Tạp chí Tin học và Điều khiển học, T. 18, S. 3 (2002), 253–261

MỘT SỐ ĐẶC TRƯNG NHẬP NHẰNG CỦA MÃ

PHAN TRUNG HUY, VŨ THÀNH NAM

Abstract. In this paper we study some ambiguous properties concerning codes of finite words in context related to infinite words. Some results and algorithms to calculate measures of ambiguities of codes are established.

Tóm tắt. Trong bài này, chúng tôi đưa vào nghiên cứu một số đặc trưng tính nhập nhằng có quan hệ đến mã từ hữu hạn trên cơ sở áp dụng một vài kết quả có liên quan tới ngôn ngữ từ vô hạn và khái niệm độ trễ hữu hạn và vô hạn. Một số tập đặc trưng cho sự nhập nhằng liên quan tới mã từ hữu hạn được đưa vào. Kết quả biểu diễn cho các tập đặc trưng nhập nhằng với mỗi mã cho trước được thiết lập. Dựa vào việc đưa vào nghiên cứu một độ đo ngôn ngữ, có thể định nghĩa các độ đo nhập nhằng của mã liên quan. Từ đó, các thuật toán hiệu quả để tính độ trễ đồng bộ và tính độ đo tính nhập nhằng cho mã chính quy được thiết lập.

1. XUẤT XỨ VẤN ĐỀ

Việc nghiên cứu các khía cạnh về độ không nhập nhằng liên quan đến mã đã được nhiều công trình đề cập, chẳng hạn [1–4]. Bài viết này đề cập một hướng nghiên cứu mới: tính nhập nhằng của mã và độ đo tính nhập nhằng mà một phần đã đề cập trong [5]. Từ đó thiết lập một số kết quả biểu diễn tính nhập nhằng và chứng tỏ rằng có thuật toán hữu hiệu để tính độ nhập nhằng của các mã chính quy từ hữu hạn. Để nhận được các biểu diễn về tập nhập nhằng, ở đây chúng tôi áp dụng một số tính chất liên quan đến ngôn ngữ từ hữu hạn và vô hạn chính qui như đã xét trong [6, 7].

Như ta đã biết, thông thường khi nói về mã, ta thường hình dung không có tính nhập nhằng khi phân tích từ. Đối với nhiều lớp mã, chỉ cần nhận được phần đầu của thông điệp, chúng ta có thể xác định phân tích các từ trong thông điệp khi giải mã (nếu có độ trễ hữu hạn, xem [3]).

Tuy nhiên không phải luôn luôn khi nhận được chỉ một phần thông điệp, chúng ta có thể xác định được phân tích từ của nó, đặc biệt với lớp mã có độ trễ vô hạn hoặc không là Z-mã (xem [2]). Chúng ta chỉ xác định được sự giải mã đúng khi đã nhận được đầy đủ thông điệp. Đây là khía cạnh của tính nhập nhằng liên quan đến mã, là đối tượng chính được nghiên cứu trong bài viết này

Việc áp dụng tính nhập nhằng đó là khả thi, ví dụ để tăng độ bảo mật của mã. Về lý thuyết, để đi đến ứng dụng, ta cần định nghĩa một độ đo tính nhập nhằng. Sau đây ta nhắc lại một số khái niệm, ký hiệu cần thiết.

Một số kí hiệu và khái niệm sơ bộ

Giả sử A là tập hữu hạn hoặc vô hạn các ký hiệu mà ta gọi là bảng chữ cái. Trong bài này, các bảng chữ giả thiết là hữu hạn.

Tập tất cả các từ hữu hạn trên A bao gồm cả từ rỗng ε được ký hiệu là A^*, tập $A^* - \{\varepsilon\} = A^+$ là nửa nhóm tự do sinh bởi A. Tích các từ là phép *nối ghép từ* (concatenation). Tập các từ vô hạn trên A được kí hiệu bởi A^ω và tập các từ có cả từ vô hạn , hữu hạn được kí hiệu bởi A^∞. Nếu thay A bởi X ta có các ký hiệu tương tự X^+, X^∞.

Tập con $X \subseteq A^+$ gọi là *mã* nếu mỗi từ $w \in A^+$ không có quá 1 phân tích trong $X (w = x_1 ... x_n, n \geq 1, \quad x_i \in X)$.

Giả sử $X \subseteq A^+, Y \subseteq A^\infty$ ta xác định tập các thương $X^{-1}Y = \{u \in A^\infty \mid \exists x \in X : ux \in Y\}$. Để đơn giản ta sẽ ký hiệu $(A^*)^{-1}X$ bởi $A^{-}X$, và giả sử $X, Y \subseteq A^\infty$ ta xác định tập các thương $YX^{-1} = \{u \in A^* \mid \exists x \in X : ux \in Y\}$. Đơn giản ta sẽ ký hiệu $X(A^*)^{-1}$ bởi XA^{-}.

Cho $X \subseteq A^\infty$, *Pref(X)* và *Suff(X)* lần lượt là tập các *khúc đầu* và tập các *khúc đuôi* của các từ $x \in X$. Như vậy ta có ngay $Pref(X) = XA^-$ và $Suff(X) = A^-X$.

Cho $X, Y \subset A^*$, tích XY là *không nhập nhằng* nếu từ $xy = x'y', \ \ x, x' \in X, \ y, y' \in Y$ suy ra $x = x'$ và $y = y'$. Luỹ thừa X^+ là *không nhập nhằng* nếu X là mã.

Một đồng cấu vị nhóm $h : A^* \to M$ từ vị nhóm tự do A^* đến vị nhóm M được gọi là *thoả ngôn ngữ* $L \subseteq A^*$ nếu $L = h^{-1}\ h(L)$ và gọi là *thoả ngôn ngữ từ vô hạn* $W \subseteq A^\omega$ nếu với mỗi cặp phần tử $e, f \in M$, nếu $h^{-1}(e)[h^{-1}(f)]^\omega \cap W \neq \varnothing$ thì $h^{-1}(e)[h^{-1}(e)][h^{-1}(f)]^\omega \subseteq W$.

1.1. Độ trễ của mã

Giả sử X là tập con của A^*. Khi đó ta nói X có *độ trễ giải mã hữu hạn* nếu tồn tại số nguyên $d \geq 0$ sao cho:

$$\forall x, \ x' \in X; \ \forall y \in X^d; \ \forall u \in A^*, \ xyu \in x'X^* \Rightarrow x = x' \tag{1.1}$$

Nếu X có độ trễ giải mã hữu hạn thì *số nguyên nhỏ nhất thoả (1.1) gọi là độ trễ giải mã* của X. Để ý rằng khái niệm được xác định với chiều từ trái sang phải, tất nhiên ta có khái niệm đối xứng nếu xét từ phải sang trái.

Nếu không tồn tại d như vậy, ta nói rằng mã có độ trễ vô hạn, xem như $d = \infty$.

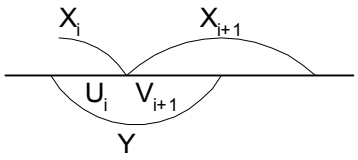
Liên quan đến độ trễ ta có định lý sau (xem [3]).

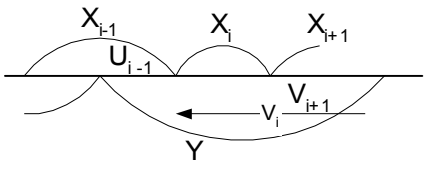
Định lý 1.1. *Mọi mã hữu hạn cực đại với độ trễ giải mã hữu hạn là mã prefix.*

Thuật toán tính độ trễ:

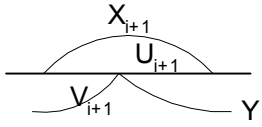
Ta có thể xây dựng thuật toán tính độ trễ hữu hạn, tương tự ý tưởng thuật toán Sardinas-Patterson. Khác với thuật toán Sardinas-Patterson, ta phân biệt 2 lớp cắt: lớp các phần dư bên phải với từ trong X_{i+1}, gọi là tập V_i và lớp các phần dư bên trái gọi là lớp U_i. Hai lớp này được xác định như sau:

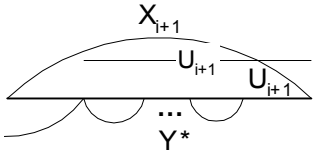
$Y = X, V_0 = Y;$ $\qquad V_1 = U_0^{-1}Y + X^{-1}V_0$
$U_0 = Y^{-1}X - \varepsilon;$ $\qquad V_{i+2} = U_{i+1}^{-1}Y + X^{-1}V_{i+1}$
$U_0 = U_0 + Y^-U_0$





$U_{i+1} = V_{i+1}^{-1}X$
$U_{i+1} = U_{i+1} + Y^-U_{i+1}$





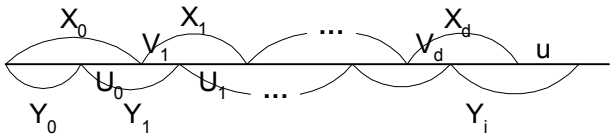
Ta có định lý sau

Định lý 1.2 *Tập X có độ trễ giải mã hữu hạn d khi và chỉ khi $V_{d+1} = \varnothing$. Với U_i và V_{i+1} xác định như trên.*

Chứng minh. Ta chứng tỏ nếu X không có độ trễ giải mã hữu hạn thì $V_n \neq \varnothing$ với mọi n.

Thật vậy, nếu X không có độ trễ giải mã hữu hạn thì với mọi n ta có tồn tại từ w có phân tích:

$$w = x_0x_1...x_nu = y_0y_1...y_m, \text{ với } x_i \in X, \ y_j \in Y, \ u \in A^+, \ x_0 \neq y_0.$$



Từ cách xác định V_i ta có $u \in V_{n+1}$ nghĩa là $V_{n+1} \neq \varnothing$.

Ngược lại, ta sẽ chứng minh nếu $V_n \neq \varnothing$ với mọi n thì X không có độ trễ giải mã hữu hạn. Điều đó dựa trên lập luận sau.

Nếu $V_n \neq \varnothing$ thì X không thể có độ trễ giải mã $d < n$. Thật vậy, nếu ngược lại thì $V_d = \varnothing$ vì theo định nghĩa ta có không tồn tại phân tích

$$w = x_0x_1...x_du = y_0y_1...y_m, \quad \text{với } x_i \in X,\ y_j \in X,\ u \in A^+,\ x_0 \neq y_0.$$

Từ cách xác định U_i, V_{i+1} ta có U_d là tập rỗng và suy ra V_{d+1} cũng là tập rỗng.
Tiếp tục từ cách xác định U_i, V_{i+1}, ta có U_{d+1} là tập rộng và suy ra V_{d+2},... dẫn đến V_n là tập rỗng, mâu thuẫn. Theo lập luận trên ta có điều cần chứng minh. ■

Mệnh đề trên là cơ sở để xây dựng thuật toán kiểm tra độ trễ hữu hạn, được mô tả như sau.

Đầu vào: tập X.

Đầu ra: độ trễ giải mã của X.

B0. $Y = X; V_0 = Y$

B1. $U_0 = Y^{-1}X - \{\varepsilon\}$
$U_0 = U_0 + Y^-U_0$
$V_1 = U_0^{-1}Y + X^{-1}V_0$

B2. Biết U_i, V_{i+1} xác định U_{i+1}, V_{i+2} nhưsau
$U_{i+1} = V_{i+1}^{-1}X$
$U_{i+1} = U_{i+1} + Y^-U_{i+1}$
$V_{i+2} = U_{i+1}^{-1}Y + X^{-1}V_{i+1}$

B3. Nếu $U_k = U_n$ hoặc $V_k = V_n$ khác rỗng thì X là mã có độ trễ giải mã vô hạn.
Kết thúc

B4 Nếu $V_{d+1} = \varnothing$ thì X có độ trễ giải mã hữu hạn d. Kết thúc.

Thuật toán đã được cài đặt trên máy tính để tính toán trên các mã hữu hạn.

Ví dụ 1

Giả sử $X = \{ca_1, c, a_1b_1, b_1a_2, a_2b_2, b_2a_3, a_3b_3\}$. Theo thuật toán ta có

$U_0 = \{a_1\}$ $V_1 = \{b_1, a_1\}$
$U_1 = \{a_2, b_1\}$ $V_2 = \{b_2, a_2\}$
$U_2 = \{a_3, b_2\}$ $V_3 = \{b_3, a_3\}$
$U_3 = \{b_3\}$ $V_4 = \varnothing$

X có độ trễ giải mã $d = 3$.

1.2 Tính nhập nhằng

Với độ trễ vô hạn, từ vô hạn có thể phân tích không duy nhất - nghĩa là nhập nhằng. Tuy nhiên tính chất đó không phải đúng cho mọi từ. Nếu lực lượng từ có phân tích nhập nhằng thấp thì nói chung tính nhập nhằng không có ý nghĩa áp dụng. Như thế chúng ta cần nghiên cứu những lớp mã mà phần lớn (tất cả) các từ phân tích trên mã này là nhập nhằng.

Ở đây khái niệm phần lớn chỉ mang ý nghĩa định tính, chúng ta cần có các độ đo để có thể đánh giá "mật độ" các từ có phân tích nhập nhằng. Chúng ta cũng cần có cách tính độ đo đó một cách hiệu quả.

Một điều nữa cần tính đến là trong thực tế, nói chung chúng ta làm việc với các từ độ dài lớn chứ không phải vô hạn. Do đó các phương pháp xác định cần áp dụng được trên lớp ngôn ngữ các từ hữu hạn. Từ đó dẫn đến định nghĩa về tính nhập nhằng mới của từ w hữu hạn theo mã X (gọi là ω - nhập nhằng). Tính nhập nhằng này liên hệ chặt chẽ với những mã không là $Z-$ mã trong

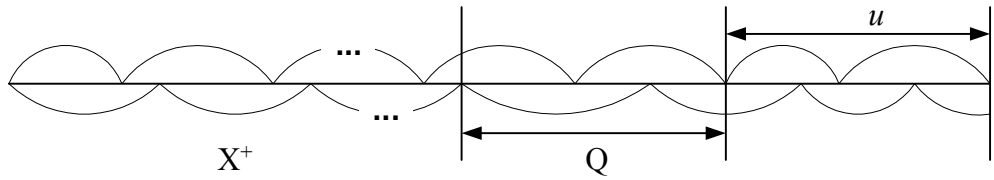
lĩnh vực từ vô hạn ([2]). Như thế, trước hết ta sẽ xác định độ đo μ cho ngôn ngữ sao cho độ đo này áp dụng được cho lớp mã chính quy. Từ đó cho phép định nghĩa độ đo nhập nhằng của mã. Trước hết ta sẽ nghiên cứu đặc trưng của tập các từ nhập nhằng.

2. CÁC TẬP NHẬP NHẰNG LIÊN QUAN MÃ

Định nghĩa 2.1. Với X là mã, từ w được gọi là có khai triển nhập nhằng ($\omega-$ nhập nhằng) trên X nếu:

(i) w có hai dạng phân tích khác nhau, một là phân tích trong X^+ , một là khúc đầu của X^+, nghĩa là có sự nhập nhằng,

(ii) tồn tại từ $u \in X^+$ sao cho wu lại có hai kiểu phân tích như trên.



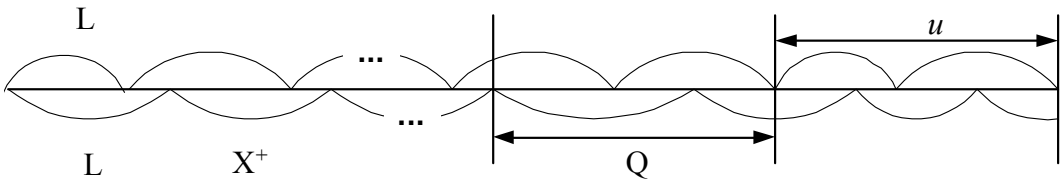
Tập những từ w như thế được gọi là tập nhập nhằng trên mã X, ký hiệu là AMB(X).

Định nghĩa 2.2. Với X là mã và L là một ngôn ngữ trên bảng chữ hữu hạn A. Từ $w \in A^*$ được gọi là có khai triển $L-$ nhập nhằng ($\omega-$ nhập nhằng) trên X nếu:

(i) w có hai dạng phân tích khác nhau, một là phân tích trong LX^+ , một là khúc đầu của LX^+, nghĩa là có sự nhập nhằng.

(ii) tồn tại từ $u \in X^+$ sao cho wu lại có hai kiểu phân tích như trên.

Tập những từ w như thế được gọi là tập $L-$ nhập nhằng trên lã X, ký hiệu là $AMB_L(X)$.



Chú ý rằng hiển nhiên ta có $AMB(X) = AMB_X(X)$

Dựa vào sơ đồ phân tích trên, ta có đặc trưng sau đây về các tập nhập nhằng.

Mệnh đề 2.1. *Cho mã X và ngôn ngữ L của các từ hữu hạn.*

(i) *Tập nhập nhằng $AMB_L(X)$ được xác định bởi công thức sau:*
$AMB_L(X) = LX^*Q \cap LX^*$, *trong đó* $Q = P \cap X^+R^{-1}$ *với*
$P = (Pref(X^+) \cap Suff(X^+) - X^+); R = (X^\omega)^{-1}X^\omega - X^+$

(ii) *Nếu X, L là chính quy thì $AMB_L(X)$ là chính quy.*

Để chứng minh , sử dụng các phân tích tổ hợp trên từ và định nghĩa không khó khăn ta có thể kiểm tra tính chất đặc trưng (i). Còn (ii) suy ra trực tiếp từ các bổ đề sau (suy ra như hệ quả trực tiếp từ những tính thoả trên ngôn ngữ chính qui từ hữu hạn và từ vô hạn như đã xét trong [2, 6]).

Bổ đề 2.1. *Cho A là bảng chữ hữu hạn và $L \subseteq A^\infty$. Khi đó L là ngôn ngữ chính qui khi và chỉ khi có một đồng cấu vị nhóm $h : A^* \to M$ từ vị nhóm tự do A^* lên vị nhóm hữu hạn M sao cho h thoả L.*

Bổ đề 2.2. *Cho A là bảng chữ hữu hạn và $X, Y \subseteq A^\infty$. Nếu đồng cấu vị nhóm $h : A^* \to M$ từ vị nhóm tự do A^* lên vị nhóm hữu hạn M thoả X, Y thì h cũng thoả các tập $pref(X), suff(X)$,*

$X \cap Y,\ X \cup Y,\ X - Y$ và các tập $Y^{-1}X,\ XY^{-1}$ miễn là chúng xác định hợp lý theo định nghĩa các tập thương.

Ví dụ 2. Xét các mã có độ trễ vô hạn sau

$$L1 = \{c,\ ca,\ bcb,\ ab,\ ba\}$$

$$L2 = \{c,\ ca,\ cbb,\ ab,\ ba,\ b^2ab^2,\ b^3ab^3\}$$

Chẳng hạn: $c(ab)^\omega = ca(ba)^\omega$ là những khai triển nhập nhằng vô hạn. Ta có thể xét $w = c(ab)^k a = ca(ba)^k,\ k > 0$ thể hiện từ hữu hạn w là $\omega-$ nhập nhằng trên $L1$. Nếu lấy $L = \{c, ca\}$ thì các khai triển trên cũng là thể hiện của tính $L-$ nhập nhằng trong sự khai triển trên mã $L1$. Việc chọn ngữ cảnh Lđể tăng sự nhập nhằng của các khai triển trên mã X là điều lý thú trong các nghiên cứu lý thuyết và ứng dụng.

3. ĐỘ ĐO CỦA NGÔN NGỮ

Trong mục này, chúng ta sẽ xây dựng độ đo μ cho ngôn ngữ trên bảng chữ cái hữu hạn.

3.1 Yêu cầu của độ đo

Ta xây dựng độ đo cho ngôn ngữ $L \subseteq A^*$
Độ đo μ sẽ xây dựng cần có các tính chất:

(A1) Độ đo là hữu hạn.

(A2) Ngôn ngữ có nhiều từ hơn thì có độ đo μ lớn hơn.

(A3) Cho X, Y là 2 ngôn ngữ sao cho tích $X.Y$ không nhập nhằng ([2]), khi đó:

$$\mu(X, Y) = \mu(X).\mu(Y)$$

Đặc biệt nếu X là mã thì $\mu(X^+) = \sum_{i=1}^{\infty}\mu(X^i) = \sum_{i=1}^{\infty}\mu(X)^i = \mu(X)/(1 - \mu(X))$

(A4) Nếu $X \cap Y = \varnothing$ thì $\mu(X \cup Y) = \mu(X) + \mu(Y)$.

3.2. Định nghĩa độ đo

Đến đây ta xây dựng một độ đo cho ngôn ngữ, từ đó có thể xác định độ nhập nhằng của mã.

Định nghĩa 3.1. Ta định nghĩa độ đo của ngôn ngữ trên $A = \{a, b\}$ như sau:

(i) $\mu(\varepsilon) = 1$

(ii) Giả sử $A = \{a,\ b\}$, khi đó $\mu(a) = \mu(b) = 1/4$
(nếu A có n chữ cái, $\mu(a) = 1/(2n)\ \forall a \in A$, tổng quát hơn $\mu(A) = 1/2$ và $\mu(a)$ có thể khác nhau).

(iii) $\mu(a_1a_2\ ...\ a_n) = \mu(a_1).\mu(a_2)\ ...\ \mu(a_n)$.

(iv) Với $L \subseteq A^+;\ \mu(L) = \Pi_{w\in L}\mu(w)$.

Ví dụ 3. $\mu(A) = \mu(a) + \mu(b) = 1/4 + 1/4 = 1/2$. Với $w = a_1a_2...a_n \in A^+$ ta có:

$$\mu(w) = 1/4\ .\ 1/4\ ...\ 1/4 = \frac{1}{4^n},\ \text{mà}\ \ |A^n| = 2^n \text{ do đó :}$$

$$\mu(A^n) = \sum_{w\in A^n}\mu(w) = 2^n\frac{1}{4^n} = \frac{1}{2^n}.$$

Cuối cùng ta chứng minh được $\mu(A^+)$:

$$\mu(A^+) = \sum_{n=1}^{\infty}\mu(A^n) = \sum_{n=1}^{\infty}\frac{1}{2^n} = 1,$$

$$\mu(A^*) = \mu(A^+) + \mu(\varepsilon) = 1 + 1 = 2,$$

$$\mu(a^+) = \sum_{n-1}^{\infty} \mu(a^n) = \sum_{n=1}^{\infty} \frac{1}{4^n} = \frac{1}{4}\sum_{n=0}^{\infty} \frac{1}{4^n} = \frac{1}{4}\frac{1}{(1-1/4)} = 1/3,$$

$$\text{và } \forall\ L \subseteq A^+,\ \mu(L) \le \mu(A^+) = 1;\ \forall\ L \subseteq A^*,\ \mu(L) \le \mu(A^*) = 2.$$

Mệnh đề 3.1. *Độ đo μ thoả các tính chất $(A1)$ đến $(A4)$ trong mục 3.1.*

Chứng minh. Ta có thể dễ dàng chứng minh các tính chất $(A1), (A2), (A4)$ theo định nghĩa μ.

Chứng minh $(A3)$:

Cho 2 ngôn ngữ A và B: $A.B$ là không nhập nhằng. Khi đó:

$$\mu(A.B) = \sum_{u\in A,\ v\in B} \mu(u.v) = \sum_{u\in A,\ v\in B} \mu(u).\mu(v) = \sum_{u\in A} \mu(u) \sum_{v\in B} \mu(v) = \mu(A)\mu(B).$$

Với $X \subset A^+$, X là mã thì

$$\mu(X^+) = \mu(X \cup X^2 \cup ...) = \sum_{i=1}^{\infty} \mu(X^i).$$

■

Sau đây là kết quả chính của bài này

Định lý 3.1. *Cho bảng chữ hữu hạn A. Cho $L \subseteq A^+$, khi đó nếu L là ngôn ngữ chính quy thì tồn tại thuật toán hiệu quả để tính $\mu(L)$.*

Chứng minh. Chúng ta sẽ chứng minh quy nạp dựa trên automat đơn định đoán nhận L.

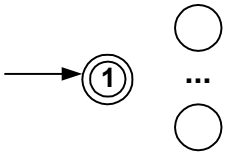
Giả sử bảng chữ cái $A = \{a, b\}$, ta sẽ thấy rằng giả sử này không hạn chế đến tính đúng đắn cho trường hợp tổng quát. Do L là chính quy nên tồn tại automat đơn định $\mathcal{A}$ đoán nhận L, $\mathcal{A} = (Q, E, A, i, T)$ trong đó i là trạng thái khởi đầu, T là tập các trạng thái kết thúc, Q là tập trạng thái, A là bảng chữ hữu hạn, $E \subseteq Q \times A \times Q$ là các tập cạnh.

Chúng ta thể hiện automat $\mathcal{A}$ qua cặp (n, m), trong đó n là số trạng thái của $\mathcal{A}$, m là số cung của $\mathcal{A}$ (từ đây ta ký hiệu $\mathcal{A}(n,\ m)$ cho tiện sử dụng)

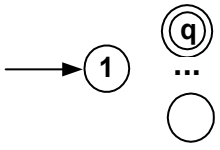
Trên cặp (n, m) xác định quan hệ thứ tự toàn phần:

$$(n, m) < (n', m') \Leftrightarrow \text{hoặc } n < n'$$
$$\text{hoặc } n = n' \text{ và } m < m'$$

Ta tiến hành quy nạp theo thứ tự trên. Với $(n,\ 0)$: Automat $\mathcal{A}$ có thể có 2 dạng sau:



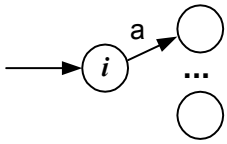
a) Automat này đoán nhận $L = \{\varepsilon\}$, do đó $\mu(L) = 1$.



b) Automat này đoán nhận $L = \{\varnothing\}$, do đó $\mu(L) = 0$.

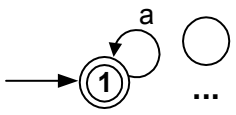
Với $(n,\ 1)$, automat $\mathcal{A}$ có dạng sau:

a) $q \in T, i \notin T$



Automat $\mathcal{A}$ có 1 cung nhãn a hoặc b đoán nhận $L = \{a\}$ hoặc $L = \{b\}$, theo Ví dụ 3 ta có $\mu(L) = 1/4$.

b) Nếu $i \notin T$

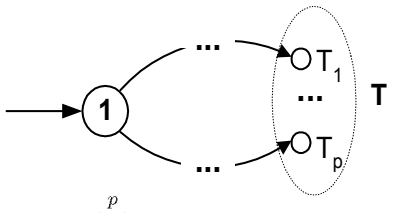


Automat $\mathcal{A}$ có 1 cung lặp nhãn a hoặc b, đoán nhận $L = \{a\}^+$ hoặc $L = \{b\}^+$. Dễ dàng tính theo ví dụ 3: $\mu(L) = 1/3$.

Nếu $i \notin T$: automat $\mathcal{A}$ đoán nhận $L = \varnothing$, do đó $\mu(L) = 0$.

Còn lại, giả sử với mọi cặp $(n', m') < (n.m)$ mỗi automat $\mathcal{A}'(n', m')$ đoán nhận L' đều tính được $\mu(L')$ một cách hiệu quả. Giả sử $\mathcal{A}(n,\ m)$ đoán nhận L, ta sẽ chứng tỏ có thể tính được $\mu(L)$ một cách hiệu quả. Chỉ cần xét trường hợp tập L không chứa $\varepsilon, L \neq A^+,\ L \neq \varnothing$. Thật vậy, do L là chính quy nên tồn tại automat $\mathcal{A}(I, Q, T)$ đơn định đoán nhận L, có 1 trạng thái đầu i. Vì $L \subseteq X^+$ không chứa $\varepsilon,\ L \neq \varnothing$ nên $i \notin T$ và $T \neq Q$. Giả sử $T = \{T_1, .., T_p :\ p \geq 1\}$

Chia $\mathcal{A}$ thành các automat con $\mathcal{A}_k$: chứa các đường đi từ i đến T_k, mỗi automat $\mathcal{A}_k$ đoán nhận tập $L_k = \{w :$ đi từ i đến $T_k\}$



Dễ thấy

$$L = \bigcup_{k=1}^{p} L_k \text{ và } L_k \cap L_{k'} = \varnothing \tag{3.1}$$

Vì hợp rời nhau, ta chỉ cần tính $\mu(L_k)$.

Đặt $P(i, j)$= {tập các con đường đi từ trạng thái i đến trạng thái j },

$P_0(i, j) =$ { tập các con đường đi từ trạng thái i đến trạng thái j lần đầu tiên},

$WP(i, j) =$ { tập các từ là nhãn đường đi $p \in P(i,\ j)$},

$WP_0(i, j) =$ { tập các từ là nhãn đường đi $p \in P_0(i,\ j)$},

$P(i, j, -q) =$ { tập các con đường đi từ trạng thái i đến trạng thái j không qua q }.

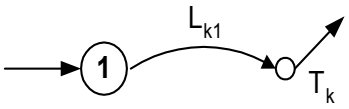
•*Xét trường hợp từ T_k có cung đi ra*

Ta có thể phân tích L_k thành các tập:

a) Tập các nhãn đường đi từ i đến T_k đúng 1 lần, ký hiệu L_{k1}

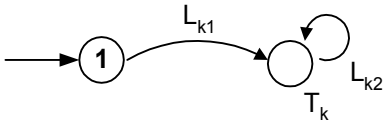
$$L_{k1} = WP_0(i,\ T_k)$$

Do automat A đơn định do đó L_{k1} là prefix và automat đoán nhận L_{k1} có số trạng thái $\leq n$ và số cạnh $< m$, do đó tính được $\mu(L_{k1})$ một cách hữu hiệu theo giả thiết quy nạp.



(3.2)

b) Tập các nhãn đường đi đến T_k và lặp tại T_k lần đầu nhưng không qua i, ký hiệu L_{k2}. $L_{k2} = WP_0(T_k,\ T_k,\ -i)$, rõ ràng L_{k2} là mã. Vì automat đoán nhận L_{k2} có số trạng thái $< n$ nên theo giả thiết, ta tính được $\mu(L_{k2})$ một cách hữu hiệu.



(3.3)

c) Tập các nhãn con đường từ T_k đến T_k lần đầu, đi qua i, ký hiệu L_{k3}

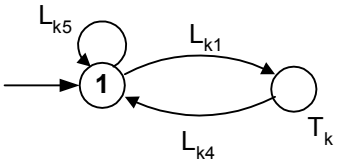
Ta có thể phân tích L_{k3} thành các tập con:

$$L_{k4} = WP_0(T_k,\ i,\ -T_k),$$
$$L_{k5} = WP_0(i,\ i,\ -T_k),$$
$$L_{k1} = WP_0(i,\ T_k),$$

Dễ thấy rằng $L_{k3} = L_{k4}.L_{k1} + L_{k4}.L_{k5}^{+}.L_{k1}$.

Các tập L_{k4}, L_{k5}, L_{k1} đoán nhận bởi các automat con của $\mathcal{A}$, có số trạng thái $< n$ nên $\mu(L_{k4})$, $\mu(L_{k5})$, $\mu(L_{k5}^{+})$ tính được một cách hữu hiệu.

Từ đó $\mu(L_{k3}) = \mu(L_{k4}).\mu(L_{k1}) + \mu(L_{k4}).\mu(L_{k4}^{+}).\mu(L_{k1})$ tính được một cách hữu hiệu do các tích không nhập nhằng.



(3.4)

Như thế, từ b), c) ta có một con đường đi từ T_k đến T_k lần đầu tiên sẽ đi theo hoặc L_{k2} hoặc L_{k3} , hơn nữa $L_{k2} \cap L_{k3} = \varnothing$. Từ đó tập các nhãn con đường đi từ T_k vào T_k (có thể lặp) sẽ là tập L_{k6}:

$$L_{k6} = (L_{k2} + L_{k3})^{+} \tag{3.5}$$

Dễ dàng thấy rằng $(L_{k2} + L_{k3})$ là mã vì không tồn tại từ (là nhãn đường lặp từ T_k vào T_k) có 2 phân tích khác nhau trong $(L_{k2} + L_{k3})$, do đó tính được một cách hữu hiệu

$$\mu(L_{k6}) = \sum_{i=1}^{\infty} (\mu(L_{k2}) + \mu(L_{k3}))^i \tag{3.6}$$

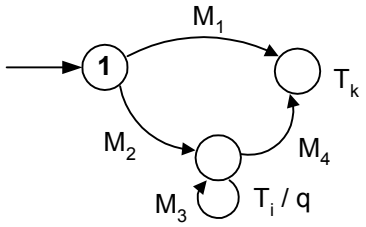
Từ kết quả (3.2) và (3.6) rút ra trong trường hợp từ T_k có cung đi ra, tính được μ cho tập $L_k = \{\ w$: nhãn con đường từ i đến T_k (có thể lặp) $\}$.

• *Trường hợp từ T_k không có cung đi ra:*

Nếu $p > 1$, khi đó tồn tại trạng thái kết thúc $T_j \neq T_k$. Còn khi $p = 1$, do $T \neq Q, i \notin T$ nên tồn tại ít nhất 1 trạng thái $q:\ q \neq i$ và $q \neq T_k$. Như thế trong cả hai trường hợp, luôn tồn tại trạng thái $q:\ q \neq i$ và $q \neq T_k$.

Từ đó xét các tập

$$M_1 = WP(i,\ j,\ -T_i),$$
$$M_2 = WP_0(i,\ T_i),$$
$$M_3 = WP_0(T_i,\ T_i),$$
$$M_4 = WP_0(T_i,\ T_k).$$



Dễ thấy

(i) M_1, M_2, M_4 là tập prefix, đoán nhận bởi automat có số trạng thái $< n$ do đó theo giả thiết tính được $\mu(M_1)$, $\mu(M_2)$ và $\mu(M_4)$ một cách hữu hiệu. (3.7)

(ii) M_3 là mã do mỗi từ trong M_3^{+} có phân tích duy nhất trong M_3, M_3 đoán nhận bởi automat có số trạng thái $< n$ do đó theo giả thiết tính được $\mu(M_3^{+})(= \mu(M_3)/(1 - \mu(M_3))$ một cách hữu hiệu. (3.8)

Mặt khác $L_k = M_1 + M_2.M_4 + M_2.M_3^{+}.M_4$ và các tích là không nhập nhằng và hợp là rời nhau nên

$$\mu(L_k) = \mu(M_1) + \mu(M_2).\mu(M_4) + \mu(M_2).\mu(M_3^{+}).\mu(M_4)$$

Từ (3.7) và (3.8) ta có $\mu(L_k)$ tính được một cách hữu hiệu.
Định lý được chứng minh. ■

3.3. Độ nhập nhằng liên quan đến mã

Dựa trên độ đo cho ngôn ngữ ta đưa vào định nghĩa độ nhập nhằng của mã như sau:

Định nghĩa 3.2. Cho bảng chữ hữu hạn A. Cho $X, Z \subseteq A^+$. Khi đó ta định nghĩa độ $Z-$ nhập nhằng của X bởi

$$\alpha z(X) = \frac{\mu(AMB_z(X))}{\mu(ZX^*)}$$

Trường hợp $Z = X$, để đơn giản ta gọi độ $X-$nhập nhằng là độ nhập nhằng của X và kí hiệu là $\alpha(X)$.

Ví dụ 4. Trên bảng chữ $A = \{a, b, c\}$, xét $L1 = \{c, ca, bcb, ab, ba\}$. Có thể kiểm tra thấy $L1$ là mã. Ta tính độ đo nhập nhằng của mã $L1$

Dễ dàng tính được $\mu(A^+) = 1$; $\mu(L1) = 1/6 + 1/36 \times 3 + 1/216 = 55/216$;

$$\mu(L1^+) = \sum_{i=1}^{\infty}(55/216)^i = 55/(215 - 55) = 55/161$$

do $P = (Pref(L1^+) \cap Suff(L1^+) - L1^+) = \{a,\ b\} = Q$ nên $\mu(P) = 1/3$.

Tích $(L1^+P)$ không nhập nhằng nên

$$\mu(AMB(L1)) = \mu(L1^+).\mu(P) = (55/161)(1/3) = 55/(161 \times 3).$$

Từ đó $\alpha(L1) = 1/3$.

4. KẾT LUẬN

Chúng ta có thể nghiên cứu phát triển theo các chủ đề sau.

1. Nghiên cứu về lý thuyết mã theo độ nhập nhằng.
2. Nghiên cứu các độ đo mới.
3. Ứng dụng độ đo trong các nghiên cứu tổ hợp khác nhau của ngôn ngữ và mã.
4. Nghiên cứu ứng dụng về độ $L-$ nhập nhằng cho các mã X để tăng độ mật của ứng dụng, ta có thể xây dựng các sơ đồ mã hoá và xác thực chữ ký mới... Ví dụ $L1$ cho ta sơ đồ mã với a, b, c xem như các khối con khoá luân phiên trên các bảng chữ a, b, c, còn L là khối c.

Lời cảm ơn

Chúng tôi xin chân thành cảm ơn tiến sĩ Nguyễn Hương Lâm về gợi ý làm cho phép chứng minh Định lý 3.4 được gọn gàng nhờ việc đưa vào quan hệ thứ tự các cặp làm cơ sở cho phép chứng minh quy nạp. Chúng tôi cũng xin bày tỏ lời cám ơn tới người phản biện đã đóng góp ý kiến cho việc trình bày để bài báo được nâng cao chất lượng.

TÀI LIỆU THAM KHẢO

[1] A. Arnold, Deterministic and non-ambiguous rational w-languages, *Lecture Notes in Computer Science* **192** (1985), 138–146.

[2] Do Long Van, Nguyen Huong Lam and Phan Trung Huy, On Code Concerning Bi-Infinitive Words, *Acta Cybernetica*, **11** (1-2) (1993).

[3] J. Berstel and D. Perrin, *Theory of Codes*, Academic Press, NewYork 1985.

[4] Phan Trung Huy, Do Long Van, *On non-ambiguous Buchi V-automata*, Preprints 39/99, Hanoi Institute of Mathematics, 1999. Report. "Workshop in Algebra and Discrete Mathematics", Chinese University of Hongkong 27-31/3/2000.

[5] Phan Trung Huy, On Ambiguities and Unambiguities Related With w-Languages, *Combinatorics and Applications* ", Hanoi Viet nam, 3-5/12/2001.

[6] Phan Trung Huy, Igo Litovsky, Do Long Van, Which finite monoids are syntactic monoids of rational omega-languages, *Information Processing Letters,* **3** (42) (1992) 127-132.

[7] S. Eilenberg, *Automata, Languages and Machines*, Acad. Pres., New York 1974.

Khoa Toán ứng dụng, Trường ĐHBK Hà nội

Nhận bài ngày 15 - 4 - 2002
Nhận lại sau khi sửa ngày 28 - 4 - 2002