中国科学 : 信息科学　2011 年　第 41 卷　第 7 期 : 777–788
www.scichina.com　info.scichina.com
《中国科学》杂志社
SCIENCE CHINA PRESS
论　文

混沌密码系统弱密钥随机性分析

尹汝明, 袁坚*, 山秀明, 王希勤

清华大学电子工程系, 北京 100084
* 通信作者. E-mail: jyuan@tsinghua.edu.cn

收稿日期: 2010–01–14; 接受日期: 2010–07–01
国家自然科学基金 (批准号: 60674048) 资助项目

摘要　弱密钥问题是混沌密码系统设计中的关键问题, 已有研究主要从混沌序列退化角度进行分析. 然而, 本文指出保证混沌序列不退化的密钥参数仍可能构成混沌密码的弱密钥. 本文提出以混沌密码序列随机性作为评价标准, 应用严格的统计检验方法对混沌密码的弱密钥进行检测. 进一步, 对一类混沌密码系统进行了弱密钥研究, 检测出了该系统大量未被发现的弱密钥. 这确证了所提出方法的有效性. 另一方面, 虽然已有较多研究采用统计检验对混沌比特序列进行测试, 但将统计检验用于分析混沌密码弱密钥或弱序列的研究还很少见. 本文给出的统计检验弱序列分析, 对当前混沌密码统计检验研究是一个很好的补充.

关键词　混沌　密码学　统计检验　弱密钥　序列随机性

1　引言

混沌系统具有高度的初值和参数敏感性, 将系统的参数用作密码系统的密钥, 混沌系统很自然地被用于密码系统的设计 [1−3]. 在密码系统中, 保密性是通过对消息进行特定的变换得到的. 变换在一个保密参数的控制下进行, 这个参数称为密码系统的密钥. 为取得较高的安全性, 要求密码系统对密钥高度敏感, 这保证了无正确密钥情况下重构原变换是十分困难的.

目前, 在混沌密码设计中, 弱密钥是一个关键的问题 [3−8]. 密码系统中密钥参数被称为弱密钥, 是指在使用特定密钥参数时系统更容易被破译 [9]. 一般, 混沌密码的破译难度依赖于混沌系统产生序列的特性. 在某些参数下, 混沌序列存在短周期 [10], 这些参数就可能构成混沌密码的弱密钥, 这类弱密钥在文献 [3, 8] 中进行了分析. 另一方面, 混沌序列非短周期参数仍可能构成弱密钥. 在多数混沌密码系统中, 混沌序列并不直接用作系统的输出, 需要对混沌序列进行变换得到比特序列. 此时, 比特序列的随机性是衡量混沌密码破译难度的重要标准 [9]. 研究发现, 在混沌序列特性良好时, 比特序列的随机性仍可能很差 [11,12], 此时, 其对应参数仍可形成密码系统的弱密钥. 采用统计检验方法对混沌比特序列进行测试可以检测该类弱密钥. 注意到, 目前已有较多的研究利用统计检验对混沌比特序列进行分析. 但是, 仍存在以下问题. 一些研究采用的统计检验不太严格, 典型的是计算相关函数或者采用 FIPS 140-2 给出的检验 [13−16]. 事实上, 较低的相关函数值只是对比特序列随机性直观上的分析, 而 FIPS 140-2 给出的检验也因为不够严格而不再被推荐使用 [17]. 在研究中采用这些不严格的检验得出

引用格式: 尹汝明, 袁坚, 山秀明, 等. 混沌密码系统弱密钥随机性分析. 中国科学: 信息科学, 2011, 41: 777–788

的结论很可能是不准确的 [12]. 最近, 开始有研究采用更加严格的检验方法如 NIST 测试套件等 [11,18], 这在很大程度上改进了混沌序列随机性的检测. 但我们发现, 这些研究多是利用统计检验验证所提出的密码序列具有良好随机性, 而很少将统计检验用于分析混沌密码的弱密钥或弱序列 [11]. 显然, 对弱序列的分析更有助于改进混沌密码并凸显出严格统计检验方法的重要价值, 因此, 当前很需要在这一方面对已有研究加以补充.

在本文中, 我们提出一种新的分析混沌密码弱密钥的方法. 该方法以混沌密码输出比特序列的严格的随机性作为评价标准, 对混沌密码的弱密钥参数进行检测. 为证实所提出分析方法更加有效, 本文以一类密码系统为例进行了研究, 检测分析了该密码系统在不严格检验下大量未被发现的弱密钥, 细致分析了弱密钥的产生机理并提出了改进方法. 相比从混沌序列退化的角度进行的弱密钥分析, 本文提出的方法能够更为全面地检测出系统潜藏的弱密钥. 另一方面, 本文给出的利用统计检验对弱密钥的有效分析, 很好地充了当前缺少弱序列分析的研究状况, 这对当前混沌密码序列统计检验的研究是一个很好的推动.

本文在简要介绍已有研究的基础上, 首先提出了应用严格的随机性统计检验分析混沌密码弱密钥的方法. 紧接着, 为证实所提出分析方法更加有效, 本文以一类密码系统为例进行了研究, 检测分析了该密码系统大量未被发现的弱密钥, 并提出了避免弱密钥的改进方法. 在此工作基础上, 我们对当前混沌比特序列统计检验方面存在的主要研究进行了分析讨论, 进一步论述了本文工作的特点及意义. 最后, 我们对文章进行了总结.

2 分析弱密钥的新方法

2.1 已有研究

对混沌密码系统弱密钥问题, 已有研究注意到, 混沌系统在某些参数区域内存在短周期窗口, 此时混沌序列发生严重退化, 这些使得混沌序列退化的参数就可能构成混沌密码的弱密钥 [5]. 文献 [3] 对这类弱密钥做了系统的分析并提出了解决方法: 一是避免使用短周期窗口对应参数作为密钥; 二是尽量采用特性较好的混沌系统来设计密码算法. 要求参数在某个连续区间内变化时, 系统不出现周期窗口和不规则的分形, 帐篷映射能很好地满足这一要求.

2.2 弱密钥分析新方法

已有研究在一定程度上避免了混沌密码中较为明显的弱密钥. 然而, 只是从避免混沌序列退化的角度去解决混沌密码弱密钥问题是不够的. 在多数混沌密码系统中, 混沌映射产生的混沌序列并不直接用作密码系统的输出, 一般需要对混沌序列进行量化得到比特序列, 以便于加密数字信息. 为保证混沌密码的高安全性, 要求得到的比特序列具有很好的伪随机特性 [9]. 而混沌序列不发生退化未必能保证得到的比特序列具有良好的伪随机性. 研究发现, 在混沌序列不发生退化的情况下, 得到比特序列的伪随机性仍可能很差, 进而严重威胁密码系统的安全 [11,12].

在本文中, 我们提出一种新的分析混沌密码弱密钥的方法. 该方法以混沌密码输出比特序列的严格的随机性作为评价标准, 对混沌密码的弱密钥参数进行检测. 在特定密钥参数下, 我们对混沌密码所产生的比特序列进行随机性统计检验, 得到序列的伪随机特性. 如果序列伪随机特性较差, 则表明其对应的密钥参数是弱密钥. 在得到系统的弱密钥之后, 我们结合混沌系统特性从理论上对弱密钥的产生机理进行分析, 进而得出避免弱密钥的改进方法.

在所提出的分析方法中, 关键一步是对混沌密码系统产生的序列进行随机性统计检验. 随机性统计检验通过计算序列的某个统计量来判断序列接近随机的程度. 为取得较高的安全性, 密码系统产生的比特序列必须通过随机性统计检验 [9]. 在现代密码系统的设计和分析中, 随机性统计检验发挥着重要的作用. 在美国高级加密标准 (advanced encryption standard, AES) 的制定中, 以及在欧洲信息安全项目 (European network of excellence for cryptology, ECRYPT) 的密码算法征集评选中, 随机性统计检验都是作为评估和分析算法安全性的重要方法而被采用 [19−21].

为了对比特序列进行随机性检验, 需要选取适当的检验工具. 我们应该选择理论坚实且对序列伪随机特性要求严格的检验工具, 以避免在检测中漏掉系统的某些弱密钥. 我们在第二部分给出的分析实例将说明选择合适检验工具的重要性. 在本文中, 我们采用广泛认可的由美国标准与技术研究院 (national institute of standards and technology, NIST) 提出的随机性统计检验工具 [22]. 该检验工具要求系统连续产生 m 个长度为 n 的比特序列, 对每个序列进行检验得到检验值 $P\text{-}value$, 如果 $P\text{-}value > \alpha$ 则当前序列通过随机性检验. 其中, α 为假设检验的显著性水平 (推荐 $\alpha = 0.01$). 根据得到的 m 个检测结果 (m 个 $P\text{-}value$) 对序列的随机性进行最终的判定. 判定方法有两个: 一是计算 m 个序列中通过检验的序列比例 Proportion, 当 Proportion 大于某个门限值 T 则判定序列为随机, 其中门限值 T 跟显著性水平 α 和测试序列的数目 m 有关 (本文取 $\alpha = 0.01$, m=100, 此时 T=0.960150); 二是通过检验已得到的 m 个 $P\text{-}value$ 分布的均匀性得到一个最终的 $P\text{-}value_T$, 当 $P\text{-}value_T$ >0.0001 则判定序列为随机. 该检验工具软件包含 15 个有效检验项目. 在本文以下给出的各次检验中, 我们都采用这 15 个项目对系统密钥流进行了统计检验. 但是, 为了表达的简洁性, 我们只选取 4 个典型的检验项目给出检验结果. 这 4 个检验项目分别是: 频率检验 (frequency test)、块内频率检验 (frequency test within a block)、游程检验 (runs test) 和块内最长游程检验 (test for the longest run of ones in a block). 对本文的测试来说, 这 4 个检验项目具有代表性: 各次检验的结果都表明, 一方面, 如果系统密钥流不能通过这几个检验, 则该密钥流也不能通过多数其他检验项目; 另一方面, 如果系统密钥流能很好的通过这几个检验, 则该密钥流也能够通过其他检验项目.

3 一个混沌密码系统的弱密钥分析

在这一部分, 我们应用提出的弱密钥分析方法对一个具体的混沌密码系统进行弱密钥分析. 一般的, 在弱密钥参数下, 密码系统产生的密钥流很可能因为具有某种规律而出现明显偏离随机的现象. 统计检验正是通过检测这种偏离随机的现象, 进而得出系统可能具有的弱密钥. 在本节给出的密码系统中, 某些参数因为导致密钥流出现周期内序列强相关或者出现较短周期, 因而形成弱密钥. 我们通过随机性统计检验发现了这些弱密钥, 并根据其产生的机制提出了避免弱密钥的改进措施. 对这个具体混沌密码系统的分析证实了文章所提出的弱密钥分析方法的有效性.

3.1 一个混沌伪随机比特发生器

文献 [13] 设计了一个混沌伪随机比特发生器, 用其可以构成混沌流密码. 该发生器包括 3 个组成部分: 离散化分段线性映射、模加运算和量化运算. 这里简要介绍一下 3 个组成部分, 算法细节见文献 [13]. 离散化分段线性映射是对原连续映射进行适当离散化得到的. 原连续映射如公式 (1) 所示, 其

中连续映射参数 $p \in (0, 0.5)$, 初值 $x(0) \in [0, 1]$.

$$x(k+1) = f(x(k), p) = \begin{cases} x(k)/p, & 0 \leqslant x(k) < p; \\ (x(k) - p)/(0.5 - p), & p \leqslant x(k) < 0.5; \\ (1 - p - x(k))/(0.5 - p), & 0.5 \leqslant x(k) < 1 - p; \\ (1 - x(k))/p, & 1 - p \leqslant x(k) < 1. \end{cases} \tag{1}$$

$$X(k+1) = F(X(k), P) = \begin{cases} \left\lfloor \dfrac{2^{n_0} \cdot X(k)}{P} \right\rfloor, & 0 \leqslant X(k) < P; \\ \left\lfloor \dfrac{2^{n_0} \cdot (X(k) - P)}{2^{n_0-1} - P} \right\rfloor, & P \leqslant X(k) < 2^{n_0-1}; \\ \left\lfloor \dfrac{2^{n_0} \cdot (2^{n_0} - X(k) - P)}{2^{n_0-1} - P} \right\rfloor, & 2^{n_0-1} \leqslant x(k) < 2^{n_0} - P; \\ \left\lfloor \dfrac{2^{n_0} \cdot (2^{n_0} - X(k))}{P} \right\rfloor, & 2^{n_0} - P \leqslant x(k) < 2^{n_0}. \end{cases} \tag{2}$$

离散后的分段线性映射如公式 (2) 所示. 其中, $\lfloor x \rfloor$ 表示对 x 做向下取整运算, 整数 $X(k) \in [0, 2^{n_0} - 1]$ 为系统 k 时刻的离散状态值, 整数 $P \in (0, 2^{n_0-1})$ 是系统的离散控制参数, $n_0 = 32$ 表示计算字长. 离散映射参数 $X(k)$、P 和连续映射参数 $x(k)$、p 的转换关系如公式 (3) 所示.

$$x(k) = \frac{X(k)}{2^{n_0}}, \quad p = \frac{P}{2^{n_0-1}}. \tag{3}$$

混沌伪随机比特发生器在离散参数 P_1, P_2 的控制下, 以 $X_1(0)$, $X_2(0)$ 为初始条件分别迭代 2 个离散混沌映射, 不断得到状态值 $X_1(k)$, $X_2(k)$, 利用公式 (4) 把 2 个状态值进行模加得到混沌伪随机数 $Y(k)$.

$$Y(k) = (X_1(k) + X_2(k)) \bmod 2^{n_0}. \tag{4}$$

最后, 量化整数 $Y(k)$ 得到伪随机比特序列 $S(k)$, 具体量化方法见公式 (5), 即舍弃 $Y(k)n_0$ 个比特中最前 l 和最后的 l 比特, 只保留中间 $n_0 - 2l$ 比特作为输出序列, 其中 $l = 4$.

$$S(k) = b_l b_{l+1} \cdots b_{n_0-l-1}, \text{当 } Y(k) = b_0 b_1 \cdots b_{n_0-1}. \tag{5}$$

2 个离散混沌映射的初始值和参数 $[P_1 \quad P_2 \quad X_1(0) \quad X_2(0)]$ 构成混沌伪随机比特发生器的密钥参数.

文献 [13] 提出的发生器包含离散混沌映射, 这有时会给弱密钥理论分析带来困难. 根据公式 (3) 给出的变换关系, 可以用连续分段线性映射代替发生器中离散分段线性映射, 这样就得到发生器连续参数模型, 如图 1 所示. 其中, $f(x, p)$ 为公式 (1) 给出的连续分段线性映射, 量化方法由公式 (6) 给出.

$$S(k) = b_l b_{l+1} \cdots b_{n_0-l-1}, \quad \text{当} \quad y(k) = 0.b_0 b_1 \cdots b_{n_0-1}. \tag{6}$$

在以下叙述中, 为避免混淆, 我们称 $x(0)$ 和 p 为发生器的连续密钥参数, 而称 $X(0)$ 和 P 为发生器的离散密钥参数.

3.2 发生器弱密钥分析

我们对上面的伪随机比特发生器进行了弱密钥分析, 得出了发生器存在的两类弱密钥, 对弱密钥产生的机理进行了理论分析, 并最终提出了避免弱密钥的改进措施.

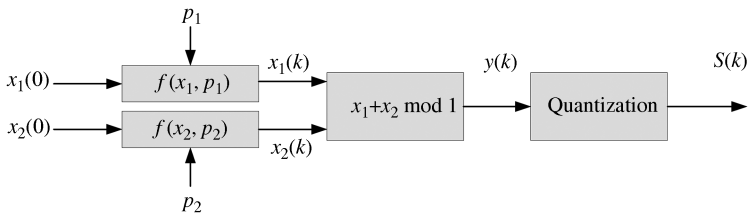

图 1 发生器连续参数模型

Figure 1 The equivalent generator with the continuous chaotic maps

表 1 参数为 [33 34 1000 1000] 的发生器序列统计检验结果

Table 1 The test results for the generator with the parameters [33 34 1000 1000]

Test name	Proportion	$P\text{-}value_T$
Frequency	0.9900	0.275709
Block-Frequency	0.9300*	0.000818
Runs	0.6100*	0.000000*
Longest-Run	0.0000*	0.000000*

"$*$" denotes that the sequence has failed the corresponding test.

3.2.1 周期内序列强相关形成的弱密钥

在某些密钥参数下, 设计者对发生器产生的序列进行了检验, 结果表明序列能够通过检验 [13]. 但是, 利用提出的方法对相同参数下的序列进行检验, 却发现该密钥参数是发生器的一个弱密钥. 我们对这个弱密钥的产生机理进行了的理论分析和仿真验证, 结果表明, 该随机比特发生器有一大类产生机制相似的弱密钥. 在分析这些弱密钥产生机制的基础上, 我们提出了避免这类弱密钥的改进方法.

表 1 给出了某一个密钥参数下, 应用 NIST 检验工具对发生器产生的比特序列进行统计检验的结果, 表中标有 * 的项表示序列不能通过该项统计检验. 表中的发生器参数取的是发生器设计者进行统计检验的原参数 [13], 即 $[P_1 \quad P_2 \quad X_1(0) \quad X_2(0)] = [33 \quad 34 \quad 1000 \quad 1000]$, 对应的连续密钥参数 $p_1 \approx p_2 \approx 7.7 \times 10^{-9}$. 可以看到, 在表 1 对应的密钥参数下, 序列除去能通过最为简单的频率检验外, 其他三项检验都不能通过, 而且各统计量的检验值和理论要求值相差很远. 显然, 表 1 对应的密钥参数是发生器的一个弱密钥.

我们针对图 1 给出的发生器连续参数模型, 从理论上分析这个弱密钥产生的原因. 表 1 中, 发生器的连续密钥参数为 $p_1 \approx p_2 \approx \frac{33}{2^{32}} \approx 7.7 \times 10^{-9}$, 取值接近于 0. 该参数下分段线性映射接近标准帐篷映射. 标准帐篷映射迭代公式由公式 (7) 给出.

$$x(k+1) = f_{\text{tent}}(x(k)) = \begin{cases} 2 \cdot x(k), & 0 \leqslant x(k) < 0.5; \\ 2 \cdot (1 - x(k)), & 0.5 \leqslant x(k) < 1. \end{cases} \tag{7}$$

由于该弱密钥参数下分段线性映射接近于标准帐篷映射, 为了理论分析的方便, 我们就以标准帐篷映射代替发生器连续模型中的连续分段线性映射, 来对弱密钥的产生机理进行分析.

在发生器连续模型中, 设 k 时刻两个帐篷映射的状态分别为 $x_1(k)$, $x_2(k)$, 经过发生器模加运算后得到 $y(k) = x_1(k) + x_2(k) \mod 1$. 考虑 $k+1$ 时刻两个帐篷映射的状态 $x_1(k+1)$, $x_2(k+1)$, 由公式

(7), $k+1$ 时刻经过发生器模加运算后得到

$$y(k+1)=\begin{cases}2\cdot(x_1(k)+x_2(k)) \bmod 1, & 0\leqslant x_1(k),x_2(k)<0.5;\\ 2\cdot(1+x_1(k)-x_2(k)) \bmod 1, & 0\leqslant x_1(k)<0.5, 0.5\leqslant x_2(k)<1;\\ 2\cdot(1+x_2(k)-x_1(k)) \bmod 1, & 0.5\leqslant x_1(k)<1, 0\leqslant x_2(k)<0.05;\\ 2\cdot(1-(x_1k)+x_2(k)) \bmod 1, & 0.5\leqslant x_1(k),x_2(k)<1.\end{cases} \tag{8}$$

进一步有

$$y(k+1)=\begin{cases}2\cdot y(k) \bmod 1, & 0\leqslant x_1(k),x_2(k)<0.5;\\ 2\cdot(2-y(k)) \bmod 1, & 0.5\leqslant x_1(k),x_2(k)<1.\end{cases} \tag{9}$$

假设 $y(k)$ 的二进制表示为 $y(k)=0.b_0^kb_1^k\cdots b_{n-1}^k$, $y(k+1)$ 的二进制表示为 $y(k+1)=0.b_0^{k+1}b_1^{k+1}\cdots b_{n-1}^{k+1}$. 当 $x_1(k)$ 和 $x_2(k)$ 同属于 $[0,0.5)$ 或 $[0.5,1)$ 区间时, 根据公式 (9), 则

$$\begin{aligned}&b_i^{k+1}=b_{i+1}^k, 0\leqslant i\leqslant n-2,\\ &b_{n-1}^{k+1}=0,\end{aligned}\quad \text{当}\quad 0\leqslant x_1(k),x_2(k)<0.5; \tag{10}$$

$$\begin{aligned}&b_i^{k+1}=\overline{b_{i+1}^k}, 0\leqslant i<I,\\ &b_i^{k+1}=b_{i+1}^k, I\leqslant i\leqslant n-2,\\ &b_{n-1}^{k+1}=0,\end{aligned}\quad \text{当}\quad 0.5\leqslant x_1(k),x_2(k)<1. \tag{11}$$

其中, I 为 $y(k)=0.b_0^kb_1^k\cdots b_{n-1}^k$ 中最后一个取值为 1 的比特的位置, 即 $b_I^k=1, b_i^k=0, I<i\leqslant n-1$.

对发生器连续模型, 在得到相邻时刻 $y(k)$ 和 $y(k+1)$ 的关系后, 要分析最终比特序列随机性的退化, 还需要考察发生器的量化方法. 发生器的量化方法由公式 (6) 给出, 图 2 直观地给出了发生器离散混沌序列得到比特序列的方法. 图中, $y(k), y(k+1), \ldots, y(k+t)$ 表示发生器在模加运算之后量化之前得到的连续 $t+1$ 个结果, $S(k), S(k+1), \ldots, S(k+t)$ 表示从每次量化得到的比特序列. 发生器把这 $t+1$ 个量化得到的比特序列顺序组合到一起就得到了最终的比特序列 S.

在图 2 所示的量化方法下, 根据公式 (10), (11) 给出的分析, 在最终的比特序列 S 中相邻时间抽取出的两个比特串 $S(k)$ 和 $S(k+1)$ 具有极强的相关性: 两比特串中的大部分比特或者是错位相等关系, 或者是错位互补关系. 此时, 最终得到的比特序列 S 显然是随机性很差的. 知道序列 S 的某一部分后, 能够很好地预测接下来的一段固定长度的序列. 这在理论上解释了在发生器连续密钥参数 p_1 和 p_2 取值接近于 0 时, 整个发生器的输出序列发生严重退化的原因.

我们实际计算发生器的输出序列对理论分析结果进行了验证. 在计算中, 取发生器的离散密钥参数为 $[P_1 \quad P_2 \quad X_1(0) \quad X_2(0)] = [33 \quad 34 \quad 1000 \quad 1000]$, 对应的连续密钥参数为 $p_1\approx p_2\approx \frac{33}{2^{32}}\approx 7.7\times 10^{-9}$, 取值接近于 0, 这正是理论分析所针对的情况. 我们发现, 较多相邻段比特序列具有的错位相等或错位互补的关系. 这个实验结果表明, 实际发生器比特序列的关联关系是和理论分析结果相一致的, 发生器所产生的序列相邻部分具有极强的相关性, 这是该参数下序列不能通过随机性统计检验的原因.

以上, 我们从理论上分析了发生器连续密钥参数 p_1 和 p_2 接近于 0 时, 所产生比特序列发生退化的原因. 更多的随机性统计检验实验表明, 当发生器的连续密钥参数 p_1 和 p_2 同时接近 $\frac{1}{2}$, $\frac{1}{4}$ 或 $\frac{1}{8}$ 时,

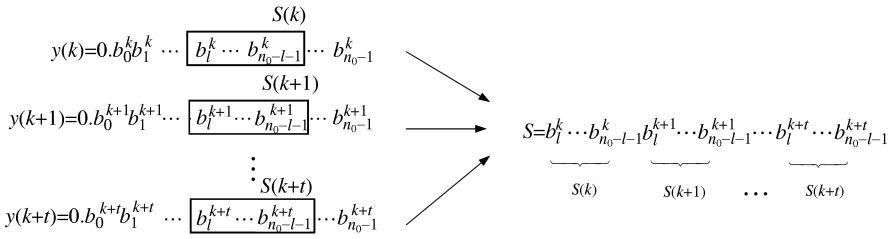

图 2 发生器量化混沌序列得到比特序列的方法

Figure 2 The bit extraction scheme adopted in the generator

发生器所产序列也会发生类似的退化, 此时序列也都不能很好地通过随机性统计检验. 另外, 当 p_1 和 p_2 接近于 $\frac{1}{2^m}, (m > 3)$ 时, 序列也都会发生不同程度的退化. 这样, 这一类密钥参数构成了发生器的一大类弱密钥.

为了避免发生器出现以上弱密钥, 可以不使用这类参数. 但这样, 一方面会使系统的密钥空间变小, 降低系统的安全性; 另外使用中检测这些弱密钥也会使给算法应用带来不便. 另一个解决办法是改进发生器中的量化运算, 使得混沌系统的这类参数不再构成弱密钥. 通过前面的分析, 我们看到, 原发生器一次从混沌序列固定位置连续抽取多个比特, 而相邻两次抽取的比特串具有较强的相关性, 这是最终比特序列随机性较差的原因. 为避免量化算法的这一缺陷, 我们可以减少一次抽取比特的数目. 设一次连续抽取比特串的长度为 L, 在 L 取不同值的条件下, 我们对离散分段线性映射产生的比特序列进行了随机性统计检验. 结果表明, 当 $L > 1$ 时, 在某些参数下 (如参数 P 取值很小时), 离散映射产生的序列随机性仍然较差, 不能通过某些项检验. 因此, 我们将原来的一次抽取多个比特改成只抽取一个比特. 在离散分段映射退化参数 $X(0) = 1000, P = 33$ 下, 对于 k 时刻的状态值 $X(k) = b_0^k b_1^k \cdots b_{n_0-1}^k$, 我们抽取 $X(k)$ 中固定位置的一个比特 $b_r^k, r \in [0, n_0 - 1]$ 组成比特序列并对序列进行了随机性检验. 对参数 $r = 16$, 改进的量化算法产生的序列具有较好的随机性, 原算法的弱密钥参数不再构成改进系统的弱密钥.

3.2.2 序列出现短周期或固定点形成的弱密钥

这一部分, 我们将结合序列随机性统计检验研究发生器序列的周期情况, 分析使发生器产生短周期序列的弱密钥参数, 并提出避免该类弱密钥的改进方法. 离散分段线性映射很可能形成短周期甚至是固定点, 这严重影响混沌密码的安全性 [9]. 为解决这一问题, 原发生器把两个离散映射的结果进行模加以期望增加序列的周期. 这在一定程度上改善了序列的周期性能. 但是, 仍存在较为严重的缺陷. 参与模加运算的两个离散映射都不具有稳定的周期长度, 而且对于某些的参数和初值, 离散映射的轨道极可能出现短周期或固定点, 这会使得最终的序列极容易形成极短的周期. 此时, 造成短周期的参数和初值就构成了发生器的弱密钥.

(1) 序列出现固定点形成弱密钥.

先讨论发生器离散映射的固定点弱密钥问题. 在公式 (2) 所示的离散分段线性映射中, 令 $X(k+1) = X(k)$, 并且先不考虑公式中的取整运算 $\lfloor x \rfloor$, 就可以得到离散分段线性映射可能存在 3 个不为 0 的固定点 $X_{fix1} = \frac{P \cdot 2^{n0}}{0.5 \times 2^{n0} + P}$, $X_{fix2} = \frac{(P - 2^{n0}) \times 2^{n0}}{P - 1.5 \times 2^{n0}}$ 和 $X_{fix3} = \frac{2^{2 \times n0}}{2^{n0} + P}$. 离散分段线性映射中与这 3 个实数点相接近的离散整数值, 很可能构成离散映射的固定点. 例如, 当离散映射参数 P 取值为 34 时, 可以得到两个固定点: $X_{fix1} = 68$, $X_{fix2} = 2\,863\,311\,523$. 当 P 取值为 33 时, 可以得到一个固定点: $X_{fix1} = 66$. 因此, 在离散密钥参数为 $[P_1 \quad P_2 \quad X_1(0) \quad X_2(0)] = [33 \quad 34 \quad 66 \quad 68]$ 或者

$[33 \quad 34 \quad 66 \quad 2863311523]$ 时, 发生器只能重复产生一个固定的比特串, 这样两个参数都是发生器的固定点弱密钥. 类似地, 对于发生器的任意离散参数 P_1 和 P_2, 都可能存在固定点弱密钥. 这类弱密钥的可能值可以按公式预先计算得到. 另外, 还有一类特殊的固定点弱密钥. 由于计算机只能表示有限精度, 因此, 在连续分段线性映射的参数 $p = \frac{1}{4}$ 时, 不论初值如何, 最终的迭代结果都会收敛到 0[23]. 这意味着当发生器离散参数 $[P_1 \quad P_2] = [2^{30} \quad 2^{30}]$ 时, 无论 $X_1(0)$ 和 $X_2(0)$ 取值如何, 整个发生器的输出都会收敛到 0. 以上固定点弱密钥都是可以预先计算得到的, 因此, 可以对每个密钥参数在使用前进行检查, 这样就可以避免这类固定点弱密钥.

(2) 序列出现短周期形成弱密钥.

除固定点外, 离散分段线性映射还很可能形成短周期 [24]. 短周期的出现使得分析者得到序列一个较短的周期后就能够完全预测出以后的序列, 严重影响密码的安全性. 通过仿真计算, 我们估计了公式 (2) 所示离散分段线性映射典型的轨道周期长度, 如表 2 所示. 表 2 中, 我们随机选取离散分段映射的参数和初值, 计算出轨道周期小于固定长度的参数的大概比例. 可以看到, 大约有 40%的密钥参数会使得离散映射的轨道周期小于 10^5, 发生器采用了两个离散映射结果模加的办法来扩展周期, 这样大约会有 16%的密钥参数使得发生器最终得到的序列周期小 10^{10}. 这样一个周期长度是不能满足当前多数加密应用需求的 [9].

短周期是序列一种明显的偏离随机性的特征. 这种严重偏离随机性的特征在适当的随机性统计检验下也能够被准确发现, 即也可以借助随机性统计检验来检测序列是否出现周期. 在某个短周期参数的条件下, 表 3 给出了发生器产生序列的统计检验结果. 其中, 参数为

$$[P_1 \quad P_2 \quad X_1(0) \quad X_2(0)] = [839032169 \quad 1049817325 \quad 2039465734 \quad 3987648544],$$

此时两个离散映射的轨道周期都不到 100. 另外, 此时发生器采用的是改进后的量化方法, 这是为了排除周期内序列强相关给检验结果带来的影响. 表 3 给出的统计检验结果表明, 序列的随机性极差, 4 项统计检验都无法通过, 从而检测出了序列具有短周期.

可以看到, 原发生器的短周期缺陷是由于参与模加运算的两个离散映射都没有固定周期导致的. 为了改善序列的周期特性, 可以利用有固定周期长度的序列去定期扰动混沌序列, 这种方法能够保证最终得到序列的最小周期长度 [25]. 据此, 我们对原发生器作了进一步的改进: 用线性同余发生器对离散混沌映射进行扰动, 改进后的发生器如图 3 所示. 注意, 图中发生器量化运算也采用了改进后的方法, 即每次抽取一个比特 (Y_k 二进制表示中第 16 个比特) 组成输出序列. 其中, 线性同余发生器如公式 (12) 所示, $Z(0) \in [1, M-1]$ 为线性同余发生器的初始状态, $A = 16807, M = 2^{31} - 1$. 线性同余发生器产生的随机数具有很好的特性 [26], 并且具有最大的周期长度 $M - 1$.

$$Z(m+1) = A \cdot Z(m) \bmod M, \quad m = 0, 1, 2, \ldots. \tag{12}$$

在图 3 中, 线性同余发生器在初始时刻对离散混沌映射进行第一次扰动, 以后迭代每隔 Δ 次迭代扰动一次, 扰动方法如公式 (13) 所示.

$$\begin{aligned} &Y(k) = X(k), && k \neq m\Delta, m = 0, 1, 2, \ldots; \\ &Y(k) = (X(k) + Z(m)) \bmod 2^{n_0}, && k = m\Delta, m = 0, 1, 2, \ldots. \end{aligned} \tag{13}$$

在这种扰动方法下, 可以证明发生器最终的输出序列周期大于 $(M-1) \cdot \Delta$[25]. 例如, 取 $\Delta = 500$, 则序列周期大于 $M \cdot \Delta = (2^{31} - 2) \cdot 500 \approx 10^{12}$, 这较原发生器的周期长度有了较大的改善. 在参数 $P = 33$,

表 2 离散分段线性映射的典型周期长度

Table 2 Typical values of the period length of the discretized chaotic map

Period length	Less than 10^3	Less than 10^4	Less than 10^5
Percentage of keys	0.006	0.5	40

表 3 某短周期参数下发生器序列的统计检验结果

Table 3 The test results for a short period sequence produced by the generator

Test name	Proportion	$P\text{-}value_T$
Frequency	1.0000	0.000000*
Block-Frequency	1.0000	0.000000*
Runs	0.0000*	0.000000*
Longest-Run	0.0000*	0.000000*

"$*$" denotes that the sequence has failed the corresponding test.

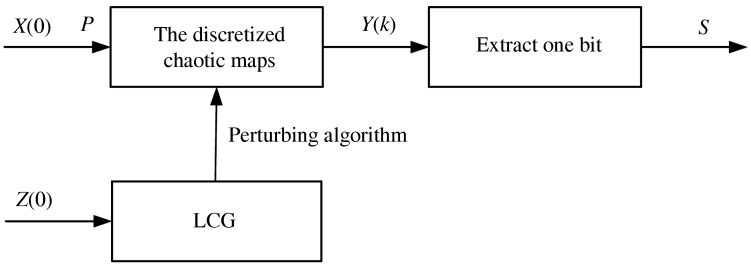

图 3 改进后的混沌随机比特发生器

Figure 3 The improved chaotic sequence generator

$X_0 = 1000$, $\Delta = 500$, $Z_0 = 3$ 条件下, 我们对改进后的序列进行了统计检验. 结果表明, 改进后的随机比特发生器在具有较长周期的同时, 也具有较好的伪随机特性.

4 对混沌比特序列统计检验相关研究的分析讨论

本文提出了应用严格的随机性统计检验分析混沌密码弱密钥的方法. 我们使用该方法对一类混沌密码系统进行了弱密钥研究. 事实上, 已有一些工作对混沌比特序列的统计特性进行了研究. 本节对这些相关研究进行分析比较, 进一步论述本研究的特点及意义. 以下, 先讨论理论方面的已有研究, 然后再讨论统计检验工具在混沌比特序列测试上的应用.

对混沌系统实数状态值进行适当量化可以得到比特序列. 目前, 关于这类混沌比特序列统计特性的理论研究比较有效 [27−30]. 然而, 这些理论成果对混沌密码设计的指导是相对有限的. 这些理论研究的基本思路是利用混沌系统的遍历性和不变密度对比特序列的相关性进行理论分析. 针对具有某种对称性的遍历的混沌映射, 文献 [27] 最早给出了产生独立同分布二进制序列的充分条件. 紧接着, 在综述文章 [28] 中, 作者总结了这一工作, 系统的论述了利用混沌系统产生马尔可夫序列和独立同分布序列的方法, 并且细致研究了这些序列在流密码和 CDMA 通信系统中的应用. 近来, 文献 [29] 进一步在理论上研究了该类比特序列的游程统计特性. 注意到, 遍历性和不变密度是研究该类比特序列相关性的理论基础, 而混沌系统实现中带来的参数误差可能影响系统的这些特性, 最近的文献 [30] 对这

个问题进行了探索. 虽然以上研究在理论上给出了产生良好特性 (比如独立同分布) 比特序列的方法, 但是, 正如理论研究者们自己所说, 用这些理论指导混沌密码的设计还存在较大的问题 [28]. 原因在于: 首先, 这些理论研究关注的主要是比特序列的相关特性, 而密码学上对安全序列有更严格的要求, 比如长周期、高的线性复杂度等, 低相关性只是最为基本的要求. 其次, 这些理论研究都针对理想的实数混沌系统, 但多数实际情况下混沌系统状态需要采用定点或者浮点表示. 此时, 理论上具有良好特性的序列很可能发生严重的退化 [23,24]. 最典型的例子, Bernoulli 移位映射在理论上将产生 0 和 1 等概率出现的独立同分布序列, 但由于计算机有限精度的影响, 输出序列在很短时间后就会变为全 0. 另外, 还有更多因素限制了该类比特序列在混沌密码中的应用, 比如基于实数混沌系统的混沌密码加解密效率较低, 基于电路的模拟混沌系统不利于充分利用混沌的参数敏感性等等.

除理论分析外, 利用统计检验工具可以有效分析序列的随机特性. 常用的统计检验包括 FIPS 140-2 统计检验、NIST 检验套件和 DIEHARD 系列检验等等. 在混沌真随机序列的研究中, 这些检验工具已经得到很广泛的使用 [31−33]. 近来, 这些工具也开始被用于测试混沌密码产生的伪随机序列 [11,13,14,16,18]. 但是, 从目前的使用情况来看, 仍存在一些问题需要进一步改善. 一方面, 一些研究采用的检验不太严格, 典型的是计算相关函数或者采用 FIPS 140-2 统计检验 [13−16]. 事实上, 较低的相关函数值只是对比特序列随机性直观上的分析, 而 FIPS 140-2 给出的检验也因为不够严格而不再被推荐使用 [17]. 在研究中采用这些不严格的检验得出的结论很可能是不准确的. 这方面, 除去本文给出的例子外, 另一个例子见文献 [12]. 文献 [12] 利用 NIST 检验套件对文献 [15] 给出的混沌比特序列进行了严格测试, 发现序列随机性极差. 然而这一序列却通过了文献 [15] 给出的简单的信息熵统计分析. 另一方面, 虽然有少数研究采用了严格的检验工具如 NIST 检验套件, 但在使用方法上仍存在一些不足之处. 首先, 一些研究随机选择系统参数或者选择系统最好的参数进行测试, 以测试得到的良好结果作为密码安全的确证 [11,18]. 这样的测试比较容易通过, 但却不满足密码设计的基本要求, 因为安全的密码需要在所有密钥参数下产生的密钥流都具有良好的随机性. 而通过这种测试的混沌密码可能仍存在弱密钥, 即某些参数下系统产生序列随机性仍可能较差. 其次, 多数研究是利用统计检验验证所提出的密码序列具有良好随机性, 而很少有研究将统计检验用于分析混沌密码的弱密钥或弱序列. 据我们所知, 只有文献 [11, 12] 利用 NIST 检验套件对混沌密码的弱序列进行了识别. 但两篇文章的重点都不在统计检验的使用上, 因而都没有就这个问题进行深入的探讨. 统计检验应该用来分析现有密码设计存在的问题进而提出改进, 但对弱序列有效分析的缺乏影响了混沌密码研究者对严格统计检验的重视和合理使用, 这对混沌密码的设计是不利的. 注意到, 在传统密码的研究中, 利用严格的统计检验分析算法存在的缺陷并改进算法这一过程已经做得很好 [19−21], 而混沌密码的研究在这一方面则还存在差距. 因此, 当前很需要在这一方面对混沌密码的已有研究加以补充.

本文强调了严格的随机性统计检验在混沌密码设计中的重要作用. 特别的, 本文以一类混沌密码系统为例进行了研究, 应用严格的检验工具分析了该系统大量的弱序列并深入研究了其产生机理. 这些弱序列在不严格的统计检验方法下是不能被发现的 [13]. 进一步, 即使对严格的统计检验, 如果检验方法不合理, 例如随机选择一个系统参数进行测试, 这类弱序列也是很难被发现的. 这样, 本文给出的利用统计检验对弱密钥的有效分析, 很好的补充了当前弱序列分析缺乏的研究状况. 这对当前混沌密码序列统计检验的研究是一个很好的推动.

5 总结

本文提出了应用严格的随机性统计检验分析混沌密码弱密钥的方法. 文章以一类密码系统为例

进行了研究, 检测分析了该密码系统大量未被发现的弱密钥. 这证实了我们提出的方法的有效性. 另一方面, 本文给出的利用统计检验对弱密钥的有效分析, 很好的补充了当前缺少弱序列分析的研究状况. 但是, 需要认识到, 利用统计检验对混沌密码弱密钥进行分析存在局限性. 由于混沌密码密钥空间巨大, 对所有密钥参数进行检测是不可能的. 因此, 迫切需要在理论上研究如何保证混沌密码序列的良好特性, 如良好的统计特性、大的周期长度和高度的非线性等等. 另外, 我们认为应该针对状态离散化的混沌系统开展理论研究, 这样取得的成果才能更有效地指导实际混沌密码的设计. 这也是我们未来工作努力的方向.

参考文献

1 Kocarev L. Chaos-based cryptography: a brief overview. IEEE Circ Syst Mag, 2001, 1: 6–21

2 Zhang Y W, Wang Y M, Shen X B. Chaos-based image encryption algorithm using alternate structure. Sci China Ser F-Inf Sci, 2007, 50: 334–341

3 Alvarez G, Li S. Some basic cryptographic requirements for chaos-based cryptosystems. Int J Bifurcat Chaos, 2006, 16: 2129–2151

4 Biham E. Cryptanalysis of the chaotic-map cryptosystem suggested at EuroCrypt'91. In: Advances in Cryptology - EuroCrypt'91. Berlin: Springer, 1991. 532–534

5 Alvarez G, Montoya F, Romera M, et al. Cryptanalysis of a discrete chaotic cryptosystem using external key. Phys Lett A, 2003, 319: 334–339

6 Skrobek A. Cryptanalysis of chaotic stream cipher. Phys Lett A, 2007, 363: 84–90

7 Li S, Alvarez G, Chen G, et al. Breaking a chaos-noise-based secure communication scheme. Chaos, 2005, 15: 013703

8 Li C, Li S, Alvarez G, et al. Cryptanalysis of a chaotic block cipher with external key and its improved version. Chaos Soliton Fract, 2008, 37: 299–307

9 Schneier B. Applied Cryptography: Protocols, Algorithms, and Source Code in C. 2nd ed. Brisbane: John Wiley and Sons, 1996. 280–282

10 Robinson R C. An introduction to Dynamical Systems: Continuous and Discrete. New Jersey: Pearson Prentice Hall, 2004. 303–366

11 Tang K W, Tang K S, Man K F. A chaos-based pseudo-random number generator and its application in voice communications. Int J Bifurcat Chaos, 2007, 17: 923–933

12 Li C, Li S, Alvarez G, et al. Cryptanalysis of two chaotic encryption schemes based on circular bit shift and XOR operations. Phys Lett A, 2007, 369: 23–30

13 Lian S, Sun J, Wang J, et al. A chaotic stream cipher and the usage in video protection. Chaos Soliton Fract, 2007, 34: 851–859

14 Li P, Li Z, Halang W A, et al. Analysis of a multiple-output pseudo-random-bit generator based on a spatiotemporal chaotic system. Int J Bifurcat Chaos, 2006, 16: 2949–2963

15 Xiang T, Liao X, Tang G, et al. A novel block cryptosystem based on iterating a chaotic map. Phys Lett A, 2006, 349: 109–115

16 Li P, Li Z, Halang W A, et al. A multiple pseudorandom-bit generator based on a spatiotemporal chaotic map. Phys Lett A, 2006, 349: 467–473

17 National Institute of Standards and Technology (NIST). Security Requirements for Cryptographic Modules. Federal Information Processing Standards Publication 140-2. 2001

18 Patidar V, Sud K K, Pareek N K. A pseudo random bit generator based on chaotic logistic map and its statistical testing. Informatica, 2009, 33: 441–452

19 Soto J. Randomness Testing of the AES Candidate Algorithms. NIST Interagency Reports 6390. 1999

20 Soto J, Bassham L. Randomness Testing of the Advanced Encryption Standard Finalist Candidates. NIST Interagency Reports 6483. 2000

21 Turan M S, Doganaksoy A, Calik C. Detailed Statistical Analysis of Synchronous Stream Ciphers. eSTREAM report 2006/043. 2006
22 Rukhin A, Soto J, Nechvatal J, et al. A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications. NIST Special Publication 800-22. 2001
23 Li S. When chaos meets computers. arXiv: nlin.CD/0405038. 2004
24 Li S, Chen G, Mou X. On the dynamical degradation of digital piecewise linear chaotic maps. Int J Bifurcat Chaos, 2005, 15: 3119–3151
25 Sang T, Wang R, Yan Y. Perturbance-based algorithm to expand cycle length of chaotic key stream. Electron Lett, 1998, 34: 873–877
26 Park S K, Miller K W. Random number generators: good ones are hard to find. Commun ACM, 1988, 31: 1192–1201
27 Kohda T, Tsuneda A. Statistics of chaotic binary sequences. IEEE Trans Inform Theory, 1997, 43: 104–112
28 Kohda T. Information sources using chaotic dynamics. Proc IEEE, 2002, 90: 641–661
29 Tsuneda A. Design of binary sequences with tunable exponential autocorrelations and run statistics based on one-dimensional chaotic maps. IEEE Trans Circ Syst-I, 2005, 52: 454–462
30 Addabbo T, Fort A, Papini D, et al. Invariant measures of tunable chaotic sources: robustness analysis and efficient estimation. IEEE Trans Circ Syst-I, 2009, 56: 806–819
31 Ergun S, Ozoguz S. Truly random number generators based on a non-autonomous chaotic oscillator. Int J Electron Commun, 2007, 61: 235–242
32 Tomassini M, Sipper M, Perrenoud M. On the generation of high-quality random numbers by two-dimensional cellular automata. IEEE Trans Comput, 2000, 49: 1146–1151
33 Addabbo T, Alioto M, Fort A, et al. A feedback strategy to improve the entropy of a chaos-based random bit generator. IEEE Trans Circ Syst-I, 2006, 53: 326–337

Weak key analysis for chaotic cipher based on randomness properties

YIN RuMing, YUAN Jian*, SHAN XiuMing & WANG XiQin

Department of Electronic Engineering, Tsinghua University, Beijing 100084*, China*
*E-mail: jyuan@tsinghua.edu.cn

Abstract Weak key analysis is a key issue in the design of chaotic ciphers. While most of the existing research focusing on the degradation of the chaotic sequences which cause weak keys, we point out that the parameters for which the chaotic sequences do not degrade are still possible to be weak keys. In this paper, we propose a new approach based on the rigorous statistical test to improve the weak key analysis. The weak keys of a specific chaotic cipher are investigated by using our method and a large number of new weak keys are detected. These results verify that our method is more effective. On the other hand, although statistical tests are now widely adopted to test the chaos-based bit sequences, there are few reports of analysis results on the weak keys or weak sequences of chaotic cipher. Thus our work may be helpful for current research on statistical tests of chaotic cipher.

Keywords chaos, cryptography, statistical test, weak keys, sequence randomness