**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

**KHOA AN TOÀN THÔNG TIN**

****

**HỌC PHẦN: CÁC KỸ THUẬT GIẤU TIN**

**MÃ HỌC PHẦN: INT14102**

**Chủ đề: Giấu tin trong âm thanh**

**Lab: lsb\_watermark\_audio**

Sinh viên thực hiện: Hồ Phan Đức Anh

Mã sinh viên: B21DCAT024

Nhóm: 04

Giảng viên hướng dẫn: Đỗ Xuân Chợ

**HÀ NỘI 2025**

**Bài lab Các kỹ thuật giấu tin: lsb\_watermark\_audio**

1. **Mục đích**

Giúp sinh viên hiểu được thuật toán LSB và phương pháp thủy vân số cùng cách kết hợp chúng để phục vụ giấu tin trong âm thanh.

1. **Yêu cầu đối với sinh viên**

Quen thuộc với hệ điều hành Linux và có kiến thức về kỹ thuật giấu tin.

1. **Nội dung lý thuyết**
   1. **Thuật toán LSB**

Phương pháp LSB (Least Significant Bit) là một kỹ thuật giấu tin trong âm thanh bằng cách thay thế các bit có trọng số thấp nhất trong mẫu âm thanh số bằng dữ liệu cần giấu. Do các bit này ít ảnh hưởng đến chất lượng âm thanh tổng thể, sự thay đổi của chúng thường không được thính giác con người nhận ra.



*Giá trị 8 bit của tín hiệu A = 218*



*Giá trị 8 bit của tín hiệu B = 219 sau khi giấu bit 1 vào LSB của A*

Khi giấu 1 bit mật vào mẫu dữ liệu thì độ lệch giá trị giữa mẫu dữ liệu gốc và dữ liệu mang tin tối đa là 1. Để tăng độ bền vững của kỹ thuật giấu này, ta có thể giấu nhiều lần chuỗi thông điệp mật hoặc điều chỉnh vị trí các bit cao hơn trong mỗi mẫu để giấu tin. Các kỹ thuật giấu dùng phương pháp điều chỉnh LSB thường chọn các bit từ vị trí 1 đến 3, tính từ bên phải sang trái

**Ưu điểm:**

* Đơn giản và hiệu quả: Dễ triển khai, cho phép giấu một lượng lớn dữ liệu mà không làm tăng kích thước tệp âm thanh.
* Khó phát hiện: Sự thay đổi ở các bit LSB thường không thể nhận biết được bằng thính giác con người.

**Nhược điểm:**

* Dễ bị tấn công: Các kỹ thuật nén mất dữ liệu hoặc chỉnh sửa âm thanh có thể làm mất hoặc hỏng thông tin được giấu.
* Thiếu tính bền vững: Do sự nhạy cảm với các biến đổi, phương pháp LSB không đảm bảo tính toàn vẹn của thông tin giấu khi tệp âm thanh bị chỉnh sửa hoặc nén.

Do đó, việc kết hợp giấu tin giữa phương pháp LSB với các thuật toán mã hóa như RSA và AES là hoàn toàn hợp lý và cần thiết.

**3.2 Phương pháp thủy vân số**

Thủy vân số là phương pháp giấu thông tin (thủy vân) vào các vật chứa. Thủy vân là một thông tin nào đó mang ý nghĩa. Yêu cầu đối với thủy vân là một lượng thông tin rất nhỏ nhưng đủ mạnh để có thể bảo vệ vật chứa thủy vân. Ứng dụng của thủy vân số hiện nay rất đa dạng và hầu hết các lĩnh vực như: bảo vệ bản quyền hoặc chống xuyên tạc nội dung,….

Theo đó, 4 phương pháp phân loại chính bao gồm:

a) Phân loại theo mức độ ổn định của thủy vân đối với các tác động

- Thủy vân số bền vững (Robust Watermarking): Là dạng thủy vân tồn tại bền vững trước các cuộc tấn công nhằm loại bỏ thủy vân. Trong trường hợp loại bỏ được thủy vân thì vật chứa tin cũng không còn giá trị sử dụng. Một ứng dụng điển hình của thủy vân bền vững chính là bảo vệ bản quyền: thuỷ vân được nhúng trong sản phẩm như một hình thức dán tem bản quyền.

- Thủy vân số nửa dễ vỡ (Semi Fragile Watermarking): Là dạng thủy vân tồn tại bền vững khi vật chứa tin bị sửa đổi vô hại như: nén, làm nhiễu, lọc,… nhưng lại nhạy cảm (dễ vỡ) khi vật chứa tin bị sửa đổi độc hại như: đổi nội dung, cắt bỏ một phần. Thủy vân nửa dễ vỡ được thiết kế để phát hiện các sửa đổi độc hại trên sản phẩm (nhằm đảm bảo tính toàn vẹn của sản phẩm), đồng thời cho phép một số hoạt động sửa đổi vô hại trên sản phẩm.

- Thủy vân số dễ vỡ (Fragile Watermarking): Là dạng thủy vân nhạy cảm (dễ vỡ) trước mọi thay đổi của vật chứa tin, dù là thay đổi nhỏ nhất. Chính vì đặc điểm nhạy cảm như vậy nên thủy vân dễ vỡ được ứng dụng nhiều vào việc xác thực nội dung. Ví dụ: Khi một tòa soạn sử dụng một bức ảnh để đưa tin, toà soạn phải xác minh bức ảnh này đúng với ảnh gốc và chưa được chỉnh sửa.

b) Phân loại theo đầu vào của máy dò

Máy dò dùng để phát hiện vật chứa có chứa thủy vân hay không hoặc vị trí của thủy vân trong vật chứa. Tùy thuộc vào đầu vào của máy dò, hệ thống thủy vân được chia thành 3 loại chính là hệ thống đóng, hệ thống mở và hệ thống nửa đóng. Trong đó:

- Hệ thống đóng: đầu vào cần vật chứa gốc X, gồm 2 loại:

+ Loại 1: So sánh vật chứa có thủy vân S với vật chứa gốc X để tìm ra vị trí chứa thủy vân.

Đầu vào:

• Vật chứa có thủy vân S.

• Vật chứa gốc (vật chứa chưa có thủy vân) X.

• Khóa K.

Đầu ra: Vị trí chứa thủy vân M.

+ Loại 2: So sánh vật chứa nghi ngờ có thủy vân S với vật chứa gốc X để tìm ra vị trí chứa thủy vân

Đầu vào:

• Vật chứa có thủy vân S.

• Vật chứa gốc (vật chứa chưa có thủy vân) X.

• Khóa K.

• Thủy vân 𝑀′ là bản sao của thủy vân M.

Đầu ra: Trả lời có (1) hoặc không (0) cho câu hỏi: “Vật chứa tin S có chứa

các thủy vân M không?”.

- Hệ thống nửa đóng: Mục đích là để kiểm tra xem vật chứa S có chứa thủy vân M hay không?

+ Đầu vào:

* Vật chứa tin S
* Khóa K.
* Thủy vân 𝑀′ là bản sao của thủy vân M.

+ Đầu ra: Trả lời có (1) hoặc không (0) cho câu hỏi: “Vật chứa tin S có chứa các thủy vân M không?”.

- Hệ thống mở: Mục đích là để kiểm tra xem vật chứa S có chứa thủy vân M hay không

+ Đầu vào:

* Vật chứa tin S.
* Khóa K.

+ Đầu ra: Thủy vân M.

c) Phân loại theo tính ẩn hay hiện

- Thủy vân hiện (Perceptible Watermarking): Là loại thuỷ vân được hiện ngay trên sản phẩm và mọi người dùng có thể nhìn thấy được. Với loại thủy vân hiện cần có biện pháp chống lại sự thay đổi hay loại bỏ thủy vân trái phép.

- Thủy vân ẩn (Imperceptible Watermarking): Khó có thể nhìn thấy thuỷ vân bằng mắt thường.

d) Phân loại theo mức độ riêng tư

- Thủy vân số riêng tư (private watermarking): chỉ có người dùng được ủy quyền có thể phát hiện ra thủy vân. Thủy vân riêng tư chống lại người dùng trái phép tìm cách tách thủy vân ra khỏi vật chứa thủy vân. Thủy vân số riêng tư được ứng dụng trong bảo vệ bản quyền (xem mục 1.4.2 xác thực nội dung).

- Thủy vân số nửa riêng tư (Semi private watermarking): cho phép mọi người đọc có thể phát hiện có thủy vân được giấu trong các vật chứa. Tuy nhiên người dùng không biết được thủy vân được giấu ở vị trí nào. Trong thủy vân nửa riêng tư mọi người đều biết quá trình phát hiện và đặc biệt là khoá phát hiện, do đó người giấu thủy vân cần sử dụng khóa bí mật để nhúng thủy vân và cung cấp khóa công khai lên mạng để mọi người xác minh thủy vân.

- Thủy vân số công khai (Public watermarking): cho phép mọi người đọc được thủy vân trong vật chứa tin nhưng không thể sửa, xóa thủy vân. Thủy vân số công khai được ứng dụng trong kiểm soát sao chép.

1. **Nội dung thực hành**

Khởi động bài lab:

Vào terminal, gõ:

*labtainer -r lsb\_watermark\_audio*

(Chú ý: Sinh viên sử dụng mã sinh viên của mình để nhập thông tin email người thực hiện bài lab khi có yêu cầu, để sử dụng khi chấm điểm)

**Nhiệm vụ 1: Tạo tệp watermark**

Chuyển 1 chuỗi thông tin thành file nhị phân để giấu vào file âm thanh

***python3 watermark\_processor.py "hidden\_message" watermark.bin***

**Nhiệm vụ 2: Giấu watermark vào file âm thanh**Ghi dữ liệu watermark nhị phân vào file âm thanh input.wav, lưu thành output.wav.

***python3 embed\_watermark.py input.wav watermark.bin output.wav***

**Nhiệm vụ 3: Kiểm tra file đầu ra với sox**

So sánh thống kê file âm thanh trước và sau khi nhúng watermark.

**sox input.wav output.wav -n stat**

**Nhiệm vụ 4: Trích xuất watermark từ file âm thanh**

Trích xuất watermark đã giấu từ file output.wav và ghi kết quả vào data\_received.txt.

**python3 extract\_watermark.py output.wav data\_received.txt**

Đọc file data\_received.txt

**cat data\_received.txt**

**Kết thúc bài lab:**

Kiểm tra checkwork:

*checkwork*

Trên terminal đầu tiên sử dụng câu lệnh sau để kết thúc bài lab:

*stoplab*