INFORMATION LINEAGE USING LIME FRAMEWORK

[1]G. Swagan, [2]M. Aparna, [3]Sarika. S, [4]S. Srilakshmi

[1,2,3]Students, [4]Assistant Professor
Department of Computer Science and Engineering
St. Martin's Engineering College, Hyderabad, Telangana, India

ABSTRACT: Now days, unknowingly the leakage of privacy and important data is one of the most serious security problems that companies experience in modern world technologies .When we are transferring data over the network there are many illegal user trying to get useful information. There should be proper security to data which is send to network. Now days Smartphone's use has been increased rapidly and the applications used in Smartphone's can get easy access to our confidential information. So as to avoid this we used the data lineage mechanism. We give the fake data to guilty agent. We developed and analyze a novel accountable data transfer protocol between two entities within a malicious environment by making an oblivious transfer, digital signature and robust Watermarking. At the end, we perform an trail evaluation to demonstrate the actual working of our protocol and apply the framework to the confidential data leakage scenarios of data outsourcing and social media. In general, we consider the lime framework for data transfer, to be a key step towards getting accountability by design.

Index Terms: Data Privacy Leakage Model, Watermarking, Data Leakage Prevention, Data Leakage Protection, Data Loss Prevention.

I. INTRODUCTION

In these days data leakage is a serious concern for the business organization in this tremendous networked world. Illegal methods may have serious problems resulting for an organization in case of both long term and short term. To prevent from all these unwanted and nasty activities from happening, an organized effort is needed to control the information flow inside and outside the organization. Here is our attempt to demystify the jargon surrounding the data leakage prevention procedures which will help you to choose and apply the best suitable option for your own business. Leakage describes an unwanted loss of data which moved from its original location and Lineage describes as data flow across multiple entities that take two characteristic, main roles (i.e., owner and consumer). We define the appropriate security guarantees required by such a data lineage mechanism toward identification of a guilty entity, and identify the simplifying non-repudiation and honesty assumptions. In the course of doing business, sometimes our private data must be given to the trusted third parties. For example, a hospital may give patient records to medical researchers who will advice new treatments. Similarly, a company can have partnership deal with other company and sometimes they require sharing customer data. Another enterprise may outsource its data processing, so data must be given to various other companies. The owner of the data can be called as distributor and the supposedly trusted third parties the agents. The goal is to detect when the distributors sensitive data have been leaked by agents.

II. LITERATURE SURVEY

Panagiotis Papadimitriou and Hector Garcia- Molina of Stanford University they develop model for data leakage detection for accessing guiltiness of agents. The data distributor to identify the malicious agent which leaked the information. In addition, they argue that current watermarking techniques are not practical, as they may embed extra information which could affect agents work and their level of robustness may be inadequate.

Hasan, Sion and Winslett- They summarizes the data provenance. They stated that a system that enforces recording of read and write actions in a tamper proof origin chains. This creates the possible way of verifying the origin of information in a document.

Pretschner, M. Hilty, F. Schurz, C. Schaefer, and T. Walter employ data usage control enforcement systems and preventive measures to ensure that data is transferred in distributed systems in a controlled manner preserving the well explained policie. N. P. Sheppard, R. Safavi-Naini, and P. Ogunbona finds out the main problem of an insider attack, where the data generator have multiple single entities and one of these publishes a version of the document.

A. Existing System:

In the digital era, information leakage through unintentional exposures, or intentional sabotage by disgruntled employees and malicious external entities, present one of the most serious threats to organizations. Confidential information is undoubtedly one of the most severe security threats that organizations face in the digital era. The threat is now extended to our personal lives: complete personal information is available to social media and smart phone providers and is indirectly transferred to untrustworthy third party and fourth party applications.

B. Disadvantages:

1. Determining the leaker is possible by forensic techniques, but these are usually time taking and expensive also do not always generate the desired results.

2. An attacker is able to break into the provenance information of a file the problem of data leakage in malicious environments is not protected by their approach.

C. Proposed system:

We provide a general accountable methodology in data transferring. This accountability can be directly connected with provably detecting a transmission history of data across multiple entities starting from its origin. This is known as data provenance, data lineage or source tracing. In this paper, we formalize this problem of provably associating the guilty party to the leakages, and work on the data lineage methodologies to solve the problem of information leakage in various leakage scenarios. We observe that entities in data flows assume one of two roles: owner or consumer. We introduce an additional role in the form of auditor, whose task is to identify the guilty party for the data leak, and define the exact communication between owner and consumer. In the process, we determine an optional non-repudiation assumption made between two owners, and an optional trust (honesty) assumption made by the auditor about the owners.

D. Advantages:

1. The main advantage of our model is that it records the transactions by design; i.e., it drives the system designer to consider different ways of data leakages and the corresponding accountability constraints at the design stage. This helps us to get over the existing problem where most lineage methodology are applied only after a leakage has happened.

2. We get the correct results and show that it is realizable by giving micro benchmarking results. By presenting a general applicable framework, we keep log recording of transactions as early as in the design phase of a data transfer infrastructure.

III. OVERVIEW

Data Leakage Prevention is the category of solution which help an organization to apply controls for preventing the unwanted accidental or malicious leakage of precise information to illegitimate entities in or outside the organization. Here sensitive information may refer to organization’s internal process documents, strategic business plans, intellectual property, financial statements, security policies, network diagrams, blueprints etc.

There are many fields where data leakage may occur, so it is very essential detect such kind of detection, following users may lead to data leakage. The security illiterate Majority of employees with little or no knowledge of security Corporate risk because of accidental breaches, gadget needs introduce a variety of devices to their work PCs Download software .The unlawful residents use the company IT resources in ways they shouldn’t by storing music, movies, or playing games. The malicious/disgruntled employees typically minority of employees gain access to areas of the IT system to which they shouldn’t Send corporate data (e.g., customer lists, RD, etc.) to third parties.

A. Architecture:

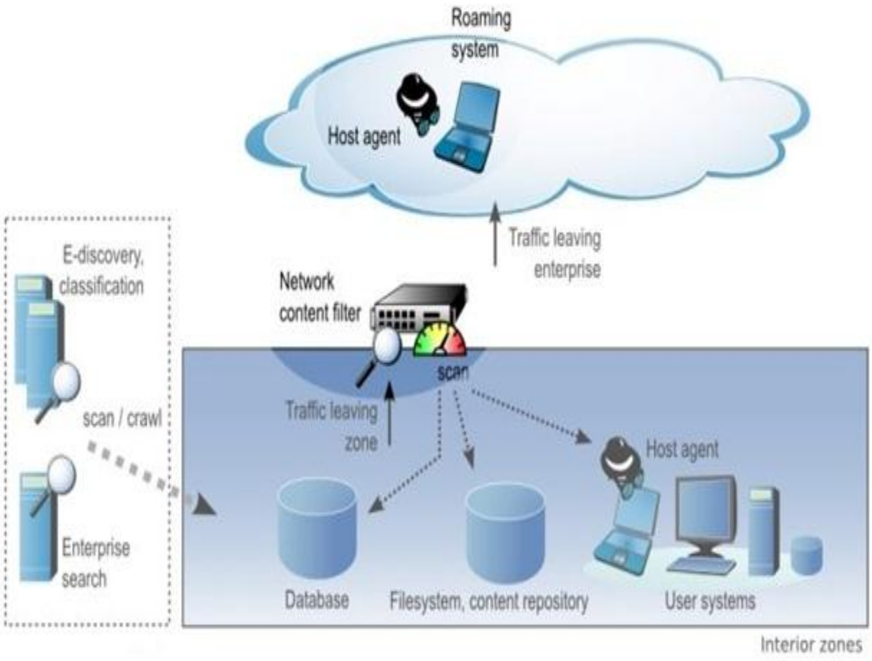

Fig: 1 System Architecture

B. Modules:

1. Data owner: The data owner is responsible for the management of documents and the consumer receives documents and can carry out some task using them.

2. Consumer: Receives the document. Consumers might transfer a document to another consumer, so we also have to consider the case of an untrusted sender. Each consumer can reveal new embedded information to the auditor to point to the next consumer and to prove his own innocence.

3. Auditor: Auditor is not involved in the transfer of documents, it is only invoked when a leakage occurs and then performs all steps that are necessary to identify the leaker.

IV. METHODOLOGY

We provide security to the file uploaded by the data owner and for this we have used AES encryption for encrypting the data owner files that are being uploaded. And the file uploaded by the data owner will have digital signature, which is used to know whether there is any attack on the file which is uploaded by the data owner as it changes when an unauthorized person tries to access the file.

The consumer need to request the key for accessing the file which is uploaded by the data owner, whereas the auditor provides the transactions done by data owner and consumers.

AES Algorithm:

The Advanced Encryption Standard is based on the concept of the symmetric block cipher. Accordingly, in order to provide protection to classified information and is implemented in software as well as hardware to encrypt sensitive data. Moreover, algorithm comprises of the three block ciphers: AES-128, AES-192 and AES-256. As the cipher encrypts and decrypts data in the blocks of 128 bits using cryptographic keys of 128, 192 and 256-bits, respectively. Certainly, AES encryption algorithm defines a varied number of transformations that are to be performed on data stored in an array. Nevertheless, the first step of the cipher is to insert the data into an array and the cipher transformations are repeated over in the form of encryption rounds. The number of rounds is determined by the help of key length, 10 rounds for 128-bit keys, 12 rounds for 192-bit keys and 14 rounds for 256-bit keys respectively.

Steps:

1. Rijndael's key is used to derive the round keys from the cipher key. It requires a 128-bit round key block for each round and adds one more.
2. The bitwise xor is used for a block of the round key as it is combined with each byte of the state.
3. According to a table where each byte is replaced with another substitution step in a non-linear manner.
4. A certain number of steps in a cyclic manner are shifted to the last three rows of the state by performing the transposition.
5. Combining the four bytes in each column. an operation is employed where it operates on the columns of the state by mixing it.
6. Finally, Add the Round-Key.
7. The final round consists of the same rounds that are mentioned above except the Mix-columns.

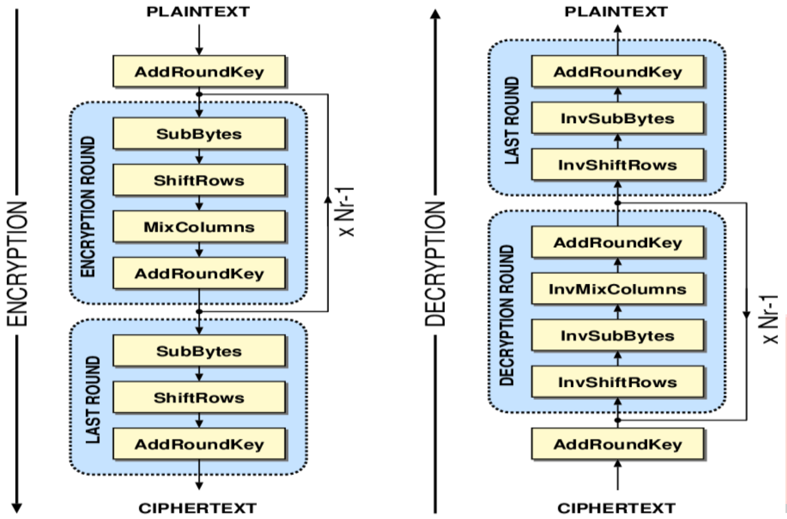

Fig: 2 Execution of AES Algorithm

V. RESULTS

The lime framework use combination of data transfer protocol, oblivious transfer and digital signature for data transfer in an encrypted Framework. LIME will determine the malicious programs who leaked the personal information or documents and provide the appropriate action to protect our data.

We prove its correctness and show that it is realizable by giving micro bench marking results. By presenting a general applicable framework, we introduce accountability as early as in the design phase of a data transfer infrastructure.

VI. CONCLUSION

We present LIME, a model for accountable data transfer across multiple entities. We define participating parties, their interrelationships and give a concrete instantiation for a data transfer protocol using a novel combination of oblivious transfer, robust watermarking and digital signatures.

Although LIME does not actively prevent data leakage, it introduces reactive accountability. Thus, it will determine malicious parties from leaking private documents and will encourage honest (but careless) parties to provide the required protection for sensitive data.

LIME is flexible as we differentiate between trusted senders (usually owners) and untrusted senders (usually consumers). In the case of the trusted sender, a very simple protocol with little overhead is possible. The untrusted sender requires a more complicated protocol, but the results are not based on trust assumptions and therefore they should be able to convince a neutral entity (e.g.a judge).

Our work also motivates further research on data leakage detection techniques for various document types and scenarios. For example, it will be an interesting future research direction to design a verifiable lineage protocol for derived data.

BIBLIOGRAPHY

[1] Mascher-Kampfer, H. St ¨ogner, and A. Uhl, “Multiple re-watermarking scenarios,” in Proceedings of the 13th International Conference on Systems, Signals, and Image Processing (IWSSIP 2006). Citeseer, 2006, pp. 53–56.

[2] P. Papadimitriou and H. Garcia-Molina, “Data leakage detection,” Knowledge and Data Engineering, IEEE Transactions on, vol. 23, no. 1, pp. 51–63, 2011

[3] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, “Secure spread spectrum watermarking for multimedia,” Image Processing, IEEE Transactions on, vol. 6, no. 12, pp. 1673–1687, 1997.

[4] B. Pfitzmann and M. Waidner, “Asymmetric fingerprinting for larger collusions,” in Proceedings of the 4th ACM conference on Computer and communications security, ser. CCS ’97, 1997, pp. 151–160.

[5] S. Goldwasser, S. Micali, and R. L. Rivest, “A digital signature scheme secure against adaptive chosen-message attacks,” SIAM J. Comput., vol. 17, no. 2, pp. 281–308, 1988.

[6] A. Adelsbach, S. Katzenbeisser, and A.-R. Sadeghi, “A computational model for watermark robustness,” in Information Hiding. Springer, 2007, pp. 145–160.

[7] M. Naor and B. Pinkas, “Efficient oblivious transfer protocols,” in Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms, 2001, pp. 448–457