ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 2, Issue 4, April 2016

Security Based Cryptographic System of Estimating Sensitive Data Using Privacy Preserving DLD

Kalyani.N[1], Umamaheswari.B[2]

PG Scholar, Department of CSE, Vandayar Engineering College, Thanjavur, India [1]

Assistant Professor, Department of CSE, Vandayar Engineering College, Thanjavur, India [2]

***Abstract*— In the recent years, security violations for the organizations, institutions have been increased. Due to the emerged technology development, the intruder/hacker can get the secret information and violates the secrecy. Among the various data leakages the human mistakes are the major problem. The requirement to maintain the secrecy can sometimes be compromised with the hacker. The digest fingerprint approach is used to provide privacy for the information and maintains secrecy. In this paper, we propose an efficient algorithm that encrypts the sensitive data and that is acknowledgeable to the DLD provider that makes the system to be accurate, efficient compare to the previous data. It produce minimum false alarm. The evaluation results show that our method can support accurate detection of leakages in the organization with minimum false alarm rate.**

***Index Terms*— Security violation, Intruder, Digest fingerprint, Encryption, DLD provider, false alarm, Prediction.**

I. INTRODUCTION

With the improvement in the network technologies, internet plays a vital role in human life. By using internet, tremendous information has been reached to the users. However, at the same time it also contains a lot of leakages in the information it contain. While they are using the internet, the confidential data has to be kept secret for the authorize users. There is a chance for the harmful and harmless leakages to the organization that can affect the systems while it is in use or not. Because of the intrusions in the organization and the system, the authenticated information has been spread out to the unauthorized users in the network. Malicious users or intruders can get into the system and get favor of the information by making the administrator fail to concentrate on the network operations in the organization. Therefore, there is a need to secure the organization's confidential information from the unauthorized users, while it is used in the network.

Network-based information leaks pose a serious threat to confidentiality. They are the primary means by which hackers extract data from compromised computers. The network can also serve as an avenue for insider leaks, which, according to a 2007 CSI/FBI survey, are the most prevalent security threat for organizations. Because the volume of legitimate network traffic is so large, it is easy for attackers to blend in with normal activity, making leak prevention difficult. In one experiment, a single computer browsing a social networking site for 30 minutes generated over 1.3 MB of legitimate request data the equivalent of about 195,000 credit card numbers. Manually analyzing network traffic for leaks would be unreasonably expensive and error-prone. Due to the heavy volume of normal traffic, limiting network traffic based on the raw byte count would only help stop large information leaks.

The leak measurement techniques presented here focus on the Hypertext Transfer Protocol (HTTP), the main protocol for web browsing. They take advantage of HTTP and its interaction with Hypertext Markup Language (HTML) documents and JavaScript code to quantify information leak capacity.

The basic idea is to compute the expected content of HTTP requests using only externally available information, including previous network requests, previous server responses, and protocol specifications. Then, the amount of *unconstrained* outbound bandwidth is equal to the edit distance (edit distance is the size of the edit list required to transform one string into another) between actual and expected requests, plus timing information.

Given correct assumptions about timing channel characteristics, these results may overestimate, but will never underestimate the true size of information leaks, thus serving as a tight upper bound on information leakage.

II. RELATED WORKS

The successful approaches for detection of leakages in the data in organization are based on the detection and security oriented encryption technique.

A. XUXIAN JIANG et.al (2)

Designing effective and accurate detection system for the data in the semi honest environment has become an important research issue. In this paper, we proposed an influential paper on virtual machine monitoring approach that presents with the design, implementation and evaluation of Virtual Machine (VM) watcher based on the malware detection and monitoring by addressing the semantic gap challenge. The view comparison based stealthy malware detection which involves

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 2, Issue 4, April 2016

comparing a VM's semantic views obtained from both inside and outside for possible discrepancy detection.

Out of the box execution has the unmodified off the shelf anti-malware software with improved detection accuracy. This is an extreme test to VM watcher's semantic gap-narrowing technique.

It enables cross-platform malware scanning where anti-malware software developed for one platform can be readily used for another platform non-intrusive system call monitoring in a production or honey pot VM, which elevates the tamper resistance of malware behavior observation and experimentation. Real world stealthy root kits and worms further demonstrate the power of the new malware detection and monitoring capabilities enabled by VM watcher. It has the advantage of detecting more data and more users for the unauthorized access and information in the organization. It has the comparing speed higher than the other methods.

B. *Fang Liu et al (5)*

The formal approach that allows a workflow architect to perform detection of exposure of sensitive data. It illustrates an efficient paper that presents the exposure of sensitive data in storage and transmission poses a serious threat to organizational and personal security. Data leak detection aims at scanning content (in storage or transmission) for exposed sensitive data. Because of the large content and data volume, such a screening algorithm needs to be scalable for a timely detection. The privacy goal of the system is to prevent the sensitive data from being exposed to DLD provider or entrusted nodes. Map Reduce is a programming frame-work for distributed data intensive applications.

It has been used to solve security problems such as spam filtering, Internet tracking analysis and log analysis. Map Reduce algorithms can be deployed on nodes in the cloud or in local computer clusters. It present a series of new Map Reduce parallel algorithms for distributed computing the sensitivity of content based on its similarity with sensitive data patterns. The similarity is based on collection intersection (a variant of set intersection that also counts duplicates).

Our detection provides the privacy enhancement to preserve the confidentiality of sensitive data during the outsourced detection. Because of this privacy enhancement, our Map Reduce algorithms can be deployed in distributed environments where the operating nodes are owned by third-party service providers. It designs new Map Reduce algorithms for computing collection intersection for data leak detection.

It uses the algorithms support a useful privacy-preserving data transformation that enables the privacy-preserving technique to minimize the exposure of sensitive data during the detection. It supports the secure out sourcing of the data leak detection to entrusted Map Reduce and cloud providers. The method identifies anomalous and drastic increase in the amount of information carried by the traffic. It has to reduce the size of the trained data and the test sensitive data. It uses the Hadoop tool that specifically performs the map reducing for the data in the large scale organization. In addition to that it reduces the space used for the storage of the data. It uses the large data set for the detection of the leakage.

C. *Heng Yin et.al (3)*

To observe the malicious information access and processing behavior is the fundamental trait of numerous malware categories breaching users privacy (including key loggers, password thieves, network sniffers, stealth backdoors, spyware and root kits), which separates these malicious applications from benign software. It proposed a system, Panorama, to detect and analyze malware by capturing this fundamental trait. In this paper, extensive experiments Panorama successfully detected all the malware samples and had very few false positives. Furthermore, by using Google Desktop as a case study, it shows that our system can accurately capture its information access and processing behavior, and it can confirm that it does send back sensitive information to remote servers in certain settings. It also gives the believe that a system such as Panorama will offer indispensable assistance to code analysts and malware researchers by enabling them to quickly comprehend the behavior and inner workings of an unknown sample. It creates the taint graph by using the information in the comparison of the training dataset and the test input data set. We have designed and developed Panorama, an end-to-end system that can automatically analyze samples for malicious information access and processing behavior. As a critical component of Panorama, we have designed and developed a whole-system, fine-grained, operating-system-aware, dynamic taint tracking system to enable us to monitor and investigate the unknown sample's information access and processing behavior to sensitive information.

By using the graph the data's can be easily find the malware and intrusions in the system. Our system detected all the malware samples and had very few false positives. The malware samples include a wide range of different classes of malware, such as key loggers, password sniffers, packet sniffers, stealth backdoors, root kits and spyware. It proposes the whole-system fine-grained taint analysis to discern fine-grained information access and processing behavior of a piece of unknown code. This behavior captures the intrinsic characteristics of a wide-spectrum of malware, including key loggers, password sniffers, packet sniffers, stealth back doors, BHO-based spyware, and root kits. Thus, the detection and analysis relying on it cannot be easily evaded. To evaluate the effectiveness of this approach, we have designed and developed a system, called Panorama.

D. *Gunter Karjoth et.al (8)*

It describes a privacy policy model that protects personal data from privacy violations by means of enforcing enterprise-wide privacy policies. By extending Jajodia et al. „s Flexible Authorization Framework (FAF) with grantors and obligations, we create a privacy control language that includes user consent, obligations, and distributed administration. Conditions impose restrictions on the use of

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 2, Issue 4, April 2016

the collected data, such as modeling guardian consent and options. Access decisions are extended with obligations, which list a set of activities that must be executed together with the access request. Grantors allow defining a separation of duty between the security officer and the privacy officer. We presented a privacy policy model for enterprises that can serve as the basis for an internal access control system to handle received data in accordance with privacy standards. Thus, the data subject providing his/her personal data has the assurance that the enterprise receiving the information will handle it according to the stated privacy policy. The enterprise as well can verify that its business practices are not in conflict with the privacy policies posted on its Web site, which are usually considered a binding contract between the site owner and the people visiting the site. It has the Access control which is a crucial concern to build secure IT systems and, more specifically, to protect the confidentiality of information. However, access control is necessary, but not sufficient. Actually, IT systems can manipulate data to provide services to users. The results of a data processing may disclose information concerning the objects used in the data processing itself. Therefore, the control of information flow results fundamental to guarantee data protection. In the last years many information flow control models have been proposed. However, these frameworks mainly focus on the detection and prevention of improper information leaks and do not provide support for the dynamical creation of new objects.

It has the separate module for flexible authorization. It has the frame structure and tree structure. It has the drawbacks that it cannot support for the DBMS and the framework is complicated to construct and implemented.

III. PROPOSED DESIGN

Data leakage can be detected at the both server side and admin side. DLD can be evaluated based on fuzzy finger print mechanism**.** Fuzzy fingerprint technique enhances data privacy during the data-leak detection operations. Our approach is based on a fast and practical one-way computation on the sensitive data (SSN records, classified documents, sensitive emails, etc.). It enables the data owner to securely delegate the content-inspection task to DLD providers without exposing the sensitive data. Using our detection method, the DLD provider, who is modeled as an honest-but-curious (aka semi-honest) adversary, can only gain limited knowledge about the sensitive data from either the released digests, or the content being inspected. Host authenticated auditing algorithm is used to enhance the security at customer side**.** Host-assisted mechanism is used for the complete data-leak detection for large-scale organizations. In preprocessing i.e., encryption, ECC algorithm is implemented. ECC is one of the first practical public-key cryptosystems and is widely used for secure data transmission. In such a cryptosystem, the encryption is public and differs from the decryption key which is kept secret. In ECC, this asymmetry is based on the practical difficulty of factoring the product of two large prime numbers, the factoring problem. The ECC algorithm was described and placed in the public domain. What others found was that while it offered greater potential security it was slow. ECC focused on its efforts on creating better implementations of the algorithm to improve its performance

A. User Registration

A user is a communication medium for the data that has been stored in the organization and the data leak detection system. They need to register into the organization and verify to be an authorized user. The user has to provide specific username and password and their fingerprint for their security of the data. It is consider being one constraint system.

B. Template Creation

For every user, the password template has been created along with individual user name and password. The template has been checked and verified by the administrator to recognize the authenticated user and its password

C. Data Collection

Data collection is a input component of an information system. Data is collected is in the form of excel data and text format. These datasets are known as historical datasets. These data has been encrypted and decrypted, so that it has to be sending and receive by the authorized user.

D. Data Leakage Detection

The data that has been stored in the organization has to be kept so secure and privacy. The data which is more sensitive like private data and confidential data should kept more secure. The leaked information will be identified by the admin with the unauthorized user access.

E. Encryption/Decryption

Alice represents her text or data to send as a point Pm Alice sends Bob a pair of points:

$$Cm= \{k*G, Pm + k*PB\}$$

Where k = randomly chosen integer

Bob decrypts the message using his private key:

$$Pm + k*P – nB (k*G) = Pm + k(nB *G) - nB (k*G) = Pm$$

F. System Architecture

Detection of data leakage in the sensitive data has been performed with the help of encryption and decryption technique for the user data. In the fig.1, the overall system architecture is depicted. Initially, the user has to register into the system with the username and password and the specified fingerprint of the user. The registered user information has been created as a template and stored in the administrator database. The registered user has to login into system with the appropriate username and password to access the data that has been already stored in the database. The user has to be authenticated with the

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 2, Issue 4, April 2016

template that they have stored in the database with the login details. The system has to compare the username, password and the fingerprint of the recognized user, if they are matched with the stored information they will be considered as a recognized user. If the information got mismatch with any of the user information like username, password and fingerprint then it will be considered as an unauthorized access to the system.

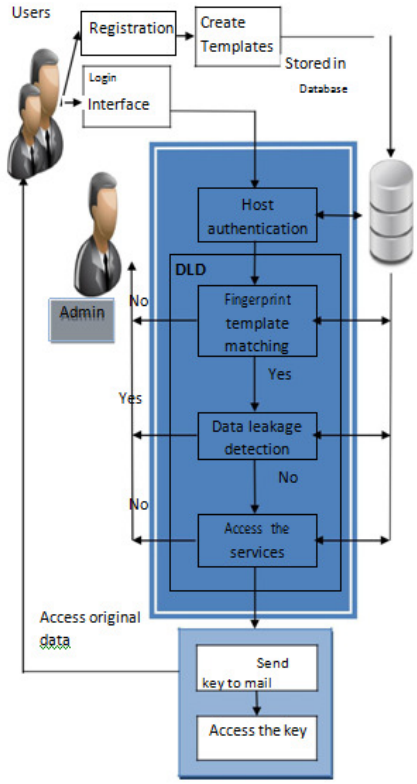

Fig 1 Architecture Overview

If the system detects any data leak in the system then the user is denied to access the files in the system. If it is an authorized access then they need to perform the sending and receiving of files from the user and the admin. The files that we have sent and receive may sometimes be compromised with the hacker so it needs to make encryption and decryption with the efficient algorithm like Elliptic Curve Cryptography. The user has to make the transaction of data to be secure with the help of the algorithms. The key transaction has to be performed by the user's personal mail and by that the key has to be kept so secret. The DLD provider will manage all the information in the database and the user has to request for the file and the authentication has to be performed and the recognized user can get the information.

IV. EXPERIMENTAL ANALYSIS

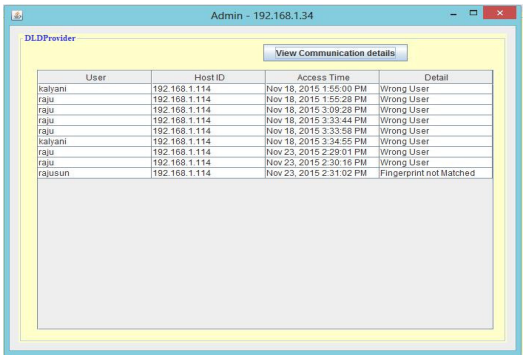

Fig 2 User Communication Details

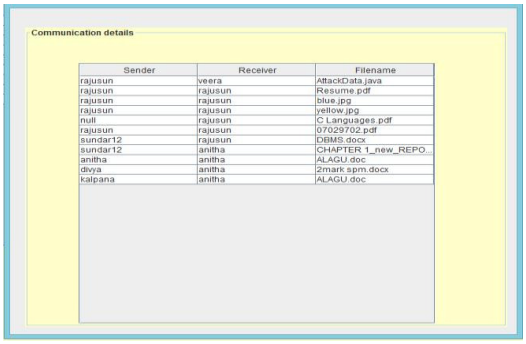

Fig 3 Data provider Details

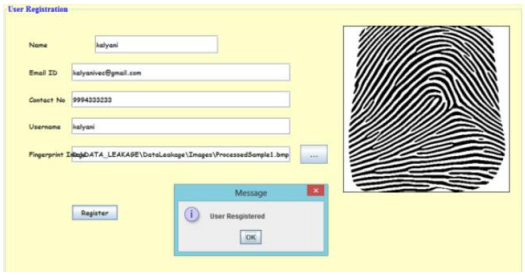

Fig 4 User Registration

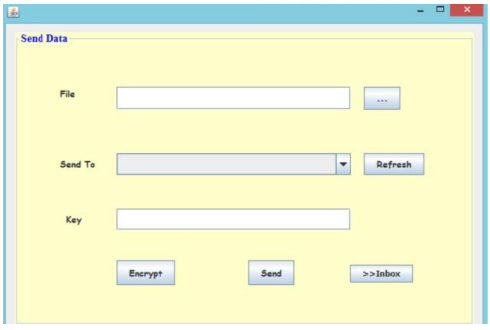

Fig 5 Data Requestor Details

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 2, Issue 4, April 2016

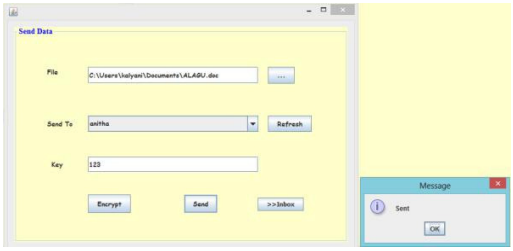

Fig 6 Encryption and Message Transmission

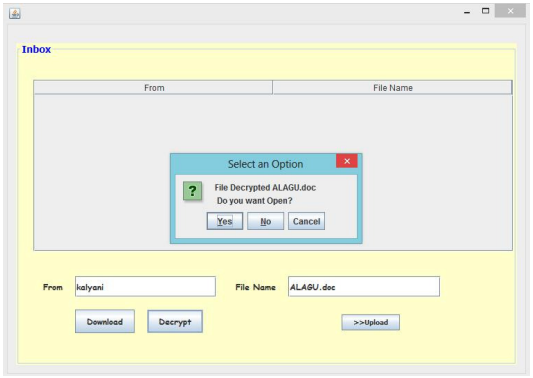

Fig 7 Decryption and Message Receive

V. CONCLUSIONS

Data services and secure transmission of sensitive data becomes an increasing number of challenges t. As a result, service-r the web service providers. Relevant data become too big to be efficiently processed by traditional approaches. DLD is done automatically using Elliptic Curve Cryptography. 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. By conducting extensive experiments to validate the accuracy, privacy, and efficiency of our solutions. From that the data has been transmitted to the appropriate user and the admin. The information that has been kept secret has been maintained.

REFERENCES

[1] Ahmadi .M, and Geravand .S "Bloom filter applications in network security: A state-of-the-art survey," Comput. Netw., vol. 57, no. 18, pp. 4047– 4064, Dec. 2013.

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

[3] Butt . A. R., Liu .F, Shu .X, Yao .D, and "Privacy-preserving scanning of big content for sensitive data exposure with MapReduce," in Proc.ACM CODASPY, 2015.

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

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

[6] Jiang .X, Wang .X, and Xu .D, "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.

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