香浓证明了”一次一密”密码体制在理论上是不可破译的, 序列密码就是所寻求的方法之一. 序列密码是军事、外交等领域的主流密码体制

1 序列密码的加、解密过程

[1] 将报文、语音、图像、数据等原始明文转换为明文数据序列

[2] 将转换后的数据序列用密钥序列进行逐位加密生成密文数据序列并发送给接收者.

[3] 接收者使用相同的密钥序列对密文数据序列进行逐位解密恢复出明文序列.

2 序列密码的发展过程

- 序列密码不存在数据扩展和错误传播, 实时性比较好, 加解密实现容易
- Vernam 密码最早的二进制序列密码序列密码系统. 当 Vernam 密码中的密钥序列为完全随机的二进制序列时, 它就是”一次一密”密码. 但其密钥产生分配和管理都极为困难, 故未得到广泛运用.
- 随着微电子技术和数学理论的发展, 基于伪随机序列的序列密码称为当前最通用的密码系统. 这种序列密码中, 加、解密的密钥序列都是伪随机序列.
- 伪随机序列是由密钥流产生器产生的. 密钥流产生器实际上就是通过给定算法产生通常是 $0 \sim 1$ 数据流的密钥流.

3 伪随机序列

只要求截获到比周期更短的一段密文时不至于泄露更多的信息, 这样的序列称为伪随机序列

序列密码的安全保密性主要依赖于密钥序列, 因此研究什么样的伪随机序列可以得到作为序列密码的密钥序列称为序列密码研究中的主要问题. 即序列密码的关键是产生密钥序列的算法.

3.1 序列密码的一般原理

假设序列密码中

- M 为明文空间是有可能的二进制数字序列组成的集合
- K 为密钥空间, $k \in K$ 为密钥序列产生算法 A 产生长密钥序列的一个短序列

  序列密码的成败取决于算法 A 的保密程度以及复杂度, 各国的核心密码都不公布算法 A. 序列密码的加、解密过程
- 对于每一个短密钥 $k \in K$, 由算法 A 确定一个二进制序列 $A(k) = k_1k_2\cdots$.
- 加密时, 当明文 $m \in M, m = m_1m_2\cdots m_n$ 时, 对 $i = 1, 2, \cdots, n$, 计算 $c_i = m \oplus k_i$, 则密文为 $c = E(m, k) = c_1c_2\cdots c_n$
- 解密时, 对 $i = 1, 2, \cdots, n$, 计算 $m_i = c_i \oplus k_i$, 恢复出明文.

  通常称密钥 k 为种子密钥, 由 k 通过算法 A 产生的 $A(k) = k_1k_2\cdots$ 序列称为密钥序列

  序列密码的安全性主要依赖于密钥序列 $A(k) = k_1k_2\cdots$, 当 $k_1k_2\cdots$ 是离散无记忆的二进制均匀分布信源产生的随机序列时, 则该密码系统是一次一密系统. 但实际上 $A(k)$ 是由 k 通过确定性算法产生的伪随机序列, 该系统不是完全保密的

  设计序列密码系统的关键是设计密钥序列 $A(k)$; 破译序列密码也只需要求出所使用的 $A(k)$ 序列密码系统中密钥序列设计应该考虑如下几个因素:

  [1] 系统的安全保密性

  [2] 密钥易于分配、保管、更换;

  [3] 产生密钥序列简单、快速

  目前, 密钥序列产生大多数是基于移位寄存器. 为达到安全保密性要求, 序列密码的密钥序列应该满足伪随机准则:

  [1] 极大的周期: 现代密码机的数据率为 $10^8 bit/s$, 如果 10 年内不使用重复的 $\{k_i\}$, 要求 $\{k_i\}$ 的周期 $> 3 \times 10^{16}$ 或者 2^{55}.

– 周期长, 是为了不至于使得通过两组密文相加的结果和语言冗余度分析就能获得一些关于明文的信息

[2] 良好的随即统计特性, 即序列中每位接近均匀分布

– 良好的随机特性是为了是密钥序列能很好地掩盖明文, 以抵抗" 已知明文攻击"

[3] 序列线性不可预测性充分大

– 线性不可预测性是为了防止从部分密钥序列通过线性关系简单的推导出整个密钥序列的测度.

4 群的概念

定义 4.1 给定一集合 $G = \{a, b, \cdots\}$ 和该集合上的运算 $*$, 满足下列四个条件的代数系统 $< G, * >$ 称为群:

- 封闭性: 若 $a, b \in G$, 则存在 $c \in C$, 使得 $a * b = c$
- 结合律成立: $a, b, c \in G$, 恒有 $(a * b) * c = a * (b * c)$
- 存在单位元素 e: 即存在 $e \in G$, 对于 $\forall a \in G$, 恒有 $a * e = e * a = a$

  存在逆元: 对于 $a \in G$, 恒有 $b \in G$, 使得 $a * b = b * a = e$, 元素 b 称为 a 的逆元, 用 a^{-1} 来表示, 即 $b = a^{-1}$.

  若 $\forall a, b \in G$, 有 $a * b = b * a$, 则称 $< G, * >$ 为阿贝尔群, 为简便起见, 简记作 $a * b$ 为 ab

5 域的概念

定义 5.1 F 是至少含有两个元素的集合, 对 F 定义了两种运算"+"和"*", 并且满足以下三个条件的代数系统 $< F, +, * >$ 称为域

- F 的元素关于运算"+" 构成阿贝尔群, 设单位元为 0
- $F/\{0\}$ 关于运算"*" 构成阿贝尔群.
- 对于 $a, b, c \in F$ 分配律成立. 即

$$(a + b) * c = a * c + b * c$$
$$c * (a + b) = c * a + c * b$$

若 F 域的元素有限个, 则称之为有限域或者伽罗瓦域. $F/\{0\}$ 表示集合 F 除去元素 $\{0\}$ 后的元素.

p 是素数, 则 $F=\{0,1,2\cdots,p-1\}$ 在 $\bmod p$ 的意义下关于"+" 和"*" 运算构成的域用 $GF(p)$ 来表示.

6 线性移位寄存器

移位寄存器是序列密码种产生密钥序列的一个主要组成部分. $GF(2)$ 上 n 即反馈移位寄存器的表示见下图

- 标有 $a_1,a_2,a_3,\cdots,a_{n-1},\ a_n$ 的小方框表示 $(0,1)$ 二值存储单元, 信号流从左向右. 这 n 个二值存储单元称为该反馈移位寄存器的级.
- 在任意时刻, 者 n 级的内容构成该反馈移位寄存器的状态, 即反馈移位寄存器的状态对应于一个 $GF(2)$ 上的 n 为向量, 共有 2^n 种可能的情况.
- 每一时刻的状态可用 n 长序列 $a_1,a_2,\cdots a_n$, 或者 n 维向量 $f(a_1,a_2,\cdots,a_n)$ 表示. 其中 a_i 为当时第 i 级存储器的内容.
- 在主时钟确定的周期区间上, 每一级存储器 a_i 都将其内容向下一级 a_{i-1} 传递, 并根据存储器当时的状态计算 $f(a_1,a_2,\cdots,a_n)$ 作为 a_n 下一时间的内容.
- 称函数 $f(a_1,a_2,\cdots,a_n)$ 为反馈函数, 它是 n 元布尔函数, 即 n 个变元 $a_1,a_2,\cdots,a_n$ 可以独立的取 0 和 1 这两个可能的值. 对 n 个变元 $a_1,a_2,\cdots,a_n$ 作与、或、取反等运算. 最后函数值也为 0 或 1 的函数. 这样的反馈函数共有 2^{2^n}.
- 在时钟脉冲时, 如果反馈移位寄存器的状态为:

$$S_t=(a_t,t_{t+1},\cdots,a_{t+n-1})$$

则

$$a_{t+n}=f(a_t,a_{t+1},\cdots,a_{t+n-1}) \tag{1}$$

a_{t+n} 是移位寄存器的输入. 在 a_{t+n} 的驱动下, 移位寄存器的各个数据向前推移一位, 使得状态变为

$$S_{t+1}=(a_{t+1},a_{t+2},\cdots,a_{t+n}) \tag{2}$$

同时整个移位寄存器的输出为 a_t, 由此得到一系列数据

$$a_1, a_2, \cdots, a_n, \cdots \tag{3}$$

满足关系式 (1), 称无穷序列 (2) 为一个反馈移位寄存器序列

定义 6.1 *序列 $a_1, a_2, \cdots, a_n, \cdots$ 成为周期序列, 若存在正整数 T 使得*

$$a_{i+T} = a_i, i = 1, 2, \cdots$$

满足 (3) 式的最小正整数 T 称为序列 a_i 的周期.

若移位寄存器的反馈函数$f(a_1, a_2, \cdots, a_n)$ 是 $a_1, a_2, \cdots, a_n$ 的线性函数, 则称为线性移位寄存器, 否则称为非线性移位寄存器

设 $f(a_1, a_2, \cdots, a_n)$ 为线性函数, 则 f 可以写成

$$f(a_1, a_2, \cdots, a_n) = C_n a_1 \oplus C_{n-1} a_2 \oplus \cdots \oplus C_1 a_n$$

其中 $C_i = 0$ 或 1, $C_1, C_2, \cdots, C_n$ 为反馈系数. 二进制下, $C_1, C_2, \cdots, C_n$ 的作用相当于一个开关, 用断开和闭合表示 0 和 1, 这样的线性函数共 2^n 个.

输出序列 a_t 满足

$$a_{n+1} = C_n a_t \oplus C_{n-1} a_{t+1} \oplus \cdots \oplus C_1 a_{n+t-1}$$

其中 t 为非负正整数

例题: 有一个三级移位寄存器如下图所示

其中初态为 $S_1 = (a_1, a_2, a_3) = (1, 0, 1)$

解:

$S_1 = (a_1, a_2, a_3) = (1, 0, 1)$, 输出为$a_1 = 1$

$S_2 = (a_2, a_3, a_4) = (0, 1, 1), a_4 = a_1 a_2 \oplus a_3 = (1 \times 0) \oplus 1 = 1$, 输出$a_2 = 0$

$S_3 = (a_3, a_4, a_5) = (1, 1, 1), a_5 = a_2 a_3 \oplus a_4 = (0 \times 1) \oplus 1 = 1$, 输出$a_3 = 1$

$S_4 = (a_4, a_5, a_6) = (1, 1, 0), a_6 = a_3 a_4 \oplus a_5 = (1 \times 1) \oplus 1 = 0$, 输出$a_4 = 1$

$S_5 = (a_5, a_6, a_7) = (1, 0, 1), a_7 = a_4 a_5 \oplus a_6 = (1 \times 1) \oplus 0 = 1$, 输出$a_5 = 1$

$S_6 = (a_6, a_7, a_8) = (0, 1, 1), a_8 = a_5 a_6 \oplus a_7 = (1 \times 0) \oplus 1 = 1$, 输出$a_6 = 0$

$S_7 = (a_7, a_8, a_9) = (1, 1, 1), a_9 = a_6 a_7 \oplus a_8 = (0 \times 1) \oplus 1 = 1$, 输出$a_7 = 1$

如下表所示 所以此移位寄存器输出序列为 $101110111011\cdots$, 是周期为

状态 (a_1, a_2, a_3)				输出
S_1	1	0	1	1
S_2	0	1	1	0
S_3	1	1	1	1
S_4	1	1	0	1
S_5	1	0	1	1
S_6	0	1	1	0
S_7	1	1	1	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

4 的序列.

例题: 有一个五级线性移位寄存器

其中 $S_1 = (a_1, a_2, a_3, a_4, a_5) = (10011)$

解: $S_1 = (a_1, a_2, a_3, a_4, a_5) = (10011)_2$, 输出为 $a_1 = 1$ $S_2 = (a_2, a_3, a_4, a_5, a_6)$, 其中 $a_6 = f(a_1, a_2, a_3, a_4, a_5)$. 对于五级线性移位寄存器应有

$$f(a_1, a_2, a_3, a_4, a_5) = C_5a_1 \oplus C_4a_2 \oplus C_3a_3 \oplus C_2a_4 \oplus C_1a_5$$

由给出的图可知

$$C_2 = C_3 = C_5 = 0, C_1 = C_4 = 1$$

即

$f(a_1, a_2, a_3, a_4, a_5) = a_1 \oplus a_4$

$a_6 = a_1 \oplus a_4 = 1 \oplus 1 = 0, S_2 = (a_2, a_3, a_4, a_5, a_6) = (00110)$, 输出为$a_2 = 0$

$S_3 = (a_3, a_4, a_5, a_6, a_7) = (01101), a_7 = a_2 \oplus a_5 = 0 \oplus 1 = 1$, 输出为$a_3 = 0$

$S_4 = (a_4, a_5, a_6, a_7, a_8) = (11010), a_8 = a_3 \oplus a_6 = 0 \oplus 0 = 0$, 输出为$a_4 = 1$

$S_5 = (a_5, a_6, a_7, a_8, a_9) = (10100), a_9 = a_4 \oplus a_7 = 1 \oplus 1 = 0$, 输出为$a_5 = 1$

$S_6 = (a_6, a_7, a_8, a_9, a_7) = (01001), a_10 = a_5 \oplus a_8 = 1 \oplus 0 = 1$, 输出为$a_6 = 0$

$S_7 = (a_7, a_8, a_9, a_10, a_11) = (10010), a_11 = a_6 \oplus a_9 = 0 \oplus 0 = 0$, 输出为$a_7 = 1$

$S_8 = (a_8, a_9, a_10, a_11, a_12) = (00100), a_12 = a_7 \oplus a_10 = 1 \oplus 1 = 0$, 输出为$a_8 = 0$

$S_9 = (a_9, a_10, a_11, a_12, a_13) = (01000), a_13 = a_8 \oplus a_11 = 0 \oplus 0 = 0$, 输出为$a_9 = 0$

$\vdots$

在线性移位寄存器种总是假定 $c_1, c_2, \cdots, c_n$ 种至少有一个系数不为 0. 若只有一个系数不为 0, 设仅有 a_j 项非 0, 实际上是一种延迟装置, 一般对于 n 级线性移位寄存器, 总假定 $c_n = 1$

线性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; 若其初始状态为非 0, 则其后继状态不会为 0
- n 级线性移位寄存器的状态周期 $\leq 2^n - 1$, 其输出序列的周期 = 状态周期 $\leq 2^n - 1$;
- 选择合适的反馈函数可是序列的周期达到最大值 $2^n - 1$, 则称此时的输出序列为最大长度线性移位寄存器, 简称 m 序列

7 线性移位寄存器的一元多项式表示

设 n 级线性移位寄存器的输出序列 a_i 满足递推关系

$$a_{k+n} = C_1 a_{k+n-1} \oplus C_2 a_{k+n-2} \oplus \cdots \oplus C_n a_k \tag{4}$$

对于任何 $k \geq 1$ 成立. 这种递推关系可用一个一元 n 次多项式表示

$$p(x) = 1 + C_1 x + C_2 x^2 + \cdots + C_n x^n \tag{5}$$

称 (5) 式为该线性移位寄存器的联系多项式或特征多项式

设 n 级线性移位寄存器对应于递推关系, 则有 2^n 个递推序列, 其中非恒为 0 的序列有 $2^n - 1$ 个. 令这非零的序列全体为 $G[P(x)]$, 对 $G[P(x)]$ 种任一序列 a_j, 有母函数

$$A(x) = \sum_{i=1}^{\infty} a_i x^{i-1}$$

.

定理 7.1 设 $p(x)1 + c_1 x + c_2 x^2 + \cdots + c_n x^n$ 是 $GF(2)$ 上的多项式, 且递推序列 $\{\}a_i \in G[P(x)]$, 令

$$A(x) = \sum_{i=1}^{\infty} a_i x^{i-1}$$

则

$$A(x) = \frac{\phi(x)}{p(x)}$$

其中

$$\phi(x) = \sum_{i=1}^{n} c_{n-i} x^{n-i} \sum_{j=1}^{i} a_j x^{j-1}$$

根据上述定理, 若序列 $\{a_t\} \in G[p_n(x)]$, 其中 $p_n(x)$ 是 n 级线性移位寄存器的特征多项式, 则母函数为 $A(x) = \frac{\phi(x)}{p(x)}$, 其中的次数低于 n, 最多为 $n-1$ 次

定理 7.2 $p(x)|q()$ 的充分必要条件是 $G[p(x)] \subset G[q(x)]$.

上述定理说明 n 级线性移位寄存器产生的序列可用级数更多的的线性移位寄存器来实现.

定义 7.1 设 $p(x)$ 为 $GF(2)$ 上的 n 次多项式, 使得 $p(x)|x^p - 1$ 的最小 p 称为 $p(x)$ 的周期或 $p(x)$ 的阶.

定理 7.3 设 $p(x)$ 为 $GF(2)$ 上的 n 次多项式, 且 $p(x)$ 是序列 $\{a_i\}$ 的特征多项式, p 为 $p(x)$ 的阶, 则 $\{a_i\}$ 的周期为 $r|p$

n 级线性移位寄存器输出序列的周期 r 不依赖于初始条件, 而依赖于特征多项式 $p(x)$.

定理 7.4 若 $p(x)$ 是 n 次不可约多项式, 且 $p(x)$ 的阶为 m, $\{a_i\} \in G[p(x)]$ 则序列 $\{a_i\}$ 的周期为 m.

上述定理说明了特征多项式满足什么条件, n 级线性移位寄存器的输出序列为 m 序列.

定理 7.5 n 级线性移位寄存器产生的状态序列最大周期 $2^n - 1$ 的必要条件是其特征多项式是不可约的.

上述定理不一定成立

定义 7.2 $p(x)$ 为 n 次不可约多项式, 若 $p(x)$ 的阶位 $2^n - 1$, 称 $p(x)$ 为 n 次本原多项式

定理 7.6 设 $\{a_i\} \in G[p(x)]$, 则 $\{a_i\}$ 为 m 序列的充要条件是 $p(x)$ 为 n 次本原多项式.

8 m 序列的伪随机性

8.1 随机序列的一般特性

设序列 $\{a_i\} = (a_1 a_2 a_3 \cdots)$ 为序列 $0-1$, 称序列种形式为 $0\underbrace{11\cdots1}_{\text{k 个 1}}0$ 为一个长为 k 的 1 游程, $1\underbrace{00\cdots0}_{\text{k 个 0}}1$ 的一段为一个长为 k 的 0 游程.

定义 8.1 *$GF(2)$ 上周期为 T 的序列 $\{a_i\}$ 的自相关函数定义为*

$$R_\alpha(\tau) = \frac{1}{T}\sum_{k=1}^{T}(-1)^{a_k}(-1)^{a_{k+\tau}}, 0 \le \tau \le T-1$$

- 周期序列 $\{a_i\}$ 的自相关函数表示序列 $\{a_i\}$ 与 $\{a_{i+\tau}\}$ 在一个周期内对应位相同的位数与对应位相异的位数之差的一个参数 ($\frac{\text{相同位的数目}-\text{相异位的数目}}{T}$).
- 当 $\tau = 0$, $R_\alpha(\tau) = 1$; 当 $\tau \neq 0$ 时, $R_\alpha(\tau)$ 称为异相自相关函数. 异相自相关函数是序列随机性的一个指标.

  伪随机周期序列应该满足以下三个随机性公设:

  [1]: 在序列的一个周期内, 0 与 1 的个数相差至多为 1.

  [2]: 在序列的一个周期内, 长为 1 的游程数占游程总数的 $\frac{1}{2}$, 长为 2 的游程总数占游程总数的 $\frac{1}{2^2}$, 长为 i 的游程数占游程总数的 $\frac{1}{2^i}$, 且在等长的游程中 0 的游程个数和 1 的游程个数相等.

  [3]: 异相自相关函数是一个常数.

- 公设 [1] 说明 $0-1$ 序列中 0 与 1 的出现概率" 基本" 相等.
- 公设 [2] 说明 0 与 1 在 n 个位置上出现的概率相同.
- 公设 [3] 说明若将 $\{a_i\}$ 与 $\{a_{i+\tau}\}$ 比较, 无法得到关于 $\{a_i\}$ 的实质性信息.

例题 7.3 假设破译者得到密文串 101101011110010 和相应的明文串 011001111111001. 同时假定攻击者也知道密钥流是使用 5 级线性移位寄存器产生的, 试破译改密码系统. 解: 有明文 (15 位), 密文 (15 位) 对可求出长为 15 位的密钥序列

有开始的 10^i 密钥流比特得到上述矩阵方程

m_i	0	1	1	0	0	1	1	1	1	1	1	1	0	0	1
c_i	1	0	1	1	0	1	0	1	1	1	1	0	0	1	0
$k_i = m_i \oplus c$	1	1	0	1	0	0	1	0	0	0	0	1	0	1	1