BẢO MẬT VĂN BẢN TIẾNG VIỆT BẰNG MẬT MÃ AFFINE KẾT HỢP VỚI MÃ OTP

Đặng Đình Đức[1,*]
[1]*Trường Đại học Công nghiệp Quảng Ninh*
[*]*Email: dangdinhduc@qui.edu.vn*

TÓM TẮT

Bài báo trình bày phương pháp kết hợp giữa mã hoá Affine và One-Time Pad (OTP) nhằm nâng cao tính bảo mật cho văn bản tiếng Việt có dấu. Do đặc thù bảng chữ cái tiếng Việt bao gồm các ký tự đặc biệt, việc áp dụng các kỹ thuật mã hóa truyền thống cần được điều chỉnh để tương thích. Mô hình đề xuất giúp tăng tính ngẫu nhiên và độ khó trong việc giải mã mà không có khóa tương ứng, đồng thời vẫn đảm bảo tính khả thi trong triển khai thực tế. Trên thực tế, mã hóa Affne là mã dễ sử dụng, tuy nhiên việc chọn các khóa sao cho thỏa mãn điều kiện yêu cầu của thuật toán là một điều rất khó, trong khi OTP là mã chỉ dùng khoá một lần. Bài báo này chúng tôi đề xuất phương pháp kết hợp giữa mã Affine và mã OTP để mã hóa và giải mã dựa trên ngôn ngữ lập trình Python để bảo mật cho văn bản tiếng Việt có dấu.

Từ khóa: *Affne, OTP, python, mức độ bảo mật, mã hóa, giải mã văn bản tiếng Việt.*

1. ĐẶT VẤN ĐỀ

Hiện nay, nhu cầu bảo vệ dữ liệu và thông tin cá nhân được trao đổi trong môi trường internet trở nên vô cùng quan trọng. Đặc biệt, với các văn bản tiếng Việt có dấu, việc mã hóa gặp phải những thách thức nhất định do bộ ký tự phức tạp hơn so với tiếng Anh thông thường. Do đó, việc phát triển và áp dụng các thuật toán mã hóa phù hợp là cần thiết.

Việc chuyển thông tin gốc (bản rõ) thành thông tin mật (ở dạng không đọc được) được gọi là *quá trình mã hoá (encryption)*. Ngược lại, việc biến đổi thông tin mật về dạng thông tin ban đầu (gốc) gọi là *quá trình giải mã (decryption)*.

Mật mã khóa đối xứng (Symmetric Key Cryptography) là một kỹ thuật mã hóa trong đó cùng một khóa được sử dụng cả để mã hóa và giải mã dữ liệu. Điều này nghĩa là cả người gửi và người nhận phải biết và giữ bí mật cùng một khóa để đảm bảo an toàn thông tin.

Xét hai hàm: *E(x)* là hàm biến đổi bản rõ (gốc) thành bản mã hoá và *D(y)* là hàm biến đổi bản mã hoá trở về bản rõ (gốc). Giả sử *x* là văn bản cần mã hóa và *y* là dạng văn bản đã được biến đổi qua việc mã hóa. Khi đó ta ký hiệu:

Mã hóa: y = Ek(x) và giải mã: x = Dk(y)

Trong các hệ mật mã đối xứng thường có 5 thành phần (*P,E,K,C,D*) sau [1]:

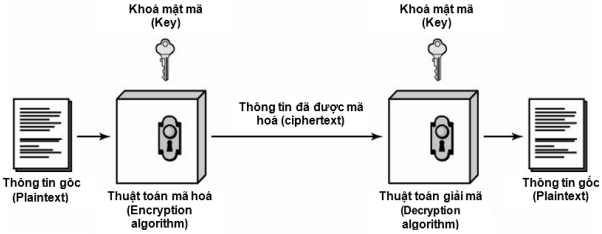
P (Plaintext) là tập hữu hạn các bản rõ và được gọi là không gian bản rõ.

E (Encrytion) là tập hợp các qui tắc mã hóa có thể.

K (Key) là khoá hay còn gọi là không gian khoá. Đối với mỗi phần tử *k* của K được gọi là một khoá (Key). Số lượng khoá phải lớn để không đủ thời gian thám mã;

C (Ciphertext) là các bản mã và được gọi là không gian các bản đã được mã hoá.

D (Decrytion) Thuật toán giải mã, thuật toán này là sự kết hợp giữa thông tin đã mã hóa (ciphertext) cùng với khóa (key) mật mã để đưa ra thông tin gốc (plaintext) ban đầu



Hình 1. Cấu trúc một hệ thống mật mã quy ước

Vì vậy, việc sử dụng một khoá duy nhất cho mã hóa và giải mã là không an toàn. Để đảm bảo an toàn tốt hơn, người ta thường kết hợp

với các hệ mật mã khác nhau.

2. MÃ HOÁ AFFINE VÀ MÃ OTP

2.1. Mã hoá Affne

Mã Affine về cơ bản giống như mã Caesar nhưng nó khác là cần hai khóa thay vì một khóa như mã Caesar [3][4]. Mã affine là một loại mã thay thế, khi mã hoá mỗi chữ cái trong bảng chữ cái được ánh xạ thành số nguyên tương ứng, sau khi sử dụng một hàm toán học thì chúng được chuyển thành một chữ cái khác. Nghĩa là mỗi chữ cái được mã hóa thành một chữ cái khác và ngược lại, mã hóa về cơ bản là một mã thay thế tiêu chuẩn với một quy tắc chi phối chữ cái nào đi với chữ cái nào.

Ở đây, các chữ cái của bảng chữ cái có kích thước m đầu tiên được ánh xạ tới các số nguyên trong phạm vi [0 ... m – 1]. Sau đó, nó sử dụng số học môđun để biến đổi số nguyên mà mỗi chữ cái văn bản thuần túy tương ứng thành một số nguyên khác tương ứng với một chữ cái văn bản mã hóa [2][3]. Hàm mã hóa cho một chữ cái duy nhất là $[E(x) = (ax + b) \bmod m]$. Trong đó môđun m là kích thước của bảng chữ cái, a và b là các khóa của mã hóa. Giá trị a phải được chọn sao cho a và m là số nguyên tố cùng nhau, b là độ lớn của sự dịch chuyển.

Hàm giải mã là $[D(x) = a^{-1}(x - b) \bmod m]$. Trong đó a^{-1} là nghịch đảo nhân môđun của a modulo m. Tức là nó thỏa mãn phương trình $1 = a*a^{-1} \bmod m$.

Từ chương trình trên ta thấy, nghịch đảo nhân của a chỉ tồn tại nếu a và m là số nguyên tố cùng nhau. Do đó, nếu không có hạn chế trên a, giải mã có thể không khả thi. Để chứng minh rằng hàm giải mã là nghịch đảo của hàm mã hóa, ta sử dụng hệ sau [2][4]:

$$
\begin{aligned}
D(E(x)) &= a^{-1}(E(x) - b) \bmod m \\
&= a^{-1}(((ax + b) \bmod m) - b) \bmod m \\
&= a^{-1}(ax + b - b) \bmod m \\
&= a^{-1} ax \bmod m \\
&= x \bmod m
\end{aligned}
$$

2.2. Mã OTP (One Time Pad)

Trong thời đại kỹ thuật số ngày nay, bảo mật thông tin là một yếu tố then chốt trong các hệ thống truyền thông. Một trong những phương pháp mã hóa đơn giản nhưng vô cùng hiệu quả và an toàn là mã hóa One-Time Pad (OTP) – được phát minh bởi Gilbert Vernam năm 1917. OTP được chứng minh là không thể phá vỡ nếu được sử dụng đúng cách, do đó, đây là một chủ đề quan trọng trong lĩnh vực mật mã học

Mã hóa One-Time Pad (OTP) là một trong những phương pháp mã hóa mạnh nhất được biết đến, đảm bảo tính bảo mật tuyệt đối khi được sử dụng đúng cách [5]. Mã OTP dựa trên nguyên lý sử dụng một khóa bí mật có độ dài bằng với thông điệp gốc, và mỗi ký tự trong thông điệp sẽ được mã hóa bằng cách kết hợp với ký tự tương ứng trong khóa.

Mã OTP gồm ba qui trình:

1. Tạo khóa (Key Generation)

Tạo ra một chuỗi khóa ngẫu nhiên có độ dài ít nhất bằng thông điệp gốc. Mỗi ký tự trong khóa được chọn ngẫu nhiên và không phụ thuộc vào bất kỳ thông tin nào. Khóa phải được dùng duy nhất một lần (One-Time).

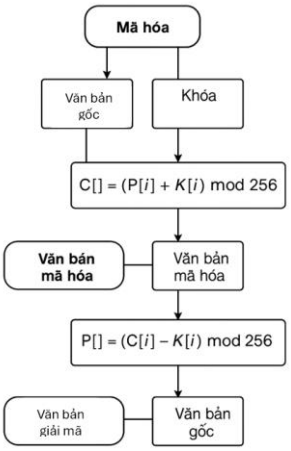
2. Mã hóa (Encryption)

Chia bản rõ P thành các khối có kích thước hàm băm SHA-256 bit. Nếu không chẵn thì phải chèn thêm cho đủ một khối. Chuyển thông điệp gốc (plaintext) thành thông điệp mã hóa (ciphertext). Mỗi ký tự trong văn bản được mã hóa bằng cách kết hợp với ký tự tương ứng trong khóa. Phép toán thường dùng là XOR (nếu làm việc với chuỗi byte) hoặc cộng modulo (nếu dùng bảng chữ cái). Sau đó ghép tất cả các khối bản mã để thu được bản mã.

3. Giải mã (Decryption)

Khôi phục lại văn bản gốc từ văn bản mã hóa và khóa. Dùng khóa giống hệt với lúc mã hóa để đảo ngược phép toán

Đây là phương pháp mã hóa duy nhất đã được chứng minh là không thể phá vỡ về mặt lý thuyết, với điều kiện khóa phải hoàn toàn ngẫu nhiên, có độ dài bằng hoặc lớn hơn thông điệp, và chỉ sử dụng một lần [6][7]. Như vậy, độ dài của khóa OTP phải bằng đúng độ dài của bản rõ và phải là một chuỗi key được sinh ngẫu nhiên, tức là mọi byte của khóa có thể nhận bất kỳ giá trị nào trong khoảng từ 0 đến 255 với xác suất như nhau và độc lập với giá trị của tất cả các byte khóa khác. [8]



Hình 2. Sơ đồ mô tả 3 bước của OTP

Quá trình mã hóa: *Ci = Pi ⊕ Ki, (với i = 1,2,3,...)*. Trong đó *Pi* là ký tự thứ *i* của bản rõ, *Ki* là byte thứ *i* của khóa được sử dụng để mã hóa bản rõ này và *Ci* là ký tự thứ *i* của bản mã kết quả, dấu ⊕ là ký hiệu của phép cộng loại trừ (XOR), phép toán này được dùng trong mã hóa OTP hoặc khi dùng cho mã nhị phân hoặc mã ASCII.

Quá trình giải mã: *Pi = Ci ⊕ Ki*.

Ta thấy, vì khóa One Time Pad là hoàn toàn ngẫu nhiên và không thể dự đoán được, nên có thể rút ra hai kết luận:

- Thứ nhất, xác suất xuất hiện của bất kỳ bit nào trong khóa One Time Pad đều bằng với xác suất xuất hiện của bất kỳ bit nào khác trong khóa đó.

- Thứ hai, việc biết tất cả các giá trị trước đó của khóa trong một chuỗi không cho chúng ta bất kỳ thông tin nào về bit khóa tiếp theo.

Như vậy, để mã OTP an toàn tuyệt đối, cần tuân thủ nghiêm ngặt 4 điều kiện sau:

1. Khóa hoàn toàn ngẫu nhiên (không tạo bằng công thức xác định).

2. Khóa dài bằng hoặc dài hơn bản rõ.

3. Khóa chỉ được sử dụng một lần duy nhất (One-Time).

4. Khóa phải được giữ bí mật tuyệt đối giữa người gửi và người nhận.

3. MÔ HÌNH KẾT HỢP AFFINE VÀ OTP

3.1. Quy trình mã hoá

1. Chuẩn hoá văn bản đầu vào: Loại bỏ ký tự không nằm trong bảng mã, chuyển sang mã số.

2. Mã hoá Affine: Áp dụng công thức Affine với tham số (a, b) và mod m.

3. Sinh OTP: Tạo khoá ngẫu nhiên K có độ dài bằng văn bản đã mã hoá.

4. XOR hoặc cộng modulo OTP: Kết hợp kết quả Affine với khoá OTP theo công thức của OTP.

5. Chuyển ngược thành ký tự mã hoá cuối cùng.

3.2. Quy trình giải mã

1. Trừ OTP: Trừ từng phần tử khóa OTP từ văn bản mã hóa (modulo).

2. Giải mã Affine: Áp dụng công thức giải mã của Affine để thu lại mã số ban đầu.

3. Chuyển mã số sang ký tự

3.3. Đánh giá độ an toàn

Mã Affine Là một mã thay thế đơn tuyến tính, đây là một mã Cổ điển, không đủ an toàn với dữ liệu hiện đại, rất dễ bị phá bằng phân tích tần suất. Không gian khoá nhỏ, với tiếng Anh (m = 26), chỉ khoảng ~312 khoá hợp lệ. cho nên Affine Cipher không an toàn trong môi trường hiện đại. Còn mã OTP (One-Time Pad) mỗi ký tự trong bản rõ được mã hoá bằng một ký tự khoá ngẫu nhiên, chỉ dùng 1 lần, về cơ bản, mã OPT an toàn tuyệt đối nếu dùng đúng.

Tuy nhiên, yêu cầu khoá khó sử dụng, khoá phải dài bằng bản rõ. Việc bị tấn công từ bên ngoài là không thể nếu không có khoá. Việc dùng OTP làm khoá cho Affine có thể làm tăng tính bảo mật cho văn bản. Nhưng nếu dùng OTP để mã hoá tiếp kết quả Affine (dạng 2 lớp), thì lúc này mức bảo mật phụ thuộc hoàn toàn vào OTP. [9].

4. KẾT QUẢ THỰC NGHIỆM

Chương trình thử nghiệm được chạy trên máy tính có cấu hình: CPU Intel Core i3 7100 2.9GHz, RAM DDR4 8GB, HDD 7200 rpm và sử dụng hệ điều hành Windows 10. Các tệp có kích thước, số ký tự khác nhau được chọn ngẫu nghiên. Việc mã hoá được tiến hành 2 lần cho mỗi tệp với mỗi tệp có dung lượng khác nhau sau đó tính thời gian trung bình.

4.1. Thực nghiệm bản rõ tiếng Việt có dung lượng 616 Byte được import từ file. txt với bộ

khóa nhỏ a = 7; b = 9; khoá OTP sinh ngẫu nhiên.

Hình 3. Mã hoá với bộ khoá nhỏ

4.2. Thực nghiệm bản rõ tiếng Việt có dung lượng 4,32 KB được import từ file. txt với bộ khóa trung bình a = 2357; b = 2432; khoá OTP sinh ngẫu nhiên.

Hình 4. Mã hoá với bộ khoá trung bình

4.3. Thực nghiệm bản rõ tiếng Việt có dung lượng 10,8 KB được import từ file. txt với bộ khóa lớn a = 7907; b = 7919; khoá OTP sinh ngẫu nhiên.

Hình 5. Mã hoá với bộ khoá lớn

5. KẾT LUẬN

Trong bài báo này, tác giả đã nghiên cứu mật mã đối xứng, cụ thể là Affine và OTP, để xây dựng được chương trình cho phép mã hoá dữ liệu trước khi chuyển đến người nhận qua mạng internet với khả năng mã hoá dữ liệu đảm bảo theo nguyên tắc bảo mật của mật mã đối xứng. Tác giả cũng đã xây dựng công cụ để thực nghiệm cho các phương án ứng dụng hệ mật mã đối xứng này.

Kết quả so sánh thời gian mã hoá với văn bản .txt được import từ ngoài vào với bộ khoá a, b, OTP khác nhau cho thấy khi kết hợp hai mã đối xứng này thời gian mã hoá giữa các bộ dữ liệu, bộ khoá chênh lệch nhau không đáng kể.

Việc kết hợp mã hóa Affine và One-Time Pad là một hướng đi khả thi trong việc nâng cao bảo mật cho văn bản tiếng Việt. Mặc dù vẫn còn một số thách thức trong thực hiện và triển khai, nhưng mô hình này mở ra nhiều tiềm năng trong nghiên cứu và ứng dụng thực tế, cần được đánh giá cẩn thận dựa trên các yêu cầu cụ thể về bảo mật và tính khả thi trong triển khai.

TÀI LIỆU THAM KHẢO

[1] Ramandeep Sharma, Richa Sharma, Harmanjit Singh, AUG, 2012, Classical Encryption Techniques, International Journal of Computers & Technology, Volume 3. No. 1

[2] https://vi.wikipedia.org/wiki/M%E1%BA%ADt_m%C3%A3_Affine

[3] Alhassan1, M. J; Hassan2, A; Sani3, S;4Alhassan, Y, Volume 7 ~ Issue 10 (2021) pp: 08-12, A Combined Technique of an Affine Cipher and Transposition Cipher. ISSN(Online) : 2394-0743 ISSN (Print): 2394-0735

[4] Mokhtary, M. (2012). Analysis and Design of Affine and Hill Cipher. Journal of Mathematic Research, 4,67-77.

[5]. N.J.Croft and M.S.Olivier (2005). “Using an approximated One-Time Pad to Secure ShortMessaging service(SMS)”. SATNAC 2005 Proceedings.

[6]. Raman Kumar,Roma Jindal, Abhinav Gupta , SagarBhalla, HarshitArora (2011). "*A Secure Authentication System-Using Enhanced One Time Pad Technique".* IJCSNS International Journal of Computer Science and Network Security, VOL.11 No.2,February 2011.

[7] SharadPatil , Ajay Kumar (2010). "Effective Secure Encryption Scheme(One Time Pad) using Complement Approach". International Journal of Computer Science & Communication, Vol.1,No.1,January-June 2010, pp.229-233.

[8]. SharadPatil, ManojDevare, Ajay Kumar (2007). "Modified One Time Pad Data Security Scheme: Random Key Generation Approach". International Journal of Computer Science and Security (IJCSS), Volume (3): Issue(2).

[9] Abdurrahman Ridho, et al 2019, Analysis of Possibility of the Combination of Affine Cipher Algorithm with One Time Pad Cipher Using the Three-Pass Protocol Method in Text Security. J. Phys.: Conf. Ser. 1255 012028

Thông tin của tác giả:

ThS. Đặng Đình Đức
Khoa CNTT, Trường Đại học Công nghiệp Quảng Ninh
Điện thoại: +(84).973.482.666 - Email: dangdinhduc@qui.edu.vn

SECURING VIETNAMESE TEXT USING AFFINE CIPHER COMBINED WITH OTP ENCRYPTION

Information about authors:

Dang Dinh Duc, Master, IT Department, Quang Ninh University of Industry, email: dangdinhduc@qui.edu.vn

ABSTRACT

This paper presents a method that combines Affine encryption and the One-Time Pad (OTP) to enhance the security of Vietnamese text with diacritical marks. Due to the specific characteristics of the Vietnamese alphabet, which includes special characters, the application of traditional encryption techniques requires adjustments to ensure compatibility. The proposed model increases randomness and the difficulty of decryption without the corresponding key, while still maintaining practical feasibility for implementation. In practice, Affine encryption is relatively easy to use; however, selecting keys that satisfy the algorithm's requirements can be quite challenging, whereas OTP uses a key only once. In this paper, we propose a hybrid method that integrates Affine and OTP encryption for encoding and decoding Vietnamese text with diacritical marks, implemented using the Python programming language.

Keywords: *Affine, OTP, Python, security level, encryption, decryption of Vietnamese text*

REFERENCES

[1] Ramandeep Sharma, Richa Sharma, Harmanjit Singh, AUG, 2012, Classical Encryption Techniques, International Journal of Computers & Technology, Volume 3. No. 1

[2] https://vi.wikipedia.org/wiki/M%E1%BA%ADt_m%C3%A3_Affine

[3] Alhassan1, M. J; Hassan2, A; Sani3, S;4Alhassan, Y, Volume 7 ~ Issue 10 (2021) pp: 08-12, A Combined Technique of an Affine Cipher and Transposition Cipher. ISSN(Online) : 2394-0743 ISSN (Print): 2394-0735

[4] Mokhtary, M. (2012). Analysis and Design of Affine and Hill Cipher. Journal of Mathematic Research, 4,67-77.

[5]. N.J.Croft and M.S.Olivier (2005). “Using an approximated One-Time Pad to Secure ShortMessaging service(SMS)”. SATNAC 2005 Proceedings.

[6]. Raman Kumar,Roma Jindal, Abhinav Gupta , SagarBhalla, HarshitArora (2011). "*A Secure Authentication System-Using Enhanced One Time Pad Technique".* IJCSNS International Journal of Computer Science and Network Security, VOL.11 No.2,February 2011.

[7] SharadPatil , Ajay Kumar (2010). "Effective Secure Encryption Scheme(One Time Pad) using Complement Approach". International Journal of Computer Science & Communication, Vol.1,No.1,January-June 2010, pp.229-233.

[8]. SharadPatil, ManojDevare, Ajay Kumar (2007). "Modified One Time Pad Data Security Scheme: Random Key Generation Approach". International Journal of Computer Science and Security (IJCSS), Volume (3): Issue(2).

[9] Abdurrahman Ridho, et al 2019, Analysis of Possibility of the Combination of Affine Cipher Algorithm with One Time Pad Cipher Using the Three-Pass Protocol Method in Text Security. J. Phys.: Conf. Ser. 1255 012028

Ngày nhận bài: 11/04/2025;
Ngày gửi phản biện: 11/04/2025;
Ngày nhận phản biện: 11/04/2025;
Ngày chấp nhận đăng: 13/04/2025.