CRYPTOGRAPHY

Prof. Dr. Sıddıka Berna Örs Yalçın

Istanbul Technical University, Department of Electronics and Communication Engineering

March 31, 2023

Contact Info

Room Number: 2318
email: siddika.ors@itu.edu.tr
web page: http://web.itu.edu.tr/~orssi/

References

1. Douglas R. Stinson, Cryptography Theory and Practice, Third Edition, CRC Press, November 2005.
2. Alfred J. Menezes, Paul C. van Oorschot and Scott A. Vanstone, Handbook of Applied Cryptography, CRC Press, ISBN: 0-8493-8523-7, October 1996, 816 pages.
3. Simon Singh, Code Book : The Science of Secrecy from Ancient Egypt to Quantum Cryptography, Westminster, MD, USA: Anchor, 2000. p xv.

Content

1. Classical cryptography: introduction: some simple cryptosystems
2. Cryptanalysis of simple cryptosystems
3. Shannon's theory: probability theory, entropy, properties of entropy
4. Product cryptosystems
5. Block ciphers: substiturion-permutation network
6. Linear cryptanalysis
7. Differential cryptanalysis
8. The data encryption standard (DES)
9. Advanced encryption standard (AES), modes of operation
10. Hash functions: collision-free hash functions, authentication codes
11. The RSA system and factoring: introduction to public-key cryptography
12. Public-key cryptosystems based on discrete logarithm problem: the ElGamal cryptosystem
13. Finite field and elliptic curve systems
14. Signature schemes: introduction, the ElGamal signature scheme
15. The digital signature algorithm (DSA), the elliptic curve digital signature algorithm (ECDSA)

Grading

1st Homework	3rd week	15 %
2nd Homework	7th week	15 %
Midterm	8th week	15 %
3rd Homework	11th week	15 %
Final		40 %

History of Cryptography

hieroglyphs
around 2000 B.C.

漢汉
字字

ideogram
ancient Chinese

Clay tablets
from Mesopotamia

History of Cryptography

hieroglyphs
around 2000 B.C.

漢 汉
字 字

ideogram
ancient Chinese

Clay tablets
from Mesopotamia

ABCDEFGHIJKLMNOPQRSTUVWXYZ
ZYXWVUTSRQPONMLKJIHGFEDCBA

Atbash cipher - around 500 to 600 BC

History of Cryptography

hieroglyphs
around 2000 B.C.

漢字 汉字

ideogram
ancient Chinese

Clay tablets
from Mesopotamia

ABCDEFGHIJKLMNOPQRSTUVWXYZ
ZYXWVUTSRQPONMLKJIHGFEDCBA

Atbash cipher - around 500 to 600 BC

Scytale - Spartan

	1	2	3	4	5
1	A	B	C	D	E
2	F	G	H	I	J
3	K	L	M	N	O
4	P	Q	R	S	T
5	U	V	W	X	Y/Z

Polybius Square - Greek method

Steganography

Secret communication achieved by hiding the existence of a message is known as steganography, derived from the Greek words steganos, meaning "covered" and graphein, meaning "to write".

- physically concealed beneath wax on wooden tablets
- a tattoo on a slave's head concealed by regrown hair

Cryptography

- Cryptography was derived from the Greek word kryptos, meaning "hidden".
- The aim of cryptography is not to hide the existence of a message, but rather to hide its meaning, a process known as encryption.
- To render a message unintelligible, it is scrambled according to a particular protocol which is agreed beforehand between the sender and the intended recipient.
- The advantage of cryptography is that if the enemy intercepts an encrypted message, then the message is unreadable.
- Without knowing the scrambling protocol, the enemy should find it difficult, if not impossible, to recreate the original message from the encrypted text.

Singh, Simon. Code Book : The Science of Secrecy from Ancient Egypt to Quantum Cryptography. Westminster, MD, USA: Anchor, 2000. p 6. `http://site.ebrary.com/lib/istanbulteknik/Doc?id=10235313&ppg=28`

Usage of Cryptography (1/2)

- Past
  - Military, Diplomatic Service, Government
  - was used as a tool to protect national secrets and strategies
- Now
  - Private sector
  - is used to protect information in digital form and to provide security services

Usage of Cryptography (2/2)

Although cryptography is now having a major impact on civilian activities, it should be noted that military cryptography remains an important subject.

It has been said that

- the First World War was the chemists' war, because mustard gas and chlorine were employed for the first time
- the Second World War was the physicists' war, because the atom bomb was detonated
- the Third World War would be the mathematicians' war, because mathematicians will have control over the next great weapon of war - information. Mathematicians have been responsible for developing the codes that are currently used to protect military information. Not surprisingly, mathematicians are also at the forefront of the battle to break these codes.

Singh, Simon. Code Book : The Science of Secrecy from Ancient Egypt to Quantum Cryptography. Westminster, MD, USA: Anchor, 2000. p xv.

`http://site.ebrary.com/lib/istanbulteknik/Doc?id=10235313&ppg=19`

CRYPTOGRAPHY

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

Basic Terminology: Domains

The cryptosystem is a five- tuple $\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D}$

- $\mathcal{P}$: a set called the plaintext space.
- $\mathcal{C}$: a set called the ciphertext space.
- $\mathcal{K}$: a set called the key space.
- For each $K \in \mathcal{K}$, there is an encryption rule $e_K \in \mathcal{E}$ and a corresponding decryption rule $d_K \in \mathcal{D}$. Each $e_K : \mathcal{P} \to \mathcal{C}$ and $d_K : \mathcal{C} \to \mathcal{P}$ are functions such that $d_K(e_K(x)) = x$ for every element $x \in \mathcal{P}$.

If a cryptosystem is to be of practical use:

1. Each e_K and d_K should be efficiently computable.
2. An opponent, upon seeing a ciphertext string y should be unable to determine the key K or the plaintext string x.

Shift Cipher

- Let $\mathcal{P} = \mathcal{C} = \mathcal{K} = \mathbb{Z}_{26}$.
- For $0 \leq K \leq 25$, define $e_K(x) = (x + K) \bmod 26$ and $d_K(y) = (y - K) \bmod 26$
- Since there are only 26 possible keys, it is easy to try every possible K until a meaningful plaintext is obtained.
- $K = 3 \implies$ is called Ceaser cipher (˜55 BC)

Caesar Cipher

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

Example 1 :

$x =$ T H I S C I P H E R I S C E R T A I N L Y N O T S E C U R E

$y = e_K(x) =$ W K L V F L S K H U L V F H U W D L Q O B Q R W V H F X U H

Cryptanalysis

The practice of changing ciphertext into plaintext without complete knowledge of the cipher.

Cryptanalysis

The practice of changing ciphertext into plaintext without complete knowledge of the cipher.

- First method : Frequency analysis
- Although it is not known who first realized that the variation in the frequencies of letters could be exploited in order to break ciphers, the earliest known description of the technique is by the ninth-century scientist Abu Yusuf Ya'qub ibn Is-haq ibn as-Sabbah ibn'omran ibn Ismail al-Kindi. Known as "the philosopher of the Arabs" al-Kindi was the author of 290 books on medicine, astronomy, mathematics, linguistics and music.
- His greatest treatise, which was rediscovered only in 1987 in the Sulaimaniyyah Ottoman Archive in Istanbul, is entitled "A Manuscript on Deciphering Cryptographic Messages".

Singh, Simon. Code Book : The Science of Secrecy from Ancient Egypt to Quantum Cryptography. Westminster, MD, USA: Anchor, 2000. p 17.
http://site.ebrary.com/lib/istanbulteknik/Doc?id=10235313&ppg=39

Attacks

Kerckhoffs' Principle: The security of a cryptosystem must not depend on keeping secret the crypto-algorithm. The security depends only on keeping secret the key.[1]

[1] It was definitively stated in 1883 by the Dutch linguist Auguste Kerckhoffs von Nieuwenhof in his book La Cryptographie militaire

Attacks

Kerckhoffs' Principle: The security of a cryptosystem must not depend on keeping secret the crypto-algorithm. The security depends only on keeping secret the key.[1]

The aim of the attacker is to read the encrypted messages, which in many cases is achieved by finding the secret key of the system.

[1] It was definitively stated in 1883 by the Dutch linguist Auguste Kerckhoffs von Nieuwenhof in his book La Cryptographie militaire

Attacks

Kerckhoffs' Principle: The security of a cryptosystem must not depend on keeping secret the crypto-algorithm. The security depends only on keeping secret the key.[1]

The aim of the attacker is to read the encrypted messages, which in many cases is achieved by finding the secret key of the system.

The efficiency of the attack is measured by

- the amount of plaintext- ciphertext pairs required,
- time spent for their analysis
- the success probability of the attack

[1] It was definitively stated in 1883 by the Dutch linguist Auguste Kerckhoffs von Nieuwenhof in his book La Cryptographie militaire

Types of Attacks

- Ciphertext- Only
- Known Plaintext
- Chosen Plaintext
- Chosen Ciphertext
- Adaptive Chosen Plaintext or Ciphertext
- Related Key
- Partial Knowledge of the Key

The Goals of Cryptanalytic Attacks

- Distinguishing Attacks
- Partial Knowledge of the Plaintext
- Decryption
- Encryption (Forgery)
- Partial Key Recovery
- Total Key Recovery

Cryptology

- Cryptanalysis: the study of mathematical techniques to break the system
- Cryptology: cryptography + cryptanalysis
- Cryptosystem: a set of cryptographic primitives, symmetric key and public key

Cryptanalysis of Example 1

W K L V F L S K H U L V F H U W D L Q O B Q R W V H F X U H

V J K U E K Q J G T K U E G T V C K P N A P Q V U G E W T G
U I J T D J P I F S J T D F S U B J O M Z O P U T F D V S F
T H I S C I P H E R I S C E R T A I N L Y N O T S E C U R E
$K = 3$

Mono- alphabetic Substitution Cipher

- Let $\mathcal{P} = \mathcal{C} = \mathbb{Z}_{26}$. $\mathcal{K}$ consists of all possible permutations of the 26 symbols. For each permutation $\pi \in \mathcal{K}$, define $e_\pi(x) = \pi(x)$ and $d_\pi(y) = \pi^{-1}(y)$ where π^{-1} is the inverse permutation to π.
- If the alphabet is the English alphabet, then the size of the key space is $26! \approx 400,000,000,000,000,000,000,000,000$
- The distribution of letter frequencies is preserved in the ciphertext.

Example

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

Example

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

$$x = \quad \text{T H I S C I P H E R I S C E R T A I N L Y N O T S E C U R E}$$
$$y = e_\pi(x) = \text{O P R M F R G P J K R M F J K O B R C X A C E O M J F Q K J}$$

Cryptanalysis of Mono- alphabetic Substitution Cipher (1/8)

Ciphertext:

VGQBQHWHUXYQVULRZUGVWBUYVHKYZTHXQBNOZY
BVYVURVEQBOYQHYVTMXTZRQHULVULYQBZOWBOZ
GYQBKBYYBOTZGYQBHKEQHMBYYQHYUZYHNZOWRZ
DKWMBPHWBZDYVGHUXZUBNVTQBTYZWBRVEQBOYQ
BTBHUWLBYHYYQBVOPBHUVULQBTQZDKWTDMTYVY
DYBYQBGZDOYQKBYYBOZGYQBHKEQHMBYUHPBKXG
ZOHUWTZYQBZYQBOT

Cryptanalysis of Mono- alphabetic Substitution Cipher (1/8)

Ciphertext:

V G Q B Q H W H U X Y Q V U L R Z U G V W B U Y V H K Y Z T H X Q B N O Z Y
B V Y V U R V E Q B O Y Q H Y V T M X T Z R Q H U L V U L Y Q B Z O W B O Z
G Y Q B K B Y Y B O T Z G Y Q B H K E Q H M B Y Y Q H Y U Z Y H N Z O W R Z
D K W M B P H W B Z D Y V G H U X Z U B N V T Q B T Y Z W B R V E Q B O Y Q
B T B H U W L B Y H Y Y Q B V O P B H U V U L Q B T Q Z D K W T D M T Y V Y
D Y B Y Q B G Z D O Y Q K B Y Y B O Z G Y Q B H K E Q H M B Y U H P B K X G
Z O H U W T Z Y Q B Z Y Q B O T

Letter Frequency in the English Language

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

Cryptanalysis of Mono- alphabetic Substitution Cipher (1/8)

Ciphertext:

VGQBQHWHUXYQVULRZUGVWBUYVHKYZTHXQBNOZY
BVYVURVEQBOYQHYVTMXTZRQHULVULYQBZOWBOZ
GYQBKBYYBOTZGYQBHKEQHMBYYQHYUZYHNZOWRZ
DKWMBPHWBZDYVGHUXZUBNVTQBTYZWBRVEQBOYQ
BTBHUWLBYHYYQBVOPBHUVULQBTQZDKWTDMTYVY
DYBYQBGZDOYQKBYYBOZGYQBHKEQHMBYUHPBKXG
ZOHUWTZYQBZYQBOT

Letter Frequency in the English Language

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

Letter Frequency in the ciphertext

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
0	33	0	6	4	0	8	19	0	0	8	5	5	3	12	2	24	5	0	12	15	15	10	5	33	19

Cryptanalysis of Mono- alphabetic Substitution Cipher (1/8)

Ciphertext:

VGQBQHWHUXYQVULRZUGVWBUYVHKYZTHXQBNOZY
BVYVURVEQBOYQHYVTMXTZRQHULVULYQBZOWBOZ
GYQBKBYYBOTZGYQBHKEQHMBYYQHYUZYHNZOWRZ
DKWMBPHWBZDYVGHUXZUBNVTQBTYZWBRVEQBOYQ
BTBHUWLBYHYYQBVOPBHUVULQBTQZDKWTDMTYVY
DYBYQBGZDOYQKBYYBOZGYQBHKEQHMBYUHPBKXG
ZOHUWTZYQBZYQBOT

Letter Frequency in the English Language

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

Letter Frequency in the ciphertext

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
0	33	0	6	4	0	8	19	0	0	8	5	5	3	12	2	24	5	0	12	15	15	10	5	33	19

$e_\pi(\mathrm{E}) = \mathrm{B}$ or Y and $e_\pi(\mathrm{T}) = \mathrm{B}$ or Y

Cryptanalysis of Mono- alphabetic Substitution Cipher (2/8)

Digraphs in the ciphertext with B

--QBQ---------------WBU--------QBN--Y
BV-------QBO-----------------QBZ-WBO-
--QBKBYYBO----QBH----MBY---------------
---MBP-WBZ----------UBN--QBT--WBR--QBO-Q
BTBH--LBY---QBV-PBH----QBT--------------
-YBYQBG-----KBYYBO---QBH---MBY--PBK--
---------QBZ-QBO-

Cryptanalysis of Mono- alphabetic Substitution Cipher (2/8)

Digraphs in the ciphertext with B

--QBQ--------------WBU---------QBN--Y
BV------QBO-----------------QBZ-WBO-
--QBKBYYBO----QBH----MBY--------------
---MBP-WBZ---------UBN--QBT--WBR--QBO-Q
BTBH--LBY---QBV-PBH----QBT-------------
-YBYQBG-----KBYYBO---QBH---MBY--PBK--
--------QBZ-QBO-

The Digraph Frequencies in the English Language

th he an in er on re ed nd ha at en es of nt ea ti to io le is ou ar as de rt ve

Digraph Frequency in the ciphertext with B

QB	BQ	WB	BU	BN	YB	BV	BO	BZ	BK	KB	BY	BH	MB	BP	UB	BT	BR	TB	LB	PB	BG
15	1	4	1	2	4	2	6	3	2	2	6	4	3	1	1	3	1	1	1	2	1

Cryptanalysis of Mono- alphabetic Substitution Cipher (2/8)

Digraphs in the ciphertext with B

--QBQ-----------------WBU---------QBN--Y
BV-------QBO------------------------QBZ-WBO-
--QBKBYYBO----QBH-----MBY-------------------
---MBP-WBZ--------UBN--QBT--WBR--QBO-Q
BTBH--LBY---QBV-PBH-----QBT----------------
-YBYQBG------KBYYBO----QBH----MBY--PBK--
---------QBZ-QBO-

The Digraph Frequencies in the English Language

th he an in er on re ed nd ha at en es of nt ea ti to io le is ou ar as de rt ve

Digraph Frequency in the ciphertext with B

QB	BQ	WB	BU	BN	YB	BV	BO	BZ	BK	KB	BY	BH	MB	BP	UB	BT	BR	TB	LB	PB	BG
15	1	4	1	2	4	2	6	3	2	2	6	4	3	1	1	3	1	1	1	2	1

$$e_\pi(\mathrm{HE}) = \mathrm{QB} \Rightarrow e_\pi(\mathrm{H}) = \mathrm{Q}$$
$$e_\pi(\mathrm{ER}) = \mathrm{BO} \text{ or } \mathrm{BY} \Rightarrow e_\pi(\mathrm{R}) = \mathrm{O} \text{ or } \mathrm{Y}$$

Cryptanalysis of Mono- alphabetic Substitution Cipher (3/8)

Trigraphs in the ciphertext such as -QB, QB-, -BO, BO-, -BY, BY-

-GQBQH------------------------XQBN---
--------EQBOY---------------------YQBZ-WBOZ
-YQBKBYYBOT--YQBH----MBYY---------------
-------------------------------------EQBOYQ
BT----LBYH-YQBV-------LQBT--------------
-YBYQBG-----KBYYBOZ-YQBH----MBYU-------
--------YQBZYQBOT

Cryptanalysis of Mono- alphabetic Substitution Cipher (3/8)

Trigraphs in the ciphertext such as -QB, QB-, -BO, BO-, -BY, BY-

- GQBQH - XQB N - - -
- - - - - - - EQBOY - YQBZ - WBOZ
- YQBKBYYBOT - - YQBH - - - - MBYY - - - - - - - - - - - - - - - -
- EQ B OYQ
BT - - - - LBYH - YQBV - - - - - - - LQBT- - - - - - - - - - - - - -
- YBYQBG - - - - - - KBYYBOZ-YQ BH - - - -MBYU - - - - - - -
- - - - - - - YQBZYQBOT

The Trigraph Frequencies in the English Language

the and tha ent ion tio for nde has nce tis oft men

The Trigraph Frequencies in the ciphertext

GQB	QBQ	BQH	XQB	QBN	EQB	QBO	BOY	YQB	QBZ	WBO	BOZ	QBK	KBY	BYY	YBO	BOT
1	1	1	1	1	2	3	2	9	2	1	2	1	2	3	2	2

QBH	MBY	QBT	YBY	BYQ	QBG	BYU
2	2	2	1	1	1	1

Cryptanalysis of Mono- alphabetic Substitution Cipher (3/8)

Trigraphs in the ciphertext such as -QB, QB-, -BO, BO-, -BY, BY-

- GQBQH - XQB N - - -
- - - - - - - EQBOY - - - - - - - - - - - - - - - - - - - YQBZ - WBOZ
- YQBKBYYBOT - - YQBH - - - - MBYY - - - - - - - - - - - - - -
- EQ B OYQ
BT - - - - LBYH - YQBV - - - - - - - LQBT - - - - - - - - - - - - - -
- YBYQBG - - - - - KBYYBOZ-YQ BH - - - - MBYU - - - - - - -
- - - - - - - YQBZYQBOT

The Trigraph Frequencies in the English Language

the and tha ent ion tio for nde has nce tis oft men

The Trigraph Frequencies in the ciphertext

GQB	QBQ	BQH	XQB	QBN	EQB	QBO	BOY	YQB	QBZ	WBO	BOZ	QBK	KBY	BYY	YBO	BOT
1	1	1	1	1	2	3	2	9	2	1	2	1	2	3	2	2

QBH	MBY	QBT	YBY	BYQ	QBG	BYU
2	2	2	1	1	1	1

$e_\pi(\mathrm{THE}) = \mathrm{YQB} \Rightarrow e_\pi(\mathrm{T}) = \mathrm{Y}$ and $e_\pi(\mathrm{H}) = \mathrm{Q} \Rightarrow e_\pi(\mathrm{R}) = \mathrm{O}$

Cryptanalysis of Mono- alphabetic Substitution Cipher (4/8)

Digraphs in the ciphertext with Y

- - - - - - - - - XYQ - - - - - - - - - UYV - KYZ - - - - - - ZY
BVYV - - - - - - OYQHYV - - - - - - - - - - - LYQ - - - - - -
GYQ - - BYYB - - - GYQ - - - - - - BYYQHYUZYH - - - - -
- - - - - - - - - - DYV - - - - - - - - - - - TYZ - - - - - - OYQ
- - - - - - - BYHYYQ - - - - - - - - - - - - - - - - - - - TYVY
DYBYQ - - - - OYQ - BYYB - - GYQ - - - - - - - BYU - - - - -
- - - - - - ZYQ - ZYQ - - -

Cryptanalysis of Mono- alphabetic Substitution Cipher (4/8)

Digraphs in the ciphertext with Y

```
---------XYQ----------UYV-KYZ-------ZY
BVYV------OYQHYV-------------LYQ-------
GYQ--BYYB---GYQ-------BYYQHYUZYH-------
----------DYV-------------TYZ--------OYQ
-------BYHYYQ---------------------TYVY
DYBYQ----OYQ-BYYB--GYQ-------BYU-------
------ZYQ-ZYQ---
```

The Digraph Frequencies in the English Language

th he an in er on re ed nd ha at en es of nt ea ti to io le is ou ar as de rt ve

Digraph Frequency in the ciphertext with Y

XY	YQ	UY	YV	KY	YZ	ZY	YB	VY	OY	HY	LY	GY	BY	YU	YH	DY	TY
1	13	1	5	1	2	4	4	2	3	3	1	3	6	2	1	2	2

Cryptanalysis of Mono- alphabetic Substitution Cipher (4/8)

Digraphs in the ciphertext with Y

\- - - - - - - - - XYQ - - - - - - - - - UYV - KYZ - - - - - - ZY
BVYV - - - - - - OYQHYV - - - - - - - - - - - LYQ - - - - - -
GYQ - - BYYB - - - GYQ - - - - - - BYYQHYUZYH - - - - -
\- - - - - - - - - - DYV - - - - - - - - - - - TYZ - - - - - - OYQ
\- - - - - - - BYHYYQ - - - - - - - - - - - - - - - - - - - TYVY
DYBYQ - - - - OYQ - BYYB - - GYQ - - - - - - - BYU - - - - -
\- - - - - - ZYQ - ZYQ - - -

The Digraph Frequencies in the English Language

th he an in er on re ed nd ha at en es of nt ea ti to io le is ou ar as de rt ve

Digraph Frequency in the ciphertext with Y

XY	YQ	UY	YV	KY	YZ	ZY	YB	VY	OY	HY	LY	GY	BY	YU	YH	DY	TY
1	13	1	5	1	2	4	4	2	3	3	1	3	6	2	1	2	2

$e_\pi(\mathrm{TH}) = \mathrm{YQ}$, $e_\pi(\mathrm{TI\ or\ TO}) = \mathrm{YV}$

Cryptanalysis of Mono- alphabetic Substitution Cipher (5/8)

Trigraphs in the ciphertext such as -YQ, YQ-, -YV, YV-

```
- - - - - - - - - XYQV - - - - - - - - - UYVH - - - - - - - - - - - -
- VYVU - - - - - OYQHYVT - - - - - - - - - - - LYQB - - - - - -
GYQB - - - - - - - - GYQB - - - - - - - YYQH - - - - - - - - - -
- - - - - - - - - - DYVG - - - - - - - - - - - - - - - - - - - OYQ
B - - - - - - - - - YYQB - - - - - - - - - - - - - - - - - - TYVY
- - BYQB - - - OYQK - - - - - - GYQB - - - - - - - - - - - - -
- - - - - - - ZYQBZYQB - -
```

Trigraphs in the ciphertext such as -YQ, YQ-, -YV, YV-

---------XYQV--------UYVH-----------

-VYVU-----OYQHYVT-----------LYQB------

GYQB--------GYQB-------YYQH----------

----------DYVG------------------OYQ

B---------YYQB------------------TYVY

--BYQB---OYQK------GYQB-------------

------ZYQBZYQB--

The Trigraph Frequencies in the English Language

the and tha ent ion tio for nde has nce tis oft men

The Trigraph Frequencies in the ciphertext

XYQ	YQV	UYV	YVH	VYV	YVU	OYQ	YQH	DYV	YVG	YQB	YYQ	TYV	YVY
1	1	1	1	1	1	3	2	1	1	6	1	1	1

BYQ	YQK	GYQ	ZYQ
1	1	1	1

$e_\pi(\mathrm{THE}) = \mathrm{YQB}$, $e_\pi(\mathrm{THA}) = \mathrm{YQH}$, $\Rightarrow e_\pi(\mathrm{A}) = \mathrm{H}$

$e_\pi(\mathrm{TIO\ or\ TIS}) = \mathrm{YVH\ or\ YVU\ or\ YVG} \Rightarrow e_\pi(\mathrm{I}) = \mathrm{V}$

Cryptanalysis of Mono- alphabetic Substitution Cipher (7/8)

B $\rightarrow$ E, Y $\rightarrow$ T, Q $\rightarrow$ H, O $\rightarrow$ R, H $\rightarrow$ A, V $\rightarrow$ I

```
VGQBQHWHUXYQVULRZUGVWBUYVHKYZTHXQBNOZY
I-HEHA-A--THI------I-E-TIA-T--A-HE-R-T
--------------------------------------
BVYVURVEQBOYQHYVTMXTZRQHULVULYQBZOWBOZ
EITI--I-HERTHATI------HA--I--THE-R-ER-
--------------------------------------
GYQBKBYYBOTZGYQBHKEQHMBYYQHYUZYHNZOWRZ
-THE-ETTER---THEA--HA-ETTHAT--TA--R---
--------------------------------------
DKWMBPHWBZDYVGHUXZUBNVTQBTYZWBRVEQBOYQ
----E-A-E--TI-A----E-I-HE-T--E-I-HERTH
--------------------------------------
BTBHUWLBYHYYQBVOPBHUVULQBTQZDKWTDMTYVY
E-EA---ETATTHEIR-EA-I--HE-H--------TIT
--------------------------------------
DYBYQBGZDOYQKBYYBOZGYQBHKEQHMBYUHPBKXG
-TETHE---RTH-ETTER--THEA--HA-ET-A-E---
--------------------------------------
ZOHUWTZYQBZYQBOT
-RA----THE-THER-
```

Cryptanalysis of Mono- alphabetic Substitution Cipher (8/8)

B $\rightarrow$ E, Y $\rightarrow$ T, Q $\rightarrow$ H, O $\rightarrow$ R, H $\rightarrow$ A, V $\rightarrow$ I, G $\rightarrow$ F, K $\rightarrow$ L, W $\rightarrow$ D

```
VGQBQHWHUXYQVULRZUGVWBUYVHKYZTHXQBNOZY
IFHEHADA--THI-----FIDE-TIALT--A-HE-R-T
--------------------------------------
BVYVURVEQBOYQHYVTMXTZRQHULVULYQBZOWBOZ
EITI--I-HERTHATI------HA--I--THE-RDER-
--------------------------------------
GYQBKBYYBOTZGYQBHKEQHMBYYQHYUZYHNZOWRZ
FTHELETTER--FTHEAL-HA-ETTHAT--TA--RD--
--------------------------------------
DKWMBPHWBZDYVGHUXZUBNVTQBTYZWBRVEQBOYQ
-LD-E-ADE--TIFA----E-I-HE-T-DE-I-HERTH
--------------------------------------
BTBHUWLBYHYYQBVOPBHUVULQBTQZDKWTDMTYVY
E-EA-D-ETATTHEIR-EA-I--HE-H--LD----TIT
--------------------------------------
DYBYQBGZDOYQKBYYBOZGYQBHKEQHMBYUHPBKXG
-TETHEF--RTHLETTER-FTHEAL-HA-ET-A-EL-F
--------------------------------------
ZOHUWTZYQBZYQBOT
-RA-D--THE-THER-
```

Cryptanalysis of Mono- alphabetic Substitution Cipher (8/8)

B → E, Y → T, Q → H, O → R, H → A, V → I, G → F, K → L,
W → D, U → N, X → Y, L → G

VGQBQHWHUXYQVULRZUGVWBUYVHKYZTHXQBNOZY
IFHEHADANYTHING--NFIDENTIALT--AYHE-R-T

BVYVURVEQBOYQHYVTMXTZRQHULVULYQBZOWBOZ
EITIN-I-HERTHATI--Y---HANGINGTHE-RDER-

GYQBKBYYBOTZGYQBHKEQHMBYYQHYUZYHNZOWRZ
FTHELETTER--FTHEAL-HA-ETTHATN-TA--RD--

DKWMBPHWBZDYVGHUXZUBNVTQBTYZWBRVEQBOYQ
-LD-E-ADE--TIFANY-NE-I-HE-T-DE-I-HERTH

BTBHUWLBYHYYQBVOPBHUVULQBTQZDKWTDMTYVY
E-EANDGETATTHEIR-EANINGHE-H--LD----TIT

DYBYQBGZDOYQKBYYBOZGYQBHKEQHMBYUHPBKXG
-TETHEF--RTHLETTER-FTHEAL-HA-ETNA-ELYF

ZOHUWTZYQBZYQBOT
-RAND--THE-THER-

Cryptanalysis of Mono- alphabetic Substitution Cipher (8/8)

B $\rightarrow$ E, Y $\rightarrow$ T, Q $\rightarrow$ H, O $\rightarrow$ R, H $\rightarrow$ A, V $\rightarrow$ I, G $\rightarrow$ F, K $\rightarrow$ L,
W $\rightarrow$ D, U $\rightarrow$ N, X $\rightarrow$ Y, L $\rightarrow$ G, R $\rightarrow$ C, Z $\rightarrow$ O, T $\rightarrow$ S

VGQBQHWHUXYQVULRZUGVWBUYVHKYZTHXQBNOZY

IFHEHADANYTHINGCONFIDENTIALTOSAYHE-ROT

BVYVURVEQBOYQHYVTMXTZRQHULVULYQBZOWBOZ

EITINCI-HERTHATIS-YSOCHANGINGTHEORDERO

GYQBKBYYBOTZGYQBHKEQHMBYYQHYUZYHNZOWRZ

FTHELETTERSOFTHEAL-HA-ETTHATNOTA-ORDCO

DKWMBPHWBZDYVGHUXZUBNVTQBTYZWBRVEQBOYQ

-LD-E-ADEO-TIFANYONE-ISHESTODECI-HERTH

BTBHUWLBYHYYQBVOPBHUVULQBTQZDKWTDMTYVY

ESEANDGETATTHEIR-EANINGHESHO-LDS--STIT

DYBYQBGZDOYQKBYYBOZGYQBHKEQHMBYUHPBKXG

-TETHEFO-RTHLETTEROFTHEAL-HA-ETNA-ELYF

ZOHUWTZYQBZYQBOT

ORANDSOTHEOTHERS

Affine Cipher (1/2)

- $\mathcal{P} = \mathcal{C} = \mathbb{Z}_{26}$
- $\mathcal{K} = \{(a, b) \in \mathbb{Z}_{26} \times \mathbb{Z}_{26} : gcd(a, 26) = 1\}$.
- For $K = (a, b) \in \mathcal{K}$, $e_K(x) = y \equiv (ax + b) \bmod 26$ and $d_K(y) = x \equiv a^{-1}\,(y - b) \bmod 26$.

Affine Cipher (1/2)

- $\mathcal{P} = \mathcal{C} = \mathbb{Z}_{26}$
- $\mathcal{K} = \{(a, b) \in \mathbb{Z}_{26} \times \mathbb{Z}_{26} : gcd(a, 26) = 1\}$.
- For $K = (a, b) \in \mathcal{K}$, $e_K(x) = y \equiv (ax + b) \bmod 26$ and $d_K(y) = x \equiv a^{-1} (y - b) \bmod 26$.

$a \in \mathbb{Z}_n$, *multiplicative inverse* of $a \bmod n$, denoted by $a^{-1} \bmod n$.
$aa^{-1} \equiv 1 \bmod n$.

Affine Cipher (1/2)

- $\mathcal{P} = \mathcal{C} = \mathbb{Z}_{26}$
- $\mathcal{K} = \{(a, b) \in \mathbb{Z}_{26} \times \mathbb{Z}_{26} : gcd(a, 26) = 1\}$.
- For $K = (a, b) \in \mathcal{K}$, $e_K(x) = y \equiv (ax + b) \bmod 26$ and $d_K(y) = x \equiv a^{-1}(y - b) \bmod 26$.

$a \in \mathbb{Z}_n$, *multiplicative inverse* of a mod n, denoted by a^{-1} mod n.
$aa^{-1} \equiv 1 \bmod n$.

For example, $a = 2 \Rightarrow 2 \times 2 \bmod 26 = 4$, $2 \times 3 \bmod 26 = 6$,
$2 \times 4 \bmod 26 = 8$, $2 \times 5 \bmod 26 = 10$, $2 \times 6 \bmod 26 = 12$,
$2 \times 7 \bmod 26 = 14$, $2 \times 8 \bmod 26 = 16$, $2 \times 9 \bmod 26 = 18$,
$2 \times 10 \bmod 26 = 20$, $2 \times 11 \bmod 26 = 22$, $2 \times 12 \bmod 26 = 24$,
$2 \times 13 \bmod 26 = 0$, $2 \times 14 \bmod 26 = 2$, $2 \times 15 \bmod 26 = 4$,
$2 \times 16 \bmod 26 = 6$, ...

Affine Cipher (1/2)

- $\mathcal{P} = \mathcal{C} = \mathbb{Z}_{26}$
- $\mathcal{K} = \{(a, b) \in \mathbb{Z}_{26} \times \mathbb{Z}_{26} : gcd(a, 26) = 1\}$.
- For $K = (a, b) \in \mathcal{K}$, $e_K(x) = y \equiv (ax + b) \bmod 26$ and $d_K(y) = x \equiv a^{-1}(y - b) \bmod 26$.

$a \in \mathbb{Z}_n$, *multiplicative inverse* of a mod n, denoted by a^{-1} mod n. $aa^{-1} \equiv 1 \bmod n$.

For example, $a = 2 \Rightarrow 2 \times 2 \bmod 26 = 4$, $2 \times 3 \bmod 26 = 6$, $2 \times 4 \bmod 26 = 8$, $2 \times 5 \bmod 26 = 10$, $2 \times 6 \bmod 26 = 12$, $2 \times 7 \bmod 26 = 14$, $2 \times 8 \bmod 26 = 16$, $2 \times 9 \bmod 26 = 18$, $2 \times 10 \bmod 26 = 20$, $2 \times 11 \bmod 26 = 22$, $2 \times 12 \bmod 26 = 24$, $2 \times 13 \bmod 26 = 0$, $2 \times 14 \bmod 26 = 2$, $2 \times 15 \bmod 26 = 4$, $2 \times 16 \bmod 26 = 6$, ...

Hence there is no x as $2 \times x \bmod 26 = 1$, then 2 does not have a multiplicative inverse mod26.

Affine Cipher (2/2)

In order that decryption is possible, for any $y \in \mathbb{Z}_{26}$
$\Longleftrightarrow gcd(a, 26) = 1$.

Affine Cipher (2/2)

In order that decryption is possible, for any $y \in \mathbb{Z}_{26}$
$\iff gcd(a, 26) = 1$.

- If $gcd(a, 26) = d > 1$, then $0 \equiv ax \bmod 26$ has to distinct solutions in $\mathbb{Z}_{26}$, namely $x = 0$ and $x = \frac{26}{d}$.
- In this case $e(x) = ax + b \bmod 26$ is not an injective function and hence not a valid encryption function.
- Since $26 = 2 \times 13$, $a = 1, 3, 5, 7, 9, 11, 15, 17, 19, 21, 23, 25$, b can be any element in $\mathbb{Z}_{26}$.
- Hence affine cipher has $12 \times 26 = 312$ possible keys.

Cryptanalysis of the Affine Cipher (1/3)

Ciphertext:
KADHLFMLNMFKVERSLDYAREHFSOORLDYAREHKRWKNHDS
XFSFUUDSRLDYAREDSTADLAMRKKREHFERRSLVORONHDS
XKARUVEPNMFTARERFSOFERTAVMRSNPIREHIRKTRRSFS
OFSODHERMFKDQRMZYEDPR

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

Letter Frequency in the ciphertext

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
6	0	0	8	8	8	0	5	1	0	6	4	4	3	4	2	0	16	8	2	2	2	0	1	2	0

Cryptanalysis of the Affine Cipher (2/3)

$E_e(\mathrm{E}) = \mathrm{R} \to 17 \equiv a4 + b \bmod 26$
$E_e(\mathrm{T}) = \mathrm{D} \to 3 \equiv a19 + b \bmod 26$ then $a = 6$, $gcd(6, 26) = 2 > 1$.
$E_e(\mathrm{T}) = \mathrm{E} \to 4 \equiv a19 + b \bmod 26$ then $a = 13$,
$gcd(13, 26) = 13 > 1$.
$E_e(\mathrm{T}) = \mathrm{F} \to 5 \equiv a19 + b \bmod 26$ then $a = 20$,
$gcd(20, 26) = 2 > 1$.
$E_e(\mathrm{T}) = \mathrm{S} \to 18 \equiv a19 + b \bmod 26$ then $a = 7$, $gcd(7, 26) = 1$ and
$b = 9$.

Decrypted message

PVOWESTEITSPYDQFEORVQDWSFXXQEORVQDWPQNPIWO
FCSFSJJOFQEORVQDOFUVOEVTQPPQDWSDQQFEYXQXIW
OFCPVQJYDMITSUVQDQSFXSDQUVYTQFIMLQDWLQPUQQ
FSFXSFXOWDQTSPOBQTGRDOMQ

Cryptanalysis of the Affine Cipher (3/3)

$E_e(\mathrm{T}) = \mathrm{K} \rightarrow 10 \equiv a19 + b$ mod 26 then $a = 3$, $gcd(3, 26) = 1$ and $b = 5$.

Decrypted message

THISCALCULATORENCIPHERSANDDECIPHERSTEXTUSI
NGANAFFINECIPHERINWHICHLETTERSAREENCODEDUS
INGTHEFORMULAWHEREANDAREWHOLENUMBERSBETWEE
NANDANDISRELATIVELYPRIME

THIS CALCULATOR ENCIPHERS AND DECIPHERS TEXT USING AN AFFINE CIPHER IN WHICH LETTERS ARE ENCODED USING THE FORMULA WHERE AND ARE WHOLE NUMBERS BETWEEN AND AND IS RELATIVELY PRIME

Alberti Cipher

All of the Western European governments used cryptography
Venice created an elaborate organization in 1452.
Leon Battista Alberti was known as “The Father of Western Cryptology” in part because of his development of polyalphabetic substitution.

Formula
The larger one is called Stabilis [stationary or fixed], the smaller one is called Mobilis [movable]
Polyalphabetic substitution is any technique which allows different ciphertext symbols to represent the same plaintext symbol.

Vigenère Cipher

$m \geq 0$ and $m \in \mathbb{Z}$.
Let $\mathcal{P} = \mathcal{C} = \mathcal{K} = (\mathbb{Z}_{26})^m$.

Vigenère Cipher

$m \geq 0$ and $m \in \mathbb{Z}$.
Let $\mathcal{P} = \mathcal{C} = \mathcal{K} = (\mathbb{Z}_{26})^m$.
For a key $K = (k_1, k_2, \ldots, k_m)$, we define
$e_K(x_1, x_2, \ldots, x_m) = (x_1 + k_1, x_2 + k_2, \ldots, x_m + k_m)$
$d_K(y_1, y_2, \ldots, y_m) = (y_1 - k_1, y_2 - k_2, \ldots, y_m - k_m)$

Vigenère Cipher

$m \geq 0$ and $m \in \mathbb{Z}$.
Let $\mathcal{P} = \mathcal{C} = \mathcal{K} = (\mathbb{Z}_{26})^m$.
For a key $K = (k_1, k_2, \ldots, k_m)$, we define
$e_K(x_1, x_2, \ldots, x_m) = (x_1 + k_1, x_2 + k_2, \ldots, x_m + k_m)$
$d_K(y_1, y_2, \ldots, y_m) = (y_1 - k_1, y_2 - k_2, \ldots, y_m - k_m)$
Example: $m = 6$. The keyword is *CIPHER*. $K = (2, 8, 15, 7, 4, 17)$.
plaintext: namedafterblaisedevigenere

$$\begin{array}{cccccccccccccccccccccccccc}
13 & 0 & 12 & 4 & 3 & 0 & 5 & 19 & 4 & 17 & 1 & 11 & 0 & 8 & 18 & 4 & 3 & 4 & 21 & 8 & 6 & 4 & 13 & 4 & 17 & 4 \\
2 & 8 & 15 & 7 & 4 & 17 & 2 & 8 & 15 & 7 & 4 & 17 & 2 & 8 & 15 & 7 & 4 & 17 & 2 & 8 & 15 & 7 & 4 & 17 & 2 & 8 \\
\hline
15 & 8 & 1 & 11 & 7 & 17 & 7 & 1 & 19 & 24 & 5 & 2 & 2 & 16 & 7 & 11 & 7 & 21 & 23 & 16 & 21 & 11 & 17 & 21 & 19 & 12
\end{array}$$

ciphertext: piblhrhbtyfccqhlhvxqvlrvtm

Vigenère Cipher

$m \geq 0$ and $m \in \mathbb{Z}$.
Let $\mathcal{P} = \mathcal{C} = \mathcal{K} = (\mathbb{Z}_{26})^m$.
For a key $K = (k_1, k_2, \ldots, k_m)$, we define
$e_K(x_1, x_2, \ldots, x_m) = (x_1 + k_1, x_2 + k_2, \ldots, x_m + k_m)$
$d_K(y_1, y_2, \ldots, y_m) = (y_1 - k_1, y_2 - k_2, \ldots, y_m - k_m)$
Example: $m = 6$. The keyword is *CIPHER*. $K = (2, 8, 15, 7, 4, 17)$.
plaintext: namedafterblaisedevigenere

13	0	12	4	3	0	5	19	4	17	1	11	0	8	18	4	3	4	21	8	6	4	13	4	17	4
2	8	15	7	4	17	2	8	15	7	4	17	2	8	15	7	4	17	2	8	15	7	4	17	2	8
15	8	1	11	7	17	7	1	19	24	5	2	2	16	7	11	7	21	23	16	21	11	17	21	19	12

ciphertext: piblhrhbtyfccqhlhvxqvlrvtm

The number of possible keywords of length m is 26^m.

Cryptanalysis of Vigenère Cipher

The first step is to determine the keyword length, m.

- Kasiski test : 1854 - Charles Babbage and 1863 - Friedrich Kasiski
- Index of coincidence

Kasiski Test

Two identical segments of plaintext will be encrypted to the same ciphertext whenever their occurrence in the plaintext is Δ positions apart, where $\Delta \equiv 0 \bmod m$.

- Search the ciphertext for pairs of identical segments of length at least three.
- Record the distance between the starting positions of the two segments.
- If we obtain several such distances, say $\Delta_1, \Delta_2, \ldots$ we would conjecture that m divides all of the Δ_i's, $m \mid gcd\,(\Delta_1, \Delta_2, \ldots)$

The reason this test works is that if a repeated string occurs in the plaintext, and the distance between them is a multiple of the keyword length, m, the keyword letters will line up in the same way with both occurrences of the string.

Example for Kasiski Test (1/2)

ciphertext:

	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	26	27	28	29	30
1	K	N	W	L	L	R	F	I	C	F	X	Y	K	V	V	J	S	E	H	Q	S	L	R	K	K	I	A	A	S	J
2	C	G	W	L	A	I	Q	B	T	P	X	Z	P	K	X	W	L	V	T	B	W	H	X	Z	U	J	N	Z	S	T
3	J	I	C	N	M	E	I	B	W	L	S	I	F	M	G	V	J	K	J	M	A	L	X	K	G	Z	H	V	J	K
4	J	M	P	S	T	Y	C	J	T	A	X	Y	C	B	C	V	X	R	Y	W	G	K	G	F	W	T	S	I	I	D
5	C	L	T	V	Y	K	K	N	P	U	C	F	P	M	L	P	W	Y	G	A	I	V	H	V	E	Q	E	O	I	I
6	V	P	T	Z	I	R	P	L	V	L	X	R	V	B	W	L	M	I	O	M	P	U	M	E	I	P	T	Z	L	F
7	W	T	S	Z	Y	S	U	B	X	A	Y	K	G	B	W	L	J	F	W	Z	I	O	P	V	V	B	T	Y	S	W
8	V	P	T	H	P	G	J	I	Q	L	X	E	C	U	T	S	C	W	Q	Z	P	U	H	J	Q	B	W	L	S	K
9	J	M	G	Z																										

Example for Kasiski Test (2/2)

ciphertext string	occurs at (index)	spacing	factors
MEI	64 172	108	2 3 4 6 9 12 18 27 36 54 108
BWL	67 163	96	2 3 4 6 8 12 16 24 32 48 96
BWL	67 193	126	2 3 6 7 9 14 18 21 42 63 126
BWLS	67 235	168	2 3 4 6 7 8 12 14 21 24 28 42 56 84 168
WLS	68 236	168	2 3 4 6 7 8 12 14 21 24 28 42 56 84 168
VJKJM	75 87	12	2 3 4 6 12
JKJM	76 88	12	2 3 4 6 12
KJM	77 89	12	2 3 4 6 12
KJM	77 239	162	2 3 6 9 18 27 54 81 162
KJM	89 239	150	2 3 5 6 10 15 25 30 50 75 150
FWTS	113 179	66	2 3 6 11 22 33 66
WTS	114 180	66	2 3 6 11 22 33 66
VPT	150 210	60	2 3 4 5 6 10 12 15 20 30 60
PTZ	151 175	24	2 3 4 6 8 12 24
BWL	163 193	30	2 3 5 6 10 15 30
BWL	163 235	72	2 3 4 6 8 9 12 18 24 36 72
BWL	193 235	42	2 3 6 7 14 21 42

Example for Kasiski Test (2/2)

ciphertext string	occurs at (index)	spacing	factors
MEI	64 172	108	2 3 4 6 9 12 18 27 36 54 108
BWL	67 163	96	2 3 4 6 8 12 16 24 32 48 96
BWL	67 193	126	2 3 6 7 9 14 18 21 42 63 126
BWLS	67 235	168	2 3 4 6 7 8 12 14 21 24 28 42 56 84 168
WLS	68 236	168	2 3 4 6 7 8 12 14 21 24 28 42 56 84 168
VJKJM	75 87	12	2 3 4 6 12
JKJM	76 88	12	2 3 4 6 12
KJM	77 89	12	2 3 4 6 12
KJM	77 239	162	2 3 6 9 18 27 54 81 162
KJM	89 239	150	2 3 5 6 10 15 25 30 50 75 150
FWTS	113 179	66	2 3 6 11 22 33 66
WTS	114 180	66	2 3 6 11 22 33 66
VPT	150 210	60	2 3 4 5 6 10 12 15 20 30 60
PTZ	151 175	24	2 3 4 6 8 12 24
BWL	163 193	30	2 3 5 6 10 15 30
BWL	163 235	72	2 3 4 6 8 9 12 18 24 36 72
BWL	193 235	42	2 3 6 7 14 21 42

Keyword length $m = 6$.

Index of Coincidence (1/3)

1920 - Friedman

Definition: Suppose $x = x_1 x_2 \ldots x_n$ is a string of n alphabetic characters. The *index of coincidence* of x, denoted by $I_c(x)$, is defined to be the probability that two random elements of x are identical.

Index of Coincidence (1/3)

1920 - Friedman

Definition: Suppose $x = x_1 x_2 \ldots x_n$ is a string of n alphabetic characters. The *index of coincidence* of x, denoted by $I_c(x)$, is defined to be the probability that two random elements of x are identical.

- Suppose the frequencies (number of occurrences) of A, B, C, ..., Z in x are $f_0, f_1, \ldots, f_{25}$.
- $I_c(x) = \frac{f_0}{n}\frac{f_0-1}{n-1} + \frac{f_1}{n}\frac{f_1-1}{n-1} + \cdots + \frac{f_{25}}{n}\frac{f_{25}-1}{n-1}$

Index of Coincidence (1/3)

1920 - Friedman

Definition: Suppose $x = x_1 x_2 \ldots x_n$ is a string of n alphabetic characters. The *index of coincidence* of x, denoted by $I_c(x)$, is defined to be the probability that two random elements of x are identical.

- Suppose the frequencies (number of occurrences) of A, B, C, ..., Z in x are $f_0, f_1, \ldots, f_{25}$.
- $I_c(x) = \frac{f_0}{n}\frac{f_0-1}{n-1} + \frac{f_1}{n}\frac{f_1-1}{n-1} + \cdots + \frac{f_{25}}{n}\frac{f_{25}-1}{n-1}$

$$I_c(x) = \frac{\sum_{i=0}^{25} f_i(f_i-1)}{n(n-1)} = \frac{\sum_{i=0}^{25} f_i^2 - \sum_{i=0}^{25} f_i}{n(n-1)} = \frac{\sum_{i=0}^{25} f_i^2 - n}{n(n-1)} = \frac{\sum_{i=0}^{25} f_i^2}{n(n-1)} - \frac{1}{n-1}$$

Index of Coincidence (1/3)

1920 - Friedman

Definition: Suppose $x = x_1 x_2 \ldots x_n$ is a string of n alphabetic characters. The *index of coincidence* of x, denoted by $I_c(x)$, is defined to be the probability that two random elements of x are identical.

- Suppose the frequencies (number of occurrences) of A, B, C, ..., Z in x are $f_0, f_1, \ldots, f_{25}$.
- $I_c(x) = \frac{f_0}{n}\frac{f_0-1}{n-1} + \frac{f_1}{n}\frac{f_1-1}{n-1} + \cdots + \frac{f_{25}}{n}\frac{f_{25}-1}{n-1}$

$I_c(x) = \frac{\sum_{i=0}^{25} f_i(f_i-1)}{n(n-1)} = \frac{\sum_{i=0}^{25} f_i^2 - \sum_{i=0}^{25} f_i}{n(n-1)} = \frac{\sum_{i=0}^{25} f_i^2 - n}{n(n-1)} = \frac{\sum_{i=0}^{25} f_i^2}{n(n-1)} - \frac{1}{n-1}$

$n \to \infty \Rightarrow I_c(x) \to \frac{\sum_{i=0}^{25} f_i^2}{n^2}$

$p_i = \frac{f_i}{n}$ then $I_c(x) \to \sum_{i=0}^{25} p_i^2$

Index of Coincidence (2/3)

letter	probability	letter	probability	letter	probability
A	0.0856	B	0.0139	C	0.0279
D	0.0378	E	0.1304	F	0.0289
G	0.0199	H	0.0528	I	0.0627
J	0.0013	K	0.0042	L	0.0339
M	0.0249	N	0.0707	O	0.0797
P	0.0199	Q	0.0012	R	0.0677
S	0.0607	T	0.1045	U	0.0249
V	0.0092	W	0.0149	X	0.0017
Y	0.0199	Z	0.0008		

Index of Coincidence (2/3)

letter	probability	letter	probability	letter	probability
A	0.0856	B	0.0139	C	0.0279
D	0.0378	E	0.1304	F	0.0289
G	0.0199	H	0.0528	I	0.0627
J	0.0013	K	0.0042	L	0.0339
M	0.0249	N	0.0707	O	0.0797
P	0.0199	Q	0.0012	R	0.0677
S	0.0607	T	0.1045	U	0.0249
V	0.0092	W	0.0149	X	0.0017
Y	0.0199	Z	0.0008		

$n \to \infty \Rightarrow I_c(x) \approx \sum_{i=0}^{25} p_i^2 = 0.065$

Index of Coincidence (2/3)

letter	probability	letter	probability	letter	probability
A	0.0856	B	0.0139	C	0.0279
D	0.0378	E	0.1304	F	0.0289
G	0.0199	H	0.0528	I	0.0627
J	0.0013	K	0.0042	L	0.0339
M	0.0249	N	0.0707	O	0.0797
P	0.0199	Q	0.0012	R	0.0677
S	0.0607	T	0.1045	U	0.0249
V	0.0092	W	0.0149	X	0.0017
Y	0.0199	Z	0.0008		

$n \to \infty \Rightarrow I_c(x) \approx \sum_{i=0}^{25} p_i^2 = 0.065$

The same reasoning applies if x is a ciphertext string obtained using any monoalphabetic cipher.

Index of Coincidence (3/3)

$y = y_1 y_2 \ldots y_n$ constructed by Vigenère Cipher.
m substrings of y, $\bar{y_1}$, $\bar{y_2}$, ..., $\bar{y_m}$ by writing out the ciphertext in columns in a rectangular array of dimensions $m \times (n/m)$.
Example: $n = 15$ and $m = 3$

$$\begin{matrix} y_1 & y_4 & y_7 & y_{10} & y_{13} \\ y_2 & y_5 & y_8 & y_{11} & y_{14} \\ y_3 & y_6 & y_9 & y_{12} & y_{15} \end{matrix} \Rightarrow \begin{matrix} \bar{y_1} & = & y_1 y_4 y_7 y_{10} y_{13} \\ \bar{y_2} & = & y_2 y_5 y_8 y_{11} y_{14} \\ \bar{y_3} & = & y_3 y_6 y_9 y_{12} y_{15} \end{matrix}$$

If $I_c(\bar{y_i}) \approx 0.065$ then m is the keyword length.
If m is not the keyword length then $\bar{y_i}$ s are random
$\sum_{i=0}^{25} f_i = n$, in random text $f_0 = f_1 = \cdots = f_{25}$, $26 f_i = n$, $f_i = \frac{n}{26}$,
$I_c(x) = \frac{26 f_i^2}{n^2} = \frac{1}{26} = 0.038$. The two values 0.065 and 0.038 are sufficiently far apart that we will often be able to determine the correct keyword length.

Permutation Cipher

Alter the plaintext characters positions by rearranging them using a permutation.

- m: positive integer
- $\mathcal{P} = \mathcal{C} = (\mathbb{Z}_{26})^m$
- $\mathcal{K}$ consist of all permutations of $1, \ldots, m$
- For a key π, $e_\pi(x_1, \ldots, x_m) = \left(x_{\pi(1)}, \ldots, x_{\pi(m)}\right)$
- $d_\pi(y_1, \ldots, y_m) = \left(y_{\pi^{-1}(1)}, \ldots, y_{\pi^{-1}(m)}\right)$
  where π^{-1} is the inverse permutation to π.

Example for Permutation Cipher

$m = 6$

Encryption: $\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 3 & 1 & 6 & 2 & 5 \end{pmatrix}$

Example for Permutation Cipher

$m = 6$

Encryption: $\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 3 & 1 & 6 & 2 & 5 \end{pmatrix}$

plaintext:

HE WALKED UP AND DOWN THE PASSAGE TWO OR THREE TIMES

Example for Permutation Cipher

$m = 6$

Encryption: $\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 3 & 1 & 6 & 2 & 5 \end{pmatrix}$

plaintext:

HE WALKED UP AND DOWN THE PASSAGE TWO OR THREE TIMES

plaintext is divided into groups of 6:

HEWALK EDUPAN DDOWNT HEPASS AGETWO ORTHRE ETIMES

Example for Permutation Cipher

$m = 6$

Encryption: $\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 3 & 1 & 6 & 2 & 5 \end{pmatrix}$

plaintext:

HE WALKED UP AND DOWN THE PASSAGE TWO OR THREE TIMES

plaintext is divided into groups of 6:

HEWALK EDUPAN DDOWNT HEPASS AGETWO ORTHRE ETIMES

ciphertext:

WLEHKAUADENPONDDTWPSEHSAEWGAOTTRROEHIETESM

Example for Permutation Cipher

$m = 6$

Encryption: $\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 3 & 1 & 6 & 2 & 5 \end{pmatrix}$

plaintext:

HE WALKED UP AND DOWN THE PASSAGE TWO OR THREE TIMES

plaintext is divided into groups of 6:

HEWALK EDUPAN DDOWNT HEPASS AGETWO ORTHRE ETIMES

ciphertext:

WLEHKAUADENPONDDTWPSEHSAEWGAOTTRROEHIETESM

Decryption: $\pi^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 5 & 2 & 1 & 6 & 4 \end{pmatrix}$

Remarks for Permutation Cipher

- Will frequency analysis work for breaking permutation cipher?

Remarks for Permutation Cipher

- Will frequency analysis work for breaking permutation cipher?
- The Permutation Cipher is not monoalphabetic.

Remarks for Permutation Cipher

- Will frequency analysis work for breaking permutation cipher?
- The Permutation Cipher is not monoalphabetic.
- In the example the first E is encrypted as L, the second E is encrypted as U and the third E is encrypted as S.
- This encryption does not change the frequency of alphabetic characters but the positions of the letters.

Remarks for Permutation Cipher

- Will frequency analysis work for breaking permutation cipher?
- The Permutation Cipher is not monoalphabetic.
- In the example the first E is encrypted as L, the second E is encrypted as U and the third E is encrypted as S.
- This encryption does not change the frequency of alphabetic characters but the positions of the letters.
- The different number of keys are $m!$.

Product Cryptosystems 1/2

Introduced by Shannon in 1949

- For simplicity; $\mathcal{C} = \mathcal{P}$: *endomorphic cryptosystem*
- Suppose $S_1 = (\mathcal{P}, \mathcal{P}, \mathcal{K}_1, \varepsilon_1, \mathcal{D}_1)$ and $S_2 = (\mathcal{P}, \mathcal{P}, \mathcal{K}_2, \varepsilon_2, \mathcal{D}_2)$ are two endomorphic cryptosystems.
- *Product cryptosystem* of S_1 and S_2:
  $S_1 \times S_2 = (\mathcal{P}, \mathcal{P}, \mathcal{K}_1 \times \mathcal{K}_2, \varepsilon, \mathcal{D})$.
- A key of the product cryptosystem: $K = (K_1, K_2)$, where $K_1 \in \mathcal{K}_1$ and $K_2 \in \mathcal{K}_2$.
- $e_{(K_1,K_2)}(x) = e_{K_2}(e_{K_1}(x))$ and $d_{(K_1,K_2)}(y) = d_{K_1}(d_{K_2}(y))$.

Product Cryptosystems 2/2

$$\begin{aligned} d_{(K_1,K_2)}\left(e_{(K_1,K_2)}(x)\right) &= d_{(K_1,K_2)}\left(e_{K_2}\left(e_{K_1}(x)\right)\right) \\ &= d_{K_1}\left(d_{K_2}\left(e_{K_2}\left(e_{K_1}(x)\right)\right)\right) \\ &= d_{K_1}\left(e_{K_1}(x)\right) \\ &= x \end{aligned}.$$

- Cryptosystems have the probability distributions associated with their keyspaces.
  $\mathbf{Pr}\left[(K_1, K_2)\right] = \mathbf{Pr}\left[K_1\right] \times \mathbf{Pr}\left[K_2\right]$.
- We choose K_1 and K_2 independently, using the probability distributions defined on $\mathcal{K}_1$ and $\mathcal{K}_2$.
- The product of a substitution cipher with another substitution cipher is another substitution cipher, so for practical purposes, we want to alternate.

Multiplicative Cipher

- $\mathcal{P} = \mathcal{C} = \mathbb{Z}_{26}$ and let $\mathcal{K} = \{a \in \mathbb{Z}_{26} : \gcd(a, 26) = 1\}$.
- For $a \in \mathcal{K}$, define $e_a(x) = ax \bmod 26$ and $d_a(y) = a^{-1}y \bmod 26$ $(x, y \in \mathbb{Z}_{26})$.

Multiplicative Cipher

- $\mathcal{P} = \mathcal{C} = \mathbb{Z}_{26}$ and let $\mathcal{K} = \{a \in \mathbb{Z}_{26} : \gcd(a, 26) = 1\}$.
- For $a \in \mathcal{K}$, define $e_a(x) = ax \bmod 26$ and $d_a(y) = a^{-1}y \bmod 26$ $(x, y \in \mathbb{Z}_{26})$.

Suppose **M** is the Multiplicative Cipher and **S** is the Shift Cipher, then **M** $\times$ **S**=**S** $\times$ **M**=Affine Cipher.

Proof:

S: $e_K(x) = (x + K) \bmod 26$, $K \in \mathbb{Z}_{26}$.

M: $e_K(x) = (ax) \bmod 26$, $a \in \mathbb{Z}_{26}$ and $\gcd(a, 26) = 1$.

M $\times$ **S**: $e_{(a,K)}(x) = (ax + K) \bmod 26$.

Multiplicative Cipher

- $\mathcal{P} = \mathcal{C} = \mathbb{Z}_{26}$ and let $\mathcal{K} = \{a \in \mathbb{Z}_{26} : \gcd(a, 26) = 1\}$.
- For $a \in \mathcal{K}$, define $e_a(x) = ax \bmod 26$ and $d_a(y) = a^{-1}y \bmod 26$ $(x, y \in \mathbb{Z}_{26})$.

Suppose **M** is the Multiplicative Cipher and **S** is the Shift Cipher, then **M** $\times$ **S**=**S** $\times$ **M**=Affine Cipher.

Proof:

S: $e_K(x) = (x + K) \bmod 26$, $K \in \mathbb{Z}_{26}$.

M: $e_K(x) = (ax) \bmod 26$, $a \in \mathbb{Z}_{26}$ and $\gcd(a, 26) = 1$.

M $\times$ **S**: $e_{(a,K)}(x) = (ax + K) \bmod 26$.

The probability of a key in Affine Cipher is $\frac{1}{312} = \frac{1}{12} \times \frac{1}{26}$.

Product Cipher

$$\underbrace{\mathbf{S} \times \mathbf{S} \times \cdots \times \mathbf{S}}_{\text{n}} = S^n$$

If $\mathbf{S}^2 = \mathbf{S}$, then $\mathbf{S}$ is *idempotent cryptosystem*.

Product Cipher

$$\underbrace{\mathbf{S} \times \mathbf{S} \times \cdots \times \mathbf{S}}_{\text{n}} = S^n$$

If $\mathbf{S}^2 = \mathbf{S}$, then $\mathbf{S}$ is *idempotent cryptosystem*.

- If a cryptosystem is not idempotent, then there is a potential increase in security by iterating it several times.
- Taking the product of substitution- type ciphers with permutation- type ciphers is a commonly used technique.

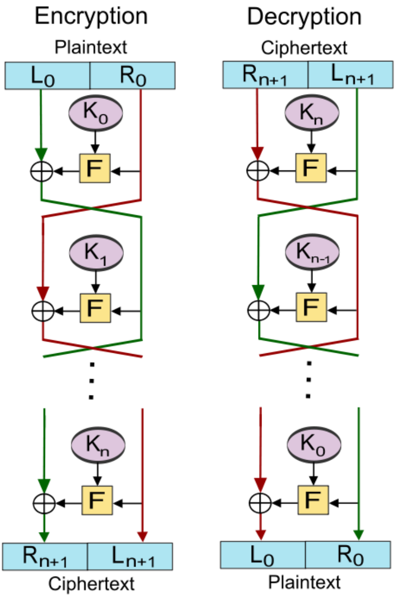
Introduction to Block Cipher 1/2

Iterated cipher: The cipher requires the specification of a *round function* and a *key schedule* and the encryption of a plaintext will proceed through N_r similar *rounds*.

- K: a random binary key.
- $K^1, \cdots, K^{N_r}$: N_r *round keys* (*subkeys*):
- Key schedule: Public algorithm for construction of the round keys from K
- $\omega^r = g\left(\omega^{r-1}, K^r\right)$
  ω^r: next state, ω^{r-1}: current state, K^r: round key, g: round function
- ω^0: plaintext, ω^{N_r}: ciphertext
- In order for decryption to be possible, the function g must be injective (one- to- one)
- If g's second argument is fixed, then $g^{-1}\left(g\left(\omega, a\right), a\right) = \omega$ for all ω and a.
- $\omega^{r-1} = g^{-1}\left(\omega^r, K^r\right)$

Feistel Cipher

- Named after the German-born physicist and cryptographer Horst Feistel who did pioneering research while working for IBM (USA)
- advantage: encryption and decryption operations are very similar, even identical in some cases
- requires only a reversal of the key schedule
- the size of the code or circuitry required to implement such a cipher is nearly halved.



Feistel Cipher - Construction details

Let F be the round function and let $K_0, K_1, \ldots, K_n$ be the sub-keys for the rounds respectively.

Encryption:

1. Split the plaintext block into two equal pieces, (L_0, R_0)
2. For $i = 0, 1, \ldots, n$, compute
   $L_{i+1} = R_i$,
   $R_{i+1} = L_i \oplus F(R_i, K_i)$.
3. The ciphertext is (R_{n+1}, L_{n+1}).

Decryption:

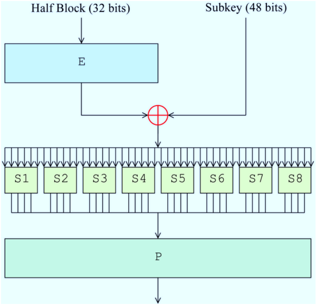
1. Split the ciphertext block into two equal pieces (R_{n+1}, L_{n+1})
2. For $i = n, n-1, \ldots, 0$, compute
   $R_i = L_{i+1}$,
   $L_i = R_{i+1} \oplus F(L_{i+1}, K_i)$.
3. The plaintext is (L_0, R_0).

One advantage of the Feistel model compared to a substitution-permutation network is that the round function does not have to be invertible.

Note the reversal of the subkey order for decryption; this is the only difference between encryption and decryption.

Data Encryption Standard

- Selected by the NBS as an official FIPS for the US in 1976.
- Was initially controversial because of classified design elements, a relatively short key length, and suspicions about a NSA backdoor.
- Insecure due to the 56-bit key size being too small
- In January, 1999, distributed.net and the Electronic Frontier Foundation collaborated to publicly break a DES key in 22 hours and 15 minutes.
- The algorithm is believed to be practically secure in the form of Triple DES, although there are theoretical attacks.



Triple Data Encryption Standard

- TDES uses a “key bundle” which comprises three DES keys, K_1, K_2 and K_3, each of 56 bits.
- The encryption: $\mathit{ciphertext} = E_{K_3}\left(D_{K_2}\left(E_{K_1}\left(\mathit{plaintext}\right)\right)\right)$
- Decryption: $\mathit{plaintext} = D_{K_1}\left(E_{K_2}\left(D_{K_3}\left(\mathit{ciphertext}\right)\right)\right)$

Triple Data Encryption Standard

- TDES uses a “key bundle” which comprises three DES keys, K_1, K_2 and K_3, each of 56 bits.
- The encryption: $ciphertext = E_{K_3}\left(D_{K_2}\left(E_{K_1}\left(plaintext\right)\right)\right)$
- Decryption: $plaintext = D_{K_1}\left(E_{K_2}\left(D_{K_3}\left(ciphertext\right)\right)\right)$

Keying options:

1. Keying option 1: All three keys are independent.
   strongest, with $3 \times 56 = 168$ independent key bits.
2. Keying option 2: K_1 and K_2 are independent, and $K_3 = K_1$.
   provides less security, with $2 \times 56 = 112$ key bits.
3. Keying option 3: All three keys are identical, i.e.
   $K_1 = K_2 = K_3$.
   equivalent to DES, with only 56 key bits

Substitution- Permutation Networks (SPNs)

- ℓ and m are positive integers
- plaintext: $x = (x_1 x_2 \cdots x_{\ell m})_2$ and ciphertext: $y = (y_1 y_2 \cdots y_{\ell m})_2$
- ℓm:block length
- S-box, $\pi_S : \{0,1\}^\ell \longrightarrow \{0,1\}^\ell$ is substitution.
  It is used to replace ℓ bits with a different set of ℓ bits.
- Permutation, π_P: $\{1,\ldots,\ell m\} \longrightarrow \{1,\ldots,\ell m\}$
  It is used to permute ℓm bits.
- $x = (x_1 x_2 \cdots x_{\ell m}) = x_{(1)} \| \cdots \| x_{(m)}$ for $1 \le i \le m$
  $x_{(i)} = (x_{i,\ell-1},\ldots,x_{i,0})$.
- The first and last operations in a round are XORs with subkeys: *whitenning*.

An Example SPN (1/2)

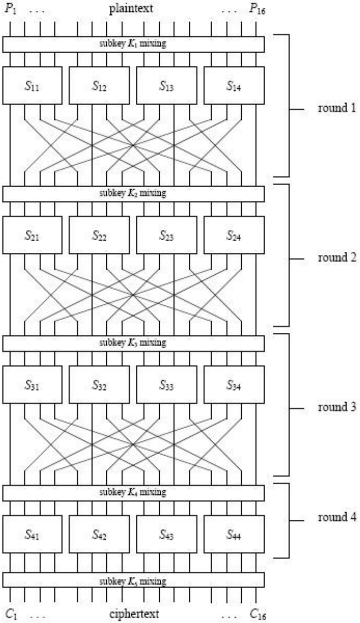
$\ell = m = N_r = 4.$

z	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
$\pi_S(z)$	E	4	D	1	2	F	B	8	3	A	6	C	5	9	0	7

z	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$\pi_P(z)$	1	5	9	13	2	6	10	14	3	7	11	15	4	8	12	16

Key schedule: $K = (k_1, \ldots, k_{32}) \in \{0,1\}^{32}$.
For $1 \leqslant r \leqslant 5$, $K^r = (k_{4r-3}, \ldots, k_{4r+12})$.

An Example SPN (2/2)



Substitution- Permutation Networks (SPNs)

- The design is simple and very efficient, in both hardware and software.
  - In software, S- box $\longrightarrow$ look- up table. Memory= $\ell 2^{\ell}$.
  - In hardware, needs smaller implementation.

In Example: Memory for S- box= $\ell 2^{\ell} = 4 \times 2^4 = 2^6$.
If the S- box would be 16 bits to 16 bits, then Memory= $\ell 2^{\ell} = 16 \times 2^{16} = 2^{20}$.

Substitution- Permutation Networks (SPNs)

- The design is simple and very efficient, in both hardware and software.
  - In software, S- box $\longrightarrow$ look- up table. Memory= $\ell 2^{\ell}$.
  - In hardware, needs smaller implementation.

In Example: Memory for S- box= $\ell 2^{\ell} = 4 \times 2^4 = 2^6$.
If the S- box would be 16 bits to 16 bits, then Memory= $\ell 2^{\ell} = 16 \times 2^{16} = 2^{20}$.

A practical secure SPN would have

- a larger key size
- a larger block length
- larger S- Box
- more rounds

Substitution- Permutation Networks (SPNs)

- The design is simple and very efficient, in both hardware and software.
  - In software, S- box $\longrightarrow$ look- up table. Memory= $\ell 2^\ell$.
  - In hardware, needs smaller implementation.

In Example: Memory for S- box= $\ell 2^\ell = 4 \times 2^4 = 2^6$.
If the S- box would be 16 bits to 16 bits, then Memory= $\ell 2^\ell = 16 \times 2^{16} = 2^{20}$.

A practical secure SPN would have

- a larger key size
- a larger block length
- larger S- Box
- more rounds

$\Longrightarrow$ Advanced Encryption Standard (AES)

Linear Cryptanalysis

Take advantage of high probability occurrences of linear expressions involving

- plaintext bits
- ciphertext bits
- subkey bits

a known plaintext attack: that is, it is premised on the attacker having information on a set of plaintexts and the corresponding ciphertexts.

Linear Cryptanalysis

The basic idea is to approximate the operation of a portion of the cipher with an expression that is linear where the linearity refers to a mod 2 bit wise addition.

$$\mathbf{Pr}\left[X_{i_1} \oplus X_{i_2} \oplus \cdots \oplus X_{i_u} \oplus Y_{j_1} \oplus Y_{j_2} \oplus \cdots \oplus Y_{j_v} = 0\right] = p_L \qquad (1)$$

The approach in linear cryptanalysis is to determine expressions of the form above which have a high or low probability of occurrence.

If a cipher displays a tendency for equation (1) to hold with high probability or not hold with high probability, this is evidence of the cipher's poor randomization abilities.

Linear probability bias: the amount by which the probability of a linear expression holding deviates from $\frac{1}{2}$.

The Pilling- up Lemma

$X_1, X_2, \ldots$: independent binary variables, hence $X_i = 0 \text{ or } 1$.

$\mathbf{Pr}[X_i = 0] = p_i$ and $\mathbf{Pr}[X_i = 1] = 1 - p_i$.

$i \neq j \rightarrow$ The independence of X_i and X_j implies that

$\mathbf{Pr}[X_i = 0, X_j = 0] = p_i p_j$

$\mathbf{Pr}[X_i = 0, X_j = 1] = p_i (1 - p_j)$

$\mathbf{Pr}[X_i = 1, X_j = 0] = (1 - p_i) p_j$

$\mathbf{Pr}[X_i = 1, X_j = 1] = (1 - p_i)(1 - p_j)$.

The Pilling- up Lemma

$X_i \oplus X_j = 0 \Rightarrow X_i = X_j$: linear expression

$\mathbf{Pr}\left[X_i \oplus X_j = 0\right] = p_i p_j + \left(1 - p_i\right)\left(1 - p_j\right)$

The Pilling- up Lemma

$X_i \oplus X_j = 0 \Rightarrow X_i = X_j$: linear expression

$$\mathbf{Pr}\left[X_i \oplus X_j = 0\right] = p_i p_j + (1 - p_i)(1 - p_j)$$

$X_i \oplus X_j = 1 \Rightarrow X_i \neq X_j$: affine expression

$$\mathbf{Pr}\left[X_i \oplus X_j = 1\right] = p_i (1 - p_j) + (1 - p_i)\, p_j$$

The Pilling- up Lemma

$X_i \oplus X_j = 0 \Rightarrow X_i = X_j$: linear expression

$\mathbf{Pr}\left[X_i \oplus X_j = 0\right] = p_i p_j + (1 - p_i)(1 - p_j)$

$X_i \oplus X_j = 1 \Rightarrow X_i \neq X_j$: affine expression

$\mathbf{Pr}\left[X_i \oplus X_j = 1\right] = p_i (1 - p_j) + (1 - p_i) p_j$

The *bias* of X_i: $\epsilon_i = p_i - \frac{1}{2}$.

- $-\frac{1}{2} \leq \epsilon_i \leq \frac{1}{2}$
- $\mathbf{Pr}\left[X_i = 0\right] = \frac{1}{2} + \epsilon_i$
- $\mathbf{Pr}\left[X_i = 1\right] = \frac{1}{2} - \epsilon_i$

The Pilling- up Lemma

Let $\epsilon_{1,2,\ldots,k}$ denote the bias of the variable $X_1 \oplus \cdots \oplus X_k$.

The Pilling- up Lemma

Let $\epsilon_{1,2,\ldots,k}$ denote the bias of the variable $X_1 \oplus \cdots \oplus X_k$.

$$
\begin{aligned}
\mathbf{Pr}\left[X_i \oplus X_j = 0\right] &= p_{i,j} \\
&= \frac{1}{2} + \epsilon_{i,j} \\
&= \left(\frac{1}{2} + \epsilon_i\right)\left(\frac{1}{2} + \epsilon_j\right) + \left(\frac{1}{2} - \epsilon_i\right)\left(\frac{1}{2} - \epsilon_j\right) \\
&= \frac{1}{2} + 2\epsilon_i\epsilon_j
\end{aligned}
$$

The Pilling- up Lemma

Let $\epsilon_{1,2,\ldots,k}$ denote the bias of the variable $X_1 \oplus \cdots \oplus X_k$.

$$\begin{aligned}
\mathbf{Pr}\left[X_i \oplus X_j = 0\right] &= p_{i,j} \\
&= \frac{1}{2} + \epsilon_{i,j} \\
&= \left(\frac{1}{2} + \epsilon_i\right)\left(\frac{1}{2} + \epsilon_j\right) + \left(\frac{1}{2} - \epsilon_i\right)\left(\frac{1}{2} - \epsilon_j\right) \\
&= \frac{1}{2} + 2\epsilon_i\epsilon_j
\end{aligned}$$

LEMMA 3.1 (Pilling- up lemma) Let $\epsilon_{1,2,\ldots,k}$ denote the bias of the variable $X_1 \oplus \cdots \oplus X_k$. Then

$$\epsilon_{1,2,\ldots,k} = 2^{k-1}\Pi_{j=1}^{k}\epsilon_j$$

.

The Pilling- up Lemma

Let $\epsilon_{1,2,\ldots,k}$ denote the bias of the variable $X_1 \oplus \cdots \oplus X_k$.

$$\begin{aligned}
\mathbf{Pr}\left[X_i \oplus X_j = 0\right] &= p_{i,j} \\
&= \frac{1}{2} + \epsilon_{i,j} \\
&= \left(\frac{1}{2} + \epsilon_i\right)\left(\frac{1}{2} + \epsilon_j\right) + \left(\frac{1}{2} - \epsilon_i\right)\left(\frac{1}{2} - \epsilon_j\right) \\
&= \frac{1}{2} + 2\epsilon_i\epsilon_j
\end{aligned}$$

LEMMA 3.1 (Pilling- up lemma) Let $\epsilon_{1,2,\ldots,k}$ denote the bias of the variable $X_1 \oplus \cdots \oplus X_k$. Then

$$\epsilon_{1,2,\ldots,k} = 2^{k-1}\Pi_{j=1}^{k}\epsilon_j$$

.

CORROLLARY 3.2 Suppose that $\epsilon_j = 0$ for some j. Then $\epsilon_{1,2,\ldots,k} = 0$.

Concatenation of Linear Expressions

Consider four independent binary variables, X_1, X_2 and X_3.

Let $\mathbf{Pr}\left[X_1 \oplus X_2 = 0\right] = \frac{1}{2} + \epsilon_{1,2}$ and $\mathbf{Pr}\left[X_2 \oplus X_3 = 0\right] = \frac{1}{2} + \epsilon_{2,3}$.

Concatenation of Linear Expressions

Consider four independent binary variables, X_1, X_2 and X_3.

Let $\mathbf{Pr}\left[X_1 \oplus X_2 = 0\right] = \frac{1}{2} + \epsilon_{1,2}$ and $\mathbf{Pr}\left[X_2 \oplus X_3 = 0\right] = \frac{1}{2} + \epsilon_{2,3}$.

$$\begin{aligned}\mathbf{Pr}\left[(X_1 \oplus X_2) \oplus (X_2 \oplus X_3) = 0\right] &= \frac{1}{2} + 2\epsilon_{1,2}\epsilon_{2,3} \\ \mathbf{Pr}\left[X_1 \oplus X_3 = 0\right] &= \frac{1}{2} + \epsilon_{1,3}\end{aligned}$$

Concatenation of Linear Expressions

Consider four independent binary variables, X_1, X_2 and X_3.

Let $\mathbf{Pr}\left[X_1 \oplus X_2 = 0\right] = \frac{1}{2} + \epsilon_{1,2}$ and $\mathbf{Pr}\left[X_2 \oplus X_3 = 0\right] = \frac{1}{2} + \epsilon_{2,3}$.

$$\mathbf{Pr}\left[(X_1 \oplus X_2) \oplus (X_2 \oplus X_3) = 0\right] = \frac{1}{2} + 2\epsilon_{1,2}\epsilon_{2,3}$$
$$\mathbf{Pr}\left[X_1 \oplus X_3 = 0\right] = \frac{1}{2} + \epsilon_{1,3}$$

We are combining linear expressions to form a new linear expression.
$\epsilon_{1,3} = 2\epsilon_{1,2}\epsilon_{2,3}$

Concatenation of Linear Expressions

Consider four independent binary variables, X_1, X_2 and X_3.

Let $\mathbf{Pr}[X_1 \oplus X_2 = 0] = \frac{1}{2} + \epsilon_{1,2}$ and $\mathbf{Pr}[X_2 \oplus X_3 = 0] = \frac{1}{2} + \epsilon_{2,3}$.

$$\mathbf{Pr}[(X_1 \oplus X_2) \oplus (X_2 \oplus X_3) = 0] = \frac{1}{2} + 2\epsilon_{1,2}\epsilon_{2,3}$$
$$\mathbf{Pr}[X_1 \oplus X_3 = 0] = \frac{1}{2} + \epsilon_{1,3}$$

We are combining linear expressions to form a new linear expression.
$\epsilon_{1,3} = 2\epsilon_{1,2}\epsilon_{2,3}$

The expression $X_1 \oplus X_2 = 0$ and $X_2 \oplus X_3 = 0$ are analogous to linear approximation of S- boxes and $X_1 \oplus X_3 = 0$ is analogous to a cipher approximation.

How do we construct expressions which are highly linear and hence can be exploited?

This is done by considering the properties of the cipher's only nonlinear component: S-box.

How do we construct expressions which are highly linear and hence can be exploited?

This is done by considering the properties of the cipher's only nonlinear component: S-box.

It is possible to concatenate linear approximations of the S-boxes together so that intermediate bits can be canceled out and we are left with a linear expression which has a large bias and involves only plaintext and the last round input bits.

Linear Approximations of S- boxes

X_1	X_2	X_3	X_4	Y_1	Y_2	Y_3	Y_4	$X_2 \oplus X_3$	$Y_1 \oplus Y_3 \oplus Y_4$	$X_1 \oplus X_4$	Y_2	$X_3 \oplus X_4$	$Y_1 \oplus Y_4$
0	0	0	0	1	1	1	0	0	0	0	1	0	1
0	0	0	1	0	1	0	0	0	0	1	1	1	0
0	0	1	0	1	1	0	1	1	0	0	1	1	0
0	0	1	1	0	0	0	1	1	1	1	0	0	1
0	1	0	0	0	0	1	0	1	1	0	0	0	0
0	1	0	1	1	1	1	1	1	1	1	1	1	0
0	1	1	0	1	0	1	1	0	1	0	0	1	0
0	1	1	1	1	0	0	0	0	1	1	0	0	1
1	0	0	0	0	0	1	1	0	0	1	0	0	1
1	0	0	1	1	0	1	0	0	0	0	0	1	1
1	0	1	0	0	1	1	0	1	1	1	1	1	0
1	0	1	1	1	1	0	0	1	1	0	1	0	1
1	1	0	0	0	1	0	1	1	1	1	1	0	1
1	1	0	1	1	0	0	1	1	0	0	0	1	0
1	1	1	0	0	0	0	0	0	0	1	0	1	0
1	1	1	1	0	1	1	1	0	0	0	1	0	1

For linear expression $X_2 \oplus X_3 \oplus Y_1 \oplus Y_3 \oplus Y_4 = 0$, 12 out of the 16 cases the expression hold true. The probability bias is $\frac{12}{16} - \frac{1}{2} = \frac{1}{4}$.

Linear Approximations of S- boxes

X_1	X_2	X_3	X_4	Y_1	Y_2	Y_3	Y_4	$X_2 \oplus X_3$	$Y_1 \oplus Y_3 \oplus Y_4$	$X_1 \oplus X_4$	Y_2	$X_3 \oplus X_4$	$Y_1 \oplus Y_4$
0	0	0	0	1	1	1	0	0	0	0	1	0	1
0	0	0	1	0	1	0	0	0	0	1	1	1	0
0	0	1	0	1	1	0	1	1	0	0	1	1	0
0	0	1	1	0	0	0	1	1	1	1	0	0	1
0	1	0	0	0	0	1	0	1	1	0	0	0	0
0	1	0	1	1	1	1	1	1	1	1	1	1	0
0	1	1	0	1	0	1	1	0	1	0	0	1	0
0	1	1	1	1	0	0	0	0	1	1	0	0	1
1	0	0	0	0	0	1	1	0	0	1	0	0	1
1	0	0	1	1	0	1	0	0	0	0	0	1	1
1	0	1	0	0	1	1	0	1	1	1	1	1	0
1	0	1	1	1	1	0	0	1	1	0	1	0	1
1	1	0	0	0	1	0	1	1	1	1	1	0	1
1	1	0	1	1	0	0	1	1	0	0	0	1	0
1	1	1	0	0	0	0	0	0	0	1	0	1	0
1	1	1	1	0	1	1	1	0	0	0	1	0	1

For linear expression $X_2 \oplus X_3 \oplus Y_1 \oplus Y_3 \oplus Y_4 = 0$, 12 out of the 16 cases the expression hold true. The probability bias is $\frac{12}{16} - \frac{1}{2} = \frac{1}{4}$.

For $X_1 \oplus X_4 = Y_2$ the probability bias is 0.

Linear Approximations of S- boxes

X_1	X_2	X_3	X_4	Y_1	Y_2	Y_3	Y_4	$X_2 \oplus X_3$	$Y_1 \oplus Y_3 \oplus Y_4$	$X_1 \oplus X_4$	Y_2	$X_3 \oplus X_4$	$Y_1 \oplus Y_4$
0	0	0	0	1	1	1	0	0	0	0	1	0	1
0	0	0	1	0	1	0	0	0	0	1	1	1	0
0	0	1	0	1	1	0	1	1	0	0	1	1	0
0	0	1	1	0	0	0	1	1	1	1	0	0	1
0	1	0	0	0	0	1	0	1	1	0	0	0	0
0	1	0	1	1	1	1	1	1	1	1	1	1	0
0	1	1	0	1	0	1	1	0	1	0	0	1	0
0	1	1	1	1	0	0	0	0	1	1	0	0	1
1	0	0	0	0	0	1	1	0	0	1	0	0	1
1	0	0	1	1	0	1	0	0	0	0	0	1	1
1	0	1	0	0	1	1	0	1	1	1	1	1	0
1	0	1	1	1	1	0	0	1	1	0	1	0	1
1	1	0	0	0	1	0	1	1	1	1	1	0	1
1	1	0	1	1	0	0	1	1	0	0	0	1	0
1	1	1	0	0	0	0	0	0	0	1	0	1	0
1	1	1	1	0	1	1	1	0	0	0	1	0	1

For linear expression $X_2 \oplus X_3 \oplus Y_1 \oplus Y_3 \oplus Y_4 = 0$, 12 out of the 16 cases the expression hold true. The probability bias is $\frac{12}{16} - \frac{1}{2} = \frac{1}{4}$.

For $X_1 \oplus X_4 = Y_2$ the probability bias is 0.

For $X_3 \oplus X_4 = Y_1 \oplus Y_4$ the probability bias is $\frac{2}{16} - \frac{1}{2} = -\frac{3}{8}$.

Linear Approximation Table

$a_1X_1 \oplus a_2X_2 \oplus a_3X_3 \oplus a_4X_4 \oplus b_1Y_1 \oplus b_2Y_2 \oplus b_3Y_3 \oplus b_4Y_4 = 0$

If $A = 1101_2 = D_{16}$ and $B = 0101_2 = 5_{16}$, then
$X_1 \oplus X_2 \oplus X_4 \oplus Y_2 \oplus Y_4 = 0$.

Linear Approximation Table

$$a_1X_1 \oplus a_2X_2 \oplus a_3X_3 \oplus a_4X_4 \oplus b_1Y_1 \oplus b_2Y_2 \oplus b_3Y_3 \oplus b_4Y_4 = 0$$

If $A = 1101_2 = D_{16}$ and $B = 0101_2 = 5_{16}$, then

$X_1 \oplus X_2 \oplus X_4 \oplus Y_2 \oplus Y_4 = 0.$

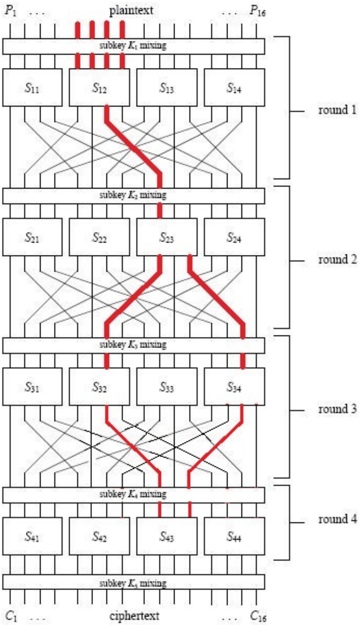
Input Sum	Output Sum 0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	+8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	−2	−2	0	0	−2	+6	+2	+2	0	0	+2	+2	0	0
2	0	0	−2	−2	0	0	−2	−2	0	0	+2	+2	0	0	−6	+2
3	0	0	0	0	0	0	0	0	+2	−6	−2	−2	+2	+2	−2	−2
4	0	+2	0	−2	−2	−4	−2	0	0	−2	0	+2	+2	−4	+2	0
5	0	−2	−2	0	−2	0	+4	+2	−2	0	−4	+2	0	−2	−2	0
6	0	+2	−2	+4	+2	0	0	+2	0	−2	+2	+4	−2	0	0	−2
7	0	−2	0	+2	+2	−4	+2	0	−2	0	+2	0	+4	+2	0	+2
8	0	0	0	0	0	0	0	0	−2	+2	+2	−2	+2	−2	−2	−6
9	0	0	−2	−2	0	0	−2	−2	−4	0	−2	+2	0	+4	+2	−2
A	0	+4	−2	+2	−4	0	+2	−2	+2	+2	0	0	+2	+2	0	0
B	0	+4	0	−4	+4	0	+4	0	0	0	0	0	0	0	0	0
C	0	−2	+4	−2	−2	0	+2	0	+2	0	+2	+4	0	+2	0	−2
D	0	+2	+2	0	−2	+4	0	+2	−4	−2	+2	0	+2	0	0	+2
E	0	+2	+2	0	−2	−4	0	+2	−2	0	0	−2	−4	+2	−2	0
F	0	−2	−4	−2	−2	0	+2	0	0	−2	+4	−2	−2	0	+2	0

Properties of the Table

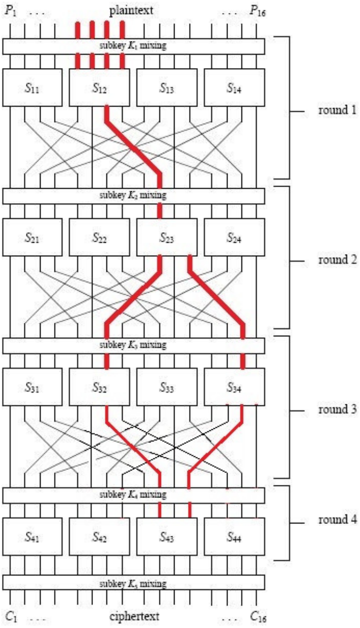
- Each element in the table represents the number of matches between the linear equation represented in hexadecimal as "Input Sum" and the sum of the output bits represented in hexadecimal as "Output Sum" minus 8.
  Example: Input Sum=A and Output Sum=6 then expression that is considered is $X_1 \oplus X_3 \oplus Y_2 \oplus Y_3 = 0$
- Hence, dividing an element value by 16 gives the probability bias for the particular linear combination of input and output bits.
- The linear combination involving no output bits (column 0) will always equal the linear combination of no input bits (row 0) resulting in a bias of $+1/2$ and a table value of $+8$ in the top left corner.
- The sum of any row or any column must be either $+8$ or - 8.

Constructing Linear Approximations for the Complete Cipher

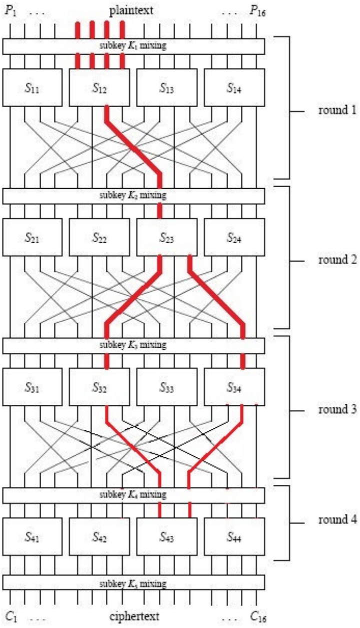
- Once the linear approximation information has been compiled for the S-boxes in an SPN, we have the data to proceed with determining linear approximations of the overall cipher of the form of equation (1).
- This can be achieved by concatenating appropriate linear approximations of S-boxes.
- By constructing a linear approximation involving plaintext bits and data bits from the output of the second last round of S-boxes, it is possible to attack the cipher by recovering a subset of the subkey bits that follow the last round.



- We would like to use as less S-Boxes as possible. The S-Boxes that are used are called active S-Boxes.
- The permutation layer distributes all the outputs of one S-Box to different S-Boxes at the next round. Hence, the best choice is to use just one output bit of an S-Box.

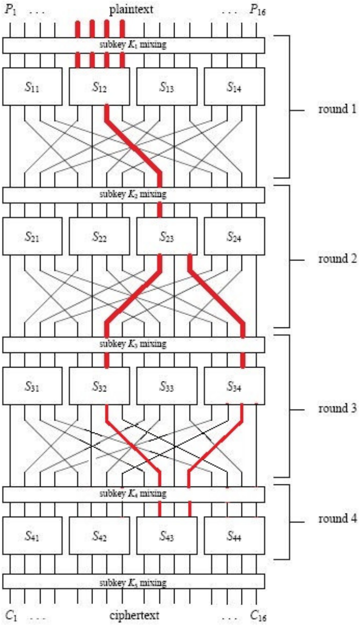


- We would like to use as less S-Boxes as possible. The S-Boxes that are used are called active S-Boxes.
- The permutation layer distributes all the outputs of one S-Box to different S-Boxes at the next round. Hence, the best choice is to use just one output bit of an S-Box.
- We consider S_{12} for the first round.

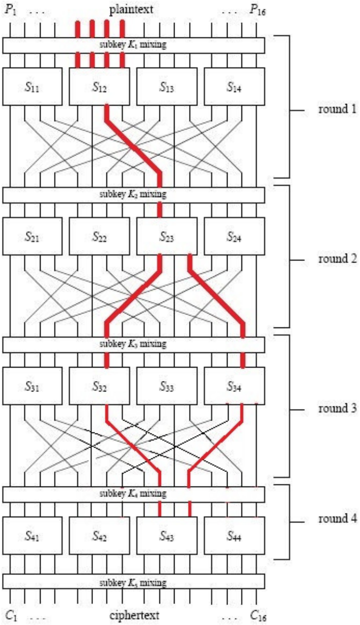


- We would like to use as less S-Boxes as possible. The S-Boxes that are used are called active S-Boxes.
- The permutation layer distributes all the outputs of one S-Box to different S-Boxes at the next round. Hence, the best choice is to use just one output bit of an S-Box.
- We consider S_{12} for the first round.
- Only Y_1, Y_2, Y_3 or Y_4 should be involved in the expression that is used for S_{12}.
- The choices are as follows

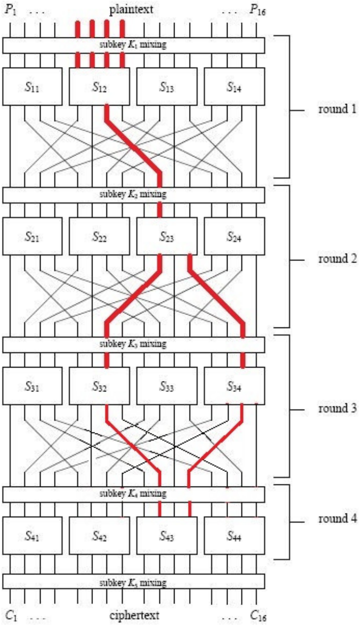
	1	2	4	8
F	-2	-4	-2	0



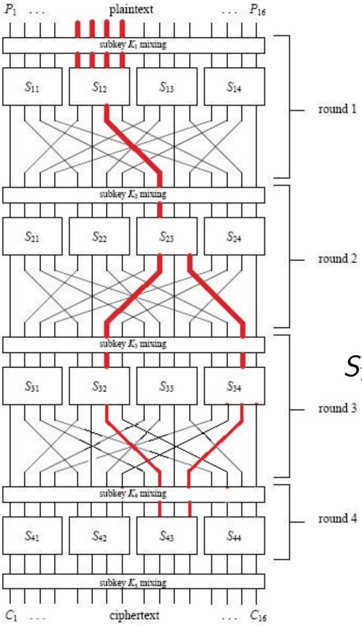
- U_i (V_i) represent the 16-bit block of bits at the input (output) of the round i S-boxes.



- U_i (V_i) represent the 16-bit block of bits at the input (output) of the round i S-boxes.
- $U_{i,j}$ ($V_{i,j}$) represent the j-th bit of block U_i (V_i).

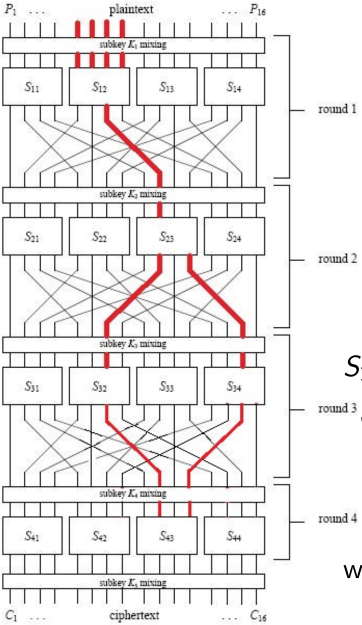


- U_i (V_i) represent the 16-bit block of bits at the input (output) of the round i S-boxes.
- $U_{i,j}$ ($V_{i,j}$) represent the j-th bit of block U_i (V_i).
- K_i represents the subkey block of bits XORed at the input to round i.



- U_i (V_i) represent the 16-bit block of bits at the input (output) of the round i S-boxes.
- $U_{i,j}$ ($V_{i,j}$) represent the j-th bit of block U_i (V_i).
- K_i represents the subkey block of bits XORed at the input to round i.

$$S_{12}\colon\ X_1 \oplus X_2 \oplus X_3 \oplus X_4 = Y_3$$

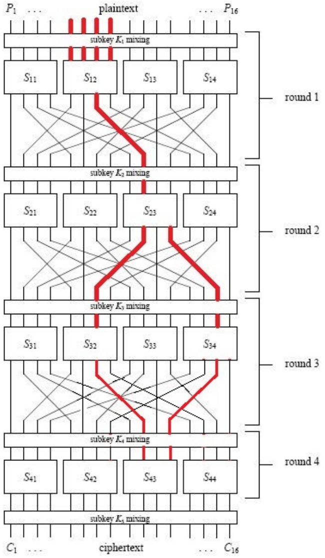


- U_i (V_i) represent the 16-bit block of bits at the input (output) of the round i S-boxes.
- $U_{i,j}$ ($V_{i,j}$) represent the j-th bit of block U_i (V_i).
- K_i represents the subkey block of bits XORed at the input to round i.

$$S_{12}\colon\ X_1 \oplus X_2 \oplus X_3 \oplus X_4 = Y_3$$

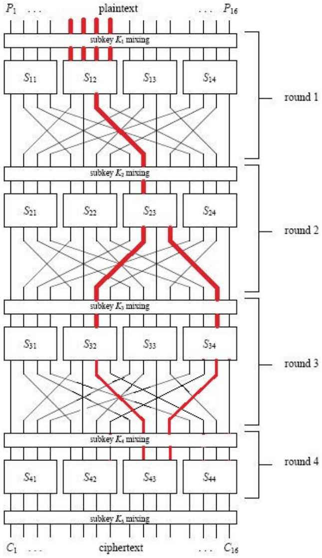
$$\begin{aligned} V_{1,7} &= U_{1,5} \oplus U_{1,6} \oplus U_{1,7} \oplus U_{1,8} \\ &= P_5 \oplus K_{1,5} \oplus P_6 \oplus K_{1,6} \oplus P_7 \\ &\quad \oplus K_{1,7} \oplus P_8 \oplus K_{1,8} \end{aligned}$$

with bias $-\frac{1}{4}$.



The choices for S_{23} are as follows:

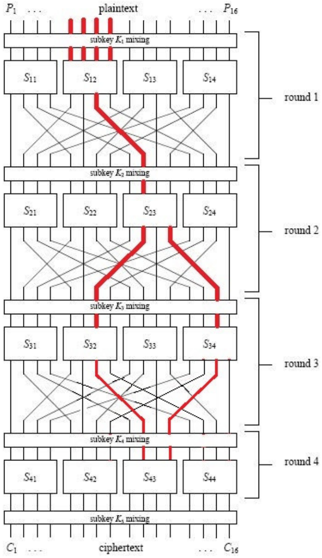
	5	D
4	-4	-4



The choices for S_{23} are as follows:

	5	D
4	-4	-4

S_{23}: $X_2 = Y_2 \oplus Y_4$



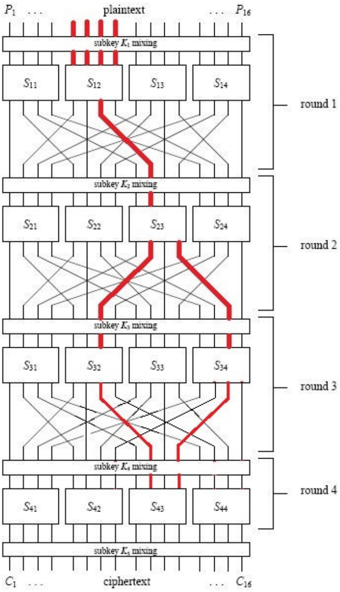
The choices for S_{23} are as follows:

	5	D
4	-4	-4

S_{23}: $X_2 = Y_2 \oplus Y_4$

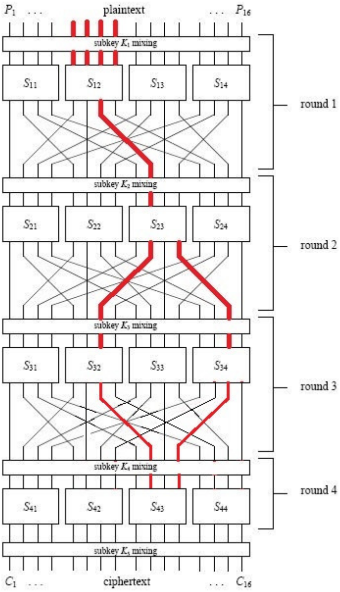
$$\begin{aligned} V_{2,10} \oplus V_{2,12} &= U_{2,10} \\ &= V_{1,7} \oplus K_{2,10} \\ &= P_5 \oplus K_{1,5} \oplus P_6 \oplus K_{1,6} \\ &\oplus P_7 \oplus K_{1,7} \oplus P_8 \oplus K_{1,8} \\ &\oplus K_{2,10} \end{aligned}$$

with bias $-\frac{1}{4}$.



The choices for S_{32} and S_{34} are as follows:

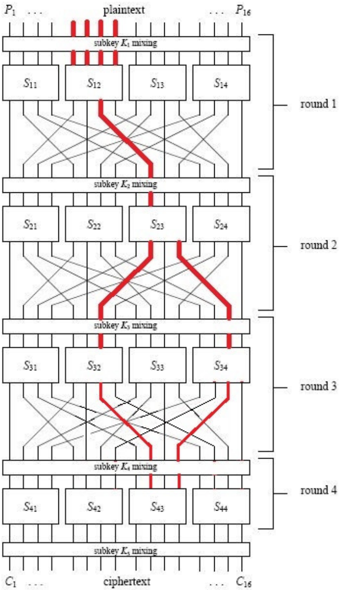
	2	3	6	A
2	-2	-2	-2	+2



The choices for S_{32} and S_{34} are as follows:

	2	3	6	A
2	-2	-2	-2	+2

S_{32}: $X_3 = Y_3$



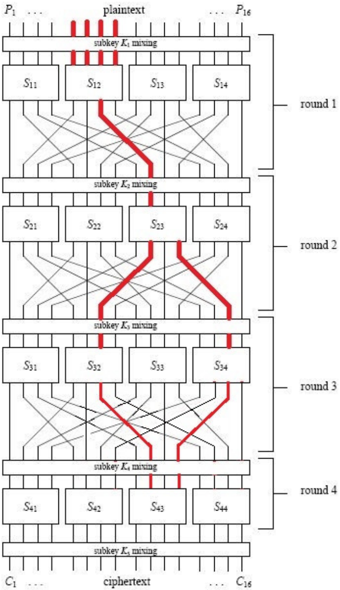
The choices for S_{32} and S_{34} are as follows:

	2	3	6	A
2	-2	-2	-2	+2

S_{32}: $X_3 = Y_3$

$V_{3,7} = U_{3,7} = V_{2,10} \oplus K_{3,7}$

with bias $-\frac{1}{8}$.



The choices for S_{32} and S_{34} are as follows:

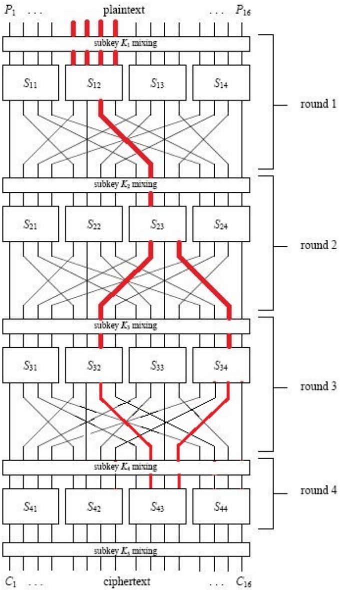
	2	3	6	A
2	-2	-2	-2	+2

S_{32}: $X_3 = Y_3$

$V_{3,7} = U_{3,7} = V_{2,10} \oplus K_{3,7}$

with bias $-\frac{1}{8}$.

S_{34}: $X_3 = Y_3$



The choices for S_{32} and S_{34} are as follows:

	2	3	6	A
2	-2	-2	-2	+2

S_{32}: $X_3 = Y_3$

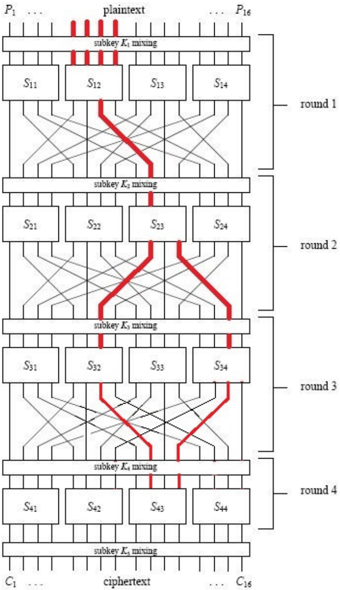
$V_{3,7} = U_{3,7} = V_{2,10} \oplus K_{3,7}$

with bias $-\frac{1}{8}$.

S_{34}: $X_3 = Y_3$

$V_{3,15} = U_{3,15} = V_{2,12} \oplus K_{3,15}$

with bias $-\frac{1}{8}$.



The choices for S_{32} and S_{34} are as follows:

	2	3	6	A
2	-2	-2	-2	+2

S_{32}: $X_3 = Y_3$

$V_{3,7} = U_{3,7} = V_{2,10} \oplus K_{3,7}$

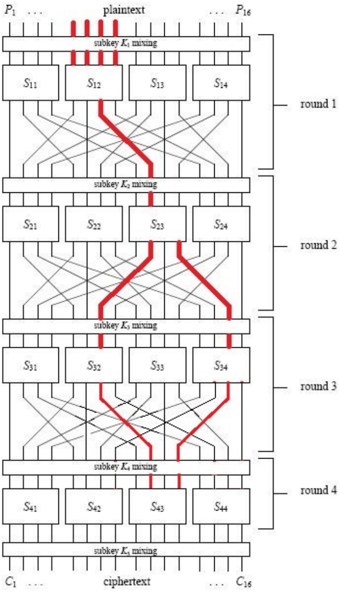
with bias $-\frac{1}{8}$.

S_{34}: $X_3 = Y_3$

$V_{3,15} = U_{3,15} = V_{2,12} \oplus K_{3,15}$

with bias $-\frac{1}{8}$.

$U_{4,10} \oplus K_{4,10} \oplus U_{4,12} \oplus K_{4,12} = V_{2,10} \oplus V_{2,12} \oplus K_{3,7} \oplus K_{3,15}$



The choices for S_{32} and S_{34} are as follows:

	2	3	6	A
2	-2	-2	-2	+2

S_{32}: $X_3 = Y_3$

$V_{3,7} = U_{3,7} = V_{2,10} \oplus K_{3,7}$

with bias $-\frac{1}{8}$.

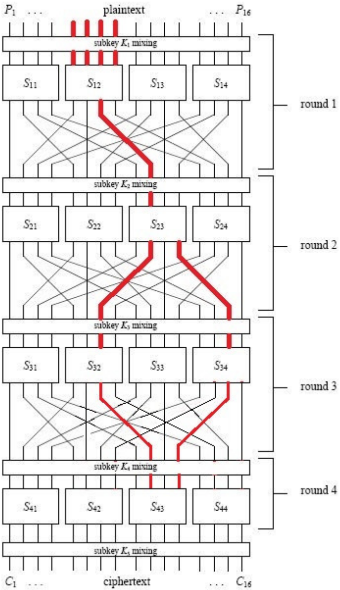
S_{34}: $X_3 = Y_3$

$V_{3,15} = U_{3,15} = V_{2,12} \oplus K_{3,15}$

with bias $-\frac{1}{8}$.

$U_{4,10} \oplus K_{4,10} \oplus U_{4,12} \oplus K_{4,12} = V_{2,10} \oplus V_{2,12} \oplus K_{3,7} \oplus K_{3,15}$

$P_5 \oplus P_6 \oplus P_7 \oplus P_8 \oplus U_{4,10} \oplus U_{4,12} \oplus K_{1,5} \oplus K_{1,6} \oplus K_{1,7} \oplus K_{1,8} \oplus K_{2,10} \oplus K_{3,7} \oplus K_{3,15} \oplus K_{4,10} \oplus K_{4,12} = 0$



The choices for S_{32} and S_{34} are as follows:

	2	3	6	A
2	-2	-2	-2	+2

S_{32}: $X_3 = Y_3$
$V_{3,7} = U_{3,7} = V_{2,10} \oplus K_{3,7}$
with bias $-\frac{1}{8}$.
S_{34}: $X_3 = Y_3$
$V_{3,15} = U_{3,15} = V_{2,12} \oplus K_{3,15}$
with bias $-\frac{1}{8}$.
$U_{4,10} \oplus K_{4,10} \oplus U_{4,12} \oplus K_{4,12} = V_{2,10} \oplus V_{2,12} \oplus K_{3,7} \oplus K_{3,15}$
$P_5 \oplus P_6 \oplus P_7 \oplus P_8 \oplus U_{4,10} \oplus U_{4,12} \oplus K_{1,5} \oplus K_{1,6} \oplus K_{1,7} \oplus K_{1,8} \oplus K_{2,10} \oplus K_{3,7} \oplus K_{3,15} \oplus K_{4,10} \oplus K_{4,12} = 0$
by application of the Piling- Up Lemma bias is $2^3 \times -\frac{1}{4} \times -\frac{1}{4} \times -\frac{1}{8} \times -\frac{1}{8} = \frac{1}{128}$.

$P_5 \oplus P_6 \oplus P_7 \oplus P_8 \oplus U_{4,10} \oplus U_{4,12} \oplus \sum K = 0$
where
$\sum K = K_{1,5} \oplus K_{1,6} \oplus K_{1,7} \oplus K_{1,8} \oplus K_{2,10} \oplus K_{3,7} \oplus K_{3,12} \oplus K_{4,10} \oplus K_{4,12}$
and $\sum K$ is fixed at either 0 or 1 depending on the key of the cipher.

- Now since $\sum K$ is fixed, we note that
  $P_5 \oplus P_6 \oplus P_7 \oplus P_8 \oplus U_{4,10} \oplus U_{4,12} = 0$
  must hold with a probability of either $p_L = \frac{1}{2} - \frac{1}{128} = \frac{63}{128}$ or $1 - \frac{63}{128} = \frac{65}{128}$, depending on whether $\sum K = 0$ or 1, respectively.
- In other words, we now have a linear approximation of the first three rounds of the cipher with a bias of magnitude $\frac{1}{128}$.
- We must now discuss how such a bias can be used to determine some of the key bits.

Meaning of the p_L

- Linear expression implicitly has subkey bits involved. If the sum of the involved subkey bits is "0", the bias will have the same sign as the bias of the expression involving the subkey sum and if the sum of the involved subkey bits is "1", the bias will have the opposite sign as the bias of the expression involving the subkey sum
- $p_L = 1$ implies that linear expression is a perfect representation of the cipher behavior and the cipher has a catastrophic weakness.
- $p_L = 0$, then linear expression represents an affine relationship.

Steps of the Linear Cryptanalysis 1/2

1. Suppose that it is possible to find a probabilistic linear relationship between a subset of plaintext bits and a subset of state bits immediately preceding the substitutions performed in the last round.
2. Assume that an attacker has a large number of plaintext-ciphertext pairs, all of which are encrypted using the same unknown key K.
3. Decrypt all the ciphertexts, using all possible candidate keys for the last round of the cipher.
4. For each candidate key, we compute the values of the relevant state bits involved in the linear relationship.
5. Determine if the above mentioned linear relationship holds.
6. Whenever it does, we increment a counter corresponding to the particular candidate key.
7. The candidate key that has a frequency count that is furthest from $1/2$ times the number of pairs contains the correct values for these key bits.

Steps of the Linear Cryptanalysis 2/2

- The linear expression affects the inputs to S-box S_{43} in the last round.
- For each plaintext/ciphertext sample, we would try all 16 values for the target partial subkey $K_{5,9}, K_{5,10}, K_{5,11}, K_{5,12}$.
- We determine the value of $U_{4,10}, U_{4,12}$ by running the data backwards through the target partial subkey and S-box S_{43} for each ciphertext.
- We determine if the linear expression $P_5 \oplus P_6 \oplus P_7 \oplus P_8 \oplus U_{4,10} \oplus U_{4,12} = 0$ holds or not and produce the following table.

	Candidate Key Value															
	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
P_1, C_1	+	+	-	+	-	+	-	-	-	+	-	+	+	-	-	+
P_2, C_2	-	+	+	+	-	-	+	-	-	-	-	-	+	+	-	+
$\dots$																
P_n, C_n	+	-	-	+	+	-	-	+	-	-	+	-	+	-	-	+

The count which deviates the largest from half of the number of plaintext/ciphertext samples is assumed to be the correct value.

Complexity of Attack

- The larger the magnitude of the bias in the S- boxes, the larger the magnitude of the bias of the overall expression.
- The fewer active S- boxes, the larger the magnitude of the overall linear expression bias.
- Let ϵ represent the bias from $1/2$ of the probability that the linear expression for the complete cipher holds. The number of known plaintexts required in the attack is proportional to ϵ^{-2} and, letting N_L represent the number of known plaintexts required, it is reasonable to approximate N_L by $N_L \approx \epsilon^{-2}$.
- In practice, it is generally reasonable to expect some small multiple of ϵ^{-2} known plaintexts are required.

Differential Cryptanalysis

- Differential cryptanalysis exploits the high probability of certain occurrences of plaintext differences and differences into the last round of the cipher.
- A system with input $X = [X_1 X_2 \ldots X_n]$ and output $Y = [Y_1 Y_2 \ldots Y_n]$.
- Let two inputs to the system be X' and X'' with the corresponding outputs Y' and Y'', respectively.
- The input difference is given by $\Delta X = X' \oplus X''$ where $\oplus$ represents a bit-wise exclusive-OR of the n-bit vectors and, hence, $\Delta X = [\Delta X_1 \Delta X_2 \ldots \Delta X_n]$ where $\Delta X_i = X_i' \oplus X_i''$.
- Similarly, $\Delta Y = Y' \oplus Y''$ is the output difference and $\Delta Y = [\Delta Y_1 \Delta Y_2 \ldots \Delta Y_n]$

Differential

- In an ideally randomizing cipher, the probability that a particular output difference ΔY occurs given a particular input difference ΔX is $\frac{1}{2}^n$, where n is the number of bits of X.
- Differential cryptanalysis seeks to exploit a scenario where a particular ΔY occurs given a particular input difference ΔX with a very high probability p_D (i.e., much greater than $\frac{1}{2}^n$).
- The pair $(\Delta X, \Delta Y)$ is referred to as a differential.
- Differential cryptanalysis is a chosen plaintext attack.
- The attacker will select pairs of inputs, X' and X'', to satisfy a particular ΔX, knowing that for that ΔX value, a particular ΔY value occurs with high probability.
- Investigate the construction of a differential $(\Delta X, \Delta Y)$.

Differential Characteristic

- Examine high likely differential characteristics where a differential characteristic is a sequence of input and output differences to the rounds so that the output difference from one round corresponds to the input difference for the next round.
- Using the highly likely differential characteristic gives us the opportunity to exploit information coming into the last round of the cipher to derive bits from the last layer of subkeys.

Example 1 for Differentials

$X, Y \in 0, 1, \ldots, 4$

$Y = f(X) = 3X + 2 \bmod 5,\ \Delta X = X' - X'' \bmod 5$

		ΔY				
X'	Y'	$\Delta X = 0$	$\Delta X = 1$	$\Delta X = 2$	$\Delta X = 3$	$\Delta X = 4$
0	2	0	3	1	4	2
1	0	0	3	1	4	2
2	3	0	3	1	4	2
3	1	0	3	1	4	2
4	4	0	3	1	4	2

		ΔY				
		0	**1**	**2**	**3**	**4**
	0	5	0	0	0	0
	1	0	0	0	5	0
ΔX	**2**	0	5	0	0	0
	3	0	0	0	0	5
	4	0	0	5	0	0

$Pr(\Delta Y = 1 | \Delta X = 2) = 1,\ Pr(\Delta Y = 2 | \Delta X = 2) = 0$

Example 2 for Differentials

$X, Y \in 0, 1, \ldots, 4$

$Y = f(X) = X^2 \bmod 5,\ \Delta X = X' - X'' \bmod 5$

		ΔY				
X'	Y'	$\Delta X = 0$	$\Delta X = 1$	$\Delta X = 2$	$\Delta X = 3$	$\Delta X = 4$
0	0	0	4	1	1	4
1	1	0	1	0	2	2
2	4	0	3	4	3	0
3	4	0	0	3	4	3
4	1	0	2	2	0	1

		ΔY				
		0	**1**	**2**	**3**	**4**
	0	5	0	0	0	0
	1	1	1	1	1	1
ΔX	**2**	1	1	1	1	1
	3	1	1	1	1	1
	4	1	1	1	1	1

$Pr(\Delta Y = 1 | \Delta X = 2) = \frac{1}{5},\ Pr(\Delta Y = 2 | \Delta X = 2) = \frac{1}{5}$

Differential of S-Box

- Examine the properties of individual S-boxes and use these properties to determine the complete differential characteristic.
- Consider the input and output differences of the S-boxes in order to determine a high probability difference pair.
- Combining S-box difference pairs from round to round so that the nonzero output difference bits from one round correspond to the non-zero input difference bits of the next round, enables us to find a high probability differential consisting of the plaintext difference and the difference of the input to the last round.

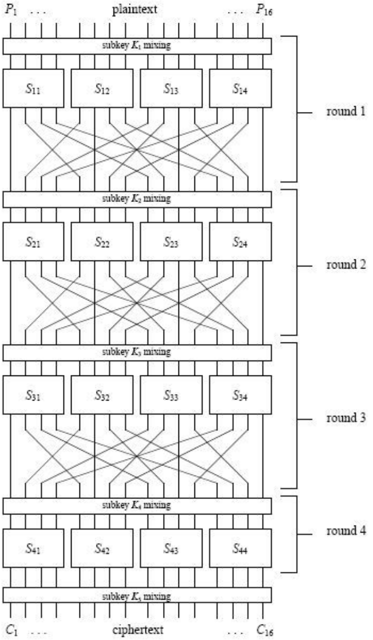
- $Pr(\Delta Y|\Delta X)$ can be derived by considering input pairs (X', X'') such that $X' \oplus X'' = \Delta X$.

z	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
$\pi_S(z)$	E	4	D	1	2	F	B	8	3	A	6	C	5	9	0	7

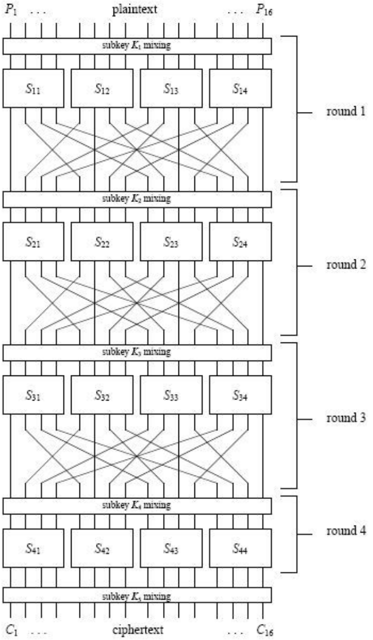
X	Y	ΔY		
		$\Delta X = 1011$	$\Delta X = 1000$	$\Delta X = 0100$
0000	1110	0010	1101	1100
0001	0100	0010	1110	1011
0010	1101	0111	0101	0110
0011	0001	0010	1011	1001
0100	0010	0101	0111	1100
0101	1111	1111	0110	1011
0110	1011	0010	1011	0110
0111	1000	1101	1111	1001
1000	0011	0010	1101	0110
1001	1010	0111	1110	0011
1010	0110	0010	0101	0110
1011	1100	0010	1011	1011
1100	0101	1101	0111	0110
1101	1001	0010	0110	0011
1110	0000	1111	1011	0110
1111	0111	0101	1111	1011

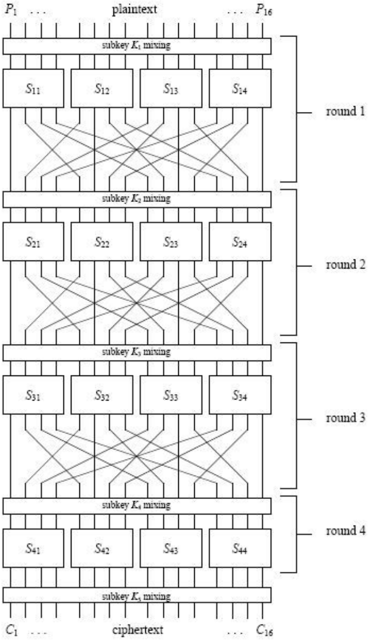
Input Difference	Output Difference															
	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	2	0	0	0	2	0	2	4	0	4	2	0	0
2	0	0	0	2	0	6	2	2	0	2	0	0	0	0	2	0
3	0	0	2	0	2	0	0	0	0	4	2	0	2	0	0	4
4	0	0	0	2	0	0	6	0	0	2	0	4	2	0	0	0
5	0	4	0	0	0	2	2	0	0	0	4	0	2	0	0	2
6	0	0	0	4	0	4	0	0	0	0	0	0	2	2	2	2
7	0	0	2	2	2	0	2	0	0	2	2	0	0	0	0	4
8	0	0	0	0	0	0	2	2	0	0	0	4	0	4	2	2
9	0	2	0	0	2	0	0	4	2	0	2	2	2	0	0	0
A	0	2	2	0	0	0	0	0	6	0	0	2	0	0	4	0
B	0	0	8	0	0	2	0	2	0	0	0	0	0	2	0	2
C	0	2	0	0	2	2	2	0	0	0	0	2	0	6	0	0
D	0	4	0	0	0	0	0	4	2	0	2	0	2	0	2	0
E	0	0	2	4	2	0	0	0	6	0	0	0	0	0	2	0
F	0	2	0	0	6	0	0	0	0	4	0	2	0	0	2	0

S-box difference distribution table


P_1 ... plaintext ... P_{16}
subkey K_1 mixing
S_{11}
S_{12}
S_{13}
S_{14}
round 1
subkey K_2 mixing
S_{21}
S_{22}
S_{23}
S_{24}
round 2
subkey K_3 mixing
S_{31}
S_{32}
S_{33}
S_{34}
round 3
subkey K_4 mixing
S_{41}
S_{42}
S_{43}
S_{44}
round 4
subkey K_5 mixing
C_1 ... ciphertext ... C_{16}

$$\Delta P = [1111\ 0000\ 0000\ 0000]$$
$$\Delta U_1 = [1111\ 0000\ 0000\ 0000]$$





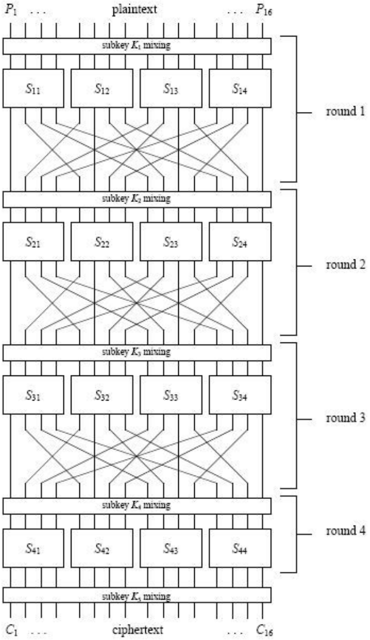
$\Delta P = [1111\,0000\,0000\,0000]$

$\Delta U_1 = [1111\,0000\,0000\,0000]$

	4
F	6

$S_{11} : Pr(\Delta Y = 4 | \Delta X = F) = \frac{6}{16}$

$\Delta V_1 = [0100\,0000\,0000\,0000]$



$\Delta P = [1111\,0000\,0000\,0000]$

$\Delta U_1 = [1111\,0000\,0000\,0000]$

	4
F	6

$S_{11} : Pr(\Delta Y = 4 | \Delta X = F) = \frac{6}{16}$

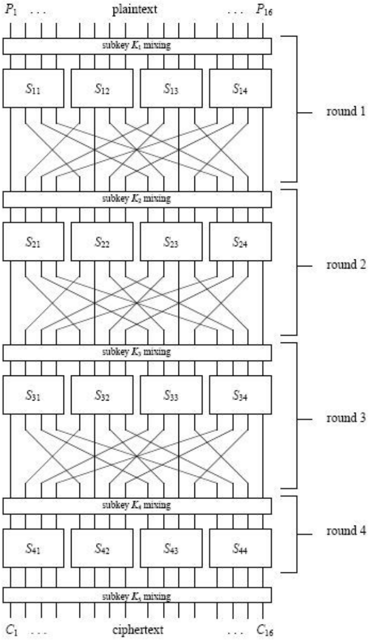
$\Delta V_1 = [0100\,0000\,0000\,0000]$

$\Delta U_2 = [0000\,1000\,0000\,0000]$

	B	D	6	7	E	F
8	4	4	2	2	2	2

$S_{22} : Pr(\Delta Y = 6 | \Delta X = 8) = \frac{2}{16}$

$\Delta V_2 = [0000\,0110\,0000\,0000]$



$\Delta P = [1111\,0000\,0000\,0000]$

$\Delta U_1 = [1111\,0000\,0000\,0000]$

	4
F	6

$S_{11} : Pr(\Delta Y = 4 | \Delta X = F) = \frac{6}{16}$

$\Delta V_1 = [0100\,0000\,0000\,0000]$

$\Delta U_2 = [0000\,1000\,0000\,0000]$

	B	D	6	7	E	F
8	4	4	2	2	2	2

$S_{22} : Pr(\Delta Y = 6 | \Delta X = 8) = \frac{2}{16}$

$\Delta V_2 = [0000\,0110\,0000\,0000]$

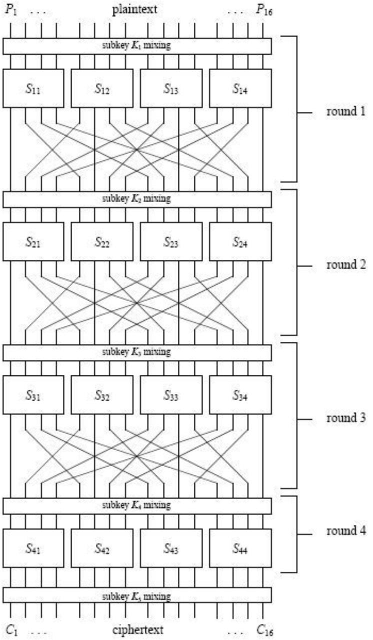
$\Delta U_3 = [0000\,0100\,0100\,0000]$

	6	B	3	9	C
4	6	4	2	2	2

$S_{32} : Pr(\Delta Y = 6 | \Delta X = 4) = \frac{6}{16}$

$S_{33} : Pr(\Delta Y = 6 | \Delta X = 4) = \frac{6}{16}$

$\Delta V_3 = [0000\,0110\,0110\,0000]$



$\Delta P = [1111\,0000\,0000\,0000]$

$\Delta U_1 = [1111\,0000\,0000\,0000]$

	4
F	6

$S_{11} : Pr(\Delta Y = 4 | \Delta X = F) = \frac{6}{16}$

$\Delta V_1 = [0100\,0000\,0000\,0000]$

$\Delta U_2 = [0000\,1000\,0000\,0000]$

	B	D	6	7	E	F
8	4	4	2	2	2	2

$S_{22} : Pr(\Delta Y = 6 | \Delta X = 8) = \frac{2}{16}$

$\Delta V_2 = [0000\,0110\,0000\,0000]$

$\Delta U_3 = [0000\,0100\,0100\,0000]$

	6	B	3	9	C
4	6	4	2	2	2

$S_{32} : Pr(\Delta Y = 6 | \Delta X = 4) = \frac{6}{16}$

$S_{33} : Pr(\Delta Y = 6 | \Delta X = 4) = \frac{6}{16}$

$\Delta V_3 = [0000\,0110\,0110\,0000]$

$\Delta U_4 = [0000\,0110\,0110\,0000]$

We reach to S_{42} S_{43} and stop here

Steps of the Differential Cryptanalysis 1/2

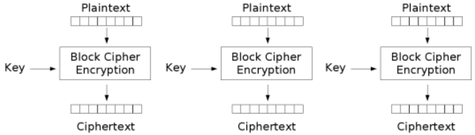
1. Suppose that it is possible to find a differential with high probability like $Pr(\Delta U_4 = [0000\,0110\,0110\,0000] | \Delta P = [1111\,0000\,0000\,0000]) = (\frac{6}{16})^3 \times \frac{2}{16} = \frac{27}{4096}$.
2. Assume that an attacker can choose plaintexts to be encrypted. Assume that the attacker has a large number of P-C and P'-C' pairs, all of which are encrypted using the same unknown key K and plaintext$\oplus$plaintext'$=\Delta P$.
3. Decrypt all the ciphertexts, using all possible candidate keys for the last round of the cipher.
4. For each candidate key, we compute the values of the relevant state bits involved in the differential.
5. Whenever the above mentioned differential holds does, we increment a counter corresponding to the particular candidate key.
6. The candidate key that has highest count is the correct key bits.

Steps of the Differential Cryptanalysis 2/2

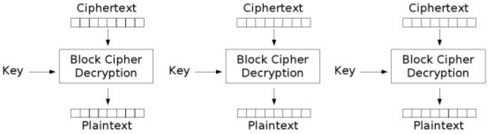
- The differential affects the inputs to S-box S_{42} and S_{43} in the last round.
- For each P-C and P'-C' sample, we would try all 256 values for the target partial subkey $K_{5,5\ldots12}$ and calculate 256 $V_{5,5\ldots12}$ and $V'_{5,5\ldots12}$.
- We determine 256 $U_{4,5\ldots12}$ and $U'_{4,,5\ldots12}$ by running Vs and V's backwards through S-boxes S_{42} and S_{43}.
- We determine if $\Delta U = U_{4,5\ldots12} \oplus U'_{4,5\ldots12} = 0110\,0110$ or not and produce the count table.
- The column which includes the highest number of $+$ is the right key value.

Modes of operation

For messages exceeding block length, n, the message is partitioned into n-bit blocks.
E_K: the encryption function
E_K^{-1}: the decryption function
$x = x_1, \ldots, x_t$: A plaintext message



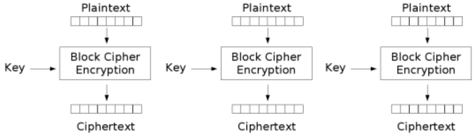
Electronic Codebook (ECB) mode encryption



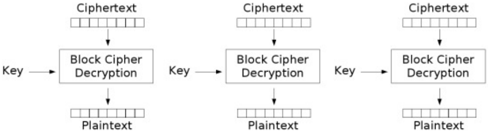
Electronic Codebook (ECB) mode decryption

Modes of operation

For messages exceeding block length, n, the message is partitioned into n-bit blocks.
E_K: the encryption function
E_K^{-1}: the decryption function
$x = x_1, \ldots, x_t$: A plaintext message



Electronic Codebook (ECB) mode encryption



Electronic Codebook (ECB) mode decryption

1. Identical plaintext blocks result in identical ciphertext.

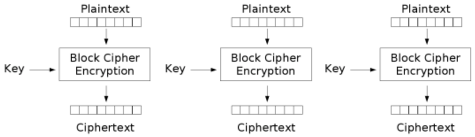
Modes of operation

For messages exceeding block length, n, the message is partitioned into n-bit blocks.

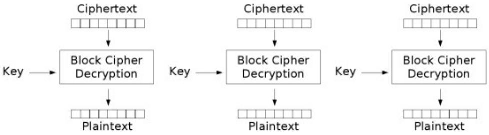
E_K: the encryption function

E_K^{-1}: the decryption function

$x = x_1, \ldots, x_t$: A plaintext message



Electronic Codebook (ECB) mode encryption

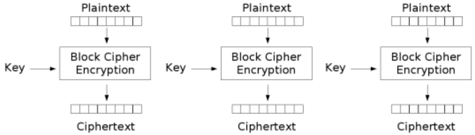


Electronic Codebook (ECB) mode decryption

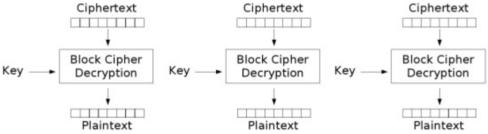
1. Identical plaintext blocks result in identical ciphertext.
2. Reordering ciphertext blocks results in correspondingly re-ordered plaintext blocks.

Modes of operation

For messages exceeding block length, n, the message is partitioned into n-bit blocks.
E_K: the encryption function
E_K^{-1}: the decryption function
$x = x_1, \ldots, x_t$: A plaintext message



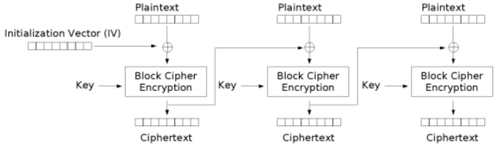
Electronic Codebook (ECB) mode encryption



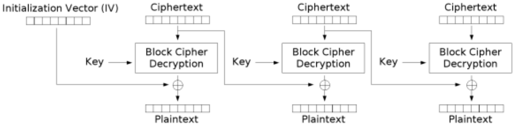
Electronic Codebook (ECB) mode decryption

1. Identical plaintext blocks result in identical ciphertext.
2. Reordering ciphertext blocks results in correspondingly re-ordered plaintext blocks.
3. One or more bit errors in a single ciphertext block affect decipherment of that block only.

Cipher Block Chaining Mode

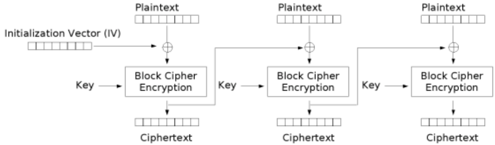


Cipher Block Chaining (CBC) mode encryption

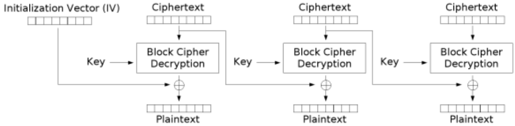


Cipher Block Chaining (CBC) mode decryption

Cipher Block Chaining Mode



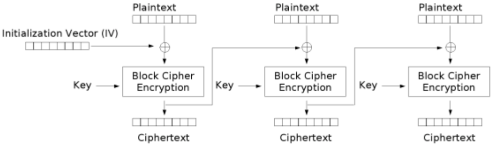
Cipher Block Chaining (CBC) mode encryption



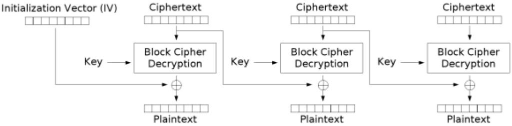
Cipher Block Chaining (CBC) mode decryption

1. Identical ciphertext blocks do not result.

Cipher Block Chaining Mode



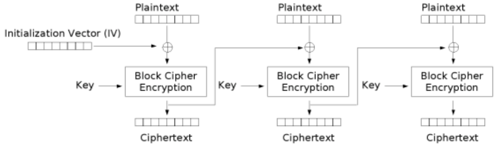
Cipher Block Chaining (CBC) mode encryption



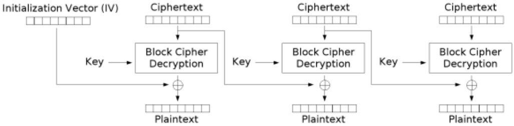
Cipher Block Chaining (CBC) mode decryption

1. Identical ciphertext blocks do not result.
2. Ciphertext c_j depends on x_j and all preceding plaintext blocks.

Cipher Block Chaining Mode



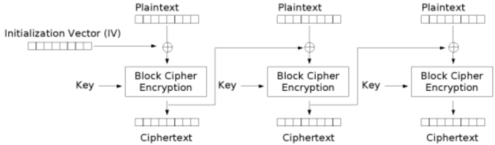
Cipher Block Chaining (CBC) mode encryption



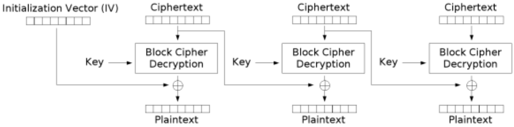
Cipher Block Chaining (CBC) mode decryption

1. Identical ciphertext blocks do not result.
2. Ciphertext c_j depends on x_j and all preceding plaintext blocks.
3. A single bit error in ciphertext block c_j affects decipherment of blocks c_j and c_{j+1}.

Cipher Block Chaining Mode



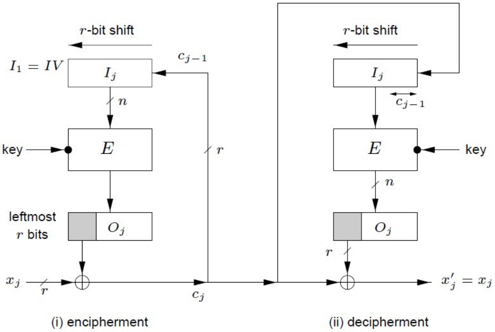
Cipher Block Chaining (CBC) mode encryption



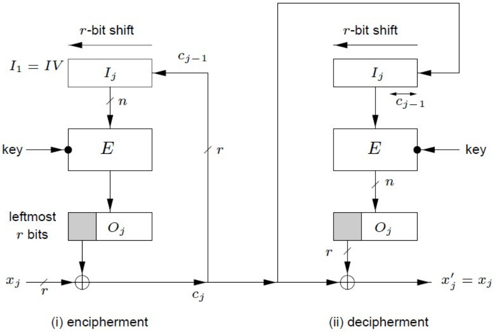
Cipher Block Chaining (CBC) mode decryption

1. Identical ciphertext blocks do not result.
2. Ciphertext c_j depends on x_j and all preceding plaintext blocks.
3. A single bit error in ciphertext block c_j affects decipherment of blocks c_j and c_{j+1}.
4. The CBC mode is self-synchronizing or ciphertext autokey.

Cipher feedback (CFB) Mode

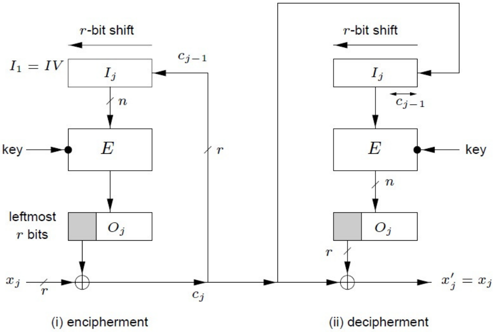


Cipher feedback (CFB) Mode



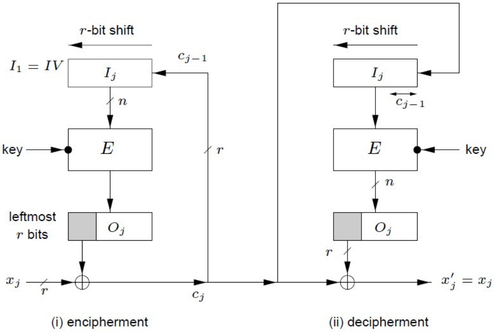
1. Identical ciphertext blocks do not result.

Cipher feedback (CFB) Mode



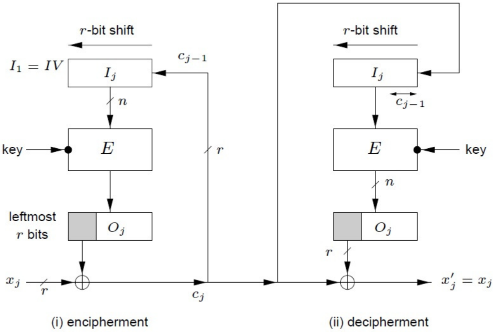
1. Identical ciphertext blocks do not result.
2. Ciphertext block c_j depends on both x_j and preceding plaintext blocks.

Cipher feedback (CFB) Mode



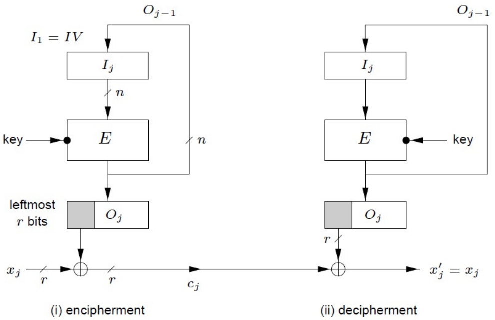
1. Identical ciphertext blocks do not result.
2. Ciphertext block c_j depends on both x_j and preceding plaintext blocks.
3. One or more bit errors in any single r-bit ciphertext block c_j affects the decipherment of that and the next $\lceil \frac{n}{r} \rceil$ ciphertext blocks. Self - synchronizing, but requires $\lceil \frac{n}{r} \rceil$ ciphertext blocks to recover.

Cipher feedback (CFB) Mode



1. Identical ciphertext blocks do not result.
2. Ciphertext block c_j depends on both x_j and preceding plaintext blocks.
3. One or more bit errors in any single r-bit ciphertext block c_j affects the decipherment of that and the next $\lceil \frac{n}{r} \rceil$ ciphertext blocks. Self - synchronizing, but requires $\lceil \frac{n}{r} \rceil$ ciphertext blocks to recover.
4. Throughput: for $r < n$, throughput is decreased by a factor of $\frac{n}{r}$.

Output feedback (OFB) Mode



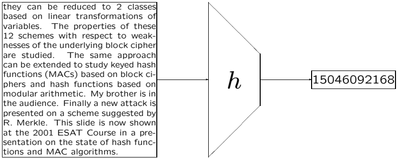
1. Identical ciphertext blocks do not result.
2. The keystream is plaintext-independent.
3. One or more bit errors in any ciphertext character c_j affects the decipherment of only that character.
4. The OFB mode recovers from ciphertext bit errors.
5. Throughput: for $r < n$, throughput is decreased. Since the keystream is independent of plaintext or ciphertext, it may be pre-computed.

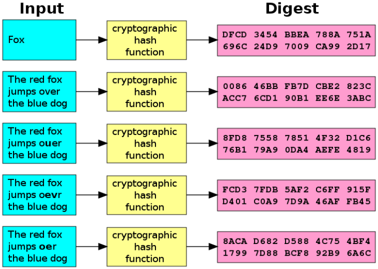
Data Integrity and Source Authentication

- Encryption does not protect data from modification by another party.
- Need a way to ensure that data arrives at destination in its original form as sent by the sender and it is coming from an authenticated source.

Hash Functions

Definition - hash function





A cryptographic hash function is a deterministic procedure that takes an arbitrary block of data and returns a fixed-size bit string, the hash value, such that an accidental or intentional change to the data will change the hash value.

The data to be encoded is often called the message and the hash value is sometimes called the message digest or simply digest.

Properties of Ideal Cryptographic Hash Functions

It is

1. easy to compute the hash value for any given message,
2. infeasible to find a message that has a given hash,
3. infeasible to modify a message without hash being changed,
4. infeasible to find two different messages with the same hash.

Even if the data is stored in an insecure place, its integrity can be checked from time to time by recomputing the digest and verifying that the digest has not changed.

Definition of Hash Family

A hash family is a four tuple $\mathcal{X}, \mathcal{Y}, \mathcal{K}, \mathcal{H}$, where teh following conditions are satisfied:

1. $\mathcal{X}$ is a set of possible messages
2. $\mathcal{Y}$ is a finite set of possible message digests
3. $\mathcal{K}$ is a finite set of possible keys
4. For each $K \in \mathcal{K}$, there is a hash function $h_K \in \mathcal{H}$. Each $h_K : \mathcal{X} \to \mathcal{Y}$.

A pair $(x, y) \in \mathcal{X} \times \mathcal{Y}$ is said to be valid under the key K if $h_K(x) = y$.
Let $\mathcal{F}^{\mathcal{X},Y}$ denote the set of all functions from $\mathcal{X}$ to $\mathcal{Y}$. Suppose that $|\mathcal{X}| = N$ and $|\mathcal{Y}| = M$. Then $|\mathcal{F}^{\mathcal{X},Y}| = M^N$. Any hash family $\mathcal{F} \subseteq \mathcal{F}^{\mathcal{X},Y}$ is termed an (N, M)-hash family.
MDC (Modification Dedection Code): An unkeyed hash function is a function $h_K : \mathcal{X} \to \mathcal{Y}$, where $|\mathcal{K}| = 1$.

Security of Cryptographic Hash Functions

A cryptographic hash function must be able to withstand all known types of cryptanalytic attacks. As a minimum, it must have the following properties:

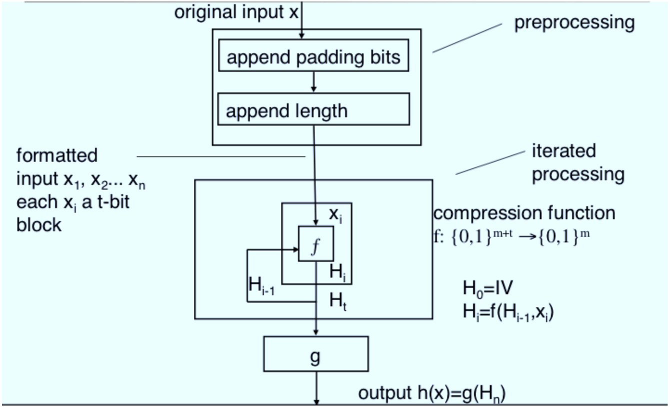
1. Preimage resistance: Given a hash y it should be difficult to find any message x such that $y = h(x)$. This concept is related to that of one-way function.
2. Second preimage resistance: Given an input x_1 it should be difficult to find another input x_2 where $x_1 \neq x_2$ such that $h(x_1) = h(x_2)$. This property is sometimes referred to as weak collision resistance.
3. Collision resistance: It should be difficult to find two different messages x_1 and x_2 such that $h(x_1) = h(x_2)$. Such a pair is called a cryptographic hash collision. This property is sometimes referred to as strong collision resistance. It requires a hash value at least twice as long as that required for preimage-resistance, otherwise collisions may be found by a birthday attack.

Uses of hash functions

- Message authentication
- Software integrity
- One-time Passwords
- Digital signature
- Timestamping
- Certificate revocation management

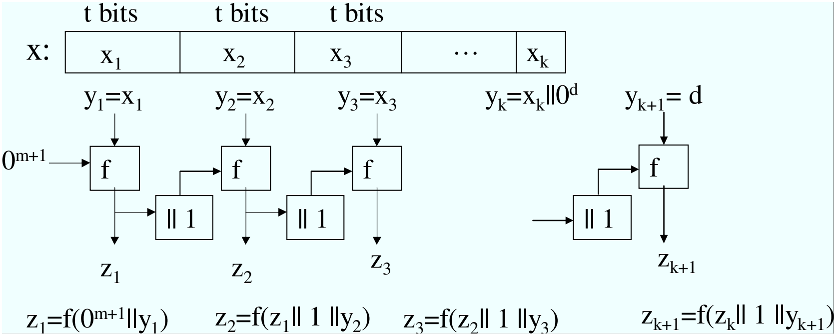
Constructing Hash Function From Compression Functions

A compression function takes a fixed-length input string and output a shorter string $f : \{0,1\}^{m+t} \to \{0,1\}^m$.



The Merkle-Damgard Construction of Hash Functions

- Goal: construct a hash function $h : \{0,1\}^\star \to \{0,1\}^m$ from a compression function $f : \{0,1\}^{m+t+1} \to \{0,1\}^m$
- Given message x of arbitrary length



Example:

- Compression function: $f : \{0,1\}^{128+512+1} \to \{0,1\}^{128}$
- Message x has 1000 bits:
  - y_1 is first 512 bits of x
  - y_2 is last 488 bits of $x||0^{24}$
  - y_3 is $0^{480}||$ 32-bit binary representation of 24
- Iteration results
  - $z_1 = f\left(0^{129}||y_1\right)$ z_1 has 128 bits
  - $z_2 = f\left(z_1||1||y_2\right)$
  - $z_3 = f\left(z_2||1||y_3\right)$ z_3 is the message digest $h(x)$

Example:

- Suppose that message x' has 488 bits and $h(x) = h(x')$ (there is a collision for h):
  - y_1' is $x'||0^{24}$
  - y_2' is $0^{480}||$ 32-bit binary representation of 24
  - $z_1' = f\left(0^{129}||y_1'\right)$ z_1' has 128 bits
  - $z_2' = f\left(z_1'||1||y_2'\right)$ z_2' is $h(x')$
- Then $f\left(z_1'||1||y_2'\right) = f\left(z_2||1||y_3\right)$ and $y_3 = y_2'$
  - if $z_1' \neq z_2$ then a collision is found for f
  - if $z_1' = z_2$ then $f\left(0^{129}||y_1'\right) = f\left(z_1||1||y_2\right)$, there is also a collision for f

Security of the Merkle- Damgard Construction

If $f : \{0,1\}^{m+t+1} \to \{0,1\}^m$ is collision resistant, then the Merkle-Damgard construction $h : \{0,1\}^\star \to \{0,1\}^m$ is collision resistant.

SHA1 (Secure Hash Algorithm)

- SHA was designed by NIST and is the US federal standard for hash functions, specified in FIPS-180 (1993).
- SHA-1, revised version of SHA, specified in FIPS-180-1 (1995) use with Secure Hash Algorithm).
- It produces 160-bit hash values.
- NIST have issued a revision FIPS 180-2 that adds 3 additional hash algorithms: SHA-256, SHA-384, SHA-512, designed for compatibility with increased security provided by AES.

SHA3 Contest

NIST announced a public competition on Nov. 2, 2007 to develop a new cryptographic hash algorithm. The winning algorithm will be named "SHA-3", and will augment the hash algorithms currently specified in the Federal Information Processing Standard (FIPS) 180-3, Secure Hash Standard.

NIST received 64 entries by October 31, 2008; and selected 51 candidate algorithms to advance to the first round on December 10, 2008, and 14 to advance to the second round on July 24, 2009.

Based on the public feedback and internal reviews of the second-round candidates, NIST selected 5 SHA-3 finalists - BLAKE, Grøstl, JH, Keccak, and Skein to advance to the third (and final) round of the competition on December 9, 2010, which ended the second round of the competition.

A one-year public comment period is planned for the finalists. NIST also plans to host a final SHA-3 Candidate Conference in the spring of 2012 to discuss the public feedback on these candidates, and select the SHA-3 winner later in 2012.

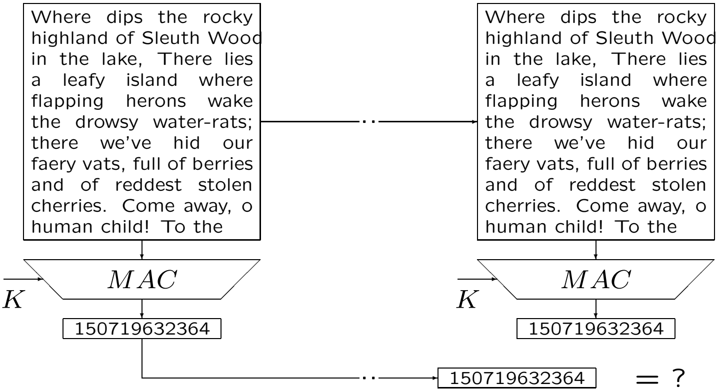
Further details of the competition are available at

Message Authentication Codes

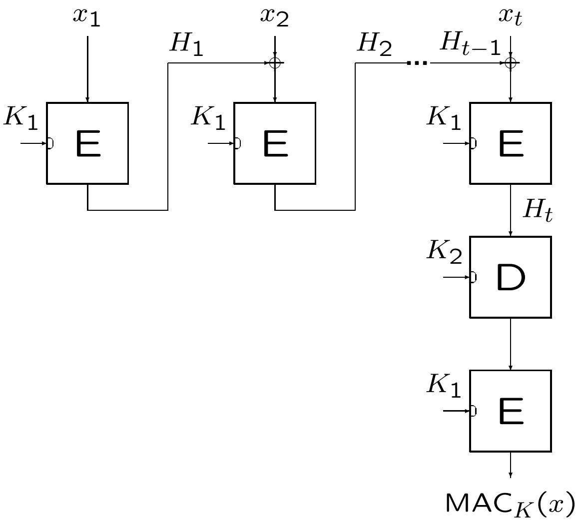
MAC (Message Authentication Code): Hash function with secret key

- hard to produce a forgery
- can only be generated and verified by someone who secret MAC-key
- do not use the same key for MAC and for encryption

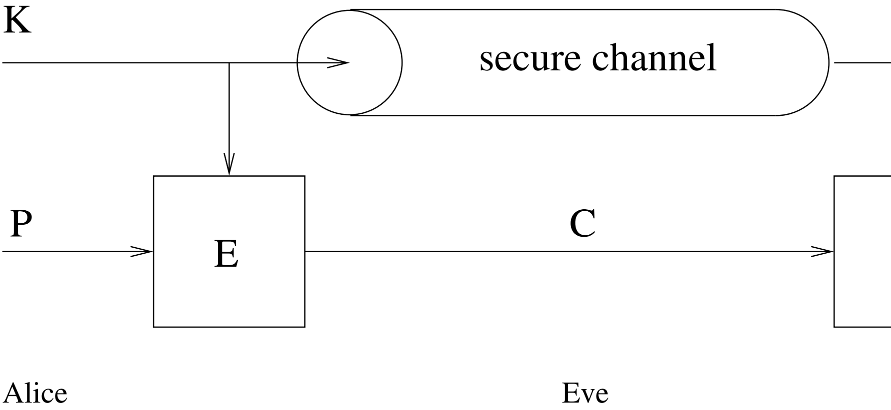
MAC = hash function with secret key



MAC based on block cipher: retail MAC



Symmetric Key Cryptography



Secret Key $\leftrightarrow$ Public Key

- key agreement
  How can 2 people who have never met share a key which is only known to these 2 people
- digital signature
  How can one be sure that a message comes from the sender who claims to have produced that message?

Public Key Cryptosystem

W. Diffie, M. Hellman, "New directions in cryptography", IEEE Transactions on Information Theory, Nov 1976, Volume: 22, Issue:6, page(s): 644 - 654.

1. for every $K \in \mathcal{K}$ e_K is the inverse of d_K,
2. for every $K \in \mathcal{K}$, $x \in \mathcal{P}$ and $y \in \mathcal{C}$ $e_K(x) = y$ and $d_K(y) = x$ are easy to compute.
3. for almost every $K \in \mathcal{K}$, each easily computed algorithm equivalent to d_K is computationally infeasible to derive from e_K,
4. for every $K \in \mathcal{K}$, it is feasible to compute inverse pairs e_K and d_K from K.

Because of the third property, a user's enciphering function e_K can be made public without compromising the security of his secret deciphering function d_K. The cryptographic system is therefore split into two parts, a family of enciphering transformations and a family of deciphering transformations in such a way that, given a member of one family, it is infeasible to find the corresponding member of the other.

Problem 1: Key-Agreement (1/3)

Diffie-Hellman Key Agreement Protocol
($f(X, Z)$: commutative one way function)

Alice		*Bob*
$Y_A = f(X_A, Z)$		
	$\xrightarrow{Y_A}$	
		$Y_B = f(X_B, Z)$
	$\xleftarrow{Y_B}$	
$K_{AB} = f(X_A, Y_B) = f(X_A, f(X_B, Z))$		$K_{BA} = f(X_B, f(X_A, Z))$

Key-Agreement (2/3)

Modular Exponentiation

- given α and a prime p with $\alpha \in [1, p-1]$
- $w = \alpha^x \bmod p$ can be computed efficiently (square and multiply)

Inverse operation (*discrete logarithm*)

- given α, p and w, find x such that

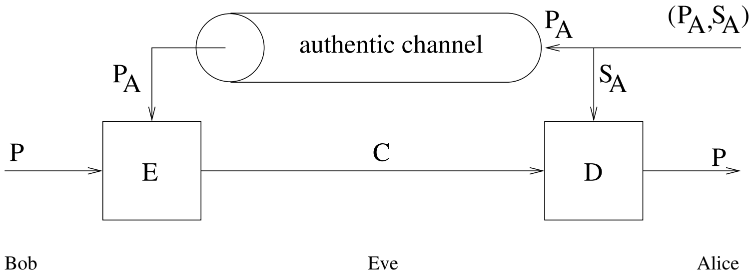
$$\alpha^x \bmod p \equiv w$$

Key-Agreement (3/3)

- $p = 37$: the integers from 0 to 36 form a field with $+$ and $\times$ mod 37
- $\alpha = 2$ is a generator of the non-zero elements: powers of 2 generate all non-zero elements: $2^0 = 1$, $2^1 = 2$, $2^3 = 8$, $2^4 = 16$, $2^5 = 32$, $2^6 = 27$, $2^7 = 17$, ..., $2^{36} = 1$
- $X_A = 10 \Rightarrow Y_A = 2^{10} \bmod 37 = 25$
- $X_B = 13 \Rightarrow Y_B = 2^{13} \bmod 37 = 15$
- $K_{AB} = (Y_B)^{X_A} = 15^{10} \bmod 37 = 15^{8+2} \bmod 37 = 7 \times 3 \bmod 37 = 21$
- $K_{BA} = (Y_A)^{X_B} = 25^{13} \bmod 37 = 25^{8+4+1} \bmod 37 = 34 \times 16 \times 25 \bmod 37 = 21$
- $K_{AB} = K_{BA} = 21$

Problem 2: Public-key cryptography (1/3)

(trapdoor one-way functions)



Public-key cryptography (2/3)

RSA public-key algorithm

trapdoor one-way function:

- given x: "easy" to compute $f(x)$
- given $f(x)$: "hard" to compute x
- given $f(x)$ and the trapdoor information: finding x is "easy"

given two large primes p and q and a public key (e, n)
$n = p \times q$ (factoring n is hard)
$f(x) = x^e \bmod n$ is a trapdoor one-way function
trapdoor information (p, q) allows to find a private key (d, n) such that
$(x^e)^d = (x^e)^{1/e} = x \bmod n$

Public-key cryptography (3/3)

RSA public-key algorithm (2): detail

key generation:

choose two primes p and q

$n = p \times q$, $\phi(n) = (p-1)(q-1)$

choose e prime w.r.t. $\phi(n)$

compute $d = e^{-1} \bmod \phi(n)$

public key $= (e, n)$

private key $= (d, n)$ or (p, q)

encrytion: $c = m^e \bmod n$

decrytion: $m = c^d \bmod n$

Modular Exponentiation

Attacks on the RSA Cryptosystem

Although 35 years of research have led to a number of fascinating attacks, none of them is devastating. They mostly illustrate the dangers of improper use of RSA. Indeed, securely implementing RSA is a nontrivial task.

Factoring Large Integers

We refer to factoring the modulus as a brute-force attack on RSA.

Factoring algorithms running time

Pollard's Rho algorithm	$O\left(\sqrt{p}\right)$
Pollard's $p-1$ algorithm	$O\left(p'\right)$ where p' is the largest prime factor of $p-1$
Pollard' s $p+1$ algorithm	$O\left(p'\right)$ where p' is the largest prime factor of $p+1$
Elliptic Curve method (ECM)	$O\left(e^{(1+o(1))(2\ln p \ln\ln p)^{1/2}}\right)$
Quadratic Sieve (Q.S.)	$O\left(e^{(1+o(1))(\ln N \ln\ln N)^{1/2}}\right)$
Number Filed Sieve (NFS)	$O\left(e^{(1.92+o(1))(\ln N)^{1/3}(\ln\ln N)^{2/3}}\right)$

Our objective is to survey attacks on RSA that decrypt messages without directly factoring the RSA modulus N.
Is breaking RSA as hard as factoring?

Chinese Remainder Theorem

The following problem was posed by Sunzi (4th century AD) in the book Sunzi Suanjing:

when a number is
repeatedly divided by 3, the remainder is 2;
by 5 the remainder is 3;
and by 7 the remainder is 2.

What will be the number?

Oystein Ore mentions another puzzle with a dramatic element from Brahma-Sphuta-Siddhanta (Brahma's Correct System) by Brahmagupta (born 598 AD):
An old woman goes to market and a horse steps on her basket and crashes the eggs.
The rider offers to pay for the damages and asks her how many eggs she had brought.
She does not remember the exact number, but when she had taken them out two at a time, there was one egg left. The same happened when she picked them out three, four, five, and six at a time, but when she took them seven at a time they came out even. What is the smallest number of eggs she could have had?

Involves a situation like the following: we are asked to find an integer x which gives a remainder of 4 when divided by 5, a remainder of 7 when divided by 8, and a remainder of 3 when divided by 9.
In other words, we want x to satisfy the following congruences.
$x \equiv 4 \bmod 5$, $x \equiv 7 \bmod 8$, $x \equiv 3 \bmod 9$
There can be any number of moduluses, but no two of them should have any factor in common. Otherwise the existence of a solution cannot be guaranteed.
The method for solving this set of three simultaneous congruences is to reduce it to three separate problems whose answers may be added together to get a solution to the original problem.
To understand this, think about why
$144 + 135 + 120$ will be a solution to the simultaneous congruences.
144 gives a reminder of 4 when divided by 5. On the other hand, 135 and 120 are multiples of 5, so adding them doesn't change this reminder.
$144 + 135 + 120 \equiv 144 \bmod 5 \equiv 4 \bmod 5$
135 gives a reminder of 7 when divided by 8. On the other hand,

Broadcast Attack

Think that Alice wants to send the same message, x to Bob, Bill and Bart, who have all the same public key, e, but different modulus, n_1, n_2, n_3. Can Eve find x without knowing the private keys?
Yes, she can by using CRT!

$$x^e \equiv a_1 \bmod n_1$$
$$x^e \equiv a_2 \bmod n_2$$
$$x^e \equiv a_3 \bmod n_3$$

Common Modulus

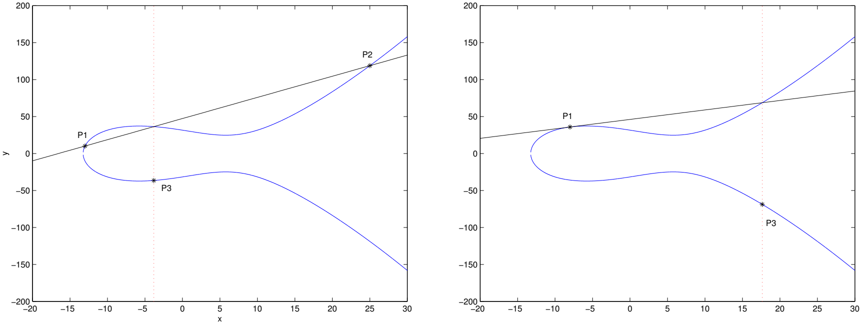
To avoid generating a different modulus $N = pq$ for each user, one may wish to fix N once and for all. The same N is used by all users. A trusted central authority could provide user i with a unique pair (e_i, d_i) from which user i forms a public key $< N, e_i >$ and a secret key $< N, d_i >$.
Fact 1: Let $\langle N, e \rangle$ be an RSA public key. Given the private key d, one can efficiently factor the modulus $N = pq$. Conversely, given the factorization of N, one can efficiently recover d.
By Fact 1 Bob can use his own exponents e_b, d_b to factor the modulus N. Once N is factored Bob can recover Alice's private key d_a from her public key e_a. This observation, due to Simmons, shows that an RSA modulus should never be used by more than one entity. Exposing the private key d and factoring N are equivalent. Hence there is no point in hiding the factorization of N from any party who knows d.

Blinding

Let $< N, d >$ be Bob's private key and $< N, e >$ his corresponding public key. Suppose Marvin wants Bob's signature on a message $M \in Z_N^*$. Being no fool, Bob refuses to sign M.
Marvin can try the following: he picks a random $r \in Z_N^*$ and sets $M' = r^e M \bmod N$. He then asks Bob to sign the random message M'. Bob may be willing to provide his signature S' on the innocent-looking M'. Marvin now simply computes $S = S'/r \bmod N$ and obtains Bob's signature S on the original M. Indeed,
$S = \frac{S'}{r} = \frac{M'^d}{r} = \frac{r^{ed} M^d}{r} = \frac{rM^d}{r} = M^d$
This technique, called blinding, enables Marvin to obtain a valid signature on a message of his choice by asking Bob to sign a random "blinded" message. Bob has no information as to what message he is actually signing.

Elliptic Curve Group over $\mathbb{R}$

Definition: set of the solutions of Weierstrass equation $E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$ over a field and the point at infinity $\mathcal{O}$.



Adding two points

Doubling a point

Elliptic Curve Point Addition and Doubling over $GF(p)$ $p > 3$

$$E : y^2 = x^3 + ax + b$$

$$P_1 = (x_1, y_1),\ P_2 = (x_2, y_2) \text{ and } P_3 = (x_3, y_3) = P_1 + P_2$$

$$\begin{aligned} x_3 &= \lambda^2 - x_1 - x_2 \\ y_3 &= \lambda(x_1 - x_3) - y_1 \end{aligned}$$

$$\lambda = \begin{cases} (y_2 - y_1)\,(x_2 - x_1)^{-1} & \text{if } P_1 \neq P_2 \\ (3x_1^2 + a)\,(2y_1)^{-1} & \text{if } P_1 = P_2 \end{cases}$$

projective coordinates are used to get rid of modular multiplicative inversion

Elliptic Curve Point Multiplication

$$[k]P = \underbrace{P + P + \cdots + P}_{k}$$

Require: EC point $P = (x, y)$, integer k, $0 < k < M$,
$k = (k_{l-1}, k_{l-2}, \cdots, k_0)_2$, $k_{l-1} = 1$ and M

Ensure: $Q = [k]P = (x', y')$

$Q \leftarrow P$

for i from $l - 2$ downto 0 **do**

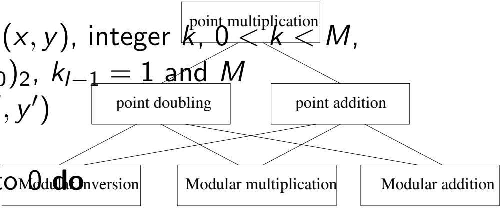
  $Q \leftarrow 2Q$

  if $k_i = 1$ **then**

    $Q \leftarrow Q + P$

  end if

end for



Elliptic Curve Point Addition and Doubling

Require: $P_1 = (x, y, 1, a)$, $P_2 = (X_2, Y_2, Z_2, aZ_2^4)$

Ensure: $P_1 + P_2 = P_3 = (X_3, Y_3, Z_3, aZ_3^4)$

1.	$T_1 \leftarrow Z_2^2$	
2.	$T_2 \leftarrow xT_1$	
3.	$T_1 \leftarrow T_1 Z_2$	$T_3 \leftarrow X_2 - T_2$
4.	$T_1 \leftarrow yT_1$	
5.	$T_4 \leftarrow T_3^2$	$T_5 \leftarrow Y_2 - T_1$
6.	$T_2 \leftarrow T_2 T_4$	
7.	$T_4 \leftarrow T_4 T_3$	$T_6 \leftarrow 2T_2$
8.	$Z_3 \leftarrow Z_2 T_3$	$T_6 \leftarrow T_4 + T_6$
9.	$T_3 \leftarrow T_5^2$	
10.	$T_1 \leftarrow T_1 T_4$	$X_3 \leftarrow T_3 - T_6$
11.	$aZ_3^4 \leftarrow Z_3^2$	$T_2 \leftarrow T_2 - X_3$
12.	$T_3 \leftarrow T_5 T_2$	
13.	$aZ_3^4 \leftarrow \left(aZ_3^4\right)^2$	$Y_3 \leftarrow T_3 - T_1$
14.	$aZ_3^4 \leftarrow a\left(aZ_3^4\right)$	

latency $= 14T_{MM}$

Require: $P_1 = (X_1, Y_1, Z_1, aZ_1^4)$

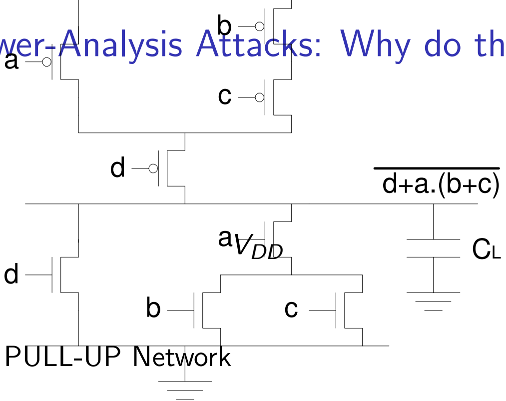
Ensure: $2P_1 = P_3 = (X_3, Y_3, Z_3, aZ_3^4)$

Modular Addition, Subtraction Circuit over $GF(p)$

Require: M, $0 \le A < M$,
  $0 \le B < M$
Ensure: $C = A + B \bmod M$
  $C' = A + B$
  $C'' = C' - M$
  if $C'' < 0$ **then**
    $C = C'$
  else
    $C = C''$
  end if

Require: M, $0 \le A < M$,
  $0 \le B < M$
Ensure: $C = A - B \bmod M$
  $C' = A - B$
  $C'' = C' + M$
  if $C' < 0$ **then**
    $C = C''$
  else
    $C = C'$

Power-Analysis Attacks: Why do they work?



PULL-UP Network

Dynamic power consumption is mainly due to the charge and discharge of the load capacitance C_L.

PULL-DOWN Network

Types of Power-Analysis Attacks

Simple Power Analysis (SPA) Attacks:

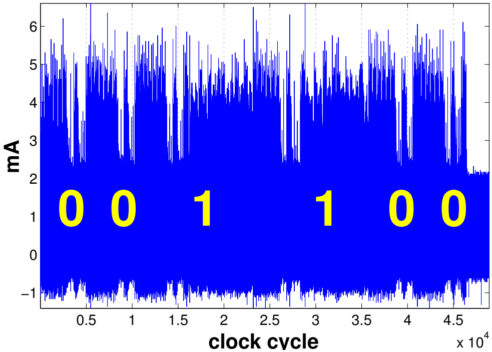
- every instruction $\Longrightarrow$ unique power-consumption trace
- one measurement

Differential Power Analysis (DPA) Attacks:

- many measurements
- statistical analysis used

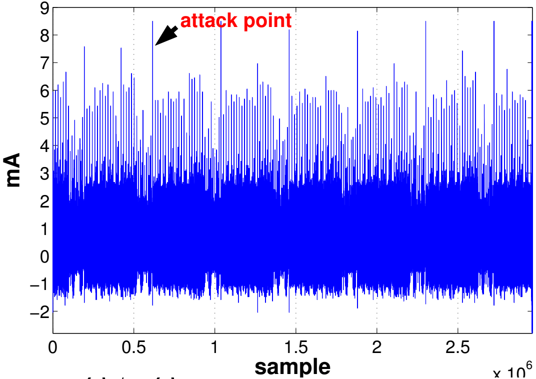
SPA Attack on Elliptic Curve Point Multiplication

Require: EC point $P = (x, y)$, integer k, $k = (k_{l-1}, k_{l-2}, \cdots, k_0)_2$,



The key used during this measurement is 1001100.

Countermeasure for SPA Attack



$= (k_{l-1}, k_{l-2}, \cdots, k_0)_2,$

$Q \leftarrow Q_2$

else

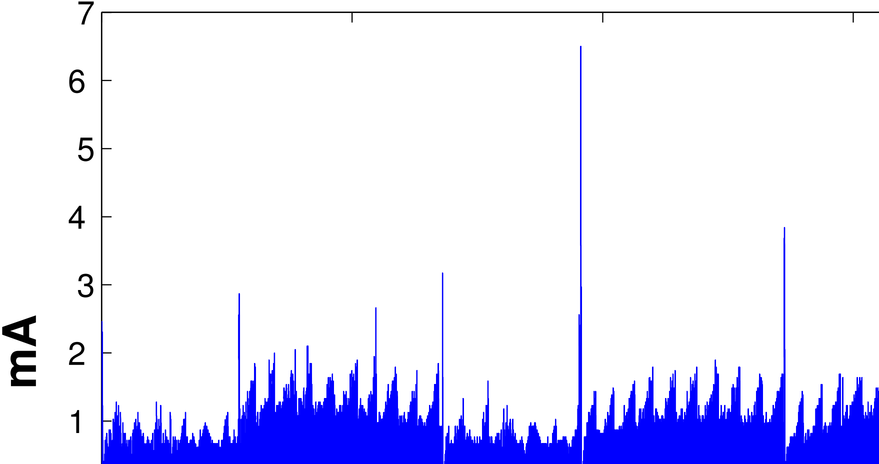
$Q \leftarrow Q_1$

end if

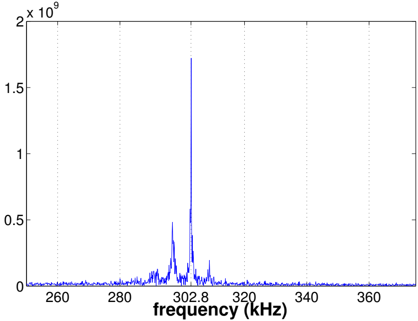
end for

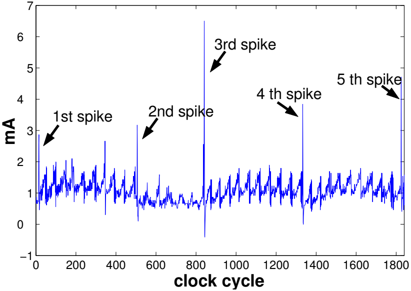
Current Consumption Measurement for DPA Attack

- Data length: 2400 clock cycles around the 2nd update of Q_1.
- Clock frequency: 300 kHz.
- Sampling frequency: 250 MHz.



Pre-Processing





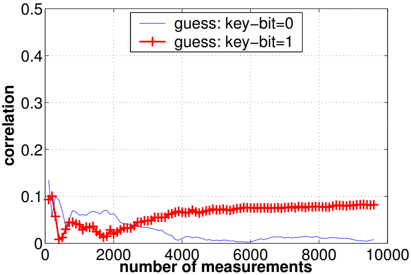
Discrete Fourier transform
between 250 kHz and 375 kHz
Clock frequency : 302.8 kHz

maximum in every clock cycle

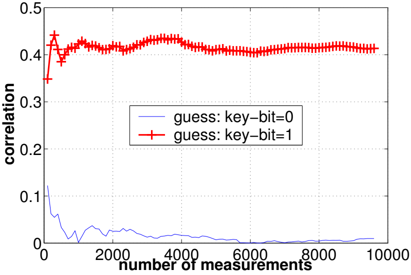
Correlation Analysis

1. hypothetical model $\Longrightarrow$ predict side-channel output for N inputs.
   Prediction is the number of bits changed from 0 to 1 from X_i to X_{i+1}
2. Prediction is for:
   - a certain moment of time
   - a certain key guess
3. Predictions are correlated with the real side-channel output.
   - Correlation is high $\Longrightarrow$ model is correct

Results of the Correlation Analysis

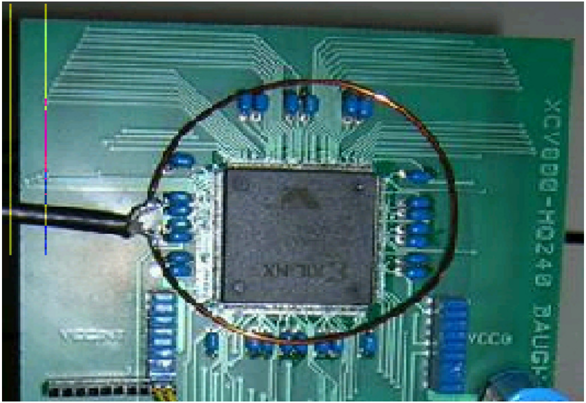


Third spike

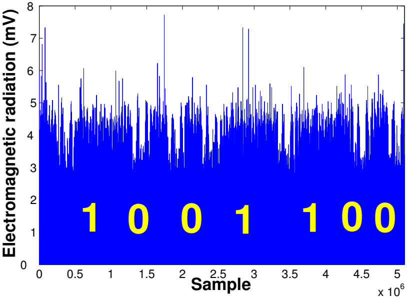


Fifth spike

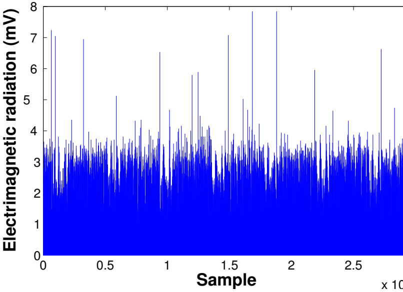
Electromagnetic Analysis of an FPGA Implementation of Elliptic Curve Cryptosystem over $GF(p)$



SEMA Attack

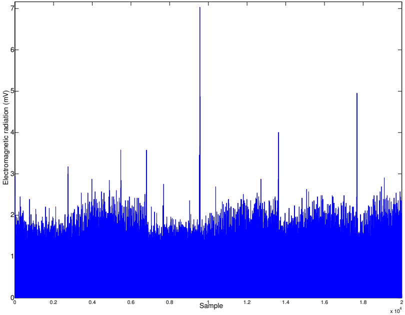


with double and add algorithm



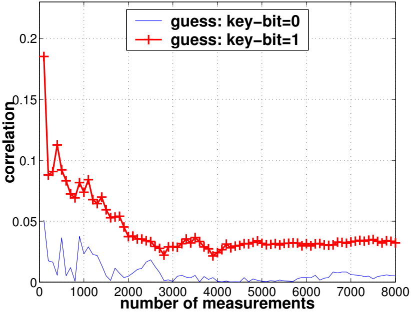
with always double and add algorith

DEMA Attack

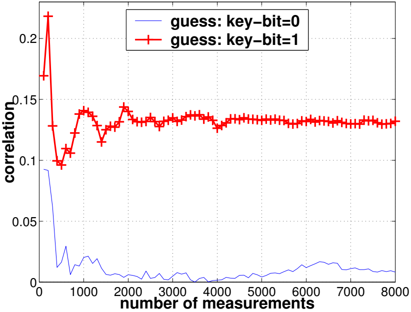


Electromagnetic radiation trace of the FPGA for the attacked point

Correlation Analysis



Third spike



Fifth spike