随机性序列的典型纠偏方法及分析*)

杨 柱 吕述望 苏桂平

(中国科学院研究生院信息安全国家重点实验室 北京100039)

The Analysis of Typical Methods to Rectify the Deviation of Random Sequence

YANG Zhu LU Shu-Wang SU Gui-Ping

(State Key Laboratory of Information Security, Graduate School of Chinese Academy of Sciences, Beijing 100039, China)

Abstract According to the definition of random experimentation and the demand of random sequence in cryptography, three typical methods to rectify the deviation of random sequence are given below. We discussed the characters of their performance, furthermore gave the results of data analysis.

Keywords Random sequence, Cryptography, Rectify the deviation, Random test

1 引言

根据密码学中的 Kerckhoff 基本假设,假设密码分析者拥有所使用的算法的全部知识,密码系统的安全性完全寓于密钥之中[1],可见密钥的生成和管理是密码系统中的关键。

随机数发生器被用做密钥管理,其主要功能是产生真随机的序列来用作密钥。通常一个密码体制提供一个很大的密钥空间,如果系统的密钥是基于物理现象随机产生的,通常的字典攻击就很难奏效,从而最大程度地保持了密钥的原有空间尺度。

好的随机数发生器的输出序列要通过严格的随机性检验。在众多的随机性检验中,频数检验是最基本的一种检验。频数检验反映序列中0、1出现概率的差的大小。

2 随机序列的定义及其特征

概率统计中对随机试验的定义是:对于可重复的试验,如果每次试验的结果事先不可预测,这样的试验叫做随机试验[2]。进一步,我们定义随机试验的输出序列为随机性序列。

从文[3]中可知:密码学中取随机性的0、1序列作为密钥,它至少要满足以下三点:

(1)看起来是随机的,即能通过所能找到的所有正确的随机性检验。

(2)这个序列是不可预测的,也就是说,即使给出产生序列的算法或者硬件设计以及以前已经产生的序列的所有知识,也不可能通过计算来预测下一个比特是什么。

(3)这个序列不能重复产生,即使在完全相同的操作条件下,用完全相同的输入对序列产生器操作两次,也将得到两个完全不同的、毫不相关的位序列。

设随机性0、1序列中,1出现的概率 $P(1)=p$,0出现的概率为 $P(0)=1-P(1)$,从信息论的角度,1比特密钥的熵为:

$$H=-\{P(1)\log_2 P(1)+P(0)\log_2 P(0)\} \quad (1)$$

从上式可以看出,只有当 $P(0)=P(1)=0.5$,也就是用作密钥的随机性序列中0和1的出现概率相等时,1比特密钥的熵才能达到最大值1。

频数检验取长度为 N 的随机性序列,计算其中0的个数 $N(0)$和1的个数 $N(1)$,其统计量为:

$$V(N)=\frac{\{N(1)-N(0)\}^2}{N} \quad (2)$$

当 $V(N)$低于某一个设定的阈值时,认为序列通过频数检验。

3 随机序列的典型纠偏方法

(1)多路异或 首先将随机性序列的0、1偏差定义为:

$$\varepsilon=|P(1)-P(0)| \quad (3)$$

相应的统计量为:

$$E(N)=\frac{|N(1)-N(0)|}{N} \quad (4)$$

假设有两路随机性信号,其偏差为 $\varepsilon_1=|P_1(1)-P_1(0)|$、$\varepsilon_2=|P_2(1)-P_2(0)|$,那么两路按位异或输出序列的偏差为:

$$\varepsilon_3=\varepsilon_1\varepsilon_2 \quad (5)$$

可见,当 $\varepsilon_1,\varepsilon_2<0.5$的时候,$\varepsilon_3<\varepsilon_1,\varepsilon_2$。

两路的处理办法可以推广到 N 路。例如采用 N 路偏差都为 ε 的随机序列相异或,输出序列的偏差

$$\varepsilon_N=\varepsilon^N \quad (6)$$

两路异或纠偏的原理如图1,其中序列 DATA1、DATA2 按位异或得到序列 CORR1。

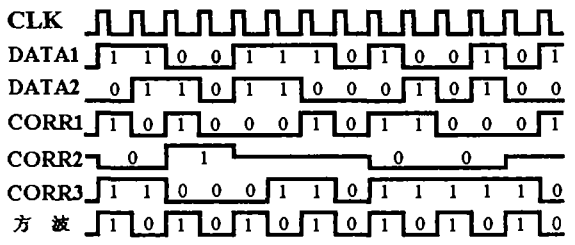

图1 纠偏方法的实例说明

多路异或是最常用的纠偏方法,效果也比较好,实际设计时,经过调节,序列的偏差往往很小,进行多路异或后偏差减小会很明显。例如采用两路偏差相同的随机数发生器相异或,每一路的偏差为0.001,则输出序列的偏差为0.000001,减小

*)基金编号:国家基础研究发展规划项目(编号:G1999035808)。杨 柱 硕士研究生,主要研究方向为信息安全和计算机应用。吕述望 教授,博士生导师,主要研究方向为信息安全。苏桂平 博士研究生,主要研究方向为信息安全。

了1000倍。

但是这种方法是通过增加系统规模的办法来换取小的偏差，因为采用 N 路相当于将系统的规模扩大了 N 倍。在随机数发生器的设计中，要从功耗、造价等方面均衡考虑来决定 N 值。

(2)诺依曼(Neumann)纠偏　在文[4]中，Jon von Neumann 提出了一种纠偏方法，即对于随机序列发生器输出的连续的2比特采取如下转换方法，本文称这种方法为诺依曼纠偏：

表1　诺依曼纠偏转换关系

输　入	输　出
00	不输出
01	0
10	1
11	不输出

其概率含义是显然的，随机性序列的各位彼此统计独立，则对于序列中长为2比特的串01和10，有：

$$P(01)=P(10)=P(0)P(1) \tag{7}$$

在图1中，序列 CORR2是序列 DATA2采用了诺依曼纠偏的结果。

式(7)表明这种纠偏方法可以实现0偏差，但是这种方法是以降低随机序列发生器的输出速率为代价的，例如在图1中，长为14比特的序列 DATA2，经过诺依曼纠偏之后，输出只有4比特。

假设序列中1和0出现的概率分别是 $P(1)$和 $P(0)$，而串01和10出现的概率为：$P_0=P(10)+P(01)=2P(0)P(1)$，则随机数发生器的输出速率降低为原来的$\frac{P_0}{2}$倍。

诺依曼纠偏另外一个突出的缺点是，随机序列发生器有否输出是不确定的，从概率的角度讲，会有连续多个节拍没有输出的情况发生，这会给使用带来很多不便。

(3)密码学处理　很多密码算法可作为伪随机数发生器，根据 Shannon 的建议，在密码算法的设计过程中采用混淆和扩散法。所谓混淆就是使明文和密文的统计独立性之间的关系复杂化。所谓扩散就是将每一位明文数字的影响尽可能迅速地散布到较多个输出的密文数字中。如果将随机数发生器的输出序列作为明文，则输出的密文的偏差可以很小。

另外，基于线性移位寄存器的伪随机序列发生器已经有深入的研究，用这种方法可以得到大周期、各态遍历性好的伪随机序列，如果用这样的序列和真随机数发生器的输出序列相异或，也可以得到偏差小的输出序列。举一个最简单的例子，当移位寄存器的长度为1的时候，输出的伪随机序列就是一个方波(当然，用这样的序列作为密钥毫无安全可言)，图1中序列 CORR3是 DATA2和方波相异或的结果。

用密码学处理纠偏的办法和前面两种纠偏的方法区别在于：异或纠偏和诺依曼纠偏是通过增加随机性序列单比特的熵而减小冗余的，而用密码算法或者伪随机序列处理过的序列的熵几乎没有变化，例如对于一个种子密钥长度为 N_k 的伪随机序列发生器，无论输出序列有多长，整个序列的熵都不会大于 N_k。

(4)多种方法的级连　上述三种方法可以级连起来使用，尤其是多路异或纠偏和诺依曼纠偏级连使用会得到更好的纠偏效果，不过第3种方法即使级联使用多次，也只能等效为一个变换。

4　实际数据分析

采用了某种随机序列发生器产生的4组长为1亿比特的序列，用上述4种方法加以处理，输出序列的统计量 $E(N)$如表2所示。

表2　数据分析结果

序号	源序列 E(N)($\times10^{-4}$)	2路异或纠偏	诺依曼纠偏		与方波异或	先异或纠偏后诺依曼纠偏($\times10^{-4}$)	
	($\times10^{-4}$)	E(N)($\times10^{-4}$)	E(N)($\times10^{-4}$)	输出序列长度(万)	E(N)($\times10^{-4}$)	E(N)($\times10^{-4}$)	输出序列长度(万)
1	137.2234	3.3936	2.5435	2344.78	5.087	0.0458	2490.42
2	134.4446		2.2773	2349.74	4.5546		
3	101.1306	0.7536	1.7207	2386.48	3.4414	0.3984	2495.47
4	74.6994		2.5059	2394.92	5.0118		

数据分析的结果表明，三种纠偏方法都能够很好地减小源序列的偏差。需要说明的一点是，对任何长度有限的随机性序列的分析，都不能完全地反映一个随机数发生器的行为，本文数据分析的目的是为了给出每一种纠偏方法的效果。

除了诺依曼纠偏之外，其他几种方法输出序列长度和源序列的长度相等。

结论和展望　本文总结了多路异或、诺依曼、密码学处理三种典型随机序列纠偏方法，随机数发生器的设计是安全信息系统的重要环节，如何设计、处理、评估随机数发生器是非常重要的问题，而纠偏是随机数发生器设计中最基本也是最重要的处理，不同的纠偏方法，对随机序列发生器的性能改善程度也不同，处理的复杂程度和系统规模也不同。基于纠偏，还可以设计一些密码学的密钥攻击策略。

参考文献

1　赵战生，等. 信息安全技术浅谈. 北京：科学出版社，1998

2　耿素云. 张立昂. 概率统计. 北京：北京大学出版社，1986

3　Schneier B. Applied Cryptography---Protocols, Algorithms, and Source Code in C, John Wiley & Sons, 1994

4　Jun B, Kocher P. The Intel Random Code Generator Cryptography Research, Inc. April 1999