IMPACT: International Journal of Research in Engineering & Technology (IMPACT: IJRET)
ISSN(E): 2321-8843; ISSN(P): 2347-4599
Vol. 2, Issue 4, Apr 2014, 75-82

M-SCORE: A MISUSEABILITY WEIGHT MEASURE

A. SARANYA

B.E., M. Tech, Department of Computer Science & Engineering, SRM University, Chennai, 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 an effort to determine the extent of damage to an organization that a user can cause using the information she has obtained. We introduce the concept of misuseability Weight. A misuseability weight measure, calculates a score that represents the sensitivity level of the data exposed to the user and by that predicts the ability of the user to maliciously exploit the data. By assigning a score that represents the sensitivity level of the data that a user is exposed to, the misuseability weight can determine the extent of damage to the organization if the data is misused.

KEYWORDS: Data Leakage, Data Misuse, Security Measures, Misuseability Weight

INTRODUCTION

A distributor has a given set of sensitive data to a number of hypothetical trusted third party agents. It 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. 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. Security-related data measures including k-Anonymity [4], l-Diversity [5], and (_, k)-Anonymity [6] are mainly used for privacy-preserving and are not relevant when the user has free access to the data. When data access is provided to insiders of organizations, the insiders may misuse it. Assuming that there is possibility of misuseability, it is essential to have a measure to know how sensitive the data exposed is.

To overcome this problem certain mechanisms to prevent data misuse and data leakage is essential. However, finding the people who commit such malicious activities on the accessed data is very challenging for many reasons. Moreover there are incidents in the history that proved that insiders misuse data for some reasons including monetary gains. In order to detect the damage misuseability is used, which assigns a sensitivity score to data sets, thereby estimating the level of harm that might be inflicted upon the organization when the data is leaked. "Misuseability measure" is to know the misuseability score of the data being exposed to insiders. This will reduce the misuse of leakage of data which is made available to the insiders. User behavior profiles are used to devise plans to mitigate such fraud. When profiles are observed,

it can be understood that the normal user's behavior is obviously different from that of the malicious user who has malicious intentions. User behavior can be analyzed using SQL commands and other features provided by the query languages. The misuseability measure is also computed based on three main factors as well. They include quality of data, quantity of data, and the distinguishing factor. We extend the concept of Measurability weight to support multiple publications with more than one distinguishing factor and sensitivity of combinations of sensitive values.

RELATED WORK

Misuse Detection in Databases

Despite the necessity of protecting information stored in database systems (DBS), existing security models are insufficient to prevent misuse, especially insider abuse by legitimate users. Almost all systems all over the world suffer from outsider and insider attacks. Outsider attacks are those that come from outside the system, however, insider attacks are those that are launched from insiders of the system. Misuse Detection- based on signatures that demonstrate the characteristics of well-known system vulnerabilities and attacks. It works well with known misuse patterns but fails with new ones, Anomaly Detection- based on the behavior of the matter, e.g., user, application, or component of a system. It's the most popular way of detection. It is better than the misuse detection methodology because it has a better opportunity to detect previously unknown attacks and Insider Misuse- sources ranging from discontent employee (database administrators, application developers, application users), who may maliciously damage the data integrity to outsider that gain access to the data. Two approaches to misuse of data.1. Syntax centric 2. Data centric. In syntax centric data request are analyzed to detect misuse. In Data centric actual data accessed is analyzed to detect misuse.

Misuseability Weight Concept

The misusability measure is also computed based on three main factors as well. They include quality of data, quantity of data, and the distinguishing factor.

The four dimensions are:

Number of Entities: This is the data size with respect to the different entities that appear in the data. Having data about more entities obviously increase the potential damage as a result of a misuse of this data.

Anonymity Level: While the number of different entities in the data can increase the misuseability weight, the anonymity level of the data can decrease it. The anonymity level is regarded as the effort that is required in order to fully identify a specific entity in the data.

Number of Properties: Data can include a variety of details, or properties, on each entity (e.g., employee salary or patient disease). Since each additional property can increase the damage as a result of a misuse, the number of different properties (i.e., amount of information on each entity) should affect the misuseability weight**.**

Values of Properties: The property value of an entity can greatly affect the misuseability level of the data.

THE M-SCORE MEASURE

To measure the misuseability weight, we propose a new algorithm—the M-score. This algorithm considers and measures different aspects related to the misuseability of the data in order to indicate the true level of damage that can result if an organization's data falls into wrong hands. The M-score measure is tailored for tabular data sets

(e.g., result sets of relational database queries) and cannot be applied to non tabular data such as intellectual property, business plans, etc. It is a domain independent measure that assigns a score, which represents the misuseability weight of each table exposed to the user, by using a sensitivity score function acquired from the domain expert. By assigning a score that represents the sensitivity level of the data that a user is exposed to, the misuseability weight can determine the extent of damage to the organization if the data is misused. We define three, nonintersecting types of attributes: quasi-identifier attribute, sensitive attributes; and other attributes.

- **Quasi Identifier**

Quasi -identifier are attributes that may be linked, possibly using an external data source, to reveal a particular entity that the specific information is about.

- **Sensitive Attribute**

Sensitive attribute are attributes that are used to evaluate the risk derived from exposing the data. The sensitive attributes are mutually excluded from the quasi-identifier attributes.

K-ANONYMITY

The k-anonymity privacy requirement for publishing micro data requires that each equivalence class (i.e., a set of records that are indistinguishable from each other with respect to certain "identifying" attributes) contains at least k records. The concept of k-anonymity tries to capture, on the private table PT to be released, one of main requirements that has been followed by the statistical community and by agencies releasing the data, and referring to which the released data should be indistinguishably related to no less than a certain number of respondents. The set of attributes included in the private table, also externally available and thus exploitable for linking, is called quasi-identifier.

This requirement just stated is then translated in the k-anonymity requirement below, which states that every tuple released cannot be related to fewer than k respondents. Each release a data must be such that every combination of values of quas-identifiers can be indistinctly matched to at least k respondents. Since it seems impossible, or highly impractical and minimizing, to make assumptions on the datasets available for linking to external attackers or curious data recipients, essentially k-anonymity takes a safe approach requiring that, in this table itself, the respondents be indistinguishable (within a given set) with respect to the set of attributes.

CALCULATING M-SCORE

We have built a framework to visualize the process of measuring M-Score for the given dataset. The framework is as shown in figure 1. First of all the given dataset is given to a module which computes raw record score. Sensitivity score function is used to computer raw record score. Afterwards the record distinguishing factor (DF) is computed. The DF is the measure to know what extent the quasi identifiers can reveal the identity of a record. Once record score is computed, the values are substituted in the formula of M-score. The computation of M-Score is as follows

$$\text{M-SCORE}=r^{1/X}*RS=r^{1/X}*MAX\ (RRS^{i}/D^{i})$$

Where RRS is the raw record source

S – Sensitivity attribute

D – Distinguish factor

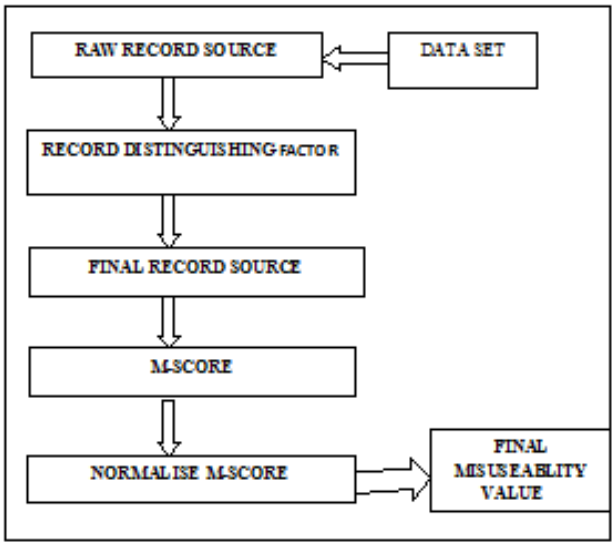

Figure 1: Architecture for Measuring M-Score

ARCHITECTURE FOR MISUSEABLITY WEIGHT MEASURE

Organization's data is extremely important and proves as a main constituent in embodying the core of the organization's power. Organization definitely wants that to this power should be preserved and retained. On the other side, this data is required for daily working on different processes. Data leakage, especially caused by insider threat is one of the important issue in information security research. This is because insider threats have potential to inflict severe damage to the organization's resources, financial assets and reputation.

Misuseability weight concept is very useful for data leakage and misuse detection since it measures the harm or the risk of damage that can be caused if the important data falls into the wrong hands. It assigns a score which estimates the sensitivity level of damage so that security measures can be applied in order to avoid data leakage and misuse detection. For this, user profile is created using the local knowledge base so that only required data can be given for avoiding data leakage and misuse. The architecture is shown in figure 2.

It is observed that though many of the suggested methods contributed a lot towards the data leakage and misuse detection, there is a need of the system which will give more efficient and improved result. Thus proposed plan of work involves following steps which will try to resolve the problems occurred in the previous system. The very first step of the system is making a user profile which is based on the information submitted by the employee as well as the kind of work done by him.

On the admin login, administrator is able to see the list as well as the complete information of all these employees. In the next step, tabular data or the database is analyzed for different dimensions of misuseability. This proves very helpful for calculating sensitivity level of the damage which is done in the next step. Measuring the sensitivity level of damage involves collection of information from the domain expert. Based on this, sensitive attributes are selected which help in calculating the sensitivity score as well as the misuseability score.

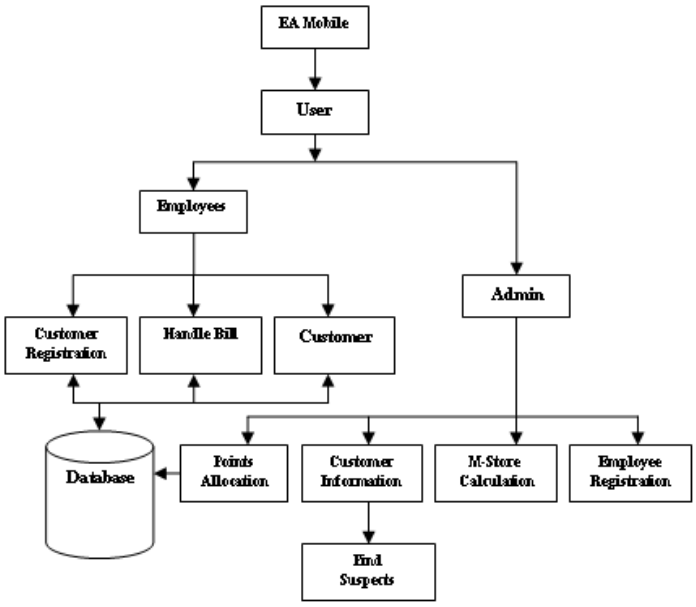

Figure 2: Architecture for Misuseablity of Data

METHODOLOGY FOR DATA LEAKAGE AND MISUSE DETECTION

Detection of an insider threat which is which mainly responsible for the data leakage and misuse, typically construct the normal user profile for each user by recording appropriate usage metrics. It is taken from the two categories.

- Host
- Network

In pure host monitoring, daily actions of the user on his/her machine are logged and the user behavior is modeled by using metrics such as the sequences of the OS commands, applications used by the user. Another way that is network monitoring focuses more on network aspect of user actions, such as database queries, file system access, printer access etc. Despite the differences among the above methods, some form of profile is constructed for each user and that profile is used as the baseline for monitoring data leakage and misuse.

EVALUATION

Insiders

The insider threat against database management systems is a dangerous security problem. Authorized users may abuse legitimate privileges to masquerade as other users or to maliciously harvest data. We model user's access patterns by profiling the *data points* that users access, in contrast to analyzing the *query expressions* in prior approaches. Our data-centric approach is based on the key observation that query syntax alone is a poor discriminator of user intent, which is much better rendered by *what* s accessed. We present a feature-extraction method to model user's access patterns.

Data-Centric User Profiles

A relational database often consists of multiple relations with attributes and relationships specified by multiple *primary key* and *foreign key* constraints. One can visualize a database as a single relation, called the *Universal Relation.* In corporating the attribute information from all the relations in the database. We compute a statistical "summary" of the

query's result tuple. The summary for a query is represented by a vector of fixed dimension regardless of how large the query's result tuple set is. This way, past queries (i.e. normal queries) from a user can be intuitively thought of as a "cluster" in some high dimensional space.

M-Score-Based Anomaly Detection

Anomaly detection involves detecting statistically significant deviations of test data from nominal distribution. In typical applications the nominal distribution is unknown and generally cannot be reliably estimated from nominal training data due to a combination of factors such as limited data size and high dimensionality. A different usage scenario arises in implementing M-score based anomaly detection. During the learning phase, the normal behavior of each user or role is extracted. The normal behavior represents the sensitivity level of the data to which users are exposed, during their regular activity within different contexts (e.g., time of day, location). During the detection phase, the M-score of each action is computed and validated against the behavioral model that was derived in the learning phase.

CONCLUSIONS AND FUTURE WORK

Thus, this paper introduces a system which will measure the risk of damage that can be caused when data is exposed to the insider. This involves collecting knowledge from the domain expert as well as use of the risk measuring algorithm. Measuring risk before data exposure will help administrator to take proper action to prevent or minimize the damage. Consequently, a new misuseability measure, the M-score, was proposed. By assigning a score that represents the sensitivity level of the data that a user is exposed to, the misuseability weight can determine the extent of damage to the organization if the data is misused. We extended the M-score basic definition. Efficiently acquiring the knowledge required for computing the M-score, and showed that the M-score is both feasible and can fulfill its main goals.

REFERENCES

1. 2010 Cyber Security Watch Survey, http://www.cert.org/archive/pdf/ecrimesummary10.pdf, 2012.
2. A. Kamra, E. Terzi, and E. Bertino, "Detecting Anomalous Access Patterns in Relational Databases," Int'l J. Very Large Databases, vol. 17, no. 5, pp. 1063-1077, 2008.
3. S. Mathew, M. Petropoulos, H.Q. Ngo, and S. Upadhyaya, "Data-Centric Approach to Insider Attack Detection in Database Systems," Proc. 13th Conf. Recent Advances in Intrusion Detection, 2010.
4. L. Sweeney, "k-Anonymity: A Model for Protecting Privacy," Int'lJ. Uncertainty, Fuzziness and Knowledge Based Systems, vol. 10, no. 5, pp. 571-588, 2002.
5. A. Machanavajjhala et al., "L-Diversity: Privacy beyond K-Anonymity," ACM Trans. Knowledge Discovery from Data, vol.1, no.1, article 1, 2007.
6. R.C. Wong, L. Jiuyong, A.W. Fu, and W. Ke, "(_, k)-Anonymity: An Enhanced k-Anonymity Model for Privacy-Preserving Data Publishing," Proc. 12th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, 2006.
7. Bishop, M.: The insider problem revisited. In: Proc. of the 2005 Workshop on New Security Paradigms (NSPW 2005), pp. 75–76 (2005).

8. Brackney, R., Anderson, R.: Understanding the Insider Threat: Proceedings of a March 2004 Workshop. RAND Corp. (2004).

9. Buneman, P., Khanna, S., Tan, W.C.: Why and where: A characterization of data provenance. In: ICDT, pp. 316–330 (2001).

10. Haas, P.J., Hellerstein, J. M.: Ripple joins for online aggregation. In: SIGMOD Conference, pp. 287–298 (1999).

11. Hu, Y., Panda, B.: Identification of malicious transactions in database systems. In: Proc. of the 7th International Database Engineering and Applications Symposium, pp. 329–335 (2003).

12. Kamra, A., Terzi, E., Bertino, E: Detecting anomalous access patterns in relational databases. The VLDB Journal 17(5), 1063–1077 (2008).