INTRUSION DETECTION SYSTEM

Ambermani Pratap Singh
Dept. of ISE
RV College of Engineering
Bangalore, Karnataka, India
ambermani.is16@rvce.edu.in

Vanishree K.
Assistant Professor Dept. of ISE
RV College of Engineering
Bangalore, Karnataka, India
vanishreek@rvce.edu.in

***Abstract*- Intrusion detection system or IDS is an application device or software that controls and monitors over the network or system activities and ensures none malicious tasks are being carried out. IDS finds and checks over any unwanted activities being carried out. The need of IDS has increased to a great extent because of growth and increased interaction of web and internet throughout the complete globe. This raise has resulted into greater cause of concern regarding the network communication and ensuring the safety of many secured digital data and information. So it's the primary job to preserve such important and secured information. Because of so much of upsurge of web globally hackers have many new techniques and practices in their bank to disregard the safety of our valuable info. So many of the intrusion detection system have devised several algorithms and techniques to help and safeguard against such attacks by hackers and intruders. The primary objective of this paper is to provide summarized study of IDS, techniques and algorithms behind the IDS, types of IDS available in the market, various ways of attacks, tools techniques and challenges faced, research and development against these challenges and many future scope of improvements in this field.**

***Keywords* - Intrusion Detection System, Machine Learning, Deep Learning, Cyber Security, Network Intrusion Detection System(NIDS).**

I. INTRODUCTION

During the present times global web network and internet security has becomes great obstacle for many companies, sectors and organizations. To ensure the safety of trusted credential data from the intruders with malicious intent. In this procedure of safety of such essential data points many data Web firewalls many cryptography authentication techniques and several VPN's have been introduced since past times to secure the architecture and infrastructure of network connectivity and web communication all over the globe. Intrusion detection techniques compared to these tools and methods is latest addition these

practices. IDS aims at finding out if any malicious activity or any intruder is involved or not.

IDS is a methodology to keep a check on such activities. IDS is an evolved technique that improves the network security and protecting the data and information of the organization. IDS generally aims at assisting the networks team by working with the network administrator. It helps the administrator by detecting intruders or malicious activities going around the network. IDS then informs the team and alerts the network to get the data safeguarded by taking efficient actions with respect to those attacks Going by the definition intrusion refers to any unwanted not authorized access or malicious use of information or any other resources. This intruder or the attacker is true entity present in this real world aiming to harm in any possible sorts by gaining unauthorized access.

Working of Intrusion detection system is very much relatable to firewall methodology. It is based on firewall security. The firewall safeguards the organization from malicious attacks through the web. IDS detects if anyone tries to access in through the firewall. If any malicious user tries to break into the firewall security and tries to have an access on any system within the organization, the ids generates and alerts system administrator.

The basic working principle of intrusion detection system is based on network traffic and connectivity. Therefore, intrusion detection system is a security system that keeps check over network traffic and systems and tries rise to analyse the incoming attacks and all other possible hostile situations originating from outside the organization through unusual traffic.

II. NEED AND MOTIVATION

In present Times internet has become the integral part of our daily life all throughout the globe also so there is significant increase in extent up to which business world is getting involved with internet everyday thousands of new in people are getting connected to internet. The business model based on internet interaction is also known as e business platform connectivity and network enhancement it has become integral need and Critical aspect of E- business platform.

Involving business with internet drinks both advantages and disadvantages to the business model. As it enhances growth rate of any business model it brings with itself lots of risk associated. When any organization provides its information on the web thinking about the harmless people involved in development of business there is increase chance that this idea can turn into a risk factor for that particular organization. Many malicious users, intruders are hackers may also get access to the organizations

internals systems in various reasons please can be for example

- Software bugs or weak points of the system.
- Not so secured administration and networks
- Relying on primitive default configuration.
- loopholes present in system security

Depending upon the situation the intruders use different types of techniques or methods such as password cracking, peer to peer attack, sniffing attack, dos attacks, is dropping attack, application layer attack etc. to work upon the system vulnerabilities mentioned in above paragraph. Therefore, there is a need for external support to these resources of the organization from the internet or even from users inside the organization.

III. TYPES OF INTRUSION DETECTION SYSTEMS

Primarily there are two types of intrusion detection systems. These are network based intrusion detection system and host based intrusion detection system. Classifications where made on the basis of medium or method of intrusion Commonly there are two kinds of IDS that exists, namely:

A. NETWORK BASED INTRUSION DETECTION SYSTEM (NIDS):

NIDS can be a platform which is independent and aims at detecting intrusions by examination of network traffic and monitoring multiple hosts. These style of system access traffic by behaving form of a parasite to the network hub, a switch specifically designed for port mirroring. For accuracy and efficiency, the sensors are situated at the strategically chosen locations, within the DMZ. These receptors, i.e. the sensors, collect all the network data and analyze the contents of packets.

B. HOST BASED INTRUSION DETECTION SYSTEM (HIDS):

In this type of IDS, the system usually contains a singular agent embedded onto variety system that detects intrusions by researching the logs, calls made by the system, modifications of file-system and other states and activities. In HIDS, a software agent can be a sensor.

Traditionally, these IDS systems make the most and make use of the signatures of known attacks to identify the inbound attacks. But due to the rapid advancement of latest kinds of malware whose signatures are unknown, the filtering of ever-increasing kinds of state of the art malware becomes difficult. This paved way for the conjunction of machine learning and cybersecurity. IDS in machine learning can be a classification problem. Therefore, possibly an unbiased dataset is utilized for training the model which model is optimized for integrating into the

IDS to predict malware more accurately. This whole prediction principle eliminates the entire dependency of malware signatures for detection, hence acting as a bonus over traditional IDS. For the prediction model to work more efficiently, we would like to feed it with many coaching data. Often these data are unlabeled and are available with improper feature engineering. The pitfall with traditional machine learning is that the training data must be labelled properly for the model to search out patterns then provide prediction results. But this will be not always possible to the vast size of the knowledge. this will be where Deep Learning (DL) has explicit advantages over conventional in algorithm scenarios of processing data that's unstructured. Pattern discovery are steer clear away the ancillary task of labelling. In classical machine learning theory, engineering features become an essential part. DL lacks the feature scrutiny. This becomes very necessary in parts where the target is to segregate fraudulent sequences versus non-fraudulent sequences and justify how we reached our conclusions.

IV. METHODOLOGY

Presently, this networks of computers are generating an unlimited amount of data traffic during a quick past. This generation of data is increasing continuously and also the kinds of data that the network traffic contains is additionally different. Current traffic of the overall public satisfies all the three V's of massive data namely, volume, velocity, and variety. Big Data can be a term implicates the large volume of data (structured & unstructured) that deluge business worldwide at a faster rate than anyone can predict. an outsized pool of data doesn't always signify that everything is useful and this needs a tool which will extract the essence from it to work out the cardinal step. This will be where the Analytics of the so called Big Data comes into play. Rapid digitization possesses to guide this world to some extent where we might run out storage capabilities soon.

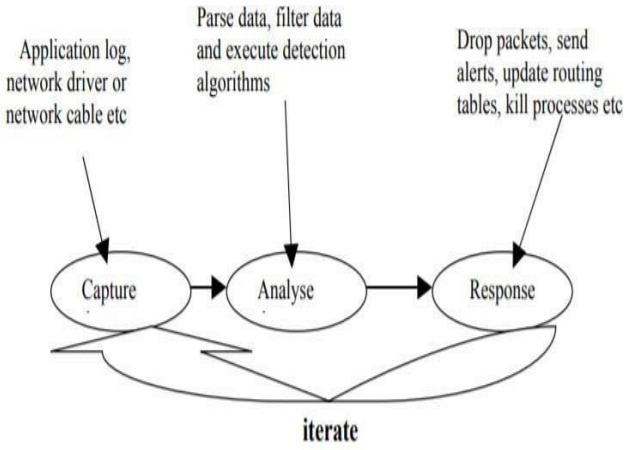
Detection of intrusion usually involves analyzing huge information. This can be outlined as a look downside that the technologies of computing that square measure thought cannot cope up with the flow of data. Even an isolated supply of the event will cause huge information challenges consistent with an analysis, a 1Gigabytes/second of continuous network information will cause problems associated with huge information for Intrusion Detection even once inspecting deep information packets. The ancient storage ways like relative information bases inherently have the capability to effectively scope against the character of massive data challenges caused by the detection of an intrusion. A typical answer accustomed handle such issues is an open sourced tool known as Hadoop. It is a storage platform that is distributed and is liberally open-sourced. It is the aptitude to run on most of the

hardware and has been exploited to raise and entertain the large information storage needs of large. Hadoop is a collective name for many different technologies like Pig, Map Reduce, Hive, HDFS, etc. Hadoop proposes the employment of those technologies to beat the hurdles in huge information and states that the 3Vs aren't sufficient to handle these problems and therefore introduces the 3Cs Construct, Cardinality, Continuity, and quality.

The process of mining this huge information and analyzing it's known as huge information analytics. The fast emendations in aspects of process, storing and analyzing of massive information in recent years includes:

- The speed decrease in storage, expense and central processing unit power.
- Cost-effectiveness and adaptability of Cloud computing and Datacenter's in accordance with storage and elastic computation.
- Development of revolutionary and new frameworks like Hadoop.

The ultimate objective of massive information analytics for IDS is to confirm promising, market prepared period intelligence, even if it's a compelling future, there square measure many issues (such as information source, privacy) that has got to be diminished so as to unleash the particular potential of it.



An IDS is composed of several components:

- Sensors which generate security events
- A Console to monitor events and alerts and control the sensors
- A central Engine that records events logged by the sensors in a database and uses a system of rules to generate alerts from security events received

Figure 1: Methodology Diagram for the working system

A. NEURAL NETWORKS

A Neural Network may be a paradigm associated with the process of data whose thought method is motivated by the approach the nervous systems of many biological components, like the brain, analyze and segregate data. The vital side of this approach is that the novel assembly of the system that processes data. it's engineered with various neurons that square measure extremely connected with one another, operating in unison to grasp the coaching dataset and solve issues specific to the set. These, like individuals, learn by example. Learning in biological atmosphere systems involves standardization of those links that between the neurons of multiple layers. this can be true of artificial neural networks still.

B. DEEP LEARNING

It is a natural extension to the present field of cubic centimeter exercised with models that square measure motivated by the assembly and performance of brain cells. one in all the key aspects of Neural networks is quantifiability that signifies that the result's superiority is directly proportional to the number of knowledge that we tend to use however it conjointly implies the requirement for considerably additional procedure time. additionally, to the advantage of straightforward scaling, another usually explicit and purpose is their intuitive ability to eliminate the burden of large-scale feature engineering. Especially, deciliter shines in eventualities that need analogue input-outputs that permits it to figure with RGB component information from pictures, text information from documents or audio information from music files instead of limiting the user to use quantities during a tabular format.

C. NEURAL NETWORKS & DEEP LEARNING TECHNIQUES FOR NETWORK BASED IDS (NIDS)

The aim of an IDS is to detect abnormalities within the traffic and mitigate a possible intrusion. must be an important efficient way considering variants of malware and kinds of attacks within the field of security. The of Signature-based and traditional if-else algorithm based detection layers are drawn in effective which imply the inclusion of within the field of AI within in the field NIDS inbound attack that's likely to cause damage or theft of digital assets. Many researchers are actively working for years now to experiment with different kinds of techniques and data sets to determine a network-based IDS that has Superhuman accuracies and capability. Deep learning and artificial neural networks have proven effective in such applications thanks to the advancement within the computational power of the commercially viable processors. To beat this of conventional intrusion detection systems, its proposed and implemented an analysis method that mixes the skills of a kernel PCA and LSTM-RNN to get the specific result features like of information, extraction of features, action of attacks are incorporated into the layer proposed by it. They need to use the NSL-KDD set for and has outperformed the capability shown by SVM, neutral network and Bayesian methods. The experiments done by [6] with LSTM-RNN's on the KD 99 has shown accuracy of about 93.82% thanks to their inherent ability to correlate to consecutive records by peeking back in time.

D. NEURAL NETWORKS & DEEP LEARNING TECHNIQUES FOR HOST BASED IDS (HIDS)

A host-based intrusion detection system (HIDS) is a technique that ensures and maintains records of tasks and activities of the computer network on which it is positioned to alert the admin in case if any of the protocols are breached. Host-based IDS can be assumed as an agent that checks and performs whether anything or anyone has run through the security policy of the system.

NIDS is generally kept at the demilitarized zone of the system network. Its aim is to check the incoming data packets and eliminate threats and malicious acts and then pass the data forward to the next stage if the existence of potential malicious packets is non-present with the transmission of data. It is more concentrated on the singular local node, unlike NIDS. [9] has made comprehensive experimentation by using DNNs and SVMs (Support Vector Machines) on the KDD-DARPA benchmarking dataset for making a Host-based IDS to overcome the issues of the traditional hard-coded logic and algorithms. The accuracy is concluded to be greater than 99% along with a training time of sec. An approach where Radial basis functions neural networks as containers are being used by [10] for host-based IDS which proves robust compared to the soft computing methods with very low training time and high accuracy in prediction. Hence, the solution to the question about superiority is very necessary. The final solution is using both NIDS and HIDS together with each other and AI integrated into them. In order for the host (and network) to remain secure, it is necessary that we should select the right IDS scheme that is appropriate for the all the instances.

V. ARCHITECTURE

An overview of planned DNN's design for all use cases is shown in Fig. This includes a hidden-layer of five and an output layer. The input layer consists of forty-one neurons. The neurons in input-layer to hidden-layer and hidden to output-layer area unit connected fully. The back-propagation mechanism is employed to coach the DNN networks. The planned network consists of totally connected layers, bias layers and dropout layers to form the network additional sturdy.

- **Input and hidden layers***:* This layer consists of forty-one neurons. These area units then fed into the hidden layers. Hidden layers use ReLU because the non-linear activation operates. Then weights area unit value-added to feed them forward to consequent hidden layer. The somatic cell count in every hidden layer is slashed steady from the primary to the output to form the outputs additional correct and at an equivalent time reducing the machine price.
- **Regularization***:* To build the complete method economical and time-saving, Dropout (0.01). The operate of the dropout is to disconnect the neurons willy-nilly, creating the model additional sturdy and therefore preventing it from over-fitting the coaching set.
- **Output layer and classification***:* The output layer consists of 2 neurons Attack and Benign. Since the 1024 neurons from the previous layer should be regenerate

into simply a pair of neurons, a sigmoid activation operate is employed. thanks to the character of the sigmoid operate, it returns solely 2 outputs, therefore favoring the binary classification that was planned in this paper.

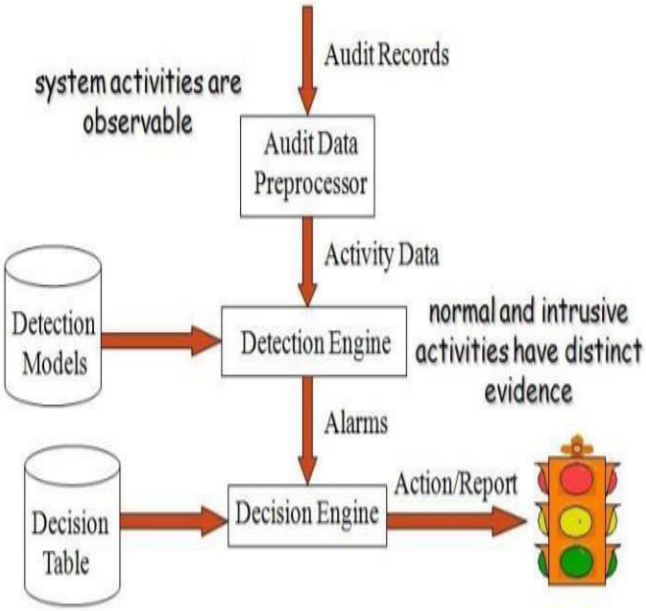
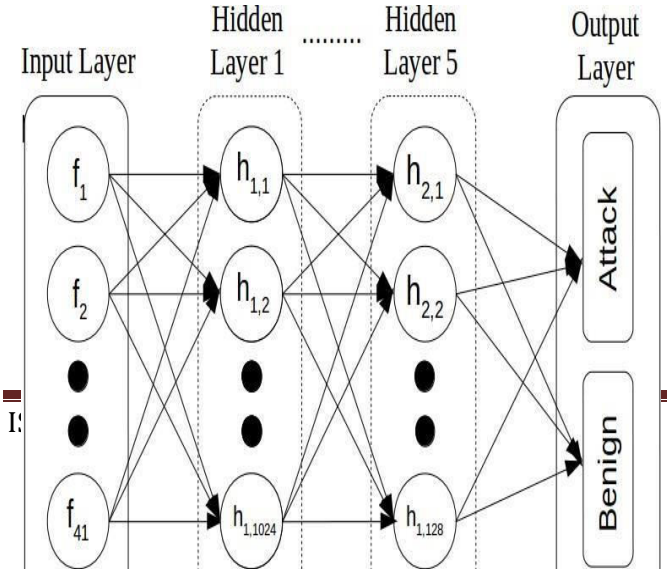


Figure 2: Components of IDS

Figure 3: Architecture Diagram for the proposed system

VI. CONCLUSION

As the importance of IDS within the day-to-day digital world is more obvious, the need for a more accurate model with no warning rate has become apparent. To aim at an algorithm that's more strong to the always-increasing style of malicious acts and threats, researchers have taken an array of paths and have succeeded in achieving exponentially high accuracy rates while benchmarking with the prevailing publicly available dataset. It is clear that, with proper hyper parameter tuning, most of the model existing within the market can do superhuman ability in identifying inbound threats, which leads us to the conjunction of most of the research papers that are made on AI based IDS. Due to the essential principle of deep learning which dictates that more data equals more accuracy, we want to train the models with a dataset that has signatures and features of more modern advanced cyber warfare attacks, so as to organize the research stage IDS into a sturdy market ready product.



ACKNOWLEDGEMENT

We are indebted to our guide Prof Vanishree K., Assistant Professor, Dept. of ISE, R.V College of Engineering for the constant guidelines, suggestion and support throughout the span of the work of this paper.

VII. REFERENCES

1. Hodo, E., Bellekens, X., Hamilton, A., Tachtatzis, C. and Atkinson, R., 2017. Shallow and deep networks intrusion detection system: A

taxonomy and survey. arXiv preprint arXiv:1701.02145.

2. Manzoor, M.A. and Morgan, Y., 2017. Network intrusion detection system using apache storm. Probe, 4107, p.4166.

3. Nassar, M., al Bouna, B., and Malluhi, Q., 2013, June. Secure outsourcing of network flow data analysis. In Big Data (BigData Congress), 2013 IEEE International Congress on (pp. 431-432). IEEE.

4. Suthaharan, S., 2014. Big data classification: Problems and challenges in network intrusion prediction with machine learning. ACM SIGMETRICS Performance Evaluation Review, 41(4), pp.70-73.

5. Meng, F., Fu, Y. and Lou, F., 2018, March. A network threat analysis method combined with kernel PCA and LSTM-RNN. In Advanced Computational Intelligence (ICACI), 2018 Tenth International Conference on (pp. 508-513). IEEE.

6. Staudemeyer, R.C., 2015. Applying long short-term memory recurrent neural networks to intrusion detection. South African Computer Journal, 56(1), pp.136-154.

7. Shone, N., Ngoc, T.N., Phai, V.D. and Shi, Q., 2018. A deep learning approach to network intrusion detection. IEEE Transactions on Emerging Topics in Computational Intelligence, 2(1), pp.41-50.

8. Lee, B., Amaresh, S., Green, C. and Engels, D., 2018. Comparative Study of Deep Learning Models for Network Intrusion Detection. SMU Data Science Review, 1(1), p.8.

9. Mukkamala, S., Janoski, G. and Sung, A., 2002. Intrusion detection using neural networks and support vector machines. In Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on (Vol. 2, pp. 1702-1707). IEEE.

10. Ahmed, U. and Masood, A., 2009, October. Host based intrusion detection using RBF neural networks. In Emerging Technologies, 2009. ICET 2009. International Conference on (pp. 48-51). IEEE.