**Machine Learning in Cyber Security**

*Pooja Sree Simhadri, 3rd Year.*

*Department of Computer Science and Engineering, SRM University – AP, Andhra Pradesh, India*

Email: [poojasree\_simhadri@srmap.edu.in](mailto:poojasree_simhadri@srmap.edu.in)

**Abstract:** Cyber security is creating security for devices from cyber attacks. Cyber attacks are mainly done when the data transaction is done between two device networks. Machine Learning is growing technology for the better future usage of technology in day-to-day life based on historical data and information. We are using Machine Learning in Cyber security to identify and be aware of threats caused in the device. The need to use machine learning in cyber security is to automatically identify and classify data by Supervised Learning. Although, day-by-day cyber threats are getting increased rapidly and more network logs are being created and it’s been a difficult task to analyze the data manually due to security loss. So, we need to get help from the machine. Normally, we use IDS (Intrusion Detection Systems) and IPS (Intrusion Prevention Systems) for the detection and prevention of threats. These are pattern based meaning only previously recorded ransomeware, malware or other cyber attacks will be detected. These can’t detect Zero day attacks (new attacks). Also, Big Data which is previously present in the Database must be protected and that can be done with help of Cyber Security using Machine Learning.

The major challenge in this transaction of the data through network is that the attackers are hiding their actions from Intrusion detection systems by editing system logs and covering tracks where user cannot know when the file got changed. We use different algorithms in the Machine Learning and will overcome the different types of Cyber threats. There are many Cyber attacks like Cyber Pharming, DNS tunnelling, XSS attacks, Denial-of- Service, SQL injection etc. These all challenges and security can be solved through following particular procedures of Machine Learning in Cyber Security.

**Keywords: Cyber Security, Machine Learning, Fraud detection, data loss prevention.**

**Introduction:**

**Cyber Security:**

Cyber security is creating security for threats causing in the network and the historical data or information resided in the database. There are various types of cyber attacks mainly Phishing (gathering of personal information and stealing money ) , Cyber Pharming ( redirecting one from a particular website to other website and asking money), Vishing ( mostly done in financial institutions like banks), Smishing ( SMS tricking ). The main target of the attackers is to seek financial gain by data theft also they target for the business disruption in the companies. Cyber attacks can be done both from within the company and also externally. Some organizations may heir some cyber attack teams in order to get above the other or the attack may be committed by cyber attackers for financial gains. These threats are known as External threats. Some threats may start within the company who have authorized and legitimate access knowingly or unknowingly. These threats are known as Internal threats. The main targets of the attackers are Business data, Customer financial data and their databases, IT infrastructure access , IT services for financial payments and sensitive personal data from government officials. There are also advanced cyber attacks like Backdoor Trojan which creates backdoor entry in victim’s system and gain remote and almost total control on the system, XSS attacks which insert malicious malware into the legitimate sources, Denial of Service which flood system’s resources by preventing responses to service requests, DNS tunnelling which is a transactional protocol that extracts data silently by establishing a communication with an unknown server. The total result of the cyber attacks is that it damages enterprise which causes valuable downtime, data manipulation and data loss, also money is being lost through ransoms which lead to major financial losses. To stop these activities, and have cyber crime free networks we need to use Cyber Security.

**Machine Learning in Cyber Security:**

Machine Learning is teaching the machine to identify, adapt and react in a particular situation. We use Machine Learning in Cyber Security for teaching the machine to identify the cyber threats from network and cyber attack methods. In simpler way, we can say that machine learning make cyber security proactive, simpler and far more effective in low cost. In Cyber Security, we follow a policy called DIKW [ Data Information Knowledge Wisdom] pyramid for security. This policy has mainly three aspects, they are Protect which protects the data or information. Assure where the Information Analysis take place and Manage where the information or the knowledge in managed and kept secured. Also, collection of knowledge is known as Insight and Insights with policies are known as Wisdom. In this process, we can manage the data securely. By using machine learning, we can analyze, develop and manipulate the data patterns with certain algorithms and anomalies. Not only quantity of the data, the quality of the data is also quite important. So, we have to concentrate more on collection, organizing and structuring the data. The quality need to be rich and textured enough to determine the details about network sensors, protocols, applications and machines to develop the patterns in machine learning algorithms. The overall agenda of our working is to identify and classify the cyber threat and acknowledge the user / software regarding the threat using different kinds of constraints.

**RELATED WORKS:**

**Types of Cyber Attacks and their Approach :**

There are many types of cyber attacks which are from basic to advanced level. We discuss about the few types of cyber attacks and their approach

1. **Cyber Pharming :** It is practice where the users are guided to different websites rather than the one preferred by the user and triggers the private information like bank account information, passwords etc. Here, the attackers redirect the user’s traffic from legitimate website to unsafe website where the DNS is under their control and succeed in their goal by causing financial loss and personal information disruption.

**Approach:**

This can be caused by malware pharming where the web traffic of user is accessed by malware such as Trojans where the attackers flood malicious mails, when the user opens the attached link or file in the mails the system will be indulged with free malware.

DNS poisoning is also a condition of cyber pharming where the access is gained on the DNS of the computer and redirected to fake websites. This can be done by one attacker to many users at a time. So, this is called most hazardous attack.

1. **Backdoor Trojan:** It is a practice where the virus is attacked to the system from backside

of the system and will gain remote access and has overall control of the system. This is done through based on phishing, the attacker’s sends malware by installing from backdoor.

**Approach:**

This is done through spamming mails through different links and files. When, the user opens the link a Trojan malware is installed from backside of the system.

1. **XSS attacks:** Cross site scripting is a type of attack where malicious scripts (data) are injected into system. This enters into web application through un trusted source via web requests. This is a limitless attack pattern. This content sent to user also include the forms of Javascript, HTML, Flash etc. So, that user can’t doubt the script. The user will have no idea of the malicious scripts and will execute the data. Now, the scripts enter into the system and can access all the cookies, tokens and other private information. It can also occur when the dynamic content sent to the web user is not being aware or validated of the malicious content. This attack will mostly invade private information by transmitting the data. There are two types of XSS attacks they are Reflected XSS attacks where the injected script is reflected off the web server resulting in an error message or search result. Stores XSS attacks where the attacks already occurred are stored permanently in a common field or database.

**Approach:**

These scripts enter the web application through un trusted source explicitly where the attacker can take complete control over the system. The attackers send some popup or error kind of messages so, that the user uses the link for the content required by him and gets attacked by the malware.

1. **Denial-of-Service:** This attack completely shut downs the system making it inaccessible for the system users. This attacks the legitimate users like account holders, employees etc. These attackers mostly target high profile organizations like banks, media , MNC’s, trade and government organizations. They result in victim’s great money and time loss. They exploit vulnerabilities which lead to service crash taking advantage of the bugs present in the system. An additional type DoS attack is DdoS where when multiple users are synced to one DoS server separately, then the attack is done for a single server as target. Due random distribution, the location is difficult to detect. One device will be attacked from different locations. Due to this, the whole system will be crashed subsequently and the attacker gains the total control of the system.

**Approach:**

Two common methods to attack the system by DoS are flooding services and crashing services. Flooding services are when there is a lot of traffic and buffer in the transaction of data which causes a slowdown in system where the system eventually stops working and th attack takes place. Some of the floods in DoS attacks are

Buffer overflow flood where the system is indulged with lot of traffic which can’t be handled. Then the list of attacks will be injected in the system along with exploiting the bugs.

ICMP flood where the networks of system are being misconfigured by sending spoofed packets. ICMP flood attack is also known as smurf attack or ping to death.

SYN flood where an appeal request is sent to connect the server but does not completes handshake. Though, manages to flood all ports so that user cannot access a legitimate server. Then, attacker gets succeeded in getting the access of user system when user is deviated to fake server.

1. **DNS Tunnelling**: DNS Tunnelling is attacking the software by malware through DNS protocol. It attacks via client server model and exploits data. These attacks are more dangerous because DNS is foundational protocol where names of domain are mapped to IP addresses. In the DNS tunnelling, attacker creates a website named under his choice and installs tunnelling malware program in the website. When a query is raised by the user, DNS resolver en-routes it to the attacker and the attacker installs the tunnelling program through server. This is an indirect connection between victim and attacker, DNS resolver acts as a medium. When the tunnel is installed it leads to the data exploitation and privacy interruption. Since, there will be no interaction between the user and attacker the location of attacker cannot be detected.

**Approach:**

This is where a tunnel is created in the DNS server making the firewall as an advantage where DNS resolvers can move in and out of the firewall. Then the DNS resolver maintains a relationship with server mapping domain names and IP addresses. Fromm this moment, when the query is raised the attacker injects malware through tunnelling.

1. **Drive-by-Download Attack:** This attack takes place where malicious content is downloaded into the system without the consent or knowledge of the user. There will an unsafe code downloaded into the system where it becomes key to cyber threat in the system. This can be entered into system if the system is not updated or has unsuccessful updates. It acts in three steps to invade your data firstly hijack the device, spy on the activity of the user and then ruin the data or completely disable the device. To overcome this, one need to keep all of the data up to date, remove unsupported and previous data, maintain the strong passwords, install web security apps to be aware of the attacks.

**Approach:**

This cyber attack in form of two types they are authorized and unauthorized form of attacks.

In authorized attack, user manually injects Trojans into system by clicking some unsafe links or files. Here, attacker creates malware, user interacts and form connection with it, malware installs and finally attackers enter the system.

In unauthorized attack, user is not directly acknowledged with the attack. When user surfs some other server or browser system gets directly attacked without any pop-ups or notifications. It is a direct entry into the system.

1. **SQL Injection Attack:** This is an attack where an injection called SQL is attacked to the database. The attacker uses SQL language code to gain access and manipulate the database. This attack leads to very serious and dangerous consequences in organizations because these are mostly dependent on relational databases leading to web development. Due to SQL injection attacks, one must compromise on the data integrity and user’s privacy. These injects are of three types based on how they cause data damage through back-end. They are

In-band SQLi which is straightforward to the attackers, they use same channel to release attacks and revert results. This is further divide into sub-categories like error-based and union based attacks.

Out-band SQLi which takes place when the attacker is unable to use the channel foor releasing attacks and reverting results or it is used when the system is too slow to run.

Inferential SQL is where the attackers use the patterns and behaviour of the server by sending more payloads of data. This is also known as Blind SQLi.

**Approach:**

SQL injections are getting entered through a database by a piece of query code or whenever the query is raised. When the additional information sent in web forms are not being stopped, attackers are trying to invade the database and manipulate the data in it. By doing this, the attacker will gain complete access overall the database and will be able to manipulate any table.

These are the different types of dangerous cyber attacks and its approach taking place in the networks.

**FRAMEWORKS OF CYBER SECURITY:**

A Cyber Security framework is a structured set of guidelines, best practices, and standards that organizations can use to manage and reduce cyber security risks. The framework provides a common language and a set of agreed-upon standards and practices that security leaders can use to communicate and collaborate effectively across different countries and industries

By adopting a cyber security framework, organizations can better understand their security postures and identify areas of weakness or vulnerability in their systems and processes. They can then define the processes and procedures needed to assess, monitor, and mitigate cyber security risks, and ensure that these measures are consistently applied across the organization.

Frameworks such as the National Institute of Standards and Technology (NIST) Cyber security Framework, the Center for Internet Security (CIS) Controls, and the ISO/IEC 27001 standard are widely recognized and adopted by organizations of all sizes and industries. These frameworks provide a comprehensive set of guidelines that can help organizations to establish a strong cyber security posture, reduce risk, and protect critical assets and data from cyber threats.

SOC2 (Service Organization Control 2) is a set of auditing standards and procedures that evaluate how well an organization manages its data and systems, including security, availability, processing integrity, confidentiality, and privacy. SOC2 reports are commonly used by service providers to demonstrate their security posture to customers and partners.

NERC-CIP (North American Electric Reliability Corporation Critical Infrastructure Protection) is a set of mandatory cyber security standards for the electric utility industry in North America. The standards define requirements for the secure operation of the power grid and protect against cyber threats that could disrupt the system.

HIPAA (Health Insurance Portability and Accountability Act) is a US federal law that regulates how sensitive patient data is stored and protected in the healthcare industry. HIPAA requires healthcare organizations and their business associates to establish and maintain privacy and security measures to protect sensitive patient health information

GDPR (General Data Protection Regulation) is a European Union regulation that sets rules for the protection of personal data of EU citizens. GDPR applies to any organization that collects, processes, or stores personal data of EU citizens, regardless of where the organization is located. GDPR includes requirements for obtaining consent, data breach notification, data portability, and the right to erasure

FISMA (Federal Information Security Management Act) is a US federal law that requires federal agencies to develop, implement, and maintain information security programs to protect government information systems, and networks. FISMA also requires federal agencies to undergo regular cyber security audits and assessments to ensure compliance with established security standards and guidelines.

**EXPERIMENTAL STUDY / RESULT ANALYSIS:**

From this experiment, we can understand the different protocols required to analyze a cyber threat. To detect the cyber threat pattern the destined Date and Time, IP address, Port numbers etc are required to get the exact required information of a cyber threat. To prevent the cyber attacks, the computer software must be aware of common attacks faced by companies. From the results, we got to know about various sub-categories in the attacks and how to acknowledge ourselves with the attack. There are attacks which are attacked through different protocols.

We got to know that there are some unique attack categories. We made an algorithm to analyze the date and time of each particular attack. So, that one can understand the way of attacking and probability of preventing attacks at that particular points. We calculated the pearson correlation of source port and destination port. We analyzed the heatmap for better understanding. We plotted the scatter plots. Then we calculated the destination port and duration so that we can understand the patterns. We conducted the chi square test and got to know that contingency table of Attack type versus Target port and the individual counts are not uniform. This allows to affirm our inference that there might be an interaction between these two variables. We can formally test whether the variations we observe reflect real differences or are just a by-product of randomness. There are many different ways to test, but we will focus on the most widely used test: the **Chi-square test**. The null hypothesis for the Chi-square test is as follows:

The attack category is independent of destination port. The 𝑝 -value is extremely small and therefore we reject the null hypothesis and conclude that the attack category and destination port are not independent. In other words, the destination port do not remain the same for different attack categories, what agrees with what we observe previously.

To visualize this relationship between these variables (source port, destination port, and attack category), we can initially make a scatterplot using the .scatterplot() function between the source and destination ports in terms of the attack category:

A certain pattern can be seen in the graph, in which attacks usually occur on ports less than 10,000 or greater than 50,000, except for shellcode-type attacks, as we had anticipated in the hypothesis test.

To see this relationship more in depth, we can visualize the the distribution of the logical origin and destination ports with a strip diagram using the .stripplot() function:

**CONCLUSION:**

Cyber security is the most required technology of today because of the evolving industries and everyone in the society directly or indirectly being affected by the technology we use. Technology can be used in the two ways optimizing the amount of work and destroying someone’s work. Everyone must be aware of the cyber threats going on in the society and try to prevent them in their way possible. We can protect our data by encrypting the information and backing up the data. The most dangerous way for a system to enter into threat is through memory. So, we must be very conscious of what’s going on the system. The most efficient methods to prevent attacks and protect from cyber threats are the DIKW [Data Information Knowledge Wisdom] Method which follows the three main policies Protect, Assure and Manage.

**Protect** which protects and safe guards the data.

**Assure** which ensures the information access whenever required and also assuring mo attack takes place.

**Manage** is managing the knowledge stored in the system providing security.

We also have other method called The CIA Traid which Confidentiality, Integrity and Availability. Following these methods can ensure and protect our data and prevent from Cyber Attacks.

**REFERENCE:**

1. **Tech Target solutions**
2. **Research papers from SANS Technology Institute**
3. **Science Gate.**