Review on Importance and Advancement in Detecting Sensitive Data Leakage in Public Network

Ms. Revathi Yegappan
PG Scholar, Computer Network Engineering
Dept of Computer Science Engineering
NHCE, Bangalore
Email: revathiyegappan@gmail.com

·Dr. S. Mohan Kumar
Associate Professor, Department of Computer Science and Engineering,
New Horizon College of Engineering,
Bangalore.
Email: drsmohankumar@gmail.com

ABSTRACT- Data is one of the most valuable assets in every organization. It is required to detect and prevent the data loss from being stolen or leaked from the organization to the outside world. Data loss may occur intentionally or unintentionally by the internal employees or by the trusted third parties due to mishandling or by mistakes. These data loss causes damage to the organization brands and reputation. Even though there are various techniques to prevent the data loss, but it is essential to detect the leakage of sensitive data as soon as possible before leaving the trusted network. Data loss prevention controls should be effectively implemented in advance in detecting and preventing the sensitive data from being leaked out of an organization.

Keywords: Information Security, Cyber-Crime, Hacks, Attacks, Data Security, Security Risks, Data Leakage, Data Leak Detection, Network Security, Privacy.

INTRODUCTION

Sensitive data could be of different types, it could be related details about a client, employee, Health records of an individual, finance, credit cards details etc. Sensitive data is the information where the disclosure is protected by laws and regulations and mainly by the organization policy. The loss of sensitive data leads to financial damage and the reputation of an organization. This loss of sensitive information affects the organization, customer and the external parties whose information are compromised. Thus the leakage of sensitive information should be prevented from unauthorized transmission of data to the public domain.

Information security controls are employed to protect the sensitive data. An organization need to categorize its data asset and define their sensitivity and identify the level of protection by means of data classification. This classification of data helps to ensure adequate controls are provided to sensitive data .To secure these sensitive data, it is important to know what kind of data is considered to be sensitive, where these sensitive data's are located and who can access those data's.

Data violation is when the sensitive, protected data is viewed, stolen or accessed by unauthorized individual. Data breach involves accessing the personal health information, personally identifiable information, intellectual property. To avoid such data violations Government compliance regulations, industry guidelines provide strict governance on sensitive and personal data.

In any organization data security program is primarily implemented on the devices and media controls. This includes policies, encryption techniques and other necessary safeguards to protect the sensitive information that are stored in storage devices, systems and transportable media. A Data security program should be implemented by considering different factors like Technical safeguards, access controls, monitoring and logging, backup and recovery, data disposal, security training and awareness, auditing and testing, and response program.

While we understand importance of the data security and data loss, it's also necessary understand the impact of the data losses in today scenario. Evolution of science and technology trends impacts business in both positive and negative ways. From our analysis we understand that, most of the current technology trends in Information Technology has created an anxiety on the information security space.

For example the rising market in today's world is the consumer market. E-commerce had made anything and everything possible to buy in today's world. While this is nice to have, it has created high risk areas on information security. Organizations are compelled to provide by accessibility by all means like, mobile, Kiosks, Laptops, I pads, etc. and consumer could be any one. So the information of organization is available in all the places wherever there is connectivity. This increases susceptibility, sharing of data thru social media and accessible of unintended data owners.

We could also add more to above trends. Cloud revolution has significantly helped the business teams to reduce their Infrastructure cost. The flip side, it has created a risk in terms of Information privacy and challenges in compliances. This one of the reasons why many organizations have not shifted their ERP's to cloud mode.

While we understand the significance of Business Continuity Programs (BCP) in many organizations which is operating 24/7, damages or failures of these systems have created major impact on the financials and loyalty of these companies. While some organization faces the resource crunch situation, human resource do open the gates for third party vendors and contractors to work for their end clients. This will end up in providing complete access and rights to these resources as like an employee. Just in case the right not reset after their contract period or if the monitoring of accessibility is not done properly, it could internally be a high risk situation for the information available in the respective organization.

RELATED RESEARCH WORK

- Statistics from security firms, research institutions and government organizations show that the numbers of data-leak instances have grown rapidly in recent years. The rising cost of data loss incidents According to a 2010 Ponemon Institute study, the average total cost per data breach has risen to $7.2 million, or $214 per record lost. In addition to the costs of incidents increasing, the number of leaks appears to be increasing every year.
- Papadimitriou.P[et.al],(2009),'A Model for Data Leakage Detection',[8]-This paper proposes data allocation strategies (across the agents) that improve the probability of identifying leakages. These methods do not rely on alterations of the released data (e.g., watermarks). In some cases "realistic but fake" data records are injected to further improve our chances of detecting leakage and identifying the guilty party.
- Marecki.J.[et.al],(2010), 'A Decision Theoretic Approach to Data Leakage Prevention',[9]-This paper focus on domains with one information source (sender) and many information sinks (recipients) where: (i) sharing is mutually beneficial for the sender and the recipients, (ii) leaking a shared information is beneficial to the recipients but undesirable to the sender, and (iii) information sharing decisions of the sender are determined using imperfect monitoring of the (un)intended information leakage by the recipients.
- Jiangjiang Wu[et.al],(2011), 'An Active Data Leakage Prevention Model for Insider Threat',[7]-This paper presents an active data leakage prevention model for insider threat that combines trusted storage with virtual isolation technologies and expresses the protection requirements from the aspect of data object. It shows an implementation framework and give formal description as well as security properties proof.
- X. Shu and D.Yao[et.al], (2012), 'Data leak detection as a service',[6]- In this paper A network-based data-leak detection (DLD) technique is adopted to detect the accidental leaks due to human errors. The algorithm minimizes the exposure of sensitive data to other network.
- Alneyadi.S[et.al],(2013), 'Adaptable N-gram classification model for data leakage prevention',[5]- This paper uses N-grams statistical analysis for data classification purposes. The method is based on using N-grams frequency to classify documents under distinct categories. It uses simple taxicap geometry to compute the similarity between documents and existing categories.
- Yan Wen[et.al],(2014),'Towards Thwarting Data Leakage with Memory Page Access Interception',[4]-This paper uses Gemini, an instrumentation-free approach, to track data propagation dynamically and then prevent data leakage. Gemini leverages the page fault interrupt mechanism of the operating system, instead of DBI, to track memory page accesses, and then thwart the data leakage. As a result, Gemini is application transparent, i.e., it solves the application compatibility issue.
- Alneyadi.S[et.al],(2015),'Detecting Data Semantic: A Data Leakage Prevention Approach',[3]-In this paper, a statistical data leakage prevention (DLP) model is presented to classify data on the basis of semantics. This study contributes to the data leakage prevention field by using data statistical analysis to detect evolved confidential data. The approach was based on using the well-known information retrieval function Term Frequency-Inverse Document Frequency (TF-IDF).
- Shu, X[et.al].,(2015),'Fast Detection of Transformed Data Leaks',[1]-In this paper, a detection is coupled with a comparable sampling algorithm it compares the similarity of two separately sampled sequences.
- Yuri Shapiro[et.al],(2013),'Content-based data leakage detection using extended fingerprinting',[2]- In this paper an extension to the fingerprinting approach is done based on sorted k-skip-n-grams.

IJERGS

OBSERVATION

- Exact string matching technique fails to detect data leak in network due to low tolerance for unknown noise.
- Comparison based on regular expression supports wild card but it is not scalable and practically difficult to deploy.
- Even though there exists many such solutions to detect and prevent the leakage of confidential data but none of the methods provide absolute protection .Thus the purpose of this research is to provide advancement in detecting the sensitive data leakage in public domain by performing deep content inspection and providing the ability to discover the data leakage as fast as possible .Alert have to be provided about the vulnerability of the data exposure to the user and the administrators.

CONCLUSION

In this paper different data leakage detection models and techniques are premeditated .Thus it is very important to implement DLP controls and information security controls to manage data loss risks. These clear set of controls should be monitored over time and the focus is on defense in depth approach. The data leakage detection and prevention should ensure sensitive data remain safe. The goal of this module is to discover the leakage of confidential data by using a real dataset in public domain and the proposed method try to improve the accuracy and better detection.

REFERENCES:

[1] Xiaokui Shu, Jing Zhang, Danfeng (Daphne) Yao ,"Fast Detection of Transformed Data Leak", ieee transactions on information forensics and security, vol. 11, no. 3, march 2016

[2]Shapira, Yuri, Bracha Shapira, and Asaf Shabtai. "Content-based data leakage detection using extended fingerprinting." arXiv preprint arXiv:1302.2028 (2013).

[3]Alneyadi, Sultan, Elankayer Sithirasenan, and Vallipuram Muthukkumarasamy. "Detecting Data Semantic: A Data Leakage Prevention Approach." Trustcom/BigDataSE/ISPA, 2015 IEEE. Vol. 1. IEEE, 2015.

[4]Wen, Yan, Jinjing Zhao, and Hua Chen. "Towards Thwarting Data Leakage with Memory Page Access Interception." Dependable, Autonomic and Secure Computing (DASC), 2014 IEEE 12th International Conference on. IEEE, 2014.

[5]Alneyadi, Sultan, Elankayer Sithirasenan, and Vallipuram Muthukkumarasamy. "Adaptable n-gram classification model for data leakage prevention." Signal Processing and Communication Systems (ICSPCS), 2013 7th International Conference on. IEEE, 2013.

[6]Shu, Xiaokui, and Danfeng Daphne Yao. "Data leak detection as a service."Security and Privacy in Communication Networks. Springer Berlin Heidelberg, 2012. 222-240.

[7]Wu, Jiangjiang, et al. "An active data leakage prevention model for insider threat." Intelligence Information Processing and Trusted Computing (IPTC), 2011 2nd International Symposium on. IEEE, 2011.

[8]Papadimitriou, Panagiotis, and Hector Garcia-Molina. "A model for data leakage detection." Data Engineering, 2009. ICDE'09. IEEE 25th International Conference on. IEEE, 2009.

[9]Marecki, Janusz, Mudhakar Srivatsa, and Pradeep Varakantham. "A Decision Theoretic Approach to Data Leakage Prevention." Social Computing (SocialCom), 2010 IEEE Second International Conference on. IEEE, 2010.