ĐẠI NAM
UNIVERSITY

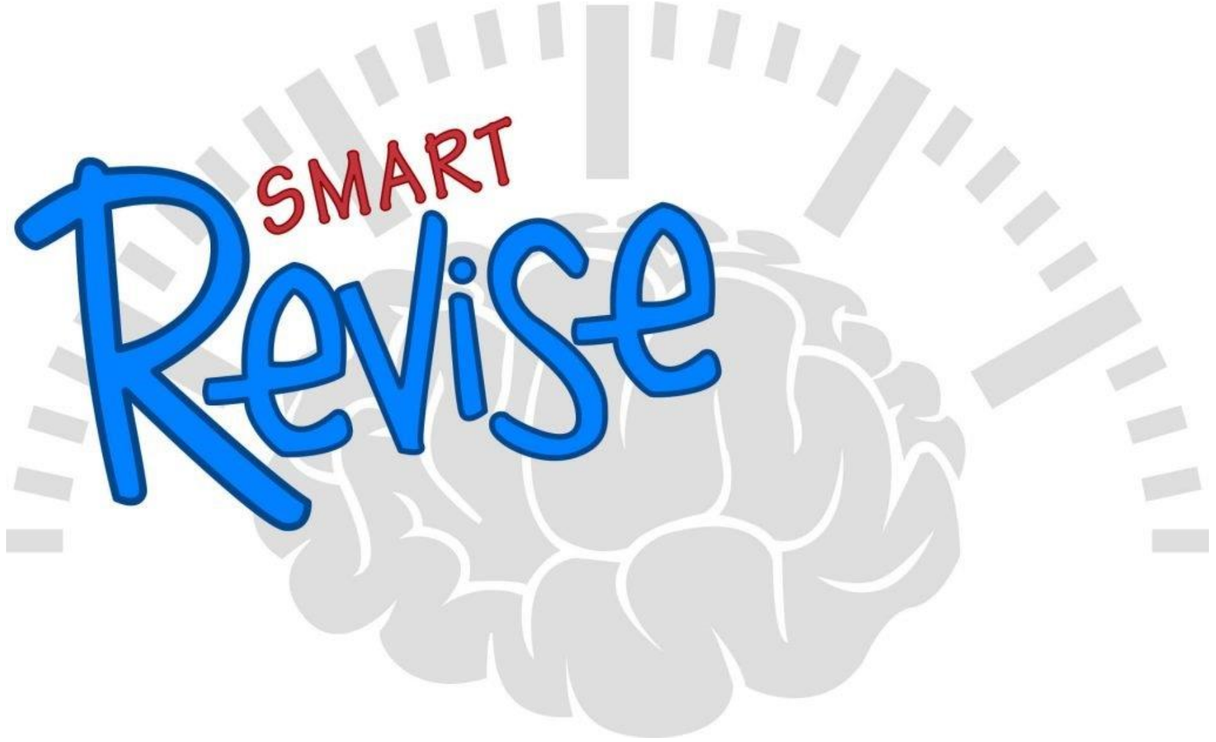
Bài 2

CƠ SỞ AN TOÀN, BẢO MẬT THÔNG TIN (tiếp)

Giảng viên: TS. Trần Quý Nam
(namtq@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

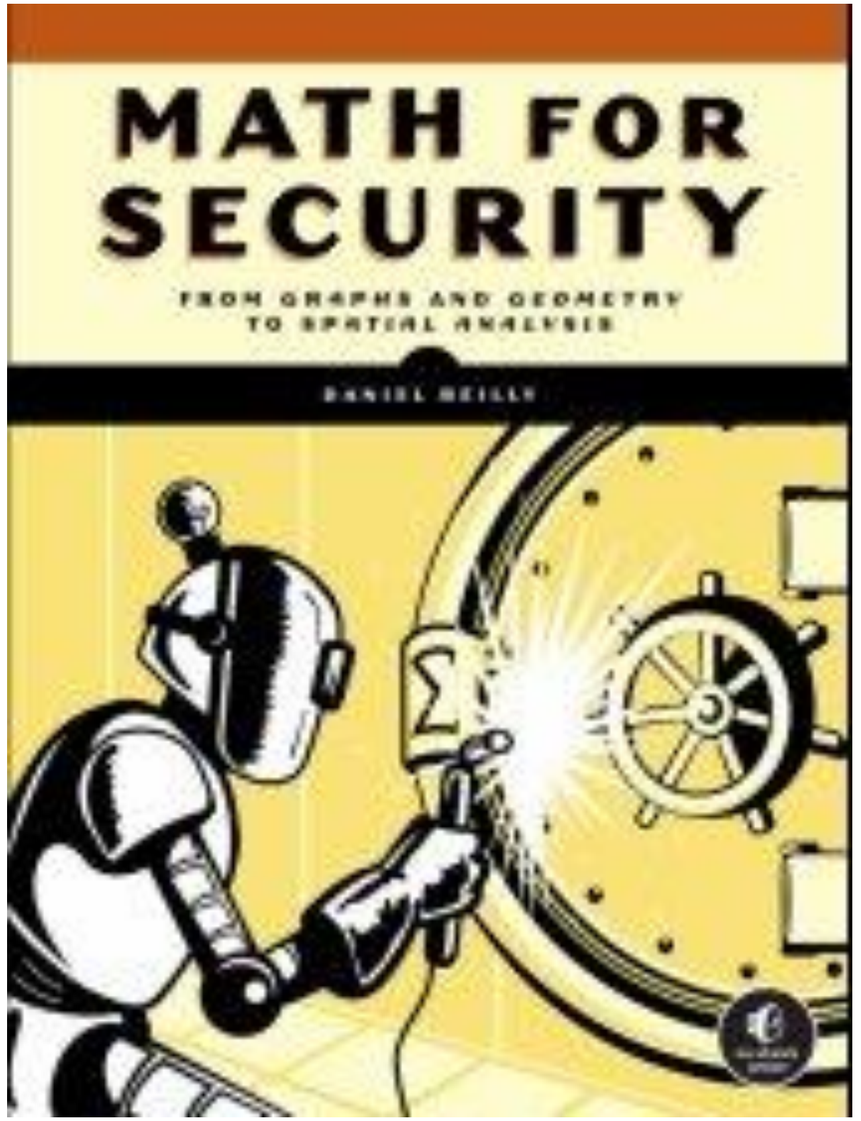


OUTLINE

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

CƠ SỞ TOÁN HỌC

- Sự cần thiết phải 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

LÝ THUYẾT THÔNG TIN

❖ Thông tin là được truyền từ đối tượng này đến đối tượng khác để báo một "điều" gì đó. Thông tin chỉ có ý nghĩa khi "điều" đó bên nhận chưa biết.

❖ Thông tin xuất hiện dưới nhiều dạng âm thanh, hình ảnh, ...

❖ Một trong những phương tiện để diễn đạt thông tin là ngôn ngữ.

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
- Trên thực tế, có 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)

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

Communication is one of the most basic human needs. From smoke signals to carrier pigeons to the telephone to television, humans have always sought methods that would allow them to communicate farther, faster and more reliably. But the engineering of communication systems was always tied to the specific source and physical medium. Shannon instead asked, “Is there a grand unified theory for communication?” In a 1939 letter to his mentor, Vannevar Bush, Shannon outlined some of his initial ideas on “fundamental properties of general systems for the transmission of intelligence.” After working on the problem for a decade, Shannon finally published his masterpiece(opens a new tab) in 1948: “A Mathematical Theory of Communication.”

Định lượng thông tin

❖ Năm 1948, Claude Shannon đã lần đầu giới thiệu thuật ngữ “bit” để làm đơn vị đo lường thông tin

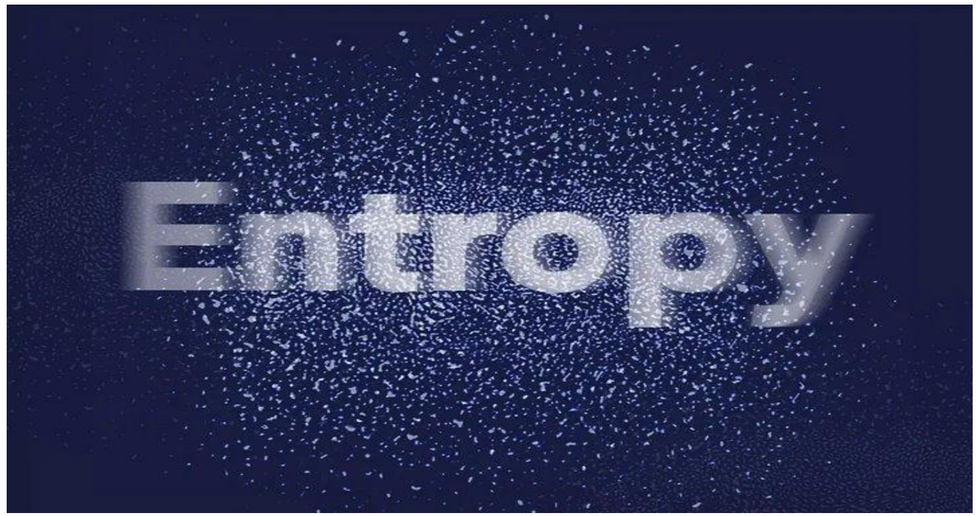
❖Lấy ví dụ đơn giản, giả sử chúng ta có một mã là một chuỗi nhị phân độ dài 5, chẳng hạn như “10001”. Khi đó, lượng tin của mã này sẽ là:

$$I(\text{“10001”}) = -\log_2 p(\text{“10001”}) = -\log_2 \frac{1}{2^5} = -(-5) = 5(\text{bits})$$

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):

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



Entropy is a scientific concept, most commonly associated with states of disorder, randomness, or uncertainty. The term and the concept are used in diverse fields, from classical thermodynamics, where it was first recognized, to the microscopic description of nature in statistical physics, and to the principles of information theory. It has found far-ranging applications in chemistry and physics, in biological systems and their relation to life, in cosmology, economics, sociology, weather science, climate change and information systems including the transmission of information in telecommunication.

Entropy

- Lý thuyết thông tin được xây dựng dựa trên nền tảng xác suất thống kê.
- Thông số quan trọng nhất là entropy: lượng thông tin chứa trong một biến ngẫu nhiên
- Entropy hợp hay entropy có điều kiện để đo lường thông tin tương hỗ (lượng thông tin chung giữa hai biến ngẫu nhiên)

Entropy

Entropy của biến X, $H(X)$, được tính bằng

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

Một trong những trường hợp thường gặp nhất của entropy cho biến ngẫu nhiên là **hàm entropy nhị phân** .tức là entropy cho biến ngẫu nhiên X có phân phối xác suất $p(x)$ với duy nhất hai khả năng $\{0, 1\}$.

$$H_{\mathbb{b}}(X) = \sum_{x \in \mathcal{X}} -p(x) \log p(x) - (1 - p(x)) \log(1 - p(x))$$

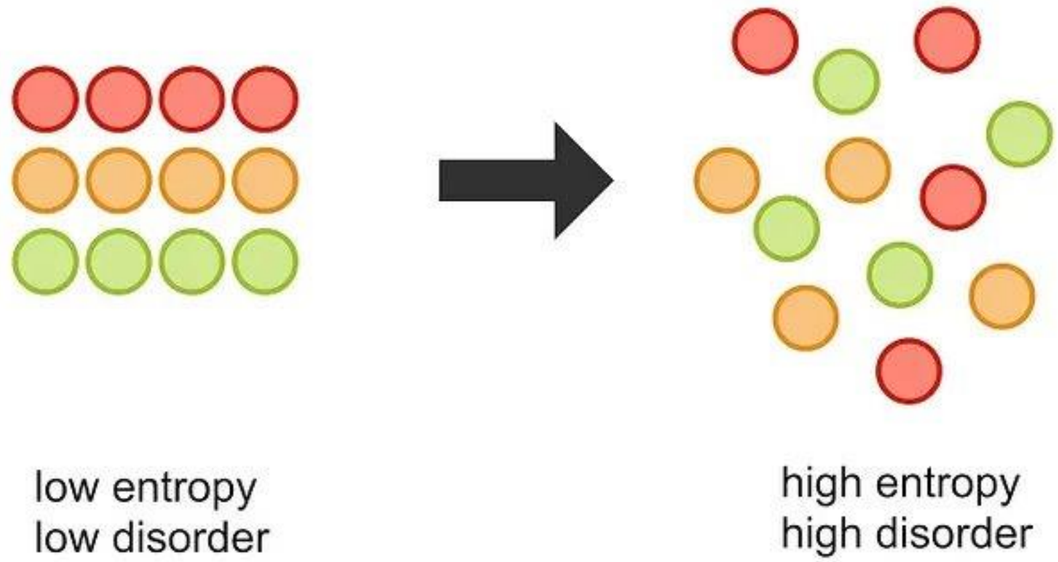
Trong trường hợp X là một biến ngẫu nhiên liên tục, entropy của X sẽ được tính theo công thức tích phân:

$$H(X) = -\int_{x \in \mathcal{X}} p(x) \log p(x) dx$$

Entropy is central to the second law of thermodynamics, which states that the entropy of an isolated system left to spontaneous evolution cannot decrease with time. As a result, isolated systems evolve toward thermodynamic equilibrium, where the entropy is highest. A consequence of the second law of thermodynamics is that certain processes are irreversible.

LÝ THUYẾT THÔNG TIN

- Tại sao không nên sử dụng mật khẩu kiểu “123456” ?
- Mật khẩu như “By8@h*9” 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? ==> Dùng AI

TỐC ĐỘ NGÔN NGỮ

- Tốc độ ngôn ngữ trong lý thuyết thông tin, còn được gọi là "entropy rate," là một khái niệm 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 $H(X)$ cũng được tính theo công thức:

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

❖ Trong đó:

➢ $L(X_1, X_2, \dots, X_n)$: là tổng số ký hiệu (âm tiết, từ, hoặc đơn vị ngôn ngữ khác) được nói trong khoảng thời gian n.

➢ n là khoảng thời gian đo lường (thường tính bằng giây).

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.

ĐỘ DƯ THỪA THÔNG TIN

❖ Độ dư thừa thông tin đo lường mức độ lãng phí hoặc thừa thãi trong quá trình mã hóa hoặc truyền tải thông tin.

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

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

r(X) là mật độ thông tin. Tính bằng:

$$r(X) = \frac{H(X)}{\log_2 M}$$

- H(X) là entropy của ngôn ngữ (lượng thông tin trung bình trên mỗi ký hiệu).
- M là số lượng ký hiệu trong bảng mã (ví dụ: 26 chữ cái trong tiếng Anh, 2 ký hiệu trong hệ nhị phân).

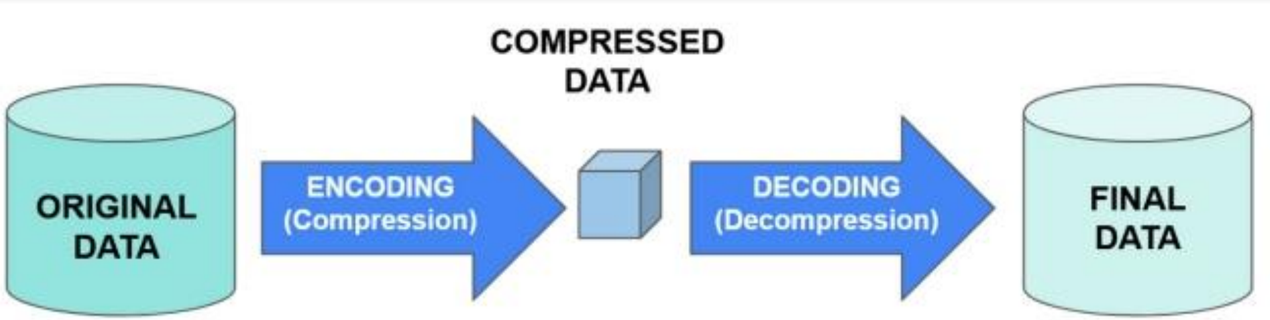
Ví dụ về độ dư thừa thông tin trong ngôn ngữ

Theo nghiên cứu của Shannon và các nhà ngôn ngữ học:

Ngôn ngữ	$R(X)$ (Âm tiết/giây)	$H(X)$ (bit/âm tiết)	Độ dư thừa (%)
Tiếng Anh	6.19	9.83	~50%
Tiếng Pháp	7.18	7.75	~60%
Tiếng Tây Ban Nha	7.82	8.01	~59%
Tiếng Nhật	7.84	5.25	~67%

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.

❖ Trong đó:

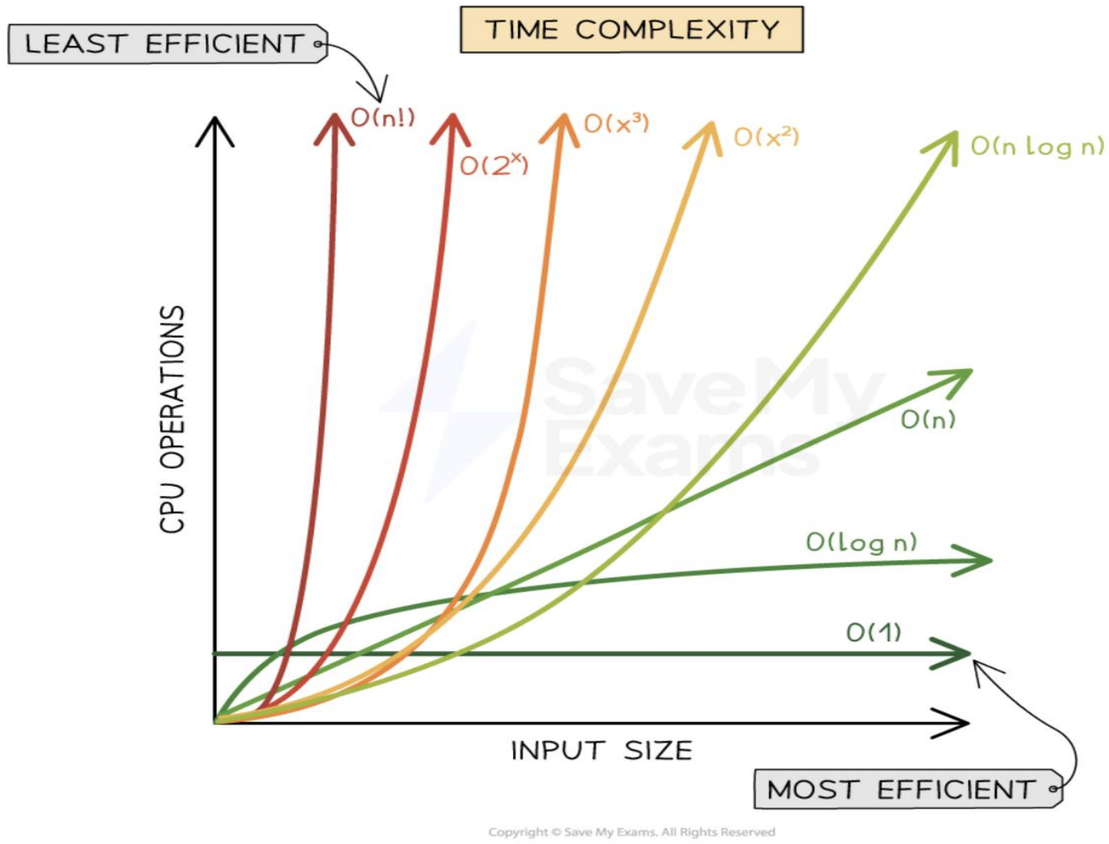
$$U = \frac{H(k)}{D}$$

- $H(k)$: 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

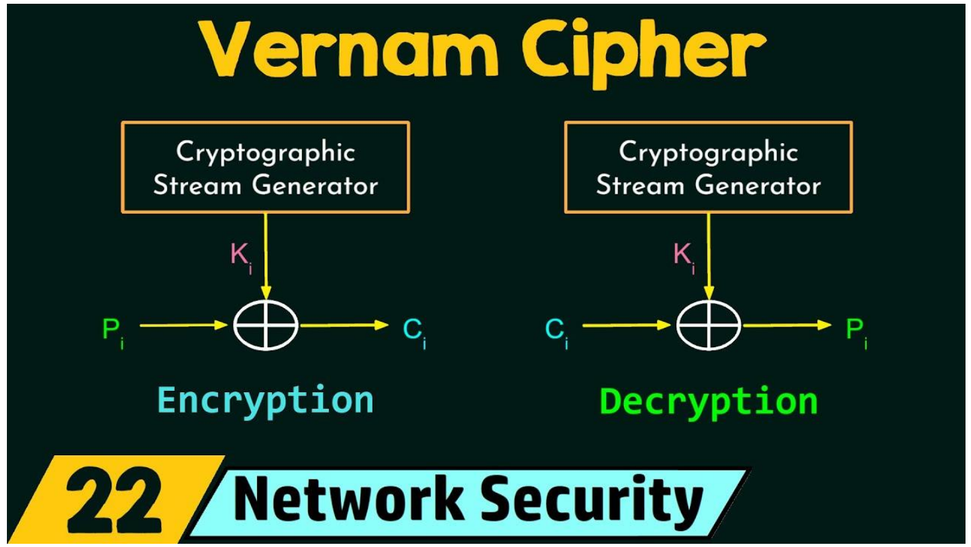


Độ an toàn tính toán

- 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

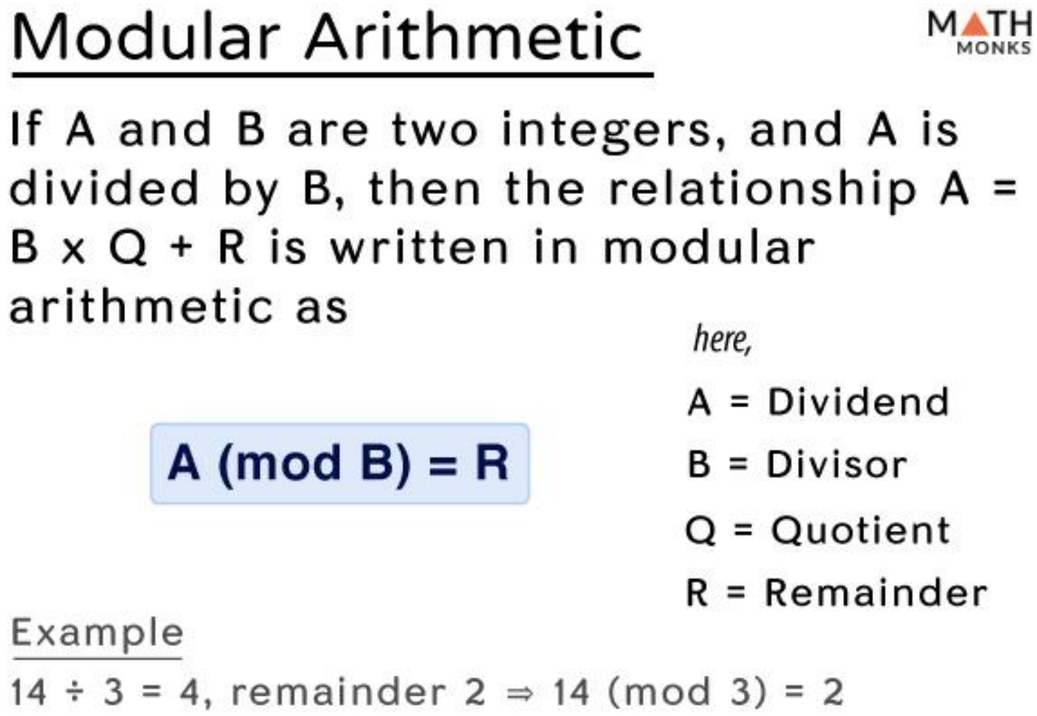


https://demonstrations.wolfram.com/VernamCipherOneTimePad/

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 mod 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.



SỐ ĐỒNG DƯ (Congruence Number)

- Số đồng dư là số có cùng một phần dư khi chia cho một số nguyên dương cho trước (gọi là modulo).
- 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.

2.3 MODULAR ARITHMETIC

The Modulus

If a is an integer and n is a positive integer, we define $a \bmod n$ to be the remainder when a is divided by n. The integer n is called the **modulus**. Thus, for any integer a, we can rewrite Equation (2.1) as follows:

$$a = qn + r \qquad 0 \leq r < n; q = \lfloor a/n \rfloor$$

$$a = \lfloor a/n \rfloor \times n + (a \bmod n)$$

$$11 \bmod 7 = 4; \qquad -11 \bmod 7 = 3$$

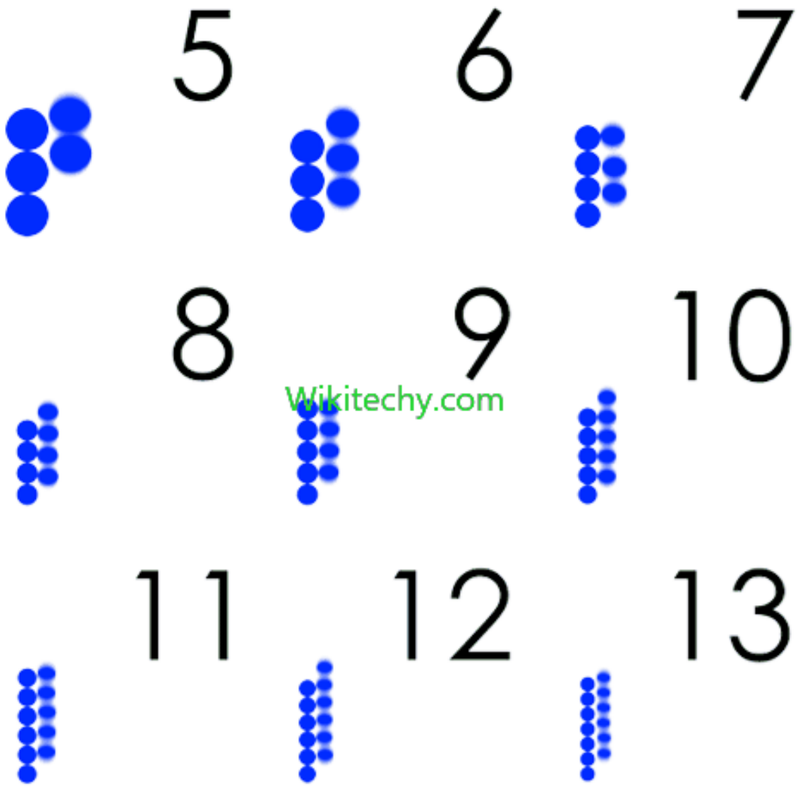
Two integers a and b are said to be **congruent modulo** n, if $(a \bmod n) = (b \bmod n)$. This is written as $a \equiv b \pmod{n}$.[2]

$$73 \equiv 4 \pmod{23}; \qquad 21 \equiv -9 \pmod{10}$$

Note that if $a \equiv 0 \pmod{n}$, then $n \mid a$.

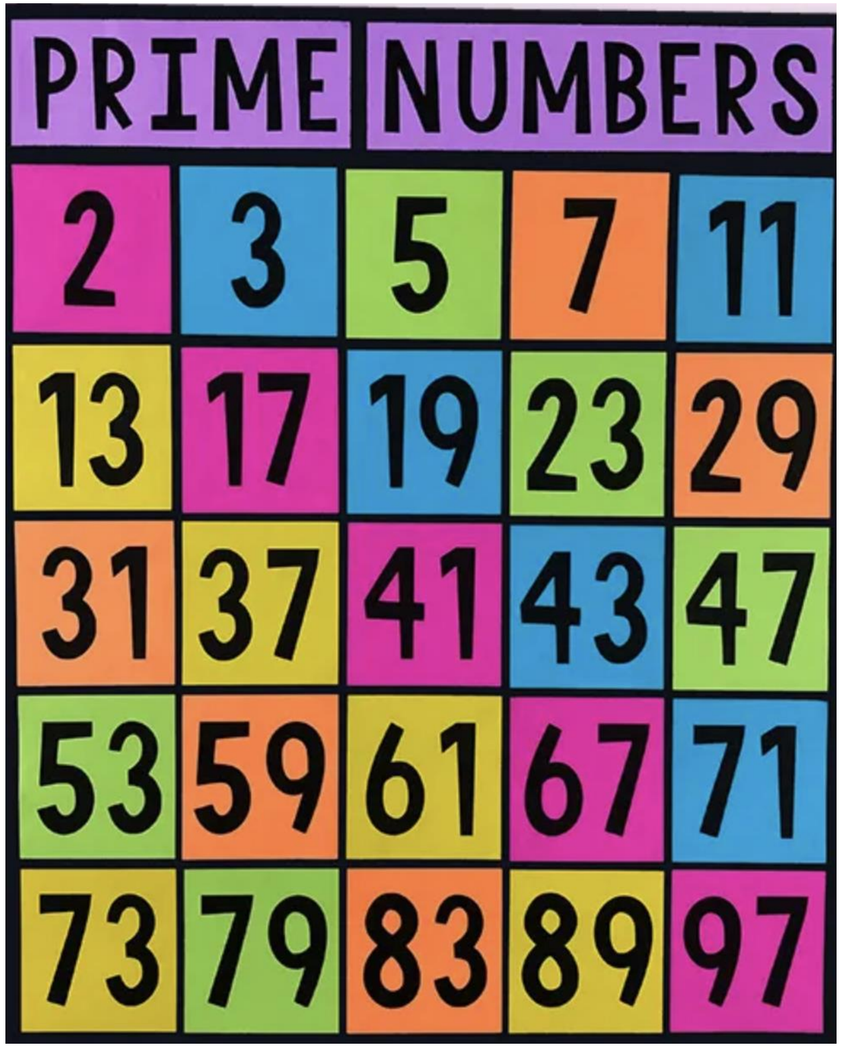
Số nguyên tố

- 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.

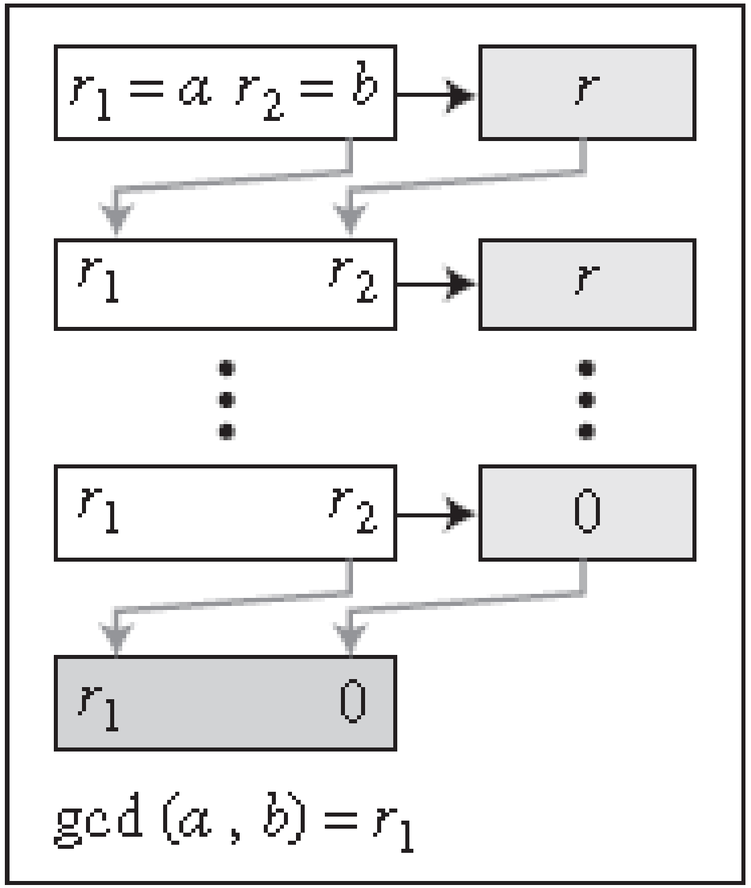
❖ Ước chung lớn nhất của hai số nguyên a và b là số nguyên lớn nhất chia hết cả hai số đó.

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 $\gcd(0, 0) = 0$.



a. Process

```
r1 = a; r2 = b;        (Initialization)
while (r2 > 0)
  {
   q = r1 / r2;
   r = r1 - q * r2;
   r1 = r2; r2 = r;
  }

   gcd (a, b) = r1
```

b. Algorithm

Thuật toán Euclid

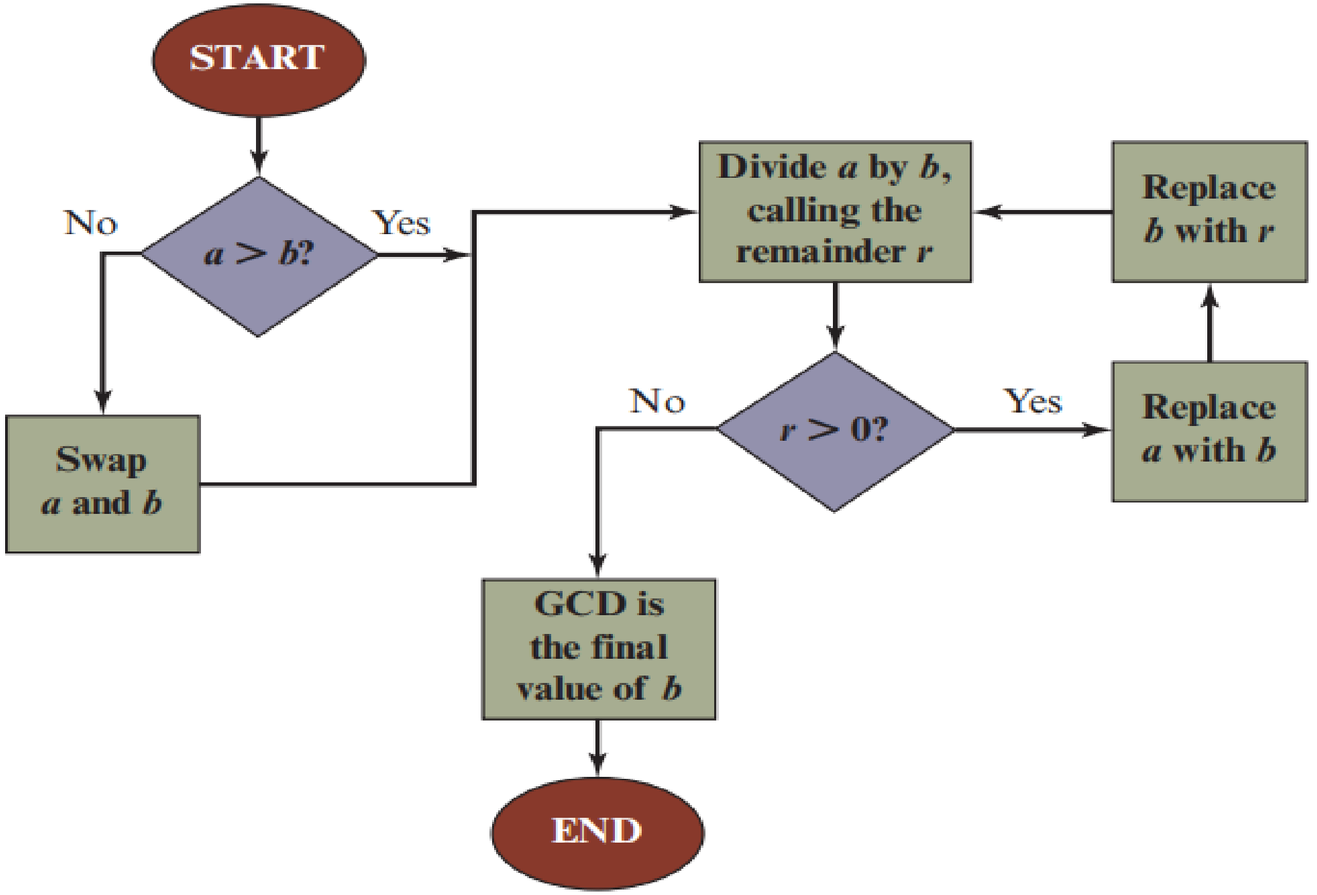


Figure 2.2 Euclidean Algorithm

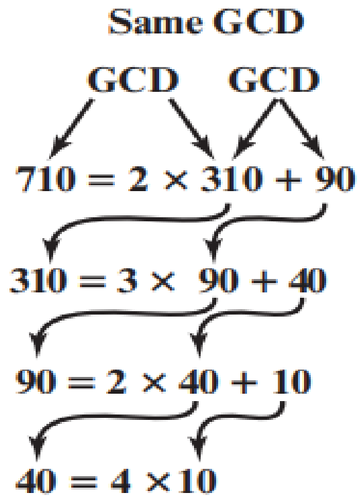


Figure 2.3 Euclidean Algorithm Example: gcd(710, 310)

Sử dụng AI (có thể ChatGPT, Google, DeepSeek để tìm hiểu về số nguyên tố cùng nhau (co-prime hay relatively prime) ?

Sinh viên trả lời tốt sẽ được cộng điểm.

Hai số nguyên a, b ∈ Z \{0} được gọi là nguyên tố cùng nhau nếu:

gcd(a, b) = 1.

Ví dụ: (5,8) , (9,14) là các cặp nguyên tố cùng nhau

- Phần tử nghịch đảo của phép nhân modulo:
- Nếu hai số nguyên *a* và *n* nguyên tố cùng nhau, thì tồn tại số nguyên *w* sao cho:

$$a.w \equiv 1 \bmod n$$

- Ta gọi *w* là phần tử nghịch đảo của *a* trong phép modulo cho *n* và ký hiệu là a^{-1}

Nghịch đảo modulo

Nghịch đảo của một số nguyên là số mà khi nhân với nó sẽ có tích là 1. Để tính thương của 2 số, ta nhân một số với nghịch đảo của số kia.

$\frac{a}{b} = a * b^{-1}$

Tương tự như vậy, nghịch đảo modulo của một số theo modulo M là số mà khi nhân với nó thì được tích chia M dư 1.

$$\text{b là nghịch đảo modulo M của a} \iff (a * b) \mod M = 1$$

hay viết cách khác:

$$\text{b là nghịch đảo modulo M của a} \iff a * b \equiv 1 \mod M$$

Ví dụ: 7 là nghịch đảo modulo 11 của 8 vì 7*8 = 56 và 56 % 11 = 1

Và phép chia modulo sẽ tương ứng với phép nhân nghịch đảo modulo.

Để tính nghịch đảo nhân modulo n, áp dụng giải thuật **Euclid mở rộng**:

$s * n + t * b = gcd(n,b) = 1$ (định lí Bézout)

- Thực hiện phép mod cả 2 vế

$(s * n + t * b) \bmod n = 1 \bmod n$

$[(s * n) \bmod n] + [(t*b) \bmod n] = 1 \bmod n$

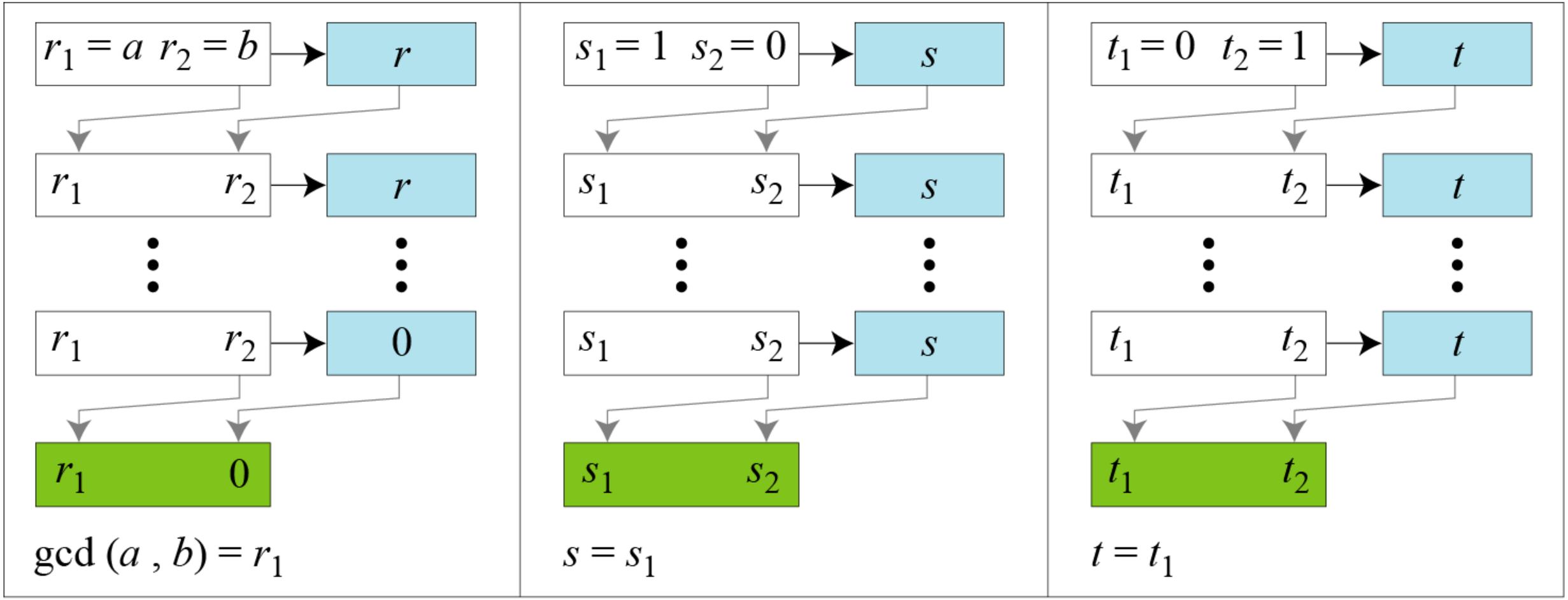
$0 + [(t*b) \bmod n] = 1$

➔ $(t * b) \bmod n = 1$ → t chính là nghịch đảo nhân của b

Thuật toán này vừa có thể tính được gcd(a,b) vừa tính được các giá trị s và t

https://en.wikipedia.org/wiki/Extended_Euclidean_algorithm#:~:text=Extended%20Euclidean%20algorithm%20also%20refers,a%20and%20b%20are%20coprime.
https://laptrinhthidau.wordpress.com/2016/08/23/thuat-toan-euclid-mo-rong/
http://nguyenduccuong.com/nckh/80-thut-toan-euclide-m-rng

Euclide mở rộng



a. Process

$r_1 \leftarrow a;$ $r_2 \leftarrow b;$
$s_1 \leftarrow 1;$ $s_2 \leftarrow 0;$
$t_1 \leftarrow 0;$ $t_2 \leftarrow 1;$

(Initialization)

while $(r_2 > 0)$

{

$q \leftarrow r_1 / r_2;$

$r \leftarrow r_1 - q \times r_2 ;$
$r_1 \leftarrow r_2 ;\ r_2 \leftarrow r ;$

(Updating r's)

$s \leftarrow s_1 - q \times s_2 ;$
$s_1 \leftarrow s_2 ;\ s_2 \leftarrow s ;$

(Updating s's)

$t \leftarrow t_1 - q \times t_2;$
$t_1 \leftarrow t_2 ;\ t_2 \leftarrow$ t;

(Updating t's)

}

gcd $(a , b) \leftarrow r_1;\ s \leftarrow s_1;\ t \leftarrow t_1$

b. Algorithm

Euclide mở rộng

Euclide mở rộng

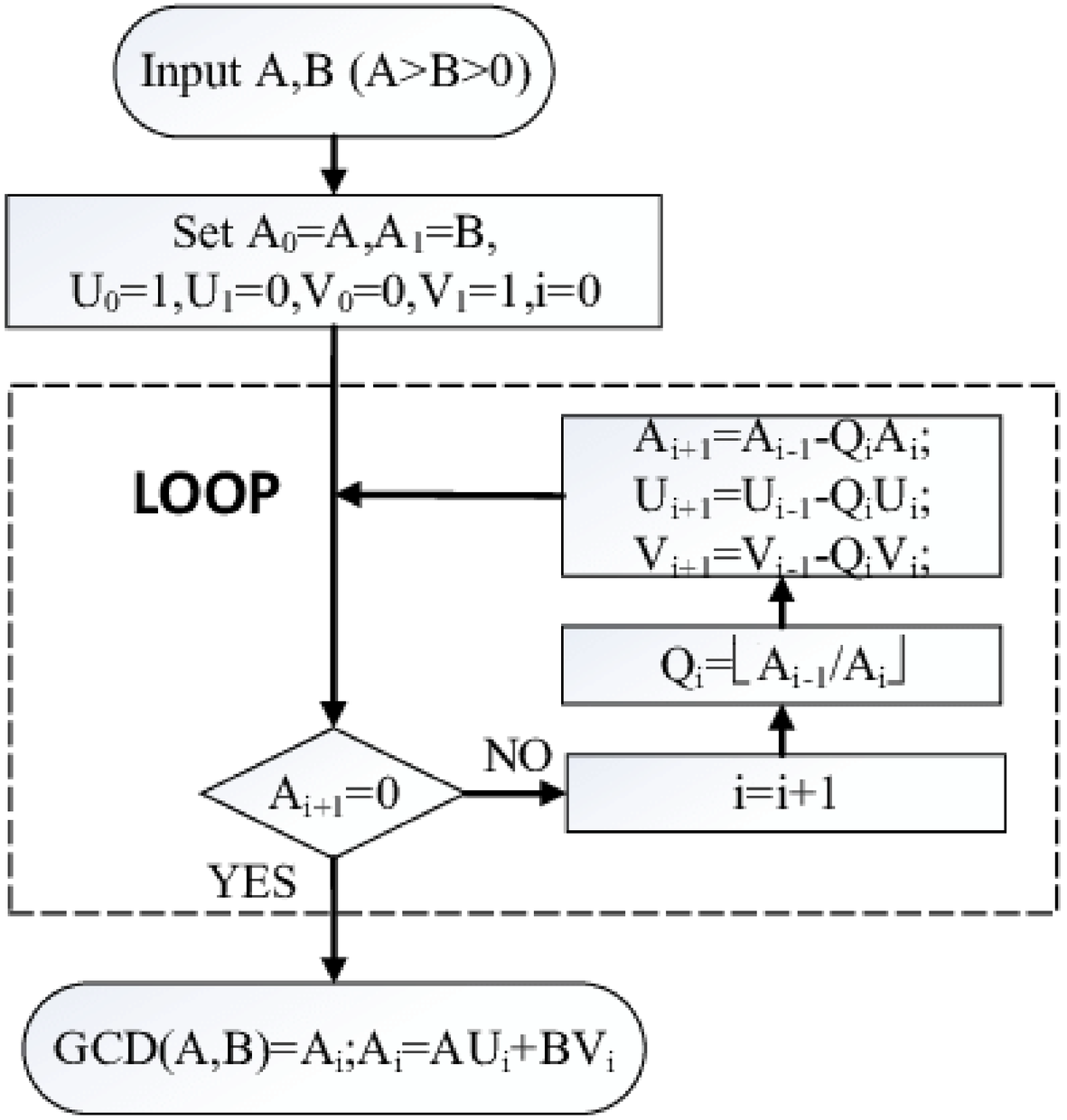
Ví dụ: a = 161 và b = 28, tìm gcd (a, b) và giá trị s và t.

Giải: $r = r_1 - q \times r_2$; $s = s_1 - q \times s_2$; $t = t_1 - q \times t_2$

q	r_1	r_2	r	s_1	s_2	s	t_1	t_2	t
5	161	28	21	1	0	1	0	1	−5
1	28	21	7	0	1	−1	1	−5	6
3	21	7	0	1	−1	4	−5	6	−23
	7	0		**−1**	4		**6**	−23	

→ gcd (161, 28) = 7, s = −1 và t = 6.

Euclide mở rộng



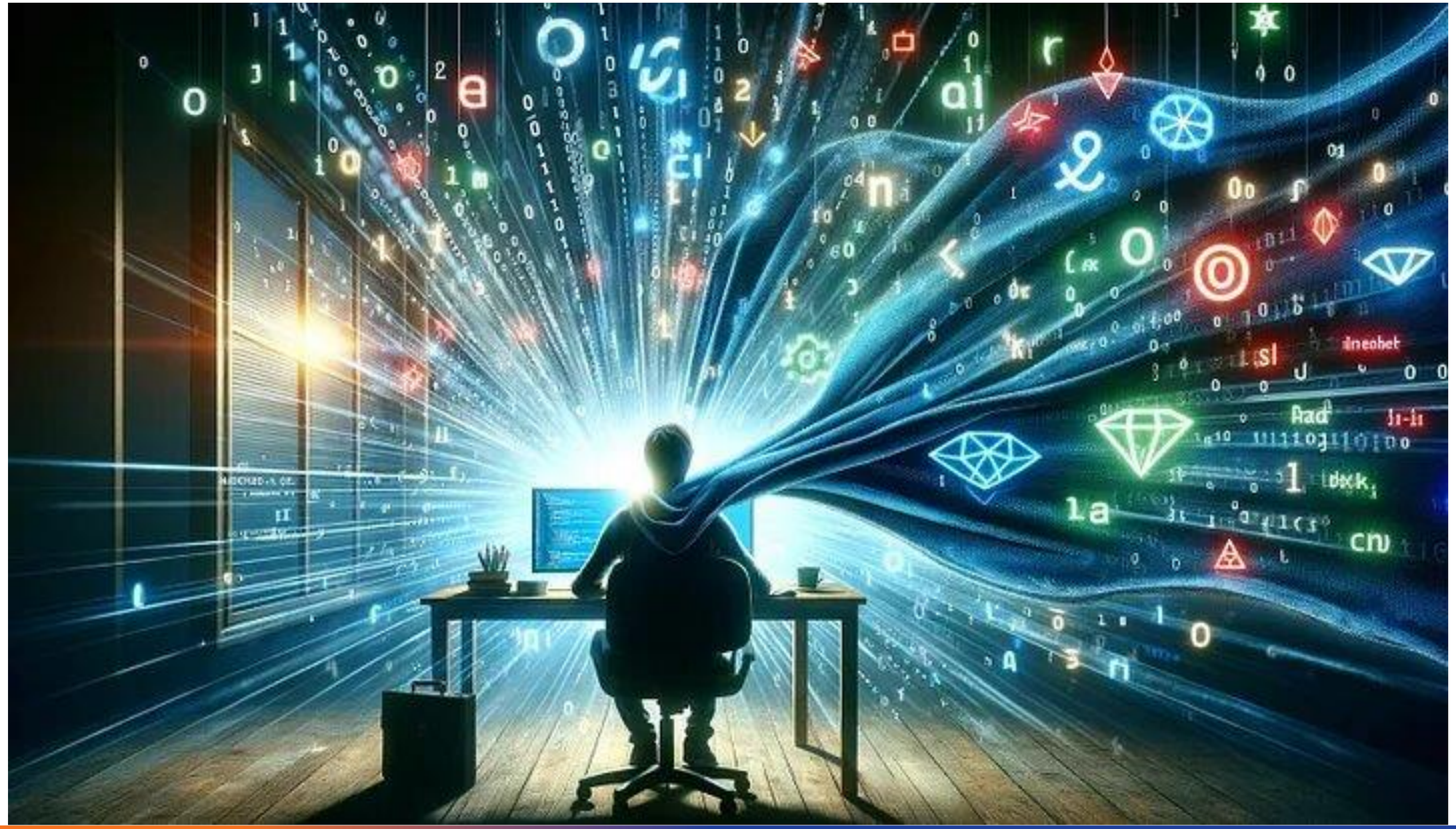
Euclide mở rộng

Input: $a, b \in R$.

Output: $g \in R$ a gcd of a and b together with $s, t \in R$ such that $g = s\,a + t\,b$.

$r_0 := a;\ s_0 := 1;\ t_0 := 0$
$r_1 := b;\ s_1 := 0;\ t_1 := 1$
$i := 2$
while $r_{i-1} \neq 0$ **repeat**
  $q_i := r_{i-2}$ quo r_{i-1}
  $r_i := r_{i-2}$ rem r_{i-1}
  $s_i := s_{i-2} - q_i\, s_{i-1}$
  $t_i := t_{i-2} - q_i\, t_{i-1}$
  $i := i + 1$
return $(r_{i-2}, s_{i-2}, t_{i-2})$

THỰC HÀNH



Bài 1: Tính Entropy của một nguồn tin

Cho một nguồn thông tin có phân phối xác suất như sau:

$P(A) = 0.4$

$P(B) = 0.3$

$P(C) = 0.2$

$P(D) = 0.1$

Hãy tính entropy của nguồn tin này theo công thức:

$$H(X) = -\sum_i P(x_i) \log_2 P(x_i)$$

Bài 2: Viết chương trình C++ để tính entropy tự động của một nguồn thông tin dựa trên xác suất xuất hiện của các ký tự.

Chương trình thực hiện:

- Nhập số lượng ký tự và xác suất của chúng.
- Tính entropy theo công thức:

$$H(X) = -\sum_{i} P(x_i) \log_2 P(x_i)$$

- In ra giá trị entropy của nguồn tin.

THỰC HÀNH

Bài 2: Viết chương trình C++ kiểm tra một số có phải số nguyên tố hay không.

Sinh viên giải thích trước cả lớp và giảng viên về ý nghĩa, hoạt động và kết quả các dòng code C++

Bài 3: Viết chương trình C++ thực hiện thuật toán Euclide (Euclidean algorithm) để tìm ước chung lớn nhất (GCD - Greatest Common Divisor) của hai số nguyên.

Sinh viên giải thích trước cả lớp và giảng viên về ý nghĩa, hoạt động và kết quả các dòng code C++

Bài 5: Viết chương trình C++ thực hiện thuật toán Euclide (Euclidean algorithm) mở rộng để tìm nghịch đảo modulo của số nguyên a theo modulo m, tức là tìm số x sao cho:

$$a \cdot x \equiv 1 \pmod{m}$$

Sinh viên giải thích trước cả lớp và giảng viên về ý nghĩa, hoạt động và kết quả các dòng code C++

TỔNG KẾT

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

ĐẠI NAM
UNIVERSITY