Một phương pháp mới thiết lập hệ mật mã dựa trên bài toán tương ứng Post

A Novel Method Based on the Post Correspondance Problem for Designing Cryptosystems

Hồ Ngọc Vinh[1*], Nguyễn Đình Hân[2]

[1] Trường Đại học Sư phạm kỹ thuật Vinh, Đường Nguyễn Viết Xuân, Thành phố Vinh, Nghệ An

[2] Trường Đại học Sư phạm kỹ thuật Vinh, Khoái Châu, Hưng Yên

Đến Tòa soạn: 19-12-2016; chấp nhận đăng: 5-9-2017

Tóm tắt

Mật mã học là một bộ phận của khoa học máy tính. Mật mã học có nhiều ứng dụng và ngày càng có vai trò quan trọng đối với cuộc sống con người. Trong lĩnh vực truyền thông, đóng góp tiêu biểu của mật mã học là nghiên cứu, phát triển các hệ mật mã cho phép con người truyền tin bí mật trong môi trường không an toàn. Tuy nhiên, những tiến bộ gần đây trong kỹ thuật phân tích mã và năng lực tính toán của máy tính đòi hỏi mật mã học phải thiết lập các hệ mật mã mới an toàn hơn. Trong bài báo này, chúng tôi đề xuất một phương pháp thiết lập các hệ mật mã mới dựa trên bẫy cửa sập là một dẫn xuất của bài toán không quyết định được PCP cho lớp ngôn ngữ từ định biên.

Từ khóa: Mật mã học, ngôn ngữ từ định biên, bài toán PCP, bẫy cửa sập.

Abstract

Cryptography is an integral part of computer science. Nowadays, cryptosystems are one of the most important applications of cryptography. In communication area, cryptosystems can be used to establish cryptographic protocols and algorithms for transmitting data securely via an unsafe enviroment such as Internet. Recent advances in cryptoanalysis and computing have exhibited a lot of weeknesses of existing cryptosystems. Therefore, it is most urgent that cryptography establishes new powerful cryptosystems. In this paper, we present a novel method to design cryptosystems as the standard approach to protect data. The obtained cryptosystems contain a trapdoor which can be reduced to an undecidable problem, namely the POST correspondance problem.

Keywords: Cryptography, POST correspondance problem, trapdoor.

1. Mở đầu

Sự phát triển của xã hội dẫn đến việc ngày nay mật mã không những chỉ được dùng trong bí mật quân sự và ngoại giao, mà còn chủ yếu dùng trong bí mật kinh tế, thương mại. Trong các ứng dụng của mật mã học, một nhiệm vụ rất quan trọng của mật mã là đảm bảo sao cho chỉ có những chủ nhân hay những người có thẩm quyền mới có thể truy nhập được thông tin lưu trữ trong một hệ thông tin nào đó. Phương pháp thực hiện là mã hóa dữ liệu và lưu trữ chúng trong hệ dưới dạng đã được mã hóa. Khi đó những người truy nhập trái phép dù có được thông tin ở dạng mã hóa, thì các thông tin này cũng vẫn là vô dụng nếu họ không biết cách giải mã. Do đó nếu giả định rằng các kỹ thuật mã hóa khó phá được dùng thì hệ sẽ là an toàn chống lại được những người sử dụng trái phép. Một hệ mật có thể được coi là *an toàn*, nếu vấn đề thám mã là một *bài toán bất trị* [1]. Vì vậy, khi xem xét một hệ mật, cùng với nó ta phải đưa vào nghiên cứu vấn đề tấn công. Nguyên lý chung trong mật mã học là không có một hệ mật nào tồn tại lâu dài trước sự tấn công, do đó luôn có nhu cầu thiết lập các hệ mật mới. Phương pháp xây dựng các hệ mật nói chung dựa trên nguyên lý *bẫy cửa sập*: thông tin đã công bố không đủ để hiểu biết về một cửa sập bí mật.

Nhắc lại rằng khi đề cập đến một số bài toán là *dễ*, *khó* hay *bất trị*, ta hiểu theo nghĩa độ phức tạp tính toán của chúng. Theo quan điểm mật mã học, bất kỳ bài toán NP-đầy đủ nào cũng sẽ được xem như là bất trị, còn một bài toán là dễ đòi hỏi một giới hạn đa thức bậc thấp [1]. Các hệ mật mã khóa công khai nổi tiếng như RSA và El Gamma được thiết kế dựa trên các bài toán khó đã được nghiên cứu khá kỹ lưỡng.

Trong bài báo này, ta thiết lập một hệ mật dựa trên một dẫn xuất của bài toán không giải được PCP - là bẫy cửa sập ở mức cao nhất trong phân bậc độ khó của các bài toán. Bài toán PCP (Post Correspondence Problem) được đề xuất bởi Emil Post vào năm 1946 [2]. Đây là bài toán thuộc vào lớp bài toán NP-khó, đã và đang được nhiều người quan tâm nghiên cứu. Nó thường được dùng để chứng minh tính không quyết định được của một số bài toán (xem thêm [3-7]). Bài báo được cấu trúc như sau. Trước hết,

[*] Địa chỉ liên hệ: Tel: (+84) 903456411
Email: hongocvinh@gmail.com

trong Phần 2 ta nhắc lại các khái niệm từ định biên trên bảng chữ cái A, $\diamond$-ngôn ngữ, $\diamond$-đồng cấu vị nhóm và thiết lập một số kết quả cơ bản về tính đoán nhận được của $\diamond$-otomat, quan hệ giữa $\diamond$-otomat, $\diamond$-đoán nhận được và $\diamond$-ngôn ngữ chính quy. Trong phần 3, ta xem xét bài toán PCP kinh điển, từ đó đưa ra một dẫn xuất của nó cho lớp ngôn ngữ từ định biên tạo nên một bẫy cửa sập có thể ứng dụng để xây dựng các hệ mật có độ bảo mật rất cao. Phần 4 ta đưa ra kết luận cho bài báo.

2. Từ định biên

Trước hết, ta nhắc lại các ký hiệu và khái niệm được trình bày chi tiết trong [8, 9]. Cho A là bảng hữu hạn các chữ cái. Tập $B = \{0, 1\}$ là tập biên. $A_\diamond = \{(i, a, j) \mid a \in A \cup e, i, j \in B\}$ là tập tất cả các từ mở rộng từ A và $A^*_\diamond = \{(i, w, j) \mid w \in A^*, i, j \in B\} \cup \{g, e\}$. Khi đó, mỗi bộ (i, w, j), $w \in A^*$ được gọi là một $\diamond$-*từ* (*từ định biên với biên i, j*), là mở rộng của từ w với $e, g \notin A^*$ là hai phần tử mới lần lượt đóng vai trò là phần tử zero và phần tử đơn vị của tập các $\diamond$-từ $A^*_\diamond$. Trên $A^*_\diamond$ ta trang bị một phép toán *tích biên* như sau: $\forall x_1 = (i_1, w_1, j_1)$, $x_2 = (i_2, w_2, j_2) \in A^*_\diamond$, nếu $j_1 = i_2$ thì $x_1.x_2 = (i_1, w_1w_2, j_2)$, ngược lại $x_1.x_2 = g$; đặc biệt, $\forall x \in A^*_\diamond$, $x.g = g.x = g$, $x.e = e.x = x$. Dễ thấy $A^*_\diamond$ lập thành một vị nhóm với phép toán tích biên có đơn vị là e.

Ta gọi $A^*_\diamond$ là $\diamond$-*vị nhóm* được định nghĩa bởi A. Một tập $L \subseteq A^*_\diamond$ được gọi là một *ngôn ngữ mở rộng* ($\diamond$-*ngôn ngữ*) trên A. Nếu không sợ hiểu nhầm thì ta cũng sử dụng ký hiệu $|x|$ là độ dài của $\diamond$-từ x, theo nghĩa $|g| = +\infty$, $|e| = 0$ và $|x| = |w|$ nếu $x = (i, w, j)$. Cho $X, Y \subseteq A^*_\diamond$, ta gọi *thương trái* (*thương phải*) *của* X *bởi* Y, ký hiệu $Y^{-1}X$ (t.ư $X.Y^{-1}$) là các tập được xác định bởi: $Y^{-1}X = \{x \in A^*_\diamond \mid \exists y \in Y : y.x \in X\}$ và $XY^{-1} = \{x \in A^*_\diamond \mid \exists y \in Y : x.y \in X\}$. Hàm chiếu *Proj*: $A^*_\diamond \to A^* \cup \{0\}, 0 \notin A^*$, là hàm được xác định bởi: $Proj(e) = e$, $Proj(g) = 0$ và $Proj(i,w,j) = w$.

Định nghĩa 1. *Cho M là một vị nhóm bất kỳ có phần tử đơn vị là 1, phần tử zero là 0 và $\varphi : A^*_\diamond \to M$ là một ánh xạ. Khi đó, φ được gọi là $\diamond$-đồng cấu vị nhóm nếu thỏa mãn các điều kiện sau*:

(*1*) $\forall x, y \in A^*_\diamond$ *mà* $x.y \neq g$ *thì*

$$\varphi(x.y) = \varphi(x).\varphi(y)$$

(*2*) $\varphi(e) = 1$

(*3*) $\varphi(g) = 0$

Định nghĩa 2. *Cho $L \subseteq A^*_\diamond$ và vị nhóm M. Nếu tồn tại một $\diamond$-đồng cấu vị nhóm $\varphi : A^*_\diamond \to M$ sao cho $L = \varphi^{-1}(N)$, với $N \subseteq M$ thì ta nói rằng vị nhóm M thỏa L (hay L thỏa bởi vị nhóm M).*

Từ Định nghĩa 2, nếu $N_1, N_2 \subseteq M$ ta có

$$\varphi^{-1}(N_1 \cap N_2) = \varphi^{-1}(N_1) \cap \varphi^{-1}(N_2),$$

$$\varphi^{-1}(N_1 \cup N_2) = \varphi^{-1}(N_1) \cup \varphi^{-1}(N_2),$$

$$\varphi^{-1}(N_1 \setminus N_2) = \varphi^{-1}(N_1) \setminus \varphi^{-1}(N_2).$$

Hơn nữa, nếu φ là toàn ánh thì:

$$\varphi^{-1}(N_1^{-1}N_2) = \varphi^{-1}(N_1)^{-1}\varphi^{-1}(N_2),$$

$$\varphi^{-1}(N_1 N_2^{-1}) = \varphi^{-1}(N_1)\varphi^{-1}(N_2)^{-1}.$$

Với mỗi $\diamond$-ngôn ngữ $L \subseteq A^*_\diamond$, áp dụng cách thức của S. Eilenberg [10] ta có thể xây dựng vị nhóm M thỏa L trên $A^*_\diamond$. Ký hiệu $R(A,M)$ là tập tất cả các $\diamond$-ngôn ngữ trên $A^*_\diamond$ thỏa bởi M. Theo [9], thì $R(A,M)$ đóng với các phép toán Boolean. Hơn nữa, nếu φ là đẳng cấu thì $R(A,M)$ đóng với phép lấy thương trái và thương phải.

Mối quan hệ giữa $\diamond$-đoán nhận được, $\diamond$-ngôn ngữ chính quy và $\diamond$-otomat được thể hiện qua định lý sau:

Định lý 1. (**[8]**) *Cho $L \subseteq A^*_\diamond$. Khi đó, các điều kiện sau đây là tương đương*:

(*i*) *L là $\diamond$-đoán nhận được bởi một $\diamond$-otomat đa định hữu hạn.*

(*ii*) *L thỏa bởi một vị nhóm hữu hạn.*

(*iii*) *L $\diamond$-đoán nhận được bởi một $\diamond$-otomat đơn định hữu hạn.*

(*iv*) *L là $\diamond$-ngôn ngữ chính quy.*

3. Bài toán PCP và ứng dụng

Đã có rất nhiều tiếp cận tới bài toán PCP (bài toán tương ứng Post) được trình bày trong các công trình khoa học khác nhau. Sau đây chúng tôi trình bày tóm lược bài toán PCP, từ đó đề xuất bài toán PCP theo tiếp cận của từ định biên và $\diamond$-otomat nhằm làm nổi bật tính khả dụng của từ định biên trong các hệ mật mã.

Bài toán PCP: Cho bảng chữ hữu hạn A và các từ $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n \in A^*$, $n \geq 1$. Tìm các dãy chỉ số $i_1, i_2, \ldots, i_l$, $1 \leq i_j \leq n, j = 1,\ldots,l$ sao cho:

$$a_{i_1}a_{i_2}\ldots a_{i_l} = b_{i_1}b_{i_2}\ldots b_{i_l}$$

Bài toán quyết định là có lời giải hay không có lời giải.

Ví dụ 1. Cho $A = \{a, b\}$ và hai dãy $a = (a, ab, bba)$, $b = (baa, aa, bb)$. Đây là một tình huống của bài toán tương ứng Post. Nếu ta ghép các từ thứ ba, thứ hai, thứ ba và thứ nhất trong hai danh sách đã cho, thì trong cả hai trường hợp ta nhận được từ *bbaabbbaa*.

$$a_3a_2a_3a_1 = bba + ab + bba + a$$

$= bb + aa + bb + baa = \beta_3\beta_2\beta_3\beta_1$

Khi đó dãy chỉ số (3, 2, 3, 1) là một *lời giải* đối với tình huống này của bài toán tương ứng Post.

Hơn nữa, từ dãy chỉ số (3, 2, 3, 1), ta có dãy "*lặp*" (3, 2, 3, 1, 3, 2, 3, 1) cũng là lời giải của bài toán. Nghĩa là, nếu tìm được dãy chỉ số $(i_1, i_2, \ldots, i_l)$ là lời giải của bài toán thì ta cũng sẽ tìm được dãy $(i_1, i_2, \ldots, i_l)^*$ cũng là lời giải của bài toán.

Mặt khác, nếu bài toán trên chỉ có hai dãy dạng α_2, α_3 và β_2, β_3 thì bài toán sẽ không có lời giải, bởi vì sẽ không tìm được cặp có cùng ký tự cuối.

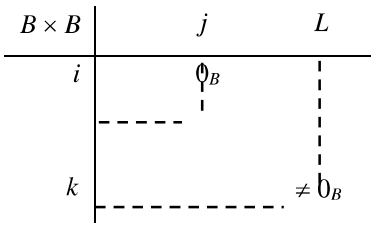
Ví dụ 2. Cho $A = \{a, b, c\}$ và hai dãy: $\alpha = (bb, ab, c)$, $\beta = (b, ba, bc)$. Khi đó các dãy chỉ số (1, 2, ..., 2, 3) và $(1, 2, ..., 2, 3)^*$ là lời giải của bài toán

Về tính quyết định của bài toán trên, ta có kết quả:

Định lý 2. ([1]) *Bài toán tương ứng Post là không quyết định được.*

Trong phần tiếp theo, chúng tôi sẽ trình bày một dạng mở rộng khác của otomat đa định hữu hạn A và nếu không sợ hiểu nhầm thì ta vẫn gọi là ◇-*otomat đa định hữu hạn* $A_\diamond$, với tập biên B không chỉ là {0, 1}, mà tập biên B được mở rộng thành $B \subseteq \mathbb{N} \cup \{0_B\}$, với 0_B là phần tử zero của nửa nhóm B và phép toán trên B được xác định "*bí mật*" theo hệ mật. Tiếp đến, ta sẽ sử dụng ◇-otomat đa định hữu hạn $A_\diamond$ để ứng dụng vào bài toán tương ứng Post nhằm xây dựng hệ mật mới, đảm bảo tính bảo mật cao.

Cho B là một nửa nhóm, ta xây dựng cấu trúc điều khiển B (bảng nhân $B\times B$) như sau: với $i, j, l, k \in B$,

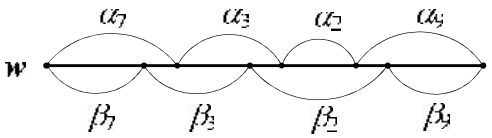


Đặc biệt: $\forall i \in B,\ i.0_B = 0_B.i = 0_B$.

Bảng nhân $B\times B$ phần lớn chứa các phần tử 0_B. Thông thường, để tăng thêm tính bảo mật thì B có số phần tử rất lớn để không thể dễ dàng tìm được cấu trúc của B.

Ví dụ 3. Cho w là từ tuyệt mật, giả sử

$$w = \alpha_7.\alpha_3.\alpha_2.\alpha_9 = \beta_7.\beta_3.\beta_2.\beta_9.$$



Hình 1. Cấu trúc từ tuyệt mật w

Với từ w được cho như trong Hình 1, ta có các điểm cắt và lưu trữ chúng dưới dạng bảng (cấu trúc bảng nhân của nửa nhóm bí mật B) như sau:

$B\times B$	1	2	3	4	5	6	7	8	9	...
1	0_B	0_B	0_B	0_B	0_B	0_B	0_B	0_B	0_B	0_B
2	0_B	0_B	0_B	0_B	0_B	0_B	0_B	0_B	$\neq 0_B$	0_B
3	0_B	$\neq 0_B$	0_B	0_B	0_B	0_B	0_B	0_B	0_B	0_B
4	0_B	0_B	0_B	0_B	0_B	0_B	0_B	0_B	0_B	0_B
5	0_B	0_B	0_B	0_B	0_B	0_B	0_B	0_B	0_B	0_B
6	0_B	0_B	$\neq 0_B$	0_B	0_B	0_B	0_B	0_B	0_B	0_B
7	0_B	0_B	0_B	0_B	0_B	0_B	0_B	0_B	0_B	0_B
8	0_B	0_B	0_B	0_B	0_B	0_B	0_B	0_B	0_B	0_B
9	0_B	0_B	0_B	0_B	0_B	0_B	0_B	0_B	0_B	0_B
. .	0_B	0_B	0_B	0_B	0_B	0_B	0_B	0_B	0_B	0_B

Chọn M là một vị nhóm rất lớn. M có chứa phần tử đơn vị là 1_M, phần tử zero là 0_M. Từ bảng nhân của B, ta xây dựng vị nhóm $M_\diamond$ được điều khiển bởi nửa nhóm B như sau: với $(i_1, m_1, j_1), (i_2, m_2, j_2) \in M^\diamond$. Nếu $j_1.i_2 \neq 0_B$ thì $(i_1, m_1, j_1).(i_2, m_2, j_2) = (i_1, m_1m_2, j_2)$, ngược lại $(i_1, m_1, j_1).(i_2, m_2, j_2) = \theta$.

Ta có $\mathrm{Card}(M_\diamond) = \mathrm{Card}(B^2) . \mathrm{Card}(M)$. Từ đồng cấu vị nhóm $\varphi : A^* \to M$, ta có thể xác định đồng cấu $\varphi_\diamond : A_\diamond^* \to M_\diamond$ như sau:

(1). $\forall a \in A, (i, a, j) \to (i, \varphi(a), j)$,

(2). $(i, e, j) \to (i, 1_M, j)$,

(3) $e \to e$,

(4) $\theta \to \theta$.

Với $M_\diamond$ được xây dựng như trên và một khóa bí mật $K = (i, m^*, j^*)$, ta xây dựng ◇-otomat $A_\diamond(K) = (A_\diamond, Q_\diamond, \delta_\diamond, I_\diamond, T_\diamond)$, với tập các trạng thái $Q_\diamond = M_\diamond$, tập các trạng thái khởi đầu $I_\diamond \subseteq Q_\diamond$, $I_\diamond = (i, 1_M, i)$, $i \in B$, tập các trạng thái kết thúc $T_\diamond = K$ và hàm chuyển $\delta_\diamond$ được xác định bởi: với mỗi trạng thái $(i,m,j) \in M_\diamond$ và ◇-từ $(k,w,l) \in A_\diamond^*$ thì $\delta_\diamond((i, m, j), (k, w, l)) = (i, m.\varphi(w), l)$ nếu $i.k \neq 0_B$, ngược lại $\delta_\diamond((i, m, j), (k, w, l)) = 0_B$.

Ta ký hiệu: nếu $i.k \neq 0_B$ thì

$$(i, m, j) \xrightarrow{(k, w, l)} (i, m.\varphi(w), l)$$

Lấy $K = (i, m^*, j^*) \in M_\diamond$ và phần tử khởi đầu $(i, 1_M, i)$, $i \in B$. Khi đó, ta gọi $w \in A^*$ là một *từ hợp lý* và đường đi có nhãn w là *đường đi thành công* nếu tồn tại dãy:

$$(i_1, a_1, j_1), (i_2, a_2, j_2), \ldots, (i_l, a_l, j_l)$$

và đường đi

$$(i, 1_M, i) \xrightarrow{(i_1,a_1,i_1).(i_2,a_2,i_2)\ldots(i_{l-1},a_l,i_l)} K$$

với $i_1 = i$, $i_l = j^*$, $a_1 a_2 \ldots a_l = w$ và $\varphi(w) = m^*$

Với bảng nhân bí mật của B có cỡ là Card(B^2) như trên. Nếu thực hiện tấn công theo kiểu vét cạn để tìm cấu trúc của B thì thuật toán có độ phức tạp thời gian $O(\text{Card}(B^2)^{\text{Card}(B)})$. Lưu ý rằng trong sơ đồ bảo mật thì hàm φ có thể thay đổi theo thời gian nhằm đảm bảo tính bí mật. Vị nhóm M có thể công khai hoặc có thể là một nhóm phép thế. Với mỗi từ w, có thể có nhiều cách chia các điểm cắt khác nhau. Nói chung người/cơ quan chứng thực (CA) phải biết cách chia này.

Sau đây ta xem xét *một dẫn xuất của bài toán tương ứng Post* cho lớp ngôn ngữ từ định biên.

Cho bảng chữ hữu hạn A, các từ $a_1, a_2, ..., a_n \in A^*$, khóa $K = (i, m^*, j^*)$ và bảng nhân bí mật B. Tìm các dãy chỉ số $i_1, i_2, ..., i_l$, $n \le i_j \le l$, $j = 1,\ldots,l$ sao cho: $w = a_{i_1}.a_{i_2} \ldots a_{i_l}$ và $(i_1, \varphi(w), i_l) = K$, với $\varphi : A^* \to M$ là một đồng cấu vị nhóm.

Nghĩa là, tìm dãy $i_1, i_2, ..., i_l$ sao cho $i_1 = i$, $i_l = j^*$, $\varphi(a_1 a_2 \ldots a_l) = m^*$.

Tiếp theo, ta biểu diễn bài toán trên máy tính và đánh giá độ phức tạp của thuật toán tìm lời giải cho bài toán này.

Từ dãy $w = a_{i_1}, a_{i_2}, ..., a_{i_l}$, ta mở rộng thành dãy $(i_1, a_{i_1}, i_1), (i_2, a_{i_2}, i_2), \ldots, (i_l, a_{i_l}, i_l)$ và lưu trữ trong bộ nhớ dưới dạng $\{(i_1, a_{i_1}), (i_2, a_{i_2}), \ldots, (i_l, a_{i_l})\}$. Đối với khóa $K = (i, m^*, j^*)$, ta lưu trữ dãy chỉ số khóa $(i_1.i_2, i_2.i_3, \ldots, i_{l-1}.i_l)$, với $i_1 = i$, $i_l = j^*$, $\varphi(w) = m^*$. Trong sơ đồ bảo mật, ta cần giữ bí mật chuỗi w và dãy chỉ số $i_1, i_2, ..., i_l$. Hàm φ có thể công khai hoặc giữ bí mật.

Sau đây, ta đánh giá độ phức tạp của thuật toán vét cạn tấn công vào sơ đồ bảo mật để tìm một lời giải hợp lý.

Đặt $p = |w|$. Khi đó, kích thước của nửa nhóm B là tập tất cả các khúc con của $\{ w + 0_B + (\neq 0_B) \}$, nghĩa là

$$\text{Card}(B) = p + (p\text{-}1) + (p\text{-}2) + ... + 1 + 1 + 1$$

$$= \frac{(p+1).p}{2} + 2$$

Số phần tử khác 0_B trong nửa nhóm B là các phần tử có độ dài 2 trở đi, do đó:

$$\text{Card}(B)_{\neq 0_B} = (p\text{-}1) + (p\text{-}2) + \ldots + 1$$

$$= \frac{p.(p-1)}{2}$$

Vì vậy, bảng nhân của B có số phần tử là:

$$Card(B \times B) = \left(\frac{(p+1).p}{2} + 2\right)^2$$

Tỷ số phần tử khác 0_B và tổng số phần tử của bảng nhân B là:

$$\frac{\frac{p.(p-1)}{2}}{\left(\frac{(p+1).p}{2} + 2\right)^2} \approx \frac{p.(p-1)}{\frac{(p+1)^2.p^2}{2}}$$

$$\approx \frac{2}{p^2}$$

Dãy hợp lý cần tìm là: $i = i_1, i_2, \ldots, i_{l-1}, i_l = j^*$ và $i_1.i_2 \neq 0_B$, $i_2.i_3 \neq 0_B, \ldots, i_{l-1}.i_l \neq 0_B$.

Khi đó thuật toán vét cạn để tìm dãy hợp lý trên có độ phức tạp thời gian cỡ:

$$Card(B^{l-2}) = \left(\frac{(p+1).p}{2}\right)^{l-2}$$

Ta còn xem xét độ phức tạp thời gian của thuật toán tấn công vào sơ đồ bảo mật theo phương pháp tấn công xác suất như sau:

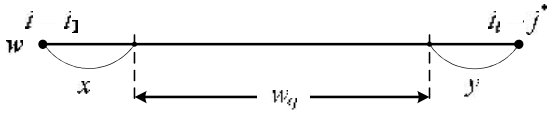
$$Card(B^{l-2}) = \left(\frac{(p+1).p}{2}\right)^{l-2}$$

Đặt $q = |w_q|$. Khi đó $l\text{-}1 \le q \le p\text{-}2$. Ta có $x w_q y = w$, với $x, y \neq e$ (Hình 2). Vậy có C_{q-1}^{l-2} cách chia các dãy khúc con để lấy ra dãy $i_2, i_3, ..., i_{l-1}$.

+ Với $|x| = 1$, $|y| = 1$ và q cố định, thì số cách để lấy w_q là:

$$p - 1 - q = p - 1 - (q - 2) = 1$$

+ Với q tùy ý, thì số cách để lấy w_q là: C_{q-1}^{l-2}



Hình 2. Chi tiết cấu trúc từ tuyệt mật w

Vậy, xác suất thắng lợi P để tìm dãy $i = i_1, i_2, ..., i_{l-1}, i_l = j^*$ là:

$$P \le \frac{C_{q-1}^{l-2}}{\left(\frac{(p+1).p}{2}\right)^{l-2}} = \frac{2^{l-2}.C_{q-1}^{l-2}}{(p+1)^{l-2}.p^{(l-2)}}$$

$$\approx \frac{2^{l-2}.C_{q-1}^{l-2}}{p^{2.(l-2)}} \approx \log_2\left(\frac{2^{l-2}.C_{q-1}^{l-2}}{p^{2.(l-2)}}\right)$$

$$\approx \log_2 C_{q-1}^{l-2} + 2.(l-2).\log_2 2 - 4.(l-2)\log_2 p$$

$$\approx \log_2\left(\frac{(q-1).(q-2)\ldots(q-l+2)}{1.2.3\ldots(l-2)}\right) + 2.(l-2).\log_2 2 - 4.(l-2)\log_2 p$$

$$\approx \log_2(q-1)+\log_2(q-2)+\ldots+\log_2(q-l+2)-\log_2 1-\log_2 2-\ldots-\log_2(l-2)+2.(l-2).\log_2 2-4.(l-2)\log_2 p$$

Với các giá trị l, p, q cụ thể, ta có P tương ứng cho trong Bảng 1.

Bảng 1. Bảng kê xác suất tìm được nghiệm của bài toán

TT	l	p	q	Tỷ lệ (xấp xỉ)
1	8	80	16	$1 / 2^{127}$
2	10	128	32	$1 / 2^{185}$
3	10	128	16	$1 / 2^{195}$
4	12	320	52	$1 / 2^{279}$
5	12	320	32	$1 / 2^{287}$
6	12	512	100	$1 / 2^{296}$
7	16	512	32	$1 / 2^{314}$
8	16	1024	32	$1 / 2^{504}$
9	16	2048	32	$1 / 2^{560}$
...	...	...	...	...

Nhận xét 1. Từ bảng đánh giá kết quả ở trên, ta thấy rằng khả năng tấn công thành công sơ đồ bảo mật bằng phương pháp vét cạn là rất nhỏ. Sơ đồ bảo mật trên có thể được sử dụng để giải quyết một số bài toán thực tiễn như xác thực người dùng.

Ví dụ 4. Cho các dữ kiện:

Mỗi người dùng được hệ thống xác thực *CA* cung cấp một dãy từ mã $\{a_1, a_2, \ldots, a_n\}$, $a_i \in A^*$. Hệ thống *CA* lưu trữ bộ thông tin:

$$\{ K_{CA} = (i, w, j^*) \in M\diamond ;\ B\times B ;\ K ; f \},$$

trong đó, K_{CA} là từ khóa bí mật của *CA*; $B\times B$ là bảng nhân bí mật; $K = \{K_C, K_D, K_E, \ldots\}$ là tập khóa của các người dùng và $f : A^*\times K_C \to A^*$ là một đồng cấu mã.

Bài toán xác thực người dùng được phát biểu như sau: *Cho biết người dùng C bất kỳ có phải là người dùng hợp pháp hay không?*

Để trả lời câu hỏi trên, ứng dụng sơ đồ bảo mật đã xây dựng, ta đề xuất sơ đồ xác thực sau:

Sơ đồ 1. Xác thực người dùng.

B_1. Hệ thống yêu cầu người dùng *C* cung cấp dãy từ mã $\{a_1, a_2, \ldots, a_n\}$.

B_2. Tìm dãy từ mã $\{a_1', a_2', \ldots, a_n'\}$, với $a_i' = f(a_i, K_C)$.

B_3. Kiểm tra, nếu $i = i_1, j^* = i_l$ và $a_{i_1}'.a_{i_2}' \ldots a_{i_l}' \in \mathcal{L}(A\diamond(K_{CA}))$ thì kết luận người dùng *C* là hợp pháp, ngược lại kết luận *C* là không hợp pháp.

Trong Sơ đồ 1, đồng cấu bí mật $f : A^*\times K \to A^*$ có thể là phép *XOR* hoặc phép toán đồng dư modulo *P*.

$$\begin{cases} a_i' = a_i \oplus K_C \\ a_i' = a_i . K_C + C \ \ mod\ P, \text{ với } (K_A, P) \text{ là số nguyên tố cùng nhau} \end{cases}$$

Nhận xét 2. Với Sơ đồ 1 xác thực người dùng, nếu kẻ gian *D* muốn tấn công hệ thống thì *D* phải giả dạng một người dùng hợp pháp khác (giả sử người dùng *C*), khi đó *D* giải bài toán tương ứng Post. Tức là, với hai dãy từ $\{b_1, b_2, \ldots, b_n\}$ và $\{g_1, g_2, \ldots, g_n\}$ tương ứng của *D* và *C*, thì *D* phải tìm dãy chỉ số $\{i_1, i_2, \ldots, i_k\}$ sao cho:

$$f(b_{i_1}, K_D).f(b_{i_2}, K_D) \ldots f(b_{i_l}, K_D) = f(g_{i_1}, K_C).f(g_{i_2}, K_C) \ldots f(g_{i_l}, K_C)$$

4. Kỹ thuật bẫy cựa sập

Trong các phần trên ta đã biết bài toán tương ứng Post không có lời giải. Dựa trên bài toán này ta xây dựng hệ mật trên lớp ngôn ngữ từ định biên. Thực tế, các hệ mật mã hiện đại được sử dụng rộng rãi, đặc biệt trong thương mại điện tử, có đặc điểm chung là chúng cùng dựa trên ý tưởng của một bài toán khó nào đó của toán học được biết thuộc lớp NP hoặc NP-Complete. Chẳng hạn hệ mật mã RSA dựa trên bài toán phân tích thừa số và hệ mật mã El Gamal dựa trên bài toán Logarit rời rạc.

Các bài toán trên được xem là không giải được theo nghĩa chỉ có lời giải trong thời gian và năng lực tính toán hữu hạn nào đó. Sau đây ta đề cập đến một bài toán không giải được thực sự. Từ đó mở ra khả năng ứng dụng bài toán này trong phát triển các hệ mật mới. Đối với lớp ngôn ngữ từ định biên chính

quy trên $A^*_\diamond$ mà ta đã xét trong bài báo, bài toán thành viên không có lời giải. Tuy nhiên, nhờ phép chiếu định nghĩa trên $A^*_\diamond$, ta nhận được ngôn ngữ chính quy trên A^* tương ứng với $\diamond$-ngôn ngữ chính quy cho trên $A^*_\diamond$. Từ đó bài toán không giải được trên $A^*_\diamond$, quy về bài toán giải được trên lớp ngôn ngữ chính quy thuộc A^*. Đây là tình huống có thể khai thác để xây dựng các hệ mật mới.

5. Kết luận

Trong bài báo chúng tôi đưa ra một dạng otomat mới được mở rộng từ otomat thông thường. Từ đó, cho ta một cách nhìn mới về biểu diễn thông tin theo mô hình toán học của otomat và các ứng dụng của nó. Ta cũng đề xuất bài toán tương ứng *Post* cho lớp ngôn ngữ từ định biên nhằm tạo ra các bẫy cửa sập cho phép xây dựng các hệ mật có tính bí mật rất cao. Ngoài ra, trong số các kết quả nhận được có nhiều khía cạnh có thể phát triển tiếp cả về mặt lý thuyết và ứng dụng. Chúng tôi sẽ tiếp tục phát triển mở rộng hướng nghiên cứu này trong các công trình sau.

References

[1] A. Salomaa, Nhập môn tin học lý thuyết tính toán và các ôtômat (Bản dịch), NXB Khoa học và Kỹ thuật (1992).

[2] E. L. Post, A variant of a recursively unsolvable problem, Bull. Amer. Math. Soc Vol. 52 (No. 4) (1946) 264-268.

[3] K. Ruohonen, "On some variants of Post's correspondence problem". Acta Informatica (Springer) Vol. 19 (No. 4) (1983) 357-367.

[4] M. R. Garey, D. S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, W. H. Freeman & Co. New York, NY, USA (1979).

[5] Y. Gurevich, Average case completeness, J. Comp. Sys. Sci. (Elsevier Science) Vol. 42 (No. 3) (1991) 346-398.

[6] V. Halava, M. Hirvensalo, R. de Wolf, Marked PCP is decidable, Theoretical Computer Science, Vol. 255 (2001) 193-204.

[7] P. Chambart, Ph. Schnoebelen, Post embedding problem is not primitive recursive, with applications to channel systems, Lecture Notes in Computer Science, Vol. 4855 (2007) 265-276.

[8] Hồ Ngọc Vinh, Phan Trung Huy, Đỗ Long Vân. $\diamond$-ngôn ngữ chính quy và mã, Hội thảo khoa học quốc gia lần thứ IV Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin (FAIR 2009), Hà Nội (2009) 13-22.

[9] H. N. Vinh, P. T. Huy, Codes of Bounded Words, Proceedings of the 3rd International Conference on Computer and Electrical Engineering (ICCEE 2010), Vol. 2 (2010) 89-95.

[10] S. Eilenberg, Automata, languages and machines, Vol. B, Academic Press, New York (1976).