- 4.密钥短语密码
- 密码表是从正常顺序字母表按某种规律变换而成
- 优点是便于记忆，缺点是密钥量小，保密强度低。
- 如果代换字母表由26个字母随机抽取排列，则共有26！种不同的排列，即共有26！个不同密钥。
- $26! \approx 4 \times 10^{26}$，是一个很大的量。
- 用穷举法进行密码分析，即便利用现代计算机，也是很困难的

- 为了保留随机代换密码密钥量大的优点，同时又克服密钥不便记忆的缺点，就产生了密钥短语密码。
- 基本思想是任意选择一个英文短语作为密钥，去掉重复字母后，将其依次写在明文字母表的下面，然后将字母表中没有在短语中出现的字母依次写在此短语后面，就可构造出一个字母代换表。

- 例2.3：取密钥短语为key phrase cipher，则其代换表为
- 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
- A'=K,E,Y,P,H,R,A,S,C,I,B,D,F,G,J,L,M,N,O,Q,T,U,V,W,X,Z
- 若明文m=key phrase cipher
- 则密文c=BHX LSNKOH YCLSHN
- 密钥短语是可以任意选择的，故可构成的代换字母表的数量是极大的
- 足以对付密码分析者用穷举法进行的攻击
- 同时密钥短语既可任意选择，又便于记忆

- 2.2.3密码分析的统计方法
- 密钥短语密码与其他单表代换密码一样，每个字母都是用另一个字母固定代换
- 因而明文中所用语言的各种统计特性都会反应到密文中，
- 这样只要所截获的密文具有适当长度，足以反映出所用语言的某些统计特性，
- 利用统计方法就可以很快破译单表代换密码。

表 4.1　字母统计表

字母	a	B	c	d	e	f	g	h	i	j	K	l	m
频率	0.0356	0.0139	0.0279	0.0378	0.1304	0.0289	0.0199	0.0528	0.0627	0.0013	0.042	0.0339	0.0249
字母	n	o	p	q	r	s	t	u	V	w	X	y	z
频率	0.0707	0.0797	0.0199	0.0012	0.0677	0.0607	0.1045	0.0249	0.0092	0.0149	0.0017	0.0199	0.0008

字母e出现的频率最高，z出现的频率最低。依据各字母出现频率大小的不同，可将26个字母划分为五组，如下表所示：

表 4.2　英文字母分类表

Ⅰ类	极高频率字母集	E
Ⅱ类	次高频率字母集	t a o i n s h r
Ⅲ类	中等频率字母集	d l
Ⅳ类	低频率字母集	c u m w f g y p b
Ⅴ类	极低频率字母集	v k j x q z

- 利用统计分析方法，破译单表代换密码的方法是：
- 首先根据密文的统计分析得到单字母频率分布表，将密文字母按频率分类，并与明文字母分类表进行比较，得到初步了解；
- 通过研究双字母、三字母或四字母的密文组合来区分元音和辅音字母，从而确定表示元音的那些密文字母；
- 采用猜字法，像beginning,committee, people,tomorrow等在密文中也会以某种形式出现，由此进行试验。
- 综合利用英语本身的各种统计特性，对单表代换进行统计分析，破译这类单表密码。
- 如何克服这类攻击？

2.3 多表代换密码

多表代换密码就是用一个以上代换表依次对明文字母进行代换的加密方法。令明文字母表为 Z_q，$\pi=(\pi_1,\pi_2,\cdots)$ 为代换系列，$m=(m_1m_2\cdots)$ 为明文字母序列，则相应的密文字母序列是：

$$c=E_k(m)=\pi(m)=(\pi_1(m_1)\pi_2(m_2)\cdots)$$

若是非周期的无限序列，则相应的密码为非周期多表代换密码。

这类密码对每个明文字母都采用不同的代换表（或密钥），称为*一次一密钥密码*，是理论上唯一不可破的密码，它可以使明文特点完全隐蔽，需要密钥量和明文信息长度相同而难于广泛使用

为了减少密钥量，在实际应用中多采用周期多表代换密码，即代换表个数有限，重复使用，此时代换序列为 $\pi=(\pi_1,\pi_2,\cdots\pi_d,\pi_1,\pi_2,\cdots\pi_d,\cdots)$，相应于明文 **m** 的密文为

$$c=E_k(m)=\pi(m)=(\pi_1(m_1)\pi_2(m_2)\cdots\pi_d(m_d)\pi_1(m_{d+1})\pi_2(m_{d+2})\cdots\pi_d(m_{d+d})\cdots)$$

- 当d=1时就退化为单表代换

- **2.3.1几种多表代换密码**
- **1.维吉尼亚密码**
- **以加法密码为基础的周期代换密码。**

d 个代换表 $\pi = \pi_1\pi_2\cdots\pi_d$ 由 **d** 个字母序列给定的密钥 $K = (k_1, k_2, \cdots k_d) \in Z_q^d$ 决定，其中 $k_i(i=1,\cdots d)$ 确定明文第 **i+td** 个字母**(t** 为正整数**)**的移位次数**,**即

$$c_{i+td}=E_{k_i}(m_{i+td})=m_{i+td}+k_i \bmod q$$

称K为用户密钥,其周期地延伸就给出了整个明文加密所需的工作密钥。

- 例 2.4 ： 令 q=26,m=user key and working key，用户密钥K=RADIO，即d=5，则有：
- 明文m=user key and working key
- 密钥K=RADI ORA DIO RADIORA DIO
- 密文c=LSHZ YVY DVR NOUSWEG NMM

- 2.博福特密码
- 博福特密码是按mod q减法运算的一种周期代换密码，即
- $c_{i+td}=E_{ki}(m_{i+td})= k_i -m_{i+td} \bmod q$
- 博福特密码以k_i加密相当于下式的维吉尼亚密码加密：
- $c_{i+td}=((q-m_{i+td})+k_i) \bmod q$

- 3.弗纳姆密码
- 周期代换密码保密性随周期d加大而增加。
- 当d的长度和明文一样长时就是滚动密钥。
- 如果其中所采用的密钥不重复就是一次一密体制。
- 一般密钥可取自一本书或一篇报告作为密钥源
- 当字母表字母数q=2时的滚动密钥密码称为弗纳姆密码。
- 该密码方案是先将英文字母编成五单元波多电码（见下表），然后随机选择二元数字流作为密钥，用$K=k_1k_2\ldots k_l\ldots(k_l\in\{0,1\})$表示。
- 明文字母表示为$(k_i\in\{0.1\}$。

表 4.3　波多电码

A	11000	B	10011	C	01110	D	10010
E	10000	F	10110	G	01011	H	00101
I	01100	J	11010	K	11110	L	01001
M	00111	N	00110	O	00011	P	01101
Q	11101	R	01010	S	10100	T	00001
U	11100	V	01111	W	11001	X	10111
Y	10101	Z	10001	α	01000	β	00010
γ	11111	δ	11011	ε	00100	η	00000

注：α：字间隔，β：回车，γ：数字→字母，δ：字母→数字，ε：空格行；η：空格

- 弗纳姆密码的加密运算就是将k和m的相应位逐位模2相加，即
- $c_i=m_i \oplus k_i \bmod 2 \quad i=1,2\ldots,$
- 解密时，可用同样的密钥对密文数字同步地逐位模2相加，便可恢复出明文的二元码序列，即
- $m_i=c_i \oplus k_i \bmod 2 \; i=1,2\ldots,$

- 2.3.2密码分析
- 单表代换可以依靠语言的统计特性来破译，
- 多表代换下，原来的统计特性通过多个表的平均作用而被隐蔽起来，因而它的破译比单表要难。
- 在周期为d的多表代换中，字母表中的每个字母将根据它在明文字母序列中的位置而有d种不同的代换字母。
- 在多表代换下，密文字母的频率分布起伏不象明文的起伏那样明显，而且随着代换表数d的加大而更加趋于平坦。
- 但可以通过定量分析，来研究多表代换与单表代换的差别。

- 频率$p_l(l=1,2,\ldots)$与均匀分布的离差的平方和，称为粗糙度，记为M.R.。若研究对象是英文字母，则

$$M.R.=\sum_{l=0}^{25}(p_l-\tfrac{1}{26})^2=\sum_{l=0}^{25}{p_l}^2-2\sum_{l=0}^{25}\frac{p_l}{26}+\frac{1}{26}\approx\sum_{l=0}^{25}{p_l}^2-0.0385$$

对于明文或单表代换密码其粗糙度M.R.≈0.027，而均匀分布的粗糙度M.R.=0。一般的密文粗糙度应在0～0.027之间变化。

- 采用近似计算方法。
- f_l表示第l个字母在密文中出现的次数，则在有N个字母的密文中任意抽到两个字母都是第l个字母的概率为

$$\frac{C_{f_l}^2}{C_N^2}=\frac{f_l(f_l-1)}{N(N-1)}$$

把上式作为第 l 个字母的 p_l^2 的近似值，则

$$\sum_{l=0}^{25}p_l^2\approx\frac{\sum_{l=0}^{25}f_l(f_l-1)}{N(N-1)},$$

该式称为重合指数，记为 I.C.，它表示在给定密文中两个字母相同的机会。

- $Y=y_1y_2\cdots y_n$是通过维吉尼亚密码加密得到。
- 把Y分成d个长为n/d的字符串，记为$Y_1Y_2\cdots Y_{n/d}$。
- 如果d就是密钥字长度，则每个I.C.值都接近0.065.
- 可采用重码分析法来初步确定d。
- 明文中有2个相同字母组在明文序列中间隔字母数恰为d的倍数时，这2个明文字母组所对应的密文字母组必相同。
- 反之，密文中有2个相同字母组在密文序列中间隔字母数恰为d的倍数时，这2个密文字母组所对应的明文字母组不一定相同。但相同的可能性较大

- 下面是用维吉尼亚密码加密得到:
- CHREEVOAHMAERATBIAXXWTNXBEEOPH
  BSBQMQEQERBWRVXUOAKXAOSXXWEAH
  BWGJMMQMNKGRFVGXWTRZXWIAKLXFP
  SKAUTEMNDCMGTSXMXBTUIADNGMGPSR
  ELXNJELXVRVPRTULHDNQWTWDTYGBPH
  XTFALJHASVBFXNGLLCHRZBWELEKMSJI
  KNBHWRJGNMGJSGLXFEYPHAGNRBIEQJT
  AMRVLCRREMNDGLXRRIMGNSNRWCHRQ
  HAEYEVTAQEBBIPEEWEVKAKOEWADREM
  XMTBHHCHRTKDNVRZCHRCLQOHPWQAII
  WXNRMGWOIIFKEE
- 1,166,236,276,286
- 间隔165，235，275，285，
- 最大公因子为5

2.4 多字母代换密码

- 2.4.1几种多字母函数
- 1.普莱费尔密码
- 这是一种著名的双码代换密码，它的密钥由5×5阶矩阵给定，将英文字母随机填入阵中，将I和J算作一个字母。
- 可选定一个密钥字，除去重复字母后依次按行填入阵中，而后将字母表中还没用上的各字母继续按行填入阵中，就得到密钥阵K。

- 将明文划分成长为L=2的组或字母对，用(m_1,m_2)表示，它们在密钥阵K中的位置用(k_{ij},k_{ln})，假设$(m_1 \neq m_2)$，则$(m_1,m_2)=(k_{ij},k_{ln})$的密文字母为：

$$(c_1,c_2)=\begin{cases}(k_{in},k_{lj}) & \text{若} i\neq l,\ \text{且} j\neq n\\ (k_{i,j+1},k_{i,n+1}) & \text{若} i=l,\ \text{且} j\neq n\\ (k_{i+1,j},k_{l+1,j}) & \text{若} i\neq l,\ \text{且} j=n\end{cases}$$

式中，下标指数按模5运算。

如果$m_1=m_2$，则可在m_1和m_2之间插入哑字母如x，则明文变为…m_1xm_2…。

- 例2.5：明文m=cryptographic system，取cryptography为密钥字，则其密钥阵就是：

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

相应加密结果是：

m=cr yp to gr ap hi cs ys te mx

c=RY PT CB EG HY IQ TL TN RK QV

2.矩阵变换密码

令明文字母表为 Z_q，若采用 L 个字母为单位进行代换，则多字母代换为映射 $f: Z_q^L \to Z_q^L$。

若映射是线性变换，可用Z_q上的L×L矩阵T表示
若T是满秩的，则变换为一一对应映射，存在逆变换T^{-1}，使$TT^{-1}=T^{-1}T=I$。
将L个字母的数字表示为Z_q上的L维向量$m=(m_1,m_2,\ldots,m_L)$，
则相应的密文向量$c=(c_1,c_2,\ldots,c_L)$为$mT=c$。
把T^{-1}作为解密矩阵，可由c恢复出相应明文$cT^{-1}=m$

- **例2.7：设q=26，L=4，选满秩阵**

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

加密时，先将字母按下述乱序表变换成Z_{26}上的整数：

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

3 15 4 10 1 13 7 24 2 16 9 14 5 12 20 6 18 25 8 19 22 21 0 17 23 11

对明文m=cryptographic system的前4个字母组变换成向量x=(4,25,23,6)

由xT(mod26)得密文y=(3,8,2,25)，相应密文字母为ASIR

类似可依次对后面明文组加密。

若最后一组明文不足4个字母，就加上哑字母（如x）凑足4个。

最后可求得密文为ASIR QGLT WCFW BMSK IPWG

- 加密中，4个字母为一个整体，变换其中任一个明文字母都会使相应的4个密文字母受到影响。例如，将cryp变为crmp所得密文就由ASIR变为YUSK。
- 该例的逆阵

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

由$yT^{-1}(mod26)$可得到x,由x及乱序表就可恢复出明文。

- 3.扩表法
- 扩表法就是将字母表Z_q扩展为Z_q^L，然后再采用单表代换。
- 将Z_q上的L个字母$m_0,m_1\ldots,m_{L-1}$表示成Z_q^L中的整数
- $x=m_0+m_1q+m_2q^2+\ldots+m_{L-1}q^{L-1}$
- 任一L长字母组都变成$Z_q^L=\{0,1,\ldots,q^{L-1}\}$上的某个元素，然后再对该元素加密。

- 例2.8：英文字母表$Z_{26}=\{0,1,\ldots,25\}$，采用双码扩表加密，则$Z_{26}^2=\{0,1,\ldots,675\}$。
- 明文字母对is可表示成Z_{26}^2中的数字$x=8+18\times 26=476$。
- 对Z_{26}^2采用仿射变换，取$k_0=576,k_1=129$
- $y\equiv 129\times 476+576\equiv 4643$
- $\equiv 17\times 26+22(\mathrm{mod}676)$,
- 相应密文字母为WR。
- 解密时先将密文字母WR变换成Z_{26}^2中的元素，即$y=17\times 26+22=464$，而后按加密的反变换求解出$x=18\times 26+8$。

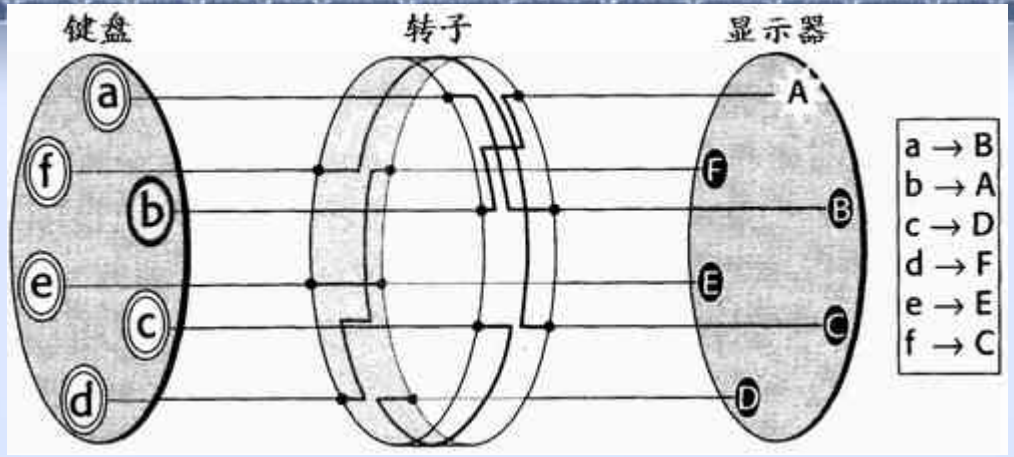
2. 4. 2密码分析

- 先计算给定密文的频率分布和I.C.值，以排除单表代换，
- 然后统计重复字母的频率、间隔和间隔的最大公钥数d，并将密文排成d行，计算各行的I.C.值，
- 如果多数与0.065相差较大时，就可排除多表代换。
- 并根据d值确定多字母加密每组的个数。
- 若密文是由扩展法实现的双字母代换，则可将两个相邻字母看作是中的一个元素。
- 对于双字母英文代换而言，就是676个字母的单表代换，其I.C.值约为0.0069。
- 如果密文按双字母出现频率计算的I.C.值与该值接近，就可假定是双字母代换，进行试译。
- 对于矩阵变换型密码的破译，可通过已知明文——密文字母组求出密钥T和T^{-1}。

ENIGMA 密码机

- 第一次世界大战结束为止，所有密码都是使用手工来编码的 。
- 谢尔比乌斯发明的加密电子机械名叫ENIGMA
- 当时最为可靠的加密系统之一

水平面板的下面部分就是键盘，一共有26个键，键盘排列接近现在使用的计算机键盘。为了使消息尽量地短和更难以破译，空格和标点符号都被省略。



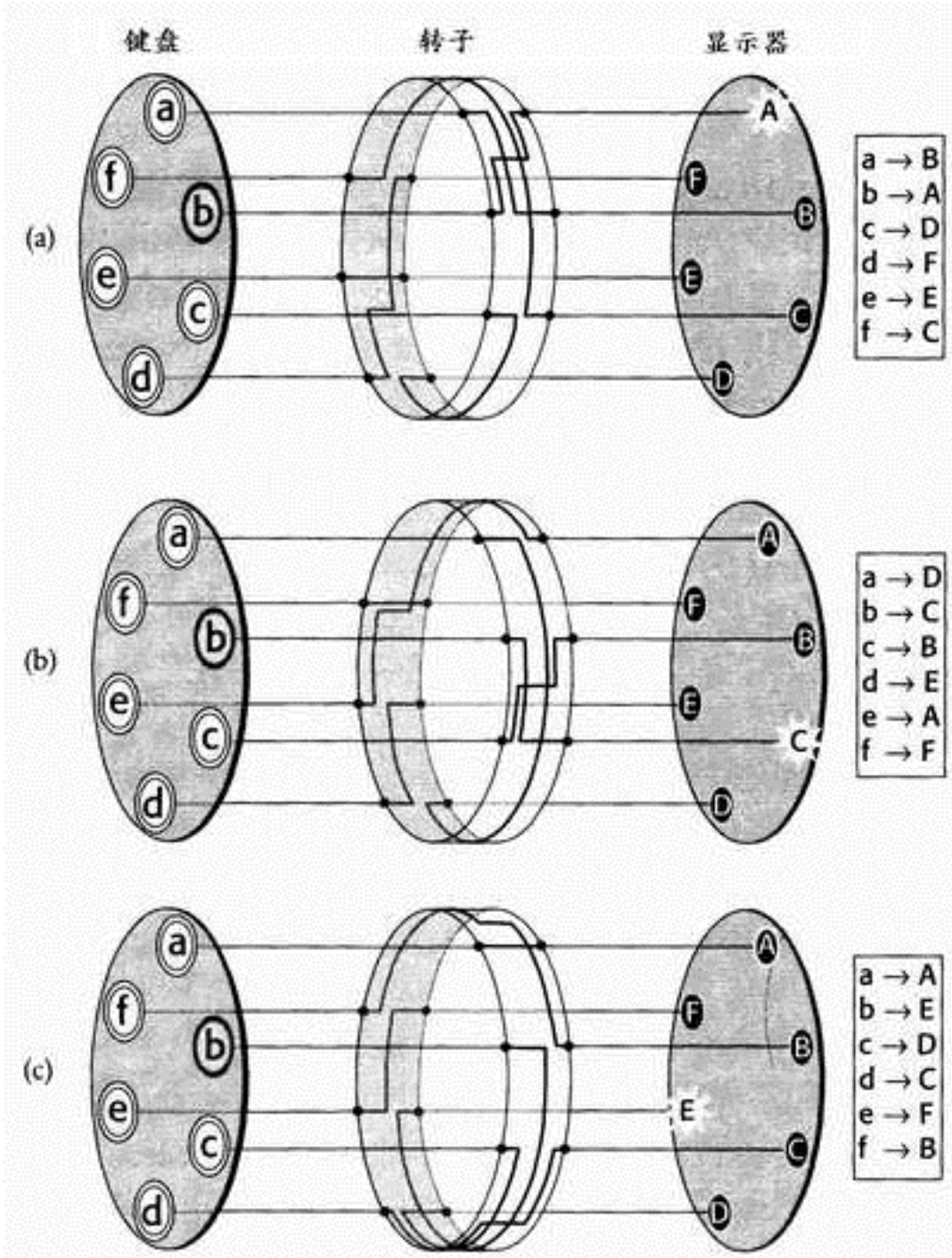
示意图六个键

实物照片中，键盘上方就是显示器，它由标示了同样字母的26个小灯组成，当键盘上的某个键被按下时，和此字母被加密后的密文相对应的小灯就在显示器上亮起来。

在示意图上只画了六个小灯。在显示器的上方是三个转子，它们的主要部分隐藏在面板之下，在示意图只画了一个转子。

- 键盘、转子和显示器由电线相连，转子本身也集成了6条线路（在实物中是26条），把键盘的信号对应到显示器不同的小灯上去。
- 如果按下a键，那么灯B就会亮，这意味着a被加密成了B。同样地b被加密成了A，c被加密成了D，d被加密成了F，e被加密成了E，f被加密成了C。
- 依次键入cafe（咖啡），显示器上就会依次显示DBCE。
- “简单替换密码”。

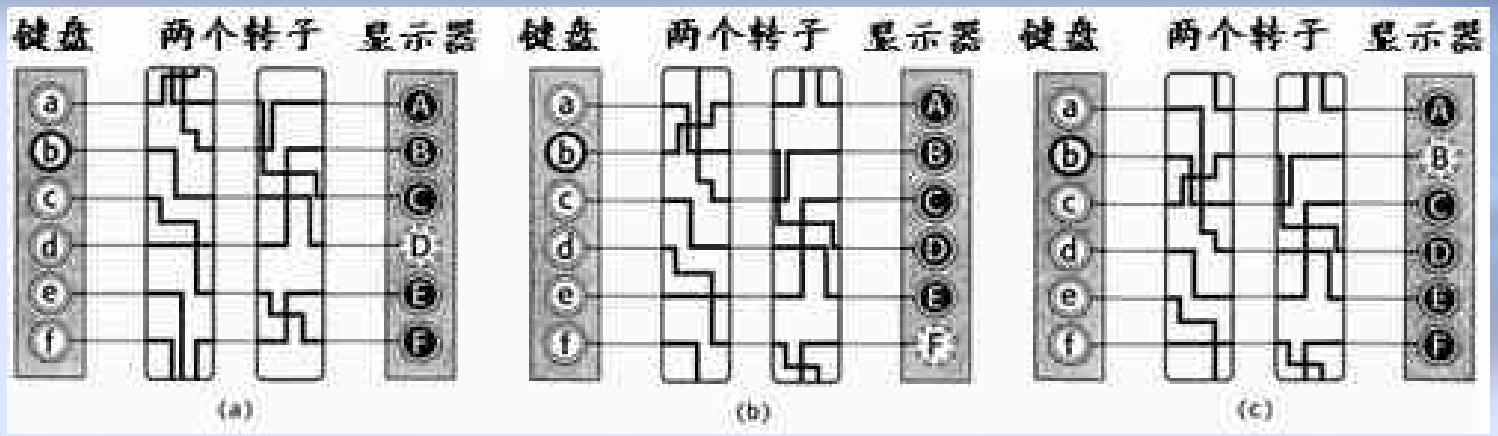
- 转子的作用若仅是把一个字母换成另一个字母，那就没有价值了。
- 但是“转子”会转动！
- 谢尔比乌斯关于ENIGMA的最重要的设计——当键盘上一个键被按下时，相应的密文在显示器上显示，然后转子的方向就自动地转动一个字母的位置（在示意图中就是转动1/6圈，而在实际中转动1/26圈）。
- 示意图表示了连续键入3个b的情况：



当第一次键入b时，信号通过转子中的连线灯A亮起来，放开键后转子转动一格，各字母所对应的密码就改变了

第二次键入b时，它所对应的字母就变成了C；同样地，第三次键入b时，灯E闪亮。

- ENIGMA加密的关键：这不是一种简单替换密码。同一个字母b在明文的不同位置时，可以被不同的字母替换，而密文中不同位置的同一个字母，可以代表明文中的不同字母，频率分析法在这里就没有用武之地了。这种加密方式被称为“复式替换密码”。
- 但是如果连续键入6个字母（实物中26个字母），转子就会整整转一圈，回到原始的方向上，这时编码就和最初重复了。
- 而在加密过程中，重复的现象是很危险的，这可以使试图破译密码的人看见规律性的东西。于是谢尔比乌斯在机器上又加了一个转子。当第一个转子转动整整一圈以后，它上面有一个齿拨动第二个转子，使得它的方向转动一个字母的位置。

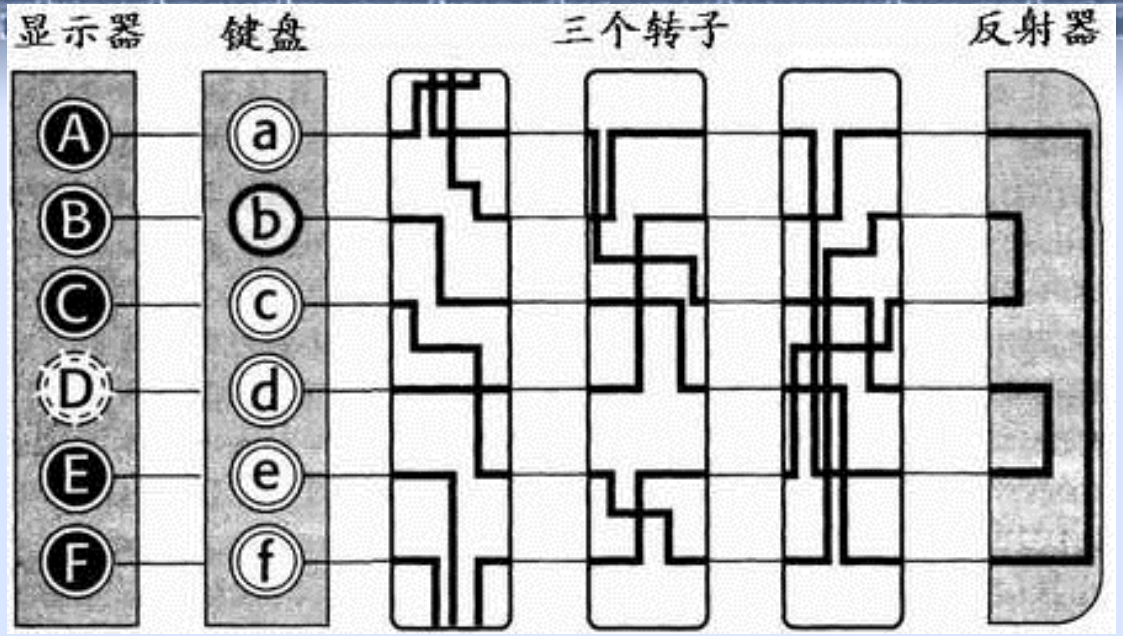


这里(a)图中假设第一个转子（左边的那个）已经整整转了一圈，按b键时显示器上D灯亮；当放开b键时第一个转子上的齿也带动第二个转子同时转动一格，

于是(b)图中第二次键入b时，加密的字母为F；而再次放开键b时，就只有第一个转子转动了，于是(c)图中第三次键入b时，与b相对应的就是字母B。

用这样的方法，要6*6=36（实物中为26*26=676）个字母后才会重复原来的编码。而事实上ENIGMA里有三个转子（二战后期德国海军用ENIGMA甚至有四个转子），不重复的方向个数达到26*26*26=17576个。

- 在此基础上谢尔比乌斯十分巧妙地在三个转子的一端加上了一个反射器，而把键盘和显示器中的相同字母用电线连在一起。反射器和转子一样，把某一个字母连在另一个字母上，但是它并不转动。



这里键盘和显示器中的相同字母由电线连在一起。

一个很巧妙的开关，当一个键被按下时，信号不是直接从键盘传到显示器（要是这样就没有加密了），而是首先通过三个转子连成的一条线路，然后经过反射器再回到三个转子，通过另一条线路再到达显示器上，比如说上图中b键被按下时，亮的是D灯。

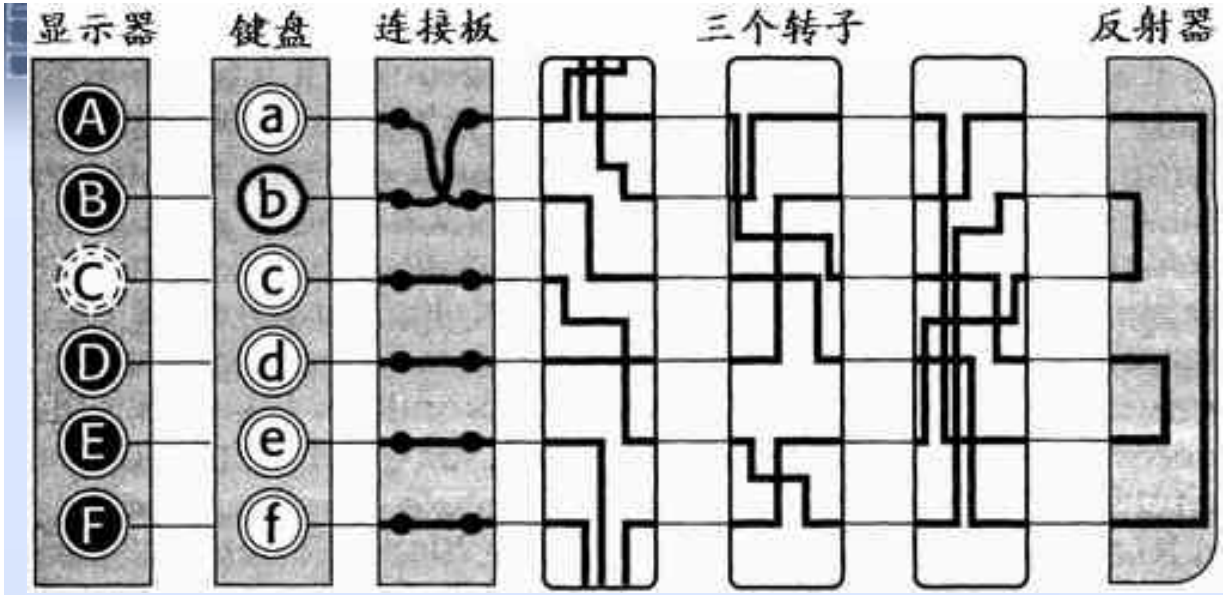
如果这时按的不是b键而是d键，那么信号恰好按照上面b键被按下时的相反方向通行，最后到达B灯。换句话说，在这种设计下，反射器虽然没有象转子那样增加可能的不重复的方向，但是它可以使译码的过程和编码的过程完全一样。

- 用ENIGMA发送一条消息。发信人首先要调节三个转子的方向，使它们处于17576个方向中的一个（事实上转子的初始方向就是密匙，这是收发双方必须预先约定好的），然后依次键入明文，并把闪亮的字母依次记下来，然后就可以把加密后的消息用比如电报的方式发送出去。当收信方收到电文后，使用一台相同的ENIGMA，按照原来的约定，把转子的方向调整到和发信方相同的初始方向上，然后依次键入收到的密文，并把闪亮的字母依次记下来，就得到了明文。于是加密和解密的过程就是完全一样的——这都是反射器起的作用。
- 反射器带来的一个副作用就是一个字母永远也不会被加密成它自己，因为反射器中一个字母总是被连接到另一个不同的字母。

- 于是转子的初始方向决定了整个密文的加密方式。如果通讯当中有敌人监听，他会收到完整的密文，但是由于不知道三个转子的初始方向，就不得不一个个方向地试验来找到这个密匙。问题在于17576个初始方向这个数目并不是太大。如果试图破译密文的人把转子调整到某一方向，然后键入密文开始的一段，看看输出是否象是有意义的信息。如果不象，那就再试转子的下一个初始方向……如果试一个方向大约要一分钟，而二十四小时日夜工作，那么在大约两星期里就可以找遍转子所有可能的初始方向。
- 如果对手用许多台机器同时破译，那么所需要的时间就会大大缩短。这种保密程度是不太足够的。

- 多加转子，但是每加一个转子初始方向的可能性只是乘以了26。尤其是，增加转子会增加ENIGMA的体积和成本。
- 加密机器便于携带的，而不是一个具有十几个转子的庞然大物。
- 三个转子做得可以拆卸下来互相交换，这样一来初
- 始方向的可能性变成了原来的六倍。假设三个转子的编号为1、2、3，那么它们可以被放成123-132-213-231-312-321六种不同位置，当然收发消息的双方除了要预先约定转子自身的初始方向，还要约定好这六种排列中的使用一种。

- 在键盘和第一转子之间增加了一个连接板。这块连接板允许使用者用一根连线把某个字母和另一个字母连接起来，这样这个字母的信号在进入转子之前就会转变为另一个字母的信号。这种连线最多可以有六根（后期的ENIGMA具有更多的连线）
- ６对字母的信号互换，其他没有插上连线的字母保持不变。
- 当然连接板上的连线状况也是收发信息的双方需要预先约定的



在上面示意图中，当b键被按下时，灯C亮。
于是转子自身的初始方向，转子之间的相互位置，以及连接板连线的状况就组成了所有可能的密匙，让我们来算一算一共到底有多少种。
三个转子不同的方向组成了26*26*26=17576种不同可能性；
三个转子间不同的相对位置为6种可能性；
连接板上两两交换6对字母的可能性数目非常巨大，有100391791500种；
于是一共有17576*6*100391791500，大约为10000000000000000，
即一亿亿种可能性。

- 只要约定好上面所说的密匙，收发双方利用ENIGMA就可以十分容易地进行加密和解密。但是如果不知道密匙，在这巨大的可能性面前，一一尝试来试图找出密匙是完全没有可能的。
- 连接板对可能性的增加贡献最大，那么为什么谢尔比乌斯要那么麻烦地设计转子之类的东西呢？
- 原因在于连接板本身其实就是一个简单替换密码系统，在整个加密过程中，连接是固定的，所以单使用它是十分容易用频率分析法来破译的。转子系统虽然提供的可能性不多，但是在加密过程中它们不停地转动，使整个系统变成了复式替换系统，频率分析法对它再也无能为力，与此同时，连接板却使得可能性数目大大增加，使得暴力破译法（即一个一个尝试所有可能性的方法）望而却步。

2.5 Shannon理论

- Shannon 在 1949 年 发 表 了 “Communication theory of secrecy system”一文，用信息论的观点对信息加密问题作了全面阐述，深化了人们对密码学的理解，使信息论成为研究密码学的一个重要理论基础。

- 在19世纪主要是发明一些更加高明的加密技术，这些技术的安全性通常依赖于用户赋予它们多大的信任程度。
- 在19世纪Kerchoffs写下了现代密码学的原理。其中一个的原理提到：加密体系的安全性并不依赖于加密的方法本身，而是依赖于所使用的密匙。
- 为近代密码学指明了方向

- Shannon在1949年发表了“Communication theory of secrecy system”一文，用信息论的观点对信息加密问题作了全面阐述，深化了人们对密码学的理解，使信息论成为研究密码学的一个重要理论基础。

- 在密码系统中，对信息m的加密变换作用类似于向信息注入噪声。
- 密文c就相当于经过有扰信道得到的接收信息，密码分析就相当于在有扰信道下去除噪声，恢复原文。
- 所不同的是，这种干扰不是信道中的自然干扰，而是发送者有意加进的，目的是使窃听者难以恢复出原来的信息。
- Shannon从概率统计观点出发研究信息的传输和保密问题，将通信系统归为图2.1，将密码系统归为图2.2。

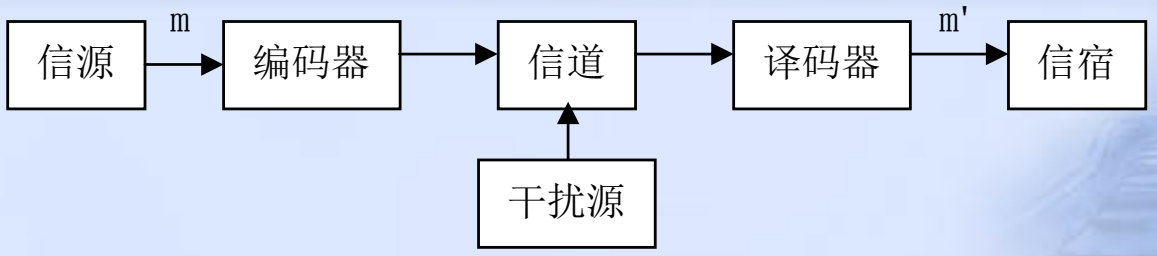


图 2.1　通信系统

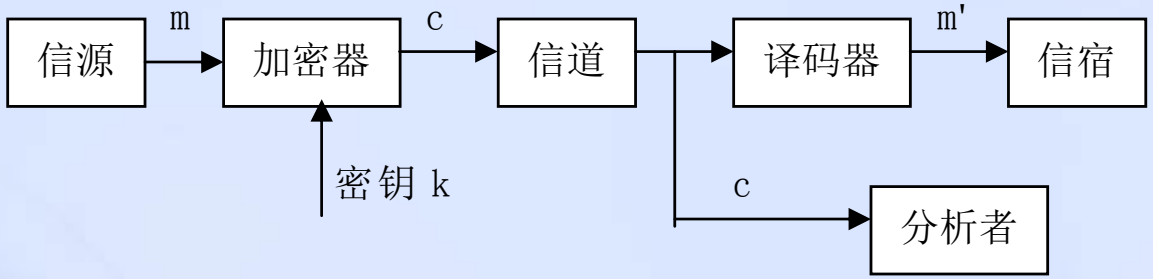


图 2.2　密码系统

通信系统设计目的是在信道有扰条件下，使接收的信息无错或差错尽可能地小。

密码系统设计的目的在于使窃听者即使在完全准确地收到了接收信号也无法恢复出原始信息。

在密码系统研究中，假定信道是没有自然干扰的。

2.4.5信息量和熵

定义 2.1：对于集合 $X=\{x_1,x_1,\cdots,x_n\}$，有 $\sum_{i=1}^{n} p(x_i)=1$。这里 $p(x_i)$为事件 x_i 出现的概率。x_i 出现给出的信息量定义为：$I(x_i)=-\log_a p(x_i)$。当 a=2 时，相应的信息单位称为比特(bit)。

信息量反映了事件x_i出现的可能性大小，也是为确定事件x_i出现所必须付出的信息量。

定义 2.2：将集合 X 中事件出现给出的信息量的统计平均值

$H(X)=-\sum_{i=1}^{n} p(x_i)\log_2 p(x_i)\geq 0$称为集合 X 的熵。当 $p(x_i)=0$ 时，定义 $p(x_i)\log_a p(x_i)=0$。

- 熵表示集合X中出现一个事件平均给出的信息量，或者是集合X中事件的平均不确定性，或者是确定集合X出现一个元素必须提供的信息量。
- 定理2.2：对任意有n个事件的集合X有：$0\leq H(X)\leq \log_2 n$。

设有两个事件集 $X=\{x_1,x_2,\cdots,x_n\}$和 $Y=\{y_1,y_2,\cdots,y_n,\}$，其联合事件集为 $XY=\{x_iy_j,i=1,\cdots n,j=1,\cdots,m\}$，$x_iy_j$ 的概率为 $p(x_iy_j)$，显然$\sum_{i=1}^{n}p(x_i)\sum_{j=1}^{m}p(y_j\mid x_i)=\sum_{j=1}^{m}p(y_j)\sum_{i=1}^{n}p(x_i\mid y_j)=1$。

定义 2.3：对于集合 X 和 Y 的联合事件集 XY 中事件出现给出的信息量的统计平均值 $H(XY)=-\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_iy_j)\log_2 p(x_iy_j)$ 称为 X 和 Y 的联合熵。称 $H(X\mid Y)=-\sum_{j=1}^{m}\sum_{i=1}^{n}p(x_iy_j)\log_2 p(x_i\mid y_j)$ 和 $H(Y\mid X)=-\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_iy_j)\log_2 p(y_j\mid x_i)$ 为条件熵，又称为含糊度。

- 定理2.3：对任意有限事件集合X、Y有：
- (1)$H(XY)=H(YX)=H(X)+H(Y|X)=H(Y)+H(X|Y)$
- (2)$H(X|Y)\leq H(X)$，等式成立当且仅当X和Y统计独立。
- (3)$H(Y|X)\leq H(Y)$，等式成立当且仅当X和Y统计独立。

定义 2.4：将 X、Y 分别看作一个系统的输入、输出空间，对于 $x_i \in X$，$y_j \in Y$，将由 y_j 得到的关于 $x_i \in X$，出现的信息量定义为：

$$I(x_i; y_j) = \log_2 \frac{p(x_i \mid y_j)}{p(x_i)}$$

称 $I(x_i; y_j)$ 为事件 xi 和 yj 之间的互信息

- 当$p(x_i|y_j)>p(x_i)$时，$I(x_i;y_j)>0$；
- 当$p(x_i|y_j)=p(x_i)$时，$I(x_i;y_j)=0$；
- 当$p(x_i|y_j)<p(x_i)$时，$I(x_i;y_j)<0$；
- 因为$p(x_iy_j)=p(x_i)p(y_j|x_i)=p(y_j)p(x_i|y_j)$；
- 所以有$I(x_i;y_j)=I(y_j;x_i)$。 。
- 这说明两个集合中的一对事件可以相互提供的信息量相等。

定义 2.5：$I(x_i;y_j)$称为事件 x_i 和 y_j 之间的互信息。它们的统计平均值 $I(X;Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_iy_j)I(x_i;y_j)$ 称为集合 X 和 Y 之间的平均互信息量。称 C=maxI(X;Y)为信道容量，它表示通过信道可传送的最大信息量。

•容易证明，I(X;Y)=H(X)-H(X|Y)=H(Y)-H(Y|X)=H(X)+ H(Y)-H(XY)

- 例 2.8 ： 令 $X=\{x_1,x_2\}$ ， $Y=\{y_1,y_2\}$ ，$I(X;Y)=H(Y)-H(Y|X)=H(Y)+p\log_2 p+(1-p)\log_2(1-p)$，
- 对于H(Y),当$p(y_1)=p(y_2)=1/2$时，H(Y)取最大值1，
- 故$C=1+p\log_2 p+(1-p)\log_2(1-p)$。
- 称这样的系统为二元对称信道。

- **2.5.2完善保密性**
- 令明文熵为$H(M^L)$，密钥熵为$H(Y)$，密文熵为$H(C^V)$(L为明文分组长度，V为密文分组长度)
- 在已知密文条件下明文的含糊度为 $H(M^L|C^V)$，在已知密文条件下密钥的含糊度为$H(Y|C^V)$。
- 从唯密文破译来看，密码分析者的任务是从截获的密文中提取有关明文的信息：$I(M^L;C^V)=H(M^L)-H(M^L|C^V)$，
- 或从密文中提取有关密钥的信息：$I(Y;C^V)=H(Y)-H(Y|C^V)$。
- 对于合法接受者来说，则是在已知密钥和密文条件下提取明文信息，
- 由加密变换的可逆性知，$H(M^L|C^VY)=0$，
- 故有$I(M^L;C^VY)=H(M^L)$。

- $H(Y|C^V)$和$H(M^L|C^V)$越大，窃听者从密文中能提取的有关明文和密钥的信息就越小。
- 定理2.4：对任意密码系统有：$I(M^L;C^V)\geq H(M^L)- H(Y)$
- 证明留作习题。
- 定义4.6：一个密码系统，若其明文与密文之间的互信息$I(M^L;C^V)=0$，则称该系统为完善的密码系统。

- $0=I(M^L;C^V)=H(M^L)-H(M^L|C^V)$,
- 即$H(M^L|C^V)=H(M^L)$。
- 由定理2.3(2)($H(X|Y)\leq H(X)$，等式成立当且仅当X和Y统计独立)知，
- M^L和C^V统计独立，
- 即对任意的$m_i\in M^L$，$c_j\in C^V$，必有$p(m|c)=p(m)$。

- 定理2.5：存在完善的密码系统。
- 证明：采用构造法。不失一般性，假定明文是二元数字序列$m=(m_1,m_2,\ldots,m_L)$，这里$m_i\in GF(2)$。
- 令密钥序列$k=(k_1,k_2,\ldots,k_r)$和密文序列$c=(c_1,c_2,\ldots,c_v)$也为二元序列，这里m和k彼此独立。
- 选$L=r=v$，并令k为随机数字序列，即对一切$k\in K$，有$p_K(k)=1/2^L$。
- 加密变换采用弗纳姆密码体制，则有$c=E_k(m)=m\oplus k$，其中加法是逐位按模2进行，即$c_l(m)=m_l\oplus k_l$。
- 解密变换为$m=D_k(c)=c\oplus k$，加法也是逐位按模2进行。

- 要证明I(M;C)=0，即证明H(M)=H(M|C)，也就是证明M和C统计独立，即对任意的m∈M，c∈C，必有p(m|c)=p(m)。
- 定理2.5构造的系统在唯密文破译下是安全的，但在已知明文攻击下是不安全的。
- 若知道明文——密文对(m',c')，则由c'=m'⊕k可求得k= m'⊕c'。
- 因此为抗已知明文攻击，就要求密钥不能重复使用，即采用一次一密体制。
- 在此体制下，任何已知明文——密文对都无助于破译以后收到的密文。

- Shannon最先证明这种体制是完善保密的，且可抗已知明文——密文对攻击。
- 但在实际使用中，每通信一次就要更换密钥，而密钥传送的安全性就构成一个薄弱环节。
- Shannon理论中一个最大的弱点就是没有考虑密钥传送问题，即认为密钥的传送不经过密码系统本身。

2.5.3实际保密性

- 在讨论完善保密性时，是假定分析者有无限时间、设备和资金条件下，研究唯密文攻击时密码系统的安全性。
- 一个密码系统的破译，如果对手有无限资源可利用，而在截获任意多密文下仍不能被破译，则在理论上是保密的。
- 实际上，密码分析者所具有资金设备和时间总是有限的，且可采用统计分析、已知明文——密文对等手段攻击。
- 在这种情况下，研究密码体制的安全性，就是研究系统的实际保密性。

- 一个密码系统的破译所需努力，如超过对手的能力时，该系统实际上是保密的
- 理论上不安全的系统可能提供实际上的安全保密性。而理论上安全的，在实际上也可能是脆弱的。
- 估计一个系统的实际保密性？
- 最主要的是须考虑两个因素：一是密码分析者的计算能力，二是他所采用的破译算法的有效性。
- 密码分析者的计算能力取决于他所拥有的资源条件。

- 在估计系统的保密性时，首先要估计破译它所需的基本运算次数和存储量，而后考虑在现有设备条件下所需时间。
- 破译算法的有效性对系统安全也是十分重要的。
- 密码分析者总是在研究新的破译方法来减少破译所需的运算量。
- 因此必须确保所使用的密码系统没有轻而易举的破译方法。

- 密码系统设计者应考虑Shannon 理论的一些重要结论，保证密钥空间足够大，同时要遵循扩散、混乱和均衡的原则。
- 所谓扩散，就是根据Shannon理论，应尽可能隐蔽明文信息，保证当某个明文位发生变化时，整个密文都应全部变化；
- 混乱则应保证明文的所有特性都被打乱，从密文中看不到明文的任何信息；
- 均衡是指密文中0,1字符均衡，而不管明文中0，1字符情况。
- 根据现有设备和资源条件，综合考虑已有和可能有的破译算法有效性，设计出满足实际保密要求的密码系统，以使密码体制在现有和不久将来具有的设备和资源条件下无法破译。

2.5 序列密码

- 若能以一种方式产生一随机序列，且能在解密时能重复产生，则就可利用这样的序列进行加密。
- 要求该序列由一特定的密钥所确定。

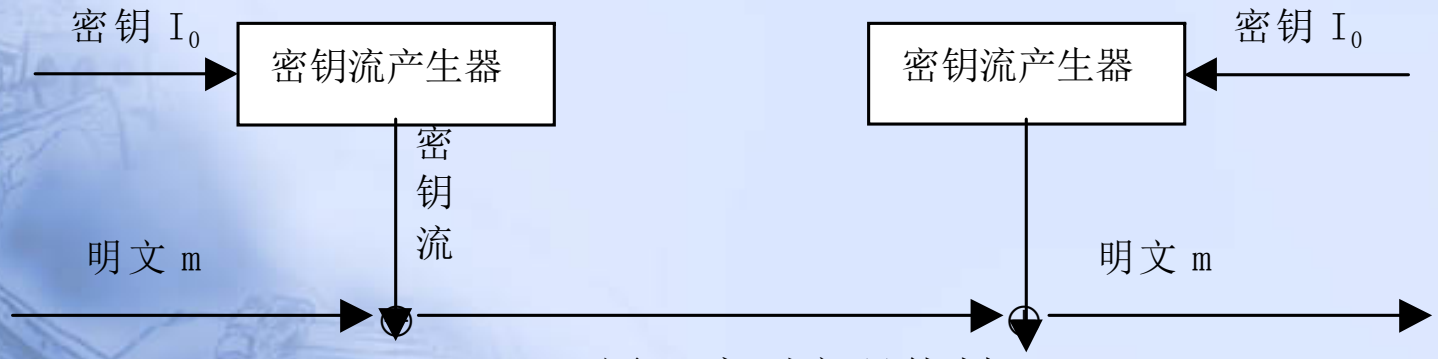


图　序列密码体制

- 密钥流产生器实际上是一给定的算法，产生的密钥流通常是0-1数据流。
- 由于没有任何有限的算法可产生真正的随机序列，因此这里的序列是有周期的，不可能做到随机，只能要求截获比周期短的一段时不会泄露更多的信息。
- 这样的序列称为伪随机序列。

- 2.5.1序列密码加密方式分类
- 序列密码加密时一般采用同步和自同步两种方法。
- 1.同步序列密码
- 同步序列密码的特征是它直接作用于信息的密钥序列（工作密钥）$K_s=k_1k_2\ldots$的产生与信息序列无关。作为工作密钥，它必须与被加密的信息等长。它由一个随机性好的主密钥通过密钥流生成器产生。

- 同一明文字符在不同时刻由于密钥不同而被加密成不同的密文字符。
- 要求收发两端的密钥流生成器的初始密钥相同，输出的密钥就一样。
- 只有保持两端精确同步才能正常工作，一旦失步就不能正确解密，这是其主要缺点。
- 正是其对失步的敏感性，使得系统在有窜扰者进行注入、删除、重放等主动攻击时异常敏感而有利于检测。
- 此类体制的优点是当传输中出现一些偶然错误时，只影响相应位的信息恢复，没有差错扩散。
- 但也是信息安全的弱点，即明文位某个信息改变只影响相应位的密文，而没有扩散效应。

- 实际使用中工作密钥是有周期的，而长度为周期的明密文对就可破译出工作密钥，因此周期应足够长，
- 并在产生工作密钥时，再引入一个初始化向量IV（称为种子量），
- 工作密钥通过以主密钥K和种子量IV生成，这样不同的初始状态就会导致不同的工作密钥。

- 2.自同步序列密码
- 明文也参加工作密钥的生成，即工作密钥流k_i历史地与$c_1c_2\cdots c_{i-1}$有关，
- 则密文c_i不仅与当前明文m_i有关，而且与以前的密文$c_1c_2\cdots c_{i-1}$有关。
- 一般在有限的n级存储器下，将与$c_{i-n}c_{i-n+1}\cdots c_{i-1}$有关。这类序列密码的密钥流可由下式表示：
- $k_i=f(K,IV,c_{i-n}c_{i-n+1}\cdots c_{i-1})$
- 此时，从某种意义上讲工作密钥流是在不断变化的

- 自同步序列密码在传输过程中有一位出错，就会影响其后n位密钥的正确性，相应地对明文信息的恢复也将连续n位受到影响。
- 但这类体制，只要收方连续正确收到n位密文，就会产生相同的密钥的，因而它具有自同步能力。
- 虽然它对窜扰者的主动攻击不像同步序列密码那样敏感，但它由于将明文每个字符扩散到密文多个字符中，强化了其抗统计分析的能力。

- 2.5.2密钥流的生成
- 需要能产生周期足够长、随机性又相当好的序列。
- 从50年代开始，以有限自动机为主流的理论和方法得到了迅速发展。
- 近年来研究的混沌密码、胞元自动机密码等，在有限精度的数字实现条件下，最终同样可归结为有限自动机来描述。

- 作业:P71 1，2，3,4,5
- 证明定理4.3(2)