International Journal on Applications in Engineering and Technology
Volume 2: Issue 2: February 2016, pp 10-13 www.aetsjournal.com ISSN (Online) : 2455-0523

Securing Data using Paillier Encryption Algorithm Based on Data Leak Detection (DLD) Technique

Athulya Subhash, Konjengbam Sarjita Devi, SaravanaPriya. D, Subbulakshmi.R, T. Palani Raja

Abstract— Data leakage is the common problem in the government or any other organizations and it has been growing rapidly, it is caused by human errors. The detection operation is secrecy in an existing system but it is difficult to satisfy in real time. During the detection process the detection server may be compromised. In proposed system to present a fuzzy fingerprint technique to solve this kinds of issues. The advantage of this technique is to improve the data protection against the unauthorized data transmission, to provide network security to the sensitive data also identifying the guilt agents. The evaluation results indicate high accuracy, accurate detection with very low false alarms.

Keywords— network security, sensitive data, data leak, privacy.

I. INTRODUCTION

A network is a telecommunication network, it gives permission to the computer for data transmission. Data leakage is big challenge in an organization, there are several algorithms are design for data security. Network data-leak-detection is a method, it performs Deep Packet Inspection (DPI) over a network channel. DPI is used to analyze the TCP/IP packet for inspecting the data, when the data found in network traffic then gives alerts to the organization. If the detection system is outsourced then it may expose the sensitive data to the unauthorized user. To proposed the Fuzzy fingerprint algorithm to solve this problem that enhances data privacy during the process. This approach is based on the one-way computation. It can support the data owner to safely delegate the detection operation without exposing the sensitive data.

In this detection operation the data owner to prepare the fingerprint and then release the fingerprint, small amount of data to the DLD provider. Data owner does not want to directly expose the sensitive data to the provider. The DLD provider continuously monitor the network channel and check any data leaks are found over a channel. If any leaks are found immediately send all data leak reports to the data owner. Now the data owner can decide whether or not it is a data leak also identifying the guilt agents. During the monitoring process the DLD provider gain exact knowledge about the sensitive data. The security goal of this method is to detect the inadvertent data leak cause by human mistakes. The privacy goal of the Fuzzy fingerprint mechanism to prevent the DLD provider from gaining the exact value about the data during the operation. It means that the DLD provider given digests of the sensitive data to the owner then the content of the network to be examined. The DLD provider should not learn the exact value of the sensitive data. The main goal of the Fuzzy fingerprint technique is to hide the sensitive data during the detection operation.

Athulya Subhash, Konjengbam Sarjita Devi, SaravanaPriya. D, Subbulakshmi. R U.G student, Department of Computer Science And Engineering, Nehru Institute of Technology, Coimbatore, India(athulyasubhash211@gmail.com, sarjitakonjengbam20@gmail.com, saravanapriyamd@gmail.com, subbulakshmir75@gmail.com)
T. Palani Raja , Assistant Professor, Department of Computer Science And Engineering, Nehru Institute of Technology, Coimbatore, India(palanirajaT@gmail.com)

II. RELATED WORKS

There are several advances in security applications, gives more security to the sensitive data. The fuzzy fingerprint mechanism to identify the outsourced DLD server and provide a systematic solution to this problem. There existing system, shingle and Rabin fingerprint technique was used for identifying the data leak and a collaborative setting. To propose the fuzzy fingerprint algorithm gives the privacy preserving data leak detection solution with convincing results .Most data leak detection products do not have the privacy preserving feature and this products are offered by the industries. The proposed system approach is different from the other approach and it can provides the data leak detection service. Using this method the data owner does not need to fully reveal the sensitive data to the DLD provider. Bloom filter is used in the network security layers from network security to application security, it is a space-saving data structure for se membership test. The fuzzy Bloom filter invented to constructs a special Bloom filter, it sets the corresponding filter bits to1's. This method is a potential privacy preserving technique. The fuzzification process is used in fuzzy fingerprint technique, it is separated from the membership test, and it is flexible to test whether the fingerprint is sensitive with or without

fuzzification. Privacy preserving keyword search or fuzzy keyword search provide string matching in semi honest environments. Anomaly detection can be used to detect data leaks in network traffic. It detects the new information in traffic, entropy analysis is used in this detection process. To present a signature based model to monitor the design can be outsourced also detect the data leaks. Both the anomaly detection and signature based detection approaches are different.

Tracing and enforcing are another approaches for data leak detection. It contains data flow and file-descriptor sharing enforcement. This approaches do not provide a remote service so this approaches are different from ours. The fuzzy fingerprint approach some other privacy preserving methods are invented for specific process, e.g., secure multi-party computation. SMC is a cryptographic mechanism it supports the string matching also complex functions. The advantage of the proposed system is its concision and efficiency.

III. EXISTING SYSTEM

The literatures on privacy preserving data mining can divide into two categories. In the first category, methods modify the data mining algorithms so that without knowing the exact values of data, they allow data mining operations on distributed dataset. In the second category, methods are modifying the values of the datasets to protect privacy of data values. In this category there are several research has been done in data distortion or data perturbation are as follow: In the year 1982, T. Dalenius they firstly proposed the idea of Data Swapping. In this technique the database is to transform by switching a subset of attributes between selected pair of records. Therefore the lower order frequency counts are preserved in such a way that the data confidentiality is not compromised. The data swapping is a great data perturbation technique for privacy protection of data values. In the year 1985, Liew proposed data distortion method based on probability distribution.

IV. PROPOSED SYSTEM

FUZZY FINGERPRINT METHOD AND PROTOCOL Shingles and Fingerprints: The DLD provider obtains digests of sensitive data from the data owner. The data owner uses a sliding window and Rabin fingerprint algorithm to generate short and hard to-reverse (i.e., one-way) digests through the fast polynomial modulus operation. The sliding window generates small fragments of the processed data (sensitive data or network traffic), which preserves the local features of the data and provides the noise tolerance property. Rabin fingerprints are computed as polynomial modulus operations, and can be implemented with fast XOR, shift, and table look-up operations. The Rabin fingerprint algorithm has a unique min-wise independence property, which supports fast random fingerprints selection (in uniform distribution) for partial fingerprints disclosure. The shingle-and-fingerprint process is defined as follows. A sliding window is used to generate q-grams on an input binary string first. The fingerprints of q-grams are then computed. A shingle (q-gram) is a fixed-size sequence of contiguous bytes. For example, the 3-gram shingle set of string abcdefgh consists of six elements {abc, bcd, cde, def, efg, fgh}. Local feature preservation is accomplished through the use of shingles.

A. Design Considerations:

1. Generate fingerprint for each sensitive data.
2. Release the fingerprint and reveal the small amount of data to the provider.
3. DLD provider monitor the network traffic.
4. Detect the data leaks.
5. Report all data leak alerts to the data owner, it enables to identify the guilt agents.
6. Data owner decide whether or not it is a true leak.

B. Description of the Proposed Algorithm:

The main goal of the proposed algorithm is to discover the appearance of the sensitive data over a supervised network channel and prevents the DLD provider to learn the content of the data. The proposed algorithm contains three main steps.

Step 1: Shingles and Fingerprints: The DLD provider obtains digests from the data owner for each sensitive data. The data owner to generate the one-way computation digests using the shingle and Rabin fingerprint. A shingle is fixed-size sequence of the bytes. For example, 2-gram shingle set of string abcde consists of four elements (abb c cd de). The use of shingles stand alone does not satisfy the one-way computation requirements. After the shingling Rabin fingerprint is used to satisfy the one way function requirements.

Step 2: Selection Criteria:

The fuzzy fingerprint is matched in network traffic, the DLD provider to detect the data leak and alerts are triggered then Reports all data leaks to the data owner. The fuzzy fingerprint is not matched the DLD provider adversarial needs to reverse the Rabin fingerprinting computation for obtain the shingle. To quantify the alert rate in the network traffic for sensitive data. The expected alert rate(R) is presented in eq(1).Where this the total number of fuzzified sensitive fingerprints t expected traffic fingerprint $ps \in (0,1)$partial disclosure rate, and α the expected rate.

Step 3: Limitations:

There three limitations are used in this method. 1. Modified data leak, the shingle has the limited power to detect fully modified data leaks. The data is modified the data leak detection failure may occur. Advanced content comparison is needed to solve this issue. 2 Dynamic sensitive data, it is used to protect the dynamically changing data.

International Journal of Innovative Research in Computer and Communication Engineering (A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 4, Issue 1, January 2016

The digests continuously need to update. Raise question to the community for this problem. 3. Selective fragments leak, false negative may occur using the subset of the sensitive data scheme (partial disclosure).

V. OVERALL ARICHITECTURE

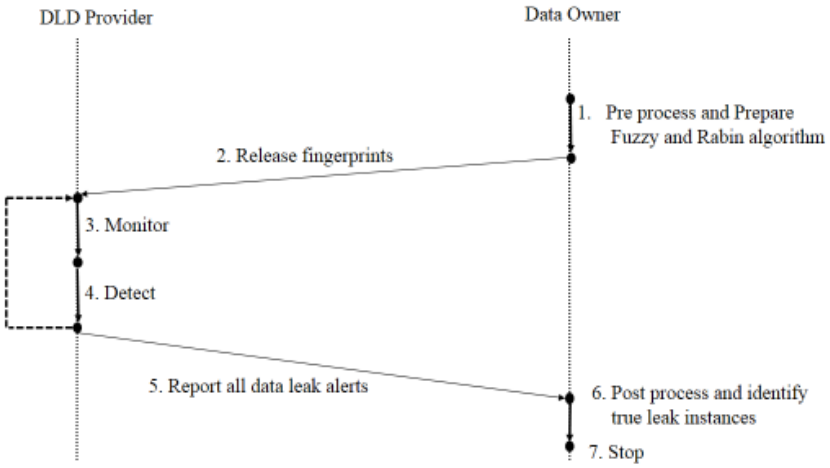


Figure 1

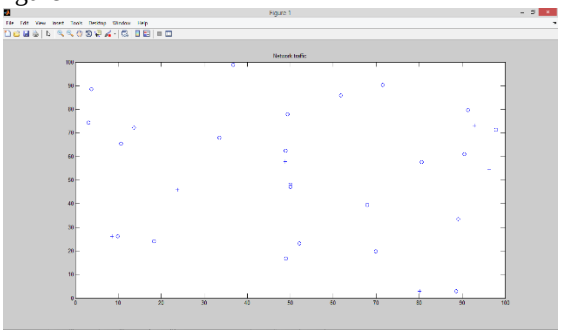


Figure 2

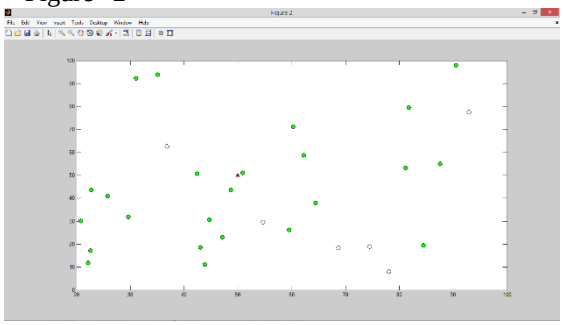
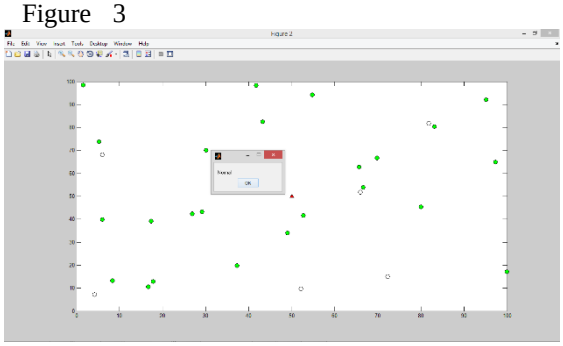


Figure 3



VI. CONCLUSION

We proposed fuzzy rules and Paillier Encryption algorithm, a privacy-preserving data-leak detection model and present its realization. Using special digests, the exposure of the sensitive data is kept to a minimum during the detection. We have conducted extensive experiments to validate the accuracy, privacy, and efficiency of our solutions. . For future work, we plan to focus on designing a host-assisted mechanism for the complete data-leak detection for large-scale organization.

References

[1]. X. Shu and D. Yao, "Data leak detection as a service," in Proc. 8th Int. Conf. Secur. Privacy Commun. Netw., 2012, pp. 222–240.

[2]. Risk Based Security. (Feb. 2014). Data Breach Quick- View: An Executive's Guide to 2013 Data Breach Trends .

[3]. Ponemon Institute. (May 2013). 2013 Cost of Data Breach Study: Global Analysis.

[4]. Identity Finder. Discover Sensitive Data Prevent Breaches DLP Data Loss Prevention.

[5]. K. Borders and A. Prakash, "Quantifying information leaks in outbound web traffic," in Proc. 30th IEEE Symp. Secur. Privacy, May 2009, pp. 129–140.

[6]. H. Yin, D. Song, M. Egele, C. Kruegel, and E. Kirda,"Panorama: Capturing system-wide information flow for malware detection and analysis," in Proc. 14th ACM Conf. Comput. Commun. Secur., 2007, pp. 116–127.

[7]. K. Borders, E. V. Weele, B. Lau, and A. Prakash, "Protecting confidential data on personal computers with storage capsules," in Proc. 18th USENIX Secur. Symp., 2009, pp. 367–382.

[8]. A. Nadkarni and W. Enck, "Preventing accidental data disclosure in modern operating systems," in Proc. 20th ACM Conf. Comput. Commun. Secur., 2013, pp. 1029–1042.

[9]. A. Kapravelos, Y. Shoshitaishvili, M. Cova, C. Kruegel, and G. Vigna, "Revolver: An automated approach to the detection of evasiveweb-based malware," n Proc. 22nd USENIX Secur. Symp., 2013, pp. 637–652.

[10]. X. Jiang, X. Wang, and D. Xu, "Stealthy malware detection and monitoring through VMM-based 'out-of-the-box' semantic view reconstruction," ACM Trans. Inf. Syst. Secur., vol. 13, no. 2, 2010, p. 12.

[11].G. Karjoth and M. Schunter, "A privacy policy model for enterprises," in Proc. 15th IEEE Comput. Secur. Found. Workshop, Jun. 2002, pp. 271–281.