ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

Privacy-Preserving Detection of Sensitive Data Exposure

Gawas Ravina, Idhate Asmita, Dahifale Aruna, Prof. Shinde Sushma

B. E Student, Dept. of Computer, Siddhant College of Engineering, Sudumbare, Pune, India

B. E Student, Dept. of Computer, Siddhant College of Engineering, Sudumbare, Pune, India

B. E Student, Dept. of Computer, Siddhant College of Engineering, Sudumbare, Pune, India

Assistant Professor, Dept. of Computer, Siddhant College of Engineering, Sudumbare, Pune, India

ABSTRACT: Research institutions and government organizations show that the number of data-leak instances have grown rapidly in recent years.Privacy Preserving of sensitive data leakage has became the most important issue in today's world. Many government organization, research institutions tells that number of data leakage instances have growing rapidly in today's world. Our Approach is to detect the data leakage by comparing the data stored on internet domain. The most important feature of this method is the data owner can safely delegate the detection operation. In some cases one can also insert the fake objects. By using the concept of fake objects we can detect the guilty agent who was responsible to leak the data to the unauthorized user. This system provides us with login, registration and authentication process. This method provides us with accurate decision with very less number of false alarms.

KEYWORDS: Network security, data leak, privacy, collection intersection, fake records, leakage model

I. INTRODUCTION

Detecting and preventing data leaks requires a set of complementary solutions, which may include data-leak detection , data confinement , stealthy malware detection, and policy enforcement.The approach of this paper is based on fast and practical one-way computation on the sensitive data which consist of sensitive emails, classified documents etc. The system consist of data owner, data agents, public server, private server and an internet domain. Here the data owner digests or fingerprints from the sensitive data.

Further it discloses only small amount of data to the data leakage detection providers. The data leakage detectors computes fingerprints from network traffic and identifies the leak in them. The purpose of this system is to identify the data leakage of sensitive data of the files or any documents. In order to avoid the leakage of sensitive data one can add random noise or can replace the values with some ranges. In some cases one can also use fake objects with are only known to the manager of the company. This fake objects are unknown to the third party, hence one can detect the guilty person. The file is divided into chunks and then uploaded. Then using secure hash algorithm we have to calculate the hash value. Finally on the basis of comparison we can detect the data has being leaked or not.

Our goal is to detect the data leakage and if possible identify the leaker of the data. In many multinational companies and business process the owner or manager gives the sensitive data to the trusted agent. This data is very sensitive and confidential and should be handled carefully. If the sensitive data is found in some unauthorized domain, it leaves the company unprotected and can also destroy the image of the company. Straightforward realizations of data-leak detection require the plaintext sensitive data. However, this requirement is undesirable, as it may threaten the confidentiality of the sensitive information. If a detection system is compromised, then it may expose the plaintext sensitive data.

II. RELATED WORK

Shingle with Rabin fingerprint was used previously for identifying similar spam messages in a collaborative setting , as well as collaborative worm containment ,virus scan, and fragment detection. In comparison, we tackle the unique data-leak detection problem in an outsourced setting where the DLD provider is not fully trusted. Such privacy

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

requirement does not exist in above models, e.g., the virus signatures are non-sensitive in the virus-scan paradigm.We propose the fuzzy fingerprint approach to meet the special privacy requirement and present genomic computation, private database query , private join operations, and distributed data mining. The provable privacy guarantees offered by SMC comes at a cost in terms of computational complexity and realization difficulty. The advantage of our approach is its concision and efficiency.

There have been several advances in understanding the privacy needs or the privacy requirement of security applications. We identify the privacy needs in an outsourced data-leak detection service and provide a systematic solution to enable privacy-preserving DLD services.

III. LITERATURE SURVEY

Literature survey gives the survey of previously proposed models for data leakage detection. One of the technique is watermarking Technique. In watermarking technique a unique code is embedded in the data or copy given to the authorized users. If this code is later found in the hands of unauthorized user then the leaker may be identified. Traditionally, leakage detection is handled by watermarking, e.g., a unique code is embedded in each distributed copy . If that copy is later discovered in the hands of an unauthorized party, the leaker can be identified. Watermarks can be very useful in some cases, but again, they involve some modification of the original data. Furthermore, watermarks can sometimes be destroyed if the data recipient is malicious, for leakage detection unobtrusive technique is required.

IV. EXISTING SYSTEM

Traditionally, leakage detection is handled by watermarking, e.g., a unique code is embedded in each distributed copy. If that copy is later discovered in the hands of an unauthorized party, the leaker can be identified. Watermarks can be very useful in some cases, but again, involve some modification of the original data. Furthermore, watermarks can sometimes be destroyed if the data recipient is malicious. Here the code is embedded in digital form like audio or video.

V. PROPOSED SYSTEM

The system consist of data owner, data agents, public server, private server, internet domain. Data owner will load the sensitive data and create chunks of the file. Then calculate the hash value and send the hash information to public as well private server. Data agents will download the data from private server and process the data. Any agent can send the data to the internet. Then the data in the public server will match the data with the internet, if the data hash is matched then the data has been leaked. In this system we will calculate the hash value by using SHA1 algorithm i.e secure hash algorithm.Figure 1 shows the architecture design of the system.

Object:

The objects involved in the projects are:

Data Owner:

Loads the sensitive data and create chunks of the file, calculate hash of each chunk and send the information to public as well as private server.

Data Agent:

Download data from private server, process the data.

Public Server:

Public Server will store the information of the Hash function and check the Data Hash.

Private Server:

Private Server will store the data.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

Admin:
Admin will compare the Hash information stored in public server and uploaded information on internet to check whether Data is leaked.

Internet:
Data agent will upload the file on Internet which is received from private server.

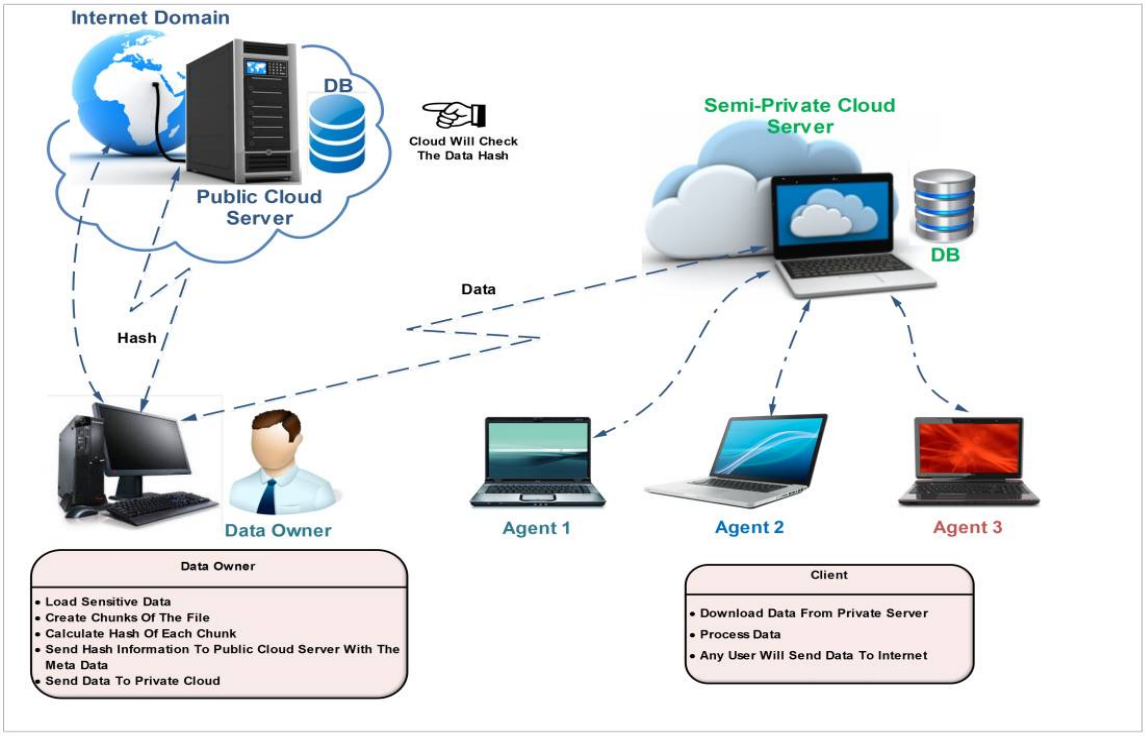

Fig 1: Architecture Design

VI. PROPOSED ALGORITHM

Secure Hash Algorithm 1(SHA 1):

Step 1: Append Padding Bits.
Step 2: Append Length.
Step 3: Prepare Processing Functions.
Step 4: Prepare Processing Constants.
Step 5: Initialize Buffers.
Step 6: Processing Message in 512-bit blocks.

VII .EXPERIMENTAL RESULTS

We have used two algorithms SHA 1 for hash calculations and AES for security. The evaluation goal is to answer the following questions:
1) Can our solution accurately detect sensitive data leak in the traffic with low false positives (false alarms) and high true positives (real leaks)?
2) Does using partial sensitive-data fingerprints reduce the detection accuracy in our system?
3) What is the performance advantage of our *fingerprint filter* over traditional Bloom filter with SHA-1?

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

4) How to choose a proper fuzzy length and make a balance between the privacy need and the number of alerts?

VIII. CONCLUSION AND FUTURE SCOPE

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 organizations.

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] K. Borders and A. Prakash, "Quantifying information leaks in outbound web traffic," in *Proc. 30th IEEE Symp. Secur. Privacy*, May 2009, pp. 129–140.
[3] 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.
[4] 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.
[5] 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.
[6] A. Kapravelos, Y. Shoshitaishvili, M. Cova, C. Kruegel, and G. Vigna, "Revolver: An automated approach to the detection of evasiveweb-based malware," in *Proc. 22nd USENIX Secur. Symp.*, 2013, pp. 637–652.
[7] 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.
[8]V. Varadharajan, "Internet filtering—Issues and challenges," *IEEE Security Privacy*, vol. 8, no. 4, pp. 62–65, Jul./Aug. 2010.

BIOGRAPHY

Gawas Ravina, Idhate Asmita, Dahifale Aruna are BE students of Siddhant College Of Engineering, Computer Department, Savitribai Phule Pune University. Our research interests are Cloud Computing, Algorithms etc.