一种基于误差快速扩散元胞自动机的加密技术*

王福来†

(浙江财经学院数学与统计学院,杭州 310012)

(2010 年 9 月 1 日收到;2011 年 1 月 16 日收到修改稿)

构造了一个具有较大密钥空间的新型一维元胞自动机. 在该元胞自动机中,密钥为采用移位映射的伪随机序列及受控扰动项,避免了数据膨胀,元胞自动机具有随机性触发规则. 该元胞自动机一次处理信息量大,避免了复杂的计算过程. 所生成的流密码在理论上被证明了具有理想的随机性与雪崩效应,误差扩散速度快. 实证分析研究表明,流密码不仅在全局上、而且在局部上都具有良好的随机性能,通过测试长度为 24000 的流密码在 400 次迭代产生的数据表明,经 χ^2 检验,在显著性水平为 5% 时,频数检验通过率超过 95%,序列检验通过率为 100%;数据敏感性检验方面,密钥在任意位置改变 1bit 导致各流密码比特数发生改变的平均值(灵敏值为 V)为 49%—51% 之间(理论上 $V=50\%$),均值为 49.99%,方差为 1.193×10^{-5}. 因此这是一个较理想的加密技术工具.

关键词: 保密通信, 元胞自动机, 伪随机序列

PACS: 05.45.Vx

1. 引 言

元胞自动机不仅广泛用于交通、经济、工程等领域,也已成为信息科学界研究的热门课题之一[1—7]. 由于采用元胞自动机加密通信有利于提高加密速度,人们提出了各种基于元胞自动机的加密方案[1—3],并且将改进的重点放在加快误差扩散速度、处理大信息量、增大密钥空间等方面. Gutowitz[1] 提出了基于触发元胞自动机的分组密码算法. 该算法采用了一种特别的触发规则,用反向迭代实现加密、正向迭代实现解密,具有简单、方便硬件实现等优点. 平萍等[2] 改进了文献[1]的算法,提出了基于复合元胞自动机系统构造分组密码的新方法,使得明文的敏感性大大增提高,且不会出现如分组密码算法[1] 中敏感性在第 5 步达到峰值后剧减的缺点;但文献[2]的方法中明文的敏感性仍不够强,在 50 次之后才能体现出雪崩效应,即在 50 步后灵敏值 $V=48\%$(理论上 $V=50\%$);并且该元胞自动机规则下流密码的随机性机理不明确. 张旭等[3] 提出了一种改进的元胞自动机,其特点是:1) 用密文直接产生流密码,实现即时同步;2) 采用元胞自动机思想进行加密,保证了该系统具有高速通信的特点. 但该方法的缺点是:由于每个元胞的状态多达 2^{24} 个,且没有有效的演化规则,导致流密码对密钥的敏感性在局部上不够敏感,局部的 V 值为 20%—80%, 波动幅度太大,这影响了流密码的质量,导致在加密过程中第 45 次出现 1bit 扰动时,第 73 次(最后一次)仍能获得基本图像. 元胞自动机的演化规则对结果的影响机理不明确.

无论是元胞自动机加密还是伪随机数的产生,最大的困难在于伪随机数的产生机理在理论上不易得到保障. Borcherds 等[9] 及笔者[10] 指出有限精度混沌系统出现的短周期行为难以进行精确的分析;文献[11]也指出文献[12]提出的随机数生成方法,对于许多初始值,生成的随机数序列不能经受卡方检验. 这些都体现出实证分析与理论分析的研究仍存在不足.

本文提出的元胞自动机模型,在理论上被证明了是可靠的,在实证中是有效的,从而增加了保密通信的安全性. 该元胞自动机优点是:1) 密钥空间大. 密钥采用了文献[13]的高质量伪随机数序列生成方法,用密文直接产生流密码,即实行同步. 2) 从理论上保证了流密码具有良好的随机性与敏感性,元胞自动机的演化规则对结果的影响机理非常明确,大大提高了抗破译能力. 3) 由于采用移位映射

* 国家自然科学基金(批准号: 10871168)资助的课题.

† E-mail: flyerwon@sina.com

及受控扰动项的方法,避免了数据膨胀,加快了加解密速度.

2. 元胞自动机模型

2.1. 相关定义

定义 1[3](元胞自动机) 一维元胞自动机是一个三元组 $CA=(S,R,f)$,其中 S 为有限状态集,R 为邻域半径,f 为映射函数,又简称为规则.

定义 2[3](全局状态配置) 设元胞自动机是由 N 个元胞构成的有限元胞自动机,各元胞排列成一行,按顺序分别编号为 1, 2, …, N,称 $G^{(t)}=(s^t(1),s^t(2),\cdots,s^t(N))$ 为元胞自动机在 t 时刻的一个全局状态配置.

定义 3[12](均匀映射) 设混沌映射为

$$x_n=f(x_{n-1})\quad(n=1,2,\cdots),\tag{1}$$

其中 x_0 为初值. 均匀映射 $g(\cdot)$ 定义为

$$y_n=g(x_n)=\frac{K(x_n)}{N}\quad(n=1,2,\cdots,N),\tag{2}$$

其中 N 为序列 $\{x_n\}$ 的长度. 长度为 N 的序列 $\{x_n\}$ 记作 $\{x_n\}_1^N$,$K(x_n)$ 为 x_n 在混沌序列 $\{x_n\}_1^N$ 中按升序排列的序号.

显然 $\{g(x_i)\}_{i=1}^N$ 为在 $[0,1]$ 区间上为标准均匀分布的序列,且 $g(\cdot)$ 为同胚映射,系统(1)与(2)在拓扑结构上是等价的,因此 $\{x_n\}_1^N$ 与 $\{y_n\}_{n=1}^N$ 保持了诸多拓扑不变性,也很大程度上保持了原有动力系统性质,如两个系统的排列熵[14,15]相同.

为生成混沌序列密码,引入不可逆转函数 $T_1(x_n)$. 转换函数为

$$T_1(x_n)=\begin{cases}0 & (x_n\in\bigcup_{d=1}^{m}B_{2d-1}^{2m}),\\ 1 & (x_n\in\bigcup_{d=1}^{m}B_{2d}^{2m}),\end{cases}\tag{3}$$

其中 $2m$ 为正整数,$B_0^{2m},B_1^{2m},B_2^{2m},\cdots,B_{2m}^{2m}$ 是 $[0,1]$ 区间的 $2m$ 个连续的等分区间. $T(x_n)$ 的作用即是将 $2m$ 个连续区间交替变成 0,1 符号序列. 一般随序列长度 N 的增大 $2m$ 也取得大些.

类似地得到以 1, 2, 3, 4 为元素的转换函数

$$T_2(x_n)=\begin{cases}1 & (x_n\in\bigcup_{d=1}^{m}B_{4d-3}^{4m}),\\ 2 & (x_n\in\bigcup_{d=1}^{m}B_{4d-2}^{4m}),\\ 3 & (x_n\in\bigcup_{d=1}^{m}B_{4d-1}^{4m}),\\ 4 & (x_n\in\bigcup_{d=1}^{m}B_{4d}^{4m}),\end{cases}\tag{4}$$

笔者[13]已在实践上给出了生成高质量伪随机序列的算法,在文献[14]中给出了序列复杂度的算法,并在理论上给出了证明,可有效地避免由于计算精度带来的弱密钥的产生. 选择序列的元素个数并没有特别要求,对结果也无影响;但是当元素个数过多时,序列的长度应适当增加.

由文献[13]的定理 1 我们容易得到相似结论:设 X 为 $[0,1]$ 上的均匀随机变量,样本序列为 $\{x_n\}_1^N$,则经符号变换(4)后,所得的序列 $\{s_n\}_1^N$ 的极限分布为"1—4"的均匀随机分布.

定义 4($S^d(\cdot)$ 移位映射) 从 $\{m_k\}_1^N$ 的元素 m_k 变为 $m_{k+d(\mathrm{mod}N)}$,这个变换过程称为 $S^d(\cdot)$ 移位映射,即 $S^d(m_k)=m_{k+d(\mathrm{mod}N)}$,$S^d(\{m_k\}_1^N)=\{m_{k+d(\mathrm{mod}N)}\}_1^N$.

定义 5(k 位扰动) 对一个"0,1"序列 $\{m_k\}_1^N$ 的第 i 位置的扰动是指将 m_i 改为 $1-m_i$,而其他符号保持不变.

显然,当 $n<N$,经过 $S^n(\cdot)$ 移位,$\{m_k^0\}_1^N$ 每个元素的位置都发生了改变. 由 $\{m_k\}_1^N$ 的随机性,由于序列长度 N 较大,执行 $S^d(\cdot)$ 移位后,$S^d(\{m_k\}_1^N)$ 仍为随机序列.

2.2. 元胞自动机演化规则

元胞自动机演化规则为

第 1 步 由相同或不同的混沌映射分别生成长度为 N 的"0,1"伪随机序列 $\{m_k^0\}_1^N$ 和"1—4"伪随机序列 $\{n_k^0\}_1^N$.

第 2 步 任意选择 $r_j^i(0\leqslant i_1,i_2\leqslant 400,1\leqslant j_1,j_2\leqslant N)$,将($\{m_k^0\}_1^N$,$\{n_k^0\}_1^N$,$r_j^i$)共同作为密钥. $\{m_k^i\}_1^N$ 表示 $\{m_k^0\}_1^N$ 的第 i 次迭代,本模型取最大迭代次数为 $N_1=400$;r_j^i 表示将 $\{m_k^i\}_1^N$ 的第 j 个数据扰动,受控扰动项可设置多个或不设.

第 3 步 首先检测 $\{m_k^t\}_1^N$ 有无被扰动,若无扰动,则 $\{n_k^{t+1}\}_1^N=S^1\{n_k^t\}_1^N$;若在非第 i 次迭代或第 r_j^i 项发生扰动,则 $\{n_k^{t+1}\}_1^N=S^2\{n_k^t\}_1^N$;若在第 i 次的非 r_j^i 项发生扰动,则 $\{n_k^{i+1}\}_1^N=S^3\{n_k^i\}_1^N$. 注意,作为密钥的 r_j^i 是受控扰动项,在不为零时是必定会发生扰动的.

第 4 步 设在 $t(t=0,1,2,\cdots)$ 时刻 $G(t)$ 的状态变为 $\{m_k^t\}_1^N$,则 $\{m_k^{t+1}\}_1^N$ 的规则为

$$m_k^{t+1}=\begin{cases}\mathrm{xor}(m_{n_k^t}^t,m_k^t) & (k=1\text{ 或 }N),\\ 1-\mathrm{xor}(\mathrm{xor}(m_{n_k^t-1}^t,m_{n_k^t+1}^t),m_k^t) & \\ \qquad(k=2,3,\cdots,N-1),\end{cases}\tag{5}$$

这里 $1-x\mathrm{or}(\cdot)$ 为 $x\mathrm{or}(\cdot)$ 的相反的结果,即同或.

即当 $t=1$ 或 N 时,先求 m_k^t 在第 n_k^t 位置的值 $m_{n_k^t}^t$,再求 $m_{n_k^t}^t$ 与 m_k^t 的异或;当 $t=2,3,\cdots,N-1$,先求 m_k^t 在第 n_k^t-1 位置和第 n_k^t+1 的值 $m_{n_k^{t+1}-1}^t$ 与 $m_{n_k^{t+1}+1}^t$ 的异或,并将结果与 m_k^t 求异或;转至第 3 步.

同或或异或产交错运用可有待于加速误差扩散,能避免出现连续多个"1"的出现.

2.3. 理论分析

设 $\{m_k^1\}_1^N$ 为随机"0,1"序列,$\{n_k^1\}_1^N$ 为随机"1—4"随机序列,因为 $m_{n_k^1}^1$ 与 m_k^1,n_k^1-1 与 n_k^1+1 都是相互独立的,故由(5)式知 $\{m_k^2\}_1^N$ 仍为"0,1"随机序列;并且由归纳法还可知,当 $t<N$ 时,$\{m_k^t\}_1^N$ 仍为"0,1"随机序列.

由于迭代次数远小于密钥的长度,即 $N_1 \ll N$ 时,如 $N_1/N=400/24000 \ll 1$,$\{n_k^0\}_1^N$ 不可能经迭代使 n_k^0 重新回到第一位置. 根据概率理论,我们进一步得到

定理 1 m_k^t 与 m_k^{t+1} 是相互独立的;m_k^t 与 m_k^{t+d} $(0\leqslant t+d\leqslant 400, d\neq 0)$ 是相互独立的.

由定理 1 可知,$P(m_k^t=m_k^{t+1})=0.5$,则总体期望值为 $0.5N$,总体均值为 50%;于是得推论 1.

推论 1 $\lim\limits_{N\to\infty}\{m_k^t\}_1^N$ 有理想的"雪崩效应",即敏感值 $V=50\%$.

2.4. 实证分析

在实证时,取密钥为由

$$x_n=\mu x_{n-1}(1-x_{n-1}) \quad (n=1,2,\cdots). \quad (6)$$

当 $\mu=3.97$,$x_0=0.8$ 时产生的序列经文献[13]的方法而得到"0,1"序列 $\{m_k^0\}_1^N$ 及"1—4"随机序列 $\{n_k^0\}_1^N$. $N=24000$,迭代次数为 $N_1=400$. r_j^i 分别设为 0(即不扰动)及 $r_j^i=r_{100}^0$ 进行试验. 由($\{m_k^0\}_1^N$,$\{n_k^0\}_1^N$,r_j^i)共同作为密钥.

2.4.1. 频数检验、序列检验

1)检测每一次迭代中 0 的个数 n_0 和 1 的个数 n_1,计算统计量 $\chi^2=\dfrac{(n_0-n_1)^2}{N}$. 显著水平为 5% 时,当 $\chi^2\leqslant 3.841$ 时为通过频数检验[3].

2)设 n_{00},n_{01},n_{10},n_{11} 分别为每一次迭代中 00, 01, 10, 11 模式出现的频数. 计算统计量 $\chi^2=\dfrac{4}{N-1}\sum\limits_{i=0}^{i=1}\sum\limits_{j=0}^{j=1}(n_{ij}^2)-\dfrac{2}{N}\sum\limits_{i=0}^{i=1}(n_i^2)+1$. 显著水平为 5% 时,当 $\chi^2\leqslant 5.991$ 时为通过序列检验[3].

分别设置扰动项为 $r_j^i=0$(即没有扰动),$r_j^i=r_{100}^0$,计算长度 $N=24000$,在 400 次迭代中频数检验的平均值及方差与序列检验通过率(见表 1).

表 1 频数检验及序列检验

通过率及方差	频数通过率均值/%	频数方差 $/10^{-5}$	序列通过率/%
$N=24000$, $r_j^i=0$	96.00	1.193	100
$N=24000$, $r_j^i=r_{100}^0$	95.25	1.2409	100
理论值	100	0	100

从表 1 可看出,不论是频数检验还是序列检验,结果都十分理想.

2.4.2. 密钥敏感性分析

分别设置扰动项为 $r_j^i=0$(即没有扰动),$r_j^i=r_{100}^0$,计算长度为 $N=24000$,在 400 次迭代中,与 $\{m_k^0\}_1^N$ 的比特数发生改变的平均值;再计算 $r_j^i=r_{100}^0$,计算长度为 $N=24000$,在 400 次迭代中,与 $\{m_k^0\}_1^{7200}$ 的比特数发生改变的平均值见图 1 及表 2. 从表 2 及图 1 可看出,均值非常接近 50% 的理想值,且方差十分小. 误差扩散非常快,能在第一次迭代时迅速达到接近 50% 的理想值,这比文献[1—3]的结果都要好.

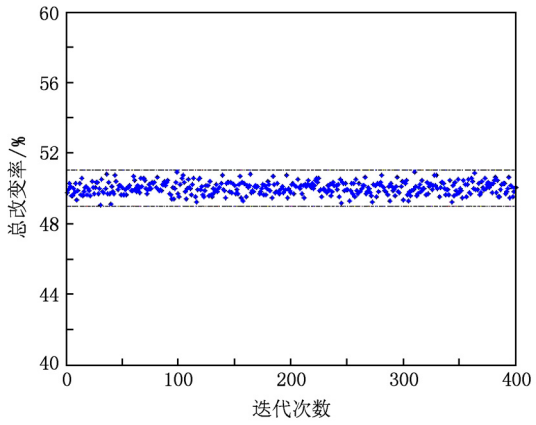

图 1 密钥改变 1bit 时各次输出流密码的总改变率

表 2 不同混沌映射下短游程的比重

参数	均值/%	方差
$N=24000$, $r_j^i=0$	49.99	1.193
$N=24000$, $r_j^i=r_{100}^0$	50.00	1.2409
$N=72000$, $r_j^i=r_{100}^0$	50.05	4.2639
理论值	50.00	0

为给像素为 256 × 256 的采用 unit8 型数组的 Lenna 图像加解密,从元胞自动机的 400 次迭代数组集合中取前 256 × 256 × 8 个数据进行加解密,如图 2 所示.

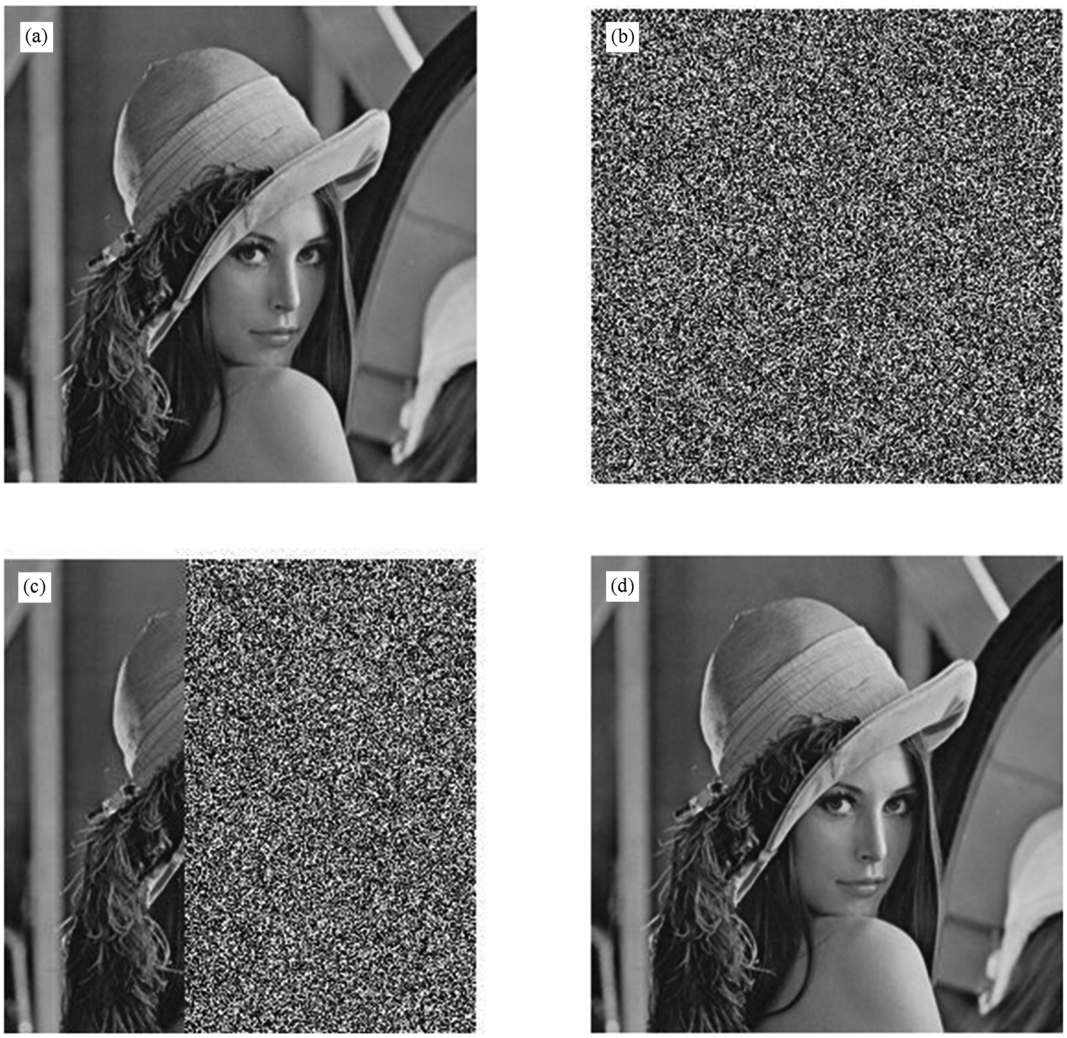

图 2 Lenna 图像的加密解密 (a) 原始 Lenna 图像;(b) 第 1 次迭代时发生 1bit 扰动后的解密图像;(c)第 13 次迭代(相当于 73 次传输过程中的第 45 次)时 1bit 发生扰动时的解密图像;(d)正确解密图像

表 3 短游程的比重

模式	00	01	10	11	000	111
$N=24000$, $r_j^i=0$ 均值	0.2500	0.2499	0.2499	0.2502	0.1251	0.1252
$N=24000$, $r_j^i=0$ 方差/10^{-6}	12.352	2.8519	2.8407	11.800	9.5017	9.2152
$N=72000$, $r_j^i=r_{100}^0$均值	0.2504	0.2498	0.2498	0.2498	0.1250	0.1253
$N=72000$, $r_j^i=r_{100}^0$方差/10^{-5}	4.0726	0.9176	4.3481	9.1769	2.9821	3.1759
理论值	0.2500	0.2500	0.2500	0.2500	0.1250	0.1250

2.4.3. *游程检验*

由文献[1—3, 17]可知这是非常理想的结果.

2.4.4. *自机关、互相关分析*

设$\{x_n\}_1^N$和$\{y_n\}_1^N$为两个长度为 N 的伪随机序列. 自相关函数定义为

$$R_{xx}(T) = \frac{1}{N}\sum_{i=0}^{N-1} x_i x_{i+T}; \tag{7}$$

互相关函数定义为

$$R_{xy}(T) = \frac{1}{N}\sum_{i=0}^{N-1} x_i y_{i+T}, \tag{8}$$

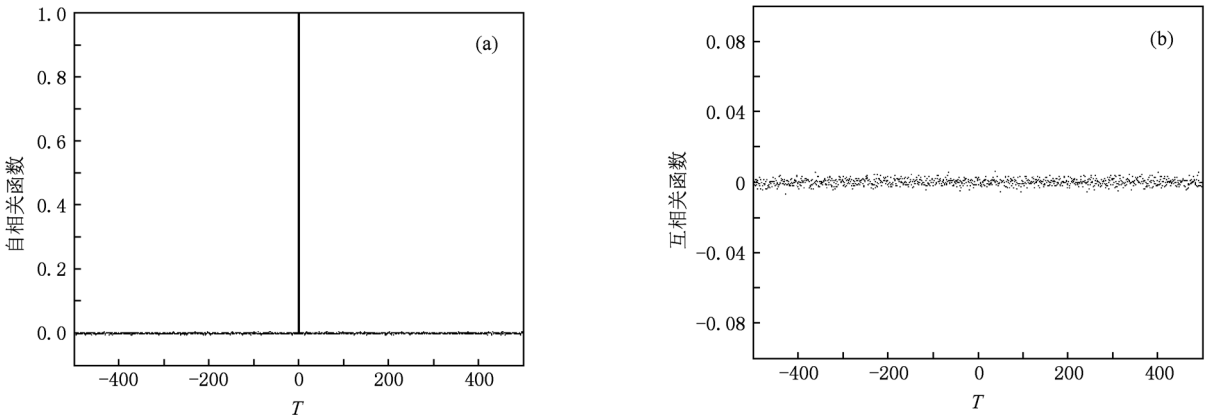

图 3 相关函数与相关间隔的关系 (a)自相关函数,(b)互相关函数

其中,N 为序列长度,T 为相关间隔.

为与以往文献作比较,本文取 $N = 2000$ 或 20000, $-500 \leqslant T \leqslant 500$,计算各序列的相关特性. 取 $N = 24000$, $r_j^i = r_{100}^0$ 在第 20,21 次迭代数据的前 40000 个数据,相关间隔 T 取不同值,按(7),(8)式计算得自相关与互相关函数见图 3.

除当 $T = 0$ 外,最大自相关值为 0.0053;最大互相关值为 0.0075. 从图 3 的结果来看,本文算法得出的伪随机序列的自相关函数近似于 δ 函数,互相关性函数近似为零,即是一个全域零相关系统[18,19]. 比较文献[17],可知本文算法所得的伪随机序列的自相关与互相关性能比文献[17]给出的优越.

3. 结 论

提出了一个新型的元胞自动机,由于密钥的产生与混沌序列产生的函数及受控扰动项有关,使得密钥空间增大,增大了破译难度,能有效地掩盖明文信息. 该元胞自动机生成同步流密码的速度快. 要保持迭代次数远小于密钥序列长度. 该元胞自动机在通信保密传输方面是一个理想的技术.

[1] Gutowitz H A 1994 *Method and Apparantus for Encryption, Decryption and Authentication Using Dynamical Systems* USA Patent: 5-395-589

[2] Ping P, Zhao X L, Zhang H, Liu F Y 2008 *Acta Phys. Sin.* **57** 6188 (in Chinese) [平 萍、赵学龙、张 宏、刘凤玉 2008 物理学报 **57** 6188]

[3] Zhang X, Ren W, Tang D N, Tang G N 2010 *Acta Phys. Sin.* **59** 5281 (in Chinese) [张 旭、任 卫、唐冬妮、唐国宁 2010 物理学报 **57** 5281]

[4] Ding J X, Huang H J 2010 *Acta Phys. Sin.* **59** 3093 (in Chinese) [丁建勋、黄海军 2010 物理学报 **59** 3093]

[5] Qian Y S, Wang H L, Wang C L 2008 *Acta Phys. Sin.* **57** 2115 (in Chinese) [钱勇生、汪海龙、王春雷 2008 物理学报 **57** 2115]

[6] Wang L, Zhou S H, Yuan J, Ren Y, Shan X M 2007 *Acta Phys. Sin.* **56** 36 (in Chinese) [王 磊、周淑华、袁 坚、任 勇、山秀明 物理学报 2007 **56** 36]

[7] Li K P, Gao Z Y 2005 *Chin. Phys.* **14** 930

[8] Qian Y S, Shi P J, Zeng Q, Ma C X, Lin F, Sun P, Wang H L 2010 *Chin. Phys.* B **19** 048201

[9] Borcherds P H, Mccauley G P 1993 *Chaos Soliton. Fract.* **3** 451

[10] Wang Fulai 2010 *Advances in Difference Equations* Doi:10.1155/2010/985982 Article ID 985982

[11] Liang H, Lui Q H, Bai F S 2005 *Comput. Math. Appl.* **49** 331

[12] Sobol I M, Levitan Y L 1999 *Comput. Math. Applic.* **37** 33

[13] Wang F L 2010 *Chin. Phys.* B **19** 090505

[14] Hou W, Feng G L, Deng W J, Li J P 2006 *Acta Phys. Sin.* **57** 37 (in Chinese) [侯 威、封国林、董文杰、李建平 2006 物理学报 **55** 2663]

[15] Cao Y H, Tung W W, Gao J B Protopopescu V A, Hively L M 2004 *Phys. Rev.* E **70** 217

[16] Wang F L 2010 *Chin Phys.* B **19** 0605151

[17] Sheng L Y, Xiao Y Y, Sheng Z 2008 *Acta Phys. Sin* **57** 4007 (in Chinese) [盛利元、肖燕予、盛 喆 2008 物理学报 **57** 4007]

[18] Wichmann B A, Hill I D 2006 *Comput. Stat. Data Anal.* **51** 1614

[19] Sánchez S, Criado R, Vega C 2005 *Math. Coput. Model.* **42** 809

A method of digital secure communication based on a cellular automata with rapid dispersion of errors*

Wang Fu-Lai†

(*School of mathematics and statistics, Zhejiang University of Finance and Economics, Hangzhou* 310012, *China*)

(Received 1 September 2010; revised manuscript received 16 January 2011)

Abstract

An improved one-dimensional cellular automata is designed in which the key space is large with pseudo random series on a shift map and perturbed terms and thus data expansion is avoided. Random triggering rules are involved. There is no complex computation, but a large amount of information can be processed every time. The stream cipher generated by the automata is proved to be of ideal randomness and avalanche effect with rapid dispersion velocity of errors. Empirical results show that the stream cipher is perfectly random both globally and locally. The chi-square test (confidence 95%) on a set of stream cipher with a length of 24000 and its 400 time iterations shows that the passing rates of frequencies and series are above 95% and 100%, respectively. To test the sensitivity of the data, 1 bit is changed at any position of the key stream and the average variation rate of total bits is 49.99%, ranging from 49% to 51% (theoretical value is 50%), and variance is $1.193 \cdot 10^{-5}$, which means that the automata is a good encryption technique.

Keywords: secure communication, cellular automata, pseudo-random series

PACS: 05.45.Vx

* Projected supported by the National Natural Science Foundation of China (Grant No. 10871168).

† E-mail: flyerwon@sina.com