

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

NGUYỄN MINH KHÔI

Chuyên ngành : KHOA HỌC MÁY TÍNH

Mã số : 60.48.01

TÌM HIỂU KỸ THUẬT NHÚNG
THÔNG TIN ĐỂ BẢO VỆ BẢN QUYỀN CƠ SỞ
DỮ LIỆU QUAN HỆ

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

NGƯỜI HƯỚNG DẪN KHOA HỌC:

PGS.TS BÙI THỂ HỒNG

Thái Nguyên, năm 2013

LỜI CẢM ƠN

Luận văn được thực hiện tại trường Đại học Công nghệ Thông tin và Truyền Thông – Đại học Thái Nguyên dưới sự hướng dẫn của PGS. TS **Bùi Thế Hồng**.

Trước hết, tôi xin bày tỏ lòng biết ơn sâu sắc tới thầy Bùi Thế Hồng, người đã có những định hướng, những kiến thức quý báu, những lời động viên và chỉ bảo giúp tôi vượt qua những khó khăn để tôi hoàn thành tốt luận văn của mình.

Tôi xin được bày tỏ lòng cảm ơn và sự kính trọng của mình đến các thầy cô giáo Trường Đại học Công nghệ Thông tin và Truyền Thông, Đại học Thái Nguyên, đặc biệt là các thầy cô giáo đã giảng dạy và giúp đỡ tôi trong suốt quá trình học tập tại trường.

Tôi cũng đặc biệt cảm ơn tới bạn bè lớp Cao học K9D, các đồng nghiệp tại Trường Đại Lương Thế Vinh tỉnh Nam Định, đã luôn động viên, giúp đỡ tôi trong quá trình học tập và công tác, để tôi hoàn thành nhiệm vụ được giao.

Nhân dịp này, tôi cũng xin gửi lời cảm ơn đến gia đình, người thân, đã tạo điều kiện giúp đỡ, động viên, trợ giúp tôi về tinh thần trong suốt quá trình học tập.

Thái Nguyên, ngày 10 tháng 03 năm 2013

Tác giả

Nguyễn Minh Khôi

LỜI CAM ĐOAN

Tôi xin cam đoan luận văn “*Tìm hiểu kỹ thuật nhúng thông tin để bảo vệ bản quyền cơ sở dữ liệu quan hệ*” này là công trình nghiên cứu của riêng tôi. Các số liệu sử dụng trong luận văn là trung thực. Các kết quả nghiên cứu được trình bày trong luận văn chưa từng được công bố tại bất kỳ công trình nghiên cứu nào khác.

Nguyễn Minh Khôi

MỤC LỤC

LỜI CẢM ƠN.....	i
LỜI CAM ĐOAN.....	iii
DANH MỤC CÁC HÌNH VẼ.....	vi
1. Đặt vấn đề.....	1
2. Đối tượng và phạm vi nghiên cứu.....	2
3. Hướng nghiên cứu của đề tài.	2
4. Những nội dung nghiên cứu chính.....	3
5. Phương pháp nghiên cứu.....	3
6. Ý nghĩa khoa học của đề tài.	3
CHƯƠNG I: TÌM HIỂU VỀ BẢN QUYỀN SẢN PHẨM SỐ VÀ KỸ THUẬT NHÚNG TIN.....	5
1.1 Bản quyền sản phẩm số.....	5
1.2. Kỹ thuật nhúng thông tin và ứng dụng	7
1.2.1 Kỹ thuật nhúng thông tin	7
1.2.2 Phân loại các kỹ thuật nhúng tin	8
1.3 Ứng dụng của kỹ thuật nhúng tin.....	10
1.4 Môi trường nhúng tin	12
1.4.1 Trong văn bản	12
1.4.3 Trong audio	13
1.4.4 Trong video	14
1.4.5 Nhúng thông tin vào cơ sở dữ liệu quan hệ.	15
1.5 Tình hình nghiên cứu kỹ thuật nhúng tin vào cơ sở dữ liệu	16
1.5.1 Theo kiểu dữ liệu (Data type)	17
1.5.2 Theo kiểu biến dạng (Distortion).....	18
1.5.3 Theo độ nhạy (Sensitivity).....	18
1.5.6 Theo thông tin thủy vân (watermark information)	18
1.5.7 Tính kiểm tra được	19
1.5.8 Theo cấu trúc dữ liệu (Data structure)	20
1.5.9 Xử lý dữ liệu theo khối (Data cube).....	21
1.5.10 Xử lý theo dòng dữ liệu (Streaming data).....	21
CHƯƠNG II: NHÚNG THÔNG TIN ĐỂ BẢO VỆ BẢN QUYỀN CƠ SỞ DỮ LIỆU.....	22
2.1. Cơ sở của giải pháp	22
2.1.1 Nhúng tin vào ảnh số.....	22

2.1.1.1	Nhúng tin bằng kỹ thuật thay thế bit có trọng số thấp nhất.	22
2.1.1.2	Các phép biến đổi miền không gian ảnh sang miền tần số.....	26
2.1.1.3	Kỹ thuật sinh chuỗi giả ngẫu nhiên	30
2.1.1.4	Các kỹ thuật trải phổ trong truyền thông	31
2.2.	Giải thuật nhúng đề xuất	33
2.3.	Giải thuật giải nhúng	41
2.4	Bảng hợp trước các loại tấn công.....	43
2.4.1	Xóa ngẫu nhiên bản ghi	44
2.4.2	Sửa đổi bảng ghi	45
2.4.3	Chèn bản ghi	47
	CHƯƠNG III: THỬ NGHIỆM GIẢI THUẬT ĐỀ XUẤT.....	48
3.1	Dữ liệu cho thử nghiệm.....	48
3.2	Xây dựng chương trình thử nghiệm.....	50
3.2.1	Tổ chức, cài đặt giải thuật	51
3.2.2	Hoạt động của chương trình.....	52
	KẾT LUẬN	55
	PHỤ LỤC.....	57
1.	Mã xác thực thông tin (MAC)	57
2.	Hàm băm	58
	TÀI LIỆU THAM KHẢO:	63

DANH MỤC CÁC KÝ HIỆU VÀ MỤC VIẾT TẮT

LSB	Least Significant Bit
DFT	Discrete Fourier Transform
DCT	Discrete Cosine Transform
DWT	Discrete Wavelet Transform
IDWT	Inverse Discrete Cosine Transform
HVS	Human Visual System
CSDL	Cơ sở dữ liệu quan hệ

DANH MỤC CÁC HÌNH VẼ

Hình 1.1. Sơ đồ biểu diễn quá trình nhúng tin.....	8
Hình 1.2. Sơ đồ biểu diễn quá trình giải mã.....	8
Hình 1.3. Phân loại kỹ thuật nhúng tin.....	9
Hình 2.1: Minh họa thủy vân ảnh số.....	22
Hình 2.1.1: Minh họa cách nhúng tin bằng kỹ thuật LSB.....	25
Hình 2.2.1 Minh họa dữ liệu thông tin sinh viên.....	34
Hình 2.2.2 Minh họa dữ liệu kết quả học tập của sinh viên.....	35
Hình 2.2.3 Bảng hợp lưu dữ liệu để nhúng tin.....	35
Hình 2.2.4: Hình mô phỏng bảng hợp chứa thông tin tìm kiếm, xác định dữ liệu.....	36
Hình: 2.2.5: Minh họa bảng hợp vào các bit sẽ nhúng.....	39
Hình 2.2.6: Bảng sinh viên sau khi nhúng tin.....	40
Hình 2.2.7 Bảng Kết quả học tập sau khi nhúng tin.....	40
Hình 2.3.1: Bảng bit sau giải nhúng.....	43
Hình 2.4.1.1: Bảng kết quả học tập sau khi bị xóa một số bản ghi.....	45
Hình 2.4.1.2: Bảng bit thu được từ giải nhúng sau khi CSDL bị xóa bản ghi.....	45
Hình 2.4.2.1: Bảng kết quả học tập sau khi bị sửa một số bản ghi.....	46
Hình 2.4.2.2: Bảng bit thu được từ giải nhúng sau khi CSDL bị sửa bản ghi.....	46
Hình 3.1.1: Biểu diễn tổ chức, lưu trữ thông tin trong CSDL thử nghiệm.....	48
Hình 3.1.2: Mối liên giữa bảng dữ liệu sinh viên và bảng điểm kết quả học tập.....	50
Hình 3.2.1: Giao diện kết nối tới CSDL.....	52
Hình 3.2.2: Giao diện chọn bảng.....	53
Hình 3.2.3: Giao diện xác định các trường lấy thông tin.....	54
Hình 3.2.4 : Thao tác nhúng và tách nhúng.....	54
Hình 4.1. Sơ đồ hoạt động của MAC.....	57
Hình 4.2. Mô tả quá trình xác thực thông tin bằng MAC.....	58
Hình 4.3. Biểu diễn hàm băm.....	61

LỜI MỞ ĐẦU

1. Đặt vấn đề.

Với sự phát triển nhanh chóng của Internet, việc sao chép và phân phối sản phẩm số bất hợp pháp đang trở thành vấn đề nhức nhối với nhiều cơ quan, tổ chức, cá nhân. Trong khi dữ liệu số lại là đối tượng dễ dàng bị nhân rộng, thay đổi và sao chép. Do đó, nếu việc vi phạm bản quyền trở nên phổ biến, thì nó sẽ làm suy yếu giá trị của thương mại của các sản phẩm số này. Điều này đã tạo ra cuộc cách mạng trong lĩnh vực bảo vệ quan điểm của quyền sở hữu trí tuệ. Ngày nay, không chỉ ngành công nghiệp hình ảnh và âm nhạc (audio), cả các bộ phim (video) đang gặp phải vấn đề vi phạm bản quyền lớn, do các trang web cộng đồng chia sẻ tập tin và sự gia tăng của băng thông Internet. Vì vậy, quản lý quyền kỹ thuật số đang ngày càng là một vấn đề đáng quan tâm.

Nhúng thông tin là một trong những kỹ thuật giấu dữ liệu hiện đại. Nó được định nghĩa như là quá trình chèn thông tin vào dữ liệu nhưng bảo đảm không cảm thụ được, nghĩa là chỉ làm thay đổi nhỏ dữ liệu gốc. Nhúng thông tin đã được ứng dụng trong nhiều lĩnh vực như bảo vệ quyền sở hữu, điều khiển việc sao chép, xác nhận giấy tờ, hay truyền đạt thông tin khác, ... trong đó ứng dụng phổ biến của nó là cung cấp bằng chứng về bản quyền tác giả của các dữ liệu số bằng cách nhúng các thông tin bản quyền. Rõ ràng trong ứng dụng này, thông tin nhúng cần phải bền vững trước các thao tác nhằm loại bỏ chúng.

Trong thời đại bùng nổ thông tin hiện nay, cơ sở dữ liệu quan hệ gặp các nguy cơ tấn công là không thể tránh khỏi. Do lo ngại ngày càng tăng về việc vi phạm bản quyền cơ sở dữ liệu quan hệ, nên việc quản lý quyền kỹ thuật số của cơ sở dữ liệu đang trở thành một lĩnh vực nghiên cứu cực kỳ quan trọng. Gần đây đã được một số công trình tiên phong trong lĩnh vực này

giúp xác lập quyền sở hữu cơ sở dữ liệu quan hệ. Giải pháp được đưa ra chính là sử dụng kỹ thuật nhúng thông tin vào cơ sở dữ liệu để cá nhân, tổ chức xác nhận quyền làm chủ sở hữu cơ sở dữ liệu mình. Trước hết hãy xem xét kịch bản người mua-người bán chung một chủ sở hữu của cơ sở dữ liệu được bán cho nhiều khách hàng khác nhau. Đối với trường hợp này, rõ ràng vấn đề chính là xác định quyền sở hữu mỗi bản sao của các dữ liệu được bán như vậy là xác định nguồn gốc chính xác và đáng tin cậy của cơ sở dữ liệu quan hệ trong trường hợp vi phạm bản quyền. Cũng với cách thức này áp dụng cho tình huống công bố cơ sở dữ liệu và vấn đề xác minh thông tin trong cơ sở dữ liệu có bị xuyên tạc hay không. Vì vậy luận văn này sẽ tìm hiểu các kiểu tấn công nhằm vào cơ sở dữ liệu quan hệ, từng tình huống và tìm hiểu các cách giải quyết từng vấn đề. Tiếp đến luận văn sẽ tìm hiểu thuật toán nhúng thông tin vào cơ sở dữ liệu quan hệ và đưa ra minh họa.

2. Đối tượng và phạm vi nghiên cứu.

- Một số cách thức vi phạm bản quyền cơ sở dữ liệu quan hệ thường gặp.
- Nhúng thông tin vào cơ sở dữ liệu quan hệ là một lĩnh vực rộng lớn và mới mẻ luận văn này chỉ tìm hiểu các giải thuật cơ sở cho việc nhúng và giải nhúng.

–

3. Hướng nghiên cứu của đề tài.

- Tìm hiểu về cơ sở dữ liệu quan hệ.
- Tìm hiểu mô hình nhúng thông tin cơ sở dữ liệu quan hệ.
- Tìm hiểu các giải thuật liên quan.

4. Những nội dung nghiên cứu chính.

CHƯƠNG I: TÌM HIỂU VỀ BẢN QUYỀN SẢN PHẨM SỐ VÀ KỸ THUẬT NHÚNG TIN

CHƯƠNG II: NHÚNG THÔNG TIN ĐỂ BẢO VỆ BẢN QUYỀN CƠ SỞ DỮ LIỆU

CHƯƠNG III: THỬ NGHIỆM GIẢI THUẬT ĐỀ XUẤT

KẾT LUẬN

5. Phương pháp nghiên cứu.

– Về lý thuyết:

+ Đọc tài liệu để tìm hiểu về cơ sở dữ liệu quan hệ, kỹ thuật nhúng thông tin bảo vệ bản quyền sản phẩm số nói chung và cơ sở dữ liệu quan hệ nói riêng.

+ Tìm hiểu cơ sở toán học và phương pháp xây dựng mô hình cho bài toán nhúng thông tin để bảo vệ bản quyền cơ sở dữ liệu quan hệ.

– Về thực nghiệm:

+ Áp dụng các thuật toán trên cho mô hình cơ sở dữ liệu để minh họa.

6. Ý nghĩa khoa học của đề tài.

– Trong thời đại ngày nay việc an toàn quyền sở hữu thông tin là một nhu cầu cấp bách không thể xa rời. An toàn sở hữu thông tin ở đây chính là bảo vệ thông tin không bị sao chép trái phép, không bị thay đổi ở mọi hình thức xuyên tạc hay bóp méo sự thật. Mà quan trọng hơn cả là khẳng định quyền làm chủ thông tin, mà đặc biệt hơn nữa trong quyền sở hữu thông tin chính là quyền sở hữu cơ sở dữ liệu quan hệ. Bởi vì mọi cá nhân, cơ quan, tổ chức,... tồn tại trong xã hội thông tin thì thông tin của mình đều được lưu trữ trong cơ sở dữ liệu quan hệ. Một khi cơ sở dữ liệu đó bị xâm phạm, bị biến đổi mà ta không thể nào chứng thực lại thông tin gốc trong cơ sở dữ liệu thì điều