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. 3, Issue 12, December 2015

A Review on Techniques of Privacy-Preserving Detection of Sensitive Data Exposure Mechanism

Santosh R. Ambre[1], Prof. MS Arti Mohanpurkar[2]

Student, Dept. of CS, Dr. D Y Patil School of Engineering & Technology, Pune, Savitribai Phule Pune University, Pune, India

Professor, Dept. of CS, Dr. D Y Patil School of Engineering & Technology, Pune, Savitribai Phule Pune University, Pune, India

ABSTRACT: Lately, insights from examination establishments, security firms and government associations demonstrate that the quantities of information break cases have become quickly. Among different information release cases, fundamental driver of information misfortune are one of the human missteps. There exist arrangements identifying incidental delicate information spills created by human oversights and to give alarms to associations. In this review paper, we introduce a security saving information spill recognition (DLD) arrangement where an extraordinary arrangement of touchy information overviews is utilized as a part of identification. The assessment results demonstrate that the discovery system can bolsters exact recognition with little number of false cautions under different information spill situations. The benefit of this technique is that it empowers the information proprietor to securely appoint the location operation without uncovering the touchy information to the supplier. Estimations from security firms, research establishments also, government affiliations exhibit that the amount of data opening events have grown rapidly starting late. Here human misunderstandings are one of the major drivers of data disaster. Such a technique generally speaking requires the recognizable proof operation to be coordinated in secret. In this paper, we present a protection saving data spill area (DLD) answer for span the issue where an exceptional plan of sensitive data outlines is used as a piece of ID. The upside of our framework is that it engages the data proprietor to safely choose the disclosure operation to a semi legitimate supplier without revealing the tricky data to the supplier. We depict how Internet organization suppliers can offer their customers DLD as an additional organization with strong insurance guarantees. Here framework can reinforce precise acknowledgment with minimal number of false alerts under diverse data spill circumstances

KEYWORD: Data Leak, Network Security, Privacy, Collection Intersection.

I. INTRODUCTION

Distinguishing and forestalling information spills requires an arrangement of correlative arrangements, which might incorporate information spill discovery, information control [2][3], stealthy malware identification and strategy authorization. System information spill identification (DLD) normally looks for any events of delicate information designs and performs profound bundle examination (DPI). DPI is a method to investigate payloads of TCP/IP parcel for examining application layer information, e.g., HTTP header/content. Alarms are activated and movement passes a limit when the measure of delicate information found.

There are two sorts of info groupings in information spill discovery model: touchy information successions and content arrangements. (1) Sensitive information contains the touchy data that can't be presented to unapproved parties, e.g.,

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. 3, Issue 12, December 2015

restrictive reports, clients' records. Touchy information can likewise be divided to little delicate information grouping. (2) Content is the information to be investigated events of touchy information designs. The location need to segment the first substance stream into substance sections [1].

In this study paper, the information spill location arrangement which can be conveyed and outsourced in a semi fair recognition environment. The fluffy unique mark procedure is utilized improves information security amid information spill identification operations. This methodology depends on handy restricted calculation on the touchy information (arranged reports, delicate messages, SSN records, and so on.). This empowers the information proprietor to safely appoints the substance investigation undertaking to DLD suppliers without uncovering the touchy information. By utilizing the discovery system, the DLD supplier, who is demonstrated as fair however inquisitive foe, and it just increase restricted information of touchy information from either the discharge digests. Utilizing these strategies, an Internet administration supplier (ISP) gives information release discovery as an extra administration for its clients. The location method of information proprietor process a unique arrangement of fingerprints or processes from the touchy data[1]. The DLD supplier registers fingerprints from system movement and recognizes potential breaks.

To prevent the DLD provider from gathering exact knowledge about the sensitive data and the collection of potential leaks is composed of noises and real leaks. The data owner who post-processes the potential leaks sent back by the DLD provider and then determines whether there is any real data leak. As indicated by a report from Risk Based Security (RBS) , the quantity of released touchy information records has expanded drastically amid the last couple of years, i.e., from 412 million in 2012 to 822 million in 2013. Purposely arranged assaults, incidental holes (e.g., sending private messages to unclassified email records), and human missteps (e.g., relegating the off-base benefit) lead to the vast majority of the information spill episode. Distinguishing and forestalling information holes requires an arrangement of integral arrangements, which may incorporate information spill recognition, information control stealthy malware recognition, and strategy requirement. Network information spill location (DLD) normally performs profound bundle review (DPI) and looks for any events of delicate information designs. DPI is a method to examine payloads of IP/TCP bundles for reviewing application layer information, e.g., HTTP header/content. Alarms are activated when the measure of touchy information found in movement passes a limit. The recognition framework can be conveyed on a switch or incorporated into existing system interruption recognition frameworks (NIDS).Direct acknowledge of information break discovery require the plaintext touchy information. Nonetheless, this prerequisite is undesirable, as it may undermine the classification of the touchy data. On the off chance that a discovery framework is traded off, at that point it may uncover the plaintext touchy information (in memory).

II. RELATED WORK

1. *A Review on Privacy Preserving Data Mining: Techniques and Research Challenges*

From this paper we Refer-

In today's world, privacy is the major concern to protect the sensitive data. People are very much concerned about their sensitive information which they don't want to share. Our survey in this paper focuses on the existing literature present in the field of Privacy Preserving Data Mining. From our analysis, we have found that there is no single technique that is consistent in all domains. All methods perform in a different way depending on the type of data as well as the type of application or domain. But still from our analysis, we can conclude that Cryptography and Random Data Perturbation methods perform better than the other existing methods. Cryptography is best technique for encryption of sensitive data. On the other hand Data Perturbation will help to preserve data and hence sensitivity is maintained .In future, we want to propose a hybrid approach of these techniques.

2. *A Survey: Privacy Preservation Techniques in Data Mining*

From this paper we Refer-

The main objective of privacy preserving data mining is developing algorithm to hide or provide privacy to certain sensitive information so that they cannot be disclosed to unauthorized parties or intruder. Although a Privacy and accuracy in case of data mining is a pair of ambiguity. Succeeding one can lead to adverse effect on other. In this, we made an effort to review

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. 3, Issue 12, December 2015

a good number of existing PPDM techniques. Finally, we conclude there does not exists a single privacy preserving data mining algorithm that outperforms all other algorithms on all possible criteria like performance, utility, cost, complexity, tolerance against data mining algorithms etc. Different algorithm may perform better than another on one particular criterion.

3. Secure Data Detection for Confidential Data Exposure

From this paper we Refer-

From this we conclude that the privacy-preserving detection method is used to secure sensitive data from the exposure. Using some special digests the disclosure of the sensitive data is kept to minimum during detection. The conducted extensive experiments to validate the accuracy, privacy, and efficiency of our solutions. We propose an active data leakage prevention model. By adding a secure data container to execute security prevention mechanism, the model can ensure that data is used in a trusted and controllable environment. Based on the model an implementation framework of active data leakage protection is given.

4. Data Leak Detection as a Service: Challenges and Solutions

From this paper we Refer-

Preventing sensitive data from being compromised is an important and practical research problem. We proposed a novel fuzzy fingerprint framework and algorithms to realize privacy-preserving data-leak detection. Using special digests, the exposure of the sensitive data is kept to a minimum during the detection. We described its application in the cloud computing environments, where the cloud provider naturally serves as the DLD provider.We defined our privacy goal by quantifying and restricting the probability that the DLD provider identifies the exact value of the sensitive data. We presented the protocols and data structures including a Bloom-filter based fuzzy fingerprint filter. Our extensive experiments validate the accuracy, privacy, and efficiency of our solutions. For future work, we will test our current solution on binary sensitive data, and then focus on designing solutions that will efficiently prevent the leakage of complex data types, especially dynamically-changing sensitive data, such as source code of programs and sensitive documents constantly being modified.

5. A Technique for Avoiding Data Leakage and Misuse

From this paper we Refer-

This paper reviews the various techniques for data leakage and misuse detection. But still none of the techniques gives the sensitivity level of the damage caused to data while providing the data to the insider.

Thus, this paper introduces a system which will measure the risk of damage that can be caused when data is exposed to the insider. This involves collecting knowledge from the domain expert as well as use of the risk measuring algorithm. Measuring risk before data exposure will help administrator to take proper action to prevent or minimize the damage. Also, this system will alter the data in such a way that the risk will be reduced and at the same time modified data will be useful for performing desired task.

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. 3, Issue 12, December 2015

III. ARCHITECTURE

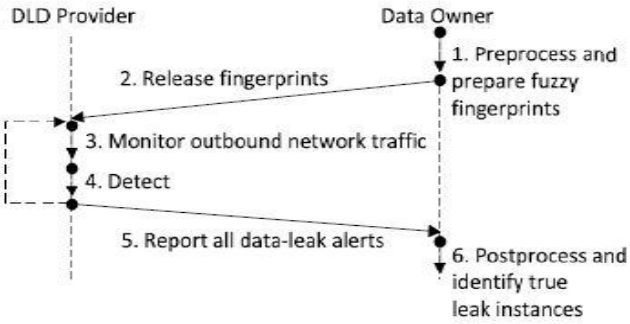

Figure 1: Our Privacy-preserving Data-Leak Detection Model.

Explanation

Overview of Privacy-Enhancing DLD, the privacy-preserving data-leak detection method minimizes the knowledge that a DLD provider may collect during the process and supports practical data-leak detection as a service. The Figure 1 lists the six operations executed by the DLD provider and the data owner in our protocol. They include PREPROCESS runs by the data owner to prepare sensitive data of digests. RELEASE for the data owner to send the digests to the DLD provider that MONITOR and then DETECT for the DLD provider to collect outgoing traffic of the organization, compute digests of traffic content and then it identify potential leaks, REPORT for the DLD provider to return data-leak alert to the data owner where there may be false alarms or positives and POSTPROCESS for the data owner to pinpoint true data-leak instances [1]. To achieve the goal of privacy the data owner generates a special type of digests, which call fuzzy fingerprints. The purpose of fuzzy fingerprint is hide the true sensitive data in a crowd. The fuzzy fingerprint prevents the DLD provider from learning its exact value. the noise tolerance property. The Rabin fingerprint algorithm has a unique min-wise independence property [6], which supports fast random fingerprints selection (in uniform distribution) for partial fingerprints disclosure. Rabin fingerprints are computed as polynomial modulus operations, and can be implemented with fast XOR, shift, and table look-up operations. The shingle-and-fingerprint process is defined as - A sliding window is used to generate first q-grams on an input binary string. The fingerprints of q-grams are then computed.

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. 3, Issue 12, December 2015

IV. PROPOSED SYSTEM MECHANISM

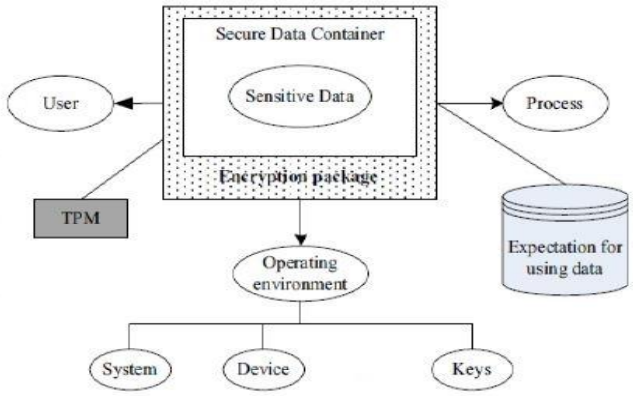

Fig. 2. Active Data Leakage Prevention Model

Explanation

The main idea of active data leakage prevention model is to add Secure Data Container (abbreviated as SDC) to achieve active security, as shown in Figure.2, SDC is equivalent to adding a protection shell for documents. Data and security attributes are encrypted andProcess can only use the decrypted data in SDC. All operations to write data to non-trusted storage or sent data to non-trusted process will be prohibited. Neither authorized normal users nor illegal processes can leak protected sensitive data out.The integrity itself and data encryption or decryption keys of the SDC are guaranteed by the underlying TPM module. When processes access sensitive data, SDCwill actively detect the integrity and security of the related usage environment, involving platforms, hardware platforms and decryption keys, etc. It ensures that data is used by authorized users in trusted environment and complies with data protection usage expectation by authenticating users and processes.

VI.CONCLUSION

From this we conclude that the privacy-preserving detection method is used to secure sensitive data from the exposure. Using some special digests the disclosure of the sensitive data is kept to minimum during detection. The conducted extensive experiments to validate the accuracy, privacy, and efficiency of our solutions. For future work, the plan to focus on designing a host-assisted mechanism for the complete data-leak detection for large-scale organizations.

REFERENCES

1. ShwetaTaneja" A Review on Privacy Preserving Data Mining:Techniques and Research Challenges" ShwetaTaneja et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 2310-2315
2. HinaVaghashia"A Survey: Privacy Preservation Techniques in Data Mining" International Journal of Computer Applications (0975 – 8887) Volume 119 – No.4, June 2015
3. S.Saranya "Secure Data Detection for Confidential Data Exposure" International Journal of Innovative Research in Engineering Science and Technology December 2015 ISSN 2320 – 981X
4. XiaokuiShu "Data Leak Detection As a Service: Challenges and Solutions" Department of Computer Science, Virginia Tech
5. Rashmi Bhat1 "A Technique for Avoiding Data Leakage and Misuse International Journal of Advance Research in Computer Science and Management Studies" Volume 2, Issue 2, February 2014