Mật mã học (buổi I)

Các khái niệm chung và một số mật mã cổ điển

Mục lục

- Các khái niệm cơ sở
- Các mô hình tấn công
- Một số mật mã cổ điển
  - Mã mono-alphabetic
  - Mã Vigenere
  - Mã One-time-pad

Mục tiêu và nguyên tắc chung

- Đảm bảo tính mật (Confidentiality)
  - Đảm bảo tài sản không bị truy cập trái phép bởi những người không có thẩm quyền
- Đảm bảo tính nguyên vẹn (Integrity)
  - Đảm bảo tài sản không thể bị sửa đổi, làm giả bởi những người không có thẩm quyền
    - Tính nguyên vẹn của dữ liệu (Data integrity)
    - Tính nguyên vẹn của chủ thể (Origin integrity)
- Tính khả dụng (Availability)
  - Đảm bảo tài sản là sẵn sàng để đáp ứng cho người có thẩm quyền

Các công cụ chung

- Mật mã
- Điều khiển bằng phần mềm
- Điều khiển bằng phần cứng
- Chính sách và các thủ tục
- Điều khiển vật lý

Thế nào là Crypto?

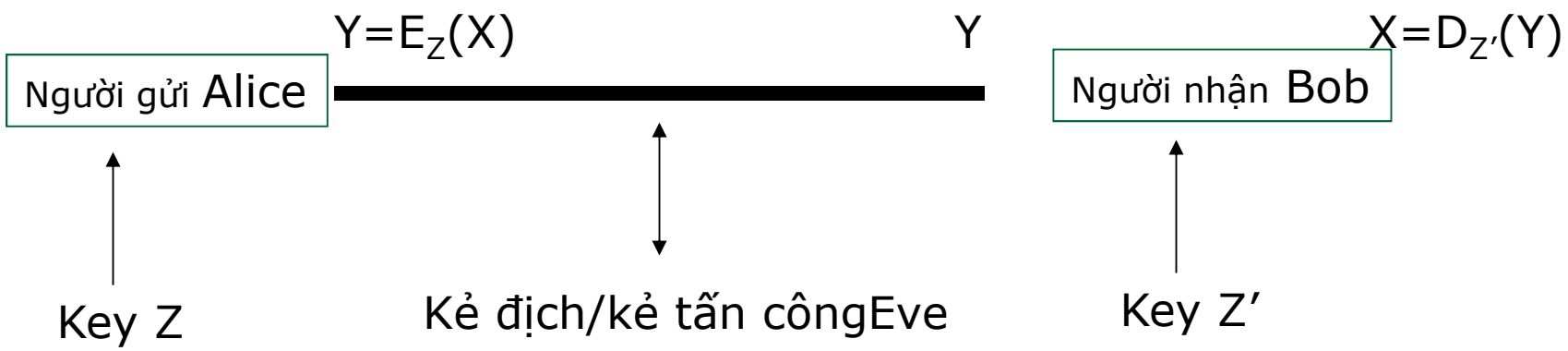
- Xây dựng và phân tích các giao thực mật mã để đạt được các mục tiêu về an toàn thông tin
- Một giao thức (hoặc một cơ chế) là một bộ các thủ tục cho biết các bên tham gia cần phải làm những gì
- Các nhà mật mã học phân tích các giao thức dưới các mô hình tấn công
  - Giả sử khả năng và các hành động có thể của kẻ tấn công
    - Chúng ta cần đứng trên vai trò của kẻ tấn công để suy nghĩ

Các thuật ngữ

- **Cryptography:** là môn học về các kỹ thuật toán học nhằm cung cấp các dịch vụ an toàn thông tin.
- **Cryptanalysis:** là môn học về các kỹ thuật toán học nhằm phá huỷ các dịch vụ an toàn thông tin.
- **Cryptology:** là môn học bao gồm cryptography và cryptanalysis.

Các thuật ngữ

- Plaintexts: bản rõ
- Ciphertexts: bản mã
- Keys: khoá
- Encryption: mã hoá
- Decryption: giải mã



Mật mã khoá bí mật: Secret-key cryptography

- Còn có tên là: mật mã khoá đối xứng -- symmetric cryptography
- Quá trình mã hoá và giải mã dùng cùng một khoá (Z=Z') → *mật mã khoá đối xứng*
- Khoá cần phải giữ bí mật → *mật mã khoá bí mật*
- Vấn đề phân phối khoá: làm sao để chia sẻ khoá giữa A và B

Mật mã khoá công khai: Public-key cryptography

- Còn gọi là: mật mã khoá phi đối xứng -- asymmetric cryptography
- Khoá dùng để mã hoá và giải mã là khác nhau
  - Không thể suy ra khoá giải mã từ khoá dùng để mã hoá và ngược lại
- Tốn chi phí hơn khoá đối xứng

Is it a secure cipher system?

- **Why insecure**
  - **just break it under a certain reasonable attack model (show failures to assure security goals)**
- Why secure:
  - Evaluate/prove that under the considered attack model, security goals are assured
  - Provable security: Formally show that (with mathematical techniques) the system is as secure as a well-known secure one (usually simpler).

Phá mã

- Có rất nhiều kiểu tấn công, phụ thuộc vào:
  - Kiểu thông tin mà kẻ tấn công có thể có
  - Tương tác với máy mã hoá
  - Năng lực tính toán của kẻ tấn công

Phá mã

- **Tấn công chỉ dựa vào bản mã (Ciphertext-only attack)**:
  - Kẻ tấn công chỉ **biết bản mã**
  - Mục tiêu: tìm được bản rõ và khoá
  - Chú ý: một hệ thống bị tấn công bởi kiểu tấn công này thì hoàn hoàn toàn không an toàn
- **Tấn công dựa vào bản rõ (Known-plaintext attack)**:
  - Kẻ tấn công biết một vài bản mã và các bản rõ tương ứng
  - Mục tiêu: tìm ra khoá đã được dùng để mã hoá
    - Hoặc tìm ra cách để giải mã các gói tin dùng cùng khoá với các gói tin đã bắt được

Phá mã…

- **Tấn công bản rõ có chọn lựa (Chosen-plaintext attack)**
  - Kẻ tấn công có thể chọn một số các bản rõ và nhận được các bản mã tương ứng
  - Mục tiêu: suy đoán khoá
- **Tấn công bản mã có chọn lựa (Chosen-ciphertext attack)**
  - Tương tự như trên nhưng kẻ tân công có thể chọn một vài bản mã và nhận được các bản rõ tương ứng.
  - Sự lựa chọn của bản mã có thể thay đổi tuỳ vào bản rõ nhận được trước đó.

Models for Evaluating Security

- **Unconditional (information-theoretic) security**
  - **Assumes that the adversary has unlimited computational resources.**
  - Plaintext and ciphertext modeled by their distribution
  - Analysis is made by using probability theory.
  - For encryption systems: **perfect secrecy**, observation of the ciphertext provides no information to an adversary.

Models for Evaluating Security

- **Provable security:**
  - Prove security properties based on assumptions that it is difficult to solve a well-known and supposedly difficult problem (NP-hard …)
    - E.g.: computation of discrete logarithms, factoring
- **Computational security (practical security)**
  - Measures the amount of computational effort required to defeat a system using the best-known attacks.
  - Sometimes related to the hard problems, but no proof of equivalence is known.

Models for Evaluating Security

- **Ad hoc security (heuristic security):**
  - Variety of convincing arguments that every successful attack requires more resources than the ones available to an attacker.
  - Unforeseen attacks remain a threat.
  - **THIS IS NOT A PROOF**

Mật mã cổ điển

Mã dịch Shift cipher (mã cộng additive cipher)

- Không gian khoá: [1 .. 25]
- Mã hoá với khoá K cho trước:
  - Mỗi ký tự của bản rõ P được mã hoá thành ký tự thứ K sau nó (dịch đi K bước về phía phải):
  - Cách định nghĩa khác: $Y=X \oplus K$ ➔ mã cộng
- Giải mã với khoá K cho trước:
  - Dịch trái

I love you -> L oryh brx

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

Y=(X+K)mod 26
X=(Y-K)mod 26

P = CRYPTOGRAPHYISFUN
K = 11
C = NCJAVZRCLASJTDQFY

Ví dụ

- Tìm bản rõ của bản mã sau
  - A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
  - Djqifs jt csplfo
  - Cipher is broken

Mã dịch: phá mã

- Duyệt toàn bộ
  - Không gian khoá nhỏ (<= 26 khoá).
  - Tìm được K → giải mã dễ dàng

Mã thế một bảng thế
General Mono-alphabetical Substitution Cipher

- Không gian khoá: Toàn bộ hoán vị của bảng chữ cái $\Sigma = \{A, B, C, \ldots, Z\}$
- Mã hoá với khoá π cho trước:
  - Mỗi ký tự X trong bản rõ P được thay thế bởi ký tự $\pi(X)$ tương ứng trong hoán vị π
- Giải mã với khoá π cho trước:
  - Mỗi ký tự Y trong bản mã C được thay thế bởi ký tự $\pi^{-1}(Y)$ tương ứng trong hoán vị π^{-1}

- **Example:**

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

π = B A D C Z H W Y G O Q X S V T R N M S K J I P F E U

BECAUSE → AZDBJSZ

Có vẻ an toàn

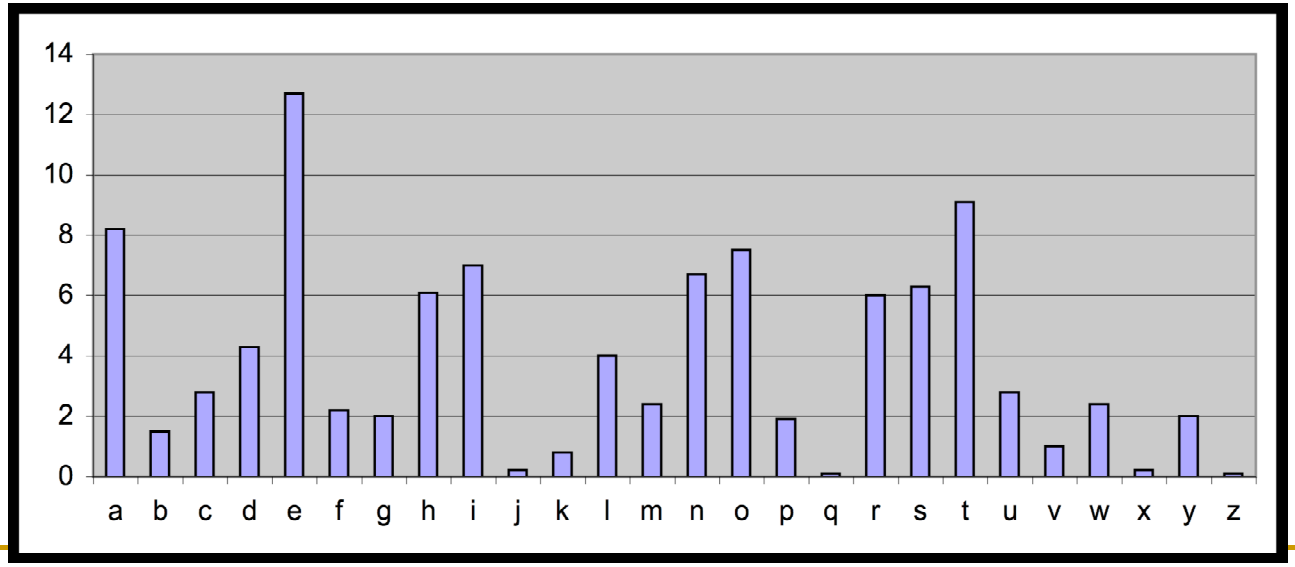
- Phương pháp duyệt toàn bộ là bất khả thi
  - Không gian khoá lớn: $26! \approx 4*10^{26}$
- Được sử dụng phổ biến ở thiên niên kỷ thứ nhất trước công nguyên
- Đã từng được cho là không thể phá được

Phá mã một bảng thế

- Ví dụ
  - Bản mã:uxwk lakkvma xvk naac vuuvbdap vcp uxwk wk kwliya
  - Biết số các ký tự trong bản rõ là:
    - a: 5, b:1, c:1 d:2, e:6, g:1, h:3, i:4, k:1, l:1, m:2, n:2, p:1, s:7, t:4, các ký tự còn lại không xuất hiện
  - Hãy tìm bản rõ ?

Phá mã một bảng thế:
phân tích tần số xuất hiện của các ký tự

- Mỗi ngôn ngữ đều có đặc trưng:
  - Tần số xuất hiện của các ký tự, của một nhóm 2 hay nhiều ký tự.
- Mã thế duy trì đặc trưng trên ➔ có nguy cơ bị tấn công bằng cách phân tích tần số xuất hiện của các ký tự



Mã thế: phá mã

- The number of different ciphertext characters or combinations are counted to determine the frequency of usage.
- The cipher text is examined for patterns, repeated series, and common combinations.
- Replace ciphertext characters with possible plaintext equivalents using known language characteristics.
- Example:

  THIS IS A PROPER SAMPLE FOR ENGLISH TEXT. THE FREQUENCIES OF LETTERS IN THIS SAMPLE IS NOT UNIFORM AND VARY FOR DIFFERENT CHARACTERS. IN GENERAL THE MOST FREQUENT LETTER IS FOLLOWED BY A SECOND GROUP. IF WE TAKE A CLOSER LOOK WE WILL NOTICE THAT FOR BIGRAMS AND TRIGRAMS THE NONUNIFORM IS EVEN MORE.

  - Observations: f_X=1 và f_A=15.

Mã đa bảng thế (polyalphabetic cipher)

- Sử dụng nhiều bảng thế
- Khóa sẽ quyết định thứ tự hòa trộn của các bảng thế này
- Ví dụ:
  - Bảng thế với từ khóa là 1
    - Tin a b c d
    - Mã B D C A
  - Bảng thế với từ khóa là 2
    - Tin a b c d
    - Mã D C A B
  - Khóa: 21
  - Tin: abcdbcda
  - Mã: ?

Mã Vigenere

- Một loại mã đa bảng thế
- **Định nghĩa**:
  - Giả sử m là một số nguyên dương, $P = C = (Z_{26})^n$, và $K = (k_1, k_2, \ldots, k_m)$ là khóa, thế thì:
- **Thuật toán mã hóa**:

$e_k (p_1, p_2 \ldots p_m) = (p_1+k_1, p_2+k_2 \ldots p_m+k_m) \pmod{26}$

- **Thuật toán giải mã**:

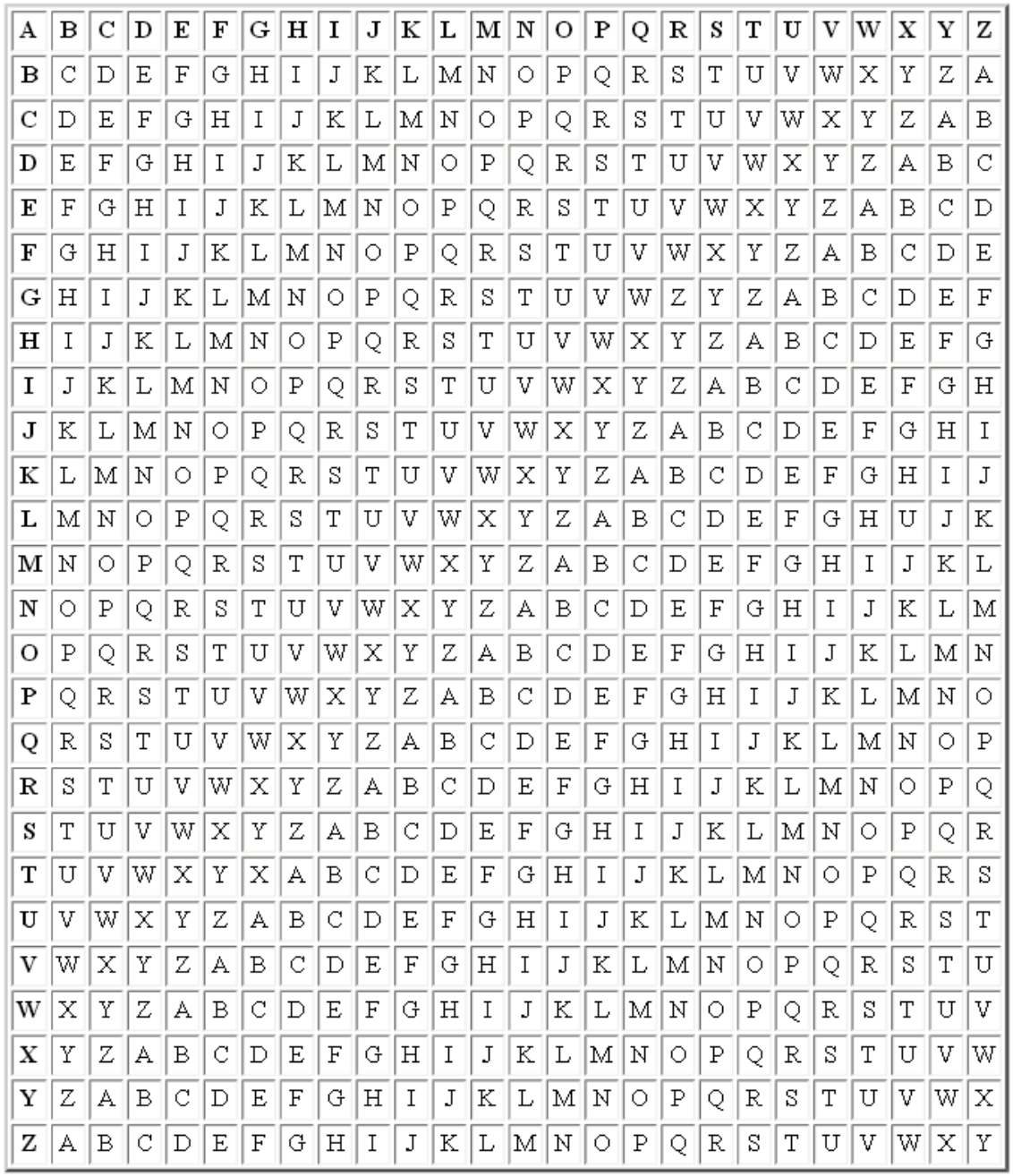
$d_k (c_1, c_2 \ldots c_m) = (c_1-k_1, c_2-k_2 \ldots c_m- k_m) \pmod{26}$

- **Example:**

Tin: C R Y P T O G R A P H Y

Khóa: L U C K L U C K L U C K

Mã: N L A Z E I I B L J J I



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	Z	Y	Z	A	B	C	D	E	F
H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	U	J	K
M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
T	U	V	W	X	Y	X	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y

Ví dụ:

- Khóa: thisisthekey
- Tin: find the cipher of this text
- Mã ?

Mã Vigenere (phá mã)

- Mấu chốt: tìm ra độ dài p của khóa
- Chia bản mã thành p nhóm
  - Với i=0,p-1, nhóm i chứa các ký tự mã ở vị trí kp+i

  ⇒ mỗi nhóm có thể xem như là 1 bản mã với dịch

  ⇒ có thể sử dụng phương pháp thống kê để phá mã theo từng nhóm

Mã Vigenere
Cách tìm độ dài khóa

- Sử dụng chỉ số trùng nhau (Index of coincidence: IC)
  - Xác xuất để hai thành phần ngẫu nhiên của một chuỗi có độ dài n là trùng nhau
  - Định nghĩa
    - $x=x\downarrow 1\ x\downarrow 2\ \ldots x\downarrow n$, thế thì $IC(x)=Pr x\downarrow i=x\downarrow j$ $với\ i\ mọi\ x\downarrow i$, $x\downarrow j$ $chọn\ ngẫu\ nhiên$

Mã Vigenere

Cách tìm độ dài khóa

- Nhận xét về chỉ số IC
  - Tần số xuất hiện của các ký tự càng bằng nhau thì chỉ số IC càng nhỏ
  - Tần số xuất hiện của các ký tự càng lệch nhau thì chỉ số IC càng lớn
  - Chỉ số IC của ngôn ngữ tự nhiên: 0.068
  - Quan hệ của chỉ số IC và p trong thống kê bản mã vigenere
    - p càng lớn ,IC càng nhỏ
    - p=1 → mã dịch, p lớn nhất

Key length (p)	1	2	3	4	5	...	10
IC	0.068	0.052	0.047	0.044	0.043	...	0.041

Mã Vigenere
Cách tìm độ dài khóa

- Tìm p sao cho chỉ số IC của các nhóm là lớn nhất
- Code

1. Set k=1

2. Check if p equals k

2.a. Devide the cipher into k letter groups as before and compute the IC of each.

2.b. If they all are quite the same and approximately equals to 0.068 then p=k

If they are quite different to each other and quite smaller than 0.068 then p>k

3. Increase k by 1 and go back to step 2

Mã Vigenere
Cách tìm độ dài khóa

- Công thích tính IC

$$IC(x) = \frac{\sum_{i=0}^{25} f_i (f_i - 1)}{n(n-1)}$$

Trong đó, f_i là tần số xuất hiện của ký tự alphabet thứ i ở trong x.

Thống kê tần suất xuất hiện của các ký tự trong tiếng anh

- The letters in the English alphabet can be divided into 5 groups of similar frequencies

  I e

  II t,a,o,i,n,s,h,r

  III d,l

  VI c,u,m,w,f,g,y,p,b

  V v,k,j,x,q,z

- Some frequently appearing bigrams or trigrams

  Th, he, in, an, re, ed, on, es, st, en at, to

  The, ing, and, hex, ent, tha, nth, was eth, for, dth.

Ví dụ phá mã

- Pjmu mu b amtjfo rfsr. Mr jbu cffi fiaowtrfg cw rjf uvcurmrvrmqi amtjfo. Wqv bof xfow nvahw. Rjf amtjfo jbu cffi coqhfi.
- The letters in the English alphabet can be divided into 5 groups of similar frequencies

  I e
  II t,a,o,i,n,s,h,r
  III d,l
  VI c,u,m,w,f,g,y,p,b
  V v,k,j,x,q,z
- Some frequently appearing bigrams or trigrams

  Th, he, in, an, re, ed, on, es, st, en at, to

  The, ing, and, hex, ent, tha, nth, was eth, for, dth.

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
B C A G F E D J M K H N L I Q T P O U R V X Z S W Y

- This is a cipher text. It has been encrypted by the substitution cipher. You are very lucky. The cipher has been broken.

One-Time Pad

- Khóa được chọn ngẫu nhiên
- Plaintext $X = (x_1\ x_2\ \ldots\ x_n)$
- Key $K = (k_1\ k_2\ \ldots\ k_n)$
- Ciphertext $Y = (y_1\ y_2\ \ldots\ y_n)$

- $e_k(X) = (x_1+k_1\ \ x_2+k_2\ \ldots\ x_n+k_n) \bmod m$
- $d_k(Y) = (x_1-k_1\ \ x_2-k_2\ \ldots\ x_n-k_n) \bmod m$

One-time pad

- Ví dụ
  - Plaintext space = Ciphtertext space =Keyspace = $\{0,1\}^n$
  - Key is chosen randomly
  - For example:
  - Plaintext is 10001011
  - Key is 00111001
  - Then ciphertext is 10110010

One-Time Pad

- Khóa chỉ được dùng duy nhất một lần
  - Sau khi dùng sẽ bị hủy
- Khóa rất dài, thường có chiều dài bằng độ dài bản tin
  - Phi thực tế