Chapter 7.7
Database Security and Statistical Database Security

Edgar R. Weippl
Secure Business Austria, Austria

INTRODUCTION

In this article we will present an introduction to issues relevant to database security and statistical database security. We will briefly cover various security models, elaborate on how data analysis in data warehouses (DWH) might compromise an individual's privacy, and explain which safeguards can be used to prevent attacks.

In most companies, databases are an essential part of IT infrastructure since they store critical business data. In the last two decades, databases have been used to process increasing amounts of transactional data, such as, a complete account of a person's purchases from a retailer or connection data from calls made on a cell phone.

As soon as this data became available from transactional databases and online transactional processing (OLTP) became well established, the next logical step was to use the knowledge contained in the vast amounts of data. Today, data warehouses (DWH) store aggregated data in an optimal way to serve queries related to business analysis.

In recent years, most people have begun to focus their attention on security. Early OLTP applications were mainly concerned with integrity of data during transactions; today privacy and secrecy are more important as databases store an increasing amount of information about individuals, and data from different systems can be aggregated. Thuraisingham (2002) summarizes the requirements briefly as "*However, we do not want the information to be used in an incorrect manner.*"

All security requirements stem from one of three basic requirements: confidentiality (aka secrecy), integrity, and availability (CIA). Confidentiality refers to the requirement that only authorized subjects, that is, people or processes should be permitted to read data. Integrity means that unauthorized modifications must not be permitted. This includes both modifications by unauthorized people and incorrect modification by authorized users. To correctly perform the services requested, the system needs to remain available; a denial-of-service compromises the requirement of availability.

Other security requirements may include privacy, non-repudiation, and separation of duties. These requirements are, however, composite requirements that can be traced back to one of the three basic requirements. Privacy, for instance, is the non-disclosure (=confidentiality) of personal data; non-repudiation refers to the integrity of transaction logs and integrity of origin. Throughout this article we will focus only on technical attacks and safeguards and not on social engineering. Social engineering is often the easiest and, in many cases, a very successful attack vector. For an in-depth coverage of social engineering we recommend (Böck, 2007).

In Section 2 we cover the most relevant access control models; in Section 3 we provide an overview of security in statistical databases. Finally, in Section 4 we highlight the essentials of securing not only the transactional and the statistical databases but the entire system.

BACKGROUND

Access Control is the most important technique or mechanism for implementing the requirements of confidentiality and integrity. Since databases were among the first large-scale systems in military applications, there is a long history of security models, dating back to the 1960s. The basic principle in all access control models is that a *subject* is or is not permitted to perform a certain *operation* on an *object*. This process is described by the triplet (s, op, o). A security policy specifies who is authorized to do what. A security mechanism allows enforcement of a chosen security policy.

One can distinguish between two fundamentally different access control mechanisms: discretionary access control (DAC) and mandatory access control (MAC). In DAC models the user decides which subject is able to access which object to perform a certain operation. In contrast, when using MAC, the system decides who is allowed to access which resource and the individual user has no discretion to decide on access rights.

Discretionary Access Control (DAC)

In relational database management systems (DBMS), the objects that need to be protected are tables and views. Modern DBMS allow a fine granularity of access control so that access to individual fields of a record can be controlled.

By default, a subject has no access. Subjects may then be *granted* access, which can be *revoked* anytime. In most systems the creator of a table or a view is automatically granted all privileges related to it. The DBMS keeps track of who subsequently gains and loses privileges, and ensures that only requests from subjects who have the required privileges—at the time the request is executed—are allowed.

Mandatory Access Control (MAC)

Mandatory Access Control is based on system-wide policies that cannot be changed by individual users. Each object in the database is automatically assigned a security class based on the access privileges of the user who created the object.

The most widely known implementation of a MAC system is a multi-level security (MLS) system. MLS systems were first described by Bell LaPadula (Bell, 1975) in the 1960s. Each subject, which could either be a user or user program, is assigned a *clearance* for a security class. Objects are assigned security *levels*. Security levels and clearances can be freely defined as long as all items are comparable pair-wise. Most common are security classes (i.e., levels and clearances), such as, top secret (TS), secret (S), confidential (C), and unclassified (U).

Rules based on security levels and clearances govern who can read or write which objects. Today, there are only a few commercially available systems that support MAC, such as, SELinux or also Oracle DBMS (Version 9 and higher)