基于一维扩展元胞自动机的伪随机数发生器研究*)

赵学龙[1] 王庆梅[1] 许满武[2] 刘凤玉[1]

(南京理工大学计算机科学与技术系 南京210094)[1] (南京大学计算机科学与技术系 南京210008)[2]

摘 要 随机数作为加密数据的载体,备受关注,自然,随机数发生器亦成为密码学的重要研究课题之一。依据元胞自动机(CA)的基本理论,构造出一维扩展元胞自动机模型,借以生成随机数发生器的元胞自动机规则,同时引进遗传算法(GA),以元胞状态序列的熵作为遗传演化的适应度,从而有效地实现最佳元胞自动机规则的搜索,即构造出特定初始条件下的最优GA-CA耦合随机数发生器。最后对其生成随机数性能进行统计检验,其结果通过美国联邦信息处理标准(FIPS140-2)。为网络安全应用提供一种新的、特别适合硬件实现的、简单快速的伪随机数发生器。

关键词 元胞自动机,演化算法,伪随机数发生器,密码学

Pseudo Random Numbers Generator Based on One-Dimensional Extended Cellular Automata

ZHAO Xue-Long[1] WANG Qing-Mei[1] XU Man-Wu[2] LIU Feng-Yu[1]

(Department of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094)[1]

(Department of Computer Science and Technology, Nanjing University, Nanjing 210008)[2]

Abstract Pseudo random numbers are very important in many fields in cryptography, and it is a hard problem to gain a good pseudo random number. Based on the basic theory of cellular automata (CA), using genetic algorithm to find the rule of CA, one dimension extended nonuniform CA is constructed to generate good quality random number. The produced random numbers pass all tests specified by Federal Information Processing Standards (FIPS) 140-2. This generator outperforms previously other one-dimensional CA generator, and it provides a constructive method for pseudo random number generator in cryptography.

Keywords Cellular automata, Genetic algorithm, Pseudo random number generator, Cryptography

1 引言

随机数在现代网络安全中有重要而广泛的应用,通常并不能获得真正的随机数,而只是采用一些方法来生成伪随机数。目前常用的伪随机数发生方法主要有线性同余发生方法,这种方法的主要问题是不适合硬件实现。另外一种流行的方法是线性反馈移位寄存器(LFSR)发生器,这种方法适合硬件实现但是在VLSI实现中很难模块化。而基于元胞自动机(CA--cellular automata)的生成方法则是这个领域的最新进展,它依据局部作用原理,非常适合硬件实现,又可以快速并行地产生伪随机数,具有非常优越的特点。

Wolfram、Hortensius、Tsalides、Nandi等人对CA在伪随机数的生成进行了详细论述。这些工作本质上都是通过对位串模式结构的研究,采用手工方法建立CA规则,构造CA伪随机数发生器,这些工作缺点在于繁杂,需要通过大量的数学试验来验证。另外一类方法就是引入优化方法来得到CA的规则,Sipper和Tomassini[1,2]引入演化算法来发现元胞自动机的规则,研究了基本元胞自动机中生成高质量伪随机数的一维和二维元胞自动机规则。Seredynski等人[3]对这种方法做了进一步研究,对CA规则的半径进行了扩展。这类方法大大减少了寻找CA规则的工作,能够自动地找出满足条件的最优CA规则。本文就是沿着这个思路,构造出一维扩展元胞自动机模型,借以生成随机数发生器的元胞自动机规则,同时引进遗传算法(GA--Genetic Algorithm),以元胞状态序列的熵作为遗传演化的适应度,从而有效地实现最佳元胞自动机规则的搜索,即构造出特定初始条件下的最优CA-GA耦合随机数发生器。最后对其生成随机数性能进行统计检验,结果通过美国联邦信息处理标准(FIPS140-2)。为网络安全应用提供一种新的、特别适合硬件实现的、简单快速的伪随机数发生器。

2 扩展元胞自动机及其熵

2.1 元胞自动机的基本原理

元胞自动机最早由冯·诺伊曼于20世纪40年代末提出,是时间、空间和状态都离散的动力系统,元胞自动机的基本思想简单地说就是,元胞在下一时步的状态是由当前其本身的状态及其邻居的状态共同决定。状态是元胞的所有可能取值,邻居是影响元胞状态演化的影响范围,元胞状态演化的函数称为状态演化函数,也就是CA的规则。基本元胞自动机规则就是邻居半径为1、元胞状态为2(0和1)的元胞自动机,其状态演化函数为:

$$s_i(t+1)=f(s_{i-1}(t),s_i(t),s_{i+1}(t)) \quad (1)$$

式中,s为元胞的状态,$s_i(t)$为元胞i在t时刻的状态,f为元胞自动机的演化函数。

*)本课题受国家自然科学基金(60273035)、江苏省攻关项目(BE2003064)资助。赵学龙 博士研究生,主要研究方向为信息安全、元胞自动机。王庆梅 博士研究生,研究方向为网络信息安全、数字水印。许满武 教授,博士生导师,研究领域为可视化对象建模。刘凤玉 教授,博士生导师,主要研究领域为人工智能与信息安全。

Wolfram 对这些规则进行了细致的研究，并根据这些规则的二进制编码的十进制数进行编号，如 $f(1\ 1\ 1)=0$，$f(1\ 1\ 0)=0$，$f(1\ 0\ 1)=0$，$f(1\ 0\ 0)=1$，$f(0\ 1\ 1)=1$，$f(0\ 1\ 0)=1$，$f(0\ 0\ 1)=1$，$f(0\ 0\ 0)=0$，其对应二进制00011110的十进制数是30，则该规则就称为规则30。Wolfram 认为元胞自动机的非线性复杂性是产生伪随机数的原因。如规则30的布尔状态演化函数为

$$s_i(t+1)=s_{i-1}(t)\mathrm{XOR}(s_i(t)\mathrm{OR}s_{i+1}(t)) \tag{2}$$

随机的位串序列从特定元胞时间函数提取(通常是中间元胞)，获得质量很好的伪随机数。

这里所有的元胞都遵循同样的规则进行演化，在对元胞自动机进行具体应用时，这个定义使元胞自动机的灵活性和应用范围受到很大限制。

2.2 扩展元胞自动机模型

定义[1] 所有元胞的规则都相同的元胞自动机称为均匀元胞自动机；每个元胞都按照不同规则进行演化的元胞自动机称为混合元胞自动机。

上面讨论的元胞自动机都是基于均匀元胞自动机的，本文主要是对混合元胞自动机进行研究，同时对 CA 进行进一步扩展，以考虑前一时刻元胞状态对当前元胞的影响，这样，元胞自动机的一般表达式就为：

$$s_i(t+1)=f_i(c_{i-r}^{t-\tau}s_{i-r}^{t-\tau},\cdots,c_{i+r}^{t-\tau}s_{i+r}^{t-\tau},\cdots,c_{i-r}^{t}s_{i-r}^{t},\cdots,c_{i+r}^{t}s_{i+r}^{t}) \tag{3}$$

式中 r 为邻居半径，t 为演化时步，$c_i^{t-\tau}$ 为影响系数，表示 $t-\tau$ 时步元胞是否影响元胞 i 在 $t+1$ 时刻的状态，这里只取0和1，即等于1时表示影响，等于0时表示不影响。当只有 $c_{i-1}^t=c_i^t=c_{i+1}^t=1$，其他系数都为0时，式(3)就转化为式(1)，也就是 Wolfram 研究的基本元胞自动机。随着元胞自动机的邻居半径和影响时步的增加，其复杂性呈指数增加。为了简化起见，我们只考虑元胞 i 在前一时步存在影响的情况，即只有 $c_i^{t-1}=1$ 的情况，这样式(3)就就简化为

$$s_i(t+1)=f_i(s_i(t-1),s_{i-1}(t),s_i(t),s_{i+1}(t)) \tag{4}$$

这就是本文要研究的元胞自动机。

2.3 元胞状态序列的熵

熵是热力学系统中重要的态函数之一。在信息论中，熵是某事件不确定性的度量，封闭系统中，熵只能增加或者保持不变，熵越大，表明越混乱，随机性也越好，因而可以用元胞状态序列的熵来评价元胞自动机生成伪随机数的随机性[1]。设 k 为每个元胞状态序列位置的可能值(这里为元胞的状态)，为了计算每个元胞状态序列(位串)的熵，把位串分成长度为 h 的子序列。长度为 h 的子序列共有 k^h 个可能排列方式。这样整个位串的熵 E_h 可由下式给出：

$$E_h=-\sum_{j=1}^{k^h}p_{h_j}\log_2 p_{h_j} \tag{5}$$

其中，h_j 是 k^h 排列中第 j 个排列，p_{h_j} 是子序列 h_j 的发生概率。当所有的排列概率 p_{h_j} 都相等即等于 $1/k^h$ 时，E_h 得到最大值 h。这样就可以根据熵的大小来初步判断元胞状态序列的随机性。

3 GA-CA 伪随机数生成模型

3.1 遗传算法的基本原理

遗传算法(GA)是基于达尔文进化论原理发展而来的一种全局求优算法，通过优胜劣汰的自然选择机制达到总体最优化。遗传算法，首先对问题的可能解的值进行编码，所有的解称为种群，每个具体的解为一个体，个体的编码称为染色体。染色体由很多基因组成，这些基因是编码的位，通常采用二进制编码。给每个染色体赋予一个适应度的值，按照适应度的值，执行选择、交叉和变异算子，完成一代遗传计算，经过多次迭代，获得最优结果。这里用 GA 来自动优化 CA 的规则，获得高质量伪随机数的 CA 规则。

3.2 GA-CA 耦合模型

考虑(4)式描述的元胞自动机，采用 Wolfram 的编号方式，即 $f(1\ 1\ 1\ 1)=l_{15}$，$f(1\ 1\ 1\ 0)=l_{14}$，$\cdots$，$f(0\ 0\ 0\ 0)=l_0$，对应的十进制值 $R=\sum_{j=0}^{15}l_j2^j$ 作为对应扩展元胞自动机的规则号。显然，可以用16位的二进制对所有规则进行编码。每个规则就是一个染色体，每个二进制位对应的就是一个基因，所有规则的集合就是种群。

首先为每个元胞随机赋予一个初始规则，给元胞自动机赋予一个随机初始构型，在这个构型下，按照指定的随机规则让元胞自动机运行 M 时步，计算每个元胞状态序列的熵，再以另外一个随机构型同样运行 M 时步，把每个元胞的熵累加起来，重复这个计算过程 C 次。把每个元胞的熵作为遗传算法中的适应度，按照局部作用原理，遗传算子也只在局部起作用，即只对相邻接的邻居元胞起作用[1,2]。

根据邻居熵的大小获得元胞 i 的新规则：如果邻居的熵都比元胞 i 本身的熵小，则元胞 i 的规则保持不变；如果有一个邻居的熵大于元胞 i 本身的熵，则用这个邻居的规则替代元胞 i 的规则；如果有两个邻居的熵都比元胞 i 本身的熵大，则把这两个邻居元胞的规则编码进行交叉操作，随机选择一个子代作为元胞 i 的规则；如果有多于两个邻居元胞的熵比元胞 i 本身的熵大(在规则半径大于1时才会有这种情况)，则随机选择两个熵比较大的邻居规则，执行交叉操作，随机选择一个子代作为元胞 i 的规则。最后判断是否满足遗传算法的终止条件，满足则停止，不满足则对每个元胞规则编码的每个基因以变异概率 p_m 执行变异算子操作，得到新的元胞规则。之后把每个元胞的熵置为0，进行新一轮的遗传演化计算。整个算法概括如下：

1. 随机初始化规则表，并每个元胞赋予一个规则，每个 CA 的适应度 $f_i=0$；

2. 初始状态计数器 $c=0$；

3. 随机产生一个初始状态构型，在这个构型上让 CA 运行 M 时步；

4. 计算每个元胞序列的适应度 f_i，同时构型计数器加1，$c=c+1$；

5. 如果运行了 C 次即 $c\mathrm{Mod}C=0$ 时，进行下一步，否则转到3；

6. 计算每个元胞的适应度的值，在邻居中应用遗传算子操作：选择、交叉和变异；

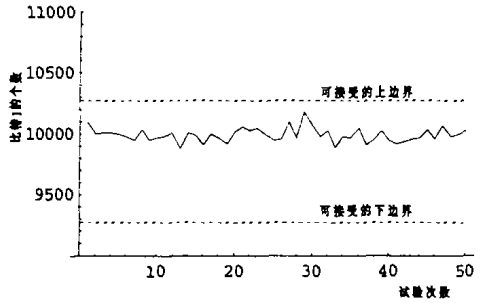
7. 如果不满足终止条件则转到2，否则停止。

4 统计检验

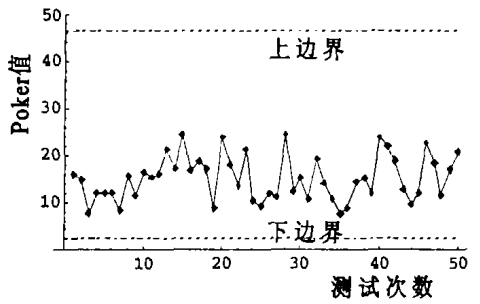
按照上面的模型在元胞空间为50，遗传代数采用50代，在变异概率 $p_m=0.001$，时步 $M=4096$，计算次数 $C=300$ 的条件下，进行了三组试验，得到三组规则。

我们采用美国联邦信息处理标准(FIPS 140-2)[5]公布的加密模块指定的伪随机数统计测试，即频率测试(单比特测试)、序列测试(双比特测试)、扑克测试、游程测试以及自相关

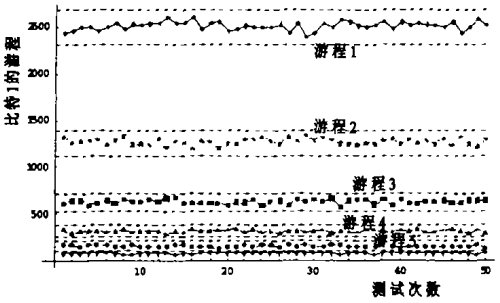
测试5项基本测试[2]。对产生的20000比特伪随机数序列进行测试，若任何一个测试没有通过，则认为没有通过 FIPS 的随机性测试。本文采用两种方法来生成检验的20000比特：a)元胞空间大小为50，运行400时步，生成长度为400的50个随机时间序列，把这50个随机序列连接起来，构成长度为20000比特的序列；b)元胞空间大小为50，运行8时步，然后把50个长度为8比特连接起来构成400比特，这个过程重复50次得到长度为20000比特的随机序列。按照这个测试标准要求，每种方法各进行了50次试验，部分试验结果如图1所示。



(a) 频率测试/单比特测试结果



(b) 扑克测试结果



(c) 游程测试结果结果

图1 GA-CA 随机数发生器生成伪随机数的统计检验结果

从测试结果可以看出，用遗传算法获得的扩展元胞自动机规则可以产生高质量的伪随机数，均全部通过。说明采用遗传算法来生成扩展元胞自动机的规则是可行的，同时也说明，生成的伪随机数满足美国联邦数据处理标准中加密模块的要求。

结论 元胞自动机固有的结构和演化规则简单，而其状态既不可预测又瞬息万变的特性，为随机现象模拟提供了得天独厚的条件。本研究利用这一基本特性，进行了元胞自动机的一维扩展，并结合遗传算法的优化搜索功能，构造出 GA-CA 模型——一维扩展元胞自动机的伪随机数发生器生成模型。试验结果表明，该发生器能生成随机性良好的伪随机数，经统计检验，完全通过美国联邦信息处理标准(FIPS140-2)，满足加密模块的要求。扩展元胞自动机可以快速地生成高质量的伪随机序列，特别适用于需要快速加密的场合。同时这种伪随机数发生器也可以方便地用硬件实现，可以用于 VLSI 内建自测试以及并行计算等多种领域。混合元胞自动机扩展了均匀元胞自动机的应用范围，拓展了元胞自动机的发展应用空间，为密码学等领域应用提供一种简单快速的伪随机数发生器。

一维扩展元胞自动机的演化状态，在一定程度上受到有限邻居的制约。如果构造出二维扩展元胞自动机模型，伪随机数的性质将会得以更大提高。

元胞自动机演化规则确定之后，初始状态则直接影响未来的演化状态，而"求优"实乃在限定空间的"最优化"。若把总体优化与局域优化兼用，例如人工神经网络、支持向量机等方法与遗传算法结合，可能构造出更优秀的随机数发生器。以上两点也正是我们进一步研究的内容。

参 考 文 献

1 Tomassini M, et al. Generating high-quality random numbers in parallel by cellular automata. Future Generation Computer Systems, 1999, 16(2-3): 291～305

2 Tomassini M, Perrenoud M. Cryptography with cellular automata. Applied Soft Computing Journal, 2001, 1(2): 151～160

3 Seredynski F, Bouvry P, Zomaya A Y. Secret Key Cryptography with Cellular Automata. In: Proc. of the Intl. Parallel and Distributed Processing Symposium (IPDPS'03), Nice, France, 2003

4 Chopard B, Droz M. 物理系统的元胞自动机模拟. 祝玉学，赵学龙译. 北京：清华大学出版社，2003

5 FIPS. FIPS140-2: Security requirements for Cryptographic Modules. http://csrc.nist.gov/publications/fips/fips140-2/fips1402.pdf. 2001

(上接第136页)

CRL 大小，改善了响应时间，减少时间碎片；又可以降低对 Base CRL 峰值请求率，降低峰值带宽和平均负荷。增量-过量模型适合于在 Delta CRL 的颁发周期和时间跨度较短、证书吊销率不高、证书有效期较长的大型 PKI 系统中。

参 考 文 献

1 Adams C, Farrell S. RFC2510 Internet X.509 Public Key Infrastructure Certificate Management Protocols [s]. RFC2510, Internet Engineer TaskForce, March 1999

2 Rivest R L. Can we eliminate certificate revocation lists. In: Rafael H, ed. Financial Cryptography. Anguilla, 1998. British West Indies: Springer, 1997. 178～183

3 Hously R, Ford W, Polk W, et al. Internet X.509 public key infrastructure certificate and CRL profile. IETFRFC2459, 1999. http://www.ietf.org/rfc/rfc2459.html

4 Micali S. Efficient certificate revocation. TechnicalMemory, MIT/LCS/TM-5426, 1996. http://www.lcs.mit.edu/pub-lications

5 Kocher P. On certificate revocation and validation. In: Hirschfeld, R., ed. Financial Cryptography-FC'98. LNCS1465, Berlin: Springer-Verlag, 1998. 171～177

6 Moni Naor, Kobbi Nissim. Certificate revocation and certificate update. IEEE Journal on Selected Areasin Communications, 2000, 18(1): 561～170

7 王尚平，张亚玲，王育民. 证书吊销的线索二叉排序 Hash 树解决方案. 软件学报，2001，12(9)：1343～1350

8 Arnes A, Just M, Knapskong S J, et al. Selecting revocation solutions for PKI. Paper Submitted to NORSEC2000, 2000

9 Cooper D A. A more efficient use of Delta-CRLs. In: Proc. of the 2000 IEEE Symposium on Security and Privacy, 2000. 190～202

10 David A C. A Model of Certificate Revoction [C]. In: Proc. 15th Annual, Computer Security Applications Conference, 1999. 256～264

11 David A C. A closer look at revocation and key compromise in public key infrastructures[ED/CD]. http://csrc.nist.gov/nissc/1998/proceedings/paperG2.pdf, 1998-10-11