基于相关攻击的 A5/1 算法识别

陈 伟[①] 胡 云[①] 杨义先[①] 钮心忻[②]
[①](北京邮电大学信息安全中心 北京 100876)
[②](北京邮电大学数字内容研究中心 北京 100876)

摘 要：利用相关攻击获得畸变的 A5/1 序列，再用统计工具对其进行处理可以得出 A5/1 统计特征，从而找到正确区分 A5/1 算法输出和伪随机序列的方法。实验验证该方法能有效地将 A5/1 算法输出和伪随机序列区分开来。

关键字：密码分析，算法识别，A5/1 算法，相关攻击，正态分布
中图分类号：TN918 **文献标识码**：A **文章编号**：1009-5896(2006)05-0827-05

Application of Correlation Attack in Algorithm Identify

Chen Wei[①] Hu Yun[①] Yang Yi-xian[①] Niu Xin-xin[②]
[①](*Information Security Center, Beijing University of Posts & Telecommunications, Beijing* 100876, *China*)
[②](*Research Center of Digital Contents, Beijing University of Posts & Telecommunications, Beijing* 100876, *China*)

Abstract A correlation attack on A5/1 algorithm can be educed by the linear filling weakness in initiate process of A5/1 algorithm. An aberrant A5/1 sequence can be obtained from it, which have treated by statistic tools to get A5/1 statistic trait, so the distinction of A5/1 output from real random sequence can be found. Test results show that this method can work effectively.

Key words Cryptanalysis, Algorithm identify, A5/1 algorithm, Correlation attack, Normal distribution

1 引言

广义而言，密码分析是指获得明文、密钥或算法的任何信息的技术手段。Lars Knudsen把破译算法分为四类：全部破译、全盘推导、实例或局部推导、信息推导[1]。我们将它按照目的分为 3 类：以获得零星信息(如算法体制、明文格式、密钥的零星位等)为目的的信息推导；通过找到等效或近似算法恢复部分或全部明文的明文推导；密钥恢复。一般认为信息推导属于最弱的一种。因此，中外专家们在密码分析学领域的工作大多集中在第 3 类分析。常用攻击法也以恢复密钥为主要目的。

信息推导之所以弱，在于对方所用的密码算法类型和资料等信息一般容易通过非技术手段获得。但在实际工作中，当通过非技术手段无法获得这些信息时，我们就必须从技术上对两个比特序列进行区分。

算法识别的目的就是根据某个算法产生的比特序列和真随机序列之间的细微差别来识别出待测序列是否由该算法产生。技术上的困难表现在两方面：其一，现代密码学理论的发展从设计上保证了算法输出的伪随机性越来越好，而商用密码算法更要通过各种统计和代数测试才能实用化，算法输出序列和真随机序列之间的差别越来越小。其二，一些常用攻击法利用了与密钥紧密相关的算法结构特征，而算法识别要利用与密钥无关或相关性很少的结构特征，因此现有大部分攻击法很难用于识别。

虽然如此，算法识别仍然需要借鉴很多密码分析法，如相关攻击等。作为算法识别研究的一个范例，本文旨在通过运用这些方法，结合统计学工具，给出 A5/1 算法输出特征。

2 A5/1 算法简介

A5/1 算法(以下简称A5/1)是欧洲数字蜂窝移动电话系统GSM中采用的流密码加密算法，用于电话手机到基站线路上的话音和数据加密。它的输入为 86 bit密钥，分为 64 bit会话密钥K_c和 22 bit帧序列号F_n，生成的 114 bit密钥流和 114 bit的明文“异或”产生 114 bit密文，“或”和 114 bit的密文“异或”产生 114 bit明文。它工作在流密码的计数器模式下[2]。

A5/1 内部由 3 个线性反馈移位寄存器(LFSR)R0, R1 和 R2 组成，级数分别为 19, 22, 23，抽头数分别为 4, 2, 4 个，将 R0, R1 和 R2 的“异或”作为位输出。其设计思想是采用“3 个 LFSR 互相控制时钟”的结构。A5/1 有 3 个钟控输入，分别为每个 LFSR 的中间位，和 3 个钟控输出，分别控制每个 LFSR 的停／走。钟控机制采用择多逻辑，在每一轮中时钟驱动 2 个 LFSR 移位。若 2 或 3 个控制比特值为“1”，则控制比特为“1”的寄存器移位；若 2 或 3 个控制比特为“0”，则控制比特为“0”的寄存器移位。这种不规则的互钟控结构使得 A5/1 的性能非常优良，可通过所有已知统计检测。

A5/1 的加解密过程是相似的。每一帧密钥流按照如下方式产生：

2004-12-20收到，2005-11-28 改回
国家自然基金(60372094, 90204017)资助

(1)初始化　首先将 3 个移存器全部清 0。然后，在 64 个时钟周期内，将每个移存器都移位 64 次(不带钟控)，并在每次移位后将密钥 bitK_i(i=0~63)置入每个移存器的最低位。第 3 步，在 22 个时钟周期内，将每个移存器都移位 22 次(不带钟控)，并在每次移位后将帧序列号 bitFrame$_i$(j=0~21)置入每个移存器的最低位；

(2)产生输出密钥流　将 3 个移存器钟控移位 100 拍，丢弃输出。然后，互钟控输出 228 bit，作为双向通信中使用的密钥流。

A5/1 的初始化过程中对密钥和帧号采用了线性填充，本意是使 3 个移存器的内部状态和每一个密钥和帧序列号比特都有关，但是却降低了安全性。本文将要利用这一弱点，用相关攻击产生畸变的 A5/1 密钥流，从而暴露出其统计特性。

3　针对 A5/1 的相关攻击和畸变密钥流的产生

因为 A5/1 在 GSM 通信中的广泛应用，它也不可避免地成为中外密码专家的研究热点之一。文献[3]提出的相关攻击主要思路是：利用其线性初始化弱点，可以将密钥信息和帧号信息的影响分开。因为每个密钥流 bit 都是原始密钥信息和帧号信息的复杂函数，如果去除了帧号信息，就可以以高概率识别出原始密钥信息。

A5/1 有“每个时钟周期内，每个 LFSR 移位的平均概率都是 3/4”的特点，所以可以用如下步骤进行攻击：首先计算出 R0, R1 和 R2 输出和 A5/1 输出的相关概率，然后将 R0, R1 和 R2 输出和 A5/1 输出的一些相关比特位“异或”，去除密钥流中的帧号信息。当密钥流中只剩下原始密钥信息后，大量帧的某个或某些位的集合将会表现出一定的 0-1 不平衡性，从而可以进行密钥信息的概率推导。

记 $K=(k_0,k_1,\cdots,k_{63}),F_n=(f_0,f_1,\cdots,f_{21})$。在规则时钟控制下，每一个移存器的后续状态都是密钥K和帧序列号F_n的线性函数。以 $u_1^0,u_2^0,u_3^0,\cdots$ 表示R0 第 3 步初始化完成后在规则钟控下的输出序列；R1, R2 输出序列也可以类似表示。则有 $u_t^0=\sum_{i=0}^{63}c_{it}^0k_i+\sum_{j=0}^{21}d_{jt}^0f_j$，其中 c_{it}^0，$i=0,1,\cdots 63$，d_{jt}^0，$j=0,1,\cdots 21$ 均为未知 0，1 常数。令 $s_t^0=\sum_{i=0}^{63}c_{it}^0k_i$，称为 u_t^0 的密钥部分；$\hat{f}_t^0=\sum_{j=0}^{21}d_{jt}^0f_j$，称为 u_t^0 的帧序列号部分，则上式可写成 $u_t^0=s_t^0+\hat{f}_t^0$，$t\ge 0$，表示R0 在t时刻的输出是密钥K和帧序列号F_n的线性函数。我们注意到，$s_0^0,s_1^0,\cdots$ 都是未知数，但在一次会话的所有帧中保持不变，而 $f_0^0,f_1^0,\cdots$ 在确定帧序列号时下可以推导出来，属于已知数，但随每一次的帧序列号改变而改变。同样，对于R1，R2 也有类似等式成立。

不妨将A5/1 初始化后要丢弃的前 100 位也作为输出计算在内。设A5/1 输出帧的第cz位z_{cz}时，R0, R1 和R2 正好分别移位了cl_0、cl_1和cl_2拍，即有等式 $z_{cz}=u_{\mathrm{cl}_0}^0\oplus u_{\mathrm{cl}_1}^1\oplus u_{\mathrm{cl}_2}^2$ 成立。将上述式子代入整理，可以得到：

$$s_{\mathrm{cl}_0}^0\oplus s_{\mathrm{cl}_1}^1\oplus s_{\mathrm{cl}_2}^2=\hat{f}_{\mathrm{cl}_0}^0\oplus\hat{f}_{\mathrm{cl}_1}^1\oplus\hat{f}_{\mathrm{cl}_2}^2\oplus z_{cz}\tag{1}$$

等式左边为密钥信息的异或，右边表示A5/1 输出帧的第cz位和R0、R1 和R2 输出序列中的第cl_0, cl_1和cl_2位“异或”运算，过滤掉帧号信息影响的结果。

当A5/1 输出某一帧的第cz位正好由R0、R1 和R2 输出序列中的第cl_0, cl_1和cl_2位异或得到时，上式成立的概率为 1；当第cz位不是由R0, R1 和R2 输出序列中第cl_0, cl_1和cl_2位“异或”得到时，上式成立的概率为 1/2。设前一种情况发生的概率为 $p_{\mathrm{cl}_0,\mathrm{cl}_1,\mathrm{cl}_2,cz}$，则后一种情况发生的概率为 $1-p_{\mathrm{cl}_0,\mathrm{cl}_1,\mathrm{cl}_2,cz}$。因此，对于同一次会话的$n$个不同帧的第cz bit 构成的$n$长 0-1 序列，(1)式成立的概率为

$$\begin{aligned}&p(s_{\mathrm{cl}_0}^0\oplus s_{\mathrm{cl}_1}^1\oplus s_{\mathrm{cl}_2}^2=\hat{f}_{\mathrm{cl}_0}^0\oplus\hat{f}_{\mathrm{cl}_1}^1\oplus\hat{f}_{\mathrm{cl}_2}^2\oplus z_{cz})\\&=p_{\mathrm{cl}_0,\mathrm{cl}_1,\mathrm{cl}_2,cz}\cdot 1+(1-p_{\mathrm{cl}_0,\mathrm{cl}_1,\mathrm{cl}_2,cz})\cdot\frac{1}{2}=\frac{1+p_{\mathrm{cl}_0,\mathrm{cl}_1,\mathrm{cl}_2,cz}}{2}\end{aligned}\tag{2}$$

上式是A5/1 输出帧的第cz位和R0, R1 和R2 输出序列中的第cl_0, cl_1和cl_2位之间的相关概率。根据A5/1 结构中的择多逻辑，我们可以导出如下概率递推公式：

$$\left.\begin{aligned}&p(\mathrm{cl}_1,\mathrm{cl}_2,\mathrm{cl}_3,0)=1,\quad 若\mathrm{cl}_1=\mathrm{cl}_2=\mathrm{cl}_3=0\\&p(\mathrm{cl}_1,\mathrm{cl}_2,\mathrm{cl}_3,cz)=0,\quad 若\mathrm{cl}_1<0或\mathrm{cl}_2<0或\mathrm{cl}_3<0\\&p(\mathrm{cl}_1,\mathrm{cl}_2,\mathrm{cl}_3,cz)=0,\quad 若\mathrm{cl}_1>cz或\mathrm{cl}_2>cz或\mathrm{cl}_3>cz\\&p(\mathrm{cl}_1,\mathrm{cl}_2,\mathrm{cl}_3,cz)=0.25p(\mathrm{cl}_1-1,\mathrm{cl}_2-1,\mathrm{cl}_3-1,cz-1)\\&\quad+0.25p(\mathrm{cl}_1-1,\mathrm{cl}_2-1,\mathrm{cl}_3,cz-1)\\&\quad+0.25p(\mathrm{cl}_1-1,\mathrm{cl}_2,\mathrm{cl}_3-1,cz-1)\\&\quad+0.25p(\mathrm{cl}_1,\mathrm{cl}_2-1,\mathrm{cl}_3-1,cz-1)\end{aligned}\right\}\tag{3}$$

对于同一次会话产生的密钥流序列，取n个不同帧的第cz位构成一条n长 0-1 序列，它和R0, R1 和R2 输出序列中对应帧的第cl_0、cl_1和cl_2位构成的三条n长 0-1 序列“异或”得到 $\hat{f}_{\mathrm{cl}_0}^0\oplus\hat{f}_{\mathrm{cl}_1}^1\oplus\hat{f}_{\mathrm{cl}_2}^2\oplus z_{cz}$，即第cz位过滤掉部分帧号信息影响后产生的$n$长 0-1 序列，我们称之为“$n$长畸变A5/1 序列”。从式(2)可知，它一定是一个不平衡的 0-1 序列，具有不可忽略的 0-1 分布不平衡偏差率。

记 $p=(1\pm p_{\mathrm{cl}_0,\mathrm{cl}_1,\mathrm{cl}_2,cz})/2$。由式(2)知，4 元组($\mathrm{cl}_0$, cl_1, cl_2, cz)畸变序列的每一bit均服从(0，1)分布，取 0 或 1 的概率值为p。从统计学角度，n长 bit串可看作是服从(0，1)分布的随机变量x经过n重贝努里试验的产物。设X表示n长畸变序列中‘1’的个数，那么它服从概率二项分布 $b(n,p)$。当n取值很大的时候，由Liapunov定理[4]，随机变量 $\left[\left(\sum_{i=1}^{n}X_i-\frac{n}{2}\right)-\left(np-\frac{n}{2}\right)\right]\Big/\sqrt{np(1-p)}$ 也近似服从标准正态分布 $N(0,1)$，即随机变量 $\left(\sum_{i=1}^{n}X_i-\frac{n}{2}\right)\Big/\sqrt{np(1-p)}$ 近似服从正态分布 $N(\pm\sqrt{n}\,p_{\mathrm{cl}_0,\mathrm{cl}_1,\mathrm{cl}_2,cz},1)$。不同 A5/1 输入密钥产生不同的 n 长畸变序列，它们的 0-1 不平衡值可以构成一条统计曲线，将之做归一化处理以后，是一条对称轴向 y 轴左侧或

右侧偏移的正态曲线。

根据文献[5]中的伪随机性要求，随机序列 0-1 偏差值一般在 48.27~51.73%，即 $p_{cl_0,cl_1,cl_2,cz}/\sqrt{n} \le 1.73\%$ 。而当 $cz > 100$ 时z_{cz}可见， $p_{cl_0,cl_1,cl_2,cz}$ 才有实用价值。但式(2)计算发现此时的 $p_{cl_0,cl_1,cl_2,cz}$ 值很小，且随cz的增大而递减， $p_{cl_0,cl_1,cl_2,cz}$ 的最大值如表 1 所示：

表 1　$cz > 100$ 时的不平衡概率最大值

Tab.1 Max value of unbalance probablebility when $cz > 100$

cl_0	cl_1	cl_2	cz	$p_{cl_0,cl_1,cl_2,cz}$
76	76	75	101	0.000974338
76	75	76	101	0.000974338
75	76	76	101	0.000974338
76	76	76	101	0.000974338

例如，当 n 取值为 90000 时， $\sqrt{n} \cdot p_{cl_0,cl_1,cl_2,cz}$ 的值仅仅为 0.2923， $p_{cl_0,cl_1,cl_2,cz}/\sqrt{n} \approx 1\%$ 。对于一条长 1282500 byte 的 A5/1 密钥流来说，其对应于 4 元组(76，76，75，101)的畸变序列的 0-1 偏差值平均仅为 ±88。因此，直接利用单个 4 元组还不足以将 A5/1 密钥流和随机比特流区分开，需要对曲线进行三次变形处理。

4　算法识别特征的产生

定理 1[4]　设正态随机变量 $X_1, X_2, \cdots, X_m$ 相互独立，数学期望和方差分别为 $E(X_k)=\mu_k, D(X_k)=\sigma_k^2 \ne 0, k=1,2,\cdots$。当 m 很大时，随机变量 $z=a_1x_1+a_2x_2+\cdots+a_mx_m$ 近似服从于正态分布 $N\left(\sum_{k=1}^{m}a_k\mu_k, \sum_{k=1}^{m}a_k^2\sigma_k^2\right)$。

因为 $s_{cl_0}^i \oplus s_{cl_1}^i \oplus s_{cl_2}^i$ 的值仅仅与 3 元组(cl_0, cl_1, cl_2)相关，对固定的 3 元组(cl_0，cl_1, cl_2)， $s_{cl_1}^i \oplus s_{cl_2}^i \oplus s_{cl_3}^i$ 的值是固定的 0 或 1，所以与一定取值区间I 上的cz对应的一组畸变序列簇的正态曲线位于y轴同侧，只是数学期望 $\sqrt{n}\cdot p_{cl_0,cl_1,cl_2,cz}$ (曲线最高点)离y轴距离不同而已。第一次处理用用线性函数 $z=\frac{x_1+x_2+\cdots+x_m}{\sqrt{m}}, (m=|I|)$ 将它们复合。根据定理 1，z 服从正态分布 $N\left(\pm\sqrt{\frac{n}{m}}\cdot\sum_{i=1}^{m}p_{cl_0,cl_1,cl_2,cz_m}, 1\right)$，其最高点离 y 轴略远。例如，对于$cl_0$=$cl_1$=$cl_2$=78，$I$={102，103，104，105，106}，$m$=5，$n$=90000，如表 2 所示：

表 2　区间 I 内的 4 元组概率表

Tab.2 Probablebility Table of 4 item group in range I

cl_0	cl_1	cl_2	cz	$p_{cl_0,cl_1,cl_2,cz}$	$\sqrt{n}p_{cl_0,cl_1,cl_2,cz}$
78	78	78	102	0.000590241	0.1770723
78	78	78	103	0.000846263	0.2538789
78	78	78	104	0.000946413	0.2839239
78	78	78	105	0.000820412	0.2461236
78	78	78	106	0.000546692	0.1640076

当 n 取值为 90000 时， $\sqrt{\frac{n}{m}}\cdot\sum_{i=1}^{m}p_{cl_0,cl_1,cl_2,cz_m}$ =0.5031。它意味着，对于一条长 1282500 byte 的 A5/1 密钥流来说，对应于 3 元组(78，78，78)和区间 I 的复合序列的 0-1 偏差值平均仅为 ±151，此值比上一值略有提高，但是仍不足以将 A5/1 密钥流和随机比特流区分开。

为进一步加大A5/1 密钥流产生的概率曲线和随机比特流概率曲线之间的距离，需要将多个 3 元组(cl_0，cl_1, cl_2)对应复合曲线进行再复合。但一个 3 元组决定一个 $s_{cl_1}^i \oplus s_{cl_2}^i \oplus s_{cl_3}^i$ 的未知值，两个不同的 3 元组复合曲线可能分布在y轴的两侧。对于 $N_1(\mu_1,1)$ $N_2(\mu_2,1), \mu_2<0<\mu_1$ ，如果仅用函数 $z=(x_1+x_2)/\sqrt{2}$ 进行线性复合，那么得到的二次复合曲线 $N_z(\mu_1+\mu_2,1)$ 离y轴更近，更不容易区分A5/1 密钥流和随机比特流。

将与 3 元组(cl_0，cl_1, cl_2)对应的概率曲线沿y轴对称向x正半轴折叠，取 $|x|$ 作为“测试点”，可以有效解决上述问题。设原概率曲线的分布函数为 $f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2}}, -\infty<x<\infty$，则 $z=|x|$ 的分布函数为

$$f(z)=\begin{cases}\frac{1}{\sqrt{2\pi}}\left(e^{-\frac{(z-\mu)^2}{2}}+e^{-\frac{(z+\mu)^2}{2}}\right), & 0\le z<\infty\\ 0, & z<0\end{cases}$$

对于随机序列对应的标准正态分布的绝对值函数(简称标准绝对函数) $x'=|x|$， $\mu=0$，有

$$E(X')=\frac{2}{\sqrt{2\pi}}\int_0^\infty xe^{-\frac{x^2}{2}}\mathrm{d}x=\sqrt{\frac{2}{\pi}}=0.7979,$$

$$E(X'^2)=\frac{2}{\sqrt{2\pi}}\int_0^\infty x^2e^{-\frac{x^2}{2}}\mathrm{d}x=1,$$

$$D(X')=E(X'^2)-E(X')^2=\frac{\pi-2}{\pi}=0.3634。$$

而对于 3 元组(cl_0，cl_1, cl_2)对应的复合概率统计曲线的绝对值函数(简称三元组绝对函数) $x''=|x|$， $\mu\ne0$，有

$$E(X'')=\frac{1}{\sqrt{2\pi}}\int_0^\infty x\left(e^{-\frac{(x-\mu)^2}{2}}+e^{-\frac{(x+\mu)^2}{2}}\right)\mathrm{d}x$$

$$=\frac{1}{\sqrt{2\pi}}\left[\int_0^\infty e^{-\frac{(x-\mu)^2}{2}}d\left(\frac{(x-\mu)^2}{2}\right)+\mu\int_0^\infty e^{-\frac{(x-\mu)^2}{2}}\mathrm{d}x\right.$$

$$\left.+\int_0^\infty e^{-\frac{(x+\mu)^2}{2}}d\left(\frac{(x+\mu)^2}{2}\right)-\mu\int_0^\infty e^{-\frac{(x+\mu)^2}{2}}\mathrm{d}x\right]$$

$$=\frac{1}{\sqrt{2\pi}}\left[\left(e^{-\frac{(x-\mu)^2}{2}}+e^{-\frac{(x+\mu)^2}{2}}\Big|_\infty^0\right)+\mu\int_{-\mu}^\infty e^{-\frac{t^2}{2}}\mathrm{d}t-\mu\int_\mu^\infty e^{-\frac{s^2}{2}}\mathrm{d}s\right]$$

$$=\frac{1}{\sqrt{2\pi}}\left(2e^{-\frac{\mu^2}{2}}+\mu\int_{-\mu}^{\mu}e^{-\frac{x^2}{2}}\mathrm{d}x\right)=\sqrt{\frac{2}{\pi}}e^{-\frac{\mu^2}{2}}+\mu[2\Phi(\mu)-1]$$

$$E(X''^2)=\frac{1}{\sqrt{2\pi}}\int_0^{\infty}x^2\left(e^{-\frac{(x-\mu)^2}{2}}+e^{-\frac{(x+\mu)^2}{2}}\right)\mathrm{d}x$$

令 $t=x-\mu, s=x+\mu$，则有：

$$\begin{aligned}E(X''^2)&=\frac{1}{\sqrt{2\pi}}\left[\int_{-\mu}^{\infty}(t^2+2t\mu+\mu^2)e^{-t^2/2}\mathrm{d}t\right.\\&\quad\left.+\int_{\mu}^{\infty}(s^2-2s\mu+\mu^2)e^{-s^2/2}\mathrm{d}s\right]\\&=\mu^2\frac{1}{\sqrt{2\pi}}\left(\int_{-\mu}^{\infty}e^{-t^2/2}\mathrm{d}t+\int_{\mu}^{\infty}e^{-s^2/2}\mathrm{d}s\right)\\&\quad+\frac{2\mu}{\sqrt{2\pi}}\left(\int_{-\mu}^{\infty}te^{-t^2/2}\mathrm{d}t-\int_{\mu}^{\infty}se^{-s^2/2}\mathrm{d}s\right)\\&\quad+\frac{1}{\sqrt{2\pi}}\left(\int_{-\mu}^{\infty}t^2e^{-t^2/2}\mathrm{d}t+\int_{\mu}^{\infty}s^2e^{-s^2/2}\mathrm{d}s\right)\\&=\mu^2+\frac{2\mu}{\sqrt{2\pi}}\left(e^{-t^2/2}\Big|_{\infty}^{-\mu}-\Big|_{\infty}^{\mu}\right)\\&\quad+\frac{1}{\sqrt{2\pi}}\left[\int_{-\mu}^{\infty}-t\,d\left(e^{-t^2/2}\right)+\int_{\mu}^{\infty}-sd\left(e^{-s^2/2}\right)\right]\\&=\mu^2+0+\frac{1}{\sqrt{2\pi}}\left(te^{-t^2/2}\Big|_{\infty}^{-\mu}+se^{-s^2/2}\Big|_{\infty}^{\mu}\right.\\&\quad\left.+\int_{-\mu}^{\infty}e^{-t^2/2}\mathrm{d}t+\int_{\mu}^{\infty}e^{-s^2/2}\mathrm{d}s\right)\\&=\mu^2+\frac{1}{\sqrt{2\pi}}(0+\sqrt{2\pi})=\mu^2+1\end{aligned}$$

$$D(X'')=E(X''^2)-E(X'')^2=\mu^2+1-\left(\sqrt{\frac{2}{\pi}}e^{-\mu^2/2}+\mu[2\Phi(\mu)-1]\right)^2$$

式中 $\mu=\sqrt{\frac{n}{m}}\sum_{i=1}^{m}p_{\mathrm{cl}_0,\mathrm{cl}_1,\mathrm{cl}_2,\mathrm{cz}_m}$ 的大小不仅与递推公式(3)有关，也与选取的帧数 n 有关。当 n=90000 时，μ =0.5031，$E(X'')$ =0.8968，$D(X'')$ = 0.446，而对于标准绝对函数，$E(X')$ =0.7979，$D(X')$ =0.3634。两种函数的统计曲线见图 1。

可以看出，3 元组绝对函数的数学期望、方差与标准绝对函数数学期望、方差相比有很大不同，且两者差异随 μ 的增大而增大。

根据 Liapunov 定理，将大量独立同分布的不同随机变量相加，结果符合一定的正态分布，其数学期望为各个随机变量数学期望的和，方差为各个随机变量方差的和。第三次处理是为了增大 A5/1 输出序列与随机序列的区别。

设“测试点”数为 l ,各个“测试点”上的 μ 值分别为 $\mu_i(0<i<l)$。将多个“测试点”上的数据叠加，进行归一化整理得：

$E'(X')=0,\quad D'(X')=1;$

$$E'(X'')=\frac{\sum_{i=1}^{l}\left(\sqrt{\frac{2}{\pi}}e^{-\mu_i^2/2}+\mu_i[2\Phi(\mu_i)-1]\right)-l\sqrt{\frac{2}{\pi}}}{\sqrt{l(1-\frac{2}{\pi})}}$$

$$D'(X'')=\frac{\sum_{i=1}^{l}\left(\mu_i{}^2+1-\left(\sqrt{\frac{2}{\pi}}e^{-\mu_i^2/2}+\mu_i[2\Phi(\mu_i)-1]\right)^2\right)}{l(1-2/\pi)}\ E'(X'')$$

$D'(X'')$ 的值同样与选取的帧数 n 有关。当 n=90000 时，$E'(X'')$ =3.365，$D'(X'')$ =1.174; 当 n= 122500 时，$E'(X'')$ = 4.555，$D'(X'')$ =1.232; 当 n=160000 时，$E'(X'')$ = 5.912，$D'(X'')$ = 1.296。可见选取的帧数 n 越大，A5/1 序列的统计特征越暴露，伪随机性越差。当 n=90000 时，畸变序列的统计特征和标准正态曲线的对比见图 2。

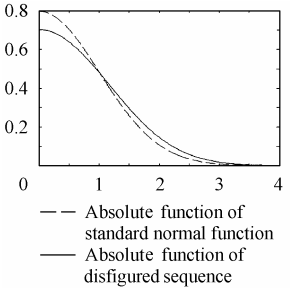

图 1 3 元组绝对函数与标准绝对函数的对比

Fig.1 Comparison of 3 item group absolute function with standard absolute function

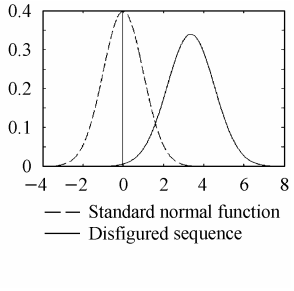

图 2 n=90000 时畸变序列产生的统计曲线与标准正态曲线的对比

Fig.2 Comparison of disfigured sequence curve with standard normal function curve when n=90000

4 实验数据

对 200 个随机产生的密钥，我们用 A5/1 算法进行加密，产生一些 A5/1 输出序列。同时我们随机产生一些随机序列。将这两种 0-1 序列都按照上述方法进行统计处理，以计算出的“归一化 0-1 偏差指数”作为横坐标，并对得出的数据按照坐标区间进行划分，以落在某个坐标区间内的频数作为纵坐标，得到的结果如图 3，图 4 所示：

由图 3，图 4 可以看出，在随机序列的计算值的实测情况中，峰值出现在－1 到 1 之间，而在 A5/1 算法输出序列的计算值的实测情况中，峰值出现在 1.6 到 3.95 之间。两者之间的区别非常明显。

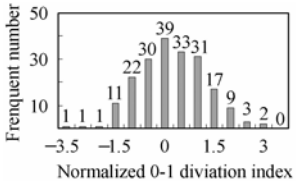

图 3 伪随机序列的计算值分布图

Fig.3 Distribution of caculated value of random sequence

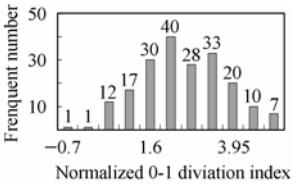

图 4 A5/1 算法输出的计算值分布图

Fig.4 Distribution of caculated value of A5/1 output sequence

5 结束语

理论分析和实验都证明，通过利用相关攻击的思想构造复合畸变密钥流，并通过对该序列的统计分析可以成功将

A5/1 输出密钥流序列和真随机序列区分开，从而达到算法识别的目的。作为密码分析的一种，算法识别中仍有很多可研究的课题，如是否存在一种通用的算法识别方法，各种密码分析手段在这一领域的具体应用，等等，都可以作为今后的进一步研究目标。

参 考 文 献

[1] Schneier B. Applied Cryptography. Second Edition: Protocols, Algorithms, and Source Code in C. New York: John Wiley & Sons, Inc. 1996：6−7.

[2] 王育民, 刘建伟. 通信网的安全——理论与技术. 西安：西安电子科技大学出版社, 1999：100−101.

[3] Ekdahl P, Johansson T. Another attack on A5/1. *IEEE Trans. on Information Theory*, January 2003, 49(1):284−289.

[4] 盛骤, 谢式千, 潘承毅. 概率论与数理统计. 高等教育出版社，第一版，1994: 131−139

[5] FIPS 140-1:Security Requirements for Cryptograghic Modules.

陈　伟：男，1973 年生，博士生，研究方向为密码分析、信息隐藏检测、NGN 安全等.

胡　云：男，1973 年生，博士生，研究方向为密码分析、信息隐藏和数字水印等.

杨义先：男，1961 年生，教授，博士生导师，主要研究方向为密码学、网络安全等.

钮心忻：女，1963 年生，教授，博士生导师，主要研究方向为信息隐藏和数字水印等.