A Ranking Based Approach in Security Alert for Handling Sensitive Data Organization

Tania Veronica Sebastian & J. Raja
Department of Computer Science and Engineering Annai Mathammal Sheela Engineering College Namakkal

Abstract – An organization undoubtedly wants to preserve and retain data stored in an organization’s computers. On the other hand, this data is necessary for daily work processes. Users within the organization’s perimeter (e.g., employees, subcontractors, or partners) perform various actions on this data (e.g., query, report, and search) and may be exposed to sensitive information embodied within the data they access. In an effort to determine the extent of damage to an organization that a user can cause using the information obtained, the concept of a ranking based approach in security alert for handling sensitive data organization is introduced. The 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. By assigning a score that represents the sensitivity level of the data that a user is exposed to, the weight can determine the extent of damage to the organization if the data is misused. Using this information, the organization can then take appropriate steps to prevent or minimize the damage.

Keywords – organization, score, sensitive data, user.

I. INTRODUCTION

A ranking based approach in security alert for handling sensitive data organization 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. Data stored in an organization’s computers is extremely important and expresses the core of the organization’s power. An organization undoubtedly wants to preserve and retain this power. On the other hand, this data is necessary for daily work processes.

There is no previously proposed method for estimating the potential harm that might be caused by misused or leaked data while considering important dimensions of the nature of the exposed data is the problem statement. The main assets of an organization are its sensitive information such as customer or patient data and business secrets constitute. This information is essential for the employees of an organization, subcontractors, or partners to perform their tasks. On the other hand if we are if the access to such type of information are limited for preserving secrecy might damage their ability to implement the actions that can best serve the organization.

Thus, there are two mechanisms that are essential in identifying malicious insiders and they are data leakage and data misuse detection mechanisms. The methods of deception become more and more sophisticate and so the task of detecting malicious insiders is very challenging. According to the 2010 Cyber Security Watch Survey 26 percent of the cyber-security events, recorded in a 12-month period, were caused by insiders. These insiders were the most damaging with 43 percent of the respondents reporting that their organization suffered data loss. Of the attacks, 16 percent were caused by theft of sensitive data and 15 percent by exposure of confidential data. The focus is on mitigating leakage or misuse incidents of data stored in databases (i.e., tabular data) by an insider having legitimate privileges to access the data. There have been numerous attempts to deal with the malicious insider scenario. The methods that have been devised are generally based on user behavioral profiles that define normal user two behaviors and issue an alert whenever a user’s behavior significantly deviates from the normal profile. The most common approach for representing user behavioral profiles is by analyzing the SQL statement submitted by an application server to the database (as a result of user requests),and extracting various features from these SQL statements. Another approach focuses on analyzing the actual data exposed to the user, i.e., the result-sets. However, none of the proposed methods consider the different sensitivity levels of the data to which an insider is exposed. This factor has a great impact in estimating the damage that can be caused to an organization when data is leaked or misused. Security-related data measures including k-Anonymity, l-Diversity, and (α, k)-Anonymity are mainly used for privacy-preserving and are not relevant when the user has free access to the data. Therefore, a ranking based approach in security
alert for handling sensitive data organization 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.

Customer or patient data and business secrets constitute the main assets of an organization. These are some examples of sensitive information. Such sensitive information is essential for the sub-contractors, organization's employees, or partners to perform their tasks. Limiting access to this type of information for preserving secrecy might lead to damage the organization’s ability to implement the actions that can best serve the organization as well as its customers.

In addition to traditional access control mechanisms the use of data leakage and data misuse detection mechanisms is essential and can also help in identifying malicious insiders which might risk the organization. The task of detecting malicious insiders becomes very challenging as the methods of deception become more and more sophisticated. A ranking based approach in security alert for handling sensitive data organization is used to assign a sensitivity score to datasets, thereby estimating the level of harm that might be inflicted upon the organization when the data is leaked. Consequently, ranking based approach in security alert for handling sensitive data organization can be used for: (1) learning the normal behavior of an insider in terms of the sensitivity level of the data she is usually exposed to by applying anomaly detection; (2) improving the process of handling leakage incidents identified by other misuse detection systems by enabling the security officer to focus on incidents involving more sensitive data; (3) designed to regulate user access to sensitive data stored in relational databases to implement a dynamic misuseability-based access control; and (4) reducing the misuseability of the data proactively.

The existing method check the table satisfies the k-anonymity that is whether the table appears for at least k-times. The differential privacy ensure that statistical (or aggregation) queries can be executed on a database with high accuracy while preserving the privacy of the entities in the database. The data-centric approach focuses on what the user is trying to access instead of how she expresses it. With this approach, an action is modeled by extracting features from the obtained result-set.

Proposed system developed a model for assessing the “guilt” of agents. The system also present algorithms for distributing objects to agents, in a way that improves the chances of identifying a leaker. Finally, it also considers the option of adding “fake” objects to the distributed set. Such objects do not correspond to real entities but appear realistic to the agents. In a sense, the fake objects acts as a type of ranking based approach in security alert for handling sensitive data organization for the entire set, without modifying any individual members. If it turns out an agent was given one or more fake objects that were leaked, then the distributor can be more confident that agent was guilty.

In proposed system, a ranking based approach in security alert is used for handling sensitive data organization, for estimating the risk emanating from data exposed to insiders. This concept focuses on assigning 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 this data. It assigns a weight to tabular data, discuss some of its properties, and demonstrate its usefulness in several leakage scenarios.

II. SYSTEM MODEL

A data distributor has given sensitive data to a set of supposedly trusted agents (third parties). Some of the data is leaked and found in an unauthorized place (e.g., on the web or somebody’s laptop). The distributor must assess the likelihood that the leaked data came from one or more agents, as opposed to having been independently gathered by other means. Data allocation strategies (across the agents) that improve the probability of identifying leakages is proposed. These methods do not rely on alterations of the released data. In some cases we can also inject “realistic but fake” data records to further improve our chances of detecting leakage and identifying the guilty party. The goal is to detect when the distributor’s sensitive data has been leaked by agents, and if possible to identify the agent that leaked the data.

The main focus of the project is the data allocation problem as how can the distributor “intelligently” give data to agents in order to improve the chances of detecting a guilty agent.

Fake objects are objects generated by the distributor in order to increase the chances of detecting agents that leak data. The distributor may be able to add fake objects to the distributed data in order to improve his effectiveness in detecting guilty agents. Our use of fake objects is inspired by the use of “trace” records in mailing lists.

The Optimization Module is the distributor’s data allocation to agents has one constraint and one objective. The distributor’s constraint is to satisfy agents’ requests, by providing them with the number of objects they request or with all available objects that satisfy their conditions. His objective is to be able to detect an agent who leaks any portion of his data. In this process we are assigning the weight for ach data.

A data distributor has given sensitive data to a set of supposedly trusted agents (third parties). Some of the data is leaked and found in an unauthorized place (e.g., on the web or somebody’s laptop). The distributor must assess the likelihood that the leaked data came from one or more agents, as opposed to having been independently gathered by other means. Based on the weight of the data we can find the sensitive of the information that is leaked.
III. ALGORITHMS

A. Explicit Data Request

In case of explicit data request with fake not allowed, the distributor is not allowed to add fake objects to the distributed data. So data allocation is fully defined by the agent’s data request. In case of explicit data request with fake allowed, the distributor cannot remove or alter the requests R from the agent. However distributor can add the fake object. In algorithm for data allocation for explicit request, the input to this is a set of request, from n agents and different conditions for requests. The e-optimal algorithm finds the agents that are eligible to receiving fake objects. Then create one fake object in iteration and allocate it to the agent selected. The e-optimal algorithm minimizes every term of the objective summation by adding maximum number of fake objects to every set yielding optimal solution.

Step 1: Calculate total fake records as sum of fake records allowed.
Step 2: While total fake objects > 0
Step 3: Select agent that will yield the greatest improvement in the sum objective
Step 4: Create fake record
Step 5: Add this fake record to the agent and also to fake record set.
Step 6: Decrement fake record from total fake record set.

Algorithm makes a greedy choice by selecting the agent that will yield the greatest improvement in the sum-objective.

B. Sample Data Request

With sample data requests, each agent may receive any T from a subset out of different ones. Hence, there are different allocations. In every allocation, the distributor can permute T objects and keep the same chances of guilty agent detection. The reason is that the guilt probability depends only on which agents have received the leaked objects and not on the identity of the leaked objects. Therefore, from the distributor’s perspective there are different allocations. An object allocation that satisfies requests and ignores the distributor’s objective is to give each agent a unique subset of T of size m. The s-max algorithm allocates to an agent the data record that yields the minimum increase of the maximum relative overlap among any pair of agents. The s-max algorithm is as follows.

Step 1: Initialize Min_overlap ← 1, the minimum out of the maximum relative overlaps that the allocations of different objects to
Step 2: for k ∈ {k | ∈ } do Initialize max_rel_ov ← 0, the maximum relative overlap between and any set that the allocation of to
Step 3: for all j = 1,..., n : j = i and ∈ do Calculate absolute overlap as abs_ov ← | ∩ | + 1 Calculate relative overlap as rel_ov ← abs_ov / min ( , )
Step 4: Find maximum relative as max_rel_ov ← MAX (max_rel_ov, rel_ov) If max_rel_ov ≤ min_overlap then min_overlap ← max_rel_ov ret_k ← k
Return ret_k

It can be shown that algorithm s-max is optimal for the sum-objective and the max-objective in problems where M ≤ |T| and n < |T|. It is also optimal for the max objective if |T| ≤ M ≤ 2 |T| or all agents request data of the same size.

It is observed that the relative performance of algorithm and main conclusion do not change. If p approaches to 0, it becomes easier to find guilty agents and algorithm performance converges. On the other hand, if p approaches 1, the relative differences among algorithms grow since more evidence is needed to find an agent guilty. The algorithm presented implements a variety of data distribution strategies that can improve the distributor’s chances of identifying a leaker. It is shown that distributing objects judiciously can make a significant difference in identifying guilty agents, especially in cases where there is large overlap in the data that agents must receive.

A ranking based approach in security alert for handling sensitive data organization proposed a score calculating the sensitivity level of the data exposed to the user. Proposed system have different approaches for efficiently acquiring the knowledge required for computing the score, and the score is both feasible and can fulfill the main goal for estimating the user. This method is very useful for protecting both individual data and statistical information.

IV. CONCLUSION

A ranking based approach in security alert for handling sensitive data organization definition was extended to consider prior knowledge the user might have and presented four applications using the extended definition. There are two important issues, which relate to the knowledge elicitation and representation: the temporal aspect of the score and the validity of the knowledge, acquired from the experts, over time; and the knowledge acquisition that might be subjective and not consistent among different experts which, in turn, may lead to an inaccurate sensitivity function. In regards to the time factor the sensitivity level of an attribute’s value will change in rare cases and especially the order of the values with respect to their sensitivity level. For example, the value of a gold customer will not change and will remain more sensitive than a silver customer. A customer’s type may change from gold to silver and this will be reflected when computing the score of the customer’s record. However, the need to validate and reacquire the knowledge from time to time, and the knowledge can be acquired accurately with relatively minimal effort (in terms of experts time) using the pair.
wise comparison approach, to explore methods for incremental learning, or post learning fine tuning of the elicited sensitivity score function in future work. With respect to the subjectivity of the elicited scoring function, experiments indicate that the methods used ensure that the acquired knowledge is not biased. In fact, using knowledge acquired from one expert is sufficient in order to calculate sound scores for the entire domain. In some cases the value of customers can be calculated by using known knowledge on the customer (e.g., how much she spends) and by predicting future revenue from the customer. In such cases, the sensitivity level of sensitive attributes can be objectively obtained by using machine learning techniques; in particular by fitting the sensitive parameter values to the customer value.

V. REFERENCES

[1] A. Barth, A. Datta, J.C. Mitchell, and H. Nissenbaum, “Privacy and Contextual Integrity: Framework and Applications,” Proc. IEEE Symp. Security and Privacy, pp. 184-198, 2006.

[2] Amir Harel, Asaf Shabtai, Lior Rokach, and Yuval Elovici,” M-Score: A Misuseability Weight Measure” IEEE transactions on dependable and secure computing, vol. 9, no. 3, May/June 2012

[3] A. Machanavajjhala et al., “L-Diversity: Privacy Beyond K-Anonymity,” ACM Trans. Knowledge Discovery from Data, vol.1, no.1, article 1, 2007.

[4] B. Carminati, E. Ferrari, J. Cao, and K. Lee Tan, “A Framework to Enforce Access Control over Data Streams,” ACM Trans. Information Systems Security, vol. 13, no. 3, pp. 1-31, 2010.

[5] E. Bertino and R. Sandhu, “Database Security-Concepts, Approaches, and Challenges,” IEEE Trans. Dependable and Secure Computing, vol. 2, no. 1, pp. 2-19, Jan.-Mar. 2005

[6] E. Celikel et al., “A Risk Management Approach to RBAC,” Risk and Decision Analysis, vol. 1, no. 2, pp. 21-33, 2009.

[7] 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.

[8] L. Sweeney, “k-Anonymity: A Model for Protecting Privacy,” Int’l J. Uncertainty, Fuzziness and Knowledge Based Systems, vol. 10, no. 5, pp. 571-588, 2002.

[9] Q. Yaseen and B. Panda, “Knowledge Acquisition and Insider Threat Prediction in Relational Database Systems,” Proc. Int’l Conf. Computational Science and Eng., pp. 450-455, 2009.

[10] R.C. Wong, L. Jiuyong, A.W. Fu, and W. Ke, “(α, 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.

[11] 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.

[12] 2010 CyberSecurity Watch Survey, http://www.cert.org/archive/pdf/ecrimesummary10.pdf, 2012.