伪随机稀疏序列的研究

胡予濮,杨 波
(西安电子科技大学 ISN 国家重点实验室,信息安全研究所,陕西西安 710071)

摘 要: 伪随机稀疏序列在信息隐藏技术中有广泛的应用. 本文给出了几类伪随机稀疏序列,它们分别是:乘积序列,自缩乘积序列,自扩序列,其于 $GF(2^m)$ 上 m 序列的稀疏序列,和基于乘方剩余符号的稀疏序列. 讨论了它们的良好伪随机性,主要是最小周期和线性复杂度.

关键词: 流密码; 信息隐藏; m 序列; 最小周期; 线性复杂度
中图分类号: TN918.1 **文献标识码**: A **文章编号**: 0372-2112 (2002) 01-0142-03

Research on Pseudo Sparse-Sequences

HU Yu-pu, YANG Bo
(*ISPI, ISN National Key Lab., Xidian University, Xi'an, Shaanxi* 710071)

Abstract: Pseudo sparse-sequence has wide applications in information hiding technology. This paper presents several classes of pseudo sparse-sequences, named respectively product sequences, self-shrinking product sequences, self-expanding sequences, sparse sequences based on $GF(2^m)$, and sparse sequences based on power-residue-symbol. Their randomness is discussed, mainly on the least periods and linear complexities.

Key words: stream cipher; information hiding; m-sequence; the least period; linear complexity

1 引言

在本文中将首次提出伪随机稀疏序列的概念.

定义 1 设 $a = a_0 a_1 a_2 \cdots$ 是 $GF(2)$ 上的序列,最小周期为 P,在一个最小周期内 1 的个数为 M,称 $D = M/P$ 是序列 a 的密度. 若 $D \ll 1/2$,则称 a 为稀疏序列.

伪随机稀疏序列指的是这样的稀疏序列,它满足一切伪随机特性. 伪随机稀疏序列当然不能用来构造密钥流,但它在数字信号处理,特别是在信息隐藏技术[1]中有广泛的应用. 当需要将秘密信息(身份标记,版权标记,数字水印等)密写埋藏于一段有意义的消息(cover)中时,埋藏地点应该是伪随机地稀疏分布于 cover 之内. 稀疏序列的伪随机性仍然包括极大的周期、高线性复杂度、良好的游程分布等.

一个最明显也最平凡的结果是:若稀疏序列的密度为 D,最长 0 游程长度为 R,线性复杂度为 L,则必有 $L \geq R + 1 \geq 1/D$. 但需要指出的是在应用于信息隐藏技术中时,稀疏序列 a 的线性复杂度应该远远大于 $1/D$. 这是由于:(1)攻击者发现一两个埋藏地点是不太困难的,而此时如果用稀疏序列 a 来分布埋藏地点,且 a 的线性复杂度接近于 $1/D$,被动攻击者(他们试图察觉或识别秘密信息)就有可能用 B-M 算法[2]破解秘密信息的所有埋藏地点;(2)为了抵抗主动攻击者(他们试图破坏或伪造秘密信息),需要使隐藏的信息具有稳健性(robustness),在各埋藏地点所埋藏的内容有相当的重复[1],当改变一两个埋藏地点的消息时,不至于使隐藏的信息受到显著破坏.

2 基于 GF(2) 上 m 序列的稀疏序列

2.1 乘积序列

定义 2 对于 $GF(2)$ 上的 n 级 m 序列 $a = a_0 a_1 a_2 \cdots$,取 $(j_1, j_2, \cdots, j_k)$, $0 \leq j_1 < j_2 < \cdots < j_k \leq n-1$. 取乘积 $b_t = a_{t+j_1} a_{t+j_2} \cdots a_{t+j_k}$, $t = 0, 1, 2, \cdots$: 称序列 $b = b_0 b_1 b_2 \cdots$ 为基于 m 序列 a 的乘积序列.

证明:乘积序列的密度 $D = 2^{n-k}/(2^n - 1)$,最小周期为 $P = 2^n - 1$. 还可以由文献[3] pp49 - 69 的分析方法,证明一些特殊情形的结论,比如当 $k = 2$ 时 $L \geq C_n^2$. 但得不到其线性复杂度 L 一般的具体结果,只知道平凡的结果 $L \geq \max(n+1, 2^k)$.

2.2 自缩乘积序列

定义 3 设有 $GF(2)$ 上的 n 级 m 序列 $a = a_0 a_1 a_2 \cdots$: 取集合 $\{1, 2, \cdots, n-1\}$. 对于子集 $J \subset \{1, 2, \cdots, n-1\}$,若 $a_t = 1$,则输出乘积 $\prod_{j \in J} a_{t+j}$(当子集 J 是空集时,定义 $\prod_{j \in J} a_{t+j} = 1$);否则放弃输出; $t = 0, 1, 2, \cdots$: 如此得到输出序列 $b_0 b_1 b_2 \cdots$,记为 $b(J)$,称其为基于 m 序列 a 的自缩乘积序列.

证明 自缩乘积序列的密度 $D = 2^{n-1-k}/(2^n - 1) = 2^{-k}$,其中 $k = |J|$ 为 J 内的元素个数;其最小周期总为 2^{n-1} 的因子. 注意:当序列的最小周期为 2^l 时其线性复杂度总大于

收稿日期:2000-08-18;修回日期:2001-04-10
基金项目:国防科技重点实验室基金(No. 99JS06.3.1.DZ0112)

2^{l-1}[3].因此关键的问题是自缩乘积序列的最小周期应该是2^{n-1}的大因子,最好能达到2^{n-1}.

定理 1　取自缩乘积序列的全体所成的集合$\{b(J), J \subset \{1,2,\cdots,n-1\}\}$.则有不多于一半的自缩乘积序列,其最小周期小于$2^{n-1}$.

证明　取J^c为J的补集.首先$b(\{1,2,\cdots,n-1\})$的最小周期为2^{n-1},这是因为$b(\{1,2,\cdots,n-1\})$在连续2^{n-1}个比特中只有一个1.其次$b(\{1,2,\cdots,n-1\}) = b(J \cup J^c) = b(J)\,b(J^c)$为乘积序列.如果序列$b(J)$和$b(J^c)$的最小周期都是$2^{n-1}$的真因子,则$b(\{1,2,\cdots,n-1\})$的最小周期也必是$2^{n-1}$的真因子,矛盾.矛盾说明序列$b(J)$和$b(J^c)$的最小周期至少有一个是$2^{n-1}$.定理1得证.

引理 1　记$B(k) = \{b(J), J\{1,2,\cdots,n-1\}, b(J)$的最小周期小于$2^{n-k}\}$;记$J_k$满足:$b(J_k) \in B(k)$,且$|J_k|$达到最大.则①$B(k) \supset B(k+1)$;②对任意$b(J) \in B(k), J \neq J_k$时都有真包含关系$J \subset J_k$;③若$b(J) \in B(k)$,则$|J| \leq n-k-1$.

证明　①是平凡的.②的证明只须用反证法:若真包含关系$J \subset J_k$不成立,且$J \neq J_k$,则$|J_k| < |J_k \cup J|$,且不难看出$b(J_k \cup J) \in B(k)$,这与最大的假设相矛盾.矛盾说明②成立.以下证③.若$|J| > n-k-1$,则由定义3可知,$b(J)$在其连续2^{n-1}个比特中1的个数不多于2^{k-1},故$b(J)$的最小周期不会小于2^{n-k},矛盾.矛盾说明③成立.引理1得证.

定理 2　$\{b(J), J \subset \{1,2,\cdots,n-1\}\}$中,最小周期小于$2^{n-k}$的自缩乘积序列所占的比例不多于$2^{-k}$.

证明　$\{b(J), J \subset \{1,2,\cdots,n-1\}\}$中共有$2^{n-1}$个序列.由引理1知$|J_k| \leq n-k-1$,因此$|B(k)| \leq 2^{n-k-1}$,$|B(k)|/2^{n-1} \leq 2^{-k}$.定理2得证.

推论 1　从$\{1,2,\cdots,n-1\}$中随机地取出m个元素组成子集J,$b(J)$的最小周期不小于2^l的概率记为$\Pr(m,l)$.则当$m \geq l$时$\Pr(m,l) = 1$;当$m < l$时$\Pr(m,l) \geq (C_{n-1}^m - C_{l-1}^m)/C_{n-1}^m$.

证明　由引理1和定理2得$|J_{n-l}| \leq l-1$.因此当$l \leq m$时必有$b(J) \notin B(n-l)$.以下设$l > m$.则$B(n-l)$中有不多于C_{n-1}^m个序列具有密度$D = 2^{-m}$.推论1得证.

2.3　自扩序列

定义 4　对于$GF(2)$上的n级m序列$a = a_0 a_1 a_2 \cdots$,取$(j_1, j_2, \cdots, j_k)$,$1 \leq j_1 < j_2 < \cdots < j_k \leq n-1$.对每个$t = 0,1,2,\cdots$,若$(a_{t+j_1}, a_{t+j_2}, \cdots, a_{t+j_k})$中有$l$个0,则输出$a_t$,然后连续输出$l$个0.如此得到输出序列$b = b_0 b_1 b_2 \cdots$,称其为基于$m$序列$a$的自扩序列.

定理 3　n级m序列a的自扩序列b的最小周期为$P = k(2^{n-1}-1) + (2^{n-1})$,其密度为$D = (2^{n-1})/(k(2^{n-1}-1) + (2^{n-1})) \approx 1/(k+2)$,其线性复杂度$L > n-1+k(k+3)/2$.

证明　首先考虑集合$\{t; t = 0,1,2,\cdots,2^n-2\}$.该集合中恰有$C_k^l 2^{n-k}$个$t$使得$(a_{t+j_1}, a_{t+j_2}, \cdots, a_{t+j_k})$中恰有$l$个0,$l = 0,1,\cdots,k-1$;恰有$C_k^l 2^{n-k}-1$个$t$使得$(a_{t+j_1}, a_{t+j_2}, \cdots, a_{t+j_k})$中全为0.因此当序列$a$走过$2^n-1$个比特时,$a$的自扩序列$b$走过$k(2^{n-1}-1) + (2^n-1)$个比特.其次我们指出在序列$b$的连续$k(2^{n-1}-1) + (2^n-1)$个比特中,有且仅有一个最长的0游程,这个0游程出现在当$(a_t, a_{t+1}, a_{t+2}, \cdots, a_{t+n-1}) = (1,0,0,\cdots,0)$时所对应的序列$b$的0游程.以下只须证明这个0游程的长度不小于$n-1+k(k+3)/2$.对应于$a_t = 1$,$b$的输出为1和连续$k$个0;对应于$a_{t+1} = 0$,$b$的输出至少为连续$k$个0;对应于$a_{t+2} = 0$,$b$的输出至少为连续$k-1$个0;…;对应于$a_{t+k} = 0$,$b$的输出至少为连续1个0;对应于$a_{t+k+1} = a_{t+k+2} = a_{t+n-1} = 0$,$b$的输出至少为连续$n-k-1$个0.因此这个0游程的长度不小于$n-1+k(k+3)/2$.定理3得证.

3　基于 GF(2^m)上 m 序列的稀疏序列

定义 5　对于$GF(2^m)$上的n级m序列$a = a_0 a_1 a_2 \cdots$,若$a_t = 1$,则输出$b_t = 1$;若$a_t \neq 1$,则输出$b_t = 0$;$t = 0,1,2,\cdots$.如此得到输出序列$b = b_0 b_1 b_2 \cdots$,称其为基于序列a的稀疏序列.

定理 4　稀疏序列b的最小周期为$P = (2^{mn}-1)$,其密度为$D = (2^{m(n-1)})/(2^{mn}-1)$.

证明　注意m序列的性质.已知序列a的最小周期为$(2^{mn}-1)$.序列b的连续$(2^{mn}-1)$个比特中,有$(2^{m(n-1)})$个1,且只有一个最长的1游程(n长).定理4得证.由定理4,已经知道序列b的线性复杂度$L \geq \max(mn+1, 2^m)$.

引理 2　设$a = a_0 a_1 a_2 \cdots$为$GF(2^m)$上的n级m序列,$c = c_0 c_1 c_2 \cdots$为a的$(2^{mn}-1)/(2^m-1)$步左平移序列,则存在$GF(2^m)$上的元素e,使得序列$(ea_0, ea_1, ea_2, \cdots) = c$.

证明　任取$GF(2^m)$上的元素y,我们知道序列$(ya_0, ya_1, ya_2, \cdots)$是$a$的平移序列,设它是$a$的$u$步左平移序列.则$(y^2 a_0, y^2 a_1, y^2 a_2, \cdots)$是$a$的$2u$步左平移序列;$(y^3 a_0, y^3 a_1, y^3 a_2, \cdots)$是$a$的$3u$步左平移序列;…;$(y^{2^m-1} a_0, y^{2^m-1} a_1, y^{2^m-1} a_2, \cdots)$是$a$的$(2^m-1)u$步左平移序列.但$(y^{2^m-1} a_0, y^{2^m-1} a_1, y^{2^m-1} a_2, \cdots) = a$,故$(2^{mn}-1) \mid (2^m-1)u$,即存在$GF(2^m)$上的元$e$,使得序列$(ea_0, ea_1, ea_2, \cdots)$为$a$的$(2^{mn}-1)/(2^m-1)$步左平移序列.又显然$e$是$GF(2^m)$上的元素.引理2得证.

引理 3　设$a = a_0 a_1 a_2 \cdots$为$GF(2^m)$上的n级m序列,序列$b = b_0 b_1 b_2 \cdots$为基于序列a的稀疏序列.记$b^{(tu)}$为b的tu步左平移序列,其中$u = (2^{mn}-1)/(2^m-1)$.则2^m-1个序列

$$\{b^{(tu)}, t = 0,1,2,\cdots,2^m-2\}$$

它们各自取值为1的位置互不重叠.

证明　由定义5、定理3和引理2即得证引理3.

定理 5　设a为$GF(2^m)$上的n级m序列,其在$GF(2^m)$上的极小多项式为$f(x)$;序列b为基于序列a的稀疏序列,其在$GF(2)$上的极小多项式为$g(x)$.则有$f(x) \mid g(x)$.

证明　记

$$S_a(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{2^{mn}-2} x^{2^{mn}-2};$$

$$S_b(x) = b_0 + b_1 x + b_2 x^2 + \cdots + b_{2^{mn}-2} x^{2^{mn}-2}.$$

由引理2和引理3知存在$GF(2^m)$上的$y = e^{-1}$,$u = (2^{mn}-$

$1)/(2^m-1)$,使

$$S_b(x)(1+(yx^u)+(yx^u)^2+\cdots+(yx^u)^{2^m-2}) \bmod (x^{2^{mn}-1}-1)=S_a(x)$$

因此有 $f(x)=\dfrac{x^{2^{mn}-1}-1}{\gcd(x^{2^{mn}-1}-1,S_a(x))}=\dfrac{yx^u-1}{\gcd(yx^u-1,S_b(x))}$;

又有 $g(x)=\dfrac{x^{2^{mn}-1}-1}{\gcd(x^{2^{mn}-1}-1,S_b(x))}$,以及 $yx^u-1 \mid x^{2^{mn}-1}-1$,故有 $f(x) \mid g(x)$. 定理 5 得证.

定理 6 设序列 b 为基于 $GF(2^m)$ 上 n 级 m 序列 a 的稀疏序列,则 b 的线性复杂度 $L(b)$ 有:$L(b) \geq n(2^m-1)$.

证明 只须证明 $GF(2^{mn})$ 上有不少于 $n(2^m-1)$ 个元素 x,使得 $S_b(x) \neq 0$.

取 $\beta \in GF(2^m)$,$\beta \neq 0$,构造集合 $A(\beta)=\{x \in GF(2^{mn}), x^u=\beta\}$,其中 $u=(2^{mn}-1)/(2^m-1)$. 共有 2^m-1 个这样的集合,它们互不相交. 我们将证明:对每一个集合 $A(\beta)$,至少有 n 个 $x \in A(\beta)$,使得 $S_b(x) \neq 0$. 首先 $A(\beta)=\{\delta^i\varepsilon;\ k=0 \sim u-1\}$,其中 δ 是 $GF(2^{mn})$ 上一个阶为 u 的固定的元素,ε 是 $A(\beta)$ 上一个固定的元. 其次将 $S_b(x)$ 记为

$$S_b(x)=S_0(x)+x^uS_1(x)+x^{2u}S_2(x)+\cdots x^{(v-1)u}S_{v-1}(x)$$

其中 $v=2^m-1$,$S_0(x)$、$S_1(x)$、$S_2(x)$、…、$S_{v-1}(x)$ 均为 $GF(2)$ 上次数低于 u 的多项式. 由引理 3 知道,$S_0(x)$、$S_1(x)$、$S_2(x)$、…、$S_{v-1}(x)$ 各自系数为 1 的项不相重合,即对每一个 j,$\{b_j, b_{u+j}, b_{2u+j}, \cdots, b_{(v-1)u+j}\}$ 中至多只有一个为 1. 而向量 $(b_0, b_1, b_2, \cdots, b_{2^{mn}-2})$ 非 0,因此以下向量必为非 0 向量:

$$\left(\sum_{i=0}^{v-1}b_{iu}\beta, \varepsilon\sum_{i=0}^{v-1}b_{iu+1}\beta, \varepsilon^2\sum_{i=0}^{v-1}b_{iu+2}\beta, \cdots, \varepsilon^{u-1}\sum_{i=0}^{v-1}b_{iu+(u-1)}\beta\right)$$

不失一般性,设 $a_{2^{mn}-2}=a_{2^{mn}-3}=\cdots a_{2^{mn}-n}=0$. 因此 $\varepsilon^{u-1}\cdot\sum_{i=0}^{v-1}b_{iu+(u-1)}\beta=\varepsilon^{u-2}\sum_{i=0}^{v-1}b_{iu+(u-2)}\beta=\cdots\varepsilon^{u-n+1}\cdot\sum_{i=0}^{v-1}b_{iu+(u-n+1)}\beta=0$. 于是有

$$\begin{bmatrix} S_b(\varepsilon) \\ S_b(\varepsilon\delta) \\ S_b(\varepsilon\delta^2) \\ \cdots \\ S_b(\varepsilon\delta^{u-1}) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \delta & \delta^2 & \cdots & \delta^{u-n} \\ 1 & \delta^2 & \delta^4 & \cdots & \delta^{2(u-n)} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & \delta^{u-1} & \delta^{2(u-1)} & \cdots & \delta^{(u-1)(u-n)} \end{bmatrix} \cdot \begin{bmatrix} \sum_{i=0}^{v-1}b_{iu}\beta \\ \varepsilon\sum_{i=0}^{v-1}b_{iu+1}\beta \\ \varepsilon^2\sum_{i=0}^{v-1}b_{iu+2}\beta \\ \cdots \\ \varepsilon^{u-n}\sum_{i=0}^{v-1}b_{iu+(u-n)}\beta \end{bmatrix}$$

上式的矩阵中任一个子方阵可逆. 如果集合 $A(\beta)$ 中有少于 n 个 $x \in A(\beta)$ 使得 $S_b(x) \neq 0$,则存在 $u-n+1$ 个 $x \in A(\beta)$ 使得 $S_b(x)=0$,这意味着上式的矩阵中某一个子方阵不可逆. 矛盾. 矛盾说明至少有 n 个 $x \in A(\beta)$ 使 $S_b(x) \neq 0$. 定理 6 得证.

4 基于乘方剩余符号的稀疏序列

定义 6 设有素数 $p=q2^m+1$. 序列 $a=a_0a_1a_2\cdots$ 定义如下:

$$a_k=\begin{cases} 0, & p \mid k \\ 0, & \gcd(p,k)=1,\ k^q \pmod p \neq 1 \\ 1, & \gcd(p,k)=1,\ k^q \pmod p = 1 \end{cases}$$

称序列 a 为基于乘方剩余符号的稀疏序列.

显然此序列的最小周期为 p,密度为 $q/p \approx 2^{-m}$.

定理 7 基于乘方剩余符号的稀疏序列 a 的线性复杂度 $L(a)$ 有:$L(a) \geq q$.

证明 记 $S_a(x)=a_0+a_1x+a_2x^2+\cdots+a_{p-1}x^{p-1}$. 记 β 为特征为 2 的域上的 p 阶元. 将 $GF(p)^*$ 乘法群按子群 $\{k \in GF(p)^* \mid k^q=1\}$ 划分陪集,易知共有 2^m 个陪集,每个陪集有 q 个元. 当 k、j 属于同一陪集时,

$$S_a(\beta^k)=S_a(\beta^j)$$

又不可能使所有 $S_a(\beta^k)$,$k \in GF(p)^*$ 均为 0(否则 $S_a(x) \equiv 0$). 故至多有 (2^m-1) q 个 $k \in GF(p)^*$,使 $S_a(\beta^k)=0$. 这说明序列 a 的极小多项式 $(x^p-1)/\gcd(x^p-1,S_a(x))$ 的次数不小于 q. 定理 7 得证.

5 结论

我们已经讨论了 5 种稀疏序列的最小周期和线性复杂度,它们是:基于 $GF(2)$ 上 m 序列的乘积序列、自缩乘积序列、自扩序列;基于 $GF(2^m)$ 上 m 序列的稀疏序列;基于乘方剩余符号的稀疏序列. 这些序列都是容易生成的,并且除了乘积序列的线性复杂度难以讨论之外,其它序列都有较好的最小周期和线性复杂度性质,适用于信息隐藏等技术的应用.

参考文献:

[1] Stefan Katzenbeisser, Fafien A P Petitcolas, et al. Information hiding techniques for steganography and digital watermarking [C]. Artech House, Inc., 2000.

[2] Massey J L. Shift-register synthesis and BCH decoding [J]. IEEE Trans. IT, Jan. 1969: 122 - 127.

[3] 丁存生,肖国镇. 流密码学及其应用 [M]. 北京:国防工业出版社,1994.

作者简介:

胡予濮 男. 1955 年 11 月生于河南. 博士,教授,硕士研究生导师,中国电子学会高级会员,现从事信息安全领域的科研工作.