Buổi 8

CHƯƠNG 5: AN TOÀN CSDL THỐNG KÊ

CSDL thống kê là gì?

- Là một CSDL được sử dụng cho mục đích phân tích thống kê.

- Là một CSDL chứa các bản ghi nhạy cảm mô tả về các cá nhân nhưng chỉ các câu truy vấn thống kê (như: COUNT, SUM, AVERAGE, MAX, MIN…) mới được trả lời, ngoài các câu truy vấn này thì những truy vấn vào các mục dữ liệu riêng sẽ không được đáp lại

- Có hai dạng SDB cơ bản:

+ Dạng CSDL quan hệ

+ Dạng CSDL vĩ mô

Tại sao phải bảo vệ SDB>?

- Vì SDB chứa dữ liệu thống kê liên quan đến thông tin nhạy cảm của nhiều cá nhân

Tấn công vào các SBD bằng cách nào?

- Tấn công suy diễn (Interference attack): kết hợp các câu truy vấn thống kê -> Thu được thông tin bí mật về 1 cá nhân

Một số kiểu tấn công suy diễn

- Tấn công trực tiếp

+ Sử dụng các câu truy vấn thông thường, không phải truy vấn thống kê

- Tấn công dựa vào đếm

+ Là loại tấn công bằng cách kết hợp giá trị đếm với giá trị tổng để thu được thông tin bí mật

- Tấn công trình theo dõi

- Tấn công hệ tuyến tính

CSDL thống kê khác CSDL quan hệ?

Các đặc tính của SDB cần được bảo vệ:

- SDB tĩnh:

+ SDB không thay đổi trong suốt thời gian tồn tại của chúng

+Ví dụ: CSDL thống kê dân số

- SDB động:

+ Thay đổi liên tục theo sự thay đổi của dữ liệu thực, cho phép sửa đổi để phản ánh các thay đổi động của thế giới thực

+ Ví dụ các CSDL nghiên cứu trực tuyến, lớp học trực tuyến khi bổ sung thành viên…

- SDB trực tuyến (online):

+ Người sử dụng nhận được các phản hồi thời gian thực cho các câu truy vấn thống kê của mình

- SDB ngoại tuyến (offline):

+ Người sử dụng không biết khi nào các thống kê của họ được xử lý, việc SDB bị lộ sẽ khó khăn

- Kiến thức làm việc (working knowledge)

+ Là tập các mục thông tin (field) và giá trị thuộc tính trong SDB và các kiểu thống kê có sẵn trong SDB mà người dùng có thể có thể biết một cách hợp lejw

- Kiến thức bổ sung của người sử dụng (sumplementary knowledge):

+ Người sử dụng có thể có kiến thức bên ngoài về các cá nhân được biểu diễn trong SDB. Người dùng hoàn toàn có thể lợi dụng kiến thức này cho các mục đích xấu để suy diễn

- Công thức đặc trưng:

+ Là một công thức logic, được ký hiệu bởi một chữ cái viết hoa (A, B, C, …), trong đó các giá trị thuộc tính được kết hợp với nhau thông qua các toán tử Boolean như OR, AND, NOT

- Tập truy vấn (query set): của một công thức đặc trưng C là tập tất cả các bản ghi thoả mãn C. Ký hiệu là X(C)

- Thống kê trên C: là các câu truy vấn thống kê trên C. Ký hiệu q(C). Ví dụ: COUNT(C), SUM(C,AJ), MIN(C,AJ), MAX(C,Aj)