

## JEPPIAAR INSTITUTE OF TECHNOLOGY "Self-Belief | Self Discipline | Self Respect"



DEPARTMENT

### OF

## COMPUTER SCIENCE AND ENGINEERING

## **LECTURE NOTES**

## **CS8792- CRYPTOGRAPHY AND NETWORK SYSTEM**

Year/Semester: IV/07 CSE 2021 – 2022

Prepared by

**H.SHINE** 

Assistant Professor/CSE

### **UNIT I INTRODUCTION**

Security trends – Legal, Ethical and Professional Aspects of Security, Need for Security at Multiple levels, Security Policies – Model of network security – Security attacks, services mechanisms – OSI security architecture – Classical encryption techniques: substitution techniques, transposition techniques, steganography- Foundations of modern cryptography: perfect security – information theory – product cryptosystem – cryptanalysis.

## Definition

Cryptography is the science of using mathematics to encrypt and decrypt data.

#### Phil Zimmermann

Cryptography is the art and science of keeping messages secure.

Bruce Schneier

The art and science of concealing the messages to introduce secrecy in information Security is recognized as cryptography.

It is the study and practice of techniques for secure communication in the presence of third parties called adversaries. Data Confidentiality, Data Integrity, Authentication and Non-repudiation are core principles of modern-day cryptography.

## Terminologies

A message is **plaintext** (sometimes called clear text). The process of disguising a message in such a way as to hide its substance is **encryption**. An encrypted message is **cipher text**. The process of turning cipher text back into plaintext is **decryption**.



A **cryptosystem** is an implementation of cryptographic techniques and their accompanying infrastructure to provide information security services. A cryptosystem is also referred to as a **cipher system**. The various components of a basic cryptosystem are as follows

- Plaintext
- Encryption Algorithm
- Cipher text
- Decryption Algorithm
- Encryption Key
- Decryption Key

While **cryptography** is the science of securing data, **cryptanalysis** is the science of analyzing and breaking secure communication. Classical cryptanalysis involves an interesting combination of analytical reasoning, application of mathematical tools, pattern finding, patience, determination, and luck. **Cryptanalysts** are also called attackers. **Cryptology** embraces both cryptography and cryptanalysis.

## **Security Trends**

#### **Definition of Computer Security**

The protection afforded to an automated information system in order to attain the applicable objectives of preserving the integrity, availability, and confidentiality of information system resources (includes hardware, software, firmware, information / data, and telecommunications)

#### Confidentiality

#### • Data confidentiality

Assures that private or confidential information is not made available or disclosed to unauthorized

• Privacy

Assures that individuals control or influence what information related to them may be collected and stored and by whom and to whom that information may be disclosed.

#### Integrity

• Data integrity

Assures that information and programs are changed only in a specified and authorized manner.

• System integrity

Assures that a system performs its intended function in an unimpaired manner, free from deliberate or inadvertent unauthorized manipulation of the system.

#### Availability

• Assures that systems work promptly and service is not denied to authorize users.

## **CIA Triad**



### Confidentiality

- Preserving authorized restrictions on information access and disclosure, including means for protecting personal privacy and proprietary information.
- A loss of confidentiality is the unauthorized disclosure of information.

### Integrity

- Guarding against improper information modification or destruction, including ensuring information nonrepudiation and authenticity.
- A loss of integrity is the unauthorized modification or destruction of information.

#### Availability

Ensuring timely and reliable access to and use of information

A loss of availability is the disruption of access to or use of information or an information system.

#### Authenticity

• The property of being genuine and being able to be verified and trusted

#### Accountability

• The security goal that generates the requirement for actions of an entity to be traced uniquely to that entity

## Legal, Ethical and Professional aspects of security

We must understand the scope of an organization's legal and ethical responsibilities. To minimize liabilities/reduce risks, the security practitioner must:

- 1. Understand current legal environment.
- 2. Stay current with laws and regulations.
- 3. Watch for new issues and emerge.

Information is endangered both by external factors, such as hackers, computer viruses, thefts, and internal ones - the loss of data as a result of improper protection, the lack of backup copies or the loss of a flash drive that contains unprotected data. An improper protection of data may result in the loss of company's reputation, its customers' trust or in financial losses. This issue is of particular importance as regards the court system due to the volume of personal data that are processed and stored in courts and their unique character (sentences, orders, and statements of reasons, convictions, and personal details of victims or land registers). They all constitute information that must be protected against theft, loss or alterations. The loss of data could affect negatively the trial and the judicial independence by possible external pressure in cases where data was lost.

#### Law and Ethics in Security

- Definition of Laws
- Rules that mandate or prohibit certain behavior
- Drawn from ethics
- Definition of Ethics
- Define socially acceptable behaviors
- I Key difference between law and ethics
- Laws carry the authority of a governing body
- Ethics do not carry the authority of a governing body
- Based on cultural mores
- I Fixed moral attitudes or customs
- Some ethics standards are universal

#### Organizational Liability and the Need for Counsel

What if an organization does not behave ethically? Even if there is no breach of criminal law, there can still be liability.

- What is Liability?
- Legal obligation of organization
- Extends beyond criminal or contract law
- Include legal obligation to restitution
- Employee acting with or without the authorization performs and illegal or unethical act that causes some degree of harm
- Employer can be held financially liable

An organization increases its liability if it refuses to take measures known as due care.

What is Due care?

Organization makes sure that every employee knows what is acceptable or unacceptable

I Knows the consequences of illegal or unethical actions

Due diligence

**Requires** 

I Make a valid effort to protect others

I Maintains the effort

In legal system, any court can assert its authority over an individual or organization if it can establish jurisdiction

What is Jurisdiction?

Court's right to hear a case if a wrong is committed

I Term – long arm

Extends across the country or around the world

## **Policy Versus law**

I Policies

I Guidelines that describe acceptable and unacceptable employee behaviors

I Functions as organizational laws

I Has penalties, judicial practices, and sanctions

Difference between policy and law

I Ignorance of policy is acceptable

I Ignorance of law is unacceptable

I Keys for a policy to be enforceable

Dissemination (distribution)

Review (reading)

Comprehension (understanding)

Compliance (agreement)

Uniform enforcement

## **Types of Law**

I Civil – govern a nation or state

Criminal – addresses activities and conduct harmful to public

Private – encompasses family, commercial, labor, and regulates the relationship between individuals and organizations

I Public – regulates the structure and administration of government agencies and their relationships with citizens, employees, and other governments

## **International Laws and Legal Bodies**

I Organizations do business on the Internet – they do business globally

I Professionals must be sensitive to the laws and ethical values of many different cultures, societies, and countries

I Few international laws relating to privacy and informational security

International laws are limited in their enforceability

## **Ethics and Security**

### The Ten Commandments of Computer Ethics®

#### From The Computer Ethics Institute

- 1. Thou shalt not use a computer to harm other people.
- 2. Thou shalt not interfere with other people's computer work.
- 3. Thou shalt not snoop around in other people's computer files.
- 4. Thou shalt not use a computer to steal.
- 5. Thou shalt not use a computer to bear false witness.
- 6. Thou shalt not copy or use proprietary software for which you have not paid.
- Thou shalt not use other people's computer resources without authorization or proper compensation.
- 8. Thou shalt not appropriate other people's intellectual output.
- Thou shalt think about the social consequences of the program you are writing or the system you are designing.
- Thou shalt always use a computer in ways that ensure consideration and respect for your fellow humans.

# Overriding factor in leveling ethical perceptions within a small population is education

Employees must be trained in expected behaviors of an ethical employee, especially in areas of information security

Proper ethical training vital to creating informed, well prepared, and low-risk system user

## **Deterrence To Unethical And Illegal Behaviours**

- Deterrence: best method for preventing an illegal or unethical activity; e.g., laws, policies, technical controls
- Laws and policies only deter if three conditions are present:
  - I Fear of penalty
  - IProbability of being caught
  - IProbability of penalty being administered

## **Codes of Ethics And Professional Organizations**

I Several professional organizations have established codes of conduct/ethics

- Codes of ethics can have positive effect unfortunately, many employers do not encourage joining of these professional organizations
- Responsibility of security professionals to act ethically and according to policies of employer, professional organization, and laws of society

## **Need for Multilevel Security**

Having information of different security levels on the same computer systems poses a real threat. It is not a straight-forward matter to isolate different information security levels, even though different users log in using different accounts, with different permissions and different access controls.

Some organizations go as far as to purchase dedicated systems for each security level. This is often prohibitively expensive, however. A mechanism is required to enable users at different security levels to access systems simultaneously, without fear of information contamination.

The term multi-level arises from the defense community's security classifications: Confidential, Secret, and Top Secret.

Individuals must be granted appropriate clearances before they can see classified information. Those with Confidential clearance are only authorized to view Confidential documents; they are not trusted to look at Secret or Top Secret information. The rules that apply to data flow operate from lower levels to higher levels, and never the reverse. This is illustrated below.



**Information Security Levels** 



Under such a system, users, computers, and networks use labels to indicate security levels. Data can flow between like levels, for example between "Secret" and "Secret", or from a lower level to a higher level. This means that users at level "Secret" can share data with one another, and can also retrieve information from Confidential-level (i.e., lower-level), users. However, data cannot flow from a higher level to a lower level. This prevents processes at the "Secret" level from viewing information classified as "Top Secret". It also prevents processes at a higher level from accidentally writing information to a lower level. This is referred to as the "no read up, no write down" model.

### **MLS and System Privilege**

MLS access rules are always combined with conventional access permissions (file permissions). For example, if a user with a security level of "Secret" uses Discretionary Access Control (DAC) to block access to a file by other users, this also blocks access by users with a security level of "Top Secret". A higher security clearance does not automatically give permission to arbitrarily browse a file system.

Users with top-level clearances do not automatically acquire administrative rights on multi-level systems. While they may have access to all information on the computer, this is different from having administrative rights.

### Security Levels, Objects and Subjects

As discussed above, subjects and objects are labeled with Security Levels (SLs), which are composed of two types of entities:

Sensitivity: — A hierarchical attribute such as "Secret" or "Top Secret".

Categories: — A set of non-hierarchical attributes such as "US Only" or "UFO".

An SL must have one sensitivity, and may have zero or more categories.

Examples of SLs are: { Secret / UFO, Crypto }, { Top Secret / UFO, Crypto, Stargate } and { Unclassified }

Note the hierarchical sensitivity followed by zero or more categories. The reason for having categories as well as sensitivities is so that sensitivities can be further compartmentalized on a need-to-know basis.

## **Security Polices**

Following are some points which help in security policy of an organization.

- Who should have access to the system?
- How it should be configured?
- How to communicate with third parties or systems?

Policies are divided in two categories -

- User policies
- IT policies.

User policies generally define the limit of the users towards the computer resources in a workplace. For example, what are they allowed to install in their computer, if they can use removable storages.

Whereas, IT policies are designed for IT department, to secure the procedures and functions of IT fields.

- **General Policies** This is the policy which defines the rights of the staff and access level to the systems. Generally, it is included even in the communication protocol as a preventive measure in case there are any disasters.
- Server Policies This defines who should have access to the specific server and with what rights. Which software's should be installed, level of access to internet, how they should be updated.
- **Firewall Access and Configuration Policies** It defines who should have access to the firewall and what type of access, like monitoring, rules change. Which ports and services should be allowed and if it should be inbound or outbound.
- **Backup Policies** It defines who is the responsible person for backup, what should be the backup, where it should be backed up, how long it should be kept and the frequency of the backup.
- VPN Policies These policies generally go with the firewall policy, it defines those users who should have a VPN access and with what rights. For site-to-site connections with partners, it defines the access level of the partner to your network, type of encryption to be set.

## **Structure of a Security Policy**

When you compile a security policy you should have in mind a basic structure in order to make something practical. Some of the main points which have to be taken into consideration are –

**Description** of the Policy and what is the usage for?

- Where this policy should be applied?
- Functions and responsibilities of the employees that are affected by this policy.
- Procedures that are involved in this policy.
- Consequences if the policy is not compatible with company standards.

## **Types of Policies**

In this section we will see the most important types of policies.

• **Permissive Policy** – It is a medium restriction policy where we as an administrator block just some well-known ports of malware regarding internet access and just some exploits are taken in consideration.

- **Prudent Policy** This is a high restriction policy where everything is blocked regarding the internet access, just a small list of websites are allowed, and now extra services are allowed in computers to be installed and logs are maintained for every user.
- Acceptance User Policy This policy regulates the behavior of the users towards a system or network or even a webpage, so it is explicitly said what a user can do and cannotin a system. Like are they allowed to share access codes, can they share resources, etc.
- User Account Policy This policy defines what a user should do in order to have or maintain another user in a specific system. For example, accessing an e-commerce webpage. To create this policy, you should answer some questions such as
  - Should the password be complex or not?
  - What age should the users have?
  - Maximum allowed tries or fails to log in?
  - When the user should be deleted, activated, blocked?
- **Information Protection Policy** This policy is to regulate access to information, hot to process information, how to store and how it should be transferred.
- **Remote Access Policy** This policy is mainly for big companies where the user and their branches are outside their headquarters. It tells what should the users access, when they can work and on which software like SSH, VPN, RDP.
- **Firewall Management Policy** This policy has explicitly to do with its management, which ports should be blocked, what updates should be taken, how to make changes in the firewall, how long should be the logs be kept.
- **Special Access Policy** This policy is intended to keep people under control and monitor the special privileges in their systems and the purpose as to why they have it. These employees can be team leaders, managers, senior managers, system administrators, and such high designation based people.
- Network Policy This policy is to restrict the access of anyone towards the network resource and make clear who all will access the network. It will also ensure whether that person should be authenticated or not. This policy also includes other aspects like, who will authorize the new devices that will be connected with network? The documentation of network changes. Web filters and the levels of access. Who should have wireless connection and the type of authentication, validity of connection session?
- Email Usage Policy This is one of the most important policies that should be done because many users use the work email for personal purposes as well. As a result information can leak outside. Some of the key points of this policy are the employees should know the importance of this system that they have the privilege to use. They should not open any attachments that look suspicious. Private and confidential data should not be sent via any encrypted email.
- Software Security Policy This policy has to do with the software's installed in the user computer and what they should have. Some of the key points of this policy are Software of the company should not be given to third parties. Only the white list of software's should

be allowed, no other software's should be installed in the computer. Warez and pirated software's should not be allowed

## **Model for Network Security**



A message is to be transferred from one party to another across some sort of Internet service. The two parties, who are the principals in this transaction, must cooperate for the exchange to take place.

A logical information channel is established by defining a route through the Internet from source to destination and by the cooperative use of communication protocols (e.g., TCP/IP) by the two principals.

#### All the techniques for providing security have two components:

A security-related transformation on the information to be sent.

Examples: encryption of the message, addition of a code based on the contents

Some secret information shared by the two principals, unknown to the opponent Example: encryption key used in conjunction with the transformation

#### A trusted third party may be needed to achieve secure transmission.

- for distributing the secret information to the two principals
- to arbitrate disputes between the two principals concerning the authenticity of a message transmission

#### Four basic tasks in designing a particular security service:

- 1. Design an algorithm for performing the security-related transformation such that an opponent cannot defeat its purpose.
- 2. Generate the secret information to be used with the algorithm.
- 3. Develop methods for the distribution and sharing of the secret information.
- 4. Specify a protocol to be used by the two principals that makes use of the security algorithm and the secret information to achieve a particular security service

## **Network Access Security Model**



- Protecting an information system from unwanted access from hacker, intruder hacker who, with no malign intent, simply gets satisfaction from breaking and entering a computer system.
- Intruder can be a disgruntled employee who wishes to do damage or a Criminal who seeks to exploit computer assets for financial gain
- placement in a computer system of logic that exploits vulnerabilities in the system and that can affect application programs as well as utility programs, such as editors and compilers

#### Two kinds of threats:

- **Information access threats**: Intercept or modify data on behalf of users who should not have access
- Service threats: Exploit service flaws in computers to inhibit use by legitimate users

Examples: Viruses and worms, spread using disks & inserted over network

## **The OSI Security Architecture**

- ITU-T Recommendation X.800, Security Architecture for OSI, defines such a systematic approach
- The OSI security architecture focuses on security attacks, mechanisms, and services.

#### Security attack

• Any action that compromises the security of information owned by an organization.

#### Security mechanism

• A process (or a device) that is designed to detect, prevent, or recover from a security attack.

#### **Security service**

- A processing or communication service that enhances the security of the data processing systems and the information transfers of an organization
- The services are intended to counter security attacks, and they make use of one or more security mechanisms to provide the service

#### **Security Attacks**

- means of classifying security attacks, used both in X.800 and RFC 2828
- A passive attack attempts to learn or make use of information but does not affect system resources.
- An active attack attempts to alter system resources or affect their operation.

#### **Passive Attacks**

- It is the nature of eavesdropping on, or monitoring of, transmissions.
- The goal is to obtain information that is being transmitted.
- very difficult to detect, because they do not involve any alteration of the data
- feasible to prevent the success of these attacks, usually by means of encryption
- emphasis in dealing with passive attacks is on prevention rather than detection

#### Two types of passive attacks

- Release of message contents
- Traffic analysis.

#### **Release of Message Contents**

• A telephone conversation, an electronic mail message, and a transferred file may contain sensitive or



confidential information

• prevent an opponent from learning the contents of these transmissions

#### **Traffic Analysis**

- observe the pattern of these messages
- The opponent could determine the location and identity of communicating hosts and could observe the frequency and length of messages being exchanged.
- This information might be useful in guessing the nature of the communication that was taking place

Active Attacks

- Active attacks involve some modification of the data stream or the creation of a false stream
- detect and to recover from any disruption or delays caused by them
- can be subdivided into four categories:
  - o masquerade,
  - o replay,
  - modification of messages
  - denial of service

#### Masquerade

• one entity pretends to be a different

entity

• usually includes one of the other forms of active attack

Example

Authentication sequences can be captured and replayed after a valid authentication

Bob

sequence

#### Replay

• passive capture of a data unit and its subsequent retransmission to produce an unauthorized effect



Message from Darth

Alice

that appears to be

from Bob

Internet or other comms facility

Darth



### **Modification of Messages**

• Some portion of a legitimate message is altered, or that messages are delayed or reordered, to produce an unauthorized effect

#### Example

• A message meaning "Allow John Smith to read confidential file accounts" is modified to mean "Allow Fred Brown to read confidential file accounts."

### **Denial of Service**

- Prevents or inhibits the normal use or management of communications facilities
- May have a specific target; for example, an entity may suppress all messages directed to a particular destination
- Disruption of an entire network, either by disabling the network or by overloading it with messages so as to degrade performance



(d) Denial of service

## **Security Services in X.800**

- X.800 defines a security service as a service that is provided by a protocol layer of communicating open systems and that ensures adequate security of the systems or of data transfers.
- RFC 2828, defines as a processing or communication service that is provided by a system to give a specific kind of protection to system resources;
- Security services implement security policies and are implemented by security mechanisms.

#### **X.800**

• divides these services into five categories and fourteen specific services

#### Authentication

- The assurance that the communicating entity is the one that it claims to be
- Two types
  - Peer Entity Authentication
  - o Data-Origin Authentication

#### **Access control**

• The prevention of unauthorized use of a resource

#### Data confidentiality

- The protection of data from unauthorized disclosure.
- Four Types
  - Connection Confidentiality
  - Connectionless Confidentiality
  - o Selective-Field Confidentiality
  - Traffic-Flow Confidentiality

#### **Data integrity**

- The assurance that data received are exactly as sent by an authorized entity (i.e., contain no modification, insertion, deletion, or replay).
- Five types
  - Connection Integrity with Recovery
  - Connection Integrity without Recovery
  - Selective-Field Connection Integrity
  - Connectionless Integrity
  - Selective-Field Connectionless Integrity

#### Nonrepudiation

- Provides protection against denial by one of the entities involved in a communication of having participated in all or part of the communication
- Two types
  - Nonrepudiation, Origin
  - Nonrepudiation, Destination

## Security Mechanisms in X.800.

#### Specific security mechanisms:

May be incorporated into the appropriate protocol layer in order to provide some of the OSI security services.

#### Encipherment

The use of mathematical algorithms to transform data into a form that is not readily intelligible. The transformation and subsequent recovery of the data depend on an algorithm and zero or more encryption keys.

#### **Digital Signature**

Data appended to, or a cryptographic transformation of, a data unit that allows a recipient of the data unit to prove the source and integrity of the data unit and protect against forgery (e.g, by the recipient).

#### **Access Control**

A variety of mechanisms that enforce access rights to resources.

#### **Data Integrity**

A variety of mechanisms used to assure the integrity of a data unit or stream of data units.

#### **Authentication Exchange**

A mechanism intended to ensure the identity of an entity by means of information exchange.

#### **Traffic Padding**

The insertion of bits into gaps in a data stream to frustrate traffic analysis attempts.

#### **Routing Control**

Enables selection of particular physically secure routes for certain data and allows routing changes, especially when a breach of security is suspected.

#### Notarization

The use of a trusted third party to assure certain properties of a data exchange.

#### **Pervasive Security Mechanisms**

Mechanisms that are not specific to any particular OSI security service or protocol layer.

#### **Trusted Functionality**

That which is perceived to be correct with respect to some criteria (e.g., as established by a security policy).

#### Security Label

The marking bound to a resource (which may be a data unit) that names or designates the security attributes of that resource.

#### **Event Detection**

Detection of security-relevant events.

#### **Security Audit Trail**

Data collected and potentially used to facilitate a security audit, which is an independent review and examination of system records and activities.

#### **Security Recovery**

Deals with requests from mechanisms, such as event handling and management functions, and takes recovery actions.

## **Classical Encryption Techniques**

- Symmetric encryption is a form of cryptosystem in which encryption and decryption are performed using the same key. It is also known as conventional encryption.
- Symmetric encryption transforms plaintext into ciphertext using a secret key and an encryption algorithm. Using the same key and a decryption algorithm, the plaintext is recovered from the ciphertext.
- Traditional (precomputer) symmetric ciphers use substitution and/or transposition techniques. Substitution techniques map plaintext elements (characters, bits) into ciphertext elements. Transposition techniques systematically transpose the positions of plaintext elements.
- Rotor machines are sophisticated precomputer hardware devices that use substitution techniques.
- Steganography is a technique for hiding a secret message within a larger one in such a way that others cannot discern the presence or contents of the hidden message.
  - Symmetric Cipher Model
- o Cryptanalysis and Brute-Force Attack
  - Substitution Techniques
    - Caesar Cipher

- Monoalphabetic Ciphers
- o Playfair Cipher
- o Hill Cipher
- o Polyalphabetic Ciphers
- One-Time Pad
- Transposition Techniques
- Rotor Machines
- Steganography

#### Introduction

- **Symmetric encryption** is a form of cryptosystem in which encryption and decryption are performed using the **same key**. It is also known as **conventional encryption**.
- Symmetric encryption transforms plaintext into ciphertext using a secret key and an encryption algorithm. Using the same key and a decryption algorithm, the plaintext is recovered from the ciphertext.
- The two types of attack on an encryption algorithm are **cryptanalysis**, based on properties of the encryption algorithm, and **brute-force**, which involves trying all possible keys.
- Traditional (precomputer) symmetric ciphers use substitution and/or transposition techniques. Substitution techniques map plaintext elements (characters, bits) into ciphertext elements. Transposition techniques systematically transpose the positions of plaintext elements.
- Rotor machines are sophisticated precomputer hardware devices that use substitution techniques.
- Steganography is a technique for hiding a secret message within a larger one in such a way that others cannot discern the presence or contents of the hidden message.
- An original message is known as the **plaintext**, while the coded message is called the **ciphertext**.
- The process of converting from plaintext to ciphertext is known as **enciphering or encryption**; restoring the plaintext from the ciphertext is **deciphering or decryption**.
- The many schemes used for encryption constitute the area of study known as **cryptography**. Such a scheme is known as a **cryptographic system or a cipher**.
- Techniques used for deciphering a message without any knowledge of the enciphering details fall into the area of cryptanalysis. **Cryptanalysis** is what the layperson calls

"breaking the code." The areas of cryptography and cryptanalysis together are called **cryptology** 

## Symmetric Cipher Model

A symmetric encryption scheme has five ingredients

- Plaintext: This is the original intelligible message or data that is fed into the algorithm as input.
- Encryption algorithm: The encryption algorithm performs various substitutions and transformations on the plaintext.
- Secret key: The secret key is also input to the encryption algorithm. The key is a value independent of the plaintext and of the algorithm. The algorithm will produce a different output depending on the specific key being used at the time. The exact substitutions and transformations performed by the algorithm depend on the key.
- **Ciphertext:** This is the scrambled message produced as output. It depends on the plaintext and the secret key. For a given message, two different keys will produce two different cipher texts. The cipher text is an apparently random stream of data and, as it stands, is unintelligible.
- **Decryption algorithm:** This is essentially the encryption algorithm run in reverse. It takes the cipher text and the secret key and produces the original plaintext



Simplified Model of Symmetric Encryption

Two requirements for secure use of conventional / symmetric encryption

• We need a strong encryption algorithm

The opponent should be unable to decrypt ciphertext or discover the key even if he or she is in possession of a number of ciphertexts together with the plaintext that produced each ciphertext

• Sender and receiver must have obtained copies of the secret key in a secure fashion and must keep the key secure. If someone can discover the key and knows the algorithm, all

communication using this key is readable, do not need to keep the algorithm secret; we need to keep only the key secret. The principal security problem is maintaining the secrecy of the key

#### Model of Conventional Cryptosystem

A source produces a message in plaintext, X = [X1, X2, ..., XM]. The M elements of X are letters in some finite alphabet. Traditionally, the alphabet usually consisted of the 26 capital letters. Nowadays, the binary alphabet {0, 1} is typically used. For encryption, a key of the form K = [K1, K2, ..., KJ] is generated. If the key is generated at the message source, then it must also be provided to the destination by means of some secure channel. Alternatively, a third party could generate the key and securely deliver it to both source and destination.



With the message X and the encryption key K as input, the encryption algorithm forms the ciphertext Y = [Y1, Y2, ..., YN]. We can write this as

#### $\mathbf{Y} = \mathbf{E}(\mathbf{K}, \mathbf{X})$

This notation indicates that Y is produced by using encryption algorithm E as a function of the plaintext X, with the specific function determined by the value of the key K.

The intended receiver, in possession of the key, is able to invert the transformation:

#### $\mathbf{X} = \mathbf{D}(\mathbf{K}, \mathbf{Y})$

An opponent, observing Y but not having access to K or X, may attempt to recover X or K or both X and K. It is assumed that the opponent knows the encryption (E) and decryption (D)

algorithms. If the opponent is interested in only this particular message, then the focus of the effort is to recover X by generating a plaintext estimate.

## **Substitution Techniques**

- A substitution technique is one in which the letters of plaintext are replaced by other letters or by numbers or symbols
- If the plaintext is viewed as a sequence of bits, then substitution involves replacing plaintext bit patterns with ciphertext bit patterns

#### 1. Caesar Cipher

The earliest known, and the simplest, use of a substitution cipher was by Julius Caesar. The Caesar cipher involves replacing each letter of the alphabet with the letter standing three places further down the alphabet. For example,

plain: meet me after the toga party

cipher: PHHW PH DIWHU WKH WRJD SDUWB

Note that the alphabet is wrapped around, so that the letter following Z is A.

plain: 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 cipher: D E F G H I J K L M N O P Q R S T U V W X Y Z A B C

a	b	с	d	e	f	g	h	i	j	k	1	m
0	1	2	3	4	5	б	7	8	9	10	11	12
n	0	р	q	r	S	t	u	v	W	х	У	Z
13	14	15	16	17	18	19	20	21	22	23	24	25

For each plaintext P substitute the ciphertext letter C

 $C = E(3, p) = (p + 3) \mod 26$ 

A shift may be of any amount, so that the general Caesar algorithm is

 $C = E(k, p) = (p + k) \mod 26$ 

where takes on a value in the range 1 to 25. The decryption algorithm is simply

 $p = D(k, C) = (C - k) \mod 26$ 

#### **Cryptanalysis of Caesar Cipher**

- only have 26 possible ciphers
- A maps to A,B,..Z
- could simply try each in turn
- a brute force search
- given ciphertext, just try all shifts of letters
- do need to recognize when have plaintext

#### **Exercise:**

Plain Text : civil engineering

Key : 7

Cipher Text :

Cipher Text : RTXYFRFENSLUJWXTSGFPMJNX

Key:5

Plain Text : ?

#### **Monoalphabetic Ciphers**

- Rather than just shifting the alphabet shuffle (jumble) the letters arbitrarily
- Each plaintext letter maps to a different random ciphertext letter
- Hence key is 26 letters long

• The "cipher" line can be any permutation of the 26 alphabetic characters, then there are 26! or greater than 4x1026 possible keys.

• This is 10 orders of magnitude greater than the key space for DES and would seem to eliminate brute- force techniques for cryptanalysis

• Monoalphabetic ciphers are easy to break because they reflect the frequency data of the original alphabet

• A countermeasure is to provide multiple substitutes, known as homophones, for a single letter.

• For example, the letter e could be assigned a number of different cipher symbols, such as 16, 74, 35, and 21, with each homophone assigned to a letter in rotation or randomly

Language Redundancy and Cryptanalysis

- human languages are redundant
- eg "th lrd s m shphrd shll nt wnt"
- letters are not equally commonly used
- in English E is by far the most common letter
- followed by T,R,N,I,O,A,S
- other letters like Z,J,K,Q,X are fairly rare
- have tables of single, double & triple letter frequencies for various languages
- two-letter combinations, known as digrams (ex: th)

### 2. Playfair Cipher

The best-known multiple-letter encryption cipher is the Playfair. The Playfair algorithm is based on the use of a  $5 \times 5$  matrix of letters constructed using a keyword.

### **Playfair Key Matrix**

- $5 \times 5$  matrix of letters constructed using a keyword
- filling in the letters of the keyword (minus duplicates) from left to right and from top to bottom,
- filling in the remainder matrix with the remaining letters in alphabetic order.
- The letters I and J count as one letter
- Example matrix using the keyword MONARCHY

Μ	0	Ν	А	R
С	Н	Y	В	D
Е	F	G	I/J	Κ
L	Р	Q	S	Т
U	V	W	Х	Ζ

In this case, the keyword is monarchy. The matrix is constructed by filling in the letters of the keyword (minus duplicates) from left to right and from top to bottom, and then filling in the remainder of the matrix with the remaining letters in alphabetic order. The letters I and J count as one letter. Plaintext is encrypted two letters at a time, according to the following rules:

1. Repeating plaintext letters that are in the same pair are separated with a filler letter, such as x, so that balloon would be treated as ba lx lo on.

2. Two plaintext letters that fall in the same row of the matrix are each replaced by the letter to the right, with the first element of the row circularly following the last. For example, ar is encrypted as RM.

3. Two plaintext letters that fall in the same column are each replaced by the letter beneath, with the top element of the column circularly following the last. For example, mu is encrypted as CM.

4. Otherwise, each plaintext letter in a pair is replaced by the letter that lies in its own row and the column occupied by the other plaintext letter. Thus, hs becomes BP and ea becomes IM / JM.

#### Exercise:

Plain Text : civil engineering Key : abishek Cipher Text : ?

Note: The letters M and N count as one letter

#### Example

Given the key MONARCHY apply Play fair cipher to plain text "FACTIONALISM"

#### Solution

- (p) FA CT IO NA LI SM
- (c) IO DL FA AR SE LA
- (d) FA CT IO NA LI SM

#### Security of Playfair Cipher

- security much improved over monoalphabetic since have  $26 \times 26 = 676$  digrams
- would need a 676 entry frequency table to analyse and correspondingly more ciphertext
- was widely used for many years eg. by US & British military in WW1

it can be broken, given a few hundred letters since still has much of plaintext structure

#### 3. Hill Cipher

Another interesting multiletter cipher is the Hill cipher, developed by the mathematician Lester Hill in 1929.

This encryption algorithm takes successive M plaintext letters and substitutes for them M ciphertext letters. The substitution is determined by linear equations in which each character is assigned a numerical value (a=0, b=1, c=2, , z=25). For M=3, the system can be described as

 $c_1 = (k_{11}p_1 + k_{12}p_2 + k_{13}p_3) \mod 26$   $c_2 = (k_{21}p_1 + k_{22}p_2 + k_{23}p_3) \mod 26$  $c_3 = (k_{31}p_1 + k_{32}p_2 + k_{33}p_3) \mod 26$ 

$$(c_1 \ c_2 \ c_3) = (p \ p_2 \ p_3) \begin{pmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{pmatrix} \mod 26$$

 $\mathbf{C} = \mathbf{PK} \mod 26$ 

where **C** and **P** are row vectors of length 3 representing the plaintext and ciphertext, and **K** is a 3 \* 3 matrix representing the encryption key. Operations are performed mod 26.

#### Example:

Plain Text : paymoremoney

 $\mathbf{K} = \begin{pmatrix} 17 & 17 & 5\\ 21 & 18 & 21\\ 2 & 2 & 19 \end{pmatrix}$ 

The first three letters of the Plain Text are represented by

Cipher Text: LNSHDLEWMTRW

#### Exercise:

Plain Text : FINALYEAR

 $\begin{vmatrix} 2 & 5 & 3 \\ 3 & 1 & 4 \\ 9 & 7 & 6 \end{vmatrix}$  Key : Cipher Text : ?

Cipher Text : XAJOCVDAIUSGDAAUPIAGDGCSGDHAFQGSXI

Key: abishek

Plain Text : ?

Note: The letters M and N count as one letter

#### Example

Encrypt the message "meet me at the usual place at ten rather than eight oclock" using the Hill cipher with the key (). Show your calculations and the result. Show the calculations for the corresponding decryption of the ciphertext to recover the original plaintext.

1) mathematically give each letter a number

abcdefghij klmnopqrstuvw x z 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 2) 1<sup>st</sup> pair from plain text "me" =>  $\binom{12}{4}$  $\binom{9}{5} \binom{4}{7}\binom{12}{4} = >\binom{9x12+4x4}{5x12+7x4} = \binom{124}{88} = > \mod 26 = >\binom{20}{10} = >\binom{u}{k}$ 3) 2<sup>nd</sup> pair fro plain text "et"  $\binom{9}{5} \binom{4}{7}\binom{4}{19} => \binom{9x4+4x19}{5x4+7x19} = \binom{112}{153} => \mod 26 => \binom{8}{23} => \binom{i}{x}$ Cipher text for "meet" is "ukix" 5) To get plain text from cipher text, we need to find the inverse of K 6) |A| = (9x7 - 5x4) => 43 7) Adj (A) =>  $\begin{pmatrix} 7 & -4 \\ -5 & 9 \end{pmatrix}$  =>  $\frac{1}{43}\begin{pmatrix} 7 & -4 \\ -5 & 9 \end{pmatrix}$  =>  $\frac{1}{17}\begin{pmatrix} 7 & -4 \\ -5 & 9 \end{pmatrix}$  ( $\because 43 \% 26 = 17$ ) 8) Find the multiplier for 17, using 17 x X = 1 mod 28 => X = 23  $9 \begin{pmatrix} 161 & -92 \\ -115 & 207 \end{pmatrix} => mod 26 => \begin{pmatrix} 5 & -14 \\ -11 & 25 \end{pmatrix} => \begin{pmatrix} 5 & 12 \\ 15 & 25 \end{pmatrix} (: Add 26 for - ive values)$ 10) P = CK<sup>-1</sup> = > For the ciper text of "uk",  $\begin{pmatrix} 5 & 12 \\ 15 & 25 \end{pmatrix} \begin{pmatrix} 20 \\ 10 \end{pmatrix} = \begin{pmatrix} 5x20 + 12x10 \\ 15x20 + 25x10 \end{pmatrix} = > \begin{pmatrix} 220 \\ 550 \end{pmatrix} \mod 26 = > \begin{pmatrix} 12 \\ 4 \end{pmatrix} = \begin{pmatrix} m \\ 6 \end{pmatrix}$ Hence the plain text is "me"

#### 4. <u>Polyalphabetic Ciphers</u>

Another way to improve on the simple monoalphabetic technique is to use different monoalphabetic substitutions as one proceeds through the plaintext message. The general name for this approach is **polyalphabetic substitution cipher**.

All these techniques have the following features in common:

- A set of related monoalphabetic substitution rules is used.
- A key determines which particular rule is chosen for a given transformation.
  - 5. <u>Vigenere Cipher</u>

#### **Encryption and Decryption**

Given a key letter X and plaintext letter Y, the ciphertext letter is at the intersection of the row labeled X and the column labled Y.

To encrypt a message, a key is needed that is as long as the message. Usually the key is repeating keyword. Decryption is simple. The key letter again identifies the row. The position of the ciphertext letter in that row determines the column, and the plaintext letter is the top of the column.

#### Example:

Key : deceptive Plain Text : we are discovered yourself

key: deceptivedeceptive plaintext: wearediscoveredsaveyourself ciphertext: ZICVTWQNGRZGVTWAVZHCQYGLMGJ

#### **Exercise:**

Plaintext : cryptography and network security Key : sectionb Ciphertext : ?



Vigenere Table

#### Note : Rows represents Plaintext and Columns represents the Key

keyword can be eliminated by using a nonrepeating keyword that is as long as the message itself. Vigenère proposed what is referred to as an **autokey system**, in which a keyword is concatenated with the plaintext itself to provide a running key. For our example,

key: deceptivewearediscoveredsav plaintext: wearediscoveredsaveyourself

ciphertext: ZICVTWQNGKZEIIGASXSTSLVVWLA

**Exercise:** Plaintext : cryptography and network security Key : sectionb Ciphertext : ?

#### **Vernam Cipher**

The ultimate defense against such a cryptanalysis is to choose a keyword that is as long as the plaintext and has no statistical relationship to it. Such a system was introduced by an AT&T engineer named Gilbert Vernam in 1918.

 $c_i = p_i \oplus k_i$ 

where

 $p_i = i$ th binary digit of plaintext

 $k_i = i$ th binary digit of key

 $c_i = i$ th binary digit of ciphertext

⊕ = exclusive-or (XOR) operation



#### 6. One-Time Pad

- improvement to the Vernam cipher that yields the ultimate in security
- using a random key that is as long as the message, so that the key need not be repeated
- the key is to be used to encrypt and decrypt a single message, and then is discarded.
- Each new message requires a new key of the same length as the new message **Example** ciphertext:ANKYODKYUREPFJBYOJDSPLREYIUNOFDOIUERFPLUYTS key: pxlmvmsydofuyrvzwc tnlebnecvgdupahfzzlmnyih plaintext: mr mustard with the candlestick in the hall

ciphertext:ANKYODKYUREPFJBYOJDSPLREYIUNOFDOIUERFPLUYTS key: mfugpmiydgaxgoufhklllmhsqdqogtewbqfgyovuhwt plaintext: miss scarlet

with the knife in the library **two fundamental difficulties** 

- problem of making large quantities of random keys
- problem of key distribution and protection

## **Transposition Techniques**

A very different kind of mapping is achieved by performing some sort of permutation on the plaintext letters

#### **Rail Fence Technique**

The simplest such cipher is the rail fence technique, in which the plaintext is written down as a sequence of diagonals and then read off as a sequence of rows.

For example, to encipher the message "meet me after the toga party" with a rail fence of depth 2, we write the following

m em atrhtgpry etefeteoaat

The encrypted message is

MEMATRHTGPRYETEFETEOAAT

#### **Pure Transposition Cipher**

Write the message in a rectangle, row by row, and read the message off, column by column, but permute the order of the columns.

The order of the columns then becomes the key to the algorithm

#### Example

Key: 4 3 1 2 5 6 7

Plaintext: a t t a c k p

ostponed untiltwo amxyz

Ciphertext: TTNAAPTMTSUOAODWCOIXKNLYPETZ

#### **Double Transposition**

performing more than one stage of transposition Example

if the foregoing message is reencrypted using the same algorithm

 Key:
 4
 3
 1
 2
 5
 6
 7

 Input:
 t
 t
 n
 a
 pt

 m
 t
 su
 o
 ao

```
d wcoix k
n ly pet z
```

Output: NSCYAUOPTTWLTMDNAOIEPAXTTOKZ

This is a much less structured permutation and is much more difficult to cryptanalyze

## Steganography

We conclude with a discussion of a technique that is, strictly speaking, not encryption, namely, steganography

A plaintext message may be hidden in one of two ways.

- [ The methods of steganography conceal the existence of the message
- <sup>[]</sup> The methods of cryptography render the message unintelligible to outsiders o by various transformations of the text

Various ways to conceal the message

## Arrangement of words or letters within an apparently innocuous text spells out the real message

#### **Character marking**

Selected letters of printed or typewritten text are overwritten in pencil. The marks are ordinarily not visible unless the paper is held at an angle to bright light.

#### Invisible ink

A number of substances can be used for writing but leave no visible trace until heat or some chemical is applied

#### **Pin punctures**

Small pin punctures on selected letters are ordinarily not visible unless the paper is held up in front of a light.

#### **Typewriter correction ribbon**

Used between lines typed with a black ribbon, the results of typing with the correction tape are visible only under a strong light

#### Hiding a message by using the least significant bits of frames on a CD

• The Kodak Photo CD format's maximum resolution is 2048 by 3072 pixels, with each pixel containing 24 bits of RGB color information.

- The least significant bit of each 24-bit pixel can be changed without greatly affecting the quality of the image
- Thus you can hide a 2.3-megabyte message in a single digital snapshot

#### Number of drawbacks

- lot of overhead to hide a relatively few bits of information
- once the system is discovered, it becomes virtually worthless
- the insertion method depends on some sort of key
- o Alternatively, a message can be first encrypted and then hidden using steganography

#### Advantage of steganography

- can be employed by parties who have something to lose should the fact of their secret communication be discovered
- Encryption flags traffic as important or secret or may identify the sender or receiver as someone with something to hide

## **Foundations of Modern Cryptography**

Modern cryptography is the cornerstone of computer and communications security. Its foundation is based on various concepts of mathematics such as number theory, computational-complexity theory, and probability theory.

#### **Characteristics of Modern Cryptography**

There are three major characteristics that separate modern cryptography from the classical approach.

Classic Cryptography	Modern Cryptography
It manipulates traditional characters, i.e., letters and digits directly.	It operates on binary bit sequences.
It is mainly based on 'security through obscurity'. The techniques employed for coding were kept secret and only the parties involved in communication knew about them.	It relies on publicly known mathematical algorithms for coding the information. Secrecy is obtained through a secrete key which is used as the seed for the algorithms. The computational difficulty of algorithms, absence of secret key, etc., make it impossible for an attacker to obtain the original information even if he knows the algorithm used for coding.
It requires the entire cryptosystem for communicating confidentially.	Modern cryptography requires parties interested in secure communication to possess the secret key only.

## **Perfect Security**

What is a "secure" cipher? Intuitively, the answer is that a secure cipher is one for which an encrypted message remains "well hidden," even after seeing its encryption.

However, turning this intuitive answer into one that is both mathematically meaningful and practically relevant is a real challenge. Indeed, although ciphers have been used for centuries, it is only in the last few decades that mathematically acceptable definitions of security have been developed.

In this section, we develop the mathematical notion of perfect security — this is the "gold standard" for security (at least, when we are only worried about encrypting a single message and do not care about integrity). We will also see that it is possible to achieve this level of security; indeed, we will show that the one-time pad satisfies the definition. However, the one-time pad is not very practical, in the sense that the keys must be as long as the messages: if Alice wants to send a 1GB file to Bob, they must already share a 1GB key! Unfortunately, this cannot be avoided: we will also prove that any perfectly secure cipher must have a key space at least as large as its message space. This fact provides the motivation for developing a definition of security that is weaker, but that is acceptable from a practical point of view, and which allows one to encrypt long messages using short keys.

If Alice encrypts a message m under a key k, and an eavesdropping adversary obtains the cipher text c, Alice only has a hope of keeping m secret if the key k is hard to guess, and that means, at the very least, that the key k should be chosen at random from a large key space. To

say that m is "well hidden" must at least mean that it is hard to completely determine m from c, without knowledge of k; however, this is not really enough. Even though the adversary may not know k, we assume that he does know the encryption algorithm and the distribution of k. In fact, we will assume that when a message is encrypted, the key k is always chosen at random, uniformly from among all keys in the key space. The adversary may also have some knowledge of the message encrypted — because of circumstances, he may know that the set of possible messages is quite small, and he may know something about how likely each possible message is. For example, suppose he knows the message m is either m0 = "ATTACK AT DAWN" or m1 = "ATTACK AT DUSK", and that based on the adversary's available intelligence, Alice is equally likely to choose either one of these two messages. This, without seeing the cipher text c, the adversary would only have a 50% chance of guessing which message Alice sent. But we are assuming the adversary does know c. Even with this knowledge, both messages may be possible; that is, there may exist keys k0 and k1 such that E(k0,m0) = c and E(k1,m1) = c, so he cannot be sure if m = m0 or m = m1.

However, he can still guess. Perhaps it is a property of the cipher that there are 800 keys k0 such that E(k0,m0) = c, and 600 keys k1 such that E(k1,m1) = c. If that is the case, the adversary's best guess would be that m = m0. Indeed, the probability that this guess is correct is equal to 800/(800 + 600)  $\ddagger$  57%, which is better than the 50% chance he would have without knowledge of the ciphertext. Our formal definition of perfect security expressly rules out the possibility that knowledge of the ciphertext increases the probability of guessing the encrypted message, or for that matter, determining any property of the message whatsoever.

Without further ado, we formally define perfect security. In this definition, we will consider a probabilistic experiment in which is key is drawn uniformly from the key space. We write k to denote the random variable representing this random key. For a message m, E(k,m) is another random variable, which represents the application of the encryption function to our random key and the message m. Thus, every message m gives rise to a different random variable E(k,m).

#### **Definition of Perfect Security**

Let  $\varepsilon = (E,D)$  be a Shannon cipher defined over (K,M, C). Consider a probabilistic experiment in which the random variable k is uniformly distributed over K. If for all m0,m1  $\in$  M, and all  $c \in$  C, we have

**Pr** [ E(k,m0) = c ] = **Pr** [ E(k,m1) = c ], then we say that  $\varepsilon$  is a perfectly secure Shannon cipher.

There are a number of equivalent formulations of perfect security

**Theorem 2.4.** Let  $\mathcal{E} = (E, D)$  be a Shannon cipher defined over  $(\mathcal{K}, \mathcal{M}, \mathcal{C})$ . Consider a random experiment in which k and m are random variables, such that

- k is uniformly distributed over K,
- m is distributed over M, and
- k and m are independent.

Define the random variable  $\mathbf{c} := E(\mathbf{k}, \mathbf{m})$ . Then we have:

- if  $\mathcal{E}$  is perfectly secure, then c and m are independent;
- conversely, if c and m are independent, and each message in M occurs with nonzero probability, then E is perfectly secure.

*Proof.* We define  $\mathcal{M}^*$  to be the set of messages that occur with nonzero probability.

We begin with a simple observation. Consider any fixed  $m \in \mathcal{M}^*$  and  $c \in \mathcal{C}$ . Then we have

$$\Pr[\mathbf{c} = c \mid \mathbf{m} = m] = \Pr[E(\mathbf{k}, m) = c \mid \mathbf{m} = m],$$

and since k and m are independent, so are  $E(\mathbf{k}, m)$  and m, and hence

$$\Pr[E(\mathbf{k}, m) = c \mid \mathbf{m} = m] = \Pr[E(\mathbf{k}, m) = c].$$

Putting this all together, we have:

$$\Pr[\mathbf{c} = c \mid \mathbf{m} = m] = \Pr[E(\mathbf{k}, m) = c].$$
(2.1)

We now prove the first implication. So assume that  $\mathcal{E}$  is perfectly secure. We want to show that c and m are independent. To to this, let  $m \in \mathcal{M}^*$  and  $c \in \mathcal{C}$  be given. It will suffice to show that

$$\Pr[\mathbf{c} = c \mid \mathbf{m} = m] = \Pr[\mathbf{c} = c].$$

We have

$$\begin{split} \Pr[\mathbf{c} = c] &= \sum_{m' \in \mathcal{M}^*} \Pr[\mathbf{c} = c \mid \mathbf{m} = m'] \Pr[\mathbf{m} = m'] \quad \text{(by total probability)} \\ &= \sum_{m' \in \mathcal{M}^*} \Pr[E(\mathbf{k}, m') = c] \Pr[\mathbf{m} = m'] \quad \text{(by (2.1))} \\ &= \sum_{m' \in \mathcal{M}} \Pr[E(\mathbf{k}, m) = c] \Pr[\mathbf{m} = m'] \quad \text{(by the definition of perfect security)} \\ &= \Pr[E(\mathbf{k}, m) = c] \sum_{m' \in \mathcal{M}^*} \Pr[\mathbf{m} = m'] \\ &= \Pr[E(\mathbf{k}, m) = c] \quad (\text{probabilities sum to 1}) \\ &= \Pr[\mathbf{c} = c \mid \mathbf{m} = m] \quad (\text{again by (2.1)}) \end{split}$$

This shows that c and m are independent.

That proves the first implication. For the second, we assume that c and m are independent, and moreover, that every message occurs with nonzero probability (so  $\mathcal{M}^* = \mathcal{M}$ ). We want to show that  $\mathcal{E}$  is perfectly secure, which means that for each  $m_0, m_1 \in \mathcal{M}$ , and each  $c \in \mathcal{C}$ , we have

$$\Pr[E](\mathbf{k}, m_0) = c] = \Pr[E(\mathbf{k}, m_1) = c].$$
(2.2)

But we have

 $\begin{aligned} \Pr[E(\mathbf{k}, m_0) = c] &= \Pr[\mathbf{c} = c \mid \mathbf{m} = m_0] \quad (by \ (2.1)) \\ &= \Pr[\mathbf{c} = c] \quad (by \ \text{independence of } \mathbf{c} \ \text{and} \ \mathbf{m}) \\ &= \Pr[\mathbf{c} = c \mid \mathbf{m} = m_1] \quad (again \ by \ \text{independence of } \mathbf{c} \ \text{and} \ \mathbf{m}) \\ &= \Pr[E(\mathbf{k}, m_1) = c] \quad (again \ by \ (2.1)). \end{aligned}$ 

That shows that  ${\mathcal E}$  is perfectly secure.  $\Box$ 

## **Information Theory**

For a long time, information theory has mainly been used in cryptography to prove lower bounds on the size of the secret key required to achieve a certain level of security in secrecy and authentication systems. In order to prove the security of a cryptographic system, a definition of security or, alternatively, of breaking the system must be given. Whether a system with provable security is satisfactory from a theoretical and practical viewpoint depends in a crucial manner on three aspects:

(1) On the acceptability and generality of the definition of security;

(2) On how realistic the two assumptions are; and

(3) On the practicality of the system.

For instance, it is trivial to \prove" the security of a cipher if we define security (irrelevantly) to mean that an adversary is unable to square a circle with straightedge and compass. It is similarly trivial to prove that an adversary cannot obtain any information about the plaintext for a system in which the legitimate receiver cannot either, or if one assumes that the adversary is unable to even receive the ciphertext.

There are two possible types of assumptions about the adversary's computing power:

- A system is called computationally-secure if it is secure against an adversary with reasonably bounded computational resources and it is called information-theoretically secure if it is secure even against adversaries with infinite computing power
- The second type of assumption, namely that an adversary has infinite computing power, implies no restriction whatsoever and therefore anticipates all arguments about model's of computation and realistic estimates of an opponent's computing power.

However, if one considers the theoretical possibility of testing all possible keys of a system at once, it appears impossible to prove a system secure under such an assumption. Here is where information theory comes into play. Shannon defined a cipher system to be perfect if the ciphertext provides no information about the plaintext or, equivalently, if plaintext and ciphertext are

statistically independent. In other words, when a perfect cipher is used to encipher a message, an adversary can do no better than guess the message without even looking at the ciphertext.

The role of information theory in cryptography can be characterized as that of deriving results on the provable security of a system, even in presence of adversaries with infinite computing power.

#### **Basics of Information Theory**

According to Shannon, the entropy of an information source S is defined as:

$$H(S) = \eta = \sum_{i} p_i \log_2 \frac{1}{p_i}$$

where pi is the probability that symbol Si in S will occur.

$$\log_2 \frac{1}{p_i}$$

• indicates the amount of information contained in Si, i.e., the number of bits needed to code Si.

• For example, in an image with uniform distribution of gray-level intensity, i.e. pi = 1/256, then the number of bits needed to code each gray level is 8 bits. The entropy of this image is 8.

#### Pessimistics Results: Lower bounds on key size

Information theory has been used in cryptography primarily to derive pessimistic results, i.e., lower bounds on the size of the secret key necessary to achieve a certain level of security. In this section we review the three most important areas for which such bounds have been derived: secrecy, Y This is a generalization of Shannon's model: it contains a secret randomizer S known only to the sender of a message X as well as a public randomizer R assumed to be available to everybody, including the eavesdropper.

#### Model of symmetric cipher with two types of Randomizer



Key Source

There are two dual and complementary security goals in communication: Confidentiality (or secrecy) and authenticity. Confidentiality means that an eavesdropper cannot obtain any useful information about the plaintext, and authenticity means that 5 an active eavesdropper cannot successfully insert a fraudulent message  $\mathbf{Y}$  that will be accepted by the receiver

#### Information-theoretic security

Information-theoretic security is a cryptosystem whose security derives purely from information theory; the system cannot be broken even if the adversary has unlimited computing power. The cryptosystem is considered cryptanalytically unbreakable if the adversary does not have enough information to break the encryption.

There are a variety of cryptographic tasks for which information-theoretic security is a meaningful and useful requirement. A few of these are:

- 1. Secret sharing schemes such as Shamir's are information-theoretically secure (and also perfectly secure) in that having less than the requisite number of shares of the secret provides no information about the secret.
- 2. More generally, secure multiparty computation protocols often have informationtheoretic security.
- 3. Private information retrieval with multiple databases can be achieved with informationtheoretic privacy for the user's query.
- 4. Symmetric encryption can be constructed under an information-theoretic notion of security called entropic security, which assumes that the adversary knows almost nothing about the message being sent. The goal here is to hide all functions of the plaintext rather than all information about it.
- 5. Quantum cryptography is largely part of information-theoretic cryptography.

## **Product Cryptosystem**

Two of the first kinds of cryptosystems that we considered were simple substitution ciphers and permutation ciphers. Each of them quickly proved vulnerable to attack. We now consider a new kind of cryptosystem that is based on them but which is considerably more difficult to attack; so difficult, in fact, that most modern cryptosystems are of the type we now consider. A product cryptosystem is a block cipher that repeatedly performs substitutions and permutations, one after the other, to produce ciphertext.

Example : DES and AES (Brief description in Unit II)

## **Cryptanalysis and Brute-Force Attack**

#### Cryptanalysis

Cryptanalysis is the science of cracking codes and decoding secrets. It is used to violate authentication schemes, to break cryptographic protocols, and, more benignly, to find and correct weaknesses in encryption algorithms.

It may be used in information warfare applications - for example, forging an encrypted signal to be accepted as authentic. Competitors who have been able to discover the key will now want to use it to their advantage, therefore they will want to send bogus encrypted messages to the source in order to gain information or gain an advantage. It could also be used to pretend to be the source in order to send bogus information to others, who now will think that it came from the official source.

#### According to Diffie and Hellman

Skill in the production of cryptanalysis has always been heavily on the side of the professionals, but innovation, particularly in the design of new types of cryptographic systems, has come primarily from amateurs.

Among the types of attacks are:

Ciphertext only attacks Known plaintext attacks Chosen plaintext attacks Chosen ciphertext attacks Man-in-the-middle attacks Side channel attacks Brute force attacks Birthday attacks

Type of Attack	Known to Cryptanalyst
Ciphertext Only	Encryption algorithm
	• Ciphertext
Known Plaintext	Encryption algorithm
	• Ciphertext
	• One or more plaintext-ciphertext pairs formed with the
	secret key
Chosen plaintext	Encryption algorithm
	• Ciphertext
	• Plaintext message chosen by cryptanalyst, together with its
	corresponding ciphertext generated with the secret key
Chosen ciphertex	Encryption algorithm

	<ul> <li>Ciphertext</li> <li>Purported ciphertext chosen by cryptanalyst, together with its corresponding decrypted plaintext generated with the secret key</li> </ul>
Chosen text	<ul> <li>Encryption algorithm</li> <li>Ciphertext</li> <li>Plaintext message chosen by cryptanalyst, together with its corresponding ciphertext generated with the secret key</li> <li>Purported ciphertext chosen by cryptanalyst, together with its corresponding decrypted plaintext generated with the secret key</li> </ul>

#### **Ciphertext Only**

A ciphertext only attack (COA) is a case in which only the encrypted message is available for attack, but because the language is known a frequency analysis could be attempted. In this situation the attacker does not know anything about the contents of the message, and must work from ciphertext only.



#### **Known Plaintext Attack**

In a known plaintext attack (KPA) both the plaintext and matching ciphertext are available for use in discovering the key.

The attacker knows or can guess the plaintext for some parts of the ciphertext. For example, maybe all secure login sessions begin with the characters LOGIN, and the next transmission may be PASSWORD. The task is to decrypt the rest of the ciphertext blocks using this information. This may be done by determining the key used to encrypt the data, or via some shortcut.



#### **Chosen Plaintext Attack**

A chosen plaintext attack (CPA) occurs when the attacker gains access to the target encryption device - if, for example, it is left unattended. The attacker then runs various pieces of plaintext though the device for encryption. This is compared to the plaintext to attempt to derive the key.

In an adaptive chosen plaintext attack (ACPA), the attacker not only has access to the plaintext and its encryption, but can adapt or modify the chosen plaintext as needed based on results of the previous encryptions.



#### **Chosen Ciphertext Attack**

In a chosen ciphertext attack (CCA), the cryptanalyst can choose different ciphertexts to be decrypted and has access to the decrypted plaintext.

This type of attack is generally applicable to attacks against public key cryptosystems.

An adaptive chosen ciphertext attack involves the attacker selecting certain ciphertexts to be decrypted, then using the results of these decryptions to select subsequent ciphertexts. The modifications in the ciphertext help in deciphering the key from the decryptions.



#### **Two schemes**

#### • Unconditionally secure

If the ciphertext generated by the scheme does not contain enough information to determine uniquely the corresponding plaintext, no matter how much ciphertext is available

#### • Computationally secure

meets either of the following criteria:

- The cost of breaking the cipher exceeds the value of the encrypted information.
- The time required to break the cipher exceeds the useful lifetime of the information.

#### **Brute-force attack**

- The attacker tries every possible key on a piece of ciphertext until an intelligible translation into plaintext is obtained.
- On average, half of all possible keys must be tried to achieve success.