

# 系统 LT 码在删除信道下的渐进性能分析及度分布设计

华 洁<sup>1,2</sup> 徐大专<sup>1,2</sup> 许生凯<sup>1,2</sup>

(1. 南京航空航天大学江苏省物联网和控制技术重点实验室, 南京, 211106; 2. 南京航空航天大学电子信息工程学院, 南京, 211106)

**摘 要:** 首先基于与或树分析法, 对系统 LT 码在删除信道下的渐进性能公式进行推导, 并给出其下限。仿真结果表明当开销足够大时, 实际误码率, 渐进性能与下限三者完美匹配。然后根据渐进性能, 提出改进的优化模型(Improved systematic linear programming, ISLP)对度分布进行优化设计。优化所得的度分布明显优于鲁棒孤波分布(Robust soliton distribution, RSD)分布与截断度分布(Truncated degree distribution, TDD)分布。另外, 优化后的度分布其渐进性能可由设定的开销与误码率进行控制, 即在所设置的开销之内达到理想的误码率, 这一特性可进一步影响完全译码时所需开销。仿真结果表明, 数据恢复时所需的开销与所设置的开销相近。对比系统 LT 码和 LT 码的误码率与恢复原始数据时所需的开销和编译码时间, 表明系统 LT 码能比 LT 码更快地恢复原始数据, 具有更优的性能。

**关键词:** 渐进性能; 度分布; 线性优化; 系统 LT 码

**中图分类号:** TN929.5 **文献标志码:** A

## Asymptotic Performance Analysis and Degree Distribution Design for Systematic Luby Transform Codes over Binary Erasure Channel

Hua Jie<sup>1,2</sup>, Xu Dazhuan<sup>1,2</sup>, Xu Shengkai<sup>1,2</sup>

(1. Jiangsu Key Laboratory of Internet of Things and Control Technologies, Nanjing University of Aeronautics and Astronautics, Nanjing, 211106, China; 2. College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, 211106, China)

**Abstract:** The asymptotic performance formula of systematic LT codes(SLT) in binary erasure channel (BEC) is firstly derived based on AND-OR tree analysis and its lower limit is given. Simulation results show that the actual bit error ratio(BER), the asymptotic performance and the lower limit match each other perfectly when the overhead is large enough. Then the optimization of degree distribution is proposed by the improved systematic linear programming(ISLP) model in accordance with the asymptotic performance. The optimized degree distribution is obviously superior to robust soliton distribution(RSD) and truncated degree distribution(TDD). Furthermore, the asymptotic behavior of the optimized degree distribution can be controlled by the given overhead and BER. In other words, the ideal BER can be obtained within the overhead we want, which also influences the overhead for complete decoding. Simula-

tion results show that the overhead required for data recovery is close to that of we set. The comparison of BER, the overhead required for data recovery and the time of encoding and decoding for LT codes and SLT codes show that SLT codes have better performance than LT codes with more quick speed in data recovery.

**Key words:** asymptotic behavior; degree distribution; linear optimization; systematic LT codes

## 引 言

二进制删除信道(Binary erasure channel, BEC)是典型的删除信道模型<sup>[1]</sup>。在网络传输过程中,数据要么被正确接收要么被丢弃,因此,因特网的传输表现出明显的删除信道特征。为解决数据在传输中的丢包问题,回复重传技术(Automatic repeat request, ARQ)应运而生。但随着用户需求的增加,ARQ所引发的“反馈风暴”也大幅度降低了网络传输效率。针对这一问题,Luby 提出了数字喷泉码的概念,但并未构造出实用的数字喷泉码<sup>[2-5]</sup>。2002年,Luby<sup>[6]</sup>设计 LT 码作为第一种实用的数字喷泉码,并为其设计度分布——鲁棒孤波分布(Robust soliton distribution, RSD),使其能在任意删除信道中逼近信道容量。

非系统是 LT 码的缺点之一,而非系统 LT 码在完成 LT 编码符号之前发送输入符号。从文献[7]已知,系统 LT 码优于非系统码。因此,本文在删除信道下针对系统 LT 码进行分析与度分布设计。Nguyen 等<sup>[7,8]</sup>对系统 LT(Systematic Luby transform, SLT)码进行定义并给出了一种截断度分布(Truncated degree distribution, TDD),该度分布能使系统 LT 码具有与 RSD 分布相近的性能。然而,TDD 分布仍不是最优分布。经过对传统系统 LT 码结构的修改,文献[9]提出了一种新的无码率码,消除了 LT 码原有的错误平底现象。文献[10]中介绍了一种复杂度更低的 LT 码——准系统参杂 LT(Quasi-systematic doped LT, QS-DLT)码。为提高效率,在高斯信道下,文献[11]通过改变 LT 码的编码方式,使得输入度分布不再服从泊松分布。文献[8]则通过分析系统 LT 码的译码特性并对 RSD 分布进行修改而得到 TDD 分布。

综上,现有的 SLT 码的研究,有的通过改变编码方式提高系统 LT 码性能;有的基于对系统 LT 的译码概率分析设计度分布;有的在原有度分布的基础上进行修改,得到新的度分布。而本文将对系统 LT 码的渐进性能进行分析,并以此为依据设计度分布,主要的工作包括:(1)对系统 LT 在删除信道下的性能进行分析。通过与或树分析方法对系统 LT 码的渐进性能进行推导,并给出明确的表达式,同时获取渐进性能的下限。理论数据与仿真结果相符合,推导成立;(2)基于渐进性能与其下限,设计优化模型,并对系统 LT 码的度分布进行优化。结果表明,优化所得的度分布性能优于 RSD 和 TDD。另外,在设置的开销内,可达到理想的误码率。这表明,可以控制度分布的误码率性能并进一步影响恢复数据时的开销。仿真结果表明,成功译码时的开销与所设置的值相近;(3)在不同删除概率下,对系统 LT 码与 LT 码进行比较。结果表明,删除信道下系统 LT 码在恢复数据上所花费的编译码时间小于 LT 码,显示了在快速恢复数据方面优于 LT 码的性能。

## 1 系统 LT 码

文献[7]已经对系统 LT 码进行了详细说明,但为便于进一步的分析与度分布优化,将对系统 LT 码提出一些自己的概念。对普通 LT 码,在使用传播置信算法(Belief propagation, BP)进行译码时,不用考虑删除信道的删除概率,这是因为 LT 码的码字可以任意数量随机产生,只要接收端接收到足够多的码字数就可完全译码<sup>[1,6]</sup>。但对于系统的 LT 码来说,提前发送的与输入符号相同的码字只能生成一次(这部分码字称为系统部分),所以对其要考虑信道的删除概率。因此对于随后发送的普通 LT 码(非系

统部分), 定义另一开销  $\bar{\epsilon}$ ,  $N$  为接收端总共接收到的码字数, 包含系统与非系统两部分, 且  $N = K(1 + \epsilon)$ , 其中  $K$  与  $\epsilon$  的定义同文献[6], 分别为输入符号码长与接收端的总开销。信道的删除概率为  $p$ , 已知其只对系统部分的编码有效, 则在统计意义上, 接收端平均能接收到的系统部分的码字数为  $K(1 - p)$ , 则  $\bar{\epsilon} = \frac{(N - K(1 - p))}{K} - 1, \epsilon = 1 + \bar{\epsilon} - p$ 。

当系统 LT 在删除信道下传输时, 删除概率  $p$  是一个非常重要的参数。当信道条件良好, 即  $p$  非常小时, 可以通过系统部分的码字对原始信息进行恢复, 这大大降低了对度分布设计的必要性, 所以本文主要对具有高删除概率 ( $p \geq 0.5$ ) 的信道进行讨论。

## 2 渐进性能

Luby 使用“与或树”对 LT 码的译码过程进行分析。

假设  $T_l$  为深度为  $2l$  的树, 深度为  $0, 2, \dots, 2l$  的节点为或节点, 对其子节点进行逻辑“或”运算; 深度为  $1, 3, \dots, 2l-1$  的节点为与节点, 对其子节点进行逻辑“与”运算。或节点有  $i$  个子节点的概率记为  $\delta_i, i \in \{1, 2, \dots, O\}$ , 与节点有  $j$  个子节点的概率记为  $\beta_j, j \in \{1, 2, \dots, A\}$ ,  $\{\delta_1, \delta_2, \dots, \delta_O\}$  和  $\{\beta_1, \beta_2, \dots, \beta_A\}$  均为概率分布。树中的每一个节点被记为 0 或者 1。0 表示该节点未被恢复 1 则相反。假设  $a$  表示或节点为 0 的概率,  $b$  表示与节点为 1 的概率, 则有如下定理<sup>[12]</sup>。

**定理 1** 与或树  $T_l$  的根节点为 0 的概率  $y_l = f(y_{l-1})$ , 其中  $y_{l-1}$  表示树  $T_{l-1}$  的根节点为 0 的概率

$$f(x) = a \cdot \delta(1 - b \cdot \beta(1 - x)) \quad (1)$$

式中:  $\delta(x) = \sum_{i=1}^O \delta_i x^i, \beta(x) = \sum_{j=1}^A \beta_j x^j$ 。即在树  $T_l$  上深度为 2 的或节点可以被看作独立树  $T_{l-1}$  的根节点, 因此  $y_l$  可由  $y_{l-1}$  经过函数迭代而来, 依次类推, 最终可以得到定理 1。

系统 LT 码的译码过程也可用与或树表示, 将输入节点记为或节点, 输出节点记为与节点。当输入节点相连的节点中有一个为 1, 则输入节点译为 1; 当输出节点相连的节点中有一个节点为 0, 则输出节点被译为 0。因此, 可以使用定理 1 的结论获取输入节点在  $l$  次迭代后仍未被恢复的概率, 即其为 0 的概率。在系统 LT 码译码过程中,  $\delta(x)$  对应边角度的输入度分布,  $\beta(x)$  为边角度的输出度分布。已知节点角度的输入度  $\Lambda(x)$  分布为泊松分布, 而节点角度的输出度分布为  $\Omega(x)$ 。由文献[6]可知,  $\delta(x) = \frac{\Lambda'(x)}{\Lambda'(1)}, \beta(x) = \frac{\Omega'(x)}{\Omega'(1)}$ 。已知或节点未被恢复的概率为  $p$ , 则  $a$  和  $y_0$  都被设为  $p$ , 由于输出节点(与节点)已知, 则认为  $b$  为 1。将上述变量代入式(1)可推得定理 2。

**定理 2** 在经过  $l$  次迭代后, 系统 LT 的恢复概率为

$$\begin{cases} y_0 = p \\ y_l = p \cdot \exp[-(1 + \bar{\epsilon})\Omega'(1 - y_{l-1})] \end{cases} \quad (2)$$

式中:  $\Omega(x) = \sum_{d=1}^{d_{\max}} \Omega_d x^d$  表示度分布, 当  $k \rightarrow \infty, y = \lim_{l \rightarrow \infty} y_l$ 。通常情况下, 随着  $\bar{\epsilon}$  的增长, 接收端恢复原始信息的可能越大, 因此将对系统 LT 在删除信道下的误码率下限进行推导。

**定理 3** 在删除概率为  $p$  的删除信道下,  $\beta$  为度分布的平均度数, 则有

$$y \geq p e^{-(1+\bar{\epsilon})\beta} \quad (3)$$

**证明:** 在第  $l$  次迭代中, 当  $l \rightarrow \infty, y = y_l = y_{l-1}$  由于  $y_{l-1}$  略大于 0, 且  $\Omega'(1) = \beta$ , 则有

$$y = y_l = p \cdot \exp[-(1 + \bar{\epsilon})\Omega'(1 - y_{l-1})] \geq p \cdot \exp[-(1 + \bar{\epsilon})\Omega'(1)] = p \cdot e^{-\beta(1+\bar{\epsilon})}$$

并对式(3)进行变形有

$$\beta \geq -\frac{\ln(\frac{y}{p})}{1+\epsilon} \tag{4}$$

对定理 2、3 进行仿真验证,取  $\Omega(x)$  为 RSD(0.3, 0.25, 1 000), 在删除概率  $p=0.5$  的删除信道下比较不同码长的系统 LT 的误码率, 结果如图 1 所示。由图 1 可知, 随着码长由 1 000 增长到  $\infty$  (即渐进), 系统 LT 码的实际误码率越来越靠近渐进性能, 渐进性能为应用此度分布所能达到的最佳性能, 在码长较大时能获得。另外, 定理 3 中的下限也绘制在图中, 明显可知, 当开销足够大时, 下限与渐进性能, 实际误码率完美重合。图 2 为取  $\Omega(x)$  为 RSD(0.3, 0.25, 3 000) 时, 在不同删除概率下系统 LT 码和 LT 码的误码率与下限的对比结果。同样, 可得当开销足够大时, 实际的误码率值与所推导的下限相符合; 但当删除概率增大时, 误码率的下限却随之减小。对这种情况, 可作以下分析: 已知  $\epsilon=1+\bar{\epsilon}-p$ , 且由图 1、2 基本可得  $y \approx p \cdot e^{-\beta(1+\bar{\epsilon})} = p \cdot e^{-\beta(\epsilon+p)}$ , 对  $p$  求导有  $\frac{dy}{dp} = (1-\beta p)e^{-(\epsilon+p\beta)}$ , 显然易得  $\frac{dy}{dp} < 0$ , 所以当误码率达到理论下限时, 其值随着删除概率的增大而减小。同时, 表 1 列出了在 3 000 次仿真中, 系统 LT 码与 LT 码在完全恢复数据时所需的开销、编码时间与译码时间的对比。观察表中数据, 虽然恢复数据所需的实际开销随着删除概率的增大而减小, 但其非系统部分的开销却明显增大, 而对于 LT 码, 编译码的时间主要受非系统部分影响, 所以随着删除概率的增大, 编译码所需的时间却大大增加。

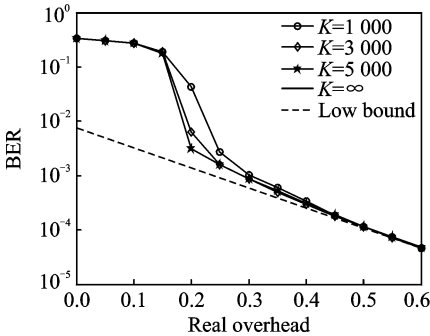


图 1 不同码长的系统 LT 码的误码率

Fig.1 BER performances of different lengths of SLT

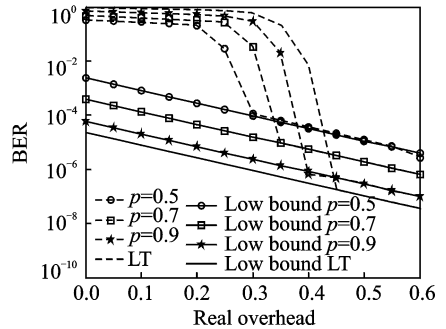


图 2 在不同删除概率下系统 LT 码与 LT 码的误码率

Fig.2 BER performances of SLT and LT under different erasure probabilities

表 1 系统 LT 与 LT 恢复数据时所需的平均开销与编译码时间

Tab.1 Average time and overheads to complete decoding of SLT and LT

编码	$p$	$\epsilon$	$\bar{\epsilon}$	编码时间/s	译码时间/s
系统 LT	0.5	0.335 9	-0.164 1	0.319 1	0.433 3
	0.7	0.286 5	-0.013 5	0.437 3	1.105 3
	0.9	0.274 5	0.174 5	0.451 1	1.204 2
LT		0.277 1	0.277 1	0.548 1	1.321 0

### 3 优化模型

本文对度分布的优化模型进行设计, 以实现: (1) 编码和译码的复杂度都与 Tanner 图中的边数呈线性关系, 因此控制度分布的平均度数能有效降低编译码的复杂度<sup>[12]</sup>; (2) 当误码率为  $y$ , 码长为  $K$  时, 恢复数据的概率  $P_s = (1-y)^K$ , 为了在理想开销时完成数据的恢复 ( $P_s \rightarrow 1$ ), 就必须保证在所设置的开销前,  $y$  值足够小。

基于以上两点, 将模型 LPA<sup>[13]</sup> 向系统 LT 码进行推广, 并对其改进。已知, 在每一次译码迭代

中,都要保证译码过程的方向正确,即  $y_{l+1} \leq y_l$ , 且根据概率分布中每一元素的非负性与和为一的特性,提出系统码的线性优化模型(Systematic linear programming, SLP),模型如式(5)所示。由第 3 节所得的渐进性能与其下限,将式(4)作为新的约束对 SLP 进行改进,得到的模型 ISLP 由式(6)所示。

$$\text{SLP: } \min \sum_{d=1}^{d_{\max}} d\Omega_d$$

$$\begin{cases} \sum_{d=1}^{d_{\max}} d\Omega_d x_i^{d-1} \geq \frac{-\ln(\frac{1-x_i}{p})}{1+\epsilon} & i=1,2,\dots,m \\ \sum_{d=1}^{d_{\max}} \Omega_d = 1 \\ 0 \leq \Omega_d \leq 1 \end{cases} \quad (5)$$

$$\text{ISLP: } \min \sum_{d=1}^{d_{\max}} d\Omega_d$$

$$\begin{cases} \sum_{d=1}^{d_{\max}} d\Omega_d x_i^{d-1} \geq -\frac{\ln(\frac{1-x_i}{p})}{1+\epsilon} & i=1,2,\dots,m \\ \sum_{d=1}^{d_{\max}} d\Omega_d \geq -\frac{\ln(\frac{y}{p})}{1+\epsilon} \\ \sum_{d=1}^{d_{\max}} \Omega_d = 1 \\ 0 \leq \Omega_d \leq 1 \end{cases} \quad (6)$$

式中:  $1-p=x_1 < x_2 < \dots < x_m=1-\delta$  为区间  $[1-p, 1-\delta]$  上的  $m$  个等分点;  $y$  和  $\delta$  均为理想中要达到的误码率,其中  $y < \delta$ ,  $y$  的设置在一定程度上降低了算法的复杂度,  $y$  的取值主要由  $P_s$  决定,为确保恢复数据的概率足够高,在设置  $y$  时将使得  $P_s > 0.5$ ;  $d_{\max}$  为度分布的最大度数。

### 4 仿真结果与分析

为验证优化模型 ISLP 优化所得的度分布能提高系统 LT 码的性能,将优化所得的度分布  $\Omega_1(x)$  与经典的 RSD 和 TDD 进行对比。3 种分布的各项参数如表 2 所示,原始码长  $K=3\ 000$ , 删除概率  $p=0.5$ 。图 3 给出了 3 种度分布的误码率曲线。使用 3 种度分布时,系统 LT 码完全恢复数据时所需的平均开销结果如表 3 所示。由图 3 所示,虽然当 3 种度分布有相近的平均度数时,三者的渐进性能下限也相近,但在“瀑布区”,  $\Omega_1(x)$  比另两种度分布下降得更快,最先到达平底,且由表 3 可知,  $\Omega_1(x)$  恢复数据所需的开销也最小。结合图 3 与表 3,可知当几种度分布有相近的渐进性能下限时,最先下降到此下限的度分布完成原始

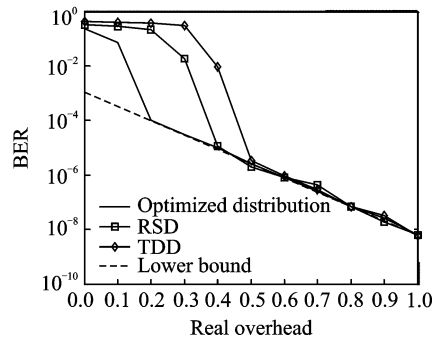


图 3 3 种度分布下系统 LT 码的误码率  
Fig. 3 BER performance of SLT with three degree distributions

表 2 3 种度分布各项参数

Tab. 2 Parameters of three degree distributions

度分布	参数	平均度数
$\Omega_1(x)$	$d_{\max} = 100, p = 0.5, \epsilon = 0.2, \delta = 10, y = 1e-3$	12.167 4
RSD	$K = 3\ 000, c = 0.19, \delta = 0.05$	12.166 1
TDD	$K = 3\ 000, c = 0.48, \delta = 0.08, \gamma = 2$	12.160 7

数据恢复所需的开销最小。

上述仿真表明,使用 ISLP 优化所得的度分布优于 RSD 与 TDD。现在,对 SLP 与 ISLP 进行对比,以验证 ISLD 中所增加的新约束大大地提高了优化效果。两种模型的参数设置为  $d_{\max} = 100, p = 0.8, y = 10^{-4}$ , 另外 SLP 中  $\delta = 10^{-4}$ , ISLP 中  $\delta = 10^{-3}$ 。比较结果如图 4 所示。从图 4 可清楚地看到, ISLP 的优化效果明显优于 SLP。观察图 4 中所标出的 3 个点,在 SLP 中,只能对点 A 进行控制,但在 ISLP 中由于新约束的作用,可以控制点 B 与点 C。点 B 表明在所希望的开销前,即能达到渐进性能的平台处,点 C 则表明在所设置的开销时可取得理想的误码率。与点 A 相比,点 B 与点 C 对恢复数据的概率更具影响意义,从而可以进一步对恢复数据的开销进行控制。表 4 给出了模型 ISLP 优化度分布恢复数据的开销,可知实现数据恢复的实际开销与所设置的开销相近,这表明模型 ISLP 对恢复数据所需的开销有一定的控制作用。

表 3 3 种度分布恢复数据时的开销

Tab. 3 Overheads to restore data of three distributions

$\Omega_1(x)$	RSD	TDD
0.150 3	0.300 9	0.310 0

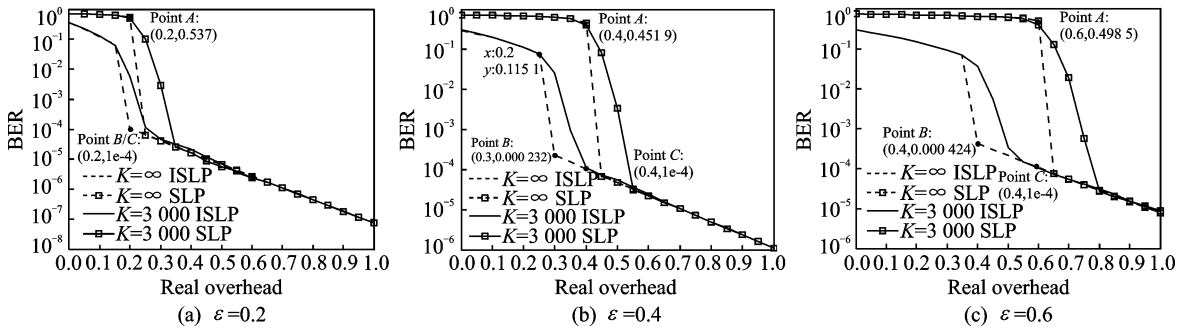


图 4 模型 ISLP 与 SLP 优化度分布的误码率

Fig. 4 BER of the distributions optimized by ISLP and SLP

表 4 ISLP 优化度分布恢复数据时的开销

Tab. 4 Average overheads to complete decoding of the distributions optimized by ISLP

$\epsilon$	数值		
设置	0.2	0.4	0.6
实际	0.198 2	0.409 7	0.610 6

## 5 结束语

在删除信道下,由于系统部分码字的作用,系统 LT 码可以比 LT 码更快地恢复数据。本文主要在大删除概率的删除信道下针对系统 LT 码进行了详细研究。通过与或树分析得到系统 LT 码在删除信道下的渐进性能并得到其下限,且相应的仿真也证明所推得的理论 LT 码渐进性能与实际性能相符。基于以上结论,本文设计了系统 LT 码的度分布优化模型,此模型优化所得的度分布有效地提高了系统 LT 码的误码率性能。仿真结果表明,相比于传统的 RSD 与 TDD,优化所得的度分布能更快地恢复源数据。模型 SLP 与 ISLP 的对比结果表明,新添加的约束可以有效地提高度分布的性能,且在使用 ISLP 模型优化度分布时,系统 LT 码的误码率性能与恢复数据时所需的开销是可控的。

## 参考文献:

[1] Elias P, Elias P. Coding for two noisy channels[C]// The 3rd London Symposium in Information Theory. London:[s. n.], 1956:61-76.  
 [2] Byers J W, Luby M, Mitzenmacher M, et al. A digital fountain approach to reliable distribution of bulk data [C]// Proceed-

ings of the ACM SIGCOMM Conference on Applications, Technologies Architectures, and Protocols of Computer Communication. Vancouver, Canada:ACM,1998,28(4):56-67.

- [3] MacKay J C. Fountain codes [J]. *IEE Proc Comm*, 2005,152(6):1062-1068.
- [4] 黄晓可, 刘洛琨, 郭虹. 一种基于短 LT 码的级联编译码算法[J]. *数据采集与处理*, 2014, 29(3):445-450.  
Huang Xiaoke, Liu Luokun, Guo Hong. A concatenated coding algorithm based on LT codes with small message length[J]. *Journal of Data Acquisition and Processing*, 2014, 29(3): 445-450.
- [5] 徐大专, 邵汉钦, 张小飞, 等. 数字喷泉码及网络喷泉码的最新进展[J]. *数据采集与处理*, 2014, 29(3):351-359.  
Xu Dazhuan, Shao Hanqin, Zhang Xiaofei, et al. Recent research progress on digital fountain codes and network fountain codes [J]. *Journal of Data Acquisition and Processing*, 2014, 29(3): 351-359.
- [6] Luby M. LT codes[C]//The 43rd Annual IEEE Symposium on Foundations of Computer Science. Vancouver, BC, Canada: IEEE, 2002:271-280.
- [7] Nguyen T D, Yang L L, Hanzo L. Systematic Luby transform codes and their soft decoding [J]. *IEEE Workshop on Signal Processing Systems*, 2007,67-72.
- [8] Nguyen T D, Yang L L, Ng S X, et al. An optimal degree distribution design and a conditional random integer generator for the systematic Luby transform coded wireless internet [C]//Wireless Communications and Networking Conference. Las Vegas, NV, USA:IEEE, 2008,243-248.
- [9] Yuan Xiaojun, Li Ping. On systematic LT codes [J]. *IEEE Commun Lett*, 2008,12(9):681-683.
- [10] Yuan Xiaojun, Li Ping. Quasi-systematic doped LT codes [J]. *IEEE J Sel Areas Commun*, 2009,27(6): 866-875.
- [11] Hayajneh K F, Yousefi S. Improved systematic fountain codes in AWGN channel [C]//13th Canadian Workshop on Information Theory. [S. l. ]:IEEE, 2013:148-152.
- [12] Luby M G, Mitzenmacher M, Shokrollahi M A. Analysis of random processes via AND-OR tree evaluations[C]//Proceedings of the 9th Annual ACM-SIAM Symposium on Discrete Algorithms. New York, USA:Society for Industrial & Applied Mathematics,1998:364-373.
- [13] Sejdinovic D, Piechocki R J, Doufexi A. AND-OR tree analysis of distributed LT codes[C]//Networking and Information Theory, IEEE Information Theory Workshop on. Volos, Greece: IEEE, 2009:261-265.

#### 作者简介:



华洁(1991-),女,硕士研究生,研究方向:数字喷泉码、网络编码。



徐大专(1963-),男,教授,博士生导师,研究方向:通信理论与信号处理, E-mail: xudazhuan@nuaa.edu.cn。



许生凯(1990-),男,博士研究生,研究方向:数字喷泉码、网络编码。

