Tutorial 7

Aims

- To discuss the different database protection needs
- To exemplify some database protection methods

Questions

Short question

How often do you back up your data?

1. Database integrity has two main aspects: schema integrity and data accuracy/integrity. What are the differences between the two?
   - Integrity means the data in the database is reliable (data not corrupted, no unauthorized modification, etc).
   - Physical protection of integrity means preventing unauthorized users from physically accessing the database servers (usb plug-in for example).
   - **Schema integrity** refers to maintaining the data structure (columns and rows in tables) correctly. For example, the records/rows in databases will not be mixed up.
   - **Data accuracy/integrity** is about having correct data values. E.g. when two users are accessing the same data element, there can be a conflict if both of them try to update (write) the data at the same time.

2. What is the difference between direct data access, direct inference and indirect inference? Illustrate it on examples not related to computing.
   - Users' **direct data access** is managed by common access control mechanisms, for example those utilising on authentication.
     Example: asking the price of an item (e.g. in a shop)
   - Authorised access can still jeopardise database security via inference attacks, where an attacker can deduce/guess sensitive information by collecting and performing analysis on publicly accessible information.
     - **Direct inference attacks** can involve queries that look general in nature. However, those queries are designed to include only very few records in the response. The queries may contain misleading components, to make detection of the attack intention difficult.
       Example: asking the price range of items, when we know there is only one or two items of that type (e.g. from someone, who may interpret it as asking about their financial situation)
     - **Indirect inference attacks** use a combination of queries that may overlap, and by analysing the answers the attacker may be able to eliminate uninterested records and obtain/guess the correct information.
       Example: some car dealers do not disclose the price of different options, only the final price with a particular option. You need to make a number of queries to find out the price of each option.

3. What are the major issues with data aggregation?

   It can be a tool for indirect inference attack. Practical dangers include linking data to each other and possibly to individuals, and that can violate the right to privacy. It is the basis of data mining. Example problem: can lead to revealing medical records.

4. Consider database integrity and access control requirements.

   (a) How are they interlinked and where do they differ?

   The interlink is that both access control and data integrity require that only authorized users can physically access databases and modify schema/data. The main difference is that integrity relates to data stored in the database, while access control is also responsible for revealing the data to users, and e.g. prevent information leaking.

   (b) Are there any mechanisms that can address both? If yes, what are they, if no, why aren't there any?

   Example: access control describing update (write) permissions

   (c) What methods are available for multilevel security?

   Encryption: using different keys for different levels
   Integrity lock: access rights for different users is protected by a hash
   Sensitivity lock: access rights for different users is ancrypted and protected by a hash