Một tinh chỉnh hiệu quả cho Bộ tạo dãy giả ngẫu nhiên Massey-Rueppel hướng phần cứng

Hoàng Đình Linh, Nguyễn Văn Long

***Tóm tắt*— Các số và các dãy ngẫu nhiên đóng một vai trò quan trọng trong mật mã. Để tạo một nguồn ngẫu nhiên vật lý thường khá tốn kém, do đó hầu hết các hệ thống hiện nay đều sử dụng các bộ sinh số giả ngẫu nhiên. Bộ tạo dãy giả ngẫu nhiên Massey-Rueppel được công bố vào năm 1984 là một trong những bộ tạo số giả ngẫu nhiên sử dụng thanh ghi dịch phản hồi tuyến tính được sử dụng rộng rãi do tính hiệu quả và đáp ứng đầy đủ các tính chất mật mã. Tuy nhiên, khi cấu hình phần cứng thì bộ tạo này chỉ thực sự hiệu quả khi số các hệ số khác 0 trong đa thức đặc trưng của nó là nhỏ. Trong bài báo này, chúng tôi đề xuất một tinh chỉnh nhằm cải thiện hiệu suất thực thi khi cấu hình phần cứng mà không cần quan tâm đến các hệ số của đa thức đặc trưng. Bài báo cũng chứng minh rằng, mô hình bộ tạo sau khi được tinh chỉnh vẫn giữ được các tính chất mật mã tốt như bộ tạo Massey-Rueppel ban đầu, trong khi hiệu suất thực thi phần cứng cải thiện hơn rất nhiều. Bài báo còn trình bày kết quả một số đánh giá thực nghiệm tính ngẫu nhiên đối với các dãy đầu ra. Kết quả thu được cho thấy, các dãy đầu ra đạt được tính giả ngẫu nhiên. Từ đó củng cố thêm cho độ an toàn đã được chứng minh.**

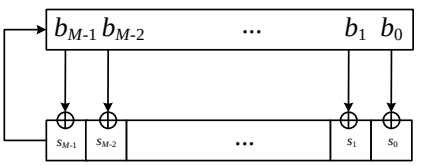
***Abstract*— Random numbers and sequences play an important role in cryptography. Generating a true random source is quite costly, so almost recent systems use pseudorandom number generators. Pseudorandom sequence generators Massey-Rueppel, which was proposed in 1984, is one of the most popular pseudorandom sequence generators using the linear feedback shift registers by its efficiency and sufficiently mathematical proofs. However, this generator is only efficient in hardware implementation when the number of non-zero coefficients in its linear feedback polynomial is small. In this paper, we proposed an improvement for this generator but never mind what coefficients of its linear feedback polynomial are. We have proved that the proposed generator also owns good cryptographic properties as same as the original one, while the efficiency is better in the hardware implementaion. Moreover, we have evaluated some statistical random tests for the output sequences. The results show that output sequences attain pseudorandomness which suitable for theoretical properties.**

***Từ khóa*— bộ tạo dãy giả ngẫu nhiên; thanh ghi dịch phản hồi tuyến tính; Massey-Ruppel; cấu hình phần cứng.**

***Keywords*— pseudorandom sequence generator; linear feedback shift register; Massey-Rueppel; hardware implementation.**

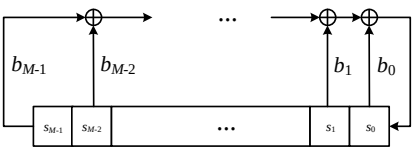
I. GIỚI THIỆU

Thanh ghi dịch phản hồi tuyến tính được ứng dụng nhiều trong các giải pháp sinh chuỗi giả ngẫu nhiên. Điều này do tính chất mật mã của bộ tạo dạng này được chứng minh một cách tường minh và đơn giản trong việc cấu hình phần cứng cũng như phần mềm của các thanh ghi dịch. Bộ sinh dãy giả ngẫu nhiên Massey-Rueppel đã được nghiên cứu và ứng dụng rộng rãi trên thế giới. Tuy nhiên các thanh ghi dịch trong bộ tạo dãy Massey-Rueppel là các thanh ghi dịch tuyến tính phản hồi dạng Fibonacci. Trong cấu hình phần cứng, thanh ghi dịch dạng này chỉ thực sự hiệu quả khi số các hệ số khác 0 trong đa thức đặc trưng của nó là số nhỏ [1].



Galois LFSR

Hình 1a. Thanh ghi dịch phản hồi tuyến tính dạng Galois



Fibonacci LFSR

Hình 1b. Thanh ghi dịch phản hồi tuyến tính dạng Fibonacci

Nhằm cải thiện hiệu suất khi cấu hình phần cứng, chúng tôi đề xuất một tinh chỉnh cho bộ tạo Massey-Rueppel. Cụ thể, chúng tôi sử dụng các thanh ghi dịch dạng Galois, thay vì Fibonacci như trong bộ tạo dãy Massey-Rueppel. Kết quả là bộ

Bài báo được nhận ngày 3/6/2017. Bài báo được nhận xét bởi phản biện thứ nhất vào ngày 17/7/2017 và được chấp nhận đăng vào ngày 1/8/2017. Bài báo được gửi cho phản biện thứ hai vào ngày 24/7/2017 và được chấp nhận đăng vào ngày 1/8/2017.

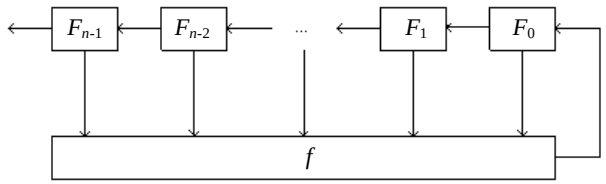
tạo mới vẫn giữ được các tính chất mật mã tốt trong khi hiệu suất phần cứng tăng lên đáng kể.

Bố cục của bài báo gồm: sau Mục Giới thiệu, Mục II trình bày một số kiến thức chuẩn bị về khái niệm tính chất của các thanh ghi dịch phản hồi tuyến tính cũng như các cấu hình của chúng. Mục III trình bày về tinh chỉnh hiệu quả cho bộ tạo dãy giả ngẫu nhiên Massey-Rueppel đối với cấu hình phần cứng. Trong Mục IV, chúng tôi đánh giá một số tính chất mật mã đối với bộ tạo sau khi được tinh chỉnh. Kết quả thu được là dãy đầu ra có các tính chất mật mã tương tự bộ tạo Massey-Rueppel, nhưng gia tăng hiệu suất phần cứng. Một số kết quả thực nghiệm được trình bày trong Mục V. Cuối cùng là mục Kết luận.

II. CƠ SỞ LÝ THUYẾT

A. Thanh ghi dịch phản hồi tuyến tính

Trong thể hiện phần cứng của một máy trạng thái hữu hạn, các flip-flops thường được sử dụng để lưu các trạng thái bên trong. Với *n* flip-flop thì máy có thể thể hiện được tối đa 2^n trạng thái. Đơn giản hơn, chúng ta có thể hình dung *n* flip-flop dưới dạng một thanh ghi dịch phản hồi độ dài *n*.



Hình 2. Một thanh ghi dịch phản hồi

Trong một thanh ghi dịch phản hồi độ dài *n* (Hình 2), gọi các flip-flop lần lượt là $F_0, F_1, \cdots, F_{n-1}$. Tại mỗi thời điểm, F_i nhận giá trị từ F_{i-1} với $i > 0$ và F_0 được cập nhật bởi hàm phản hồi $f: \{0,1\}^n \to \{0,1\}$. Ở đây, chúng ta giả sử rằng giá trị F_{n-1} là đầu ra của thanh ghi dịch. Trong thực tế, có thể lấy giá trị của một flip-flop bất kỳ làm đầu ra cho thanh ghi dịch.

Trong toán học, một dãy $\{a_i\}_{i\in\mathbb{N}}$ được sinh bởi một thanh ghi dịch chỉ là một dãy thoả mãn phép đệ quy *n*-biến

$$a_{i+n} = f(a_i, \cdots, a_{i+n-1}) . \quad (1)$$

Định nghĩa này không chỉ hạn chế cho các dãy nhị phân, mà hầu hết các kết quả sẽ đúng cho các dãy ghi dịch được xác định trên một trường (hữu hạn) bất kỳ, hoặc đôi khi cả với các dãy được xác định trên các vành.

Một thanh ghi dịch phản hồi là tuyến tính nếu hàm phản hồi là tuyến tính. Do đó:

Định nghĩa 1, [1]. *Một dãy ghi dịch phản hồi tuyến tính (LFSR) là một dãy* $\{a_i\}_{i\in\mathbb{N}}$ *thoả mãn phép đệ quy*

$$a_{i+n} = \sum_{j=0}^{n-1} c_j a_{i+j} . \quad (2)$$

Một điều kiện cần cho tính an toàn của một hệ thống đó là các dãy giả ngẫu nhiên được tạo ra phải có chu kỳ lớn. Do đó các dãy có chu kỳ cực đại được quan tâm đặc biệt.

Trong [1] đã đưa ra và chứng minh chi tiết một kết quả như sau:

Định lý 1 ([1]). *Cho* $\{a_i\}_{i\in\mathbb{N}}$ *là một dãy LFSR trên* $\mathbb{F}_q$ *và* ξ *là một nghiệm của đa thức đặc trưng bất khả quy của LFSR.*

Khi đó, tồn tại $\alpha \in \mathbb{F}_{q^n}$ *sao cho*

$$a_i = Tr_{\mathbb{F}_{q^n}/\mathbb{F}_q}(\alpha\xi^i) . \quad (3)$$

Từ đó, chúng tôi đưa ra một hệ quả trực tiếp từ định lý trên và sử dụng hệ quả này để chứng minh một số tính chất của mô hình được đề xuất trong phần sau.

Hệ quả 1. *Cho* $\{a_i\}_{i\in\mathbb{N}}$ *là một m-dãy LFSR trên* $\mathbb{F}_q$ *và* ξ *là một nghiệm của đa thức đặc trưng nguyên thuỷ của LFSR. Khi đó, tồn tại* $\alpha \in \mathbb{F}_{q^n}$ *sao cho*

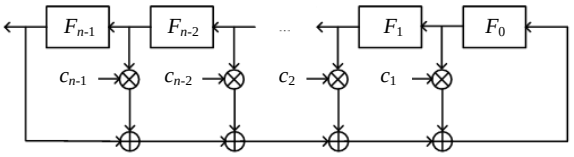
$$a_i = Tr_{\mathbb{F}_{q^n}/\mathbb{F}_q}(\alpha\xi^i) . \quad (4)$$

Phần còn lại của bài báo chỉ xem xét trường hợp $q = 2$ và ký hiệu Tr^L để biểu thị hàm vết $Tr_{\mathbb{F}_{2^L}/\mathbb{F}_2}$.

B. Các cấu hình cho thanh ghi dịch phản hồi tuyến tính

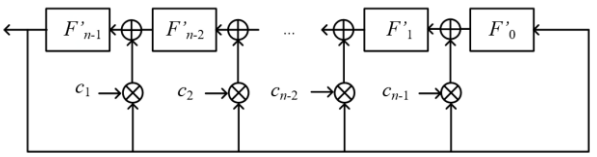
Như đã trình bày trong phần mở đầu, các *m*-dãy sinh bởi các LFSR được ứng dụng nhiều trong các bộ sinh dãy giả ngẫu nhiên hoặc mã dòng sử dụng trong các ứng dụng mật mã. Bởi vì chúng có thể được cài đặt rất hiệu quả.

Để cài đặt các LFSR trong phần cứng, có hai cách cơ bản. Thứ nhất, chúng ta có thể chuyển Hình 2 trực tiếp vào phần cứng, khi đó nó được mô tả như Hình 3. Cách cài đặt này được gọi là cấu hình Fibonacci.



Hình 3. Cấu hình Fibonacci của một LFSR

Cách cài đặt thứ hai có tên gọi là cấu hình Galois (Hình 4) có ưu điểm là tất cả các tín hiệu chỉ phải đi qua nhiều nhất là 1 cổng XOR. Ngược lại, trong cấu hình Fibonacci, nếu đa thức đặc trưng có mật độ dày các bit 1 thì tín hiệu phản hồi phải đi qua xấp xỉ $n/2$ cổng XOR.



Hình 4. Cấu hình Galois của một LFSR

Một điểm thú vị là cấu hình Galois cũng sẽ tạo ra cùng một dãy như cấu hình Fibonacci nếu lựa chọn giá trị khởi đầu phù hợp.

Ta có mệnh đề sau:

Mệnh đề 1. *Cấu hình Galois sẽ tạo ra cùng một dãy như cấu hình Fibonacci nếu nó được khởi tạo với* $F'_{n-1-i} = \sum_{j=0}^{i} F_{n-1-i+j}c_j,\ 0 \le i \le n-1$. *Trong đó* $F_0,\cdots,F_{n-1}$ *là trạng thái ban đầu của cấu hình Fibonacci của LFSR.*

Chứng minh.

Với $i=0$ ta có $F'_{n-1} = F_{n-1}$, vậy bit đầu ra đầu tiên là giống nhau trong cả hai cấu hình.

Chúng ta cần chứng minh rằng trạng thái tiếp theo $\hat{F}_{n-1},\cdots,\hat{F}_0$ của cấu hình Fibonacci và trạng thái tiếp theo $\hat{F}'_{n-1},\cdots,\hat{F}'_0$ của cấu hình Galois cũng phải thoả mãn

$$F'_{n-1-i} = \sum_{j=0}^{i} F_{n-1-i+j}c_j,\ 0 \le i \le n-1\ .$$

Với $1 \le i \le n-2$ ta có

$$\begin{aligned} F'_{n-1-i} &= F'_{n-2-i} + F'_{n-1}c_{i+1} \\ &= \sum_{j=0}^{i+1} F_{n-2-i+j}c_j + F'_{n-1}c_{i+1} \\ &= \sum_{j=0}^{i} F_{n-2-i+j}c_j + F_{n-1}c_{i+1} + F'_{n-1}c_{i+1} \\ &= \sum_{j=0}^{i} F_{n-2-i+j}c_j = \sum_{j=0}^{i} F_{n-1-i+j}c_j. \end{aligned}$$

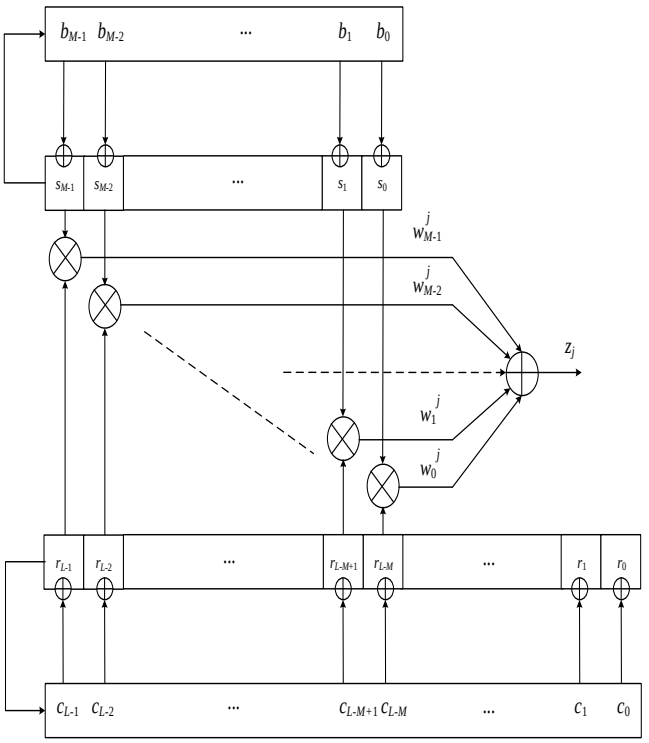
Với $i = n-1$ ta có

$$\begin{aligned} F'_0 &= F'_{n-1} = F_{n-1} \\ &= \sum_{i=0}^{n-1} F_i c_{i+1} + \sum_{i=0}^{n-2} F_i c_{i+1} = F_0 + \sum_{i=0}^{n-2} F_i c_{i+1} \\ &= F_0 + \sum_{i=0}^{n-2} F_{i+1} c_{i+1} = \sum_{j=0}^{n-1} F_j c_j. \end{aligned}$$

Do đó, cả hai cấu hình đều cùng đưa ra một kết quả.

III. BỘ TẠO ĐỀ XUẤT

Giả sử ta có hai thanh ghi dịch phản hồi tuyến tính dạng Galois (Galois LFSR - GLFSR) có chu kỳ cực đại với đa thức nguyên thủy trong vai trò đa thức đặc trưng có bậc lần lượt là *M*, *L*, trong đó, $M < L$ và $M, L = 1$. Khi đó, chúng tôi xây dựng mô hình bộ tạo sau:



Hình 5. Bộ tạo dãy Massey-Rueppel sau khi được tinh chỉnh

Hai GLFSR này sinh ra các dãy tương ứng là

Dãy 1:

$$x_i^j = x_{i-1}^{j-1} + x_{M-1}^{j-1} b_i \pmod 2 \text{ với } 1 \le i \le M-1,$$
$$x_0^j = x_{M-1}^{j-1} b_0 .$$

Hay có biểu diễn ma trận

$$\begin{pmatrix} x_0^j \\ x_1^j \\ \vdots \\ x_{M-1}^j \end{pmatrix} = \begin{pmatrix} 0 & 0 & \cdots & 0 & b_0 \\ 1 & 0 & \cdots & 0 & b_1 \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & 1 & b_{M-1} \end{pmatrix} \begin{pmatrix} x_0^{j-1} \\ x_1^{j-1} \\ \vdots \\ x_{M-1}^{j-1} \end{pmatrix}$$

Dãy 2:

$$r_i^j = r_{i-1}^{j-1} + r_{L-1}^{j-1} c_i \pmod 2 \text{ với } 1 \le i \le L-1,$$
$$r_0^j = r_{L-1}^{j-1} c_0 .$$

Hay có biểu diễn ma trận

$$\begin{pmatrix} r_0^j \\ r_1^j \\ \vdots \\ r_{L-1}^j \end{pmatrix} = \begin{pmatrix} 0 & 0 & \cdots & 0 & c_0 \\ 1 & 0 & \cdots & 0 & c_1 \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & 1 & c_{L-1} \end{pmatrix} \begin{pmatrix} r_0^{j-1} \\ r_1^{j-1} \\ \vdots \\ r_{L-1}^{j-1} \end{pmatrix}$$

Dãy đầu ra được tạo bởi công thức sau:

$$z_j = \sum_{i=0}^{M-1} w_i^j \mod 2 = \left(\sum_{i=0}^{M-1} x_i^j r_{i+L-M}^j \right) \mod 2 . \quad (5)$$

IV. MỘT SỐ TÍNH CHẤT MẬT MÃ CỦA BỘ TẠO

Trong phần này chúng tôi xem xét một số tính chất mật mã đối với bộ tạo đã đề xuất. Cụ thể, xem xét chu kỳ, độ phức tạp tuyến tính và tính cân bằng của dãy đầu ra được sinh bởi bộ tạo đề xuất.

A. Chu kỳ và độ phức tạp tuyến tính

Độ phức tạp tuyến tính $\Lambda\ z$ của dãy z là bậc của đa thức đặc trưng có bậc nhỏ nhất trong số các đa thức đặc trưng của LFSR mà tạo ra dãy z, tức là độ dài của LFSR ngắn nhất tạo ra z. Độ phức tạp tuyến tính thường được sử dụng trong các phân tích mật mã. Cần tính độ phức tạp tuyến tính của dãy đầu ra được sinh bởi bộ tạo khi hai đa thức đặc trưng của 2 LFSR thành phần là các số nguyên tố cùng nhau *L* và *M*.

Ký hiệu $p = 2^M - 1$ và $q = 2^L - 1$ là chu kỳ của các *m*-dãy $X_t = x_t^j{}_{j=0}^{\infty}$ và $R_t = r_t^j{}_{j=0}^{\infty}$ như mô tả trong mô hình bộ tạo. Khi đó ta có kết quả sau:

Mệnh đề 2. *Độ phức tạp tuyến tính* $\mathcal{L}$ *và chu kỳ T của dãy đầu ra* $Z = z_j$ *được tính bởi công thức sau:*

$$\mathcal{L} = L \times M,$$
$$T = p \times q.$$

Để chứng minh mệnh đề 2, chúng tôi sử dụng hai bổ đề được đưa ra trong [2] như sau:

Bổ đề 1, ([2]): *Nếu* γ *và* δ *lần lượt thuộc* $\mathbb{F}_{2^L}$ *và* $\mathbb{F}_{2^M}$, *trong đó* $L, M = 1$ *thì*

$$Tr^L\ \gamma\ Tr^M\ \delta = Tr^{LM}\ \gamma\delta .$$

Bổ đề 2, ([2]): *Nếu các đa thức tối tiểu của* γ *và* δ *trên trường* $\mathbb{F}_2$ *có các bậc là L và M, và* $\gcd L, M = 1$, *thì đa thức tối tiểu của* $\gamma\delta$ *trên trường* $\mathbb{F}_2$ *có bậc là LM.*

Ngoài ra, chúng ta cũng có thể chứng minh kết quả sau:

Bổ đề 3: *Cho* a, b *là các số nguyên dương. Khi đó* $a, b = 1$ *khi và chỉ khi* $2^a - 1, 2^b - 1 = 1$.

Chứng minh:

Điều kiện cần: Giả sử phản chứng $a, b = 1$ và $\gcd\ 2^a - 1, 2^b - 1 = d > 1$. Ta có $2^a \equiv 1 \mod d$ suy ra bậc của 2 mod *d* là ước của *a*.

Tương tự, $2^b \equiv 1 \mod d$ suy ra bậc của 2 mod *d* là ước của *b*. Suy ra $ord_d\ 2$ là ước chung của a và b.

Mặt khác, vì d lẻ nên suy ra $ord_d\ 2 \neq 1$ mâu thuẫn với giả thiết $a, b = 1$. Vậy suy ra $d = 1$.

Điều kiện đủ: Giả sử phản chứng rằng $2^a - 1, 2^b - 1 = 1$ và $a, b = d > 1$.

Ta có $a, b = d$ tức là:

$$\begin{cases} a = k_1 d, \\ b = k_2 d, \end{cases}$$

với $k_1 \neq k_2$ là các số nguyên dương nào đó.

Khi đó

$$\begin{cases} 2^a - 1 = 2^{k_1 d} - 1 = \left(2^d - 1\right)\left(2^{d(k_1-1)} + 2^{d(k_1-2)} + \cdots + 1\right), \\ 2^b - 1 = 2^{k_2 d} - 1 = \left(2^d - 1\right)\left(2^{d(k_2-1)} + 2^{d(k_2-2)} + \cdots + 1\right). \end{cases}$$

Từ đó suy ra $2^d - 1 > 1$ là ước của $2^a - 1$ và $2^b - 1$. Mẫu thuẫn với giả thiết $2^a - 1, 2^b - 1 = 1$. Vậy suy ra $d = 1$ □

Chứng minh mệnh đề 2:

Giả sử α và β lần lượt là các nghiệm của đa thức đặc trưng sinh ra các dãy $X_t = x_t^j$ và $R_t = r_t^j$ trên các trường $\mathbb{F}_{2^M}$ và $\mathbb{F}_{2^L}$. Ta có:

$$x_t^j = Tr^M(A_t\alpha^j), A_t \in \mathbb{F}_{2^M},$$
$$r_{t+L-M}^j = Tr^L(B_{t+L-M}\beta^j), B_{t+L-M} \in \mathbb{F}_{2^L}.$$

Nên ta có

$$w_t^j = x_t r_{t+L-M} = Tr^M(A_t\alpha^j)\,Tr^L(B_{t+L-M}\beta^j).$$

Vì $L, M = 1$ nên áp dụng bổ đề 1, ta có

$$w_t^j = Tr^M(A_t\alpha^j)\,Tr^L(B_{t+6}\beta^j) = Tr^{LM}\left(A_tB_{t+L-M}(\alpha\beta)^j\right).$$

Từ đó suy ra

$$z_j = \sum_{t=0}^{M-1} w_t^j = \sum_{t=0}^{M-1} Tr^{ML}\left(A_tB_{t+L-M}(\alpha\beta)^j\right) = Tr^{ML}\left(A'B'(\alpha\beta)^j\right) = Tr^{ML}(C\gamma^j),$$

với $\gamma = \alpha\beta \in \mathbb{F}_{2^{ML}}$.

Từ bổ đề 2 ta có, đa thức tối tiểu của $\alpha\beta$ có bậc là ML và dãy z_j nhận đa thức tối tiểu đó làm đa thức đặc trưng. Từ biểu diễn toán tử vết của z_j suy ra độ phức tạp tuyến tính của dãy đầu ra bằng bậc của đa thức đặc trưng tối tiểu, hay chính là LM.

Ta thấy rằng chu kỳ của dãy z_j là bậc của phần tử $\alpha\beta$ trong nhóm nhân $\mathbb{F}_{2^{ML}} \setminus 0$. Vì $\alpha \neq 0, \beta \neq 0$ nên $\alpha\beta \neq 0$. Gọi *k* là số tự nhiên nhỏ nhất sao cho $(\alpha\beta)^k = 1$. Khi đó $\alpha^k = \beta^{-k}$.

Nếu $\alpha^k = \beta^{-k} = 1$ thì *k* là bội chung nhỏ nhất của bậc của α và β. Khi đó chu kỳ $T = lcm(2^M - 1, 2^L - 1)$.

Mà $M, L = 1$ theo bổ đề 3 suy ra $(2^M - 1, 2^L - 1) = 1$. Do đó $T = (2^M - 1)(2^L - 1)$.

Nếu $\alpha^k = \beta^{-k} \neq 1$ tức là có một phần tử khác 0 và 1 thuộc $\mathbb{F}_{2^M} \cap \mathbb{F}_{2^L}$. Vì $M, L = 1$ và $\mathbb{F}_{2^M}$ và $\mathbb{F}_{2^L}$ được sinh bởi nghiệm α và β của các đa thức đặc trưng khác nhau nên $\mathbb{F}_{2^M} \cap \mathbb{F}_{2^L} = \mathbb{F}_2$. Suy ra $\alpha^k = \beta^{-k} \in \mathbb{F}_2$ (vô lý). Vậy chu kỳ của dãy z_j là $(2^M - 1)(2^L - 1)$ □

B. Tính cân bằng

Chúng tôi chứng minh được rằng sai khác giữa số bit 0 và số bit 1 của dãy đầu ra là $1/q$. Cụ thể ta có mệnh đề sau:

Mệnh đề 3: *Với các giả thiết như trên, ta có trong một chu kỳ của dãy z_j :*

- *Số các bit 0 ký hiệu là* $n_0 = \dfrac{p(q-1)}{2}$,

- *Số các bit 1 ký hiệu là* $n_1 = \dfrac{p(q+1)}{2}$.

Chứng minh:

Xét các dãy $X_t = x_t^j$, theo tính chất *m*-dãy thì có $2^M - 1$ bộ *M*-bit dạng $x_0^j, x_1^j, \cdots, x_{M-1}^j$ khác nhau và khác 0, mỗi bộ xuất hiện đúng 1 lần trong chu kỳ $2^M - 1$ của dãy *X*.

Xét các dãy $R_t = r_t^j$, theo tính chất *m*-dãy thì có $2^L - 1$ bộ *L*-bit dạng $r_0^j, r_1^j, \cdots, r_{L-1}^j$ khác nhau và khác 0, mỗi bộ xuất hiện đúng 1 lần trong chu kỳ $2^L - 1$ của dãy *R*. Vì $M < L$ nên có $2^M - 1$ bộ *M*-bit khác nhau và khác không, mỗi bộ xuất hiện đúng 2^{L-M} lần, và bộ *M*-bit $0, 0, \cdots, 0$ xuất hiện $2^{L-M} - 1$ lần trong một chu kỳ $2^L - 1$ của dãy *R*.

Vì $M, L = 1$ và chu kỳ $T = pq$ nên áp dụng định lý dãy đồng dư tuyến tính cho chỉ số của 2 dãy sau:

$$x = x_0, x_1, \cdots, x_{2^M-1}$$
$$r = r_0, r_1, \cdots, r_{2^L-1}$$

thì xét dãy tích x, r trong chu kỳ $T = pq$, mỗi phần tử của dãy *x* sẽ được nhân toàn bộ với $2^L - 1$ phần tử của dãy *r*.

Do đó, mỗi bộ *M*-bit $x_0^j, x_1^j, \cdots, x_{M-1}^j$ của dãy *X* sẽ được nhân kết hợp với toàn bộ $2^L - 1$ bộ *M*-bit dạng $r_{L-M}^j, r_{L-M+1}^j, \cdots, r_{L-1}^j$ của dãy *R* trong chu kỳ *T* của dãy *Z*.

Xét dãy

$$z_j = \sum_{i=0}^{M-1} x_i^j r_{i+L-M}^j$$

Tính số các bit 0 của dãy z_j như sau:

Giả sử $x = x_0, x_1, \cdots, x_{M-1}$, $wt\ x = \omega > 0$, $r = r_{L-M}, r_{L-M+1}, \cdots, r_{L-1}$. Khi *x* cố định, ta sẽ tính tất cả các véc tơ *r* thoả mãn

$$\langle x, r \rangle = \sum_{i=0}^{M-1} x_i r_{i+L-M} = 0.$$

Giả sử

$$\begin{cases} x_{j_t} = 1, t = 1,2,\cdots,\omega \\ x_j = 0, j \notin \{j_1, j_2, \cdots, j_\omega\} \end{cases}.$$

Khi đó, $\langle x, r \rangle = 0$ khi và chỉ khi trong bộ $\tilde{r} = r_{j_1}, r_{j_2}, \cdots, r_{j_\omega}$ có số các bit 1 là số chẵn. Có tổng cộng $2^{\omega-1}$ bộ $\tilde{r}$ như vậy, trong đó có đúng 1 véc tơ không chứa bit 1.

Do các giá trị của $r_j, j \notin \{j_1, j_2, \cdots, j_\omega\}$ là tuỳ ý, nên có tất cả $2^{\omega-1} \times 2^{M-\omega} = 2^{M-1}$ véc tơ *r* thoả mãn $\langle x, r \rangle = 0$ bao gồm cả véc tơ $r^* = (0,0,\cdots,0)$.

Vì trong một chu kỳ *T*, mỗi bộ *M*-bit khác không của dãy *R* xuất hiện đúng 2^{L-M} lần, còn bộ *M*-bit $0,0,\cdots,0$ xuất hiện đúng $2^{L-M}-1$. Nên với mỗi bộ *M*-bit khác 0 của dãy *X*, có tất cả

$$\begin{aligned} (2^{M-1}-1) \times 2^{L-M} + 1 \times (2^{L-M}-1) &= 2^{L-1} - 2^{L-M} + 2^{L-M} - 1 \\ &= 2^{L-1} - 1. \end{aligned}$$

số bit 0 của dãy *Z*. Do đó, có tổng số các bit 0 trong dãy *Z* là:

$$n_0 = (2^M - 1) \times (2^{L-1} - 1) = \frac{p(q-1)}{2}.$$

Từ đó suy ra tổng số các bit 1 trong dãy *Z* là

$$\begin{aligned} n_1 &= (2^M - 1)(2^L - 1) - n_0 \\ &= (2^M - 1)(2^L - 1) - (2^M - 1)(2^{L-1} - 1) \\ &= (2^M - 1)(2^L - 2^{L-1}) \\ &= (2^M - 1)2^{L-1} \\ &= \frac{p(q+1)}{2}. \end{aligned}$$

Từ các kết quả trên cho thấy dãy đầu ra là tuần hoàn với chu kỳ $T = (2^M - 1)(2^L - 1)$ cân bằng và có độ phức tạp tuyến tính là $\mathcal{L} = ML$. Tính tương quan của dãy đầu ra chưa được kiểm tra.

V. KẾT QUẢ THỰC NGHIỆM

Ngoài việc đánh giá lý thuyết đối với độ phức tạp tuyến tính, chu kỳ và tính cân bằng của dãy đầu ra, chúng tôi đã thực hiện một số thử nghiệm cho dãy đầu ra của bộ tạo tinh chỉnh.

Kịch bản thử nghiệm: Chúng tôi lựa chọn $L = 61, M = 67$ và tạo ra 10 file dữ liệu MAU01.txt đến MAU10.txt, mỗi file kích thước 128 KB và sử dụng 06 tiêu chuẩn kiểm tra cho dãy độ dài 128 bit gồm: Tiêu chuẩn tần số, tiêu chuẩn Serial, tiêu chuẩn tự tương quan [3], tiêu chuẩn Markov, tiêu chuẩn 3-type [3], tiêu chuẩn độ phức tạp Lempel-Ziv[4]. Chi tiết xem tại Phụ lục cuối bài.

Mục đích thử nghiệm: Đánh giá tính giả ngẫu nhiên đối với dãy đầu ra của bộ tạo tinh chỉnh và so sánh với bộ tạo ban đầu. Do hạn chế về mặt không gian, ở đây chúng tôi chỉ trình bày kết quả 2 mẫu.

Kết quả thu được như sau:

BẢNG 1. KẾT QUẢ THỬ NGHIỆM MẪU 01

MẪU 01					
Tiêu chuẩn kiểm tra	***Tổng số dãy***	***Bộ tạo Massey-Ruepel gốc***		***Bộ tạo tinh chỉnh***	
		Số dãy đạt	***Tỷ lệ đạt***	***Số dãy đạt***	***Tỷ lệ đạt***
Tiêu chuẩn tần số	8192	8064	98.44%	8075	98.57%
Tiêu chuẩn Serial	8192	8041	98.16%	8033	98.06%
Tiêu chuẩn tự tương quan	8192	8016	97.85%	8028	98.00%
Tiêu chuẩn Markov	8192	8025	97.96%	8003	97.69%
Tiêu chuẩn 3-type	8192	8040	98.14%	8018	97.88%
Tiêu chuẩn độ phức tạp Lempel-Ziv	8192	8024	97.95%	8000	97.66%

BẢNG 2. KẾT QUẢ THỬ NGHIỆM MẪU 02

MẪU 02					
Tiêu chuẩn kiểm tra	***Tổng số dãy***	***Bộ tạo Massey-Ruepel gốc***		***Bộ tạo tinh chỉnh***	
		Số dãy đạt	***Tỷ lệ đạt***	***Số dãy đạt***	***Tỷ lệ đạt***
Tiêu chuẩn tần số	8192	8069	98.50%	8032	98.05%
Tiêu chuẩn Serial	8192	8049	98.25%	8020	97.90%
Tiêu chuẩn tự tương quan	8192	8006	97.73%	8010	97.78%
Tiêu chuẩn Markov	8192	8032	98.05%	8018	97.88%
Tiêu chuẩn 3-type	8192	8037	98.11%	8028	98.00%
Tiêu chuẩn độ phức tạp Lempel-Ziv	8192	8010	97.78%	7993	97.57%

Kết quả cho thấy các dãy đầu ra đạt được tính ngẫu nhiên, do đó củng cố thêm cho các kết quả lý thuyết đã trình bày như trên.

VI. KẾT LUẬN

Trong bài báo này, chúng tôi đã đề xuất một tinh chỉnh hiệu quả cho bộ tạo dãy giả ngẫu nhiên Massey-Rueppel hướng phần cứng. Chúng tôi đã chứng minh lý thuyết rằng bộ tạo đề xuất cũng đạt được một số tính chất mật mã như bộ tạo Massey-Rueppel. Hơn thế nữa, bộ tạo đề xuất hiệu quả hơn bộ tạo dãy Massey-Rueppel khi cấu hình phần cứng. Chúng tôi đã sử dụng các kiểm tra thống kê để đánh giá tính ngẫu nhiên của dãy đầu ra. Kết quả thu được cho thấy các dãy đầu ra đạt tính ngẫu nhiên, từ đó củng cố thêm cho các kết quả lý thuyết đã được trình bày.

TÀI LIỆU THAM KHẢO

[1]. Klein, A, "Stream ciphers", Springer, 2013.

[2]. Massey, J.L. and R.A. Rueppel. "Linear Ciphers and Random Sequence Generators with Multiple Clocks" in EUROCRYPT. Springer, 1984.

[3]. Kunuth, D, "The Art of Computer Programming" vol. 2 Seminumerical Algorithms, Reading, Massachusetts: Addison Wesley, 1998.

[4]. Doğanaksoy, A. and F. Göloğlu. "On Lempel-Ziv complexity of sequences". in International Conference on Sequences and Their Applications, Springer, 2006.

SƠ LƯỢC VỀ TÁC GIẢ

TS. Nguyễn Văn Long

Đơn vị công tác: Viện Khoa học - Công nghệ mật mã, Ban Cơ yếu Chính phủ.

Email: longnv@bcy.gov.vn

Quá trình đào tạo: Nhận bằng Kỹ sư chuyên nghành An toàn thông tin các hệ thống viễn thông tại Học viện FSO - Liên bang Nga năm 2008. Bảo vệ thành công luận án Tiến sĩ tại học viện FSO Liên bang Nga theo chuyên ngành Các phương pháp bảo vệ thông tin năm 2015.

Hướng nghiên cứu hiện nay: Nghiên cứu, cài đặt, thiết kế các thuật toán mã khối.

CN. Hoàng Đình Linh

Đơn vị công tác: Viện Khoa học - Công nghệ Mật mã, Ban Cơ yếu Chính phủ.

Email: linhhd@bcy.gov.vn

Quá trình đào tạo: Nhận bằng cử nhân Toán học tại Đại học Khoa học tự nhiên - Đại học Quốc gia Hà Nội năm 2014.

Hướng nghiên cứu hiện nay: Nghiên cứu, thiết kế, đánh giá độ an toàn chứng minh được của các thuật toán mã hóa đối xứng.

Phụ lục: Mô tả một số kiểm tra tính ngẫu nhiên theo thống kê

Tiêu chuẩn tần số đơn

Kiểm tra tần số đơn là kiểm tra đo sự sai khác giữa số các số 1 và số các số 0 trong một dãy *n*-bit. Tính cân bằng là một đặc điểm cơ bản cho một đầu ra của một thuật toán. Do đó, kiểm tra tần số được sử dụng như một bước ban đầu đối với hầu hết tất cả các bộ kiểm tra. Nếu một thuật toán không vượt qua kiểm tra tần số thì một số kiểm tra khác thậm chí không thể áp dụng được.

Cho $s = s_0 s_1 \cdots s_{n-1}$ với $s_i \in \{0,1\}, \forall i = \overline{0, n-1}$ là dãy bit ngẫu nhiên độc lập có cùng phân bố đều. Ký hiệu n_1 là số bit 1 trong dãy *s* và n_0 là số bit 0 trong dãy *s*. Tính

$$X_1 = (n_1 - n_0)^2/n.$$

Khi đó X_1 có phân bố giới hạn là χ^2 với 1 bậc tự do. Thật vậy, có

$$X_1 = \frac{(2n_1 - n)^2}{n} = \left(\frac{n_1 - n/2}{\sqrt{n \cdot \frac{1}{4}}} \right)^2.$$

Ta có $n_1 = s_0 + s_1 + ... + s_{n-1}$ với $E(s_i) = 1/2$ và $\sigma^2(s_i) = 1/4$. Suy ra $\frac{n_1 - n/2}{\sqrt{n/4}}$ có phân bố giới hạn là phân bố chuẩn theo Định lý Giới hạn Trung tâm, hay X_1 có phân bố giới hạn là χ^2 với 1 bậc tự do khi $n \geq 10$.

Tiêu chuẩn Serial

Mục đích của kiểm tra này là xác định xem khi nào số lần xuất hiện các mẫu 00, 01, 10, 11 trong dãy *s* là xấp xỉ với số lần xuất hiện của chúng trong dãy ngẫu nhiên. Ký hiệu n_1, n_0 lần lượt là số các bit 1 và số các bit 0 của dãy *s*. Tương tự, ký hiệu $n_{00}, n_{01}, n_{10}, n_{11}$ lần lượt là số lần xuất hiện các mẫu 00, 01, 10, 11 trong dãy *s*.

Ta có:

$$n_0 + n_1 = n,$$
$$n_{00} + n_{01} + n_{10} + n_{11} = n - 1.$$

Khi đó,

$$X = \frac{4}{n-1}\left(n_{00}^2 + n_{01}^2 + n_{10}^2 + n_{11}^2\right) - \frac{2}{n}\left(n_0^2 + n_1^2\right) + 1$$

sẽ tiệm cận với phân bố χ^2 với $2^2 - 2 = 2$ bậc tự do khi $n \geq 21$.

Tiêu chuẩn tự tương quan

Hệ số tự tương quan (autocorrelation) *A*(*j*) của dãy *s* được định nghĩa là:

$$A(j) = \sum_{i=0}^{n-1-j} s_i \oplus s_{i+j}.$$

Chúng ta đã chứng minh được rằng

$$X = 2\left(A(j) - \frac{n-j}{2}\right)\Big/\sqrt{n-j},$$

sẽ tuân theo phân bố chuẩn □(0, 1) theo Định lý Giới hạn Trung tâm.

Tiêu chuẩn xích Markov

Tiêu chuẩn xích Markov xem xét giá trị thống kê sau

$$X = \frac{S_{n-1} - (n-1)/2}{\sqrt{(n-1)/4}},$$

trong đó $S_{n-1} = \sum_{j=0}^{n-2} (s_j - s_{j+1})^2$. Chúng ta có thể chứng minh được rằng *X* tuân theo phân bố chuẩn $\mathcal{N}(0, 1)$ theo Định lý Giới hạn Trung tâm.

Tiêu chuẩn m-type

Tiêu chuẩn *m*-type xem xét số lần xuất hiện của một mẫu *m*-bit trong dãy độ dài *n*-bit.

$$X = \frac{2^m}{k}\left(\sum_{i=1}^{2^m} n_i^2\right) - k;$$

trong đó $k = n/m$, và n_i là số lần xuất hiện của mẫu thứ *i*. Chúng ta có thể chứng minh được rằng *X* tuân theo phân bố chuẩn $\mathcal{N}(0, 1)$ theo Định lý Giới hạn Trung tâm.