Abstract Submission for the Paper

Misuse Detection in Database Systems Through User Profiling

Christina Yip Chung, Michael Gertz, Karl Levitt
Department of Computer Science, University of California at Davis
{chungy| gertz | levitt @cs.ucdavis.edu}

Topic Categories:
Misuse/anomaly detection
Data mining
New IDS methodologies and technologies

Christina Yip Chung

Ms. Chung received her B.Sci.(Computer Science) degree from the University of Hong Kong in 1996 and her M.Sci.(Computer Science) degree from the University of California, Davis in 1998. She is currently a Ph.D. candidate in the Department of Computer Science at the University of California, Davis. Her research interests include data mining and knowledge discovery, misuse detection and database security.

Michael Gertz

Dr.Gertz received his M.S. degree in Computer Science from the University of Dortmund, Germany. In 1996 he received his Ph.D. in Computer Science from the University of Hannover, Germany, where he worked as a research and teaching assistant from December 1991 to November 1997. Since October 1997 Dr.Gertz is an Assistant Professor in the Department of Computer Science at the University of California, Davis.

Dr.Gertz's research interests are in the design and implementation of distributed, heterogeneous and multidatabase systems with a particular focus on the management of data quality and correctness. His current research is on database integration techniques, database interoperability, constraint maintenance in multidatabase systems and integrity enforcement in temporal databases. Dr.Gertz's research interests also include database security, the design of user interfaces to database systems and database administration tools

Karl Levitt

Professor Levitt received his Ph.D. from New York University in 1966. He is head of the Computer Security Laboratory in UC Davis.

Professor Levitt conducts research in the areas of computer security, automated verification, and software engineering. With respect to computer security he is working on techniques to detect malicious code (viruses, worms, time bombs, etc.) in programs and to detect attempts to penetrate or misuse computer systems, especially computer networks. With respect to verification, he is applying an automated theorem prover (Higher Order Logic - HOL) to the verification of hardware and software systems, especially operating systems for safety-critical embedded systems. With respect to software engineering, he is working on new methods for testing programs that make use of heuristic techniques and methods for automating the generation of operating system code from templates.

Motivation

Concepts for misuse detection in DBS have not been adequately addressed by existing intrusion detection systems (IDS) which reside on the operating system and/or network. Auditing the user behavior at these layers is unsuited for misuse detection at the DBS level because the semantics and structure of the data are not reflected in such low-level audit logs. Intrusion attempts that IDSs fail to detect at the operating system layer may be detected as anomalous events at the database system layer.

The success of IDS heavily depends on the domain knowledge available to the system. Domain knowledge of database systems, such as the structure and semantics of the data, is specified explicitly in the data. Database management systems provide a rich set of functionality designed to efficiently process large sets of data and are suitable for processing audit logs.

In the proposed approach we derive profiles to describe the typical behavior for users and roles in a relational database system. The profiles generated can be used in detecting misuse behavior. Very often, security officers (SO) do not use the available means to guard against the information stored in the DBS because the security policies are not well known. The profiles derived by our system can serve as a valuable tool for *security re-engineering* of an organization by helping the SO to define and refine policies and to verify existing policies.

Approach

Working scopes and distance measure: We conjecture that a user typically will not access all attributes and data in a database schema. Attributes used in a query are related through primary and foreign key dependencies among the relations in a schema. Therefore, the access patterns of users will form some *working scopes* which are sets of attributes that are usually referenced together with some values. Based on this conjecture, we define the notion of *distance measure* between sets of attributes that considers both the structure of the data and the user behavior. Our notion of distance measure is used to guide the search for regular patterns in audit data and to describe the typical user behavior.

Architecture: Our IDS consists of four components: (1) Auditor, (2) Data Processor, (3) Profiler, and (4) Detector. The Auditor is responsible for selecting the set of features to audit, such as the type of query, the set of attributes referenced by the query, or new and old values of modified data. The Data Processor groups data in the audit log into separate audit sessions. The Profiler derives a profile for each audit session. Based on how the data in the audit log are grouped into audit sessions, different types of profiles are generated, such as profiles for users and roles. The Detector detects misuse by comparing new information about the user behavior against his corresponding profile, and by comparing the profiles with the security policies.

Frequent itemset profiler: We propose a noval approach based on the idea of *frequent itemsets* and the notion of distance measure to discover the working scopes of users and database applications. Our frequent itemsets approach is similar to the clustering approach. Unlike clustering approaches, which are non-scalable, our frequent itemsets profiler is implemented by SQL queries and is scalable to the size of the audit session. The patterns discovered by our frequent itemset profiler are sets of features with their typical values, which are of finer level of granularity than clusters discovered by clustering algorithms which are only sets of objects.

Evaluation: We have conducted an analytical analysis on the correctness and completeness of our frequent itemset profiler and conducted preliminary study on its performance and effectiveness.