一种混沌伪随机序列均匀化普适算法的改进

李佩玥[1],石俊霞[2],郭嘉亮[1],陈 雪[1],杨怀江[1]
(1. 中国科学院长春光学精密机械与物理研究所应用光学国家重点实验室,吉林长春 130033;
2. 中国科学院长春光学精密机械与物理研究所,吉林长春 130033)

摘 要: 为了分析盛利元等所述算法的安全性与普适性,从信息论的角度提出了单轮迭代信息损失量和动力学系统平均信息损失速度的概念,分析结果表明,第二类比特位变换的单轮迭代信息损失量为12比特,标准第二类比特位变换的单轮迭代信息损失量与指数 e 有关,存在信息损失量较小的可能性,将1023-e 作为移位位数,使得标准第二类比特位变换无法遍历[-1,1]区间内的所有浮点数. 进一步提出了暂态数据和第一类暂态变换的概念,并对文献[14]中所述算法进行了改进,改进后算法能够将任意混沌输出序列转换为至[0,1]区间内的浮点数,转换过程的信息损失量为 L-1 比特,接近有限计算精度为 L 时的最大信息损失速度 $I_{max}=L$,且通过 χ 检验可证明转换后的混沌输出序列服从均匀分布.

关键词: 数字混沌; 均匀性; 暂态数据; 暂态变换
中图分类号: TN911.21, O415.5 **文献标识码:** A **文章编号:** 0372-2112 (2015)04-0753-07
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2015.04.018

Improvement of a Universal Algorithm for Uniformization of Chaotic Pseudo-Random Sequences

LI Pei-yue[1], SHI Jun-xia[2], GUO Jia-liang[1], CHEN Xue[1], YANG Huai-jiang[1]
(1. *State Key Laboratory of Applied Optics, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, Jilin* 130033, *China*; 2. *Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, Jilin* 130033, *China*)

Abstract: In order to analyze the security and universality of the arithmetic proposed by Sheng et al, the concept of information loss in single iteration and average speed of information loss in dynamic system is proposed based on information theory. It is shown that the information loss of the 2nd bit-operation transformation is 12 bits, and which of the standard 2nd bit-operation transformation is related to the exponent e. It is possible that the information loss of the 2nd bit-operation transformation is so small. Not all of the float number in [-1,1] can be traversed by the standard 2nd bit-operation transformation just because the 1023-e is used as the shift number. The concept of transient data and 1st transient transformation is proposed further, and the arithmetic proposed in literature[14] is improved as well. The output sequence of random digital chaotic system can be transformed as float number in [0,1] by the improved arithmetic. The information loss of this transformation is L-1 bits, which is approached to the maximum speed of information loss $I_{max}=L$ under the computing precision L. The transformed sequence is uniform distributed which can be proved by χ-verification.

Key words: digital chaotic; uniformity; transient data; transient transformation

1 引言

伪随机序列是现代密码学中的重要组件之一,其广泛应用于数据加解密、数字签名与认证、统计分析、分布式计算等领域. 伪随机序列的基本特性有两种:

(1)随机性,即能够通过所能找到的所有正确的随机性检验.

(2)不可预测性,即给出产生序列的算法或者硬件设计和以前产生序列的所有知识,也不可能通过计算来预测下一个随机元素是什么. 伪随机序列发生器

收稿日期:2014-01-28;修回日期:2014-08-07;责任编辑:马兰英
基金项目:国家重点基础研究发展计划 973 项目基金(No. 2007CB311201);吉林省科技发展计划(No. 20130522120JH)

(Pseudo-Random Number Generator, PRNG)的设计与性能分析需尽可能地满足以上条件.

近年来,除通常意义上的设计方法(如基于格子结构与分布[1]、基于线性反馈移位寄存器的动态反馈策略[2]、基于符号方法[3]等)外,基于非线性动力学系统,特别是混沌系统的 PRNG 设计方法层出不穷:文献[4]提出了一种基于 Logistic 映射的伪随机序列发生器,并对产生的伪随机序列进行了多方面的性能测试;文献[5]则采用针对性设计的编码算法解决了采样后的 Chen 系统相空间分布不均匀的问题,并以此为基础,完成了 PRNG 的设计;文献[6]在没有使用比例缩放或离散操作的前提下,采用数字电路实现了 ITCM 混沌映射,并实现了均匀分布的二进制序列的输出;文献[7~9]分别采用了耦合映射格子(CML-MPRBG)、Chebyshev 映射以及三混合混沌映射输出伪随机序列. 但上述算法多以改善混沌输出序列的均匀性为目标,并未过多考虑算法的安全性问题,导致对算法所生成伪随机序列的分析较为片面(如文献[10]中的谱熵算法对 Logistic 映射、Gaussian 映射和 TD-ERCS 系统产生的混沌伪随机序列复杂度的分析,文献[11]对伪随机序列线性复杂度的分析等),并未从本质上提高所生成伪随机序列的密码学特性,使得算法本身极易受到攻击. 文献[12]仅使用统计测试的方法即对文献[5]中所述算法实现了有效攻击,而文献[13]则利用有限精度实现浮点数的周期性证明了文献[4]中所述算法的不安全性.

相比于近期发表的上述文献,文献[14]中提出的一种混沌伪随机序列均匀化普适算法是目前较为有效的混沌伪随机序列生成方法之一,该算法基于计算机浮点数表示的 bit 位操作,不针对任何具体对象,可将任意分布(只要求连续或分段连续)的实型绥机变量转换成均匀分布的随机变量,并且运算速度快,易于硬件实现[15]. 但是,该算法却同样存在上述算法的安全性问题,本文从信息论的角度对该算法的安全性进行了定量分析,将信息论中信息及信息熵的概念引入到混沌伪随机序列发生器的评价中,并提出了动力学系统单轮迭代信息损失量和动力学系统平均信息损失速度的概念.

分析结果表明,该算法并不能生成高安全性伪随机序列,其主要原因是算法中使用了大量可逆运算,整个算法的逆向推导过程的不确定性较小,此外,本文也从算法实现的角度证明了文献[14]的标准第二类比特位变换并不能遍历[-1,1]区间内的所有浮点数,算法本身不具有普适性. 在文献[14]部分结论的支撑下,本文采用暂态变换的方法对其算法进行了改进,使得改进后算法既能够满足伪随机序列均匀性的要求,又可以将序列相邻元素间的信息关联性降到最小,以增强其序列元素间的不可预测性.

2 混沌伪随机序列均匀化普适算法

IEEE754 标准规定,一个实数的双精度二进制表示由三部分组成:1-bit 符号位(用 s 表示),11-bit 有偏指数位(用 e 表示),52-bit 尾数位(用 f 表示),由

$$(-1)^s \times 2^{(e-1023)} \times 1.f, 0 < e < 2047, s \in \{0,1\} \quad (1)$$

换算到十进制数或二进制数.

为了更好的阐述混沌伪随机序列均匀化普适算法,根据上述表示方法,文献[14]中做了如下定义:

定义 1 如果尾数 f 中第 51-bit,50-bit,…,(51-b+1)-bit 均为"0",而(51-b)-bit 为"1",则将 f 的前 b 位依次移到 f 的右边构成一个新尾数 f,这样的操作称为左移位 b 操作,新尾数 f 记为 $f_{\leftarrow b}$.

定义 2 将尾数 f 中第 0-bit 设置为"1",第 1-bit,2-bit,…,(b-1)-bit 均设置为"0",然后将此 b 位依次移到 f 的左边构成一个新尾数 f,这样的操作称为右移位 b 操作,新尾数记为 $f_{\rightarrow b}$.

定义 3 f_H 与 f_L 对应 bit 位进行异或运算,记为

$$f_H' = f_H \oplus f_L = f_{51}' f_{50}' \cdots f_{27}' f_{26}' \quad (2)$$

其中 $f_i' = f_{51-i} \oplus f_i, i = 0,1,2,\cdots,24,25$,再将 f_H' 与 f 按照高位和低位合并构成一个新的尾数,即

$$f' = f_H' f_L = f_{51}' f_{50}' \cdots f_{27}' f_{26} f_{25} f_{24} \cdots f_1 f_0 \quad (3)$$

称 f' 为尾数 f 的 bit 位变换,记为 Bit$\{f\}$,即 $f' = \text{Bit}\{f\}$,f' 称为 bit 位变换的变换核.

定义 4 $\{s,e,\text{Bit}\{f\}\}$ 称为 $\{s,e,f\}$ 的第一类 bit 位变换,用 $B_1(x)$ 表示.

定义 5 对 Bit$\{f\}$ 进行左移位 b 操作后得 Bit$\{f\}_{\leftarrow b}$,且令 $e' = 1023 - b, s = 0$,新的实数 $\{0,e',\text{Bit}\{f\}_{\leftarrow b}\}$ 称为 $\{s,e,f\}$ 的第二类 bit 位变换,用 $B_2(x)$ 表示.

定义 6 若 $\{s,e,f\} \in [-1,1]$,$\{0,1023-b,\text{Bit}\{f_{\rightarrow(1023-e)}\}_{\leftarrow b}\}$ 称为 $\{s,e,f\}$ 的标准第二类 bit 位变换,仍用 $B_2(x)$ 表示.

定理 1 设实型随机变量 $\xi \in G$ 的分布函数 $F_\xi(x) = P\{\xi < x\}$ 连续(或分段连续),若对 ξ 进行第二类 bit 位变换(或标准的第二类 bit 位变换)成随机变量 η,则 $\eta \in [0,1]$,且其分布函数 $F_\eta(x)$ 以不大于 2^{-52} 的理想偏差服从均匀分布,即

$$P_\eta(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_\eta(x) = 1 \quad (4)$$

该定理表明,对一个实数表示的随机变量,只要采用第二类 bit 位变换(或标准的第二类 bit 位变换),就能获得均匀分布的随机数,简单地说,若 ξ 是一个实的随机变量,则 $B_2(\xi)$ 是[0,1]上均匀分布的随机变量,这就是混沌伪随机序列均匀化普适算法.

3　文献[14]中算法的分析

3.1　动力学系统的信息损失

在信息论中,当信源发出的消息通过信道传输给接收者后,才能消除不确定性并获得信息. 如果信源中某一消息发生的不确定性越大,一旦它发生,并为接收者收到后,消除的不确定性就越大,获得的信息也就越大,换言之,事件发生的概率越小,不确定性就越大,所获得的信息量就越大,反之,则越小. 通常情况下,信息量的多少可通过下式计算:

$$I(a_i)=\log_2\frac{1}{P(a_i)} \tag{5}$$

其中,$P(a_i)$ 为事件 a_i 发生的概率.

而对于输出序列 $\{a_i\}_{i=0}^{\infty}$ 的信源 X 而言,其平均信息量(信息熵)可定义为:

$$H(X)=E\left[\log_2\frac{1}{P(a_i)}\right]=-\sum_{i=1}^{\infty}P(a_i)\log_2P(a_i) \tag{6}$$

在有限计算精度下,对于任意动力学系统 S 而言,若其第 n 次迭代的输入、输出状态分别为 $\boldsymbol{x}_n$ 和 $\boldsymbol{x}_{n+1}$,即 $\boldsymbol{x}_{n+1}=S(\boldsymbol{x}_n)$,由信息量与信息熵的定义可知,动力学系统 S 每次迭代的反向推导过程 $\boldsymbol{x}_{n+1}\rightarrow\boldsymbol{x}_n$ 均是一个熵增大的过程,而其正向推导过程 $\boldsymbol{x}_n\rightarrow\boldsymbol{x}_{n+1}$ 则是一个原始信息减少、新信息引入的过程. 若假设由 $\boldsymbol{x}_{n+1}$ 反向推导 $\boldsymbol{x}_n$ 有 u_n 种可能性,则定义本次迭代的正向迭代信息损失量(或反向推导的不确定性)为:

$$I_S(n)=\log_2u_n \tag{7}$$

动力学系统 S 的平均信息损失速度为:

$$K_S=H(S)=E[I_S(n)] \tag{8}$$

特别地,当动力学系统 S 为混沌系统时,由 Kolmogorov 熵的物理意义可知,其信息损失速度即为混沌动力学系统的 Kolmogorov 熵.

对于任何在计算机上实现的数字动力学系统 S_D,若假设计算机的有限计算精度为 L 比特,则其每轮迭代的正向迭代信息损失量(或反向推导的不确定性)的最大值为 $I_{\max}=L$,即每轮迭代都将损失全部的原始信息,已知本轮迭代结果,反向推导本轮迭代初值将有 2^L 种可能,此时,动力学系统 S_D 的平均信息损失速度 $K_{S_D}=L$. 从信息论的角度来说,在这种情况下,迭代初值与迭代结果之间已完全没有信息上的关联,由该动力学系统生成的伪随机时间序列将具有最优的密码学特性. 但是,这种动力学系统是不存在的,其仅可作为密码学中伪随机序列发生器设计的最终目标.

命题 1　在有限计算精度下,对于由混沌系统 S_1 和任意动力学系统 S_2 构成的复合动力学系统 S:

$$\begin{cases}\boldsymbol{x}'_{n+1}=S_1(\boldsymbol{x}_n)\\ \boldsymbol{x}_{n+1}=S_2(\boldsymbol{x}'_{n+1})\end{cases} \tag{9}$$

若假设混沌系统 S_1 的平均信息损失速度为 K_{S_1},动力学系统 S_2 的平均信息损失速度为 K_{S_2},则该复合动力学系统的平均信息损失速度为:

$$K_S=K_{S_1}+K_{S_2} \tag{10}$$

证明　易知式[9]所示动力学系统的单步迭代过程为 $\boldsymbol{x}_n\rightarrow\boldsymbol{x}'_{n+1}\rightarrow x_{n+1}$.

对于有限计算精度实现的混沌系统 S_1 而言,其混沌轨道已不仅仅具备无限精细的分形结构,截断误差的存在使得混沌轨道中出现了大量的短周期轨道和多对一映射,此时,混沌轨道每步迭代的反演过程将具有较大的不确定性. 而任意动力学系统 S_2 也存在引入多对一映射的可能性,其单步迭代的反演过程也将具有一定的不确定性.

假设已知 $\boldsymbol{x}_{n+1}$,求解 $\boldsymbol{x}'_{n+1}$ 共有 u_n 种可能性,则反向迭代过程 $\boldsymbol{x}_{n+1}\rightarrow\boldsymbol{x}'_{n+1}$ 的不确定性为:

$$I'(n)=\log_2u_n \tag{11}$$

假设已知 $\boldsymbol{x}'_{n+1}$,求解 $\boldsymbol{x}_n$ 共有 v_n 种可能性,则反向迭代过程 $\boldsymbol{x}'_{n+1}\rightarrow\boldsymbol{x}_n$ 的不确定性为:

$$I''(n)=\log_2v_n \tag{12}$$

由组合代数中的乘法原理可知,若已知 $\boldsymbol{x}_{n+1}$,则 $\boldsymbol{x}_n$ 共有 u_nv_n 种可能性,因此,反向迭代过程 $\boldsymbol{x}_{n+1}\rightarrow\boldsymbol{x}'_{n+1}\rightarrow\boldsymbol{x}_n$ 的不确定性为:

$$I(n)=\log_2u_nv_n=\log_2u_n+\log_2v_n=I'(n)+I''(n) \tag{13}$$

则由式[8]易知 $K_S=K_{S_1}+K_{S_2}$. 证毕.

3.2　两类比特位变换的分析

第二类比特位变换 $\{s,e,f\}\rightarrow\{0,1023-b,\mathrm{Bit}\{f\}_{\leftarrow b}\}$,已知变换后的浮点数 $\{0,1023-b,\mathrm{Bit}\{f\}_{\leftarrow b}\}$,易计算求得 b 值,再根据左移 b 位操作的定义可求解 $\mathrm{Bit}\{f\}$,而由于比特位变换 $f'=\mathrm{Bit}\{f\}$ 是可逆变换,因此,尾数 f 的计算不存在任何不确定性. 由双精度浮点数的定义可知,符号位 s 和指数位 e 共有 2^{12} 种可能,换言之,文献[14]中第二类比特位变换的正变换信息损失量(或反变换的不确定性)为 $I'_{B_2}=\log_22^{12}=12$.

标准第二类比特位变换 $\{s,e,f\}\rightarrow\{0,1023-b,\mathrm{Bit}\{f_{\rightarrow(1023-e)}\}_{\leftarrow b}\}$,已知变换后的浮点数 $\{0,1023-b,\mathrm{Bit}\{f_{\rightarrow(1023-e)}\}_{\leftarrow b}\}$,易计算求得 b 值,再根据左移 b 位操作的定义可求解 $\mathrm{Bit}\{f_{\rightarrow(1023-e)}\}$,由于比特位变换 $f'=\mathrm{Bit}$

$\{f\}$ 的可逆性，$f_{\rightarrow(1023-e)}$ 的计算不存在任何不确定性. 由右移操作的定义，则可根据 $f_{\rightarrow(1023-e)}$ 中高位 0 的个数确定右移操作的移位数 $(1023-e)$，即可求得指数 e. 但是，在右移操作中，需要将尾数中的 b 比特位强制置为“0”或“1”，因此，已知 $f_{\rightarrow(1023-e)}$ 求解 f 仍存在 2^{1023-e} 种可能，符号位 s 仍存在“0”和“1”两种可能，即文献[14]中标准第二类比特位变换的正变换信息损失量（或反变换的不确定性）为 $I''_{B_2}=\log_2 2^{1024-e}=1024-e$，且 $I''_{B_2}\in[1,52]$. 此外，由 $\{s,e,f\}\in[-1,1]$ 可知，$e-1023<0$，而移位位数 $1023-e$ 应满足 $1023-e<52$，即 $e\in(1023-52,1023)$，显然，在此条件的限制下，$\{s,e,f\}$ 并无法遍历 $[-1,1]$ 区间内的所有浮点数，算法本身并无普适性.

综上所述，文献[14]中所述混沌伪随机序列均匀化普适算法由于采用了“$\leftarrow b$”和“Bit $\{f\}$”等可逆变换，使得变换前后的浮点数间存在很强的信息上的关联，其定量表现为正变换信息损失量（或反变换的不确定性）较小或变化范围较大，逆向求解的可能性较大. 将上述变换应用于任何数字混沌动力学系统，由命题 1 可知，变换后的系统仅可保证输出混沌序列的均匀性得到改善，并无法彻底消除序列元素间信息上的关联，即无法保证输出混沌序列的密码学要求，并不适于高密级环境中的密码设计. 特别地，对于标准第二类比特位变换，由于将 $1023-e$ 作为右移操作的移位位数，限制了指数 e 的范围，导致该算法无法遍历 $[-1,1]$ 区间内的所有浮点数，降低了算法的普适性.

4 算法改进

对于浮点数计算而言，除非引入与尾数无关且具有良好密码学特性的二进制序列，否则无论使用哪种变换，均无法将原浮点数中的原始信息快速损失完全，而一旦有方法产生这种二进制序列，那么，再将其应用于浮点数变换以重新产生伪随机序列的必要性是值得商榷的. 换句话来说，对于浮点数实现的混沌系统，若既要保证输出混沌序列均匀性，又要使得该混沌系统的平均信息损失速度接近 $I_{\max}$ 是很难实现的. 因此，本文采用定点数实现的方法，对文献[14]中的混沌伪随机序列均匀化普适算法进行了改进.

假设有限计算精度为 L 比特，R_L 为该有限计算精度下的实数集，对于任意定点数 $x_{fx}\in R_L$，其均可由以下三部分表示：符号位 s，小数点位置 p 和尾数 b，并通过 $x_{fx}=(-1)^s\times 2^{-p}\times b$ 换算到十进制或二进制数. 通常情况下，有限计算精度下的定点数运算仅发生在小数点位置 p 相同的定点数之间，故而定点数的小数点尾数 p 并不占用其存储空间，而仅是存放在其他寄存器中，即定点数 x_{fx} 在有限计算精度 L 比特下的存储格式可表示为 $sb_{L-2}b_{L-1}\cdots b_2b_1b_0$，$s,b_i\in[0,1]$. 其中，二进制字符串 $b_{L-2}b_{L-1}\cdots b_2b_1b_0$ 为长度为 $L-1$ 比特的尾数 b.

定义 7 假设定点数 $u,v,w\in R_L$，且 u,v,w 为分别可表示为 $\{s_u,p_u,b_u\}$，$\{s_v,p_v,b_v\}$，$\{s_w,p_w,b_w\}$，若 $w=u\times v$，即有 $p_u=p_v=p_w$，$s_w=(-1)^{s_u+s_v}$，则称 b'_w 为 $w=u\times v$ 过程中的暂态数据，如果 b'_w 满足：

(1) b'_w 为长度为 $L-2$ 比特的二进制字符串.

(2) b'_w 可表示为 $\omega_{L-3}\omega_{L-4}\cdots\omega_2\omega_1\omega_0$.

其中，$\omega_i,i=0,1,2,\cdots,L-3$ 为结果 $\omega=b_u\times b_v$ 的低 $L-2$ 比特. b_u,b_v 均为长度为 $L-1$ 比特的二进制比特串，而由乘法运算的性质可知，$\omega=b_u\times b_v$ 应为长度为 $2L-3$ 的二进制比特串，但是，有限的计算精度使得定点数运算 $w=u\times v$ 最终只保留了 ω 中高 $L-1$ 比特的计算结果作为乘积 w 的尾数 b_w，即 $\omega_{2L-4}\omega_{2L-5}\cdots\omega_L\omega_{L-1}\omega_{L-2}$，而舍弃了低 $L-2$ 比特的二进制比特串 b'_w. 文献[14]中已证明以下结论：设实型随机变量 $\xi\in[0,1]$，具有连续（或分段连续）的概率密度函数 $P_\xi(x)$，将 ξ 表示成二进制形式 $0.\xi_1\xi_2\xi_3\cdots\xi_i\cdots$，则 $\xi_i\in\{0,1\}$，$i=1,2,3,\cdots$ 是一个二值随机变量序列，且当 $i\rightarrow\infty$ 时，ξ_i 趋于均匀分布，即 $\lim\limits_{n\rightarrow\infty}P(\xi_n=0)=\lim\limits_{n\rightarrow\infty}P(\xi_n=1)$. 因此，之所以称 b'_w 为暂态数据，是因为其并不作为运算结果输出，攻击者无法通过算法的外部输出数据进行攻击，但相比于 b_w 而言，b'_w 却具有更好的均匀性，虽然“暂态数据”的定义由乘法运算得到，但易证明混沌系统中常用到的乘方、除法、微分、积分等运算在数值计算的过程中均将不可避免的产生“暂态数据”. 从另外一个角度来说，在混沌系统的迭代过程中，对于密码算法而言，定点运算过程中产生的“暂态数据”将是服从均匀分布的理想伪随机信号源.

定义 8 假设存在定点数 $u,v,w\in R_L$，w 为 u,v 相互运算的结果，其可表示为 $\{s_w,p_w,b_w\}$，其中 b_w 为 $L-1$ 比特的二进制比特串，即有 $b_w=\alpha_{L-2}\alpha_{L-3}\cdots\alpha_2\alpha_1\alpha_0$，$\alpha_i\in\{0,1\}$，若 u,v 相互运算过程中产生的 $L-2$ 比特暂态数据为 $b''_w=\beta_{L-3}\beta_{L-4}\cdots\beta_2\beta_1\beta_0$，$\beta_i\in\{0,1\}$，则将 b''_w 低位填“0”后与 b_w 逐位异或，并记作 b'_w，即：$b'_w=\alpha_{L-2}\alpha_{L-3}\cdots\alpha_2\alpha_1\alpha_0\oplus\beta_{L-3}\beta_{L-4}\cdots\beta_2\beta_1\beta_00$，称 b'_w 为 b_w 的暂态变换，记作 $b'_w=\mathrm{Trans}(b_w)$.

定义 9 $\{0,L,\mathrm{Trans}(b_w)\}$ 称为 $\{s_w,p_w,b_w\}$ 的标准第一类暂态变换，用 $T_1(x)$ 表示，显然变换后的定点数 $x'_{fx}\in[0,1]$.

定理 2 在有限计算精度 L 的定点数实现条件下，假设数字混沌系统 S_D 第 n 轮迭代的输出状态为 $\boldsymbol{x}_n$，若对 $\boldsymbol{x}_n$ 进行第一类暂态变换成为新的输出状态 $\boldsymbol{x}'_{n+1}$，则

数字混沌系统 S_D在相空间中任一维度上的每轮迭代输出均将至少以 $L-1$ 比特的平均信息损失速度损失原始信息,且输出状态序列$\{\boldsymbol{x}'_n\}_{n=0}^{\infty}$在相空间中任一维度上服从均匀分布.

假设数字混沌系统 S_D的相空间维度为 d,则其第 n 轮迭代的输出状态 $\boldsymbol{x}_n$ 可表示为 $\boldsymbol{x}_n=(\boldsymbol{x}_n^{(1)},\boldsymbol{x}_n^{(2)},\cdots,\boldsymbol{x}_n^{(d)})$,若定点数 $\boldsymbol{x}_n^i$ 可表示为$\{s_n^{(i)},p_n^{(i)},b_n^{(i)}\}$,其中 $i=1,2,\cdots,d$,则第一类暂态变换后的相应输出状态可表示为$\{0,L,\mathrm{Trans}(b_n^{(i)})\}$. 由暂态数据的定义可知,对尾数 $b_n^{(i)}$ 的暂态变换需要使用 $L-2$ 比特的暂态数据,对于算法攻击者而言,此部分数据为不可见数据,这将为由 $\mathrm{Trans}(b_n^{(i)})$反向推导 $b_n^{(i)}$ 带来 $2^{(L-2)}$种可能性,此外,符号位 s 仍会带来两种可能性,显然,数字混沌系统 S_D在相空间中任一维度上的每轮迭代均将以 $L-1$ 比特的平均信息损失速度损失原始信息. 此外,文献[14]中的定理 1 - 定理 3 已经对输出状态序列$\{\boldsymbol{x}'_n\}_{n=0}^{\infty}$在相空间中任一维度上的均匀性进行了证明,在此不再赘述.

该定理表明,对于定点数实现的任意数字混沌系统,只要在其每轮迭代后对相空间状态 $\boldsymbol{x}_i$ 采用第一类暂态变换,就能获得[0,1]区间均匀分布的随机数,且能保证混沌输出状态序列的相邻元素间仅存在不大于 $I_{\max}-(L-1)=1$ 比特的信息关联性,这就是改进后的混沌伪随机序列均匀化普适算法.

5　改进后算法的验证

与文献[14]的一样,本文以 Logistic 映射、Henon 映射、Lorenz 系统为例,采取第一类暂态变换,分别统计变换前后的概率密度. 系统采用 128 比特定点数实现,进行 20000 次迭代,并采用假设检验的方法判断混沌输出状态序列是否服从均匀分布,检验方法如下:

(1)假设 $H_0:x_1,x_2,x_3,\cdots,x_{N-1},x_N$ 服从均匀分布.

(2)将相空间平均分为 m 个小区间,$x_1,x_2,x_3,\cdots,x_{N-1},x_N$ 落入第 i 个小区间的数目为 n_i,第 i 个小区间的理论频数为 $\mu_i=\dfrac{N}{m}$,其中,$i=1,2,\cdots,m$.

(3)计算统计量 $V=\sum_{i=1}^{m}\dfrac{(n_i-\mu_i)^2}{\mu_i}$.

统计量 V 渐进服从 χ^2_{m-1}分布,若定义显著性水平为 α(统计量 V 落入拒绝域的概率),则当 $V<\chi^2_{m-1}(\alpha)$ 时接受 H_0,当 $V>\chi^2_{m-1}(\alpha)$时拒绝 H_0,为了能够正确评价混沌输出序列的均匀性,本文选取 $\alpha=0.05$,$m=100$,此时$\chi^2_{99}(0.05)=123.23$.

5.1　Logistic 映射

Logistic 映射采用形式 $x_{n+1}=1-\mu x_n^2$,$x_n\in[-1,1]$,$n=0,1,2,\cdots$,当 $\mu=2$ 时,采用改进后的混沌伪随机序列均匀化普适算法,可得 $x'_n=T_1(x_n)$,$x'_n\in[0,1]$,改进前后的统计结果对比如图 1 所示. 统计表明,输出混沌状态序列$\{x'_n\}_{n=0}^{\infty}$的统计量 $V=97.45<\chi^2_{99}(0.05)=123.23$,接收假设 H_0,即其服从均匀分布. 对比实验表明:本文算法可实现文献[14]中所述算法的均匀化效果,同时,暂态数据的应用使得伪随机序列的相邻元素间存在不大于 1 比特的信息关联,从理论上了保证了该算法的安全性高于文献[14]所述算法.

5.2　Henon 映射

Henon 映射的形式为:

$$\begin{cases}x_{n+1}=1-ax_n^2+y_n\\ y_{n+1}=bx_n\end{cases}$$

其中 $n=0,1,2,\cdots$,取 $a=1.4$,$b=0.3$,$x_0=0.3345$,$y_0=0.01$. 采用改进后的混沌伪随机序列均匀化普适算法,可得 $x'_n=T_1(x_n)$,$x'_n\in[0,1]$,$y'_n=T_1(y_n)$,$y'_n\in[0,1]$,改进前后的统计结果对比如图 2 所示. 统计表明,输出混沌状态序列$\{x'_n\}_{n=0}^{\infty}$和$\{y'_n\}_{n=0}^{\infty}$的统计量 $V_x=91.36<\chi^2_{99}(0.05)=123.23$,$V_y=99.24<\chi^2_{99}(0.05)=123.23$,接收假设 H_0,即二者均服从均匀分布. 对比实验表明:本文算法可实现文献[14]中所述算法的均匀化效果,同时,暂态数据的应用使得伪随机序列的相邻元素间存在不大于 1 比特的信息关联,从理论上了保证了该算法的安全性高于文献[14]所述算法.

5.3　Lorenz 系统

Lorenz 系统的形式为:

$$\begin{cases}\dot{x}=\sigma(y-x)\\ \dot{y}=rx-xz-y\\ \dot{z}=xy-bz\end{cases}$$

取系统参数 $\sigma=10$,$r=28$,$b=8/3$,初始点 $x_0=0.3$,$y_0=0.4$,$z_0=0.5$,采用改进后的混沌伪随机序列均匀化普适算法,可得 $x'_n=T_1(x_n)$,$x'_n\in[0,1]$,$y'_n=T_1(y_n)$,$y'_n\in[0,1]$,$z'_n=T_1(z_n)$,$z'_n\in[0,1]$,改进前后的统计结果对比如图 3 所示. 统计表明,输出混沌状态序列$\{x'_n\}_{n=0}^{\infty}$、$\{y'_n\}_{n=0}^{\infty}$、$\{z'_n\}_{n=0}^{\infty}$的统计量 $V_x=98.64<\chi^2_{99}(0.05)=123.23$,$V_y=101.56<\chi^2_{99}(0.05)=123.23$,$V_z=112.69<\chi^2_{99}(0.05)=123.23$ 接收假设 H_0,即三者均服从均匀分布. 对比实验表明:本文算法可实现文献[14]中所述算法的均匀化效果,同时,暂态数据的应用使得伪随机序列的相邻元素间存在不大于 1 比特的信息关联,从理论上了保证了该算法的安全性高于文献[14]所述算法.

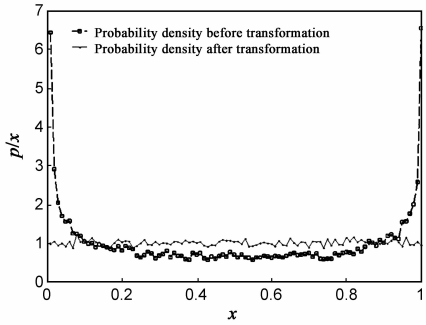
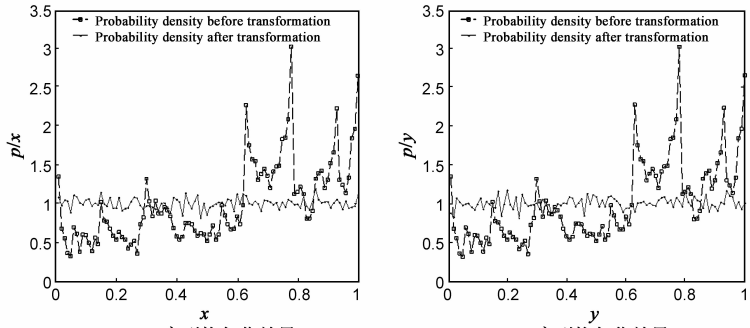
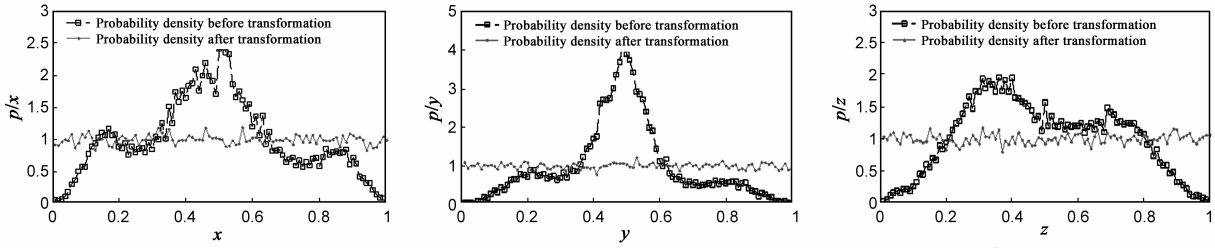


图1 改进后算法对Logistic映射序列均匀化的实验结果



(*a*) *x*序列均匀化效果 (*b*) *y*序列均匀化效果

图2 改进后算法对Henon映射序列均匀化的实验结果



(*a*) *x*序列均匀化效果 (*b*) *y*序列均匀化效果 (*c*) *z*序列均匀化效果

图3 改进后算法对Lorenz系统序列均匀化的实验结果

6 结论

本文从信息损失的角度对文献[14]中的算法展开了分析,分析结果表明:该算法的信息损失速度较小,输出序列相邻元素间的信息关联性较大,存在被有效攻击的可能性,且该算法的标准第二类比特位变换中采用1023-*e*作为移位位数,限制了指数*e*的范围,导致其并不能遍历[-1,1]区间内的所有浮点数,算法本身并不具有普适性.

通过理论分析可知,对于浮点数实现的混沌系统,若既要保证输出混沌序列均匀性,又要使得该混沌系统的平均信息损失速度接近$I_{\max}$是很难实现的. 本文基于定点数的实现方法对文献[14]中的算法进行了改进,首次将定点数计算过程中产生的被截断数据定义为暂态数据,并将此暂态数据应用于混沌输出状态序列均匀化的过程中. 理论与分析结果表明,该方法可以有效的将任意混沌系统的输出状态序列转换为[0,1]区间内的浮点数,重新生成的浮点数序列将服从均匀分布,可以实现文献[14]所述算法的全部功能. 此外,本文算法所得到混沌伪随机序列的相邻元素间仅存在不大于1比特的信息上的关联,可保证所生成序列具有较高的安全性.

参考文献

[1] Gottlieb Pirsic, Arne Winterhof. On the structure of digital explicit nonlinear and inversive pseudorandom number generators[J]. Journal of Complexity, 2010, 26(1): 43-50.

[2] A Peinado, A Fuster-Sabater. Generation of pseudorandom binary sequences by means of linear feedback shift registers (LFSRS) with dynamic feedback[J]. Mathematical and Computer Modelling, 2013, 57(11-12): 2596-2604.

[3] Kit-Ho Mak. More constructions of pseudorandom sequences of k symbols[J]. Finite Fields and Their Applications, 2014, 25(1): 222-233.

[4] A Kanso, N Smaoui. Logistic chaotic maps for binary numbers generations[J]. Chaos, Solitons and Fractals, 2009, 40(5): 2557-2568.

[5] Hanping Hu, LingFeng Liu, NaiDa Ding. Pseudorandom sequence generator based on the Chen chaotic system[J]. Computer Physics Communications, 2013, 184(3): 765-768.

[6] L Palacios-Luengas, G Delgado-Gutierrez, M Cruz-Irisson, J L Del-Rio-Correa, RVazquez-Medina. Digital noise produced by a non discretized tent chaotic map[J]. Microelectronic Engineering, 2013, 112(1): 264-268.

[7] Ping Li, Zhong Li, Wolfgang A Halang, Guanrong Chen. A multiple pseudorandom-bit generator based on a spatiotemporal chaotic map[J]. Physics Letters A, 2006, 349(6): 467-473.

[8] Liu Nian-sheng. Pesudo-randomness and complexity of binary sequences generated by the chaotic system[J]. Communications in Nonlinear Science and Numerical Simulation, 2011, 16(2): 761-768.

[9] M Francois, T Grosges, D Barchiesi, R Erra. Pseudo-random number generator based onmixing of three chaotic maps[J]. Communications in Nonlinear Science and Numerical Simulation,2014,19(4):887 - 895.

[10] 孙克辉,贺少波,何毅,尹林子. 混沌伪随机序列的谱熵复杂性分析[J]. 物理学报,2013,62(1):010501 - 1 - 010501 - 8.
Kehui Sun, Shaobo He, Yi He, Linzi Yin. Complexity analysis of chaotic pseudo-random sequences based on spectral entropy algorithm[J]. Acta Physica Sinica,2013, 62(1):010501 - 1 - 010501 - 8. (in Chinese)

[11] Xiaoni Du, Andrew Klapper, Zhixiong Chen. Linear complexity of pseudorandom sequence generated by Fermat quotients and their generalizations [J]. Information Processing Letters, 2012, 112 (6): 233 - 237.

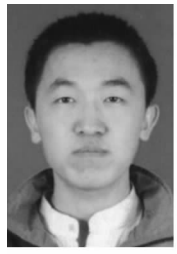
[12] Fatih Ozkaynak, Sirma Yavuz. Security problems for a pseudorandom sequence generator based on the Chen chaotic system[J]. Computer Physics Communications, 2013,184(9):2178 - 2181.

[13] K J Persohn, R J Povinelli. Analyzing logistic map pseudorandom number generators for periodicity induced by finite precision floating-point representation [J]. Chaos, Solitons and Fractals,2012,45(3):238 - 245.

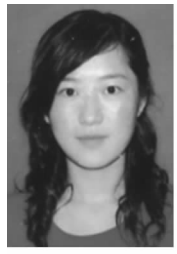
[14] 盛利元,肖燕予,等. 将混沌序列变换成均匀伪随机序列的普适算法[J]. 物理学报,2008,5(7):4007 - 4013.
Sheng Liyuan, Xiao Yanyu, et al. A universal algorithm for transforming chaoticsequences into uniform pseudo-random sequences[J]. Acta Physica Sinica,2008,5(7): 4007 - 4013. (in Chinese)

[15] 张占锋,盛利元,刘长水. 混沌伪随机序列均匀化普适算法的 FPGA 实现[J]. 计算机测量与控制,2009,17(12):2525 - 2554.
Zhang Zhanfeng, Sheng Liyun, liu Changshui. FPGA implementation of a universal algorithm for uniformization of chaotic pseudo-random sequences [J]. Computer Measurement & Control,2009,17(12):2525 - 2554. (in Chinese)

作者简介



李佩玥　男,1985 年 6 月出生于吉林磐石,博士,助理研究员,主要研究方向:混沌密码学、网络安全、精密驱动与控制技术.
E-mail:lipy@ sklao. ac. cn



石俊霞(通信作者)　女,1984 年 10 月出生于内蒙古乌兰察布,博士,助理研究员,主要研究方向:计算机理论、空间成像理论、图形图像处理.
E-mail:shijx@ ciomp. ac. cn