基于抛物线映射的混沌 LT 编码算法

黄　诚[①②]　　易本顺[①]
[①](武汉大学电子信息学院　武汉　430072)
[②](中国电信集团武汉分公司　武汉　430071)

摘　要：该文提出一种基于抛物线映射和混沌置乱方法的 LT 编码算法。首先用混沌初始值作为密钥，采用抛物线映射产生混沌序列并转换为类均匀分布序列，再通过位置置乱算法生成 LT 码的度分布和度邻接数据序列，较传统的重要抽样方法具有更高的灵敏度，保留了理论分布的结构。实验结果表明，该算法具有实现结构简单、分组头部开销小、保密性好及高于传统重要抽样方法的性能。

关键词：混沌；抛物线映射；混沌置乱；无码率码；LT 码
中图分类号：TN911.22　　**文献标识码**：A　　**文章编号**：1009-5896(2009)10-2527-05

Chaotic LT Encoding Algorithm Based on Parabolic Map

Huang Cheng[①②]　　Yi Ben-shun[①]
[①](*Electronic Information School*, *Wuhan University*, *Wuhan* 430072, *China*)
[②](*China Telecom Group*, *Wuhan Branch Company*, *Wuhan* 430071, *China*)

Abstract: A Luby Transform (LT) encoding algorithm based on parabolic map by using chaotic scrambling method is proposed. Firstly, chaotic sequences are produced by using parabolic map and then transformed into uniform-like sequences. The degree distribution and data set of neighbors of LT codes are generated by using position scrambling algorithm which is more sensitive than traditional importance sampling method keeping the construction of theoretical distribution. Experimental results show that the algorithm has more simple construction, smaller header costs of the packets, better encryption effect and furthermore is outperformed the traditional importance sampling method.

Key words: Chaos; Parabolic map; Chaotic scrambling; Rateless codes; Luby Transform(LT) codes

1　引言

无码率码(又称喷泉码、数字喷泉)以 LT(Luby Transform)码为典型代表，其特点是编码码率不为定值，同时也是一种基于分组的信道编码方式，发送端通过无码率编码可将原始分组序列转换为无限长的编码分组序列，而接收端只需从中接收 K 个略大于原始分组数量 k 的分组，即能以很高的概率译码恢复，很少或不需反馈重传控制，适用于分组擦除信道或有损信道的实时应用，特别适用于单点对多点通信[1–3]。

混沌在通信中主要应用于保密和抗干扰，如信息加密、图像置乱、跳频码的生成等。目前也有部分与信道编码技术结合的例子，文献[4]将混沌理论引入了 Turbo 码和 LDPC(Low Density Parity Check)码的设计与实现，属于混沌与固定码率编码结合的实例。本文则将混沌理论引入 LT 码的编码环节，实现混沌与无码率编码的结合，具有 3 种有别于传统方法的特点：(1)传统随机发生器生成度分布数据需要将度分布数据和度邻接数据加入 LT 编码分组作为头部开销，接收方依据头部开销数据来恢复译码，对于度分布的平均度数 $H(d)$，需要加入 $H(d)\ln k$ 的头部开销，而本文提出的混沌 LT 编码算法同步时不需要额外开销；在基于分组交换的擦除信道中，开销值是与编码分组上限 N 有关的 $\ln N$，一般情况下远小于 $H(d)\ln k$。(2)利用抛物线映射实现对 LT 码的度分布选取，同时生成类似均匀分布的数据序列，再采用排序置乱方法生成度邻接数据，与文献[5–9]中直接使用均匀分布的方法进行对比，仿真表明能得到更好的译码性能。(3)由于接收端必须获得与编码端相同的混沌初始值即密钥才能实施译码，因此混沌 LT 编码方法保密性能要高于传统方法。

基于上述思路，文章第 2 节介绍 LT 码的编译码原理，第 3 节阐述抛物线映射的 LT 码实现方案，在分析 LT 码原理的基础上，提出基于抛物线映射的实现方法并进行分析和对比，在第 4 节对基于抛

2008-11-03 收到，2009-03-26 改回

物线映射的 LT 编码方法进行仿真实验并对比传统方法，讨论了混沌 LT 编码方法的相关特性，得出结论。

2 LT 码原理

文献[1]对 LT 编码中用到的度分布设计进行了推导和分析，该分布称为鲁棒孤子分布(Robust Soliton distribution)。如图 1 所示，空心方块为原始分组，黑色实心方块为 LT 编码分组，LT 编码的算法步骤如下：

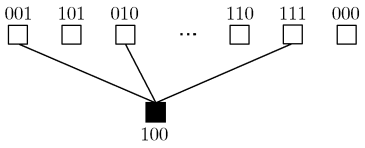

图 1 度数为 3 的 LT 分组编码过程

(1)从鲁棒孤子分布 $\mu(\bullet)$中任意选取一个度值 d_i。

(2)从 k 个原始分组中随机且均匀地选取 d_i 个不同的原始分组。

(3)将这 d_i 个原始分组进行异或运算，生成一个编码分组。

(4)重复(1)到(3)，不断生成 LT 编码分组。

收到 K 个编码分组后开始译码算法步骤，一般 K 略大于原始分组个数 k。如图 2 所示，图上部的空心方块为原始分组，下部的黑色实心方块为接收到的编码分组，LT 译码算法步骤如下：

(1)在接收到一定数量的编码分组后，若存在度数为 1 的编码分组，即可开始译码，将该编码分组直接复制给 S_1，同时将处理完成的分组删除，如图 2(a)。

(2)已赋值的 S_1 分组还与两个编码分组相连，将 S_1 分组的值与这两个分组进行异或运算，同时移除连接关系，如图 2(b)和 2(c)。

重复(1)和(2)，也从度数为 1 的编码分组开始继续译码，即最终完成译码过程，如图 2(d)–2(e)。

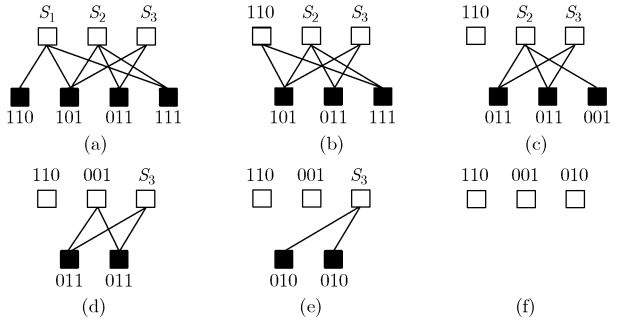

图 2 原始分组为 3，接收 LT 编码分组为 4 的译码过程

3 基于抛物线映射的 LT 码实现

3.1 基于抛物线映射的混沌置乱方法

对于抛物线映射，其迭代表达式为

$$x_{n+1}=1-\mu x_n^2,\ x_n\in(-1,1) \tag{1}$$

当 μ=2 时系统处于混沌状态，其概率密度函数为

$$\rho(x)=\frac{1}{\pi\sqrt{1-x^2}} \tag{2}$$

将此概率密度函数经过下式的变换，可得到类均匀分布的混沌数列。

$$y_n=\frac{1}{\pi}\arcsin x_n+0.5 \tag{3}$$

取一个初值 x_0，就可以得到混沌序列 y_n，y_n为 0–1 区间的类似均匀分布的随机数序列[10]。

按照文献[5–9]中的方法，要解决两个问题，即(1)对于生成的 K 个编码分组，应保证度数据符合设计的度分布，才能保证性能也接近理论值。若按照传统的 IS(Importance Sampling)即重要抽样方法，利用均匀分布进行映射转换来选取度分布数据必然存在灵敏度上的误差，即实际选取的数据分布并不严格符合理论值，存在一定偏离，改变了度分布结构。(2)每次从原始分组中随机且均匀的选取 d_i 个不同的原始分组，即必须保证选取原始分组时不能重复。由于每次选取时连续 n 次选中同一分组的概率为$(1/k)^n$，而两个相同的原始分组异或后的结果为 0，实际度数变成了(d_i–2)，会破坏鲁棒孤子分布的实际分布，造成译码性能下降，若每次选取分组后再与以前的选取数据进行比较，则要生成 K 个 LT 编码分组，计算量为$\sum_{i=1}^{K}\sum_{j=1}^{d_i-1}j$，编码效率较低。

为了解决以上的两个问题，本文采用基于矢量位置系数置乱的方法：

(1)令 $\boldsymbol{M}_i(i=1,2,\cdots,n)$为 n 维矢量。

(2)令 $\boldsymbol{N}_i(i=1,2,\cdots,n)$为由式(3)生成的 n 维混沌矢量，$\boldsymbol{H}_i(i=1,2,\cdots,n)$为与 $\boldsymbol{M}_i$和 $\boldsymbol{N}_i$的位置系数对应的 n 维矢量。

(3)对 $\boldsymbol{N}_i$进行升序或降序排列，则 $\boldsymbol{H}_i$的值相应发生变化生成新的矢量 $\boldsymbol{H}_i'$。

(4)以 $\boldsymbol{H}_i'$的值为矢量位置系数依次从 $\boldsymbol{M}_i$中取值生成新的矢量 $\boldsymbol{M}_i'$。

n 维矢量 $\boldsymbol{M}_i'$即为基于矢量位置系数置乱生成的置乱矢量，可以确保包含所有的原始数据且不重复，并使数据的随机分布状况满足理论要求，还能将生成的置乱矢量数据存入存储设备，用存储复杂度换取计算复杂度。

3.2 鲁棒孤子分布序列生成算法

实际应用中 LT 编码分组不可能无限生成，需

要指定分组生成长度上限 N。在确定上限 N 后，通过抛物线映射生成符合要求的鲁棒孤子分布的度分布数据序列，算法实现步骤如下：

(1)得到符合要求的鲁棒孤子分布 $\mu(\cdot)$后，将各度数 d_i 与编码分组生成上限 N 相乘得出各度数个数按顺序存入一维矢量 $\boldsymbol{M}_0$。

(2)利用混沌序列 y_n 生成 N 个迭代数据，存入一维矢量 $\boldsymbol{M}_y$。

(3)将矢量 $\boldsymbol{M}_0$ 和矢量 $\boldsymbol{M}_y$ 的行列系数一一对应。

(4)将矢量 $\boldsymbol{M}_y$ 升序或降序排列生成新矢量 $\boldsymbol{M}_y^{'}$，此时 $\boldsymbol{M}_y$ 的行列系数被重新排列，将这些系数值计入矢量为 $\boldsymbol{I}_y$。

(5)按照矢量 $\boldsymbol{I}_y$ 的值作为行列系数取出矢量 $\boldsymbol{M}_0$ 的度数据 d_i。

完成步骤(1)至步骤(5)即生成鲁棒孤子度分布数据序列。

3.3 度邻接分布序列生成算法

度邻接数据是影响 LT 编译码性能的重要因素，编码运算量也取决于此。下面采用混沌置乱的方法来生成度邻接数据序列。令鲁棒孤子分布 $\mu(\cdot)$最大度数为 m，则生成 N 个编码分组时需要选取 S_d 个原始分组即度邻接数据的个数为

$$S_d = \sum_{i=1}^{m} \mathrm{floor}(\mu(i)\cdot N) \tag{4}$$

对 k 个原始分组，需要将 S_d 个度邻接数据分为 L 组数据和 L_m 个剩余的度邻接数据。

$$L = \mathrm{floor}(S_d/k) \tag{5}$$

$$L_m = S_d \bmod k \tag{6}$$

度邻接分布序列生成算法的实现步骤如下：

(1)根据需要生成的 N 个编码计算出相应的度邻接数据 S_d。

(2)将 S_d 分解为 L 组数据和 L_m 个剩余的度邻接数据。

(3)利用混沌序列 y_n 生成 k 个迭代数据，存入一维矢量 $\boldsymbol{L}_y$。

(4)将 k 个原始分组数据矢量 $\boldsymbol{K}_0$ 和矢量 $\boldsymbol{L}_y$ 的行列系数一一对应。

(5)将矢量 $\boldsymbol{L}_y$ 升序或降序排列生成新矢量 $\boldsymbol{L}_y^{'}$，此时 $\boldsymbol{L}_y$ 的行列系数被重新排列，将这些系数值计入矢量为 $\boldsymbol{J}_y$。

(6)用矢量 $\boldsymbol{K}_0$ 的原始分组数据生成置乱矢量 $\boldsymbol{K}_1$，按照矢量 $\boldsymbol{J}_y$ 的值作为行列系数依次从 $\boldsymbol{K}_1$ 中取值，当 $\boldsymbol{J}_y$ 的值大于 $\boldsymbol{K}_1$ 中剩余的值时，实施截断并舍弃本组余下的数据，从下一组数据开始取值。

(7)重复步骤(3)- 步骤(6)，直到所有的 L 组数据全部处理完成，同时再处理一组数据按照 L_m 的长度进行截断。

(8)将(7)以前步骤生成的度邻接序列依次合并生成度邻接矢量 N_d。

至此完成所有度邻接数据的处理，然后从 3.2 节的度分布序列中依次取值，再从度邻接矢量 $\boldsymbol{N}_d$ 中取出相应数值的原始分组进行异或操作，不断生成编码分组。

从上述的算法步骤可以得知，由于编码的度分布序列和度邻接序列均采用抛物线映射生成，所以在完全同步的情况下，只需要知道抛物线映射的初始值 x_0，就可以根据分组接收的顺序恢复出相应的度分布序列和度邻接序列，不必向编码分组头部加入这些数据。由于分组交换网络是异步传输的，必须向分组头部加入编码的顺序信息，此时要传递的是分组在全部 N 个编码分组中的生成顺序，开销需要的 bit 数为 $\ln N$，以文献[11]中的实验数据为例，相关头部开销的比较如表 1。

在表 1 中，$H(d)$是度分布的平均度数，k 为原始分组长度，N_1、N_2 分别为设定生成编码分组的上限值，令 N_1 为 k 的 10 倍，N_2 为 k 的 10^{20} 倍，$H(d)\ln k$ 为传统方法的头部平均开销，$\ln N_1$、$\ln N_2$ 分别为不同数量级的上限值下混沌方法的头部开销。在设定生成的编码分组为原始分组 10 倍时，传统方法的头部开销大约是混沌方法的 5 倍，即使编码分组达到原始分组的 10^{20} 倍，尽管已超出实际需要，混沌编码的分组头部开销仍小于传统方法。

表 1 传统 LT 编码与混沌 LT 编码头部开销比较

$H(d)$	k	$H(d)\ln k$	N_1	$\ln N_1$	N_2	$\ln N_2$
5.870283	65536	65.103522	655360	13.392940	65536×10^{20}	57.142057
5.511558	80000	62.224288	800000	13.592367	80000×10^{20}	57.341484
5.849934	100000	67.349854	1000000	13.815511	100000×10^{20}	57.564627
5.825016	120000	68.125001	1200000	13.997832	120000×10^{20}	57.746949

4 实验结果及分析

仿真原理图如图 3，采用混沌方法对 256×256×8 的 Lena 灰度图进行数据传输实验。编码分组上限值为 N=10240，即原始分组的 10 倍，由于分组信道异步传输更复杂，故实验选取分组信道进行测试。

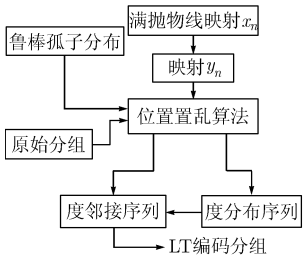

图 3 系统原理框图

由于一帧 Lena 图像为 524288 bit，实验设定原始分组个数 k =1024，则每个分组的数据长度为 512 bit，按照文献[1]中的相关参数设置要求，令 $c=0.1$，δ =0.005，则符合要求的度分布如图 4。

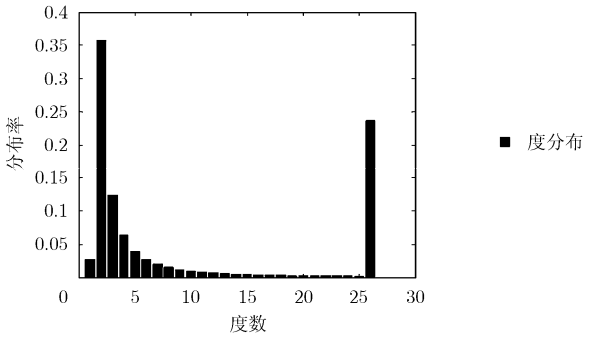

图 4 k =1024，c =0.1，δ =0.005 的度分布

在满足上述参数的条件下，为比较两种方法的性能，分别进行 100 次 Monte Carlo 仿真。传统方法采用以均匀分布为基础的 IS 方法生成度分布数据，然后同样以均匀分布为基础的 IS 方法生成度邻接数据，以这两个数据序列生成编码分组。混沌方法则完全遵循第 3 节的内容，绘出波纹尺寸-迭代次数实验结果如图 5，对于相同码率下两种不同方法的图像恢复效果如图 6。

从图 5 中可以看到，混沌方法迭代译码成功次数集中在 25–30 次区间，译码瀑布区相对集中，而 IS 方法集中在 30–45 次区间，瀑布区相对分散，混沌方法译码成功迭代次数小于 IS 方法。从图 6 可以看出，在译码码率同为 1.45 时，混沌方法已恢复了 Lena 图像，而 IS 方法的恢复图像仍有干扰。

最后还要解决计算精度有限长问题，按照文献[10]中关于有限精度门限值 L_T和序列长度 N的关系，可由下列经验公式来计算：

$$L_T(N)=2(\log_2 N+2) \tag{7}$$

一般情况下计算机的双精度数据为 64 bit，则可实现的混沌序列长度约为 1.0737×10^9，这对本文的仿真实验完全足够，但在涉及序列长度更长的应用时，应考虑有限精度的影响，将每帧数据的设计长度与有限精度门限值结合起来，充分发挥混沌系统性能。

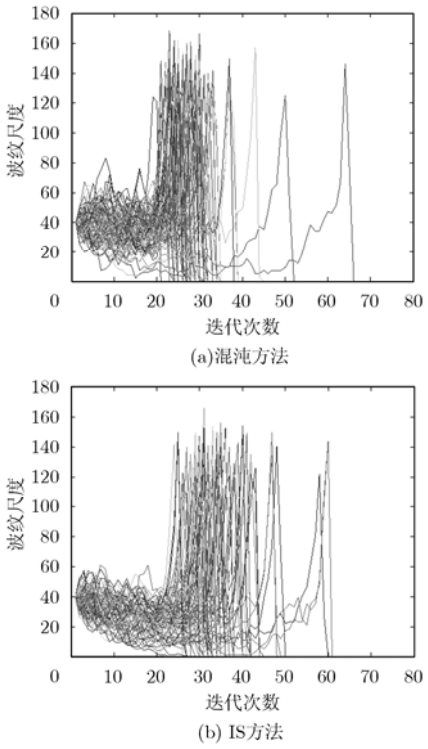

图 5 混沌方法与传统方法的 Monte Carlo 仿真比较

图 6 码率同为 1.45 时 Lena 图像恢复比较

5 结束语

本文提出了基于抛物线映射的混沌 LT 编码方法，给定混沌初始值并代入抛物线映射的迭代公式，通过生成的混沌序列及位置置乱算法，能够生成严

格满足理论设计的 LT 度分布和度邻接关系，通过仿真实验，证实了该方法具有结构简单、开销小、保密性好及高于传统重要抽样方法的性能，最后讨论了数据精度对混沌序列的影响，在 64 bit 的计算机双精度数据下，能够较好地实现符合要求的混沌 LT 编码。

参 考 文 献

[1] Luby M. LT codes[C]. Proceedings of The 43rd Annual IEEE Symposium on Foundations of Computer Science, Vancouver, CA, 2002: 271–282.

[2] Makay D J. Fountain codes[J]. *Proceedings of IEEE, Communications*, 2005, 152(6): 1062–1068.

[3] 林广荣, 林新荣, 依那等. 基于 LDPC 码的数字喷泉编码[J]. 电子与信息学报, 2008, 30(4): 822–825.
Lin Guang-rong, Lin Xin-rong, and Yi Na, *et al*.. Digital fountain based on LDPC code[J]. *Journal of Electronics & Imformation Technology*, 2008, 30(4): 822–825.

[4] 肖东亮, 焦秉立, 林春蕾等. 混沌理论在现代信道编码技术中的应用[J]. 电子学报, 2007, 35(10): 1961–1967.
Xiao D L, Jiao B L, and Lin C L, *et al*.. Application study on chaotic theory for modern channel coding[J]. *Acta Electronica Sinica*, 2007, 35(10): 1961–1967.

[5] Park Dohyung and Chung Sae-Young. Performance—complexity tradeoffs of rateless codes [C]. 2008 IEEE International Symposium on Information Theory, Toronto, CA, 2008, 7: 2056–2060.

[6] Venkiah A, Piantanida P, and Poullia C, *et al*.. Rateless coding for quasi-static fading channels using channel estimation accuracy [C]. 2008 IEEE International Symposium on Information Theory, Toronto, CA, 2008, 7: 2257–2261.

[7] Agarwal S, Hagedorn A, and Trachtenberg A. Adaptive rateless coding under partial information[C]. Information Theory and Applications Workshop, San Diego, USA , 2008, 2: 5–11.

[8] Ming Xiao, Aulin T, and Medard M. Systematic binary deterministic rateless codes[C]. 2008 IEEE International Symposium on Information Theory, Toronto, CA, 2008, 7: 2066–2070.

[9] Tarus H, Bush J, Irvine J, and Dunlop J. Exploiting Redundancies to Improve Performance of LT Decoding [C]. Communication Networks and Services Research Conference 2008 6th Annual, Halifax, CA, 2008, 5: 198–202.

[10] 蔡国权, 宋国文, 于大鹏. Logistic 映射混沌扩频序列的性能分析[J]. 通信学报, 2000, 21(1): 60–63.
Cai G Q, Song G W, and Yu D P. On properties of logistic-map chaotic spread spectrum sequences[J]. *Journal on Communications*, 2000, 21(1): 60–63.

[11] Shokrollahi A. Raptor codes[J]. *IEEE Transactions on Information Theory*, 2006, 52(6): 2551–2567.

黄　诚：　男，1977 年生，博士生，研究方向为信道编码技术、混沌理论和无线通信.

易本顺：　男，1965 年生，教授，博士，博士生导师，研究方向为光纤通信、无线通信、图像处理.