PREVENTING DATA OVER-COLLECTION USING DYNAMIC PERMISSION MAPPING IN MOBILE CLOUD FRAMEWORKS

Mrs.Geetha[1],
[1]Asso.Professor,
Department of Computer Science,
Rajalakshmi institute of Technology ,India.
geetha.m@ritchennai.edu.in

Chandru.T[2],
[2], UG Student,
Department of Computer Science ,
Rajalakshmi institute of Technology, India.
chandruakash02@gmail.com

Dinesh Ram.S[3],
[3], UG Student,
Department of Computer Science ,
Rajalakshmi institute of Technology, India.
dineshram98409@gmail.com

Praveen .A[4]
[4] UG Student,
Department of Computer Science ,
Rajalakshmi institute of Technology, India.
praveeen2705@gmail.com

ABSTRACT: The sensitive data leaks on the computer system are the major threats to the organisations.The major causes for data loss are done by the human errors and the lack of proper detection mechanism. The organizations need some of the requirements which is used identify the data leaks that are stored on and also checking them when they are being transformed. We help in securing sensitive data using data leak detection technique . There was no security in existing system, Therefore the data provider can enter the data without the permission of the data-owner and also it provides a sequence comparing technique which leads to data traffic and takes more time in finding the sensitive data. We propose a data-leak detection technique which uses the Lucene search engine framework and Levenshtein-distance technique to avoid data leak and provide security to sensitive data and also the safe outsourcing of resources .Hence our technique helps in preventing transforming data leaks and provide security to sensitive data.

KEYWORDS:

Data leak detection, lucene search engine, levenshtein-distance technique .

I.INTRODUCTION:

According to recent reports from the security firms the number of data-leak incidents have grown rapidly in research institutions and government organizations .Human errors are one of the main causes of data leaks among various data leak cases .The number of data leak records are maintained as the records by the Risk Based Security(RBS). The data leaks are mostly caused by planned attacks by assigning the wrong privilege. Stopping data leaks from occuring requires a data confinement stealthy malware detection. Data-leak detection (DLD) typically performs Deep packet inspection (DPI) which finds whether there are any occurrences of sensitive data patterns. If the sensitive data's have been detected they are prevented from sending it to the unauthorized persons.

II.EXISTING WORK:

Sequence search method is currently being used for data leak detection. However, this requirement is undesirable, as it may affect the confidentiality of the sensitive information. Sequence search method requires a substantial amount of time to check the whole plaintext in the document. This leads to increase in data traffic. There was no privacy preserving in existing system, so providers can access the information without data-owner's permission. The searching process used in this process may take a large amount of time resulting in heavy data traffic.

Problem Definition

1. **Inadvertent data leak :**

A responsible user accidentally leaks the sensitive data in the outbound traffic.Unintentional forwarding of an internal email and attachment to a outsider and not using a encryption are caused by unmindful individuals

2. **Malicious data leak :**

The users within the organization may steal sensitive data or information from the organization or the host.

3. **Data Traffic and Time consumption:**

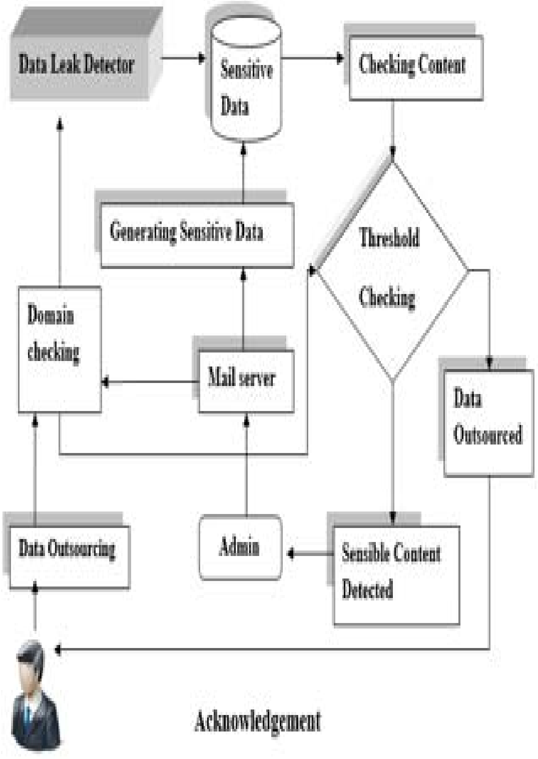
Data traffic on proxy server and mail server affects the performance of detection technique and time delay of common legitimate users.

978-1-5090-4855-7/17/$31.00 ©2017 IEEE

4. **Static filtering of authorized users:** This approach of static filtering technique for the authorized users affect the efficient of DLD.

III.PROPOSED SYSTEM:

- In our proposed system we use a data-leak detection solution, which can be outsourced from an organization, we design and implement Lucene search engine framework Levenshtein-distance technique to avoid data leak and also provide privacy preserving to sensitive data. Two important key personnels in our proposed model are the data owners and the mail server.
- The sensitive data that are to be transformed is provided by the data owners and they also maintain the network traffic of the organization for the data leak users within the organisation.
- Mail Server - DLD provider checks for the most data leaks by maintain a constant view over through the network traffic. The main aim of the process is to detect the data leaks and the sensitive data's are available for inspecting the data's. The plaintext or data present in the organized channel can either be a encrypted channel or a unencrypted channel and it can be extracted and checked by an authority. Authority has the threshold for every categorized position of users.
- In our security model, analysis system is safe and reliable. Privacy-preserved data-leak detection can be achieved by computation steps. It other method for the detection system.
- We use the web service to manage the users judicious content instead of data bases because of static performance and rough data handling. Even the sensible data storage have to safeguarded from vulnerabilities in existing system. For that purpose we used to maintain the sensible data in cloud.



IV.IMPLEMENTATION:

Build data leakage detection framework:

•The Sensitive data that is maintained in the cloud is provided by the data owner of the mail server, creating the catalogue for lucene search framework and other data leakage detectors. The accredited customer's details and server details are maintained in the data owners cloud. The Data Leak Detector helps in processing the sensitive information. The DLD performs detection mechanism.

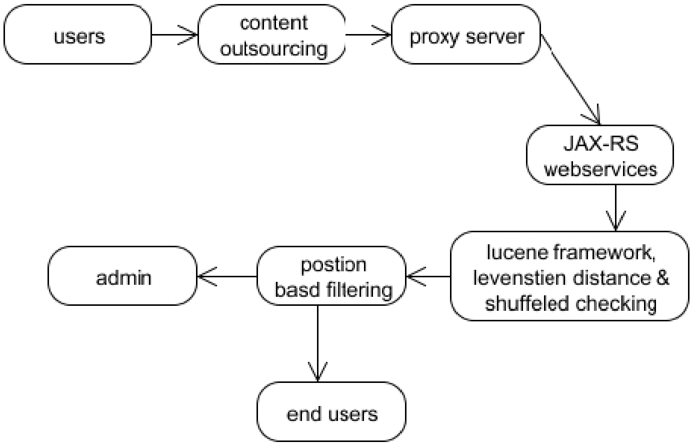
• The DLD consist of lucene search engine framework, levenshtein distance algorithm and our own shuffled checking algorithm. Every data outsourcing from authorized user transformation are configured with the cloud directly by the DLD.

Content Outsourcing with DLD Checker:

- The content that is sent to other organizations are checked by the DLD.They are verified with the contents that are stored .All the stored data are conserved in the index file. Using the Index file it will identify the sensitive data and assign the threshold to the employees. DLD will not permit the leakage of any confidential data to other organization.
- In proxy mail server the filteration occurres by using the users email domain. The recovery of users details are made by using their mail from the cloud. Then thresholds are given for the users based on their designation and the transferred content has been tested by lucene framework search engine, levenshtein distance checking .

Sensitive Data Detection and Quote Request:

- The DLD framework checks the content that are being sent outside, if any data leak occurs the DLD will detect the confidential data. Here DLD will check not only the confidential data but also it will check the access privilege. Every data owner maintain common access privilege every file. For example, all the contents are encrypted before they are sent outside. If DLD identifies any important information being sent outside the organization then it will detect the Sensitive content in between of the file outsourcing.
- For the purpose of fallacious alert, we use a threshold for every users designation. If the Sensitive content percentage of transferred file exceeds the threshold percentage which prompt alert mail to Admin of the proxy mail server. Alert mail consists of the whole information about the users even what are the confidential contents are tings from the transferred content by the DLD framework.
- After the filtration of mail in Mail server, the user can claim or quote the request to admin with the quote reasons details. Finally the Mail server admin can go through the quote request mails from the quote users whether he can pass through the mails or not.



V.CONCLUSION

Hence we proposed and developed fast detection of data-leakage framework to avoid sensitive data exposure and also provide privacy-preserving to sensitive data. The following methods have been used

- Lucene search framework to detect the sensible data easily using indexing technique.
- 2. Levenshtein distance algorithm to detect the shuffling of transferred mail content.
- 3. We implement the own logics for detect sampling of n-grams in transferred mail content appropriately.

We implement threshold rate based on assigning and checking domains based on user filtering technique.

VI.REFERENCES:

[1] Jing Zhang et all., " Fast Detection of Transformed Data Leaks", IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 11, MARCH 2016.

[2] X. Shu et all, "Rapid screening of transformed data leaks with efficient algorithms and parallel computing," in Proc. 5th ACM Conf. Data Appl. Secure. Privacy (CODASPY), San Antonio, TX, USA, Mar. 2015.

[3] D. Yao et all., "Privacy-preserving detection of sensitive data exposure," IEEE Trans. Inf. Forensics Security, vol. 10, no. 5, pp. 1092–1103, May 2015.

[4] L. De Carli et all., "Beyond pattern matching: A concurrency model for stateful deep packet inspection," in Proc. ACM SIGSAC Conf. Comput. Commun. Secur., 2014.

[5] A. V. Ah et all., “Efficient string matching: An aid to bibliographic search,” *Commun. ACM*, vol. 18, no. 6, pp. 333–340, Jun. 1975.

[6] R. S. Boyer et all., “A fast string searching algorithm,” *Commun. ACM*, vol. 20, no. 10, pp. 762–772, Oct. 1977.

[7] Pandithurai, O&Sureshkumar, C 2016, ‘High performance multipath routing algorithm using CPEGASIS protocol in Wireless sensor cloud Environment’ Circuits and Systems (CS)acceptance date on 30 may 2016 Published Online August, vol. 7, no.10,pp. 3246-3252.

[8] S. Kumar et all., “Curing regular expressions matching algorithms from insomnia, amnesia, and acalculia,” in *Proc. 3rd ACM/IEEE Symp. Archit. Netw. Commun. Syst. (ANCS)*, 2007, pp. 155–164.

[9] Pandithurai O&Sureshkumar, C 2016,‘CPEGASIS: Efficient Data retrieval in wireless sensor networks using Cloud computing Environment in Asian journal of information technology’, Year: 2016 | Volume: 15 | Issue: 19 | Page No.: 3725-3729

[10] H. A. Kholidy et all., “DDSGA: A data-driven semi-global alignment approach for detecting masquerade attacks,” *IEEE Trans. Dependable Secure Comput.*, vol. 12, no. 2, pp. 164–178, Mar./Apr. 2015.

[11] S. F. Altschul et all., “Basic local alignment search tool,” *J. Molecular Biol.*, vol. 215, no. 3, pp. 403–410, Oct. 1990.