



# 现代密码学

- 主讲人: 张凤荣 ([zhangfengrong@xidian.edu.cn](mailto:zhangfengrong@xidian.edu.cn))

赵臻

- 西安电子科技大学

## 引言--信息安全的威胁

- 为什么需要信息（网络）安全？

- ❖ “无论是否愿意承认，只要在家里或办公室里连上了Internet，我们就都是易受攻击的（Vulnerable）。”
- ❖ “Internet的互联性使‘天涯若比邻’成为现实。我们的‘街坊’现在可能是一个身处异国不怀好意的黑客，也可能是一个年幼而聪明的孩子，正在寻找我们系统的漏洞，只是为了好玩儿。”

----选自《黑客大曝光》

# 网络不安全的原因

自身缺陷

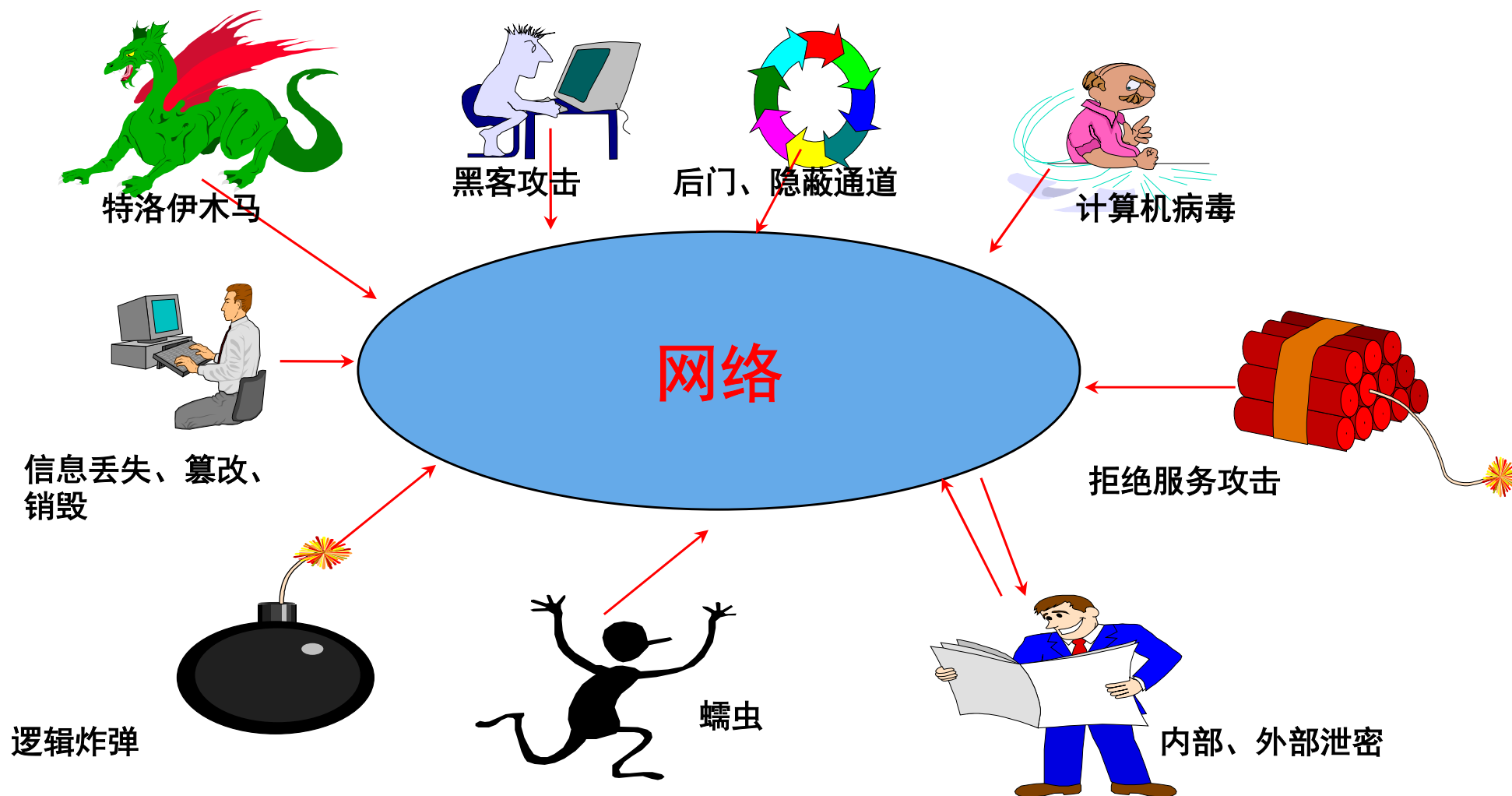



开放性



黑客攻击

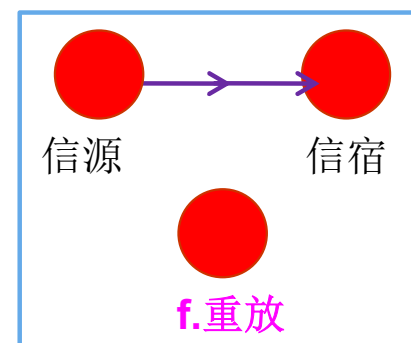
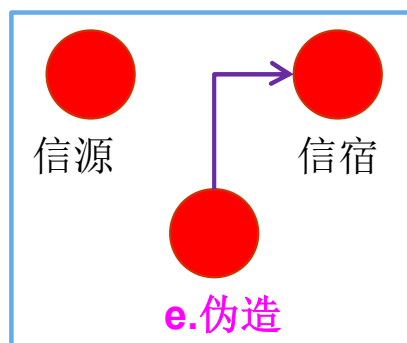
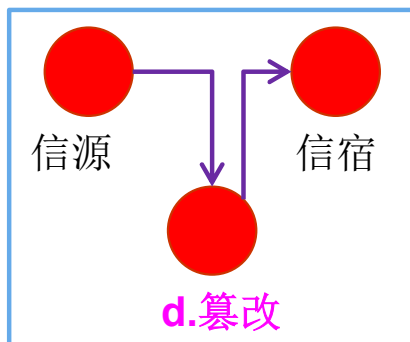
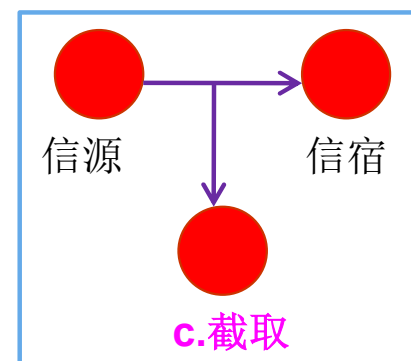
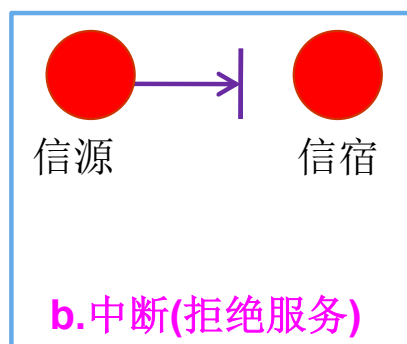
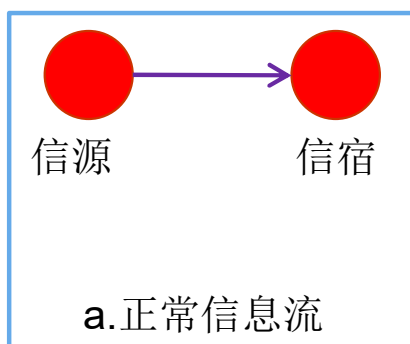
# 网络安全主要威胁来源





Impossible  $\Leftrightarrow$  I'm possible

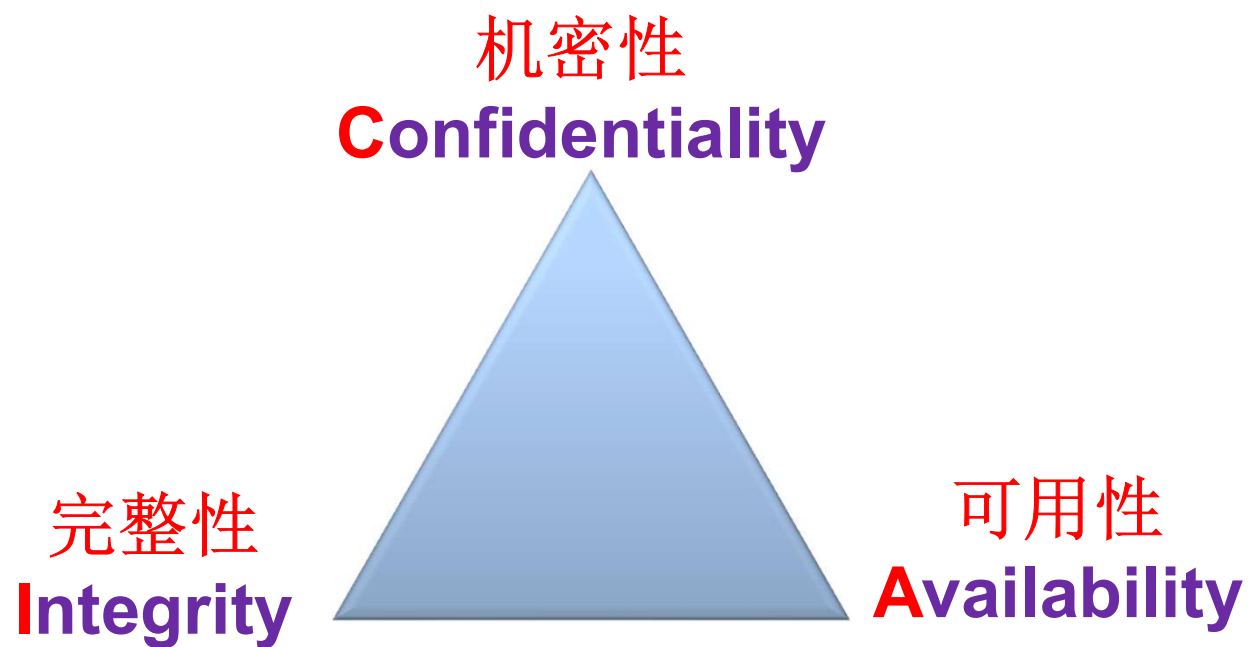
# 信息安全的攻击方式有哪些？？



# 信息安全是什么？？

- ❖ **信息安全**：防止任何对**数据**进行未授权访问的措施，或者防止造成信息有意无意泄露、破坏、丢失等问题的发生，让数据处于远离危险、免于威胁的状态或特性
- ❖ **网络安全**：计算机网络环境下的信息安全

# CIA三元组





# 信息安全的目标

1. **机密性 (Confidentiality)** : 保证信息不泄露给非授权的用户和实体
2. **完整性 (Integrity)** : 保证信息处于“保持完整或一种未受损的状态”
  - 防止任何对信息应有特性或状态的中断、窃取、篡改、伪造等
3. **可用性 (Availability)** : 授权用户按需随时访问所需信息而不被非法拒绝
4. **认证性 (Authentication)** : 确保消息的来源, 可分为消息认证和实体认证
5. **不可否认性 (Non-Repudiation)** : 保障用户无法在事后否认曾经对消息的生成、签发、接收等

# 举例说明

1. **机密性 (Confidentiality) :**
  - 你与我说话时，别人能不能偷听？
2. **完整性 (Integrity) :**
  - 收到的传真不太清楚？传送过程中有人篡改过没有？
3. **认证性 (Authentication) :**
  - 我不认识你，你是谁？
  - 我怎么确认你就是你，要是别人冒充你怎么办？
4. **不可否认性 (Non-Repudiation) :**
  - 我收到货后，不想付款，想抵赖怎么办？
  - 我将钱寄给你后，你不想发货，又如何？

# 密码学在信息安全中的作用

## ❖ 现实生活：

- 阻止窃贼闯入—安装防盗门
- 从银行骗取钱财—验证身份证
- 签署合同—签名

## ❖ 信息时代？？

- 信息被窃取—（密码学的）**加密技术**
- 数据完整性—（密码学的）**散列函数（Hash函数）**
- 认证性—（密码学的）**身份鉴别技术**
- 不可否认性—（密码学的）**数字签名**
- .....

密码技术是信息安全的主要手段之一，但绝不是确保信息安全的唯一技术，也不能解决信息安全中的所有问题

# 密码、口令 (password)



QQ login interface. It features a penguin icon on the left. The main area has a text input field labeled "QQ号码/手机/邮箱" with a dropdown arrow, a "注册帐号" link, a "密码" input field with a "找回密码" link, and checkboxes for "记住密码" and "自动登录". A "登录" button is at the bottom.



Google account login interface. It has a "Google 帐户" header. Below it are "用户名:" and "密码:" labels, each followed by a text input field. An example email "ex: pat@example.com" is shown below the username field. There is a checkbox for "保持登录状态" and a "登录" button.



Office system login interface. It has a computer icon and the title "办公系统". Below are "用户名:" and "密码:" labels, each followed by a text input field. A "登录" button with a key icon is on the right.



Email system login interface. It has an envelope icon and the title "邮件系统". Below are "用户名:" and "密码:" labels, each followed by a text input field. On the right are "登录" and "注册" buttons.

**演示：获取用户名和密码等**



❖ 《密码法》从2020年1月1日执行。



# 参考书目

1. 现代密码学,杨波 编著,清华大学出版社
2. 现代密码学教程(第2版) 谷利泽,郑世慧,杨义先著





# 考核方式

48学时

考试成绩60%

平时成绩40%



# 第一讲：保密学基础

- 一、保密学的基本概念
- 二、密码体制分类
- 三、古典密码
- 四、初等密码分析





# 一、保密学的基本概念

**保密学** (Cryptology): 研究信息系统安全保密的科学。它包含两个分支:

**密码学** (Cryptography), 对信息进行编码实现隐蔽信息的科学。

**密码分析学** (Cryptanalytics), 研究分析破译密码的科学。两者相互对立, 而又互相促进地向前发展。



# 几个概念 (一)

- 明文 (消息) (Plaintext) : 被隐蔽消息。
- 密文 (Ciphertext) 或密报 (Cryptogram): 明文经密码变换成的一种隐蔽形式。
- 加密 (Encryption): 将明文变换为密文的过程。
- 解密 (Decryption): 加密的逆过程, 即由密文恢复出原明文的过程。
- 加密员或密码员 (Cryptographer): 对明文进行加密操作的人员称作。



## 几个概念 (二)

- **加密算法** (Encryption algorithm): 密码员对明文进行加密时所采用的一组规则。
- **接收者** (Receiver): 传送消息的预定对象。
- **解密算法**: 接收者对密文进行解密时所采用的一组规则。
- **密钥** (Key): 控制加密和解密算法操作的数据处理, 分别称作加密密钥和解密密钥。
- **截收者** (Eavesdropper): 在信息传输和处理系统中的非授权者, 通过搭线窃听、电磁窃听、声音窃听等来窃取机密信息。



## 几个概念 (三)

- **密码分析** (Cryptanalysis): 截收者试图通过分析从截获的密文推断出原来的明文或密钥。
- **密码分析员** (Cryptanalyst): 从事密码分析的人。
- **被动攻击** (Passive attack): 对一个保密系统采取截获密文进行分析的攻击。
- **主动攻击** (Active attack): **非法入侵者** (Tamper)、**攻击者** (Attacker) 或**黑客** (Hacker) 主动向系统窜扰, 采用删除、增添、重放、伪造等窜改手段向系统注入假消息, 达到利己害人的目的。



# 保密系统模型

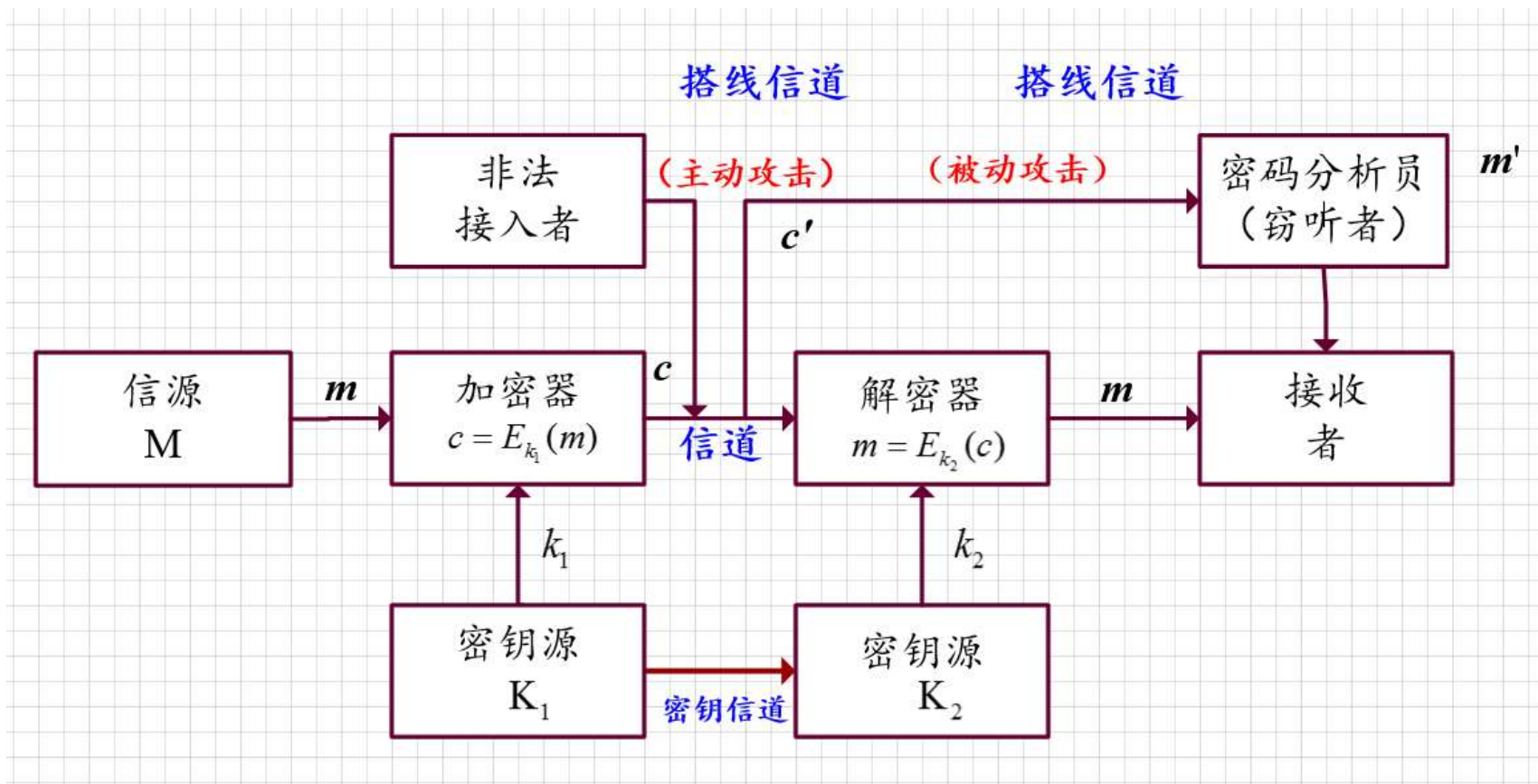


图1 保密系统模型图



# 保密系统模型

**保密系统** (Secrecy System):  $(M, C, K_1, K_2, E_{k_1}, D_{k_2})$

- 明文消息空间  $M$
- 密文消息空间  $C$
- 密钥空间  $K_1$  和  $K_2$ 、在单钥体制下  $K_1 = K_2 = K$ ，此时密钥  $k \in K$  需经安全的密钥信道由发方传给收方。
- 加密变换:  $E_{k_1} \in E$ ,  $m \rightarrow c = E_{k_1}(m)$ , 其中  $k_1 \in K_1$ ,  $m \in M$ ,  $c \in C$ , 由加密器完成。
- 解密变换:  $D_{k_2} \in D$ ,  $c \rightarrow m = D_{k_2}(c)$ , 其中  $k_2 \in K_2$ ,  $m \in M$ ,  $c \in C$ , 由解密器实现。



# 保密系统应满足下述要求

- 系统即使达不到理论上是不可破的，即  $\Pr\{m' = m\} = 0$ ，也应当为实际上不可破的。就是说，从截获的密文或某些已知明文密文对，要决定密钥或任意明文在计算上是不可行的。
- 系统的保密性不依赖于对加密体制或算法的保密，而依赖于密钥。这是著名的 Kerckhoff 原则。
- 加密和解密算法适用于所有密钥空间中的元素。
- 系统便于实现和使用方便。





# 认证与认证系统

- **认证系统** (Authentication system): 为了防止消息被篡改、删除、重放和伪造的一种有效方法使发送的消息具有被验证的能力, 使接收者或第三者能够识别和确认消息的真伪。实现这类功能的密码系统称作认证系统。
- **保密性**: 保密性是使截获者在不知密钥条件下不能解读密文的内容。
- **认证性**: 使任何不知密钥的人不能构造一个密报, 使意定的接收者脱密成一个可理解的消息 (**合法的消息**)。





# 认证与认证系统

认证理论和技术是保密学研究的一个重要领域。如传统的手书签字正在被更迅速、更经济和更安全的数字签字 (Digital signature) 代替。



# 安全认证系统应满足下述条件

- 意定接收者能够检验和证实消息的合法性和真实性。
  - 消息的发送者对所发送的消息不能抵赖。
  - 除了合法消息发送者外，其它人不能伪造合法的消息。
- 而且在已知合法密文  $c$  和相应消息  $m$  下，要确定加密密钥或系统地伪造合法密文在计算上是不可行的。
- 必要时可由第三者作出仲裁。



# 完整性

**完整性**(Integrity): 在有自然和人为干扰条件下, 系统保持检测错误和恢复消息和原来发送消息一致性的能力。实际中常常借助于纠、检错技术和杂凑技术来保证消息的完整性。



## 二、密码体制分类

### 密码体制有2大类：

- 单钥体制 (One-key system)：加密密钥和解密密钥相同
- 双钥体制 (Two key system)：加密密钥和解密密钥不同

### 单钥体制主要问题（网络环境下）：

- 密钥产生 (Key generation)
- 密钥管理 (Key management)



# 单钥体制

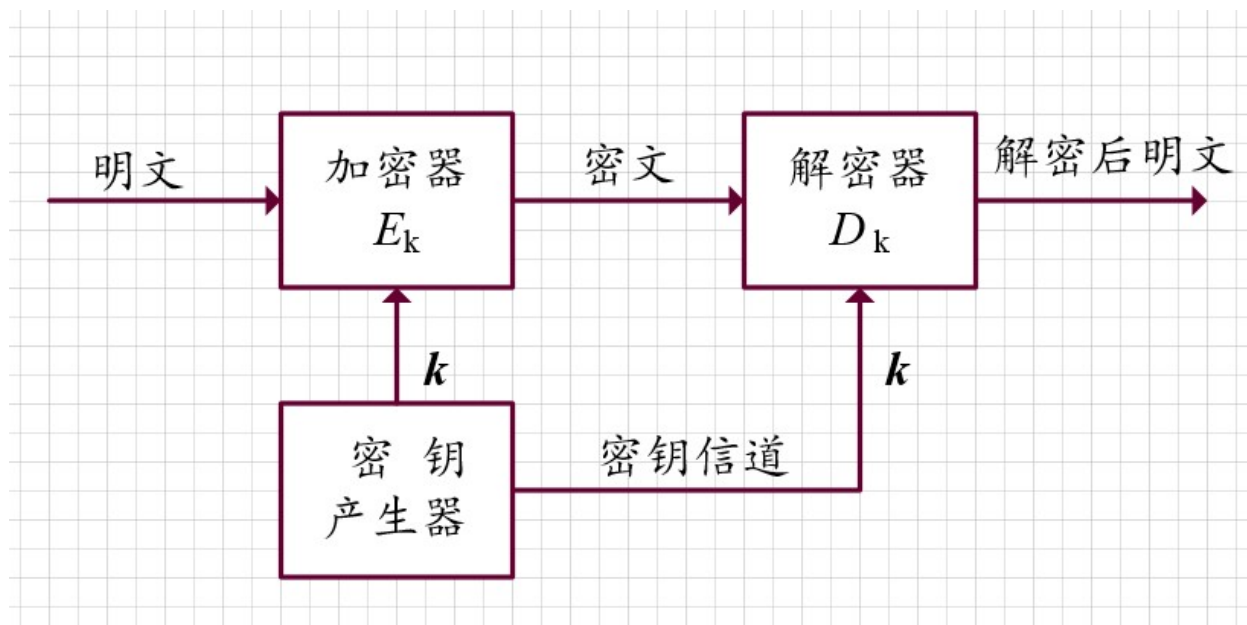


图2 单钥保密体制

**分类：** **流密码** (Stream cipher)、**分组密码** (Block cipher)  
单钥体制不仅可用于数据加密，也可用于消息的认证。



# 双钥体制或公钥体制

Diffie 和 Hellman 1976 年首次提出，每个用户都有一对选定的密钥（公钥 $K_1$ ；私钥 $K_2$ ），公开的密钥 $K_1$ 可以像电话号码一样进行注册公布。

**主要特点：** 加密和解密能力分开

- 多个用户加密的消息只能由一个用户解读，可用于公共网络中实现保密通信。
- 只能由一个用户“加密”消息而使多个用户可以解读，可用于认证系统中对消息进行数字签字。
- 无需事先分配密钥。

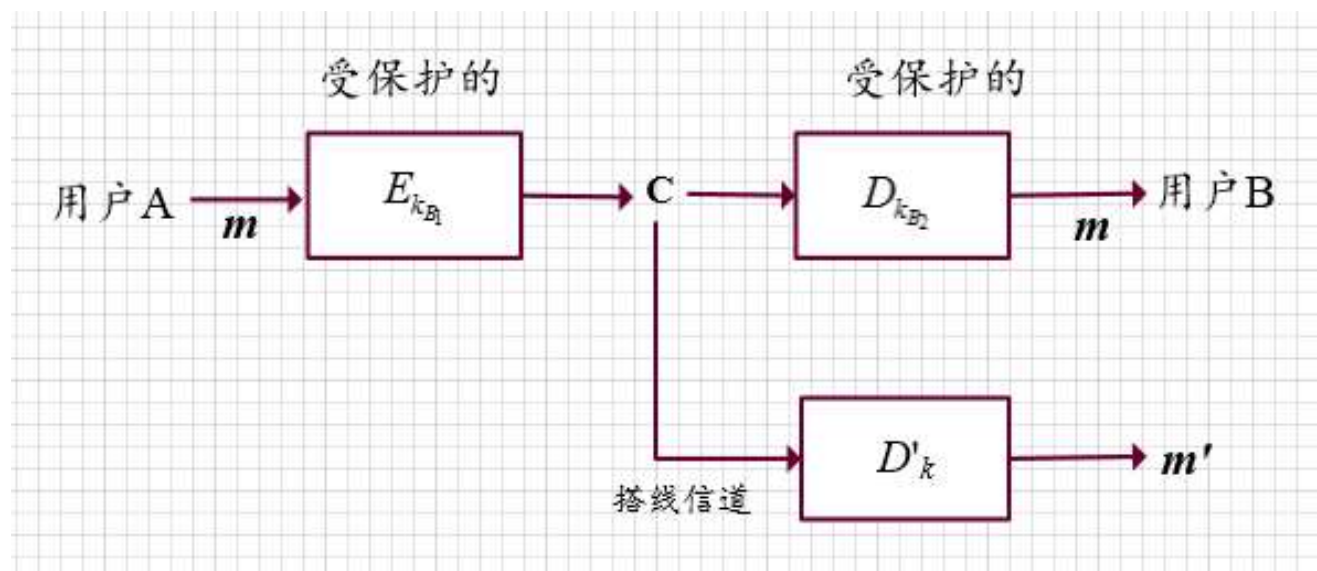


图3 双钥保密体制

- **双钥体制加解密**:  $m = D_{k_{B_2}}(c) = D_{k_{B_2}}(E_{k_{B_1}}(m))$
- **安全保障**: 从公开钥  $k_{B_1}$  和密文  $c$  要推出明文  $m$  或解密密钥  $k_{B_2}$  在计算上是不可行的。
- 由于任一用户都可用用户 B 的公开钥  $k_{B_1}$  向他发送机密消息, 因而密文  $c$  不具有认证性。



# 双钥认证体制

**双钥认证体制：**用户A以自己的秘密钥 $k_{A_2}$ 对消息 $m$ 进行A的专用变换 $D_{k_{A_2}}$ ，A计算密文： $c = D_{k_{A_2}}(m)$ 送给用户B，B验证 $m$ ：

$$m = E_{k_{A_1}}(c) = E_{k_{A_1}}(D_{k_{A_2}}(m)) \quad (5)$$

**安全性：**由于 $k_{A_2}$ 是保密的，其他人都不可能伪造密文 $c$ ，可用A的公开钥解密时得到有意义的消息 $m$ 。因此可以验证消息 $m$ 来自A而不是其他人，而实现了对A所发消息的认证。





# 双钥认证体制

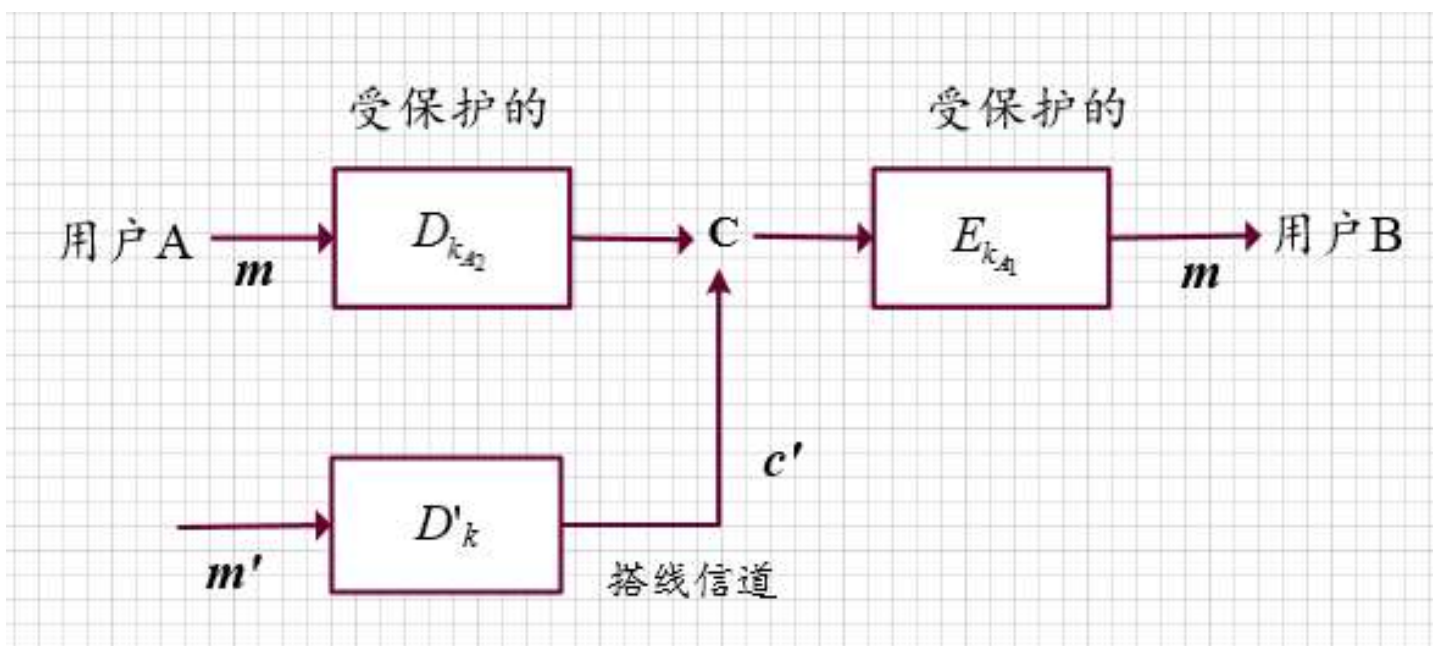


图4 双钥认证体制



# 双钥保密和认证体制

为了要同时实现保密性和确证性，要采用双重加、解密，如图5 所示：

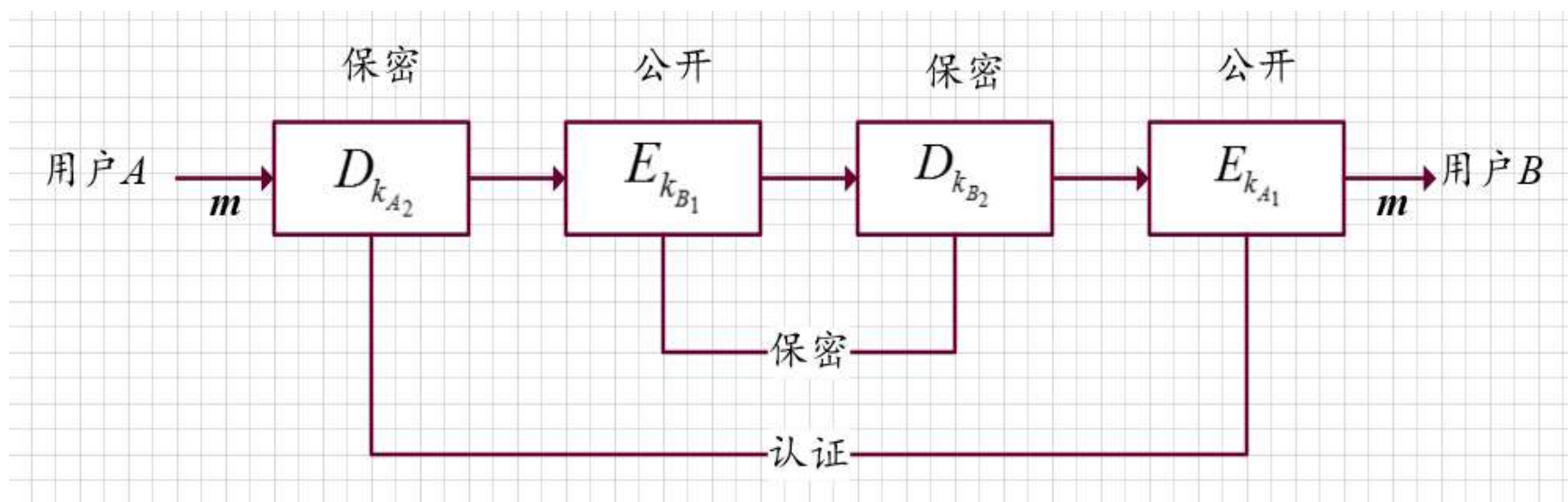


图5 双钥保密和认证体制



# 双钥保密和认证体制

- A 计算的密文:

$$c = E_{k_{B_1}}(D_{k_{A_2}}(m)) \quad (6)$$

- B 验证和解密:

$$\begin{aligned} m &= E_{k_{A_1}}(D_{k_{B_2}}(c)) \\ &= E_{k_{A_1}}(D_{k_{B_2}}(E_{k_{B_1}}(D_{k_{A_2}}(m)))) \\ &= E_{k_{A_1}}(D_{k_{A_2}}(m)) \end{aligned} \quad (7)$$



## 三 古典密码

➤ 古典密码体制都是**对称密码体制**，密钥由安全信道传递

➤ 经典密码体制可分为：

✓ **替换密码/代换密码** (Substitution) —— 用一个符号代替另一个符号

### □ 单表替换密码

- 移位代换密码
- 乘数密码
- 仿射密码

### □ 多表替换密码

✓ **置换密码/换位密码** (Permutation) —— 对符号进行重新排序



# 替换密码

- **替换密码**根据预先建立的**替换表**，将明文依次通过查表，替换为相应字符，生成密文，**替换密码的密钥就是替换表**
- **单表替代密码**：使用一个固定的替换表—明文、密文字符**一一对应**
- **多表替代密码**：使用多个替换表



## ❖ 单表替换密码（单字母代替）：

令明文  $m = m_0 m_1 \dots$ ，则相应密文为  
 $c = E(m) = c_0 c_1 \dots = f(m_0) f(m_1) \dots$

### • 实例1：移位密码（加同余密码； shift substitution cipher）

加密变换：  $E_k(i) = (i + k) \bmod q = j$

其中  $0 \leq i, j < q$ ,  $K = \{k \mid 0 \leq k < q\}$

解密变换：  $D_k(j) = (j - k) \bmod q = i$



- 例：Kaiser（凯撒）密码

若令26个字母分别对应于整数0~25

字母	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
数字	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

kaiser密码的加密：

$$c = (m + k) \bmod 26$$

解密：

$$m = (c - k) \bmod 26$$



- $k = 5$ 时:

明文: a b c d e f g h i j k l m

密文: F G H I J K L M N O P Q R

明文: n o p q r s t u v w x y z

密文: S T U V W X Y Z A B C D E

- 明文:

data security has evolved rapidly

密文:

I F Y F X J H Z W N Y D M F X J A T Q A J I W F U N I Q  
D





- 实例2：乘数密码 (multiplicative cipher)

加密变换:

$$E_k(i) = i * k \bmod q = j \quad 0 \leq j < q$$

- 仅当  $(k, q) = 1$  时，明文/密文字母才一一对应

解密变换:

$$D_k(j) = j * k^{-1} \bmod q = i \quad 0 \leq j < q$$



- 例：英文字母表 $q = 26$ ，取 $k = 9$ ，则有如下明文密文字母对应表。 $(a=0, \dots, z=25)$

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

明文:

$M$  = multiplicative cipher

密文:

$C$  = EYVPUFVUSAPUHK SUFLKX



- **实例3**: 仿射密码(affine cipher; 线性同余密码)
  - 移位密码和乘数密码的组合

- 选取 $k_1$ ,  $k_2$ 两个参数, 其中 $\gcd(k_1, 26) = 1$

加密变换:  $C = k_1 * m + k_2 \bmod 26$

- $K_1 = 1$ 时, 移位密码
- $K_2 = 1$ 时, 乘数密码

解密变换:  $m = (C - k_2) * k_1^{-1} \bmod 26$



例:  $k_1 = 7$ ,  $k_2 = 10$

明文: please send moneys

对应数据为:

16 12 5 1 19 5 19 5 14 4 13 15 14 5 25 19

通过变换  $c = 7m + 10 \bmod 26$  可得:

18 16 19 17 13 19 13 19 4 12 23 11 4 19 3 13

密文:

R P S Q M S M S D L W K D S C M



- 例: **Playfair密码** (一战期间**英国**采用)
  - 思想: 双字母转换
  - 密钥:  $5 \times 5$  矩阵 (由一个**关键词**构造)
    - 比如 FIVESTARS, 将单词中**重复的字母去掉**, 可以得到 FIVESTAR, 将剩下的字母排列成  $5 \times 5$  矩阵的起始部分, 矩阵的剩余部分则用 26 个字母表中未出现的字母顺序填充
  - I 和 J 作为一个字母来对待



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

对每一对明文 $m_1, m_2$ 加密如下:

1. 若 $m_1$ 和 $m_2$ 同行, 则密文 $c_1$ 和 $c_2$ 分别紧靠 $m_1$ ,  $m_2$ 右端的字母, 其中第一列看做最后一列的右方
2. 若 $m_1$ 和 $m_2$ 同列, 则密文 $c_1$ 和 $c_2$ 分别是紧靠 $m_1$ ,  $m_2$ 下方的字母, 其中第一行看做最后一行的下方
3. 若 $m_1$ 和 $m_2$ 不同行也不同列, 则 $c_1$ 和 $c_2$ 是 $m_1$ ,  $m_2$ 确定的矩形的其他两角的字母, 并 $c_1$ 和 $m_1$ ,  $c_2$ 和 $m_2$ 同行



- M=Playfair cipher was actually
  - 先将明文分解成两个字母一对：

P a y f a i r c i p h e r w a s a c t u a || y

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

- 查表得密文： C= QK BW IT VA AS OK VB IG IC TA WT



4. 若出现重复字母，即 $m_1 = m_2$ ，则在其中插入字母Q
5. 如明文字母是单数，将Q放在明文的末端

lly  $\rightarrow$  lq ly

查表：

QZ KZ

**密钥空间：25!**





- 例： **ADFGX密码** （一战期间**德国**采用）
  - **第一步**：将字母表中字母组成 $5 \times 5$ 矩阵，字母/和/被认为是同一个字母，矩阵的行和列用字母*A*, *D*, *F*, *G*, *X*标记，矩阵可能是

	<i>A</i>	<i>D</i>	<i>F</i>	<i>G</i>	<i>X</i>
<i>A</i>	<i>p</i>	<i>g</i>	<i>c</i>	<i>e</i>	<i>n</i>
<i>D</i>	<i>b</i>	<i>q</i>	<i>o</i>	<i>z</i>	<i>r</i>
<i>F</i>	<i>s</i>	<i>l</i>	<i>a</i>	<i>f</i>	<i>t</i>
<i>G</i>	<i>m</i>	<i>d</i>	<i>v</i>	<i>i</i>	<i>w</i>
<i>X</i>	<i>k</i>	<i>u</i>	<i>y</i>	<i>x</i>	<i>h</i>



- 每一明文字母用它所在行和列的标记代替，如 *s* 变成了 *FA*，*z* 变成了 *DG*，假设明文是

*Kaiser Wilhelm*

- 第一步的结果就是

*XA FF GG FA AG DX GX GG FD XX AG FD GA*

	<i>A</i>	<i>D</i>	<i>F</i>	<i>G</i>	<i>X</i>
<i>A</i>	<i>p</i>	<i>g</i>	<i>c</i>	<i>e</i>	<i>n</i>
<i>D</i>	<i>b</i>	<i>q</i>	<i>o</i>	<i>z</i>	<i>r</i>
<i>F</i>	<i>s</i>	<i>l</i>	<i>a</i>	<i>f</i>	<i>t</i>
<i>G</i>	<i>m</i>	<i>d</i>	<i>v</i>	<i>i</i>	<i>w</i>
<i>X</i>	<i>k</i>	<i>u</i>	<i>y</i>	<i>x</i>	<i>h</i>



- **第二步**：选择一个**关键字**，比如 *Rhein*，用关键字字母来标记矩阵的列，将步一的结果组成矩阵：

<i>R</i>	<i>H</i>	<i>E</i>	<i>I</i>	<i>N</i>
<i>X</i>	<i>A</i>	<i>F</i>	<i>F</i>	<i>G</i>
<i>G</i>	<i>F</i>	<i>A</i>	<i>A</i>	<i>G</i>
<i>D</i>	<i>X</i>	<i>G</i>	<i>X</i>	<i>G</i>
<i>G</i>	<i>F</i>	<i>D</i>	<i>X</i>	<i>X</i>
<i>A</i>	<i>G</i>	<i>F</i>	<i>D</i>	<i>G</i>
<i>A</i>				



- 第三步：重新调整列，使列的标记按字母表的顺序排列

<i>E</i>	<i>H</i>	<i>I</i>	<i>N</i>	<i>R</i>
<i>F</i>	<i>A</i>	<i>F</i>	<i>G</i>	<i>X</i>
<i>A</i>	<i>F</i>	<i>A</i>	<i>G</i>	<i>G</i>
<i>G</i>	<i>X</i>	<i>X</i>	<i>G</i>	<i>D</i>
<i>D</i>	<i>F</i>	<i>X</i>	<i>X</i>	<i>G</i>
<i>F</i>	<i>G</i>	<i>D</i>	<i>G</i>	<i>A</i>
				<i>A</i>

按列读字母可得密文：

*FAGDFAFXFGFAXXDGGGXGXGDGAA*

- 已知关键字，解密是容易的：从关键字长度和密文长度可确定列长度，字母被放置到列中，重新排序可以与关键字匹配，然后用初始矩阵恢复明文



## ❖ 多表替换密码：

- 以两个或以上**替换表**依次对明文字母进行替换的加密方法
- 如使用有5个简单替换表的替换密码，明文第1个字母用第1个替换表，第2个字母用第2个表，以此类推，循环使用这五张替换表
- 由莱昂·巴蒂斯塔于1568年发明，著名的有：
  - **维吉尼亚密码**
  - **Hill密码**



- 一次一密 (one-time pad cipher) : 每个明文字母都采用不同的替换表或密钥进行加密
  - 令明文字母表为  $Z_q$ , 令  $\pi = (\pi_1, \pi_2, \dots)$  为替换序列, 明文为  $m = m_1 m_2 \dots$ , 则相应的密文为
$$c = E_k(m) = \pi(m) = \pi_1(m_1) \pi_2(m_2) \dots$$
- 若  $\pi$  为非周期无限序列, 则相应的密码为非周期多表替换密码



- 该体制是Major Joseph Mauborgne和AT&T公司的Gilbert Vernam在1917年为电报通信设计的一种密码，又称Vernam密码
- 该体制理论上不可破译，但实际应用中却受到很大限制，主要原因：
  - 密钥是真正的随机序列；
  - 密钥长度 $\geq$ 明文长度；
  - 每个密钥只用一次（一次一密）



- 为减少密钥量，多采用周期多表替换密码，即替换表个数有限，重复使用
- 代换序列

$$\pi = \pi_1 \pi_2 \dots \pi_d \pi_1 \pi_2 \dots \pi_d$$

明文序列  $m$  的密文为：

$$\begin{aligned} C &= E_k(m) = \pi(m) \\ &= \pi_1(m_1) \pi_2(m_2) \dots \pi_d(m_d) \pi_1(m_{d+1}) \dots \pi_d(m_{2d}) \end{aligned}$$

- $d$  称为周期； $d = 1$  时退化为单表替换





# 维吉尼亚密码

- 1858年, 法国密码学家维吉尼亚提出
- 维吉尼亚方阵是  $26 \times 26$  的方阵
  - 第1行是a到z按正常顺序排列, 第2行由第1行循环左移1位得到, 第3行由第2行循环左移1位得到, 依此类推……

	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



- 加密过程：
  - **密钥字**: encryption
  - **明文**: public key distribution
  - 由于密钥字比明文短，所以要重复书写密钥字以得与明文等长的密钥序列

密钥	e	n	c	r	y	p	t	i	o	n	e	n	c	r	y	p	t	i	o	n	e
明文	p	u	b	l	i	c	k	e	y	d	i	s	t	r	i	b	u	t	i	o	n
密文	T	H	D	C	G	R	D	M	M	Q	M	F	V	R	G	Q	N	B	W	B	R



- 数学解释:

- 密钥:  $k = k_1 k_2 k_3 \cdots k_n$

- 明文:  $m = m_1 m_2 m_3 \cdots m_n$

- 加密:  $E(m) = C = c_1 c_2 c_3 \cdots c_n$

$$c_i = m_i + k_i \bmod 26$$

- 解密:  $m_i = c_i - k_i \bmod 26$



# 希尔 (Hill) 密码

- Lester S. Hill在1929年发明
- 基本思想:将 $n$ 个明文字母通过线性变换, 将它们转换为 $n$ 个密文字母。解密只需做一次逆变换即可
- 算法的密钥  $K = \{ Z_{26} \text{ 上的 } n \times n \text{ 可逆矩阵} \}$ , 明文  $M$  与密文  $C$  均为 $n$ 维向量, 加密和解密变换分别为:

$$C = K \cdot M \bmod 26 \quad M = K^{-1} \cdot C \bmod 26$$

$K^{-1}$ 为 $K$ 在模26上的逆矩阵, 满足

$$KK^{-1} = K^{-1}K = I \pmod{26}$$

$I$  为单位矩阵



• 例:  $n = 4$ ,

$$K = \begin{pmatrix} 8 & 6 & 9 & 10 \\ 6 & 9 & 5 & 10 \\ 5 & 8 & 4 & 9 \\ 10 & 6 & 11 & 4 \end{pmatrix}$$

$$K^{-1} = \begin{pmatrix} 23 & 20 & 5 & 1 \\ 2 & 11 & 18 & 1 \\ 2 & 20 & 6 & 25 \\ 25 & 2 & 22 & 25 \end{pmatrix}$$

不难验证:

$$\begin{pmatrix} 8 & 6 & 9 & 10 \\ 6 & 9 & 5 & 10 \\ 5 & 8 & 4 & 9 \\ 10 & 6 & 11 & 4 \end{pmatrix} \begin{pmatrix} 23 & 20 & 5 & 1 \\ 2 & 11 & 18 & 1 \\ 2 & 20 & 6 & 25 \\ 25 & 2 & 22 & 25 \end{pmatrix} = \begin{pmatrix} 339 & 416 & 312 & 364 \\ 416 & 339 & 442 & 390 \\ 364 & 286 & 339 & 338 \\ 364 & 494 & 332 & 131 \end{pmatrix}$$
$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \text{mod } 26$$



• 例：设明文消息为**good**，试用 $n = 2$ ，密钥  $K = \begin{pmatrix} 11 & 8 \\ 3 & 7 \end{pmatrix}$  的 Hill 密码对其加密，然后再解密

• 解：将明文划分为两组：**(g, o)** 和 **(o, d)**，即 **(6, 14)** 和 **(14, 3)**

加密过程如下：

$$\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = K \begin{pmatrix} m_1 \\ m_2 \end{pmatrix} = \begin{pmatrix} 11 & 8 \\ 3 & 7 \end{pmatrix} \begin{pmatrix} 6 \\ 14 \end{pmatrix} = \begin{pmatrix} 178 \\ 116 \end{pmatrix} \equiv \begin{pmatrix} 22 \\ 12 \end{pmatrix} \pmod{26} \Rightarrow \begin{pmatrix} w \\ m \end{pmatrix}$$

$$\begin{pmatrix} c_3 \\ c_4 \end{pmatrix} = K \begin{pmatrix} m_3 \\ m_4 \end{pmatrix} = \begin{pmatrix} 11 & 8 \\ 3 & 7 \end{pmatrix} \begin{pmatrix} 14 \\ 3 \end{pmatrix} = \begin{pmatrix} 178 \\ 63 \end{pmatrix} \equiv \begin{pmatrix} 22 \\ 11 \end{pmatrix} \pmod{26} \Rightarrow \begin{pmatrix} w \\ l \end{pmatrix}$$

• **good** 的加密结果是 **wmwl**。显然，明文不同位置的字母“o”加密成的密文字母不同



## • 解密过程:

$$\begin{pmatrix} m_1 \\ m_2 \end{pmatrix} = K^{-1} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 7 & 18 \\ 23 & 11 \end{pmatrix} \begin{pmatrix} 22 \\ 12 \end{pmatrix} = \begin{pmatrix} 370 \\ 638 \end{pmatrix} \equiv \begin{pmatrix} 6 \\ 14 \end{pmatrix} \pmod{26} \Rightarrow \begin{pmatrix} g \\ o \end{pmatrix}$$
$$\begin{pmatrix} m_3 \\ m_4 \end{pmatrix} = K^{-1} \begin{pmatrix} c_3 \\ c_4 \end{pmatrix} = \begin{pmatrix} 7 & 18 \\ 23 & 11 \end{pmatrix} \begin{pmatrix} 22 \\ 11 \end{pmatrix} = \begin{pmatrix} 352 \\ 627 \end{pmatrix} \equiv \begin{pmatrix} 14 \\ 3 \end{pmatrix} \pmod{26} \Rightarrow \begin{pmatrix} o \\ d \end{pmatrix}$$

## • Hill密码特点:

- ◆ 可以较好地抑制自然语言的统计特性，不再有单字母替换的一一对应关系，对抗“惟密文攻击”有较高安全强度
- ◆ 密钥空间较大，在忽略密钥矩阵 $K$ 可逆限制条件下， $|K|=26^{n \times n}$
- ◆ 易受已知明文攻击及选择明文攻击





# 替换密码小结

- 单表替换密码
  - 移位密码（凯撒密码）
  - 乘数密码
  - 仿射密码
  - 其他：playfair, ADFGX
- 多表替换密码
  - 一次一密（Vernam）
  - 维吉尼亚密码
  - 希尔（Hill）密码





# 置换密码体制

- 置换加密（换位密码）：明文字符集保持不变，但顺序被打乱

- 例：明文

COMPUTERGRAPHICSMAYBESLOWBUTATLEASTITSEXPNENSIVE

COMPUTERGR  
AP HI CSMAYB  
ES LOWBUTAT  
LE AS T ITSEX  
P E N S I V E

密文:

CAELPOPSEEMHLANPIOSSUCWTITSBIVEMUTERATSGYAERBTX



- 例（深度为2的栅栏技术）：

- 消息“meet me after the toga party”
- 写出如下形式：

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

- 被加密的消息是：

MEMATRHTGPRYETEFETEOAAT



# 古典密码体制的分析

- **穷举分析：**

- **加同余密码：**明文为英文字母表时， $k$  只有25种可能值。明文为8位扩展ASCII码时， $k$  只有255种值。穷举即可
- **乘数密码**更易破译。密钥 $k$ 要满足条件 $(n, k) = 1$ ，即 $k$ 只有 $\leq \phi(n)$ 个不同取值。去掉 $k = 1$ 恒等情况， $k$ 只有 $\phi(n) - 1$ 种
  - 明文为英文字母表时， $k$  只能取3, 5, 7, 9, 11, 15, 17, 19, 21, 23, 25共11种不同值
- **仿射密码**保密性好一些。但密钥也只有 $n\phi(n) - 1$ 种。  
( $m_i * k_1 + k_2$  , 去掉  $m_i * 1 + 0$ )
  - 对英文字母表的情况，密钥只有 $26 \times 12 - 1 = 311$ 种
  - 古代密码分析者用穷举法破译可能会有一定困难，但对计算机来说微不足道



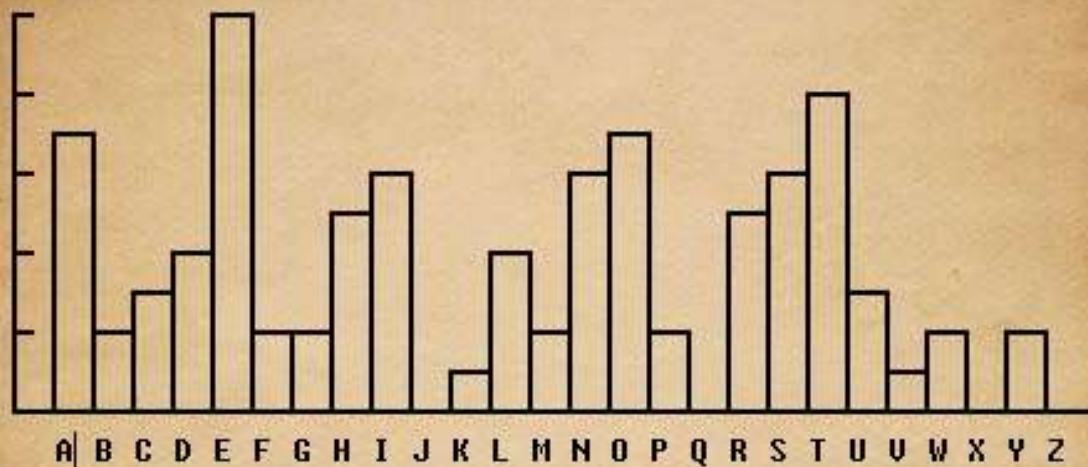
# 古典密码体制的分析

- 统计特性：

- 任何自然语言都有固有的统计特性。如果这种统计特性在密文中有所反映，便可以通过分析明文和密文的统计规律来破译密码。许多古典密码都可用统计分析法破译
- 根据各字母频率可将英文字母分为几组。不仅单字母以相当稳定的频率出现，双字母组和三字母组同样如此



英文字母频率分布表



键盘上的分布





极高频率字母组	E
次高频率字母组	TAOINSHR
中等频率字母组	DL
低频率字母组	CUMWFGYPB
甚低高频率字母组	VKJXQZ

- 出现频率最高的30个双字母组合依次是:

th he in er an re ed on  
es st en at to nt ha nd  
ou ea ng as or ti is et  
lt ar te se hi of

- 出现频率最高的20个三字母组合依次是:

the (比ing高3倍) ing and her ere  
ent tha nth was eth for  
dth hat she ion int his  
sth ers ver



- **统计数据**是通过非专业性文献中的字母进行统计得到的
- 对明文相关知识的掌握对破译密码也十分重要
  - 计算机程序文件的字符频率分布与报纸的字符频率分布有显著不同





## 分析实例—单表密码体制的统计分析

例：设某一段明文经单表替代密码加密后的密文如下：

YIFQ FMZR WQFY VECF MDZP CVMR ZWNM DZVE JBTX CDDUMJ  
NDIF EFMD ZCDM QZKC EYFC JMYR NCWJ CSZR EXCH ZUNMXZ  
NZUC DRJX YYSM RTME YIFZ WDYV ZVYF ZUMR ZCRW NZDZJJ  
XZWG CHSM RNMD HNCM FQCH ZJMX JZWI EJYU CFWD JNZDIR

试分析出对应明文。





**解：**将加密变换记为 $E_k$ ，解密变换记为 $D_k$ ，密文中共有168个字母

**第1步：**统计密文中字母的出现次数和频率

字母	出现次数	出现频率	字母	出现次数	出现频率
A	0	0.000	N	9	0.054
B	1	0.006	O	0	0.000
C	15	0.089	P	1	0.006
D	13	0.077	Q	4	0.024
E	7	0.042	R	10	0.060
F	11	0.065	S	3	0.018
G	1	0.006	T	2	0.012
H	4	0.024	U	5	0.030
I	5	0.030	V	5	0.030
J	11	0.065	W	8	0.048
K	1	0.006	X	6	0.036
L	0	0.000	Y	10	0.060
M	16	0.095	Z	20	0.119



第2步：从出现频率最高的几个字母及双字母组合、三字母组合开始，并假定它们是英语中出现频率较高的字母及字母组合对应的密文，逐步推测各密文字母对应的明文字母

- 密文字母Z出现次数最高（出现频率约为0.12）。可以猜测： $D_k(Z) = e$ 。除Z外，出现至少10次的密文字母为C, D, F, J, M, R, Y，出现频率在0.06~0.095之间，可以猜测密文字母{C, D, F, J, M, R, Y}可能对应于明文字母集合{t, a, o, i, n, s, h, r}中的字母，但不能肯定是哪一个
- 因为已经假设密文Z解密成e，现在考虑密文中形如—Z和Z—的双字母组合：



$-Z$	出现次数	$Z-$	出现次数
DZ	4	ZW	4
NZ	3	ZU	3
RZ	2	ZR	2
HZ	2	ZV	2
XZ	2	ZC	2
FZ	2	ZD	2
		ZJ	2

- DZ和ZW出现4次；NZ和ZU出现3次；RZ，HZ，XZ，FZ，ZR，ZV，ZC，ZD和ZJ出现2次
- 由于ZW出现4次而WZ一次也未出现，同时W出现的频率为0.048，故可猜测

$$D_k(W) = d$$



**第3步：**同理，可很容易确定其余密文/明文字母的对应关系，最后加上标点，得到完整的明文：

Our friend from Paris examined his empty glass with surprise, as if evaporation had taken place while he wasn't looking. I poured some more wine and he settled back in his chair, face tilted up towards the sun.

**注记：**以上是破解一般单表替代密码的统计分析方法，如果已知所用的密码体制（例如加法密码、乘法密码和仿射密码等），则相应的分析工作会更简单



## 结论—破译单表替代密码的大致过程：

1. 统计密文的各种统计特征，如果密文量比较多，则完成这步后便可确定出大部分密文字母
2. 分析双字母、三字母密文组，以区分元音和辅音字母
3. 分析字母较多的密文，在这一过程中大胆使用猜测的方法，如果猜对一个或几个词，就会大大加快破译过程



## 分析实例—多表密码体制的统计分析

- 多表替代密码一定程度上隐藏了明文消息的一些统计特征，破译相对较难
  - 由于每一个明文字母都有多个不同的密文字母来代替，因此密文字母频率分布是比较平坦的，所以它们的保密性比单表替换密码高
- 例： 维吉尼亚（Vigenere）密码的统计分析



- 多表替代密码的分析主要包括两个步骤：
  1. 确定密钥的长度 $d$  — 即确定使用的加密表个数
  2. 确定具体的密钥字
- 确定密钥长度的两种常用方法：
  - Kasiski测试法 (Kasiski test) — 由普鲁士军官Kasiski在1863年提出
  - 重合指数法 (index of coincidence)



## • Kasiski测试法的基本原理:

- 若用给定的  $d$  个密钥表周期地对明文字母加密, 则当明文中两个相同字母组的间隔为  $d$  的倍数时, 这两个明文字母组对应的密文字母组必相同
- 两个相同的密文段, 对应的明文段不一定相同, 但相同的可能性大。将密文中相同的字母组找出来, 并找出相同字母组距离的最大公因子, 就有可能提取出密钥的长度  $d$





考虑下面一个维吉尼亚密码的简单例子

明文: requests additional test

密钥: TELEXTEL EXTELEXTEL EXTE

密文: CAVKTBLT EUQWSWJGEA LTBL

- 明文包含字母序列est两次，而又碰巧被同样的密钥段加密，因而对应的密文都是TBL
- 反映了如下事实：序列est位于密钥长度（或周期）的整数倍处。相同字母组的距离反映了密钥长度n的相关信息



- Kasiski测试过程如下：

- 搜索长度至少为2的相邻的一对对相同的密文段，记下它们之间的距离。而密钥长度 $d$ 可能就是这些距离的最大公因子



ISWZPNQCKMYYYYJKAYYEZFFSWEESSPGZXQAHF  
ISWZPNQCKMTVYJOACVEHAESA ZRLTPQIZMXOT  
QSWMCVUDSIJGGDEUWAZRSFXWILKUEJQLDACB  
GDL YJXMYLMDQKZMPLDILQEMWF SWDP AZEZQNW  
DYWDZXFSAEAAZJDUELVPMTCEKWSEEFURZFSW  
DPXACQAFKMXWAWVEZFSD BGD LAYUQXGDPEKWS  
EEFURZF SWDP OUEZKZMYLQNPQQDEMJTQYGUVA  
ZOGRWAWPVUEQAFJQ JGG COMJZAHQAFKTJDKAD  
MNWP JGG CWKPKAYEQZZPTVKZMQGWDVFAHLTLL  
USSP XAZPGZJGGOSDWAZRKA EZQCWKZMMCWITL  
TEZMEDAZCAYQAFJRLUQLKUQQAFJQYWH PJT FJ  
FLKUQQAFJQYWH PJPZOZDZMWDUMWFSWAYWRZJ  
KZMISGBTF OSEEJGGD GREDKMMFDMDPARQJAHF  
UDKTZOZEZQYAITDXVFAHLTLLKZMMCWZZVDPS  
YPJ

在里面重复序列  
有ISWZPN  
QCKM, BG  
DL, SWDP,  
JGGC等;

如果每个重复间  
隔都能被3整除,  
关键词应该有三  
个字母



## 重合指数法 (index of coincidence)

**定义 (重合指数)** : 设  $x = x_1 x_2 \dots x_n$  是  $n$  个字母的串,  $x$  的重合指数是指  $x$  中两个随机元素相同的概率, 记为  $I_c(x)$ 。

- 假定  $f_0, f_1, \dots, f_{25}$  分别表示  $x$  中字母 A, B, ..., Z 出现的频率。我们能以  $\binom{n}{2}$  种方法选择  $x$  中的两个元素, 对每一个  $i$ ,  $0 \leq i \leq 25$ ,  $x$  中的两个元素都被选择为  $i$  的方法有  $\binom{f_i}{2}$  种



因此 $x$ 的重合指数:

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



- 记字母A, B, …, Z出现的期望概率分别为 $p_0, p_1, \dots, p_{25}$ 。通过对大量的小说、杂志、报纸等的统计, 人们已经获得了英文的26个字母的概率分布的一个估计。期望值为

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

- 因为两个随机元素都是A的概率为 $p_0^2$ , 都是B的概率为 $p_1^2$ , 等等



- 如果 $x$ 是利用任何单表替换密码获得的一个密文，那么在这种情况下，各个概率将只是被作了一个置换，但量 $\sum_{i=0}^{25} p_i^2$ 不变。因此，对用单表替换密码获得的一个密文 $x$ ，也有

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



- $$\begin{array}{l} Y_1 = y_1 y_{d+1} y_{2d+1} y_{3d+1} \cdots \\ Y_2 = y_2 y_{d+2} y_{2d+2} y_{3d+2} \cdots \\ \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \\ Y_d = y_d y_{2d} y_{3d} y_{4d} \cdots \end{array}$$

- 如果  $d$  是密钥字长度, 那么每个  $I_c(Y_i)$  ( $1 \leq i \leq d$ ) 都将大概等于 0.065





- 如果 $d$ 不是密钥字的长度，那么串 $Y_i$ 将看起来更随机些，因为它们是采用不同的密钥移位加密获得的

- 而一个完全随机的串 $x$ ,

$$I_c(x) \approx 26 \cdot (1/26)^2 = 1/26 = 0.038.$$

值0.065和0.038间隔充分远，所以能确定正确的密钥字长度



- 例：Vigenere密码加密的密文：

CHREEVOAHMAERATBIAXXWTNXBEEOPHBSBQMQEQRBWVRVXUOAKXAOSXXWEAHBWGJMM  
QMNKGRFVGXWTRZXWIAKLXFPSKAUTEMNDCMGTSMXBTUIADNGMGPSRELXNJELXVRVPRTU  
LHDNQWTWDTYGBPHXTFAUHASVBFXNGLLCHRZBWELEKMSJIKNBHWRJGNMGJSGLXFEYPHAG  
NRBIEQJTAMRVLCRREMNDGLXRRIMGNSNRWCHRQHAHEYVTAQEBBIPEEWEVKAKOEWADREMX  
MTBHHCHRTKDNVRZCHRCLQOHPWQAIWXXNRMGWOIIFKEE

- Kasiski测试法：密文串CHR出现四次，起始位置在1，166，236和286。从第1个到其它3个的距离分别为165，235和285， $\gcd(165, 235, 285) = 5$ ，所以5可能是密钥字长度



## • 重合指数法:

1. 当 $d = 1$ 时, 重合指数为0.0455
2. 当 $d = 2$ 时, 两个重合指数分别为0.046和0.041;
3. 当 $d = 3$ 时, 三个重合指数分别为0.043, 0.050和0.047;
4. 当 $d = 4$ 时, 四个重合指数分别为0.042, 0.039, 0.046和0.040,
5. 当 $d = 5$ 时, 五个重合指数分别为0.063, 0.068, 0.069, 0.061和0.072, 这为密钥字长度是5提供了有力证据



- **第二步—确定密钥字：重合互指数法** (mutual index of coincidence)
- **定义（重合互指数）**：假定  $x = x_1 x_2 \dots x_n$  和  $y = y_1 y_2 \dots y_{n'}$ , 分别是长为  $n$  和  $n'$  的字母串。 $x$  和  $y$  的重合互指数是指  $x$  的一个随机元素等于  $y$  的一个随机元素的概率，记为  $MI_c(x, y)$



- 将 $x$ 和 $y$ 中的字母A,B,C,……,Z出现的次数分别表示为 $f_0, f_1, \dots, f_{25}$ 和 $f_0', f_1', \dots, f_{25}'$ , 那么

$$MI_c(x, y) = \sum_{i=0}^{25} \frac{f_i}{n} \frac{f_i'}{n'} = \frac{\sum_{i=0}^{25} f_i f_i'}{nn'}$$



- **$d$ 的值已确定**，假定  $K = (k_1, k_2, \dots, k_d)$  是密钥字。我们来估计  $MI_c(Y_i, Y_j)$
- $Y_i$  中的一个随机字母和  $Y_j$  中的一个随机字母都是 A 的概率是  $p_{-k_i} p_{-k_j} (= p_{26-k_i} p_{26-k_j})$ 。因此，我们可估计

$$MI_c(Y_i, Y_j) \approx \sum_{h=0}^{25} p_{h-k_i} p_{h-k_j} = \sum_{h=0}^{25} p_h p_{h+k_i-k_j}$$

- 易知， $MI_c(Y_i, Y_j)$  的估计值只依赖于差  $(k_i - k_j) \bmod 26$ ，我们称该差为  **$Y_i$  和  $Y_j$  的相对移位**



# 期望的重合指数

相对位移	$MI_c$ 的期望值
0	0.065
1(25)	0.039
2(24)	0.032
3(23)	0.034
4(22)	0.044
5(21)	0.033
6(20)	0.036
7(19)	0.039
8(18)	0.034
9(17)	0.034
10(16)	0.038
11(15)	0.045
12(14)	0.039
13	0.043

- 当相对移位不是0时，估计值均在0.031~0.045之间，而是0时，估计值是0.065。我们能使用这个表来推测 $Y_i$ 和 $Y_j$ 的相对移位  
$$l = (k_i - k_j) \bmod 26$$



- 固定 $Y_i$ , 考虑由长为 $d$ 的密钥字 $e_0 = (0, 0, \dots, 0)$ ,  $e_1 = (1, 1, \dots, 1)$ ,  $e_2 = (2, 2, \dots, 2)$ ,  
.....  $e_{25} = (25, 25, \dots, 25)$ 分别加密 $Y_j$ , 将加密结果 分别记为 $Y^0_j$ ,  $Y^1_j$ ,  $Y^2_j$ , .....  $Y^{25}_j$ , 则容易计算

$$MI_c(Y_i, Y_j^d) = \frac{\sum_{i=0}^{25} f_i f_{i-d}}{nn'} \quad (0 \leq d \leq 25)$$





- 当  $d = l$  时,  $MI_c$  将接近 0.065, 因为  $Y_i$  和  $Y_j - l$  的相对移位是 0
- 对  $d \neq l$  的值,  $MI_c$  的值将在 0.031 和 0.045 之间
- 使用重合互指数, 我们能获得子串  $Y_1, Y_2, \dots$  中的任何两个的相对移位



- 利用计算机不难算出260个值 $MI_c(Y_i, Y^d_j)$ ,  $0 \leq i < j \leq 5, 0 \leq d \leq 25$ 。

表 3.8 估计的重合指数

i	j	MI(Y <sub>i</sub> , Y <sup>d</sup> <sub>j</sub> ) 的值									
		0	1	2	3	4	5	6	7	8	9
1	2	0.018	0.027	0.028	0.036	0.039	0.047	0.028	0.025	0.022	0.028
		0.038	0.044	0.036	0.037	0.043	0.037	0.043	0.037	0.039	0.039
		0.042	0.041	0.034	0.037	0.035	0.045	0.042	0.046	0.046	0.046
1	3	0.029	0.032	0.040	0.034	0.029	0.032	0.038	0.033	0.033	0.039
		0.036	0.030	0.045	0.036	0.040	0.034	0.037	0.037	0.037	0.027
		0.039	0.030	0.031	0.037	0.033	0.039	0.034	0.037	0.037	0.037
1	4	0.034	0.043	0.025	0.027	0.038	0.043	0.040	0.032	0.029	0.029
		0.034	0.028	0.044	0.044	0.034	0.039	0.043	0.043	0.037	0.037
		0.035	0.047	0.022	0.027	0.039	0.037	0.039	0.033	0.033	0.033
1	5	0.042	0.033	0.028	0.044	0.043	0.044	0.039	0.031	0.029	0.029
		0.030	0.036	0.040	0.041	0.029	0.039	0.048	0.039	0.044	0.044
		0.028	0.036	0.044	0.042	0.047	0.033	0.028	0.040	0.040	0.040
2	3	0.044	0.048	0.041	0.032	0.039	0.033	0.036	0.030	0.029	0.029
		0.039	0.034	0.029	0.040	0.037	0.041	0.032	0.037	0.043	0.043
		0.035	0.033	0.027	0.032	0.035	0.032	0.042	0.030	0.030	0.030
2	4	0.048	0.034	0.043	0.044	0.034	0.031	0.040	0.043	0.040	0.040
		0.048	0.044	0.023	0.034	0.028	0.042	0.038	0.028	0.039	0.039
		0.039	0.033	0.032	0.040	0.034	0.043	0.028	0.028	0.028	0.028
2	5	0.033	0.033	0.034	0.048	0.039	0.038	0.043	0.036	0.036	0.036
		0.029	0.033	0.041	0.039	0.037	0.032	0.026	0.031	0.038	0.038
		0.042	0.037	0.041	0.046	0.045	0.041	0.035	0.035	0.035	0.035
3	4	0.038	0.038	0.040	0.033	0.030	0.040	0.035	0.041	0.039	0.039
		0.038	0.037	0.033	0.034	0.033	0.039	0.032	0.039	0.038	0.038
		0.038	0.029	0.044	0.032	0.031	0.032	0.034	0.030	0.030	0.030
3	5	0.031	0.034	0.034	0.034	0.030	0.043	0.043	0.030	0.025	0.025
		0.041	0.033	0.036	0.033	0.032	0.039	0.032	0.032	0.032	0.031
		0.027	0.030	0.032	0.035	0.034	0.032	0.032	0.032	0.032	0.032
4	5	0.022	0.034	0.033	0.038	0.041	0.033	0.037	0.038	0.038	0.038
		0.028	0.034	0.031	0.032	0.033	0.032	0.032	0.034	0.034	0.032
		0.029	0.043	0.033	0.027	0.026	0.039	0.048	0.035	0.035	0.035



- 对每对 $(i, j)$ , 找出接近于0.065的 $MI_c(Y_i, Y_j^d)$ 的值:
  - $Y_1$ 和 $Y_2$ 的相对移位可能是9;
  - $Y_1$ 和 $Y_5$ 的相对移位可能是16;
  - $Y_2$ 和 $Y_4$ 的相对移位可能是13;
  - $Y_1$ 和 $Y_5$ 的相对移位可能是7;
  - $Y_3$ 和 $Y_5$ 的相对移位可能是20;
  - $Y_4$ 和 $Y_5$ 的相对移位可能是11



- 这给出了关于未知量 $k_1, k_2, k_3, k_4, k_5$ 的如下关系式:  
$$k_1 - k_2 = 9; \quad k_1 - k_5 = 16; \quad k_2 - k_3 = 13;$$
$$k_2 - k_5 = 7; \quad k_3 - k_5 = 20; \quad k_4 - k_5 = 11$$
- **即** $k_2 = k_1 + 17, \quad k_3 = k_1 + 4, \quad k_4 = k_1 + 21,$   
 $k_5 = k_1 + 10$ 。密钥字的可能形式为  
 $(k_1, k_1 + 17, k_1 + 4, k_1 + 21, k_1 + 10), \quad k_1 \in Z_{26}$
- 不难确定密钥字是**JANET** (最多穷举26种可能)



- 将密文可解密为：

The almond tree was intentative blossom. The days were longer, often ending with magnificent evenings of corrugated pink skies. The hunting season was over, with hounds and guns put away for six months . The vineyards were busy again as the well-organized farmers treated their vines and the more lackadaisical neighbors hurried to do the pruning they should have done in November



# 本章小结

- 经典密码体制定义和分类
- 经典密码学实例
  - 替换密码体制
  - 置换密码体制
- 古典密码的统计分析
  - 单表替代密码分析
  - 多表替代密码分析



## 四、初等密码分析

**密码分析学**：研究分析解密规律的科学。即截收者在不知道解密密钥及通信者所采用的加密体制的细节条件下，对密文进行分析，试图获取机密信息。

- 密码分析在外交、军事、公安、商业等方面都具有重要作用，也是研究历史、考古、古语言学和古乐理论的重要手段之一。
- 密码设计和密码分析是共生的、又是互逆的，两者密切有关但追求的目标相反。两者解决问题的途径有很大差别。密码设计是利用数学来构造密码，而密码分析除了依靠数学、工程背景、语言学等知识外，还要靠经验、统计、测试、眼力、直觉判断能力……，有时还靠点运气。密码分析过程通常经验、统计、假设、推断和证实等步骤。
- Shannon 在1949年开创的信息论理论第一次透彻地阐明的密码分析的基本问题。密码分析之所以能够破译密码，最根本的是依赖于明文中的多余度。



# 初等密码分析

**破译 (Break ) 或攻击 (Attack) 密码方法:**

- 穷举破译法 (Exhaustive Attack Method), 又称作强力法 (Brute-force Method)。
- 分析法, 有确定性和统计性两类。





# 初等密码分析

## 穷举破译法

对截收的密报依次用各种可解的密钥试译，直到得到有意义的明文；或在不变密钥下，对所有可能的明文加密直到得到与截获密报一致为止，此法又称为完全试凑法 (Complete trial-and-error Method)。只要有足够多的计算时间和存储容量，原则上穷举法总是可以成功的。但实际中，任何一种能保障安全要求的实用密码都会设计得使这一方法在实际上是不可行的。

为了减少搜索计算量，可以采用较有效的改进试凑法。它将密钥空间划分成几个 (例如 $q$  个) 等可能的子集，对密钥可能落入哪个子集进行判断，至多需进行 $q$  次试验。关键在于如何实现密钥空间的等概子集的划分。



# 初等密码分析

## 分析破译法

- **确定性分析法**：利用一个或几个已知量（如已知密文或明文-密文对）用数学关系式表示出所求未知量（如密钥等）。已知量和未知量的关系视加密和解密算法而定，寻求这种关系是确定性分析法的关键步骤。例如，以 $n$ 级线性移存器序列作为密钥流的流密码，就可在已知 $2n$  bit密文下，通过求解线性方程组破译。
- **统计分析法**：利用明文的已知统计规律进行破译的方法。密码破译者对截收的密文进行统计分析，总结出其间的统计规律，并与明文的统计规律进行对照比较，从中提取出明文和密文之间的对应或变换信息。



# 初等密码分析

## 密码可能经受的不同水平的攻击

- **惟密文破译** (Ciphertext Only Attacks)。分析者从仅知道的截获密文进行分析，试图得出明文或密钥。
- **已知明文破译** (Know Plaintext Attacks)。分析者除了有截获的密文外，还有一些已知的明文-密文对(通过各种手段得到的)，试图从中得出明文或密钥。
- **选择明文破译** (Chosen Plaintext Attacks)。分析者可以选定任何明文-密文对来进行攻击，以确定未知的密钥。
- **选择密文攻击** (Chosen Ciphertext Attack)。分析者可以利用解密机，按他所选的密文解密出相应的明文。双钥体制下，类似于选择明文攻击，他可以得到任意多的密文对密码进行分析。

**这几类攻击的强度依次增大，唯密文攻击最弱。**



# 初等密码分析

攻击类型	攻击者拥有的资源
惟密文攻击	<ul style="list-style-type: none"><li>■加密算法</li><li>■截获的部分密文</li></ul>
已知明文攻击	<ul style="list-style-type: none"><li>■加密算法</li><li>■截获的部分密文和对应的明文</li></ul>
选择明文攻击	<ul style="list-style-type: none"><li>■加密算法</li><li>■加密黑盒子，可加密任意明文得到相应的密文</li></ul>
选择密文攻击	<ul style="list-style-type: none"><li>■加密算法</li><li>■解密黑盒子，可解密任意密文得到相应的明文</li></ul>



# 初等密码分析

## 无条件安全和计算安全：无条件安全（完善保密）

- 对密码体制的任何攻击，都不优于（对明文）完全盲目的猜测，这样的密码体制就称为无条件安全的（或完善保密的）。
- 一次一密的加密方式容易实现无条件安全性。因为密钥时时更新，所以以往得到的任何明文/密文对，对于破译新的密文没有任何帮助，只能做完全盲目的猜测。



# 初等密码分析

## 无条件安全和计算安全：计算安全

计算安全是一个模糊的概念。我们可以给出以下三个级别的定义

- 对密码体制的任何攻击，虽然可能优于完全盲目的猜测，但超出了攻击者的计算能力。这是最高级别的计算安全。
- 对密码体制的任何攻击，虽然可能没有超出攻击者的计算能力，但所付出的代价远远大于破译成功所得到的利益。这是第二级别的计算安全。
- 对密码体制的任何攻击，虽然可能没有超出攻击者的计算能力，但破译成功所需要的时间远远大于明文本身的有效期限。这也是第二级别的计算安全。



# 初等密码分析

- 密码分析的成功除了靠上述的数学演绎和归纳法外，还要利用大胆的猜测和对一些特殊或异常情况的敏感性。
- 一个保密系统是否被“攻破”，并无严格的标准。
- 密码史表明：密码分析者的成就似乎远比密码设计者的成就更令人赞叹！许多开始时被设计者吹为“百年或千年难破”的密码，没过多久就被密码分析者巧妙地攻破了。
- 在第二次世界大战中，美军破译了日本的“紫密”，使得日本在中途岛战役中大败。一些专家们估计，同盟军在密码破译上的成功至少使第二次世界大战缩短了8年。





# 初等密码分析-破译实例

## 1. 珍珠港事件 (1941.12.7, 7:30AM)

1941.12.7早上1点28分，西雅图布里奇岛海军情报站截获日本给驻美大使紫密报，9分钟后海军部大楼1649室收到转发报(OP-20-GY)。上午5点已脱密成日文：“请贵大使在当地时间7日下午1时(Hawaii早上7:30)将我国答复交美国政府。9:30分交海军作战部斯塔克上将和诺克斯海军部长”。克雷默将译报带向国务院(日大使馆的译电员在一小时后才将电报脱密)，9:15到达白宫；罗斯福用10分钟看完13部分电报说：这意味着战争。10:30左右，克来默回到办公室看到第14部电报译稿；10:45分将此电报送到白宫。马歇尔仔细看了此报，想向太平洋地区发出警告。经过加密，46分钟后才到达檀香山美国无线电公司(当地上午7:33分)。7:55分日机开始轰炸。日大使下午2:20才走进白宫提交照会。11:45分最后一批日机才离开珍珠港。





# 初等密码分析-破译实例

## 2. 中途岛海战 (代号AF)

Joseph John Rochefort 中校1942年春成功破译日海军JN25b密本，得知日要进攻中途岛。舰队总司令尼米兹上将建议授“卓越服务奖章”。遭海军部反对，直到1976年 Rochefort 逝世后，1985海军部长 John Lehman才发布命令，由总统授奖章给他女儿。

日本由于准备来不及原定4月1日更换密码为JN25c，但一直延到5月1日才启用，而使美军大量破译了秘密情报JN25b，精确推出日将于6月3日开始攻击中途岛。由于情报准确，1942.5.7美机炸沉日航空母舰《祥风》，另一艘的飞行甲板被炸弯，而使多架飞机投入大海中。中途岛作战日本的“赤城”、《加贺》、《苍龙》、《飞龙》号航空母舰被击沉，从而使日本丧失了进行大海这战的能力。尼米兹上将认为，中途岛战本质上是情报的胜利。



# 初等密码分析-破译实例

## 3. 冲绳岛战役

冲绳岛战役中密码分析者破译了72000吨《大和》号超级战列舰决死出击令，及其位置，舰载机群于1945.4.7日12时32分将《大和》号击中，经过二小时反复轰炸沉入海底。船上2767人全部战死。

## 4. 山本五十六之死

山本五十六于1943年4月13日下午5时55分到所罗门岛视察的日程播给第一基地部队等，被美军截获并通过破译JN25密码的专用IBM设备破译出。尼米兹上将经研究决定出击。4月13日7时25分出动18架战斗机将山本座机击落。5月21日日本才广播这一消息。



# 初等密码分析-破译实例

## 5. 日本商船的毁灭

二次大战中，美国密码分析人员破开75种日本海军密码。美国击沉日本商船总吨位的2/3，美潜艇用鱼雷击沉日本油船110艘，日本商船队的毁灭是东条列举日本战败的三个原因之一(其他两个为越岛战略和快速航空母舰作战)。

## 6. 德国潜艇指挥部密码的破译

德国潜艇指挥部德尼茨的B机关泄露了太多的军事情报。1944.6.4德《U-505》潜艇受到美海军22.3特遣大队反潜深炸弹攻击，受伤浮起后美军冲入无线电室，缴获了密码机和大量明、密报，并秘密将《U-505》拖回美国。德军误认为《U-505》沉没海底而未换密码，欧战结束前11个月，几乎每天击沉一艘潜艇，共计击沉300多艘。



# 初等密码分析-破译实例

## 7. 瑞典人破译德国外交和军事密码

1941年春，瑞典人破译德国外交和军事密码，分析出德军将于6月20~25日入侵俄车。潘汉年也从日将内部得到这一情报，并通过延安告知斯大林。

信号截收和密码分析在战争中的作用极大，大大缩短了战争(WW2缩短了8年)。拯救了千千万万人的生命。

## 8. 苏联 8.19 事件

美国国家的安全局在1991年8.19事件破译了政变领导人克格勃主席克留奇科夫与国防部长亚佐夫的保密电话，并将情报告诉叶利钦，使其准确掌握了苏军各级军官支持与反对者名单，美使馆还派出通信保密专家帮助叶利钦的保持他与支持者的通信安全，美驻俄使馆第10层的电子侦察中心至今仍加强对俄军进行全面侦察。



# 初等密码分析-破译实例

## 9. 海湾战争

1990.7 ~ 1991年初的海湾战争充分显示了电子侦察和通信保密的作用。

## 10. RSA-129 破译

Rivest等最初悬赏\$100的RSA-129，已由包括五大洲43个国家600多人参加，用1600台机器同时产生820条指令数据，通过Internet网，耗时8个月，于1994.4.2利用二次筛法分解出为64位和65位的两个因子，原来估计要用4亿亿年。所给密文的译文为“这些魔文是容易受惊的鱼鹰”。这是有史以来最大规模的数学运算。RSA-130于1996.4.10利用数域筛法分解出来，目前正在向更正大的数、特别是512 bits RSA，即RSA-155冲击[Cowie等1996]。



# 初等密码分析-破译实例

## 11. RSA-155 破译:

1999年8月22日, 来自六个不同国家的科学家们在CWI (CWI是在荷兰的一个数学和计算机科学的国家研究学会) 的Herman te Riele的带领下, 在对RSA-155的攻击中, 利用数域筛法 (NFS) 成功的分解出了512-bit RSA模的素因子。要分解RSA-155所需的CPU时间大约为8400MIPS年(MIPS-年指以每秒执行1000000条指令的计算机运行一年), 这大约为分解RSA-140所需时间的4倍。分解RSA-155总共用了7个月的时间。密码分析者估计在3年内分解RSA-155所用到的算法和计算技术至少在科技界将会得到普及, 因此RSA-155将不再安全。并且人们预计在十年内RSA-232也将被攻破。

**Factorization of a 768-bit RSA modulus**, <http://eprint.iacr.org/2010/006>



thank you