Intrusion Detection System with Machine Learning Algorithms and Comparison Analysis

JAIN KARAN ANAND

AIT-CSE, CHANDIGARH UNIVERSITY, INDIA

Abstract: The incremental increase in the usage of technology has led to an increase in the amount of data that is being processed over the Internet significantly over the time period. With the huge amount of data that is being flown over the Internet, comes the scenario of providing security to the data, and this is where an Intrusion Detection System (IDS) comes into the picture and helps in detecting any virtual security threats. The IDS which is being proposed is being implemented using latest technologies such as Machine Learning Algorithms to classify the attacks and detecting them whenever an attack happens and also to find which machine learning algorithm is best suitable for identifying the attack.

Keywords: **Intrusion Detection System (IDS), Machine Learning, Correlation Engine, Vulnerability Assessment, True Alarm, False Positive.**

1. INTRODUCTION

The computer systems and computer networks play a very important role with the increased growth of the technology in today's world. The old manual systems of businesses, enterprises and organizations are being automated with computerized solutions due to rapid changes that are happening in the world. Thus, due to lots of processing, huge amount of data/information gets generated and is stored in the systems, which needs security. Here, Information Security comes into action. Information Security helps in preventing unauthorized access, modification of information, destroying the information. The Information may be physical or electrical. The objectives of Information Security are CIA (C-Confidentiality, I-Integrity, A-Availability). Information Security has been evolved over the years and offers many areas of specialization. Areas such as Securing Networks, Allied Infrastructure, etc comes under Information Security. Some of the advantages of Information Security are protecting the system form malicious attacks, securing confidential information, etc.

An Intrusion Detection System (IDS) helps to monitor network and systems for detecting any malicious activity. An Intrusion Detection System helps to prevent intrusions with the help of set of rules that have been defined. An Intrusion Detection System helps the system to be reliable and secure. In general, IDS tries to detect the intrusions and an alert gets generated based on the set of rules that have been defined for any such intrusion that occurs within a system or a network.

Intrusion Detection System generally can be categorized into Host based Intrusion Detection System (HIDS) and Network based Intrusion Detection System (NIDS). HIDS deals with monitoring the internals of a computing system and analyzing them. NIDS deals with monitoring the network traffic and look for any unauthorized access or a suspicious activity. The detection method of an Intrusion Detection System can be either a Signature based Method or an Anomaly based Method. Though the Intrusion Detection System monitors for malicious activities, they are prone to generate false alarms. When IDS is developed, the false alarm rate should be less when compared to any other existing Intrusion Detection Systems.

The detection of an intrusion begins where a firewall ends. Prevention of an unauthorized access is best, but it is not always possible. Intrusion Detection can be defined as real time monitoring and analysis of data and network analysis for any possible vulnerabilities and attacks that are in progress. Major limitation of many existing Intrusion Detection Systems is filtering of false alarms and detecting true attacks. Many machine learning algorithms are being used for Intrusion Detection now-a-days. These machine learning algorithms when implemented can detect attacks which may be either a true attack or a false alarm. An Intrusion Detection System needs to be accurate for detection of true attacks.

The existing Intrusion Detection Systems use either a signature based detection or an anomaly based detection as the detection method for intrusions. These detection methods mentioned above have some problems in them. The signature based detection method sometimes tends to generate false positive due to its capability of detecting only known attacks. While the anomaly based detection method generate alerts for even a legitimate events and network traffic resulting in higher false positive rate. Real time attack detection is still a challenge for the existing Intrusion Detection Systems. The existing systems cannot authenticate remote login users. When the existing Intrusion Systems

will give higher rate of false positives then the security analysts cannot depend on them. Zero day attack rate has been increasing significantly over the time, which are acting as threats for the existing intrusion detection systems. At present there is no perfect solution in the Information Security industry for detecting the intrusions.

The proposed Intrusion Detection System uses Machine Learning Algorithms to solve the classification problem of pattern recognition and intrusion identification. This system provides us the comparison analysis of the different machine learning algorithms that are implemented for classification of attacks and helps in identifying the intrusions when occurred. This system will help in overcoming the disadvantage of the existing Intrusion Detection Systems. The proposed system even helps in improving the accuracy of detection rate when compared to the existing intrusion detection systems. The true positive rate is also increased thus resulting in better detection of intrusion that are true and are not false.

2. LITERATURE SURVEY

Existing Tools for Intrusion Detection System:

- Snort: Snort is one of the Intrusion Detection tool and it is programmed in C language. Martin Roesch developed Snort in 1998. It is open source software and is currently under Cicso. Snort acts as a real-time traffic monitor.

- OSSEC: OSSEC known as Open Source HIDS SECurity. It helps in performing log analysis. OSSEC also performs integrity checking, rootkit detection, alerting in real time and active response.

- Prelude SIEM: Prelude SIEM is a hybrid Intrusion Detection System. Prelude SIEM generates alerts when an Intrusion or any security threats occurs within a network in real time.

- Suricata: Suricata acts as an intrusion detection and prevention system. It acts as a complete network security monitoring system. When compared with Snort, the advantage with Suricata is that it works all its way to the application layer.

- AIDE: AIDE known as Advanced Intrusion Detection Environment is another free intrusion detection system. Its main focus is on rootkit detection and file signature comparisons. It uses both signature based analysis and anomaly based analysis which is run on demand. It is Host based Intrusion Detection System for UNIX and LINUX.

- Zeek: Zeek was called formerly as Bro. It is free and open source software for Network Intrusion Detection. With the help of Zeek we can also perform live analysis of network events.

SNORT:

Snort is the tool that we are going to use for detecting Intrusions for the system that has been proposed. Snort as mentioned above is owned by CISCO Systems and it is an open source software and can be used free of cost. This software can be installed in Windows, UNIX and Linux Operating Systems. Snort is used to detect any intrusions that occur within a network. Snort can be operated in three modes Sniffer mode, Packet Logger mode and Intrusion Detection Mode. In the Intrusion Detection mode, Snort can be able to monitor the network traffic against a rule set that has been defined and it takes action based on what that has been identified. Snort allows users to define their own rules in the rules file. Generation of alerts is also a feature that is available in Snort and which makes it more efficient.

3. METHODOLOGY

The system that has been proposed is going to be implemented in Python and KDD Datasets are going to be used. The datasets that have been downloaded contains training dataset and testing datasets with four different classes for the Intrusions. The attacks that are going to be detected are Denial of Service attack, User to Root (U2R) attack, Remote to Local (R2L) attack and Probe attack. These are the four attacks that are classified into classes of intrusions.

The downloaded datasets for the Intrusion Detection System are applied on the Machine Learning Algorithms. The machine learning algorithms that are being implemented in the system that has been proposed are Decision Tree Algorithm, Random Forest Algorithm, Logistic Regression Algorithm and KNN Algorithms. These four machine learning algorithms that are being implemented are classification machine learning algorithms that are used for Classification purpose.

The Decision Tree Algorithm is a supervised machine learning algorithm, mostly used for the purpose of Classification. In this machine learning algorithm we split the sample into two or more homogeneous sets based on the most significant splitter or differentiator in input variables. In the Decision Tree Algorithm, the internal node represents a test on the attribute, while the branch depicts the outcome and leaf represents the decision that has been made after computing attribute.

The Logistic Regression Algorithm is also a supervised machine learning algorithm and used for Classification. Logistic Regression is a regression model, which builds a model to predict the probability that the data belongs to a particular category. The threshold value setting is very important aspect of the Logistic Regression because Logistic Regression becomes a classification technique only when decision threshold value is set.

The Random Forest Algorithm is a machine learning algorithm that is capable of performing classification as well as regression based on the requirement. It uses Bootstrap Aggregation (also known as Bagging) for performing classification or regression with the help of multiple decision trees. The output is determined by combining multiple decision trees rather than depending on a single decision tree.

The Random Forest Algorithm is a machine learning algorithm that is capable of performing classification as well as regression based on the requirement. It uses Bootstrap Aggregation (also known as Bagging) for performing classification or regression with the help of multiple decision trees. The output is determined by combining multiple decision trees rather than depending on a single decision tree.

The KNN (K Nearest Neighbor) Algorithm is another supervised machine learning algorithm that is used for solving classification problems. In this algorithm the distance between a query scenario and a set of scenarios is calculated by using a distance function, for example Euclidian Distance formula. It is a competitive learning algorithm because it internally uses competition between the data instances in order to take predictive decisions.

The datasets that have been downloaded are taken as input by the four machine learning algorithms that are being implemented. The input is processed and attacks are classified. The testing data set will not be of the same probability distribution as that of the training dataset. The testing dataset also has attacks that are not present in the training dataset. Thus it makes the task to be more realistic.

The accuracy of intrusion detection is calculated for the four machine learning algorithms that are being implemented. A comparison analysis report is generated for the accuracy rate in percentage for the four machine learning algorithms in the form of a graph. From the graph we can then clearly analyze which algorithm is more efficient for the attacks that have been classified as classes of intrusion earlier. Not only the accuracy rate of intrusion detection is calculated, the mean square error rate, mean average error rate are also being calculated for a proper analysis.

We have rule-sets defined in Snort in snort.conf file. When Snort is configured to Intrusion Detection mode, based on the rule-sets defined in the snort.conf file the network is monitored to detect an intrusion when occurred so that an action can be taken based on what has been identified. Snort here acts as Vulnerability Assessment tool.

The comparison analysis graph generated for the four machine learning algorithms helps us to understand which algorithm works efficiently for the attacks that we have considered for the system that has been proposed. The detection rate of each attack is plotted in a graph for better understanding.

Snort can also perform protocol analysis with which detection of different types of attacks and probes can ne possible. The snort config file is available at / etc / snort / snort.conf. We can include some rules in rule file. Snort rule structure is as follows:

<Rule Actions> <Protocol> <Source IP Address> <Source Port> <Direction Operator> <Destination Port>

We can insert some snort rules in our rule file:

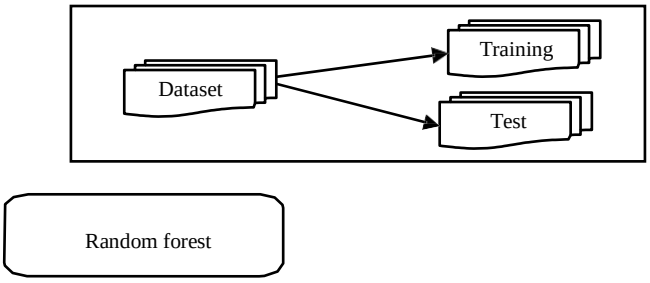
1. For alert

Example: alert icmp any any -> any any (msg: ”Ping Detected”; sid:1001466; rev:1)
This rule makes alerts on each and every icmp packet as well as shows source and destination port addresses.

2. To Run snort

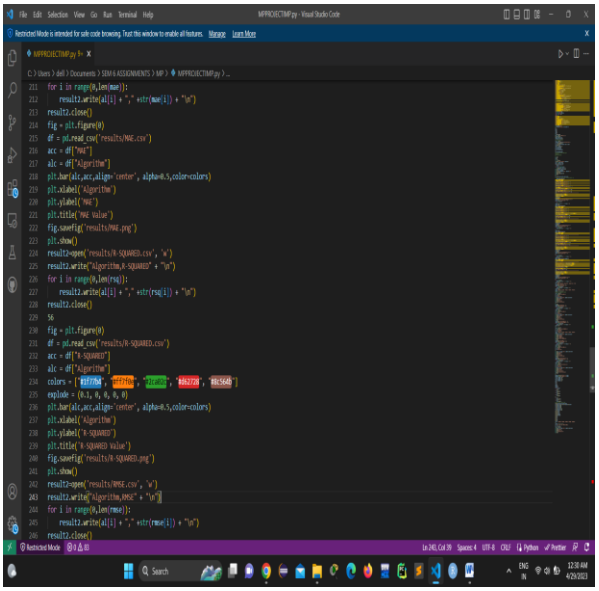
Snort –dev –c /etc/snort/snort.conf -1 /var/log/snort/ -i eth0 –A full –k none

Architecture of the system:

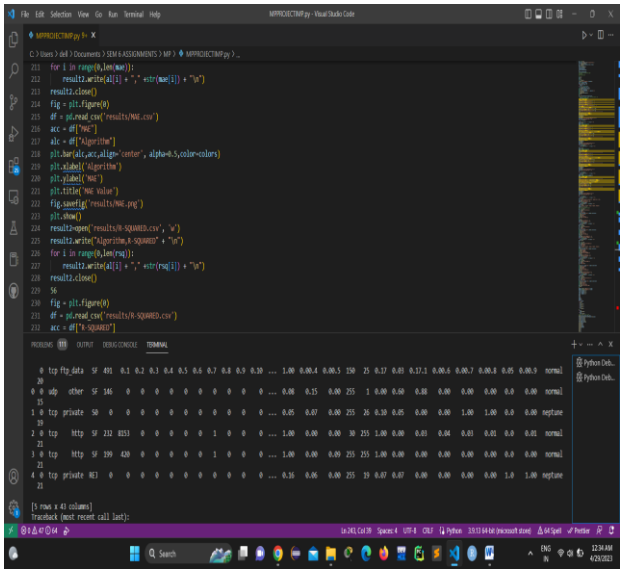


4. RESULT & DISCUSSION

The main disadvantage in the existing IDSs is the higher rate of false alarms. To excel in the reliability of a security system, we need to lower the rate of false alarms or false positives. The system that has been proposed helps to detect the true attacks because of the four machine learning algorithms that have been implemented. Even if a particular attack is missed from getting detected in an algorithm, it would get detected in any of the other algorithms, thus the detection rate of the intrusion is high. With the help of Snort, the Vulnerability Assessment is performed with defined rule-sets, for which an action is taken for an intrusion that has been identified. The comparison analysis helps us to clearly understand which algorithm is efficiently detecting the attacks. The use of machine learning algorithms helps us to classify the true attacks and false alarms, thus helping us to build good IDS that can decrease the false alarm rate and can provide reliability and security.



Traditional IDSs available today has its own relative weaknesses and strengths. While one solution may be strong at host-based intrusion detection, the other solution may be strong at network-based intrusion detection. The organizations are highly concerned about their network and system performance; hence they use multiple IDSs from various vendors as they do not wish to take a chance with security. Different IDSs generate alert events in different formats, as well as use different protocols. If the outputs alerts are not integrated properly, false positive rates may increase hence interrupting the legitimate performance of a system or a network. False alarms caused by the large volume of IDSs is intolerable to the administrators as it delays the smooth functioning of an organization. It is necessary to decrease the excessive of false alarms to reduce the operational cost and excel in the reliability of a security system. Hence, this research was conducted intending to advance a procedure to obtain alerts from different sensors and standardizes them into IDMEF.



5. 5. FUTURE WORKS & CONCLUSION

The research intended to introduce an advanced machine learning and rule-based, HIDS and NIDS correlated intrusion detection system. The system gives an optimized and reliable output which creates a fewer false positive rate compared to the past researches and existing IDS solutions. Further research can be conducted in developing an advanced intrusion detection system using the proposed approach. There are various open source IDS tools which can further be integrated with the proposed architecture to compare findings to fin the best possible combination. The overall objective is to achieve a more successful result in order to persevere against the modern types of attacks, which cannot be discovered by the traditional standalone Intrusion Detection System.

The system that has been proposed can be made more reliable and efficient by implementing other machine learning algorithms along with the ones that already have been implemented so that intrusion can be detected easily. Also the other types of attacks can also be classified as the classes of intrusion to identify more attacks and provide more security and reliability. Thus further development of the system can help to increase

the detection rate and lower the false positive rates.

6. REFERENCES

[1] A B. Athira, V. Pathari, "Standardisation and Classification of Alerts Generated by Intrusion Detection Systems", IJCI, International Journal on Cybernetics & Informatics, Vol 5 Issue 2, 2016.

[2] Johansson Daniel, Andersson Par, "Intrusion Detection Systems with Correlation Capabilities"

[3] Yasm Curt, "Prelude as a Hybrid IDS Framework", March, 2009

[4] Kumar Vinod, Sangwan Prakash Om, "Signature Based Intrusion Detection System Using SNORT", IJCAIT, International Journal of Computer Applications & Information Technology, Vol. I, Issue III, November 2012.

[5] Singh Deepak Kumar, Gupta Jitendra Kumar, "An approach for Anomaly based Intrusion detection System using SNORT", IJSER, International Journal of Scientific & Engineering Research, Volume 4, Issue 9, September 2013.

[6] S, Vijayarani, and Maria Sylviaa S. "Intrusion Detection System – A Study", IJSPTM, International Journal of Security, Privacy and Trust Management ,Vol 4, Issue 1, pp. 31–44, February 2015.

[7] Yang Guangming, Chen Dongming, Xu Jian, Zhu Zhiliang, "Research of Intrusion Detection System Based on Vulnerability Scanner", ICACC, Advanced Computer Control, March 2010.

[8] Chakraborty Nilotpal, "Intrusion Detection System and Intrusion Prevention System: A Comparative Study", IJCBR, International Journal of Computing and Business Research , Volume 4 Issue 2, May 2013.

[9] **[9]** Kothari Pravin, "Intrusion Detection Interoperability and Standardization", February, 2002.

[10] TIMOFTE Jack, "Intrusion Detection using Open Source Tools", Revista Informatica Economică nr.2(46), pp. 75-79, 2008