MẬT MÃ HỌC

Phan Quốc Tín – tinpq@uit.edu.vn

Tài liệu tham khảo

- Slide bài giảng
  - Được upload theo từng buổi học.
- Sách tham khảo tiếng Anh
  - *"Introduction to Modern Cryptography"*, Jonathan Katz and Yehuda Lindell.
  - *"An Introduction to Cryptography"*, Richard A. Mollin.
- Sách tham khảo Tiếng Việt
  - *"Mã hóa và Ứng dụng"*, Dương Anh Đức, Trần Minh Triết.
  - *"Mã hóa Thông tin: lý thuyết và ứng dụng"*, Bùi Doãn Khanh, Nguyễn Đình Thúc, Hoàng Đức Hải.

Đánh giá môn học

- Thực hành + bài tập: 30%
- Giữa kỳ: 20%
- Cuối kỳ: 50%

3

Thực hành

- Upload nội dung thực hành chậm nhất 2 ngày trước buổi học.
- Làm bài tập trong 5 tiết.
  - Hình thức đánh giá: vấn đáp.
  - Hoàn thành tại lớp: 100% số điểm được nhận.
  - Không hoàn thành, về nhà làm: 80% số điểm được nhận.
  - Sao chép: 0 điểm

4

Bài tập

- Tại lớp
- Về nhà

5

Thi giữa kỳ, cuối kỳ

- Trắc nghiệm
- Không sử dụng tài liệu

6

Chương 1. GIỚI THIỆU MẬT MÃ HỌC

Phan Quốc Tín – tinpq@uit.edu.vn

Lịch sử

- Kỹ thuật giấu thư (steganography): che giấu sự tồn tại của thư tín
  - Ưu điểm: đơn giản.
  - Khuyết điểm: nội dung bị lộ khi người đưa thư bị khám xét gắt gao hoặc thư tín bị phát hiện.
- Khoa học mật mã (cryptography): che giấu nội dung của thư tín
  - Ưu điểm: bảo vệ được nội dung dù cho thư tín bị phát hiện.
  - Khuyết điểm: phức tạp.

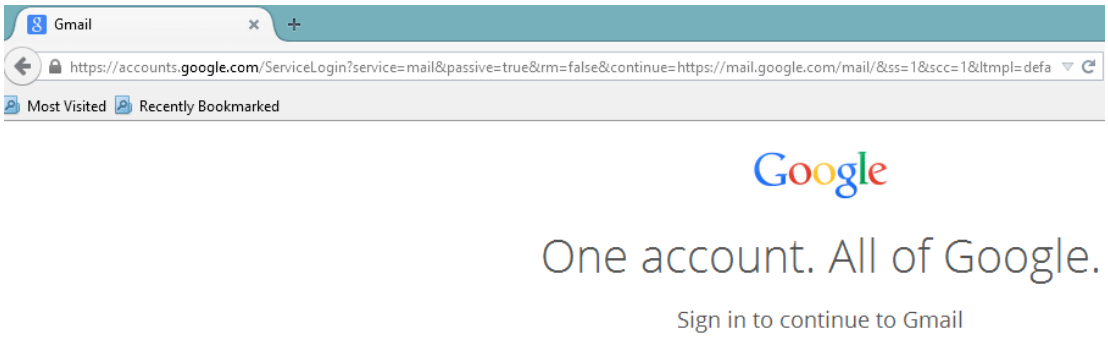
8

Mật mã học (Cryptography)

- **Mật mã học** là một ngành khoa học chuyên nghiên cứu các phương pháp bảo vệ thông tin dựa trên các kỹ thuật toán-tin học.
- **Mã hóa** là một cách thức để chuyển đổi thông tin từ dạng thông thường có thể nhận thức được sang dạng không thể nhận thức được.
- Những điều không đúng về **mật mã học**
  - Là giải pháp cho tất cả các vấn đề liên quan đến an toàn thông tin.
  - Mật mã học đáng tin cậy dù cho được cài đặt và sử dụng không đúng.
  - Là những kỹ thuật chúng ta có thể tự phát minh ra

9

Ứng dụng của mật mã học



- **Https** sử dụng giao thức SSL để gửi password đến server một cách an toàn
- **SSL** là một sản phẩm của mật mã học

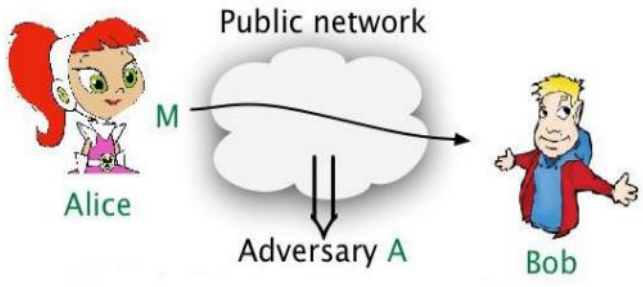
10

Ứng dụng của mật mã học

- Những ứng dụng khác
  - Các máy ATM
  - Ngân hàng điện tử
  - Kết nối an toàn sử dụng SSH
  - Bầu cử
  - …
- Theo thống kê, có 11,748 ứng dụng trên android sử dụng API về mật mã, 10,327 trong số đó sử dụng sai (88%)

11

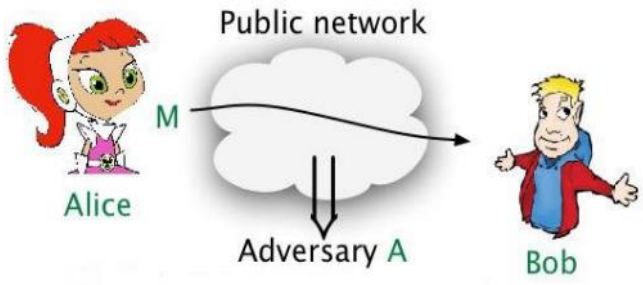
Mục tiêu của mật mã học



- Adversary: một người thông minh và có máy tính cấu hình cao
- Mục tiêu
  - Đảm bảo **tính riêng tư (privacy)** của dữ liệu
  - Đảm bảo **tính toàn vẹn (integrity)** và **tính xác thực (authenticity)**

12

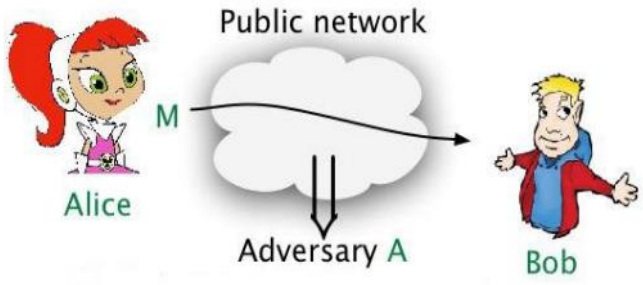
Tính riêng tư



- Mục tiêu là đảm bảo A không thể xem được dữ liệu (thông điệp) M
- Ví dụ: M có thể là thông tin thẻ tín dụng Alice gửi đến server và chúng ta không muốn tin tặc biết được

13

Tính toàn vẹn và xác thực



- Mục tiêu là để đảm bảo
  - M được gửi bởi Alice chứ không phải người khác
  - M không bị thay đổi trong quá trình truyền

14

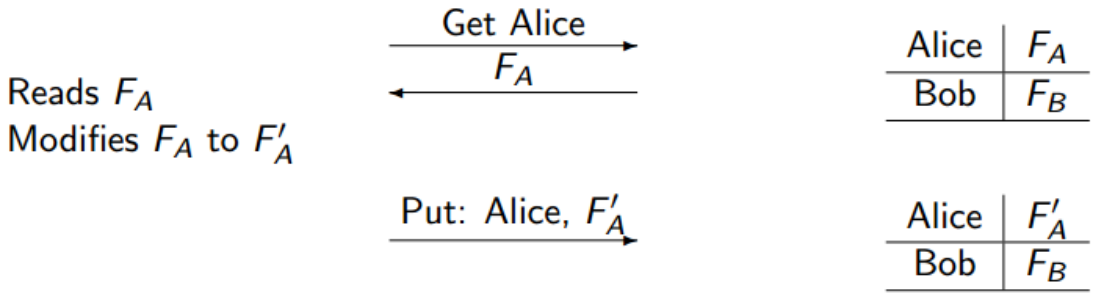
Tính toàn vẹn và xác thực

Alice chuyển 10.000.000
đến Charlie

- Tin tặc Eve có thể
  - Thay đổi "Charlie" thành "Eve"
  - Thay đổi "10.000.000" thành "20.000.000"
- Những hành vi trên ảnh hưởng đến tính toàn vẹn

15

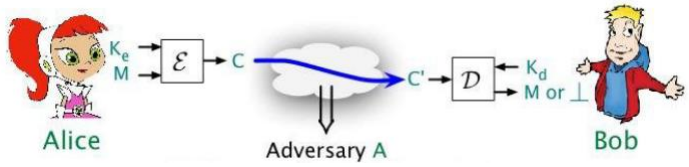
Tính toàn vẹn và xác thực



- F_A và F'_A là thông tin riêng tư, không thể để lộ
- Tính toàn vẹn và xác thực:
  - Bác sĩ phải được xác thực khi lấy dữ liệu của Alice
  - F_A, F'_A không bị thay đổi trong quá trình truyền
  - F_A thật sự được gửi từ CSDL
  - F'_A thật sự được gửi từ bác sĩ

16

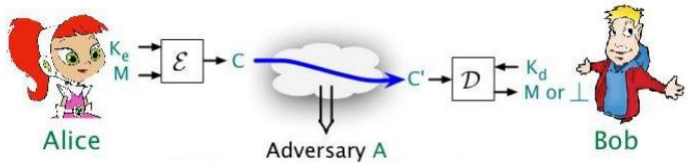
Các hệ mật mã



- E: thuật toán mã hóa (encryption algorithm)
- D: thuật toán giải mã (decryption algorithm)
- K_e: khóa mã hóa (encryption key)
- K_d: khóa giải mã (decryption key)
- M: thông điệp (message)
- C: bản mã (ciphertext)

17

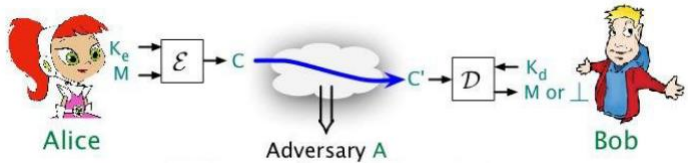
Các hệ mật mã



- Các thuật toán: tiêu chuẩn, đã được cài đặt.
- Thuật toán mã hóa và giải mã là công khai (nguyên tắc Kerchhoffs).
- Khóa công khai (bất đối xứng)
  - K_e: công khai, K_d: bí mật
- Khóa bí mật (đối xứng)
  - $K_e = K_d$: bí mật

18

Các hệ mật mã



- Chúng ta quan tâm đến
  - Làm sao thiết kế các thuật toán để đạt được các mục tiêu an toàn
- Mật mã học khó bởi vì:
  - Người lập mã không thể biết trước được khả năng của tin tặc.
  - "Testing" dường như là không thể.

19

Các phương pháp mã hóa cổ điển

- Mã thay thế (Substitution cipher) / mã Caesar
  - $K_e = K_d = \pi$
  - Ánh xạ: $\Sigma \rightarrow \Sigma$ là bí mật
  - Ví dụ: Σ = {A,B,C,…} và π là bảng thay thế sau

σ	A	B	C	D	⋯
$\pi(\sigma)$	E	A	Z	U	⋯

  - $E_\pi(CAB) = \pi(C)\ \pi(A)\ \pi(B) = Z\ E\ A$
  - $D_\pi(ZEA) = \pi^{-1}(Z)\ \pi^{-1}(E)\ \pi^{-1}(A) = C\ A\ B$

20

Các phương pháp mã hóa cổ điển

- Mã thay thế / mã Caesar
  - Ví dụ:
    - Cho trước K (bảng thay thế)

σ	A	B	C	D	E	F	G	H	I	J	K	L	M
$\pi(\sigma)$	B	U	P	W	I	Z	L	A	F	N	S	G	K
σ	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
$\pi(\sigma)$	D	H	T	J	X	C	M	Y	O	V	E	Q	R

Hãy mã hóa thông điệp: "HELLO I AM A STUDENT"

- Cho trước K như sau: mỗi ký tự sẽ bị dịch đi 3. VD: A sẽ được thay thế bằng D. Hãy mã hóa thông điệp "HELLO I AM A STUDENT"

21

Các phương pháp mã hóa cổ điển

- Mã thay thế / mã Caesar
  - Giả sử có 26 ký tự, không gian khóa của mã Caesar là bao nhiêu?
    - $|K| = 26$
    - $|K| = 26!$
    - $|K| = 2^{26}$
    - $|K| = 26^2$
  - Mã Caesar có thể bị tấn công dựa vào tần suất xuất hiện của các chữ cái
    - VD: "e": 12.7% "t": 9.1% "a": 8.1%

22

Các phương pháp mã hóa cổ điển

- Mã thay thế / mã Caesar
  - Minh họa một tấn công

UKBYBIPOUZBCUFEEBORUKBYBHOBBRFESPVKBWFOFERVNBCVBZPRUBOFERVNBCVBPCYYFVUFOFEIKNWF
RFIKJNUPWRFIPOUNVNIPUBRNCUKBEFWWFDNCHXCYBOHOPYXPUBNCUBOYNRVNIWNCPOJIOFHOPZRVFZ
IXUBORJRUBZRBCHNCBBONCHRJZSFWNVRJRUBZRPCYZPUKBZPUNVPWPCYVFZIXUPUNFCPWRVNBCVBRPY
YNUNFCPWWJUKBYBIPOUZBCUIPOUNVNIPUBRNCHOPYXPUBNCUBOYNRVNIWNCPOJIOFHOPZRNCRVNBC
UNENVVFZIXUNCHPCYVFZIXUPUNFCPWZPUKBZPUNVR

B	36
N	34
U	33
P	32
C	26

→ E
→ T
→ A

NC	11
PU	10
UB	10
UN	9

digrams

→ IN
→ AT

UKB	6
RVN	6
FZI	4

trigrams

→ THE

Các phương pháp mã hóa cổ điển

- Mã Vigenère (thế kỷ 16, Rome)

k = C R Y P T O C R Y P T O C R Y P T
m = W H A T A N I C E D A Y T O D A Y (+ mod 26)

c = Y Y Y I T B K T C S T M V E B O R

Có thể bị phá bằng cách đi tìm chiều dài của từ khóa và suy đoán dựa trên tần suất xuất hiện của ký tự

Các phương pháp mã hóa cổ điển

- Mã Vigenère
  - Cho K = “HELLO”
  - Mã hóa thông điệp “MY NAME IS TAM”

25

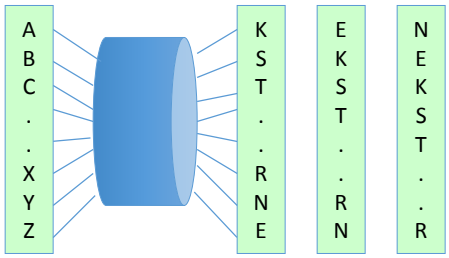
Các phương pháp mã hóa cổ điển

- Mã Vigenère
  - Cho K = “HELLO”
  - Mã hóa thông điệp “MY NAME IS TAM”
  - Có thể sử dụng bảng thay thế

	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
A	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	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	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	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	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	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	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	F
H	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	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	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	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	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	K
M	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	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	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	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	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	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	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	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	S
U	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	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	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	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	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	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

Các phương pháp mã hóa cổ điển

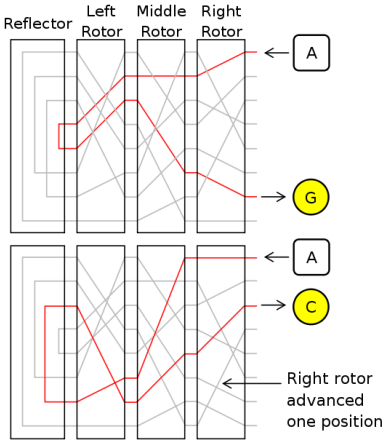
- Các máy Rotor
  - Máy Hebern: rotor đơn



27

Các phương pháp mã hóa cổ điển

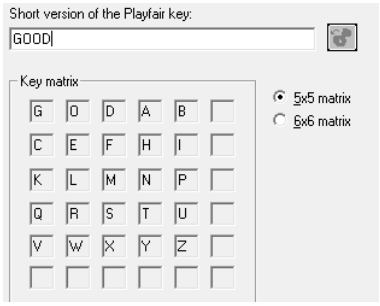
- Các máy Rotor
  - Enigma: 3-5 rotors
  - Mã hóa: gõ các ký tự của bản rõ, đèn tương ứng của bản mã sáng lên.
  - Giải mã: ngược lại.



28

Các phương pháp mã hóa cổ điển

- Mã Playfair
  - Mật mã hóa đa ký tự (mỗi lần mã hóa 2 ký tự liên tiếp nhau)
  - Dựa trên một ma trận các chữ cái 5x5 được xây dựng từ các chữ cái (khóa)
    - Lần lượt thêm từng ký tự của khóa vào ma trận
    - Nếu ma trận chưa đầy, thêm các ký tự còn lại trong bảng chữ cái vào ma trận theo thứ tự A->Z
    - I và J xem như một ký tự



29

Các phương pháp mã hóa cổ điển

- Mã Playfair
  - Mã hóa
    - Mã hóa từng cặp 2 ký tự liên tiếp nhau
    - Nếu 2 ký tự này giống nhau thì thêm ký tự 'x' vào giữa. VD: 'intelligent' sẽ được tách thành 'in' 'te' 'lx' 'li' 'ge' 'nt'
    - Nếu dư 1 ký tự thì thêm ký tự 'q' vào cuối
    - Nếu 2 ký tự nằm cùng dòng thì được thay thế bằng 2 ký tự tương ứng bên phải
    - Nếu 2 ký tự nằm cùng cột thì được thay thế bằng 2 ký tự tương ứng bên dưới
    - Nếu 2 ký tự lập thành hình chữ nhật được thay thế bằng 2 ký tự tương ứng trên cùng dòng ở 2 góc còn lại

30

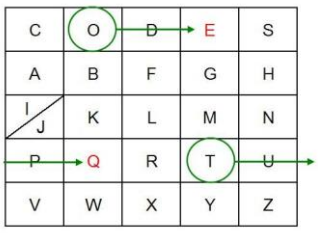
Các phương pháp mã hóa cổ điển

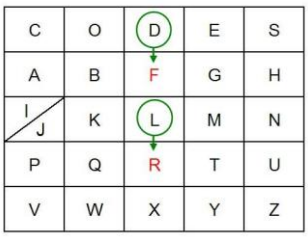
- Mã Playfair
  - Mã hóa
    - Mã hóa từng cặp 2 ký tự liên tiếp nhau
    - Nếu 2 ký tự này giống nhau thì thêm ký tự ‘x’ vào giữa. VD: ‘intelligent’ sẽ được tách thành ‘in’ ‘te’ ‘lx’ ‘li’ ‘ge’ ‘nt’
    - Nếu dư 1 ký tự thì thêm ký tự ‘q’ vào cuối
    - Nếu 2 ký tự nằm cùng dòng thì được thay thế bằng 2 ký tự tương ứng bên phải
    - Nếu 2 ký tự nằm cùng cột thì được thay thế bằng 2 ký tự tương ứng bên dưới
    - Nếu 2 ký tự lập thành hình chữ nhật được thay thế bằng 2 ký tự tương ứng trên cùng dòng ở 2 góc còn lại

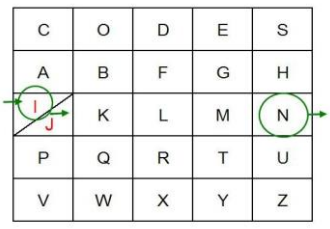
31

Các phương pháp mã hóa cổ điển

- Mã Playfair
  - Mã hóa







32

Các phương pháp mã hóa cổ điển

- Mã Playfair
  - Cho biết khóa là từ khóa "FOOTBALL", hãy tìm ma trận 5x5 và mã hóa thông điệp "MY NAME IS TAM"

33

Mã hóa kỹ thuật số

- 1974: chuẩn mã hóa dữ liệu (Data Encryption Standard - DES)
  - Số lượng khóa: 2^{56}
  - Kích thước khối: 64 bits
- 2001: chuẩn mã hóa tiên tiến (Advanced Encryption Standard - AES)
  - Số lượng khóa: 2^{128} (2^{192}, 2^{256})
  - Kích thước khối: 128 bits.

34

XÁC SUẤT RỜI RẠC

Phan Quốc Tín – tinpq@uit.edu.vn

Nguyên tắc Kerckhoffs

"The system must be practically, if not mathematically, indecipherable" – Kerckhoffs (1883).

Phân bố xác suất

- U: tập hữu hạn (ví dụ: U = $\{0,1\}^n$)
- Định nghĩa: Phân bố xác suất P trên U là một ánh xạ

P: U → [0,1] sao cho $\sum_{x \in U} P(x) = 1$

- Ví dụ:

1. Phân bố đều (Uniform distribution):

$\forall x \in U$: P(x) = 1/|U|

2. Phân bố điểm (Point distribution) tại x_0:

$P(x_0) = 1$, $\forall x \neq x_0$: P(x) = 0

37

Sự kiện

- Cho tập A ⊆ U: Pr[A] = $\sum_{x \in A} P(x) \in [0,1]$.

Ghi chú: Pr[U] = 1

- Tập A được gọi là một sự kiện.
- Ví dụ: U = $\{0,1\}^8$

A = {tất cả x thuộc U sao cho lsb_2 = 11} ⊆ U

Nếu phân bố xác xuất trên U là phân bố đều thì:

Pr[A] = 

38

Tổng xác suất

- Cho 2 sự kiện A_1 và A_2

$$Pr[A_1 \cup A_2] \leq Pr[A_1] + Pr[A_2]$$

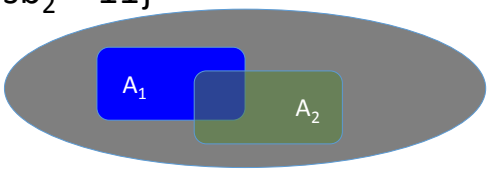
Nếu $A_1 \cap A_2 = \emptyset$ thì $Pr[A_1 \cup A_2] =$ [redacted]

Ví dụ: A_1 = {tất cả x thuộc $\{0,1\}^8$ sao cho $lsb_2 = 11$}

A_2 = {tất cả x thuộc $\{0,1\}^8$ sao cho $msb_2 = 11$}

$Pr[lsb_2(x) = 11$ hoặc $msb_2(x) = 11]$

$= Pr[A_1 \cup A_2] \leq$ [redacted]

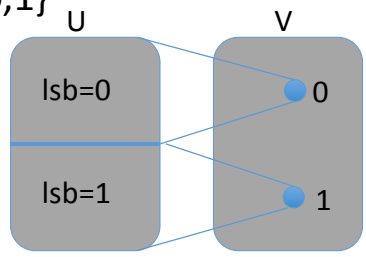


39

Biến ngẫu nhiên

- Định nghĩa: một biến ngẫu nhiên X có thể được mô hình hóa bằng một hàm $X: U \longrightarrow V$

Ví dụ: $X: \{0,1\}^n \longrightarrow \{0,1\}$; $X(y) = lsb(y) \in \{0,1\}$



- Đối với phân bố xác suất đều trên U:

$Pr[X=0] = 1/2$, $Pr[X=1] = 1/2$

Thuật toán ngẫu nhiên

- Thuật toán tất định: $y \longleftarrow A(m)$

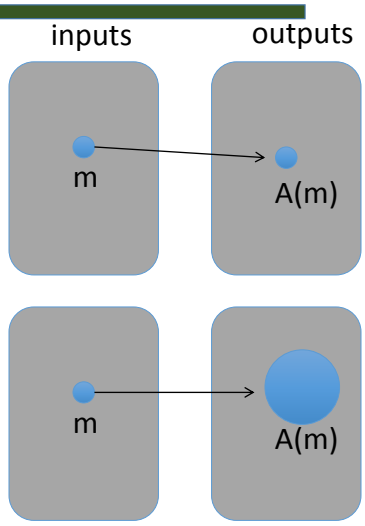
- Thuật toán ngẫu nhiên

  $y \longleftarrow A(m ; r)$ với $r \xleftarrow{R} \{0,1\}^n$

  output là một biến ngẫu nhiên

  $y \xleftarrow{R} A(m)$

Ví dụ: $A(m ; k) = E(k, m)$, $y \xleftarrow{R} A(m)$



Sự độc lập

- Định nghĩa: 2 sự kiện A và B là độc lập nếu

  $Pr[A \text{ and } B] = Pr[A] \cdot Pr[B]$

2 biến X,Y nhận giá trị trong V là độc lập nếu

$\forall a,b \in V$: $Pr[X=a \text{ and } Y=b] = Pr[X=a] \cdot Pr[Y=b]$

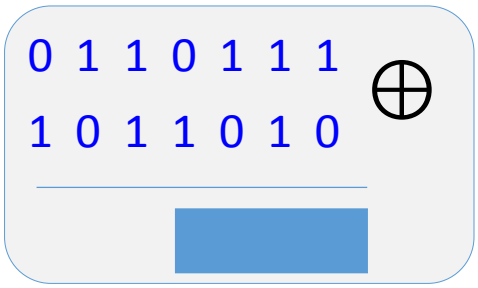
- Ví dụ: $U = \{0,1\}^2 = \{00, 01, 10, 11\}$ và $r \xleftarrow{R} U$

Biến ngẫu nhiên X và Y được định nghĩa: $X = lsb(r)$, $Y = msb(r)$

$Pr[X=0 \text{ and } Y=0] = Pr[r=00] = ¼ = Pr[X=0] \cdot Pr[Y=0]$

Nhắc lại: XOR

XOR của 2 chuỗi thuộc $\{0,1\}^n$ là phép cộng modulo cho 2



Một tính chất quan trọng của XOR

Định lý: Y là một biến ngẫu nhiên trên $\{0,1\}^n$, X là một biến độc lập với Y và phân bố đều trên $\{0,1\}^n$

Thì $Z := Y \oplus X$ phân bố đều trên $\{0,1\}^n$

Proof: (với n=1)

Pr[Z=0] = Pr[(X,Y) = (0,0) hoặc (X,Y) = (1,1)] = Pr[(X,Y) = (0,0)] + Pr[(X,Y) = (1,1)] = $P_0 / 2 + P_1 / 2 = 1/2$

Y	Pr
0	P_0
1	P_1

X	Pr
0	1/2
1	1/2

X	Y	Pr
0	0	$P_0 / 2$
0	1	$P_1 / 2$
1	0	$P_0 / 2$
1	1	$P_1 / 2$

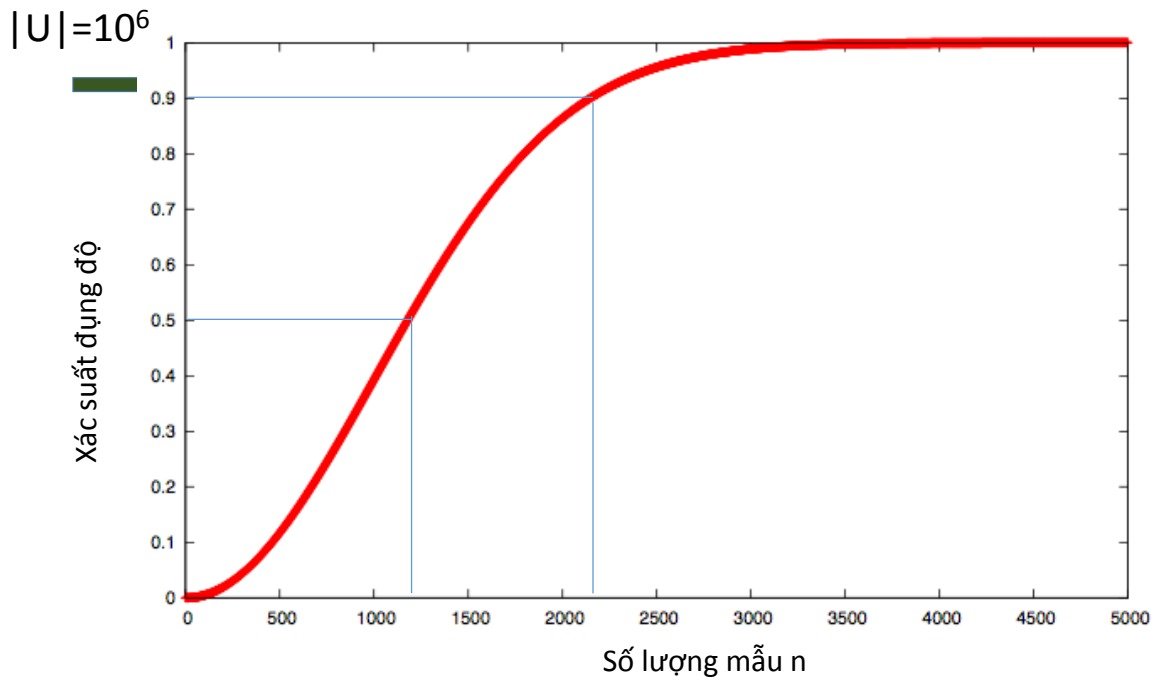
Nghịch lý "ngày sinh nhật"

- Cho $r_1, \ldots, r_n \in U$ là các biến ngẫu nhiên độc lập nhau.

Định lý: khi $n = 1.2 \times |U|^{1/2}$ thì $Pr[\exists i \neq j: r_i = r_j] \geq ½$

Nghịch lý: cho U là tập ngày sinh thì |U| = 365. Khi n = 1.2 x $\sqrt{365}$ ≈ 23.

Theo nguyên lý chuồng bồ câu, xác suất đạt 100% khi số người đạt 367 (vì có 366 ngày sinh khả dĩ, kể cả ngày 29 tháng 2). Tuy nhiên, xác suất 99.9% đạt được chỉ với 70 người và 50% với 23 người.



HẾT CHƯƠNG 1