NGHIÊN CỨU ỨNG DỤNG MÃ BCH XÂY DỰNG HỆ MẬT

A SECURE NIEDREITER CRYPTOSYSTEM'S VARIANT BASE ON BCH CODES

Lê Văn Thái

TÓM TẮT

Nội dung bài báo đề xuất giải pháp sử dụng ghép các mã BCH thành phần nhằm giảm kích thước khóa của hệ mật mã dựa trên mã. Để mở rộng khả năng sửa lỗi của mã BCH và ứng dụng vào xây dựng hệ mật, bài báo sử dụng phương pháp chuẩn syndrome giải mã mã BCH. Hệ mật đề xuất có kích thước khóa công khai giảm 5,7 lần so với hệ mật Niederreiter trong đề xuất gốc ở cùng mức an ninh và đảm bảo an toàn chống lại các cuộc tấn công cấu trúc và tấn công giải mã.

Từ khóa: *Hệ mật McEliece, hệ mật Niederreiter, chuẩn syndrome, mã BCH, hệ mật dựa trên mã.*

ABSTRACT

In this paper, we propose a solution to merge BCH codes into chained BCH codes and applications to build the cryptosystem. The proposed method reduced the key size by 5.7 times compared to the Niederreiter cryptosystem at the same level of security. The proposed cryptosystem guarantees security against structural attacks and decryption attacks. At the same time, the article also presented the norm-syndrome based decoding method of BCH code. This method has increased the ratio of the number of syndromes that can be decoded out of the total number of possible syndromes, extending the application range of the BCH code.

Keywords*: McEliece Cryptosystem, Niderrreiter Cryptosystem, Norm syndrome, BCH Codes, Code based cryptosystem.*

Trường Đại học Công nghiệp Hà Nội
Email: thailv@haui.edu.vn
Ngày nhận bài: 15/01/2019
Ngày nhận bài sửa sau phản biện: 03/4/2019
Ngày chấp nhận đăng: 25/4/2019

1. ĐẶT VẤN ĐỀ

Hệ thống mã hóa khóa công khai hiện nay hầu hết dựa trên độ khó của các bài toán lý thuyết số như bài toán phân tích ra thừa số, bài toán logarit rời rạc trên trường hữu hạn. Kết quả nghiên cứu của Peter Shor năm 1994 và thuật toán tìm kiếm trên dữ liệu không có cấu trúc của Grover năm 1996 đã cảnh báo các hệ mật khóa công khai RSA, ElGamal,… sẽ không an toàn khi máy tính lượng tử với quy mô đủ lớn xây dựng thành công [1]. Hệ mật mã dựa trên mã McEliece được giới thiệu năm 1978 [2]. Đây là sơ đồ hệ mật đầu tiên sử dụng tính ngẫu nhiên trong mã hóa. An ninh của hệ mật này dựa trên độ khó của bài toán giải mã theo syndrome và đã được chứng minh là bài toán NP đầy đủ [3]. Đề xuất ban đầu sử dụng mã nhị phân Goppa và thuật toán giải mã Patterson. Ưu điểm của hệ mật này là tính bảo mật cao, thời gian thực hiện mã hoá và giải mã nhanh, yêu cầu thiết bị thực hiện đơn giản. Hơn nữa, hệ mật này được chứng minh là có khả năng chống lại sự tấn công lượng tử [4]. Tuy nhiên, hệ mật này chưa được áp dụng trong thực tế xuất phát từ nhược điểm cơ bản của nó là tỷ lệ mã hóa thấp, kích thước khóa khá lớn.

Năm 1986, biến thể của hệ mật McEliece là hệ mật Niederreiter được đề xuất [5]. Hệ mật Niederreiter sử dụng ma trận kiểm tra H để làm khóa và sử dụng vector lỗi để giải mã. An ninh của hệ mật McEliece và hệ mật Niederreiter khi sử dụng mã nhị phân Goppa được chứng minh là hoàn toàn tương đương [6]. Ưu điểm của hệ mật Niederreiter là có khả năng áp dụng để xây dựng sơ đồ chữ ký số, ứng dụng trong thực tế [7].

Trong quá trình phát triển của hệ mật mã dựa trên mã. Đã có nhiều đề xuất thay thế mã Goppa trong hệ mật gốc bằng các mã khác nhằm giảm kích thước khóa. Năm 1994, Sidelnikov đã đề xuất sử dụng mã Reed-Muller áp dụng cho hệ mật Niederreiter. Năm 1996, Heeralal Janwa và Oscar Moreno đã đề xuất hệ mật sử dụng mã AG (algebraic-geometric). Năm 2005, Berger và Loidreau đã đề xuất sử dụng mã *quasi-cyclic alternant* làm ẩn cấu trúc khóa mật. Những năm gần đây có nhiều đề xuất sử dụng các họ mã và phương pháp giải mã mới nhằm làm giảm kích thước khóa. Monico và cộng sự đề xuất sử dụng mã kiểm tra mật độ thấp (LDPC). Năm 2007, Baldi và cộng sự đề xuất một biến thể mới dựa trên mã *quasi-cyclic* (QC-LDPC). Năm 2013, Misoczki và cộng sự đề xuất sử dụng mã kiểm tra mật độ trung bình (QC-MDPC). Năm 2016, Moufek đã đề xuất kết hợp mã QC-LDPC và QC-MDPC và sử dụng bộ tạo số giả ngẫu nhiên để tạo ma trận sinh. Tuy nhiên, các nghiên cứu mới về tấn công đã chỉ ra các đề xuất này không an toàn với tấn công cấu trúc [8 , 9 , 10].

Trong bài báo này, tác giả đề xuất sử dụng ghép các mã BCH thành chuỗi mã và áp dụng cho biến thể Niederreiter để xây dựng hệ mật. Sơ đồ hệ mật đề xuất cho phép giảm kích thước khóa, đảm bảo an toàn với các tấn công giải mã và tấn công cấu trúc. Đồng thời trong bài báo, tác giả cũng đề xuất phương pháp chuẩn syndrome giải mã mã BCH nhằm mở rộng khả năng sửa lỗi và phạm vi ứng dụng của mã BCH, cho phép ứng dụng mã BCH để xây dựng hệ mật mã dựa trên mã.

Phần còn lại của bài báo được tổ chức như sau: trong phần 2, trình bày giải pháp nâng cao hiệu quả sửa lỗi của

mã BCH sử dụng phương pháp chuẩn syndrome. Phần 3, đề xuất xây dựng hệ mật sử dụng mã ghép BCH. Phần này cũng giới thiệu về hệ mật McEliece và Niederreiter. Phần 4, đánh giá độ phức tạp và độ an toàn của hệ mật đề xuất.

2. NÂNG CAO HIỆU QUẢ CỦA MÃ BCH SỬ DỤNG PHƯƠNG PHÁP CHUẨN SYNDROME

Các hệ mật dựa trên mã sửa lỗi không thể áp dụng để mã hóa cho một bản tin bất kỳ. Bởi vì một syndrome ngẫu nhiên hầu như tương ứng với vector lỗi có trọng lượng lớn hơn khả năng sửa lỗi của mã. Do đó, để áp dụng mã BCH vào xây dựng hệ mật, nhiệm vụ đầu tiên là xây dựng phương pháp giải mã để nâng cao hiệu quả sửa lỗi của mã. Trong nội dung tiếp theo tác giả xây dựng phương pháp giải mã mã BCH theo chuẩn syndrome (Norm Syndrome). Khi phân hoạch các vector lỗi thành các lớp không giao nhau có chuẩn syndrome phân biệt cho phép mở rộng khả năng sửa lỗi của mã BCH.

Tham số chuẩn syndrome được xây dựng dựa trên cấu trúc của mã BCH và các biến thể. Đặc điểm của chuẩn syndrome là tính bất biến với tác động của nhóm các dịch vòng. Syndrome của các nhóm khác nhau thì khác nhau. Khi sử dụng chuẩn syndrome để giải mã, có thể sửa được lỗi ngẫu nhiên và lỗi cụm. Vì khi chọn đa thức sinh thích hợp thì chuẩn syndrome của các vector lỗi ngẫu nhiên và một số cấu hình lỗi cụm độ dài nhỏ, lỗi cụm đồng pha là không trùng nhau.

Giả thiết σ là phép thế dịch vòng, dưới tác động của σ, vector lỗi dịch phải một vị trí. Tập hợp tất cả các vector khác nhau đôi một $\sigma^i(e)$ với $0 \le i \le n-1$ của vector lỗi e bất kỳ được gọi là σ-*orbit* của nó. Các phần tử của σ-*orbit* chuyển hóa lẫn nhau dưới tác động của phép dịch vòng. Mỗi σ-*orbit* có một vector sinh, tọa độ đầu tiên của vector này luôn khác 0.

Ta có $\sigma^\lambda(e) = e$ với λ là số tự nhiên $1 \le \lambda \le n$. Với một vector lỗi e bất kỳ σ-*orbit* chứa k phần tử trong đó $\lambda = n$ hoặc λ là ước của nó. Khi đó cấu trúc của σ-*orbit* có dạng:

$$\sigma(e) = \langle e \rangle = \left\{ e, \sigma(e), ..., \sigma^{\lambda-1}(e) \right\} \quad (1)$$

Hai vector lỗi tùy ý e và e' thì các σ-*orbit* $<e>$, $<e'>$ hoặc là trùng nhau hoặc không giao nhau. Do vậy dưới tác động của nhóm các phép dịch vòng không gian vector lỗi phân chia thành các lớp σ-*orbit* không giao nhau.

Ma trận kiểm tra của mã BCH tổng quát với $\delta = 2t + 1$ có dạng [11]:

$$H = \left[\beta^{bi}, \beta^{(b+1)i}, ..., \beta^{(b+\delta-2)i} \right]^T, 0 \le i \le n-1. \quad (2)$$

Giả sử hạng của ma trận kiểm tra là $m(\delta\text{-}1)$, tức là các hàng của ma trận H là độc lập tuyến tính. Khi đó, syndrome của vector lỗi e gồm δ-1 thành phần trong trường GF(2^m) có dạng $S(e) = (s_1, s_2, \ldots, s_{\delta-1})$.

Cho e là vector lỗi tùy ý, với mã BCH có ma trận kiểm tra (2) ta có:

$$S(\sigma(e)) = (\beta^b s_1, \beta^{b+1} s_2, ..., \beta^{b+\delta-2} s_{\delta-1}). \quad (3)$$

Như vậy, chuẩn syndrome là vector $N(S)$ có $C^2_{\delta-1}$ tọa độ $N_{ij}(1 \le i \le j \le \delta\text{-}1)$, được xác định theo công thức [12]:

$$\begin{aligned} N_{ij} &= s_j^{(b+i-1)/h_{ij}} / s_i^{(b+j-1)/h_{ij}} \text{ if } s_i \ne 0,\ h_{ij} = \gcd(b+i-1,\ b+j-1); \\ N_{ij} &= \infty \qquad \text{if } s_j \ne 0;\ s_i = 0; \\ N_{ij} &= - \qquad \text{if } s_i = s_j = 0. \end{aligned} \quad (4)$$

Tính chất của chuẩn syndrome là tính bất biến của nó với phép thế dịch vòng. Từ công thức (3, 4) ta có, đối với mọi mã vector lỗi e của mã BCH luôn thỏa mãn:

$$N(s(\sigma(e))) = N(s(e)) \quad (5)$$

Bản chất của phương pháp giải mã theo chuẩn syndrome là các phần tử của σ-*orbit* chuyển hóa lẫn nhau dưới tác động của phép dịch vòng. Chuẩn syndrome sẽ chỉ ra σ-*orbit* mà vector lỗi nằm trong đó. Do đó xác định được vector sinh e_0 tương ứng, so sánh syndrome nhận được S và $S(e_0)$ ta xác định được lượng dịch vòng để biến đổi e_0 thành e, do đó sẽ tìm được chính xác vector lỗi [12].

Các bước thực hiện giải mã theo phương pháp chuẩn syndrome:

+ Tính syndrome $S(e) = (s_1, s_2, \ldots, s_t)$ với s_i là phần tử của trường Galoa GF(2^m).

+ Tính bậc của chuẩn syndrome N: Tính $\deg s_j$, $\deg s_i$ là bậc thành phần s_j, s_i của syndrome $S(e) = (s_1, s_2, \ldots, s_i, s_j, \ldots, s_t)$ với $1 \le i \le j \le t$. Chuẩn syndrome của syndrome $S(e)$ tính theo công thức (4), xác định bậc của nó $\deg N_{ij}$.

+ Theo $\deg N_{ij}$ xác định vector sinh và bậc i_0 của thành phần syndrome đầu tiên s^0_i ứng với vector sinh.

+ Tính số thứ tự bit lỗi đầu tiên theo công thức $L_i = (\deg s_i - \deg s^0_i) \bmod n$.

+ Tìm vector lỗi e bằng cách dịch vòng vector sinh đi L_i nhịp.

+ Sửa tín hiệu nhận được: Cộng tín hiệu nhận được với vector lỗi e.

Khảo sát trên trường GF(2^m) với các đa thức sinh khác nhau, dựa trên phương pháp chuẩn syndrome, chúng ta có thể phân hoạch tập các vector lỗi thành các tập con không giao nhau. Khi đó mã BCH và biến thể của nó có thể sửa được một số cấu hình lỗi ngoài khoảng cách mã. Vì vậy khi sử dụng phương pháp chuẩn syndrome giải mã mã BCH ta có thể áp dụng mã BCH để xây dựng hệ mật mã dựa trên mã [13].

3. ĐỀ XUẤT XÂY DỰNG HỆ MẬT SỬ DỤNG MÃ BCH

3.1. Hệ mật McEliece và Niederreiter

***Hệ mật mã khóa công khai McEliece* [2]:**

Tạo khóa: Chọn một mã tuyến tính nhị phân C có khả năng sửa được t lỗi. Ma trận sinh G kích thước $K \times N$. Chọn một ma trận nhị phân khả nghịch Q kích thước $K \times K$. Chọn một ma trận hoán vị ngẫu nhiên P kích thước $N \times N$. Tính toán khóa công khai $G' = Q.G.P$ kích thước $K \times N$. Các ma trận (Q, G, P) là khóa bí mật.

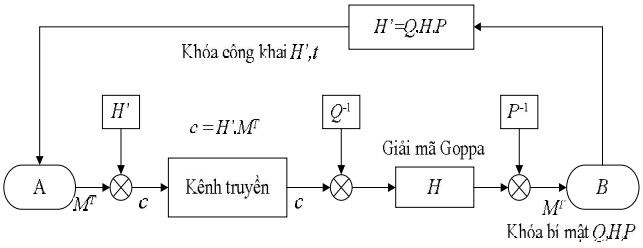
Mã hóa: Khi muốn gửi bản tin M tới bên nhận thông qua khóa công khai (G', t). Biểu diễn bản tin M như một chuỗi

nhị phân có độ dài *k* bit. Tạo một vector *e* ngẫu nhiên có độ dài *N* và có trọng số $w(e) \le t$. Tính toán bản mã $c = MG' + e$ và gửi cho bên nhận.

Giải mã: Sau khi nhận được *c*, bên nhận thực hiện giải mã bản tin. Tính $cP^{-1} = M(QGP)P^{-1} + eP^{-1} = MQG + eP^{-1}$. Sử dụng thuật toán giải mã sửa lỗi đối với cP^{-1} để tìm được *MQ*. Tính $M = (MQ)Q^{-1}$, xác định được bản tin gốc ban đầu.

***Hệ mật khóa công khai Niederreiter* [5]:**

Hệ mật Niederreiter là một biến thể của hệ mật McEliece. Điểm khác là nó sử dụng ma trận kiểm tra *H* của mã Goppa để làm khóa thay thế cho ma trận sinh *G* trong hệ mật McEliece gốc. Sơ đồ hệ mật được thể hiện trên hình 1.



Hình 1. Sơ đồ hệ mật mã khóa công khai Niederreiter

Tạo khóa: Chọn mã Goppa (*N,K*) có khả năng sửa *t* lỗi, có ma trận kiểm tra *H* kích thước (*N-K*)×*N*. Chọn ma trận khả nghịch *Q* kích thước (*N-K*)×(*N-K*). Chọn ma trận chuyển vị *P*(*N*×*N*). Tính khóa công khai $H' = Q.H.P$. Các ma trận (*Q, H, P*) là các khóa bí mật.

Mã hóa: Với khóa công khai (*H', t*), bản tin *M* cho dưới dạng chuỗi nhị phân dài *N* bit có trọng số nhỏ hơn hoặc bằng *t*, bên gửi sẽ thực hiện tính bản mã: $c = H'.M^T$.

Giải mã: Bên nhận sở hữu khóa mật tiến hành thực hiện tính:

$c' = Q^{-1}c = Q^{-1}H'.M^T = Q^{-1}.Q.H.P.M^T = H.P.M^T$;

Trong đó, *c'* là một trong các syndrome của mã Goppa được sử dụng.

Sử dụng thuật toán giải mã theo syndrome cho mã (*N,K*) ta tìm được $M' = P.M^T$.

Tính bản tin $M^T = P^{-1}.M'$ và xác định bản tin gốc *M*.

3.2. Xây dựng hệ mật sử dụng mã BCH

Để giảm kích thước khóa, khắc phục nhược điểm của hệ mật Niederreiter, tác giả đề xuất sử dụng giải pháp ghép các mã BCH thành phần. Các mã thành phần với chiều dài và kích thước mã không lớn, sử dụng phương pháp giải mã dựa trên chuẩn syndrome mở rộng khả năng sửa ngoài giới hạn khoảng cách mã, nâng cao được tỷ lệ số syndrome có thể giải mã được trên tổng số syndrome có thể có của mã BCH.

Để xây dựng một hệ mật mã dựa trên mã ghép BCH, cần sử dụng một họ mã tuyến tính với đặc điểm mã hóa tốt. Mỗi mã của mã ghép này cần có một thuật toán giải mã với độ phức tạp đa thức. Ký hiệu Γ là họ mã tuyến tính. Một mã $C_i \in \Gamma$ sẽ được định nghĩa bởi độ dài n_i, số bit thông tin k_i và khả năng sửa lỗi t_i. Mã ghép này phải đủ lớn để chống lại tấn công vét cạn và mỗi mã C_i của mã ghép được xác định bởi ma trận kiểm tra H_i. Hệ thống xây dựng được từ các mã thành phần này được gọi là mã ma trận kiểm tra tổng. Giả thiết họ này có ℓ mã, ma trận kiểm tra có dạng là một ma trận đường chéo chính với các phần tử là các ma trận H_i ($i = 1 \div \ell$).

Giả sử dụng các mã BCH nhị phân và các biến thể (mã BCH thuận nghịch và mở rộng) làm các mã thành phần. Các giá trị chuẩn syndrome và vector sinh được sắp xếp trong các bảng. Để giải mã từ mã, thực hiện tính syndrome và chuẩn syndrome của nó. Từ chuẩn syndrome ta tìm được vector sinh và dựa vào bậc của syndrome thành phần s_1 ta tính được số lượng dịch vòng. Do đó ta xác định vector lỗi tương ứng. Các thuật toán sử dụng trong hệ mật đề xuất như sau:

Tạo khóa: Chọn một họ ℓ mã BCH và mã BCH mở rộng, H_i là ma trận kiểm tra và t_i là số lỗi có thể sửa được, với $i = 1 \div \ell$. Ma trận kiểm tra của các mã thành phần được sắp xếp theo đường chéo chính tạo thành ma trận kiểm tra *H* có cấu trúc như sau:

$$H\left[(N-K)\times N\right] = \begin{bmatrix} H_1 & 0 & 0 \\ 0 & H_i & 0 \\ 0 & 0 & H_\ell \end{bmatrix} \quad (6)$$

- Chọn một ma trận khả nghịch *Q*[(*N-K*)×(*N-K*)], và chọn một ma trận hoán vị *P*[*N*× *N*] trong trường GF(2). Trong đó ma trận *P* là ma trận đường chéo chính với các thành phần P_i ($i = 1 \div \ell$) là các ma trận hoán vị cấp n_i.

- Tính khóa công khai *H'*: $H' = Q.H.P$. Đây là ma trận kiểm tra của một mã tương đương với mã ghép BCH.

- Khóa công khai là (*H',t*). Trong đó *t* là tổng số lỗi có thể sửa được $t = \sum t_i$ ($i = 1 \div \ell$).

- Khóa mật là các ma trận (*Q,H,P*). Trong đó *H* là ma trận kiểm tra của mã ghép BCH.

Mã hóa: Bản tin cần truyền đi *M* được biểu diễn dưới dạng chuỗi nhị phân dài *N* bit có cấu trúc dạng $M_1||M_2||\ldots||M_l$ với độ dài đoạn M_i là n_i bit có trọng số nhỏ hơn hoặc bằng t_i. Phía gửi thực hiện tính bản mã $c = H'.M^T$.

Giải mã: Để giải mã bản mã *c*, Phía nhận sử dụng khóa mật và phương pháp chuẩn syndrome thực hiện giải mã theo các bước sau:

- Tính $c' = Q^{-1}.c = H.P.M^T$; *c'* là một trong các syndrome của mã được sử dụng.

- Từ *c'* thực hiện tính $M' = P.M^T$. Sử dụng thuật toán giải mã BCH dựa theo chuẩn syndrome.

- Từ *M'* xác định bản tin M^T: $M^T = P^{-1}.M'$. Từ đó ta khôi phục bản tin gốc *M*.

4. ĐÁNH GIÁ ĐỘ PHỨC TẠP VÀ AN NINH HỆ MẬT

4.1. Lựa chọn tham số

Hệ mật sử dụng mã ghép BCH đề xuất, cho phép sử dụng các bộ mã với các tham số mã thành phần n_i, k_i, t_i khác nhau. Tuy nhiên thuật toán giải mã phải đảm bảo có độ phức tạp đa thức và các tham số của bộ mã tổng phải đảm bảo đủ lớn để chống lại tấn công vét cạn. Việc đánh giá độ an toàn bảo mật và độ phức tạp thực hiện của hệ mật phụ

thuộc vào bộ tham số lựa chọn. Qua khảo sát sự phụ thuộc của các tham số mã *N*, *K*, *t* vào độ bảo mật của sơ đồ hệ mật bằng việc xác định giới hạn độ phức tạp (*WF*) của các thuật toán tấn công Canteaut-Chabaud [14] và thuật toán tấn công ngày sinh nhật [16] và để đảm bảo độ bảo mật của hệ mật đề xuất, tác giả chọn bộ mã gồm 10 mã BCH thành phần, gồm: Một mã C_5(31,21) và mã thuận nghịch mở rộng C_6(32,21), ba mã C_7(31,16), một mã C_8(32,16), hai mã C_7(63,45) trên trường GF(2^6), hai mã C_7(127,106), mỗi mã nói trên cho phép mở rộng khả năng sửa thêm 1 lỗi, ngoại trừ mã C_7(31,16) có khả năng sửa đến 5 lỗi. Khi đó tổng số bit kiểm tra bằng 160, tổng chiều dài mã hóa $N = \sum n_i = 568$ và $K = \sum k_i = 408$. Tốc độ mã hóa $R = K/N = 0{,}72$. Số lượng lỗi có thể sửa được tối đa: $t = \sum t_i = 41$.

Bộ mã ghép BCH gồm 10 mã thành phần với các tham số trên, đã đáp ứng các yêu cầu mức độ an toàn của hệ mật. Thông qua việc mã hóa, giải mã các mã thành phần, có chiều dài và kích thước không lớn, hệ mật đề xuất đã giảm được độ phức tạp thực hiện, tăng được tốc độ thực hiện mã hóa và giải mã. Đồng thời khi ghép các mã thành phần thành mã tổng làm tăng độ phức tạp của các thuật toán tấn công vào hệ mật đề xuất.

4.2. Đánh giá độ phức tạp thực hiện của hệ mật đề xuất

Độ phức tạp thực hiện của hệ mật phụ thuộc vào thuật toán giải mã các mã BCH thành phần. Giả thiết các mã thành phần có cùng tham số ($n_i = n$, $k_i = k$). Mỗi cặp mã BCH và mã thuận nghịch sử dụng cùng đa thức sinh của trường GF(2^m) số lượng chuẩn syndrome được xác định theo công thức (7):

$$T_{Syndrome} \approx \sum_{j=1}^{t_i} C_n^j \Big/ n \qquad (7)$$

Việc thực hiện tính toán chuẩn syndrome tương đương với việc phải sử dụng một bộ nhớ có dung lượng $m.2^m = n.\log_2 n$. Trong sơ đồ hệ mật dựa trên mã ghép BCH với họ mã sử dụng gồm ℓ mã thành phần. Do đó, độ phức tạp để thực hiện giải mã cho ℓ mã thành phần được xác định theo công thức (8):

$$WF_1 = \ell.(\sum_{j=1}^{t_i} C_n^j / n).n.\log_2 n = \ell.\sum_{j=1}^{t_i} C_n^j.\log_2 n \qquad (8)$$

Phần còn lại là các phép nhân ma trận nhị phân với độ phức tạp $(N)^2/2$ và $(N\text{-}K)^2/2$ phép toán nhị phân, trong đó $N = \sum n_i$, $K = \sum k_i$ với $i = 1 \div \ell$. Do đó độ phức tạp thực hiện của hệ mật đề xuất là:

$$WF_{ghep} = \ell.\sum_{j=1}^{t_i} C_n^j.\log_2 n + \frac{(n.\ell)^2}{2} + \frac{(n-k)^2.\ell^2}{2} \qquad (9)$$

Với bộ tham số đề xuất lựa chọn, ước lượng độ phức tạp thực hiện của hệ mật sử dụng mã ghép BCH xác định theo công thức (9) $WF_{ghep} = 2^{24,6}$.

4.3. Đánh giá độ bảo mật của hệ mật đề xuất

4.3.1. Tấn công giải mã

Thuật toán tấn công hiệu quả nhất với hệ mật mã dựa trên mã là giải mã tập thông tin ISD (*Information-Set Decoding*). Giải pháp thực hiện của thuật toán ISD được mô tả như sau:

Phía tấn công không biết cấu trúc của mã bí mật vì vậy phải giải mã một mã ngẫu nhiên. Để thực hiện tấn công, phía tấn công chọn ngẫu nhiên *k* trong *n* tọa độ của *c* và ký hiệu là vector c_k (k bit). Ký hiệu G'_k và e_k lần lượt là *k* cột của *G'* và các vị trí tương ứng của *e*. Ta có $c_k = MG'_k + e_k$ hay $(c_k + e_k)(G'_k)^{-1} = M$. Nếu *k* thành phần của e_k bằng 0 thì ta có $c_k(G'_k)^{-1} = M$ và có thể khôi phục lại thông điệp gốc mà không cần giải mã.

Lee và Brickell là các tác giả đầu tiên phân tích an ninh của hệ mật mã dựa trên mã. Trên cơ sở tính toán khoảng cách mã tối thiểu Leon đã phát triển cách tấn công này bằng cách tìm kiếm từ mã trọng số thấp. Phương pháp này tiếp tục được Stern tối ưu bằng cách chia tập thông tin thành 2 phần, do đó làm tăng được tốc độ tìm kiếm các từ mã có trọng số thấp dựa trên thuật toán tấn công ngày sinh nhật. Một số cải tiến khác cũng đã được đề xuất: Canteaut và Chabaud [14], Bernstein và các cộng sự [15], Finiasz và Sendrier [16]. Trong [17] đã chỉ ra xác suất để thực hiện giải mã thành công cho một lần lặp của thuật toán tương ứng với các trọng số khác nhau của Lee và Brickell (P_{LB}), Leon (P_L), Stern (P_S), công thức (10):

$$P_{LB} = \frac{C_k^p.C_{n-k}^{t-p}}{C_n^t},\quad P_L = \frac{C_k^p.C_{n-k-v}^{t-p}}{C_n^t},\quad P_S = \frac{\left(C_{\lceil k/2 \rceil}^p\right)^2.C_{n-k-v}^{t-2p}}{C_n^t} \qquad (10)$$

Thuật toán giải mã Canteaut-Chabaud

Cho *C* là một mã có chiều dài *n* trên trường F_2 và $y \in F_2^n$ có khoảng cách *t* so với một từ mã $c \in C$, thì *y-c* là phần tử trọng số *t* của mã *C*+{0,*y*}. Vì vậy nếu *C* là mã dài *n* trong F_2 với khoảng cách mã tối thiểu lớn hơn *t*, thì một phần tử $e \in C$+{0,*y*} trọng số *t* không thể thuộc mã *C*, cho nên nó phải thuộc mã *C*+{*y*}; nghĩa là *y-c* là một phần tử của mã *C* có khoảng cách *t* so với *y*. Bản mã của hệ mật McEliece $y \in F_2^n$ có khoảng cách *t* với từ mã gần nhất $c \in C$ có khoảng cách mã tối thiểu ít nhất là 2*t*+1. Phía tấn công biết khóa công khai của hệ mật McEliece là ma trận sinh của *C* và có thể tìm *y* với tập các ma trận sinh của *C*+{0,*y*}. Chỉ có từ mã trọng số *t* trong *C*+{0,*y*} là *y-c* bằng cách tìm từ mã này phía tấn công tìm được *c* và từ đó dễ dàng khôi phục được bản rõ.

Tình huống tương tự có thể áp dụng với hệ mật Niederreiter với khóa công khai là ma trận kiểm tra của *C*. Bằng các biến đổi tuyến tính phía tấn công sẽ dễ dàng tìm được ma trận sinh của *C* và tiến hành xử lý bằng phương pháp như trên. Với bản mã đã cho của hệ mật Niederreiter sử dụng đại số tuyến tính phía tấn công tìm từ mã thỏa mãn khi nhân với ma trận kiểm tra tạo ra bản mã đặc biệt. Điểm mấu chốt của các tấn công trên là tìm từ mã có trọng số *t* trong *C*+{0,*y*}. Giới hạn dưới độ phức tạp *WF* (*work factor*) của thuật toán Canteaut-Chabaud được trình bày theo công thức (11) [17]:

$$WF(n,k,t) \geq \min_p \left(\frac{3.\ell}{2^\ell} \frac{C_n^t}{C_{n-k-\ell}^{t-2p}} \right),\ \ell = \log_2\left(C_{\lceil k/2 \rceil}^p\right). \qquad (11)$$

Tấn công ngày sinh nhật

Bài toán giải mã syndrome (Computational Syndrome Decoding - CSD): Cho trước ma trận $H \in \{0,1\}^{(n-k)\times n}$, một từ $s \in \{0,1\}^{(n-k)}$ và một số nguyên dương t, tìm từ $e \in \{0,1\}^n$ với trọng số Hamming nhỏ hơn t sao cho $H.e^T = s$.

Ký hiệu bài toán trên là CSD(H,s,t). Bài toán này tương đương với giải mã sửa t lỗi bằng mã có ma trận kiểm tra H. Bài toán giải mã syndrome như vậy đã được chứng minh là NP-đầy đủ [3]. Với tấn công ngày sinh nhật giới hạn dưới độ phức tạp được xác định bởi công thức (12) [16]:

$$WF_{BA}(n,k,t) \approx 2L\log(2.t.L),\ L = \min\left(\sqrt{C_n^t}, 2^{(n-k)/2}\right) \qquad (12)$$

Với t lẻ và $C_n^{\lceil t/2\rceil} < L$, công thức (12) chỉ là giới hạn dưới, tính toán chính xác được xác định theo công thức (13):

$$WF_{BA}(n,k,t) \approx 2L'\log_2\left(2.t.\frac{L^2}{L'}\right),\ L' = \frac{\left(C_n^{\lceil t/2\rceil}\right)^2 + L^2}{2C_n^{\lceil t/2\rceil}} \qquad (13)$$

Cho tới nay đã có nhiều thuật toán mới tấn công vào hệ mật mã dựa trên mã. Mặc dù chưa có thuật toán nào thực sự hiệu quả, song có thể giúp các nhà mật mã học có những lựa chọn các tham số bảo mật một cách phù hợp cho từng mục đích ứng dụng.

Đánh giá độ phức tạp của tấn công giải mã vào mã tổng hệ mật đề xuất:

Với hệ mật dựa trên mã ghép BCH, phía tấn công không biết độ dài của các mã thành phần và khả năng sửa lỗi của chúng (n_i, t_i), không biết mã thành phần nào được sử dụng. Phía tấn công biết được khóa công khai là ma trận kiểm tra H' tham số t là tổng trọng số của từ mã. Nghĩa là chỉ biết các tham số (N,K,t). Do đó, khi áp dụng thuật toán tấn công Canteaut-Chabaud, hay tấn công ngày sinh nhật, có thể áp dụng các công thức đánh giá độ phức tạp tấn công cho một mã.

Hệ mật đề xuất khi sử dụng bộ tham số lựa chọn như trên, khi đó $N = 568$ và $K = 408$. Áp dụng phương pháp giải mã theo chuẩn syndrome cho từng mã thành phần, cho phép mở rộng khả năng sửa lỗi của mã tổng lên đến $t = 41$. Khi đó, độ phức tạp của tấn công giải mã theo thuật toán giải mã *Canteaut-Chabaud* công thức (11) có giá trị đạt tới $WF = 2^{84,2}$ và độ phức tạp của tấn công theo thuật toán *tấn công ngày sinh nhật* công thức (12,13) có độ phức tạp lên tới $WF = 2^{127}$.

Đánh giá độ phức tạp của tấn công giải mã vào các mã thành phần:

Xét độ an toàn của hệ mật dựa trên mã ghép BCH, khi tấn công giải mã vào các mã thành phần. Giả sử trong trường hợp tối nhất kẻ tấn công xác định được các tham số n_i, k_i, t_i từ tham số công khai của hệ mật. Với mỗi mã thành phần, phía tấn công sử dụng tấn công giải mã ISD với độ phức tạp thực hiện là $WF(n_i, k_i, t_i)$. Xác suất chọn được k_i tọa độ không lỗi trong đoạn n_i bit theo tấn công Canteaut-Chabaud, công thức (10) là:

$$p_i = \frac{\left(C_{\lceil k_i/2\rceil}^p\right)^2 . C_{n_i-k_i-\nu_i}^{t_i-2p}}{C_{n_i}^{t_i}} \qquad (14)$$

Biến đổi qua công thức (11) khi n nhỏ giá trị tối ưu $p = 2$, ta có:

$$p_i = \frac{P(k_i)}{WF_i} \qquad (15)$$

trong đó $P(k_i)$ là chi phí thực hiện một lần lặp được xác định:

$$P(k_i) \approx 3C_{[k_i/2]}^2 . \log_2 C_{[k_i/2]}^2 \qquad (16)$$

Do đó xác suất chọn được $K = \sum k_i$ với $i = 1 \div \ell$ tọa độ không lỗi là:

$$p = \prod_{i=1}^{\ell} p_i = \prod_{i=1}^{\ell} \frac{P(k_i)}{WF_i} \qquad (17)$$

Từ đó suy ra độ phức tạp tấn công ISD với hệ mật dựa trên chuỗi mã theo kiểu tấn công vào từng mã thành phần là:

$$\begin{aligned} WF_{ac} &= \frac{P(K)}{p} = P(K)\prod_{i=1}^{\ell}\frac{WF_i}{P(k_i)} \\ &= \prod_{i=1}^{\ell} WF(n_i,\ k_i,\ t_i).P(K).\prod_{i=1}^{\ell}\frac{1}{P(k_i)} \end{aligned} \qquad (18)$$

Với bộ mã gồm 10 mã thành phần lựa chọn trên độ phức tạp của tấn công vào từng mã thành phần theo công thức (18) $WF_{ac} = 2^{87,8}$. Độ phức tạp này cao hơn so với trường hợp tấn công vào mã tổng $WF = 2^{84,2}$.

Xét với bộ mã khác: Giả sử chọn bộ mã gồm 14 mã thành phần với các tham số mã cụ thể như sau: một mã $C_6(32,21)$, bốn mã $C_8(32,16)$, một mã $C_8(64,45)$, ba mã $C_8(128,106)$, năm mã $C_{10}(32,11)$. Khi đó, ta có $N = 768$, $K = 503$, $t = 55$. Sử dụng công thức (18) để đánh giá độ an toàn của hệ mật đối với tấn công vào từng mã thành phần khi sử dụng bộ mã gồm 14 mã kết quả $WF_{ac} = 2^{97,3}$. Trong khi đó độ phức tạp tấn công Canteaut-Chabaud là $WF = 2^{93,5}$ khi tấn công vào hệ mật với tham số tổng.

Như vậy, việc tấn công vào từng mã thành phần có độ phức tạp cao hơn so với tấn công vào mã tổng. Trong cả hai trường hợp, hệ mật đề xuất đảm bảo được độ phức tạp tấn công trên 2^{80}, đáp ứng yêu cầu về độ bảo mật của một hệ mật.

4.3.2. Tấn công cấu trúc

Độ phức tạp của tấn công cấu trúc đối với khóa công khai của sơ đồ hệ mật dựa trên mã ghép BCH đề xuất có thể định lượng bằng cách dò tìm toàn bộ tổ hợp có thể có của ma trận hoán vị $P(N!)$, mã bí mật và ma trận khả nghịch $Q(0,29\times 2^{(N-K)})$.

Giả sử tấn công cho phép xác định được ma trận H và Q, khi đó phía tấn công sẽ tìm được ma trận P. Tiếp theo với mỗi khóa mật phải kiểm tra cho tới khi khóa này là khóa đúng. Đối với sơ đồ hệ mật đề xuất, độ phức tạp của phương pháp tấn công này sẽ tăng theo độ phức tạp của các mã BCH. Bởi vì các mã thành phần: mã BCH, mã BCH mở

rộng, mã thuận nghịch có độ dài khác nhau với các đa thức sinh khác nhau. Ngoài ra, trong hệ mật đề xuất có thể áp dụng hoán vị đối với các mã BCH thành phần để tăng thêm độ phức tạp tấn công.

Để tấn công cấu trúc trong trường hợp thuận lợi nhất là xác định được tham số n_i, k_i của mỗi mã thành phần, từ đó xác định được việc sử dụng các mã thành phần.

Giả sử thay đổi tham số *b* để bí mật ma trận mã BCH thành phần (có khoảng cách cấu trúc d = 5, 7), cho công khai các đa thức sinh của trường GF(2^m), m = 5, 6, 7. Ngoài ra ở đây còn sử dụng các biến thể như mã BCH mở rộng, mã thuận nghịch và mở rộng của nó nên số lượng mã có thể chọn có thể tăng đột biến. Mặt khác tương ứng có 6; 6; 14 đa thức nguyên thủy bậc 5, 6, 7. Các mã được sắp xếp thành chuỗi theo một thứ tự ngẫu nhiên. Vì vậy, số lượng mã thành phần khác nhau là 10668 mã. Khi đó, độ phức tạp tấn công để xác định cấu trúc của 10 mã thành phần có giá trị lên tới 2^{137}.

Bảng 1. So sánh sơ đồ hệ mật dựa trên mã ghép BCH với sơ đồ Niederreiter

Sơ đồ hệ mật	*N*	*K*	*t*	Kích thước khóa (bytes)	Tấn công ISD (bit)
Hệ mật Niederreiter [18]	2048	1751	27	65.006	81
Hệ mật sử dụng mã BCH	568	408	41	11.360	84,3

Bảng 1 thể hiện kết quả so sánh kích thước khóa của hệ mật sử dụng mã ghép BCH đề xuất mới và hệ mật Niederreiter ở cùng một mức an ninh. Trong đó kích thước khoá của hệ mật đề xuất là 11.360 bytes giảm 5,7 lần so với kích thước của hệ mật Niederreiter 65.006 bytes. Hệ mật sử dụng mã ghép BCH đã đề xuất, khi kết hợp sử dụng phương pháp giải mã theo chuẩn syndrome cho phép tăng tỷ lệ mã hóa, nâng cao được số lỗi có thể sửa của các mã thành phần, do đó khắc phục được nhược điểm của hệ mật gốc Niederreiter.

Hệ mật đề xuất an toàn với các tấn công giải mã và tấn công cấu trúc. Độ bảo mật của hệ mật đề xuất được khẳng định thông qua kết quả khảo sát các dạng tấn công điển hình vào hệ mật: Độ phức tạp của tấn công giải mã $2^{84,2}$ và tấn công cấu trúc 2^{137}. Hệ mật đề xuất, mặc dù chi phí cho việc giải mã các mã thành phần (độ phức tạp thực hiện) còn khá lớn $2^{24,6}$; tuy nhiên với ưu điểm kích thước khóa đã được giảm nhỏ và khả năng sửa lỗi của mã được nâng cao, do đó có thể áp dụng hệ mật để xây dựng sơ đồ chữ ký số ứng dụng trong thực tế.

5. KẾT LUẬN

Bài báo đề xuất hệ mật mã dựa trên mã, biến thể mới của hệ mật Niederreiter dựa trên cấu trúc ghép các mã BCH thành phần có độ dài và kích thước khác nhau. Hệ mật đề xuất cho phép giảm được kích thước khóa 5,7 lần so với hệ mật Niederreiter ở cùng một mức an ninh. Bài báo ứng dụng phương pháp chuẩn syndrome giải mã mã BCH đã cho phép tăng tỷ lệ số lượng các syndrome có thể giải mã được trên tổng số các syndrome có thể có. Do đó, nâng cao hiệu quả sửa lỗi của mã BCH và khi áp dụng vào xây dựng hệ mật đã khắc phục được nhược điểm căn bản của hệ mật Niederreiter.

TÀI LIỆU THAM KHẢO

[1]. Mosca M., 2015. "*Cybersecurity in an era with quantum computers: will we be ready?*". The IACR Cryptology ePrint Archive Report 2015/1075.

[2]. McEliece R. J., 1978. *A Public-Key Cryptosystem Based on Algebraic Coding Theory*. The Deep Space Network Progress Report, pp: 114-116.

[3]. Berlekamp E., McEliece R., Tilborg H.v., 1978. *"On the Inherent Intractability of Certain Coding Problems"*. IEEE Transactions on Information Theory, *24*(3), pp: 384-386.

[4]. Bernstein D. J., Lange T., and Peters C., 2008. *Attacking and defending the McEliece cryptosystem.* Post-Quantum Cryptography, Second International Workshop, PQCrypto2008, Cincinnati, OH, USA, pp: 31-46.

[5]. Niederreiter H., 1986. *"Knapsack-type Cryptosystems and Algebraic Coding Theory"*. Problems of Control and Information Theory, *15*(2), pp: 159-166.

[6]. Li Y. X., Deng R. H., and Wang X. M., 1994. *"On the equivalence of McEliece's and Niederreiter's public-key cryptosystems"*. IEEE Transactions on Information Theory, *40*(1), pp: 271-273.

[7]. Courtois N., Finiasz M., Sendrier N., 2001. *How to achieve a mceliece based digital signature scheme.* Advances in Cryptology - ASIACRYPT 2001, Lecture Notes in Computer Science, pp: 157-174.

[8]. Minder L., Shokrollahi A., 2007. *Cryptanalysis of the Sidelnikov Cryptosystem*. Advances in Cryptology - EUROCRYPT 2007, 26th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Barcelona, Spain 2007. Lecture Notes in Computer Science, pp: 347-360.

[9]. Otmani A., Tillich J.-P., and Dallot L., 2010. *"Cryptanalysis of Two McEliece Cryptosystems Based on Quasi-Cyclic Codes"*. Mathematics in Computer Science, Vol 3(2), pp: 129-140.

[10]. Wieschebrink C., 2010. *Cryptanalysis of the Niederreiter public key scheme based on GRS subcodes.* Post-Quantum Cryptography, Third International Workshop, PQCrypto 2010, Darmstadt, Germany, May 25-28, 2010, pp: 61-72.

[11]. Moon T. K., 2005. *"Error correction coding mathematical methods and algorithms"*. John Wiley & Sons Ltd.

[12]. Липницкий В. А., and Конопелько В. К., 2007. *Норменное декодирование помехоустойчивых кодов и алгебраические уравнения*. Минск : Изд. центр БГУ.

[13]. Pham Khac Hoan, Le Van Thai, Vu Son Ha, 2013. *Simultaneous correction of random and burst errors using norm syndrome for BCH codes*. Hội thảo quốc gia REV- KC01 2013, tháng 12/2013, Tr 154-158.

[14]. Canteaut A., and Chabaud F., 1998. *"A new Algorithm for Finding Minimum Weight Words in a Linear Code: Application to McEliece's Cryptosystem and to Narrow-Sense BCH Codes of Length 511"*. IEEE Transactions on Information Theory, 44(1), pp: 367-378.

[15]. Bernstein D. J., Lange T., and Peters C., 2008. *Attacking and defending the McEliece cryptosystem.* Post-Quantum Cryptography, Second International Workshop, PQCrypto2008, Cincinnati, OH, USA, October 17-19, 2008, pp: 31-46.

[16]. Finiasz M., and Sendrier N., 2009. *Security Bounds for the Design of Code-Based Cryptosystems.* Advances in Cryptology ASIACRYPT 2009, Lecture Notes in Computer Science, pp: 88-105.

[17]. Bernstein D. J., Buchmann J., and Dahmen E., 2009. *Post-quantum cryptography*. Springer-Verlag Berlin Heidelberg, pp: 95-145.

[18]. Siim S., 2015., "*Study of McEliece cryptosystem*". The MTAT. 07.022 Research Seminar in Cryptography, Spring 2015.

AUTHOR INFORMATION

Le Van Thai

Hanoi University of Industry