Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 2, No. 1, February : 2023

DETECTION AND IDENTIFICATION OF CYBER- ATTACKS WITH MACHINE LEARNING APPROACH

Prof. B V R V PRASAD Professor, Department of ECE, NRI Institute of Technology
V. NEELIMA, T. SIVALEELA, SK. GOWSIA PYARI, V. JOHN, T. HARSHA VARDHAN
B. Tech, Student, Department of ECE, NRI Institute of Technology

ABSTRACT

Artificial intelligence (AI) technologies have given the cyber security industry a huge leverage with the possibility of having significantly autonomous models that can detect and prevent cyberattacks – even though there still exist some degree of human interventions. Because of the existing methods of attack and the dynamic nature of malware or other unwanted software (adware etc.) it is important to create, update and approve malicious packages that can be available to the public automatically and systematically Contrasted with the past, improvements in PC and correspondence innovations have given broad and propelled changes. The use of new innovations gives incredible advantages to people, organizations, and governments, be that as it may, messes some up against them. For instance, the protection of significant data, security of put away information stages, accessibility of information and so forth. Contingent upon these issues, digital fear-based oppression is one of the most significant issues currently. Digital fear, which made a great deal of issues people and establishments, has arrived at a level that could undermine open and nation security by different gatherings, for example, criminal association, proficient people, and digital activists. An Intrusion Detection System (IDS) is a network security technology originally built for detecting vulnerability exploits against a target application or computer. Along these lines, Intrusion Detection Systems (IDS) has been created to maintain a strategic distance from digital assaults. Right now, we are using support vector machine (SVM), Random Forest and ANN algorithms. According to results, SVM accuracy is 97.800005%, Random forest accuracy is 97.816655% and ANN accuracy is 94.22222%. Experimental analysis shows us that machine learning algorithms can detect attacks with higher performance than usual methods and can make cyber security simpler, more proactive, less expensive, and far more effective.

Keywords: *Cyber Security, Denial of Service, SVM, Random Forest*

I. INTRODUCTION

Recent advances in technology have led to the introduction of cyber-physical systems, which due to their better computational and communicational ability and integration between physical and cyber-components, has led to significant advances in many dynamic applications. But this improvement comes at the cost of being vulnerable to cyber-attacks. Cyber-physical systems are made up of logical elements and embedded computers, which communicate with communication channels such as the Internet of Things (IoT). More specifically, these systems include digital or cyber components, analog components, physical devices, and humans that designed to operate between physical and cyber parts. In other words, a cyber-physical system is any system that includes cyber and physical components and humans and can trade between the physical and cyber parts. In cyber-physical systems, the security of these types of systems becomes more important due to the addition of the physical part.

One of the most important challenges of a cyber-physical system, in its physical part is the presence of a large number of sensors in the environment, which collect so much data, with so much variety, and at high speed. Also, the connection between the sensors and the necessary calculations and the analysis of the obtained data will be among the main challenges. Therefore, one of the most

important features of a cyber-physical system is to communicate between these sensors, compute and control the system. AI technologies have been utilized in gathering data which can then be processed into information that are valuable in the prevention of cyberattacks. These AI-based cybersecurity frameworks have commendable scalability about them and can detect malicious activities within the cyberspace in a prompter and more efficient manner than conventional security architectures. However, our one or two completed studies did not provide a complete and clear analyses to apply different machine learning algorithms on different media systems.

The security of cyber-physical systems to detect cyber-attacks is an important issue in these systems. It should be noted that cyber-attacks occur in irregular ways, and it is not possible to describe these attacks in a regular manner. However, in the deception attacks that inject false data to system, which are carried out by abusing system components, such as sensors or controllers and it can corrupt data or enter incorrect information into the system and cause misbehaving. These attacks can be detected and monitoring by system. But if the attacker can plan a high-level attack to prevent himself from being identified, these attacks are called stealthy deception attacks, and other common methods of counteracting attacks will not work.

Therefore, it is important to be aware of the attacks that occur to respond in a timely manner to attackers. In other words, the security system must be aware of the attack, otherwise it will not be able to identify attacks efficiently. Cyber defense can be improved by using security analytic to search for hidden patterns and how to deceive. The main purpose of this study is to investigate the functionality of deception attacks from the system's point of view and to identify and control them. In deception attacks, either the data generated in the system is attacked or the data is applied to the system is false, which is like the desired data and the system recognizes them as valid. Therefore, analyzing the data in the system that is exchanged between nodes will be necessary for its security. Therefore, this motivates the use of machine learning methods in cyber security.

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane:SVM algorithm can be used for Face detection, image classification, text categorization, etc.

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

Artificial neural networks (ANNs), usually simply called neural networks (NNs) or neural nets, are computing systems inspired by the biological neural networks that constitute animal brains. An ANN is based on a collection of connected units or nodes called artificial neurons, which loosely model the neurons in a biological brain. Each connection, like the synapses in a biological brain, can transmit a signal to other neurons. An artificial neuron receives signals then processes them and can

signal neurons connected to it. The signal at a connection is a real number, and the output of each neuron is computed by some non-linear function of the sum of its inputs. The connections are called edges. Neurons and edges typically have a weight that adjusts as learning proceeds. The weight increases or decreases the strength of the signal at a connection. Neurons may have a threshold such that a signal is sent only if the aggregate signal crosses that threshold. Typically, neurons are aggregated into layers. Different layers may perform different transformations on their inputs. Signals travel from the first layer (the input layer), to the last layer (the output layer), possibly after traversing the layers multiple times.

II. LITERATURE SURVEY

An IDS (Intrusion Detection System) generally must deal with problems such as large network traffic volumes, highly uneven data distribution, the difficulty to realize decision boundaries between normal and abnormal behavior, and a requirement for continuous adaptation to a constantly changing environment. In general, the challenge is to efficiently capture and classify various behaviors in a computer network. Strategies for classification of network behaviors are typically divided into two categories.

They are misuse detection and anomaly detection. Misuse detection techniques examine both network and system activity for known instances of misuse using signature matching algorithms. This technique is effective at detecting attacks that are already known. However, novel attacks are often missed giving rise to false negatives. Alerts may be generated by the IDS, but reaction to every alert waste time and resources leading to instability of the system. To overcome this problem, IDS should not start elimination procedure as soon as the first symptom has been detected but rather it should be patient enough to collect alerts and decide based on the correlation of them. Some research statistics with regards to the impact of cyber security to businesses, organizations.

In recent years, cybercrime has been responsible for more than $400 billion in funds stolen and costs to mitigate damages caused by crimes. It has been predicted that a shortage of over 1.8 million cybersecurity workers will be experienced by 2022. It's been predicted that organizations globally will spend at least $100 billion annually on cyber security protection. Attackers currently make over $1 billion in annual revenue from Ransomware attacks, such as WannaCry and Crypto Wall attacks.

III. PROPOSED SYSTEM:

The main objective of the proposed system is to develop a Cyber Attack Detection Model (CADM) which is a system or tool that intends to deal with malicious activities by attackers. The main idea behind this model is to analyze the malicious activates and compare with different machine learning algorithms like SVM, Random Forest & ANN model and design a efficient front end tools which can identify the malicious activity. Cyber security has become a major concern in the era of modern technology.

The important steps of the algorithm are given in below.

1) Normalization of every dataset.
2) Convert that dataset into the testing and training.
3) Form IDS models with the help of using RF, ANN and SVM algorithms.
4) Evaluate every model's performance.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 2, No. 1, February : 2023

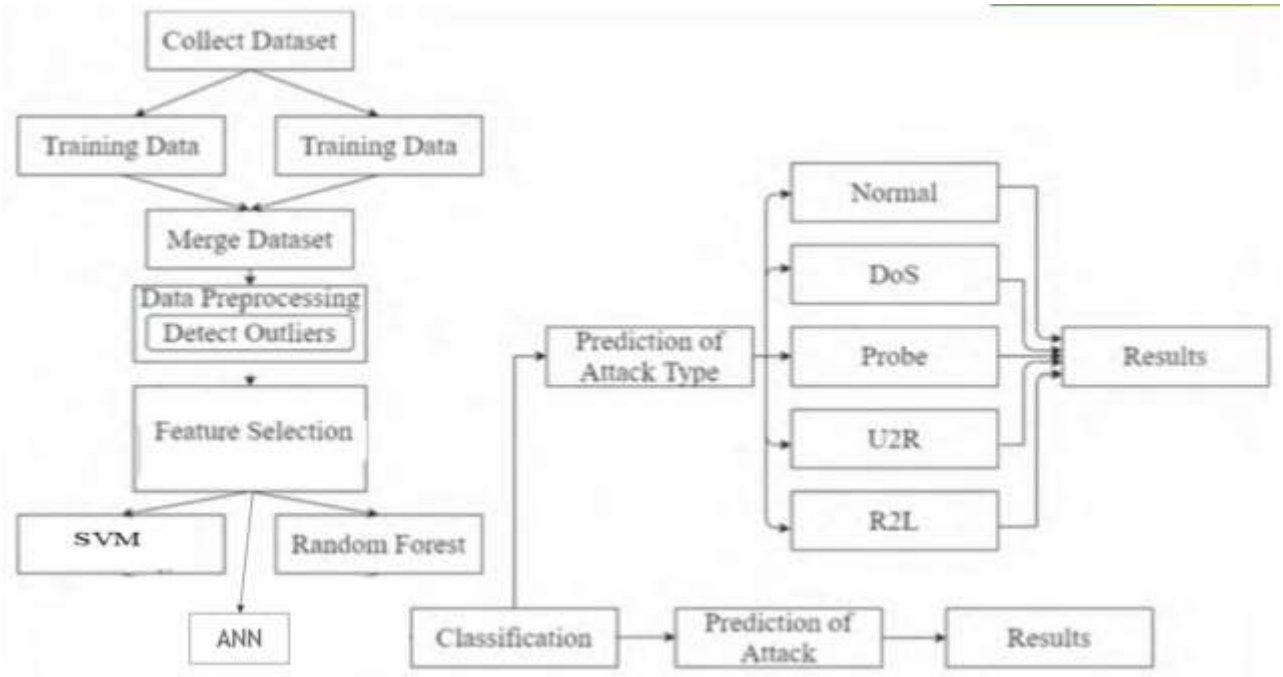

Figure 1. Block Diagram of Cyber Attack Detection Model (CADM)

The block diagram shown in Figure 1 is a Cyber Attack Detection Model (CADM) with machine learning approach. The implementation of the Cyber Attack Detection Model (CADM) with machine learning approach contains following steps.

Data Collection: Collect sufficient data samples and legitimate software samples.

Train and Test Modeling: Split the data into train and test data Train will be used for training the model and Test data to check the performance.

Data Preprocessing: Data augmented techniques will be used for better performance.

Attack Detection Model: Based on the model trained algorithm will detect whether the given transaction is anomalous or not

Feature selection, prediction and classification: Implemented by SVM, RF and ANN Algorithms

IV. RESULTS

The results of proposed Cyber Attack Detection Model (CADM) at different steps described below with figures 1 to 6. This proposed model tested with twenty different website links and these links listed in table1 with the details of website attacked or not and type of attack. The random results of tabulated website links were analyzed with mean, SD and variance parameters. The accuracy and processing time parameters used for analyzing the performance of proposed CADM model. The accuracy of this model with SVM, RF and ANN machine learning algorithms tabulated in table2 and processing time tabulated in table 3.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 2, No. 1, February : 2023

Figure 2. Home Page

The Figure 2. shows about the Home-page of cyber-attack detection. Here we need to enter the username and password. Then we need to login into the Homepage. If we do not have an account, then we need create account. Then only we can login into the page.

Figure 3. Add Data

The Figure 3. shows about the add data page of cyber-attack detection. Here we need to enter the details of the problem we are facing. We need to enter entity, year, records, organization type, methods, add data and time. Finally, we have to press the submit button.

Figure 4. Cyber Attack Data

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 2, No. 1, February : 2023

The Figure 4. shows about the Cyber Attack Data page of cyber-attack detection. If cyber-attack is detected, then the data is added in the cyber-attack data. In the above figure we can observe the cyber-attack data

Figure 5. Non-Cyber Attack Data

The Figure 5. shows about the Non-Cyber Attack Data page of cyber-attack detection. If cyber-attack is not detected, then the data is added in the non-cyber-attack data. In the above figure we can observe the non-cyber-attack data.

Figure 6. Cyber Attack Analysis

The Fig 6. shows about the Cyber Attack Analysis page of cyber-attack detection. In this page, we can analyse the cyber-attacks. Here it shows the type of the cyber-attack that is affected by the system.

Type	**Website Link**	**Cyber-attack/not**	**Cyber-attack name**
NRI College	http://nriit.edu.in/	Non-Cyber attack	Unmalware
NRI exam cell	https://www.nriitexamcell.com/autonomous/	Non-Cyber attack	Unmalware
NRI pharmacy	https://collegedunia.com/college/58515-nri-college-of-pharmacy-agiripalle	Non-Cyber attack	Unmalware
Health care	https://www.linkedin.com/jobs/view/930124877/?refId=d3493ec8-privateid.37f8-4218-92b2-	Cyber attack	Main in middle attack

	4656b383a955&trk =eml-jymbii-organic-job-card&midToken=serverattack /AQHBnYxQHAJchw&trkEmail=eml-jobs_jymbii_ digest-null-4-null-null-9xzoen~joem4ler~uj-null-jobs~ view&lipi=urn%3Ali%3Apage%3Aemail_jobs_jymbii_digest%3BCxmcCwrxR62ABhqSrl2dYA%3D%3D		
Advertising	https://www.google.co.in/search?q=python+free+online+course+certification&oq=p&aqs/2F4=chrome.0.69i59j69i60l4j69i57.2378j0j7&sourceid=c/tcpoffset/hrome&ie=UTF-8	Cyber attack	Drive by attack
Retail	http://127.0.0.1:8000/user/userpage/portid	Cyber attack	Eavesdropping attack
Banking	https://stackoverflow.com/questions/43727583/expected-string-or-bytes-like-object/2C	Cyber attack	Password attack
Youtube	https://www.youtube.com/	Non-Cyber attack	Unmalware
Linkedin	https://www.linkedin.com/checkpoint/challenge/AgEm4oMya4FT4QAAAYZzAY-gyW5gc3Wpx2mvVtLvz7pZV_t8s6nkNSmIPxSuY8Wo1-0ha0VKSIA2dSpElifPEBMtnkAxCw?ut=0ZZqJ8atu5uqE1	Non-Cyber attack	Unmalware
Instagram	https://www.instagram.com/	Non-Cyber attack	Unmalware
Tech	https://www.google.co.in/search?q=edi&oq=edi&aqs=chrome..69i57j69i61l3j0l2.1854j0j9&sourceid=chrome&ie=UTF-8/NID/monlist	Cyber attack	SQL injection attack
Telecoms	https://www.google.co.in/search?q=dsv&oq=dsv&aqs=chrome..69i57j0l5.1403j0j7&sourceid=chrome&ie=UTF-8/fragflag	Cyber attack	Man-in-the-middle attack
Gaming	https://www.bayt.com/en/job-seekers/create-account/?url_id=1&2F4/utm_medium=associate&utm_source=walkinupdates%2ecom+1880861/malwareid	Cyber attack	Eavesdropping attack
Mana Badi	http://www.manabadi.co.in/boards/apsbtet-results-apsbtet-diploma-results-sbtet-results.asp	Non-Cyber attack	Unmalware
E-Aadhaar	https://myaadhaar.uidai.gov.in/genricDownloadAadhaar	Non-Cyber attack	Unmalware
Mail	https://mail.google.com/mail/u/0/#inbox	Non-Cyber attack	Unmalware
Financial	https://mail.google.com/mail/u/0/#inbox/2C	Cyber attack	Password attack
Government	https://127.0.0.1:8000/user/userpage/portid	Cyber attack	Password attack

Academic	https://stackoverflow.com/questions/43727583/expected-string-or-bytes-like-object/ECSID/c	Cyber attack	Drive-by attack
Military	https://www.google.co.in/search?q=edi&oq=edi&aqs=chrome..2F4;69i57j69i61l3j0l2.1854j0j9&sourceid=chrome&ie=UTF-8	Cyber attack	Phishing and spear phishing attack

Table 1. List of websites taken for the testing

The above table 1 shows about the list of 20 websites taken for the testing. In these twenty-website links, nine website links are non-cyber-attack links and eleven website links are cyber-attack links. We have several cyber-attack types, they are Main in middle attack, Drive by attack, Eavesdropping attack, Password attack, Phishing and spear phishing attack and SQL injection attack. Here we can observe that there are two Main in middle attacks, two Drive by attacks, two Eavesdropping attacks and three Password attacks.

For the above website links, proposed model applied with SVM, RF and ANN algorithms. The Accuracy and Processing time of model with machine algorithms were analysed with the parameters Mean, Standard Deviation. The evaluated and compared results of accuracy tabulated in table2 and processing time in tabel3.

Parameters	SVM	Random Forest	ANN
Mean	97.800005	97.816655	94.22222
Standard Deviation	1.1372725	1.180758301	2.3726968
Variance	1.2933886	1.3941902	5.6296903

Table 2 Measurement of accuracy of SVM, RF and ANN algorithms

The table 2 shows that the measurement of accuracy of model with SVM, RF and ANN algorithms. Observed that the accuracy of RF and SVM is almost similar. But the accuracy of ANN is less compared to SVM and RF. If we consider accuracy is the main parameter, we can prefer the RF algorithm.

Parameters	SVM	Random Forest	ANN
Mean	2.6149398	4.2615391	0.40252215
Standard Deviation	0.3636191	4.0927822	0.15924593
Variance	0.1322188	16.750866	0.02535927

Table 3. Measurement of processing time of SVM, RF and ANN algorithms

The table 3 shows that the measurement of processing time of model with SVM, RF and ANN algorithms. Observed that RF takes more processing time than the SVM algorithm. If the processing time is the main parameter and speed is required then we can prefer SVM otherwise RF. ANN also takes the less processing time but the accuracy is low compared to SVM and RF algorithms.

V. CONCLUSION & FUTURE SCOPE

In this study, an attempt was made to use the resilient control consensus method in complex discrete cyber-physical networks with several local attacks. By applying this control method, it was observed that even in the presence of cyber-attacks, the system can remain stable and isolate the attacked node and the performance of the system is not weakened. Using the neural network used in this study, it was observed that with a deep neural network, with 7 hidden layers, the system shows better performance. At the present time, assessments of help vector machine, ANN, Random Forest and significant learning estimations reliant upon dataset were presented moderately. Results show that the significant learning estimation performed generally best results over SVM, ANN, RF. It

occurs in the manner that when we think about long back a long time there might be such countless assaults occurred so when these assaults are perceived then the highlights at which esteems these assaults are going on will be put away in some datasets. So by utilizing these datasets we will anticipate if digital assault is finished. These forecasts should be possible by three calculations like SVM, ANN, RF this paper assists with distinguishing which calculation predicts the best precision rates which assists with foreseeing best outcomes to recognize the digital assaults occurred or not.

Machine learning can intelligently identify previously unknown forms of malware and attacks to help protect organizations from potential zero-day attacks. Insights at scale: With data and application in many different locations, being able to identify trends across large volumes of devices is just not humanly possible.

REFERENCES

1. Ibrahimi and M. Ouaddane, "Management of intrusion detection systems based-kdd99: Analysis with lda and pca," in Wireless Networks and Mobile Communications (WINCOM), 2017 International Conference on. IEEE, 2017, pp.1–6.
2. Girish L, Rao SKN (2020) "Quantifying sensitivity and performance degradation of virtual machines using machine learning.", Journal of Computational and Theoretical Nanoscience, Volume 17, Numbers 9- 10, September/October 2020, pp. 4055-4060(6)https://doi.org/10.1166/jctn.2020.9019
3. 2020 11th International Conference on Information and Knowledge Technology (IKT) December 22-23, 2020; Shahid Beheshti University - Tehran, Iran in Detection and Identification of Cyber-Attacks in Cyber-Physical Systems Based on Machine Learning Methods
4. 4 International Journal of Advanced Scientific Innovation, Volume 01 Issue 02, May 2021, ISSN: 2582-8436: Detection of Cyber Attack in Network using Machine Learning Techniques.
5. Francois Chollet, Google AI researcher and creator of the popular Keras deep learning library, published his book, Deep Learning with Python in October 2017.
6. Deep Learning for Computer Vision with Python by Dr. Adrian Rosebrock
7. A Guide to MATLAB for beginners and experienced users by Brain R. Hunt, Ronald L. Lipsman, Jonathan M. Rosenberg, Kevin R. Combes, John E. Osborn, Garrett J. Stuck
8. https://www-geeksforgeeks-org.cdn.ampproject.org/v/s/www.geeksforgeeks.org/deep-learning-with-python-opencv/amp/?amp_gsa=1&amp_js_v=a9&usqp=mq331AQKKAFQArABIIACAw%3D%3D#amp_tf=From%20%251%24s&aoh=16622602770557&referrer=https%3A%2F%2Fwww.google.com&ampshare=https%3A%2F%2Fwww.geeksforgeeks.org%2Fdeep-learning-with-python-opencv%2F
9. https://www.tutorialspoint.com/dip/index.htm
10. https://www.geeksforgeeks.org/python-programming-language/