I.J. Modern Education and Computer Science, **2018, 9, 26-32**
Published Online September 2018 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijmecs.2018.09.04

The Power of Anonymization and Sensitive Knowledge Hiding Using Sanitization Approach

T.Satyanarayana Murthy, N.P.Gopalan
National Institute of Technology /CA, Tiruchirapalli, 620015, INDIA.
Email: murthyteki@gmail.com,npgopalan@nitt.edu

Datta Sai Krishna Alla
Vignan University/CSE, Guntur, INDIA
Email: dattasaialla11@gmail.com

Received: 03 August 2018; Accepted: 19 August 2018; Published: 08 September 2018

***Abstract*—**In recent day's huge rapid growth of corporate industries professional are based on the online marketing. These markets are associated with millions of online transactions which contain the details of the items, number of items, price and additional information like working details, salary information and personal information. The customers associated with these transactions are concerned about privacy issues. This manuscript aims to concentrates more on the additional information about the customer apart from dealing with the items. More analysis helps in knowing the sensitive information about an individual. In this article two algorithms were used, out of which first algorithm has been used to hide the sensitive information about an individual and other proposed algorithm has been used to hide the sensitive transaction information. These algorithms are proposed based on k-Anonymity and association rule hiding techniques. A novel algorithm has been proposed for association rule hiding algorithm to reduce the side effects such as Sensitive item-set hiding failure, Non-sensitive misses, extra item-set generations and Database dissimilarities along with the reduction of running time and complexities through transaction deletion.

***Index Terms*—**Association Rule Hiding, Anonymization, k-Anonymity, Sensitive Items, Non-Sensitive Items.

I. Introduction

The Association rule hiding has been a blooming topic in the field of Data Mining with the introduction of Privacy Preserving Data Mining (PPDM) by Agrawal in the year 2006 [1,2,3,4,5]. PPDM which is knowledge hiding method to protect the end results can be used in association rule mining, clustering and classification. The main objective of this area is to find better ways and techniques to get near optimal result of hiding sensitive rules which pose a threat to privacy (O'Leary 1991,1995).A solution that hides all sensitive data without producing any side effects is optimal solution. The association rule hiding was first introduced by Atallah in 1999. Many optimization algorithms exist to address this problem. Some of the approaches for sanitization are heuristic, exact, boarder and evolutionary. Some of the evolutionary based algorithms are:a) GeneticAlgorithm [6,7,8] based algorithms(Holland, 1992), such as cpGA2DT(Lin et al. 2014a),spGA2DT(Lin et al. 2014b),pGA2DT(Lin et al., 2015).Apart from hiding the sensitive rules huge data collection has been evolving from multiple sources in large volumes and acquiring in various forms requires data anonymization. These data have privacy concerns may hide the sensitive information using un-realization algorithms lead to leakage of data. Dataset is made up of tuples. Each tuple associated with a collection of attributes like A1, A2, A3...An. These attributes are divided into identifying, non-sensitive, sensitive and Quasi Identifiers. Identifying attributes are directly used to identify a person. Sensitive attributes contain confidential information about an individual like a disease, city, etc. , Quasi-identifiers are a combination of attributes that are combined to identify an individual. Non-sensitive attributes contain public information. Anonymization techniques are k-anonymity, l-diversity, t-closeness are mainly used to anonymize the data. It concentrates on Anonymization of sensitive identity information. Let a dataset released for the research purpose by removing the identifying attributes and sensitive attributes, but an un-authorized user tries to disclose the identity of the individuals by using the quasi-identifiers and non-sensitive data. Anonymization methods are k-Anonymity, l-diversity, and t-closeness fail in the better way of hiding the data. These techniques lead to a homogenous attack, background knowledge attack, and similarity attack. Anonymity method accepts the non identified data as an input and produces anonymized data as an output without subjecting to attacks. In this article we applied both the anonymization and rule hiding techniques together for achieving better results in anonymizing the data with less execution time and hiding with lower side effects..

This paper contains 8 sections.Section-2 contains the literature survey about the various evolutionary

algorithms.Section-3 specifies the preliminaries and basics of association rule hiding. Section 4 and 5 describe the problem definition and k-anonymity algorithm.Section-6 and 7 states the proposed algorithm and its performance parameters.Section-8 contains experimental results and the last Section gives the conclusion and future work.

II. Literature Survey

In 1999 Atallah et al., [9] proposed a Disclosure limitation of sensitive rules uses cyclic based approach for handling the sensitive rules. In 2001 three algorithms 1.a, 1.b ,2.a were presented by Dasseni et al.,[10] in which 1.a, 1.b was based on the concept of increasing support of the antecedent of the rule and thus reducing the confidence of the rule. The algorithm 2.a was based on the concept of decreasing the support of the itemset of rule generated. Some algorithms similar to previous three Confidence reduction(CR),CR2, and Generating Itemset Hiding(GIH) were proposed by Saygin et al.In 2004 [11] the work of Dasseni et al.(2001) was extended by Verykios et al.(2004b) and introduced algorithm 2.b in order to hide generating itemset of sensitive rules. In 2007 [12] Wang et al., proposed hiding predictive association rules, proposed an algorithm Decrease Support and Confidence (DSC) algorithm. The constraint Satisfaction problem proposed by Menon et al. in 2005 was extended by Menon and Sarkar to reduce Not-To-Hide rules and number of sanitized transactions. In 2013 B.Keshava Murty [13] proposed a Privacy preserving association rule mining over distributed databases using genetic algorithm for hiding the sensitive rules. In 2014 Lin et al.,[14] in order to select transactions for hiding itemset used Genetic algorithms for the first time in which Lin et al (2014a) contain Compact Pre-large GA-based algorithm to Delete Transactions(cpGA2DT) that deletes the transactions specified and Lin et al., (2014b) contains an algorithm proposed to produce and insert new transactions . On the basis of PISA platform (Bleuler et al., 2003) an Evolution based Multi-objective Optimization –base Rule Hiding (EMO-RH) algorithm was proposed by Cheng et al. (2014).In 2015 Lin et al.,[15,16,17] proposed algorithms Simple Genetic Algorithm to Delete Transactions (sGA2DT) and Pre-large Genetic Algorithm to Delete transactions .In 2016 [18,19] Particle swarm optimization based algorithm to Delete Transaction (PSO2DT) was proposed by Lin et al.(2016) with less parameters and also to find the number of transactions to be deleted to minimize side effects which was earlier done manually. Cuckoo Optimization Algorithm for Association Rule Hiding (COA4ARH) by Afsari et al., In (2016)[20] was proposed to hide sensitive rules using Cuckoo Optimization Algorithm(Yang & Deb,2009).The algorithm varies in the concept of performing a pre-processing operation that contain 2 phases in order to reduce number of loops and time to get an optimal solution.In 2017 Telikani & Shahbahrami [21] improved MaxMin solution (Moustakides &Verykios, 2006, 2008) with 2 heuristics to hide association rules.This algorithm is named DCR(Decrease the confidence of Rule).J.M.-T.Wu et al.: Ant Colony System Sanitization approach for hiding sensitive itemsets.,(2017).JIMMY MING-TAI WU [22]proposed an Ant Colony Optimization based algorithm for association rule hiding. T.Satyanarayana Murthy et al.,[23,24,25,26,27]proposed meta heuristic based algorithms for better way of association rule hiding. Un-realization Algorithm plays vital a role in data modification, data perturbation, data swapping and Statistical based Techniques. Aggarwal and Yu P.S [28] proposed perturbation and randomization approaches. k-Anonymity [29] a data modification technique achieving privacy by using modified attributes values. These modifications are based on generalization and suppression. Data Perturbation techniques accept the original dataset and add noise so that a sanitized dataset is produced, but reconstruction of the original dataset from the perturbed data was compromised. Random substitutions are an advanced perturbation approach where randomly change the values [30]. Data Swapping [31] uses a t-order statistics model to perturb the dataset by swapping the values in a dataset. Chong k.Liew et al., [32] use probability distribution for data distortion. The major operations for data distortion are, identify density function and generate a disturbing series and reconstruction of the original series. Rakesh Agrawal et al., proposed a perturbation technique uses Gaussian distribution, noise adding and random substitution. It cannot handle periodic updates on the data, and it cannot apply to categorical attributes. Jaideep Vaidya et al., [33] proposed a privacy-preserving scalar product protocol for privacy-preserving data mining. Lindell et al., [34] suggested a SMC algorithm for data hiding. It uses ID3 algorithm to achieve the privacy. Hilol Kargupta at al.,[35] proposed a data perturbation approach based on random matrix based filtering. C. Aggarwal et al.,[36] proposed a condensation approach which uses nearest neighbor classifier technique for hiding the sensitive data. Ashwin et al., [b37] proposed an l-diversity technique for data Anonymization. It resolves the limitations of k-anonymity. It cannot handle continuous sensitive attributes. It performs better than k-anonymity, and still, it has limitations. They are similarity attack and skewness attack. Ninghui et al. in proposed a t-closeness approach overcome the barriers of l-diversity approach. It produces better results than l-diversity. It mainly calculates the distances between the sensitive attribute distribution and the attribute distribution. Salva kisleirch et al. proposed a KATCUS algorithm for k-anonymity. It used decision tree approach. Pui K. Fong et al.,[38,39] proposed a dataset complementation approach and proved by using decision tree classifier for hiding the data. It uses ID3 algorithm and modified ID3 algorithm for achieving the privacy. It cannot handle periodic updates in the dataset. More Storage complexity required because it uses universal set approach. Differential privacy and t-closeness both are used to anonymize the data. In In this

paper proposed two algorithms, both yields better results for Anonymization and hiding.

III. Association Rule Hiding Process

In Figure.1 shows the relationship between the side effects of Association rule hiding process. ABCD represents original dataset and EOHD is the sanitized dataset i.e the output. The various side effects are described thereafter. In ABCD, ABC is the non frequent itemset (FIs)i.eSupport less than Min Support threshold.ACD represent set of itemsets that are frequent in the database i:e Support Min Support threshold. AOD is the set of sensitive data items specified by user or expert; OCD is the non sensitive itemsets.

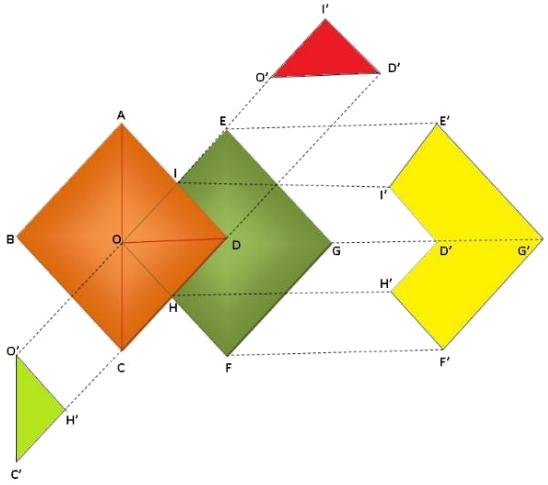

Fig.1. Association Rule Hiding Process

OCH is the data items which were supposed to be present in sanitized database but a miss occurs and it remains hidden. This is called Non sensitive Miss (NSM).In EOFG (Sanitized dataset), IOD is the sensitive dataset of original dataset that was supposed to be hidden in sanitized dataset but the sanitization algorithm could not hide these items. Thus a Sensitive Hidden Failure (SHF) occurs. EIDHFG is another side effect where the extra or ghost rules are generated due to sanitization process. This is called ERG represented as EIDHFG is the extra or ghost rules that are generated as a side effect of sanitization process. Let the original database be Db and the sanitized database be Db* and then the similarity in Db* and Db must be maintained high, with value not less than 90, which is the currently achievable value in most of the algorithms. Database similarity = Db*-Db . A high value of database similarity indicates that the sanitized database still reflects the original database structure and the deletion and insertions did not affect the overall database to larger extend.

IV. Problem Definition

Online shopping data consists of tuples where each T(T1, T2....Tn).These tuples are associated with attributes like A(A1, A2, A3...An).The privacy of customer information has been anonymized by an anonymization algorithm and the association rules are sanitized so the original database is followed by the side effects Sensitive data hiding failure, Non sensitive Misses and the Extra item-sets being generated. Thus to reduce the side effect a novel algorithm has been developed to delete the transactions to obtain sanitized database. To address this problem, architecture has been proposed by using a k-anonymity and a rule hiding algorithms are proposed.

V. K-ANOnymity Technique

K-Anonymity techniques used to resolve the identity disclosure. Consider a data holder such as a bank or hospital, it wants to share their data with the other organizations but the sensitive information needs to be hidden, for solving this problem a formal protection mode named k-anonymity and a set of policies was developed based on suppression and generalization techniques. If the information for each person contained in the release cannot be separated or distinguished from at least k-1 individuals whose information also there in the release. These are the possible attacks against k-anonymity: 1. Unsorted matching attack 2. Complementary release attack 3. Temporal attack

Table 1. Customer Dataset from on-line store

Sno	Cust ID	Customer Name	Address	Mobile	Trans Id
1	101	Shyam	Thuvakudi	9025887845	211014
2	102	Sagar	KK Nagar	9025887846	211015
3	103	Naveen	Katur	9025887847	211016
4	104	Vidhya	Tiruverumbur	9025887848	211017
5	105	Satya	Oil Mill	9025887849	211018
6	106	Sree	NITT	9025887850	211019
7	107	pandu	Thilai Nagar	9025887851	211020
8	108	prasad	Sri rangam	9025887852	211021
9	109	niharika	TVS Toll gate	9025887853	211022
10	110	praneeth	Mannapuram	9025887854	211023
11	111	prasanth	Jk Nagar	9025887855	211024

Table-1 and Table-2 consists of the customer data of the online Shoppe. The data should be protected from the attackers who access the data to know the details of the person or an individual. We need to prevent the sensitive information of the individuals from being disclosed. By using suppression technique the values of attributes are

replaced by *. In this data we have changed the identifying attribute to *. The last digits of customer ID, Mobile number, Transaction ID are suppressed as *. This will be difficult to access the data and their anonymized representation given in Table-2. The address is changed to Trichy because it may have many villages and towns. So it is difficult to know the address of the person.

Table 2. Anonymized Customer Dataset from on-line store

Sno	Cust ID	Customer Name	Address	Mobile	Trans Id
1	10*	*	Trichy	90258878**	2110**
2	10*	*	Trichy	90258878**	2110**
3	10*	*	Trichy	90258878**	2110**
4	10*	*	Trichy	90258878**	2110**
5	10*	*	Trichy	90258878**	2110**
6	10*	*	Trichy	90258878**	2110**
7	10*	*	Trichy	90258878**	2110**
8	10*	*	Trichy	90258878**	2110**
9	10*	*	Trichy	90258878**	2110**
10	10*	*	Trichy	90258878**	2110**
11	10*	*	Trichy	90258878**	2110**

that contains all the Transactions and the Item sets. For the given Database, Apply the Association rule Hiding Process.

Table 3. Customer Dataset with items

Trans Id	Item1	Item2	Item3	Item4	Item5
211014	A	B	C	D	E
211015	B	C			
211016	E	F	C		
211017	C	D	A		
211018	A	C	D		
211019	A	D			
211020	A	E	F		
211021	A	B	C	F	
211022	A	C			
211023	A	D			
211024	A	E			

and give the result. Consider the min_conf value is 70% and the_min sup value is 40%.Finally, the rules that are generated from the above L2 table are:

Table 4. Association rules from Customer Dataset

Trans Id	Item1	Item2
AC >D	30%	80%
AD >C	30%	75%
CD >A	30%	100%
A >C	50%	62.5%
A >D	40%	50%
C >A	50%	71%
D >A	40%	100%

Out of the seven rules that are generated from the above Transactions Table, Rules 1, 2, 3, 4, 5 are initially hidden because either their confidence or support values are less than the given min support or min conf values. The remaining rules are the Rule 6 & Rule 7. In order to hide them, we will use the bitmap notations concept. We have to choose what rule that has to be hidden and select the transaction & make the value of any item to 0. Now we will hide the rule, C >A. Its support value is 50% and confidence value is 71%. The transaction which supports the above rules is 211023 I.e., AC. Its bitmap Notation is < 211023; 101000; 2 >. Now change the Value of C in the derived bit map notation to 0. Then it becomes, < 211023; 100000; 1 >. After modifying the Value in the BitMap, its support and confidence values are 40% and 66.66% respectively. As the confidence value is lesser than the min-confidence value, that particular rule is hidden. Similarly, for the rule D >A, change the value of the D to 0. Then the resultant will be as follows, I.e., < 211019; 100000; 1 >. After modifying, its support and confidence values will be 30% and 100% respectively. So that rule can be also hidden.

VI. MPSO2DT Technique

1. Let D be a dataset.
2. Use the dataset D to generate the rules.
3. Rules R=SAR+NSR
4. Calculate the support S, confidence C for all the rules where s>MST and c>MCT .
5. A support set S=s1,s2...snand confidence set C=c1,c2,...cn.
6. Determine the MAXSUP and MAXCONF constant values.
7. Determine the sensitive rules based on MAXSUP and MAXCONF.
   for(rules R, i variable) do
   Let an array B[i]=if(Rule(i).S(i)>MAXSUP and Rule(i).C(i)>MAXCONF)
   until ..No Rules.
8. B[i]=b1,b2,b3,b4...bn are the sensitive rules.
9. modified perturbation().
10. Compare D1 with D for losses

VII. Performance Parameters

Some of the most predominant performance parameters of association rule hiding are.

1. **Hiding Failure:** Some sensitive rules fail to be hidden in the sanitized database.
2. **Lost rules:** Some of the frequent itemsets that are non sensitive remain hidden in sanitized database. This leads to unnecessary hiding of data.
3. **Ghost rules:** Ghost rules are the additional rules which are generated because of sanitization process.
4. **Database dissimilarity:** The ratio between the sanitized database Db* and the original database

Db must be kept minimum i.e the deletion or change in the transactions of the original database must be minimum. Database similarity ratio must be more than 90%.

5. **Computational complexity:** The computational time complexity must be minimum i.e the sanitization algorithm must be able to produce the resultant database with lesser time than other algorithms. The lesser the complexity, more efficient is the algorithm.
6. **Accuracy:** The degradation of data accuracy leads to resulting of the knowledge extracted from the sanitized database to be useless. The accuracy is inversely proportional to the dissimilarity in the databases.

VIII. Experimental Results

Experiments are conducted on Core-i3 processor and ap-plied k-anonymity techniques and ARH technique .The pro-posed algorithm MPSO2DT has been compared with PSO2DT and cpGA2DT in achieving better results.

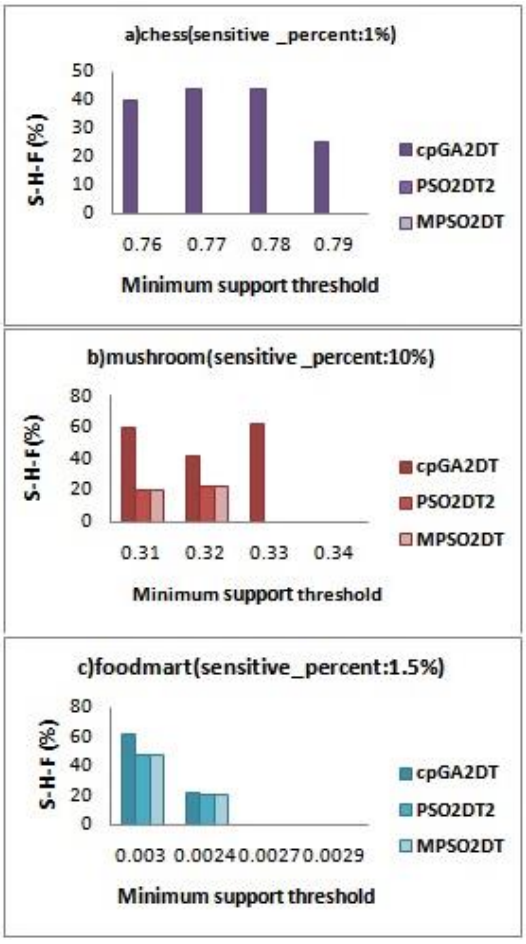

Fig.2. S-H-F

A. S-H-F

The S-H-F is an evaluation metric to find the extend of the Sensitive data item-sets that could not be hidden in Db* . It is given by

$$S\ H\ F = SIs^{*}\text{-}SI$$

Where SIs* is the sensitive data item-sets that are still present in sanitized database which was supposed to be hidden. And SIs* is a superset of SIs* which is the sensitive dataset of original database shown in Figure 2.

References

[1] Rakesh Agrawal, Tomasz Imielinski, Arun Swami," Mining association rules between sets of items in large databases". *ACM SIGMOD international conference on Management of data SIGMOD*, pp. 207, 1993.

[2] Agrawal,R.,Srikant,R., a.QuestSyntheticDataGenerator . IBM AlmadenRe- searchCenter ⟨http://www.Almaden.ibm.com/cs/quest/syndata.html⟩,1994.

[3] Chen, M.S.,Han,J.,Yu,P.S.,"Datamining:An overview from a database per spective". *IEEE Trans. Knowl. DataEng*. 8(6),866–883,1996.

[4] Aggarwal,C.C.,Pei,J.,Zhang,B.,."On privacy preservation against adversarial data mining", *ACM SIGKDD International Conference on Knowledge Dis- covery and Data Mining*, pp.510–516, 2006.

[5] Goldberg, D.E., "Genetic Algorithms in Search,Optimization and Machine Learning". Addison-Wesley Longman PublishingCo., Inc, Boston, MA, USA, 2002.

[6] Goldberg, David ," The Design of Innovation: Lessons from and for Competent Genetic Algorithms", *Norwell, MA: Kluwer Academic Publishers*,2002.

[7] Kennedy,J.,Eberhart,R," Particle swarm optimization ", *IEEE International Conference on Neural Networks*, pp.1942–1948,1995.

[8] Kennedy, J.,Eberhart,R.,."A discrete binary version of particle swarm algorithm". *In IEEE International Conference on Systems,Man,and Cybernetics*, pp. 4104–4108,1997.

[9] Atallah, M. , Bertino, E. , Elmagarmid, A. , Ibrahim, M. , Verykios, V., "Disclosure limitation of sensitive rules", *In IEEE knowledge and data engineering exchange*, pp. 45–52,1999.

[10] Dasseni,E.,Verykios,V.S.,Elmagarmid,A.K.,Bertino,E.,"H iding association rules by using confidence and support", *International Workshop on Information Hiding*, pp.369–383,2001.

[11] Verykios, V. S. , Elmagarmid, A. K. , Bertino, E. , Saygin, Y. , Dasseni, E.," Association rule hiding", *IEEE Transactions on Knowledge and Data Engineering*, pp. 434–447,2004 .

[12] Wang, S.-L. , Parikh, B. , & Jafari, A. "Hiding informative association rule sets". *Expert Systems with Applications*, 33 , pp-316–323 ,2007.

[13] B.Kesava Murthy ,Asad M.Khan," Privacy preserving association rule mining over distributed databases using genetic algorithm", *Neural Computing and Application* , 2013.

[14] C.-W. Lin, T.-P. Hong, C.-C. Chang, and S.-L. Wang, "A greedy-based approach for hiding sensitive itemsets by transaction insertion," *J. Inf. Hiding Multimedia Signal Process*., vol. 4, no. 4, pp. 201–227, 2013.
[15] C.-W. Lin, B. Zhang, K.-T. Yang, and T.-P. Hong, "Efficiently hiding sensitive itemsets with transaction deletion based on genetic algorithms, "*Sci. World J.*, vol. 2014, pp. 1–13, Sep. 2014.
[16] C.-W. Lin, T.-P. Hong, K.-T. Yang, and S.-L. Wang, "The GA-based algorithms for optimizing hiding sensitive itemsets through transaction deletion," *Appl. Intell.*, vol. 42, no. 2, pp. 210–230, 2015.
[17] J. C.-W. Lin, L. Yang, P. Fournier–Viger, M.-T. Wu, T.-P. Hong, and L. S.-L. Wang, "A swarm-based approach to mine high-utility itemsets," *in Proc. Multidisciplinary Social Netw. Res.*, pp. 572–581,2015.
[18] J. C.-W. Lin, Q. Liu, P. Fournier-Viger, T.-P. Hong, M. Voznak, and J. Zhan," *A sanitization approach for hiding sensitive itemsets based on particle swarm optimization," Eng. Appl. Artif. Intell.*, vol. 53, pp. 1–18, Apr. 2016.
[19] J.C.W. Lin, Q. Liu, P. Fournier-Viger, T.P. Hong, M. Voznak, J. Zhan," A sanitization approach for hiding sensitive itemsets based on particle swarm optimization", *Engineering Applications of Artificial Intelligence*, pp 1-18, 2016.
[20] Mahtab Hossein Afshari ,M.N.Dehkordi,M.Akbari" Association rule hiding using cuckoo optimization algorithm" *Expert Systems With Applications*, vol 64, pp 340–351,2016.
[21] Telikani, A., Shahbahrami, A., Optimizing association rule hiding using combination of border and heuristic approaches. Applied Intelligence 47, 544–557,2017.
[22] Jimmy Ming-Tai Wu,Justin Zhan and Jerry Chui Wei Lin "Ant Colony System Sanitization Approach to Hiding Sensitive Itemsets" in IEEEAccess,2017.
[23] N.P.Gopalan, T.Satyanarayana Murthy, Yalla Venkateswarlu, "Hiding Critical Transactions using Un-realization Approach", IJPAM, Vol 118, No.7 ,629-633,2018.
[24] T.Satyanarayana Murthy, N.P.Gopalan, "A Novel Algorithm for Association Rule Hiding", *International Journal of Information Engineering and Electronic Business (IJIEEB)*, Vol.10, No.3, pp. 45-50, 2018. DOI: 10.5815/ijieeb.2018.03.06
[25] T.Satyanarayana Murthy, N.P.Gopalan, Yalla Venkateswarlu," An efficient method for hiding association rules with additional parameter metrics ", *IJPAM* ,Vol 118,No.7, 285-290,2018.
[26] T.Satyanarayana Murthy, N.P.Gopalan, "Association rule hiding using chemical reaction optimization",SCOPRO 2017 Conference,IIT Bhubaneswar, 2017, (Accepted).
[27] T.Satyanarayana Murthy.,"Privacy Preserving for expertise data using K-anonymity technique to advise the farmers", *International Journal of Electrical, Electronics and Data Communication*, Volume-1, Issue-10, 2013.
[28] Aggarwal C.C., Yu P.S, Privacy preserving Data Mining:, Models and Algorithms.Springer, 2008.
[29] Sweeney,L. : k-anonymity: A Model for Protecting Privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems 10(5), 557–570, 2002.
[30] J. Dowd, S. Xu, W. Zhang, "Privacy-preserving decision tree mining based on random substitutions", P*roc. Int. Conf. Emerg. Trends Inf. Commun. Security*, pp. 145-159, 2006.
[31] T. Dalenius and S.P. Reiss, "Data-Swapping: A Technique for Disclosure Control," *Journal of Statistical Planning and Inference*, vol. 6, pp. 73-85, 1982.
[32] C. K.Liew ,U.J.Choi , C.J.Liew, "A Data distortion by probability distribution" ACM TODS ,pp 395-411,1985.
[33] J. Vaidya and C. Clifton, "Privacy preserving association rule mining in vertically partitioned data", in The Eighth ACM SIGKDD International conference on Knowledge Discovery and Data Mining, Edmonton, Alberta, CA, July 2002, IEEE 2002.
[34] Y. Lindell, B.Pinkas, "Privacy preserving data mining", in proceedings of Journal of Cryptology, 5(3), 2000.
[35] H. Kargupta and S. Datta, Q. Wang and K. Sivakumar, "On thePrivacy Preserving Properties of Random Data PerturbationTechniques", *in proceedings of the Third IEEE International Conference on Data Mining*, IEEE 2003.
[36] C. Aggarwal , P.S. Yu, "A condensation approach to privacy preserving data mining", in proceedings of International Conference on Extending Database Technology (EDBT), pp. 183–199, 2004.
[37] Machanavajjhala, J.Gehrke, D. Kifer and M. Venkitasubramaniam, "I-Diversity: Privacy Beyond k-Anonymity", Proc. Int'l Con! Data Eng. (ICDE), p. 24, 2006.
[38] Pui K. Fong and Jens H. Weber-Jahnke, Senior Member ,"Privacy Preserving Decision Tree Learning Using Unrealized Data Sets ", *Proceeding of the IEEE Transactions On Knowledge And Data Engineering*,VOL. 24, NO. 2, pp. 353 -364, FEB 2012.
[39] P.K. Fong, J.H.Weber-Jahnke "Privacy preservation for training data sets in database: Application to decision tree learning, " M.SC Thesis,2012.

Authors' Profiles

N.P.Gopalan Professor at Department of Computer Applications, National Institute of Technology, Tiruchirappalli, Tamil Nadu, India. He obtained his Ph.D. from Indian Institute of Science, Bangalore, India. His research interests are in Data Mining, Distributed Computing, Cellular Automata, Theoretical Computer Science, Image Processing and Machine Intelligence.

T.Satyanarayana Murthy Department of Computer Applications, National Institute of Technology, Tiruchirappalli. He obtained his B. Tech. in Information Technology from Jawaharlal Nehru Technological University, Hyderabad, Andhra Pradesh, India in 2006 and M. Tech. in Computer Science Engineering from Andhra University, Visakhapatnam, Andhra Pradesh, India in 2010. His current research interests include Data Mining, Privacy Preserving, CBIR, Big Data, Soft Computing.

Alla Guru Datta Sai Krishna Studying B.Tech in Vignans University. His research interests towards Data Analytics and Privacy Preserving Data Mining issues.

How to cite this paper: T.Satyanarayana Murthy, N.P.Gopalan, Datta Sai Krishna Alla, "The Power of Anonymization and Sensitive Knowledge Hiding Using Sanitization Approach", International Journal of Modern Education and Computer Science(IJMECS), Vol.10, No.9, pp. 26-32, 2018.DOI: 10.5815/ijmecs.2018.09.04

