Data Lineage in Malicious Environment (DLIME) for Text Data by using AES, SHA

Prof. Abhijit Pawar¹, Miss. Sonali Kadam², Miss. Pranali Jagtap³, Miss. Priyanka Bhosale⁴, Miss. Supriya Kadam⁵
¹,²,³,⁴,⁵Department of Information Technology, SVPM’s College of Engineering, Malegaon (Bk), Tal- Baramati, Dist-Pune, Savitribai Phule Pune University.

Abstract: Intentional or unintentional confidential data is leaked and it is undoubtedly one of the most severe security threats that organizations are facing in this digital era. The threats now extending to our personal lives and professional lives, information is available to social networks and smart phone providers is indirectly transferred(hacked or transferred) to untrustworthy third party and fourth party applications. Presenting a data lineage framework DLIME for identifying a guilt entity who leaked the data across public cloud which takes three characteristic, principal roles (i.e. owner, consumer and auditor). DLIME allows to identify guilty entity who leaked the data within a malicious environment by building upon fake record, digital signature and encrypting the modified record.

Keywords: Owner, Consumer, Auditor.

I. INTRODUCTION

Large amounts of digital data can be copied at almost no cost and can be spread through the internet in very short time and easily. Additionally, the risk of getting caught who leaked the data is very low, as there are currently almost no accountability for getting caught that thieves. For these reasons, the problem of data leakage has reached now a day. Not only companies are affected by data leakage, it is also a concern to individual’s personal lives. The rise of smart phones and social networks has made the situation worse. Through smart phones and social media does not have the full security, the providers directly or indirectly (hacked or transferred) to untrusted third party and forth party applications.

A generic data lineage framework LIME for data flow across multiple entities that take three characteristic, principal roles (i.e. Owner, Auditor and consumer). The system is define the exact security guarantees required by such a data leakage data transfer protocol between two entities i.e. owner and consumer within a malicious environment by building upon fake record, digital signature and encrypting the modified record.

A. Problem Statement

Intentional or unintentional confidential data is leaked and it is undoubtedly one of the most severe security threats that organizations are facing in this digital era.

B. Aim of Projects

The aim is to detect when the owners sensitive data have been leaked by agents and identify the agent that leaked the data on public server.

II. PROPOSED SYSTEM

In an organization there are three different roles that can be involved in this system: first is data owner, second is data consumer (agent) and third auditor. The data owner is responsible for the distributing data on private cloud by adding unique fake records, calculating digital hash by digital SHA algorithm, encrypting records by AES, and also sending the digital hash to auditor for auditing and then uploading it on private server and also send the key by email to agent who is performing a survey on that records and the consumer receives documents, decrypt it and can carry out some task using them. The auditor is not involved in the transfer of documents, he is only invoked when a leakage occurs and then performs all steps that are necessary to identify the leaker.
A. **Steps for Data Leakage Detection**

1) Owner will have a database of guaranteed security. This data will be given to agents for processing (analysis purpose) and the data will be in the form of a table.

2) Owner will generate fake records for each agent (unique). For example, f1, f2, f3 for agent 1 and f4, f5, f6 for agent 2. Save fake records separately in the database individually.

3) Calculate the digital signature of the table and send the digital signature to the auditor. (using SHA algorithm)

4) Now encrypt data using a key, send the encrypted data to the storage server and send the key through email to the agent.

5) The agent will download data from the server. Using the key, decrypt the data. An agent can perform analysis on the data.

6) Any agent can upload data to the public server.

7) Owner will start verification request to the auditor.

8) The auditor will check the public server for data. If data is found, compare hash with the hash database. If matched, report the owner with the hash of the matched record.

9) Finally, the owner will match the hash with the unique fake record in the database, and detect the agent who leaked the data.

### III. MODULES

**A. Owner**

Owner will have a database of a general survey and this data will be given to agents for processing and this data will be in the form of a table. Owner will generate fake records for each agent, calculate digital signatures of the table, and send it to the auditor. Encrypt the data using a key and send the encrypted data to the private server.

**B. Agent**

Agent will download data from the private server and decrypt the data using the key. Survey can be performed on the data and any malicious data will leak the data.

**C. Auditor**

Auditor will check the public server for data. If data is found, compare hash with the hash database. If matched, report the owner with the hash of the matched record.
matched record.

IV. ALGORITHMS

A. SHA (Secure Hash Algorithm)

Steps
1) Append Padding bits: Padding means addition of bits to the original message. To make length of original message to a value 64 bits less than multiple of 512. The message is padded to make the length of message 448 mod 512.
2) Append Length: A block 64 bit is appended to a message. 64 bits of original message is appended to the result is (original message + Padding).
3) Initialize MD5 Buffer: A 160-bit buffer is used to store the intermediate as well as final result.

B. AES (Advance Encryption Standard)

Steps
1) Sub Bytes: Sub Bytes() consists of replacement of each byte using a fixed S-box lookup table to achieve non-linearity into the 4*4 array(16 bytes).
2) Shift Rows: The o/p of the subbyte transformation is i/p to the shiftrows transformation which consists of rotation of each byte of the state array in the order of a row of data matrix.
3) Mix Column: Mix column performs operation on the state array obtained from shiftrows column-by-column is multiplied with row of a fixed matrix.
4) Add Round Key: The round key is added by combining each byte of the state array using bitwise XOR operations. The actual 'encryption' is performed in the AddRoundKey() function.

V. CONCLUSION

DLIME, Finding guilt entity by adding fake records and calculate digital signatures and encrypting records so that no other third party can have access to read it. By doing these it gets easy to identify the leaker when has leaked the data on public cloud by matching the digital hash on public server and hash from the auditors database which was given by owner after he added fake record and calculated hash. Although LIME does not actively prevent data leakage, but it allows us to find guilt entity who leaks the data on public cloud.

VI. ACKNOWLEDGEMENT

We take this opportunity to thank our project guide Prof. Pawar A.H. and other guidance teacher and Head of the Department Prof. Mali J.N and Principal Dr.Yadav D.M for their valuable guidance and for providing all the necessary facilities, which were indispensable in the completion of this project report. We are also thankful to all the staff members of the Department of Information Technology of SVPM's College of Engineering, Malegaon (Bk) for their valuable time, support, comments, suggestions and persuasion. We would also like to thank the institute for providing the required facilities, Internet access and important books.

REFERENCES