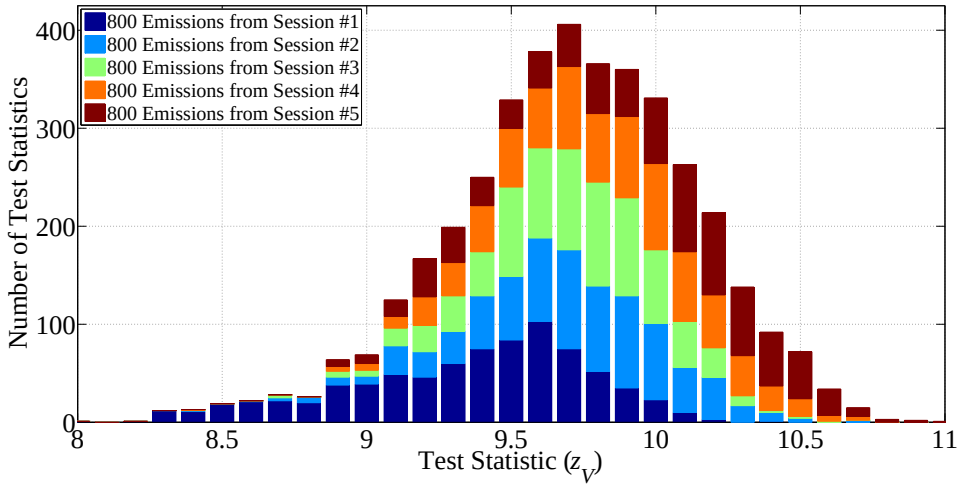

Figure 10. Stacked histogram of matched filter test statistics for *device classification* from N_{Tst}=4000 testing emissions collected equally over N_{Acq}=5 acquisition sessions. The matched filter template was generated from N_{Tng}=1000 separate training emissions which were collected during the same N_{Acq}=5 acquisition sessions. The results shown were produced using Auth1 and are representative of other devices.

Separate software routines were created for each of the $N_{Op} = 12$ operations to be evaluated which began by generating a signal pulse on a General Purpose Input/Output (GPIO) line of the MSP430 DUT to trigger acquisition and then repeated the operation of interest for a set number of times. Each of the $N_{Op} = 12$ software routines were executed enough times to collect $N_B = 5000$ emissions for each operation which were then separated into two sets: $N_{Tng} = 1000$ training emissions and $N_{Tst} = 4000$ testing emissions. All $N_B \times N_{Op} = 5000 \times 12 = 60000$ emissions were collected in a single session without moving the probe or DUT to avoid collection bias due to repositioning.

3.2 RF Signal Collection

This sections provides details about the hardware and configuration used for the acquisition system.

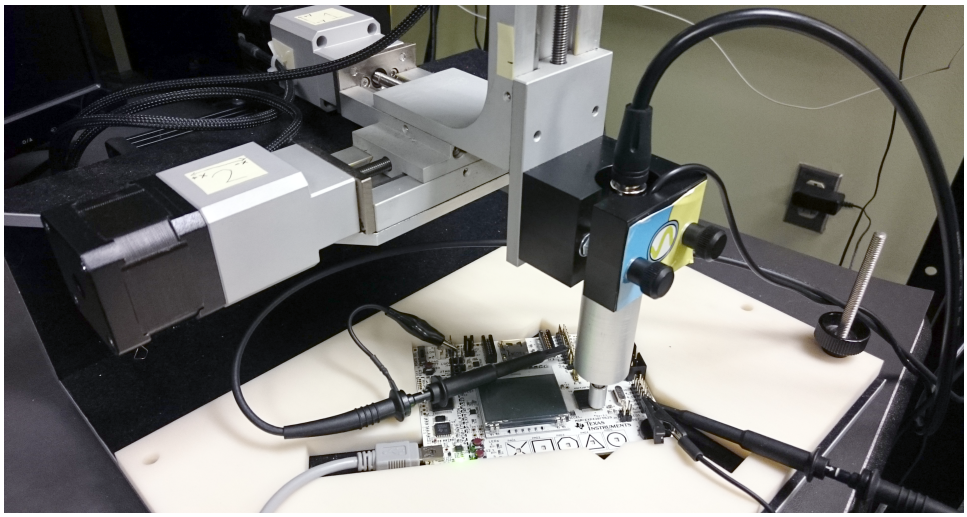

Figure 11. XYZ table, RF near-field probe, and DUT in custom mount used for URE collection. This is the same acquisition system used in [100, 101, 128, 129].

3.2.1 Acquisition System.

A computer controlled XYZ table was used to position a Riscure high-sensitivity RF near-field probe [94] above the DUT which was held stationary in the custom built device mount shown in Figure 11. The probe was lowered to rest above the surface of the MSP430 chip package and the metal shroud was lowered to shield the probe tip from the surrounding environment.

The RF near-field probe was sampled using a Teledyne LeCroy WavePro 760Zi-A oscilloscope which is capable of sample rates up to $f_S = 40$ Gsps with 8-bit of resolution and 6 GHz of bandwidth. Previous Air Force Institute of Technology (AFIT) RF-DNA research efforts typically used sample rates around $f_S = 200$ Msps to $f_S = 1$ Gsps.

Initial experiments for this research effort sampled URE at a higher rate of $f_S = 10$ Gsps and then down sampled the data to lower rates of $f_S = 1$ Gsps by taking every tenth sample and $f_S = 250$ Msps by taking every fortieth sample. The matched filtering classification process described in this chapter was used to eval-

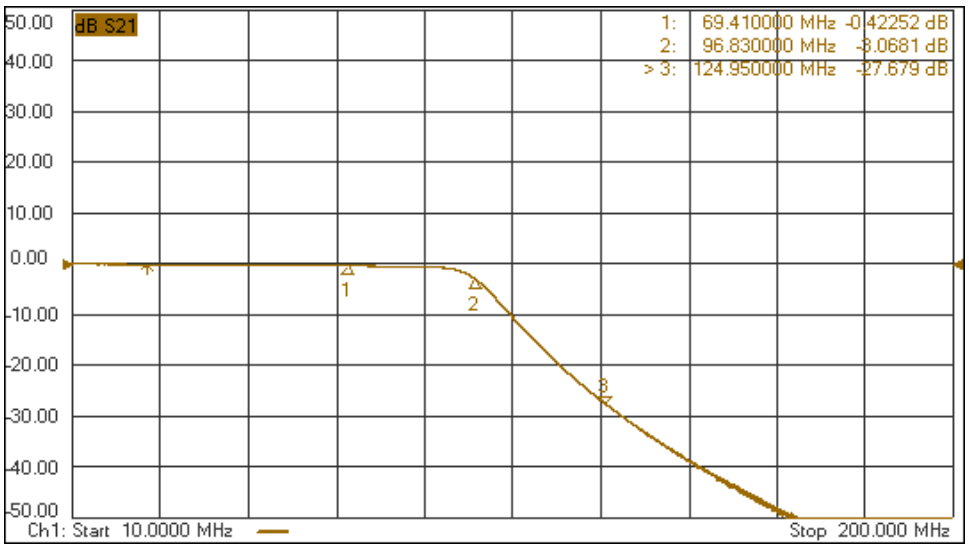

Figure 12. Measured frequency response of Mini-Circuits BLP-90+ Coaxial LPF with nominal -3 dB cutoff frequency f_{CO}=90 MHz.

uate the data at each of the three sample rates. The data that was down sampled to $f_S = 1$ Gsps achieved classification performance that was statistically equivalent to the original $f_S = 10$ Gsps data with confidence intervals of $CI = 95\%$ and the $f_S = 250$ Msps data performed statistically worse than the other two sample rates. Based on these results, all future URE acquisitions were collected and analyzed at $f_S = 1$ Gsps. To prevent aliasing, a Mini-Circuits BLP-90+ coaxial Low Pass Filter (LPF) with a passband spanning DC to 81 MHz and a nominal -3 dB cutoff frequency of $f_{CO} = 90$ MHz was inserted between the probe and the oscilloscope [79]. The frequency response of the LPF is shown in Figure 12 as measured using a network analyzer.

A significant amount of clock jitter was observed in the MSP430 MCLK signal during initial experiments. The DUT had a nominal operating frequency of $f_{MCLK} = 1.049MHz$ which corresponds to a clock period of $T_{MCLK} = 953.3$ ns; however, clock periods were observed in the range of $T_{MCLK} \in [908, 1002]$ ns as shown in Table 9. The varying clock frequency may have been a side effect of the

Table 9. Minimum, Maximum, and Mean Observed MCLK Periods from N_B=5000 Acquisitions of N_{Dev}=10 DUTs Executing the Routine for Device Discrimination

DUT Name	Min T_{MCLK} (ns)	Max T_{MCLK} (ns)	Mean T_{MCLK} (ns)
Auth1	920	969	945.6
Auth2	940	988	954.7
Auth3	921	968	952.3
Auth4	914	949	944.5
Auth5	925	972	943.6
Auth6	943	990	951.0
Auth7	908	955	950.1
Auth8	932	980	949.3
RogueA	949	1002	955.4
RogueB	917	954	946.2

FLL used to generate the MCLK signal. It is also possible that the jitter was an intentional design feature to curtail the harmonic content of the square wave. Using a dithered clock which intentionally varies the clock frequency by a small amount is a recommended technique for spreading emissions out in the frequency spectrum to reduce the strength at any single frequency [86].

The signal processing techniques used in this research required precise alignment of sampled URE with clock cycle edges to subdivide and truncate signals. GPIO line 7.7 of the MSP430 DUT was configured to output the MSP430 MCLK signal which was sampled and recorded on a secondary oscilloscope channel via a contact probe. Another contact probe was also connected to GPIO line 1.0 of the MSP430 DUT to trigger acquisitions based on the software generated pulse. All signal collections were stored and processed using MATLAB® software.

3.2.2 RF Near-Field Probe Placement.

To determine the collection location that would provide the best discrimination between devices using matched filtering classification, the RF probe was physically located over a 25 × 25 grid spanning the surface of three arbitrarily chosen DUTs

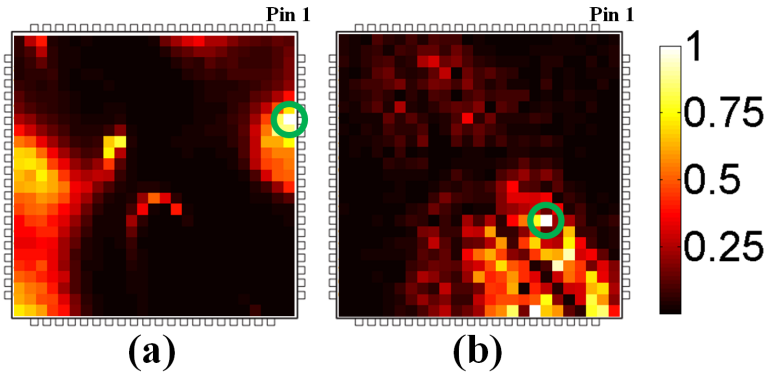

Figure 13. Summed variance of URE between (a) DUT ∈{Auth1, Auth2, RogueA} and (b) operation ∈ {ADD.B, AND.B, MOV.B, XOR.B,} at 25×25 locations above the MSP430F5529. The circled locations were selected to collect emissions for (a) device discrimination and (b) operation identification based on the maximum summed variance between classes. The grid values were normalized for display purposes with lighter colors representing higher summed variance values.

(Auth1, Auth2, and RogueA) executing a repeated `XOR.B` operation. The URE collected from each DUT at each location, L, within the 25×25 grid were evaluated using

$$\underset{L}{\operatorname{argmax}} \left(\sum_{k=1}^{N} Var(x_{Auth1,L}[k], x_{Auth2,L}[k], x_{RogueA,L}[k]) \right) \tag{3}$$

where Var denotes variance and N is the number of samples in each collected emission, x. The probe location that provided the maximum summed *inter-device* variance is circled in Figure 13a and was used to collect URE for the purpose of *device discrimination*.

A similar process was used to determine the best collection location for *operation identification* using an arbitrarily chosen DUT, RogueA. URE were collected at each location within a 25×25 grid while RogueA repeatedly executed one of four arbitrarily chosen operations (`ADD.B`, `AND.B`, `MOV.B`, `XOR.B`). The probe location that provided the maximum summed *inter-operation* variance

$$\underset{L}{\operatorname{argmax}} \left(\sum_{k=1}^{N} Var(x_{\texttt{ADD.B},L}[k], x_{\texttt{AND.B},L}[k], x_{\texttt{MOV.B},L}[k], x_{\texttt{XOR.B},L}[k]) \right) \tag{4}$$

was selected as the collection location for the purpose of *operation identification* and is circled in Figure 13b.

3.3 Post Collection Processing

This section provides details about the post collection signal processing steps shown in Figure 14. All collected emissions had the DC offset removed, simulated noise added, and the average power normalized. Two separate processes were then used to produce: 1) RF-DNA fingerprints to be used with MDA/ML and GRLVQI classification and 2) Time Domain (TD) signals to be used with matched filtering classification.

3.3.1 DC Bias Removal and Power Normalization.

Slight differences in the acquisition system configuration and probe position between URE collection sessions can impact the observed DC offset and signal power. To reduce the impact of DC bias, the amplitude of each emission was centered by subtracting the signal mean to remove the DC offset. This DC removal process was accomplished prior to calculating the collected signal power to generate appropriately scaled simulated noise.

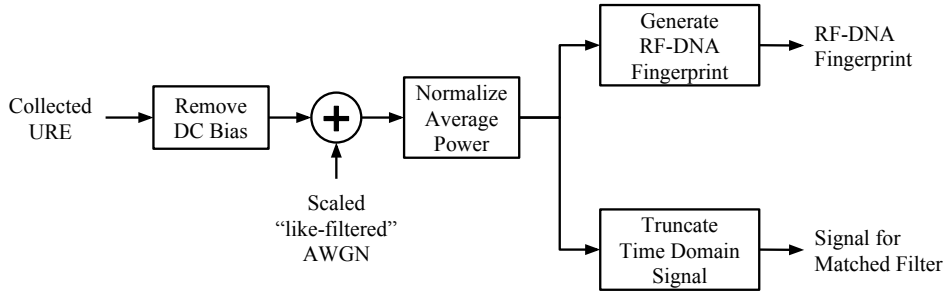

Figure 14. Block diagram outline the digital signal processing steps applied to all collected URE to generate RF-DNA fingerprints and truncated waveforms for matched filtering.

Additionally, after adding simulated noise, the average power of each emission was also normalized to reduce the effect of collection bias. This normalization process was based on average power rather than maximum amplitude so that the correlation process used for matched filtering would be sensitive to differences in signal shape and not be impacted by signals having different average powers.

Other RF-DNA and Correlation-Based Anomaly Detection (CBAD) research efforts have used additional post-collection signal processing stages including a Band Pass Filter (BPF), Digital Down Conversion (DDC), and Hilbert Transform (HT) [102, 105]. Preliminary experiments for this research effort used the matched filtering techniques to classify emissions from $N_{Op} = 12$ operations after applying every permutation of the following processing stages:

- BPF $\in \{On, Off\}$ with center frequency $f_C \in \{3.5, 4.0, 4.5, ..., 33.5\}$ MHz and bandwidth $W_{-3dB} \in \{1, 2, 3, 4, 5, 6\}$ MHz
- DDC $\in \{On, Off\}$
- HT $\in \{On, Off\}$

After analyzing all 1468 possible permutations of additional processing stages it was determined that the best matched filtering classification performance was achieved when none of the BPF, DDC, or HT stages were applied. Therefore, the only post-collection processing applied to the collected emissions prior to RF-DNA fingerprint generation or signal truncation for matched filtering were DC offset removal and average power normalization.

3.3.2 SNR Scaling.

Since URE are themselves a type of noise, for this research the Signal-to-Noise Ratio (SNR) at signal collection, SNR_C, is assumed to be infinite with the entire

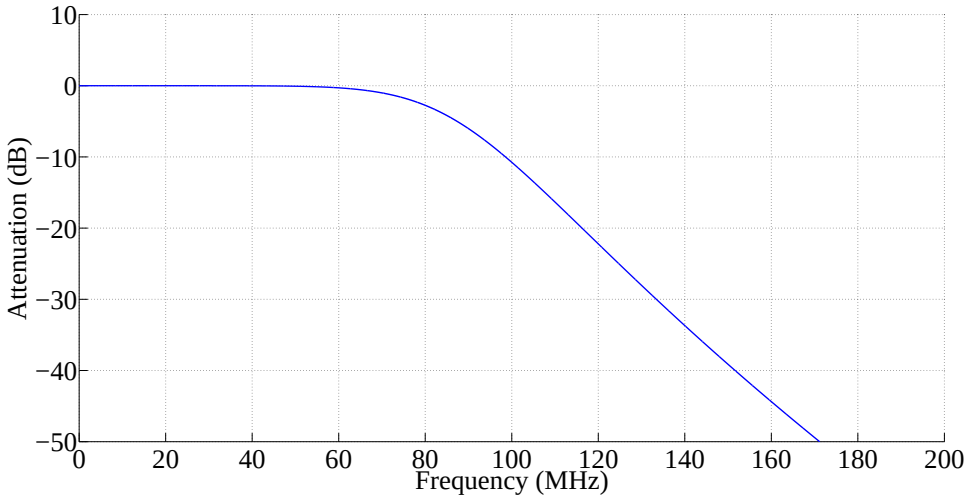

Figure 15. Simulated frequency response of 8th order Butterworth LPF with f_{CO}=90 MHz. Filter coefficients were generated using the Matlab® `butter` function.

acquired waveform representing useful signal content. To evaluate the impact of SNR on *classification* and *verification*, like-filtered Additive White Gaussian Noise (AWGN) was generated, scaled, and added to each collected waveform to simulate a range of analysis SNRs, $SNR_A \in [-30, 30]$ in $SNR_{A\Delta} = 5$ dB intervals. Prior work has shown that Gaussian distributed noise is an appropriate model for emission noise [39].

The MATLAB® `randn` function was used to generate $N_{Nz} = 1$ independent realization of normally distributed psuedorandom values for each of the $N_B = 5000$ emissions collected per device or operation. The filter whose frequency response is shown in Figure 15 was applied to the noise realization using the MATLAB® `filtfilt` function to simulate the response of the in-line, anti-aliasing LPF used during URE collection. The same filtered noise realizations were scaled to achieve each of the desired SNR_A points and added to the corresponding $N_B = 5000$ emissions for each device or operation.

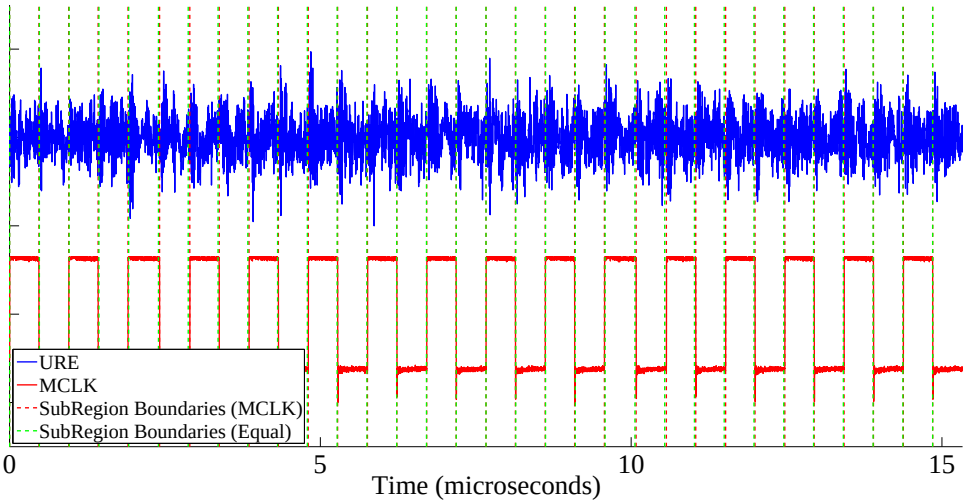

Figure 16. Collected URE and MCLK signal from RogueA divided into N_{SR}=32 subregions at the rising and falling edges of MCLK to generate RF-DNA fingerprints for device discrimination.

3.3.3 RF-DNA Fingerprint Generation.

The signal Region Of Interest (ROI) used for *device discrimination* corresponded to the 16 clock cycle duration of the program described in Section 3.1.3. Prior research determined that partitioning samples into subregions corresponding to integer multiples of the number of clock cycles in the ROI yielded statistically superior results relative to partitioning based on fractional clock cycles [14]. Due to the observed clock jitter, the two methods illustrated by Figure 16 were considered for choosing the ROI and dividing it into subregions.

The first method used the recorded MCLK signal to individually select the ROI for each emission that corresponded to exactly 16 clock cycles. The ROI was then divided into N_{SR} subregions precisely at the rising and falling edges of MCLK. This resulted in a ROI and subregions that *varied in length* for each evaluated emission. The second considered method truncated the ROI such that all recorded emissions were the same length as the shortest observed emission and then divided the ROI into N_{SR} *equal length* subregions.

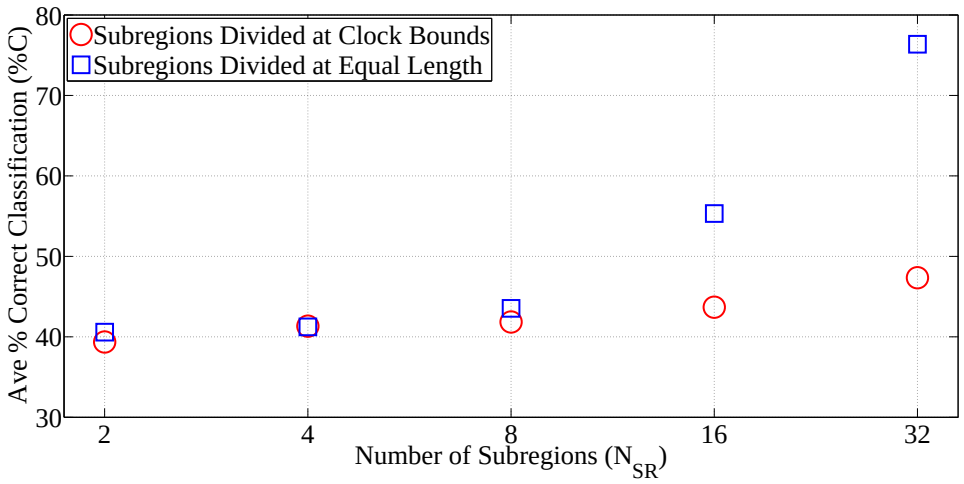

Figure 17. Average MDA/ML classification performance evaluated at SNR_C for $N_{DevAuth}$=8 device classes using RF-DNA fingerprints generated with $N_{SR}\in\{2,4,8,16,32\}$ subregions divided at MCLK boundaries (red) or into equal intervals (blue). Confidence intervals of CI=95% are approximately the same size or smaller than the marker shapes and have been omitted for visual clarity.

To determine which ROI and subregion selection method produced the best *device classification* performance, $N_{FP} = 5000$ RF-DNA fingerprints were generated at SNR_C using each method with $N_{SR} \in \{2, 4, 8, 16, 32\}$ subregions for each of the $N_{DevAuth} = 8$ authorized device classes. Those fingerprints where then evaluated using MDA/ML classification as described in Section 3.4.2 and Section 3.5.1 to compare the performance of each method. As shown in Figure 17, dividing the ROI in $N_{SR} = 32$ *equal length* subregions (blue) produced statistically better classification performance when compared to the other evaluated methods (red). Therefore, the results presented in Chapter IV for MDA/ML and GRLVQI *device classification* and *verification* where generated using RF-DNA fingerprints with $N_{SR} = 32$ equally divided subregions.

For the purpose of *operation identification*, the ROI was selected to correspond to a single MCLK cycle since all of the $N_{Op} = 12$ register-to-register instructions have an effective duration of one clock cycle [117]. To determine the optimal ROI

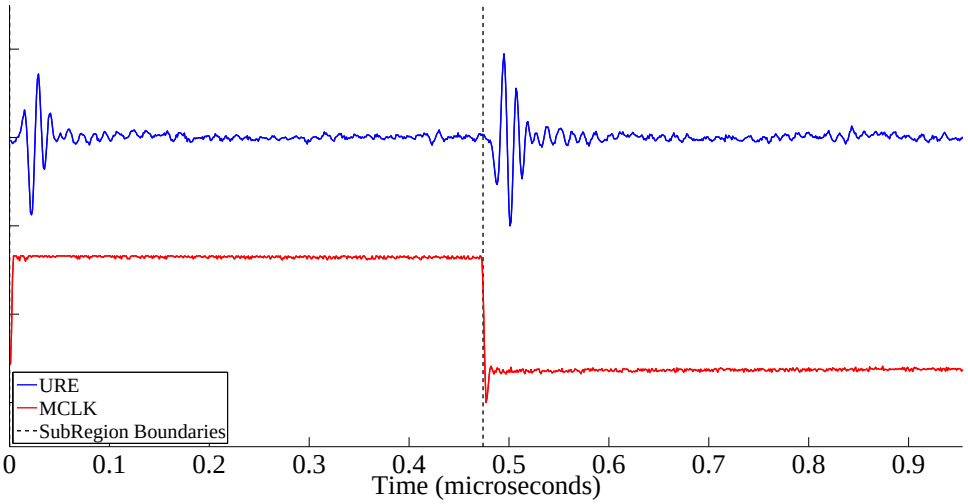

Figure 18. Collected URE and MCLK signals corresponding to an `XOR.B` operation divided into N_{SR}=2 subregions at the rising and falling edges of MCLK to generate RF-DNA fingerprints for operation identification. The `XOR.B` operation is representative of other MSP430 Format I instructions.

and subregion selection method to use for *operation identification*, the previously described process was used to evaluate the MDA/ML classification performance of differing subregion selection methods with $N_{SR} = 2$ subregions. Dividing the ROI into $N_{SR} = 2$ subregions at *MCLK edges* achieved $\%C_{Ave} = 9.8\%$ whereas dividing the ROI into $N_{SR} = 2$ *equal* subregions only achieved $\%C_{Ave} = 8.9\%$ for $N_{Op} = 12$ operation classes. Therefore, for the purpose of *operation identification*, the RF-DNA fingerprints used to evaluate MDA/ML and GRLVQI classification were generated using $N_{SR} = 2$ subregions divided at *MCLK edges* as shown in Figure 18.

RF-DNA fingerprints were generated for each of the $N_B = 5000$ emissions per device or operation using TD features in accordance with previous AFIT RF-DNA fingerprinting research [127]. The RF-DNA process is based on complex signals having In-phase (I) and Quadrature-phase (Q) components; however, the emissions used in this research were collected and stored as real-valued TD sequences. The MATLAB® `hilbert` function was used to transform each real-valued sequence, $x[n]$, into the complex IQ representation $x_{IQ}[n] = x_{re}[n] + x_{im}[n]$. The instantaneous

amplitude $a[n]$, phase $\phi[n]$, and frequency $f[n]$ was calculated for each of the N_{SR} subregions and the total ROI as

$$a[n] = \sqrt{x_{re}[n]^2 + x_{im}[n]^2} \tag{5}$$

$$\phi[n] = \tan^{-1}\left[\frac{x_{im}[n]}{x_{re}[n]}\right], x_{re} \neq 0 \tag{6}$$

$$f[n] = \frac{1}{2\pi}\left[\frac{d\phi[n]}{dn}\right] \tag{7}$$

for the complex signal $x_{IQ}[n] = x_{re}[n] + x_{im}[n]$. Then the standard deviation (σ), variance (σ^2), skewness (γ), and kurtosis (κ) of the instantaneous amplitude, phase, and frequency of the signal was calculated within each of the N_{SR} subregions as well as the total ROI to produce $N_f = 4 \times 3 \times (N_{SR} + 1)$ statistical features.

The $N_f = 4 \times 3 \times (32 + 1) = 396$ statistical features resulting from the $N_{SR} = 32$ subregions used for *device discrimination* were combined into a single vector as shown in Table 10 to construct the RF-DNA fingerprint.

Table 10. Construction of RF-DNA fingerprint as used for *device discrimination* with N_{DevSR}=32 subregions to produce N_f=396 statistical features.

Statistic:	σ	σ^2	γ	κ	σ	σ^2	γ	κ	...	σ	σ^2	γ	κ	σ	σ^2	γ	κ	σ	σ^2	γ	κ	...	σ	σ^2	γ	κ	σ	σ^2	γ	κ	...	σ	σ^2	γ	κ
Region:	SR=1				SR=2				...	SR=32				Total				SR=1				...	Total				SR=1				...	Total			
Attribute:	Amplitude																	Phase									Frequency								

Similarly, the $N_f = 4 \times 3 \times (2 + 1) = 36$ statistical features resulting from the $N_{SR} = 2$ subregions used for *operation identification* were combined into a single vector as shown in Table 11.

Table 11. Construction of RF-DNA fingerprint as used for *operation identification* with N_{SR}=2 subregions to produce N_f=36 statistical features.

Statistic:	σ	σ^2	γ	κ	σ	σ^2	γ	κ	σ	σ^2	γ	κ	σ	σ^2	γ	κ	σ	σ^2	γ	κ	σ	σ^2	γ	κ	σ	σ^2	γ	κ	σ	σ^2	γ	κ
Region:	SR=1				SR=2				Total				SR=1				SR=2				Total				SR=1				SR=2			
Attribute:	Amplitude												Phase												Frequency							

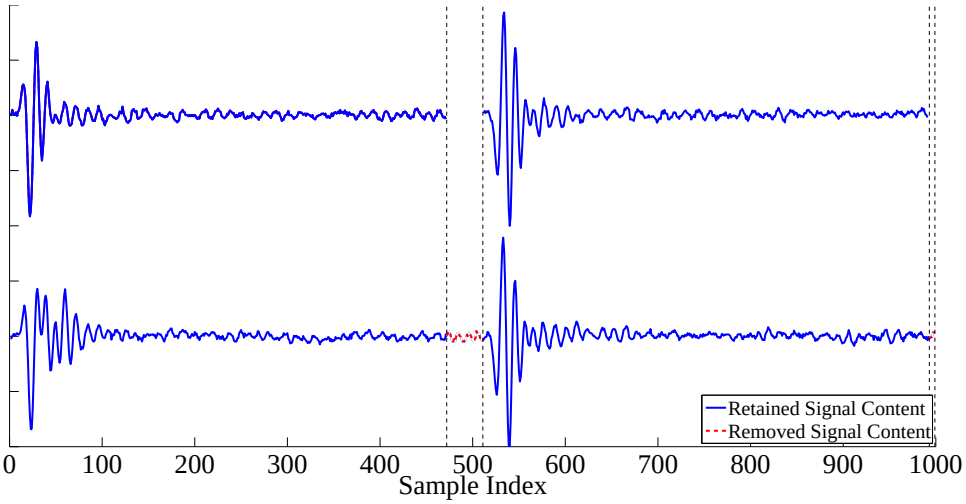

Figure 19. Regions of an `XOR.B` operation having a longer clock period (bottom) which were removed to match the length of the shortest observed waveform (top).

3.3.4 Matched Filtering Signal Truncation.

The correlation process used for matched filtering classification described in Section 3.5.3 required that all of the waveforms have the same length. To remove the effect of clock jitter shown in Table 9, the collected emissions were subdivided at rising and falling clock edges using the recorded MCLK signal, then all of the half-cycle segments were individually truncated to have the same length as the shortest recorded segment. This truncation process is illustrated in Figure 19 using two `XOR.B` operations and was applied to the emissions for both *device discrimination* and *operation identification*.

3.4 Model Development

The MDA/ML, GRLVQI, and matched filter classification models where generated using a k-Fold Cross-Validation (kF-CV) process with $N_{Tng} = 1000$ training emissions for each of the $N_C = N_{DevAuth} = 8$ classes for *device classification* or $N_C = N_{Op} = 12$ classes for *operation classification*. The model for MDA/ML in-

cludes a projection matrix, GRLVQI utilizes prototype vectors, and matched filtering requires reference templates for each of the N_C possible classes. For each of the three classification techniques, a single model was developed at the highest evaluated $SNR_A = 30$ dB and then used to evaluate each of the $N_{Tst} = 4000$ emissions per class across the full range of $SNR_A \in [-30, 30]$ in $SNR_{A\Delta} = 5$ dB intervals.

3.4.1 k-Fold Cross Validation.

In standard kF-CV, N_{Tng} training emissions are subdivided into k equal groups, where k is a factor of N_{Tng}. One of the k subgroups is withheld for validation and the remaining $k - 1$ groups are used to generate a model. The withheld validation group is used to evaluate that model's performance and the process is repeated until all k subgroups have been used as the validation group. The model that produces the best validation performance is chosen as the model for classification [95]. This standard kF-CV process was used in prior RF-DNA research [127].

A modified version of kF-CV was also considered for this research in which each of the k subgroups was individually used to develop a model with the remaining $k - 1$ groups withheld for validation. This modified kF-CV process was considered because it is similar to the reference sequence selection process used in prior CBAD research. For the CBAD process, a set of $N_{Pot} = N_{Tng} + 1$ *potential* reference sequences was constructed to consist of the N_{Tng} training sequences and the sequence calculated as the average of the N_{Tng} sequences. Each of the N_{Pot} sequences was considered individually as a potential reference and was evaluated against the remaining $N_{Pot} - 1$ sequences [102].

To determine which k value and kF-CV method provided the best performance, the $N_B = 5000$ emissions per device where evaluated using the matched filter classification methods described in Section 3.4.4 and Section 3.5.3 with reference tem-

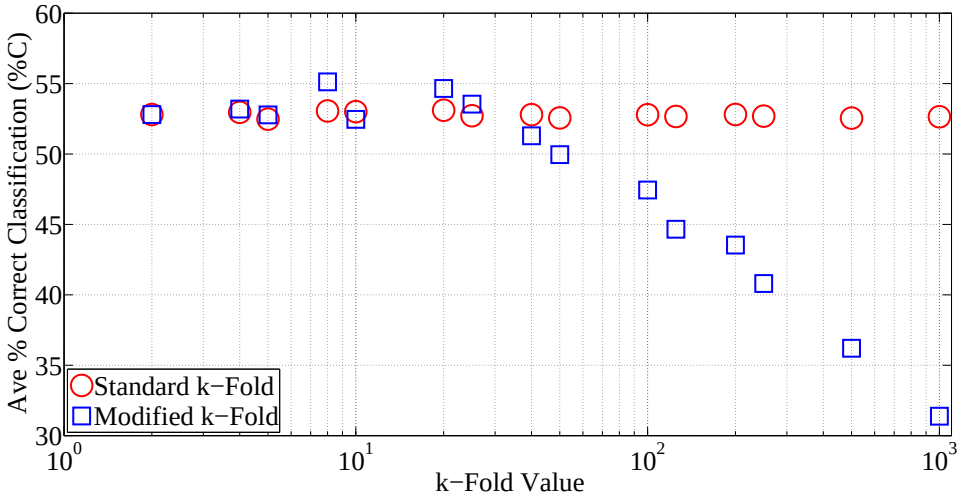

Figure 20. Average matched filtering *device classification* performance for $N_{DevAuth}$=8 authorized device classes. The reference templates were generated from N_{Tng}=1000 emissions per device using standard and modified kF-CV processes with k∈{2,4,5,8,10,20,25,40,50,100,125,200,250,500,1000}. Classification performance was evaluated at the collection SNR, SNR_C, using N_{Tst}=4000 testing emissions per device. Confidence intervals of CI=95% are approximately the same size or smaller than the marker shapes and have been omitted for visual clarity.

plates developed using the standard and modified kF-CV methods for all valid values of $k \in \{2, 4, 5, 8, 10, 20, 25, 40, 50, 100, 125, 200, 250, 500, 1000\}$. The results in Figure 20 show that the standard kF-CV process maintained similar classification performance across all evaluated k values and outperformed the modified kF-CV process for values of $k \leq 40$.

Although the modified kF-CV did perform slightly better than the standard kF-CV process at $k = 8$ and $k = 20$, it did not demonstrate consistently better performance at lower values of $k < 40$. Since the modified kF-CV did not provide a large or consistent performance gain over the standard kF-CV process, the decision was made to use the standard kF-CV process as used in previous AFIT RF-DNA research efforts.

Prior research recommended the use of $k = 5$ or $k = 10$ when the use of repeated cross-validation for larger k values is not feasible [95]. Therefore, the models

for all three evaluated classification methods were generated using the standard kF-CV process with $k = 5$. Although the decision to use standard kF-CV with $k = 5$ was based on the evaluated performance of $N_{Tst} = 4000$ testing emissions, the process of generating and choosing the reference waveforms, given a specific kF-CV method, only considered the $N_{Tng} = 1000$ training emissions to generate and evaluate potential reference signals.

3.4.2 MDA/ML Projection Matrix Generation.

The model for MDA/ML classification is a projection matrix, $\mathbb{W}$, which projects an N_f-dimensional RF-DNA fingerprint vector, $\mathbf{F}$, onto $N_C - 1$ space [74]. For this research, the $N_f = 396$-dimensional fingerprints used for *device discrimination* with $N_{DevAuth} = 8$ classes are projected into $N_{DevAuth} - 1 = 7$-dimensional space. The $N_f = 36$-dimensional fingerprints used for *operation identification* with $N_{Op} = 12$ classes are projected into $N_{Op} - 1 = 11$-dimensional space. The Multiple Discriminant Analysis (MDA) model development process was conducted in accordance with prior AFIT RF-DNA research [14, 29, 127].

The MDA transformation begins by finding the intra-class, $\mathbb{S}_w$, and inter-class, $\mathbb{S}_b$, scatter matrices for a given set of training RF-DNA fingerprint vectors

$$\mathbb{S}_w = \sum_{i=1}^{N_C} P_i \Sigma_i \tag{8}$$

$$\mathbb{S}_b = \sum_{i=1}^{N_C} P_i (\mu_i - \mu_0)(\mu_i - \mu_0)^T \tag{9}$$

where Σ_i is the class covariance matrix and P_i is the prior probability for class C_i [121]. For this research, P_i was assumed to be equal for all classes.

The $N - 1$ eigenvectors of $\mathbb{S}_w^{-1}\mathbb{S}_b$ form the $N_f \times (N_C - 1)$-dimensional projection matrix $\mathbb{W}$ which optimally maximizes the ratio of inter-class distance and intra-

class variance [121]. An individual fingerprint vector, $\mathbf{F}_j$, is then projected onto $(N_C - 1)$-dimensional MDA space by

$$\mathbf{F}_j^{\mathbb{W}} = \mathbb{W}^T \mathbf{F}_j \tag{10}$$

The full MDA-projected training matrix, $\mathbb{F}_T^{\mathbb{W}}$, is the combination of MDA-projected training fingerprints, $\mathbf{F}_j^{\mathbb{W}}$, where $j \in \{1, 2, ..., N_{Tng}\}$

$$\mathbb{F}_T^{\mathbb{W}} = \begin{bmatrix} \mathbf{F}_1^{\mathbb{W}} \\ \mathbf{F}_2^{\mathbb{W}} \\ \vdots \\ \mathbf{F}_{N_{Tng}}^{\mathbb{W}} \end{bmatrix}_{N_{Tng} \times N_f} \tag{11}$$

A multivariate normal distribution is fitted to the MDA-projected data and the estimated distribution parameters (mean vector, $\hat{\mu}_i^{\mathbb{W}}$, and covariance matrix, $\hat{\Sigma}_i^{\mathbb{W}}$) for each class C_i. In accordance with previous efforts, a pooled estimate of the covariance matrix is used instead of individual covariance matrices for each class [121]. The output from the MDA model training process includes the following:

- MDA projection matrix ($\mathbb{W}$)
- N_C sets of MDA-projected training fingerprints ($\mathbb{F}_{C_i}^{\mathbb{W}}$)
- N_C estimated mean vectors ($\hat{\mu}_i^{\mathbb{W}}$)
- pooled estimate of the covariance matrix ($\hat{\Sigma}_i^{\mathbb{W}}$)

The mean vector and covariance matrix for each device is the *reference template* for that device [14].

3.4.3 GRLVQI Prototype Vector Generation.

The GRLVQI model development process was conducted as described in prior work [90] with the goal of defining *classification* boundaries that minimize the Bayes risk. $N_P = 10$ prototype vectors consisting of N_f RF-DNA fingerprint features were defined to represent each of the N_C classes. Matrix $\mathbf{P}$ of dimension $(N_C \cdot N_P) \times N_f$ is formed from the collection of prototype vectors, $\mathbf{p}^n$. The best *in-class*, $\mathbf{p}^I$, and *out-of-class*, $\mathbf{p}^O$, prototype vectors are *differentially shifted* by a distortion value, d_λ^n, computed as [45]

$$d_\lambda^n = \sum_{i=1}^{N_f} \lambda_i (\mathbf{f}_i^m - \mathbf{p}_i^n)^2 \tag{12}$$

where $n \in \{1, 2, \ldots, N_P\}$, $\mathbf{f}^m$ is a randomly selected input fingerprint, $\mathbf{p}^n \in \mathbf{P}$, and λ_i is the relevance weights of the i^{th} feature, normalized such that $||\lambda||_1 = 1$ and $\lambda_i \geq 0 \forall \;\; i \in \{1, \ldots, N_f\}$ [45].

The relevance weight, λ_i, is initialized to a random set of values at the beginning of the classifier training process. To minimize the utilization of poor prototype vectors, a bias parameter, B^n is calculated as

$$B^n = \psi(\frac{1}{N_P} - F_{old}^n) \tag{13}$$

where ψ is a user selected amount to scale the bias parameter and F_{old}^n is the frequency at which a prototype vector is selected as the "best" prototype vector [77]. The bias parameter is applied to the original distortion value by

$$d_{Bias}^n = d_\lambda^n - B^n \tag{14}$$

The best *in-class* prototype vector, $\mathbf{p}^I$, is the $\mathbf{p}^n$ with the *same* class label as the randomly selected $\mathbf{f}^m$ for which d_{Bias}^n is the smallest and the best *out-of-class*

prototype vector, $\mathbf{p}^O$, is the $\mathbf{p}^n$ with a *different* class label than $\mathbf{f}^m$ for which d^n_{Bias} is the smallest. The *in-class* distortion $d^I = d^n_{Bias}$ is the distortion value that resulted in the selection of $\mathbf{p}^I$. Similarly, the *out-of-class* distortion $d^O = d^n_{Bias}$ is the distortion value that resulted in the selection of $\mathbf{p}^O$.

After selecting the best *in-class* and *out-of-class* prototype vectors, the prototype vectors are updated by [45]

$$\mathbf{p}^I(t+1) = \mathbf{p}^I(t) + \frac{4\alpha^I(t) f'|_{\mu(\mathbf{f}^m),\tau} d^O}{(d^I + d^O)^2} \mathbf{\Lambda}(\mathbf{f}^m - \mathbf{p}^I(t)) \tag{15}$$

$$\mathbf{p}^O(t+1) = \mathbf{p}^O(t) + \frac{4\alpha^O(t) f'|_{\mu(\mathbf{f}^m),\tau} d^I}{(d^I + d^O)^2} \mathbf{\Lambda}(\mathbf{f}^m - \mathbf{p}^O(t)) \tag{16}$$

where $\mathbf{\Lambda}_{i,i} = \lambda_i$,$\alpha^I$ and α^O are the learn rates for *in-class* and *out-of-class* prototypes, respectively, and τ is a time decay term [77]. $f'|_{\mu(\mathbf{f}^m),\tau}$ is the first derivative of the sigmoid loss function

$$f(\mu(\mathbf{f}^m), \tau) = \frac{1}{1 + e^{-\tau\mu(\mathbf{f}^m)}} \tag{17}$$

where $\mu(\mathbf{f}^m)$ is the misclassification measure, defined as

$$\mu(\mathbf{f}^m) = (\frac{d^I - d^O}{d^I + d^O}) \tag{18}$$

Correct classification occurs if $\mu(\mathbf{f}^m) < 0$, *misclassification* occurs if $\mu(\mathbf{f}^m) \geq 0$, and $\mu(\mathbf{f}^m) = -1$ corresponds to *perfect* classification [96].

To minimize potential divergence of the prototype vectors, GRLVQI implements a conditional update rule under which the winning *in-class* and *out-of-class* prototype vectors are *only* updated if the input sample is misclassified; otherwise, only the in-class prototype vector is updated [77].

After selecting $\mathbf{p}^I$ and $\mathbf{p}^O$, the α^I and α^O learn rates are adjusted and the λ_i

relevances adjusted using [45]

$$\Delta\lambda_i = -\frac{2\alpha(t)^{\boldsymbol{\lambda}} f'|_{\mu(\mathbf{f}^m),\tau}\left[d^O(\mathbf{f}^m - \mathbf{p}^I)\right]}{(d^I + d^O)^2} + \frac{2\alpha(t)^{\boldsymbol{\lambda}} f'|_{\mu(\mathbf{f}^m),\tau}\left[d^I(\mathbf{f}^m - \mathbf{p}^O)\right]}{(d^I + d^O)^2} \tag{19}$$

The process of distorting prototype vectors and updating the relevances is repeated for $N_I = 600$ iterations, or until other termination criteria are satisfied. Following termination, the output from the GRLVQI model training process includes the following:

- prototype vectors representing the best fit model
- relevance rankings, $\boldsymbol{\lambda}$, associated with best fit prototype vectors

The corresponding "best" *Relevance Vector*, given by

$$\boldsymbol{\lambda}_B = \left[\lambda_1, \lambda_2, \ldots, \lambda_{N_f}\right] \tag{20}$$

can be used for feature Dimensional Reduction Analysis (DRA) as evaluated in [66, 90–93, 102, 127]; however, this research effort did not consider the impact of applying DRA to GRLVQI processing.

3.4.4 Matched Filter Template Generation.

The matched filter reference templates used for *device discrimination* (x_{Auth1}, x_{Auth2}, x_{Auth3}, etc...) and *operation identification* ($x_{\texttt{ADD.B}}$, $x_{\texttt{ADDC.B}}$, $x_{\texttt{AND.B}}$, etc...) were generated for each class by calculating the mean of $N_{kTng} = 800$ TD training emissions belonging to one of the $k = 5$ groups for kF-CV. Although the average power of all collected emissions was normalized in a prior post-processing step, the process of averaging $N_{kTng} = 800$ normalized signals can produce a mean signal with a different average power. Therefore, the average power of the resulting mean signal was

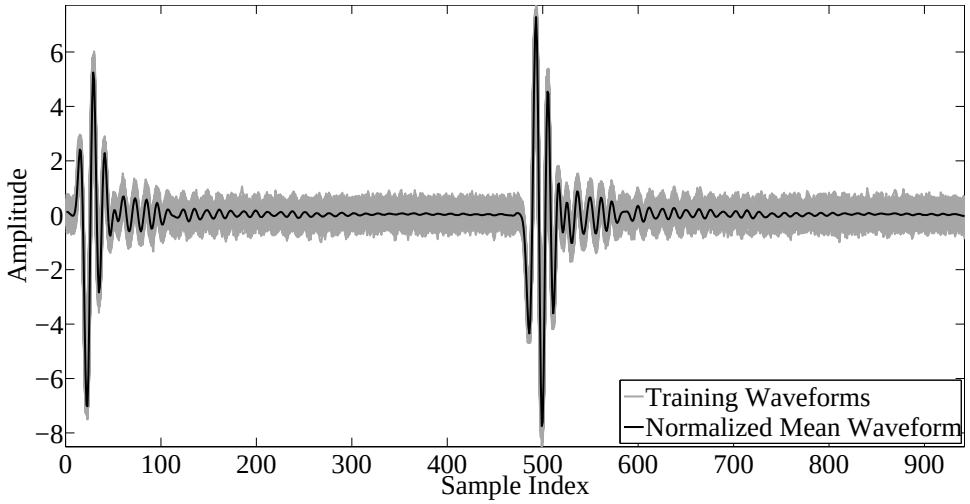

Figure 21. Mean of N_{kTng}=800 training waveforms with normalized average power used as a matched filter template for *operation classification*. The template shown corresponds to the `XOR.B` operation and is representative of other operations.

also normalized so that all of the matched filter template waveforms would have the same average power. Figure 21 shows $N_{kTng} = 800$ emissions (grey) collected from RogueA executing the `XOR.B` operation and the normalized mean waveform (black) used as the matched filter template, $x_{\texttt{XOR.B}}$.

3.5 Classification Evaluation

Classification performs a one-vs.-many comparison of an *unknown* test emission, x_{Tst}, to $N_{DevAuth} = 8$ device classes for *device discrimination* or $N_{Op} = 12$ operation classes for *operation identification*. The classifier generates N_C test statistic values, z_V, corresponding to how much a given test emission resembles each of the N_C possible classes. The class yielding the highest probability is chosen as the estimated device or operation for that test emission. To evaluate the performance of each classification method, $N_{Tst} \times N_{DevAuth} = 4000 \times 8 = 32000$ testing emissions for *device discrimination* and $N_{Tst} \times N_{Op} = 4000 \times 12 = 48000$ testing emissions for *operation identification* were classified at all SNR_A points using the models devel-

oped at $SNR_A = 30$ dB. The percent correct classification ($\%C$) was calculated for the individual classes as well as the average percent correct classification ($\%C_{Ave}$) achieved across all N_C classes.

3.5.1 MDA/ML Classification.

Each RF-DNA fingerprint, $\mathbf{F}$, is classified using MDA/ML by projecting the fingerprint into MDA space using (10) and then computing a measure of similarity comparing the projected fingerprint, $\mathbf{F}^{\mathbb{W}}$, to each of the N_C classes. As in [14], the Bayesian posterior probability was used as the measure of similarity under the assumptions of equal prior probabilities and uniform costs, an approach which optimally minimizes the classification error probability [121]. A projected fingerprint, $\mathbf{F}^{\mathbb{W}}$, is assigned to class w_i where $i \in \{1, 2, ..., N_C\}$

$$P(w_i|\mathbf{F}^{\mathbb{W}}) > P(w_j|\mathbf{F}^{\mathbb{W}}) \quad \forall j \neq i \tag{21}$$

where $P(w_i|\mathbf{F}^{\mathbb{W}})$ is the conditional *posterior probability* that $\mathbf{F}^{\mathbb{W}}$ belongs to class w_i. Using *Baye's Theorem*, the conditional probabilities in (21) can be expressed as

$$P(w_i|\mathbf{F}^{\mathbb{W}}) = \frac{P(\mathbf{F}^{\mathbb{W}}|w_i)P(w_i)}{P(\mathbf{F}^{\mathbb{W}})} \tag{22}$$

Under the assumption of equal prior probabilities for all classes, $P(w_i) = 1/N_C$ remains constant in the numerator of (22). Similarly, the denominator remains constant across all w_i for a given fingerprint, $\mathbf{F}^{\mathbb{W}}$, and can be neglected when evaluating the relative probabilities in (21). Therefore, the decision criteria in (21) can be reduced to maximizing the *likelihood* $P(\mathbf{F}^{\mathbb{W}}|w_i)\ \forall\ w_i$ [14].

3.5.2 GRLVQI Classification.

The GRLVQI *classification* process determines a one-vs.-many "best match" with a minimum Euclidean distance metric that has been successfully used in previous research [45, 77, 90]. Each RF-DNA fingerprint, $\mathbf{F}$, is declared as belonging to class C according to

$$\underset{C}{\operatorname{argmin}} \left(\sqrt{\sum_{i=1}^{N_f} \lambda_i \left(F_i - p_i^{n,C} \right)^2} \right) \tag{23}$$

where F_i is the i^{th} feature of $\mathbf{F}$, $\lambda_i \in \boldsymbol{\lambda}$ is the relevance ranking of the i^{th} feature, and $n \in \{1, 2, ..., N_P\}$ with $\mathbf{p}^{n,C}$ being the n^{th} prototype vector associated with class model C.

3.5.3 Matched Filter Classification.

The matched filter *classification* process uses a bank of parallel matched filters to evaluate an *unknown* test input, x_{Tst}, as illustrated in Figure 22 corresponding to each of the (a) $N_{DevAuth} = 8$ "authorized" device classes or (b) $N_{Op} = 12$ Format I operation classes. The output test statistics, z_V, corresponding to each class are compared and the matched filter which produced the largest z_V value is chosen as the estimated class, C, according to

$$\underset{C}{\operatorname{argmax}} \left(\sum_{k=0}^{N} x_C[k] x_{Tst}[k] \right) \tag{24}$$

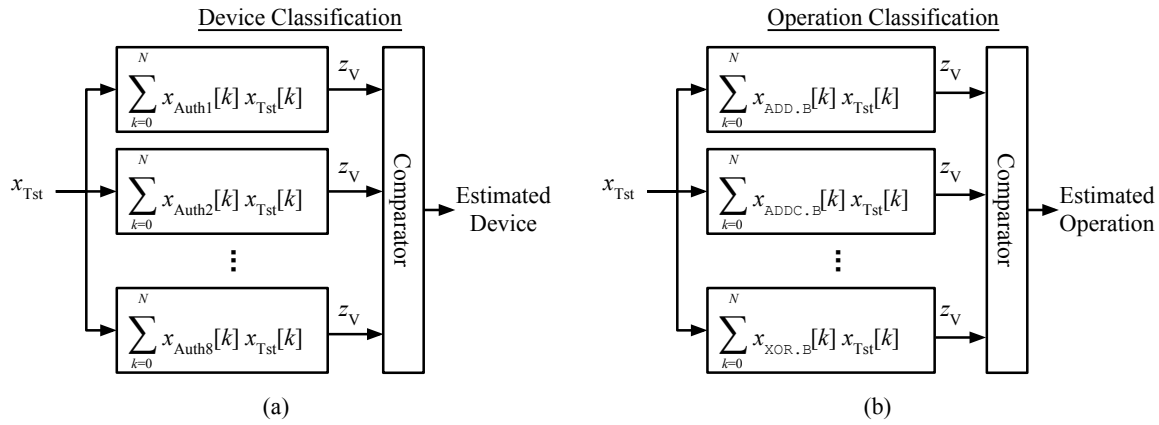

Figure 22. Bank of parallel matched filters used for (a) device discrimination and (b) operation identification.

3.5.4 Confusion Matrices and Performance Curves.

Classification performance is evaluated based on percent correct classification ($\%C$), calculated for each class as the number of correctly classified test emissions divided by the total number of test emissions evaluated for that class. For example, using the representative confusion matrix in Table 12 with $N_C = 3$ classes and $N_{Tst} = 4000$ test emissions per class, the percent correct classification for class C1 is calculated as $\%C = 3801/4000 = 95.0\%$. Similarly, $\%C = 3602/4000 = 90.1\%$ for class C2 and $\%C = 3476/4000 = 86.9\%$ for class C3.

Confusion matrices were generated at each evaluated SNR_A to facilitate analysis of identification errors and have been included in Appendix B for the highest evaluated $SNR_A = 30$ dB. The $\%C$ values calculated from the confusion matrix at each SNR_A were plotted as a function of SNR_A to produce the *classification* performance curves presented in Chapter IV.

Table 12. Representative confusion matrix for N_C=3 classes with N_{Tst}=4000 test emissions per class.

Actual Class \ *Declared Class*	**C1**	**C2**	**C3**
C1	**3801**	127	72
C2	197	**3602**	201
C3	14	510	**3476**

3.6 Verification Evaluation

Verification performs a one-vs.-one comparison to determine how much a device "looks like" a claimed identity, similar to how a person's identity is verified through the use of an ID card. To protect against unauthorized "rogue" devices impersonating authorized devices a verification threshold, t_V, can be used to either grant or deny system access to the device in question. For this research effort the "rogue" devices represent counterfeit chips attempting to pass as authentic hardware. In reality, all evaluated devices were genuine MSP-EXP430F5529 boards and the $N_{DevRogue} = 2$ boards selected to represent "rogue" devices where chosen because they came from a different manufacturing lot and have a different revision MSP430 chip than the other $N_{DevAuth} = 8$ boards representing "authentic" devices.

The MDA/ML, GRLVQI, and matched filtering classifier models are trained using $N_{Tng} = 1000$ training emissions from the $N_{DevAuth} = 8$ authorized devices as previously described. The verification performance is then evaluated by presenting $N_{Tst} = 4000$ previously "unevaluated" testing emissions from the $N_{DevAuth} = 8$ authorized devices as well as $N_{DevRogue} = 2$ rogue devices for verification.

Table 13. Combinations of actual and claimed identities for device verification with the corresponding outcomes and verification rates based on the accept/reject decision. Correct decisions are highlighted in green and incorrect decisions are highlighted in red.

Actual Identity	Claimed Identity	Decision	Outcome	Verification Rate
Auth$X$	Auth$X$	Accepted	True Verification	$TVR = 1 - FRR$
Auth$X$	Auth$X$	Rejected	False Rejection	$FRR = 1 - TVR$
Auth$Y$	Auth$X$	Accepted	False Verification	$FVR = 1 - TRR$
Auth$Y$	Auth$X$	Rejected	True Rejection	$TRR = 1 - FVR$
Rogue$Z$	Auth$X$	Accepted	Rogue Acceptance	$RAR = 1 - RRR$
Rogue$Z$	Auth$X$	Rejected	Rogue Rejection	$RRR = 1 - RAR$

Performance is characterized using the verification rate values described in Table 13. When an authorized device is presented for verification claiming its true identity, the rate at which it is correctly verified and granted access is the True Verification Rate (TVR). The False Rejection Rate (FRR) is equivalent to $1 - TVR$ and corresponds to when an authorized device claiming its true identity is incorrectly rejected and denied system access. The False Verification Rate (FVR) corresponds to when an authorized device is presented for verification claiming the identity of a different authorized device and it is incorrectly granted access. When the verification process correctly detects the identity mismatch and rejects the authorized device claiming a different identity than its own, that corresponds to the True Rejection Rate (TRR) which is equivalent to $1 - FVR$.

The rate at which a previously unevaluated "rogue" device is presented for verification claiming the identity of an authentic device and it is incorrectly accepted is the Rogue Accept Rate (RAR). When the rogue device is correctly rejected that corresponds to the Rogue Rejection Rate (RRR) which is equivalent to $1 - RAR$. The RAR and RRR metrics are used when attempting to verify *rogue devices* and have similar meaning to the FVR and TRR metrics, respectively, which are used when performing verification of *authentic devices*.

This research effort did not perform a verification assessment for "rogue operations" as was done for rogue devices because a MCU can only execute operations from a finite, known set. To verify that a MCU is executing the authorized program, observed URE can be classified as originating from one of the possible operations associated with the authorized program. A sequence of estimated operations can then be compared to the expected sequence of operations to detect anomalous execution.

3.6.1 ROC Curve Generation.

Verification performance is documented using Receiver Operating Characteristics (ROC) curves which illustrate the performance of a binary classifier (accept/reject) as the decision threshold, t_V, is varied. ROC curves are generated by plotting TVR vs. FVR when evaluating verification of *authorized devices* and TVR vs. RAR when evaluating the rejection of *rogue devices*. The TVR is calculated using the Probability Mass Function (PMF) of z_V values from the comparison of an authentic device vs. its own identity (Auth$X$ vs. Auth$X$). A representative PMF for an "A vs. A" comparison is shown in blue and red at the top of Figure 23a. The verification threshold, t_V, is varied across the range of PMF bins and the percentage of z_V values that are accepted at a given t_V value correspond to the TVR at that threshold.

A similar process is used to calculate FVR from the PMF comparing other authentic devices vs. the claimed identity (Auth Y vs. Auth$X$) and to calculate RAR from the PMF comparing a rogue device vs. the claimed identity of an authentic device (Rogue$Z$ vs. Auth$X$). A representative PMF for "B vs. A" comparison is shown in yellow and green at the bottom of Figure 23a. ROC curves having a higher TVR and a lower FVR/RAR correspond to better performance. The point

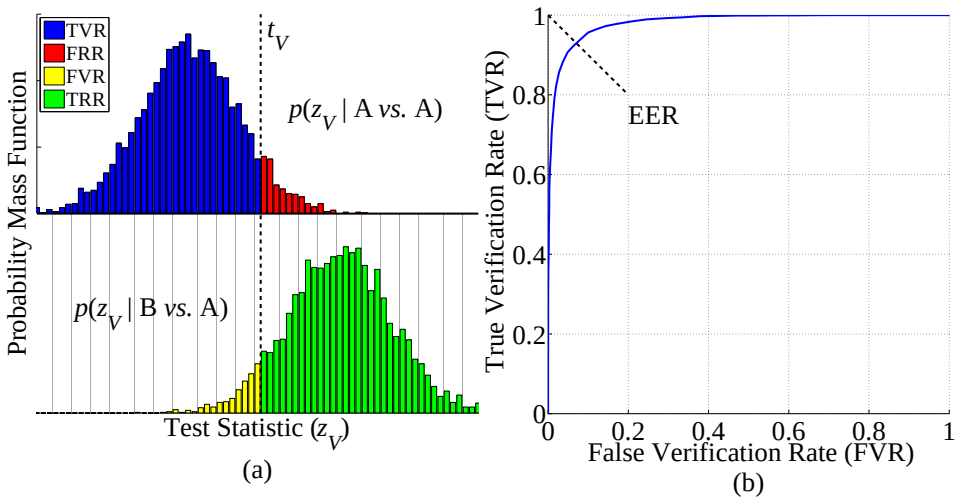

Figure 23. Representative (a) PMFs of test statistic, z_V, values for "A vs. A" (blue/red) and "B vs. A" (yellow/green) comparisons with (b) the corresponding ROC curve for *authorized device verification*. The verification rate regions are color coded with accepted z_V values on the left side of t_V and rejected z_V values on the right side of t_V. The shown threshold, t_V, corresponds to the EER.

at which the FRR is equivalent to the FVR/RAR is the Equal Error Rate (EER). The EER provides a single metric with which to evaluate the *verification* process, with a lower EER representing better performance [56].

3.7 Computational Time Evaluation

To compare the computational costs of the three classification methods, the execution time was recorded for the three shaded process regions in Figure 24. These three regions were selected for comparison because they represent portions of the classification process that differ between the three methods. The steps to remove DC bias, generate like-filtered AWGN, and normalize signal power were not considered for computational cost comparison because they are identical for all three methods. The computational times presented in Chapter IV represent the elapsed time to complete each processing region at a single SNR_A. The times were recorded while processing a "batch" of emissions as described in this section rather than a

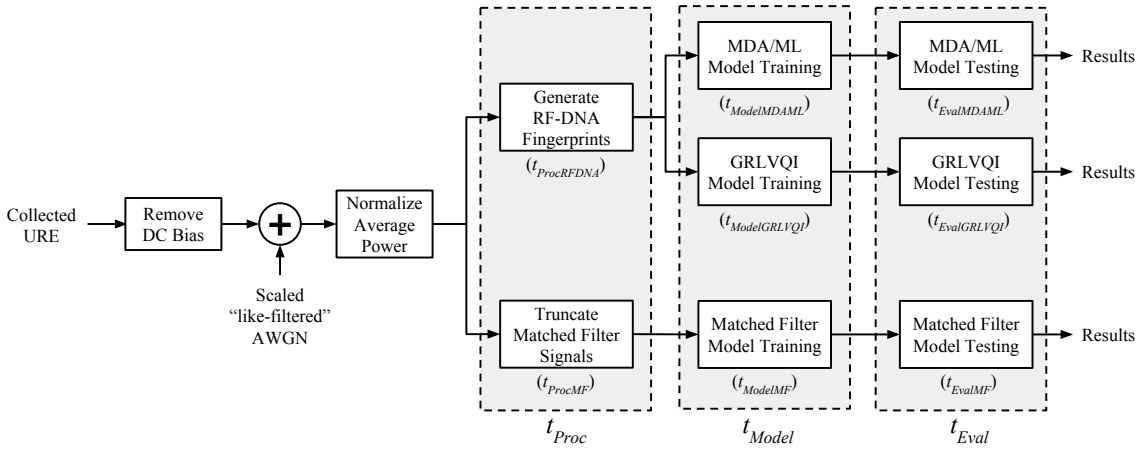

Figure 24. Block diagram highlighting the three process regions evaluated for computational time: post collection signal processing (t_{Proc}), model development (t_{Model}), and model evaluation (t_{Eval}).

single emission because the time required to process a single emission was too small to measure with useful accuracy in the MATLAB® environment.

The signal processing (t_{Proc}) region of Figure 24 contains the method-specific post-collection processing applied to the normalized signals. For the RF-DNA methods, post-collection processing ($t_{ProcRFDNA}$) includes the process of dividing the signal ROI into N_{SR} sub-regions and calculating the statistical values for amplitude, phase, and frequency to generate the RF-DNA fingerprint vector. For matched filtering, the post-collection processing (t_{ProcMF}) includes truncating signals at rising and falling clock edges to be the same length. The execution time for the t_{Proc} region was evaluated to complete the processing of $N_B \times N_{DevAuth} = 5000 \times 8 = 40000$ emissions for *device discrimination* and $N_B \times N_{Op} = 5000 \times 12 = 60000$ emissions for *operation identification*.

The model development (t_{Model}) region of Figure 24 corresponds to the k-fold model development phase using $N_{Tng} = 1000$ fingerprints/waveforms for each of $N_{DevAuth} = 8$ authorized devices or $N_{Op} = 12$ operations. The MDA/ML model development process ($t_{ModelMDAML}$) used a standard kF-CV process with $k = 5$ to

generate a projection matrix. The GRLVQI model development process ($t_{ModelGRLVQI}$) used a standard kF-CV process with $k = 5$ to generate prototype vectors. The matched filtering model development process ($t_{ModelMF}$) used a standard kF-CV process with $k = 5$ to generate reference templates.

The model evaluation (t_{Eval}) region of Figure 24 corresponds to the performance evaluation phase which uses the model developed for each classification method to evaluate $N_{Tst} \times N_{DevAuth} = 4000 \times 8 = 32000$ emissions for *device discrimination* and $N_{Tst} \times N_{Op} = 4000 \times 12 = 48000$ emissions for *operation identification*.

The "total computational time" was calculated as $t_{TotalMDAML} = t_{ProcRFDNA} + t_{ModelMDAML} + t_{EvalMDAML}$ for RF-DNA fingerprint generation and MDA/ML classification, $t_{TotalGRLVQI} = t_{ProcRFDNA} + t_{ModelGRLVQI} + t_{EvalGRLVQI}$ for RF-DNA fingerprint generation and GRLVQI classification, and $t_{TotalMF} = t_{ProcMF} + t_{ModelMF} + t_{EvalMF}$ for TD signal truncation and matched filtering classification.

When implementing these classification techniques in a real-world monitoring system the time required for model development is of less importance because the model is developed "offline" using training emissions prior to system deployment. The computational cost to process future observed URE and evaluate them with the previously developed model provide a better representation of the computational cost in a real-time monitoring system. Therefore, an "alternative computational time" ($t_{AltTotal} = t_{Proc} + t_{Eval}$) was also calculated using the corresponding regions for each method which excluded the model development time, representing an "as deployed" solution.

The execution times for each process region were recorded using the MATLAB® `tic` and `toc` functions. All processing was executed using MATLAB® 2014b on a desktop computer running Microsoft Windows 7 Enterprise Edition 64-bit Operating System (OS) with 192 GB of RAM and dual Intel® Xeon® E5-2697 v2 pro-

cessors (24-cores/48-threads total) clocked at 2.70 GHz. Since Windows is a non-deterministic OS, all benchmark tests were repeated five times and the median time was selected as the result to mitigate the impact of variations in the recorded time due to OS process scheduling.

The execution time of a computer program is dependent on how it is implemented. To make the comparison of execution time between classification methods "fair," efforts were made to make the software implementations for each method similar where possible. The code used for the MDA/ML and GRLVQI model development and evaluation was derived from the common body of MATLAB® code used for prior AFIT RF-DNA research whereas the matched filtering model development and evaluation code was uniquely created for this effort. A custom RF-DNA fingerprint generation program was used for this research which excluded the use of the MATLAB® `parfor` loops which were used in the existing RF-DNA fingerprint code and mirrored the program structure used in the signal truncation routine for matched filtering. This simpler program structure without parallel for loops is similar to what might be required to implement these algorithms in a deployed solution using low-cost embedded hardware with limited processing power.

IV. Results

This chapter provides research results for *device classification* in Section 4.1 and *operation classification* in Section 4.2 which were generated using the methodology described in Chapter III. Section 4.3 presents *device ID verification* results and Section 4.4 compares the computational time required for each of the three evaluated classification methods.

4.1 Device Classification

Average *device classification* performance ($\%C_{Ave}$) across $N_C = N_{DevAuth} = 8$ authorized device classes is presented in Figure 25 using Radio-Frequency (RF) Distinct Native Attribute (RF-DNA) features paired with Multiple Discriminant Analysis/Maximum Likelihood (MDA/ML) classification, RF-DNA features paired with Generalized Relevance Learning Vector Quantized-Improved (GRLVQI) classification, and Time Domain (TD) signals with matched filtering classification. Confidence intervals of $CI = 95\%$ were calculated and used to evaluate the statistical significance of classification results; however, they have been omitted from plots to improve visual clarity.

Using RF-DNA with MDA/ML resulted in average classification performance $\%C_{Ave} \geq 75.38\%$ for $SNR_A \geq 20$ dB. Using RF-DNA with GRLVQI resulted in average classification performance $\%C_{Ave} \geq 60.85\%$ for $SNR_A \geq 20$ dB. Using matched filtering resulted in average classification performance $\%C_{Ave} \geq 52.38\%$ for $SNR_A \geq 20$ dB. RF-DNA with MDA/ML had the best average classification performance for $SNR_A \geq 0$ dB. However, matched filtering had the best average classification performance for low Signal-to-Noise Ratio (SNR) of $SNR_A < 0$ dB, possibly due to its optimal performance in the presence of noise [89, 99].

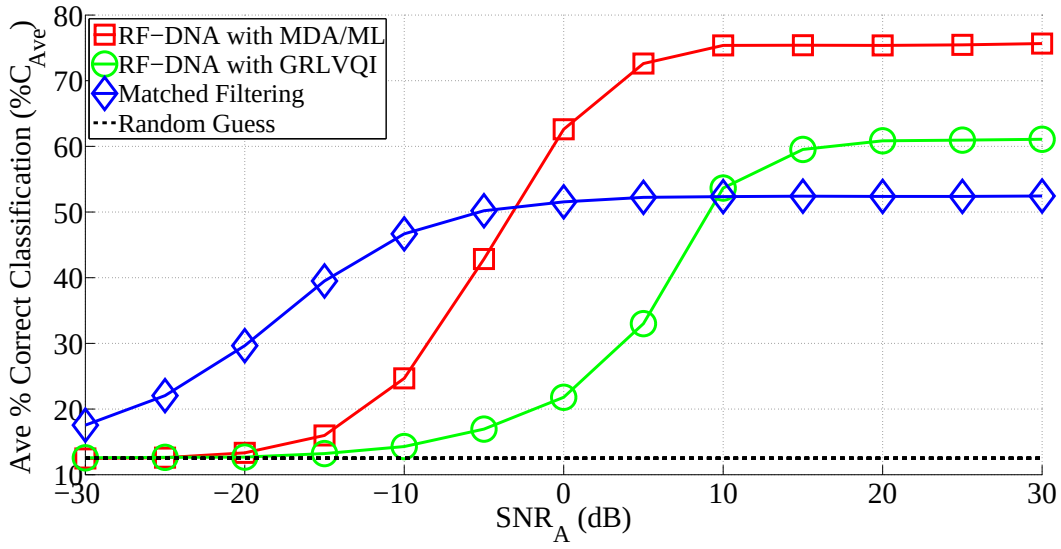

Figure 25. Average *device classification* performance ($\%C_{Ave}$) for N_C=$N_{DevAuth}$=8 devices using RF-DNA with MDA/ML, RF-DNA with GRLVQI, and matched filtering techniques to evaluate N_{Tst}=4000 testing emissions per device class. Confidence intervals of CI=95% are approximately the same size or smaller than the marker shapes and have been omitted for visual clarity. RF-DNA with MDA/ML achieved the best performance of the techniques considered for $SNR_A \geq 0$ dB.

Additional plots showing the *device classification* performance of individual device classes and the confusion matrices at $SNR_A = 30$ dB for each classification method are included in Appendix B, but have been omitted here for brevity. Table 14 provides a comparison of the individual class results at the highest evaluated $SNR_A = 30$ dB and highlights the classes that achieved better than "random guess" classification performance ($\%C_{Random} = 12.5\%$) for each method. All 8 devices were classified with $\%C \geq 12.5\%$ for all three evaluated classification techniques.

For all three classification methods, the Device Under Test (DUT) named "Auth1" achieved the highest percent correct classification out of the $N_{DevAuth} = 8$ evaluated devices. It is worth noting that Auth1 was the DUT used during the early investigation portion of this research effort and had therefore spent many more hours powered up and operating than the other DUTs. This research effort did not

Table 14. Comparison of *device classification* performance for N_C=$N_{DevAuth}$=8 devices using RF-DNA with MDA/ML, RF-DNA with GRLVQI, and matched filtering techniques to evaluate N_{Tst}=4000 testing emissions per device class at SNR_A=30 dB. Cells are colored based on whether they are greater than (green) or less than (red) "random guess" $\%C_{Random}$=12.50%. RF-DNA with MDA/ML achieved the best performance with $\%C_{Ave}$=75.66%

	RF-DNA w/ MDA/ML	RF-DNA w/ GRLVQI	Matched Filtering
Auth1	97.40%	88.78%	71.75%
Auth2	77.35%	64.78%	59.63%
Auth3	78.78%	55.48%	31.18%
Auth4	71.28%	61.63%	56.80%
Auth5	80.73%	70.65%	62.25%
Auth6	61.96%	41.93%	43.20%
Auth7	80.83%	73.23%	33.00%
Auth8	56.95%	32.18%	61.78%
$\%C_{Ave}$	75.66%	61.08%	52.45%

maintain formal device usage logs; however, as an estimate, Auth1 had *at least* two months worth of non-consecutive lifetime usage whereas the other devices each had less than 24 hours of non-consecutive lifetime usage prior to acquiring the Unintentional RF Emissions (URE) used to generate the results presented here. As semiconductor devices age, internal changes such as electro-migration, time dependent dielectric breakdown [87], and bias temperature instability [57, 71] alter a device's switching characteristics. The longer operating time of Auth1 may have influenced its physical characteristics, and thus URE, in a way that made it "more unique" and distinguishable from the other devices.

Another possible reason for the higher classification performance of Auth1 is that it was a member of the set of devices (Auth1, Auth2, RogueA) used to determine the collection location for *device discrimination* as described in Section 3.2. The influence of Auth1 in that collection location decision may have contributed to its better performance. However, Auth2 was also part of the set used to determine the collection location and its performance was near the middle of the $N_{DevAuth} = 8$

evaluated devices. If the device's operational age impacted URE as previously suggested, Auth2 would have been representative of the other devices (Auth3-8) which had been used for a similar amount of time. Therefore, the process used to choose a collection location may have selected the location that maximized the variance between "old" devices (Auth1) and "young" devices (Auth2-8). The DUT named "RogueA" was also included in the set of devices used to choose a collection location; however, excluding it from the set did not change the chosen location which maximized inter-device variance.

Table 15 contains the normalized cross-correlation ($C_{X,Y}$) of the $N_{DevAuth} = 8$ reference templates used for matched filter *device classification*. This provides insight into how much the template for each device class "looks like" the templates for the other classes, with a lower value representing a greater difference and more distinguishability between classes. The DUTs Auth7 and Auth8 looked the most alike with a normalized cross-correlation $C_{Auth7,Auth8} = 0.9879$. The DUTs Auth1 and Auth5 were the most different from each other with a normalized cross-correlation $C_{Auth1,Auth5} = 0.8300$. As expected due to its superior classification performance, the template for Auth1 was the most unique by having the lowest average normalized cross-correlation with other devices of $C_{Ave} = 0.9077$.

Table 15. Normalized cross-correlation ($\mathbf{C_{X,Y}}$) of *matched filter* reference templates used for *device classification* of N_C=$N_{DevAuth}$=8 devices. Reference templates were generated from N_{Tng}=1000 training emissions per device using standard k-fold cross-validation with k=5.

	Auth1	**Auth2**	**Auth3**	**Auth4**	**Auth5**	**Auth6**	**Auth7**	**Auth8**
Auth1	**1**	.9363	.8608	.8536	.8300	.9346	.9747	.9636
Auth2	.9363	**1**	.9600	.9680	.9456	.9724	.9534	.9430
Auth3	.8608	.9600	**1**	.9720	.9861	.9395	.8875	.8661
Auth4	.8536	.9680	.9720	**1**	.9805	.9438	.8843	.8804
Auth5	.8300	.9456	.9861	.9805	**1**	.9344	.8635	.8456
Auth6	.9346	.9724	.9395	.9438	.9344	**1**	.9689	.9579
Auth7	.9747	.9534	.8875	.8843	.8635	.9689	**1**	.9879
Auth8	.9636	.9430	.8661	.8804	.8456	.9579	.9879	**1**

4.2 Operation Classification

Average *operation classification* performance ($\%C_{Ave}$) across $N_C = N_{Op} = 12$ Format I operation classes is presented in Figure 26 for the RF-DNA with MDA/ML, RF-DNA with GRLVQI, and matched filtering classification techniques. Confidence intervals of $CI = 95\%$ were calculated and used to evaluate the statistical significance of classification results; however, they have been omitted from plots to improve visual clarity.

Using RF-DNA with MDA/ML resulted in average classification performance $\%C_{Ave} \geq 10.73\%$ for $SNR_A \geq 20$ dB. Using RF-DNA with GRLVQI resulted in average classification performance $\%C_{Ave} \geq 8.62\%$ for $SNR_A \geq 20$ dB. Using matched filtering resulted in average classification performance $\%C_{Ave} \geq 11.09\%$ for $SNR_A \geq 20$ dB. Matched filtering had the best average classification performance for all evaluated $SNR_A \in [-30, 30]$ dB.

Additional plots showing the *operation classification* performance of individual operation classes and the confusion matrices at $SNR_A = 30$ dB for each classifica-

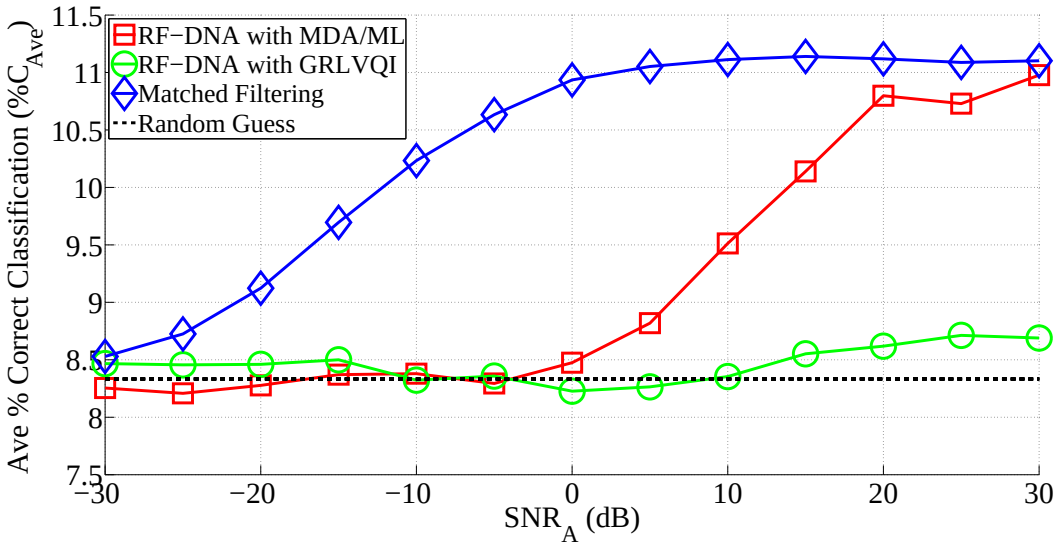

Figure 26. Average *operation identification* performance ($\%C_{Ave}$) for N_C=N_{Op}=12 operations using RF-DNA with MDA/ML, RF-DNA with GRLVQI, and matched filtering techniques to evaluate N_{Tst}=4000 testing emissions per operation class. Confidence intervals of CI=95% are approximately the same size or smaller than the marker shapes and have been omitted for visual clarity. Matched filtering achieved the best performance of the techniques considered for $SNR_A \in [-30, 30]$ dB.

tion method are included in Appendix B, but have been omitted here for brevity. Table 16 provides a comparison of the individual class results at the highest evaluated $SNR_A = 30$ dB and highlights the classes that achieved better than "random guess" classification performance ($\%C_{Random} = 8.33\%$) for each method. Using RF-DNA with MDA/ML, all 12 operations were classified with $\%C \geq 8.33\%$. Using RF-DNA with GRLVQI, 5 out of 12 operations were classified with $\%C \geq 8.33\%$. Using matched filtering, 3 out of 12 operations were classified with $\%C \geq 8.33\%$.

There was not a single operation class that consistently achieved better classification performance than the other classes for all of the evaluated methods. RF-DNA with MDA/ML achieved the best classification performance for `BIT.B` with $\%C = 13.83\%$. RF-DNA with GRLVQI achieved the best classification performance for `SUBC.B` with $\%C = 19.15\%$. Matched filtering achieved the best classification performance for `CMP.B` with $\%C = 54.70\%$.

Table 16. Comparison of *operation classification* performance for N_C=N_{Op}=12 operations using RF-DNA with MDA/ML, RF-DNA with GRLVQI, and matched filtering techniques to evaluate N_{Tst}=4000 testing emissions per operation class at SNR_A=30 dB. Cells are colored based on whether they are greater than (green) or less than (red) "random guess" $\%C_{Random}$=8.33%. Matched filtering achieved the best performance with $\%C_{Ave}$=11.10%

	RF-DNA w/ MDA/ML	RF-DNA w/ GRLVQI	Matched Filtering
ADD.B	12.25%	15.40%	1.90%
ADDC.B	11.7%	7.28%	0.78%
AND.B	12.78%	6.98%	0.73%
BIC.B	9.05%	3.98%	10.20%
BIS.B	13.68%	8.23%	1.70%
BIT.B	13.83%	3.20%	0.85%
CMP.B	8.68%	9.43%	54.70%
DADD.B	9.60%	5.18%	5.78%
MOV.B	9.40%	8.93%	50.33%
SUB.B	12.08%	9.15%	0.90%
SUBC.B	9.15%	19.15%	4.88%
XOR.B	9.55%	7.40%	0.50%
$\%C_{Ave}$	10.98%	8.69%	11.10%

The classification performance for individual operations using RF-DNA with MDA/ML and RF-DNA with GRLVQI did not have any apparent outliers and remained within roughly 10% of the average classification performance. Matched filtering had two outlier classes, `CMP.B` and `MOV.B`, that achieved roughly 40% better than average classification while the remainder of the operations achieved below average performance. Examination of the confusion matrix for matched filtering in Table 17 revealed that a significant number of test emissions belonging to the ten lowest performing operation classes were incorrectly declared as `CMP.B` or `MOV.B`.

Figure 27 shows the $N_{Op} = 12$ reference templates used for matched filter *operation classification*. At the macro level, shown in Figure 27a, the $N_{Op} = 12$ reference templates appear to be nearly identical to each other. Closer inspection of the waveform peaks, as shown in Figure 27b, revealed that the rising edge of the second peak of the second half-clock cycle of the `MOV.B` operation occurred *before* the

Table 17. Confusion matrix for *operation classification* of N_C=N_{Op}=12 operations using *matched filtering* to evaluate N_{Tst}=4000 testing emissions per operation at SNR_A=30 dB.

Actual Class \ *Declared Class*	ADD.B	ADDC.B	AND.B	BIC.B	BIS.B	BIT.B	CMP.B	DADD.B	MOV.B	SUB.B	SUBC.B	XOR.B
ADD.B	**76**	15	12	242	51	33	1438	153	1805	25	131	19
ADDC.B	89	**31**	28	313	77	32	1783	200	1218	26	184	19
AND.B	93	33	**29**	322	85	42	1911	226	1018	23	199	19
BIC.B	110	31	25	**408**	82	42	2184	228	620	30	215	25
BIS.B	76	24	36	299	**68**	41	1821	190	1214	31	182	18
BIT.B	77	31	22	372	77	**34**	1901	206	1024	32	198	26
CMP.B	91	24	29	389	87	41	**2188**	269	619	29	215	19
DADD.B	94	31	20	303	67	36	1783	**231**	1219	23	175	18
MOV.B	64	22	11	221	50	25	1303	137	**2013**	10	133	11
SUB.B	95	22	25	312	75	31	1984	199	1012	**36**	183	26
SUBC.B	98	23	22	362	92	41	2077	232	821	22	**195**	15
XOR.B	93	17	15	314	69	38	1811	201	1209	27	186	**20**

other operations and the falling edge of the CMP.B occurred *after* the other operations. Additionally, MOV.B and CMP.B have the greatest amplitude at samples 505 and 506, respectively. Although Figure 27b only shows a single peak, this trend was present on the other major peaks as well. To verify that the MOV.B and CMP.B operations were not "randomly" at the two extremes due to the specific set of training emissions that was used, reference templates were also generated using the same process with four other independent sets of N_{Tng} = 1000 training emissions per operation, drawn from the group previously set aside as testing emissions. In all four cases, the relative position of the peaks was similar to the templates generated with original training set.

Since the templates for all N_{Op} = 12 Format I operations had the same general shape, as shown in Figure 27a, if the peaks for an unknown testing emission occurred *sooner* than the peaks in the MOV.B template, possibly due to noise or clock jitter, then that emission would likely achieve a higher correlation with the MOV.B

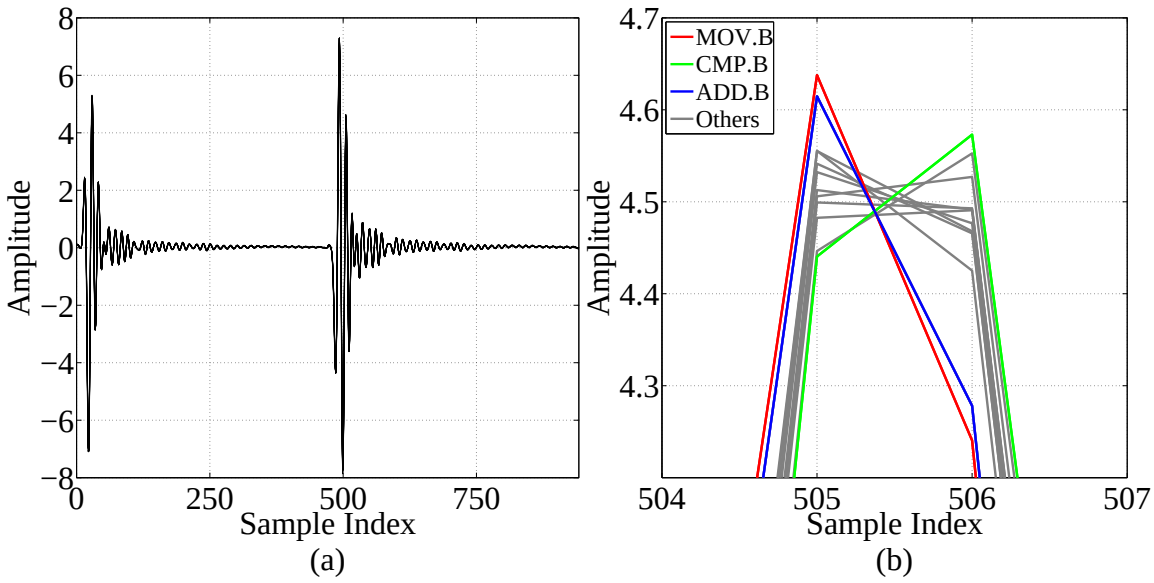

Figure 27. Matched filter reference templates for N_{Op}=12 operation classes showing (a) the complete templates and (b) an expanded view of the second peak of the second half-clock cycle.

template than with the others. Similarly, if the peaks for an unknown testing emission occurred *later* than the peaks in the `CMP.B` template then that emission would likely achieve a higher correlation with the `CMP.B` template than with the others. This potential bias towards the `MOV.B` and `CMP.B` operations due to their extreme positions relative to the other templates may have contributed to higher classification performance.

Table 18 contains the normalized cross-correlation ($C_{X,Y}$) of the $N_{Op} = 12$ reference templates used for matched filter *operation classification*. The `BIC.B` and `CMP.B` operations looked the most like each other with a normalized cross-correlation of $C_{\texttt{BIC.B,CMP.B}} = 0.9999$. The `CMP.B` and `MOV.B` operations were the most different from each other with a normalized cross-correlation of $C_{\texttt{CMP.B,MOV.B}} = 0.9869$.

The template for `MOV.B` had the lowest average normalized cross-correlation $C_{Ave} = 0.9933$ and the template for `ADD.B` had the second lowest average normalized cross-correlation $C_{Ave} = 0.9953$. As visible in Figure 27b, the `MOV.B` and `ADD.B`

Table 18. Normalized cross-correlation ($C_{X,Y}$) of *matched filter* reference templates used for *operation classification* of N_C=N_{Op}=12 operations. Reference templates were generated from N_{Tng}=1000 training emissions per operation using standard k-fold cross-validation with k=5.

	ADD.B	ADDC.B	AND.B	BIC.B	BIS.B	BIT.B	CMP.B	DADD.B	MOV.B	SUB.B	SUBC.B	XOR.B
ADD.B	1	.9974	.9956	.9905	.9968	.9953	.9901	.9977	.9997	.9948	.9938	.9970
ADDC.B	.9974	1	.9997	.9978	.9999	.9996	.9975	.9999	.9957	.9995	.9992	.9999
AND.B	.9956	.9997	1	.9990	.9998	.9999	.9988	.9996	.9934	.9999	.9998	.9998
BIC.B	.9905	.9978	.9990	1	.9982	.9991	.9999	.9974	.9875	.9993	.9996	.9981
BIS.B	.9968	.9999	.9998	.9982	1	.9998	.9980	.9998	.9950	.9996	.9994	.9999
BIT.B	.9953	.9996	.9999	.9991	.9998	1	.9989	.9995	.9931	.9999	.9998	.9997
CMP.B	.9901	.9975	.9988	.9999	.9980	.9989	1	.9972	.9869	.9992	.9995	.9979
DADD.B	.9977	.9999	.9996	.9974	.9998	.9995	.9972	1	.9960	.9993	.9990	.9999
MOV.B	.9997	.9957	.9934	.9875	.9950	.9931	.9869	.9960	1	.9924	.9913	.9952
SUB.B	.9948	.9995	.9999	.9993	.9996	.9999	.9992	.9993	.9924	1	.9999	.9996
SUBC.B	.9938	.9992	.9998	.9996	.9994	.9998	.9995	.9990	.9913	.9999	1	.9993
XOR.B	.9970	.9999	.9998	.9981	.9999	.9997	.9979	.9999	.9952	.9996	.9993	1

templates were closely grouped together and separated from the other ten operation templates which were also closely grouped. Even though the MOV.B and ADD.B resembled each other with $C_{\texttt{MOV.B,ADD.B}} = 0.9997$, their distance from the rest of the operation templates gave them the lowest average cross-correlation. However, despite having the second lowest average cross-correlation, the ADD.B operation achieved poor matched filtering classification performance $\%C = 1.90\%$. This poor performance of the ADD.B operation may be due to the previously mentioned matched filter bias towards the MOV.B operation due to its earlier peak position.

4.3 Device ID Verification

Device verification results are presented using Receiver Operating Characteristics (ROC) curves which were generated at the highest evaluated $SNR_A = 30$ dB

operating point using the process described in Section 3.6.1. The notation A:B is used in the plot legends to indicate how much device A "looks like" device B, where A is the actual identity and B is the claimed identity. The ROC curves for *authorized device verification* represent how much each of the $N_{DevAuth} = 8$ authorized devices "looks like" itself (Auth:Auth). *Rogue device rejection* performance is presenting using separate ROC curves comparing each of the $N_{DevRogue} = 2$ "rogue" devices to each of the $N_{DevAuth} = 8$ authorized devices (Rogue:Auth), all of which were actually authentic MSP-EXP430F5529 boards from differing manufacturing lots representing "authorized" and "rogue" devices.

This research assessed ROC curves for *authorized device verification* to represent "successful verification" using the benchmark of $TVR > 90\%$ and $FVR < 10\%$, which is equivalent to $EER \leq 10\%$. Similarly, the ROC curves for *rogue device rejection* were assessed to represent "successful rejection" when $TVR > 90\%$ and $RAR < 10\%$. This $EER \leq 10\%$ benchmark was used to maintain consistency with prior Air Force Institute of Technology (AFIT) RF-DNA research efforts [102, 127], allowing a comparison of results. The region corresponding to $EER \leq 10\%$ in the presented plots has been outlined with a red box. Solid ROC curve lines indicate A:B pairs which met the benchmark of $EER \leq 10\%$ and dashed lines indicate A:B pairs which failed to meet the benchmark. Figure 28 shows an overlay of all *authorized device verification* ROC curves using RF-DNA with MDA/ML (red), RF-DNA with GRLVQI (green), and matched filtering (blue) for comparison.

The Equal Error Rate (EER) when verifying each authorized device is given in Table 19 along with the average EER_{Ave} achieved across all $N_{DevAuth} = 8$ authorized device classes. Using RF-DNA with MDA/ML, one device was successfully verified with $EER \leq 10\%$ and the average EER was $EER_{Ave} = 16.23\%$. Using RF-DNA with GRLVQI, none of the devices were successfully verified with

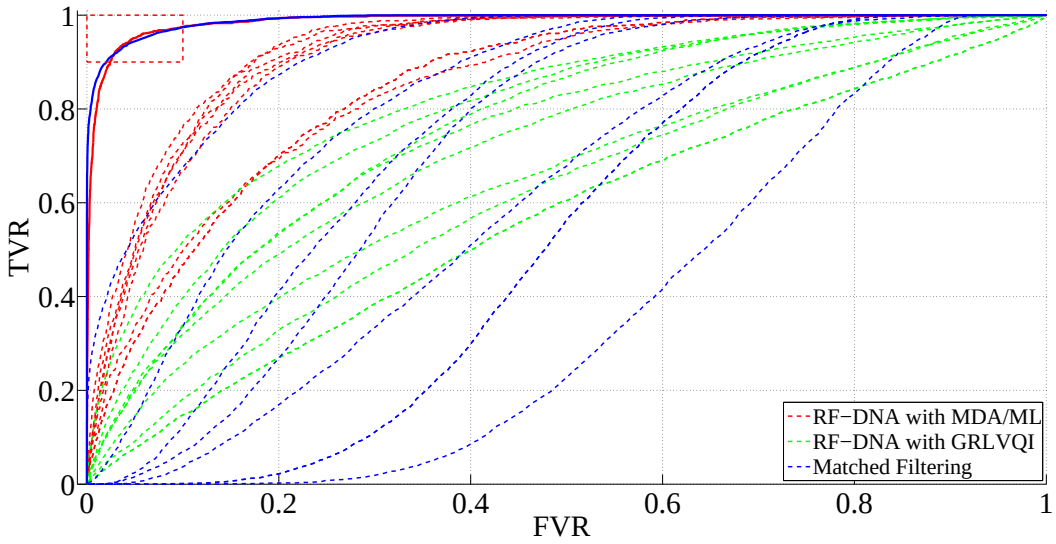

Figure 28. Comparison of *authorized device verification* performance for $N_{DevAuth}$=8 authentic devices using RF-DNA with MDA/ML, RF-DNA with GRLVQI, and matched filtering techniques to evaluate N_{Tst}=4000 testing emissions per device at SNR_A=30 dB. The arbitrary benchmark of $EER \leq 10\%$, outlined in the top-left corner, was met by one device using RF-DNA with MDA/ML and one device using matched filtering.

Table 19. Comparison of EER and average EER_{Ave} for *authentic device verification* using RF-DNA with MDA/ML, RF-DNA with GRLVQI, and matched filtering techniques. Results obtained from N_{Tst}=4000 emissions per device at SNR_A=30 dB. Cells corresponding to "successful verification" ($EER \leq 10\%$) have been highlighted in green. RF-DNA with MDA/ML achieved the best performance with an average EER of EER_{Ave}=16.23%.

	RF-DNA w/ MDA/ML	RF-DNA w/ GRLVQI	Matched Filtering
Auth1:Auth1	4.91%	25.89%	5.28%
Auth2:Auth2	14.48%	31.39%	25.72%
Auth3:Auth3	15.37%	28.35%	17.22%
Auth4:Auth4	16.62%	39.35%	34.15%
Auth5:Auth5	14.74%	33.73%	31.60%
Auth6:Auth6	24.39%	41.93%	43.14%
Auth7:Auth7	15.52%	31.05%	59.32%
Auth8:Auth8	23.84%	44.78%	48.19%
Average	16.23%	34.56%	33.08%

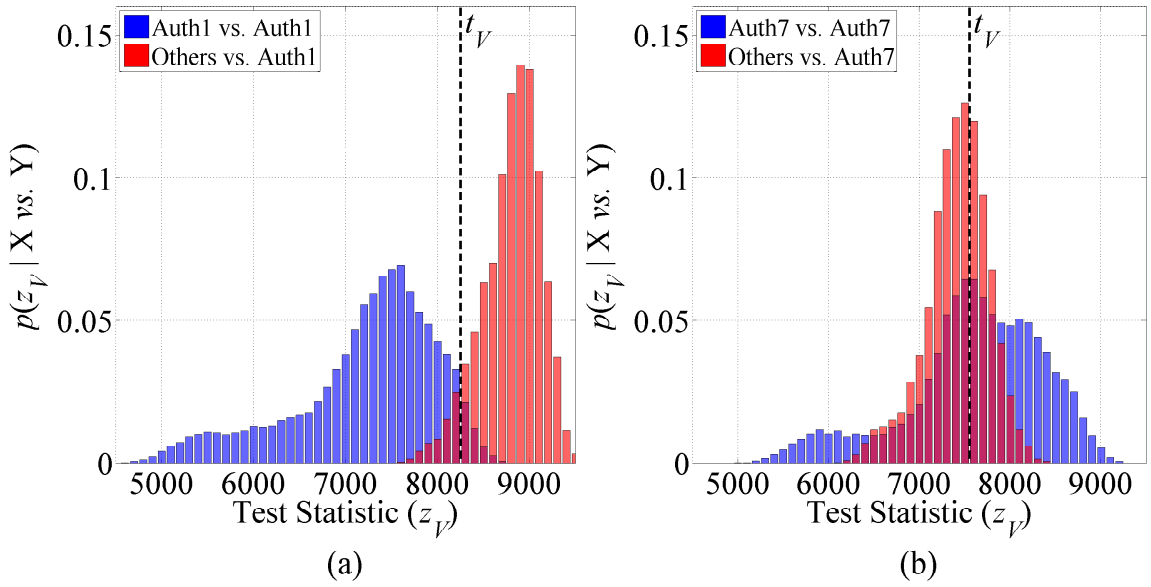

Figure 29. Normalized PMFs showing the distribution of test statistic z_V values used to evaluate *authorized device verification* with matched filtering for (a) Auth1, which achieved the best performance with EER=5.28%, and (b) Auth7, which achieved the worst performance with EER=59.32%. The shown thresholds, t_V, correspond to the EER with accepted z_V values on the right side of t_V and rejected z_V values on the left side.

$EER \leq 10\%$ and the average EER was $EER_{Ave} = 34.56\%$. Using matched filtering, one device was successfully verified with $EER \leq 10\%$ and the average EER was $EER_{Ave} = 33.08\%$.

Auth1 was the only device to be successfully verified with $EER \leq 10\%$ and it achieved the lowest EER of all $N_{DevAuth} = 8$ for all three evaluated methods. This remains consistent with the observation made during *device classification* that Auth1 achieved the best *classification performance*, possibly because it had spent more hours powered and operating than the other DUTs which may have influenced its URE in a way that differentiated it from the other devices. The normalized Probability Mass Functions (PMFs) in Figure 29a show the distribution of test statistics using matched filtering to compare how much Auth1 "looks like" Auth1 and how much the other seven authorized DUTs "look like" Auth1. The PMFs for

Auth1 verification are more distinguishable than the PMFs for Auth7 verification shown in Figure 29b and only have a small amount of overlap at the EER threshold.

The *authorized device verification* of Auth7 using matched filtering achieved $EER = 59.32\%$ which is worse than "random guess" performance. As shown in Section 4.1, Auth7 had the second lowest matched filtering *device classification* performance of $\%C = 33.00\%$ which means it was difficult to distinguish Auth7 from other devices. Figure 29b shows that the distribution of test statistics comparing how much Auth7 "looks like" Auth7 exists completely within the range of test statistics comparing how much the other seven authorized DUTs "look like" Auth7. More test statistics in the Auth7 vs. Auth7 PMF exist on the left side of the EER threshold and are therefore incorrectly rejected than exist on the right side of the threshold.

Figure 30 shows an overlay of all *rogue device rejection* ROC curves for RogueA and RogueB using RF-DNA with MDA/ML (red), RF-DNA with GRLVQI (green), and matched filtering (blue) for comparison. The EER when presenting RogueA for verification as an "authorized device" is given in Table 20 along with the average EER_{Ave} achieved across all $N_{DevAuth} = 8$ authorized device classes. Using RF-DNA with MDA/ML, RogueA was successfully rejected when compared to three of the authorized devices and $EER_{Ave} = 14.17\%$. Using RF-DNA with GRLVQI, RogueA was not successfully rejected when compared to any of the authorized devices and $EER_{Ave} = 31.53\%$. Using matched filtering, RogueA was successfully rejected when compared to six of the authorized devices and $EER_{Ave} = 4.81\%$.

The EER when presenting RogueB for verification as an "authorized device" is given in Table 21 along with the average EER_{Ave} achieved across all $N_{DevAuth} = 8$ authorized device classes. Using RF-DNA with MDA/ML, RogueB was successfully

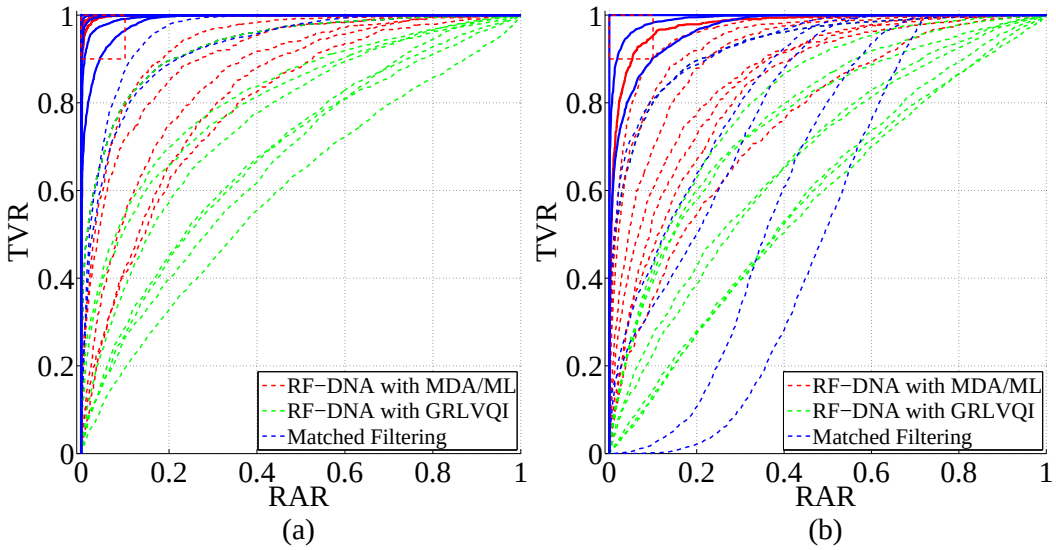

Figure 30. Comparison of *rogue device rejection* performance for a) RogueA and b) RogueB using RF-DNA with MDA/ML, RF-DNA with GRLVQI, and matched filtering techniques. Results obtained from N_{Tst}=4000 testing emissions per device at SNR_A=30 dB. The arbitrary benchmark of $EER \leq 10\%$, outlined in the top-left corner, was met by three devices using RF-DNA with MDA/ML and six device using matched filtering for RogueA, and it was met by two devices using RF-DNA with MDA/ML and three devices using matched filtering for RogueB.

Table 20. Comparison of EER and average EER_{Ave} for *rogue device rejection* of RogueA using RF-DNA with MDA/ML, RF-DNA with GRLVQI, and matched filtering techniques. Results obtained from N_{Tst}=4000 emissions per device at SNR_A=30 dB. Cells corresponding to "successful rejection" ($EER \leq 10\%$) have been highlighted in green. Matched filtering achieved the best performance with an average EER of EER_{Ave}=4.81%.

	RF-DNA w/ MDA/ML	RF-DNA w/ GRLVQI	Matched Filtering
RogueA:Auth1	1.42%	14.60%	0.03%
RogueA:Auth2	28.14%	37.03%	0.05%
RogueA:Auth3	17.72%	30.25%	0.65%
RogueA:Auth4	2.16%	27.18%	3.30%
RogueA:Auth5	1.20%	25.45%	6.70%
RogueA:Auth6	13.98%	36.25%	1.55%
RogueA:Auth7	26.00%	39.13%	15.27%
RogueA:Auth8	22.72%	42.33%	10.95%
Average	14.17%	31.53%	4.81%

Table 21. Comparison of EER and average EER_{Ave} for *rogue device rejection* of RogueB using RF-DNA with MDA/ML, RF-DNA with GRLVQI, and matched filtering techniques. Results obtained from N_{Tst}=4000 emissions per device at SNR_A=30 dB. Cells corresponding to "successful rejection" ($EER \leq 10\%$) have been highlighted in green. RF-DNA with MDA/ML achieved the best performance with an average EER of EER_{Ave}=18.40%.

	RF-DNA w/ MDA/ML	RF-DNA w/ GRLVQI	Matched Filtering
RogueB:Auth1	24.66%	37.25%	0.00%
RogueB:Auth2	7.10%	29.65%	9.85%
RogueB:Auth3	13.22%	29.48%	30.00%
RogueB:Auth4	23.20%	44.73%	39.68%
RogueB:Auth5	16.84%	37.48%	49.48%
RogueB:Auth6	31.24%	43.33%	25.93%
RogueB:Auth7	10.02%	27.00%	14.88%
RogueB:Auth8	20.94%	43.45%	4.60%
Average	18.40%	36.54%	21.80%

rejected when compared to one of the authorized devices and $EER_{Ave} = 18.40\%$. Using RF-DNA with GRLVQI, RogueB was not successfully rejected when compared to any of the authorized devices and $EER_{Ave} = 36.54\%$. Using matched filtering, RogueB was successfully rejected when compared to three of the authorized devices and $EER_{Ave} = 21.80\%$.

RF-DNA with MDA/ML achieved the best performance (lowest average EER) of the three evaluated techniques for *authorized device verification* and *rogue device rejection* when presented with RogueB for verification. Matched filtering achieved the best performance for *rogue device rejection* when presented with RogueA for verification. RogueA achieved a lower average EER than RogueB for all three evaluated techniques. The MSP-EXP430F5529 boards for RogueA and RogueB came from separate manufacturing lots which may have caused differences between the devices which made RogueA more distinguishable from the $N_{DevAuth} = 8$ "authorized" DUTs than RogueB. It is also worth noting that RogueA was included in the arbitrarily chosen set of devices (Auth1, Auth2, RogueA) used to determine the

Table 22. Comparison to computational times to process N_B=5000 emissions for each of $N_{DevAuth}$=8 devices, generate classification models, and evaluate *device classification* performance using RF-DNA with MDA/ML, RF-DNA with GRLVQI, and matched filtering. Results shown are median values obtained from N_{Trial}=5 independent trials for each scenario. All times are formatted as *H:MM:SS.s*.

	RF-DNA w/ MDA/ML	**RF-DNA w/ GRLVQI**	**Matched Filtering**
t_{Proc} ($\%t_{Proc}$)	0:52:13.6 (98.4%)	0:52:13.6 (23.3%)	0:15:34.9 (85.3%)
t_{Model} ($\%t_{Model}$)	0:00:38.8 (1.2%)	2:51:26.9 (76.6%)	0:02:20.0 (12.8%)
t_{Eval} ($\%t_{Eval}$)	0:00:11.5 (0.4%)	0:00:04.7 (0.04%)	0:00:21.4 (1.9%)
t_{Total}	**0:53:04.0**	**3:43:43.2**	**0:18:15.8**

probe location to collection URE for *device discrimination*. However, as mentioned in Section 4.1, whether or not RogueA was included in that set of devices did not change the resulting collection location chosen by the process.

4.4 Computational Time

The *computational times* required to complete the post-collection processing (t_{Proc}), model development (t_{Model}), and performance evaluation (t_{Eval}) are presented in Table 22 and Table 23 for *device classification* and *operation classification*, respectively. The results presented in this chapter were produced as described in Section 3.7 and represent the median observed time out of $N_{Trial} = 5$ independent trials for each scenario. The complete record of observed times from all trials is included in Appendix B, but has been omitted here for brevity.

The total time (t_{Total}) presented in Table 22 and Table 23 was calculated as the sum of the three processing regions for each classification method using $t_{Total} = t_{Proc} + t_{Model} + t_{Eval}$. The percentage of t_{Total} that was required for each processing region is also included in parentheses. For example, the percent of time required for post-collection processing ($\%t_{Proc}$) was calculated as $\%t_{Proc} = 100 \times t_{Proc}/t_{Total}$.

Table 23. Comparison to computational times to process N_B=5000 emissions for each of N_{Op}=12 operations, generate classification models, and evaluate *operation classification* performance using RF-DNA with MDA/ML, RF-DNA with GRLVQI, and matched filtering. Results shown are median values obtained from N_{Trial}=5 independent trials for each scenario. All times are formatted as *H:MM:SS.s*.

	RF-DNA w/ MDA/ML	**RF-DNA w/ GRLVQI**	**Matched Filtering**
t_{Proc} ($\%t_{Proc}$)	0:05:32.8 (96.1%)	0:05:32.8 (33.6%)	0:00:03.1 (12.4%)
t_{Model} ($\%t_{Model}$)	0:00:08.0 (2.3%)	0:10:56.9 (66.4%)	0:00:18.3 (73.6%)
t_{Eval} ($\%t_{Eval}$)	0:00:05.5 (1.6%)	0:00:00.2 (0.02%)	0:00:03.5 (14.0%)
t_{Total}	**0:05:46.4**	**0:16:56.1**	**0:00:24.9**

The *computational times* for *device classification* are compared in Table 22 for each of the three classification methods. The majority of the time required for RF-DNA with MDA/ML classification was spent processing the collected emissions to generate RF-DNA fingerprints. The majority of the time required for RF-DNA with GRLVQI classification was spent during the model development phase to generate prototype vectors. The majority of the time required for matched filtering classification was spent processing the collected TD signals to truncate them at rising and falling clock edges. RF-DNA with GRLVQI classification required the greatest total time, taking roughly 4 times longer than RF-DNA with MDA/ML and 12 times longer than matched filtering.

The *computational times* for *operation classification* are compared in Table 23 for each of the three classification methods. The majority of the time required for RF-DNA with MDA/ML classification was spent processing the collected emissions to generate RF-DNA fingerprints. The majority of the time required for RF-DNA with GRLVQI classification was spent during the model development phase to generate prototype vectors. The majority of the time required for matched filtering classification was spent during the model development phase to generate reference

Table 24. Comparison of alternative computational times ($t_{AltTotal}$=t_{Proc}+t_{Eval}) using RF-DNA with MDA/ML, RF-DNA with GRLVQI, and matched filtering techniques for device and operation classification. All times are formatted as *H:MM:SS.s*.

	RF-DNA w/ MDA/ML	RF-DNA w/ GRLVQI	Matched Filtering
Device Classification	0:52:25.2	0:52:18.4	0:15:56.2
Operation Classification	0:05:38.4	0:05:33.0	0:00:06.6

waveform templates. RF-DNA with GRLVQI classification required the greatest total time, taking nearly 3 times longer than RF-DNA with MDA/ML and 42 times longer than matched filtering.

When considering the "alternative computational times" ($t_{AltTotal} = t_{Proc} + t_{Eval}$) presented in Table 24, matched filtering outperformed both of the RF-DNA based classification methods with the lowest $t_{AltTotal}$ time for *device* and *operation* classification. For *device classification* the RF-DNA techniques required roughly 3 times more $t_{AltTotal}$ to complete than matched filtering and for *operation classification* the RF-DNA techniques required roughly 50 times more $t_{AltTotal}$ than matched filtering.

V. Conclusion

This chapter provides a summary of research activities and results comparing: 1) Radio-Frequency (RF) Distinct Native Attribute (RF-DNA) features paired with Multiple Discriminant Analysis/Maximum Likelihood (MDA/ML) classification, 2) RF-DNA features paired with Generalized Relevance Learning Vector Quantized-Improved (GRLVQI) classification, and 3) Time Domain (TD) signals with matched filtering classification for the purposes of *device discrimination* and *operation identification* based on Unintentional RF Emissions (URE) from Micro-Controller Units (MCUs). Section 5.1 provides a summary of the research motivation and goals for this effort. Section 5.1.1 and Section 5.1.2 summarize the *classification* and *verification* performance, respectively, as presented in Chapter IV using each of the three evaluated techniques. This chapter concludes with future research recommendations in Section 5.2 for continued investigation and improvement of the methods demonstrated in this work.

5.1 Research Summary

Supervisory Control And Data Acquisition (SCADA) systems are an integral part of critical military and civilian infrastructures around the world and are therefore a prime target for cyber attack. In the past decade, hackers have compromised SCADA systems to inflict physical damage and shutdown targets including nuclear centrifuges [131], steel mills [132], and city power grids [75]. The growing threat of cyber attack against critical infrastructure in the United States motivated President Obama to issue Executive Order 13636 with the goal of "improving critical infrastructure cybersecurity" [84].

One serious threat to the security and reliability of systems performing critical

functions is the presence of counterfeit hardware in the supply chain [106]. Driven by the desire for cheaper electronic components, most Integrated Circuit (IC) manufacturers have outsourced the fabrication of semiconductor devices to Asian countries with a low cost of labor which raises concerns regarding the authenticity of devices [1]. Reports of counterfeit hardware have increased four fold from 2009 to 2011 [54] and counterfeit hardware has been discovered in systems ranging from common network routers to high-altitude missile computers [40]. This reliance on foreign semiconductor suppliers creates the opportunity for malicious hardware Trojans to be implanted in devices [85]. Techniques to identify and authenticate semiconductor devices are necessary to prevent counterfeit components and hardware Trojans from entering the supply chain for critical systems.

The goal of the *device discrimination* aspect of this research was to compare the effectiveness of three techniques to *classify* and *verify* the identity of individual MCU devices that could be applied to detect counterfeit or Trojan hardware. RF-DNA fingerprints paired with the MDA/ML and GRLVQI classifiers have been used in prior Air Force Institute of Technology (AFIT) research to classify devices from URE [14–16, 25, 26, 100–105, 127–129]. This research effort also evaluated TD signals with matched filtering for *device discrimination* as a computationally simpler alternative to the established RF-DNA techniques.

In additional to guarding against counterfeit and Trojan hardware, it is important to defend critical system against software vulnerabilities and exploitation. Modern worms and viruses, such as the infamous Stuxnet virus, may take advantage of previously unknown "zero-day" vulnerabilities to infect SCADA systems and then hide their existence from the operator by reporting a false status while performing nefarious deeds [112]. Modern methods for detecting unauthorized activity on information systems focus on analyzing data within the upper Applica-

tion (level 1) and Network (level 5) layers of the 7-layer Open Systems Interconnect (OSI) model. Virus detection methods often rely on definitions derived from previously seen malware and would therefore not be able to detect zero-day exploits. Additionally, these detection methods require processing resources beyond normal operation which is problematic for many fielded systems due to the limited on-board computing resources of the embedded processors used in Programmable Logic Controller (PLC) devices [102].

The goal of the *operation identification* aspect of this research was to compare the effectiveness of three techniques to estimate the individual operations executed by a MCU from URE. The series of estimated operation could be compared with the authorized program to verify the device execution. The MDA/ML and GR-LVQI classifiers with RF-DNA fingerprints that have been established as successful methods for *device discrimination* where also evaluated for *operation identification*. However, due to their high computational cost, the RF-DNA methods are not well suited for implementation in embedded systems with limited processing resources. Therefore, matched filtering with TD signals is also considered as a potential alternative to the RF-DNA techniques due to its relatively low computational cost.

The MSP430F5529 16-bit MCU manufactured by Texas Instruments [117] was selected as the Device Under Test (DUT) for this research because it is widely used and representative of modern MCU architecture and semiconductor manufacturing processes [119]. URE were collected from the DUT using a high-sensitivity RF near-field probe with a Teledyne LeCroy WavePro 760Zi-A oscilloscope sampling at a rate of $f_S = 1$ Gsps. The collected emissions were then processed using MATLAB® to develop classification models and evaluate their performance for each of the three techniques. Section 5.1.1 presents a summary of the performance for *device* and *operation* classification and the required computational cost using each of the

evaluated techniques. Section 5.1.2 summarizes the *authorized device verification* and *rogue device rejection* performance achieved using each technique.

5.1.1 Classification Performance.

Table 25 provides a summary of the *classification* results presented in Section 4.1, Section 4.2, and Section 4.4. The *classification performance* values represent the average percent correct classification ($\%C_{Ave}$) achieved at the highest evaluated $SNR_A = 30$ dB. The *computational time* values represent the "alternative computational time" ($t_{AltTotal} = t_{Proc} + t_{Eval}$) which excludes the model development time, representing an "as deployed" solution. While an actual deployed system might acquire, process, and evaluate emissions on an individual basis, the times shown here represent the computational time to process a "batch" of emissions as described in Section 3.7 because the time required to process individual emissions could not be measured with reasonable accuracy in the MATLAB® environment. The implementation of the classification techniques in a deployed solution would likely differ from the MATLAB® implementation used in this research; however, these times provide an initial way to compare the *relative* computational cost of the three techniques.

Table 25. Summary of the *classification performance* ($\%C_{Ave}$ at SNR_A=30 dB) and *computational time* ($t_{AltTotal}$=t_{Proc}+t_{Eval}) using RF-DNA with MDA/ML, RF-DNA with GRLVQI, and matched filtering techniques for device and operation classification. Cells are colored according to the relative performance of the three methods with green representing the best, yellow the middle, and red the worst performance.

	RF-DNA w/ MDA/ML	RF-DNA w/ GRLVQI	Matched Filtering
Device Classification Performance	75.66%	61.08%	52.45%
Operation Classification Performance	10.98%	8.69%	11.10%
Device Classification Computational Time	0:52:25.2	0:52:18.4	0:15:56.2
Operation Classification Computational Time	0:05:38.4	0:05:33.0	0:00:06.6

For *device classification*, RF-DNA with MDA/ML provided the best performance, RF-DNA with GRLVQI was second best, and matched filtering was the lowest. It is not surprising that both RF-DNA techniques out-performed matched filtering for *device classification* since they were developed for that purpose through prior research; however, the RF-DNA techniques did not perform as well here as they have in other AFIT research efforts. For example, prior research using MDA/ML classification with RF-DNA fingerprints successfully classified 40 near-identical PIC MCUs with a correct identification rate of better than 90% for $SNR_A \geq 15$ dB [14]. The benchmark of percent correct classification $\%C \geq 90\%$ has been used in previous AFIT RF-DNA to represent "successful classification" [14, 90, 127]. Although none of the evaluated techniques achieved an average percent correct classification of $\%C_{Ave} \geq 90\%$, Auth1 did surpass the benchmark with $\%C = 97.40\%$ at $SNR_A = 30$ dB using RF-DNA with MDA/ML. As previously described in Section 4.1, the URE from Auth1 had characteristics that made it more unique and distinguishable from the other devices, possibly related to Auth1 having spent more time powered up and operating than the other evaluated DUTs.

There are several potential factors that may have contributed to the lower classification performance of the RF-DNA techniques in this research compared to prior efforts. The process used by Texas Instruments to manufacture MSP430 MCUs might not produce as much variation between devices as the process used by Microchip Technology to manufacture the PIC MCUs used in previous RF-DNA research [14], thus making it more difficult to discriminate between individual MSP430 devices. Additionally, in this research effort factors such as probe placement and post-collection signal processing steps were chosen to maximize the matched filtering classification performance rather than the RF-DNA performance. Future research should consider the positive and negative impacts of those process decisions.

For *operation classification*, matched filtering provided the best performance, RF-DNA with MDA/ML was second best, and RF-DNA with GRLVQI had the lowest performance – barely exceeding "random guess" accuracy. This poor performance was unsatisfactory and would not be useful for real-world applications. It was significantly lower than the *operation identification* performance achieved in previous Side Channel Analysis (SCA) for Reverse Engineering (SCARE) research efforts. For example, a prior SCARE effort analyzing the power consumption of an 8-bit ATMega163 MCU claimed to achieve 100% correct classification when identifying 39 commonly used instructions [81, 82]. Another effort analyzing URE from a decapsulated 8-bit PICF687 MCU achieved instruction recognition rates up to 96.24% [108]. There were multiple aspects of the current research effort that made the task of *operation identification* more difficult than in previous SCARE research.

1. The 16-bit MSP430 MCU has a more complex architecture and instruction set than simple 8-bit ATMega and PIC MCUs. The $N_{Op} = 12$ evaluated MSP430 operations execute in a single clock cycle whereas the 8-bit PIC MCU operations have a duration of four clock cycles, providing a longer period for observation.

2. One goal of this research was to use a *non-contact*, *non-destructive* method to collect side-channel information for analysis. Power consumption is a commonly used side-channel for SCARE; however, it requires contact measurement of the DUT power supply and was therefore not used. Additionally, since this research effort did not decapsulate the DUT as in previous URE-based SCARE [108], it was not possible to achieve the same close and precise positioning of the near-field probe over specific circuit elements.

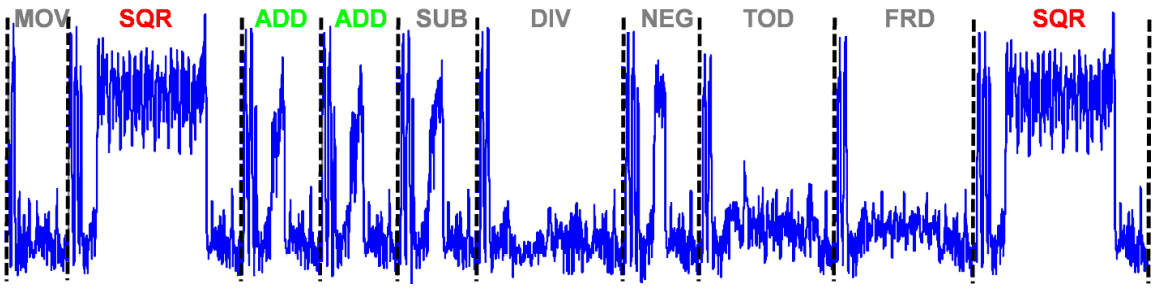

Figure 31. Sequence of URE collected from an Allen Bradley PLC executing a ladder logic program [102].

3. Executing the $N_{Op} = 12$ Format I operations produced URE which were difficult to distinguish between because they had a nearly identical signal shape. Other types of devices may produce operation-dependent URE which is more unique. For example, Figure 31 shows that the individual ladder logic instruction executed on a PLC for previous Correlation-Based Anomaly Detection (CBAD) research are visibly recognizable from the URE [102].

The combination of these additional factors increased the complexity of the current research and contributed to the significantly lower performance compared to previous efforts. Future research evaluating RF-DNA and matched filtering techniques for *operation identification* should scale back the aforementioned factors by using a simpler DUT or different side-channels that provide better differentiation between operations.

The *computational time* comparison in Table 25 shows that it takes significantly less time to process and evaluate URE with matched filtering classification than with the RF-DNA based classification technique. These results suggest that matched filtering may serve as a computationally low cost alternative to the complex RF-DNA fingerprint generation, MDA/ML, and GRLVQI algorithms if its classification performance can be improved to meet application requirements.

5.1.2 Verification Performance.

Table 26 provides a summary of the *verification* results presented in Section 4.3. The *successful verification* and *successful rejection* numbers represent the number of Receiver Operating Characteristics (ROC) curves that achieved the $EER \leq 10\%$ benchmark for each scenario. The EER_{Ave} values represent the average EER of the $N_{AuthDev} = 8$ ROC curves for each scenario.

Table 26. Summary of the number of successes ($EER \geq 10\%$) and average EER_{Ave} for *authentic device verification* and *rogue device rejection* performance using RF-DNA with MDA/ML, RF-DNA with GRLVQI, and matched filtering techniques. Cells are colored according to the relative performance of the three methods with green representing the best, yellow the middle, and red the worst performance. All results were evaluated at SNR_A=30 dB.

	RF-DNA w/ MDA/ML	**RF-DNA w/ GRLVQI**	**Matched Filtering**
Authorized Device Successful Verification	1 of 8	0 of 8	1 of 8
RogueA Device Successful Rejection	3 of 8	0 of 8	6 of 8
RogueB Device Successful Rejection	1 of 8	0 of 8	3 of 8
Authorized Device Verification EER_{Ave}	16.23%	34.56%	33.08%
RogueA Device Rejection EER_{Ave}	14.17%	31.53%	4.81%
RogueB Device Rejection EER_{Ave}	18.40%	36.54%	21.80%

Both RF-DNA with MDA/ML and matched filtering techniques were able to successfully verify one out of eight authentic devices and reject a rogue device compared to at least one out of the eight authentic devices. RF-DNA with GRLVQI was not able to successfully verify or reject any authentic or rogue devices. Considering that matched filtering had the lowest *device classification* performance of the three techniques, it was surprising that matched filtering successfully accepted/rejected more authentic/rogue devices than either of the RF-DNA based techniques.

Additionally, matched filtering was the only technique that achieved an average Equal Error Rate (EER) of $EER_{Ave} \leq 10\%$ which occurred when evaluating *rogue device rejection* with RogueA. However, it is worth noting that the ROC curves for *authentic device verification* and *rogue device rejection* using matched filtering spanned a wider range of EERs than either of the RF-DNA based techniques. For example, the EERs when evaluating RogueB with matched filtering ranged from $EER = 0.00\%$ for RogueB:Auth1 to $EER = 49.48\%$ for RogueB:Auth5 – more than twice the EER range of the other methods. Therefore, despite appearing to be the best *verification* technique based on the number of "successful" verifications and rejections, the inconsistent performance of matched filtering in terms of EER makes it a less desirable technique compared to RF-DNA with MDA/ML.

5.2 Future Research Recommendations

The *classification* and *verification* results presented here had much lower performance than previous related efforts and leaves significant room for improvement, especially for *operation identification*. Despite the poor *operation classification* performance achieved in this research effort, the fact that average percent correct classification was better than "random guess" for all three techniques suggests that there is characteristic signal content in URE that can be used to differentiate operations. The following recommendations are for related research avenues that warrant additional investigation to improve performance.

1. Improve URE Collection Process: There are several improvements that could be made to the acquisition system and collection process used in this research to further reduce to impact of noise and collection bias. Performing the collections in a temperature controlled anechoic chamber would reduce environmental noise. Additionally, powering the DUT from a DC power supply

rather than the USB cable would remove the possibility of noise from the computer affecting the DUT. Since each DUT was mounted on its own MSP-EXP430F5529 board, differences between those boards could have influenced the URE for this research. It would be better to use a single board for all collections with a Zero Insertion Force (ZIF) socket that allows just the DUT chip to be changed. Using a high quality external source to clock the DUT would reduce the impact of clock jitter. Finally, a larger set of devices and operations should be evaluated to determine the optimal location for URE collection. It was not realized until the end of this research effort that the "most unique" device, Auth1, was one of the three that had been arbitrarily used to choose the collection location.

2. Study Impact of Device Age on URE: This research effort did not maintain device usage logs that would be necessary to provide formal analysis, but the observations about Auth1 and its distinguishability suggest that future research into the effects of device usage on URE may be worthwhile.

3. Optimize Processing for RF-DNA: For this analysis, several factors such as probe placement and post-collection signal processing were chosen to optimize *matched filtering* performance. Using a different probe location or additional post-collection processing steps such as a Band Pass Filter (BPF), Digital Down Conversion (DDC), and/or Hilbert Transform (HT) may improve the *classification* and *verification* performance using RF-DNA based techniques. Shifting the location of the RF-DNA subregion so they do not match the clock boundaries may improve performance. Also, it may be beneficial to use subregions that are smaller than half clock periods to focus on "higher activity" portions of the signal, especially when evaluating short, single cycle operations.

4. Consider Alternative Operation "Classes": The $N_{Op} = 12$ operation classes considered for this research were all Format I instructions using register-to-register addressing; however, it is possible to define the classes in other ways. For example, the classes could be grouped based on the addressing modes (register-to-register, register-to-memory, etc...) or by the Hamming weights of the opcodes and/or operands. Future research should consider alternative class groupings that are more distinguishable from each other to provide better classification performance for *operation identification*.

5. Prior Probabilities for Operation Identification: This research effort assumed an equal prior probability, P_i, for all *device* and *operation* classes to maintain consistency with prior AFIT RF-DNA work. While that assumption made sense for this research because all classes were represented equally, in real-world code certain operations tend to appear more frequently than others. Using code profiling and compiler statistics to adjust the prior probabilities for operation classes may improve the *operation identification* performance when analyzing real-world code segments.

6. Consider Preceding/Succeeding Operations: The operations immediately preceding and succeeding the operation of interest can potentially influence the URE during its execution. It may be beneficial to create sub-classes of operations which considers the preceding and succeeding instructions, similar to the grouping process used for Constellation-Based Distinct Native Attributes (CB-DNA) which was based on preceding and succeeding symbol estimations [12].

7. Decision Tree for Multi-Cycle Operation Identification: The $N_{Op} = 12$ Format I operations analyzed here used the register-to-register addressing mode and

therefore had an effective duration of one clock cycle. However, the MSP430 has operations and addressing modes which range in execution time from one to six clock cycles. To determine the duration of an *unknown* operation from observed URE it may be necessary to use a decision tree that makes a logical conclusion about the duration of the operation based on certain signal characteristics.

8. Classify Operation Sequences: This research effort attempted to identify individual MCU operations, effectively reverse-engineering the executed program. Rather than attempting to classify individual instructions it may be beneficial to create classes that represent common sequences of operations which can exclude any impossible (or unlikely) operation sequences. This approach could take advantage of hidden Markov chains to adjust model probabilities based on observed sequences as done in previous SCARE research efforts [107].

9. Consider the Effect of Data Values: The program sequences used here for *operation identification* removed the impact of data values on URE by initializing all of the registers and operands to zero. As demonstrated by prior research [39] and early experiments [100], the Hamming weight and Hamming distance of consecutive instructions and data values can have a characteristic effect on URE. Future research should continue to investigate the effects of data on URE to determine if it is possible to identify both an executed operation and the processed values.

10. Clock Edge Alignment from URE: One way that this research effort differentiated itself from prior SCARE research is that it used a non-contact RF probe to collect URE. However, due to the observed clock jitter, a contact probe was also used in this effort to record the Master CLocK (MCLK) to facilitate

signal alignment. Future research should evaluate the effectiveness of using signal processing techniques to align clock edges based solely on the observed URE to avoid the need for the additional contact probe.

11. Hardware Implementation of Matched Filtering: One of the main reasons for evaluating the relatively simple correlation-based matched filtering process as an alternative to algorithms like MDA/ML and GRLVQI was that correlation is well suited to implementation in realtime, inexpensive hardware [102]. The MATLAB® based matched filtering implementation used in this research out performed both RF-DNA techniques in regards to *computational time*. A Field-Programmable Gate Array (FPGA) based implementation of the matched filtering classification process would serve as a proof-of-concept to demonstrate its speed and performance using custom hardware. Similarly, a MCU based matched filtering implementation would demonstrate its applicability to use in low-cost, embedded processors.

Appendix A. MSP430 Format I Instructions

Table 27. MSP430 Source and Destination Addressing Modes [118].

Addressing Mode	Syntax	Description
Register	Rn	Register contents are operand.
Indexed	X(Rn)	(Rn + X) points to the operand. X is stored in the next word, or stored in combination of the precending extension word and the next word.
Symbolic	ADDR	(PC + X) points to the operand. X is stored in the next word, or stored in combination of the preceding extension word and the next word. Indexed mode X(PC) is used.
Absolute	&ADDR	The word following the instruction contains the absolute address. X is stored in the next word, or stored in combination of the preceding extension word and the next word. Indexed mode X(SR) is used.
Indirect Register	@Rn	Rn is used as a pointer to the operand.
Indirect Autoincrement	@Rn+	Rn is used as a pointer to the operand. Rn is incremented afterwards by 1 for .B instructions.
Immediate	#N	N is stored in the next word, or stored in combination of the preceding extension word and the next word. Indirect autoincrement mode @PC+ is used.

Appendix B. Additional Results

This appendix contains additional result data that was not presented in Chapter IV including the *computational times* observed for each of the independent trials, the *classification performance* plots and confusion matrices for the individual device and operation classes, and separate *verification performance* receiver operating characteristics (ROC) curves for each of the three evaluated methods.

Table 28. Computational times to generate N_B=5000 RF-DNA fingerprints for each of $N_{DevAuth}$=8 devices, generate the MDA/ML projection matrix using standard k=5 kF-CV with N_{Tng}=1000 training fingerprints per device, and evaluate the MDA/ML classification performance using N_{Tst}=4000 testing fingerprints per device. All times are formatted as *H:MM:SS.s*.

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Median
t_{Proc}	0:55:04.3	0:54:29.9	0:52:11.6	0:50:35.2	0:52:13.6	**0:52:13.6**
t_{Model}	0:00:41.5	0:00:38.8	0:00:38.6	0:00:38.8	0:00:38.9	**0:00:38.8**
t_{Eval}	0:00:11.7	0:00:11.5	0:00:11.5	0:00:11.7	0:00:11.4	**0:00:11.5**
t_{Total}	0:55:57.4	0:55:20.2	0:53:01.7	0:51:25.7	0:53:04.0	**0:53:04.0**

Table 29. Computational times to generate N_B=5000 RF-DNA fingerprints for each of $N_{DevAuth}$=8 devices, generate GRLVQI prototype vectors using standard k=5 kF-CV with N_{Tng}=1000 training fingerprints per device, and evaluate the GRLVQI classification performance using N_{Tst}=4000 testing fingerprints per device. All times are formatted as *H:MM:SS.s*.

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Median
t_{Proc}	0:55:04.3	0:54:29.9	0:52:11.6	0:50:35.2	0:52:13.6	**0:52:13.6**
t_{Model}	2:51:33.4	2:52:20.1	2:51:26.6	2:51:07.4	2:51:21.9	**2:51:26.6**
t_{Eval}	0:00:04.7	0:00:04.3	0:00:05.0	0:00:05.2	0:00:04.7	**0:00:04.7**
t_{Total}	3:46:42.3	3:46:54.3	3:43:43.2	3:41:47.8	3:43:40.3	**3:43:43.2**

Table 30. Computational times to process and truncate N_B=5000 emissions for each of $N_{DevAuth}$=8 devices, generate matched filter reference templates using standard k=5 kF-CV with N_{Tng}=1000 training waveforms per device, and evaluate the matched filter classification performance using N_{Tst}=4000 testing waveforms per device. All times are formatted as *H:MM:SS.s*.

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Median
t_{Proc}	0:15:32.5	0:15:35.1	0:15:34.5	0:15:37.9	0:15:34.9	**0:15:34.9**
t_{Model}	0:02:20.2	0:02:20.4	0:02:19.9	0:02:20.0	0:02:19.3	**0:02:20.0**
t_{Eval}	0:00:21.0	0:00:21.4	0:00:21.4	0:00:21.5	0:00:21.1	**0:00:21.4**
t_{Total}	0:18:13.7	0:18:16.8	0:18:15.8	0:18:19.4	0:18:15.3	**0:18:15.8**

Table 31. Computational times to generate N_B=5000 RF-DNA fingerprints for each of N_{Op}=12 operations, generate the MDA/ML projection matrix using standard k=5 kF-CV with N_{Tng}=1000 training fingerprints per operation, and evaluate the MDA/ML classification performance using N_{Tst}=4000 testing fingerprints per operation. All times are formatted as *H:MM:SS.s*.

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Median
t_{Proc}	0:05:59.0	0:06:02.9	0:05:32.8	0:05:22.5	0:05:31.1	**0:05:32.8**
t_{Model}	0:00:08.0	0:00:07.9	0:00:08.0	0:00:08.2	0:00:08.2	**0:00:08.0**
t_{Eval}	0:00:05.6	0:00:05.6	0:00:05.5	0:00:05.5	0:00:05.5	**0:00:05.5**
t_{Total}	0:06:12.6	0:06:16.4	0:05:46.4	0:05:36.1	0:05:44.8	**0:05:46.4**

Table 32. Computational times to generate N_B=5000 RF-DNA fingerprints for each of N_{Op}=12 operations, generate GRLVQI prototype vectors using standard k=5 kF-CV with N_{Tng}=1000 training fingerprints per operation, and evaluate the GRLVQI classification performance using N_{Tst}=4000 testing fingerprints per operation. All times are formatted as *H:MM:SS.s*.

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Median
t_{Proc}	0:05:59.0	0:06:02.9	0:05:32.8	0:05:22.5	0:05:31.1	**0:05:32.8**
t_{Model}	0:10:56.9	0:11:08.7	0:10:46.1	0:10:24.5	0:11:35.4	**0:10:56.9**
t_{Eval}	0:00:00.2	0:00:00.2	0:00:00.2	0:00:00.2	0:00:00.2	**0:00:00.2**
t_{Total}	0:16:56.1	0:17:11.8	0:16:19.1	0:15:47.2	0:17:06.7	**0:16:56.1**

Table 33. Computational times to process and truncate N_B=5000 emissions for each of N_{Op}=12 operations, generate matched filter reference templates using standard k=5 kF-CV with N_{Tng}=1000 training waveforms per operation, and evaluate the matched filter classification performance using N_{Tst}=4000 testing waveforms per operation. All times are formatted as *H:MM:SS.s*.

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Median
t_{Proc}	0:00:03.0	0:00:03.1	0:00:03.1	0:00:03.2	0:00:03.1	**0:00:03.1**
t_{Model}	0:00:18.3	0:00:18.5	0:00:18.3	0:00:18.3	0:00:18.3	**0:00:18.3**
t_{Eval}	0:00:03.5	0:00:03.5	0:00:03.5	0:00:03.5	0:00:03.5	**0:00:03.5**
t_{Total}	0:00:24.8	0:00:25.0	0:00:24.9	0:00:25.0	0:00:24.9	**0:00:24.9**

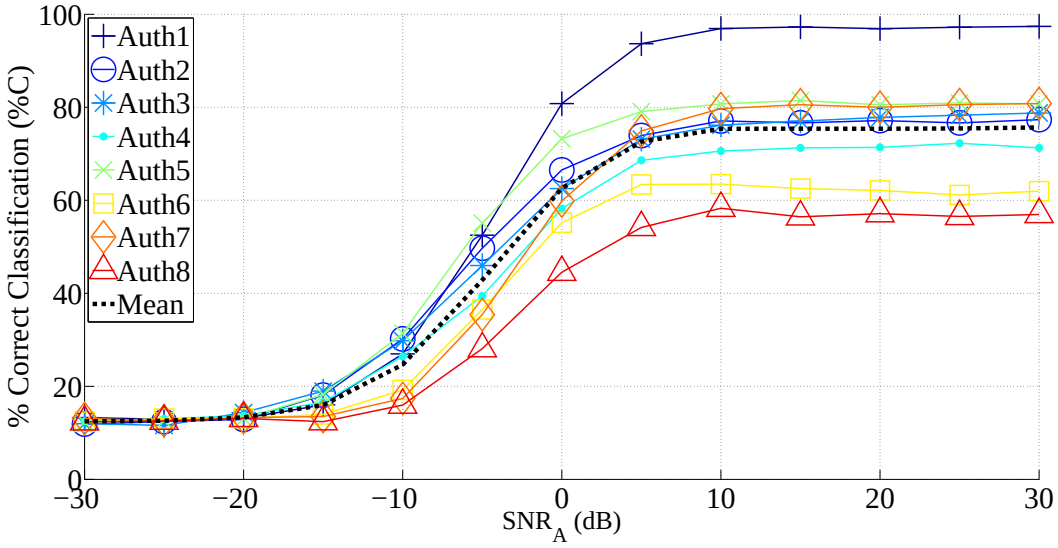

Figure 32. ***Device classification*** **performance for N_C=$N_{DevAuth}$=8 devices using *RF-DNA with MDA/ML* to evaluate N_{Tst}=4000 testing fingerprints per device class. Confidence intervals of CI=95% are approximately the same size or smaller than the marker shapes and have been omitted for visual clarity.**

Table 34. Confusion matrix for *device classification* of N_C=$N_{DevAuth}$=8 devices using *RF-DNA with MDA/ML* to evaluate N_{Tst}=4000 testing emissions per device at SNR_A=30 dB.

Actual Class \ *Declared Class*	Auth1	Auth2	Auth3	Auth4	Auth5	Auth6	Auth7	Auth8
Auth1	**3688**	5	113	55	77	41	7	14
Auth2	34	**2769**	737	8	2	136	145	169
Auth3	135	707	**2790**	5	3	117	155	88
Auth4	74	5	10	**2609**	654	403	36	209
Auth5	90	2	4	707	**2885**	222	15	75
Auth6	54	159	108	404	255	**2225**	217	578
Auth7	4	201	112	53	19	243	**2837**	531
Auth8	17	212	95	301	133	591	728	**1923**

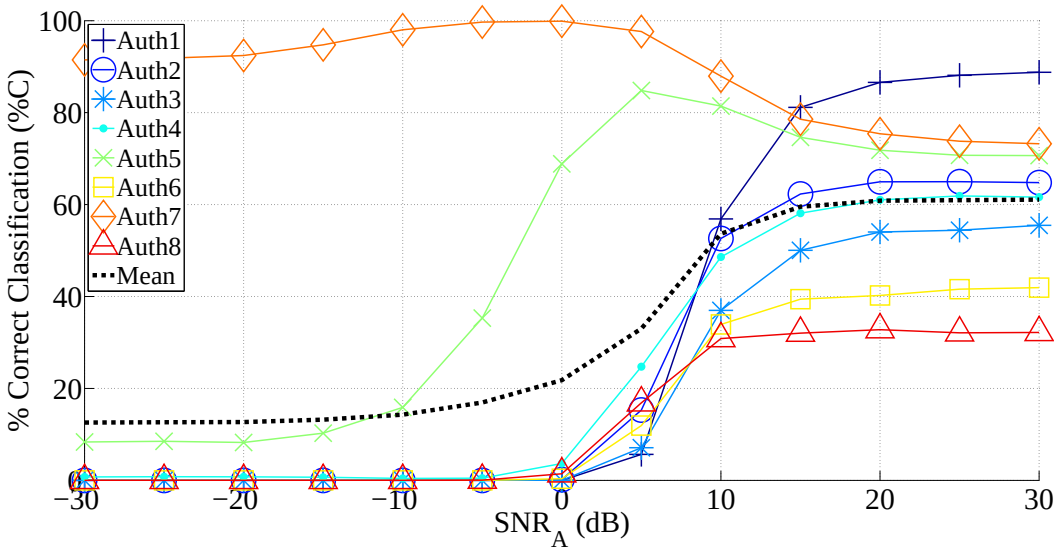

Figure 33. ***Device classification*** **performance for N_C=$N_{DevAuth}$=8 devices using *RF-DNA with GRLVQI* to evaluate N_{Tst}=4000 testing fingerprints per device class. Confidence intervals of CI=95% are approximately the same size or smaller than the marker shapes and have been omitted for visual clarity.**

Table 35. Confusion matrix for *device classification* of N_C=$N_{DevAuth}$=8 devices using *RF-DNA with GRLVQI* to evaluate N_{Tst}=4000 testing emissions per device at SNR_A=30 dB.

Actual Class \ *Declared Class*	Auth1	Auth2	Auth3	Auth4	Auth5	Auth6	Auth7	Auth8
Auth1	**3551**	29	103	79	126	66	10	36
Auth2	71	**2591**	801	11	3	117	243	163
Auth3	289	784	**2219**	8	11	140	308	241
Auth4	194	26	12	**2465**	747	333	63	160
Auth5	160	2	5	689	**2826**	162	76	80
Auth6	185	230	183	447	328	**1677**	311	639
Auth7	6	200	89	35	28	186	**2929**	527
Auth8	97	318	133	354	211	592	1008	**1287**

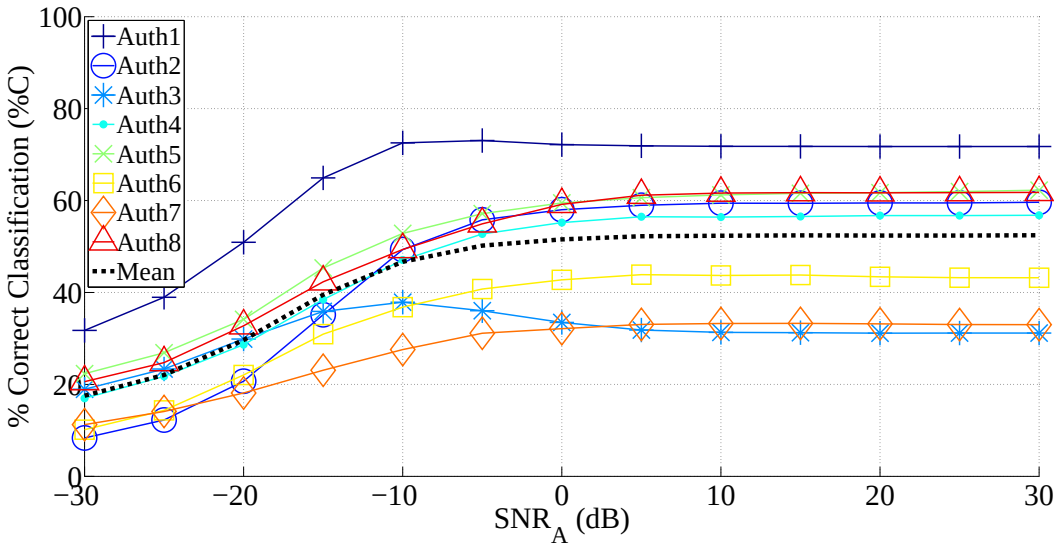

Figure 34. ***Device classification*** **performance for N_C=$N_{DevAuth}$=8 devices using** ***matched filtering*** **to evaluate N_{Tst}=4000 testing emissions per device class. Confidence intervals of CI=95% are approximately the same size or smaller than the marker shapes and have been omitted for visual clarity.**

Table 36. Confusion matrix for ***device classification*** **of N_C=$N_{DevAuth}$=8 devices using** ***matched filtering*** **to evaluate N_{Tst}=4000 testing emissions per device at SNR_A=30 dB.**

Actual Class \ *Declared Class*	**Auth1**	**Auth2**	**Auth3**	**Auth4**	**Auth5**	**Auth6**	**Auth7**	**Auth8**
Auth1	**2870**	0	0	0	0	1127	3	0
Auth2	177	**2385**	35	286	834	137	43	103
Auth3	0	1503	**1247**	0	1250	0	0	0
Auth4	0	878	8	**2272**	842	0	0	0
Auth5	0	184	1041	285	**2490**	0	0	0
Auth6	12	59	0	0	488	**1728**	1585	128
Auth7	546	21	5	0	232	898	**1320**	978
Auth8	66	0	0	0	0	1374	89	**2471**

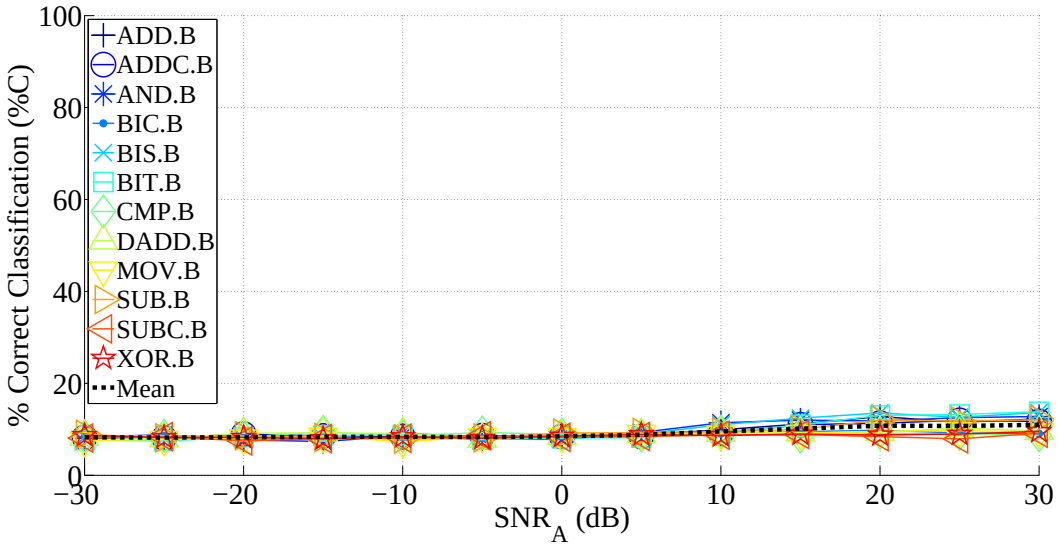

Figure 35. ***Operation classification*** **performance for N_C=N_{Op}=12 operations using *RF-DNA with MDA/ML* to evaluate N_{Tst}=4000 testing fingerprints per operation class. Confidence intervals of CI=95% are approximately the same size or smaller than the marker shapes and have been omitted for visual clarity.**

Table 37. Confusion matrix for *operation classification* of N_C=N_{Op}=12 operations using *RF-DNA with MDA/ML* to evaluate N_{Tst}=4000 testing emissions per operation at SNR_A=30 dB.

Actual Class \ *Declared Class*	ADD.B	ADDC.B	AND.B	BIC.B	BIS.B	BIT.B	CMP.B	DADD.B	MOV.B	SUB.B	SUBC.B	XOR.B
ADD.B	**263**	259	145	481	273	396	528	354	389	207	389	316
ADDC.B	247	**239**	149	525	317	369	490	381	388	180	397	318
AND.B	199	195	**178**	513	427	297	494	371	362	171	444	349
BIC.B	100	138	129	**643**	226	186	625	415	490	114	521	413
BIS.B	216	207	177	494	**490**	306	485	369	352	171	402	331
BIT.B	225	240	137	531	276	**407**	499	346	423	185	407	324
CMP.B	96	120	135	613	230	209	**638**	459	442	109	512	437
DADD.B	113	123	133	625	239	189	667	**427**	443	103	543	395
MOV.B	108	126	141	634	221	185	674	449	**456**	123	502	381
SUB.B	237	229	122	448	323	388	511	370	387	**215**	420	350
SUBC.B	124	145	130	645	226	168	607	422	519	107	**525**	382
XOR.B	113	138	141	625	217	161	650	438	471	112	531	**403**

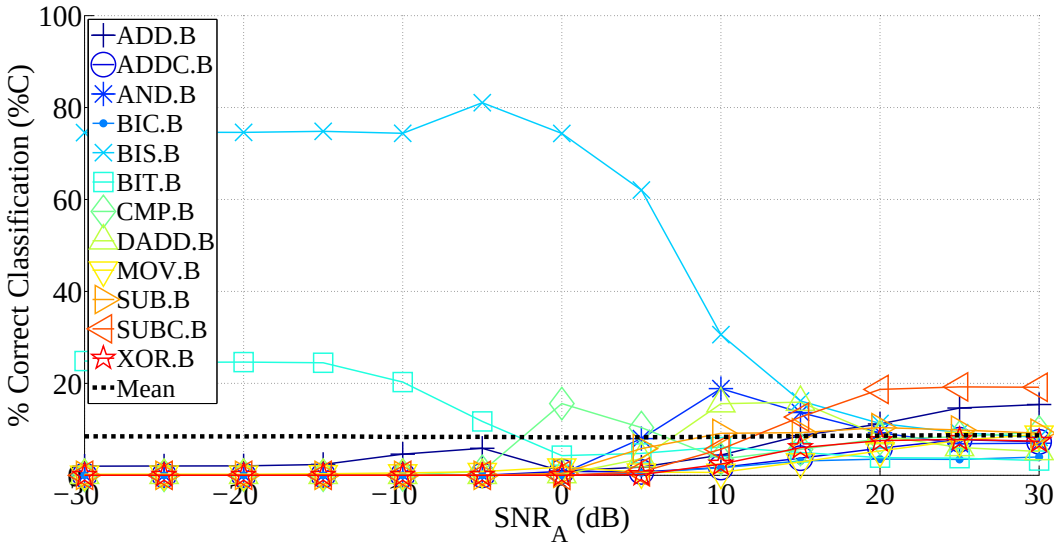

Figure 36. ***Operation classification*** **performance for N_C=N_{Op}=12 operations using *RF-DNA with GRLVQI* to evaluate N_{Tst}=4000 testing fingerprints per operation class. Confidence intervals of CI=95% are approximately the same size or smaller than the marker shapes and have been omitted for visual clarity.**

Table 38. Confusion matrix for *operation classification* of N_C=N_{Op}=12 operations using *RF-DNA with GRLVQI* to evaluate N_{Tst}=4000 testing emissions per operation at SNR_A=30 dB.

Actual Class \ *Declared Class*	ADD.B	ADDC.B	AND.B	BIC.B	BIS.B	BIT.B	CMP.B	DADD.B	MOV.B	SUB.B	SUBC.B	XOR.B
ADD.B	**616**	335	303	110	269	145	320	203	262	368	764	305
ADDC.B	650	**291**	272	123	286	141	279	219	273	401	749	316
AND.B	674	300	**279**	105	307	136	317	227	224	332	734	255
BIC.B	604	286	170	**159**	253	126	384	197	329	407	808	277
BIS.B	680	273	288	108	**329**	100	321	246	320	388	696	251
BIT.B	601	324	301	123	291	**128**	352	217	272	363	754	274
CMP.B	568	290	179	146	250	137	**377**	230	323	410	803	287
DADD.B	594	283	182	145	271	99	377	**207**	314	447	803	278
MOV.B	585	293	176	120	237	132	392	218	**357**	362	815	313
SUB.B	651	345	269	131	268	109	397	242	284	**366**	767	271
SUBC.B	598	291	186	141	279	129	370	218	347	365	**766**	316
XOR.B	577	259	193	145	250	119	401	221	328	380	831	**296**

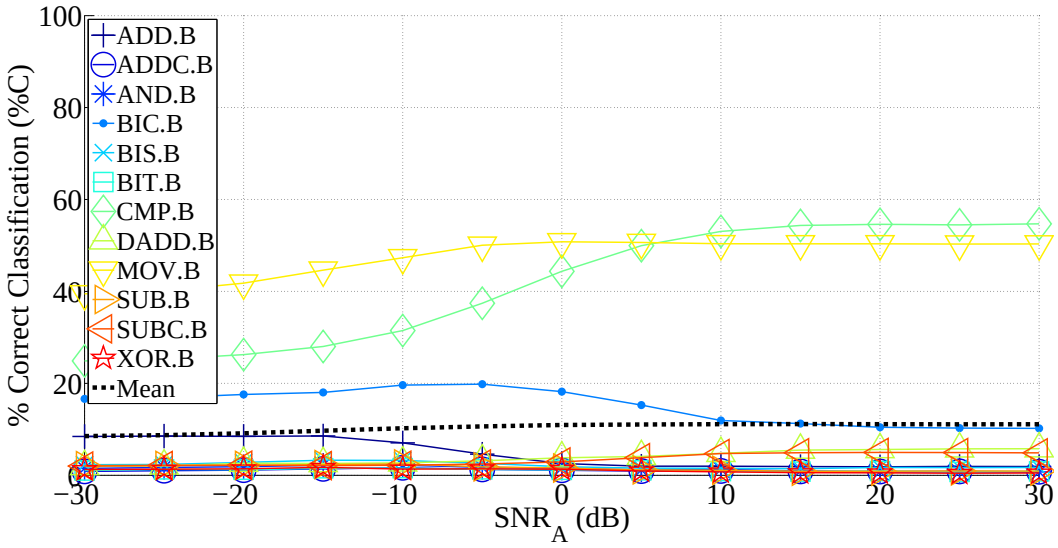

Figure 37. ***Operation classification*** **performance for N_C=N_{Op}=12 operations using *matched filtering* to evaluate N_{Tst}=4000 testing emissions per operation class. Confidence intervals of CI=95% are approximately the same size or smaller than the marker shapes and have been omitted for visual clarity.**

Table 39. Confusion matrix for *operation classification* of N_C=N_{Op}=12 operations using *matched filtering* to evaluate N_{Tst}=4000 testing emissions per operation at SNR_A=30 dB.

Actual Class \ *Declared Class*	ADD.B	ADDC.B	AND.B	BIC.B	BIS.B	BIT.B	CMP.B	DADD.B	MOV.B	SUB.B	SUBC.B	XOR.B
ADD.B	**76**	15	12	242	51	33	1438	153	1805	25	131	19
ADDC.B	89	**31**	28	313	77	32	1783	200	1218	26	184	19
AND.B	93	33	**29**	322	85	42	1911	226	1018	23	199	19
BIC.B	110	31	25	**408**	82	42	2184	228	620	30	215	25
BIS.B	76	24	36	299	**68**	41	1821	190	1214	31	182	18
BIT.B	77	31	22	372	77	**34**	1901	206	1024	32	198	26
CMP.B	91	24	29	389	87	41	**2188**	269	619	29	215	19
DADD.B	94	31	20	303	67	36	1783	**231**	1219	23	175	18
MOV.B	64	22	11	221	50	25	1303	137	**2013**	10	133	11
SUB.B	95	22	25	312	75	31	1984	199	1012	**36**	183	26
SUBC.B	98	23	22	362	92	41	2077	232	821	22	**195**	15
XOR.B	93	17	15	314	69	38	1811	201	1209	27	186	**20**

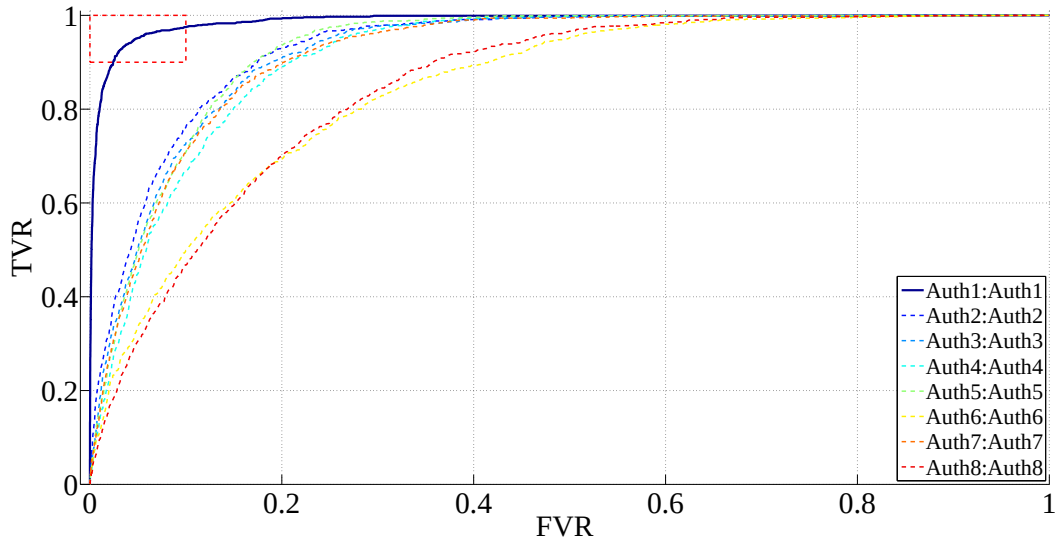

Figure 38. ***Authorized device verification*** **performance for $N_{DevAuth}$=8 authentic devices using *RF-DNA with MDA/ML*. Results obtained from N_{Tst}=4000 testing fingerprints per device at SNR_A=30 dB. Only one device met the arbitrary benchmark of $EER \leq 10\%$, outlined in the top-left corner.**

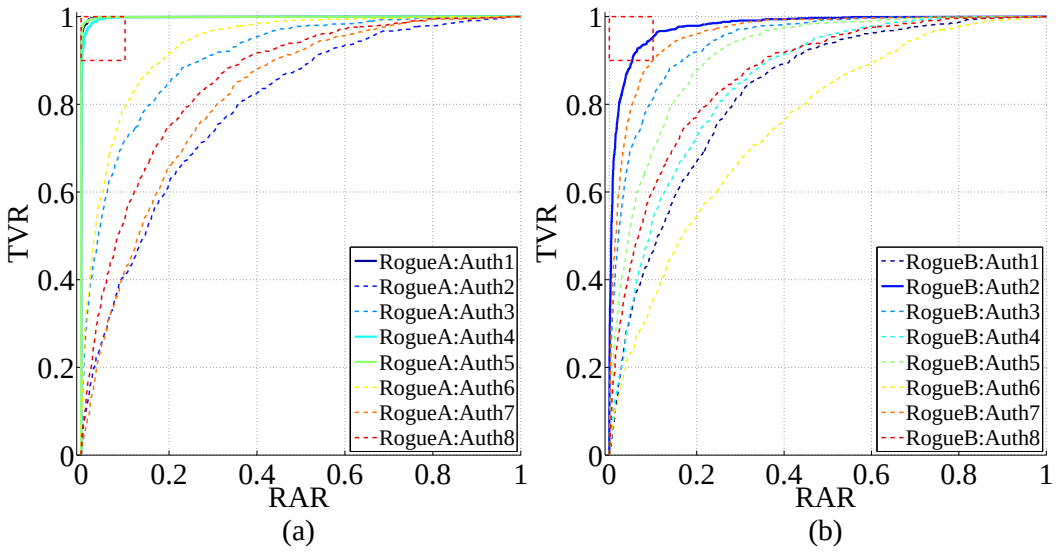

Figure 39. ***Rogue device rejection*** **performance for a) RogueA and b) RogueB using *RF-DNA with MDA/ML*. Results obtained from N_{Tst}=4000 testing fingerprints per device at SNR_A=30 dB. Only three devices met the arbitrary benchmark of $EER \leq 10\%$, outlined in the top-left corner, for RogueA and one devices met the benchmark for RogueB.**

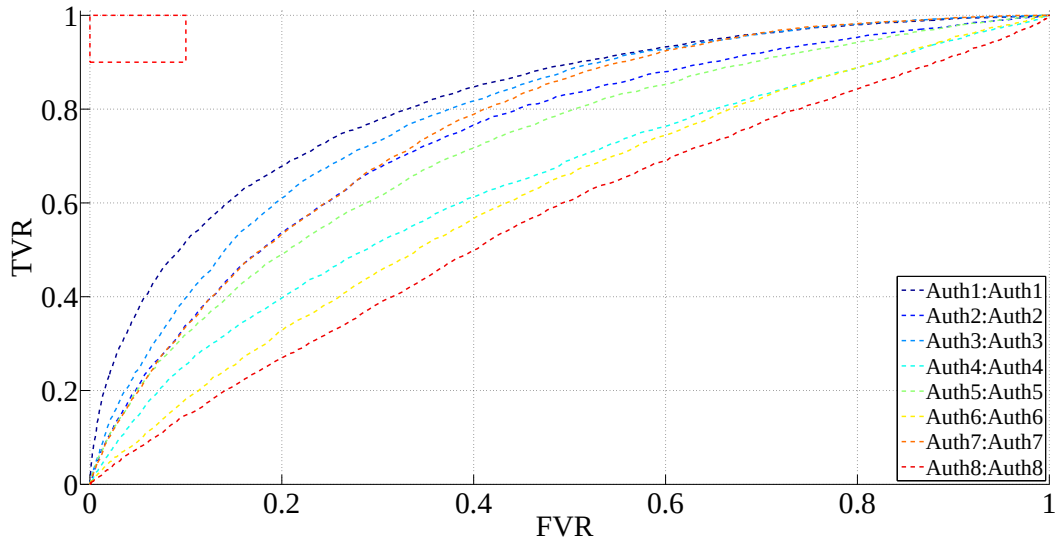

Figure 40. ***Authorized device verification*** **performance for $N_{DevAuth}$=8 authentic devices using *RF-DNA with GRLVQI*. Results obtained from N_{Tst}=4000 testing fingerprints per device at SNR_A=30 dB. No devices met the arbitrary benchmark of $EER \leq 10\%$, outlined in the top-left corner.**

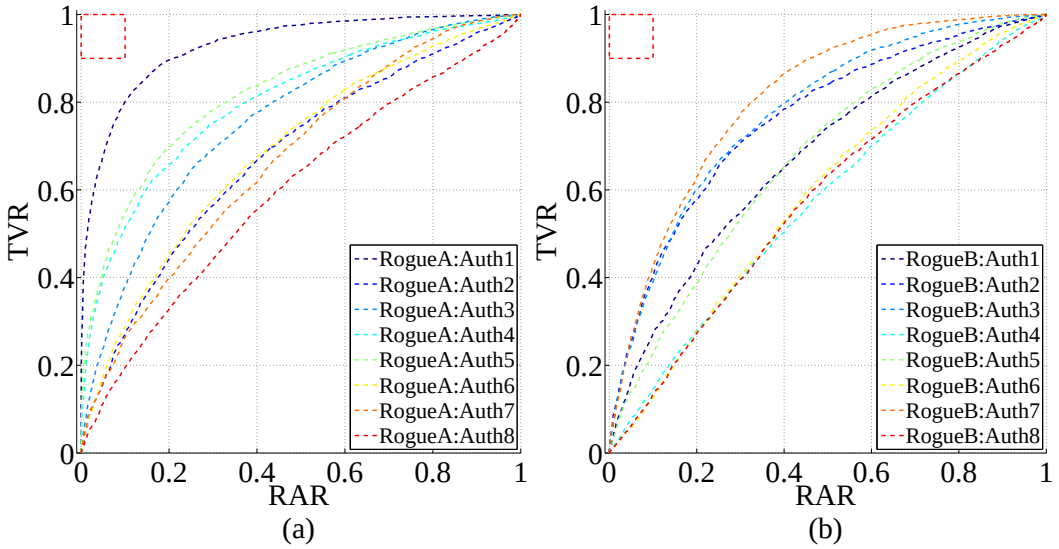

Figure 41. ***Rogue device rejection*** **performance for a) RogueA and b) RogueB using *RF-DNA with GRLVQI*. Results obtained from N_{Tst}=4000 testing fingerprints per device at SNR_A=30 dB. No devices met the arbitrary benchmark of $EER \leq 10\%$, outlined in the top-left corner, for either RogueA or RogueB.**

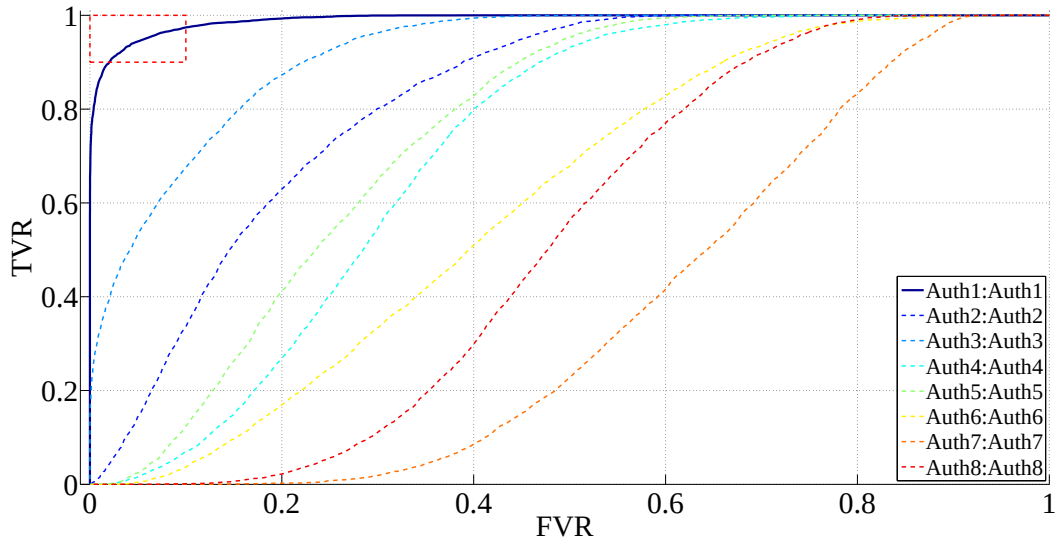

Figure 42. ***Authorized device verification*** **performance for $N_{DevAuth}$=8 authentic devices using** ***matched filtering.*** **Results obtained from N_{Tst}=4000 testing emissions per device at SNR_A=30 dB. Only one device met the arbitrary benchmark of $EER \leq 10\%$, outlined in the top-left corner.**

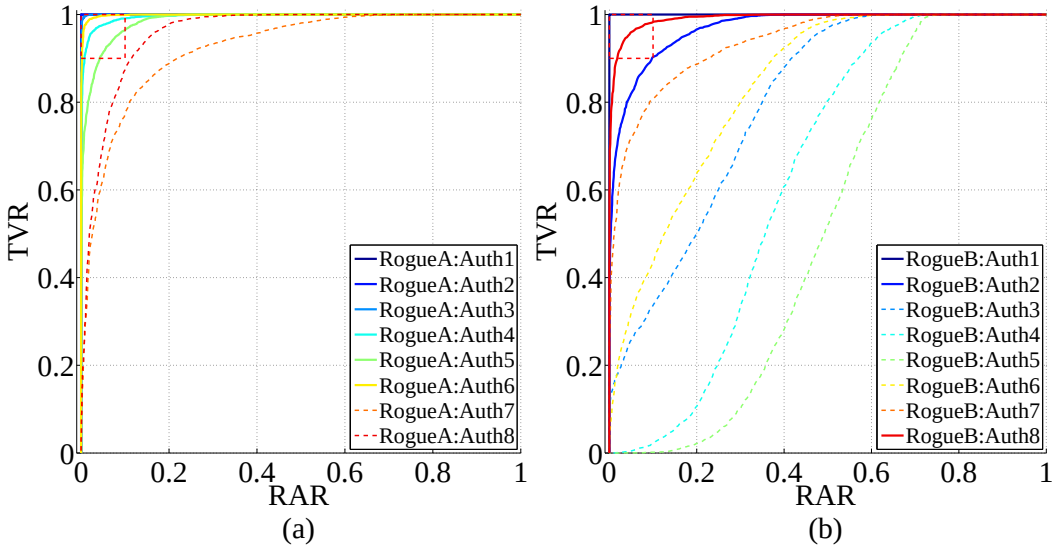

Figure 43. ***Rogue device rejection*** **performance for a) RogueA and b) RogueB using** ***matched filtering.*** **Results obtained from N_{Tst}=4000 testing emissions per device at SNR_A=30 dB. Only six devices met the arbitrary benchmark of $EER \leq 10\%$, outlined in the top-left corner, for RogueA and three devices met the benchmark for RogueB.**

Bibliography

1. Adee, S. The Hunt For The Kill Switch. *Spectrum, IEEE*, 45(5):34–39, May 2008.

2. Agrawal, D., B. Archambeault, J. Rao, and P. Rohatgi. The EM SideChannel(s). In *Cryptographic Hardware and Embedded Systems-CHES 2002*, pages 29–45. Springer, 2003.

3. Agrawal, D., B. Archambeault, S. Suresh, P. Rohatgi, and J. Rao. Advances in Side-Channel Cryptanalysis, Electromagnetic Analysis and Template Attacks. *Cryptobytes, RSA Laboratories*, 6(1):20–32, 2003.

4. Amiel, F., B. Feix, and K. Villegas. In Carlisle Adams, Ali Miri, and Michael Wiener, editors, *Selected Areas in Cryptography*, volume 4876 of *Lecture Notes in Computer Science*. 2007.

5. Archambeau, C., E. Peeters, F. Standaert, and J. Quisquater. Template Attacks in Principal Subspaces. In Louis Goubin and Mitsuru Matsui, editors, *Cryptographic Hardware and Embedded Systems - CHES 2006*, volume 4249 of *Lecture Notes in Computer Science*, pages 1–14. Springer Berlin Heidelberg, 2006.

6. Baranski, W., A. Wytyczak-Partyka, and T. Walkowiak. Computational Complexity Reduction in PCA-Based Face Recognition. *Institute of Computer Engineering, Control and Robotics, Wroclaw University of Technology, Poland*, 2007.

7. Beaumont, M., B. Hopkins, and T. Newby. Hardware Trojans - Prevention, Detection, Countermeasures (A Literature Review). Technical report, Defence Science and Technology Organisation Edinburgh (Australia) Command Control Communications and Intelligence Div, July 2011.

8. Becker, G. *Intentional and Unintentional Side-Channels in Embedded Systems*. PhD thesis, University of Massachusetts Amherst, 2014.

9. Bernstein, K. Trusted Integrated Circuits (TRUST).

10. Boak, D. A History of U.S. Communications Security: The David G. Boak Lectures, July 1973.

11. Brik, V., 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, pages 116–127, New York, NY, USA, 2008. ACM.

12. Carbino, T., M. Temple, and T. Bihl. Ethernet Card Discrimination Using Unintentional Cable Emissions and Constellation-Based Fingerprinting. In *Computing, Networking and Communications (ICNC), 2015 International Conference on*, pages 369–373, 2015.

13. Clark, S., B. Ransford, A. Rahmati, S. Guineau, J. Sorber, K. Fu, and W. Xu. WattsUpDoc: Power Side Channels to Nonintrusively Discover Untargeted Malware on Embedded Medical Devices. In *Proceedings of USENIX Workshop on Health Information Technologies*, volume 2013, 2013.

14. Cobb, W. *Exploitation of Unintentional Information Leakage from Integrated Circuits*. PhD thesis, Air Force Institute of Technology, 2011.

15. Cobb, W., E. Garcia, M. Temple, R. Baldwin, and Y. Kim. Physical Layer Identification of Embedded Devices Using RF-DNA Fingerprinting. In *MILITARY COMMUNICATIONS CONFERENCE, 2010 - MILCOM 2010*, pages 2168–2173, Oct 2010.

16. Cobb, W., E. Laspe, R. Baldwin, M. Temple, and Y. Kim. Intrinsic Physical-Layer Authentication of Integrated Circuits. *Information Forensics and Security, IEEE Transactions on*, 7(1):14–24, Feb 2012.

17. Crawford, M., T. Telesco, C. Nelson, J. Bolton, K. Bagin, and B. Botwin. Defense Industrial Base Assessment: Counterfeit Electronics. Technical Report GAO-13-462T, U.S. Department of Commerce: Bureau of Industry and Security Office of Technology Evaluation, January 2010.

18. CrySyS. Duqu: A Stuxnet-like Malware Found in the Wild. Technical report, Budapest University of Technology and Economics Department of Telecommunications, October 2011.

19. CrySyS. sKyWIper (a.k.a. Flame a.k.a. Flamer): A Complex Malware for Targeted Attacks. Technical report, Budapest University of Technology and Economics Department of Telecommunications, May 2012.

20. Danev, B. and S. Capkun. Transient-Based Identification of Wireless Sensor Nodes. In *Proceedings of the 2009 International Conference on Information Processing in Sensor Networks*, IPSN '09, pages 25–36, Washington, DC, USA, 2009. IEEE Computer Society.

21. Danev, B., H. Luecken, S. Capkun, and K. Defrawy. Attacks on Physical-Layer Identification. In *Proceedings of the Third ACM Conference on Wireless Network Security*, WiSec '10, pages 89–98, New York, NY, USA, 2010. ACM.

22. Danev, B., T. Heydt-Benjamin, and S. Capkun. Attacks on Physical-layer Identification. In *18th Conf on USENIX Security Symposium*, SSYM'09, pages 199–214, 2009.

23. Daudigny, R., H. Ledig, F. Muller, and F. Valette. SCARE of the DES. In John Ioannidis, Angelos Keromytis, and Moti Yung, editors, *Applied Cryptography and Network Security*, volume 3531 of *Lecture Notes in Computer Science*, pages 393–406. Springer Berlin Heidelberg, 2005.

24. DeJean, G. and D. Kirovski. RF-DNA: Radio-Frequency Certificates of Authenticity. In Pascal Paillier and Ingrid Verbauwhede, editors, *Cryptographic Hardware and Embedded Systems - CHES 2007*, volume 4727 of *Lecture Notes in Computer Science*, pages 346–363. Springer Berlin Heidelberg, 2007.

25. Deppensmith, R. and S. Stone. Integrated Circuit (IC) Aging Effects on Radio-Frequency Distinct Native Attributes (RF-DNA). In *Aerospace and Electronics Conference, NAECON 2014 - IEEE National*, pages 331–333, June 2014.

26. Deppensmith, R. and S. Stone. Optimized Fingerprint Generation Using Unintentional Emission Radio-Frequency Distinct Native Attributes (RF-DNA). In *Aerospace and Electronics Conference, NAECON 2014 - IEEE National*, pages 327–330, June 2014.

27. Desmond, L., C. Yuan, T. Pheng, and R. Lee. Identifying Unique Devices Through Wireless Fingerprinting. In *Proceedings of the First ACM Conference on Wireless Network Security*, WiSec '08, pages 46–55, New York, NY, USA, 2008. ACM.

28. Downie, J. and J. Walkup. Optimal Correlation Filters for Images with Signal-Dependent Noise. *J. Opt. Soc. Am. A*, 11(5):1599–1609, May 1994.

29. Dubendorfer, C., B. Ramsey, and M. Temple. An RF-DNA Verification Process for ZigBee Networks. In *MILITARY COMMUNICATIONS CONFERENCE, 2012 - MILCOM 2012*, pages 1–6, Oct 2012.

30. Dudczyk, J. and M. Wnuk. The Utilization of Unintentional Radiation for Identification of the Radiation Sources. In *Microwave Conference, 2004. 34th European*, volume 2, pages 777–780, Oct 2004.

31. Dudczyk, J., J. Matuszewski, and M. Wnuk. Applying the Radiated Emission to the Specific Emitter Identification. In *Microwaves, Radar and Wireless Communications, 2004. MIKON-2004. 15th International Conference on*, volume 2, pages 431–434 Vol.2, May 2004.

32. Edwards, N. Hardware Intrusion Detection for Supply-Chain Threats to Critical Infrastructure Embedded Systems. Master's thesis, University of Illinois at Urbana-Champaign, 2012.

33. Eisenbarth, T., C. Paar, and B. Weghenkel. Building a Side Channel Based Disassembler. In MarinaL. Gavrilova, C.J.Kenneth Tan, and EdwardDavid

Moreno, editors, *Transactions on Computational Science X*, volume 6340 of *Lecture Notes in Computer Science*. 2010.

34. Ellis, K. and N. Serinken. Characteristics of Radio Transmitter Fingerprints. *Radio Science*, 36(4):585–597, 2001.

35. Emilio, I. Are Cyber Weapons Effective Military Tools? *Military and Strategic Affairs*, 7(1), March.

36. Erez, N. and A. Wool. Control Variable Classification, Modeling and Anomaly Detection in Modbus/TCP {SCADA} Systems. *International Journal of Critical Infrastructure Protection*, 10:59–70, September.

37. Garcia, E. Evaluation of the Single Keybit Template Attack. Master's thesis, Air Force Institute of Technology, 2011.

38. Gerdes, R., T. Daniels, M. Mina, and S. Russell. Device Identification via Analog Signal Fingerprinting: A Matched Filter Approach. In *NDSS*, 2006.

39. Goldack, M. and C. Paar. Side-Channel Based Reverse Engineering for Microcontrollers. Master's thesis, Ruhr-Universität Bochum, 2008.

40. Goldman, D. Fake Tech Gear has Infiltrated the U.S. Government, 2012.

41. Gonzalez, C. and J. Reed. Detecting Unauthorized Software Execution in SDR Using Power Fingerprinting. In *MILITARY COMMUNICATIONS CONFERENCE, 2010 - MILCOM 2010*, pages 2211–2216, Oct 2010.

42. Goubin, L. and M. Matsui. Templates vs. Stochastic Methods. In *Cryptographic Hardware and Embedded Systems - CHES 2006*, volume 4249 of *Lecture Notes in Computer Science*, pages 15–29. Springer Berlin Heidelberg, 2006.

43. Guin, U., D. DiMase, and M. Tehranipoor. Counterfeit Integrated Circuits: Detection, Avoidance, and the Challenges Ahead. *Journal of Electronic Testing*, 30(1):9–23, 2014.

44. Hale, G. Stuxnet Effect: Iran Still Reeling, August 2011.

45. Hammer, B. and T. Villmann. Generalized Relevance Learning Vector Quantization. *Neural Networks*, 15(89):1059 – 1068, 2002.

46. Harmer, P. and M. Temple. An Improved LFS Engine for Physical Layer Security Augmentation in Cognitive Networks. In *Computing, Networking and Communications (ICNC), 2013 International Conference on*, pages 719–723, 2013.

47. Harmer, P., M. Temple, M. Buckner, and E. Farquahar. Using Differential Evolution to Optimize 'Learning from Signals' and Enhance Network Security. In *Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation*, GECCO '11, pages 1811–1818, New York, NY, USA, 2011. ACM.

48. Harmer, P., M. Temple, M. Buckner, and E. Farquhar. 4G Security Using Physical Layer RF-DNA with DE-Optimized LFS Classification. *JCM*, 6(9):671–681, 2011.

49. Harmer, P., M. Williams, and M. Temple. Using DE-Optimized LFS Processing to Enhance 4G Communication Security. In *Computer Communications and Networks (ICCCN), 2011 Proceedings of 20th International Conference on*, pages 1 –8, Aug 2011.

50. Hodson, H. Hackers Accessed City Infrastructure via SCADA FBI, November 2011.

51. Holcombe, B. Government Smart Card Handbook. Technical report, US General Service Administration (GSA), February 2004.

52. IC Insights Inc. MCU Market on Migration Path to 32-bit and ARM-based Devices, April 2013.

53. ICS-CERT. Ongoing Sophisticated Malware Campaign Compromising ICS (Update B). Technical Report ICS-ALERT-14-281-01B, Industrial Control Systems Cyber Emergency Readiness Team, December 2014.

54. IHS Technology. Reports of Counterfeit Parts Quadruple Since 2009, Challenging US Defense Industry and National Security, February 2012.

55. IHS Technology. Top 5 Most Counterfeited Parts Represent a $169 Billion Potential Challenge for Global Semiconductor Market, April 2012.

56. Jain, A., A. Ross, and S. Prabhakar. An Introduction to Biometric Recognition. *Circuits and Systems for Video Technology, IEEE Transactions on*, 14(1):4–20, Jan 2004.

57. Ji, Z., S. Hatta, J. Zhang, J. Ma, W. Zhang, N. Soin, B. Kaczer, S. De Gendt, and G. Groeseneken. Negative Bias Temperature Instability Lifetime Prediction: Problems and Solutions. In *Electron Devices Meeting (IEDM), 2013 IEEE International*, pages 15.6.1–15.6.4, Dec 2013.

58. Jin, Y. and Y. Makris. Hardware Trojan Detection Using Path Delay Fingerprint. In *Hardware-Oriented Security and Trust, 2008. HOST 2008. IEEE International Workshop on*, pages 51–57, June 2008.

59. Kam, T. and M. Basu. Complexity Measures of Supervised Classification Problems. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(3):289–300, Mar 2002.

60. Kasmi, M., A. Mostafa, and J. Lanet. Methodology to Reverse Engineer a Scrambled Java Card Virtual Machine Using Electromagnetic Analysis. In *Next Generation Networks and Services (NGNS), 2014 Fifth International Conference on*, pages 278–281, May 2014.

61. Kasper, M., T. Kasper, A. Moradi, and C. Paar. Breaking KeeLoq in a Flash: On Extracting Keys at Lightning Speed. In Bart Preneel, editor, *Progress in Cryptology - AFRICACRYPT 2009*, volume 5580 of *Lecture Notes in Computer Science*, pages 403–420. Springer Berlin Heidelberg, 2009.

62. Keller, W. *Advanced Detection of Electronic Counterfeits (ADEC)*, June 2014.

63. Keller, W. and B. Pathak. Integrated Circuit with Electromagnetic Energy Anomaly Detection and Processing, March 2012. US Patent App. 13/410,909.

64. Keller, W., S. Freeman, and J. Galyardt. System and Method for Physically Detecting Counterfeit Electronics, September 6 2012. US Patent App. 13/410,797.

65. Kheir, M., H. Kreft, I. Hölken, and R. Knöchel. On the Physical Robustness of RF On-Chip Nanostructured Security. *Journal of Information Security and Applications*, 19(4):301–307, 2014.

66. Klein, R. *Application of Dual-Tree Complex Wavelet Transforms to Burst Detection and RF Fingerprint Classification*. PhD thesis, Air Force Institute of Technology, Sep 2009.

67. Klein, R., M. Temple, and M. Mendenhall. Application of Wavelet-Based RF Fingerprinting to Enhance Wireless Network Security. *Journal of Communications and Networks*, 11(6):544; 12; 114–555, Dec 2009.

68. Klein, R., M. Temple, M. Mendenhall, and D. Reising. Sensitivity Analysis of Burst Detection and RF Fingerprinting Classification Performance. In *IEEE International Conference on Communications, 2009. ICC '09.*, pages 1 –5, Jun 2009.

69. Kocher, P. Timing Attacks on Implementations of Diffie-Hellman, RSA, DSS, and Other Systems. In Neal Koblitz, editor, *Advances in Cryptology CRYPTO 96*, volume 1109 of *Lecture Notes in Computer Science*, pages 104–113. Springer Berlin Heidelberg, 1996.

70. Kocher, P., J. Jaffe, and B. Jun. Differential Power Analysis. In Michael Wiener, editor, *Advances in Cryptology CRYPTO 99*, volume 1666 of *Lecture Notes in Computer Science*, pages 388–397. Springer Berlin Heidelberg, 1999.

71. Krishnan, S., V. Narayanan, E. Cartier, D. Ioannou, K. Zhao, T. Ando, U. Kwon, B. Linder, J. Stathis, M. Chudzik, A. Kerber, and K. Choi. Bias Temperature Instability in High-K/Metal Gate Transistors - Gate Stack Scaling Trends. In *Reliability Physics Symposium (IRPS), 2012 IEEE International*, pages 5A.1.1–5A.1.6, April 2012.

72. Kumhyr, D., G. Johnson, W. Bulko, J. Vargas, and Y. Young. Electromagnetic Profiling to Validate Electronic Device Authenticity, Aug 2012. US Patent 8,242,793.

73. Kuo, S., B. Lee, and W. Tian. *Real-Time Digital Signal Processing: Fundamentals, Implementations and Applications*. John Wiley & Sons, second edition, 2006.

74. Li, T., S. Zhu, and M. Ogihara. Using Discriminant Analysis for Multi-Class Classification: an Experimental Investigation. *Knowledge and Information Systems*, 10(4):453–472, 2006.

75. Lipovsky, R. and A. Cherepanov. BlackEnergy Trojan Strikes Again: Attacks Ukrainian Electric Power Industry, January 2016.

76. Mackenzie, H. Shamoon Malware and SCADA Security What are the Impacts?, October 2012.

77. Mendenhall, M. and E. Merenyi. Relevance-Based Feature Extraction for Hyperspectral Images. *Neural Networks, IEEE Transactions on*, 19(4):658–672, April 2008.

78. Messerges, T., E. Dabbish, and R. Sloan. Power Analysis Attacks of Modular Exponentiation in Smartcards. In etinK. Ko and Christof Paar, editors, *Cryptographic Hardware and Embedded Systems*, volume 1717 of *Lecture Notes in Computer Science*, pages 144–157. Springer Berlin Heidelberg, 1999.

79. Mini-Circuits. *BLP-90+ Coaxial Low Pass Filter Datasheet*.

80. Montminy, D. *Enhancing Electromagnetic Side-Channel Analysis in an Operational Environment*. PhD thesis, Air Force Institute of Technology, 2013.

81. Msgna, M., K. Markantonakis, and K. Mayes. Precise Instruction-Level Side Channel Profiling of Embedded Processors. In Xinyi Huang and Jianying Zhou, editors, *Information Security Practice and Experience*, volume 8434 of *Lecture Notes in Computer Science*, pages 129–143. Springer International Publishing, 2014.

82. Msgna, M., K. Markantonakis, D. Naccache, and K. Mayes. Verifying Software Integrity in Embedded Systems: A Side Channel Approach. In Emmanuel Prouff, editor, *Constructive Side-Channel Analysis and Secure Design*, volume

8622 of *Lecture Notes in Computer Science*, pages 261–280. Springer International Publishing, 2014.

83. Novak, R. Side-Channel Based Reverse Engineering of Secret Algorithms. In *Proceedings of the Twelfth International Electrotechnical and Computer Science Conference (ERK 2003), Ljubljana, Slovenia, September*, pages 25–26. Citeseer, 2003.

84. Obama, B. Executive Order 13636: Improving Critical Infrastructure Cybersecurity. Technical report, The White House, Washington DC: U.S. Government, February 2013.

85. Office of the Under Secretary of Defense for Acquisition, Technology and Logistics. Report of the Defense Science Board Task Force on High Performance Microchip Supply, 2005.

86. Ott, H. *Electromagnetic Compatibility*. John Wiley and Sons, Inc., 2009.

87. Pey, K., N. Raghavan, X. Li, W. Liu, K. Shubhakar, X. Wu, and M. Bosman. New Insight into the TDDB and Breakdown Reliability of Novel High-K Gate Dielectric Stacks. In *Reliability Physics Symposium (IRPS), 2010 IEEE International*, pages 354–363, May 2010.

88. PFP Cybersecurity Inc. *Supply Chain Protection: A White Paper on Counterfeit Detection*, 2015.

89. Proakis, J. *Digital Communications*. McGraw-Hill, 2000.

90. Reising, D. *Exploitation of RF-DNA for Device Classification and Verification Using GRLVQI Processing*. PhD thesis, Air Force Institute of Technology, 2012.

91. Reising, D. and M. Temple. WiMAX Mobile Subscriber Verification Using Gabor-Based RF-DNA Fingerprints. In *Communications (ICC), 2012 IEEE International Conference on*, pages 1005–1010, June 2012.

92. Reising, D., M. Temple, and J. Jackson. Authorized and Rogue Device Discrimination Using Dimensionally Reduced RF-DNA Fingerprints. *Information Forensics and Security, IEEE Transactions on*, 10(6):1180–1192, June 2015.

93. Reising, D., M. Temple, and M. Oxley. Gabor-based RF-DNA Fingerprinting for Classifying 802.16e WiMAX Mobile Subscribers. In *Computing, Networking and Communications (ICNC), 2012 International Conference on*, pages 7 –13, Feb 2012.

94. Riscure. *EM Probe Station Datasheet*, 2011.

95. Rodriguez, J., A. Perez, and J. Lozano. Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(3):569–575, March 2010.

96. Sato, A. and K. Yamada. Generalized Learning Vector Quantization. *Advances in neural information processing systems*, pages 423–429, 1996.

97. Scarfone, K. and P. Mell. Guide to Intrusion Detection and Prevention Systems (IDPS). *NIST special publication*, 800(2007):94, 2007.

98. Skahill, P. Department of Defense Gives UConn Millions for Hardware Security, May 2014.

99. Sklar, B. *Digital Communications*, volume 2. Prentice Hall NJ, 2001.

100. Stone, B. and S. Stone. Radio Frequency Based Reverse Engineering of Microcontroller Program Execution. In *Aerospace and Electronics Conference, NAECON 2015 - IEEE National*, 2015.

101. Stone, B. and S. Stone. Comparison of Radio Frequency Based Techniques for Device Discrimination and Operation Identification. In *International Conference on Cyber Warfare and Security ICCWS-2016*, 2016.

102. Stone, S. *Radio Frequency Based Programmable Logic Controller Anomaly Detection*. PhD thesis, Air Force Institute of Technology, 2013.

103. Stone, S. and M. Temple. Radio-Frequency-Based Anomaly Detection for Programmable Logic Controllers in the Critical Infrastructure. *International Journal of Critical Infrastructure Protection*, 5(2):66 – 73, 2012.

104. Stone, S. and M. Temple. Detecting Anomalous SCADA Operation Using RF-Based Hilbert Transforms. *International Journal of Critical Infrastructure Protection*, 5(2):11–33, July 2013.

105. Stone, S., M. Temple, and R. Baldwin. Detecting Anomalous Programmable Logic Controller Behavior Using RF-Based Hilbert Transform Features and a Correlation-Based Verification Process. *International Journal of Critical Infrastructure Protection*, (0):–, 2015.

106. Stradley, J. and D. Karraker. The Electronic Part Supply Chain and Risks of Counterfeit Parts in Defense Applications. *Components and Packaging Technologies, IEEE Transactions on*, 29(3):703–705, Sept 2006.

107. Strobel, D. *Novel Applications for Side-Channel Analyses of Embedded Microcontrollers*. PhD thesis, Ruhr-Universität Bochum, 2014.

108. Strobel, D., F. Bache, D. Oswald, F. Schellenberg, and C. Paar. SCANDALee: A Side-channel-based Disassembler Using Local Electromagnetic Emanations. In *Proceedings of the 2015 Design, Automation & Test in Europe Conference & Exhibition*, DATE '15, pages 139–144, San Jose, CA, USA, 2015. EDA Consortium.

109. Suski, W., M. Temple, M. Mendenhall, and R. Mills. Radio Frequency Fingerprinting Commercial Communication Devices to Enhance Electronic Security. *Int. J. Electron. Secur. Digit. Forensic*, 1:301–322, Oct 2008.

110. Suski, W., M. Temple, M. Mendenhall, and R. Mills. Using Spectral Fingerprints to Improve Wireless Network Security. In *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, pages 1–5, Dec 2008.

111. Symantec Security Response. W32.Duqu: The Precursor to the Next Stuxnet. Technical report, Symantec.

112. Symantec Security Response. W32.Stuxnet Dossier. Technical report, Symantec, February 2011.

113. Tech-FAQ. The OSI Model What It Is; Why It Matters; Why It Doesnt Matter.

114. Texas Instruments Inc. *AES128 A C Implementation for Encryption and Decryption*, Mar 2009.

115. Texas Instruments Inc. *MSP-EXP430F5529 Experimenter Board User's Guide*, Jun 2011.

116. Texas Instruments Inc. Msp430 Von Neumann or Harvard, 2012.

117. Texas Instruments Inc. *MSP430F551x, MSP430F552x Mixed Signal Microcontroller*, May 2013.

118. Texas Instruments Inc. *MSP430x5xx and MSP430x6xx Family User's Guide*, May 2014.

119. Texas Instruments Inc. Microcontrollers with Innovation at the Core, 2015.

120. Texas Instruments Inc. *MSP430F5529 Device Erratasheet*, December 2015.

121. Theodoridis, S. and K. Koutroumbas. *Pattern Recognition*. Academic Press, fourth edition, 2009.

122. USAF AFCERT. AFCERT Operations Metrics. Technical report, San Antonio, TX: USAF, December 2011.

123. Vermoen, D., M. Witteman, and G. Gaydadjiev. Reverse Engineering Java Card Applets Using Power Analysis. In Damien Sauveron, Konstantinos Markantonakis, Angelos Bilas, and Jean-Jacques Quisquater, editors, *Information Security Theory and Practices. Smart Cards, Mobile and Ubiquitous Computing Systems*, volume 4462 of *Lecture Notes in Computer Science*, pages 138–149. Springer Berlin Heidelberg, 2007.

124. Williams, M., M. Temple, and D. Reising. Augmenting Bit-Level Network Security Using Physical Layer RF-DNA Fingerprinting. In *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*, pages 1–6. IEEE, 2010.

125. Williams, M., S. Munns, M. Temple, and M. Mendenhall. RF-DNA Fingerprinting for Airport WiMax Communications Security. In *Network and System Security (NSS), 2010 4th International Conference on*, pages 32–39, Sep 2010.

126. Wilshusen, G. Cybersecurity: A Better Defined and Implemented National Strategy Is Needed to Address Persistent Challenges. Technical Report GAO-13-462T, U.S. Government Accountability Office, March 2013.

127. Wright, B. PLC Hardware Discrimination using RF-DNA fingerprinting. Master's thesis, Air Force Institute of Technology, 2014.

128. Wylie, J. and S. Stone. Detecting Anomalous Behavior in Microcontrollers Using Unintentional Radio Frequency Emissions. In *Aerospace and Electronics Conference, NAECON 2015 - IEEE National*, 2015.

129. Wylie, J., S. Stone, and B. Mullins. Detecting a Weakened Encryption Algorithm in Microcontrollers Using CorrelationBased Anomaly Detection. In *International Conference on Cyber Warfare and Security ICCWS-2016*, 2016.

130. Yudken, J. Manufacturing Insecurity: Americas Manufacturing Crisis and the Erosion of the U.S. Defense Industrial Base. Technical report, AFL-CIO Industrial Union Council, September 2010.

131. Zetter, K. How Digital Detectives Deciphered Stuxnet, the Most Menacing Malware in History. *Wired Magazine*, 11:1–8, 2011.

132. Zetter, K. Cyberattack Causes Physical Damage for the Second Time Ever, January 2015.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
24–03–2016	Master's Thesis	Dec 2014 – Feb 2016

4. TITLE AND SUBTITLE

Comparison of Radio Frequency Distinct Native Attribute and Matched Filtering Techniques for Device Discrimination and Operation Identification

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Stone, Barron D., 1st Lt, 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-ENG-MS-16-M-048

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

Air Force Research Laboratory, Integrated Electronic & Net-Centric Warfare Div
Attn: Yong C. Kim
2241 Avionics Circle
Wright-Patterson AFB, OH 45433-7322
(937) 528-8026 (DSN 798-8062) yong.kim@us.af.mil

10. SPONSOR/MONITOR'S ACRONYM(S)

AFRL/RYWA

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

The research presented here provides a comparison of the classification, verification, and computational time performance of three techniques to analyze unintentional radio-frequency (RF) emissions (URE) from semiconductor devices for the purposes of device discrimination and operation identification. URE from ten MSP430F5529 16-bit microcontrollers were analyzed using: 1) RF distinct native attributes (RF-DNA) fingerprints paired with multiple discriminant analysis/maximum likelihood (MDA/ML) classification, 2) RF-DNA fingerprints paired with generalized relevance learning vector quantized-improved (GRLVQI) classification, and 3) time domain (TD) signals paired with matched filtering. These techniques were considered for potential applications to detect counterfeit/Trojan hardware infiltrating supply chains and to defend against cyber attacks by monitoring the executed operations of embedded systems in critical supervisory control and data acquisition (SCADA) networks.

15. SUBJECT TERMS

RF-DNA, MDA/ML, GRLVQI, Matched Filter, Hardware Discrimination, SCARE, MSP430, Unintentional Emissions

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Maj Samuel Stone, AFIT/ENG
U	U	U	U	147	19b. TELEPHONE NUMBER *(include area code)* (937) 785-3636, x6137; samuel.stone@afit.edu