***Deep Learning Approach for Intelligent***

***Intrusion Detection System***

IT systems Networks are dealing with swarms of exposed sensitive user data Internal and external intruders.

Since computers and networks are assembled and pro-grammed by humans, there are possibilities for bugs in both the hardware and software. These human errors and bugs can lead to vulnerabilities. Confidentiality, data integrity and availability are main pillars of information security. Au-thenticity and accountability are also plays an important rolein information security. Generally attacks against the confi-dentiality addresses passive attacks for example eavesdrop-ping, integrity addresses active attacks for example systemscanning attacks i.e. ’Probe’ and availability addresses the attacks related to making network resources down so thesewill be unavailable for normal users for example denial of service (’DoS’) and distributed denial of service (’DDoS’).IDS systems have limited capability to detect attacks related.

ICT system is considerably more complex, con-nected and involved in generating extremely large volumeof data, typically called as big data. This is primarily due tothe advancement in technologies and rapid deployments oflarge number of applications. Big data is a buzzword which contains techniques to extract important information from large volume of data. Allowing access to big data technology in the domain cyber security particularly IDS is of paramountimportance [44]. The advancement in big data technologyfacilitates to extract various patterns of legitimate and ma-licious activities from large volume of network and system activities data in a timely manner that in turn facilitates to improve the performance of IDS. However, processing of big data by using the conventional technologies is often difficult . The purpose of this section is to describe the comput-ing architecture and the advanced ethods adopted in the proposed framework, such as text representation methods, deep neural networks (DNNs) and the training mechanisms employed in DNNs.

SCALABLE COMPUTING ARCHITECTURE The technologies such as Hadoop Map reduce and Apache Spark in the field of high performance computing is found to be an effective solution to process the big data and to provide timely actions. We have developed scalable framework based on big data techniques, Apache Spark cluster computing platform .

TEXT REPRESENTATION METHODS System calls are essential in any operating system depicting the computer processes and they constitute a humongous amount of unstructured and fragmented texts that a typical HIDS uses to detect intrusions and cyber-attacks. In this research we consider text representation methods to classify the process behaviours using system call trace.

DEEP NEURAL NETWORK (DNN) We employ an artificial neural network (ANN) approach as the computational model since it is influenced by the characteristics of biological neural networks to incorporate intelligence in our proposed method.

PROBLEM FORMULATION, DATASET LIMITATIONS AND STATISTICAL MEASURES A. PROBLEM FORMULATION FOR NIDS Generally, the network traffic data is collected and stored in raw TCP dump format. Later, this data can be preprocessed and converted into connection records DATASET LIMITATIONS Most of the datasets which epresents the current network traffic attacks are private due to privacy and security issues. On the other direction, the datasets which are publicly avail-able are laboriously anonymized and suffer from various issues. In particular they failed to validate that their datasets typically exhibit the real-world network traffic profile. Generally, the units in input to hidden layer and hidden to output layer are fully connected. The DNN is composed of various omponents,a brief description of each component is given below.Fully connected layer: This layer is called as fully con-nected layer since the units in this layer have connection to every other unit in the succeeding layer. Generally, the fully connected layers map the data into high dimensions. The output will be more accurate, when the dimension of data is more. It uses ReLU as the non-linear activation function.

Packet and system call processing module: In this module, the networks which are to be monitored are connected to a single port mirroring switch, which replicates the flow of the entire network traffic of all the switches. The proposed SHIA IDS is required to monitor a network composed of different subnets, each with n number of different machines. The monitored networks are hosts composed of computer machines which allow users to communicate and transfer data. All the traffic generated by the internet were collected, without considering the internal traffic between net-works.The SHIA framework with the network where one ofthe switches is used as a port mirroring switch and connected to a traffic collector. This module collects network traffic data using Netmap packet capturing tool and stores it in NoSQL database. Feature vec-tors are passed into the DNN module, and a copy of data is passed into NoSQL database. Likewise, the system calls and configuration files are collected in a distributed manner by following the methodology provided in . These are passed to NoSQL database and followed by the text representation method to map the system calls to feature vectors.

In this paper, we proposed a hybrid intrusion detection alert system using a highly scalable framework on ommodity hardware server which has the capability to analyze the network and host-level activities. The framework employed distributed deep learning model with DNNs for handling and analyzing very large scale data in real time. The DNN model was chosen by comprehensively evaluating their performance in comparison to classical machine learning classifiers on various benchmark IDS datasets. In addition, we collected host-based and network-based features in real-time and em-ployed the proposed DNN model for detecting attacks and intrusions. In all the cases, we observed that DNNs exceeded in performance when compared to the classical machine

learning classifiers. Our proposed architecture is able to per-form better than previously implemented classical machine learning classifiers in both HIDS and NIDS. To the best of our knowledge this is the only framework which has the capability to collect network-level and host-level activities in a distributed manner using DNNs to detect attack more accurately. The performance of the proposed framework can be further enhanced by adding a module for monitoring the DNS