Number Theory and Cryptography
ทฤษฎีจำนวนและการเข้ารหัส

3. Primes and Greatest Common Divisors

Primes

- **Definition**: A positive integer *p* greater than 1 is called prime if the only positive factors of *p* are 1 and p. A positive integer that is greater than 1 and is not prime is called composite.
- Example: The integer 7 is prime because its only positive factors are 1 and 7, but 9 is composite because it is divisible by 3.

3.1 Primes

- The study of prime numbers goes back to ancient times.
- Thousands of years ago it was known that there are infinitely many primes; the proof of this fact, found in the works of **Euclid**, is famous for its elegance and beauty.
- Primes have become essential in modern cryptographic systems (i.e. finding large primes)
    - Length of time required to factor large integers into their prime factors indicate strength of the systems.

The Fundamental Theorem of Arithmetic

- **Theorem 1**: Every integer > 1 can be written uniquely as a prime or as the product of two or more primes, where the prime factors are written in order of non-decreasing size.
- Examples:
    - $100 = 2 * 2 * 5 * 5 = 2^2 * 5^2$
    - $641 = 641$
    - $999 = 3 * 3 * 3 * 37 = 3^3 * 37$
    - $1024 = 2^{10}$

3.2 Trial Division

- How to show that a given integer is prime?
- **Theorem 2**: If n is a composite integer, then n has a prime divisor less than or equal to $\sqrt{n}$.
- **Trial division** algorithm is a brute-force algorithm which divide n by all primes not exceeding $\sqrt{n}$ and conclude that n is prime if it is not divisible by any of these primes.
- Examples: Show that 101 is prime.
  - The only primes not exceeding $\sqrt{101}$ are 2, 3, 5, and 7. They do not divide 101, it follows that 101 is prime.

Prime Factorisation

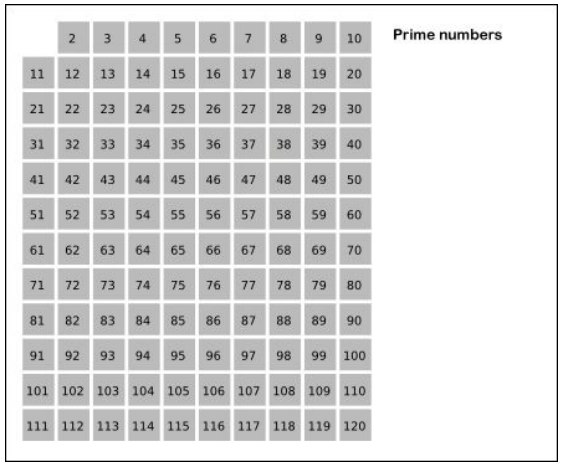
- Every integer has a prime factorisation, but how to find it?
- **Problem**: finding the prime factorisation of n.
- Steps:
  - Begin by dividing n by successive primes, starting with the smallest prime 2.
    - If n has a prime factor p, then p not exceeding $\sqrt{n}$ will be found. If not, then n is prime.
  - Continue by factoring n/p by successive primes, starting from p, until the factorisation has been reduced to a prime.

Prime Factorisation

- Example: Find the prime factorisation of 7007.
  - 7007 / 2 -> 7007 /3 -> 7007 / 5 (Not divisible)
  - 7007 / 7 = 1001 (divisible)
  - 1001 / 7 = 143 (divisible)
  - 143 / 7 (Not divisible)
  - 143 / 11 = 13 (divisible)
  - 13 is prime, the procedure is completed.
- 7007 = 7 * 1001 = 7 * 7 * 143 = 7 * 7 * 11 * 13

3.3 The Sieve of Eratosthenes

- Find all primes not exceeding a specified positive integer.
- Example, between 1 and 120.
  - Delete all the integers divisible by 2 (except 2).
  - Delete all the integers divisible by 3 (except 3).
  - ... until a prime is > $\sqrt{n}$



Infinitude of Primes

Euclid
(325 B.C.E. – 265 B.C.E.)

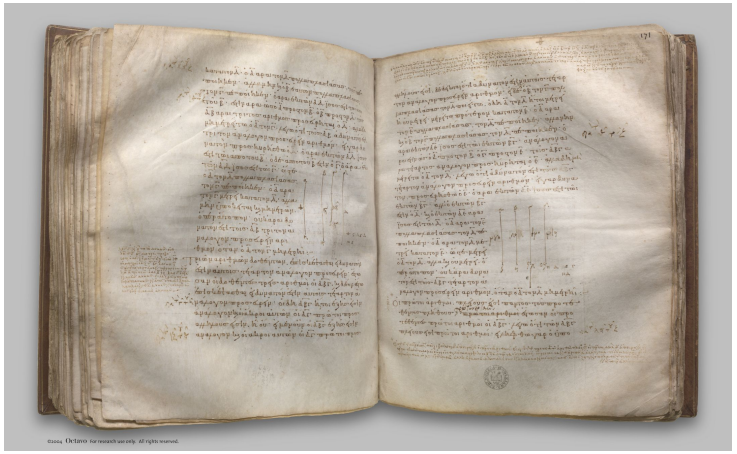
Theorem 3: There are infinitely many primes. (Euclid)

Proof by contradiction: Assume finitely many primes: $p_1, p_2, \ldots, p_n$

- Let $Q = p_1 p_2 \cdots p_n + 1$
- By the fundamental theorem of arithmetic, Q is prime or else it is a product of primes.
- But none of the primes p_j divides Q since if $p_j \mid Q$, then p_j divides $Q - p_1 p_2 \cdots p_n = 1$.
- Hence, there is a prime not on the list $p_1, p_2, \ldots, p_n$.
- It is either Q, or if Q is composite, it is a prime factor of Q. This contradicts the assumption that $p_1, p_2, \ldots, p_n$ are all the primes.
- Consequently, there are infinitely many primes.

Infinitude of Primes

Euclid
(325 B.C.E. – 265 B.C.E.)



Elements (Book IX, Proposition 20)

Obtained from: https://www.claymath.org/library/historical/euclid/images100/175eucmsd21.jpg

Mersene Primes

- Because there are infinitely many primes, given any positive integer, there are primes greater than this integer.
- **Definition**: Prime numbers of the form $2^p - 1$, where p is prime, are called **Mersene primes**.
- $2^2 - 1 = 3$, $2^3 - 1 = 7$, $2^5 - 1 = 37$, and $2^7 - 1 = 127$ are Mersene primes.
- $2^{11} - 1 = 2047$ is not a Mersene prime
  - since $2047 = 23 * 89$.

The Quest to Discover Largest Prime

- There is an efficient test for determining if $2^p - 1$ is prime.
- As of 2018, 51 Mersene primes were known, the largest is $\mathbf{2^{82,589,933} - 1}$, which has nearly 25 million decimal digits (24,862,048).
- The Great Internet Mersene Prime Search (GIMPS) is a distributed computing project to search for new Mersene Primes. - https://www.mersenne.org/

Distribution of Primes

- How many primes are less than a positive number x?
- Mathematicians have been interested in the distribution of prime numbers among the positive integers.
- In the 19th century, the prime number theorem was proved which gives an asymptotic estimate for the number of primes not exceeding x.

Prime Number Theorem: The ratio of the number of primes not exceeding x and $x/\ln x$ approaches 1 as x grows without bound. ($\ln x$ is the natural logarithm of x)

Distribution of Primes

Prime Number Theorem: The ratio of $\pi(x)$, the number of primes not exceeding x, and $x/\ln x$ approaches 1 as x grows without bound. ($\ln x$ is the natural logarithm of x)

- The number of primes not exceeding x, can be approximated by $x/\ln x$.
- $\pi(x)$ can be estimated with other functions, but unsolved questions remained.
- The odds that a randomly selected positive integer less than n is prime are approximately $(n/\ln n)/n = 1/\ln n$.

Distribution of Primes

TABLE 2 **Approximating $\pi(x)$ by $x/\ln x$.**

x	$\pi(x)$	$x/\ln x$	$\pi(x)/(x/\ln x)$
10^3	168	144.8	1.161
10^4	1229	1085.7	1.132
10^5	9592	8685.9	1.104
10^6	78,498	72,382.4	1.084
10^7	664,579	620,420.7	1.071
10^8	5,761,455	5,428,681.0	1.061
10^9	50,847,534	48,254,942.4	1.054
10^{10}	455,052,512	434,294,481.9	1.048

Generating Primes

- Finding large primes is important in cryptography.
- So far, no useful closed formula that always produces primes has been found.
- There is no simple function $f(n)$ such that $f(n)$ is prime for all positive integers n.
- But $f(n) = n^2 - n + 41$ is prime for all integers 1,2,..., 40. Because of this, we might conjecture that f(n) is prime for all positive integers n. But $f(41) = 41^2$ is not prime.
- More generally, there is no polynomial with integer coefficients such that $f(n)$ is prime for all positive integers n.

3.4 Conjectures about Primes

- Even though primes have been studied extensively for centuries, many conjectures about them are unresolved, including:
- **Goldbach's Conjecture**: Every even integer n, $n > 2$, is the sum of two primes.
- It has been verified by computer for all positive even integers up to $4 * 10^{18}$.
- Although no proof has been found, the conjecture is believed to be true by most mathematicians.

Conjectures about Primes

- There are infinitely many primes of the form $n^2 + 1$, where n is a positive integer.
- But it has been shown that there are infinitely many positive integers n such that $n^2 + 1$ is prime or the product of at most two primes.

Conjectures about Primes

- **The Twin Prime Conjecture**: The twin prime conjecture is that there are infinitely many pairs of twin primes.
- Twin primes are pairs of primes that differ by 2.
  - p and $p + 2$ where p is prime and $p + 2$ is prime or the product of 2 primes
- Examples are 3 and 5, 5 and 7, 11 and 13, 4967 and 4969, etc.
- The current world's record for twin primes (as of mid 2018) consists of numbers $2{,}996{,}863{,}034{,}895 * 2^{1{,}290{,}000} \pm 1$, which have 388,342 decimal digits.

3.5 Greatest Common Divisor (GCD)

- The largest integer that divides both of two integers.
- **Definition:** Let a and b be integers, not both zero. The largest integer d such that $d \mid a$ and also $d \mid b$ is called the *greatest common divisor* of a and b. The greatest common divisor of a and b is denoted by $\gcd(a,b)$.
- One way to find the gcd is to find all the positive common divisors of both integers and then take the largest divisor.

Greatest Common Divisor (GCD)

Example 1: What is the greatest common divisor of 24 and 36?

Solution: The positive common divisors of 23 and 36 are

1, 2, 3, 4, 6, and 12.

Thus, gcd(24, 36) = 12

Example 2: What is the greatest common divisor of 17 and 22?

Solution: No other than 1, so gcd(17,22) = 1

Relative Prime Number

- **Definition:** The integers *a* and *b* are *relatively prime* if their greatest common divisor is 1.
- **Example 1**: 17 and 22 are relatively prime because gcd(17, 22) = 1
- **Definition**: The integers $a_1, a_2, \ldots, a_n$ are *pairwise relatively prime* if $\gcd(a_i, a_j) = 1$ whenever $1 \leq i < j \leq n$.
- **Example 2**: 10, 17, and 21 are pairwise relatively prime because gcd(10, 17) = 1, gcd(10, 21) = 1, and gcd(17, 21) = 1.
- **Example 3**: 10, 19, and 24 are not pairwise relatively prime because gcd(10, 24) = 2 which is > 1.

Greatest Common Divisor (GCD) using Prime Factorisation

- Suppose the prime factorisations of *a* and *b* are:

$$a = p_1^{a_1} p_2^{a_2} \cdots p_n^{a_n}, \qquad b = p_1^{b_1} p_2^{b_2} \cdots p_n^{b_n},$$

where each exponent is a nonnegative integer, and where all primes occurring in either prime factorisation are included in both. Then:

$$\gcd(a, b) = p_1^{\min(a_1, b_1)} p_2^{\min(a_2, b_2)} \cdots p_n^{\min(a_n, b_n)}.$$

Greatest Common Divisor (GCD) using Prime Factorisation

- This formula is valid since the integer on the right (of the equals sign) divides both *a* and *b*. No larger integer can divide both *a* and *b*.
- **Example**: $120 = 2^3 * 3 * 5 \qquad 500 = 2^2 * 5^3$

  $\gcd(120,500) = 2^{\min(3,2)} * 3^{\min(1,0)} * 5^{\min(1,3)} = 2^2 * 3^0 * 5^1 = 20$
- Finding the gcd of two positive integers using their **prime factorisations** is not efficient because there is no efficient algorithm for finding the prime factorisation of a positive integer.

Greatest Common Divisor (GCD) using Prime Factorisation

- **Example**: gcd(120,500)
  - 120 = 2 * 2 * 2 * 3 * 5
  - 500 = 2 * 2 * 5 * 5 * 5
  - Integers that both divides 120 and 500 are 2 * 2 * 5 = 20

Greatest Common Divisor (GCD) Exercise

Exercise 1: What is the gcd of 84 and 96?

Exercise 2: What is the gcd of 48 and 72?

Exercise 3: What is the gcd of 33 and 77?

Least Common Multiple (LCM)

- **Definition:** The *least common multiple* of the positive integers *a* and *b* is the smallest positive integer that is divisible by both *a* and *b*. It is denoted by lcm(*a*,*b*).
- Prime factorisations can also be used to find the LCM of two integers.

$$\operatorname{lcm}(a,b) = p_1^{\max(a_1,b_1)} p_2^{\max(a_2,b_2)} \cdots p_n^{\max(a_n,b_n)}$$

- This number is divided by both *a* and *b* and no smaller number is divided by *a* and *b*.

Least Common Multiple (LCM)

Example: What is the lcm of $2^3 3^5 7^2$ and $2^4 3^3$?

Solution: $\operatorname{lcm}(2^3 3^5 7^2, 2^4 3^3) = 2^{\max(3,4)}\ 3^{\max(5,3)}\ 7^{\max(2,0)} = 2^4 * 3^5 * 7^2$.

- The greatest common divisor and the least common multiple of two integers are related by:
- **Theorem 5**: Let *a* and *b* be positive integers. Then

  $ab = \gcd(a,b) * \operatorname{lcm}(a,b)$

3.6 Euclidean Algorithm

- The **Euclidian algorithm** is an efficient method for computing the gcd of two integers.
- **Lemma 1**: Let $a = bq + r$,

  gcd(a,b) = gcd(b,r) where a > b and $r = a$ **mod** b.
- **Example**: Find gcd(91, 287):
  - 287 = (91 * 3) + 14
  - 91 = (14 * 6) + 7
  - 14 = (7 * 2) + 0
- gcd(91, 287) = gcd(91, 14) = gcd(14, 7) = 7

Euclid (325 B.C.E. – 265 B.C.E.)

- **Euclid** of Alexandria was a Greek mathematician, often referred to as the "founder of geometry" or the "father of geometry"

Euclidean Algorithm

Algorithm gcd(a, b: positive integers with $a > b$)
x = a, y = b
while (y ≠ 0):
    r = x **mod** y
    x = y
    y = r
return x where {gcd(a,b) is x}

- The number of divisions required to find the gcd is O(log b).

Euclidean Algorithm for GCD Exercise

Exercise 1: What is the gcd of 2322 and 654?

Exercise 2: What is the gcd of 27664 and 2054?

4. Applications of Congruences

- Hashing Functions
- Pseudorandom Numbers
- Check Digits (Parity Check Bits, UPCs, ISBNs)

4.1 Hashing Functions

- Suppose that a customer identification number is ten digits long.
- To retrieve customer files quickly, instead of using that ten-digit id number to assign a memory location, a smaller integer **key** associated to the id can be used.
- This can be done by **hashing**.
- Records or id numbers are identified using a **key**.

Hashing Functions

Definition: A hashing function h assigns memory location $h(k)$ to the record that has k as its key.

- A common hashing function is $h(k) = k$ **mod** m,
  where m is the number of available memory locations.

Hashing Functions

- A common hashing function is $h(k) = k$ **mod** m,

Example: Let $h(k) = k$ **mod** 111. This hashing function assigns the records of customers with id numbers as keys to memory locations in the following manner:

h(064212848) = 064212848 **mod** 111 = 14

h(037149212) = 037149212 **mod** 111 = 65

h(107405723) = 107405723 **mod** 111 = 14

but since location 14 is already occupied, the record is assigned to the next available position, which is 15.

Hashing Functions

- The hashing function is not one-to-one as there are many more possible keys than memory locations.
- When more than one record is assigned to the same location, a **collision** occurs.
- One way to solve a collision is to assign the record to the first free location after the occupied memory.
- For collision resolution, a **linear probing function** can be used:

  $h(k, i) = (h(k) + i)$ **mod** m, where i runs from 0 to $m - 1$.

- There are many other methods of handling with collisions.

4.2 Pseudorandom Numbers

- Randomly chosen numbers are needed for many purposes, including computer simulations.
- *Pseudorandom numbers* are not truly random since they are generated by systematic methods.
- The **linear congruential method** $ax \equiv b \pmod{m}$ is one commonly used procedure for generating pseudorandom numbers.
- Four integers are needed: the **modulus** m, the **multiplier** a, the **increment** c, and **seed** x_0, with $2 \le a < m$, $0 \le c < m$, $0 \le 0_x < m$

Pseudorandom Numbers

- We generate a sequence of pseudorandom numbers $\{x_n\}$, with $0 \le x_n < m$ for all n, by successively using the recursively defined function

  $$x_{n+1} = (a_{xn} + c) \textbf{ mod } m.$$

- If psudorandom numbers between 0 and 1 are needed, then the generated numbers are divided by the modulus, x_n / m.

Pseudorandom Numbers

- **Example**: Find the sequence of pseudorandom numbers generated by the linear congruential method with modulus $m = 9$, multiplier $a = 7$, increment $c = 4$, and seed $x_0 = 3$.

4.3 Check Digits: UPCs

- A common method of detecting errors in strings of digits is to add an extra digit at the end, which is evaluated using a function. If the final digit is not correct, then the string is assumed not to be correct.
- **Example**: Retail products are identified by their Universal Product Codes (UPCs). Usually these have 12 decimal digits, the last one being the check digit. The check digit is determined by the congruence:

  $3x_1 + x_2 + 3x_3 + x_4 + 3x_5 + x_6 + 3x_7 + x_8 + 3x_9 + x_{10} + 3x_{11} + x_{12} \equiv 0 \pmod{10}$.
- Suppose that the first 11 digits of the UPC are 79357343104. What is the check digit?
- Is 041331021641 a valid UPC?

5. Cryptography

- The subject of transforming information so that it cannot be easily recovered without special knowledge
- Number theory is the basis of many classical ciphers.
- These ciphers encrypt messages by changing each letter to a different letter. This can applies to block of letters.

5. Cryptography

- Classical Cryptography
  - Knowing how to encrypt allows decrypting.
- Cryptosystems
- Public Key Cryptography
  - Knowing how to encrypt and two large primes allow decrypting.
- RSA Cryptosystem
- Cryptographic Protocols

5.1 Classical Cryptography

Caesar Cipher

- Julius Caesar created secret messages by shifting each letter three letters forward in the alphabet (sending the last three letters to the first three letters.)
- For example, the letter B is replaced by E and the letter X is replaced by A. This process of making a message secret is an example of **encryption**.

Caesar Cipher

- Here is how the Caesar's encryption process works mathematically:
  - Replace each letter by an integer from $\mathbf{Z}_{26}$, that is an integer from 0 to 25 representing one less than its position in the alphabet.
  - For example, replace A by 0, and Z by 25.
  - The encryption function is $f(p) = (p + 3)$ **mod** 26.
  - It replaces each integer p in the set $\{0,1,2,\ldots,25\}$ by $f(p)$.
  - Replace each integer p by the letter with the position $p + 1$ in the alphabet.

Caesar Cipher Example

Example: What is the secret message produced from the message "WINTER IS COMING" using the Caesar cipher?

- First, replace the letters in the message with numbers.
- Replace each of these numbers p by $f(p) = (p + 3)$ **mod** 26.
- Translate back to letters

Caesar Cipher

- To recover the original message, use

  $f^{-1}(p) = (p - 3)$ **mod** 26.
- So, each letter in the coded message is shifted back three letters in the alphabet, with the first three letters sent to the last three letters.
- This process of recovering the original message from the encrypted message is called **decryption**.

Shift Cipher

- The Caesar cipher is one of a family of ciphers called **shift ciphers**.
- Letters can be shifted by an integer k, with 3 being just one possibility.
- The encryption function is

  $f(p) = (p + k)$ **mod** 26

  and the decryption function is

  $f^{-1}(p) = (p - k)$ **mod** 26

  The integer k is called a **key**.

Shift Cipher Example

Example: Decrypt the ciphertext message "FVB ZOHSS UVA WHZZ" encrypted with the shift cipher with shift k = 7?

- First, translate the letters back to elements of $\mathbf{Z}_{26}$.
- Shift each of these numbers by $-k$ = -7 **mod** 26.
- Translate back to letters

Affine Ciphers

- Shift ciphers are a special case of **affine ciphers** which use functions of the form

  $f(p) = (ap + b)$ **mod** 26,

  where a and b are integers, chosen so that f is a bijection (1:1).
- The function is a bijection if and only if gcd(a,26) = 1.
- **Example**: What letter replaces the letter K when the function $f(p) = (7p + 3)$ **mod** 26 is used for encryption?
- 10 = K, therefore, $f(10) = ((7 * 10) + 3)$ **mod** 26 = 21, which is V.

Affine Ciphers

- To decrypt a message encrypted by a shift cipher, the congruence $c \equiv ap + b \pmod{26}$ needs to be solved for p.
  - Subtract b from both sides to obtain $c - b \equiv ap \pmod{26}$.
  - Multiply both sides by the inverse $\bar{a}$ of a modulo 26, which exists since gcd(a, 26) = 1.
  - $\bar{a}(c - b) \equiv \bar{a}ap \pmod{26}$
  - $\bar{a}(c - b) \equiv p \pmod{26}$.
- $p \equiv \bar{a}(c - b) \pmod{26}$ is used to determine p in $\mathbf{Z}_{26}$.

Cryptanalysis of Shift Ciphers

- The process of recovering plaintext from ciphertext without knowledge both of the encryption method and the key is known as **cryptanalysis** or **breaking codes**.
- If we know that a ciphertext message was produced by a shift cipher, we can try to recover the message by shifting all characters of the ciphertext by each of the 26 possible shifts.
    - One of these is guaranteed to be the plaintext.

Cryptanalysis of Shift Ciphers

- Alternatively, an important tool for cryptanalysing ciphertext produced with a shift ciphers is the relative frequencies of letters.
- The nine most common letters in the English texts are E 13%, T 9%, A 8%, O 8%, I 7%, N 7%, S 7%, H 6%, and R 6%.

Cryptanalysis of Shift Ciphers

To analyse ciphertext:

- Find the relative frequency of the letters in the ciphertext.
- List the most common letters in frequency order.
- Hypothesize that the most frequent letter is produced by encrypting E.
- If the value of the shift from E to the most frequent letter is k, shift the ciphertext by $-k$ and see if it makes sense.
- If not, try T as a hypothesis and continue.

Cryptanalysis of Shift Ciphers Example

Example: We intercepted the message “ZNK KGXRE HOXJ MKZY ZNK CUXS” that we know was produced by a shift cipher. What was the original plaintext message?

Block Ciphers

- Ciphers that replace each letter of the alphabet by another letter are called *character* or *monoalphabetic* ciphers.
- They are vulnerable to cryptanalysis based on letter frequency. ***Block ciphers*** avoid this problem, by replacing blocks of letters with other blocks of letters.
- A simple type of block cipher is called the ***transposition cipher***.
- The key is a permutation σ of the set {1,2,...,m}, where m is an integer, that is a one-to-one function from {1,2,...,m} to itself.

Block Ciphers

- To encrypt a message, split the letters into blocks of size *m*, adding additional letters to fill out the final block.
- Encrypt $p_1,p_2,...,p_m$ as $c_1,c_2,...,c_m = p_{\sigma(1)},p_{\sigma(2)},...,p_{\sigma(m)}$.
- To decrypt the $c_1,c_2,...,c_m$, transpose the letters using the inverse permutation σ^{-1}.

Block Ciphers Example

- **Example**: Using the transposition cipher based on the permutation σ of the set {1,2,3,4} with σ(1) = 3, σ(2) = 1, σ(3) = 4, σ(4) = 2,
  a. Encrypt the plaintext PIRATE ATTACK
  b. Decrypt the ciphertext message SWUE TRAEOEHS, which was encrypted using the same cipher.

Cryptosystems

Definition: A cryptosystem is a five-tuple (*P*, *C*, *K*, *E*, *D*), where

- *P* is the set of plaintext strings,
- *C* is the set of ciphertext strings,
- *K* is the *keyspace* (set of all possible keys),
- *E* is the set of encryption functions, and
- *D* is the set of decryption functions.

Cryptosystems

- The encryption function in E corresponding to the key k is denoted by E_k.
- The decryption function in D that decrypts ciphertext that was encrypted using E_k is denoted by D_k.
- Therefore:

  $D_k(E_k(p)) = p$, for all plaintext strings p.

Shift Ciphers as a Cryptosystem

Assume the messages are strings consisting of elements in $\mathbf{Z}_{26}$.

- P is the set of strings of elements in $\mathbf{Z}_{26}$,
- C is the set of strings of elements in $\mathbf{Z}_{26}$,
- $K = \mathbf{Z}_{26}$,
- E consists of functions of the form
  $E_k(p) = (p + k)$ **mod** 26 , and
- D is the same as E where $D_k(p) = (p - k)$ **mod** 26.

5.2 Public Key Cryptography

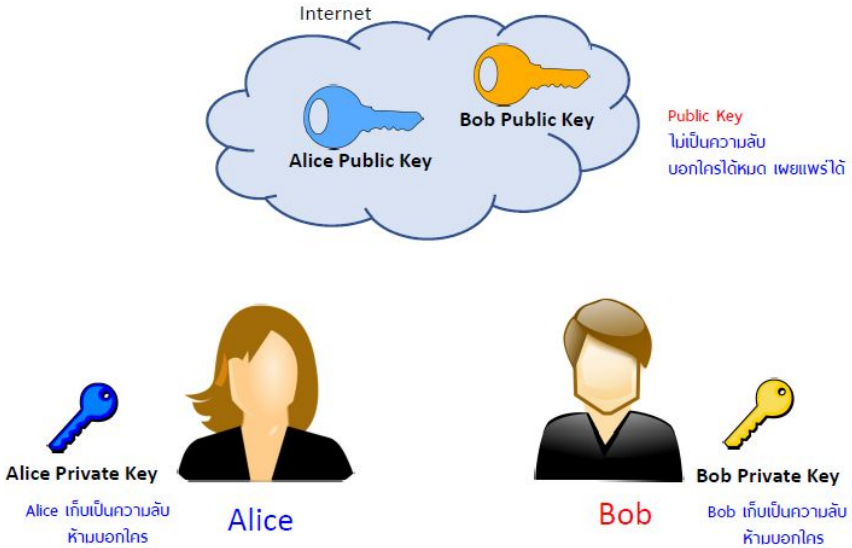
- All classical ciphers, including shift and affine ciphers, are **private key cryptosystems**. Knowing the encryption key allows one to quickly determine the decryption key.
- All parties who wish to communicate using a private key cryptosystem must share the key and keep it a secret.
- The shift cipher and affine cipher cryptosystems are quite simple and are extremely vulnerable to cryptanalysis.
- However, the Advanced Encryption Standard (AES), is extremely complex and is highly resistant to cryptanalysis.

Public Key Cryptography

To avoid the need for keys to be shared/exchanged:

- In public key cryptosystems, first invented in the 1970s, knowing how to encrypt a message does not help one to decrypt the message.
- Therefore, everyone can have a publicly known encryption key.
- The only key that needs to be kept secret is the decryption key.
- However, encryption and decryption can be extremely time-consuming. Impractical for many applications.

Most commonly used public key cryptosystem for secure data transmission is **RSA system**.



5.3 The RSA Cryptosystem

- A public key cryptosystem, now known as the **RSA system** was introduced in 1977 by three researchers at MIT.

Ronald Rivest (Born 1948) 

Adi Shamir (Born 1952) 

Leonard Adelman (Born 1945) 

- It is now known that the method was discovered earlier in 1973 by Clifford Cocks, working secretly for the UK government's GCHQ.

The RSA Cryptosystem

- Each individual creates and then publishes a public encryption key (n, e) based on two large prime numbers.
- $n = pq$ (the modulus) is the product of two large (300 digits) primes p and q, and an exponent e that is relatively prime to $(p - 1)(q - 1)$.
- The prime numbers must be kept secret.
- Anyone can use the public key to encrypt a message, but only someone with knowledge of the prime numbers (p and q) can decode the message.

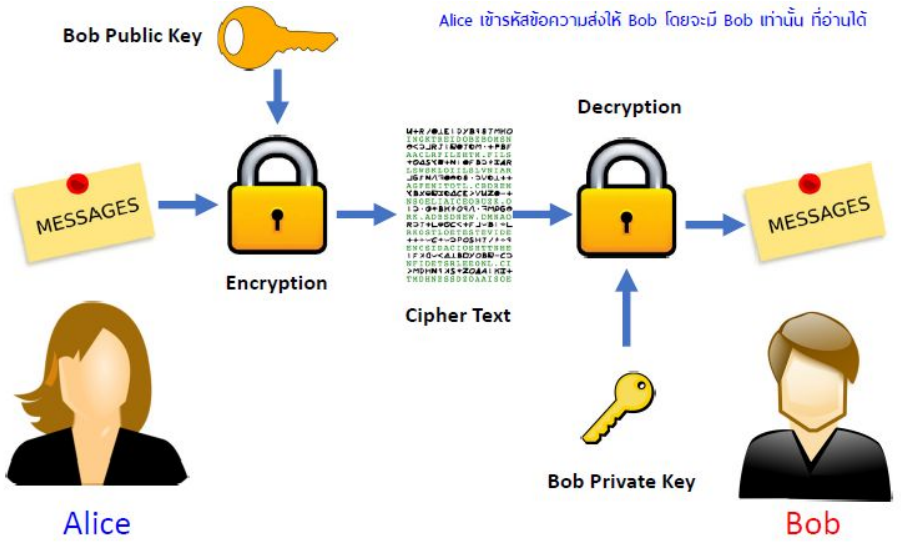
The RSA Cryptosystem

- To produce a usable key, two large primes (p and q) can be quickly found using probabilistic primality tests.
- However, $n = pq$, with approximately 600 digits, cannot be factored in a reasonable length of time (without a decryption key).
- Knowing the factorisation of the modulus n, that is p and q, the inverse d of e can be found and subsequently decrypt messages.

5.4 RSA Encryption

To encrypt a message using RSA using a key (n, e):

1. Translate the plaintext message M into sequences of two digit integers representing the letters.
   a. Use 00 for A, 01 for B, ..., and 09 for J.
2. Concatenate the two digit integers into strings of digits.
3. Divide this string into equally sized blocks of $2N$ digits where $2N$ is the largest even number 2525...25 with $2N$ digits that does not exceed n.
4. The plaintext message M is now a sequence of integers $m_1, m_2, \ldots, m_k$.
5. Each block m_i (an integer) is encrypted to a ciphertext block c_i using the function $c = m^e$ **mod** n.



RSA Encryption Example

Example: Encrypt the message STOP using the RSA cryptosystem with key (2537,13).

- n = 2537 = 43 * 59, p = 43 and q = 59 are primes and
- gcd(e,(p − 1)(q − 1)) = gcd(13, 42 * 58) = 1.
- Translate letters into numerical equivalents.
- Divide into blocks of four digits (because 2525 < 2537 < 252525)
- Then, encrypt each block using the mapping
  $c = m^{13}$ **mod** 2537.

5.5 RSA Decryption

- To decrypt a RSA ciphertext message, the decryption key d, an inverse of e modulo (p − 1)(q − 1) is needed.
- The inverse exists since gcd(e,(p − 1)(q −1)) = gcd(13, 42 * 58) = 1.
- With the decryption key d, we can decrypt each block with the computation $m = c^d$ **mod** p*q.
- RSA works as a public key system since the only known method of finding d is based on a factorization of n into primes. There is currently no known feasible method for factoring large numbers into primes.

RSA Decryption Example

Example: Decrypt the encrypted message 0981 0461 using the RSA cryptosystem with key (2537,13).

- n = 2537 = 43 * 59, p = 43 and q = 59 are primes and
- An inverse of 13 modulo 42 * 58 is 937 = d.

To decrypt a block c, compute

$m = c^{937}$ **mod** 2537.

5.6 RSA as a Public Key System

- No known method to decrypt messages without factorisation of n.
- Finding factorisation of n is a difficult problem, as opposed to finding large primes p and q, which can be done quickly.
- The most efficient factorisation methods known (as of 2017) require billions of years to factor 600-digit integers.
- When p and q are 300-digit primes, messages encrypted using $n = pq$ as the modulus cannot be decrypted in a reasonable time unless the primes p and q are known.
- RSA system is a block cipher, encrypting blocks of characters into blocks of characters.

RSA as a Public Key System

- When new factorisation techniques on large integers/primes are found, it is necessary to use large primes to ensure the secrecy of messages.
- With the steady increase of the speed of computers, the recommended size of the primes p and q used to produce a RSA public key has increased.
  - Large n results in slower RSA encryption and decryption.
  - It is less commonly used to directly encrypt user data.
- RSA system will be insecure once quantum computing is available.

RSA as a Public Key System

- More often, RSA passes encrypted shared/private keys for symmetric key cryptography which in turn can perform bulk encryption-decryption operations at much higher speed.
- The private keys are distributed to pairs of individuals when they wish to communicate.
  - This symmetric private key is used for both the encryption and decryption of messages.

5.7 Cryptographic Protocols

- **Cryptographic protocols** are exchanges of messages carried out by two or more parties to achieve a particular security goal.
- ***Key exchange*** is a protocol by which two parties/people can exchange a secret key over an insecure channel without having shared any information in the past.
- For example, for two people to send secure messages to each other using a private key cryptosystem they need to share a common key.

Cryptographic Protocols: Key Exchange

Diffie-Hellman key agreement protocol is described by example.

1. Suppose that Alice and Bob want to share a common key.
2. Alice and Bob agree to use a prime p and a primitive root a of p.
3. Alice chooses a secret integer k_1 and sends a^{k1} **mod** p to Bob.
4. Bob chooses a secret integer k_2 and sends a^{k2} **mod** p to Alice.
5. Alice computes $(a^{k2})^{k1}$ **mod** p. Bob computes $(a^{k1})^{k2}$ **mod** p.

At the end of the protocol, Alice and Bob have computed their shared key,

$(a^{k2})^{k1}$ **mod** $p = (a^{k1})^{k2}$ **mod** p.

Cryptographic Protocols: Key Exchange

- The messages sent in step 1 till 3 are assumed to be sent insecurely.
  - p, a, a^{k1} **mod** pi, and a^{k2} **mod** p are assumed to be public information.
- The protocol ensures that k_1, k_2, and the common key $(a^{k2})^{k1}$ **mod** $p = (a^{k1})^{k2}$ **mod** p are kept secret.
- To find the secret information from the public information would require the adversary to find k_1 and k_2 from a^{k1} **mod** p and a^{k2} **mod** p respectively.
  - This is an instance of the discrete logarithm problem, considered to be computationally infeasible when p and a are sufficiently large.

Cryptographic Protocols: Digital Signatures

- How a message can be sent so that a recipient of the message will be sure that the message came from the purported sender of the message?
- This can be accomplished using the RSA cryptosystem to apply a *digital signature* to a message.

Cryptographic Protocols: Digital Signatures

- Suppose that Alice's RSA public key is (n, e) and her private key is d.
- Alice encrypts a plain text message x using $E_{(n,e)}(x) = x^e$ **mod** n.
- She decrypts a ciphertext message y using $D_{(n,e)}(x) = y^d$ **mod** n.

Cryptographic Protocols: Digital Signatures

- Alice wants to send a message M so that everyone who receives the message knows that it came from her.
  - She translates the letters into their numerical equivalents and splits into blocks $\{m_1, m_2, ..., m_k\}$, just as in RSA encryption.
  - She then applies her decryption function $D_{(n,e)}$ to the blocks and sends the results to all intended recipients.
  - The recipients apply Alice's encryption function to each block.
    - Alice's key (n, e) is public information, anyone has this.
    - The result is the original plain text since $E_{(n,e)}(D_{(n,e)}(x)) = x$.

Cryptographic Protocols: Digital Signatures

- Example: Suppose Alice's RSA cryptosystem has a key(2537,13) in which $n = 2537 = 43 * 59$, so $p = 43$ and $q = 59$ are primes and exponent $e = 13$ as $\gcd(e,(p-1)(q-1)) = \gcd(13, 42 * 58) = 1$.
- Her decryption key is d = 937.
- She wants to send the message "MEET AT NOON" to her friends so that they can be certain that the message is from her.
- What should she send?