International Journal on Applications in Engineering and Technology
Volume 2: Issue 8: August 2016, pp 14 - 16. www.aetsjournal.com ISSN (Online) : 2455 - 0523

A Survey on Data Leakage Identification Through Leak Filter Modeling and File Sealing Technique

M. Mansurabegam , P. Ezhilarasu , D. Satheesh Kumar , S.Jeevitha

Abstract— Cloud computing is the delivery of computing as a service rather than a product, whereby shared resources, software, and information are provided to computers and other devices as a utility over a network. Clouds can be classified as public, private or hybrid. Cloud computing relies on sharing of resources to achieve coherence and economies of scale, similar to a utility over a network. The leak of sensitive data on computer systems poses a serious threat to organizational security. The lack of proper encryption on files and communications due to human errors is one of the leading causes of data loss. In this paper, we need sequence alignment techniques for detecting complex data-leak patterns. Our algorithm is designed for detecting long and inexact sensitive data. So we introduce a parallelized version of our algorithms in graphics processing unit that achieves high analysis throughput. We determine the high multithreading scalability of our data leak detection method required by a sizable organization.

Keywords— Data leak detection, content inspection, sampling, alignment, dynamic programming and parallelism.

I. INTRODUCTION

There are the number of leaked sensitive data records has grown 10 times in the last 4 years, and it reached a record high of 1.1 billion in 2014. A significant portion of the data leak incidents are due to human errors, for example, a lost or stolen laptop containing unencrypted sensitive files or transmitting sensitive data without using end-to-end encryption such as PGP. A recent Kaspersky Lab survey shows that accidental leak by staff is the leading cause for internal data leaks in corporate. In order 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. Data leak detection differs from the anti-virus (AV) scanning (e.g., scanning file systems for malware signatures) or the network intrusion detection systems (NIDS) (e.g., scanning traffic payload for malicious patterns. It also focus on the production based companies for best transactions between company and the client. Now a day's automation products are getting increased and need of the automotive products also. Still the existing system are facing more drawbacks in security in the current software, order placing, manufacturing the products according to the order and delivering the products. The most important factor is computerizing all the data in a centralized server and taking backups of old records.

Our solution to the detection of transformed data leaks is a sequence alignment algorithm, executed on the sampled sensitive data sequence and the sampled content being inspected. The alignment produces scores indicating the amount of sensitive data contained in the content. We solve the scalability issue by sampling both the sensitive data and content sequences before aligning them. It produces the pair of a comparable sampling algorithm and a sampling-oblivious alignment algorithm. In this paper, we formalize and expand the description and analysis of our comparable sampling algorithm and sampling-oblivious alignment algorithm. In this paper, we focus on detecting inadvertent data leaks, and we assume the content in file system or network traffic (over supervised network channels) is available to the inspection system.

The objective of this paper is to propose securely communicate with the production unit to the administration unit, sales unit, quality check unit and store house unit and to avoid data leakage. A Data distributor has given sensitive data to supposedly trusted agents (unauthorized party).Some data has been leaked and found in an unauthorized place(e.g. on web or in somebody's laptop).The distributor must assess the likelihood that the leaked data came from one or more agents, as opposed to having been gathered by other independent means. Hence we design in such a way that the data on reaching any destination say agent or unauthorized party, data as well as the IP address of the receiver will reach the distributor. If the distributor receives an IP address other an agent's address, the distributor finds out that the data has been leaked. Fake object is inserted along with the data at the time

M.Mansurabegam , PG scholar, Department of Computer Science and Engineering Hindusthan College of Engineering and Technology, Coimbatore, Tamilnadu, India (Email : mansurabegam241@gmail.com)

Dr.P Ezhilarasu , Associate Professor, Department of Computer Science and Engineering Hindusthan College of Engineering and Technology, Coimbatore, Tamilnadu, India (Email : prof.p.ezhilarasu@gmail.com)

Mr. D. Satheesh Kumar , Assistant Professor, Department of Computer Science and Engineering ,Hindusthan College of Engineering and Technology, Coimbatore, Tamilnadu, India
(Email : hicetsatheesh@gmail.com)

S.Jeevitha , PG scholar, Department of Computer Science and Engineering , Hindusthan College of Engineering and Technology, Coimbatore, Tamilnadu, India (Email : jeevithaess@gmail.com)

of distribution for detecting the guilty agent. From the data received, the distributor compares with his database and finds out which agent has leaked the data by calculating the probability for each agent on the leaked data. Some of the major contributions are given here.

- This process will find out in case of any employee trying to leak any confidential data to the outside companies.
- Intrusion detection system (IDS) was introduced in the centralized server. This leads to avoid third party hacker or intruder to getting in to the centralized server.
- IP synchronization used for unique data usage. This application will be getting synchronized with the current IP address of the company.

II. VARIOUS METHODS

Table 1 Analysis of various method

S.NO	TITLE	PROCESS	FUTURE WORK
1.	Traffic Morphing: An Efficient Defense against Statistical Traffic Analysis [1]	To identify security threats and providing efficient management of network resources in network traffic analysis.	To reduce the VoIP classifier's accuracy from 81% to 20% with only15.4% overhead on average.
2.	Optimum Packet Length Masking [2]	To find the optimal solution of the masking problem in case of two application types. An explicit efficient algorithm are FTP-control and VoIP traffic.	To minimizing the required overhead.
3.	Sealing Information Leaks with Browser Side Obfuscation of Encrypted Flows [3]	To provides a comprehensive and configurable suite of traffic transformation techniques for a browser to defeat traffic analysis without requiring any server-side modifications.	To prevent information leaks and offer much better scalability and flexibility.
4.	Predicted Packet Padding for Anonymous Web Browsing Against Traffic Analysis Attacks [4]	Low latency systems often provide better performance and are intended for real-time applications, particularly for web browsing.	To minimizing the delays and bandwidth cost.
5.	Phonotactic Reconstruction of Encrypted VoIP Conversations: Hookt on Fon-iks [5]	The two common design decisions made in VoIP protocols—namely, the use of variable-bit-rate (VBR) codecs for speech encoding and length-preserving streammultaneously segmenting words.	To apply a similar methodology when tackling the problem of reconstructing words from strings of phonemes. .
6.	Data leak detection as a service [6]	To describe a network-based data-leak detection (DLD) technique, the main feature of which is that the detection does not require the data owner to reveal the content of the sensitive data	To provide a quantifiable method to measure the privacy guarantee offered by our fuzzy fingerprint framework.
7.	Sampling techniques to accelerate pattern matching in network intrusion detection systems [7]	To perform deep packet inspection at high speed for security and application specific services. the problem of the large memory consumption of DFAs has been solved in many different ways, only a few works have focused on increasing the lookup speed.	To increase the speed of automata, which is still limited by the processing required by every single byte.
8.	Snort-Light weight intrusion detection for networks [8].	To provide a layer of defense which monitors network traffic for predefined suspicious activity or patterns, and alert system administrators when potential hostile traffic is detected.	To increase a specific security solution in a short amount of time.

III. CONCLUSION

The current developed system is found to be working accurately. The system is found to be delightful running under the web application system. The programming technique used in the design of the system provides a scope for further expansion and implementation of any change which may occur in future. The present work has been completed for web application to technical terms used in this application. This is because all technical fields and process are about to implement in the future enhancement.

REFERENCES

[1] Xiaokui shu , Jing Zhang , Danfeng (Daphne) yao."Fast detection of transformed data leaks" IEEE Transactions on information Forensics and security,VOL.,11,NO.3 march 2016.

[2] Kaspersky Lab. (2014). Global Corporate IT Security Risks. [Online]. Available:http://media.kaspersky.com/en/business-security/
[3] 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, pp. 191–196 K.
[4] Lee, H. Lin, and W.-C. Feng, "Performance characterization of data-intensive kernels on AMD fusion architectures," Comput. Sci.-Res. Develop., vol. 28, no. 2, pp. 175–184, May 2013.
[5] X. Shu, D. Yao, and E. Bertino, "Privacy-preserving detection of sensitive data exposure," IEEE Trans. Inf. Forensics Security, vol. 10, no. 5, pp. 1092–1103, May 2015.
[6] L. Yang, R. Karim, V. Ganapathy, and R. Smith, "Improving NFA-based signature matching using ordered binary decision diagrams," in Proc. 13th Int. Symp. Recent Adv. Intrusion Detect., Sep. 2010, pp. 58–78.
[7] Alistair McMonnies, "Object-oriented programming in ASP.NET", Pearson Education, and ISBN: 81-297-0649-0, First Indian Reprint 2004.
[8] Jittery R.Shapiro, "The Complete Reference ASP.NET" Edition 2002, Tata McGraw-Hill, Publishing Company Limited, New Delhi.
[9] Robert D.Schneider, Jettey R.Garbus, "Optimizing SQL Server", Second Edition, Pearson Education Asia, SBN: 981-4035-20-3.