iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** III **Month of publication:** March 2017

DOI: http://doi.org/10.22214/ijraset.2017.3234

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

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

Prof. Abhijit Pawar[1], Miss. Sonali Kadam[2], Miss. Pranali Jagtap[3], Miss. Priyanka Bhosale[4], Miss. Supriya Kadam[5]
[1,2,3,4,5]*Department 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 consumer). 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.

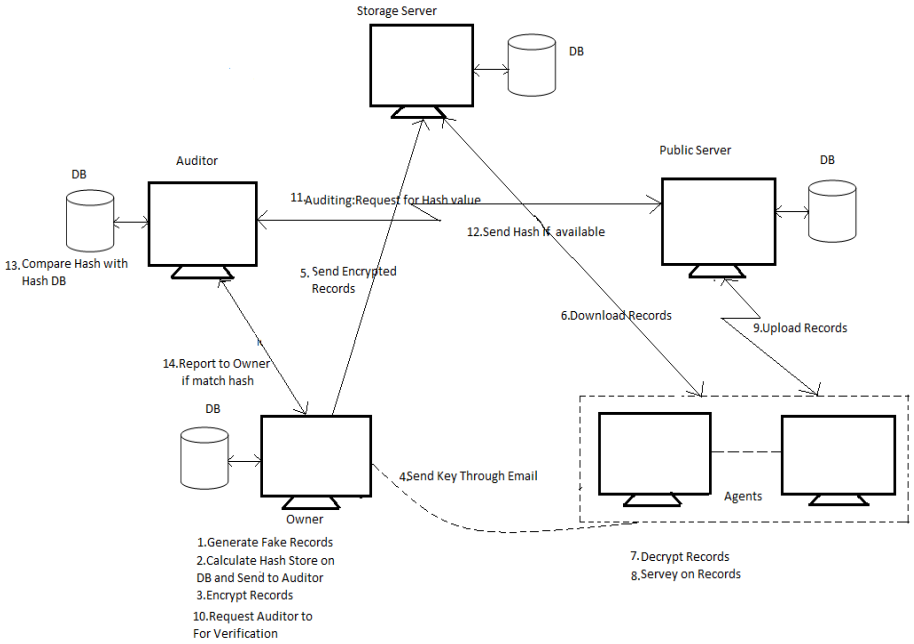
B. Aim of Projects

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.

System Architecture



A. Steps for Data Leakage Detection

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

2) Owner will generate fake record for each agent (unique). for eg. f1, f2,f3 for agent1 and f4,f5,f6 for agent2. Save fake record in the database individually.

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

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

5) Agent will download data from the server.Using key agent decrypt the data. An agent can make analysis on the data.

6) Any agent will upload data to public server.

7) Owner will start verification request to auditor.

8) Auditor will check public server for data. If data found- compare hash with the hash db. If matched owner with hash of matched record.

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

III. MODULES

A. Owner

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

B. Agent

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

C. Auditor

Auditor will check public server for data if data found compare hash with the hash DB. If matched report the owner with hash of

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.
4) Process Message in 512 bit It consist of four rounds of 20-step each.Each round takes i/p 512-bit block processed it and produced 160 bit o/p.

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

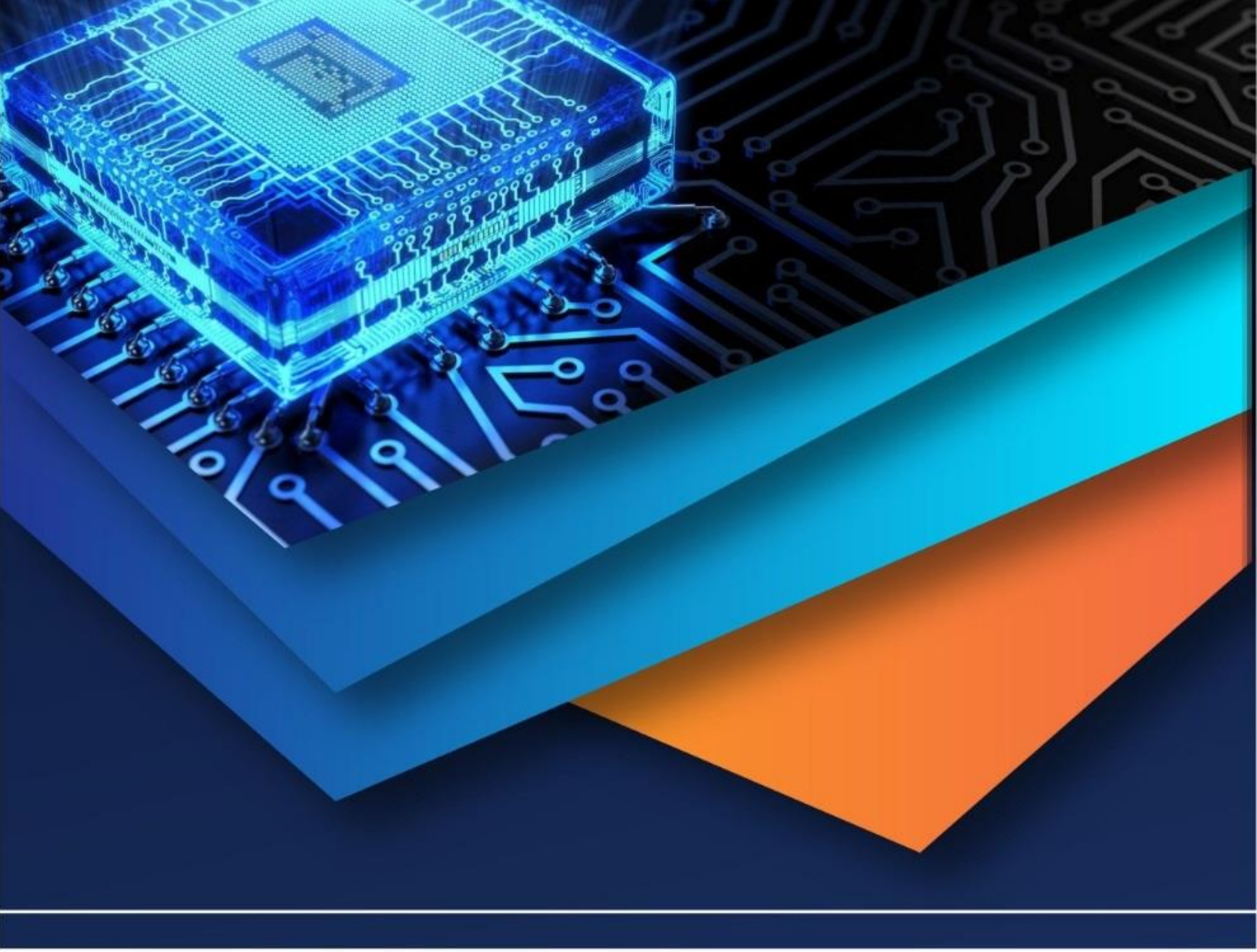
DLIME, Finding guilt entity by adding fake records and calculate digital signatures and encrypting records so that no other third party can have access to read it. By doing these it gets easy to identify the leaker when has leaked the data on public cloud by matching the digital hash on public server 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 leakes 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 thasnk the institute for providing the required facilities, Internet access and important books

REFERENCES

[1] P. Papadimitriou and H. Garcia-Molina, ”Data leakage detection”, Knowledge and Data Engineering, IEEE Transactions on, vol.23,no.1,pp.5163,2011.
[2] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, ”Secure spread spectrum watermarking for multimedia”, Image Processing, IEEE Transaction son, vol. 6, no. 12, pp. 16731687, 1997.
[3] M. Backes, N. Grimm, and A. Kate, ”Lime: Data lineage in the malicious environment”, in Security and Trust Management - 10th International Workshop, STM 2014, Wroclaw,Poland,September 10-11, 2014.Proceedings,2014,pp.183187.
[4] R. Parviainen and P. Parnes, ”Large scale distributed watermarking of multicast media through encryption”, in Proceedings of the IFIP TC6/TC11 International Conference on Communications and Multimedia Security Issues of the New Century, vol.192, 2001,pp.
[5] V. M. Potdar, S. Han, and E. Chang, ”A survey of digital image watermarking techniques”, in International Conference on Industrial Informatics,2005.INDIN05.20053 r IEEE.2005,pp.
[6] M. A. Alsalami and M. M. Al-Akaidi, ”Digital audio watermarking:survey”, School of Engineering and Technology, De Montfort University,UK 2003.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY