组合生成器的多线性相关攻击

张卫明[1,2], 李世取[1]
(1. 信息工程大学信息研究系, 河南郑州市 450002; 2. 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: 本文对组合生成器提出了一种相关攻击方法, 这种方法同时利用组合生成器输入与输出之间多个线性关系的信息来恢复密钥, 我们从理论上证明了该方法可有效的减少攻击所需的数据量. 特别地, 我们将这种方法用于攻击"蓝牙组合生成器", 使攻击复杂度得到显著降低.

关键词: 流密码; 组合生成器; 相关攻击; 蓝牙组合生成器
中图分类号: TN918.1 文献标识码: A 文章编号: 0372-2112 (2005) 03-0427-06

Multi-Linear Correlation Attack on Combiners

ZHANG Wei-ming[1,2], LI Shi-qu[1]
(1. *Dept. of Information Research, Information Engineering University, Zhengzhou, Henan 450002, China;*
2. *State Key Laboratory of Information Security, Graduate School of the Chinese Academy Sciences, Beijing 100039, China*)

Abstract: A correlation attack method on combiners is presented, which is based on multi-linear correlations between the inputs and outputs of combiners. This method can effectively reduce the amount of data needed in correlation attacks. Specially, we apply this method to bluetooth combiners, and reduce the complexity of attack effectively.

Key words: stream cipher; combiners; correlation attack; bluetooth combiner

1 引言

组合生成器是流密码系统中使用的一类重要的密钥流生成器. 分为无记忆组合生成器和带记忆组合生成器. 无记忆组合生成器由一组线性移位寄存器(LFSRs)和一个非线性组合函数组成, 而带记忆组合函数还需要另外有一组非线性组合函数做反馈. 目前针对这种系统主要的一种攻击方法是相关攻击, 关于相关攻击已有大量的研究结果, 相关攻击的基础一般是在输入和输出之间找一个具有大的相关性的线性对, 这个线性对的相关性的大小很大程度上决定了攻击所需的密文量, 和攻击算法的复杂度. 但是现实中使用的密钥流生成器都不会存在相关性很大的线性关系. 例如蓝牙协议[1]中的密钥流生成器——蓝牙生成器[2], 是一个带 4 比特记忆的组合生成器, 每次最多输出 2745 比特就重新初始化, 而且其所有的输入输出线性对的相关性都不大, 所以即使选一个相关性达到最大的线性对进行相关攻击, 所需的明密文量也远大于 2745 比特.

本文的主要目的是考察如何同时使用多个具有相关性的线性对攻击组合生成器, 从而约减攻击所需的数据量. 我们利用组合生成器的多个线性关系构造了一个统计量, 并考察了此统计量的近似分布. 在此基础上, 给出了一个一般性的相关攻击方法, 并对此方法所需的数据量进行了分析, 证明了同时利用多个线性关系可以有效的约减数据量. 最后我们把此方法用于分析蓝牙组合生成器, 结果表明, 如果计算和存储资源足够, 可以将所需的数据量降到了 1000 比特以内, 从而突破蓝牙协议中密钥流输出长度不超过 2745 比特的限制.

2 无记忆组合生成器序列的近似分布

本文用大写字母表示布尔随机变量, 小写字母表示其取值或一般布尔变量. 对 $w=(w_1, \cdots, w_n) \in GF^n(2)$ 和 $x=(x_1, \cdots, x_n) \in GF^n(2)$, 用 $w \cdot x$ 表示 $w_1x_1 \oplus \cdots \oplus w_nx_n$.

具有 n 个输入的无记忆非线性组合生成器定义为:

$$z_j = f(x^{(j)}), j \geqslant 1,$$

其中, $f(x)$, $x \in GF^n(2)$ 为输出函数, $x^{(j)}=(x_{1j}, x_{2j}, \cdots, x_{nj})$ 为时刻 j 的输入向量.

密钥流生成器的输入要求有很好的伪随机性, 所以可假设输入是独立均匀分布的, 从而有如下概率模型. 设 $X^{(1)}=(X_{11}, X_{21}, \cdots, X_{n1})$, $X^{(2)}=(X_{12}, X_{22}, \cdots, X_{n2})$, $\cdots$, $X^{(m)}=(X_{1m}, X_{2m}, \cdots, X_{nm})$, $\cdots$是定义在同一概率空间上的相互独立的 n 维布尔随机向量序列, 且对任一 $j \geqslant 1$, $(X_{1j}, X_{2j}, \cdots, X_{nj})$ 中的 $X_{1j}, X_{2j}, \cdots, X_{nj}$ 是相互独立且都具有均匀分布的布尔随机变量, 由此可以得到独立布尔随机变量序列

$$f(X^{(1)}), f(X^{(2)}), \cdots, f(X^{(m)}), \cdots\cdots$$

收稿日期: 2003-09-05; 修回日期: 2004-11-19
基金项目: 国家自然科学基金(No. 60473022)

设 $X=(X_1, X_2, \cdots, X_n)$ 为布尔随机向量, 其中 $X_1, X_2, \cdots, X_n$ 是任一概率空间上的相互独立且都具有均匀分布的布尔随机变量. 称 $a_1\oplus w^{(1)}\cdot x, w^{(1)}\in GF^n(2)$ 是 $f(x)$ 的最优仿射逼近, 如果

$$P\{f(X)=a_1\oplus w^{(1)}\cdot X\}=\max\{P\{f(X)=b\oplus w\cdot X\}: w\in GF^n(2), b\in GF(2)\},$$

称 $p_1=P\{f(X)=a_1\oplus w^{(1)}\cdot X\}$ 为符合概率. 由能量守恒定理[3]知总有 $p_1>\frac{1}{2}$.

由模型假设知 $P\{f(X^{(j)})=a_1\oplus w^{(1)}\cdot X^{(j)}\}=p_1, j\geqslant 1$, 且 $(f(X^{(1)}), w^{(1)}\cdot X^{(1)}), (f(X^{(2)}), w^{(1)}\cdot X^{(2)}), \cdots, (f(X^{(m)}), w^{(1)}\cdot X^{(m)}), \cdots$是独立且同分布的二维布尔随机向量序列. 因而, 若令

$$\xi_j=\begin{cases}1, & \text{若} f(X^{(j)})=a_1\oplus w^{(1)}\cdot X^{(j)}\\ -1, & \text{若} f(X^{(j)})\neq a_1\oplus w^{(1)}\cdot X^{(j)}\end{cases}, j\geqslant 1,$$

则所得 $\xi_1, \xi_2, \cdots, \xi_m, \cdots$是独立且同分布的取值都为 $-1, 1$ 的二值随机变量序列:

$$P(\xi_j=1)=P\{f(X^{(j)})=a_1\oplus w^{(1)}\cdot X^{(j)}\}=p_1, j\geqslant 1,$$

因为对任意的 ξ_j, 其数学期望是 $2p_1-1$, 方差是 $1-(2p_1-1)^2=4p_1(1-p_1)$, 由中心极限定理可知, m 充分大时, $\xi_1+\xi_2+\cdots+\xi_m$ 近似服从正态分布 $N(m(2p_1-1), 4mp_1(1-p_1))$. 下面我们来讨论一般的情况: 设 $0\neq w^{(i)}\in GF^n(2)$ 和 $a_i\in GF(2), 1\leqslant i\leqslant k$. $a_i+w^{(i)}\cdot x, 1\leqslant i\leqslant k$ 是与 $f(x)$ 符合概率最大的前 k 个仿射函数. 设 $p_i=P(f(X)=a_i\oplus w^{(i)}\cdot X\}, 1\leqslant i\leqslant k$. 由前述, 若令

$$\xi_{ij}=\begin{cases}1, & \text{若} f(X^{(j)})=a_i\oplus w^{(i)}\cdot X^{(j)}\\ -1, & \text{若} f(X^{(j)})\neq a_i\oplus w^{(i)}\cdot X^{(j)}\end{cases}, j\geqslant 1, 1\leqslant i\leqslant k,$$

则对每个 $1\leqslant i\leqslant k$ 所得是 $\xi_{i1}, \xi_{i2}, \cdots, \xi_{im}, \cdots\cdots$是独立且同分布的取值都为 $-1, 1$ 的二值随机变量序列:

$$P\{\xi_{ij}=1\}=P(f(X^{(j)})=a_i\oplus w^{(i)}\cdot X^{(j)}\}=p_i, j\geqslant 1.$$

同上, 易知

$$\begin{gathered}(f(X^{(1)}), w^{(1)}\cdot X^{(1)}, \cdots, w^{(k)}\cdot X^{(1)}),\\ (f(X^{(2)}), w^{(1)}\cdot X^{(2)}, \cdots, w^{(k)}\cdot X^{(2)}),\\ \vdots\\ (f(X^{(m)}), w^{(1)}\cdot X^{(m)}, \cdots, w^{(k)}\cdot X^{(m)}),\\ \vdots\end{gathered}$$

是独立的 $k+1$ 维布尔随机向量序列. 由此易知

$$(\xi_{11}, \cdots, \xi_{k1}), (\xi_{12}, \cdots, \xi_{k2}), \cdots, (\xi_{1m}, \cdots, \xi_{km})\cdots$$

是独立且同分布的 $k+1$ 维随机变量序列. 令

$$\eta_j=\xi_{1j}+\xi_{2j}+\cdots+\xi_{kj}, j\geqslant 1$$

则 $\eta_1, \eta_2, \cdots, \eta_m, \cdots$是一条独立同分布的随机变量序列. 下面的定理给出了此序列的极限分布.

定理 1 对无记忆组合生成器, 取输入的 k 个仿射函数 $a_i+w^{(i)}\cdot x, 1\leqslant i\leqslant k$, 满足 $p_i=P\{f(X)=a_i+w^{(i)}\cdot X\}, 1\leqslant i\leqslant k$, 则对如上构造的随机变量序列 $\eta_1, \eta_2, \cdots, \eta_m, \cdots$当 m 充分大时, $C_m=\frac{1}{m}(\eta_1+\eta_2+\cdots+\eta_m)$ 近似服从正态分布 $N(\mu_m, \sigma_m^2)$, 其中

$$\mu_m=2\sum_{i=1}^{k}\left(p_i-\frac{1}{2}\right),$$

$$\sigma_m^2=\frac{4}{m}\sum_{i=1}^{k}p_i(1-p_i)-\frac{2}{m}\sum_{1\leqslant i<l\leqslant k}(2p_i-1)(2p_l-1).$$

证明 首先, 对 $j>1$, 求 $\eta_j=\xi_{1j}+\xi_{2j}+\cdots+\xi_{kj}$ 的数字特征. 对 $1\leqslant i<l\leqslant k$ 由布尔随机变量联合分布的分解式[3]知

$$\begin{aligned}P\{f(X)=a_i\oplus w^{(i)}\cdot X, f(X)=a_l\oplus w^{(l)}\cdot X\}&=\frac{1}{2}[P\{f(X)=a_i\oplus w^{(i)}\cdot X\}+P\{f(X)=a_l\oplus w^{(l)}\cdot X\}\\&\quad+\cdots+P\{w^{(i)}\cdot X\oplus w^{(l)}\cdot X=a_i\oplus a_l\}-1]\\&=\frac{1}{2}\left(p_i+p_l-\frac{1}{2}\right),\end{aligned}$$

$$\begin{aligned}P\{f(X)=a_i\oplus w^{(i)}\cdot X, f(X)\neq a_l\oplus w^{(l)}\cdot X\}&=P\{f(X)=a_i\oplus w^{(i)}\cdot X\}-P\{f(X)=a_i\oplus w^{(i)}\cdot X, f(X)=a_l\oplus w^{(l)}\cdot X\}\\&=p_i-\frac{1}{2}\left(p_i+p_l-\frac{1}{2}\right)\\&=\frac{1}{2}\left(p_i-p_l+\frac{1}{2}\right),\end{aligned}$$

同理可得

$$\begin{aligned}P\{f(X)\neq a_i\oplus w^{(i)}\cdot X, f(X)=a_l\oplus w^{(l)}\cdot X\}&=\frac{1}{2}\left(p_l-p_i+\frac{1}{2}\right),\\P\{f(X)\neq a_i\oplus w^{(i)}\cdot X, f(X)\neq a_l\oplus w^{(l)}\cdot X\}&=\frac{1}{2}\left(\frac{3}{2}-p_i-p_l\right).\end{aligned}$$

故对任意的 $j\geqslant 1$ 和 $1\leqslant i<l\leqslant k$, (ξ_{ij}, ξ_{lj}) 的联合分布都是

$$\begin{aligned}P\{\xi_{ij}=-1, \xi_{lj}=-1\}&=\frac{1}{2}\left(\frac{3}{2}-p_i-p_l\right);\\P\{\xi_{ij}=1, \xi_{lj}=-1\}&=\frac{1}{2}\left(p_i-p_l+\frac{1}{2}\right);\\P\{\xi_{ij}=-1, \xi_{lj}=1\}&=\frac{1}{2}\left(p_l-p_i+\frac{1}{2}\right);\\P\{\xi_{ij}=1, \xi_{lj}=1\}&=\frac{1}{2}\left(p_i+p_l-\frac{1}{2}\right).\end{aligned}$$

于是有

$$\begin{aligned}P\{\xi_{ij}+\xi_{lj}=-2\}&=\frac{1}{2}\left(\frac{3}{2}-p_i-p_l\right);\\P\{\xi_{ij}+\xi_{lj}=0\}&=\frac{1}{2};\\P\{\xi_{ij}+\xi_{lj}=2\}&=\frac{1}{2}\left(p_i+p_l-\frac{1}{2}\right).\end{aligned}$$

所以

$$\begin{aligned}\mathrm{Cov}(\xi_{ij}, \xi_{lj})&=E\xi_{ij}\xi_{lj}-E\xi_{ij}E\xi_{lj}\\&=P\{\xi_{ij}\xi_{lj}=1\}-P\{\xi_{ij}\xi_{lj}=-1\}-(2p_i-1)(2p_l-1)\\&=-(2p_i-1)(2p_l-1).\text{所以}\end{aligned}$$

$$E\eta_j=E(\xi_{1j}+\xi_{2j}+\cdots+\xi_{kj})=2\sum_{i=1}^{k}\left(p_i-\frac{1}{2}\right),$$

$$D\eta_j=D(\xi_{1j}+\xi_{2j}+\cdots+\xi_{kj})=\sum_{i=1}^{k}D\xi_{ij}+2\sum_{1\leqslant i<l\leqslant k}\mathrm{Cov}(\xi_{ij}, \xi_{lj})$$

$$=\sum_{i=1}^{k}4p_i(1-p_i)-2\sum_{1\leqslant i<l\leqslant k}(2p_i-1)(2p_l-1).$$

由 $\eta_1, \eta_2, \cdots, \eta_m, \cdots$是一条独立同分布的随机变量序列知

$$\mu_m = E\left[\frac{1}{m}(\eta_1+\eta_2+\cdots+\eta_m)\right] = 2\sum_{i=1}^{k}\left(p_i-\frac{1}{2}\right),$$

$$\sigma_m^2 = D\left[\frac{1}{m}(\eta_1+\eta_2+\cdots+\eta_m)\right]$$

$$= \frac{4}{m}\sum_{i=1}^{k}p_i(1-p_i) - \frac{2}{m}\sum_{1\leqslant i<l\leqslant k}(2p_i-1)(2p_l-1).$$

由中心极限定理, 当 m 充分大时, $\frac{1}{m}(\eta_1+\eta_2+\cdots+\eta_m)$ 近似服从正态分布 $N(\mu_m, \sigma_m^2)$.

3　带记忆组合生成器序列的近似分布

具有 r 比特记忆和 n 个输入的非线性组合生成器定义为:

$$y^{(j)} = V(x^{(j)}, y^{(j-1)}), j\geqslant 1;$$
$$z_j = f(x^{(j)}, y^{(j-1)}), j\geqslant 1.$$

其中, $V: GF^n(2)\times GF^r(2)\rightarrow GF^r(2)$ 为状态向量函数, $f: GF^n(2)\times GF^r(2)\rightarrow GF(2)$ 为输出函数, $y^{(j)} = (y_{1j}, y_{2j}, \cdots, y_{rj})$ 是时刻 j 的状态向量, y^0 为初始状态, $x^j = (x_{1j}, x_{2j}, \cdots, x_{nj})$ 为时刻 j 的输入向量.

我们做如下模型假设: 作为密钥流生成器, 可假设 $f(x_1, x_2, \cdots, x_n, x_{n+1}, \cdots, x_{n+r})$ 是平衡的 $n+r$ 元布尔函数. $V(x_1, x_2, \cdots, x_n, x_{n+1}, \cdots, x_{n+r}) = (g_1(x_1, x_2, \cdots, x_n, x_{n+1}, \cdots, x_{n+r}), \cdots, g_r(x_1, x_2, \cdots, x_n, x_{n+1}, \cdots, x_{n+r}))$ 是平衡的 $n+r$ 元 r 维布尔向量函数. 对输入随机变量序列 $X^{(1)} = (X_{11}, X_{21}, \cdots, X_{n1})$, $X^{(2)} = (X_{12}, X_{22}, \cdots, X_{n2})$, $\cdots$, $X^{(m)} = (X_{1m}, X_{2m}, \cdots, X_{nm})$, …的假设同第 2 节; 设初始状态 $Y_{10}, \cdots, Y_{r0}$ 是相互独立, 都具均匀分布的布尔随机变量, 且 $Y^{(0)} = (Y_{10}, \cdots, Y_{r0})$, $X^{(1)}, X^{(2)}, \cdots, X^{(m)}$, …也是相互独立的.

最后记

$$Y^{(j)} = (g_1(X^{(j)}, Y^{(j-1)}, \cdots, g_r(X^{(j)}, \cdots, Y^{(j-1)})), j = 1, 2, \cdots,$$

$$Z_j = f(X^{(j)}, Y^{(j-1)}),\quad j = 1, 2, \cdots,$$

由假设可推知对每个 $j\geqslant 1$, $Y^{(j)}$ 都与 $Y^{(0)}$ 一样, 是分量之间相互独立且每个分量都具有均匀分布的 r 维布尔随机向量; 而且对每个 $j\geqslant 1$, Z_j 都是具有均匀分布的布尔随机变量.

带记忆组合生成器是对无记忆组合生成器的推广, 通过选取适当的输出函数和状态向量函数, 带记忆组合生成器可以使 j 时刻的输出与以前所有的输入都没有相关性, 即 Z_j 与 $(X^{(1)}, \cdots, X^{(j)})$ 相互独立. 但 Golic[9] 的结果表明: 对带 r 比特记忆的组合生成器, 连续 $r+1$ 个输出与对应的连续 $r+1$ 组输入之间必然存在相关性. 所以对带记忆组合生成器做相关攻击, 一般首先要在连续的一段输入与输出之间找具有大的相关性的线性对. 我们把关于连续 t 长输入和输出的仿射函数称为组合生成器的 t 长仿射函数, 注意 $t-1$ 长仿射函数也可看作是 t 长仿射函数.

设 $A_1^t, \cdots, A_k^t$ 是 k 个 t 长仿射函数, 即 k 个关于 $t+mt$ 个变元的仿射函数. 对 $1\leqslant l\leqslant k$, 由 A_l^t 和 $Z_j, \cdots, Z_{j+t-1}, X^{(j)}, \cdots, X^{(j+t-1)}$ 可定义布尔随机变量

$$A_{lj}^t = A_l^t(Z_j, \cdots, Z_{j+t-1}, X^{(j)}, \cdots, X^{(j+t-1)}), j\geqslant 1.$$

从而得布尔随机变量序列 $A_{l1}^t, A_{l2}^t, \cdots, A_{lm}^t, \cdots$, 由模型可知此序列是同分布的, 设 $P\{A_{lj}^t = 0\} = p_l, 1\leqslant l\leqslant k, j\geqslant 1$, 我们把 A_{lj}^t 与 0 的符合概率 p_l 简称为仿射函数 A_l^t 的符合概率. 令 $\xi_{lj} = \begin{cases}1, \text{若 } A_{lj}^t = 0;\\ -1, \text{若 } A_{lj}^t = 1\end{cases}$ $1\leqslant l\leqslant k, j\geqslant 1$. 则对 $1\leqslant l\leqslant k$, $\xi_{l1}, \xi_{l2}, \cdots, \xi_{lm}, \cdots$是同分布的取值都为 $-1, 1$ 的二值随机变量序列, 满足: $P\{\xi_{lj} = 1\} = p_l, j\geqslant 1$. 令 $\eta_j = \xi_{1j}+\xi_{2j}+\cdots+\xi_{kj}, j\geqslant 1$, 则 $\eta_1, \eta_2, \cdots, \eta_m, \cdots$是一条同分布的随机变量序列. 类似无记忆的情况, 我们希望得到此序列的极限分布. 根据模型假设, 一般而言, $\eta_1, \eta_2, \cdots, \eta_m, \cdots$并不是一条独立的随机变量序列, 但是实际中使用的带记忆组合生成器, 作为安全的密钥流生成器, 序列 $\eta_1, \eta_2, \cdots, \eta_m, \cdots$的相依关系应该比较弱, 可近似看作一条独立的随机变量序列. 所以当 $A_{1j}^t, A_{2j}^t, \cdots, A_{kj}^t$ 的任意非零线性和都平衡时, 由定理 1 的证明过程可知, 当 m 充分大时, 我们可以假定随机变量 $C_m = \frac{1}{m}(\eta_1, \eta_2, \cdots, \eta_m)$ 近似服从正态分布 $N(\mu_m, \sigma_m^2)$, 其中

$$\mu_m = 2\sum_{i=1}^{k}\left(p_i-\frac{1}{2}\right),$$

$$\sigma_m^2 = \frac{4}{m}\sum_{i=1}^{k}p_i(1-p_i) - \frac{2}{m}\sum_{1\leqslant i<l\leqslant k}(2p_i-1)(2p_l-1).$$

4　组合生成器的多线性相关攻击

下面我们对组合生成器描述一种一般形式的相关攻击方法, 旨在说明如何利用输入和输出的多个线性关系的信息来约减攻击所需的数据量. 由前两节的分析可看出, 对带记忆组合生成器进行相关攻击, 要在连续 t 长的一段输入和输出间找符合概率大的线性关系, 而无记忆组合生成器可看成是特殊的带记忆组合生成器, 对无记忆组合生成器做相关攻击, 即令 $t=1$, 所以我们只需对带记忆的情况进行说明即可.

4.1　参数假定

对如上定义的具有 r 比特记忆和 n 个输入的非线性组合生成器, 我们假定已知如下参数:

(1)所有 n 个移位寄存器的生成多项式 f_i 和级数 r_i ($i=1, 2, \cdots, n$);

(2)输出函数 f 和状态向量函数 V;

(3)一定长的密钥流序列 $z_j, z_{j+1}, \cdots$.

4.2　目的

利用输出序列和输入序列之间的相关性还原密钥流生成器的初态 α_{ij} ($i=1, \cdots, n; j=1, \cdots, r_i$).

4.3　方法

(1)对某个 $t\geqslant 1$, 寻找 k 个 t 长仿射函数 $A_1^t, \cdots, A_k^t$, 满足对 $j\geqslant 1$, $P\{A_{lj}^t = 0\} = p_l, p_l > \frac{1}{2}, 1\leqslant l\leqslant k$, 且 $A_{1j}^t, A_{2j}^t, \cdots, A_{kj}^t$ 的任意非零线性和都是平衡的;

(2)由于每 t 组输入 $x^{(j)}, \cdots, x^{(j+t-1)}$ 均可由移位寄存器的初态唯一线性表出, 所以对每个时刻 j 我们可以以一定的概率列出如下方程组:

$$\sum_{i=1}^{n}\sum_{m=1}^{r_i}b_{lm}^{ij}\alpha_{im} = \sum_{i=0}^{t-1}v_{j+i}^{l}z_{j+i},\quad l=1, 2, \cdots, k \qquad (1)$$

其中系数$(b_{i1}^{lj}, \cdots, b_{ir_i}^{lj})$和$(v_j^l, \cdots, v_{j+t-1}^l)$由生成多项式$f_i$和仿射函数$A_l^t$决定. 记

$$\alpha=(\alpha_{11}, \cdots, \alpha_{1r_1}, \cdots, \alpha_{n1}, \cdots, \alpha_{nr_n}),$$

$$B_j=\begin{pmatrix} b_{11}^{1j} & \cdots & b_{1r_1}^{1j} & \cdots & b_{n1}^{1j} & \cdots & b_{nr_n}^{1j} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{11}^{kj} & \cdots & b_{1r_1}^{kj} & \cdots & b_{n1}^{kj} & \cdots & b_{nr_n}^{kj} \end{pmatrix},$$

$$z_j^{j+t-1}=(z_j, \cdots, z_{j+t-1}),$$

$$V_j=\begin{pmatrix} v_j^1 & \cdots & v_{j+t-1}^1 \\ \vdots & \vdots & \vdots \\ v_j^k & \cdots & v_{j+t-1}^k \end{pmatrix}.$$

则方程组(1)可表成

$$B_j\alpha^{\mathrm{T}}=V_jz_{j,t}^T.$$

(3)令$j=j+1$, 重复第(2)步N次, 即得到一个含kN个方程和$R\sum_{i=1}^{n}r_i$个变量的方程组

$$B_{j+i}\alpha^{\mathrm{T}}=V_{j+i}z_{j+i,t}^T,\ i=0, 1, \cdots, N-1 \tag{2}$$

(4)构造统计量

$$C_N(\alpha)=\frac{1}{N}\sum_{i=0}^{N-1}(\#\{B_{j+i}\alpha^T=V_{j+i}z_{j+i,t}^T\}-\#\{B_{j+i}\alpha^T\neq V_{j+i}z_{j+i,t}^T\}).$$

解方程组(2)得解集$B=\{\alpha=(\alpha_{11}, \cdots, \alpha_{1r_1}, \alpha_{n1}, \cdots, \alpha_{nr_n})\mid C_N(\alpha)>T\}$, 其中$T$为决策门限值;

(5)验证第(4)步中所得解集中的解, 如果找到正确的初态, 则结束; 否则, 适当增大N, 返回(3).

4.3 数据量分析

做如下假设检验问题:

H_0: α是正确的初态; H_1: α不是正确的初态.

由第二节和第三节的分析知: 若α为正确的初态时, 当N充分大, $C_N(\alpha)$近似服从正态分布$N(\mu_0, \sigma_0^2)$, 其中, $\mu_0=2\sum_{i=1}^{k}\left(p_i-\frac{1}{2}\right)$, $\sigma_0^2=\frac{4}{N}\sum_{i=1}^{k}p_i(1-p_i)-\frac{2}{N}\sum_{1\leqslant i<l\leqslant k}(2p_i-1)(2p_l-1)$. 若α不是正确的初态, 此时得到的输入与我们观测到的密钥流序列相互独立, 即相当于$p_1=\cdots=p_k=\frac{1}{2}$, 当$N$充分大, $C_N(\alpha)$近似服从正态分布$N(\mu_1, \sigma_1^2)$, 其中, $\mu_1=0$, $\sigma_1^2=\frac{k}{N}$.

设决策门限值为T, 当$C_N(\alpha)\geqslant T$时, 接受H_0; 当$C_N(\alpha)<T$时接受H_1. 我们可能犯两类错误: "假真错误"和"真假错误", 称犯假真错误的概率为取伪概率, 记为P_f, 称犯真假错误的概率为弃真概率, 记为P_m, 则

$$P_f=\int_T^{\infty}\frac{1}{\sigma_1\sqrt{2\pi}}\exp\left[\frac{-(x-\mu_1)^2}{2\sigma_1^2}\right]dx,$$

$$P_m=\int_{-\infty}^{T}\frac{1}{\sigma_0\sqrt{2\pi}}\exp\left[\frac{-(x-\mu_0)^2}{2\sigma_0^2}\right]dx.$$

引入函数$Q(x)=\frac{1}{\sqrt{2\pi}}\int_x^{\infty}\exp\frac{-y^2}{2}dy$, 则有

$$P_f=Q\left(T\sqrt{\frac{N}{K}}\right) \tag{3}$$

$$P_m=Q\left(\frac{2\sum_{i=1}^{k}\left(p_i-\frac{1}{2}\right)-T}{\left[4\sum_{i=1}^{k}p_i(1-p_i)-2\sum_{1\leqslant i<l\leqslant k}(2p_i-1)(2p_l-1)\right]^{1/2}}\sqrt{N}\right) \tag{4}$$

此处我们主要关心取伪概率P_f, 因为要恢复的初态总长度为$R=r_1+\cdots+r_n$, 为了保证攻击算法第四步所得的解集包含真解, 我们需要足够长的密钥流, 即足够大的N, 使$P_f\leqslant\frac{1}{2^R}$, 一般取$P_f=\frac{1}{2^R}$即可, 同时限定真假错误概率$P_m$, 并查表得$w_f$和$w_m$满足: $\frac{1}{2^R}=Q(w_f)$, $P_m=Q(w_m)$. 由此利用式(3)和(4)我们就可以估计所需密钥流的长度N.

定理**2** 在对组合生成器的如上攻击算法中, 限定真假错误概率为P_m, 由正态分布表得w_f和w_m满足: $\frac{1}{2^R}=Q(w_f)$, $P_m=Q(w_m)$, 则攻击算法中的决策门限

$$T=\frac{2w_f\sqrt{k}\sum_{i=1}^{k}\left(p_i-\frac{1}{2}\right)}{w_f\sqrt{k}-w_m\left[4\sum_{i=1}^{k}p_i(1-p_i)+2\sum_{1\leqslant i<l\leqslant k}(2p_i-1)(2p_l-1)\right]^{1/2}},$$

所需密钥流长度

$$N=\left(\frac{w_f\sqrt{k}+w_m\left[4\sum_{i=1}^{k}p_i(1-p_i)-2\sum_{1\leqslant i<l\leqslant k}(2p_i-1)(2p_l-1)\right]^{1/2}}{2\sum_{i=1}^{k}\left(p_i-\frac{1}{2}\right)}\right)^2$$

下面我们分析攻击算法中采用的仿射函数的个数k及其符合概率$p_1, p_2, \cdots, p_k$对所需数据量N, 和所要解方程组的方程量的影响. 我们取的仿射函数都满足$p_i>\frac{1}{2}$, $1\leqslant i\leqslant k$. 称$\alpha_i=p_i-\frac{1}{2}$为第i个仿射函数的符合优势, $1\leqslant i\leqslant k$, 并设$a=\min\{a_i, \cdots, a_k\}$. 事实上由式(3)和(4)可推得

$$\sqrt{N}=\frac{w_f\sqrt{k}+w_m\left[4\sum_{i=1}^{k}p_i(1-p_i)-2\sum_{1\leqslant i<l\leqslant k}(2p_i-1)(2p_l-1)\right]^{1/2}}{2\sum_{i=1}^{k}\left(p_i-\frac{1}{2}\right)}$$

$$\leqslant\frac{w_f\sqrt{k}+w_m\left[4\sum_{i=1}^{k}p_i(1-p_i)\right]^{1/2}}{2\sum_{i=1}^{k}\left(p_i-\frac{1}{2}\right)}$$

$$\leqslant\frac{(w_f+w_m)\sqrt{k}}{2\sum_{i=1}^{k}a_i}\leqslant\frac{w_f+w_m}{2a\sqrt{k}}.$$

所以

$$N\leqslant\frac{(w_f+w_m)^2}{4a^2k} \tag{5}$$

所要解方程组的方程量

$$kN\leqslant\frac{(w_f+w_m)^2}{4a^2} \tag{6}$$

注意到对给定的组合生成器, 初态长度 R 是定值, 而 w_f 由 $Q(w_f)=\frac{1}{2^R}$ 决定, 所以 w_f 是一个定值. 对限定的真假错误概率 P_m, 由 $P_m=Q(w_m)$ 知 w_m 也是定值, 由式(5)知, 相关攻击所需的数据量随着 a^2k 的增大而减少. 也就是说, 所用的仿射函数的符合优势越大, 个数越多, 所需数据量越少. 另外, 由式(6)知, 如果 a 不变, k 的变化对方程量的影响不大, 方程量主要与符合概率 a 有关, 它与 a^2 成反比. 下面我们以攻击“蓝牙生成器”为例, 来说明上述方法的实际效果.

5　蓝牙生成器的多线性相关攻击

“蓝牙协议”是一个短距离无线通信协议, 1999 年“蓝牙特别兴趣组”公布了“蓝牙技术标准 1.0”[1], 其安全机制中采用的密钥流生成算法 E_0, 事实上就是一个有 4 个输入带 4bit 记忆的组合生成器, 被称作“蓝牙组合生成器”[2]. 作为输入的 4 个 LFSR 的长度分别为 25, 31, 33, 39, 其反馈多项式均是本原的, 抽头数均为 5 个. 记 j 时刻四个移位寄存器的输出为 $x^{(j)}=(x_{1j}, x_{2j}, x_{3j}, x_{4j})$, 状态向量为 $c^{(j)}=(c_{j+1}^0, c_{j+1}^1, c_j^0, c_j^1)$, 生成器的输出为 z_j. 则此生成器的定义如下:

$$z_j=x_{1j}\oplus x_{2j}\oplus x_{3j}\oplus x_{4j}\oplus c_j^0,$$
$$c_{j+1}^0=s_{j+1}^0\oplus c_j^0\oplus c_{j-1}^0\oplus c_{j-1}^1,$$
$$c_{j+1}^1=s_{j+1}^1\oplus c_j^1\oplus c_{j-1}^0,$$

其中 $(s_{t+1}^1, s_{t+1}^0)=\lfloor(x_{1j}+x_{2j}+x_{3j}+x_{4j}+2c_j^1+c_j^0)/2\rfloor\in\{0,1,2,3\}$, $(c_0^0, c_0^1, c_{-1}^0, c_{-1}^1)$ 为 4 比特记忆的初态.

按照第 3 节一般的带记忆组合生成器的模型, 蓝牙生成器可表示如下形式:

输出函数为 $z_j=f(x^{(j)}, c^{(j-1)})=x_{1j}\oplus x_{2j}\oplus x_{3j}\oplus x_{4j}\oplus c_j^0$. 状态向量函数为

$$c^{(j)}=V(x^{(j)}, c^{(j-1)})=(g_1(x^{(j)}, c^{(j-1)}), g_2(x^{(j)}, c^{(j-1)}), g_3(x^{(j)}, c^{(j-1)}), g_4(x^{(j)}, c^{(j-1)}))$$
$$g_1(x^{(j)}, c^{(j-1)})=c_j^1\oplus H_1(j)\,c_j^0\oplus H_2(j)\oplus c_j^0\oplus c_{j-1}^0\oplus c_{j-1}^1,$$
$$g_2(x^{(j)}, c^{(j-1)})=H_1(j)\,c_j^0c_j^1\oplus H_1(j)\,c_j^1\oplus H_3(j)\,c_j^0\oplus H_4(j)\oplus c_j^1\oplus c_{j-1}^0,$$
$$g_3(x^{(j)}, c^{(j-1)})=c_j^0,$$
$$g_4(x^{(j)}, c^{(j-1)})=c_j^1.$$

其中

$$H_1(j)=x_{1j}\oplus x_{2j}\oplus x_{3j}\oplus x_{4j},$$
$$H_2(j)=x_{1j}x_{2j}\oplus x_{1j}x_{3j}\oplus x_{1j}x_{4j}\oplus x_{2j}x_{3j}\oplus x_{2j}x_{4j}\oplus x_{3j}x_{4j},$$
$$H_3(j)=x_{1j}x_{2j}x_{3j}\oplus x_{1j}x_{2j}x_{4j}\oplus x_{1j}x_{3j}x_{4j}\oplus x_{2j}x_{3j}x_{4j},$$
$$H_4(j)=x_{1j}x_{2j}x_{3j}x_{4j}.$$

蓝牙生成器的初态总长为 $R=25+31+33+39=128$ 比特. 其运行过程分为两级: 第一级是初始化, 产生 128 比特初态, 第二级产生长度不超过 2745 比特的密钥流, 然后又重新初始化.

我们下面分析如何用上一节的多线性相关攻击方法来攻击蓝牙生成器的第二级, 即利用不超过 2745 比特的密钥流恢复 128 比特初态.

首先我们利用文[6]中的相关系数计算方法, 通过计算我们发现对蓝牙生成器所有仿射函数所能达到的最大符合概率为 0.54883(记为 p_1), 符合概率达到此值的仿射函数总共有 32 个, 包括 16 个 5 长仿射函数和 16 个 6 长仿射函数. 另外, 符合概率达到次大值 0.53125(记为 p_2)的仿射函数有 96 个, 都是 4 长仿射函数. 我们可以把这 128 个函数都看成 6 长仿射函数. 32 个符合概率达到最大值的仿射函数的任意非零相性和都是平衡的, 验证所有这 128 个函数的任意非零线性和是否都平衡比较困难, 但是对蓝牙生成器, 所有的 6 长仿射函数中, 除了这 128 个以外, 其它的符合概率都与 1/2 非常接近, 所以我们总可以使用第三节中得到的近似分布.

因为初态长为 128, 所以此处取伪概率 $p_f=\frac{1}{128}$, 为稳妥我们相应的取 $w_f=14$, 查正态分布表可知 $Q(14)<\frac{1}{128}$, 设定真假错误概率 $p_m=0.01$, 查表得 $w_m=2.33$, $Q(2.33)=0.01$.

若按照传统的方法, 只取一个符合概率达到最大值 p_1 的仿射函数做相关攻击, 由定理 2 计算可知需要的数据量为 27921 比特, 方程量为 27921. 但是蓝牙生成器最多输出 2745 比特密钥就重新初始化, 所以我们考虑使用多个仿射函数.

由定理 2, 通过计算可知当用 11 个符合概率达到 p_1 的仿射函数做相关攻击时, 所需的数据量为 2503 比特, 低于蓝牙生成器的最大输出长度, 此时所需解方程组的方程量为 2503 × 11= 27533.

我们把各种情况下作相关攻击所需数据量和所解方程组的方程量在表 1 中列出. 由上表可看出, 随着所用符合概率达到 p_1 的仿射函数的个数的增加, 所需数据量降低, 所解方程组的方程量也稍有减少. 当加入符合概率为 p_2 的仿射函数后, 虽然所需数据量减少, 但减少的速度变慢, 而且方程量增加了, 这是因为 p_2 比较小的缘故. 这与我们在第四节的分析相吻合.

表 1　仿射函数个数与相关攻击所需数据量和方程量的关系

所用仿射函数	数据量	方程量
1 个符合概率达到 p_1 的仿射函数	27921	27921
11 个符合概率达到 p_1 的仿射函数	2053	27533
32 个符合概率达到 p_1 的仿射函数	833	26656
全部 128 个仿射函数(32 个符合概率为 p_1 的, 96 个符合概率为 p_2)	388	49664

6　结论

本文从理论上分析了同时利用多个线性关系对组合生成器进行相关攻击的可行性, 给出了一种一般性的攻击方法, 证明了该方法可以有效的约减攻击所需的数据量, 对蓝牙组合生成器的分析进一步说明了这种攻击的效果. 但是需要注意, 第 5 节表 1 所列的数据是理论上的, 前提条件是攻击者有足够的计算资源和存储资源. 而对一种具体的密码体制的攻击, 需要根据攻击者的能力寻求计算复杂度和空间复杂度的折衷.

现在对蓝牙生成器的攻击结果很多,但都还是理论上的.如 Ekdahl 和 Johansson[5]的方法在已知 $O(2^{50})$ 的明文的条件下得到 128 比特密钥的复杂度为 $O(2^{60})$. Golic[6]的方法在拥有 2^{80} 的 103bit 字库的条件下利用 45 个数据包做攻击的复杂度为 $O(2^{70})$. Fluhrer 和 Lucks[7]的方法在已知 132 比特明文时得到 128 比特密钥的复杂度为 $O(2^{84})$,在已知 2^{43} 比特明文时的复杂度为 $O(2^{73})$. Canniere 等[8]的方法在已知大约 $O(2^{32})$ 比特明文的前提下恢复 128 比特密钥的复杂度为 $O(2^{76})$. 我们基于本文的多线性相关攻击思想,利用蓝牙组合生成器在已知输出的条件下输入的多个线性函数与零函数的相关性,并结合卷积码快速相关攻击算法[9]对其进行了攻击,结果表明需要大约 $O(2^{34.311})$ 输出序列,$O(2^{31.311})$ 存储空间,攻击复杂度约为 $O(2^{64.2})$,成功概率 $\leqslant 0.99999994$,攻击的综合效率优于以往的攻击方法.

参考文献:

[1] Bluetooth™ SIG. The Bluetooth Specification Version1.0[S]. 1999.

[2] HERMELIN M, NYBERG K. Correlation properties of bluetooth combiner generator[A]. SONG J. The 2th International Conference on Information Security and Cryptology (ICISC' 99)[C]. LNCS1787. Berlin: Springer-Verlag, 2000. 17-29.

[3] 李世取, 曾本胜, 等. 密码学中的逻辑函数[M]. 北京: 中软电子出版社, 2003.

[4] 张卫明, 李世取. 带记忆组合生成器的相关免疫性[A]. 王育民. 密码学进展——Chinacrypt' 2002[C]. 北京: 电子工业出版社, 2002. 21-30.

[5] EKDAHL K, JOHANSSON T. Some Results on Correlations in the Bluetooth Stream Cipher[EB/OL]. http://www.it.lth.se/patrik/papers/bluetooth.ps, 2004-11-7.

[6] GOLIC J DJ, BAGINI V, MORGARI G. Linear cryptanalysis of bluetooth stream cipher[A]. KNUDSEN L. EUROCRYPT 2002[C]. LNCS 2332. Berlin: Springer-Verlag, 2002. 238-255.

[7] FLUHRER S R, LUCKS S. Analysis of the E0 encryption system[A]. SERGE V. Selected Areas in Cryptography-SAC2001[C]. LNCS 2259. Berlin: Springer-Verlag, 2001. 38-41.

[8] CANNIERE C D, JOHANSSON T, PRENEEL B. Cryptanalysis of the Bluetooth Stream Cipher [EB/OL]. http://www.cosic.esat.kuleuven.ac.be/publications/article-22.pdf, 2004-11-7.

[9] JONSSON F. Some Results on Fast Correlation Attacks[D]. Lund Sweden: Department of Information Technology, Lund University, 2002.

作者简介:

张卫明 男, 1976 年生于河北省定州市, 现为解放军信息工程大学信息研究系博士生, 主要研究方向为密码学和信息隐藏. E-mail: nlxdweiming@sohu.com

李世取 男, 1945 年生于重庆市, 现为解放军信息工程大学信息研究系教授, 博士生导师, 主要研究方向为密码学理论.