Enhancing Information Leakage in Multi-Cloud Storage Services

Sanket Aher[1], Rohit Raut[2], Akash Tidke[3], Shubham Vanarase[4], Prof. H. R. Agashe[5]

Students, Department of Information Technology[1,2,3,4]

Professor, Department of Information Technology[5]

Matoshri College of Engineering and Research Center, Nashik, Maharashtra, India

Abstract*:* *Data leakage is a growing insider threat in information security among organizations and individuals. A series of methods have been developed to address the problem of data leakage prevention (DLP). However, large amounts of unstructured data need to be tested in the Big Data era. As the volume of data grows dramatically and the forms of data become much complicated, it is a new challenge for DLP to deal with large amounts of transformed data. We propose an Adaptive Weighted Graph Walk model (AGW) to solve this problem by mapping it to the dimension of weighted graphs. Our approach solves this problem in three steps. First, the adaptive weighted graphs are built to quantify the sensitivity of tested data based on its context. Then, the improved label propagation is used to enhance the scalability for fresh data. Finally, a low-complexity score walk algorithm is proposed to determine the ultimate sensitivity. Experimental results show that the proposed method can detect leaks of transformed or fresh data fast and efficiently.*

Keywords: Data Leakage, Big data, Adaptive Weighed Graph, Low-complexity, etc..

I. INTRODUCTION

Data leakage is defined as the accidental or unintentional distribution of private or sensitive data to unauthorized entity. Sensitive data of companies and organizations includes intellectual property (IP), financial information, patient information, personal credit-card data, and other information depending on the business and the industry. Furthermore, in many cases, sensitive data is shared among various stakeholders such as employees working from outside the organizational premises (e.g., on laptops), business partners and customers. This increases the risk of confidential information falling into unauthorized hands. Whether caused by malicious intent, or an inadvertent mistake, by an insider or outsider, exposed sensitive information can seriously hurt an organization.

In our system, we have assumed that distributor's sensitive data is in the form of "Relational Database" and agent (trusted party) is going to request for a part of original database say data object. Before requesting required data object, agent needs to Sign Up by filling registration form. While filling this registration form, unique ID is assigned to each registering agent by system itself. After Sign UP, agent can Sign In to send request for data object.

Unique ID assigned to an agent is used to create fake object (watermark)which is going to be embedded in requested data object and this modified Data Object+ Fake Object is given to an agent, so that even if this modified data is leaked by agent and similar data is found by distributor somewhere, then the distributor must assess the likelihood that the leaked data came from one or more agents.

II. LITERATURE SURVEY

An enterprise data leak is a scary proposition. Security practitioners have always had to deal with data leakage issues that arise from email and other Internet channels. But now with the use of mobile technology, it's easier for data loss to occur, whether accidentally or maliciously. The guilty detection approach we present is related to the data provenance problem tracing the lineage of S objects implies essentially the detection of the guilty agents [2]. And assume some prior knowledge on the way a data view is created out of data sources. Our problem formulation with objects and sets is more general as far as the data allocation.

DOI: 10.48175/568

Strategies are concerned; our work is mostly relevant to watermarking (Stenography) [1] that is used as a means of establishing original ownership of distributed objects. Watermarking is a unique code is embedded in distribute copy [5]. Data leakages can be identified using these original data [5]. Thus watermarking is a useful methodology. But sometimes the watermarks can be destroyed if the data recipient is malicious [5]. Hence this technique proves to be inefficient. Finally, there are also lots of other works on mechanisms that allow only authorized users to access sensitive data through access control policies [2]. Such approaches prevent in some sense data leakage by sharing information only with trusted parties. However, these policies are restrictive and may make it impossible to satisfy agent's requests.

III. PROBLEM STATEMENT

The leak of sensitive data on computer systems poses a serious threat to organizational security. Statistics show that the lack of proper encryption on files and communications due to human errors is one of the leading causes of data loss. But mostof them worried about security; frequently they used to keep the data in single data server or data chunk. In this case if the data is lost or hacked in a sense entire data will be loose. To circumvent these kinds of vulnerabilities and to achieve better security we are offering of multi-instances data storage technique where the data will be stored in different databases or data chunks means instances.

IV. GOALS AND OBJECTIVES

The objectives of the "Data Leakage Detection" are as follows:

- To detect the agent who leaked the confidential data and send alert message to the distributor.
- Detection of guilty agent.
- Send message or email to the distributor with identification of guilty agent.
- Send alert message to the guilty agent.
- Take legal action on agent when he/she break rule after the alert message.

V. SYSTEM DESIGN

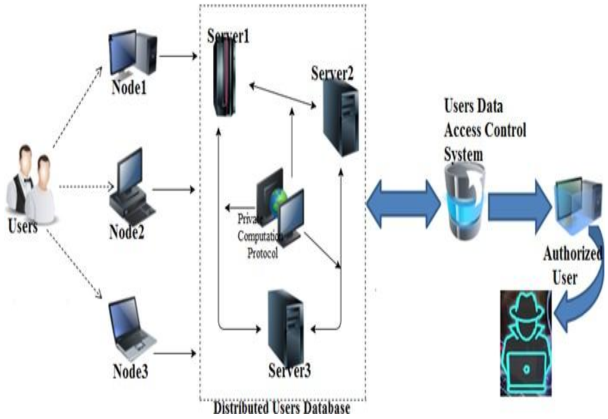

Fig. 1.System Architecture

B. Proposed System

In recent years, wireless network networks have been widely used in healthcare, banking, educational, and industrial applications, such as hospital and home patient monitoring, and also sensitive user information. Wireless networks are more vulnerable to eavesdropping, modification, impersonation and replaying attacks than the wired networks. A lot of work has been done to secure wireless networks. The existing solutions can protect the user data during transmission, but cannot stop the inside attack where the administrator of the user database reveals the sensitive user data. In this proposed system, we propose a practical approach to prevent the inside attack by using multiple data servers to store user data. The main contribution of this system is securely distributing the user's or organization-related data in multiple data servers and employing the SHA crypto-systems to perform statistical analysis on the user/organizational data without compromising its privacy

VI. ALGORITHMS USED

In this proposed system we implement any one of the following algorithm as per dataset.

A. AES Algorithm:

The encryption process uses a set of specially derived keys called round keys. These are applied, along with other operations, on an array of data that holds exactly one block of data to be encrypted. This array we call the statearray. You take the following AES steps of encryption for a 128-bit block:

- Derive the set of round keys from the cipher key.
- Initialize the state array with the block data (plaintext).
- Add the initial round key to the starting state array.
- Perform nine rounds of state manipulation.
- Perform the tenth and final round of state manipulation.
- Copy the final state array out as the encrypted data (ciphertext).

The reason that the rounds have been listed as “nine followed by a final tenthround” is that the tenth round involves a slightly different manipulation from the others.

B. Shamir Secret Sharing:

In cryptography, secret-sharing schemes are schemes that split shares of a given secret among a set of trusted participants. This secret could be a very important piece of information that may be needed in the future but needs to be kept private and secure. These shares are completely useless on their own,however, when put together reconstruct the secret and make it apparent. Asa thought experiment; think of secret sharing schemes as a puzzle where the puzzle pieces are split among ten players but are completely blank. The imagethat the puzzle creates is only visible once all the pieces are put together.

The kind of data that is best suited for a secret-sharing algorithm is information that must be kept absolutely private, but must also be stored securely andnever lost. Typically, you use secret sharing for access keys to accounts withhighly sensitive information in them. The goal is to spread the key out from one geographic location into multiple, so in order to compromise a system, you need to compromise devices in distinct locations first. There are multiple kinds of secret sharing algorithms.

VII. CONCLUSION

We have investigated the security and privacy as well as data leakage issues in the wireless network data collection storage and queries and presented a complete solution for the privacy-preserving wireless network. To secure the communication between user and data servers.

To keep the privacy of the user’s data, we proposed a new data collection protocol that splits the user’s data into three numbers and stores them in three data servers, respectively. As long as one data server is not compromised, the privacy of the user’sdata can be preserved. For the legitimate user e.g. receiver to access the user’s data,we proposed an access control protocol, where three data servers cooperate to pro- vide the user with the user’s data, but do not know what it is.

VIII. ACKNOWLEDGMENT

We would also like to show our gratitude to, *Prof. H. R. Agashe (professor, department of Information Technology, Matoshri College of Engineering and Research Center, Nashik, Maharashtra, India.)* for sharing their pearls of wisdom with us during the course of this research. We are also immensely grateful to him for his comments on an earlier version of the manuscript, although any errors are our own and should not tarnish the reputations of these esteemed persons.

DOI: 10.48175/568

REFERENCES

[1]. R. Agrawal and J. Kiernan, “Watermarking Relational Databases,”Proc. 28th Int’l Conf. Very Large Data Bases (VLDB ’02), VLDBEndowment, pp. 155-166, 2002. IEEE Transaction and knowledge and data engineering, Vol.23, No.1, January 2011.

[2]. P. Bonatti, S.D.C. di Vimercati, and P. Samarati, “An Algebra for Composing Access Control Policies,” ACM Trans. Information and System Security, vol. 5, no. 1, pp. 1-35, 2002.

[3]. P. Buneman, S. Khanna, and W.C. Tan, “Why and Where: Characterization of Data Provenance,” Proc. Eighth Int’l Conf. Database Theory (ICDT ’01), J.V. den Bussche and V. Vianu, eds., pp. 316-330, Jan. 2001.

[4]. P. Buneman and W.-C.Tan, “Provenance in Databases,” Proc.ACM SIGMOD, pp. 1171-1173, 2007.

[5]. Ms. Aishwarya Potdar1, Ms. Rutuja Phalke2, Ms. Monica Adsul3, Ms.Prachi Gholap4 B.E, Department of Computer Engineering, KJCOEMR, Pune University, Pune, India, International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 4, April 2013.

[6]. R. Agrawal and J. Kiernan. Watermarking relational databases. In VLDB ’02: Proceedings of the 28th international conference on Very Large Data Bases, pages 155–166. VLDB Endowment, 2002.

[7]. P. Bonatti, S. D. C. di Vimercati, and P. Samarati. An algebra for composing access control policies. ACM Trans. Inf. Syst. Secur., 5(1):1–35, 2002.

[8]. P. Buneman, S. Khanna, and W. C. Tan. Why and where: A characterization of data provenance. In J. V. den Bussche and V. Vianu, editors, Database Theory - ICDT 2001, 8th International Conference, London, UK, January 4-6, 2001, Proceedings, volume 1973 of Lecture Notes in Computer Science, pages 316–330. Springer, 2001.

[9]. P. Buneman and W.-C. Tan. Provenance in databases. In SIGMOD ’07: Proceedings of the 2007 ACM SIGMOD international conference on Management of data, pages 1171–1173, New York, NY, USA, 2007. ACM.

[10]. Y. Cui and J. Widom. Lineage tracing for general data warehouse transformations. In The VLDB Journal, pages 471–480, 2001.

DOI: 10.48175/568