A Survey-Various Methods Of Data Leak Detection in Network

Athira G
P G Scholar
Department of Information Technology
Govt.Engineering college
Bartonhill, Trivandrum

Sreedivya R S
Assistant Professor
Department of Information Technology
Govt.Engineering college
Bartonhill, Trivandrum

Abstract

Data leakage is the unauthorized transmission of sensitive data or information . ie, Data leakage is de-fined as the accidental or intentional distribution of sen-sitive data to an unauthorized entity. Sensitive data of companies and organization includes intellectual prop-erty, employee information, financial information, pa-tient information, personal credit card data etc. The no of data leakage incidents are growing day by day. Many researchers and government organizations show that the number of data-leak instances have grown rapidly in recent years. Privacy Preserving of sensitive data leakage has became the most important issue in todays world. So, the importance of detection of data leak also increases.

KEYWORDS— Data leak, Network Security, Privacy, Fake Object

1. INTRODUCTION

A data leakage or data breach may be the inten-tional or unintentional release of secure information in an unsecure environment. According to various reports from Risk Based Security, the number of leaked sensi-tive data records have increased dramatically. Attacks, inadvertent leaks, and human mistakes lead to most of the data-leakage incidents. Detecting and preventing data leaks requires total solutions, which may include data-leak detection. The leakage of important data, whether it is accidental or intentional, huge loss will happen to the data owner. It is very hard for any sys-tem administrator to trace out the data leaker among the system users[1]. Typical approaches to preventing data leak are under two categories host-based solutions and network-based solutions. Host-based approaches may include encrypting data when not used and enforcing policies to restrict the transfer of sensitive data[5].

Data leakage poses a serious issue for companies as the number of leak incidents increases and the cost to maintain the original data increases. Mainly data leak-age occurs through e-mails, instant messaging, website forums, and file transfers among others, which are un-regulated and un-monitored [3]. In many cases organizational data is shared among customers, various stake holders, employees working in other organizations etc. This increases the chance to get sensitive information or important data falling into unauthorized parties. [6] discussed about a method, Optimality results are presented for an end-to-end inference approach to correct(i.e., diagnose and repair) probabilistic network faults at minimum expected cost. One motivating application of using this end-to-end inference approach is an externally managed overlay network, where we cannot directly access and monitor nodes that are independently operated by different administrative domains, but instead we must infer failures via end to-end measurements.

2. VARIOUS DATA LEAK DETECTION METHODS

2.1. Perturbation Technique

The Perturbation is most traditional and available technique where the data are modified and made less sensitive before sending for communication. We can add random noise to certain attributes, or can replace exact values by ranges[2]. However, in some cases, it is

athirag1893@gmail.com

important not to alter the original data owners data. For example, if an outsourcer is doing our payroll, he must have the exact salary and customer bank account num-bers. In medical cases, doctors or medical researchers need accurate data for treating patients.

2.2. Water Marking Technique

Water Marking is the traditional technique used for data leak detection. e.g., a unique code is embedded in each copy of the data which is distributed. If that copy is later found in the hands of an unauthorized person, the leaker or spy can be identified. Watermarks could be very useful in some cases[4]. If we consider some Real life example, then watermarking can also be used for compressed video data. Encryption and watermarking are two techniques used to provide copy protection and copyright protection for digital video and audio data. Encryption techniques can be used to protect digital data during the transmission from the sender to the re-ceiver. But, if the receiver has decrypted the data, then the actual data is no longer protected. Watermarking techniques can be used as a complement technique for encryption by embedding a secret imperceptible signal called a watermark, directly into the clear data. This wa-termark signal is embedded in such a way that it cannot be removed without affecting the quality of the audio or video data[2]. The watermark signal is also useful for copyright protection as this signal can hide informa-tion about the author or owner in the data. To trace the source of illegal copies watermark signal can be used by means of fingerprinting techniques. Disadvantage-Watermarks involve some modification of the original data. If the data recipient is malicious, then sometimes watermarks can be destroyed.

2.3. Fake Object Technique

Fake object is generally alteration or change in ac-tual data which apparently improves the probability To find guilty agents. The data owner may be able to add fake objects to the original data in order to improve its effectiveness in detecting guilty agents. The method of modifying data to detect leakage of data is not new. However, in most cases, individual objects are modi-fied, e.g., by adding random noise to sensitive data, or adding a watermark to an image. Considering about real time use of fake objects our study says we can consider trace records one of them. Trace records are basically owned addresses by entities[4]. Let us con-sider two companies A and B, now suppose company A sells a mailing list to company B for saying advertise-ment. At that time company A adds trace records to the mailing list because of which every time when company B uses that mailing list company A receives a copy of it. This therefore can identify unauthorized use of data. Disadvantage- it leads to, alteration of original data. Al-tering original data may not be suitable every time for e.g suppose we have Financial information like budget or employees salary, such sensitive information cannot be altered as any alteration can lead to financial crises of the company.

2.4. Steganography

The technique of hiding information or data by em-bedding messages within other, ie, not harmful mes-sages. Steganography is used when encryption is not allowed. Or, more commonly, steganography is used to supplement encryption. An encrypted file or docu-ment also can hide data using steganography, so even if the encrypted file is decrypted, the hidden message is not seen by any one. Steganography is a tech-nique which embeds sensitive information with other data using some rules and methods because of which we can identify authorized person or entity[4]. Appar-ently watermarking technique is mostly widely used in steganography. When come to watermarking, it has two types: visible and invisible watermarking. Visible wa-termarking is one which is visible, it is a text or logo which depicts the authorization. For example the logo of famous company like BMW, Audi, Mercedes very apparently indicates their particular cars. These logo can even be for advertisement and promotion of these firm. Coming to invisible watermarks, they can be em-bedded in audio, video, images as well as text. The remarkable aspect about invisible marking is that they appears to be the original object and besides this tech-nique can be used for copyright prevention which au-thorizes authors, creators, writers etc. There is list of Real time application of steganography like in the busi-ness world it can used to hide secret chemical or plans for new inventions. Even terrorist can use steganogra-phy for secret communication and attack plans. Tradi-tionally steganography was in used while making maps where the cartographers adds tiny fictional street to pre-vent the map from copycats. Disadvantage: Message is hard to recover if the image is subjected to attack such translation and rotation besides it relative easy to detect.

2.5. Data Leak Detection Provider

In this technique there will be a data leak detection provider. There will be a data owner who owns the data. Data owner computes a special set of digests or finger-prints using fuzzy fingerprint method from the sensitive data. Then discloses only a small amount of them to the DLD provider. The DLD provider computes finger-prints from network traffic and identifies potential leaks in them. To prevent DLD provider from gathering ex-act knowledge about the sensitive data, the collection of potential leak is composed of real leaks and noises. It is the data owner, who post-processes the potential leaks sent back by the DLD provider and determines whether there is any real data leak [7]. This technique focuses on detecting inadvertent data leak. Inadvertent data leak may be due to human errors such as forgetting to use encryption, carelessly forwarding an internal email and attachments to outsiders. Using this technique, an Inter-net Service Provider (ISP) can perform detection on its customers.

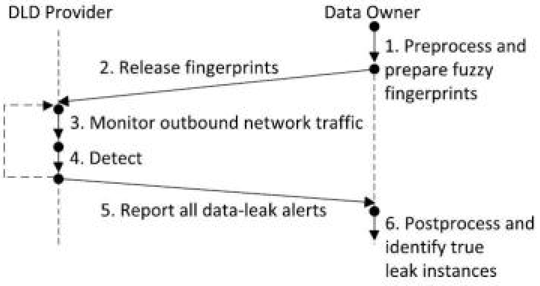
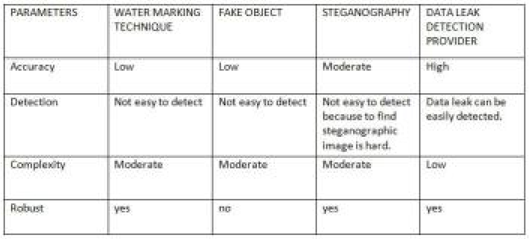

DLD Provider
Data Owner
1. Preprocess and prepare fuzzy fingerprints
2. Release fingerprints
3. Monitor outbound network traffic
4. Detect
5. Report all data-leak alerts
6. Postprocess and identify true leak instances

Figure 1. Data-Leak Detection Model

3. RESULT AND COMPARISON



PARAMETERS	WATER MARKING TECHNIQUE	FAKE OBJECT	STEGANOGRAPHY	DATA LEAK DETECTION PROVIDER
Accuracy	Low	Low	Moderate	High
Detection	Not easy to detect	Not easy to detect	Not easy to detect because to find steganographic image is hard.	Data leak can be easily detected.
Complexity	Moderate	Moderate	Moderate	Low
Robust	yes	no	yes	yes

Figure 2. Comparison of different Methods

4. CONCLUSION

From this study, we can understand that there are various kinds of data leak detection techniques and can understand that data leak detection and prevention is important. Analyzing different methods we can conclude that a third party involvement technique is better compared to others. ie, Data leak detection provider . In this method data leakage can be effectively determined. The possibility of getting actual data to the data leak de-tection provider is less. An effective fuzzy fingerprint technique is used.

References

[1] Aho, A. V., and Corasick, M. J. Efficient string matching: an aid to bibliographic search. Commun.ACM (1975).

[2] R. Agrawal and J. Kiernan, Watermarking Relational Databases,Proc. 28th Intl Conf. Very Large Data Bases (VLDB 02), VLDB Endowment, pp. 155-166, 2002

[3] Technical Report TR-BGU-2409-2010 24 Sept. 2010 1 A Survey of Data Leakage Detection and Prevention So-lutions P.P (1-5, 24- 25) A. Shabtai, a. Gershman, M. Kopeetsky, y. Elovici Deutsche Telekom Laboratories at Ben-Gurion University, Israel.

[4] Steganography, Cryptography, Watermarking: A Comparative Study Hardikkumar V. Desai (B.Sc., MCA)Research Scholar, Singhania University, Volume 3, No. 12, December 2012 Journal of Global Research in Computer Science.

[5] Sandip A.Kale, Prof. Kulkarni S.V. (Department Of Computer Sci. and Engg,MIT College of Engg), Dr.B.A.M.University, Aurangabad(M.S), India, Data Leakage Detection: A Survey, (IOSR Journal of Computer Engineering (IOSRJCE)ISSN : 2278-0661 Volume 1, Issue 6 (July-Aug 2012), PP 32-35 www.iosrjournals.org.

[6] Christo Ananth, Mona, Kamali, Kausalya, Muthulakshmi, P.Arthy, “Efficient Cost Correction of Faulty Overlay nodes”, International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE), Volume 1,Issue 1, August 2015,pp:26-28

[7] Shu, Xiaokui, Danfeng Yao, and Elisa Bertino. ”Privacy-preserving detection of sensitive data exposure.” Infor-mation Forensics and Security, IEEE Transactions on 10.5 (2015): 1092-1103.