Multi-Sharing Control for Privacy-Preserving for Big Data Storage

Mr. Nikhil M. Raghatate
Computer Science And Engineering
Ballarpur Institute of Technology
Chandrapur, India

Prof. Mrunali Vaidya
Computer Science And Engineering
Ballarpur Institute of Technology
Chandrapur, India

***Abstract*— 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. For better attack detection, big data incorporates attack graph analytical procedures into the intrusion detection processes. We must note that the design of does not intend to improve any of the existing intrusion detection algorithms; indeed, employs a reconfigurable virtual networking approach to detect and counter the attempts to compromise VMs, thus preventing zombie VMs.**

***Index Term*— PPSSI, attack graph, intrusion detection, zombie VMs.**

I INTRODUCTION.

In today’s increasingly digital world, there is often a tension between safeguarding privacy and sharing information. Although, in general, sensitive data clearly needs to be kept confidential, data owners are often motivated, or forced, to share sensitive information 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. 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 untraced ability of digital activities. The disclosure of private information yields an increasing number of legal, monetary, practical, or even emotional, privacy issues. However, the need for controlled (privacy-preserving) sharing of sensitive information occurs in many realistic scenarios, ranging from national security to individual privacy protection.

II. LITERATURE SURVEY

Big data analysis system concept for detecting unknown attacks:Sung-Hwan Ahn Nam-Uk Kim ,Tai-Myoung Chung. 16-19 Feb. 2014(IEEE). 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.[1]

Big Data Analytics with Hadoop to analyze Targeted Attacks on Enterprise Data: Bhawna Gupta, Dr.Kiran Joyti in 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.[2]

Zero Day Attack Signatures Detection Using Honeypot: Musca, Mirica, E. ; Deaconescu, R. 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. [3]
Cloud Model based Outlier Detection Algorithm for Categorical Data: Dajiang Lei Liping Zhang And Lisheng Zhang, 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 detectioncan help us make good decisions on erroneous data or prevent the negative influence of malicious and faulty behavior. Many data mining techniques try to reduce the influence of outliers or eliminate them entirely. The in forementionmanner may result in the loss of important hidden information.[4][5]
Cloud Computing-Based Forensic Analysis for Collaborative Network Security Management System: Zhen Chen*, Fuye Han, Junwei Cao, Xin Jiang, and Shuo Chen.1, 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 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 can not identify the intrustion.[6]

III. PROPOSED SYSTEM

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.
For better attack detection, big data incorporates attack graph analytical procedures into the intrusion detection processes. We must note that the design of does not intend to improve any of the existing intrusion detection algorithms; indeed, employs a reconfigurable virtual networking approach to detect and counter the attempts to compromise VMs, thus preventing zombie VMs.

The proposed method has several advantages.
1. To avoid the attacker.
2. Secrecy of the data should be maintained.
3. The scheme is robust to withstand brute force attacks.

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 if personal data are used for other purposes subsequent to the original transaction between an individual and an organization when the information was collected (Culnan, 1993). 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 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. However, unlike as is the case for statistical databases, another objective of collective privacy preservation is to protect sensitive knowledge that can provide competitive advantage in the business world.

IV. SYSTEM DESIGN

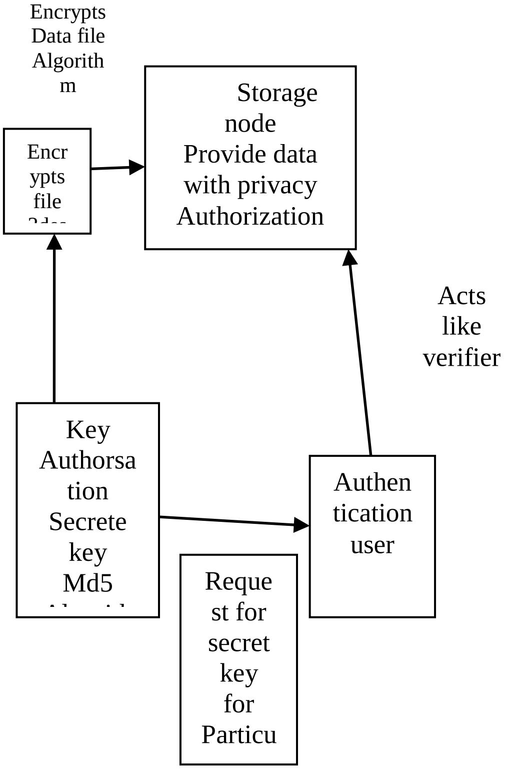

Figure1. Basic Block Diagram

The A typical setting involves two parties: one that seeks information from the other that is either motivated, or compelled, to share (only) the requested information. Consequently, in numerous occasions, there is a tension between information sharing and privacy. On the one hand, sensitive data needs to be kept confidential; on the other hand, data owners may be willing, or forced, to share information.

A social network user (Alice) wants to find out whether there are any other users nearby with whom she shares friends or group memberships, without relying on a third-party. Some of this information might be very sensitive, e.g., it might reveal Alice's medical issues or sexual orientation. Today, Alice would have to broadcast her information in order to discover a nearby "match", thus compromising her privacy. Whereas, Alice might be willing to disclose sensitive information only to users with a matching profile.

Interest Sharing:
Two or more users would like to share their common interests and activities, e.g., to discover matching locations, routes, preferences, or availabilities, without exposing any other information beyond the matching interests.

These examples motivate the need for privacy-preserving sharing of sensitive information and pose two main technical challenges: (1) how to enable this type of sharing such that parties learn no information beyond what they are entitled to, and (2) how to do so efficiently, in real-world practical terms.
Cryptographic Protocols and Open Problems

Technology advances have radically influenced our modes of communication and haveequally prompted a number of privacy challenges. As a result, there has recently been a lot of research activities in the context of Privacy-Enhancing Technologies (PETs).

Modern Cryptography
has played a key role within PETs, producing a number of effective cryptographic protocols for privacy protection

V. CONCLUSIONS

The developed system is powerful enough we proposed a novel architecture for Privacy-Preserving Sharing of Sensitive Information (PPSSI), based on efficient PSI techniques. It enables a client and a server to exchange information without leaking more than the required minimum. Privacy guarantees are formally defined and achieved with provable security. Experimental results show that our approach is sufficiently efficient for real-world applications. 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 TPA 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 TPA upon delegations from multi-users.

REFERENCES

[1] Asonov, D., Freytag, J.-C.: Almost optimal private information retrieval. In: PETS

[2] Boneh, D., Di Crescenzo, G., Ostrovsky, R., Persiano, G.: Public key encryption with keyword search. In: Cachin, C., Camenisch, J.L. (eds.) EUROCRYPT 2004. LNCS, vol. 3027, pp. 506–522. Springer, Heidelberg (2004)

[3] zaslon Analytics. Consumer Data Losses, http://www.caslon.com.au/ datalossnote.htm

[4] Chor, B., Gilboa, N., Naor, M.: Private information retrieval by keywords. Manuscript (1998)

[5] Chor, B., Kushilevitz, E., Goldreich, O., Sudan, M.: Private information retrieval. Journal of the ACM 45(6), 965–981 (1998)

[6] Davidoff, S.: What Does DHS Know About You?, http://tinyurl.com/what-dhs-knows

[7] De Cristofaro, E., Jarecki, S., Kim, J., Tsudik, G.: Privacy-preserving policy-based information transfer. In: Goldberg, I., Atallah, M.J. (eds.) PETS 2009. LNCS, vol. 5672, pp. 164–184. Springer, Heidelberg (2009)

[8] De Cristofaro, E., Lu, Y., Tsudik, G.: Efficient techniques for privacy-preserving sharing of sensitive information. Cryptology ePrint Archive, http://eprint.iacr.org/2011/113

[9] De Cristofaro, E., Tsudik, G.: Practical private set intersection protocols with linear complexity. In: Sion, R. (ed.) FC 2010. LNCS, vol. 6052, pp. 143–159. Springer, Heidelberg (2010)

[10] Feige, U., Killian, J., Naor, M.: A minimal model for secure computation (extended abstract). In: STOC (1994)