www.ijcrt.org © 2017 IJCRT | International Conference Proceeding ICGTETM Dec 2017 | ISSN: 2320-2882
IJCRT Publish Paper record is available at DOI: http://doi.one/10.1727/IJCRT.17185

The Study and Review of Detection of Sensitive Data Leakage for Privacy Preserving

Rucha J Patil, Yogeshwari S. Borse
Department of Computer Engineering, SSBT's College Of Engineering and Technology, Bambhori, Jalgaon [M.S],India
Department of Computer Engineering, SSBT's College Of Engineering and Technology, Bambhori, Jalgaon [M.S],India

Abstract - **According to Risk Base Security (RBS),leakage of sensitive data record instance has grown now a days. Human mistakes plays an important role in cause of data loss among various data leak. There are various method to detect the data leak cause by human mistakes and prevent the data by generating an alert . Among various approaches, monitoring the data which is transmit for expose of sensitive information is common. Also it consider all data as sensitive and perform detection operation for all those data. However this makes the detection process difficult and detection time to increase. In addition, the data owner may require to provide detection report to the DLD provider . But there is possibility that the provider can read the sensitive data. So there is a need of new data detection solution that allow provider to scan the content for leak without learning information. Therefore one need methods that gives accurate detection with very small number of false alarm under various leak scenario and result shows that the method improve the detection time .**

Keywords - Data Leak, Network Security, Privacy, Fingerprint.

I. INTRODUCTION

Today's era, most of the leaked sensitive data record has increase dramatically. Data leakage means unauthorized transmission of sensitive data or information from within an organization to an external destination where the confidentiality of information is compromise. A common approach is to monitor the data in storage and transmission for expose sensitive information. Also it consider all data as sensitive and perform detection operation for all those data. However this makes the detection process difficult and detection time to increase. In addition, the data owner may require to provide detection report to the DLD provider . But there is possibility that the provider can read the sensitive data. In order to minimize the leakage of the sensitive data, organization needs to prevent cleartext sensitive data from appearing in the storage. A screening tool is use to scan the files. Therefore one need a new data detection solution that allow provider to scan the content for leak without learning information. Therefore one need methods that gives accurate detection with very small number of false alarm under various leak scenario.

Human mistakes plays an important role in cause of data loss among various data leak.. There are various method to detect the data leak cause by human mistakes and prevent the data by generating an alert . Among various approaches, monitoring the data which is transmit for expose of sensitive information is common. Also it consider all data as sensitive and perform detection operation for all those data. However this makes the detection process difficult and detection time to increase .So there is a need of new data detection solution that allow provider to scan the content for leak without learning information. Therefore one need methods that gives accurate detection with very small number of false alarm under various leak scenario and result shows that the method improve the detection time .

In order improve the detection time and detection of sensitive data packet , host assisted mechanism is used which checks the frequency of occurrence of data. Highly differentiated values are considered as sensitive and fingerprints are generated for them. Repeated values are ignored in this method. Statistical approach is use to generate sensitive data and it is stored in table. The fingerprints are generated by data leak detection (DLD) provider and identifies potential leaks by matching the fingerprints. The potential leak consist of real leaks and noises so that no one can get exact information about the sensitive data. Then data owner post process data send by the DLD provider to check where there is any leak in the sensitive data The objectives are to improve the detection time and to improve detection of sensitive packet.

II . RELATED WORK

G. Karjoth et al. in [4], privacy policy specification is hotbed of the research as use of internet is increased in recent years.The number of user participating in online activity is increased. P3P and EPAL is use to represent the privacy policies specified in quality criteria of software requirement specification.

A. Broder et al. in [2] ,have proposed the Bloom Filter. Bloomfilter is data structure which is space efficient and used for generating a set in order to support the membership queries .Bloom filters allow false positives but space savings often have more weight than specified.

H. Yin, D. Song et al. in [3], Propose a system Panorama as malware is increased in recent years. Malware is detected by capturing fundamental trait. In the experiment, Panorama successfully detected all malware sample but had a few false rate for analyzing unknown code sample.

K. Borders et al. in [5],introduce Storages Capsule. It is used to protect confidential file on personal computer. Storage Capsule use cryptographic key to encrypted file. so that it will prevent the confidentiality of the data. Checkpoint of the current system state are use to keep the track also disabling device output is use to achieve the goal before allowing access to storage capsule But it do not rely on high integrity.

X. Shu et al.in [8], introduce a network based data leak detection technique. Data owner does not play an important role in this technique as it uses digest to identify the sensitive data. To measure the privacy for fingerprint framework they provide a quantifiable method. But not efficient enough for practical data leak inspection in the setting.

A. Nadkarni and W. Enck in [6], propose aquifer for preventing accidental information exposure in modern operating system framework and system. In aquifer, the entire user interface workflow is protectd using secrecy restriction that define by the application developers . But it has lack of application seperation .

Y. Jang et al. in [7], have proposed a way to capture the system behavior that matches with richer semantics of the users intent. In this method text-based applications is used for observation. Based on this idea, they have implemented of prototype called s Gyrus2.It will capture the user intent . but it will not capture the time of event generated by user.

Xiaokui Shu et al. in [1] , has proposed fuzzy fingerprint technique. In this technique DLD provider use special set of the sensitive data digests. Set intersection method is used between the digest. But set intersection is an orderless as ordered of digest is not analyze every time it may mismatch. so sometimes it generate false alert rate.

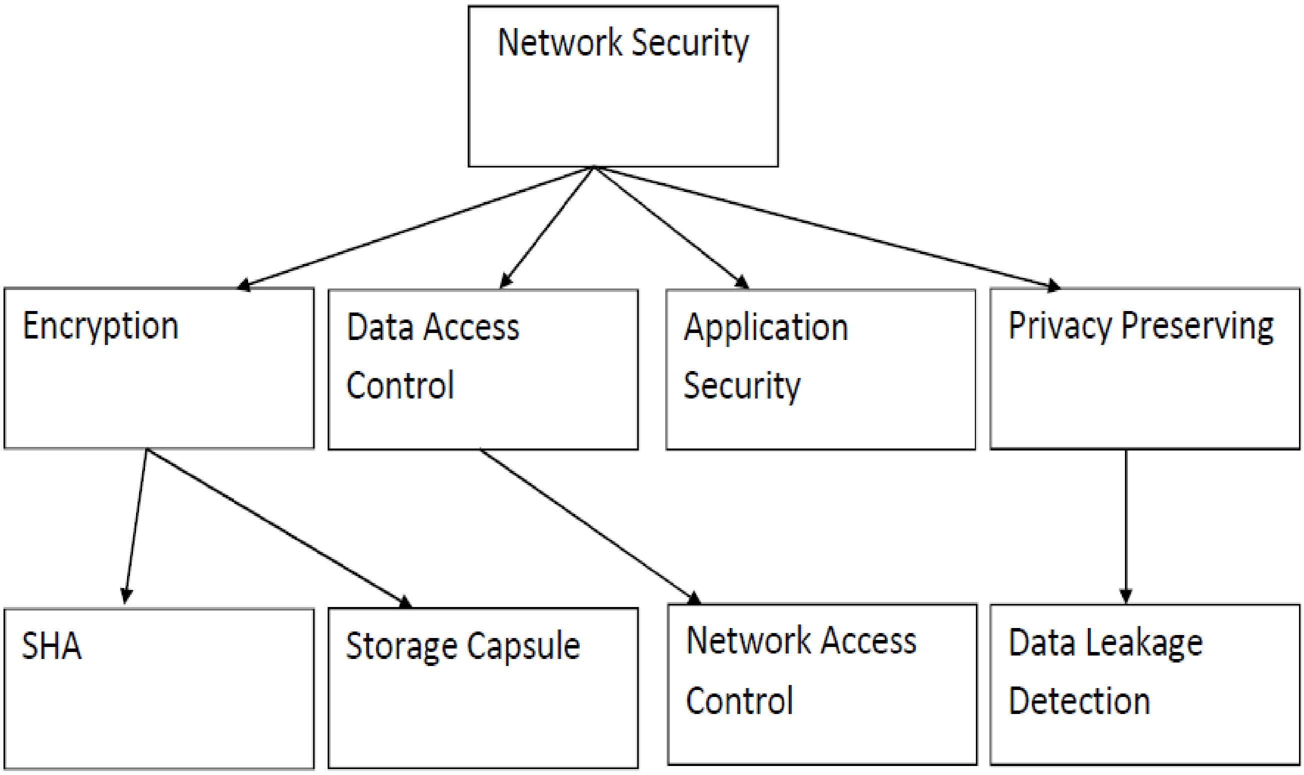

Figure 1. Structure of Literature Survey

III. PROPOSED SOLUTION

In the proposed system, host assisted mechanism is used which checks the frequency of occurrence of data. Highly differentiated values are considered as sensitive and fingerprints are generated for them. Repeated values are ignored in this method. Statistical approach is use to generate sensitive data and it is stored in table. . The fingerprints are generated by data leak detection (DLD) provider and identifies potential leaks by matching the fingerprints. The potential leak consist of real leaks and noises so that no one can get exact information about the sensitive data. Then data owner post process data send by the DLD provider to check where there is any leak in the sensitive data .The objectives are to improve the detection time and to improve detection of sensitive packet.

The architecture of the proposed system is shown in Figure 2. They list the following operations executed.

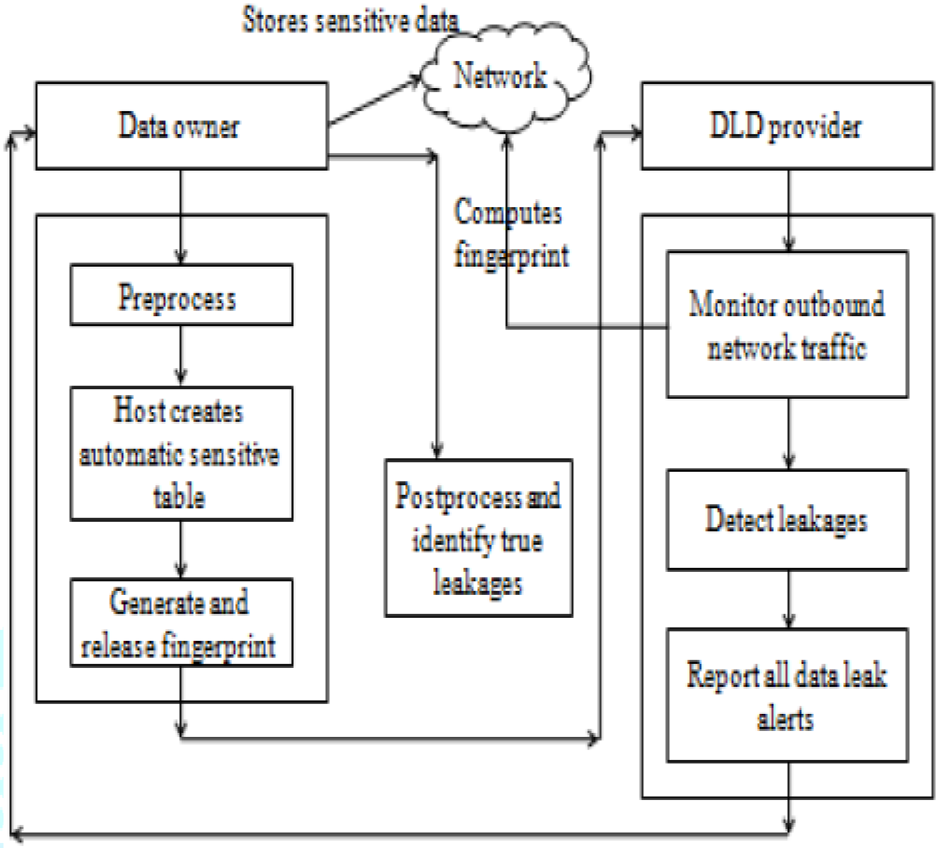
Stores sensitive data
Network
Data owner
DLD provider
Computes fingerprint
Preprocess
Host creates automatic sensitive table
Generate and release fingerprint
Postprocess and identify true leakages
Monitor outbound network traffic
Detect leakages
Report all data leak alerts

Figure 2. Architecture of the Propose System

1. Data owner PREPROCESS the sensitive data. The data owner stores their sensitive data in their network. They need their data to be in a protected way. They can't able to check the data frequently.
2. Host assisted mechanism is used after the data have been stored in the network. This checks the frequency of occurrence of data. Highly differentiated values are considered as sensitive and fingerprints are generated for them. Repeated values are ignored in this method. . Statistical approach is use to generate sensitive data and it is stored in table.. Only the important data are considered here so the detection will be very effective through this mechanism.
3. Data owner release fingerprint to DLD provider.
4. The fingerprints are generated by data leak detection (DLD) provider and identifies potential leaks by matching the fingerprints. The potential leak consist of real leaks and noises so that no one can get exact information about the sensitive data..
5. Then they will be match fingerprints from sensitive data and network traffic. It they will match then the data leak detection provider will send an alert to the data owner.
6. Data owner post-processes the data sent back by the DLD provider and check whether there is any real data leak.

Algorithm

The design section presents algorithm of system. Algorithm describes the working of the system.

The pseudo code for host assisted mechanism is as follows. Host assisted mechanism is used which takes only highly differentiated data as sensitive. Then fingerprints are generated for them. The DLD provider will perform operation for the sensitive data fingerprints and find out the leakages quickly.

Algorithm 1: Host assisted Mechanism Algorithm

```
Begin

1. Enter the data
2. Covert data into ASCII code
3. View ASCII code
4. Find mean and standard deviation
5. Consider highly differentiated values as sensitive data

End
```

IV. CONCLUSION

The host assisted mechanism is use to find out the data leakage within less time . Existing detection system takes all the data to conduct the detection operation. But in the proposed system the data owner the sensitive data is kept to a minimum level. It helps to delegate the detection operation to DLD provider without revealing sensitive data. The privacy is achieved and detection operation is done efficiently. Detection time will be reduced through this mechanism.

REFERENCES

[1] Danfeng Yao ,Xiaokui Shu and Elisa Bertino, Fellow IEEE,,”Privacy-Preserving Detection of Sensitive Data Exposure”, IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 10, NO. 5, MAY 2015

[2] A. Broder and M. Mitzenmacher, Network applications of bloom filters: A survey,Internet Math., vol. 1, no. 4, pp. 485509, 2004..

[3] H. Yin, M. Egele, D. Song, C. Kruegel, and E. Kirda, “Panorama: Capturing system wide information flow for malware detection and analysis”, in Proc. 14th Association for Computer Machinery Conf. Comput. Commun. Secur., 2007, pp. 116127.

[4] G. Karjoth and M. Schunter, ‘A privacy policy model for enterprises’, in Proc. *15th IEEE Comput. Secur. Found. Workshop*, Jun. 2002, pp. 271281.

[5] K. Borders, B. Lau, , E. V. Weele and A. Prakash, “Protecting confidential data on personal computers with storage capsules”, in Proc. 18th USENIX security Symp., 2009, pp. 367382.

[6] A. Nadkarni and W. Enck, Preventing accidental data disclosure in modern operating systems, in Proc. 20th Association for Computer Machinery Conf. Comput. Commun. Secur., 2013, pp 10291042.

[7] Y. Jang, B. D. Payne,, S. P. Chung and W. Lee,” Gyrus: A framework for user-intent monitoring of text-based networked applications”, in Proc. 23rd USENIX Security Symp., 2014, pp. 7993.

[8] X. Shu and D. Yao, *Data leak detection as a service*, in Proc. 8th Int. Conf. Secur. Privacy Commun. Netw., 2012, pp. 222240.

[9] 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. 8397, May 2013.