第 33 卷 第 2 期 计 算 机 工 程 2007 年 1 月
Vol.33 No.2 Computer Engineering January 2007
·软件技术与数据库· 文章编号：1000—3428(2007)02—0071—03 文献标识码：A 中图分类号：T39

随机数发生器探讨及一种真随机数发生器实现

宋 勇，陈贤富，姚海东
(中国科学技术大学电子科学与技术系，合肥 230027)

摘 要：高质量随机数在信息安全中具有重要的应用。该文利用双重随机思想，基于真随机源对一个高质量伪随机数发生器进行参数重置，同时对该伪随机数发生器产生的随机数的特定二进制位进行采样从而获得 0-1 序列，将之填充到一个整数单元，归一化后作为输出。分析表明，该真随机数发生器具有良好的性能。

关键词：真随机数发生器；信息安全；双重随机

Discussion on High-quality RNG and Scheme of True RNG

SONG Yong, CHEN Xianfu, YAO Haidong
(Dept. of Electronic Sci. and Tech., Univ. of Sci. & Tech. of China, Hefei 230027)

【Abstract】 High quality random numbers play an important role in information security. Based on the idea of double randomizing, this paper uses the true random source to reset the parameters of a high quality pseudo-random number generator (p-rng), by which random numbers are produced continually. From these numbers, a certain binary bit is sampled respectively and meanwhile filled into an integer unit randomly which is normalized until being filled. Analysis shows that this true random number generator has good performance.

【Key words】 True random number generator; Information security; Double randomizing

密码被广泛应用于信息系统中，以实现系统信息的保密性、完整性、可用性、可控性和不可否认性。随机数在密码学中扮演着极其重要的角色，比如密钥管理、密码学协议、数字签名、身份认证等都需要用到随机数。如何产生高质量的随机数一直是一个经久不衰的话题，如今，它已成为密码学乃至信息安全领域的一个重要研究方向。本文首先对高质量随机数产生方法进行了探讨，然后给出一种真随机数实现方案，理论和实验分析显示利用这种方法产生的随机数质量较高。

1 高质量随机数发生器探讨

一般而言，高质量的随机数序列具有分布均匀、周期长、序列无关等特性。检验序列质量的方法有跟随性、游程、均匀性、独立性、相关性等一系列检验以及谱分析[9]、ENT(一种随机数性能检测程序)等。

生成随机数的方法繁多，从产生机理来说，可分为数学方法和物理方法两种，其所产生的随机数分别被称之为伪随机数和真随机数，前者易被破解，后者取自物理世界的真实随机源，难以破解，但这并不代表基于真随机源产生的随机数质量就很高，要取决于产生算法如何利用这个真随机源，相反的，许多用数学方法产生的随机数质量比较好。因此，若能将数学方法和物理方法结合起来，则可能产生高质量的真随机数。从实现方法来说，有以软件为主、以硬件为主以及软硬结合等方法[2~6,10]。

相比于伪随机数发生器的研究而言，真随机数发生器的研究还相当初步。设计一个真随机数发生器包括两步：首先是获取真随机源；然后是利用真随机源依照特定的数学方法获得真随机数。真随机源广泛存在于现实世界中，比如计算机网络中IP包到达的时间、随机噪音、计算机当前的秒级时钟、键盘反应时间、热噪声、操作系统的进程信息、纠缠光子对的纠缠[7]、光量子的偏振[12]等，获取方法可以通过调用系统函数或者硬件电路来实现[2,3,5]，比如文献[2,3]利用硬件电路产生热噪声、文献[5]则通过构建一个混沌电路来产生随机序列、文献[4,6]则采用系统调用的方法来获取系统进程和线程的随机特性。利用真随机源产生真随机数的方法有很多，一种最简单的方法是直接利用真随机源的奇偶特性来产生 0-1 序列。为了增加序列的随机性，往往还对产生的 0-1 序列进行一系列的变换，比如归一化、非线性映射、移位、加密等。比如，文献[8]利用计算机上可以获取的一些真随机源，将其随机位存放到一个缓冲区中，并不断执行这类操作，同时使用CRC-32 多项式来更好地混合缓冲区。当请求某个随机数时，就从缓冲区中读取随机数源，首先计算缓冲区中内容的MD5 值或SHA值，然后将MD5 值或SHA值反馈进缓冲区以保证下次取到的不是相同的值，同时将输出值折半以掩藏用户反馈进缓冲区的内容，防止别人根据本次输出值和反馈进缓冲区的内容计算得到下次输出值。用这种方法产生的随机数经统计分析具有良好的随机性。

真随机源的产生需要考虑可行性、经济性、速度等问题。有的硬件电路易于描述，但难以实现，有的方法实现成本较高。某些随机源必须涉及到人机交互，比如文献[8]提到的键盘随机性、鼠标随机性、中断随机性等，当不存在交互时，其随机源的随机性就大打折扣了。有的随机事件发生频率很低，利用这类随机源生成随机数的速度可能比较低，不适合于实时性要求高的场合。

以上讨论了真随机数的产生方法，另外一类不容忽视的

基金项目：国家自然科学基金资助项目(70071043)
作者简介：宋 勇(1979－)，男，博士生，主研方向：计算智能和复杂自适应系统；陈贤富，博士、副教授；姚海东，硕士生
收稿日期：2006-01-10 **E-mail：**skywhite@mail.ustc.edu.cn

方法是基于非线性的原理来产生高质量随机数，如基于混沌映射、元胞自动机、分形等原理的随机数发生器[10]。混沌系统的保密性比较好，因为它对初始条件十分敏感性，即使在映射关系被截取的情况下，只要种子的小数点后位数取得足够多，破解序列的可能性还是比较低的，但不容忽视的是，由于该类系统本质上是确定性的，如果递推公式和其中的某个状态被破解，后续的状态就会被破解，因此对于基于混沌原理的随机数发生器要求：(1)种子精度足够高；(2)映射关系足够复杂。如可以通过随机改变混沌映射的参数来提高混沌的复杂性[10]。一般而言，混沌映射仍然是确定性的，而有关文献则给出了一种不确定性的混沌映射，它不存在迭代关系、不依赖于初始条件，这为基于非确定性混沌原理产生高质量随机数奠定了基础。

元胞自动机(CA)也常用于产生随机数。CA 具有大规模并行性、局部交互性、细胞结构的简单性等优良特征，因而非常适合于快速、高效的硬件实现。Wolfram 第 1 个将 CA 应用到随机数发生器上，他深入研究了均匀规则 30 的动力学问题以及在随机数发生器上的应用，后来基于非均匀 CA 产生随机数成为主流，这方面的工作可参考文献[11]。

除了上述方法外，组合随机数发生器也是研究的热点[14]；Press等指出，双重随机化技术是获取高质量随机数的一种重要方法[13]，另有文献在分析一维游走理论时对其进行了验证。

2 一种真随机数发生器实现方案及分析

本文所提出的真随机数实现方案如图 1 所示。

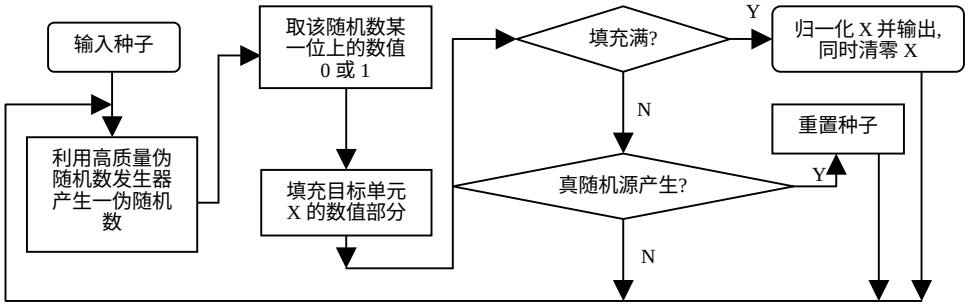

图 1　一种真随机数发生器实现方案

对该方案的补充说明：

(1)图中所示的伪随机数发生器可以利用有关文献给出的具有不可预测特性的混沌映射以及MINSTD线性同余发生器[1]等来实现。

(2)目标单元 X 为一个 32 位或者大于 32 位的整数单元，从伪随机数上采样的 0 或 1 随机地填充到该单元中，直到填充满为止。

(3)考虑到经济性、可行性等因素，真随机源采用网络 IP 包到达本机这一事件，对于具有迭代形式的伪随机数发生器，可以对 IP 包到达时间进行处理(比如取时间的低 16 位)作为伪随机数发生器的新种子。

(4)双重随机体现在伪随机和真随机两个方面，这使破解难度大大加大。

对[0,1]区间上均匀分布的随机数序列进行适当变换后，可以获得满足其它分布规律如高斯分布、泊松分布的随机数序列。那么利用上述随机数发生器产生的随机数序列是否满足[0,1]区间上的均匀分布特性呢？下面从理论上进行分析。

从统计角度说，计算机存储单元的某一二进制位满足 p=0.5 的 0-1 分布规律，用 x 表示该二进制位，则其期望和方差分别为 $E(x)=0.5$、$D(x)=0.25$。将采样得到的 x 逐位填充到一个含 N 位比特的目标单元 X 中，则有 $X=\sum_{i=0}^{N-1} x_i 2^i$。

x_i 表示填充的二进制位，对 X 归一化有

$$Y = X/(2^N-1) = \sum_{i=0}^{N-1} x_i \frac{2^i}{2^N-1}$$

显然，$0 \le Y \le 1$。现在的问题就是分析变量 Y 是否满足均匀分布特征。由概率论得

$$E(X) = E(x_i) = E(x) = 0.5$$

$$D(Y) = \frac{D(x_i)}{3}\frac{4^N-1}{(2^N-1)^2} = \frac{D(x_i)}{3}\frac{1-\frac{1}{4^N}}{1-\frac{1}{2^{N-1}}+\frac{1}{4^N}}$$

当 $N \to \infty$ 时，$D(Y) \to D(x_i) = D(x) = \frac{1}{12}$，当 N 取 32 时，$D(Y)$ 已经十分接近 $\frac{1}{12}$，因此采用这种方法得到的随机数序列具有[0,1]区间上均匀分布的特征，但这并不说明变量 Y 在[0,1]区间上是均匀分布的，用数学证明相对而言比较繁琐，这里采用对随机数序列样本进行假设检验的方法来验证其均匀分布性，下一节将对此进行说明。

3 实验及结果分析

本文所采用的高质量随机数发生器基于一种混沌映射，其形式如下：

$$X_n = \sin^2(\pi\theta z^n) \qquad (1)$$

它是 $X_{n+1} = \sin^2(z \arcsin\sqrt{X_n})$ 的精确解(z 取整数时)，当 z 分别取整数、分数、有理数、无理数等类型时，对应的第一回归映射(first return map)截然不同。当 z 为无理数时，其第一回归映射是毫无规律的，而且式(1)与一般的确定性混沌映射不同，它不依赖于初始条件，而仅与参数θ、z 和变量 n 取值有关，因此其具有不可预测特性，可以用来生成随机数。

式(1)中θ可动态改变，比如根据一个简单的混沌映射来确定，每生成一定数目(设为 M)的随机数即改变一次θ值，同时重置式(1)中的 n 为 0，但通过该式获得的随机数序列均匀性稍差，用式(2)进行转换后可以得到改善。

$$Y_n = 2/\pi \arcsin\sqrt{X_n} \qquad (2)$$

将这种方法应用到第 2 节所述的方案中，当真随机源产生时，重置种子就变成改变θ值，并重置 n 值，另外每隔 M 次改变一次θ和 n_0。由于 n 常被重置，使 n 不至于过大，式(1)的计算就不会很复杂，另外研究也表明采用改变θ值获得的整个随机数序列的随机性与不改变θ所获序列的随机性是相同的。

基于上述方案，采样了 262 144 个随机数(注意：这里 N 取 32，M 取 500)，用时 40s，其中共重置 117 次，以此为样本对其所代表总体的均匀分布性、独立性等进行检测。

(1)[0,1]分布均匀性检验

用 χ^2拟合优度检验法对样本作分布均匀性进行检验。将样本划分到 24 个等宽区间内，统计实际落在每个区间i内的样本个数n_i(见表 1)。

表 1 各个区间中的随机数数目

I	n_i	I	n_i
1	10 810	13	10 751
2	10 957	14	11 017
3	11 105	15	10 914
4	10 850	16	10 967
5	10 779	17	10 807
6	11 003	18	11 018
7	10 911	19	11 114
8	11 024	20	10 843
9	10 938	21	10 939
10	11 040	22	10 676
11	10 978	23	10 763
12	10 937	24	11 003

对于均匀分布,各随机数落在第 i 个区间内的概率 $p_i = \frac{1}{24}$。

按 χ^2拟合法，取统计量 $V = \sum_{i=1}^{l} \frac{(n_i - np_i)^2}{np_i}$，其中$l$为区间个数，n为样本数，计算得 $V = 28.02$。取 $\alpha = 0.05$，查自由度 $k = l - 1 = 23$ 的 χ^2分布表 $\chi^2_{0.052\,3} = 35.172 > V$，故应接受显著水平为 $\alpha = 0.05$ 的均匀性假设，因此可以认为这些样本所代表的总体服从 $u(0,1)$ 分布。

另外，作者也用柯尔莫哥洛夫提出的 Dn 检验(一种均匀性检测方法)法进行了检验，Dn＝0.072 9 < $D_{1\,000.10}$=0.120 67，故可接受均匀分布的假设检验。由上述理论分析和试验分析可知，利用本文所提出的随机数发生器产生的随机数满足[0,1]区间上均匀分布特性。

(2)独立性检验

用游程检验法来检验样本的独立性。将上述样本中各个值分别与 0.5 相减，取其符号得到一个正负号序列，统计其中正号个数n_1、负号个数n_2及游程数r如下：

$n_1 = 130\,910$， $n_2 = 131\,234$， $r = 131\,213$

假设这些随机数是独立分布的，取统计量 r，则有

$$E(r) = \frac{2n_1n_2}{n_1 + n_2} + \frac{1}{2} = 131\,072.799\,77,$$

$$D(r) = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)} = 65\,535.549\,8$$

r 近似满足正态分布，由于 $|Z_r| = \frac{r - E(r)}{\sqrt{D(r)}} = 0.547\,7$，查正态分布表 $Z_{\frac{\alpha}{2}} = 1.96 > |Z_r|$，因此可以认为随机数序列满足独立分布特性。

本文对 20 个相同大小的样本进行检测，均通过上述检验。同时，观察了其中一个样本的第一回归映射，如图 2 所示：由图可知，该映射毫无规律，也就不能根据历史状态推测出后续的状态，这对于信息安全是至关重要的。文献[9]给出了一种谱测试的方法用于分析随机数序列的性能，其二维散布图即上述的第一回归映射，但这种方法适合于具有迭代形式的随机数发生器，这里就不再讨论。

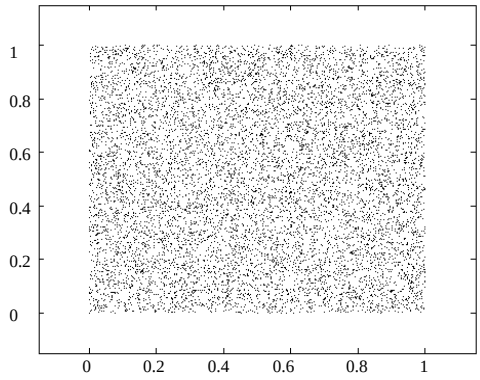

图 2 第一回归映射

4 结论

本文对高质量随机数发生器进行了探讨，并基于双重随机化思想提出了一种真随机数发生器方案，理论和实验分析显示该方法简单易行，所产生的随机数序列具有均匀性好、独立性高、周期长、速度快等特点，在信息安全以及随机过程模拟等方面具有较高的实际应用价值。

参考文献

1 Park S K, Miller K W. Random Number Generators: Good Ones Are Hard to Find[J]. Comm. of the ACM, 1988, 31(10): 1192-1201.

2 辛　茜, 曾晓洋, 张国权, 等. 真随机数发生器的系统建模与仿真[J]. 系统仿真学报, 2005, 17(1): 53-56.

3 王　莱, 刘松强. 真随机数发生器的设计和实现[J]. 核电子学与探测技术, 1998, 18(6): 452-455.

4 梁金千, 张　跃. 在计算机上产生真随机数的探讨[J]. 计算机工程, 2003, 29(15): 176-177.

5 俞　俊, 沈海斌, 严晓浪. 基于混沌的高速真随机数发生器的设计与实现[J]. 半导体学报, 2004, 25(8): 1013-1018.

6 黄　枫, 申　洪. 基于 Inter RNG 的真随机数生成器研究[J]. 第一军医大学学报, 2004, 24(9): 1091-1095.

7 马海强, 常　君, 吴令安. 基于纠缠光子对的真随机数源[C]. 第十届全国量子光学学术报告会论文摘要集, 2002.

8 袁卫忠, 谢俊元, 谢　立, 等. 网络安全中随机数技术分析与应用[J]. 计算机工程, 2001, 27(6): 116-117, 142.

9 Knuth D E. The Art of Computer Programming[M]. Addison-Wesley, 1981.

10 王相生, 甘骏人. 一种基于混沌的序列密码生成方法[J]. 计算机学报, 2002, 25(4): 351-356.

11 Hortensius P D, Mcleod R D, Card H C. Parallel Random Number Generator for VLSI Systems Using Cellular Automata[J]. IEEE Transactions on Computers, 1989, 38(10): 1466-1473.

12 冯明明, 秦小林, 周春源, 等. 偏振光量子随机源[J]. 物理学报, 2003, 53(1): 72-76.

13 Press W H, Flannery B P, Teukolsky S A, et al. Numerical Recipes in Fortran: The Art of Scientific Computing[M]. Cambridge University Press, 1996.

14 杨自强, 魏公毅. 常见随机数发生器的缺陷及组合随机数发生器的理论与实践[J]. 数理统计与管理, 2001, 20(1): 45-51.