International Journal for Rapid Research in Engineering Technology & Applied Science
Vol IX Issue I January 2023
ISSN Online:- (2455-4723)

A Review on Intrusion Detection System using Machine Learning

Nitika jain , Dr Jitendra Singh Chouhan
Department of computer science and engineering, with specialization in software engineering Aravali Aravali institute of technical studies, udaipur, rajasthan rajasthan technical university, kota, rajasthan
nitjain06@gmail.com , Jitendrasingchauhan1984@gmail.com

Abstract: An intrusion detection system (IDS) are devices or software's that are used to monitors networks for any unkind activities that bridge the normal functionality of systems hence causing some policy violation. This paper reviews some of the intrusion detection systems and software's highlighting their main classifications and their performance evaluations and measure. As in today's developing network environment there is threat of new type of attacks daily in the network. So, the network administration system is also needed to be updated regularly for up gradation of security level. One of the network packet monitoring system is Intrusion detection systems (IDS).There are many techniques in the literature for developing these defense systems. However, it is also important to examine the improvement of the datasets used to train and test these security systems. Enhanced datasets extend the detection capabilities of offline and online intrusion detection models. Standard datasets such as KDD 99 and NSL-KDD are obsolete and do not contain data on current attacks such as denial of service. Therefore, they are not suitable for evaluation. This article presents an in-depth analysis of IDS records and presents the challenges of IDS. This article also provides an overview of the deep learning approach that can be used to develop a better network intrusion detection system.

Keywords: Intrusion Detection, Security System, Deep learning, Attack

I Introduction

compromise the overall integrity and confidentiality of a resource. The goal therefore of intrusion detection is to identify accessors that attempt to intrude and compromise systems security controls. Current IDS examine the entire data features to detect any intrusion and misuse patterns, although some of the features may be redundant and may contribute less to the detection process [1]. Current anomaly based intrusion detection systems and many other technical approaches have been developed and deployed to track novel attacks on systems. 98% detection rates at a high and 1% at a low alarm rate can therefore be achieved by using these techniques [2]. This paper review the various intrusion detection systems by evaluating their performance measures.

According to V. Jyothsna[3] there are three main types of intrusion detection systems: -signaturebased (SBS), anomaly-based (ABS) intrusion detection systems and Network Intrusion Detection System (NIDS). SBS systems such as Snort [3]make use of pattern recognition techniques by maintaining the database of signatures of previously known attacks to compare them with newly analyzed data. An alarm is raised when similarities are established. On the other hand ABS systems such as PAYL [4] build a statistical model to describe the normal network traffic, where any abnormal behavior that deviates from the model are identified. On the contrary anomaly-based systems have the advantage that they can detect zero-day attacks [2]. a) Signature based Detection With the explosion of internet commerce, e-business services on the web, e-banking and other high profile applications, organizations providing this services need to prepare themselves to the best possible protection against unauthorized penetration [5]. Signature detection involves searching network traffic for a series of malicious bytes or packet sequences. The main advantage of this technique is that signatures are very easy to develop and understand if we know what network behavior we are trying to identify. The events generated by signature based IDS can communicate the cause of the alert. As pattern matching can be done more efficiently on modern systems so the amount of power needed to perform this matching is minimal for a rule set. This technique can be easily deceived because they are only based on regular expressions and string matching. These mechanisms only

look for strings within packets transmitting over wire. More over signatures work well against only the fixed behavioral pattern, they fail to deal with attacks created by human or a worm with self-modifying behavioral characteristics. Signature based detection system (also called misuse based), this type of detection is very effective against known attacks, and it depends on the receiving of regular updates of patterns [6]. But signature based detection does not work well when the user uses advanced technologies like NOP generators, payload encoders and encrypted data channels. The efficiency of the signature based systems is greatly decreased, as it has to create a new signature for every variation. As the signatures keep on increasing, the system engine performance decreases. Due to this, many intrusion detection engines are deployed on systems with multi processors and multi Gigabit network cards. IDS developers develop the new signatures before the attacker does, so as to prevent the novel attacks on the system. The difference of speed of creation of the new signatures between the developers and attackers determine the efficiency of the system [2]. b) Anomaly based Detection An anomaly-based intrusion detection system is an intrusion detection system for detecting both network and computer intrusions and misuse by monitoring system activity and classifying it as either normal or anomalous. The classification is based on heuristics or rules, rather than patterns or signatures, and attempts to detect any type of misuse that falls out of normal system operation. This is as opposed to signature-based systems, which can only detect attacks for which a signature has previously been created [7]. The anomaly based detection is based on defining the network behavior. The network behavior is in accordance with the predefined behavior, then it is accepted or else it triggers I 11 © 2017 Global Journals Inc. (US) ( ) C Global Journal of Computer Science and Technology Volume XVII Issue III Version I Year 2017 II. CLASSIFICATION OF IDS Author α σ: Jomo Kenyatta University of Agriculture and Technology. e-mails: Jneyole434@gmail.com, ymuchelule@gmail.com Keywords: IDSs. performance measure and performance measures. the event in the anomaly detection. The accepted network behavior is prepared or learned by the specifications of the network administrators. The important phase in defining the network behavior is the IDS engine capability to cut through the various protocols at all levels. The Engine must be able to process the protocols and understand its goal. Though this protocol analysis is computationally expensive, the benefits it generates like increasing the rule set helps in less false positive alarms. The major drawback of anomaly detection is defining its rule set. The efficiency of the system depends on how well it is implemented and tested on all protocols. Rule defining process is also affected by various protocols used by various vendors. Apart from these, custom protocols also make rule defining a difficult job. For detection to occur correctly, the detailed knowledge about the accepted network behavior need to be developed by the administrators. But once the rules are defined and protocol is built then anomaly detection systems works well. c) Network Intrusion Detection System NIDS are deployed on strategic point in network infrastructure. The NIDS can capture and analyze data to detect known attacks by comparing patterns or signatures of the database or detection of illegal activities by scanning traffic for anomalous activity. NIDS are also referred as "packet-sniffers", because it captures the packets passing through the of communication mediums [6]. The network IDS usually has two logical components: the sensor and the management station. The sensor sits on a network segment, monitoring it for suspicious traffic. The management station receives alarms from the sensor(s) and displays them to an operator. The sensors are usually dedicated systems that exist only to monitor the network. They have a network interface in promiscuous mode, which means they receive all network traffic not just that destined for their IP address and they capture passing network traffic for analysis. If they detect something that looks unusual, they pass it back to the analysis station. The analysis station can display the alarms or do additional analysis. A fundamental problem for network intrusion detection systems (NIDSs) that passively monitor a network link is the ability of a skilled attacker to evade detection by exploiting ambiguitiesin the traffic stream as seen by the NIDS [8]

II Headings and Footnotes

Tommaso Zoppi (2020) et.al Anomaly detection aims at identifying patterns in data that do not conform to the expected behavior, relying on machine-learning algorithms that are suited for binary classification. It has been arising as one of the most promising techniques to suspect intrusions, zero-day attacks and, under certain conditions, failures. This tutorial aims to instruct the attendees to the principles, application and evaluation of anomaly-based techniques for intrusion detection, with a focus on unsupervised algorithms, which are able to classify normal and anomalous behaviors without relying on input data with labeled attacks.

Anushka Srivastava (2019) et.al Intrusion Detection is a vastly growing area. Traditionally supervised learning techniques were used for detecting intrusions in the network traffic data. But nowadays not only the rate of traffic has increased tremendously, but the nature of

network attacks is also changing. Detecting these new types of attacks requires improvements in detection techniques. Machine learning algorithms are well researched by researchers for detecting anomalies in the network traffic. New datasets have been added in the public repositories. In this paper, we have used novel feature reduction-based machine learning algorithms for detecting anomalous patterns in the recently provided dataset. High accuracy of 86.15 percent has been achieved.

P Illavarason (2019) et.al Network security is the most challenging task of the modern digital era. Due to the development in internet, the number of network attacks has also increased, this is prevented by access control, key manager, and intrusion detection system. Among these the most challenging task is intrusion detection system that ensures the network security. The current approach focuses on the important issues in intrusion detection system, which will identify the unwanted attacks and unauthorized access in the network. The comprehensive overview of the detailed survey is analyzed with the existing data set for identifying the unusual attacks that can understand the current issues in intrusion detection problems. The detailed investigation is reported for observing several issues on the intrusive performance by using the machine learning classification. Here machine learning classification algorithm is used for detecting the several categories of attacks. Furthermore, this study evaluates the performance criteria based on the feature extraction and machine learning classification techniques algorithm. Finally, based on the results observed we recommend some important features by using machine learning classification in order to find out the efficient method for detecting the particular attack.

Wei Zhong (2020) et.al With vast amounts of data being generated daily and the ever increasing interconnectivity of the world's internet infrastructures, a machine learning based Intrusion Detection Systems (IDS) has become a vital component to protect our economic and national security. Previous shallow learning and deep learning strategies adopt the single learning model approach for intrusion detection. The single learning model approach may experience problems to understand increasingly complicated data distribution of intrusion patterns. Particularly, the single deep learning model may not be effective to capture unique patterns from intrusive attacks having a small number of samples. In order to further enhance the performance of machine learning based IDS, we propose the Big Data based Hierarchical Deep Learning System (BDHDLS). BDHDLS utilizes behavioral features and content features to understand both network traffic characteristics and information stored in the payload. Each deep learning model in the BDHDLS concentrates its efforts to learn the unique data distribution in one cluster. This strategy can increase the detection rate of intrusive attacks as compared to the previous single learning model approaches. Based on parallel training strategy and big data techniques, the model construction time of BDHDLS is reduced substantially when multiple machines are deployed.

Nimmy Krishnan (2018) et.al Recent times have seen a steady shift of technology from traditional software models to the cloud. The substantial growth in the number of applications using cloud based infrastructures calls for the need of security mechanisms for their protection. Intrusion detection systems are one of the most suitable security solutions for protecting cloud based environments. Although there are several approaches to intrusion detection, such as signature-based and anomaly-based, machine learning (ML) based approaches have emerged as a recent interest and research area. With their robust learning models, and data centric approach, ML based security solutions for cloud environments have been proven effective. Attack features are extracted from network and application logs. Attack presence is confirmed by performing Machine learning techniques such as logistic regression and belief propagation. Performance measures such as average detection time is used to evaluate the performance of the approach.

S. Shinly Swarna Sugi (2020) et.al Internet of Things (IoT) combines the internet and physical objects to transfer information among the objects. In the emerging IoT networks, providing security is the major issue. IoT device is exposed to various security issues due to its low computational efficiency. In recent years, the Intrusion Detection System valuable tool deployed to secure the information in the network. This article exposes the Intrusion Detection System (IDS) based on deep learning and machine learning to overcome the security attacks in IoT networks. Long Short-Term Memory (LSTM) and K-Nearest Neighbor (KNN) are used in the attack detection model and performances of those algorithms are compared with each other based on detection time, kappa statistic, geometric mean, and sensitivity. The effectiveness of the developed IDS is evaluated by using Bot-IoT datasets.

Indrajit Das (2021) et.al Cyber-attacks have been the major concern with the growing advancement in technology. Complex security models have been developed to combat these attacks, yet none exhibit a full-proof performance. Recently, several machine learning (ML) methods have gained significant popularity in offering effective and efficient intrusion detection schemes which assist in proactive detection of multiple network intrusions, such as Denial of Service (DoS), Probe, Remote to User (R2L), User to Root attack (U2R).

Multiple research works have been surveyed based on adopted ML methods (either signature-based or anomaly detection) and some of the useful observations, performance analysis and comparative study are highlighted in this paper. Among the different ML algorithms in survey, PSO-SVM algorithm has shown maximum accuracy. Using RBF-based classifier and C-means clustering algorithm, a new model i.e., combination of serial and parallel IDS is proposed in this paper. The detection rate to detect known and unknown intrusion is 99.5% and false positive rate is 1.3%. In PIDS (known intrusion classifier), the detection rate for DOS, probe, U2R and R2L is 99.7%, 98.8%, 99.4% and 98.5% and the False positive rate is 0.6%, 0.2%, 3% and 2.8% respectively. In SIDS (unknown intrusion classifier), the rate of intrusion detection is 99.1% and false positive rate is 1.62%. This proposed model has known intrusion detection accuracy similar to PSO - SVM and is better than all other models. Finally, the future research directions relevant to this domain and contributions have been discussed.

Abhinav Singhal (2021) et.al This paper outlines an approach to build an Intrusion detection system for a network interface device. This research work has developed a hybrid intrusion detection system which involves various machine learning techniques along with inference detection for a comparative analysis. It is explained in 2 phases: Training (Model Training and Inference Network Building) and Detection phase (Working phase). This aims to solve all the current real-life problem that exists in machine learning algorithms as machine learning techniques are stiff they have their respective classification region outside which they cease to work properly. This paper aims to provide the best working machine learning technique out of the many used. The machine learning techniques used in comparative analysis are Decision Tree, Naïve Bayes, K-Nearest Neighbors (KNN) and Support Vector Machines (SVM) along with NSLKDD dataset for testing and training of our Network Intrusion Detection Model. The accuracy recorded for Decision Tree, Naïve Bayes, K-Nearest Neighbors (KNN) and Support Vector Machines(SVM) respectively when tested independently are 98.088%, 82.971%, 95.75%, 81.971% and when tested with inference detection model are 98.554%, 66.687%, 97.605%, 93.914%. Therefore, it can be concluded that our inference detection model helps in improving certain factors which are not detected using conventional machine learning techniques.

Toya Acharya (2021) et.al The internet-based services undoubtedly led the worldwide revolution with exponential growth, but security breaches resulting personal digital asset losses which need for a comprehensive cybersecurity solution. Traditionally, signature-based network intrusion detection is employed to capture attributes of normal and abnormal traffics in a network, but it fails to detect the zero-day attack. The machine learning-based approach is attractive among various known NIDS methods to circumvent the shortcoming because machine learning based approach can efficiently analyze the big network traffic data and efficiently detect the zero-day attack. The imbalanced NIDS dataset does not provide better performance on practical implementation scenarios. Reducing the number of target classes into a new target class creates a balanced NIDS and improved classifier performance. In this paper, we present the efficacy of several machine learning algorithms, including Random forest (RF), J48, Naïve Bayes, Bayesian Network, Bagging, AdaBoost, and Support Vector Machine (SVM) using network logs traffic (KDD99, UNSW-NB15, and CIC-IDS2017) using WEKA. This paper examined the impact of changing the number of output classes of the publicly available network intrusion datasets on sensitivity (True Positive Rate), False Positive Rate (FPR), Area under the ROC curve (AUC) and incorrectly identified percentage. Interestingly, the efficiency of these classifiers has increased, adding strongly correlated features to the target classes. The experimented results reveal that the machine learning classifiers performance improved when the number of target classes decreased. The addition of a highly correlated feature to the output class increases the performance of the classifiers.

Chung-Ming Ou (2019) et.al An adaptable agent-based IDS (AAIDS) inspired by the danger theory of artificial immune system is proposed. The learning mechanism of AAIDS is designed by emulating how dendritic cells (DC) in immune systems detect and classify danger signals. AG agent, DC agent and TC agent coordinate together and respond to system calls directly rather than analyze network packets. Simulations show AAIDS can determine several critical scenarios of the system behaviors where packet analysis is impractical.

Monika D. Rokade (2021) et.al Computer network and virtual machine security is very essential in today's era. Various architectures have been proposed for network security or prevent malicious access of internal or external users. Various existing systems have already developed to detect malicious activity on victim machines; sometimes any external user creates some malicious behavior and gets unauthorized access of victim machines to such a behavior system considered as malicious activities or Intruder. Numerous machine learning and soft computing techniques design to detect the activities in real-time

network log audit data. KKDDCUP99 and NLSKDD most utilized data set to detect the Intruder on benchmark data set. In this paper, we proposed the identification of intruders using machine learning algorithms. Two different techniques have been proposed like a signature with detection and anomaly-based detection. In the experimental analysis, demonstrates SVM, Naïve Bayes and ANN algorithm with various data sets and demonstrate system performance on the real-time network environment.

The majority of published documents claiming to evaluate IDSs are conducted as comparisons, rather than evaluations. Evaluation should be considered to be a determination of the level to which a particular IDS meets specified performance targets [9].The basic task in intrusion detection system is to classify network activities as normal or abnormal while minimizing misclassification [10]. Many problems exist in IDS and need to be addressed, such as the low detection capability against the unknown network attack, high false alarm rate, and insufficient analysis capability. Generally, intrusion detection is targeted as classification problem, to distinguish between the normal activities and the malicious activities [11].

According to the NSS publication "Intrusion Detection Systems Group Test(2001), the evaluation of each IDS consists of two components. The first component is a qualitative analysis of the various features and functions of each product. The comments and analysis of the various features are well considered and unbiased [12]. The group further established that the quantitative component of consisted of four tests of the NIDSs on a controlled laboratory network. These test focused upon specific performance indicators, attack recognition, performance under load, ability to detect evasion techniques and a stateful operation test.

The performance measures used by these evaluation were: a ratio of attack detection to false positive, ability to detect new and stealthy attacks, a comparison of host vs. network based systems to detect different types of attacks, the ability of anomaly detection techniques to detect new

attacks, improvements between 1998 and 1999, and the ability of systems to accurately identify attacks. The research also attempted to establish the reason each IDS failed to detect an attack, or generated a false positive. Both the 1998 and 1999 evaluations identified a number of weaknesses with existing IDSs.

A number of these issues have since been resolved, while others are still valid. The testing process used sample of generated network traffic, audit logs, system logs and file system information. This information was then distributed to various evaluators who would provide the appropriate data to the Intrusion Detection Systems. This ensured each system was provided with identical data, whilst allowing proper configuration of each system.

Ranum (2001) extract established that constructing good benchmarks and tests for IDS was difficult and in order to accurately measure IDS complexity one needed to expand considerable efforts in designing tests by ensuring that the tests weren't inherently biased or inaccurate. This was a challenge to the IDS especially as they depend on operation environment. He further concluded that if tests were to be made they were to base on qualitative and comparative measures. In his summary he presented some experiences in benchmarking IDS with a focus on poorly designed tests and their effects. And a technology continue to advance the IDS management systems would become increasingly inefficient [13].

Alessandri [14] proposed the use of a systematic description scheme for regulating the descriptions used to describe IDS functions. This approach should allow for an evaluation of IDSs based upon their descriptions, without necessitating experimentation. The disadvantage of this approach is the requirement of accurate descriptions. Currently such an approach does not exist so implementing it is not possible. This approach does hold a certain promise for the future

INTRUSION DETECTION DATASETS

Evaluation records play an important role in validating an IDS approach by allowing us to evaluate the ability of the proposed method to detect intrusive behavior. The datasets used for analyzing network packets in commercial products are not readily available for data protection reasons. However, there are publicly available records such as DARPA, KDD, NSL-KDD and ADFA-LD which are commonly used as a reference. The existing datasets

used to create and compare IDSs are explained in this section with their functions and restrictions. A. KDD Cup 99 Dataset The first attempts to create an IDS record were made in 1998 by the Defense Advanced Research Project Agency (DARPA) and created the Knowledge Discovery and Data Mining (KDD) record. In 1998, DARPA introduced a programmer to the MIT Lincoln labs to provide a complete and realistic IDS benchmarking environment (MIT Lincoln Laboratory, 1999). Although this dataset was an important contribution to IDS research, its accuracy and ability to take actual conditions into account has been widely criticized. These datasets were collected using multiple computers connected to the Internet to model a small American air base with limited personnel. Network packets and host log files were collected. Lincoln Labs created an experimental test environment to get a snapshot of the 2-month TCP packet for a Local Area Network (LAN) using a typical US Air Force LAN. They modeled the LAN as if it were a real Air Force environment, but nested it with multiple simulated interventions. The network packets collected were approximately four gigabytes in size and contained approximately 4,900,000 datasets. The 2-week test data contained approximately 2 million connection records, each with 41 characteristics and were classified as normal or abnormal. The extracted data are a series of TCP sessions that begin and end at specific times, among which the data flows from and to a source IP address to a destination IP address that contains a variety of attacks that occur in an environment of military network. The 1998 DARPA dataset was used as a basis for deriving the KDD Cup99 dataset, which was used in the third international competition for knowledge discovery and data mining tools (KDD, 1999). The 1999 KDD Cup dataset was used for the third international competition for the discovery of knowledge and data mining tools. Each connection instance is described by 41 attributes (38 continuous or discrete numeric attributes and 3 symbolic attributes). Each instance is called a normal attack or a specific type of attack. These attacks fall into one of four categories: DoS, Probe, U2R and R2L. The 1999 KDD Cup provided training and test data sets, called 10% KDD or correct data sets. The 10% KDD dataset contains 22 types of attacks, while the correct dataset contains the same 22 types of attacks and 17 types of additional attacks. These records are obsolete because they do not contain any records of recent malware attacks. For example, the behavior of attackers differs between different network topologies, operating systems and different criminal software and toolkits. Nonetheless, KDD99 continues to be used as a reference within the IDS research community and is currently still used by researchers. B. NSL-**KDD Dataset**

NSL-KDD is a public dataset, which has been developed from the earlier KDD-99 dataset. Statistical analysis of the cup99 dataset resulted in significant problems that significantly affect the accuracy of intrusion detection and lead to a misleading assessment of AIDS. The main problem with registering KDD is the large number of duplicate packages. Tavallaee et al.analyzed the KDD training and test sets and found that approximately 78% and 75% of network packets are duplicated in training and test data sets. This huge number of duplicate instances in the training set would affect the machine learning methods aimed at normal instances, preventing them from learning from irregular instances that normally damage the computer system. Tavallaee et al. created the 2009 NSLKDD dataset from the KDD Cup'99 dataset to solve the above problems by eliminating duplicate datasets. The NSL-KDD train data record contains 125,973 data records and the test data record contains 22,544 data records. The size of the NSL-KDD record is sufficient to facilitate the use of the entire NSL-KDD record without the need for a random sample. This has led to consistent and comparable results from various research projects. The NSL_KDD dataset includes 22 training intrusion attacks and 41 attributes (i.e. characteristics). In this dataset, 21 attributes refer to the connection itself and 19 attributes describe the type of connection within the same host.

CHALLENGES DURING DEVELOPMENT OF IDS

some challenges of IDS are discussed as below: A. Challenges related to nature of datasets Two main categories of challenges arise during the development of NIDSs to obscure future attacks. The first challenge is suitable feature selection. The second challenge is inaccessibility of labeled traffic datasets. Following challenges are related to nature of dataset:

- The imbalanced and diverse nature of the datasets.
- Unavailability of labeled traffic dataset from real networks.
- Misclassification of targeted input pattern.

B. Challenges related to data processing increase the use of new technologies in field of network communication leads to lower imperfection rates and therefore generates a huge amount of network data. The following challenges concern data processing.

- The processing of data is increased.
- Difficulty to process big data.
- Long time processing or computational complexity is high.
- Difficulty to process large network data for packet classification as there exists millions of packets.

C. Challenges related to security Challenges in security are divided into two categories. Firstly, the security of every machine presented currently to the Internet can be compromised and attacked. External attacks constantly threaten important data. Hackers discover new methods to steal or damage valuable data in every organization every day. Secondly, security demand is crucial to many companies and organizations that depend on a database to safeguard sensitive data. The value of certain data is worth millions. Thus, strong data protection must be guaranteed.

D. Challenges related to growth of new attacks Smartphone malware is currently used in daily life activities, such as entertainment, controlling smart homes and paying bills. Smart phones have demonstrated a considerable increase in growth rate given their mobility and ever-expanding capabilities. Android is an ideal platform for legitimate developers and attackers creating malware given its large market share and room for development

PROBLEM DOMAIN It is well known that anomaly-based IDS suffer from the high rate of false alarms. Continuous efforts are being made to reduce the high false positive rate. We believe that intrusion detection is a data analysis process and can be studied as a problem of classifying data correctly. From this standpoint, it can also be observed that any classification scheme is as good as the data presented to it as input. Cleaner the data, higher accurate results are likely to be obtained. From anomaly-based IDS point of view, it implies that if we can extract features that demarcate normal data from abnormal one properly, false positive rate can be reduced to a great extent. On the similar lines, we observe that most of the data mining and machine learning based methods in intrusion detection make use of wellknown tools and techniques. It may turn out that these general techniques are not very effective in classifying data as normal or abnormal with very high accuracy. There is a need to customize those techniques according to the requirement of intrusion detection. Apart from the problems mentioned above, the fast detection of attacks remains one of the focal points to be worried about. With the present complexity and variety of attacks, we need a huge amount of data to analyze and produce results. But larger the amount of data, longer the time to analyze it, which delays the detection of attacks. An IDS will be of more use if it can trigger an alarm early enough to reduce the damage that an ongoing attack can do. Thus, there is a need to make IDS as fast as to operate on-line. It is believed that this can be achieved if we can reduce the data, to be analyzed, without degrading its quality.

IV Conclusions

Information security has become a legitimate concern for organizations and computer users due to the growing trust in computers and electronic transactions. Various techniques are used to ensure a company's security against threats or attacks. On the other hand, attackers are discovering new techniques and ways to violate these security guidelines. The main types of IDS technologies - network-based, wireless and host-based offer substantially different functions. This paper reviews and analyses the research area for intrusion detection systems (IDSs) based on deep learning (DL) techniques into a coherent taxonomy and identifies the gap in this pivotal research area

Insert acknowledgment, if any. The preferred spelling of the word “acknowledgment” in American English is without an “e” after the “g.” Use the singular heading even if you have many acknowledgments. Avoid expressions such as “One of us (S.B.A.) would like to thank ....” Instead, write “F. A. Author thanks ... .” Sponsor and financial support acknowledgments are also placed here.

References

1. Srilatha Chebrolua, Ajith Abrahama, Johnson P. Thomasa,. (2005). Feature deduction and ensemble design of intrusion detection systems. ELSEVIER, Pp. 295–307

2. V. Jyothsna, V. V. Rama Prasad, K. Munivara Prasad. (2011). A Review of Anomaly based Intrusion Detection Systems. International Journal of Computer Applications, pp. 26-36.

3. Shirazi, H. M. (2009). ”Anomaly Intrusion Detection System using Information Theory, K-NN and KMC Algorithms. Australian Journal of Basic and Applied Sciences, pp. 2581-2597

4. Wang. K and Stolfo.S.J. (2004). Anomalous Payloadbased Network Intrusion Detection. 7th Symposium on Recent Advances in Intrusion Detection (pp. pp. 203–222). USA: LNCS Springer-Verlag.
5. Brox, A. (2002, May 01st). THE CYBER SECURITY SOURCE. Retrieved December 20th, 2016, from SC Magazine US: https://www.scmagazine.com/signature-based-or-anomaly-based-intrusiondetection-the-practice-and-pitfalls/article/548733/
6. Asmaa Shaker Ashoor, Prof. Sharad Gore. (2005). Importance of Intrusion Detection System (IDS).

International Journal of Scientific Engineering Research, pp. 1-7.

7. Anomaly-based intrusion detection system. (2016, July 16th). Retrieved December 20th, 2016, from Wikipedia Encyclopedia: https://en.wikipedia.

8. Mark Handley, Vern Paxson and Christian Kreibich. (2001). Network Intrusion Detection: Evasion, Traffic Normalization, and End-to-End Protocol Semantics. Berkeley, CA 94704 USA: International Computer Science Institute

9. Wilkison, M. (2002, June 10th). IDFAQ: How to Evaluate Network Intrusion Detection Systems? Retrieved from SANS Technology Institute: https:// www.sans.org/security-resources/idfaq/how-toevaluate-network-intrusion-detection-systems/8/10

10. Leila Mohammadpour, Mehdi Hussain, Alihossein Aryanfar, Vahid Maleki Raee and Fahad Sattar. (2015). Evaluating Performance of Intrusion Detection System using Support Vector Machines: Review. International Journal of Security and Its Applications, pp.225-234.

11. Kuang, F., Xu, W., & Zhang, S. (2014). A novel hybrid KPCA and SVM with GA model for intrusion detection. Applied Soft Computing, pp. 178-184.

12. . E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors", *Comput. Sci.*, vol. 3, no. 4, pp. 212-223, 2012.

13. R. Vinayakumar, K. P. Soman, P. Poornachandran, "Applying convolutional neural network for network intrusion detection", *Proc. Int. Conf. Adv. Comput. Commun. Inform. (ICACCI)*, pp. 1222-1228, Sep. 2017.

14. R. Vinayakumar, K. P. Soman, P. Poornachandran, "Evaluating effectiveness of shallow and deep networks to intrusion detection system", *Proc. Int. Conf. Adv. Comput. Commun. Inform. (ICACCI)*, pp. 1282-1289, Sep. 2017.

15. G. Wang, X. Hao, J. Ma, L. Huang, "A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering", *Expert Syst. Appl.*, vol. 37, no. 9, pp. 6225-6232, 2010.

16. G. Ding, Q. Wu, L. Zhang, Y. Lin, T. A. Tsiftsis, Y.-D. Yao, "An amateur drone surveillance system based on the cognitive Internet of Things", *IEEE Commun. Mag.*, vol. 56, no. 1, pp. 29-35, Jan. 2018.

17. L. Huang, A. D. Joseph, B. Nelson, B. I. P. Rubinstein, and J. D. Tygar, "Adversarial Machine Learning," in ACM workshop on Security and artificial intelligence, 2011.

18. F. Pierazzi, G. Apruzzese, M. Colajanni, A. Guido, and M. Marchetti, "Scalable architecture for online prioritization of cyber threats," in International Conference on Cyber Conflict (CyCon), 2017.

19. J. Kim, J. Kim, H. L. T. Thu, and H. Kim, "Long Short Term Memory Recurrent Neural Network Classifier for Intrusion Detection," in IEEE International Conference on Platform Technology and Service (PlatCon),2016.

20. P. Torres, C. Catania, S. Garcia, and C. G. Garino, "An analysis of Recurrent Neural Networks for Botnet detection behavior," in IEEE Biennial Congress of Argentina (ARGENCON), 2016.

21. G. E. Dahl, J. W. Stokes, L. Deng, and D. Yu, "Large-scale malware classification using random projections and neural networks," in IEEE International Conference on Acoustics, Speech and SignalProcessing (ICASSP), 2013.

22. 12. G. D. Hill and X. J. Bellekens, "Deep Learning Based Cryptographic Primitive Classification," arXiv preprint, 2017.

23. M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," Science, 2015.