An Efficient Intrusion Detection System by Using Behaviour Profiling and Statistical Approach Model

Rajagopal Devarajan and Padmanabhan Rao
PG and Research Department of Computer Science and Applications, Vivekanandha College of Arts and Sciences for Women (Autonomous), India

Abstract: *Unauthorized access in a personal computer or single system of a network for tracking the system access or theft the information is called attack/ hacking. An Intrusion detection System defined as an effective security technology, it detect, prevent and possibly react to computer related malicious activities. For protecting computer systems and networks from abuse used mechanism named Intrusion detection system. The aim of the study is to know the possibilities of Intrusion detection and highly efficient and effective prevent technique. Using this model identified the efficient algorithm for intrusion detection Behaviour Profiling Algorithm and to perform dynamic analysis using Statistical Approach model using log file which provides vital information about systems and the activities on them. The proposed algorithm implemented model it produced above 90%, 96% and 98% in the wired, wireless and cloud network respectively. This study concluded that, the efficient algorithm to detect the intrusion is behaviour profiling algorithm, while join with the statistical approach model, it produces efficient result. In further research, possibility to identify which programming technique used to store the activity log into the database. Next identify which algorithm is opt to implement the intrusion detection and prevention system by using big data even the network is wired, wireless or cloud network.*

Keywords: *IDS, IPS, behaviour profiling algorithm, statistical approach model, NIDS, HIDS.*

Received September 12, 2019; accepted May 9, 2020
https://doi.org/10.34028/iajit/18/1/13

1. Introduction

Modern internet services implemented as complex and large scale distributed system [28]. Security has become a most important issue in recent years [3]. Today's tendency in computer security shows an increased amount of work being done in database security research. The reason behind such an increase is because in traditional security mechanism such as the use of firewall is no longer effective in today's database security [31]. The intrusion detection concept has introduced by Anderson in the year 1980 and he defined the 'Intrusion' as unauthorized access information, additional information manipulation, or render the system unusable or unreliable [26]. Mitali *et al*. [12] defined 'Intrusion' as an act of trespassing without any permission and hence resulting in loss and destruction. An Intrusion detection System defined as an effective security technology, which can detect, prevent and possibly react to computer related malicious activities [7]. In Intrusion, Intrude means a person involved into some process or uninvited task [12]. Intruder is a person violates perfectly working computer system with a criminal mindset [12]. There are two types of Intruder named internal and external. Internal Intruder is insider who has entrusted with authorized access to the machine or network. External Intruder is outsider who has unauthorized remote access to machine or network. Hacker/Attacker is a person who is unauthorized access in a PC or single system of a network for trace the system access or theft the information [12]. Unauthorized access in a Personal Computer (PC) or single system of a network for tracking the system access or theft the information is called attack/ hacking. The attack is classified into two types namely passive and active. In passive, attack the attacker monitoring the transformation of information, process of the system without modification. In active attack the attacker stolen the data from the system and modified the files which has system [7]. Recently attacks have been sophisticated affecting different parts and applications as well as several systems simultaneously and are carried out over a period. Detecting and preventing such attacks by analyzing each log individually. Either a log file is a file that records events that occur in an Operating System (OS) or other applications or it is a message between different users of communication software [20]. A log file provides vital information about systems and the activities occurring on them [20]. Evidence of malicious insider activity covered within large data streams, such as system logs accumulated over the months or years. Traditional approach to the

insider threat detection problem is supervised learning [15].

1.1. Objective of the Paper

The contribution of the work has listed as follows:

1. Studied about the efficient intrusion detection and prevention algorithms from previously published articles in reputed journal.
2. Identified the efficient algorithm for intrusion detection "Behaviour profiling algorithm" and to perform dynamic analysis "Statistical approach model".
3. Identified the intrusion detection and prevention have been performed in various types of network like wired, wireless, cloud etc.
4. In this paper concentrated to develop a proposed model to perform intrusion detection in wired network using resource monitor and gathered evaluation metrics like Log files, CPU Usage etc. The gathered data has included identifying the efficiency.
5. To identifying the efficiency the data compared with existing model produced output and finally stated the percentage of efficiency.

2. Literature Survey

Seva *et al*. [26], made the survey on Intrusion Detection (IDS) System to overcome the pitfalls of former IDS like signature, anomaly, and Knowledge based. The aim of their survey is to secure the data, which had the system and others to know the advantages and disadvantages of among the Intrusion Detection System.

For that in their paper, they discussed about Host Based IDS, Network Based IDS, Anomaly Based IDS, Signature Based IDS, Specification Based IDS, Knowledge Based IDS, and Hybrid IDS. Mitali *et al*. [12], studied about Intrusion Detection and Prevention System (IDPS). Through that, they discussed the matters like Intrusion, various techniques and systems using to detect and prevent the intrusion in an internet and they described the main features of several IDPs systems/ Platforms in concise manner. Presented information constitutes to start the research in the field of IDPS. Dan et.al contributed in their paper general-purpose techniques that used for configurable summarization of time series data that scales to high event rates. Additionally scalable infrastructure for normalize and summarize log data for troubleshoots of complex distributed systems. Their anomaly detector had developed using Grid technique and aimed at middleware and applications with consistent performance that includes predictable user computation, large file transfer, visualization services, and streaming data analysis.

Sigelman *et al*. [28], implemented large-scale system for modern internet services. They introduced dapper design, Google's production distributed systems tracing infrastructure. Dapper shares conceptual similarities with other tracing system and it deployed across virtually all of Google's systems and it has allowed largest workloads to trace without need of any other applications. Legg *et al*. [9] described an automated system that capable of detecting insider threats within an organization for that they defined tree structure profiling approach that incorporates the details of activities and provides the description of the users' behaviour. Deviations assessed and founded that the anomalous behaviour. Kumar [7] examined the impact of applying IDPS in cloud environment and tried to how IDPS helps to maintain the resources of cloud with brief explanation about the types of IDPS. The author concluded and discussed existing solutions for intrusion detection in cloud, types of cloud based IDs that includes network based, distribution-based, host based and virtual machine introspection based systems.

Parveen *et al*. [15] achieved highly accurate anomaly detection with the help of Ensemble based stream mining leverages multiple classification models and this stream is unlabeled, unbounded, and evolving. This technique is a supervised learning approach and it combines the power of one class Support Vector Machines (SVMS). Raut [20] developed a Host based Intrusion Detection System (HIDS) with the help of Logs generated files. The Logs monitored and analyzed any suspicious activity. Borkar and Patil [3] proposed a HIDS and it called as Post Attack Intrusion Detection and they investigating the system log files. The system log file contains the log of all system call. The HIDS has two features, i.e., reduces the time to locate log with intruder activities by factoring, a classifier classify the normal behaviour form malicious one. Sequitur method is used to factor the logs, Hidden Markov Model (HMM) is used for classification and k-means is used to classify the behaviour. Swapna and Srivatsav [31] proposed the system for detects the attacks from internet by using web server logs and signatures and they stated that Network Intrusion Detection System (NIDS) could not fully protect the database from the attacks. Bhayani [2] tend to propose the model for detecting the anomaly user using log files supported dynamic rule creation. The outcome of the model conceivable error rates below 15%. Frequent P-growth and apriori algorithm used. It limits the communication that made by human. He compared the implementation, which has developed by java, and Hadoop, Hadoop cluster is 10 times faster.

Rahayu *et al*. [19] proposed the technique on tracing the blaster attack from various logs that has different Open Systems Interconnection (OSI) layers. They intended preliminary investigation upon attacks using signature based technique. Robert *et al*. [21] states that cyber attacks continuously grow daily-basis and Office

Personal Management (OPM) breach those attacks. For suspecting, the anomaly usage break up network events into time segment blocks. Sharma *et al*. [27] proposed to identify anomalies used cognitive tokens. It provides intelligent sensing model for anomaly detection and it caused less error within the range of 0.1 % to 2.8%. Saratkar and Richariya [23] made the survey on genetic based anomaly detection that is compared the tracing data with actual data. Zhang *et al*. [36] states that log files are the main sources for security analysis; it is not user friendly and laborious work need to obtain useful information from log files, concluded that with assistance of visualization system administrators can detect the anomaly users.

Malviya *et al*. [10] proposed to convert unstructured weblogs into structured weblogs using data cleaning to remove noise. From this structured weblog classified the user as normal, suspected and attacker using firewall access lists. Rupam *et al*. [22] shows packet sniffer operations that had provide the service like monitoring and troubleshooting network traffic for using this service can identify the intruder. Corney *et al*. [5] describes the identification of anomaly events and events pattern that have manifest by computer system logs, which reduced the false alerts. Raghavan and Raghavan [18] developed a method to identify the origin of the files that downloaded from internet using metadata association. It reveals that high order relationships such as system and log files. Using log files can identify different browser logs generated during users' online activity.

Wagner *et al*. [34] demonstrates the effectiveness in discovering the anomaly users while using multiple databases. For detecting the attack, they used audit logs, which is a stored executed Structured Query Language (SQL) command. Read only query detection analyses caching behaviour of multiple Database Management System (DBMS). Sipola *et al*. [29] presented detecting the security attacks from network log data. Unsupervised data from diffusion maps have applied. For analyse textual data, log file have used and it transformed into feature space. Principal Component Analysis (PCA) used to extract orthogonal components. Patcha and Park [16] discussed recent technologies available for detecting anomaly access and presented three generic approaches to detect intrusion detection. Best *et al*. [1] used Machine-Independent Audit Trail Analyser (MIATA) system, which is knowledge base capturing rejected users' activity, and it updates user profiles. Wagner *et al*. [34] presented anomaly detection techniques, which had developed by using pattern recognition and statistical approaches, and it can detect the attacks in cloud environment distributed system application. Sari [25] made the comparative study on IDS in cloud networks and concluded that Advanced Encryption Standard (AES) encryption algorithm and 2-step authentication process helps to secure the data in the cloud storage. Virushabadoss and Bhuvaneswari [33] discussed the behaviour profiling algorithm technique. It has used to solve the problems, which generated in fog systems. Using statistical metrics the system identify the rouge node and anonymous behaviour of actions of the node and it is 90% to detecting anomaly users' behaviors. Chaudhary *et al*. [4] proposed a system to process the system logs and generate graphical and tabular format of user activity. Using web mining it has classified for administering the users activity in server.

Lee *et al*. [8] dedicated a work to resolve the journaling of journal anomaly in android Input/Outpu (IO) Stack and developed a model Write Ahead Logging Direct Input/Output (WALDIO). It reduces the IO Volume against WAL mode. Saraydaryan *et al*. [24] presented a framework for monitoring IS. The user behaviour analysis detecting more than 80% legitimate actions of attack. Pore and Bartere [17] reviewed new security threats for mobile devices, various techniques for detecting malware attacks in mobile and concluded that Android OS provides more security for mobile devices and spy camera can play to attack. Mishra *et al*. [11] presents a System Call Analysis approach to detect malware attacks in virtual machines and which applies machine learning. It generalizes the behaviour and improves adaptability and efficiency. Motghare and Nikose [13] presents an approach, which has developed by using hybrid algorithm and it, detect intrusions in the distributed network. Suganya and Kathiresan [30] introduces binding approach based on Improved Binary Black Hole Optimization Algorithm (IBBHOA) in SVM classifier. It used to classify anomaly behaviors, Domain Name System (DNS) failed requests. Yu *et al*. [35] made the survey on social media anomaly detection techniques followed. For analysis activity and graph based approaches used. Time dynamics of the social graph consider by graph based approach.

3. Intrusion Detection

Anomaly is an unexpected lack in behaviour that adversely affects application performance [6]. For protecting computer systems and networks from abuse used an important security mechanism named IDS [26]. IDPS is a Software or Hardware system that has all the capabilities of Intrusion detection and can react effectively in case of possible intrusions [7]. The primary goal of intrusion detection and prevention system is to provide a view of unusual activity happening in the network and to generate alerts notifying administrators or blocking a suspected connection [7]. IDS used as a countermeasure to preserve data integrity [32]. Table 1 depicts the analysis intrusion detection and prevention methods.

3.1. Host based Intrusion Detection System

The events like file accessing and applications used for

execution is examined by HIDS and it analyse the data the originates on computer system such as operating system and application software file attributes and event logs [26].

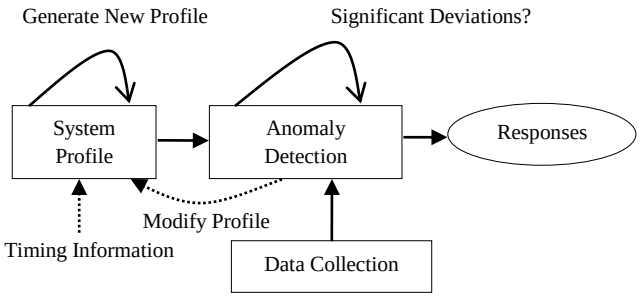

Figure 1. Anomaly based Intrusion detection system.

Host based Intrusion detection system combines log file analysis technology and Back Propagation (BP) neural network technology. Three steps are following for log file analysis i.e., pre-decoding log file, decoding log file and analysis log file. BP neural network technology establish through system behaviour characteristics profile in advance [26].

3.2. Network based Intrusion Detection System

Analyse the packets flow in a network by NIDS. Network intrusions have been carried out in different forms such as Spam, Trojan horse, Viruses, and Worms etc., it detecting long-term attacks such as sniffer programs. NIDS provides wide-ranging defense against information mining, network hacking and theft identification [26].

3.3. Anomaly based Intrusion Detection System

In an Anomaly based Intrusion Detection System (AIDS) normal behaviour of users' activity compared with other defined behaviour. Each similarities has founded is called attack or intrusion. Not every deviation is an intrusion. This may result into false negative [26]. Figure 1 illustrates that the Anomaly based Intrusion Detection System.

Table 1. Analysis of intrusion detection and prevention methods.

Type	Method	Approach	Anomaly Detection/ Prevention	Signature Detection/ Prevention	Advantages	Disadvantages
HIDPS and NIDPS	Signature based	Sequence matching, Malicious matching	No	Yes	Automated response to malicious attacks	Unable to detect and respond to anomaly behaviour
	Signature and anomaly based	Peer to Peer	Yes	Yes	Efficient, Reliable, Trusted	Implementation and Memory Issue
		In and Out Source			Secured	Well trained analysts are required
		Operating system and application level approach			Reduce human effort, Automatic response	Ineffective Cost for implementation, Monitoring issues
		Network Layer to application layer level			Customize flexibility, Cost effective	Well trained analysts are required
HIDPS	Signature and anomaly based	Secure mobile agent	Yes	Yes	Reduce human effort and Real time Response	Need to be adopt other technique, Security of mobile agent
		Operating system and application level approach			Strong IDPS mechanism	Large amount of memory required

3.4. Signature based Intrusion Detection System

Signature based Intrusion Detection System (SIDS) compares the packet information with the known virus signatures and it encapsulates the rules, which is used to detect threats. Snort is SIDS and it is an Open Source Software (OSS) [26].

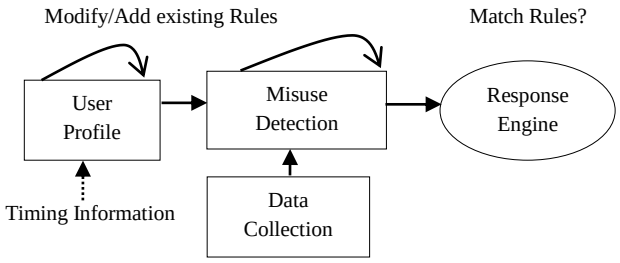

Figure 2. Signature based intrusion detection system.

The concept behind signature detection is that it stores the pattern, signature of attacks and tries to detect abnormal behaviour by analyzing the given traffic and matching several different rules [12]. Figure 2 illustrates that the signature based Intrusion Detection System.

3.5. Specification based Intrusion Detection System

Specification based Intrusion Detection System (SPECIDS) works similarly Anomaly based IDS [26].

3.6. Hybrid Intrusion Detection System

Hybrid Intrusion Detection System (HyIDS) obtained by mingle Packet Header Anomaly Detection (PHAD) and Network Traffic Anomaly Detection (NETAD). NETAD is used to find suspicious packets based on unusual byte values in network packets. PHAD acts as a pre processor and added to Snort [26]. Table 2 depicts the comparison of different intrusion detection methods

Table 2. Comparison of different intrusion detection methods.

Analyzing Techniques	Operation
Pattern Matching	Pattern matching approach implementation is very easy. This approach is used to identify the contents of log files and which is used to find from a sequence of bytes for pattern matching
State-full Pattern Matching	State-full pattern matching method used to identify the pattern for entire data stream.
Protocol Decode-Based Analysis	Protocol Decode Based Analysis approach used to identify the infringement against rules. It has classified by Internet standards and it makes by the state-full pattern matching.
Heuristic Based Analysis	This kind of algorithms based on statistical evaluations of network traffic contents and it makes their decision upon the pre-programmed logic.
Anomaly Detection	Anomaly detection approach is gathering information from previous activity through training with patterns and it assumed as normal, it has tried to anomalous actions.

4. Proposed Model

Through the above survey decided to identify the intrusion, improve the efficiency and accuracy of existing model which had been used for intrusion detection the behaviour profiling algorithm technique have been implemented. Using behaviour profiling algorithm the network administrator have been collect the composite data of the network users behaviour which is in the form of log files.

4.1. Proposed Behaviour Profiling Algorithm:

- *Step* 1: Logs entropy (or Information gain) from Server.
- *Step* 2: Identify the Process and respective Process Identification Number (PID).
- *Step* 3: Identify the Internet Protocol (IP) Address for each Process using PID.
- *Step* 4: Identify the Protocol which is used for transformation Process Transmission Control Protocol (TCP) or User Datagram Protocol (UDP).
- *Step* 5: Identify the Active Time and Response time for each process.
- *Step* 6: Identify the Data Packet transformation and Data Packet Loss.
- *Step* 7: Trace the IP Address and its behaviour of Transformation.
- *Step* 8: Classify the Authorised and Unauthorised Transformation
- *Step* 9: Make the Prevention from the unauthorised transformation.

To implement the proposed algorithm as an expression the following acronyms have been used. The Nodes in the Wired Network is represented as (NWiNet), the Nodes in the Wireless Network is marked as (NWLNet), the Number of Servers in the Wired Network is mentioned as (NSWiNet), the Number of Servers in the Wireless Network is noted as (NSWLNet), the Number of Normal client Nodes in the Wired Network is noticed as (NNNWiNet), the Number of Normal Nodes in the Wireless Network is denoted as (NNNWLNet), the Number of Nodes connected in the Server is represented as (NNS), the Number of Process Running in Nodes is marked as (NRPN), the Number of Process using TCP is mentioned as (NPRTCP), the Number of Process using UDP is represented as (NPRUDP), the Number of Data Packets Transferred is denoted as (NDPTrans), the Number of Data Packets Loosed is marked as (NDPloss), the Authorised Transformation is mentioned as (ATrans) and Unauthorsed Transformation is denoted as (UTrans).

By identifying the number of servers NSWiNet that are connected in the wired network and the number of client nodes NNNWiNet connected in the network which leads to identify the nodes NWiNet are connected in the wired network.

$$\sum NWiNet = \sum NSWiNet + \sum NNNWiNet \quad (1)$$

By identifying the number of servers NSWLNet that are connected in the wireless network and the number of client nodes NNNWLNet connected in the network which lead to identify the nodes NWLNet are connected in the wireless network.

$$\sum NWLNet = \sum NSWLNet + \sum NNNWLNet \quad (2)$$

By identifying the total number of servers NSW(i)LNet are connected in the network and the total number of client nodes are connected in the each server nodes NNNW(i)LNet in the network which leads to identify the total number of nodes NNS are connected in the wired and wireless network.

$$\sum NNS = \sum NSW(i)\,LNet + \sum NNNW(i)LNet \quad (3)$$

By identifying the NPRTCP and the NPRUDP which leads to identify the NRPN which are connected in the server.

$$\sum NRPN = \sum NPRTCP + \sum NPRUDP \quad (4)$$

The NPRTCP can be identified by the difference between the NRPN which are connected in the server and the NPRUDP

$$\sum NPRTCP = \sum NRPN - \sum NPRUDP \quad (5)$$

The NPRUDP can be identified by the difference between the NRPN which are connected in the server and the NPRTCP

$$\sum NPRUDP = \sum NRPN - \sum NPRTCP \quad (6)$$

To identify the number of authorized transaction (∑ATrans) processes running in a personal computer or intranet server along with TCP, first identify the NPRTCP, next identify the total NRPN. The Sum of difference between NPRTCP and NRPN identify by using the IP. To identify the number of authorized transaction (∑ATrans) processes running in a web server along with TCP, first identify the Total Number of Processes Are Running along with TCP (∑NPRTCP), next identify the Total Number of

Process Running in each nodes (∑NRPN). The sum of difference between ∑NPRTCP and ∑*NRPN* identified using each node IP.

$$\sum ATrans = IP \in NPRTCP \sim \sum NRPN \quad \text{(or)} \quad IP \in \sum NPRTCP \sim \sum NRPN \tag{7}$$

To identify the number of authorized transaction (∑*ATrans*) processes running in a personal computer or intranet server along with UDP, first identify the NPRUDP, next identify the total NRPN. The Sum of difference between NPRUDP and NRPN has been identified using the IP. To identify the number of authorized transaction (∑*ATrans*) processes running in a web server along with UDP, first identify the ∑*NPRUDP*, next identify the in each nodes ∑*NRPN*. The sums of difference between ∑*NPRUDP* and ∑*NRPN* identified using each node IP.

$$\sum ATrans = IP \in NPRUDP \sim \sum NRPN \quad \text{(or)} \quad IP \in \sum NPRUDP \sim \sum NRPN \tag{8}$$

To identify the Number of Unauthorized Transaction (∑UTrans) processes running in a personal computer or intranet or web or internet server, first identify which IP is entering and processing its process in each nodes in the network except the Administrator IP. Next analyse the IP's behaviour like the IP downloads and uploads its file to the nodes, it makes the data transaction losses, it generates the deadlock conditions in the network, it traces the system supporting files. If an IP performs one of these activities in the network, the administrator simply locks this IP activity because the IP address is attacker/intruder.

$$\begin{aligned} \sum Utrans &= IP \sim NNS \\ &= IP \sim NWiNet \\ &= IP \sim NWLNet \\ &= IP \sim NSWiNet \\ &= IP \sim NSWLNet \\ &= IP \sim NNNWiNet \\ &= IP \sim NNNWLNet \end{aligned} \tag{9}$$

5. Related Work

5.1. Tracking Log Files

Using resource monitor the network intrusion detection has been identified. Resource monitor consists of 5 different tracking facilities. They are Overview, Central Processing Unit (CPU), Memory, Disk and Network. In this each tracking facility consist of other 4 facilities. Disk facility can used to identify the PID number, which generated by an operating system. It is a unique number that identifies each running process in an operating system. Overview tab contains Disk facility, which the user gathering the information about Read [B/Sec], Write [B/Sec], Total [B/Sec], I/O Priority, and Response time. From the Figure 3, track the PID-1028 and the mentioned PID has executed svchost.exe. The same PID has been tracked using Network tab. The following Figure 4 shows the Disk and Network tabs.

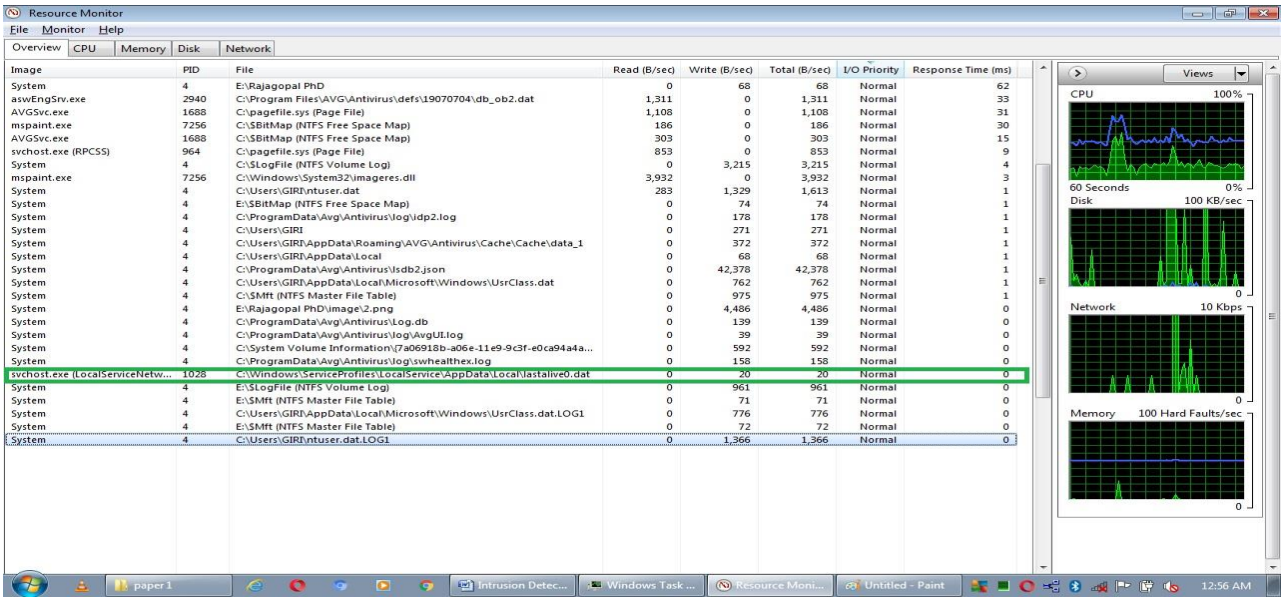

Figure 3. Resource monitor overview tab.

The Network tab has the fields like PID, Address, Send [B/sec], Receive [B/sec], Total [B/sec]. Through the PID and Address fields, the user can identify the process running system. NIDS are easy to secure and difficult to identify the attacker. NIDS required promiscuous network access in order to analyze all unicast traffic.

The difference between NIDS, Network based Intrusion Prevention System (NIPS) is that the monitoring interface is read / write. The read and write operation in a disk or in a network has been monitor through the tab Disk facility. In the network tab it has listening ports tab. It shows in the Figure 5. This tab has six fields like image, PID, Address, port, protocol

and firewall status. The activity tab has the field address. Through the listening ports the user can identify whether the IP address is IPV4 family or IPV6 family, port address, protocol and firewall status can identify.

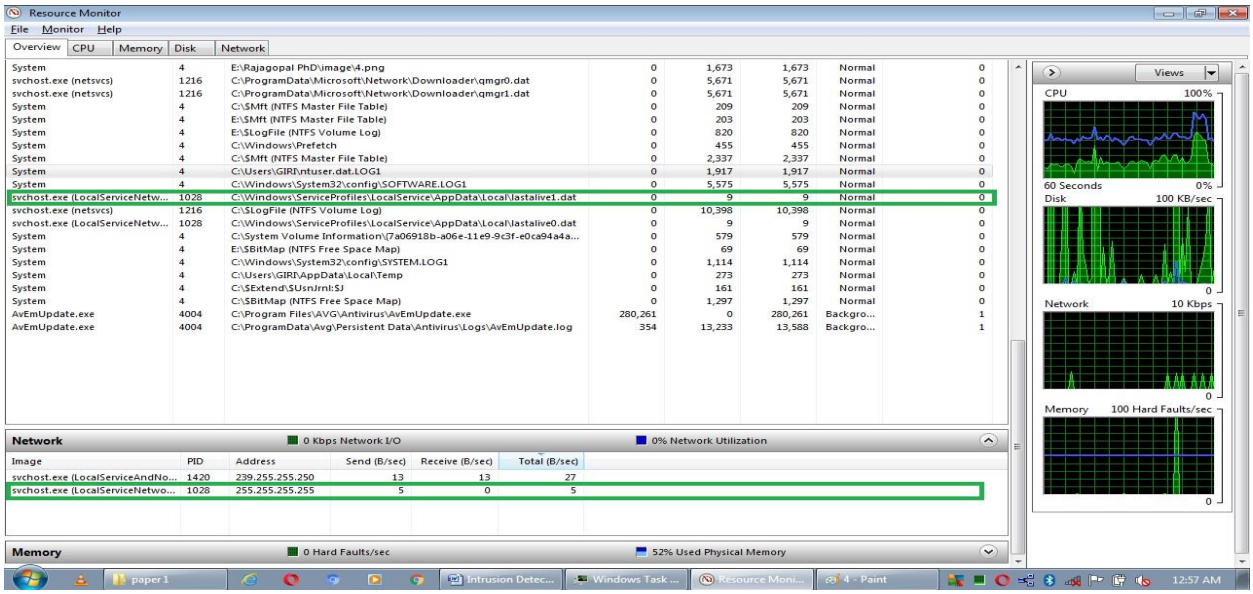

Figure 4. Disk and network tab in the resource monitor.

There are several methods for detecting network traffic and anomalies but one of the most common methods is to checking the network traffic for compliance with different protocol standards like TCP/IP for the underlying traffic and application layer protocols such as Hyper Text Transfer Protocol (HTTP) for web traffic, Simple Mail Transfer Protocol (SMTP) for email and so on.

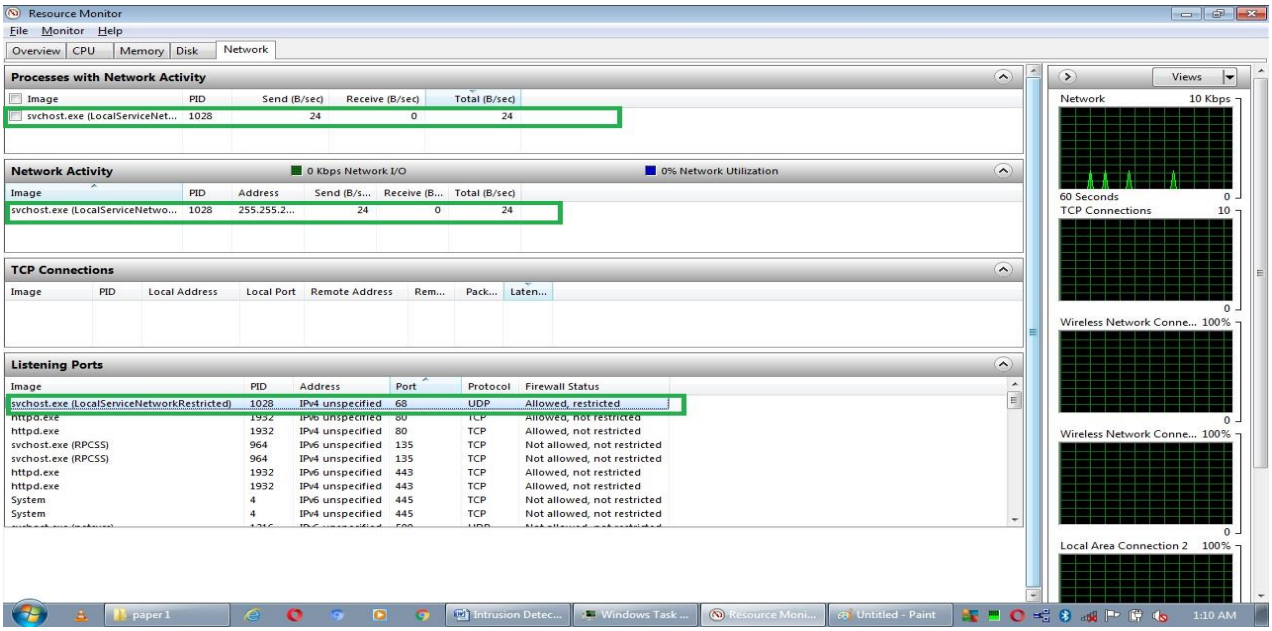

Figure 5. Network tab in resource monitor.

There are entire classes of attacks that do not violate any protocol standard and it will not be detected by the model of anomaly detection. Another method for identify the traffic is to build a model for user behaviour and to generate an alarm when a user deviates from the normal patterns.

6. Experimental Result

In order to implement the behaviour profiling algorithm for identify the intrusion detection 1000 user behaviour log files taken from 1440 data points. The data computation, experiments done in windows 2008 R2 Server along with 40 nodes. In that each node had 1 Gigabyte (GB) Random Access Memory (RAM) and 80 GB Hard Disk Drive (HDD). Every 120 seconds 5 nodes has included in the server to evaluate the existing and proposed model through the metrics like CPU Usage, logs requests in the web server, Weak Data and Synthetic Data. The Table 3 illustrates that

the quantitative metrics evaluation data which has obtained from the existing model and the following Figure 6 shows the time series representation of the dataset.

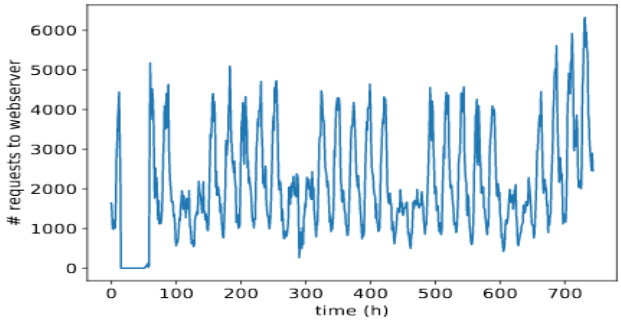

Figure 6. Original logs.

Table 3. Existing model evaluation metrics.

No of Nodes	Logs requests in Web Server	CPU Usage (%)	Weak Data	Synthetic Data
5	1286	172	301	985
10	2579	198	538	2041
15	4048	254	798	3250
20	6580	272	1763	4817
25	8781	287	1992	6789
30	10521	346	2089	8432
35	12482	380	1640	10842
40	16754	421	1989	14765

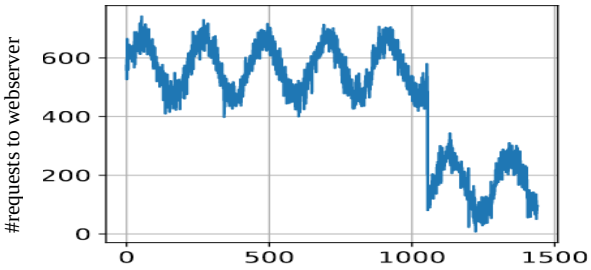

Figure 7. Synthetic week data.

To find the performance and Efficiency of the Existing and proposed model (ExE) identified by using the model weak data and Synthetic data and find the Percentage of the Existing model (PExE) by using ExE

$$ExE = \frac{\varepsilon\, Weak\ Data}{\varepsilon\, Synthetic\ Data} * 100 \quad (10)$$

$$PExE = 100 - ExE \quad (11)$$

Through the Table 3 founded that the existing model produces above 75% accuracy. The above Figure 7 shows what the generated data looks like for a week window. The Table 4 illustrates that the quantitative metrics evaluation data which has obtained from the Proposed model. Through that table founded that the proposed model produces above 90% accuracy.

For analysis used the CPU and Memory usage. The above Figure 8 shows a sample of CPU usage for a user in the trace. The above Figure 9 shows what the generated data looks like for CPU usage of different number of application runs.

Table 4. Proposed model evaluation metrics.

No of Nodes	Logs requests in Web Server	CPU Usage (%)	Weak Data	Synthetic Data
5	1321	128	122	1199
10	2647	57	178	2469
15	3948	132	320	3628
20	6893	157	492	6401
25	9012	138	581	8431
30	11142	198	648	10497
35	12987	212	706	12281
40	17358	230	912	16446

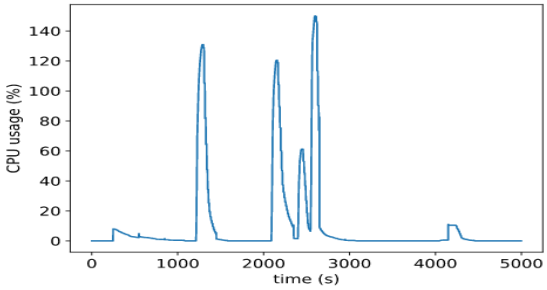

Figure 8. Original data.

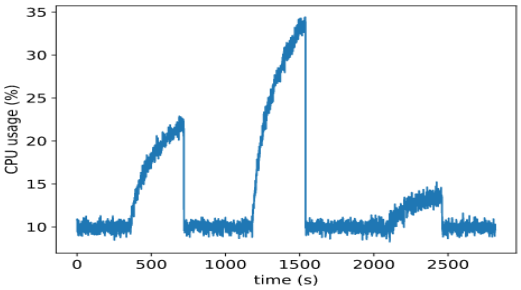

Figure 9. Synthetic data.

7. Conclusions

From the study identified that, using resource monitor the user can identify the intrusion and the process has taken the bits for sending and receiving. According to the study, to get accurate and efficient result the user has to store these data into a database. The resource monitor generates multiple records during short period. To analyse and handle those data, use the big data even it is wired or wireless network. It supports cloud network too. From the existing model while using 5 nodes in the wired network 1286 log requests were received in the web server, in that 301, 985 weak and synthetic data identified respectively with the CPU usage of 172%. In order to using proposed system 1321 log requests were received in the web server, in that 122, 1199 weak and synthetic data identified respectively with the CPU usage of 128%. While using 10 nodes 2579, 2647 log requests received from the web server in the existing and proposed model respectively. In existing model 538, 2041 weak and synthetic data identified respectively with the CPU usage of 198%. In proposed model 178, 2469 weak and synthetic data identified respectively. While using 25 nodes in the wired network in the existing model totally 8781 log requests received from the web server with 1992 weak data and 6789 synthetic data. In the

proposed model 9012 log requests received from the web server. In that 581 and 8431 requests are weak and synthetic data respectively. CPU usage of the proposed system is 138%.

While using 40 nodes in the wired network in the existing model totally 16754 log requests received from the web server with 1989 weak data and 14765 synthetic data. In the proposed model 17358 log requests received from the web server. In that 912 and 16446 requests are weak and synthetic data respectively. CPU usage of the proposed system is 230%. Saraydaryan *et al*. [24], framework for monitoring IS and which has the user behaviour analysis detecting more than 80% legitimate actions of attack. Virushabadoss *et al*. [33], using behaviour profiling algorithm and statistical metrics in the system identify the rouge node and anonymous behaviour of actions of the node and it is 90% to detecting anomaly users' behaviors. The conclusion of the paper is that the efficient algorithm to detect the intrusion is behaviour profiling algorithm, the algorithm while join with the statistical approach model, it produces above 90% of the efficiency in the wired network, above 96% of efficiency in the wireless network and above 98% of efficiency in the cloud network. In further research, possibility to identify which programming technique used to store the activity log into the database, identify the performance analysis of algorithm which is opt to implement the intrusion detection and prevention system by using big data even the network is wired or wireless or cloud network.

References

[1] Best J., Mohay G., and Anderson A., "Machine-Independent Audit Trail Analysis-A Decision Support Tool for Continuous Audit Assurance," *International Journal of Intelligent Systems in Accounting, Finance and Management*, vol. 12, no. 2, pp. 85-102, 2004.

[2] Bhayani D., "Identification of Security Breaches in Log Records using Data Mining Techniques," *International Journal of Pure and Applied Mathematics*, vol. 119, no. 15, pp. 743-756, 2018.

[3] Borkar B. and Patil A., "Post Attack Detection Using Log Files Analysis," *International Journal of Innovative Research in Science, Engineering and Technology*, vol. 2, no. 4, pp. 1195-1199, 2013.

[4] Chaudhary P., Ghuge J., Phalke S., and Nirjal S., "Web Log Pre-processing for Web usage Mining," *International Journal for Scientific Research and Development*, vol. 2, no. 12, pp. 604-606, 2015.

[5] Corney M., Mohay G., and Clark A., "Detection of Anomalies from User Profiles Generated from System Logs," *in Proceedings of the 9th Australasian Information Security Conference*, Australia, pp. 23-32, 2011.

[6] Gunter D., Tierney B., Brown A., Swany D., Bresnahan J., and Schopf J., "Log Summarization and Anomaly Detection for Troubleshooting Distributed Systems," *in Proceedings of 8th IEEE/ACM International Conference on Grid Computing*, Austin, pp. 19-21, 2007.

[7] Kumar K., "Intrusion Detection and Prevention System in Enhancing Security of Cloud Environment," *International Journal of Advanced Research in Computer Engineering and Technology*, vol. 6, no. 8, pp. 1138-1152, 2017.

[8] Lee W., Lee K., and Son H., "WALDIO: Eliminating the Filesystem Journaling in Resolving the Journaling of Journal Anomaly," *in Proceedings of the USENIX Annual Technical Conference*, Santa Clara, PP. 235-247, 2015.

[9] Legg P., Buckley O., Goldsmith M., and Creese S., "Automated Insider Threat Detection System using User and Role-based Profile Assessment," *IEEE Systems Journal*, vol. 11, no. 2, pp. 503-512, 2017.

[10] Malviya M., Jain A., and Gupta N., "Improving Security by Predicting Anomaly User through Web Mining: A Review," *International Journal of Advances in Engineering and Technology*, vol. 1, no. 2, pp. 28-32, 2011.

[11] Mishra P., Pilli E., Varadharajan V., and Tupakula U., "Securing Virtual Machines from Anomalies using Program-Behavior Analysis in cloud Environment," *in Proceedings of IEEE 18th International Conference on High Performance Computing and Communication*s, Sydney, pp. 991-998, 2016.

[12] Mittal M., Khan A., and Agrawal C., "A Study of Different Intrusion Detection and Prevent System," *International Journal of Scientific and Engineering Research*, vol. 3, no. 8, pp. 1526-1531, 2013.

[13] Motghare A. and Nikose A., "A Survey to Track Intrusion Detection in the System by using Data Mining," *International Research Journal of Engineering and Technology*, vol. 6, no. 1, pp. 1583-1586, 2019.

[14] Oppermann A., Toro F., Thiel F., and Seifert J., "Anomaly Detection Approaches for Secure Cloud Reference Architectures in Legal Metrology," *in Proceedings of the 8th International Conference on Cloud Computing and Services Science*, Funchal, pp. 549-556, 2018.

[15] Parveen P., Mcdaniel N., Weger Z., Evans J., Thuraisingham B., Hamlen K., and Khan L., "Evolving Insider Threat Detection Stream Mining Perspective," *International Journal of*

Artificial Intelligence Tools, vol. 22, no. 5, pp. 1-24, 2013.

[16] Patcha A. and Park J., "An Overview of Anomaly Detection Techniques: Existing Solutions and Latest Technological Trends," *Computer Networks*, vol. 51, no. 12, pp. 3448-3470, 2007.

[17] Pore A. and Bartere M., "A Review on Camera Based Attacks on Android Smart Phones," *International Journal of Computer Science*, vol. 6, no. 1, pp. 88-92, 2015.

[18] Raghavan S. and Raghavan S., "Determining the Origin of Downloaded files Using Metadata Associations," *Journal of Communications*, vol. 8, no. 12, pp. 902-910, 2013.

[19] Rahayu S., Robiah Y., Sahib S., Abdollah M., Masud Z., and Roslan I., "Tracing Technique for Blaster Attack," *International Journal of Computer Science and Information Security*, vol. 4, no. 1, pp. 1-8, 2009.

[20] Raut U., "Log Based Intrusion Detection System," *IOSR Journal of Computer Engineering*, vol. 20, no. 5, pp. 15-22, 2018.

[21] Robert J., Bradley G., Boehmke, Bauer K., Saie C., and Bihl T., "Anomaly Detection: Implementation of Augmented Network Log Anomaly Detection Procedures," *The R Journal, Contributed Research Article*, vol. 9-2, pp. 354-365, 2017.

[22] Rupam., Verma A., and Singh A., "An Approach to Detect Packets Using Packet Sniffing," *International Journal of Computer Science and Engineering Survey*, vol. 4, no. 3, pp. 21-33, 2013.

[23] Saratkar K. and Richariya P., "Classification and Genetic based Anomaly Detection," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 6, no. 2, pp. 504-507, 2016.

[24] Saraydaryan J., Fatiha B., Ubeda S., and Legrand V., "Comprehensive Security Framework for Global Threats Analysis," *International Journal of Computer Science Issues*, vol. 2, no. 1, pp. 18-32, 2009.

[25] Sari A., "A Review of Anomaly Detection Systems in Cloud Networks and Survey of Cloud Security Measures in Cloud Storage Applications," *Journal of Information Security*, vol. 6, no. 2, pp.142-154, 2015.

[26] Seva N., Budhwani P., Talekar S., Borle S., and Jadhav N., "Survey on Intrusion Detection System," *International Journal of Advanced Research in Computer Science and Management Studies*, vol. 2, no. 1, pp. 101-109, 2014.

[27] Sharma V., You I., and Kumar R., "ISMA: Intelligent Sensing Model for Anomalies Detection in Cross Platform OSNs with a Case Study on IoT," *IEEE Access*, vol. 5, pp. 3284-3301, 2017.

[28] Sigelman B., Barroso L., Burrows M., Stephenson P., Plakal M., Beaver D., Jaspan S., and Shanbhag C., "Dapper, a Large-Scale Distributed Systems Tracing Infrastructure," Google Technical Report dapper, 2010.

[29] Sipola T., Juvonen A., and Lehtonen J., "Dimensionality Reduction Framework for Detecting Anomalies from Network Logs," *Engineering Intelligent Systems*, vol. 20, no. 1, pp. 87-97, 2012.

[30] Suganya S. and Kathiresan V., "Anomaly Detection in DNS Query Logs using Improved Binary Black Hole Optimization Algorithm," *International Journal of Engineering and Technology*, vol. 9, no. 4, pp. 3058-3065, 2017.

[31] Swapna G. and Srivatsav R., "Securing Web Applications by Analyzing the Logs of the Database Server or Web Server," *International Journal of Engineering Research and Applications*, vol. 2, no. 6, pp. 432-435, 2012.

[32] Tabash M., Abd Allah M., and Tawfik B., "Intrusion Detection Model Using Naïve Bayes and Deep Learning Technique," *The International Arab Journal of Information Technology*, vol. 17, no. 2, pp. 215-224, 2020.

[33] Virushabadoss S, Bhuvaneswari C, "Analysis of Behavior Profiling Algorithm to Detect Usage Anomalies in Fog Computing," *International Journal of Engineering Science Invention*, pp. 14-19, 2018.

[34] Wagner J., Rasin A., Glavic B., Heart K., Furst J., Bressan L., and Grier J., "Carving Database Storage to Detect and Trace Security Breaches," *Digital Investigation*, vol. 22, pp. 127-136, 2017.

[35] Yu R., Qiu H., Wen Z., Lin C., and Liu Y., "A Survey on Social Media Anomaly Detection," *SIGKDD Explorations Newsletter*, vol. 18, no. 1, pp. 1-14, 2016.

[36] Zhang Z., Xiao Y., Chen M., Zhang J., and Deng H., "A Survey of Security Visualization for Computer Network Logs," *Security and Communication Networks*, vol. 5, PP. 404-421, 2011.

Rajagopal Devarajan completed his Bachelor of Computer Science degree and completed his Master of Computer Applications degree in Periyar University in the year 2003 and 2006 respectively. He has completed his Master of Philosophy in PRIST University in the year 2012. He has 3 Years and 7 Months Experience in the field of Software Development and 9 Years 2 months Experience in Teaching. He has published 19 articles in different International Journals and 2 books. He presented 2 papers in international conference and 2 papers in national conference. He participated in International seminar and a national seminar. His article published as a chapter in "Cognitive Science and Technology" by Ella Hunter. One of his articles has placed newly opened science library in Konkuk University, South Korea. Currently he is working as an Assistant Professor in the PG and Research Department of Computer Science and Computer Applications, Vivekanandha College of Arts and Sciences for women (Autonomous), Tiruchengode, Namakkal DT, India. He is pursuing his Ph.D degree in Periyar University. His areas of interest are Computer Networks, Data Structures and Algorithms, Programming Languages. He is the life time member of ISTE and IAENG.

Padmanabhan Rao completed Master of Computer Applications degree in the year 1998 and obtained Ph.D in the year 2015. He joined as the faculty in the department of computer science in the year 1999. His areas of interest are Sensor Networks, and Distributed Network. He is member of IAENG and IACSIT.