

CSE 4351: Computer Security



Instructor: Dr. MD Armanuzzaman

Cryptography



What is Cryptography

- Greek: “krypto” = hide
- Cryptography – secret writing. Originally, it is the study of encryption principles and methods
 - The most basic problem of cryptography is to ensure security of communication over insecure media
- Cryptographer

What is Cryptography

- Cryptanalysis – analyzing and breaking secrets. Originally, the study of principles and methods of deciphering ciphertext without knowing key
- Cryptanalyst

Cryptography

Can help

- **Confidentiality**
 - Obscure a message from eavesdroppers
- **Integrity**
 - Assure recipient that the message was not altered
- **Authenticity**
 - Verify the identity of the source of a message
- **Non-repudiation**
 - Convince a 3rd party that what was said is accurate

Cryptography & Security

- Most people argue *cryptology is a branch of mathematics*
- Security is about math, engineering, hardware, software, people, etc.
- **Attackers try find the weakest link.** In most cases, this is not the mathematics.
- Example: HeartBleed

Cryptography & Security

- Cryptographic tools are essential in designing secure solutions and their understanding is crucial to correct usage
- We'll look at these types of cryptographic tools
 - symmetric encryption
 - hash functions and message authentication codes
 - public-key encryption
 - digital signatures and certificates
 - pseudo-random number generators

Terminology

- *Plaintext*: the message to be transmitted or stored.
- *Ciphertext*: the disguised message.

- *Key*: Sequence that controls the operation and behavior of the cryptographic algorithm
- *Keyspace*: Total number of possible values of keys in a crypto algorithm

Terminology

- *Encryption*: the process of **disguising** a message so as to hide the information it contains; this process can include both encoding and enciphering.
- *Protocol*: an algorithm, defined by a **sequence of steps**, precisely specifying the actions of multiple parties in order to achieve an objective.
- *Cryptosystem*: The **combination of algorithm, key, and key management** functions used to perform cryptographic operations

Kerckhoff's Principle

- French handbook of military cryptography, 1883
- A cryptosystem should be secure even if everything about the system, except the key, is public knowledge.
- Don't assume enemy won't know algorithm
 - Can capture machines, find patents, etc.
 - Too expensive to invent new algorithm if it might have been compromised

Security by obscurity is never a good solution

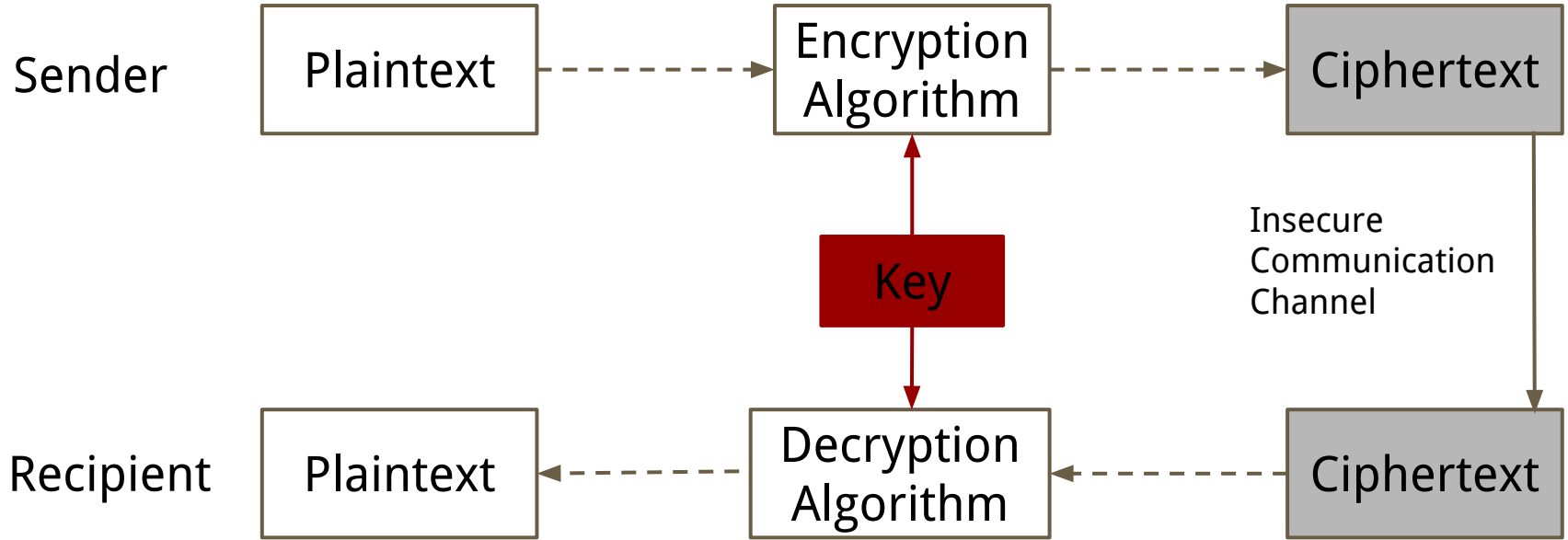
Symmetric Encryption I



Symmetric Encryption

- Symmetric (or secret-key) encryption means that the **same key** is used both for encryption and decryption
- The key must **remain secret** at both ends
- Such algorithms are:
 - normally very fast
 - can be used as primitives in more complex cryptographic protocols
 - the key often has a short lifetime

Symmetric Encryption



$$\text{Ciphertext} = E(K, \text{Plaintext})$$
$$\text{Plaintext} = D(K, \text{Ciphertext})$$

Symmetric Encryption

$Ciphertext = E(K, Plaintext)$

$Plaintext = D(K, Ciphertext)$

Desired properties:

1. Kerckhoff's: secrecy depends only on K
2. Without knowing K ; must be “hard” to invert
3. Easy to compute E and D

All cryptosystems until 1970s were like this.

Asymmetric cryptosystems allow encryption and decryption keys to be different.

Symmetric Encryption Formally

More formally, a computationally secure symmetric key encryption scheme is defined as:

- a private-key encryption scheme consists of polynomial-time algorithms (Gen , Enc , Dec) such that
 - 1. Gen : on input the security parameter n , outputs key k
 - 2. Enc : on input a key k and a message $m \in \{0, 1\}^*$, outputs ciphertext c
 - 3. Dec : on input a key k and ciphertext c , outputs plaintext m
- we write $k \leftarrow Gen(1^n)$, $c \leftarrow Enc_k(m)$, and $m := Dec_k(c)$

Classic Symmetric Encryption Algorithms



Caesar Cipher

- Caesar used to *encrypt* his messages to communicate with his generals
- He would take each letter of the alphabet and replace it with a letter a certain distance away from that letter. When he got to the end, he would wrap back around to the beginning.
- Example with a shift of 3

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
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

Caesar Cipher

- Shift table

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
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

- Key: 3
- Plaintext: ATTACK
- Ciphertext: DWWDFN
- **Keyspace: 25**

Caesar Cipher

- Linux tr command

```
tr A-Za-z D-ZA-Cd-za-c
```

How to Attack Caesar Cipher?

- Attackers
 - Don't know the key or the shift table
 - Only knows Ciphertext
- Make a list of all possible keys and corresponding plaintext. If you expected the unencrypted text to be in English, you could easily figure out which word was right
- Small key space (25)

Extend Caesar to Simple Substitution Cipher

- Substitute each letter based on mapping

- Key is alphabet mapping:

a -> J, b -> L, c -> B, d -> R, ..., z -> F

Substitution Cipher

- Number of possible keys

26 (ways to choose what a maps to)

* 25 (b can map to anything else)

* 24 (c can map to anything else)

... * 1 (only one choice left for z)

= $26! = 403291461126605635584000000$

If every person on earth tried one per second, it would take 5B years to try them all.

How secure is this cipher?

How to Attack Simple Substitution Cipher?

VKY THSZYAPSVR NO VYJDP DV YC LDPN (TVYL) SP D LTQCSU AYPYDAUK THSZYAPSVR
CNUDVYX SH YC LDPN, VYJDP. ONTHXYX SH 1914, TVYL SP D GYGQYA NO VKY THSZYAPSVR
NO VYJDP PRPVYG DHX PYAZYP DP D GDFNA AYPYDAUK DHX YXTUDVSNHDC
SHPVSVTVSNH SH VKY YC LDPN AYMSNH. VKY THSZYAPSVR NOOYAP D ISXY ADHMY NO
THXYAMADXTDVY, MADXTDVY, DHX XNUVNADC LANMADGP DHX SP AYUNMHSWYX ONA
SVP UNGGSVGYHV VN AYPYDAUK, SHHNZDVSNH, DHX PVTXYHV PTUUYP.

Letter Frequency Analysis

VKY THSZYAPSVR NO VYJDP DV YC LDPN (TVYL) SP D LTQCSU AYPYDAUK THSZYAPSVR
CNUDVYX SH YC LDPN, VYJDP. ONTHXYX SH 1914, TVYL SP D GYGQYA NO VKY THSZYAPSVR
NO VYJDP PRPVYG DHX PYAZYP DP D GDFNA AYPYDAUK DHX YXTUDVSNHDC
SHPVSVTVSNH SH VKY YC LDPN AYMSNH. VKY THSZYAPSVR NOOYAP D ISXY ADHMY NO
THXYAMADXTDVY, MADXTDVY, DHX XNUVNADC LANMADGP DHX SP AYUNMHSWYX ONA
SVP UNGGSVGYHV VN AYPYDAUK, SHHNZDVSNH, DHX PVTXYHV PTUUYPP.

e 0.12702	s 0.06327	u 0.02758	p 0.01929	q 0.00095
t 0.09056	h 0.06094	m 0.02406	b 0.01492	z 0.00074
a 0.08167	r 0.05987	w 0.02360	v 0.00978	
o 0.07507	d 0.04253	f 0.02228	k 0.00772	
i 0.06966	l 0.04025	g 0.02015	j 0.00153	
n 0.06749	c 0.02782	y 0.01974	x 0.00150	

Word Frequency Analysis

VKY THSZYAPSVR NO VYJDP DV YC LDPN (TVYL) SP D LTQCSU AYPYDAUK THSZYAPSVR
CNUDVYX SH YC LDPN, VYJDP. ONTHXYX SH 1914, TVYL SP D GYGQYA NO VKY THSZYAPSVR
NO VYJDP PRPVYG DHX PYAZYP DP D GDFNA AYPYDAUK DHX YXTUDVSNHDC
SHPVSVTVSNH SH VKY YC LDPN AYMSNH. VKY THSZYAPSVR NOOYAP D ISXY ADHMY NO
THXYAMADXTDVY, MADXTDVY, DHX XNUVNADC LANMADGP DHX SP AYUNMHSWYX ONA
SVP UNGGSVGYHV VN AYPYDAUK, SHHNZDVSNH, DHX PVTXYHV PTUUYPP.

Most common trigrams in English:

the = 6.4%

and = 3.4%

Word Frequency Analysis

THE THSZEAPSTR NO TEJDP DT EC LDPN (TTEL) SP D LTQCSU AEPEDAUH THSZEAPSTR
CNUDTEX SH EC LDPN, TEJDP. ONTHXEX SH 1914, TTEL SP D GEGQEA NO THE THSZEAPSTR
NO TEJDP PRPTEG DHX PEAZEP DP D GDFNA AEPEDAUH DHX EXTUDTSNHDC SHPTSTTTSNH
SH THE EC LDPN AEMSNH. THE THSZEAPSTR NOOEAP D ISXE ADHME NO THXEAMADXTDTE,
MADXTDTE, DHX XNUTNADC LANMADGP DHX SP AEUNMHSWEX ONA STP UNGGSTGEHT TN
AEPEDAUH, SHHNZDTSNH, DHX PTTXEHT PTUUEPP.

tr VKY THE

Guess?

THE THSZEAPSTR NO TEJDP DT EC LDPN (TTEL) SP D LTQCSU AEPEDAUH THSZEAPSTR
CNUDTEX SH EC LDPN, TEJDP. ONTHXEX SH 1914, TTEL SP D GEGQEA NO THE THSZEAPSTR
NO TEJDP PRPTEG DHX PEAZEP DP D GDFNA AEPEDAUH DHX EXTUDTSNHDC SHPTSTTTSNH
SH THE EC LDPN AEMSNH. THE THSZEAPSTR NOOEAP D ISXE ADHME NO THXEAMADXTDTE,
MADXTDTE, DHX XNUTNADC LANMADGP DHX SP AEUNMHSWEX ONA STP UNGGSTGEHT TN
AEPEDAUH, SHHNZDTSNH, DHX PTTXEHT PTUUEPP.

Guess?

THE THSZEAPSTR NO TEJAP AT EC LAPN (TTEL) SP A LTQCSU AEPEAAUH THSZEAPSTR
CNUATEX SH EC LAPN, TEJAP. ONTHXEX SH 1914, TTEL SP A GEGQEA NO THE THSZEAPSTR NO
TEJAP PRPTEG AHX PEAZEP AP A GAFNA AEPEAAUH AHX EXTUATSNHAC SHPTSTTTSNH SH
THE EC LAPN AEMSNH. THE THSZEAPSTR NOOEAP A ISXE AAHME NO THXEAMAAXTATE,
MAAXTATE, AHX XNUTNAAC LANMAAGP AHX SP AEUNMHSWEX ONA STP UNGGSTGEHT TN
AEPEAAUH, SHHNZATSNH, AHX PTTXEHT PTUUEPP.

tr VKYD THEA

Guess?

THE THSZEAPSTR NO TEJAP AT EC LAPN (TTEL) SP A LTQCSU AEPEAAUH THSZEAPSTR
CNUATEX SH EC LAPN, TEJAP. ONTHXEX SH 1914, TTEL SP A GEGQEA NO THE THSZEAPSTR
NO TEJAP PRPTEG AHX PEAZEP AP A GAFNA AEPEAAUH AHX EXTUATSNHAC SHPTSTTTSNH
SH THE EC LAPN AEMSNH. THE THSZEAPSTR NOOEAP A ISXE AAHME NO THXEAMAAXTATE,
MAAXTATE, AHX XNUTNAAC LANMAAGP AHX SP AEUNMHSWEX ONA STP UNGGSTGEHT TN
AEPEAAUH, SHHNZATSNH, AHX PTTXEHT PTUUEPP.

tr VKYD THEA

Guess?

THE THSZEASSTR NO TEXAS AT EC LASN (TTEL) SS A LTQCSU AESEAAUH THSZEASSTR
CNUATEX SH EC LASN, TEXAS. ONTHXEX SH 1914, TTEL SS A GEGQEA NO THE THSZEASSTR NO
TEXAS SRSTEG AHX SEAZES AS A GAFNA AESEAAUH AHX EXTUATSNHAC SHSTSTTTSNH SH
THE EC LASN AEMSNH. THE THSZEASSTR NOOEAS A ISXE AAHME NO THXEAMAAXTATE,
MAAXTATE, AHX XNUTNAAC LANMAAGS AHX SS AEUNMHSWEX ONA STS UNGGSTGEHT TN
AESEAAUH, SHHNZATSNH, AHX STTXEHT STUUESS.

tr VKYDJP THEAXS

Pattern Analysis and Guess

THE UNIVERSITY OF TEXAS AT EL PASO (UTEP) IS A PUBLIC RESEARCH UNIVERSITY LOCATED IN EL PASO, TEXAS. FOUNDED IN 1914, UTEP IS A MEMBER OF THE UNIVERSITY OF TEXAS SYSTEM AND SERVES AS A MAJOR RESEARCH AND EDUCATIONAL INSTITUTION IN THE EL PASO REGION. THE UNIVERSITY OFFERS A WIDE RANGE OF UNDERGRADUATE, GRADUATE, AND DOCTORAL PROGRAMS AND IS RECOGNIZED FOR ITS COMMITMENT TO RESEARCH, INNOVATION, AND STUDENT SUCCESS.

tr A-Z DQUXYOMKSFBCGHNLEAPVTZIJRW

tr DQUXYOMKSFBCGHNLEAPVTZIJRW A-Z

Simple Substitution Doesn't Work

- A large space of keys is not enough
- Monoalphabetic
 - The same plaintext letters are always replaced by the same ciphertext letters
- Doesn't hide statistical properties of plaintext. Doesn't hide relationships in plaintext
- Natural languages are very redundant

Make it Harder?

- Hide statistical properties
 - Encrypt “e” with 12 different symbols, “t” with 9 different symbols, etc.
- Polyalphabetic cipher
 - Use different substitutions
- Transposition (permutation)
 - Scramble order of units; reorder units of plaintext

Transposition Cipher

A Transposition Cipher is a cipher in which the plaintext message is rearranged by some means.

These hide the message by rearranging the letter order, without altering the actual letters used.

Also called **permutation ciphers**.

Transposition Cipher

T O D A Y

+ I S + M

O N D A Y

Write across rows. Read in a counter clockwise spiral from top-left.

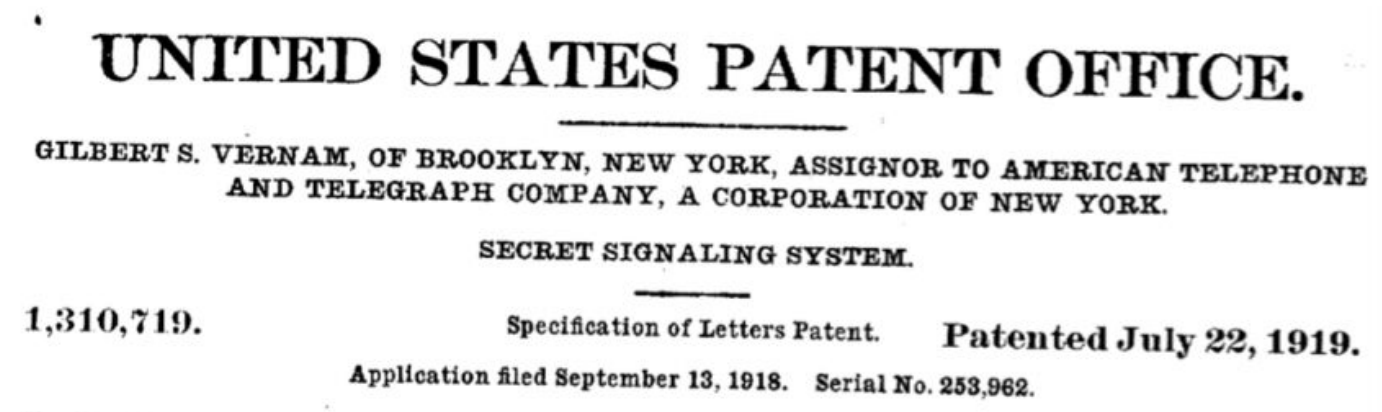
Encrypt(TODAY+IS+MONDAY) = T+ONDAYMYADOIS+

Perfectly Secure Cipher: One-Time Pad

- 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.

Perfectly Secure Cipher: One-Time Pad

- In 1919, Gilbert Verman got a patent that states a cipher in which each character in a message was electrically combined with a character on a paper tape key



Perfectly Secure Cipher: One-Time Pad

- Later, U.S. Army captain Joseph Mauborgne recognized that if the character sequence on the key tape could be completely random, cryptanalysis would be more difficult
- Proved perfectly secure by Claude Shannon in “Communication Theory of Secrecy Systems”, 1949

Communication Theory of Secrecy Systems*

By C. E. SHANNON

1. INTRODUCTION AND SUMMARY

THE problems of cryptography and secrecy systems furnish an interesting application of communication theory.¹ In this paper a theory of secrecy systems is developed. The approach is on a theoretical level and is intended to complement the treatment found in standard works on cryptography.² There, a detailed study is made of the many standard types of codes and ciphers, and of the ways of breaking them. We will be more concerned with the general mathematical structure and properties of secrecy systems.

The treatment is limited in certain ways. First, there are three general types of secrecy system: (1) concealment systems, including such methods as invisible ink, concealing a message in an innocent text, or in a fake covering cryptogram, or other methods in which the existence of the message is concealed from the enemy; (2) privacy systems, for example speech inversion, in which special equipment is required to recover the message; (3) “true” secrecy systems where the meaning of the message is concealed by cipher, code, etc., although its existence is not hidden, and the enemy is assumed to have any special equipment necessary to intercept and record the transmitted signal. We consider only the third type—concealment systems are primarily a psychological problem, and privacy systems a technological one.

Secondly, the treatment is limited to the case of discrete information, where the message to be enciphered consists of a sequence of discrete symbols, each chosen from a finite set. These symbols may be letters in a language, words of a language, amplitude levels of a “quantized” speech or video signal, etc., but the main emphasis and thinking has been concerned with the case of letters.

The paper is divided into three parts. The main results will now be briefly summarized. The first part deals with the basic mathematical structure of secrecy systems. As in communication theory a language is considered to

* The material in this paper appeared originally in a confidential report “A Mathematical Theory of Cryptography” dated Sept. 1, 1945, which has now been declassified.

¹ Shannon, C. E., “A Mathematical Theory of Communication,” *Bell System Technical Journal*, July 1948, p. 379; Oct. 1948, p. 623.

² See, for example, H. F. Gaines, “Elementary Cryptanalysis,” or M. Givierge, “Cours de Cryptographie.”

Perfectly Secure Cipher: One-Time Pad

- It produces random output that bears no statistical relationship to the plaintext
- Because the ciphertext contains no information whatsoever about the plaintext (except the length), there is simply no way to break the code
- For any given ciphertext, all plaintexts are equally possible

Is it practical?

- Need to generate truly random pad values as long as all messages
- Generating truly random keys/numbers/values is hard
- Need to securely distribute the one-time pad values

Modern Symmetric Encryption Algorithms



From Classic to Modern Ciphers

Consider using several ciphers in succession to make harder

- Two substitutions make a more complex substitution
- Two transpositions make more complex transposition
- But a substitution followed by a transposition makes a new much harder cipher

Principle of Modern Ciphers

- Use both substitution and transposition
- Formalized/proposed by C. E. Shannon
 - Using information theory in 1945
 - Product ciphers
 - combines two or more transformations in a manner intending that the resulting cipher is more secure.
- Formulate the principles of “confusion” (standing for substitution) and “diffusion” (standing for transposition)



Confusion and Diffusion

Modern substitution ciphers take in N bits and substitute N bits with M bits using lookup table: called ***S-Boxes***

Modern transposition ciphers take in N bits and permute using lookup table: called ***P-Boxes***

Symmetric Encryption

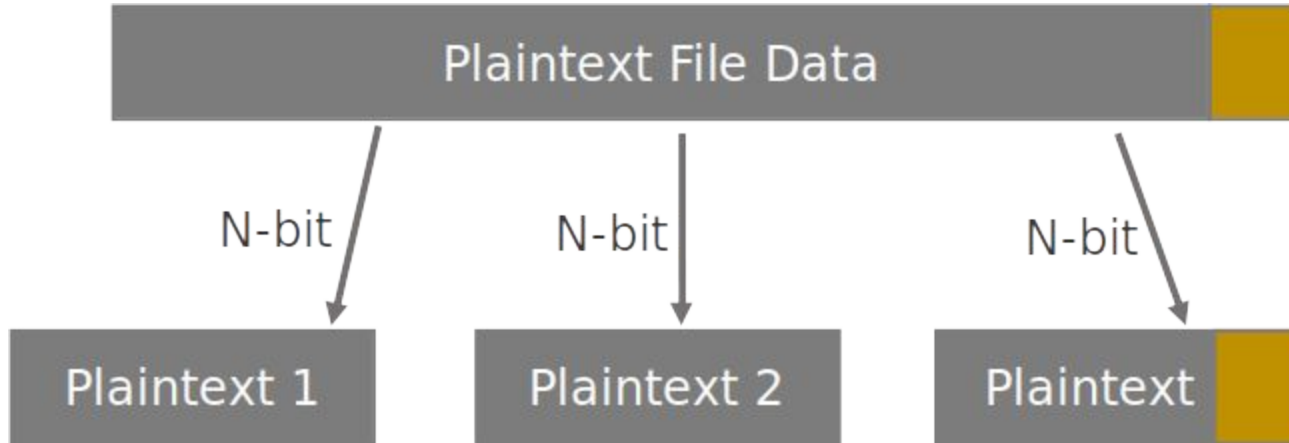
- There are two types of symmetric key algorithms:
 - block ciphers
 - the key has a fixed size
 - prior to encryption, the message is partitioned into blocks
 - each block is encrypted and decrypted separately
 - stream ciphers
 - the message is processed as a stream
 - pseudo-random generator is used to produce a long key stream from a short key

Block Ciphers

- The algorithm maps an *n -bit plaintext block* to an *n -bit ciphertext block*
- Most modern block ciphers are *product ciphers*
 - we sequentially apply more than one operation to the message
- Often a sequence of permutations and substitutions is used
- A common design for an algorithm is to proceed in iterations
 - one iteration is called a *round*
 - each round consists of similar operations
 - *i th* round key k_i is derived from the secret key k using a fixed, public algorithm

Block Ciphers

- Divide input bit stream into n-bit sections



- In a good block cipher, each output bit is a function of all n input bits and all k key bits

Stream Ciphers

- Process message bit by bit (as a stream)
- Have a pseudo random keystream combined (XOR) with plaintext bit by bit
- Randomness of stream key completely destroys statistical properties in message

$$\begin{array}{rcl} & 11001100 & \text{plaintext} \\ \oplus & \underline{01101100} & \text{key stream} \\ & 10100000 & \text{ciphertext} \end{array}$$

Attacks Against Symmetric Encryption

- Encryption and decryption algorithms are assumed to be known to the adversary
- Types of attacks
 - **ciphertext only attack**: adversary knows a number of ciphertexts
 - **known plaintext attack**: adversary knows some pairs of ciphertexts and corresponding plaintexts
 - **chosen plaintext attack**: adversary knows ciphertexts for messages of its choice
 - **chosen ciphertext attack**: adversary knows plaintexts for ciphertexts of its choice
- We want a general-purpose algorithm to sustain all types of attacks

Security Against Chosen-Plaintext Attacks

- In chosen-plaintext attack (CPA), adversary A is allowed to ask for encryptions of messages of its choice – it is active and adaptive
- A is given black-box access to encryption oracle and can query it on different messages
- A is asked to distinguish between encryptions of messages of its choice