Webology, Volume 18, Number 4, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I4/142

USING MACHINE LEARNING FOR CYBER SECURITY ENHANCEMENT

Noor Mohd[1], Shruti Bhatla[1], Deepak Upadhyay[2]

[1]Department of Computer Science & Engineering Graphic Era Deemed to be University, Dehradun, Uttarakhand, India
[2]Department of Computer Science & Engineering, Graphic Era Hill University, Dehradun, Uttarakhand India, 248002

ABSTRACT

Cyber security is a major problem of modern society since Vulnerabilities of computer Network has become easy with the help of technologies and human skills. Currently different type of attacks are occurring for example DOS attack, Probing, R2U, R2L virus, port scans, buffer overflow, CGI Attack and flooding etc. We need a platform where a system can be developed for recognition and prevention of these attacks. In This paper, most of the latest methods are summarized to implement IDS for cyber security. Intrusion Detection Systems is a most suitable solution for cyber-attacks. Machine learning based Intrusion Detection Systems have high accuracy, in rapidly changing environment. This paper also discusses the ML technique with the lowest accuracy and explores some research area for researchers.

Keywords: Cyber security, Machine learning, Intrusion detection system

INTRODUCTION

Cyber criminals have a big advantage in the cyber war, out of many attacks, attacker needs one right attempt, and for security personal needs success rate 100%. Researchers show that in 2018, many businesses, individuals, organizations and company were victimized by cyber criminals [1]. Stolen data include intelligence data, financial records, and personal data, for detecting attackers attempts, they are successful or not, Intrusion detection plays an vital role in the network security and forensic analysis [2], and it can detect many types of attacks, however, Internet environment increasing networks complexity, its structure, and diverse network model, instead of it attackers are also updating technology for attacks, so traditional IDS is difficult to meet security needs. We need an advance IDE for detection and prevention network attacks. Machine Learning techniques are one of the famous techniques used for detecting network attacks.

Artificial intelligence has many branches Machine learning is one of them, It has the capability to self learning on the basis of previous data and it can improve system automatically without being explicitly programmed [3]. ML techniques depends on mathematical model and take decision after analysing patterns in datasets, after that IDS predict result for new inputted data. Machine Learning has many application and span across a vast area. Including ecommerce, where ML used to recommend customer based on their behaviour, health care where ML application are used for

Webology, Volume 18, Number 4, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I4/142

recommending customer based on patient symptoms. Machine learning algorithms are divided into 3 types.

SUPERVISED LEARNING:- Main function of supervised learning is that learn a function which map input data to output data, based on input-output pairs. It predicts a function from labelled data. Some supervised learning algorithms are artificial neural network, Regression, Bayesian Statistics, Gaussian. Decision Tree, Support Vector Machine, Bayesian Statistics, Preceptor, Gaussian, Random forest, K-nearest neighbour and Naive Bayes.

UNSUPERVISED LEARNING:- Unsupervised learning Techniques use unlabelled data instances. Clustering is used by this technique. Some of the common unsupervised learning methods are Cluster analysis, Apriori algorithm, Eclat algorithm and Outlier detection [4].

REINFORCEMENT LEARNING:- In this technique system interact with an environment to achieve goal. This approach ask user to set label for unlabeled instances.

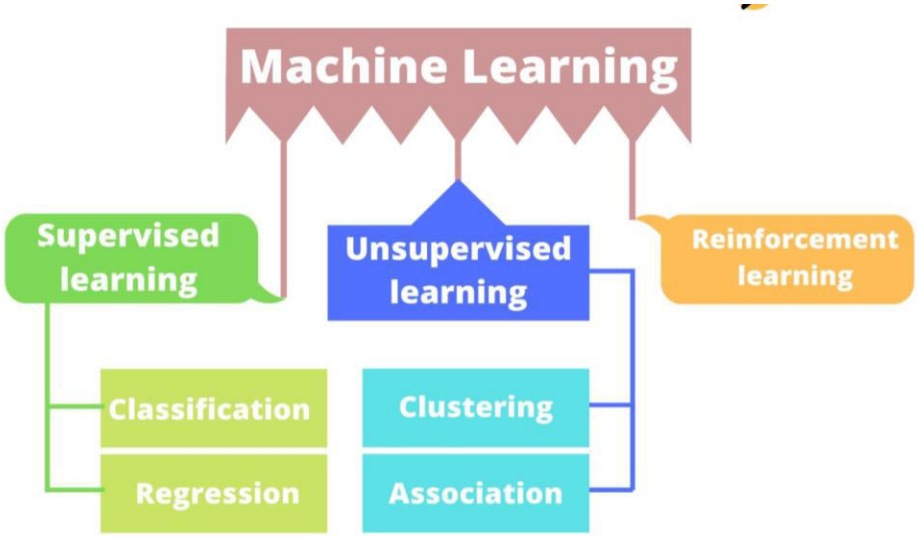

Figure 3: Type of Machine Learning Algorithms

CAUSE, Types and solution for CYBER ATTACKS

Cyber criminals select easiest way to earn big money. Mostly they target bank, MNC's, financial firms where chances of knowing sensitive information get increased, to Identify these type of criminal is important, so cyber low is introducing across the global for such type of activities. Some reasons are below for vulnerability of computer.

- **Easy to access** – Main problem to restrict hackers to access unauthorized gain on machine, so that there are many possibilities of breach the security due to complex technology. Hacker can find codes, retina images, and voice recorder or even can make duplicate biometric system and can bypass firewall. It became is easy to gain access on System or network.

Webology, Volume 18, Number 4, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I4/142

- **Storing data in small space**– All computers store sensitive data in a small space. Reasoned being it becomes easy for hackers to copy that data into other secondary devices for own profit.
- **Complex** – Computers run on Operating System, And an Operating System programmed by thousands of lines. For human mind it is not easy to remove all gaps or threads From Operating System. Because of this hacker uses these gaps for their own profit.
- **Negligence** – Negligence of human mind is a common nature; it leaves a path for hacker to get access on machine. Programmer tries to develop an error less pc but still there may be a bug in Operating System.
- **Loss of evidence** – Data related to cyber crime can be easily removed by expert hacker reason being forensic investigation of cyber crime becomes a challenging task.

Cyber attacks or hacking is an act that tends to harm secure data, steal sensitive information and interrupt digital life. Cyber attackers use many techniques like computer viruses, Ransomware, Spyware, Installs malware, Phishing, SQL injection, Denial of Service. Since last few decades different type of business around the world become victim to hackers including companies such as HSBC, Sony resulting in thousands of records related to consumers got exposed. These market related to exposed company get failed. Name of Cyber Attacks are given below.

1. **Probing Attack**:- Probing is a technique for attacks where attacker collect data or find possible vulnerabilities of computer networks. In the network many services and path of computers are available for attacks. In probe techniques some use social engineering, it is commonly heard, and can be used with a small expertise.
2. **Denial of service attacks**:- DOS is another technique for attackers, in which targeted systems memory got busy by Buffer overflow attacks, ICMP flood and SYN flood. Resulting of this legitimate user enable to access machine. Attackers focus of targeted implementations bugs, or by exploiting the system's miss configuration bugs.
3. **User to root attacks**:- In these techniques attackers try to root access to the local system. And exploit the information by unauthorized access.
4. **Remote to user attacks**- In a remote to user attack, attackers send packets to a targeted machine through network and try to gain access on remote machine. There are many techniques for R2L in social engineering is most commonly used. Now problem is that how to protect user's data or personal information and restricts attackers to enter into the network.

Their May be many possible solution of cyber crime, Machine learning is one of them, and Machine Learning have many defensive method and algorithms for cybercrime. Machine learning is a sub part of Artificial Intelligence. It work on training data set and these training data set depends on known facts from past experience. Prediction is the main task of Machine Learning. Machine learning's methods are divided into 3 types: - supervised learning, unsupervised learning, and reinforcement learning, by using one of the algorithms from these categories some type of cyber attacks can be detect.

Intrusion Detection System is an active device which analyses and scans the network activity and detects any unauthorized access, spam and viruses. [10,11,12] talks about intrusion detection system and machine learning algorithms If any then send alert signal. IDS may be software, hardware or a

Webology, Volume 18, Number 4, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I4/142

combination of these two. IDS have only one goal to catch attacker in the act before they do real damage to information or data. IDS secure a network from attackers. It monitors network, audit network and configure for vulnerabilities and analyze network data, IDS is a important component in the network security toolbox, An IDS gives three important functions: monitoring, detection and generating a signal. Intelligent IDS techniques include Machine Learning, Genetic Algorithm, Support Vector Machine [5], Decision Tree, and Artificial Neural Network, For testing these algorithms, we need Data set and Intrusion Detection Systems. Collecting data from Computer network is very time consuming, developer test their IDS using available dataset. These Dataset contain all type of possible tested Data and training data, 80% of data is related to attacker's.

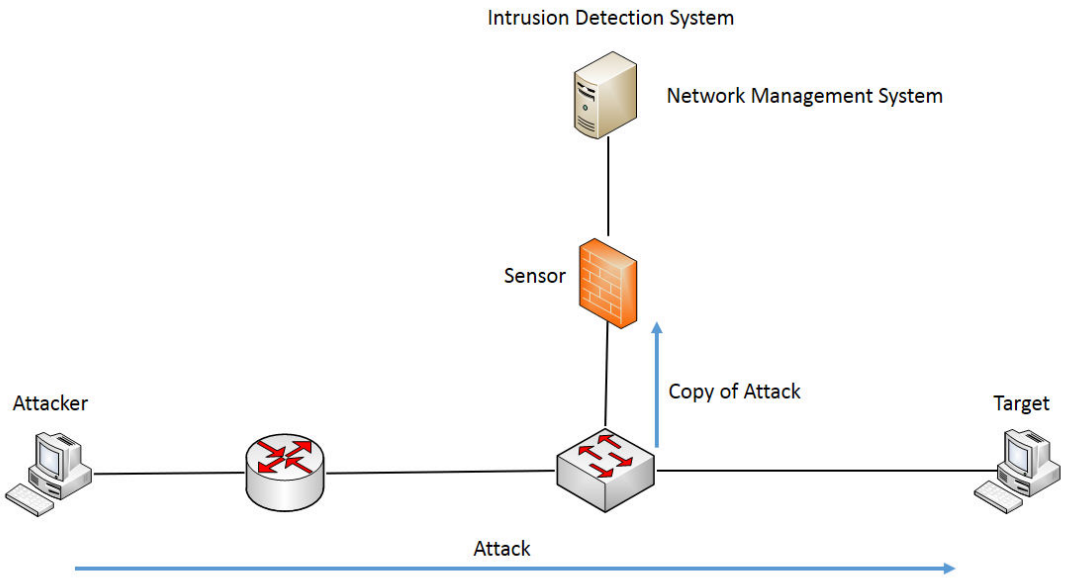

Figure1: Intrusion Detection System in a Artificial Neural Network

According to Gozde Karatas [6] IDS can be implemented by following ML Techniques:

- Artificial Neural Network
- Support vector Machines
- Data Mining
- Rule-Based System
- Fuzzy logic
- Statistical

Datasets: - To determine the performance of IDS developer needs a dataset, Dataset is a collection of different type of attacks on the basis IDS check its performance. Now day's lot of Datasets are available For example:

- KDD cup99
- NSL-KDD
- CIC IDS 2017
- CIC IDS 2017
- CSE-CIC IDS 2017

Webology, Volume 18, Number 4, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I4/142

- Benign with MCFP Bot Traffic

ANALYSIS OF MACHINE LEARNING METHOD ON DIFFERENT ATTACKS

Some of the algorithms are compared here with different cyber-attacks and summary of those attacks where ML algorithm is not used.

	Probing	DoS	U2R	R2L	Phishing attack [7]
Decision Tree	No	yes	yes	No	yes
Naive Bayes	yes	yes	yes	Yes	No
K-nearest neighbour	No	yes	No	No	No
Artificial Neural Network	Yes	Yes	No	No	No
Support Vector Machines	yes	yes	Yes	Yes	No
Fuzzy Logic	yes	yes	Yes	Yes	No
Particle Swarm Optimization Algorithm	Yes	Yes	Yes	Yes	No

Figure 1: ML Algorithms applied for Cyber Attacks

In the above figure some algorithm is not yet implemented on different cyber-attacks like k- nearest neighbor in not implemented in probing, U2R, R2L, Decision Tree is applied only for probing and R2L, Artificial Neural Network applied only for probing and DOS. Phishing attack can detect with good accuracy by decision tree, Decision tree has give best accuracy for phishing attack.

Some more attacks and its prevention methods are describe below [8]:-

- **Man in the Middle (MITM)**:- This type of attacks includes Session hijacking, IP Spoofing and and its prevention are attention on security warning and always access HTTPS site. Phishing attack this type of attacks are spear phishing and prevention methods are be cautious, Use antivirus software and phishing detection tools.
- **SQL injection**:- This type of attacks are Union based sql injection, Error based sql injection and blind sql injection , and its prevention methods are attack surface has to be reduced , use firewall , always monitor statement of sql based query.
- **Cross site scripting (XSS) attack**: - This type of attacks are store Cross site scripting, DOM Cross site scripting, prevention methods are encoding/decoding, input validation and filter user input.
- **Password attack**:- This type of attacks are Brute force , Dictionary attack, and prevention methods are Create strong & unique & secure passwords , change passwords after a time period, passwords should be different for different account.
- **Cross site request forgery (CSRF) attack**: - Prevention of this type attacks are Disable scripting in Browser, Never save your login in the browser.
- **Malware Attack [9]**:- In this attack hacker installs malicious software into targeted PC example virus, warm, Trojan, ransom ware and spyware use antivirus software for it.

Webology, Volume 18, Number 4, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I4/142

CONCLUSION AND FUTURE WORK

This paper describes different type of cyber-attacks and Machine learning techniques. It also provide summary of Machine learning algorithms used for detection of cyber-attacks. An over view of IDS and is working is also given. The significance of database in Intrusion Detection System is also highlighted and finally list of machine learning algorithm ,having best accuracy in the form of a table, against some specific cyber-attacks ,and list of latest cyber-attack and its prevention methods are describe. Each approach has some advantages and disadvantages. It is still a research area to identify new cyber-attacks and update our methodologies to protect our computer network.

References

1. S. Larson. 10 biggest hacks of 2017. 2017, December 20. Retrieved: November 3, 2018.
2. Dinil Mon, Divakaran, et al. "Evidence gathering for network security and forensics," Digital Investigation ,2017, pp.56–65.
3. S. Dolev and S. Lodha, "Cyber Security Cryptography and Machine Learning", In Proceedings of the First International Conference, CSCML 2017, Beer-Sheva, Israel, June 29-30, 2017.
4. Ayon Dey Department of CSE, Gautam Buddha University, Greater Noida, Uttar Pradesh, India. "Machine Learning Algorithms: A Review". Vol, 7, 1174-1179, 2016.
5. I. Zaharakis, S. B. Kotsiantis and P. Pintelas. "Supervised machine learning: Emerging artificial intelligence applications in computer engineering", 160, 3-24, 2007.
6. Gozde Karatas, Onder Demir, Ozgur Koray Sahingoz "Deep Learning in Intrusion Detection Systems", International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism, Dec, 2018.
7. R. Kiruthiga, D. Akila," Phishing Websites Detection Using Machine Learning", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-2S11, September 2019.
8. Jibi Mariam Biju1, Neethu Gopal2, Anju J Prakash3," CYBER ATTACKS AND ITS DIFFERENT TYPES", International Research Journal of Engineering and Technology (IRJET), Volume: 06 Issue: 03 | Mar 2019. International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-12S3, October 2019 74 Retrieval Number: L101910812S319/2019©BEIESP DOI: 10.35940/ijitee.L1019.10812S319 Published By: Blue Eyes Intelligence Engineering and Sciences Publication.
9. Richa Adlakha, Shobhit Sharma, Aman Rawat, Kamlesh Sharma," Cyber Security Goal's, Issue's, Categorization & Data Breaches" , 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con), India, 2019.
10. Mohd, N., Singh, A., & Bhadauria, H. S. (2021). Intrusion Detection System Based on Hybrid Hierarchical Classifiers. *Wireless Personal Communications*, *121*(1), 659-686.
11. Tiwari, P., Upadhyay, D., Pant, B., & Mohd, N. (2022). Multiclass Classification in Machine Learning Algorithms for Disease Prediction. In *International Conference on Advanced Informatics for Computing Research* (pp. 102-111). Springer, Cham.
12. Mohd, N., Singh, A., & Bhadauria, H. S. (2020). A novel SVM based IDS for distributed denial of sleep strike in wireless sensor networks. *Wireless Personal Communications*, *111*(3), 1999-2022.