Độ không nhập nhằng của ngôn ngữ và ứng dụng

On Unambiguity of Languages and Applications

Nguyễn Đình Hân, Đặng Quyết Thắng, Hồ Ngọc Vinh, Phan Trung Huy

***Abstract**: The classification of languages based on unambiguous product of words and codes contains a gap. Our work aims at investigations to fill up the gap. The classes of k-unambiguous languages is considered as extensions of codes, in which a code is k-unambiguous for all $k \geq 0$, the unambiguous product can be used to define k-unambiguous languages with $k \leq 2$. Given a regular language X, its unambiguous value k can be determined by an $O(n^2)$ time complexity algorithm, where n is the finite index of the syntactic congruence of X. The k-unambiguous languages with k is large enough, can be used in information encryption, and can provide us an encryption schema with high enough security since their ambiguous characteristics.*

I. GIỚI THIỆU

Mã có vai trò quan trọng trong lý thuyết và ứng dụng, đặc biệt trong lý thuyết thông tin và mật mã học. Khái niệm *tích không nhập nhằng* của hai ngôn ngữ được đề xuất bởi Schützenberger (1955) là cơ sở để xây dựng khái niệm mã. Cho hai ngôn ngữ $X, Y \subseteq A^*$, tích XY là *không nhập nhằng* nếu $w = xy = x'y'$ với $x, x' \in X$, $y, y' \in Y$ thì $x = x'$ và $y = y'$. Một tập X là mã nếu sự phân tích một từ w tùy ý thành tích các từ thuộc X là không nhập nhằng. Có nghĩa là nếu $w = x_1x_2 \ldots x_n = y_1y_2 \ldots y_m$, với $x_i, y_j \in X$ thì suy ra $n = m$, $x_i = y_i$, $i = 1,\ldots,n$. Đây là mối quan hệ giữa tích không nhập nhằng và mã. Nghiên cứu các tính chất không nhập nhằng liên quan đến sự phân tích mã là bài toán cơ bản của lý thuyết mã. Các tính chất đại số của mã dựa trên tích không nhập nhằng được nghiên cứu sâu sắc bởi Schützenberger (1955), Gilbert and Moore (1959) và các tác giả khác (xem [1-6]). Gần đây, nghiên cứu lý thuyết mã có xu hướng đưa vào các yếu tố điều khiển, đa trị, nhập nhằng để mở rộng khái niệm tích, từ đó xây dựng những lớp mã mới. Chẳng hạn z–mã dựa trên tích *zigzag* đưa vào bởi Anselmo [7,8], *T*–mã dựa trên tích trộn có điều khiển [9], *C*–mã dựa trên tích nhập nhằng sử dụng yếu tố ngữ cảnh [10,11] và ◇-mã [12].

Giữa định nghĩa của mã và tích không nhập nhằng có một khoảng trống. Nghiên cứu của chúng tôi sẽ làm rõ thông tin về khoảng trống này, đồng thời thiết lập những kết quả mới. Trong bài này, chúng tôi đề xuất khái niệm ngôn ngữ có *độ không nhập nhằng k*, với $0 \leq k \leq \infty$, tạo nên sự phân bậc toàn bộ các ngôn ngữ. Với $k = 0$ là lớp tất cả các ngôn ngữ, $k = \infty$ là lớp ω-mã, $k = 2$ liên quan đến tích không nhập nhằng. Từ đó khái niệm này làm mịn khoảng trống trong liên hệ giữa mã và tích không nhập nhằng và ta có thể sử dụng những ngôn ngữ không phải là mã để mã hóa thông tin với những giá trị k đủ lớn. Lưu ý rằng trong các hệ mật, mã là đối tượng bị 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.

Dưới góc độ đại số, mỗi mã là cơ sở của một vị nhóm tự do nào đó. Trong [13], cho trước một ngôn ngữ chính quy X bất kỳ bởi bộ ba (φ, M, B), với $\varphi: A^* \to M$ là một đồng cấu vị nhóm thỏa X, M là vị nhóm hữu hạn, $B \subseteq M$ sao cho $X = \varphi^{-1}(B)$, các tác giả đã mở rộng thuật toán Sardinas-Patterson để nhận được một thuật toán hiệu quả cỡ tuyến tính kiểm tra tính chất mã của một ngôn ngữ chính quy thay vì cỡ hàm mũ của

thuật toán kinh điển Sardinas-Patterson. Bằng kỹ thuật tương tự, chúng tôi đề xuất một thuật toán mới xác định độ không nhập nhằng k của một ngôn ngữ chính quy X bất kỳ. Kết hợp phương pháp tính toán trên một cấu trúc dữ liệu đặc biệt, chúng tôi nhận được một thuật toán có độ phức tạp thời gian $O(n^2)$ để xác định giá trị k với n là *chỉ số* của X.

II. MỞ ĐẦU

Trước khi giới thiệu khái niệm độ không nhập nhằng của ngôn ngữ, chúng tôi nhắc lại một số khái niệm và ký hiệu được sử dụng trong bài báo, chi tiết xem trong [2, 3]. Cho A là một bảng hữu hạn các chữ cái. A^* là vị nhóm tự do của tất cả các từ hữu hạn sinh bởi A. Từ rỗng được ký hiệu là ε và $A^+ = A^* - \{\varepsilon\}$. Một ngôn ngữ trên A là một tập con của A^*.

Cho $X \subseteq A^*$, ta nói X thỏa bởi đồng cấu vị nhóm $\varphi: A^* \to M$ nếu tồn tại $B \subseteq M$ sao cho $X = \varphi^{-1}(B)$. Trong trường hợp này, ta cũng nói X cho bởi một bộ ba (φ, M, B) khi φ, M, B xác định. Nếu X là ngôn ngữ chính quy, ta luôn chọn được vị nhóm cú pháp M_X hữu hạn và đồng cấu vị nhóm $\varphi_X : A^* \to M_X$ thỏa X. Ta gọi $k = \mathrm{Card}(M_X)$ là *chỉ số của tương đẳng cú pháp* của X hay ngắn gọn, k là *chỉ số* của X.

Sau đây chúng tôi định nghĩa khái niệm k-không nhập nhằng, k-nhập nhằng và độ không nhập nhằng của ngôn ngữ.

Định nghĩa 2.1. *Cho $X \subseteq A^+$ và một số tự nhiên $k \geq 0$. Khi đó,*

(i) *Tập X được gọi là k-không nhập nhằng nếu X thỏa mãn điều kiện: với mọi số nguyên $m \geq 1$ và với mọi phần tử $x_1, x_2, \ldots, x_k, y_1, y_2, \ldots, y_m$ thuộc X, nếu có $x_1x_2 \ldots x_k = y_1y_2 \ldots y_m$ thì suy ra $k = m$ và $x_i = y_i$ với $i=1,\ldots,k$.*

Ngược lại nếu X không thỏa mãn điều kiện trên, thì X được gọi là k-nhập nhằng.

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

Chú ý: Điều kiện X là k-nhập nhằng trong định nghĩa trên tương đương điều kiện tồn tại từ $w \in A^*$ mà $w = x_1x_2 \ldots x_k = y_1y_2 \ldots y_m$ với $x_1 \neq y_1$.

Nếu X có độ không nhập nhằng là $k \geq 0$ hữu hạn thì X là l-không nhập nhằng với mọi $0 \leq l \leq k$ và X là $(k+1)$-nhập nhằng. Rõ ràng, X là k-không nhập nhằng với mọi $k \geq 0$ khi và chỉ khi X là mã. Quy ước, tất cả các ngôn ngữ là 0-không nhập nhằng.

Ví dụ 2.1. Cho $A = \{c, a_1, a_2, a_3, b_1, b_2\}$, $X = \{c, ca_1, a_1b_1, b_1a_2, a_2b_2, b_2a_3, a_3\}$. Ta có thể kiểm tra bằng định nghĩa X là k-không nhập nhằng với mọi $0 \leq k \leq 2$ nhưng X là 3-nhập nhằng vì tồn tại từ

$$w = (ca_1)(b_1a_2)(b_2a_3) = (c)(a_1b_1)(a_2b_2)(a_3).$$

Ví dụ *2.2.* Cho $k \geq 1$ là một số tự nhiên tùy ý. Xem xét bảng chữ $A = \{c, a_1, b_1, \ldots, a_k, b_k\}$ và $X = \{c, ca_1, a_1b_1, b_1a_2, \ldots, b_{k-1}a_k, a_kb_k, b_k\}$. Rõ ràng, X là k-không nhập nhằng và X là $(k+1)$-nhập nhằng.

Phân bậc ngôn ngữ theo tính chất không nhập nhằng: Ta ký hiệu $\mathcal{L}_k$ *là lớp ngôn* ngữ *k-không nhập nhằng. Khi đó* $\mathcal{L}_0$ là lớp tất cả các ngôn ngữ, $\mathcal{L}_\infty$ là lớp ω-mã, và $\mathcal{L}_{Code}$ là lớp mã.

Vì $X \subseteq A^+$ là k-không nhập nhằng với mọi $k \geq 0$ khi và chỉ khi X là mã. Do đó nếu $X \in \mathcal{L}_{Code}$ thì $X \in \mathcal{L}_i$ với mọi $i \geq 0$. Ta có $\mathcal{L}_{Code} = \bigcap_{i \geq 0} \mathcal{L}_i$.

Từ Ví dụ 2.2, ta nhận được phân bậc toàn bộ các ngôn ngữ:

$$\mathcal{L}_\infty \subsetneq \mathcal{L}_{Code} \subsetneq \ldots \subsetneq \mathcal{L}_{k+1} \subsetneq \mathcal{L}_k \subsetneq \ldots \subsetneq \mathcal{L}_1 \subsetneq \mathcal{L}_0.$$

III. THUẬT TOÁN XÁC ĐỊNH ĐỘ KHÔNG NHẬP NHẰNG CỦA NGÔN NGỮ CHÍNH QUY

Cho $X, Y \subseteq A^*$. *Thương trái* và *thương phải* của X bởi Y được định nghĩa như sau:

$$Y^{-1}X = \{ w \in A^* \mid \exists y \in Y, yw \in X \},$$

$$XY^{-1} = \{ w \in A^* \mid \exists y \in Y, wy \in X \}.$$

Ký hiệu $u^{-1}X$, Xu^{-1} được sử dụng khi tập $Y = \{u\}$ chỉ có một phần tử.

Cho $X \subseteq A^+$, dựa vào định nghĩa k-không nhập nhằng, trong thủ tục sau đây ta tính toán các thương trái để tìm hai X-phân tích khác nhau của một xâu w bất kỳ thuộc A^*. Ta xem xét hai dãy các tập U_i, V_{i+1} được định nghĩa đệ quy như sau:

$$\begin{aligned} U_0 &= (X^+)^{-1}X - \{\varepsilon\}, \\ V_1 &= U_0^{-1}X \cup (X^{-1}X - \{\varepsilon\}), \\ U_i &= (V_iX^*)^{-1}X, \\ V_{i+1} &= U_i^{-1}X \cup X^{-1}V_i \cup V_i,\ i \geq 1. \end{aligned} \qquad (3.1)$$

Tính đúng đắn của thủ tục dựa trên các kết quả sau đây.

Bổ đề 3.1. *Cho $X \subseteq A^+$ và cho dãy các tập U_i, V_{i+1} ($i \geq 1$) được định nghĩa theo công thức* (3.1). *Nếu có $k \geq 1$ sao cho $z \in V_k$ và $z \notin V_i$ với mọi $i<k$ thì tồn tại $m \geq 1$ và các phần tử $x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m$ thuộc X sao cho $x_1x_2 \ldots x_kz = y_1y_2 \ldots y_m$ với $x_1 \neq y_1$.*

Chứng minh. Ta chứng minh quy nạp theo k điều khẳng định trên.

+ Với $k = 1$, giả sử $z \in V_1 = U_0^{-1}X \cup (X^{-1}X - \{\varepsilon\})$. Khi đó $z \in U_0^{-1}X$ (trường hợp 1) hoặc $z \in X^{-1}X - \{\varepsilon\}$ (trường hợp 2).

- Trường hợp 1: $z \in U_0^{-1}X$. Ta có $u_0 \in U_0$ và $x \in X$ sao cho $u_0z = x$. Từ $u_0 \in U_0=(X^+)^{-1}X - \{\varepsilon\}$, ta có $x_1x_2\cdots x_n \in X^+$, $n \geq 1$, $y_1 \in X$ sao cho $x_1x_2\cdots x_nu_0 = y_1$. Vì $u_0 \neq \varepsilon$ ta suy ra $x_1 \neq y_1$. Khi đó $x_1x_2\cdots x_nu_0 = y_1$ với $x_1 \neq y_1$. Nhân ghép hai vế của biểu thức trên với z và thay u_0z bởi x ta nhận được

$$x_1x_2\cdots x_nx = y_1z \text{ với } x_1 \neq y_1.$$

- Trường hợp 2: $z \in X^{-1}X - \{\varepsilon\}$. Ta có $x_1, y_1 \in X$ sao cho $x_1z = y_1$ với $x_1 \neq y_1$.

Vậy điều khẳng định đúng với $k = 1$.

+ Giả sử điều khẳng định đúng với $k \geq 1$, ta chứng minh nó cũng đúng với $k+1$.

Giả sử $z \in V_{k+1} = U_k^{-1}X \cup X^{-1}V_k \cup V_k$ và $z \notin V_i$ với mọi $i \leq k$. Khi đó $z \in U_k^{-1}X$ (trường hợp 1) hoặc $z \in X^{-1}V_k$ (trường hợp 2).

- Trường hợp 1: $z \in U_k^{-1}X$. Ta có $u_k \in U_k$ và $y \in X$ sao cho $u_kz = y$. Với $u_k \in U_k= (V_kX^*)^{-1}X$, tồn tại $v_k \in V_k$, $y'_1y'_2 \ldots y'_p \in X^*$, $p \geq 0$, $x_{k+1} \in X$ sao cho $y'_1y'_2 \ldots y'_pu_k = x_{k+1}$. Mặt khác, từ điều kiện $v_k \in V_k$, theo giả thiết quy nạp ta có:

$$x_1x_2 \ldots x_kv_k = y_1y_2 \ldots y_m \text{ với } x_1 \neq y_1.$$

Nhân ghép hai vế của biểu thức trên với $y'_1y'_2 \ldots y'_pu_kz$ và thay $y'_1y'_2 \ldots y'_pu_k$ bởi x_{k+1}, đồng thời thay u_kz bởi y, ta nhận được:

$$x_1x_2 \ldots x_kx_{k+1}z = y_1y_2 \ldots y_my'_1y'_2 \ldots y'_py,$$

với $x_1 \neq y_1$.

- Trường hợp 2: $z \in X^{-1}V_k$. Ta có $x_{k+1} \in X$ và $v_k \in V_k$ sao cho $x_{k+1}z = v_k$. Từ $v_k \in V_k$, theo giả thiết quy nạp ta có

$$x_1x_2 \ldots x_kv_k = y_1y_2 \ldots y_m \text{ với } x_1 \neq y_1.$$

Thay v_k bởi $x_{k+1}z$, ta có hệ thức $x_1x_2 \ldots x_{k+1}z = y_1y_2 \ldots y_m$ với $x_1 \neq y_1$.

Vậy điều khẳng định đúng với mọi $k \geq 1$.

Bổ đề 3.2. *Cho $X \subseteq A^+$ và cho dãy các tập U_i, V_{i+1} ($i \geq 1$) được định nghĩa theo công thức* (3.1). *Nếu có $m, n \geq 1$, $x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m \in X$ và $z \in A^*$ tùy ý sao cho $x_1x_2 \ldots x_nz = y_1y_2 \ldots y_m$ với $x_1 \neq y_1$ và $|z|<|y_m|$ thì $z \in V_n$.*

Chứng minh. Ta chứng minh quy nạp theo n điều khẳng định trên.

+ Với $n = 1$, giả sử có

$$x_1z = y_1y_2 \ldots y_m \text{ với } x_1 \neq y_1 \text{ và } |z|<|y_m|.$$

Từ $|z|<|y_m|$ ta có $y_m = uz$ với $u \in A^+$. Khi đó ta có quan hệ $x_1 = y_1y_2 \ldots y_{m-1}u$ với $x_1 \neq y_1$. Từ đó suy ra $u =$

$(y_1y_2 \dots y_{m-1})^{-1}x_1 \in (X^+)^{-1}X - \{\varepsilon\} = U_0$. Ta có $z = u^{-1}y_m \in U_0^{-1}X \subseteq V_1$. Vậy điều khẳng định đúng với trường hợp $n=1$.

+ Giả sử điều khẳng định đúng với $n \geq 1$, ta chứng minh nó cũng đúng với $n+1$. Giả sử có $x_1x_2 \dots x_{n+1}z = y_1y_2 \dots y_m$ với $x_1 \neq y_1$ và $|z| < |y_m|$. So sánh độ dài của từ $x_{n+1}z$ và từ y_m, ta phân biệt ba trường hợp :

- Trường hợp 1: $|x_{n+1}z| < |y_m|$. Theo giả thiết quy nạp, $x_{n+1}z \in V_n$. Khi đó với $x_{n+1}z \in V_n$ ta suy ra $z = x_{n+1}^{-1}(x_{n+1}z) \in X^{-1}V_n \subseteq V_{n+1}$. Vì vậy $z \in V_{n+1}$.
- Trường hợp 2: $x_{n+1}z = y_m$. Theo giả thiết quy nạp $\varepsilon \in V_n$. Khi đó $\varepsilon^{-1}x_{n+1} = x_{n+1} \in (V_nX^*)^{-1}X = U_n$, ta suy ra $z = x_{n+1}^{-1}y_m \in U_n^{-1}X \subseteq V_{n+1}$.
- Trường hợp 3: $|x_{n+1}z| > |y_m|$. Tồn tại $s \geq 1$, $z' \in A^*$, $|z'| < |y_s|$, $u \in A^+$ sao cho $x_{n+1} = z'y_{s+1}y_{s+2} \dots y_{m-1}u$ và $y_m = uz$. Điều này có nghĩa là $z' \in V_n$. Ta có $u = (z'y_{s+1}y_{s+2} \dots y_{m-1})^{-1}x_{n+1} \in (V_nX^*)^{-1}X = U_n$. Khi đó $z = u^{-1}y_m \in U_n^{-1}X \subseteq V_{n+1}$.

Vậy điều khẳng định đúng với mọi $n \geq 1$.

Trong định lý sau đây, ta thiết lập một điều kiện cần và đủ để một ngôn ngữ chính quy bất kỳ có độ không nhập nhằng hữu hạn. Định lý này được suy trực tiếp từ hai bổ đề trên.

Định lý 3.1. *Cho $X \subseteq A^+$ là một ngôn ngữ chính quy và cho dãy các tập U_i, V_{i+1} ($i \geq 1$) được định nghĩa theo công thức* (3.1). *Khi đó X có độ không nhập nhằng là k-1 khi và chỉ khi tồn tại số nguyên $k \geq 1$ sao cho $\varepsilon \in V_k$ và $\varepsilon \notin V_i$* với mọi $i < k$.

Đối với một ngôn ngữ X thuộc lớp ngôn ngữ chính quy, số các dãy U_i, V_{i+1} ($i \geq 1$) tương ứng với X là hữu hạn. Do đó Định lý 3.1 cung cấp cho ta một thuật toán xác định độ không nhập nhằng của X.

Từ Định lý 3.1 và các bổ đề trên, ta suy ra các tính chất tương đương sau đây:

(*i*) *X không là mã.*

(*ii*) *Tồn tại $i \geq 1$ sao cho $\varepsilon \in V_i$.*

(*iii*) *Tồn tại $k \geq 1$ sao cho X là k-nhập nhằng.*

Nhận xét 3.1. *Chú ý rằng các lớp ngôn ngữ k-không nhập nhằng hữu hạn mặc dù không là mã nhưng vẫn có thể sử dụng được để mã hóa thông tin.*

Ví dụ 3.1. Cho $A = \{c, a_1, a_2, a_3, b_1, b_2\}$, $X = \{c, ca_1, a_1b_1, b_1a_2, a_2b_2, b_2a_3, a_3\}$. Ta có $U_0 = \{a_1\}$, $V_1 = \{a_1, b_1\}$, $U_1 = \{b_1, a_2\}$, $V_2 = \{a_1, b_1, a_2, b_2\}$, $U_2 = \{b_1, a_2, b_2, a_3\}$, $V_3 = \{a_1, b_1, a_2, b_2, \varepsilon\}$. Vì $\varepsilon \in V_3$ *và* $\varepsilon \notin V_i$ với mọi $i \leq 2$, theo Định lý 3.1, độ không nhập nhằng của X là 2.

Giả sử có thông điệp $w = ca_1b_1a_2b_2a_3$. Nếu ta mã hóa thông điệp w thành các đoạn mã gồm không quá 2 từ của X thì phép giải mã đảm bảo tính duy nhất. Ví dụ w được mã thành hai đoạn mã w_1 và w_2: $w_1 = (c)(a_1b_1)$ *và* $w_2 = (a_2b_2)(a_3)$. Khi đó phép giải mã cho kết quả duy nhất:

$$w = w_1w_2 = (c)(a_1b_1)(a_2b_2)(a_3).$$

Ngược lại, phép giải mã là nhập nhằng. Ví dụ w được mã thành một đoạn mã có độ dài 3, $w_1 = (ca_1)(b_1a_2)(b_2a_3)$. Khi đó, phép giải mã cho hai khả năng: $w = (c)(a_1b_1)(a_2b_2)(a_3)$ hoặc $w = (ca_1)(b_1a_2)(b_2a_3)$.

Nhận xét 3.2. *Trong thực tiễn, có ngôn ngữ X thuộc lớp ngôn ngữ k-không nhập nhằng nhưng X chỉ có một số ít dãy nhập nhằng. Chẳng hạn, trong Ví dụ 2.2, X chỉ có một dãy nhập nhằng là $ca_1a_1b_1b_1a_2 \dots b_{k-1}a_kb_k$. Vì vậy, để nâng cao tính bảo mật cho các sơ đồ ứng dụng, ta phải xây dựng được X có nhiều dãy nhập nhằng. Bổ đề sau đây cho phép ta tìm được X là k-không nhập nhằng có ít nhất m dãy nhập nhằng, với $m \geq 1$.*

Bổ đề 3.3. *Cho $m \geq 1$, $k \geq 0$ và X^i ($i = 1,\dots,m$) là các ngôn ngữ k-không nhập nhằng, $X^i = \{c^i, c^ia^i_1, a^i_1b^i_1, b^i_1a^i_2, \dots, b^i_{k-1}a^i_k, a^i_kb^i_k, b^i_k\}$ sao cho $\bigcap_{i=1}^{m} X^i = \varnothing$. Khi đó ngôn ngữ $X = \bigcup_{i=1}^{m} X^i$* là *k-không nhập nhằng.*

Chứng minh. Giả sử X là k-nhập nhằng. Khi đó tồn tại dãy $x_1x_2 \ldots x_k = y_1y_2 \ldots y_m$, $m \geq 1$, $x_1, x_2, \ldots, x_k, y_1, y_2, \ldots, y_m \in X$, với $x_1 \neq y_1$. Không mất tính tổng quát, giả sử $|x_1|<|y_1|$, khi đó tồn tại $u \in A^*$ sao cho $x_1u = y_1$. Theo cách xác định các X^i thì chỉ có một khả năng $x_1 = c^i$ và $y_1 = c^i a^i_1$, với $c^i, c^i a^i_1 \in X^i$ nào đó. Tương tự, ta phải có $x_j, y_l \in X^i$, $j = 1,\ldots,k$, $l = 1,\ldots,m$. Từ đó suy ra X^i là k-nhập nhằng, trái với giả thiết X^i là k-không nhập nhằng. Vậy X là k-không nhập nhằng.

IV. ĐỘ PHỨC TẠP CỦA THUẬT TOÁN XÁC ĐỊNH ĐỘ KHÔNG NHẬP NHẰNG CỦA NGÔN NGỮ CHÍNH QUY

Trong phần này, với đầu vào là ngôn ngữ chính quy X biểu diễn bởi đồng cấu đại số, ta xây dựng một thuật toán có độ phức tạp thời gian $O(n^2)$ để xác định độ không nhập nhằng của X. Cho trước X bởi bộ ba (φ, M, B), Bổ đề 4.1 và Bổ đề 4.2 sau đây cho phép chúng ta xác định độ không nhập nhằng của X nhờ thực hiện tính toán trên vị nhóm hữu hạn thỏa X thay vì tính toán trực tiếp trên X.

Bổ đề 4.1. *Cho $X,Y \subseteq A^+$ là các ngôn ngữ chính quy, P là một vị nhóm hữu hạn và h: $A^* \to P$ là một toàn cấu vị nhóm. Nếu X và Y thỏa bởi h thì $(X^+)^{-1}Y$ thỏa bởi h.*

Chứng minh. Giả sử X và Y thỏa bởi h, theo định nghĩa có $K, L \subseteq P$ sao cho $X = h^{-1}(K)$, $Y = h^{-1}(L)$. Đặt $h(X^+) = T$, ta có $K^+ = T$. Bằng định nghĩa ta có thể kiểm tra $(X^+)^{-1}Y = h^{-1}(T^{-1}L)$.

Xét toàn cấu h: $A^* \to P$ thỏa X và $Y = \{\varepsilon\}$, trong đó P là vị nhóm hữu hạn. $X = h^{-1}(K)$, $Y = h^{-1}(L)$, $L = \{1_P\}$, $T = K^+$, $S = K^*$ với $K, L \subseteq P$. Ta xem xét dãy các tập Q_i, R_{i+1} được định nghĩa đệ quy như sau:

$$\begin{aligned} &Q_0 = T^{-1}K - L, \\ &R_1 = Q_0^{-1}K \cup (K^{-1}K - L), \\ &Q_i = (R_iS)^{-1}K, \\ &R_{i+1} = Q_i^{-1}K \cup K^{-1}R_i \cup R_i,\ i \geq 1. \end{aligned} \qquad (4.1)$$

Ta có thể kiểm tra trực tiếp bằng định nghĩa bổ đề sau đây.

Bổ đề 4.2. *Cho $X \subseteq A^+$ là một ngôn ngữ chính quy và cho dãy các tập U_i, V_{i+1} $(i \geq 1)$ được định nghĩa theo công thức* (3.1). *Đặt $U_i = h^{-1}(Q_i)$* và *$V_i = h^{-1}(R_i)$, với Q_i và R_{i+1} được định nghĩa theo công thức* (4.1). *Khi đó với mọi $k \geq 1$, $\varepsilon \in V_k$ khi và chỉ khi $1_P \in R_k$, với 1_P là phần tử đơn vị của vị nhóm P.*

Từ đó ta thiết lập được thuật toán như sau:

Thuật toán xác định độ không nhập nhằng của ngôn ngữ chính quy

Dữ liệu vào: Ngôn ngữ chính quy $X \subseteq A^+$ cho bởi một bộ ba (φ, M, B), với φ: $A^* \to M$ là một đồng cấu vị nhóm thỏa X, M là vị nhóm hữu hạn, $B \subseteq M$, $X = \varphi^{-1}(B)$.

Dữ liệu ra: k là độ không nhập nhằng của X.

Bước 1. Từ bộ ba (φ, M, B), ta xây dựng toàn cấu vị nhóm h: $A^* \to P \subseteq M \times Z_2$, với $Z_2 = \{0, 1\}$ là vị nhóm với phép nhân, thỏa X và $\{\varepsilon\}$ định nghĩa bởi: $\forall a \in A$, $h(a) = (\varphi(a), 0)$, $1_P = h(\varepsilon) = (1_M, 1)$. Đặt $\overline{K} = h(A) \cup 1_P = \{(h(a), 0) \mid a \in A\} \cup 1_P$. Tính $\overline{K}^* = P$.

Bước 2. Tính $K = P \cap B \times Z_2 = \{(b,m) \in P \mid b \in B\}$, ta có $h^{-1}(K) = X$, *và* $L = \{(1_M, 1)\}$, ta có $h^{-1}(L) = \{\varepsilon\}$.

Bước 3. Tính $T = K^+$, $S = K^*$.

Bước 4. Tính $Q_0 = T^{-1}K - L$, $R_1 = Q_0^{-1}K \cup (K^{-1}K - L)$ và đặt n=2.

If $Q_0 = \varnothing$ Then *Return* $k = \infty$.

Bước 5 (Lặp).

Tính $Q_n = (R_nS)^{-1}K$, $R_{n+1} = Q_n^{-1}K \cup K^{-1}R_n \cup R_n$.

If $1_P \in R_n$ Then *Return* $k = n$-1.

If $R_n = R_{n-1}$ Then *Return* $k = \infty$

Else đặt $n = n+1$, và quay về Bước 5.

Chi tiết độ phức tạp của thuật toán:

Biểu diễn chi tiết thuật toán bằng cấu trúc dữ liệu, khởi đầu, ta xây dựng một tập các node, mỗi node z bao gồm {*z.left, z.right, z.father, z.brother, z.flag*$_1$, *z.flag*$_2$, *z.flag*$_3$, *z.flag*$_4$, *z.flag*$_5$} sao cho với các phần tử x, y bất kỳ thuộc M, nếu $z=xy$, thì ta đặt *z.left* = x, *z.right* = y, *x.father* = *y.father* =z, *y.brother* = x. Quá

trình này được thực hiện trong thời gian $O(n^2)$ với n = Card(M).

Để thực hiện Bước 1, ta sử dụng hai stack ký hiệu lần lượt là *stack*1 và *stack*2. Ở bước con thứ nhất, với mỗi phần tử z thuộc $\overline{K}$, ta đặt $z.flag_1 = n+1$ và đẩy z vào *stack*1. Với mỗi phần tử z thuộc $M - \overline{K}$, ta đặt $z.flag_1 = n+2$ và đẩy z vào *stack*2. Ở bước con kế tiếp, với mọi z thuộc *stack*2 thỏa điều kiện $z.left.flag_1 = z.right.flag_1 = n+1$, ta đặt $z.flag_1 = n+1$ và đẩy z vào *stack*1. Quá trình này sẽ dừng sau tối đa n bước con và khi đó *stack*1 chứa các phần tử của P. Bước 1 có thể hoàn thiện trong thời gian $O(n^2)$.

Trong Bước 2, với mỗi phần tử z của *stack*1, nếu z thuộc K, thì ta đặt $z.flag_1=0$. Công việc này đòi hỏi thời gian tuyến tính theo cỡ của P, tức là $O(n)$.

Để thực hiện Bước 3, bằng phương pháp tương tự Bước 1, với K nằm trong *stack*1 và dùng *stack*2 như bộ nhớ tạm, ta nhận được $T=K^+$ và $S=K^*$. Ta có: $z\in T \Leftrightarrow z.flag_2=0$ và $z\in S \Leftrightarrow z.flag_3=0$, với z là phần tử của *stack*1. Bước 3 có độ phức tạp thời gian là $O(n^2)$.

Để thực hiện Bước 4, trên *stack*1, ta tính Q_0 bởi điều kiện: $z\in Q_0 \Leftrightarrow z.flag_4=0 \Leftrightarrow (z \neq 1_P$ và $z.father.flag_1=0$ và $z.brother.flag_2=0)$ và ta tính R_1 bởi điều kiện: $z\in R_1 \Leftrightarrow z.flag_5=1 \Leftrightarrow ((z \neq 1_P$ và $z.father.flag_1=0$ và $z.brother.flag_1=0)$ hoặc $(z.father.flag_1=0$ và $z.brother.flag_4=0))$ với z là phần tử của *stack*1. Bước này có thể hoàn thiện trong thời gian $O(n)$.

Trong Bước 5, ta thực hiện một vòng lặp để tính dãy các tập Q_n, R_{n+1}, $n\geq 1$ với các giá trị khởi đầu K, S, Q_0, R_1 nằm trong *stack*1. Ở bước con thứ $i>1$, với mọi phần tử z thuộc P, ta có: $z\in Q_i \Leftrightarrow z.flag_4=i \Leftrightarrow z.father.flag_1= 0$ và $z.brother.left.flag_5\geq 1$ và $z.brother.right.flag_3=0$. Sau khi tính Q_i, ta tính tiếp R_{i+1}. Ta có: $z\in R_{i+1} \Leftrightarrow z.flag_5=i+1 \Leftrightarrow z.flag_5\geq 1$ hoặc $(z.brother.flag_4= i$ và $z.father.flag_1= 0)$ hoặc $(z.brother.flag_1= 0$ và $z.father.flag_5\geq 1)$. Quá trình này sẽ dừng sau tối đa n bước con. Ta có: $1_P\in R_n \Leftrightarrow 1_P.flag_5\geq 1$. Tất cả công việc này có thể hoàn thiện trong thời gian $O(n^2)$.

Vậy, thời gian thực hiện của toàn bộ 5 bước của thuật toán là $O(n^2)$.

Lưu ý rằng độ phức tạp $O(n^2)$ của thuật toán trên đây cũng có thể nhận được nhờ phương pháp trong [14]. Khi đó thay vì sử dụng kỹ thuật stack, ta sẽ áp dụng một số kỹ thuật trên đồ thị hữu hạn (chi tiết xem trong [14]).

Từ đó, ta có kết quả sau:

Mệnh đề 4.1. *Cho $X \subseteq A^+$ là một ngôn ngữ chính quy chỉ số n. Tồn tại một thuật toán có độ phức tạp thời gian $O(n^2)$ xác định độ không nhập nhằng của X.*

V. KẾT LUẬN

Trong bài báo, chúng tôi giới thiệu khái niệm *k*-không nhập nhằng, *k*-nhập nhằng với *k* là một số tự nhiên và độ không nhập nhằng của ngôn ngữ. Những khái niệm này làm mịn khoảng trống trong liên hệ giữa mã và tích không nhập nhằng. Một phân bậc ngôn ngữ theo tính chất không nhập nhằng đã được thiết lập và ta có thể sử dụng các lớp ngôn ngữ *k*-không nhập nhằng để mã hóa thông tin với giá trị *k* đủ lớn. Từ đó cho thấy đặc trưng độ không nhập nhằng có vai trò ứng dụng trong mã hóa và việc nghiên cứu những đặc trưng này của ngôn ngữ là cần thiết. Thuật toán sử dụng kỹ thuật tính toán trên vị nhóm biểu diễn ngôn ngữ có độ phức tạp cỡ đa thức được xây dựng cho trường hợp ngôn ngữ chính quy và Bổ đề 3.3 cho phép chúng ta đề xuất những sơ đồ ứng dụng cụ thể.

LỜI CẢM ƠN

Nhóm tác giả trân trọng cảm ơn các phản biện đã xem xét kỹ và đóng góp ý kiến quý báu giúp cho chất lượng trình bày của bài báo được nâng cao.

TÀI LIỆU THAM KHẢO

[1] E. N. GILBERT, E. F. MOORE, *"Variable length binary encodings"*. Bell System Tech, J., pp. 933-967, 1959.

[2] G. LALLEMENT, *Semigroups and Combinational Applications*. John Wiley & Sons, Inc, 1979.

[3] J. BERSTEL, D. PERRIN, C. REUTENAUER, *Theory of Codes*. Academic Press, Inc, 1985.

[4] J. DEVOLDER, M. LATTEUX, I. LITOVSKY, L. STAIGER, *"Codes and infinite words"*, Acta Cybernetica, Vol. 11, No. 4, 1994. Szeged.

[5] M. ANSELMO, *"A non-ambiguous language factorization problem"*, In Proceedings of Developments in Language Theory, pp. 141-152, 1999.

[6] M. ANSELMO, *"A non-ambiguous decomposition of regular languages and factorizing codes"*, Discrete Applied Mathematics. Vol. 126, Issues 2-3, pp. 129-165, 2003.

[7] D.L. VAN, B. LESAËC, I. LITOVSKY, *"On coding morphisms for zigzag codes"*, Informatique théorique et applications, 26 (6), pp. 565-580, 1992.

[8] M. ANSELMO, *"On zigzag codes and their decidability"*, Theoretical Computer Science, 74, pp. 341-354, 1990.

[9] A. MATEESCU, G.D. MATEESCU, G. ROZENBERG, A.SALOMAA, *"Shuffle-Like Operations on ω-words"*, New Trends in Formal Languages, Lecture Notes in Computer Science, Vol. 1218, pp. 395-411, 1990. Springer-Verlag, Berlin, Heidelberg.

[10] PHAN TRUNG HUY, VŨ THÀNH NAM, *"Về một hình thức mã mới"*, Kỷ yếu Hội thảo Quốc gia lần thứ VI *"Một số vấn đề chọn lọc của công nghệ thông tin"*, Thái Nguyên, 8/2003, pp. 164-170.

[11] PHAN TRUNG HUY, VŨ THÀNH NAM, *"Mã luân phiên và mã tiền ngữ cảnh"*, Kỷ yếu Hội thảo quốc gia lần thứ VIII *"Một số vấn đề chọn lọc của công nghệ thông tin"*, Đà Nẵng, 8/2004, pp. 188-197.

[12] HỒ NGỌC VINH, NGUYỄN ĐÌNH HÂN, PHAN TRUNG HUY, *"Mã với tích biên và Độ trễ giải mã"*, Tạp chí Công nghệ Thông tin và Truyền thông, Tập V-1, Số 4 (24), pp. 46-56, 2010.

[13] NGUYỄN ĐÌNH HÂN, HỒ NGỌC VINH, PHAN TRUNG HUY, ĐỖ LONG VÂN, *"Thuật toán xác định tính chất mã của ngôn ngữ chính quy"*, Tạp chí Tin học và Điều khiển học, Tập 27, Số 1, pp.1-8, 2011.

[14] NGUYỄN ĐÌNH HÂN, PHAN TRUNG HUY, ĐẶNG QUYẾT THẮNG, *A quadratic algorithm for testing of omega-codes*. In: J.-S. Pan, S.-M. Chen, N.T. Nguyen (Eds.): ACIIDS 2012, Part I, LNAI 7196, pp. 338-347. Springer, Heidelberg, 2012.

Ngày nhận bài: 07/07/2011

SƠ LƯỢC VỀ TÁC GIẢ

NGUYỄN ĐÌNH HÂN

Sinh năm 1977 tại Hưng Yên.

Tốt nghiệp Trường ĐH Quốc gia Hà Nội năm 2000, tốt nghiệp Cao học tại AIT năm 2005.

Hiện đang công tác tại Trường Đại học Sư phạm Kỹ thuật Hưng Yên.

Lĩnh vực nghiên cứu: lý thuyết mã và ứng dụng, ngôn ngữ hình thức, an toàn và bảo mật thông tin.

Email: hannguyen@utehy.edu.vn

ĐẶNG QUYẾT THẮNG

Sinh năm 1970 tại Lai Châu.

Tốt nghiệp Đại học Sư phạm Hà Nội 2 - ngành Toán năm 1991, Đại học Bách Khoa Hà Nội - ngành Tin học năm 1998; tốt nghiệp Cao học tại Học viện Kỹ thuật Quân sự năm 2005.

Hiện đang công tác tại Trường Đại học Sư phạm Kỹ thuật Nam Định.

Lĩnh vực nghiên cứu: lý thuyết mã, ngôn ngữ hình thức, an toàn và bảo mật thông tin.

Email: thangdgqt@gmail.com

HỒ NGỌC VINH

Sinh năm 1977 tại Nghệ An.

Tốt nghiệp Đại học Quốc gia Hà Nội năm 1999, tốt nghiệp Cao học tại Trường Đại học Bách khoa Hà Nội năm 2003.

Hiện đang công tác tại Khoa CNTT, Trường Đại học Sư phạm Kỹ thuật Vinh, Nghệ An.

Lĩnh vực nghiên cứu: lý thuyết mã, lý thuyết nửa nhóm, vị nhóm, ngôn ngữ hình thức, an toàn và bảo mật thông tin.

Email: hnvinh.skv@moet.edu.vn

PHAN TRUNG HUY

Tốt nghiệp Đại học Sư phạm I Hà Nội - Khoa Toán năm 1976. Nhận học vị Tiến sĩ năm 1992 ngành Tin học - Đại số tại Viện Toán học Việt Nam. Được phong chức danh Phó Giáo sư năm 2004.

Hiện công tác tại Viện Toán - Tin ứng dụng, Trường Đại học Bách Khoa Hà Nội.

Lĩnh vực nghiên cứu: Tin học - Đại số- về Lý thuyết đa tạp các vị nhóm hữu hạn và đa tạp các ngôn ngữ từ vô hạn; Ngôn ngữ hình thức từ hữu hạn và vô hạn; Otomat và thuật toán; Lý thuyết mã và ứng dụng; Mô phỏng tính toán lượng tử; Giấu tin trong ảnh. Đồ thị và tổ hợp trên từ.

Email: huyfr2002@yahoo.com