Detection of cyber attacks using machine learning

N Vamshi Krishna[1],Dr.Manjula Sanjay Koti[2]

[1] Student, MCA,Dayananda sagar academy of technology and managment,Karnatka,india
[2] Prof&Head,MCA, Dayananda sagar academy of technology and management,Karnataka, india

ABSTRACT

The prevalence of hostile network traffic activities has considerably increased them need for precise intrusion detection. Intrusion detection systems can automatically identifysecurity violations such denial of service attacks, malware, port scanning, buffer overflows,CGI attacks, and flooding. Also provided is a thorough analysis of benchmark cybersecurity datasets. This essay's objective is to give readers a road map for understanding the approach used by intrusion detection systems and cyber security. Information and communication technology integration enables smart metres to participate in real-time in the operation of electricity systems.[1] However, cyber dangers might have an effect on contemporary metering infrastructure. Cyber intrusions could have an effect on both utilities and power users.

Keyword: - *intrusion detection, Cgi attacks,smart meters*

1. DETECTION OF CYBER ATTACKS USING MACHINE LEARNING

essential piece of information management equipment called a network intrusion 2 detection system helps identify and stop security breaches including unauthorised access, system tampering, duplication, or any other type of harm to an information system. The network-based IDS analyses if a certain activity is appropriate or intrusive based on network audit data. Most web IDS are compatible with security tools like firewalls and antivirus products. The NIDS may keep an eye on incoming traffic signals and then check them for any suspicious activity or probes.[2] If you can identify the malicious packets in advance, we can stop the destruction by dropping them or by taking the necessary action on them. Only automated systems capable of performing detection and prevention with little or no human intervention and high levels of scalability. Snort is a potent non-proprietary system for preventing intrusion that can recognise and analyse network data in real time. It is based on rule fingerprinting and anomaly testing techniques. The AMI network signals were encrypted and decrypted with the least amount of processing and communication latency.To detect malicious human-initiated activity, smart metres can be used with an IDS with two sensor processes.[3] The creation of a complete anomaly-based detection system enables the decrease of erroneous alerts and issues caused by unbalanced occurrences.

1.2 Algorithms

1. rule-based algorithms

2. statistics-based algorithms

3. machine learning algorithm

2. RELATED WORKS

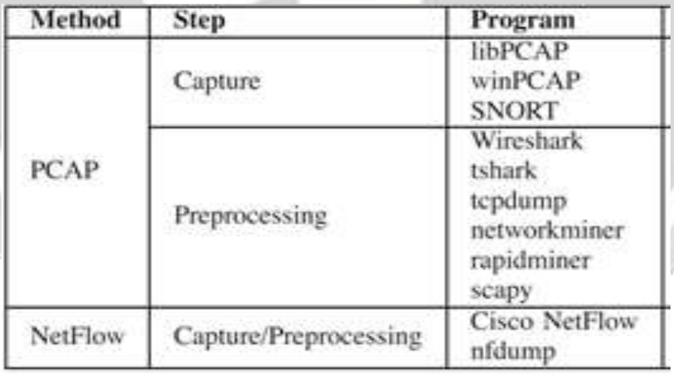
The two components of an intrusion detection system are anomaly detection and signaturebased or abusive intrusion detection. Attacks that use pattern matching or signatures are defended against by signature-based IDS. One of the greatest ways to detect the known hazards is to use signatures. A pre-programmed list of known threats and associated indicators of compromise (IOCs) is used to operate it. It keeps track of the packets moving across the network, checks them against a list of approved IOCs, and notifies the user of any odd conduct. It starts the non-intrusive, avoidable activities. An anomaly-based detection system can warn of unknown dangerous activities. Analytical-based IDS employs machine learning techniques to educate it to recognise the normalised baseline rather than looking for known threats.[4] As a result, all network activity is compared to this baseline, which represents the system's regular behaviour. Pattern matching and the signature IDS are used to check data. Reliable signatures' network security is maintained using it. The elements of each signature in this group are linked together. From a technical, economical, or administrative perspective, implementing an intrusion detection system should be simpler than doing so with competing open source options. By decoding network packets and comparing them to predefined rules, it makes it possible to evaluate and identify a variety of dangers.[5] A block diagram can be used to demonstrate the packet filtering for the stream.

2.1 DATA COLLECTION

There are two methods for gathering data. The second one handles system calls, whereas the first one processes packet headers and payloads derived from bundles of network traffic and from protocols like the TCP/IP communication stack. The Netflow protocol and packet capture (PCAP) are two tools for gathering network traffic. The use of PCAP mandates the gathering 9 of all packet headers for data delivery and makes it simpler to gather more precise network information.[6] Utilising NetFlow, it is possible to compile summaries of data on packet flow in a network.These software tools are employed to record network traffic.

2.2 FEATURES OF IDS

Data about network activities is used to build network-based data. TCP/IP connections' core characteristics are isolated and categorised as header-based, connection-based, and flowbased. features that flow analysis determines for flow-based testing. IP protocols, source and destination port numbers, IP header length, and IP headers are all included in the headerbased feature, which is connected to packet headers. Features that relate to traffic are associated with either a certain time period. The same host or service can be used to extract these features. The request numbers, request type, and the number of unsuccessful login attempts are a few examples of content-based attributes that can be extracted from data 8 included in different data parts of packets.



Method	Step	Program
PCAP	Capture	libPCAP winPCAP SNORT
	Preprocessing	Wireshark tshark tcpdump networkminer rapidminer scapy
NetFlow	Capture/Preprocessing	Cisco NetFlow nfdump

2.3 ATTACK TYPES

The numerous IDS attacks are thoroughly covered in this section. 1.DDOS assaults work by severely overburdening the server with requests, rendering it unresponsive. 2.Utr exploits include taking on the persona of a regular user to locate security gaps and get root access.[8] 3. R2L attacks use the remote system to try to gain unauthorised access to the target system. 4. Probe attacks gather information about the system while sending scan packets throughout the entire network to look for vulnerabilities. 5. To acquire unauthorised access and steal information, injection attacks use scripts that insert queries.

3. DATA MINING MODEL/ALGORITHM

1.KNN ALGORITHM The K-means approach, which separates the items into k pieces (k n), is the most straightforward and fundamental way for clustering by splitting. A centroid-based approach is K-means. The k-means is effective for spotting outliers since the mean value of the cluster is 3 larger when a value deviates from the median of the given data. Normal behavioural patterns are anticipated to occur noticeably more frequently than outliers or deviant activity in this outlier identification experiment.

2.AES(Advanced Standard Encryption) ALGORITHM Cryptography is used in this algorithm. The symmetric block cypher with a chunk capacity of 128 bits is used by the AES Encryption technique to encrypt data. These various blocks are 5 converted using keys that are 128, 192, and 256 bits long. These blocks are initially individually encrypted, and only then are they joined to create the ciphertext. An SP network, sometimes referred to as a substitution-permutation network, serves as its base.[10] It is made up of several interconnected processes, some of which call for bit shuffles while others call for permutations or the replacement of particular outputs for inputs.

3. DES(Data Encryption Standard) ALGORITHM DES is vulnerable to serious attacks. Each block has a size of 64 bits. The input for DES and 64-bit cypher text is plain text. One algorithm and key can be used for encryption and decryption with very little modification. The key's exact size is 56 bits. Bit positions 8, 16, 24, 32, 40, 48, 56, and 64 serve no function.[11] A wide range of permutations are used to achieve encryption and decryption.

4.DATA AGGREGATION ALGORITHM A successful data aggregation method will boost network robustness while reducing node energy usage. Despite these benefits, data aggregation lengthens the distance packets must travel via the network.[12] Data aggregation contributes to reduced traffic and energy consumption. The network becomes more dependable as a result. The main objective of data aggregation is to reduce redundant data by choosing the relevant information from the incoming data and providing it to the end nodes with the help of data aggregation techniques like MEAN MIN,MAX,MEDIAN, and others.

5.ANN (Artificial Neural Network) Massive numbers of basic neural units (also known as artificial neurons) form the foundation of the ANN algorithm, and these artificial neurons closely mimic the observed behaviour of the axons in a real brain. The principles of ANN algorithms were inspired by biological neurons and their behaviours. Their weight is processed for the output to identify the concurrent layers, and they may contain one or more hidden -layers.[13] Clearly complicated and non-linear relationships between exposure and controllable variables are captureable by ANN algorithms. Because they are taught rather than formally implemented, these systems work well in circumstances where it is difficult to express the solution or feature identification in a traditional computer code.

6.BLOW FISH ALGORITHM Blow Fish algorithm, a 64-bit block cipher which is symmetric and is of variable length. It is a general-purpose algorithm providing quick and unpaid alternative for DES and IDEA Encryption algorithms. Blow Fish is remarkably quicker than DES and IDEA. All the, its small block size and this being insecure is one of the major issue. This algorithm includes two vital parts; 1. Data Encryption: It includes key dependent permutation and data dependent substitution. It uses logic gates. 2.Key Expansion and Sub Keys: Bit keys are transformed [14] into sub key arrays. Sub keys are pre contradicted well before encryption are decryption.

4. CONCLUSIONS

In this work, we implement the methods for improving intrusion detection system performance. The IDS sensor typically compares each packet to each signature in a signaturebased IDS. This approach significantly lowers false positives while having minimal effect on increasing detection rates.[15] Our research shows that the principal pattern makes our technique more effective in terms of detection rate and lowered false positives by decreasing the need to compare rule signatures and enhancing detection. Use benchmark datasets for intrusion detections as a point of comparison for cutting-edge cybersecurity techniques. The distribution of feature and attack types, data gathering techniques, and requirements for dataset dependability have all been taken into account.

5. REFERENCES

[1] Horng, S.-J.; Su, M.-Y.; Chen, Y.-H.; Kao, T.-W.; Chen, R.-J.; Lai, J.-L. & Perkasa, C. D.,"A novel intrusion detection system based on hierarchical clustering and support vector machines," Expert systems with Applications, Elsevier, 2011, 38, 306-313

[2] Lacroix, A. B.; Langlois, J. P.; Boyer, F.-R.; Gosselin, A. & Bois, G.,"Node configuration for the Aho-Corasick algorithm in intrusion detection systems Architectures for Networking and Communications Systems (ANCS)," 2016 ACM/IEEE Symposium on, 2016, 121-122

[3] Liao, H.-J.; Lin, C.-H. R.; Lin, Y.-C. & Tung, K. Y.,"Intrusion detection system: A comprehensive review," Journal of Network and Computer Applications, Elsevier, 2013, 36, 16-24

[4] IDS. [online]. Available: http://www.snort.org [5] Scarfone, K. & Mell, P., "Guide to intrusion detection and prevention systems (idps)," NIST special publication, 2007, 800, 94

[5] Scarfone, K. & Mell, P., "Guide to intrusion detection and prevention systems (idps)," NIST special publication, 2007, 800, 94

[6] Buczak A. L. & Guven, E.,"A survey of data mining and machine learning methods for cyber security intrusion detection," IEEE Communications Surveys & Tutorials, IEEE, 2016, 18,1153-1176