Comparison on Privacy-Preserving Detection of Sensitive Data Exposure
P.Lalitha¹ G.Saranya²
¹Assistant Professor ²MPhil Scholar
¹,²Department of Computer Science
¹,²Hindusthan College of Arts & Science, Coimbatore - 28

Abstract— Recently security firms, Government organizations and other research scholars identifies that data leakage is common in fields where sensitive data are stored and processed. And it is analyzed that data leakage is been growing rapidly, in which human errors are one of the main causes of data loss. The common approach in detecting data leak screens the content in storage and transmission that expose sensitive information. The approach usually requires the detection operation to conduct the secrecy. And there exists lot of techniques, algorithms to find data leakage and provide alerts to the organization. The data leak detection(DTD) privacy preserving solution solves through the special set of sensitive data digest in the detection. The data leak detection systems were proposed to detect data in small scale and large scale organizations with fuzzy fingerprints. In this paper, we review the data leak detection mechanisms.

Key words: Data Leak, Network Security, Privacy, Sensitive Data

I. INTRODUCTION
The Risk Based Security (RBS) [1]reported that number of leak of sensitive data records has increased dramatically during last few years. Deliberately planned attacks, inadvertent leaks and human mistakes like assigning the wrong privilege lead to most of the data-leak incidents. The detecting and preventing data leaks requires the set of complementary solutions, which may include data leak detection [2, 3] data confinement [4, 5, 6] stealthy malware detection [7,8] and policy enforcement [9].

Several algorithms are designed for data security to avoid data leakage. The approaches designed for network based data leak are efficient than the host based detection which analyzes on content[17].And there exists some semi automatic detection system with DLD mechanism. Network data Leak detection is one such method; where it performs deep packet inspection (DPI) over a network channel. The DPI analyzes the TCP/IP packets and inspects the data, when the inspected data found in network then give alerts to the organization.

In this study paper, privacy preserving solutions detecting the data leak with different approaches are analyzed. The network based approach depends on a handy restricted calculation on the touchy information. This empowers the information proprietor to safely appoint the substance investigation undertaking to DLD suppliers without uncovering the touchy information. The host assisted mechanism is used which generates an automatic table containing the sensitive data. Automatic table is generated using statistical approach with highly differentiated values as sensitive and fingerprints. This approach is based on a fast and practical one-way computation on the sensitive data. The host assisted and network based approaches are reviewed.

II. LITERATURE SURVEY
Xiaokui Shu, Danfeng Yao and Elisa Bertino [10] had analyzed that human mistakes are one of the main causes of data loss in organization. They present the privacy-preserving data- leak detection (DLD) solution to solve the problem where a special set of sensitive data digests is used in the detection. The advantage of the method is that it enables the data owner delegate to detect with a semihonest provider without revealing the sensitive data to the provider. The results show that the method can support the accurate detection with very small number of false alarms under various data-leak scenarios. This method is difficult for huge database.

X. Shu and D. Yao [11] focuses on the latter kind of services, in which the location information are used to determine the membership of one or more geographic sets. The problem is addressed with Bloom Filters (BF)[15] which filters allow false positives but space savings often outweigh this drawback when the probability of an error is made sufficiently low which works efficiently geographics sets.They [11] presented an extension Spatial Bloom Filter (SBF) which manages the spatial, geographical information in a space efficient way that suits for location-aware applications based on homomorphic properties with public key encryption schemes.

Prakash [12] describes the routes of information leakage, for example, human, paper, the Internet, and USB flash memory and he analyzed that 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. Malware has become a significant, complex, and widespread problem in the computer industry [13] and classification model is used to detect malware commonalities based on dimensions of the persistence and stealth.

Y. Jang, S. P. Chung, B. D. Payne, and W. Lee [14] have implemented a prototype called Gyrus2 which captures the richer semantics of the user’s intent. It enforces correct behavior of the applications by capturing user intent and based on its attack agnostic, it scales better than the traditional security systems.

R. Chen, B. C. M. Fung, N. Mohammed, B. C. Desai, and K. Wang [16] studies about privacy-preserving data publishing (PPDP) methods that provides tools for publishing information and meanwhile preserving data privacy. The study challenges in practical data publishing and clarify the differences and requirements of PPDP from other related problems, this pays the ways for future research directions.

Rashmi Bhat et al [18] studies about various techniques for data leakage and misuse detection. A system is introduced which measures the risk of damage caused when data is exposed to the insider and it also alter the data in such a way that the risk will be reduced and at the same
time modified data will be useful for performing desired task.

S. Saranya [19] proposed special digests (privacy preserving mechanism) Active data leakage model that disclosure the sensitive data during detection. Experiments validate the accuracy, privacy, and efficiency of the solution and the model can ensure that the data is used in a trusted and controllable environment.

III. PROPOSED SYSTEM

A. Network Based Approach:

Network based 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. Fig. 1 illustrates the six operations between the data owner and the DLD provider in our protocol, which include PREPROCESS run by the data owner to prepare the digests of sensitive data, data owner RELEASE 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 dataowner to pinpoint true data leak instances.

To achieve the goal of privacy the data owner generates a special type of digests, which call fuzzy fingerprints.

The purpose of fuzzy fingerprint is hiding the true sensitive data in a crowd. The fuzzy fingerprint prevents the DLD provider from learning its exact value.

B. Host Assisted Approach:

Fig. 2 shows the host assisted mechanism for DLD model. Host-assisted mechanism provides complete data-leak detection for large-scale organizations. In this mechanism the data owner computes the special set of digests or fingerprints from the sensitive data and it discloses only a small amount of them to the DLD provider. For data preprocessing the initial filtering steps takes considerable amount of processing time which eliminates the complete, noisy and inconsistent data. Fuzzy Fingerprint enhances data privacy during data-leak detection operations and the detection system it can be deployed on a router or integrated into existing network intrusion detection systems (NIDS)[20].

C. Comparison:

<table>
<thead>
<tr>
<th>S. No</th>
<th>Characteristics</th>
<th>Privacy-Preserving Detection of Sensitive Data Exposure</th>
<th>Detection and Avoidance of Sensitive Data in Host-assisted Mechanism using Fuzzy Fingerprint Technique</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Data leakage detection</td>
<td>Data leak detection for small scale organizations</td>
<td>Complete data leak detection for large scale organizations</td>
</tr>
<tr>
<td>2</td>
<td>Detection operation</td>
<td>semi honest provider that does not reveal sensitive data to the provider</td>
<td>semi honest provider without revealing sensitive data to the provider</td>
</tr>
<tr>
<td>3</td>
<td>Network security</td>
<td>SHA/MD5 with fingerprint</td>
<td>Data signature and fuzzy fingerprint.</td>
</tr>
<tr>
<td>4</td>
<td>Mechanism</td>
<td>Network based DLD approach</td>
<td>Host assisted Approach</td>
</tr>
<tr>
<td>5</td>
<td>ISP</td>
<td>It can perform detection on its customers’ traffic securely and provide data-leak detection as an add-on service for its customers</td>
<td>It also offers their customers DLD as an add-on service with strong privacy guarantees</td>
</tr>
</tbody>
</table>

Table 1: Comparison

IV. CONCLUSION

We conclude that the privacy-preserving detection method is used to secure sensitive data from the exposure. The host assisted mechanism provides better performance than network based mechanism and host assisted provide complete data leak detection for large scale organization which is difficult in network based mechanism. So from the comparison host assisted mechanism performs better than network based approach.

REFERENCES

Comparison on Privacy-Preserving Detection of Sensitive Data Exposure

All rights reserved by www.ijsrd.com