Mật mã và độ phức tạp thuật toán (Complexity and Cryptography) Chủ đề 5: Hệ mật đa trị và nhập nhằng

PGS.TS. Nguyễn Đình Hân
(Mobile: 0915.046.320; Email: han.nguyendinh@hust.edu.vn)

Viện Toán ứng dụng và Tin học
Trường Đại học Bách khoa Hà Nội

Hệ mật đa trị và nhập nhằng

1. Tiếp cận và phương pháp
2. Ngôn ngữ k-không nhập nhằng và phân bậc ngôn ngữ
3. Phép mã hóa đa trị và phép mã hóa đa trị hạn chế
4. Hệ mật đa trị và nhập nhằng MAS
5. Đánh giá độ an toàn của hệ mật MAS
6. Cài đặt hệ mật MAS

Tiếp cận và phương pháp

- Trong mật mã học, không có hệ mật nào tồn tại lâu dài trước sự tấn công. Do đó, luôn có nhu cầu thiết lập các hệ mật mới.
- Các hệ mật truyền thống được thiết lập trên cơ sở một bộ mã nhất định. Khi đó, mã là mục tiêu bị đối phương tấn công. Nếu ta sử dụng *ngôn ngữ không phải là mã* thì sẽ nâng cao được hiệu quả an toàn chống tấn công cho các hệ mật.
- Mặt khác, phép mã hóa mà ta đã xét ở các phần trước có tính chất đơn trị. Tức là, thủ tục mã hóa kết hợp một *bản rõ* với một *bản mã* duy nhất. Nếu ta có thể kết hợp một bản rõ với nhiều bản mã thì sẽ tăng cường được khả năng chống thám mã. Đây là đặc tính của phép mã hóa đa trị.
- Ta sẽ xem xét hệ mật đa trị và nhập nhằng MAS (Multi-valued and Ambiguous Scheme) đáp ứng các điều kiện trên đây.

Ngôn ngữ k-không nhập nhằng

Định nghĩa 5.1 Cho $X \subseteq \Sigma^+$ và một số tự nhiên $k \geq 0$. Khi đó

(i) Tập X được gọi là k-không nhập nhằng kiểu 1 nếu với mọi số nguyên $m \geq 1$ và với mọi $x_1, x_2, \ldots, x_k, y_1, y_2, \ldots, y_m \in X$, nếu có $x_1 x_2 \cdots x_k = y_1 y_2 \cdots y_m$ thì suy ra $k = m$ và $x_i = y_i$ với $i = 1, \ldots, k$.

Ngược lại (nếu có $\omega \in \Sigma^*$ mà $\omega = x_1 x_2 \cdots x_k = y_1 y_2 \cdots y_m$ với $x_1 \neq y_1$) thì tập X được gọi là k-nhập nhằng kiểu 1.

(ii) Nếu tồn tại số k lớn nhất sao cho X là k-không nhập nhằng kiểu 1 thì k được gọi là độ không nhập nhằng kiểu 1 của X. Trường hợp ngược lại, với $k \geq 0$ bất kỳ, X là k-không nhập nhằng kiểu 1, thì X được gọi là có độ không nhập nhằng kiểu 1 vô hạn.

Ngôn ngữ k-không nhập nhằng

Ví dụ 5.1 Cho Σ là bảng hữu hạn các chữ cái. Khi đó

1) Giả sử $\Sigma = \{a, b\}$ và $X = \{a, b, aaab\}$. Ta có thể kiểm tra bằng định nghĩa X là 0-không nhập nhằng kiểu 1. Hơn nữa, X là 1-nhập nhằng kiểu 1 vì tồn tại từ $\omega = (aaab) = (a)(a)(a)(b)$. Vậy, X có độ không nhập nhằng kiểu 1 là $k = 0$.

2) Giả sử $\Sigma = \{c, a_1, b_1, \ldots, a_k, b_k\}$ với $k \geq 1$ là một số tự nhiên tùy ý và giả sử $Y = \{c, ca_1, a_1b_1, b_1a_2, \ldots, b_{k-1}a_k, a_kb_k, b_k\}$. Rõ ràng, k là độ không nhập nhằng kiểu 1 của Y.

Lưu ý $\rightsquigarrow$ Ta có thể sử dụng thuật toán kiểm định mã để kiểm tra tính chất mã của X và Y.

Xác định độ không nhập nhằng kiểu 1

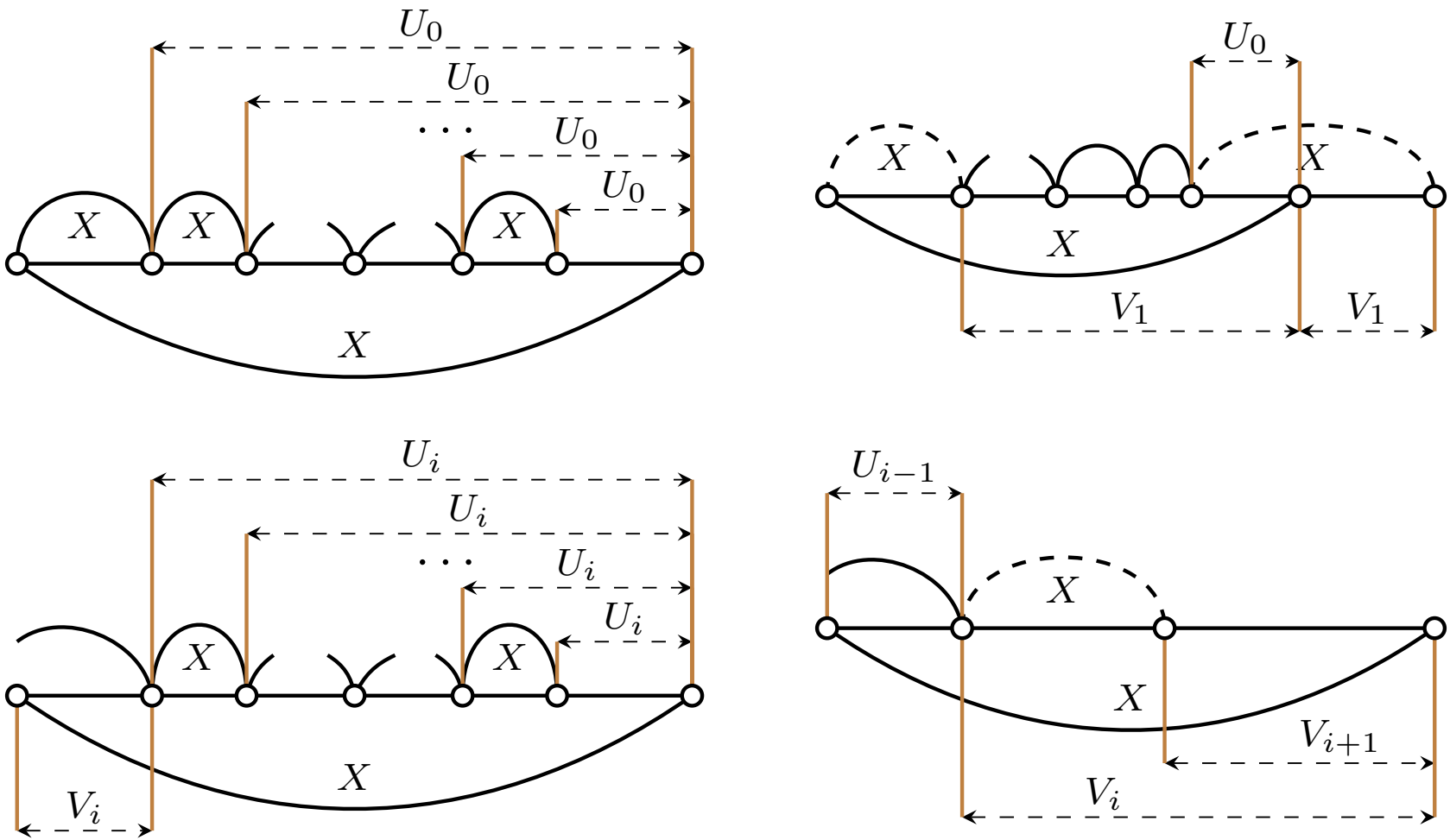
Cho $X \subseteq \Sigma^+$, để xác định độ không nhập nhằng kiểu 1 của X, ta xem xét hai tập phần dư U_i, V_{i+1} kết hợp với X được định nghĩa đệ quy như sau

$$\begin{aligned} U_0 &= (X^+)^{-1}X - \{\varepsilon\}, V_1 = U_0^{-1}X \cup (X^{-1}X - \{\varepsilon\}), \\ U_i &= (V_iX^*)^{-1}X, V_{i+1} = U_i^{-1}X \cup X^{-1}V_i \cup V_i,\ i \geq 1. \end{aligned} \tag{5.1}$$

Khi đó, định lý sau đây cung cấp một tiêu chuẩn cần và đủ để một ngôn ngữ có độ không nhập nhằng kiểu 1 hữu hạn.

Định lý 5.1 *Cho $X \subseteq \Sigma^+$ và cho $U_i, V_{i+1}\ (i \geq 1)$ được định nghĩa theo công thức (5.1). Khi đó X có độ không nhập nhằng kiểu 1 là $k-1$ khi và chỉ khi tồn tại một số nguyên $k \geq 1$ sao cho $\varepsilon \in V_k$ và $\varepsilon \notin V_i$ với mọi $i < k$.*

Xác định độ không nhập nhằng kiểu 1



Hình 5.1 Minh họa các bước tính toán các tập U_i, V_{i+1}

Phân bậc ngôn ngữ theo độ không nhập nhằng kiểu 1

Ta ký hiệu $\mathscr{L}_k$ là lớp ngôn ngữ k-không nhập nhằng kiểu 1. Khi đó $\mathscr{L}_0$ là lớp tất cả các ngôn ngữ, $\mathscr{L}_\infty$ là lớp ω-mã, và $\mathscr{L}_{Code}$ là lớp mã.

Vì $X \subseteq \Sigma^+$ là k-không nhập nhằng kiểu 1 với mọi $k \geq 0$ khi và chỉ khi X là mã. Do đó nếu X thuộc lớp $\mathscr{L}_{Code}$ thì X thuộc lớp $\mathscr{L}_i$ với mọi $i \geq 0$. Ta có

$$\mathscr{L}_{Code} = \bigcap_{i \geq 0} \mathscr{L}_i$$

và phân bậc ngôn ngữ

$$\mathscr{L}_\infty \subsetneq \mathscr{L}_{Code} \subsetneq \cdots \subsetneq \mathscr{L}_2 \subsetneq \mathscr{L}_1 \subsetneq \mathscr{L}_0.$$

Định nghĩa 5.2

(i) Một đồng cấu đa trị là một ánh xạ $f : \Sigma^* \to \Gamma^*$ tương ứng mỗi chữ cái $a \in \Sigma$ với một tập con X_a của Γ^* và $f(a_1a_2 \dots a_n) = f(a_1)f(a_2) \dots f(a_n)$ với mọi $a_1, a_2, \dots, a_n \in \Sigma$.

(ii) Đồng cấu đa trị f được gọi là một phép mã hóa đa trị nếu với mọi $\omega, \omega' \in \Sigma^*$, $\omega \neq \omega'$ thì $f(\omega) \cap f(\omega') = \emptyset$.

(iii) Đồng cấu đa trị f được gọi là một phép mã hóa đa trị hạn chế nếu có số nguyên $k > 0$ và với mọi $\omega, \omega' \in \Sigma^{\leq k}$, $\omega \neq \omega'$ thì $f(\omega) \cap f(\omega') = \emptyset$.

Phép mã hóa đa trị và phép mã hóa đa trị hạn chế

Kết quả cơ bản sau đây cung cấp một điều kiện cần và đủ để một đồng cấu là phép mã hóa đa trị, phép mã hóa đa trị hạn chế.

Mệnh đề 5.1 Giả sử Σ, Γ là các bảng hữu hạn các chữ cái. Cho đồng cấu đa trị $f : \Sigma^* \to \Gamma^*$, tương ứng mỗi chữ cái $a \in \Sigma$ với một tập con X_a của Γ^* và số nguyên $k > 0$. Khi đó,

(i) điều kiện cần và đủ để f là phép mã hóa đa trị là: nếu $X_1X_2 \dots X_p \cap X'_1X'_2 \dots X'_q \neq \emptyset$ thì $p = q$ và $X_i = X'_j$, với $i, j = 1, \dots, p$.

(ii) điều kiện cần và đủ để f là phép mã hóa đa trị hạn chế là: nếu $X_1X_2 \dots X_p \cap X'_1X'_2 \dots X'_q \neq \emptyset$ với $p, q \leq k$ thì $p = q$ và $X_i = X'_j$, với $i, j = 1, \dots, p$.

Phép mã hóa đa trị và phép mã hóa đa trị hạn chế

Để có thể sử dụng các ngôn ngữ không là mã trong các hệ mật đa trị và nhập nhằng MAS, ta thiết lập một kết quả cơ sở như sau:

Mệnh đề 5.2 Cho bảng chữ $\Sigma = \{a_1, a_2, \ldots, a_n\}$ và một số nguyên $k > 0$. Xét ngôn ngữ X có độ không nhập nhằng kiểu 1 là k, sao cho có thể phân hoạch X thành n tập con đôi một rời nhau $X_1, X_2, \ldots, X_n$, $X_i \cap X_j = \emptyset$, $\forall i \neq j, X_1 \cup X_2 \cup \cdots \cup X_n = X$.

Giả sử đồng cấu $g : \Sigma^* \to X^*$, tương ứng mỗi chữ cái $a_i \in \Sigma$ với một tập con X_i và $g(\omega\omega') = g(\omega)g(\omega')$ với mọi $\omega, \omega' \in \Sigma^{\leq k}$. Khi đó g là phép mã hóa đa trị hạn chế.

Hệ mật đa trị và nhập nhằng

Giả sử $\Sigma = \{a_1, a_2, \ldots, a_n\}$ là bảng chữ cái, X là ngôn ngữ có độ không nhập nhằng kiểu 1 là $k, k > 0$ sao cho có thể phân hoạch X thành m tập con đôi một rời nhau $X_1, X_2, \ldots, X_m$, $X_i \cap X_j = \emptyset$, $\forall i \neq j, X_1 \cup X_2 \cup \cdots \cup X_m = X$, $m \geq \mathrm{Card}(\Sigma)$. Ký hiệu X_P là tập tất cả các phân hoạch của X. Ta có

Định nghĩa 5.3 (Hệ mật MAS) Đặt $\mathcal{P} = \Sigma^{\leq k}$, $\mathcal{C} = X^*$. $\mathcal{K}$ gồm tất cả các đơn ánh $g : \Sigma \to X_P = \{X_1, X_2, \ldots, X_m\}$. Với mỗi $g \in \mathcal{K}$, định nghĩa

$$e_g(x) = w \in g(x),$$

và định nghĩa

$$d_g(w) = \{y \mid w \in g(y)\}.$$

Ví dụ 5.2 Cho bảng chữ $\Sigma = \{u_1, u_2, u_3, u_4, u_5\}$ và xét ngôn ngữ $X = \{c, ca_1, a_1b_1, b_1a_2, a_2b_2, b_2a_3, a_3\}$ có độ không nhập nhằng kiểu 1 là $k = 2$.

Một phương án phân hoạch tập X và ánh xạ g như sau: $X_1 = \{c\}$, $X_2 = \{ca_1, a_1b_1\}$, $X_3 = \{b_1a_2\}$, $X_4 = \{a_2b_2\}$, $X_5 = \{b_2a_3, a_3\}$, và ta có $g(u_i) \in X_i$, $i = 1, \ldots, 5$.

Giả sử, từ bản rõ cần mã hóa là $\omega = u_2u_3u_5$. Vì từ này có độ dài 3 nên ta tách ω thành 2 từ $\omega_1 = u_2u_3$ và $\omega_2 = u_5$ để đảm bảo độ dài của chúng nhỏ hơn hoặc bằng k.

Với phép mã hóa g được định nghĩa như trên, các từ mã nhận được là: $ca_1b_1a_2$ và a_3, hoặc $ca_1b_1a_2$ và b_2a_3, hoặc $a_1b_1b_1a_2$ và a_3, hoặc $a_1b_1b_1a_2$ và b_2a_3.

Ví dụ 5.2 (tiếp) Giải mã từ $ca_1b_1a_2$ ta sẽ nhận được hai từ của X, tương ứng là $ca_1 \in X_2$ và $b_1a_2 \in X_3$. Vì $ca_1 \in g(u_2)$ và $b_1a_2 \in g(u_3)$, từ bản rõ nhận được là u_2u_3.

Kết quả giải mã từ a_3 là từ bản rõ u_5. Ghép hai từ bản rõ nhận được, ta có từ bản rõ ban đầu. Các trường hợp khác giải mã tương tự và cho kết quả duy nhất là từ bản rõ ω.

Sự nhập nhằng có thể xảy ra khi ta mã hóa các từ bản rõ có độ dài lớn hơn k. Ví dụ trường hợp mã toàn bộ từ $\omega = u_2u_3u_5$, với g được định nghĩa như trên thì một từ mã mà ta nhận được có thể là $ca_1b_1a_2b_2a_3$. Khi đó giải mã sẽ cho các khả năng: $(c)(a_1b_1)(a_2b_2)(a_3)$ với $c \in X_1, a_1b_1 \in X_2, a_2b_2 \in X_4, a_3 \in X_5$, hoặc $(ca_1)(b_1a_2)(b_2a_3)$ với $ca_1 \in X_2, b_1a_2 \in X_3, b_2a_3 \in X_5$.

Tương ứng với hai từ bản rõ $u_1u_2u_4u_5$ và $u_2u_3u_5$.

Đánh giá độ an toàn của hệ mật MAS

Căn cứ nguyên tắc thiết kế của hệ mật, MAS có thể là mục tiêu của hai hình thức tấn công thám mã sau đây.

Hình thức 1. Kẻ tấn công không có thông tin về X và g (tức là, kiểu tấn công chỉ biết bản mã). Khi đó, thực hiện thám mã hệ mật MAS tương đương với nỗ lực giải bài toán tương ứng Post - là bài toán không quyết định được.

Hình thức 2. Kẻ tấn công không có thông tin về g (tức là, kiểu tấn công biết bản rõ). Khi đó, để xác định g cần thực hiện $S(m,n) \times n!$ phép thử khóa. Với n và m lần lượt là số phần tử của bảng chữ Σ và ngôn ngữ X, $S(m,n)$ là số Stirling loại 2. Ví dụ, với $n = 5, m = 8$ như trong Ví dụ 5.2, ta có $S(8,5) = 1050$ và $5! = 120$. Vậy, số cách chọn g là 126.000.

Cài đặt hệ mật đa trị và nhập nhằng

– Ta xem Σ và Γ là các tập con hữu hạn của tập $\{0,1\}^*$. Để đảm bảo yêu cầu về độ an toàn, ta chọn số phần tử của Σ là 128. Theo đó, mỗi phần tử của Σ được biểu diễn bởi một xâu 7 bit.

– Ta còn phải thiết kế bảng chữ Γ để từ đó xây dựng được ngôn ngữ bí mật X thỏa mãn ba điều kiện sau đây:

(1) Số phần tử của X phải lớn hơn 128 sao cho có thể phân hoạch X thành 128 tập con không rỗng.
(2) Độ dài trung bình các từ của X gần với 7 nhất (tức là, phép mã hóa X đạt hiệu quả tối ưu).
(3) X có độ không nhập nhằng kiểu 1 là $k, k > 0$.

Lưu ý $\rightsquigarrow$ Nhằm đáp ứng mục tiêu giải mã nhanh, tập Γ phải thỏa mãn tính chất "prefix-free", nghĩa là không có phần tử nào của Γ là *tiền tố* của một từ khác cũng thuộc Γ.

Cài đặt hệ mật đa trị và nhập nhằng

Với những yêu cầu trên đây, ta sẽ dùng mã Huffman có độ dài biến đổi để biểu diễn các phần tử của Γ. Cụ thể như sau

Letter	Huffman Code	Letter	Huffman Code
c	0001	d_9	00100000
a_1	1000	d_{10}	00100001
a_2	1001	d_{11}	00100010
a_3	0100	d_{12}	00100011
b_1	0101	$\cdots$	$\cdots$
b_2	1110	$\cdots$	$\cdots$
b_3	0110	d_{135}	11111110
a_4	0000	d_{136}	11111111

Cài đặt hệ mật đa trị và nhập nhằng

- Tiếp theo, ta sẽ dùng 128 phần tử của Γ để thiết kế X. Những phần tử khác chưa sử dụng của Γ sẽ dùng vào mục đích gây nhiễu.
- Ngôn ngữ bí mật X thỏa những điều kiện đặt ra là:
$X = \{c, ca_1, a_1b_1, b_1a_1, b_1a_2, a_2a_1, a_2b_1,$
$a_2b_2, b_2a_1, b_2a_2, b_2a_3, b_2b_1, a_3a_1, a_3a_2, a_3b_1,$
$a_3b_2, a_3b_3, b_3a_1, b_3a_2, b_3a_3, b_3b_1, b_3b_2, b_3c\} \cup$
$\{d_9, d_{10}, \ldots, d_{129}\}$.
- Số phần tử của X là 144. Ta có thể dễ dàng kiểm tra độ không nhập nhằng kiểu 1 của X là 4.

Cài đặt hệ mật đa trị và nhập nhằng

- Cho m là một số nguyên dương cố định và S là một xâu bit bí mật có độ dài m. Ta chọn một ngôn ngữ bí mật $X \subseteq \{0,1\}^*$ có độ không nhập nhằng kiểu 1 là $k > 0$ thỏa mãn điều kiện: với mọi $x_1, x_2, \ldots, x_k \in X$, ta có $|x_1| + |x_2| + \cdots + |x_k| \leq m$.
- Với X và Σ xác định như trên, ta thiết lập e_g và d_g như trong Định nghĩa 5.3.
- Đến đây, ta có thể mô tả một phương án cài đặt hệ mật MAS với hai thủ tục chính: ENCODE và DECODE.

Thủ tục mã hóa

procedure ENCODE(u)

$i = 1, j = 1$;

while $i \leq n$ do

$count = 1$;

while $(count \leq k)$ and $(|\omega_j| < m)$ do

if $|\omega_j e_g(u_i)| \leq m$ then

$\omega_j = \omega_j e_g(u_i)$, $count = count + 1, i = i + 1$

else PAD(ω_j);

if $|\omega_j| < m$ then PAD(ω_j);

if $j == 1$ then $\omega'_j = \omega_j \oplus S$ else $\omega'_j = \omega_{j-1} \oplus \omega_j$;

$j = j + 1$;

return $\omega = \omega'_1 \omega'_2 \cdots \omega'_{j-1}$

Thủ tục giải mã

procedure DECODE(ω)

$i = 1, j = 1$;

while $j \leq q$ do

if $j == 1$ then $\omega_j = \omega'_j \oplus S$ else $\omega_j = \omega_{j-1} \oplus \omega'_j$;

EXTRACT(ω_j, tmp);

$count = 1$;

while $(count \leq length(tmp))$ do

$u_i = d_g(tmp[count]), count = count + 1, i = i + 1$;

$j = j + 1$;

return $u = u_1 u_2 \cdots u_{i-1}$

Ví dụ 5.3 Cho bảng hữu hạn các chữ cái $\Sigma = \{u_1, u_2, u_3, u_4, u_5\}$ và $\Gamma = \{c, a_1, a_2, a_3, b_1, b_2, b_3, b_4\}$.

- Giả sử các phần tử của Γ được chọn theo một phân phối đều $1/8$, khi đó mã Huffman biểu diễn $c, a_1, a_2, a_3, b_1, b_2, b_3$ và b_4 lần lượt là $110, 001, 010, 011, 100, 101, 000$ và 111.
- Chọn $X = \{c, ca_1, a_1b_1, b_1a_2, a_2b_2, b_2a_3, a_3, ca_1a_3b_1\} \subseteq \Gamma^*$ có độ không nhập nhằng kiểu 1 là $k = 3$. Một trong các phân hoạch của X là: $X_1 = \{c\}$, $X_2 = \{ca_1, a_1b_1\}$, $X_3 = \{b_1a_2, ca_1a_3b_1\}$, $X_4 = \{a_2b_2\}$, $X_5 = \{b_2a_3, a_3\}$. Ta sẽ định nghĩa g bởi $g(u_i) \in X_i$, $i = 1, \ldots, 5$.
- Lấy $m = 18, S = 101000110110101100$ và giả sử bản rõ cần mã hóa là $u = u_2u_3u_5u_3u_4u_5u_2u_1u_3u_5$. Khi đó, một trong các bản mã nhận được từ ENCODE là $\omega = \omega'_1\omega'_2\omega'_3\omega'_4$.

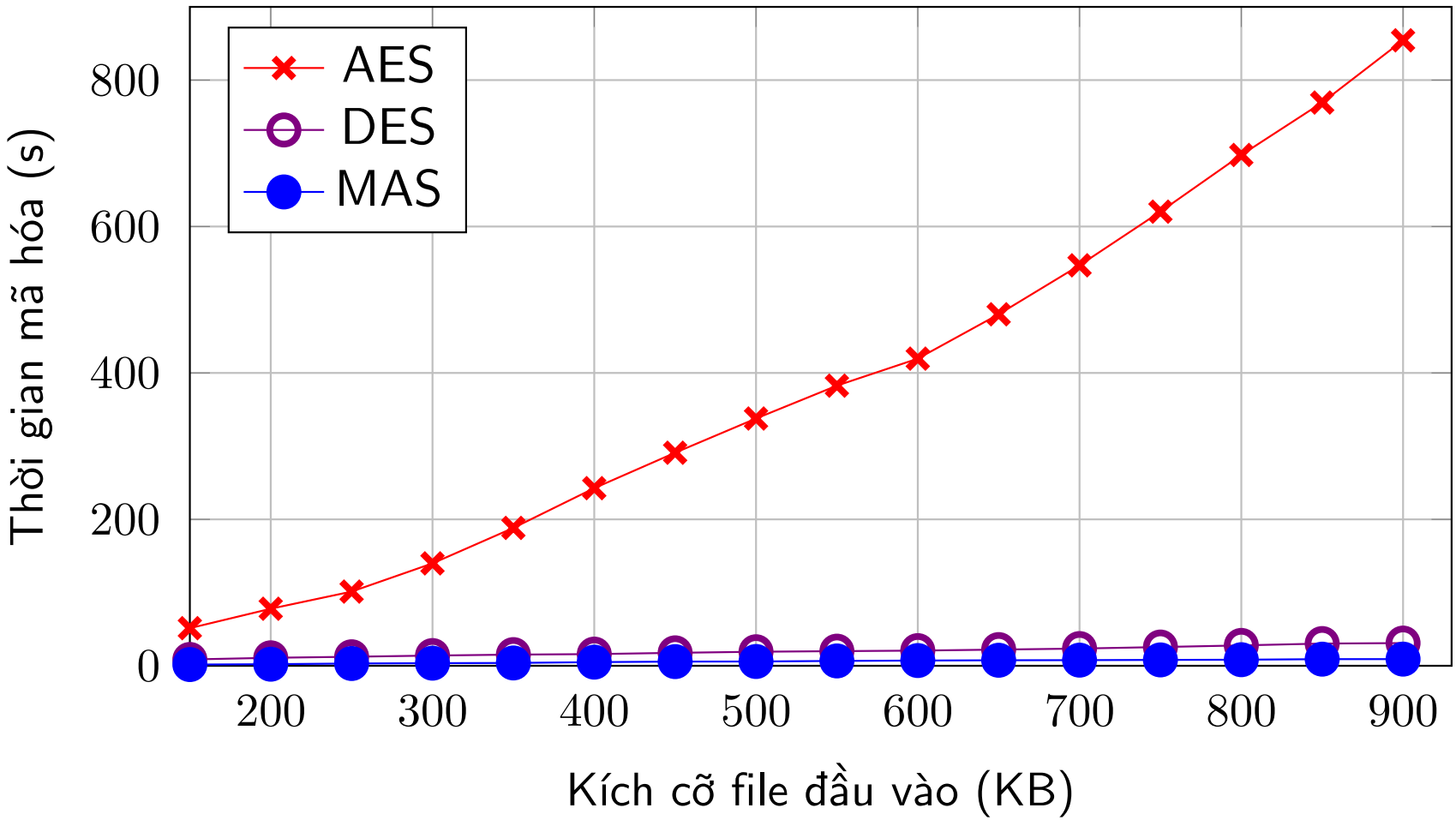
i	j	u_i	$e_g(u_i)$	w_j	PAD(w_j)	w'_j
1	1	u_2	ca_1	ca_1		
2	1	u_3	b_1a_2	$ca_1b_1a_2$		
3	1	u_5	b_2a_3	$ca_1b_1a_2b_2a_3$		01100101...
4	2	u_3	$ca_1a_3b_1$	$ca_1a_3b_1$		
5	2	u_4	a_2b_2	$ca_1a_3b_1a_2b_2$		00000011...
6	3	u_5	a_3	a_3		
7	3	u_2	a_1b_1	$a_3a_1b_1$		
8	3	u_1	c	$a_3a_1b_1c$	$a_3\underline{b_3}a_1b_1c\underline{b_3}$	10100101...
9	4	u_3	b_1a_2	b_1a_2		
10	4	u_5	b_2a_3	$b_1a_2b_2a_3$	$b_1a_2b_2\underline{b_3}a_3\underline{b_4}$	11101010...

Hình 5.2 Các bước tính toán của thủ tục ENCODE trong Ví dụ 5.3

i	j	$count$	ω_j	tmp	$tmp[count]$	u_i
1	1	1	11000110...	$ca_1b_1a_2b_2a_3$	ca_1	u_2
2	1	2	11000110...	$ca_1b_1a_2b_2a_3$	b_1a_2	u_3
3	1	3	11000110...	$ca_1b_1a_2b_2a_3$	b_2a_3	u_5
4	2	1	11000101...	$ca_1a_3b_1a_2b_2$	$ca_1a_3b_1$	u_3
5	2	2	11000101...	$ca_1a_3b_1a_2b_2$	a_2b_2	u_4
6	3	1	01100000...	$a_3a_1b_1c$	a_3	u_5
7	3	2	01100000...	$a_3a_1b_1c$	a_1b_1	u_2
8	3	3	01100000...	$a_3a_1b_1c$	c	u_1
9	4	1	10001010...	$b_1a_2b_2a_3$	b_1a_2	u_3
10	4	2	10001010...	$b_1a_2b_2a_3$	b_2a_3	u_5

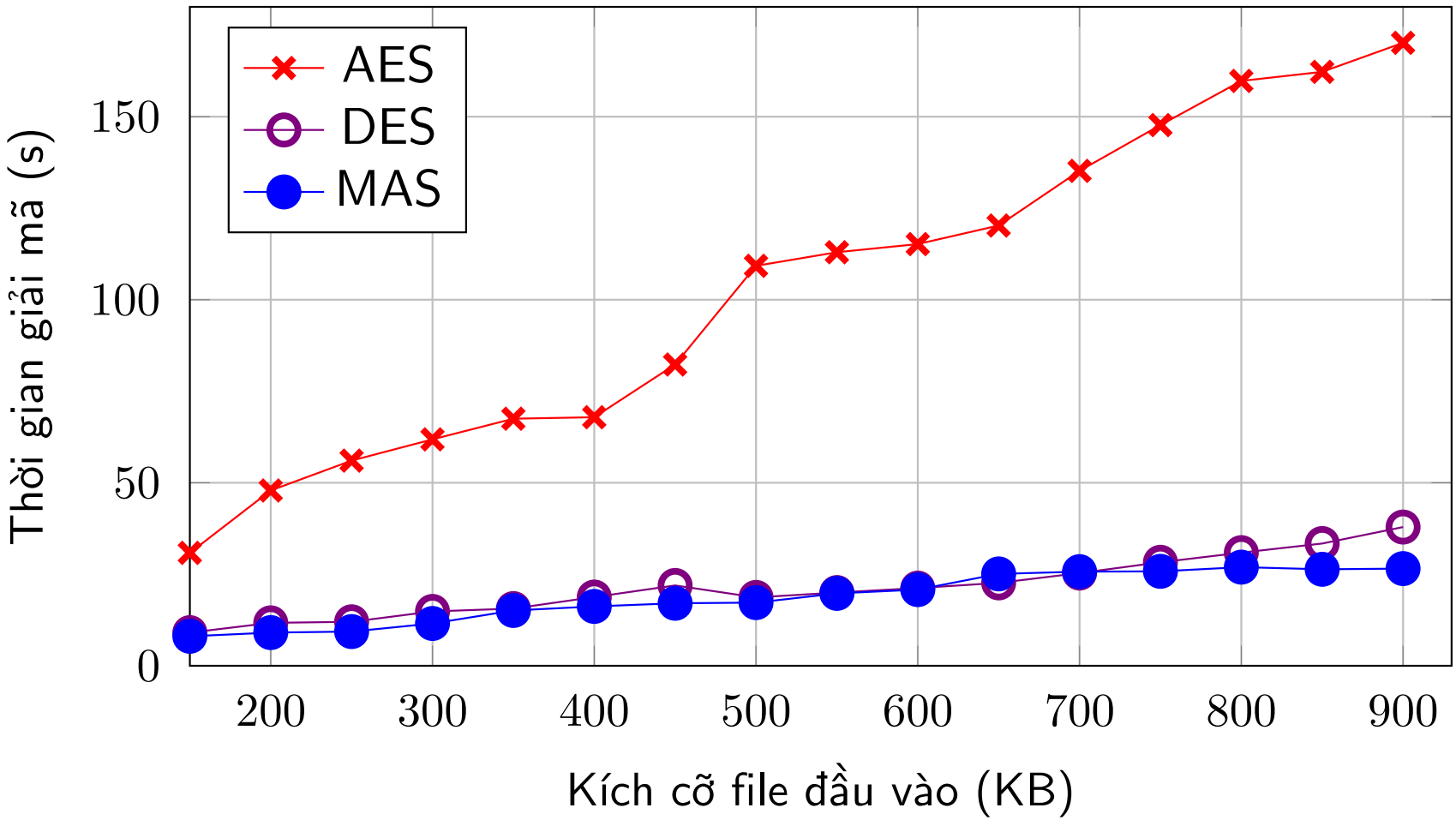
Hình 5.3 Các bước tính toán của thủ tục DECODE trong Ví dụ 5.3

Kết quả thực nghiệm



Hình 5.4 So sánh thời gian mã hóa của các hệ mật MAS, DES và AES

Kết quả thực nghiệm



Hình 5.5 So sánh thời gian giải mã của các hệ mật MAS, DES và AES

TRÂN TRỌNG CẢM ƠN!