Symmetric Ciphers对称密码

二. 传统加密技术

Classical Encryption Techniques

WHY？ 为什么需要密码算法

- 信息的存储:存放在公开的地方
- 信息的交换:使用非隐秘介质
- 信息的传输:通过不安全信道

Basic Terminology 基本术语

- 明文(plaintext) - 原文 (original message)
- 密文(ciphertext) - 加密后的消息 (coded message)
- 密码算法(cipher) -明文与密文转换的算法(algorithm)
- 密钥(key) -密码算法的输入,实现明文与密文的转换
- 加密算法(encrypt) -明文转换为密文的算法
- 解密算法(decrypt) -密文转换为明文的算法
- 密码编码学(cryptography) -研究密码的原理及方法的理论
- 密码分析学(cryptanalysis)-研究破译密码获得信息及密钥的学科
- 密码学(cryptology) -密码编码学和密码分析学

密码学

□ 研究与信息安全相关理论

✓ 机密性、完整性、实体鉴别、抗否认等的数学理论

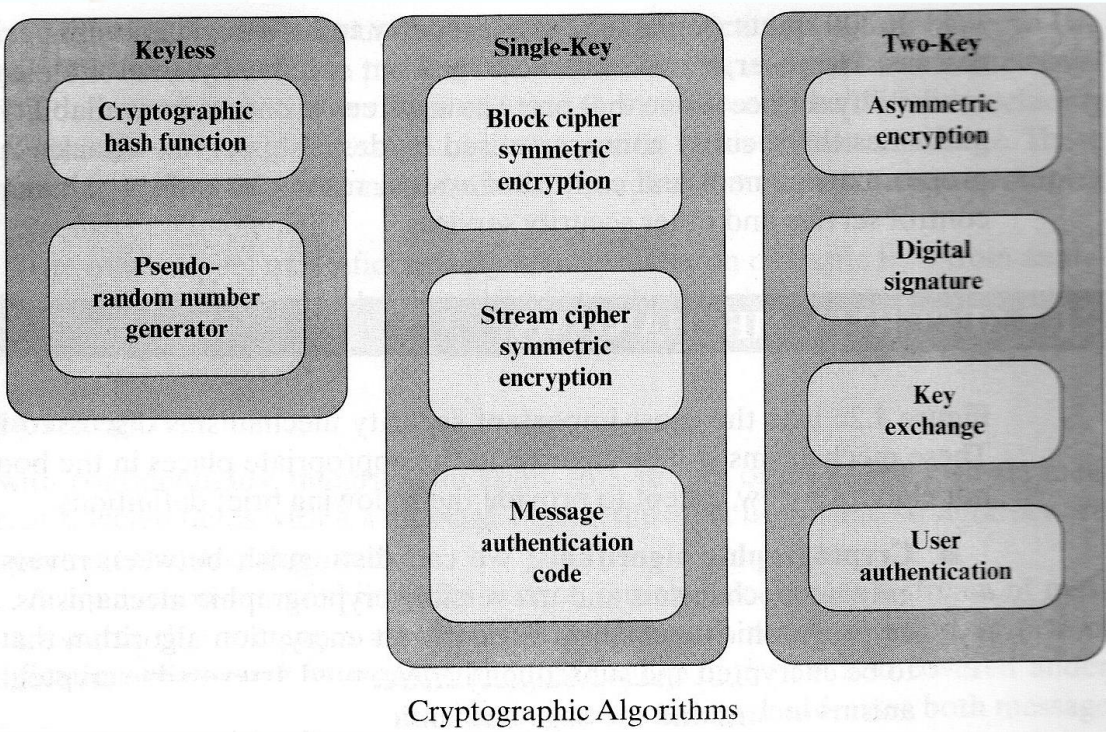
✓ 由**密码编码学**和**密码分析学**构成

- **密码编码学的基本目标**

机密性、数据完整性、鉴别、抗否认

- **基本密码方法**

加密、散列函数、数字签名



Cryptographic Algorithms

密码编码 Cryptography

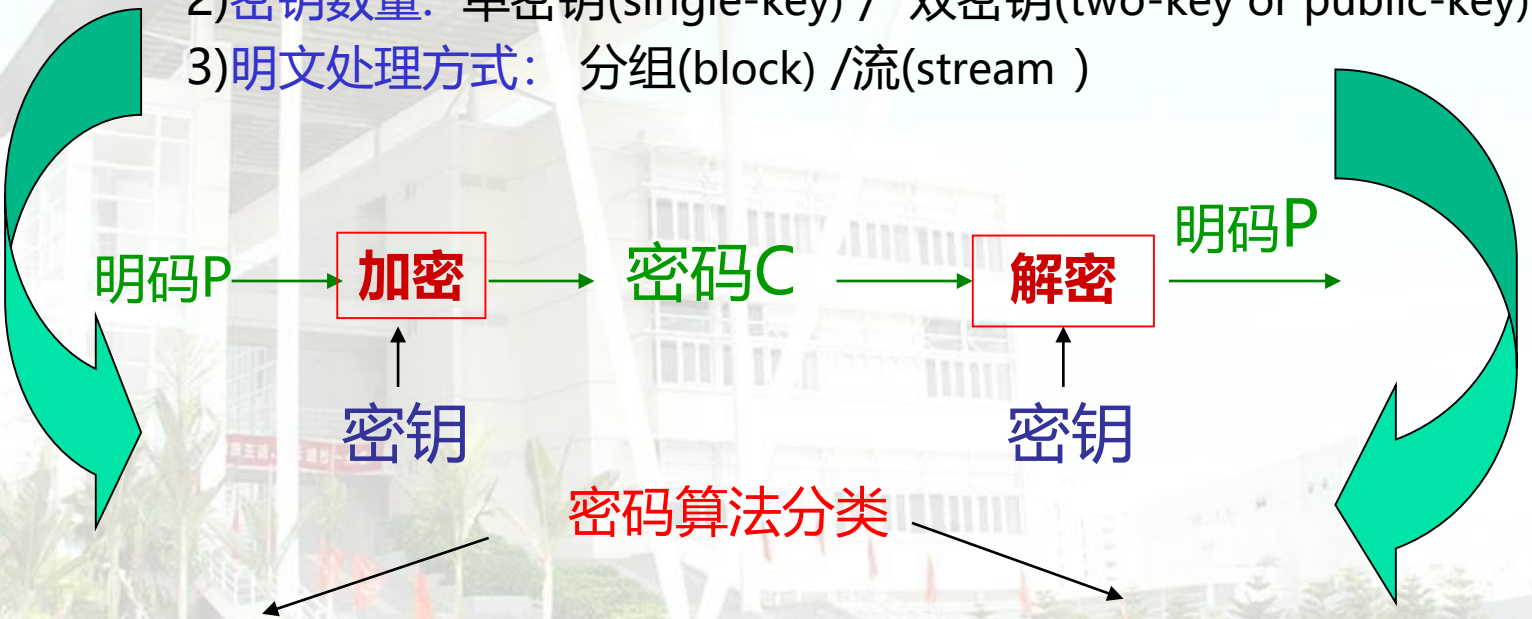
characterize cryptographic system by:

1)明密转换类型:

- 替代(substitution)： 明文中的元素映射成另一元素。
- 置换(transposition)： 明文中的元素被重新排列。

2)密钥数量: 单密钥(single-key) / 双密钥(two-key or public-key)

3)明文处理方式： 分组(block) /流(stream)



密码算法

- 受限（restricted)算法

  保密性基于对算法的保密

- 基于密钥（key-based)的算法

  保密性基于对密钥的保密

密码算法原理

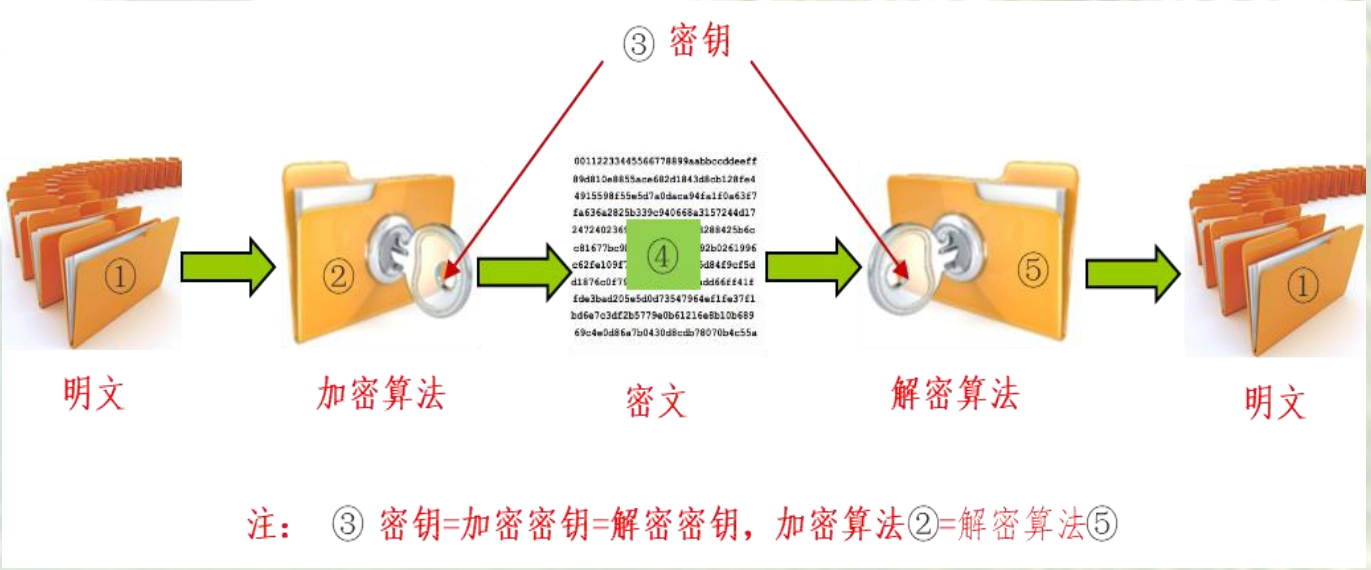
➢数据加密技术

- ✓ 数据加密（Data Encryption）：利用加密密钥（Encryption key）通过加密算法将明文数据（Plain text）转换成密文数据（Cipher text）
- ✓ 收到密文的接收方利用解密密钥（Decryption key）通过解密算法还原成明文数据

- ✓ 加密技术是网路与信息安全的基础，将加密算法和解密算法通称为密码算法
- ✓ 分为对称密码算法及非对称密码算法两大类

对称加密算法（Symmetric Encryption）

- 对称密码算法：加密运算与解密运算使用同一把密钥，对称密码模型如图所示



- ✓ 由5部分组成：①明文、②加密算法、③密钥、④密文、⑤解密算法
- ✓ 加密算法与解密算法采用同一算法，加密密钥与解密密钥为同一把密钥
- ✓ 常见的对称加密算法有AES、3DES、以及SM4

对称加密算法

- 如果用*X*代表明文，*Y*代表密文，*E*代表加密算法，*D*代表解密密算法，则加解密过程可描述为：

$$Y = E_K(X)$$

$$X = D_K(Y)$$

- 优点

（1）加解密速度快,适用于直接对大量数据进行加密

（2）保密性主要取决于密钥的安全性

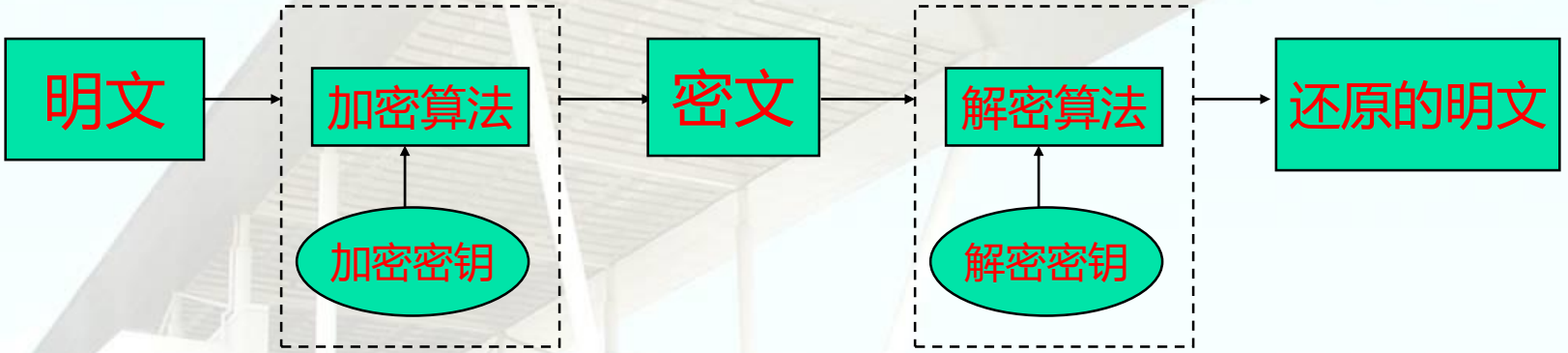
（3）发送及接收双方需事先约定共有密钥

- 挑战

如何在公开及分布的电脑网路上安全保管及大量分发密钥

双方使用相同密钥，因此无法实现数据签名和不可否认性等功能

对称密码算法特点

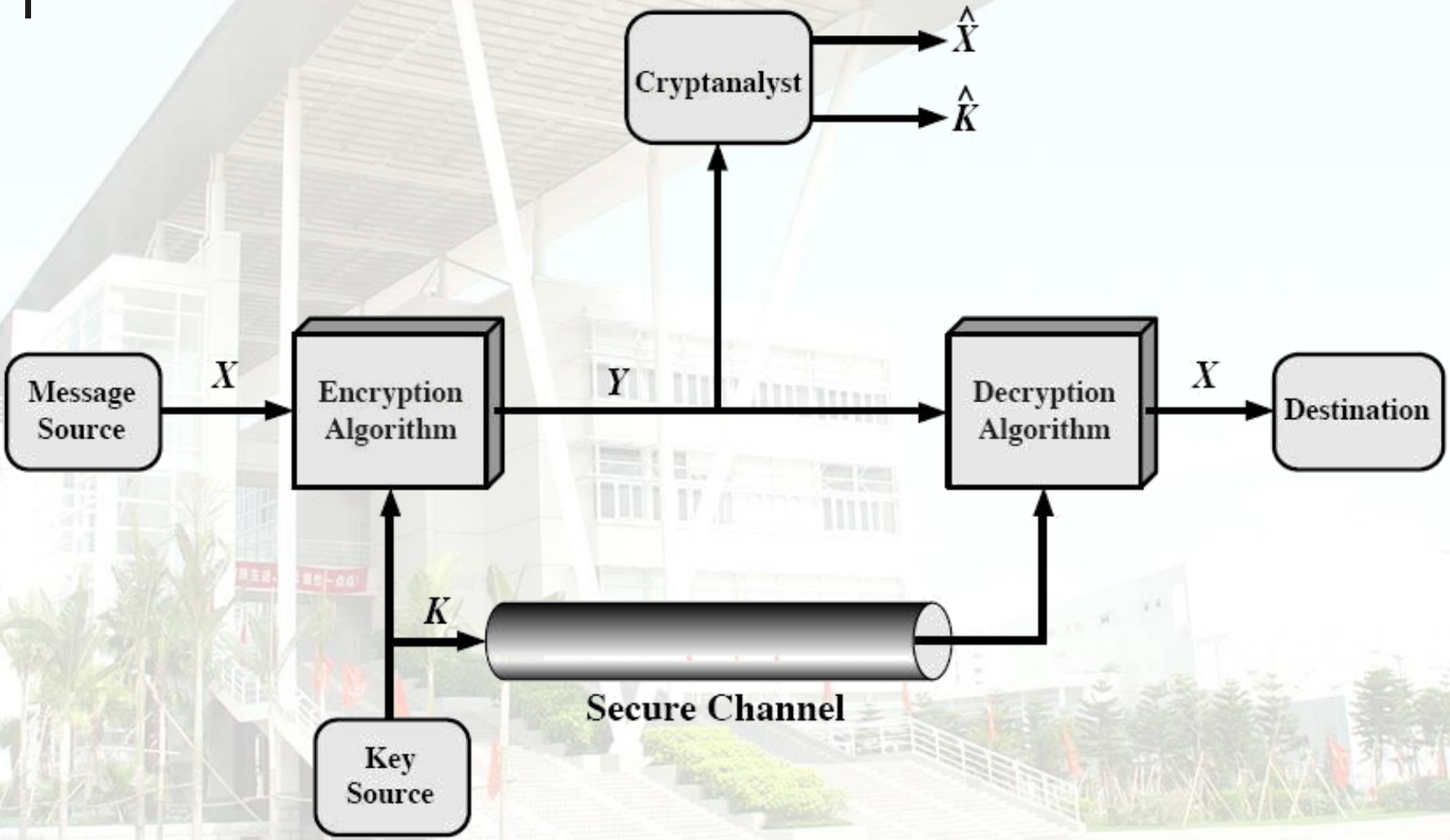


- 加密算法和解密算法公开(known)
- 加密密钥和解密密钥是同一支密钥(single-key)，称对称密钥

two requirements

- 加密算法足够强大(strong encryption algorithm)，仅依靠密文不可能译出明文
- 安全性依赖于密钥的安全性，而不是算法安全性。
  密钥的安全性包括：密钥空间、随机性、保密性。
  need a secure channel to distribute key !!!
- 算法符号描述： $Y = E_K(X)$, $X = D_K(Y)$

Model of cryptosystem



密码算法公开性(known)

- Make it feasible for widespread use 便于广泛使用
- Low-cost chip implementations 低成本芯片实现
- Maintaining the secrecy of the key 密钥的管理

密码**分析学** Cryptanalysis

- objective to recover key not just message only
- general approaches **方法:**

cryptanalytic attack **密码分析攻击**

系统分析法（统计分析法）:利用明文的统计规律

确定性分析法

brute-force attack **强力法** 或称 **穷举攻击**

对截获的密文依次用各种可能的密钥破译。

对所有可能的明文加密直到与截获的密文一致为止

攻击类型

Table 2.1 Types of Attacks on Encrypted Messages

Type of Attack	Known to Cryptanalyst
Ciphertext only	•Encryption algorithm •Ciphertext
Known plaintext	•Encryption algorithm •Ciphertext •One or more plaintext-ciphertext pairs formed with the secret key
Chosen plaintext	•Encryption algorithm •Ciphertext •Plaintext message chosen by cryptanalyst, together with its corresponding ciphertext generated with the secret key
Chosen ciphertext	•Encryption algorithm •Ciphertext •Purported ciphertext chosen by cryptanalyst, together with its corresponding decrypted plaintext generated with the secret key (破译)
Chosen text	•Encryption algorithm •Ciphertext •Plaintext message chosen by cryptanalyst, together with its corresponding ciphertext generated with the secret key •Purported ciphertext chosen by cryptanalyst, together with its corresponding decrypted plaintext generated with the secret key (破译)

已知密文

已知一些明文

选择明文

选择密文

选择文本

Cryptanalytic Attacks- know algorithm

- **ciphertext only (惟密文)**
  - only know ciphertext, **难攻**
- **known plaintext (已知明文)**
  - Know ciphertext & plaintext
- **chosen plaintext (选择明文)**
  - select plaintext and obtain ciphertext
- **chosen ciphertext (选择密文) (少见)**
  - select ciphertext and obtain plaintext
- **chosen text (选择文本) (少见)**
  - select plaintext or ciphertext to en/decrypt

安全性

- **unconditional security 无条件安全**

  no matter how much computer power or time is available,
  the cipher cannot be broken

- **computational security 计算上安全**

  given limited computing resources

  - ✓ 破译密码成本超过信息价值
  - ✓ 破译时间超过信息有效生命周期

  For all reasonable encryption algorithms, we have to assume computational security where it either takes too long, or is too expensive.

攻击的复杂性分析

- **数据复杂性**（data complexity）

  指用作攻击输入所需要的数据

- **处理复杂性**（processing complexity）

  完成攻击所需要的时间

- **存储需求**（storage requirement）

  进行攻击所需要的存储量

强力攻击花费时间
Brute Force Search- Average Time Required

Average Time Required for Exhaustive Key Search

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

DES/3DES/AES

- On average, half of all possible keys must be tried to achieve success.
- proportional to key size
- assume : know plaintext

20世纪早期密码机

对称密码的两个基本运算

- 代换 (Substitution)
- 置换 (Transposition or Permutation)

代换技术Substitution Techniques

- where letters of plaintext are replaced by other letters or by numbers or symbols

  明文的字母由其他字母、数字或符号所代替

- if plaintext is viewed as a sequence of bits, then substitution involves replacing plaintext bit patterns with ciphertext bit patterns

  由密文位串代换明文位串

Caesar Cipher (恺撒密码)

- **earliest known substitution cipher by Julius Caesar**
- **first attested use in military affairs （ in *Gallic Wars* ）**
- **replaces each letter by 3rd letter by defining transformation as:**

```
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:

```
meet me after the toga party
PHHW PH DIWHU WKH WRJD SDUWB
```

Caesar Cipher (恺撒密码)

- mathematically assign each letter a number

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	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

then we have Caesar cipher as:

密文: $c = E(p) = (p + 3) \bmod (26)$

解密: $p = D(c) = (c - 3) \bmod (26)$

推广: 如果将移位3 推广到任意数K [1，25]， **则**

$c = E(p) = (p + k) \bmod (26)$

$p = D(c) = (c - k) \bmod (26)$

Cryptanalysis of Caesar Cipher 恺撒密码分析

- only have 26 possible ciphers
  but only 25 can be used
- **brute force search** by simply trying each in turn
  given ciphertext, just try all shifts of letters
  until can recognize the plaintext

  - **攻击手段**:
    1. 算法简单，密钥K的数量仅为25个 -》强力攻击
    2. 明文单词构造有规律，可以根据字母频率 -》分析攻击

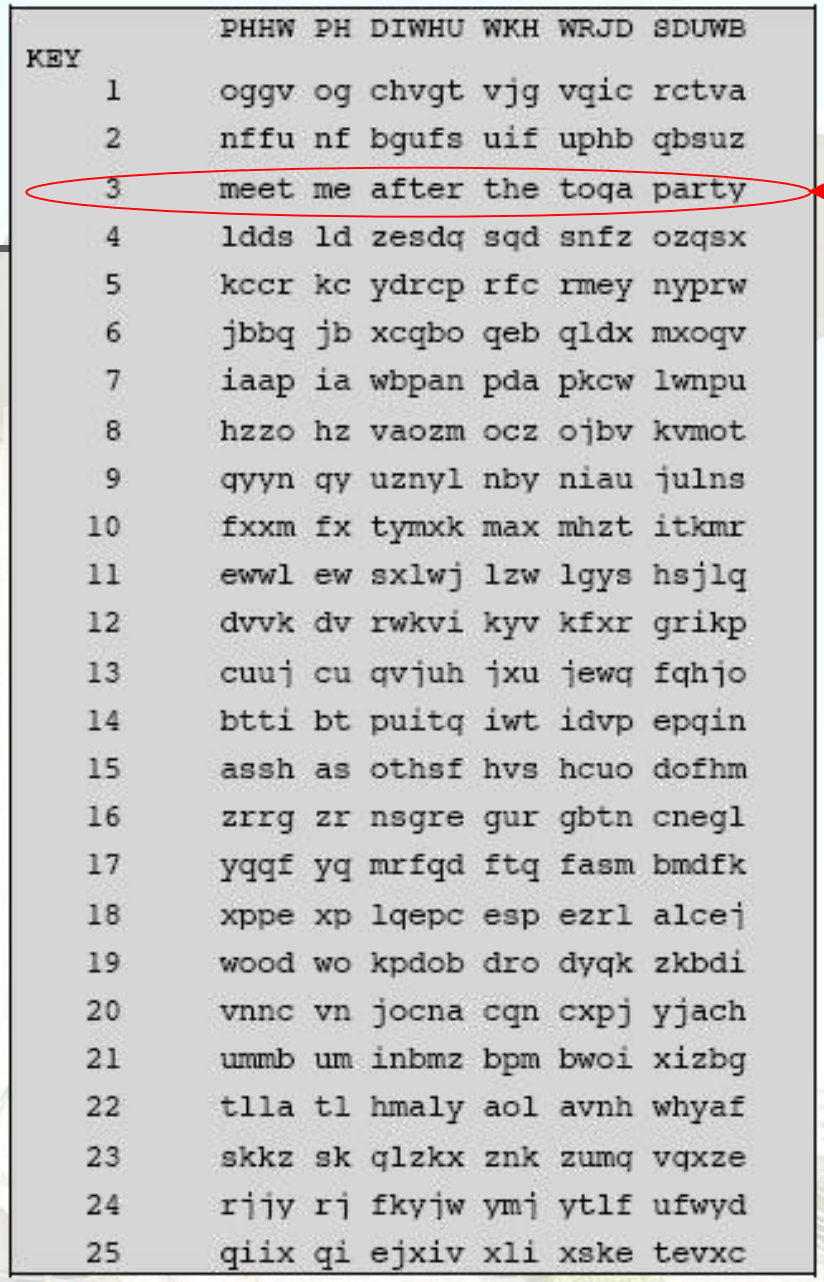


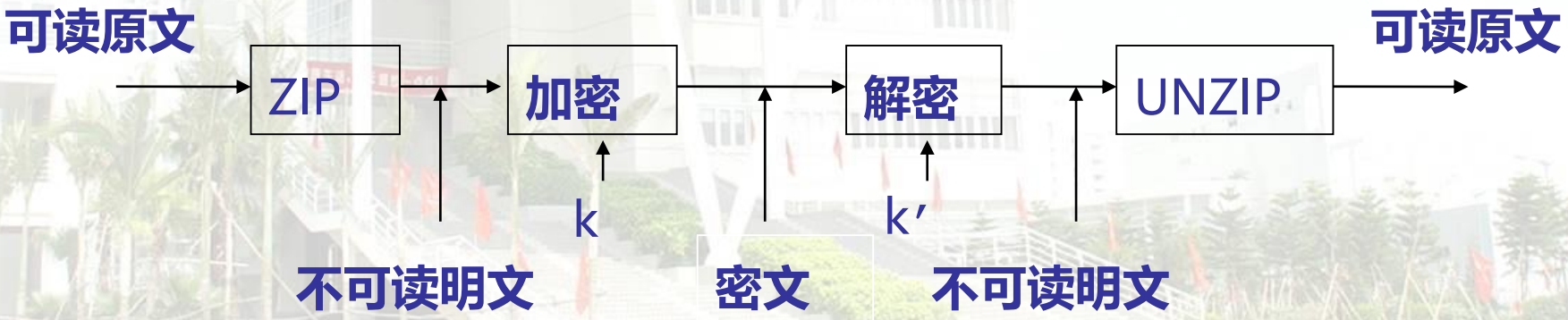
Figure 2.3 Brute-Force Cryptanalysis of Caesar Cipher

Caesar Cipher (恺撒密码)

明文经ZIP之后，增加识别难度

```
˜+Wμ"— Ω—O)≤4{∞ǂ ¸ ë~Ω%ràu·¯Í ◊¯Z-
Ú≠2Ò#Åæ∂ œ«q7 ¸Ωn·®3N◊Ú Œz′Y-ƒ∞Í[±Û_ èΩ,<NO¬±«ˇxã  Åä£èü3Å
x}ö§k º Â
_ÿÍ ˆΔÉ]  ‘  J⁄°iTê&ı'c<uΩ-
¯ÄD(G WÄC~y_ïõÄW PÔı«ÎÜ†ç], ¡ˇÌ^üÑ
πˇ≈ˇLˇ9OgflOˇ&Œ≤¬≤ØÔ§˝: ˇŒ!SGqèvoˆ úError!
```

Figure 2.4 Sample of Compressed Text



但并没增加安全性！！！

改进：

1. 变换方法可以采用26字母的任意排列，使得可用密钥数量剧增。

$26! = 4 * 1026$

2. 使用多字母组合

安全性

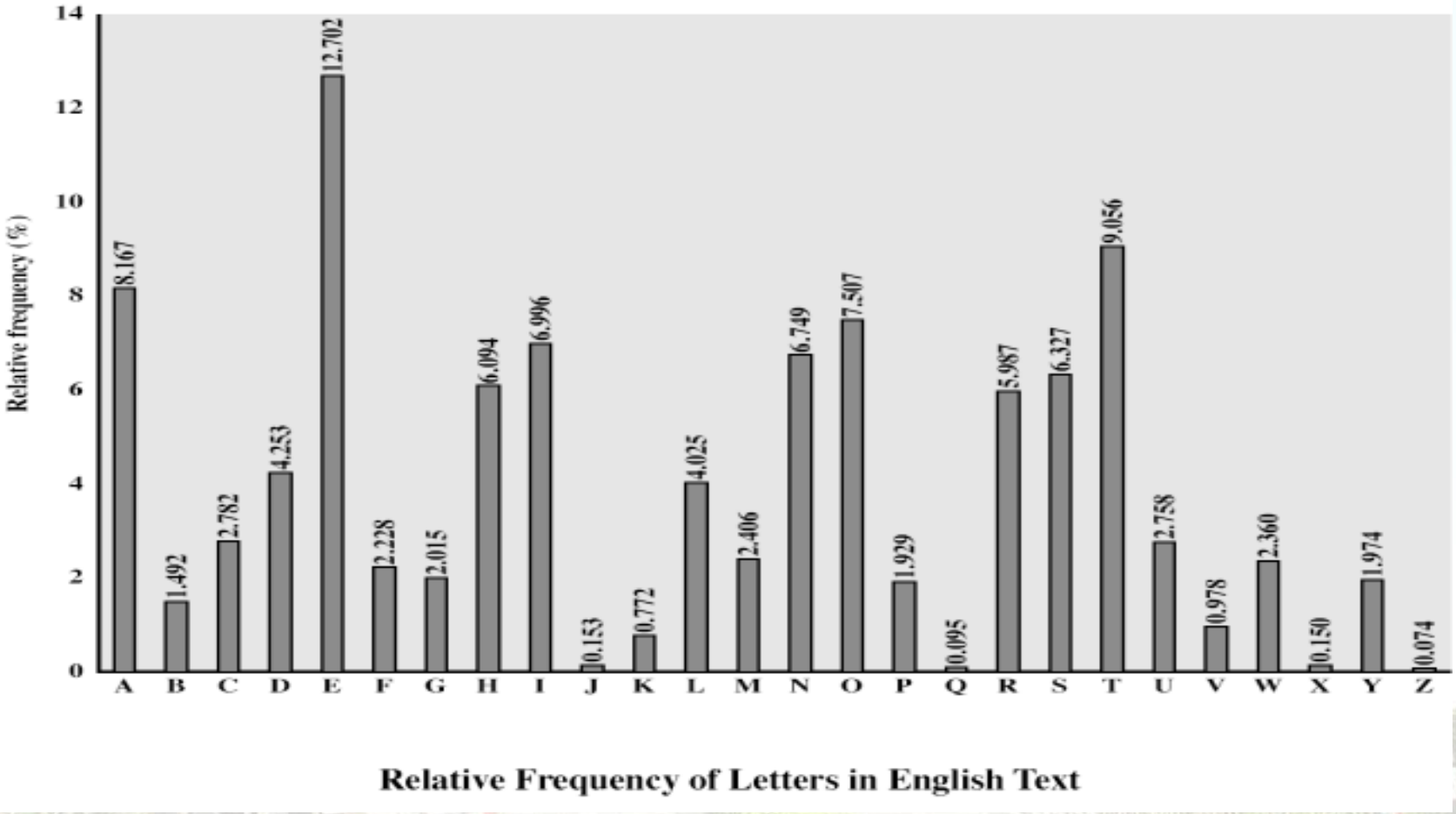
- now have a total of 26! = 4 x 10^26 keys
- with so many keys, is it secure ?

NO !!!

- problem is language characteristics

单表代替密码 容易被攻破，携带了语言的统计学特性。

英语文本中字母的使用频率统计



Playfair Cipher

广泛用于第一次及二次世界大战

- not even the large number of keys in a monoalphabetic (单表代替) cipher provides security

  密钥多并不代表安全

- improving security is to encrypt multiple letters

  the **Playfair Cipher** is an example

  invented by Charles Wheatstone in 1854,

  but named after his friend Baron Playfair

Playfair Key Matrix 密钥矩阵

- a 5X5 matrix of letters based on a keyword
- fill in letters of keyword
- fill rest of matrix with other letters， with I/J used as a single letter.

eg. using the keyword MONARCHY

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

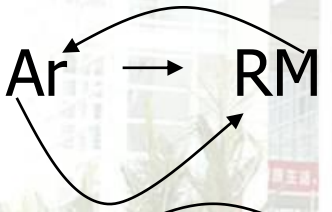
Encrypting and Decrypting 加解密过程

- plaintext is encrypted two letters at a time
  一次加密两个字母

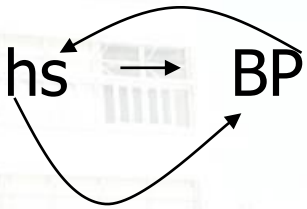
1. if a pair is a repeated letter, insert filler like 'X' 重复填充X
2. if both letters fall in the same row, replace each with letter to right **同行取右边**
3. if both letters fall in the same column, replace each with the letter below it **同列取下边**
4. otherwise each letter is replaced by the letter in the same row and in the column of the other letter of the pair
   其它取交叉

Playfair Key Matrix 密钥矩阵

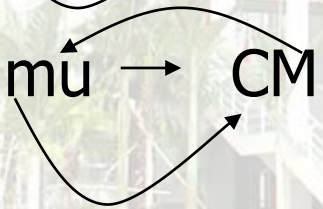
M	O	N	A	R
C	H	Y	B	D
E	F	G	I/J	K
L	P	Q	S	T
U	V	W	X	Z



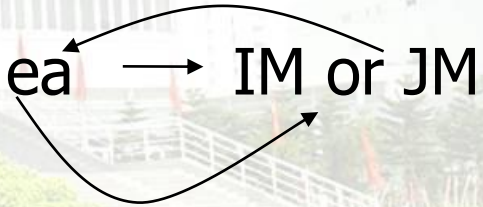
同行取右边



其它取交叉



同列取下边



Playfair密码的安全性

- security much improved over monoalphabetic
- 26 x 26 = 676 tries to analyse (verses 26 for a monoalphabetic)
- was widely used for many years
  - eg. by US & British military in WW1/WW2
- it **can** be broken, given a few hundred letters
- since still has much of plaintext structure

例子

- 取「playfair exm」为密匙，
- 将Q去除，或将I和J视作同一字

- 要加密的讯息: “Hide the gold in the tree stump”
- HI DE TH EG OL DI NT HE TR EX ES TU MP
- 密文: BM ND ZB XD KY BE JV DM UI XM MN UV IF

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

Hill 密码(1929 Lester Hill)

m个连续的明文字母用m个密文字母代替，由m个线性方程决定。这里假设m=3:

$$C1 = (k_{11}\ p1 + k_{12}\ p2 + k_{13}\ p3) \bmod 26$$
$$C2 = (k_{21}\ p1 + k_{22}\ p2 + k_{23}\ p3) \bmod 26$$
$$C3 = (k_{31}\ p1 + k_{32}\ p2 + k_{33}\ p3) \bmod 26$$

即：

$$\begin{pmatrix} C1 \\ C2 \\ C3 \end{pmatrix} = \begin{pmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{pmatrix} \begin{pmatrix} p1 \\ p2 \\ p3 \end{pmatrix}$$

或 C＝KP

Hill密码

基于矩阵的线性（linear）变换

$$C = KP$$

$$P = K^{-1}C$$

$$\begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix} = \begin{pmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} \bmod 26$$

例如：**明文** “paymoremoney”，

$$K = \begin{pmatrix} 17 & 17 & 5 \\ 21 & 18 & 21 \\ 2 & 2 & 19 \end{pmatrix} \qquad K^{-1} = \begin{pmatrix} 4 & 9 & 15 \\ 15 & 17 & 6 \\ 24 & 0 & 17 \end{pmatrix}$$

$$K(15 \quad 0 \quad 24) = (11 \quad 13 \quad 18) = LNS$$

密文 为 “LNSHDLEWMTRW”，

优点：完全隐藏了单字母的频率。

缺点： 如果得到m个明密对，则可以计算出K，从而破密。

Polyalphabetic Ciphers多表代替密码

polyalphabetic substitution ciphers

- improve security using multiple cipher alphabets 多表
- make cryptanalysis harder to guess 较难猜
- use a key to select which alphabet is used 密钥选表

1) A set of related monoalphabetic substitution rules
2) A key determines which particular rule is chosen for a given transformation.

Vigenère(维热纳尔密码) Cipher

- simplest polyalphabetic substitution cipher **最简单**
- effectively multiple caesar ciphers

  多重恺撒编码
- key is multiple letters long $K = k_1\ k_2\ \ldots\ k_d$

  多字母密钥
- i^{th} letter specifies i^{th} alphabet to use

  第i字母确定i表
- decryption simply works in reverse

  解密同理反向

Vigenère替换表 Table 2.3

Table 2.3 The Modern Vigenère Tableau

明文

Key \ Plaintext	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
a	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	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
c	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
d	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
e	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
f	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
g	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
h	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
i	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
j	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
k	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
l	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
m	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
n	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
o	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
p	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
q	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
r	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
s	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
t	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
u	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
v	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
w	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
x	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
y	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
z	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y

密钥

Example of Vigenère Cipher

- write the plaintext out
- write the keyword repeated above it
- use each key letter as a caesar cipher key
- encrypt the corresponding plaintext letter

eg using keyword *deceptive*

```
key:        deceptivedeceptivedeceptive
plaintext:  wearediscoveredsaveyourself
ciphertext:ZICVTWQNGRZGVTWAVZHCQYGLMGJ
```

Security of Vigenère Ciphers 安全性

- have multiple ciphertext letters for each plaintext letter

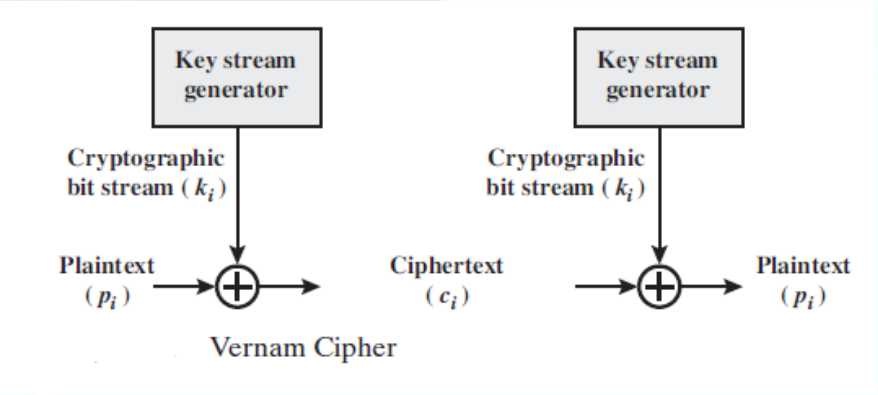
  一个明文字母可对应多个密文字母

- letter frequencies are obscured

  字母频率特性微弱

异或（Exclusive - or）加密 (Vernam)

XOR Operation $\oplus$:



encryption:

$$Ci = Pi \oplus ki$$

Decryption:

$$Pi = Ci \oplus ki$$

Where

Pi = i^{th} binary digit of plaintext

ki = i^{th} binary digit of key

Ci = i^{th} binary digit of ciphertext

One-Time Pad 一次一密

- if a truly random key as long as the message is used, the cipher will be secure， called a One-Time pad

  如果密钥的长度与明文的长度一致，则达到最大强度

- is unbreakable since ciphertext bears no statistical relationship to the plaintext

  密文与明文没有统计关系，牢不可破

- How to Generate large quantities of random keys?

  随机密钥的产生

- How to distribute the key safely?

  密钥的分发

对称密码基本运算

- 置换 **(Transposition or Permutation)**

rearranging the letter order

明文字母的某种置换，形成新的排列

通过执行对明文字母的某种置换，取得一种类型完全不同的映射-Mapping

。

Rail Fence (栅栏) cipher

原文：meet me after the toga party

write message out as:

m e m a t r h t g p r y

e t e f e t e o a a t

giving ciphertext：

MEMATRHTGPRYETEFETEOAAT

Row Transposition Ciphers

Example:
密钥:列的读出顺序。
算法：以一个矩阵形式逐行写出明文，再逐列读出该消息，并以行的顺序排列。

- **key:** 4 3 1 2 5 6 7

plaintext:

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

ciphertext:TTNAAPTMTSUOAODWCOIXKNLYPETZ

-

使用多轮置换加密可提高安全性

key: **4 3 1 2 5 6 7**

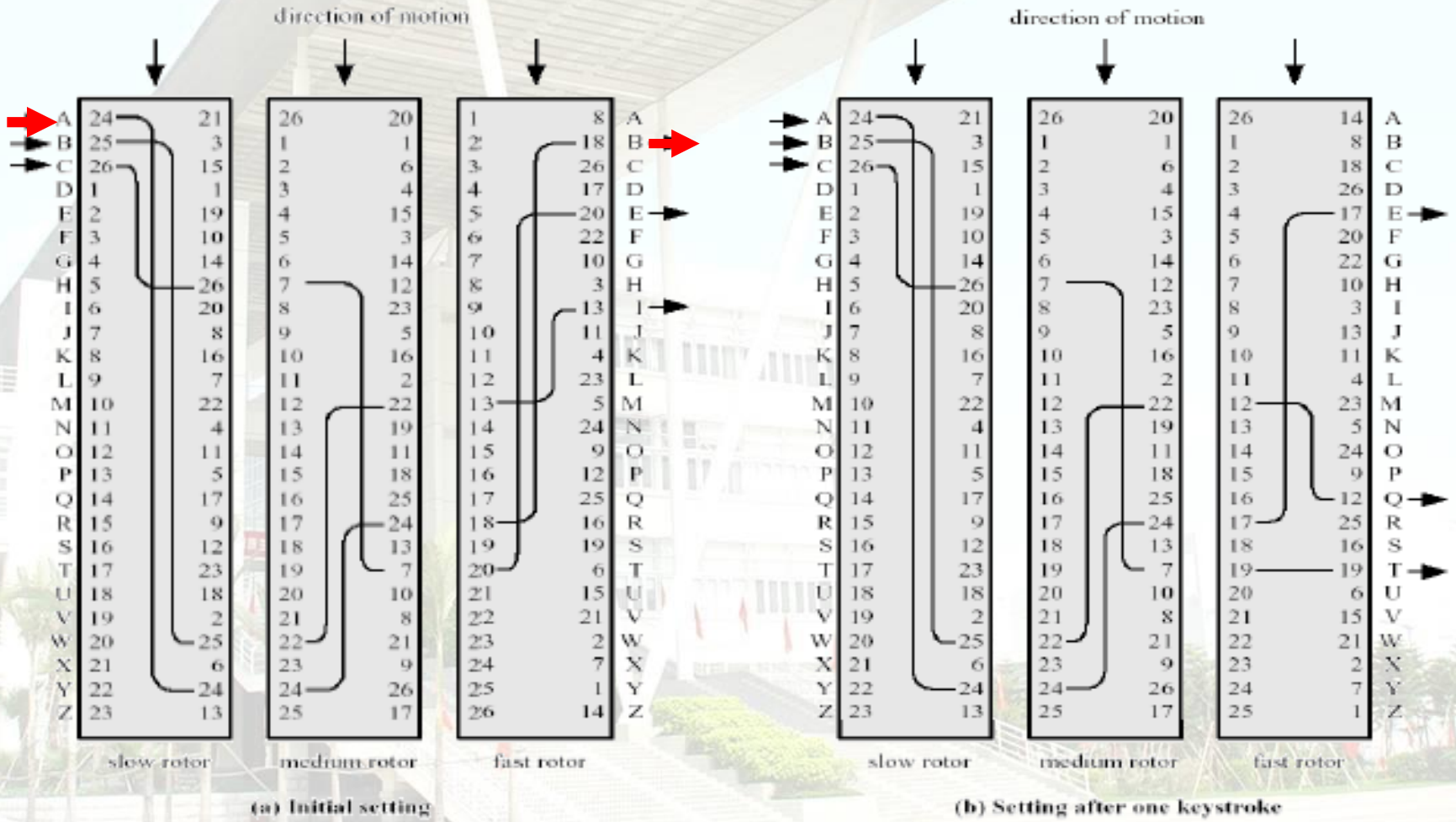
t	t	n	a	a	p	t
m	t	s	u	o	a	0
d	w	c	o	l	x	k
n	l	y	p	e	t	z

ciphertext:

NSCYAUOPTTWLTMDNAOIEPAXTTOKZ

Rotor machine 转子机

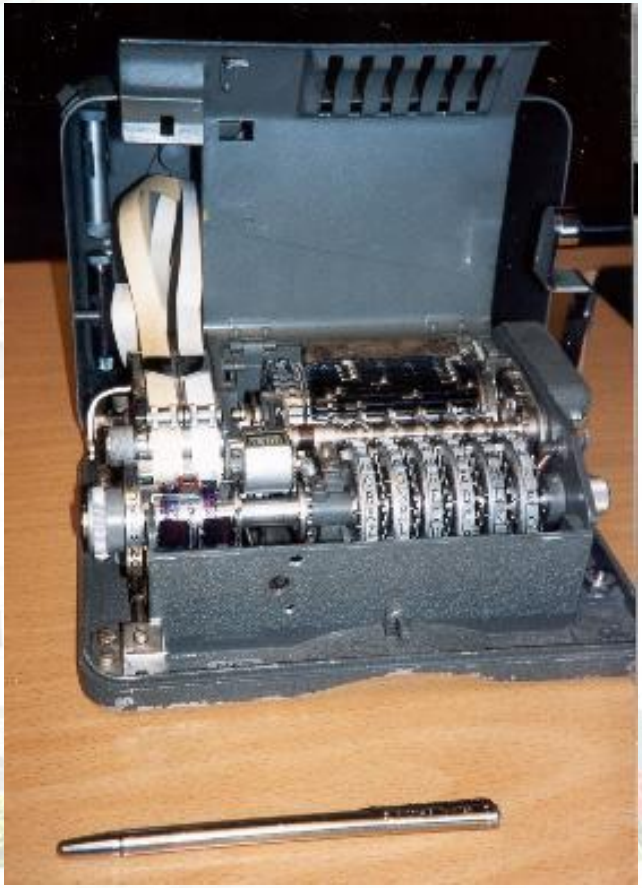
- 通过多个转子，完成字母的多次转换。



with 3 cylinders have

26^3=17576 alphabets
替换字母表

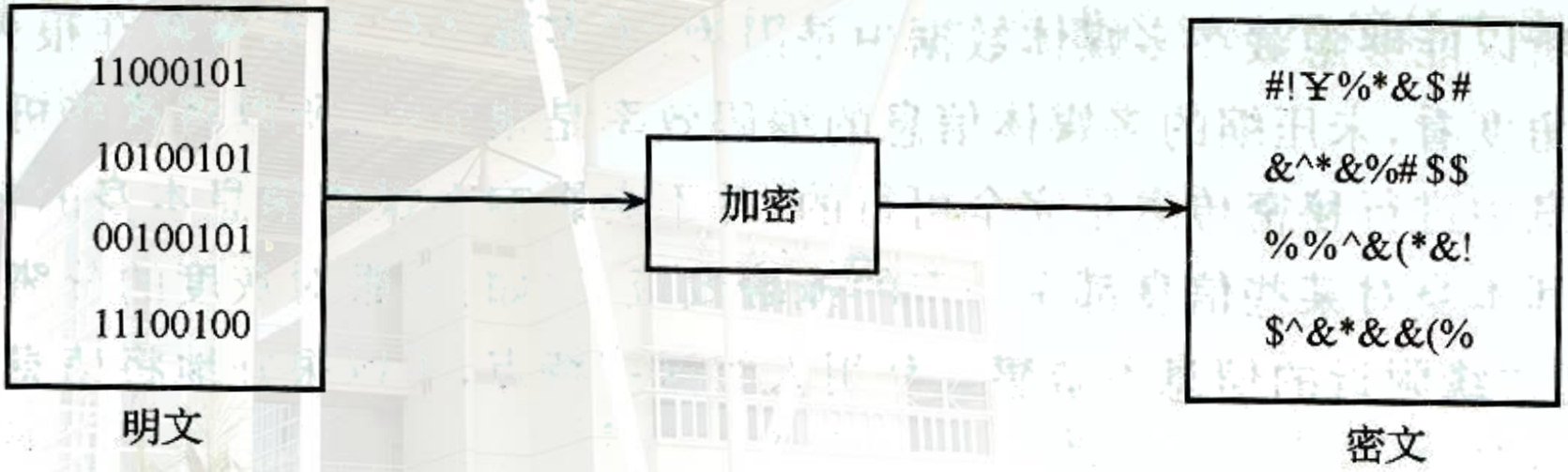
widely used in WW2
Hagelin Rotor Machine



Steganography 隐藏术

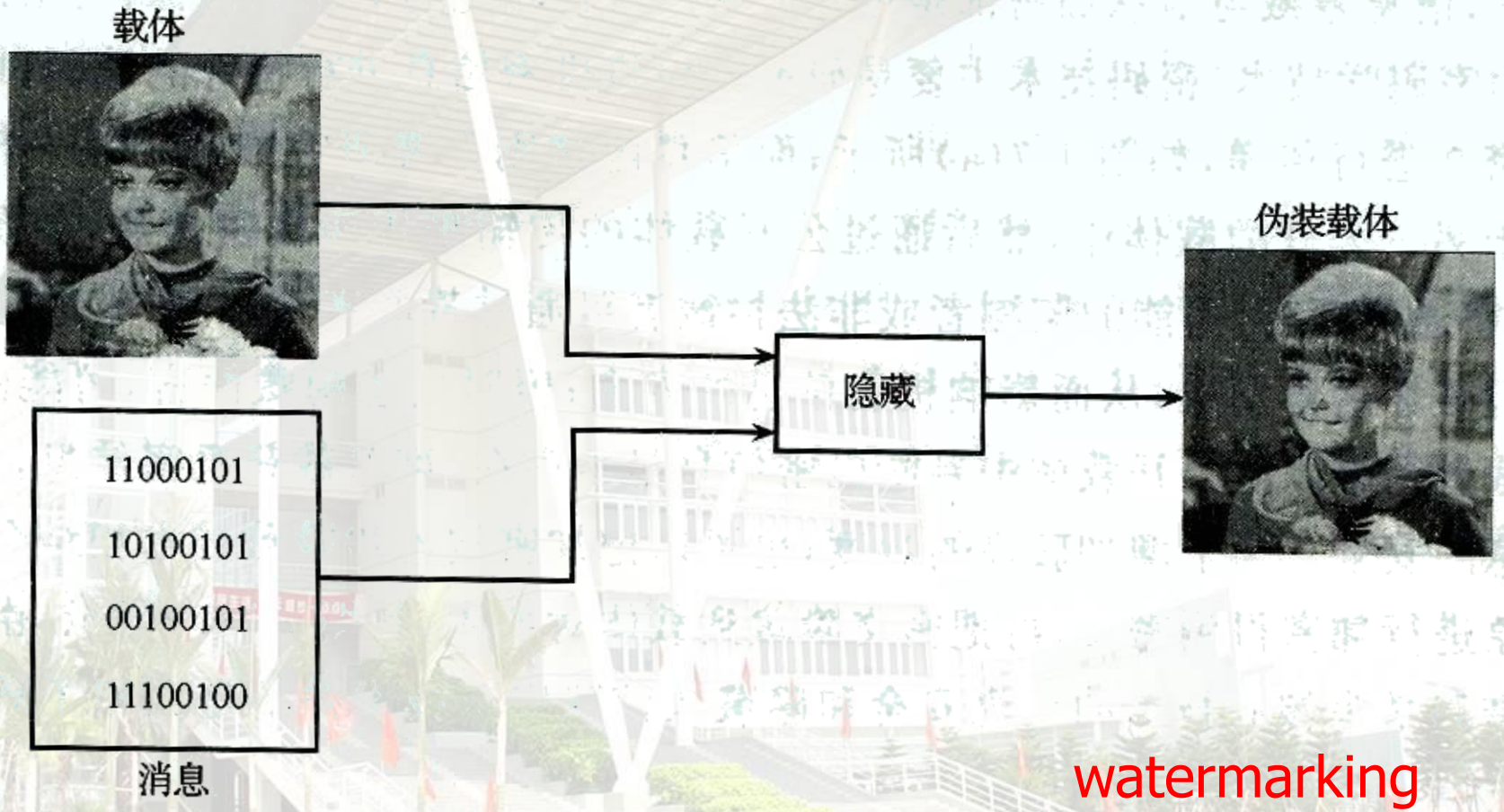
- an alternative to encryption
- hides existence of message
- has drawbacks
  - high overhead to hide relatively few info bits

信息加密 VS信息隐藏



信息加密

信息加密 VS信息隐藏 （数字水印）



watermarking

隐藏

信息加密 VS信息隐密（数字水印）

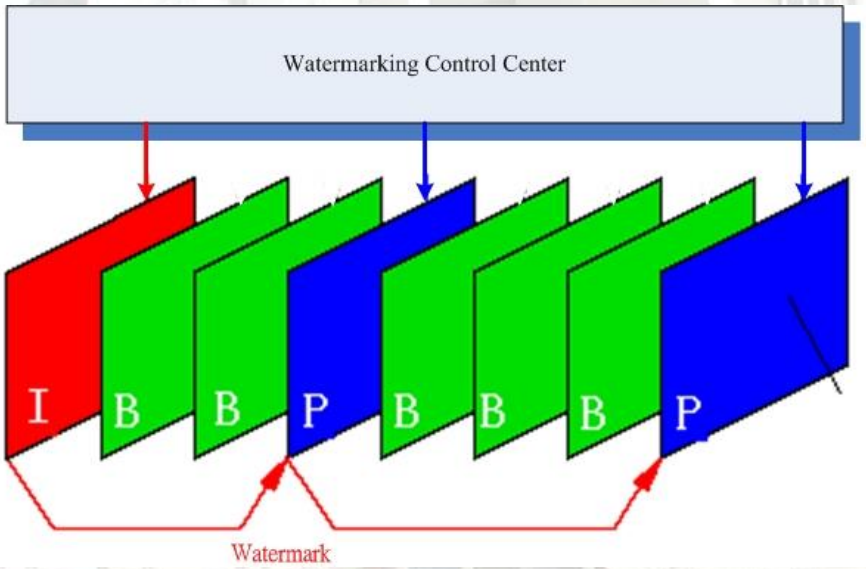
watermarking

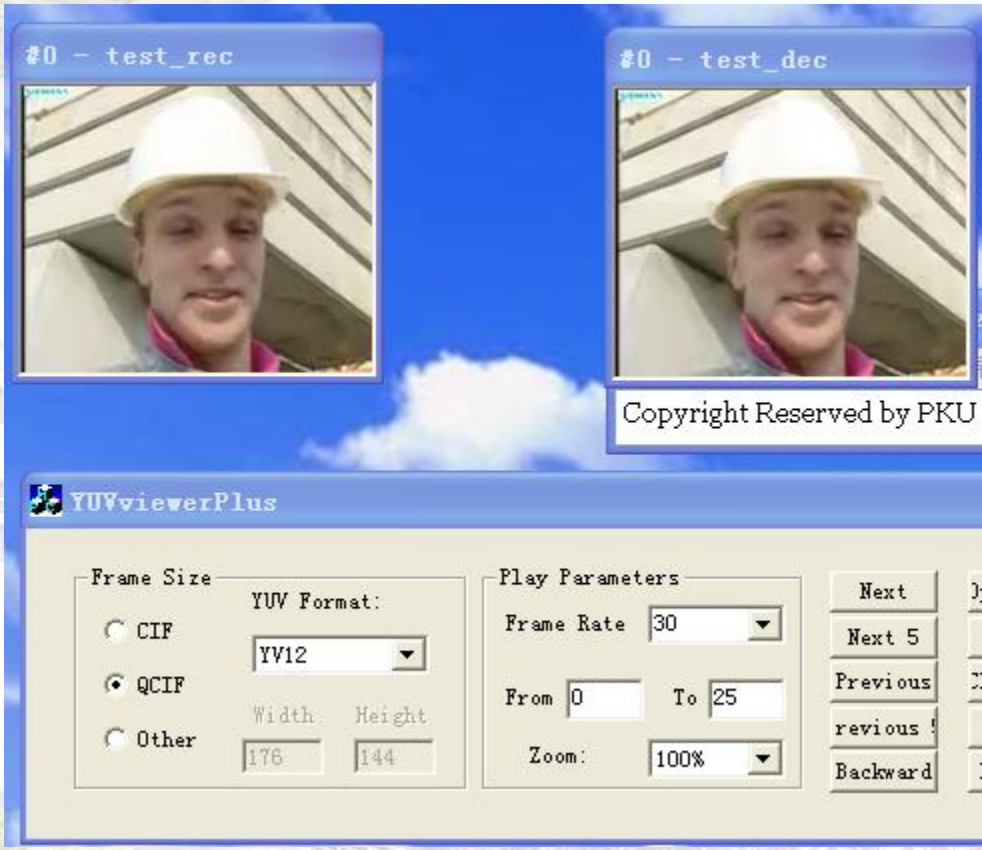
隐密

基于数字水印的H.264视频版权保护系统

系统功能：
信息隐藏
数字版权保护
授权播放

系统特性：
同时对I,P帧进行水印嵌入
嵌入信息加密，扩频预处理
运动矢量的安全保护





安全的图像数字水印平台

系统功能：

信息隐藏

数字版权保护

图像完整性认证

身份认证

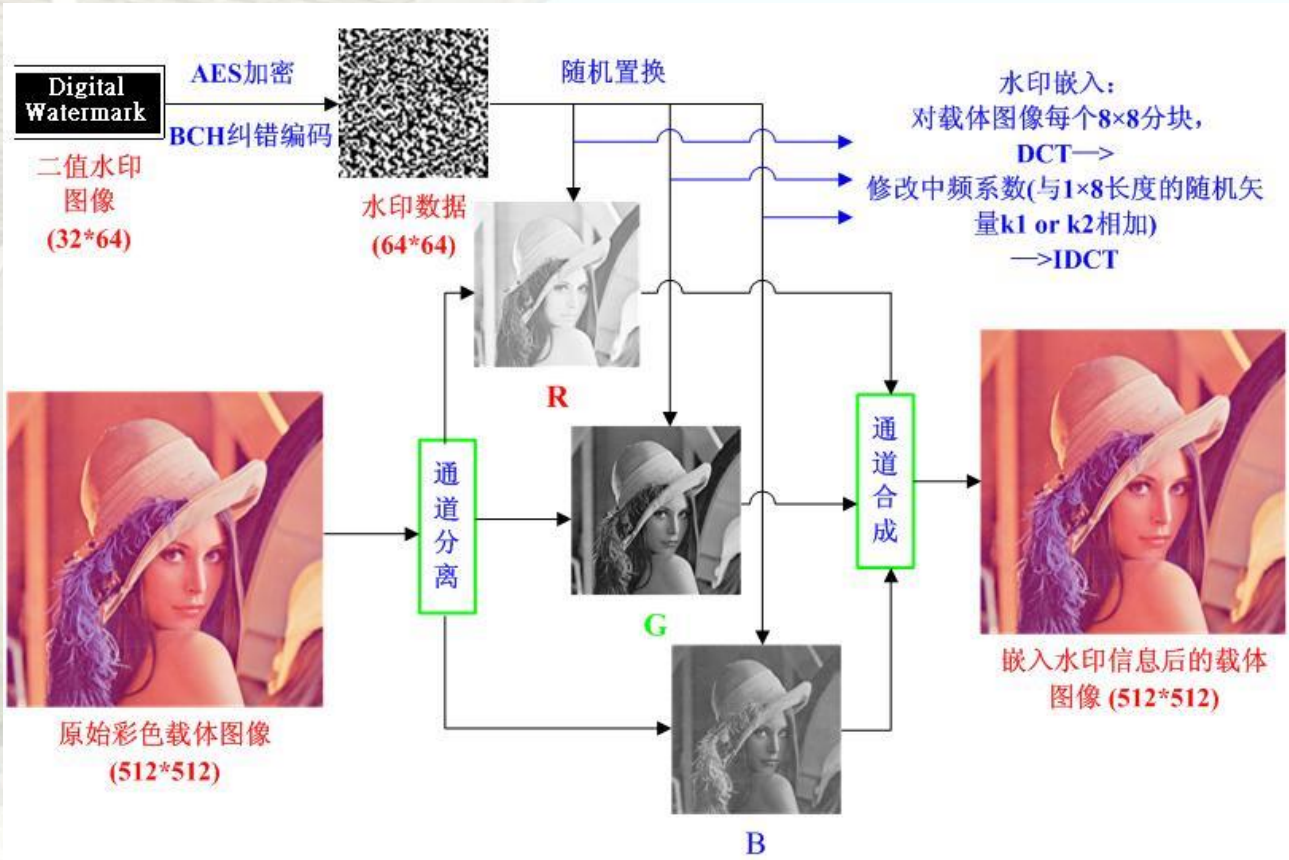
系统特性：

可定位图像受攻击位置

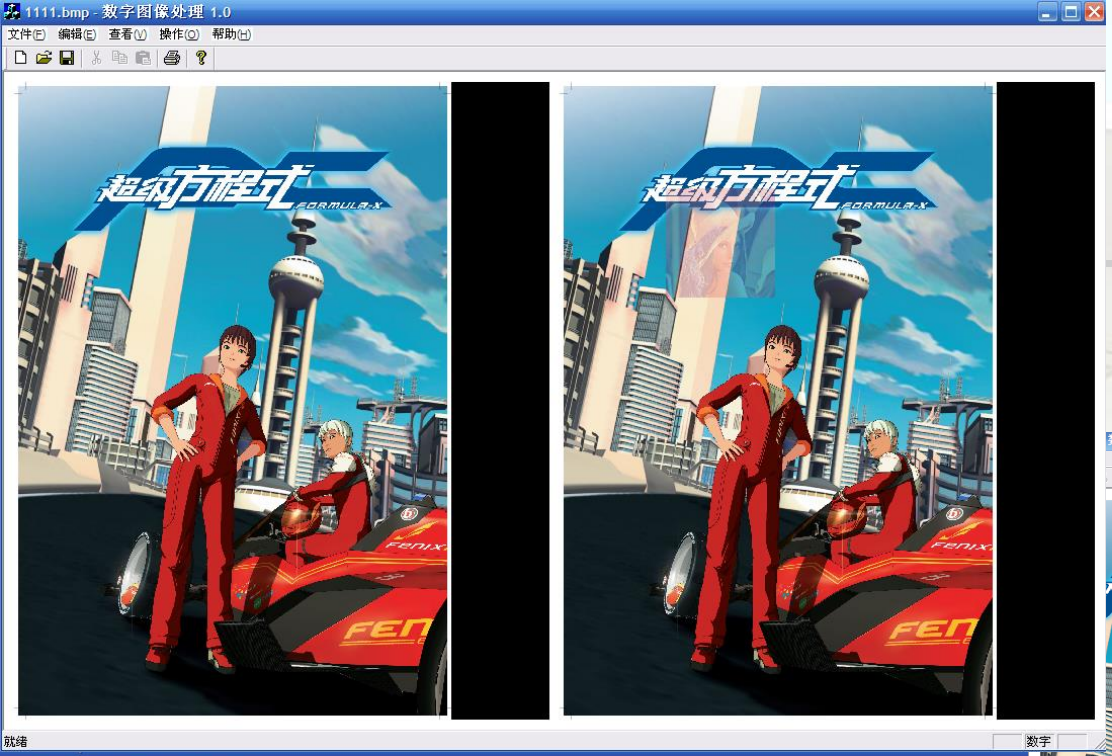
对嵌入信息进行加密，扩频等预处理

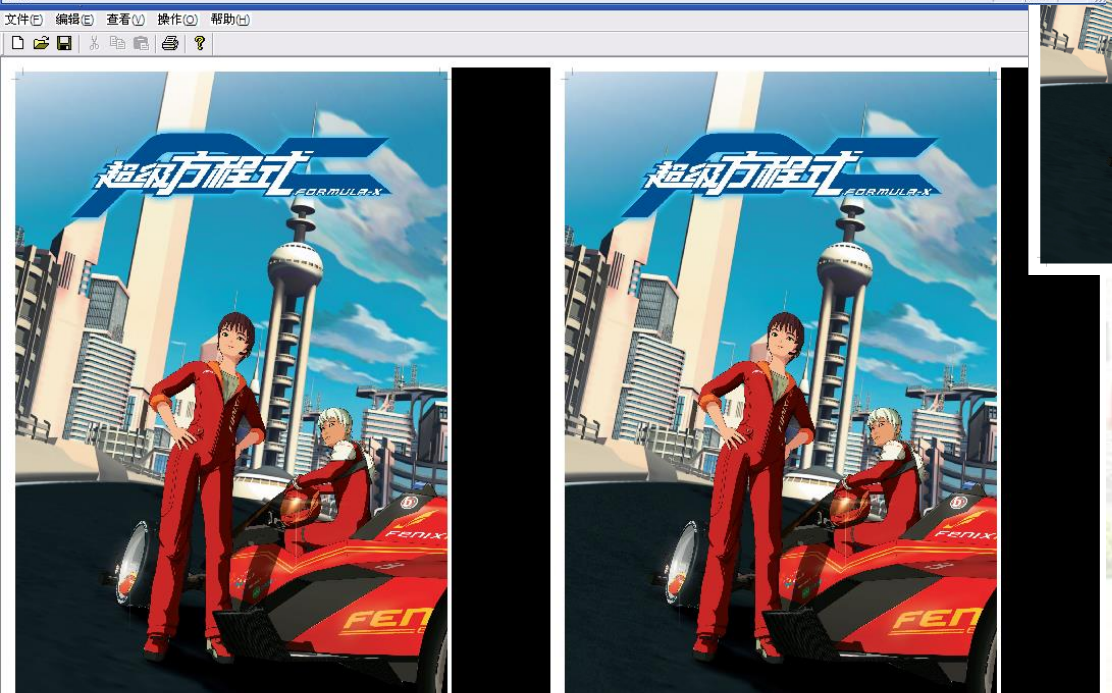
使用公钥密码体系保护，RSA,ECC等

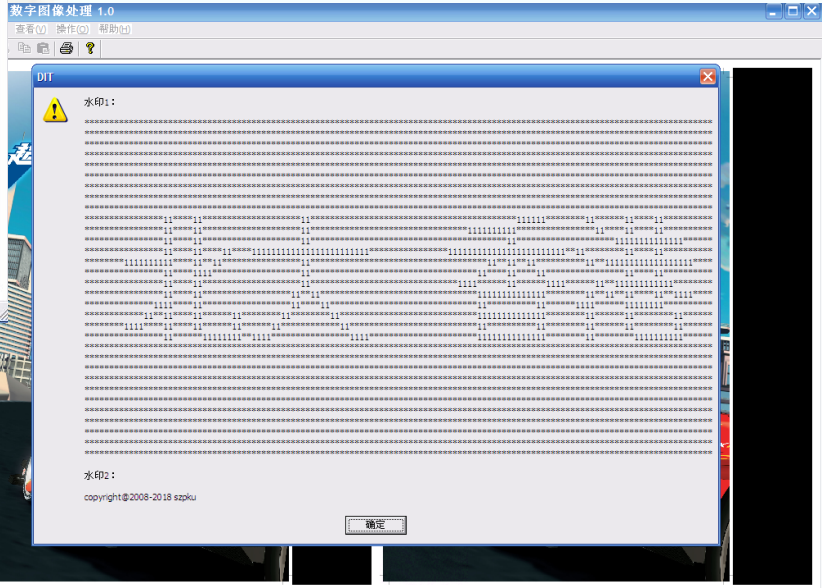
集成鲁棒水印与脆弱水印特性。



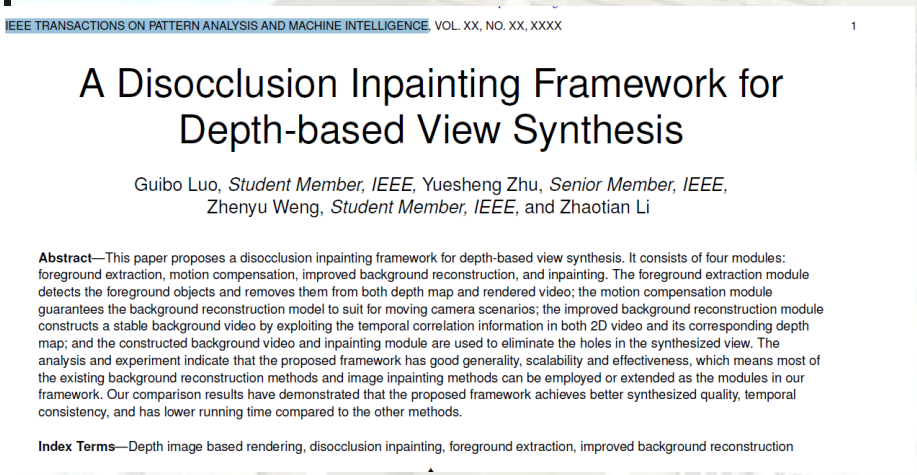
For the curious: 'lena' or 'lenna' is a digitized Playboy centerfold, from November 1972. (Lenna is the spelling in Playboy, Lena is the Swedish spelling of the name.) Lena Soderberg (ne Sjööblom) was last reported living in her native Sweden, happily married with three kids and a job with the state liquor monopoly. In 1988, she was interviewed by some Swedish computer related publication, and she was pleasantly amused by what had happened to her picture. That was the first she knew of the use of that picture in the computer business.


1111.bmp - 数字图像处理 1.0
文件(F) 编辑(E) 查看(V) 操作(O) 帮助(H)
超级方程式 FORMULA-X
就绪
数字

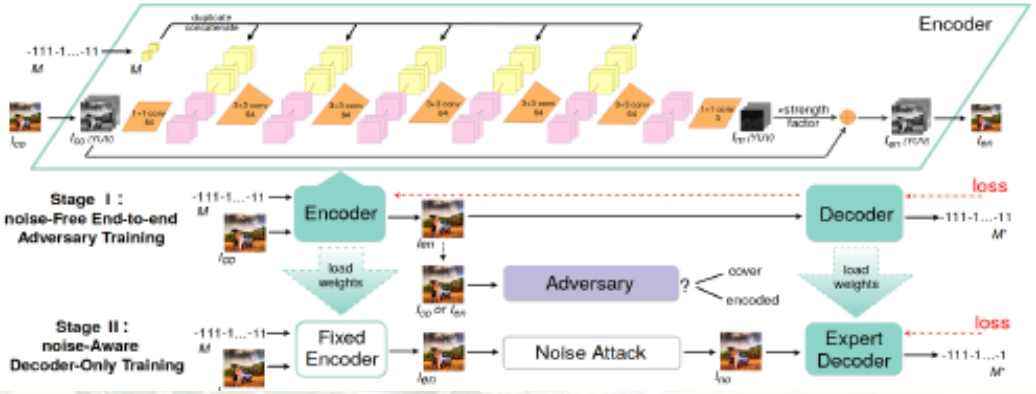

文件(F) 编辑(E) 查看(V) 操作(O) 帮助(H)
超级方程式 FORMULA-X


数字图像处理 1.0
查看(V) 操作(O) 帮助(H)
DIT
水印1：
水印2：
copyright@2008-2018 szpku
确定

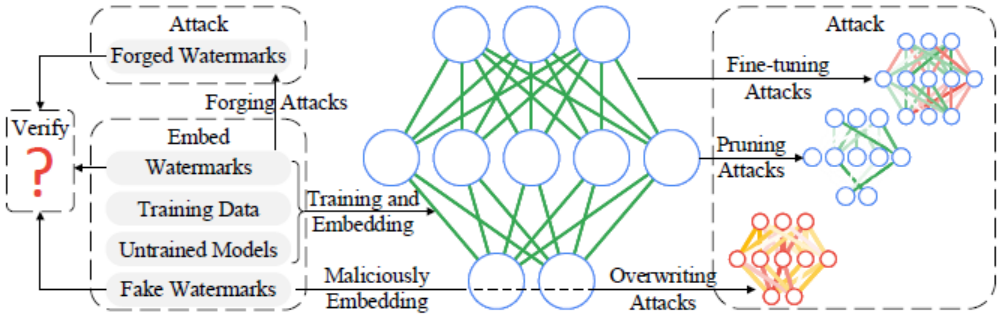
面向多媒体内容处理及数字权益(DRM)保护的深度学习方法



一种新的盲水印嵌入及检测方法 (ACM-2019)



面向云服务的AI模型/算法版权安全保护



基于水印技术的机器学习模型/算法版权保护

- 将原创人版权信息嵌入到计算模型中
- 出现算法被窃用时提取版权信息进行取证
- 适用与公有云的AI 计算服务

Watermarking Deep Neural Networks with Greedy Residuals

Hanwen Liu [1] Zhenyu Weng [1] Yuesheng Zhu [1]

Abstract

Deep neural networks (DNNs) are considered as intellectual property of their corresponding owners and thus are in urgent need of ownership protection, due to the massive amount of time and resources invested in designing, tuning and training them. In this paper, we propose a novel watermark-based ownership protection method by using the residuals of important parameters. Different from other watermark-based ownership protection methods that rely on some specific neural network architectures and during verification require external data source, namely ownership indicators, our method does not explicitly use ownership indicators for verification to defeat various attacks against DNN watermarks. Specifically, we greedily select a few and important model parameters for embedding so that the impairment caused by the changed parameters can be reduced and the robustness against different attacks can be improved as the selected parameters can well preserve the model information. Also, without the external data sources for verification, the adversary can hardly cast doubts on ownership verification by forging counterfeit watermarks. The extensive experiments show that our method outperforms previous state-of-the-art methods in five tasks.

1. Introduction

Due to the impressive performance on predictive tasks, machine learning, and deep neural networks (DNNs) particularly have become an increasingly popular method for a variety of usages in real-world applications. Such applications need models to be well-trained, which requires massive training data and computing resources (Strubell et al., 2019). The process of training data collecting, cleansing, storing, and model training can be quite troublesome and time-consuming, not to mention the potential security and privacy issues (Shokri et al., 2017; Song et al., 2017; Salem et al., 2019), and thus the models are considered to be the valuable intellectual property of their legitimate owners. However, since the models are supposed to be exposed and serve users with some helpful information for commercial use, the threat of well-trained model theft seems to be inevitable (Oh et al., 2018; Orekondy et al., 2019). Also, although well-trained models are closely protected in general, if the cost of the theft is much lower than that of legal purchase, the value incorporated in such models makes stealing a lucrative task for malicious adversaries.

One solution to the model theft problem is watermarking DNN models (Uchida et al., 2017), which is an insightful way to identify model ownership of the legal owner. Watermarking objects is a well-studied problem in the security community under the general theme of digital watermarking. Essentially, the aim of watermarking is to mark the model with some kind of identity information, in favor of further ownership tracking. To ensure that the embedded identity information will not greatly affect the model performance, this watermarking process usually takes place during model training. An external data source, e.g., a carefully selected picture (Adi et al., 2018), is usually used for ownership verification of the watermarked model. We refer to this data source as the *ownership indicator* by analogy with the acid-based indicator in elementary chemistry.

Designing a practical yet robust watermarking method for DNN models, however, is not easy. On the one hand, with the embedding of additional information, the model may not be able to maintain its original performance. On the other hand, the model can be easily modified due to the intrinsic nature of DNN, making it possible to perform removal attacks, such as fine-tuning for transfer learning (Pan & Yang, 2010) and pruning for model compression (Li et al., 2017). Besides, in some cases, the adversary can always make an adversarial example of the ownership indicator (Fan et al., 2019) or directly overwriting the existing watermark to cast doubt on the ownership verification. The various attacks that DNN watermarks need to face are summarized in Figure 1 and described in Section 3.1 in detail. Furthermore, as for practicality and robustness, most current methods have to rely on specific DNN architectures, which greatly limits the application scenarios of watermarking.

[1]School of Electronic and Computer Engineering, Peking University. Correspondence to: Yuesheng Zhu <zhuys@pku.edu.cn>.

Proceedings of the 38th International Conference on Machine Learning, PMLR 139, 2021.

Watermarking Deep Neural Networks with Greedy Residuals，Thirty-eighth International Conference on Machine Learning , 2021，

「生物认证」技术 Biometrics Technology

- 利用使用者身上与生俱有的特征辨识使用者的身份

生物特征大体可分为身体特征和行为特征两类。

身体特征：指纹、掌纹、眼纹、静脉纹、脑纹、唇纹、皮纹、脸型、脚印、DNA、耳纹、虹膜等；

行为特征: 签名、击键、声纹、握手、行走步态等。

多模态/多因子信息的安全识别与鉴别

A Robust Single-sensor Face and Iris Biometric Identification System based on Multimodal Feature Extraction Network

Zhengding Luo, Qinghua Gu, Yuesheng Zhu, Zhiqiang Bai
Communication and Information Security Lab
Shenzhen Graduate School, Peking University, China
Emails: {luozd, guqh, zhuys, baizq}@pku.edu.cn

***Abstract*—Joint face-iris identification can integrate complementary information from face and iris to fulfill the requirement of performance improvement and security. However, most of the current face-iris multimodal biometric systems acquire face and iris with different sensors which brings about the increase of capturing complexity and device cost. Besides, they are limited by the identification performance degradation under non-ideal scenarios. In order to address these problems, a robust single-sensor face and iris biometric identification system based on multimodal feature extraction (MFE) network is proposed. Only a single sensor is needed to obtain face and iris images in the proposed system, with the goal of improving recognition performance while minimizing sensor cost and acquisition time. The MFE network is designed as a general network module to extract both face and iris features and it is trained with a triplet framework to reduce intra-class variations and enlarge inter-class variations. Our experimental results on CASIA.v4-distance and FRGC v2.0 non-ideal datasets show that the proposed system achieves better identification performance in terms of Equal Error Rate (EER) and False Reject Rate (FFR), etc. compared with other unimodal and multimodal biometric systems.**

***Index Terms*—multimodal biometrics, face and iris recognition, non-ideal biometrics, deep learning, a single sensor**

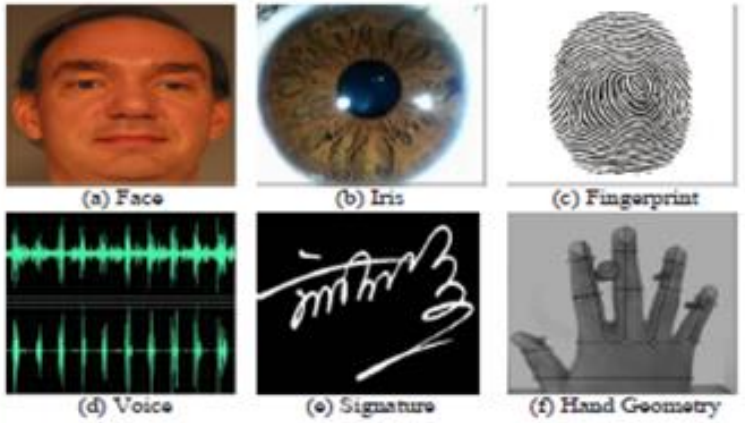
I. Introduction

Traditional identification techniques include password-based schemes and token-based schemes. However, these schemes are vulnerable to attacks when the passwords are divulged or the tokens are stolen. Biometric recognition refers to automatic identification using certain physiological or behavioral traits associated with an individual. These biometric traits include face, fingerprint, iris, palmprint and voice, etc. which have an edge over traditional identification approaches because they cannot be stolen or shared [1]. But unimodal biometric systems are limited by some inherent drawbacks such as lack of uniqueness, restricted degrees of freedom, non-universality, sensitivity to noisy data, vulnerability to spoofing and unacceptable error rates [2]. Multimodal biometrics can fuse information from multiple modalities to overcome the limitations of single modality and enhance discriminant ability [3]. Apart from that, multi-biometric systems increase the resistance to spoofing attacks by making it difficult to spoof multiple modalities simultaneously [4].

Among biometric traits, face and iris have received significant attention because of many outstanding characteristics. Face recognition is the most natural and acceptable way in biometric recognition, whereas photographs, videos, 3D masks and other spoofing ways make face recognition less reliable [5] [6]. Iris is one of the most promising biometric traits due to its unique textures, which is stable until the end of human life unless there are accidents [7]. Iris images are usually acquired within a short distance under near-infrared illumination [8]. However, iris recognition may suffer from identification performance degradation when image acquisition is not constrained strictly. Therefore, the pros and cons of face and iris can complement each other to enhance the overall recognition performance and security [9].

While research into face-iris multimodal biometrics has received a large increase over recent years, many related experiments are based on **chimeric datasets** (i.e. face modality and paired iris modality come from different users) due to a lack of available **real-user datasets** (i.e. face modality and corresponding iris modality come from the same person) [10]. The independent acquisition of each modality from different sensors may increase sensor cost, data acquisition time and the risk of spoofing in chimeric datasets. It is much more desirable to acquire multiple modalities from a single sensor for security and usability reasons in practice [11], [12]. In addition, since iris images can be extracted from face images without incurring additional hardware cost as shown in Fig. 1, it is economical and convenient to obtain face and iris samples using a single sensor device.

It is common and unavoidable to deal with some noisy factors such as off-angles, reflections, illumination changes and blurred images under less-constrained or non-ideal scenarios. These noisy factors result in the recognition performance drop and lack of security in face-iris multimodal biometric systems. Therefore, development of a robust face-iris multimodal biometric system is highly desirable. The key contributions of this paper are summarized as follows: (1) We propose a face-iris multimodal biometric system combining information from face and iris to enhance the limited discriminant ability of unimodal biometrics. Besides, the proposed system exhibits superior robustness under non-cooperative environments. (2) A multimodal feature extraction (MFE) network is developed to extract face and iris features in our system. The MFE network is a modality-general network and it is trained with a triplet



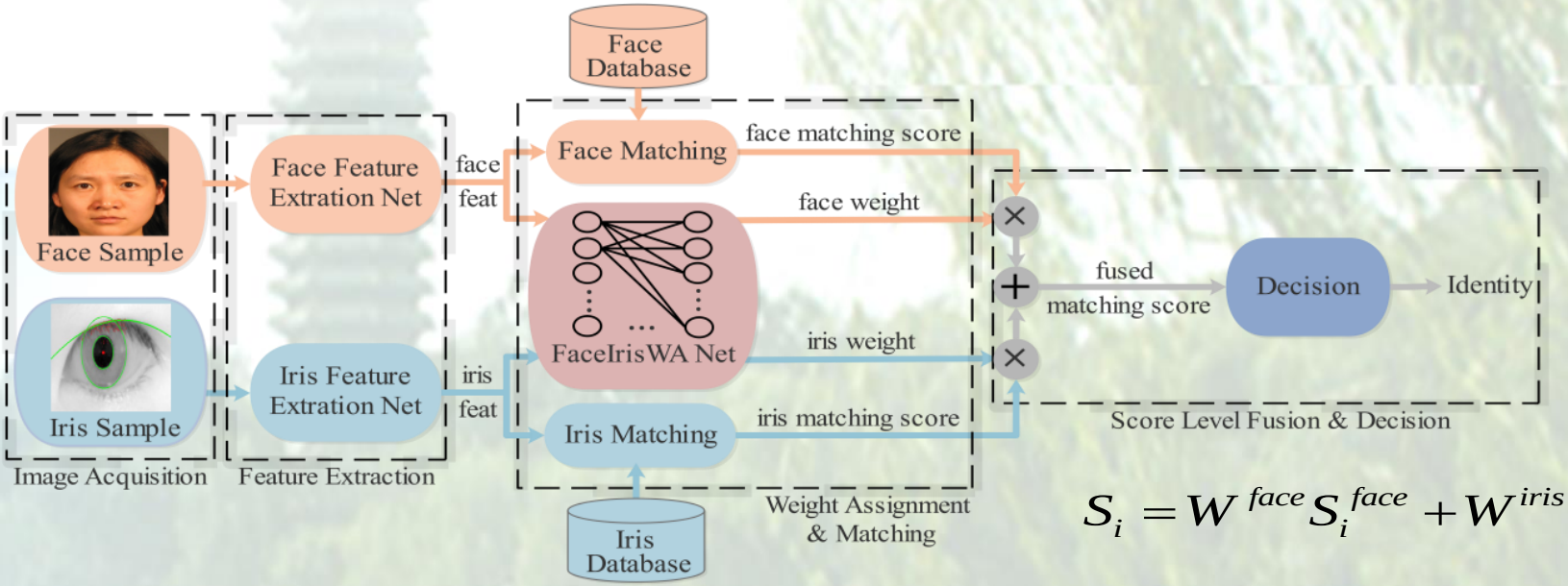
生物特征识别

- 人脸 – GAN 攻击　　3D打印人脸套

- 疫情场景（口罩）的身份安全识别与鉴别

虹膜与人脸融合安全识别与鉴别

自适应的人脸-虹膜分数层融合



$$S_i = W^{face} S_i^{face} + W^{iris} S_i^{iris}$$

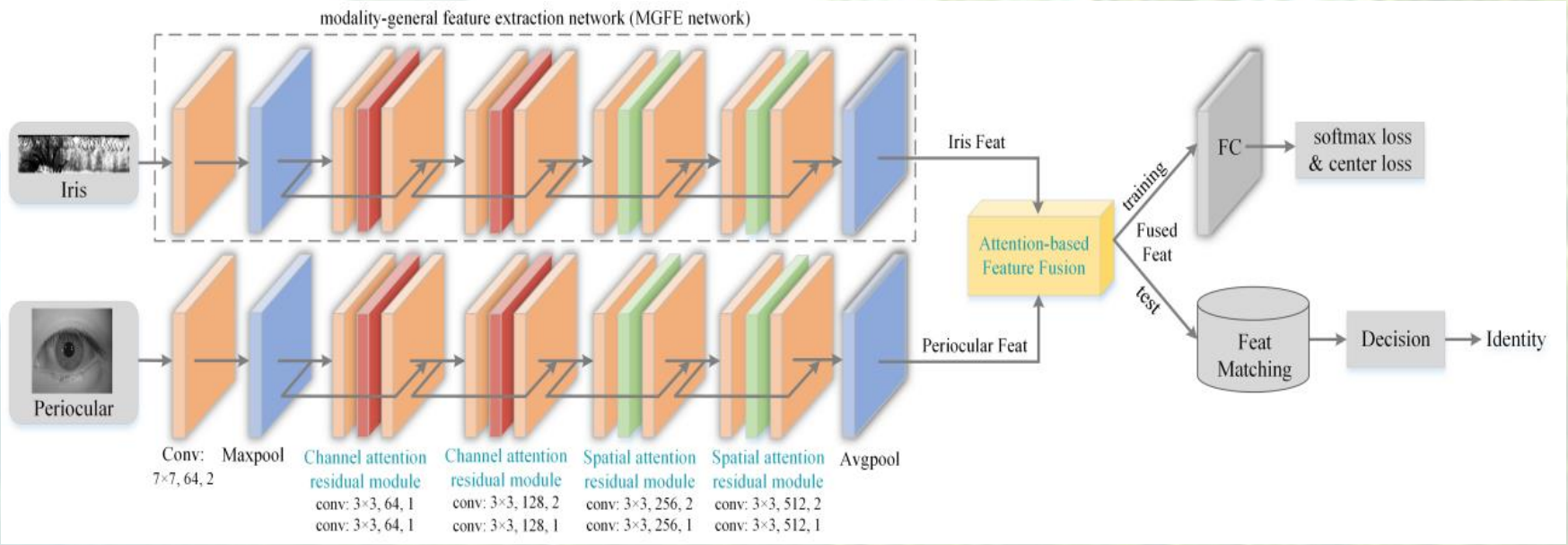
匹配分数层融合，设计用于自适应分配人脸和虹膜权重的Face-Iris Weight Assignment Network (FaceIrisWA Net)，实现更加灵活高效的权重分配，

Luo, Z., Zhu, Y., et al. "A Robust Single-Sensor Face and Iris Biometric Identification System Based on Multimodal Feature Extraction Network", in the IEEE 31st International Conference on Tools with Artificial Intelligence, 2019, pp. 1237-1244.

Luo, Z., Gu, Q., et al. "An Adaptive Face-Iris Multimodal Identification System based on Quality Assessment Network", in the 27th International Conference on Multimedia Modeling, 2020.

虹膜与眼周融合安全识别与鉴别

基于注意力机制的虹膜-眼周特征层融合



由特征提取网络（MGFE network）和基于注意力的特征融合模块组成。既可提取虹膜也可提取眼周特征。MGFE网络包括卷积层、最大池化层、四个残差模块和平均池化层

Luo, Z., Zhu, Y., et al. "An Efficient Deep Learning Framework based on Multiple Attention Mechanisms for Joint Iris-Periocular Biometric Recognition", IEEE Signal Processing Letters, 2021.

应用

- 1.门禁系统(辨公室、家庭)
- 2.身份鉴定(自动提款机、电子商务)
- 3.电脑开机、文件管理
- 4.自动安全监控
- 5.打击犯罪
- 6.海关检查