ISSN(Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

M-Score: A Misusability Weight Measure for Insider Attack Detection in DBMS

A.Aruna[1], M.Sayee kumar[2,] T.Aravind[3]

ME-CSE (Final year), Muthayammal Engineering College, Rasipuram, India[1]

Assistant Professor, Muthayammal Engineering College, Rasipuram, India[2]

Assistant Professor, Muthayammal Engineering College, Rasipuram, India[3]

ABSTRACT: Data stored in an organization's computers is extremely important and embodies 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. 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 she has obtained, we introduce the concept of Misuseability Weight.

The M-score measure is tailored for tabular data sets (e.g., result sets of relational database queries) and cannot be applied to no tabular data such as intellectual property, business plans, etc. It is a domain independent measure that assigns a score, which represents the misuse ability 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. Using this information, the organization can then take appropriate steps to prevent or minimize the damage.

I.INTRODUCTION

Relational database management systems (RDBMS) are the fundamental means of data organization, storage and access in most organizations, services, and applications. Naturally, the ubiquity of RDBMSs led to the prevalence of security threats against these systems. An intruder from the outside, for example, may be able to gain unauthorized access to data by sending carefully crafted queries to a back-end database of a Web application. An insider attack against an RDBMS, however, is much more difficult tdetect, and potentially much more dangerous. By definition, detecting insider attacks by specifying explicit rules or policies is a moot point: an insider is always defined relative to a set of policies. Consequently, we believe that the most effective method to deal with the insider threat problem is to statistically profile normal users' (computing) behaviors and raise a flag when a user deviates from his/her routine. Intuitively, a good statisticalprofiler should be able to detect non-stealthy sabotage attacks, quick data harvesting attacks, or masquerading attacks, because the computing footprints of those actions should be significantly different from day-to-day activities, from a statistical point of view.

II.RELATED WORKS

2.1 Database Security—Concepts, Approaches, And Challenges

As organizations increase their reliance on, possibly distributed, information systems for daily business, they become more vulnerable to security breaches even as they gain productivity and efficiency advantages. Though a number of techniques, such as encryption and electronic signatures, are currently available to protect data when transmitted across sites, a truly comprehensive approach for data protection must also include mechanisms for enforcing access control policies based on data contents, subject qualifications and characteristics, and other relevant contextual information, such as time. It is well understood today that the semantics of data must be taken into account in order to specify effective access control policies.

ISSN(Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

2.2 Knowledge Acquisition And Insider Threat Prediction In Relational Database System

This paper investigates the problem of knowledge acquisition by an unauthorized insider using dependencies between objects in relational databases. It defines various types of knowledge. In addition, it introduces the Neural Dependency and Inference Graph (NDIG), which shows dependencies among objects and the amount of knowledge that can be inferred about them using dependency relationships. Moreover, it introduces an algorithm to determine the knowledgebase of an insider and explains how insiders can broaden their knowledge about various relational database objects to which they lack appropriate access privileges. In addition, it demonstrates how NDIGs and knowledge graphs help in assessment of insider threats and what security officers can do to avoid such threats.

2.3. A Security Punctuation Framework For Enforcing Access Control On Streaming Data

`The management of privacy and security in the context of data stream management systems (DSMS) remains largely an unaddressed problem to date. Unlike in traditional DBMSs where access control policies are persistently stored on the server and tend to remain stable, in streaming applications the contexts and with them the access control policies on the real-time data may rapidly change. A person entering a casino may want to immediately block others from knowing his current whereabouts. We thus propose a novel "stream-centric" approach, where security restrictions are not persistently stored on the DSMS server, but rather streamed together with the data.

Here, the access control policies are expressed via security constraints (called security punctuations, or short, sps) and are embedded into data streams. The advantages of the sp model include flexibility, dynamicity and speed of enforcement. DSMSs can adapt to not only data-related but also security-related selectivity's, which helps reduce the waste of resources, when few subjects have access to data. We propose a security-aware query algebra and new equivalence rules together with cost estimations to guide the security-aware query plan optimization. We have implemented the sp framework in a real DSMS. Our experimental results show the validity and the performance advantages of our sp model as compared to alternative access control enforcement solutions for DSMSs.

2.4. Evolution Of Privacy-Preserving Data Publishing

To achieve privacy protection better in data publishing, data must be sanitized before release. Research on protecting individual privacy and data confidentiality has received contributions from many fields. In order to grasp the development of privacy preserving data publishing, we discussed the evolution of this theme, focused on privacy mechanism, data utility and its metrics. The privacy mechanism, such as k anonymity ,l-diversity and t-closeness, provides formal safety guarantees and data utility preserve useful information while publishing data. Meantime, we discussed social network privacy and location based service. Finally, we made a conclusion with respect to privacy preserving data publishing, and given further research directions.

Fake objects are objects generated by the distributor that are not in set T. The objects are designed to look like real objects, and are distributed to agents together with T objects, in order to increase the chances of detecting agents that leak data.

Fake Objects

The distributor may be able to add fake objects to the distributed data in order to improve his effectiveness in detecting guilty agents. However, fake objects may impact the correctness of what agents do, so they may not always be allowable.

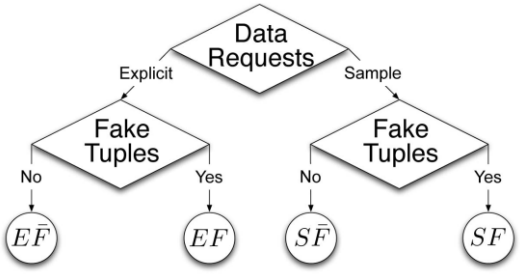


Fig 1:Classification of fake tuples

ISSN(Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

In some applications, fake objects may cause fewer problems that perturbing real objects. For example, say that the distributed data objects are medical records and the agents are hospitals. In this case, even small modifications to the records of actual patients may be undesirable. However, the addition of some fake medical records may be acceptable, since no patient matches these records, and hence, no one will ever be treated based on fake records.

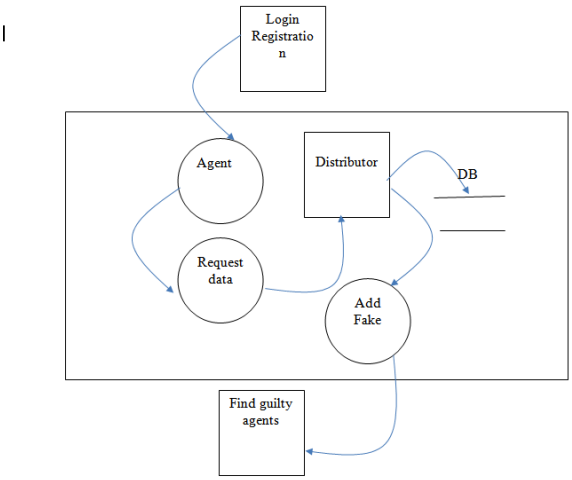
Our use of fake objects is inspired by the use of "trace" records in mailing lists. In this case, company A sells to company B a mailing list to be used once (e.g., to send advertisements). Company A adds trace records that contain addresses owned by company A. Thus, each time company B uses the purchased mailing list, A receives copies of the mailing. These records are a type of fake objects that help identify improper use of data. The distributor creates and adds fake objects to the data that he distributes to agents. We let Fi _ Ri be the subset of fake objects that agent Ui receives. As discussed below, fake objects must be created carefully so that agents cannot distinguish them from real objects.

In many cases, the distributor may be limited in how many fake objects he can create. For example, objects may contain e-mail addresses, and each fake e-mail address may require the creation of an actual inbox (otherwise, the agent may discover that the object is fake). The inboxes can actually be monitored by the distributor: if e-mail is received from someone other than the agent who was given the address, it is evident that the address was leaked. Since creating and monitoring e-mail accounts consumes resources, the distributor may have a limit of fake objects. If there is a limit, we denote it by B fake objects.

III.PROPOSED METHODOLOGY

This system present a new concept, Misuseability Weight, 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 misuseability weight to tabular data, discuss some of its properties, and demonstrate its usefulness in several leakage scenarios.There is no previously proposed method for estimating the potential harm that might be caused by leaked or misused data while considering important dimensions of the nature of the exposed data. Only our proposed one for calculating M-score, can solve the above problems.In proposed system have different approaches for efficiently acquiring the knowledge required for computing the M-score, and the M-score is both feasible and can fulfill the main goal for estimating the user.This M-score method is very useful for protecting both individual data and statistical information.

Level 0:



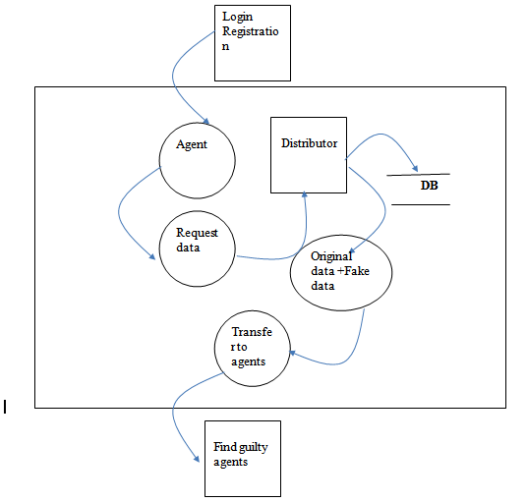
ISSN(Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

Level 1:



IV .MODULE DESCRIPTION

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. We propose 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 we can also inject "realistic but fake" data records to further improve our chances of detecting leakage and identifying the guilty party. Our goal is to detectwhen the distributor's sensitive data has been leaked by agents, and if possible to identify the agent that leaked the data.

4.1.Data Allocation Module:

The main focus of our 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.

4.2.Fake Object Module:

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.

4.3. Optimization Module:

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.

4.4. Data Distributor:

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.

ISSN(Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

V.CONCLUSION

The system introduced a new concept of misuseability weight anddiscussed the importance of measuring the sensitivity levelof the data that an insider is exposed to. We defined fourdimensions that a misuseability weight measure mustconsider. To the best of our knowledge and based on theliterature survey we conducted, there is no previouslyproposed method for estimating the potential harm thatmight be caused by leaked or misused data whileconsidering important dimensions of the nature of theexposed data. Consequently, a new misuseabilitymeasure,the M-score, was proposed. We extended the M-score basicdefinition to consider prior knowledge the user might haveand presented four applications using the extended definition.Finally, we explored different approaches for efficientlyacquiring the knowledge required for computing theM-score, and showed that the M-score is both feasible and can fulfill its main goals.

VI. FUTURE WORK

We can also plan to extend the M-score to support multiplepublications with Di >1, and the sensitivity of combinationsof sensitive values; evaluate the measure using datafrom different domains, such as patient medical records,and using multiple contexts; and investigating othermisuseabilityweight measures that are designed to handledata in formats other than tabular data.

REFERENCES

[1] Agrawal, R., Srikant, R.: Privacy-preserving data mining. In: Proc. of the ACMSIGMOD Conference on Management of Data (SIGMOD 2000), pp. 439–450 (2000)
[2] Babcock, B., Chaudhuri, S., Das, G.: Dynamic sample selection for approximatequery processing. In: SIGMOD Conference, pp. 539–550 (2003)
[3] Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, Heidelberg(October 2007)
[4] Bishop, M.: The insider problem revisited. In: Proc. of the 2005 Workshop on NewSecurity Paradigms (NSPW 2005), pp. 75–76 (2005)
[5] Brackney, R., Anderson, R.: Understanding the Insider Threat: Proceedings of aMarch 2004 Workshop. RAND Corp. (2004)
[6] Lee, S.Y., Low, W.L., Wong, P.Y.: Learning fingerprints for a database intrusiondetection system. In: Gollmann, D., Karjoth, G., Waidner, M. (eds.) ESORICS2002. LNCS, vol. 2502, pp. 264–280. Springer, Heidelberg (2002)
[7] Ramasubramanian, P., Kannan, A.: Intelligent multi-agent based database hybridintrusion prevention system. In: Bencz´ur, A.A., Demetrovics, J., Gottlob, G. (eds.)ADBIS 2004. LNCS, vol. 3255, pp. 393–408. Springer, Heidelberg (2004)
[8] Schonlau, M., DuMouchel, W., Ju, W., Karr, A., Theus, M., Vardi, Y.: Computerintrusion: Detecting masquerades. Statistical Science 16(1), 58–74 (2001)
[9] Spalka, A., Lehnhardt, J.: A comprehensive approach to anomaly detection inrelational databases. In: DBSec, pp. 207–221 (2005)
[10] Srivastava, A., Sural, S., Majumdar, A.K.: Database intrusion detection usingweighted sequence mining. Journal of Computers 1(4), 8–17 (2006)
[11] Stonebraker, M.: Implementation of integrity constraints and views by query modification.In: SIGMOD Conference, pp. 65–78 (1975)
[12] Wenhui, S., Tan, D.: A novel intrusion detection system model for securing webbaseddatabase systems. In: Proc. of the 25th International Computer Softwareand Applications Conference on Invigorating Software Development (COMPSAC2001), p. 249 (2001)
[13] Yao, Q., An, A., Huang, X.: Finding and analyzing database user sessions. In: Proc.of Database Systems for Advanced Applications, pp. 283–308 (2005)