

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

NGUYỄN THỊ PHƯƠNG CHI

**KỸ THUẬT GIẤU TIN
TRONG TỆP VĂN BẢN VÀ ỨNG DỤNG**

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

THÁI NGUYÊN, 2017

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

NGUYỄN THỊ PHƯƠNG CHI

**KỸ THUẬT GIẤU TIN
TRONG TỆP VĂN BẢN VÀ ỨNG DỤNG**

Chuyên ngành: Khoa học máy tính

Mã số: 60 48 01 01

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Người hướng dẫn khoa học: PGS.TS. ĐẶNG VĂN ĐỨC

THÁI NGUYÊN, 2017

LỜI CAM ĐOAN

Tôi xin cam đoan luận văn “*Kỹ thuật giấu tin trong tệp văn bản và ứng dụng*” là sản phẩm của riêng cá nhân, không sao chép lại của người khác. Trong toàn bộ nội dung của luận văn, những điều được trình bày hoặc là của cá nhân hoặc là được tổng hợp, nghiên cứu từ nhiều nguồn tài liệu. Tất cả các tài liệu tham khảo đều có xuất xứ và trích dẫn rõ ràng.

Tôi xin hoàn toàn chịu trách nhiệm và chịu mọi hình thức kỷ luật theo quy định cho lời cam đoan của mình.

Thái Nguyên, ngày 25 tháng 07 năm 2017

Học viên

Nguyễn Thị Phương Chi

LỜI CẢM ƠN

Lời đầu tiên, tôi xin bày tỏ lòng biết ơn đến PGS.TS. Đặng Văn Đức - Viện Công Nghệ Thông Tin, người đã tận tình hướng dẫn, chỉ bảo và giúp đỡ tôi trong suốt quá trình nghiên cứu và hoàn thành luận văn này.

Tôi xin chân thành cảm ơn các thầy cô giáo trường Đại học Công nghệ Thông tin và Truyền thông - Đại học Thái Nguyên đã giảng dạy và cung cấp cho chúng tôi những kiến thức rất bổ ích trong thời gian học cao học, giúp tôi có nền tảng tri thức để phục vụ nghiên cứu khoa học sau này.

Tôi cũng xin cảm ơn Lãnh đạo và đồng nghiệp tại đơn vị trường THPT Bãi Cháy nơi tôi công tác đã tạo điều kiện và giúp đỡ tôi trong suốt quá trình nghiên cứu và hoàn thành luận văn. Tôi cũng xin bày tỏ lòng cảm ơn đến gia đình và bạn bè, những người luôn quan tâm, động viên và khuyến khích tôi trong quá trình học tập.

Thái Nguyên, ngày 25 tháng 06 năm 2017

Nguyễn Thị Phương Chi

MỤC LỤC

MỞ ĐẦU	1
1. Nhu cầu bảo mật thông tin	1
2. Lý do lựa chọn đề tài	2
3. Mục tiêu nghiên cứu	2
4. Đối tượng và phạm vi nghiên cứu	3
5. Phương pháp nghiên cứu	3
6. Cấu trúc của luận văn	3
CHƯƠNG 1: TỔNG QUAN VỀ GIẤU TIN	5
1.1. Tổng quan về giấu tin	5
1.1.1 Vài nét về lịch sử giấu tin	5
1.1.2. Khái niệm cơ bản về giấu tin	6
1.1.3. Phân loại kỹ thuật giấu tin cơ bản	8
1.1.4. Mô hình kỹ thuật giấu tin	9
1.1.5. Môi trường giấu tin	13
1.1.6. Khả năng ứng dụng của giấu tin	14
1.1.7. Tổng kết chương 1	16
CHƯƠNG 2. CÁC KỸ THUẬT GIẤU TIN TRONG VĂN BẢN	17
2.1 Các loại bảng mã tiếng Việt	17
2.1.1. Bộ gõ Unicode	17
2.1.2. Bộ gõ VISCII	17
2.1.3. Bộ gõ TCVN3	18
2.1.4. Bộ gõ VNI	18
2.2 Các loại văn bản	18
2.2.1. Bảng mã ASCII	18
2.2.2. Siêu văn bản HTML	20

2.2.3 Văn bản theo chuẩn pdf	21
2.2.4 Văn bản theo chuẩn Microsoft Word (.doc hoặc .docx).....	22
2.3. Các tiệm cận của giấu tin trong văn bản.....	22
2.3.1. Watermarking trực tiếp trên văn bản	22
2.2.2. Watermarking trên văn bản đã định dạng.....	26
2.2.3. Watermarking trên tập tin hình ảnh quét từ văn bản	27
2.4. Giấu tin trong tệp MS Word	28
2.4.1. Lựa chọn thuộc tính giấu tin trong tệp MS Word.....	28
2.4.2 Mô tả thuật toán	32
2.4.3 Bảo mật	34
2.5. Tổng kết chương 2	36
CHƯƠNG 3. XÂY DỰNG CHƯƠNG TRÌNH GIẤU TIN TRONG	
TÀI LIỆU	37
3.1. Mô tả bài toán giấu tin trong tệp văn bản MS Word	37
3.2. Lựa chọn ngôn ngữ và cài đặt	37
3.3 Giao diện chương trình	37
3.3.1 Mẫu thiết kế	37
3.3.2. Quy trình công thức xử lý.....	38
3.4. Demo chương trình	38
3.4.1. Thông tin giấu	38
3.4.2. Giấu tin.....	39
3.4.3. Giải mã tin đã giấu.....	39
3.3.4. Trích xuất tin.....	40
3.5. Chạy thực nghiệm	42
3.6. Đánh giá kết quả đạt được.....	48
KẾT LUẬN VÀ KHUYẾN NGHỊ.....	49
TÀI LIỆU THAM KHẢO	50

DANH MỤC HÌNH VẼ

Hình 1.1. Lược đồ chung cho giấu tin	6
Hình 1.2. Kỹ thuật giấu tin.....	7
Hình 1.3. Một cách phân loại các kỹ thuật giấu tin	8
Hình 1.4. Lược đồ chung cho quá trình giấu tin.	12
Hình 1.5. Sơ đồ quá trình giải mã	12
Hình 2.1. mô tả nhúng thông tin bản quyền vào các khoảng trống sau mỗi dòng. ...	24
Hình 2.1.a Đoạn văn bản trước khi nhúng.....	24
Hình 2.1.b. Đoạn văn bản sau khi nhúng.....	24
Hình 2.2. Đoạn văn bản chứa dữ liệu nhúng sử dụng các khoảng trống sau mỗi từ.....	25
Hình 2.3. Ví dụ nhúng bản quyền bằng phương pháp cú pháp	25
Hình 2.4. Ví dụ về các cặp từ đồng nghĩa.....	26
Hình 2.5. Ví dụ về dịch chuyển dòng	26
Hình 2.6. Ví dụ dịch chuyển từ.....	27
Hình 2.7. Ví dụ về dịch chuyển ký tự.....	27
Hình 2.8. Ví dụ thay đổi đặc trưng.....	27
Hình 2.9. Ví dụ thay đổi màu sắc các ký tự.....	28
Hình 2.10. Ví dụ thay đổi màu sắc các ký tự.....	29
Hình 2.11. Ví dụ thay đổi khoảng trống các ký tự.....	30
Hình 3.1. Giao diện chương trình	37
Hình 3.2. Thông tin giấu	38
Hình 3.3. Giấu tin.....	39
Hình 3.4. Nhập mật khẩu giấu tin	39

HÌNH 3.5. GIẢI MÃ TIN ĐÃ GIẤU 40

MỞ ĐẦU

1. Nhu cầu bảo mật thông tin

Cuộc cách mạng thông tin dữ liệu số đã đem lại những thay đổi sâu sắc trong xã hội và trong cuộc sống. Những thuận lợi thông tin dữ liệu số mang lại cũng đề ra những thách thức và cơ hội mới cho quá trình đổi mới. Sự ra đời những phần mềm có tính năng mạnh, các thiết bị mới như máy ảnh kỹ thuật số, máy quét chất lượng cao, máy in, máy ghi âm kỹ thuật số, ... đã vọt tới thế giới tiêu dùng rộng lớn để sáng tạo, xử lý và thương thức các dữ liệu số.

Cùng với sự bùng nổ của Internet - phương tiện multimedia, các dữ liệu số được lan truyền và sao chép rất nhanh chóng dễ dàng, những vấn nạn như ăn cắp bản quyền, xuyên tạc thông tin, truy nhập thông tin trái phép... Vậy, vấn đề bảo mật thông tin số ngày càng trở nên cấp bách, đòi hỏi phải không ngừng tìm kiếm các giải pháp mới, hữu hiệu cho an toàn và bảo mật thông tin. Ngày nay, cùng với sự phát triển mạnh mẽ của ngành khoa học công nghệ thông tin, internet đã trở thành một phương tiện, một nhu cầu không thể thiếu đối với mọi người, nhu cầu trao đổi thông tin qua mạng ngày càng lớn vì vậy việc đảm bảo an toàn và bảo mật thông tin cho dữ liệu truyền trên mạng là rất quan trọng. Nhiều kỹ thuật mã hóa thông tin ra đời nhằm quyết vấn bảo mật dữ liệu. Tuy nhiên một văn bản được mã hóa sẽ dễ dàng gây sự chú ý, các tin tặc sẽ tìm mọi cách để giải mã, đánh cắp thông tin [2].

Một công nghệ mới phân nào giải quyết được những khó khăn trên là kỹ thuật giấu thông tin trong các nguồn đa phương tiện như: Âm thanh, hình ảnh, văn bản.... Xét theo khía cạnh mục đích và cách thực hiện thì giấu tin cũng là một dạng mật mã nhằm đảm bảo tính an toàn của thông tin, nhưng phương pháp này ưu điểm ở chỗ là giảm được khả năng phát hiện ra sự tồn tại của các tin cần giấu trong các dữ liệu được truyền đi trên mạng [2][3]. Ngoài

các nhà khoa học nước ngoài quan tâm đến lĩnh vực nghiên cứu này [3], các nhà khoa học trong nước cũng đã có nhiều công bố về các lĩnh vực liên quan.

2. Lý do lựa chọn đề tài

Từ trước đến nay, nhiều phương pháp bảo vệ thông tin đã được đưa ra, trong đó giải pháp dùng mật mã được ứng dụng rộng rãi nhất. Thông tin ban đầu được mã hoá để lưu trữ hoặc truyền trên mạng, sau đó bên nhận sẽ được giải mã nhờ khoá của hệ mã để có thông tin rõ. Kỹ thuật cổ điển này đã và đang được sử dụng rộng rãi và có nhiều ứng dụng trong thực tế. Cho đến nay đã có rất nhiều hệ mã đối xứng, bất đối xứng..., phức tạp được sử dụng như DES, RSA, NAPSACK..., rất hiệu quả và phổ biến.

Một phương pháp mới khác đã và đang được nghiên cứu và ứng dụng mạnh mẽ ở nhiều nước trên thế giới, đó là phương pháp giấu tin (DataHiding) được quan tâm phát triển trong nhiều năm nay. Giấu tin là một công nghệ mới rất phức tạp, đang được tập trung nghiên cứu ở nhiều nước trên thế giới như Đức, Mỹ, Italia, Canada, Nhật Bản... Giấu tin là nhúng thông tin vào trong một tệp dữ liệu khác. Tệp dữ liệu khác có thể là ảnh, âm thanh, video và văn bản (dữ liệu đa phương tiện). Đã có nhiều phương pháp kèm theo các thuật toán hiệu quả đã được công bố trên thế giới về chủ đề này. Tuy nhiên, những kết quả thực nghiệm cho thấy cần phải có thêm thời gian để nghiên cứu thẩm định, tuy nhiên các nhà khoa học khẳng định rằng đây là một công nghệ mới đầy hứa hẹn cho vấn đề an toàn và bảo mật thông tin.

Với mục đích tìm hiểu học hỏi về các mô hình, kỹ thuật giấu thông tin nhằm bảo vệ thông tin quan trọng trong các dữ liệu đa phương tiện, trong đó tập trung vào dữ liệu văn bản, học viên đã chọn thực hiện đề tài luận văn: ***“Kỹ thuật giấu tin trong tệp văn bản và ứng dụng”***.

3. Mục tiêu nghiên cứu

Trong luận văn, ngoài việc nghiên cứu tổng quan kỹ thuật giấu tin nói chung, sẽ trung sâu hơn vào giấu tin trong tệp văn bản. Sau đó, học viên sẽ thực nghiệm với tệp văn bản MS Word vì tệp văn bản MS Word là một trong những khuôn dạng tệp văn bản được sử dụng nhiều nhất hiện nay.

4. Đối tượng và phạm vi nghiên cứu

Trong khuôn khổ của luận văn, tác giả chỉ tập trung nghiên cứu và tìm hiểu tổng quan về các phương pháp các kỹ thuật giấu tin trong tệp văn bản. Nghiên cứu khả năng ứng dụng trong một mô hình thử nghiệm cụ thể là giấu tin trong tệp văn bản MS Word .

5. Phương pháp nghiên cứu

Thực hiện nghiên cứu lý thuyết thông qua tổng hợp các bài báo khoa học đã được công bố trong và ngoài nước và một số luận văn liên quan đến chủ đề này đã được thực hiện kết hợp với thực nghiệm thông qua chương trình thử nghiệm giấu tin trong tệp Word.

6. Cấu trúc của luận văn

Ngoài phần mở đầu trình bày về mục tiêu luận văn, bài toán cần giải quyết, phương pháp nghiên cứu ... và phần kết luận trình bày tóm tắt các kết quả đạt được của luận văn và các nhiệm vụ cần giải quyết tiếp theo, luận văn được trình bày thành ba chương như sau:

Chương 1: Tổng quan về giấu tin trong dữ liệu đa phương tiện

- 1.1. Lịch sử giấu tin
- 1.2. Khái niệm giấu tin
- 1.3. Mô hình giấu tin cơ bản
- 1.4. Một số kỹ thuật giấu tin
- 1.5. Khả năng ứng dụng giấu tin

Chương 2: Các kỹ thuật giấu tin trong văn bản