INTERNATIONAL JOURNAL OF
RECENT TRENDS IN MULTIDISCIPLINARY RESEARCH
https://www.doi.org/10.59256/ijrtmr.20240402018
March-April 2024, Vol 4(02), 105-109.
ISSN No: 2583-0368

A Review: Machine Learning Approaches for Zero-Day Attacks Recognition

Sivakumar Nagarajan
Technical Architect, I & I Software Inc, 2571 Baglyos Circle, Suite B-32, Bethlehem, PA-18020, USA

OPEN ACCESS
Article Citation:
Sivakumar Nagarajan "A Review: Machine Learning Approaches for Zero-Day Attacks Recognition", International Journal of Recent Trends in Multidisciplinary Research,
March-April 2024, Vol 4(02), 105-109

Publishedby5thDimension Research Publication

Abstract: Advancements and rapid technological developments in information age promoted computerization of day to day activities in all walks of life. As technology is growing rapidly, the cybercrime rate also increases both in number and complexity. Since a variety of attacks evolves regularly with complex patterns and varied signatures the task of securing cyberspace becomes more and more difficult and challenging. To minimize the impact of cybercrime through early detection of intrusions, network activity in terms of network traffic, is monitored in real-time thus accumulating huge data which is sometimes erroneous. Therefore, synergizing the concepts of cybersecurity and data analytics is essential to develop effective security algorithms for attack detection. Existing security measures like firewalls are no longer sufficient to deal with these emerging attacks. Because the firewall only checks the header of the data packet, it doesn't go through the details of the packet. Intrusion Detection Systems (IDSs) are the systems introduced as a second line of security after firewalls to handle cyber intrusions more efficiently. IDSs play a key role in protecting cyberspace by examining the entire details of the traffic packets to detect intrusions.

Key Words: Intrusion Detection Systems, network traffic, machine learning, Zero-day attack.

1. Introduction

The tremendous growth in the technological developments has resulted in the digitization of data in many aspects of life. Ease of internet usage and its availability at affordable costs to connect internet users globally facilitating huge data transfer in different forms and the lack of disciplined behavior in cyberspace, etc., paves the way towards vulnerabilities and threats related to digital information security. The process of exploiting the vulnerabilities leads to cyberattacks/ cyber-crimes specifically causing the breach of Confidentiality, Integrity, and Availability known as "CIA triad". Cyber-attacks are described as any unauthorized or illegal activity of penetrating the cyberspace by breaking the security bounds. Different types of popular cyber-attacks are trojans, viruses, worms, spyware, rootkit, etc. Trojans are the suspicious software programs that hide into a legitimate host and allow attackers to access the host. Viruses and worms are malicious programs that can replicate themselves into multiple files and perform malicious operations.

They usually enter the system through an external storage device like hard disk, pen-drive, etc. or by clicking unknown links or by downloading malicious email attachments, etc. Upon entering into the host system, viruses start affecting the system immediately after their activation by the host system, whereas, worms are standalone as they are activated and start affecting the host independently. Elk Cloner, Blaster, Nimda, etc. are familiar examples of viruses. whereas Stuxnet, Witty, SQL Slammer, Code Red, Conifer, etc. are familiar examples of worms. Spyware is software that covertly hides in a victim's system and runs in background continuously thereby collects and reports all sensitive information to the attacker. A rootkit is a dangerous software program that hides its existence and allows attackers to take control of the victim's system by gaining root privileges.

The motive behind the cyber-crime may be anything like wrongful financial gains, political enmity, terrorist act, casual hobby for some hackers, etc. Different types of cyber-attacks aim to attack different types of victims. Specifically, attacks like spamming, password sniffing, computer sabotage, cyber-stalking, cyber defamation, phishing, email spoofing, etc. aim at individuals. Attacks like creditcard frauds, internet-time theft, violation of intellectual property rights, etc. aim at a wrongful gain of resources/finance. Attacks like Denial of Service (DoS), virus, mail bomb, salami attack, logic bomb, trojan horse, data

diddling, industrial espionage, etc. are aiming against organizations and finally, attacks like forgery, web jacking, cyber terrorism, etc. are aiming at the society at large. Cyber-attacks pose serious national and economic security challenges and hence, effective security mechanisms like Intrusion Detection Systems need to be developed to control the damage caused by these cyber-attacks.

2. Intrusion Detection Systems

Intrusion Detection Systems (IDSs) are the hardware or software systems that constantly monitor the cyberspace for detecting the intrusions [Fun17]. On detection of an intrusion, IDS generates an alert to the system administrator to follow-up further action in response to the alert. Intrusion detection systems are deployed next to the firewalls in the detection process and their main motto is defense-in-depth. Firewalls don't go through the internal details of the traffic, whereas, intrusion detection systems employ detailed detection procedures by capturing the internal and implicit details like the signature of the traffic. The basic architecture of an Intrusion Detection System is shown in the Figure 1. As shown in the figure, IDS consists of many components; monitoring system, data gathering devise, ID engine, knowledge base, configuration device, and response component. Data gathering devise receives raw data from the monitoring system and transforms it to identify events which will be submitted to the ID engine for detection of possible intrusions. ID engine analyses the submitted data by taking required information from the knowledge base and configuration devices where, knowledgebase provides information about known attacks like their signatures, characteristics, etc. and configuration devices provide information about the system or network configurations details. Upon detection of an attack, the ID engine sends an alert to the response component which carries out further action.

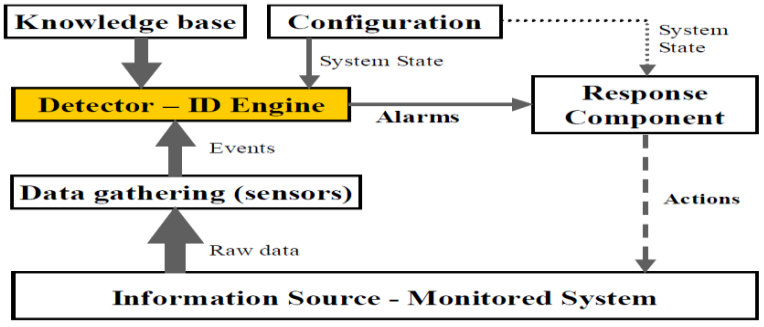


Figure 1: Basic architecture of IDS [Laz05]

3. Types of Intrusion Detection Systems

Depending on the scope of defense there are two types of intrusion detection systems. One is the Host-based intrusion detection system and the other is am Network-based intrusion detection system.

Host-based Intrusion Detection System (HIDS)

The host-based intrusion detection system is deployed on an individual host that is supposed to be susceptible to possible outbreaks. HIDS examines the host for identifying suspicious and unauthorized events. The examples of the host-based intrusion detection systems are OS-SEC, Tripwire, etc. The scenario of host-basedn IDS is shown in Figure 2 Host-based intrusion detection systems generally aim to detect attacks made by the insiders by focusing on the misuse of privileges. To identify intrusions HIDS monitors the activities of the host at a higher level of detail like audit trails, operating system commands, system calls, input traffic packets targeting the host, noutput packets from the host, etc. HIDS extensively relies on the host operating system and monitors audit trails as their main source of evidence. Audit trails are collected periodically or continuously in real-time, and the collected audit trails are compared to the existing security norms for suspecting/identifying intrusions.

With the help of audit trails, HIDSs can detect attacks with higher accuracy and blower false-positive rates. But, at the same time, this analysis of the audit trails puts substantial overheads on the performance of the host. Specifically, it increases the server load and consumes more memory, CPU time, storage, and other resources of the host which can be very expensive.

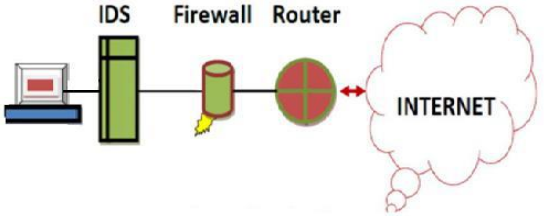


Figure 2: Scenario of HIDS

Network-based Intrusion Detection System (NIDS)

The Network-based Intrusion Detection System is deployed on an entire network and it is often aimed to detect attacks made by the outsiders by focusing on the exploitation of vulnerabilities. For identifying suspicious activity, NIDS monitors both

the inward and outward traffic of the network. Specifically, NIDS inspect events as packets of information exchange between hosts. For effective detection of network intrusions, a NIDS is placed at the network boundaries or between the network and the server. Snort, Bro, etc. are examples of network-based intrusion detection systems. The scenario of NIDS is shown in Figure 3

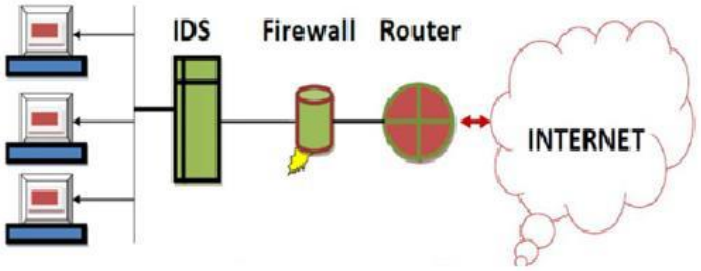


Figure 1.3: Scenario of NIDS

4. Detection Strategies of IDS

Depending on the detection strategies/approaches used for intrusion detection, IDSs are broadly categorized into Signature-based Intrusion Detection Systems and Anomaly-based Detection Systems.

Signature-based Detection Approach

Signature captures the pattern of a group of packets which is defined in terms of their shared features. Attack signatures capture the patterns of attack packets. Signature-based intrusion detection is done by matching attack signatures. The matching process involves comparing the features of a new traffic packet with those presented in the existing set of attack signatures. On detecting the matched signature an alert is sent to the system administrator regarding the detected intrusion. The advantage of a signature-based detection approach is that it can detect known attacks with high accuracy and minimal false-positive rate. This advantage follows a major drawback also; the signature-based approach is applicable to identify known attacks based on their signatures. It cannot detect unknown attacks for which signatures are not available in its knowledge base.

When a new attack is observed, the IDS developers will wait until a sufficient number of new attack packets are collected to create its signature and accordingly update the IDS knowledge base. The IDS which implements the signature-based detection approach is referred to as signature-based IDS.

Anomaly-based Detection Approach

In Anomaly-based intrusion detection system the normal network behavior/activity is modeled in terms of the shared features of genuine traffic packets like the number of ports, the number of devices connected, bandwidth, protocols defined, etc. and detection is done by observing the deviations of a packet activity from the observed normal behavior model. Specifically, any packe that deviates from the baseline of normal behavior is suspected as an attack packet. The suspected anomaly could be either a known attack or an unknown attack, the details of which are given in the next sections of this chapter. The IDS that implements an anomaly-based detection approach is referred to as anomaly-based IDS and is shown in Figure 1.4.

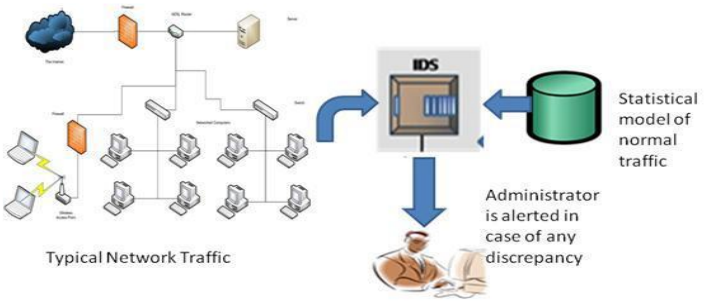


Figure 1.4: Anomaly-based IDS

Though anomaly-based intrusion detection systems are deployed to detect unknown attacks they run into a bottleneck on sensitivity control. Less sensitive anomaly-based IDSs may miss some critical intrusions whereas sensitive IDSs may raise too many false alarms, as a result of their higher False Positive Rate (FPR) control setting. Each false alarm is expensive as it calls for security expert intervention to decide whether it is an attack or not. So, it is better to reduce false alarms while keeping the detection accuracy as high as possible in an intrusion detection system.

Hybrid Approach to Intrusion Detection

To combine the advantages of both signature-based and anomaly-based approaches, some of the researchers proposed hybrid systems where a hybrid IDS implements both signature-based and anomaly-based approaches. Hybrid IDS is implemented in two phases. Initially, all known attacks are detected using a signature-based detection approach in the first phase. Later, all the benign traffic filtered from the first phase is sent to the anomaly-based detection approach in the second phase, where suspicious unknown attack packets are separated from the benign packets. In other words, once the signature-

based IDS filters out the known attack packets, the residual benign traffic when passed through the anomaly-based IDS gets separated into normal traffic and the suspicious unknown attack traffic. This combined advantage gives more robustness in detecting attacks, although perhaps at the cost of operating a bit slowly, possibly, causing a lag in detection.

5. ML for Supervisory Learning in IDS

Intrusion detection systems need to analyze huge amounts of traffic information for extracting hidden patterns of features to form attack signatures and hence require Machine Learning (ML) techniques. Machine learning methods are broadly categorized as supervised and unsupervised methods. Supervised methods predict class labels of instances based on the knowledge learned from a large collection of labeled data. Whenever labeled data is not available unsupervised methods like clustering, statistical modeling, etc., are used as they do not require labeled data for pattern extraction. Since intrusion detection systems generally maintain labeled data related to known attacks and genuine traffic packets (benign), supervised methods are mostly preferred for the construction of signature-based IDS.

Machine learning has proven its potential for developing effective IDS for detecting known attacks using the signature-based approach. Many supervisory learning algorithms like K-Nearest Neighbor (KNN), Decision Tree (DT) [Ari17], Random Forest (RF), Support Vector Machine (SVM), Naïve Bayes (NB) [Wah15], Logistic Regression (LR), etc., are extensively used for implementing signature-based IDS, and a few of them are described below in brief.

6. Data Pre-processing

Data pre-processing should be done to prepare the data before building a classifier.Pre-processing is the task of cleaning and transforming the dataset and making itready for model construction. Models constructed without proper pre-processing may not give reliable results. Pre-processing generally includes the following tasks:

- **Handling Missing Values:** Data may include some missing values; those values should be handled either by removing them or replacing them with some suitable value like mean.
- **Removing Insignificant Fields:** Data may contain some sparse fields dominated by zero or null values and contribute nothing to define a pattern. Those fields should be identified and removed.
- **Handling Mixed Type of Features:** Some machine learning algorithms like KNN deals with numeric data only. But real-world network traffic data may also include some categorical features. So, these categorical features need to be transformed into a numeric form by adopting proper conversion methods.
- **Data Normalization:** Normalization is the process of scaling the descriptive features of the dataset into a uniform range. Min-Max and Zscore are the popular methods used for data normalization. Min-Max normalizes the values of features into a linear scale bounded by a given range and it is suitable for uniformly distributed features. Z-score normalizes the data points from Gaussian distribution to Normal distribution so that most of the values of a feature are close to 0.

7. Machine Learning based IDS Framework

The general framework of machine learning based IDS is given in Figure. 9. The collection of labeled data is partitioned into training, validation, and testing data. As shown in the figure, data is initially sent to the pre-processing module where appropriate feature engineering is done to prepare the data for building classification models. The classifier is learned using the pre-processed training data, and the classification model is refined by varying the optimal model parameters which are fixed by validating the model using the validation data. Finally, the generalization accuracy of the classifier is assessed using the test data, and if it is acceptable, the classifier will be used to make predictions on unknown data.

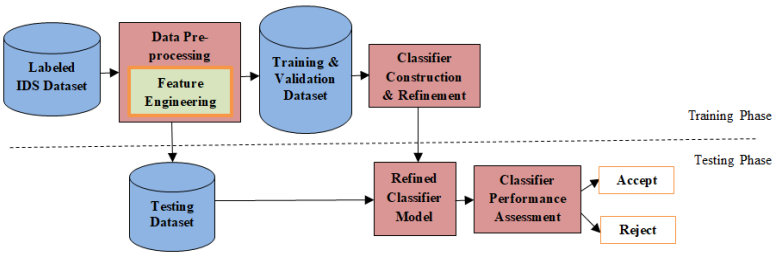


Figure 9: Stand-alone ML approach to IDS

In essence, the classifier learns attack patterns/signatures in the training phase using known labeled data and uses them for making predictions on the incoming network traffic data with an implicit assumption that both known labeled data and the new incoming network traffic data possess the same characteristics in terms of feature spaces and data distributions.

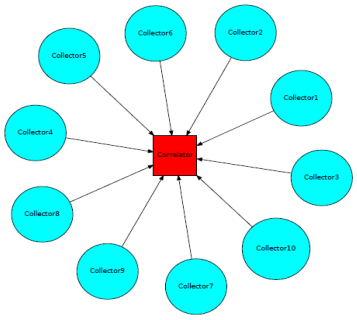
Along with this stand-alone approach of using a single classifier for implementing intrusion detection systems, there exist two other approaches for implementing intrusion detection systems; one is the hybrid learning approach and the other is the ensemble approach. The hybrid learning approach makes use of a combination of ML (clustering/classification) algorithms for performing attack detection by processing input through multiple levels. The ensemble approach combines the advantages of several classification algorithms trained on different training samples to improve the detection accuracy. Both hybrid and ensemble approaches are used in attack detection to get a more accurate prediction. Some researchers use hybrid approaches whereas others prefer ensemble approaches for implementing intrusion detection systems. Every method performs well in its

way but the specific method is selected based on the specific problem to be solved by the researcher. The ultimate goal is to construct a model that efficiently classifies traffic packets between attack and benign classes in case of binary classification, between specific attack and benign classes in case of multi-class classification.

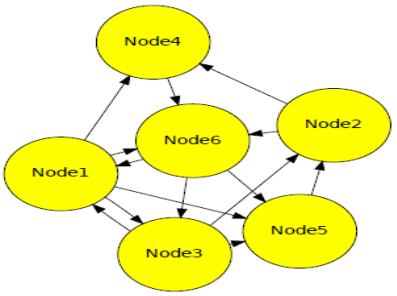
8. Intrusion Detection Networks

Intrusion detection networks (IDNs) [are the networks connecting multiple IDSs. Specifically, different collaborative IDSs formed as a network for sharing information such as intrusion alerts, signatures, malicious files, etc. constitute an Intrusion Detection Network wherein each IDS acts as a node in the network. Intrusion detection networks allow each IDS to share attack knowledge with other IDS such that combining the information from multiple detectors might help detect new attacks thereby increasing the detection accuracy while handling unknown attacks. Based on the co-operation topology, IDN follows two architectures; one is centralized architecture and the other is distributed architecture. IDN architectures. In the centralized architecture, there exist central IDS that gather, maintain, and disseminate the latest attack information to make more accurate judgments about attack identification. However, the centralized IDN architecture suffers from heavy traffic clog near the central IDS server and also exhibits reduced reliability due to a single point of failure based on the central IDS.

Examples of centralized IDN are Shield, CRIM, etc. represents the decentralized IDN architecture. In the decentralized architecture, IDSs at all nodes participate equally in knowledge sharing and analysis. As each IDS participates equally there is no single point of failure but, due to the lack of complete data at each IDS, detection accuracy may be compromised to some extent. Examples of de-centralized IDN are Indra, HBCIDS, etShield, etc.



(a) Centralized IDN



(b) Distributed IDN

9. Applications of Transfer Learning

Transfer learning can be successfully applied to various disciplines. Specifically, in the computer science discipline transfer learning is popular in the areas of Computer Vision (CV) and Natural Language Processing. Popular transfer learning models in CV include Google's inception, Oxford VGG, Microsoft ResNet. Similarly, popular applications of transfer learning in NLP include sentiment analysis, malicious email identification, multi-lingual text classification, identification of topic words in reviews, etc. The details of other disciplines where transfer learning is successfully applied are given. there was comparatively limited research work reported on the transfer learning approaches towards security and network intrusion detection problems.

10. Conclusion

Its work explores the applicability of recent advancements of machine learning towards the development of Intrusion Detection Systems to achieve high detection accuracy and low false-positive rate. In the context of network intrusion detection, different attack scenarios exist leading to their categorization into three groups; known attacks, new attacks, and zero-day attacks. Signature-based attack detection is preferred whenever it is possible to extract attack signatures and maintain them to match with the incoming traffic packets in the IDS.

Reference

1. *Ariafar, Elham, and RasoulKiani. "Intrusion detection system using an optimized framework based on datamining techniques." 2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI). IEEE, 2017.*
2. *Belavagi, Manjula C., and BalachandraMuniyal. "Performance evaluation of supervised machine learning algorithms for intrusion detection." Procedia Computer Science 89.2016 (2016): 117-123.*
3. *Chiba, Zouhair, et al. "A cooperative and hybrid network intrusion detection framework in cloud computing based on snort and optimized back propagation neural network." Procedia Computer Science 83 (2016): 1200- 1206.*
4. *Fung, Carol, and Raouf Boutaba, "Intrusion Detection", Intrusion detection networks: a key to collaborative security. Auerbach Publications, (2017): 21-37.*
5. *Hachmi, Fatma, KhadoujaBoujenfa, and Mohamed Limam. "Enhancing the Accuracy of Intrusion Detection Systems by Reducing the Rates of False Positives and False Negatives Through Multi-Objective Optimization." Journal of Network and Systems Management 27.1 (2019): 93-120.*
6. *Pan, SinnoJialin, and Qiang Yang. "A survey on transfer learning." IEEE Transactions on knowledge and data engineering 22.10 (2009): 1345-1359.*
7. *Saleh, Ahmed I., Fatma M. Talaat, and Labib M. Labib. "A hybrid intrusion detection system (HIDS) based on prioritized k-nearest neighbors and optimized SVM classifiers." Artificial Intelligence Review 51.3 (2019): 403-443.*