MỘT SỐ VẤN ĐỀ VỀ GIẤU THÔNG TIN

ThS. Nguyễn Văn Lập
Trường Cao đẳng Lào Cai
Tác giả liên hệ: lap@ylc.edu.vn

Ngày nhận: 10/12/2023
Ngày nhận bản sửa: 20/12/2023
Ngày duyệt đăng: 21/12/2023

Tóm tắt

Nhu cầu cần che giấu thông tin đã có cách đây hàng nghìn năm về trước, tuy nhiên, kỹ thuật này được dùng chủ yếu trong quân đội và trong các cơ quan tình báo. Gần đây, theo nhu cầu của cuộc sống, giấu thông tin mới được các nhà nghiên cứu quan tâm, từ đó, đã có nhiều công trình nghiên cứu về vấn đề này. Bên cạnh đó, cuộc cách mạng số hoá thông tin và sự phát triển nhanh chóng của mạng truyền thông là nguyên nhân chính dẫn đến sự thay đổi này. Theo một số nghiên cứu trước đây, giấu thông tin là nhúng một số thông tin quan trọng vào trong một đối tượng dữ liệu số khác nhằm giữ bí mật và tính chính xác của thông tin. Thông tin sẽ được giấu vào trong một phương tiện chứa nhờ một bộ nhúng. Bộ nhúng là thuật toán giấu tin, được thực hiện với một khoá bí mật hoặc được mã hóa trước khi giấu. Sau khi giấu tin, các đối tượng chứa tin sẽ được gửi đi hoặc được phát tán trên mạng nhưng khó có thể nhận biết được sự tồn tại của các thông tin quan trọng đang được giấu trong đó.
Từ khóa: *Giấu thông tin, kỹ thuật giấu thông tin, thủy vân số, bộ nhúng, dữ liệu số.*

Some Issues on data hiding

MA. Nguyen Van Lap
Lao Cai College
Corresponding Author: lap@ylc.edu.vn

Abstract

The need to conceal information has been present for thousands of years, but historically, this technique has predominantly been utilized by the military and intelligence agencies. However, in recent times, driven by the demands of modern life, the concept of steganography has piqued the interest of researchers and information technology institutions, leading to an abundance of research on this subject. Furthermore, the digital information revolution and the rapid evolution of communication networks have been pivotal factors contributing to this paradigm shift. According to previous studies, information concealment involves discreetly embedding vital information within another digital data object to ensure both secrecy and the integrity of the information. This concealment process employs a specialized algorithm, executed with either a secret key or pre-encryption, prior to hiding the information. Once concealed, these carrier objects are transmitted or disseminated across networks, rendering the concealed information virtually undetectable therein.
Keywords: *Data hiding, steganography, watermarking, embedding, concealing, digital data.*

Mở đầu

Theo nhu cầu của cuộc sống, giấu thông tin luôn được con người quan tâm từ thời xa xưa. Ngày nay, nhờ cuộc cách mạng số hoá thông tin và sự phát triển nhanh chóng của mạng truyền thông nên việc truyền thông tin qua mạng Internet ngày càng phát triển. Tuy nhiên, chính từ đó, nguy cơ dữ liệu bị truy cập trái phép cũng tăng lên; vì vậy, vấn đề bảo đảm an toàn và bảo mật thông tin cho dữ liệu truyền trên mạng là rất cần thiết. Do đó, để đảm bảo an toàn và bí mật cho một thông điệp truyền đi, người ta có thể giấu thông tin vào trong một đối tượng dữ liệu số khác nhằm giữ bí mật và tính chính xác của thông tin. Thông tin sẽ được giấu vào trong một phương tiện chứa nhờ một bộ nhúng. Bộ nhúng là thuật toán giấu tin, được thực hiện với một khoá bí mật hoặc được mã hóa trước khi giấu. Sau khi giấu tin, các đối tượng chứa tin sẽ được gửi đi hoặc được phát tán trên mạng nhưng người khác khó có thể nhận biết được sự tồn tại của các thông tin quan trọng đang được giấu trong đó. Bên cạnh đó, việc kết hợp các kỹ thuật giấu tin với các kỹ thuật mã hóa có thể nâng cao độ an toàn cho việc truyền thông tin.

1. Tổng quan về kỹ thuật giấu thông tin

1.1. Khái niệm về giấu thông tin

Trong một thời gian dài, mã hóa dữ liệu được sử dụng như một giải pháp chính để giải quyết vấn đề bảo mật thông tin. Mã hóa dữ liệu là việc chuyển đổi dữ liệu từ định dạng có thể đọc được sang định dạng được mã hóa. Dữ liệu được mã hóa chỉ có thể đọc được hoặc xử lý sau khi được giải mã. Bên cạnh đó, còn có cách khác để bảo mật thông tin, đó là thực hiện các giao dịch ngầm bên trong các giao dịch công khai, gọi là giấu tin (data hiding) hay giấu thông tin.

"Giấu thông tin" (steganography) là kỹ thuật nhúng thông tin *(embeding)* vào trong một nguồn đa phương tiện gọi là các phương tiện chứa (host data) mà không nhận biết được sự tồn tại của thông tin được giấu *(invisible)* trong đó. Giấu tin khác mã hóa ở chỗ là giấu tin giấu đi sự hiện diện của thông tin, trong khi mã hóa giấu đi ý nghĩa của thông tin.

Như vậy, có thể định nghĩa khái quát như sau: Giấu thông tin là kỹ thuật nhúng một số thông tin, tài liệu vào trong một đối tượng dữ liệu khác [1, tr.7]. Giấu thông tin nhằm hai mục đích: (i) bảo mật cho dữ liệu đang giấu; (ii) bảo mật cho chính đối tượng được dùng để chứa thông tin. Khuynh hướng thứ nhất là giấu bí mật (steganography), tập trung vào các kỹ thuật giấu tin sao cho thông tin giấu được nhiều và quan trọng là người thám tin khó phát hiện được một đối tượng có chứa tin giấu là trong hay không. Khuynh hướng thứ hai là thủy vân số (watermaking), nghiên cứu cách thức đánh dấu đối tượng chứa nhằm khẳng định bản quyền sở hữu hay phát hiện xuyên tạc thông tin. Nhiều kiểu dữ liệu số có thể được chọn làm dữ liệu chứa cho bài toán giấu tin như ảnh, video, âm thanh, văn bản hay các gói tin độc được truyền trong mạng.

1.2. Sơ lược về lịch sử giấu tin

Trong tiếng Hy Lạp, từ "Steganography" có nghĩa là tài liệu được phủ (covered writing), và đến ngày nay, nó đang được sử dụng. Theo sử gia người Hy Lạp Herodotus cổ đại, kỹ thuật giấu thông tin được truyền

qua nhiều thế hệ và được ghi chép lại từ rất sớm vào khoảng thế kỷ thứ V trước Công nguyên, khi mà bạo chúa Histiaeus bị vua Darius bắt giữ ở Susa, ông đã bí mật gửi tin bí mật cho con rể mình bằng cách cạo trọc đầu người nô lệ và xăm lên đó những thông tin để thông báo cho con rể của mình là Aristagoras ở Miletus. Một thời gian, sau khi tóc của người nô lệ này đã mọc lại bình thường thì anh ta mới được gửi tới Miletus để thông báo thông tin mà anh đang được mang trên đầu.

Trong Hy Lạp cổ đại, cũng có một câu chuyện khác tương tự do Herodotus ghi lại là dùng các viên thuốc được bọc trong sáp ong để ghi lại các thông tin. Một người Hy Lạp khác có tên là Demeratus muốn gửi thông tin cho Sparta rằng Xerxes đang có ý định xâm chiếm Hy Lạp. Để truyền tin đi, anh ta đã dùng cách bóc bỏ đi lớp sáp này của viên thuốc, sau đó, ghi thông tin lên bề mặt các viên thuốc, và cho bọc các viên thuốc lại bằng một lớp sáp mới. Những viên thuốc với lớp bọc mới này đã dễ dàng vượt qua được các cuộc kiểm tra.

Cũng theo nhiều nghiên cứu cho biết, cách sử dụng mực không màu để ghi lại thông tin là một trong các cách giấu tin khá phổ biến từng được sử dụng. Cách giấu này đã được sử dụng trong khoảng thời gian dài và tương đối là hữu hiệu cho giấu thông tin. Từ xa xưa, người La Mã cổ đã biết cách sử dụng những chất có sẵn trong tự nhiên như nước quả, nước tiểu và sữa để viết các thông tin, thông báo chèn vào giữa các loại văn tự thông thường. Khi muốn đọc các thông tin, thông báo này thì người ta chỉ cần hơ nóng lên là có thể dễ dàng đọc được.

Từ những dẫn chứng trên, cho thấy các ý tưởng về việc che giấu thông tin đã có từ hàng nghìn năm về trước, nhưng kỹ thuật này chủ yếu được dùng trong quân đội và trong các cơ quan tình báo, không được sử dụng thông dụng trong đời sống người dân. Gần đây, theo nhu cầu của cuộc sống, giấu thông tin được các nhà nghiên cứu quan tâm và đã có nhiều công trình nghiên cứu về vấn đề này.

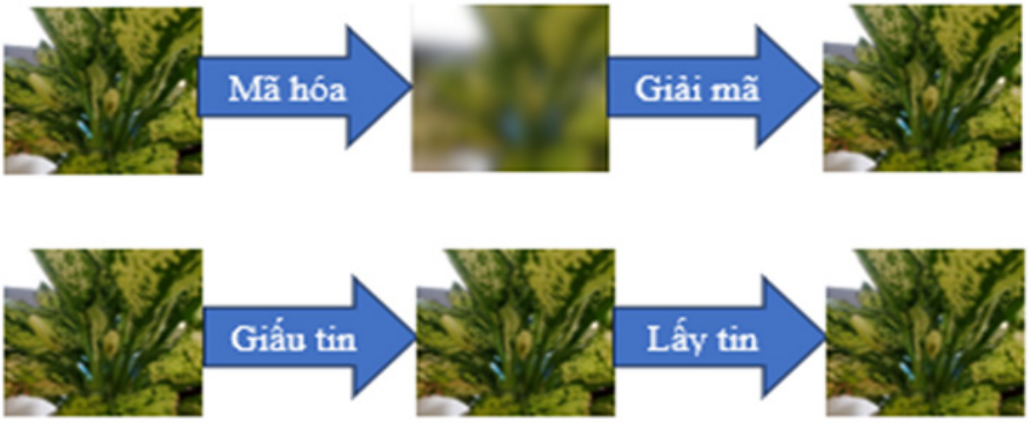
Bên cạnh đó, cùng với sự bùng nổ của cuộc cách mạng khoa học công nghệ, đặc biệt là cuộc cách mạng số và sự phát triển nhanh chóng của hệ thống mạng truyền thông là một trong những tác nhân chính cho sự thay đổi này. Công nghệ ngày càng phát triển thì đi kèm với nó là tạo ra những bản sao ngày một tinh vi và hoàn chỉnh nhờ vào các kỹ thuật chỉnh sửa, thay thế. Ngày nay, sự truyền tải thông tin trên mạng với dữ liệu đa phương tiện đã phát sinh rất nhiều vấn đề như: ăn cắp bản quyền, thông tin bị xuyên tạc, thậm chí là phát tán thông tin trái phép... Từ những thực trạng trên, đã đưa đến tính cấp thiết cần phải giấu thông tin trong dữ liệu đa phương tiện để phục vụ trong đời sống, cũng như bảo vệ dữ liệu đa phương tiện.

1.3. Điểm khác biệt giữa giấu thông tin và mã hóa thông tin

Theo các nghiên cứu thì sự khác biệt giữa giấu thông tin với mã hoá thông tin ở chỗ: thông tin được hiện rõ hay thông tin bị che giấu. Mã hoá thông tin là các thông tin hiện rõ là nó có được mã hoá hay không. Giấu thông tin thì rất khó biết được là có hay không thông tin được giấu bên trong.

Trên thực tế, khi mà thông tin mã

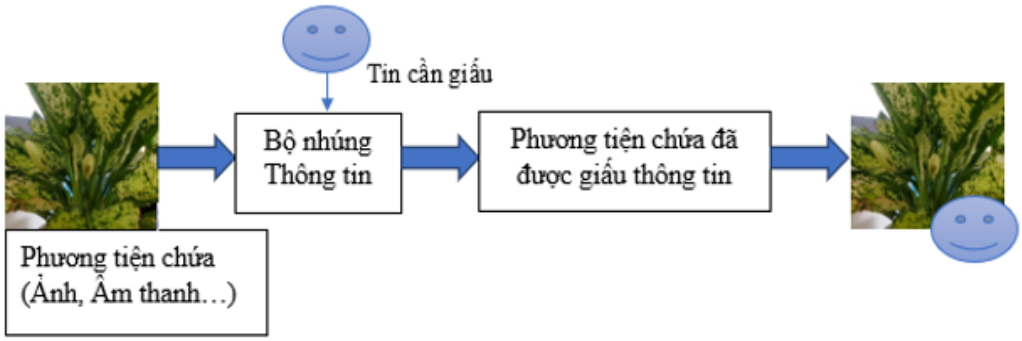
hoá bị phát hiện thì tạo sự thu hút cho đội tin tặc và chúng sẽ tìm mọi cách để phá. Từ đó, dẫn tới các cuộc chạy đua giữa một bên là đội bảo vệ thông tin và một bên là đội tin tặc, cuộc đua này vẫn đang diễn ra và chưa có hồi kết, chưa bên nào chiếm được lợi thế tuyệt đối. Từ thực trạng như hiện nay, phương pháp giấu thông tin thực sự rất cần thiết. Hình 1 cho chúng ta thấy rõ những điểm khác nhau giữa mà hóa và giấu tin.



Hình 1. Các điểm khác nhau giữa mã hóa và giấu tin

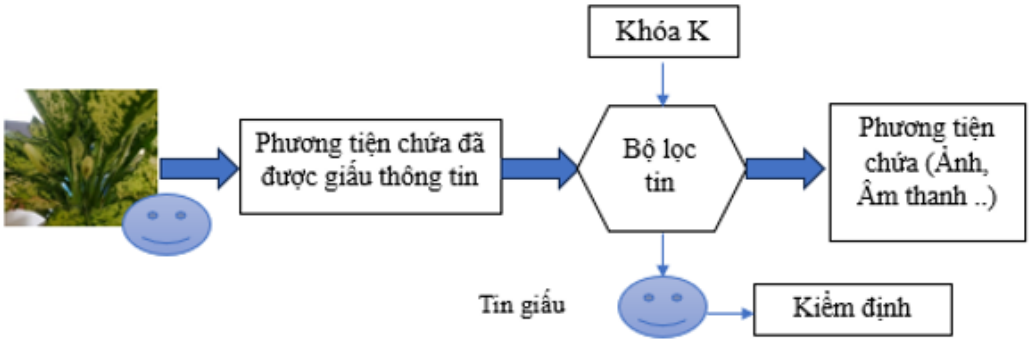
2. Một số mô hình giấu tin cơ bản

Hình 2 mô tả chung cho quá trình giấu thông tin một cách cơ bản nhất. Tùy thuộc vào từng mục đích sử dụng mà thông tin giấu trong dữ liệu có thể là bản quyền tác giả, có thể là tài liệu bí mật cần truyền đi. Người ta sẽ sử dụng một kỹ thuật (bộ nhúng) để đưa thông tin vào trong dữ liệu. Bộ nhúng có thể là chương trình hay thuật toán để đưa thông tin và dữ liệu và thường được thực hiện kèm theo một khoá bí mật (mã mật). Sau khi người ta đã giấu tin và dữ liệu (phương tiện chứa) sẽ được gửi đến người nhận hoặc được truyền lên trên mạng Internet.



Hình 2. Lược đồ chung cho quá trình giấu tin

Hình 3 mô tả chung quá trình tách tin đã giấu trong dữ liệu (phương tiện chứa). Trong quá trình giải mã, cần có một bộ giải mã tương ứng với bộ nhúng và khoá K đã đưa vào trong quá trình nhúng. Kết thúc quá trình giải mã, thu được dữ liệu gốc và tin giấu trong dữ liệu. Tin giấu thu được sau khi xử lý, kiểm định, được so sánh với các tin ban đầu đem đi giấu.



Hình 3. Mô tả chung cho quá trình tách tin

Giải thích một số thuật ngữ cơ bản:

Giấu tin (data hiding): Đây là thuật ngữ dùng để chỉ kỹ thuật giấu tin nói chung, bao gồm cả giấu tin mật và thuỷ vân số [2, tr.27].

Giấu tin mật (steganography): Thuật ngữ chỉ những kỹ thuật để giấu tin mật vào trong một đối tượng [2, tr.23].

Thuỷ vân số (watermarking): Dùng để chỉ những kỹ thuật giấu tin, được sử dụng để bảo vệ đối tượng chứa thông tin giấu [2, tr.23].

Phương tiện chứa (host signal): Chính là dữ liệu gốc được dùng để chứa thông tin cần giấu. Ví dụ: Giấu tin trong bức ảnh thì bức ảnh đó sẽ được gọi là ảnh chứa hoặc tin được giấu trong audio thì audio đó được gọi là audio chứa…

Thông tin cần giấu (embeded data): là thông tin được giấu (nhúng) vào trong dữ liệu. Thông tin được chia làm 2 loại như sau: (i) Giấu tin mật: thông tin giấu là các thông điệp (message); (ii) Kỹ thuật thuỷ vân số: thông tin là các thuỷ vân (công khai hoặc bí mật).

3. Một số dữ liệu có thể giấu thông tin

Công nghệ thông tin càng phát triển thì dữ liệu đa phương tiện ngày càng đa dạng, phong phú và đang tồn tại trong các lĩnh vực, các ứng dụng của công nghệ thông tin. Ngày nay, dữ liệu đa phương tiện thường xuyên được sử dụng để làm phương tiện trong quá trình giấu thông tin. Tuy nhiên, để lựa chọn dữ liệu làm phương tiện chứa, người sử dụng cần phải tìm hiểu rõ về các cấu trúc của những loại dữ liệu đó. Dữ liệu thường xuyên được sử dụng để giấu tin hiện nay như: Ảnh, audio, video...

3.1. Dữ liệu ảnh số

Theo khái niệm, thông tin giấu trong ảnh số là một trong số các phương pháp giấu thông tin, ảnh số sẽ được sử dụng làm phương tiện chứa.

Trong nhiều phương pháp hiện nay đang được sử dụng thì giấu thông tin trong ảnh chính là phương pháp chiếm tỉ lệ lớn nhất trong các chương trình ứng dụng hay các phần mềm giấu tin trong dữ liệu đa phương tiện, bởi lượng thông tin được trao đổi trong hình ảnh là rất lớn. Bên cạnh đó, việc giấu thông tin trong ảnh có vai trò quan trọng trong việc bảo vệ an toàn thông tin như: xác thực thông tin, xác định xem thông tin có bị làm thay đổi không, bảo vệ bản quyền tác giả, điều khiển truy cập, giấu thông tin mật... Từ những ưu điểm trên mà giải pháp giấu thông tin trong ảnh đã nhận được rất nhiều sự quan tâm từ các cá nhân, các tổ chức trên thế giới cùng nghiên cứu.

Bên cạnh đó, cùng với công nghệ phát triển như ngày nay, ảnh số đang được sử dụng rất nhiều trong các lĩnh vực đời sống xã hội thì giấu thông tin

trong ảnh là thực sự cần thiết. Tại các quốc gia phát triển, người ta đã sử dụng chữ ký số thay cho chữ ký tay và được số hoá để sử dụng như là hồ sơ cá nhân của các dịch vụ của người dân.

Ngoài ra, trong một số ứng dụng về nhận diện thẻ căn cước, hộ chiếu..., người ta có thể giấu thông tin trên các ảnh để xác định thông tin thực. Đây là xu hướng chung của nhiều nước trên thế giới cũng như tại Việt Nam. Thông tin được giấu trên các ảnh thẻ để xác định thông tin thực, tránh trường hợp bị giả mạo.

Đặc biệt, nhờ vào tính "vô hình" của giấu thông tin trong ảnh, đây chính là cách thức truyền thông tin mật cho nhau mà người khác không thể biết được, bởi sau khi giấu thông tin, chất lượng ảnh hầu như không bị thay đổi, đặc biệt đối với ảnh màu như hiện nay.

3.2. Dữ liệu Audio

Khác với giấu thông tin trong các đối tượng đa phương tiện, giấu thông tin trong audio có những đặc điểm riêng biệt. Cụ thể, một trong những yêu cầu cơ bản của giấu thông tin là đảm bảo tính chất "ẩn" của thông tin được giấu, đồng thời, không làm ảnh hưởng đến chất lượng của dữ liệu. Do đó, giấu thông tin trong ảnh hoàn toàn phụ thuộc vào thị giác của con người (HVS - Human Vision System), trong khi đó, giấu thông tin vào audio lại hoàn toàn phụ thuộc vào thính giác (HAS - Human Auditory system) của con người.

Tuy nhiên, lại có vấn đề khó khăn, vướng mắc khác ở đây chính là thính giác của con người chỉ có thể nghe được các tín hiệu ở các dải tần rộng và công suất lớn, điều đó dẫn tới nhiều khó khăn đối với kỹ thuật giấu thông tin trong audio. Có điều thuận lợi là thính giác của con người lại rất khó khăn trong việc nghe được những sự khác biệt các dải tần và công suất. Do đó, các âm thanh lớn, cao tần thì có thể che giấu được các âm thanh nhỏ, tần số thấp một cách dễ dàng.Trong nhiều tài liệu nghiên cứu và phân tích tâm lý đã đưa ra các điểm yếu trên, và điều này sẽ làm cơ sở cho việc chọn các dạng audio thích hợp để giấu tin.

Bên cạnh đó, khó khăn thứ hai của việc giấu thông tin trong audio là việc lựa chọn kênh thông tin để truyền tải dữ liệu. Cụ thể, kênh truyền hay băng thông chậm sẽ gây ảnh hưởng đến chất lượng của thông tin sau khi được giấu. Nếu nhúng một đoạn phần mềm java applet vào trong một đoạn audio (16 bit, 44.100hz) có chiều dài bình thường thì các phương pháp thông thường cũng cần ít nhất tốc độ đường truyền là 20bps [3, tr.19].

Ngoài ra, việc giấu thông tin trong audio còn đòi hỏi yêu cầu rất cao về tính đồng bộ và tính an toàn của thông tin. Hiện các phương pháp giấu thông tin cho audio đều lợi dụng điểm yếu trong hệ thống thính giác của con người. Vì vậy, phương pháp sử dụng audio để giấu thông tin hiện nay vẫn còn nhiều hạn chế và chưa có nhiều công trình được công bố trong lĩnh vực này.

3.3. Dữ liệu video

Tương tự như các kỹ thuật giấu thông tin trong ảnh hay trong audio, giấu thông tin trong video đã được nghiên cứu và ngày càng phát triển để ứng dụng vào thực tế. Hiện nay, giấu thông tin trong dữ liệu video được ứng dụng nhiều vào các lĩnh vực như: điều khiển truy cập thông tin, xác thực thông tin và bảo vệ bản quyền tác phẩm, tác giả...

Cùng với sự phát triển mạnh mẽ của việc giấu thông tin trong video, nhiều kỹ thuật giấu tin trong video cũng được nghiên cứu và phát triển mạnh mẽ theo hai khuynh hướng là thuỷ vân số và giấu dữ liệu.

Hiện nay, người ta thường sử dụng các phân bổ đều để triển khai kỹ thuật giấu thông tin trong video, điều này đã được chỉ ra bởi Cox. Theo phương pháp này, thông tin được giấu sẽ được chia đều theo tần số của dữ liệu gốc. Hiện nay, để giấu thông tin trong video, các nhà nghiên cứu đã sử dụng những hàm số cosin riêng và các hệ số truyền sống riêng để triển khai phương pháp này.

Bên cạnh đó, trong các thuật toán khởi nguồn thì thường chỉ có các kỹ thuật cho phép giấu các tài liệu hoặc ảnh vào trong video. Hiện nay, công nghệ được phát triển cùng với các kỹ thuật mới đã cho phép người dùng có thể giấu cả âm thanh lẫn hình ảnh vào trong video.

Một số kết luận

Tóm lại, cùng với sự phát triển của khoa học công nghệ và sự bùng nổ của cuộc Cách mạng công nghiệp 4.0 đã giúp kỹ thuật giấu thông tin được áp dụng trong nhiều loại đối tượng, chứ không riêng dữ liệu đa phương tiện như ảnh, audio, video. Theo đó, đã có một số nghiên cứu giấu tin trong cơ sở dữ liệu quan hệ. Tuy nhiên, để đi sâu vào lĩnh vực công nghệ phức tạp này, cần có nhiều nghiên cứu và giải pháp chuyên sâu hơn nữa.

Để nâng cao dung lượng giấu tin trong ảnh và để giấu được bản tin có dung lượng lớn, chúng ta có thể sử dụng phương án giấu bản tin trong nhiều bức ảnh liên tiếp. Ngoài ra, đối với phương pháp giấu tin trong ảnh, để nâng cao độ bảo mật của thông tin, ta có thể cải tiến một kỹ thuật giấu tin trong ảnh bằng cách kết hợp với các phương pháp mã với khóa sử dụng một lần (OTP) hoặc mã hóa bản tin 2 lần để tăng mức độ bảo mật của thông tin được giấu.

Tài liệu tham khảo

[1] Nguyễn Văn Tảo, Đỗ Trung Tuấn, Bùi Thế Hồng. *Một số thuật toán giấu tin và áp dụng giấu tin mật trong ảnh.* Kỷ yếu hội thảo RDA8.

[2] Bùi Thế Hồng, Nguyễn Văn Tảo (2007). "Về một kỹ thuật thủy vân sử dụng phép biến đổi sóng nhỏ rời rạc và ma trận số giả ngẫu nhiên", *Tạp chí Khoa học và Công nghệ,* tập 45 - số 03.

[3] Nguyễn Văn Tảo (2009). *Nghiên cứu một số kỹ thuật giấu tin và ứng dụng*. Luận án tiến sĩ Toán học.