Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 12, December 2014, pg.111 – 117

RESEARCH ARTICLE

Collaborative Data Mining using Incentives and Multi-Party Secure Communications

PEETHALA ASHA[1], A.GANESH KUMAR[2]

Assistant Professor, Department of CSE, Gokul Institute of Technology and Sciences, Piridi village.
Bobbili mandalam ,Vizianagaram dt. Jntu kakinada university A.P, India[1]
Student, Department of CSE, Gokul Institute of Technology and Sciences, Piridi village.
Bobbili mandalam ,Vizianagaram dt. Jntu kakinada university A.P, India[2]

ashapeetala@gmail.com [1], ganeshkumar_gokul@gmail.com [2]

Abstract-- Security and privacy play an important role in collaborative data mining when multiple competing parties are involved. The environment is expected to be a non-cooperative and there is no guarantee that all parties provide compatible or correct inputs. To encourage them to provide correct inputs, many techniques came into existence. One such method was recently proposed by Kantarcioglu et al. based on the incentives. Incentives let the competing parties to give meaningful inputs. In this paper we build mechanisms based on incentives and secure multi-party communications to ensure security and privacy in collaborative data mining. Our prototype application demonstrates the proof of concept. It takes data from multiple competing parties and verify the inputs for their correctness. Based on the correct inputs, the incentives of the party which gives input will be increased. When misbehavior is encountered, the incentives are reduced and the party is asked to provide correct inputs again. On giving the correct input, the reduced incentives are added again. This kind of technique and also a mechanism for secure multi-party communications makes the system useful for secure and privacy preserving collaborative data mining. The empirical results are encouraging.

Index Terms – Security, privacy-preserving data mining, horizontally partitioned data, vertically partitioned data

I. RELATED WORKS

This section reviews the literature on related works. Privacy preserving data mining is very important as the data mining should not disclose sensitive details. The researchers in [23], [15], [14] and [16] focused on horizontally partitioned data. Other researches in [22], [6] and [13] focused on vertically partitioned data. Many privacy preserving data mining solutions [2] assumed that the participants provide data truthfully. Horizontal partitioning is done through distributed partition of data. The data is horizontally split into various sites. It does mean that the data is saved in various locations. The collaborative data mining where multiple participants or competing parties provide compatible data in a distributed environment it is important that all parties provide genuine data. The vertical data is

split into multiple databases and that can be used as and when required. Non cooperative computation models and other related models were explored in [20] and [21]. More details are provided on NCC model in [18].

Another important research area explored in literature was about Secure Multi-Party Computations (SMC) that helps multiple competing parties to contribute in providing data in secure manner. The competing parties are supposed to give genuine data as explored in [25], [24], and [9]. SMC protocols take care of the secure communications among multiple parties who involve in the data mining process. In this paper we follow an approach that provides incentives to the genuine data givers. Among the competing parties, it is observed for any incompatible data. When incompatible data is provided by any party, the incentives are reduced otherwise the incentives are increased. The idea is to ensure that competing parties are encouraged to provide genuine data.

II. PROTOTYPE IMPLEMENTATION

We built a prototype application that facilitates collaborative data mining. Multiple parties can collaborate and provide data for mining in distributed environment. Privacy preserving data mining is the important research area. In this paper we focused on building a data mining application that allows multiple parties to give compatible inputs in order to mine the data to produce business intelligence. Association rule mining is demonstrated with the application. However, it can be extended to support all mining operations. The collaborative data mining along with privacy preserving data analysis is possible with the help of the prototype application. The application has the notion of incentives. Incentives allow the competing parties to behave well. When they misbehave, the incentives will be decreased that will affect the credibility of that party for participating collaborative computing. This is the motivation the competing parties get so as to provide genuine inputs to the system.

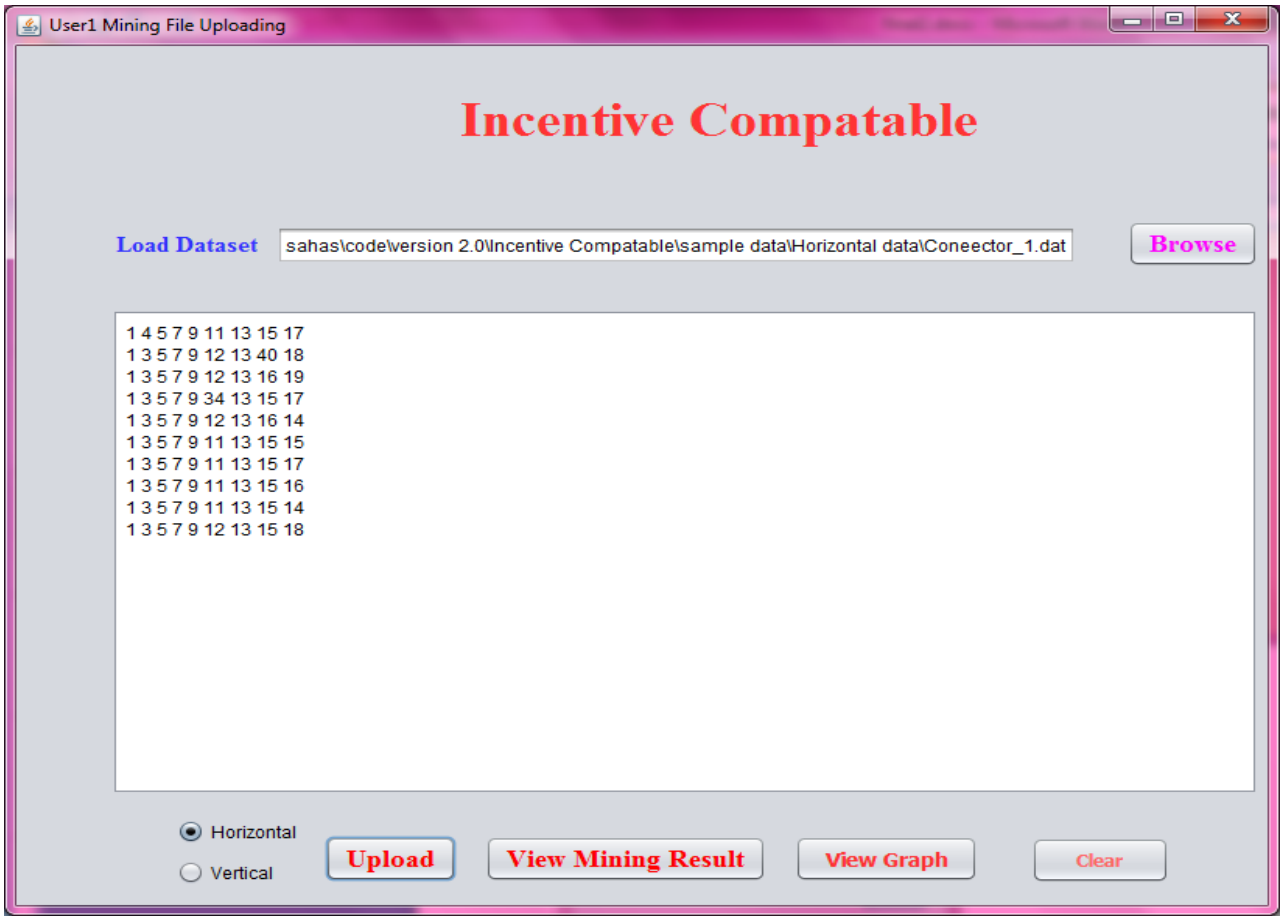

Figure 3 – User 1 loads either horizontal or vertically partitioned data

As can be seen in Figure 3, it is evident that the ends user is capable of providing inputs. This way multiple parties can provide inputs for collaborative data mining. In this application three users are considered for experiments. The UI for uploading data also provide functionalities for viewing data mining results and have a graph to understand the data analysis details just by a glance. In the same fashion, three users can provide inputs that are compatible. On clicking view mining results, the application generates the association rules as presented in Figure 4.

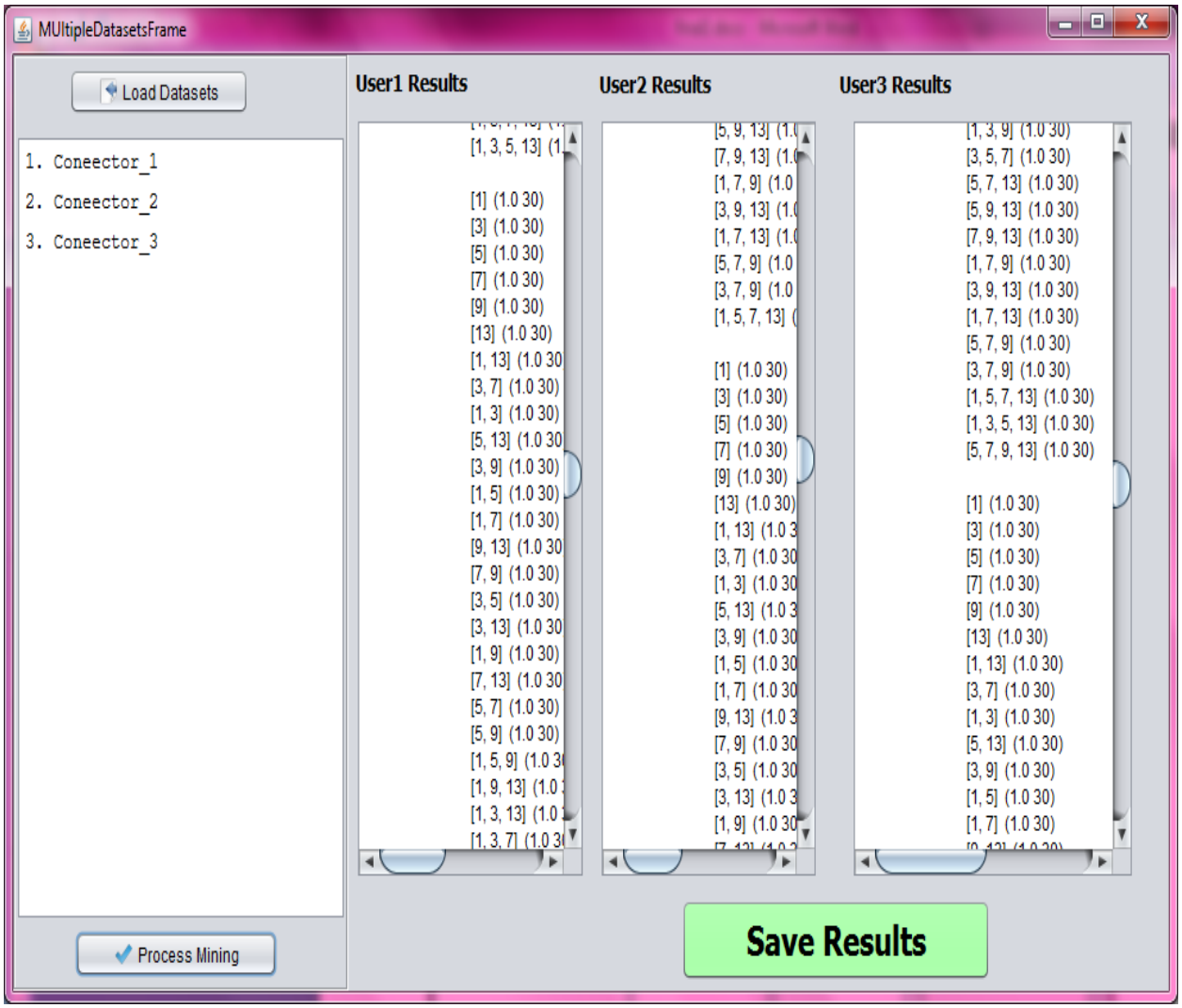

Figure 4 – The results of collaborative data mining

As can be seen in Figure 4 it is evident that the mining results are presented. However before mining results the data given by the competing parties is revivified for compatibility. If any inconsistency with respect to given values is suspected, that particular party is notified besides reducing incentives. Then one more chance is given to provide correct data. When the party gives correct data again the incentives will be restored and the process of mining gets completed. The consolidated results along with incentives are provided as shown in Figure 5.

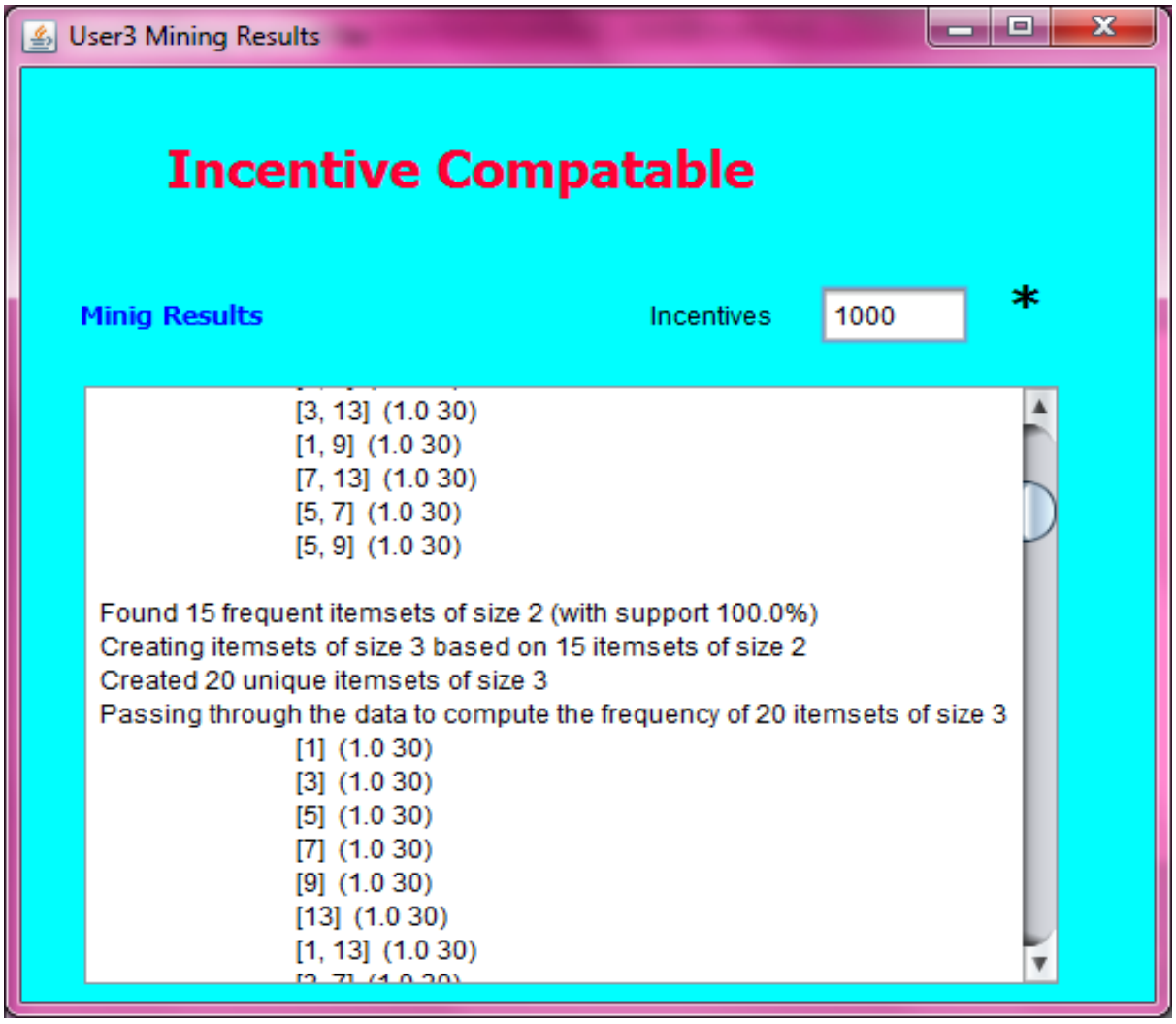

Figure 5 – Consolidated results of collaborative and privacy preserving data analysis

While performing the data mining after taking data from competing parties, the security is applied and each party needs to provide pre-distributed key so as to validate users in a secure distributed environment. Figure 6 presents the results of incompatible data provided by one of the parties. Then the incentives are reduced so as to encourage the party to give correct input. This will reduce the chances of misbehavior on parties thus achieving privacy preserving collaborative data analysis.

Figure 6 – Incentives of different competing parties

As can be seen in Figure 6, it is evident that the parties involved in the privacy preserving collaborative data analysis are given incentives for every genuine act. For every misbehaving act, incentives are reduced. This makes them get motivated to provide genuine results so as to achieve the purpose of collaborative data analysis in distributed environment.

III. CONCLUSIONS AND FUTURE WORK

In this paper, we studied the problem of collaborative data analysis. When multiple competing parties are involved in providing data for privacy preserving data analysis, the parties might provide either genuine data or they provide incompatible data. In this paper we explored the possibilities to encourage all competing parties to provide genuine data so as to serve the purpose of collaborative data mining in order to acquire new business intelligence and device new business models. We presented secure and incentive compatible approach that lets competing parties to get incentives for genuine data. When incompatible data is suspected, the incentives are reduced. This is the motivation for the competing parties to behave well and provide genuine data. We built a prototype application that demonstrates the proof of concept. The empirical results revealed that the proposed solution is good for collaborative and privacy preserving data analysis in distributed environment. In future we extend this research into other aspects such as Big Data mining with MapReduce programming using distributed programming frameworks like Hadoop.

REFERENCES

[1] I. Abraham, D. Dolev, R. Gonen, and J. Halpern, "Distributed Computing Meets Game Theory: Robust Mechanisms for Rational Secret Sharing and Multiparty Computation," Proc. 25th Ann. ACM Symp. Principles of Distributed Computing, pp. 53-62, 2006.
[2] R. Agrawal and E. Terzi, "On Honesty in Sovereign Information Sharing," Proc. Int'l Conf. Advances in Database Technology, pp. 240- 256, 2006.
[4] I. Ashlagi, A. Klinger, and M. Tenneholtz, "K-NCC: Stability Against Group Deviations in Non-Cooperative Computation," Proc. Third Int'l Conf. Internet and Network Economics, pp. 564-569, 2007.
[5] M.J. Atallah, M. Bykova, J. Li, and M. Karahan, "Private Collaborative Forecasting and Benchmarking," Proc. Second ACM Workshop Privacy in the Electronic Soc. (WPES), Oct. 2004.
[7] Doe News, www.doe.gov, Feb. 2005.
[9] "Directive 95/46/EC of the European Parliament and of the Council of 24 Oct. 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data," Official J. European Communities, vol. 281, pp. 31-50, Oct. 1995.
[11] O. Goldreich, S. Micali, and A. Wigderson, "How to Play Any Mental Game - A Completeness Theorem for Protocols with Honest Majority," Proc. 19th ACM Symp. the Theory of Computing, pp. 218-229, 1987.
[13] S.D. Gordon and J. Katz, "Rational Secret Sharing, Revisited," Proc. Int'l Conf. Security and Cryptography for Networks, p. 229, 2006.
[14] J. Halpern and V. Teague, "Rational Secret Sharing and Multiparty Computation: Extended Abstract," Proc. Ann. ACM Symp. Theory of Computing (STOC '04), pp. 623-632, 2004.
[16] S. Han and W.K. Ng, "Preemptive Measures against Malicious Party in Privacy-Preserving Data Mining," Proc. SIAM Int'l Conf. Data Mining (SDM), pp. 375-386, 2008.
[17] "Standard for Privacy of Individually Identifiable Health Information," Fed. Register, vol. 67, no. 157, pp. 53181-53273, Aug. 2002.
[18] S. Izmalkov, S. Micali, and M. Lepinski, "Rational Secure Computation and Ideal Mechanism Design," Proc. 46th Ann. IEEE Symp. Foundations of Computer Science (FOCS '05), pp. 585-594, 2005.
[19] G. Jagannathan and R.N. Wright, "Privacy-Preserving Distributed k-Means Clustering over Arbitrarily Partitioned Data," Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 593- 599, Aug. 2005.
[20] W. Jiang, C. Clifton, and M. Kantarco_glu, "Transforming Semi- Honest Protocols to Ensure Accountability," Data and Knowledge Eng., vol. 65, no. 1, pp. 57-74, 2008.
[24] M. Kantarcioglu and O. Kardes, "Privacy-Preserving Data Mining in the Malicious Model," Int'l J. Information and Computer Security, vol. 2, pp. 353-375, Jan. 2009.
[25] M. Kantarcioglu and R. Nix, "Incentive Compatible Distributed Data Mining," Proc. IEEE Int'l Conf. Soc. Computing/IEEE Int'l Conf. Privacy, Security, Risk and Trust, pp. 735-742, 2010.
[26] M. Kantarco_glu and C. Clifton, "Privacy-Preserving Distributed Mining of Association Rules on Horizontally Partitioned Data," IEEE Trans. Knowledge and Data Eng., vol. 16, no. 9, pp. 1026-1037, Sept. 2004.
[27] H. Kargupta, K. Das, and K. Liu, "A Game Theoretic Approach toward Multi-Party Privacy-Preserving Distributed Data Mining," Proc. 11th European Conf. Principles and Practice of Knowledge Discovery in Databases, pp. 523-531, Sept. 2007.
[28] J. Katz, "Bridging Game Theory and Cryptography: Recent Results and Future Directions," Proc. Fifth Conf. Theory of Cryptography, p. 251, 2008.
[29] G. Kol and M. Naor, "Cryptography and Game Theory: Designing Protocols for Exchanging Information," Proc. Conf. Theory of Cryptography, p. 320, 2008.
[30] R. Layfield, M. Kantarcioglu, and B. Thuraisingham, "Incentive and Trust Issues in Assured Information Sharing," Proc. Fourth Int'l Conf. Collaborative Computing: Networking, Applications and Worksharing, p. 113, 2009.
[31] X. Lin, C. Clifton, and M. Zhu, "Privacy Preserving Clustering with Distributed EM Mixture Modeling," Knowledge and Information Systems, vol. 8, no. 1, pp. 68-81, July 2005.
[32] Y. Lindell and B. Pinkas, "Privacy Preserving Data Mining," Proc. Int'l Conf. Advances in Cryptology (CRYPTO '00), pp. 36-54, Aug.2000.
[33] Y. Lindell and B. Pinkas, "Privacy Preserving Data Mining," J. Cryptology, vol. 15, no. 3, pp. 177-206, 2002.
[34] A. Lysyanskaya and N. Triandopoulos, "Rationality and Adversarial Behavior in Multi-Party Computation," Proc. Ann. Int'l Conf. Advances in Cryptology, pp. 180-197, 2006.

[35] R. McGrew, R. Porter, and Y. Shoham, "Towards a General Theory of Non-Cooperative Computation (Extended Abstract)," Proc. Conf. Theoretical Aspects of Rationality and Knowledge (TARK IX), 2003.
[37] M. Naor, B. Pinkas, and R. Sumner, "Privacy Preserving Auctions and Mechanism Design," Proc. First ACM Conf. Electronic Commerce, 1999.
[38] N. Nisan and A. Ronen, "Algorithmic Mechanism Design (Extended Abstract)," Proc. Ann. ACM Symp. Theory of Computing (STOC '99), pp. 129-140, 1999.
[39] S.J. Ong, D. Parkes, A. Rosen, and S. Vadhan, "Fairness with an Honest Minority and a Rational Majority," Proc. Sixth Theory of Cryptography Conf. (TCC), 2009.
[40] Y. Shoham and M. Tennenholtz, "Non-Cooperative Computation: Boolean Functions with Correctness and Exclusivity," Theoretical Computer Science, vol. 343, nos. 1/2, pp. 97-113, 2005.
[41] J. Vaidya and C. Clifton, "Privacy Preserving Association Rule Mining in Vertically Partitioned Data," Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (SIGKDD '02), pp. 639- 644, July 2002
[43] V.S. Verykios, E. Bertino, I.N. Fovino, L.P. Provenza, Y. Saygin, and Y. Theodoridis, "State-of-the-Art in Privacy Preserving Data Mining," SIGMOD Record, vol. 33, no. 1, pp. 50-57, 2004.
[44] A.C. Yao, "Protocols for Secure Computation," Proc. 23rd IEEE Symp. Foundations of Computer Science, pp. 160-164, 1982.
[45] A.C. Yao, "How to Generate and Exchange Secrets," Proc. 27th IEEE Symp. Foundations of Computer Science, pp. 162-167, 1986.

AUTHORS

Peethala Asha is currently working towards her M.Tech degree in Gokul Institute of Technology and Sciences, Piridi village, Bobbili mandalam ,Vizianagaram dt, A.P, India. Her research interests include Data Mining and cloud computing.

A.Ganesh Kumar is working as an Assistant professor in Gokul Institute of Technology and Sciences, Piridi village, Bobbili mandalam ,Vizianagaram dt, A.P,India. His main research interests are data mining and big data mining.

