

序列密码

本讲主要内容

- 序列密码的介绍
- 线性反馈移位寄存器
- 非线性序列
- 序列密码的算法举例(A5、RC4)

序列密码的简介

序列密码通常认为起源于20世纪20年代的**维尔姆 (Vernam) 密码**，Vernam密码中的密钥序列要求是随机的序列，由于**随机**的密钥序列的产生、存储以及分配等方面存在一定的困难，Vernam体制在当时并没有得到广泛的应用。随着微电子技术和数学理论的发展和完善，基于伪随机序列的序列密码得到了长足的发展和应用，其产生有比较成熟的数学理论工具。如果密钥序列是随机的序列，则序列密码就是“一次一密”密码体制。商农已经证明**“一次一密”**密码体制在理论上是不可破译的。目前，序列密码是世界各国的军事和外交等重要领域中使用的主要密码体制之一。

在序列密码中，加密和解密所用的密钥序列都是伪随机序列，序列密码是对称密码体制中的一类，又称为**流密码**。

序列密码的描述

在序列密码中，将明文消息按一定长度（长度较小）分组，然后对**各组用相关但不同**的密钥进行加密，产生相应的密文，相同的明文分组会因在明文序列中的位置不同而对应于不同的密文分组。解密用相同的密钥序列对密文序列进行分组解密以恢复出明文序列。

➤ 设明文为 $p = p_0 p_1 p_2 \dots$ $p_i \in GF(2), i \geq 0$

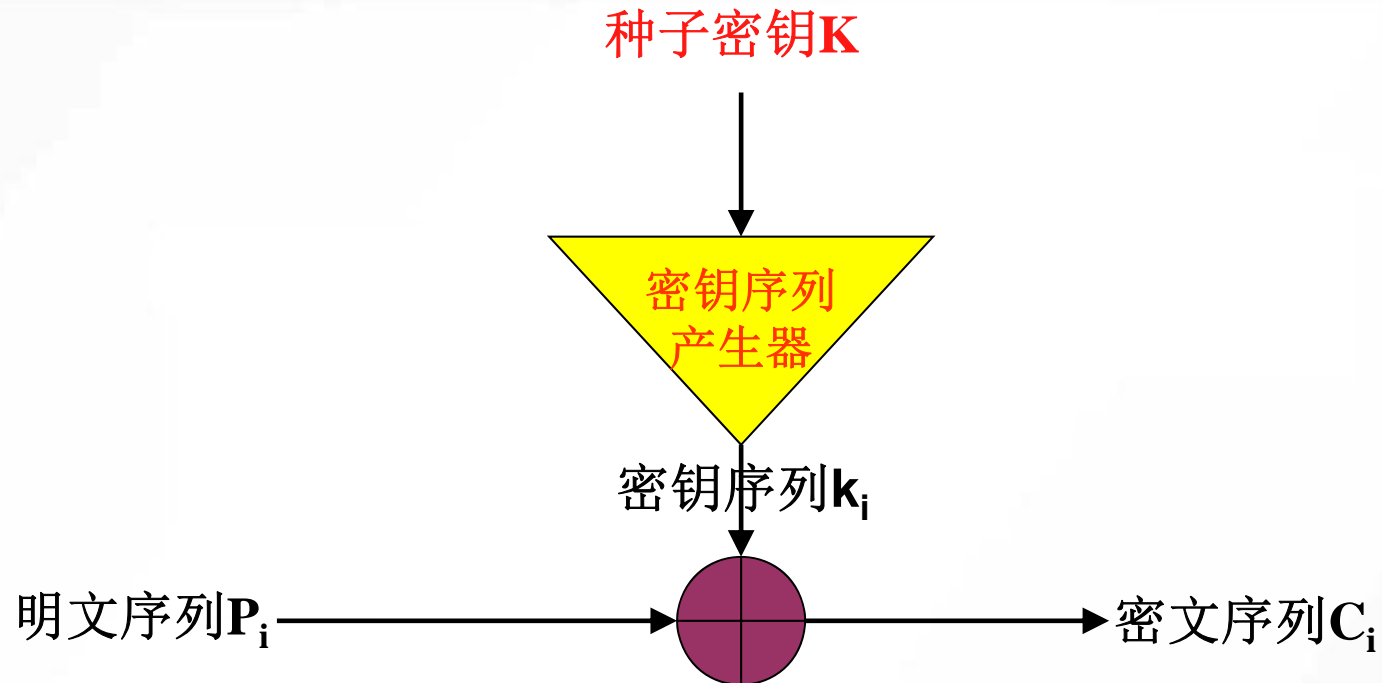
➤ 设密钥为 $k = k_0 k_1 k_2 \dots$ $k_i \in GF(2), i \geq 0$

➤ 设密文为 $c = c_0 c_1 c_2 \dots$ $y_i \in GF(2), i \geq 0$

➤ 则加密变换为 $c_i = E_{k_i}(p_i)$ $i \geq 0$

➤ 则解密变换为 $p_i = D_{k_i}(c_i)$ $i \geq 0$

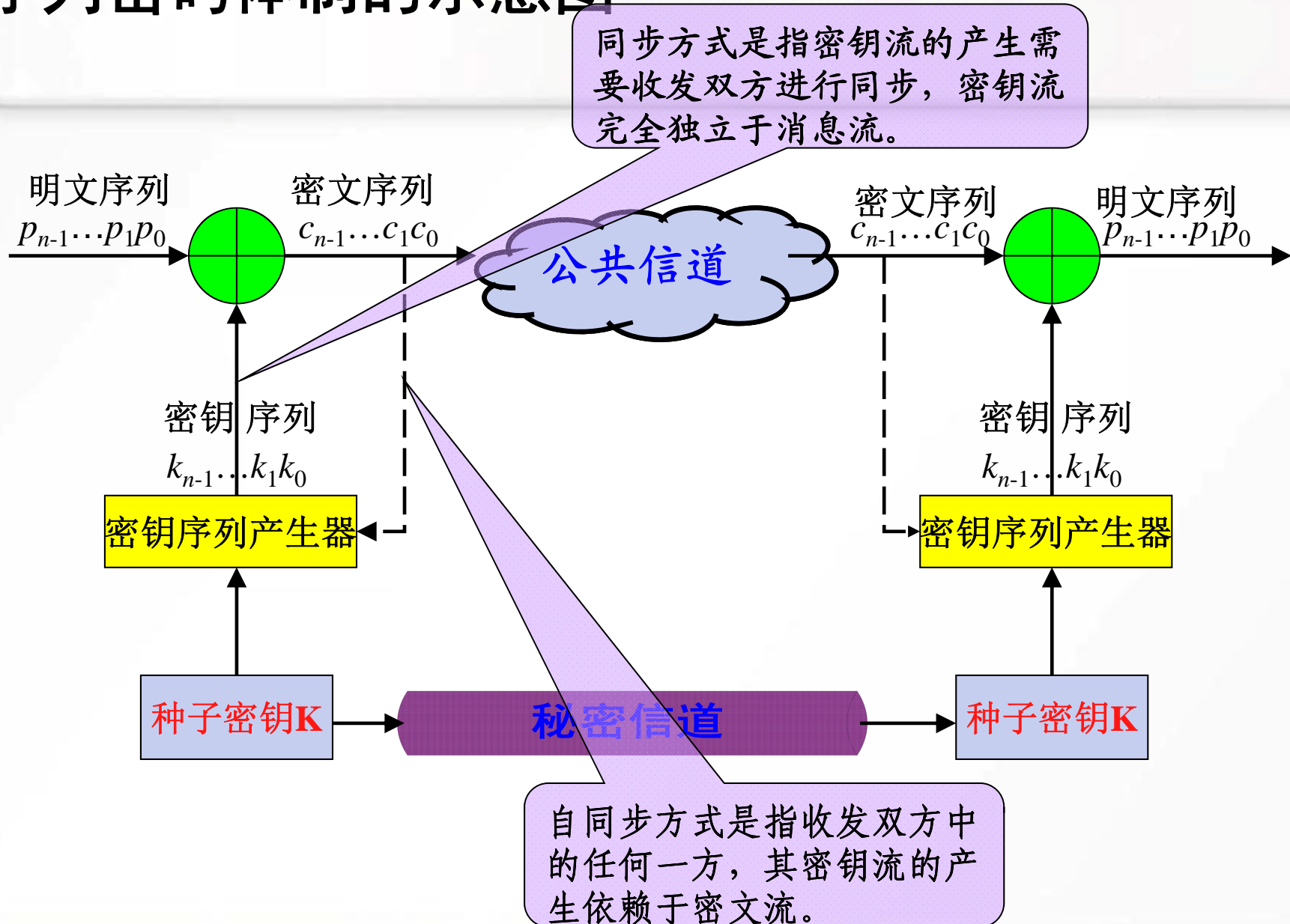
序列密码的原理



特点:

- 加解密运算只是简单的模二加运算。
- 密码安全强度主要依赖密钥流的安全性。

序列密码体制的示意图



密钥序列产生器 (KG) 基本要求

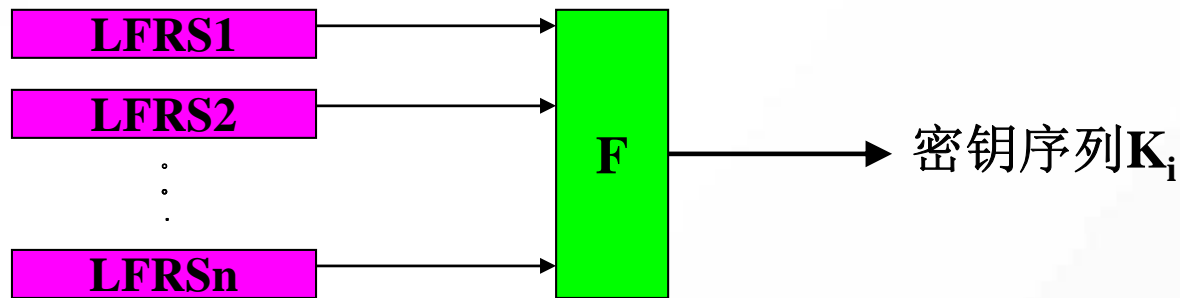
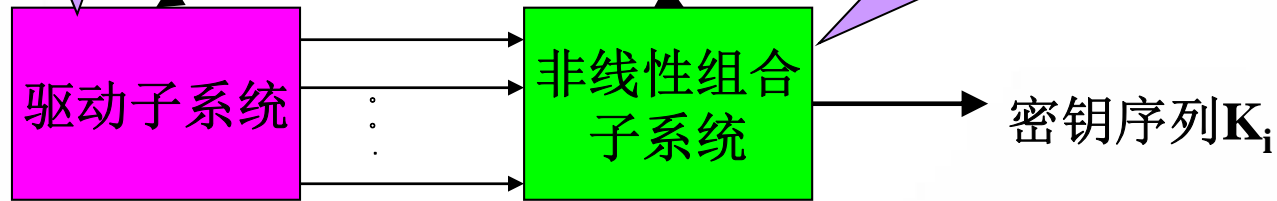
- 种子密钥K的长度足够大，一般应在128位以上；
- KG生成的密钥序列 $\{k_i\}$ 具极大周期；
- $\{k_i\}$ 具有均匀的n-元分布，即在一个周期环上，某特定形式的n-长bit串与其求反，两者出现的频数大抵相当(例如，均匀的游程分布)；
- 利用统计方法由 $\{k_i\}$ 提取关于KG结构或K的信息在计算上不可行；
- 雪崩效应。即K任一位的改变要引起 $\{k_i\}$ 在全貌上的变化。
- 密钥流 $\{k_i\}$ 不可预测的。密文及相应的明文的部分信息，不能确定整个 $\{k_i\}$ 。

密钥序列产生器的分解

一般由m-序列生成器构成，提供若干周期大、分布特性好的序列。

种子密钥K

把产生的多条驱动序列综合在一起的一些非线性编码手段，目的是有效地破坏和掩盖多条驱动序列中存在的规律或关系，提高复杂度。



常用的密钥序列产生器

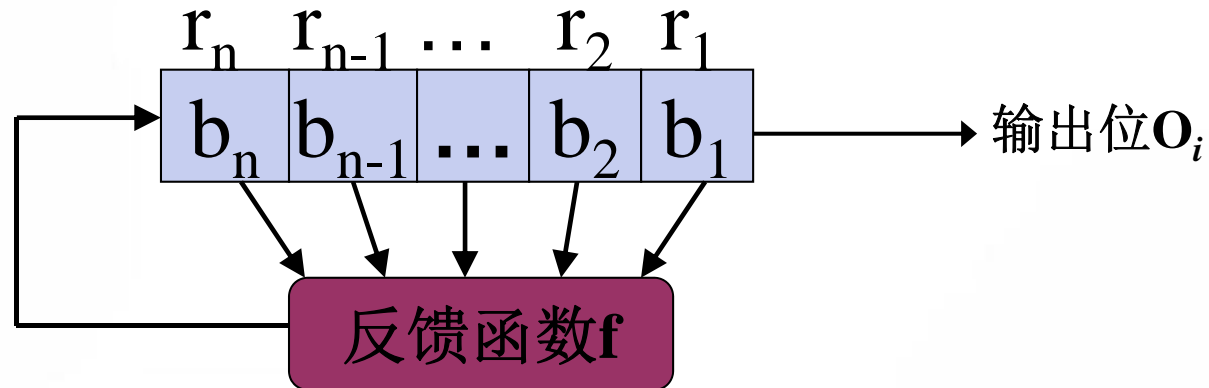
线性反馈移位寄存器理论的简介

序列密码的关键是设计一个随机性好的密钥流发生器，为了研究密钥流发生器，挪威政府的首席密码学家 **Ernst Selmer** 于**1965**年提出了移位寄存器理论，它是序列密码中研究随机密钥流的主要数学工具。

尤其，线性反馈移位寄存器因其实现简单、速度快、有较为成熟的理论等优点而成为构造密码流生成器的最重要的部件之一。

反馈移位寄存器

反馈移位寄存器(**feedback shift register,FSR**)是由n位的寄存器和反馈函数(**feedback function**)组成,如下图所示,n位的寄存器中的初始值称为**移位寄存器的初态**。

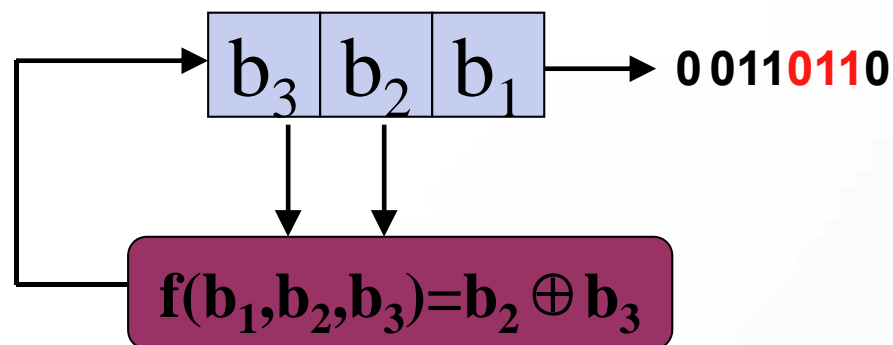


工作原理: 移位寄存器中所有位的值右移1位,最右边的一个寄存器移出的值是输出位,最左边一个寄存器的值由反馈函数的输出值填充,此过程称为进动1拍。反馈函数 f 是 n 个变元(b_1, b_2, \dots, b_n)的布尔函数。移位寄存器根据需要不断地进动 m 拍,便有 m 位的输出,形成输出序列 O_1, O_2, \dots, O_m 。

反馈移位寄存器（举例）

一个3-级的反馈移位寄存器，反馈函数 $f(x)=b_2 \oplus b_3$ ，初态为100，输出序列生成过程如下：

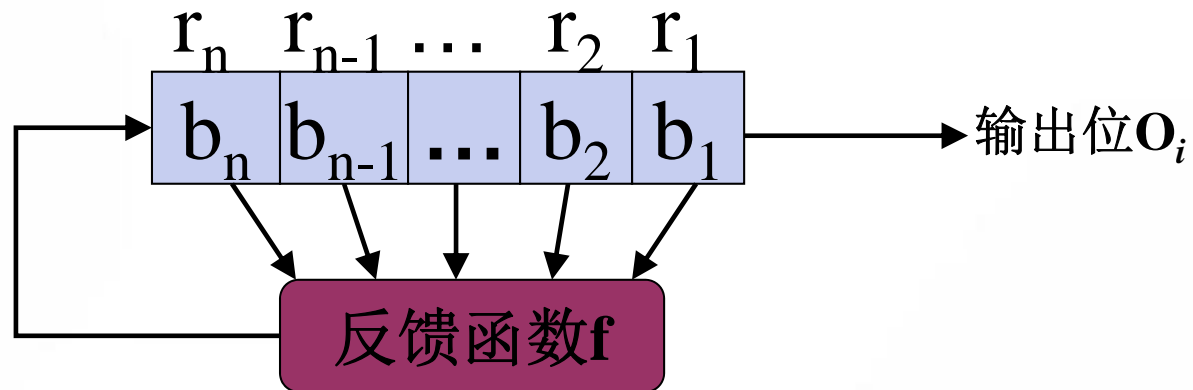
状态	输出位
100	0
110	0
011	1
101	1
110	0
011	1
101	1
110	0



移位寄存器的周期是指输出序列中连续且重复出现部分的长度。上面输出序列周期长度为3。

线性反馈移位寄存器

线性反馈移位寄存器(**linear feedback shift register, LFSR**)是一种特殊的**FSR**，其反馈函数是移位寄存器中某些位的异或，参与运算的这些位叫做抽头位。其一般组成结构如下图所示。

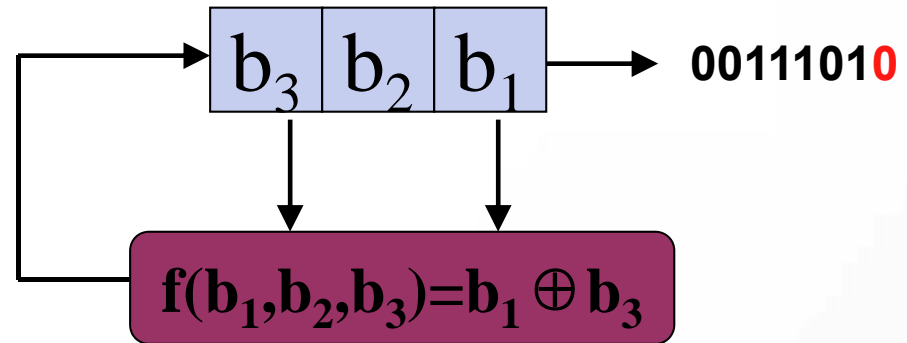


特征：n-级 **LFSR**的有效状态为 2^n-1 （全0状态除外，因全0状态的输出序列一直为全0），也即理论上能够产生周期为 2^n-1 的伪随机序列。但要产生最大周期的输出序列，抽头位要有要求。

线性反馈移位寄存器（举例）

一个3-级的反馈移位寄存器，反馈函数 $f(x)=b_1 \oplus b_3$ ，初态为100，输出序列生成过程如下：

状态	输出位
100	0
110	0
111	1
011	1
101	1
010	0
001	1
100	0



- ✓ 上面输出序列周期长度为7。
- ✓ 抽头位对输出周期长度的影响起着决定性的作用。
- ✓ 初态对输出周期长度没影响。

m序列的简介

线性反馈移位寄存器输出序列的性质完全由其反馈函数决定。 n 级线性反馈移位寄存器最多有 2^n 个不同的状态。若其初始状态为0，则其状态恒为0。若其初始状态非0，则其后继状态不会为0。因此， n 级线性反馈移位寄存器的状态周期小于等于 2^n-1 ，其输出序列的周期与状态周期相等，也小于等于 2^n-1 。

只要选择合适的反馈函数便可使序列的周期达到最大值 2^n-1 ，周期达到最大值的序列称为m序列。

m-序列特性

- **0,1平衡性**: 在一个周期内, **0**、**1**出现的次数分别为 $2^{n-1}-1$ 和 2^{n-1} 。
- **游程特性**: 在一个周期内, 总游程数为 2^{n-1} ; 对 $1 \leq i \leq n-2$, 长为 i 的游程有 2^{n-i-1} 个, 且**0**、**1**游程各半; 长为 $n-1$ 的**0**游程一个, 长为 n 的**1**游程一个。
- **$\{x_i\}$ 的自相关函数为**

$$R(\tau) = \begin{cases} 1, & \tau=0 \\ -\frac{1}{2^n-1}, & 0 < \tau \leq 2^n-2 \end{cases}$$

$$R(j) = \frac{A-D}{P}, j=1,2,\dots; p \text{ 为序列 } \{x_i\} \text{ 的周期。}$$

$$\text{其中: } A = |\{0 \leq i < p, x_i = x_{i+j}\}|, D = |\{0 \leq i < p, x_i \neq x_{i+j}\}|,$$

关于m序列的自相关函数

设序列 $\{a_i\}$ 满足递推关系:

$$a_{h+n} = c_1 a_{h+n-1} \oplus c_2 a_{h+n-2} \oplus \cdots \oplus c_n a_h$$

$$\therefore a_{h+n+\tau} = c_1 a_{h+n+\tau-1} \oplus c_2 a_{h+n+\tau-2} \oplus \cdots \oplus c_n a_{h+\tau}$$

$$\therefore a_{h+n} \oplus a_{h+n+\tau} = c_1 (a_{h+n-1} \oplus a_{h+n+\tau-1}) \oplus c_2 (a_{h+n-2} \oplus a_{h+n+\tau-2}) \oplus \cdots \oplus c_n (a_h \oplus a_{h+\tau})$$

令 $b_j = a_j \oplus a_{j+\tau}$, 可得序列 $\{b_i\}$ 满足递推关系:

$$b_{h+n} = c_1 b_{h+n-1} \oplus c_2 b_{h+n-2} \oplus \cdots \oplus c_n b_h$$

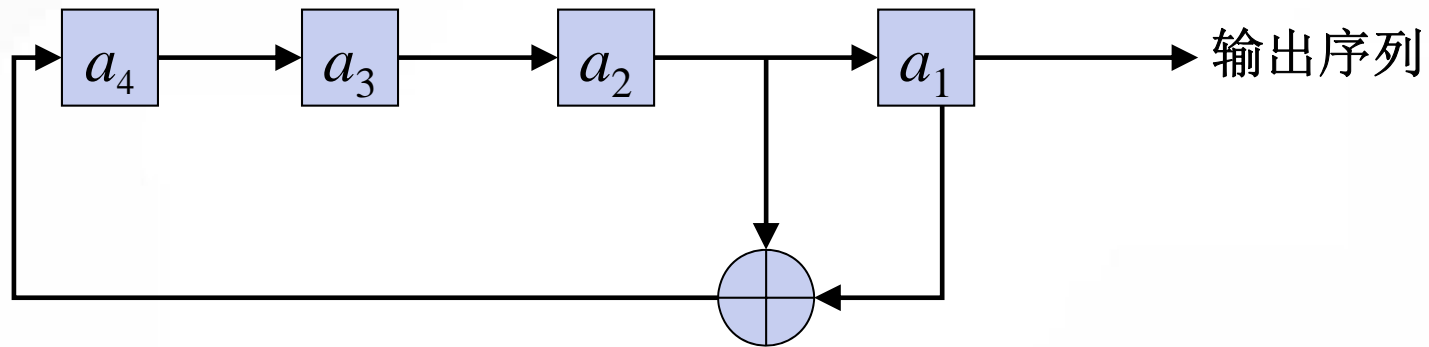
$\therefore \{b_i\}$ 也是m序列。其中在 $\{b_i\}$ 一个周期中0的个数为A,1的个数为D。

m-序列生成的简介

- 为了使LFSR成为m-序列，由抽头序列加上常数1形成的多项式必须是**本原多项式**模2。多项式的阶即移位寄存器的长度。
- n次本原多项式的个数为 $\frac{\phi(2^n - 1)}{n}$ ，其中 ϕ 为欧拉函数。
- 本原多项式的最高次数就是m-序列的长度，除了 $x^0 = 1$ 以外的其他项的次数指明了抽头序列，这些抽头从移位寄存器的左边开始计数。简而言之：**本原多项式中的抽头的阶数越低，越靠近移位寄存器的左边。**

m序列的举例

本原多项式: $p(x) = (1 + x^3 + x^4)$



初态: $a_4 a_3 a_2 a_1 = 1001$

输出序列: (100110101111000)

按前面的性质有: 游程的总数为8, 分别为 1, 00, 11, 0, 1, 0, 1111, 000。其中有一半的游程长度为2, 长度为2的游程为四分之一, 最后有一个长度为4的游程和一个长度为3的游程。

m序列密码的破译

设敌手知道一段长为 $2n$ 的明密文对，即已知

$$x = x_1 x_2 \cdots x_{2n} \quad y = y_1 y_2 \cdots y_{2n}$$

于是可求出一段长为 $2n$ 的密钥序列

$$z = z_1 z_2 \cdots z_{2n}$$

其中

$$y_i = x_i \oplus z_i, \text{ 可得 } z_i = x_i \oplus y_i$$

m序列密码的破译(续)

由此可推出线性反馈移位寄存器连续的**n+1**个状态:

$$S_1 = (z_1 z_2 \cdots z_n) \stackrel{\text{记为}}{=} (a_1 a_2 \cdots a_n)$$

$$S_2 = (z_2 z_3 \cdots z_{n+1}) \stackrel{\text{记为}}{=} (a_2 a_3 \cdots a_{n+1})$$

...

$$S_{n+1} = (z_{n+1} z_{n+2} \cdots z_{2n}) \stackrel{\text{记为}}{=} (a_{n+1} a_{n+2} \cdots a_{2n})$$

序列 $\{a_i\}$ 满足线性递推关系:

$$a_{h+n} = c_1 a_{h+n-1} \oplus c_2 a_{h+n-2} \oplus \cdots \oplus c_n a_h \quad (1 \leq h \leq n)$$

m序列密码的破译(续)

生成矩阵 $X = (S_1 \ S_2 \ \cdots \ S_n)$

而

$$\begin{aligned} (a_{n+1} \ a_{n+2} \ \cdots \ a_{2n}) &= (c_n \ c_{n-1} \ \cdots \ c_1) \begin{pmatrix} a_1 & a_2 & \cdots & a_n \\ a_2 & a_3 & \cdots & a_{n+1} \\ \cdots & & & \\ a_n & a_{n+1} & \cdots & a_{2n-1} \end{pmatrix} \\ &= (c_n \ c_{n-1} \ \cdots \ c_1) X \end{aligned}$$

若 X 可逆, 则

$$(c_n \ c_{n-1} \ \cdots \ c_1) = (a_{n+1} \ a_{n+2} \ \cdots \ a_{2n}) X^{-1}$$

m序列密码的破译举例

例： 设敌手得到密文串**101101011110010**和相应的明文串**011001111111001**，因此可计算出相应的密钥流为**110100100001011**。进一步假定敌手还知道密钥流是使用**5**级线性反馈移位寄存器产生的，那么敌手可分别用密文串中的前**10**个比特和明文串中的前**10**个比特建立如下方程

$$(a_6 \ a_7 \ a_8 \ a_9 \ a_{10}) = (c_5 \ c_4 \ c_3 \ c_2 \ c_1) \begin{pmatrix} a_1 & a_2 & a_3 & a_4 & a_5 \\ a_2 & a_3 & a_4 & a_5 & a_6 \\ a_3 & a_4 & a_5 & a_6 & a_7 \\ a_4 & a_5 & a_6 & a_7 & a_8 \\ a_5 & a_6 & a_7 & a_8 & a_9 \end{pmatrix}$$

m序列密码的破译举例(续一)

即

$$(0 \ 1 \ 0 \ 0 \ 0) = (c_5 \ c_4 \ c_3 \ c_2 \ c_1) \begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

而

$$\begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}^{-1} = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{pmatrix}$$

m序列密码的破译举例(续二)

从而得到

$$(c_5 \ c_4 \ c_3 \ c_2 \ c_1) = (0 \ 1 \ 0 \ 0 \ 0) \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{pmatrix}$$

$$\text{所以 } (c_5 \ c_4 \ c_3 \ c_2 \ c_1) = (1 \ 0 \ 0 \ 1 \ 0)$$

密钥流的递推关系为

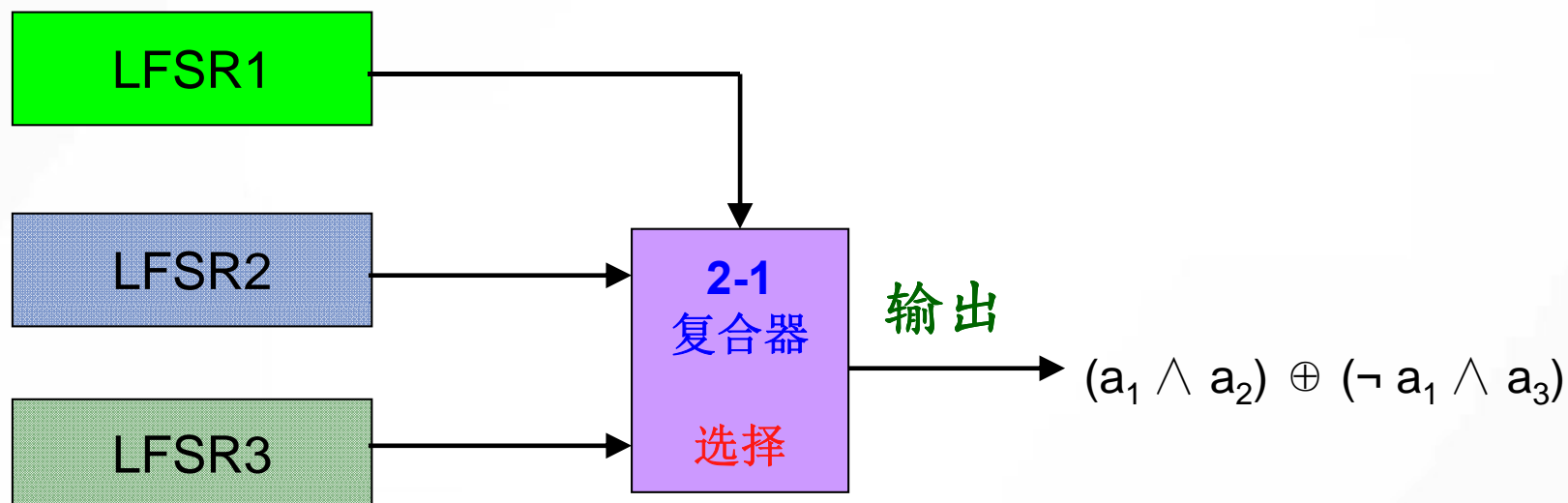
$$a_{i+5} = c_5 a_i \oplus c_2 a_{i+3} = a_i \oplus a_{i+3}$$

非线性序列

为了使密钥流生成器输出的二元序列尽可能复杂，应保证其周期尽可能大、线性复杂度和不可预测性尽可能高，因此常使用多个LFSR来构造二元序列，称每个LFSR的输出序列为**驱动序列**，显然密钥流生成器输出序列的周期**不大于**各驱动序列周期的乘积，因此，提高输出序列的线性复杂度应从极大化其周期开始。

- **Geffe**序列生成器
- **J-K**触发器
- **Pless**生成器
- 门限发生器

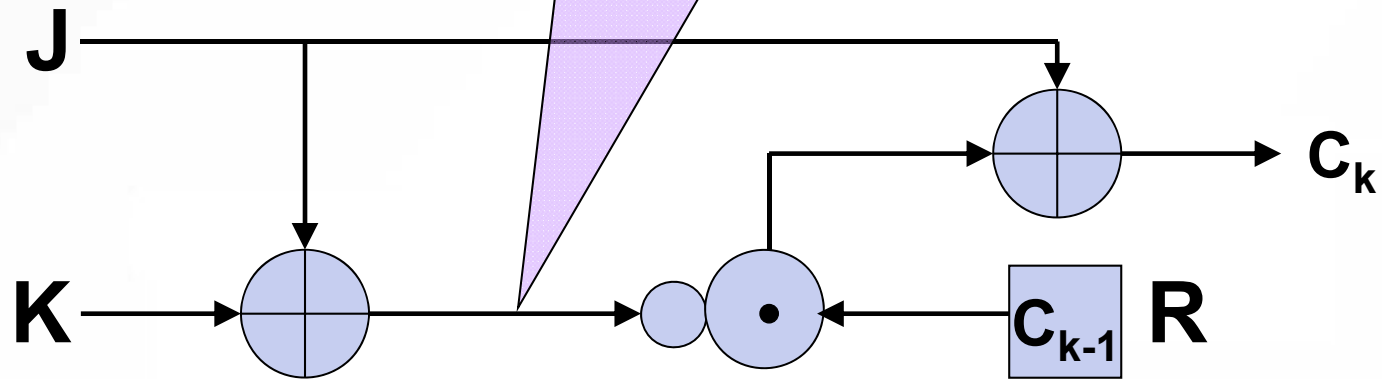
Geffe序列生成器



这个发生器的周期是三个**LFSR**的周期的最小公倍数。假设三个本原反馈多项式的阶数互素，那么这个发生器的周期是三个**LFSR**的周期之积。

J-K触发器

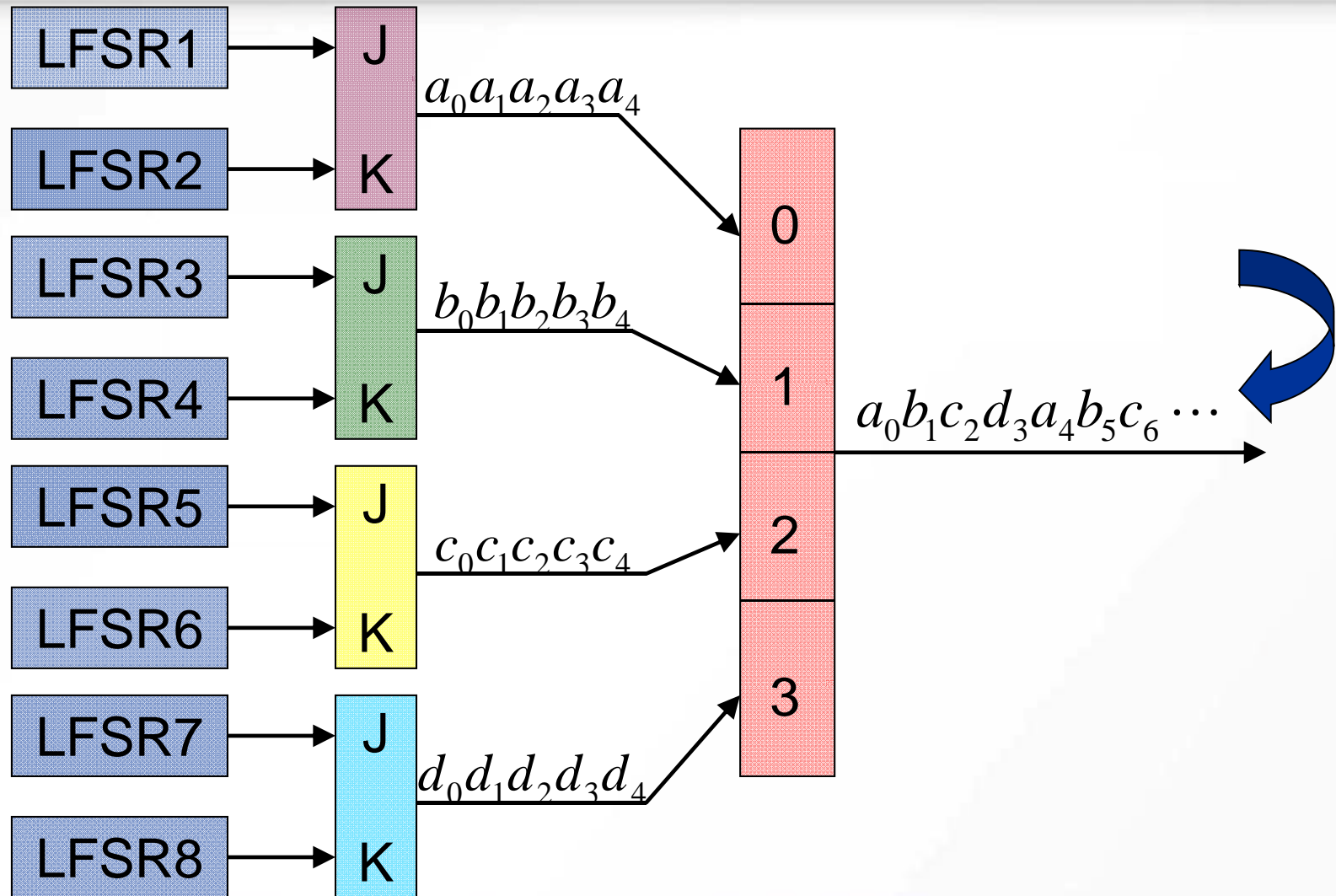
输入为0时，输出 C_{k-1} ；
输入为1时，输出0；



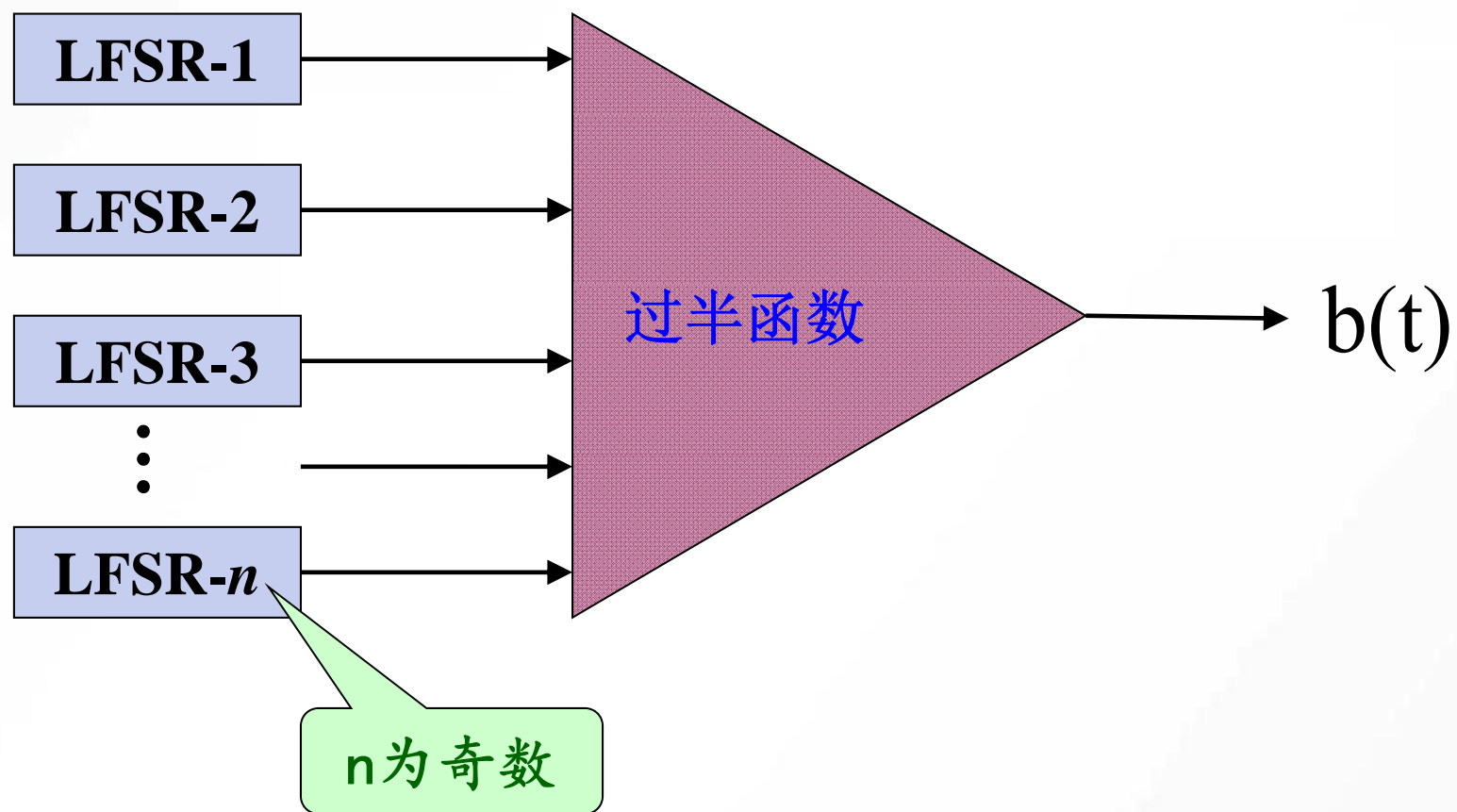
J-K触发器真值表

J	K	C_k
0	0	C_{k-1}
0	1	0
1	0	1
1	1	非 C_{k-1}

Pless生成器



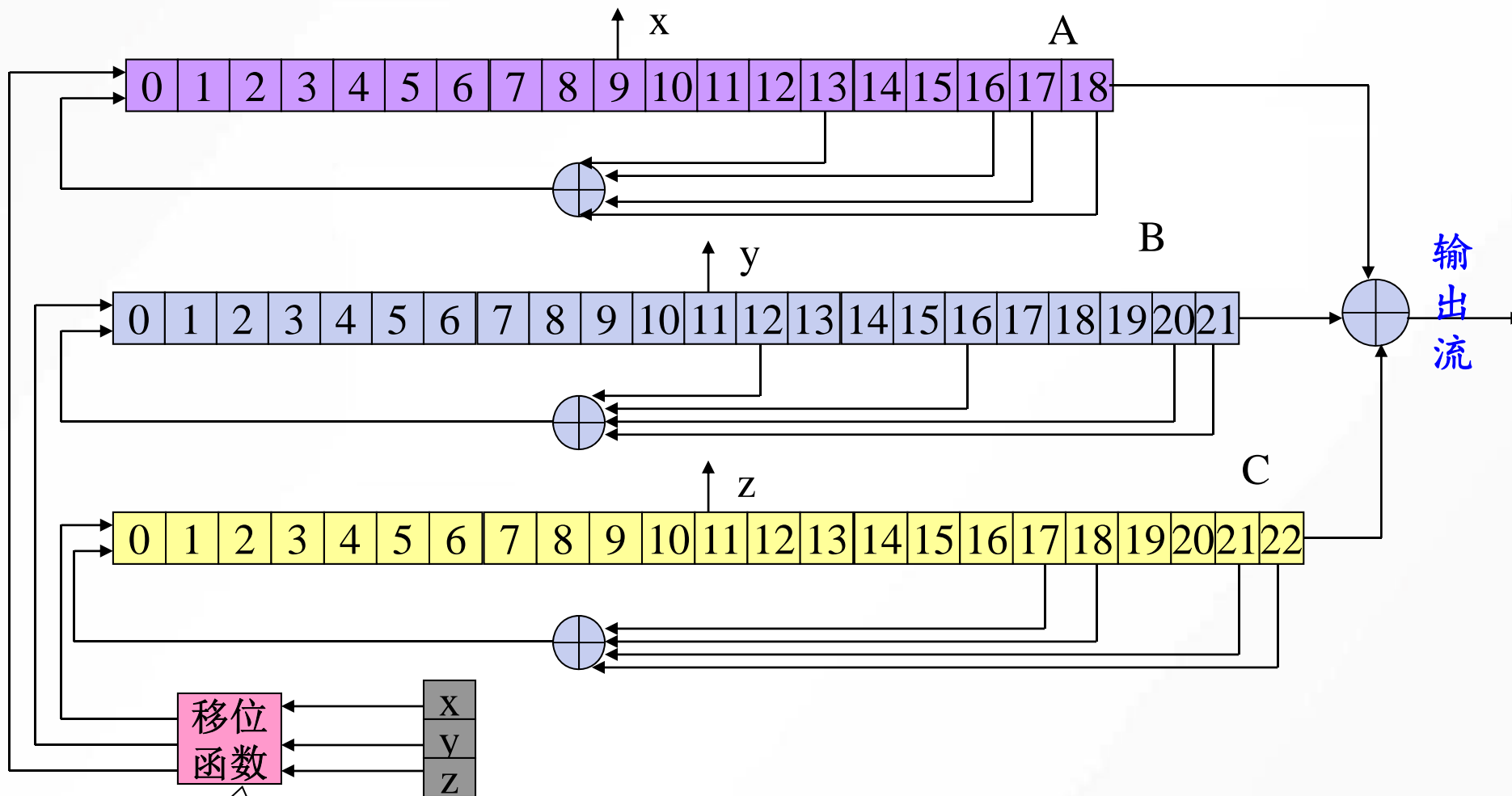
门限发生器



A5算法

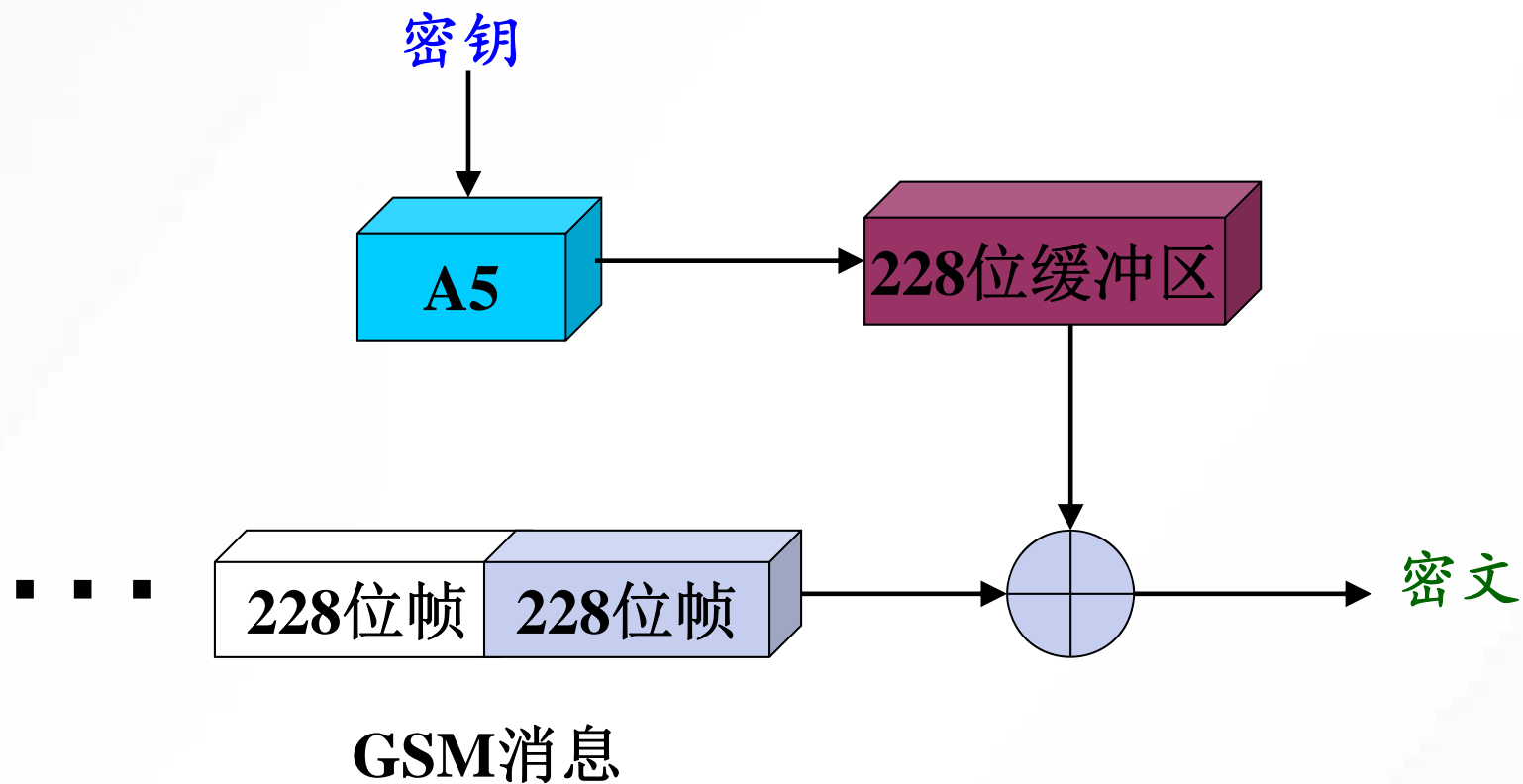
- **A5**算法已经被应用于**GSM**通信系统中，用于加密从手机到基站的连接，用来保护语音通信。一个**GSM**语言消息被转换成一系列的帧，每帧具有**228**位。每帧用**A5**进行加密。
- **A5**算法的主要组成部分是三个长度不同的线性移位寄存器，即**A**，**B**，**C**。其中**A**有**19**位，**B**有**22**位，**C**有**23**位。
- 移位是由时钟控制的，且遵循“择多”的原则。即从每个寄存器中取出一个中间位并进行判断，三个数中占多数的寄存器参加移位，其余的不移位。比如在取出的三个中间位中有两个为“1”，则为“1”的寄存器进行一次移位，而为“0”的不移。反过来，若三个中间位中有两个为“0”，则为“0”的寄存器进行一次移位，而为“1”的不移。

A5算法的示意图



遵循“多数原则”

GSM中使用A5流加密法



RC4的简介

- **RC4 (Ron Rivest Cipher)** 是**RSA**数据安全公司开发的可变密钥长度的序列密码，是世界上使用**最广泛的序列密码之一**，其优点是很容易用软件实现，加解密速度快（大约比**DES快10倍**）。
- **与基于移位寄存器的序列密码不同**，**RC4**密码是一种基于非线性数据表变换的序列密码。它以一个足够大的数据表为基础，对表进行非线性变换，产生非线性的密钥流序列。
- 一个**可变密钥长度**、面向**字节**操作的序列密码，该算法以**随机置换**作为基础。分析显示该密码的周期大于 **10^{100}** 。

RC4算法简介

- **RC4**的大小根据参数**n**的值而变化，通常**n=8**，这样**RC4**可以生成总共有**256** (2^8) 个元素的数据表**S**: S_0, S_1, \dots, S_{255} 。
- 种子密钥长度为从**1~256**个字节（即**8~2048**位）的可变长度密钥，用于初始化一个**256**个字节的**状态矢量S**。
- **RC4**有两个主要的算法：
密钥调度算法**KSA(Key-Scheduling Algorithm)**
伪随机数生成算法**PRGA (Pseudo Random-Generation Algorithm)**

RC算法的基本思想

- 依据种子密钥，利用密钥调度算法对数据表S进行重新排列。
- 利用伪随机数生成算法，从已重新排列的数据表S中取出一个字节。每取出一个字节，数据表S将发生变化。

RC算法

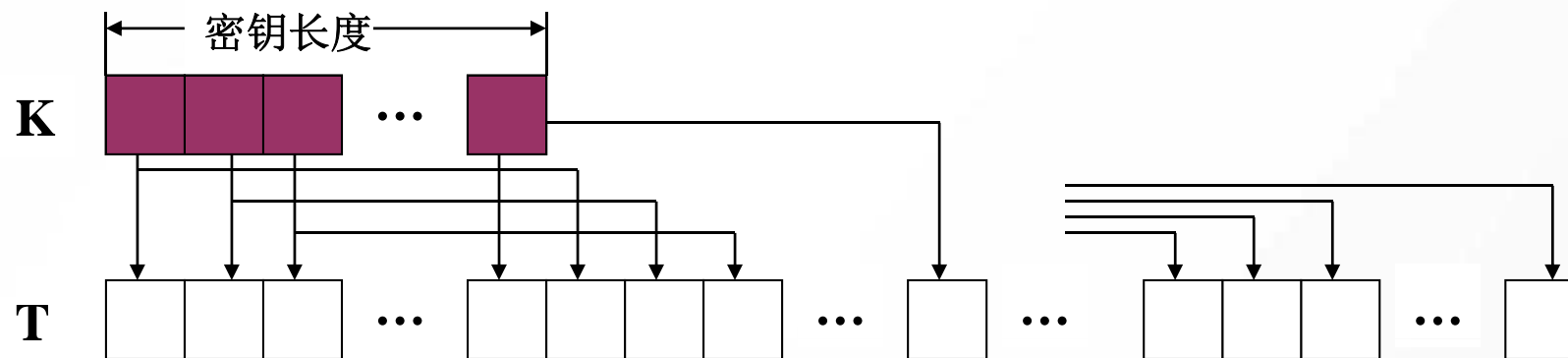
- 数据表**S**的初始状态
- 数据表**S**的初始置换
- 密钥流的生成

数据表S的初始状态

初始化S，即 $S(i) := i$ (一个字节), $0 \leq i \leq 255$ 。

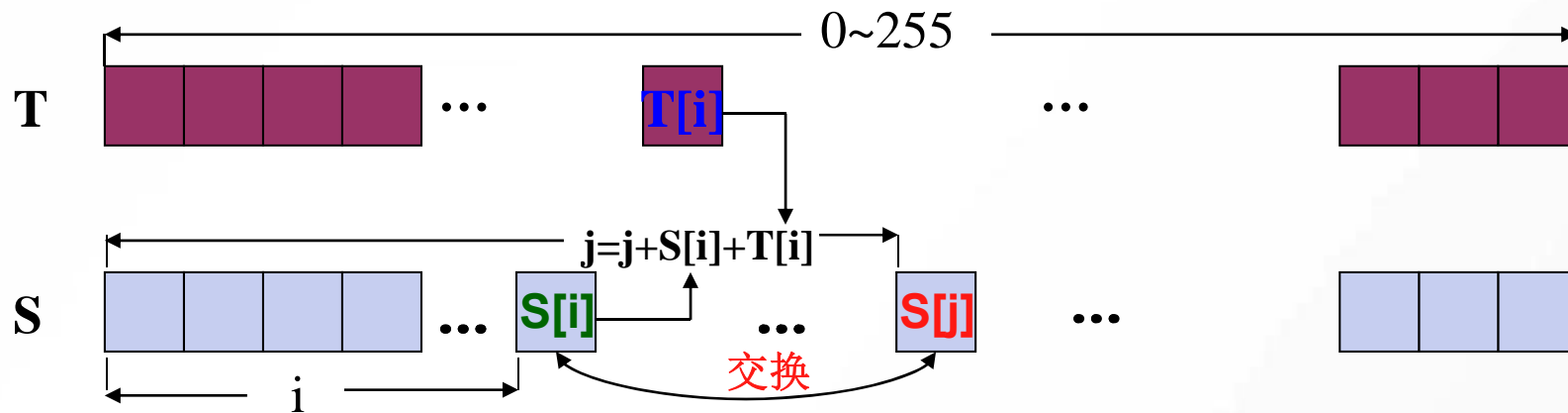


用主(种子)密钥K按字节填充另一个T表，
即 $T(i) := K(i \bmod \text{keylen})$, $0 \leq i \leq 255$ 。



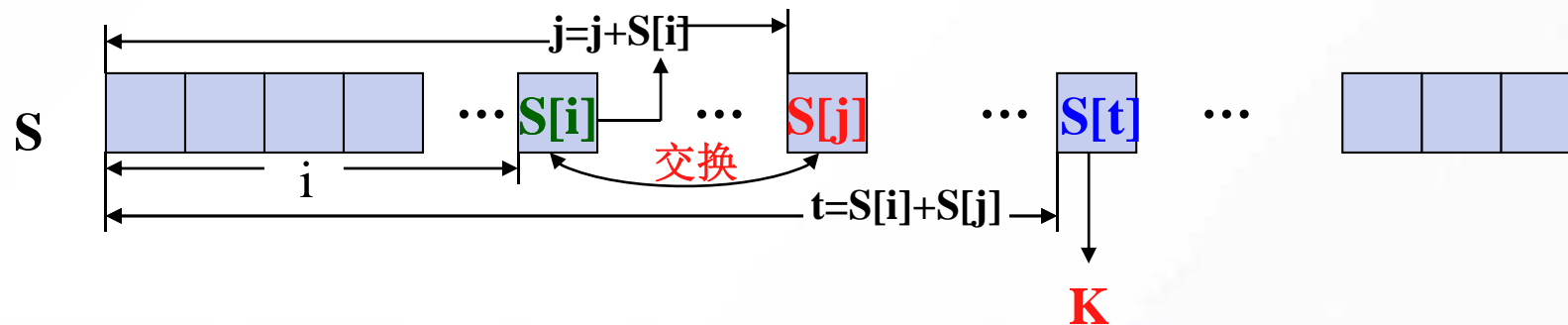
数据表S的初始置换

```
j:=0;  
for i := 0 to 255 do  
  begin  
    j := (j + S[i] + T[i]) (mod 256);  
    swap(S[i], S[j]); // 交换S(i)和S(j)的内容;  
  end
```



密钥流的生成

```
i,j:=0  
while (true)  
  begin  
    i:= i + 1 (mod 256);  
    j := j + S[i] (mod 256);  
    swap(S[i], S[j]);  
    t := S[i] + S[j] (mod 256);  
    k := S[t];  
  end
```



RC4算法说明

- 加密时，将**k**的值与明文字节异或；解密时，将**k**的值与密文字节异或。
- 为了保证安全强度，目前的**RC4**至少使用**128位密钥**。
- **RC4**算法可看成是一个有限状态自动机，把**S**表和*i,j*索引的具体取值称为**RC4**的一个状态： $T=(S_0, S_1, \dots, S_{255}, i, j)$ 。对状态**T**进行非线性变化，产生出新的状态，并输出密钥序列中的一个字节**k**。大约有 2^{1700} ($256! * 256^2$) 中可能状态，一个巨大的数字。
- 用大的数据表**S**和字长来实现这个思想是可能，即定义**16-位RC4**。

举例说明

假如使用3位（从0到7）的RC4，其操作是对8取模（而不是对256取模）。数据表S只有8个元素，初始化为：

S	0	1	2	3	4	5	6	7
	0	1	2	3	4	5	6	7

选取一个密钥，该密钥是由0到7的数以任意顺序组成的。例如选取5、6和7作为密钥。该密钥如下填入密钥数据表中：

K	5	6	7	5	6	7	5	6
	0	1	2	3	4	5	6	7

密钥调度算法KSA(举例)

然后利用如下循环构建实际的S数据表:

j:=0;

for i=0 to 7 do

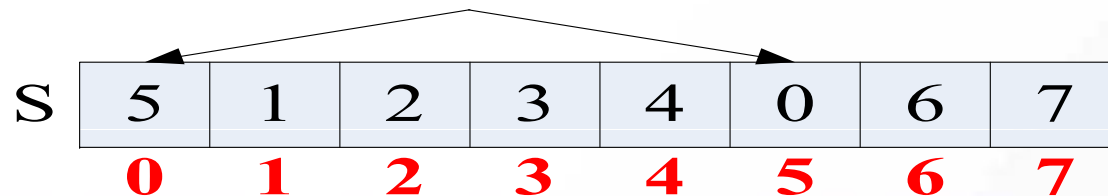
j:=(j+s(i)+k(i)) mod 8;

swap(S(i),S(j));

该循环以j=0和i=0开始。使用更新公式后j为:

$$j=(0+S(0)+K(0)) \bmod 8=5$$

因此，S数据表的第一个操作是将S(0)与S(5)互换。

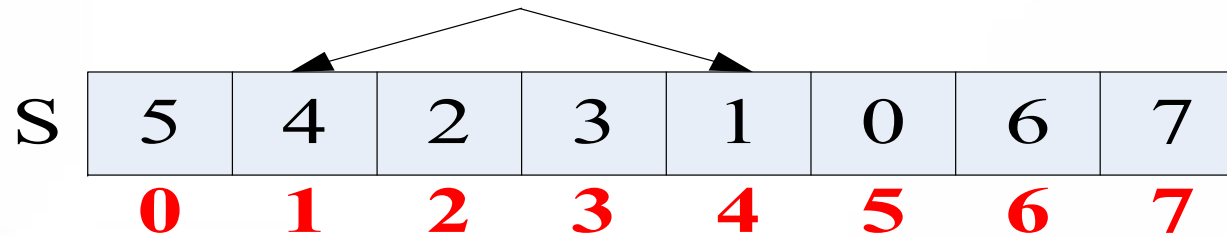


密钥调度算法KSA(举例)续

索引*i*加1后，*j*的下一个值为：

$$j=(5+S(1)+K(1)) \bmod 8=(5+1=6) \bmod 8=4$$

即将**S**数据表的**S(1)**和**S(4)**互换：



当该循环执行完后，数据表**S**就被随机化：



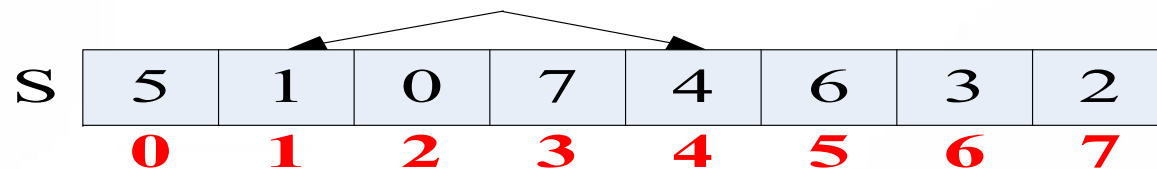
伪随机数生成算法PRGA（举例）

这样数据表**S**就可以用来生成随机的密钥流序列。从**j=0**和**i=0**开始，**RC4**如下计算第一个密钥字：

$$i=(i+1) \bmod 8=(0+1) \bmod 8 =1$$

$$j=(j+s(i)) \bmod 8=(0+s(1)) \bmod 8=(0+4) \bmod 8=4$$

swap **S(1)**和**S(4)**



然后如下计算**t**和**k**：

$$t=(S(j)+S(i)) \bmod 8=(S(4)+S(1)) \bmod 8=(1+4) \bmod 8=5$$

$$k=S(t)=S(5)=6$$

第一个密钥字为**6**，其二进制表示为**110**。反复进行该过程，直到生成的二进制的数量等于明文位的数量。

序列密码的优点（相对分组密码）

- 在硬件实施上，序列密码的速度一般要比分组密码快，而且不需要有很复杂的硬件电路。
- 在某些情况下（例如某些电信上的应用），当缓冲不足或必须对收到字符进行逐一处理时，序列密码就显得更加必要和恰当。
- 序列密码有较理想的数学分析。
- 序列密码能较好地隐藏明文的统计特性。

序列密码小结

- 安全强度取决于密钥序列的随机性和不可预测性
- 线性反馈移位寄存器理论上能够产生周期为 $2^{|n|}-1$ 的伪随机序列
- 不存在数据扩展和错误转播
- 实时性好，运算速度快，加密、解密易实现
- 密钥分配困难

本讲主要内容

- 序列密码的介绍
- 线性反馈移位寄存器
- 非线性序列
- 序列密码的算法举例 (**A5**、**RC4**)

谢谢！

