MỘT SỐ PHƯƠNG PHÁP DIỀU KHIỂN TRUY NHẬP TRONG HỆ THỐNG BẢO VỆ THÔNG TIN CƠ SỞ DỮ LIỆU

Đào Thị Hồng Hạnh

Bộ giáo dục và Đào tạo

I. Đặt vấn đề

Điều khiển truy nhập gồm hai vấn đề:

1. Hệ thống bảo vệ thông tin (HTBVTT) xác nhận chính xác người được quyền thâm nhập vào hệ thống.

2. HTBVTT xác nhận được quyền của người sử dụng trên các file hệ thống quản lý.

Bài báo này trình bày một phương pháp xác nhận người sử dụng bằng chữ ký số dựa vào khoá công khai. Hệ thống không cần lưu trữ password của người sử dụng nhằm đảm bảo sự an toàn của hệ thống được cao. Đồng thời đưa ra một số thuật toán điều khiển quyền truy nhập của người sử dụng (nsd) trên các tập dữ liệu hệ thống quản lý.

Giả sử HTBVTT quản lý quyền của m người sử dụng trên n tệp dữ liệu. Ta xây dựng ma trận điều khiển truy nhập (xem trong [1]) $A_{m\times n} = (a_{ij})_{1\le i\le n, 1\le j\le n}$, trong đó a_{ij} biểu diễn quyền của nsd i trên tệp j được mã như sau:

$a_{ij} = 0$	không có quyền
$a_{ij} = 1$	quyền chỉ đọc
$a_{ij} = 2$	quyền bổ xung
$a_{ij} = 3$	quyền sửa
$a_{ij} = 4$	quyền xoá
$a_{ij} = 5$	quyền xem cấu trúc file
$a_{ij} = 6$	quyền sửa cấu trúc file

Bài báo [1] đã chỉ ra một phương pháp nhận biết a_{ij} nhờ khoá đơn nguyên cấp cho mỗi người sử dụng. Bài báo này đưa ra một thuật toán mới để HTBVTT nhận biết a_{ij} nhờ xây dựng được hệ số thực C đặc trưng cho ma trận điều khiển truy nhập.

Khi đó $a_{ij} = f(C, i, j)$.

Nếu xét theo quan điểm cơ sở dữ liệu phân chia là trên mỗi tệp dữ liệu người sử dụng hoặc có mọi quyền hoặc không có quyền thì mỗi nsd i được cấp một véc tơ nhị phân $a_i = (a_{i1}, a_{i2}, ..., a_{in})$. Trong đó a_{ij} nhận giá trị 0 nếu không có quyền hoặc giá trị 1 nếu có quyền. Bài báo cũng chỉ ra một thuật toán thay vì cấp véc tơ nhị phân a_i bởi một khoá đơn nguyên có độ an toàn cuả hệ mã công khai Merkle-Helman Knapsack.

Bài báo gồm một số phần tiếp sau:

Phần 2. Trình bày một phương pháp xác nhận người sử dụng bằng chữ ký số dựa vào khoá công khai. (Khái niệm khoá công khai xem trong [2]).

Phần 3. Trình bày một phương pháp điều khiển ma trận truy nhập dựa vào mã số học.

Phần 4. Trình bày một phương pháp xác nhận véc tơ nhị phân điều khiển quyền truy nhập trong cơ sở dữ liệu phân chia dựa vào hệ mã công khai Merkle-Helman Knapsack.

Phần 5. Cài đặt các thuật toán chỉ ra trên máy vi tính.

Phần 6. Kết luận.

2. Một phương pháp xác nhận người sử dụng bằng chữ ký số dựa vào khoá công khai

Cái mới cuả phương pháp này là HTBVTT không cần phải lưu mật khẩu cuả người sử dụng. HTBVTT giao cho mỗi người sử dụng i mật khẩu pw_i và nhờ thủ tục xác nhận HTBVTT biết được người sử dụng i có hợp pháp hay không.

2.1 Chữ ký số dựa vào khoá công khai

Cho U và X là một tập hữu hạn nào đó. Cho khoá công khai $e \in U$ (khoá xác thực chữ ký) và khoá mật (khoá sinh chữ ký d) $d \in U$.

Gọi $f_d^{-1} : X \to X$ là hàm sinh chữ ký bởi khoá d,

Gọi $f_e : X \to X$ là hàm xác nhận chữ ký bởi khoá e.

Hai hàm sinh chữ ký và hàm xác nhận chữ ký có quan hệ với nhau như sau:

$$f_e(f_d^{-1}(x)) = x, \quad \forall x \in X. \tag{1}$$

Gọi $g : X \times U \to X$ là hàm một chiều có tính chất giao hoán (với $X \in U$):

$$g_r(f_d^{-1}(x)) = f_d^{-1}(g_r(x)), \tag{2}$$

$$g_r(g_q(x)) = g_q(g_r(x)). \tag{3}$$

2.2 Thủ tục xác nhận người sử dụng

HTBVTT sinh ra một khoá công khai e và một khoá mật d trong sơ đồ chữ ký (U, X, f, f^{-1}) $(X \in U)$

$$\text{HTBVTT cấp cho nsd } i \text{ mật khẩu } \quad pw_i = f_d^{-1}(h(ID_i)), \tag{4}$$

trong đó ID_i là số hiệu của người sử dụng i, h là hàm một chiều nào đó sao cho $h(ID_i) \in X$.

HTBCTT xác nhận người sử dụng như sau:

Bước 1. HTBVTT yêu cầu nsd i xuất trình mật khẩu pw_i.

Bước 2. HTBVTT sinh ra một số ngẫu nhiên $r \in U$ và gửi r cho nsd i.

Bước 3. Nsd i sinh ra một số ngẫu nhiên $q \in U$ và sinh ra S_i, T_i thoả mãn đẳng thức

$$T_i = g_q(h(ID_i)), \tag{5}$$

$$S_i = pw_i \,.\, g_r(g_q(gT_i(pw_i))), \tag{6}$$

trong đó phép toán . là một ánh xạ: $X \times X \to X$ thoả mãn bất đẳng thức

$$f_d^{-1}(v.u) = f_d^{-1}(v).f_d^{-1}(u)(v, u \in X) \tag{7}$$

Nsd i gửi S_i, T_i, ID_i tới HTBVTT.

Bước 4. HTBVTT kiểm tra đẳng thức sau

$$f_e(S_i) = h(ID_i).g_r(gT_i)) \tag{8}$$

Nếu đúng thì chấp nhận nsd i là hợp pháp, ngược lại thì từ chối yêu cầu truy nhập hệ thống.

Chứng minh. Thật vậy

$$\begin{aligned} f_e(S_i) &= f_e(pw_i.g_r(g_q(gT_i(pw_i)))), \\ &= f_e(f_d^{-1}(h(ID_i)).g_r(g_q(gT_i(f_d^{-1}(h(ID_i)))))), \\ &= f_e(f_d^{-1}(h(ID_i)).f_d^{-1}g_r(gT_i(g_q(f_d^{-1}(h(ID_i)))))), \\ &= f_e(f_d^{-1}(h(ID_i)).(g_r(gT_i(g_q(f_d^{-1}(h(ID_i))))))), \\ &= (h(ID_i)).g_r(gT_i(T_i)), \end{aligned}$$

Ta có thể cài đặt các hàm sau thoả mãn các yêu cầu trên

$$f_e(x) = x^e \bmod n$$

$$f_d^{-1}(x) = x^d \bmod n$$

$$g_r(x) = x^r \bmod n$$

$$x.y = x.y \bmod n$$

trong đó n là tích của hai số nguyên tố đủ lớn khác nhau, e và d thoả đẳng thức $e.d \bmod \Phi(n) = 1$. (Theo sơ đồ chữ ký RSA).

3. Một phương pháp điều khiển ma trận truy nhập dựa vào mã số học

3.1. Khái niệm mã số học (Arithmetic coding)

Mã số học AC được Pasco lần đầu tiên trình bày và đã được Witen, Neal áp dụng để nén số liệu. Tư tưởng cơ bản cuả AC là mỗi thông báo được biểu diễn bằng một số thực thuộc khoảng [0,1). Mỗi ký hiệu trong bảng chữ cái được gắn với một khoảng con nửa mở gọi là khoảng ký hiệu (viết tắt là S-range). Cách mã hoá và giải mã thông báo được trình bày qua ví dụ cụ thể sau:

Giả sử mỗi ký hiệu trong bảng chữ cái $\{a, b, c, ..., z\}$ tương ứng với khoảng xác định sau

a	[0, 1/26)
b	[1/26, 2/26)
c	[2/26, 3/26)
d	[3/26, 4/26)
...	.....
z	[25/26, 1)

Ta muốn mã thông baó "bac" thực hiện các bước sau

- Chữ b trong khoảng $[l_i, r_i) = [1/26, 2/26)$
- Chữ a trong khoảng $[l, r) = [0, 1/26)$

Khi đó ta có

$$l_{i+1} = l_i + (r_i - l_i)l,$$

$$r_{i+1} = l_i + (r_i - l_i)r.$$

Sau khi mà chữ a thì khoảng cuả thông báo (ký hiệu là M-range) là [0.04;0.0399) Mã tiếp chữ c trong khoảng [l,r)=[2/26,3/26) được M-range là [0.038575;0.038632). Ta chọn 1 số thực l trong khoảng [0.038575, 0.038632), ví dụ R= 0.0386 đủ để thông báo biểu diễn "bac". Để giải mã thông baó ta làm như sau:

Ta thấy rằng R thuộc khoảng [1/26, 2/26) nên chữ đầu tiên phải là chữ b. Khi ký tự tiếp theo ta thấy chỉ có chữ a mới làm cho khoảng [1/26,2/26) giảm thành [0.04, 0.0399) có chứa R. Quá trình giải mà cứ tiếp tục như vậy cho đến khi thông báo "bac" được phục hồi. Sau đây trình bày ứng dụng mã số học vào điều khiển ma trận truy nhập.

Gọi $A_{m\times n}$ là ma trận điều khiển truy nhập. a_{ij} biểu diễn quyền cuả nsd i trên tệp j.

Ta xây dựng hàm

$$\underset{1\le i\le m, 1\le j\le n}{T(i,j)} = (i-1).n + j,$$

M là tổng số người sử dụng, n là tổng số tệp. Hàm này đặt tương ứng mỗi phần tử (i,j) với mỗi thứ tự tuyến tính từ 1 đến nm. Ta xây dựng dãy

$$\{H_{T(i,j)}\}. \qquad (1)$$

Trong đó $H_{T(i,j)} = a_{ij}$. Sử dụng kỹ thuật AC mã hoá thông báo 1. Gọi P là tổng số quyền truy nhập cuả USER trên file. Rõ ràng mỗi phần tử $H_{T(i,j)}$ nhận giá trị trong tập $A = \{0, 1, 2, ..., p-1\}$. Sau đó ta gán S-range [0,1/p), [1/p,2/p), ..., [p-1/p,1) cho tập A một cách tương ứng. Ta thấy rằng thông báo (1) có thể được mã hoá bởi 1 số thực C, $0 \leq C < 1$.

$$C = \sum_{T(i,j)=1}^{nm} H_{T(i,j)} * x(1/p)^{T(i,j)}. \tag{2}$$

Định lý 3.2. *Cho $A_{m\times n}$ là một ma trận điều khiển truy nhập, p là tổng số quyền cuả sử dụng đối với tệp dữ liệu. Tồn tại một hằng số thực $C, 0 \leq C \leq 1$ sao cho các a_{ij} được biểu diễn như sau*

$$a_{i,j} = [C/(1/p)]^{T(i,j)} \ mod\ p.$$

Trong đó phép toán $[x]$ lấy phần nguyên cuả số x.

Chứng minh. Không giảm tổng quát ta đặt $T(i,j) = k$, $1 \leq k \leq nm$. Thay vào (2) ở trên ta có

$$C = \sum_{k=1}^{nm} H_k * (1/p)^k.$$

Như vậy thì

$$\begin{aligned} C/(1/p)^k &= (\sum_{k=1}^{nm} H_k p^{-l})/p^{-l} \\ &= \sum_{k=1}^{nm} H_u p^{k-u} + \sum_{v=k+1}^{nm} H_v p^{k-v} + H_k \end{aligned}$$

Đặt

$$j_1 = \sum_{u=1}^{nm} H_u p^{k-u},\ e = \sum_{v=k+1}^{nm} H_v p^{k-v}$$

Do $H_v \leq p-1$, $\forall 1 \leq v \leq nm$ ta có

$$e \leq (p-1)(\sum_{v=k+1}^{nm} p^{k-v}) = 1 - p^{k-n}.$$

Vậy $0 \leq e \leq 1$. Vậy ta có $[C/(1/p)^k] \bmod p = H_k$ hoặc là

$$[C/(1/p)^k] \bmod p = H_{T(i,j)} = a_{i,j}.$$

3.2. Nhận xét

HTBVTT chỉ cần nsd i khai báo số hjiệu nsd và số hiệu tập dữ liệu là hệthống tính được $a_{i,j}$. Từ đó hệ thống ra quyết định cho phép hay từ chối yêu cầu truy nhập.

4. Phương pháp xác nhận véc tơ nhị phân điều khiển quyền truy nhập trong cơ sở dữ liệu phân chia dựa vào hệ mã công khai Merkle-Helman Knapsack

4.1. Khái niệm hệ mã công khai Merkle-Helman Knapsack

Bài toán ba lô: Cho một số nguyên dương C và một véc tơ $A = (a_1, a_2, ..., a_n)$ trong đó các a_i là số nguyên dương.

Tìm véc tơ nhị phân $M = (m_1, ..., m_n)$ sao cho

$$C = AM, \quad \text{hoặc} \quad C = \sum_{i=1}^{n} a_i m_i.$$

Đây là bài toán NP đầy đủ. Thuật toán tốt nhất để giải quyết bài toán này có độ phức tạp tính toán là $O(2^{n/2})$. Denning trong [2] đã chỉ ra thuật toán thời gian tuyết tính nếu chọn các phần tử cuả véc tơ A sao cho

$$a_i > \sum_{j=1}^{i-1} a_j, \quad \text{với} \quad i = 1, 2, ..., n. \tag{1}$$

Thuật toán snap (CA)

```
for i:= n downto 1 do
   begin
      if C≥ a_i then m_i = 1 else m_i = 0;
         C := C − a_i m_i
   end;
if C = 0 then snap := M else "bài toán vô nghiệm".
```

Merkle và Hellman biến dổi bài toán Knapsack sao cho rất khó giải quyết nếu không biết thông tin phụ.

Bước 1. Chọn véc tơ Knapsack.

Chọn $B = (b_1, b_2, ..., b_n)$ thoả (1), nghiả là

$$b_i = \sum_{j=1}^{i-1} b_j \quad i = 2, 3, ..., n.$$

Bước 2. Chọn số nguyên u sao cho

$$U > 2b_n > \sum_{i=1}^{n} b_i.$$

Bước 3. Chọn một số nguyên W sao cho USCLN(U,W) =1 và tính giá trị ngược W^{-1} cuả W theo mod u.

Bước 4. Xây dựng véc tơ $A = WB$, tức là $a_i = Wb_i \bmod U$.

Ta thấy rằng để giải $C = AM$ là khó khăn nhưng nếu biết thông tin cửa sập W^{-1} và U thì có thể giải dễ dàng. Véc tơ $A = (a_1, a_2, ..., a_n)$ được gọi là véc tơ Knapsack khó được dùng làm khóa công khai. Khoá mật là véc tơ Knapsack B và thông tin cửa sập U, W^{-1}. Trong đó $B = W^{-1}A \bmod U$.

Phép mã hoá $C = E_A(M) = AM$.

Phép giải mã $D_A(C) = snap(W^{-1}C \bmod U,\ B) = M$.

Hệ mã này áp dụng cho điều khiển truy nhập trong cơ sở dữ liệu phân chia như sau: Giả sử hệ thống điều khiển quyền truy nhập của m người sử dụng trên n tập dữ liệu khác nhau. Mỗi nsd i được cấp một véc tơ nhị phân

$$M_i = (m_{i1}, m_{i2}, ..., m_{in}).$$

Trong đó $m_{ij} = 1$ nếu nsd có quyền trên tập j,

$m_{ij} = 0$ nếu nsd không có quyền trên tập j,

HTBVTT sẽ cấp cho nsd khoá nguyên $K = AM_i$ thay vì phải cấp véc tơ nhị phân M_i, trong đó A là véc tơ Knapsack khó. Hệ thống yêu cầu nsd xuất trình khoá nguyên K và hệ thống giải mã để tìm lại véc tơ nhị phân M_i

$$M_i = D_A(K) = \text{snap}(W^{-1}K \bmod U, B).$$

Ví dụ: Giả sử véc tơ $M_i = (1,1,0,1)$ điều khiển quyền của nsd i trên 4 tệp dữ liệu. HTBVTT chọn véc tơ $B = (1,3,5,10)$ thoả (1), $U = 20$, $W = 7$. Khi đó ta tính được $W^{-1} = 3$ nhờ giải phương trình đồng dư

$$WX \bmod U = 1.$$

Véc tơ Knapsack khó

$$A = (7 * 1 \bmod 20,\ 7 * 5 \bmod 20,\ 7 * 10 \bmod 20)$$

$$A = (7, 1, 15, 10).$$

HTBVTT cấp cho nsd i khoá đơn $K = AM_i$

$$K = \sum_{j=1}^{4} a_j m_{ij} = 18,$$

nsd i muốn truy nhập hệ thống thì phải xuất trình khoá đơn K. HTBVTT giải mã để tìm véc tơ M_i nhờ thuật toán snap (3*18 mod 20, B) = (1,1,0,1).

5. Cài đặt các thuật toán trên máy vi tính

HTBVTT gồm các menu sau:

5.1. Menu xác nhận người sử dụng

Menu gồm các chức năng sau:

1. HTBVTT chọn các hàm một chiều $h(x)$, $f_e(x)$, $f_d^{-1}(x)$, $g_r(x)$ thoả mãn điều kiện (1)-(3) của phần II như sau

$$n = 53 * 61 = 3233,\ e = 71\quad d = 791.$$
$$h(x) = x^8 \bmod 3233$$
$$f_e(x) = x^{71} \bmod 3233$$
$$f_d^{-1}(x) = x^{791} \bmod 3233$$
$$g_r(x) = x^r \bmod 3233$$

2. HTBVTT cấp cho mỗi người sử dụng một số hiệu ID_i và mật khẩu PW_i

3. Thực hiện thủ tục xác nhận người sử dụng tuân theo 4 bước ở phần 2. Từ đó HTBVTT ra quyết định cho phép hay từ chối yêu cầu truy nhập hệ thống của nsd i.

5.2. Menu điều khiển ma trận truy nhập dựa vào mã số học

Menu này gồm các chức năng sau:

1 - Nhập ma trận điều khiển truy nhập

2 - Tính thông số C đặc trưng cho ma trận điều khiển truy nhập

3 - Xác nhận quyền của nsd trên file, từ đó cho phép hay từ chối yêu cầu truy nhập tệp dữ liệu hệ thống quản lý.

5.3. Menu điều khiển quyền truy nhập trong cơ sở dữ liệu phân chia

Menu gồm các chức năng sau:

1 - Xây dựng hệ véc tơ KNAPSACK B thoả (1) của phần 4

2 - Xây dựng hệ véc tơ KNAPSACK khó A và các thông tin cửa sập U, W^{-1}

3 - Nhập các véc tơ nhị phân điều khiển quyền truy nhập của từng nsd mà hệ thống quản lý

4 - Tính khoá đơn cấp cho mỗi nsd

5 - Xác nhận quyền của nsd trên tập dữ liệu nhờ giải mã khoá đơn.

6. Kết luận

Bài báo đã đưa ra được một phương pháp xác nhận nsd mà không cần phải lưu password. Độ an toàn của hệ thống cao nhờ sử dụng khoá công khai. Đồng thời cũng xây dựng được một thông số đặc trưng cho ma trận điều khiển truy nhập giúp cho việc xác nhận quyền của nsd trên các tập dữ liệu hệ thống quản lý. Cuối cùng bài báo chỉ ra một sơ đồ điều khiển quyền truy nhập trong cơ sở dữ liệu phân

chia nhờ khoá nguyên có độ an toàn cuả hệ mã công khai Merkle-Knapsack. Các thuật toán trên đà được cụ thể hoá trên máy vi tính.

Tài liệu tham khảo

1. Đào Thị Hồng Hạnh, *Về một phương pháp điều khiển truy nhập dùng khoá đơn trong hệ thống bảo vệ thông tin cơ sở dữ liệu*, Tạp chí Tin học và Điều khiển 2, 1993.
2. Denning D. E.,*Cryptography and data security* Addison Weslay reading, MA 1982.
3. Davies D. W., *Applying the RSA digital signature to electric mail*, IEEE Computer, 1983, 55-62.

Abstract

Some methods of access control in protected system for database

This paper presents a new user authentication scheme, which does not require a management file for user's password. It has a high security, realizing the authentication of a large number of users by a single public key. This paper presents a access matrix control by a real number C between 0 and 1. Finally, the paper proposes an access control scheme in distributed database with a single key, which has security of public key Merkle - Knapsack.