文章编号:1001 - 2486(2006)04 - 0078 - 05

非线性组合序列的差分分析*

李 超[1,2],王文玲[1],胡朋松[1]
(1. 国防科技大学 理学院,湖南 长沙 410073;
2. 东南大学 移动通信国家重点实验室,江苏 南京 210018)

摘 要:针对非线性组合流密码的设计特点,结合明文编码与统计特性,分析了密文序列的差分性质,提出了攻击非线性组合序列的差分攻击算法。利用该算法,最后给出了攻击实例。

关键词:组合生成器;差分分析;差分位置集;差分有效性
中图分类号:TN918.1 文献标识码:A

Differential Attack on Nonlinear Combined Sequences

LI Chao[1,2], WANG Wen-ling[1], HU Peng-song[1]
(1. College of Science, National Univ. of Defense Technology, Changsha 410073, China;
2. State Key Laboratory of Mobile, Southeast University, Nanjing 210018, China)

Abstract: By combining with the properties of coding and statistic of the plaintext, we discussed the differential properties of the ciphertext sequences which are generated by a nonlinear combined generator. A difference attack algorithm which attacks the nonlinear combined sequences was proposed. In the end of the paper, an attack example of the difference attack algorithm was supplied.

Key words: combined generator; differential cryptanalysis; differential positional set; differential validity

非线性组合生成器在序列密码设计与分析中有着广泛的应用,它由 n 个线性移位寄存器($LFSR$)和一个非线性组合函数组成。目前针对这种系统较为有效的攻击方法主要有相关攻击和线性攻击,其基本思想都是利用组合生成器中组合函数 f 的输出和输入分量之间存在一定的相关性来进行分析。文献[1]针对该系统提出了“分别征服”的相关攻击方法,文献[2]提出了一种利用低密度校验码进行迭代概率译码的快速相关攻击方法,文献[3 - 4]则是对相关攻击的一系列改进。

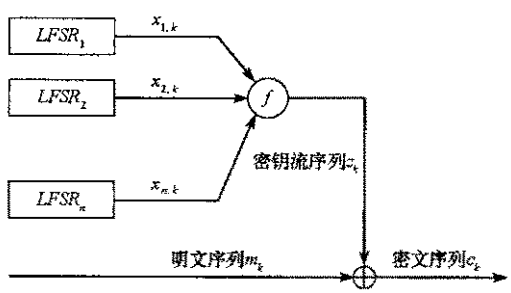

图 1 非线性组合序列密码
Fig.1 The nonlinear combined sequence cipher

差分密码分析自上世纪 90 年代初由 Eli. Biham 和 Adi. Shamir[5] 提出来之后,在分组密码的研究中已经得到了相当广泛的应用[6]。然而长期以来却很少有人把这种方法应用到序列密码的分析中去。本文所探讨的差分分析是一种唯密文攻击,基本思想是结合明文编码及其统计特性,针对非线性组合序列密码的设计特点,对密文序列进行差分分析,提出了相应的差分攻击算法,并给出了攻击实例。

1 差分的基本概念

定义 1 设集合 $A=\{a_1,\cdots,a_r\}$,$a_i\in F_2(1\leqslant i\leqslant r)$,集合 $S=\{(i_1,j_1),\cdots,(i_t,j_t)\}$ 中的元素是正整数集中的二维向量。如果对任意的 $1\leqslant k\leqslant t$,都有 $1\leqslant i_k,j_k\leqslant r$,则称集合 $\{a_{ik}\oplus a_{jk}\mid(i_k,j_k)\in S\}$ 为集合 A 关于 S 的差分集,记为 $\Delta_S A$。这时称 S 为 A 的差分位置集。集合 A 的差分位置集的全体记为 $\mathscr{U}_A$。

* 收稿日期:2005 - 12 - 28
基金项目:国家自然科学基金资助项目(60573028);东南大学移动通信国家重点实验室开放基金资助项目(A200503)
作者简介:李超(1966—),男,教授,博士。

定义 2　设集合 $A=\{a_1,\cdots,a_r\}$,$a_i\in F_2(1\leqslant i\leqslant r)$,$S$ 是 A 的一个差分位置集,$\Delta_S A$ 是 A 关于 S 的差分集,记 $P_S A=P\{x=0\mid x\in\Delta_S A\}=1/2+\delta$。若 $|\delta|>0$,则称集合 A 关于差分位置集 S 是 δ – 差分有效的,否则称为差分无效。类似地,也可以通过 $P\{x=1\mid x\in\Delta_S A\}$ 来定义集合 A 关于差分位置集 S 的差分有效性。为讨论方便,本文均采用前一种形式。

由于以下的讨论涉及到明文及编码的性质,这里对明文和编码的特点作一个简要的介绍。众所周知,在对明文信息进行加密前,人们总是需要事先对它进行编码,形成代表明文信息的二进制流数据,即所谓的明文序列。可见明文序列总是代表了一定环境中指定的信息,从而使得它具有相应的统计特性,例如文字编码序列具有明显的行文特征,语音信号和图像编码也有突出的特点。因此,差分分析之前总是假定在某一特定的编码下存在这样一条明文序列 $M=\{m_1,\cdots,m_r\}$,M 具有性质 T:明文序列 M 是该编码方式下所能表达的所有有效信息的总和。由于明文序列总是处于特定的编码方式之下,为便于讨论,将不再强调明文的编码方式。

定义 3　设明文序列 $M=\{m_1,\cdots,m_r\}$ 具有性质 T,$\mathscr{U}_M$ 是 M 的差分位置集的全体。对自然数 i,如果差分位置集 $S^*\in\mathscr{U}_M$ 满足 $|S^*|=i$,且 $|P_{S^*}M-1/2|=\max\{|P_S M-1/2|\mid S\in\mathscr{U}_M,|S|=i\}$,则称明文在统计量 i 下关于 S^* 是差分最优的。

2　差分分析

首先,建立分析的条件和模型。非线性组合序列密码如图 1 所示,设移存器的特征多项式已知,且均为本原多项式,级数分别为 $r_i(1\leqslant i\leqslant n)$;组合函数 $f(x_1,\cdots,x_n)$ 为非相关免疫平衡函数;密文序列为 $C=\{c_1,c_2,\cdots,c_N\}$。记前 N 拍移存器的输出符号分别为 $\{x_1^i,\cdots,x_N^i\}(1\leqslant i\leqslant n)$,对应的初态为 X^i,则 $X^i=(x_1^i,\cdots,x_{r_i}^i)\in F_2^{r_i}$。记 $f(x_1,\cdots,x_n)$ 的输出即密钥流序列为 $Z=\{z_1,z_2,\cdots,z_N\}$。

在上述建立的模型及条件下,由于移存器的特征多项式为本原多项式,故而在非零初态下由这些移存器产生的输出序列均是极大周期的,所以可以认为组合函数 f 的输入端是一些相互独立且服从同一分布的二元随机变量,而且对任意的 $1\leqslant i\leqslant n$ 和任意的 $1\leqslant k\leqslant N$ 都有 $P(x_k^i=0)=P(x_k^i=1)=1/2$。一般而言,作为密钥流序列,它首先应该具有良好的伪随机性质,所以组合函数 f 的输出序列也可视为一些相互独立且服从同一分布的二元随机变量,且对所有的 $k(1\leqslant k\leqslant N)$ 都近似地满足 $P(z_k=0)=P(z_k=1)=1/2$。

定义 4　若序列 $\{x_1^i,\cdots,x_N^i\}(1\leqslant i\leqslant n)$ 是由组合生成器中的第 i 个移存器在初态 $X^i\in F_2^{r_i}$ 下产生的,定义二维向量集合 $\lambda_{X^i}(1\leqslant i\leqslant n)$ 如下:

$$\lambda_{X^i}=\{(k,l)\mid x_k^i\oplus x_l^i=0,1\leqslant k,l\leqslant N\}\tag{1}$$

引理 1　如果密钥流序列 $Z=\{z_1,\cdots,z_N\}$ 是由上述模型在初态 $X^i\in F_2^{r_i}(1\leqslant i\leqslant n)$ 下产生的,λ_{X^i} 是由式(1)定义的集合。那么一定存在一个 $i(1\leqslant i\leqslant n)$ 及 $\varepsilon(0<\varepsilon\leqslant 1/2)$,使得下式成立:

$$P\{z_k\oplus z_l=0\mid(k,l)\in\lambda_{X^i}\}=1/2+\varepsilon$$

证明　因为 $x_k^i(1\leqslant i\leqslant n,1\leqslant k\leqslant N)$ 可视为一些相互独立且服从同一分布的二元随机变量,$P(x_k^i=0)=P(x_k^i=1)=1/2$,而 $f(x_1,\cdots,x_n)$ 又是非相关免疫的平衡函数,故存在一个 $i(1\leqslant i\leqslant n)$ 及 $\xi(0<|\xi|\leqslant 1/2)$ 使得 $P\{f(x_k^1,\cdots,x_k^n)\oplus x_k^i=0\}=1/2+\xi$,所以

$$\begin{aligned}
&P\{z_k\oplus z_l=0\mid(k,l)\in\lambda_{X^i}\}\\
&=P\{f(x_k^1,\cdots,x_k^n)\oplus f(x_l^1,\cdots,x_l^n)=0\mid(k,l)\in\lambda_{X^i}\}\\
&=P\{f(x_k^1,\cdots,x_k^n)\oplus x_k^i\oplus f(x_l^1,\cdots,x_l^n)\oplus x_l^i=0\mid(k,l)\in\lambda_{X^i}\}\\
&=P\{f(x_k^1,\cdots,x_k^n)\oplus x_k^i=0\}\cdot P\{f(x_l^1,\cdots,x_l^n)\oplus x_l^i=0\}\\
&\quad+P\{f(x_k^1,\cdots,x_k^n)\oplus x_k^i=1\}\cdot P\{f(x_l^1,\cdots,x_l^n)\oplus x_l^i=1\}
\end{aligned}$$

$$= (1/2 + \xi)^2 + (1/2 - \xi)^2$$
$$= 1/2 + 2\xi^2$$

因为 $0 < |\xi| \leqslant 1/2$,所以 $0 < 2\xi^2 \leqslant 1/2$。令 $\varepsilon = 2\xi^2$ 即引理得证。 □

引理2 设密钥流序列 $Z = \{z_1, \cdots, z_N\}$ 是由上述模型在初态 $X^i \in F_2^{r_i} (1 \leqslant i \leqslant n)$ 下产生的,则对任意的 i 及 $\not{X}^i \in F_2^{r_i}$,如果 $\not{X}^i \neq X^i$,那么 $P\{z_k \oplus z_l = 0 | (k,l) \in \lambda_{\not{X}^i}\} = 1/2$。其中 $\lambda_{\not{X}^i}$ 是由式(1)定义的集合。

证明 对任意的 i 及 $\not{X}^i \in F_2^{r_i}$,设 $\{\not{z}_1^i, \cdots, \not{z}_N^i\}$ 是由上述模型中的第 i 个移存器在初态 $\not{X}^i$ 下产生的序列,则

$$\begin{aligned}
&P\{z_k \oplus z_l = 0 | (k,l) \in \lambda_{\not{X}^i}\} \\
&= P\{f(x_k^1, \cdots, x_k^n) \oplus f(x_l^1, \cdots, x_l^n) = 0 | (k,l) \in \lambda_{\not{X}^i}\} \\
&= P\{f(x_k^1, \cdots, x_k^n) \oplus f(x_l^1, \cdots, x_l^n) \oplus \not{z}_k^i \oplus \not{z}_l^i = 0 | (k,l) \in \lambda_{\not{X}^i}\} \\
&= P\{f(x_k^1, \cdots, x_k^n) \oplus \not{z}_k^i = 0\} \cdot P\{f(x_l^1, \cdots, x_l^n) \oplus \not{z}_k^i = 0\} \\
&\quad + P\{f(x_k^1, \cdots, x_k^n) \oplus \not{z}_k^i = 1\} \cdot P\{f(x_l^1, \cdots, x_l^n) \oplus \not{z}_k^i = 1\} \\
&= (1+2)^2 + (1/2)^2 = 1/2
\end{aligned}$$ □

定理1 设密文序列为 $C = \{c_1, \cdots, c_N\}$,明文关于差分位置集 S 是 δ_S - 差分有效的,若加密明文的密钥流序列 $Z = \{z_1, \cdots, z_N\}$ 是由上述模型在初态 $X^i \in F_2^{r_i} (1 \leqslant i \leqslant n)$ 下产生的,那么一定存在一个 $i (1 \leqslant i \leqslant n)$ 及 $\varepsilon (0 < \varepsilon \leqslant 1/2)$,使得 $P\{c_k \oplus c_l = 0 | (k,l) \in \lambda_{X^i} \cap S\} = 1/2 + 2\delta_S\varepsilon$。

证明 明文序列记为 $M = \{m_1, \cdots, m_N\}$,则 $C = \{c_1, \cdots, c_N\} = \{z_1 \oplus m_1, \cdots, z_N \oplus m_N\}$。由于明文关于差分位置集 S 是 δ_S - 差分有效的,所以 $P_S M = P\{m_k \oplus m_l = 0 | (k,l) \in S\} = 1/2 + \delta_S$。再由引理1可知存在一个 i 及 $\varepsilon (0 < \varepsilon \leqslant 1/2)$,使得 $P\{z_k \oplus z_l = 0 | (k,l) \in \lambda_{X^i}\} = 1/2 + \varepsilon$。故

$$\begin{aligned}
&P\{c_k \oplus c_1 = 0 | (k,l) \in \lambda_{X^i} \cup S\} \\
&= P\{z_k \oplus m_k \oplus z_l \oplus m_l = 0 | (k,l) \in \lambda_{X^i} \cap S\} \\
&= P\{z_k \oplus z_l = 0\} \cdot P\{m_k \oplus m_l = 0\} + P\{z_k \oplus z_l = 1\} \cdot P\{m_k \oplus m_l = 1\} \\
&= (1/2 + \varepsilon)(1/2 + \delta_S) + (1/2 - \varepsilon)(1/2 - \delta_S) \\
&= 1/2 + 2\delta_S\varepsilon
\end{aligned}$$ □

定理2 设密文序列为 $C = \{c_1, \cdots, c_N\}$,若加密明文的密钥流序列 $Z = \{z_1, \cdots, z_N\}$ 是由上述模型在初态 $X^i \in F_2^{r_i} (1 \leqslant i \leqslant n)$ 下产生的,则对任意的 i 及 $\not{X}^i \in F_2^{r_i}$,如果 $\not{X}^i \neq X^i$,那么

$$P\{c_k \oplus c_l = 0 | (k,l) \in \lambda_{\not{X}^i} \cap S\} = 1/2$$

证明 设明文序列记为 $M = \{m_1, \cdots, m_N\}$,则 $C = \{c_1, \cdots, c_N\} = \{z_1 \oplus m_1, \cdots, z_N \oplus m_N\}$。由引理2可知对任意 $\not{X}^i \in F_2^{r_i}$,$\not{X}^i \neq X^i (1 \leqslant i \leqslant n)$,有 $P\{z_k \oplus z_l = 0 | (k,l) \in \lambda_{\not{X}^i}\} = 1/2$。故

$$\begin{aligned}
&P\{c_k \oplus c_l = 0 | (k,l) \in \lambda_{\not{X}^i} \cap S\} \\
&= P\{Z_k \oplus m_k \oplus z_l \oplus m_l = 0 | (k,l) \in \lambda_{\not{X}^i} \cap S\} \\
&= P\{z_k \oplus z_l = 0\} \cdot P\{m_k \oplus m_l = 0\} + P\{z_k \oplus z_l = 1\} \cdot P\{m_k \oplus m_l = 1\} \\
&= \frac{1}{2} P\{m_k \oplus m_l = 0\} + \frac{1}{2} P\{m_k \oplus m_l = 1\} = \frac{1}{2}
\end{aligned}$$ □

由定理1和定理2可得到以下推论:

推论1 设密文序列为 $C = \{c_1, \cdots, c_N\}$,明文关于差分位置集 S 是 δ_S - 差分有效的,若加密明文的密钥流序列 $Z = \{z_1, \cdots, z_N\}$ 是由上述模型在初态 $X^i \in F_2^{r_i} (1 \leqslant i \leqslant n)$ 下产生的,令 $\Delta_N^i = \Delta_{\lambda_{X^i} \cap S} C$,那么一定存在一个 $i (1 \leqslant i \leqslant n)$ 及 $\varepsilon (0 < \varepsilon \leqslant 1/2)$,使得

$$\lim_{N \to \infty} \frac{\sum\limits_{x \in \Delta_N^i} x}{|\Delta_N^i|} = \frac{1}{2} - 2\delta_S\varepsilon$$

推论2 设密文序列为 $C = \{c_1, \cdots, c_N\}$,若加密明文的密钥流序列 $Z = \{z_1, \cdots, z_N\}$ 由上述模型在初

态 $X^i \in F_2^{r_i}(1 \leqslant i \leqslant n)$ 下产生,则对 $\forall\ i$ 及 $\tilde{X}^i \in F_2^{r_i}$,令 $\tilde{\Delta}_N^i = \Delta_{\lambda_{\tilde{X}^i} \cap S} C$,如果 $\tilde{X}^i \neq X^i$,则

$$\lim_{N \to \infty} \frac{\sum\limits_{x \in \tilde{\Delta}_N^i} x}{|\tilde{\Delta}_N^i|} = \frac{1}{2}$$

3　差分攻击算法

在如图 1 所示的序列密码中,设明文在差分位置集 S 下是满足 δ_S - 差分有效的,又根据函数 $f(x_1, \cdots, x_n)$ 是非相关免疫函数的前提,从而确定出与函数 f 的输出符号具有一定相关性的输入端 x_i 并计算 $\xi = P\{f(x_1, \cdots, x_n) \oplus x_i = 0\} - 1/2$ 的大小,进而确定 $2\delta_S\varepsilon$。一般认为被加密的明文序列同样具有类似 δ_S 差分有效的性质,那么由定理 1、2 及其推论,通过已知的密文序列 $C = \{c_1, \cdots, c_N\}$ 来还原移存器 $LFSR_i$ 的初态,可以按照下述步骤进行:

Step 1　任选一个初态 $X^i \in F_2^{r_i}, X^i \neq 0$,通过移存器 $LFSR_i$ 生成一条周期为 $2^{r_i} - 1$ 的序列 $\{x_1^i, \cdots, x_r^i\}$,其中 $r = 2^{r_i} - 1$。

Step 2　初始化位置 $WZ = 0$

Step 3　以 WZ 为起点,截取序列 $\{x_1^i, \cdots, x_r^i\}$ 中 N 个符号 $\{y_1, \cdots, y_N\} = \{x_{WZ+1}^i, \cdots, x_{WZ+N}^i\}$,以 $Y = (y_1, \cdots, y_{r_i})$ 根据明文的差分位置集 S 及定义 4 确定二维向量集 $\lambda_Y \cap S$。

Step 4　计算密文 $C = \{c_1, \cdots, c_N\}$ 关于新的差分位置集 $\lambda_Y \cap S$ 的差分集 $\Delta_{\lambda_Y \cap S} C$,并统计该差分集中 0 的个数所占的比例,记为 P_{WZ}。若 $P_{WZ} \to 1/2 + 2\delta_S\varepsilon$,则接受 Step3 中截取的 $(y_1, \cdots, y_{r_i})$ 为移存器 $LFSR_i$ 的初态,即子密钥。

Step 5　重新赋值 $WZ = WZ + 1$ 后,若 $WZ > 2^{r_i} - 1$,则进入 Step6 结束算法,否则跳至 Step3。

Step 6　结束算法。

上述差分算法实质上也是一种分别征服、各个击破的攻击方法,通过有条件的判断来确定某个移存器 $LFSR_i$ 的子密钥。算法得以实施是依赖于第 2 小节中建立的模型中所设定的各种条件,事实上这些条件的假设也是合理的[7]。

4　实 例

设上述分析模型中移存器的个数为 8,反馈多项式分别为:$g_1(x) = x^{17} + x^3 + 1$,$g_2(x) = x^{20} + x^3 + 1$,$g_3(x) = x^{23} + x^{20} + 1$,$g_4(x) = x^{29} + x^{26} + 1$,$g_5(x) = x^{25} + x^{20} + x^{12} + x^8 + 1$,$g_6(x) = x^{31} + x^{24} + x^{16} + x^{12} + 1$,$g_7(x) = x^{33} + x^{28} + x^{24} + x^4 + 1$,$g_8(x) = x^{39} + x^{36} + x^{28} + x^4 + 1$。组合函数 $f = \{$E25047C342F8061638519519862819536 5DD937F5BF4B702B17EFED92E2D83B7$\}$

为方便分析,抽取本文引言中的文字部分作为明文序列,并采用 GB 码对其进行编码,这时明文序列 $M = \{$B7C7CFDFD0D4D7E9BACFC9FAB3C9C6F7D4DAD0F2C1D0C3DCC$\cdots\}$。

下面随机选取一组初态对以上明文序列进行加密,随机选取的移存器初态如下:

187F7、856BD、F2ACF、649F525、19F037E、31F62B97、1974D1905、64F89DDB22

上述初态中每一个字符均表示 4 位二进制数,如果初态的长度超过寄存器的长度,则去掉前面的若干位 如果初态的长度小于寄存器的长度,则前面的若干位补 0。

统计发现,对正整数 k,明文序列 M 关于位置集 $S_k = \{8(i, i+j) \mid 1 \leqslant j \leqslant k, 8(i+j) \leqslant |M|\}$ 是 $1/2$ - 差分有效的。对于组合函数 f 也有以下结论:(1) f 是平衡函数;(2) 若 f 的输入端为相互独立且均匀分布的二元随机变量,则 $P\{f \oplus x_1 = 0\} \approx 0.60156$。在以下实例中,取明文的差分位置集 $S = S_3$,进而 $2\delta_S\varepsilon \approx 0.02063$。

根据第 4 节所描述的算法,首先任取 $X^1 \in F_2^{17}, X^1 \neq 0$ 生成长度为 $2^{17} - 1$ 的序列 $\{x_1^1, \cdots, x_{2^{17}-1}^1\}$,然后对该序列的每一个位置计算 P_{WZ},并统计 P_{WZ} 在各个概率小区间出现的频次分布(实验数据中假设可

获取密文的数据量为连续的10,000比特),P_{WZ}在各个概率小区间出现的频次分布见下表。

分布区间	P_{WZ}出现频次
[0.4985, 0.4995)	69
[0.4995, 0.5005)	7924
[0.5005, 0.5015)	50950
[0.5015, 0.5025)	44295
[0.5025, 0.5035)	18598
[0.5035, 0.5045)	6338
[0.5045, 0.5055)	1982
[0.5055, 0.5065)	634
[0.5065, 0.5075)	189
[0.5075, 0.5085)	68
[0.5085, 0.5095)	15
[0.5095, 0.5105)	6
[0.5105, 0.5115)	1
[0.5125, 0.5135)	2
[0.5155, 0.5165)	1

其中对应 $LFSR_1$ 的正确初态的位置所计算的 $P_{WZ} \approx 0.51647 \rightarrow 0.5 + 0.02063$,而对应其他位置的 P_{WZ}则全部集中在区间[0.4895,0.5135]内部。该实验数据表明,本文所提出的差分攻击算法对还原上述模型中移存器序列的初态(子密钥)是行之有效的。

5 总结

本文讨论的是一种唯密文攻击方法。该方法结合明文编码和统计特性,通过考察密文序列的差分性质对非线性组合序列密码进行了研究和探讨,提出了一种差分攻击算法,并证明了如果 $2\delta_S\varepsilon \neq 0$,只要获得的密文序列长度充分大,该方法总能成功。事实上,对于本文所提出的算法仍有一些问题有待进一步解决,例如如何选取明文的差分位置集并使其达到在某个统计量下差分最优;一般要获得的密文长度达到多大的时候就可以使得上述差分分析算法奏效;如何利用移存器的推移关系快速确定密文 C 的差分位置集 $\lambda_Y \cap S$ 和计算 P_{WZ}的问题。

参考文献:

[1] Siegenthaler T. Correlation Immunity of Nonlinear Combining Function for Cryptographic Application[J]. IEEE Trans on Information Theory, 1984, 30(6): 776 – 780.

[2] Meier W, Staffelbach O. Fast Correlation Attacks on Certain Stream Ciphers[A]. Advances in Cryptology- EUROCRYPT' 88[C]. Berlin: Springer-Verlag, 1988. 301 – 314.

[3] Golic JD, Salmasizadeh M, Dawson E. Fast Correlation Attacks on the Summation Generator[J]. Journal of Cryptology. 2000.13: 245 – 262.

[4] 张卫明,李世取. 组合生成器的多线性相关攻击[J]. 电子学报, 2005,33(3): 427 ~ 432.

[5] Biham E, Shamir A. Differential Cryptanalysis of the Data Encryption Standard[M]. Berlin: Springer-Verlag, 1993.

[6] 冯登国,吴文玲. 分组密码的设计与分析[M]. 北京:清华大学出版社,2000.

[7] 冯登国. 密码分析学[M]. 北京:清华大学出版社,2000.

[8] 李超,黄建忠,项攀攀. 差分分析在序列密码攻击中的应用[J]. 应用科学学报,2004,22(2): 127 ~ 131.