REVIEW OF INTERNAL CYBER ATTACKS IN NUCLEAR FACILITIES AND AN ARTIFICIAL NEURAL NETWORK MODEL FOR IMPLEMENTING INTERNAL CYBERFORENSICS

by

Brandyn M. CAMPOS and Miltiadis ALAMANIOTIS*

Department of Electrical and Computer Engineering,
The University of Texas at San Antonio, San Antonio, Tex., USA

Scientific paper
https://doi.org/10.2298/NTRP2102128C

Deployment of digital technologies within a modern shift in cyber defense systems is essential for protecting the energy production units. One of the important components of defense is cyberforensics: once an attack has been detected to locate its origin. In this paper, a review of well-known cyberattacks in nuclear facilities is provided, with the lessons learned leading to the development of a machine learning approach implementing identification of internal attacks in the facility's data networks. Our approach may be seen as one of the layers in a defense-in-depth strategy that identifies if the attack comes from inside, which may result in identifying faster the attacker's origin. The presented model exploits network packet examination to cast accurate predictions on detailing the origin of malicious network connections. The approach fuses multiple mathematical functions within an artificial neural network to provide a response in the form of 0/1, *i. e.*, whether the attack is identified as internal or not. The utilization of a variety of test cases is developed to explore the relevance and validity of the predictive approach. The proposed implementation is examined with network data packet variance, and the results obtained exhibit a highly accurate detection rate.

Keywords: cyberforensics, digital forensics, nuclear power plant, internal attack, neural network

INTRODUCTION

Defending the nuclear energy facilities against any type of malicious attacks is a critical issue to secure a sustainable future energy supply and retain the feeling of safety to the public. Among many types of attacks, cyberattacks are of special interest due to the ubiquitous deployment of digital connectivity and information technologies. Concerning nuclear facilities, such as nuclear power plants (NPP), cyberattacks are becoming more frequent as instrumentation and control (I&C) systems are being transitioned from analogous to digital technology.

The design of cyberforensic defense methods is exceedingly more important as the number of attacks against nuclear facilities increases in occurrence [1]. The nuclear defense systems within all NPP could be an extremely serious condition and the remedies are very important in conducting safe plant operations [2].

The research field of cyberforensics, more commonly known as digital forensics, for nuclear facilities, is imperative due to the critical nature of the infrastructures being protected. As technology continues to advance, the attack strategies against NPP are also becoming more sophisticated [3]. Therefore, the design of an intelligent cyber defense system to actively defend against all cyber threats (known and unknown) and enhance the structure of defense-in-depth, is crucial for the overall nuclear facility operation and security [4].

The US Department of Homeland Security has recently reported that cyber hackers have been able to exploit insecure vendors within the power industry to gain access to privileged, air-gapped systems inside America's electric utilities [5]. It is assumed that air-gap measures are a sufficient defensive measure against malicious intrusions inside critical systems. The vulnerabilities for any given NPP exist on the networks of the vendors who supply their technologies. If one of the vendor's systems has weak points that hackers utilize to develop attack plans, then the entire nuclear infrastructure has a higher probability of being attacked internally. As hackers begin to identify the vendors for US utilities, the more likely an intrusion into critical systems is to happen. When a vendor is identified, conventional tools including spear-phishing are used to enter the computer networks. Once inside the vendor's computer network, hackers will steal critical information such as secret credentials and access to utility networks

* Corresponding author; e-mail: miltos.alamaniotis@utsa.edu

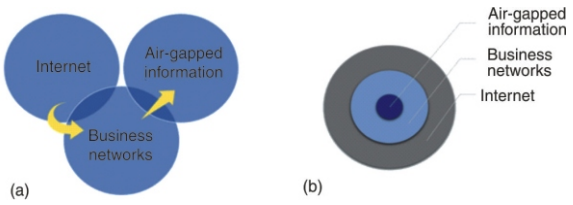

Figure 1. (a) Flow of information within the networks of a NPP, and (b) interfaces between the networks

and operations. The physical representation of network information architecture (information flow and interfaces) is detailed in fig. 1.

The cyberspace and information that exists between the NPP inner and business networks have major vulnerabilities that have been exploited by cyber attackers to develop specific cyberattack plans. The information that exists about the NPP critical control systems can be reached through attacking their business networks and probing these networks for available information about the NPP [5]. The information available on these networks is crucial for implementing an internal cyberattack. If we consider that many of the network servers being utilized within the NPP are air-gapped, meaning that they are not connected to any external networks, then it becomes apparent in which areas, if any, this information can be located. The data being collected includes but is not restricted to the Programmable Logic Controller (PLC) data such as serial numbers, version numbers, timing executions, and actively running logic code [4]. This information has been used in the past to conduct highly invasive internal cyberattacks against NPP. More importantly, is determining whether the Linux operating systems are conforming to the demands of the standards IEC 61513:2011, ISO/IEC 23360-1:2006, and ISO/IEC 27032:2012 [6].

A considerable method for developing an attack of this nature is by probing critical system information from the vendor networks and deploying sophisticated, customized internal cyberattacks that match the exact PLC configurations for a given NPP [7]. There exists the opportunity to develop an improved, co-ordinated defense system that can identify whether probes have originated internally on the network as well as flag an anomalous connection packet on the business networks. Cybersecurity assessment results of industrial I&C systems used in NPP are mainly based on the assessment of expert judgment and do not take into account all features of FPGA (*i. e.*, Field-Programmable Gate Arrays) technologies [8] adopted in the implementation of I&C systems [9]. When more FPGA features are being assessed, there is a higher probability that an attack can be identified and flagged before any damages can be conducted.

The key challenges in digital forensics are now being mitigated by combining artificial intelligence (AI) tools and cyber systems. Another research method of NPP surveillance includes utilizing electroencephalogram (EEG) signals analysis of plant workers to describe the possibility of identifying a non-initiating insider who has not yet committed an action [10]. The EEG are a method for monitoring electromagnetic signals in the brain, and with this data, an algorithm can detect irregularities in signals. The next paradigm in creating advanced cyberforensics systems includes integrating AI with our current cyber defense system models [3]. The capability of an AI system to interpret real-time data flow on a network will increase the defense system's ability to detect malicious connections to the business networks and add one more layer in the defense-in-depth of information systems in NPP [11-13]. Given that intrusions are co-ordinated and unique, there is the prospect to implement an AI digital forensics system that analyzes network packages in real-time and can differentiate between malicious behavior and normal activity on the network. The information load across the network includes high traffic and high-speed connections. The data flow of the network is currently being surveyed by automated defense systems that require human intervention for surveying networks, but even these are not optimized for high traffic volumes [14]. Therefore, it is beneficial for newer cyber defense systems to develop an intelligent system that compliments the security system by analyzing high-speed data flow across all business networks.

The important challenges currently present in cyberforensics are being addressed by utilizing programming tools such as AI. Specifically, the two methods of cyber defenses being analyzed are distributed denial of service (DDoS) attacks and intrusion detection and intrusion prevention systems (IDPS) [15, 16]. The DDoS cyberattacks' main goal is to overload a server by coordinating infected computers (botnet) to directly take up the computational resources of the server. Generally, DDoS attacks occur to produce weaknesses in cyber architecture for future attacks and to cause damages in the form of economic and resource disruption. In particular, in IDPS technology, one of the main challenges is setting up and configuring IDPS tools to correctly identify unauthorized access to a network. Currently, IDPS generates many false alerts or *false positives* [3]. The methodology is as such; a database is provided with key malicious signatures that are compared to all network connections being made in the data flow of the network. Once one of the key signatures is detected, the system will notify the users of the malicious connection [3]. Co-ordinating a defense system with databased key detection also provides certain vulnerabilities, mainly internet protocol (IP) masking and subnet masking. Notably, malicious connections to the network may not be flagged due to being masked with a key signature from the acceptable connections database. Also, there are anomaly-based detection systems that assess the behavior of the network and predict

anomalous connections and determine whether they are a threat to the network. Utilizing machine learning with anomaly-based detection systems provides an efficient method for defending NPP networks [17].

In this paper, a new intelligent forensic model for identifying internal cyber threats to NPP networks is presented and discusses the forensics model for optimizing the identification method for cyberattacks on the NPP business networks. The proposed forensic model utilizes an artificial neural network as the mean of analyzing network packet data and distinguishing internal malicious connections for network safety. The contribution of this paper contains:

– reviewing of the world around internal cyberattacks conducted against nuclear facilities,
– development and application of neural network model for analyzing network traffic data,
– a model for utilizing network data to predict malicious connections that come from inside the nuclear facility (internal attacks in NPP), and
– testing of the model on real-time network data sets.

The aforementioned processes have laid the foundation for the development of an AI model implementation that provides an optimized solution for identifying malicious internal connections that are dangerous to NPP networks. The concept of classification of anomalous behavior has been previously applied to identifying key features of data and characterizing the data into labeled categories. Thus, we propose that a cyberforensics model implemented by an AI system will increase active defenses and post-attack analysis by identifying internal network connections attacks.

The roadmap of this paper is as follows. In the next two sections, a review of known internal cyberattacks at nuclear power plants is given, while the developed methodology is presented. Following is a section where the methodology is tested, and the obtained results are discussed. The last section concludes the paper by emphasizing its main points.

SURVEY OF INTERNAL CYBERATTACKS ON NUCLEAR FACILITIES

In this section, we review the origins of critical internal cyberattacks conducted against NPP and move to present a brief examination of the STUXNET industrial attack on Iranian Nuclear Facilities. The goal is to explain the effectiveness of the attackers' preferred methodology for conducting an internal cyberattack [17]. It is important to observe the specific methods utilized by the attackers to obtain supervisory control and data acquisition (SCADA) information from computer systems that are air-gapped within nuclear facilities [18]. Establishing, implementing, and maintaining the cyber security program is accomplished by using formal assessment methods conducted by qualified staff at nuclear power reactors acting under the authority of the site's approved policy and the supervision of senior site management [19]. Therefore, critically reviewing the mitigation plan utilized by the facility will be instrumental in the development of more robust cyberforensics systems for nuclear power reactors. This survey of internal cyberattacks will deliver signature information on the lessons learned from previous cyberattacks and enable the ability to develop a digital forensics methodology for mitigating any future cyber threats.

Ignalina

Ignalina Power Plant is the first known attempt of an internal cyberattack against a nuclear facility. The control system network was affected directly by an internal injection of malicious content with the intent to harm the facility. The methodology for the attack included distributing the hazardous software through the insertion of a universal serial bus (USB) drive onto a command control system computer. Delivery of payload by USB is a considerably efficient method for injecting malicious software onto targeted computer systems. Even where nuclear facilities are air-gapped, this safeguard can be breached with nothing more than a flash drive [20]. Critical system sabotage, as well as substantial economic losses for the Ignalina nuclear facility, were the immediate results of the cyberattack. The development of newer technologies within the nuclear field has standardized the implementation of segmenting the control networks to slow the effects of an internal injection.

Stuxnet

The STUXNET is a sophisticated malware designed to obtain supervisory control and target Siemens' SCADA control software over an entire industrial control system (ICS). It was discovered in June 2010 and is widely suspected of targeting Iran's uranium enrichment program [21]. The virus was introduced by plant personnel into a control system computer via USB. The delivery of the software was highly effective, and the injection was successful. Detection of the virus was difficult due to the advanced programming techniques used to conceal itself from users and anti-malware software [22]. Virus BlockAda, an antivirus company based in Belarus, discovered the Stuxnet worm after the company receives a sample of malware causing a computer in Iran to continually reboot itself [23].

Once infected, Stuxnet spreads via multiple trajectories towards control system PLC, through other control system computers connected to the local network and across entire ICS networks [14]. The worm utilizes probed password information to access Windows operating systems that run the WinCC and PCS 7 programs. These affected PLC programs mutate in-

structions that manage industrial plants. About the Iranian facilities, the PLC programs were modified by setting the centrifuge rotors at too-low and too-high frequencies resulting in significant slow-downs and speed-ups [24]. Through the exploitation of digital vulnerabilities in industrial systems, multiple nuclear facilities were similarly affected in Iran [25]. The concentration of information needed to develop Stuxnet's main program architecture exists on the vulnerable network space between nuclear facilities and their vendor networks.

Critical system information is needed to develop PLC-specific version updates that conceal changes to control system monitors. The hackers begin by identifying the vendors of the utility company which has special access to sensitive networks [5]. The vendors are targeted with conventional methods such as spear-phishing, a social-engineering technique aimed at creating vulnerabilities by utilizing the flaws in human error. Generally delivered by email, malicious content is aimed to prescribe familiarity with the user to deceive decision-making processes [26]. Such sophisticated phishing emails are difficult for email protection systems to detect [27]. Once network probes have been enabled, attackers can run software diagnostics on equipment, update software, and affect other services for critical systems [5].

Bradwell nuclear power plant

An employed guard conducted a cyberattack in June 1999, at the Bradwell Nuclear Facility located in the United Kingdom. The employed guard of Bradwell NPP, altered and destroyed data on an internal server. The attacker of the Bradwell facility managed to gain direct access to an internal server and while connected to the facility server, was able to alter and destroy secure database information [28]. This was completed entirely from inside of the facility. The employee attempted to sabotage secure database information and was believed to have hacked into Bradwell's computer system to alter sensitive information. Information deletion of this mode can cause a system-wide error. It is unclear what documentation was destroyed, but the information deleted was crucial for conducting business within the NPP.

Despite the intrusion, the automated access control systems within the facility were able to shut down due to a high-level security alert triggered by the attack. The security monitoring of crucial servers and systems was instrumental in actively defending against this internal cyberattack.

Energy future holdings

Comanche Peak NPP owned by energy future holdings (EFH), is located in Texas, United States. In March 2008, an employee attempted an internal cyberattack against the facility. The EFH company had fired an employee based on performance reasons but had forgotten to shut off their virtual private network (VPN) access to the facility network [29]. Using this account, the disgruntled employee managed to log onto the corporate network and began emailing out proprietary data to the personal email account linked to the employee. While connected to the private network, the employee could modify and delete files on the server. Also, an email was sent to an engineering group asking about the safety of the nuclear reactor when increasing to 99 % capacity output.

The EFH management system was directly affected and rendered inoperable, resulting in being unable to accurately forecast the parameters necessary to operate the business on the following day [30]. Damage against the NPP was purely financial and led to more strict guidelines on handling the operations for firing an employee. Overall, autonomous systems were able to identify and stop any further damages from occurring.

Lessons learned from internal cyberattacks

The review of the internal cyberattacks leads to drawing some conclusions. In particular:

- Typically, internal attacks target specific users or employees of a company. In particular, the employees and specific users that belong to nuclear facility vendors are at an increased probability of attack considering the vulnerabilities presented to the critical system infrastructure.
- It is increasingly important to develop a surveying method for monitoring signature anomalous behavior on these networks.
- Mitigating a cyber threat begins with the development of a digital forensic methodology capable of distinguishing malicious connections to vendor and business networks.

The following section is a detailed methodology for an AI system for surveying and identifying anomalous connections to nuclear vendor networks.

CYBERFORENSICS METHODOLOGY

Problem statement

The limitations of implementing AI with digital forensics stem from the ability of a system to acquire the data sets for training, analyzing that data, and performing computational functions to predict behavior. The main framework for a discussed model includes smart acquisition, smart analysis, and smart presentation [30]. The smart acquisition phase is particularly important due to the limited availability of NPP data sets, and its importance in the I&C components of the NPP [31]. Moreover, another limitation of data acquisition is the size of the

training data sets being utilized by the artificial neural network (ANN). While true that there are restrictions when it comes to training data set size and dependency, the cyberforensics model proposed in this paper is capable of constructing an identification method that utilizes reduced datasets for training.

Artificial neural network for forensics model

In this section, the digital forensics system for NPP that implements an ANN approach is presented. The proposed ANN follows an expertly trained architecture, which is depicted in fig. 2, therefore, the datasets being utilized will consist of pre-process data preparation components as the main system inputs. The network inputs are developed using two methods:

– expert constructed database of known safe connections, and
– network connection data packets containing all network connection information.

After calculations are completed, the outputs from the network are loss, accuracy, and probability of internal connection.

The aim of the digital forensics ANN is to identify internal malicious connections from general network traffic datasets. The system utilizes network packet database information to train the neural network to identify malicious connections that exist on the current network data flow. The goal of the neural network is to survey all network connections through a digital analyzer and determine the probability of a malicious connection being identified. The output of the ANN consists of a probability between the values of 0.0 and 1.0, where 0.0 denotes absolute certainty that an internal malicious connection has not been identified, while 1.0 denotes absolute certainty that one has been identified. Given this information, data analytics teams can easily identify cyber probes phishing for critical information. As the system is unable to interpret network packets directly from the network, an extra data preparation step is necessary for implementing this solution.

To obtain the necessary information from network packets, a network packet analyzer software is utilized to develop the input datasets for the neural network database. The network packet information is gathered and analyzed using third-party software called *WireShark*, a free and open-source packet analyzer [32]. This program is given IP network packets as input and then outputs the detailed packet information in a spreadsheet format. Once all network packet data is converted to a readable input for the neural network, it is possible to train the network on this data set. The inputs for the neural network contain all key signature information developed from the network packets. After training the neural network, the model allows for the detection of benign and unknown connections to a network by outputting a probability of a malicious connection being made to the online network.

Forensics method

The IP network packets are organized in a manner that is not ideal for neural network interpretation. The data within the network packet must be converted from its original form into an excel data table to ensure accurate data input. The packet contents being sent and received across the network are composed of both a header and payload.

In particular, the system we are proposing utilizes the transport control protocol (TCP) information that resides within every IP packet [32]. As the network devices produce network packets, a network analyzer is utilized to interpret and convert all packet information into a readable input source for the neural network. An example of the WireShark output is presented in fig. 3.

WireShark is a powerful network analyzer that is used as a data preparation tool for the TCP information stream being received on the network. This is accomplished by analyzing the payload for each IP packet and developing a unique features data table. Once the

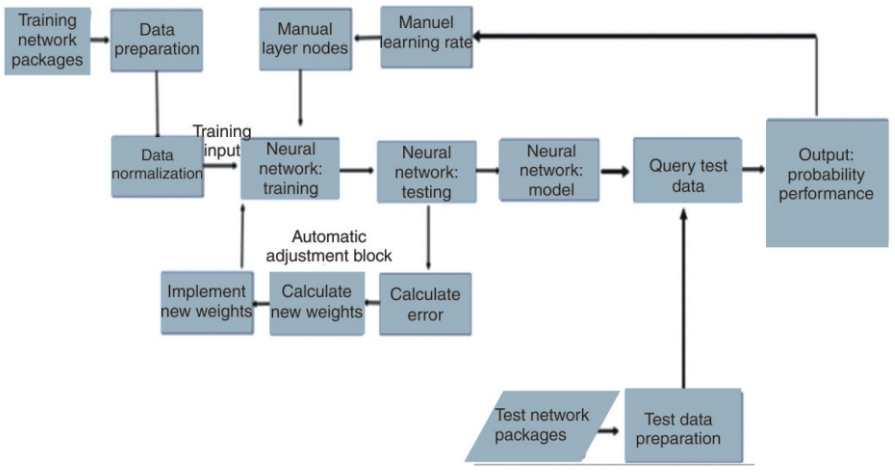

Figure 2. Abstract neural network model: Wireshark is utilized during the data preparation stages to construct inputs. Network outputs are saved to a file for viewing

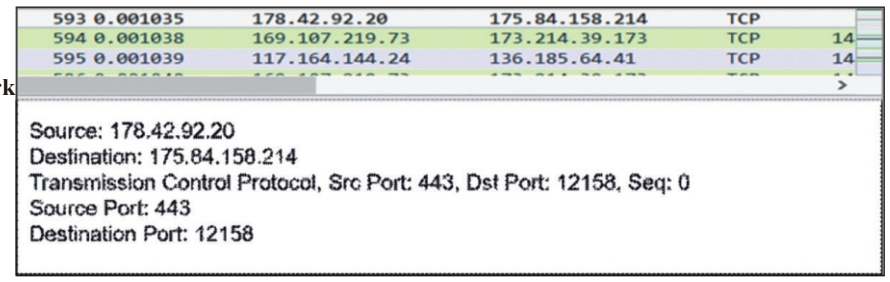

Figure 3. WireShark network packages output

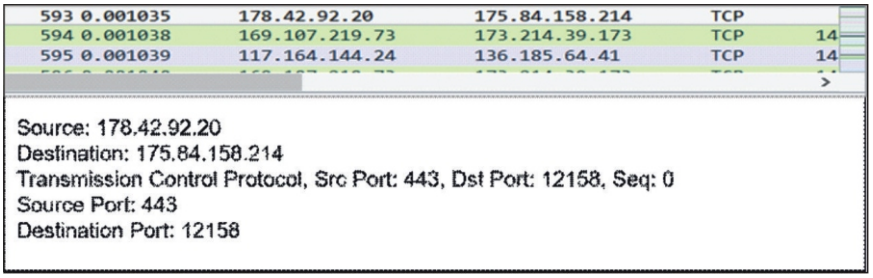

Figure 4. Internal connections data table

Input/ Output	IP Address	Source Port	Destination Port	Sequence Number	Source Address	Destination Address	Ack. Number	TCP Length
0	37.0.246.9	5938	58081	1	37.0.246.9	170.251.22	1	0
0	144.102.18	7172	58514	1	144.102.18	162.168.63	1	38
0	126.231.11	3389	64884	1	126.231.11	136.28.63.	1	59
0	169.221.14	39608	53	1	169.221.14	175.241.95	1	0
0	55.82.191.	9339	58197	1	55.82.191.	171.235.12	1	1428
0	69.35.149.	43464	443	1	69.35.149.	175.241.10	0	0
0	55.82.191.	9339	58197	1	55.82.191.	171.235.12	1	1428
0	55.82.191.	9339	58197	1	55.82.191.	171.235.12	1	1428
0	55.39.190.	1626	61942	1	55.39.190.	170.251.78	1	134
0	177.125.23	9380	34914	1	177.125.23	162.67.191	1	1400
0	200.69.87.	1361	445	1	200.69.87.	117.154.16	1	0
0	193.68.57.	58152	59033	1	193.68.57.	170.227.25	1	0

Figure 5. External connections data table

network data is available in a spreadsheet format, we then convert this table into a comma-separated values file (CSV). The information presented in fig. 4 and fig. 5 is the semi-final format of the data inputs for the neural network before transferring into a CSV.

All data collected for the training and testing of the neural network was directly obtained through the Center for Applied Internet Data Analysis (CAIDA) database. The specific data set that was utilized is the CAIDA Anonymized Internet Traces Dataset 2008 provided by the CAIDA official organization website [33]. From this dataset, specific location-based IP packages were concentrated to represent similar connections of a business on a local network. The training data directly coordinated with a series of local network connections that directly correlate with network packets that were sent and received in a localized network area. Also included in the data training set was another series of non-localized connections representing network packets that did not originate within the local network. The information included within the packet data table includes the *source, destination addresses, acknowledgment number, flags, length, the sequence number, TCP header length, window data, checksum, and actual user data*.

The neural network relies on an expertly constructed database containing the information included within the packet data tables derived from the network connections. Data tables are constantly updated with safe and reliable connections that have been made on the network previously. Both the training data and the testing data exist in this database as inputs for the AI model. The inputs for the neural network are directly delivered via database and data preparation methods are conducted on the data before entering the database. All data tables are converted to comma split files that are ideal for neural network training and interpretation. All information delivered from WireShark is stored and managed by database programs. To maintain a consistent interpretation of the network packet data, each connection must be divided into equal parts. Meaning that each IP address is divided into three parts and then maintained through its entry in the table within the database. This guarantees that each data packet that is sent into the database can be read by the neural network as an input variable with multiple nodes analyzing each part of the connection.

Our model utilizes an ANN [34] with three layers as depicted in fig. 6. Each layer utilizes dynamic activation functions to map the input nodes' values to their outputs. This model is developed with two types of non-linear activations including the rectified linear unit (Relu) and sigmoid activation functions, whose respective analytical functions are given below:

$$f(x) \quad \begin{matrix} 0 & \text{for } x \quad 0 \\ x & \text{for } x \quad 0 \end{matrix} \tag{1}$$

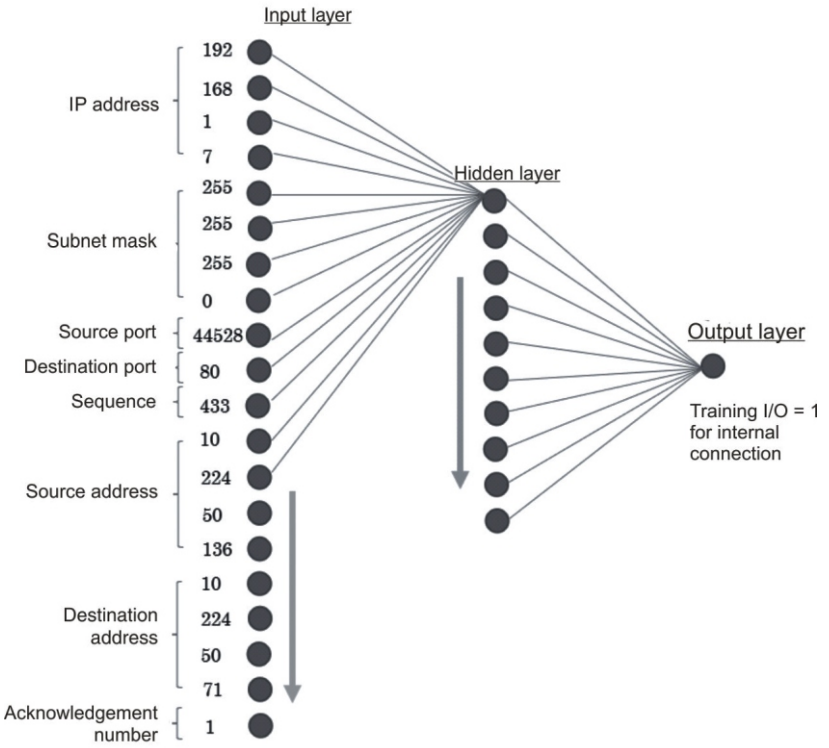

Figure 6. Cyber forensic neural network layer architecture with data table

Input/ Output	IP Address	Subnet Mask	Source Port	Destination	Sequence	Source Address	Destination Address	Ack. Number
1	192.168.1.7	255.255.255.0	44528	80	433	10.224.50.136	10.224.50.71	1

Figure 7. Data table for the ANN input layer

$$f(x) \quad \frac{1}{1 \quad e^{\,x}} \tag{2}$$

where x is the input to the function and e is the exponential term [34]. These particular activation functions are being chosen because of the observation driven on the desired results on a set of validation data. The data is non-linear as well as the correlations being non-linear, therefore we expect the inputs and outputs to have non-linear results.

The input layer of the ANN consists of 20 input nodes that require the input data entries to be maintained by the database and separated equally by requiring a pre-processing technique called CSV data array input arrangement. The input data for each node, as shown in fig. 7, include such information as IP addresses for the source and destination computers and formats them into equal sections such that each node accounts for only a particular section of each network packet. The dataset being used as inputs by the neural network program splits the data into the target and data fields. The target is what correctly distinguishes the difference between an internal and external connection being benign or anomalous. Each data part is divided into equal and normalized decimal values ranging from 0.01 to 0.9 for each corresponding input node. Normalizing each data member for neural network input nodes is essential for utilizing the sigmoid function within the multiple layers of the model. Normalization functions of the data inputs include the NumPy library that is available through open-source software – written for the Python programming language. These functions add calculation support for large, multi-dimensional arrays and matrices to return the data as an array. The preprocessing library included in the scikit-learn package, a free software machine learning library for Python (*developed by David Cournapeau*), provides several common utility functions and transformer classes utilized for supporting machine learning programs. The function, MinMaxScaler() being part of the preprocessing library, is a common requirement to normalize the data into a format that is ideal for the neural network inputs. The number of inputs and nodes utilized is identical for every connection made on the network. Generally, the information contained in the input nodes would be of correct separation and format due to the preprocessing steps that preceded the data being input into the neural

network. After the Relu function is activated on the input layer, the output from these nodes is sent into the hidden layer.

The second layer in the neural network is the hidden layer that consists of 18 nodes (*i. e*., neurons). Node quantity was decided based upon generated output from the neural network. Maintaining the highest level of accuracy and lowest total loss of the function played essential roles for initializing the hidden layer with 18 nodes. The Relu activation function works on the data being input into the hidden layer and is considered the most used activation function when designing a convolutional neural network or deep learning model. The activation function does not have a curved shape to its slope as compared to the sigmoid function. Instead, the Relu function utilizes a half rectified linear function which has an output between 0 or 1 for each input. Also, any input values into the hidden layer that approach negative values are immediately set to a value of 0. In our model, the data inputs from each layer should never reach below the value of 0. The output values from the hidden layer nodes are sent as inputs into the output layer.

The third and final output layer consists of one node that determines whether an internal connection has been made that is not recognized as safe. The output from the neural network is a value between 0 and 1 that determines the probability that a particular connection was abnormal and unsafe in comparison to the database connections.

The final layer consists of a sigmoid activation function that determines the value that the output will take. The sigmoid activation is especially used for models where we have to predict the probability as an output. This is due to the probability of anything existing only between the range of 0 and 1. The sigmoid activation function $S(x)$ is essential because it provides an efficient mathematical determination for the probability values being output from the final layer. Besides, the output layer takes the probability value computed by the sigmoid activation function and compares it to a threshold value $T = 0.5$.

When the probability value is equal to or above the threshold T, then the output layer provides a value of 1 indicating that the connection being analyzed has a higher probability of originating internally on the network. In the case that the probability value is less than 0.5, the output layer provides a value of 0 indicating that an internal attack has not been detected. The detection function $D(x)$ is given below:

$$D(x) \quad \begin{matrix} 1, & \text{if } S(x) \quad 0 \\ 0, & \text{if } S(x) \quad 0 \end{matrix} \tag{3}$$

where $D(x)$ represents the forensics function output for the input data x, and $S(x)$ is the activation function of our neural network model (as explained earlier).

For compilation and training of the neural network our model utilizes the Adam optimizer [35], which is used with a learning rate of 0.2 and a binary cross-entropy loss function that outputs over an accuracy metric. For training the model we utilize a batch size of 50 data points based on standardized computer performance. Also, this model has been programmed to train with 200 epochs. This was determined by observing the training results and optimizing the model for the highest accuracy. By graphing the training and testing data against the number of epochs, it was possible to choose an ideal amount for fitting the model to predict the anomalous behavior of the suggested testing data.

TESTING RESULTS

In this section, the presented ANN cyberforensic system is applied to the test set comprised of network connections that are depicted previously in figs. 4 and 5. Each of the network connections is divided into equal and different parts that were discussed in the previous section, *e. g*., *source and destination* IP addresses, *acknowledgement number*, *subnet mask*, *source port*, and *sequence number*. The test set is constructed using the datasets previously discussed within the previous sections. Each dataset consists of 150 network connections ranging from internal to external connections being made on the network. The forensic system model experiment was conducted by testing the outputs of known connections against the prediction model.

Due to the limited number of available datasets, we have divided the dataset into training and testing data by using 10 different ways to split them. It is expected that every run on a different portion of the dataset will provide different output results. To overcome the hurdle of the limited number of available datasets, while extensively testing our methodology is a larger variety of data, we have divided the dataset into training and testing data by using 10 different ways to split them. Every case is run 10 times and the main statistics of the results are obtained: mean and standard deviation. This approach is known as the *K*-Fold cross-validation testing [34] presented in fig. 8, wherein our case $K = 10$.

The utilization of *K*-fold cross-validation testing is detrimental to obtaining the mean and standard deviations of the resulting neural network. It is commonly used as a statistical method to estimate the skill of machine learning models. Also, the results from the *K*-Fold cross-validation testing were applied to compare and select the correct model for our given predictive modeling problem. The *K*-Fold procedure has the main network components from the dataset separated into groups. The number of groups utilized in our analysis was 10 equal splits across the dataset. Also, the batch size defined for each fold is set to split the information as 80 % for training and 20 % for testing data.

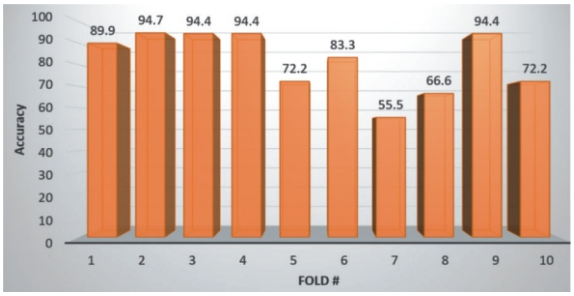

Figure 8. Accuracy *vs.* fold

This was performed resulting in a less biased estimate of the model's skills. The *K*-Fold cross-validation was developed using the scikit-learn library that provides an implementation that will split the given data sample equally. Within the scikit_learn library, the function *kFold* was called to perform the functions needed for *K*-Fold cross-validation.

Accuracy for the predictive model maintained an efficient average of 81.76 %. This details the ability of the forensic system to accurately predict correct outputs on different test datasets. As seen in fig. 8, the accuracy over the multiple folds has a minimum accuracy of 55.5 % and a maximum accuracy of 94.7 %. Overall, the accuracy of the forensics system manages to detect internal connections with high accuracy and dependency.

The results have shown that generally there are model optimizations available to increase the prediction accuracy and decrease system loss for the neural network forensics model. Fortunately, the existing parameters for prediction models detail that this model's accuracy developed through training is significant enough to utilize for accurately predicting network packets in NPP system traffic.

CONCLUSIONS

In this paper, the importance of internal cyberattacks in nuclear facilities was emphasized by providing a review of all the well-known attacks of this type. In addition, lessons learned from the review led us to the development of a cyberforensics model for NPP that implements an ANN approach as presented in the current paper. The ANN forensics aims to designate whether the detected attack comes from internal malicious connections based on general network traffic datasets.

Our methodology was developed for specifically advancing the search for the source of a cyberattack inside of a nuclear facility. By having convincing evidence that the cyberattack is from an internal network connection, then the forensics search for it will be much shorter and the problem mitigated with minimum damage loss. The system utilizes network packet database information to train the neural network to identify malicious connections that exist on the current network data flow. The model is based on the use of ANN to maintain surveillance over real-time network data flow. The goal of the neural network is to survey all network connections through a digital analyzer and determine the probability of an internal malicious connection being identified.

The proposed ANN model has been tested in a set of 150 real-world network patterns. Simulations of network traffic analysis and behavior were done with the aid of randomizer processes that provided equal data implementation opportunities. Obtained results support the belief that the ANN forensics system provides accurate detection of malicious connections to the internal networks of the businesses that support NPP functionality – *i. e.*, the average accuracy in 1- fold testing was equal to 81.76 % -. Furthermore, the proposed system is fully autonomous and requires no human intervention at any stage other than data input, while being computationally inexpensive.

As the technologies of the current cyberspace will continue to advance, the benefits of utilizing forensics systems to actively locate the origin of cyberattacks will become more apparent [36, 37]. An advantage of the presented approach is the use of ANN to locate whether the attack is internal or not may significantly reduce the time to identify the attacker by restraining the suspects within the nuclear facility premises.

ACKNOWLEDGMENT

This work was performed partially under the auspices of the Consortium on Nuclear Security Technologies (CONNECT) supported by the U.S. Department of Energy/National Nuclear Security Administration under Award Number(s) DE-NA0003948.

AUTHORS' CONTRIBUTIONS

M. Alamaniotis proposed the concept for the implementation which was conducted by B. Campos. All authors analyzed the results and participated in the preparation of the final version of the manuscript under the supervision and guidelines of M. Alamaniotis.

REFERENCES

[1] Gawand, H. L., *et al*., Securing a Cyber-Physical System in Nuclear Power Plants Using Least Square Approximation and Computational Geometric Approach, *Nuclear Engineering and Technology, 49* (2017), 3, pp. 484-494

[2] Cho, H. S., *et al*., Cybersecurity in Nuclear Industry-Analytic Study from the Terror Incident in Nuclear Power Plants (NPP), *Annals of Nuclear Energy, 99* (2017), pp. 47-53

[3] Calderon, R., The Benefits of Artificial Intelligence in Cybersecurity, *Economic Crime Forensics Capstones, 36*, 2019, Thesis, available at https://digitalcommons.lasalle.edu/ecf_capstones/36

[4] Trautman, L. J., Ormerod, P., Industrial Cyber Vulnerabilities: Lessons from Stuxnet and the Internet of Things, *72 University of Miami Law Review, 761* (2018), 3, pp. 1-66
[5] Michener, J., Beating the Air-Gap: How Attackers Can Gain Access to Supposedly Isolated Systems, *Energy Central Journal,* 2018, available at https://energycentral.com/c/iu/beating-air-gap-how-attackers-can-gain-access-supposedly-isolated-systems
[6] Andryushin, A., *et al*., Issues of Operating Systems Usage for Nuclear Power Plants, *Annals of Nuclear Energy, 70* (2014), pp. 87-89
[7] Demertzis, K., *et al*., The Next Generation Cognitive Security Operations Center: Network Flow Forensics Using Cybersecurity Intelligence, *Big Data and Cognitive Computing, 2* (2018), 35, pp. 1-17
[8] Illiashenko, O. A., *et al*., Cybersecurity Case for FPGA-Based NPP Instrumentation and Control Systems, *Proceedings*, ASME 24th International Conference on Nuclear Engineering, Charlotte, NC, USA, 2016
[9] Roh, J. H., *et al*., Cyber Security System with FPGA-based Network Intrusion Detector for Nuclear Power Plant, *Proceedings*, The 46th Annual Conference of the IEEE Industrial Electronics Society, Singapore, 2020
[10] Park, J. W., Lee, S. J., Probabilistic Safety Assessment-Based Importance Analysis of Cyber-Attacks on Nuclear Power Plants, *Nuclear Engineering and Technology, 51* (2019), 1, pp. 138-145
[11] Lee, S., Huh, J. H., An Effective Security Measures for Nuclear Power Plant Using Big Data Analysis Approach, *The Journal of Supercomputing, 75* (2019), 8, pp. 4267-4294
[12] Park, J., *et al*., Implementation of Cyber Security for Safety Systems of Nuclear Facilities, *Progress in Nuclear Energy, 88* (2016), pp. 88-94
[13] Guo, Y., *et al*., Cybersecurity Risk Analysis and Technical Defense Architecture: Research of ICS in Nuclear Power Plant Construction Stage, *Proceedings*, 3rd IAEA International Conference on Nuclear Security: Sustaining and Strengthening Efforts, Vienna, Austria, 2020
[14] Gera, J., Battula, B. P., Detection of Spoofed and Non-Spoofed DDoS Attacks and Discriminating Them from Flash Crowds, *EURASIP Journal on Information Security, 9* (2018), pp. 1-12
[15] Zhang, F., *et al*., A Robust Cybersecurity Solution Platform Architecture for Digital Instrumentation and Control Systems in Nuclear Power Facilities, *Nuclear Technology, 206* (2020), 7, pp. 939-950
[16] Silva, R. B. E., *et al*., Development of the Asherah Nuclear Power Plant Simulator for Cyber Security Assessment, *Proceedings*, 3rd IAEA International Conference on Nuclear Security: Sustaining and Strengthening Efforts, Vienna, Austria, 2020
[17] Campos, B. M., Alamaniotis, M., Lessons Learned about Network Defenses of Nuclear Power Plants: A Critical Analysis of Internal Cyberattacks, *American Nuclear Society Transactions, 121* (2019), 1, pp. 511-514
[18] Youseff, Z., *et al*., Proposal of Countermeasure against Attacks Similar to Stuxnet, *International Journal of Computer Science and Information Security, 12* (2014), 5, pp. 1-7
[19] Chenoweth, C., *et al*., The U.S. Nuclear Regulatory Commission's Cyber Security Regulatory Framework for Nuclear Power Reactors NUREG/CR-7141, United States Nuclear Regulatory Commission, 2014
[20] Baylon, C., *et al*., Cyber Security at Civil Nuclear Facilities: Understanding the Risks, Chatham House for the Royal Institute of International Affairs, London, UK, 2015
[21] Mueller, P., Yadegari B., The Stuxnet Worm, Tutorial from the University of Arizona, available at https://www2. cs. arizona, edu/č collberg/Teaching/466-566/2012/Resources/presentations/topic9-final/report.pdf, 2012
[22] Farwell, J. P., Rohozinski, R., Stuxnet and the Future of Cyberwar, *Survival, 53* (2011), 1, pp. 23-40
[23] Baezner, M., Iranian Cyber-activities in the Context of Regional Rivalries and International Tensions, CSS Cyberdefense Hotspot Analyses, Report published by ETH Zurich, 2019
[24] Kerr, P. K., *et al*., The Stuxnet Computer Worm: Harbinger of an Emerging Warfare Capability, Washington, DC: Congressional Research Service, 2010
[25] Bullee, J. W., *et al*., Spear Phishing in Organisations Explained, *Information & Computer Security, 25* (2017), 5, pp. 593-613
[26] Stembert, N., *et al*., A Study of Preventing Email (Spear) Phishing by Enabling Human Intelligence, *Proceedings*, European Intelligence and Security Informatics Conference, pp. 113-120, Manchester, UK, 2015
[27] Maguire, K., Guard Tries to Sabotage Nuclear Reactor, the Guardian, available at http://www.theguardian.com/uk/2001/jan/09/nuclear.world, 2001
[28] Poulsen, K., Ex-Employee Fingered in Texas Power Company Hack, Wired, available at http://www.wired.com/2009/ 05/efh/, 2009
[29] Rughani, P., Artificial Intelligence Based Digital Forensics Framework, *International Journal of Advanced Research in Computer Science, 8* (2017), pp. 10-14
[30] Shin, J., *et al*., Methodology on Cyber Security Evaluation in Nuclear Facilities Considering I & C Architecture, *Proceedings*, 10th International Topical Meeting on Nuclear Plant Instrumentation, Control, and Human-Machine Interface Technologies, San Francisco, Ca., USA, pp. 927-934, 2017
[31] Chappell, L., Wireshark 101: Essential Skills for Network Analysis – Second Edition Wireshark Solution Series, Laura Chappell University, San Jose, Cal., USA, 2017
[32] Bertsekas, D. P., *et al*., *Data Networks*, Vol. 2, Prentice-Hall International, New Jersey, N. Y., USA, 1992
[33] **, CAIDA Website: https://www.caida.org/data/passive/passive_2008_dataset.xml
[34] Bishop, C. M., *Pattern Recognition and Machine Learning*, Springer, Berlin, 2006
[35] Kingma, D. P., Ba, J., Adam: A Method for Stochastic Optimization, *Proceedings*, 3rd International Conference for Learning Representations (ICLR), San Diego, Cal., USA, 2015
[36] Jones, R. A., *et al*., System-Aware Security for Nuclear Power Systems, *Proceedings*, 2011 IEEE International Conference on Technologies for Homeland Security (HST), Boston, Mass., USA, 2011, pp. 224-229
[37] Gupta, D., *et al*., Cyber Threat Scenarios for Electrical Systems of Nuclear Power Plants, *Proceedings*, International Conference on Nuclear Engineering, vol. 51531, p. V009T16A085, 2018, London, UK, 2018

Received on March 31, 2021
Accepted on August 12, 2021

Брандин М. КАМПОС, Милтиадис АЛАМАНИОТИС

ПРЕГЛЕД УНУТРАШЊИХ САЈБЕР НАПАДА У НУКЛЕАРНИМ ОБЈЕКТИМА И ВЕШТАЧКИ НЕУРОНСКИ МРЕЖНИ МОДЕЛ ЗА СПРОВОЂЕЊЕ УНУТАРЊЕ САЈБЕР ФОРЕНЗИКЕ

Примена дигиталних технологија у оквиру модерног развоја у системима сајбер одбране од суштинског је значаја за заштиту постројења за производњу енергије. Једна од важних компоненти одбране је сајбер форензика: када се једном открије напад, да се утврди његово порекло. У овом раду дат је преглед познатих сајбер напада у нуклеарним објектима, са поукама из њих које су довеле до развоја приступа машинског учења који примењује идентификацију унутрашњих напада у мрежама података објекта. Наш приступ може се сматрати једним од слојева у стратегији дубинске одбране која идентификује да ли је напад изнутра, што може резултирати бржом идентификацијом порекла нападача. Представљени модел користи испитивање мрежних пакета да би дао тачна предвиђања о детаљима порекла злонамерних мрежних веза. Приступ спаја више математичких функција унутар вештачке неуронске мреже да би пружио одговор у облику 0/1, тј. да ли је напад идентификован као унутрашњи или не. Употреба разних рест случајева развијена је за истраживање релевантности и ваљаности предиктивног приступа. Предложена примена испитује се са варијансом мрежног пакета података, а добијени резултати показују високу тачну стопу откривања.

Кључне речи: сајбер сигурност, дигитална форензика, нуклеарна електрана, унутрашњи напад, неуронска мрежа