BỘ GIÁO DỤC VÀ ĐÀO TẠO

TRƯỜNG ĐẠI HỌC VĂN HIẾN

TÌM HIỂU MẬT MÃ ONE-TIME PAD VÀ VÍ DỤ MINH HỌA

TIỂU LUẬN MÔN HỌC BẢO MẬT & AN TOÀN THÔNG TIN

GVHD: Hồ Văn Ngọc

SVTH: Bùi Minh Nhật

MSSV: 201A290002

Buổi: Thứ sáu, ca 1

TP. HỒ CHÍ MINH, 2021

Ký hiệu

Để đơn giản, tiểu luận sử dụng bảng mã ASCII. Theo đó, mỗi kí tự tương ứng với một số từ 0 đến 255. Các con số này được biểu diễn dưới dạng nhị phân 8-bit. Ví dụ, ký tự `A` là tương đương với `01000001` vì `A` có mã là `65` và $65_{10} = 01000001_2$. Mỗi kí tự được xem là một khối (dãy con) 8-bit. Mỗi bản rõ, khóa, và bản mã được xem là một dãy (mảng) nhị phân có độ dài λ, với λ là bội số của 8.

Ký hiệu	**Ý nghĩa**
M	Bản rõ (thông điệp cần mã hóa, *M*essenger)
K	Khóa (*K*ey)
C	Bản mã (*C*iphertext)
λ	Độ dài của bản rõ, khóa, và bản mã
$X[i]$	Phần tử thứ i của dãy (mảng) X

Mục lục

Chương 0

Mở đầu

If you go to Heaven without being naturally qualified for it, you will not enjoy yourself there.

— George Bernard Shaw

0.1 Mục đích của tiểu luận

Tiểu luận *Tìm hiểu mật mã One-time Pad và ví dụ minh họa* được viết để đáp ứng một phần yêu cầu của học phần An toàn thông tin. Đồng thời, việc thực hiện đề tài là cơ hội để tác giả tìm hiểu cơ sở lý thuyết và công nghệ để xây dựng hệ thống mã hoá.

0.2 Đối tượng và phạm vi tìm hiểu

Tiểu luận tập trung vào mật mã One-time Pad như một ý tưởng chung. Bên cạnh đó là những ví dụ về việc sử dụng OTP trong lịch sử. Để cụ thể hóa ý tưởng OTP cần sử dụng đến những thuật toán mã hóa như Vernam và phương pháp sinh khóa.

0.3 Phương pháp tìm hiểu

Tiểu luận tiếp cận các vấn đề đặt ra theo góc nhìn chặt chẽ và chính xác của toán học.

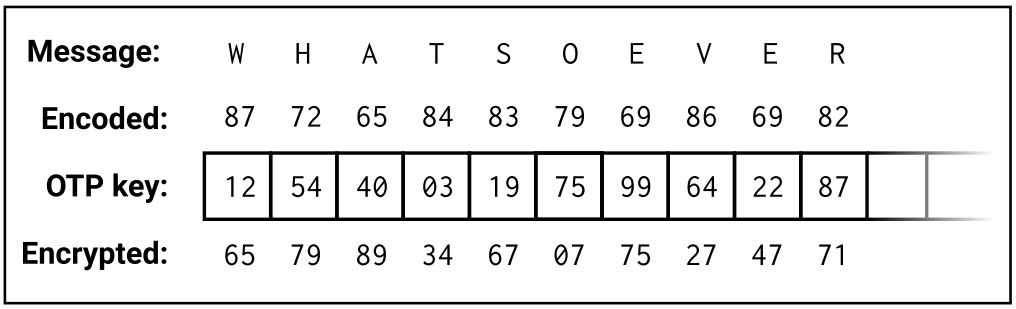
Chương I

Tổng quan về One-time Pad

> ***Calvin:*** *There! I finished our secret code!*
> ***Hobbes:*** *Let's see.*
> ***Calvin:*** *I assigned each letter a totally random number, so the code will be hard to crack. For letter "A", you write 3,004,572,688. "B" is 28,731,569½.*
> ***Hobbes:*** *That's a good code all right.*
>
> — Bill Watterson, "Calvin and Hobbes" (August 23, 1990)

1.1 Ý tưởng

One-time Pad (OTP) về bản chất không phải là một thuật toán. Nó là một tư tưởng. OTP là một khái niệm chỉ các thuật toán mã hóa, với mỗi khi mã hóa thì khóa được sinh ngẫu nhiên đồng bộ và sử dụng đúng một lần.



Hình 1.1.1: One-time Pad sử dụng thuật toán MagicCipher. Thông điệp được encode theo bảng mã ASCII. OTP key được sinh ngẫu nhiên một cách đồng bộ ra từ key stream.

Yêu cầu tối quan trọng đối với mật mã OTP là:

- Độ dài khóa tối thiểu phải bằng độ dài thông điệp.
- Khóa phải được sinh ngẫu nhiên tuân theo phân phối đồng bộ và sinh độc lập với bản rõ.
- Mỗi khóa chỉ được sử dụng đúng một lần.

Nếu khóa thỏa tất cả điều kiện trên, OTP là mã hóa hoàn hảo.

Phân phối đồng bộ (uniform distribution) là phân phối gán xác suất như nhau cho mọi phần tử trong tập hợp khác rỗng và hữu hạn S. Ví dụ, nếu biến ngẫu nhiên X được phân phối một cách đồng bộ qua việc gieo một con xúc sắc, thì $\Pr[X = x] = 1/6$ với mỗi x là một mặt của xúc sắc. Tương tự, việc gieo một đồng xu cân đối tuân theo phân phối đồng bộ sẽ có $\Pr[head] = \Pr[tail] = 1/2$. Không may, nhiều người lầm tưởng rằng đây là định nghĩa của "ngẫu nhiên".

1.2 Sử dụng One-time Pad trong lịch sử

OTP nhiều khi được gọi là "Vernam's cipher" theo tên Gilbert Vernam, một kỹ sư điện báo đã được cấp bằng sáng chế cho ý tưởng này vào năm 1919 [5]. Tuy nhiên, một mô tả trước đó về OTP đã được Frank Miller viết trong một văn bản năm 1882 về mã hóa điện báo [1].

Dưới đây là một số trường hợp OTP được sử dụng:

- OTP được sử dụng để liên lạc bằng giọng nói vô tuyến cấp cao nhất giữa các đồng minh trong Thế chiến II trong một hệ thống có tên Sigsaly [6].
- Đường dây nóng giữa Moscow và Washington được thành lập vào năm 1963 sau cuộc khủng hoảng tên lửa Cuba năm 1962, sử dụng các máy in từ xa. Mỗi quốc gia chuẩn bị các cuộn băng dùng để mã hóa thông điệp của mình và gửi chúng qua đại sứ quán của họ ở quốc gia kia [3].
- Lực lượng Đặc biệt của Quân đội Hoa Kỳ đã sử dụng OTP ở Việt Nam. Bằng cách sử dụng mã Morse với OTP và truyền vô tuyến sóng liên tục (sóng mang mã Morse), họ đã trao đổi được thông tin liên lạc bí mật và tin cậy [2].

1.3 Tính mã hóa hoàn hảo

Mã hóa hoàn hảo là mã hóa mà xác suất kẻ tấn công tính được bản rõ cũng bằng với xác suất kẻ đó đoán. Nói cách khác, kẻ tấn công không thể tính toán được bất kỳ thông tin nào của bản rõ.

Ta nói một hệ mã hóa là hoàn hảo nếu:

$$\Pr[M = m | C = c] = \Pr[M = m] \tag{1.1}$$

cho mọi thông điệp m và mọi bản mã c. Nói cách khác, sự hiểu biết về bản mã không làm thay đổi xác suất mà bản rõ xuất hiện.

Tính chất 1.3.1. *Nếu khóa k được chọn ngẫu nhiên một cách đồng bộ từ tập hợp tất cả khả năng mà khóa đó có thể nhận, thì OTP là mã hóa hoàn hảo.*

Chứng minh: Ta cần chứng minh $\Pr[M = m | C = c] = \Pr[M = m]$ cho mọi cặp m, c.

Giả sử có N khóa, mỗi khóa có xác suất $1/N$. Chúng ta bắt đầu bằng cách chỉ ra rằng mỗi bản mã c cũng có xác suất $1/N$. Nếu ta có c, ta biết rằng nó được mã hóa từ một khóa nào đó. Các lựa chọn khóa khác nhau là các biến cố rời rạc, vì vậy

$$\Pr[C = c] = \sum_k \Pr[C = c \wedge K = k]. \tag{1.2}$$

Ta chọn khóa độc lập với bản rõ, do đó khóa cũng độc lập với bản mã, sử dụng toán tử XOR (mã hóa Vernam), ta thay đổi phương trình (1.2) thành:

$$\begin{aligned} \Pr[C = c | K = k] &= \Pr[M = c \oplus k \wedge K = k] && (1.3)\\ &= \Pr[M = c \oplus k] \cdot \Pr[K = k] && (1.4)\\ &= \Pr[M = c \oplus k] \cdot (1/N). && (1.5) \end{aligned}$$

Ta biết rằng k chạy qua tất cả các khả năng mà khóa có thể nhận, do đó $c \oplus k$ chạy qua tất cả các khả năng mà thông điệp có thể nhận. Điều này nghĩa là:

$$\sum_k \Pr[M = c \oplus k] = 1. \tag{1.6}$$

Kết hợp hai phương trình (1.5) và (1.6) ta được:

$$\Pr[C = c] = \sum_k \Pr(C = c \wedge K = k) = (1/N) \sum_k \Pr[M = c \oplus k] = 1/N. \tag{1.7}$$

Từ định nghĩa của xác suất điều kiện và sự độc lập của K với M:

$$\Pr[M = m | C = c] \cdot \Pr[C = c] = \Pr[C = c \wedge M = m] \tag{1.8}$$
$$= \Pr[K = c \oplus m \wedge M = m] \tag{1.9}$$
$$= \Pr[K = c \oplus m] \cdot \Pr[M = m]. \tag{1.10}$$

Vì $\Pr[C = c] = 1/N = \Pr[K = c \oplus m]$, ta có thể nhân với N để được:

$$\Pr[M = m | C = c] = \Pr[M = m], \tag{1.11}$$

điều cần chứng minh. □

Ta sử dụng mã Vigenère với độ dài khóa bằng với độ dài bản rõ. Giả sử ta nhận được bản mã với nội dung `QWE` trước đó đã được mã hóa bằng một từ khóa gồm ba chữ cái. Nếu không cần bản mã, ta có $1/26^3$ cơ hội đoán đúng. Nhưng kể cả khi có bản mã, cơ hội đúng cũng không khá hơn chút nào. Tất cả các chuỗi gồm 3 kí tự đều có khả năng xảy ra như nhau và sự hiểu biết về bản mã không thay đổi gì cả. Đây là điều làm nên sự hoàn hảo của OTP.

1.4 Hạn chế của One-time Pad

- Độ dài khóa tối thiểu phải bằng độ dài của bản rõ. Chi phí cho việc phân phối cặp bản mã – khóa là 2λ cho một thông điệp cỡ λ bit.
- Mỗi khóa chỉ được sử dụng đúng một lần. Chi phí cho việc phân phối và hủy khóa OTP là cực kỳ tốn kém so với các phương pháp mã hóa khác. Vấn đề phân phối khóa làm OTP trở nên bất khả thi với nhiều ứng dụng thực tế.
- Việc sinh khóa ngẫu nhiên đồng bộ yêu cầu phần cứng chuyên dụng, làm ta khó cài đặt OTP trên nhiều thiết bị hiện đại nhỏ gọn như điện thoại.

Chương II

Cài đặt One-time Pad

The tree which fills the arms grew from the tiniest sprout;
the tower of nine storeys rose from a (small) heap of earth;
the journey of a thousand li commenced with a single step.

— Lao-Tzu, *Tao Te Ching*, chapter 64 (6th century BCE),
translated by James Legge (1891)

2.1 Mã hóa Vernam

Mã hóa Vernam được đặt theo tên của Gilbert Vernam (1890-1960). Vào năm 1917, ông đồng phát minh ra OTP [5]. Ý tưởng của Vernam là trộn mỗi bit của bản rõ với mỗi bit của khóa. Hàm trộn được sử dụng là toán tử XOR. Trong ví dụ của mình, Vernam không viết là "XOR". Ông cho bản rõ là "++−−−", khóa là "+−−++", bản mã là "−+−++". Cách kí hiệu và kết quả của quá trình trộn này là tương đương với bit và XOR. Toán tử XOR hai bit a và b, kí hiệu $\oplus$, được định nghĩa là:

$$a \oplus b \equiv (a + b) \bmod 2. \tag{2.1}$$

Cho thông điệp M và khóa K biểu diễn dưới dạng dãy bit, thuật toán Vernam được trình bày bằng mã giả sau:

```
VernamCipher(M, K)
  for i ← 1 to λ
      C[i] ← M[i] ⊕ K[i]
  return C
```

Mã hóa Vernam là mã hóa hai chiều. Thuật toán VernamCipher có thể dùng vừa để mã hóa và vừa để giải mã với cùng một khóa K. Ta có bản mã $C = \text{VernamCipher}(M, K)$, từ đây có thể dịch ra thông điệp $M = \text{VernamCipher}(C, K)$. Sự đối xứng của thuật

toán này được đảm bảo bằng tính chất của toán tử XOR qua từng bit một.

Tính chất 2.1.1. *Với hai bit a và b thì* $(a \oplus b) \oplus b = a$.

Chứng minh: Gọi a và b là hai bit bất kỳ: $a, b \in \{1, 0\}$. Ta có:

$$\begin{aligned}(a \oplus b) \oplus b &= [(a + b) \bmod 2 + b] \bmod 2 \\ &= [(a + b) \bmod 2 \bmod 2 + b \bmod 2] \bmod 2 \\ &= (a + b + b) \bmod 2 \\ &= (a + 2b) \bmod 2 \\ &= [a \bmod 2 + (2b) \bmod 2] \bmod 2 \\ &= a \bmod 2 = a.\end{aligned}$$

Vậy $(a \oplus b) \oplus b = a$. □

Minh họa thuật toán Vernam

Encrypt:	**Decrypt:**
$M = 0\ 1\ 0\ 0\ 0\ 1\ 0\ 1$	$C = 1\ 0\ 0\ 1\ 0\ 0\ 0\ 1$
$K = 1\ 1\ 0\ 1\ 0\ 1\ 0\ 0$	$K = 1\ 1\ 0\ 1\ 0\ 1\ 0\ 0$
$C = 1\ 0\ 0\ 1\ 0\ 0\ 0\ 1$	$M = 0\ 1\ 0\ 0\ 0\ 1\ 0\ 1$

2.2 Sinh bit của khóa

Tính chất mã hóa hoàn hảo của OTP phụ thuộc vào khóa. Khóa sinh ra phải tuân theo phân phối đồng bộ, tức $\Pr[bit = 1] = \Pr[bit = 0] = 1/2$. Chúng ta có thể sử dụng hàm RANDOM được lập trình sẵn để sinh bit. Song vấn đề là chúng ta không rõ hàm RANDOM có thực sự tuân theo phân phối đồng bộ hay không. Nói cách khác, hàm RANDOM *có thể thiên vị* khả năng xuất hiện bit 0 cao hơn bit 1 hoặc ngược lại. Để thực sự mã hóa hoàn hảo, ta phải tự lập trình một thuật toán sinh bit mà ta *biết rõ* bit sinh ra tuân theo phân phối đồng bộ.

Năm 1951, John von Neumann đã khám phá ra một phương pháp để loại bỏ thiên vị từ việc gieo một đồng xu *lệch* mà không cần quan tâm đến "độ lệch". Ta gieo

đồng xu hai lần. Nếu hai lần gieo mang kết quả khác nhau, thì lấy kết quả đầu tiên; nếu không, thì bỏ đi kết quả và lặp lại thí nghiệm từ đầu [4].

```
NeumannBit()
    a ← Random(0, 1)
    b ← Random(0, 1)
    if a ≠ b
        return a
    else
        return NeumannBit()
```

Tính chất 2.2.1. *Nếu thuật toán* NeumannBit *dừng, bit sinh ra bởi nó được phân phối ngẫu nhiên một cách đồng bộ.*

Để chứng minh tính chất trên, chúng ta cần hai giả định sau:

- Hàm Random, nếu có thiên vị, thì độ thiên vị này luôn giữ nguyên ở mọi lần random.
- Bit `0` và bit `1` đều có khả năng được sinh ra từ hàm Random, hay $0 < \Pr[bit = 0], \Pr[bit = 1] < 1$.

Chứng minh: Gọi xác suất Random ra bit `0` là p, xác suất Random ra bit `1` là $q = 1 - p$. (Lưu ý rằng ta đã giả sử rằng cả bit `0` và bit `1` đều có khả năng xuất hiện, do đó $pq \neq 0$.)

Hai lần Random là độc lập với nhau, nên:

$$\Pr[a = 0 \wedge b = 1] = \Pr[a = 1 \wedge b = 0] = pq. \tag{2.2}$$

Từ phương trình (2.2) nhận thấy

$$\Pr[a \neq b] = \Pr[a = 0 \wedge b = 1] + \Pr[a = 1 \wedge b = 0] = 2pq. \tag{2.3}$$

Khi Random được $a = 0$ và $b = 1$, hoặc $a = 1$ và $b = 0$, thì hiển nhiên $a \neq b$, nên:

$$\Pr[a = 0 \wedge b = 1 \wedge a \neq b] = \Pr[a = 1 \wedge b = 0 \wedge a \neq b] = pq. \tag{2.4}$$

Sử dụng công thức xác suất điều kiện với $a = 0$ và $b = 1$:

$$\Pr[a = 0 \wedge b = 1 \mid a \neq b] = \frac{\Pr[a = 0 \wedge b = 1 \wedge a \neq b]}{\Pr[a \neq b]} = \frac{pq}{2pq} = \frac{1}{2}. \tag{2.5}$$

Tương tự, sử dụng công thức xác suất điều kiện với $a = 1$ và $b = 0$:

$$\Pr[a = 1 \wedge b = 0 \mid a \neq b] = \frac{\Pr[a = 1 \wedge b = 0 \wedge a \neq b]}{\Pr[a \neq b]} = \frac{pq}{2pq} = \frac{1}{2}. \tag{2.6}$$

Cả hai trường hợp NeumannBit trả về bit ($a = 0$ và $b = 1$, hoặc $a = 1$ và $b = 0$), thì bit trả về được phân phối ngẫu nhiên một cách đồng bộ (phương trình (2.5) và (2.6)). Bởi vì các lần Random là độc lập với nhau, ta dùng cùng một phân tích cho mọi lần gọi đệ quy đến NeumannBit. Do đó, nếu bất kỳ lần gọi đệ quy nào trả về một bit, thì bit đó được phân phối đồng bộ. □

2.3 Thuật toán mã hóa One-time Pad

Sử dụng hai thuật toán VernamCipher và NeumannBit, chúng ta xây dựng được một thuật toán OTP dùng để mã hóa thông điệp M bất kỳ:

```
VernamOTP(M)
  for i ← 1 to λ
      K[i] ← NeumannBit()
  C ← VernamCipher(M, K)
  return (C, K)
```

Chúng ta không cần xây dựng key stream. Về bản chất, mỗi bit của khóa sinh ra đều được phân phối đồng bộ, tức không thể đoán trước. Nếu bộ đôi (C, K) được sử dụng đúng một lần, mã hóa của chúng ta là mã hóa hoàn hảo.

Chương III

Kết luận

One-time Pad là ý tưởng mã hóa đơn giản nhưng là an toàn nhất mà con người từng nghĩ đến. Nếu khóa được sinh và sử dụng đúng cách, thì OTP không thể bị thám mã, ngay cả trên phương diện lý thuyết. Tuy nhiên, yêu cầu để đạt được độ bảo mật, như phân phối khóa, là quá tốn kém khiến OTP không phù hợp với thực tiễn hiện nay.

Tài liệu tham khảo

The secret to productivity is getting dead people to do your work for you.

— Robert J. Lang (2009)

[1] Steven M. Bellovin. Frank miller: Inventor of the one-time pad. Cryptologia 35(3):203–222. Taylor & Francis, 2011.

[2] Phan Dương Hiệu and Neal Koblitz. Cryptography during the french and american wars in vietnam. Cryptologia 41(6):491–511. Taylor & Francis, 2017. ⟨https://doi.org/10.1080/01611194.2017.1292825⟩.

[3] David Kahn. The codebreakers. the story of secret writing. p. 715, 1967. New York: Macmillan.

[4] John von Neumann. Various techniques used in connection with random digits. Monte Carlo Method, chapter 13, 36–38, 1951. National Bureau of Standards Applied Mathematics Series 12, US Government Printing Office.

[5] Gilbert Sandford Vernam. Us patent 1,310,719. , 13 September, 1918. ⟨https://patentimages.storage.googleapis.com/5d/ae/f5/1256151a84830e/US1310719.pdf⟩.

[6] Patrick D. Weadon. Sigsaly story. ⟨https://www.nsa.gov/about/cryptologic-heritage/historical-figures-publications/publications/wwii/sigsaly-story/⟩.