

一类密钥流生成器的相关分析¹

马卫局* 冯登国** 巫治平* 张斌***

*(中国科学院研究生院信息安全国家重点实验室 北京 100039)

** (中国科学院软件研究所 北京 100080)

*** (新加坡国立大学信息通信实验室 新加坡 119613)

摘要: 多输出逻辑函数是构造密码系统的重要工具, 相关免疫性是设计安全逻辑函数的重要准则. 该文利用一种较为简单的方法证明了多输出逻辑函数相关免疫性两种刻划的等价性. 还对一类利用多输出逻辑函数相关免疫函数构造的密钥流生成器进行了相关性分析, 证明了这种构造方法是不成立的, 并不能达到构造者期望的相关免疫性, 并且分别利用 Walsh 变换技术和线性序列电路逼近方法找出了这类密钥流生成器的漏洞, 从而说明这类生成器在相关攻击下是脆弱的.

关键词: 相关系数, 相关免疫性, 密钥流生成器

中图分类号: TN918.1 **文献标识码:** A **文章编号:** 1009-5896(2004)08-1282-06

Cryptanalysis on a Kind of Keystream Generator

Ma Wei-ju* Feng Deng-guo** Wu Zhi-ping* Zhang Bin***

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

** (Institute of Software, The Chinese Academy of Sciences, Beijing 100080, China)

*** (Institute for Info. Communication Research, National Univ. of Singapore 119613)

Abstract Multiple outputs logic function is an important tool for constructing cryptography systems, and correlation immunity is a critical criterion in designing secure logic functions. In this paper, a very simple method is used to prove that two descriptions of correlation immunity of multiple outputs logic function are equivalent indeed. The correlation properties of a kind of keystream generator are analyzed, which is defined by multiple outputs logic functions. It is found that this constructing method is wrong, which means that the immunity expected by the construction cannot be obtained. Walsh transform technique and linear sequential circuit approximation method are applied to find the weakness of this kind keystream generator respectively. That is to say the keystream generator constructed by this method is vulnerable.

Key words Correlation coefficients, Correlation immunity, Keystream generator

1 引言

在流密码的设计当中, 我们经常以线性反馈移位寄存器 (LFSR) 作为输入, 非线性函数作为组合器来形成密钥流. Siegenthaler 在文献 [1] 中发现如果输出的密钥流中如果泄露了某一个移位寄存器的信息的话, 就可以对这个密钥流生成器进行分别征服攻击. 为此文献 [2, 3] 提出了相关免疫性的概念, 从此相关免疫性成为设计流密码的一个重要指标.

多输出逻辑函数在私钥体制的设计中也有非常广泛的应用, 例如用来设计分组密码. 在流密码的设计中如果我们利用多输出逻辑函数, 虽然安全性会有所降低, 但每个时刻可以产生多

¹ 2003-04-18 收到, 2003-08-07 改回

国家攀登计划 973 项目 (G1999035802) 和国家杰出青年科学基金 (60025205) 资助课题

个密钥比特, 就会提高软硬件实现的速度. 因此对多输出逻辑的相关免疫性的研究也是一个重要的课题. 文献 [4-6] 分别引入了两种不同的多输出相关免疫函数的概念. 文献 [7] 对这两种概念进行了等价性证明, 但证明方法较为复杂, 本文在第 2 节中给出了一种更为简单的证明. 第 3 节我们介绍了有记忆的密钥流生成器. 文献 [7] 还利用多输出相关免疫函数构造了一种密钥流生成器, 认为它可以获得很好的相关免疫性, 但本文在第 4 节中指出它就是一种有记忆的多输出密钥流生成器, 并证明了这种组合器在记忆比特位数小于等于输入的个数的时候, 这种相关免疫性是不成立的. 文中还针对这种组合器给出了几种不同的相关性分析方法. 结论在第 5 节中给出.

2 两种相关免疫概念的等价证明

首先给出两种多输出相关免疫函数的概念.

定义 1^[4] 设 x_1, \dots, x_n 是 n 个相互独立的均匀分布二元随机变量, $F(x_1, \dots, x_n): F_2^n \rightarrow F_2^m$ 是一个多输出逻辑函数, $1 \leq t \leq n$, 如果对于任意的 $1 \leq i_1 < \dots < i_t \leq n$, 有 $F(x_1, \dots, x_n)$ 与随机向量 $(x_{i_1}, \dots, x_{i_t})$ 统计独立, 则称 F 是 (n, m, t) 第一类相关免疫函数.

定义 2^[6] 设 x_1, \dots, x_n 是 n 个相互独立的均匀分布二元随机变量, $F(x_1, \dots, x_n): F_2^n \rightarrow F_2^m$ 是一个多输出逻辑函数, 如果对于任意 $v \in F_2^m$, $v \cdot F$ 都是 t 阶相关免疫的布尔函数, 则称 F 是 (n, m, t) 第二类相关免疫函数.

文献 [7] 证明了这两个定义是等价的, 但方法较为复杂, 以下我们给出一种较为简单的证明.

引理 1 一个多输出函数 $F(X, Y)$ 与 X 统计独立当且仅当 F 的分量的所有非平凡线性组合与 X 的所有线性函数统计独立.

证明 设 $F = (f_1, \dots, f_m)$, $X \in F_2^{n_1}$, $Y \in F_2^{n_2}$.

必要性 $\forall c, d \in F_2, \alpha, Z \in F_2^m, \omega, V \in F_2^{n_1}$

$$\begin{aligned} P\left\{\sum_{k=1}^m \alpha_k f_k = c, \omega \cdot X = d\right\} &= \sum_{\{Z | Z \cdot \alpha = c\}} \sum_{\{V | V \cdot \omega = d\}} P\{F(X, Y) = Z, X = V\} \\ &= \sum_{\{Z | Z \cdot \alpha = c\}} \sum_{\{V | V \cdot \omega = d\}} P\{F(X, Y) = Z\} P\{X = V\} \\ &= P\left\{\sum_{k=1}^m \alpha_k f_k = c\right\} P\{\omega \cdot X = d\} \end{aligned}$$

充分性 $\forall Z \in F_2^m, V \in F_2^{n_1}$, 因为 $\alpha \cdot F$ 与 $\omega \cdot X$ 独立, 故有

$$\begin{aligned} P\{\alpha \cdot F(X, Y) = \alpha \cdot Z + d, \omega \cdot X = \omega \cdot V + c - d\} &= P\{\alpha \cdot F(X, Y) \\ &= \alpha \cdot Z + d\} P\{\omega \cdot X = \omega \cdot V + c - d\} \end{aligned}$$

故

$$\begin{aligned}
 P\{F(X, Y) = Z, \omega \cdot X = V\} &= \frac{1}{2^{m+n}} \sum_{(\alpha, \omega) \in F_2^{m+n+1}} \sum_{c=0}^1 P\{\alpha \cdot F(X, Y) + \omega \cdot X \\
 &= \alpha \cdot Z + \omega \cdot V + c\} (-1)^c \\
 &= \left(\frac{1}{2^m} \sum_{\alpha \in F_2^m} \sum_{d=0}^1 P\{\alpha \cdot F(X, Y) = \alpha \cdot Z + d\} (-1)^d \right) \\
 &\quad \times \left(\frac{1}{2^{n+1}} \sum_{\omega \in F_2^{n+1}} \sum_{c=0}^1 P\{\omega \cdot X = \omega \cdot V + c - d\} (-1)^{c-d} \right) \\
 &= P\{F(X, Y) = Z\} P\{X = V\}
 \end{aligned}$$

定理 1 设 x_1, \dots, x_n 是 n 个相互独立的均匀分布二元随机变量, $F(x_1, \dots, x_n) : F_2^n \rightarrow F_2^m$ 是一个多输出逻辑函数, 则 F 是 (n, m, t) 第一类相关免疫函数当且仅当 F 是 (n, m, t) 第二类相关免疫函数.

证明 此定理是引理 1 的直接推论.

3 带记忆的密钥流生成器

我们称仅以非线性函数作为组合器的密钥流生成器为无记忆的密钥流生成器. 在无记忆的密钥流生成器中, Meier 和 Staffelbach 在文献 [8] 中指出输出比特与所有输入的线性函数之间的相关系数的平方和满足

$$\sum_i c_i^2 = 1 \quad (1)$$

选择具有一定相关免疫阶的组合函数就意味着某些 c_i 会消失. 然而由式 (1), 这就导致其他一些相关系数的增加. 所以在无记忆的组合器中, 相关免疫性与非线性度总有一定的制约关系. Rueppel 在文献 [9] 中建议引入记忆来消除这种制约关系. 他发现只需要引入 1 bit 记忆就可以使相关免疫阶和非线性度同时达到最大. 并且他给出求和生成器作为例子.

一个有 N 个输入、 M 比特记忆的密钥流生成器定义如下:

$$S_{t+1} = F(X_t, S_t), \quad t \geq 0 \quad (2)$$

$$y_t = f(X_t, S_t), \quad t \geq 0 \quad (3)$$

其中 $F : F_2^N \times F_2^M \rightarrow F_2^M$ 是状态向量函数, $f : F_2^N \times F_2^M \rightarrow F_2$ 是输出函数. $S_t = (s_{1,t}, \dots, s_{M,t})$ 是在时刻 t 的状态向量, S_0 是初态, $X_t = (x_{1,t}, \dots, x_{N,t})$ 是在时刻 t 的输入向量, y_t 是时刻 t 时的输出比特. 在理论分析中我们总假设输入是相互独立且平衡的随机变量序列. Golić 在文献 [10] 中分析了有记忆密钥流生成器的相关性. 他证明了必然存在一个 $M+1$ 个连续输出比特的线性函数与某个 $M+1$ 个连续输入的线性函数统计相关. 并对有记忆的密钥流生成器提出了线性序列电路逼近的相关性分析方法.

4 一类密钥流生成器的相关性分析

文献 [7] 中利用多输出相关免疫函数构造了一类密钥流生成器, 它的输出函数和状态函数分别如下:

$$Y_t = F(X_t, S_t), \quad t \geq 0 \quad (4)$$

$$S_{t+1} = G(X_t, S_t), \quad t \geq 0 \quad (5)$$

其中 $F(\mathbf{X}_t, \mathbf{S}_t)$ 是具有 N 个输出的多值逻辑函数, $G(\mathbf{X}_t, \mathbf{S}_t)$ 是一个具有 M 个输出的多值逻辑函数, 在时刻 t 的输出向量为 $\mathbf{Y}_t = (y_{1,t}, y_{2,t}, \dots, y_{N,t})$, 输入向量为 \mathbf{X}_t , 记忆向量为 \mathbf{S}_t . 文献 [7] 认为 (F, G) 可以是 $(N + M, N + M, N)$ 相关免疫函数, 因此该密钥流生成器输出符号 f_1, f_2, \dots, f_N 的任意线性组合 $\bigoplus_{i=1}^N u_i f_i$, $u_i \in F_2$ 与各线性反馈移位寄存器的输出符号 x_1, x_2, \dots, x_n 的任意线性组合 $\bigoplus_{i=1}^N v_i x_i$, $v_i \in F_2$ 之间的相关系数都为 0, 这样就可以避免利用相关攻击或者线性逼近攻击来获得线性反馈移位寄存器的初态, 即该密钥流生成器的密钥. 然而经过我们的分析, 这种组合器并不能达到该文作者预期的目的, 而且在安全性上存在着很大的漏洞. 以下, 我们给出 3 种不同的相关性分析方法.

4.1 当前的输出与当前的输入相关

首先我们给出关于多输出相关免疫函数的一个定理.

定理 2^[11] 对于任何 (n, m) , $n - 2 \geq m > 1$, 不存在 $(n, m, n - m)$ 相关免疫函数.

由定理 2, 我们得知对于上述密钥流生成器, 当记忆比特个数 M 小于等于输入个数 N 的时候, 输出向量的函数 (f_1, f_2, \dots, f_N) 并不是 $(N + M, N, N)$ 相关免疫函数. 也就是说必然存在输出符号 f_1, f_2, \dots, f_N 的某个线性组合 $\bigoplus_{i=1}^N u_i f_i$ 与各线性反馈移位寄存器的输出符号 x_1, x_2, \dots, x_N 的某个线性组合 $\bigoplus_{i=1}^N v_i x_i$ 之间的相关系数不为 0. 这样, 我们就可以利用快速相关攻击得到线性移位寄存器的初态. 这个算法的复杂度为 $O[(M + N)2^{M+2N}]$.

4.2 穷举 Walsh 谱值

如果 4.1 节的方法获得的相关系数比较小, 或者是当 $M > N$ 时, 该方法对于快速相关攻击就不是十分有效了. 通常的情况下, 我们预期的效果是获得具有最大绝对值的相关系数的输出输入线性函数对, 有时候我们还希望获得不止一对具有较大绝对值的相关系数的输出输入线性函数. 给定一个输出的线性函数, 它与一个输入的线性函数之间的相关系数可以通过 Walsh 变换技术获得. 对于任意的正整数 m 以及任意的 $t \geq m - 1$, 用 $\mathbf{Y}_t^m = (Y_t, \dots, Y_{t-m+1})$ 与 $\mathbf{X}_t^m = (X_t, \dots, X_{t-m+1})$ 分别表示 m 个连续的输出向量和 m 个连续的输入向量. 这样就得到

$$\mathbf{Y}_t^m = F_m(\mathbf{X}_t^m, \mathbf{S}_{t-m+1}), \quad t \geq m - 1 \quad (6)$$

因为状态函数 G 是一个平衡函数, 状态 \mathbf{S}_t 是一个平稳遍历的马尔可夫链, 所以即使初态 \mathbf{S}_0 是一个固定的向量, 随着 t 的增大 \mathbf{S}_t 会快速的收敛到均匀分布的随机变量, 因此我们不妨假设 \mathbf{S}_0 是均匀分布的. 这样要通过穷举的方法来获得所有输入的线性函数与所有输出的线性函数之间的相关系数的计算复杂度为 $O((mN + M)2^{2mN+M})$. 通常情况下我们取 $m = M + 1$, 当 MN 较大的时候, 这个计算方法是不可行的, 这个时候我们可以采用下面的分析方法.

4.3 线性序列电路逼近方法

Golić 在文献 [10] 中提出用线性序列电路逼近 (LSCA) 方法来求出在单个输出的带记忆密钥流生成器中, 具有相对较大相关系数的输出输入的线性函数对. 同样我们也可以把这个方法应用于多输出的带记忆密钥流生成器当中. 我们不妨对每一个输出函数的分量 f_1, f_2, \dots, f_N 进行线性序列电路逼近. 分析方法如下:

首先, 对于 $1 \leq i \leq N$, 找到 f_i 以及状态函数 G 每个分量的一个线性逼近, 这就等价于把这 $M + 1$ 个函数表示成一个线性函数和一个非平衡函数的和. 如果被分解的函数本来就是非平衡的, 那么就可以把线性函数取成零. 再根据引理 1, 如果被分解的函数和输入变量的某个子集统计独立, 那么它的每个线性逼近必须最少包括一个这个子集之外的变量. 因此最基本的要求就是对应的相关系数不为零. 当然可以运用 Walsh 变换技术来求取具有 $M + N$ 个输入的 $M + 1$ 个布尔函数的相关系数, 计算复杂度为 $O((M + 1)(M + N)2^{M+N})$. 其次, 给定线性逼近之后, 把状态函数和输出 $y_{i,t}$ 表示成矩阵的形式:

$$\mathbf{S}_{t+1} = \mathbf{A}\mathbf{S}_t + \mathbf{B}\mathbf{X}_t + \mathbf{\Delta}(\mathbf{X}_t + \mathbf{S}_t), \quad t \geq 0 \quad (7)$$

$$y_{i,t} = \mathbf{C}_i\mathbf{S}_t + \mathbf{D}_i\mathbf{X}_t + \varepsilon_i(\mathbf{X}_t, \mathbf{S}_t), \quad t \geq 0 \quad (8)$$

其中向量被看成是一列的矩阵, A, B, C_i 和 D_i 都是二元矩阵, ε_i 和 $\Delta = (\delta_1, \dots, \delta_M)$ 的分量都是非平衡的布尔函数, 称之为噪声函数. 现在主要的思想就是把 $\{\varepsilon_i(X_t, S_t)\}_{t=0}^{\infty}$ 和 $\{\delta_j(X_t, S_t)\}_{t=0}^{\infty}$, $q \leq j \leq M$, 当作输入序列, 这样式 (7) 和式 (8) 就定义了一个线性序列电路, 称之为有记忆组合器的线性序列电路逼近. 下面用文献 [10] 中的生成函数 (D -变换) 技术线性来分析这个线性序列电路. 设 $S, X, \Delta, \varepsilon_t$ 和 y_i 分别是序列 $\{S_t\}, \{X_t\}, \{\Delta(X_t, S_t)\}, \{\varepsilon_i(X_t, S_t)\}$ 和 $\{y_{i,t}\}$ 以 z 为变量的生成函数. 那么可以得到

$$S = zAS + zBX + z\Delta + S_0 \quad (9)$$

$$y_i = C_iS + D_iX + \varepsilon \quad (10)$$

式 (9) 和式 (10) 的解为

$$y_i = \left(D_i - \frac{C_i \text{adj}(zA - I)B}{\det(zA - I)} \right) X - \frac{C_i \text{adj}(zA - I)}{\det(zA - I)} (z\Delta + S_0) + \varepsilon_i \quad (11)$$

其中 I 是单位矩阵, $\det(zA - I) = \varphi(z)$, $\varphi(0) = 1$, 是状态变换矩阵 A 的特征多项式的逆, $\text{rank}(A) \leq M$, 矩阵 $\text{adj}(zA - I)$ 的元素是次数最大为 $M - 1$ 的多项式. 得到式 (11) 的计算复杂度为 $O(M^3(N + 1))$. 式 (11) 可以变成以下的形式:

$$y_i = \frac{1}{\varphi(z)} \sum_{p=1}^N g_{i,p}(z) x_p + \frac{1}{\varphi(z)} \sum_{j=1}^M h_{i,j}(z) (z\delta_j + s_{j0}) + \varepsilon_i \quad (12)$$

其中 x_p 和 δ_j 分别表示 $\{x_{p,t}\}$ 和 $\{\delta_j(X_t, S_t)\}$, 多项式 $g_{i,p}(z)$ 和 $h_{i,j}(z)$ 的次数分别不大于 M 和 $M - 1$, $1 \leq i \leq N$, $1 \leq j \leq M$. 令 $\varphi(z) = \sum_{k=0}^M \varphi_k z^k$, $g_{i,p}(z) = \sum_{k=0}^M g_{i,p,k} z^k$ 以及 $h_{i,j}(z) = \sum_{k=0}^{M-1} h_{i,j,k} z^k$, 把式 (12) 用时域的形式表达

$$\sum_{k=0}^M \varphi_k y_{i,t-k} = \sum_{i=1}^N \sum_{k=0}^M g_{i,p,k} x_{i,t-k} + e_i(X_t^{M+1}, S_{t-M}), \quad t \geq M \quad (13)$$

$$e_i(X_t^{M+1}, S_{t-M}) = \sum_{j=1}^M \sum_{k=0}^{M-1} h_{i,p,k} \delta_j(X_{t-1-k}, S_{t-1-k}) + \sum_{k=0}^M \varphi_k \varepsilon_i(X_{t-k}, S_{t-k}), \quad t \geq M \quad (14)$$

其中对任意 $0 \leq k \leq M - 1$, 假设状态向量 S_{t-k} 是 $(X_{t-k-1}^{M-k}, S_{t-M})$ 的函数. 在式 (13) 中的输出和输入函数统计相关当且仅当噪声函数 e_i 是非平衡的. 如果状态函数是非平衡的, 则相关系数与时间无关, 否则相关系数是依赖于时间的, 因为对于任意 $t \geq 0$, S_t 不再是平衡的函数. 当状态函数是平衡时, 式 (14) 中的噪声函数 e_i 定义成各个非平衡噪声函数的和. 因为各个噪声函数不一定是独立的, 所以在原则上 e_i 和常量零之间的相关系数不可能是零或者接近零. 这样我们就找出了具有相对较大相关系数的输出输入的线性函数对.

上述的分析方法的计算复杂度为 $O(NM^3(N + 1)(M + 1)(M + N)2^{M+N})$. 当然, 为了找出具有更大相关系数的输出输入的线性函数对, 我们也可以对 f_1, f_2, \dots, f_N 的任意线性组合 $\oplus_{i=1}^N u_i f_i$ 进行线性序列电路逼近, 这样计算复杂度就会增大 $(1/N)2^N$ 倍, 如果 N 不是很大的, 我们只需要花费少量的计算代价.

5 结论

本文给出多输逻辑函数相关免疫性的两个定义的等价性证明, 并对文献 [7] 提出的一种密钥流生成器给出了几种相关性分析的方法. 我们可以肯定地说这种多输出密钥流生成器比单个输出的密钥流生成器的安全性要低, 因为输出多了, 关于线性反馈移位寄存器的初态的熵漏也大了.

参 考 文 献

- [1] Siegenthaler T. Decrypting a class of stream ciphers using ciphertext only. *IEEE Trans. on Computers*, 1985, C-36(1): 81-85.
- [2] Siegenthaler T, Correlation immunity of nonlinear combining functions for cryptographic applications. *IEEE Trans. on Information Theory*, 1984, IT-30(9): 776-780.
- [3] Rueppel R A. Correlation immunity and the summation generator. *Advances in Cryptology-Crypto'86*. Berlin: Springer-Verlag, 1986: 260-272.
- [4] 丁存生, 肖国镇. 流密码学及其应用. 北京: 国防工业出版社, 1994: 169-173.
- [5] Gopalakrishan K, Stinson D R. Three characterizations of non-binary correlation-immune and resilient functions. *Designs, Codes and Cryptography*, 1995, 5(3): 241-251.
- [6] 陈鲁生. 多输出布尔函数的密码学性质. [博士论文]. 天津: 南开大学, 2000.
- [7] 徐汉良, 吕述望. 多输出相关免疫逻辑函数的等价刻画及其应用. *电子与信息学报*, 2002, 24(9): 1185-1189.
- [8] Meier W, Staffelbach O. Nonlinear criteria for cryptographic functions. *Advance in Cryptology-Eurocrypt'89*. Berlin: Springer-Verlag, 1990: 549-562.
- [9] Rueppel R A. Correlation immunity and the summation generator. *Advances in Cryptology-Crypto'86*. Berlin: Springer-Verlag, 1986: 260-272.
- [10] J. Dj. Golić. Correlation properties of a general binary combiner with memory. *Journal of Cryptology*, 1996, 9(2): 111-126.
- [11] 胡玉濮, 肖国镇, 张玉清. 对称密码学. 北京: 机械工业出版社, 2002: 32-33.

马卫局: 男, 1978 年生, 博士生, 主要研究方向是密码学和信息安全.

冯登国: 男, 1965 年生, 研究员, 博士生导师, 主要研究领域是分组密码算法的设计与分析、非线性函数的构造与分析、安全认证协议的设计与分析、计算机通信网络安全等方面.

巫治平: 男, 1978 年生, 博士生, 主要研究方向是密码学和信息安全.

张 斌: 男, 1976 年生, 博士生, 主要研究方向是密码学和信息安全.