BẢO VỆ SẢN PHẨM VĂN HÓA SỐ VỚI PHƯƠNG PHÁP GIẤU TIN TRONG DỮ LIỆU ĐA PHƯƠNG TIỆN

LÊ THỊ CẨM BÌNH

Tóm tắt

Giấu tin là phương pháp nhúng hoặc làm ẩn thông tin trong một đối tượng thông tin khác. Đây là phương pháp đang được ứng dụng rộng rãi và đem lại hiệu quả thiết thực trong các ngành như: công nghiệp phần mềm, âm nhạc, phim ảnh, sách báo... Bài viết sau đây chủ yếu tập trung vào vấn đề ứng dụng giấu tin trong cơ sở dữ liệu đa phương tiện, nhằm mục đích bảo vệ sản phẩm văn hóa dạng kỹ thuật số như: bảo vệ bản quyền tác giả, phát hiện xuyên tạc thông tin, chống sao chép và bảo mật thông tin trên mạng internet nói riêng cũng như khi truyền thông tin nói chung.

Từ khóa: Giấu tin, internet, đa phương tiện, sản phẩm văn hoá số.

Abstract

Hiding information is the method of dipping or hiding information in one another information object. This is the method that is being applied widely and brings actual effectiveness in sectors such as: software industry, music, movies, books and newspaper, etc. The following article mainly concentrates on the application on hiding information in the multimedia database to protect the digital cultural products such as protecting copyright, discovering information distort, fighting against copying and securing information on internet in particular as well as information transmission in general.

Keyword: Hiding information, internet, multimedia, digital cultural products.

1. Quá trình lịch sử

Ý tưởng che giấu thông tin để truyền đi đã được con người nghĩ ra và sử dụng từ hàng ngàn năm trước đây. Tài liệu sớm nhất liên quan về vấn đề này được tìm thấy là của sử gia Herodotus (1) chép lại những câu chuyện từ thời Hy Lạp cổ. Một trong số ghi chép đó vào năm 440 trước Công nguyên kể về bạo chúa Histaiacus bị vua Darius bắt và giam giữ cẩn mật. Để có thể liên lạc với con rể là Aristagoras ở Miletus, ông đã cạo đầu một sứ giả tin cậy và xăm trên da đầu của người đó một thông điệp thúc giục con rể nổi dậy chống lại nhà vua. Đến khi tóc của người sứ giả mọc ra đủ dài để che hình xăm thì anh ta được gửi tới nơi cần đến. Thời kỳ này, các kỹ thuật giấu tin được áp dụng chủ yếu để truyền thông tin bí mật trong chiến tranh và một số ít trong các lĩnh vực khác. Gần đây, Ủy ban di sản văn hoá quốc gia Italia đã công bố một thông tin làm xôn xao dư luận và gây tranh cãi trong giới khoa học về bức hoạ nổi tiếng có nụ cười bí ẩn của nàng Mona Lisa. Khi các sử gia dùng kính lúp

có độ phóng đại cao soi bức tranh nguyên gốc hiện đang trưng bày ở bảo tàng Louvre, họ cho rằng đã tìm thấy những ký tự và con số nhỏ xíu trong đôi mắt nàng (2). Có thể, một thiên tài như Leonardo Da Vinci trong khi vẽ tác phẩm này đã bí mật truyền đi một thông điệp nào đó mà hiện nay người ta vẫn chưa giải mã được.

Nếu như quá trình lịch sử của thời kỳ đầu cho thấy, giấu tin thường áp dụng để truyền thông tin mật thì ngày nay, sự phát triển của công nghệ thông tin đã dẫn đến việc các kỹ thuật giấu tin (chủ yếu là thủy vân số) ứng dụng trong bảo vệ sản phẩm văn hóa dạng kỹ thuật số ngày càng gia tăng. Cho đến nay, đã có nhiều bài báo và các công trình nghiên cứu khoa học khác nhau đề cập đến vấn đề này.

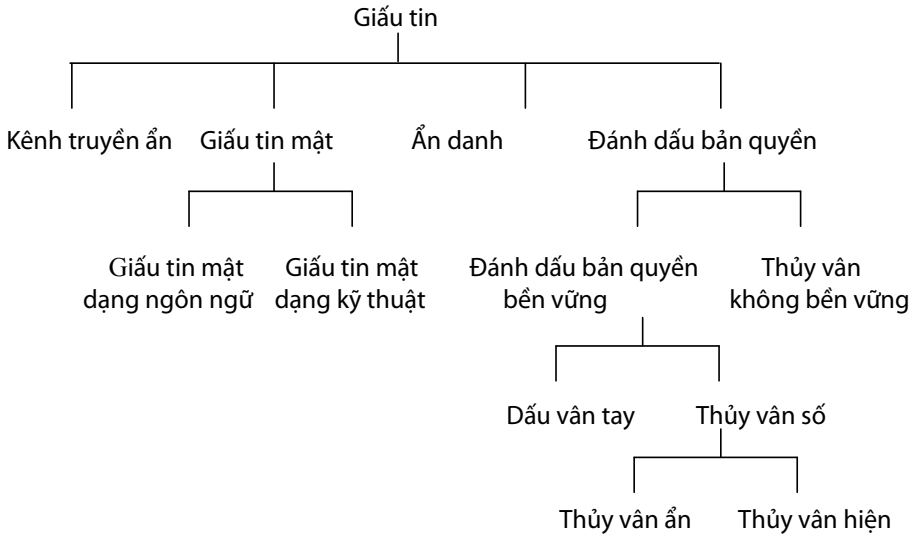
2. Phân loại giấu tin

Có nhiều cách để phân loại giấu tin: theo đặc tính, theo kỹ thuật... theo Fabien A. P. Petitcolas, Ross J. Anderson và Markus G. Kuhn (3) trong bài báo đăng vào tháng 7 năm 1999 thì cách phân loại theo kỹ thuật mà hiện nay đa số các nhà khoa học chấp nhận, được mô tả trong sơ đồ sau:

Trong sơ đồ trên, cách phân loại theo kỹ thuật chủ yếu dựa vào hai mục đích sử dụng là bảo mật dữ liệu giấu (hay dữ liệu nhúng-embedded data) và bảo vệ dữ liệu chứa (host data). Từ hai mục đích này, người ta phân chia lĩnh vực giấu tin thành hai hướng nghiên cứu chính là kỹ thuật thuỷ vân số (watermarking) và truyền thông tin mật (steganography).

Kỹ thuật thuỷ vân số (watermarking) là kỹ thuật nhúng nhãn hiệu (trademark), thẻ (tag) hay nhãn (label)... trong dữ liệu đa phương tiện hoặc đối tượng khác sao cho có thể tách chúng ra sau này (4). Khái niệm *watermark* bắt nguồn từ việc viết thông điệp bằng thứ mực vô hình lên giấy, và chỉ có thể đọc được khi nhúng nó xuống nước. Thủy vân số có hai loại là thủy vân ẩn (Imperceptible watermarking) và thủy vân hiện (Visible watermarking). Đối với thủy vân ẩn thì yêu cầu đặt ra là thông tin nhúng bị che dấu để người khác không phát hiện được. Thủy vân ẩn thường dùng là nhúng các thông tin về bản quyền sản phẩm. Ngược lại, với thủy vân hiện thì thông tin hiển thị công khai trên sản phẩm để người khác có thể phát hiện được. Thủy vân hiện thường nhúng thông tin như logo, tên tác giả, địa chỉ website...

Sơ đồ phân loại giấu tin theo đặc tính kỹ thuật



Kỹ thuật giấu thông tin mật (steganography) là kỹ thuật truyền tin mà trong đó thông tin ẩn được giấu trong thông tin chính. Khái niệm "steganography" bắt nguồn từ tiếng Hy Lạp, là sự kết hợp của từ (*) và (**) (xem phần chú thích) có nghĩa là "tài liệu được phủ" (covered writing) (5). Như vậy, thông tin mật được truyền từ người gửi tới người nhận cần được đảm bảo không làm cho người thứ ba có thể phát hiện được. Steganography có thể dùng thêm khóa (Intrinsic Steganography) để tăng tính bảo mật cho thông tin, hoặc là giấu tin thuần túy (Pure Steganography) không dùng khóa. Giấu thông tin mật có hai loại là giấu tin mật dạng ngôn ngữ (Linguistic Steganography), nghĩa là dùng ngôn ngữ thông thường để gửi thông tin bí mật, ví dụ bạn có thể ngụy trang thông tin mật ẩn trong các thông tin rác (thông tin có nội dung mở và không gây sự chú ý đối với người khác) hoặc dùng ngôn ngữ có qui ước ngầm...; còn giấu tin mật dạng kỹ thuật (Technical steganography) là kỹ thuật sử dụng các phương pháp khoa học để làm ẩn thông tin, ví dụ như dùng mực hóa học để che giấu thông tin có từ xa xưa hay các kỹ thuật sử dụng thông tin dư thừa trong văn bản, hình ảnh, âm thanh, video...

Như vậy có thể thấy sự khác biệt cơ bản giữa hai kỹ thuật nêu trên là ở chỗ, thủy vân số tập trung chủ yếu trong ứng dụng bảo vệ các đối tượng chứa, dữ liệu nhúng chủ yếu là các thông tin về bản quyền đối với sản phẩm số nên dung lượng dữ liệu nhúng thường không lớn, có thể hiện hoặc ẩn trong đối tượng chứa; trong khi đó kỹ thuật truyền thông tin mật lại quan tâm đến việc dữ liệu nhúng có dung lượng lớn, luôn ẩn trong đối tượng chứa sao cho không bị người khác phát hiện.

3. Môi trường và kỹ thuật giấu tin

Môi trường giấu tin được áp dụng hiện nay chủ yếu là các dạng dữ liệu đa phương tiện như dạng văn bản, hình ảnh, âm thanh và video với các kỹ thuật thủy vân số và giấu tin mật, bao gồm:

- *Giấu tin trong văn bản:* mặc dù dữ liệu dạng văn bản chiếm tỷ lệ lớn trên hệ thống máy tính và truyền trên mạng nhưng kỹ thuật giấu tin trong văn bản lại dễ bị phát hiện vì tính chất văn bản thuần túy rất dễ phát hiện sự thay đổi. Kỹ thuật áp dụng cho văn bản thường là phương pháp đưa thông tin ẩn vào giữa khoảng trống của các từ, đoạn hoặc các định dạng văn bản… Do lượng thông tin dư thừa đối với dữ liệu dạng văn bản là ít và dễ bị phát hiện nên người ta ít áp dụng giấu tin trong môi trường này so với các dữ liệu khác.

- *Giấu tin trong ảnh*: ảnh trên máy tính được tạo thành từ các điểm ảnh nhỏ (pixel) có màu sắc. Kỹ thuật giấu tin trong ảnh thường được thực hiện bằng cách thay thế một vài pixel ít quan trọng nhất trong ảnh gốc, nhằm mục đích không làm ảnh hưởng đến chất lượng ảnh hoặc không thể nhận thấy sự thay đổi sau khi giấu tin so với ảnh gốc bằng mắt thường. Do lượng thông tin được truyền có định dạng hình ảnh là rất lớn, có vai trò quan trọng, ví dụ như nhận thực, xác định xuyên tạc thông tin, bảo vệ bản quyền tác giả, chữ ký số... nên các ứng dụng liên quan đến giấu tin trong ảnh chiếm tỷ lệ lớn. Đây cũng là một kỹ thuật được trùm khủng bố Osama bin Laden dùng để liên lạc với đồng bọn trong vụ tấn công ngày 11 tháng 9 năm 2001 và đã qua mặt được các cơ quan an ninh.

- *Giấu tin trong âm thanh*: các định dạng âm thanh số phổ biến như MP3 áp dụng thuật toán nén bằng cách loại bỏ những sóng âm thanh mà con người không cảm thấy, nhằm giảm dung lượng của chúng mà không ảnh hưởng đến chất lượng âm thanh. Mặt khác, những âm thanh có tần số cao có thể che lấp âm thanh có tần số nhỏ. Do đó khi hiện diện cả hai loại

thì tai người khó phát hiện ra tần số nhỏ. Dựa vào nghiên cứu những đặc điểm nêu trên và dựa vào hệ thống thính giác của con người (7), kỹ thuật giấu tin trong âm thanh thường áp dụng phương pháp bổ sung thông tin ẩn vào những đặc trưng âm thanh có phạm vi nằm ngoài mức nhận biết của tai người, hoặc giấu tin vào phạm vi có tần số nhỏ khi hiện diện tần số lớn trong âm thanh, vì vậy người ta không thể phát hiện sự khác biệt khi nghe âm thanh gốc so với âm thanh gốc đã được nhúng thông tin ẩn.

- *Giấu tin trong video*: kỹ thuật giấu tin trong video dựa trên nghiên cứu về đặc điểm hệ thống thính giác và thị giác của con người. Tương tự như lĩnh vực giấu tin trong ảnh và âm thanh, giấu tin trong video được áp dụng rộng rãi trong các lĩnh vực như điều khiển truy cập, xác thực thông tin, bảo vệ quyền tác giả... Để có thể giấu dữ liệu như hình ảnh, âm thanh hoặc thậm chí cả video trong một đối tượng video khác, người ta áp dụng phương pháp như: phân bổ đều (J. Cox) (6) để phân phối thông tin giấu theo tần số của dữ liệu gốc, hoặc cấu trúc lưới đa chiều (Mukherjee)...

- *Các môi trường giấu tin khác*: ngoài các dữ liệu đa phương tiện, hiện nay các kỹ thuật giấu tin hướng đến các đối tượng khác như: các hệ quản trị cơ sở dữ liệu, các phương thức truyền thông tin,…

4. Ứng dụng giấu tin trong bảo vệ tác phẩm văn hóa số

4.1. Bảo vệ bản quyền tác giả (copy-right protection): theo Luật sở hữu trí tuệ số 50/2005/QH11, quyền tác giả là quyền của tổ chức, cá nhân đối với tác phẩm do mình sáng tạo ra hoặc sở hữu. Theo đó, thông tin mang ý nghĩa sở hữu quyền tác giả (thủy vân) được nhúng vào sản phẩm, nhằm mục đích giống như dán tem bản quyền của người chủ sở hữu và được pháp luật bảo vệ. Khi các tác phẩm văn hóa số (phim ảnh, âm nhạc, tác phẩm văn học…) được lưu thông trên thị trường thì thủy vân chính là nhân tố nhằm xác định chính xác chủ sở hữu hợp pháp. Kỹ thuật thủy vân bền vững cung cấp một chức năng rất quan trọng trong vấn đề bảo vệ bản quyền tác giả, bởi vì thủy vân cần phải bền vững như sản phẩm nhằm chống lại hành động giả mạo, tẩy xóa hay phá hủy nó.

4.2. Xác thực thông tin (authentication): xác thực thông tin nhằm xác định trong trường hợp chủ sở hữu quyền tác giả muốn kiểm tra sản phẩm của mình có bị thay đổi bởi một bên thứ ba hay không. Trong lĩnh vực này, người ta áp dụng kỹ thuật thủy vân không bền vững. Khi có sự tác động nào đó làm thay đổi sản phẩm thì dữ liệu nhúng sẽ không còn nguyên vẹn như ban đầu.

4.3. Phát hiện giả mạo thông tin (tamper detection): nhằm mục đích kiểm tra sản phẩm đó có phải là giả mạo hay không (ví dụ, khách hàng muốn kiểm tra tác phẩm mình muốn mua). Để phát hiện sự giả mạo, người chủ sở hữu sẽ nhúng thuỷ vân vào tác phẩm của mình, việc phát hiện được thực hiện bởi người mua căn cứ vào thuỷ vân sử dụng để bảo mật.

4.4. Dấu vân tay (fingerprinting): Dấu vân tay chứa dữ liệu nhúng có nội dung là thông tin (ví dụ như số serial hay khóa phần mềm) mang tính duy nhất cho mỗi giao dịch của nhà phân phối cung cấp cho người mua. Sau khi mua sản phẩm văn hóa số, người tiêu dùng sẽ sử dụng thông tin đó để giải mã và được xác nhận là người chủ hợp pháp của sản phẩm đó.

4.5. Dán nhãn (labeling): dữ liệu nhúng có thể là tiêu đề, tên tác giả, địa điểm, thời gian, chú thích… nhằm cung cấp các thông tin liên quan đến sản phẩm văn hóa số hoặc được sử dụng cho mục đích tìm kiếm chúng sau này.

Với những ứng dụng này thì yêu cầu thủy vân phải có độ an toàn cao và không bị xoá cho các thuỷ vân trong quá trình lưu thông sản phẩm.

4.6. Giấu tin mật (steganography): trong trường hợp tác phẩm văn hóa số cần được truyền một cách bí mật cho người nhận thì người ta áp dụng kỹ thuật giấu tin mật nhằm tránh người khác phát hiện.

4.7. Điều khiển truy cập (copy control): kỹ thuật thủy vân áp dụng trong việc điều khiển truy cập các sản phẩm số. Theo đó, các nhà cung cấp sản phẩm có thể sử dụng hệ thống điều khiển đọc và ghi, hoạt động theo cơ chế kiểm soát thông tin bản sao của sản phẩm, nhằm ngăn cấm việc sao chép bất hợp pháp bản gốc.

5. Kết luận

Ngành kinh doanh các sản phẩm văn hoá số là một thị trường có tiềm năng lớn. Tuy nhiên, việc phát tán dễ dàng mà không làm mất đi chất lượng, cũng như tốn tiền bản quyền tác giả đối với các sản phẩm này trên các hệ thống máy tính cũng như lưu thông trên mạng internet đã và đang làm suy yếu ngành kinh doanh này. Trước những thách thức đó, các công trình nghiên cứu về kỹ thuật giấu tin với những kết quả khả quan và được ứng dụng hiệu quả đã và đang là một giải pháp tốt, đáng được lựa chọn để giải quyết vấn đề nêu trên

L.T.C.B

(ThS, Khoa LLCT & KHCB)

Chú thích

* στεγανος ** γραφειν

Tài liệu tham khảo

1. Vũ Văn Phúc (2012), *Năng lực cạnh tranh của doanh nghiệp Việt Nam sau 5 năm gia nhập WTO*, Nxb Chính trị quốc gia.

1. James C. Judge: *Steganography: Past, Present, Future*. https://www.sans.org/reading-room/whitepapers/stenganography/steganography--past-present-future-552

2.http://www.artdaily.com/index.asp?int_sec=2&int_new=44112#.UVUKvjeK29g

3.http://gray-world.net/pl/papers/petitcolas99information.pdf

4. Đặng Văn Đức (2005), *Giáo trình Đồ họa máy tính*, bài 9, slide 4/32

5.http://en.wikipedia.org/wiki/Steganography

6. I. J. C,ox J. Kilian, T. Leighton, T. Shamoon, *A secure, robust watermark for multimedia* http://link.springer.com/chapter/10.1007/3-540-61996-8_41#page-1

7. http://tapchi.vnu.edu.vn/tn_2_09/b1.pdf

8. Petitcolas, FAP; Anderson RJ; Kuhn MG (1999). "*Information Hiding: A survey*" (pdf). Proceedings of the IEEE (special issue).http://www.petitcolas.net/fabien/publications/ieee99-infohiding.pdf

9. Gary C. Kessler: A*n Overview of Steganography for the Computer Forensics Examiner*; July 2004 – Volume 6 – Number 3.

http://www.au.af.mil/au/awc/awcgate/fbi/2004_03_research01.htm

10. Stefano Cacciaguerra & Stefano Ferretti: *Data hiding: Steganograpy and copyright marking*

http://www.bo.ingv.it/~scacciag/home_files/teach/datahiding.pdf

11. Nguyễn Hạnh Phúc, *Tổng quan về kỹ thuật giấu tin và giấu tin trong ảnh số*

http://khcn.vimaru.edu.vn/tckh/sites/default/files/data/So_09_04_2007/96_Ky%20thuat%20giau%20tin.pdf

Ngày nhận bài: 17- 4- 2013

Ngày phản biện, đánh giá: 10 - 8- 2013

Ngày chấp nhận đăng:12 - 8 - 2013