Corso di Sicurezza dei Sistemi
Prof. Salvatore D'Antonio

Cryptography

Basic cryptography

- Classical cryptography
- Public key cryptography

Overview

- Security requirements
- Classical cryptography
  - Caesar cipher
  - Vigenere cipher
  - DES
  - 3DES
- Public key cryptography
  - RSA

Security requirements

- Confidentiality
  - Only the owner knows the private key so that enciphered text cannot be read by anyone, except the owner of the private key
- Authentication
  - Only the owner knows the private key. This means that a text enciphered by using that private key has been generated by the owner of the private key.
- Integrity
  - Enciphered words or characters of a text cannot be undetectably changed without knowing private key
- Non-repudiation
  1. A message enciphered by using a private key has been sent by someone who knows that private key

Cryptosystem

- (E, D, M, K, C)
  - M set of plaintexts
  - K set of keys
  - C set of ciphertexts
  - E set of encryption functions $e: M \times K \rightarrow C$
  - D set of decryption functions $d: C \times K \rightarrow C$

Example

- Caesar cipher
  - M = {sequences of characters}
  - K = {i | i is an integer and $0 \le i \le 25$}
  - $E = \{E_k \mid k \in K$ and for any character m $E_k(m) = (m+k)$ *mod* $26\}$
  - $D = \{E_k \mid k \in K$ and for any character c $D_k(c) = (26 + c - k)$ *mod* $26\}$
  - $C = M$

Cyber-attacks

- Adversary is the person/system whose aim is to break the cryptosystem
  - Assume that the adversary knows the used encryption algorithm, but not the key
- Three types of cyber-attacks:
  - Ciphertext only: adversary only has the ciphertext; the objective is to discover the plaintext, and possibly the key
  - Known plaintext: adversary has the ciphertext ad the corresponding plaintext; the objective is to find the key
  - Chosen plaintext: the adversary may provide plaintexts and obtain the corresponding ciphertexts; the objective is to find the key

Cyber-attacks strategies

- Mathematical attacks
  - Based on analysis of underlying mathematics
- Statistical attacks
  - Make assumptions about the distribution of letters, pairs of letters (digrams), triplets of letters (trigrams), etc.
  - Called models of the language
  - Examine ciphertext, correlate properties with the assumptions

More definitions

- unconditional security
  - no matter how much computer power or time is available, the cipher cannot be broken since the ciphertext provides insufficient information to uniquely determine the corresponding plaintext
- computational security
  - given limited computing resources (eg time needed for calculations is greater than age of universe), the cipher cannot be broken

Brute force search

- Simply try any possible key
- Basic attack, its difficulty is proportional to key size

Key Size (bits)	Number of Alternative Keys	Time required at 1 decryption/µs	Time required at 10^6 decryptions/µs
32	$2^{32} = 4.3 \times 10^9$	2^{31} µs = 35.8 minutes	2.15 milliseconds
56 (DES)	$2^{56} = 7.2 \times 10^{16}$	2^{55} µs = 1142 years	10.01 hours
128 (AES)	$2^{128} = 3.4 \times 10^{38}$	2^{127} µs = 5.4×10^{24} years	5.4×10^{18} years
168 (3-DES)	$2^{168} = 3.7 \times 10^{50}$	2^{167} µs = 5.9×10^{36} years	5.9×10^{30} years
26 characters (permutation)	$26! = 4 \times 10^{26}$	2×10^{26} µs = 6.4×10^{12} years	6.4×10^6 years

Classical cryptography

- Sender and receiver share a common key
  - Key could be the same, or it could be trivial to derive one from another
  - Sometimes called symmetric cryptography
- Two basic types plus…
  - Transposition ciphers
  - Substitution ciphers
  - Combinations are called product ciphers

Transposition ciphers

- Rearrange letters in plaintext to produce ciphertext
- Example (Rail-Fence Cipher)
  - Plaintext is HELLO WORLD
  - Rearrange as

    HLOOL

    ELWRD
  - Ciphertext is HLOOL ELWRD

Attacking the cipher

- Anagramming
  - If 1-gram frequencies match English frequencies, but other n-gram frequencies do not, probably we are dealing with transposition
  - Rearrange letters to form n-grams with highest frequencies

Example

- Ciphertext: HLOOLELWRD
- Frequencies of 2-grams beginning with H
  - HE 0.0305
  - HO 0.0043
  - HL, HW, HR, HD < 0.0010
- Frequencies of 2-grams ending in H
  - WH 0.0026
  - EH, LH, OH, RH, DH ≤ 0.0002
- Implies E follows H

Example

- Arrange the ciphertext in order to have H and E adjacent

HE

LL

OW

OR

LD

- Read across and then down to get the plaintext

Row transposition ciphers

- a more complex transposition
- write letters of message out in rows over a specified number of columns
- then reorder the columns according to some key before reading off the rows
- Plaintext: attack postponed to two am

a	t	t	a	c	k	p
o	s	t	p	o	n	e
d	u	n	t	i	l	t
w	o	a	m	x	y	z

- key: 3421567
- ttnaaptmtsuoaodwcoixknlypetz

Substitution ciphers

- Change characters in plaintext to produce ciphertext;
- Note on char codes: A=0, B=1, C=2,....
- Example (Cæsar cipher)
- Plaintext is HELLO WORLD
- Change each letter to the third letter following it (X goes to A, Y to B, Z to C)
- Key is 3
- Ciphertext is KHOOR ZRUOG

Attacking the cipher

- Exhaustive search
  - If the key space is small enough, try all possible keys until you find the right one
  - Cæsar cipher has 26 possible keys
- Statistical analysis
  - Compare to 1-gram model of English

Attack based on statistical analysis

- Compute frequency of each letter in ciphertext:
  - G 0.1 H 0.1 K 0.1 O 0.3 R 0.2 U 0.1 Z 0.1
- Apply 1-gram model of English
- Frequency of characters (1-grams) in English is on next slide

Character frequencies

a	0.080	h	0.060	n	0.070	t	0.090
b	0.015	i	0.065	o	0.080	u	0.030
c	0.030	j	0.005	p	0.020	v	0.010
d	0.040	k	0.005	q	0.002	w	0.015
e	0.130	l	0.035	r	0.065	x	0.005
f	0.020	m	0.030	s	0.060	y	0.020
g	0.015					z	0.002

Statistical analysis

- $f(c)$ frequency of character *c* in ciphertext;
- $p(x)$ is frequency of character *x* in English;
- $\phi(i)$ correlation of frequency of letters in ciphertext with corresponding letters in English, assuming key is *i*
- $\phi(i) = \Sigma_{0 \leq c \leq 25} f(c)p(c-i)$
- $\phi(i) = 0.1p(6-i) + 0.1p(7-i) + 0.1p(10-i) + 0.3p(14-i) + 0.2p(17-i) + 0.1p(20-i) + 0.1p(25-i)$
- We need to maximize the correlation function

Correlation function

i	$\varphi(i)$	i	$\varphi(i)$	i	$\varphi(i)$	i	$\varphi(i)$
0	0.0482	7	0.0442	13	0.0520	19	0.0315
1	0.0364	8	0.0202	14	0.0535	20	0.0302
2	0.0410	9	0.0267	15	0.0226	21	0.0517
3	0.0575	10	0.0635	16	0.0322	22	0.0380
4	0.0252	11	0.0262	17	0.0392	23	0.0370
5	0.0190	12	0.0325	18	0.0299	24	0.0316
6	0.0660					25	0.0430

Results

- Most probable keys, based on ϕ:
  - i= 6, $\phi(i)$ = 0.0660 ; plaintext EBIIL TLOLA
  - i= 10, $\phi(i)$ = 0.0635; plaintext AXEEH PHKEW
  - i= 3, $\phi(i)$ = 0.0575; plaintext HELLO WORLD
  - i= 14, $\phi(i)$ = 0.0535; plaintext WTAAD LDGAS
- The only English phrase is for i= 3
- That's the key (3 or 'D')

Caesar's problem

- Key is too short
  - Can be found by exhaustive search
  - Statistical frequencies not concealed well
  - They look too much like regular English letters
- Make it longer
  - Multiple letters in key
  - Idea is to smooth the statistical frequencies to make cryptanalysis harder

Vigènere cipher

- Like Cæsar cipher, but use a phrase
- Example
  - Message THE BOY HAS THE BALL
  - Key VIG
  - Encipher using Cæsar cipher for each letter:

Key	V	I	G	V	I	G	V	I	G	V	I	G	V	I	G	V
Plaintext	T	H	E	B	O	Y	H	A	S	T	H	E	B	A	L	L
Ciphertext	O	P	K	W	W	E	C	I	Y	O	P	K	W	I	R	G

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

Vigenere cipher - ciphering

- Note that each row of the table corresponds to a Caesar Cipher. The first row is a shift of 0; the second is a shift of 1; and the last is a shift of 25.
- The Vigenere cipher uses this table together with a keyword to encipher a message. For example, suppose we wish to encipher the plaintext message:
- TO BE OR NOT TO BE THAT IS THE QUESTION
- using the keyword RELATIONS
- We begin by writing the keyword, repeated as many times as necessary, above the plaintext message.

Vigenere cipher - ciphering

- To derive the ciphertext using the table, for each letter in the plaintext, one finds the intersection of the row given by the corresponding keyword letter and the column given by the plaintext letter itself to pick out the ciphertext letter.

Keyword: RELAT IONSR ELATI ONSRE LATIO NSREL
Plaintext: TOBEO RNOTT OBETH ATIST HEQUE STION
Ciphertext: KSMEH ZBBLK SMEMP OGAJX SEJCS FLZSY

Vigenere cipher - deciphering

- Deciphering an encrypted message is equally straightforward. One writes the keyword repeatedly above the message.
- This time one uses the keyword letter to pick a column of the table and then traces down the column to the row containing the ciphertext letter. The index of that row is the plaintext letter.

Keyword: RELAT IONSR ELATI ONSRE LATIO NSREL
Ciphertext: KSMEH ZBBLK SMEMP OGAJX SEJCS FLZSY
Plaintext: TOBEO RNOTT OBETH ATIST HEQUE STION

One-time pad

- A Vigenère cipher with a random key having at least the same length as the message
- Provably unbreakable
- Why? Look at ciphertext DXQR.
  - It Equally likely corresponds to plaintext DOIT(key AJIY) and to plaintext DONT(key AJDY) and any other 4 letters
  - Warning: keys must be random, or you can attack the cipher by trying to regenerate the key
  - Approximations, such as using pseudorandom number generators to generate keys, are not random

Product ciphers

- ciphers using substitutions or transpositions are not secure because of language characteristics
- hence consider a sequence of several ciphers to make cryptanalysis harder, however:
- two substitutions generate a more complex substitution
- two transpositions generate a more complex transposition
- substitution followed by a transposition makes a new much harder cipher
- this is bridge from classical to modern ciphers

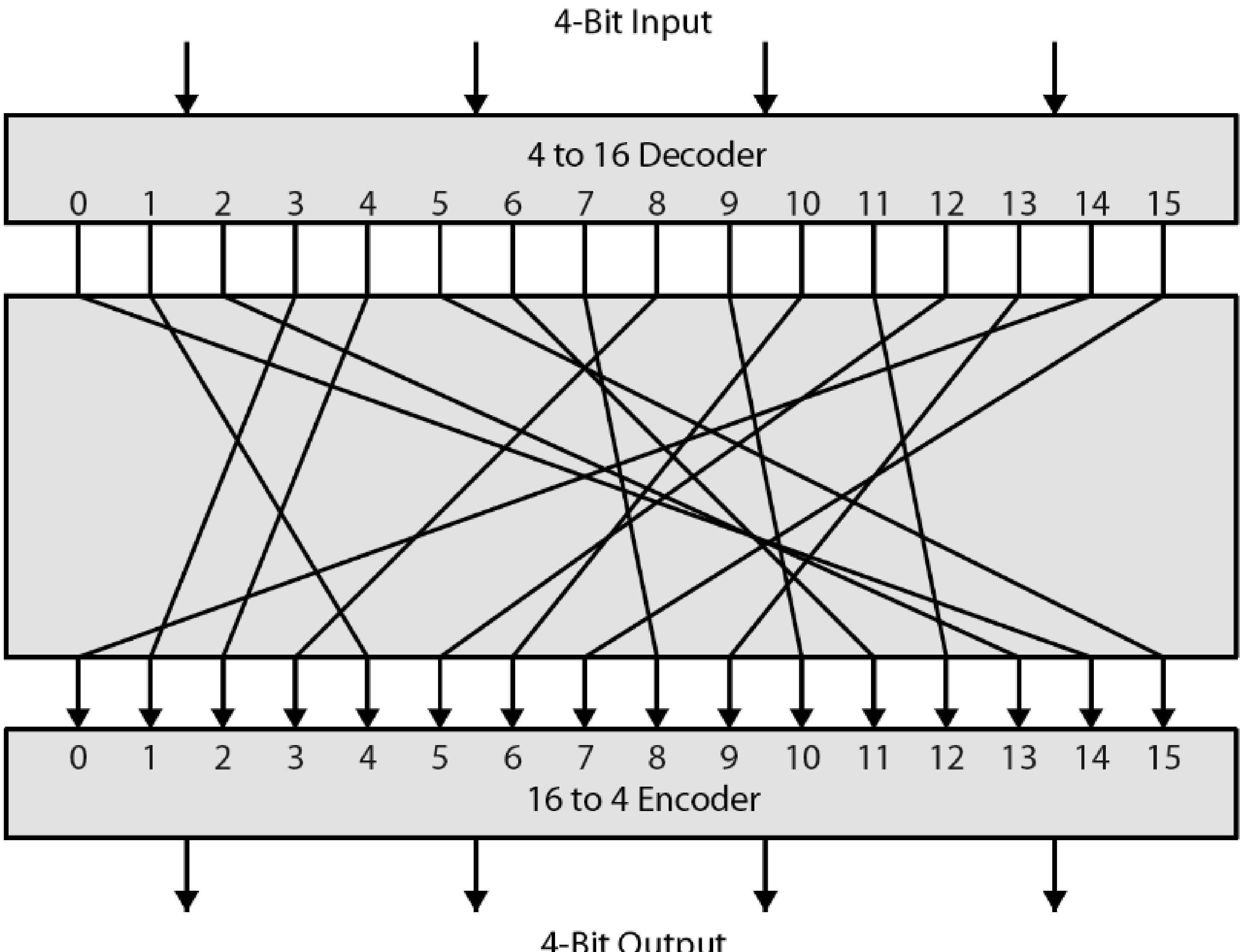
Block vs stream ciphers

- block ciphers process messages in blocks, each of them is encrypted/decrypted
  - like a substitution on very long characters (64-bits or more)
- stream ciphers process messages a bit or byte at a time when encrypting/decrypting
- many current ciphers are block ciphers
- broader range of applications

Block cipher principles

- most symmetric block ciphers are based on a Feistel Cipher Structure
- Efficient decryption to recover messages
- block ciphers look like an extremely large substitution
- It would need table of 264 entries for a 64-bit block
- Better option is to start from smaller building blocks by exploiting the idea of a product cipher

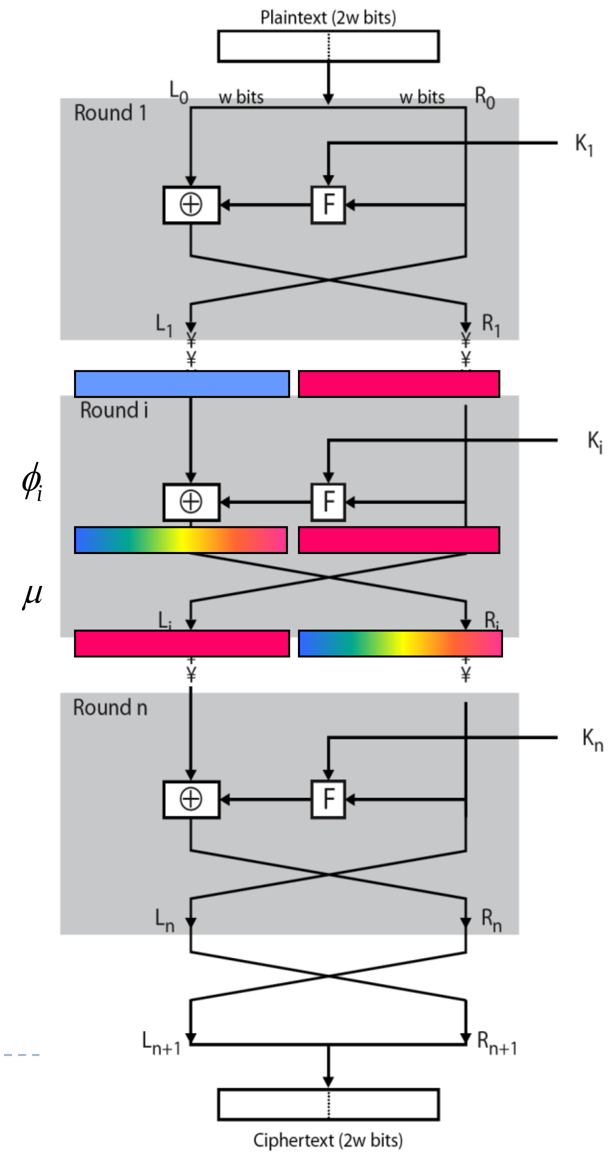
Ideal block cipher



Feistel cipher structure

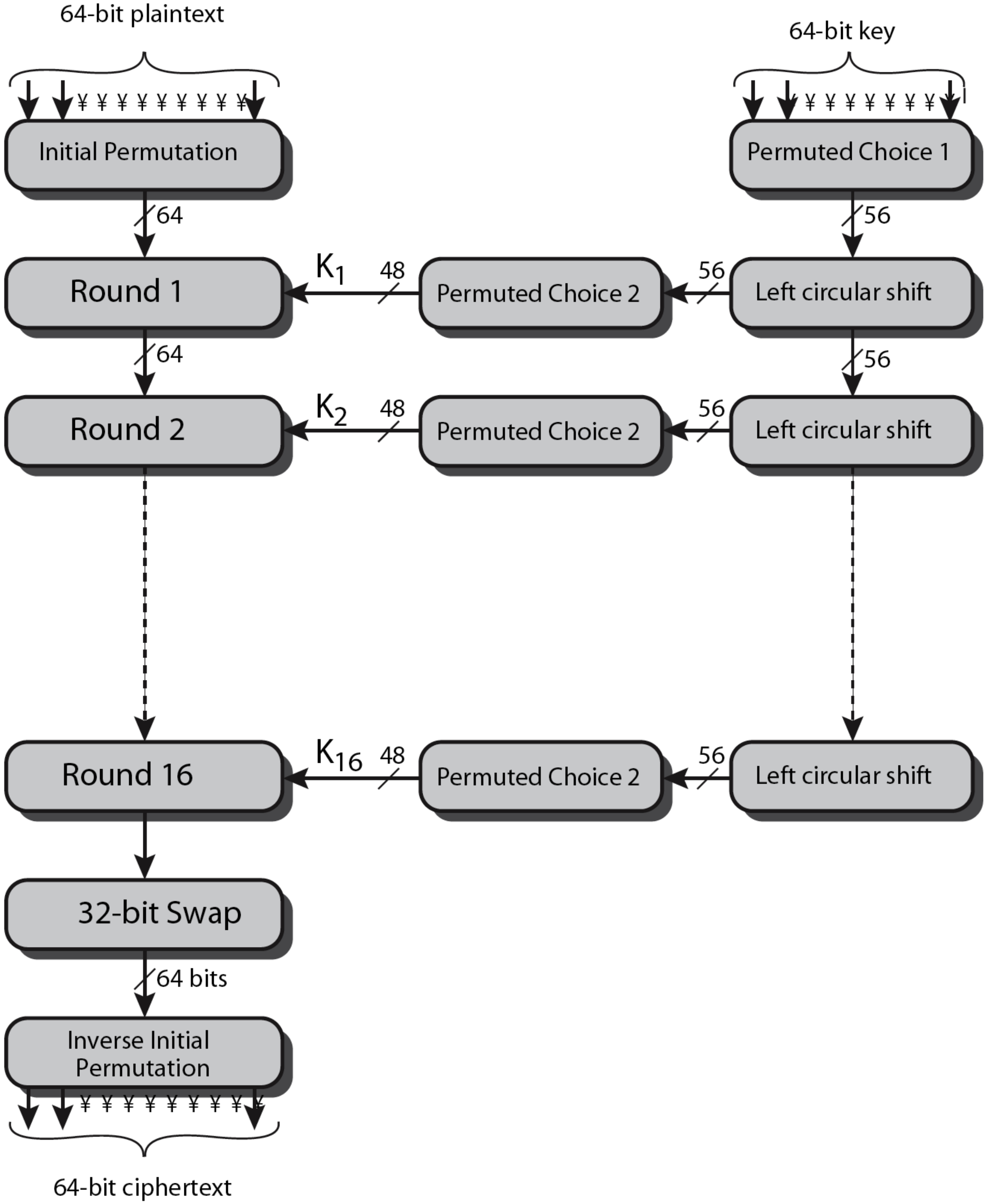
- Horst Feistel devised the feistel cipher
  - based on concept of invertible product cipher
  - partitions input block into two halves
  - processes through multiple rounds
  - performs a substitution on left half
  - based on round function of right half & subkey
  - permutation of the halves
  - implements Shannon's S-P net concept

The Feistel cipher structure



Overview of Data Encryption Standard (DES)

- A block cipher:
  - encrypts blocks of 64 bits using a 64 bit key
  - outputs 64 bits of ciphertext
- A product cipher
  - basic unit is the bit
  - performs both substitution and transposition (permutation) on the bits
  - Cipher consists of 16 rounds (iterations) each with a round key generated from the user-supplied key


64-bit plaintext
64-bit key
Initial Permutation
Permuted Choice 1
64
56
Round 1
K_1
48
Permuted Choice 2
56
Left circular shift
64
56
Round 2
K_2
48
Permuted Choice 2
56
Left circular shift
Round 16
K_{16}
48
Permuted Choice 2
56
Left circular shift
32-bit Swap
64 bits
Inverse Initial Permutation
64-bit ciphertext

Permuted Choice One (PC1)

left	57	49	41	33	25	17	9
	1	58	50	42	34	26	18
	10	2	59	51	43	35	27
	19	11	3	60	52	44	36
right	63	55	47	39	31	23	15
	7	62	54	46	38	30	22
	14	6	61	53	45	37	29
	21	13	5	28	20	12	4

The "Left" and "Right" halves of the table show which bits from the input key form the left and right sections of the key schedule state. Note that only 56 bits of the 64 bits of the input are selected; the remaining eight (8, 16, 24, 32, 40, 48, 56, 64) were specified for use as parity bits.

Permuted Choice Two (PC2)

14	17	11	24	1	5	3	28
15	6	21	10	23	19	12	4
26	8	16	7	27	20	13	2
41	52	31	37	47	55	30	40
51	45	33	48	44	49	39	56
34	53	46	42	50	36	29	32

This permutation selects the 48-bit subkey for each round from the 56-bit key-schedule state.

Initial Permutation

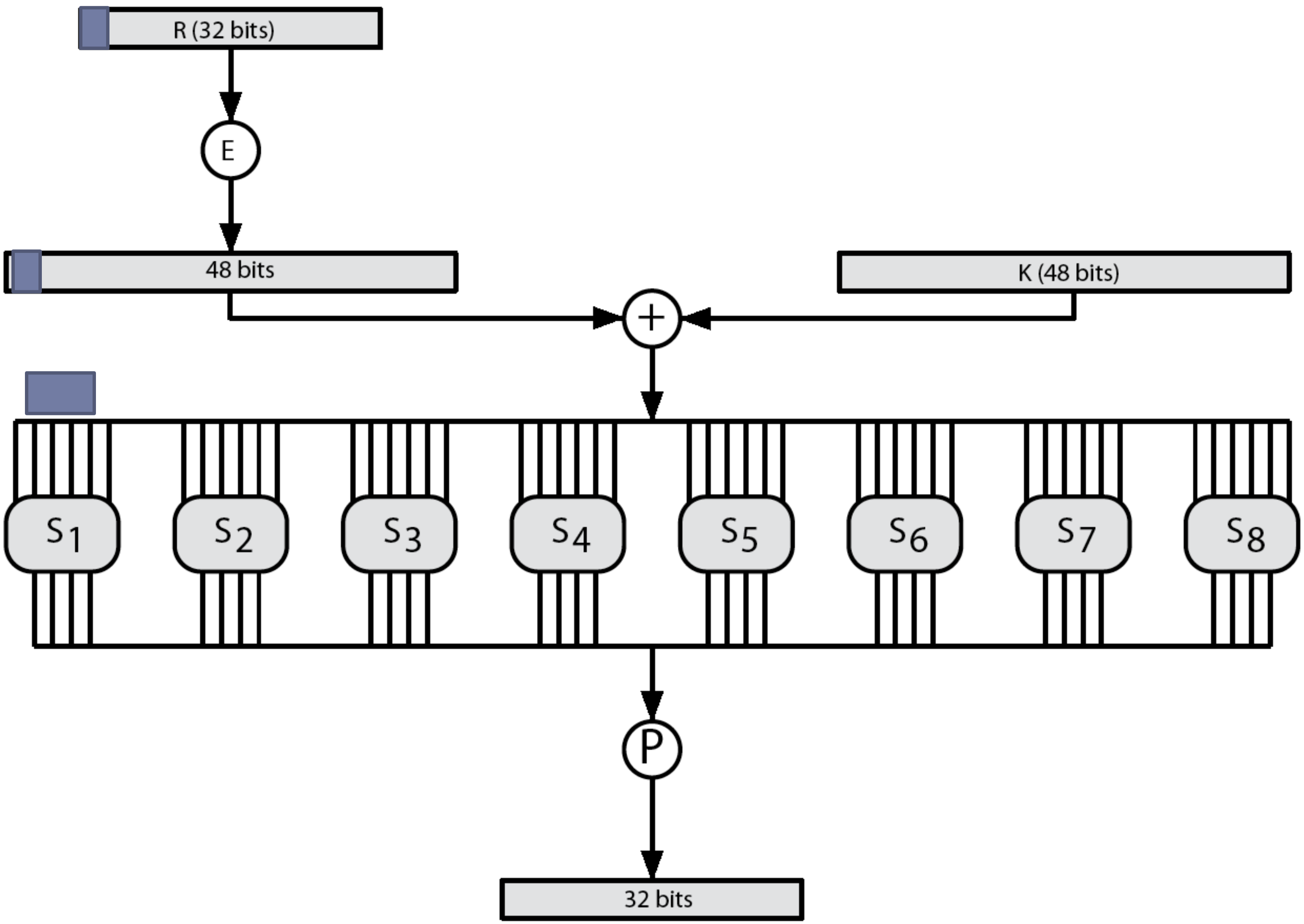
58	50	42	34	26	18	10	2
60	52	44	36	28	20	12	4
62	54	46	38	30	22	14	6
64	56	48	40	32	24	16	8
57	49	41	33	25	17	9	1
59	51	43	35	27	19	11	3
61	53	45	37	29	21	13	5
63	55	47	39	31	23	15	7

This table specifies the input permutation on a 64-bit block. The meaning is as follows: the first bit of the output is taken from the 58th bit of the input; the second bit from the 50th bit, and so on, with the last bit of the output taken from the 7th bit of the input.
This information is presented as a table for ease of presentation; it is a vector, not a matrix.

Final Permutation (IP^{-1})

40	8	48	16	56	24	64	32
39	7	47	15	55	23	63	31
38	6	46	14	54	22	62	30
37	5	45	13	53	21	61	29
36	4	44	12	52	20	60	28
35	3	43	11	51	19	59	27
34	2	42	10	50	18	58	26
33	1	41	9	49	17	57	25

The F function of DES



The Expansion Permutation E

32	1	2	3	4	5
4	5	6	7	8	9
8	9	10	11	12	13
12	13	14	15	16	17
16	17	18	19	20	21
20	21	22	23	24	25
24	25	26	27	28	29
28	29	30	31	32	1

The S-Boxes

- Eight S-boxes each map 6 to 4 bits
- Each S-box is specified as a 4 x 16 table
  - each row is a permutation of 0-15
  - outer bits 1 & 6 of input are used to select one of the four rows
  - inner 4 bits of input are used to select a column
- All the eight boxes are different.

Box S_1

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	14	4	13	1	2	15	11	8	3	10	6	12	5	9	0	7
1	0	15	7	4	14	2	13	1	10	6	12	11	6	5	3	8
2	4	1	14	8	13	6	2	11	15	12	9	7	3	10	5	0
3	15	12	8	2	4	9	1	7	5	11	3	14	10	0	6	13

- For example, $S_1(101010) = 6 = 0110$.

Permutation Function P

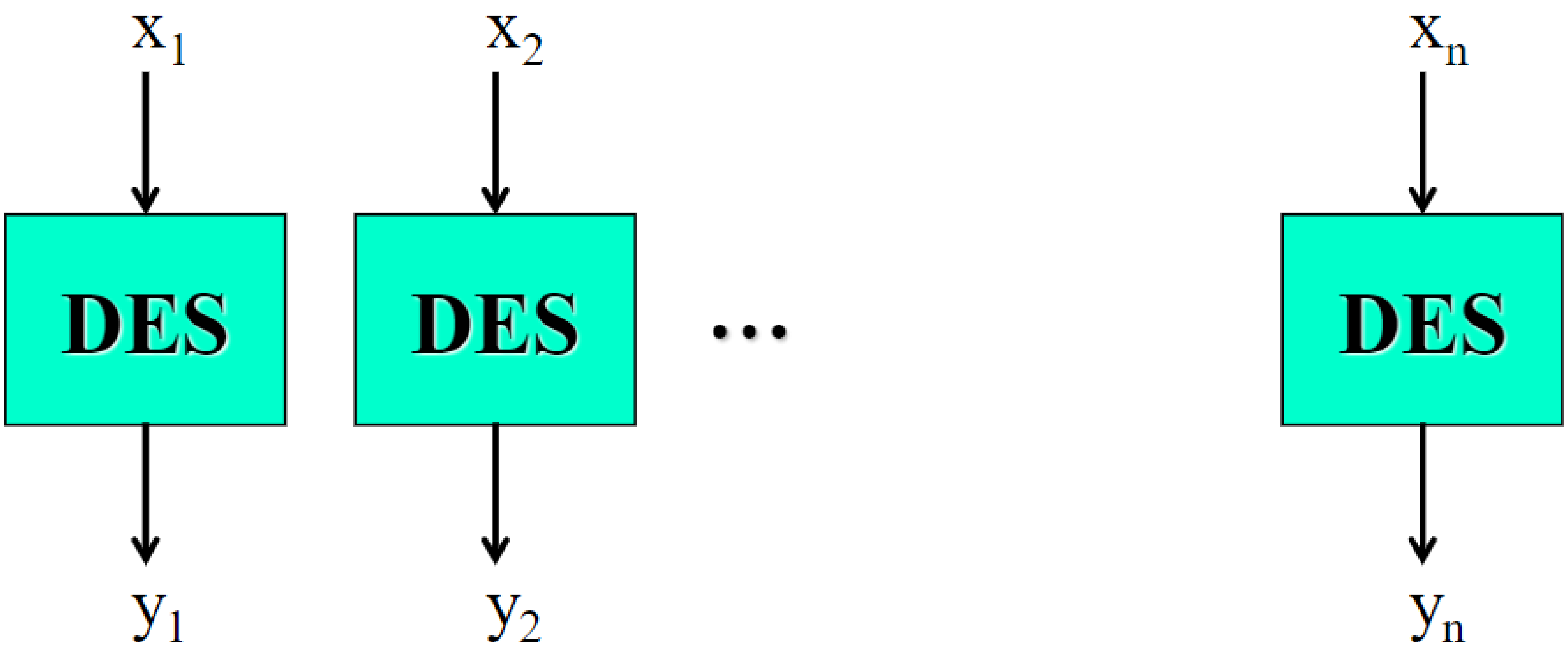
P			
16	7	20	21
29	12	28	17
1	15	23	26
5	18	31	10
2	8	24	14
32	27	3	9
19	13	30	6
22	11	4	25

DES modes

- How to cipher text longer than 64 bit?
  - Electronic codebook chaining (ECB)
  - Cipher block chaining (CBC)
  - Cipher feedback (CFB)
  - Output feedback (OFB)
  - Counter (CTR) mode

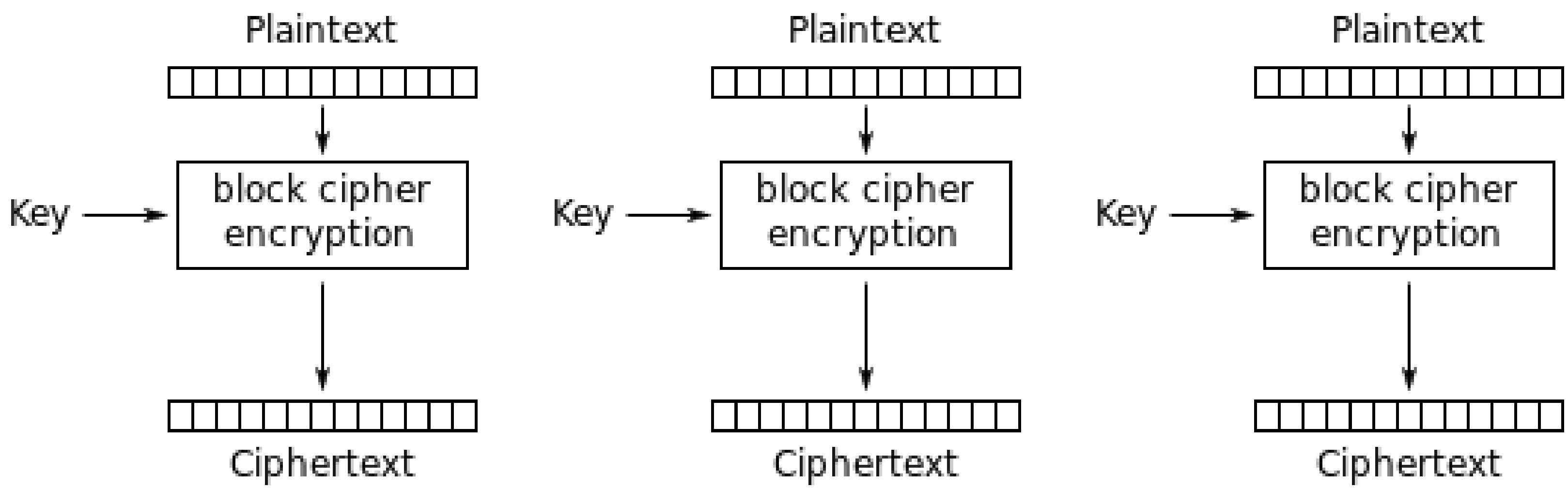
Electronic codebook chaining (ECB)

plaintext $x = x_1x_2 \ldots x_n$ (in n blocks of 64 bit)



Ciphered text $y = y_1y_2 \ldots y_n$

Electronic Codebook chaining



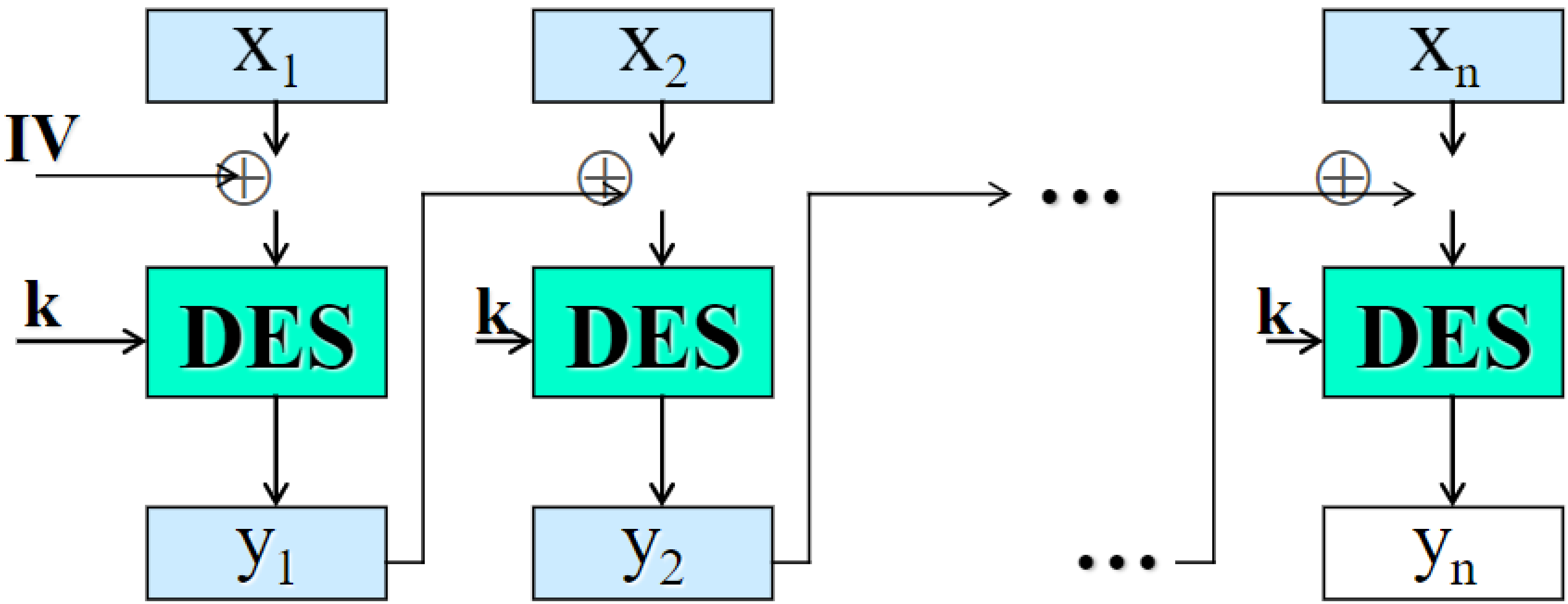
Electronic Codebook (ECB) mode encryption

Electronic codebook chaining

- Pros
  - ECB is fast
  - Errors do not propagate
- Cons
  - Blocks are independent
  - Substitution attacks are possible

Cipher Block Chaining (CBC)

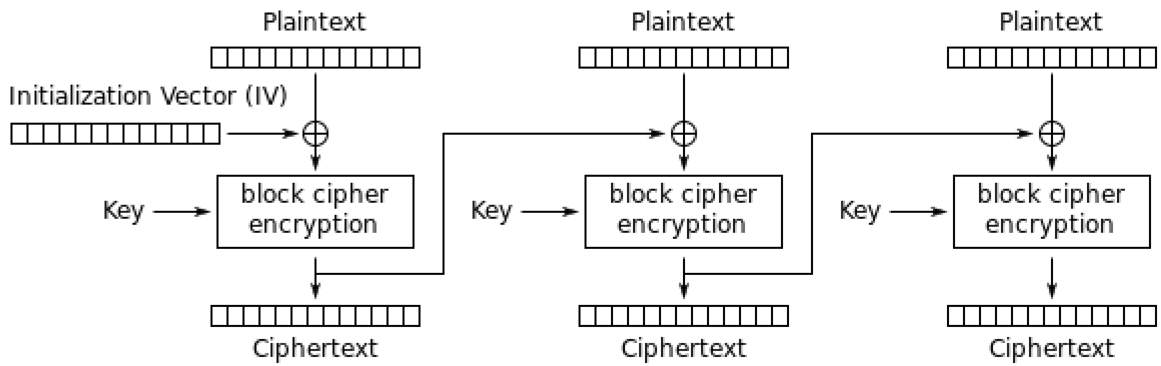
- Plaintext $X=X_1X_2X_3\ldots X_n$ (in n blocks of 64 bits)
- Ciphertext=$Y_1Y_2Y_3\ldots Y_n$
- IV = initialization vector



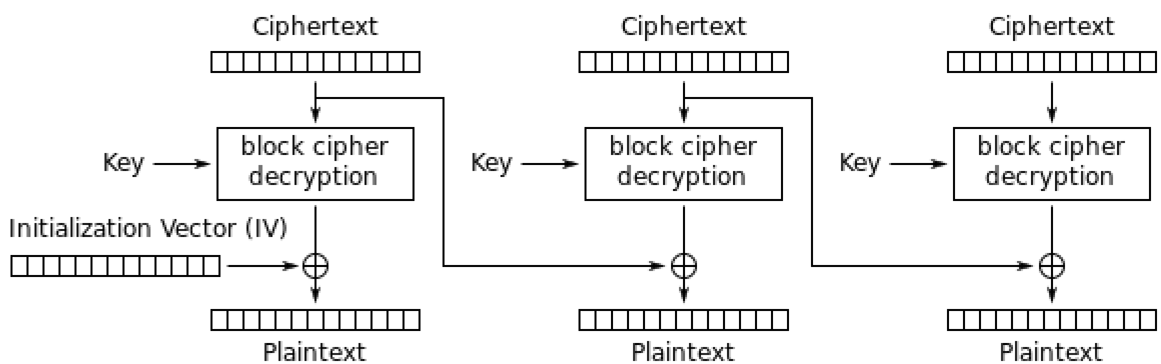
Initialization Vector

- An initialization vector (IV) is a block of bits that is used to randomize the encryption and hence to produce distinct ciphertexts even if the same plaintext is encrypted multiple times (without the need for a slower re-keying process).
- IV needs to be secret (new attacks!!)

Cipher Block Chaining mode



Cipher Block Chaining (CBC) mode encryption

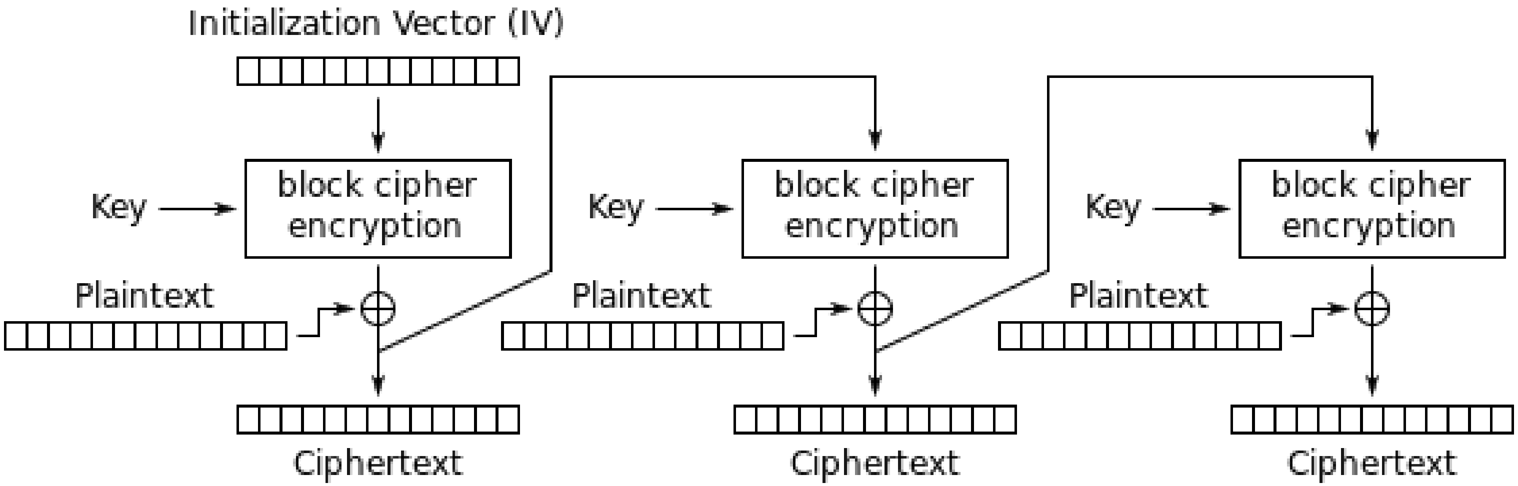


Cipher Block Chaining (CBC) mode decryption

Cipher Block Chaining (CBC)

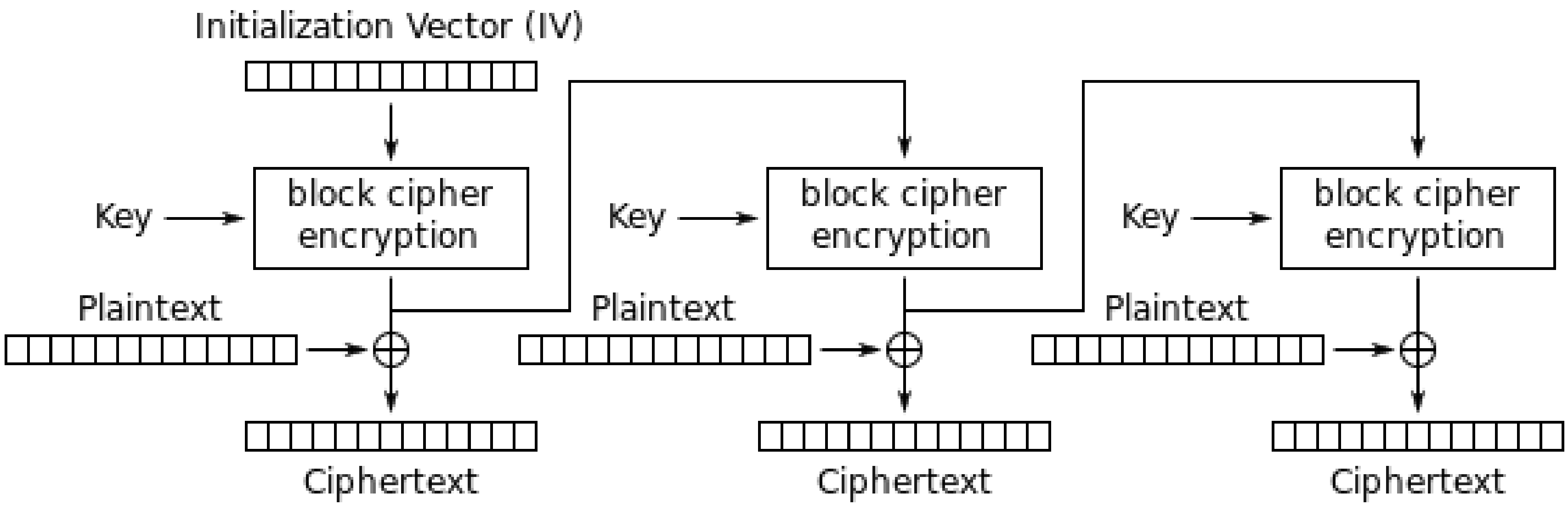
- Pros
  - Blocks depend on each other
    - No substitution attacks
- Cons
  - Slower than ECB
  - Errors propagate

Cipher Feedback mode



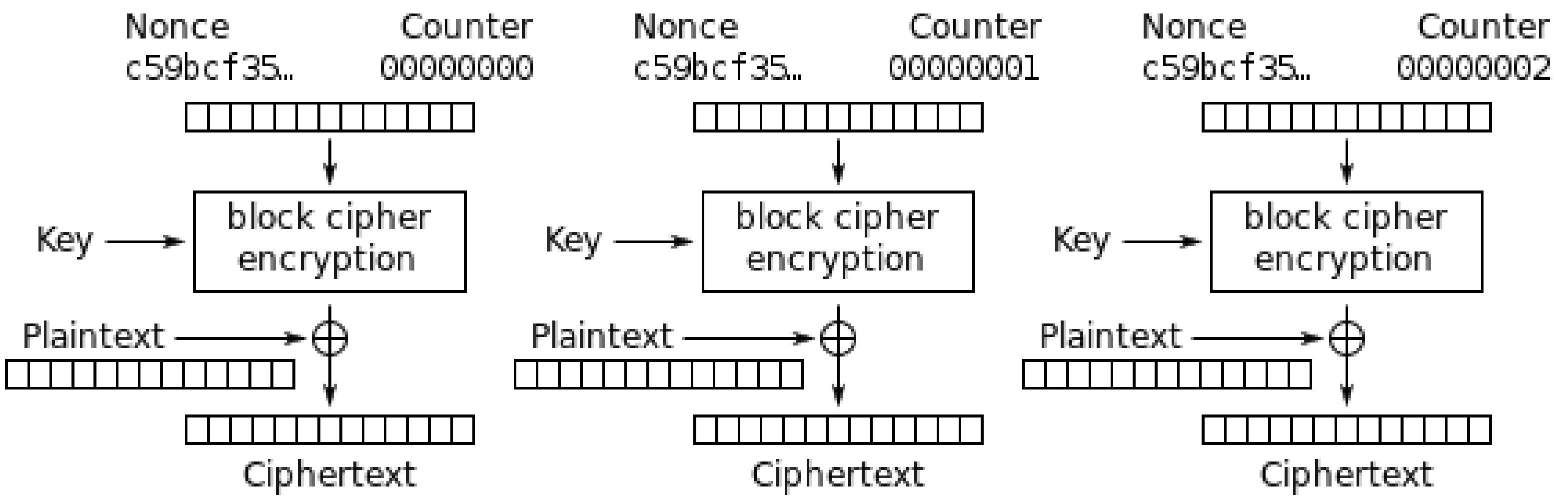
Cipher Feedback (CFB) mode encryption

Output Feedback mode



Output Feedback (OFB) mode encryption

Counter mode



Counter (CTR) mode encryption

Padding

- A block cipher works on units of a fixed size (known as a block size), but messages come in a variety of lengths. So some modes (namely ECB and CBC) require that the final block be padded before encryption. Several padding schemes exist.
- The simplest is to add null bytes to the plaintext to bring its length up to a multiple of the block size.