ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 5, Issue 12, December 2016

PREVENTION OF DATA CONTENT LEAKAGE WITH SECURED ENCRYPTION ALGORITHM

Mr. Sagar Prasad, Ms.Malti Nagle, Mr.Tarique Zeya Khan

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 have the high multithreading scalability of the data leak detection.**

***Index Terms*— Information leak detection, Content inspection, Sampling alignment, Dynamic programming, etc..**

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. 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 loss by 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. PROPOSED SYSTEM

The purpose of this proposed work is to provide the approach functions as a privacy shield to protect parties from disclosing more than the required minimum of their respective sensitive information. PPSSI deployment prompts several challenges, which are addressed in this project. Extensive experimental results attest to the practicality of attained privacy features and show that our approach incurs quite low overhead. For better attack detection, big data incorporates attack graph analytical procedures into the intrusion detection processes. We must note that the design of does not intend to improve any of the existing intrusion detection algorithms; indeed, employs a reconfigurable virtual networking approach to detect and counter the attempts to compromise VMs, thus preventing zombie VMs.

The proposed method has several advantages.

1. To avoid the attacker.
2. Secrecy of the data should be maintained.
3. The scheme is robust to withstand brute force attacks.

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. the absence of adequate safeguards, violate informational privacy. Privacy can be

violated if personal data are used for other purposes subsequent to the original transaction between an individual and an organization when the information was collected (Culnan, 1993).

One of the sources of privacy violation is called data magnets (Rezgui et al., 2003). Data magnets are techniques and tools used to collect personal data. Examples of data magnets include explicitly collecting information through on-line registration, identifying users through IP addresses, software downloads that require registration, and indirectly collecting information for secondary usage. In many cases, users may or may not be aware that information is being collected or do not know how that information is collected. In particular, collected personal data can be used for secondary usage largely beyond the users' control and privacy laws. This scenario has led to an uncontrollable privacy violation not because of data mining itself, but fundamentally because of the misuse of data.

- *Individual privacy preservation*: The primary goal of data privacy is the protection of personally identifiable information. In general, information is considered personally identifiable if it can be linked, directly or indirectly, to an individual person. Thus, when personal data are subjected to mining, the attribute values associated with individuals are private and must be protected from disclosure. Miners are then able to learn from global models rather than from the characteristics of a particular individual.

- *Collective privacy preservation*: Protecting personal data may not be enough. Sometimes, we may need to protect against learning sensitive knowledge representing the activities of a group. We refer to the protection of sensitive knowledge as collective privacy preservation. The goal here is quite similar to that one for statistical databases, in which

security control mechanisms provide aggregate information about groups (population) and, at the same time, prevent disclosure of confidential information about individuals. However, unlike as is the case for statistical databases, another objective of collective privacy preservation is to protect sensitive knowledge that can provide competitive advantage in the business world.

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.

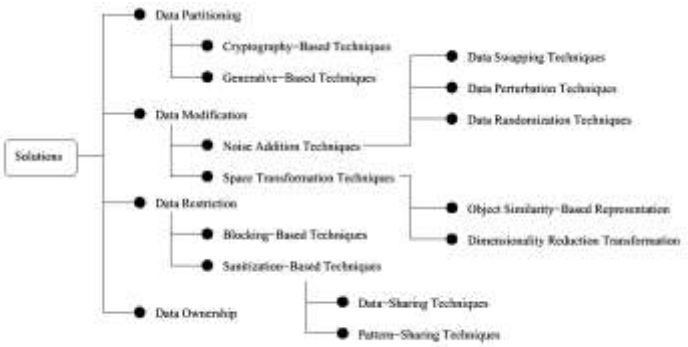

Figure 1. A taxonomy of PPDM techniques

To increase the security level this proposed scheme overcomes the limitation of "Hybrid encryption algorithm proposed . The proposed enhanced scheme includes Triple DES,
RSA and MD5.Triple DES (Variant of DES) strengthens the security of Data transmission. Reason behind for selecting triple DES rather than Double DES is that in double DES algorithm the key used for encryption and decryption is suspected to meet-in-middle attack. key distribution problem and in addition to this, MD5 to verify the integrity of the message. Use of message digest algorithm in combination of cryptographic algorithm.

III. RESEARCH METHOD

A. *ENCRYPTION ALOLGORITHM*

Steps

1. Take a file packet [N]
2. Encrypt the plaintext blocks using single DES with E key K_1.
3. Now decrypt the output of step 1 DP using single DES with key K_2.
4. Finally, encrypt the output of step 2 using single DES with key K_3.
5. The output of step 3 is the ciphertext. (CT)
6. 128 key by md5 (MD) KD
7. CT= N K_3 (DP K_2(E K_1))
8. Iblock= CT+KD\
9. Iblock is send to battalion receiver.

Triple DES as an encrypt process, it is possible to use a 3TDES (hardware) implementation for single DES by setting K_1, K_2, and K_3 to be the same value. This provides backwards compatibility with DES. Second variant of Triple DES (2TDES) is identical to 3TDES except that $K_3$is replaced by K_1. In other words, user encrypt plaintext blocks with key K_1, then decrypt with key K_2, and finally encrypt with K_1 again. Total Iblock Length is 192 bits Values returned by a hash function are called **message digest** or simply **hash values**. It is a 128-bit hash function.

B. *DECRYPTION ALGORITHM*

1. Received Iblock= CT+KD
2. first and second secret keys or second and third secret keys are the same
3. Whichever key.
4. c = E3 (KD1 (K_1 (N))) = E3(N)
5. c = E3(KD3(K_2 (N))) = K_2 (N)
6. It is possible to use 3DES cipher with a secret 128 bit key.
7. In this case first and third secret keys are the same.
8. c = K 1(KD2(K1(N))
9. If key match data decryption

   m = D1 (K_2(KD3(c)))

IV. RESULT AND ANALYSIS

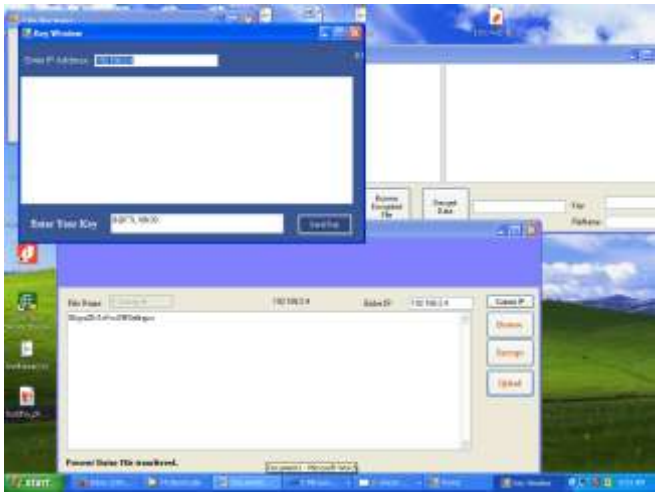

Fig 1. key Authorization

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective. The implementation stage involves careful planning, investigation of the existing system and it's constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods. We categorize three causes for sensitive data to appear on the outbound traffic of an organization, including the legitimate data use by the employees. Case I Inadvertent data leak: The sensitive data is accidentally leaked in the outbound traffic by a legitimate user. This paper focuses on detecting this type of accidental data leaks over supervised network channels. Inadvertent data leak may be due to human errors such as forgetting to use encryption, carelessly forwarding an internal email and attachments to outsiders.

Formally defined and achieved with provable security. Experimental results show that our approach is sufficiently 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. We also show how to extent our main scheme to support batch auditing for upon delegations from multi-users.

V. CONCLUSION

Privacy guarantees are formally defined and achieved with provable security. Experimental results show that our approach is sufficiently 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] Hiroki Nishiyama, Desmond Fomo, Zubair Md. Fadlullah,, and NeiKato,Fellow," Traffic Pattern Based Content Leakage Detection for Trusted Cont ent Delivery Networks" *IEEE Transaction on Parallel and Distributed System , Volume 25, No 2 Feb 2014*.

[2] K. Ramya, D. RamyaDorai, Dr. M. Rajaram "Tracing Illegal Redistributors of Streaming Contents using Traffic Patterns" *IJC A 2011*.

[3] A. Asano, H. Nishiyama, and N. Kato, "The Effect of Packet Reordering and EncryptedTraffic on Streaming Content Leakage Detection" Proc. *Int'l Conf. Computer Comm. Networks (ICCCN '10), pp. 1 6, Aug. 2010*.

[4] Y. Chu, S.G. Rao, S. Seshan, and H. Zhang, "Enabling Conferencing Applications on the Internet Using an Overlay Multicast Architecture," Proc.ACM SIGCOMM, pp. 55 67,Aug. 2010 .

[5] O. Adeyinka, "Analysis of IPSec VPNs Performance in a Multimedia Environment," Proc. Fourth *Int'l Conf. Intell igent Environments, pp. 25 - 30, 2008*

[6] M. Dobashi, H. Nakayama, N. Kato, Y. Nemoto, and A. Jamalipour, "Traitor Tracing Technology of Streaming Contents Delivery Using Traffic Pattern in Wired/Wireless Environments," Proc. *IEEE Global Telecomm. Conf., pp. 1 5, Nov./Dec. 2006.*

[7] S. Amarasing and M. Lertwatechakul, "The Study of Streaming Traffic Behavior," KKU Eng. J., vol. 33, no. 5, pp. 541 553, Sept./Oct. 2006.