采样攻击的最短采样距分析

金晨辉 史建红 邓 辉
(解放军信息工程大学电子技术学院 郑州 450004)

摘 要：采样攻击是针对序列密码的一种攻击方法。该文对本原线性反馈移存器 (LFSR)序列的采样攻击方法进行了研究，给出了采样距与被采序列和采出序列的线性复杂度之间的制约关系，给出了能使采出序列的线性复杂度小于被采序列的线性复杂度的最短采样距，给出了能成功实施采样攻击需要的最少已知明文量，并据此分析了对本原 LFSR 序列进行采样攻击的实际可行性，证明了只有当本原 LFSR 的级数很小时，该方法才可能有实用价值。

关键词：序列密码；线性反馈移位寄存器；m 序列；采样攻击
中图分类号：TN918.1 **文献标识码**：A **文章编号**：1009-5896(2008)03-0665-03

Analysis of the Minimum Decimation Distance of Decimation Attack

Jin Chen-hui Shi Jian-hong Deng Hui
(*Institute of Electronic Technology, the PLA Information Engineering University, Zhengzhou* 450004, *China*)

Abstract: Decimation attack is one attack method of stream ciphers. In this paper, the decimation attack to prime Linear Feedback Shift Register(LFSR) sequences is investigated. The connection of decimation distance and the linear complexity of the original sequence and the decimate sequence is presented. The minimum decimate distance that makes the linear complexity of the decimate sequence less than that of the original sequence is obtained. The minimum known plaintext amount for decimation attack is given, and the practical feasibility of the decimation attack to prime LFSR's is analyzed. It is proved that the decimation attack to prime LFSR is useful possibly only in the case that the degree of LFSR is very small.

Key words: Stream cipher; Linear feedback shift register; m sequnce; Decimation attack

1 引言

线性反馈移存器(LFSR)序列是序列密码算法常用的初始乱源。在同步型序列密码算法中，LFSR 的初态通常就是密钥或密钥的一部分。对同步型序列密码算法的一类基本攻击方法就是利用已知条件构造出该 LFSR 序列的相似序列，并在 LFSR 的反馈结构已知的条件下，求出该 LFSR 序列，从而求出该序列包含的密钥因素。相关攻击和快速相关攻击是实现上述求解过程的两种基本方法，这两种攻击方法的实现性能主要由两条序列的相关优势，LFSR 的级数和已知序列的长度决定，而快速相关攻击还受 LFSR 的项数的制约。在两条序列的相关优势确定的条件下，LFSR 的级数越小，相关攻击算法的计算复杂性越低且成功率越高。随着 LFSR 的级数的增加，相关攻击的计算复杂性将呈指数级增加，因而难以奏效。

针对这个问题，Filiol 在 2000 年的印度密码年会上提出了采样攻击方法[1]该方法的思想是通过选择一个采样距 d，对 L 级 LFSR 序列进行采样，得到一个 L_d ($L_d < L$)级 LFSR 序列，然后再利用 L 级 LFSR 序列的已知相似序列的 d 采样序列攻击 L_d 级 LFSR 序列，得到其初态，从而获得 L_d 比特的密钥，最后再在这 L_d 比特密钥已知的条件下，求出 L 级 LFSR 序列的初态，从而达到降低相关攻击的计算复杂性的目的。 Filiol 给出了成功实施采样攻击的例子，并证明了只要 LFSR 的级数是素数，就一定能抵抗采样攻击。

数据复杂性是衡量破译算法性能优劣的一个重要指标。本文将针对本原 LFSR，给出能实施采样攻击的最短采样距，分析其数据复杂性的下界，并据此证明只有当本原 LFSR 的级数很小时，该方法才可能有实用价值。

在本文中，GF(2) = {0,1} 均表示二元域，$x \bmod T$ 均表示整数 x 模 T 运算的结果。

2 LFSR 序列的定长采样

定义 1[2,3] 设 $f(x) = 1 \oplus c_1 x \oplus \cdots \oplus c_L x^L$ 是二元域上的 L 次多项式，二元序列 $a = \{a_i\}_{i=0}^{\infty}$ 满足

$$a_i = c_1 a_{i-1} \oplus c_2 a_{i-2} \oplus \cdots \oplus c_L a_{i-L}, \quad i \geq L \tag{1}$$

则称序列 $a = \{a_i\}_{i=0}^{\infty}$ 是一条 LFSR 序列，称 $f(x)$ 为序列 a 的一个生成多项式。序列 a 的次数最低的生成多项式的互反多项式[2]称为 a 的极小多项式，其次数称为序列 a 的线性复杂度。特别地，如果序列 a 的周期为 $2^L - 1$，则称序列 a 是一条 L 级 m 序列，并称 $f(x)$ 为本原多项式，相应的 LFSR 称为 L 级本原 LFSR。

定义 2[2,3] 设 $a = \{a_i\}_{i=0}^{\infty}$ 为二元序列， d 为正整数，则

2006-08-28 收到，2006-12-20 改回
河南省杰出青年科学基金(0312001800)资助课题

称 $a^{(d,j)}=\{a_{di+j}\}_{i=0}^{\infty}$ 为序列 a 的起点为 j 的 d 采样序列，称 $a=\{a_i\}_{i=0}^{\infty}$ 为被采序列，称 $a^{(d,j)}=\{a_{di+j}\}_{i=0}^{\infty}$ 为采出序列。简记 $a^{(d)}=\{a_{di}\}_{i=0}^{\infty}$。

令 $b_i=a_{i+j}$，则 $a_i=b_{i-j}$，因而 $a_{di+j}=b_{di}$，即 $a^{(d,j)}=b^{(d)}$，故以下只需研究序列 $a^{(d)}$ 即可。

定义 3[4] 设 $Z/(N)$ 是模 N 剩余类环，$x\in Z/(N)$。如果存在正整数 t，使得 $x^t \bmod N=1$，则称 x 的乘法阶存在，并称使 $x^t \bmod N=1$ 成立的最小的正整数 t 为 x 在 $Z/(N)$ 中的乘法阶。

定义 4[2] 设 K 是特征为 2 的有限域，$\alpha\in K$，则 α 在 GF(2) 中的极小多项式是指二元域上以 α 为根且首项系数为 1 的次数最低的多项式。

引理 1[2, 3] 设二元域上的 LFSR 序列 a 的极小多项式 $p(x)$ 是 L 次不可约多项式，α 是 $p(x)$ 的一个根，T 是 $p(x)$ 的周期，d 是正整数，记 $p^*(x)$ 是采出序列 $a^{(d)}$ 的极小多项式，则有

(1) $p^*(x)$ 是 $\alpha^d\in \mathrm{GF}(2^L)$ 在二元域上的极小多项式；

(2) $p^*(x)$ 的周期 $T_d=\dfrac{T}{\gcd(d,T)}$；

(3) $p^*(x)$ 的次数 L_d 等于 2 在 $Z/(T_d)$ 中的乘法阶。

推论 1 设二元序列 a 的周期为 T 且其极小多项式是 L 次不可约多项式，d 是正整数，则有

(1) 序列 $a^{(d)}$ 的周期 $T_d=\dfrac{T}{\gcd(d,T)}$；

(2) 序列 $a^{(d)}$ 的线性复杂度 L_d 等于 2 在 $Z/(T_d)$ 中的乘法阶；

(3) 序列 $a^{(d)}$ 的极小多项式是 L_d 次不可约多项式。

证明 由于二元非零线性递归序列的周期就是其极小多项式的周期，故由引理 1 之(2)知(1)成立；由 $p^*(x)$ 的次数就是序列 $\underline{a}^{(d)}$ 的线性复杂度知(2)成立，由引理 1 之(1)知(3)成立。

推论 2 设二元序列 $\underline{a}$ 的周期为 T 且其极小多项式是 L 次不可约多项式，d 是正整数，若 $\gcd(d,T)=d_1$，则序列 $a^{(d)}$ 与序列 $a^{(d_1)}$ 具有相同的周期和线性复杂度。

由于采样距越小，采出序列越长，因而攻击时可利用的信息越多，故推论 2 说明在进行采样攻击时，采样距 d 应选择为周期 T 的因子。

当 $\gcd(d,T)=1$ 时，$T_d=T$，故由推论 1 之(2)知序列 $a^{(d)}$ 的线性复杂度等于序列 a 的线性复杂度。因此[1]，d 采样序列 $a^{(d)}$ 的线性复杂度小于序列 a 的线性复杂度的必要条件是 d 与 a 的周期 T 不互素。

3 能降低采出序列线性复杂度的最短采样距

下面首先给出采出序列的线性复杂度与被采序列的线性复杂度之间的关系。

定理 1 设周期为 T 的二元序列 a 的极小多项式为 L 次不可约多项式，则采出序列 $a^{(d)}$ 的线性复杂度 L_d 整除 L。

证明 在引理 1 中取 $d=1$，则由引理 1 之(3)知 L 是 2 在 $Z/(T)$ 中的乘法阶，因而 $2^L \bmod T=1$，故 $T\mid(2^L-1)$。再由 T_d 整除 T 知 $2^L \bmod T_d=1$，从而由 L_d 是 2 在 $Z/(T_d)$ 中的乘法阶知 L_d 整除 L。 证毕

接着分析采样距与被采序列的周期和采出序列的线性复杂度之间的制约关系。

定理 2 设周期为 T 的二元序列 a 的极小多项式为不可约多项式，其采出序列 $a^{(d)}$ 的线性复杂度为 n，则 $d\ge T/(2^n-1)$。

证明 记采出序列 $a^{(d)}$ 的周期为 T_d，则由引理 1 的推论 1 知 $2^n \bmod T_d=1$，且 $T_d=T/\gcd(d,T)$，因而有 $2^n-1\ge T_d=T/\gcd(d,T)$，这说明 $d\ge\gcd(d,T)\ge T/(2^n-1)$。证毕

最后给出能使 L 级本原 LFSR 序列的采出序列的线性复杂度小于 L 的最小采样距。

定理 3 设二元序列 a 是 L 级 $\boldsymbol{m}$ 序列，n 是 L 的正因子，则使采出序列 $a^{(d)}$ 的线性复杂度为 n 的最小采样距为 $d=(2^L-1)/(2^n-1)$，且序列 $a^{(d)}$ 是 n 级 $\boldsymbol{m}$ 序列。

证明 由 a 是 $\boldsymbol{m}$ 序列知其周期 $T=2^L-1$，故由定理 2 知，如果序列 $a^{(d)}$ 的线性复杂度为 n，则有 $d\ge(2^L-1)/(2^n-1)$。此外，由于 $\gcd(2^n-1,2^L-1)=2^{\gcd(n,L)}-1=2^n-1$，故 2^n-1 整除 2^L-1。取 $d=(2^L-1)/(2^n-1)$，则由引理 1 知序列 $a^{(d)}$ 的周期 $T_d=(2^L-1)/d=2^n-1$，且其线性复杂度是 2 在 $Z/(T_d)=Z/(2^n-1)$ 中的阶 n。 证毕

推论 3 设 LFSR 的反馈多项式是 L 次本原多项式，则对所有正整数 d，$0<d<2^L-1$，LFSR 的输出序列的 d 采样序列的线性复杂度都是 L 的充要条件是 L 为素数。

证明 充分性的证明见文献[1]。必要性由定理 1 和定理 3 即知。

定理 3 说明，只要 LFSR 的级数不是素数，就可使采出序列的线性复杂度缩小，从而能够实施采样攻击。

推论 4 设二元序列 a 是 L 级 m 序列，p 是 L 的最小素因子，则序列 a 的 d 采样序列的线性复杂度小于 L 的必要条件是 $d>2^{\frac{p-1}{p}L}$。

证明 由 p 是 L 的最小素因子知 L 的最大真因子是 L/p，故序列 a 的 d 采样序列的线性复杂度 $L_d\le L/p$。再由定理 3 知

$$d\ge\frac{2^L-1}{2^{L_d}-1}>\frac{2^L}{2^{L_d}}=2^{L-L_d}\ge 2^{L-\frac{L}{p}}=2^{\frac{p-1}{p}L} \tag{2}$$

下面从对本原 LFSR 序列实施采样攻击所需的最短信号量入手，分析采样攻击的实际可行性。

定理 4 设已知 L 级二元 m 序列 a 的一条长度为 N 的相似序列 b，p 是合数 L 的最小素因子，则利用序列 b 的 d 采样序列能够实施采样攻击的必要条件是 $N>2^{\frac{p-1}{p}L}\dfrac{L}{p}\ge 2^{1+\frac{L}{2}}$。

证明 设序列 a 的 d 采样序列的线性复杂度为 L_d，则序列 b 的 d 采样序列 $b^{(d)}$ 的长度近似为 N/d，故能够利用序列

$b^{(d)}$ 求出序列 $a^{(d)}$ 的初态的必要条件是 $N/d > L_d$，因而 $N > dL_d$。

另一方面，由定理 3 知 $d \ge \dfrac{2^L-1}{2^{L_d}-1}$，从而有 $N > dL_d \ge \dfrac{2^L-1}{2^{L_d}-1}L_d$。由于 $f(x)=\dfrac{x}{2^x-1}$ 在 $x \ge 2$ 时是减函数，由定理 3 之推论 2 知 $L_d \le L/p$，故有

$$N > \frac{2^L-1}{2^{L_d}-1}L_d \ge \frac{(2^L-1)L}{(2^{L/p}-1)p} > 2^{L-\frac{L}{p}}\frac{L}{p} \ge 2^{\frac{p-1}{p}L+1} \ge 2^{\frac{L}{2}+1}$$

证毕

由于采样攻击在利用 d 采样序列实施攻击时，只能利用被采序列的信号量的 d 分之一，因而是以付出已知信号量的巨大牺牲为代价，换取了被攻击的 LFSR 级数的减低。显然，当 d 很大时，采样攻击方法就无法在实际的破译中实施。

当本原 LFSR 的级数是偶数时，$2n$ 级 m 序列的 $d=2^n+1$ 采样序列都是 n 级 m 序列[3]，且由定理 3 知 $d=2^n+1$ 是使 d 采样序列的级数小于 $2n$ 的最短采样距，故由定理 4 知，能对 $2n$ 级的 LFSR 序列成功实施采样攻击的前提条件是至少获得长度大于 $2^n n$ 的已知明文序列。

例如，要成功对 60 级的 m 序列实施采样攻击，由定理 4 知，至少需要长度为 30Gbit 的已知明文序列，这在实际中是难以实现的。当本原 LFSR 的级数 ≥ 41 且是偶数时，由定理 4 知，至少需要 20Mbit 的已知明文序列才能实施采样攻击。因此，此时只要限定序列密码算法在使用中一个密钥所加密的明文数据不超过 20Mbit，就能抵抗采样攻击。

根据定理 4，当本原线性反馈移存器的级数 ≥ 100 时，能成功实施采样攻击的必要条件是获取的信号长度 $N > 2^{51}$，这在现实是不可能的，故级数 ≥ 100 的本原线性反馈移存器实际上都能抵抗采样攻击。

下面对级数在 40 至 100 之间的奇合数 L，考查定理 4 中 $(p-1)L/p$ 的值，从而考查本原 LFSR 的抗采样攻击能力。(表 1)。

表 1

级数	45	49	51	55	57	63	65	69	75	77	81	85	87	91	93	95	99
最小素数 p	3	7	3	5	3	3	5	3	5	7	3	5	3	7	3	5	3
$(p-1)L/p$	30	42	34	44	38	42	52	46	60	66	54	68	58	78	62	76	66

由于当 $p=69$ 时，$2^{\frac{p-1}{p}L}\dfrac{L}{p} \ge 2^{\frac{3-1}{3}\times 69}\dfrac{69}{3} \approx 1.4\times 2^{50}$，故由定理 4 知，当本原 LFSR 的级数为 64 至 100 之间的奇合数时，成功实施采样攻击所需要的已知明文的长度都大于 2^{50}，因而实际上都能抵抗采样攻击；当本原 LFSR 的级数为 41 至 63 之间的奇合数时，除 45 和 51 外，成功实施采样攻击所需的已知明文的长度都大于 2^{42}，而当本原 LFSR 的级数为 45 和 51 时，成功实施采样攻击所需的已知明文的长度必须大于 2^{33} 和 2^{38}，因而绝大多数的实际应用都能抵抗采样攻击。

因此，采样攻击只可能对级数 ≤ 40 的 m 序列有效。但因此时初态的穷举复杂度 $\le 2^{40}-1$，故在现有计算能力下对级数 ≤ 40 的相关攻击是可实现的，也没有必要实施采样攻击。

然而，当 L 级 LFSR 序列的极小多项式不是本原多项式时，其输出序列的周期 T 可能远远小于 2^L-1，因而定理 3 中的周期 T 可能接近于 2^n-1，此时有可能取到较小的采样距 d，使得周期为 T 的 d 采样序列的线性复杂度小于 L，从而在不牺牲过多的已知明文量的条件下，成功实施采样攻击。对于该条件下采样攻击的具体性能，还有待进一步研究。这也说明，在序列密码的设计中，我们应将 LFSR 选择为本原 LFSR。

4　结束语

采样攻击是一类颇有创新的攻击方法。本文对本原 LFSR 的采样攻击方法进行了研究，给出了能够使采出序列的线性复杂度小于被采序列的线性复杂度的最短采样距，并据此给出了对本原 LFSR 的采样攻击的数据复杂性下界，进而根据能成功实施采样攻击需要的最少已知明文量，分析了对本原 LFSR 序列进行采样攻击的实际可行性，指出该方法是以牺牲可利用的已知明文量为代价，换取被攻击的 LFSR 序列的级数的降低；证明了只有当本原 LFSR 的级数很小时，该方法才可能有实用价值。

参 考 文 献

[1] Filiol E. Decimation attack of stream ciphers[A]. In: Proceedings of the First International Conference in India-INDOCRYPT' 2000, Lecture Notes in Computer Science 1977, Springer Verlag, 2000. Also available from http://eprint. iacr. org/2000/040. ps.

[2] 肖国镇, 梁传甲, 王育民. 伪随机序列及其应用[M]. 北京: 国防工业出版社, 1985年, 第二章.

[3] Rueppel R A. Analysis and Design of Stream Ciphers[M]. Springer Verlag, 1986, Ch. 6. .

[4] 陈欣, 李保红. 模n剩余类环中元素的周期分布规律[J]. 信阳师范学院学报, 2000, 13(1): 4-6.

[5] 潘承洞, 潘承彪. 初等数论[M]. 北京: 北京大学出版社, 1992年, 第五章.

金晨辉:　男，1965 年生，博士，教授，博士生导师，主要研究方向为密码理论和信息安全.

史建红:　男，1975 年生，博士生，讲师，研究方向为密码理论.

邓　辉:　男，1980 年生，硕士，研究方向为密码理论.