Privacy Preserving Approach for Preventing From Access to Unauthorized Users

Anuradha Padala, Yarramalle Srinivas, M. H. M. Krishna Prasad

Abstract: ***Today with the usage of sophisticated systems, most of the people are relying in transmitting the data using high speed internet technologies. However it is resulting into a challenging issue with respect to both storage and transmission because of security features and intelligent unethical persons, it needs to protect the data. The present article highlights a system model which aims at preserving the data within the organizations from unethical persons.***

Index Terms: ***Privacy Preservation, storage, transmission, unethical, GMM***.

I. INTRODUCTION

Privacy Preservation data mining techniques are the most usage techniques used for transmission of data from source to destination. However most of the techniques of privacy preservation aimed at storing or preventing the sensitive information and transmitting the irrelevant information which is of less importance so as to secure the vital information. However these privacy mechanisms are not considered while transmitting the data within groups [1, 2]. In general in most of the organizations are group mails that are being used in the organizations try to transmit the information to the groups. Enough security measures are taken to uphold the security issues. However due to the transfer of information from a group may be revealed to undisclosed persons [3, 4]. This concept is generally considered as leakage of data from within the organization or a group [5, 6]. Therefore it is necessary to develop mechanisms that prevail the transfer of vital information. Many technologies have been highlighted in the literature using the concepts of privacy preservation like K-anonymity, L-divergence etc. with the very objective of converting the core information into a block which cannot be identified by the third person and transferring the other related information across the globe. However this method have some disadvantages has they try to preserve some of the core information which could be pruned after modification of some transactions. Since every concept is concealed it indirectly makes the system failure. On the other hand to overcome this disadvantages geometric data transformation methods are considered for maintaining the privacy of data.

Here the data is converted into chunks where the data is transferred randomly. However when it is needed to share this data, the regrouping becomes a complex issue. To overcome this disadvantage researches have proposed a perturbation techniques based on data mining approaches where the sensitive data is transferred across by adding a noise and at the receivers end this data is rebuilt by filtering the noise. However the main disadvantage of these methodologies is that whenever the noise is added to data, it always leads towards loss of information. Techniques based on genetic algorithms are also considered where a multi-objective function is generated for hiding the sensitive information by using association-rule mining techniques. However there is no guarantee that the information is secured because if a small leakage has taken place, it leads to loss of confidentiality. Also in this mechanism, the data is to be altered by using the rules generated using association-rule mining. However if applications of rules failed automatically it leads to loss of information. Therefore in most of the technologies presented by the reviewers there are some limitations and therefore the present article proposes a methodology to overrule the disadvantages presented in the earlier works and come up with a technology that assures maximum confidentiality and also guarantees the reproduction of original data without loss of information. The article is structured as follows. In section 2, a brief highlight about the considered framework is presented, the concept of link analysis together with a brief note on type of anomalies are presented in section 3, the Gaussian mixture model considered for this article is presented in section 4. The section 5 the methodology of the proposed system is highlighted and experimentation together with results are presented in section 6. The article is summarized with conclusion in section 7.

II. LINK ANALYSIS

In the majority of the circumstances, if the links belonging to a user are posed, definite questions that come up are:

Who are the most leading persons within the network?

The relevant patterns which are common in friends?

Who are the like-minded users and how can we find these similar individuals?

As a solution to these questions, one first needs to define measures for computing centrality, level of associations, and similarity, among other traits.

Revised Manuscript Received on June 15, 2019

Padala Anuradha, Department of Computer Science Engineering, GITAM (Deemed to be University), Visakhapatnam, India.

Yarramalle Srinivas, Department of Information Technology, GITAM (Deemed to be University), Visakhapatnam, India.

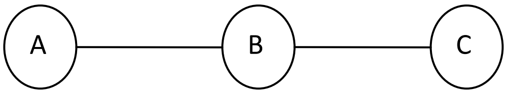
M.H.M Krishna Prasad, Department of Computer Science Engineering, JNT University, Kakinada, India.

These events are considered as input during a social interaction, such as friendships, from which the adjacency matrix values are computed.

To recognize the most interacting entity, centrality measures are considered. These measures help to classify a range of types of central nodes in a network. As a resolution to the other queries, measures are projected based to identify and the degree of centrality, which quantifies the communication patterns among the individuals and help in identifying the like-minded users.

In order to identify the most fitting users, Link Analysis plays a imperative role. The relative users inside the groups can be acknowledged by Cosine value methodology and statistical technique is considered for calculating the average interactions and the degree of cardinality.

In Centrality identification, the Cosine values are calculated as follows



In this thesis, we have utilized the averaging method for identifying the relativeness.

III. TYPES OF ANOMALIES

A. *Users Anonymilies*

Accessing of users accounts in most of the social networking sites are carried out by using their unique names across the world. On the other hand, the genuineness of the users is exposed to both the authenticated users as well as the others who are accessing the social networkings. The most frequent user details are listed which can be accessed through search.

The main disadvantages associated here is, the anonymizer or attacker could know the visitors name, and his profile and can easily navigated and retrieve the relevant information from the social networking's.

B. *De-Anonymization Attack*

In this attack, the attacker tries to identify the group membership information by identifying the history file and tries to intrude an interact with the particular groups. Hence it is the easiest way of attacking the individual's information. In many of the social networking sites the actual name of the user is floated and this name can be accessed by all the users publicly on the social networking sites along with the other individuals who are online. Users are rated depending upon the time-stamp and basing upon the frequency, the user spending more time in the social network media is top ranked and this users profile can be indexed in the search results. The attacker or anonymizer tries to search for users profile and try to interact with the other users in the network. Among the other attacks, de-anonymization attack and neighborhood attack are mostly rated. In the de-anonymization attack, the membership information of group can be obtained from the history by using a stealing technique from this , the anonymizer tries to praise the victim group.

In any social networking media it is easier to intrude into a group compared to the individuals since the number of groups are less in number compared to the number of individuals. The attackers using the history stealing method tries to identify the victims visited URL's and using this history the related links are identified. The retrieval links can be either static or dynamic. Every URL accessed by the user can be obtained from the history stealing method through group directory provided social networking sites. The attacker tries to look into the browsing history of the victim by using conditional logic in cascading style-sheets which is possible through java script. Compared to static link, dynamic links, contains the unique information about the users or group thereby this dynamic link is more utilized for the attacked by the intruder.

C. *Neighborhood Attack*

In general the social network can be represented by using a graph where the relations between the users can be accessed through the path converting the nodes called as link. In the neighborhood attack if A is a user, and B is the users friend, the friend of B can be C, D etc. In the neighborhood attack, the intruder tries to find the friends of A and using this links the sub-links of the friend of A are attacked.

D. *Users Personal Profile Information*

Whenever the intruder attacks into the social network, the personal details of the user can be easily accessed and these information includes the sensitive information such as date of birth, age, relation, background etc. The main source of this attack is leakage of information which is mainly due to inadequate privacy settings which indirectly help the attacker intrude into the system. Among the other attacks, leakage of information to the third party is mostly visualized. In this attack the victim details are stolen and by concealing the actual information, the intruder mis-guides and tries to access the other parties sensitive information. This can be considered as profile cloning.

E. *Profile Cloning*

Profile cloning is considered as stealing the user's identity and duplicating or creating a copy of the victim's identity. The profile cloning can be categorized into existing profile cloning and cross-site profile cloning. In existing profile cloning the attacker creates a duplicate profile of the victim along his photo and sends for friends request to the friends of the victim. This is mostly a successful way of intruding, since most of the friends assumes the victims as the actual and try to share the information.

In cross-site profile cloning the attack from one networking site register with the victims profiles and using this profile tries to access the data from the others networking sites.

F. *Spam Attacks*

The traditional mechanism involved in this process is, the attacker tries to genera+te different e-mail addresses to the users in different public sites and most of these e-mails can be reached into the victims spam.

In the spamming attack the victim checks the spam box and tries to delete the spams and since this spam mail contains malware and other harmful fishing sites, the attacker utilizes these malware to intrude into the system. To other spam mail attacks includes broadcast spam, e-mail based spam and content spam. In these attacks the user either broadcast the malwares across most visited sites or URL’s, broadcast the malware to all the e-mails and using the details of the victim tries to attack the other users within the group.

From the various attacks highlighted in the section 3 it can be understood that there is a desperate need of safe-guarding the users private information and shielding the transactions taking place between the groups.

IV. GENERALIZED GAUSSIAN MIXTURE MODEL

This segment of the work briefly analyses the “probability distribution” of GGMM and its attributes utilized in the criminal identification algorithm. Let the image information values (intensities) in the total Social Network group are obtained randomly, assuming that it follows a “Generalized Gaussian Mixture Distribution”. The total image collection can be considered as the group of “K” Social Network group regions, then the Social Network group values intensities in each Social Network group section go after a Gaussian distribution with generalization. The “Probability density function” is

$$f(Z \mid \mu, \sigma, \rho) = \frac{1}{2\Gamma(1+\frac{1}{\rho})A(\rho,\sigma)} e^{-\left|\frac{(Z_i-\mu_i)}{A(\rho,\sigma)}\right|^{\rho}}$$

V. PROPOSED METHODOLOGY

In order to highlight the methodology, each used identity is considered and the groups are identified. Each group is associated with a tag and in this case, the link between one user to another is attributed as a tag. A user A is considered and his visits/chatting with the user B is rated and correspondingly the number of clique pattern from A to the other related users within the group are identified. To each of these clique patterns, the Eigen vectors are estimated as proposed in section 5.2. basing on these clique pattern the average value between A and B, A and C, and A to the other links is estimated, the Eigen vector corresponding to the link is assumed to be most related to the user A, like-wise the other links between the users of the group are identified and these values are given as input to the PDF’s of GGMM proposed in section 5.6. From these PDF values the maximum and minimum values can be attributed, the user interested to associate with particular groups, needs to satisfy the maximum probability range i.e. the users cardinality is estimated, the Eigen value is generated and the PDF is estimated. This PDF value is compared to the other PDF value ranges and if it is in the maximum likelihood of the highest probability range in any of the group, the user can be linked or permitted to access the details within that group.

VI. EXPERIMENTATION AND RESULTS

In order to experiment the proposed methodology we have considered the dataset of Flicker. The cardinality averaging values together with pdf values are presented in the table 6.1.

Table 6.1: Showing the Group id with the security key

Group ID	User ID	Mean	Cosine	PDF
1	4152	3888.2	1174.8	368
1	932	3888.2	1174.8	368
1	4431	3888.2	1174.8	368
1	5681	3888.2	1174.8	368
1	4245	3888.2	1174.8	368
2	5676	4756.11111111111	306.88888888889	716
2	5465	4756.11111111111	306.88888888889	716
2	5456	4756.11111111111	306.88888888889	716
2	6767	4756.11111111111	306.88888888889	716
3	7832	5063.7	-0.7	819
5	6756	4992	71	797
5	3585	4992	71	797
5	4343	4992	71	797
5	4567	4992	71	797
6	11600	6094.55555555556	-1031.55555555556	877
6	11651	6094.55555555556	-1031.55555555556	877
6	12411	6094.55555555556	-1031.55555555556	877
6	4152	6094.55555555556	-1031.55555555556	877
9	1913	5788.35	-725.35	914
9	4152	5788.35	-725.35	914
10	4152	5863.95454545454	-800.95454545454	909
10	9088	5863.95454545454	-800.95454545454	909
13	4152	5511.48	-448.48	907
13	4152	5511.48	-448.48	907
13	476	5511.48	-448.48	907
16	4579	5347.14814814815	-284.14814814815	885
16	2007	5347.14814814815	-284.14814814815	885
17	7155	5340.06666666667	-277.06666666667	884
17	4152	5340.06666666667	-277.06666666667	884
17	4522	5340.06666666667	-277.06666666667	884
19	2004	5067.63636363636	-4.63636363636	820
19	788	5067.63636363636	-4.63636363636	820
19	4238	5067.63636363636	-4.63636363636	820
22	1084	4962.1282051282	100.8717948718	788
22	6894	4962.1282051282	100.8717948718	788
22	11083	4962.1282051282	100.8717948718	788
22	4152	4962.1282051282	100.8717948718	788
22	580	4962.1282051282	100.8717948718	788
22	2498	4962.1282051282	100.8717948718	788
23	1802	4923.5652173913	139.4347826087	775
23	5217	4923.5652173913	139.4347826087	775
23	5170	4923.5652173913	139.4347826087	775
23	7793	4923.5652173913	139.4347826087	775
23	1809	4923.5652173913	139.4347826087	775
23	1527	4923.5652173913	139.4347826087	775
23	9643	4923.5652173913	139.4347826087	775
24	11434	5558.39622641509	-495.39622641509	911

24	9088	5558.39622641509	-495.39622641509	911
24	11136	5558.39622641509	-495.39622641509	911
24	12401	5558.39622641509	-495.39622641509	911
24	6508	5558.39622641509	-495.39622641509	911
24	8935	5558.39622641509	-495.39622641509	911
24	8609	5558.39622641509	-495.39622641509	911
26	5872	5610.72131147541	-547.72131147541	914
26	8512	5610.72131147541	-547.72131147541	914
26	4040	5610.72131147541	-547.72131147541	914
26	3109	5610.72131147541	-547.72131147541	914
26	9167	5610.72131147541	-547.72131147541	914
26	6041	5610.72131147541	-547.72131147541	914
26	3019	5610.72131147541	-547.72131147541	914
26	7899	5610.72131147541	-547.72131147541	914
28	9088	5666.8064516129	-603.8064516129	916
30	6128	5674.12698412698	-611.12698412698	916
31	8491	5645.38461538462	-582.38461538462	915
31	989	5645.38461538462	-582.38461538462	915
32	4152	5622.75757575758	-559.75757575758	915
34	1591	5562.58208955224	-499.58208955224	911
35	2387	5515.88235294118	-452.88235294118	908
36	884	5448.75362318841	-385.75362318841	900
37	1297	5389.44285714286	-326.44285714286	892
39	9088	5441.53521126761	-378.53521126761	899
40	4152	5423.625	-360.625	897
41	1686	5372.42465753425	-309.42465753425	889
42	234	5302.98648648649	-239.98648648649	877
44	9088	5353.45333333333	-290.45333333333	886
45	4834	5451.16666666667	-388.16666666667	901
45	10376	5451.16666666667	-388.16666666667	901
45	8472	5451.16666666667	-388.16666666667	901
46	12335	5555.2380952381	-492.2380952381	911
46	6313	5555.2380952381	-492.2380952381	911
46	9975	5555.2380952381	-492.2380952381	911
46	4152	5555.2380952381	-492.2380952381	911
46	4152	5555.2380952381	-492.2380952381	911
46	4522	5555.2380952381	-492.2380952381	911
48	1759	5510.57647058824	-447.57647058824	907
49	4152	5494.77906976744	-431.77906976744	906
50	8124	5525	-462	908
52	6765	5539.09090909091	-476.09090909091	910
53	7294	5615.95555555556	-552.95555555556	914
53	10702	5615.95555555556	-552.95555555556	914
54	2473	5616.12658227848	-553.12658227848	914
54	1448	5616.12658227848	-553.12658227848	914
54	11479	5616.12658227848	-553.12658227848	914
54	12407	5616.12658227848	-553.12658227848	914
54	3217	5616.12658227848	-553.12658227848	914
54	2992	5616.12658227848	-553.12658227848	914
54	1297	5616.12658227848	-553.12658227848	914
54	3109	5616.12658227848	-553.12658227848	914

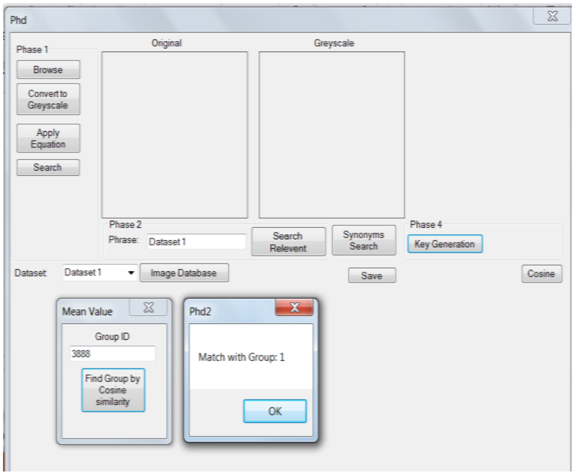
54	4743	5616.12658227848	-553.12658227848	914
54	5938	5616.12658227848	-553.12658227848	914
54	9766	5616.12658227848	-553.12658227848	914
54	3890	5616.12658227848	-553.12658227848	914
54	2843	5616.12658227848	-553.12658227848	914
54	2896	5616.12658227848	-553.12658227848	914
54	5732	5616.12658227848	-553.12658227848	914
54	5272	5616.12658227848	-553.12658227848	914
54	7526	5616.12658227848	-553.12658227848	914
54	7810	5616.12658227848	-553.12658227848	914
54	1295	5616.12658227848	-553.12658227848	914
54	1138	5616.12658227848	-553.12658227848	914
54	3587	5616.12658227848	-553.12658227848	914
54	3640	5616.12658227848	-553.12658227848	914
54	3899	5616.12658227848	-553.12658227848	914
54	9009	5616.12658227848	-553.12658227848	914
54	8170	5616.12658227848	-553.12658227848	914
54	11747	5616.12658227848	-553.12658227848	914
54	9247	5616.12658227848	-553.12658227848	914
54	1016	5616.12658227848	-553.12658227848	914
54	795	5616.12658227848	-553.12658227848	914
54	11695	5616.12658227848	-553.12658227848	914
54	1949	5616.12658227848	-553.12658227848	914
54	5418	5616.12658227848	-553.12658227848	914
54	7943	5616.12658227848	-553.12658227848	914
54	4341	5616.12658227848	-553.12658227848	914
54	721	5616.12658227848	-553.12658227848	914
54	561	5616.12658227848	-553.12658227848	914
54	6278	5616.12658227848	-553.12658227848	914
54	304	5616.12658227848	-553.12658227848	914
54	6540	5616.12658227848	-553.12658227848	914
54	9185	5616.12658227848	-553.12658227848	914
54	8512	5616.12658227848	-553.12658227848	914
54	11235	5616.12658227848	-553.12658227848	914
54	694	5616.12658227848	-553.12658227848	914
54	3312	5616.12658227848	-553.12658227848	914
54	4991	5616.12658227848	-553.12658227848	914
54	1622	5616.12658227848	-553.12658227848	914
54	4939	5616.12658227848	-553.12658227848	914
54	78	5616.12658227848	-553.12658227848	914
54	1430	5616.12658227848	-553.12658227848	914
54	7025	5616.12658227848	-553.12658227848	914
54	1624	5616.12658227848	-553.12658227848	914
54	10673	5616.12658227848	-553.12658227848	914
54	4825	5616.12658227848	-553.12658227848	914
54	364	5616.12658227848	-553.12658227848	914
54	6094	5616.12658227848	-553.12658227848	914
54	9408	5616.12658227848	-553.12658227848	914
54	808	5616.12658227848	-553.12658227848	914
54	10441	5616.12658227848	-553.12658227848	914

54	7843	5616.12658227848	-553.12658227848	914
54	5297	5616.12658227848	-553.12658227848	914
54	9032	5616.12658227848	-553.12658227848	914
54	8876	5616.12658227848	-553.12658227848	914
54	12590	5616.12658227848	-553.12658227848	914
54	8983	5616.12658227848	-553.12658227848	914
54	11247	5616.12658227848	-553.12658227848	914
54	10095	5616.12658227848	-553.12658227848	914
54	10686	5616.12658227848	-553.12658227848	914
54	5872	5616.12658227848	-553.12658227848	914
55	3780	5666	-603	916
55	10777	5666	-603	916
55	12403	5666	-603	916
55	1898	5666	-603	916
55	4630	5666	-603	916
55	7188	5666	-603	916
55	7691	5666	-603	916
55	4841	5666	-603	916
56	11930	5667.92485549133	-604.92485549133	916
56	2295	5667.92485549133	-604.92485549133	916
56	9291	5667.92485549133	-604.92485549133	916
56	6236	5667.92485549133	-604.92485549133	916
56	3652	5667.92485549133	-604.92485549133	916
56	5272	5667.92485549133	-604.92485549133	916
56	1319	5667.92485549133	-604.92485549133	916
57	4152	5659.21264367816	-596.21264367816	916
59	11326	5691.59428571428	-628.59428571428	916
60	10196	5733.29608938548	-670.29608938548	916
60	6394	5733.29608938548	-670.29608938548	916
60	6887	5733.29608938548	-670.29608938548	916
60	6754	5733.29608938548	-670.29608938548	916
61	10811	5775.98404255319	-712.98404255319	915
61	9288	5775.98404255319	-712.98404255319	915
61	9088	5775.98404255319	-712.98404255319	915
61	5145	5775.98404255319	-712.98404255319	915
61	1071	5775.98404255319	-712.98404255319	915
61	1622	5775.98404255319	-712.98404255319	915
61	8917	5775.98404255319	-712.98404255319	915
61	4152	5775.98404255319	-712.98404255319	915
61	9531	5775.98404255319	-712.98404255319	915
62	11083	5804.06349206349	-741.06349206349	913
63	4214	5795.6947368421	-732.6947368421	914
64	6754	5771.40625	-708.40625	915
64	174	5771.40625	-708.40625	915
65	12404	5805.77202072539	-742.77202072539	913

67	4152	5830.36224489796	-767.36224489796	912
67	12213	5830.36224489796	-767.36224489796	912
67	5872	5830.36224489796	-767.36224489796	912
69	6943	5958.01731601732	-895.01731601732	899
69	5010	5958.01731601732	-895.01731601732	899
69	10199	5958.01731601732	-895.01731601732	899
69	8598	5958.01731601732	-895.01731601732	899

From the table 6.1 the most predominant features between the users can be identified.

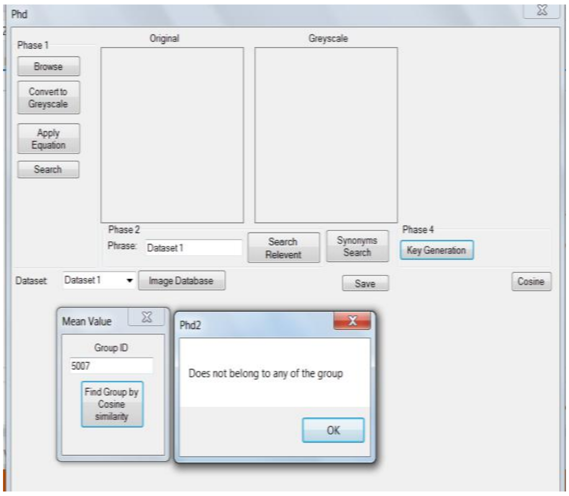
In order to authenticate a new user, his/her user ID is considered and is mapped to the average ID of every group. For this purpose Cosine Similarity is used. Based on this similarity index, the most relevant user will be allowed to interact with the group, thereby the authentication is established, and so that anonymizer cannot intrude into the system.



3888

Match with Group: 1

Fig. 6.2(a): Showing to insert User Id to identify the Group.



5007

Does not belong to any of the group

Fig. 6.2(b): Showing that the users is not matched with any group

From the above figure 6.2(a) and figure 6.2(b) we can easily identify the attackers and safe guard the information being shared by the outsiders indirectly reflects the efficiency of the model.

VII. SUMMARY

In this article a methodology is highlighted using the cardinality values obtained by calculating the average values. These values are used as tags to authenticate the user. The proposed methodology helps to overcome the de-anonymizer method.

REFERENCES

1. Gayathiri. P, Dr. B Poorna" Association Rule Hiding for Privacy Preserving Data Mining : A Survey on Algorithmic Classifications" International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 23 (2017) pp. 13917-13926
2. Rajesh N., Sujatha K. and Arul Lawrence A Selvakumar. Article: Survey on Privacy Preserving Data Mining Techniques using Recent Algorithms. International Journal of Computer Applications 133(7):30-33, January 2016. Published by Foundation of Computer Science (FCS), NY, USA
3. Anuradha.P, Y.Srinivas, MHM Krishna Prasad"A Frame Work for Preserving Privacy in Social Media using Generalized Gaussian Mixture Model" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 6, No. 7, 2015
4. Manish Sharma, Atul Chaudhary, Manish Mathuria, Shalini Chaudhary, Santosh Kumar " An efficient approach for privacy preserving in data mining" 2014 International Conference on Signal Propagation and Computer Technology (ICSPCT 2014) 28 August 2014, IEEE xplore
5. Shweta Taneja, Shashank Khanna, Sugandha Tilwalia, Ankita, 2014, A Review on Privacy Preserving Data Mining: Techniques and Research Challenges, International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, pp: 2310-2315
6. Li, Xueyun & Yan, Zheng & Zhang, Peng. (2014). A Review on Privacy-Preserving Data Mining. Proceedings - 2014 IEEE International Conference on Computer and Information Technology, CIT 2014. 769-774. 10.1109/CIT.2014.135.
7. A. Hussien, N. Hamza and H. Hefny, 2013 , Attacks on Anonymization-Based Privacy-Preserving: A Survey for Data Mining and Data Publishing, 5ournal of Information Security, Vol. 4 No. 2, 2013, pp. 101-112. doi:10.4236/jis.2013.42012.
8. Gayatri Nayak, Swagatika Dev "A Survey On Privacy Preserving Data Mining: Aproaches And Techniques" International Journal of Engineering Science and Technology (IJEST) Vol. 3 No. 3 March 2011
9. S. Vijayrani, A. Tamilarasi, M. Sampoorna, "Analysis of Privacy Preserving k-anonymity Methods and Techniques", Proceeding of the International Conference on Communication and Computational Intelligence, pp. 540-545, December 2010.
10. Narayanan, V. Shmatikov, ―De-anonymyzing social networks‖, In Proc of 30th IEEE Symposium on Security and Privacy, Berkely, CA, pp 173-187,2009
11. A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam,"L-diversity: Privacy beyond k-anonymity," ACM Transaction onKnowledge Discovery from Data, vol.1, Article No.3, Mar 2007.
12. Agarwal, R. and Shrikant, R. "Privacy Preserving Data Mining", Proceeding of Special Interest Group on Management of Dat, pp. 439-450, 2000