Bài 2

CƠ SỞ AN TOÀN, BẢO MẬT THÔNG TIN

Giảng viên: Nguyễn Văn Nhân

Điện thoại: 0346542854

Email: nhannv@dainam.edu.vn

NHẮC LẠI BÀI CŨ

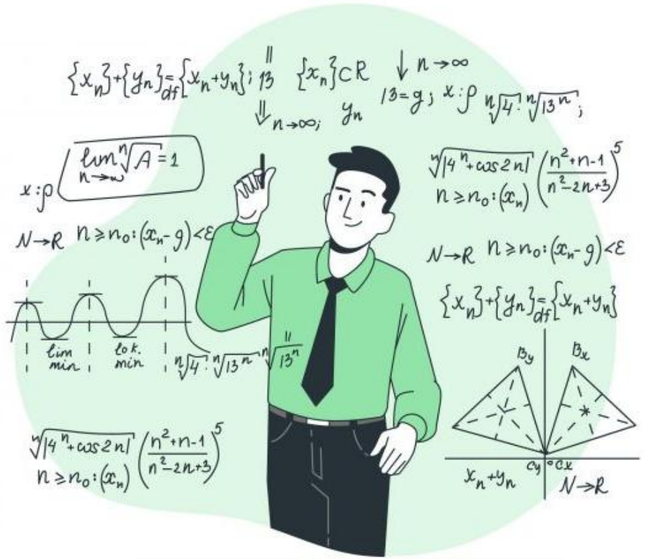
- ✓ **Tổng quan về an toàn, bảo mật thông tin**
- ✓ **Mật mã học**
- ✓ **Một số thuật toán mật mã**

NỘI DUNG

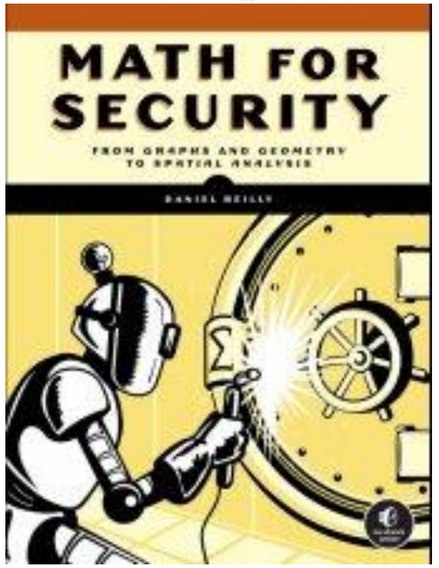
1. Giới thiệu về cơ sở toán học
2. Lý thuyết thông tin
3. Lý thuyết số

GIỚI THIỆU VỀ CƠ SỞ TOÁN HỌC

GIỚI THIỆU VỀ CƠ SỞ TOÁN HỌC



- Sự cần thiết việc nghiên cứu về toán học với mật mã
- Một số vấn đề toán học trong mật mã: Lý thuyết thông tin, lý thuyết số

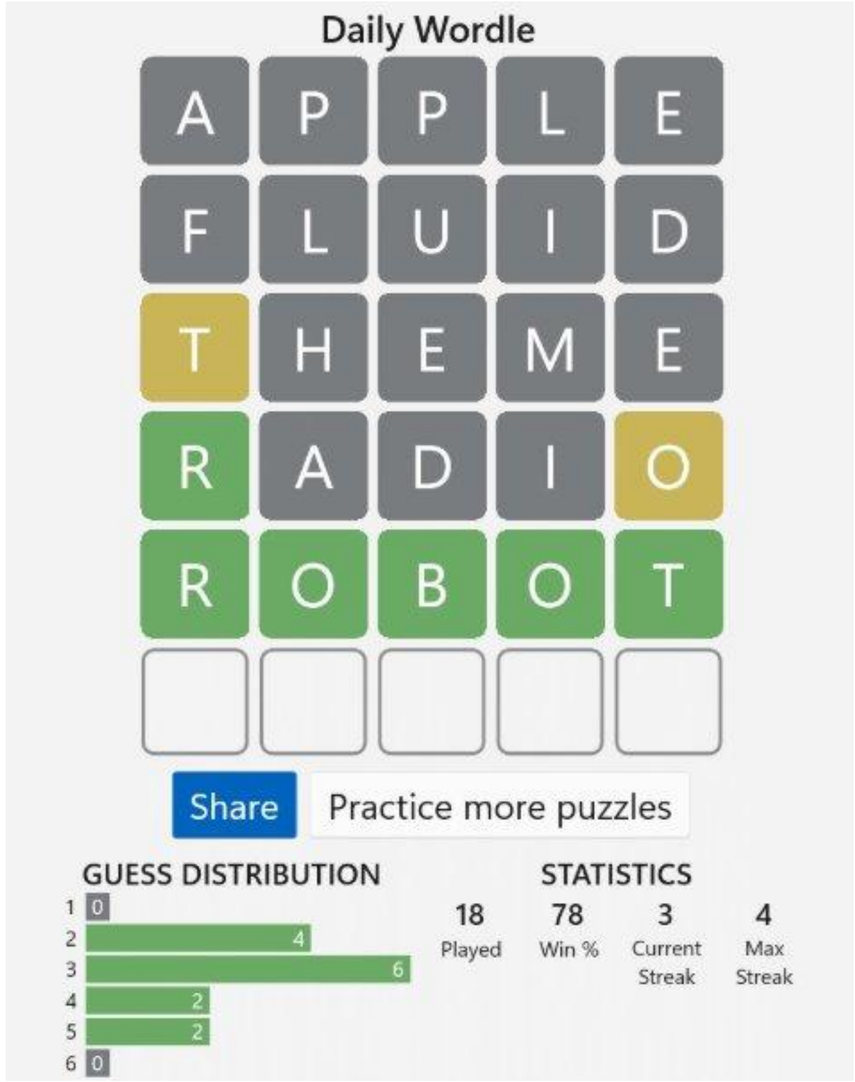


LÝ THUYẾT THÔNG TIN

Trò chơi đoán từ Wordle

Đoán một từ có 5 chữ cái trong 6 lần

- Nếu có chữ đúng và đúng vị trí → **màu xanh**
- Nếu có chữ đúng nhưng sai vị trí → **màu vàng**
- Nếu không có chữ trong từ bí mật → **màu xám**

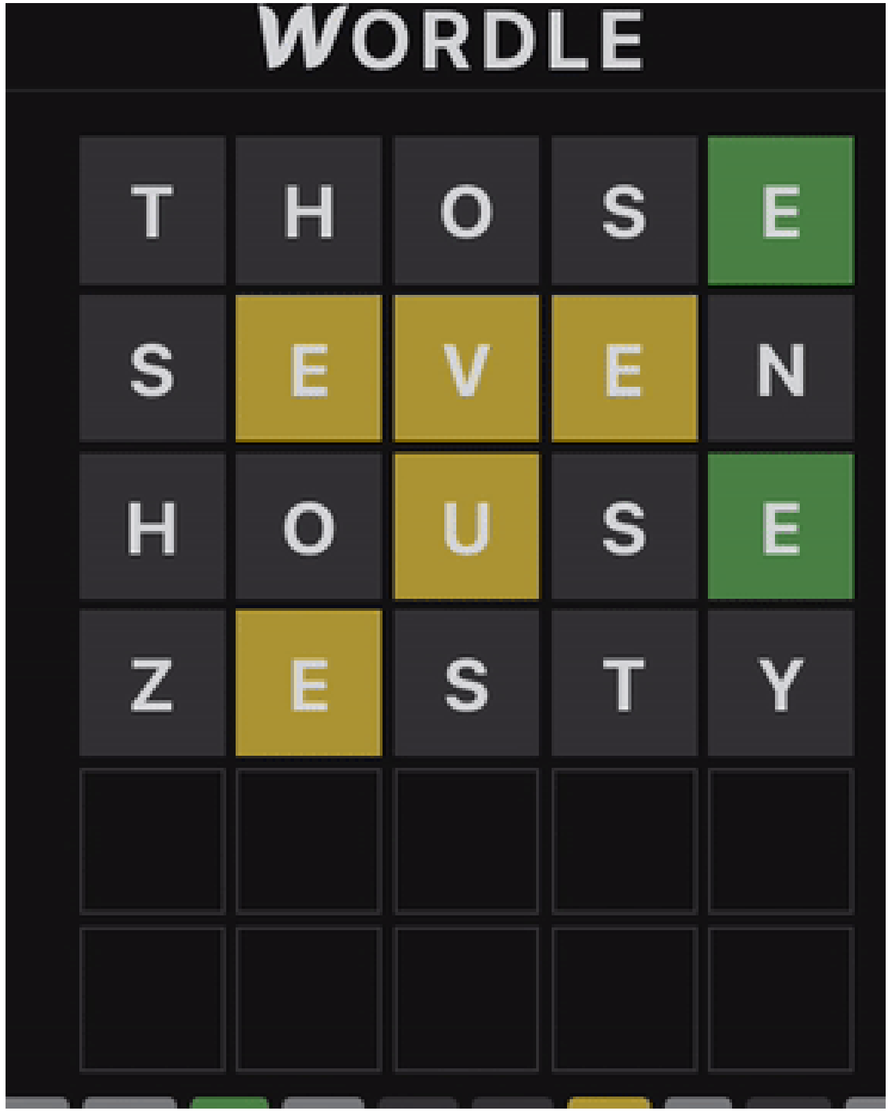


LÝ THUYẾT THÔNG TIN

Trò chơi đoán từ Wordle

- Dần thu hẹp ‘bí mật’ (từ 5 chữ) dựa trên thông tin phản hồi (màu xanh, vàng, xám).
- Mỗi lần đoán giúp giảm bớt độ bất định và thu gọn tập các khả năng.

=> Dựa vào những ‘gợi ý’ (feedback) để tìm ra lời giải chính xác.



Thông tin là gì?

- Thông tin là tất cả những gì đem lại hiểu biết, là nguồn gốc của nhận thức
- Một sự kiện ít xảy ra chứa nhiều thông tin hơn sự kiện thường xảy ra
- Thông tin có thể vô giá trị nếu đã được biết

Định lượng thông tin

- Bảng chữ cái tiếng Anh bao gồm 26 chữ cái, giả sử các chữ cái xuất hiện với tần suất như nhau, thì mỗi chữ chứa lượng thông tin như nhau.
- Những chữ cái xuất hiện thường xuyên hơn chữ cái khác (ví dụ chữ e).

=> Định lượng thông tin bằng cách nào?

Định lượng thông tin

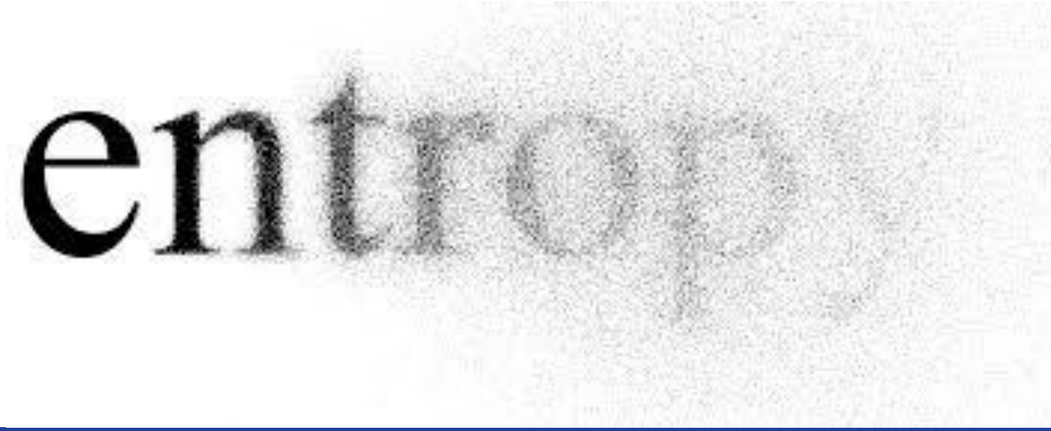
- Lý thuyết thông tin được Claude Elmwood Shannon đưa ra vào năm 1948.
- Lý thuyết thông tin là nghiên cứu toán học về định lượng, lưu trữ và truyền đạt thông tin, được xây dựng dựa trên nền tảng xác suất thống kê.
- Công thức lượng hoá thông tin (lượng tin).

$$I(x) = -log_2 p(x)$$

Entropy

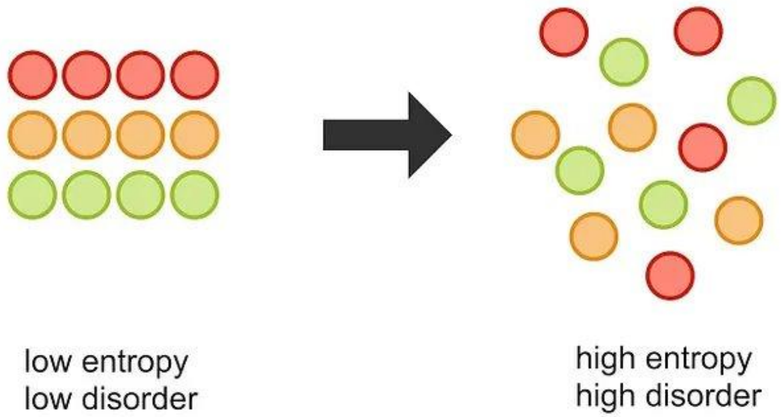
- Claude Shannon đã phát triển khái niệm entropy trong lý thuyết thông tin như là một cách đo lường độ bất định hoặc lượng thông tin trung bình mà một nguồn thông tin có thể tạo ra.
- Entropy trong ngữ cảnh này đo lường lượng thông tin mà một tín hiệu hoặc một chuỗi ký tự có thể mang theo.
- Công thức lượng hoá thông tin (lượng tin):

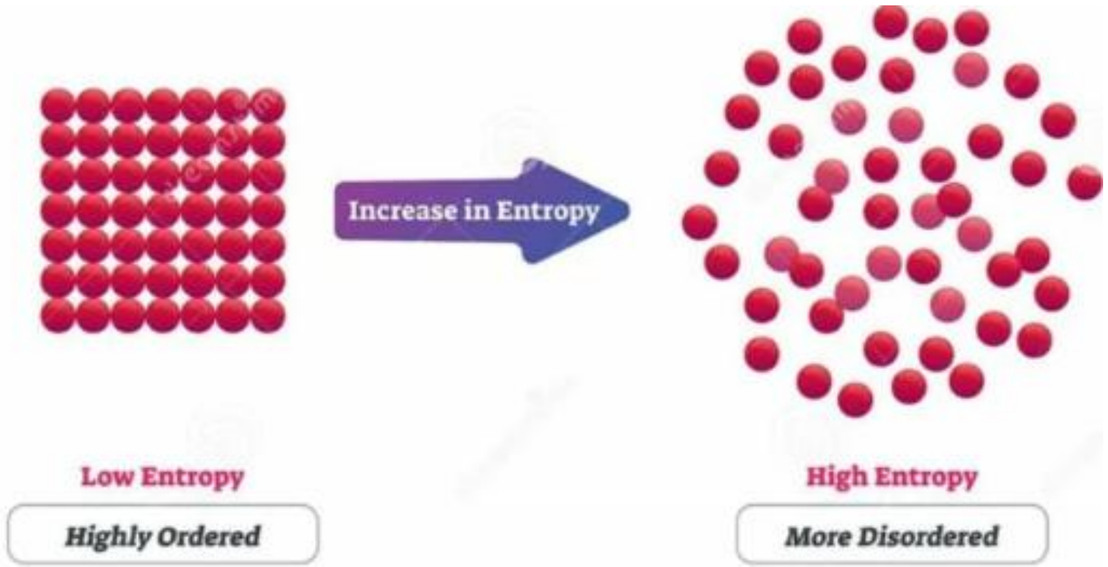
$$H(x) = E[I(x)] = -\sum_{x} p(x) log\ p(x)$$



LÝ THUYẾT THÔNG TIN

- Tại sao không nên sử dụng mật khẩu kiểu “123456”, “111222” ?
- Mật khẩu như “xY9*#@” khó bị hack?





https://timcutting.co.uk/tools/password-entropy

Tốc độ ngôn ngữ

❖ Để diễn đạt cùng một nội dung, người nói tiếng Pháp thường mất nhiều thời gian người nói tiếng Anh.

- Tiếng Anh: “I like this book”
- Tiếng Pháp: “J’aime bien ce livre”

Tại sao?

Tốc độ ngôn ngữ

- Tốc độ ngôn ngữ trong lý thuyết thông tin, còn được gọi là "entropy rate" mô tả lượng thông tin trung bình mà một nguồn thông tin tạo ra trên mỗi ký tự hoặc mỗi đơn vị thời gian.
- Tốc độ ngôn ngữ được tính toán bằng Entropy của toàn bộ chuỗi chia cho số lượng ký tự trong chuỗi.

Tốc độ ngôn ngữ

- Tốc độ ngôn ngữ $H(X)$ được tính theo công thức

$$R(X) = \lim_{x \to \infty} \frac{1}{n} H(X_1, X_2, \dots, X_n)$$

- Trong đó:
  - $H(X_1, X_2, \dots, X_n)$: entropy của chuỗi các biến ngẫu nhiên $X_1, X_2, \dots, X_n$
  - n: số lượng ký tự hoặc đơn vị trong chuỗi

Tốc độ ngôn ngữ

- Tốc độ ngôn ngữ tuyệt đối: Tốc độ ngôn ngữ $R(X)$ được tính như sau:

$$R(X) = log(L)$$

- Độ dư thừa thông tin: Độ dư thừa thông tin được tính như sau:

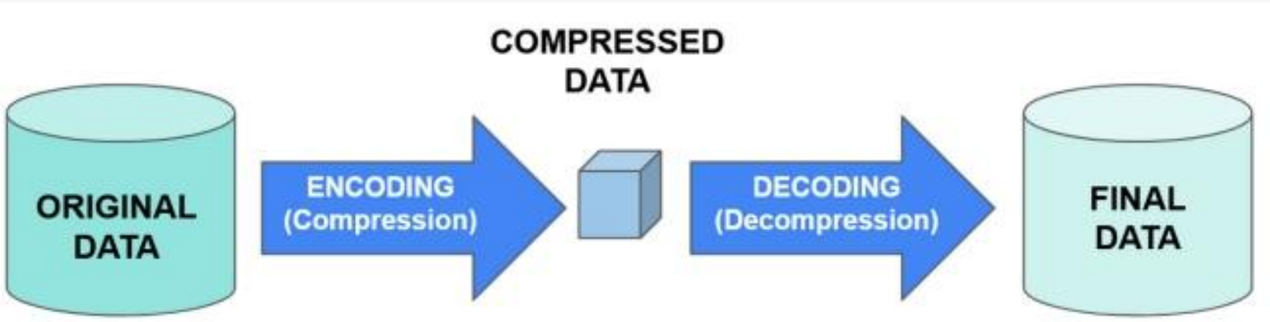
$$D = R(X) - r(X)$$

Ý nghĩa của tốc độ ngôn ngữ

- Tốc độ ngôn ngữ cao: Điều này có nghĩa là chuỗi thông tin chứa nhiều sự bất định hoặc thông tin trên mỗi ký tự, ví dụ như một ngôn ngữ phức tạp với nhiều từ vựng.
- Tốc độ ngôn ngữ thấp: Điều này có nghĩa là chuỗi thông tin chứa ít sự bất định hoặc thông tin trên mỗi ký tự, chẳng hạn như một ngôn ngữ có cấu trúc đơn giản hoặc rất dự đoán được.

Ứng dụng của sự dư thừa thông tin

- **Nén dữ liệu**: Bằng cách loại bỏ hoặc giảm bớt độ dư thừa, các thuật toán nén dữ liệu có thể giảm kích thước của tập dữ liệu mà không làm mất thông tin cần thiết.
- **Phát hiện và sửa lỗi**: Độ dư thừa thông tin cũng có thể được sử dụng để phát hiện và sửa lỗi trong truyền thông tin. Ví dụ, mã sửa lỗi (error-correcting codes) thêm độ dư thừa vào dữ liệu để có thể phát hiện và sửa các lỗi xảy ra trong quá trình truyền tải.



Tính an toàn của hệ thống mật mã

- Tại sao hệ thống mật mã phải đảm bảo độ an toàn?
- Đánh giá tính an toàn của hệ thống mật mã như thế nào?

Tính an toàn của hệ thống mật mã

- Tính an toàn của hệ thống mã hóa là một yếu tố quan trọng trong bảo mật thông tin. Đây là khả năng của hệ thống mã hóa trong việc bảo vệ dữ liệu khỏi sự truy cập trái phép, đảm bảo rằng chỉ những người có quyền mới có thể giải mã và truy cập thông tin gốc.
- Tính an toàn của một hệ thống mã hóa được đánh giá dựa trên nhiều yếu tố khác nhau

Tính an toàn của hệ thống mật mã

- Khoảng cách duy nhất (unicity distance): là số lượng tối thiểu các bản mã cần thiết, để có thể tiến hành thám mã bằng cách thử tất cả các khoá có thể (brute-force attack) thành công.

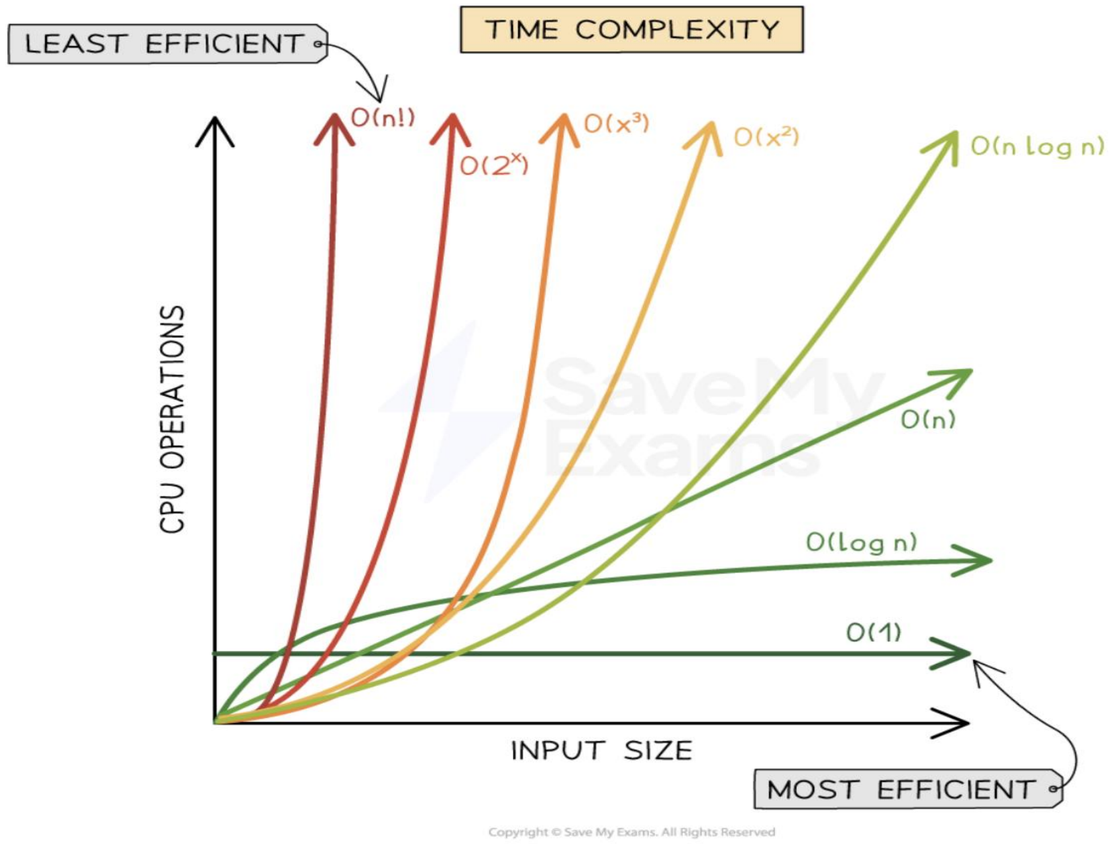
$$U = \frac{H(k)}{D}$$

- Trong đó:
  - $H(k)$: entropy của khoá
  - D: mức độ dư thừa mỗi ký tự của ngôn ngữ

Claude E.Shannon “Communication Theory of Secrecy Systems”, Bell System Technical Journal, vol.28-4, page 656—715, Oct. 1949.

Tính an toàn của hệ thống mật mã

- Độ phức tạp của thuật toán mã hóa: Độ dài khóa, độ phức tạp thuật toán
- Tính bảo mật của khóa mã hóa: Quản lý khóa, Bảo mật của khóa
- Tính ngẫu nhiên và không đoán trước được: Khóa ngẫu nhiên, Dữ liệu ngẫu nhiên

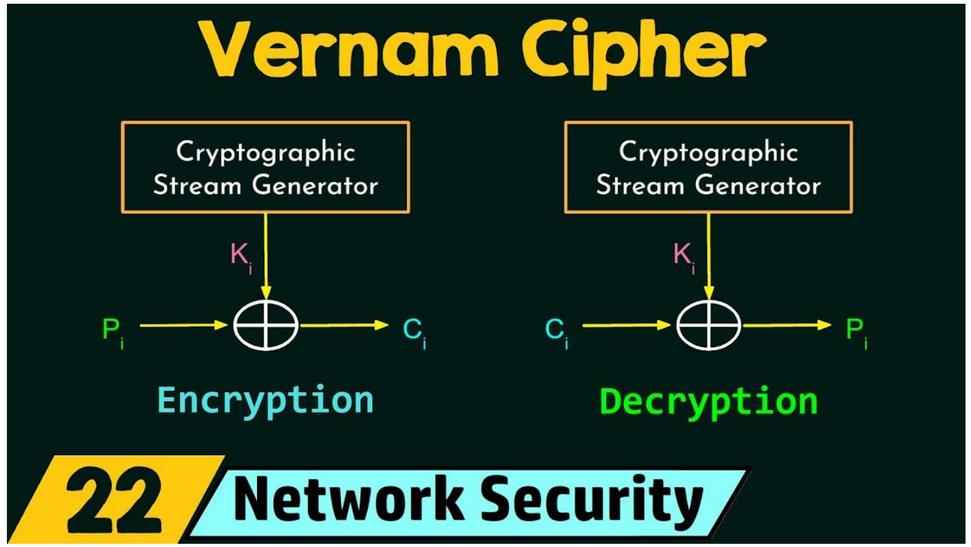


Độ an toàn tính toán

- Một hệ mật được gọi là an toàn về mặt tính toán nếu có một thuật toán tốt nhất để phá nó thì cần ít nhất N phép toán, với N là một số rất lớn nào đó.
- Mô tả mức độ an toàn của một hệ thống mã hóa dựa trên các giả định về sức mạnh tính toán hiện có và khả năng phá vỡ mã hóa của các đối thủ tiềm năng.
- Một hệ thống mã hóa được xem là có độ an toàn tính toán nếu việc phá vỡ nó đòi hỏi một lượng tài nguyên tính toán vượt quá khả năng của các đối thủ trong một khoảng thời gian hợp lý.

Độ an toàn không điều kiện (Unconditional security):

- Một hệ mật được coi là an toàn không điều kiện khi nó không thể bị phá ngay cả với khả năng tính toán không hạn chế.
  - Mã hóa Vernam (One-time Pad) là một ví dụ điển hình về hệ thống mã hóa có độ an toàn không điều kiện



https://demonstrations.wolfram.com/VernamCipherOneTimePad/

LÝ THUYẾT SỐ

Modulo số học (Modular arithmetic):

- Modulo số học: là hệ thống tính toán với số nguyên, nơi các số "quay vòng" sau khi đạt đến một giá trị nhất định
- Trong modulo số học, biểu thức $a \bmod b$ cho kết quả là phần dư không âm và nhỏ hơn b, khi lấy a chia cho b, với điều kiện b là số dương.

Modulo số học (Modular arithmetic)

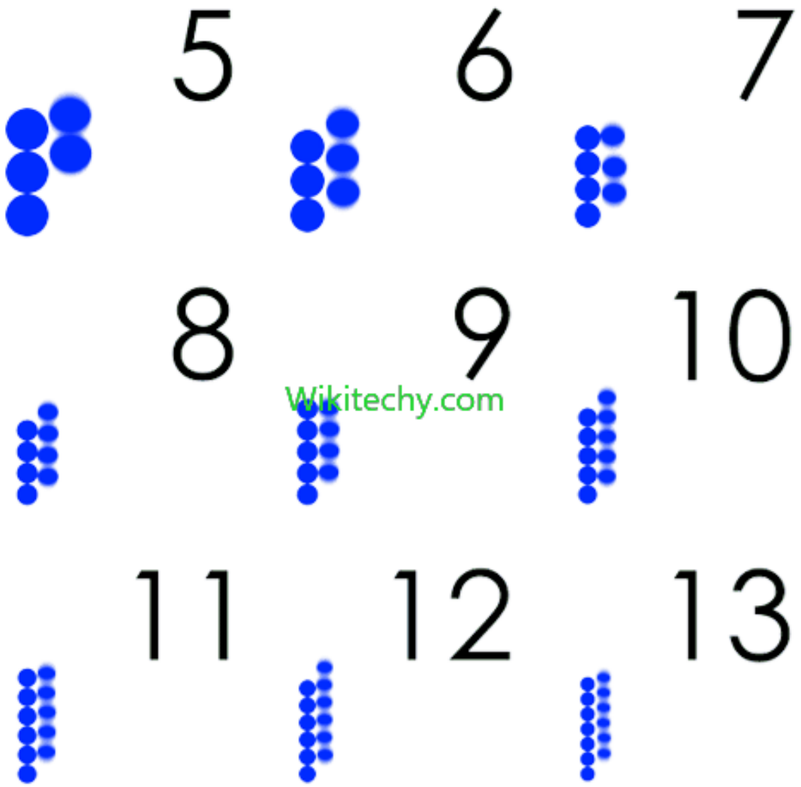
- Trong modulo số học, với hai số nguyên a và b, và một số nguyên dương n, ta nói rằng a đồng dư với b theo modulo n nếu hiệu của chúng là bội số nguyên của n:

$$a \equiv b \,(mod)\, n$$

- Modulo số học là nền tảng trong mật mã học, đặc biệt là trong các hệ mã hoá khoá công khai như RSA, dựa trên độ khó của việc phân tích thừa số các số lớn và các tính chất của phép luỹ thừa modulo.

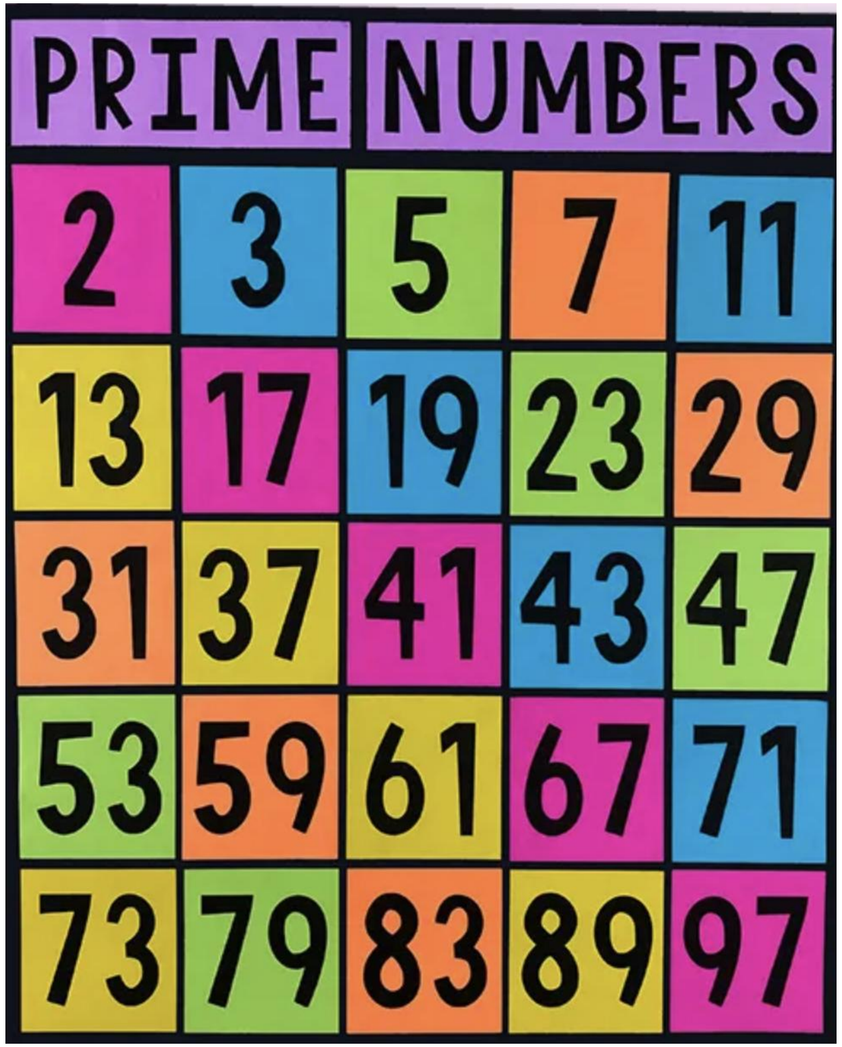
Số nguyên tố

- Số nguyên tố: là một số nguyên lớn hơn 1, chỉ có hai ước số dương là 1 và chính nó
- Mốt số nguyên dương p được gọi là số nguyên tố nếu và chỉ nếu nó chỉ có hai ước là 1 và chính nó
- Một số nguyên dương lớn hơn 1 mà có nhiều hơn hai ước số dương thì được gọi là số hợp (non-prime), nghĩa là nó có thể được phân tích thành tích của hai số nguyên dương nhỏ hơn.



Số nguyên tố

- Số nguyên tố đóng vai trò quan trọng trong các hệ thống mã hóa như RSA, nơi mà bảo mật của hệ thống dựa trên việc phân tích một số lớn thành các số nguyên tố



- An integer $p > 1$ is a prime number if and only if its only divisors are ± 1 and $\pm p$. All numbers other than ± 1 and the prime numbers are composite numbers. In other words, composite numbers are those which are the product of at least two prime numbers.
- An interger $a > 1$ can be factored in a unique way as: $a = p_1^{a_1} \times p_2^{a_2} \times \cdots \times p_t^{a_t}$ where $p_1 < p_2 < \cdots < p_n$ are prime numbers and where each a_i is a positive integer.

Thuật toán Euclid

- Thuật toán Euclide (Euclidean algorithm) là một phương pháp hiệu quả để tìm ước chung lớn nhất (GCD - Greatest Common Divisor) của hai số nguyên.
- Ước chung lớn nhất của hai số nguyên a và b là số nguyên lớn nhất chia hết cả hai số đó.

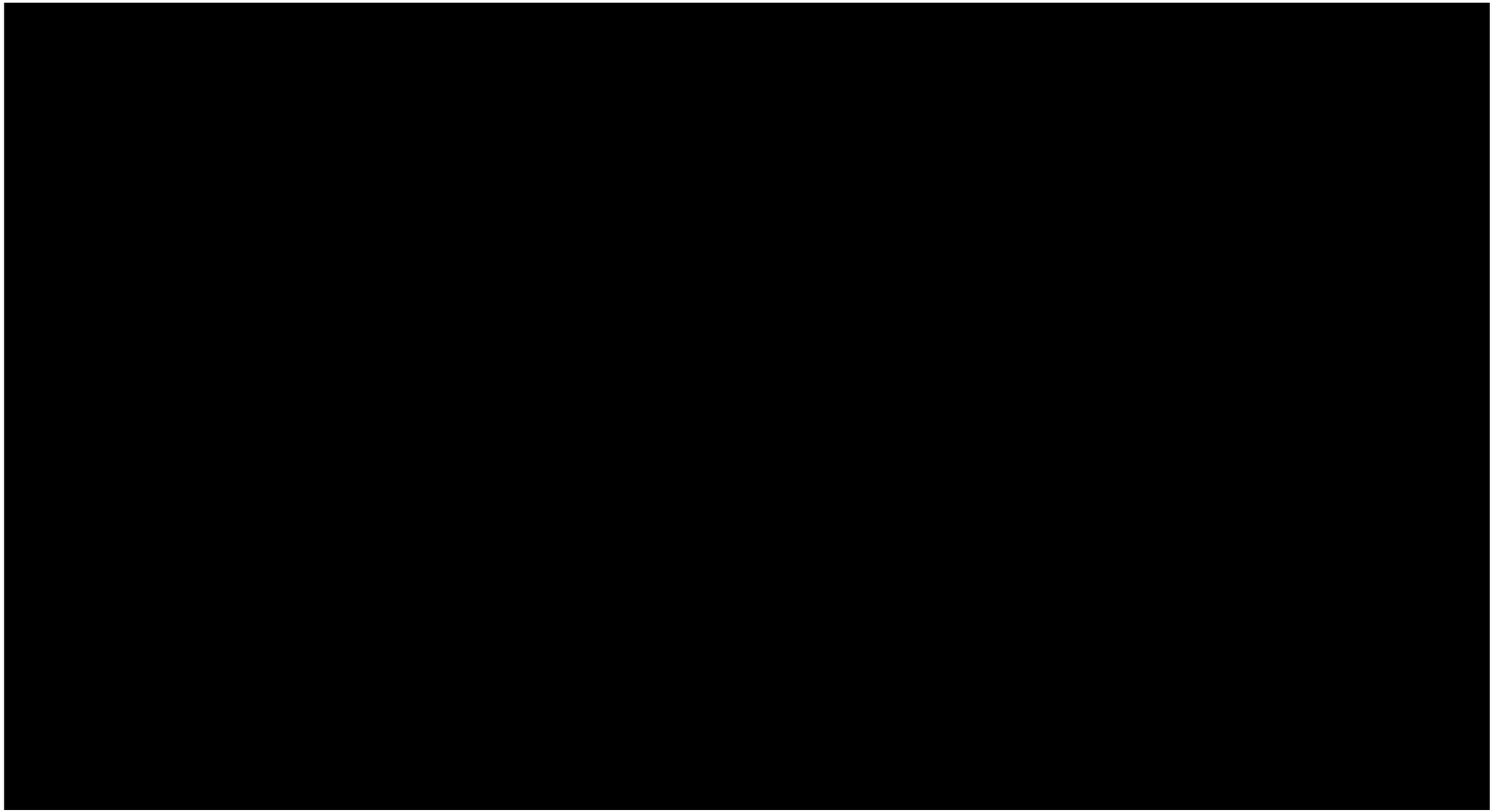
2.2 THE EUCLIDEAN ALGORITHM

One of the basic techniques of number theory is the Euclidean algorithm, which is a simple procedure for determining the greatest common divisor of two positive integers. First, we need a simple definition: Two integers are **relatively prime** if and only if their only common positive integer factor is 1.

Greatest Common Divisor

Recall that nonzero b is defined to be a divisor of a if $a = mb$ for some m, where a, b, and m are integers. We will use the notation $\gcd(a, b)$ to mean the **greatest common divisor** of a and b. The greatest common divisor of a and b is the largest integer that divides both a and b. We also define $\gcd(0, 0) = 0$.

LÝ THUYẾT SỐ



Thuật toán Euclid

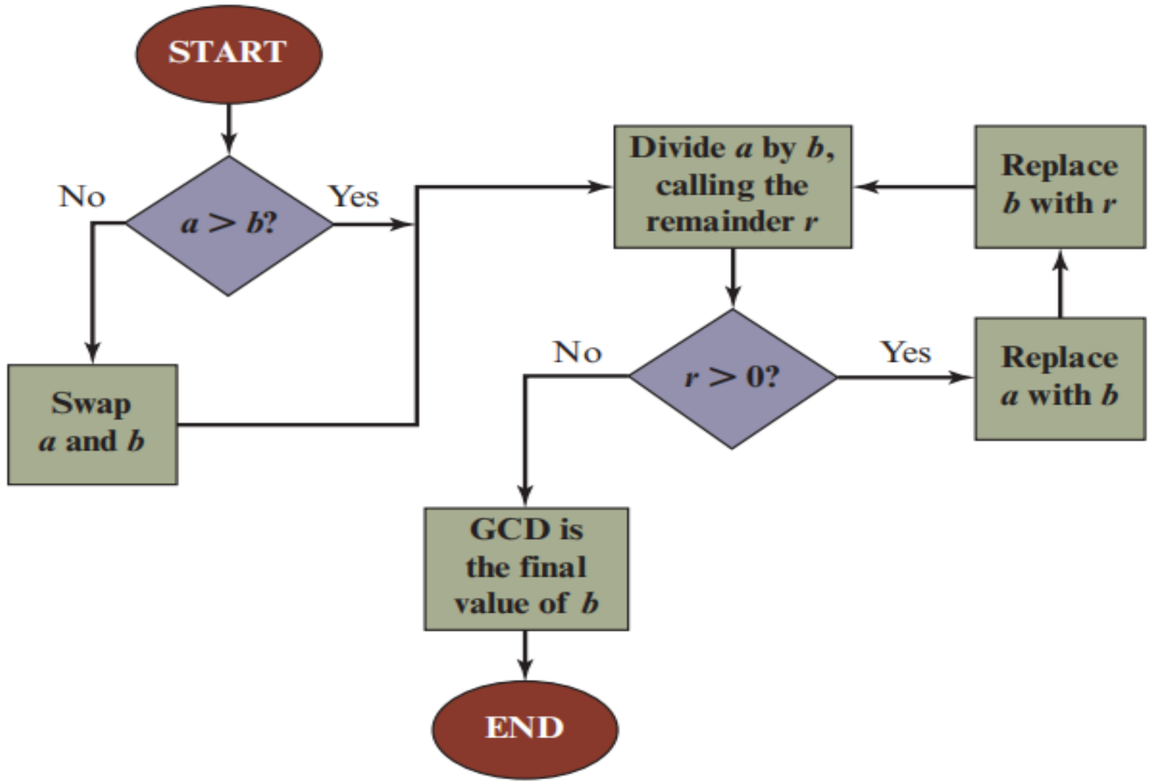


Figure 2.2 Euclidean Algorithm

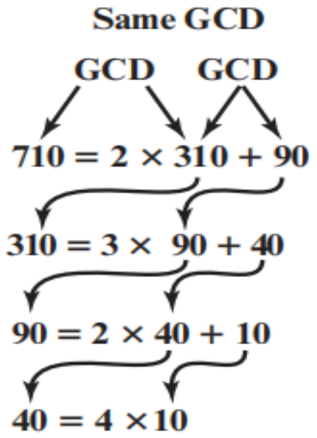


Figure 2.3 Euclidean Algorithm Example: gcd(710, 310)

Find GCF or GCD using the Euclidean Algorithm

Example:

Find GCD of 12 and 30

30 ÷ 12 = 2 remainder 6

12 ÷ 6 = 2 remainder 0

GCD

The GCD of 12 and 30 is 6

Find GCD of 123 and 36

123 ÷ 36 = 3 remainder 15

36 ÷ 15 = 2 remainder 6

15 ÷ 6 = 2 remainder 3

6 ÷ 3 = 2 remainder 0

GCD

The GCD of 123 and 36 is 3

THỰC HÀNH

TỔNG KẾT

- ✓ Lý thuyết thông tin
- ✓ Lý thuyết số
- ✓ Thực hành

ĐẠI NAM

UNIVERSITY

Thank You