序列密码的复合攻击法 [1]

孙林红 叶顶锋 吕述望 冯登国

(中国科学院研究生院信息安全国家重点实验室 北京 100039)

摘 要: 该文针对序列密码，提出了复合攻击的思想，并用相关攻击和求逆攻击复合对前馈网络进行了分析，理论和实验结果给出了复合攻击法的计算复杂度，并与常用的攻击法进行了比较。

关键词: 前馈网络，求逆攻击，相关攻击

中图分类号: TN918.1 **文献标识码**: A **文章编号**: 1009-5896(2004)01-0072-05

Composite Attack on Sequence Cipher

Sun Lin-hong Ye Ding-feng Lü Shu-wang Feng Deng-guo

(*State Key Lab of Info. Security, Chinese Academy of Sciences, Beijing* 100039, *China*)

Abstract The composite attack on sequence cipher is developed and forward feedback network is analyzed by the composition of correlation attack and inversion attack, both theory and experimental results show its complexity. Finally, comparison with normal attack is given.

Key words Forward feedback network, Inversion attack, Correlation attack

1 引言

在序列密码的研究中，主要进行各种序列密码生成器的设计和分析，前馈网络密码系统是其中比较重要的一种，因此人们对一些特殊的前馈模型进行了大量的研究，并且得出了一系列重要成果，特别是前馈网络密码系统的设计，但在分析方面比较深入的研究结果尚不多见。前馈网络密码系统由单个本原的移位寄存器和非线性布尔函数组成，非线性布尔函数 f 的输入取自本原移位寄存器中的某些相位。一个前馈网络应能抵抗目前已知的实用密码攻击。密码攻击的目的主要是从一个足够长的密钥流序列获取控制移位寄存器初态的秘密密钥。目前已有一些设计密码准则，如长周期、非线性复杂度、统计特性、抗快速相关攻击 [1−3]、条件相关攻击 [4,5]、求逆攻击 [6] 等。前馈网络密码系统如图 1 所示。

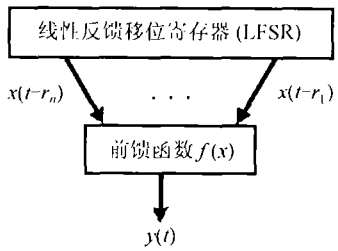

图 1 前馈网络密码系统

设 r 为线性移位寄存器的长度，n 为非退化前馈函数的输入变量的个数，$\gamma=(\gamma_i)_{i=1}^{n}$ 表示对线性移位寄存器抽头的位置，$M=\gamma_n-\gamma_1+1$ 表示前馈网络的输入存储空间的大小。如果 r 和 M 很大，则单独使用相关攻击和求逆攻击的时间复杂度都会很大。无论哪种攻击方法，其目的是恢复线性反馈移位寄存器 (LFSR) 的初始状态，所以可以将相关攻击和求逆攻击复合使用，先对初始

[1] 2002-09-24 收到，2003-04-30 改回
国家高技术研究发展计划 (863) 资助项目 (2001AA140101)

状态中 k 个比特进行相关攻击，然后对剩下的 $r-k$ 个比特进行求逆攻击，这样时间复杂度就会有所降低。本文在第 2 节介绍相关攻击和求逆攻击，第 3 节给出相关攻击和求逆攻击的复合算法，第 4 节对复合攻击法进行理论和实验分析，第 5 节讨论需要进一步研究的问题。

2 相关攻击和求逆攻击

2.1 相关攻击

在前馈网络中，密钥流 $\{y_t\}$ 和 LFSR 的输出序列 $\{x_t\}$ 之间存在着一定的相关性，所以可以将 LFSR 的输出序列看着经过一个二元无记忆对称信道 (BSC) 而得到密钥流，BSC 的误码率为 $p<1/2$，$\varepsilon=1/2-p$ 通常很小，在这种情况下，N 长 LFSR 输出序列可以看成 (N,L) 二元线性码，每一个码字对应于一个初始状态，其数量等于 LFSR 初态的数量，这样密码分析就演变为对受噪声干扰的 BSC 的译码问题。

不妨设 $P(x_t=y_t)=1-p=1/2+\varepsilon$，LFSR 的联接多项式为 $g(x)$，具体如图 2 所示。

Siegenthaler[7] 利用穷举攻击进行译码，该算法采用极大似然译码法 (ML)，该算法的复杂度为 $O(2^r r/C(p))$，$C(p)=1-H(p)$，$H(x)=-x\log_2 x-(1-x)\log_2(1-x)\cdots$

Meier 和 Staffelbach[3] 给出了当连接多项式的重量较小时译码的方法——快速相关攻击算法，本质上是利用了重复译码技术，其计算复杂度为 $O(2^{\alpha r})$, $\alpha<1$。

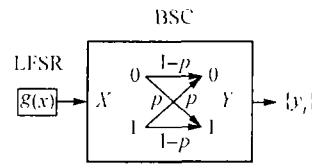

图 2　二元无记忆对称信道 (BSC)

2.2 求逆攻击

设 $x=(x(t))_{t=-r}^{\infty}$ 为二元最大长度序列，周期为 2^r-1，$(x(t))_{t=-r}^{-1}$ 为 LFSR 的初态，$f(z_1,\cdots,z_n)$ 为一非退化单输出布尔函数，$n\le r$，设 $\gamma=(\gamma_i)_{i=1}^{n}$ 为一递增的非负整数序列，$\gamma_1=0$，$\gamma_n\le r-1$，单输出前馈网络的输出序列 $y=(y(t))_{t=0}^{\infty}$ 定义如下：

$$y(t)=f(x(t-\gamma_1),\cdots,x(t-\gamma_n)),\qquad t\ge 0 \tag{1}$$

求逆攻击的目的是从密钥流的截段构造出 LFSR 的初态，分别采用前向或后向攻击法，以前向攻击为例：

$$x(t)=y(t)+g(x(t-\gamma_2),\cdots,x(t-\gamma_n)),\qquad t\ge 0 \tag{2}$$

前向攻击的过程如下：

(1) 穷举 r 比特初态中的 M 个比特 $(x(t))_{t=-M}^{-1}$

(2) 利用式 (2)，由 $(y(t))_{t=0}^{r-M-1}$ 密钥流序列确定所有可能的 $(x(t))_{t=0}^{r-M-1}$ 输入序列

(3) 利用 LFSR 的线性递归，由 $(x(t))_{t=-M}^{r-M-1}$ 生成 $(x(t))_{t=r-M}^{N-1}$

(4) 利用式 (1)，由 $(x(t))_{t=r-2M}^{N-1}$ 计算 $(y'(t))_{t=r-M}^{N-1}$，比较 $(y'(t))_{t=r-M}^{N-1}$ 和 $(y(t))_{t=r-M}^{N-1}$，如果相同则接受，否则返回到 (1)

3 复合攻击

复合攻击主要是先对初始状态中 k 个比特进行相关攻击，然后对剩下的 $r-k$ 个比特进行求逆攻击，k 的选择应在 $M/2$ 左右。为了对 k 个比特进行相关攻击，必须构造一个维数为 k 的线性码 C_2, 构造方法 [8] 如下：

LFSR 的生成多项式 $g(D)=c_0+c_1D+\cdots+c_rD^r(c_0=c_r=1)$, 输出序列为 $(x_{-r},\cdots,x_{-1},x_0,\cdots,x_{N-1})$ 因为每一 x_i 符号是 r 个初始状态的线性组合，$(x_{-r},\cdots,x_{-1},x_0,\cdots,x_{N-1})$ 的

集合构成线性码 C, 所有的字称为码字, 字 $(x_{-r},\cdots,x_{-1})$ 成为信息字, 符号 $x_{-r},\cdots,x_{-1}$ 成为信息符号, 从而有

$$\begin{cases} c_r x_{-r} + c_{r-1}x_{-r+1} + \cdots + c_1 x_{-1} + c_0 x_0 = 0 \\ \quad\vdots \\ c_r x_{N-1-r} + c_{r-1}x_{N-r} + \cdots + c_1 x_{N-2} + c_0 x_{N-1} = 0 \end{cases}$$

由上式可以将 $(x_0,\cdots,x_{N-1})$ 写成下式:

$$\left.\begin{aligned} x_0 &= h_0^1 x_{-r} + h_0^2 x_{-r+1} + \cdots + h_0^r x_{-1} \\ &\vdots \\ x_{N-1} &= h_{N-1}^1 x_{-r} + h_{N-1}^2 x_{-r+1} + \cdots + h_{N-1}^r x_{-1} \end{aligned}\right\} \tag{3}$$

如果表示 $h_i(D) = h_i^1 + h_i^2 D + \cdots + h_i^l D^{l-1}$, 则有 $h_i(D) = D^{i+r} \bmod g(D), i = 0,\cdots,N-1$ 。

假设 N 充分大, 从式 (3) 中寻找下述对:

$$h_i^1 = h_j^1, \quad h_i^2 = h_j^2, \quad \cdots, \quad h_i^{r-k} = h_j^{r-k}, \qquad 0 \le i \ne j \le N-1 \tag{4}$$

寻找等式中所有这些对, 记这些不同对的个数为 n_2 , 则所有对为 $\{i_1,j_1\},\{i_2,j_2\},\cdots,\{i_{n_2},j_{n_2}\}$ 。

如果 i 和 j 满足式 (4) , $x_i + x_j$ 为 $x_{-k},x_{-k+1},\cdots,x_{-1}$ 的线性组合, 与其他信息符号 $x_{-r},\cdots,x_{-k-1}$ 无关。这意味着序列 $(X_1,X_2,\cdots,X_{n_2}) = (x_{i_1}+x_{j_1}, x_{i_2}+x_{j_2},\cdots,x_{i_{n_2}}+x_{j_{n_2}})$ 构成一个 (n_2,k) 的线性码, 记为 C_2, 其信息符号为 $x_{-k},x_{-k+1},\cdots,x_{-1}$, 维数为 k 。

记 $Y_1 = y_{i_1} + y_{j_1}$, $Y_2 = y_{i_2} + y_{j_2},\cdots,Y_{n_2} = y_{i_{n_2}} + y_{j_{n_2}}$, 因为我们可观察到输出符号 $y_{i_1},y_{j_1},y_{i_2},y_{j_2},\cdots,y_{i_{n_2}},y_{j_{n_2}}$ 可计算 $Y_1,Y_2,\cdots,Y_{n_2}$, 则 $Y_1,Y_2,\cdots,Y_{n_2}$ 可以看作 $(X_1,X_2,\cdots,X_{n_2})$ 经过一个二元无记忆对称信道 BSC 的输出, BSC 的误码率为

$$p_2 = 2p(1-p) = 1/2 - 2\varepsilon^2 \tag{5}$$

这样就构造了一个新的维数较小的线性码 C_2 , 为了恢复符号 $x_{-k},x_{-k+1},\cdots,x_{-1}$, 我们只需对 C_2 译码。

由于 $x_{-k},x_{-k+1},\cdots,x_{-1}$ 可通过上述相关攻击得到, 所以下一步只需对剩下的 $r-k$ 个比特进行求逆攻击, 根据求逆攻击的原理, 只需对 $M-k$ 个比特进行求逆攻击. 整个算法描述如下:

预计算 选定 k, 构造线性码 C_2 。

译码

输入 观察到的比特 $(y_1,y_2,\cdots,y_N)$

第 1 步 计算 $Y_1,Y_2,\cdots,Y_{n_2}$;

第 2 步 对 C_2 的所有 2^k 个码字进行穷尽搜索, 选择具有最高概率的信息字 $(x_{-k},x_{-k+1},\cdots,x_{-1})$;

求逆

第 3 步 穷举 M 个比特 $(x(t))_{t=-M}^{-1}$ 中的 $M-k$ 个比特 $(x(t))_{t=-M}^{-k-1}$;

第 4 步 利用式 (2) , 由 $(y(t))_{t=0}^{r-M-1}$ 密钥流序列确定所有可能的 $(x(t))_{t=0}^{r-M-1}$ 输入序列;

第 5 步 利用 LFSR 的线性递归, 由 $(x(t))_{t=-M}^{r-M-1}$ 生成 $(x(t))_{t=r-M}^{N-1}$;

第 6 步　利用式 (1)，由 $(x(t))_{t=r-2M}^{N-1}$ 计算 $(y'(t))_{t=r-M}^{N-1}$，比较 $(y'(t))_{t=r-M}^{N-1}$ 和 $(y(t))_{t=r-M}^{N-1}$，如果相同则接受，否则返回到第 3 步。

4 理论和实验分析

首先计算复合攻击中相关攻击的计算复杂度。

定理 1[7]　设前馈网络中 LFSR 的长度为 r，前馈函数 $f(x)$ 输入和输出的符合率为 $(1/2)+\varepsilon$，相关攻击 LFSR 初始状态中 k 个比特所需观察序列的长度 N 为

$$N \approx (1/2) \cdot \sqrt{k(\ln 2)} \cdot \varepsilon^{-2} \cdot 2^{\frac{r-k}{2}} \tag{6}$$

定理 2[7]　复合攻击中相关攻击的复杂度为 $2^k k \ln 2/(8\varepsilon^4)$。

对于复合攻击中求逆攻击，对每一个初始存储状态 $(x(t))_{t=-M}^{-1}$，所获得二叉树 $(x(t))_{t=0}^{r-M-1}$ 的所有可能解与已知的输出序列 $(y(t))_{t=0}^{r-M-1}$ 相对应。对于 $1 \le n \le r-M$，Z_n 表示二叉树中第 n 节的节数，也就是 $(x(t))_{t=0}^{n-1}$ 的所有可能解的个数。初始节点 $n=0$ 包含唯一的节点，表示初始存储状态 $(x(t))_{t=-M}^{-1}$，而树中的每一个节点代表连续 M 个输入比特的内部存储状态。设 $Y_n = \sum_{l=1}^{n} Z_l$ 表示为树中直到第 n 节的所有节点数，不包括初始状态，则构造树的时间复杂度和空间复杂度可以表示为 $\sum Y_{r-M}/(r-M)$ 和 $\max\{Y_{r-M}\}$，此处的求和与求最大值针对于所有的 2^{M-k} 初始存储状态，由临界分支过程理论 [8] 知：

设 p 为 $f(z_1,\cdots,z_n)+f(z_1,\cdots,z_n+1)$ 等于 0 概率，LFSR 的初态随机选取，任意 $M+1$ 个输入比特平衡且独立，对于任意 $t \ge 0$，$x(t)$ 可能解的个数为一非负整数，记为随机变量 Z，则 $\Pr\{Z=0\}=p/2$，$\Pr\{Z=1\}=1-p$，$\Pr\{Z=2\}=p/2$，期望值和方差为 $\mu=1$，$\sigma^2=p$。

定理 3[8]　求逆攻击的计算复杂度为

$$T = O(q_{r-M}^{-1} 2^{M-k}), \qquad q_{r-M} \approx 1-\left(1-\frac{2}{p(r-M)}\right)^{2^{M-k}} \tag{7}$$

综合定理 2 和定理 3，则有推论 1

推论 1　给定 k, r, ε，复合攻击的计算复杂度为

$$T = O\left(q_{r-M}^{-1} 2^{M-k} + 2^k k \frac{\ln 2}{8\varepsilon^4}\right), \qquad q_{r-M} \approx 1-\left(1-\frac{2}{p(r-M)}\right)^{2^{M-k}} \tag{8}$$

如果 $k=M/2$，则 $T=O(2^{M/2})$，复合攻击与相关攻击和求逆攻击的计算复杂度比较如表 1。

表 1　复合攻击与相关攻击和求逆攻击的计算复杂度比较

攻击方法	计算复杂度
相关攻击	$O(2^r r/C(p)) = O(2^r r\varepsilon^{-2}), (r \ge M)$
求逆攻击	$T = O(q_{r-M}^{-1} 2^M), q_{r-M} \approx 1-(1-\frac{2}{p(r-M)})^{2^M}$
复合攻击	$T = O(q_{r-M}^{-1} 2^{M/2} + 2^{M/2}\frac{M\ln 2}{16\varepsilon^4}), q_{r-M} \approx 1-(1-\frac{2}{p(r-M)})^{2^{M/2}}$

对于两种攻击，实验结果比较如下：

实验的条件　移位寄存器的长度 r=60，随机选取联接多项式，前馈函数 f 的 $n=5$，$(r,n,M)=(60,5,40)$，$\gamma=(0,7,21,32,39)$，随机选取 50 个 LFSR 的初态。实验结果为相对于 LFSR 的初态的平均值，时间复杂度的单位为秒，平台为 P3-800。

实验值　$p=0.3$，复合攻击中相关攻击的译码时间如表 2 所示。$p=0.125$，复合攻击中的求逆攻击的时间如表 3 所示。

表 2 $p=0.3$, 复合攻击中相关攻击的译码时间

k	N	n_2	译码时间 (s)
17	1.46×10^7	874	0.42
20	5×10^7	1138	1.5
23	1.85×10^7	1281	12

表 3 $p=0.125$, 复合攻击中的求逆攻击的时间

$M-k$	时间复杂度 (s)
23	13.217
20	3.635
17	0.824

5 讨论

本文主要针对序列密码, 提出了复合攻击的思想, 并用相关攻击与求逆攻击复合对前馈网络进行了分析。在任何一个密码攻击中都存在成功率的问题, 由文献 [7] 知: 如果 $N=2n_0$, 则成功译码的概率接近于 1, $n_0\approx0.35r\varepsilon^{-2}$ 为译码的临界值, 所以复合攻击相对于单独的相关攻击所需的密钥流数据量更大; 即使如此, 译码存在收敛性, 也就是当译码迭代达到一定的次数后, 成功的概率不会随迭代的次数增加而增大, 所以译码本身始终存在一定的误码率, 由于复合攻击第 2 步为求逆攻击, 所以求逆攻击又可以进一步提高成功率, 不妨设译码的成功率为 $\rho_{\text{相关攻击}}$, 求逆攻击的成功率为 $\rho_{\text{求逆攻击}}$, 则复合攻击成功的概率为: $\rho_{\text{复合攻击}}=1-(1-\rho_{\text{相关攻击}})(1-\rho_{\text{求逆攻击}})=\rho_{\text{相关攻击}}+\rho_{\text{求逆攻击}}-\rho_{\text{相关攻击}}\rho_{\text{求逆攻击}}>\max\{\rho_{\text{相关攻击}},\rho_{\text{相关攻击}}\}$。所以在密码分析中, 如果能够灵活复合各种攻击方法, 攻击的效果会优于单个攻击方法的使用。

参 考 文 献

[1] Zeng K, Yang C H, Rao T R N. An improved linear syndrome algorithm in cryptanalysis with applications. In: Advances in Cryptology- EUROCRYPT'90, Berlin, Springer-Verlag, 1991: 34-47.

[2] Zeng K, Huang H. On the linear syndrome method in cryptanalysis. In: Advances in Cryptology-EUROCRYPT'88, London, Springer-Verlag, 1989: 164-171.

[3] Meier W, Staffelbach O. Fast correlation attacks on stream ciphers. In: Advances in Cryptology-EUROCRYPT'88, Brelin, Springer-Verlag, 1989: 301-314

[4] Zeng K, Yang C H, Rao T R N. On the linear consistency test(LCT) in cryptanalysis with applications. In: Advances in Cryptology-EUROCRYPT'89, Barcelona, Springer-Verlag, 1990: 186-193.

[5] Anderson R J. Searching for the optimum correlation attack. Fast Software Encryption-Leuven'94, Lecture Notes in Computer Science, vol.1008, B. Preneel, Springer-Verlag, 1995: 137-143

[6] Golic J Dj. On the security of nonlinear filter generators. Fast Software Encryption-Cambridge'96, Lecture Notes in Computer Science, vol.1039, D. Gollmanned, Springer-Verlag, 1996: 173-188.

[7] Chepyzhov V, Johansson T, Smeets B. A simple algorithm for fast correlation attacks on stream ciphers. Fast Software Encryption, FSE'2000, Lecture Notes in Computer Science, Springer-Verlag, 2000: 213-223.

[8] Athreya K B, Ney P E. Branching Processes. Berlin: Springer-Verlag, 1972.

孙林红: 男, 1969 年生, 博士生, 主要研究方向为密码理论.

叶顶锋: 男, 1966 年生, 教授, 博士生导师, 主要研究方向为密码理论.

吕述望: 男, 1941 年生, 教授, 博士生导师, 主要研究方向为密码理论.

冯登国: 男, 1965 年生, 教授, 博士生导师, 主要研究方向为密码理论.