Algorithm 7 classifyInstance(instId)

Require: m, the number of random trees built (let $nodeid_j$ represent the root node of the jth tree)

1: **for** $i = 1 \ldots p$ **do**
2: Randomly choose r from the range of random numbers for the cryptographic system
3: $lv_i \leftarrow Encrypt(0, r)$ {Random encryption of 0}
4: **end for**
5: **for** $j = 1 \ldots m$ **do**
6: $currstats \leftarrow retriveStats(instId, nodeId_j)$
7: **for** $i = 1 \ldots p$ **do**
8: $lv_i \leftarrow lv_i * currstats_i$
9: **end for**
10: **end for**
11: Collaboratively decrypt each lv_i and divide by m to get actual statistics

7. Conclusion and Future Work

In this paper, we studied the technical feasibility of realizing privacy-preserving data mining. RDTs can be used to generate equivalent, accurate and sometimes better models with much smaller cost; we are using distributed privacy-preserving RDTs. Our approach leverages the fact that randomness in structure can provide strong privacy with less computation. In the future, we plan to develop general solutions that can work for arbitrarily partitioned data and overlapping transaction.

References

[1] Jaideep Vaidya, Senior Member, IEEE, Basit Shafiq, Member, IEEE, Wei Fan, Member, IEEE, Danish Mehmood, And David Lorenzi "A Random Decision Tree Framework Or Privacy-Preserving Data Mining" Proc. IEEE Transactions On Dependable And Secure Computing, Vol. 11, No. 5, September/October 2014
[2] J. Vaidya, C. Clifton, and M. Zhu, Privacy-Preserving Data Mining.ser. Advances in Information Security first ed., vol. 19, Springer-Verlag, 2005.
[3] W. Fan, H. Wang, P.S. Yu, and S. Ma, "Is Random Model Better? On Its Accuracy and Efficiency," Proc. Third IEEE Int'l Conf. Data Mining (ICDM '03), pp. 51-58, 2003.
[4] W. Fan, J. McCloskey, and P. S. Yu, "A General Framework for Accurate and Fast Regression by Data Summarization in Random Decision Trees," Proc. 12th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '06), pp. 136-146, 2006.
[5] X. Zhang, Q. Yuan, S. Zhao, W. Fan, W. Zheng, and Z. Wang,"Multi-Label Classification without the Multi-Label Cost," Proc. SIAM Int'l Conf. Data Mining (SDM '10), pp. 778-789, 2010.
[6] A. Dhurandhar and A. Dobra, "Probabilistic Characterization of Random Decision Trees," J. Machine Learning Research, vol. 9, pp. 2321-2348, 2008.
[7] G. Jagannathan, K. Pillaipakkamnatt, and R.N. Wright, "A Practical Differentially Private Random Decision Tree Classifier," Proc. IEEE Int'l Conf. Data Mining Workshops (ICDMW '09), pp. 114-121,2009.
[8] J. Vaidya, C. Clifton, M. Kantarcioglu, and A.S. Patterson," Privacy-Preserving Decision Trees over Vertically Partitioned Data," ACM Trans. Knowledge Discovery from Data, vol. 2, no. 3,pp. 1-27, 2008.
[9] O. Goldreich, "General Cryptographic Protocols," The Foundations of Cryptography, vol. 2, pp. 599-764, Cambridge Univ. Press, 2004.

Author Profile

Prof. Vina M. Lomte, received the B.E. degree in Computer Science and Engineering from Amravati University, BNCOE, Pusad and M.E. degree in Computer Engineering from Mumbai University, MGMCET, Kamothe. Currently working as Assistant Professor of Computer Engineering Department in RMD SSOE Pune, Maharashtra, India.

Hemlata B. Deorukhakar received the B.Tech. Degree in Computer Science and Engineering from IGNOU, New Delhi in 2013. Currently appearing M.E. 2nd year Computer Engineering in RMD SSOE Pune, Maharashtra, India