FUZZY LOGIC FOR INTRUSION DETECTION SYSTEM

MUHAMED JAMSHIR M[1], AJEESHA M I[2]

[1]Adhoc Faculty, NSS College of Engineering, Palakkad, Kerala, India.

[2]Research Scholar, School of Computer Studies, Rathnavel Subramaniam College of Arts and Science, Coimbatore, Tamilnadu, India.

ABSTRACT

An Intrusion Detection System is a system that monitors network traffic for detecting unauthorized access and issues alerts. It is a software application that scans a network or a system for harmful activity or policy breaching. IDS search for suspicious activity and known threats, sending up alerts when it finds such items. Each IDS is programmed to analyze traffic and identify patterns in that traffic that may indicate a cyberattack of various sorts. The prediction process may produce false alarms in many anomaly based intrusion detection systems. With the concept of fuzzy logic, the false alarm rate in establishing intrusive activities can be reduced. A set of efficient fuzzy rules can be used to define the normal and abnormal behaviors in a computer network. The KDDCup 99 dataset is used for the experiments and evaluation. This paper describes an intrusion detection system that utilize fuzzy logic.

KEYWORDS: Intrusion detection system, Fuzzy logic, Machine Learning, KDDCup99.

INTRODUCTION

In the early 2000's IDS started becoming a security best practice. In today's day and age, security is the buzz of the tech world. Privacy and the safety of users personal information are always being brought up time and time again, with good reason. Users are frequently updating their personal information, their location, and more on a daily basis from their phones. Intrusion Detection Systems (IDS) are the first line of defense for any system or network. These systems detect and classify traffic as good or malicious, and are the inherent gatekeepers of the system/network. Today, many of these IDS systems are becoming AI-based due to the speed of new advancements in technology[2]. Intrusion detection systems are turning out to be progressively significant in maintaining adequate network protection. Intrusion prevention systems also monitor network packets inbound the system to check the malicious activities involved in it and at once sends the warning notifications. The detection methods of IDS are Anomaly based method and Signature based method. Signature-based and anomaly-based detections are the two main methods of identifying and alerting on threats.

Anomaly Based Method:

Anomaly based IDS detects the unknown malware attacks as new malware are developed rapidly. The machine learning creates a trustful activity model. It comprises of a statistical model of normal network traffic with the bandwidth used, the protocols defined for the traffic, the ports, and devices which are part of the network. It frequently monitors the network traffic and compares it with the statistical model. Anomaly testing requires more hardware spread further across the network than is required with signature-based IDS

Signature Based Model:

Signature-based IDS detects the attacks on the basis of the specific patterns such as number of bytes or number of 1's or number of 0's in the network traffic. It also detects on the basis of the already known malicious instruction sequence that is used by the malware. The detected patterns in the IDS are known as signatures. Signature-based IDS rely on a database of known attacks but it is difficult to detect new attacks, for which no pattern is available. The principle of SIDS is matching. The data is analyzed and compared with the signature of known attacks. A usual intrusion detection system is demonstrated in Fig 1. The arrow lines symbolize the amount of information flowing from one component to another.

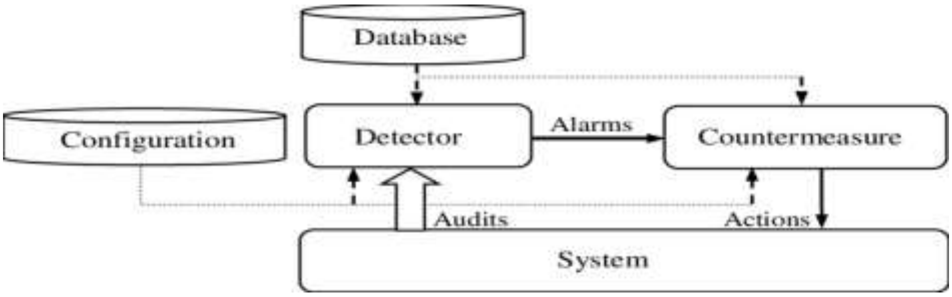


Fig 1: Basic Intrusion Detection system

IDS can be categorized based on its monitoring scope and detection techniques. Intrusion Detection Systems are broadly classified into two types. They are host-based and network-based intrusion detection systems. A HIDS is capable of verifying all parts of dynamic behavior and the state of a computer system based on the configuration. HIDS are installed in a host and they monitor traffics that are originating and coming to that particular hosts only. If the attack is on any other part of network they will not be detected by HIDS. It monitors all the user activities. They do not require any extra hardware since they can be installed in the existing host servers. In small scale network HIDS is preferred. NIDS is used to monitor and analyze network traffic to protect a system from network based threats. To capture all the data passing through the network, you need to position your IDS at the entry and exit point of data from your network to the outside world. The fuzzy logic minimizes the false alarm rate in determining intrusive activities. A set of fuzzy rules can be employed to identify normal and abnormal behavior in computer networks. The fuzzy inference logic can be applied for determining an intrusion. The purpose of introducing fuzzy logic is to deal with the fuzzy boundary between the normal and abnormal classes.

KDDCUP99 DATASET

Since 1999, KDDCup99 has been the most wildly used data set for the evaluation of anomaly detection methods. In 1998, DARPA in concert with Lincoln Laboratory at MIT launched the DARPA 1998 dataset for evaluating IDS. The refined version of DARPA dataset which contains only network data (i.e. Tcpdump data) is termed as KDD dataset. The KDD training dataset consist of 10% of original dataset that is approximately 494,020 single connection vectors each of which contains 41 features and is labelled with exact one specific attack type i.e., either normal or an attack. Any deviation from this Normal behavior is said to be an attack. Each vector is labelled as either normal or an attack, with exactly one specific attack type. A smaller version 10% training dataset is also provided for memory constrained machine learning methods. The training dataset has 19.69% normal and 80.31% attack connections. KDD CUP 99 has been most widely used in attacks on network. The attack falls in one of the following four categories:

TABLE-1 TYPE OF ATTACKS IN KDDCup 99 DATASET

CATEGORY	ATTACK TYPE
Denial of attack	Back, Land, Neptune, Pod, Smurf, Teardrop
Probe	Ipsweep, nmap, portsweep, satan
Root to Local	Fp_write, guess_password, imap, multihop, phf, spy, warezclient, warezmaster
User to Remote	Buffer_overflow, loadmodule, perl, rootkit

Increasing the accuracy detection rate for anomalies and improving the efficiency of intrusion detection models is the major objective. It detects the training dataset consisted as normal, probe, U2L, R2L and DOS. The 10% of KDD 99 training dataset has three distinct protocols namely TCP, UDP, and ICMP.

METRICS FOR IDS EVALUATION:

The Intrusion Detection System can be evaluated in many ways. The IDS can be generally evaluated using Efficiency and Effectiveness.
Efficiency: It measures the resources needed for the system including CPU cycles and main memory.
Effectiveness: It represents the ability of the system to distinguish between intrusive and non-intrusive activities.

CONFUSION MATRIX:

Confusion matrix represents the result of classification. It represents true or false classification results. IDS accuracy can be defined in terms of:
True Positive (TP): Number of intrusions correctly detected
True Negative (TN): Number of non-intrusions correctly detected
False Positive (FP): Number of non-intrusions incorrectly detected
False Negative (FN): Number of intrusions incorrectly detected

PERFORMANCE MATRIX:

Accuracy = TP + TN / TP + TN + FP + FN

Accuracy is the ratio of total number of correctly predicted instances to the total number of instances.

Precision = TP / TP + FP

The precision measures the number of correct classifications penalized by the number of incorrect classifications.

Recall = TP / TP+FN

Recall is the ratio of correctly predicted positive observations to the all observations in actual class.

F-Score

F-Score is calculated by considering both the metrics of precision and recall equally.

TABLE 2: The 41 features of KDD Cup99 dataset

Feature Index	Feature name	Description	Type
1	duration	length (number of seconds) of the connection	continuous
2	protocol_type	type of the protocol, e.g. tcp, udp, etc.	symbolic
3	service	network service on the destination, e.g., http, telnet, etc.	symbolic
4	flag	normal or error status of the connection	symbolic
5	src_bytes	number of data bytes from source to destination	continuous
6	dst_bytes	number of data bytes from destination to source	continuous
7	Land	1 if connection is from/to the same host/port; 0 otherwise	symbolic
8	wrong_fragment	number of ``wrong" fragments	continuous
9	urgent	number of urgent packets	continuous
10	hot	number of ``hot" indicators	continuous
11	num_failed_logins	number of failed login attempts	continuous
12	logged_in	1 if successfully logged in; 0 otherwise	symbolic
13	num_compromised	number of ``compromised" conditions	continuous
14	root_shell	1 if root shell is obtained; 0 otherwise	continuous
15	su_attempted	1 if ``su root" command attempted; 0 otherwise	continuous
16	num_root	number of ``root" accesses	continuous
17	num_file_creations	number of file creation operations	continuous
18	num_shells	number of shell prompts	continuous
19	num_access_files	number of operations on access control files	continuous
20	num_outbound_cmds	number of outbound commands in an ftp session	continuous
21	is_hot_login	1 if the login belongs to the ``hot" list; 0 otherwise	symbolic
22	is_guest_login	1 if the login is a ``guest" login; 0 otherwise	symbolic
23	count	number of connections to the same host as the current connection in the past two seconds	continuous
24	srv_count	number of connections to the same service as the current connection in the past two seconds	continuous
25	serror_rate	% of connections that have ``SYN" errors	continuous
26	srv_serror_rate	% of connections that have ``SYN" errors	continuous
27	rerror_rate	% of connections that have ``REJ" errors	continuous
28	srv_rerror_rate	% of connections that have ``REJ" errors	continuous
29	same_srv_rate	% of connections to the same service	continuous
30	diff_srv_rate	% of connections to different services	continuous
31	srv_diff_host_rate	% of connections to different hosts	continuous
32	dst_host_count	count for destination host	continuous
33	dst_host_srv_count	srv_count for destination host	continuous
34	dst_host_same_srv_rate	same_srv_rate for destination host	continuous

35	dst_host_diff_srv_rate	diff_srv_rate for destination host	continuous
36	dst_host_same_src_port_rate	same_src_port_rate for destination host	continuous
37	dst_host_srv_diff_host_rate	diff_host_rate for destination host	continuous
38	dst_host_serror_rate	serror_rate for destination host	continuous
39	dst_host_srv_serror_rate	srv_serror_rate for destination host	continuous
40	dst_host_rerror_rate	rerror_rate for destination host	continuous
41	dst_host_srv_rerror_rate	srv_serror_rate for destination host	continuous

FUZZY LOGIC

Fuzzy set was published in 1965 by Lotfi A. Zadeh. Fuzzy logic is based on possibility theory. The fuzzy logic provides a work space for computation with words and offers a hand in managing uncertainty during the designing of expert systems. It has now become an unavoidable part of machine learning as it can handle imprecise and uncertain situation. Machine learning tools are extensively used for intrusion detection. These algorithms build detection model. Fuzzy logic is a superset of Boolean logic that has been extended to handle the concept of truth values between completely true and completely false. While boolean logic only allows true or false, fuzzy logic allows all things in between.

FUZZY LOGIC ARCHITECTURE

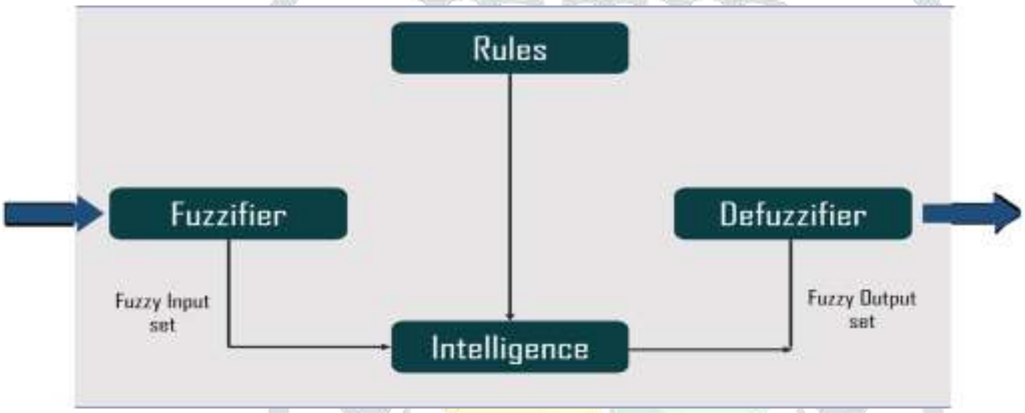


Fig 2: Fuzzy Logic Architecture

The fuzzy logic architecture consists of four main parts:

Rules– It contains all the rules and the if-then conditions offered by the experts to control the decision-making system. The recent update in the fuzzy theory provides different effective methods for the design and tuning of fuzzy controllers. Usually, these developments reduce the number of fuzzy rules.

Fuzzification– This step converts inputs or the crisp numbers into fuzzy sets. You can measure the crisp inputs by sensors and pass them into the control system for further processing. It splits the input signal into five steps such as-

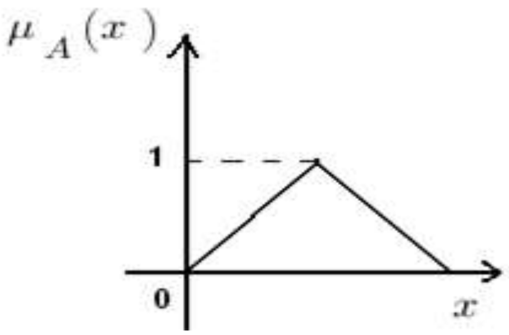
Inference Engine– It determines the degree of match between fuzzy input and the rules. According to the input field, it will decide the rules that are to be fired. Combining the fired rules, form the control actions.

Defuzzification– The Defuzzification process converts the fuzzy sets into a crisp value. There are different types of techniques available, and you need to select the best-suited one with an expert system.

Membership Function

The membership function is a graph that defines how each point in the input space is mapped to membership value between 0 and 1. It allows you to quantify linguistic terms and represent a fuzzy set graphically. A membership function for a fuzzy set A on the universe of discourse X is defined as $\mu A:X \rightarrow [0,1]$

It quantifies the degree of membership of the element in X to the fuzzy set A.



-
- X-axis represents the universe of discourse.
- Y-axis represents the degrees of membership in the [0, 1] interval.

There can be multiple membership functions applicable to fuzzify a numerical value. Simple membership functions are used as the complex functions do not add precision in the output. The triangular membership function shapes are most common among various other membership function shapes.

CONCLUSION

In this paper, I have described an overview of fuzzy logic for intrusion detection system. A fuzzy logic based system can be able to detect the intrusion behavior within a network. An effectual fuzzy rule makes an effective intrusion detection. The rules are identified by fuzzification then given to fuzzy system for classifying the test data. The KDDCup99 dataset is mostly used for intrusion detection. The Fuzzy logic is used in various fields such as automotive systems, domestic goods, environment control, etc.

REFERENCES

[1] Ajeesha M I, Dr. D Francis Xavier Christopher,*"*MACHINE LEARNING TECHNIQUES FOR INTRUSION DETECTION*"* 2019 JETIR June 2019, Volume 6, Issue 6.

[2] *Fuzzy Membership Functions and its Features (tech-wonders.com)*

[3] Gulshan Kumar, "Evaluation Metrics for Intrusion Detection Systems- A Study", International Journal of Computer Science and Mobile Applications, Vol.2 Issue. 11, November- 2014, pg. 11-17

[4] John E. Dickerson, Julie A. Dickerson, *"*Fuzzy Network Profiling for Intrusion Detection*"*, DOI: 10.1109/NAFIPS.2000.877441 · Source: IEEE Xplore, FEBRUARY 2000

[5] MOHAMMAD ALMSEIDIN, 2SZILVESZTER KOVACS, *"*INTRUSION DETECTION MECHANISM USING FUZZY RULE INTERPOLATION*"*.

[6] Mostaque Md. Morshedur Hassan, *"*CURRENT STUDIES ON INTRUSION DETECTION SYSTEM, GENETIC ALGORITHM AND FUZZY LOGIC*"*, International Journal of Distributed and Parallel Systems (IJDPS) Vol.4, No.2, March 2013.

[7] R. Shanmugavadivu, Dr. N. Nagarajan, "NETWORK INTRUSION DETECTION SYSTEM USING FUZZY LOGIC", Indian Journal of Computer Science and Engineering (IJCSE)

[8] Shailesh P. Thakare, Dr. M. S. Ali "Introducing Fuzzy Logic in Network Intrusion Detection System" International Journal of Advanced Research in Computer Science, Volume 3, No. 3, May-June 2012.

[9] Tahir Mehmood1 and Helmi B Md Rais2,"Machine learning algorithms in context of intrusion detection", 2016 3rd International Conference On Computer And Information Sciences", 2016 3rd International Conference On Computer And Information Sciences (ICCOINS)

[10] Vishnu Balan E, Priyan M K, Gokulnath C, Prof. Usha Devi G, *"*Fuzzy Based Intrusion Detection Systems in MANET*"*, Procedia Computer Science 50(2015) 109 – 114.

[11] What is Fuzzy Logic in AI and What are its Applications? | Edureka

ABOUT THE AUTHOR

Mr. Muhamed Jamshir M received his M. Tech in the area of Electronics Design and Technology from National Institute of Technology, Calicut in 2012. He obtained his B. Tech, in Electronics and communication engineering from Government Engineering College, Trissur, in 2010 from Calicut University. At present he is working as an Ad hoc Faculty, NSS College of Engineering Akathethara, Palakkad. His research interest lies in the area of Networking and Machine Learning. He served as a key note speaker for various seminars country wide.

Ms. Ajeesha M I currently pursuing PhD in the area of Data mining/Machine Learning from Rathnavel Subramaniam College of Arts & Science affiliated to Bharathiar University Coimbatore. She obtained M. Phil in the area of Data Mining from Rathnavel Subramaniam College of Arts & Science affiliated to Bharathiar University Coimbatore in 2018. Her research interests are Data Mining and Machine Learning.