

■ Nội dung môn học

1. Tổng quan về giấu tin

2. Ẩn mã

3. Phân tích ẩn mã

4. Thủy văn số

■ Chương 3. Phân tích ẩn mã

- Phân tích giấu tin hay còn gọi là tấn công một hệ giấu tin là phương pháp để phát hiện, trích rút, phá hủy hay sửa đổi thông tin đã giấu.
- Việc phân tích được coi là thành công hay không còn tùy theo ứng dụng. Đối với việc liên lạc bí mật, việc phát hiện và chứng minh một ảnh có chứa tin mật được coi là thành công.
- Đối với bảo vệ bản quyền số hay chống giả mạo thì việc phân tích được coi là thành công nếu không chỉ phát hiện ra thủy vân mà còn phá hủy hay sửa đổi nó, nhưng không làm giảm chất lượng ảnh mang.
- Bài toán chúng ta xem xét trong chương này là phân tích ẩn mã (steganalysis) đối với các ứng dụng trên hệ giấu tin mật.

■ Chương 3. Phân tích ẩn mã

■ Bài toán phân tích ẩn mã giải quyết 2 vấn đề sau:

- Phát hiện thông tin được giấu trong dữ liệu quan sát (phân tích bị động)
- Lấy được thông tin được giấu trong dữ liệu được quan sát (phân tích chủ động)

■ Một số thuật toán phân tích ảnh mã

- Nhúng LSB và tấn công biểu đồ
- Phân tích cặp mẫu
- Ảnh mã mù ảnh JPEG sử dụng hiệu chỉnh
- Phân tích ảnh mã mù trong miền không gian

■ Một số thuật toán phân tích ẩn mã

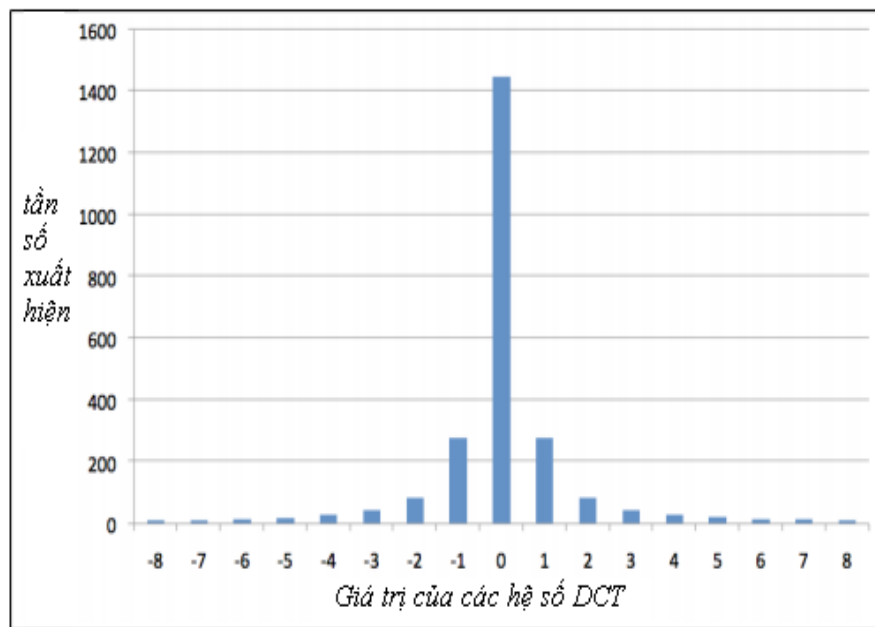
■ Nhúng LSB và tấn công biểu đồ:

□ Khái niệm cặp giá trị (PoV – Pairs of values) được Pfitzman & Westfeld đưa ra:

- Cho một ảnh I . Gọi j là giá trị của điểm ảnh trên I . Nếu I là ảnh đa cấp xám 8 bit thì $j \in [0, 255]$. Nếu j chẵn ($j = 2i$) thì sau phép lật bit giá trị của j là $2i + 1$, nếu j là lẻ ($j = 2i + 1$) thì sau phép lật bit giá trị của j là $2i$
- *PoV là một cặp hai giá trị điểm ảnh $(2i, 2i + 1)$ và hai giá trị trong cặp này chỉ sai khác nhau ở bit thấp nhất*

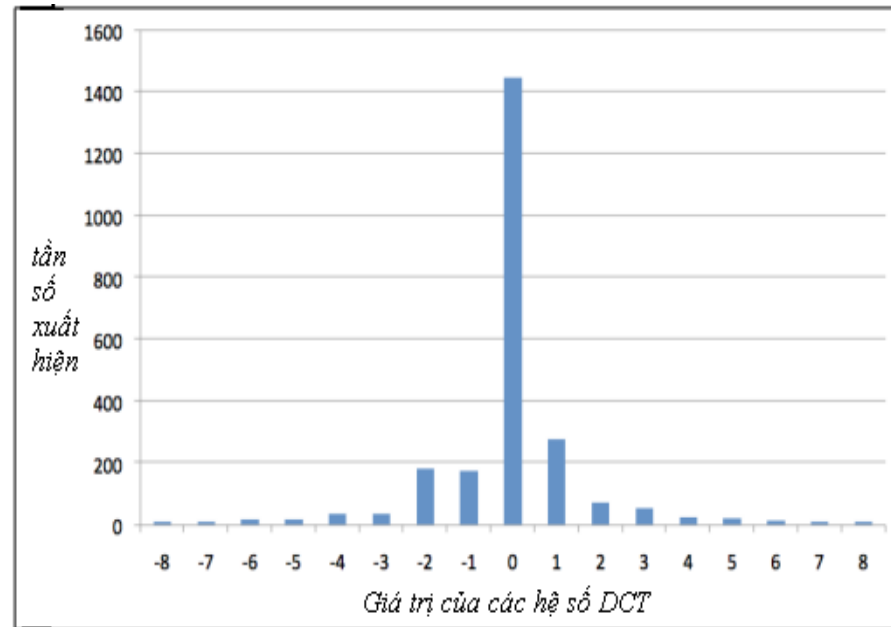
■ Một số thuật toán phân tích ẩn mã

- Việc nhúng LSB sẽ tạo ra các cặp PoV và các thay đổi trong biểu đồ biểu diễn đặc trưng điểm ảnh. Do đó, ta có thể sử dụng quan sát này để thực hiện tấn công biểu đồ



(a)

Biểu đồ 8 bit hình ảnh JPEG khi chưa nhúng



(b)

Biểu đồ của 8 bit hình ảnh JPEG khi dùng Jsteg

■ Một số thuật toán phân tích ẩn mã

- Kỹ thuật PoVs còn được gọi là phương pháp thống kê χ^2 và được áp dụng rất thành công đối với việc phát hiện giấu tin mật LSB một cách tuần tự.
- Có rất nhiều kỹ thuật PoV khác nhau như PoV2, PoV2r, PoV3.

■ Một số thuật toán phân tích ẩn mã

■ Thuật toán PoV3:

- **Ý tưởng:** Với 1 ảnh I cần kiểm tra, trước tiên ta thống kê tần số của các giá trị điểm ảnh chẵn, lẻ có mặt trong ảnh I. Ta xác định xác suất giấu tin của ảnh thông qua việc áp dụng tiêu chuẩn phân phối χ^2 đối với tần số của các cặp PoV
- **Input:** Ảnh I cần kiểm tra
- **Output:** p (xác suất giấu tin trong ảnh I)

■ Một số thuật toán phân tích ảnh mã

■ Thuật toán:

- Bước 1: Đọc dữ liệu ảnh vào một ma trận $M_{m \times n}$.
- Bước 2: Khởi tạo giá trị ban đầu cho vecto X, Y

For each $k \in [0, 127]$

$X[k] = 0;$

$Y[k] = 0$

Bước 3:

Tính $X[k]$ là tần số xuất hiện của các điểm ảnh có giá trị chẵn (cụ thể là $2k$) trên ảnh

Tính $Y[k]$ là tần số xuất hiện của các điểm ảnh có giá trị lẻ (cụ thể là $2k+1$) trên ảnh

■ Một số thuật toán phân tích ẩn mã

■ Bước 4: Giả sử ta có N cặp PoV

Với mọi k

Nếu $(X[k] + Y[k]) \leq 4$ thì

$$X[k] = Y[k] = 0$$

$$N = N - 1$$

■ Bước 5:

For each k

$$Z[k] = (X[k] + Y[k])/2$$

■ Một số thuật toán phân tích ẩn mã

- Bước 6: Giả sử ta có N cặp PoV, theo phương pháp thống kê khi bình phương với N – 1 bậc tự do ta tính:

$$\chi_{N-1}^2 = \sum_{k=0}^{127} \frac{(X[k] - Z[k])^2}{Z[k]} \quad (1)$$

- Bước 7: Tính p là xác suất của việc giấu tin:

$$p = 1 - \frac{1}{2^{\frac{N-1}{2}} \Gamma\left(\frac{N-1}{2}\right)} \int_0^{\chi_{N-1}^2} e^{\frac{-x}{2}} x^{\frac{N-1}{2}-1} dx \quad (2)$$



Im001.png



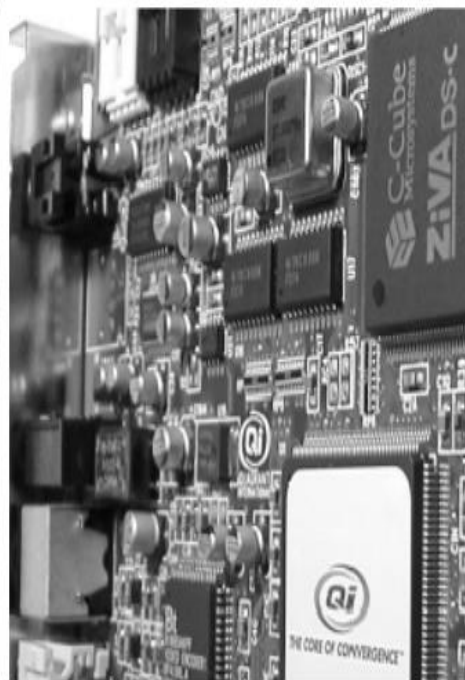
Im002.png



Im003.png



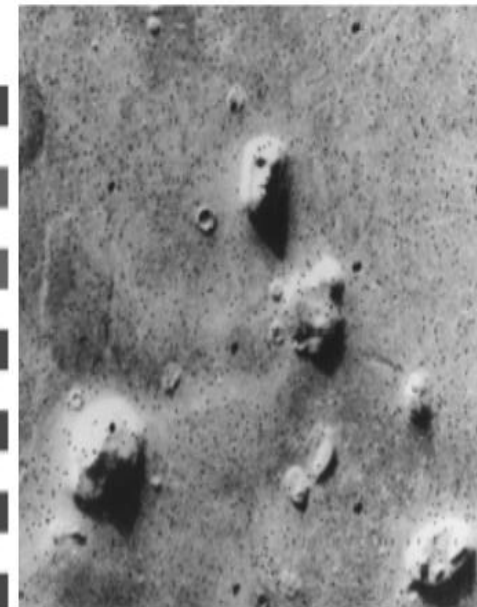
Im004.png



Im006.png

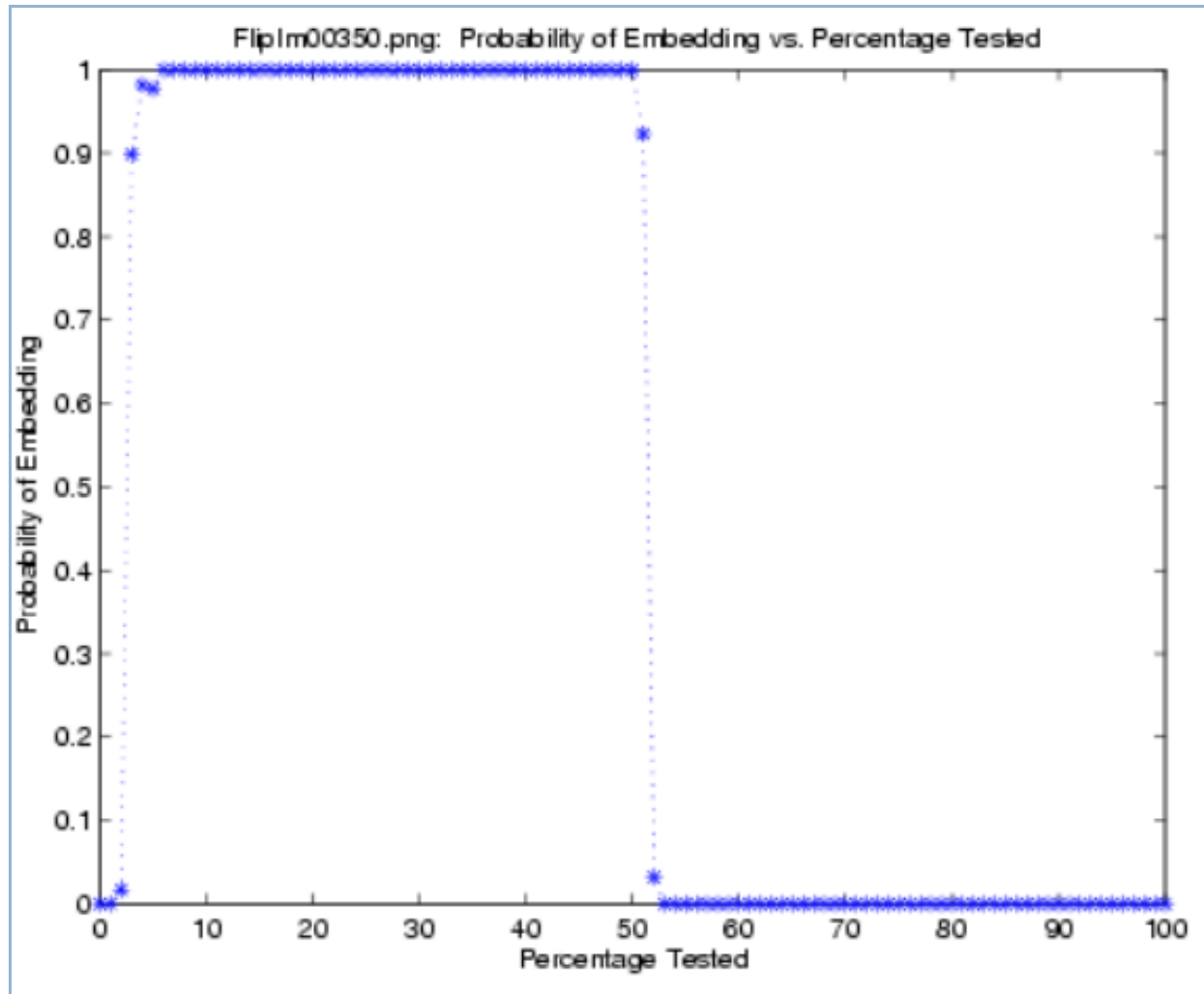


Im005.png



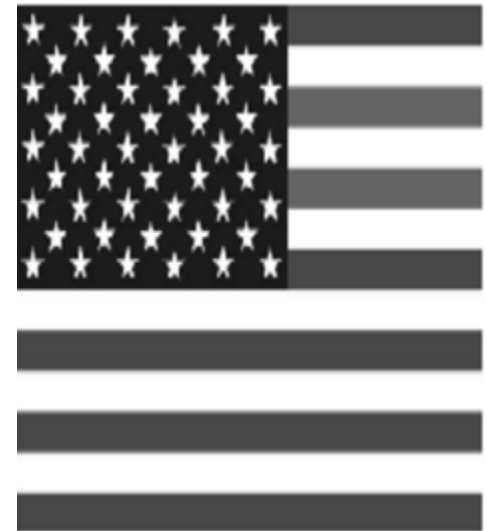
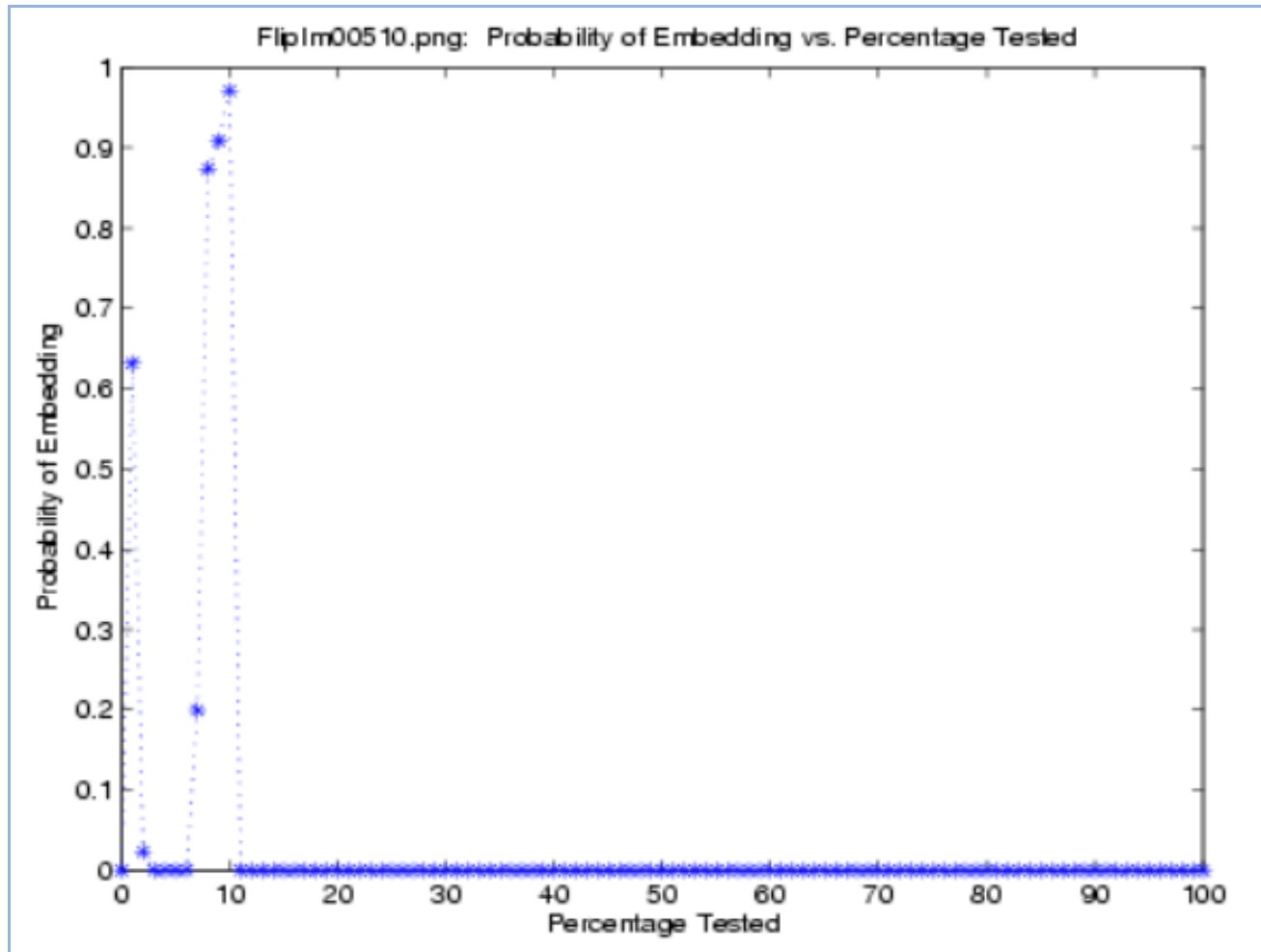
Im007.png

Một số thuật toán phân tích ẩn mã



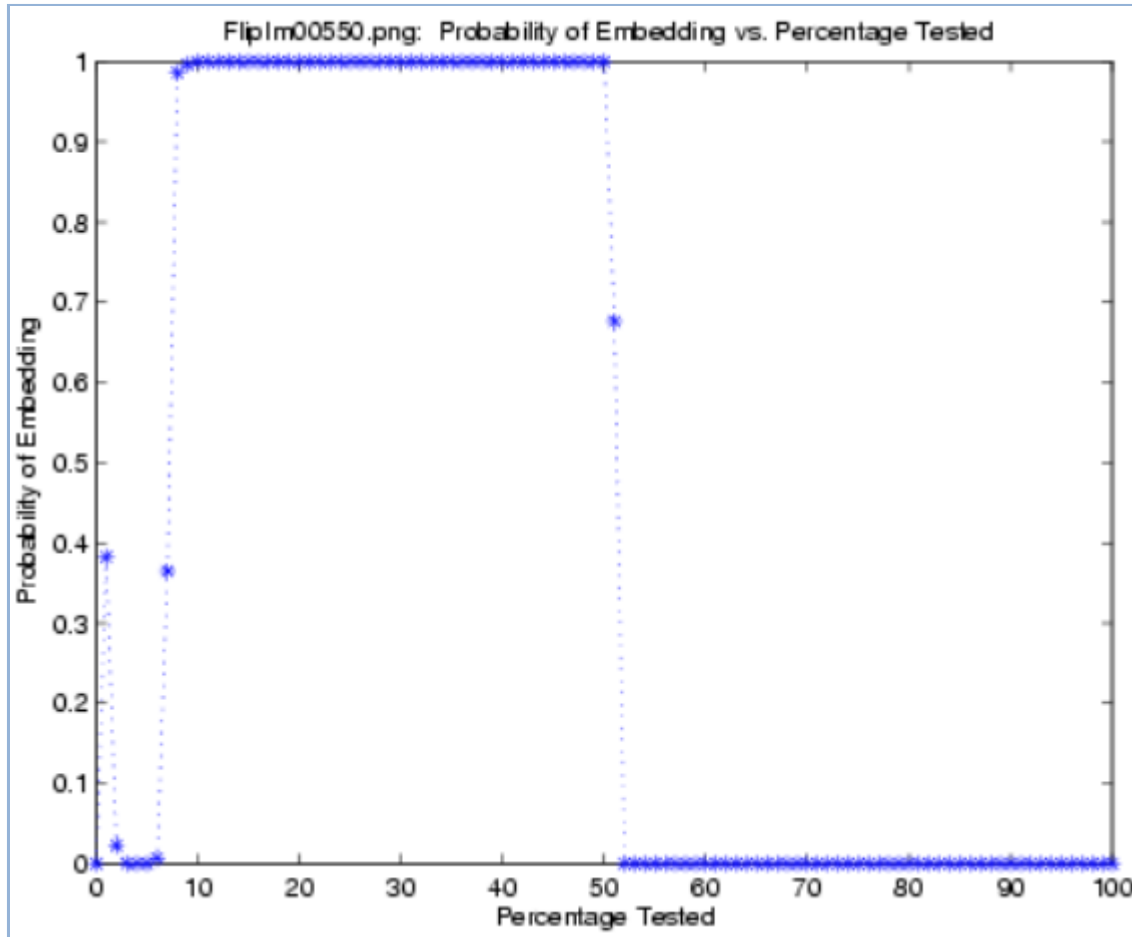
Xác suất sau khi giấu 50% ảnh Im003

Một số thuật toán phân tích ẩn mã



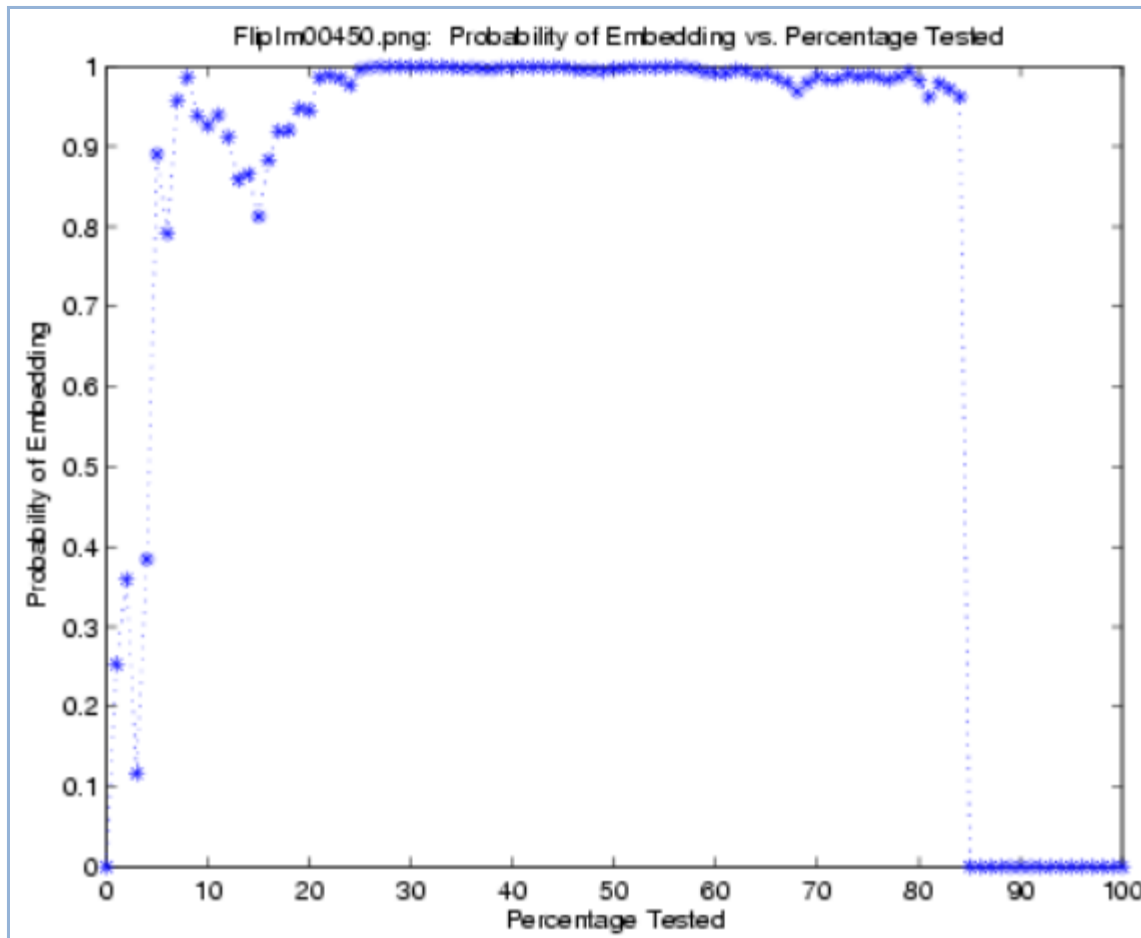
Xác suất sau khi giấu 10% ảnh Im005

Một số thuật toán phân tích ẩn mã



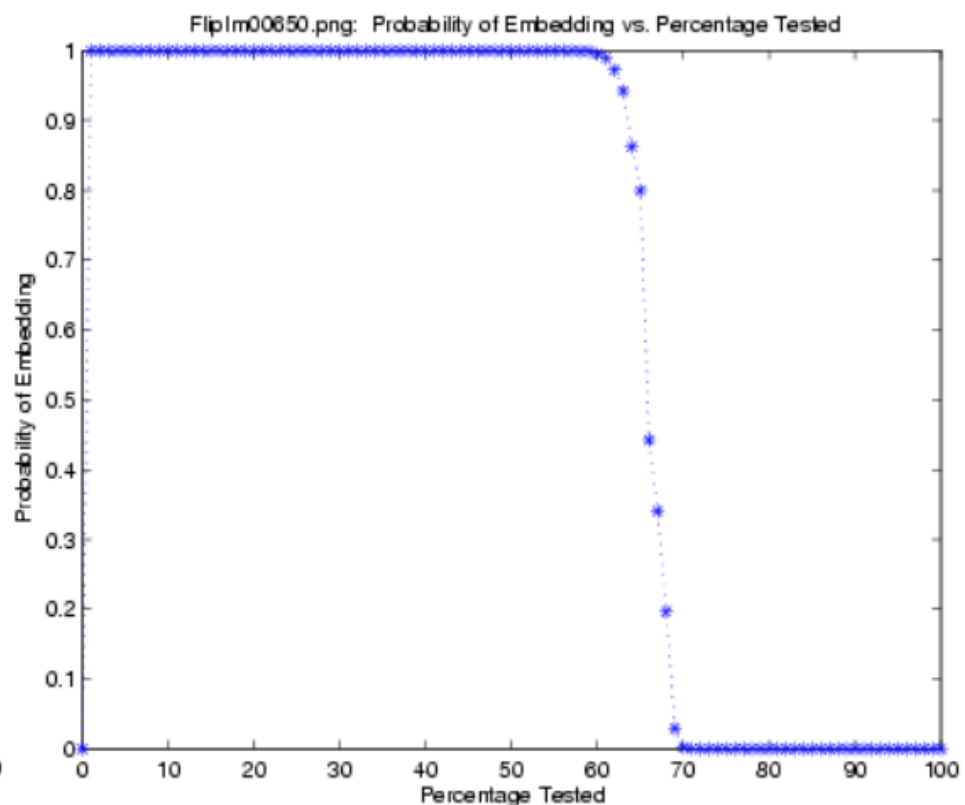
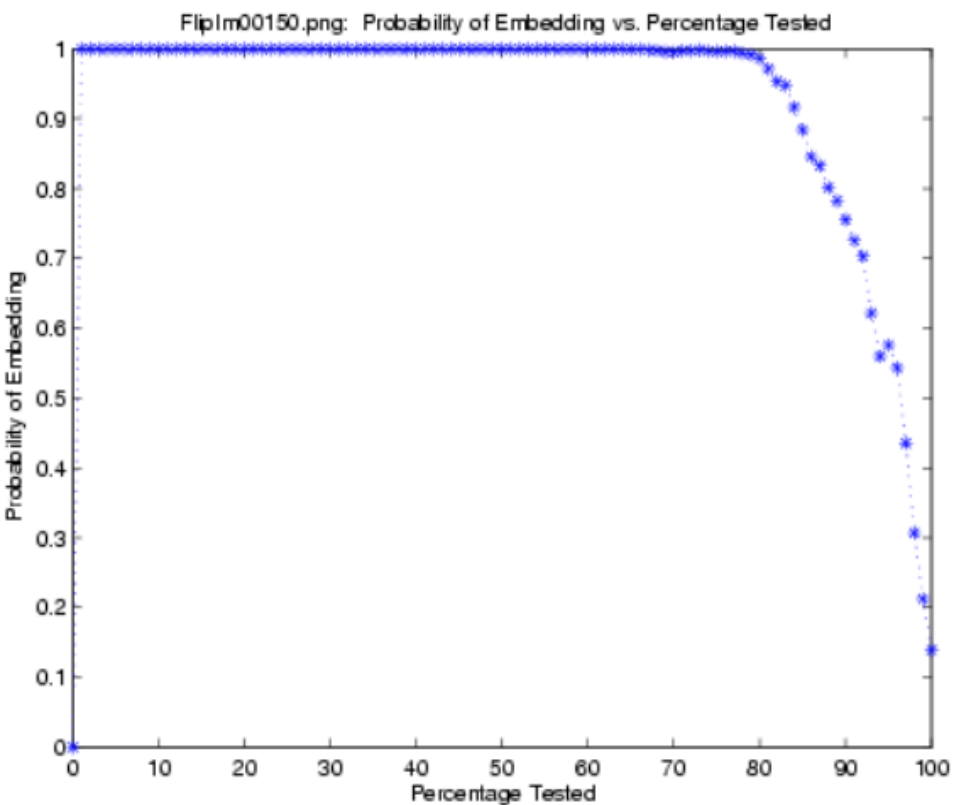
Xác suất sau khi giấu 50% ảnh Im005

Một số thuật toán phân tích ẩn mã

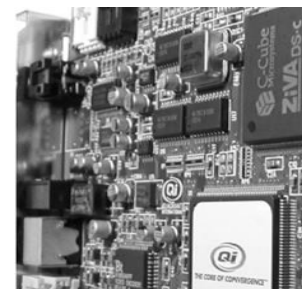


Xác suất sau khi giấu 50% ảnh Im004

Một số thuật toán phân tích ẩn mã

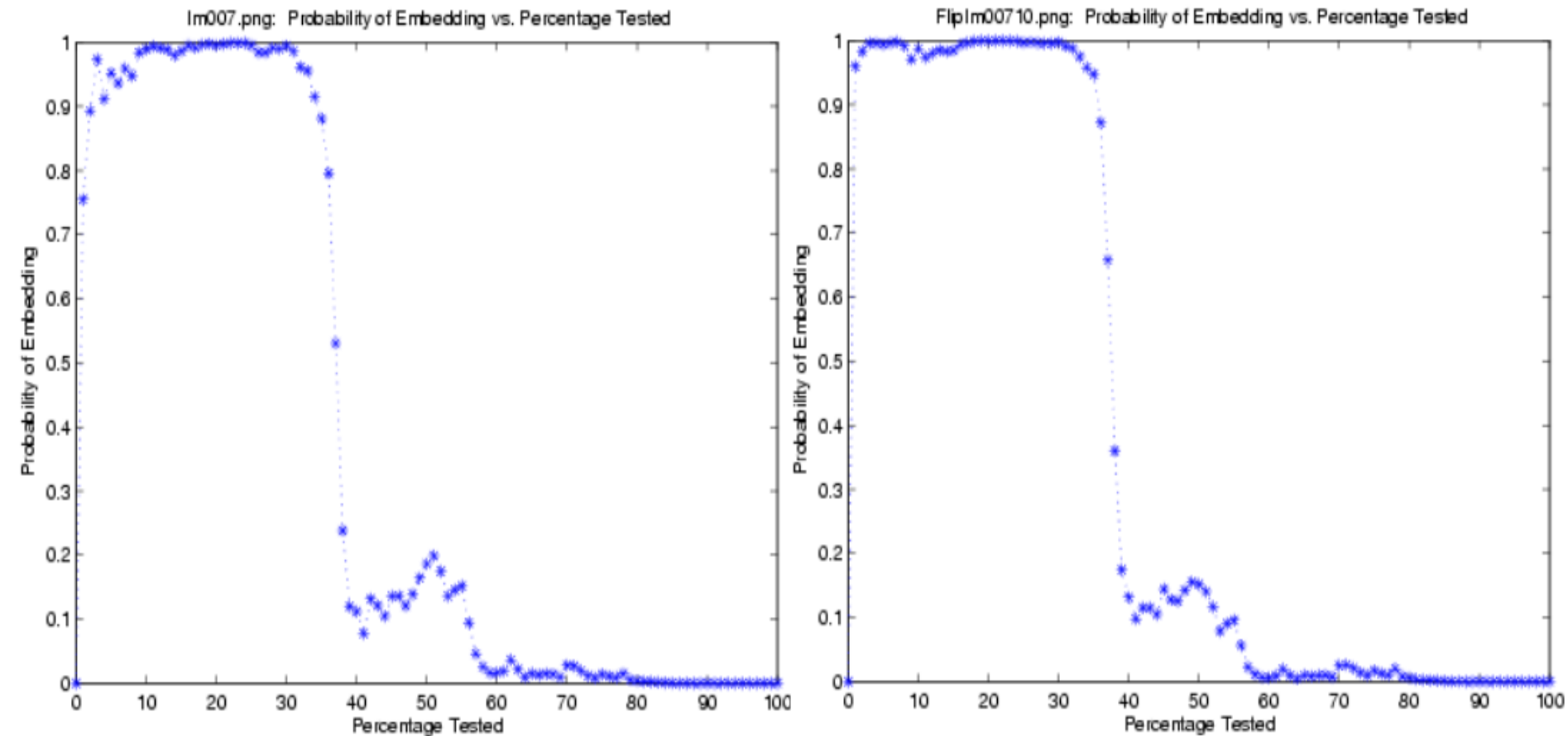


Xác suất sau khi giấu 50% ảnh Im001



Xác suất sau khi giấu 50% ảnh Im006

Một số thuật toán phân tích ẩn mã



Ảnh Im007 không nhúng dữ liệu

Xác suất sau khi giấu 10% ảnh Im007

Một số thuật toán phân tích ảnh mã

Image:	Im001	Im002	Im003	Im004	Im005	Im006	Im007
% Tested							
85%	YES	YES	YES	YES	YES	YES	YES
90%	YES	YES	YES	YES	YES	YES	YES
95%	YES	YES	YES	YES	YES	YES	YES
100%	YES	YES	YES	YES	YES	YES	YES

YES, if $p_r \geq 0.90$; NO, if $p_r \leq 0.10$; INConclusive, if $0.10 < p_r < 0.90$

Đánh giá độ tin cậy của phương pháp phân tích cặp giá trị (pp tấn công khi bình phương):

Image:	Im001	Im002	Im003	Im004	Im005	Im006	Im007
% Tested							
1%	YES	INC	NO	INC	INC	YES	YES
5%	YES	YES	YES	INC	NO	YES	YES
10%	YES	YES	YES	YES	YES	YES	YES
15%	YES	INC	NO	NO	NO	YES	YES
20%	YES	NO	NO	NO	NO	INC	YES
25%	INC	NO	NO	NO	NO	INC	YES
30%	NO	NO	NO	NO	NO	NO	YES
35%	NO	NO	NO	NO	NO	NO	YES

YES, if $p_r \geq 0.90$; NO, if $p_r \leq 0.10$; INConclusive, if $0.10 < p_r < 0.90$

Một số thuật toán phân tích ảnh mã

Image:	Im001	Im002	Im003	Im004	Im005	Im006	Im007
% Tested							
35%	YES	YES	YES	YES	YES	YES	YES
40%	YES	YES	YES	YES	YES	YES	YES
45%	YES	YES	YES	YES	YES	YES	YES
50%	YES	YES	YES	YES	YES	YES	YES
55%	YES	NO	NO	YES	NO	YES	YES
60%	YES	NO	NO	YES	NO	YES	YES
65%	YES	NO	NO	YES	NO	INC	YES
70%	YES	NO	NO	YES	NO	NO	YES

YES, if $p_r \geq 0.90$; NO, if $p_r \leq 0.10$; INConclusive, if $0.10 < p_r < 0.90$

Image:	Im001	Im002	Im003	Im004	Im005	Im006	Im007
% Tested							
1%	YES	NO	NO	NO	NO	NO	INC
5%	NO	NO	NO	NO	NO	NO	YES
10%	NO	NO	NO	NO	NO	NO	YES
15%	NO	NO	NO	NO	NO	NO	YES
20%	NO	NO	NO	NO	NO	NO	YES
25%	NO	NO	NO	NO	NO	NO	YES

YES, if $p_r \geq 0.90$; NO, if $p_r \leq 0.10$; INConclusive, if $0.10 < p_r < 0.90$

■ Một số thuật toán phân tích ẩn mã

■ Chống lại phương pháp tấn công khi bình phương:

- Chọn ảnh
- Chọn điểm ảnh để nhúng
- Tránh tạo các cặp giá trị PoV

■ Một số thuật toán phân tích ẩn mã

■ Phân tích cặp mẫu:

□ Giới thiệu thuật toán:

- Kỹ thuật SPA dựa trên lý thuyết về xích hữu hạn trạng thái
- Trước khi giấu tin, các phần tử trong cặp có quan hệ với nhau theo một độ đo nào đó. Nhưng sau khi giấu tin LSB một cách ngẫu nhiên thì các tập này sẽ thay đổi và nó dẫn đến những thay đổi các quan hệ thống kê.

■ Một số thuật toán phân tích ảnh

- KH: s_1, s_2, \dots, s_N là các chỉ số thể hiện vị trí của một mẫu trên ảnh
- (s_i, s_j) là một cặp mẫu, $1 \leq i, j \leq N$
- P là tập tất cả các cặp mẫu được lấy ra từ một ảnh. Có thể coi như là một tập hỗn hợp (multiset) của các bộ hai (u, v) , trong đó u và v là các giá trị của hai mẫu

■ Một số thuật toán phân tích ẩn mã

- Định nghĩa $D_n = \{(u, v) \in P \mid |u - v| = n\}$ là một tập con (submultiset) của P chứa cặp mẫu có dạng $(u, u + n)$ hoặc $(u + n, u)$
- Trong đó n là một số nguyên cố định $0 \leq n \leq 2^b - 1$, b là số bit nhị phân biểu diễn mỗi giá trị mẫu.
- Các cặp mẫu trong D_n sai khác nhau một lượng bằng n . Từ việc giấu tin chỉ ảnh hưởng tới các bit LSB nên ta sử dụng nhiều nhất là $(b - 1)$ bit tín hiệu trong việc chọn lựa các tập hỗn hợp đóng này.

■ Một số thuật toán phân tích ẩn mã

- Với mỗi số nguyên $m, 0 \leq m \leq 2^{b-1} - 1$ ta định nghĩa tập C_m là tập con (submultiset) của P có chứa các cặp mẫu mà giá trị của nó chỉ sai khác nhau m trong $(b - 1)$ bit đầu tiên.

$$C_m = \left\{ (u, v) \in P \mid \frac{|u - v|}{2} = m \right\}$$

$$\text{với } 0 \leq m \leq 2^{b-1} - 1$$

■ Một số thuật toán phân tích ẩn mã

■ Mỗi quan hệ giữa D_n và C_m

- Ta có: C_m chứa D_{2m}
- $D_{2m+1} = C_m \cap C_{m+1}$
- Ta phân hoạch D_{2m+1} thành hai tập con X_{2m+1} và Y_{2m+1}

• Trong đó:

$$X_{2m+1} = D_{2m+1} \cap C_{m+1}$$

$$Y_{2m+1} = D_{2m+1} \cap C_m \text{ với } 0 \leq m \leq 2^{b-1} - 2, X_{2^{b-1}} = \emptyset, Y_{2^{b-1}} = D_{2^{b-1}}.$$

- Cả hai tập X_{2m+1} và Y_{2m+1} đều là những tập con của P .
- Tập X_{2m+1} chứa các cặp (u, v) có dạng $(2k - 2m - 1, 2k)$ hoặc $(2k, 2k - 2m - 1)$.
- Tập Y_{2m+1} chứa các cặp (u, v) có dạng $(2k - 2m, 2k + 1)$ hoặc $(2k + 1, 2k - 2m)$.

■ Một số thuật toán phân tích ẩn mã

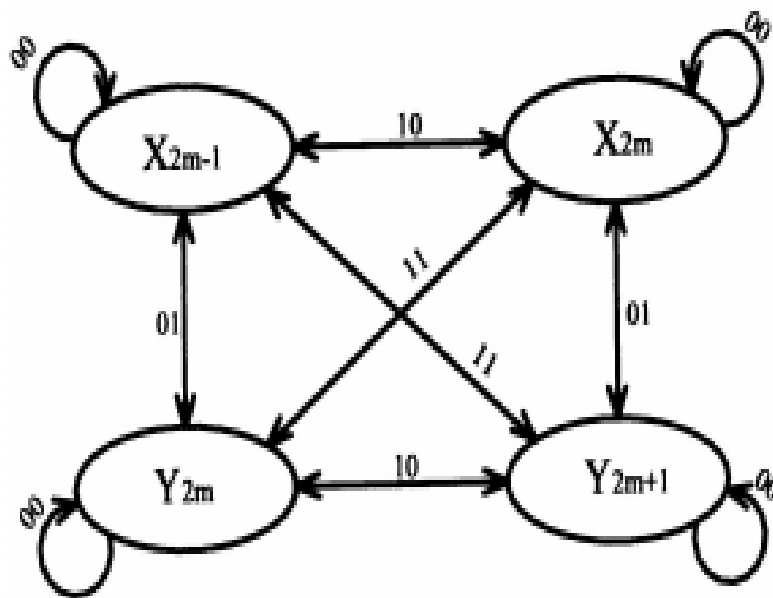
- Để phân tích ảnh hưởng của việc giấu tin LSB trên các cặp mẫu ta xem xét 4 trường hợp có thể của việc “lật” bit LSB theo mẫu
- Gọi mẫu $\pi \in \{00,01,10,11\}$ với 1 biểu thị cho một (hoặc nhiều) mẫu trong một cặp có bị đảo bit, 0 biểu thị cho một (hoặc nhiều) mẫu vẫn giữ nguyên (không bị đảo bit).

■ Một số thuật toán phân tích nhân mã

- Với mỗi m , $0 \leq m \leq 2^{b-1} - 1$, tập C_m được phân hoạch thành X_{2m-1} , D_{2m} , Y_{2m+1} .
- Lấy một cặp mẫu (u, v) tùy ý của X_{2m-1} thì (u, v) có thể có dạng $(2k - 2m + 1, 2k)$ hoặc $(2k, 2k - 2m + 1)$.
- Bằng việc chuyển đổi cặp mẫu (u, v) qua mẫu $\pi = 10$, ta thu được mẫu $(u', v') = (2k - 2m, 2k)$ hoặc $(u', v') = (2k + 1, 2k - 2m + 1)$.
- Tương tự như vậy, nếu (u, v) được thay đổi thông qua mẫu 01 thì $(u', v') = (2k - 2m + 1, 2k + 1)$ hoặc $(u', v') = (2k, 2k - 2m)$. Rõ ràng X_{2m} và Y_{2m} tạo thành một phân hoạch của D_{2m} .

■ Một số thuật toán phân tích ẩn mã

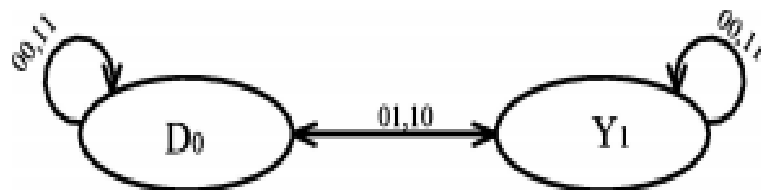
- Như vậy, C_m với $0 \leq m \leq 2^{b-1} - 1$ có thể được phân hoạch thành bốn tập con X_{2m-1} , X_{2m} , Y_{2m} và Y_{2m+1} được gọi là các tập con dấu vết (trace submultiset) của C_m



Xích hữu hạn trạng thái với các trạng thái là các tập con của C_m ($m > 0$)

■ Một số thuật toán phân tích ảnh mã

- Tập C_0 là đóng đối với phép giấu tin LSB và có thể được phân hoạch thành hai tập Y_1 và D_0



Máy trạng thái cho C_0

- Ý nghĩa của các xích hữu hạn trạng thái là có thể đo (một cách thống kê) số các tập con trước và sau khi giấu tin bằng cách sử dụng các xác suất của các mẫu thay đổi trong mỗi tập. Hơn nữa, nếu việc giấu tin LSB được làm một cách ngẫu nhiên trong ảnh thì các xác suất là các độ dài của thông điệp ảnh.

■ Một số thuật toán phân tích ẩn mã

- Với mỗi mẫu chuyển đổi $\pi \in \{00, 10, 01, 11\}$ và với bất kì tập con $A \in P$, ta định nghĩa xác suất $p(\pi, A)$ là xác suất các cặp mẫu của A bị thay đổi theo mẫu π .
- Đặt p là chiều dài thông điệp bị giấu trong các bit bị chia bởi tổng số các mẫu trong các ảnh. Ta có:

$$\begin{aligned}p(00, P) &= \left(1 - p/2\right)^2 \\p(01, P) &= p(10, P) = p/2 \left(1 - p/2\right) \\p(11, P) &= \left(p/2\right)^2\end{aligned}$$

- Đặt A và B là hai tập con của P sao cho $A \in B$. Ta nói rằng tập A là không chệch đối với tập B nếu $p(\pi, A) = p(\pi, B)$ ứng với mỗi mẫu biến đổi $\pi \in \{00, 10, 01, 11\}$. Khi $B = P$ ta nói rằng A là không chệch. Nếu tất cả bốn tập con của C_m là không chệch thì ta nói rằng C_m là không chệch.

■ Một số thuật toán phân tích ảnh mã

■ Thuật toán SPA phát hiện thông điệp nhúng LSB:

□ Đầu vào: Ảnh I cần kiểm tra

□ Đầu ra: Xác suất giấu tin p

□ Thuật toán:

● *Bước 1: Đọc vào ảnh I*

● *Bước 2: Đọc giá trị các điểm ảnh vào một ma trận A .*

● *Bước 3: Chia ma trận A thành dãy S gồm N mẫu liên tiếp nhau s_1, s_2, \dots, s_N . Mỗi mẫu s_i là một giá trị điểm ảnh.*

● *Bước 4: Xác định*

$$P = \{(s_i, s_j)\} \text{ với } 1 \leq i, j \leq N$$

■ **Bước 5:** $b = \text{độ dài xâu nhị phân biểu diễn mỗi mẫu.}$

■ **Bước 6 :** Với số nguyên n cố định, $0 \leq n \leq 2^b - 1$

Mỗi $(u, v) \in P$ nếu $|u - v| = n$ thì $D_n = D_n \cup \{(u, v)\};$

■ **Bước 7:** Với số nguyên m , $0 \leq m \leq 2^{b-1} - 1$

Mỗi $(u, v) \in P$ nếu $|u - v|/2 = m$ thì $C_m = C_m \cup \{(u, v)\};$

■ **Bước 8:** Xác định các tập giá trị sau:

$$X_{2^b-1} = \emptyset, Y_{2^b-1} = D_{2^b-1}$$

$$\text{Với } 0 \leq m \leq 2^{b-1} - 2: X_{2m+1} = D_{2m+1} \cap C_{m+1}, Y_{2m+1} = D_{2m+1} \cap C_m$$

■ **Bước 9:** Đặt $\pi \in \{00, 01, 10, 11\}$

■ **Bước 10:**

Nếu $\pi = 00$ hoặc $\pi = 10$ thì tập $X'_{2m-1} \cup X'_{2m}$ chứa các cặp mẫu của tập $X_{2m-1} \cup X_{2m}$ bị thay đổi thông qua các mẫu 00 hoặc 10

Nếu $\pi = 01$ hoặc $\pi = 11$ thì tập $Y'_{2m} \cup Y'_{2m+1}$ chứa các cặp mẫu của tập $Y_{2m} \cup Y_{2m+1}$ bị thay đổi thông qua các mẫu 01 hoặc 11

■ Một số thuật toán phân tích ẩn mã

■ *Bước 11:* Tính p

Nếu $E\{|X_{2m+1}|\} = E\{|Y_{2m+1}|\}$ thì xác định p là nghiệm nhỏ hơn của các phương trình sau

Nếu $m = 0$:

$$\frac{(2|C_0| - |C_1|)p^2}{4} - \frac{(2|D'_0| - |D'_2| + 2|Y'_1| - 2|X'_1|)p}{4} + |Y'_1| + |X'_1| = 0$$

Nếu $m \geq 1$

$$\frac{(|C_m| - |C_{m+1}|)p^2}{4} - \frac{(|D'_{2m}| - |D'_{2m+2}| + 2|Y'_{2m+1}| - 2|X'_{2m+1}|)p}{2} + |Y'_{2m+1}| + |X'_{2m+1}| = 0$$

■ Một số thuật toán phân tích ảnh mã

■ Ảnh mã mù ảnh JPEG sử dụng hiệu chỉnh

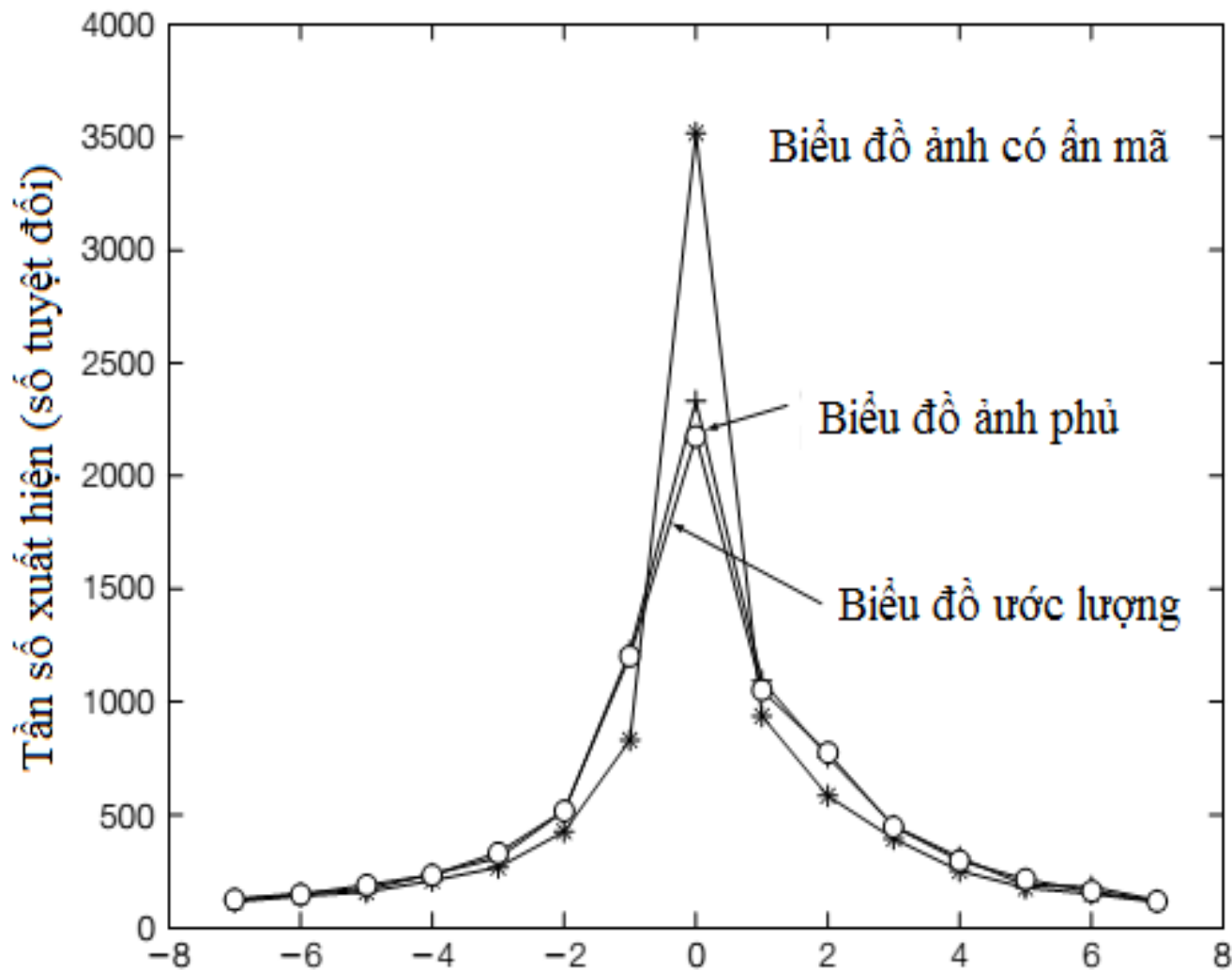
- Hầu hết các phương pháp ảnh mã ảnh được thiết kế cho định dạng JPEG bằng việc thực hiện các hệ số DCT lượng tử hóa
- Một kỹ thuật phổ biến sử dụng các thuật toán ảnh mã mù là sử dụng hiệu chỉnh, để cố gắng tính toán được vật phủ từ các vật nghi ngờ có nhúng tin
- Ước lượng vật phủ từ vật có nhúng tin là hoàn toàn có thể đối với ảnh JPEG bởi vì các hệ số DCT lượng tử hóa là bền vững với méo ảnh nhỏ do nhúng ảnh mã ảnh

■ Một số thuật toán phân tích ảnh mã

■ Hiệu chỉnh gồm các bước:

- Giải nén ảnh có nhúng tin thành miền không gian
- Cắt ảnh thành 4 cột (hoặc sử dụng phép biến đổi hình học) và thực hiện nén lại lần nữa sử dụng ma trận lượng tử như của ảnh có nhúng tin.
- Kết quả ảnh JPEG là một ước lượng của ảnh phủ mà có thể sử dụng để hiệu chỉnh một số giá trị lượng tử của ảnh gốc, như biểu đồ DCT.

Một số thuật toán phân tích ảnh mã



Biểu đồ của hệ số (1,2) của ảnh phủ (+), ảnh có ẩn tin theo thuật toán F5 và ảnh được hiệu chỉnh

■ Một số thuật toán phân tích ẩn mã

- Bằng việc phân tích cơ chế nhúng, có thể xác định được mối quan hệ giữa biểu đồ ảnh phủ và biểu đồ ảnh có nhúng tin cũng như một hàm của độ dài thông điệp được nhúng.
- Biểu đồ của ảnh có nhúng tin và biểu đồ ảnh phủ được ước lượng sẽ cho giới hạn cho độ dài thông điệp có thể ước lượng.

■ Một số thuật toán phân tích ẩn mã

- Ngoài biểu đồ của các hệ số DCT, việc hiệu chỉnh có thể tính toán được các thuộc tính khác của vật phủ
- Việc xây dựng các đặc trưng được hiệu chỉnh thông qua một hàm chức năng F .
- Khi xác định được tập các đặc trưng, phân tích ẩn mã mù sẽ được tiến hành để phân lớp.
- Sau khi lựa chọn phân lớp phù hợp, phân lớp sẽ biểu diễn với một tập huấn luyện của các đặc trưng được trích xuất từ vật phủ và vật có nhúng tin bằng thuật toán.

■ Một số thuật toán phân tích ẩn mã

- Mục đích của phân lớp là xác định các tham số bên trong để có thể phân biệt rõ ràng giữa 2 tập đặc trưng giúp kiểm tra trên tập vật phủ và vật có những tin xem đã từng được phân loại trước đây chưa.

■ Một số thuật toán phân tích ẩn mã

■ Phân tích ẩn mã mù trong miền không gian

- Phân tích ẩn mã mù trong miền không gian cũng tương tự như trong miền ảnh JPEG.
- Tuy nhiên, các đặc trưng được sử dụng là khác nhau và hiệu chỉnh dựa vào phương pháp được mô tả ở miền ảnh JPEG là không khả thi.

■ Một số thuật toán phân tích ảnh

- Hầu hết các phương pháp ảnh trong miền không gian có thể được hiểu là thêm nhiễu với các thuộc tính riêng.
- Thêm nhiễu trong miền không gian tương ứng với bộ lọc thông thấp (low-pass) của biểu đồ ảnh

■ Một số thuật toán phân tích ẩn mã

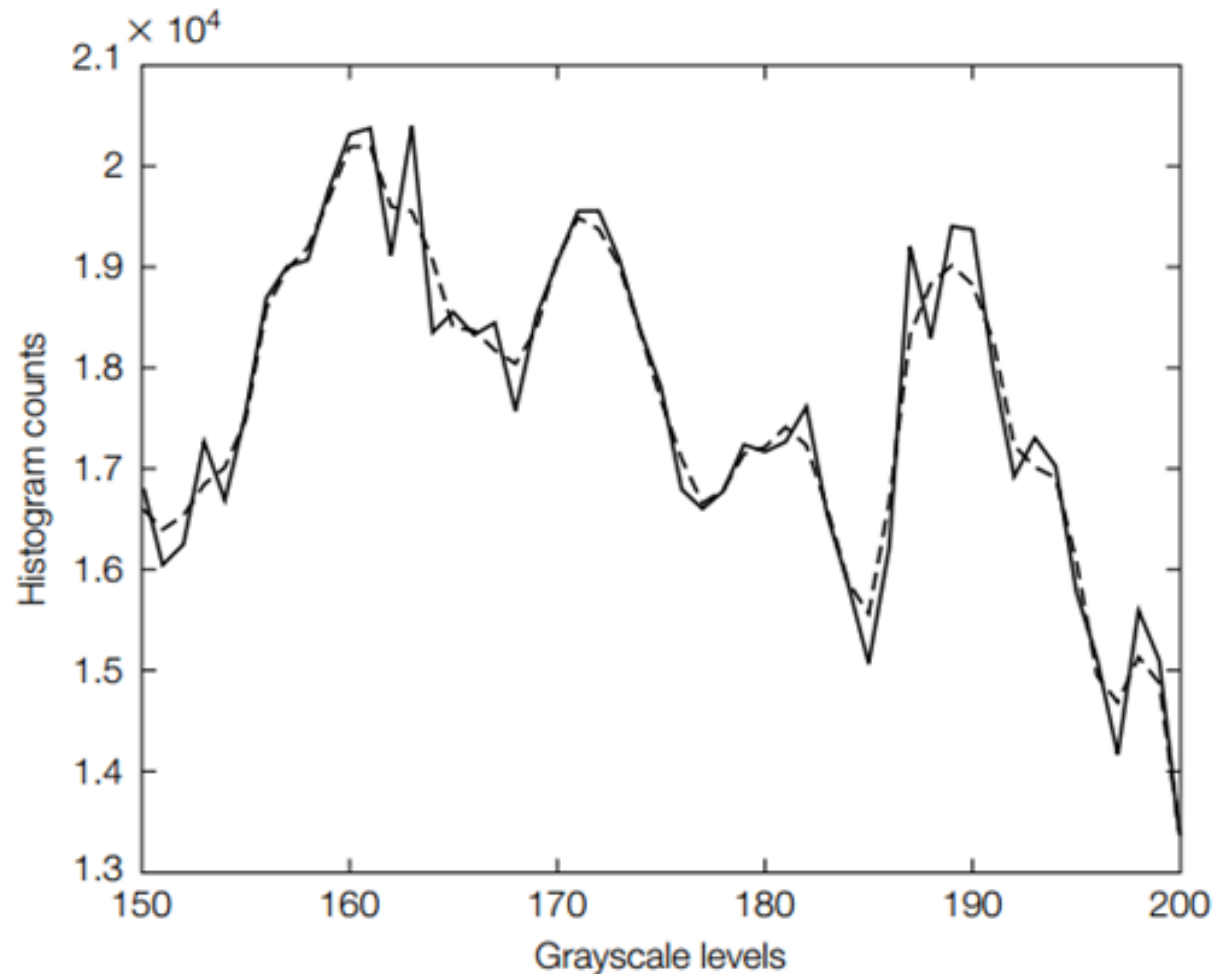
- Gọi $hc[i]$ và $hs[i]$ lần lượt là biểu đồ của ảnh phủ xám và ảnh có nhúng tin xám.
- Giả sử dấu hiệu tin nhúng được thêm vào ảnh là một dấu hiệu ngẫu nhiên không phụ thuộc vào ảnh với hàm khối lượng xác suất $f[j]$

$$\sum_j f[j] = 1$$

- Biểu đồ của ảnh có nhúng tin h_s :

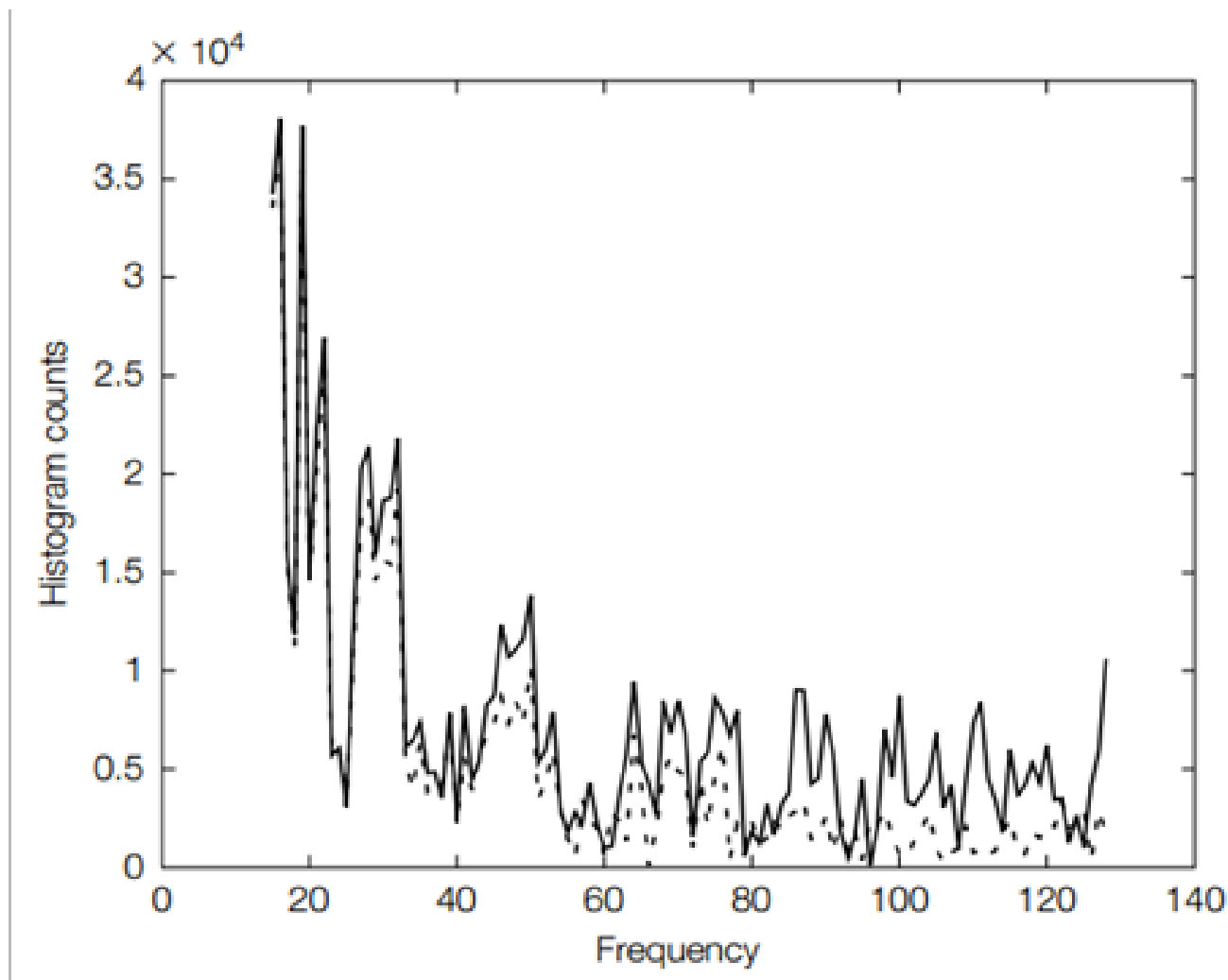
$$h_s = h_c * f$$

Một số thuật toán phân tích ảnh mã



Lược đồ của ảnh phủ và ảnh có nhúng tin sau khi được nhúng với phép nhúng 1

■ Một số thuật toán phân tích ảnh mã



Hàm đặc trưng lược đồ của ảnh phủ và ảnh có nhiễu tin sau khi được nhúng với phép nhúng ± 1

■ Một số thuật toán phân tích ẩn mã

- Lí tưởng nhất là có thể ước lượng ảnh phủ và áp dụng một hiệu chỉnh tương tự tới các đặc trưng có nguồn gốc từ miền không gian.
- Vì các thuật toán ẩn mã miền không gian được mô hình hóa giống như việc thêm nhiễu tần số cao, bộ lọc thông thấp hoặc thuật toán khử nhiễu được sử dụng cho mục đích này.