Algorithm based network intrusion detection system

Prof. Mary M Dsouza, Associate Professor, Acharya Institute of Technology, Bengaluru, India

A Sai Vigneswara R, Deepak Hegde, K Siva Narayana R, Praveena
Students, Acharya Institute of Technology, Bengaluru, India

Abstract: As hostile network traffic activities have become more prevalent, the significance of reliable intrusion detection is increasing dramatically. Intrusion Detection Systems offer automatic detection for security violations such as denial of service, malware, port scanning, buffer overflow, CGI attacks, and flooding. Additionally, athorough analysis of benchmark cybersecurity datasets is provided. This article aims to give readersa roadmap for understanding the potential of approaches for cybersecurity and intrusion detectionsystems. Real-time communication for smart metresto participate in power system operations is made possible by the integration of information and communication technology. However, Advanced Metering Infrastructures are susceptible to online threats. Cyber intrusions could affect electricity users as well as utilities.

I.INTRODUCTION

The network intrusion detection system is a vital piece of information management equipment that aids in detecting and preventing security breaches including illegal access, system alteration, duplicating, or any other kind of information system damage. Based on network audit data, the network- based IDS determines if a certain behaviour is legitimate or invasive. The majority of web IDS are compatible with security solutions like firewalls and antivirus software.

Incoming traffic signals can be monitored by the NIDS, which can then analyse them to find malicious activities and probes. We can stop the destruction by dropping the malicious packets or taking appropriate action on them if you can identify them in advance. Only systems with high levels of scalability and little human involvement can perform prevention and detection. Based on rule fingerprint and anomaly testing procedures, Snort is a powerful open - source system for intrusion prevention and detection system that can analyse network data in real time.

It has been developed to encrypt and decode AMI network signals with the least amount of processing and communication latency possible.For smart metres, an IDS with two sensor processes can be used to spot malicious human-initiated behaviour. The development of a thorough anomaly-based detection algorithm minimises false warning signal and due to unbalanced incidents.

These algorithms can be based on three major drives,

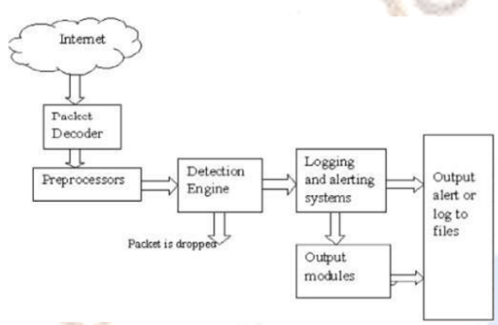
1. Rule-based algorithms
2. Statistics based algorithms
3. Machine learning algorithms.

These algorithms are discussed below which are proven valuable for the NIDS.

II. RELATED WORK

The 2 paradimse of the Intrusion detection system are Signature based or misuse intrusion detection, and anamaly detection. Signature based IDS represents attack in the form of pattern matching or signature. signature is one of the best method approach for identifying the known threats. It operates by using a pre-programmed list of known threats and their indicators of compromise (IOCs). It moniters the packets traversing the network an dit compares the packets to the database of known IOCs and flags any suspesious behavior. It triggers the non-intrusive activities which should be avoided. Anomaly based detecton can alert from the malcicious behaviour that is not known. Anamaly based IDS instead of searching fro the known threats it searches unknown threts it uses machine learning algorithm to train the IDS to recognize the normalized baseline. So this baseline represents the systems normal behavior and all network activity is compared to the baseline. Pattern matching works with the signature IDS and audit the data. It is used to maintain the relablity of the network security for the signatures. This group of signatures are in interrelation with the signature elements.

Intrusion detection system should be technically,financially or administratively easier to implement than other open source tools. It decodes the network packets and compares them with the predefined rules and eneble the inspects and detect many different types of attacks. The packet filtering for the traffic can be dipected by block diagram



Block diagram of packet filtering

Data Collection

The collection of data can be done in 2 ways. The first one is based on the processing the system calls and the other is packet headers and payloads extracted from network traffic packages and from protocols such as TCP/IP communication stack. The approaches used to collect the network traffic is packet capture (PCAP) and the Netflow protocol. PCAP enables collection of more detailed datafrom the network and it involves the collection ofall packet headers for information to be transmitted. NetFlow enables the collection of summary information and related to the flow of packets in a network.These are the programs to capture the network traffic.

Method	Step	Program	Ref.
PCAP	Capture	libPCAP	[32]
		winPCAP	[33]
		SNORT	[34]
	Preprocessing	Wireshark	[35]
		tshark	[36]
		tcpdump	[37]
		networkminer	[38]
		rapidminer	[39]
		scapy	[40]
NetFlow	Capture/Preprocessing	Cisco NetFlow	[41]
		nfdump	[42]

Features of IDS:

Network based data are obtained by collecting network traffic data. Basic features are extracted from TCP/IP connections and it can be classified as header-based, connection based and flow based. Flow-based testing attributes computed through analysis of the flow. Header based feature is related to packet header and it includes source and destination port numbers, IP protocols, IP header length. Traffic-based features are associated with either a specific time intervals. This features can be extracted by considering the same host or service. Content-based features are extracted from data embedded in different data portions of packets and include the request numbers, request type and the number of failed logins.

Attack types:

This section deals with the various types of attacks in IDSs.

1. DDOS attacks are based on the large flooding on the server and make it unavailable to respond by overloading it many service requests.
2. UTR attacks involve behaving as normal user with the aim of detecting systems vulnarabilities and gaining the root access.
3. R2L attacks attempt to use the remote system to gain unauthorised access to damage the target system.
4. Probe attacks are based on the searching for vulnarables throughout the whole network by sending scan packets and gaining information about system.
5. Injection attacks use the scripts that inject queries for purpose of gaining. unauthorised access and stealing information.

Attack name	Examples	Description
Denial of Service (DoS) [45]	Botnet, Slowloris, smurf, SYN flood	Temporarily blocks the normal use of network utilities by flooding the network with traffic.
Distributed DoS (DDoS) [46]	LAND, ping of death, RUDY, teardrop	Floods the server and makes it nonresponsive to users by overloading it with service requests. Unlike in DoS attacks, the flooding originates from many sources.
User-to-Root (U2R) [45]	Buffer overflow, rootkit, Perl, loadmodule	Behaves as a normal user with the aim of detecting system vulnerabilities and gaining root access.
Remote-to-Local (R2L) [45]	SSH brute force, warezmaster, multihop, imap, spy	Gains local access via a remote system and damages the system. May be combined with U2R attacks, thus making these attacks difficult to differentiate.
Probe [45]	Satan, IP sweep, port sweep	Searches for vulnerabilities throughout the whole network via IP addresses by sending scan packets and gaining information about the system.
Password [18]	Brute force FTP-Patator, brute force SSH-Patator	Gains access to the system after stealing passwords by guessing.
Injection [47]	SQL injection, Cross-Site Scripting (XSS)	Uses a script to inject commands/queries to gain unauthorized access and steal information.

III. DATA MINING MODEL/ALGORITHM

1. KNN algo

The easiest and most basic method of clustering by splitting, known as K-means, divides the objects into k parts (k n). K-means is a centroid-based methodology. Because the mean value of the cluster is higher when a value is distant from the median of the data, the k-means is particularly useful for locating outliers.

It is expected in this outlier identification model that normal behavioural patterns occur far more frequently than outliers or aberrant behaviours.

2. AES(Advanced Standard Encryption) algo

This is a Cryptographic algorithm With a chunk size of 128 bits, the AES Encryption technique is a symmetric block cypher. These distinct blocks are converted using keys that are 128, 192, and 256 bits long. It then connects these blocks to create the ciphertext after encrypting each one separately. It is founded on an SP network, also referred to as a substitution-permutation network. It comprises of a number of interconnected processes, some of which involve bit shuffles and others involve substituting feeds with particular outcomes (substitutions) (permutations).

3. DES(Data Encryption Standard) algo

DES is unsafe to very eminent attacks. It is a block cipher where 64-bits block size each. Plain text is the input to DES and cipher text of size 64-bits. A single algorithm and key with minor changes can be used for both Encryption and Decryption. The size of key is 56-bits. The bit position 8,16,24,32,40,48,56&64 are abandoned. Encryption and Decryption are formed in a sequence of permutations.

4. Data Aggregation algo

The decrease in node energy usage and improvement of network lifetime are qualities of a successful data aggregation algorithm. Despite these advantages, data aggregation also results in a rise in the network's packet transmission time.

Data aggregation contributes to energy conservation and traffic load reduction. It helps the network become more resilient. With the aid of data aggregation algorithms like MEAN MIN,MAX,MEDIAN and others, the primary goal of data aggregation is to decrease redundant data by extracting the pertinent information from the obtained data and sending it to the end nodes.

5. ANN (Artificial Neural Network)

ANN algorithm is based on a large number of basic neural units (artificial neurons), which are roughly equivalent to the observed behavior of the axons in a real brain. Biological neurons and their behaviors have inspired for the basis of ANN algorithms. They include one or more hidden -layers, their weight is processed for the output to decide the concurrent layers. ANN algorithms capture distinctly complex and relationships that are non-linear joining both controlabel and exposure variables. These systems thrive in areas where the solution or feature identification is challenging to describe in a conventional computer programme because they are self-learning and taughtrather than explicitly coded.

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 into sub key arrays. Sub keys are pre contradicted well before encryption are decryption.

Tabel 1:

ALGORITHM	CREATED BY	KEY SIZE (BITS)	BLOCK SIZE (BITS)
DES	IBM IN 1978	56	64
3DES	IBM IN 1978	112 OR 168	64
RIJNDAEL	JOAN DAEMEN & VINCENT RIJMEN IN 1998	256	128
BLOWFISH	BRUCE SCHNEIER IN 1993	32-448 (128 BY DEFAULT)	64

Tabel 2:

SI.NO	Reference No.	Techniques /Methods used	Accuracy(%)
1	17	KNN	85.53
2	20	AES	96.5
3	21	DES	99.2
4	22	DATA AGGREGATION	86.6
5	55	ANN	
6	66	BLOW FISH	

Conclusion:

In this study, we put the techniques for enhancing intrusion detection systems performance into practice. In a signature-based IDS, the IDS sensor typically checks each packet against all signatures. This strategy is effective in reducing false positives greatly but ineffective in raising detection rates. Due to the primary pattern's role in minimizing the need to compare rule signatures and improving detection, our research demonstrates that our strategy performs better in terms of detection rate and lowering false positives. Use benchmark datasets for intrusion detections to provide as a point of comparison for modern cybersecurity techniques. We have taken into account the data collection methods, the distribution of feature and attack kinds, and the dataset reliability standards. researchers working on ML and DL for cybersecurity applications a useful road map.

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

[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

[7] V. Gomez, C. Hernandez, and F. Martinez, "Energy policies in smart grids," *Contemp. Eng. Sci.*, vol. 10, no. 20, pp. 987–999, 2017.

[8] B. Li, S. Lv, and Q. Pan, "The Internet of Things and smart grid," in *Proc. IOP Conf. Earth Environ. Sci.*, vol. 113, 2018, pp. 12–38.

[9] H. Jiang, K. Wang, Y. Wang, M. Gao, and Y. Zhang, "Energy big data: A survey," *IEEE Access*, vol. 4, pp. 3844–3861, 2017.

[10] F. Pasqualetti, F. Dorfler, and F. Bullo, "Control-theoretic methods for cyberphysical security: Geometric principles for optimal cross- layer resilient control systems," *IEEE Control Syst. Mag.*, vol. 35, no. 1, pp. 110–127, Feb. 2015.

[11] N. F. S. Baker and K. Timlin, *In the Dark: Crucial Industries Confront Cyber Attacks*, McAfee, Santa Clara, CA, USA, 2012.

Bounding inequalities and augmented Lyapunov–Krasovskii functionals," *IEEE Trans. Autom. Control*, vol. 62, no. 10, pp. 5331–5336, Oct. 2017.

[13] S. McLaughlin, B. Holbert, A. Fawaz, R. Berthier, and S. Zonouz, "A Multi-Sensor Energy Theft Detection Framework for Advanced Metering Infrastructures," *IEEE J. Selected Areas in Communications*, vol. 31, no. 7, pp. 1319-1330, Jul. 2013.

[14] Y. Liu, S. Hu, and T. Ho, "Leveraging Strategic Detection Techniques for Smart Home Pricing Cyberattacks," *IEEE Trans. Dependable and SecureComputing*, vol. 13, no. 2, pp. 220-235, 1 Apr. 2016.

[15] X. Liu, P. Zhu, Y. Zhang, and K. Chen, "A Collaborative Intrusion Detection Mechanism Against False Data Injection Attack in Advanced Metering Infrastructure," *IEEE Trans. Smart Grid*, vol. 6, no. 5, pp. 2435-2443, Sept. 2015.

[16] R. Berthier and W.H. Sanders, "Specification-Based Intrusion Detection for Advanced Metering Infrastructures," *IEEE Pacific Rim International Symposium on Dependable Computing (PRDC)*, Pasadena, CA, USA, pp. 184-193, Dec. 2011.

[17] M. A. Faisal, Z. Aung, J. R. Williams, and A. Sanchez, "Data-Stream- Based Intrusion Detection System for Advanced Metering Infrastructure in Smart Grid: A Feasibility Study," *IEEE Syst. J.*, vol. 9, no. 1, pp. 31- 44, Mar. 2015.

[18] R. Ullah, Y. Faheem, and B. Kim, "Energy and Congestion-Aware Routing Metric for Smart Grid AMI Networks in Smart City," *IEEE Access*, vol. 5, pp. 13799-13810, 2017.

[19] Albin, E. & Rowe, N. C.," A realistic experimental comparison of the Suricata and Snort intrusion-detection systems," *Advanced Information Networking and Applications Workshops (WAINA), 2012 26th International Conference on*, 2012, 122-127

[20] Saboor, A.; Akhlaq, M. & Aslam, B. ,"Experimental evaluation of Snort against DDoS attacks under different hardware configurations *Information Assurance (NCIA), 2013 2nd National Conference on*, 2013, 31-37

[21] White, J. S.; Fitzsimmons, T. & Matthews, J. N.," Quantitative analysis of intrusion detection systems: Snort and Suricata *SPIE Defense, Security, and Sensing," International Society for Optics and Photonics*, 2013, 875704-875704

[22] Victor, G. J.; Rao, M. S. & Venkaiah, V. C.," Intrusion detection systems-analysis and containment of false positives alerts," *Int. J. Comput. Appl*, 2010, *5*, 27-33

[23] Huang, C.; Xiong, J. & Peng, Z.," Applied research on snort intrusion detection model in the campus network," *Robotics and Applications (ISRA), 2012 IEEE Symposium on*, 2012, 596-599

[24] Zammit, D.,"A machine learning based approach for intrusion prevention using honeypot interaction patterns as training data,"
University of Malta, *University of Malta*, 2016

[25] Anwar, S.; Mohamad Zain, J.; Zolkipli, M. F.; Inayat, Z.; Khan, S.; Anthony, B. & Chang, V.,"From Intrusion Detection to an Intrusion Response System: Fundamentals, Requirements, and Future Directions,"
Algorithms, Multidisciplinary Digital Publishing Institute, 2017, *10*, 39