A Survey on Effective Machine Learning Algorithm for Intrusion Detection System

Anita Verma
M.Tech Schollers CSE SIRTS, Bhopal

Dr. Aumreesh Kumar Saxena
CSE Dept SIRT, Bhopal

M. Arsad
CSE Dept SIRT, Bhopal

Abstract— Computer networks security plays an important role in modern computer systems. In order to enforce high protection levels against threats, a number of software tools are currently developed. Intrusion Detection Systems (IDS) aims to detect intruder or anomaly in the computer networks. Software model protects a computer networks from unauthorized users through detecting intruders in the network. In this we build a machine learning classifier and trained the model on the NSL-KDD dataset, after training the model are able to detect or classify the attacks in to category like normal or attack. Recently there is already work done by data mining techniques to accurately detect the malicious activities. So to further improve the accuracy of this intrusion detection system we proposed a deep learning machine.

Keywords— Network security, Intrusion Detection system, data mining algorithm, machine learning techniques, Anomaly detection, SVM, Ensemble Learning.

I. INTRODUCTION

With the advancement in the technology, millions of people are now connected with each other through one or other form of network where they share lots of important data. Hence the need of security to safeguard data integrity and confidentiality is increased rapidly. Although effort have been made to secure data transmission but at the same time, attack technique for breaching the network continued to evolve. Thus it leads to the need of such a system which can adapt with these ever changing attack techniques. Attacks will be varied during a long vary like Brute Force Attack, Heartbleed Attack, DoS Attack, DDoS Attack, net Attack etc. The information measure of the network is increasing apace because the variety of users of the web square measure increasing. There's a large variation of normal speed these days that is from 1Gbps to 10Gbps for a mean knowledge center. The transfer speed and transfer speed is completely different for large school. Firms like Google, Facebook etc., or huge company firms that are from forty Gbps to 100Gbps [1-2]. Network-based Intrusion Detection System may be a security tool that protects from an enclosed attack, outside attack and unauthorized access into the network [2].That is intended by package and/or hardware. The foremost acquainted idea is firewall that is made to shield the complete network from unauthorized access by information processing address and port variety and managing these activities by NIDS. it's intensive and wide-range operating applications which incorporates distinguishing the quantity of intrusion makes an attempt on the network for instance, denial of service attack hacking activities which can compromise the safety of any single pc or whole network by observation the traffic NIDS is mostly placed outside the firewall wherever the complete external traffic will be monitored by sensing and police investigation the anomaly activities [2]. Once during an advanced network, for instance, a tool connected to a thousand nodes, because of the quality of network, it's the most effective call to prefer AN NIDS to stay track of adjusting network atmosphere [2]. That brings to a conclusion as just one ID in any network will compromise of Confidential or Sensitive knowledge. It might build difficulties to method the massive quantity of traffic owing to just one entry point of a network turnout additional specifically after we use DPI (Deep packet Inspection) that works for matching the pattern against signature packet rules [2]. There square measure massive sorts of machine learning algorithms are wide wont to discover the Anomaly Detection NIDS. for instance, Artificial Neural Network (ANN), SVM (Support Vector Machine), Random Forest, Self Organized, Naive-Bayesian, and Deep learning. There has been a sequent development of Network Intrusion Detection System as classifiers to differentiate any anomaly from traditional traffic.

A. Intrusion Detection System

Intrusion Detection System or IDS is software, hardware or combination of both used to detect intruder activity. Snort is an open source IDS available to the general public. IDS may have different capabilities depending upon how complex and sophisticated the components are. IDS appliances that are a combination of hardware and software are available from many companies. As mentioned earlier, IDS may use signatures, anomaly-based techniques or both [3].

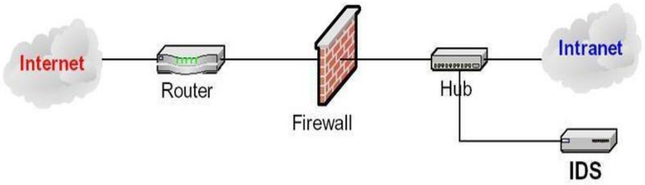

Figure 1.1 Intrusion detection System [1]

Signature is the pattern that you look for inside a data packet. A signature is used to detect one or multiple types of attacks. For example, the presence of “scripts/iisadmin” in a packet going to your web server may indicate an intruder activity. Signatures may be present in different parts of a data packet depending upon the nature of the attack [4]. For example, you can find signatures in the IP header, transport layer header (TCP or UDP header) and/or application layer header or payload.

B. Data Mining Algorithms for Intrusion Detection

The growth of data mining methods has consequently brought forth a wide range of algorithms drawn from areas as pattern recognition, machine learning and database analysis.

There are many types of algorithms that may be used to

Published By:
International Journal of Engineering Research in Current Trends

mine audit data. Data algorithms a set of heuristics designs between data mining models. These results of analysis are later used by the algorithm for defining optimal parameters to create the selected mining model. The parameters are applied across the dataset, together with selected patterns and detailed statistics [5]. Numerous studies indicate that classification techniques and clustering are by far the most widely used data mining techniques. The hybrid technique is considered shortly after together with the Association technique [5]

C. Machine Learning Aspects

Machine learning could be a technique that has to give an enormous quantity of knowledge for coaching the model wherever to predict the long run aspects. Once the model learns from the info absolutely, there's a high chance to predict the long run properly [6]. Machine learning techniques square measure commonly used once any downside cannot be solved by any mathematical calculation or writing any script alone. There are 2 classes of machine learning issues which will be self-addressed. One is supervised learning and alternative is unsupervised learning [7].

Supervised Learning: In supervised learning, predefined dataset has been provided before coaching the algorithms. Firstly, these datasets area unit labelled and supported the labels or tags, the algorithms learn. Once learning from the dataset, model will predict any future expectations [8].

Unsupervised Learning: We propose unattended NIDS with reinforcement learning algorithmic program that is compatible with the noted attack still as AN unknown attack [9].That we have a tendency to decision zero-day attack. As a result of supported the Deep Q Learning algorithmic program, that doesn’t want any past expertise, sees each attack, i.e., noted attack or unknown attack as a brand new attack. Our planned Model’s initial half has the potential to discover numerous kinds of new attacks, as an example, DoS, DDoS, Heartbleed, port scanning or the other kinds of attack which can cause an enormous quantity of network traffic [10]. At intervals that point, pattern or behaviour of network traffic has been analysed by the persona non grata who can cause AN attack. Supported previous analysis on this, NIDS wants longer to visualize the traffic to convey correct call. In this paper, we have analyses of existing IDS system which is based on various mechanisms. Our aim is to find the issues in existing IDS which cannot predict the type of network attack and having lowest accuracy.

II. LITERATURE REVIEW

2.1. Literature Survey

According [11] IDS classified the intrusion detection system into 2 sorts particularly Network primarily based IDS and Host IDS. The latter monitors all the activities of inspected packets and resources that are being utilized by the programs. Just in case of any alteration in networks, user gets a network alert. HIDS is incorporated into the pc framework to sight the abnormalities and shield the knowledge from the trespasser. On the opposite hand, NIDS is that the attribute perform of target system. It uses anti-thread software package to manage incoming and outgoing threads. It consists of signature-based classification, that facilitate in distinctive the abnormalities by comparison it with log files and former signature. The authors of [12] projected Associate in Nursing AI primarily based Intrusion detection system employing a deep neural network. Neural networks consisting of 4 hidden layers and a hundred hidden units were used for the intrusion detection system. They used non-linear Rely because the activation operates for the hidden layer neurons to reinforce the model’s performance. They adopt random improvement technique for learning in DNN. For the coaching and testing of their model they used KDD CUP ninety nine dataset. They were able to reach the accuracy of ninety nine for all the cases. They need projected a NIDS (Network Intrusion Detection System) that relies on a feature choice technique referred to as algorithmic Feature Addition (RFA) and written word technique. They tested the model on the ISCX 2012 information set. Moreover, they need projected a written word technique to encrypt payload string options into a helpful illustration that may be employed in feature choice. they need additionally projected a replacement analysis metric referred to as that mixes accuracy, detection rate and warning rate during a method that helps in comparison completely different systems and choosing the simplest among them.

In [13] They have planned a replacement intrusion detection system and self-addressed the matter of ability within the field of intrusion detection. The planned IDS is associate degree adaptation answer that provides the potential of detection famed and novel attacks further being updated in step with the new input from human consultants in an exceedingly cost-efficient manner. It deals with the analysis and applied math analysis of tagged flow primarily based CIDDS-001 dataset used for evaluating Anomaly based (NIDS) Network Intrusion Detection Systems. They essentially used 2 techniques; k-means clump and k-nearest neighbor classification to live the complexness in terms of outstanding metrics. supported analysis, they ended that each k-means clump k-nearest neighbor classification perform spill CIDDS-001 dataset in terms of used outstanding metrics. Thence the dataset are often used for the analysis of Anomaly based mostly Network Intrusion Detection Systems.

As per [14] The IDS is predicated on anomaly detection technique. In such technique, a system tries to estimate the ‘normal’ state of the network associate degreed generates an alert once any activities deviate from this ‘normal’ state. The most advantage of anomaly-based system is that it's ready to discover antecedently unseen intrusion events. They need classified detection techniques into 3 classes applied mathematics primarily based, knowledge-based, and machine learning-based. In applied mathematics primarily based technique, a random viewpoint is employed to represent the behavior of the system. Whereas information primarily based technique, utilize the offered system knowledge to capture the behavior of system. Finally, the machine learning primarily based technique uses a certain or implicit model to modify categorization of the analyzed pattern. Various machine-learning techniques may result in higher detection rates, lower warning rates, affordable computation, and communication prices in intrusion detection. During [15], Mahdi Zamani and Mahnush Movahedi studied many such technique and schemes to match all their performance. They divide the schemes into strategies supported classical procedure intelligence (CI) and AI (AI). They make a case for however many options of CI techniques may be wont to build trendy and economical

IDS. Firstly, network attacks square measure known and also the performance of the algorithms square measure compared. The Dimension Reduction focuses on victimization info obtained KDD Cup ninety nine knowledge set for the choice of attributes to spot the kind of attacks. The spatial property reduction is first of all performed on forty one attributes to fourteen and seven attributes supported Best initial Search technique so two-classification algorithmic program square measure applied.

Author	Objective	Tool Used	Algorithm Used	Accuracy	Result
[11]	Gives review on the Data Fusion for network IDS.	-	-	-	They provide various techniques which is applied or helpful in intrusion detection.
[12]	Port Scan detection trying the Analysis of Deep Learning and machine learning algorithms	-	Deep Learning and SVM	97.80% and 69.79% Precision 99% and 80%	Results show that the deep learning algorithm performed significantly better results than SVM
[13]	IDS using various data mining techniques	Weka tool	AODE algorithm	97.19% Detection rate 98%	Result prove that accuracy, DR and MCC for four types of attacks are increased by the proposed method.
[14]	Intrusion Detection In Computer Networks By using Decision Tree Algorithm	Python	Decision tree	98.04% Precision 68% Recall 61%	it is said that Decision Tree model takes less time for training because it creates a tree to handle attributes for prediction outcomes and affects the final classification results
[15]	Is based on predicting attacks into two labels using ML algorithms.	Weka tool	Random tree, J48 and Naïve Bayes	99.7666, 99.7785 and 90.4384	They use Semi-supervised algorithm for classifying the attack into two labels normal and attack.

III. PROBLEM DEFINITION

Computer networks are widely used by industry, business and various fields of the human life. Therefore, building reliable networks is a very important task for IT administrators. On the other hand, the rapid development of information technology produced several challenges to build reliable networks which are a very difficult task. There are many types of attacks threatening the availability, integrity and confidentiality of computer networks. The Denial of service attack (DOS) considered as one of the most common harmful attacks. It can be surprising at first to realize that despite extensive academic research efforts on anomaly detection, the success of such systems in operational environments has been very limited. In other domains, the very same machine learning tools that form the basis of anomaly detection systems have proven to work with great success, and are regularly used in commercial settings where large quantities of data render manual inspection infeasible. We believe that this “success discrepancy” arises because the intrusion detection domain exhibits particular characteristics that make the effective deployment of machine learning approaches fundamentally harder than in many other contexts. In the following we identify these differences, with an aim of raising the community’s awareness of the unique challenges anomaly detection faces when operating on network traffic. We note that our examples from other domains are primarily for illustration, as there is of course a continuous spectrum for many of the properties discussed (e.g., spam detection faces a similarly adversarial environment as intrusion detection does). We also note that we are network security researchers, not experts on machine-learning, and thus we argue mostly at an intuitive level rather than attempting to frame our statements in the formalisms employed for machine learning. However, based on discussions with colleagues who work with machine learning on a daily basis, we believe these intuitive arguments match well with what a more formal analysis would yield. For an anomaly detection system, a thorough evaluation is particularly crucial to perform, as experience shows that many promising approaches turn out in practice to fall short of one's expectations. That said, devising sound evaluation schemes is not easy, and in fact turns out to be more difficult than building the detector itself. Due to the opacity of the detection process, the results of an anomaly detection system are harder to predict than for a misuse detector.

IV. CONCLUSION

The rise of the internet services along with the continued growth of access around the world, network traffic security is becoming a major issue in computer network system. Every day the number of attacks is increasing in computer network. For the reason that Intrusion detection in network is very important to detect and prevent intrusions and analyse huge number of network data and classify all of these network data into anomaly and normal data but traditional IDS suffer from different problems that limit their effectiveness and efficiency. A Machine learning researcher is to design more efficient IDS (in terms of both time and space) and practical general purpose learning methods that can perform better over a widespread domain. In the context of Machine learning, the efficiency with which a method utilises data resources that is also an important performance paradigm along with time and space complexity. Higher accuracy of prediction and humanly interpretable prediction rules are also of high importance.

REFERENCES

[01] David Ahmad Effendy, Kusrini Kusrini, Sudarmawan Sudarmawan, "Classification of Intrusion Detection System (IDS) Based on Computer Network" in 2017 IEEE.

[02] Amreen Sultana, M.A.Jabbar, "Intelligent Network Intrusion Detection System is using Data Mining Techniques" in IEEE 2016.

[03] Dr. Uma Kumari, Uma Soni, "A Review of Intrusion Detection using Anomaly based Detection" in Proceedings of the 2nd International Conference on Communication and Electronics Systems (ICCES 2017) .

[04] James P. Anderson, "Computer security threat monitoring and surveillance," Technical Report 98-17, James P. Anderson Co., Fort Washington, Pennsylvania, USA, April 1980.

[05] Nawfal Turki Obeis and Wesam Bhaya, "Review of Data Mining Techniques for Malicious Detetion", Research journal of Applied Sciences 11(10):942-947, 2016.

[06] Jau-Hwang WANG, Peter S. DENG, "Virus Detection Using Data Mining Techniques", TAO-Yuar, Taiwan, ROC333.

[07] Chi Zhang and Jinyuan Sun, "Privacy and Security for Online Social Networks: Challenges and Opportunity", Yuguang Fang, University of Florida and Xidian University.

[08] Uma Salunkhe and Suresh N. Mali, "Enrichment in Intrusion Detection System Using Ensemble", Journal of Electrical and computer Engineering.

[09] Q.S. Qassim, A. M. Zin and M. J. Ab Aziz, "Anomalies classification approach for network- based intrusion detection system", International Journal of Network Security, pp.1159-1171, 2016.

[10] O.Y.Al-Jarrah, O. Alhussein, P.D.Yoo, S. Muhaidat, K.Taha and K. Kim, " Data Randomization and Cluster-based Partitioning for botnet intrusion detection", IEEE Transactions on Cybernetics, vol. 46, no. 8, pp. 1796-1806, 2016.

[11] Solane Duque, Dr. Mohd. Nizam Bin Omar, "Using Data Mining Algorithm for Developing a Model for Intrusion Detection System (IDS)", procedia Computer Science 61 (2015) 46-51

[12] Crescenzo, G. D., Ghosh, A., And Talpade, R. "Towards a theory of intrusion detection". In 10th European Symposium on Research in Computer Security ESORICS (2005), pp. 267–286

[13] Wenying Feng, Q. Z. (2014) "Mining network data for intrusion detection through combining SVMs with ant colony networks". Future Generation Computer Systems, ELSEVIER.

[14] Mazyar Mohammadi Lisehroodi, Z. M. (2013) "A hybrid framework based on neural network mlp and kmeans clustering for intrusion detection system". Proceedings of the 4th International Conference on Computing and Informatics, ICOCI 2013 (p. Paper No. 020). Sarawak, Malaysia: Universiti Utara Malaysia.

[15] A.M.Chandrashekhar, K. (2013) "Fortification of hybrid intrusion detection system using variants of neural networks & support vector machines". International Journal of Network Security & Its Applications (IJNSA).

[16] Denning, D. E. "An intrusion-detection model. IEEE Transactions on Software Engineering" Special issue on computer security and privacy 13, 2 (Feb. 1987), 222–232

[17] Corchado, E., And Herrero, "A Neural visualization of network traffic data for intrusion detection". Applied Soft Computing 11, 2 (Mar. 2011), 2042–2056.

[18] Levent Koc, T. A. (2012) "A network intrusion detection system based on a Hidden Naïve Bayes multiclass classifier". Expert Systems with Applications, ELSEVIER.

[19] Mazyar Mohammadi Lisehroodi, Z. M. (2013) "A hybrid framework based on neural network mlp and kmeans clustering for intrusion detection system". Proceedings of the 4th International Conference on Computing and Informatics, ICOCI 2013 (p. Paper No. 020). Sarawak, Malaysia: Universiti Utara Malaysia

[20] Crescenzo, G. D., Ghosh, A., And Talpade, R. "Towards a theory of intrusion detection". In 10th European Symposium on Research in Computer Security ESORICS (2005), pp. 267–286