ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

SMC and Proxy server: Privacy Preserving Data Sharing over DDS

Tinimol Andrews, Nimmy Manuel

M. Tech Student, Dept. of CSE, Mangalam College of Engineering, Kottayam, Kerala, India

Assistant Professor, Dept. of CSE., Mangalam College of Engineering, Kottayam, Kerala, India

ABSTRACT:Data should collect safely and accurately while we consider both the performance and privacy of data. Distributed data sharing over providers requires a great challenge over attackers. Data demander(authorised user's) requesting data from multiple distributed data providers, the providers access the data from different base stations. Identity of providers are unknown to the demanders. Noise mixed with the data during transmission and at the reception side, efficient data recovery is carried out. AES half encryption is done in provider side and key server perform authentication function. Shadow matrix computation performed over data for efficient transmission of data. Experimental results shows that privacy preservation of data over DDS (Distributed data stream) using SMC (Shadow matrix computation) and proxy server is more efficient and time consuming compared to traditional approach.

KEYWORDS: Proxy server, distributed data sharing, shadow coding, privacy preserving, data mining.

I. INTRODUCTION

In a distributed environment, collecting data from multiple providers is a challenging task. Data should be collected and stored in secure and intellectual way. Data demander, Multiple providers, Base stations are the basic framework of a distributed environment. A key generator or a server in the middle of demander and provider perform the authentication function. In a normal distributed environment, key server act as an exchange operator between demander and provider. When the demander request file to the provider using demander key, authentication key kept in user side. At that time the provider checks the key and requested file sent to demander from base stations. In the traditional environment, sometimes the key server may become an attacker and catch the demander key in an unauthorised way and request file to provider and access file in a normal way.

In the proposed approach, proxy server act as an authenticated key server. Proxy server not only perform a bridge between the demander and provider but also perform a half encryption on the file for secure transmission. Shadow coding used for privacy preserving and efficient transmission and recoverability of data over distributed data streams. Proxy controls all providers and all files to transfer. Each authenticated users can have the right to access the right files through the authenticated proxy. It provides a fast data recovery within short time and data privacy ensured by double encryption in provider and proxy on files and single decryption at demander side.

II. RELATED WORK

The rapid growth of graphs raises big challenges to the database community. Nowadays, graph structured data are used in numerous applications(e.g.,web graphs,social networks, biological and chemical pathways, transportation networks). High efficiency of graph operations is essential to applications. However, even primitive operations[1] on a graph can be very time-consuming due to the complexity of structural connectivities and graph size. Moreover, real graph datasets are growing rapidly in size, making the attainment of high efficiency even harder.

A table entry[2] is either empty, or it contains one of records, in which case it is fun. It can look for a record with a given key by exhaustively examining all entries of the table. Similarly, a new record can be inserted into the table by searching for an empty position. It is clear that the searches in question can become quite protected for a large collection of records.

Privacy is becoming an increasingly important issue in many data mining applications that deal with health care, security, financial, behavioral, and other types of sensitive data. It is particularly becoming important in counter-terrorism and homeland defense-related applications. These applications may require creating profiles, constructing social network models, and detecting terrorist communications among others from privacy sensitive data[3].

Propose an algorithm to securely integrate sensitive data[4] , which is horizontally divided among two parties over the same set of attributes, whereby the integrated data still retains the essential information for supporting data mining tasks. The following scenario further motivates the problem. At the same time, new knowledge that results from the integration process should not be misused by adversaries to reveal sensitive information that has not been available before the data integration.

Explosive progress in networking, storage, and processor technologies has led to the creation of ultra large databases [5]that record unprecedented amount of transactional information. In tandem with this dramatic increase in digital data, concerns about informational privacy have emerged globally. The concerns over massive collection of data are naturally extending to analytic tools applied to data. Data mining, with its promise to efficiently discover valuable, non-obvious information from large databases, is particularly vulnerable to misuse.

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 private knowledge remain private even after the mining process. The problem that arises when confidential information can be derived from released data by unauthorized users is also commonly called the “database inference” problem[6].

A user can have a clear estimation of the knowledge that an attacker can know about him. The knowledge an attacker uses to find the privacy information of a user is called the background knowledge[7]. To provide different levels of privacy protection, allow users to set personalized privacy requirements based on their own assumptions about the attacker’s background knowledge.

III. PROPOSED METHOD

Data collection is an important task over distributed environment. Data demander request data to providers. Provider collect the data from multiple base stations. Provider sent data along with noise to demander in a half encrypted form. The key server in the middle perform half more encryption using AES algorithm. Keys are shared via a secured channel over UDP connection. Several techniques are widely accepted to protect the privacy of individual data. Noise can be added along with data to improve security and a high cryptographic algorithms are needed to decrypt the noisy data.

PERTURBATION

To achieve improved privacy over the data to send, adding noise to the actual data. The collected data at demander side remove these added noise and perform decryption to achieve the actual data. There is a less chance of attack to the data when it contain added noise. Provider collect the data from multiple base stations and the location of providers and base stations are hidden to the demander. Better performance can achieve by these perturbation technique and data can be transmitted and recovered in a more secure way.

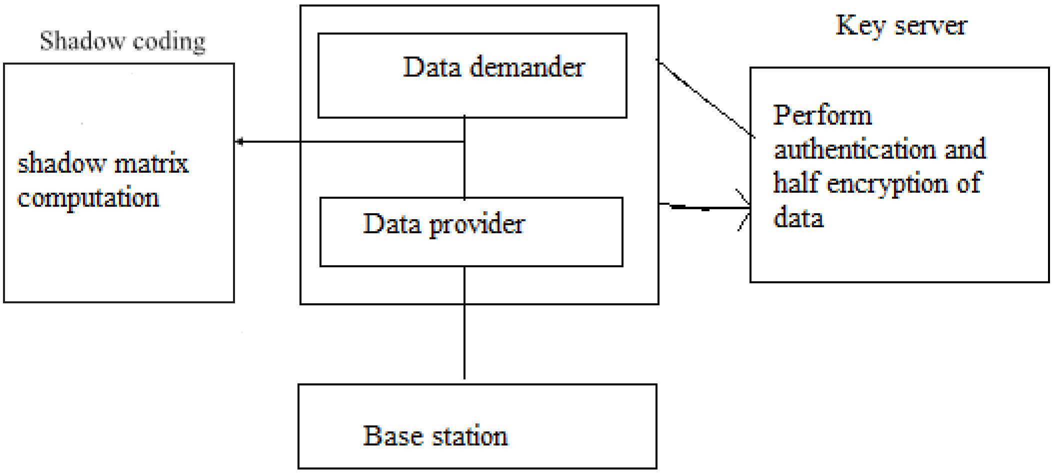
ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

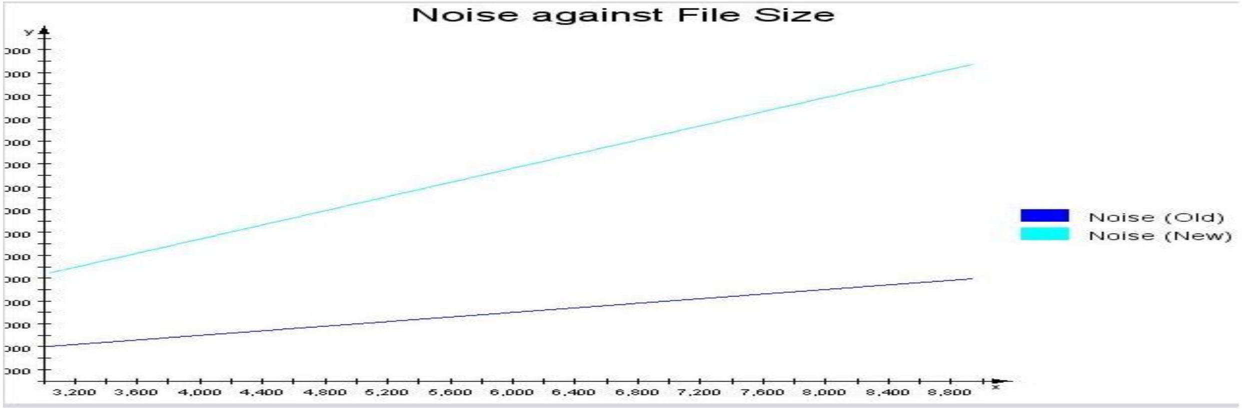
IV. ARCHITECTURE


Shadow coding
shadow matrix computation
Data demander
Data provider
Key server
Perform authentication and half encryption of data
Base station

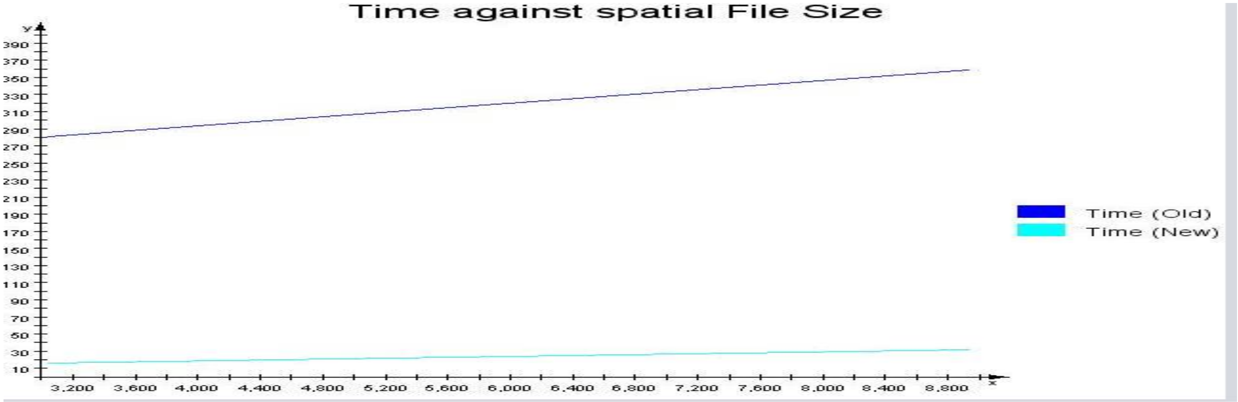
Demander request data from data providers. Provider collect data from multiple distributed base stations. Shadow matrix computation is performed for the efficient transmission and reception of data. To establish a connection with the provider, demander sent the private key to the key server for authentication. Key server send the public key of provider to demander. Demander access the corresponding provider using that key. Provider generate p1, p2, p3 for provider, Keyser and the demander. P1 and P2 for encryption at provider and key server side and p3 for decryption at demander side. During data transmission at provider side, data encrypted with p1 and send to key server. Key server again encrypt it with p2 and at last the demander decrypt it by using p3.

Perturbation is applied over data for better performance. Experimental results shows that the processing time is less compared to traditional approach and data retrieved in faster way

V. EXPERIMENTAL RESULTS



When data is mixed with noise, the length of data will increase. In Traditional approach, Noise mixing is limited to an extent because of the difficulty to recover the actual data. AES algorithm perform better encryption and decryption for efficient transmission and reception of data.



When the file size increase due to mixing of noise, the length of data increases but the time taken for transmission and reception is in a short interval of time compared to traditional approach. Experimental results shows that new approach is comparatively better than old approach.

VI. CONCLUSION AND FUTURE SCOPE

Data collection is a complex task in data mining field. In a distributed environment, complex algorithms are required for efficient transmission of data. Demander request data from providers and provider collect it from multiple distributed base stations. Identity of providers and base stations are hidden from demanders. Key server in the middle of demander and provider act as an authenticated server. Shadow coding applied to data for efficient transmission of data over providers. AES half encryption is done on provider and key server side. AES decryption is done on demander side. Noise is added along with data to achieve better privacy and security over data. Experimental results shows that new approach takes less time to retrieve data from base station and have less chance of attack.

In future, more algorithms and methods can be used to solve asynchronous data sharing over distributed environment.

REFERENCES

1. Jun Gao Jeffrey Xu Yu Ruoming Jin Jiashuai Zhou Tengjiao Wang Dongqing Yang ”Outsourcing shortest distance with privacy protection ” The VLDB Journal (2013) 22:543559 DOI 10.1007/s00778 012-0304-8.
2. Leo j Guibas ”Analysis of double hashing” Journal of computer and system sciences.
3. Kun Liu, Hillol Kargupta, and Jessica Ryan ” Random Projection-based Multiplicative Data Perturbation for Privacy Preserving Distributed Data Mining ” Fraser University 8888 University Drive, Burnaby BC, Canada
4. wangk@cs.sfu.ca.
5. Wenliang Du, Mikhail J. Atallah ”PRIVACY-PRESERVING COOPER ATIVE SCIENTIFIC COMPUTATIONS” Center for Education and Research in Information Assurance and Security, Purdue University, WestLafayette, IN 47907
6. Anna C. Squicciarini Mohamed Shehab Joshua Wede ” Privacy policies for shared content in social network sites” The VLDB Journal DOI10.1007/s00778-010-0193-7
7. Dima Alhadidi, Noman Mohammed, Benjamin C.M. Fung, and Mourad Debbabi ”Secure Distributed Framework for Achieving ε-Dierential Privacy ” Concordia Institute for Information Systems Engineering, Concordia University, Montreal, Quebec, Canada
8. Rakesh Agrawal Ramakrishnan Srikant ” Privacy-Preserving Data Mining ”IBM Almaden Research Center 650 Harry Road, San Jose, CA 95120
9. Eliazer eskin, ” Anomalie detection over noisy data using learned probability distribution ”IBM Almaden Research Center 650 Harry Road, SanJose, CA 95120

BIOGRAPHY

Tinimol Andrews doing MTech CSE in mangalam college of engineering. Receives bachelor of technology degree from mangalam college of engineering in 2013 . Area of interest is Datamining security.

Nimmy manuel works as an assistant professor in mangalam college of engineering. Receives bachelor degree from Mahatma Gandhi University Kottayam in 2006 and master degree from M.S University Thirunelveli in 2012.She has 7 years teaching experience in Computer Science. Area of interest is datamining.