Detection of Data Leakage using Misuseability Weight Measure

Mohanapriya.C
School of information technology and engineering
VIT University, vellore-632014
Tamil nadu, India

Thanapal.P
Assistant Proffessor(Senior),
School of information technology and engineering
VIT University, vellore-632014
Tamil nadu, India

ABSTRACT

Data leakage is a leakage which the information's were leaked out unknowingly. Especially company has partnership with other companies that need to share the information's together. A distributor has a set of given sensitive data to hypothetical trusted third party agents. While sharing the transactions data leakage may occur at any place. Preventing the data leakage is a serious challenge for organizations. In order to overcome these problems an M-score measurability weight measure is introduced. The goal of weight measure is to detect the data leakage when the sensitive data has been leaked, and if possible to identify the agent that leaked the data. This regards applications where the data cannot be perturbed. Perturbation is a useful technique where the data is modified and made less sensitive before being hand over to agents.

General Terms

Sensitive information, security.

Keywords

Data leakage, Weight measure

1. INTRODUCTION

A distributor has a given set of sensitive data to a number of hypothetical trusted third party agents. uses unpretentious technique for identifying data leakage of a record set. The weight measure considers the various levels of the data to which an insider is exposed out. The anomaly detection method is used to apply for learning the normal behavior of an insider in terms of sensitive level of information is usually exposed to. To improve the process of leakage handling incidents calculated by other misuse detection systems by activating the security officer to concentrates on incidents involving more sensitive data's.

2. MISUSEABILITY WEIGHT CONCEPT

Data's stored in company computers is very essential one. A company needs to secure and sustain the power undoubtedly. The data is important for regular basis work process in another hand. The sensitive information may be exposed by users who access the data in an organization. In order to detect the damage misuseability is introduced here. The sensitive level of data is revealed to, with the help of assigning scores prohibition model for insider threat that corporate trusted storage. The main goal of the project is to prevent the data leakage with the help of active defense. Bernhard Riedl, Veronika Grascher, Stefan Fenz, Thomas Neubauer[5], to each and every data. The misuseability weight can detect the extreme spoilage to the companies if the data is misbehaved. Based on this information an organization can take actions for those damages.

3. RELATED WORK

Amir Harel, Asaf Shabtai, Lior Rokach and Yuval Elovici[1], This author introduced a new methodology, Misuseability weight is used to detecting the data leakage. It also enables the data leakage prevention. This misuseability weight method assigns a unique score that presents the sensitive level of data revealed to the data. M-score measure is an algorithm used to applying this concept. It can determine the misuseability weight for single publication without considering the prior knowledge. It can be a previous publication and knowledge on the description of the publication. Later we can extend the M-score to the multiple publication and the sensitive combination sensitive values.

William Eberle and Lawrence Holder[2], This author proposed a methodology to find anomaly in business transactions and it measures using a graph based presentation. In our graph based anomalies detection (GBAD) methodology, anomaly objects of structural patterns are found in data that represents, relationships, entities and activity. The graphical based anomaly detection is nothing but a discrete-event simulation of real-world business transactions and it measures were followed by empirical results. The anomalies were successfully found with the minimal description length principles and probabilistic methodologies.

Zhang Xiaosong, Liu Fei , Chen Ting, Li Hua[3], This author explained a model novel oriented windows file systems filter driver for preventing sensitive data's from leakages. Its main goal is, system files can encrypted automatically based on encryption methodologies. Its mandatory features and transpicuous encryption, which can understand the inconsistency among user's flexibility and data security. This filter driver is especially for transparent encryption which gives a merit to the system compatibility. It will be utilized in intranet networks.

Jiangjiang Wu, Jie Zhou, Jun Ma, Songzhu Mei, Jiangchun Ren[4], Data leakage threat is the major issues in the information security. A research mainly concentrates on sensing and determines the data leakage without defense qualification. The author proposed an effective data leakage

Electronic health records(HER) promises to enhance the communication among the health care providers, thus leads to better quality treatments for patient's and cost reduction. Patient information can be highly sensitive and provides a

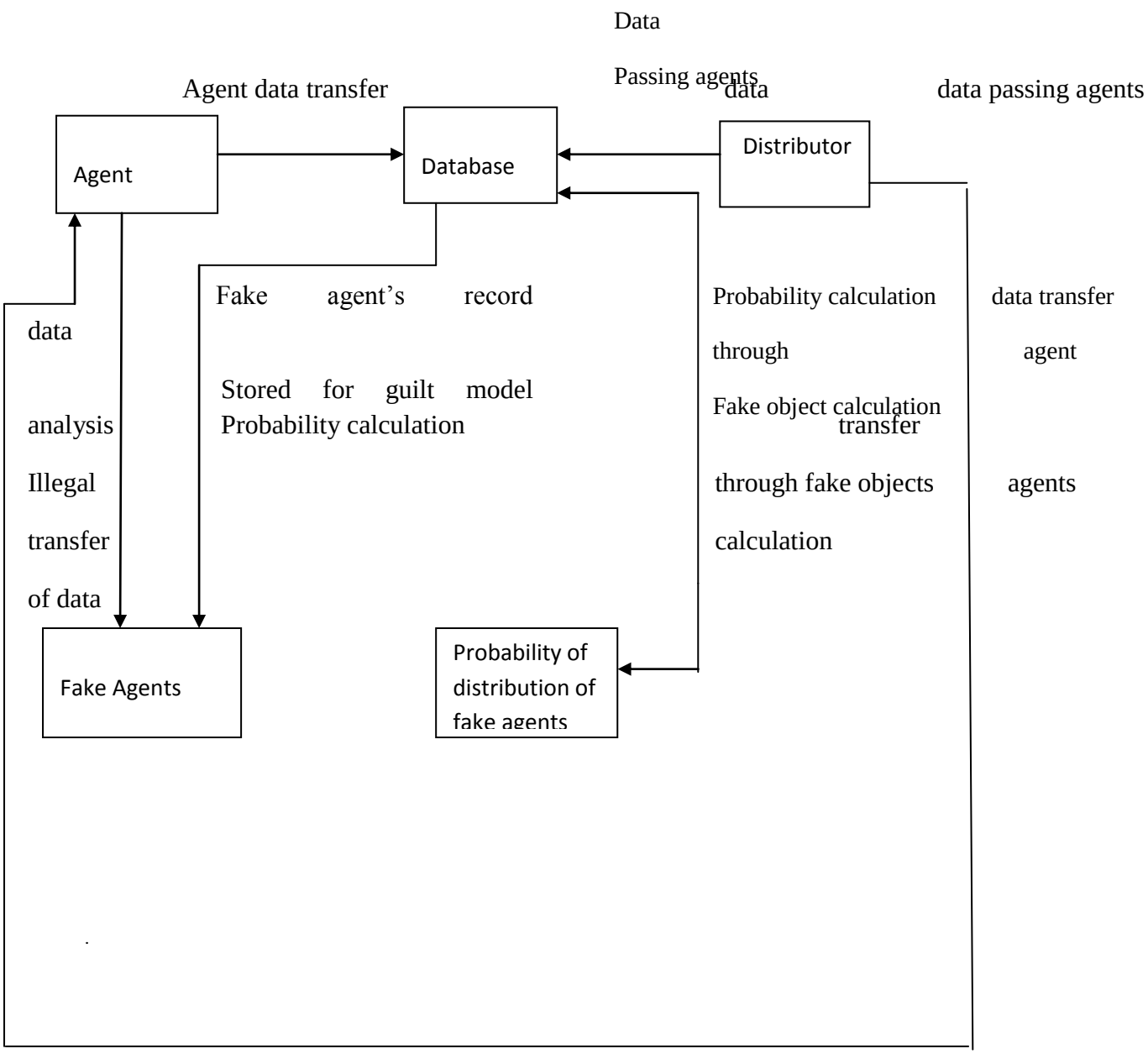
target for attackers. There is a more social and political involvements relate to the security of health data misuse. In order to reduce these problems the new system PIPE (Pseudonymization of information for privacy in e-Health) is introduced. Its consolidates primary and secondary usage data securely. It does not guarantee only promises on a superior level quality service for patients, and also decrease the cost for society in terms of social insurances. Majid Raissi-Dehkordi and David Carr [6], Insider threats plays an major role in categories of attacks to identify. Insider threats damages the security and results irreversible financial status in government. In order to overcome these problems a multi-perspective methedolgy for detection of insider threats were introduced in this project. This author focus to monitor the multiple detection engines in network movement from various perspectives and also use to combine the information to modify their detection sensitivities. This methodology reduces the false alarm probability and also improves the performance to finding the insider threats. Panagiotis Papadimitriou, Hector Garcia-molina[7], A distributor has a given set of sensitive data to a group of believable third party agents. The leaked data can be determines in any unauthorized places. The distributor must determine the probability that the data leaked from one or more agents, as against to independently collected by some other means. We introduce the data allocation techniques to improve the chance of determining leakages of data. This technique does not depend on any alterations of the published data (e.g. water marks). In the world there would not to be hand over the sensitive data to agents that may leaks unknowingly. If we had to hand over the sensitive data to a perfect world based on the watermarks. So we can trace its origin. In some cases we can't trust the agent 100%. In order to overcome these problems overlapping concept is introduced to identify data leakage.

Khanh viet, Brajendra panda, Yi Hu[8], The ambition of information security helps to enable the integrity, secretly and possibility of the data. Insider attacks are often checked by security techniques while compare to outsider threats. This author notifies the issue of collaborative insider attacks. The compromise of critical data in the system is occurred while two or more insiders work with each other. Collaborative attacks are very harder and complex to determine than insider attacks. The useful notation transaction's distance to data item, bridge data items, and mutual-access-records are specified to finding the malicious information and also used to determine single-step attacks. This method can achieve its target as less false positive rates for finding the collaborative insider attacks. C.Nithiyanandam, D.Tamilselvan, S.Balaji, V.Sivaguru[9], Preventing the misuse by third party agent is the hardest data protection challenge to rectify. IT security determines its unauthorized network is ineffective. A most powerful security tool is needed to defend against the insider threats. The introduced frame work provides a higher level security for intranets. Its main thing is detection prevention, preemption in virtual and physical environment both. Here we are going to locate the insider threat by observing and governing both user rights and user access to data based on the normal, anomalous and suspicious misbehavior. IDS concepts are used here to skip out insider threats over the distributed environment. IDS help us to conflict against the malicious misbehavior in an efficient manner.

4. OBJECTIVE

The main objective of this concept is to detect the data leakages and also identify where the leakage is located. Based on the misuseability concept data leakage can be prevented.

5. PROPOSED ARCHITECTURE



6. PROBLEM DEFINITION

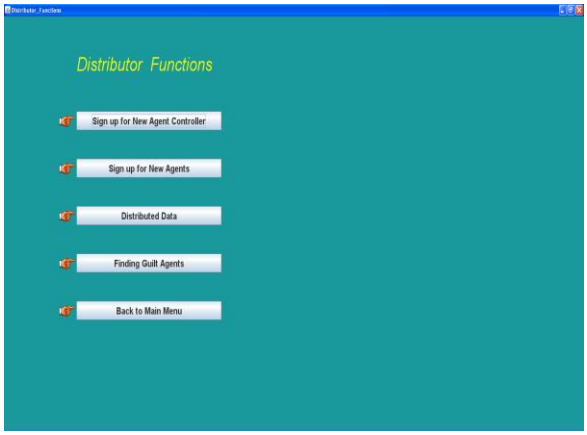
A distributor is owner of the data. It owns a set T= {t1…..tm} of valuable objects. The distributor likes to share the objects with a set of agents, but does not like to leak the objects to the third party agents. After giving a number of objects to agents, the distributor finds a set of similar objects in unauthorized location. At this moment the distributor can determine the probability of data leakage from one or more number of agents. The agent requires details of the objects from database. It should give request to the distributor. Based on the agents request the distributor can serve the data's. While

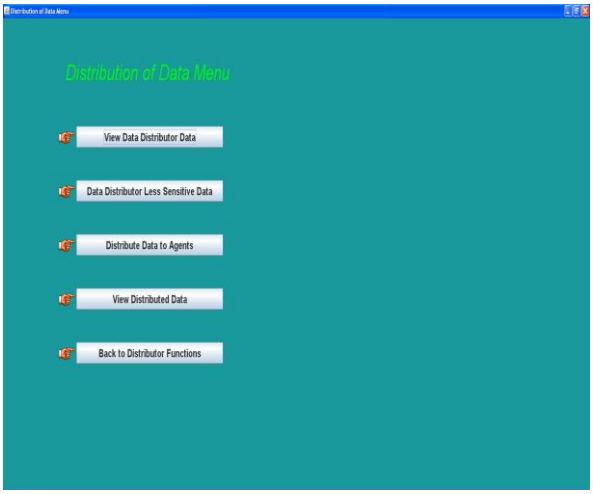
distributing the data's it will generate unique fake objects to each and every object provided to agents. Each copy of transferred data is stored in the distributor database. In order to enhance the effectiveness of finding the guilt agent, a fake object is added to the distributed data. The subset of objects can receive by an agent. If an agent leaks his data, it will be evenly guilt with a respected agent. The distributor's allocation has one objective and constraint. The distributor condition is to fulfill their agent requests, by distributing the set of objects agent requests. His goal is adept to find an agent who exposures the data.

Once the distributor discover his data in unauthorized place, he can able to made comparison between the released data and copy of its data, which is served to agents, then he can easily determine his guilt agent. Fake object is mainly used to determine the fault agents and along with its probability. Fake object should be unique for each and every agent and tuple. For computing this probability, they need to calculate the possibilities that the objects can be guessed or targeted one. Based on guessing probability he can determine the agents those who leaked the information. When the distributor allocates the data based on the agent requests, there may be a possibility of asking same object by more than one object. While providing same tuple to more than one agent an overlapping minimization takes place. A distributor can able to determine the guilt agent while data is releasing. Once the distributor finds his data in unauthorized places like website over internet, he can able to determine the guilt agent. For this distributor can calculate its probability based on the number of records found in unauthorized places. According to this probability we can conclude those who have more probability must be a guilt agent.

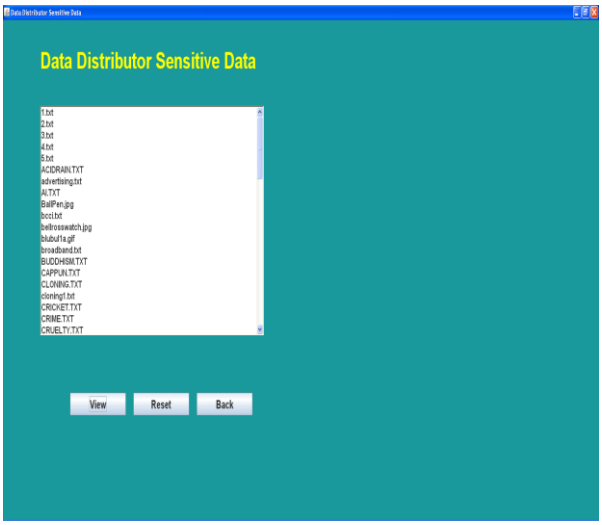
7. RESULT AND ANALYSIS

This section will discuss about the results.





.



The distributor, agent controller, and agents are the modules involved in this project. The distributor can distribute the data based on the agent needs. While sending the data to others the distributor embedded a secret key to each and every data. When comparing to other methods secret key should not be available. Based on the secret key we can either lock or unlock a file. A secret key should be confidential, so unauthorized persons can't access the file. With the help of M-score concept we can find the leakage data and those who leak the data as guilt agent. A misuseabilty weight measure helps us to detect the data leakage and also able to find the agent those who leaks the data. This concept provides a secured data in efficient manner.

There are three databases involved in this project and named as distributor database, agent database, and agent controller database.

Table.1 Distributor details

Column Name	Data Type	Length
Distributor Id	Text	5
Distributor Name	Text	50
Password	Text	50
E-mail	Text	50

The distributor database maintains the distributor details and it is owner of the data. . Based on the agents request the distributor can serve the data's.

Table 2.controller details

Column Name	Data Type	Length
Controller Id	Text	5
Controller Name	Text	50
Password	Text	50
E-mail	Text	50

The controller database maintains all the agent details respectively.

Table 3.Agent details

Column Name	Data Type	Length
Agent Id	Text	5
Agent Name	Text	50
Password	Text	50
E-mail	Text	50

The agent database maintains the agent files and also provides leaker files and agent details respectively.

8. CONCLUSION

In this world there would no need handover the sensitive data to agents unknowingly. The concept of misuseability weight examines the needs of sensitivity level of data measuring insider to expose out. Water marking is one of a method used to handle the data leakages traditionally, and for example each distributor copy has embedded with a unique code respectively. If that copy is detected in the place of an unauthorized party, the leaker can be detected. In certain cases watermarks can be needful, and sometimes can be deleted if the recipient data is malicious one. So that it M-score concept is used to detect and prevent the data leakage and also able to a find a guilt agent who leaks the data. M-score is used to enables the quality of data, quantity of data, and distinguishing factor. It can determine the misuseability weight for single publication without considering the prior knowledge. It can be a previous publication and knowledge on the description of the publication. Later we can extend the M-score to the multiple publication and the sensitive combination sensitive values.

9. ACKNOWLEDGMENTS

It is my pleasure to express with deep sense of gratitude to Thanapal.P, Assistant Professor (senior), School of Information Technology and Engineering, VIT University, for his constant guidance, continual encouragement, understanding, more than all, he taught me patience in my endeavor. My association with him is not confined to academics only, but it is a great opportunity on my part of work with him.

10.REFERENCES

[1] Amir Harel, Asaf Shabtai, Lior Rokach and Yuval Elovici, 2009, M-score: A Misuseability weight measure, IEEE Transactions on Dependable and Secure Computing.

[2**]** William Eberle and Lawrence Holder,2009, Mining for insider threats in business transaction and process, IEEE Transaction .

[3] Zhang Xiaosong, Liu Fei , Chen Ting, Li Hua,2009, Research and application of the transparent data encryption in intranet data leakage prevention, International conference on computational intelligence and security.

[4] Jiangjiang Wu, Jie Zhou, Jun Ma, Songzhu Mei, Jiangchun Ren, 2011, An active data leakage prevention model for insider threat, International symposium on intelligence information

[5] Bernhard Riedl, Veronika Grascher, Stefan Fenz, Thomas Neubauer, 2008, Pseudonymization for improving the privacy

in e-Health applications, Proceedings of the 41st Hawaii international conference on system sciences.

[6] Majid Raissi-Dehkordi and David Carr,20ll A multi-perspective approach to insider threat detection, The military communications conference-track 3-cyber security and network operations.

[7] Panagiotis Papadimitriou, Hector Garcia-Molina, 2005, A model for data leakage detection, IEEE international conference on data engineering.

[8] Khanh viet, Brajendra panda, Yi Hu, 2012 Detecting collaborative insider attacks in information systems, IEEE international conference on systems, man, and cybernetics.

[9] C.Nithiyandam, D.Tamilselvan, S.Balaji, V.Sivaguru, 2002, Advanced framework of defense system for prevention of insider's malicious behaviors.