Chương 5. Mật mã học
Phần 1 – Lịch sử và các khái niệm cơ bản

Người trình bày: NGUYỄN ĐỨC TUẤN

Khoa Công nghệ Thông tin

Hà nội, 25/02/2022

Mật mã học

1. Lịch sử
2. Các khái niệm cơ bản
3. Các thuật toán mật mã cổ điển

1. Lich sử

- Từ xa xưa, để truyền tải thông tin có giá trị một cách bí mật
  - Mã hoá thông tin bằng cách sử dụng các ký hiệu
  - Sử dụng khói với các màu sắc khác nhau
  - Thường được gọi là các mật mã cổ điển

https://www.tutorialspoint.com/cryptography/origin_of_cryptography.htm

1. Lịch sử

- Từ xa xưa, để truyền tải thông tin có giá trị một cách bí mật
  - Hoặc sử dụng các dụng cụ đặc biệt để tạo ra các thông điệp bí mật
  - Có thể đơn giản là thay thế các ký tự trong thông điệp bởi ký tự cách nó k vị trí trong bảng chữ cái (thuật toán Caesar – do Julius Caesar phát minh)
  - Các thuật toán mật mã cổ điển vẫn tồn tại đến ngày nay

https://daily.jstor.org/tales-history-cryptography/

1. Lịch sử

- Khi có sự xuất hiện của các thiết bị tính toán hiện đại
  - có độ an toàn cao hơn, cần nhiều thời gian hơn để có thể phá vỡ
  - như cỗ máy Enigma được người Đức phát triển vào năm 1936 để mã hoá các thông tin quan trọng trong WWII

https://www.theverge.com/2012/4/16/2952810/enigma-machine-alan-turing-centenary-photos

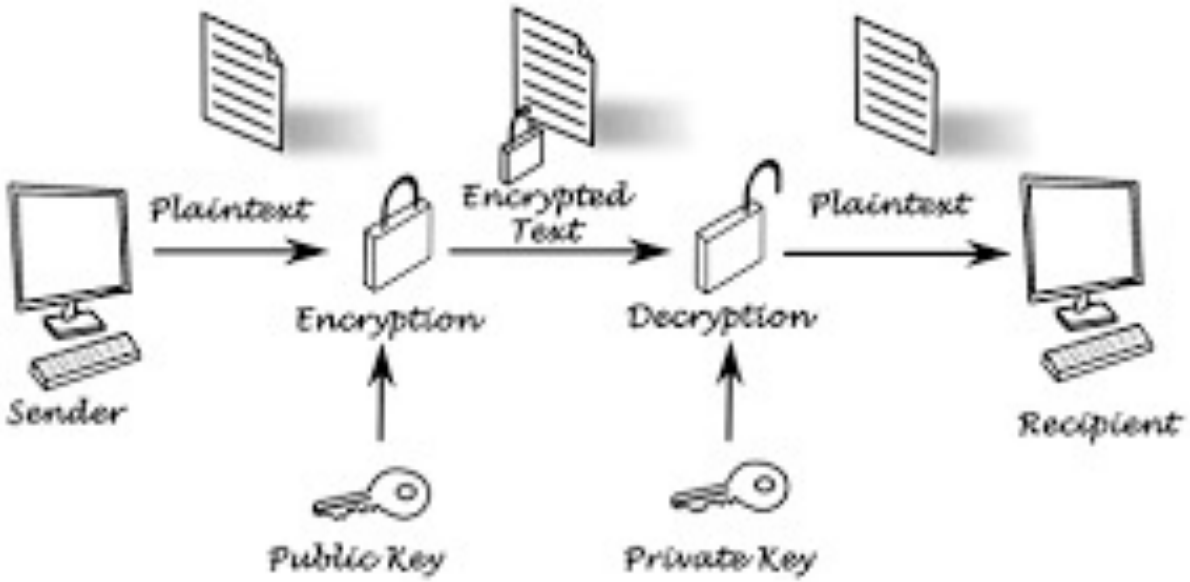
1. Lịch sử

- Khi có sự xuất hiện của máy tính
  - các hàm toán học được sử dụng để biến đổi (xáo trộn) dữ liệu
  - dữ liệu được biến đổi phức tạp, cần nhiều sức mạnh tính toán và thời gian hơn để bẽ gãy các thuật toán này

https://www.theverge.com/2012/4/16/2952810/enigma-machine-alan-turing-centenary-photos

2. Các khái niệm cơ bản

- Mật mã (Encryption – Cryptography)
  - Sự áp dụng các biến đổi dựa trên các thủ tục toán học để chuyển dữ liệu ở dạng đọc được sang dạng không thể đọc hiểu được



2. Các khái niệm cơ bản

- Mật mã (Encryption – Cryptography)
  - Được sử dụng trong các hệ thống để đảm bảo tính bí mật (riêng tư) của thông tin
  - Việc đảm bảo này chỉ có thể duy trì trong một khoảng thời gian nhất định

2. Các khái niệm cơ bản

- Mật mã (Encryption – Cryptography)
  - Thông điệp là đoạn văn bản rõ nghĩa (**plaintext** – đôi khi còn được gọi là **cleartext**) hay **M (message), hoặc văn bản trơn**
  - Bản mã (hay là dữ liệu sau khi được mã hoá) gọi là **Cipher text** (ký hiệu là **C**) hoặc **văn bản mã hoá**
  - Quá trình biến đổi bản rõ trở thành thông điệp vô nghĩa hay khác nghĩa hoàn toàn được gọi là **mã hóa** (**encryption**).

2. Các khái niệm cơ bản

- Mật mã (Encryption – Cryptography)
  - Thông điệp đã được mã hóa gọi là **ciphertext** (bản mã), quá trình biến đổi **ciphertext** trở lại **plaintext** gọi là **giải mã** (**decryption**)
  - Kỹ thuật và khoa học giữ cho thông điệp được an toàn như vậy gọi là mật mã (viết mã – **cryptography**), và nó được thực hiện bởi người viết mã (**cryptographer**).

2. Các khái niệm cơ bản

- Mật mã (Encryption – Cryptography)
  - Thủ tục mã hóa được ký hiệu là *E* hoạt động với dữ liệu là *M*, ký hiệu trong toán học là: *E*(*M*) = *C*.
  - Thủ tục giải mã được ký hiệu là *D* được thực hiện với dữ liệu là *C* tạo ra *M*: $D(C) = M$.

2. Các khái niệm cơ bản

- Mật mã (Encryption – Cryptography)
  - Thuật toán mã hóa (Crytography Algorithms/enciphering algorithm) hay còn được gọi là viết mã (Cipher), là thủ tục toán học được sử dụng để mã hóa và giải mã.
  - Thuật toán giới hạn (Restricted Algorithms): là các thuật toán mà sự an toàn của nó phụ thuộc vào việc giữ bí mật nguyên lý thực hiện.

2. Các khái niệm cơ bản

- Mật mã (Encryption – Cryptography)
  - Khóa (enciphering key) có thể là một con số hay một dãy các ký tự được dùng để mã hóa và giải mã thông điệp.
  - Thuật toán mã hoá đối xứng: khoá để giải mã có thể tính ra được từ khoá được dùng để mã hoá.
  - Sản phẩm mật mã (Crytography Product) bao gồm các hệ thống thiết bị, mô đun, mạch tích hợp và các chương trình phần mềm mã hoá chuyên dụng

2. Các khái niệm cơ bản

- Thuật toán bất đối xứng (khoá công khai)
  - Sử dụng hai khoá khác nhau
  - Mỗi khoá được sử dụng cho mỗi hoạt động tương ứng
  - Không thể tính được khoá này từ khoá kia

2. Các khái niệm cơ bản

- Thám mã (Crytanalysis) / phân tích mật mã
  - là ngành học nghiên cứu các phương thức để tìm được ý nghĩa của các thông điệp đã được mã hoá
  - bằng cách phân tích bản mã, thuật toán mã hoá và các hệ thống mật mã để tìm hiểu cách thức chúng làm việc
    - để nâng cao tính an toàn của các thuật toán
    - hoặc để tìm lại bản rõ ban đầu

2. Các khái niệm cơ bản

- Thám mã (Crytanalysis) / phân tích mật mã
  - phân tích khi biết bản rõ (know plain-text analysis)
    - khi người phân tích có một mẫu văn bản đã giải mã (trước đó được mã bởi một thuật toán cụ thể)
    - sẽ có thể tìm ra được khoá được sử dụng để mã hoá → giải mã các thông điệp khác

2. Các khái niệm cơ bản

- Thám mã (Crytanalysis) / phân tích mật mã
  - phân tích khi biết bản mã (ciphertext only analysis)
    - thực hiện việc phân tích văn bản mã để tìm bản rõ hoặc khoá khi chỉ có các mẫu dữ liệu đã được mã hoá
    - thường được thực hiện trên một tập mẫu được thu thập
    - hoặc lựa chọn các văn bản mã có các đặc trưng

3. Các thuật toán mã hoá cổ điển

- Các thuật toán
  - được thực hiện dựa trên các phép biến đổi đơn giản
- Các thuật toán mật mã dựa trên việc đổi chỗ
- Các thuật toán mật mã dựa trên việc thay thế

3.1 Các thuật toán mật mã dựa trên việc đổi chỗ

- Route Cipher (mã vòng)
- Rail Fence Cipher (mật mã hàng rào)
- Đổi chỗ theo cột (Columnar Transposition)
- Đổi chỗ theo cột kép (Double Columnar Transposition)
- Mật mã dịch chuyển Myszkowski

3.1.1 Route Cipher (mã vòng)

- Nguyên lý – viết mã (mã hoá)
  - Các ký tự trong thông điệp sẽ được điền vào các vị trí trong một khối với kích thước xác định
  - Kích thước của khối được xác định dựa trên số lượng ký tự của thông điệp cần mã hoá
  - hoặc người thực hiện sẽ tự xác định số dòng và số cột của khối
  - các ký tự sẽ được điền theo chiều từ trái – phải, từ trên xuống dưới cho đến khi hết số ký tự
  - đọc các ký tự từ khối theo một trật tự nào đó (được quy ước giữa bên gửi và bên nhận)

3.1.1 Route Cipher (mã vòng)

- Nguyên lý – giải mã
  - Dựa trên cụm từ hoặc hình vẽ gợi ý về trật tự các ký tự được điền vào khối
  - Người nhận sẽ điền các ký tự của bản mã vào khối theo đúng trật tự được gợi ý
  - Sau đó đọc ra theo chiều mà các ký tự được điền vào khi thực hiện mã

3.1.1 Route Cipher (mã vòng)

- Ví dụ – mã hoá
  - M = “Abort the mission, you have been spotted”
  - K (cụm từ gợi ý): bắt đầu từ góc trên bên trái dạng xoắn chôn ốc theo chiều kim đồng hồ”
- C = “ABORT INAET XXDET NVYST HEMOH EOPSE OSIUB“
  - Giải mã: thực hiện tương tự với cụm từ gợi ý

A →	B →	O →	R →	T ↓
T	H	E	M	I ↓
S	S	I	O	N
Y	O	U	H	A
V	E	B →	E	E
N	S	P	O	T
T	E	D	X	X

3.1.2 Rail Fence Cipher (hang rào)

- Nguyên lý – mã hoá
  - Các ký tự trong thông điệp sẽ được điền vào các cột (hàng rào)
  - Mỗi ký tự trong một cột
  - Các ký tự được điền trên *k* dòng (k là khoá)
  - Khi một ký tự nào đó được viết ở dòng k, thì ký tự tiếp theo sẽ được viết ở cột tiếp theo và dòng phía trên
  - Các ký tự trong hàng rào sẽ được đọc ra theo chiều từ trái sang phải, trên xuống dưới

D				**N**				**E**				**T**				**L**
	E		**E**		**D**		**H**		**E**		**S**		**W**		**L**	
		F				**T**				**A**				**A**		

3.1.2 Rail Fence Cipher (mã vòng)

- Ví dụ mã hoá
  - Thông điệp **M** = “DEFEND THE EAST WALL”
  - Khoá **k** = 3

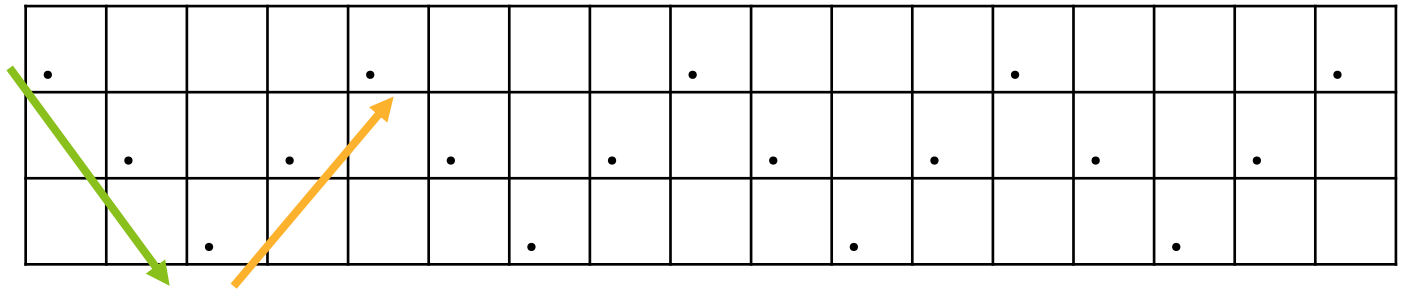
D				**N**				**E**				**T**				**L**
	E		**E**		**D**		**H**		**E**		**S**		**W**		**L**	
		F				**T**				**A**				**A**		

Đọc theo chiều từ trái sang phải, trên xuống dưới, viết thành từng cụm 5 ký tự

C = DNETL EEDHE SWLFT AA

3.1.2 Rail Fence Cipher (mã vòng)

- Ví dụ - giải mã
  - Bản mã ***C*** = "**DNETL EEDHE SWLFT AA**" và khoá ***k*** = 3
  - Tạo bảng với số cột bằng với số ký tự trong ***C*** và có ***k*** dòng
  - Điền các dấu chấm (.) vào các vị trí theo như cách thực hiện mã hoá
    - Xác định được số lượng ký tự trên mỗi dòng
    - Điền các ký tự trong C vào các dấu chấm (.) theo dòng
    - Đọc ra theo chiều được ghi vào (từ hàng 1 – 3, sau đó lại từ 3 lên 1



3.1.3 Các vấn đề với Route và Rail Fence

- Thuật toán dạng quy ước
  - bên gửi và nhận phải thống nhất trước về cách xây dựng khối các ký tự
  - thiếu độ chính xác: phụ thuộc vào kinh nghiệm của người thực hiện
  - an ninh thấp: kẻ tấn công có thể dự đoán cách thực hiện và đảo ngược sau vài lần thử

3.1.4 Đổi chỗ theo cột (Columnar Transposition)

- Nguyên lý
  - Các ký tự sẽ được viết thành m (hàng) x n (cột) dựa trên một từ khoá
  - ***m*** và ***n*** được xác định bởi từ khoá
  - Được viết theo chiều từ trái sang phải, từ trên xuống dưới cho hết số ký tự trong ***M***
  - Giả sử *K* = ZEBRAS thì *n* sẽ là 6

3.1.4 Đổi chỗ theo cột (Columnar Transposition)

- Nguyên lý
  - Nếu cột nào còn trống thì có thể thực hiện:
    - chèn ký tự null (để trống)
    - chèn thêm các ký tự theo nguyên lý: 1 cột trống điền ‘A’, 2 cột trống điền ‘B’ vào cả 2 cột, 3 cột trống điền ‘C’ vào cả 3 cột...
  - Đọc các ký tự ra theo trật tự được xác định bởi từ khoá

3.1.4 Đổi chỗ theo cột (Columnar Transposition)

- Nguyên lý
  - Xác định trật tự của các ký tự trong từ khoá
  - Giả sử K = ZEBRAS
    - dựa trên vị trí của các ký tự so với các ký tự khác trong ký tự theo trật tự trong bảng chữ cái tiếng Anh (alphabet)
    - lần lượt xác định: Z (6), E (3), B(2), R(4), A(1), S(5)

3.1.4 Đổi chỗ theo cột (Columnar Transposition)

- Ví dụ – mã hoá (không chèn thêm ký tự)
  - Cho *M* = WE ARE DISCOVERED. FLEE AT ONCE
  - Đọc ra các ký tự theo trật tự các cột (xác định bởi các ký tự trong từ khoá)

6	3	2	4	1	5
W	E	A	R	E	D
I	S	C	O	V	E
R	E	D	F	L	E
E	A	T	O	N	C
E					

C = EVLNA CDTES EAROF ODEEC WIREE

3.1.4 Đổi chỗ theo cột (Columnar Transposition)

- Ví dụ – mã hoá (chèn thêm ký tự)
  - Cho *M* = “WE ARE DISCOVERED. FLEE AT ONCE”
  - Đọc ra các ký tự theo trật tự các cột (xác định bởi các ký tự trong từ khoá)

C = EVLNE ACDTE ESEAE ROFOE DEECE WIREE

6	3	2	4	1	5
W	E	A	R	E	D
I	S	C	O	V	E
R	E	D	F	L	E
E	A	T	O	N	C
E	E	E	E	E	E

3.1.4 Đổi chỗ theo cột (Columnar Transposition)

- Ví dụ – giải mã (không chèn thêm ký tự)
  - C = EVLNA CDTES EAROF ODEEC WIREE và K = ZEBRAS
  - Sử dụng từ khoá đã cho để xác định trật tự các cột
  - Xác định số lượng ký tự trên mỗi cột
    - Do các ký được điền vào theo trật từ trái – phải, trên xuống dưới nên các cột bên phải sẽ khuyết (nếu số ký tự trong thông điệp không lấp đầy được)
    - Ví dụ: với C đã cho, có 25 ký và có 6 cột, nên sẽ có 5 cột 4 và 1 cột 5 ký tự
    - Dựa trên tính toán như vậy, điền các ký tự trong C vào khối

3.1.4 Đổi chỗ theo cột (Columnar Transposition)

- Ví dụ – giải mã (có chèn thêm ký tự)
  - C = EVLNE ACDTE ESEAE ROFOE DEECE WIREE và K = ZEBRAS
  - Sử dụng từ khoá đã cho để xác định trật tự các cột
  - Viết lần lượt các ký tự trong C vào khối 6 có cột, mỗi cột 5 ký tự
  - Đọc ra theo trật tự như mã hoá, loại bỏ 5 ký tự E (vì có 5 cột khuyết)

3.1.5 Đổi chỗ theo cột kép (Double Columnar Transposition)

- Nguyên lý
  - Được sử dụng bởi người Đức trong thế chiến 1
  - Áp dụng thuật toán đổi chỗ theo cột 2 lần với 2 từ khoá khác nhau
  - C_1 sẽ là M_2 của lần thực hiện thứ hai
  - Mọi thao tác đều thực hiện giống thuật toán đổi chỗ theo cột (Columnar Transposition)

3.1.5 Đổi chỗ theo cột kép (Double Columnar Transposition)

- Ví dụ
  - M = “WE ARE DISCOVERED. FLEE AT ONCE”
  - K_1 = ”ZEBRAS” và K_2 = “STRIDE”

C1 = EVLNA CDTES EAROF ODEEC WIREE

5	6	4	2	3	1
E	V	L	N	A	C
D	T	E	S	E	A
R	O	F	O	D	E
E	C	W	I	R	E
E					

6	3	2	4	1	5
W	E	A	R	E	D
I	S	C	O	V	E
R	E	D	F	L	E
E	A	T	O	N	C
E					

C2 = CAEEN SOIAE DRLEF WEDRE EVTOC

3.1.5 Đổi chỗ theo cột kép (Double Columnar Transposition)

- Với 2 lần mã hoá, các ký tự trong M được hoán đổi và không còn dấu hiệu (các ký tự liền nhau trong M đã được đổi chỗ)
- Giải mã
  - Được thực hiện tương tự như với thuật toán đổi chỗ
  - Lưu ý: áp dụng K_2 trước

C1 = EVLNA CDTES EAROF ODEEC WIREE

C2 = CAEEN SOIAE DRLEF WEDRE EVTOC

→ Nếu ***K*** tồn tại nhiều hơn 2 ký tự giống nhau?

3.1.6 Thuật toán đổi chỗ Myszkowski

- Được giới thiệu bởi Émile Victor Théodore Myszkowski vào năm 1902
- Nguyên lí thực hiện giống với thuật toán đổi chỗ theo cột
  - áp dụng cho từ khoá có các ký tự lặp lại
  - các ký tự giống nhau sẽ được gán cùng giá trị vị trí như nhau
  - khi mã, các ký tự sẽ được đọc đồng thời trên các cột có giá trị vị trí giống nhau theo chiều từ trái sang phải

3.1.6 Thuật toán đổi chỗ Myszkowski

- Ví dụ:
  - *M* = “WE ARE DISCOVERED. FLEE AT ONCE” với *K* = “TOMATO”
  - Từ khoá có 2 ký tự T và O
  - Chuỗi các giá trị vị trí của *K* là: 432143
  - Mã hoá sẽ được thực hiện bằng cách đọc đồng thời các ký tự trên các cột có giá trị vị trí giống nhau (4 và 3)

C = ROFOA CDTED SEEEA CWEIV RLENE

4	3	2	1	4	3
W	E	A	R	E	D
I	S	C	O	V	E
R	E	D	F	L	E
E	A	T	O	N	C
E					

Bài tập

- Bài 1 – Mã hoá thông điệp dưới đây bằng thuật toán Columnar Transposition
  - *M* = “ATTACK AT MIDNIGHT” với *K* = “WORLD”
- Bài 2 – Giải mã thông điệp dưới đây bằng thuật toán Double Columnar Transposition, với K_1 = DESCRIBE, K_2 = COASTLINE
  - *C* = “NDODR WTRFH ASEER AERMR OFLBE OERSA YEAEI HMRAL UTERH MTTYS OSU”

Bài tập

- Bài 3 – Cho bản mã dưới đây được mã bằng thuật toán Rail Fence, hãy giải mã với $k = 3$
  - M = “TBNXMVTL H RW O UPOE H AYDGEOFJ REZO”
- Bài 4 – Cho bản mã được mã hoá bởi thuật toán Route cipher với cụm từ gợi ý “từ góc trên bên phải, ngược chiều kim đồng hồ"
  - C = “MANYD IEECE XXXAT HTETW RPLNI”
- Bài 5:
  - làm thế nào để có thể ứng dụng các thuật toán trên cho tiếng Việt?