Artificial Immune System (AIS) Based Intrusion Detection System (IDS) for Smart Grid Advanced Metering Infrastructure (AMI) Networks

Final Report
CS 4624: Multimedia, Hypertext, and Information Access
Virginia Polytechnic Institute and State University
Blacksburg, VA 24061
Instructor: Dr. Edward A. Fox
Client: James R. Morris-King
May 9, 2018

Kevin Song, Paul Kim, Shivani Rajasekaran, Vedant Tyagi

Table of Contents

Table of Figures

Table of Tables

Executive Summary

The Smart Grid is a large system consisting of many components that contribute to the bidirectional exchange of power. The reason for it being "smart" is because vast amounts of data are transferred between the meter components and the control systems which manage the data. The scale of the smart grid is too large to micromanage. That is why smart grids must learn to use Artificial Intelligence (AI) to be resilient and self-healing against cyber-attacks that occur on a daily basis. Unlike traditional cyber defense methods, Artificial Immune System (AIS) principles have an advantage because they can detect attacks from inside the network and stop them before they occur.

The goal of the report is to provide a proof of concept that an AIS can be implemented on smart grid AMI (Advanced Metering Infrastructure) networks and furthermore be able to detect intrusions and anomalies in the network data. The report describes a proof of concept implementation of an AIS system for intrusion detection with a synthetic packet capture (pcap) dataset containing common Internet protocols used in Smart grid AMI networks.

An intention of the report is to provide the necessary background for understanding the implementation in the later sections. The background section defines what a smart grid is and how its Advanced Metering Infrastructure (AMI) works, describing all three networks the AMI consists of. The Wide Area Network (WAN) is one of the three networks and we were scoping down to WAN for our project. The report goes on to discuss the current cyber threats as well as defense solutions related to the smart grid network infrastructure today. One of the most widely used defense mechanisms is the Intrusion Detection System (IDS), which has many important techniques that can be used in the AIS based IDS implementation of this report.

The most commonly used AIS algorithms are defined. Specifically, the Negative Selection Algorithm (NSA) is used for our implementation. The NSA algorithm components used in the implementation section are thoroughly explained and the AIS based IDS framework is defined. A list of AIS usages/values in enterprise networks is presented as well as research on current NSA use in AIS implementations.

The latter portion of the report consists of the design and implementation. Due to data constraints and various other limitations, the team wasn't able to complete the initial implementation successfully. Therefore, a second implementation design was created, leading to the main implementation which meets the project's objective. The implementation employs a proof of concept approach using a C# console application which performs all steps of an AIS on user created network data.

In conclusion, the second implementation has the ability to detect intrusions in a synthetic dataset of "man-made" network data. This proves the AIS algorithm works and furthers the understanding that if the implementation was scaled up and used on real-time WAN network data it would run successfully and prevent attacks. The report also documents the limitations and problems one can run into when attempting to implement a solution of this scale. The ending sections of the report consists of the Requirements, Assessment, Assumptions, Results, and lessons learned followed by the Acknowledgments to MITRE Corporation which helped immensely throughout the development of the report.

01. Introduction

The mentality of fending off cyber-attacks has always been “Harden the system”. But recent events have allowed us to realize that no matter the defense we put up, cyber criminals have always found a way to get into a network and cause harm. This report will focus on how to break from this repeated cycle of learning about attacks after they have transpired, and move to sensing an attack from the inside and stopping it before it occurs. This mentality helps to stop a threat before it becomes a serious attack.

In this report we will first discuss the smart grid and define its advanced metering infrastructure (AMI) network system which contains home area network (HAN), neighborhood area network (NAN), and wide area network (WAN). We also talk about the current cyber defense technologies for the AMI and the problems they face. Next, we discuss the traditional methods of an Intrusion Detection System and how it works, with supporting examples. We end with an explanation of an AIS inspired IDS and describe briefly the main algorithms and methods it consists of. Going further along this project the next steps would be to apply and optimize the AIS algorithms and apply it to the network data from the three networks for a better resulting IDS to defend against AMI related cyber-attacks.

We focus on securing the Automatic Metering Infrastructure (AMI) networks in the smart grid using artificial immune system-based intrusion detection. So first, why is the smart grid in danger of security vulnerabilities? The new smart grid incorporates the traditional power grid with communication and information technologies. The smart grid consists of a vast network which is complex in that it contains millions of devices which are connected to each other. A vast network such as a smart grid comes with many security concerns and vulnerabilities. Current cyber defense technologies such as encryption and firewalls aren’t enough to fully defend against cyber-attacks. That is why we must look at using artificial immune system-based intrusion detection, because this system not only has a great pattern recognition system to find attacks but also fights back.

02. Problem Statement

We want to address the security concerns in Smart Grid Automatic Metering Infrastructure (AMI) networks. Previously in the power grids the control networks were not as well automated, and they ran over private communication networks. The new upgrade to smart grid system demands a significant increase in secure and intelligent communication infrastructures. The massive size of the smart grid and the increased capabilities in communication make it more prone to cyber-attacks. Defending against these attacks is vital in protecting customers secure data and in general keeping the smart grid network unharmed.

Security risk reasons in the Smart Grid, from National Institute of Standards and Technology (NIST), include [1].

- Increased complexity of the grid can facilitate attacks and create vulnerabilities.
- Interconnected networks can cause Internet Protocol (IP) communications network vulnerabilities.
- Compromised software or hardware can cause various attacks.
- Increased points of entry into Smart Grid Systems can facilitate attacks.
- Increase of new technology means increased chances of new attacks.
- Expansion of the amount of data can lead to compromise of customer secure data.

Current defense mechanisms against this issue include mainly an Intrusion Based Detection System (IDS). IDS is a way to distinguish malicious vs. benign intrusions and it is done by monitoring a network for such behaviors. This technique detects cyber attacks but unfortunately mostly finds them after they have already entered the network and caused damage. What this report proposes using is an Artificial Immune System (AIS) technique called Negative Selection Algorithm (NSA) in defending against attacks which are common to smart grid AMI networks. The advantage of AIS based IDS is its use of biologically influenced concepts in computation to stop a network attack by determining malicious patterns even before the attack happens.

03. Objective

The major goal of this report is to provide a proof of concept that an AIS can be implemented on Smart Grid WAN networks using AIS and network communications concepts. The implementation will be a proof of concept if it is able to pick up on network intrusions on the test data used. Although this is the major goal, the report will also discuss research on:

- Smart Grid AMI Infrastructure
- Intrusion Detection Systems and most common attacks and defenses
- Current AIS research in the field of cyber security,
- AIS framework for IDS and various AIS algorithms, specifically Negative Selection which is used in the implementation
- Workings of communications network Internet Protocols (IP) commonly used in WAN

04. Background

a. What is Smart Grid?

The Smart Grid consists of a network, substations, transformers, and more that deliver electricity from the power plant to our home or business. Smart grid is the inclusion of digital technology that allows for two-way communication between the utility and its customers, and the sensing along the transmission lines. Like the Internet, the Smart Grid will consist of controls, computers, automation, and new technologies and equipment working together with the electrical grid to respond digitally to the quickly changing electric demand. [2]

i. What is Advanced Metering Infrastructure (AMI)?

Advanced Metering Infrastructure (AMI) refers to systems that measure, collect, and analyze energy usage, and communicate with metering devices such as electricity meters, gas meters, heat meters, and water meters, either on request or on a schedule. In case of smart grids, The Federal Energy Regulatory Committee (FERC) defines AMI as “a metering system that records customer consumption hourly or more frequently and that provides for daily or more frequent transmittal of measurements over a communication network to a central collection point.”[3]

ii. AMI Networks- (HAN, NAN, WAN):

The infrastructure consists of several networks which could rely on different media and a multitude of protocols. In total, three networks are commonly described when referring to the AMI.

WAN - The wide area network (WAN) or field area network (FAN) are often referred to as the back end of the smart grid network infrastructure. These networks provide communications from the utility head end out to devices in the field, in this case from the network to the home devices. The WAN is also used for individual direct connect meters. WAN in a robust IP network should be able to support many different communication technologies such as the power grids. [4]

NAN - The neighborhood area network (NAN) provides sub-networks of meters, typically extending the reach to the majority of the meter population—especially residential meters. They can be in the form of Power Line Carrier (PLC) networks that form the base for the HAN. [4]

HAN - The home area network (HAN) provide interfaces into the home and business for energy consumption monitoring and to support demand response functionality. The HAN includes the communication network from the meter to devices inside the consumer’s home (or commercial building). Most HAN traffic occurs between the meter and the in-home display (IHD) and load control devices. Like the NAN, HAN communications can also be transmitted via PLC technologies. [4]

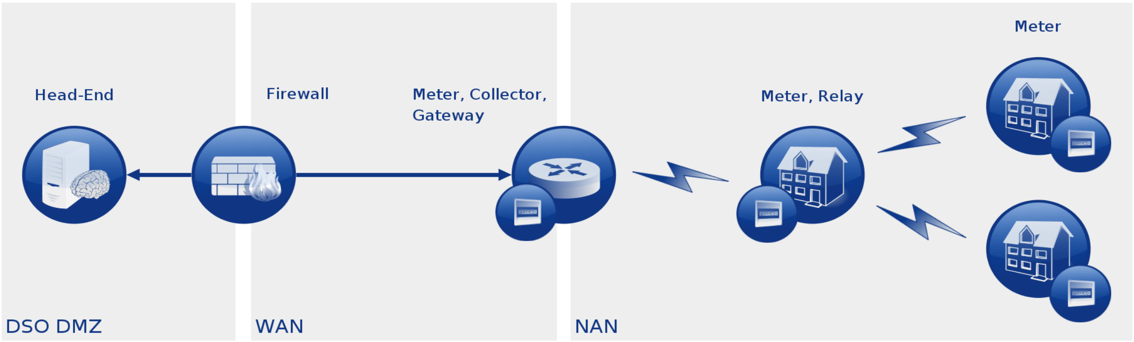


Figure 1: Advanced Metering Infrastructure [5]

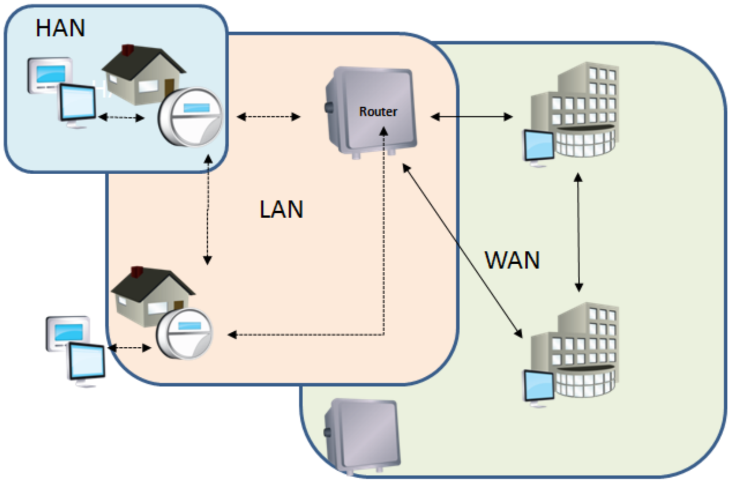


Figure 2: Advanced Metering Infrastructure (AMI) Networks [6]

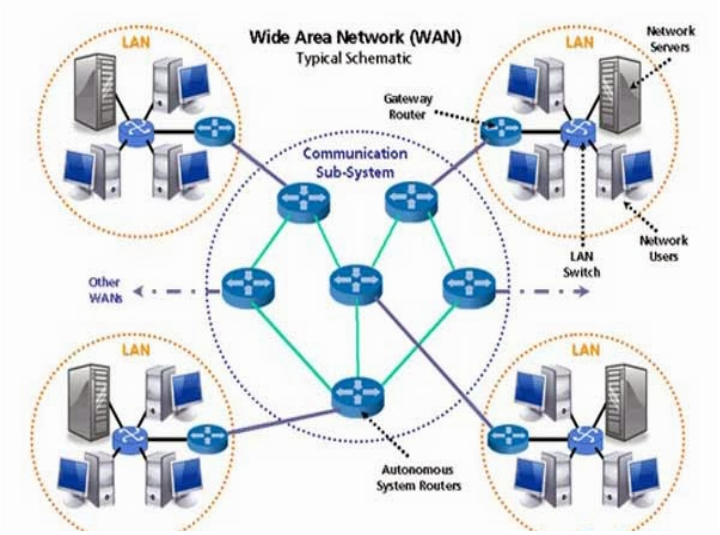


Figure 3: Wide Area Network (WAN) diagram [7]

Figure 1 and figure 2 are the diagrams of the Advanced Metering Infrastructure and its components which are categorized in accordance with the belonging network area. And Figure 3 is a diagram of Wide Area Network and its components. Wide Area Network consists of a set of Local Area Network and It doesn't matter what the WAN joins together or how far apart the networks are; the end result is always intended to allow different smaller networks from different locations to communicate with one another.

Since WANs, by definition, cover a larger distance than LANs, it makes sense to connect the various parts of the WAN using a virtual private network (VPN). A VPN creates a secure tunnel that protects your data and allows all traffic, voice or data to pass through a public WAN as if it was on a private circuit or LAN. This provides protected communications between sites, which is necessary given that the data transfers are happening over the internet. Although VPNs provide reasonable levels of security for business uses, a public internet connection does not always provide the predictable levels of performance that a dedicated WAN link can. This is why fiber optic cables are sometimes used to facilitate communication between the WAN links. [8]

b. Current cyber-attack problems in Smart Grid

Type of Attacks	Characteristics	Target
Man-in-the-Middle Attacks	Attacker becomes the middle-man to oversee the user actions.	Network/Communication
Denial-of-Service Attacks	Attacker tries to prevent access by the user.	Network/Computer
SQL Injection Attacks	Attacker injects a SQL Query to disrupt the functionality by the user.	Network/Computer
Application-Layered Attacks	Attacker targets an application by deliberately causing an error.	Network/Application
Identity-Spoofing Attacks	Attacker falsifies a source IP to enter user system.	Network/Computer
Compromised Key Attacks	Attacker obtains a key for securing the access to user progress.	Network/Communication

Table 1: Different Types of Network Attacks in Cyber Security [9]

c. Current cyber defense solutions in Smart Grid

Cryptographic mechanism is the most widely used current form of cyber defense. There are some examples of processing by cryptographic hash functions. See figure 4. Cryptography is almost always used to refer to electronic scrambling of data, but in a historical context, cryptography refers to using written secret codes. In addition, cryptography is the branch of information security which covers the study of algorithms and protocols that secure data on transmission over the Internet and on static computer systems. [10]

Some solutions of core cryptography security are confidentiality, integrity, authentication, and authorization.

Confidentiality - Encryption and decryption are the ways to make sure that information remains confidential while it’s stored and transmitted. Encryption converts information into code that makes it unreadable until it is decrypted.

Integrity - Integrity ensures that changes can’t be made to data without appropriate permission. If a system has integrity, it means that the data in the system is moved and processed in predictable ways. Also, there is use of a cryptographic hash function. It takes a message as an input, and returns a fixed-sized string.

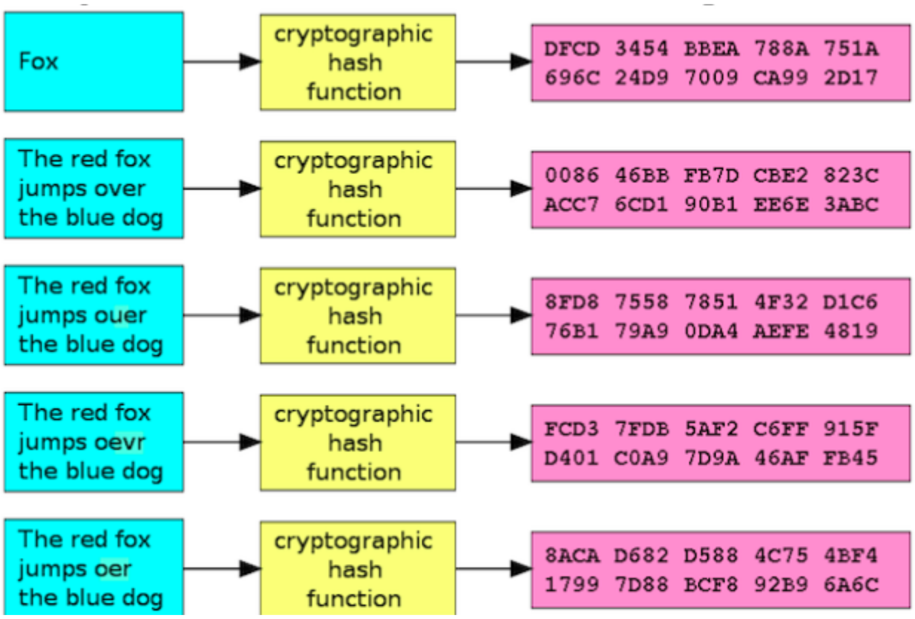

Fox
cryptographic hash function
DFCD 3454 BBEA 788A 751A 696C 24D9 7009 CA99 2D17
The red fox jumps over the blue dog
cryptographic hash function
0086 46BB FB7D CBE2 823C ACC7 6CD1 90B1 EE6E 3ABC
The red fox jumps ouer the blue dog
cryptographic hash function
8FD8 7558 7851 4F32 D1C6 76B1 79A9 0DA4 AEFE 4819
The red fox jumps oevr the blue dog
cryptographic hash function
FCD3 7FDB 5AF2 C6FF 915F D401 C0A9 7D9A 46AF FB45
The red fox jumps oer the blue dog
cryptographic hash function
8ACA D682 D588 4C75 4BF4 1799 7D88 BCF8 92B9 6A6C

Figure 4: Examples of processing by cryptographic hash function [11]

d. What is Intrusion Detection System (IDS)?

Intrusion detection systems (IDSs) monitor network traffic in order to detect when an intrusion is being carried out by unauthorized entities. An intrusion detection system (IDS) is a type of security software designed to automatically alert administrators when someone or something is trying to compromise information system through malicious activities or through security policy violations. [12]

An IDS works by monitoring system activity through examining vulnerabilities in the system and the integrity of files, and by conducting an analysis of patterns based on already known attacks. It also automatically monitors the Internet to search for any of the latest threats which could result in a future attack. [12]

There are two types of IDS: Network based (NIDS) and Host based (HIDS) intrusion detection systems. [13] Figure 5 is a diagram of the two commonly classified IDS and it shows how it generally works. See Figure 5.

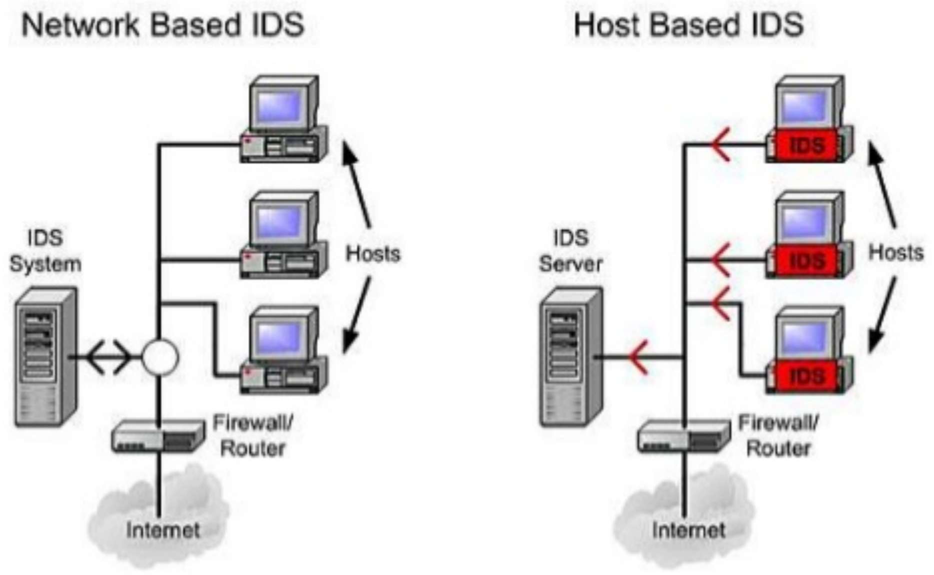


Figure 5: Network based and host-based IDS [14]

IDS normally detect on two different ideas, which are forbidden and suspicious activities, or normal activities. In other words, the system needs to learn what is normal vs. suspicious. After that, the system is able to distinguish what is normal and what is unusual. Because of this, a user must define the activities by tuning AI, so that it does not try to keep false negatives, at a minimum. AI focuses on detecting activities that are allowed. Here are the different types of algorithms and techniques that are used to detect intrusion:

- Fuzzy logic
- Probability reasoning
- Neural networks
- Genetic algorithms

i. How does IDS work?

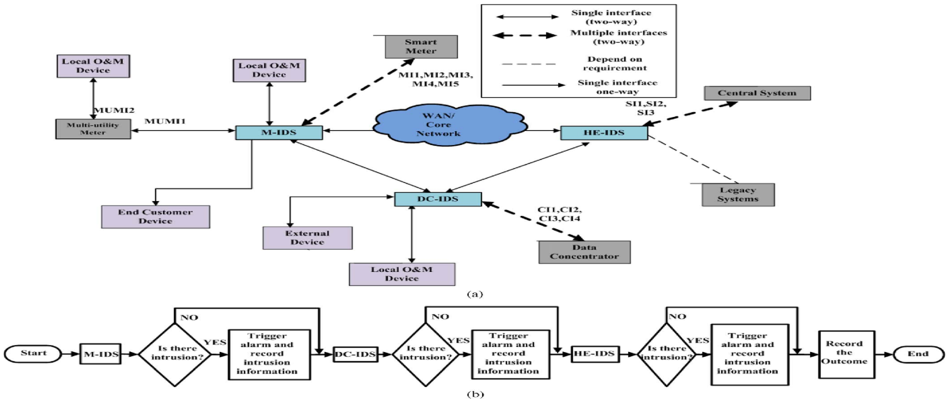


Figure 6: Intrusion Detection System diagram and its working process [37]

Intrusion detection can be either network- or host-based. A network-based IDS settles in the network and a host-based intrusion detection system is installed on the client computer. Figure 6 is a diagram of IDS and its working process on the bottom. See figure 6.

A good network-based intrusion detection systems is SNORT. SNORT was released by Martin Roesch who is the founder and CTO of SourceFire. SNORT, which is an open source intrusion detection system, analyses the data traffic in the network. [15]

e. What is Artificial Immune Systems (AIS)?

The main function of a biological immune system is to arm the body against foreign molecules known as antigens. The immune system has a wonderful recognition system which can detect changes in patterns and report abnormal behaviors in the system. AIS utilizes a machine language which incorporates the functionality of the biological immune system.

Research on AIS started in the mid 1980s by Farmer, Packard, and Perelson's study. They created a new kind of computer science learning which followed the immune system structure. Researchers in computer security started to propose the immune system as an analogy for IDS, which a lot of engineers started using [15]. AIS works by learning patterns for use in areas such as security, anomaly detection, optimization, and fault detection [16]. This report considers using AIS for anomaly detection in communication network data.

i. Basic AIS terms [17]

- Lymphocytes: Major body's defense mechanism. Can contain one or more antibodies attached to it.
- Antibody: Detectors on Lymphocytes that bind to harmful antigens and hence neutralize or destroy them.
- Antigen: A foreign substance, usually proteins which can cause an immune response
- Self-Antigen: A non-harmful antigen

ii. Framework for AIS based IDS

Three steps need to be followed to be able to apply the AIS framework based on IDS. The first step is to use an immune language and form to represent the elements in the network and their interactions. The reason for this is to show the ID elements by creating abstract models of immunology particles such as cells and molecules. In this case we can quantify the interaction between the elements by affinity measures. For example, antigen is used in AIS to show the abnormal activity in IDS. The second step is to generate the algorithms and lastly the third step is to optimize the algorithms. You can visualize these steps in Figure 7. This framework can be used as a procedure for design by an engineer who wants to work with AIS inspired IDS [18].

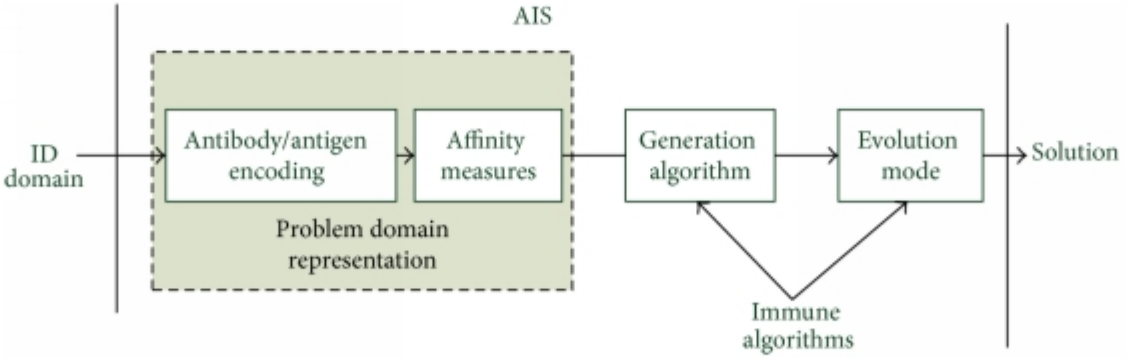


Figure 7: AIS based IDS framework [19]

iii. AIS Algorithms:

The three main theories of AIS algorithms are affinity, clonal selection, and negative selection.

1. Affinity:

Different models are used in AIS to calculate the affinity between antibodies (defending) and antigens (attacking). The affinity model is very important because the capability of detection relies on the affinity between the antigen and the detector. Let's assume that the coordinates of an antibody are given by Ab=(Ab1, Ab2, …, Abn) and the coordinates of an antigen are Ag=(Ag1, Ag2, …, Agn); the distance between them, D, is the affinity. [9]

There are various algorithms to determine the affinity. See Figure 8

Let $R=<r_0, r_1 \dots r_m>$, $S=<s_0, s_1 \dots s_m>$,

- Euclidean $D = \sqrt{\sum_{i=0}^{m} (r_0 - s_0)^2}$
- Manhattan $D = \sum_{i=0}^{m} |r_0 - s_0|$
- Hamming $D = \sum_{i-0}^{m} \delta = 1$ *if* $r_0 \neq s_0$, 0 *if otherwise*

Figure 8: Affinity algorithms [20]

2. Clonal Selection Algorithm:

Clonal selection describes a response to an antigen by the immune system. The antibodies that can recognize the antigens multiply and are chosen over ones that do not. This allows detectors to clone their parents by a mutation mechanism with high rates while the antibodies which are self-reactive get eliminated. This act is known as clonal selection. CLONALG was the algorithm created by De Castro which is based on clonal selection [21]. This algorithm takes into account all counts about cloning the best antibodies, affinity maturation, taking out non-stimulated antibodies, and maintaining diversity. Clonal selection has a great strategy for optimization and pattern recognition. This helps evolve the immune system, so it can recognize the antigens that it met in the past [22].

3. Negative Selection Algorithm:

This algorithm was first created by Forrest in AIS. The main goal of this algorithm is to provide tolerance for self-antigens. It develops an ability to differentiate against harmful antigens that are not part of the self-antigens. Once the normal pattern is defined antigens are formed which detects anomalies on random antigens. This algorithm is a model for normal and abnormal discrimination in the process of detector maturation. This allows the AIS to take out immature detectors that have a larger affinity with normal or self-samples than the threshold which is predetermined. Based on this theory the affinity between antigen (abnormal) vs. self-samples should be lower than the user specified threshold [23].

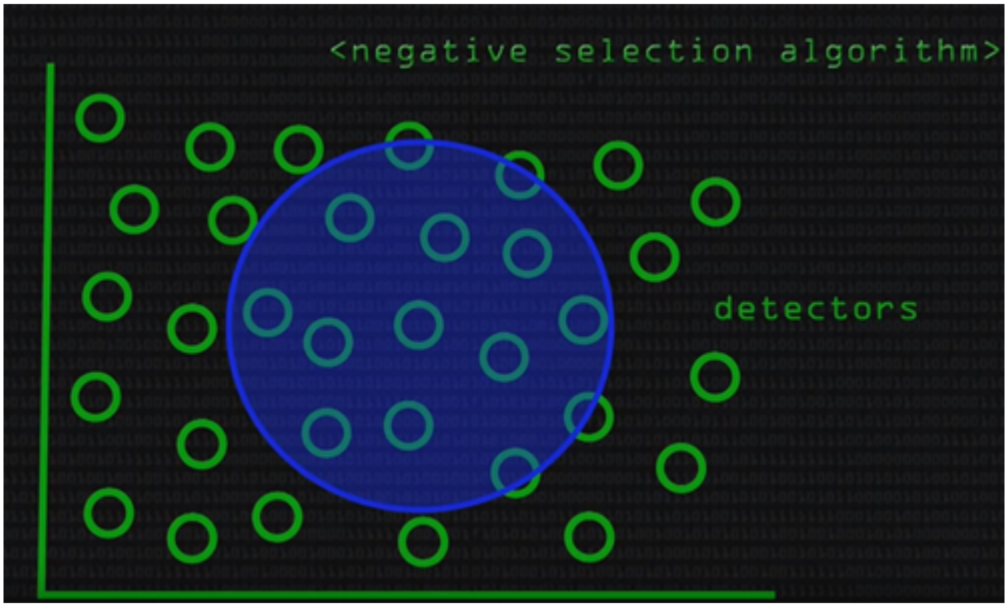


Figure 9: Negative Selection Algorithm Process [24]

Figure 9 shows detectors (green circles) being plotted or placed in a feature space which represents the data, which could be binary bits. This space maps attributes from the feature vector. Everything inside the blue circle represents the normal region. The next step in the algorithm is to delete all the detectors that touch the blue circle line, so we end up with only the abnormal region, so we can next classify the anomalies. This is a basic description of the NSA functionality. Now the detector formation will be explained in depth to understand how anomalies are found.

a. NSA process

Detector formation

The Negative Selection anomaly detection is based on the generation of detectors; these detectors are then used to find patterns of anomalies in the network.

The general steps of the NSA algorithm are given as below:

1. Create a set of self-strings (S).
2. Generate a random set of strings (R).
3. For $r_0 \sum r_0$, if under certain algorithms, $r_0$does not match with any $s \sum S$, then $r_0$will be saved in the detector set (D), otherwise, $r_0$will be rejected.

Figure 10 describes the steps mentioned above:

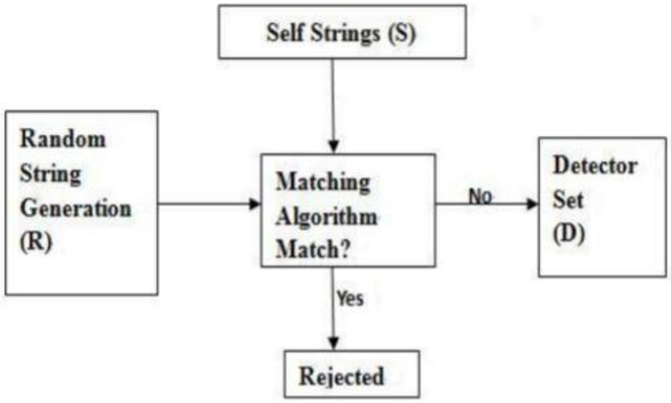


Figure 10: The negative selection algorithm detector formation [21]

The self-string (S) will take the inputs which are shown to be in “normal behavior” in the network. The random string (R) will be network data generated which might contain intrusions.

In Figure 10 it is easy to notice that the test “Matching Algorithm Match?” is the main functionality of this algorithm. The purpose of the matching algorithm is to find the affinity between the self-string (S) and the random string (R).

There are various algorithms to determine the affinity. But for the purposes of the implementation we will be using the Knuth Morris Pratt (KMP) or R-chunk bit matching formula to calculate our affinity between the self-string (S) and the random string (R).

Knuth Morris Pratt (KMP) algorithm functionality example [25]:

Input:

Txt: 010110110111
Pattern: 1101

Output: Pattern found at index 3

This algorithm will be used to detect anomalies in network data later in the implementation.

b. Why Negative Selection Algorithm was chosen for implementation

Negative Selection Algorithm is one of the most successful methods in AIS, and its typical applications include change detection, fault detection, and network intrusion detection. One of the NSA’s strengths is that the low-level representation of detectors prevents the extraction of meaningful domain knowledge. For example, a computer administrator often stops some network services and new network attacks always occur [25]. The implementation also has a well-defined self-set, so it is easier to use negative selection to form patterns that do not fit the self-set.

c. AIS use/value in enterprise networks

The role of AIS is to get there first and take measured action to mitigate risks, before the human arrives on the scene. Attacks are becoming more aggressive, so defenders need to directly fight back at machine speed.

1. Darktrace: Antigena
   Darktrace Antigena acts as a digital antibody, taking only very targeted remedial action against in-progress threats – for example, it can slow down or stop a compromised connection or device, but does not impact normal business operations. [26]

2. Symantec Corporation: The digital immune system
   The Digital Immune System will include tools and utilities for systems and policy management, virus protection, server performance, desktop configuration, diagnostics,

system stability, remote system operation, management of remote users, and disaster recovery -- all from a single management console. [27]

3. Cylance Inc.: Alpha Locker
   Alpha locker is written in C#; it has a minimum weight of up to 50 kilobytes. Locker encrypts all drives connected to the pc. Continues to encrypt files when the computer is turned off. Decryption can decrypt the chosen file or an entire folder. Admin panel has statistics and general information. The scripts backup and restore the database, which increases data reliability.

4. Current AIS Research Involving Negative Selection

a. Application of Artificial Immune System in Swarm Robotic Systems

Swarm Robotics is the study of how large numbers of relatively simple physically embodied agents can be designed such that a desired collective behavior emerges from the local interactions among agents and between the agents and the environment.

The solution was divided into 3 parts: Immune Modelling, Theoretical AIS, Applied AIS. Immune modelling is focused on mathematical models and simulations of natural and artificial immune systems. Theoretical AIS is concerned with the theoretical aspects including mathematical modelling of algorithms, convergence analysis, and performance and complexity analysis of such algorithms. Applied AIS includes working on immune-inspired algorithms, and building immune-inspired computer systems, to apply AISs to diverse real-world applications. [28]

This application was made possible by combining concepts from different fields of work such as computational application development techniques like feature extraction, pattern recognition, memory, learning, classification, adaption for utilization in computer security, fraud detection, machine learning, data analysis, and optimization algorithms.

AIS algorithms provide help in the integration of AIS and Swarm Robotics by developing a very clear understanding of immune system structures and associated functions. At this moment, a clear understanding of principles and responses of the immune system is still required for application of AIS to Swarm Robotics, but an effort has been made to use the principles of the immune system to translate the knowledge into AIS applications. [29]

Application of swarm robotics in a wide range of real life fields cannot be ignored for serious consideration for further developments. These applications include biomedical applications for developing nano-systems by applying high-resolution monitoring, fast prototyping of micro-environments such as by application of micro-nano robots (e.g., for on-site drug delivery), and development of unconventional robots.

Furthermore, swarm robots can disperse and perform multiple tasks at difficult and inaccessible sites such as in forests, lakes, hilly areas, etc. Swarms of robots, because of the robustness of the swarm, can prove highly useful for dangerous tasks including monitoring and mitigating of environmental hazards, like a leakage of a chemical substance, and for clearing off of an environment from hazardous wastes.

b. KDD Cup 1999 Data

The KDD data set is a well-known benchmark in the research of Intrusion Detection techniques. The KDD data set can be used to find and compare the Detection Rate (DR) and False Alarm Rate (FAR) for an Intrusion Detection System (IDS) based on analysis of Basic, Content, Traffic and Host classes in which all data attributes can be categorized. [30]
The classes can be defined as:

1. Basic (B) Features are the attributes of individual TCP connections.
2. Content (C) features are the attributes within a connection suggested by the domain knowledge.
3. Traffic (T) features are the attributes computed using a two-second time window.
4. Host (H) features are the attributes designed to assess attacks which last for more than two seconds.

The NSL-KDD data set with 42 attributes is used for the empirical study. This data set has a number of versions available, out of which 20% of the training data is used which is identified as KDDTrain+_20Percent with a total number of 25192 instances. Different configurations of this data set are available with variation in number of instances, but the number of attributes in each case is 42. The attribute labeled 42 in the data set is the ‘class’ attribute which indicates whether a given instance is a normal connection instance or an attack. [31]

Fifteen sets of training and test data files were processed with the Random Tree algorithm in the Weka toolit. The results were analyzed to study dominance of each class of attributes in improving the Detection Rate (DR) and minimizing the False Alarm Rate (FAR).
In the future, this study can help increase the suitability of the data set so that higher DR can be achieved with minimum FAR. Hence, future work can lead to an improved data set that can be utilized for online intrusion detection.

f. Communication Network Concepts

What is Transmission Control Protocol (TCP)?

TCP allows for the exchange of streams of data. It is a connection-oriented protocol which means it needs to have a source and destination. Individual bytes of data such as from an application or session layer protocol are put in memory buffers and transmitted by TCP in transport Protocol Data Units commonly known as “segments”. TCP is a transport layer of the TCP/IP protocol suite and is an upper level layer to the network layer. This means that for the majority of networks 90% of their current traffic uses this transport service. Within the 90% is the WAN network of the smart grid [32].

What is a Wireshark and Packet Capture (pcap) file?

Wireshark includes a Graphical User Interface (GUI) which can analyze network protocols. By using Wireshark, you can browse and interact with packet data from a live network or even an already captured file. The format that Wireshark generally uses to capture files is the pcap format [29].

In order to determine if a pcap file is “acting in its normal behaviour” the TCP connection establishment needs to be studied. The DNP3 protocol sits on a serial bus connection or in this case the TCP/IP network in order to send the protocol messages between the client and server.

Transmission Control Protocol (TCP) connection is established by using three steps as explained in the *International Journal of Advanced Research in Computer Science and Software Engineering*: [33]

1) SYN bit from host A (client) to host B (server)
2) SYN+ACK bit from host B (server) to host A (client)
3) ACK bit from host A (client) to host B (server)

Figure 11 visualizes the data flowing in TCP network between the server and client machine which is explained above. See figure 11.

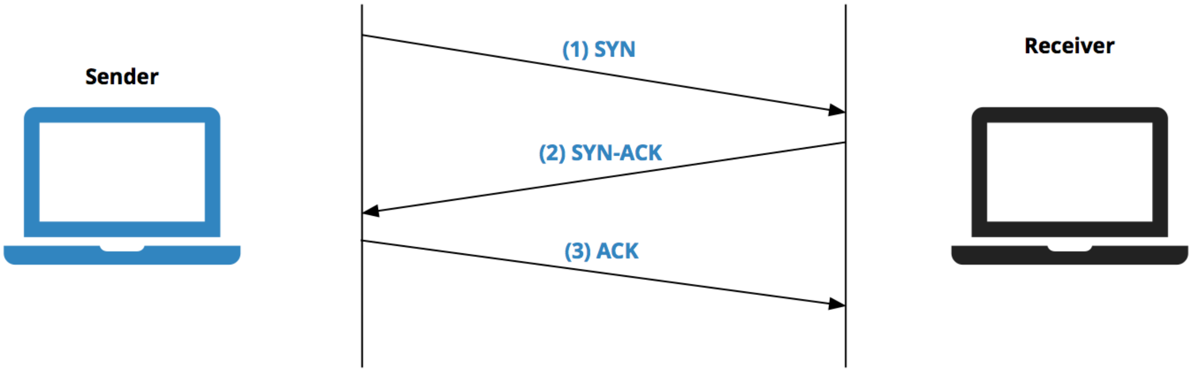
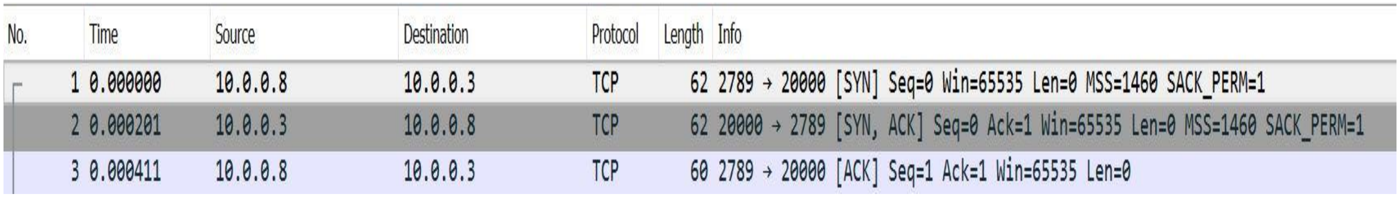


Figure 11: TCP “Handshake” connection

Whether there's an intrusion in the system or not based on these three steps. If any one of these three steps in the TCP “handshake” doesn’t occur, this shows that the connection isn’t there between the server and the client and there is an intrusion in the network [34].

Figure 12 is a screenshot of a TCP “handshake” from DNP3-TestDataPart1.pcap used [32].



No.	Time	Source	Destination	Protocol	Length	Info
1	0.000000	10.0.0.8	10.0.0.3	TCP	62	2789 → 20000 [SYN] Seq=0 Win=65535 Len=0 MSS=1460 SACK_PERM=1
2	0.000201	10.0.0.3	10.0.0.8	TCP	62	20000 → 2789 [SYN, ACK] Seq=0 Ack=1 Win=65535 Len=0 MSS=1460 SACK_PERM=1
3	0.000411	10.0.0.8	10.0.0.3	TCP	60	2789 → 20000 [ACK] Seq=1 Ack=1 Win=65535 Len=0

Figure 12: DNP3-TestDataPart1.pcap TCP "Handshake"

05. Initial Implementation

Our initial goal for implementation was to read the set of data from a pcap file and find data flowing patterns based on the detected complete data handshake among the TCP protocol. We imported pcap data to our parser.c file, and we got the necessary information reading the pcap data. We used a built-in function to get the information for each packet.

a. Limitations

There needs to be a huge set of data, and time to find a pattern of data flowing, to generate intrusion detection. Therefore, with limited time and data set, our initial goal for the implementation was to read the set of data and find the complete data handshake among the TCP protocol, which is the first step to generate intrusion detector. Given our understanding of reading pcap data in Wireshark, pcap data consists of a set of raw bytes. The type of bit flow can be detected by parsing the set of raw bytes. However, all the source code used build-in functions which returns the parsed data, not the raw bytes of imported pcap data. So, we could not find a solution to get and read the raw byte from our sources. And that preclude us from checking the type of bit. Since it is essential to check the type of bit to detect a complete handshake, our initial goal for the implementation was revealed to be impossible with the limited source. We could get the information of each packet with the built-in function. However, as we can read the packet's information with Wireshare easily and more clearly, we concluded that there is no use of parsing each part of pcap data without detecting a complete handshake using c programming language.

06. Requirements of Current implementation

The proof of concept of AIS based IDS on network data related to the WAN must:

- Incorporate AIS Negative Selection techniques and algorithms in design
- Be tested on test network data similar to the one the WAN uses
- Have the ability to catch intrusion on the test network data

07. Assumptions

We aren't able to obtain real-time smart grid AMI Wide area network (WAN) data. Since our proof of concept isn't an actual implementation of an AIS on real time smart grid networks there are many assumptions to consider about the data and implementation:

- We assume the self set pattern to be of the same size as an actual self set pattern in WAN networks.
- We assume that an intrusion or anomaly in the network is caused by an intrusion when it could be just a network fault error.
- We assume the laboratory environment of proof of implementation to be consistent with a real-world network but it is not. AIS could perform well in the laboratory environment but it might not function the same way for real networks.
- We assume the scale of the test network communications to relate to the real-world AMI WAN network communications, and also the amount of data/stress testing to be similar.
- We assume that the mappings of TCP "Handshake" parameters to negative selection terminology to be the correct mapping based on previous research.

08. Design

This design section will first explain an optimal design consisting of the necessary steps of implementing a NSA based AIS implementation. This design can be useful for researchers and industries that want to understand the basics of how to design an implementation for detecting anomalies in real time network traffic data.

The next part will describe the design of the implementation that was conducted as per the requirements for this specific project. But due to the limitations faced during this project development as stated in the sections above, the optimal design wasn't feasible. Instead the current design will follow a theoretically similar approach to the optimal general design. The current design is a scaled down approach and acts as a proof of concept using just the resources available to this project. Below will be an explanation of the optimal design as well as the design of the current implementation.

a. Optimal design of the implementation:

An intrusion detection system based on AIS is shown as the following steps [35]:

1. Collect a normal data set: The first step is to collect the parameters of the network data set when the system is acting in its "normal behavior". These parameters will be stored and used later on, in the negative selection process.

2. Generate detectors: Here the IDS detectors will be formed according to the negative selection algorithm and will be stored and kept in the system memory.

3. Real-time detection: Now the IDS detectors that were created will monitor a live network stream. The detectors and the live network data will then be compared in order to find any abnormalities in the network.

b. Current Design of the implementation

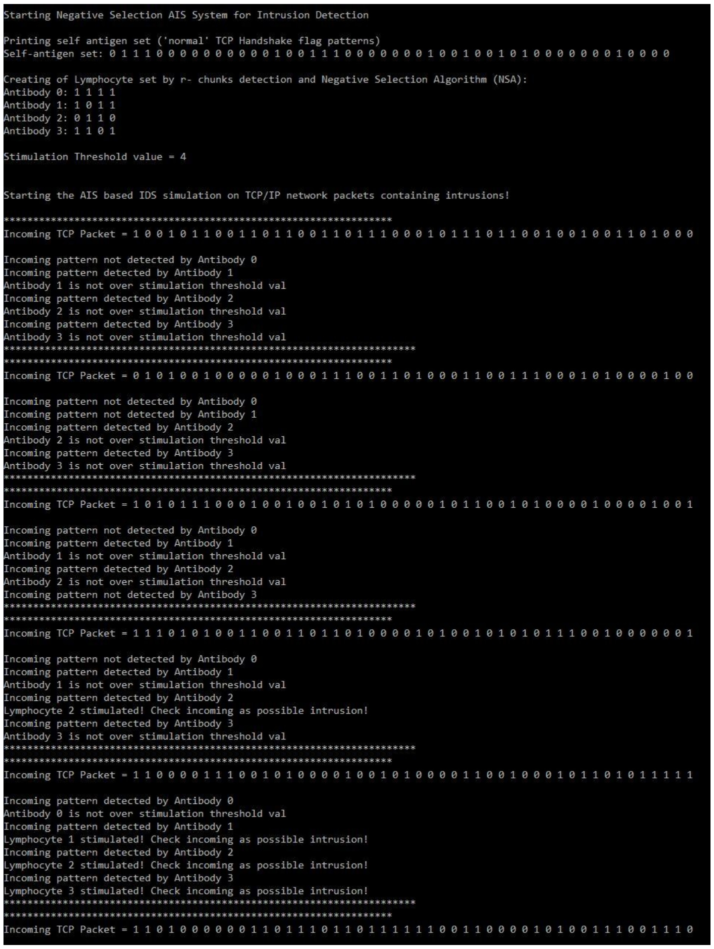
The current design follows all of the steps as the Optimal Design except step three, real-time detection. A live network wasn't feasible to create because of secure smart grid WAN networks. So for this step, the test data was human entered. This will be explained in the implementation section.

09. Implementation

The code was used from https://msdn.microsoft.com/en-us/magazine/msdnmag0113 and modified based on the research done by our team.

a. Objective of Implementation

This is a C# console application made on Visual Studios. This deliverable will not allow us to create a realistic AIS system, but it contains all the processes that are necessary in doing so such as implementation of Negative Selection Algorithm (NSA) and r-chunk bit matching. This implementation is a scaled down approach of a real world AIS to catch intrusions in real time network data. For the purpose of constructing an implementation in a timely and feasible manner, we will be using synthetic TCP/IP network data which will be user inputted instead of a real time network stream.



```
Starting Negative Selection AIS System for Intrusion Detection

Printing self antigen set ('normal' TCP Handshake flag patterns)
Self-antigen set: 0 1 1 1 0 0 0 0 0 0 0 0 0 0 1 0 0 1 1 1 0 0 0 0 0 0 0 1 0 0 1 0 0 1 0 1 0 0 0 0 0 0 0 1 0 0 0 0

Creating of Lymphocyte set by r- chunks detection and Negative Selection Algorithm (NSA):
Antibody 0: 1 1 1 1
Antibody 1: 1 0 1 1
Antibody 2: 0 1 1 0
Antibody 3: 1 1 0 1

Stimulation Threshold value = 4


Starting the AIS based IDS simulation on TCP/IP network packets containing intrusions!

******************************************************************
Incoming TCP Packet = 1 0 0 1 0 1 1 0 0 1 1 0 1 1 0 0 1 1 0 1 1 1 0 0 0 1 0 1 1 1 0 1 1 0 0 1 0 0 1 0 0 1 1 0 1 0 0 0

Incoming pattern not detected by Antibody 0
Incoming pattern detected by Antibody 1
Antibody 1 is not over stimulation threshold val
Incoming pattern detected by Antibody 2
Antibody 2 is not over stimulation threshold val
Incoming pattern detected by Antibody 3
Antibody 3 is not over stimulation threshold val
**********************************************************************
******************************************************************
Incoming TCP Packet = 0 1 0 1 0 0 1 0 0 0 0 0 1 0 0 0 1 1 1 0 0 1 1 0 1 0 0 0 1 1 0 0 1 1 1 0 0 0 1 0 1 0 0 0 0 1 0 0

Incoming pattern not detected by Antibody 0
Incoming pattern not detected by Antibody 1
Incoming pattern detected by Antibody 2
Antibody 2 is not over stimulation threshold val
Incoming pattern detected by Antibody 3
Antibody 3 is not over stimulation threshold val
**********************************************************************
******************************************************************
Incoming TCP Packet = 1 0 1 0 1 1 1 0 0 0 1 0 0 1 0 0 1 0 1 0 1 0 0 0 0 0 1 0 1 1 0 0 1 0 1 0 0 0 0 1 0 0 0 0 1 0 0 1

Incoming pattern not detected by Antibody 0
Incoming pattern detected by Antibody 1
Antibody 1 is not over stimulation threshold val
Incoming pattern detected by Antibody 2
Antibody 2 is not over stimulation threshold val
Incoming pattern not detected by Antibody 3
**********************************************************************
******************************************************************
Incoming TCP Packet = 1 1 1 0 1 0 1 0 0 1 1 0 0 1 1 0 1 1 0 1 0 0 0 0 1 0 1 0 0 1 0 1 0 1 0 1 1 1 0 0 1 0 0 0 0 0 0 1

Incoming pattern not detected by Antibody 0
Incoming pattern detected by Antibody 1
Antibody 1 is not over stimulation threshold val
Incoming pattern detected by Antibody 2
Lymphocyte 2 stimulated! Check incoming as possible intrusion!
Incoming pattern detected by Antibody 3
Antibody 3 is not over stimulation threshold val
**********************************************************************
******************************************************************
Incoming TCP Packet = 1 1 0 0 0 0 1 1 1 0 0 1 0 1 0 0 0 0 1 0 0 1 0 1 0 0 0 0 1 1 0 0 1 0 0 0 1 0 1 1 0 1 0 1 1 1 1 1

Incoming pattern detected by Antibody 0
Antibody 0 is not over stimulation threshold val
Incoming pattern detected by Antibody 1
Lymphocyte 1 stimulated! Check incoming as possible intrusion!
Incoming pattern detected by Antibody 2
Lymphocyte 2 stimulated! Check incoming as possible intrusion!
Incoming pattern detected by Antibody 3
Lymphocyte 3 stimulated! Check incoming as possible intrusion!
**********************************************************************
******************************************************************
Incoming TCP Packet = 1 1 0 1 0 0 0 0 0 0 1 1 0 1 1 1 0 1 1 0 1 1 1 1 1 1 0 0 1 1 0 0 0 0 1 0 1 0 0 1 1 1 0 0 1 1 1 0
```

Figure 13: Complete Console Application

Figure 13 is a screenshot of a console, which presents the output of the detecting application. The application reads the incoming TCP packet line by line, and shows the result of detecting process on the console.

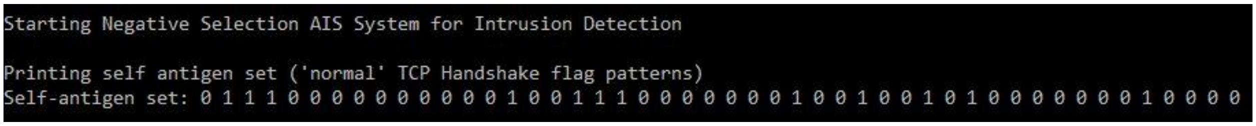
b. STEP 1: Creating the Self – Antigen Set

The human immune system contains two kinds of antigens, self-antigens and just antigens, which are considered harmful. The self-antigens which contain no anomalies are considered "normal".

In the beginning of the simulation the self-antigen is the "Normal TCP handshake" and is hard coded. The TCP handshake is explained further below.

TCP 3-way handshake:

Before understanding this implementation, we need to understand how a TCP 3-way handshake works. As a recap from the report look at page 22.



```
Starting Negative Selection AIS System for Intrusion Detection

Printing self antigen set ('normal' TCP Handshake flag patterns)
Self-antigen set: 0 1 1 1 0 0 0 0 0 0 0 0 0 0 1 0 0 1 1 1 0 0 0 0 0 0 0 1 0 0 1 0 0 1 0 1 0 0 0 0 0 0 0 1 0 0 0 0
```

Figure 14: The self-antigen set

The self-antigen set in Figure 14 represents the bytes in a normal TCP handshake. This set consists of 48 bits divided into 3 separate sections of 16 bits each, as shown below.

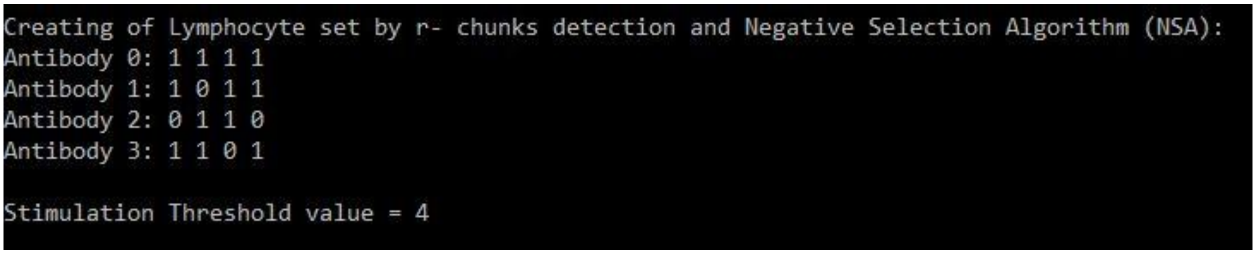
1. Syn flag: 0111000000000010
2. Syn, Ack flag: 0111000000010010
3. Ack flag: 0101000000010000

Putting all the bits together in consecutive order produces a correct complete handshake. This complete handshake is shown by the self-antigen set in Figure 3.

c. STEP 2: Creating the Lymphocyte set containing the Antibody Detectors

Antibodies are the receptors on a lymphocyte which detect harmful antigens. Therefore, a single lymphocyte can contain more than one antibody. In our case the simulation has a lymphocyte which contains 4 antibodies. Each antibody matches to an antigen. This antibody detection methodology is usually an approximation so only after a certain threshold of hits will the system trigger an attack status. In the simulation this threshold value is 4.

Due to our negative selection algorithm no antibody can detect a self-antigen. If this does occur, then that antibody will be deleted from the detector set. So, you can notice in the simulation the self-antigen set contains no pattern that follows any patterns of the antigens in the lymphocyte set. For example, the 4 consecutive 1 bits never occur in the self-set so they are made into an antibody. R-chunks or the Knuth-Morris-Pratt substring algorithm is basically a way of detection using literally “chunks” of patterns such as the antibodies as a subsection to check for matching. The algorithm takes in an input pattern such as 01011010101011101 and returns true if the current object’s antigen, such as 1101, matches the pattern. In this case it would match starting at index 3.



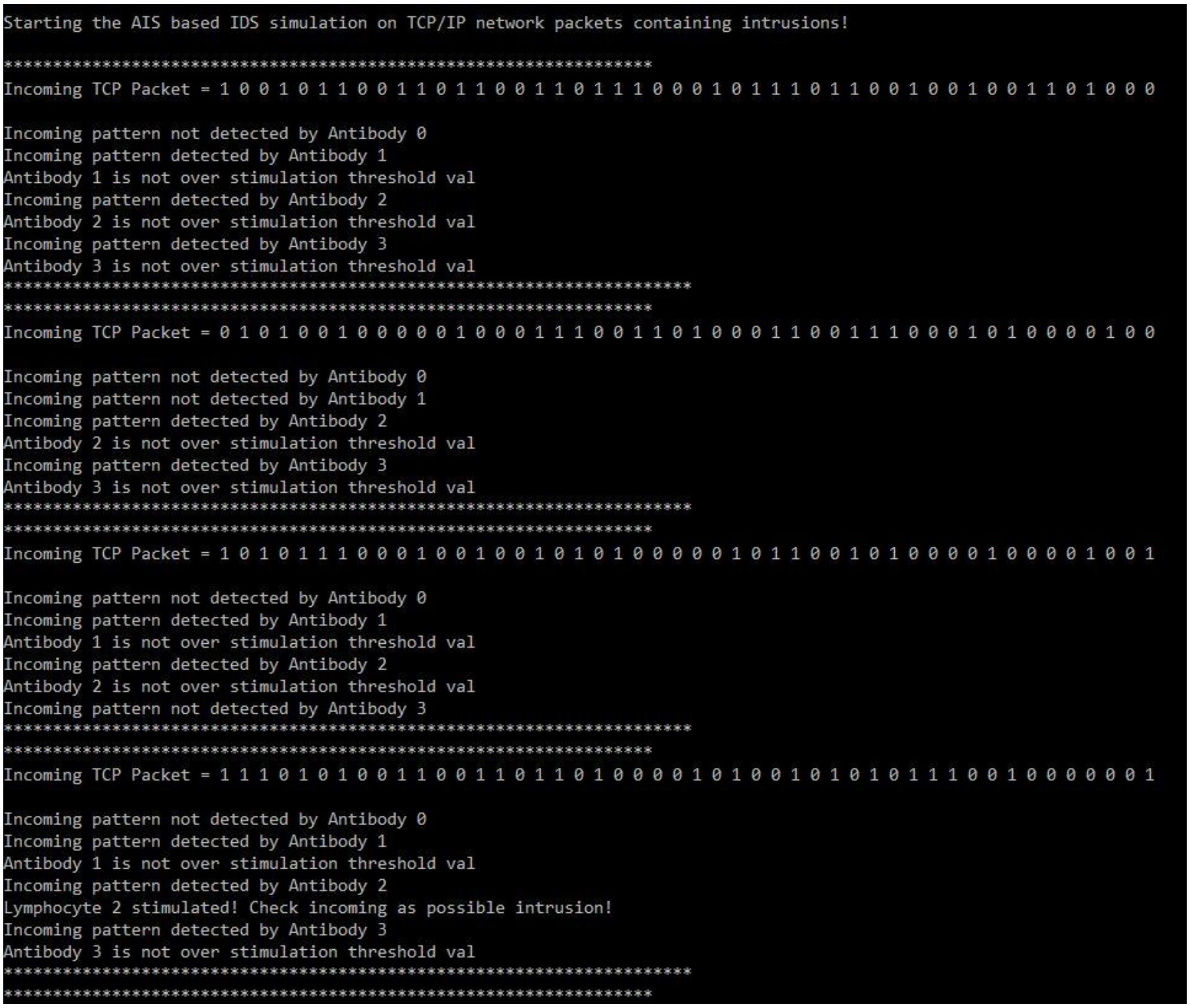
```
Creating of Lymphocyte set by r- chunks detection and Negative Selection Algorithm (NSA):
Antibody 0: 1 1 1 1
Antibody 1: 1 0 1 1
Antibody 2: 0 1 1 0
Antibody 3: 1 1 0 1

Stimulation Threshold value = 4
```

Figure 15: Lymphocyte set and Antibody creation

STEP 3: Detecting Intrusions on Incoming TCP Packets

When the incoming TCP packets containing harmful antigens were introduced in the simulation, the antibodies started to detect these harmful patterns in the packet bits. This is shown in Figure 16.



```
Starting the AIS based IDS simulation on TCP/IP network packets containing intrusions!

******************************************************************
Incoming TCP Packet = 1 0 0 1 0 1 1 0 0 1 1 0 1 1 0 0 1 1 0 1 1 1 0 0 0 1 0 1 1 1 0 1 1 0 0 1 0 0 1 0 0 1 1 0 1 0 0 0

Incoming pattern not detected by Antibody 0
Incoming pattern detected by Antibody 1
Antibody 1 is not over stimulation threshold val
Incoming pattern detected by Antibody 2
Antibody 2 is not over stimulation threshold val
Incoming pattern detected by Antibody 3
Antibody 3 is not over stimulation threshold val
**********************************************************************
******************************************************************
Incoming TCP Packet = 0 1 0 1 0 0 1 0 0 0 0 0 1 0 0 0 1 1 1 0 0 1 1 0 1 0 0 0 1 1 0 0 1 1 1 0 0 0 1 0 1 0 0 0 0 1 0 0

Incoming pattern not detected by Antibody 0
Incoming pattern not detected by Antibody 1
Incoming pattern detected by Antibody 2
Antibody 2 is not over stimulation threshold val
Incoming pattern detected by Antibody 3
Antibody 3 is not over stimulation threshold val
**********************************************************************
******************************************************************
Incoming TCP Packet = 1 0 1 0 1 1 1 0 0 0 1 0 0 1 0 0 1 0 1 0 1 0 0 0 0 0 1 0 1 1 0 0 1 0 1 0 0 0 0 1 0 0 0 0 1 0 0 1

Incoming pattern not detected by Antibody 0
Incoming pattern detected by Antibody 1
Antibody 1 is not over stimulation threshold val
Incoming pattern detected by Antibody 2
Antibody 2 is not over stimulation threshold val
Incoming pattern not detected by Antibody 3
**********************************************************************
******************************************************************
Incoming TCP Packet = 1 1 1 0 1 0 1 0 0 1 1 0 0 1 1 0 1 1 0 1 0 0 0 0 1 0 1 0 0 1 0 1 0 1 0 1 1 1 0 0 1 0 0 0 0 0 0 1

Incoming pattern not detected by Antibody 0
Incoming pattern detected by Antibody 1
Antibody 1 is not over stimulation threshold val
Incoming pattern detected by Antibody 2
Lymphocyte 2 stimulated! Check incoming as possible intrusion!
Incoming pattern detected by Antibody 3
Antibody 3 is not over stimulation threshold val
**********************************************************************
******************************************************************
```

Figure 16: Incoming TCP packet detection

The first incoming TCP packet was detected by antibody 2 but didn’t stimulate an attack status until the fifth packet came along. This is because antibody 2 hit the threshold limit (which is 4) at that iteration.

10. Results

Our test will be deemed as a success if our Negative Selection AIS algorithm is able to catch a network intrusion in the test network data used.

This implementation is a proof of concept. Since this implementation runs and successfully catches intrusions it shows that, if this AIS implementation were to be scaled up and run on networks like the WAN in smart grid AMI (which uses TCP/IP network pcap data), it would also be able to catch intrusions in the smart grid.

This could be a first chapter to having a robust system which catches attacks within a huge network from within the network itself.

11. Future work

The future of the artificial immune system (AIS) approach looks positive as the scholars and researchers share knowledge and information to move forward together as a computer science community. Continuous work is being done in this field to come up with better algorithms and improve the efficiency of the immune systems. A high false detection rate is also a problem that the researchers are trying to find a solution for, as it leads to low work functionality. Another area of possible future research is the use of Danger algorithms and their applications. Some researchers are coming up with a Dendritic cell concept in the immune systems. In this concept, one of the devices in the immune system is accustomed to a virus. The immune system is provided with a learning curve to update all the devices if any one of the devices in the system is attacked. This way all the devices are secured for that particular type of network attack. [36]

12. Lessons learned

a. Timeline/schedule

Weekly Status updates	Thursdays from 2pm-3pm
Milestone 1	Presentation: Introduce our project and the research we will conduct Due date: Feb 6th
Milestone 2	Finish all background research on topic and get started pulling data Due date: Feb 20th
Milestone 3	Start analyzing data input for AI. Due date: March 1st
Milestone 4	Create the test and training data set for the AI technology. Due date: March 29th
Milestone 5	Rough draft approval. Due date: April 12th
Milestone 6	Final report approval. Due date: April 26th
Exams Start	3rd May, 2018

Table 2: Timeline

b. Problems

As artificial intelligence in cyber security area is very new and advanced, it was our struggle to find appropriate resources for our project. Especially, we had a difficult time to find the data sets which are used in an enterprise. Also, in implementation, we had to change our goal in the process of making cogent output with limited sources and data sets. The number of packets and data flowing in between server and client machine were not enough to compute accurate data flowing patterns.

c. Solution

We felt limited regarding producing compelling output given the shortage of sources and data throughout the implementation. Also, we had limited time to understand the source codes and implement what we were planning with the limited sources. Since we were not able to parse the given pcap file and produce the data pattern using c programming language, we had to change our goal for implementation. Our solution to this was to create a proof of concept implementation which showcased all the processes of an AIS system catching intrusions on network data. We believe that this implementation, if it were to be scaled up using algorithms such as Negative Selection and Knuth Morris Pratt, could be run on real smart grid networks such as the wide area network (WAN).

13. Acknowledgements

This project was supported by Morris-King, James R. and Wynne, Devin M. from the Mitre Corporation. We'd especially like to thank James for all of his help, insight, and guidance, without which we would not have been able to accomplish the work we completed this semester.

We also want to thank our professor, Dr. Fox, for helping us throughout the milestones of this project, and all of his hard work to help guide us with this research report.

14. References

[1] Guidelines for Smart Grid Cybersecurity. (2014, September). *The Smart Grid Interoperability Panel – Smart Grid Cybersecurity Committee,1*. Retrieved April 5, 2018, from https://nvlpubs.nist.gov/nistpubs/ir/2014/NIST.IR.7628r1.pdf.

[2] What is the Smart Grid? (n.d.). Retrieved April 18, 2018, from https://www.smartgrid.gov/the_smart_grid/smart_grid.html

[3] Advanced Metering Infrastructure (AMI). (n.d.). *AMI*. Retrieved April 23, 2018, from https://www.ferc.gov/CalendarFiles/20070423091846-EPRI Advanced Metering.pdf.

[4] Shaw, K. (2018, January 18). Https://www.lifewire.com/wide-area-network-816383. Retrieved March 20, 2018, from https://www.networkworld.com/article/3248989/lan-wan/wide-area-networks-what-wans-are-and-where-theyre-headed.html

[5] Advanced Metering Infrastructure Architecture and Components. (2013, February 28). Retrieved April 20, 2018, from https://blog.compass-security.com/2013/02/advanced-metering-infrastructure-architecture-and-components/

[6] Mesawriter B. (2013, March 29). HAN, WAN, FAN, NAN, PAN, LAN. Retrieved April 19, 2018, from https://powerprimer.wordpress.com/2013/03/29/han-wan-fan-nan-pan-lan/

[7] THOMAS, J. (2016, August 25). The Advantages and Disadvantages of WANs. Retrieved April 11, 2018, from https://purple.ai/blogs/advantages-disadvantages-wans/

[8] Woodcock, JoAnne. (1999). *WAN Technologies*. Microsoft Press. Retrieved April 10, 2018, from https://technet.microsoft.com/en-us/library/bb962087.aspx

[9] M. (2012, July 18). Common Types of Network Attacks. Retrieved April 20, 2018, from https://docs.microsoft.com/en-us/previous-versions/windows/it-pro/windows-2000-server/cc959354(v=technet.10)

[10] Routledge. (2016, May 27). *Survival 56.2: Global Politics and Strategy* (1st ed.). The International Institute for Strategic Studies.

[11] Northcutt, S. (2010, January 10). Hash Functions. Retrieved March 23, 2018, from https://www.sans.edu/cyber-research/security-laboratory/article/hash-functions

[12] What is an Intrusion Detection System (IDS)? - Definition from Techopedia. (n.d.). Retrieved April 20, 2018, from https://www.techopedia.com/definition/3988/intrusion-detection-system-ids

[13] Rouse, Margret. What is intrusion detection system (IDS)? - Definition from WhatIs.com. (2018, January). Retrieved March 21, 2018, from https://searchsecurity.techtarget.com/definition/intrusion-detection-system

[14] Figure 3 from: Nam, B. V. (2016, November) KHCN. Retrieved March 22, 2018, from http://khcn.cinet.vn/articledetail.aspx?articleid=1867&sitepageid=452

[15] Casey, B. (2012, April 09). Snort and the Value of Detecting the Undetectable. Retrieved March 24, 2018, from https://www.techopedia.com/2/28294/security/snort-and-the-value-of-detecting-the-undetectable

[16] Daudi, J. (2015, February 14). An Overview of Application of Artificial Immune System in Swarm Robotic Systems. *Artificial Immune Systems,4*(1), 1000127th ser., 2. Retrieved March 18, 2018, from https://www.omicsonline.org/open-access/an-overview-of-application-of-artificial-immune-system-in-swarm-roboticsystems-2168-9695-1000127.php?aid=53117.

[17] Gordon, Dan. (2006, January). The Dana Sourcebook of Immunology: Resources for Secondary and Post-Secondary Teachers and Students, Retrieved from http://www.dana.org/Publications/ReportDetails.aspx?id=44170

[18] Singh1, S., & Singh2, J. (2013, June). A Review: AIS Based Intrusion Detection System. *Adapted IDS Model,2*(6), 2. Retrieved March 18, 2018, from https://www.ijarcce.com/upload/2013/june/12-SandeepSingh-AReviewAISBasedIntrusionDetectionSystem.pdf

[19] Figure 1 from: Yang, H. (2014, March 23). A Survey of Artificial Immune System Based Intrusion Detection. *2014*, 3. Retrieved March 18, 2018, from https://www.hindawi.com/journals/tswj/2014/156790/

[20] Shen, J. (n.d.). *Network Intrusion Detection By Artificial Immune System*. RMIT UNIVERSITY.

[21] Figure 2f from: Irimia, E. Ramona., & Gottschling, Marc. (2016) Taxonomic revision of Rochefortia Sw. (Ehretiaceae, Boraginales). Biodiversity Data Journal 4: E7720. Retrieved March 14, 2018, from https://doi.org/10.3897/BDJ.4.e7720.(n.d.).doi:10.3897/bdj.4.e7720

[22] Irimia, E. Ramona., & Gottschling, Marc. (2016) Taxonomic revision of Rochefortia Sw. (Ehretiaceae, Boraginales). Biodiversity Data Journal 4: E7720. Retrieved March 13, 2018, from https://doi.org/10.3897/BDJ.4.e7720.doi:10.3897/bdj.4.e7720

[23] Greensmith, J. (2014, July 15). Artificial Immune Systems - Computerphile. Retrieved from https://www.youtube.com/watch?v=u2qRUtg2k3Y

[24] Llama. (n.d.). Introduction to String Searching Algorithms – topcoder. Retrieved April 20, 2018, from https://www.topcoder.com/community/data-science/data-science-tutorials/introduction-to-string-searching-algorithms/

[25] ICS 161: Design and Analysis of Algorithms Lecture notes for February 27, 1996. (1996, February 27). Retrieved March 21, 2018, from https://www.ics.uci.edu/~eppstein/161/960227.html

[26] Darktrace Antigena Named 'Best AI Product in Security'. (2017, June 21). Retrieved April 18, 2018, from https://www.darktrace.com/press/2017/174/

[27] Symantec. (1999, May 14). Symantec unveils digital immune system strategy for unprecedented level of managed, intelligent protection and control. Retrieved March 3, 2017, from https://www.itweb.co.za/content/JBwEr7n5GdLv6Db2

[28] J, Daudi. (2015, February 14). An Overview of Application of Artificial Immune System in Swarm Robotic Systems. Retrieved March 18, 2018, from https://www.omicsonline.org/open-access/an-overview-of-application-of-artificial-immune-system-in-swarm-roboticsystems-2168-9695-1000127.php?aid=53117

[29] Aggarwal, Preeti., & Sharma, K. Sudhir. (2015, August 21). Analysis of KDD Dataset Attributes - Class wise for Intrusion Detection. Retrieved March 18, 2018, from https://www.sciencedirect.com/science/article/pii/S1877050915020190

[30] Tavallaee, Mahbod., Bagheri, Ebrahim., Lu, Wei., & Ghorbani, A. Ali. (2009, July). A detailed analysis of the KDD CUP 99 data set. Retrieved April 3, 2018, from https://dl.acm.org/citation.cfm?id=1736489

[31] Fairhust, Gorry. (2003, December 17). Transmission Control Protocol (TCP). Retrieved April 23, 2018, from http://www.erg.abdn.ac.uk/users/gorry/course/inet-pages/tcp.html

[32] Combs, Gerald. (n.d.). Wireshark. Retrieved March 13, 2018, from https://www.wireshark.org/docs/wsug_html/

[33] InetDaemon. (2013, September 01). TCP 3-way Handshake (SYM, SYN-ACK, ACK). *INETDAEMON.COM*. Retrieved March 18, 2018, from http://www.inetdaemon.com/tutorials/internet/tcp/3-way_handshake.shtml

[34] Automayt. (2016, June 26). Automayt/ICS-pcap. GitHub. Retrieved April 14, 2018, from https://github.com/automayt/ICS-pcap

[35] Kim, J., Bentley, P., Aickelin, U., Greensmith, J., & G. (n.d.). Immune System Approaches to Intrusion Detection. Retrieved March 26, 2018, from https://arxiv.org/ftp/arxiv/papers/0804/0804.1266.pdf.

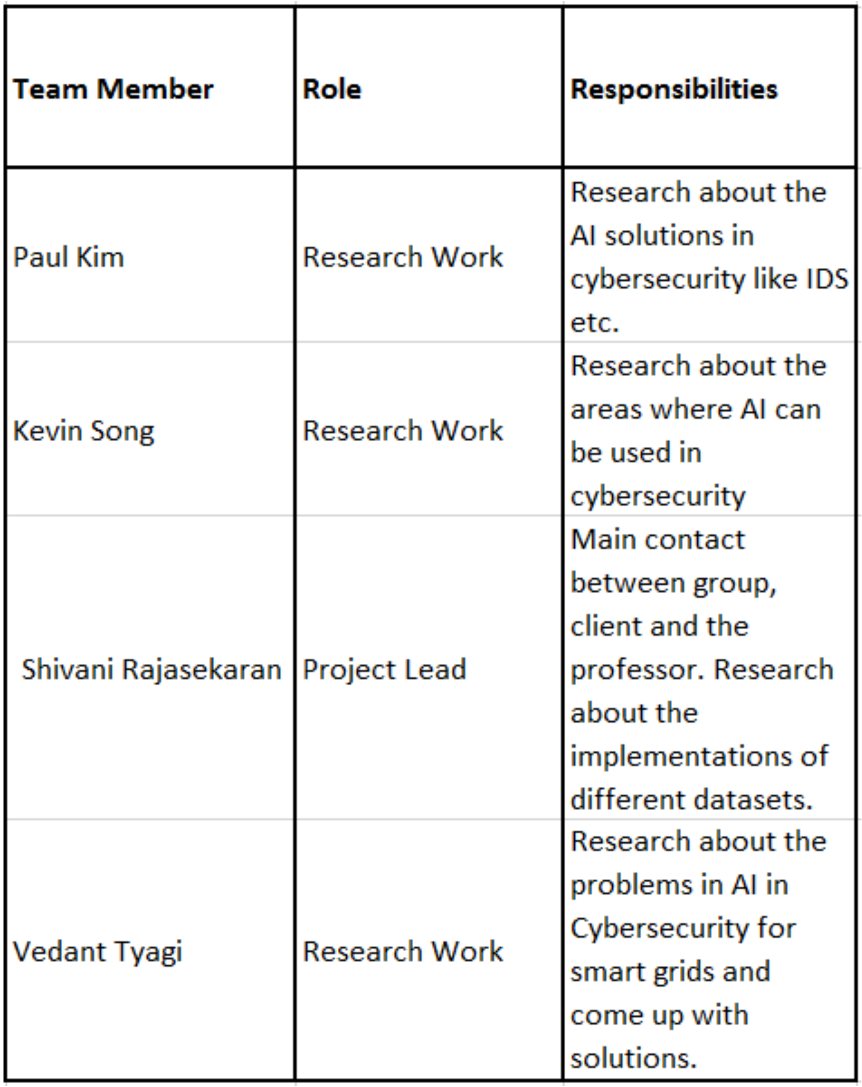
[36] Greensmith, J., Twycross, J., & Aickelin, U. (2006, July). *Dendritic Cells for Anomaly Detection*. Vancouver, BC, Canada: IEEE. doi:10.1109/CEC.2006.1688374

[37] Faisal, M. A., Aung, Z., Williams, J. R., & Sanchez, A. (2012). Securing Advanced Metering Infrastructure Using Intrusion Detection System with Data Stream Mining. *Intelligence and Security Informatics Lecture Notes in Computer Science,*96-111. doi:10.1007/978-3-642-30428-6_8

[38] Wolfe, Franklin. (2017, August 28). How Artificial Intelligence Will Revolutionize the Energy Industry. Retrieved April 23, 2018, from http://sitn.hms.harvard.edu/flash/2017/artificial-intelligence-will-revolutionize-energy-industry/

15. Appendices

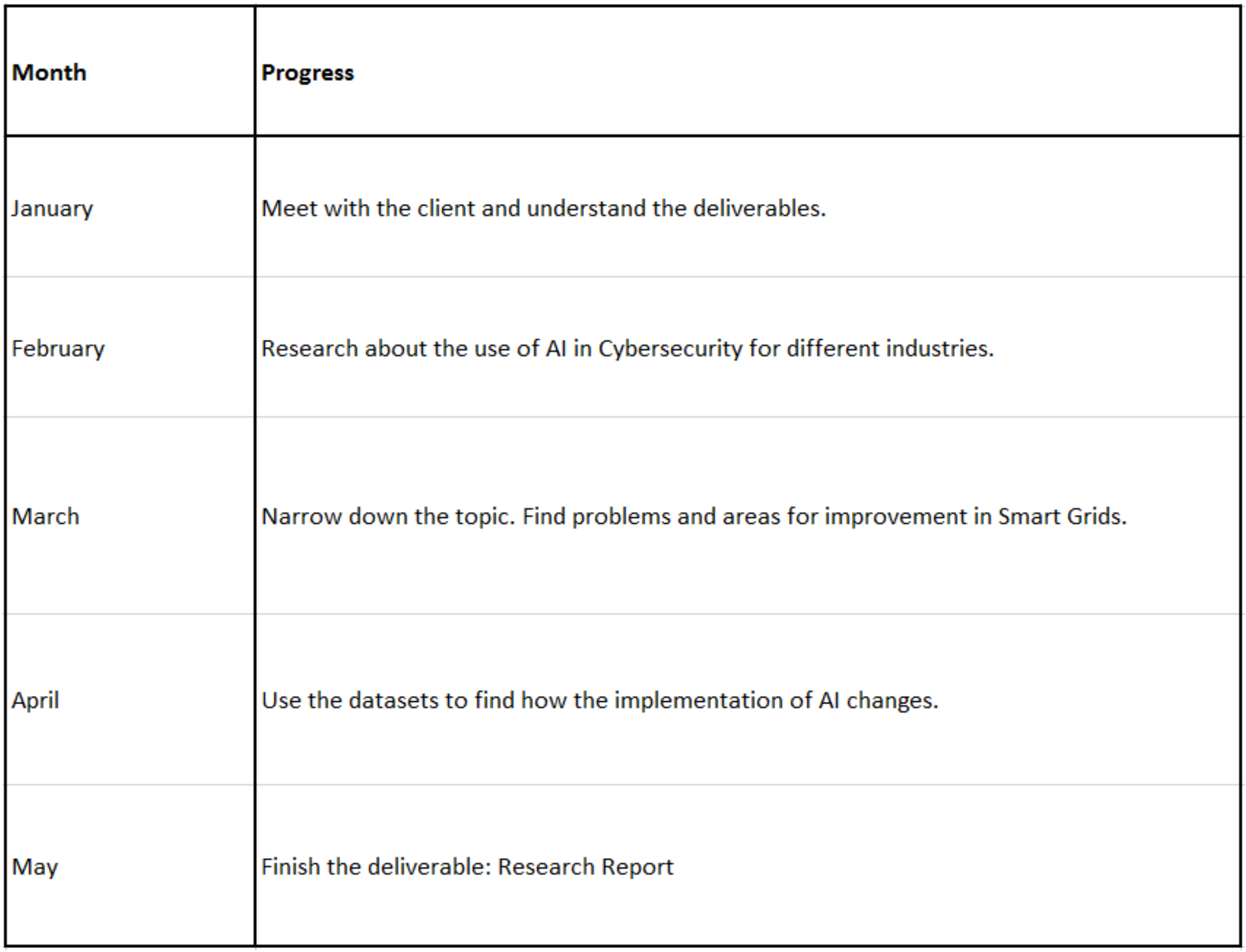
Appendix A



Team Member	Role	Responsibilities
Paul Kim	Research Work	Research about the AI solutions in cybersecurity like IDS etc.
Kevin Song	Research Work	Research about the areas where AI can be used in cybersecurity
Shivani Rajasekaran	Project Lead	Main contact between group, client and the professor. Research about the implementations of different datasets.
Vedant Tyagi	Research Work	Research about the problems in AI in Cybersecurity for smart grids and come up with solutions.

Table 3: Team Member Roles and Responsibilities

Appendix B



Month	Progress
January	Meet with the client and understand the deliverables.
February	Research about the use of AI in Cybersecurity for different industries.
March	Narrow down the topic. Find problems and areas for improvement in Smart Grids.
April	Use the datasets to find how the implementation of AI changes.
May	Finish the deliverable: Research Report

Table 4: Team Progress

Appendix C: Background of AI in Energy Based on our Approval Document

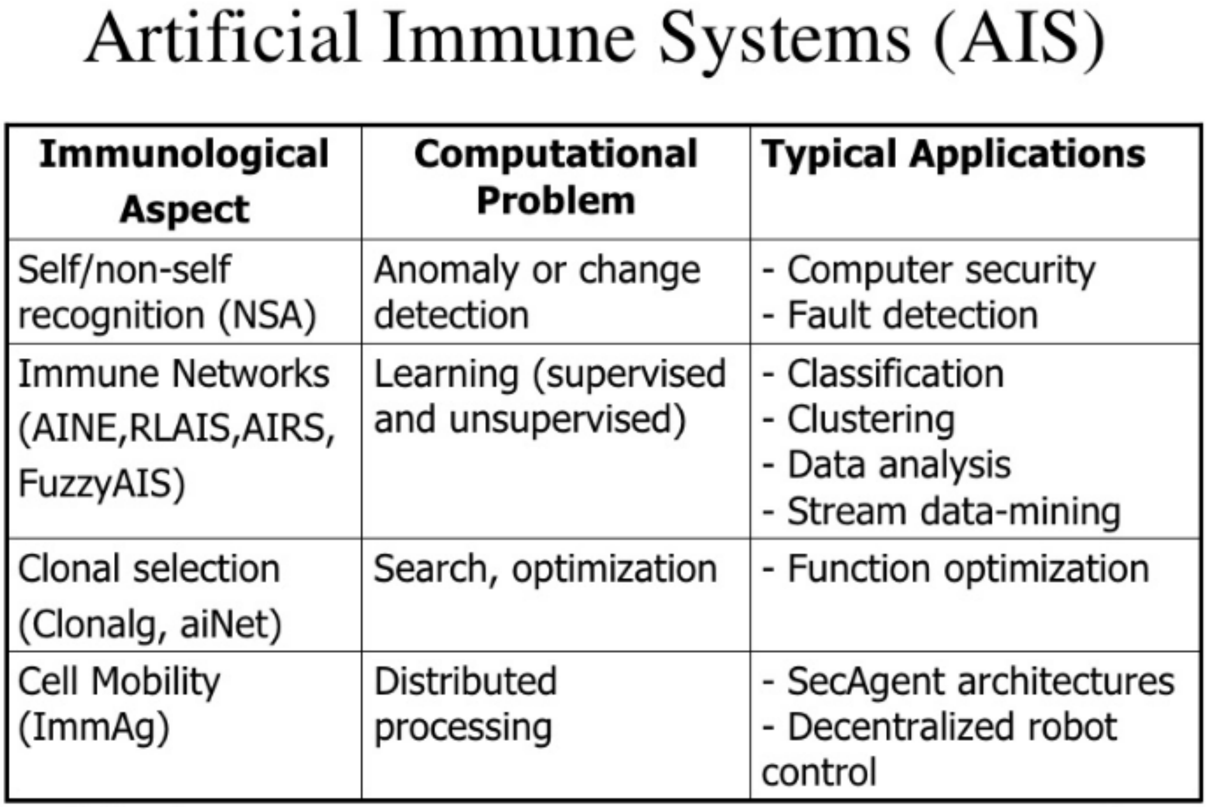
AI in Energy

The U.S contains 5,800 power plants and more than 2.7 million miles of power lines. The average age of these power plants exceeds 30 years. The transmission system is rusting away, and this has been the main cause of the 2003 Northeast blackout which was a huge disaster in U.S history. Another issue is that while the distributed energy resources created from renewable energy increases, the idea of supply and demand gets complicated. This increase in different energy renewable sources makes utility companies buy more than needed energy from private users. The private users then make a surplus amount of energy, more than they use. They then send the excess energy to the grid. Whenever this process occurs, demand outruns supply. Therefore, the utility companies turn back on the fossil fuel plants. This process is both wasteful and costly for these companies. This also leads to a higher emission of greenhouse gas into the air and costlier electricity bills for customers.

In order to find a solution to this issue the U.S. Department of Energy (DOE) requires that the smart grid should become a U.S policy. AI is the main technology that will control the smart grid. What this technology will do is use smart sensors to make decisions in a timely manner on how and when to better allocate energy resources. The AI technology will do this by using 'deep learning' algorithms and collecting and synthesizing large amounts of data from these sensors all over the United States. The technology will have to learn on its own by differentiating the norm vs. anomalies in these datasets.

One major issue with the smart grid is the rise of the use of Information Technology. This technology relies on the internet which has become a large contributor to the emission of greenhouse gas. In order to start the smart grid and process the vast amount of data, we will need more machines and computing power. This is a problem that definitely needs to be a focus for the AI industry working with the smart grids [38].

Appendix D



Artificial Immune Systems (AIS)

Immunological Aspect	Computational Problem	Typical Applications
Self/non-self recognition (NSA)	Anomaly or change detection	- Computer security - Fault detection
Immune Networks (AINE,RLAIS,AIRS, FuzzyAIS)	Learning (supervised and unsupervised)	- Classification - Clustering - Data analysis - Stream data-mining
Clonal selection (Clonalg, aiNet)	Search, optimization	- Function optimization
Cell Mobility (ImmAg)	Distributed processing	- SecAgent architectures - Decentralized robot control

Table 5: Information on AIS