Performance Evaluation of Cyber Criminal Detection Techniques

K. Veena [1], K. Meena [2]

[1] Dhanalakshmi College of Engineering
[1, 2] *Department of Computer Science and Engineering*
Veltech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology
Chennai, India
[1] veenakanagaraj07@gmail.com, [2] meen.nandhu@gmail.com

***Abstract* – Computer technology has advanced to a greater extent which leads to increase in cyber crime committed in recent years. The detection of cyber crime is not an easy task. From the literature, many researchers used various technologies to detect the cyber crime. In this paper, performance evaluation of various techniques are analysed to determine the cyber criminal. Firstly the detection of synthetic identity theft is checked. Secondly, the intrusion detection is checked using the honey pot security mechanism. Thirdly, the detection is further strengthened using the lie detection technique where the false speech of a person is determined. Finally by analysing the user profile, the detection of cyber crime is done using the clustering techniques. Synthetic Identity Theft method performs better than the remaining methods when considered for evaluation. Experimental results show that comparison of the final list of criminal users and the list of criminals determined, the number of genuine users eliminated are 41 out of 100 users, where as the number of genuine users eliminated from other methods are 16, 36 and 38 only. The number of attributes used is only 4, where as the number of attributes used for other methods are 5, 10 and 25. The percentage of performance metrics is also 37.1 and gradient is 31.1 which are better compared to other methods considered for performance analysis.**

Keywords - Cyber crime, lie detection, synthetic identity theft, neural networks, cluster, crime data, genuine data

I. INTRODUCTION

The growth of the internet and the use of digital data has increased exponentially in the past few years, so also has the growth of cyber crime. The data of every person accessing the internet is being collected,[1] which includes text messages, chat, images, places visited, frequency of accessing the internet, communications links, date and time of access, etc. The data stored can be easily used for destructive purposes such as committing cyber crimes. There are various types of crimes committed. At least one cyber crime is estimated to be committed in India every 10 minutes [2] in the first six months of 2017, which was only 12 minutes in the year 2016. According to the Indian Computer Emergency Team (CERT-IN), 27,482 cases have been reported from January to June 2016. Hence, it is vital to detect and prevent the cyber crime.

Cyber crime is an act that deals with computer and networking, in which the people are a victim. It is a crime wherein the mental torture of a person is done resulting with loss of property, name, identity, self respect and sometimes even leading to the death of the victim either by murdering or inducing the victim to committing suicide. The modern technology has given rise to the greater usage of internet. Since internet is vast, the numbers of users are millions and hence the detection of cyber crime is not a easy task. [3].Hence various technical methods are needed to detect the cyber crime. Thus a need to detect the cyber crime has arisen.[4]

Organization of the Paper: The paper is organized as follows. Section 1 includes introduction and previous work. Section 2, the methodology used is given which consists of description of all methods which is considered for experimental evaluation and algorithm used for experimental analysis. Implementation and experimental analysis are given in Section3. Finally, the conclusion , acknowledgement and references are given.

II. LITERATURE REVIEW AND WEAKNESSES OF CURRENT SYSTEMS

Nazura Abdul Manapet al [2015][5], the identity thief uses the data of another person to commit a crime. AtefehTajpour et al [2013][6] [7], it provides a comprehensive description of the identity theft and the different methods of crime. They concluded that a person should be educated and be aware of such crimes so that he will not become a victim of cyber crime. The algorithm also proved that the lack of awareness leads to the cyber crime. Identity crime uses personal information for committing the crime. E. Vasilomanolakis et al [2015],[8] the paper determines the relevant requirements for the collaborative intrusion detection system. The attacks that evade the collaborative intrusion detection system and attacks on the availability of the collaborative intrusion detection system are discussed.

Akshay et al [2016],[9] observed that Intrusion Detection System (IDS) monitors the entire network and all the system activities for malicious activities and reports

to a Management station. There is a high rate of false alarms in the current intrusion detection systems. It can be solved by using the honeypots to increase the security and the reliability of the network. Honeypots capture the required information and are also easy to use. They are mainly used by the corporate companies to secure the networks. The development and use of a handset detector and score normalization to greatly improve verification performance is described and discussed. Finally, representative performance benchmarks and system behaviour experiments on NIST SRE corpora are also presented (Reynolds, [2016]) [10],

Bhuvaneswari and J. Satheesh Kumar, [2015] [11], used various methods for lie or deception analysis. They explained that the influence of statistical features that discriminates thinking patterns from normal signal. Roshni et. al [2015] [12] examined that deception detection has legal and also medical applications. Using a classifier two classes of feature was extracted and the efficiency achieved was 83%.Sanjana et al [2016],[13] focused on the study of different clustering algorithms which highlights the characteristics of the big data. Various clustering algorithms are grouped under partitioning algorithm, hierarchical algorithm, density algorithm and grid based algorithm.

Jyh-Jian Sheu[2017],[14] used uncomplicated decision tree data mining algorithm to find the association rules about the pornographic and medical web pages. From the experiments it is proved Reynolds [2016] [10], as the efficacy assessment indexes reached a satisfactory value. Therefore the proposed method has good performance and effectiveness. Prakash Singh et al [2015], [15] utilized software tools to perform the analysis. Various algorithms and techniques used for cluster analysis are done. Clustering algorithms in terms of their execution time, number of iterations, sum of squared error and log likelihood are analysed and studied and based on the results obtained efficiency is calculated.

III. THE PROPOSED TECHNIQUE AND METHODOLOGY USED

A. Novelty of Work

In this paper, various cyber criminal detection techniques are used to identify the criminal. Hence the criminal is being identified easily and effectively. Many of the researchers used one technique to identify the criminal. Also a mere analysis is done to identify the criminal. In this paper, four techniques are used, using various attributes as per the technique requirement. Hence a criminal cannot go unseen. In each technique the users are graded depending on their activities. The genuine users are eliminated in each technique. Hence a genuine user is not punished. The criminal users are going through various process such as method 1, method 2, method 3, analysis of true negative and false positive, clustering methods and finally classification. If the criminal cannot be identified, another set of 25 attributes are used to determine the criminal. Hence in this method the criminal can be determined easily. He might have left any proof, in any of his activities. Hence a small clue is enough to determine the criminal.

B. Synthetic Identity Theft Detection

In this method, Synthetic identity theft is first determined. Synthetic identity theft is a crime that uses the identity of another person and acting like the other person. It also combines the real data with unreal data, thereby creating a new identity. Such crimes are committed to get a passport, credit /debit card and to withdraw money. The various attributes used here are theft, identity theft, cybertheft and computer related offences to analyze the crime. In this method, Input data set(A), Normal data set(B) and Target data set(C) are taken. [16]The input attributes are classified in the range 0-10. If the user is genuine, then his percentage of genuine data is updated else his percentage of Criminal data is updated. The data for the research is taken from http://kdd.ics.uci.edu/ and http://www.kdnuggets.com/datasets.

C. Intrusion Detection System

After the user is analysed with the first method, the second method Intrusion Detection System (IDS) is implemented. The honeypot security mechanism is used here that identifies the incoming data / traffic from the clients and the data / traffic coming from the attackers. In this method Load Balancer is designed and implemented. The Balancer discovers the attack on the server while it is forwarding the request and directs it to a different/alternate server called Honey-Pot. The attributes used to determine the crime here are Malicious Software, Number of times the proxy server is used, Malicious Code Presence, Password Violations and Data Forwarding to determine the crime. If the user /data is genuine, then his percentage of Genuine data is updated else his percentage of Criminal data is updated.

D. Lie Detection System

In this method, lie detection system is used to identify the false speech of a human being. Pre processing is used to assist in the reduction of noise and the plotting of the original artefact EEG signals. This method focuses mainly on the neural network used in the recognition phase and feature extraction technique carried out by the Mel Frequency Cepstrum Coefficients (MFCC). In this method the attributes used are child soliciting, abuse, assault by threat, advertising through the internet, soliciting harlotry through the internet, drug sales, excess privileges, publication irrelevant content, transmission of obscene

content and sexually explicit content. After determination of the lie, if the user is genuine, then his percentage of genuine data is updated else his percentage of Criminal data is updated.

E. Analysing User Profile using Clustering Technique

After the three methods are executed, then the analysis is done using the cluster technique, using the attributes set 1. The attribute set1 is given in table I.

TABLE I: ATTRIBUTE SET 1

Attributes Set 1 [17]				
1. Hacking	2. Theft	3. Cyber Stalking	4. Identity Theft	5. Malicious Software
6. Child soliciting	7. Abuse	8. Assault by Threat	9. Child Pornography	10. Cyber illegal imports
11. Cyber laundering	12. Cyber terrorism	13. Cybertheft	14. Advertising through the internet	15. Solicitingharlotry through the internet
16. Drug sales	17. Number of times the proxy server is used	18. Malicious Code Presence	19. Password Violations	20. Excess Privileges
21. Data Forwarding	22. Computer related offences	23. Publication irrelevant content	24. Transmission of obscene content	25. Sexually explicit content

In this method clustering technique is used to determine the percentage of genuineness or criminal. After analyses, the criminal is detected.

Algorithm Used

Step 1: Start
Step 2: Input User data
Step 3: Implement method 1 with user data,
If (crime data) update the percentage of criminal data
Else update the percentage of genuine data.
Step 4: Repeat step 3 for method 2 and method 3
Step 5: Implement method 4 after the user profile percentage is updated.
Step 6: Determine the criminal.
Step 7: Stop

F. Justification of Attributes used in Each Method

In synthetic identity theft, the attributes such as theft, identity theft, cyber theft and computer related offences are used to analyze the crime. These attributes are used to determine the synthetic identity theft, the theft record of criminals which are required to analyze the synthetic identity. Identity theft is also used as attribute because we need to determine if any previous identity theft was committed by the person. Cyber theft is also used here because, it is a activity that is used to steal a computer. This activity is also related to breaking and entering the DNS cache poisoning, embezzlement and unlawful appropriation, espionage, identity theft, fraud, malicious hacking, plagiarism and piracy.

In Intrusion detection System, the attributes used to determine the crime are malicious software, number of times the proxy server used, malicious code presence, password violations and data forwarding to determine the crime. In this method Malicious software is used as attribute because the user is checked if he has gained access to the system using the malicious software. The attribute Number of times the proxy server is used as a attribute because if one or more Internet sites are frequently requested, then these are likely to be in the proxy's cache. The attribute Malicious Code Presence is used because the spread of malicious codes is achieved by the built-in ability to self-replicate through the Internet and Computer media. Since most legitimate codes do not self-replicate and the number of ways to achieve self-replication is limited to the order of fifty, the detection of malicious codes could be reduced to the detection of the "gene of self-replication" in the code in question. The attribute Password Violations is used as attribute here because**,** some governments have national authentication frameworks that define requirements for user authentication to government services, including requirements for passwords. The attribute Data Forwarding is used as a attribute because a data hazard can lead to a pipeline stall when the current operation has to wait for the results of an earlier operation which has not yet finished.

In the Lie detection Method, attributes considered for experimentation are Child soliciting, Abuse, Assault by Threat, Advertising through the internet, Soliciting harlotry through the internet, Drug sales, Excess Privileges, Publication irrelevant content, Transmission of obscene content and Sexually explicit content. In this method the attribute Child soliciting and Abuse is used as a attribute because in this type of crime the criminals seek minor children via chat rooms for the purpose of child pornography. The attribute Assault by Threat is used because in this kind of threat the criminal threatens the life of the victim or his family or his organization through the phone, email or videos. The attribute Advertising or soliciting prostitution through the internet is used as a attribute because it is against the law to access prostitution through the internet (including in the state of Nevada in the United States) because the process of accessing the internet

crosses state and sometimes national borders.The attribute Drug sales is used as a attribute because only a customer through a state-licensed pharmacy based in the United States can sale the drugs. Both illegal and prescription drug sales through the internet are illegal. The attribute Excess Privilege is used as a attribute because the Privilege is defined as the allocation of authority over a computer system in computing.

IV. IMPLEMENTATION AND RESULTS

A. Results of Synthetic Identity Theft Detection

In this experiment, first the synthetic identity theft is determined. It is analysed with 100 users/suspects. The experiment is repeated for 10 victims. For each user data, 25 attributes are taken. In this method, Synthetic identity theft is first determined. Synthetic identity theft is a crime that uses the identity of another person and acting like the other person. It also combines the real data with unreal data, thereby creating a new identity. Such crimes are committed to get a passport, credit /debit card and to withdraw money. The synthetic identity theft of a person is determined using the various attributes of the user. In this method, Input data set(A), Normal data set(B) and Target data set(C) are taken. [16], [18]The input attributes are classified in the range 0-10. If the user is genuine, then his percentage of genuine data is updated else his percentage of Criminal data is updated.

B. Results of Intrusion Detection System

After the user is analysed with synthetic identity theft detection method, Intrusion Detection System (IDS) is implemented and checked. The honeypot security mechanism is used that identifies the incoming data / traffic from the clients and the data / traffic coming from the attackers. In this method Load Balancer is designed and implemented. The Balancer discovers the attack on the server while it is forwarding the request and directs it to a different/alternate server called Honey-Pot. If the user /data is genuine, then his percentage of genuine data is updated else his percentage of Criminal data is updated.

C. Results of Lie Detection System

In the third method lie detection system is used that is used to identify the false speech of a human being. In this method the pre processing is used to assist in the reduction of noise and the plotting of the original artefact EEG signals. This method focuses mainly on the Neural network used in the recognition phase and Feature Extraction Technique carried out by the MFCC- Mel Frequency Cepstrum Coefficients. After determination of the lie, if the user is genuine, then his percentage of genuine data is updated else his percentage of Criminal data is updated.

D. Results for Analysing User Profile Using Clustering Technique

After the three methods are executed, then the user is passed through the analyses of the profile to determine the cyber criminal. In this method various clustering techniques are used to determine the percentage of genuineness or criminal. The user data is analysed using two set of attributes. After analyses, the criminal is detected. For the clustering technique, the attributes used are, Hacking, Theft, Cyber Stalking, Identity Theft, Malicious Software, Child soliciting, Abuse, Assault by Threat, Child Pornography, Cyber illegal imports, Cyber laundering, Cyber terrorism, Cyber theft, Advertising through the internet, Soliciting harlotry through the internet, Drug sales , Number of times the proxy server is used, Malicious Code Presence, Password Violations, Excess Privileges, Data Forwarding, Computer related offences, Publication irrelevant content, Transmission of obscene content and Sexually explicit content. The number of users used are 100 with 25 attributes. The percentage of trueness of attribute of a particular user varies from 0 to 9. After the analysis the user is determined as crime data or genuine data. The input from the synthetic identity theft, intrusion detection system and lie detection technique and the cluster analyses is given as the input. If the result is greater then 2, then the user is crime data otherwise it is genuine data. The crime users after the analysis are 1,8,9,12-15, 17-18, 21-24, 26, 30-31, 34-37, 40, 43-46, 48-49, 51-61, 64-67, 70, 71, 73-75, 77-78, 80, 82-83, 86, 88-89, 91, 92, 95, 97-99 and the total number of crime users are 61. (Before analysis of true negative and false positive).

TABLE II: ANALYSIS OF VARIOUS METHODS FOR 100 USERS

Analysis of various methods for 100 users								
Technique Used	Attributes Set 1		Crime data	Total Number Of Crime users	Total Number of Genuine users	Final Crime users (61)	Criminal User after TN and FP determination (45)	Final Criminal
Method 1	2	Theft	2,9,12,13,14,15,18, 21,22,23,24,26,28-31,34-37,40,43-46,51-55,57-61,54-67,70-71,74-78,83,86-89,92,95-100	59	41	1,8,9,12-15,17-18,21-24,26,30-31,34-37,40,43-46,48-49,51-61,64-67,70,71,73-75,77-78,80,82-83,86,88-89,91-92,95,97-99	2,5,10,12,13,15,19, 21,22,24,27,32,34,35, 37,41,43,44,46, 48,51,52,53,54,55, 57,58,60, 61,62,64,65,67,70, 71,74,75,78,83,86, 89, 92,95,98,99	32 41
	4	Identity theft						
	13	Cyber theft						
	22	Computer-related offences						
Method 2	5	Malicious Software	1,3,5-6,8-15,17-25,29-37,39-52,54,56,58-67,69-71,73-78,80,82-89,91-100	84	16			
	17	Number of times the proxy server is used						
	18	Malicious Code Presence						
	19	Password Violations						
	21	Data Forwarding						
Method 3	6	Child soliciting	1,3,7-9,11,14-18,20,23-28,36-38,40,45-46,48-53,55-61,63-64,66-68,70-74,77-83.85,88-92,84,97-98	64	36			
	7	Abuse						
	8	Assault by Threat						
	14	Advertising through the internet						
	15	Soliciting harlotry through the internet						
	16	Drug sales						
	20	Excess Privileges						
	23	Publication irrelevant content						
	24	Transmission of obscene content						
	25	Sexually explicit content						
Method 4	Attributes 1 to 25 are used		1, 2,8-9,12-15,17-18,21-24,26-27,30-31,34-37,40,43-46,48-49,51-62,64-67,69-71,73-75,78,80,82,83,86,89, 91-92,95,98-99	62	38			

From Table II, the various attribute used for the various technique used are shown. In this method, the crime users are eliminated from the genuine users. The crime users are further analysed to determine the criminal. The number of criminal users before the analysis of true negative and false positive is 65. After the analysis of true negative and false positive, the number of crime users are reduced to 45. After the classification is done the number of crime users are 2. After using the attribute set 2, the criminal is determined User 32.

E. Implementation Steps

E1. Profile: the profile of 100 users and 25 attributes is given as the input. The sum and average [17] is determined. If the average is greater than 4 represents crime data (1) otherwise genuine data (0). This is the general analysis.

E2. Analysis: using the synthetic identity theft the attributes used are theft, identity theft, cybertheft and computer related offences to analyse the crime. From this data, the crime data and genuine data are determined using the sum and average method. If the average is greater than 4 represents crime data (1) otherwise genuine data (0).

E3. Intrusion detection system: the attributes used are the attributes used to determine the crime here are Malicious Software, Number of times the proxy server is used, Malicious Code Presence, Password Violations and Data Forwarding to determine the crime. From this data, the crime data and genuine data are determined using the sum and average method. If the average is greater than 4 represents crime data (1) otherwise genuine data (0).

E4. Lie detection method: the attributes used are Child soliciting, Abuse, Assault by Threat, Advertising through the internet, Soliciting harlotry through the internet, Drug sales, Excess Privileges, Publication irrelevant content, Transmission of obscene content and Sexually explicit content. From this data, the crime data and genuine data are determined using the sum and average method. If the average is greater than 4 represents crime data (1) otherwise genuine data (0).

E5. Analysis of True Negative and False Positive:

a. The Result of method 1, method 2, method 3 and method 4 is Result 1 and if the data is greater than 2, then it is crime data, otherwise it is genuine data. For the analysis of true negative and false positive, if the result is equal is 2, then it is crime data, otherwise it is genuine data. The final result is based on cluster result and result 2. If result 2 is 1 and cluster result is 1, then final result is 1, otherwise 0. Thus the criminal data is finalized. The criminal users are: 2, 5, 10, 12, 13, 15, 19, 21, 22, 24, 27, 32, 34, 35, 37, 41, 43, 44, 46, 48, 51, 52, 53, 54, 55, 57, 58, 60, 61, 62, 64, 65, 67, 70, 71, 74, 75, 78, 83, 86, 89, 92, 95, 98, 99 (45 users).

b. This set is used for further analysis. The crime data is further classified as Class X, Class Y and Class Z.

c. In Class X , the various attributes with the count 0-2 is counted, if the count range is 11 or 12, that particular user is used for further analysis.

d. In Class Y, the number of attributes with count 3, 4 and 6 is counted, the sum is further taken, if the sum is greater than 16 , that particular user is used for further analysis.

e. In Class Z, the number of attributes with count greater than 6 is counted, if the count is greater than 5, that particular user is used for further analysis.

f. If a particular user is also highlighted in Class X and Class Y , he is used for further analysis.

g. Class X is further analysed to determine the criminal.

h. If there is more than one criminal user , then those particular users are further classified with another set of 25 attributes and the criminal is determined. The criminal is User 32.

G. Performance Metrics

By analysing the performance of various methods using the matlab, it is noted that the attributes used are different for the various users. The input, hidden layer, output layer and the output are the same. The epoch value, performance and gradient are different. The time complexity is the same for all the techniques. The validation is 0, only for Method 1, otherwise it is 6.

TABLE III: THE TABLE SHOWS THE PERFORMANCE METRICS OF VARIOUS METHODS USED

	Attributes	Input	Hidden Layer	Output layer	Output	Epoch	Time	Performance	Gradient	Validation Checks
Method 1-SID	4	45	20	45	45	34	0:00:00	37.1	31.1	0
Method 2-HP	5	45	20	45	45	9	0:00:00	37.3	17.7	6
Method 3-LD	110	45	20	45	45	18	0:00:00	41.9	14.8	6
Method 4-Clustering	225	45	20	45	45	9	0:00:00	41	8.96	6

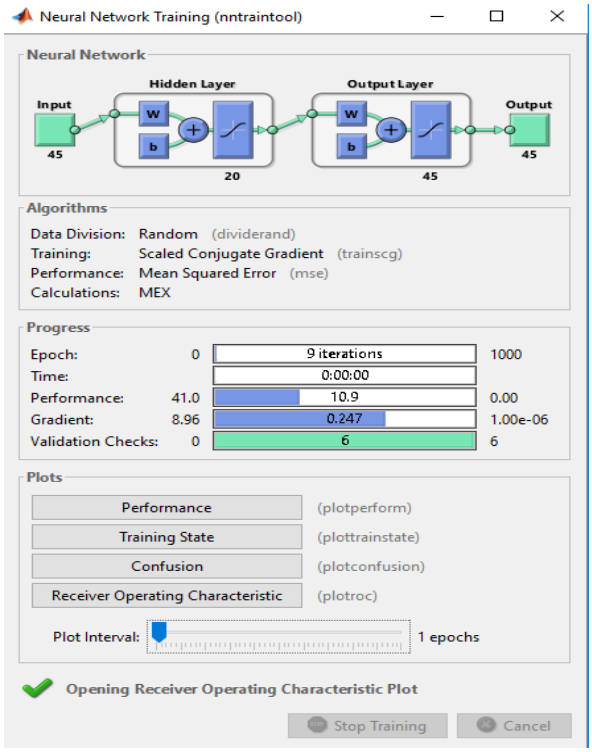

Fig. 1 Neural Network Training

The above figure 1 shows the Neural Network Training for the 45 inputs and 45 outputs (crime users) with the hidden layer and output layer as 20 and 45. The algorithm used is Data division which is random, the training as Scaled conjugate gradient , the performance as Mean Squared Error with calculations as MEX. The epoch is obtained with 9 iterations , with the performance as 10.9, Gradient as 0.247 with validation checks at 6.

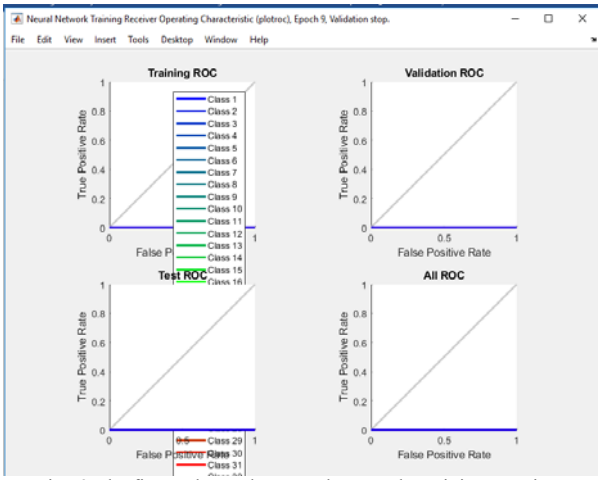

Fig. 2 The figure shows the Neural Network Training Receiver Operating Characteristics

Figure 2 shows the Training Receiver Operating Characteristics (ROC), Validation ROC, Test ROC and All ROC for the False positive rate.

Fig. 3 Crime Users v/s Class X , Class Y and Class Z Classification

Figure 3 shows the crime users v/s the various classification techniques. In Class X , the various attributes with the count 0-2 is counted. In Class Y, the number of attributes with count 3, 4 and 6 is counted.In Class Z, the number of attributes with count greater than 6 is counted.

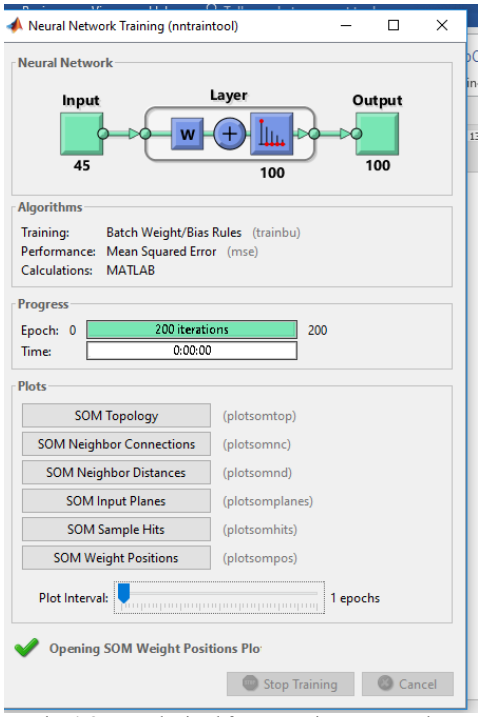

Fig. 4 Output obtained for Input given X, Y and Z

The above Figure 4 shows the neural network training for the input X, Y and Z, where X is the input value , Y is the normal value and Z is the output obtained for 100 users and 25 attributes. The figure also shows the neural network training for the four method as the input. The graph is plotted for the SOM topology, SOM neighbour connections, SOM neighbour distances, SOM Input planes, SOM sample hits and SOM weight positions where SOM is the self organizing map.

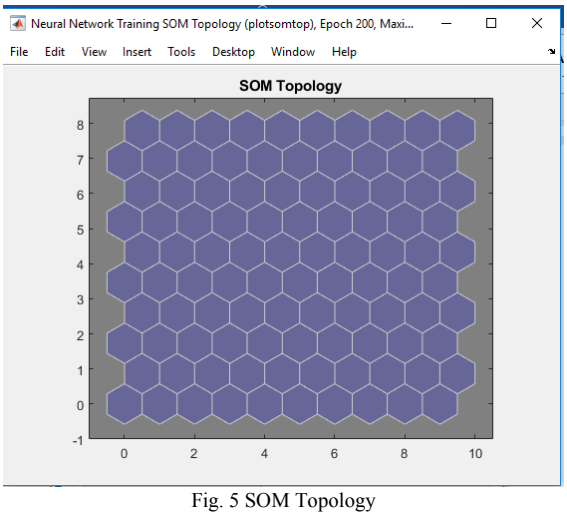

Fig. 5 SOM Topology

Figure 5 shows the SOM topology for 100 users and 25 attributes and also for the four methods used.

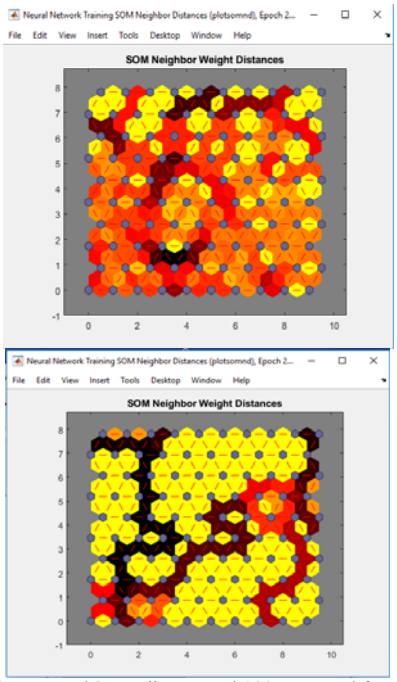

Fig. 6 100 users and 25 attributes and 100 users and four methods

Figure 6 shows the 100 users and 25 attributes and 100 users and four methods, the SOM neighbour weight distances are shown.

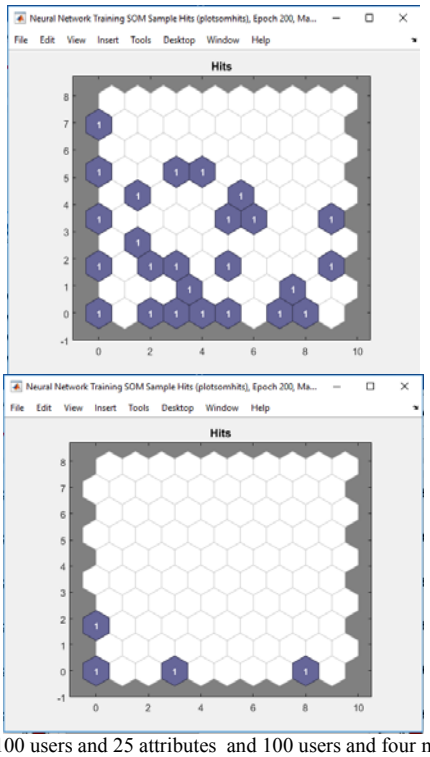

Fig. 7. 100 users and 25 attributes and 100 users and four methods

Figure 7 shows the 100 users and 25 attribute and also for the four methods, the number of Hita are shown here.

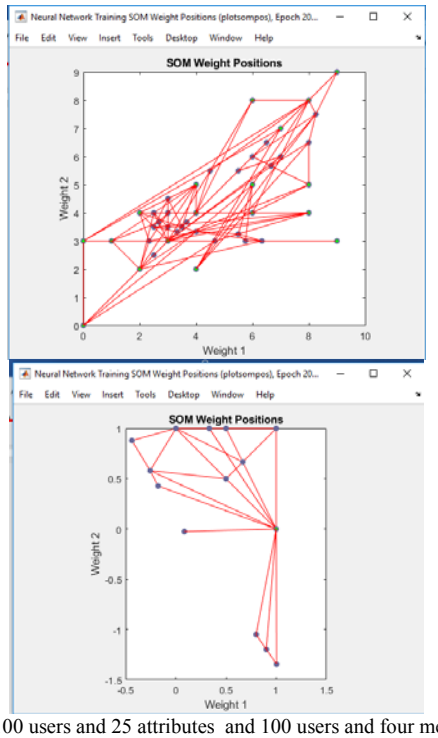

Fig.8 100 users and 25 attributes and 100 users and four methods

Figure 8 shows the 100 users and 25 attributes and 100 users and four methods for the SOM weight positions. The figure is clumsy for 100 users and 25 attributes and clear for 100 users and four methods.

H. Determination of the Criminal

The profile of 100 users and 25 attributes is given as the input. The sum and average is determined. If the average is greater than 4 shows crime data (1) otherwise genuine data(0). For the analysis using the synthetic identity theft the attributes used are theft, identity theft, cybertheft and computer related offences to analyse the crime. From this data, the crime data and genuine data are determined using the sum and average method. If the average is greater than 4 it is crime data (1) otherwise it is genuine data (0).

For the intrusion detection system, the attributes used are the attributes used to determine the crime here are Malicious Software, Number of times the proxy server is used, Malicious Code Presence, Password Violations and Data Forwarding to determine the crime. From this data, the crime data and genuine data are determined using the sum and average method. If the average is greater than 4 shows crime data (1) otherwise genuine data (0).

For the lie detection method, ,the attributes used are Child soliciting, Abuse, Assault by Threat, Advertising through the internet, Soliciting harlotry through the internet, Drug sales, Excess Privileges, Publication irrelevant content, Transmission of obscene content and Sexually explicit content. From this data, the crime data and genuine data are determined using the sum and average method. If the average is greater than 4 shows crime data (1)otherwise genuine data (0).

The criminal users are: 2, 5, 10, 12, 13, 15, 19, 21, 22, 24, 27, 32, 34, 35, 37, 41, 43, 44, 46, 48, 51, 52, 53, 54, 55, 57, 58, 60, 61, 62, 64, 65, 67, 70, 71, 74, 75, 78, 83, 86, 89, 92, 95, 98, 99 (45 users). This set is used for further analysis.

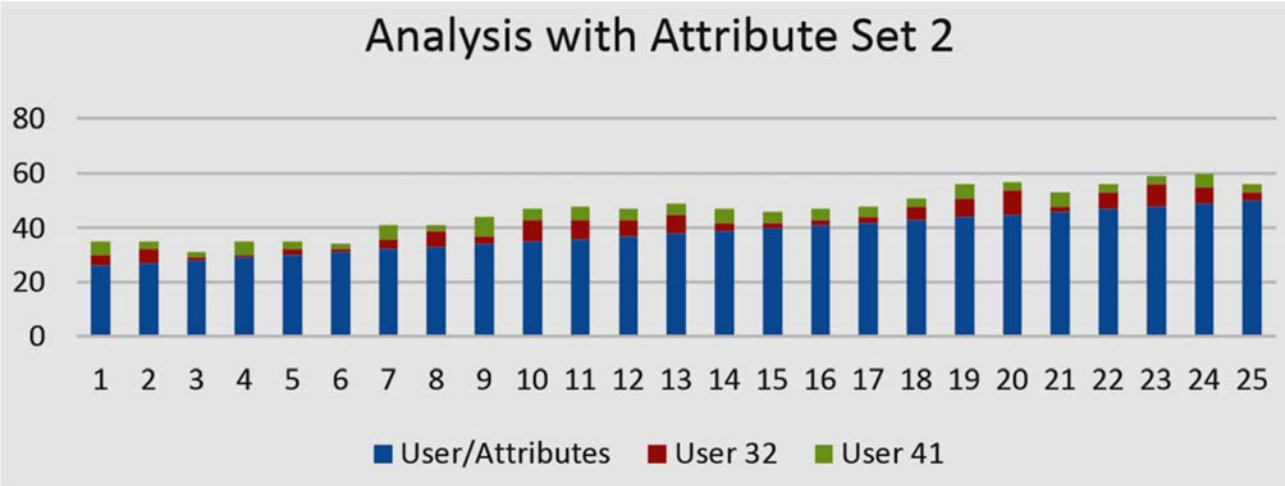

Fig. 9. Analysis with attribute set 2

Finally there are two user that are analysed as crime data, i.e., user 32 and 41. The user 32 and 41 are analysed with another set of 25 attributes and the criminal determined is user 32.

V. CONCLUSION

After analysing the various users with all the four different methods, the cyber criminal is determined. In the first method the synthetic identity is determined. If the user was genuine, then the percentage of genuine data is updated else the percentage of Criminal data was updated. In Intrusion Detection System (IDS) was implemented using the honeypot security mechanism that identifies the incoming traffic from the clients and the traffic coming from the attackers. In this method Load Balancer was designed and implemented. The Balancer discovered the attack on the server while itforwarded the request and when it directed it to different/alternate server called Honey-Pot. If the user was genuine, then the percentage of genuine data was updated else the percentage of Criminal data was updated.

In the third method lie detection system was used to identify the false speech of a human being. In that method the pre processing was used to assist in the reduction of noise and the plotting of the original artefact EEG signals. This methodused focuses mainly on the Neural network used in the recognition phase and Feature Extraction Technique carried out by the MFCC- Mel Frequency Cepstrum Coefficients. After determination of the lie, if the user was genuine, then his percentage of genuine data was updated else the percentage of Criminal data was updated.

After the three methods were executed, then the user was passed through the analyses of the profile to determine the cyber criminal. In that method various clustering techniques was used to determine the percentage of genuineness or criminal. The user data was analysed using two set of attributes. After analyses, the criminal was detected.

Since the criminal was not been able to be determined with attribute set 1, attribute set 2 was used to determine the criminal. The graph was plotted with Crime Users v/s Class A , Class B and Class C Classification, which indicated that Class B has more dominance than Class A and Class C.

The graph was plotted using matlab for the input X, Y and Z, for 100 users and 25 attributes and for the four methods as the input. The iteration was 200 in both the cases. The graph was also plotted for the SOM topology, SOM neighbour connections, SOM neighbour distances, SOM Input planes, SOM sample hits and SOM weight positions.

ACKNOWLEDGEMENT

I would like to thank all the persons who have helped me in writing this paper. I am greatly indebted to my guide who has helped me a lot in completing this paper. Since this topic cannot be spoken openly, I would like to help the victims to be relieved of mental tensions, of narrating the incident again and again. I hope using this technology the cyber criminal can be determined.

REFERENCES

[1] Jonathan Fairtlough J.D, "Introduction to cyber crime investigation", Law tech publishing group, 2015

[2] Chethan Kumar, "One cyber crime in India every 10 minutes", The times of India, 22/07/2017

[3] Cyber-Criminal Activity and Analysis, White Paper Fall 2005, Group 2 ,NilkundAseef (naseef@microsoft.com) Pamela Davis (pdavis@berkeley.edu) Manish Mittal (manishm@microsoft.com) Khaled Sedky (khaleds@microsoft.com) Ahmed Tolba (ahmedt@microsoft.com)

[4] https://blog.ipleaders.in/cyber-crime-detection

[5] Nazura Abdul Manap ,Anita Abdul Rahim , Hossein Taji, " Cyberspace Identity Theft: The Conceptual Framework", Mediterranean Journal of Social Sciences, MCSER Publishing, Rome-Italy, Vol 6 No 4 S3 August 2015, Doi:10.5901/mjss.2015.v6n4s3p595

[6] AtefehTajpour, Suhaimi Ibrahim, MazdakZamani, "Identity Theft Methods and Fraud Types, https://www.researchgate.net/publication/273259976, Article: October 2013

[7] AtefehTajpour, Suhaimi Ibrahim, MazdakZamani , "E-Commerce and Identity Theft Issues", Article October 2013, https://www.researchgate.net/publication/320006684

[8] E. Vasilomanolakis, S. Karuppayah, M. M¨uhlh¨auser, and M. Fischer (2015), "Taxonomy and survey of collaborative intrusion detection," ACM Computing Surveys (CSUR), vol. 47, no. 4, p. 55.

[9] Akshay A. Somwanshi, Prof. S.A. Joshi (2016), "Implementation of Honeypots for Server Security",International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056 Volume: 03 Issue: 03 | Mar.

[10] D. A. Reynolds, "Speaker verification using adapted Gaussian mixture model," Digital Signal Processing, vol. 10, pp. 19-41, 2016.

[11] P. Bhuvaneswari, J. Satheesh Kumar, 2015, "A Note on Methods Used for Deception Analysis and Influence of Thinking Stimulus in Deception Detection", International Journal of Engineering and Technology (IJET) Vol 7 No 1 Feb-Mar 2015.

[12] Roshni D. Tale, B.P.Harne, 2015, "Deception Detection Method using Independent Component Analysis of EEG signals", International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE) Volume 4, Issue 5 May 2015.

[13] T.Sanjana, CM Sheela and KV Narayana, "A survey on Clustering techniques for Big Data Mining", Indian Journal of Science and Technology, Jan 2016

[14] Jyh-Jian Sheu ," Distinguishing Medical Web Pages from Pornographic Ones: An Efficient Pornography Websites Filtering Method", College of Communication, National Chengchi University Taipei City 11605, 2017, . International Journal of Network Security, Vol . 19, No.5,PP839-850, Sept.2017(DOI:10.6633/IJNS.201709.19(5).22)

[15] Prakash Singh etl., ,"Performance Analysis of Clustering algorithms in Data Mining in Weka" ,International Journal of Advances in Engineering &Technology , Jan 2015, ISSN 22311963

[16] K.Veena and K.Meena ," Determination of performance to verify the synthetic identity theft by training the neural networks,Vel Tech. Dr. RR and Dr. SR Technical University, 2017 IEEE International Conference on Smart Technologies And Management For Computing, Communication, Controls, Energy And Materials (ICSTM)

[17] K.Veena and K.Meena ," Identification of cyber criminal by analyzing users profile", VeltechRangarajanDr.Sagunthala R&D

Institute of Science &Technology , International Journal of Network Security,Vol20, PP 738-745, July 2018

[18] Identify Crime Detection Using Data Mining Techniques by K.S.Arthisree and A .Jaganraj, International Journal of Advanced research in Computer Science and Software Engineering. August 2013.

BIOGRAPHY

K.Veena is a Research Scholar in Veltech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai. She received her M.E.,(I.T) degree from Vinayaka Missions University, Salem, Tamilnadu in 2007 and B.E., degree from B.V.Bhoomraddi College of Engineering and Technology, Karnatak University, Karnataka. At present she is a full time Assistant Professor at Dhanalakshmi College of Engineering. Her main interest is in Security issues regarding the safety of women.

Dr K. Meena received her Ph.D Degree in Computer Science from the Department of Computer Science and Engineering, Manonmaniam Sundaranar University, India in 2014. She received the B.E. (Electronics and Communication Engineering) and M.E. (Computer Science and Engineering) Degrees from Manonmaniam Sundaranar University in 2002 and 2009. She is now Associate Professor in Computer Science & Engineering Department, Veltech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai. Her research interests include Machine Learning, Pattern Recognition, Image Processing, Biometrics and Cyber security. She has published 30 papers in international journals, 7 papers in international conferences and 11 papers in national conferences.