Preventation of Data Content Leakage with Secured Encryption Alogorthim

Sagar Prasad, Malti Nagle, Tarique Zeya Khan
Computer Science & Engineering, SCET College,
Bhopal, Madhya Pradesh, India

Abstract— The information leak of sensitive data on systems has a serious threat to organization data security. Statistics show that the improper encryption on files and communications due to human errors is one of the leading causes of information loss. So there a need tools to identify the exposure of sensitive data by monitoring the content in storage and transmission. However, detecting the exposure of sensitive data information is challenging due to data transformation in the content. Transformations result in highly unpredictable leak patterns. In this paper, it is utilize sequence alignment techniques used for detecting complex data-leak patterns. This algorithm is designed for detecting long and inexact sensitive data patterns. This detection is paired with a comparable sampling algorithm, which allows one to compare the similarity of two separately sampled sequences. The system achieves good detection accuracy in recognizing transformed leaks. It implement a parallelized version of our algorithms in graphics processing unit to achieves high analysis data. In the case of collective privacy preservation, organizations have to cope with some interesting conflicts. For instance, when personal information undergoes analysis processes that produce new facts about users' shopping patterns, hobbies, or preferences, these facts could be used in recommender systems to predict or affect their future shopping patterns. In general, this scenario is beneficial to both users and organizations. However, when organizations share data in a collaborative project, the goal is not only to protect personally identifiable information but also sensitive knowledge represented by some strategic patterns. To demonstrate the high multithreading scalability of the data leak detection method required by a requirement of organization.

Keywords— Information leak detection, content inspection, sampling, alignment, dynamic programming, sensitive data patterns.

I. INTRODUCTION

To minimize the exposure of sensitive data and documents, an organization needs to prevent clear text sensitive data from appearing in the storage or communication. In today's increasingly digital world, there is often a tension between safeguarding privacy and sharing information. Although, in general, sensitive data clearly needs to be kept confidential data owners are often motivated, or forced, to share sensitive information Privacy-Preserving Sharing of Sensitive Information , and proposes one efficient and secure instantiation that functions as a privacy shield to protect parties from disclosing more than the required minimum of sensitive information. We model in the context of simple database-querying applications with two parties: a server that has a database, and a client, performing simple disjunctive equality queries Detecting the exposure of sensitive information is challenging due to data transformation in the content. Transformations result in highly unpredictable leak patterns. In this paper, we utilize sequence alignment techniques for detecting complex data-leak asymmetric cryptography, facilitate the creation of a verifiable association between a public key and the identity other attributes of the holder of the corresponding private key, for uses such as authenticating the identity of a specific entity, ensuring the integrity of information, providing support for non repudiation, and establishing an encrypted communications section.

II. BRIEF LITERATURE SURVEY

Big data analysis system concept for detecting unknown attacks:Sung-Hwan Ahn Nam-Uk Kim ,Tai-Myoung Chung. 16-19 Feb. 2014(IEEE). Unknown cyber-attacks are increasing because existing security systems are not able to detect them. Big data analysis techniques that can extract information from a variety of sources to detect future attacks. The event of new and previously unknown attacks, detection rate becomes very low and false negative increases. To defend against these unknown attacks. Does not detect future Advanced Persistent Threat (APT) detection.[1]

Big Data Analytics with Hadoop to analyze Targeted Attacks on Enterprise Data: Bhawna Gupta, Dr.Kiran Joyti in Journal of Computer Science and Information Technologies, Vol.5, 2014, (IEEE). Big data security analytics is used for the growing practice of organization to gather and analyze security data to detect vulnerabilities and intrusions. Security and Information Event Monitoring (SIEM) system. The malicious and targeted attacks have become main subject for government, organization or indust. Big data analytics is the process of analyzing big data to find hidden patterns, unknown correlations and other useful information that can be extracted to make better decisions. It is used effectively and at the same time, hackers can leave their targets forever.[2] Zero Day Attack Signatures Detection Using Honeypot: Musca, Mirica, E. ; Deaconescu, R. IEEE 29-31 May 2013. Unexpected behavior. Fault distribution studies

show that there is a correlation between the number of lines of code and the number of faults. LCS algorithm on the packet content of a number of connections going to the same services. Zero-day attack or threat is a computer threat that tries to exploit computer application vulnerabilities that are unknown to others or undisclosed to the software developer. Vulnerability window which is the time between the first exploitation of vulnerability and when software developers start to develop a countermeasure to that threat. [3]

Cloud Model based Outlier Detection Algorithm for Categorical Data: Dajiang Lei Liping Zhang And Lisheng Zhang, Vol. 6, No. 4, August, 2013. Numerical data but there will be a large number of categorical data in real life. Some outlier detection algorithm shave been designed. for categorical data. There are two main problems of outlier detection for categorical data, which are the similarity measure between categorical data objects and the detection efficiency. Outlier detection algorithm for categorical data. Efficient outlier detectioncan help us make good decisions on erroneous data or prevent the negative influence of malicious and faulty behavior. Many data mining techniques try to reduce the influence of outliers or eliminate them entirely. The in foremention manner may result in the loss of important hidden information.[4][5]

Cloud Computing-Based Forensic Analysis for Collaborative Network Security Management System: Zhen Chen*, Fuye Han, Junwei Cao, Xin Jiang, and Shuo Chen.1, February 2013. Internet security problems remain a major challenge with many security concerns such as Internet worms, spam, and phishing attacks. Botnets, well-organized distributed network attacks, consist of a large number of bots that generate huge volumes of spam or launch Distributed Denial of Service (DDoS) attacks on victim hosts. A distributed security overlay network with a centralized security center leverages a peer-to-peer communication protocol used in the UTMs collaborative module. These new security rules are enforced by collaborative UTM and the feedback events of such rules are returned to the security center. Collaborative network security management system can not identify the intrustion.[6] Understanding privacy in data mining requires understanding how privacy can be violated and the possible means for preventing privacy violation. In general, one major factor contributes to privacy violation in data mining: the misuse of data. Users' privacy can be violated in different ways and with different intentions. Although data mining can be extremely valuable in many applications (e.g., business, medical analysis, etc), it can also, in the absence of adequate safeguards, violate informational privacy. Privacy can be violated if personal data
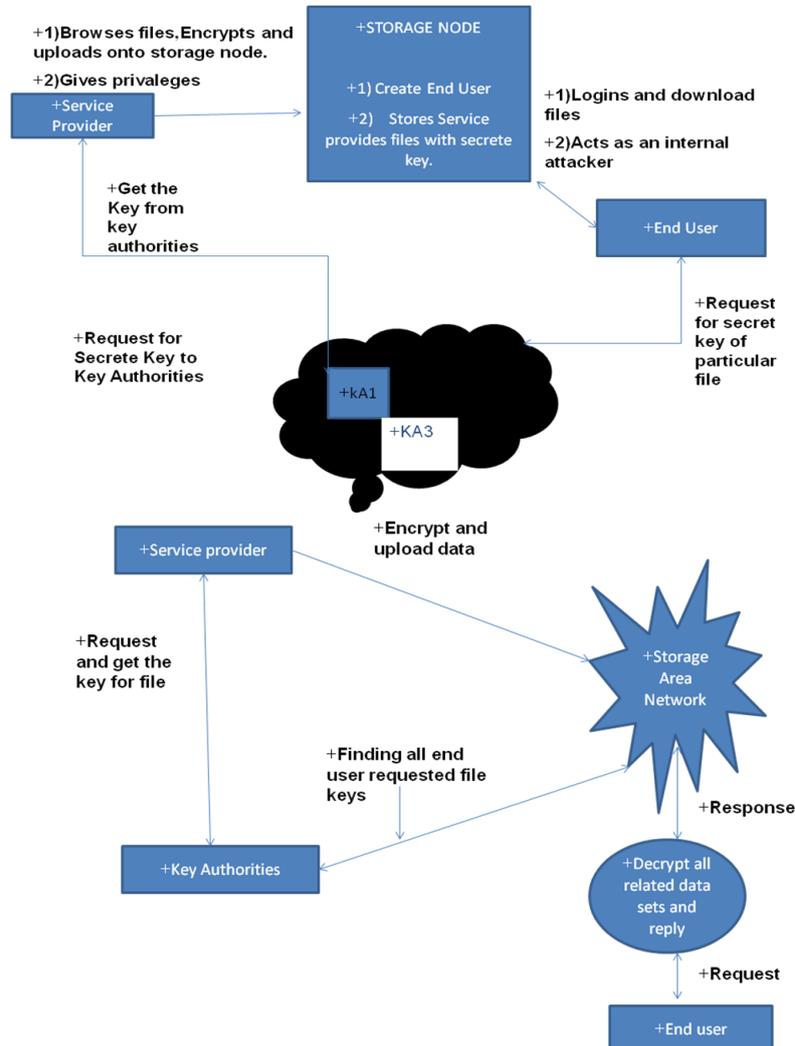
III. FLOW CHART

Fig 1. Dataflow diagram

The A typical setting involves two parties: one that seeks information from the other that is either motivated, or compelled, to share (only) the requested information. Consequently, in numerous occasions, there is a tension between information sharing and privacy. On the one hand, sensitive data needs to be kept confidential; on the other hand, data owners may be willing, or forced, to share information.

IV. PROBLEM DETECTED

We extensively evaluate the accuracy of our solution with several types of datasets under a multitude of real-world data leak scenarios. This module allows the user to register their identity into the system with proper input parameters. The key generation centers play a vital role in it, which generates public/ secret parameters . The key authorities consist of a central authority and multiple local authorities. Assume that there are secure and reliable communication channels between a central authority and each local authority during the initial key setup and generation phase. Each local authority manages different attributes and issues corresponding attribute keys to users. They grant differential access rights to individual users based on the users' attributes. The key authorities are assumed to be honest but curious. That is, they will honestly execute the assigned tasks in the system however they would like to learn information of encrypted contents as much as possible.

V. RESEARCH METHOD

1) Identity Key Generation The key generation module helps the users to share the information between source and destination. After getting the confirmation response from the receiver side the sender fix the information and encrypt it. At this time a key will be generated and sent to the receiver area. That key is useful for decrypt the data at receiver end. As well as an entity that stores data from senders and provide corresponding access to users. It may be mobile or static. Similar to the previous schemes, and also assume the storage node to be semi trusted that is honest but curious.

2) 3DES Based Encryption In Ciphertext Policy Attribute based Encryption scheme, the encryptor can fix the policy, who can decrypt the encrypted message. The policy can be formed with the help of attributes. In access policy is sent along with the ciphertext. We propose a method in which the access policy need not be sent along with the ciphertext, by which we are able to preserve the privacy of the encryptor. This techniques encrypted data can be kept confidential even if the storage server is untrusted. Moreover, our methods are secure against collusion attacks. Previous Attribute Based Encryption systems used attributes to describe the encrypted data and built policies into user's keys; while in our system attributes are used to describe a user's credentials, and a party encrypting data determines a policy for who can decrypt.

3) Confidential Data Interchange This is an entity who owns confidential messages or data and wishes to store them into the external data storage node for ease of sharing or for reliable delivery to users in the extreme networking environments. A sender is responsible for defining (attribute based) access policy and enforcing it on its own data by encrypting the data under the policy before storing it to the storage node. This is a mobile node who wants to access the data stored at the storage node (e.g., a soldier). If a user possesses a set of attributes satisfying the access policy of theencrypted data defined by the sender, and is not revoked in any of the attributes, then he will be able to decrypt the ciphertext and obtain the data.

4) Administrative Access Controller The administrator owns full access rights of this entire site. Once the administrator find out any illegal activity or other misusing happens into the way of transaction between the respective sender and receiver then the admin immediately block the user access rights to transact using this site. The block will be unblocked after getting meaningful reason from the user end.

VI. CONCLUSION

Detecting multiple common data leak scenarios. The parallel versions of our prototype provide substantial speedup and indicate high scalability of our design. For future work, we plan to explore data-movement tracking approaches for data leak prevention on a host. Privacy guarantees are formally defined and achieved with provable security. Experimental results show that our approach is efficient for real-world applications. To support efficient handling of multiple auditing tasks, we further explore the technique of bilinear aggregate signature to extend our main result into a multi-user setting, where can perform multiple auditing tasks simultaneously. Extensive security and performance analysis shows the proposed schemes are provably secure and highly efficient.

REFERENCES

[1] X. Shu, J. Zhang, D. Yao, and W.-C. Feng, "Rapid and parallel content screening for detecting transformed data exposure," in *Proc. 3rd Int. Workshop Secur. Privacy Big Data (BigSecurity)*, Apr./May 2015.

[2] X. Shu, J. Zhang, D. Yao, and W.-C. Feng, "Rapid screening of transformed data leaks with efficient algorithms and parallel computing," in *Proc. 5th ACM Conf. Data Appl. Secur. Privacy* , Mar. 2015.

[4] Kaspersky Lab. (2014). *Global Corporate IT Security Risks*. http://media.kaspersky.com/en/business-security/ Kaspersky_Global_IT_Security_Risks_Survey_report_Eng_final.pdf 2014

[5] L. De Carli, R. Sommer, and S. Jha, "Beyond pattern matching: A concurrency model for stateful deep packet inspection," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2014, pp. 1378–1390.

[6]   A. V. Aho and M. J. Corasick, "Efficient string matching: An aid to bibliographic search," *Commun. ACM*, vol. 18, no. 6, pp. 333–340, Jun. 1975.

[7]   R. S. Boyer and J. S. Moore, "A fast string searching algorithm," *Commun. ACM*, vol. 20, no. 10, pp. 762–772, Oct. 1977.

[8]   S. Kumar, B. Chandrasekaran, J. Turner, and G. Varghese, "Curing regular expressions matching algorithms from insomnia, amnesia, and acalculia," in *Proc. 3rd ACM/IEEE Symp. Archit. Net w. Commun. Syst. (ANCS)*, 2007, pp. 155–164.