e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 13, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| | Monthly Peer Reviewed & Refereed Journal |

|| Volume 7, Issue 13, April 2024 ||

International Conference on Intelligent Computing & Information Technology (ICIT-24)

Organized by

Erode Sengunthar Engineering College, Erode, Tamilnadu, India

Protecting Confidential Data Against Data Breach Using Data Leak Detection (DLD)

S.Revathi[1], S.Saranya[2], M.Prem Kumar[3]

Assistant Professor, Artificial Engineering and Machine Learning, Mahendra Institute of Technology, Tamil Nadu, India[1]

Assistant Professor, Department of Information Technology, Mahendra Institute of Technology, Tamil Nadu, India[2,3]

ABSTRACT : Statistics from tackle security problems, research institutions and government organizations show that the number of data-leak instances have grown-up rapidly in recent years. There exist solutions detecting unintentional sensitive Data leaks caused by human mistakes and to provide alerts for organizations conventional technologies for data leakage avoidance rely on the terminal or boundary control which is difficult for data leakage in spread environment. However, this secrecy requirement is hard to satisfy in practice, as detection servers may be compromised or outsourced. In this paper, we present a privacy preserving data-leak detection (DLD) solution to solve the issue where a special set of sensitive data digests is used in detection. The advantage of our method is that it enables the data owner to securely hand over the detection operation to a semi honest provider without revealing the sensitive data to the provider. For this, user profile is created using the local knowledge base so that only required data can be given for avoiding data leakage and misuse.

KEYWORDS: Data Leak, Network Security, Privacy, Collection Intersection, user profile.

I. INTRODUCTION

Detecting and preventing data leaks requires a set of corresponding solutions, which may include data-leak detection, data confinement, stealthy malware detection and policy enforcement. Detection or avoidance of data leakage and misuse is a great difficult issue for organizations. Network data-leak detection (dld) typically searches for any incidence of sensitive data patterns and performs deep packet inspection (dpi). Dpi is a technique to analyze payloads of tcp/ip packet for inspecting application layer data, e.g., http header/content. Alerts are triggered and traffic passes a threshold when the amount of sensitive data found. This challenge becomes more difficult when trying to detect and/or prevent data leakage and misuse performed by an insider having legal permissions to access the organization's systems and its sensitive data.

The data-leak detection solution which can be deployed and outsourced in a semi honest detection environment. The fuzzy fingerprint technique is used enhances data privacy during data-leak detection operations.
By compressing data security prevention boundary to data itself, they make all kinds of security control mechanisms associated with data usage more closely, and change the static and passive data protection conception. "Kang" [1] also designed a hardware architecture which integrates data and signature management software especially for data leakage protection of mobile storage device. "Kuhn" [2] applied trusted computing to disk encryption and secure latent control, preventing data from leaking in work-in progress. "Yin Fan" [3] proposed a reliability-based distributed data leakage protection model to prevent extend files from leaking. Berger's [4] trusted virtual datacenter (Trusted Virtual Datacenter, TVDc) structure that is grouping the virtual machine and the underlying data resources further based on TVDs according to security needs of centralized data services, etc.

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.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| | Monthly Peer Reviewed & Refereed Journal |

|| Volume 7, Issue 13, April 2024 ||

International Conference on Intelligent Computing & Information Technology (ICIT-24)

Organized by

Erode Sengunthar Engineering College, Erode, Tamilnadu, India

II. MODEL AND OVERVIEW

We abstract the privacy-preserving data-leak detection problem with a threat model, a security goal and a privacy goal.

A. Security Goal and Threat Model

We categorize three causes for sensitive data to appear on the outbound traffic of an organization, including the legitimate data use by the employees.

Case I Inadvertent data leak:
The sensitive data is accidentally leaked in the outbound traffic by a legitimate user. This paper focuses on detecting this type of accidental data leaks over supervised network channels. Inadvertent data leak may be due to human errors such as forgetting to use encryption, inaccurately forwarding an internal email and attachments to outsiders, or due to application flaws. A supervised network channel could be an unencrypted channel or an encrypted channel where the content in it can be extracted and checked by an authority

Case II Malicious data leak:
A rogue insider or a piece of stealthy software may take sensitive personal or organizational data from a host. Because the malicious opponent can use strong private encryption, steganography or secret channels to disable content-based traffic inspection, this type of leaks is out of the scope of our network-based solution.

Case III Legitimate and intended data transfer:
The sensitive data is sent by a legitimate user intended for legitimate purposes. In this paper, we assume that the data owner is aware of legitimate data transfers and permits such transfers. So the data owner can tell whether a piece of sensitive data in the network traffic is a leak using legitimate data transfer policies.

B. Privacy Goal and Threat Model

To prevent the DLD provider from ahead knowledge of sensitive data during the detection process, we need to set up a privacy goal that is corresponding to the security goal above. We model the DLD provider as a semi-honest challenger, who follows our protocol to carry out the operations, but may attempt to gain knowledge about the sensitive data of the data owner. Our privacy goal is defined as follows. The DLD provider is given digest of sensitive data from the data owner and content of network traffic to be examined.. We present a privacy-preserving DLD model with a new fuzzy fingerprint mechanism to improve the data defense against semi-honest DLD provider. We generate digests of sensitive data through a one-way function, and then hide the sensitive values among other non-sensitive values via fuzzification.

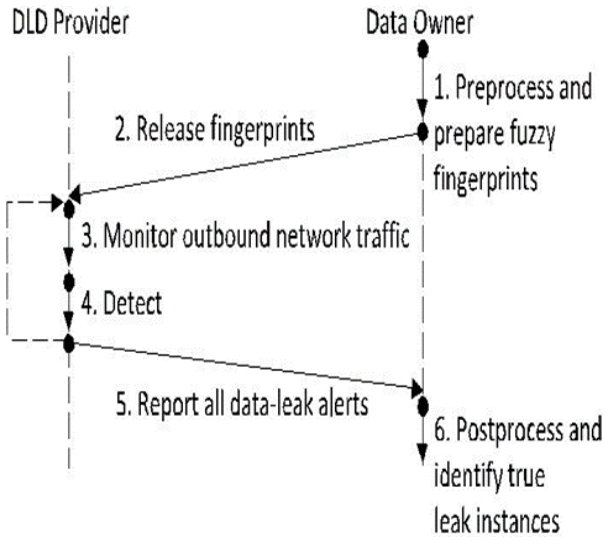


Fig.1. Our Privacy-preserving Data-Leak Detection Model.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| | Monthly Peer Reviewed & Refereed Journal |

|| Volume 7, Issue 13, April 2024 ||

International Conference on Intelligent Computing & Information Technology (ICIT-24)

Organized by

Erode Sengunthar Engineering College, Erode, Tamilnadu, India

The traffic content is accessible by the DLD provider in plaintext. Therefore, in the event of a data leak, the DLD provider may learn sensitive information from the traffic, which is to be expected for all deep packet inspection approaches. Our solution limits the amount of maximal information learned during the detection and provides quantitative certification for data privacy.Our goal is to offer DLD provider solutions to scan massive content for sensitive data exposure and minimize the possibility that the DLD provider learns about the sensitive information

Scalability: the ability to process content at a variety of scales, e.g., megabytes to terabytes, enabling the DLD provider to offer on-demand content inspection.

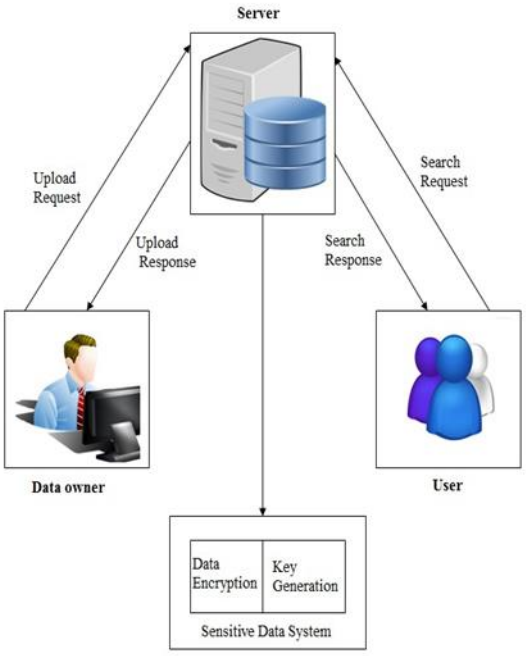
Privacy: the ability to keep the sensitive data confidential, not disclosed to the DLD provider or any attacker breaking into the detection system.

Accuracy: the ability to identify all leaks and only real leaks in the content, which implies low false negative/positive rates for the detection.

III. PRIVACY-PRESERVING DATA-LEAK DETECTION

Network-based data-leak detection (DLD) technique, the main feature of which is that the detection does not reveal the content of the sensitive data. Instead, only a small amount of specialized digests are needed. Our technique referred to as the *fuzzy fingerprint* detection – can be used to detect accidental data leaks due to human errors or application flaws. The privacy-preserving feature of our algorithms minimizes the exposure of sensitive data and enables the data owner to safely delegate the detection to others.

The privacy-preserving data-leak detection problem with a threat model, a security goal and a privacy goal.The two most important players in our abstract model: the organization (i.e., data owner) and the data-leak detection(Server).DLD provider inspects the network traffic for potential data leaks. The inspection can be performed offline without causing any real-time delay in routing the packets. The DLD provider may attempt to gain knowledge about the sensitive data.



| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| | Monthly Peer Reviewed & Refereed Journal |

|| Volume 7, Issue 13, April 2024 ||

International Conference on Intelligent Computing & Information Technology (ICIT-24)

Organized by

Erode Sengunthar Engineering College, Erode, Tamilnadu, India

A.High Secure Encryption:

The process of making data unreadable by other humans or computers for the purpose of preventing others from gaining access to its contents. Encrypted data is generated using an encryption program such as PGP, encryption machine, or a simple encryption key and appears as garbage until it is decrypted. As first publicly accessible, from the NSA for the classification "top secret" approved cipher, the Advanced Encryption Standard (AES) is one of the most frequently used and most secure encryption algorithms available today.

B.Collection Intersection:

For protecting dynamically changing data such as source code or documents under constant development or keystroke data, the digests need to be continuously updated for detection, which may not be efficient or practical. We raise the issue of how to efficiently detect dynamic data with a network-based approach as an open problem to investigate by the community.

C.Data Leakage

Our privacy-preserving data-leak detection method supports practical data-leak detection as a service and minimizes the knowledge that a DLD provider may gain during the process. Lists the six operations executed by the data owner and the DLD provider in our protocol. They include PREPROCESS run by the data owner to prepare the digests of sensitive data, RELEASE for the data owner to send the digests to the DLD provider, MONITOR and DETECT for the DLD provider to collect outgoing traffic of the organization, compute digests of traffic content, and identify potential leaks, REPORT for the DLD provider to return data-leak alerts to the data owner where there may be false positives (i.e., false alarms), and POSTPROCESS for the data owner to pinpoint true data-leak instances. When trying to detect and/or prevent data leakage and misuse performed by an insider having legitimate permissions to access the organization's systems and its sensitive data.

IV. AN ACTIVE DATA LEAKAGE PREVENTION MODEL

1.Model idea

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 and packaged, which are transparent to the upper applications. SDC is a dynamic virtual isolation environment for processes, controlling file access, network access and inter-process communication for processes accessing sensitive content.

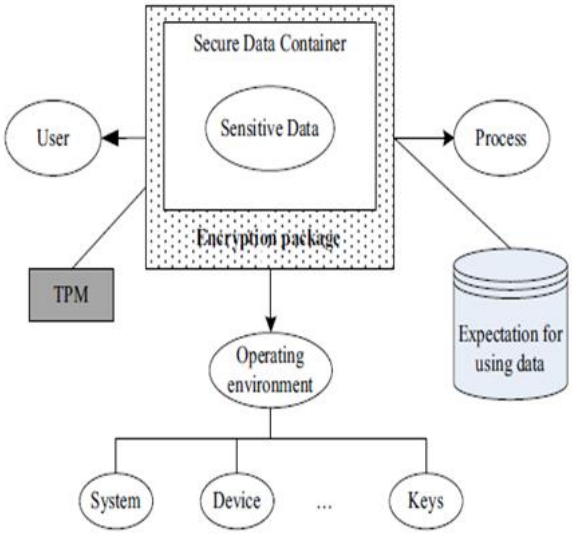


Fig. 3 Active Data Leakage Prevention Model

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| | Monthly Peer Reviewed & Refereed Journal |

|| Volume 7, Issue 13, April 2024 ||

International Conference on Intelligent Computing & Information Technology (ICIT-24)

Organized by

Erode Sengunthar Engineering College, Erode, Tamilnadu, India

Process 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, SDC will 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.

V. CONCLUSIONS

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.

REFERENCES

[1].Aho A. V. and Corasick M. J. (1998), 'Efficient String Matching: An Aid to Bibliographic Search', Commun. ACM, vol. 18, no. 6, pp.

[2] Borders K. and Prakash A. (2009), 'Quantifying Information Leaks In OutboundWeb Traffic', in Proc. 30th IEEE Symp. Secur, pp. 129–140.

[3] Borders K. and Weele E.V. (2009), 'Protecting Confidential Data On Personal Computers With Storage Capsules', in Proc.18th USENIX Secur. Symp, pp. 367–382.

[4] Burkhart M. and Strasser M. (2010), 'SEPIA: Privacy-Preserving Aggregation Of Multi-Domain Network Events And Statistics', in Proc.19th USENIX Conf. Secur. Symp., p. 15.

[5] Cai M. and Hwang k. (2005), 'Collaborative Internet Worm Containment', IEEE Security Privacy, vol. 3, no. 3, pp. 25–33.

[6] Croft J. and Caesar M. (2011), 'Towards Practical Avoidance Of Information Leakage In Enterprise Networks', in Proc. 6th USENIX Conf. Hot Topics Secur.(HotSec), p. 7.

[7] Geravand S. and Ahmadi M. (2013), 'Bloom Filter Applications In Network Security: A State-Of-The-Art Survey', Comput. Netw., vol. 57, no. 18,pp. 4047–4064.

[8] Jagannathan G. and Wright R. N. (2005), 'Privacy-Preserving Distributed K-means Clustering Over Arbitrarily Partitioned Data', in Proc. 11th Int. Conf. Knowl. Discovery Data Mining, pp.

[9] Jung J. and Sheth A. and Greenstein B. (2008), 'Privacy oracle: A System For Finding Application Leaks With Black Box Differential Testing', in Proc. 15th ACM Conf. Comput.Commun. Secur., pp. 279–288.

[10] Kleinberg J. and Papadimitriou C. H. (2001), 'On The Value Of Private Information', in Proc. 8th Conf. Theoretical Aspects Rationality Knowl., pp. 249–257.

[11] Karjoth G. and Schunter M. (2002), 'A Privacy Policy Model For Enterprises',in Proc. 15th IEEE Comput. Secur. Found. Workshop, Jun., pp. 271–281.

[12] Li K. and Zhong Z. (2009), 'Privacy-Aware Collaborative Spam Filtering," IEEE Trans. Parallel Distrib. Syst., vol. 20, no. 5, pp. 725–739.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com