Side Channel Anomaly Detection in Industrial Control Systems Using Physical Characteristics of End Devices

THESIS

Ryan D. Harris

AFIT-ENG-MS-19-M-032

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENG-MS-19-M-032

SIDE CHANNEL ANOMALY DETECTION IN INDUSTRIAL CONTROL SYSTEMS USING PHYSICAL CHARACTERISTICS OF END DEVICES

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

Ryan D. Harris, B.S.

March 2019

AFIT-ENG-MS-19-M-032

SIDE CHANNEL ANOMALY DETECTION IN INDUSTRIAL CONTROL SYSTEMS USING PHYSICAL CHARACTERISTICS OF END DEVICES

THESIS

Ryan D. Harris, B.S.

Committee Membership:

Robert F. Mills, Ph.D.
Chair

Barry E. Mullins, Ph.D., P.E.
Member

Mr. Stephen J. Dunlap
Member

AFIT-ENG-MS-19-M-032

Abstract

Industial Control Systems (ICS) are described by the Department of Homeland Security as systems that are so “vital to the United States that their incapacity or destruction would have a debilitating impact on our physical or economic security.” Attacks like Stuxnet show that these systems are vulnerable. The end goal for Stuxnet was to spin centrifuges at a frequency rate outside of normal operation and hide its activity from the ICS operator. This research aims to provide a proof of concept for an anomaly detection system that would be able to detect an attack like Stuxnet by measuring the physical change in vibration caused by the attack. The attack can hide what is reported to the operator, but it cannot hide the physical changes caused by the attack. This research uses a piezoelectric vibration sensor to collect vibration data coming from a centrifugal pump and flow meter on an ICS training system at each operating level. The collected data is then fingerprinted and classified using established RF-DNA techniques to determine if it can differentiate between the vibrations produced at each of the operating level. A clear differentiation between operating levels indicates that an ADS is feasible.

Table of Contents

Page

List of Figures

Figure Page

List of Tables

List of Acronyms

AC	Alternating Current	4
ADC	Analog to Digital Converter	6
ADS	Anomaly Detection System	3
CSV	comma-separated values	35
DHS	Department of Homeland Security	1
DFT	Discrete Fourier Transform	22
FLD	Fisher's Linear Discriminant	24
HMI	Human Machine Interface	10
ICS	Industrial Control Systems	2
IT	Information Technology	2
IDE	Integrated Development Environment	33
MDA/ML	Multiple Discriminant Analysis/Maximum Likelihood	23
NIST	National Institute of Standards and Technology	10
PLC	Programmable Logic Controllers	2
PSD	Power Spectral Density	22
RF	Radio Frequency	13
RF-DNA	Radio Frequency Distinct Native Attribute	7
ROI	Region of Interest	8
RPM	Revolutions Per Minute	4
SCADA	Supervisory Control and Data Acquisition	2
SNR	Signal-to-Noise Ratio	23

SD	Spectral Domain	7
TD	Time Domain	22
WS-DNA	Wired Signal Distinct Native Attribute	13

SIDE CHANNEL ANOMALY DETECTION IN INDUSTRIAL CONTROL SYSTEMS USING PHYSICAL CHARACTERISTICS OF END DEVICES

I. Introduction

1.1 Introduction

The Department of Homeland Security (DHS) describes critical infrastructure as "the physical and cyber systems and assets that are so vital to the United States that their incapacity or destruction would have a debilitating impact on our physical or economic security or public health or safety"[1]. Presidential Policy Directive 21 [2] identifies sixteen critical infrastructure sectors, they are:

- Chemical
- Commercial Facilities
- Communications
- Critical Manufacturing
- Dams
- Defense Industrial Base
- Emergency Services
- Energy
- Financial Services
- Food and Agriculture
- Government Facilities
- Healthcare and Public Health
- Information Technology
- Nuclear Reactors, Materials, and Waste
- Transportation Systems
- Water and Wastewater Systems

In short, critical infrastructure permeates nearly every aspect of our day to day lives in some way or another. Many of these sectors are controlled and run by Industrial Control Systems (ICS). ICS is a general term that covers different types of control systems such as Supervisory Control and Data Acquisition (SCADA) and devices like Programmable Logic Controllers (PLC)'s [3]. Protecting from and detecting attacks against ICS is extremely important for national security.

1.2 Motivation

Originally, ICSs were closed off and isolated from external networks. An operator would have to travel to the site to access and control the system. This has changed over time with the introduction of standard Information Technology (IT) networks into the ICS networks. This has been done in part because it can make the systems more efficient, more cost effective, and easier to manage as it can provide remote access into the system [3]. Along with the new capabilities this merger provides, it also brings new threats. An attacker can now be thousands of miles away while they launch their attack. Preventing these attacks should be one of the primary goals for securing ICS. However, what if an attack is successful in penetrating into the ICS network? Detecting that attack in order to mitigate any damage done is also immensely important.

One of the most famous and successful attacks against ICS is Stuxnet. Based on examinations of Stuxnet [4], it was determined that it was specifically targeting Iran's nuclear enrichment facilities. In particular its goal was to change the rate in which the facilities centrifuges would spin. At the same time, it reported back to the operator that everything was normal and that the centrifuges were spinning at the correct rate. It is this type of activity that this research aims to identify by sensing the physical difference, in this case vibration, between normal operation and not normal operation (an anomaly). The attack can hide/change what values are reported back to the operator, but it cannot hide the physical effects that the attack causes in the target device.

1.3 Research Goals

The goals of this research are the following:

1. Determine if the physical effects of devices in an ICS can be measured, fingerprinted, and classified with enough detail to be able to differentiate between the different levels of operation.
2. Determine the viability of an Anomaly Detection System (ADS) based on the results of the data fingerprinting and classification.

The results from these goals will provide the proof of concept for a side channel ADS that runs on a closed network that is out of band from the system that it is monitoring.

1.4 Approach

1.4.1 Testing Environment.

A Lab-Volt® 3531 pressure, flow, level, and temperature process training system is used for the testing environment. Figure 1 shows the training system. For this research the training system is configured to circulate water through a flow meter and back to the reservoir tanks. However, during data collection on the centrifugal pump the flow meter is not necessary and therefore a bypass is used to circulate water directly back to the tanks without flowing through the flow meter. The bypass is used to eliminate as many unknown variables as possible during pump data collection.

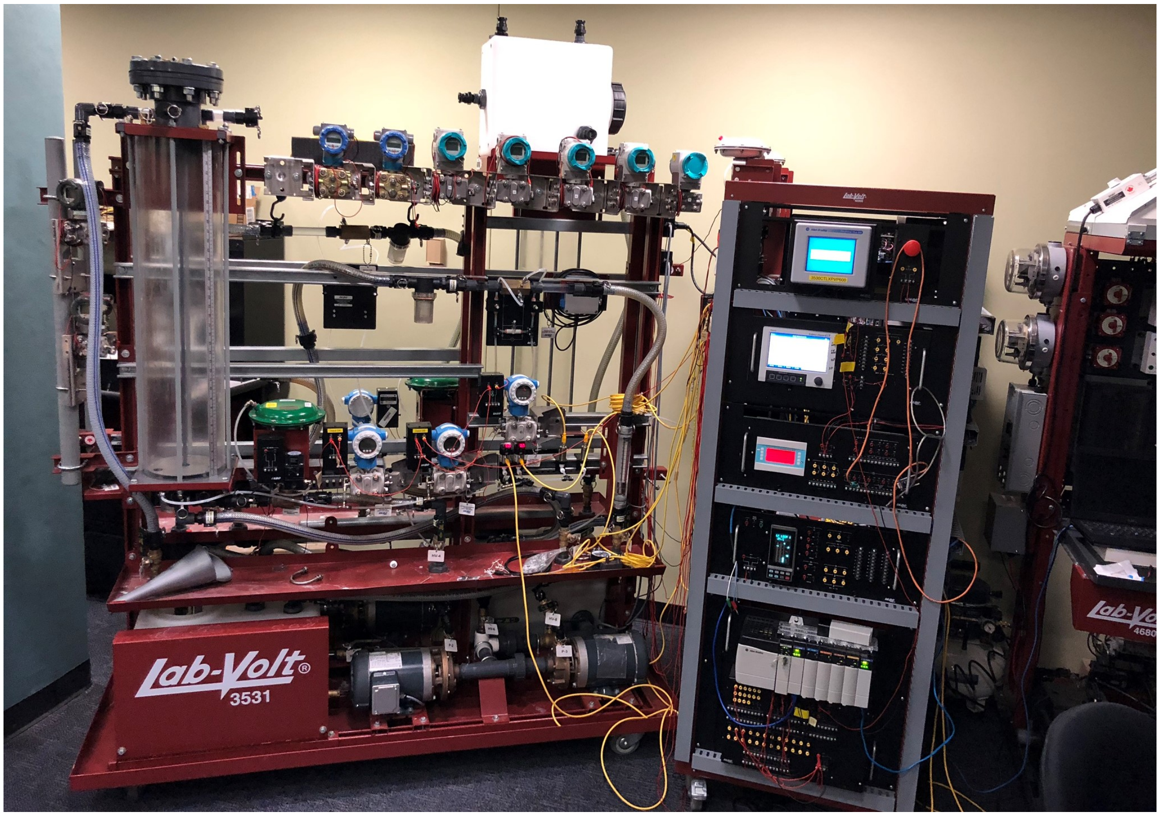


Figure 1. Lab-Volt® 3531 Training System

1.4.1.1 Centrifugal Pump.

The Marathon Electric® centrifugal pump as shown in Figure 2 is the main focus of this research. It uses a 1 horse power three-phase electric motor that spins at a maximum of 3450 Revolutions Per Minute (RPM). The pump has 60 different operating states based on the frequency of the Alternating Current (AC) power being sent to it. The slowest state is at 1 Hz and the highest is at 60 Hz. Vibration data will be collected for each of the 60 operating states.

1.4.1.2 Flow Meter.

The other device used for this research is the flow meter. With the bypass turned off, water is redirect through the flow meter before returning to the reservoir tanks. The water flowing through this meter and causing a vibration is the effect being

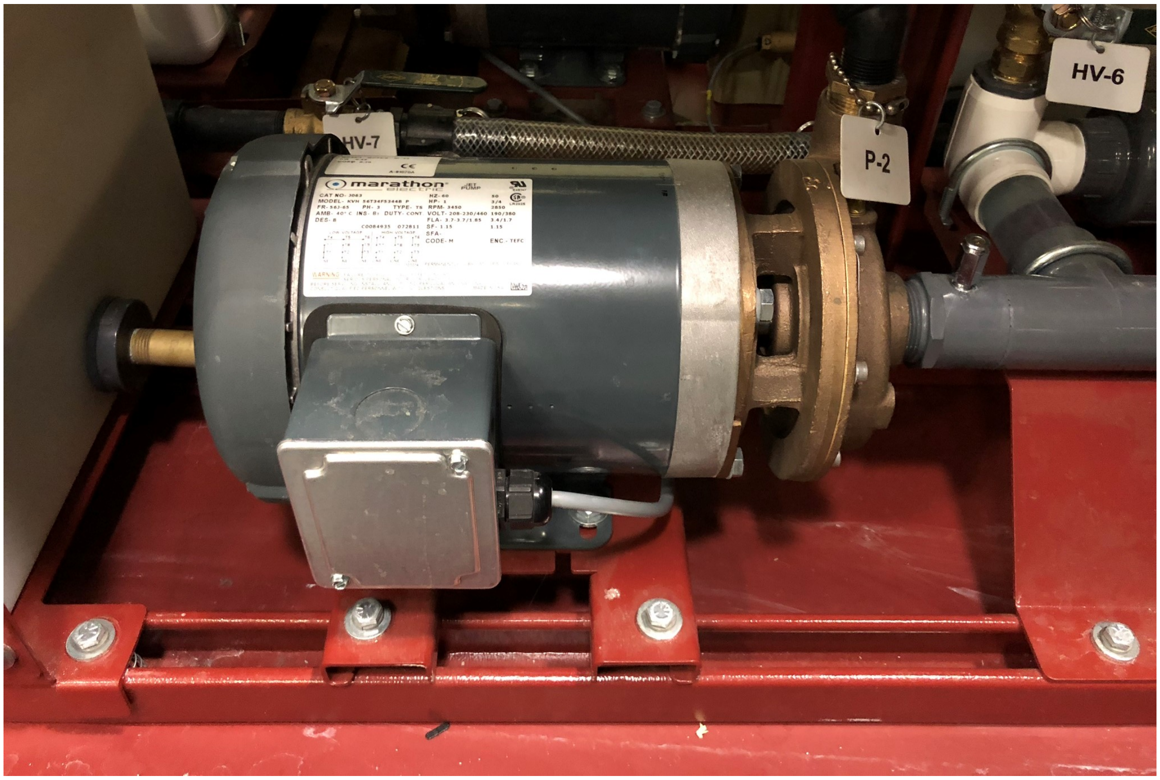


Figure 2. Marathon Electric® Centrifugal Pump

measured for this research. The flow meter returns the flow rate as liters per minute. Figure 3 shows the part of the flow meter that the vibration measurements are taken from.

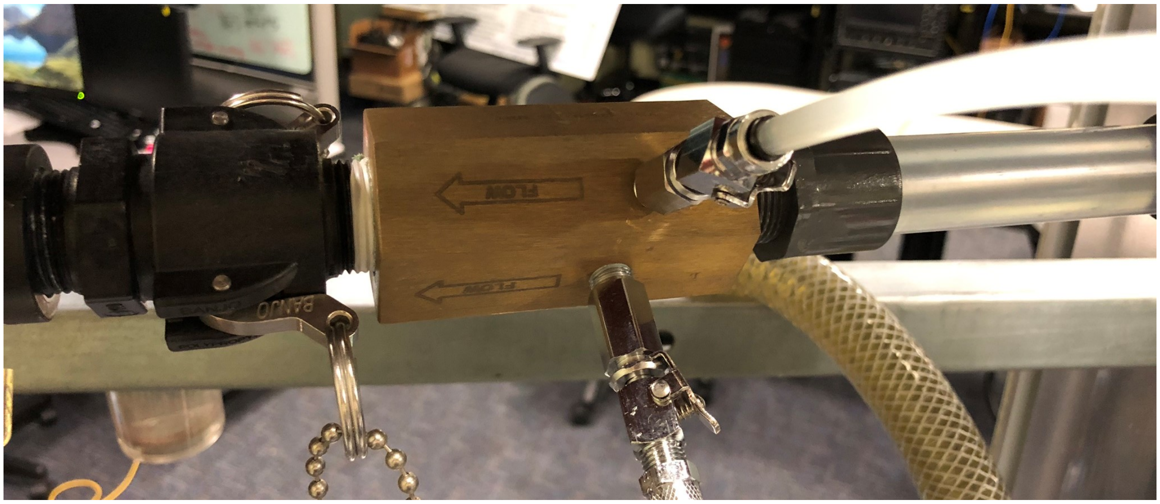


Figure 3. Flow Meter

1.4.2 Testing Equipment.

An Arduino Uno is used to provide the platform for the sensor. The Arduino was chosen because it has a built in 10-bit Analog to Digital Converter (ADC) that allows for the analog input from the sensor to be converted to a digital value between 0 and 1023 based on the analog input.

A piezoelectric disc is used for the vibration sensor. Using the piezoelectric effect, the sensor is able to generate an electric charge when a mechanical stress is applied to it [5]. The electric charge is then sent to the ADC of the Arduino for conversion to a digital value.

A custom Arduino shield is used to boost the signal from the vibration sensor. This makes the sensor more sensitive and makes the data collected easier to work with and analyze. Figure 4 shows the entire assembled sensor.

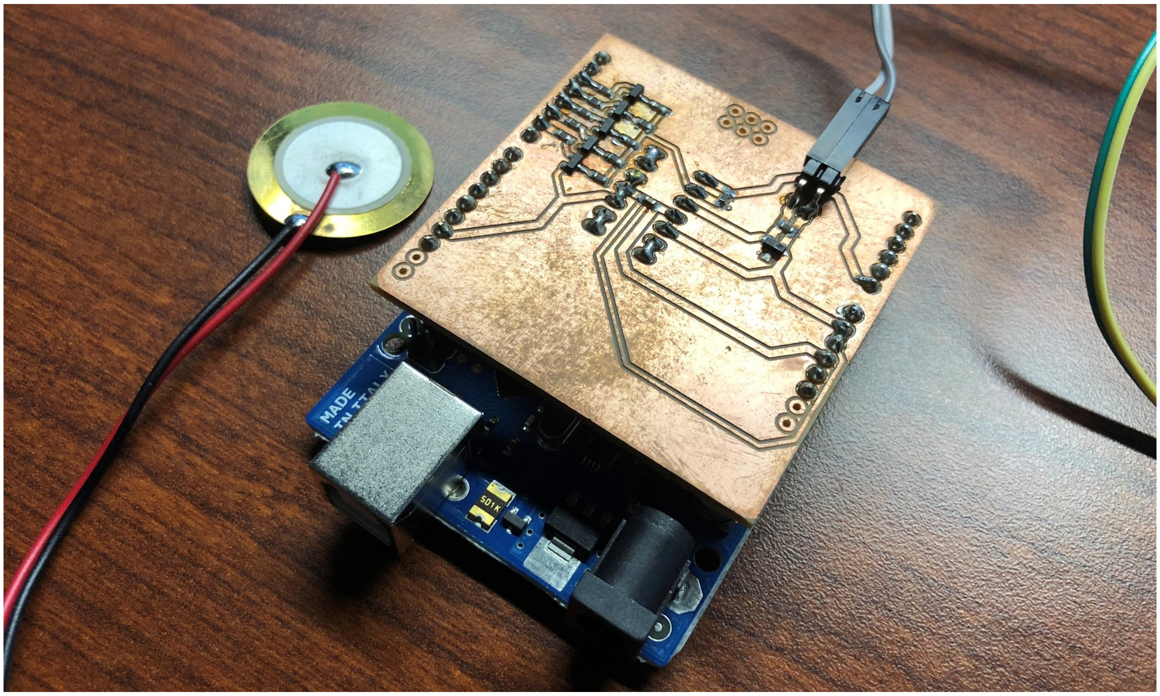


Figure 4. Vibration Sensor

1.4.3 Experimentation.

Data collection with the vibration sensor takes place in two locations on the system. The first is attached directly to the centrifugal pump where the sensor can measure the vibration from the rotation of the rotor inside the pump. The second is attached to a flow meter that has been isolated from the frame of the system so that it picks up the vibration from the flow of water moving through the pipes with as little interference from the pump as possible. Measurements are taken at every operating level/state of the pump. The pump has 60 operating states from barely running to full speed. Only 51 measurements are taken from the flow meter because when the pump is below level 10 it is not running fast enough to force water through the system. Therefore, no water is pushed through the flow meter.

After the data is collected it is in the form of 60 data captures containing 600,000 measurements/samples each for the pump data and 51 data captures of the same size for flow meter data. The data captures are then formatted for analysis. This involves restructuring the data for each capture into burst with 3000 samples per burst.

Data analysis is performed in MATLAB® using Radio Frequency Distinct Native Attribute (RF-DNA) tools and techniques in the Spectral Domain (SD). Every data capture representing an operating state of the pump is fingerprinted and classified to produce a confusion as with the final result from classification. More details are in Chapter III.

1.5 Assumptions and Limitations

This research provides a proof of concept for using the physical effects of ICS devices to implement an ADS. This section describes some of the assumptions and limitations for this research.

1.5.1 Testing Environment.

The testing environment for this research is a training system in a controlled lab environment. All parts of the system used for testing are attached to the same frame and may subject the vibration sensor to unwanted readings from other devices on the system. Steps have been taken to isolate the flow meter from the frame but it was unpractical to move the pump. Additionally, there are other research projects taking place in the lab that may introduce unwanted noise.

Real world ICS environments are typically very large, complicated, and exposed to harsh environments that could introduce a number of unknown and unpredictable variables. The use of the training system in this research limits these variables.

1.5.2 Data Collection.

The specific vibration frequency produced from the vibration on the pump and flow meter is unknown. Therefore a high sample rate of 2807 Hz is used for data collection. The sample rate was chosen based on the highest rate achievable without overwhelming the serial connection that data was being logged to.

1.5.3 Data Analysis.

With previous RF-DNA studies, there is usually a known part of the signal that becomes the Region of Interest (ROI) for analysis. Because this is an unknown in this research, the entire collected signal will be used as the ROI.

1.5.4 Scope.

The scope of this research is narrow and focused on the proof of concept for an ADS using the physical effects created by the operation of ICS devices. The first part of this research is on the ability to measure the physical vibration made by the pump and the vibration made by the flow meter as water passes through it. The second part is on the ability of each data capture representing a devices operation state to be uniquely fingerprinted and classified using RF-DNA techniques. The final part is to determine if the results from classification indicate that if this process can be used for an ADS in the future.

1.6 Thesis Overview

Chapter II contains a more detailed background and an overview of related work. Chapter III presents the methodology used for this research. Chapter IV covers the analysis implementation and results. Chapter V summarizes the research and provide recommendations for future work.

II. Background

2.1 Introduction

This chapter provides the background information and related works used for this research. Section 2.2 gives the background information on ICS. Section 2.3 provides an overview of Stuxnet and how it is a motivation for this research. Section 2.4 gives an overview of the related work for this research. Section 2.5 provides the background on electric motors like the one used in this research. Section 2.6 gives the background on the piezoelectric effect that the vibration sensor in this research takes advantage of.

2.2 Industrial Control Systems

According to National Institute of Standards and Technology (NIST) [3], "ICS are critical to the operation of the U.S. critical infrastructures that are often highly interconnected and mutually dependent systems." Many ICS installations have evolved with the insertion of traditional IT capabilities into them. With this insertion comes all of the security risks and vulnerabilities that are associated with IT systems that were previously not a concern for ICS. Cost and performance improvements have been the driving factor for this evolution with securing these newly connected system as somewhat of an afterthought.

"ICS is a general term that encompasses several types of control systems, including supervisory control and data acquisition (SCADA) systems, distributed control systems (DCS), and other control system configurations such as Programmable Logic Controllers (PLC) often found in the industrial sectors and critical infrastructures [3]." Figure 5 shows the general layout for a SCADA system.

The main components related to this research are the Human Machine Interface

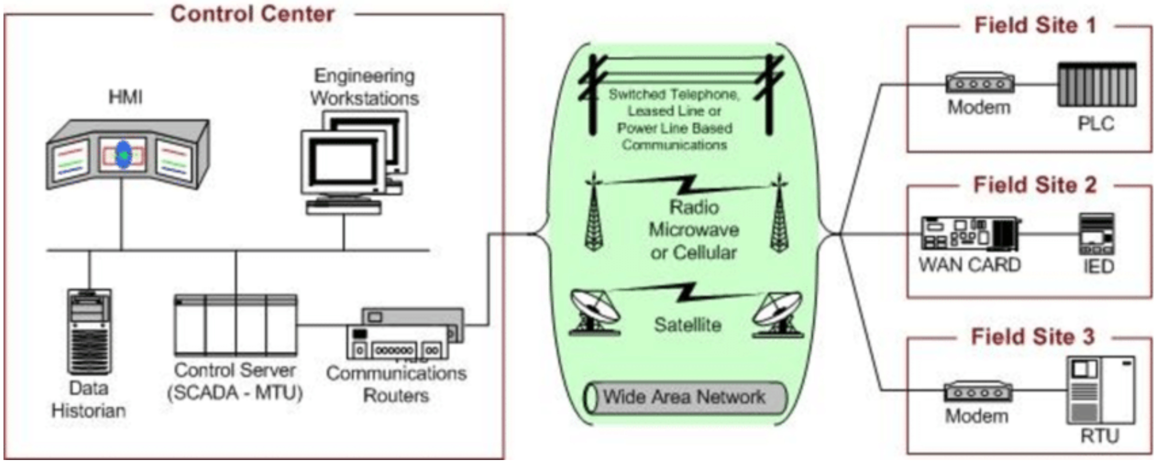


Figure 5. SCADA System General Layout [3]

(HMI) and the end devices that the operator controls through the HMI. Attacks like Stuxnet that target ICS will sometime try to hide the attack from the operator by changing the values that the operator sees on the HMI. This can make the operator think that everything is running normally when in fact the attack is underway and could be causing major damage. This research looks at the ability to measure the physical effects of the controlled devices to see if it is possible to map that to the value seen by the ICS operator. Essentially, the goal is to verify that what the operator is seeing on the HMI is what is actually happening, and if it is not, be able to alert the operator to possible anomalous behavior.

2.3 Stuxnet

Stuxnet is one of the main motivations for this research. Stuxnet was a targeted attack aimed at ICS, specifically the spin rate of centrifuges at a uranium enrichment facility. The W32.Stuxnet Dossier [4] from Symantec gives a detailed analysis of Stuxnet. Stuxnet had four main goals:

1. Replicate and spread on its own.

2. Discover the proper computers that it was targeting.
3. Disrupt the normal operation of the centrifuges.
4. Evade detection.

While the first two are very impressive in how it was executed, this research is only concerned about the last two. One of the ways in which Stuxnet disrupted the normal operation of the centrifuges is by following the sequence listed below [4].

- Examine records from frequency converter slaves for approximately 13 days to ensure that the system has been operating between 807 Hz and 1210 Hz. Which is the normal operating frequency.
- Set the frequency 1410 Hz.
- Return to normal operation.
- Wait for 27 days.
- Set the frequency to 2 Hz and then to 1064 Hz.
- Return to normal operation.
- Wait for 27 days.
- Return to setting the frequency to 1410 Hz and repeat the process.

Stuxnet hides this activity from the operator by recording normal activity and replaying those recordings to the operator during its attacks.

This research aims to provide a proof of concept for an ADS that could have detected an attack such as Stuxnet by detecting the physical changes in vibration between the normal frequency that an HMI is showing to an operator and the actual frequency that the centrifuge is operating at. This research uses a centrifugal pump and a flow meter from an ICS training system as a stand in for the centrifuges.

2.4 Related Work

2.4.1 RF-DNA.

This research has adopted established tools and techniques from previous RF-DNA research efforts. The main idea behind RF-DNA is to capture, analyze, and quantify the variance in Radio Frequency (RF) emissions as it relates to the variance in between different devices or device states and generally involves the following steps [6]:

1. Signal Collection
2. Burst Detection
3. Feature Extraction
4. RF Fingerprint Generation
5. Device Classification

For this research, signal collection will not be a RF signal, instead it will be vibration data captured at different operational states from a pump and flow meter on an ICS training system. Some of the prior work have used Wired Signal Distinct Native Attribute (WS-DNA) techniques for data collection instead of RF [7]. That research used wired communication signals for its analysis. This research does not rely on communication signals or specific parts of a signal and therefore burst detection will not be used. Feature extraction is done by performing statistical calculation against the collected data. Fingerprints are generated as a result of the feature calculations and are used essentially to identify the device or operational state of interest. Classification uses part of the generated fingerprints to train the classifier while the other part is used for blind testing with the goal of training the system to know what a given device or state looks like. Testing data will be classified as belonging to a certain device or state based on which it matches to the best. This research follows techniques and processes from prior RF-DNA work [6, 7, 8, 9, 10, 11, 12, 13, 14, 15].

2.4.2 Machine Condition Monitoring.

Machine condition monitoring is regularly used in the industrial world. The idea is to be able to detect a failure or imminent failure of a machine to prevent further possible damage or indicate when a part may need replaced. This research uses machine condition monitoring techniques to measure the vibration from a pump and flow meter on an ICS training system. However, instead of detecting machine failure, this research aims to determine if the vibration between different operating states can be detected and used in a potential ADS. This research follows techniques used in previous studies for monitoring electric motors through vibration sensors over wireless networks [16, 17, 18].

2.5 Electric Motors

The pump used on the Lab-Volt® training system is a Marathon Electric® three phase electric induction motor with a centrifugal pump attached. Figure 6 shows a breakout of a typical three-phase electric motor. This section will cover the parts of the electric motor and centrifugal pump and how they work.

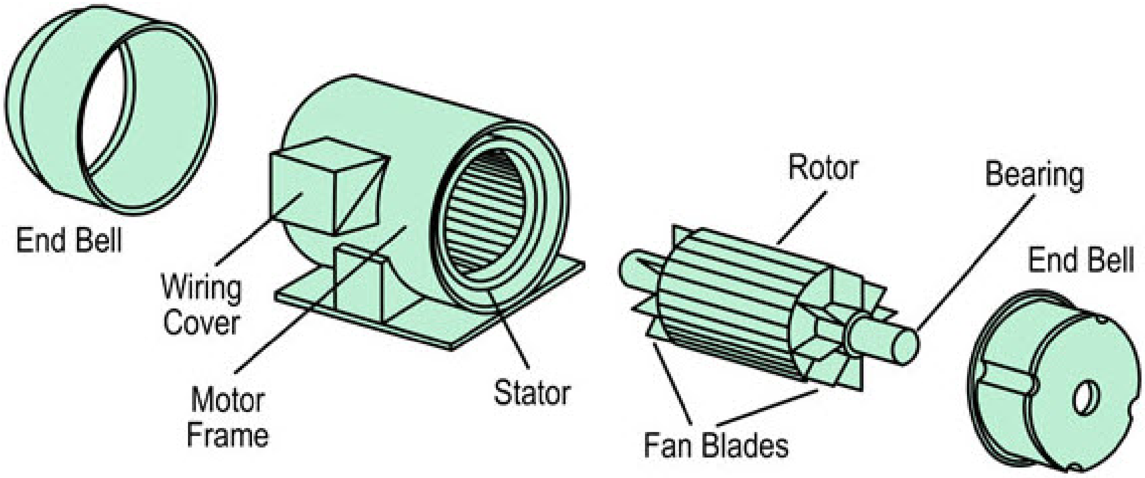


Figure 6. Three-Phase Induction Motor [19]

2.5.1 Stator.

The stator is the stationary part of the motor housed inside the frame. The stator contains the coils used to generate an electromagnetic field. The coils are placed with half of the coil on one side and the other half on the opposite side. When AC power is supplied to the coil, the magnetic field generated will alternate sides due to sinusoidal nature of AC, meaning that magnetic north with switch sides of the coil with the AC wave [19, 20].

2.5.2 Three-Phase Power.

Three-Phase AC is used in order create a rotating magnetic field which will cause the rotor to spin and ultimately drive the pump. Each phase is attached to one of the coils in the stator and then phase shifted 120° to produce rotation. When three-phase AC is applied, the different phases cause the magnetic north to travel in a circular motion and generates a rotating magnetic field. Figure 7 illustrates the the sine wave for three-phase AC shifted 120°.

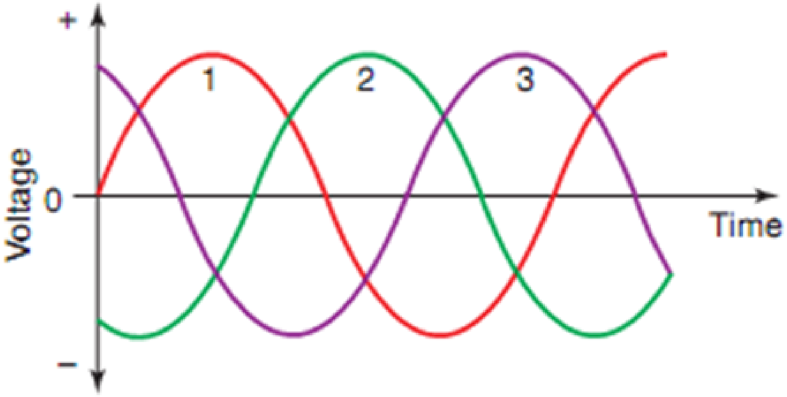


Figure 7. Three-Phase AC Sine Waves - Shifted 120°[21]

2.5.3 Rotor.

The rotor is the part that physically spins in the motor. The speed at which it spins is determined based on the frequency of the AC power that is supplied to the coils, thus controlling the speed of the rotating magnetic field. For the Lab-Volt® training system, the frequency can range from 0 Hz (off) to 60 Hz (full power). The rotor itself is made up of a shaft with rings and bearings at either end with conductor bars slotted between the rings [22].

2.5.4 Centrifugal Pump.

The centrifugal pump attaches to the shaft from the motor in order to rotate the impeller. The middle part of the impeller, called the eye, has the inlet from the water reservoir tanks attached. When the motor is powered, the drive shaft rotates the impeller and spins the water from the inlet. Due to centrifugal force, the water is pushed to the outer sides of the pump where it exits through the outlet [23]. The whole process creates a suction of water from the inlet to the outlet with the amount of flow being determined by speed of the impeller. Figure 8 displays the parts of a centrifugal pump.

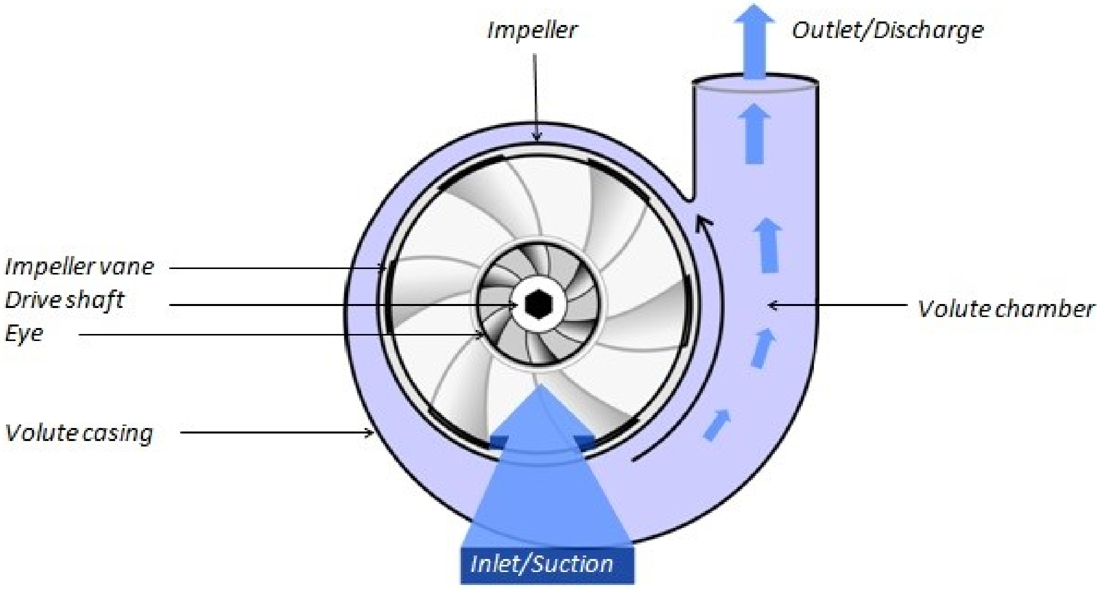


Figure 8. Centrifugal Pump [23]

2.6 Piezoelectric Effect

The piezoelectric effect describes how certain materials, such as quartz crystals, will naturally generate an electric charge when subjected to mechanical stress. The piezoelectric effect was first discovered by two brothers, French scientists Jacques and Pierre Curie, in 1880 by realizing that applying pressure to quartz and other types of crystals would create an electrical charge in the material. The first major application of the piezoelectric effect came during the first World War where the effect was used in sonar technology [5, 24]. Today, the piezoelectric effect can be found in a wide range of applications. It is used in cell phones, as acoustic guitar pickups, as vibration sensors, in grill igniters, and even in musical greeting cards to name a few [25]. Figure 9 shows how compressing a piezoelectric material creates an electric charge.

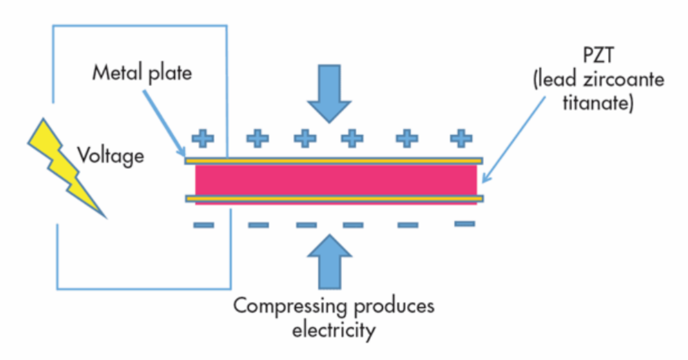


Figure 9. Piezoelectric Effect [24]

The piezoelectric effect also works in the opposite direction, known as the inverse piezoelectric effect. In this case you can feed an electric charge to a piezoelectric material and cause a mechanical stress on the material. The mechanical stress on the material can then produce acoustic sound waves. This effect is often used to turn a piezoelectric device into a cheap speaker like what can be found in musical greeting cards. Figure 10 shows the inverse effect.

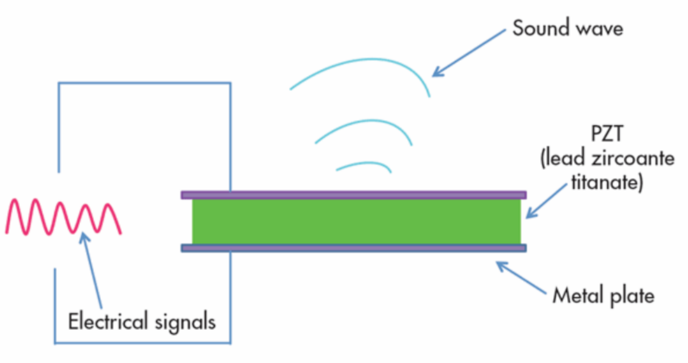


Figure 10. Inverse Piezoelectric Effect [24]

III. Methodology

3.1 Introduction

This chapter covers the methodology used for this research. Section 3.2 describes the problem definition and restates the goals of the research. Section 3.3 covers the RF-DNA methodology and process used for this research. Section 3.4 gives the overall experimental design used for the research. Section 3.5 describes the implementation for the hardware, software, and process used during data collection. Section 3.6 describes the process for preparing the data for the analysis tool.

3.2 Problem Definition

The intent of this research is to use RF-DNA fingerprinting and classification techniques to provide a proof of concept for an ADS that is capable of detecting unauthorized changes to the operation of ICS devices based on the change in the physical characteristics caused by the unauthorized change. Specifically, the goals are to:

1. Determine if the physical effects of devices in an ICS can be measured, fingerprinted, and classified with enough detail to be able to differentiate between the different levels of operation.

2. Determine the viability of an ADS based on the results of the data fingerprinting and classification.

3.3 Radio Frequency Distingct Native Attribute (RF-DNA)

This research takes advantage of well established RF-DNA tools and techniques [6, 7, 8, 9, 10, 11, 12, 13, 14, 15] in order to distinguish between different levels of operation for ICS devices. The results of this classification will determine the viability of an ADS based on the physical effects given off by the device. The basic process for RF-DNA used in this research is listed below. Each of these are discussed in more detail in the following sections.

1. Data Collection
2. Statistical Fingerprinting
3. Training and Testing
4. Classification

3.3.1 Data Collection.

Vibration measurements from a centrifugal pump, powered by a three-phase induction motor, and a flow meter are used for this research. A piezoelectric disc is used for analog vibration detection with digital values generated and logged through an ADC on an Arduino Uno.

Vibration measurements are stored into classes, N_c, based on the power level that the pump is running at during collection. Sixty classes, covering the entire operating range, are collected from the pump while only fifty-one classes are collected from the flow meter because water does not flow through the system until the pump reaches approximately power level ten. Each class contains $N_s = 600,000$ collected samples as shown in Figure 11.

Following the collection of each N_c, the samples are segmented into $N_b = 200$ bursts with $N_s = 3000$ samples each as shown in Figure 12. Previous research efforts [12, 13] focused bursts around a specific section of collected signal such as a preamble section from a wireless signal. This research assumes all data collected in a given N_c is relatively consistent across all collected signal due to the consistency of the physical

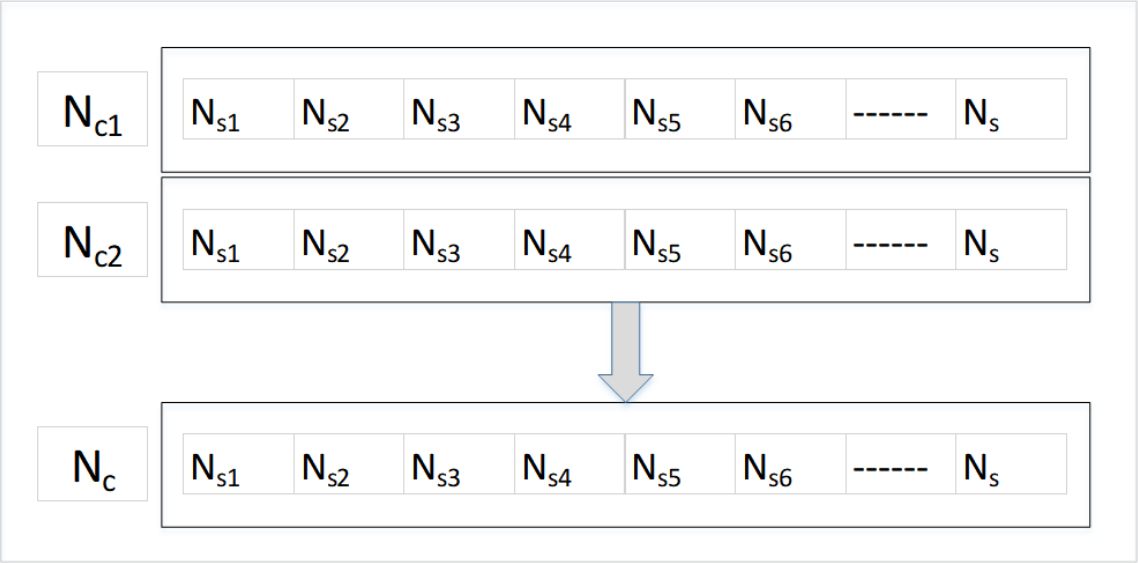


Figure 11. Data Collection for N_c Classes With N_s Samples

vibration on the device being measured, therefore, bursts are subsets of all N_s taken in consecutive sequences.

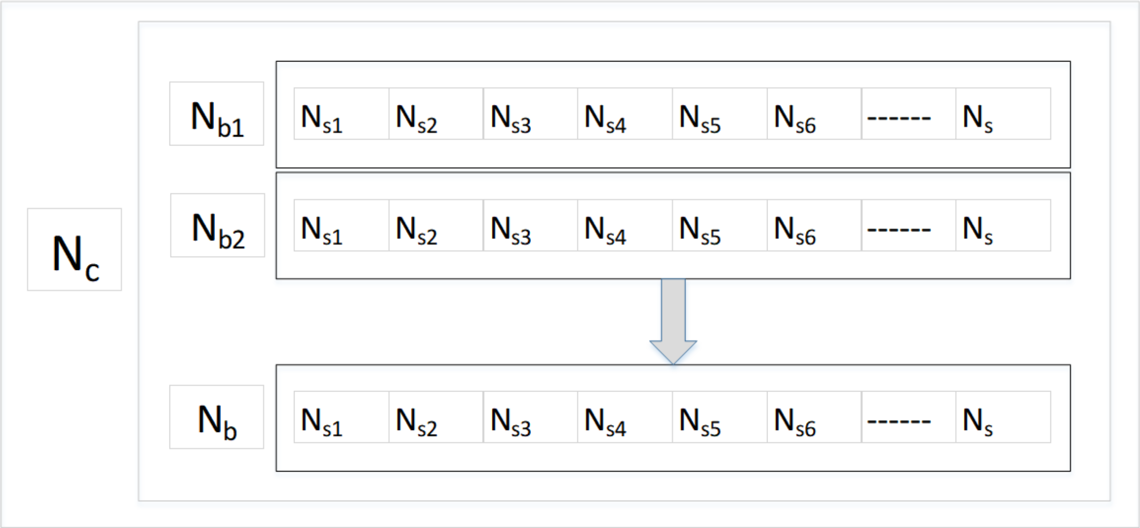


Figure 12. Example N_c Class With N_s Samples Segmented Into N_b Bursts

3.3.2 Fingerprinting.

Statistical RF-DNA fingerprints (f) are created in the SD (f_{SD}) using methods consistant with [12, 13]. First, a ROI is selected from an arbitrary N_b in the Time Domain (TD). For this research, the ROI is the entire collected signal given the assumption from Section 3.3.1 about the consistency of collected data in a given N_c, see Figure 13.

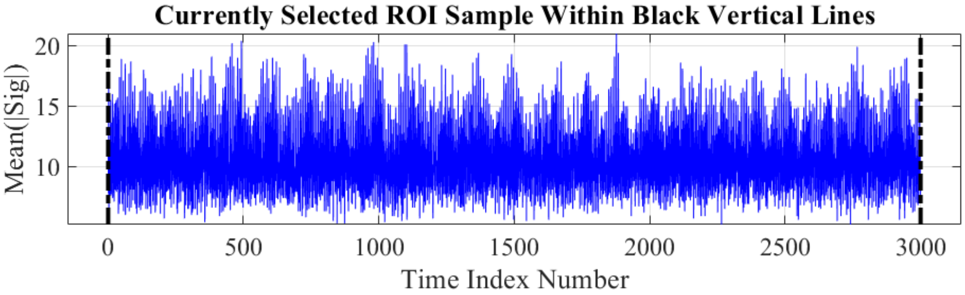


Figure 13. Selected ROI From Arbitrary N_b

A power normalized Power Spectral Density (PSD) feature sequence ($\{\bar{p}(k)\}$) is created across the selected ROI by performing a Discrete Fourier Transform (DFT) and dividing by the signals average power. Normalization is used to minimize the potential effects of collection that could bias classification results. $N_R = 26$ equal length subregions are selected from the ROI for feature generation. Features are created by performing statistical calculations for standard deviation (σ), variance (σ^2), skewness (γ), and kurtosis (κ) for each N_R and for the entire region ($N_R + 1$) shown by Figure 14. The statistics for each subregion are arranged as:

$$f_{R_i} = [\sigma_{R_i}, \sigma^2_{R_i}, \gamma_{R_i}, \kappa_{R_i}]_{1\times 4} \tag{1}$$

where $i = 1, 2, ..., N_R + 1$. Full fingerprints (f_{SD}) are formed by concatenating statistics given by:

$$f_{SD} = [f_{R_1} \vdots f_{R_2}, \vdots f_{R_3}, \cdots f_{R_{N_R+1}}]_{1\times 4(N_R+1)} \tag{2}$$

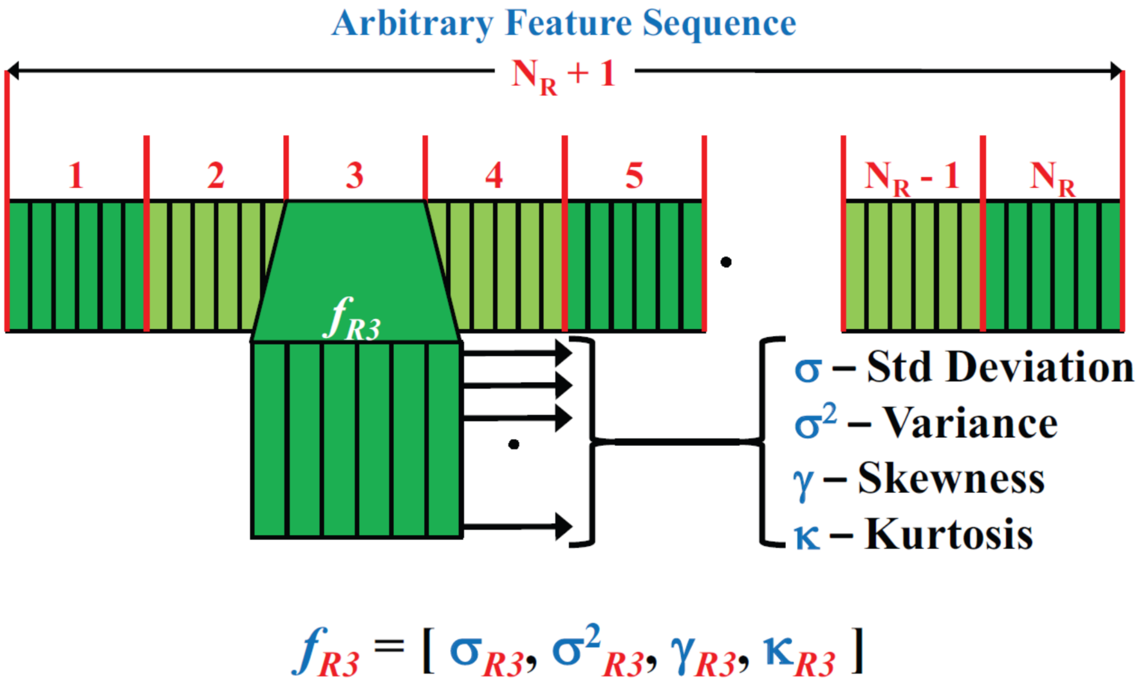


Figure 14. Regional Fingerprint Generation for $N_R + 1$ Total Regions [12, 13]

Multiple sets of fingerprints with their given features can be generated with artificial noise added to get to a desired analysis Signal-to-Noise Ratio (SNR). This can be useful when trying to classify RF signals due to the inherent nature of wireless signals. This research's focus is on the collected signal itself without any added noise profiles and therefore uses a high dB SNR to represent the collected signal.

3.3.3 Training and Testing.

Fingerprints with their given features as discussed in Section 3.3.2 are input into a Multiple Discriminant Analysis/Maximum Likelihood (MDA/ML) classifier. This research follows the methods laid out in [6, 12, 13, 14, 7] and are described at a high level here. MDA is an extension to Fisher's Linear Discriminant (FLD) for when more than two classes are being considered. Fingerprint data is first split into two separate data sets, one for training and one for testing, chosen in an interleaving pattern from the bursts in each class, see Figure 15.

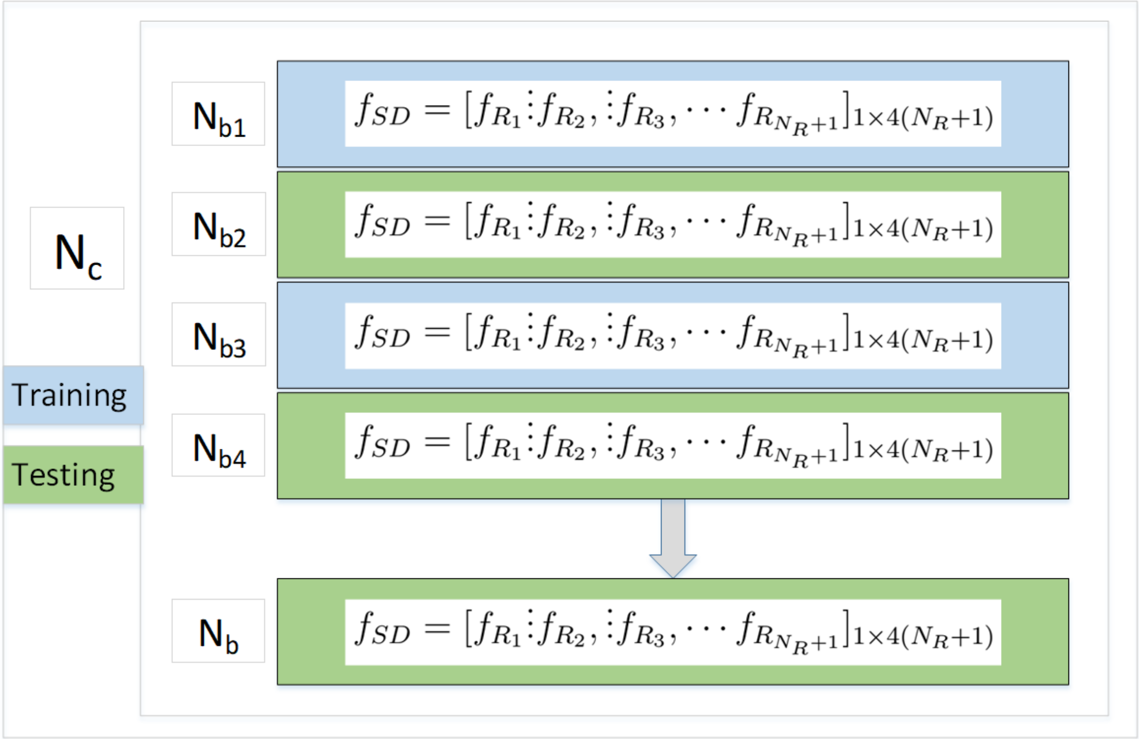

N_c
Training
Testing
N_{b1}
N_{b2}
N_{b3}
N_{b4}
N_b
$f_{SD} = [f_{R_1} \vdots f_{R_2}, \vdots f_{R_3}, \cdots f_{R_{N_R+1}}]_{1\times 4(N_R+1)}$

Figure 15. Training and Testing Data Sets Chosen In An Interleaving Pattern

The goal of MDA is to create a projection matrix that maximizes inter-class distances and minimizes intra-class spread, Figure 16 shows this visually with three classes. The first set of data is used for supervised training while the second set is used for "blind" testing. Supervised training means that the data is known data. Basically, the system is being trained by telling it which class the training data belongs to and generating a projection matrix specified by W_n that finds the "best" W_n that separates the classes. Testing data is input into the system as unknown data, and using the Bayesian posterior probability method it is assigned to a certain class based on a best match determination.

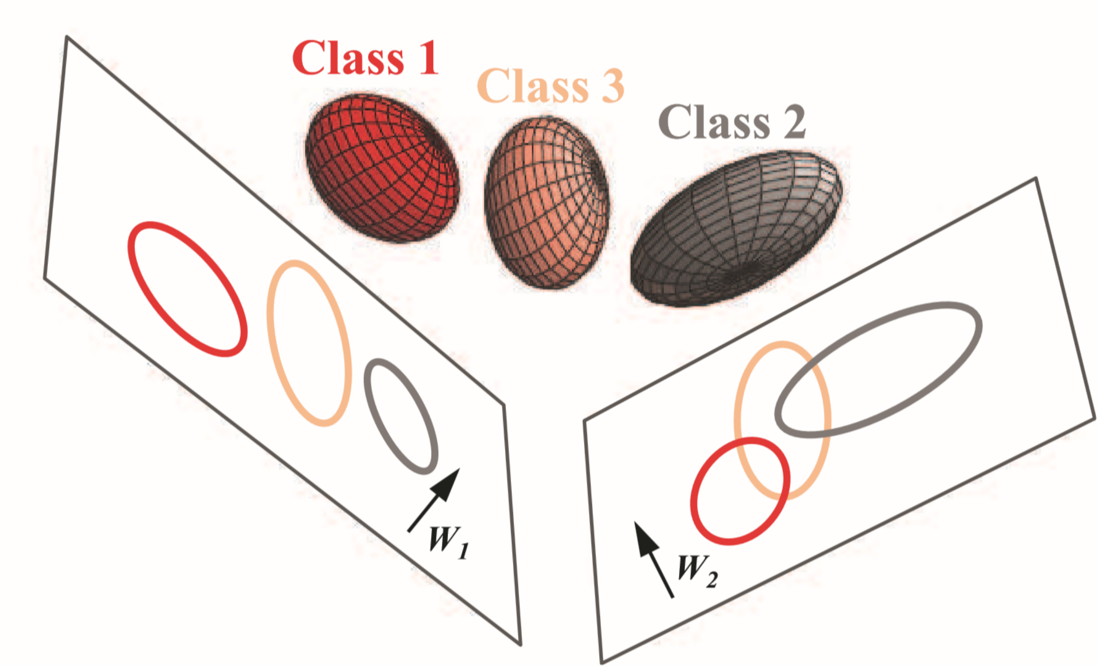


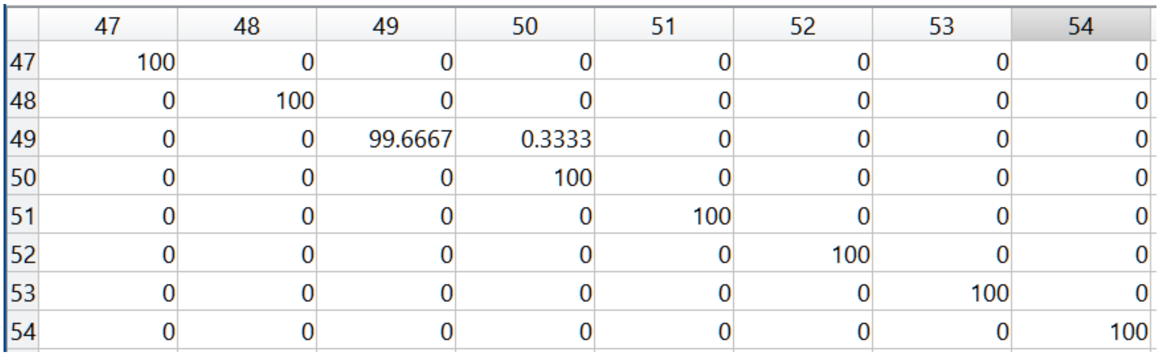
Figure 16. Representative MDA Projection For Three Classes [12]

3.3.4 Classification.

The output for the MDA/ML classifier is a confusion matrix with the percentages of correctness for each class listed for the class that its testing data matched to. Figure 17 displays a partial confusion matrix that shows classes 47 through 54. The y axis represents the class that is being tested while the x axis represents the class that was best matched to. All of the classes in this example match correctly to its own class except for class 49, which matched correctly 99.6667% to its own class and 0.3333% to class 50. The information contained in the confusion matrix is used to answer the questions posed in the research goals.

3.4 Experimental Design

The system being tested in this research is the process for collecting vibration data from ICS devices at different operating levels and using RF-DNA techniques for



	47	48	49	50	51	52	53	54
47	100	0	0	0	0	0	0	0
48	0	100	0	0	0	0	0	0
49	0	0	99.6667	0.3333	0	0	0	0
50	0	0	0	100	0	0	0	0
51	0	0	0	0	100	0	0	0
52	0	0	0	0	0	100	0	0
53	0	0	0	0	0	0	100	0
54	0	0	0	0	0	0	0	100

Figure 17. Partial Confusion Matrix

fingerprinting and classifying the data with the end result being the output from the classification. Figure 18 displays the system under test.

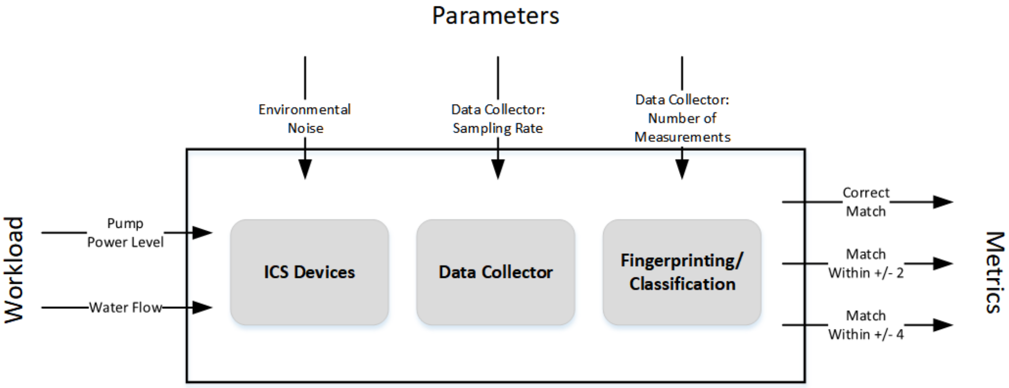


Figure 18. System Under Test

3.4.1 System Components.

The components of the system are made up the ICS devices on the training system, the data collector, and the fingerprinting and classification. The ICS devices specifically are the pump and flow meter where vibration data will be collected from. The data collector is the collection of devices that will be used for capturing the vibration data like the piezoelectric sensor, the Arduino Uno, and the Surface Pro 4. Finally, the fingerprinting and classification is made up of the RF-DNA tools and techniques that will be performed inside the MATLAB® environment.

3.4.2 Workload.

The workload parameters of the system consist of the operating levels of the pump and the amount of water flowing through the flow meter at a given pump operating level.

The operating levels for the pump are set on the HMI of the ICS training system. The levels are a range from 0 Hz (off) to 60 Hz (full power) based on the frequency of the AC power being supplied to it. As described in Chapter II, the AC frequency determines the speed of the rotation in the electric motor and ultimately of the centrifugal pump at the end of the motor. The speed of the rotation also affects the physical vibration emanating from the pump. The faster the rotation the more intense the vibration

The flow rate of the water running through the flow meter is directly related to the speed of the pump. The faster the pump is running, the faster the water is forced through the flow meter. The amount and speed of the water moving through the flow meter affects the vibration intensity that can be detected by the vibration sensor on the flow meter.

3.4.3 System Parameters.

The system parameters consist of the environmental noise, the sampling rate for the data collection, and the number of measurements taken from the vibration sensor.

The environmental noise is the unintended signal collected by the vibration sensor that did not come from the source that is trying to be measured. The noise can come from a number of sources like electrical interference, the air conditioner running in the lab, or even people walking around causing the floor to vibrate.

The sampling rate is the rate (in Hz) at which measurements from the vibration sensor are to be taken. A sampling rate of 2807 Hz was chosen for this research because the exact vibration frequencies of interest that are given off at the pump and flow meter are unknown, thus a high sample rate is warranted. The sample rate is also low enough that the serial connection to the Surface Pro 4 that is used for logging the data does not get overwhelmed and cause delays and missed data collection. The specific sample rate was calculated after finding the balance between these two concerns.

The number of measurements taken is how many readings are taken from the vibration sensor for a given operating level that the pump is running at. For this research there are 600,000 measurements taken at each level. At the given sample rate, one collection takes approximately 214 seconds. 600,000 was chosen for this research because it provides a relatively large representative profile for the vibration at that level and because the raw data collected reshapes well into a 200x3000 structured format used for analysis.

3.4.4 Performance Metrics.

The performance metrics are determined from the confusion matrix. The confusion matrix shows how the testing data was matched to the training data and the percentage of the match. The classes in a confusion matrix are mapped to the operating levels of the pump in Hz. Using this information the performance metrics are:

- The percentage that a class is correctly classified to its own class
- The percentage that a class is classified within +/- 2 classes
- The percentage that a class is classified within +/- 4 classes

The desired results for an ADS is that each class matches to itself. However, it is also good for an ADS to know if it was close to an exact match. Therefore, a range for +/- 2, and +/- 4 classes is used. These results can then be used to answer the questions posed by the research goals.

3.5 Data Collection Hardware

This section covers the hardware used during the implementation of data collection.

3.5.1 Arduino Uno.

An Arduino Uno revision 2 is used as the platform for collecting data. The Arduino was chosen because it has a built in ADC that will allow input from the analog sensor to be converted to digital values for later analysis. The main specifications for the Arduino are [26]:

- ATMEGA4809 Microcontroller
- 5V operating voltage
- 14 digital I/O pins
- 6 analog input pins
- 16 MHz clock speed

The Arduino Uno has a built in 10-bit ADC. The ADC allows for the analog input from the piezoelectric sensor to be converted to a digital value. Since the ADC is 10-bit, it can detect 1,024 (2^{10}) discrete analog levels, which can also be called the ADC's resolution [27]. The ADC uses the following equation for the conversion from analog voltage to digital value:

$$\frac{1023}{5} = \frac{ADC\ Reading}{Analog\ Voltage\ Measured}$$

Figure 19. ADC Calculation [27]

The left side of the equation shown in Figure 19 represents the resolution of the ADC over the system voltage (5V for the Arduino Uno). The right side shows the ADC reading, or value to be assigned, over the analog voltage measure, or the voltage being supplied by the piezoelectric vibration sensor. Using this equation, the ADC is able to assign a digital value between 0 and 1023 based on the supplied voltage from the sensor.

3.5.2 Piezoelectric Sensor.

A piezoelectric disc was chosen for the vibration sensor. Using the piezoelectric effect as described in Chapter II, this sensor is capable of detecting vibration intensity and translate it into an electric charge. The disc used has the following specifications:

- Plate material: Brass
- Resonant frequency (kHz): 4.6 +/- 0.5 KHz
- Resonant impedance (ohms): 300 max
- Capacitance (nF): 20.0 +/- 30
- Diameter plate: approx.2.7cm / 1.06 inch
- Lead Length: approx.33cm / 12.9 inch

3.5.3 Custom Arduino Shield.

During initial testing with the vibration sensor, it was found that the signal needed to be amplified prior to digitization. A custom Arduino shield was designed to solve this problem. Arduino shields are modular circuit boards that install on top of the Arduino in order to extend its functionality. In this case, the shield has been designed to amplify the signal by a factor of up to six, thus, increasing the sensitivity of the sensor. The practical effect of this boost is to allow the vibration sensor to be more able to report the differences between vibrations that are close to each other in intensity, essentially increasing the resolution of the Arduino's ADC. For example, detecting the difference in vibration on the pump between level four and level five.

The shield uses an operational amplifier, op-amp for short, in a non-inverting configuration. Figure 20 shows a standard non-inverting op-amp with the equation used to calculate the output voltage based on the input voltage. The input voltage comes from the vibration sensor and passes through a diode and resistor to limit the voltage sent to the op-amp to avoid damaging it.

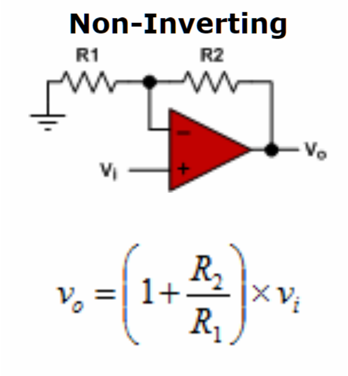


Figure 20. Non-Inverting Operational Amplifier with Equation [28]

Using this equation, R_2 is the feedback resistor, and R_1 goes to ground. Figure 21 shows the schematic for the custom Arduino sheild where R_2 is fixed at $10K\Omega$ and R_1 is a combination of up to five $10K\Omega$ resistors running in parallel (switched on or off by transistors that connect them to ground when activated). The combination resistance of R_1 is:

$$R_1 = \frac{1}{\sum_n \frac{1}{10K\Omega}} = \frac{1}{n \times \frac{1}{10K\Omega}} = \frac{10K\Omega}{n} \tag{3}$$

thus,

$$v_o = (1 + \frac{R_2}{R_1})v_i = (1 + (10K\Omega)\frac{n}{10K\Omega})v_i = (1 + n)v_i \qquad (4)$$

where n is the number of resistors that are activated. After testing the custom shield with different amplification settings, it was decided to keep all five resistors activate for the data collection in this research. Thus $n = 5$, and $v_o = 6v_i$.

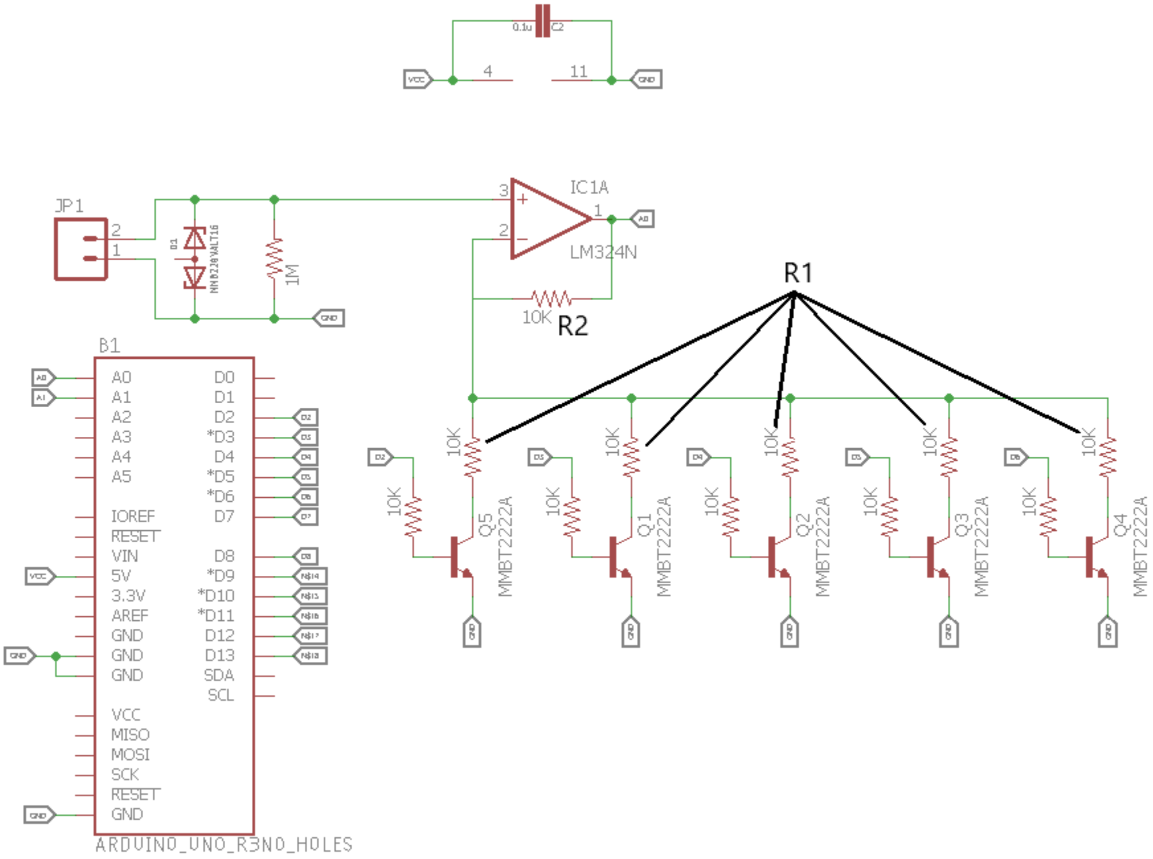


Figure 21. Schematic for Custom Arduino Shield

3.5.4 Surface Pro 4.

A Microsoft Surface Pro 4 is used to collect the data from and power the Arduino Uno through a USB port. The Surface Pro was chosen for its portability and contains the following specifications:

- Windows 10 Professional version 1803 64-Bit
- Intel Core i7-6650U CPU @ 2.20GHz
- 16 GB Memory

In addition, the two main applications used are the Arduino Integrated Development Environment (IDE) for writing and uploading the code for the Arduino Uno and Putty, which is used to log the ADC readings to a file.

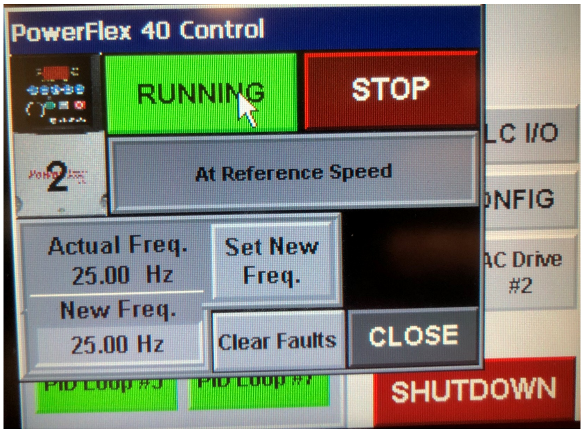
3.5.5 Arduino Software.

The software running on the Arduino Uno is designed specifically for collecting vibration data from the analog signal produced by the piezoelectric disc. The software can tune the amplification provided by the custom Ardunio shield by taking a combination of five digital pins high or low depending on the desired level of amplification. The main function of the software is to provide a simple interface to the user over the serial connection and to loop through the desired number sensor readings at a given sampling rate.

Due to the built in delay functions for Arduino being an unknown and not trustworthy, a custom delay was written to provide more precise and known timing between sensor readings. The delay works by setting an exact amount of time, in microseconds, that each measurement in the loop will take before proceeding. Ensuring that all measurement take the same amount of time.

3.5.6 Data Collection Process.

The data collection process is fairly straightforward once all of the device and sensor setup is completed. First the piezoelectric disc is attached to a flat surface on the pump or flow meter with a velcro strap to keep the sensor in place. The pump is then set at the desired operating level (see Figure 22). A serial connection is made to the Arduino Uno using Putty on the Surface Pro 4 and a key is pressed to start the data capture for that power level. When this is finished (takes approximately three minutes and thirty-four seconds) the Putty log is saved and a new session is started for the next operating level on the pump.



At Reference Speed

Actual Freq. 25.00 Hz

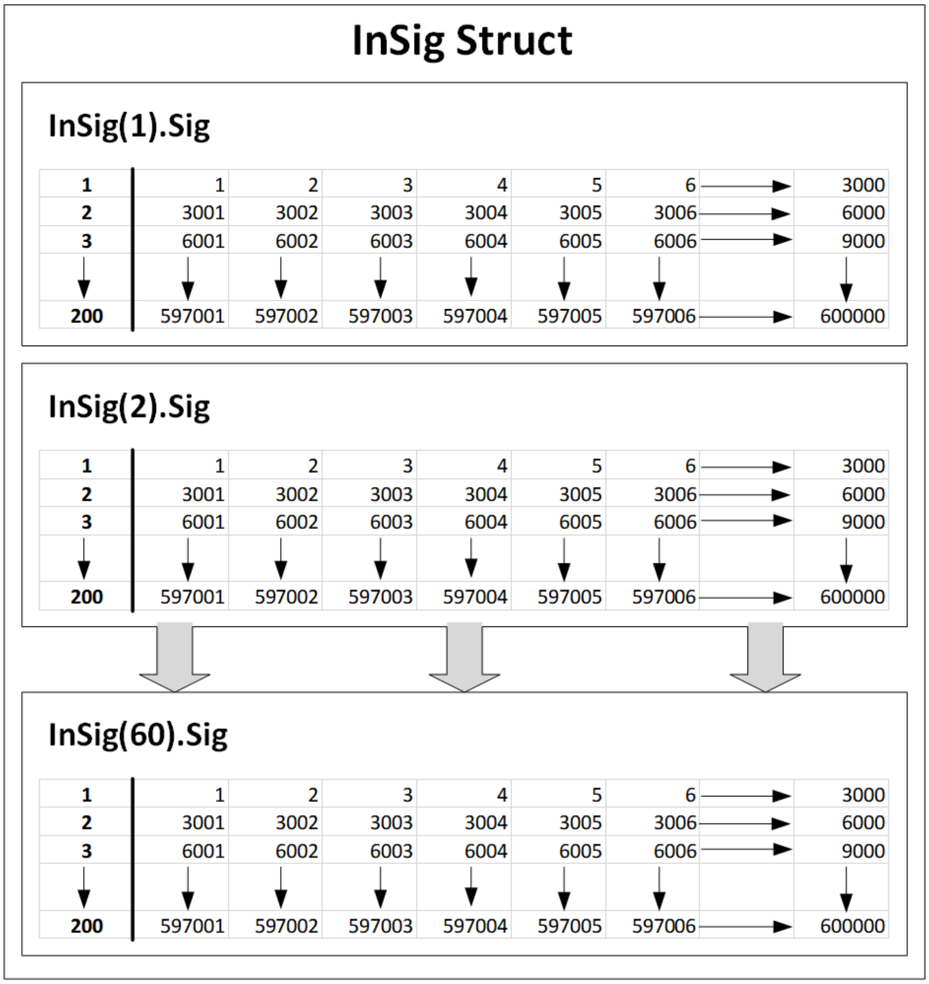
New Freq. 25.00 Hz

Figure 22. HMI Interface for Controlling Pump

3.6 Data Preparation

After the data is collected, it must be rearranged into a specific format for analysis. At this point the data is just a series of log files that contain 600,000 sensor readings with text at the top and bottom from user interaction during the collection process. RF-DNA fingerprinting in MATLAB® expects the data to be stored in a structure (known as a struct in MATLAB®) named InSig with a specific format as shown in Figure 23.

This struct stores all of the sensor data for all operational levels of the pump (called classes during analysis) into individual fields. The data in each field is reshaped to be double-precision floating-point values of size 200x3000 for the 600,000 sensor readings. The size means that there are two hundred rows (bursts) each with three thousand



InSig(1).Sig

1	1	2	3	4	5	6	→	3000
2	3001	3002	3003	3004	3005	3006	→	6000
3	6001	6002	6003	6004	6005	6006	→	9000
↓	↓	↓	↓	↓	↓	↓		↓
200	597001	597002	597003	597004	597005	597006	→	600000

InSig(2).Sig

1	1	2	3	4	5	6	→	3000
2	3001	3002	3003	3004	3005	3006	→	6000
3	6001	6002	6003	6004	6005	6006	→	9000
↓	↓	↓	↓	↓	↓	↓		↓
200	597001	597002	597003	597004	597005	597006	→	600000

InSig(60).Sig

1	1	2	3	4	5	6	→	3000
2	3001	3002	3003	3004	3005	3006	→	6000
3	6001	6002	6003	6004	6005	6006	→	9000
↓	↓	↓	↓	↓	↓	↓		↓
200	597001	597002	597003	597004	597005	597006	→	600000

Figure 23. Data Format For InSig Struct

samples from data collection.

Getting the data into the correct format starts by stripping the text from the top and bottom of each log file and changing the extension of the file to be a comma-separated values (CSV) file. The data is then read into MATLAB® using the following code:

```
clear all;
close all;

```

```
idx2 = 1;

for idx = 1:60
    % set filenames
    filename = sprintf('d%d.csv',idx);
    filename2 = sprintf('d%d-1.csv',idx);
    filename3 = sprintf('d%d-2.csv',idx);

    % ensure at least the first file exists
    if isfile(filename)
        temp = csvread(filename);
        temp2 = csvread(filename2);
        temp3 = csvread(filename3);

        % append each set of data to each other
        temp = [temp, temp2, temp3];

        % temp storage struct before data is reshaped
        test(idx).Sig = temp;

        % temp storage struct after data is reshaped
        rtest(idx).Sig = reshape(temp,3000,[]).';

        % Assign reshaped to data to MySig only where there is real data
        MySig(idx2).Sig = rtest(idx).Sig;
        idx2 = idx2 + 1;
    end
```

```
end
```

This code results in a struct called MySig that is in the correct format for the fingerprinting tool. MySig can then be copied and renamed to InSig for fingerprinting while maintaining the original MySig file for reference and archive.

IV. Analysis and Results

4.1 Introduction

This chapter presents the data analysis implementation and results for vibration data collected from the pump and flow meter on the Lab-Volt® ICS training system. Section 4.2 covers the analysis implementation and result details for data collected from the pump. Section 4.3 covers the analysis implementation and result details for data collected from the flow meter. Section 4.4 provides a summary of the chapter.

4.2 Analysis - Pump

This section describes the analysis implementation for vibration data collected from the centrifugal pump on the Lab-Volt® ICS training system. The analysis follows the overall process as described in Section 3.3.

4.2.1 Collected Data.

Three complete sets of vibration data were captured during the course of this research. Each set containing all $N_c = 60$ classes with $N_s = 600000$ samples collected per class as described in Section 3.3.1. The three data sets were appended together in order to take advantage of all collected data for analysis. The resulting analysis data contains $N_c = 60$ classes with each N_c containing $N_b = 600$ bursts and each N_b containing $N_s = 3000$ samples. In total, $N_s = 108$ Million samples were collected from the pump and used for analysis.

4.2.2 Fingerprinting.

Fingerprinting of the collected pump data is done according to the process laid out in Section 3.3.2. An overview of that process using collected data from the pump is presented here.

1. A ROI is selected from an arbitrary N_b burst in the TD. This research uses the entire $N_s = 3000$ range for the given burst. Refer back to Figure 13.

2. $N_R = 26$ subregions are selected from the power normalized PSD in the selected ROI as shown in Figure 24.

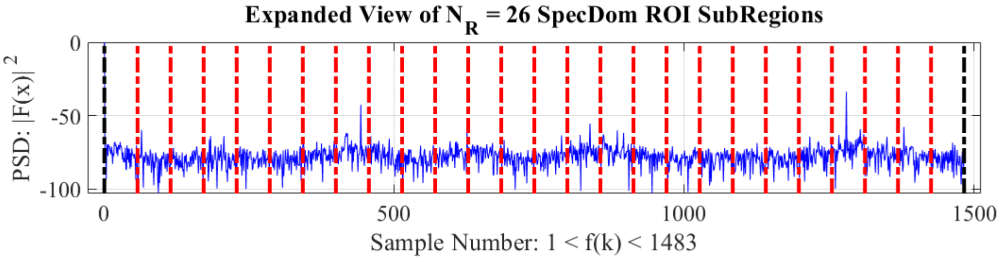


Figure 24. Subregion Selection From PSD

3. Features are created for each N_R+1 subregion plus the entire burst by performing statistical calculations for standard deviation (σ), variance (σ^2), skewness (γ), and kurtosis (κ). The number of features is calculated as follows:

$$Features = (N_R + 1) \times statistics \tag{5}$$

$$Features = (26 + 1) \times 4 \tag{6}$$

$$Features = 108 \tag{7}$$

4. Full fingerprints (f_{SD}) are created at desired analysis SNR's by concatenating features together. The collected signal without added noise is used for this research by using a high SNR.

4.2.3 Classification Results.

The MDA/ML classification process takes place in two steps, training and testing. The process for training and testing is outlined in Section 3.3.3 with a brief overview given here.

1. Fingerprint (f_{SD}) data is split into two data sets, training and testing. The split is performed by interleaving through the bursts (N_b) from each class (N_c).
2. Supervised training is performed on the training data set to generate a projection matrix.
3. "Blind" testing is performed with the testing data and classified to a given class based on a best match determination.

Classification results are generated in the form of a confusion matrix. Table 1 presents the data from the confusion matrix by the percentage that a testing class was correctly matched to the same training class, the percent that it was correctly matched +/- 2 classes, and the percent that it was correctly matched +/- 4 classes. The results are presented this way because one, it is easier to digest than a 60 class confusion matrix, and two, knowing how close a class comes to an exact match in a given range is useful for an ADS. Several observations can be made by looking at the results given in Table 1:

1. From class 19 and higher the percentage for an exact match never drops below 99% with most in that range matching at 100%. This indicates that as the pump runs faster it becomes the dominant source of vibration, thus making it easier to correctly classify. Alternatively, below class 19 the vibration is not strong enough to be completely distinguishable from the environmental noise in the signal.

2. The lowest percentage for an exact match is 61.3% at class 11, however, that percentage jumps to 98% for matching +/- 2 classes. This indicates that even when it is not an exact match, it is matched with the classes nearby.

3. Only two classes fall below 98% in the +/- 4 range. Class 5 at 95.3% and class 10 at 90.7%.

Class N_c	**Match**	**+/- 2**	**+/- 4**
1	85.7%	94%	98.7%
2	73.7%	94.7%	100%
3	70%	96.3%	99%
4	69%	98.3%	99.3%
5	87.7%	95%	95.3%
6	68.3%	95.3%	99.7%
7	71%	97.3%	100%
8	79%	94.3%	98.7%
9	71%	85.7%	98.3%
10	82%	89.3%	90.7%
11	61.3%	98%	98.7%
12	67.7%	95.3%	99%
13	69.3%	92.3%	99.3%
14	80.7%	100%	100%
15	97.3%	99.3%	99.3%
16	94%	99.7%	100%
17	92.3%	100%	100%
18	94.3%	100%	100%
19	99.7%	100%	100%
20	100%	100%	100%
21	100%	100%	100%
22	99.7%	100%	100%
23	99.3%	100%	100%
24	99%	100%	100%
25	100%	100%	100%
26	99.7%	100%	100%
27	99.7%	100%	100%
28	99.3%	100%	100%
29	100%	100%	100%
30	99.3%	99.3%	99.3%
31	99.7%	100%	100%
32	100%	100%	100%
33	100%	100%	100%
34	100%	100%	100%
35	100%	100%	100%
36	100%	100%	100%
37	100%	100%	100%
38	100%	100%	100%
39	99.7%	100%	100%
40	100%	100%	100%
41	99.3%	100%	100%
42	99.7%	99.7%	99.7%
43	100%	100%	100%
44	100%	100%	100%
45	100%	100%	100%
46	100%	100%	100%
47	100%	100%	100%
48	100%	100%	100%
49	99.7%	100%	100%
50	100%	100%	100%
51	100%	100%	100%
52	100%	100%	100%
53	100%	100%	100%
54	100%	100%	100%
55	99.3%	100%	100%
56	100%	100%	100%
57	99.7%	100%	100%
58	100%	100%	100%
59	100%	100%	100%
60	100%	100%	100%

Table 1. Pump Classification Results

Class N_c	**Match**	**+/- 2**	**+/-4**
All	93.45%	98.73%	99.58%
1 - 20	80.70%	96.25%	98.80%
21 - 40	99.83%	99.98%	99.98%

Table 2. Pump Classification Averages

Table 2 presents the average percentages for all classes, for classes 1 through 20, and for classes 21 through 40. The averages are grouped this way to more clearly display the differences between the pump running at a faster rate, thus generating more vibration, and at a slower rate while generating less vibration. Classes 21 through 40 provide an exact match on average 99.83% while classes 1 through 20 provide an exact match on average only 80.7% of the time.

What makes these numbers even more meaningful is when it is considered that the centrifugal pump on the Lab-Volt® training system used for this research in general will not be operating at a level lower than 10. This is because at levels lower than that the pump is not running fast enough for water to be forced through the system.

4.2.4 Focused Results.

MDA/ML classification works on a best match methodology. Meaning that with 60 training classes, a given set of testing data will be classified to one of those 60 classes based on which one it matches to the closest. The same would happen if only 10 classes where used for training. The testing data will be matched to one of those 10 classes. It can be inferred then that the more classes used for training, the higher the chance of being incorrectly matched. Table 3 shows the confusion matrix for training and testing with only classes 7, 8, and 9 from the original 60 classes.

This first observation from this confusion matrix is that the percentage for an exact match increased from the original testing with all 60 classes. Class 7 increased from 71% to 98%, class 8 from 79% to 94%, and class 9 from 71% to 97.3%. Figure 25

Class N_c	**7**	**8**	**9**
7	98.0%	0.7%	1.3%
8	4.0%	94.0%	2.0%
9	0.7%	2.0%	97.3%

Table 3. Confusion Matrix For Classes 7, 8, and 9

displays a scatter plot for these results. Each color in the plot represents a different class with each point representing the result from testing data. For the most part each class is grouped together as indicated by the confusion matrix. However, there are a some points that are closer to the wrong color. This the testing data that was incorrectly classified to a different class.

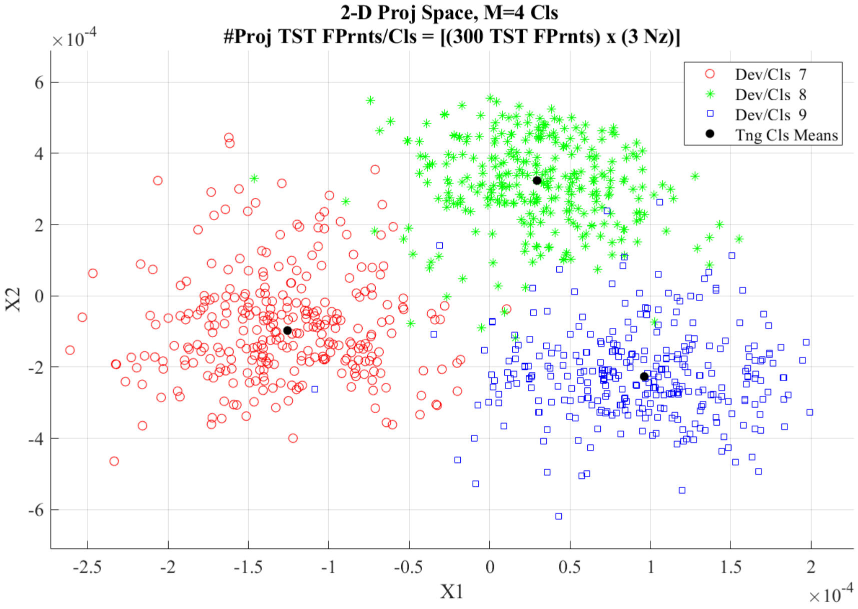


Figure 25. Scatter Plot For Classes 7, 8, and 9

Table 4 shows the confusion matrix for classes 52, 53, and 54 from the original 60. As expected based on the results from all 60 classes, each class here is matched at 100%. Figure 26 shows the scatter plot for these classes. This is were the difference between the pump running at lower levels and higher levels becomes clear. The classes are in a much tighter group with no bleed over into other classes, reinforcing the notion that the pump becomes the dominant source of vibration as it runs faster (class 19 and above).

Class N_c	**52**	**53**	**54**
52	100%	0%	0%
53	0%	100%	0%
54	0%	0%	100%

Table 4. Confusion Matrix For Classes 52, 53, and 54

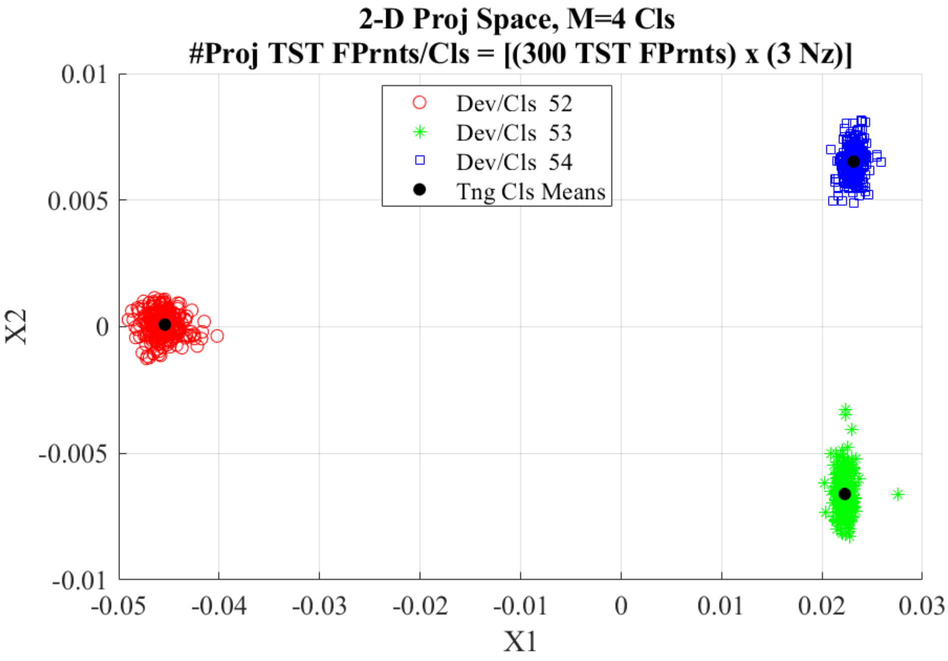


Figure 26. Scatter Plot at Pump Levels 52, 53, and 54

4.3 Analysis - Flow Meter

This section describes the analysis implementation for vibration data collected from the flow meter on the Lab-Volt® ICS training system. The analysis follows the overall process as described in Section 3.3.

4.3.1 Collected Data.

Vibration data was captured from the flow meter starting with the pump running at level 10 due to that being the lowest level where water is still forced through the system. This gives the flow meter data set $N_c = 51$ total classes with $N_s = 600000$ samples collected per class as described in Section 3.3.1. Final analysis data is comprised of $N_c = 51$ with each N_c containing $N_b = 200$ bursts and each N_b containing $N_s = 3000$ samples. In total, $N_s = 30.6$ Million samples were collected from the flow meter and used for analysis. Table 5 shows the mapping for class number to its corresponding pump level and the flow rate reported by the flow meter at that given pump level.

4.3.2 Fingerprinting.

Fingerprinting of the collected flow meter data is done according to the process laid out in in Section 3.3.2. The overview for this process is the same as it is for the pump data covered in Section 4.2.2 and will not be covered again here.

4.3.3 Classification Results.

The MDA/ML classification takes place in two steps, training and testing. The process for training and testing is outlined in Section 3.3.3 and a brief overview can be found in the classification results section for the pump data, Section 4.2.3.

Classification results are generated in the form of a confusion matrix. Table 6 presents the data from the confusion matrix by the percentage that a testing class was correctly matched to the same training class, the percent that it is correctly matched +/- 2 classes, and the percent that it was correctly matched +/- 4 classes. The results are presented this way because one, it is easier to digest than a 51 class confusion matrix, and two, knowing how close a class comes to an exact match in a

Class N_c	Pump Lvl	Flow (l/min)		Class N_c	Pump Lvl	Flow (l/min)
1	10	1.8		**27**	36	30.9
2	11	3.2		**28**	37	31.8
3	12	5.3		**29**	38	32.8
4	13	6.4		**30**	39	33.7
5	14	7.6		**31**	40	34.6
6	15	8.7		**32**	41	35.5
7	16	9.9		**33**	42	36.3
8	17	11.2		**34**	43	37.2
9	18	12.3		**35**	44	38.1
10	19	13.4		**36**	45	38.9
11	20	14.5		**37**	46	39.8
12	21	15.6		**38**	47	40.7
13	22	16.7		**39**	48	41.6
14	23	17.8		**40**	49	42.6
15	24	18.9		**41**	50	43.5
16	25	20		**42**	51	44.5
17	26	21		**43**	52	45.5
18	27	22		**44**	53	46.4
19	28	23		**45**	54	47.4
20	29	24		**46**	55	48.4
21	30	25.1		**47**	56	49.3
22	31	26.1		**48**	57	50.2
23	32	27		**49**	58	51.2
24	33	28		**50**	59	52.1
25	34	29		**51**	60	53
26	35	29.9				

Table 5. Class Mapping To Pump Level and Flow Rate

given range is useful for an ADS. Several observations can be made by looking at the results given in Table 6:

1. The percentages for an exact match are quite low and indicate that the amount of difference in vibration between each class is not enough to differentiate between classes.

2. Even at the +/- 4 class mark the percentages rarely get out of the 30's or 40's. This indicates that even at a range of nine classes it is difficult to differentiate.

3. The 100% match at class 1 is odd given the poor performance for the rest of the classes. This is probably due to class 1 looking very close to the environmental noise. At class 1 the water is barely able to flow through the system and is likely to not cause a large amount of vibration when flowing through the flow meter.

Class N_c	**Match**	**+/- 2**	**+/- 4**		**Class N_c**	**Match**	**+/- 2**	**+/- 4**
1	100%	100%	100%		**27**	4%	28%	40%
2	22%	33%	49%		**28**	6%	27%	46%
3	19%	38%	41%		**29**	4%	26%	49%
4	13%	30%	36%		**30**	5%	24%	42%
5	6%	24%	37%		**31**	8%	22%	52%
6	59%	62%	85%		**32**	11%	24%	38%
7	7%	19%	35%		**33**	11%	24%	38%
8	12%	30%	45%		**34**	2%	25%	42%
9	6%	23%	34%		**35**	6%	24%	39%
10	8%	34%	53%		**36**	4%	31%	42%
11	5%	20%	54%		**37**	7%	23%	34%
12	6%	30%	57%		**38**	3%	15%	36%
13	6%	32%	49%		**39**	6%	24%	34%
14	3%	27%	42%		**40**	7%	28%	49%
15	8%	28%	42%		**41**	6%	26%	50%
16	5%	31%	42%		**42**	3%	27%	41%
17	6%	18%	49%		**43**	8%	32%	56%
18	14%	31%	46%		**44**	2%	35%	53%
19	12%	27%	36%		**45**	11%	39%	60%
20	1%	19%	36%		**46**	13%	55%	68%
21	4%	19%	30%		**47**	11%	52%	73%
22	3%	10%	24%		**48**	13%	38%	66%
23	4%	14%	19%		**49**	12%	67%	85%
24	1%	20%	36%		**50**	19%	54%	79%
25	5%	26%	40%		**51**	22%	59%	79%
26	7%	24%	42%					

Table 6. Flow Meter Classification Results

Table 7 shows the averages for all of the classes put together. The averages paint the same picture as observed from the full set of classes with the average for an exact match at 11%, +/- 2 at 31%, and +/- 4 at 48%. Even at a range of nine classes, on average a correct match to that range happens only 48% of the time. This would likely not be sufficient for an ADS that needs to detect differences between operating states.

Class N_c	**Match**	**+/- 2**	**+/-4**
All	11%	31%	48%

Table 7. Flow Meter Classification Averages

4.3.4 Focused Results.

As was done with the results from the pump data. The analysis can be focused on a smaller set of classes to help improve the chances of classes being correctly matched. Table 8 displays the confusion matrix for training and testing with only classes 7, 8, and 9 from the original 51 classes.

Class N_c	**7**	**8**	**9**
7	47%	22%	31%
8	43%	27%	30%
9	36%	38%	26%

Table 8. Confusion Matrix For Classes 7, 8, and 9. Flow Meter

It is clear that the focused results did help a little, given that each classes percentage for an exact match did increase. However, none of them made it above 50%. This further indicates that the the vibration differences between neighboring classes is not distinct enough to differentiate between. This is visualized clearly by the scatter plot in Figure 27 where it becomes obvious that the points for each given class are all mixed together with no clear grouping as was seen with the pump results.

Table 9 shows the confusion matrix for classes 42, 43, and 44. With the higher classes from the pump data it became easier to differentiate between classes because the vibration on the pump had become the dominant source of vibration with a clear

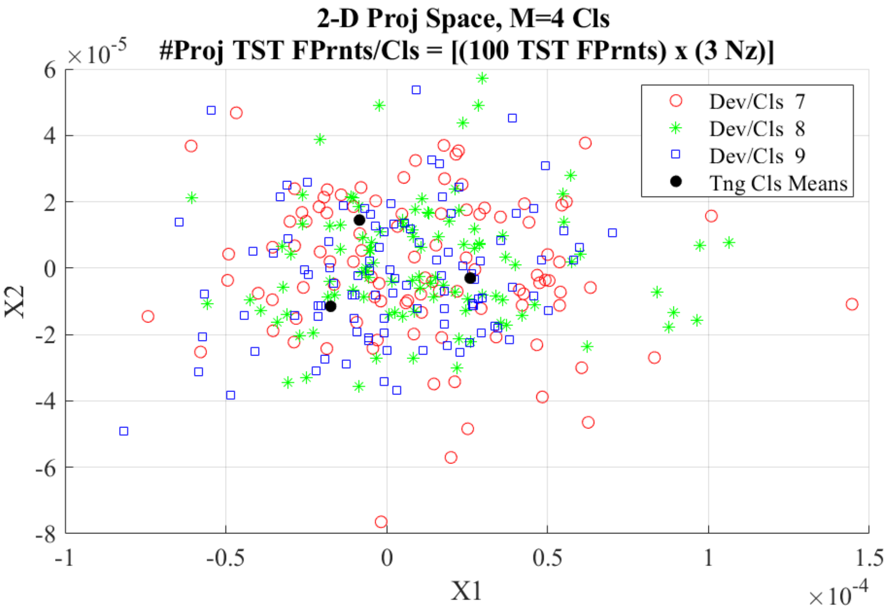


Figure 27. Scatter Plot at Classes 7, 8, and 9. Flow Meter

difference between classes. That is not the case with higher classes for the flow meter as even at the higher classes and focusing the training and testing, percentages for an exact match still do not get above 50%. Figure 28 displays the scatter plot for these classes. As with the scatter plot for the lower classes in Figure 27, this one shows no visible separation between class and no class groupings.

Class N_c	**42**	**43**	**44**
42	41%	34%	25%
43	23%	41%	36%
44	28%	38%	34%

Table 9. Confusion Matrix For Classes 42, 43, and 44. Flow Meter

Clearly there is not enough vibration differences between individual classes that are next to each other. Table 10 displays the confusion matrix for classes 5, 25, and 45. These results show that there is a somewhat significant difference between classes when there is a twenty class separation between them. Class 5 has the lowest percentage for and exact match at 74%. While class 25 and 45 come in at 87% and 92% respectively for percentage of exact match. Perhaps more telling is that class 5

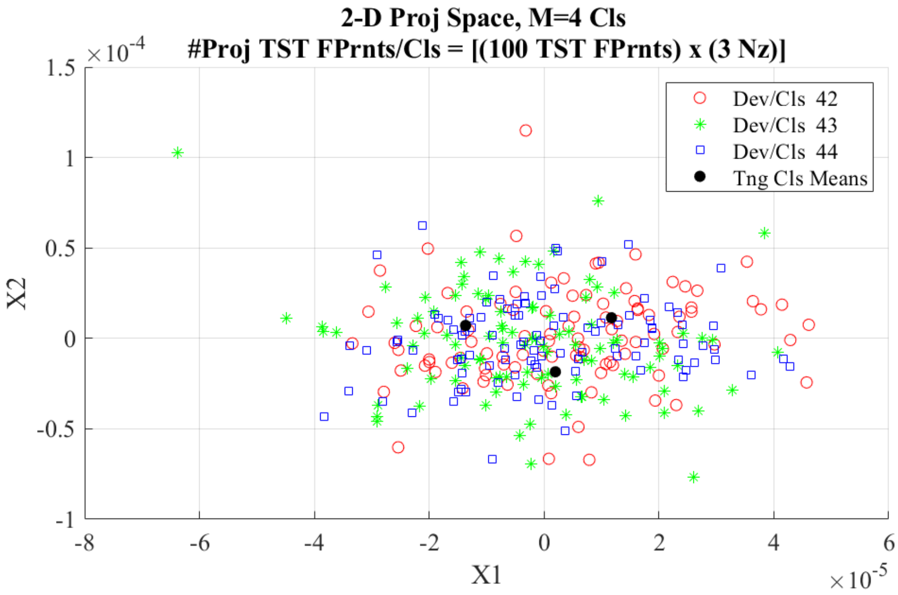


Figure 28. Scatter Plot at Classes 42, 43, and 44. Flow Meter

and 45 do not match to each other at all. Indicating that there is in fact a difference in vibration between classes, but that the difference is so gradual that it is hard to notice.

Class N_c	**5**	**25**	**45**
5	74%	26%	0%
25	7%	87%	6%
45	0%	8%	92%

Table 10. Confusion Matrix For Classes 5, 25, and 45. Flow Meter

Figure 29 displays the scatter plot for classes 5, 25, and 45. Unlike with the focused results with classes that are next to each other. This scatter plot clearly shows grouping for and individual class and some separation between the classes. This indicates that it is possible to distinguish between higher operating level with more flow and a lower operating level with less flow.

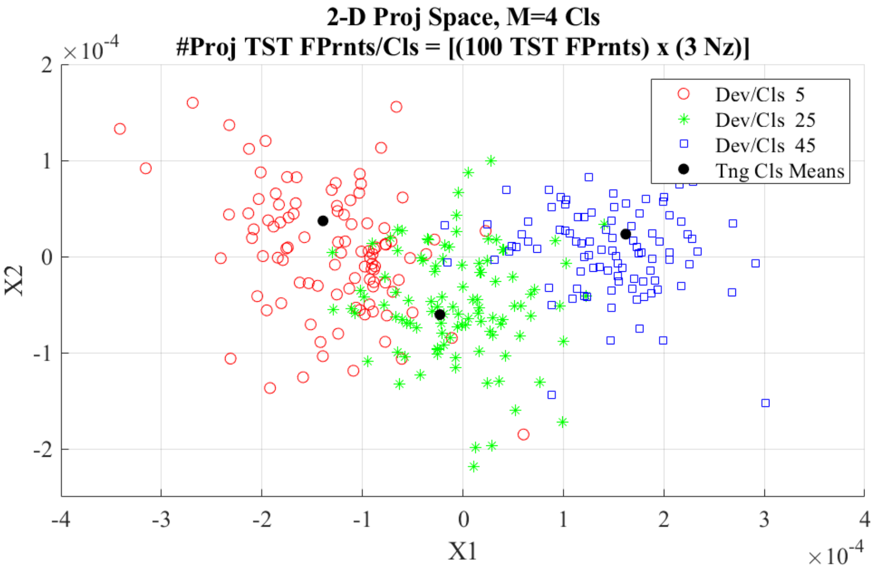


Figure 29. Scatter Plot at Classes 5, 25, and 45. Flow Meter

4.4 Summary

In summary, MDA/ML classification was performed on two sets of vibration data. One set was collected from a centrifugal pump and the other from a flow meter. Each set of data was split in half to form training and testing data sets to provide a classification based on a best match methodology. Classification results from the pump show a clear distinction between classes even at lower operating levels. Classification results from the flow meter are less clear between classes but do show a difference when there is a good amount of separation between classes being analyzed. Chapter V will present the final conclusions from analysis.

V. Conclusion

5.1 Introduction

This chapter presents the final conclusions for this research research. Section 5.2 presents the research goals and answers the questions they pose based on the results from the research. Section 5.3 describes areas for future work. Section 5.4 presents the overall summary for the research.

5.2 Conclusions of Research

This section presents the research goals at the heart of this research. The questions posed in the research goals are answered for both the results from the pump data analysis and from the flow meter data analysis.

The goals for this research are:

- Determine if the physical effects of devices in an ICS can be measured, fingerprinted, and classified with enough detail to be able to differentiate between the different levels of operation.
- Determine the viability of an ADS based on the results of the data fingerprinting and classification.

5.2.1 Pump Analysis Conclusion.

It is clear from the results presented in Section 4.2.3 that the answer to the first research goal is yes. The vibration data was able to be measured, fingerprinted, and classified. The detail of the classification was very good as well. Classification results show that operation/class levels higher than 20 can be distinctly identified at an average of 99.83% and at an average of 93.45% when all levels are considered.

With these results, an ADS is certainly viable.

5.2.2 Flow Meter Analysis Conclusion.

The conclusion is less clear for the flow meter results that are presented in Section 4.3.3. While the vibration data collected from the flow meter was able to be measure, fingerprinted, and classified. The detail of the classification left a lot to be desired. The average for an exact match were only 11% and only made it up to 48% in a +/- 4 class range. This issue appears to be that the difference in vibration levels between classes were not distinct enough to be able to differentiate between them. The only promising area shown in the flow meter results were when comparing classes that had a twenty class separation.

Given these results, an ADS using the vibration from the flow meter is likely not a viable option.

5.3 Future Work

This section presents some of the areas for future work from this research.

5.3.1 Data Collection.

Possible future work for data collection include:

- Narrow down the exact frequencies given off by the pump when running at different operating levels.
- When frequencies are known, identify a proper sample rate for data collection at those frequencies.
- Determine if further signal amplification can provide better results from the flow meter.

- Refine the design of the sensor so that it can be permanently placed on the target device.

5.3.2 Data Analysis.

Possible future work for data collection include:

- Perform analysis in a domain other than the SD. TD for example.
- Fine tune subregion and statistical feature calculation.
- Determine if a more narrow ROI in the signal can be used.

5.3.3 Anomaly Detection System (ADS).

The results from the pump analysis provided the proof of concept to move forward with implementing an ADS. Some consideration should be made before implementing:

- This research used a continuous training style. Meaning that some of the collected data was used for training and the rest for testing. An ADS would have to train with known good data and test live data coming in.
- Data collection for this research was a manual process. This will need to be automated for use in an ADS

5.4 Summary

The goal of this research was to provide a proof of concept for an ADS that could be able to detect an attack such as Stuxnet. This research was able to collect vibration data from ICS devices at different operating levels. Statistical fingerprints were able to be generated from that data to be used for training and testing in a classification system. The results from classification indicate that for at least one area of the research, an ADS based on the process is possible and future research into the area should be performed.

Bibliography

1. “Infrastructure security,” Nov 2018. [Online]. Available: https://www.dhs.gov/topic/critical-infrastructure-security

2. B. Obama, “Presidential policy directive 21: Critical infrastructure security and resilience,” *Washington, DC*, 2013.

3. K. Stouffer, J. Falco, and K. Scarfone, “Guide to industrial control systems (ics) security,” *NIST special publication*, vol. 800, no. 82, pp. 16–16, 2011.

4. N. Falliere, L. O. Murchu, and E. Chien, “W32. stuxnet dossier,” *White paper, Symantec Corp., Security Response*, vol. 5, no. 6, p. 29, 2011.

5. J. E. Company, “The piezoelectric effect - piezoelectric motors & motion systems,” Aug 2018. [Online]. Available: https://www.nanomotion.com/piezo-ceramic-motor-technology/piezoelectric-effect/

6. S. J. Stone, “Radio frequency based programmable logic controller anomaly detection,” AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH GRADUATE SCHOOL OF . . . , Tech. Rep., 2013.

7. J. Lopez, M. A. Temple, and B. E. Mullins, “Exploitation of hart wired signal distinct native attribute (ws-dna) features to verify field device identity and infer operating state,” in *International Conference on Critical Information Infrastructures Security.* Springer, 2014, pp. 24–30.

8. C. M. Talbot, M. A. Temple, T. J. Carbino, and J. A. Betances, “Detecting rogue attacks on commercial wireless insteon home automation systems,” *Computers & Security*, vol. 74, pp. 296–307, 2018.

9. T. J. Carbino, “Exploitation of unintentional ethernet cable emissions using constellation based-distinct native attribute (cb-dna) fingerprints to enhance network security,” AIR FORCE INSTITUTE OF TECHNOLOGY WRIGHT-PATTERSON AFB OH GRADUATE SCHOOL OF . . . , Tech. Rep., 2015.

10. C. K. Dubendorfer, “Using rf-dna fingerprints to discriminate zigbee devices in an operational environment,” AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH GRADUATE SCHOOL OF . . . , Tech. Rep., 2013.

11. D. Reising, “Classifying emissions from global system for mobile (gsm) communication devices using radio frequency (rf) fingerprints,” Ph.D. dissertation, Master’s thesis, Air Force Institute of Technology, 2009.

12. D. R. Reising, “Exploitation of rf-dna for device classification and verification using grlvqi processing,” AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH GRADUATE SCHOOL OF . . . , Tech. Rep., 2012.

13. M. D. Williams, S. A. Munns, M. A. Temple, and M. J. Mendenhall, “Rf-dna fingerprinting for airport wimax communications security,” in *2010 Fourth International Conference on Network and System Security*. IEEE, 2010, pp. 32–39.

14. W. E. Cobb, E. W. Garcia, M. A. Temple, R. O. Baldwin, and Y. C. Kim, “Physical layer identification of embedded devices using rf-dna fingerprinting,” in *MILITARY COMMUNICATIONS CONFERENCE, 2010-MILCOM 2010*. IEEE, 2010, pp. 2168–2173.

15. R. W. Klein, M. A. Temple, and M. J. Mendenhall, “Application of wavelet-based rf fingerprinting to enhance wireless network security,” *Journal of Communications and Networks*, vol. 11, no. 6, pp. 544–555, 2009.

16. G. Ionescu, O. Ionescu, S. Popovici, S. Costea, V. Dumitru, M. Brezeanu, G. Stan, and I. Pasuk, “Wireless ain sensor for condition based monitoring of industrial equipment,” in *CAS 2013 (International Semiconductor Conference)*, vol. 1. IEEE, 2013, pp. 55–58.

17. X. Xue, V. Sundararajan, and W. P. Brithinee, “The application of wireless sensor networks for condition monitoring in three-phase induction motors,” in *2007 Electrical Insulation Conference and Electrical Manufacturing Expo*. IEEE, 2007, pp. 445–448.

18. S. Sridhar, K. U. Rao, M. Nihaal, and A. S. Aashik, “Real time wireless condition monitoring of induction motor,” in *2016 IEEE Industrial Electronics and Applications Conference (IEACon)*. IEEE, 2016, pp. 173–178.

19. “3 phase ac induction motor working and its controlling using svpwm,” Oct 2014. [Online]. Available: https://www.elprocus.com/three-phase-ac-induction-motor-control-using-svpwm/

20. E. Osmanbasic, “Three-phase electric power explained.” [Online]. Available: https://www.engineering.com/ElectronicsDesign/ElectronicsDesignArticles/ArticleID/15848/Three-Phase-Electric-Power-Explained.aspx

21. “Hsc physics - motors and generators - dot point notes.” [Online]. Available: https://dc.edu.au/dot-point-summary-motors-and-generators/

22. Anish and K. Patil, “Construction and working of 3 phase induction motor on ship,” Mar 2018. [Online]. Available: https://www.marineinsight.com/marine-electrical/construction-and-working-of-3-phase-induction-motor-on-ship/

23. “Useful information on centrifugal pumps.” [Online]. Available: https://www.michael-smith-engineers.co.uk/resources/useful-info/centrifugal-pumps

24. C. E. Yang, “What is the piezoelectric effect?” Mar 2017. [Online]. Available: https://www.electronicdesign.com/power/what-piezoelectric-effect

25. Apc, “Apc international,” Nov 2018. [Online]. Available: https://www.americanpiezo.com/blog/top-uses-of-piezoelectricity-in-everyday-applications/

26. “Arduino uno wifi rev2.” [Online]. Available: https://store.arduino.cc/usa/arduino-uno-wifi-rev2

27. “Analog to digital conversion.” [Online]. Available: https://learn.sparkfun.com/tutorials/analog-to-digital-conversion

28. “Operational amplifiers (opamps).” [Online]. Available: http://www.rfcafe.com/references/electrical/opamps.htm

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704–0188*

1. REPORT DATE *(DD–MM–YYYY)* 21–03–2019	**2. REPORT TYPE** Master's Thesis	**3. DATES COVERED** *(From — To)* Jan 2013 — Mar 2019

4. TITLE AND SUBTITLE SIDE CHANNEL ANOMALY DETECTION IN INDUSTRIAL CONTROL SYSTEMS USING PHYSICAL CHARACTERISTICS OF END DEVICES	**5a. CONTRACT NUMBER** **5b. GRANT NUMBER** **5c. PROGRAM ELEMENT NUMBER**
6. AUTHOR(S) Harris, Ryan D., Civilian	**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-19-M-032
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Intentionally Left Blank	**10. SPONSOR/MONITOR'S ACRONYM(S)** **11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

12. DISTRIBUTION / AVAILABILITY STATEMENT

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

13. SUPPLEMENTARY NOTES

This work is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

14. ABSTRACT

Industrial Control Systems (ICS) are described by the Dept of Homeland Security as systems so "vital to the United States that their incapacity or destruction would have a debilitating impact on our physical or economic security." Attacks like Stuxnet show these systems are vulnerable. The end goal for Stuxnet was to operate centrifuges outside their normal parameters and hide the activity from the ICS operator. This research provides a proof of concept for an anomaly detection system that would be able to detect an attack like Stuxnet by measuring the physical change in vibration caused by the attack. The attack can hide what is reported to the operator, but it cannot hide the physical changes caused by the attack. This research uses a piezoelectric vibration sensor to collect vibration data coming from a centrifugal pump and flow meter on an ICS training system at each operating level. The collected data is then fingerprinted and classified using established RF-DNA techniques to determine if it can differentiate between the vibrations produced at each of the operating level. A clear differentiation between operating levels indicates that an ADS is feasible.

15. SUBJECT TERMS

ICS, Vibration Sensor, RF-DNA

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Dr. Robert F. Mills, AFIT/ENG
U	U	U	UU	72	**19b. TELEPHONE NUMBER** *(include area code)* (937) 255-3636, x4527; robert.mills@afit.edu