Volume 3, Issue 2, February-2016, pp. 57-61 ISSN (O): 2349-7084

International Journal of Computer Engineering In Research Trends

Available online at: www.ijcert.org

Detection and Avoidance of Sensitive Data in Host-assisted Mechanism using Fuzzy Fingerprint Technique

[1]Ms. Patil Deepali E., [2]Prof. Takmare Sachin B.

[1]Pursuing M.E, CSE Branch, Dept of CSE

[2]Assistant Professor Department of Computer Science and Engineering,

Bharati Vidyapeeth's College of Engineering, Kolhapur.

Maharashtra, India.

Abstract— The Data-leak cases, human mistakes are one of the causes of data loss. Deliberately planned attacks, inadvertent and human mistakes lead to most of the data-leak incidents. The detecting solutions of inadvertent sensitive data leaks caused by human mistakes and provide alerts for organizations. A common approach is to screen content in the storage and transmission for exposed sensitive information. Such an approach requires the detection operation to be conducted in secrecy. The data-leak detection (DLD) privacy- preserving solution to solve the special set of sensitive data digests is used in detection. The advantage of data owner is safely delegate the detection operation to a semihonest provider without revealing sensitive data to the provider. Internet service providers can offer their customers DLD as an add-on service with strong privacy guarantees. Evaluation results support accurate detection with very small number of false alarms under various data-leak scenarios. Host-assisted mechanism for the complete data-leak detection for large-scale organizations. To design the Host-assisted mechanism for DLD, using data signature and fuzzy fingerprint.

Keywords— Data leak, network security, privacy, fuzzy fingerprint, data-leak detection.

1. INTRODUCTION

The Risk Based Security (RBS) number of leaked the sensitive data records has increased dramatically during last few years. The detecting and preventing data leaks requires the set of complementary solutions, which may include data-leak detection data confinement stealthy malware detection and policy enforcement. The network data-leak detection (DLD) performs deep packet inspection (DPI).These searches for any occurrences of the sensitive data patterns. The technique of DPI is analyzed to payloads of IP/TCP packets for inspecting application layer data. Alerts are triggered when the amount of sensitive data found in traffic passes a threshold.

The straightforward realizations of data-leak detection require the plaintext sensitive data. This undesirable requirement, threaten it may the confidentiality of the sensitive information. The data owner may need to outsource the data-leak detection to providers, the plaintext sensitive data to them. One needs new data-leak detection solutions those providers to scan the content for leaks without learning sensitive information. The designs, implement, and evaluate the fuzzy fingerprint technique that enhances data privacy during the data-leakage detection operations. The fast and practical one-way computation this is based on the sensitive data. Using this detection method, the DLD provider, who is modeled as an honest-but-curious adversary, can gain limited knowledge about the sensitive data from either the released digests, or the content being inspected.

These techniques, an Internet service provider (ISP) can perform detection on it's the customers' traffic securely and provide the data-leak detection as an add-on service for its customers. In another scenario, the individuals can mark their own sensitive data and the ask administrator of their local network to detect data leaks for them. The DLD provider computes fingerprints from network traffic and identifies potential leaks in them. To prevent the DLD provider from gathering the exact knowledge about sensitive data, the collection of potential leaks is composed of

noises and real leaks. It is the data owner, who post-processes the potential leaks sent back by the DLD provider and determines whether there is any the real data leak. This model supports detection operation delegation. The ISPs can provide data-leak detection as an add-on the service to their customers using this model. The design, implements, and evaluates an efficient technique the fuzzy fingerprint, for privacy-preserving data-leak detection.

Fuzzy fingerprints are special sensitive data digests prepared by data owner for release to the DLD provider. These results indicate high accuracy achieved by this underlying scheme with very low false positive rate. The filtering steps and data preparation can take considerable amount of processing time but once preprocessing is done the data become more reliable and robust results are achieved. They have conducted extensive experiments to validate the accuracy, efficiency and privacy of these solutions. The result provide by host log detect the sensitive data leak detection. The host-assisted mechanism for data-leak detection the complete for large-scale organizations. To design the Host-assisted mechanism for DLD, using data signature and fuzzy fingerprint.

2. LITERATURE REVIEW

Xiaokui Shu, Danfeng Yao and Elisa Bertino, Fellow (2015) [1] has studied that among multiple data-leak cases, human mistakes are one of the main causes of data loss. Detecting inadvertent sensitive data leaks caused by the human mistakes and to provide alerts for organizations. They present privacy- preserving data-leak detection (DLD) solution to solve the where a special set of the sensitive data digests is used in detection. The advantage of method is that it enables the data owner delegate to safely the detection operation is to a semihonest provider without revealing the sensitive data to the provider. The internet service providers can offer their customers DLD as add-on service with the strong privacy guarantees. The evaluation results show that method can support accurate the detection with very small number of false alarms under various data-leak scenarios.

X. Shu and D. Yao (2012) [2] the focus on the latter kind of services, where location information is essentially used to determine the membership of one or more geographic sets. This address problem using Bloom Filters (BF), a compact data structure for representing sets. In particular present an extension of the original Bloom filter idea: the Spatial Bloom Filter (SBF). The SBF's are designed to manage the spatial, geographical information in a space efficient way, and are well-suited for enabling the privacy in location-aware applications. This show by providing two multi-party protocols for the privacy-preserving computation of location information, based on the known homomorphic properties of public key encryption schemes.

K. Borders and A. Prakash (2009) [3] routes of information leakage are various, for example, human, paper, the Internet, and USB flash memory. It is difficult to find information leakage by calculating the number of characters of HTTP requests in cases where the leaked number of characters is not large. If calculate the approximate entropy, the value is small on the whole because ignore a lot of repeated information.

H. Yin, D. Song, M. Egele, C. Kruegel, and E. Kirda (2007) [4] malware has become a significant, complex, and widespread problem within the computer industry. The classification model is based on an examination of eight the malware samples and it identifies four malware commonalities and classifications based on dimensions of the persistence and stealth. The article goal is to provide a better understanding of when the cyber-conflict will happen and to help defenders better mitigate the potential damage.

K. Borders, E. V. Weele, B. Lau, and A. Prakash (2009) [5] practical and powerful device based isolation approach for the information security and application of demonstrate in preserving the confidentiality of cryptographic keys. The device-based isolation is defined by isolating the storage and operations related to data with different security requirements through computing multiple devices. The isolation should not hinder the use and access of the data for practical applications.

A. Nadkarni and W. Enck (2013) [6] the exposure of sensitive data in storage and transmission poses a serious threat to the organizational and personal security. Auto-FBI guarantees the secure access of sensitive data on the web. It achieves this guarantee by automatically generating a new browser instance for sensitive content. Aquifer is a policy framework and system. It helps prevent accidental information disclosure in OS.

G. Karjoth and M. Schunter (2002) [7] privacy policy specification and enforcement has become a hotbed of the research activity over past few years as Internet use has been on the rise around the globe. The number of consumers participating in grows online activities; it

becomes increasingly imperative for the organizations to express their privacy practices in an accurate, accessible, and useful way. The quality criteria used in the software requirements specification can be used to evaluate the privacy policies specified using P3P and EPAL.

Y. Jang, S. P. Chung, B. D. Payne, and W. Lee (2014) [8] have proposed a way to capture richer semantics of the user's intent. The method is based on the observation that for the most text-based applications, user's intent will be displayed the entirely on screen, text, and the user will make modifications. Based on this idea, they have implemented of prototype called Gyrus2 which enforces correct behavior the applications by capturing user intent. Since this is attack agnostic, it will scale better than the traditional security systems.

A. Broder and M. Mitzenmacher (2004) [9] have described the mathematics behind Bloom filters, their history, and some important variations. Bloomfilter is a simple space-efficient randomized the data structure for representing a set in order to support the membership queries. The Bloom filters allow false positives but space savings often outweigh this drawback when the probability of an error is made sufficiently low. This ways in which Bloom filters have been used and modified the variety of network problems, with aim of providing a unified mathematical and practical framework for them and stimulating their use in future applications.

R. Chen, B. C. M. Fung, N. Mohammed, B. C. Desai, and K. Wang (2013) [10] data in its original form, however, the typically contains sensitive information about individuals, and publishing such data will violate individual privacy. The data publishing relies in current practice mainly on policies and guidelines as to what types of the data can be published and on agreements on the use of the published data. Privacy-preserving data publishing (PPDP) provides methods and the tools for publishing useful information while preserving data privacy. This systematically summarize and different approaches to evaluate PPDP, study the challenges in practical data publishing, clarify the differences and requirements that distinguish of PPDP from other related problems, and propose future research directions.

3. ARCHITECTURE DESIGN

1. Host-assisted mechanism: Host-assisted mechanism for the complete data-leak detection for large-scale organizations. The data owner computes the special set of digests or fingerprints from the sensitive data and then discloses only a small amount of them to the DLD provider. This implement detection system and perform extensive experimental evaluation on 2.6 GB Enron dataset, Internet surfing traffic of 20 users, and also 5 simulated real-world data-leak scenarios to measure its privacy guarantee, efficiency and detection rate.

2. Data Preprocessing: Sentiment or Emotion analysis of social networking data involves a lot of data preprocessing. The data preparation and filtering steps can take considerable amount of processing time but once preprocessing is done the data become reliable and robust results are achieved. Data preprocessing is done to eliminate the incomplete, noisy and inconsistent data.

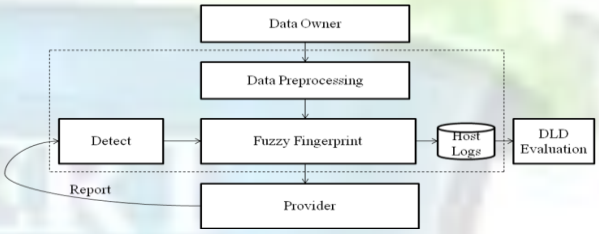

Fig.1 Architecture of Host-assisted Mechanism for Data-leak Detection model

3. Fuzzy Fingerprint: Fuzzy fingerprint technique that enhances data privacy during data-leak detection operations. Fuzzy fingerprints are special sensitive data digests prepared by the data owner for release to the DLD provider. Fuzzy fingerprint mechanism improves the data protection against semi-honest DLD provider.

4. Provider: This describes how Internet service is providers can offer their customers DLD as an add-on service with strong privacy guarantees. Using Fuzzy fingerprint techniques, an Internet service provider (ISP) can perform detection on its customers' traffic securely and provide the data-leak detection as an add-on service for its customers. The gateway dumps the network traffic and sends to a DLD server/provider (Linux).

5. Detect: A common approach is to screen content in storage and transmission for exposed the sensitive information. The detection operation conducted in secrecy. The detection system it can be deployed on a router or integrated into existing network intrusion detection systems (NIDS).

6. Host Logs: This describe the server check out by logs and decide to which model to send in a system.

7. DLD Evaluation: The evaluation results show support accurate detection with very small number of false alarms under the various data-leak scenarios. Rabin fingerprints use with variety of modulus's in fingerprint filter as the hash functions, and perform extensive experimental evaluation on fingerprint filter and bloom filter with MD5/SHA.

The privacy-preserving data-leak detection to supports practical data-leak detection as a service and minimizes the knowledge that a DLD provider may gain the process. Data owner to send the digests to the DLD provider, MONITOR,DETECT for the DLD provider is collect to outgoing traffic of the organization this compute digests of traffic content, identify potential leaks, and REPORT for the DLD provider to return data-leak alerts to the data owner where there may be false positives. Fingerprint Filter develop extension to use Bloom filter in the DETECT operation for efficient set intersection test. Bloom filter is a well-known space-saving data structure for performing set-membership test. The multiple hash functions apply to each of the set elements and stores the resulting values in a bit vector; to test whether a value v belongs to the set, filter checks each corresponding the bit mapped with each hash function.

Rabin fingerprints use variety of the modulus's in fingerprint filter as the hash functions, and perform extensive experimental evaluation on fingerprint filter and bloom filter with MD5/SHA. The analyze security and privacy guarantees provided by this data-leak detection system, as discuss the sources of possible false negatives data leak cases being overlooked and false positives legitimate traffic misclassified as data leak in the detection. Fuzzy fingerprint approach is more flexible from deployment the perspective, as the data owner can adjust and fine-tune the privacy and accuracy in the detection without recomputing the fingerprints. They have conducted extensive experiments to validate the accuracy, efficiency, and privacy of these solutions.

4. CONCLUSION

Fuzzy fingerprint is a privacy-preserving data-leak detection model and present its realization. Using special digests, the exposure of sensitive data is kept to a minimum during the detection. They have conducted extensive experiments to validate the accuracy, privacy, and efficiency of these solutions. Designs a host-assisted mechanism for complete the data-leak detection for large-scale organizations.

ACKNOWLEDGEMENT

The authors are grateful to express the sincere thanks and gratitude to Computer Department of Engineering, BVCOEK for the encouragement and facilities that were offered to us for carrying out this project. The authors would like to thank Prof. Chougule A. B. (Head of computer science and Engineering (BVCOEK)) & Prof. S. B. Takmare.

REFERENCES

[1] Xiaokui Shu, Danfeng Yao and Elisa Bertino, Fellow,"Privacy-Preserving Detection of Sensitive Data Exposure" IEEE Trans. on Information Forensics and Security, vol. 10, no. 5, May 2015,pp.1092-1103.

[2] X. Shu and D. Yao, "Data leak detection as a service," in Proc. 8th Int. Conf. Secur. Privacy Commun. Netw., 2012, pp. 222–240.

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

[4] 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.

[5] 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. 6. 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.

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

[8] Y. Jang, S. P. Chung, B. D. Payne, and W. Lee, "Gyrus: A framework for user-intent monitoring of text-based networked applications," in Proc. 23rd USENIX Secur. Symp., 2014, pp. 79–93.

[9] A. Broder and M. Mitzenmacher, "Network applications of bloom filters: A survey,"Internet Math., vol. 1, no. 4, pp. 485–509, 2004.

[10] R. Chen, B. C. M. Fung, N. Mohammed, B. C. Desai, and K. Wang, "Privacy-preserving trajectory

data publishing by local suppression," Inf. Sci., vol. 231, pp. 83–97, May 2013.

AUTHOR PROFILE

Ms. Patil Deepali E. is a M.E. student in Bharati Vidyapeeth's College of Engineering, Kolhapur. Maharashtra, India. Her research interest lies in Networking and Network security. She has published paper in National Level Conference.

Mr. Sachin Balawant Takmare is working as assistant professor in Computer Science and Engineering department of Bharati Vidyapeeth's College of Engineering, Kolhapur with Teaching experience of about 10 years. He has published about 3 International Papers and 5 National Papers.

