EasyChair Preprint

№ 2427

Intrusion Detection System and vulnerability identification using various Machine learning Algorithms

Gauri Rasane and Sunil Rathod

January 20, 2020

Intrusion Detection System and vulnerability identification using various Machine learning Algorithm

Gauri Vilas Rasane

P.G. Student, Department of Computer Engineering,
Dr. D. Y. Patil School of Engg, Pune, India
Email: gauri.rasane5@gmail.com

Dr. Sunil Rathod

Assistant Professor, Department of Computer Engineering,
Dr. D. Y. Patil School of Engg, Pune, India
Email: sunil.rathod@dypic.in

Abstract: Network security is very essential in today's environment in data security, cloud security as well as all the resources security which is shared in network environment. Basically IDS is the such kind of program which takes unauthorized access of vulnerable resources. It has categorized into Network base IDS and Host base IDS. Intrusions and abuse are constantly threatening to comprehensive internet service use. Therefore, the system for intrusion detection is the most important component of the machine and its network security. Intrusion Detection System (IDS) is an algorithm-focused computer network surveillance system that detects the presence of malevolent interference in the network. The IDS system has been recognized for maintaining high standards of safety, meaning that information is exchanged with confidence and security amongst dissimilar organizations. Systems for intrusion detection divide user activity into two main categories: regular, and distrustful. This paper system proposed an approach with machine learning algorithms for GA-FLN base IDS program. Several intrusion detection opportunities have been suggested before, but none shows acceptable results so systems are investigating for a better outcome in this region. The research suggested even takes a description of different kinds of structure techniques for Intrusion Detection System. System additionally research in these extraordinary methodologies, their exactness and also false positive proportions.

Keyword: Intrusion Detection system, soft computing, classification techniques.

1 INTRODUCTION

The massive rise in the use of computer systems in today's general public environment, increasing e-business every day, in such caseit is mandatory to provide the highest security to such instances which stored sensitive information on Virtual Machine (VM). Since it is not actually feasible to produce a program without any flaws, an integral field of research has been recognized is nothing but a intruder. Safety through darkness, network-based security and host-based security are the three most commonly used safety advances. Host-based security is hard to implement, taking into account the reality that it is often used by the majority. The hard part is to implement it on each and every mechanism of the entity. While this is quite convenient in a small network of the same machines but as the network expands and becomes heterogeneous, from an administrative point of view it becomes a big headache.

Various IDS systems [5] [6] violate integrity, confidentiality and availability of a system resource provided by the program. In the software it is not possible to respond about data being stolen or incomplete. Consequently, intrusion detection systems (IDSs) are required to reduce the serious impact of such assaults. The intrusion detection program is described as the device or software method for detecting unauthorized access to a network or a computer

system. IDS can detect all types of attacks, such as disruptive, dangerous attacks, weaknesses, data-driven attacks, host attacks, for example, authority breaches, confidential file access. System need IDS once system have a firewall, since firewall networks were not designed to detect network and application layer threats like worms, malware, Denial of Services (DoS), Distributed Denial of Services (DDoS) and Trojans. The firewall acts to prohibit the intrusion of external traffic into internal traffic network.

2 LITERATURE SURVEY

This section focused on study of various existing systems, and the work has done by various existing research in IDS.

2.1 RELATED WORK

IDS have categorized into two types called Network and Host IDS, which defines as special characterization and execution into different environments. It has also done the work with signature base and rule base anomaly detection [3]. Various researches have done the work on network as well as host based intrusion detection systems. Data mining algorithms and machine learning algorithms introduces accepted level accuracy on synthetic data set like KDDCUP99 [5] [6]. But still the some issues have left to detect the network environment malicious activities.KDDCUP99 consist 41 attributes with 23 sub attacks which basically holds master attacks like DOS, PROBE, U2R and R2L respectively. The another data sheet is introduced NSLKDD in 2010 it is quite similar like KDDCUP99 but it contains 38 attacks which provides drastic supervision and attack detection in both environments.

2.3 EXISTING METHODOLOGY

ParisaAlaeiet. Al.(2018) [1] , in this paper ,the approach implies a technique to solve malicious attack detection problem by reviewing data sets online. This is done through the use of an incrementally naïve Bayesian classifier. In comparison, active learning allows the problem to be answered through the use of a small set of defined data points which are often quite expensive to acquire. The proposed technique comprises of two acting classes, i.e., offline and online. The former includes results pre-processing, while the latter uses the NADAL electronics. The proposed solution is similar to the formal, naïve classification using Bayesian the NSL-KDD standard dataset.

Alsughayyir, Bayan et al.(2019)[2], according to this paper, a deep learning (DL) is used which can create a better and more efficient architecture for intrusion detection (ID). The purpose approach focuses on classifying normal behavior from anomalous network activity. The Intrusion Detection System (IDS) is one of the tools used to detect unauthorized network or system operation and protect the system from attacks on the network. Attacks are observed in the device by discriminating between common and irregular network behavior and functionality. This work also defines numerous methods to produce IDS which has they utilized in experiment analysis.

Borkar, Amol et Al.(2017),According to [3] Consists of the literature review of the Inner Intrusion Detection System (IIDS) and the Intrusion Detection System (IDS), which uses various data collection techniques and testing approaches for the system to function in real time. Data mining techniques are being developed for cyber analytics to allow intrusion prevention. Techniques used before like firewall, and IDS failed to detect the real-time attackers that happened in the absence of the manager, without his knowledge. Recognizing the attacker in real time is challenging, because it can produce duplicate IP and attack packets. A computer network is a Software and Hardware mix. Each part carries risks, Poor security, and deficiencies. The assault against the Ransomware leaves data vulnerable. Those who learn programming and programs can find

out about the various activities conducted on the systems quickly from the log files. We'll help to ensure security.

Bhosale, Karuna S. et al.(2018),According to [4] deep neural network (DNN), It is studied as a type of deep learning framework to build scalable and powerful IDSs to detect and recognize accidental and unexpected cyber attacks. The continuous change in network activity and the rapid development of attacks make it necessary to examine multiple databases that are created over the years via static and dynamic approaches. Such type of study makes it easier to identify the right algorithm which can work effectively to predict potential cyber-attacks. On numerous publicly available test malware databases, a detailed assessment of DNN studies and other classical machine learning classifiers is shown. The Optimum network parameters and topologies for DNNs are chosen using hyper parameter filtering methods KDDCup 99 dataset.

Chamou et al. (2019)[5],A large number of businesses around the world are being targeted and threatened by the daily emergence of new and evolving threats. It is for this reason that the scientific community has drawn attention to the nature and improvement of the performance of Intrusion Detection Systems. It is an innovative method to track malicious activity using deep learning techniques in terms of DDoS and ransomware cyber-threats. Cyber security accomplishment, data protection and secure communication are considered essential owing to the rapid growth of Web apps and their use by most Internet users.. At the same time, increased exposure to more sophisticated cyber-threats has been observed over the Internet and computer networks, in the academic and industry digital world, especially in Small-Medium Enterprises (SME), with financial costs.

Christos, et al.(2019), according to [6] Self-using Novel Network Intrusion Prevention Program Organizing with Support an enhanced neural Vector system network. The proposed system, because of its design, does not provide a security solution that is neither signed nor dependent on rules, and is capable of mitigating known and unknown risks with high accuracy. Based on our experimental results The suggested design can use the NSL KDD dataset to access specialized online education, so it fits for efficient and scalable industrial applications. Machine-based NIDS / NIPS programming partially Signature-and address the above bugs Suffering from Rule-based Industrial NIDS / NIPS. Not to be lost .

Liang, Wei, et al. (2019), according to [7]For multifunctional performance, effectively increase the detection rate and the real-time output of detecting abnormal behavior in industrial networks. The novel apps are dual in order to quickly select a node with a high security coefficient as the heart of the cluster and align the multifeature data around the center in a cluster. Experimental results indicate that the suggested algorithm is of strong quality in terms of the detection rate and time relative to other algorithms. In the networking sector, the sensitivity of identification of suspicious data exceeds 97.8%, and the incorrect detection result falls by 8.8%. Intrusion detection systems can successfully identify and monitor intruder incidents, although difficult with network security technology. So, the use of intrusion detection systems in industrial networks will overcome the limitations of conventional network security techniques, so perfecting the entire industrial safety system networks.

Loganathan, Gobinath et al.(2018), according to [8] Present a new multi-attribute method to estimate a network packet sequence based on past packets using the Sequence-to-Sequence (Seq2Seq) encoder-decoder algorithm. This model is used in an attack-free dataset to know the regular sequence of packets in TCP communications, and is then used to identify anomalous packets in TCP traffic. They show that in the DARPA 1999 dataset, the experimental multi-attribute model Seq2Seq detects anomalous raw TCP packets that are

part of 97 per cent intrusions into accuracy. It can also recognise selected intrusions with 100% real-time precision and outperform current algorithms that rely on repeated neural network models, like LSTM. The Detecting irregularities in raw TCP packets through a Seq2Seq algorithm specifically designed for sequences of different attributes. To train the model system use packets in links segregated from regular network traffic. To science, anomalies are known as actual packets which deviate significantly from the planned packets. Training the model on normal traffic instead of intrusion traffic gives access to extensive training data and allows the model to detect even new unknown threats that deviate from regular traffic pattern.

Mayank Agarwal et al.(2017),According to [9], This paper describes a system for intrusion detection of PS-Poll DOS infiltration in 802.11 networks, using a distinct case structure in real time. This methodology utilizes RTDES to track DOS attack on a single Event System in real-time. High detection rate and accuracy rate are one of the significant advantages, but shortage of frames is one of the major drawbacks. This system also able to detect software as well as hardware attacks simalteniously.

SaeidSoheilyKhah et al. (2018)[10],thisSystem, network intrusion detection(ID) is tackled by unattended and unattended hybrid mining-a comprehensive case study on the ISCX dataset. This proposes a hybrid intrusion detection (kM-RF) which generally outperforms an alternative technique in terms of false alarm rate, precision, and detection time. ISCX(A standard intrusion detection dataset) is used to assess the efficacy of kM-RF, and an in-depth analysis is performed to check the effects of any pre-processing characteristics or characteristics detected. It also uses a special pre-treatment method for categorical transformationTools or attributes for numerical data and to create more segregated groups from raw data. Some new features or applications to find payloads, clustered attacks and IP scans and a mix of k-means and random forest classifiers to prevent further interference effectively.

3 PROPOSED SYSTEM DETAILS

3.1 PROBLEM STATEMENT

In this research work, system aim to design and develop an approach for Intrusion Detection for fast learning base neural network as well as machine learning approach to evaluate the proposed system evaluation on different network dataset that will produce the classification accuracy of system.

3.2 OBJECTIVES

The main objectives of this project are itemized as follows:

- To design and develop an approach for IDS using various machine learning algorithm for network dataset.
- To generate strong and dynamic rules depending upon the real time behavior of the packet in training phase.
- The Intrusion Prevention System prevents the all Anomaly detection and misuse detection into NIDS and HIDS environment respectively.
- To explore and validate the proposed system with multiple network dataset as well as compare with existing approaches.

3.3 SYSTEM ARCHITECTURE

The goal of proposed anomaly network intrusion detection system is to maximize the detection accuracy, to minimize false positive rate and detector generation time. Basically there are two phase in the proposed system, system have taken NSLKDD dataset for system training as well testing purpose.

The proposed System worked with an ensemble configuration. When two or more combinations form a new model commonly called an ensemble model. This ensemble model incorporates input from multiple classifiers and has produced a single composite classification. Our conceptual structure consists of numbers for the classifiers. First, the software receives data from various outlets, both online and offline. Once the data is collected by software, other data mining strategies will be applied in different classifications approaches.

Training Phase:

1. Upload training data for feature extraction.

2. Apply PSO for rule creation
3. Create rules set as normal pool as well as intrusion pools set.

Testing Phase:

1. Upload Testing data or any packet which is collected from network environment.
2. Extract all features using attribute selection.
3. Apply Normalization approach on dataset.
4. Apply ensemble approach on all train as test features.
5. Show results with classification accuracy.
6. Classify all attacks.
7. Show detection results.

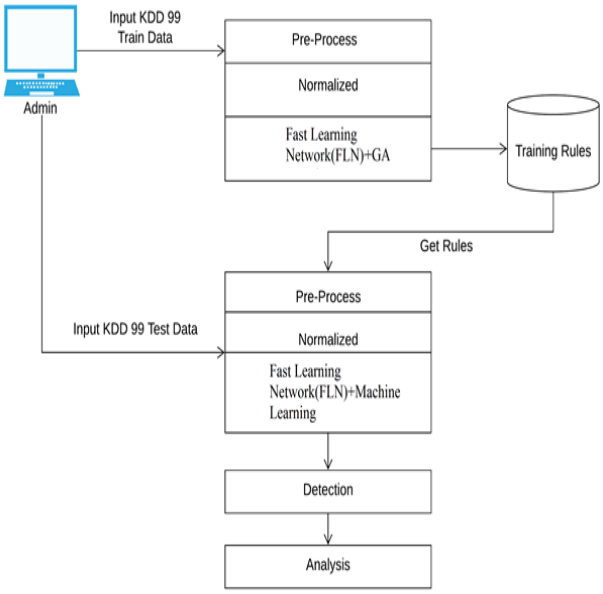

Figure 1: Proposed System architecture

System initially collects the input packet from various sources like KDD CUP, NSL KDD, ISCX and real time network packets. The entireexecution holds three different phases which are listed as below.

Module 1: Intrusion Detection System (IDS) this phase executes the first Genetic Algorithm and fuzzy algorithm for feature extraction and for creating background rules.
Module 2: Intrusion Prevention System (IPS) This work for prevention of the known attacks which is already generated by remote sources. Some classification algorithm are used for prevention of the system. Naïve Bayes, ANN, J48 weight calculation algorithms work for finding same network flow as well as packet signature.
Module 3: Intrusion Response System (IRS)It works for provide the security from different types of unknown attacks. The system holds the ensemble modules for detecting malicious activity.

3.4 MATHEMATICAL MODEL

System = { Train, Test, Analysis }

Train = { GA, Fuzzy, ANN, J48, NB }

GA = { Crossover -> Mutate -> Fitness -> Selection }

Fuzzy = { Probability, {0,1}}

{ GA -> Fuzzy -> ARM -> } {0,1}

Test = { Pattern Match, Th, Weight, Subclass }

Class = { Input ->BkRule ->Weight } -> { Normal, Attack } -> { subattacks }

Analysis = { dos,probe,U2R,R2L,Normal,unknown }

3.5 ALGORITHM DESIGN

Naïve Bayes Algorithm

Input: Feature of BK rules TrainDataF[], features if test record TestDataF[]

Output: highest Similarity score for class label

Step 1: Read all training rules from DB for each (Record R into TrainData[])!=Null

Step 2: training_items [] split(R)

Step 3: test_items1 [] split(TestF)

Step 4: score=Calculate_Score(training_items[i], test_items1[i])

Step 5: Return w;

ANN(Artificial Neural Network)

Step 1: for all (T in HidenLayer [] !=null) do

Step 2: training_items[] split(T)

Step 3: test_items1[] split(InputNueron)

Step4: w=Calculate_Score(HidenLayer[i], InputNueron)

Step 5: Return w;

J48 Weigh calculation

Input: Feature of BK rules Train_DatasetF [], features if test record Test_DatesetF[]

Output: highest Similarity weight for class label

Step 1: for all (T in TrainF [] !=null) do

Step 2: TrainData [] split(T)

Step 3: TestF [] split(TestF)

Step 4: w = classifyToAll (TrainData [], TestF[], Label)

Step 5: Return w;

4 RESULTS AND DISCUSSION

The proposed research basically focuses on the soft computing approach and classification-based detection, basically both methods having the good detection rate but generating sometimes more false positive ratio. In real-time environments, some systems are also not applicable, and some may not focus on misclassified anomalies. As noted, the mark is still missing in most applications because there is no program that currently provides a discovery rate of 100% and the sky is the limit

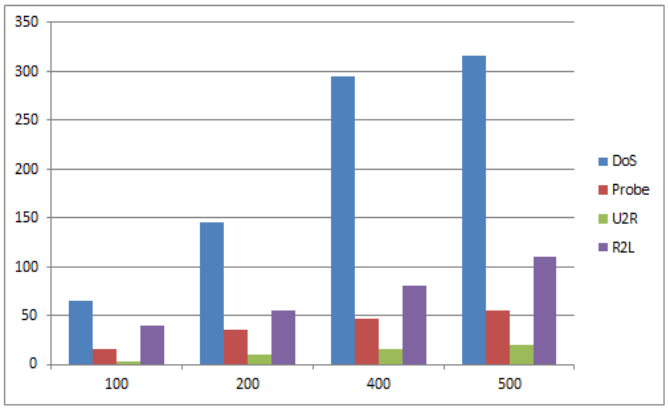

Figure 2 : Overall attack found ratio in ensemble approach

5 CONCLUSIONS

System suggested ensemble method for network traffic anomaly detection in this research work. Our approach focused on building the model of anomaly detection normal traffic profile. System also showed through experiments that some features of the NSL-KDD and ISCX dataset with the normal profile are efficient. With input training data, system propose a K-means clustering algorithm to reduce noise. The experiments showed that our Approach works good, even with a small training sample, including precise recognition. They often call for a new architecture to merge anomaly detection system with signature-dependent detection system, along with some

improvements to the usual performance profile of buildings. In our future plan, system will use an open source IDS to build and play with the proposed model in actual network. Several different ideas have emerged to confront this problem since about ten years ago the intrusion detection concept began to gain momentum in the security community. Detection systems for intrusion vary in the approaches used to collect the data and in the specific techniques used to analyze these data.

6 FUTURE WORKS

To evaluate the system with some combination of network and synthetic dataset and generate the dynamic rules for strongly unknown attack detection in vulnerable environment.

REFERENCES

[1] Alaei P, Noorbehbahani F. Incremental anomaly-based intrusion detection system using limited labeled data. In Web Research (ICWR), 2017 3th International Conference on 2017 Apr 19 (pp. 178-184). IEEE.

[2] Alsughayyir, Bayan, Ali Mustafa Qamar, and Rehanullah Khan. "Developing a Network Attack Detection System Using Deep Learning." 2019 International Conference on Computer and Information Sciences (ICCIS). IEEE, 2019.

[3] Borkar, Amol, AkshayDonode, and Anjali Kumari. "A survey on Intrusion Detection System (IDS) and Internal Intrusion Detection and protection system (IIDPS)." 2017 International Conference on Inventive Computing and Informatics (ICICI). IEEE, 2017.

[4] Bhosale, Karuna S., Maria Nenova, and GeorgiIliev. "Modified Naive Bayes Intrusion Detection System (MNBIDS)." 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS). IEEE, 2018.

[5] Chamou, Dimitra, et al. "Intrusion Detection System Based on Network Traffic Using Deep Neural Networks." 2019 IEEE 24th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD). IEEE, 2019.

[6] Constantinides, Christos, et al. "A novel online incremental learning intrusion prevention system." 2019 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS). IEEE, 2019.

[7] Liang, Wei, et al. "An Industrial Network Intrusion Detection Algorithm based on Multi-Feature Data Clustering Optimization Model." IEEE Transactions on Industrial Informatics (2019).

[8] Loganathan, Gobinath, JagathSamarabandu, and Xianbin Wang. "Sequence to sequence pattern learning algorithm for real-time anomaly detection in network traffic." 2018 IEEE Canadian Conference on Electrical & Computer Engineering (CCECE). IEEE, 2018.

[9] Mayank Agarwal, SankethPurwar, Santosh Biswas, Sukumar Nandi, "Internal Detection System for PS-Poll DOS attack in 802.11 networks using real-time discrete event system",IEEE,vol.4,issue4,2017.

[10]Sedjelmaci H, Senouci SM, Ansari N. A hierarchical detection and response system to enhance security against lethal cyber-attacks in UAV networks. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2018 Sep;48(9):1594-606.