第4章 密码学基础及密码算法

东南大学 网络空间安全学科
胡爱群 教授/博导

主要内容

密码学概述

- 密码学的基本概念（密码编码、密码分析、密码管理）
- 基本编码原理（移位密码、代替密码）
- 代替密码分析（统计规律、单表代替分析、多表代替分析）

保密理论

- 信息论简介
- Shannon保密理论
- 计算复杂性理论

1.1 密码学引言

- 网络空间安全的理论基础，大多数应用离不开密码理论的支撑；
- 1949年以前密码技术是一种技巧；
- 1949年Shannon发表《保密系统的通信理论》；
- 1970年IBM推出DES（数据加密标准）密码算法；
- 1976年Diffie和Hellman发表了《密码学的新方向》，引入公钥密码的概念，基于公开信息的密钥交换和互不信任双方的信息认证问题成为可能；
- 广泛应用于信息加密、真实性认证、完整性保护、产品防伪等。

1.2 密码学基本概念

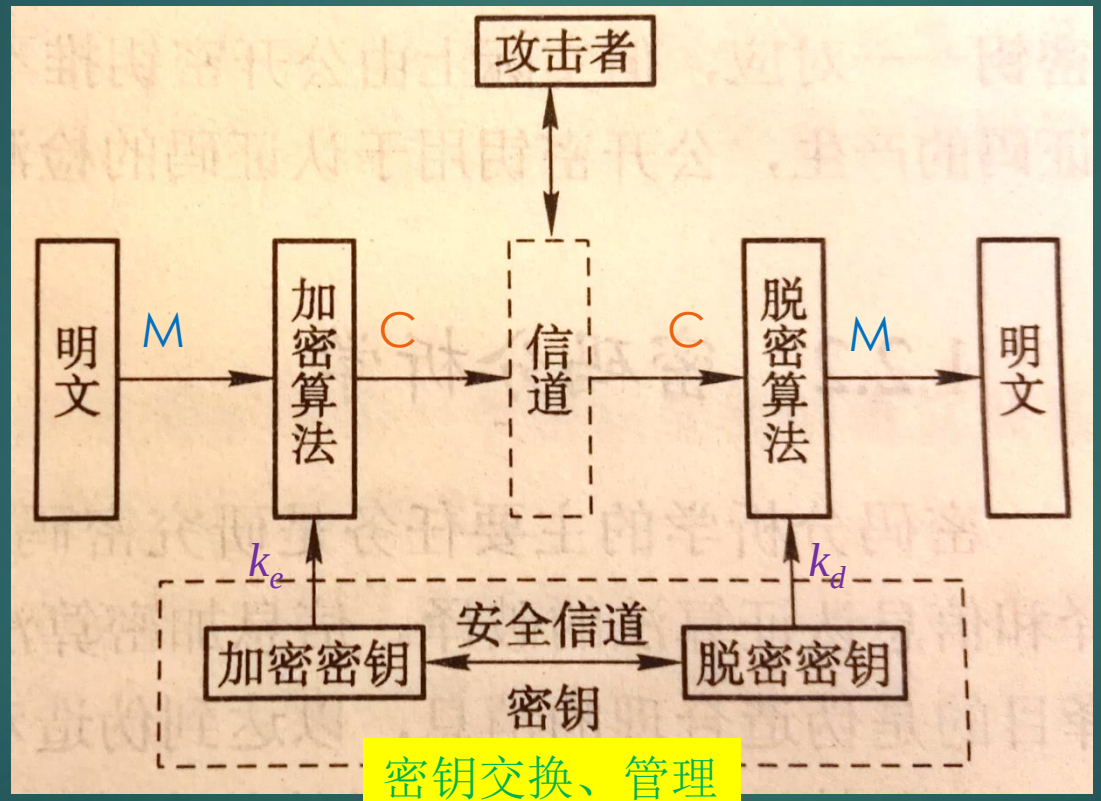
- 密码学目标： 机密性、真实性、不可否认性；
- 密码学包括：密码编码学、密码分析学和密钥管理学；

密码编码学：研究安全、高效的信息加密算法和信息认证算法的设计理论与技术；

一个密码体制：明文空间**M**、密文空间**C**、密钥空间**K**、加密算**E**、解密算法**D**五个部分组成。

$$\begin{cases} \mathrm{c} = E_{k_e}(\mathrm{m}) \\ m = D_{k_d}(c) \end{cases} \qquad \forall m \in M, c \in C, k_e \in K, k_d \in K$$

密码通信系统基本结构



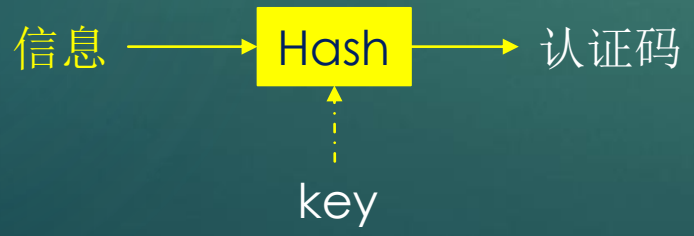
密码体制应满足以下要求:

- 即使达不到理论上的不可破，也应当是实际不可破的。即应能抵挡各种可能的攻击；
- 一切秘密蕴含于密钥之中。即只要敌手不知道密钥，就不能由已知信息推出未知明文；
- 加密算法和解密算法必须适用密钥空间中所有可能的值，即弱密钥应尽可能少；
- 应具有很好的实现性能，即应满足实际需要。

- 对称密码体制：加密密钥与解密密钥相同，即单密钥密码体制； $k_e = k_d$
- 公钥密码体制：加密密钥与解密密钥不同，即非对称密码体制、双钥密码体制。 $k_e \neq k_d$

信息认证算法

- 对信息进行一种变换，变换的结果即为认证码。当信息与其认证码不匹配时，即可判断信息与认证码至少有一个是不真实的。
- 信息认证算法就是为信息产生人为的冗余，并利用这种冗余来检测信息的真实性；
- 信息认证算法包括认证码生成算法和认证码检验算法。
- 信息认证算法包括：无密钥认证算法、单密钥认证算法和双密钥认证算法。



密码分析学

- 研究密码破译的理论与技术，包括信息加密算法和信息认证算法的破译；
- 信息加密算法的破译是要获取未授权的信息（密钥或明文）；
- 信息认证算法的破译是要伪造合理的消息，达到伪造和欺骗的目的
- 密码算法抗破译的能力称为密码强度；密码破译又称密码分析。

Kerckhoffs假设：敌手知道除密钥以外的任何信息。

密码破译基于以下假定：

- 所使用的密码体制
- 明文概率分布规律
- 密钥概率分布规律
- 所有破译方法

密码破译可利用的规律

三大规律：密码规律、文字规律、情况规律。

- 密码规律：明文、密文的对照关系，密钥与明文的对照关系；
- 文字规律：明文的文意和格式的规律；
- 情况规律：明文的内容与当前发生的事件的关系。

密码破译：发掘出密码算法的信息泄露规律及其利用方法，借助文字规律和情况规律，恢复出密钥或明文。

字母	空格	E	T	O	A	N	I	R	S
频率	0.2	0.105	0.071	0.0644	0.063	0.059	0.054	0.053	0.052
字母	H	D	L	C	F	U	M	P	Y
频率	0.047	0.035	0.029	0.023	0.0221	0.0225	0.021	0.0175	0.012
字母	W	G	B	V	K	X	J	Q	Z
频率	0.012	0.011	0.0105	0.008	0.003	0.002	0.001	0.001	0.001

英文字母出现概率

对加密算法的攻击类型

除具有前面的基本条件外：

- 唯密文攻击：具有足够多的采用同一密钥加密的密文。
- 已知明文攻击：不仅具有唯密文攻击条件，还具有足够多的采用同一密钥加密的密文和对应的明文。
- 选择明文攻击：不仅具有上面的条件，还可以选择对破译有利的明文及对应密文。
- 选择密文攻击：不仅具有已知明文攻击的条件，还可以选择对破译有利的密文及对应明文；
- 相关密钥攻击：不仅具有上述条件，而且具有所求密钥的相关密钥及所对应的明文和密文。

- 穷举攻击：穷尽所有密钥，解密密文，并检测所得明文是否正确；
- 解析攻击：针对密码算法设计所依赖的数学问题，利用数学求解方法破解；
- 统计攻击：利用明文、密文内在统计规律破解密码；对称密码算法的破解大都基于此；
- 代数攻击：将密码破解问题归结为有限域上的某个低次的多元代数方程组的求解问题。

穷举攻击的可能性

- 如果密钥空间是2^n，则平均需测试2^{n-1}次就可找到正确的密钥。
- 目前64位密钥的密码算法是不安全的，而128位密钥是比较安全的。

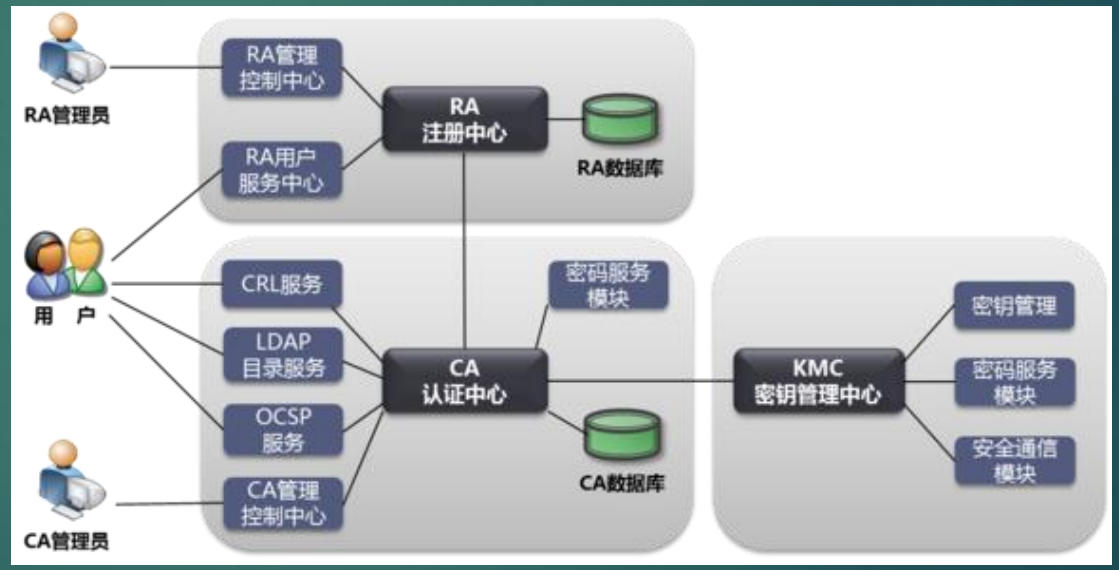
假设密钥空间为2^{64}，一台计算机每秒能测试的2^{30}个密钥，

- 则每年可测试$365*24*3600*2^{30}=2^{55}$个密钥
- 则检测完2^{64}个密钥需要512年。
- 如果用2048台机器同时做，需要三个月。

密码管理学

- 随机数生成理论与技术
- 密钥分配理论与方法
- 密钥分散管理技术
- 密钥分层管理技术
- 秘密共享技术
- 密钥托管技术
- 密钥销毁技术
- 密钥协议设计与分析技术

许多密码系统被攻破往往是密钥管理不当造成的。



1.3 密码编码基本原理

- 移位密码：在密钥的控制下，对一帧中的明文字符的位置进行移动。（不改变明文字符）

明文：wewil lmeet

加密：

密文字符位置	1	2	3	4	5
明文字符位置	2	5	4	1	3

密文：eliww mtele

解密：

明文字符位置	1	2	3	4	5
密文字符位置	4	1	5	3	2

不足：

- 明文字符形态不变，容易发现移位规律；
- 字符的统计规律没有变化，容易用字符频率统计法破译。

代替密码

- 在密钥的控制下，对明文逐字符进行替代。

单表替代：

“晨五点总攻”用区位码将其转换为四码一组的十进制数：
明文：1931 4669 2167 5560 1505

加密：

明文数字	0	1	2	3	4	5	6	7	8	9
密文数字	5	4	8	2	1	0	9	7	3	6

密文：4624 1996 8497 0095 4050

解密：

密文数字	0	1	2	3	4	5	6	7	8	9
明文数字	5	4	3	8	1	0	9	7	2	6

明文：1931 4669 2167 5560 1505

加法密码

- 加密：$c = E_k(m) = (m + k)\ mod\ q$
- 解密：$m = D_k(c) = (c - k)\ mod\ q$

选定固定的正整数q和k，
明文和密文空间：$Z_q = \{0,1, \dots, q-1\}$

凯撒密码：用加法密码构成的密码表对明文逐字替代。

先把明文字母转化为字母序号（0-25），再用加法密码进行加密，再将序号转为密文字母。

单表替换的缺点：

- 一份明文中的字符采用同一个替换表；
- 明文字符的统计规律没有被破坏，容易通过统计分析方法找到替换关系；
- 密钥与加密算法不分。知道了替代表，也就破译了替代密码。

可以通过多表替换打乱统计规律。

- 将明文分成多个片段，每个片段采用不同的替换表；
- 替换表的产生、使用由密钥控制；
- 在密钥保密且使用正确的条件下，替换表的公开与否与多表代替密码的安全没有影响。

多表替换

- 根据密钥的指示，来选择加密时使用的代替表。

例如： 加密变换 $c = E_k(m) = (m + k)\ mod\ 10$

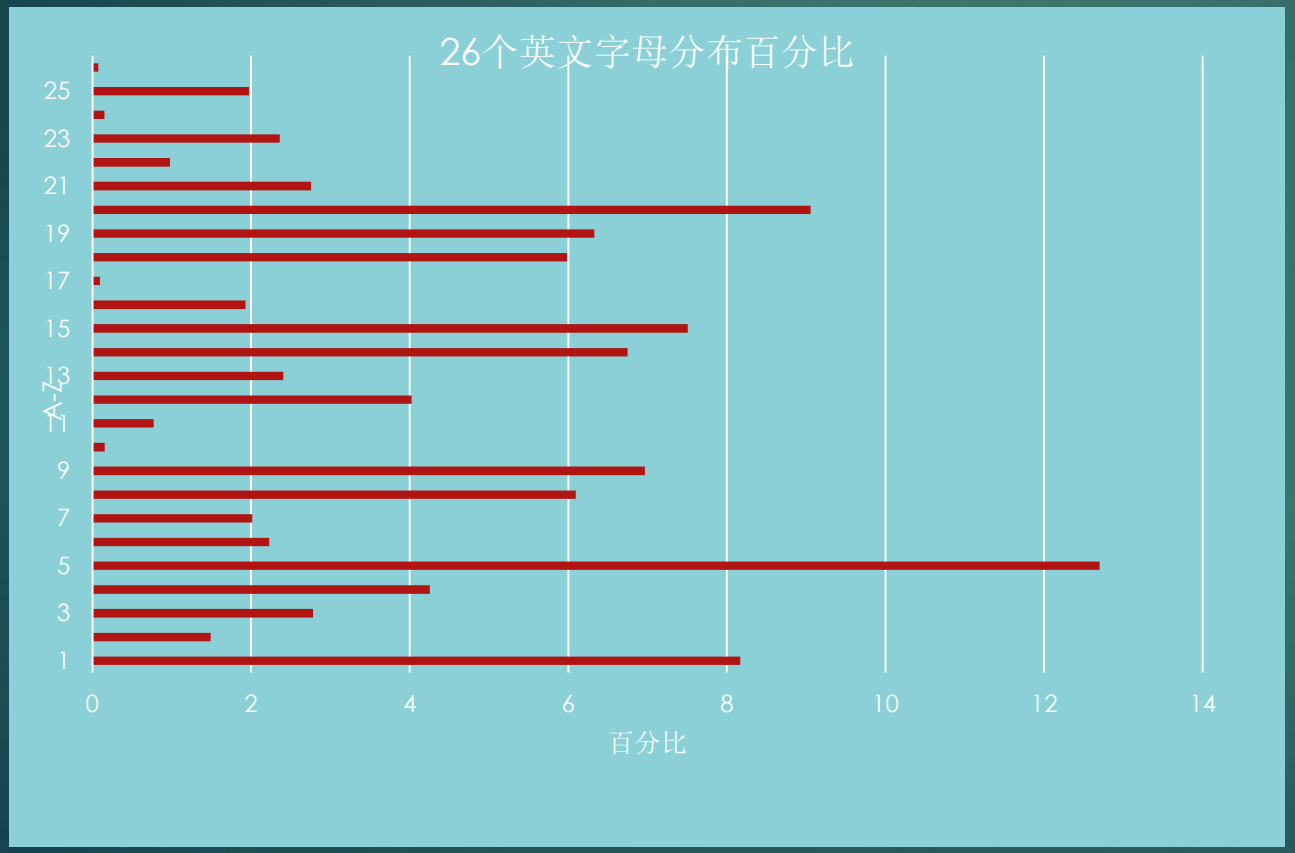
明　　文：　晨　　五　　点　　总　　攻
对应序列：1931 4669 2167 5560 1505
密钥序列：4321 5378 4322 3109 1107
密文序列：5252 9937 6489 8669 2602

明文的统计规律不直接反映在密文中。如果密钥序列是随机的，则密文序列也是随机的。此时达到“一次一密”完全保密。

但如果密钥序列的具有周期性T，则可以通过对周期T的穷举，把多表替换的破译问题变为单表替换的破译问题。

语音的内在规律

- 在一篇非专业的英文文献中，26个英文字母出现的概率为：



极高概率字母：*e*
次高概率字母：*t,a,o,I,n,s,h,r*
中等概率字母: *d,l*
低概率字母：*c,u,m,w,f,g,y,p,b*
极低概率字母：*v,k,j,x,q,z*

单表替换密码破译方法

- 单表代替是指以m个字符为一个分组进行替代的密码。

- 对密文字母的出现次数进行统计，得到密文字母频次表；可据此推断是否为单字母替换；
- 把密文字母（或几个连续字母）频次与明文字母（单词）频次进行对照，推断明文字母或单词；
- 利用文字、语义规律进一步推断明文；
- 利用模式字或模式短语进行猜字。

- 密文越长，统计越符合规律，破译越准确。
- 密文较短时，需要猜测的字母越多，可以先试译，再填空。这样可大大加快破译速度。

多表替代及其密码分析

- 利用多张替代表进行字符的替换。根据密钥序列的指示，逐字符对明文进行替代。
- 多表替换的目的是使得密文序列中各字符分布更均匀。
- 通常将明文分成若干个段，每个段用一个替代表。究竟用那个替代表，由密钥序列控制。
- 为减少密钥数量，通常将有限长的密钥字符序列周期地重复使用。这就给破译提供了机会。

粗造度—密文序列中各字符分布的均匀程度

- 设$p_c(\alpha)$是密文字符α的出现概率，则用统计量χ表示密文序列的粗造度：

$$\chi = \sum_{\alpha}\left[p_c(\alpha) - \frac{1}{26}\right]^2$$

即一个英文的密文文章中，各个字母的分布概率平均偏离1/26的程度。

如果一段密文中，各个字母分布均匀，则粗造度为0.
对于单表替换，由于密文中各字母分布概率与明文相同，则可计算出单表替换密文的粗造度为0.027.

重合指数—刻画密文序列中两个相同字符的重合程度

- 设密文字符α的出现频次为f_α,则从N个密文字符中随机选取两个字符相同的概率：

$$IC = \frac{1}{N（\mathrm{N}-1)} \sum_{\alpha=\mathrm{A}}^{\mathrm{Z}} f_\alpha(f_\alpha - 1)$$

称为多表代替密码的重合指数。

- 当每个密文字母出现的频次都相等时，重合指数达到最小值1/26≈0.0385.
- 采用d张表进行替换加密时，可以证明，d与重合指数IC关系为：

$$d \approx \frac{0.027N}{(N-1)IC - 0.0385N + 0.0655}$$

替代表数d	1	2	3	4	5
重合指数IC	0.0655	0.052	0.043	0.04	0.0385

- 根据重合指数，可以推断所用的替代表数。
- 当替代表数量大于等于5时，就无法利用重合指数检测密文字母的统计规律。

多表替代中密钥长度的确定

- 多表替代中密钥的长度即替代表的长度；
- 关键是求出密钥序列的周期t；
- 求出t后就转变为破译单表替换的问题。

对于长度为N的密文序列，

1. 移位法：右移序列d位，统计移位前后两个序列的重码数。如果d是t的整数倍，则重码数接近（N-d）*0.0655.
2. 普鲁士军官Kasiski提出的重码分析法：当且仅当两个字符的间隔是周期的倍数时，他们才是同一表的加密结果。利用高频的密文串进行分析。
3. 密表匹配技术：如果两份密文所使用的代替表相同，它们就具有相近的单码频次分布规律。

保密理论

- 随机事件的信息量和概率分布的熵:

  - 随机事件x_i的信息量定义为：

$$I(x_i) = -\log p(x_i)$$

设某密码算法有128比特的密钥，每个密钥的出现概率为2^{-128}.
密钥的信息量可以计算出来为128.
也就是说，一旦破译该密钥，就获得了128比特的信息。

  - 设$X=\{x_1,x_2,\ldots,x_n\}$, x_i出现的概率为$p(x_i)$, 则事件x_i出现时提供的平均信息量为：

$$H（X） = \sum_{i=1}^{n} p(x_i)I(x_i) = -\sum_{i=1}^{n} p(x_i)\log p(x_i)$$

定义为概率分布p的熵。

计算密钥熵的例子:

例：设密钥k为256比特，且某个破译算法共输出密钥k的4个可能值，假设这4个可能值是正确密钥值的概率分别为2^{-1}、2^{-2}、2^{-3}、2^{-4}，且其它可能值是正确密钥的概率都相等。求经过该破译算法后密钥的熵。

解：先求出扣除已有4个可能密钥值的其它密钥的总概率

$$1-\sum_{i=1}^{4}p(k=k_i)=1-\sum_{i=1}^{4}2^{-i}=2^{-4}$$

在剩余空间中，每个密钥概率均等，则任一密钥的概率为

$$p(k=\alpha)=\left(1-\sum_{i=1}^{4}p(k=k_i)\right)\Big/(2^{256}-4)=2^{-260}$$

则破译该算法后密钥的熵为

$$H(K)=[\text{已有可能密钥的熵}+\text{剩余空间密钥的熵}]$$

$$=-\sum_{i=1}^{4}p(k=k_i)\,log_2p(k=k_i)-\sum_{\alpha}p(k=\alpha)log_2p(k=\alpha)$$

=17.875

Shannon保密理论

- 一个密码体制有五个部分组成：明文空间$\boldsymbol{M}$、密文空间$\boldsymbol{C}$、密钥空间$\boldsymbol{K}$、加密算法$\boldsymbol{E}(k,m)$和解密算法$\boldsymbol{D}(k,c)$。

假定：明文、密文和密钥是各自空间上的一个随机变量，且密钥与明文相互独立。

定义：完全保密或理论保密—$\boldsymbol{I}(\boldsymbol{M};\boldsymbol{C})=0$ (明文与密文的互信息为0)

5个等价条件：

- 一个密码体制是完全保密的；
- $\boldsymbol{I(M;C)=0;}$
- $\boldsymbol{H(M|C)=H(M);}$
- $\boldsymbol{H(C|M)=H(C);}$
- 明文与密文独立。

对于完全保密的密码体制，有$H(K)\geq H(M)$.

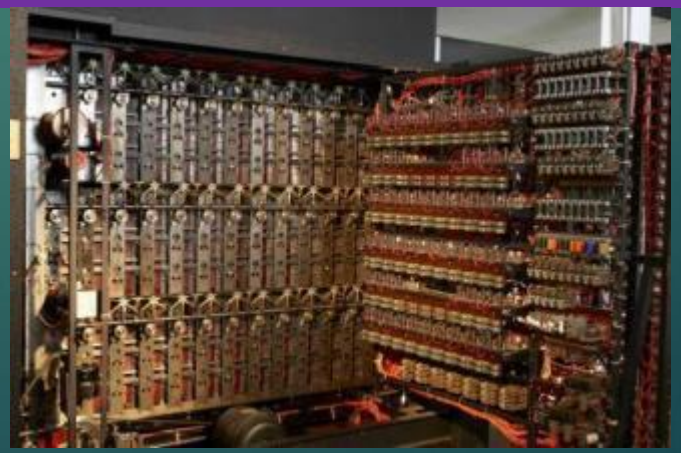
即如果要加密L比特的明文，至少需要L比特的密钥加密该明文才有可能达到完全保密。

计算复杂性

- 与信息论方法不同，计算复杂性是研究密码保密性的另一种方法。研究一个问题是否具有可行的求解方法，以及所需的计算时间和硬件资源。

- 如果破译一个密码体制所需的代价，超过了破译者的能力（时间、资源等），则这个密码体制实际是保密的。
- 现代密码体制主要考察实际保密性。理论上不保密的，实际上可能保密。
- 而理论上保密的，在实际上因为密钥管理等原因，可能是脆弱的。
- 密钥的传输、存储、销毁等过程可能不能保证密钥的秘密性，密钥的复用也导致不保密。

分析密码的实际保密性主要考察：
- 密码分析者的计算能力；
- 密码分析算法的有效性。



因式分解问题

- 要求对给定的合数n进行因式分解，即求出素数p_1、p_2、…、p_m和自然数e_1、e_2、…、e_m，使得

$$n = p_1^{e_1} p_2^{e_2} \dots p_m^{e_m}$$

二元域上的多元二次方程组的求解问题

- 求出满足

$$\begin{cases} f_1(x_1,\cdots,x_n)=b_1 \\ f_2(x_1,\cdots,x_n)=b_2 \\ \\ \\ \\ f_m(x_1,\cdots,x_n)=b_m \end{cases}$$

的一个n维二元向量（x_1,x_2,…,x_n）,其中

$$f_k(x_1,x_2,\cdots,x_n)=\sum_{i=1}^{n}\sum_{j=1}^{n}a_{k,i,j}x_i\,x_j$$

背包问题

- 给定n个整数$a_1,a_2,\ldots,a_n$和整数S, 求n维二元向量（$x_1,x_2,\ldots,x_n$）， 满足

$$a_1x_1+a_2x_2+\ldots+a_nx_n=S$$

这里（$a_1,a_2,\ldots,a_n$）称为背包向量。

有n种物品和一个容量为S的背包，每种物品都有无限件可用。第i种物品的体积是a_i，价值是w_i。将哪些物品x_i装入背包可使这些物品的体积总和不超过背包容量S，且价值总和最大。

离散对数问题

- 给定一个质数p和有限域Zp上的一个本原元a，对Zp上整数b，寻找唯一的整数c，使得

$$a^c \equiv b (\mathrm{mod}\ p)$$

- 一般的，如果仔细选择p，则认为该问题是难解的，且目前还没有找到计算离散对数问题的多项式时间算法。
- 为了抵抗已知的攻击，p至少应该是150位的十进制整数，且p-1至少有一个大的素数因子。

算法的运行时间

- 时间复杂度考察的是随着输入规模n的增大，算法的运行时间增加的规律。

定义：
设$f(n)$和$g(n)$是两个正的实数值函数，则

- 如果存在正常数c和正整数N, 使得$n \geq N$时，都有$0 \leq \frac{f(n)}{g(n)} \leq c$, 则称$f(n) = O(g(n))$

即$f(n)$与$g(n)$是在忽略常数倍的前提下，$f(n)$比$g(n)$渐近增长得慢。

- 如果存在正常数c_1、c_2和正整数N, 使得$n \geq N$时，都有$c_1 \leq \frac{f(n)}{g(n)} \leq c_2$, 则称$f(n) = \Theta(g(n))$

即$f(n)$与$g(n)$是在忽略常数倍的前提下，$f(n)$比$g(n)$渐近增长一样快。

$n^3 = O(3n^3)$　　$n^3 = \Theta(3n^3)$

多项式时间

- 设n是输入的规模，一个多项式时间的算法是指存在正整数k，使得时间复杂度为$O(n^k)$的算法。

- 多项式时间的算法意味着有效的或好的算法。
- 在密码学中，通常将平均时间复杂性是多项式函数的算法称为多项式时间算法。

例：对于前面的背包问题，如果背包向量（$a_1,a_2,\ldots,a_n$）为（$1,a,a^2,\ldots,a^n$）,则称为超递增背包问题。
可以验证，求解该问题的复杂性为$\Theta(n)$.

求解n元线性方程组问题的高斯消元法的时间复杂性为$O(n^3)$;
n个数的排序算法时间复杂性为$O(n^2)$.

问题的复杂性

- 一个密码的破译困难程度有时可以归结为一个典型问题。

- 对于一个问题，如果存在一个多项式时间的求解算法，则称该问题为P问题。（如前述的几个问题）
- 如果存在一个多项式时间的算法，对问题的解的每个猜测，都能判断解是否正确，则称该问题为NP问题。
- P问题就是实际可求解的问题；NP问题就是对其解的猜测可进行验证的问题。
- 通常P问题是NP问题的一个子问题。

- 背包问题是NP问题，但递增背包是一个P问题；
- 因式分解问题是NP问题，但尚未证明其是一个P问题。

密码学中的求解问题一般为NP问题。

密码算法的安全性依赖一些困难问题的难解性。求解密钥的问题应是难解问题，不能是P问题，应是NP问题。

习题

1、计算采用单表替换的密文序列的粗造度。

2、证明：对于完全保密的密码体制，有$H(K) \geq H(M)$。

3、证明：如果$f(n)=\Theta(g(n))$, 则$f(n)=O(g(n))$。

流密码（序列密码）

- 前述理论上的保密体制是存在的；
- 用随机的密钥序列对明文序列加密得到密文序列，且密钥序列与明文序列等长。这在实际上难以做到。
- 用少量的真随机数产生伪随机序列代替真正的随机序列，这就是序列密码。
- 这种“少量的真随机数”也称之为“种子密钥”。
- 如何刻画密钥序列的“伪随机性”？

游程

- 一条二元序列100...01的片段，称为该序列的一个0游程；0的个数为0游程的长度；
- 一条二元序列011...10的片段，称为该序列的一个1游程；1的个数为1游程的长度；
- 以0开头的信号片段0...01，为一个0游程；以1开头的信号片段11...10为一个1游程；
- 以0结束的信号片段10...0，为一个0游程；以1结束的信号片段011...1为一个1游程；

0001是一个长度为3的0游程；011是长度为2的1游程。

互相关系数和自相关系数

- 设$\{a_i\}_{i=0}^{\infty}$和$\{b_i\}_{i=0}^{\infty}$是两条周期为p的二元序列，则称：

$$C_{a,b} = \frac{1}{p}\sum_{i=0}^{p-1}(-1)^{a_i \oplus b_i}$$

为这两个序列的互相关系数。反映了这两个序列对应比特的相等程度。

类似的，称：

$$C_a\tau) = \frac{1}{p}\sum_{i=0}^{p-1}(-1)^{a_i \oplus a_{i+\tau}}$$

为序列$\{a_i\}_{i=0}^{\infty}$的自相关系数，反映了这个序列移位τ后对应比特的相等程度。

自相关函数的一般定义为：$R_a\tau) = \frac{1}{p}\sum_{i=0}^{p-1} a_i a_{i+\tau}$　在二进制序列情况下，与上面等价。

伪随机性的Golomb三假设

- 平衡假设

  序列的一个周期中“0”的个数和“1”的个数差值不超过1.

- 游程假设

  长度为m的游程个数占游程总数的1/2m；“0”游程的个数和“1”游程的个数相等。

- 自相关假设

  自相关函数在非零点τ的值是一个固定的常数。

m序列满足Golomb随机性假设，但基于m序列的加密算法却是可破的。

随机序列的其它密码学指标

- 完全性

每个随机数应是所有密钥比特的足够复杂的函数，否则会导致对该密码算法的分割攻击；

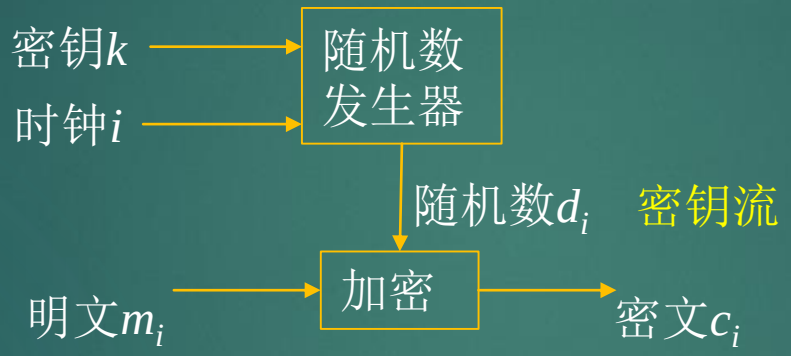
- 周期性

要求随机序列的周期对所有密钥都足够大；

- 线性复杂度

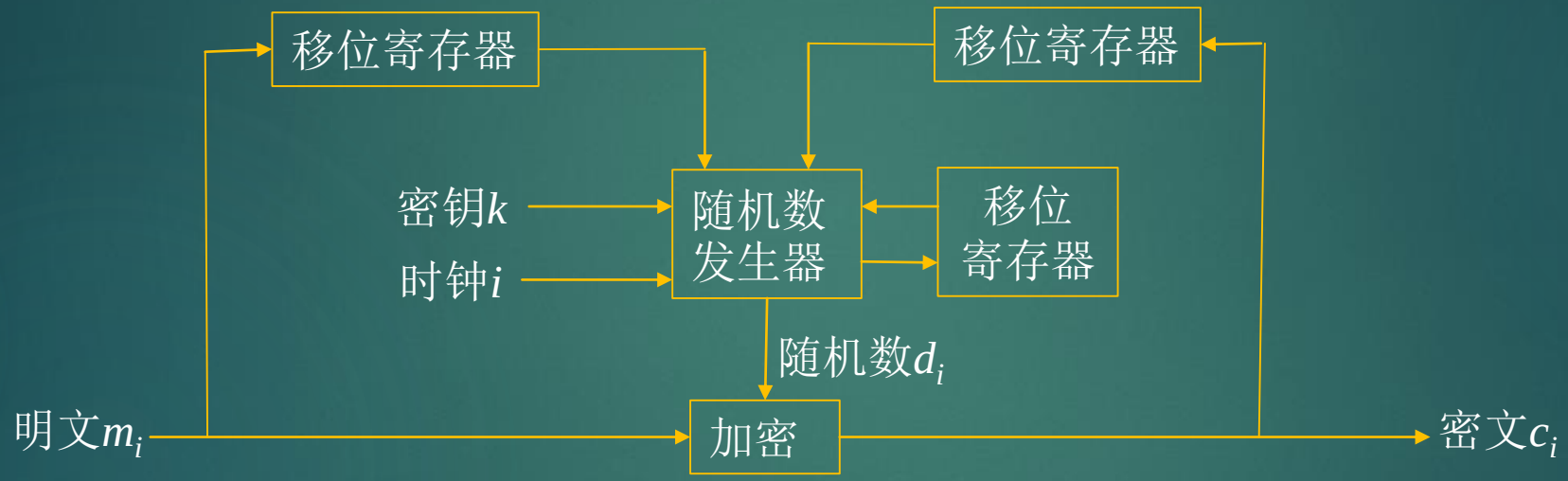
要求随机序列的线性复杂度对所有密钥都足够大。

序列密码一般模型



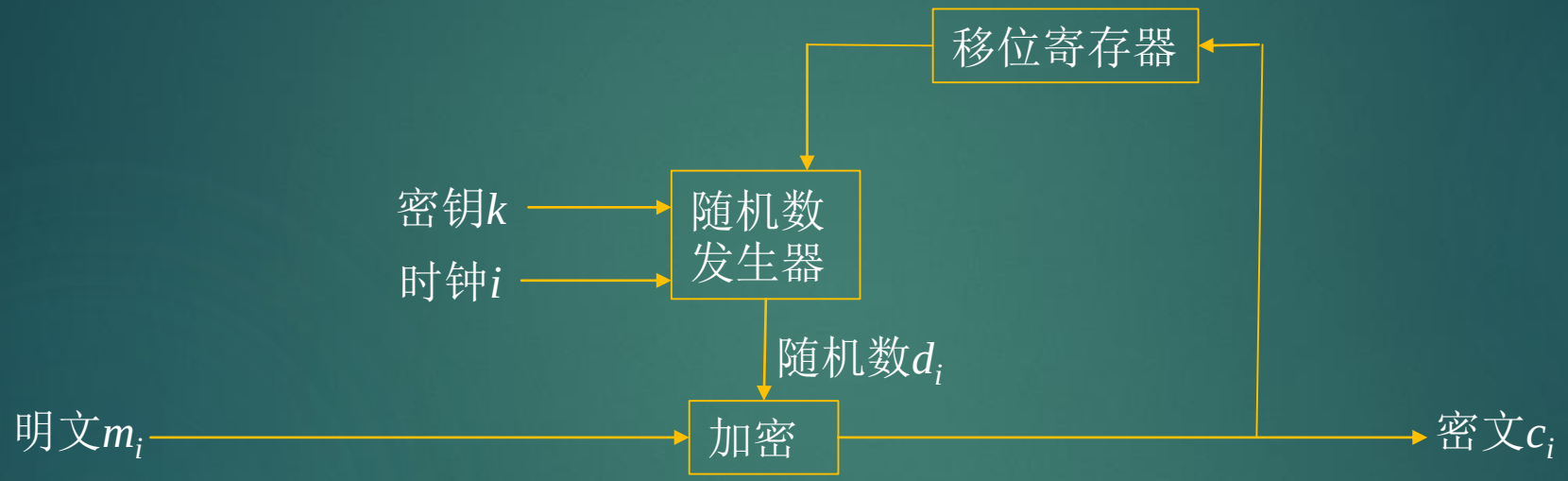
一般明文与密钥流按比特异或。

明密文反馈记忆的序列密码模型



优点：减轻密钥重用带来的安全隐患
缺点：接收端序列一旦出现错码或丢码，就无法解密。

自同步流密码模型



解密端丢失一段数据后，过一段时间，就可以恢复正确解密。

伪随机序列的产生----线性递归序列

- 函数f: [GF(q)]^n→ GF(q), $\forall a_i \geq 0$,都有$a_i \in GF(q)$,且$\forall m \geq n$,有

$$a_m = f(a_{m-1}, a_{m-2}, \cdots, a_{m-n})$$

则$\{a_i\}$为n级递归序列，f为递归函数。

如

$$a_m = c_1 a_{m-1} + c_2 a_{m-2} + \cdots + c_n a_{m-n}$$

$\{a_i\}$为n级线性递归序列。

如 $f(x_1,x_2,\ldots,x_5)=x_2 \oplus x_5$, 则$a_m=a_{m-2} \oplus a_{m-5}$, $m \geq 5$
以10011开头的递归序列为：1001101001000010101110110001111100110…

RC4序列密码算法

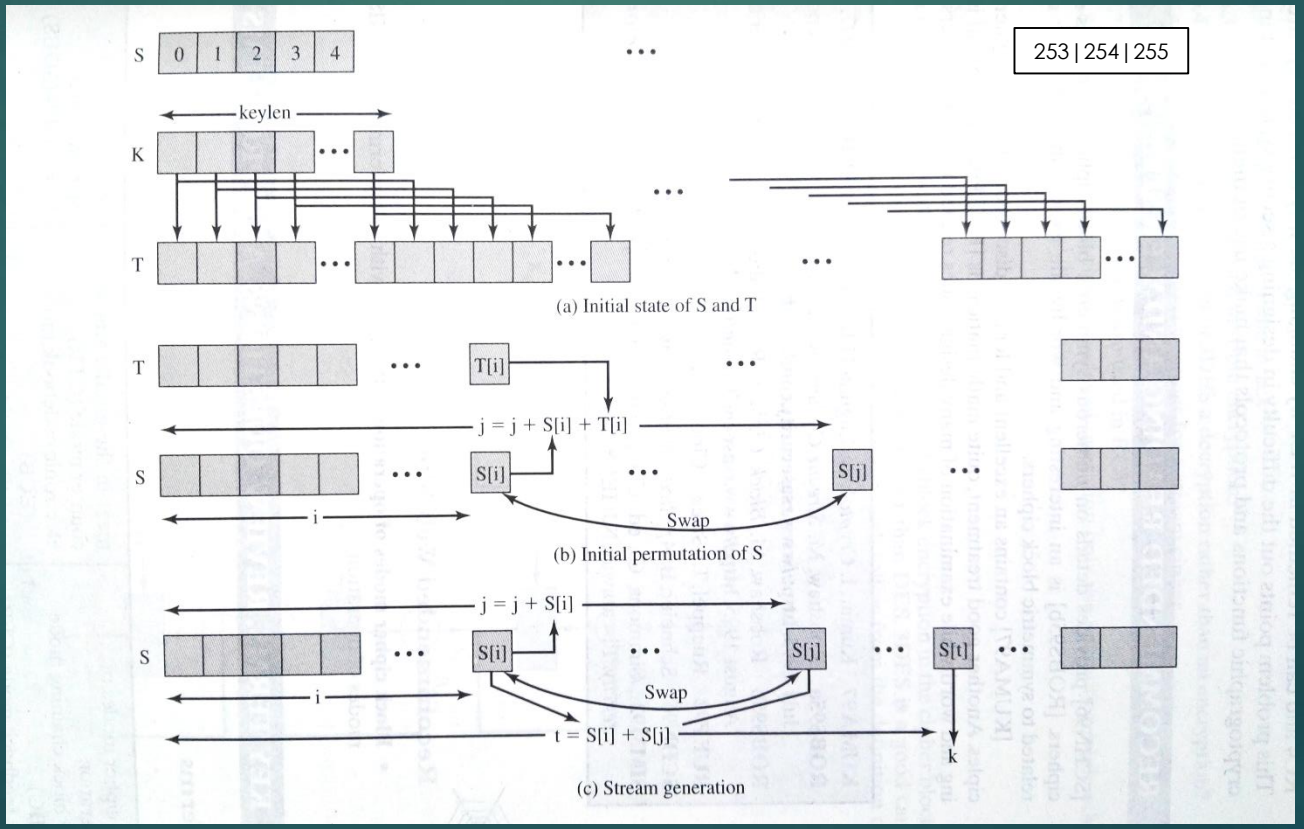
- Ron Rivest 在1987年为RSA公司开发的密钥长度可变的序列密码算法；
- 广泛应用于商业密码中，如Wi-Fi WEP安全协议；

S寄存器组初装0-255

T寄存器组装密钥

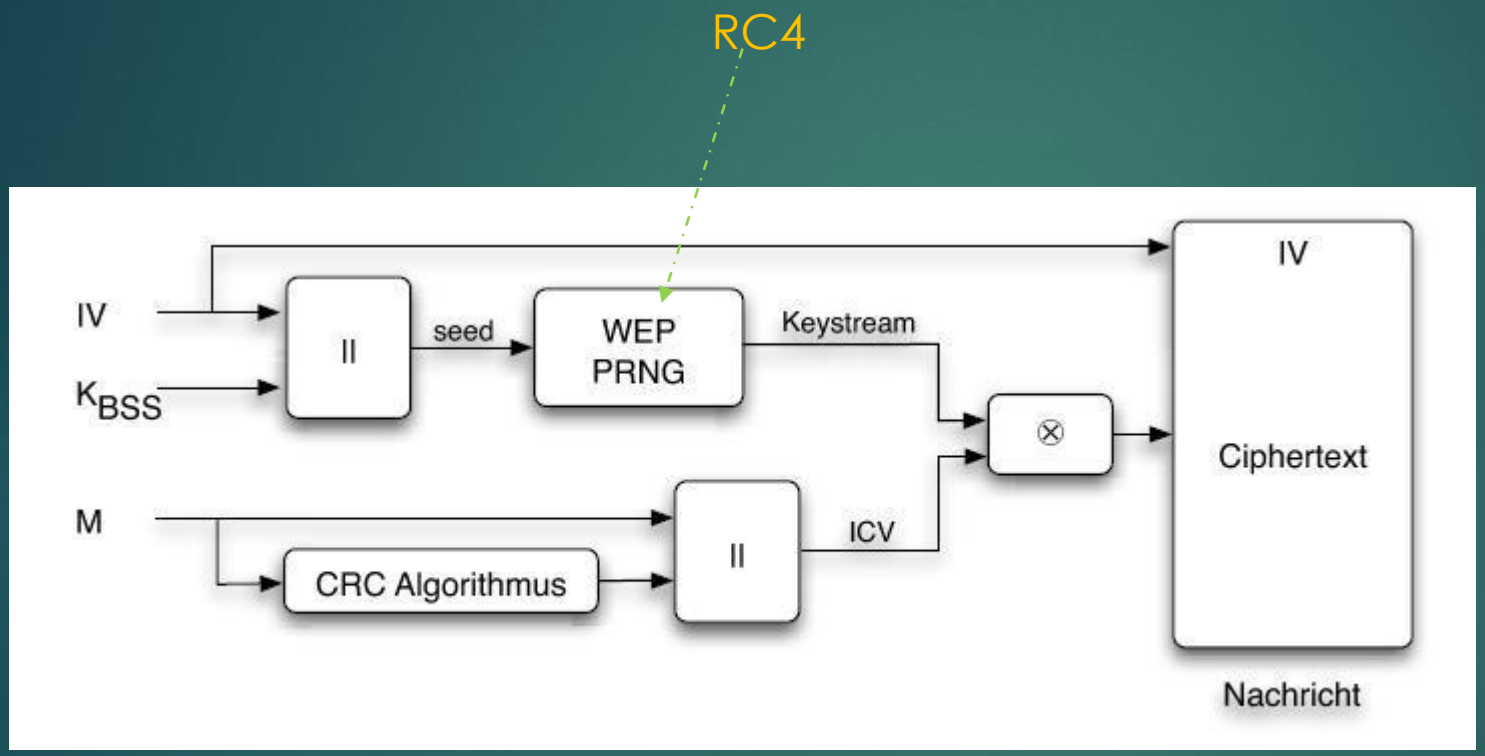
S寄存器组交换位置

S寄存器组再次交换位置



输出0-255的随机数列k

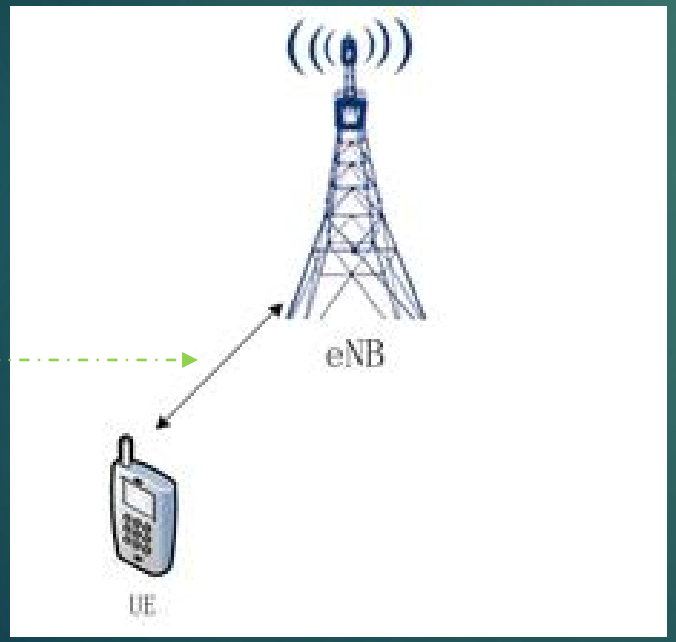
RC4算法加密过程



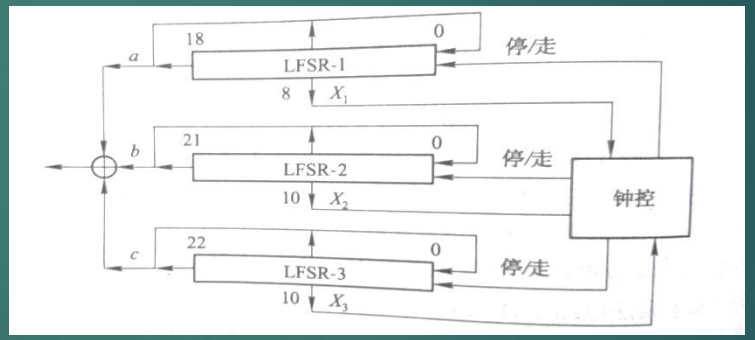
RC4算法目前可彻底破解，寻找S盒的一个较弱的初始设置。
结合了多次加密时IV的公开性进行的。

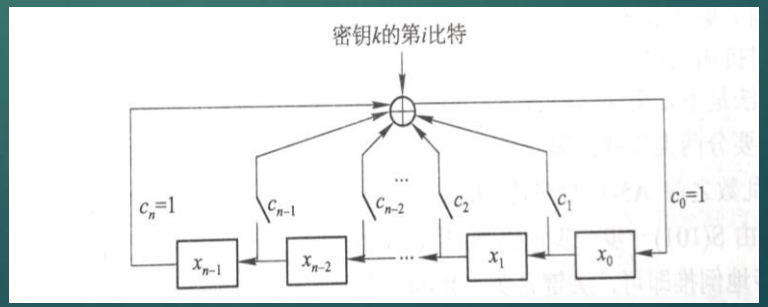
A5算法

- 用于GSM加密的序列密码算法；
- 用于手机到基站之间的通信加密；
- 用三组LFSR（线性反馈移位寄存器）序列合成伪随机序列。



设计简单，采用分割攻击的方法可破解。

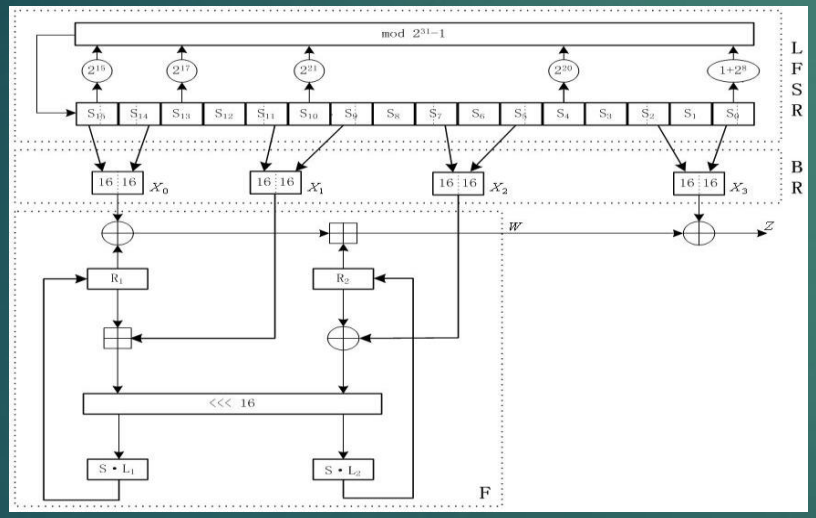


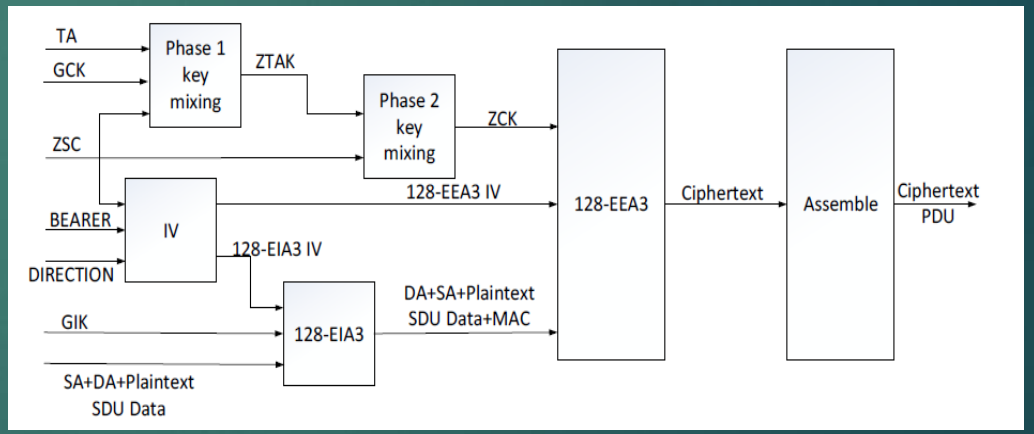


密钥参与控制移位寄存器

ZUC序列密码算法

- ZUC算法是中国自主设计的流密码算法，现已被3GPP LTE采纳为国际加密标准，即第四代移动通信加密标准。





三层结构：

- LFSR层
- 比特重组层
- 非线性迭代层

很高的安全性。

序列密码算法总结

- 生成随机性好、密码学特征好的伪随机序列；
- 设计好的加密方案；
- 密钥的强度。

习题

- 已知3级移位寄存器的线性递推式$a_n=a_{n-2}+a_{n-3}, n\geq 3$,

（1）试给出以101开头的输出序列及周期；

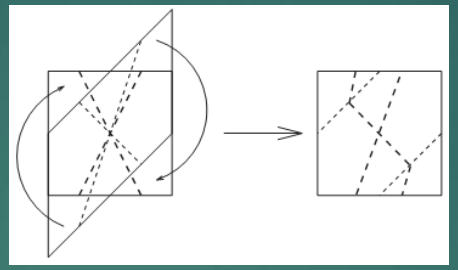
（2）分析该序列的密码学特性。

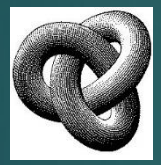
乘积密码

- 由于语言的统计特性使得使用替换或置换进行加密并不安全
- 可以交替的使用多种加密方式:
  - 两种替换得到更复杂的替换结果
  - 两种置换得到更复杂的置换结果
  - 替换后再进行置换得到的结果相对更复杂!
- 这种思想是从经典密码到现代密码体制的桥梁!

图形多轮变换例子----乘积密码的思想

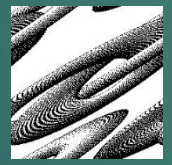
操作方式：



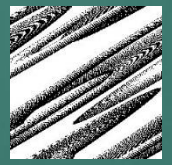


第 1 轮

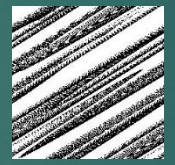
第 2 轮



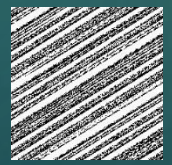
第 3 轮



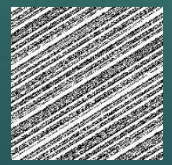
第 4 轮



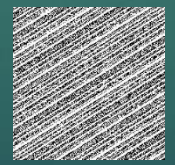
第 5 轮



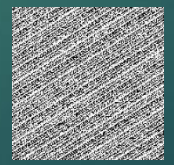
第 6 轮



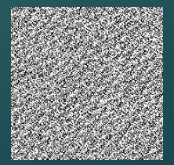
第 7 轮



第 8 轮

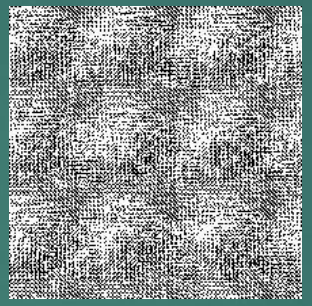


第 9 轮

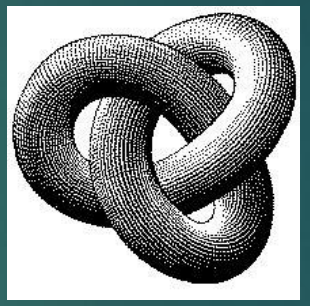


第 1 6 轮

第24轮

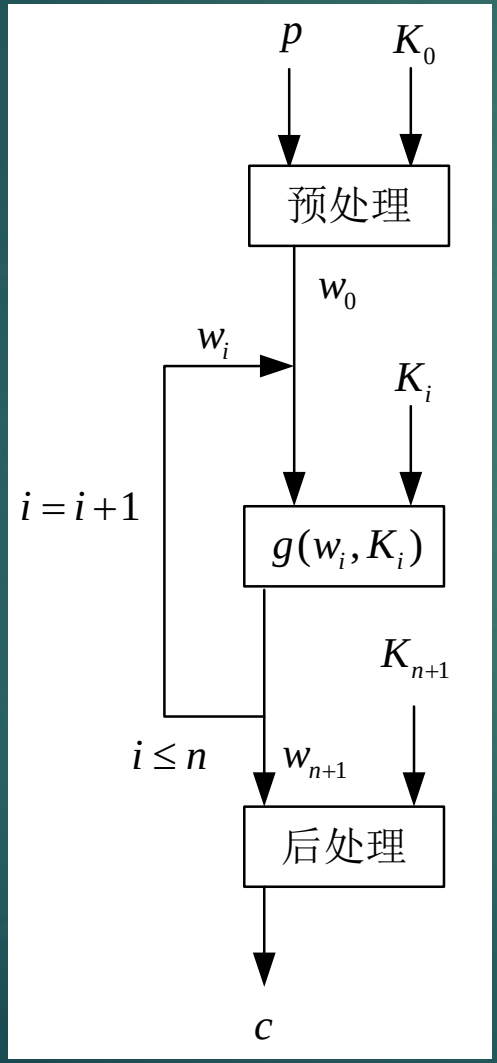


第48轮



第72轮

乘积密码的一般结构



$$e_k = e_{(k1,k2)}(m) = e_{k2}(e_{k1}(m))$$

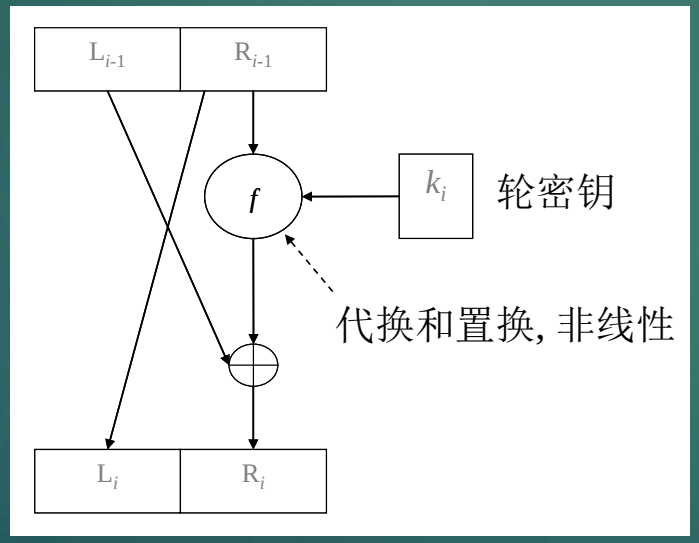
$$d_k = d_{(k1,k2)}(c) = d_{k1}(d_{k2}(c))$$

$$d_k(e_k(m)) = d_{(k1,k2)}(e_{(k1,k2)}(m)) = d_{k1}(d_{k2}(e_{k2}(e_{k1}(m)))) = d_{k1}(e_{k1}(m)) = m$$

- 多次重复执行一个简单的密码函数，密码函数每被执行一次称为一轮(round)
- 密码函数称之为轮函数g，它使用的密钥称为轮密钥
- 轮密钥通常从主密钥K导出，每轮都不同，避免出现迭代返回现象。

典型的轮函数

- Feistel网络：著名的DES算法采用该轮函数。



正函数：

$$L_i = R_{i-1}$$
$$R_i = L_{i-1} \oplus f(R_{i-1}, K_i)$$

逆函数：

$$R_{i-1} = L_i$$
$$L_{i-1} = R_i \oplus f(R_{i-1}, K_i)$$

使得加密和解密可以使用相同的结构

代换-置换网络（SPN）

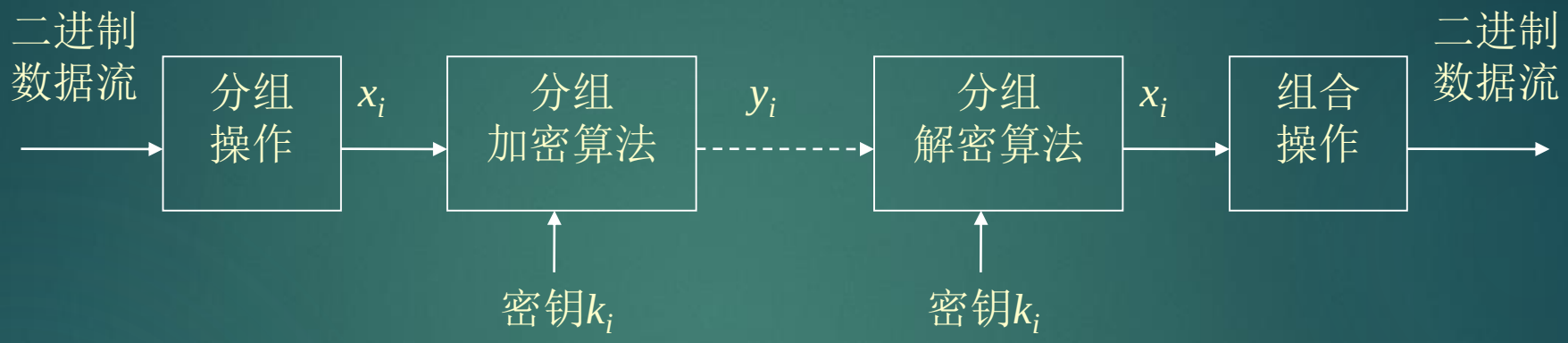
- SPN轮函数执行：代换、置换和密钥混合。
- S盒和P盒的操作通常是固定的，和轮密钥无关。在每一轮中，轮密钥的操作通常采用简单的位异或方式和输入值进行绑定。
- 一个设计良好的SPN应满足的两个属性：***扩散***和***混淆***。
- **扩散**使明文的每个比特与密钥的每个比特对密文的每个比特都产生影响。
- **混淆**使明文和密文之间的统计关系变得尽可能复杂，使用复杂的非线形代换算法可得预期的混淆效果。

代换与置换的作用--扩散与混淆

明文　　密钥

密文

分组密码加/解密过程



经过5轮迭代后，密文的每一位基本上是所有明文和密钥位的函数，而经过8轮迭代后，密文基本上是所有明文和密钥位的随机函数。更多轮的迭代可以防止差分密码分析.

为什么需要分组密码?

- 安全强度较流密码高;
- 运算量适中,适合实时加密运用.
- 适合芯片高速实现.

分组密码的轮迭代

明文

轮函数

$$w^0 \leftarrow x$$
$$w^1 \leftarrow g(w^0, k^1)$$
$$w^2 \leftarrow g(w^1, k^2)$$
$$\vdots$$
$$w^{N_r} \leftarrow g(w^{N_r-1}, k^{N_r})$$
$$y \leftarrow w^{N_r}$$

轮密钥

密文

加密

解密

$$w^{N_r} \leftarrow y$$
$$w^{N_r-1} \leftarrow g^{-1}(w^{N_r}, k^{N_r})$$
$$\vdots$$
$$w^1 \leftarrow g^{-1}(w^2, k^2)$$
$$w^0 \leftarrow g^{-1}(w^1, k^1)$$
$$y \leftarrow w^0$$

几种典型的分组密码

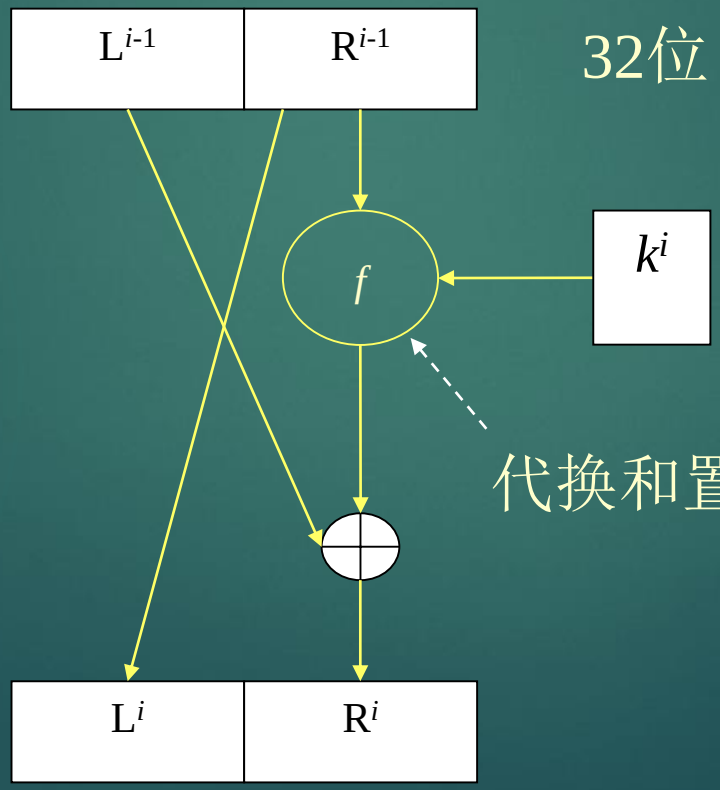
- DES（由IBM提出,分组长度64,密钥长度56位）
- AES（Rijndael由比利时Daemen和Rijmen提出,分组长度128,密钥长度128\192\256）
- SMS4（国产算法是用于WAPI的分组密码算法，是国内官方公布的第一个商用密码算法）

DES

- 第i轮函数

$$g\left(L^{i-1},R^{i-1},k^{i}\right)=\left(L^{i},R^{i}\right)$$

32位



32位

轮密钥48位

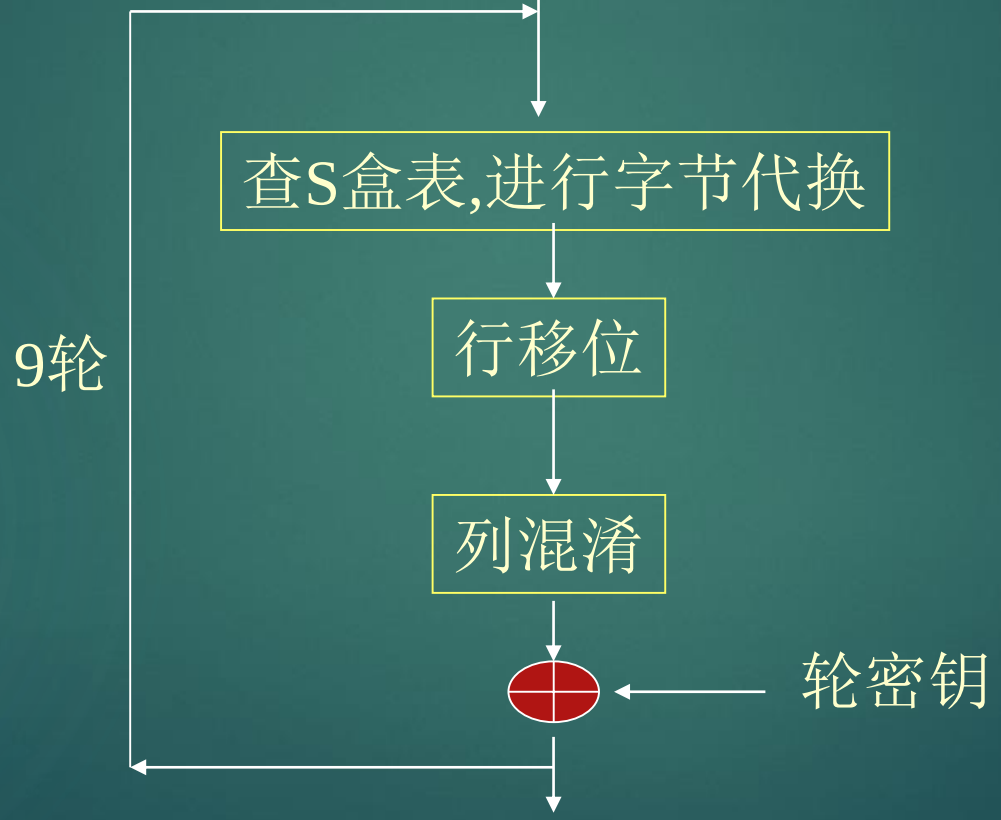
共16轮迭代

代换和置换, S盒是非线性部件

AES

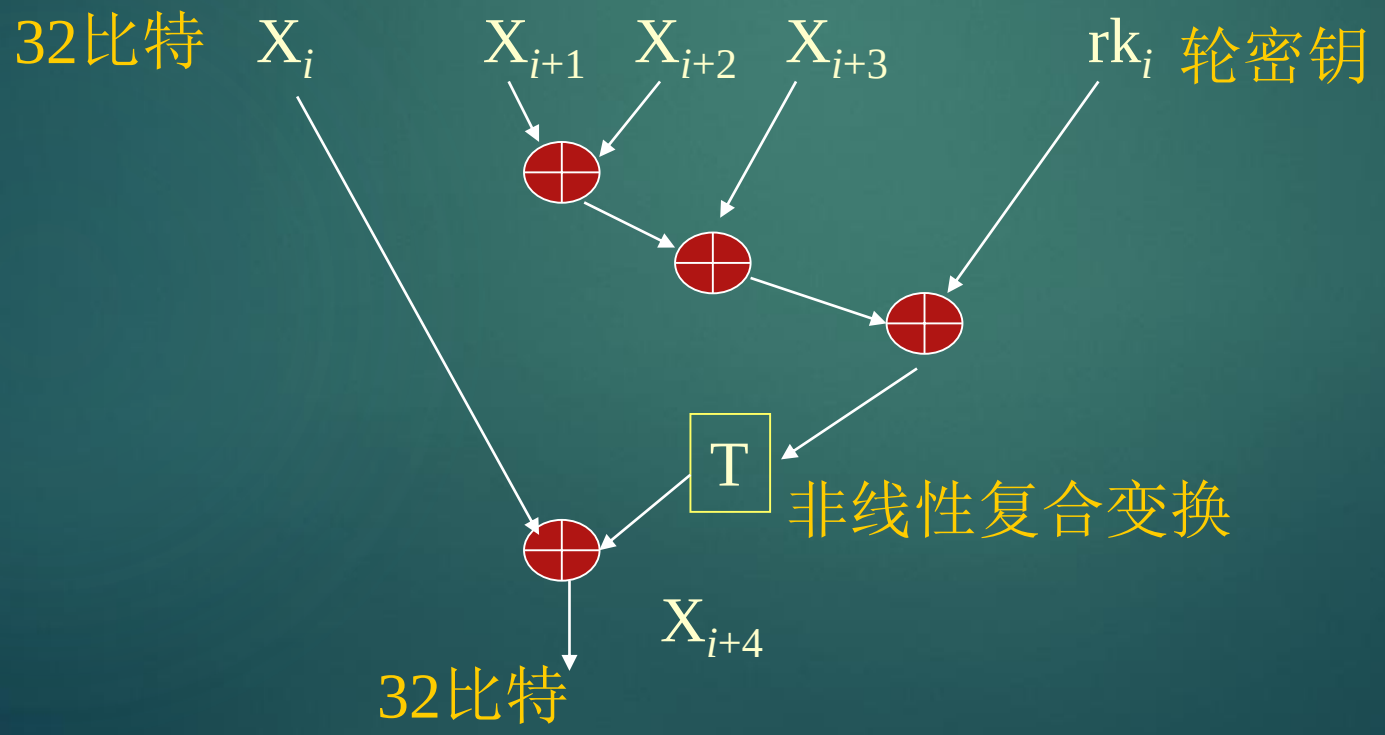
128=16*8=(4*4)*8,以字节为单位

轮运算:



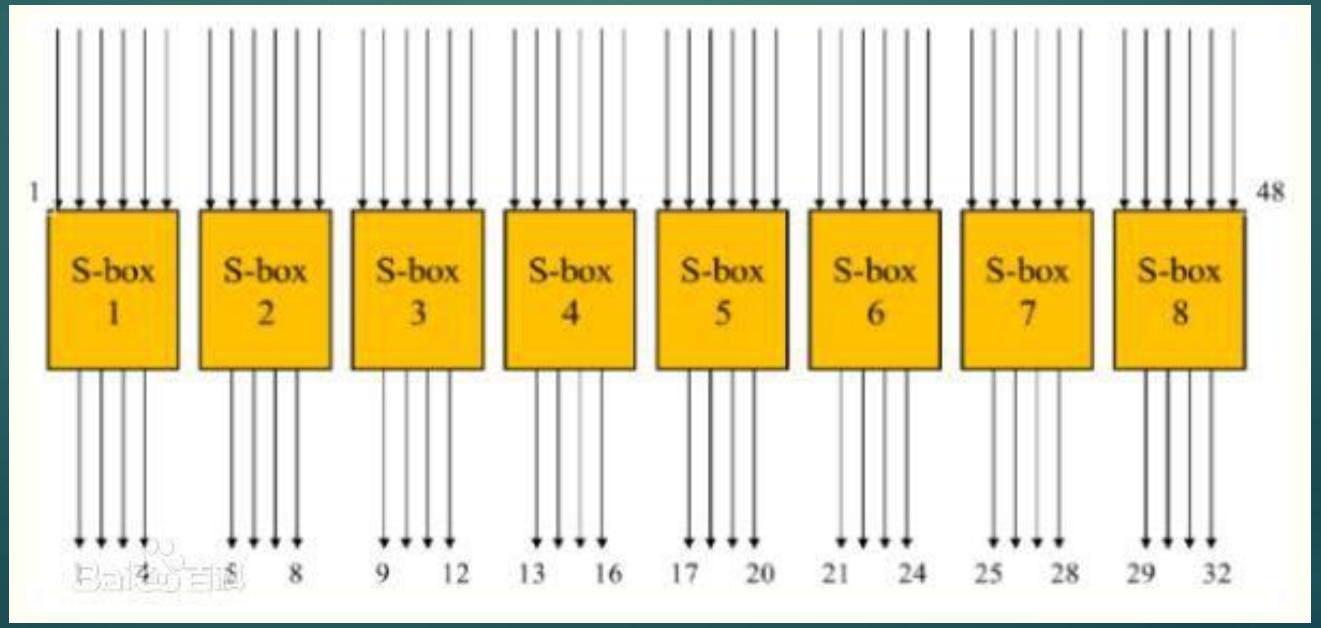
SMS4

$$X_{i+4}=F(X_i,X_{i+1},X_{i+2},X_{i+3},rk_i)=$$
$$X_i\oplus T(X_{i+1}\oplus X_{i+2}\oplus X_{i+3}\oplus rk_i),$$
$$i=0,1,\cdots,31.$$



S盒的作用

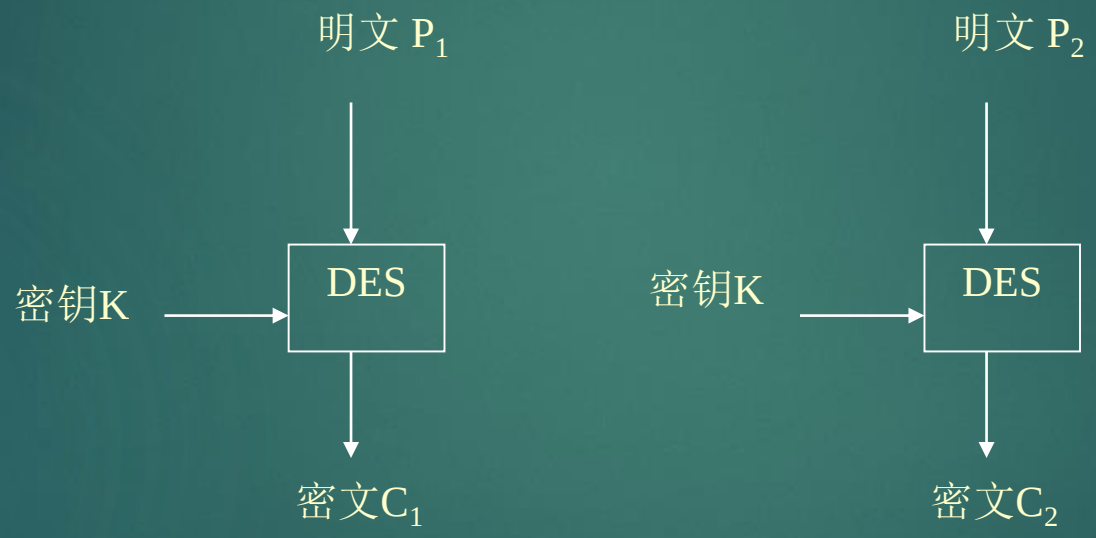
- S盒(Substitution-box)是对称密钥算法 执行置换计算的基本结构。
- S盒用在分组密码算法中，是唯一的非线性结构，其S盒的指标的好坏直接决定了密码算法的好坏；
- S盒的功能就是一种简单的“代替”操作。S盒是将48比特压缩成32比特输出；
- 替代由8个不同的S盒完成，每个S盒有6位输入4位输出。48位输入分为8个6位的分组，一个分组对应一个S盒，对应的S盒对各组进行代替操作。



密码工作模式

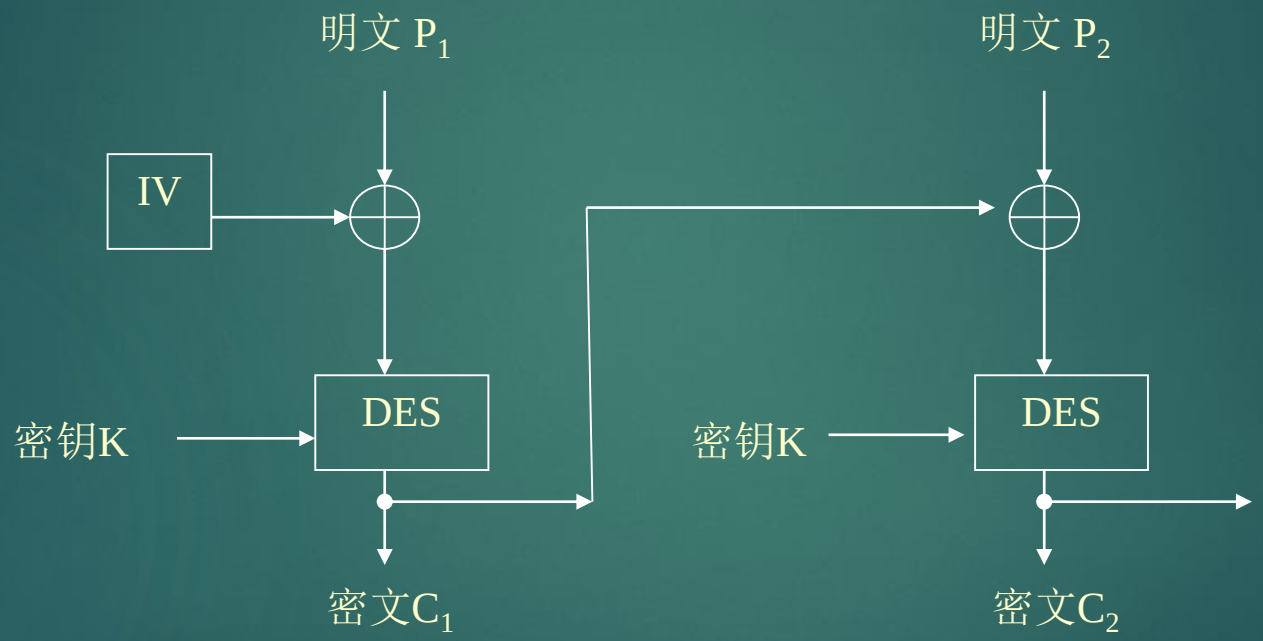
- 电码本模式（ECB, Electronic Codebook）
- 密码反馈模式（CFB, Cipher Feedbacb ）
- 密码分组链接模式（CBC, Cipher Block Chaining ）
- 输出反馈模式（OFB, Output Feedback）

ECB：电子码本模式



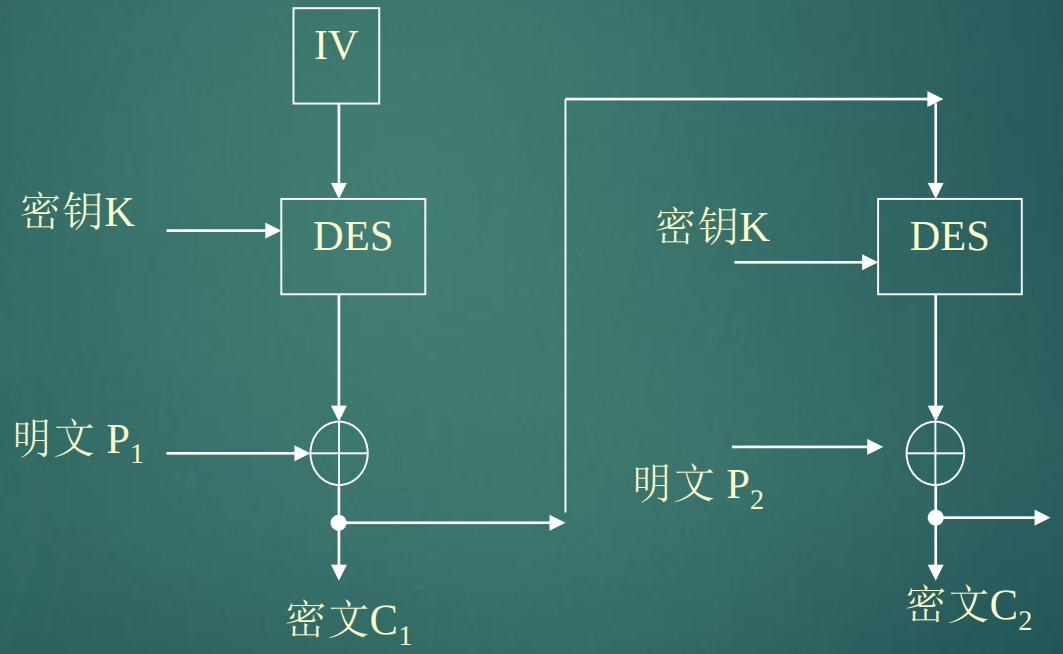
各分组之间独立， 攻击者按组分析容易找出规律。

CBC：加密块链接模式



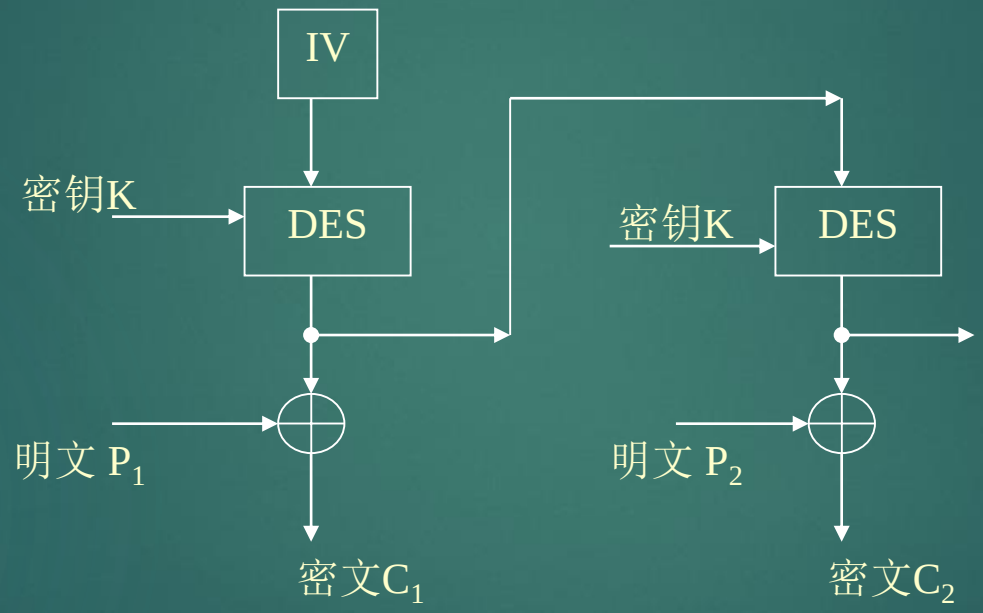
安全性提高了，但容易造成错误传播。

CFB: 加密反馈模式



安全性提高了，但也容易造成错误传播。

OFB:输出反馈模式



加密速度快，但安全性低。（相当于流密码）

分组密码加密方法总结

- 多轮迭代
- 关键是轮函数的设计
- 在轮函数中通过置换和迭代将信息与密钥充分混合
- 安全性取决于轮函数中的S盒设计（替代查表，非线性）
- 选取适当的密码工作模式，增强安全性。

分组密码攻击方法

- 线性攻击
- 差分攻击
- 故障攻击
- 功耗分析

线性攻击

- 用线性方程逼近S盒的特性。
- 1994年Matsui给出了相应的利用r－2轮最佳线性逼近来攻击r轮DES的算法。
- 之后Matsui利用此算法和两个14轮最佳线性逼近做了一次攻击DES的实验，在2^{43} 个明密文对的情况下得到26比特密钥的值，再通过穷举其它30比特而得到全部56比特密钥值。
- Burton S. Kaliski Jr. and M. J. B. Robshawl 提出了多重线性逼近攻击算法。

差分故障攻击

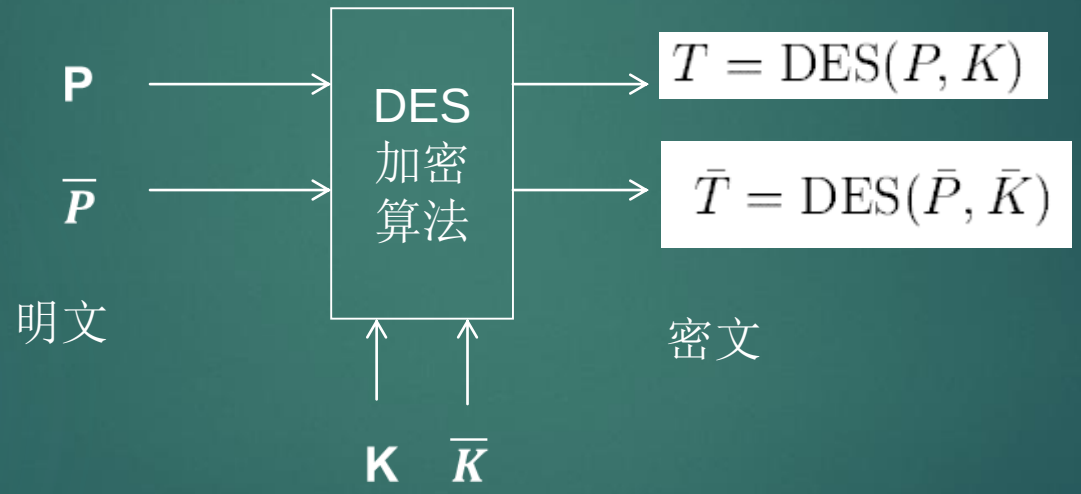
- “故障攻击”的概念是1996年由Boneh等人首次提出的，该方法利用了密码计算过程中的错误。这种攻击方法一经提出立即引起了人们的广泛关注，并展示出了其对密码体制安全性的极大破坏性。
- 1997年，Biham 和Shamir将这
体制，首次提出了“差分故障攻
了DES算法。
- SMS4算法的差分故障攻击就是利
结合差分分析实现的。面向字节
备存储中间值的存储单元进行故
字节错误。
- 利用该攻击方法，理论上仅需要
复出SMS4的128bit加密密钥。

(1)攻击者每次可以诱发存储中间值的存储单元发生任意的单字节错误，但是攻击者不知道错误发生的字节位置以及具体的错误值.

(2)对于同一个明文P而言，攻击者可以获得在同一个密钥K作用下的正确密文C和错误密文C^*.

差分分析方法

- 分析输入的明文差分对对加密后的密文对的影响。

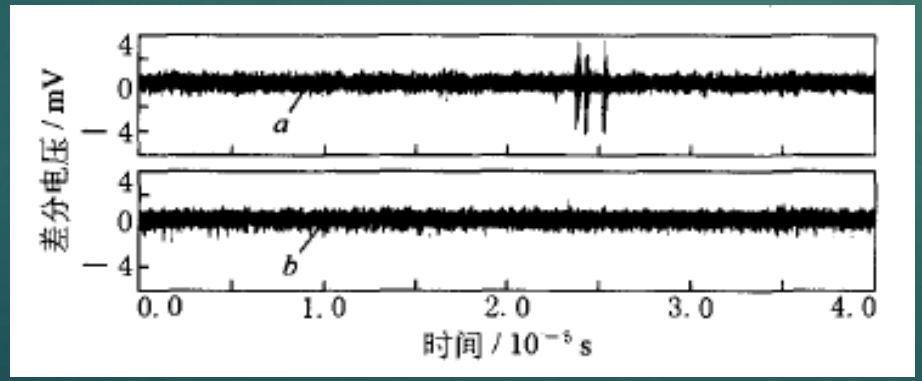


差分破解步骤

1. Choose an appropriate plaintext XOR.
2. Create an appropriate number of plaintext pairs with the chosen plaintext XOR, encrypt them and keep only the resultant ciphertext pairs.
3. For each pair derive the expected output XOR of as many S boxes in the last round as possible from the plaintext XOR and the ciphertext pair. (Note that the input pair of the last round is known since it appears as part of the ciphertext pair).
4. For each possible key value, count the number of pairs that result with the expected output XOR using this key value in the last round.
5. The right key value is the (hopefully unique) key value suggested by all the pairs.

功耗分析

- 根据密码芯片的功耗情况（电源的波动）分析出密码算法的执行情况，从而破解密码。
- 可通过监测密码系统功耗等物理泄漏信号，对其进行分析，减小密钥穷举空间，之后在小范围内进行穷举攻击，进而获得密码系统密钥。



一般采用差分功耗分析

- 插入《暗算》第21集密码破解示例。

保密通信

- 根据应用需求设计适度安全等级；
- 根据安全等级、计算能力及功耗要求，设计加解密算法；
- 设计密钥/密钥流的同步方法；
- 设计密钥交换管理机制。

保密通信形式

- 点到点通信（point to point）
- 端到端通信(peer to peer)
- 点到多点（point to multi-point）
- 网络通信（分布式通信，distribution）

安全通信要素

- 双方或多方通信者，通信者身份的确认
- 通信者之间的加密/解密设备
- 通信者之间拥有安全通信密钥
- 密钥的协商与更新机制
- 设备的配置与管理

通信者身份的确认

- 通常以拥有合法密钥作为身份确认标识；
- 通过身份认证协议确认双方的身份。

保密原则

- 即使做不到理论上是安全的，也要设计成实际是不可破或很困难的；
- 不依赖于对加密体制或算法的保密，只依赖密钥的保密。Kerckhoff原则。
- 加密和解密算法适合所有密钥空间的元素。
- 便于实现和使用。

密钥管理

- 密钥管理包括：产生、存储、备份/恢复、装入、分配、保护、更新、控制、丢失、吊销、销毁；
- 密钥类型：

  基本密钥（Base Key）或初始密钥(Primary Key)或用户密钥(User Key)；

  会话密钥(Session Key):在一次通信时所用的密钥，通过密钥交换协议产生；

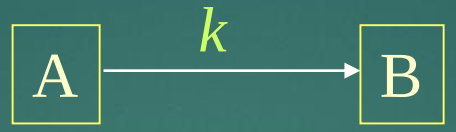
  密钥加密密钥（Key Encrypting Key）或密钥传递密钥（Key Transport Key）；

密钥的分配

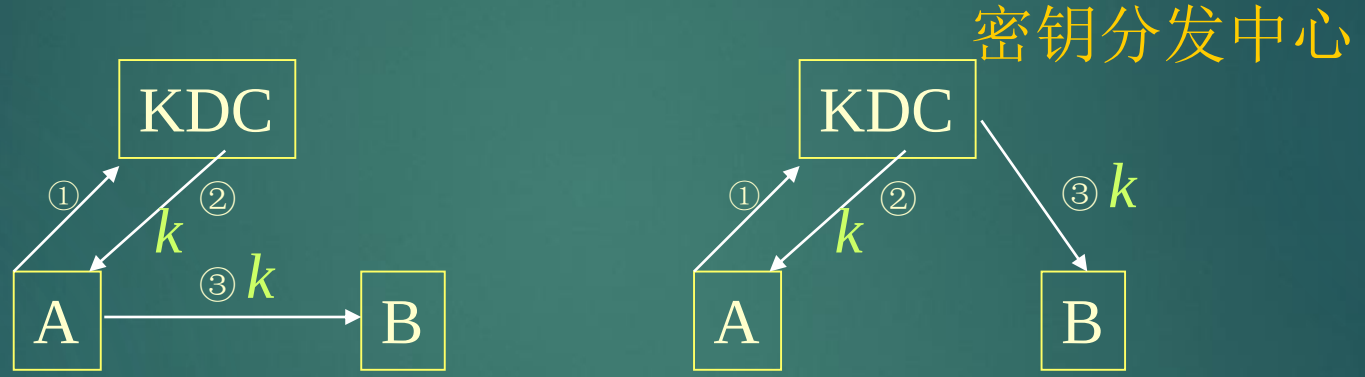
- 利用安全信道实现，如人工传递；
- 利用数学上求逆的困难性，即各种双钥体制所建立的安全信道实现；
- 利用物理现象实现，如量子通信信道；
- 利用异构信道，如电话线。

密钥分配的基本模式

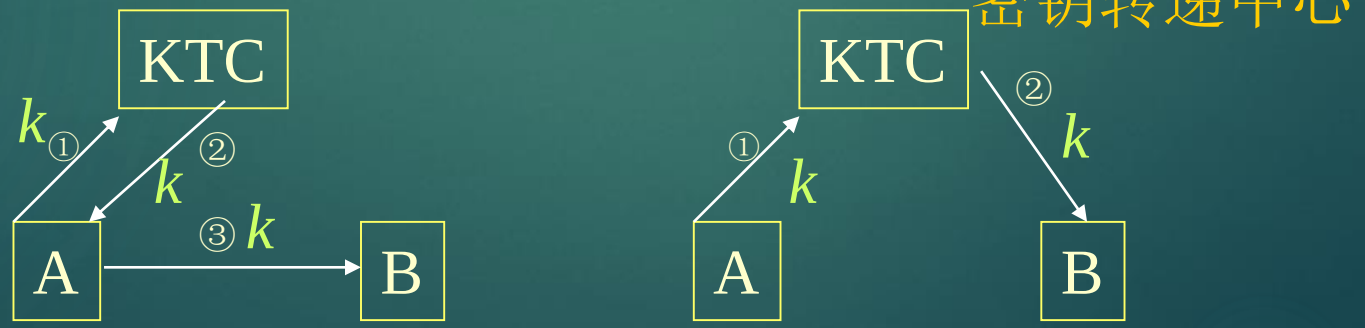
(a)



(b)

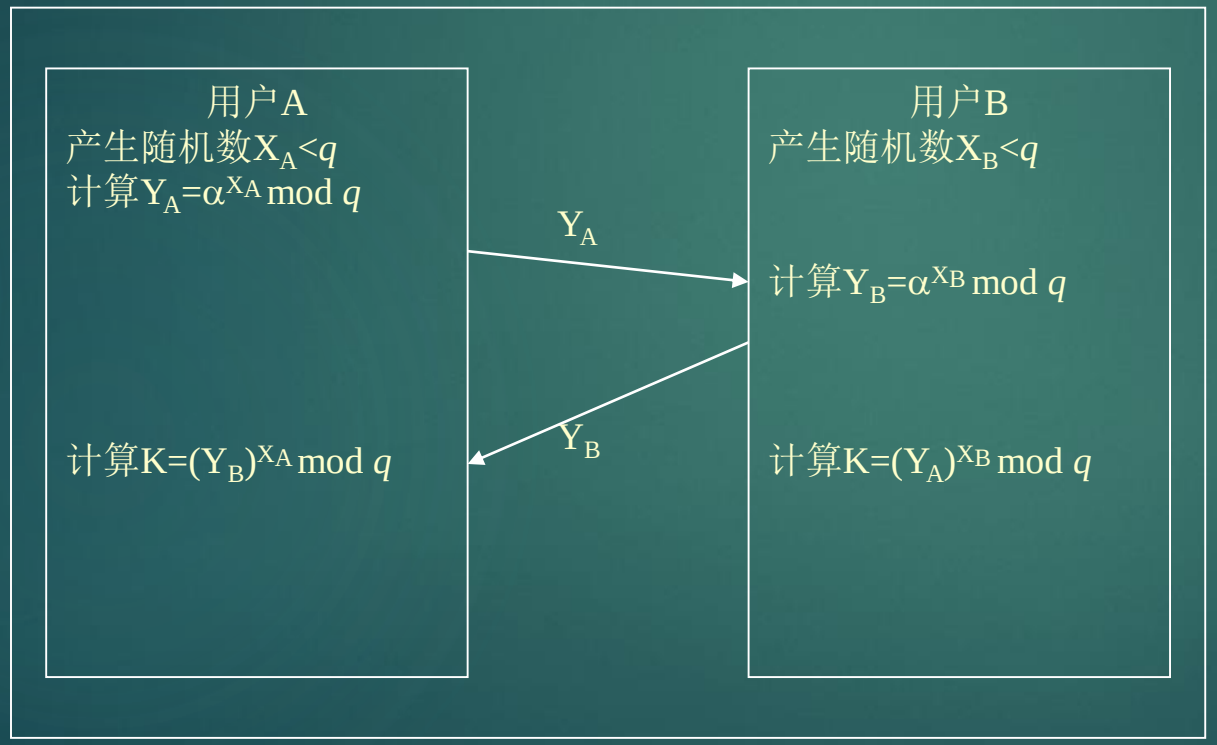


(c)



密钥交换协议

- Diffie-Hellman密钥交换协议



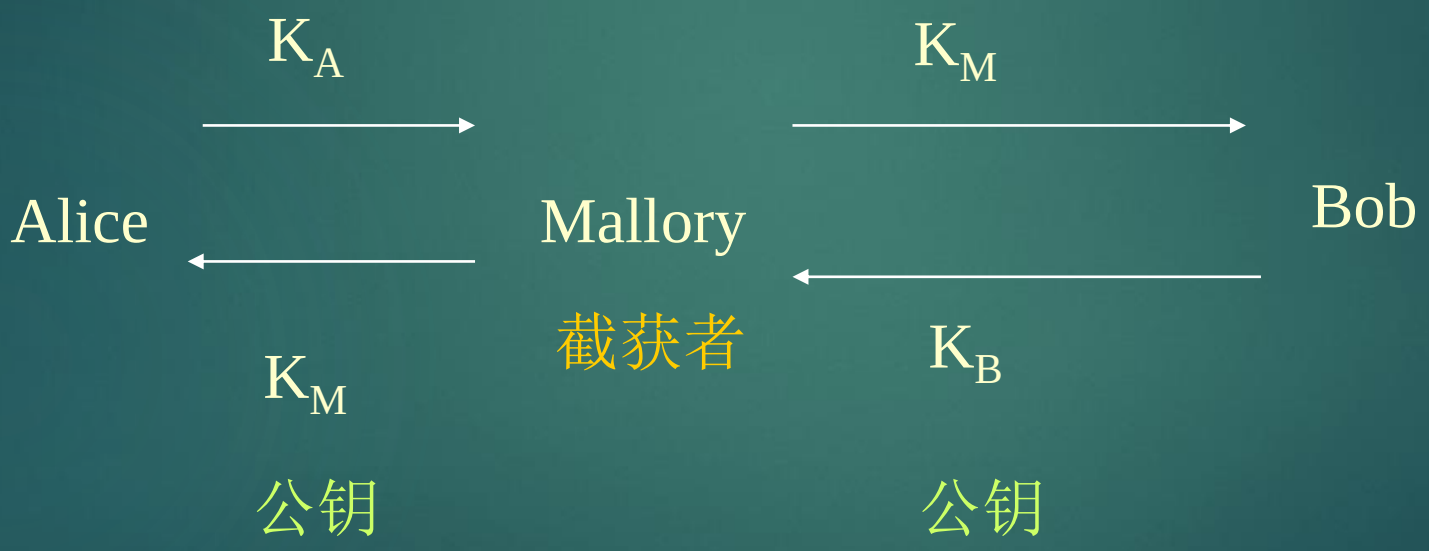
公钥

q、α为全局公开值
X_A为A的私钥
X_B为B的私钥

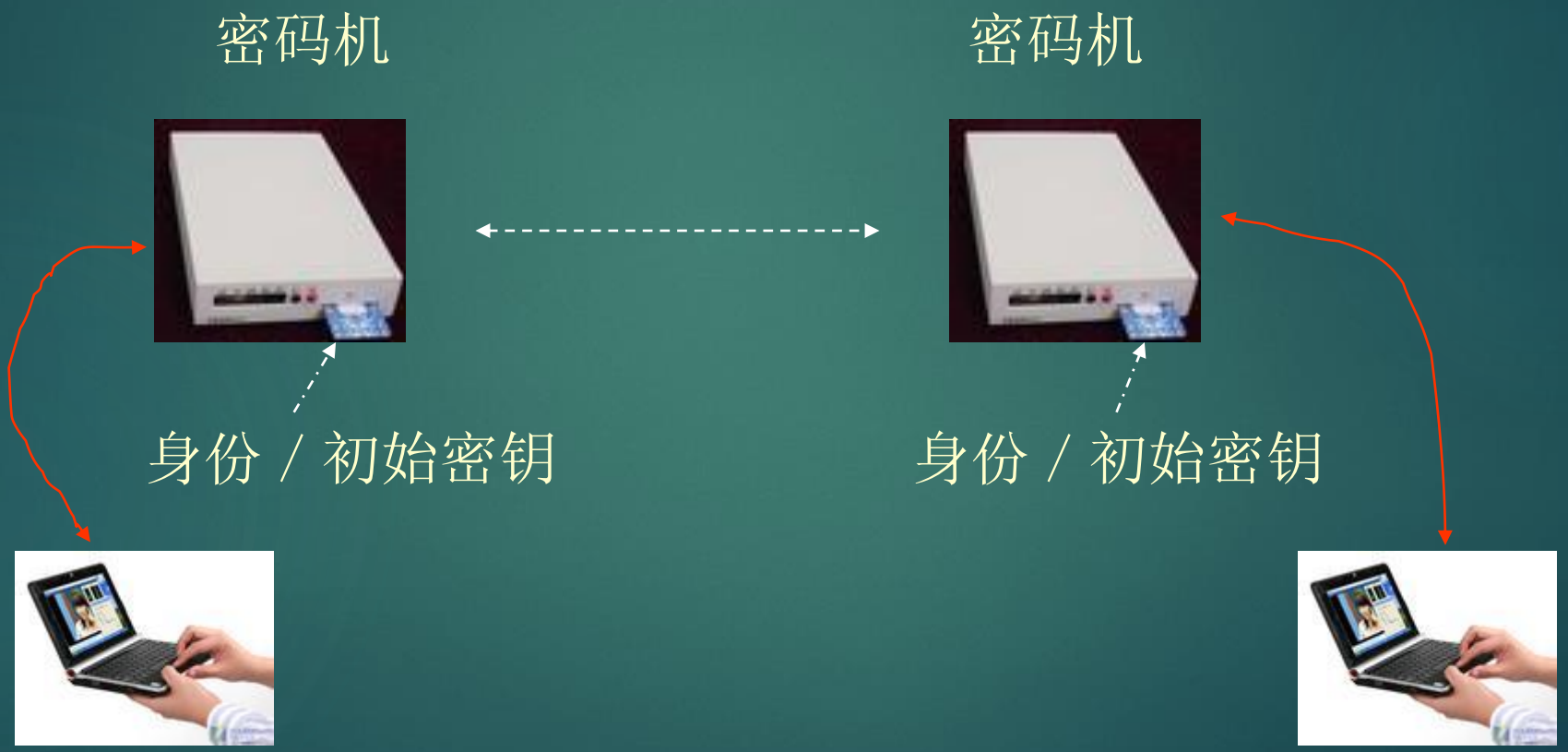
优势：
按需产生密钥，无须长久保存；
只有全局参数约定；

不足：
没有涉及双方身份信息；
容易受中间人攻击

Diffie-Hellman密钥交换协议中间人攻击

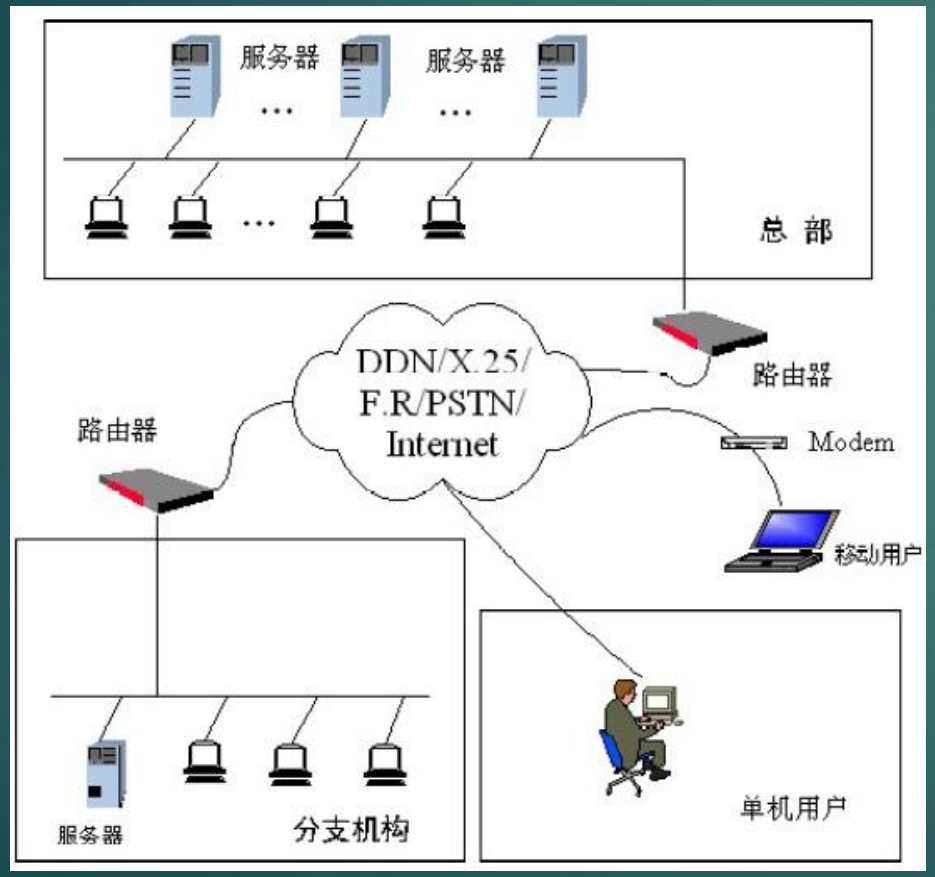


保密通信系统构成



保密通信系统实例—VPN(虚拟专用网)

在开放的互联网上建立两个节点之间的安全通道.



安全需求:

- 除了发送方和接收方外，数据不能被其他未经授权的人读取（数据的保密性）；
- 传送过程中不能被篡改（数据的一致性）；
- 发送方需要能确信接收方身份的真实性（身份验证）；
- 发送方不能否认自己的发送行为（数据的抗抵赖性）。

VPN的身份认证

- 预共享密钥：预共享密钥方式是当前采用比较多也是最不安全的一种方式。它预先为VPN连接双方设定同样的密钥，以此作为VPN连接双方建立VPN前相互信任的凭据。双方相信预先设定的密钥。
- 数字证书：数字证书方式是当前用的比较少也是最有应用前景的一种方式。双方相信同一个信任机构签发的证书。

VPN系统

- VPN 加密网关
- IPSec密码支持系统
- VPN客户端软件包
- VPN安全策略管理中心软件包
- 企业级CA子系统。

VPN 加密网关

- 是符合工业标准的硬件设备，集VPN和防火墙功能于一身，主要用于局域网络之间的互连。
- 它的内部采用VPN专用密码芯片SSP02-A，以实现大流量、高速数据传输的要求，保证用户在网上传递信息的机密性、完整性和有效性。

技术指标(100M网络环境中)：

- VPN通道数量：2048条
- 防火墙安全过滤带宽：98Mbps
- 网络加解密通信速率：最大92Mbps
- 客户端最大VPN通道数：10条
- 密钥长度：对称算法：128位，非对称算法：1024/2048位
- 支持的协议：TCP/IP、IPSec、ESP、IKE等
- 支持的标准加密算法：VPN专用密码算法，AES，HMAC-SHA-1，Diffie-Hellman，1024/2048-bit RSA
- 适用通信环境：可接入DDN、ISDN、PSTN、ADSL、宽带IP等多种通信信道。

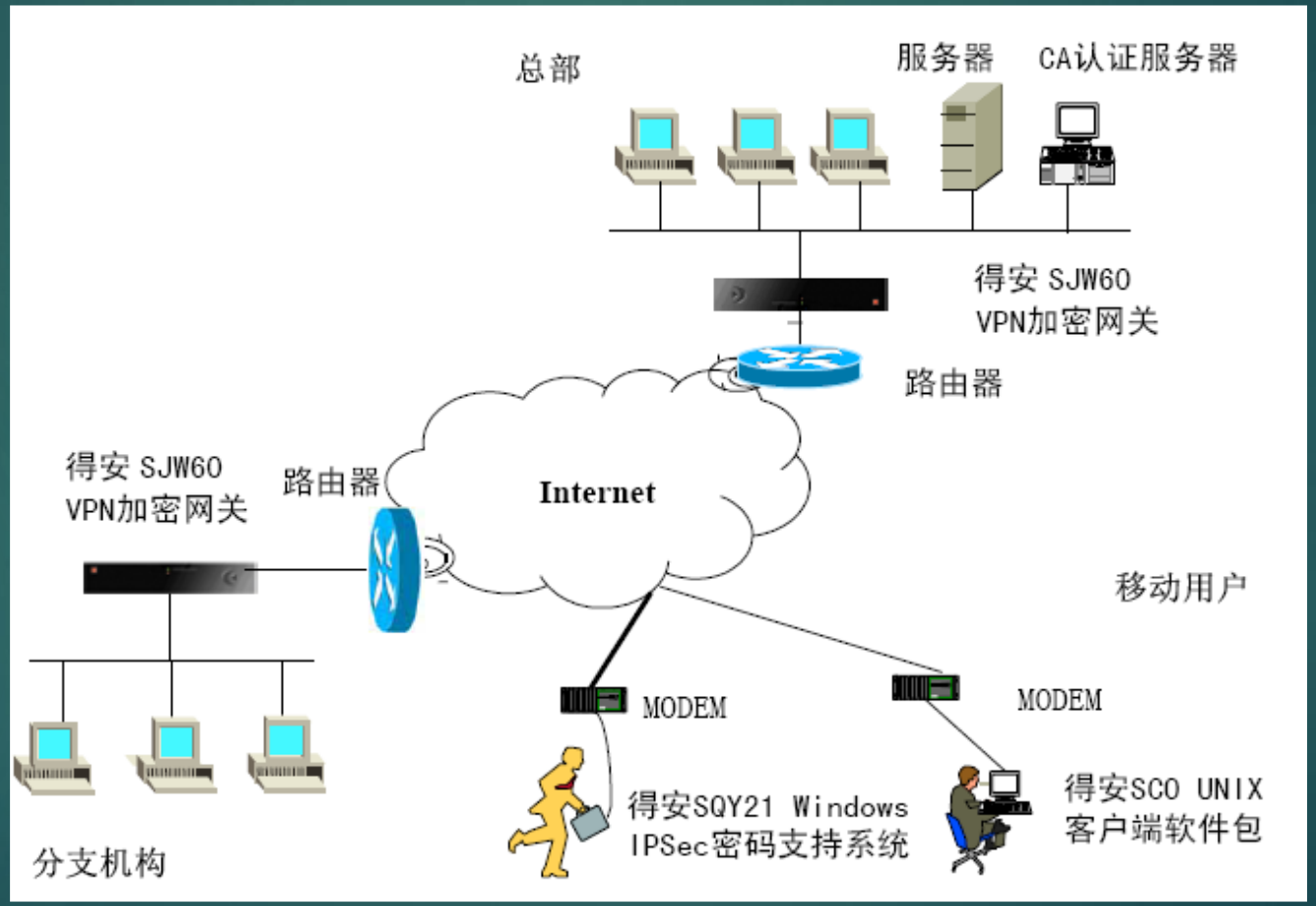
企业级CA子系统

- 采用基于PKI体系结构的“集中认证，分布协商”的密钥管理方案，支持X.509体系的身份鉴别认证机制。任何拥有由同一个CA签发的合法证书的VPN加密网关或VPN客户端，都会被这个VPN网络环境中其它的通信实体认为是可信任的。
- 一个VPN应用环境中的任意两台加密网关建立安全隧道之前通过出示数字证书来相互确认身份，其内容包括接受认证申请、审查申请人的资格、生成并发放数字证书。

VPN策略中心软件包

- 可以对整个VPN网络的安全策略进行集中管理和配置。
- 这能有效防止用户在对每台加密网关进行单独的策略设置时出现的策略配置不一致等情况。

VPN的部署示意图


总部
服务器
CA认证服务器
得安 SJW60
VPN加密网关
路由器
Internet
得安 SJW60
VPN加密网关
路由器
移动用户
MODEM
MODEM
得安SQY21 Windows
IPSec密码支持系统
得安SCO UNIX
客户端软件包
分支机构

本部分小结

- 设计保密通信系统的核心问题是寻找安全强度满足要求的加密算法；
- 保密通信系统要解决通信双方的身份认证问题；
- 密码算法是可以公开的，密钥（私钥）是保密的；
- 密钥的管理、更新是保障通信安全很关键的要素；

其它关心的问题

- 加密/解密速度对系统性能的影响
- 安全特性的显性化
- 安全代价的合理化（加密算法、密钥长度、密钥更新周期、加密/解密速度等的确定需要依据具体应用需求而定）
- 非密码的安全通信技术（如量子通信）

第三次 习题

1、阐述IPsec VPN的基本原理，包括AH模式和ESP模式、身份认证、数据加密、密钥协商以及组网方式。

2、设素数p=23，有限域F_{23}上的椭圆曲线方程$y^2=x^3+x+4$，且已知P=(4,7)、Q=(13,11)为该椭圆曲线上的两点，求$P+Q$=？（结果以有限域上的整数表示）