Identification of Cyber Criminal by Analyzing Users Profile

K. Veena and K. Meena
(Corresponding author: K. Veena)
Department of Computer Science and Engineering, VelTech Dr. RR and Dr. SR University
400 Feet, Outer Ring Road, Avadi, Chennai 600 062, India
(Email: veenakanagaraj07@gmail.com)
(Received Mar. 27, 2017; revised and accepted July 29, 2017)

Abstract

This paper presents a method to analyze the feedback from various users and thus determine the cyber criminal. The cyber criminals are effectively identified by applying the various clustering techniques for different number of users with different attributes (characters). The purpose of clustering is to identify natural groupings of data from a large data set to represent the system's behavior. The clustering is done using the various techniques like the Gaussian technique, K Means Clustering, Fuzzy C Means Clustering and Fuzzy Clustering. Clustering of numerical data forms the basis of many classification and system modeling algorithms. The data that true without any false information is take as the called the genuine data and the data that contains false information is taken the crime data. By clustering, the genuine data (Cluster 0) is eliminated and only the crime data (Cluster 1) is taken. From the genuine data the false positive is taken as the crime data. From the criminal data the true negative is also eliminated. The criminal data is further analyzed using the various classes and then the criminal is detected. Many of the researchers used minimum number of attributes to identify the criminal. In order to increase the crime identification rate, this paper uses 25 various attributes which are collected from 25 users in different scenarios. In this paper, the profile of the person involved in cyber crime is analyzed for further calculations. By identification of the profile of the cyber criminal, the detection of the crime can be done. In this paper 25 users along with 25 attributes is taken as experimental investigation.

Keywords: Cyber Criminal; Fuzzy Clustering; Identification

1 Introduction

A computer and a network can be used to commit a crime refers to as Cyber Crime. The meeting on Cyber crime was the first international treaty which was conducted to understand the Computer crime and Internet crimes. The convention on cyber crime was the first international treaty that seek to address the computer crime and internet crimes by harmonizing the national laws [7], thus improving the investigative techniques [17] and increasing the cooperation among nations.

Cyber crime is a world wide criminal phenomenon which confuses the customary distinction between fear to criminal and terrorist activity i.e internal and military i.e external security and does not respond to single authority approaches to policing.The liability of networks to exploitation for a number of different ends, and the ease with which individuals may move from one type of illegal activity to another suggests that territorialism in all its forms (both of nations and regions, and specific authorities within nations) hinders efforts to successfully combat the misuse of communications technology. There has been a rise to the industrialization of a type of crime where the commodity, personal information, moves far too quickly for conventional law enforcement methods to keep pace. Stolen personal and financial data - used, for example, to gain access to existing bank accounts and credit cards, or to fraudulently establish new lines of credit - has a monetary value [8, 11].

Whenever a cyber crime is committed the victim suffers silently. He/she is not able to speak openly and accept that he/she is a victim of cyber crime. If the victim is an Indian her case is more ridiculous. She is blamed first, hence they do not express their difficulty outside. In this paper my aim is to help such victims who suffer silently. They should just give a complaint and the set of suspects. The criminal has to be detected [2].

In the paper, GPS Spoofing Detection Based on Decision Fusion with a K-out-of-N Rule [18] to get a high detection probability of the GPS spoofing, decision fusion is proposed and three classifiers are used and the results are fused with K-out-of-N decision rule and the final classification is obtained.

In the paper, Sheu, "Distinguishing Medical Web

Pages from Pornographic Ones: An Efficient Pornography Websites Filtering Method" [15], the uncomplicated decision tree data mining algorithm is used to determine the association rules about the pornographic and medical web pages.

In the paper, "Clustering based K-anonymity algorithm for Privacy preservation" [9] K-anonymity is used as a effective model for protecting privacy while publishing data. A clustering based K-anonymity algorithm is used and it is optimized with parallelization.

In the paper, "A Quantitative and Qualitative Analysis-based Security Risk Assessment for Multimedia Social Networks", much attention is given to the spreading and sharing of personal information in the social media. Social media can be used to follow a person [22].

Cyber crime is a truly global criminal phenomenon which blurs the traditional distinction between threats to internal (criminality and terrorist activity) and external (*i.e.* military) security and does not respond to single jurisdiction approaches to policing.

Here the various clustering techniques are applied for different number of users with different attributes (characters). The input data is further analyzed for different threshold values. The profile of the person involved in cyber crime is analyzed for further calculations [19] by determining the cluster formed using the various clustering techniques. The profile of the cyber criminal is identified. The psychology of cyber criminology directs its attention towards the application of the physical, psychological, social relationships and mental characteristics, as well as towards the evidence of the cybercrime [5,10,20,21]. The computer may have been used in the commission of a crime, or it may be the target, or the user of the computer might have been the target.

The present paper presents a method to analyse the feedback from various users and thus determine the cyber criminal. By clustering, the genuine data (Cluster 0) is eliminated and only the crime data (Cluster 1) is taken [21]. From the genuine data the false positive is also taken as the crime data. From the Crime data the true negative is taken as Genuine data and added to the Cluster 0. The criminal data is further analyzed [14] using the various classes as Class as None, Soft and Hard and then the criminal is detected.

The various clustering techniques are applied for different number of users with different attributes (characters). The data is further analyzed for different threshold values [1]. The profile of the person involved in cyber crime is analyzed for further calculations.

The paper is organized as follows. In Section 1, the abstract is given and a little review of the entire paper. It is followed by the introduction which gives the necessity of detection of cyber crime. Then the motivation and contribution of the proposed work is given. Then the justification of clustering methods are given which consists of the algorithm used.

In Section 2, the methodology used is given which consists of different users, attributes and threshold value, the various clusters identified. Then the comparative results and the results of various clustering techniques are given.

In Section 3, the various classification of clustering is given such as Cluster and Fuzzy C-Means (FCM), Formation of Clusters using the Gaussian Mixture Models. The reasons for choosing Gaussian Clustering Technique is also given.

In Section 4, the implementation and the results are given. The identification of the criminal is also given. A brief description of the Gaussian Clustering Analysis, K Means Clustering Analysis, Fuzzy C Means Clustering Analysis and Fuzzy Clustering Analysis with various Users and Attributes [4, 6, 12, 13, 16]. are given. Finally, the conclusion, acknowledgement and references are given.

2 Justification of Clustering Method

2.1 The Proposed Method

The various clustering methods used here are the Gaussian technique, K Means Clustering, Fuzzy C Means Clustering and Fuzzy Clustering. The clusters are formed based on these clustering techniques. The users profile which consists of attribute set 1 with 25 users are analyzed with the analyzer 1. After the determination of the clusters the clusters are classified as Cluster 0 and Cluster 1. The false positive data is removed in the cluster 0 and the true negative is removed in the cluster 1. The data is further classified as Soft and Alert, Hard and Criminal and None and Genuine based on the average rate as 4-6, 7-9 and 0-3. After classification if the criminal cannot be determined then it is further checked with Analyzer 2, which consists of another set of 15 attributes.

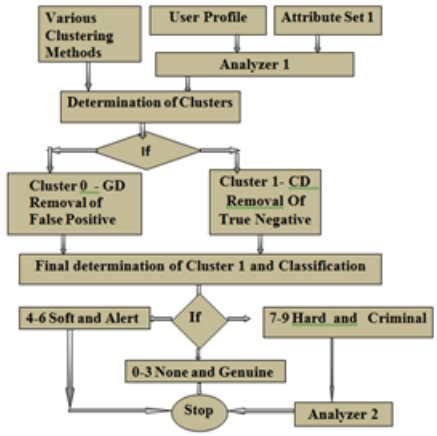

Figure 1: Cyber crime detection flowchart

2.2 Algorithm Used

To analyse the information from various users, the term entropy is the measure of disorder or user data in normal state. It contains the positive and negative values of the cluster formed.

$$\text{Entrophy}(S) = -p_{(+)} \log_2 p_{(+)} - p_{(-)} \log_2 p_{(-)} \quad (1)$$

Algorithm 1 Determination of the criminal

```
Begin
Initialize User Profile and Attribute Set 1
Input data to Analyzer 1, go to Step 5
Sent request to Analyze Data using various clustering methods
Determine the clusters as Genuine data(0) and Crime data(1)
if Cluster 0 then
    print as Genuine data and remove False Positive
    Goto Step 8
else
    print as Crime data and remove True Negative
end if
Final determination of Cluster 1 and Classification
if result is in range 0-3 then
    print as None and Genuine Data
    goto Step 13
end if
if result is in range 4-6 then
    print as Soft and Alert Data
    goto Step 13
else
    if result is in range 7-9 then
        print as Hard and Crime Data
    end if
end if
Goto Analyzer 2
Stop
```

3 Methodology

3.1 Optimization of Attributes Used

Here the data (`http//www.kdnuggets.com/datasets`) taken is for 25 user with various with 25 different attributes (attributes set 1). The various types of Cybercrime which are used as attributes are, given in Table 1. The percentage of the attributes differs in different regions, which is indicated in the table. The attributes set 2 is taken if the criminal cannot be determined with attribute set 1. The attribute set 2 consists of the data consecutive four years before the crime was committed, three months before, three days before the crime, three days after the crime, the day the crime was committed and the relation ship between the criminal and the victim.

Analysis of various attributes with their locations in percentage Table 1.

3.2 Method Used

The data is taken for different number of users, with various attributes and different threshold values. The clusters are formed based on the cut off values. If the cluster falls below the threshold value, the cluster is in "0", otherwise the cluster is in "1". The Cluster 0 is taken as the "Genuine Data" and the Cluster 1 is taken as the "Crime data".

The Different Users, Attributes and Threshold Value Table 2.

3.3 Classification of Clustering

Cluster is a group of objects that belongs to the same class. In other words, similar objects are grouped in one cluster (legal) and dissimilar objects (illegal) are grouped in another cluster. A cluster of data objects can be treated as one group. While doing cluster analysis, we first partition the set of data into groups based on data similarity and then assign the labels to the groups. The main advantage of clustering over classification is that, it is adaptable to changes and helps single out useful features that distinguish different groups. Clustering also helps in classifying documents on the web for information discovery. Clustering is also used in outlier detection applications such as detection of credit card fraud [3]. As a data mining function, cluster analysis serves as a tool to gain insight into the distribution of data to observe characteristics of each cluster. The various Clustering Methods are Partitioning Method, Hierarchical Method, Density-based Method, Grid-Based Method, Model-Based Method, Constraint-based Method and Fuzzy C Means Clustering.

Clustering with Gaussian Mixtures

The Gaussian mixture distributions can be used for clustering data, by realizing that the multivariate normal components of the fitted model can represent the clusters. To demonstrate the process, first some simulated data is generated from a mixture of two bivariate Gaussian distributions using the **mvnrnd** function: The probability density function of the d-dimensional multivariate normal distribution is given by the formula, where

$$y = f(x, \mu, \Sigma) = \frac{1}{\sqrt{|\Sigma|(2\pi)^d}} e^{-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)'}$$

where x and μ are 1-by-d vectors and Σ is a d-by-d symmetric positive definite matrix. Only random vector generation is supported for the singular case.

Partition into Clusters

Then fit the two-component Gaussian mixture distribution. Here the correct number of components

Table 1: Analysis of various attributes with their locations in percentage

Attributes set1	**Browsing Centre**	**Institution**	**Household**	**Mobile**	Medical Shop
Hacking	100	100	100	100	100
Theft	100	100	100	100	100
Cyber Stalking	100	100	100	100	100
Identity theft	100	100	100	100	100
Malicious Software	85	85	85	85	85
Child Soliciting	100	100	50	100	50
Child Abuse	100	100	100	100	100
Assault by Threat	100	100	100	100	100
Child Pornography	100	100	100	100	100
Cyber Illegal imports	85	85	85	85	85
Cyber Laundering	100	100	50	85	85
Cyber Terrorism	100	100	50	85	85
Cybertheft	100	100	50	85	85
Advertising	25	25	25	25	25
Soliciting harlotry	100	100	100	100	100
Drug Sales	25	100	85	25	25
Frequency	25	25	25	25	25
Malicious Code	50	50	25	50	50
Password Violations	85	85	25	85	85
Excess Privileges	50	85	25	50	50
Data Forwarding	25	25	25	25	25
Computer related offences	100	85	85	85	85
Publication irrelevant content	100	85	50	50	50
Transmission of obscene conten	100	100	100	100	100
Sexually explicit content	100	100	100	100	100

Table 2: The different users, attributes and threshold value

Properties	**Data Set1**	**Data Set2**	**Data Set3**	**Remarks**
Users	25	15	5	The number of users are decreased
Attributes	25	15	7	The number of attributes are decreased
Threshold Value	5	3	3	Threshold Value is gradually decreased
Number of Cluster 0	11	3	1	Users and Clusters are propotional
Number of Cluster 1	14	12	4	Users and Clusters are propotional
Remarks	1Greater	1 Greater	1 Greater	

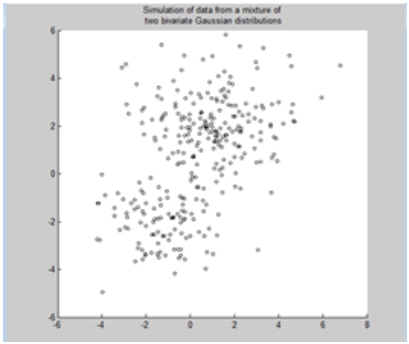

Figure 2: Simulation of data from a mixture of two bivarieties Gaussian distribution

Figure 3: Partition into clusters

is used is two. This data displays 28 iterations, log-likelihood = 1223.66.

Then plot the estimated probability density contours for the two-component mixture distribution. The two bivariate normal components overlap, but their peaks are distinct. From this data it can be concluded that the data could be divided into two clusters. Partition the data into clusters using the cluster method for the fitted mixture distribution. The cluster method assigns each point to one of the two components in the mixture distribution.

3.4 Reasons For Choosing Gaussian Clustering Technique

Analysing the various clustering techniques, it is clear that the Gaussian Clustering Technique is useful compared to the other techniques. In the Gaussian Clustering with various Users and Various Attributes, the number of Cluster 0 and Cluster 1 obtained is different. The iteration Count and the log-like hood remains the same if the users is equal to the number of attributes. The iteration count and the log-like hood differs if the users if different from the attributes. For different Users and attributes with the same threshold value, the iteration count and the log-like hood is different.

4 Implementation and Results

Here various clustering techniques like Gaussian Clustering, K Means Clustering, Fuzzy Means Clustering and Fuzzy Clustering are taken with different data sets. The iteration count decreases in Gaussian Clustering, remains the same in K Means Clustering, increases in Fuzzy C Means Clustering and reduces drastically for Fuzzy Clustering.

Performance Evidence of Various Clustering Techniques Table 3.

In the Gaussian Clustering with various Users and Various Attributes, the number of Cluster 0 and Cluster 1 obtained is different. The iteration Count and the log-like hood remains the same if the users is equal to the number of attributes. The iteration count and the log-like hood differs if the users if different from the attributes. For different Users and attributes with the same threshold value, the iteration count and the log-like hood is different. In the K Means Clustering Analysis with various Users and Attributes, the iteration count is the same and it is 5. But the iteration occurs at different values. The random data obtained also differs. In Fuzzy C Means Clustering the iteration is different for Users. If the number of Users are less then the number of Iteration Count increases. The Object Function is the same, if the threshold value is the same. In Fuzzy Clustering for different users there is different iteration count and different object function. The performance and the time taken for the various data set is shown in table IV. The training used is Scaled Conjugate Gradient, the performance used is Mean Squared Error and the Data Division used is Random.

Neural Network Training for various data set Table 4.

The experimental analysis is done with 25 users and each user has 25 attributes. The various users $1, 2, \cdots, 25$. For every user 25 attributes are taken and numbered as $1, 2, \cdots, 25$. The sum of all the attributes for each user is calculated. The average value of the attributes for each user is calculated. If the average value o is greater than 4 then it is cluster 1 called the (crime data) or o if it is less than 4 then it is cluster 0 called the (genuine data) The number of Cluster 0 is 11 and the various users are 3, 4, 5, 6, 7, 10, 11, 16, 19, 20, 25 and the number of Cluster 1 is 14 and the various users are 1, 2, 8, 9, 12, 13, 14, 15, 17, 18, 21, 22, 23, 24. The number of attribute with value 1 is 51, the number of attribute with value 2 is 194, the number of attribute with value 3 is 452, the number of attribute with value 4 is 132, the number of attribute with value 5 is 320, the number of attribute with value 6 is 146, the number of attribute with value 7 is 75, the number of attribute with value 8 is 94, the number of attribute with value 9 is 60 and the number of attribute with value 10 is 0.

Since the highest is number 3 with value 452 and the lowest is number 1 with value 51,to determine the false positive, the reference is taken as attribute 3 and attribute 1. The number of value 1 and value 3 is counted in each user and then the average is taken. In Cluster 1,

Table 3: Performance evidence of various clustering techniques

Set	**Gaussian**	**K Means**	**Fuzzy C**	**Fuzzy**
Data Set	Count and Log	Count and Data	Count and Object	Count and Object
Data Set1	28 and -1223.66	5(4,5,6,7,8) and 302.923	41 and 787.28	50 and 1180.92
Data Set2	28 and-1223.66	5(3,6,7,8,10) and 295.87	46 and 301.29	100 and 451.93
Data Set3	26 and -1215.09	5(4,4,5,6,6) and 310.605	96 and 301.29	22 and 42.34

Table 4: Neural network training for various data set

Progress	**Epoch**	**Time**	**Performance**	**Gradient**	**Validation Checks**	**Best Validation Performance**
Data Set1	8	0:00:01	19.9/41.0	1.84	6	21.3396 at Epoch 2
Data Set2	8	0:00:01	23.5/40.7	2.15	6	33.4325 at Epoch 2
Data Set3	25	0:00:08	0.000872/0.384	0.000304	6	0.025708 at Epoch 19

the genuine data is to be eliminated. The least number is taken as Genuine data *i.e.*, false positive and added to the Genuine data. It is added to cluster 0. The genuine users are user 1, user 9 and user 18.

In Cluster 0, the crime data is to be added. The highest number is taken as the Crime data *i.e.* true negative and added to the crime data. It is added to cluster 1. The crime data users are user 6, user 20 and user 25.

The user 1, user 9 and user 18 are Genuine data so add them to cluster 0, the user 6, user 20 and user 25 are criminal data so add them to cluster 1. The final users in the crime data are 2, 8, 12, 13, 14, 15, 17, 21, 22, 23, 24, 6, 20, 25. To determine the decision algorithm, the classification is done as if the sum of the range of attributes is from 0-3 it is none classification and the data is genuine, if the sum of the range of attributes is from 4-6 the classification is soft and the data is alert and if the sum of the range of attributes is from 7-9 the classification is hard and the data is criminal. The number of count in each criteria is taken and then the highest number is taken in each classification.

The Result is based on "If no cell is selected then it is NONE". Otherwise "it is either HARD or SOFT". The Result 1is based on "If all cell is selected it is HARD", "if any two cell is selected it also HARD". Otherwise "it is SOFT". The Result is based on if all three cells are selected it is HARD and that user is the criminal. If only two cell is selected, then it is HARD. Whether that user is the criminal, has to be analyzed further. If there is a tie with the users such as, more number of users are in the HARD classification, Then it is further classified with another 15 attributes, and then the criminal is detected. Table 5 shows the determination of the criminal.

From the above formed Cluster 1, the genuine data is removed, the average of Count 1 and Count 3 is taken. The least of the average is taken as the Genuine data. The User 1, User 9 and User 18 are with the least average and hence they are considered as the Genuine Data. The User 1, User 9 and User 18 are added to the Cluster 0. From the formed Cluster 0, the crime data is removed the sum of Count 1 and Count 3 is taken. The least of the Sum is taken as the Crime data. The User 6, User 20 and User 25 are with the highest sum and hence they are considered as the Crime Data. The User 6, User 20 and User 25 are added to the Cluster 1. The table shows the cluster of the Crime data which is used to determine the crime. The True Negative is removed and false positive is added. The data is further classified as Soft, Alert and Hard. Here the user 8 and User 17 are in the same a range and hence further analysis is to be done with the Attribute set 2. Since User 8 and User 17 are having the same features, they are further analyzed with another attributes set 2 and the criminal is detected. It is **User 8**.

5 Conclusions

Analysing the various clustering techniques, it is clear that the Gaussian Clustering Technique is useful compared to the other techniques. In the Gaussian Clustering with various Users and Various Attributes, the number of Cluster 0 and Cluster 1 obtained is different. The iteration Count and the log-like hood remains the same if the users is equal to the number of attributes. The iteration count and the log-like hood differs if the users if different from the attributes. For different Users and attributes with the same threshold value, the iteration count and the log-like hood is different.

1) The feedback from various users are analyzed and thus the cyber criminal is determined. The Clustering of numerical data was used as the basis of classification and system modeling algorithms. The purpose of clustering was to identify natural groupings of data from a large data set to produce a concise representation of a system's behavior. After clustering the genuine data (Cluster 0) is eliminated and only the

Table 5: Determination of the criminal

Users	**Average**	**Cluster**	**GD Removal**	**CD Add**	Classification	Result	Classification2	Result2
1	4.52	CD	GD					
2	4.28	CD			CD	HARD/SOFT	SOFT	
3	4	GD						
4	3.28	GD						
5	3.96	GD						
6	3.2	GD		CD	CD	HARD/SOFT	SOFT	
7	3.6	GD						
8	4.32	CD			CD	HARD/SOFT	HARD	CRIMINAL
9	4.76	CD	GD					
10	3.92	GD						
11	3.72	GD						
12	4.28	CD			CD	NONE		
13	4.12	CD			CD	HARD/SOFT	SOFT	
14	4.16	CD			CD	NONE		
15	4.56	CD			CD	HARD/SOFT	SOFT	
16	3.6	GD						
17	4.32	CD			CD	HARD/SOFT	HARD	
18	4.76	CD	GD					
19	3.92	GD						
20	3.72	GD		CD	CD	NONE		
21	4.28	CD			CD	NONE		
22	4.12	CD			CD	HARD/SOFT	SOFT	
23	4.16	CD			CD	NONE		
24	4.56	CD			CD	HARD/SOFT	SOFT	
25	3.72	GD		CD	GD	NONE		

crime data (Cluster 1) was taken. From the genuine data the false positive was also taken as the crime data. Before the criminal is detected, from the criminal data the true negative was also eliminated. The criminal data was further analyzed using the various classes and then the cyber criminal was detected.

2) The various clustering techniques was applied for different number of users with different attributes (characters). The data was further analyzed for different threshold values. The profile of the person involved in cyber crime was analyzed for further calculations.

3) The reasons for choosing each classification are given below. In the Gaussian Clustering with various Users and Various Attributes, the number of Cluster 0 and Cluster 1 obtained is different. The iteration Count and the log-like hood remains the same if the users is equal to the number of attributes. The iteration count and the log-like hood differs if the users if different from the attributes. For different Users and attributes with the same threshold value, the iteration count and the log-like hood is different. In the K Means Clustering Analysis with various Users and Attributes, the iteration count is the same and it is 5. But the iteration occurs at different values. The random data obtained also differs. In Fuzzy C Means Clustering the iteration is different for Users. If the number of Users are less then the number of Iteration Count increases. The Object Function is the same, if the threshold value is the same. In Fuzzy Clustering for different users there is different iteration count and different object function.

Acknowledgments

The cyber crime and security is a sensitive topic and many of the victims may not wish to speak about it openly. Hence as I would greatly like to help people in this regard using data mining techniques. I owe a great many thanks to a great many people who helped and supported me during the writing of this paper. The author wishes to express her gratefulness to the reviewers and for the chance to profit from the considerate and useful comments. The author wishes generally to emphasize that she is indebted to many ideas raised by other literature sources and is grateful for further and suggestions.

References

[1] O. M. A. Abbas, "Comparisons between data clustering algorithms," *International Arab Journal of Information Technology*, vol. 5, no. 3, pp. 320–325, July 2008.

[2] K. S. Arthisree and A. Jaganraj, "Identify crime detection using data mining techniques," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol 3/8, pp. 977–983, Aug. 2013.

[3] M. Bakhshi, M. R. Feizi-Derakhshi, E. Zafarani, "Review and comparison between clustering algorithms with duplicate entities detection purpose," *International Journal of Computer Science Emerging Tech*, vol. 3, no. 3, pp. 108–114, 2012.

[4] M. Enache, M. Hulea and T. S. Letia, "A new approach in bloggers classification with hybrid of K nearest neighbour and artificial neural network algorithms by training neural network for construction of informatics offender profile," *Indian Journal of Science and Technology*, vol. 8, no. 3, pp. 237–246, Feb. 2015.

[5] Z. Eslami, M. Noroozi, S. K. Rad, "Provably secure group key exchange protocol in the presence of dishonest insiders," *International Journal of Network Security*, vol. 18, no. 1, pp. 33–42, Jan. 2016.

[6] F. S. Gharehchopogh, S. R. Khaze and I. Maleki, "A new approach in bloggers classification with hybrid of K nearest neighbour and artificial neural networks algorithms," *Indian Journal of Science and Technology*, vol. 8, no. 3, pp. 237-246, Feb. 2015.

[7] J. Herhalt, "Cyber crime - A growing challenge for governments," *KPMG International Issues Monitor*, vol. 8, pp. 1–21, July 2011.

[8] X. Li, X. Liang, R. Lu, X. Shen, X. Lin, H. Zhu, "Securing smart grid: Cyber attacks, countermeasures, and challenges," *IEEE Communications*, vol. 50, no. 8, pp. 38–45, Aug. 2012.

[9] S. Ni, M. Xie, Q. Qian, "Clustering based K-anonymity algorithm for privacy preservation, school of computer science and engineering," *International Journal of Network Security*, vol. 19, no. 6, pp. 1062-1071, Nov. 2017.

[10] S. Ni, M. Xie, Q. Qian, "Clustering based K anonymity algorithm for privacy preservation," *International Journal of Network Security*, vol. 19, no. 6, pp. 1062–1071, 2017.

[11] C. Phua, K. Smith-Miles, V. Lee, R. Gayler, "Resilient identity crime detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 3, pp. 533–547, 2012.

[12] S. Revathi, T. Nalini, "Performance comparison of various clustering algorithms," *International Journal of Advanced Research Computer Science and Software Engineering*, vol. 3, no. 2, pp. 67–72, Feb. 2013.

[13] G. Sehgal, K. Garg, "Comparisons of various clustering algorithms," *International Journal of Computer science and Information Techologies*, vol. 5, no. 3, pp. 3074–3076, 2014.

[14] T. Sanjana, C. M. Sheela and K. V. Narayana, "A survey on clustering techniques for big data mining," *Indian Journal of Science and Technology*, vol. 9, no. 3, pp. 1-12, Jan. 2016.

[15] J. J. Sheu, "Distinguishing medical web pages from pornographic ones: An efficient pornography websites filtering method," *International Journal of Network Security*, vol. 19, no. 5, pp. 839–850, Sept. 2017.

[16] P. Singh, A. Surya, "Performance analysis of clustering algorithms in data mining in Weka," *International Journal of Advances in Engineering & Technology*, vol. 7, no. 6, pp. 1866–1873, Jan. 2015.

[17] J. R. Sun, M. L. Shih, M. S. Hwang, "A survey of digital evidences forensic and cyber crime investigation procedure," *International Journal of Network Security*, vol. 17, no. 5, pp. 497–509, 2015.

[18] M. Sun, Y. Qin, J. Bao and X. Yu, "GPS spoofing detection based on decision fusion with a k-out-of-n rule, school of information science and engineering, south east university," *International Journal of Network Security*, vol. 19, no. 5, pp. 670–674, Sept. 2017.

[19] M. Uma, G. Padmavathi, "A survey on various cyber attacks and their classification," *International Journal of Network Security*, vol. 15, no. 5, pp. 390–396, Dec. 2011.

[20] C. Valentine, C. Hay, K. M. Beaver, T. G. Blomberg, "Through a computational lens : using dual computer criminology degree programs to advance the study of criminology and criminal justice practice," *Security Informatics*, DOI: 10.1186/2190-8532-2-2, Jan. 2013.

[21] Z. Zhan, M. Xu, S. Xu, "Characterizing honeypot-captured cyber attacks : Statistical Frame work and Case study," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 11, pp. 1775–1790, 2013.

[22] Z. Zhang, L. Yang, H. Li, and F. Xiang, "A quantitative and qualitative analysis-based security risk assessment for multimedia social networks," *International Journal of Network Security*, vol. 18, no. 1, pp. 43–51, Jan. 2016.

Biography

K. Veena 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.Her main interest is in Security issues regarding the safety of women.

K. Meena obtained her Ph.D., in Manonmaniam Sundaranar University, Tirunelveli. She is a full time Associate Professor at Vel Tech Dr. RR and Dr.SR Technical University. Her research interests are Pattern Recognition, Biometrics,Image Processing, Wireless Sensor Network, Cryptography and Network Security, High Speed Networks and Computer Networks.She is having 16 International Journal, 7 international Conferences and 11 National Conferences to her credit.