

Information Security

CENG418

week-2



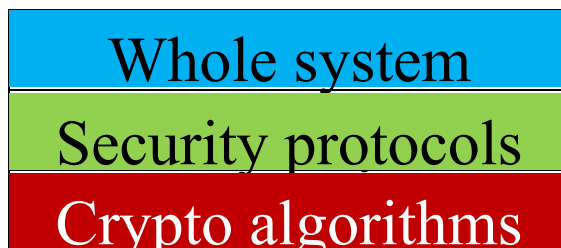
Cryptography: Terminology & Classic Ciphers

Information Security

InfoSec:

- Computer Security: deals mostly with access control
- Network Security: deals with communications security

Layers of a Security System:

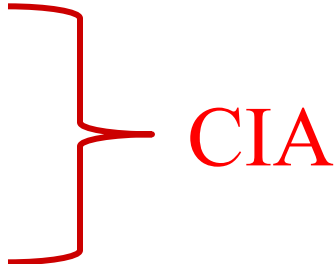


includes everything: policy management, administration, personnel training, etc.

how to achieve a certain functionality using the crypto algorithms

fundamental building blocks

Information Security

- At the core of information security is information assurance,
 - the act of maintaining the
 - confidentiality,
 - integrity, and
 - availability
- 

Main Issues

- confidentiality, privacy, secrecy
- authentication
- data integrity
- anonymity
- non-repudiation
- availability
- traceability

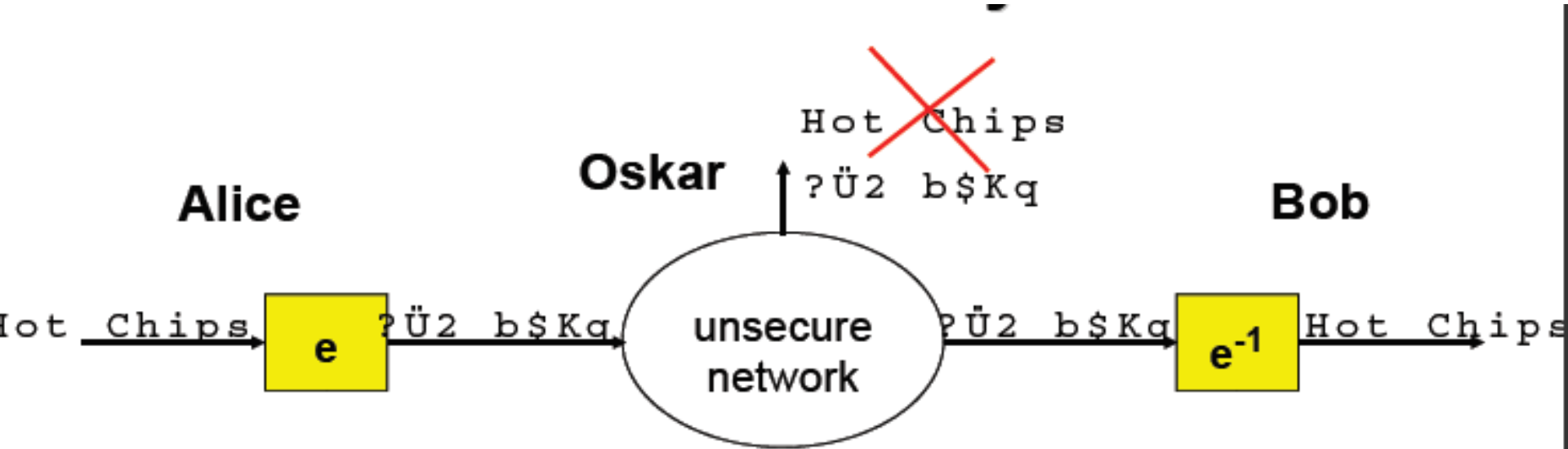
Cryptographic Goals

- **Confidentiality** is a service used to keep the content of information from all, but those authorized to have it.

Secrecy is a term synonymous with **confidentiality** and **privacy**.

- There are numerous approaches to providing confidentiality, ranging from physical protection to mathematical algorithms which **render data unintelligible**.

Confidentiality

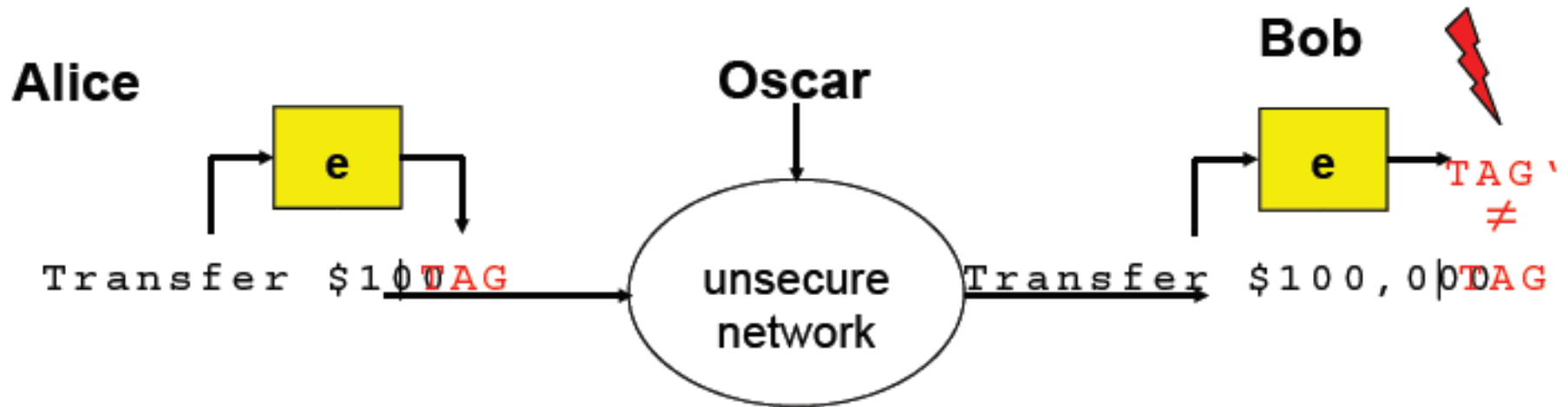


Encryption ensures **confidentiality** of messages

Cryptographic Goals

- **Data integrity** is a service which addresses the unauthorized alteration of data.
 - To assure data integrity, one must have the ability to **detect data manipulation by unauthorized parties**.
 - Data manipulation includes such things as **insertion**, **deletion**, and **substitution**.

Integrity of Messages



Cryptographic authentication tags:

2. Message Authentication Codes (MAC), or
 3. digital signatures
- ensure the **integrity** of messages

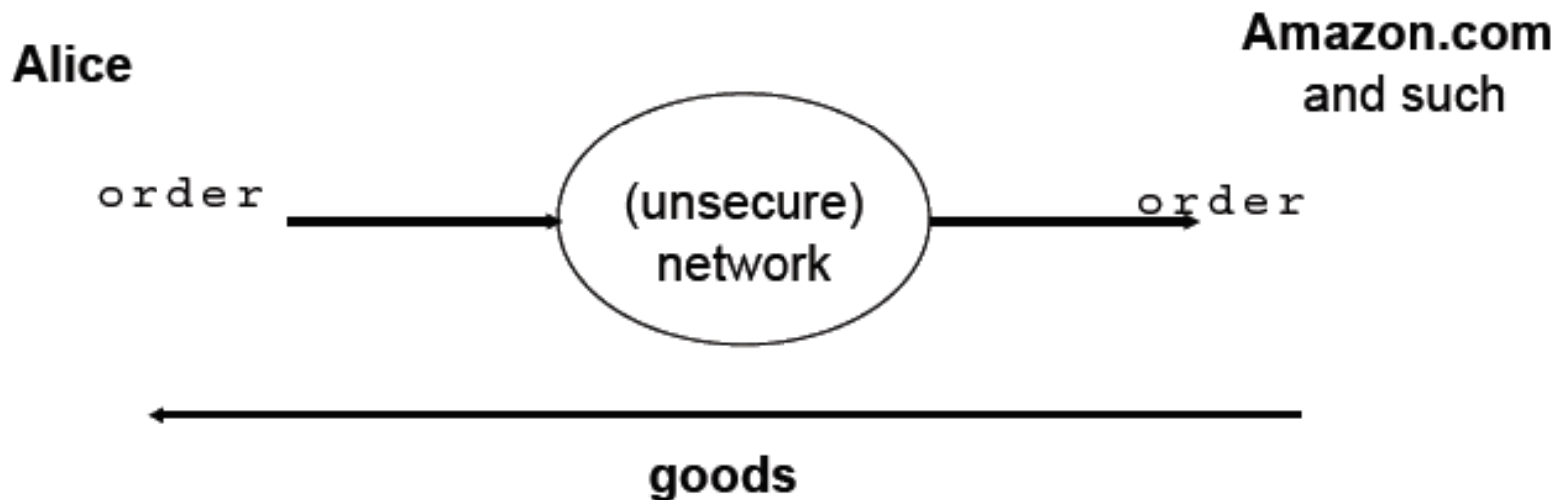
Cryptographic Goals

- **Authentication** is a service related to **identification**.
 - Two parties entering into a communication should **identify each other**.
 - Information delivered over a channel should be authenticated as to **origin, date of origin, data content, time sent, etc.**

Cryptographic Goals

- **Non-repudiation** *is a service which prevents an entity from denying previous commitments or actions.*
 - When disputes arise due to an entity denying that certain actions were taken, a means to resolve the situation is necessary.
 - For example, one entity may authorize the purchase of property by another entity and later deny such authorization was granted. A procedure involving a trusted third party is needed to resolve the dispute.

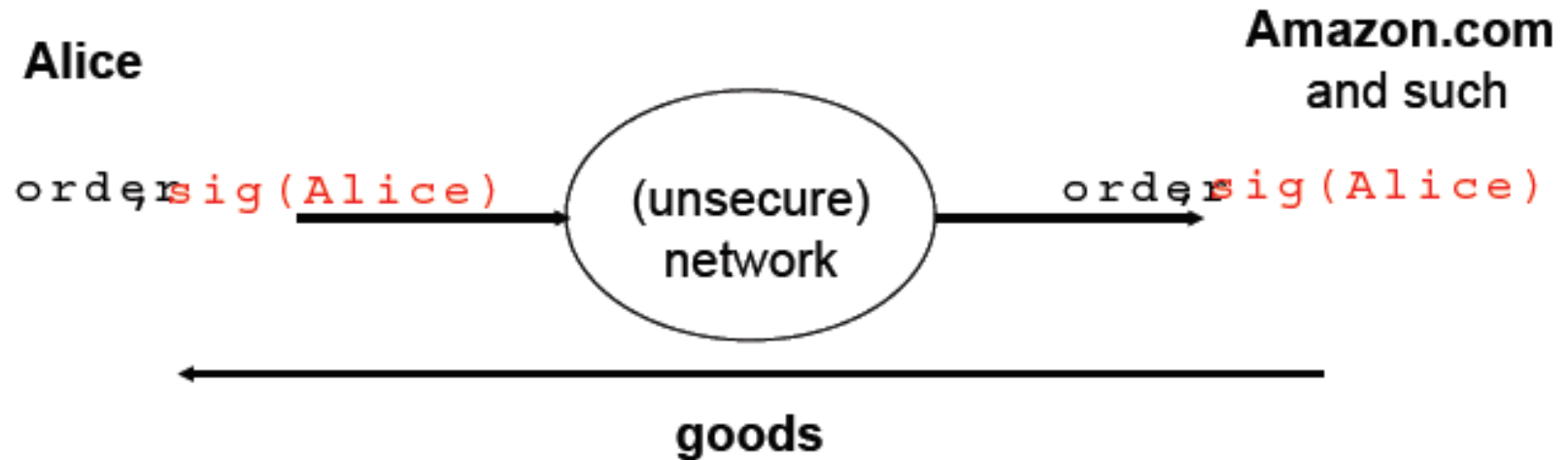
Non-Repudiation: Why we need it?



without non-repudiation:

2. Alice orders at favorite eCommerce vendor
3. stuff gets delivered
4. Alice doesn't feel like buying: „I never ordered this“
5. vendor can not **proof** it (big monetary issue if vendor = BMW.com)

Non-Repudiation: How it works?



with non-repudiation:

2. Alice orders at favorite eCommerce vendor
3. stuff gets delivered
4. Alice doesn't feel like buying: „I never ordered this“
5. vendor sues Alice: **proof** of order through Alice's signature

Non-repudiation is strong point of digital signatures

Security Goals

- Confidentiality (secrecy, privacy)
 - only those who are authorized to know can know
- Integrity
 - only modified by authorized parties and in authorized ways
- Availability
 - those authorized to access can get access

Tools for Information Security

- Cryptography
- Access control
- Hardware/software architecture for separation
- Processes and tools for developing more secure software
- Monitoring and analysis
- Recovery and response

What is the Cryptography?

- Definition of Cryptography
 - *Cryptography* is the study of mathematical techniques related to aspects of information security such as confidentiality, data integrity, entity authentication, and data origin authentication.
 - Cryptography is about the prevention and detection of cheating and other malicious activities.

Goals of Cryptography

- The most fundamental problem cryptography addresses: **ensure security of communication over insecure medium**
- What does secure communication mean?
 - confidentiality (privacy, secrecy)
 - only the intended recipient can see the communication
 - integrity (authenticity)
 - the communication is generated by the alleged sender
- What does insecure medium mean?
 - the adversary can eavesdrop
 - the adversary has full control over the communications

Basic Terminology and Concepts

Encryption domains and codomains

- “A” denotes a finite set called the *alphabet of definition*. For example, $A = \{0; 1\}$, the binary alphabet, is a frequently used alphabet of definition. Note that any alphabet can be encoded in terms of the binary alphabet.
- “M” denotes a set called the *message space*. M consists of strings of symbols from an alphabet of definition. An element of M is called a *plaintext message* or simply a *plaintext*. For example, M may consist of binary strings, English text, computer code, etc.
- “C” denotes a set called the *ciphertext space*. C consists of strings of symbols from an alphabet of definition, which may differ from the alphabet of definition for M. An element of C is called a *ciphertext*.

Basic Terminology and Concepts

Encryption and decryption transformations

- K denotes a set called the *key space*. *An element of K is called a key.*
- Each element $e \in K$ uniquely determines a bijection from M to C , denoted by E_e .
- E_e is called an *encryption function* or an *encryption transformation*. *Note that E_e must be a bijection if the process is to be reversed and a unique plaintext message recovered for each distinct ciphertext.*
- For each $d \in K$, D_d denotes a bijection from C to M (i.e., $D_d : C \rightarrow M$). D_d is called a *decryption function* or *decryption transformation*.
- The process of applying the transformation E_e to a message $m \in M$ is usually referred to as *encrypting m* or *the encryption of m* .
- The process of applying the transformation D_d to a ciphertext c is usually referred to as *decrypting c* or *the decryption of c* .

Basic Terminology and Concepts

- An *encryption scheme* consists of a set $\{E_e: e \in \mathcal{K}\}$ of encryption transformations and a corresponding set $\{D_d: d \in \mathcal{K}\}$ of decryption transformations with the property that for each $e \in \mathcal{K}$ there is a unique key $d \in \mathcal{K}$ such that $D_d = E_e^{-1}$; that is, $D_d(E_e(m)) = m$ for all $m \in \mathcal{M}$. An encryption scheme is sometimes referred to as a *cipher*.
- The keys e and d in the preceding definition are referred to as a *key pair* and sometimes denoted by (e, d) . Note that e and d could be the same.
- To *construct* an encryption scheme requires one to select a message space \mathcal{M} , a ciphertext space \mathcal{C} , a key space \mathcal{K} , a set of encryption transformations $\{E_e: e \in \mathcal{K}\}$, and a corresponding set of decryption transformations $\{D_d: d \in \mathcal{K}\}$.

Basic Terminology and Concepts

Example (*encryption scheme*) Let $\mathcal{M} = \{m_1, m_2, m_3\}$ and $\mathcal{C} = \{c_1, c_2, c_3\}$. There are precisely $3! = 6$ bijections from \mathcal{M} to \mathcal{C} . The key space $\mathcal{K} = \{1, 2, 3, 4, 5, 6\}$ has six elements in it, each specifying one of the transformations. Figure 1.5 illustrates the six encryption functions which are denoted by E_i , $1 \leq i \leq 6$. Alice and Bob agree on a trans-

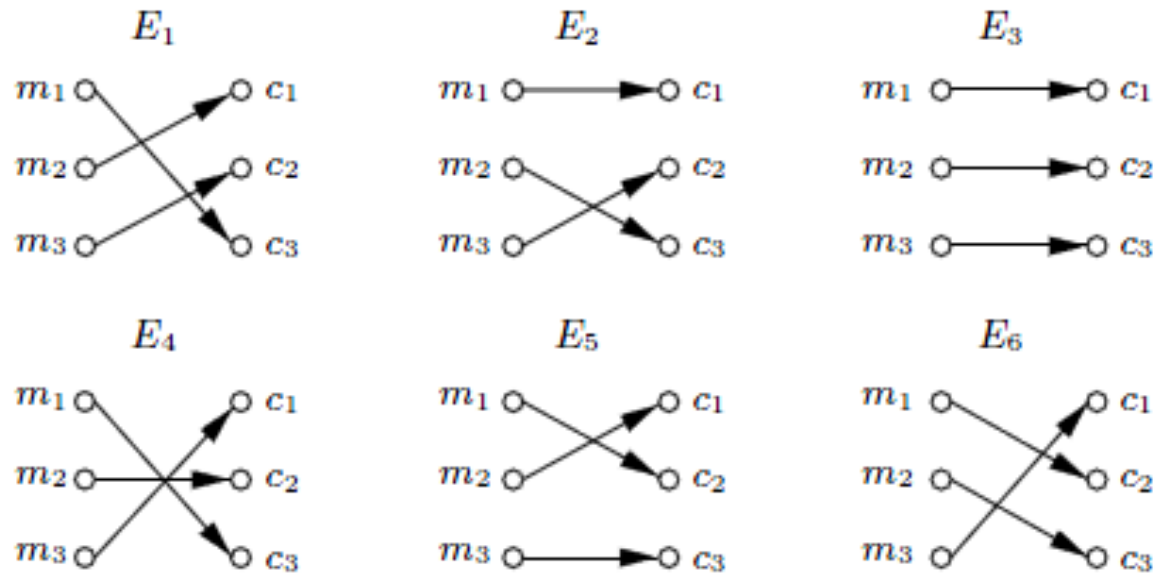


Figure 1.5: Schematic of a simple encryption scheme.

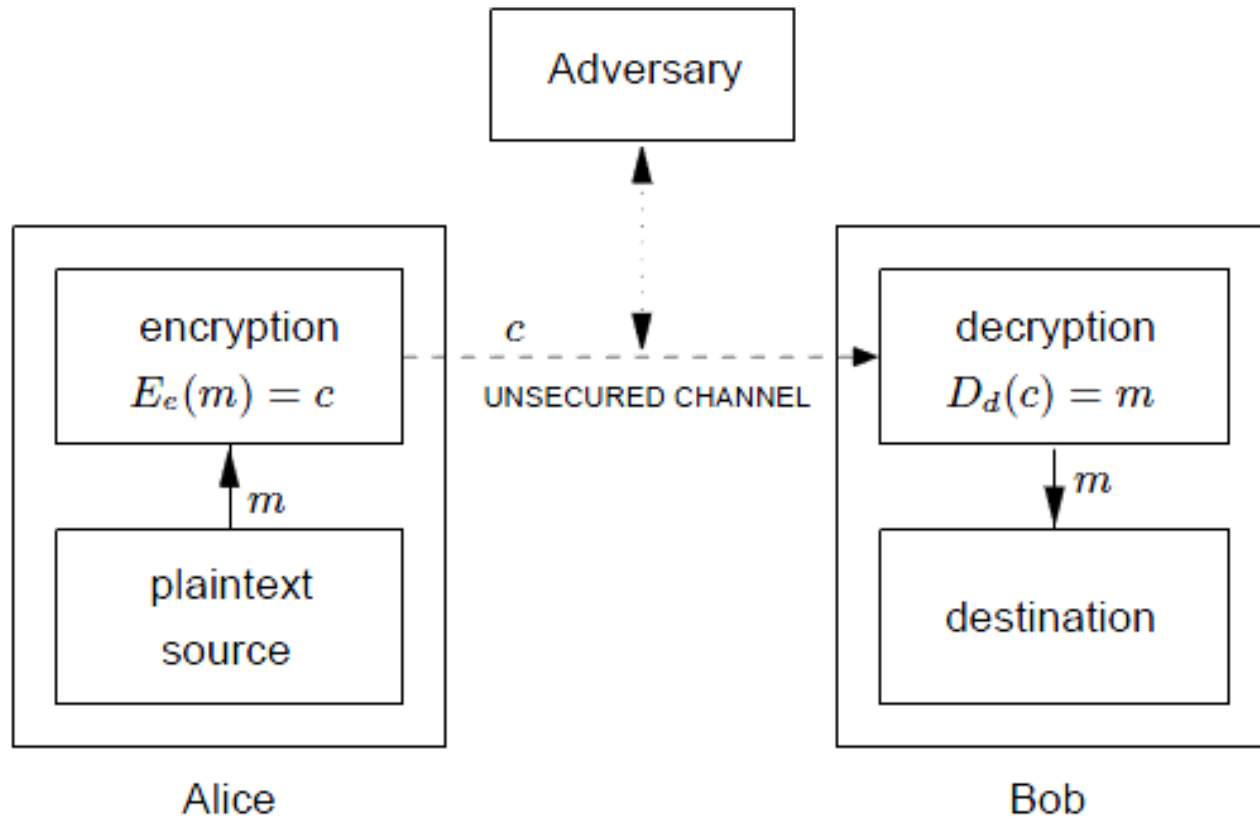
formation, say E_1 . To encrypt the message m_1 , Alice computes $E_1(m_1) = c_3$ and sends c_3 to Bob. Bob decrypts c_3 by reversing the arrows on the diagram for E_1 and observing that c_3 points to m_1 .

Basic Terminology and Concepts

- In cryptography, the set M is typically of astronomical proportions and, in these cases, new mathematical algorithms should use to describe the encryption and decryption transformations.

Basic Terminology and Concepts

- Schematic of a two-party communication using encryption.*



Basic Terminology and Concepts

Communication participants

- An *entity* or *party* is someone or something which sends, receives, or manipulates information. Alice and Bob are entities. An entity may be a person, a computer terminal, etc.
- A *sender* is an entity in a two-party communication which is the *legitimate transmitter of information*. In figure, the sender is Alice.
- A *receiver* is an entity in a two-party communication which is the *intended recipient* of information. In Figure, the receiver is Bob.
- An *adversary* is an entity in a two-party communication which is *neither the sender nor receiver*. Various other names are synonymous with adversary such as *enemy, attacker, opponent, tapper, eavesdropper, intruder, and interloper*.
 - An adversary will often attempt to play the role of either the legitimate sender or the legitimate receiver.

Basic Terminology and Concepts

Channels

- A *channel* is a means of conveying information from one entity to another.
- A *physically secure channel* or secure channel is one which *is not physically accessible* to the adversary.
- An *unsecured channel* is one from which parties other than those for which the information is intended can reorder, delete, insert, or read.
- A *secured channel* is one from which an adversary does not have the ability to reorder, delete, insert, or read.

Basic Terminology and Concepts

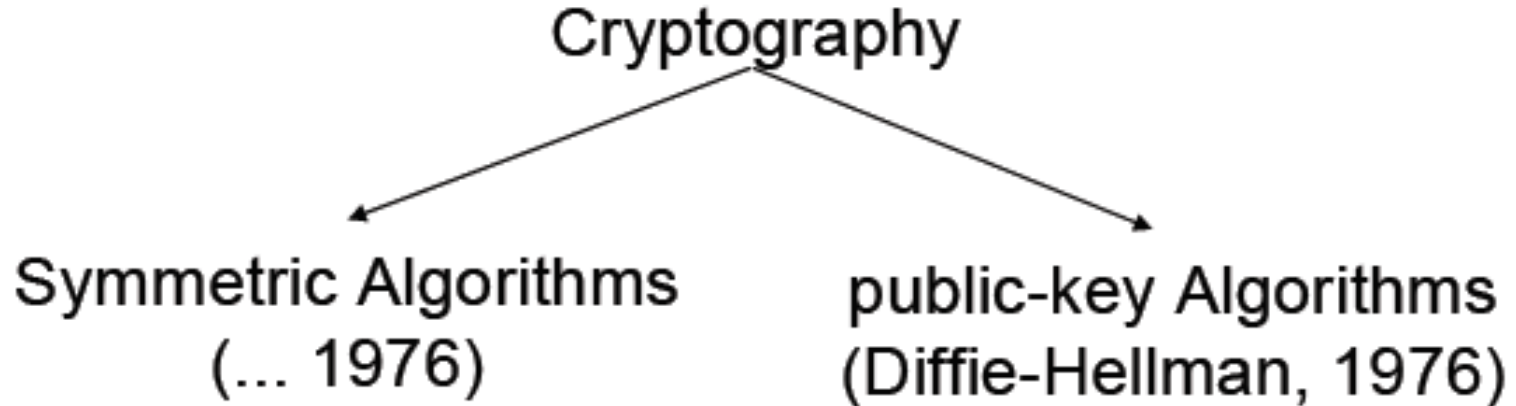
- An *information security service* is a method to provide some specific aspect of security.
 - For example, integrity of transmitted data is a security objective, and a method to ensure this aspect is an information security service.
- *Breaking an information security service* (which often involves more than simply encryption) implies defeating the objective of the intended service.
- A *passive adversary* is an adversary who *is capable only of reading information* from an unsecured channel.
- An *active adversary* is an adversary *who may also transmit, alter, or delete information* on an unsecured channel.

Basic Terminology and Concepts

- Cryptography,
 - Traditionally, designing algorithms/protocols
 - Nowadays, often synonym with cryptology
- Cryptanalysis
 - Breaking cryptography
- Cryptology: both cryptography & cryptanalysis
 - Becoming less common,

Information Security and Cryptography

1. IT Security \neq Cryptography
2. **but:** cryptography is an important **tool** for achieving IT security



History of Cryptography

- 2500+ years
- An ongoing battle between codemakers and codebreakers
- Driven by communication & computation technology
 - paper and ink
 - cryptographic engine & telegram, radio
 - modern cryptography: computers & digital communication

Basic Terminology

- Plaintext original message
- Ciphertext transformed message
- Key secret used in transformation
- Encryption
- Decryption
- Cipher algorithm for encryption/decryption

Basic Information

- Plaintext will be written in lowercase letters.
- CIPHERTEXT will be written in capital letters.
- The letters of alphabet are assigned numbers as follows:

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
q	r	s	t	u	v	w	x	y	z						
16	17	18	19	20	21	22	23	24	25						

- Spaces and punctuation are omitted.

Shift Cipher

- The Key Space:
 - [1 .. 25]
- Encryption given a key K:
 - each letter in the plaintext P is replaced with the K'th letter following corresponding number (shift right)
- Decryption given K:
 - shift left

History: $K = 3$, Caesar's cipher



Shift Ciphers – Julius Caesar

- He shifted each letter by 3 places so *a* become *D* , *b* become *E*.

gaul is divided into three parts



JDXOLVGLYLGHGLQWRWKUHHSDUWV

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
q	r	s	t	u	v	w	x	y	z						
16	17	18	19	20	21	22	23	24	25						

Gaul; an ancient region of western Europe that included what is now northern Italy and France and Belgium and part of Germany and the Netherlands

Shift Ciphers – Julius Caesar

- Label the letters as integer from 0 to 25. The key is an integer K , with $0 \leq K \leq 25$.
- The encryption process is: $y \rightarrow x + K \pmod{26}$
- The decryption process is: $x \rightarrow y - K \pmod{26}$

Shift Ciphers – Julius Caesar

Types of attack work:

- **Ciphertext only**; Eve has only the cipher text
 - An exhaustive search, since there are only 26 possible keys.
 - If message is longer than a few letters; try to find some words of 4 or 5 letters that are shifts of each other.
 - If message is sufficiently long, is to do a frequency count for the various letters. The letter e occurs most frequently in English texts. Suppose the letter L appears most frequently in the ciphertext. Since $e=4$ and $L=11$, a reasonable quest is that $K=11-4=7$.

Shift Ciphers – Julius Caesar

Types of attack work:

- **Known plaintext**; if you know just one letter of the plaintext along with the corresponding letter ciphertext, you can deduce the key. For instance; If you know $t(=19)$ encrypt to $D(=3)$ then the key $K \equiv 3 - 19 \equiv -16 \equiv 10 \pmod{26}$.
- **Chosen plaintext**; Chose the letter a as the plaintext. The ciphertext gives the key. For example if the ciphertext is H then the key is 7.
- **Chosen ciphertext**; Chose a letter A as the ciphertext. The plaintext is the negative of the key. For example, if the plaintext is h , the key is $-7 \equiv 19 \pmod{26}$.

Shift Cipher: Cryptanalysis

- Can an attacker find K?
 - YES: by a brute force attack through exhaustive key search,
 - key space is small (≤ 26 possible keys).
- Once K is found, very easy to decrypt

General Mono-alphabetic Substitution Cipher

- The key space: all permutations of $\Sigma = \{A, B, C, \dots, Z\}$
- Encryption given a key π :
 - each letter X in the plaintext P is replaced with $\pi(X)$
- Decryption given a key π :
 - each letter Y in the ciphertext P is replaced with $\pi^{-1}(Y)$

Example:

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

BECAUSE \rightarrow AZDBJSZ

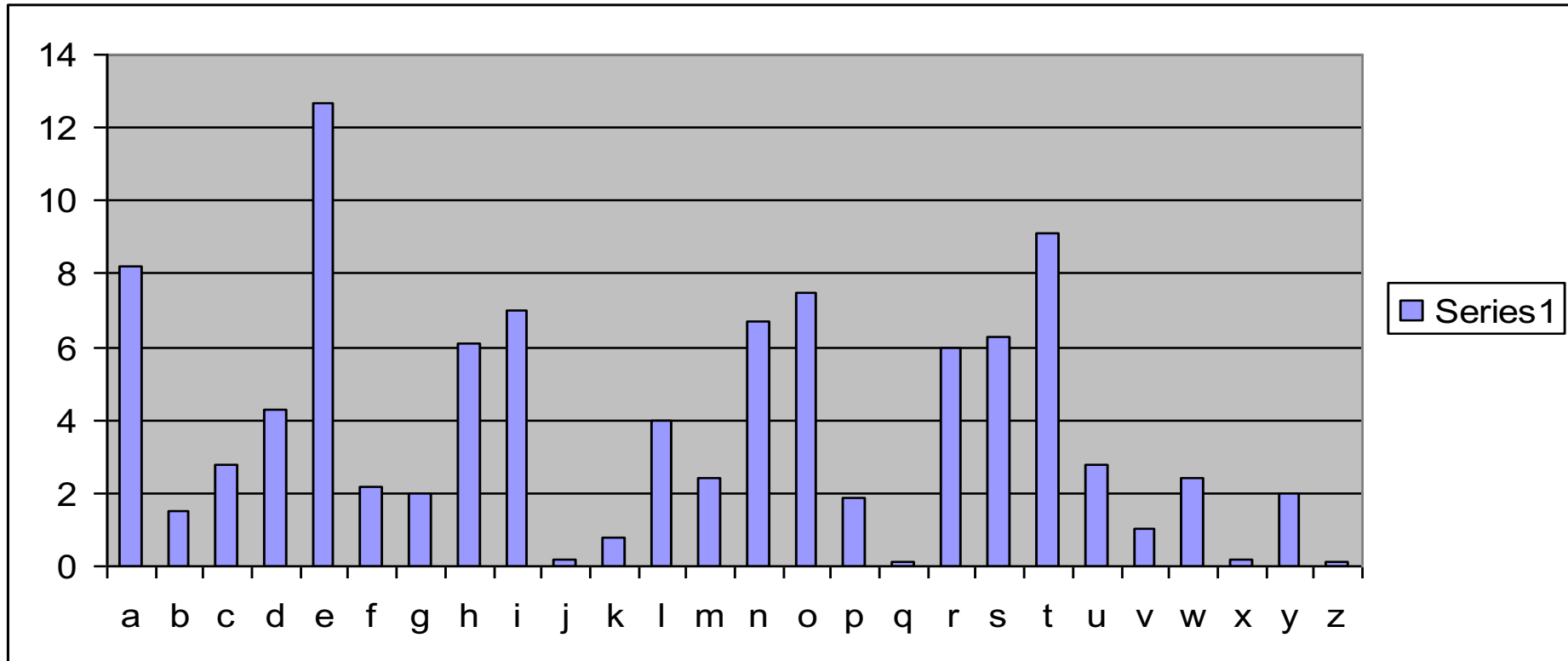
Strength of the General Substitution Cipher

- Exhaustive search is difficult
 - key space size is $26! \approx 4 \times 10^{26}$
- Dominates the art of secret writing throughout the first millennium
- Thought to be unbreakable by many back then

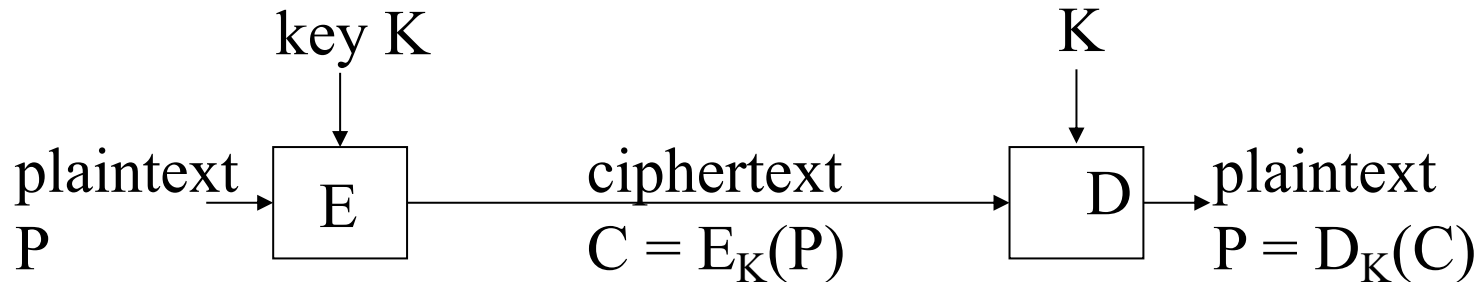
Cryptanalysis of Substitution Ciphers: Frequency Analysis

- Basic ideas:
 - Each language has certain features: frequency of letters, or of groups of two or more letters.
 - Substitution ciphers preserve the language features.
 - Substitution ciphers are vulnerable to frequency analysis attacks.

Frequency of Letters in English



Security Principles



- Security by obscurity doesn't work
- Should assume that the adversary knows the algorithm; the only secret the adversary is assumed to not know is the key
- **Kerckhoffs' principle (1883):**
Security should not rely on the secrecy of the algorithm; everything may be known but the key.
- **Key:** An easy-to-change, variable parameter of the encryption algorithm.

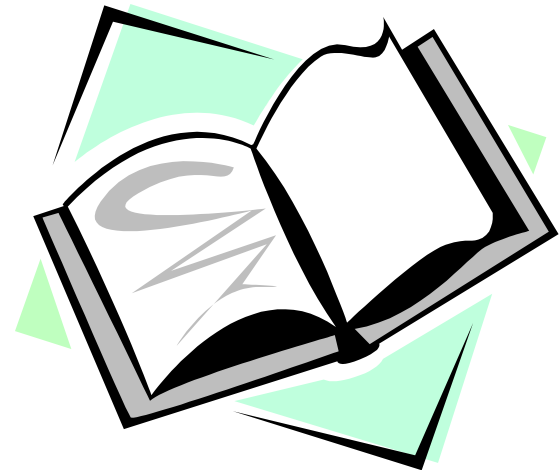
Readings for This Lecture

Readings:

- Security in Computing
 - Chapter 2: Basic Encryption and Decryption
- Cryptography engineering design principles and practical applications

Additional readings:

- Cryptography on Wikipedia



Homework

- Write Python functions that implement Ceasar encryption and decryption.
- Can you write a `break(encrypted_str)` function that tries to break the Ceasar encryption and returns the plaintext.
- The following ciphertext was generated with a monoalphabetic substitution. Can you breake it?

Break this cyphertext:

o zvfhs evve gl zfe cvvy do bglz wd zfe bv
cvvy lzvjv zgd rgcvjls gd qwrr oq lzjvfld lo
frr lo sow sowjdvrq lo wd lo vhwjs oyv frfd
zob dzfrr lzgd crooes evve cv fydbvjve gl
bgrr cv rfge lo wd bzodv njohgevypv dzowre
zfhv tvnl dzojl jvdljfggyve fye owl oq zfwyl
lzgd afe sowyk afy cwl do awpz bfd owj rohv
bv bowre yol wyevjdlfye bzfl bfd aodl qgl
cwl rgtv lzv obyvj oq f qowr egdvfdv lo tvvn
gl qjoa eghwrkgyk rvl gl qvve vhwj oy lzv
nglz oq rgqv bzvjjv gd zv koyv

HINT:

1. Study frequencies
2. Take into account the rules and restrictions in the language

```
# This code performs frequency analysis on string named corpus
```

```
import matplotlib.pyplot as plt
import pandas as pd
```

```
alphabet = "abcdefghijklmnopqrstuvwxyz"
```

```
df = pd.DataFrame({
    "letter": list(alphabet),
    "count": [corpus.lower().count(letter) for letter in alphabet]
})
```

```
df = df.sort_values("count")
plt.bar(df["letter"], df["count"])
```

Coming Attractions ...

- Cryptography: One-time Pad, Informational Theoretical Security, Stream Ciphers