ECEN 227 - Machine Learning in Cybersecurity

Dr. Mahmoud Nabil

Dr. Mahmoud Nabil
mnmahmoud@ncat.edu

North Carolina A & T State University

October 28, 2020

Talk Overview

1. Introduction

2. Mathematical Background
   - Set of Residues
   - Finite Groups
   - Finite Fields

3. Basic Cryptographic Primitives
   - Symmetric key Cryptography
   - Asymmetric key Cryptography

Outline

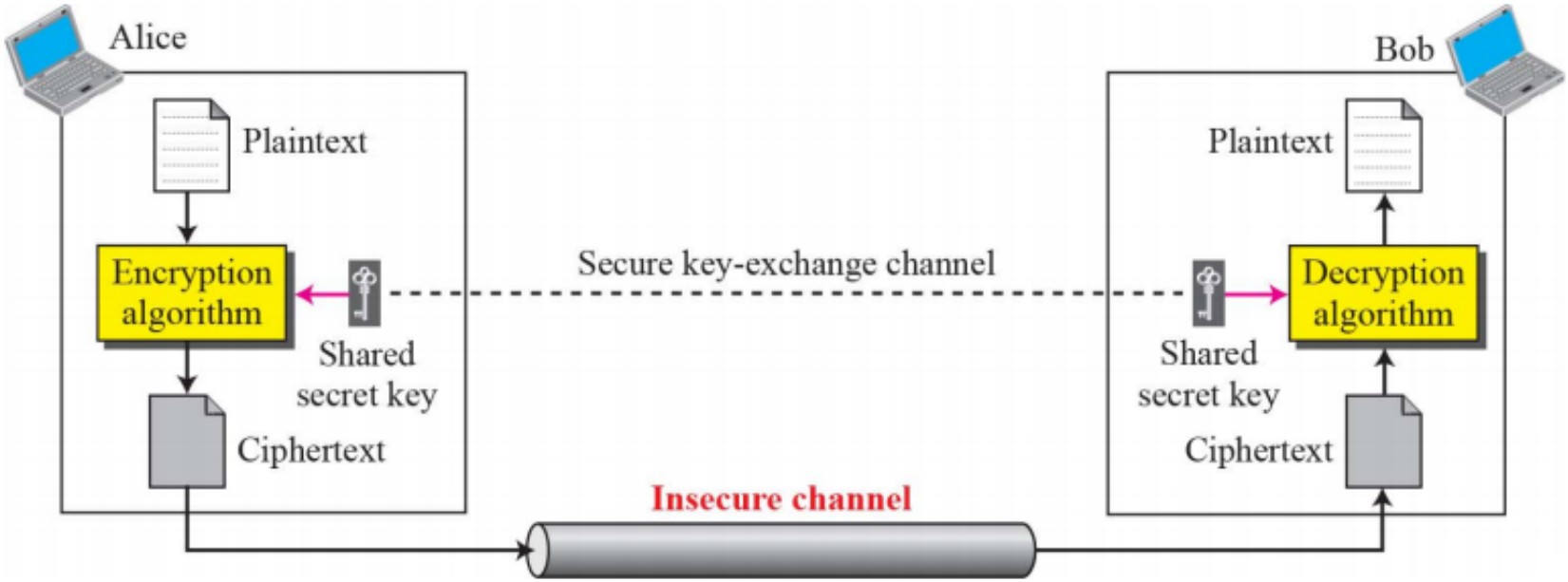
What is Cryptology?

- Cryptology
  - Cryptography
  - Cryptanalysis
- Cryptography, a word with Greek origins, means **secret writing**.
- However, we use the term to refer to the science and art of transforming messages to make them secure and immune to attacks.
- A cipher is a function which transforms a plaintext message into a ciphertext by a process called encryption.
- Plaintext is recovered from the ciphertext by a process called decryption.

Encryption / Decryption

“attack at midnight” - *plaintext*

“buubdl bu njeojhiu” - *ciphertext*



Cryptanalysis

- The science and study of breaking ciphers, i.e., the process of determining the plaintext message from the ciphertext
- Objective to recover key not just message
- A common approach is brute-force attack - try all possible cases

So what is security?

Security

Security is about how to prevent attacks, or – if prevention is not possible – how to detect attacks and recover from them.

- **Passive Attacks**
  - **Eavesdropping:** The attacker simply listens and tries to interpret the data being exchanged - if the data is in-the- clear, they succeed
  - **Traffic Analysis:** the attacker gains information by determining how much activity is there, where access points are located,
  - Difficult to detect. Should be prevented.
- **Active Attacks**
  - Attempts to alter system resources or affect their operation, examples: masquerade (spoofing), replay, modification (substitution, insertion, destruction), denial of service.
  - Difficult to prevent. Should be detected.

Conditional/Uncoditional Security

- **Conditional security** means that the system is secure under certain conditions and assumptions for instance:
  - Computationally secure systems needs hundred of years to be broken
  - Most of the existing systems are computationally secure.
- **UnConditional security** means no matter the attacker capabilities the system can withstand any attack
  - Post quantum cryptography.

Kerckhoff's Principle

Generally assumed that the attacker knows everything about the cryptosystem except the key.

Note

If, for security, the system requires that details of the system be kept secret, it is not considered secure.

Outline

Outline

Set of Residues

One of the most important structures in crypt: Residues modulo n

Let n be a positive integer $n > 1$ and Z_n represent the set of reminder of all integers on division n, then

$$Z_n = \{0, 1, 2, \ldots, n-1\}$$

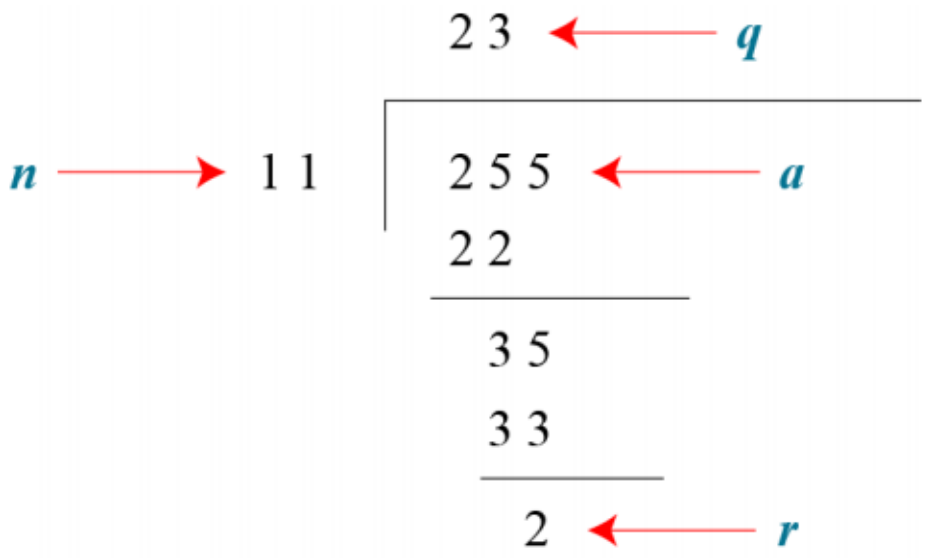
We define $a + b$ and $a \times b$ the ordinary sum and product of a and b reduced by modulo n respectively. Let

$$Z_n^* = \{a \in Z_n | a \neq 0\}$$

Example

255 mod 11 = 2

27 mod 5 = 2
36 mod 12 = 0



First Property:	$(a+b) \bmod n = [(a \bmod n) + (b \bmod n)] \bmod n$
Second Property:	$(a-b) \bmod n = [(a \bmod n) - (b \bmod n)] \bmod n$
Third Property:	$(a \times b) \bmod n = [(a \bmod n) \times (b \bmod n)] \bmod n$

(a+n) mod n = a

Set of Residues

The modulo operation creates a set, which in modular arithmetic is referred to as the set of least residues modulo n, or Z_n.

$$Z_n = \{0, 1, 2, \ldots, n-1\}$$

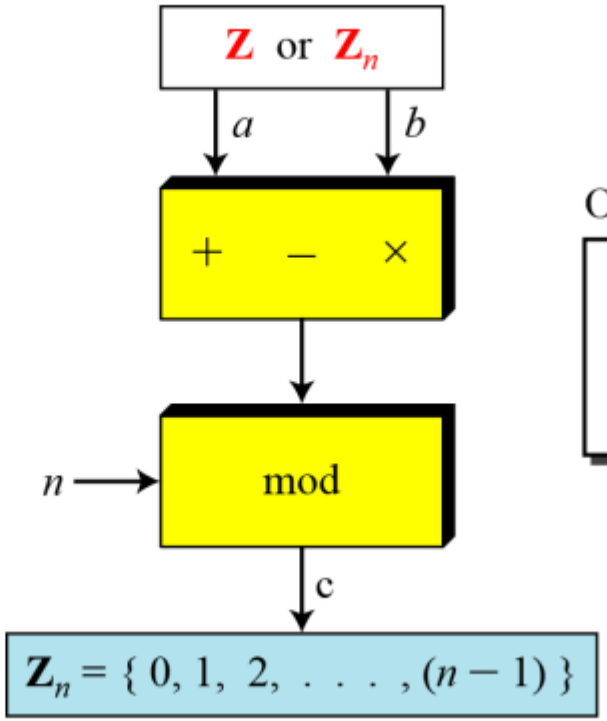
Ex.

- $Z_2 = \{0, 1, 2\}$ prime residue
- $Z_6 = \{0, 1, 2, 3, 4, 5, 6\}$
- $Z_{11} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}$ prime residue

Operation in Z_n

The three binary operations that we discussed for the set Z can also be defined for the set Z_n. The result may need to be mapped to Z_n using the mod operator.

Binary operations in Z_n



Operations

$(a + b) \bmod n = c$

$(a - b) \bmod n = c$

$(a \times b) \bmod n = c$

Example

Perform the following operations (the inputs come from Zn):

1. Add 7 to 14 in $\mathbf{Z}_{15}$.
2. Subtract 11 from 7 in $\mathbf{Z}_{13}$.
3. Multiply 11 by 7 in $\mathbf{Z}_{20}$.

Sol.

1. (14+7) mod 15 $\rightarrow$ 21 mod 15 = 6
2. (7-11) mod 13 $\rightarrow$ (-4) mod 13 = 9
3. (7×11) mod 20 $\rightarrow$ 77 mod 20 = 17

Inverses

When we are working in modular arithmetic, we often need to find the inverse of a number relative to an operation.

- An additive inverse (relative to an addition operation)
- A multiplicative inverse (relative to a multiplication operation).

Additive Inverse

In Z_n, two numbers a and b are additive inverses of each other if

$$a + b \equiv 0 \mod n$$

In modular arithmetic, each integer has an additive inverse. The sum of an integer and its additive inverse is congruent to 0 modulo n.

Ex.
Find all additive inverse pairs in Z_{10}
Sol.

The six pairs of additive inverses are (0, 0), (1, 9), (2, 8), (3, 7), (4, 6), and (5, 5).

Multiplicative Inverse

In Z_n, two numbers a and b are the multiplicative inverse of each other if

$$a \times b \equiv 1 \bmod n$$

In modular arithmetic, an integer may or may not have a multiplicative inverse. When it does, the product of the integer and its multiplicative inverse is congruent to 1 modulo n.

Ex.
Find all multiplicative inverse pairs in Z_{11}
Sol.

The seven pairs of multiplicative inverses are: (1, 1), (2, 6), (3, 4), (5, 9), (7, 8), (9, 5), and (10, 10).

Cryptography often uses these two sets: Z_p and Z_p^*. The modulus in these two sets is a prime number.

Outline

Groups

A set of objects, along with a binary operation (meaning an operation that is applied to two objects at a time) on the elements of the set, must satisfy/has the following four properties.

1. **Closure** with respect to the operation. Closure means that if a and b are in the set, then the element $a \circ b = c$ is also in the set. The symbol $\circ$ denotes the operator for the desired operation.
2. **Associativity** with respect to the operation. Associativity means that $(a \circ b) \circ c = a \circ (b \circ c)$
3. **A unique identity element** with regard to the operation $\circ$. An element i would be called an identity element if for every a in the set, we have $a \circ i = a$.
4. **An Inverse element** for each element with regard to the operation. That is, for every a in the set, the set must also contain an element b such that $a \circ b = i$. assuming that i is the identity element.

Groups

- In general, a group is denoted by $\{G, \circ, i\}$ where G is the set of objects and $\circ$ is the operator.
- Infinite groups, meaning groups based on sets of infinite size.
- A finite group contains finite number of elements. The number of elements in G is called the group order and is denoted as $|G|$

Ex.

- The set of all integers: positive, negative, and zero - along with the operation of arithmetic addition constitutes a group
- The set of all even integers : positive, negative, and zero under the operation of arithmetic addition is a group If the operation on the set elements is commutative, the group is called an abelian group. An operation $\circ$ is commutative if $a \circ b = b \circ a$.

Example

- $\mathbf{Z}$, the set consisting of all integers.
- $\mathbf{Q}$, the set consisting of all rational numbers.
- $+$ and $\times$ are ordinary addition and multiplication.

Then

- $(\mathbf{Z}, +, 0)$, $(\mathbf{Q}, +, 0)$, $(\mathbf{Q^*}, \times, 1)$ are all groups where $\mathbf{Q^*}$ is the set of all nonzero rational numbers.
- Furthermore, the are abelian.

How about $(\mathbf{Z^*}, \times, 1)$?

Notes on Groups

If the group operation is addition, the group also allows for subtraction. Similarly, multiplicative groups allow division.

- A group is guaranteed to have a special element called the identity element. The identity element of a group is frequently denoted by the symbol 0.

Notes on Groups

If the group operation is addition, the group also allows for subtraction. Similarly, multiplicative groups allow division.

- A group is guaranteed to have a special element called the identity element. The identity element of a group is frequently denoted by the symbol 0.
- As you now know, for every element a, the group must contain its inverse element b such that $a + b = 0$, where the operator $+$ is the group operator.

Notes on Groups

If the group operation is addition, the group also allows for subtraction. Similarly, multiplicative groups allow division.

- A group is guaranteed to have a special element called the identity element. The identity element of a group is frequently denoted by the symbol 0.
- As you now know, for every element a, the group must contain its inverse element b such that $a + b = 0$, where the operator $+$ is the group operator.
- So if we maintain the illusion that we want to refer to the group operation as addition, we can think of b in the above equation as the additive inverse of a and even denote it by -a. We can therefore write $a + (-a) = 0$ or more compactly as $a - a = 0$.

Notes on Groups

If the group operation is addition, the group also allows for subtraction. Similarly, multiplicative groups allow division.

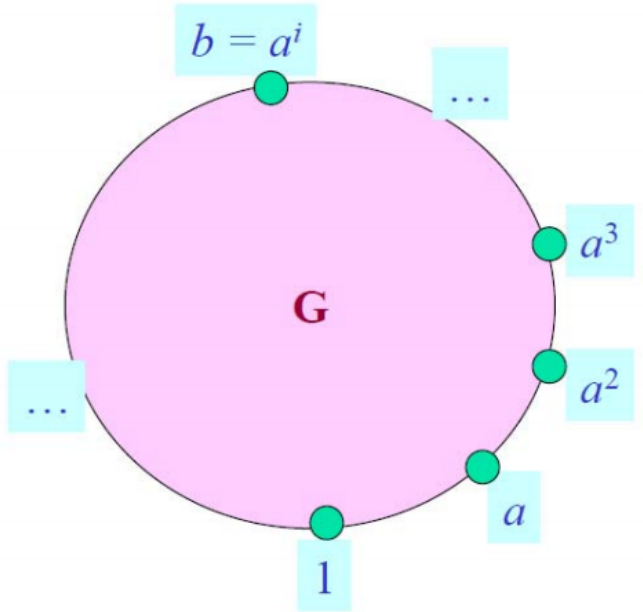
- A group is guaranteed to have a special element called the identity element. The identity element of a group is frequently denoted by the symbol 0.
- As you now know, for every element a, the group must contain its inverse element b such that $a + b = 0$, where the operator $+$ is the group operator.
- So if we maintain the illusion that we want to refer to the group operation as addition, we can think of b in the above equation as the additive inverse of a and even denote it by -a. We can therefore write $a + (-a) = 0$ or more compactly as $a - a = 0$.
- In general $a - b = a + (-b)$ where -b is the additive inverse of b with respect to the group operator $+$. We may now refer to an expression of the sort a - b as representing subtraction

Cyclic Group

Cyclic Group

A multiplicative group is said to be cyclic if there is an element $a \in G$ such that for any $b \in G$ there is some integer i with $b = a^i$ Such an element a is called the group generator of the cyclic group, and we write $G =< a >$

Ex.



Examples

- $(G_3^*, \times, 1)$, cyclic group with generator 2.

$$G_3^* = \{1,2\} = <2> = \{2^0 = 1, 2^1 = 2\},\ 2^2 = 1 \bmod 3$$

- $(G_7^*, \times, 1)$, cyclic group with generator 3.

$$G_7^* = \{0,\ 1,\ 2,\ 3,\ 4,\ 5,\ 6\}$$

$$G_7^* = <3> = \{3^0 = 1,\ 3^2 = 2,\ 3^1 = 3,\ 3^4 = 4,\ 3^5 = 5,\ 3^3 = 6,\ 3^6 = 1\}$$

- $(G_5^*, \times, 1)$, cyclic group with generator 3.

$$G_5^* = \{0,\ 1,\ 2,\ 3,\ 4\}$$

$$G_5^* = <2> = \{2^0 = 1,\ 2^1 = 2,\ 2^3 = 3\ ,2^2 = 4\}$$

Problems that are believed to be hard in G_p^*

- Let g be a generator of G_p^*. Given x $\in G_p^*$ find an r such that x $= g^r$ mod p. This is known as the **Discrete log problem**.
- Let g be a generator of G_p^*. Given x, y $\in G_p^*$ where x $= g^{r_1}$ and y $= g^{r_2}$. Find z $= g^{r_1 r_2}$. This is known as the **Diffie-Hellman problem**.

Outline

Fields

A field is a set of elements with addition and multiplication operations satisfying these rules:

1. **Two operations** (addition and multiplication) are defined on every element in the field.
2. **Closure** The addition/multiplication of two elements in the field gives an element in the field.
3. **Associativity** a+(b+c) = (a+b)+c and a(bc) = (ab)c, where a, b and c are elements in the field.
4. **Commutativity** a+b = b+a and ab = ba
5. **Additive identity** Theres a single element denoted 0 such that a + 0 = a for all elements in the field.
6. **Multiplicative identity** There is an element denoted 1 that a1 = a for every element in the field.
7. **Multiplicative inverse** For a given a, where a≠ 0 the multiplicative inverse is designated as a^{-1}

Examples of Fields

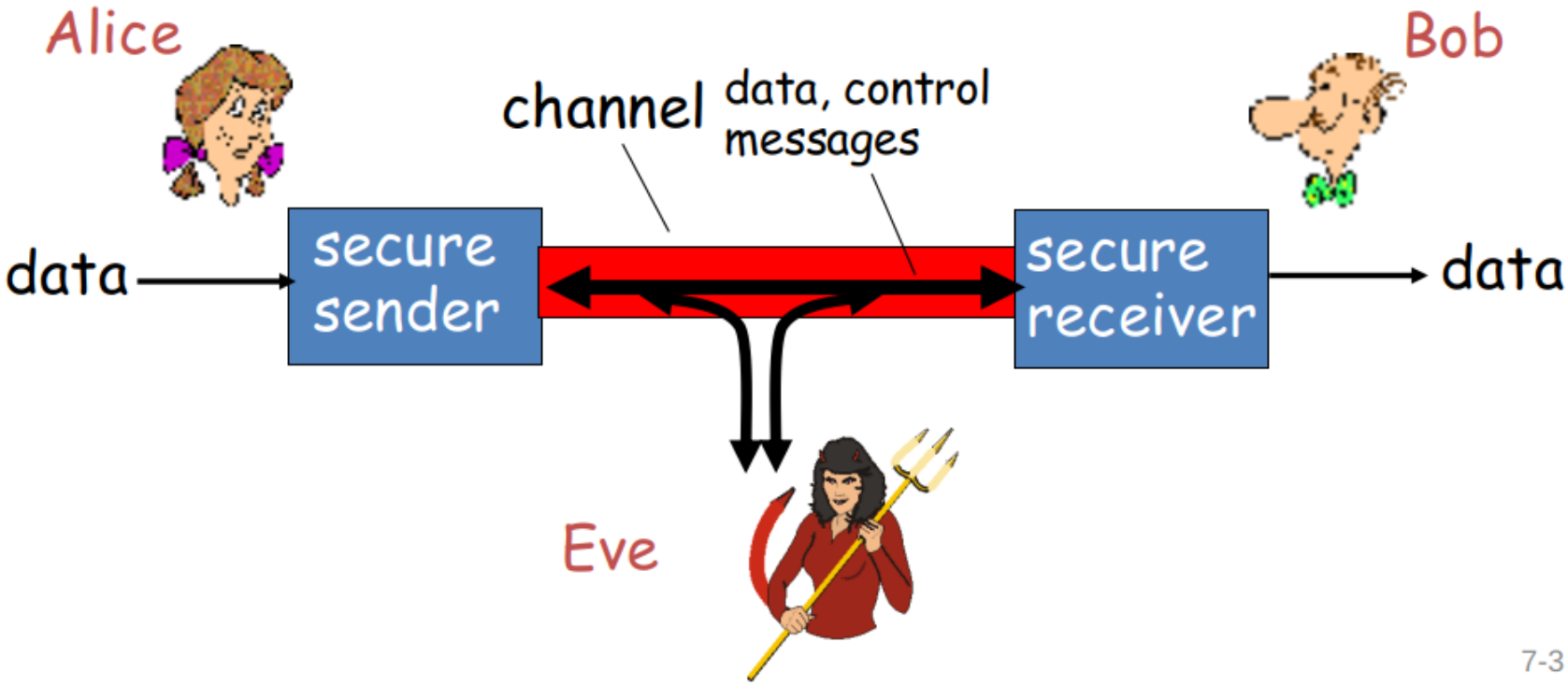
1. The set of all real numbers under the operations of arithmetic addition and multiplication
   - Is a field.
2. The set of all rational numbers under the operations of arithmetic addition and multiplication
   - Is a field.
3. The set of all even integers, positive, negative, and zero, under the operations arithmetic addition and multiplication.
   - **NOT** a field.
4. The set of all integers under the operations of arithmetic addition and multiplication
   - **NOT** a field.

Outline

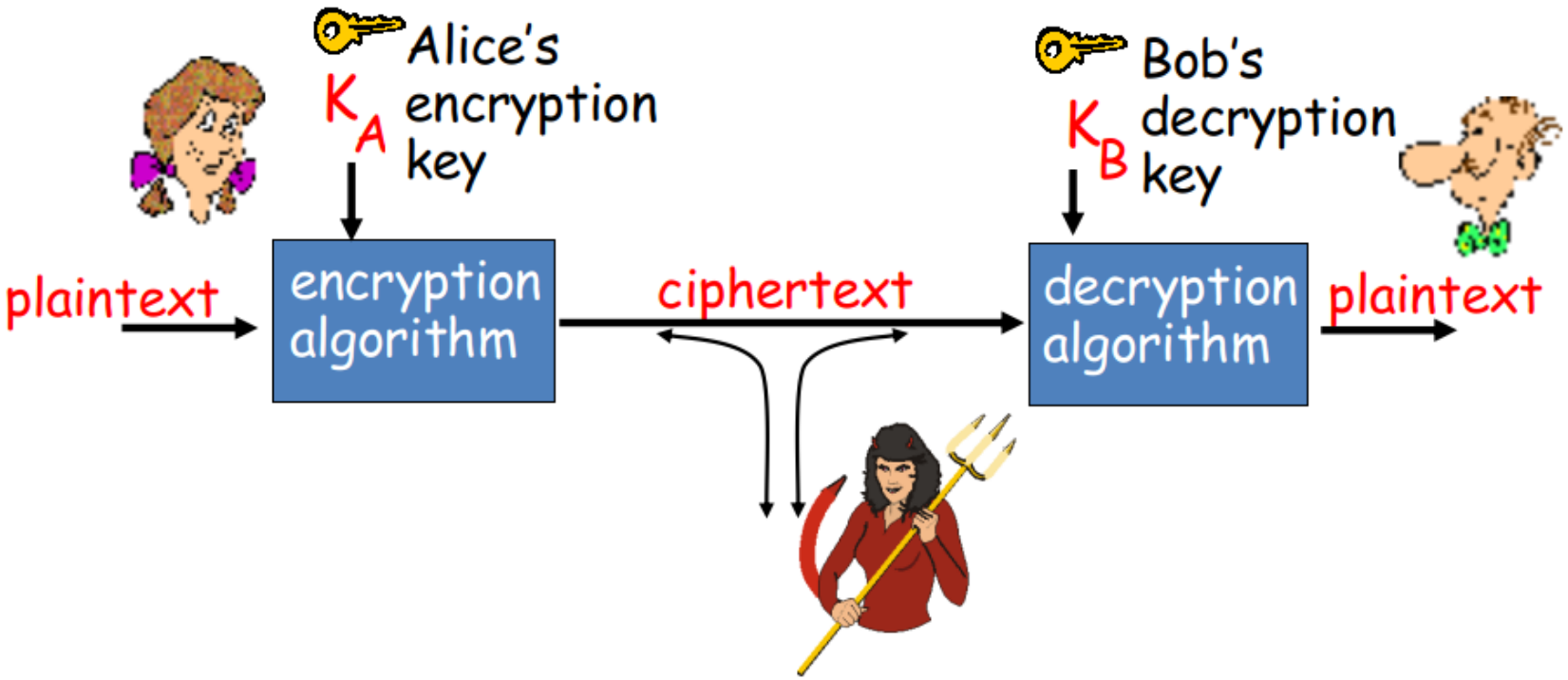
Outline

Friends and Enemies

- Bob, Alice (lovers!) want to communicate securely
- Eve (or Trudy, intruder) may intercept, delete, add messages



The Language of Cryptography



- Symmetric key crypto: sender, receiver keys identical
- Public-key crypto: encryption key (public), decryption key secret (private)

Diffie-Hellman Key Exchange Protocol

Public System Parameters

p: is a large prime number
g: a generator of Z_p^*

Alice

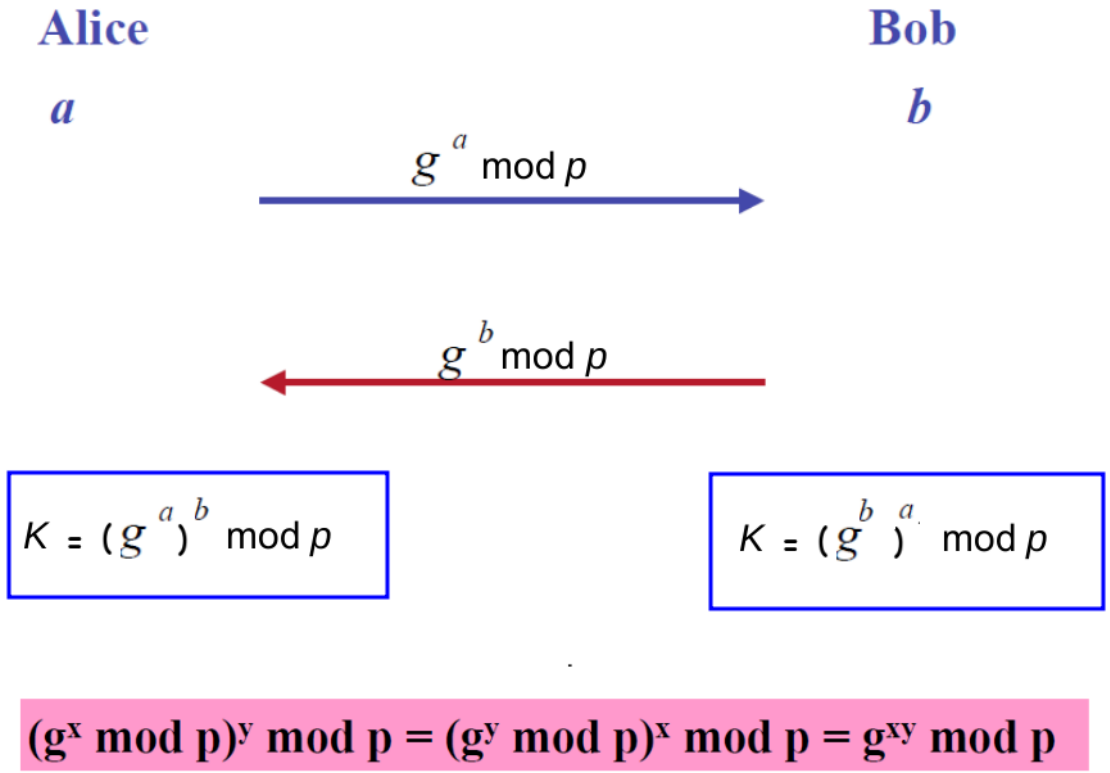
Private key: a, $0 < a < p$
Public key: g^a

Bob

Private key: b, $0 < b < p$
Public key: g^b

a and b are secret and should be large
p and g are public

Diffie-Hellman Key Exchange Protocol

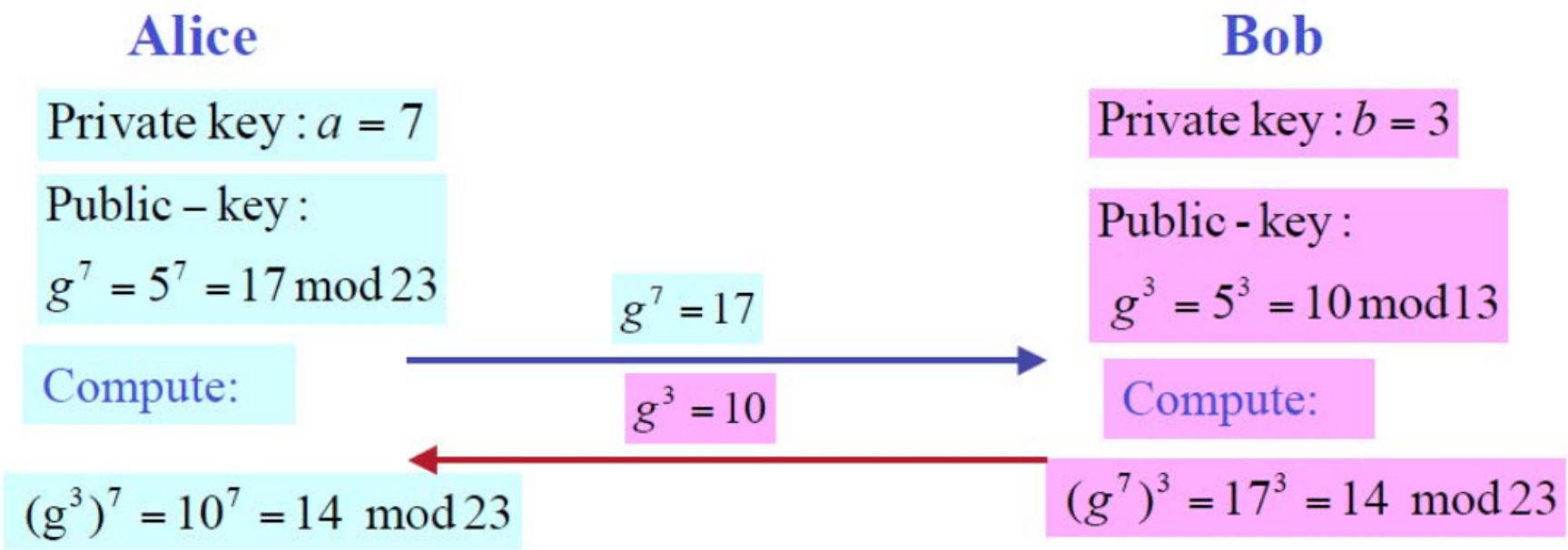


$$(g^x \bmod p)^y \bmod p = (g^y \bmod p)^x \bmod p = g^{xy} \bmod p$$

- K is the session symmetric key.
- a and b are not sent in clear because they are secret.
- Sending g^a and g^b is secure because it is not feasible to know a given g and g^a

Diffie-Hellman Example

Example Let $p = 23$. Then $g = 5$ is a primitive element of GF(p).



The secret information shared by Alice and Bob is 14.

Attacker: known $\left.\begin{matrix} g^7 = 17 \\ g^3 = 10 \end{matrix}\right\} \Rightarrow g^{21} = 14?$

Diffie-Hellman Summary

Diffe-Hellman Problem:

Given g^a and g^b, compute g^{ab}.

Thus the Diffe-Hellman key exchange scheme is secure if the DH problem is computationally infeasible.

The DH problem is computational feasible if the solving discrete logarithm in $GF(p)$ is computationally feasible.

Thus, we may say that the security of the DH key exchange scheme is based on the difficulty of solving discrete logarithm in the finite field GF(p).

Brute Force Attack

- Try all possible keys K and determine if $D_K(C)$ is a likely plaintext
  - Requires some knowledge of the structure of the plaintext (e.g., PDF file or email message)
- Key should be a sufficiently long random value to make exhaustive search attacks unfeasible

Symmetric Cryptosystem

- **Scenario**
  - Alice wants to send a message (plaintext P) to Bob.
  - The communication channel is insecure and can be eavesdropped
  - If Alice and Bob have previously agreed on a symmetric encryption scheme and a secret key K, the message can be sent encrypted (ciphertext C)
- **Issues**
  - What is a good symmetric encryption scheme?
  - What is the complexity of encrypting/decrypting?
  - What is the size of the ciphertext, relative to the plaintext?

Symmetric Key Cryptography Basics

- **Notation**
  - Secret key K
  - Encryption function $E_K(P)$
  - Decryption function $D_K(C)$
  - Plaintext length typically the same as ciphertext length
  - Encryption and decryption are one-to-one mapping functions on the set of all n-bit arrays
- **Efficiency**
  - functions E_K and D_K should have efficient algorithms
- **Consistency**
  - Decrypting the ciphertext yields the plaintext
  - $D_K(E_K(P)) = P$

Classical Cryptography

- Transposition Cipher
- Substitution Cipher
  - Simple substitution cipher (Caesar cipher)
  - Vigenere cipher
  - One-time pad

Transposition Cipher: rail fence

- Write plaintext in two or more rows
- Generate ciphertext in column order

Example:
"HELLOWORLD"

HLOOL
ELWRD

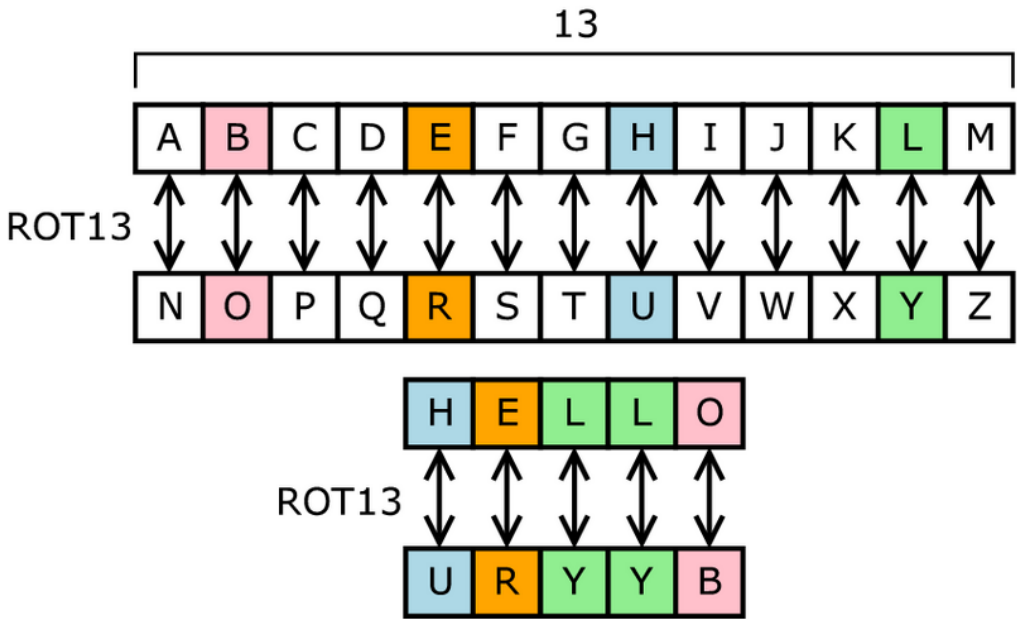
ciphertext: HLOOLELWRD

Note

Problem: does not affect the frequency of individual symbols

Substitution Ciphers

- Each letter is uniquely replaced by another.
- There are 26! possible substitution ciphers for English language.
- Also know as **Caesar Cipher**
- One popular substitution cipher for some Internet posts is ROT13.

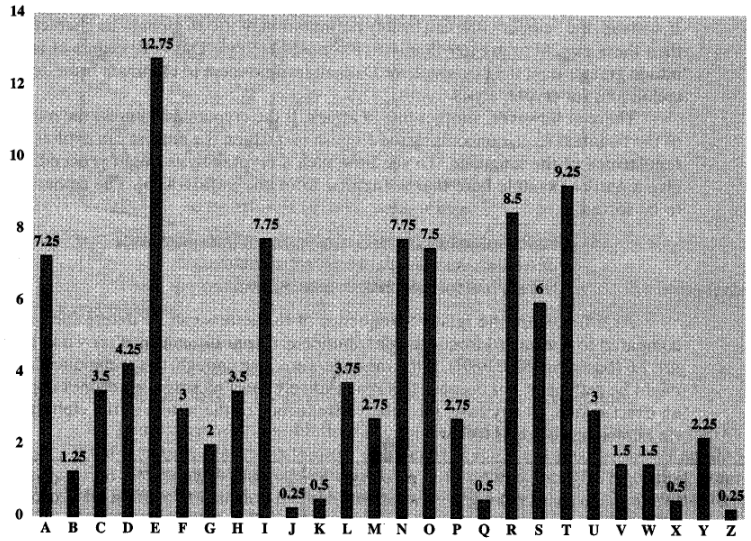


Frequency Analysis

- Letters in a natural language, like English, are not uniformly distributed.
- Knowledge of letter frequencies, including pairs and triples can be used in cryptologic attacks against substitution ciphers.

a:	8.05%	b:	1.67%	c:	2.23%	d:	5.10%
e:	12.22%	f:	2.14%	g:	2.30%	h:	6.62%
i:	6.28%	j:	0.19%	k:	0.95%	l:	4.08%
m:	2.33%	n:	6.95%	o:	7.63%	p:	1.66%
q:	0.06%	r:	5.29%	s:	6.02%	t:	9.67%
u:	2.92%	v:	0.82%	w:	2.60%	x:	0.11%
y:	2.04%	z:	0.06%				

8.1: Letter frequencies in the book *The Adventures of Tom Sawyer*, by Twain.



Vigenere Cipher

- Idea: Uses Caesar's cipher with various different shifts, in order to hide the distribution of the letters.
- A key defines the shift used in each letter in the text
- A key is repeated as many times as required to become the same length of the plaintext.

Example.

Plain text: l a t t a c k
Key: 2 3 4 2 3 4 2 (key is "234")
Cipher text: K d x v d g m

Problem of Vigenere Cipher

- Vigenere is easy to break (Kasiski, 1863):
- Assume we know the length of the key. We can organize the ciphertext in rows with the same length of the key. Then, every column can be seen as encrypted using Caesar's cipher.
- Length of the key can be induced using several techniques.

One-Time Pads

- Extended from Vigenere cipher
- There is one type of substitution cipher that is absolutely unbreakable.
- The one-time pad was invented in 1917 by Joseph Mauborgne and Gilbert Vernam We use a block of shift keys, $(k_1, k_2, ..., k_n)$, to encrypt a plaintext, M, of length n, with each shift key being chosen uniformly at random.
- Since each shift is random, every ciphertext is equally likely for any plaintext.

Note that

One time pad is not practical to implement. Why?

- The key has to be as long as the plaintext

Desirable Properties of Ciphers

Security

1. **Diffusion** Process of spreading effect of plaintext or key as widely as possible over ciphertext

Avalanche effect

Approximately half of the ciphertext bits change (at random) in response to a one bit change in the plaintext or the key.

2. **Confusion** The relationship between key and ciphertext bits should be complicated. Ciphertext and plaintext should appear to be statistically independent

Efficiency

1. High encryption and decryption rate.
2. Simplicity (easier to implement and analyze).
3. Suitability for hardware or software.
4. Key size should be small, but large enough to preclude exhaustive key search.

Block Ciphers in Practice

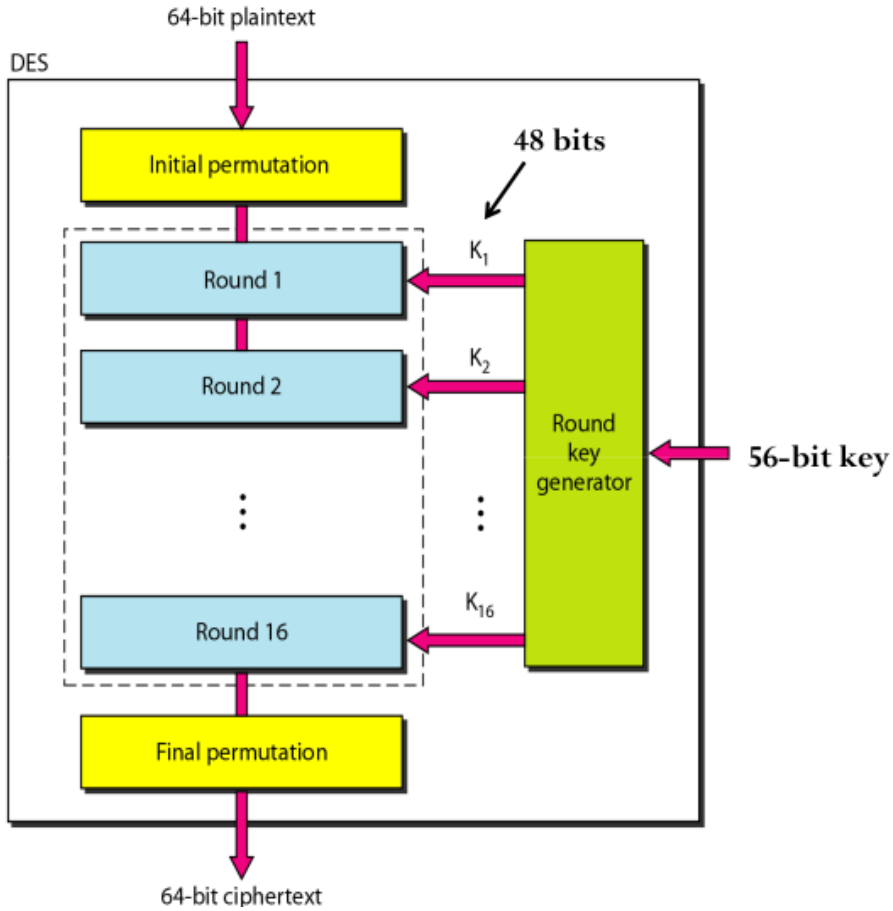
- **Data Encryption Standard (DES)**
  - Developed by IBM and adopted by NIST in 1977 64-bit blocks and 56-bit keys
  - Small key space makes exhaustive search attack feasible since late 90s
- **Advanced Encryption Standard (AES)**
  - Selected by NIST in 2001 through open international competition and public discussion
  - Several possible key lengths: 128, 192 and 256 bits
  - Exhaustive search attack not currently possible
  - AES-256 is the symmetric encryption algorithm of choice

Data Encryption Standard (DES)

Underling principle: Take something simple and use it several times; hope that the result is complicated.

- Easy to implement. The code of one round can be repeated.

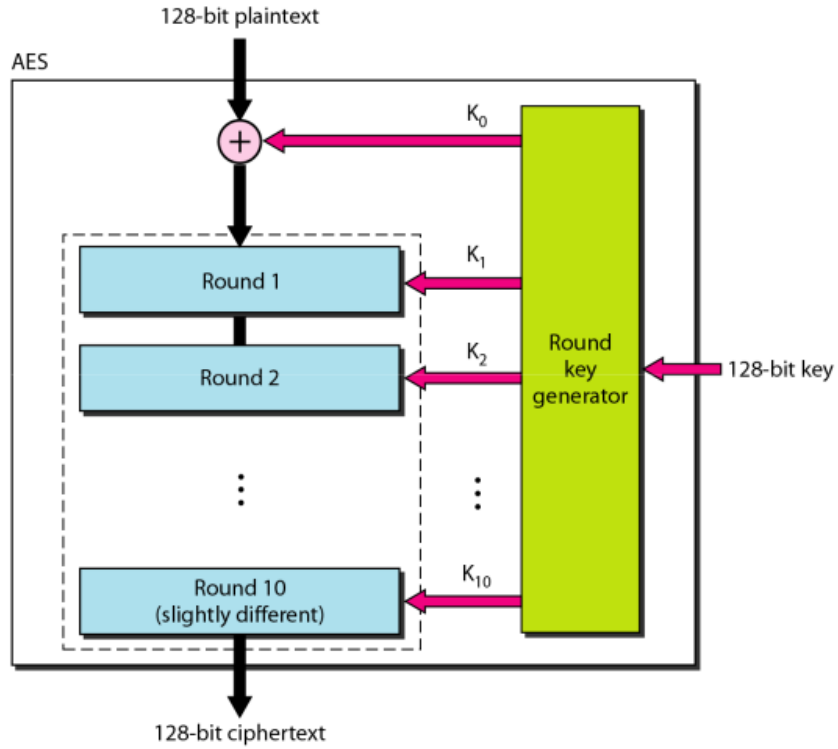
It was observed that alternating rounds of simple substitutions and transpositions could produce a strong cipher (even though individual operations are not strong).



Each round is simply some substitution and permutation operation

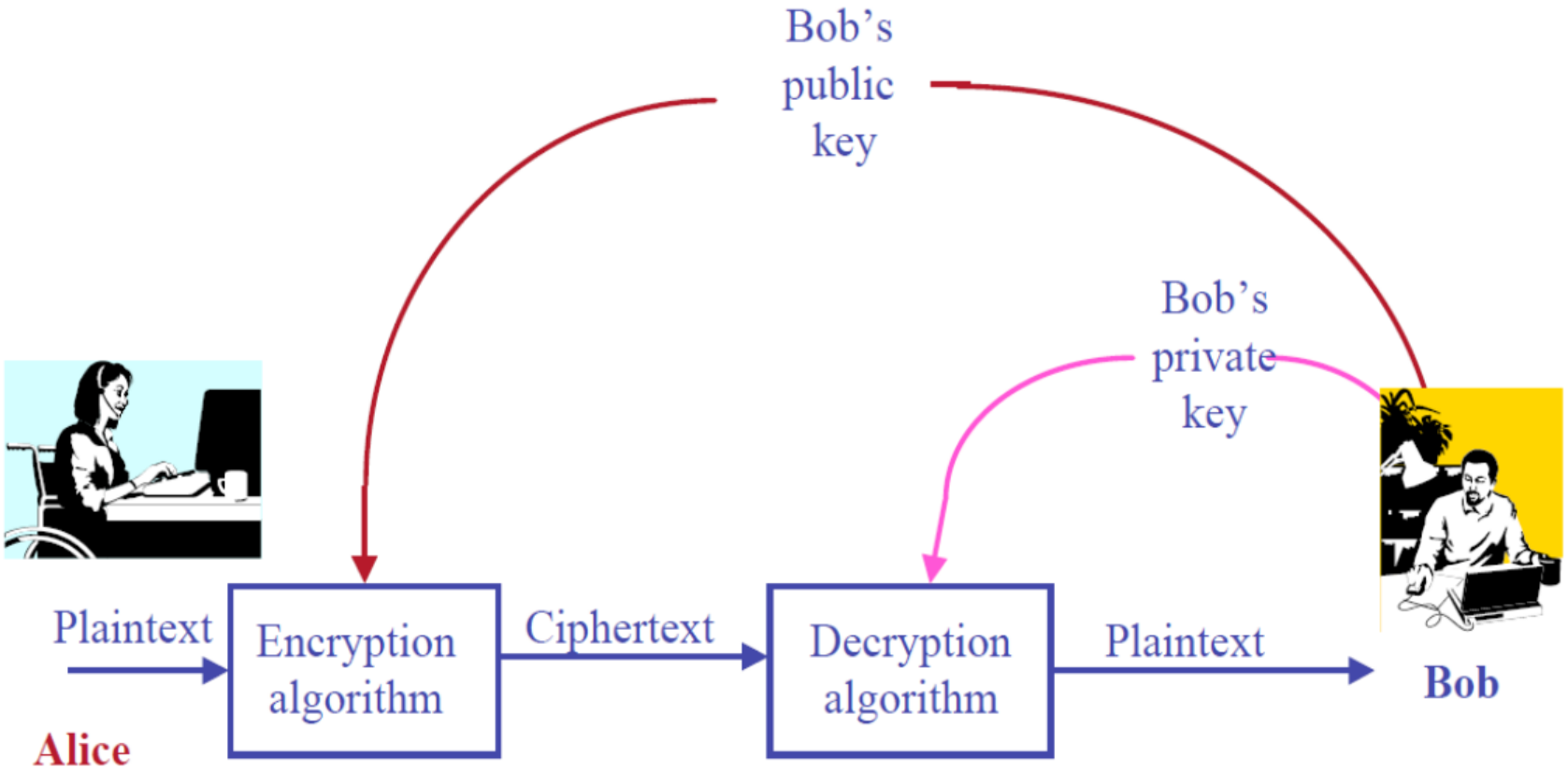
Advanced Encryption Standard (AES)

- Same underlying principle
- AES operates on 128-bit blocks. It is designed to be used with keys that are 128, 192, or 256 bits long, yielding ciphers known as AES-128, AES-192, and AES-256.



Each round is simply some substitution and permutation operation

Outline



- Symmetric-key and asymmetric-key ciphers are complements of each other; the advantages of one can compensate for the disadvantages of the other.
- Asymmetric-key ciphers are sometimes called public-key ciphers.

Public Key Cryptosystms Security

Public Key cryptosystems are base on the difficulty of a computational problems.

1. **Ex. Factoring Large Integers**
   - It is easy to compute n $= p \times q$ given two large primes p and q, but it is hard to find p and q given n. Used in RSA

RSA

- RSA is one of the first public-key cryptosystems. First published 1977.
- Depends on the difficulty of factorizing large prime numbers.

Choosing Keys:

1. Choose two large prime numbers p, q. (e.g., 1024 bits each)
2. Compute n = pq, z = (p-1)(q-1)
3. Choose e (with e<n) that has no common factors with z. (e, z are relatively prime).
4. Choose d such that ed mod z = 1 . (i.e., e and d are multiplicative inverses with respect to modulo z)
5. Public key is (n,e). Private key is (n,d)

RSA: Encryption, decryption

1. Given (n, e) and (n, d) as computed above
2. To encrypt bit pattern, m, compute $c = m^e \bmod n$ (i.e., remainder when m is divided by n)
3. To decrypt received bit pattern, c, compute $m = c^d \bmod n$ (i.e., remainder when c is divided by n)

Magic Happens

$$m = \underbrace{(m^e \bmod n)^d}_{c} \bmod n$$

RSA example

Bob chooses p=5, q=7. Then n=35, z=24.

e=5 (so e, z are relatively prime).
d=29 (so ed = 1 mod z).

- **Encrypt**
  - Let m =12
  - $m^e = 12^5 = 1524832$
  - m^e mod n = 17
- **Decrypt**
  - Let c =17
  - $c^d = 17^{29} = 481968572106750915091411825223071697$
  - c^d mod n = 12

Computationally expensive!

Why it works?

$$m = \underbrace{(m^e \bmod n)^d}_{c} \bmod n$$

Useful number theory theorem: If p,q prime and n = pq, then:

$$x^y \bmod n = x^{y \bmod (p-1)(q-1)} \bmod n$$

$$\begin{aligned}(m^e \bmod n)^d \bmod n &= m^{ed} \bmod n \\ &= m^{ed \bmod (p-1)(q-1)} \bmod n \\ &= m^1 \bmod n \\ &= m\end{aligned}$$

remember ed are multiplicative inverses for modulo (p-1)(q-1)

RSA First Property

$$E_e(D_d(m)) = m = E_d(D_e(m))$$

1. use public key first, followed by private key
2. use private key first, followed by public key

Result is the same in both cases

- Encryption with private key is used for signing messages
  - In digital signature, only one entity can sign a message **(i.e., private key holder)**, and any one can verify the signature **(i.e., public key holder(s))**.

Ex.

- Alice send (m, Enc_e(m))
- Bob check $Dec_d(Enc_e(m)) \stackrel{?}{=}$ m
- Entity authentication and data interity are achieved.

RSA Second Property

$$\begin{aligned} E(m_1).E(m_2) &= m_1^e.m_2^e \bmod n \\ &= (m_1.m_2)^e \bmod n \\ &= E(m_1.m_2) \end{aligned}$$

RSA is Partially homomorphic cryptosystem
(Allow multiplication on encrypted data)

Other Homomorphic Schemes

- Paillier Scheme: Allow homomorphic addition.

$$E(m_1).E(m_2) = E(m_1 + m_2)$$

$$E(m_1)^k = E(k.m_1)$$

- Fully Homomorphic Encryption:
  - Allow homomorphic addition and homomorphic.
  - Based on lattice based cryptography.
  - Crypto Nets [1] uses homomorphic encryption to do encrypted prediction for neural networks.

[1] https://arxiv.org/pdf/1412.6181.pdf

Comparison between Symmetric and Asymmetric key Cryptography

Advantages of Symmetric Key Cryptography

- Symmetric-key algorithms are generally fast, but Key management is a problem.

Drawbacks of Symmetric Key:

- Key Establishment problem
- Key management problem How?
- Does not allow homomorphic properties.

That does not mean symmetric key cryptography is useless. It is used with asymmetric key cryptography as will be explained later.

Comparison between Symmetric and Asymmetric key Cryptography

Advantages of Asymmetric Key Cryptography

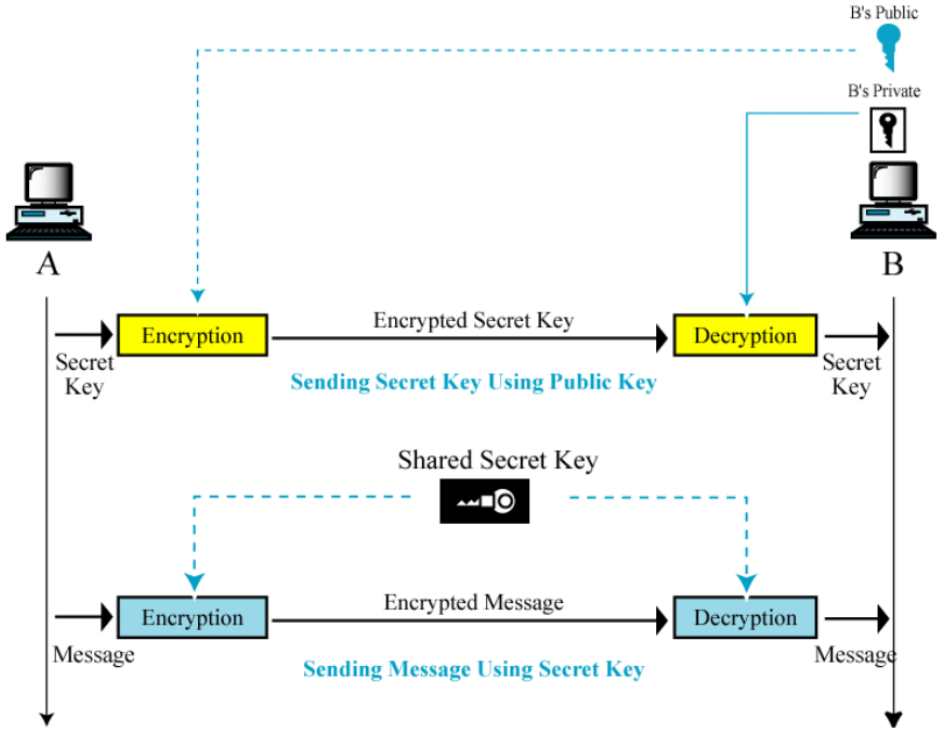
- Allow some homomorphic operations on the encrypted data.
- No requirements for a secret channel for key transfer.
- Each user has only one key pair which simplifies the key management

Drawbacks of Asymmetric Key:

- Asymmetric keys are typically larger than symmetric keys.
- Asymmetric key schemes are slower than symmetric key counterparts.
  - In practice, asymmetric key algorithms are typically hundreds to thousands times slower than symmetric key algorithms.

Hybrid Schemes

Public key cryptography is used to share a symmetric key and symmetric key cryptography is used for data encryption because it is more efficient.



- DES (or AES) for encrypting actual data
- RSA for encrypting corresponding DES session key

Questions