ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

A Survey on Privacy-Preserving of Big data Storage Using Hybrid Triple DES and MD5

Anagha V. Raut [1], Prof. Sonal Honale [2]

P.G. Student, Department of Computer Engineering, AGPCE Engineering College, Nagpur, Maharashtra, India[1]

Associate Professor, Department of Computer Engineering, AGPCE Engineering College, Nagpur, Maharashtra, India[2]

ABSTRACT: The aim of this proposed work is to provide security Privacy-Preserving Sharing of Sensitive Information. The purpose of this proposed work is to provide the approach functions as a privacy shield to protect parties from disclosing more than the required minimum of their respective sensitive information. PPSSI deployment prompts several challenges, which are addressed in this project. Extensive experimental results attest to the practicality of attained privacy features and show that our approach incurs quite low overhead.

KEYWORDS: Privacy-Preserving, PPSSI, Extensive, Privacy shield, overhead.

I. INTRODUCTION

This template, In today's, sensitive data clearly needs to be kept confidential, data owners are often motivated, or forced, to share sensitive information increasingly digital world, there is often a tension between safeguarding privacy and sharing information. Although, in general Privacy-Preserving Sharing of Sensitive Information (PPSSI), and proposes one efficient and secure instantiation that functions as a privacy shield to protect parties from disclosing more than the required minimum of sensitive information. We model PPSSI in the context of simple database-querying applications with two parties: a server that has a database, and a client, performing simple disjunctive equality queries. In terms of the airline safety example above, the airline (server) has a database with passenger information, while DHS (client) poses queries corresponding to its TWL Intended Contributions. We explore the notion of Privacy-Preserving Sharing of Sensitive Information (PPSSI).

Our main building blocks are Private Set Intersection (PSI) techniques. As part of PPSSI design, we address several challenges stemming from adapting PSI to realistic database settings. In particular, we propose a novel encryption method to handle "multi-sets" and "data pointers" challenges and propose a new architecture with an Isolated Box to deal with "bandwidth" and "liability" challenges. Our experimental evaluation demonstrates that our approach incurs very low overhead: about 10% slower than standard (not privacy-preserving). All source code is publicly available.

The notion of privacy is commonly described as the ability of an individual or a group to seclude information about themselves, and thereby reveal it selectively. In many nations, laws or constitutions protect privacy as a fundamental individual right. The availability of information about an individual may result in having power over that individual, hence, generating concerns on potential misuse by governments, corporations, or other individuals.

In recent years, advances in computer and communication technologies have significantly amplified privacy risks. Nowadays, data is routinely exchanged electronically and collected by third parties. Privacy concerns are no longer limited to the anonymity and untraceability of digital activities. The disclosure of private information yields an increasing number of legal, monetary, practical, or even emotional, privacy issues.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

II. LITERATURE SURVEY

Kaitai Liang, Willy Susilo, Senior Member, IEEE "Privacy-Preserving Ciphertext Multi-Sharing Control for Big Data Storage", IEEE Transaction on Information Forensis and Security, Vol.10. No.8, AUG 2015.

In this paper, for the first time propose a privacy-preserving ciphertext multi-sharing mechanism. It combines the merits of proxy re-encryption with anonymous technique in which a ciphertext can be securely and conditionally shared multiple time without leaking both the knowledge of underlying message and the identity information of ciphertext senders/ recipients.[1]

Sung-Hwan Ahn Nam-Uk Kim ,Tai-Myoung Chung, "Big data analysis system concept for detecting unknown attacks",16-19 Feb. 2014(IEEE).

In this paper, Unknown cyber-attacks are increasing because existing security systems are not able to detect them, big data analysis techniques that can extract information from a variety of sources to detect future attacks. The event of new and previously unknown attacks, detection rate becomes very low and false negative increases to defend against these unknown attacks. Does not detect future Advanced Persistent Threat (APT) detection.[2]

Bhawna Gupta, Dr. Kiran Joyti, "Big Data Analytics with Hadoop to analyze Targeted Attacks on Enterprise Data", Journal of Computer Science and Information Technologies, Vol.5, 2014,(IEEE).

Big data security analytics is used for the growing practice of organization to gather and analyze security data to detect vulnerabilities and intrusions. Security and Information Event Monitoring (SIEM) system. The malicious and targeted attacks have become main subject for government, organization or indust. Big data analytics is the process of analyzing big data to find hidden patterns, unknown correlations and other useful information that can be extracted to make better decisions. It is used effectively and at the same time, hackers can leave their targets forever.[3]

Musca, Mirica, E.; Deaconescu, R.," Zero Day Attack Signatures Detection Using Honeypot", IEEE 29-31 May 2013. Unexpected behavior.

Fault distribution studies show that there is a correlation between the number of lines of code and the number of faults. LCS algorithm on the packet content of a number of connections going to the same services. Zero-day attack or threat is a computer threat that tries to exploit computer application vulnerabilities that are unknown to others or undisclosed to the software developer. Vulnerability window which is the time between the first exploitation of vulnerability and when software developers start to develop a countermeasure to that threat.[4]

Dajiang Lei Liping Zhang and Lisheng Zhang, "Cloud Model based Outlier Detection Algorithm for Categorical Data", Vol. 6, No. 4, August, 2013.

Numerical data but there will be a large number of categorical data in real life. Some outlier detection algorithm shave been designed for categorical data. There are two main problems of outlier detection for categorical data, which are the similarity measure between categorical data objects and the detection efficiency. Outlier detection algorithm for categorical data. Efficient outlier detection can help us make good decisions on erroneous data or prevent the negative influence of malicious and faulty behaviour. Many data mining techniques try to reduce the influence of outliers or eliminate them entirely. The information manner may result in the loss of important hidden information.[5][6].

Zhen Chen, Fuye Han, Junwei Cao, Xin Jiang, and Shuo Chen.1," Cloud Computing-Based Forensic Analysis for Collaborative Network Security Management System", February 2013.

Internet security problems remain a major challenge with many security concerns such as Internet worms, spam, and phishing attacks. Botnets, well-organized distributed network attacks, consist of a large number of bots that generate huge volumes of spam or launch Distributed Denial of Service (DDoS) attacks on victim hosts. A distributed security overlay network with a centralized security center leverages a peer-to-peer communication protocol used in the

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

UTMs collaborative module. These new security rules are enforced by collaborative UTM and the feedback events of such rules are returned to the security center. Collaborative network security management system cannot identify the intrusion.[7]

III .RESEARCH METHODOLOGY

The need for privacy-preserving sharing of sensitive information occurs in many different and realistic everyday scenarios, ranging from national security to social networking. A typical setting involves two parties: one seeks information from the other without revealing the interest while the latter is either willing or compelled, to share only the requested information. This poses the challenges. How to enable sharing such that parties learn no information beyond, what they are entitled, How to do so efficiently, in real-world practical terms. This explores the notion of Privacy-Preserving Sharing of Sensitive Information (PPSSI), and provides a concrete and efficient instantiation, modeled in the context of simple database querying. Proposed approach functions as a privacy shield to protect parties from disclosing more than the required minimum of their respective sensitive information. In this proposed system aims to provide information sharing while maintaining privacy of the information being shared in the big data storage. The system uses heterogeneous data storage. The hybrid encryption algorithm is proposed for data privacy. The data transmission is secured by proposing Triple DES and MD5. The proposed system is proposed to be implemented in java development kit (jdk), Net beans.

IV. PROPOSED APPROACH

The purpose of this proposed work is to provide the approach functions as a privacy shield to protect parties from disclosing more than the required minimum of their respective sensitive information. Multi-Sharing Control for Privacy-Preserving deployment prompts several challenges, which are addressed in this project. Extensive experimental results attest to the practicality of attained privacy features and show that our approach incurs quite low overhead.
For better attack detection, big data incorporates attack graph analytical procedures into the intrusion detection processes. The proposed method has several advantages:

- To avoid the attacker.
- Secrecy of the data should be maintained.
- The scheme is robust to withstand brute force attacks.

One of the sources of privacy violation is called data magnets (Rezgui et al., 2003). Data magnets are techniques and tools used to collect personal data. Examples of data magnets include explicitly collecting information through on-line registration, identifying users through IP addresses, software downloads that require registration, and indirectly collecting information for secondary usage. In many cases, users may or may not be aware that information is being collected or do not know how that information is collected. In particular, collected personal data can be used for secondary usage largely beyond the users' control and privacy laws. This scenario has led to an uncontrollable privacy violation not because of data mining itself, but fundamentally because of the misuse of data.

- Individual privacy preservation: The primary goal of data privacy is the protection of personally identifiable information. In general, information is considered personally identifiable if it can be linked, directly or indirectly, to an individual person. Thus, when personal data are subjected to mining, the attribute values associated with individuals are private and must be protected from disclosure. Miners are then able to learn from global models rather than from the characteristics of a particular individual.
- Collective privacy preservation: Protecting personal data may not be enough. Sometimes, we may need to protect against learning sensitive knowledge representing the activities of a group. We refer to the protection of sensitive knowledge as collective privacy preservation. The goal here is quite similar to that one for

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

statistical databases, in which security control mechanisms provide aggregate information about groups (population) and, at the same time, prevent disclosure of confidential information about individuals.

- Understanding privacy in data mining requires understanding how privacy can be violated and the possible means for preventing privacy violation. In general, one major factor contributes to privacy violation in data mining: the misuse of data.
- Users' privacy can be violated in different ways and with different intentions. Although data mining can be extremely valuable in many applications (e.g., business, medical analysis, etc), it can also, in the absence of adequate safeguards, violate informational privacy. Privacy can be violated of personal data.

V. PROJECT STRUCTURE

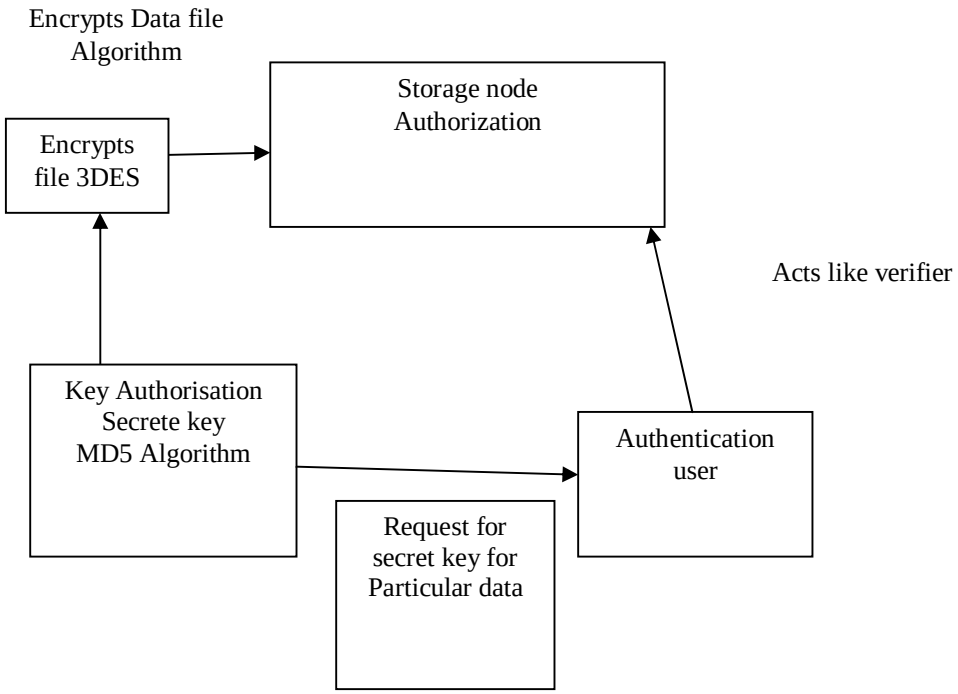

Fig. Basic block diagram

To increase the security level this proposed scheme overcomes the limitation of “Hybrid encryption algorithm proposed. The proposed enhanced scheme includes Triple DES, RSA and MD5.Triple DES (Variant of DES) strengthens the security of Data transmission. Reason behind for selecting triple DES rather than Double DES is that in double DES algorithm the key used for encryption and decryption is suspected to meet-in-middle attack. key distribution problem and in addition to this, MD5 to verify the integrity of the message. Use of message digest algorithm in combination of cryptographic algorithm.

MD5Crypto Service Provider class to generate the Hash string. I encapsulated the use of the method in the Compute Hash(byte[]objectAsBytes) method.MD5CryptoServiceProvider class wants a byte array as input. It does not accept an object directly. What you get out of it is not a string as we would like to have, but a byte array. Therefore I added the conversion from byte array to Hex. The conversion is done by using the ByteTo String() method. The method accepts a format string as input. And "X2" here means that each byte is converted into a two-char-string-sequence (e.g. 01011100

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

=> 5C or 00000111 => 07). Now there is still the question as to how to convert an object into a byte array. We know that our object is serializable. So we can serialize it into the memory (using a Memory Stream and a Binary Formatter) and getting out of the memory the needed byte array. Because the whole thing should be thread-safe, we lock the Serialization of the object.

VI. CONCLUSION

To support efficient handling of multiple auditing tasks, we further explore the technique of bilinear aggregate signature to extend our main result into a multi-user setting, where we can perform multiple auditing tasks simultaneously. Extensive security and performance analysis shows the proposed schemes are provably secure and highly efficient. We also show how to extent our main scheme to support batch auditing for delegation from multi-users.

REFERENCES

[1] Kaitai Liang, Willy Susilo, Senior Member, IEEE "Privacy-Preserving Ciphertext Multi-Sharing Control for Big Data Storage", IEEE Transaction on Information Forensis and Security, Vol.10. No.8, Aug .2015.
[2] Sung-Hwan AhnNam-Uk Kim, Tai-Myoung Chung, "Big data analysis system concept for detecting unknown attacks", 16th International Conference on Advance Communication Technology, 14197809, 16-19 Feb.2014.
[3] Bhawna Gupta, Dr. Kiran Joyti, "Big Data Analytics with Hadoop to analyze Targeted Attacks on Enterprise Data", Journal of Computer Science and Information Technologies, Vol.5, Issue 3, ISSN .0975-9646, 2014.
[4] Musca, Mirica, E. ; Deaconescu, R., " Zero Day Attack Signatures Detection Using Honeypot", IEEE Conference on Control System and Computer Science, 13683312, 29-31 May 2013.
[5] Dajiang Lei, Liping Zhang and Lisheng Zhang, "Cloud Model based Outlier Detection Algorithm for Categorical Data", International Journal of Database Theory and Application, Vol. 6, No. 4, August, 2013.
[6] Zhen Chen, Fuye Han, Junwei Cao, Xin Jiang, and Shuo Chen.1, " Cloud Computing-Based Forensic Analysis for Collaborative Network Security Management System", TSINGHUA Science AND Technology,Vol.8.No.1,ISSN 1007-0214 pp.40-50, February 2013.
[7] Dmitri Asonov, Johann-Christoph Freytag, "Almost optimal private information retrieval", International Workshop on PETS, Vol.2482, pp.209-223, 2002.
[8] D.Boneh and X.Boyen, "Introduction". "ID secures identity-based encryption", Berlin, Germany: Springer-Verlag, vol.3027, pp. 223–238, 2007.