International Journal of Research Instinct
(Peer-reviewed Journal)

ISSN ONLINE 2348-2095
Research Article

PRIVACY-PRESERVING DETECTION OF SENSITIVE DATA EXPOSURE

Mrs.M.Preethi.,

.M.Phil. Research Scholar

Srimad Andavan Arts Science College(Autonomous), Trichy-620005.

maheepree@gmail.com

ABSTRACT

This article "**Privacy-Preserving Detection of Sensitive Data Exposure"** is used to detect the data leakage of sensitive data. Most of the organizations transfer data over the internet. So the administrators are faced with the task of keeping confidential information from leaving their networks. Mostly the data loss is caused mainly by human mistakes. Recently Government organizations show that the numbers of data-leak instances have grown quickly. One of the important issues in the information security research is data leakage or data loss especially caused by insider threat as insider threats have potential to thrust severe damage to the organization's resources, financial assets and reputation. However, privacy and secrecy consideration can prohibit the organizations from willing to share the data from each other and this is one of the major tasks in the information security. In this article we are using the data digest method for reducing data leakage and misuse detection. The data is breaking into packets while transferring over the internet. So the message will rise if the data leakage occurs. Finally the data can be merged and viewed by the owner.

KEYWORDS: Data leakage, Data misuse, Insiders, Network Security, Privacy.

1. INTRODUCTION

Organization's data is very important and proves as a main constituent in embodying the core of the organization's power and this power should be preserved and maintained. On the other side, this data is required for daily working on different processes. Consumers within the

organization such as employees or partners perform different procedures on this data and may be exposed to the important information while accessing the data. Due to this processing and action, it may lead to data leakage and misuse. Detecting and preventing data leaks perform some steps such as data-leak detection [1], data In short, the risk to data security from insider's threat is becoming more and more critical because of the endless use of the computers and also communication systems. Various methods have been proposed for defending data from outer attacks but those mechanisms fail to protect data from authorized users who may misuse their privileges in carrying out malicious activities.

2. LITERATURE SURVEY

Literature survey gives the survey of previously proposed models for data leakage detection. In the recent years, several methods have been proposed to deal with the problem of data leakage and misuse in database systems, especially caused by the insider.

2.1 PREVIOUS WORK

In this section we present analysis of previous related work carried out for the privacy preserving data leakage detection problems with a threat model, a security goal and a privacy goal.

A. Data owner or organization owns the sensitive data handles the DLD provider to inspect the network traffic from the organizational networks for inadvertent data leak. However, the data owner does not want to reveal the sensitive data to the DLD provider directly.

B. 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. However, the DLD provider may attempt to gain knowledge about the sensitive data [1] [2].

C. Security Goal, Threat Model & Privacy Model Case 1-Inadvertent data leakage: The sensitive data is accidentally leaked to the outside world by a unauthorized user. This paper focuses on type of data leaks, the main causes of inadvertent data leak includes human errors such as forgetting to use encryption, carelessly forwarding an internal mail and attachments to outsiders, or due to application flaws.

Case 2-Malicious data leakage: A piece of stealthy software may steal sensitive personal or organizational information from a host because malicious adversary can use strong encryption or stenography to disable content based traffic inspection, thus this type of leaks are out of scope of our network based solution .

Case 3-Legitimate and intended data transfer: The sensitive data is sent by a legitimate user for legitimate purposes. In this paper, we assume that legitimate data transfers use data encryption such as SSL, which allows one to distinguish it from the inadvertent data leak. Therefore, in what follows we assume that plaintext sensitive data appearing in network traffic is only due to inadvertent data leaks [1] [2] [5].The security goal in this paper is to detect Case 1 leaks that are inadvertent data leaks. In other words, we aim to detect sensitive data appearance in traffic (attached email) over supervised network channels. We assume that 1) plaintext data in supervised network channels can be extracted for inspection 2) the data owner is aware of legitimate data transfers 3) whenever sensitive data is found over network traffic, the data owner can decide whether or not it is a data leak The privacy goal in our fuzzy fingerprint mechanism is to prevent the DLD provider from inferring the exact knowledge of the sensitive data; the DLD provider is given the fingerprints of sensitive data and the content of network traffic which may or may not contain data leak. In our model, we aim to hide the sensitive values among other no sensitive values, so that the DLD provider is unable to pinpoint sensitive data among them even under data-leak scenarios.Our privacy goal is defined as follows. The DLD provider is given digests of sensitive data from the data owner and the content of network traffic to be examined. The DLD provider should not find out the exact value of a piece of sensitive data with more than 1/K probability, where K is an integer representing the number of all possible sensitive data candidates that can be inferred by the DLD provider.

2. MOTIVATION

Now a days, the most commonly employed “privacy protection procedure” is to simply remove the explicit identifier of the record holders before releasing the data... Sweeney showed a real-life example of privacy attack on William Weld, who is a former governor of the state of Massachusetts.

This research, privacy-preserving data publishing, is a study of preventing this kind of linking attack. Its goal is to prevent linking some record holder to a specific (or a small number of) data record and sensitive information in the released data while, at the same time, preserving the useful information in the released data. This thesis identifies a collection of privacy threats in various real life data publishing problems, and presents a unified anonymization algorithm for removing these threats. Releasing the data analysis or data mining result [52] such as a classifier, instead of the data, could be an option if the data publisher knows exactly how the data miner may analyze the data. This information, however, often is unknown at the moment of release.

For example, in visual data mining, the data recipient needs to visualize data records in order to produce a classifier that makes sense, and in the k-nearest neighbor classification the data itself is the classifier. In these cases, releasing data records is essential. In other cases, some classifiers are preferred for accuracy, some for precision/recall, some for interpretability, and yet some for certain domain-specific properties. The data publisher (such as a hospital) does not have the expertise to make such decisions for the data recipient (such as biomedical researchers) due to the lack of domain knowledge and sophisticated data mining techniques. Publishing the data provides the recipient a greater flexibility of data analysis.

3. PROPOSED SYSTEM

In this a data-leak detection solution which can be outsourced and be deployed in a semi honest detection environment. In this system, Blowfish algorithm is used to encrypt the data. Blowfish is used to enhance data privacy during data-leak detection operations. Blowfish is one of the fastest block ciphers. Slowness kept Blowfish from being used in some applications. n the detection procedure, the data owner computes a encryption process for sensitive data and then discloses only a small amount of them to the DLD provider.

ADVANTAGES

- The data owner to securely delegate the content-inspection task to DLD providers without exposing the sensitive data.
- Using our techniques, an Internet service provider (ISP) can perform detection on its customers' traffic securely and provide data-leak detection as an add-on service for its customers

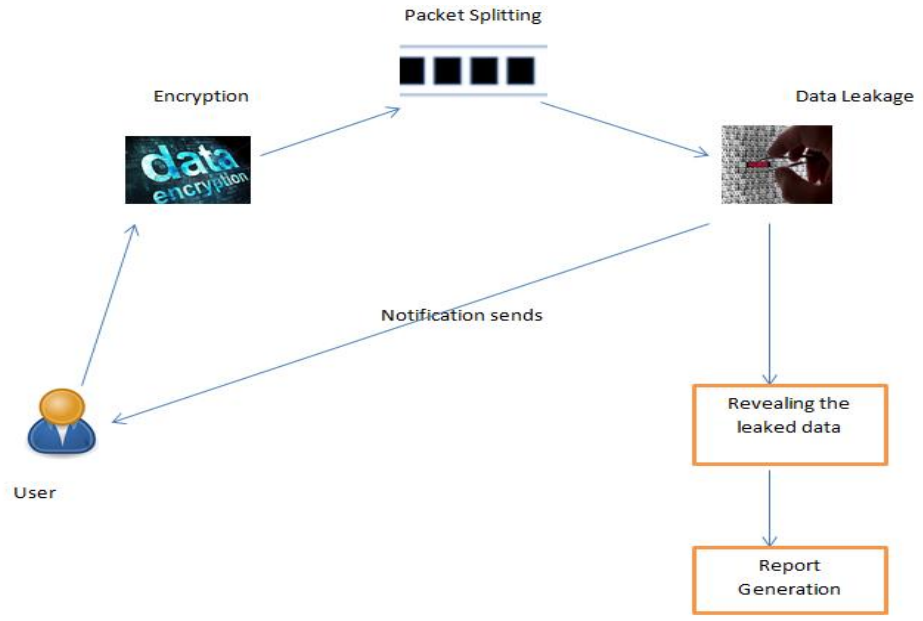
- This approach works well especially in the case where consecutive data blocks are leaked

PROPOSED SYSTEM ALGORITHM

- Blowfish is a variable-length key, 64-bit block cipher. The algorithm consists of two parts: a key-expansion part and a data- encryption part.
- Key expansion converts a key of at most 448 bits into several subkey arrays totaling 4168 bytes.
- Data encryption occurs via a 16-round Feistel network. Each round consists of a key dependent permutation, and a key- and data-dependent substitution.
- All operations are XORs and additions on 32-bit words. The only additional operations are four indexed array data lookups per round.
- Blowfish algorithm with others in terms of the throughput, processing time and power consumption. More the throughput, more the speed of the algorithm & less will be the power consumption.

SYSTEM DESIGN

4. Architecture Diagram



5. ANALYSIS AND DISCUSSION

In this section we present the analysis of security and privacy guarantees provided by considering in the section 2 data-leak detection systems. In the following section we present privacy analysis. We consider static web server in our implementation in order to extract the sensitive data and match the digest in the considerable Poisson distribution network traffic. We identify the limitations associated with proposed approach. A. Privacy Analysis Our privacy goal is to prevent the DLD provider from inferring the exact knowledge of all sensitive data, both the outsourced sensitive data and the matched digests in network traffic. We quantify the probability for the DLD provider to infer the sensitive shingles as follows.

6. CONCLUSION

Privacy-preserving data-leak detection model and present its realization. Using special digests, the exposure of the sensitive data is kept to a minimum during the detection. We have conducted extensive experiments to validate the accuracy, privacy, and efficiency of our solutions. For future work, we plan to focus on designing a host-assisted mechanism for the complete data-leak detection for large-scale organizations. The above results show the superiority of Blowfish algorithm with others in terms of the throughput, processing time and power consumption. More the throughput, more the speed of the algorithm & less will be the power consumption. Secondly, AES has advantage over the other 3DES and DES in terms of throughput & decryption time. Third point is that 3DES has the least performance among all the algorithms mentioned here. Finally we can conclude that Blowfish is the best of all.

7. References

- Jawahar Thakur, NageshKumar,"DES,AESandBlowfish: Symmetric Key Cryptography Algorithms Simulation Based Performance Analysis," in International Journal of Emerging Technology and Advanced Engineering (ISSN 2250-2459, Volume 1, Issue 2, December 2011), pp.6-12
- E. Thambiraj,G. Ramesh,Dr. R. Umarani , "A survey on Various Most Common Encryption Technique " International Journal of Advanced Research in Computer Science and Software Engineering , Volume 2, Issue 7, July 2012,pp226-233

- HimaniAgrawal and Monisha Sharma "Implementation and analysis various symmetric cryptosystems " in indian Journal of Science and Technology in Vol. 3 No.12 (Dec 2010) ISSN: 0974- 846 ,p.1173-1176
- **Data leak detection as a service,"** X. Shu and D. Yao have been proposed that the analysis technique on controlled test cases and on real web traffic from 10 useers over 30 days
- **Panorama Capturing system-wide information flow for malware detection and analysis,"** H. Yin, D. Song, M. Egele, C. Kruegel, and E. Kirda have been proposed that Malware has brought along serious security and privacy threats.