ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 6, June 2019

Light Weight Attack Defence System using Machine Learning Ensemble Approach

Miss. Pooja Tambat[1], Prof. Brijendra Gupta[2]

P.G. Student, Department of Computer Engineering, SCOE, Pune, Maharashtra, India[1]

Assistant Professor, Department of Computer Engineering, SCOE, Pune, Maharashtra, India[2]

ABSTRACT: In n/w IDS (intrusion detection framework) ask around, one standard procedure for finding assaults is watching a n/w's activity for anomalies: deviations from profiles of commonness as of late picked up from agreeable traffic, consistently recognized using gadgets acquired from the AI organize. Information security is an issue of certifiable overall concern. The multifaceted nature, accessibility, and responsiveness of the Internet have served to assemble the security risk of information frameworks enormously. This paper concerns intrusion detection. We delineate approaches to manage intrusion detection using NN and Machine Learning Ensemble approach. The key contemplations are to discover profitable models or features that delineate customer direct on a framework, and use the course of action of relevant features to build classifiers that can see anomalies and known intrusions, in a perfect world dynamically. Using a ton of benchmark information from a KDD competition organized by DARPA, we demonstrate that eficient and accurate classifiers can be attempted to recognize intrusions. We dissect the execution of neural frameworks based, and AI gathering, frameworks for intrusion detection.

KEYWORDS:IDS, GA, Ensemble Learning, Machine Learning.

I. INTRODUCTION

Customarily, NIDS (n/w intrusion detection systems) are comprehensively grouped dependent on the style of detection they are utilizing: systems depending on misuse-detection monitor activity with exact depictions of known pernicious conduct, while anomaly detection systems have a thought of ordinary movement and banner deviations from that profile. The two methodologies have been widely examined by the exploration network for a long time. In any case, as far as genuine arrangements, we watch a striking lopsidedness: in operational settings, of these two primary classes we find solely just abuse indicators being used—most normally as signature systems that examine organize traffic for trademark byte groupings.

This paper censures intrusion detection, an imperative issue in cautious data. We present the utilization of neural n/ws and different Machine learning approaches called Ensemble Learning for intrusion detection of data frameworks. Since the vast majority of the intrusions can be situated by looking at examples of client exercises, numerous IDSs (Intrusion Detection Systems) have been worked by using the perceived attack and abuse designs. Utilizing neural n/w for intrusion detection has been done inside the security network [1,7,8,10,11]. In our tests, the neural n/w's and Machine Learning Ensemble are prepared with typical client movement and attack designs. The information we utilized started from MIT's Lincoln Labs. It was created for KDD rivalry by DARPA and is viewed as a standard benchmark for intrusion detection assessments. Our objective for intrusion detection is to distinguish the two irregularities and abuses. The methodology is to prepare the neural networks and Machine Learning Ensemble to become familiar with the typical conduct and attack designs; at that point noteworthy deviations from ordinary conduct are hailed as attacks. We start by giving fundamental definitions and terms in the following segment.

II. RELATED WORK

Marir, Naila, et al. proposes a novel distributed approach for the identification of abnormal conduct in large-scale n/w. The created model finds the abnormal conduct from large-scale n/w traffic information utilizing a blend of a multi-layer

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 6, June 2019

ensemble Support Vector Machines (SVM) and deep feature extraction distributedly. To start with, we play out a non-direct dimensionality reduction, accomplished through a distributed Deep Belief n/w on large-scale n/w traffic information. At that point, the got features are encouraged to the multi-layer ensemble SVM. The development of the ensemble is practiced through the iterative lessen worldview dependent on Spark. Exact outcomes demonstrate a promising addition in execution contrasted and other existing models [1].

Ali, Mohammed Hasan, et al. a created learning model for Fast Learning n/w (FLN) in view of PSO (particle swarm optimization) has been proposed and named as PSO-FLN. The model has been connected to the issue of interruption location and approved dependent on the well known dataset KDD99. Our created model has been analyzed against a wide scope of meta-heuristic calculations for preparing ELM, and FLN classifier. PSO-FLN has beated other learning approaches in the testing exactness of the learning [2].

Quiring, Erwin, Daniel Arp, and KonradRieck.Demonstrates a first exertion to unite these networks. To this end, we present a bound together documentation of blackbox attacks against machine learning and watermarking. To show its viability, we apply ideas from watermarking to machine learning and the other way around. We demonstrate that countermeasures from watermarking can relieve late model-extraction attacks and, comparatively, that systems for solidifying machine learning can battle off prophet attacks against watermarks. We further exhibit a novel danger for watermarking plans dependent on ongoing deep learning attacks from antagonistic learning. Our work gives an applied connection between two research fields and in this manner opens novel headings for enhancing the security of both, machine learning and computerized watermarking [3].

Feng, Ming, and HaoXu.proposes the online ideal cyber-defense issue has been explored for CPS (Cyber-Physical Systems) with obscure cyber-attacks. Right off the bat, a novel cyber state elements has been created that can assess the continuous effects from current cyber-attack and defense procedures viably and progressively. Next, receiving amusement hypothesis strategy, the thought ideal defense configuration can be gotten by utilizing the full information of cyber-state elements. To loosen up the prerequisite about cyberstate elements, an amusement hypothetical performing artist faultfinder NN (neural n/w) structure was produced to proficiently become familiar with the ideal cyber defense methodology online [4].

Borkar, Amol, AkshayDonode, and Anjali Kumari. Propose, far and wide, billions of individuals get to the web today. Intrusion detection innovation is another age of security innovation that screen system to maintain a strategic distance from malevolent exercises. The paper comprises of the writing study of IIDS (Internal Intrusion Detection System) and IDS (Intrusion Detection System) that utilizes different information mining and legal methods calculations for the system to work progressively. Information digging techniques are proposed for cyber analytics in help of intrusion detection [5].

III. PROPOSED ALGORITHM

A. *System Architecture:*

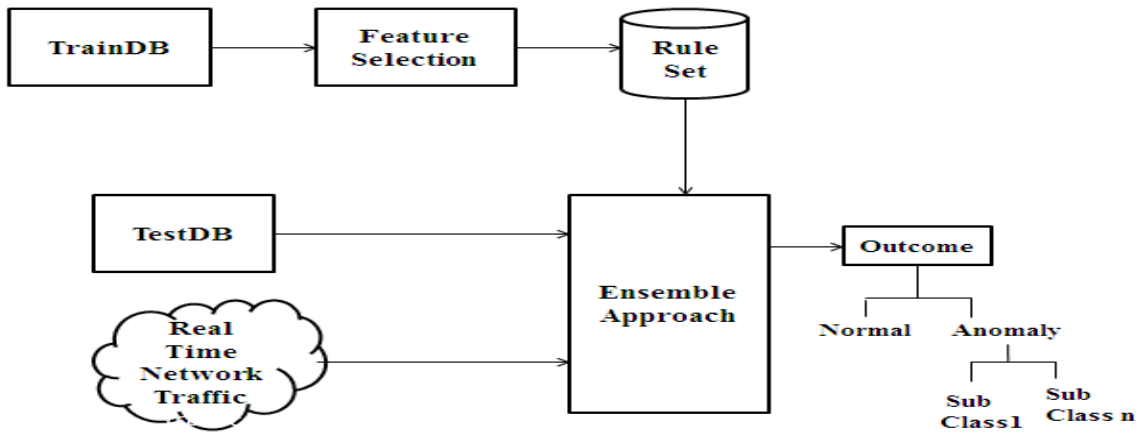

Fig. 1. Proposed System Architecture

Intrusion detection (ID) is a type of security management system for computers and networks. An ID system gathers and analyzes information from various areas within a computer or a network to identify possible security breaches, which include both intrusions (attacks from outside the organization) and misuse (attacks from within the organization).

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 6, June 2019

ID uses vulnerability assessment (sometimes referred to as scanning), which is a technology developed to assess the security of a computer system or network.

The four main forms of attacks are
1. DoS (Denial of Service): A DoS attack is a type of attack in which the hacker makes a computing or memory resources too busy or too full to serve legitimate networking requests and hence denying users access to a machine e.g. apache, smurf, neptune, ping of death, back, mail bomb, UDP storm etc. are all DoS attacks.
2. U2R (User to Root): These attacks are exploitations in which the hacker starts off on the system with a normal user account and attempts to abuse vulnerabilities in the system in order to gain super user privileges e.g. perl, xterm.
3. R2L (Root to Local): These attacks are exploitations in which the hacker starts off on the system with a normal user account and attempts to abuse vulnerabilities in the system in order to gain super user privileges e.g. perl, xterm.
4. Probe: Probing is an attack in which the hacker scans a machine or a networking device in order to determine weaknesses or vulnerabilities that may later be exploited so as to compromise the system.

B. *Description of the Proposed Algorithm:*

The proposed system worked with ensemble approach, system first collect data from different online as well as offline sources. Once data has collected by system it will apply some data mining strategies with different classification approaches. The above figure shows the proposed system architecture and execution. Basically there are two phase in the proposed system, we have taken NSL KDD dataset for system training as well testing purpose.

1) Data Preprocessing:Data preprocessing done by Weka tool. This is offline method. Data preprocessing includes following three main task:
    a) Converting non-numerical features of NSL-KDD dataset into numerical values.
    b) At the end transferring attack types into numerical values.
    c) Finally preparing proper dataset.
2) Feature Selection and Rule Creation:Normalization done in this phase. Min-Max normalization used to normalizing the features. Information Gain (IG) is use to reduce the features. Information Gain (IG) is applies on Feature selection phase. Information Gain is nothing but attribute selection mechanism in both training and testing dataset. Genetic algorithm is used for creating a rule set as a normal pool and intrusion pool. This rule set is used as a background knowledge when ensemble model is running in system.
3) Ensemble Model:Building the hybrid model consisting number of classifiers. Our system uses Naive Bayes, Artificial Neural Network and J48 algorithm for ensemble model. Mapping TestDB with Rule set and applies on ensemble method. Developing a model that exhibits the best performance and accuracy. Compare the accuracy of each classifiers and select best model.
4) Result Generation:Finally results are generated whether incoming packet is normal or anomaly. If it is anomaly then it also finds subclasses of that anomaly.

IV. PSEUDO CODE

Algorithm 1: Genetic Algorithm (GA) For Rules Creation
Input: Set of network packet which consist 41 attributes with class label
Output: Set of normal as well intrusion rules
Step 1: Initialize randomly population with 41 chromosomes.
Step 2: Initialize N (total number of training dataset records).
Step 3: For each chromosome creates a new population.
Step 4: Apply Crossover to best selected chromosome.
Step 5: Apply Mutation for each chromosome to generate new population.
Step 6: Calculate fitness= F(x) /sum (F(x)).
Step 7: Select best fit chromosome as 50 % and delete worse fit chromosome.
Step 8: End for

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 6, June 2019

Algorithm 2: Naive Bayes (NB)
Input: Feature of BK rules TrainF[], Features if test record TestF[]
Output: The highest similarity weight for class label
Step 1: Read all training rules from DB for each (Rec R into TrainF[])!=Null
Step 2: items [] <- split(R)
Step 3: items1 [] <- split(TestF)
Step 4: CalculateWeight(DB [i], items1)
Step 5: Return w;

Algorithm 3: Artificial Neural Network (ANN)
Step 1: for all (T in HidenLayer [] !=null) do
Step 2: items[] <- split(T)
Step 3: items1[] <- split(InputNueron)
Step4: w=CalculateWeight(HidenLayer[i], InputNueron)
Step 5: Return w;

Algorithm 4: J48
Input: Feature of BK rules TrainF[], features if test record TestF[]
Output: The highest similarity weight for class label
Step 1: for all (T in TrainF []!=null) do
Step 2: items [] <- split(T)
Step 3: items1 [] <- split(TestF)
Step 4: w = classifyToAll(Train,TestF[], Label)
Step 5: Return w;

V. SIMULATION RESULTS

For this experiment we have done different population base experiments eg (pop size 100,500,1000, 2000,3000,4000 and 5000) the each accuracy graph show the detection accuracy of each phase. In the first investigation, we utilized our fluffy hereditary calculation to group typical system information and assault. At that point, we indicate identification rate acquired for KDD99 dataset. I characterize them into two classes which are ordinary and assault. In the second analysis, we utilized the fluffy hereditary calculation to arrange sorts of assaults in the online continuous sniffer dataset. Framework reflects about both methodology results attractive level, underneath figures, examination table and result charts demonstrate the general framework stream and discovery productivity of framework.

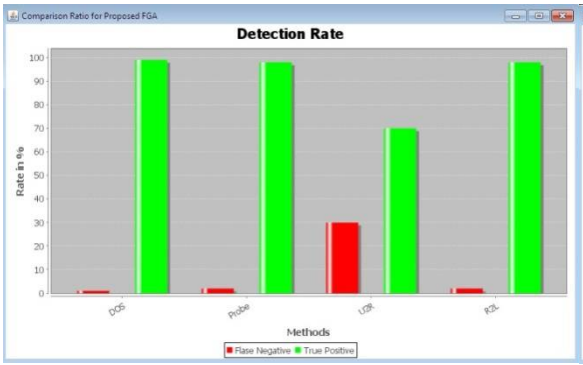

Fig.2.Accuracy when population size is 100

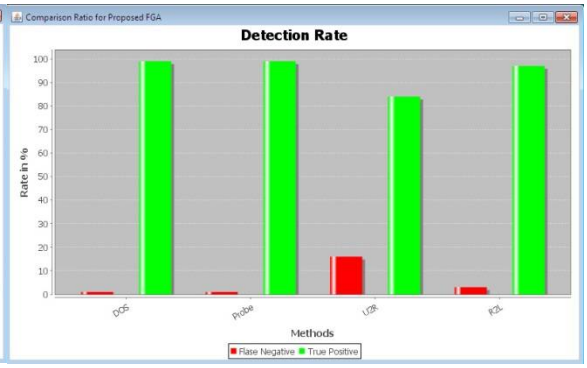

Fig.3. Accuracy when population size is 500

The below table shows the proposed system accuracy for each attack type with false ratio as well as accuracy.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 6, June 2019

Net Size	DOS		Probe		U2R		R2L	
	TP	FN	TP	FN	TP	FN	TP	FN
100	99%	1%	98%	2%	70%	30%	98%	2%
500	99%	1%	99%	2%	84%	16%	97%	3%
1000	99%	1%	99%	1%	87%	13%	98%	2%
2000	99%	1%	98%	2%	71%	29%	99%	1%
3000	99%	1%	98%	2%	76%	24%	97%	3%
4000	99%	1%	98%	2%	77%	23%	99%	1%
5000	99%	1%	98%	2%	67%	33%	98%	2%

Table.1.Accuracy of each attack type

VI. CONCLUSION AND FUTURE WORK

We have built IDS (intrusion detection systems) utilizing neural n/w and Machine learning ensemble, and tried their execution on a lot of benchmark DARPA information. It is seen that both the neural n/w and Machine Learning Ensemble approach convey profoundly exact outcomes (more prominent than 98% precision on testing set) and show perfect dimension of execution. Statistical learning techniques are being utilized all the more widely in ongoing intrusion detection systems, inferable from their flexibility and their speculation capacity with respect to new assault marks that would should be 'adapted' rapidly? When found? by an IDS. Then again, neural systems have effectively turned out to be valuable in numerous IDSs, and are particularly suited for multi-classification orders.

For future move up to recognize DDOS attacks likewise different attacks other than DOS on the controller of the Software portrayed n/w. There are diverse methods to perceive attacks and trying to distinguish strike subject to Entropy. By executing entropy as acknowledgment technique, will prepared to perceive attacks on one host or a subnet of hosts in a n/w.

REFERENCES

1. Marir, Naila, et al. "Distributed Abnormal Behavior Detection Approach Based on Deep Belief Network and Ensemble SVM Using Spark." IEEE Access 6 (2018): 59657-59671.
2. Ali, Mohammed Hasan, et al. "A new intrusion detection system based on Fast Learning Network and Particle swarm optimization." IEEE Access 6 (2018): 20255-20261.
3. Quiring, Erwin, Daniel Arp, and KonradRieck. "Forgotten Siblings: Unifying Attacks on Machine Learning and Digital Watermarking." IEEE European Symposium on Security and Privacy. 2018.
4. Feng, Ming, and HaoXu. "Deep reinforecement learning based optimal defense for cyber-physical system in presence of unknown cyber-attack." Computational Intelligence (SSCI), 2017 IEEE Symposium Series on. IEEE, 2017.
5. Borkar, Amol, AkshayDonode, and Anjali Kumari. "A survey on Intrusion Detection System (IDS) and Internal Intrusion Detection and protection system (IIDPS)." Inventive Computing and Informatics (ICICI), International Conference on. IEEE, 2017.
6. Liao, Hung-Jen, et al. "Intrusion detection system: A comprehensive review." Journal of Network and Computer Applications 36.1 (2013): 16-24.
7. Pamukov, Marin E., and Vladimir K. Poulkov. "Multiple negative selection algorithm: Improving detection error rates in IoT intrusion detection systems." Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), 2017 9th IEEE International Conference on. Vol. 1. IEEE, 2017.
8. Khannous, Anass, et al. "Manet security: An intrusion detection system based on the combination of negative selection and danger theory concepts." Next Generation Networks and Services (NGNS), 2014 Fifth International Conference on. IEEE, 2014.
9. Kumar, Manoj, and Robin Mathur. "Unsupervised outlier detection technique for intrusion detection in cloud computing." Convergence of Technology (I2CT), 2014 International Conference for. IEEE, 2014.
10. Badgujar, Tejaswini, and Priyanka More. "An Intrusion Detection System implementing Host based attacks using Layered Framework." Innovations in Information, Embedded and Communication Systems (ICIIECS), 2015 International Conference on. IEEE, 2015.
11. BahmanRashidi, Carol Fung, Elisa Bertino,"A Collaborative DDoS Defence Framework using Network Function Virtualization", IEEE Transactions on Information Forensics and Security,2017.
12. Lara, A. Kolasani, and B. Ramamurthy, "Network innovation using OpenFlow: A survey," IEEE Commun. Surveys Tuts., vol. 16, no. 1, pp. 493–512, Feb. 2014.
13. D. Huang, L. Xu, C. Chung T. Xing, "SnortFlow: A openflow-based Intrusion Prevention System in Cloud Environment," Second GENI Research nad Educational Experiment Workshop , pp. 89-92, 2013.
14. Lara and B. Ramamurthy, "OpenSec: A framework for implementing security policies using OpenFlow," in Proc. IEEE Globecom Conf., Austin, TX, USA, Dec. 2014, pp. 781–786.