ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 11, March 2016

DETECTION AND PREVENTION OF DATA EXFILTRATION IN LARGE SCALE ORGANISATIONS

Alamelu.V
PG Scholar
Department of Computer Science and Engineering
M.Kumarasamy college of Engineering
Thalavapalayam, Karur
alamuaccet@gmail.com

Prem Kumar.K
Assisstant Professor
Department of Computer Science and Engineering
M.Kumarasamy college of Engineering
Thalavapalayam, Karur.
premkumark.cse@mkce.ac.in

ABSTRACT

In recent years, survey from security firms, investigate and government organizations shows that number of information leaks have grown rapidly. Among various information leak cases, one of the major cause for information leak are human mistake. There are solutions to detect unintentional sensitive information leaks caused by human mistakes and provide alerts for organizations. One such method is privacy preserving Data Leak Detection(DLD) solution where a particular set of responsive information digest is used in detection of data leaks. Privacy preserving DLD representation intended for preventing inadvertent information leak in system traffic. Such a representation yields a powerful and delegatable data leak detection framework. However this method is challenging to satisfy in practice, as DLD may be compromised. To overcome this, an additional security is provided by encryption. In cloud computing environment the cloud source can perform data leak detection as an add-on service to its clients. The advantage of this method is that it enables the information owner to safely hand over the detection operation without revealing the sensitive information to the provider.

I. INTRODUCTION

Detecting and preventing information leaks requires a set of corresponding solutions, which may include data-leak detection, information confinement, stealthy malware detection and policy enforcement. Network data-leak detection (DLD) typically search for some occurrence of responsive information patterns and performs deep packet inspection (DPI). DPI is a technique to analyze payloads of TCP/IP packet for inspecting application layer information, e.g., HTTP header/content. Alerts are triggered and traffic passes a threshold when the amount of sensitive information found. There are two types of input sequences in data-leak detection model: sensitive information sequences and content sequences. Sensitive information contain the sensitive information that cannot be exposed

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 11, March 2016

to unauthorized parties, e.g., proprietary documents, customers records. Sensitive information can also be partitioned to small sensitive information sequence. Content is the information to be inspected occurrences of sensitive information patterns. The detection need to divide the unique content stream into content segment. Christo Ananth et al. [7] proposed a secure hash message authentication code. A secure hash message authentication code to avoid certificate revocation list checking is proposed for vehicular ad hoc networks (VANETs). The group signature scheme is widely used in VANETs for secure communication, the existing systems based on group signature scheme provides verification delay in certificate revocation list checking. In order to overcome this delay this paper uses a Hash message authentication code (HMAC). It is used to avoid time consuming CRL checking and it also ensures the integrity of messages. The Hash message authentication code and digital signature algorithm are used to make it more secure . In this scheme the group private keys are distributed by the roadside units (RSUs) and it also manages the vehicles in a localized manner. Finally, cooperative message authentication is used among entities, in which each vehicle only needs to verify a small number of messages, thus greatly alleviating the authentication burden. The DLD source computes fingerprints from network traffic and identifies potential leaks. To stop the DLD supplier from assembly exact knowledge about the sensitive information and the collection of potential leaks is composed of noises and real leaks. The information owner who post-processes the potential leaks sent back by the DLD source and then determines whether there is any real information leak.

II.RELATED WORK

Stealthy malware detection and monitoring Through VMM-based "out of the box" semantic View reconstruction in this system is proposed a VMM-based approach that enables "out of the box" malware detection and monitoring by addressing the semantic gap challenge. More specifically, VMwatcher achieves stronger tamper-resistance by moving anti-malware facilities out of the monitored VM while maintaining the native semantic view of the VM via external semantic view reconstruction. Our evaluation of the VMwatcher prototype demonstrates its practicality and effectiveness. In particular, our experiments with real-world stealthy root kits and worms further demonstrate the power of the new malware detection and monitoring capabilities enabled by VMwatcher.

Privacy preserving similarity detection for information analysis Encryption in this system is proposed a method for privacy-preserving document similarity detection. It should identify either semantically or syntactically similar documents. As the result two methods were developed. Both of them have the following structure. At first, the areas the documents are related to be found. Then documents are transformed into the set of distinct meaningful words. These sets as well as

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 11, March 2016

documents subject areas are compared in a secure way. In the first method the original privacy-preserving information comparison protocol was used for secure comparison. In the second method the modified private-matching scheme was used for same purpose. Based on the comparison results the type of similarity between documents is identified. Both of the methods provide privacy protection for the documents content of the parties.

Information Leak Detection as a Service: Challenges and Solutions in this system proposed a novel fuzzy fingerprint framework and algorithms to realize privacy-preserving information-leak detection. Using particular digests, the exposure of the sensitive information is kept to a smallest amount during the discovery. Described its application in the cloud computing environments, where the cloud source naturally serves as the DLD source. Defined our privacy goal by quantifying and restricting the probability that the DLD source identifies the exact value of the sensitive information. Presented the protocols and information structures including a Bloom-filter based fuzzy fingerprint filter. Our extensive experiments validate the accurateness, privacy, and efficiency of our solutions**.**

Quantifying Information Leaks in Outbound Web Traffic In this system introduce a novel approach for quantifying information leaks in web traffic. Instead of inspecting a message's information, the goal was to quantify its information content. The algorithms in this paper achieve precise results by discounting fields that are repeated or constrained by the protocol. This work focuses on web traffic, but similar principles can apply to other protocols. examination engine process static fields in HTTP, HTML, and JavaScript to create a distribution of expected request content. It also executes dynamic scripts in an emulated browser environment to obtain complex request values. Evaluate our analysis techniques on controlled test cases and on real web traffic from 10 users over a 30-day period. For the controlled tests, the measurement techniques yielded byte counts that ranged from 0.32%-1.12% of the raw message size. These tests highlighted some limitations of our approach, such as being unable to filter parts of URLs that contain random numbers to prevent caching. For the real web traffic evaluation, the precise unconstrained byte counts averaged 1.48% of the corresponding raw values. This was significantly better than a generic compression algorithm, which averaged 9.87% of the raw size for each request.

A Privacy Policy Model for Enterprises in this paper proposed a formal model for approval organization and access control in privacy protecting system. Taking a systems view of privacy, Elaborate on technical mechanisms to make sure that personal information is used only for authorized purposes. Our model is capable of precisely

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 11, March 2016

capturing the meanings of a wide variety of such privacy policies. Then define a privacy language whose semantics is defined with reference to the model. This ensures that each privacy policy has a clear and unambiguous interpretation that is defined without reference to some particular implementation of a privacy protecting system.

III.PROPOSED SYSTEM

Two important roles in proposed model

- Organization owns the sensitive information and it authorizes the DLD provider to inspect the network for anomalies, namely inadvertent data leak.
- DLD provider inspects the network traffic potential data leaks. However, DLD provider attempt to gain knowledge about the sensitive data. Inspection can be performed offline to avoid the time delay in routing the packets.

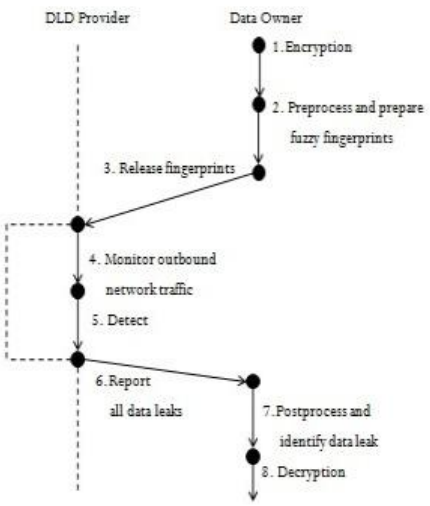

Fig 1: Privacy-preserving DLD model

Our privacy-preserving data-leak detection method supports DLD as a service and minimizes the information that a DLD provider can gain during the process. Fig 1 lists eight operations executed by data owner and DLD provider in our protocol. They include ENCRYPTION which encrypts the data to be transferred, PREPROCESS run by data owner to prepare the digest of sensitive data, RELEASE for the data owner to send the digest to DLD provider, MONITOR and DETECT for the DLD provider to collect outgoing traffic of the organization, compute digest of traffic content, and identify potential data leaks, REPORT for identify the true data leak instances, DECRYPTION of leaked instances to assure the leak of sensitive data.

Encryption and Decryption

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 11, March 2016

Elliptic Curve Cryptography (ECC) is to transplant original encryption algorithm to elliptic curve. It not only realizes key exchange protocol and public key encryption and decryption, but also realizes digital signature. The essence of ECC is to transplant classical encryption algorithm to secure elliptic curve for its implementation. For example, in RSA, they are common modular addition, modular multiplication previously, now after it is transplanted to elliptic curve, they become point addition and point multiplication, which are also called elliptic addition, elliptic multiplication.

Shingles and Fingerprint

DLD provider receives the digest of sensitive data from data owner. The data owner uses sliding window and Rabin fingerprint algorithm to generate small and hard-to-reverse digest. The sliding window generates small fragments of encrypted and processed data which preserves the local features of data and provide noise tolerance property. Rabin fingerprint algorithm has a unique min-wise independent property, which supports fast random fingerprint selection for partial fingerprint disclosure.

A shingle is a fixed size sequence of contiguous bytes. The use of shingle alone does not satisfy the hard-to-reverse requirement. Rabin fingerprint is required to satisfy such requirement after shingling.

IV.CONCLUSION

We proposed a privacy preserving DLD provider that uses a special digest to detect the leak of sensitive data. Using this special digest and encryption, the exposure of sensitive data is minimum during the detection.

V.REFERENCES

[1] Stealthy Malware Detection through VMM-Based "Out-of-the-Box" Semantic View Reconstruction Xuxian Jiang, Xinyuan Wang Dongyan Xu

[2] Privacy preserving similarity detection for information analysis Iraklis Leontiadis, Melek Onen, Refik Molva ¨ Networking and Security Department EURECOM, Sophia-Antipolis,France {leontiad,onen,molva}@eurecom.fr M.J. Chorley, G.B. Colomb

[3] Information Leak Detection As a Service: Challenges and Solutions Xiaokui Shu Danfeng (Daphne) Yao Department of Computer Science, Virginia Tech

[4] A Privacy Policy Model for Enterprises Gunter Karjoth and Matthias Schunter ¨ IBM Research

[5] Quantifying information leaks in outbound web traffic K Borders, A Prakash - Security and Privacy, 2009 30th IEEE 2009 - ieeexplore.ieee.org

[6] 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

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 11, March 2016

Conf. Comput. Commun. Secur., 2007, pp. 116–127.

[7] Christo Ananth, M.Danya Priyadharshini, "A Secure Hash Message Authentication Code to avoid Certificate Revocation list Checking in Vehicular Adhoc networks", International Journal of Applied Engineering Research (IJAER), Volume 10, Special Issue 2, 2015,(1250-1254)

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

[9] 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