

BẢO VỆ SỰ TOÀN VỆN CỦA CƠ SỞ DỮ LIỆU QUAN HỆ BẰNG KỸ THUẬT THỦY VÂN

Lưu Thị Bích Hương

Viện Công nghệ thông tin, Trường Đại học Sư phạm Hà Nội 2

Ngày nhận bài 21/01/2021, ngày nhận đăng 05/3/2021

Tóm tắt: Bài báo đề xuất một lược đồ thủy vân dùng để bảo vệ sự toàn vẹn cho cơ sở dữ liệu quan hệ. Trong lược đồ này, thủy vân được tạo ra từ các thuộc tính trong cơ sở dữ liệu và được nhúng vào một số thuộc tính của cơ sở dữ liệu quan hệ. Cơ sở dữ liệu quan hệ sau khi nhúng thủy vân thì giá trị sử dụng của các thuộc tính không thay đổi. Bài báo cũng chứng minh được tính đúng đắn của lược đồ thủy vân đề xuất.

Từ khóa: Toàn vẹn dữ liệu; thủy vân; cơ sở dữ liệu.

1. Mở đầu

Trong thời đại của cách mạng công nghệ 4.0, kỹ thuật số ngày càng trở nên phổ biến và vai trò quan trọng trong cuộc sống. Theo báo cáo Digital 2020 (kỹ thuật số 2020) toàn cầu của We are social, Việt Nam có 96,9 triệu dân, số lượng thuê bao di động là 145,8 triệu thuê bao chiếm tỷ lệ 150% so với tổng dân số cả nước, số lượng người dùng internet là 68,17 triệu người, chiếm tỷ lệ 70% số dân, số lượng người dùng mạng xã hội là 65 triệu người chiếm tỷ lệ 67% số dân. Điều này cho thấy phần đông người dân Việt Nam đã tiếp cận với điện thoại di động thông minh và cũng không ít người sử dụng 2-3 chiếc điện thoại cùng một lúc để phục vụ cho cuộc sống. Việc con người ngày càng phụ thuộc vào Internet trong các hoạt động hàng ngày, việc quảng bá sản phẩm, thương hiệu theo hình thức trực tuyến là điều mà bất cứ doanh nghiệp hay tổ chức nào cũng cần thực hiện. Qua thống kê cho thấy số người dùng mạng xã hội đang tăng nhanh và đây cũng là công cụ truyền thông phổ biến của hầu hết tất cả các doanh nghiệp, tổ chức ở Việt Nam.

Giống như mọi quy luật tự nhiên, mọi sự tốt đẹp luôn đi kèm theo mặt trái của nó. Thông tin, dữ liệu dễ tìm thấy đồng nghĩa với dễ bị ăn cắp bản quyền, truy cập trái phép, nhiều nguồn đưa ra thông tin sai lệch, mất đi tính đúng đắn của thông tin... Vì vậy, vấn đề bảo vệ sự toàn vẹn của cơ sở dữ liệu là những vấn đề nóng đang được xã hội quan tâm nhiều hơn cả.

Hiện nay, có khá nhiều các nghiên cứu về thủy vân để bảo vệ sự toàn vẹn của dữ liệu, như nghiên cứu về thủy vân với dữ liệu là ảnh số 0, nghiên cứu về thủy vân cơ sở dữ liệu quan hệ với các thuộc tính số [2, 3, 8] hay các nghiên cứu về thủy vân cơ sở dữ liệu với các thuộc tính không phải kiểu số [4, 6, 7]. Với nghiên cứu này, chúng tôi đề xuất một lược đồ thủy vân cơ sở dữ liệu quan hệ với các thuộc tính kiểu số và không phải kiểu số. Trong lược đồ này, thủy vân được tạo ra dựa vào các thuộc tính và sẽ được nhúng vào cơ sở dữ liệu.

Trong phần tiếp theo chúng tôi trình bày một số định nghĩa. Phần 3 là lược đồ thủy vân, trong phần này sẽ trình bày tư tưởng của lược đồ thủy vân cơ sở dữ liệu quan hệ, thuật toán nhúng thủy vân, thuật toán phát hiện thủy vân và xác minh sự toàn vẹn. Phần 4 là chứng minh tính đúng đắn của lược đồ thủy vân đề xuất. Phần cuối cùng là kết luận.

2. Định nghĩa

Trong một lược đồ quan hệ, có một số thuộc tính có ý nghĩa quan trọng và một số thuộc tính khác có ảnh hưởng không lớn đến giá trị sử dụng cũng như ý nghĩa thực tế. Sau đây chúng tôi sẽ đưa ra hai định nghĩa để làm rõ hơn về các loại thuộc tính này.

Định nghĩa 1: Thuộc tính có tác động cao

Một thuộc tính được gọi là thuộc tính tác động cao hoặc thuộc tính có tác động cao hoặc thuộc tính có ảnh hưởng cao hay còn gọi là thuộc tính có ý nghĩa quan trọng khi bất kỳ một thay đổi nào đối với giá trị của thuộc tính thì giá trị đặc trưng của chúng cũng bị thay đổi theo.

Định nghĩa 2: Thuộc tính có tác động thấp

Một thuộc tính được gọi là thuộc tính tác động thấp hoặc thuộc tính có tác động thấp hoặc thuộc tính có ảnh hưởng thấp hay còn gọi là thuộc tính không có ý nghĩa quan trọng nếu có một thay đổi nhỏ đối với giá trị của thuộc tính sẽ có ảnh hưởng không lớn đến giá trị sử dụng cũng như giá trị ý nghĩa thực tế của thuộc tính đó.

Ví dụ, trong một lược đồ quan hệ nhân sự thì các thuộc tính ghi số chứng minh thư, họ tên, năm sinh, giới tính, ngày tăng lương là những thuộc tính quan trọng, có ảnh hưởng lớn đối với đương sự. Các thuộc tính như quê quán, nơi sinh có ảnh hưởng không lớn đối với đương sự. Thông thường các thuộc tính tác động thấp là các thuộc tính không phải kiểu số.

3. Lược đồ thủy vân

Như trên đã trình bày, có khá nhiều bài báo tập trung nghiên cứu các kỹ thuật thủy vân cơ sở dữ liệu quan hệ có các thuộc tính kiểu số [2, 3, 8], một số tác giả nghiên cứu về thủy vân các cơ sở dữ liệu quan hệ có các thuộc tính không phải kiểu số [4, 5] và có bài báo nghiên cứu về thủy vân cơ sở dữ liệu với dữ liệu phân loại 0.

Trong 0, lược đồ thủy vân được sử dụng cho dữ liệu không phải số. Ý tưởng của lược đồ này là dựa vào tất cả các thuộc tính không phải số để tính khóa bí mật, sau đó dùng khóa bí mật để nhúng vào thuộc tính tác động thấp. Trong lược đồ này nếu để lộ thuật toán nhúng thủy vân thì cơ sở dữ liệu quan hệ không còn được đảm bảo.

Trong 0, lược đồ thủy vân được sử dụng cho dữ liệu không phải số. Ý tưởng của lược đồ này là dựa vào sự ràng buộc ngữ nghĩa của các thuộc tính văn bản để thủy vân. Việc xác định ngữ nghĩa của các bộ giá trị rất khó thực hiện đối với tiếng Việt.

Trong 0, lược đồ thủy vân được sử dụng cho dữ liệu phân loại. Điểm mấu chốt của kỹ thuật này là việc đổi thứ tự của các bộ trong cơ sở dữ liệu. Tuy nhiên, lược đồ này có tính an toàn chưa cao do việc đổi thứ tự của các bộ chỉ được thực hiện trên các cặp đã định trước.

Lược đồ thủy vân đề xuất sử dụng cho cơ sở dữ liệu quan hệ có các thuộc tính số và thuộc tính không số. Tư tưởng của lược đồ là tính giá trị trên tất cả các thuộc tính, sau đó lấy giá trị này kết hợp với khóa thủy vân cho trước để xác định thủy vân rồi nhúng vào các thuộc tính không phải số tác động thấp.

Lược đồ thủy vân được thiết kế để bảo vệ sự toàn vẹn cho các quan hệ thuộc lược đồ quan hệ có dạng:

$$R(H_1, H_2, \dots, H_m, L_1, L_2, \dots, L_n)$$

Trong đó, m thuộc tính H_1, H_2, \dots, H_m , là thuộc tính tác động cao còn L_1, L_2, \dots, L_n là n thuộc tính tác động thấp. Không mất tính tổng quát, giả sử trong lược đồ quan hệ trên mỗi quan hệ có ω bộ, K là khóa thủy vân.

Lược đồ thủy vân có sử dụng một số ký hiệu được liệt kê trong Bảng 1.

Bảng 1: Các ký hiệu được sử dụng trong lược đồ thủy vân

Ký hiệu	Ý nghĩa
R	Lược đồ quan hệ
r	Quan hệ thuộc lược đồ R
r_i	Bộ thứ i của quan hệ r
$r_i.L_j$	Giá trị của thuộc tính L_j thuộc bộ r_i
ω	Số bộ trong quan hệ r
K	Khóa thủy vân
n	Số thuộc tính tác động thấp
m	Số thuộc tính tác động cao

Lược đồ thủy vân đề xuất dùng để bảo vệ sự toàn vẹn cho các cơ sở dữ liệu quan hệ được thực hiện dựa vào hai thuật toán:

- Thuật toán nhúng thủy vân vào quan hệ
- Thuật toán phát hiện thủy vân và xác minh sự toàn vẹn.

3.1. Thuật toán nhúng thủy vân

Ý tưởng của thuật toán nhúng thủy vân bao gồm các bước sau:

- Sinh thủy vân từ các bộ của lược đồ cơ sở dữ liệu quan hệ R
- Với mỗi quan hệ $r \in R$, chia ω bộ của r thành g nhóm
- Mỗi một nhóm G_i sinh ra xác định 1 ký tự thủy vân nhúng vào thuộc tính tác động thấp tại vị trí bất kì trong nhóm.

3.1.1. Sinh thủy vân

- Tìm thủy vân từ các bộ của quan hệ bằng cách dựa vào khóa thủy vân, các thuộc tính của cơ sở quan hệ.

- Đối với m thuộc tính tác động cao, tính tổng mã Unicode của toàn bộ các thuộc tính và ký hiệu là:

$$A^H = \sum_{i,j} \{ \sum \text{ASCII của } r_i.H_j, 1 \leq i \leq \omega; 1 \leq j \leq m \}$$

- Đối với n thuộc tính tác động thấp, tính tổng mã Unicode của tất cả các bộ theo từng thuộc tính

$$A^L_j = \sum_i \{ \sum \text{ASCII của } r_i.L_j, 1 \leq i \leq \omega \}$$

- Xây dựng một ma trận đặt tên là D bao gồm 2 hàng và n cột với các thành phần là D_{i1}, D_{i2} , (với $i = 1, 2, \dots, n$) được tính như sau:

$$D_{i1} = A^H + A^L_i$$

$$D_{i2} = A^L_i$$

- Xây dựng ma trận thủy vân W bằng cách nhân ma trận D với ma trận chuyển vị D^T của nó. Ma trận W thu được là một ma trận vuông kích thước 2×2 với các giá trị trên đường chéo chính là e_1, e_2 . Đây chính là các giá trị đặc trưng của ma trận này.

- Điểm khác của lược đồ đề xuất là dùng hàm băm các giá trị e_j sau khi ghép với khóa thủy vân K . Chuyển các giá trị băm thành các ký tự thủy vân W_j theo công thức:

$$W_j = \text{ATOC}(\text{HASH}(e_j, K) \text{ MOD } 224 + 32), j = 1, 2$$

Trong đó $\text{ATOC}()$ là hàm chuyển mã ASCII thành ký tự tương ứng. Sở dĩ phải cộng thêm 32 là vì 31 ký tự đầu tiên trong bảng mã ASCII là các ký tự không in ra được. Khóa K là bí mật và đối xứng, chỉ người chủ cơ sở dữ liệu được biết và được dùng ở cả quá trình nhúng thủy vân và phát hiện thủy vân. Hàm băm được sử dụng ở đây để đảm bảo rằng khi có bất cứ một thay đổi nào xảy ra trong cơ sở dữ liệu thì các ký tự thủy vân W_j cũng thay đổi theo. Đây chính là điều mong muốn đối với các lược đồ thủy vân dùng để bảo vệ sự toàn vẹn của các cơ sở dữ liệu quan hệ.

Procedure SinhTV(R, K)

Input: R, K

Output: W

```

1.    $A^H = 0$ 
2.   for  $i = 1$  to  $\omega$  do
3.       for  $j = 1$  to  $m$  do
4.            $A^H = A^H + \sum \text{ASCII của } r_i.H_j$ 
5.       end for
6.   end for
7.   for  $j = 1$  to  $n$  do
8.        $A^L_j = 0$ 
9.       for  $i = 1$  to  $\omega$  do
10.           $A^L_j = A^L_j + \sum \text{ASCII của } r_i.L_j$ 
11.       end for
12.   end for
13.   for  $i = 1$  to  $n$  do // Xây dựng ma trận D
14.        $D_{i1} = A^L_i + A^H$ 
15.        $D_{i2} = A^L_i$ 
16.   end for
17.    $W = D * D^T$  // Sinh các ký tự thủy vân
18.   for  $j = 1$  to 2 do
19.        $W_j = \text{ATOC}(\text{HASH}(e_j, K) \text{ MOD } 224) + 32$  //Ký tự thủy vân
20.   end for
21.   Sắp thứ tự  $W_j$ 
22.   return W

```

3.1.2. Chia nhóm các bộ của quan hệ

Cho quan hệ r thuộc R với ω bộ dữ liệu, khi đây việc phân chia ω bộ của quan hệ r thành các g nhóm dựa vào khóa chính của bộ và khóa thủy vân K . Cách phân chia này sẽ làm tăng tính ngẫu nhiên khi chọn các bộ và phân các bộ vào các nhóm riêng rẽ. Tính ngẫu nhiên này có độ bảo mật cao được đảm bảo bằng hàm băm mật mã $H()$.

Mục đích của việc phân chia nhằm tăng khả năng bền vững của thủy vân trước các tấn công và phát hiện giả mạo nếu có. Với số lượng nhóm g , khóa thủy vân K chỉ chủ sở hữu cơ sở dữ liệu quan hệ biết, việc phân chia các bộ của quan hệ vào nhóm G_k ($k = 0, 1, \dots, g - 1$) sẽ được thực hiện bằng thủ tục ChiaNhóm() sau:

Procedure ChiaNhóm(r, K, g)

1. **for** $k = 0$ **to** $g-1$ **do** // khởi tạo các chỉ số và các nhóm
2. $q_k = 0$
3. $G_k = \phi$
4. **end for**
5. **for** $i = 1$ **to** ω **do**
6. $k = H(K \parallel r_i.P) \bmod g$
7. $G_k = G_k \cup \{r_i\}$
8. $q_k = q_k + 1$ // q_k là số bộ trong nhóm G_k
9. **end for**

3.1.3. Nhúng thủy vân vào các thuộc tính tác động thấp

Đầu tiên, tiến hành lựa chọn bộ để nhúng thủy vân trong nhóm G_j của quan hệ r :

$$t = \text{CODE}(W_{(j \bmod 2)+1}) \bmod q_j$$

Trong đó q_j là số bộ của nhóm G_j , hàm CODE là hàm chuyển ký tự thành mã Unicode tương ứng.

Sau đó, tiến hành xác định thuộc tính và vị trí nhúng để nhúng ký tự thủy vân trong bộ đã chọn.

Trong lược đồ thủy vân có sử dụng thủ tục nhúng thủy vân vào các thuộc tính tác động thấp trong một nhóm, thủ tục đó được xây dựng như sau:

Procedure NhungTV(G_j, W)

1. // Chọn bộ thứ t trong G_j để nhúng
2. $t = \text{CODE}(W_{(j \bmod 2)+1}) \bmod q_j$
3. //Xác định thuộc tính nhúng
4. $e = t \bmod n$
5. $vtrinhung = \text{Converter}(H(K)) \% \text{length}(r_t.L_e)$
6. Chèn $W_{(j \bmod 2)+1}$ vào $r_t.L_e$ tại $vtrinhung$

3.1.4. Nhúng thủy vân vào cơ sở dữ liệu quan hệ

Thuật toán 3.1: Nhúng thủy vân

Input:

- Lược đồ $R(H_1, H_2, \dots, H_m, L_1, L_2, \dots, L_n)$. Trong đó H_1, H_2, \dots, H_m là các thuộc tính có tác động cao, còn L_1, L_2, \dots, L_n là các thuộc tính có tác động thấp.

- Khóa thủy vân K , số nhóm g .

Output: Lược đồ R đã nhúng thủy vân.

1. $W = \text{SinhTV}(R, K)$
2. **for** $r_i \in R$ **do**
3. ChiaNhóm(r_i, K, g)
4. **for** $j = 0$ **to** $g-1$ **do**
5. NhungTV(G_j, W)
6. **end for**
7. **End for**

3.2. Phát hiện thủy vân và xác minh sự toàn vẹn

Quan hệ r thuộc R sau khi nhúng thủy vân có thể lưu thông bình thường trong môi trường công cộng. Khi có nghi ngờ về một sự xuyên tạc hay giả mạo nào đó đối với quan hệ này, người chủ sở hữu quan hệ có thể tiến hành xác minh bằng thuật toán phát hiện thủy vân.

Giả sử r' là một quan hệ thuộc lược đồ R . Cần kiểm tra xem r' có phải là bản giả mạo của quan hệ r đã thủy vân hay không. Thuật toán phát hiện thủy vân và xác minh sự toàn vẹn được chia làm 3 phần cơ bản:

- Thực hiện qui trình sinh thủy vân trên r' giống như đã thực hiện trên r trong thuật toán nhúng thủy vân. Gọi các ký tự được sinh ra từ r' là W_j với $j = 1, 2$.
- Theo các qui tắc nhúng thủy vân vào các thuộc tính tác động thấp trong thuật toán 3.1.2, trích ra các ký tự đã nhúng và gọi các ký tự này là W'_j với $j = 1, 2$.
- So sánh các ký tự của W' với các ký tự của W với nhau. Nếu chúng trùng khớp thì R' chính là R và không bị sửa đổi. Nếu ngược lại thì kết luận là R' chính là R nhưng đã bị sửa đổi.

Thuật toán 3.2. Phát hiện thủy vân

Input:

- Lược đồ $R(H_1, H_2, \dots, H_m, L_1, L_2, \dots, L_n)$, trong đó H_1, H_2, \dots, H_m là các thuộc tính có tác động cao, còn L_1, L_2, \dots, L_n là các thuộc tính có tác động thấp.
- Khóa thủy vân K , số nhóm g .

Output: Lược đồ R toàn vẹn hoặc đã bị sửa đổi.

1. $W = \text{SinhTV}(R', K)$
2. **For** $r_i \in R'$ **do**
3. ChiaNhom(r_i, K, g)
4. **for** $j = 0$ **to** $g-1$ **do**
5. NhưngTV(G_j, W)
6. $W'_j = \text{substring}(r'_i, L_e, \text{vtrinhung}, 1)$
7. **end for**
8. Sắp thứ tự W'
9. **If** $W \triangleleft W'$ **then**
10. return “Lược đồ bị sửa đổi”
11. **end if**
12. **End for**

Giả sử có trung bình p ký tự trong mỗi thuộc tính của một bộ và mỗi bộ có $m + n$ thuộc tính. Khi đó, việc tính tổng các ký tự có trong một bộ tốn khoảng $p(m + n)$ đơn vị thời gian. Vậy thời gian tính toán cho lược đồ thủy vân là $\omega p(m + n)$ đơn vị thời gian hay độ phức tạp là $O(\omega p(m + n))$.

4. Tính đúng đắn

Để chứng minh tính đúng đắn của lược đồ đề xuất chúng tôi đưa ra các định lý sau:

Định lý 4.1: Lược đồ quan hệ R có các thuộc tính tác động cao và các thuộc tính tác động thấp được nhúng thủy vân bằng thuật toán 3.1.

Chứng minh:

Trong thuật toán 3.1, đầu tiên sẽ tìm thủy vân từ các bộ của quan hệ bằng cách dựa vào khóa thủy vân, các thuộc tính có tác động cao và các thuộc tính có tác động thấp. Sau đó, sẽ nhúng thủy vân vào các thuộc tính có tác động thấp. Do đó để chứng minh định lý 4.1 ta cần phải tìm được ma trận thủy vân W và cách nhúng thủy vân W vào thuộc tính tác động thấp. Thật vậy:

- Đối với m thuộc tính có tác động cao, tính tổng mã ASCII trong toàn bộ các thuộc tính, cụ thể là:

$$A^H = \sum_{i,j} \{ \sum \text{ASCII của } r_i.H_j, 1 \leq i \leq \omega; 1 \leq j \leq m \}$$

- Đối với n thuộc tính có tác động thấp, tính tổng mã ASCII của thuộc tính, cụ thể là:

$$A^L_j = \sum_i \{ \sum \text{ASCII của } r_i.L_j, 1 \leq i \leq \omega \}$$

- Ma trận D bao gồm 2 hàng và n cột với các thành phần là D_{i1}, D_{i2} (với $i = 1, 2, \dots, n$) được tính như sau:

$$\begin{aligned} D_{i1} &= A^H + A^L_i \\ D_{i2} &= A^L_i \\ \Rightarrow W &= D \times D^T. \end{aligned} \quad (1)$$

- Ma trận W có các giá trị trên đường chéo chính là e_1, e_2 .

- Dùng hàm băm các giá trị e_j sau khi ghép với khóa thủy vân K . Chuyển các giá trị băm thành các ký tự thủy vân W_j theo công thức:

$$W_j = \text{ATOC}(\text{HASH}(e_j, K) \text{ MOD } 224 + 32); \quad j = 1, 2$$

- Tiến hành lựa chọn ký tự để nhúng vào trong thuộc tính:

- Nếu $n \leq 2, w^*_i = W_i$ ($i = 1, 2, \dots, n$)

- Nếu $n > 2, w^*_i = W_{(i \text{ MOD } 2)+1}$ ($i = 1, 2, \dots, n$)

- Xác định vị trí nhúng trong các thuộc tính bằng cách sử dụng hàm băm với khóa K . (2)

Từ (1) và (2) ta có điều phải chứng minh.

Định lý 4.2: Thuật toán 3.2 xác định sự toàn vẹn của lược đồ quan hệ R nếu lược đồ quan hệ R không bị sửa đổi.

Chứng minh:

Để chứng minh định lý ta sẽ chứng minh W_j và W'_j là trùng khớp. Thật vậy :

- Dựa vào m thuộc tính tác động cao và n thuộc tính tác động thấp tính được ma trận D bao gồm 2 hàng và n cột với các thành phần là D_{i1}, D_{i2} (với $i = 1, 2, \dots, n$) với:

$$\begin{aligned} D_{i1} &= A^H + A^L_i \\ D_{i2} &= A^L_i \end{aligned}$$

Trong đó A^H, A^L_i là tổng mã ASCII của các thuộc tính tác động cao và thuộc tính tác động thấp.

$$\Rightarrow W_j \quad (j = 1, 2) \quad (1)$$

- Theo thuật toán 3.2, ta có:

$$W'_i = \text{substring}(r'_j.L_i, j \text{ MOD } (\text{length}(r'_j.L_i) - 1), 1)$$

Hay W'_i sẽ lấy ra chuỗi con gồm một ký tự từ $r'_j.L_i$

$$\Rightarrow \text{bit thủy vân } W'_j \quad (j = 1, 2) \quad (2)$$

- Mặt khác, theo giả thiết quan hệ r không bị sửa đổi nên

$$\Rightarrow r'_j.L_i \text{ trùng với } r_j.L_i \quad (3)$$

Từ (1), (2) và (3) $\Rightarrow W_j$ trùng với $W'_j \Rightarrow$ Điều phải chứng minh.

5. Kết luận

Lược đồ thủy vân đề xuất nhằm bảo vệ sự toàn vẹn cho cơ sở dữ liệu quan hệ có các thuộc tính số và thuộc tính không số có những ưu điểm sau:

- Bền vững: Các giá trị đặc trưng của lược đồ có thể được nhúng trong thuộc tính có tác động thấp ở khắp nơi trong quan hệ. Vì vậy rất khó để có thể gỡ bỏ hết các ký tự thủy vân đã nhúng.

- Nhạy cảm: Mọi thay đổi trên quan hệ bất kỳ của lược đồ đều ảnh hưởng đến ký tự thủy vân được sinh ra điều này có nghĩa ta có thể phát hiện được sự thay đổi trên cả các quan hệ không nhúng thủy vân lên trên.

- Phát hiện mù: Quá trình xác minh sự toàn vẹn của lược đồ không đòi hỏi lược đồ gốc và thủy vân gốc.

- Không hiện: Thủy vân được nhúng là các ký tự có đặc tính là không hình dạng (rỗng) và không chiếm chỗ.

Nhược điểm của lược đồ là việc sử dụng các ký tự không chiếm chỗ không hiện hình và chèn các ký tự này vào vị trí bất kỳ trên các thuộc tính tác động thấp khiến cho việc khôi phục dữ liệu gốc trở lên khó khăn hơn.

Lời cảm ơn: Nghiên cứu này được tài trợ bởi Trường Đại học Sư phạm Hà Nội 2 qua Đề tài có mã số HPU2.CS-2021.03.

TÀI LIỆU THAM KHẢO

- [1] Phuoc Hung Vo, Thai Son Nguyen, Van Thanh Huynh, Thanh C Vo, Thanh Nghi Do, *Secure and Robust Watermarking Scheme in Frequency Domain Using Chaotic Logistic Map Encoding*, Advanced Computational Methods for Knowledge Engineering, Springer, 2019.
- [2] Chitla Arathi, "Literature Survey on Distortion based Watermarking Techniques for Databases," *International Journal of Computer Science & Communication Networks*, Vol. 2(4), pp. 456-463.
- [3] Al-Haj, A. and Odeh, A., "Robust and blind watermarking of relational database systems," *Journal of Computer Science*, 4: 1024-1029, 2008.
- [4] Bedi R., Thengade A., Wadhai V., "A New Watermarking Approach for Non Numeric Relational Database," *International Journal of Computer Applications* (0975 - 8887), Vol. 13, No. 7, pp. 37-40, 2011.
- [5] Javier Franco Contreras and Gouenou Coatrieux, "Protection of Relational Databases by Means of Watermarking: Recent Advances and Challenges," *Advances in Security in Computing and Communications*, 2017. DOI: 10.5772/intechopen.68412.

- [6] Murugan R, Jaseena K U, John T Abraham, “An Invisible Watermarking Technique for Integrity and Right Protection of Relational Databases,” *International Journal of Applied Engineering Research*, Vol. 12, No. 24 (2017), pp. 15754-15758.
- [7] Yingjiu Li, Huiping Guo, Sushil Jajodia, *Tamper Detection and Localization for Categorical Data Using Fragile Watermarks*, DRM '04: Proceedings of the 4th ACM workshop on Digital rights management New York, NY, USA: ACM (2004), pp. 73-82.
- [8] Shuguang Yuan, J. Yu, Peisong Shen, C. Chen, *Verify a Valid Message in Single Tuple: A Watermarking Technique for Relational Database*, Computer Science, DASFAA, 2020.

SUMMARY

PROTECTING THE INTEGRITY OF RELATIONAL DATABASE BY WATERMARKING TECHNIQUE

Luu Thi Bich Huong

The Institute for Information Technology, Hanoi Pedagogical University 2

Received on 21/01/2021, accepted for publication on 05/3/2021

The paper proposed a watermarked scheme used to protect integrity for relational databases. In this schema, watermark is created from attributes in the database and embedded in some properties of the relational database. After embedding the watermark in relational database, the use value of the attributes does not change. The paper proves the correctness of the proposed watermarked scheme.

Keywords: Data integrity; watermark; relation database.