

Detection of Cyber Attacks and Network Attacks Using Machine Learning Algorithms

Rohit Khedkar¹, Ganesh Mahajan², Mitali Bhujbal³, Kirti karade⁴, Prof. S K Hiremath⁵

^{1,2,3,4} Students and ⁵ Professor of Department of Computer Engineering, JSPM, Jayawantrao Sawant College of Engineering, Savitribai Phule Pune University, Pune

Abstract: Now a days cyber crime growing and has a big effect everywhere globally. ethical hackers are normally involved in identifying flaws and recommending mitigation measures. the cyber safety international, there's a pressing need for the improvement of powerful techniques. Because of the effectiveness of machine learning in cyber security issues, machine learning for cyber security has recently become a hot topic. In cyber security, machine learning approaches have been utilized to handle important concerns such as intrusion detection, malware classification and detection, spam detection, and phishing detection. Although ML cannot fully automate a cyber-security system, it can identify cyber-security threats more efficiently than other software-oriented approaches, relieving security analysts of their burden. As a result, effective adaptive methods, such as machine learning techniques, can yield higher detection rates, lower false alarm rates, and cheaper computing and transmission costs. Our key goal is that the challenge of detecting attacks is fundamentally different from those of these other applications, making it substantially more difficult for the intrusion detection community to apply machine learning effectively. In this study, the CPS is modeled as a network of agents that move in unison with one another, with one agent acting as a leader and commanding the other agents. The proposed strategy in this study is to employ the structure of deep neural networks for the detection phase, which should tell the system of the attack's existence in the early stages of the attack. The use of robust control algorithms in the network to isolate the misbehaving agent in the leader-follower mechanism has been researched. Following the attack detection phase with a deep neural network, the control system uses the reputation algorithm to isolate the misbehaving agent in the presented control method. Experiment results show that deep learning algorithms can detect attacks more effectively than traditional methods, making cyber security simpler, more proactive, and less expensive and more expensive.

Keywords: Network Protocols, Wireless Networks, Cyber-crime, Machine learning techniques, cyber-security systems, attacks, SQL Injection, Cross-Site Scripting (XSS), Phishing Attacks, and Intrusion Detection Attacks (IDS), etc.

1. Introduction

In this modern era of statistics and verbal exchange technology, physical items at the moment are related to every different through cyber networks and are together referred to as a cyber bodily system. The stateful firewall, also known as assault detection and prevention, identifies and blocks assaults in network traffic. An exploit can be a statistics-gathering probe

or an assault aimed toward compromising, disabling, or harming a community or network aid. the line between the 2 making the most targets may be hazy in some circumstances. for instance, a barrage of TCP SYN segments might be an IP coping with a sweep with the reason of triggering responses from lively hosts, or it might be an SYN flood attack with the rationale of overwhelming a network so that it can now not feature well. furthermore, due to the fact an attacker typically precedes an assault via acting reconnaissance on the target, we can recall records-gathering efforts as a precursor to a forthcoming assault—this is, they represent the primary level of an attack. consequently, the period makes the most encompasses both reconnaissance and attack activities, and the difference between the 2 isn't always usually clear. The net and laptop networks have turned out to be an important part of our corporations and normal lifestyles. With the boom in our dependence on computers and communique networks, malicious sports have grown to be an increasing number of standards. Network attacks are an important trouble in nowadays's communique environments. The network traffic must be monitored and analyzed to come across malicious activities and attacks to ensure the reliable functionality of the networks and the safety of customers' information.

These days, gadget-mastering techniques had been implemented for the detection of community attacks. gadget studying fashions can extract similarities and styles among the community visitors. in contrast to signature-based totally techniques, there's no need for manual analyses to extract assault patterns. applying gadgets gaining knowledge of algorithms can robotically build predictive fashions for the detection of community assaults. New threats and difficulties to wifi conversation systems have advanced due to the development of 5th-generation networks and artificial intelligence technologies, specifically in cyber security. We offer an overview of assault detection techniques utilizing the energy of deep mastering strategies on this gadget. specifically, we first summarize the fundamental troubles of network security and attack detection and introduce numerous successful associated programs with the use of deep gaining knowledge of shape. We recognition on assault detection systems built on several forms of architectures, including auto-encoders, generative adverse networks, recurrent neural networks, and convolutional neural networks, based totally on the type of deep studying methodologies. Following that, we give some benchmark datasets with descriptions and evaluate the overall performance of numerous representation approaches to illustrate the modern state of attack detection strategies using deep studying structures. subsequently, we summarize this work and discuss some ways to improve the overall performance of assault detection under thoughts of using deep getting-to-know structures.

2. Literature Survey

1. Nutjahan, Farhana Nizam, Shudarshon Chaki, Shamim Al Mamun, M. Shamim Kaiser, "Assault Location and Avoidance within the Cyber-Physical System". [2016] [1] In this paper proposes Cyber-Physical System cyber-attack detection and prevention To detect distributed denial of service and false data injection attacks, the Chi-square detector and Fuzzy logic-based attack classifier (FLAC) were utilized. Activity profiling, average packet rate, change

point detection algorithm, cusum algorithm, unexpired user sessions, injected incomplete information, and reuse of session key are some of the fuzzy features used to choose the attacks described. An illustration situation has been made utilizing the OpNET Test system.. Chi-square detectors and FLAC can detect cyber-physical attacks with high accuracy, according to simulation results.

2. .Yong Fang, Cheng Huang, Yijia Xu, and Yang Li, “RLXSS: Optimizing XSS Location Show to Protect Against Antagonistic Assaults Based on Fortification Learning”. [2019] [2]. In this research, we introduce RLXSS, a reinforcement learning-based strategy for optimizing the XSS detection model to defend against adversarial attacks. To begin with, the ill-disposed tests of the discovery show are mined by the ill-disposed assault demonstration based on fortification learning. Besides, the discovery demonstration and the ill-disposed show are then again prepared. After each circular, the newly-excavated antagonistic tests are stamped as malevolent tests and are utilized to retrain the location demonstrate. The proposed RLXSS model successfully mines adversarial samples that avoid black-box and white-box detection while retaining aggressive features, according to experimental data. Furthermore, by alternating training between the detection model and the confronting assault model, the detection model's escape rate is continuously reduced, indicating that the model can increase the detection model's ability to defend against attacks.

3. Rishikesh Mahajan, Irfan Siddavatam, “Phishing Website Detection using Machine Learning Algorithms”. [2018] [3] Phishing is the foremost fundamental strategy for getting delicate data from clueless shoppers. The objective of phishers is to get touchy data such as usernames, passwords, and bank account data. Cybersecurity professionals are now looking for dependable and consistent detection solutions for phishing websites. The purpose of this work is to discuss machine learning technology for detecting phishing URLs by extracting and analyzing various aspects of authentic and phishing URLs. To detect phishing websites, the Decision Tree, Random Forest, and Support Vector Machine algorithms are used. The goal of this study is to detect phishing URLs as well as to narrow down the best machine learning method by analyzing each algorithm's accuracy rate, false positive and false negative rate.

4. Vishnu. B. A, Ms. Jevitha. K. P., “Prediction of Cross-Site Scripting Assault Utilizing Machine Learning Algorithms”. [2018] [4] Cross-site scripting (XSS) is one of the most frequently occurring types of attacks on web applications, hence is of importance in information security. XSS occurs when an attacker injects malicious code, usually JavaScript, into a web application such that it can be executed in the user's browser. Detecting malicious scripts is an important aspect of an online application's defense. This study studies the use of SVM, k-NN, and Random Forests to detect and limit known and undiscovered assaults on JavaScript code by developing classifiers. It has shown that using an interesting feature set that combines language syntax and behavioral information resulted in classifiers that provide excellent accuracy and precision on huge real-world data sets without focusing solely on

obfuscation.

5. Zohre Nasiri Zarandi, Iman Sharif, "Detection and Distinguishing proof of Cyber-Attacks in Cyber-Physical Frameworks Based on Machine Learning Methods". [2020] [5] The CPS is modeled in this study as a network of agents that move in unison with one another, with one agent acting as a leader and the other agents being ordered by the leader. In this study, the proposed strategy is to employ the structure of deep neural networks for the detection phase, which should tell the system of the existence of the attack control method, after the attack detection phase with the use of a deep neural network, the control system uses the reputation algorithm to isolate the misbehave agent. Experiments reveal that deep learning algorithms outperform traditional approaches in detecting assaults, making cyber security simpler, more proactive, less expensive, and considerably more successful.

3. Objectives Of The Project

The goals of the machine are-

- * to overcome these shortcomings, there's a want to accumulate representative intrusion detection facts to increase and examine detection mechanisms for pc community attacks.
- * Similar to a consultant's regular facts, it should also contain a right diversity of various kinds of assaults.
- * To discover network assaults via making use of machine learning techniques.
- * To reduce operational time.
- * To expanded accuracy and reliability.
- * To accelerate operational performance.
- * To provide information security.

4. Scope Of The Project

Device learning is a subfield of pc science, which makes use of sample popularity and synthetic intelligence strategies to organize and extract behaviors and entities from the records. these formerly regarded styles and relationships trained through gadget-getting-to-know algorithms may be used to do prediction tasks on new records. With these days' technology, systems gaining knowledge of algorithms contact our ordinary lifestyles by being used in a huge range of applications.

This venture has a massive scope because it has the following functions which help in making it easy to use, recognize and adjust it:

- * Easy to detection of cyber and community attacks.
- * No want to do separate configurations to deal with attacks.
- * To keep the environment with the aid of the use of gadgets gaining knowledge of strategies
- * To boom the accuracy and efficiency of the assaults detection system.
- * control of Kaggle datasets and their feature choice.

5. Proposed System

Cybercrime is spreading throughout the world, using any sort of weak spot in the computing environment. ethical hackers are usually involved with assessing vulnerabilities and offering mitigation techniques. The improvement of effective techniques is a pressing need in the cybersecurity community. maximum techniques used in nowadays's IDS aren't capable of addressing the dynamic and complicated nature of cyber-assaults on laptop networks. gadget studying for cyber safety has emerged as a trouble of first-rate significance recently because of the effectiveness of system learning in cyber safety issues. ML tactics have been used to deal with critical problems in cyber security, including intrusion detection, malware type and detection, junk mail detection, and phishing detection. even though machines gaining knowledge cannot fully automate a cyber-security device, it can identify cyber safety threats with greater effects than other software-orientated methods, easing the pressure on security analysts. As a result, effective adaptive procedures, which include machine-gaining knowledge of strategies, can bring about higher detection costs, decrease fake alarm charges, and cheaper computing and transmission prices. Our principal intention is to show that the trouble of detecting assaults is fundamentally one of a kind from those other programs, making it far extra hard for the intrusion detection network to properly use machine-gaining knowledge of gadgets gaining knowledge of algorithms may be used to teach and hit upon if there was a cyber assault. As quickly as the assault is detected, an e-mail notification can be dispatched to the safety customers. Any category set of rules may be used to decide whether or not or not an attack is a DoS/DDoS assault. assist Vector system (SVM), a supervised gaining knowledge of method that analyses information and recognizes styles, is one example of a category set of rules. for the reason that we can not expect whilst, when, or how an assault will occur, and absolute prevention can not be assured, our excellent guess in the meantime is early discovery, which allows you to assist reduce the danger of irreparable harm such occurrences can do. corporations can use current answers or build their very own to come across cyber-assaults at a very early stage to minimize the effect. Any gadget that calls for minimum human intervention could be ideal.

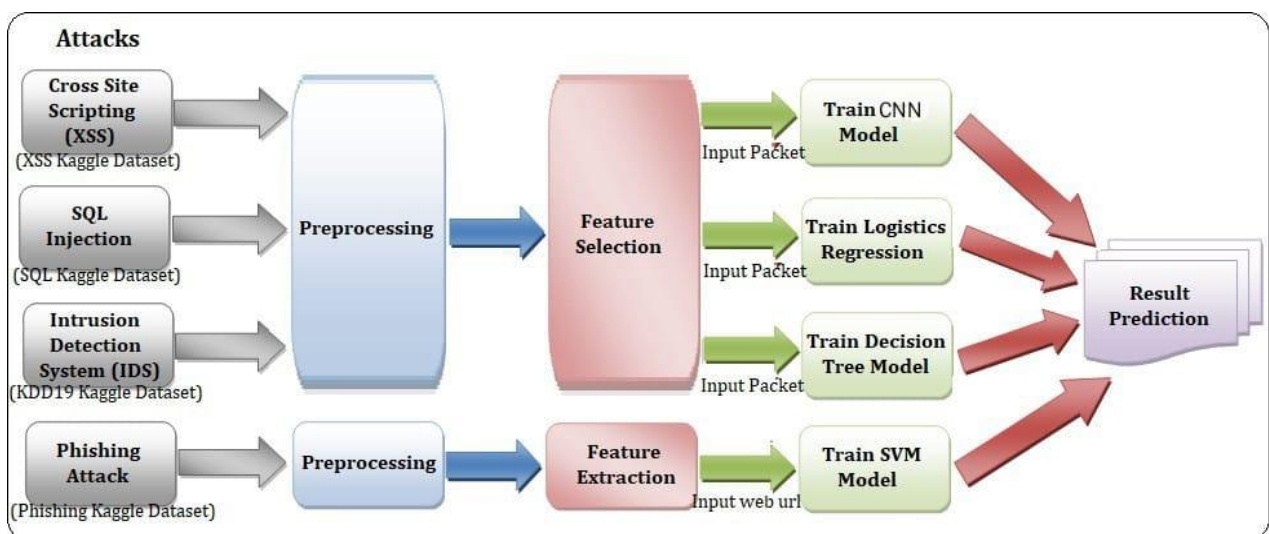


Fig.1: System Architecture

5.1 Preprocessing

Fact preprocessing is a crucial step within the information mining procedure that includes manipulating or dropping statistics earlier than it is used to make certain or improve overall performance. In records mining and device studying initiatives, the word "garbage in, trash out" is specifically apt. information collection strategies are frequently uncontrolled, resulting in out-of-range values (for example, earnings: a hundred), not possible information mixtures (for instance, intercourse: Male, Pregnant: sure), and lacking values, amongst other things. reading records that haven't been thoroughly checked for such issues can cause false conclusions. As a result, earlier than doing any evaluation, the representation and best information ought to come first. facts preprocessing is frequently the most crucial level of a machine mastering task. as an example: on this mission for Detection of pass site Scripting(XSS) attacks, some characters have values very huge eg 8221, and some are Chinese letters, so we're getting rid of letters having values extra than 8222 and for the rest, we will be considering values extra than 128 and much less than 8222 and assign the values of the one so that they can be normalized. For the Intrusion Detection attack, we can be extracting numerical attributes and could scale them to have zero mean and unit variance. in addition, we can flip the result back into a data frame. Then we can extract categorical attributes from both education and take a look at units, encode the categorical attributes, and separate the goal column from encoded information.

5.2 Feature Selection And Feature Extraction

In the case of a Network Intrusion Attack, for extracting important features we have used Random Forest Classifier. Here we have extracted the following 15 attributes from the dataset :

```
['src_bytes','dst_bytes', 'logged_in', 'count', 'srv_count', 'same_srv_rate', 'diff_srv_rate',  
'dst_host_srv_count', 'dst_host_same_srv_rate','dst_host_diff_srv_rate',  
'dst_host_same_src_port_rate','dst_host_srv_diff_host_rate', 'protocol_type', 'service', 'flag']
```

In case of Phishing attacks, we have extracted the following features are,

- 1) Address bar-based features like Using IP addresses, Long URLs to hide suspicious parts, URLs having @ symbol, Redirecting using // etc
- 2) Abnormal-based features like URL of Anchor, Links in <meta>,<script> and <link> tags, server form handler (SFH), submitting information to email, etc
- 3) HTML and JavaScript-based features like website forwarding, status bar customization, disabling right-clicking, using pop-up windows, etc.
- 4) Domain-based features like age of the domain, DNS record, website traffic, Pagerank, etc.

5.3 Different Training Models

- CNN Model is used to detect Cross-Site Scripting (XSS) attacks.
- Logistics Regression Model is used to detect SQL Injection attacks.
- The decision tree model is used to detect Intrusion Detection (IDS) attacks.

- SVM Model is used to detect phishing attacks.

5.4 Result Prediction

For IDS if the output is an anomaly then it will be considered an attack, on the other hand, if the output is normal then it is a legitimate packet.

For SQL Injection, Phishing attacks, and Cross Site Scripting attacks the output is in the format of 0 and 1, where 0 is not an attack and 1 will be considered malicious.

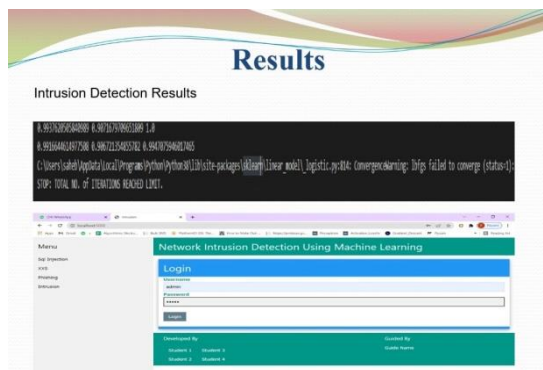


Fig.2: ABC

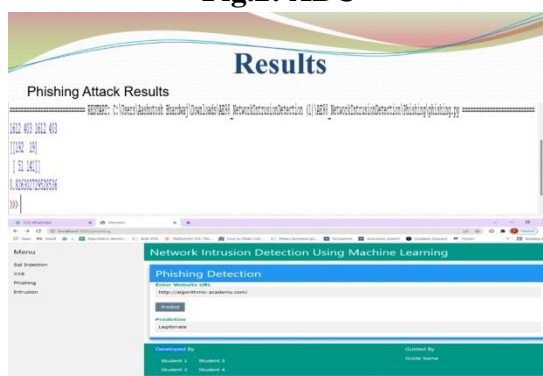


Fig.4: ABC

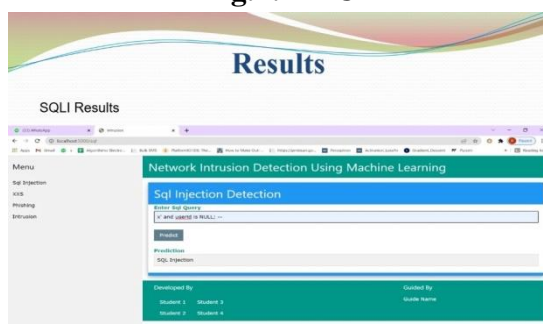


Fig.6: ABC

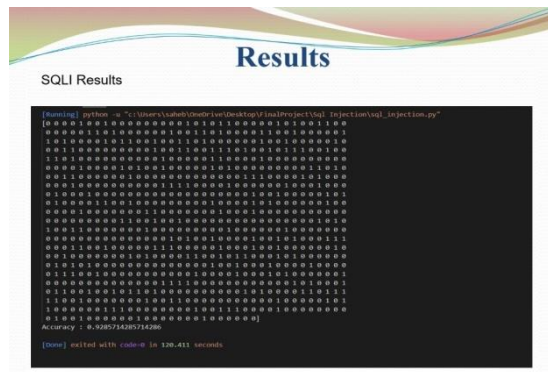


Fig.3: ABC

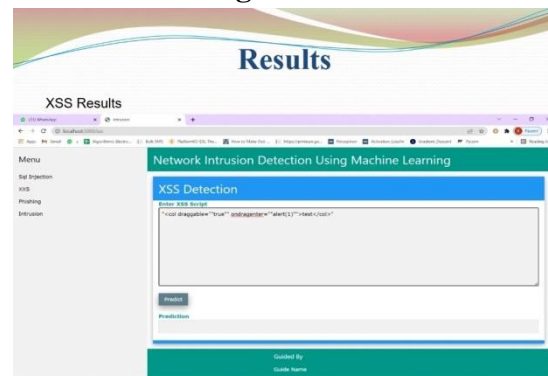


Fig.5: ABC



Fig.7: ABC

5.5 Math's

A.SQL Injection:

SQL injection, also called SQLI, is a commonplace assault vector that makes use of malicious sq. code for backend database manipulation to access statistics that turned into not supposed to be displayed. This fact can also consist of any range of items, which include sensitive employer statistics, person lists, or non-public consumer info. Examples of square injection:

- o Retrieving hidden facts, wherein you may adjust an sq. question to return extra results.
- o Subverting application logic, in which you could change a question to intervene with the utility's good judgment.
- o UNION assaults, wherein you can retrieve statistics from extraordinary database tables. analyzing the database, where you can extract information approximately the version and structure of the database. Blind SQL injection, in which the effects of a query you manage aren't returned in the software's responses. To clear up this

attack we use the Logistic Regression version to train in the machine learning platform.

B.Cross Site Scripting:

go-web page scripting assaults, additionally known as XSS assaults, are a sort of injection assault that injects malicious code into otherwise safe websites. An attacker will use a flaw in a goal internet software to send some type of malicious code, most commonly patron-aspect JavaScript, to a given-up person. place of concentrating on the software's host itself, XSS attacks normally target the software's customers immediately. businesses and companies strolling web packages can depart the door open for XSS assaults if they show content material from users or untrusted resources without proper escaping or validation.

- A. XSS occurs when an attacker tricks a web application into sending information in a shape that a consumer's browser can execute. maximum usually, this is an aggregate of HTML and XSS provided by way of the attacker, however, XSS also can be used to deliver malicious downloads, plugins, or media content material.

C.Phishing Attack:

Phishing assaults are the exercise of sending fraudulent communications that seem to return from a reputable source. it is also accomplished through email. The purpose is to scouse borrow touchy statistics like credit score card and login information, or to put in malware at the sufferer's gadget.

The message is made to appear as though it comes from a dependent sender. If it fools the sufferer, she or he is coaxed into offering exclusive data, regularly on a rip-off website. occasionally malware is also downloaded onto the target's computer.

Phishing is a form of social engineering assault regularly used to thief user data, which includes login credentials and credit score card numbers. It happens whilst an attacker, masquerading as a trusted entity, dupes a victim into opening an e-mail, instantaneous message, or textual content message. The recipient is then tricked into clicking a malicious link, which could lead to the set up of malware, the freezing of the device as a part of a Ransomware assault, or the unveiling of sensitive information.

D.Intrusion Detection System:

An intrusion detection system (IDS) is a tool or software program application that video displays units on a network for malicious activity or policy violations. Any malicious pastime or violation is normally reported or amassed centrally with the use of a security facts and event control device. Intrusion detection structures are designed to perceive suspicious and malicious pastimes via community visitors, and an intrusion detection system (IDS) allows you to discover whether your network is being attacked. Intrusion detection structures are used to detect anomalies to catch hackers before they do real harm to a network. They can be either community- or host-based. ... Intrusion detection structures work using both seeking out signatures of recognized assaults or deviations from the normal hobby. To remedy this attack we use the selection Tree version to educate in the system mastering platform.

6. Results And Evaluation

Attack	Algorithm	Accuracy
Intrusion Detection	Decision Tree	99.47%
	KNN Classifier	99.16%
	BNB Classifier	90.67%
SQL Injection Attack	Logistic Regression	92.85%
Cross Site Scripting Attack (XSS)	Convolutional Neural Network	98.59%
Phishing Attack	Support Vector Machine	82.63%

Table.1: Comparative Analysis

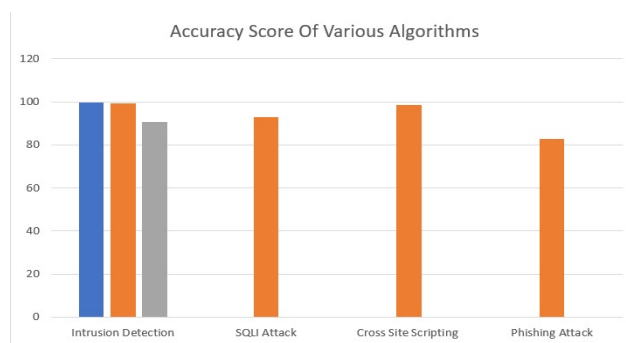


Fig.8: Accuracy Score of various algorithms

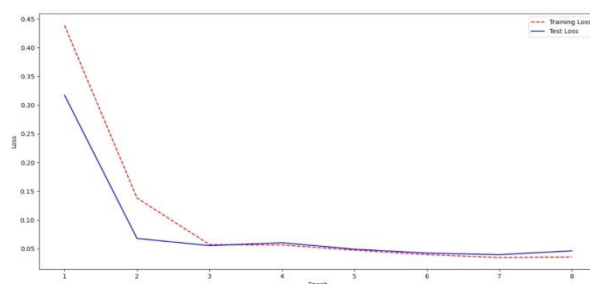


Fig.9: ABC

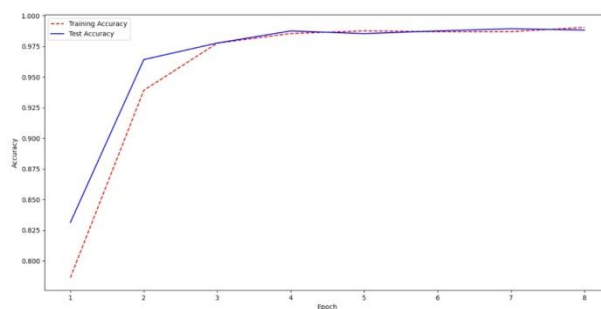


Fig.10: ABC

7. Limitations

- 1) No attack detection software is present at all.
- 2) Time consumed in accessing the attacks in real-time.
- 3) Manual creation of lists for various activities.
- 4) The general problem throughout this work is the detection of network attacks.
- 5) The problem definition gets more specific for any attack type and includes an expanded

definition of the attack and its behavior.

8. Conclusion

In this study, an attempt was made to use the resilient control consensus method in complex discrete cyber-physical networks with several local attacks off. 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. Also in a recurrent neural network integrated with a deep neural network, a deep layer network with a linear function performs better. So With the deep learning method, systems can analyze patterns and learn from them to help prevent similar attacks and respond to changing behavior. To summarise, ML has the potential to make cyber security simpler, more proactive, less expensive, and considerably more successful. After observing the state of the system reported by the neural network, the control system makes decisions based on it and, if there is an attack, detects it and isolates it, so as not to have a detrimental effect on the behavior of other agents. As a result, effective adaptive approaches, such as machine learning techniques, can result in higher detection rates, lower false alarm rates, and cheaper computing and transmission costs. We reviewed several influential algorithms for attack detection based on various ML techniques. Because of the characteristics of ML approaches, it is feasible to construct attacks with high detection rates and low false positive rates, while the system rapidly adapts to changing hostile behaviors. One thing is sure, any organization failing to adopt these techniques now or in the immediate future risk compromising data or worse servers.

Acknowledgements

We would prefer to give thanks to the researchers likewise publishers for creating their resources available. We are conjointly grateful to the guide, and reviewer for their valuable suggestions and also thank the college authorities for providing the required infrastructure and support.

References

- [1] Z. N. Zarandi and I. Sharifi, "Detection and Identification of Cyber-Attacks in Cyber-Physical Systems Based on Machine Learning Methods," 2020 11th International Conference on Information and Knowledge Technology (IKT), 2020, pp. 107-112, doi: 10.1109/IKT51791.2020.9345627.
- [2] Nurjahan, F. Nizam, S. Chaki, S. Al Mamun and M. S. Kaiser, "Attack detection and prevention in the Cyber-Physical System," 2016 International Conference on Computer Communication and Informatics (ICCCI), 2016, pp. 1-6, doi: 10.1109/ICCCI.2016.7480022.
- [3] Ding Chen, Qiseng Yan, Chunwang Wu, and Jun Zhao, "SQL Injection Attack Detection and Prevention Techniques Using Deep Learning," Journal of Physics: Conference Series, Volume 1757, International Conference on Computer Data and Artificial Intelligence (ICCBDAI 2020) October 2020, Changsha, China

- [4] Ercan NurcanYılmaz, SerkanGönen, “Attack detection/prevention system against cyber-attack in industrial control systems,” *Computers & Security* Volume 77, August 2018, pp 94-105
- [5] Arpitha. B, Sharan. R, Brunda. B. M, Indrakumar. D. M, Ramesh. B. E, “Cyber Attack Detection and notifying system using ML Techniques,” *International Journal of Engineering Science and Computing (IJESC)*, Volume 11, Issue No.06
- [6] Yirui Wu, Dabao Wei, and Jun Feng, “Network Attacks Detection Methods Based on Deep Learning Techniques: A Survey,” *Security Threats to Artificial Intelligence-Driven Wireless Communication Systems*, 2020.
- [7] Rafał Kozik, Michał Choraś, “Machine Learning Techniques for Cyber Attacks Detection,” *Image Processing and Communications Challenges 5*, pp 391-398, Springer International Publishing Switzerland 2014.
- [8] Nutjahan, Farhana Nizam, Shudarshon Chaki, Shamim Al Mamun, M. Shamim, “Attack Detection and Prevention in the Cyber Physical System,” *2016 International Conference on Computer Communication and Informatics (IEEE -2016)*, Jan. 07 - 09, 2016, Coimbatore, India
- [9] Anupong, W., Yi-Chia, L., Jagdish, M., Kumar, R., Selvam, P. D., Saravanakumar, R., & Dhabliya, D. (2022). Hybrid distributed energy sources providing climate security to the agriculture environment and enhancing the yield. *Sustainable Energy Technologies and Assessments*, 52 doi:10.1016/j.seta.2022.102142
- [10] Yong Fang, Cheng Huang, Yijia Xu and Yang Li, “RLXSS: Optimizing XSS Detection Model to Defend Against Adversarial Attacks Based on Reinforcement Learning,” *Future Internet* 2019.
- [11] Pratik Rajendra Chougule, Aniket Sanjay Kumbhar, Vinayak Vasant Pachange, Karan Dinkar Phonde, S. P. Phadtare, “Phishing Websites Detection using Python,” *Journal of Web Development and Web Designing*, Volume-5, Issue-2 (May-August, 2020)
- [12] Rishikesh Mahajan, Irfan Siddavatam, “Phishing Website Detection using Machine Learning Algorithms,” *International Journal of Computer Applications (0975 – 8887)* Volume 181 – No. 23, October 2018
- [13] Vishnu. B. A, Ms. Jevitha. “Prediction of Cross-Site Scripting Attacks Using Machine Learning Algorithms,” *Conference Paper • October 2014*.
- [14] Aoudni, Y., Donald, C., Farouk, A., Sahay, K. B., Babu, D. V., Tripathi, V., & Dhabliya, D. (2022). Cloud security based attack detection using transductive learning integrated with hidden markov model. *Pattern Recognition Letters*, 157, 16-26. doi:10.1016/j.patrec.2022.02.012
- [15] Shinelle Hutchinson, Zhaohe Zhang, and Qingzhong Liu, “Detecting Phishing Websites with Random Forest,” *Third International Conference, MLICOM 2018, Hangzhou, China, July 6-8, 2018, Proceedings*
- [16] Ines Jemal, Omar Cheikhrouhou, Habib Hamam and Adel Mahfoudhi, “SQL Injection Attack Detection and Prevention Techniques Using Machine Learning,” *International Journal of Applied Engineering Research* ISSN 0973-4562 Volume 15, Number 6 (2020) pp. 569-580