A Novel Approach for Preserving Privacy in Malicious Model

Mr.Abhishek B.Kale[1]
[1]M.Tech,Department of Computer Science & Engineering
Priyadarshini Bhagwati College of Engineering,
Nagpur(M.S),India
abhishekkale1121@gmail.com

Ms.A.A.Nikose[2]
[2]Asst.Professor, Department of Computer Science &
Engineering, Priyadarshini Bhagwati College of
Engineering, Nagpur(M.S),India
archunikose@rediffmail.com

***Abstract*—** Data mining is a well-known technique for automatically and intelligently extracting information or knowledge from a large amount of data, however, it can also disclose sensitive information about individuals compromising the individual's right to privacy. Most of the privacy preserving solutions based on cryptography proposed by different researchers are in semi-honest model approach, where participating parties always follow the protocol. But providing stronger solutions considering malicious approach would be more useful for many practical applications because it tries to protect a protocol from arbitrary malicious behaviour using cryptography tools. An novel approach having a new protocol for privacy preserving in multiparty communication is designed for maintaining data integrity & implemented in malicious approach by using threshold homomorphic encryption & non interactive zero knowledge protocols according to real world paradigm.

Keywords*—*malicious model, privacy preserving,threshold homomorphic encryption

*****

I. Introduction

Data mining is a well-known technique for automatically and intelligently extracting information or knowledge from a large amount of data, however, it can also disclosure sensitive information about individuals compromising the individual's right to privacy [1]. Privacy preserving data mining is a novel research direction in data mining. In order to make a publicly available system secure, we must ensure not only that private sensitive data have been trimmed out, but also to make sure that certain inference channels have been blocked as well. A number of effective methods for privacy preserving data mining have been proposed [2-10]. But most of these methods might result in information loss and side-effects in some extent, such as data utility-reduced, data mining efficiency-downgraded, etc. Privacy can be invaded by the adversaries. In cryptography literature, adversaries have been classified into the following two categories in general.

Semi-honest adversary: A semi-honest adversary follows the protocol specification faithfully and tries to learn extra information from the message transcript during the execution.

Malicious adversary: A malicious adversary does not follow the protocol specification and can alter the input from the message transcript during the execution. On the basis of the above specifications, we have constructed our protocol in malicious model so that we can provide stronger security guarantee.

On the basis of the above specifications, we have constructed our protocol in malicious model so that we can provide stronger security guarantee.

II. Related Work

Most of the existing privacy-preserving *k*-means schemes (either two-party or multiparty "privacy-preserving distributed mining of association rules on horizontally partitioned data", kantarcioglu and clifton [1] & "privacy preserving data mining" lindell, yehuda, pinkas [2] deal with the semi honest model. Also they have some security problems "a secure protocol for computing dot-products in clustered and distributed environments", i. Ioannidis, a. Grama, m.j. atallah [4] & "privacy preserving naive bayes classifier for vertically partitioned data", vaidya and clifton introduced a solution [6] based on t. Okamoto and s. Uchiyama, "a new public-key cryptosystem as secure as factoring" homomorphic cryptosystem [13] which is not secure without random padding. Jagannathan et al.'s proposal [7] applies a scheme [8] whose security is not proven and the whole protocol is inefficient in clustering update computation.

"Privacy-preserving distributed k means clustering over arbitrarily partitioned data" , G. Jagannathan & R. N. Wright[14] proposed a scheme for secure scalar production protocol is used which has the serious weakness that a leakage of some database entries, can reveal the whole database as well as it is very inefficient when the database is large. In "Non-cryptographic fault-tolerant computing in constant number of rounds of interaction" J. Bar-Ilan and D. Beaver[15] proposed circuit technique is used which is time consuming. In "Privacy preserving clustering in malicious model" S. Jha, L. Kruger, and P. McDaniel [16] the above problems are covered but none of these schemes can deal with malicious adversaries. The cost of communication in these protocols are mostly reasonable . So, we cannot ignore the matter of cost. In "A new privacy preserving distributed k-clustering algorithm" G. Jagannathan, K. Pillaipakkamnatt, and R. N. Wright [17] an idea about how to convert the *k*-means clustering process from semi-honest model to malicious model without giving any kind of details about the protocols. In "Privacy-preserving two-party k-means clustering via secure approximation" C. Su, F. Bao, J. Zhou, T. Takagi, and K. Sakurai [20] the protocols have been constructed considering more than two participating parties in malicious model. They have used secret sharing with code based identification scheme. But secret sharing scheme is not efficient in some cases.

III. Proposed Work

A simple network is created of multiple nodes which perform data communication through specified protocol. Next, in the network a malicious node is detected by using weighted trust evaluation algorithm. A weighted-trust evaluation (WTE) algorithm detects the compromised nodes by monitoring its reported data. It is a light-weighted algorithm that would incur little overhead & detects misbehaved nodes accurately with very short delay. After detecting malicious node in the network, privacy preserving protocol is used this consists of four sub protocols which preserves the data integrity & data privacy from malicious nodes. The privacy preserving protocol is constructed using threshold homomorphic encryption & non-interactive zero knowledge protocol. The data preserved using privacy preserving protocol is needed to be routed to other nodes in the network, this is done by calculating shortest path between the surrounding nodes & malicious node using Dijkstra's algorithm to find the shortest path. The algorithm finds the path with lowest cost (i.e. the shortest path) between that vertex and every other vertex. It can also be used for finding costs of shortest paths from a single vertex to a single destination vertex by stopping the algorithm once the shortest path to the destination vertex has been determined.

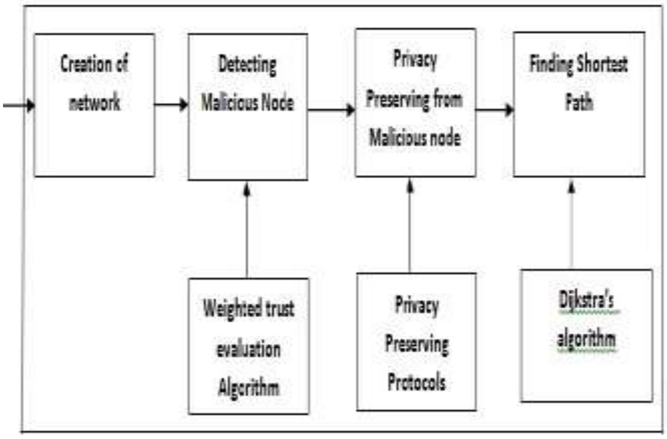
Creation of network
Detecting Malicious Node
Privacy Preserving from Malicious node
Finding Shortest Path
Weighted trust evaluation Algorithm
Privacy Preserving Protocols
Dijkstra's algorithm

Fig1: System Architecture

A. Creation of network nodes

A wireless sensor network (WSN) consists of a large number of small sensor nodes that are deployed in the area in which a factor is to be monitored. In wireless sensor network, energy model is one of the optional attributes of a node. The energy model denotes the level of energy in a mobile node. If the node is a sensor, the energy model should include a special component called "sense Power". It denotes the energy consumed during the sensing operation. In this module, multiple network nodes are created using Network Simulator2 (NS2). It consists of three types of sensor nodes Low-power "Sensor Nodes (SN)" with limited functionality, Higher-power "Forwarding Nodes (FN)" that forward the data obtained from sensor nodes to upper layer, "Access Points (AP)", or called "Base Stations (BS)" that route data between wireless networks and the wired infrastructure. Sensor nodes can be imagined as small computers. They usually consist of a processing unit with limited computational power and limited memory, a communication device and a power source.

B. Detecting malicious nodes

In this module, a malicious node is detected by using weighted trust evaluation algorithm. The malicious node may be defined as a node which does not follow the exact behavior. Most of the attacks are accomplished by modifying a message or simply not to forward the message which it is supposed to forward. A weighted-trust evaluation (WTE) algorithm detects the compromised nodes by monitoring its reported data. It is a light-weighted algorithm that would incur little overhead & detects misbehaved nodes accurately with very short delay. As shown in the fig 2, a weight *W* is assigned to each sensor node.Updating the weight of each sensor node has two purposes.First, if a sensor node is compromised (becomes a malicious node) and frequently sends its report inconsistent with the final decision, its weight is likely to be decreased. Then if a sensor node's weight is lower than a specific threshold, we can identify it as a malicious node. Second, the weight also decides how much a report may contribute to the final decision.

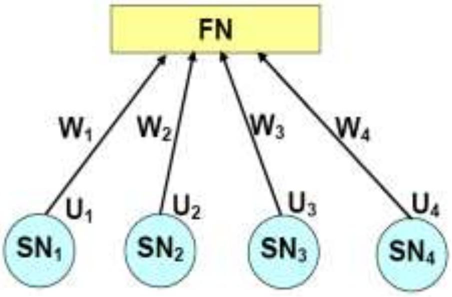

Fig 2 : A weight based network

C. Privacy preserving in malicious model

In this module, privacy preserving protocol is used consisting of four sub protocols which preserves the data integrity & data privacy from malicious nodes. The privacy preserving protocol is constructed using threshold homomorphic encryption which allows specific types of computation to be performed on encrypted data without compromising the encryption & obtain encrypted result & non-interactive zero knowledge protocol to prove that the actions taken by the party are correct.

(I) Secure Data Standardization Protocol Checks correctness problem from the malicious node by calculating mean & standard deviation between the two parties thus getting the standardized data. In case of multi-attribute data, it computes the distance function using the Euclidean distance.

(II) Secure Distance Measuring Protocol using threshold encryption computes distance between two closest nodes with their secret keys & private data input. The distances between a data object & all other nodes are found as a distance dataset.

(III) Secure Clustering Update Communication Protocol assigns data object to closest nodes by finding out the most proper clustering center for each data object .When data objects have been assigned to their nearest clusters, the cluster shares have also been changed. The protocol is used to calculate new center of cluster.

(IV) Secure Iteration Stopping Protocol stops the iterations after the all the provisions for securing privacy of data is done & output satisfies our requirements this is done by running the iterations until the difference of Euclidean distance between two consequent calculations.

D. *Finding shortest path*

It finds the shortest path between the malicious node & neighboring nodes to maintain the data privacy & route the data to other nodes using Dijkstra’s shortest path algorithm. For a given source vertex (node) in the graph, the algorithm finds the path with lowest cost (i.e. the shortest path) between that vertex and every other vertex. It can also be used for finding costs of shortest paths from a single vertex to a single destination vertex by stopping the algorithm once the shortest path to the destination vertex has been determined. It picks the unvisited vertex with the lowest-distance, calculates the distance through it to each unvisited neighbor, and updates the neighbor's distance if smaller & mark visited when done with neighbors.

E. *Simulation Results*

In order to evaluate the performance of the approach, the Network Simulator (NS2) version 2.32 with wireless extension is used. NS-2 is a discrete event simulator targeted at networking research. It is one of the most popular simulation packages used in the literature because of its availability (free license), and possibility to implement and test new protocols and applications . NS-2 is a commonly used package to simulate several schemes (such as routing and multicast protocols) in wired and wireless domains as it supports several networking protocols and standards . Also, it implements modules for wireless stations, which can move in a 2D environment.

To simulate the network, an OTcl script is used to initiates the event scheduler, and sets up the network topology using the network objects.

a. Simulation Environment

In the simulation experiments, the underlying scheme is used with DSR. Moreover, a network with 1000m x 800m area and 25, 50 and 100 mobile nodes was simulated. The simulation time is 1000 seconds. The mobile nodes move within the network space according to the Random Waypoint model. The communication patterns used are 6 Constants Bit Rate (CBR) connections with a data rate of 10 packets per second. 10%, 20%, 30% and 40% of the total number of nodes were chosen randomly as malicious nodes. Those malicious nodes drop, alter and forward packets

b. Performance Metrics

For evaluating the Scheme , three commonly used metrics are used to evaluate the performance of the three protocols and Packet Delivery Ratio (PDR), Routing Overhead, and Average Delay to show the strength of the proposed scheme.

Table 1:Simulation Parameters.

Simulator	**NS-2**
Simulator duration	1000 sec
Simulator area	1000mm * 800 mm
Number of nodes	25,50 Nodes
Transmission range	250 m
Movement model	Random waypoint
Traffic type	CBR
Data size	512 bytes
Packet rate	10 pkt /sec
Maximum speed	5 m/s
Pause time	60 sec

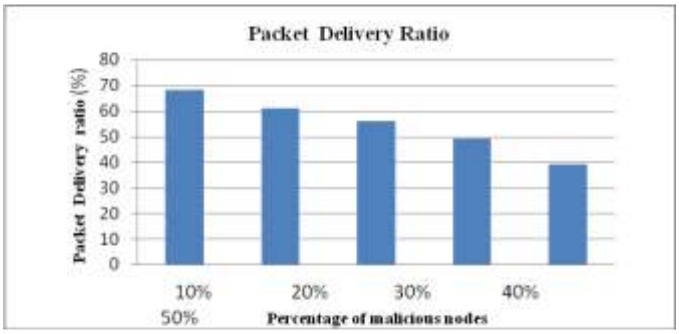

Figure 3: Packet Delivery Ratio

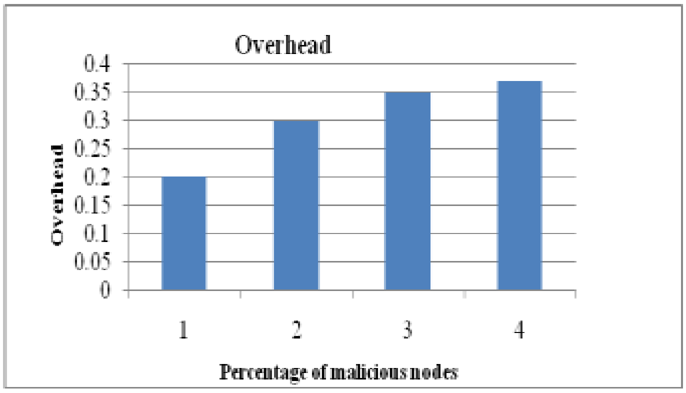

Figure 4: Overhead

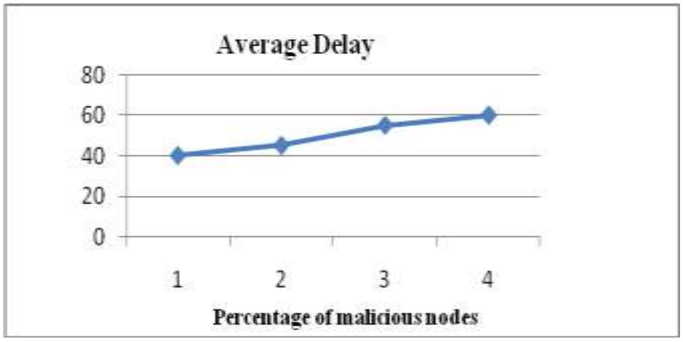

Figure 5:Average Delay

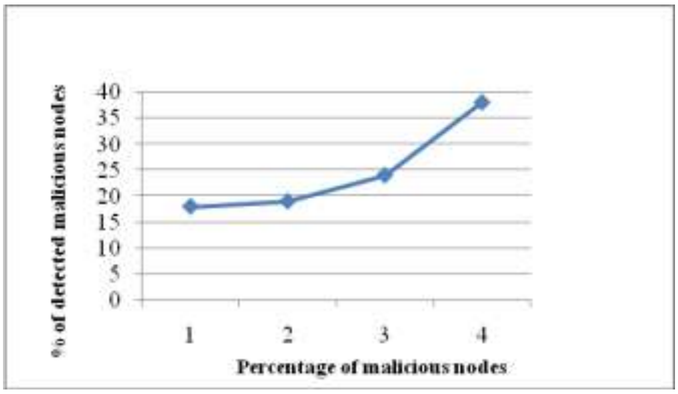

Figure 6:Detected malicious nodes(%)

F. Conclusion

We have proposed a new scheme based on NIZK proofs for privacy-preserving between two parties in malicious node. Most of the existing privacy-preserving schemes deal with the semi-honest model having various security and correctness problems. Ours is a challenging work because it guarantees stronger security and has overcome such problems. Moreover, we have minimized the communication cost. As ours is a two-party model, it will be a great future work in considering more than two parties in the presence of malicious adversaries.

G. Future work

The protocols have been constructed considering more than two participating parties in malicious model. They have used secret sharing with code based identification scheme. But secret sharing scheme is not efficient in some cases. A stronger security guarantee with reasonable cost is needed to achieve with higher reliability & optimum efficiency. Further, the investigation can be done on security and improvement proposal for better secure communication in network environment.

E. References

[1] M. Kantarcioglu, C. Clifton, “Privacy-Preserving Distributed Mining of Association Rules on Horizontally Partitioned Data”, IEEE Transactions on Knowledge and Data Engineering, vol.16, no.9, pp.1026-1037, 2004.

[2] Lindell, Yehuda, Pinkas, “Privacy preserving data mining”, In Proceedings of the Advances in Cryptology–CRYPTO, pp.36–54,2000.[3] J. Vaidya, C. Clifton, “Privacy Preserving Association Rule Mining in Vertically Partitioned Data”, In Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.639-644, 2002.

[3] J. Vaidya, C. Clifton, “Privacy Preserving Association Rule Mining in Vertically Partitioned Data”, In Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.639-644, 2002.

[4] I. Ioannidis, A. Grama, M.J. Atallah, “A Secure Protocol for Computing Dot-Products in Clustered and Distributed Environments”, In Proceedings of the 31st International Conference on Parallel Processing, pp.379-384, 2002.

[5] W.L. Du, Z.J. Zhan, “Building Decision Tree Classifier on Private Data”, In Proceedings of the IEEE International Conference on Data Mining Workshop on Privacy, Security, and Data Mining, pp.1-8, 2002.

[6] J. Vaidya, C. Clifton, “Privacy Preserving Naive Bayes Classifier for Vertically Partitioned Data”, In Proceedings of the 2004 SIAM International Conference on Data Mining, pp.522–526, 2004.

[7] J. Vaidya, C. Clifton, “Privacy-Preserving k-Means Clustering over Vertically Partitioned Data”, In Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.206–215, 2003.

[8] Rahena Akhter, Rownak Jahan Chowdhury,”Privacy-Preserving Two-Party *k*-Means Clustering in Malicious Model” In Proceedings of the 2013 IEEE 37th Annual Computer Software and Applications Conference Workshops, pp.121–126,2013.

[9] T.-K. Yu, D. T. Lee, S.-M. Chang, and J. Zhan, “Multi-party k-means clustering with privacy consideration,” in *ISPA*, 2010, pp. 200–207.

[10] C. Su, F. Bao, J. Zhou, T. Takagi, and K. Sakurai, “Security and correctness analysis on privacy-preserving *k*-means clustering schemes,” *IEICE Transactions*, vol. 92-A, no. 4, pp. 1246–1250, 2009.

[11] J. Vaidya and C. Clifton, “Privacy-preserving *k*-means clustering over vertically partitioned data,” in *KDD*, 2003, pp. 206–215.

[12] T. Okamoto and S. Uchiyama, “A new public-key cryptosystem as secure as factoring,” in *EUROCRYPT*, 1998, pp. 308–318.

[13] G. Jagannathan and R. N. Wright, “Privacy-preserving distributed kmeans clustering over arbitrarily partitioned data,” in *KDD*, 2005, pp.593–599.

[14] J. Bar-Ilan and D. Beaver, “Non-cryptographic fault-tolerant computing in constant number of rounds of interaction,” in *PODC*, 1989, pp. 201–209.

[15] S. Jha, L. Kruger, and P. McDaniel, “Privacy preserving clustering in malicious model” in *ESORICS*, 2005, pp. 397–417.

[16] G. Jagannathan, K. Pillaipakkamnatt, and R. N. Wright, “A new privacy preserving distributed k-clustering algorithm,” in *SDM*, 2006.