BAN CƠ YẾU CHÍNH PHỦ

**HỌC VIỆN KỸ THUẬT MẬT MÃ**

-------------------------------



BÀI TẬP LỚN MÔN

**AN TOÀN CƠ SỞ DỮ LIỆU**

**Đề tài:**

**TÌM HIỂU CÁC KỸ THUẬT THỦY VÂN SỐ TRONG CƠ SỞ DỮ LIỆU QUAN HỆ**

**Giảng viên hướng dẫn**: Thầy Lại Minh Tuấn

**Nhóm sinh viên thực hiện**: Ngô Quang Triều – AT160554

Nguyễn Thị Lệ Thu – AT160447

Nguyễn Văn Cường – AT160605

Trần Văn Tiến – AT160551

Hà Nội, 2021

**LỜI NÓI ĐẦU**

Ngày nay, với sự phát triển mạnh mẽ của các mạng máy tính tốc độ cao, đặc biệt là Internet, các phương tiện kỹ thuật số như phương tiện lưu trữ , phương tiện truyền thông, đã mở ra một kỷ nguyên mới – kỷ nguyên thông tin số. Hầu hết các thông tin ngày nay đều được lưu trữ dưới dạng số hóa. Việc trao đổi, phân bố, sao chép và xử lý các sản phẩm số này ngày càng nhanh chóng, đơn giản, nằm ngoài tầm kiểm soát của các tổ chức. Việc đặt ra cho tất cả các phương thức, kinh doanh, phân phối tài nguyên số trên mạng là tuân thử các nguyên tắc về quyền sở hữu trí tuệ, và không cản trở quá trình phân phối, trao đổi tài nguyên số. Nhu cầu được bảo vệ bản quyền, và sở hữu trí tuệ các sản phẩm số đã trở thành một vấn đề quan trọng và đang được quan tâm. Dữ liệu là một tài sản vô cùng quan trọng đối với chủ thể sở hữu của nó. Chia sẻ dữ liệu số đang trở thành một xu hướng mới ở cả cấp độ cá nhân và tổ chức. Dữ liệu được chia sẻ phải tuân thủ tính bí mật (để ngăn chặn việc tiết lộ dữ liệu trái phép), tính toàn vẹn (sửa đổi dữ liệu độc hại) và tính khả dụng (khả năng khôi phục từ lỗi dữ liệu). Dữ liệu kỹ thuật số có thể được sao chép, thay đổi và có thể được phân phối lại với các mục đích khác nhau và điều này có thể vi phạm bản quyền của chủ sở hữu. Sự gia tăng đột biến gần đây của Internet, những tài sản kỹ thuật số như hình ảnh, video, âm thanh, nội dung cơ sở dữ liệu,.. một cách dễ dàng những người bình thường trên khắp thế giới có thể truy cập được như chia sẻ, mua, phân hối hoặc nhiều mục đích hoạt động khác. Do đó, các sản phẩm kỹ thuật số như vậy đối mặt với những thách thức nghiêm trọng như vi phạm bản quyền, phân phối bất hợp phát, yêu cầu quyền sở hữu, giả mạo, trộm cắp,… . Trong bối cảnh này, công nghệ thủy vân số là một giải pháp hiệu quả để đáp ứng thách thức như vậy. Hình mờ được coi là một số loại thông tin được nhúng vào dữ liệu cơ bản để phát hiện giả mạo, chứng minh quyền sở hữu, truy tìm kẻ phản bội,…

Đề tài “Tìm hiểu về kỹ thuật thủy vân số trong cơ sở dữ liệu quan hệ” sẽ trình bày:

Chương I: Tổng quan về thủy vân số bao gồm khái niệm, phân loại thủy vân số và một số đặc điểm , yêu cầu đối với thủy vân số.

Chương II: Một số kỹ thuật thủy vân số cho các cơ sở dữ liệu quan hệ , chương này chủ yếu trình bày về hai kỹ thuật chính là thủy vân số biến dạng và thủy vân số không biến dạng.

Chương III: Thực nghiệm phương pháp thủy vân số cơ sở dữ liệu quan hệ MySQL dựa vào kỹ thuật thủy vân số biến dạng.

**MỤC LỤC**

**CHƯƠNG I. TỔNG QUAN VỀ THỦY VÂN SỐ**

1. **Khái niệm về thủy vân số 4**
2. **Phân loại thủy vân số 5**
   1. Thủy vân số dễ vỡ (Fragile materking) 5
   2. Thủy vân số bền vững (Roburst Watermark) 6
3. **Các ứng dụng của thủy vân số 6**
   1. Xác định quyền sở hữu 6
   2. Lấy dấu vân tay 7
   3. Phát hiện gian lận giả mạo 7
4. **Yêu cầu của thủy vân số 7**
5. **Đặc tính của thủy vân số 8**

**Chương II. MỘT SỐ KỸ THUẬT THỦY VÂN SỐ 9**

1. **Thủy vân số biến dạng 9**
   1. Thủy vân số dựa trên thuộc tính kiểu dữ liệu số 10
   2. Thủy vân số dựa trên thuộc tính kiểu dữ liệu phân loại 10
   3. Thủy vân số dựa trên bộ giá trị hoặc chèn thuộc tính 11
2. **Thủy vân số không biến dạng 12**
   1. Trích xuất giá trị băm dưới dạng thông tin mờ 12
   2. Kếp hợp dấu hiệu của chut sở hữu và đặc tính CSDL 13
   3. Chuyển đổi quan hệ CSDL thành dạng nhị phân 13
   4. Hoán vị dựa trên R-tree 14

**CHƯƠNG 3. THỰC NGHIỆM 15**

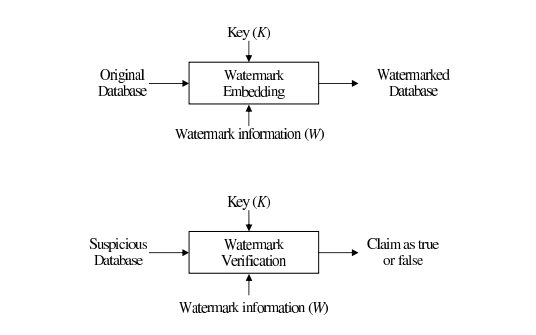
**TÀI LIỆU THAM KHẢO 16**

**CHƯƠNG I. TỔNG QUAN VỀ THỦY VÂN SỐ**

1. **Khái niệm về thủy vân số**

Thuật ngữ watermark được bắt nguồn từ một loại mực hình được viết trên giấy và chỉ hiển thị khi nhúng giấy đó vào nước. Thuật ngữ thủy vân số được cộng đồng thế giới chấp nhận rộng rãi vào đầu thập niên 1990. Khoảng năm 1995, sự quan tâm rộng rãi đến thủy vân số bắt dầu phát triển nhanh.

Thủy vân số hay Digital Watermarking là một loại dấu ấn thủy vân được nhúng một cách tinh vi vào dữ liệu gốc ban đầu (như ảnh số, audio, video, hình ảnh, text hay kể cả một cơ sở dữ liệu). Nhằm xác định thông tin bản quyền đối với dữ liệu đó. Về cơ bản, thủy vân số cơ sở dữ liệu bao gồm hai giai đoạn: giai đoạn nhúng thủy vân số và giai đoạn xác minh thủy vân số.



**Hình  1.1 Sơ đồ mô tả kỹ thuật đánh dấu cơ sở dữ liệu cơ bản**

Trong giai đoạn nhúng thủy vân số, một khóa riêng tư K (khóa chỉ người sở hữu biết) được sử dụng để nhúng hình mờ W vào trong cơ sở dữ liệu gốc. Cơ sở dữ liệu này sau khi được đánh dấu thủy vân sẽ được công khai rộng rãi. Để xác minh quyền sở hữu cơ sở dữ liệu đáng ngờ bằng cách sử dụng khóa K (khóa được sử dụng trong giai đoạn nhúng) được trích xuất và so sánh thông tin hình mờ này với thông tin hình mờ gốc ban đầu.

Dữ liệu quan hệ định nghĩa từ các dữ liệu đa phương tiện theo nhiều khía cạnh:

- Dữ liệu dự phòng: Các đối tượng đa phương tiện bao gồm só lượng lớn các bit cung cấp dùng để bao phủ, che giấu thông tin về hình mờ W, trong khi đối tượng cơ sở dữ liệu là một tập hợp các đối tượng độc lập, được gọi là các bộ giá trị. Hình mờ phải được nhúng vào các bộ giá trị này.

- Dữ liệu quan hệ không theo thứ tự: vị trí tương đối của các đối tượng dữ liệu đa phương tiện không được thanh đổi trong khi có một sự sắp xếp vị trí nào đó cho dữ liệu quan hệ được thiết lập

- Cập nhật thường xuyên: Không có một phần nào của các đối tượng đa phương tiện bị thay thế trong khi các bộ giá trị có thể được chèn, xóa hoặc cập nhật trong các hoạt động cơ sở dữ liệu bình thường.

Do những khác biệt này giữa dữ liệu quan hệ và dữ liệu đa phương tiện nên không có phương pháp đánh dấu hình ảnh hoặc âm thanh nào phù hợp cho việc tạo hình mờ của cơ sở dữ liệu quan hệ.

1. **Phân loại thủy vân số**
   1. ***Thủy vân số dễ vỡ (Fragile materking)***

Thủy vân số “dễ vỡ” (Fragile materking) là kỹ thuật nhúng watermark vào trong cơ sở dữ liệu sao cho khi truyền dữ lệu đó trên môi trường công cộng. Nếu có bất cứ một phép biến đổi nào làm thay đổi dữ liệu gốc đã được đánh dấu thủy vân thì dữ liệu nhận được sẽ không còn nguyên vẹn so với dữ liệu watermark ban đầu. Các kỹ thuật watermark này được sử dụng trong các ứng dụng nhận thực thông tin và phát hiện xuyên tạc thông tin. Thủy vân số dễ vỡ yêu cầu dấu thủy vân phải dễ bị biến đổi trước sự tấn công trên dữ liệu thủy vân. Do đó, thủy vân dễ vỡ thường dùng để xác thực tính toàn vẹn của sản phẩm chứa dấu thủy vân. Để xác thực tính toàn vẹn của các sản phẩm chứ dấu thủy vân, ta so sánh sự sai khác giữa dấu thủy vân gốc (W) và dấu thủy vân trích xuất được (W’). Nếu W ≠[https://thuthuatnhanh.com/thuthuatnhanh.jpg](https://thuthuatnhanh.com/cach-viet-dau-khac-trong-excel/) W’ thì kết luận sản phẩm chứ dấu thủy vân đã bị tấn công. Ngược lại, thì kết luận sản phẩm chưa bị tấn công ( toàn vẹn ) .

* 1. ***Thủy vân số bền vững (Roburst Watermark)***

Thủy vân số bền vững hay còn gọi là Roburst Watermark thường được dùng trong ứng dụng bảo vệ bản quyền. Trong ứng dụng đó, watermark đóng vai trò là thông tin của chủ sở hữu hợp pháp. Dấu thủy vân được nhúng trong ảnh số. Roburst Watermark như một hình thức dán tem bản quyền, chống lại việc tẩy xóa, làm giả hay các biến đổi phá hủy watermark, do đó watermark yêu cầu phải tồn tại bền vững (ít bị biến đổi).

Thủy vân số bền vững được chia làm hai loại là thủy vân ẩn và thủy vân hiện. Thủy vân hiện là loại thủy vân được hiển thị ngay trên sản phẩm và người dùng có thể nhìn thấy được, thông tin bản quyền được hiện ngay trên sản phẩm. Còn với thủy vân ẩn thì bằng mắt thường không thể nhìn thấy được dấu thủy vân. Trong việc bảo vệ bản quyền, chủ sở hữu sẽ chứng minh sản phẩm đang bị đánh cắp bằng cách chỉ ra thủy vân ẩn đã được nhúng trong sản phẩm đó.

1. **Các ứng dụng của thủy vân số trong cơ sở dữ liệu**
   1. ***Xác định quyền sở hữu***

Mặc dù đã có nhiều quy định về bảo vệ bản quyền và có những chuyển biến tích cực nhưng những hành động xâm phạm bản quyền tác tỉa vẫn còn diễn ra tràn lan, tinh vi và công khai trước chủ sở hữu. Trong việc mua bán hay trao đổi các tắc phẩm số này sẽ nảy sinh một số vấn đề như:

* Phải đảo bảo quyền của tác giả: để bảo vệ được bản quyền thì chủ sử hữu dữ liệu đó phải chúng minh được nó thuộc quyền sở hữu của mình bằng cách đưa ra những thông tin đặc biệt của dữ liệu đó.
* Đảm bảo thông tin : thông tin phải sẵn sàng cho người dùng hợp pháp, chống phân phốt bât hợp pháp các nội dung của tác phẩm.
* Lần vết thông tin, phát hiện người phân phối sản phẩm bất hợp pháp: khi vấn đề vi phạm bản quyền xảy ra hoặc chủ sở hữu nghi có bản sao sản phẩm là không hợp lệ. Do đó thủy vân phải không thể nhìn thấy hoặc không thể tách rời sản phẩm chứa nó. Dấu thủy vân ( một thông tin nào đó mang ý nghĩa quyền sở hữu tác giả) sẽ được nhúng vào trong tác phẩm này và chỉ có người chủ sở hữu hợp pháp các sản phẩm đó biết dùng để minh chúng cho bản quyền sản phẩm.

Ví dụ: Chủ sở hữu A có thể nhúng watermark vào cơ sở dữ liệu R của cô ấy một khóa bí mật K (khóa chỉ người sở hữu A biết).Sau đó watermark có thể được công khai. Giả sử A nghi ngờ người dùng B có bản sao chép không hợp lệ từ quan hệ R, tập các bộ giá trị S có thể là tập con của R. Để đánh bại quyền sở hữu của B, A có thể chứng minh sự hiện diện của watermark trong cơ sở dữ liệu của cô ấy trong dữ liệu quan hệ của B.

* 1. ***Lấy dấu vân tay***

Việc lấy dấu vân tay nhằm mục đích xác định kẻ phản bội. Trong các ứng dụng nội dung cơ sở dữ liệu được công bộ rộng rãi qua mạng, thủy vân này được sử dụn để nhận diện người gửi hay người nhận một thông tin nào đó. Ví dụ các thủy vân khác nhau sẽ được nhúng vào các bản copy khác nhau của thông tin gốc trước khi chuyển cho nhiều người. Những ứng dụng này yêu cầu phải đảm bảo độ an toàn cao cho các thủy vân tránh bị xóa dấu vết trong khi phân phối.

* 1. ***Phát hiện gian lận và giả mạo***

Dấu thủy vân trong trường hợp này được dùng để phát hiện dữ liệu bị xuyên tạc. Dấu thủy vân sẽ được nhúng để trong một tác phẩm sau đó đc trích xuất ra để so sánh với dấu thủy vân ban đầu, nếu có sử sai khác chứng tỏ sản phẩm đã bị tấn công và xuyên tạc, do đó các thủy vân nên được ẩn để tránh sự tò mò của đối phương. Ngoài ra thủy vân này có thể tìm được vị trí bị xuyên tạc cũng như phân biệt được các thay đổi. Yêu cầu dối với ứng dụng này là khả năng giấu thông tin cao và thủy vân không bền vững.

1. **Yêu cầu cơ bản của thủy vân trên cơ sở dữ liệu quan hệ**

Khi thực hiện thủy vân dữ liệu cần phải có một số tiêu chí để đánh giá chất lượng của giải thuật. Thông thường người ta sẽ dựa trên các tính chất sau:

*Đảm bảo tính vô hình*: Quá trình thủy vân sẽ làm biến đổi ảnh do mang thủy vân được nhúng vào. Tính vô hình thể hiện ở mức độ biiến đổi ảnh mang làm cho người khác khó thể nhận ra, do vậy sẽ đảm bảo được tính bí mật của thủy vân.

*Khả năng chống giả mạo (tính toàn vẹn):* Đối với thủy vân thì khả năng chống giả mạo là yêu cầu vô cùng quan trọng vì như vậy mới bảo vệ được bản quyền. Để có thể chống lại sự giả mạo thì bất cứ sự thay đổi nào về nội dung của các dữ liệu thì thủy vân sẽ hủy đi dẫn đến rất khó làm giả các dữ liệu có chứa thủy vân.

*Tính bền vững*: Thủy vân phải có khả năng tồn tại cao để chống lại các hình thức tấn công có chủ đích hoặc không có chủ đích. Để đảm bảo cho điều này, thủy vân phải được dấu trong các vùng quan trọng đối với trực giác. Phương pháp này phải đảm bảo sao cho việc không thể lấy lại thủy vân tương đương với dữ liệu bị biến đổi quá nhiều.

1. **Đặc tính của thủy vân số**

Để đảm bảo cân bằng giữa các thộc tính là thực sử cần thiết và vấn đề đảm bảo cân bằng phải dựa trên sự phân tích một cách cẩn thận.

*Độ trung thực*: nghĩa là người theo dõi không thể phát hiện ra dấu thủy vân. Để tín hiệu thực sự là không thể cảm thấy thì thông tin phải được nhúng vào những bit ít quan trọng. Tuy nhiên, tín hiệu lại dễ dàng bị loại bỏ trong quá trình nên có tổn thất thông tin. Các nghiên cứu trước đây về thủy vân đề tập trung vào việc thiết kế thủy vân không thể thấy được và thường nhúng thủy vân vào vùng tín hiệu ít quan trọng về mặt cảm nhận. Tuy nhiên, gần đây các kỹ thuật khác lại chèn giấu thủy vân không thấy được vào vùng tín hiệu quan trọng về mặt cảm nhận, có thể nâng cao tính bền vững chống lại các quá trình xử lý tín hiệu.

*Tính bền vững*: dữ liệu được thủy vân số có thể phải trải qua nhiều loại biến đổi khác nhau, dấu thủy vân phải có tính bền vững mới có thể chịu được các pháp biến đổi ảnh cũng như biến đổi tín hiệu số thành tín hiệu tương tự. Ngoài ra dữ liệu chứa thủy vân phải chịu các phép biến đổi hình học như di chuyển vị trí, co dãn kích thước, cắt xén. Thủy vân đạt được tính bền vững thực sự khi: dấu thủy vân vẫn còn trong dữ liệu sau khi biến đổi và trích xuất vẫn có thể phát hiện ra thủy vân. Ví dụ, dấu thủy vân vẫn còn tồn tại trong dữ liệu nhưng khi trích xuất lại không thể phát hiện để đưa ra thủy vân. Trong trường hợp không xác định rõ phép biến đổi để thực hiện biến đổi ngược thì trích xuất không thể phát hiện đưa ra thủy vân mặc dù thủy vân vẫn tồn tại.

*Tính dễ hòng*: là thuộc tính ngược hoàn toàn với tính bền vững của thủy vân. Thuộc tính này được dùng trong các lược đồ thủy vân. Lược đồ này yêu cầu đặt ra là dấu thủy vân hoặc bị phá hủy bởi cứ phương pháp sao chép nào ngoại trừ phương pháp sao chép hợp pháp.

*Tính dư thừa*: Tính dư thùa liên quan đến một thực tế là thủy vân được lặp lại ở những vùng tần số khác nhau. Do đó nếu có một lỗi trên một vùng tần số thì vẫn có thể khôi phục thông điệp từ các dải tần khác.

*Độ phức tạp tính toán*: Độ phức tạp tính toán của lược đồ thủy vân đều rất quan trọng. Điều này đúng khi xử lý các dữ liệu thời gian thực. Mặt khác, cần phải xem xét tính co dãn của độ phức tạp tính toán. Ví dụ, lược đồ thủy vân thế hệ đầu tiên có độ phức tạp tính toán không lớn nhưng độ tin cật không cao so với lược đồ thủy vân thế hệ tiếp theo. Nhưng khi giải quyết một số vấn đề tính toán lớn thì lược đồ thủy vân ở thế hệ sau lại làm việc tốn hơn.

**CHƯƠNG II. MỘT SỐ KỸ THUẬT THỦY VÂN SỐ CƠ SỞ DỮ LIỆU**

1. **Thủy vân số biến dạng (Distortion-based Watermarking)**
   1. ***Thủy vân số dựa trên thuộc tính kiểu dữ liệu số***
      1. *Sử dụng nhóm bit vô nghĩa làm thông tin hình mờ*

Các phương án watermarking được đề xuất bởi [Agrawal et all] (còn được gọi là thuật toán AHK) dửa trên thuộc tính kiểu dữ liệu số và việc đánh dấu được thực hiện ở mức bit. Việc đánh dấu ở đây tức là chọn ra các bộ, các thuộc tính và giá trị tương ứng với các bộ, các thuộc tính đó. Sau đó, ta sẽ thay đổi các bit ít ý nghữa nhất của giá trị đó. Những thay đổi này sẽ tạo thành thủy vân. Ý tưởng cơ bản của kỹ thuật này là đảm bảo tại một số vị trí bit của một số thuộc tính trong một số bộ có chứa giá trị nhất định. Các bộ, các thuộc tính trong bộ các vị trí bit trong bộ thuộc tính và các giá trị bit nhất định này đều phải được kiểm soát ở một khóa bí mật K và các tham số riêng  γ, ν, ξ của chủ sở hữu quan hệ. Mẫu bit này sẽ hình thành ra thủy vân, Chỉ duy nhất chủ sở hữu của khóa bí mật này mới có thủy tìm lại được thủy vân với xác suất cao. Các tham số γ, ν, ξ và K đại diện cho số bộ giá trị cần đánh dấu, số lượng thuộc tính có sẵn để đánh dấu, số bit quan trọng nhất có sẵn để đánh dấu trong một thuộc tính và khóa bí mật tương ứng. Mật mã MAC với hàm H (K || H (K || r.P)) trong đó r.P là khóa chính của bộ r và || đại diện cho hoạt động nối, được sử dụng để xác định vị trí bit ứng viên. Hàm HASH H (K || r.P) được sử dụng để xác định các giá trị bit được nhúng tại các vị trí đó. Sự lựa chọn MAC và HASH là do một chiều đặc điểm chức năng và xác suất va chạm ít hơn.

* + 1. *Sử dụng hình ảnh làm thông tin hình mờ*

Wang và cộng sự 2008a, mô tả sơ đồ đánh dấu bằng hình ảnh dựa trên hình ảnh thay vì nhúng hình ảnh gốc làm hình mờ. Trong giai đoạn nhúng, ảnh gốc có kích thước N x N đầu tiên được chuyển đổi thành hình ảnh xáo trộn, sau đó được biểu diễn bằng một chuỗi nhị phân *b s* có độ dài L = N x N. Thứ hai, tất cả các bộ giá trị trong mối quan hệ được nhóm lại thành L nhóm. Giá trị hàm băm được tính bằng khóa chính của bộ giá trị, bí mật khóa và thứ tự của ảnh, xác định nhóm mà mỗi bộ thuộc về. Cuối cùng, bit thứ i của *b s* được nhúng vào vị trí bit được chọn theo thuật toán của giá trị thuộc tính cho các bộ giá trị trong nhóm thứ i đáp ứng một tiêu chuẩn. Giai đoạn phát hiện tuân theo kỹ thuật biểu quyết đa số. Tuy nhiên, bảo mật của lược đồ này được cải thiện vì nó không chỉ dựa vào khóa bí mật mà còn số xáo trộn *d* và thứ tự của hình ảnh *N* . Thay vì nhúng hình ảnh xáo trộn, kỹ thuật watermarking nhúng hình ảnh gốc bằng cách chuyển đổi nó thành một luồng bit có độ dài nhất định.

* + 1. *Sử dụng đặc điểm nội dung làm thông tin hình mờ*

Giai đoạn chèn hình mờ sẽ trích xuất một số bit, được gọi là đặc tính cục bộ, từ thuộc tính đặc trưng A1 của bộ giá trị t và nhúng các bit đó vào thuộc tính hình mờ A2 của cùng một bộ giá trị. Việc lựa chọn các bộ giá trị phụ thuộc vào việc giá trị ngẫu nhiên được tạo ra ( từ 0 đến 1) nỏ hơn tỉ lệ nhúngα của cơ sở dữ liệu quan hệ và yêu cầu not NULL của giá trị thuộc tính đặc trưng. Trong phát hiện hình mờ, bằng cách làm theo quy trình tương tự, đặc tính cục bộ của thuộc tính đặc trưng được trích xuất và so sánh với các bit cuối cùng của hình mờ.

Guo và cộng sự 2006b, các tác giả đề xuất một lương đồ thủy vân dễ vỡ có thể xác minh tính toàn vẹn của của quan hệ cơ sở dữ liệu. Trong kế hoạch được đề xuất, tất cả các bộ giá trị trong một quan hệ cơ sở dữ liệu trước tiên được chia thành các nhóm và được sắp xếp một cách an toàn.

- Trong mỗi nhóm, có hai loại hình mờ đươc nhúng: thuộc tính hình mờ W1 bao gồm γ hình mờ có độ dài ν và hình chìm nhiều W2 bao gồm ν hình mờ có độ dài γ. Trong đó γ và ν là số thuộc tính trong bộ giá trị và số lượng bộ giá trị trung bình trong mỗi nhóm tương ứng.

- W1 và W2 được tạo ra bằng cách trích xuất chuỗi bit từ giá trị hàm băm. Vì hình mờ thuộc tính W1, giá trị hàm băm được tạo theo thông báo mã xác thực cùng một thuộc tính của tất cả các bộ giá trị trong một nhóm trong khi đổi với bộ giá trị hình mờ W2, nó được hình thành từ cùng một mã xác thực tin nhắn và tất cả thuộc tính của cùng một bộ giá trị. Hãy quan sát điều đó, trong quá trình nhúng và giai đoạn phát hiện, chúng bỏ qua ít nhất hai bit quan trọng của tất cả các thuộc tính của kiểu số ngoại trừ khóa chính khi tính toán các giá trị hàm băm. Thuộc tính hình mờ được nhúng ở mức LSB, trong khi các bộ giá trị hình mờ được nhúng ở bên cạnh mức LSB. Bằng cách này, các hình mờ được nhúng thực sự tạo thành lưới hình mờ, giúp phát hiện, xác định vị trí và mô tả đặc điểm của các sửa đổi.

* 1. ***Thủy vân số dựa trên thuộc tính kiểu dữ liệu phân loại***

Không giống như các sơ đồ thủy vân trên, nơi đánh dấu watermark dựa trên thuộc tính số. Do Sion và cộng sự đề xuất dựa trên dữ liệu kiểu phân loại. Quá trình nhúng watermark bắt đầu với một mối quan hệ có ít nhất một thuộc tính phân loại A (dùng để tạp hình mờ). Hình mờ *wm*, bộ khóa bí mật *(k1, k2)* và các tham số khác như e để xác định phần trăm bộ giá trị cần đánh dấu. Sử dụng khóa chính K và khóa bí mật *k1* và các tham số *e*, nó phát hiện ra một tập hợp bộ giá trị “vừa vặn” được sử dụng để mã hóa nhãn hiệu. Quá trình lựa các bộ giá trị phù hợp là giống như thuật toán AHK. Giả sử quan hệ cơ sở dữ liệu có  η bộ giá trị, sau đó khớp các bộ giá trị tập chứa trong khoảng  *η / e* bộ giá trị. *Vm* hình mờ ngắn hơn được chuyển thành *dữ liệu vm* có độ dài bằng  *η / e* bằng cách triển khai mã sửa lỗi (ECC).

Thuật toán đánh dấu tạo ra một giá trị bí mật của số bit cần thiết để đại diện cho tất cả các giá trị phân loại cso thể có cho thuộc tính A tùy thuộc vào khóa chính và *k1*, sau đó, buộc bit quan trọng nhất của nó thành một giá trị đến vị trí tương ứng (ngẫu nhiên, tùy thuộc vào khóa chính và khóa *k2*). Trong dữ liệu wm, tính chất giả ngẫu nhiên của hàm băm *H ( T i ( K ) , k 2 ))* trung bình đảm bảo rằng phần lớn các bit trong *dữ liệu vm* sẽ được nhúng ít nhất một lần. Việc sử dụng các phím *k1* và *k2* khác nhau đảm bảo rằng không có mối tương quan giữa các bộ giá trị đã chọn để nhúng (được chọn bởi *k1*) và các giá trị bit tương ứng trong *dữ liệu vm* (được chọn bởi *k2*). Họ cũng đề xuất thực hiện nhúng dựa trên nhiều thuộc tính phân loại bằng cách xem xét không chỉ mối liên hệ giữa khóa và thuộc tính phân loại đơn A mà tất cả các mối liên kết giữa khóa chính và thuộc tính phân loại các thuộc tính để tăng tính mạnh mẽ của chương trình. Mặc dù kế hoạch này được tuyên bố mạnh mẽ để chống lại các cuộc tấn công nghiêm trọng (*ví dụ*: các cuộc tấn công ngẫu nhiên), tuy nhiên, kế hoạch không phù hợp với các quan hệ cơ sở dữ liệu cần cập nhập thường xuyên, vì nó rất tốn kém để đánh dấu lại các quan hệ cơ sở dữ liệu được cập nhập. Dù chỉ nhỏ một phần của các bộ giá trị đã chọn bị ảnh hưởng bởi việc nhúng hình mờ, các sửa đổi thuộc tính phân loại (*ví dụ*: thay đổi từ “đỏ” sang “xanh lam”) trong một số ứng dụng có thể quá lớn để có thể chấp nhận được. Kỹ thuật watermarking này được áp dụng cho dữ liệu y tế được phân loại theo thứ bậc.

* 1. ***Thủy vân số dựa trên bộ giá trị chèn thuộc tính***
     1. *Sử dụng bộ giá trị làm thông tin hình mờ*

Cách tiếp cận trong [Pournaghshband, 2008] nhằm mục đích tạo ra các bộ giá trị giả và chèn nhúng vào cơ sở dữ liệu một các sai lầm. Thuật toán tạo bộ giá trị giả sử dụng quan tâm đến các thuộc tính chính của ứng viên và mức dộ nhạy cảm của các thuộc tính không phải ứng viên. Người sử dụng xác suất lấy mẫu Bernoulli pi cho thuộc tính không phải ứng của viên thứ *i* . *Ai* để quyết định giá trị giả của nó có thể được chọn đồng nhất hoặc giá trị với tần suất xuất hiện cao hơn trong tập giá trị hiện có của *Ai* trong mối quan hệ.

Không giống như các thuật toán khác, thuật toán phát hiện không phải là thuật toán nghịch đảo để thuật toán tạo hình mờ và thuật toán chèn có tính xác suất trong thiên nhiên. Thuật toán phát hiện kiểm tra xam tài liệu các bộ giá trị giả được chèn vào trong giai đoạn chèn watermark, tồn tại hoặc đã được thay đổi. Nó kiểm tra thông qua chính chìa khóa. Ngay sau khi nó tìm thấy một kết quả trùng khớp (tức là các bộ giá trị giống hệt nhau hoặc tương tự), việc phát hiện sẽ xong. Việc phát hiện sẽ không thành công đối với cơ sở dữ liệu hình mờ khi tất cả đề là giả các bộ giá trị được xóa bằng cách xóa lành tính. Số lượng bộ giá trị giả sẽ được chèn vào được quyết định bởi chủ sở hữu cơ sở dữ liệu. Tuy nhiên, giai đoạn chèn hình mờ phải lưu ý đến thực tế là các giá trị của dấu bộ giá trị giả không nên bằng mọi cách làm giảm chất lượng của dữ liệu trong cơ sở dữ liệu và không nên tác động đến kết quả truy vấn. Một lợi thế của chương trình này là quyền sở hữu có thể được xác minh công khai nhiều lần cho đến khi tất cả các bộ giá trị giả được tiết lộ và chương trình không bị tăng khả năng cập nhật.

* + 1. *Sử dụng thuộc tính ảo làm thông tin hình mờ*

Thay vì chèn các bộ giá trị , Prasannakumari, 2009 đề xuất một kỹ thuật đánh dấu watermark bằng cách chèn một thuộc tính ản trong quan hệ cơ sở dữ liệu sẽ đóng vai trò là hình mờ chứa tổng kiểm tra chãn lẻ của tất cả các thuộc tính và giá trị tổng hợp thu được bất kỳ một trong các thuộc tính số của tất cả các bộ giá trị. Quá trình chèn thuộc tính ảo được thực hiện độc lập cho mỗi phân vùng không chồng chéo thu được từ quan hệ ban đầu. Lược đồ này được thiết kế để xác thực biên nhận chống giả mạo của cơ sở dữ liệu qua một kênh liên lạc không an toàn. Mặc dù cách tiếp cận này là mong manh và có thể dễ dàng phát hiện bất kỳ cuộc tấn công xóa hoặc chèn hoặc thay đổi nào.

1. **Thủy vân số không biến dạng (Distortion-free Watermarking)**

Hầu hết các kỹ thuật watermarking không bị biến dạng đều mong manh theo nghĩa là ngoài việc xác nhận quyền sở hữu, họ còn hướng đến việc duy trì tính toàn vẹn của thông tin trong cơ sở dữ liệu. Giai đoạn chèn hình mờ không phụ thuộc trền bất kỳ loại thuojc tính cụ thể nào và không gây ra bất kỳ biến dạng nào trong dữ liệu cơ bản của cơ sở dữ liệu.

* 1. ***Trích xuất giá trị hàm băm dưới dạng thông tin mờ***

Để đạt được mục đích của hình mờ mỏng manh, các tác giả trong [Li và cộng sự 2004] và [Bhattacharya và Cortesi, 2009a] đề xuất các phướng án đánh dấu watermarking có thể phát hiện bất kỳ sửa đổi nào được thực hiện đối với quan hệ cơ sở dữ liệu. Những kế họach được thiết kế cho dữ liệu phân loại không thể chịu được sự biến dạng, do đó, loại dữ liệu này cần phải dùng watermark không gây ra biến dạng. Trong “Li et al, 2004”, phân vùng các bộ giá trị là dựa trên giá trị băm được tham số hóa với khóa chính và khóa bí mật, nhưng trong “Bhattacharya and Cortesi, 2009a”  phân vùng dựa trên các giá trị thuộc tính phân loại. Sau khi phân vùng, các giá trị băm cấp bộ và cấp nhóm cho mỗi nhóm được tính toán. Trong “Li et al, 2004”, thủy vân số có chiều dài bằng số lượng cặp tuple trong nhóm, được trích xuất từ ​​giá trị băm cấp nhóm và đối với mỗi cặp tuple, thứ tự của hai tuple này sẽ thay đổi hoặc không thay đổi tùy theo giá trị băm tuple của chúng và bit watermark tương ứng. Hơn nữa, Li “Li, 2007” đề xuất thực hiện việc hoán đổi vị trí của các bộ giá trị dựa trên thuật  toán hủy xếp hạng hoán vị tuyến tính Myrvold và Ruskeys để tăng khả năng nhúng. Trong các sơ đồ thiết kề này, bất kỳ sửa đổi nào của giá trị thuộc tính đều ảnh hưởng đến hình mờ trong hai nhóm vì tuple đã sửa đổi có thể bị xóa từ một nhóm và được thêm vào nhóm khác.

* 1. ***Kết hợp dấu hiệu của chủ sở hữu và đặc tính của cơ sở dữ liệu***

Đề án được đề xuất bởi Tsai et al. “Tsai et al, 2006” nhằm duy trì tính toàn vẹn của thông tin trong cơ sở dữ liệu và dựa trên xác thực công khai cơ chế. Ý tưởng đằng sau kế hoạch này là, đầu tiên là một hình mờ *W*   được tạo ra là một  √n × √n hình ảnh trắng, ở đâu *n* là số lượng tuple  trong quan hệ, ngoài bốn góc có dấu của chủ sở hữu. Nó tạo ra một giá trị *Ci* (0 ≤ *Ci* ≤ 255) cho mỗi tuple *ti* trong cơ sở dữ liệu bằng cách sử dụng hàm băm MD5 và hoạt động XOR. Nếu có *n* bộ dữ liệu trong cơ sở dữ liệu, nó tạo ra một đặc tính *C* chiều dài *n* bằng cách kết hợp tất cả Ci theo thứ tự. Cuối cùng, một mã chứng nhận *R* tạo ra bằng cách XOR *C* và *W*. Dạng mã hóa của *R* sử dụng khóa riêng tư được cung cấp công khai. Trong quá trình xác minh tính toàn vẹn của mối quan hệ *T* ′, theo cách tương tự, nó tạo ra đặc tính *C*′ từ *T* ′. Sau khi giải mã bằng khóa công khai, mã chứng nhận *R* được thực hiện XOR với *C*′, từ đó có được hình mờ *W*′. Tính toàn vẹn của hình mờ được trích xuất này sẽ chứng minh tính toàn vẹn của cơ sở dữ liệu.

* 1. ***Hoán vị dựa trên R-tree***

Trái ngược với các lược đồ thủy vân số truyền thống, kỹ thuật thủy vân số dựa trên cấu trúc dữ liệu R-tree hay còn gọi là cây quan hệ đã được đề xuất trong “Kamel, 2009”. Kỹ thuật được đề xuất tận dụng lợi thế của thực tế  R-tree không đặt các điều kiện về thứ tự của các mục bên trong nút. Trong lược đồ được đề xuất, các mục nhập bên trong các nút R-tree được sắp xếp lại, tùy theo thứ tự ban đầu bí mật (khóa bí mật), theo cách tương ứng với giá trị của hình mờ. Để đạt được điều đó, họ đề xuất ánh xạ 1-1 giữa tất cả các hoán vị có thể có của các mục trong nút R-tree và tất cả các giá trị có thể có của hình mờ. Không làm mất tính tổng quát, hình mờ được giả định là các giá trị số. Ánh xạ được đề xuất sử dụng một hệ thống đánh số sử dụng cơ sở biến với giá trị giai thừa. Tỷ lệ phát hiện các cuộc tấn công độc hại phụ thuộc vào bản chất của cuộc tấn công, phân phối dữ liệu và kích thước của nút R-tree. Kỹ thuật watermarking được đề xuất có các tính năng mong muốn sau:

- Nó không thay đổi các giá trị của dữ liệu trong nút R-tree mà ẩn hình mờ theo thứ tự tương đối của các mục bên trong nút R-tree.

- Chi phí hiệu suất là tối thiểu.

- Việc kiểm tra tính toàn vẹn không yêu cầu kiến thức về dữ liệu chưa được đánh dấu.

**CHƯƠNG III. THỰC NGHIỆM**

1. **Cơ sở lý thuyết**

Quá trình thực nghiệm dựa trên cơ sở dữ liệu quan hệ MySQL với dữ liệu dưới dạng float. Cơ sở dữ liệu này được đánh dấu nước thông qua kỹ thuật thủy vân số biến dạng, sử dụng nhóm bit vô nghĩa để chèn thông tin hình mờ.

Để tìm kiếm nhóm bit phù hợp trong cơ sở dữ liệu, ta thực hiện đối với mỗi bộ giá trị tuple trong cơ sở dữ liệu

* Tính toán giá trị hàm băm r.MAC = H(K o r.P) , trong đó MAC là từ viết tắt của Message Authentication Code, K là khóa riêng tư, r.P là khóa chính của tuple đó
* Khởi tạo các tham biến m – vị trí lựa chọn để đánh dấu CSDL theo khóa chính , v – lựa chọn thuộc tính chứa giá trị, e – lựa chọn vị trí bit
* Khi tìm được vị trí của các giá trị phù hợp, ta thực hiện thay đổi một bit là 1 nếu như r.MAC là số chắn và 0 nếu r.MAC là số lẻ

Thuật toán để tìm kiếm như sau

match-ori = 0

for each t in D:

r.MAC = H(K ο r.P)

if (r.MAC mod m) == 0: // Select primary key

match\_ori ++

i = r.MAC mod v // Select attribute

b = r.MAC mod e // Select bit

if r.MAC mod 2 == 0

set bit b of r.Ai

else

clear bit b of r.Ai

Bởi đây là kỹ thuật thủy vân số dựa trên kỹ thuật thủy vân số biến dạng, nên dữ liệu sẽ bị biến đổi một khoảng so với dữ liệu ban đầu. Do vậy để xác minh được chủ sở hữu của một CSDL ta thực hiện cùng một thuật tóa thủy vân trên A nhưng tại đây, thay vì biến đổi giá trị của bit thì kiểm tra tại vị trí bit đó có bị thay đổi hay không , và đếm số lượng bản ghi bị thay đổi bit, sau đó sao sánh với số bản ghi bị thay đổi dựa trên CSDL gốc. Nếu cả 2 giá trị này giống nhau hoặc bị xê dịch không đáng kể thì có thể chứng minh A là cơ sở dữ liệu gốc.

match = 0

for each t in D:

r.MAC = H(K ο r.P)

if (r.MAC mod m) == 0: // Select primary key

i = r.MAC mod v // Select attribute

b = r.MAC mod e // Select bit

if r.MAC mod 2 == 0

match ++

else

match ++

if match == match\_ori:

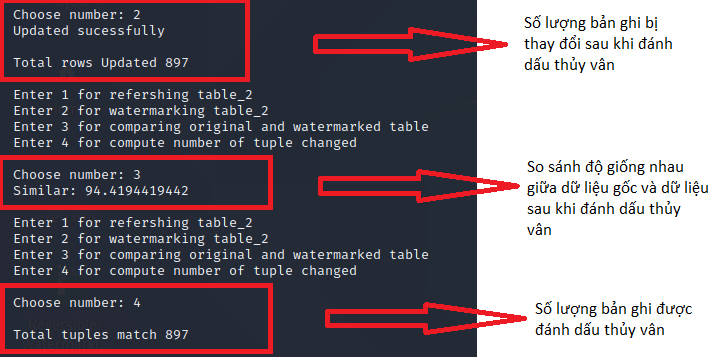
verifycation successfully

1. **Kết quả thực nghiệm**

Các giá trị biến được lựa chọn như sau:

|  |  |  |
| --- | --- | --- |
| STT | Tên biến | Giá trị |
| 1 | m | 11 |
| 2 | v | 16 |
| 3 | e | 3 |

Kết quả thực hiện thủy vân số biến dạng:



**TÀI LIỆU THAM KHẢO**

[Agrawal et al., 2003a] Agrawal, R., Haas, P. J., and Kiernan, J. (2003a). A system for watermarking relational databases. In Proceedings of the 2003 ACM SIGMOD international conference on Management of data (SIGMOD ’03), pages 674–674, San Diego, California. ACM Press

[Bertino et al., 2005] Bertino, E., Ooi, B. C., Yang, Y., and Deng, R. H. (2005). Privacy and ownership preserving of outsourced medical data. In Proceedings of the 21st International Conference on Data Engineering (ICDE ’05), pages 521–532, Tokyo, Japan. IEEE Computer Society

[Bhattacharya and Cortesi, 2009a] Bhattacharya, S. and Cortesi, A. (2009a). A distortion free watermark framework for relational databases. In Proceedings of the 4th International Conference on Software and Data Technologies (ICSOFT ’09), pages 229–234, Sofia, Bulgaria. INSTICC Press.

[Li et al., 2004] Li, Y., Guo, H., and Jajodia, S. (2004). Tamper detection and localization for categorical data using fragile watermarks. In Proceedings of the 4th ACM workshop on Digital rights management (DRM ’04), pages 73–82, Washington, DC, USA. ACM Press.

[1] Raju Halder, Agostino Cortesi. Watermaking Techniques for Relational Databases: Survey, Classification and Comparison. Journal of Universal Computer Science · January 2010.

[2] Yaxiao Song, Watermaking of Relational Databases. B.E Harbin Institute of Technology, China. 2004.