Volume 5, Issue 7, July 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Secure Communication Using Privacy Preserving in a Data Mining

Sandhya D. Patil, Prof. Nita M. Thakare, Prof. Sheikh Firoj
Department of Computer Technology, Priyadarshini College
Nagpur, Maharashtra, India

Abstract— This paper describes the problem of Privacy Preserving Data Mining (PPDM). Data mining is the process of extracting hidden information from the database. The current trend in business collaboration shares the data and mine results to gain mutual benefit. Privacy preserving data mining has become increasingly popular because it is allowing the sharing of private sensitive data for analysis purposes. It describes some of the common cryptographic tools and constructs used in several PPDM.

Keywords: — PPDM, anonymous id , privacy preservation, AIDA algorithm

I. INTRODUCTION

The word Privacy refers here the information about us that we feel is personal, confidential or private should not be unnecessarily distributed or publicly known Our personal or private information should not be misused (whatever that means) Our private information should be kept secure so any other person can not use it in a Bad manner.

Privacy preserving means maintaining private information secure from the intruders. When something is private to a *person*, it usually means there is something to them inherently special or sensitive.

The anonymous communication plays a vital role in internet's popularity for both personal and business purposes.. The disadvantages of sharing private data are being studied in detail. Other available applications like searching of patient medical records, maintaining security about social networking, electronic voting and many more.

To distinguish between anonymous communication and anonymous ID assignment, consider a situation where N parties wish to display their data, in N slots on a third party site, anonymous ID assignment method assigns N slots to the users whereas anonymous communication allows the users to conceal their identities In our network the identities of the parties are known but not the original identity.

Data mining refers to the techniques of extracting rules and patterns from data. It is also commonly known as KDD (Knowledge Discovery from Data). Traditional data mining operates on the data warehouse model of gathering all data into a central site and then running an algorithm against that warehouse. This model works well when the entire data is owned by a single custodian who generates and uses a data mining model without disclosing the results to any third party. However, in a lot of real life application of data mining, privacy concerns may prevent this approach.

Data mining is the process of extracting bulk of data from the large data bases. It is very difficult to retrieve the records of multiple data from large data and keeping it confidential so we use privacy preserving here. PPDM are preserving accuracy of the data and the generated models and the performance of the mining process while maintaining the privacy constraints.

The data is altered before delivering it to the data miner. The data is distributed between two or more sites, which cooperate using a semi-honest protocol to learn global data mining results without revealing any information about the data at their individual sites.

While using a model to classify data, the classification results are only revealed to the designated party, who does not learn anything else other that the classification. here we are going to implement an algorithm AIDA and with this algorithm we are providing security to large amount of data by assigning a private key to a authorized user and a public key to an unauthorised users .the use of anonymous id assisgnment technique which makes information confidential we are using security algorithms with database management system to keep records of hospitals confidential.

Data mining research deals with the extraction of potentially useful information from large collections of data with a variety of application areas such as customer relationship management, market basket analysis, and bioinformatics. The extracted information could be in the form of patterns, clusters or classification models.

In response to that, data mining researchers started to address privacy concerns by developing special data mining techniques under the framework of privacy preserving data mining. Opposed to regular data mining techniques, privacy preserving data mining can be applied to databases without violating the privacy of individuals.

Recent advances in data collection, data dissemination and related technologies have inaugurated a new era of research where existing data mining algorithms should be reconsidered from a different point of view, this of privacy preservation.

Privacy preserving data mining is a novel research direction in data mining and statistical databases, where data mining results are analyzed for the side-effects they in data privacy. The main consideration in privacy preserving data mining is two fold.

First, sensitive raw data should be modified or trimmed out from the original database, in order for the recipient of the data not to be able to compromise privacy.Second, sensitive knowledge which can be mined from a database by using data mining algorithms should also be excluded.

The main objective in privacy preserving data mining is to develop algorithms for modifying the original data in some way, so that the private data and knowledge remain private even after the mining process. In a nutshell,
the privacy preserving mining methods modify the data.

II. RELATED WORK

Existing and new algorithms for assigning anonymous IDs are examined with respect to trade-offs between communication and computational requirements. suppose that access to the database is strictly controlled because data are used for certain experiments that need to be maintained confidential ,which allows Alice directly to read the contents of the tuple breaks the privacy of Bob; the confidentiality of the database managed by Alice is violated once Bob has access to the contents of the database, without informing Alice and Bob know the contents of the tuple and the database respectively. Such functions are useful in data mining applications and also helps characterize the complexities of the secure multiparty computation. Our main algorithm will be based on a method for anonymously sharing simple data and results in methods for efficient sharing of complex data. So many applications are exists that require dynamic unique IDs for network.

The use of newtons and sturms algorithm minimizes the problems of an existing system but they are not going to implement a complex data structure, no doubt they are using secure sum function but still they are having data accessing problem as data goes on increasing day by day.

Every time they need to secure sum algorithm as data is increasing that increases the overhead of calculating sum and everytime display function of it. so it is directly affect on accessing and calculation of n data among various inputs with respect to their output.So to minimize this overhead we are going to implement an aida alogirithm with ecc curve that directly calculate the records of the data without using secure computation function and without using newtons and sturms theorems.

That means we are going to exploring the four stages stages work into in only one step, which reduces overhead and saves times to calculate a frequent number of time secure sum function.

III. PROPESD PLAN OF A WORK

Our proposed work will be worked in four phases as follows:

Phase 1: In this phase the system is dealing with implementation of different security algorithms and processing of data which is a communication module.

Phase 2: Creating an application to let the user sharing the data through secure channel and assignment of different IDs. we are going to implement data base connection.

Phase 3: This phase consists of the implementation and performance evaluation of the approach. Effectiveness of our propose approach will be measured in terms of accuracy. Evaluate its performance in terms of localizable rate, location error and computational overhead. That is number of users are going to sharing plus providing security.

Phase 4: we are analyzing how system works that is each users performance of quering to server**.**

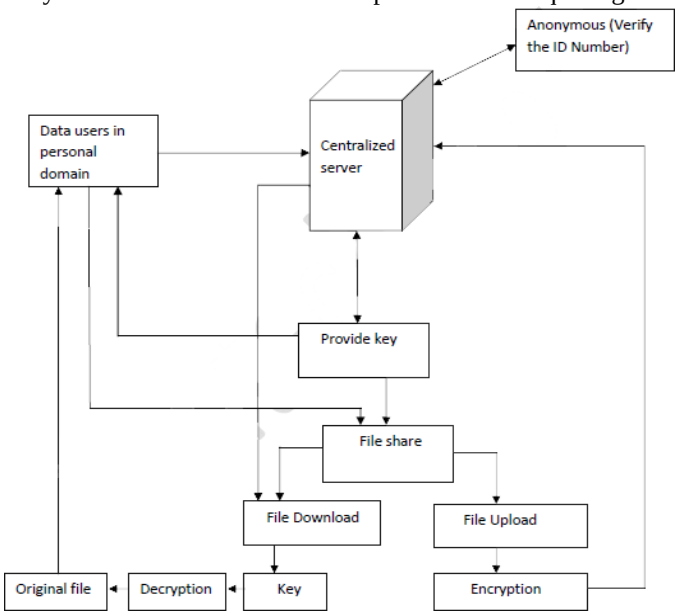

Fig.1. Actual working of a system

IV. WORKING OF PROPOSED SYSTEM

we are implementing a system that is basically work on client server architecture. As shown in above figure the centralised server is a data central server that containing the information about the all medical records of the patient.The

clients who want to access the information must be authorized themselves with a private key provided by the server to which the server provides private key.

After requesting from the client a server must show the current activity of the clients cpu like usage of cpu, how many times of request are coming from the client etc. The server must also show the information about client like IP address of a client, cpu usage, requesting load of cpu etc.

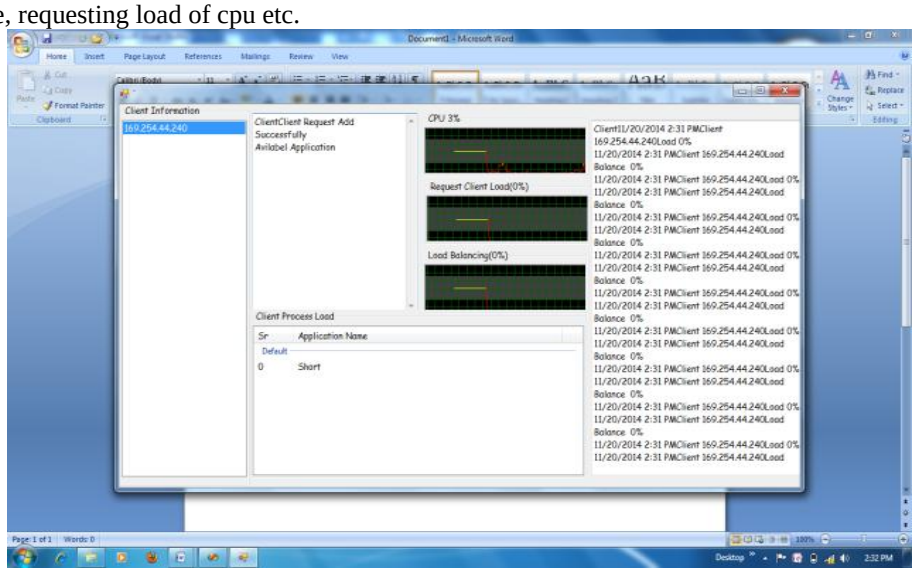

Fig2 output of a server side.

The server will maintained all processes of a client, when we are entering a bulk of data it will properly dividing the data into databases and will show output accordingly. It will shows all client information related to a particular client.

The client side shows following window

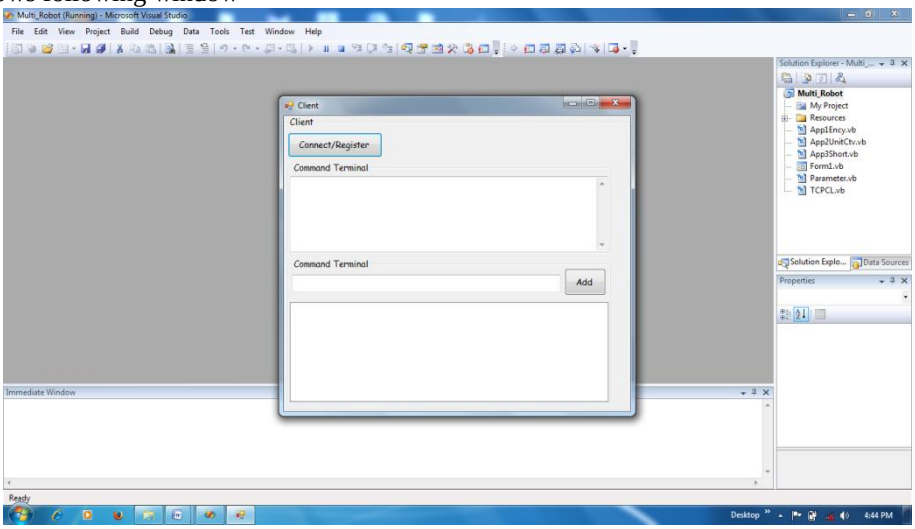

Fig.3 output of a client side

Before establishing connection client should have to connect with server. The server keeps records of all clients which are connecting to it.

The work reported in this paper further explores the connection between sharing secrets in an anonymous manner, distributed secure multiparty computation and anonymous ID assignment. The use of the term "anonymous" here differs from its meaning in research dealing with symmetry breaking and leader election in anonymous networks. Our network is not anonymous and the participants are identifiable in that they are known to and can be addressed by the others. Methods for assigning and using sets of pseudonyms have been developed for anonymous communication in mobile networks. The methods developed in these works generally require a trusted administrator, as written, and their end products generally differ from ours in form and/or in statistical properties.

V. CONCLUSION

By using above method we are going to implementing a system that is work on aida algorithm with the use of ecc curve directly, and exploring the usage of ecc curve to accessing medical records. This paper isbasically used for a survey of medical patients for a particular disease. Acces is made easy for a particular patient with efficient and in a secure manner.

REFERENCES

[1] Sarbanes–Oxley Act of 2002, Title 29, Code of Federal Regulations,Part 1980, 2003.[1]
[2] White Paper—The EssentialGuide toWeb AnalyticsVendor Selection, IBM [Online]. Available: http://measure.coremetrics.com/corem/getform/reg/wp-evaluation-guide[2]
[3] Shamir, "How to share a secret," Commun. ACM, vol. 22, no. 11,pp. 612–613, 1979.[3]
[4] Friedman, R. Wolff, and A. Schuster, "Providing k-anonymity "[4]
[5] F. Baiardi, A. Falleni, R. Granchi, F. Martinelli, M. Petrocchi, and A.Vaccarelli, "Seas, secure e-voting protocol: Design and implementation,"Comput. Security, vol. 24, no. 8, pp. 642–652, Nov. 2005.[5]
[6] D. Chaum, "Untraceable electronic mail, return address and digitalpseudonyms," Commun. ACM, vol. 24, no. 2, pp. 84–88, Feb. 1981.[6]
[7] Q. Xie and U. Hengartner, "Privacy-preserving matchmaking for mobilesocial networking secure against malicious users," in Proc. 9th Ann. IEEE Conf. Privacy, Security and Trust, Jul. 2011, pp. 252–259.[7]
[8] O. Goldreich, S. Micali, and A. Wigderson, "How to play any mentalgame," in Proc. 19th Ann. ACM Conf. Theory of Computing, Jan. 1987,pp. 218–229, ACM Press.[8]

