www.ijecs.in
International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 5 Issue 10 Oct. 2016, Page No. 18338-18342

Privacy Preserving in Data Mining to Protect Privileged information in Ad-hoc Network

Dillip Kumar Swain[1] , Sarojananda Mishra[2], Sasmita Mishra[3]
Department of CSEA, IGIT Sarang, Odisha, India, dkswain@yahoo.com[1]
Department of CSEA, IGIT Sarang, Odisha, India, sarose.mishra@gmail.com[2]
Department of CSEA, IGIT Sarang, Odisha, India, sasmita.mishra.csea@gmail.com[3]

Abstract:

Privacy of data is one of the major issues in every network. The network may be wired or wireless and the network may be ad-hoc, pear to pear or distributed network privacy of data is an important factor in each case. In each case in a distributed or shared network all people connected in a network work with the protocol. But there is no guarantee of security threat and data lost or stolen. So precautions needed in each case to maintain the network and to maintain the private data. An Ad-hoc network is a Local Area Network (LAN) which is built spontaneously as devices connect. Instead of depending upon a base station to manage the dataflow, each node in a network forward data packets to each other. As all most all nodes are responsible for data flow, so the probable data lost or stolen of information cannot be neglected. In order overcome these problems in data mining there should be some proper protocols as well as some special precautions should be taken. There are many experiments and developments already done in this aspect to secure data as well as to secure the network. This paper provides a review of various security challenges that are already taken for protection of privileged information in various peer to peer network, shared or distributed network. We have also proposed a technique for privacy preservation of privileged information in case of an Ad-hoc network. This technique may provide a better security for the privileged information in an Ad-hoc network.

Keywords: LAN, Ad-hoc Network, Data Mining, Data packets, Protocol

I. INTRODUCTION

Data mining can be called as the knowledge discovery. It is one of the most popular topics for researcher now a day. There are several applications of data mining in the field of information sharing and networking. If we consider the term networking it means there is the sharing of data among different nodes or computer. So the security of data cannot be neglected in any network like distributed or ad-hoc network. Privacy preserving for data mining for the privileged and private information is one of the major issue now a days. There are several institutions or several organizations which are now being connected in the form of Local Area Network. Several organizations are also using ad-hoc networks for better utilization of resources as well as on demand service for the real time necessity of nodes. Privacy preserving technique is becoming more popular in the field of shared and ad-hoc networks where the data is being shared by several users. As several users are concerned with the data sharing so preservation of private data as well as security of privileged information is very much necessary in each networking. The parties associated with the network may be honest or not. So parties may unknowingly or intentionally take part in fraudulent activities in case of a multiparty network. So there is the necessity of a proper and secure algorithms, technique or technology to avoid these types of security issues. The ad-hoc networks are generally developed for sharing of data and information in a LAN with on demand basis. The ad-hoc networks depend upon the technology of forwarding data packet form node to node. These data packets may be in different sizes depending upon the protocol or topology of the network but in these matter fraudulent activities to the privileged information should be avoided. In this paper we have discussed several developments that are already done for the privacy preserving in data mining as well as various algorithms and techniques that are being applied in this field. We have also proposed a new technique for privacy preserving in ad-hoc networks. It will helpful for securing privileged information in different nodes and it may provide a better security to these ad-hoc local area networks.

This paper is organized as follows. In section II we have provided a brief literature review on various technologies or developments that are being developed

for privacy preservation in data mining. In section III we have discussed about the working techniques of Ad-hoc network and its association with data mining. We have proposed a technique for privacy preservation of privileged information in case of an Ad-hoc network in section IV. Finally the paper concluded in section V with conclusion and our future step.

II. PRIVACY PRESERVATION IN DATA MINING

Data mining techniques can be defined as identifying trends and patterns from huge amount of data. Data mining and Data warehousing both terms are associated with each other. This process is mostly gathering huge amount of data into a common platform and then running an algorithm to find the useful and meaningful information. In [1] K. Saranya et al. provides a brief review on privacy preservation in data mining. Generally when an honest node in a network wants to access the private information of another node it can be treated as a fraudulent activity. Then the node can be considered as a dishonest or semi-honest node. In generally the size of distributed network and the presence of heterogeneous node create privacy preservation most complicated. The intention of the participating nodes depends upon their activities. In [2][11] they have also discussed about the various approach and necessity of privacy preserving in data mining.

Cryptographic techniques are the most useful techniques for securing data from fraudulent activities. By using cryptographic techniques for storing sensitive data and providing access to these data are mostly secure. In[3] they have applied cryptographic techniques for privacy preservation in data mining. In this they have discussed results of a part of cryptographic research how number of nodes can jointly communicate any function without interfering to one another information. Security problem can be solved if two nodes access the same function without accessing their personal information. A cryptographic technique can provide this facility. Oblivious transfer is the basic principle of secure computation. The oblivious polynomial evaluation" (OPE) contains a sender and a receiver. The senders input is a polynomial Q of degree k over some finite field f and the receiver input is an element z ε f (the degree k of Q is public). The intension of the technique is that the receiver obtains Q (z) without accessing anything about the polynomial Q, similarly the sender also learns nothing. The problem is considered as the private computation of the function (Q, z) → (λ, Q (z))[3]. In this they have considered two cases that is when the condition is two party case and multiparty case.

In the information sharing and organizing sector managing the unstructured data is a major issue. Unstructured data may not have a particular data model or it is not possible to access by data mining. In [4] they have provided a frame work for these unstructured data. In this paper they have proposed a technique for these unstructured data. So in this case the unstructured data is being converted to structured data. In association with conversion of the unstructured data to structured data these unstructured data are being converted to Xml and node identification and storage with large amount of data. After conversion of unstructured to structured data they have proposed a model designed using VB.NET for privacy preservation of Meta data. In the implementation and testing purpose they have used two text documents. The total occurrence words in the text document are designed by Rapid Mirror tool. In the example data set they have taken two examples, four special attributes, 2199 regular attributes. By considering the two documents as m1 and m2 equation is applied for calculation of result.

$$Z_c = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)[4]$$

Where in equation (1) Z_c is the value of standard normal variate. The $\bar{x}$ and $\bar{y}$ are the mean values of the words of occurrences in the two documents. The σ_1^2 *and* σ_2^2 are the variances of the words occurrences in the documents. n_1 and n_2 considered as the number of regular attributes of documents m1 and m2 respectively[4].

In case of decentralized network privacy preservation is also a major issue. There are various nodes in a decentralized network. Some nodes are honest and some nodes are semi-honest. There may be a penalty issue for these semi-honest nodes. Identification of these semi-honest nodes and applying a penalty for these semi-honest nodes is very much necessary. So in [5] they have discussed an alternative technique to recognize these nodes who creates disturbance. They have focused on privacy preservation with penalty upon a decentralized network in a multiparty computation. The assumption of semi-honesty in participant behavior is sub-optimal and in this paper they propose a penalty based mechanism for a series of secure average computations. They have applied an algorithm for this purpose to identify these semi-honest nodes. Sometimes the penalty is very harsh that the semi-honest parties may force to terminate from the system if they violate the protocol repeatedly.

According to this proposed technique the following rules may be implemented in the framework to penalize the semi-honest party.

Rule-1: The party who violates the protocol, it has threshold chances to participate in the system. As the protocol the party must bound to change their strategies if it disobeys the rule.

Rule-2: If the party tries to violate the protocol continuously beyond threshold value, it must be terminated from the system[5].

According to the simulation with 100 nodes and where nodes are randomly chosen as honest and semi honest they observe that after applying this penalty technique the number of fraudulent activity in a decentralized network is decreasing time to time. The figure 1. shows the simulation result.

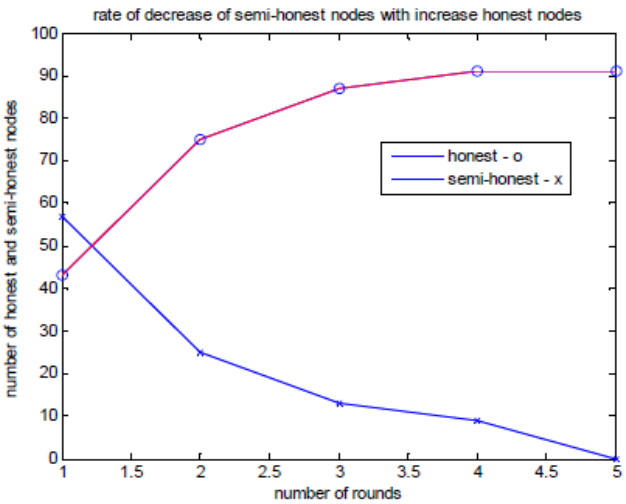

Figure 1. Change in number of bad nodes in the network over successive rounds [5].

In [6][8] they have discussed about privacy preserving in distributed data mining protocols. He has provided a brief review upon the developments of these protocols over a time period of 1999 to 2009 as well as performance matrix. In this paper he has discussed about Public-key encryption scheme, oblivious transfer protocol, Secret sharing scheme, Randomization techniques. He has also provides a brief review upon the various data mining algorithms like classification, association rule[10], clustering that are applied in various research articles [6].

Peer-to peer network is also a very popular network in the phenomenon of network topology. In this peer to peer network privacy preservation is also a major issue. As there are large number of parties or nodes are being associated with this network, fraudulent activity or misbehaving nodes may present in these network. Privileged data may be stolen or misguided by these fraud nodes. In [7] H.K.Bhuyan et al. discussed about the privacy preservation issue in a peer-to-peer network. This research basically focuses on developing secure computational model for privacy preservation of the privileged and valuable data by performing multiparty computation. According to the authors participating parties should attached to the coordinator of the peer-to-peer network with a specified path and maintain privacy by performing certain application.

III. DATA SHARING IN AD-HOC NETWORK

A mobile ad-hoc network is a collection of wireless mobile hosts which form a temporary
Network without the presence of any centralized administration or standard support services. In these networks nodes enter and leave the network spontaneously. There is no particular centralized control or fixed infrastructure to for configuration or reconfiguration for these networks. These are having limited communication range and reduce battery drain. Various network protocol like proactive and reactive protocols are exist for these mobile ad-hoc networks which deals with data communication and forwarding packets over node to node.

Ad-hoc networks are mostly used in military sector where these networks are being designed when necessary. In [13] they have provides a brief review about the various ad-hoc networks. Starting from familiar wireless network architectures, they have progressed towards independent operation of the network nodes for defining the notion of ad-hoc networking. Typically, these networks operate with distributed functions and allow traffic to pass over multiple radio hops between source and destination. Moreover they have discussed some of the typical properties of ad hoc networks, such as routing algorithms, the implications of radio layers etc. In this they have discussed about the IEEE 802.11 specification that deal with the wireless LAN communication [13].

Location based service is becoming more popular in the recent years for wireless and also mobile networks. In recent LBS systems, Service Providers (SPs) require users report and their accurate locations, which may be illegally used by adversaries to bypass sensitive information and data of users. So in this manner privacy concern discloser creates more and more problems for LBS. In [14] they have discussed about Privacy Preservation for Location-based Services in Mobile Networks. In this paper they have proposed a technique of pseudonym change and it works for non-

cooperative users with the absence of intermediate system. This is basically for ad-hoc anonymity[14].

In [15][16] they have discussed about vehicular ad-hoc networks. In these papers they have focused on security concern and preservation of privileged information. Vehicular Ad hoc Networks are having great influence with improving road safety as well as driving conditions. So approaching these necessities these VANETS are becoming more and more popular in the recent years. But though it provides very sensitive services to the society but severe fraudulent activities and severe malicious attacks cannot be neglected which may cause severe accidents. In [15] they have provided a brief review about the various security issues in VANETs like privacy preservation, message authentication, message non-repudiation, access control, liability identification, entity authentication etc. In [16] they have proposed a secure data downloading protocol with privacy preservation in vehicular ad hoc networks. In these applications vehicle sends requests regarding various issues like where is the nearest gas station. It may receive the information from a Road Side Unit. In this paper they have proposed a protocol which enables vehicles to download data securely from RSUs even if one or more RSUs may be in contact. According to this protocol it guarantees vehicles exclusive access to their privileged data without losing their private information of the vehicles [16].

IV. PRIVACY PRISERIVING IN AD-HOC NETWORK

An ad-hoc network is basically a small network that contains limited number of nodes. This may create a Local Area Network having coverage up to a few kilometers. Some nodes may create fraudulent activities in the network like stolen of private and privileged information. Generally the data transferred in the network in the form of data packets. In between the sender and receiver node there is the in between nodes which helps in forwarding data packets. These nodes may create unwanted activities in the data packets. So the data packets should have a two step encryption method with two keys like private and public key. The data packets can the decrypted using public key by the middle nodes to know the destination address and the final receiver node should have the private key to decrypt the original packet to get the necessary data. From the initial network setup for testing purpose we can take up to five nodes and we can calculate the average time by calculating the total time dividing by five. This average time should be captured from genuine nodes in a simulation atmosphere. If the average time is "A" the following algorithm can be applied to detect the malicious nodes.

Algorithm:

1. Use of two step encryption techniques for the encryption of data packets and use two keys one for public and private.
2. Destination address should be available to a node after decryption using public key.
3. The destination node should know the private key to decrypt the original packets.
4. In each node the receiving time and resending time should be captured in each time.
5. This time should be compared with "A" and be noted in each time.
6. The number of nodes visited by the packet should be tracked.
7. If the total number of node visited by the packet in "n" so the total time spent by the packet (Let it is "T") for reaching the destination can be compared with A x n.
8. If there is a comparable difference between these times we can check the individual time taken by the nodes for receiving and resending the nodes.
9. If a node takes more time to resend the packet we can check that node and this node can be suspected as a malicious node.

So in this way a suspected node can be identified and if it is found to be a malicious node or if the node is associated with fraudulent activities the node can be dismissed. Furthermore penalties and punishment can be taken for that particular node. The nodes found out using this technique may not associate with fraudulent activities as there may be the delay in network or other problems but these nodes can be considered as suspected node and further examining of these nodes can secure the network and provide better service.

V. CONCLUSION AND FUTURE WORK

Privacy preservation is a major issue in each type of network. Starting from wire to wireless and LAN to WAN in each network security of data is very much essential. Data mining is associated with every type of network. Instead of searching a large amount of data maximum network sharing a limited data as well as the basic data for public use or to share in the network. So instead of accessing the shared data some malicious node may create fraudulent activity and may steal the privileged information. So though ad-hoc networks are small networks still these types of suspicious activities cannot be ignored. So in this paper we have provided an algorithm and a proposed idea to remove fraudulent activities as well as to secure the privileged information in data mining of ad-hoc networking. In our future research we will provide a real time model as well as protocol in this aspect for privacy

preservation of data mining to secure privileged information in ad-hoc networks. Furthermore we will take the simulation environment to calculate the average time and we will apply this average time with one thousand nodes with randomly choosing malicious nodes.

REFERENCES

[1] K. Saranya, K. Premalatha, S. S. Rajasekar, "A Survey on Privacy Preserving Data Mining", Second International Conference on Electronics and Communication Systems, IEEE, 2015.
[2] M. Sharma, A. Chaudhary, M. Mathuria, S. Chaudhary, S. Kumar, "An Efficient Approach for Privacy Preserving in Data Mining", International Conference on Signal Propagation and Computer Technology, IEEE, 2014.
[3] Anand Sharma, Vibha Ojha, "Implementation of Cryptography for Privacy Preserving Data Mining", International Journal of Database Management Systems Vol.2 (3), 2010.
[4] V. Thavavel, S. Sivakumar, "A generalized Framework of Privacy Preservation in Distributed Data mining for Unstructured Data Environment", International Journal of Computer Science Issues, Vol. 9(1), 2012.
[5] H. K. Bhuyan, S. K. Dash, S. Roy, D. K. Swain, "Privacy Preservation with Penalty in Decentralized Network using Multiparty Computation", International Journal of Advancements in Computing Technology, Volume4(1), 2012.
[6] Zhuojia Xu, "Analysis of Privacy Preserving Distributed Data Mining Protocols", School of Engineering and Science, 2011.
[7] H. K. Bhuyan, N. K. Kamila, S. K. Dash, "An Approach for Privacy Preservation of Distributed Data in Peer-to-Peer Network using Multiparty Computation", International Journal of Computer Science Issues, Vol. 8(4), 2011.
[8] R. N. Wright, Z. Yang, S. Zong, "Distributed Data Mining Protocols for Privacy: A Review of Some Recent Results", MADNES'05, Springer LNCS, 2006.
[9] Adriano A. Veloso, W. Meira Jr, S. Parthasarathy, M. Bunte de Carvalho, "Efficient, Accurate and Privacy-Preserving Data Mining for Frequent Item sets in Distributed Databases", Article of Ohio-State University, 2003
[10] N. V. Muthu lakshmi, K. S. Rani, "Privacy Preserving Association Rule Mining in Horizontally Partitioned Databases Using Cryptography Techniques", International Journal of Computer Science and Information Technologies, Vol. 3 (1) , 2012.
[11] Yehuda Lindell, Benny Pinkas*y*, "Privacy Preserving Data Mining", Journal of cryptology ACM digital library, Vol. 15(3), 2000.
[12] W. Xiao-Dan, Y. Dian-min, L. Feng-li, W. Yun-feng, C. Chao-Hsein, "Privacy Preserving Data Mining Algorithms by Data Distortion", International Conference on Management Science and Engineering, IEEE, 2006.
[13] M. Frodigh, P. Johansson, P. Larsson, "Wireless ad hoc networking: The art of networking without a network", Ericsson Review(4), 2000.
[14] L. Junliang, Z. Yang, Y. Liu, "Ad-hoc Anonymity: Privacy Preservation for Location-based Services in Mobile Networks", 18th International Conference on Parallel and Distributed Systems, IEEE, 2012.
[15] S. Behera, B. Mishra, P Nayak, D. Jena, "A secure and efficient message authentication protocol for vehicular Ad hoc Networks with privacy preservation(MAPWPP)" 5th International Conference on Internet Multimedia Systems Architecture and Applications, IEEE, 2011.
[16] Y Hao, J Tang, C. Zhou, "Secure Data Downloading with Privacy Preservation in Vehicular Ad Hoc Networks", International Conference on Communications, IEEE, 2010.
[17] C. Clifton, M. Kantarcioglu, X. Lin, M. Y. Zhu, "Tools for Privacy Preserving Distributed Data Mining", SIGKDD Explorations, Vol. 4(2), 2006.