

数字喷泉码度分布优化设计的最新研究进展

徐大专 许生凯 华洁 程凯

(南京航空航天大学电子信息工程学院, 南京, 210016)

摘要: 数字喷泉码是一种无速率码, 能够有效避免大规模网络广播组播传输中的“反馈风暴”问题, 在大数据的网络传输中具有广阔的应用前景。本文论述了当前数字喷泉码的研究现状, 分别介绍了不同信道条件下几种典型的数字喷泉码。度分布是数字喷泉码设计的基础, 本文总结了现有数字喷泉码度分布优化设计方面的研究工作, 重点介绍了不同信道下基于密度演化方法的度分布设计模型, 并分析了存在的不足及改进之处。最后, 对数字喷泉码的设计、应用和发展趋势给出了分析和展望。

关键词: 数字喷泉码; 度分布; 优化设计; 密度演化

中图分类号: TN911.22 **文献标志码:** A

Recent Progress on Optimization Design of Degree Distributions in Digital Fountain Codes

Xu Dazhuan, Xu Shengkai, Hua Jie, Cheng Kai

(College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, China)

Abstract: Digital fountain codes are an effective kind of rateless codes to avoid “feedback storm” in network broadcast/multicast with a wide horizon of development in the network transmission of big data. Therefore, the recent research progress of digital fountain codes is summarized and some typical digital fountain codes are introduced in different channels. Since degree distributions are fundamental to design digital fountain codes, previous work about optimization design of degree distributions is summarized and some linear optimization techniques over different channels based on density evolution are presented in particular. Meanwhile, the problems of optimization design are pointed out and some possible improvements are proposed. Finally, the outlook of optimization design in degree distributions and applications of digital fountain codes is analyzed.

Key words: digital fountain codes; degree distribution; optimization design; density evolution

引言

近年来随着因特网业务的迅猛发展, 用户对音视频等流媒体服务的要求越来越高, 即对数据传输的准确性提出了更高的要求。有限的网络带宽与日益增长的网络数据流量之间的矛盾严重地制约了网络数据业务的发展。传统的分组码由于码率固定的缺陷将不能很好地适应网络广播组播中的不同用户信

道以及信道变化的情况。数字喷泉技术为提高网络数据传播的有效性和可靠性提供了新的思路^[1-2]。

数字喷泉码^[3-6]是针对大规模网络数据分发和可靠性传输而提出的一种新的信道编码方式,其编译码较为简单,易于实现。数字喷泉码利用整数集上的某个概率分布来随机选取度数进行编码,同时对编码符号的个数没有限制,可以根据需要产生任意数量的编码符号。因此数字喷泉码是一种无速率码。接收端只需要接收到足够多的编码符号,而无需关心接收的顺序,即可完成译码。当译码成功后,接收端只需给发送端一个反馈信号,从而避免了分组码带来的“反馈风暴”问题。此外,利用数字喷泉码的上述优点可以实现数据的分布式备份。数字喷泉码具有广阔的应用前景,目前已被 DVB-H 和 3GPP TS 26.346^[7-8]等国际标准采用,并且正在参与其他多项国际标准的制定。数字喷泉码设计的关键是度分布的选择。本文围绕数字喷泉码的基本原理及其度分布的优化设计工作展开论述,以期能够推动对数字喷泉码的进一步研究和讨论,进而推广数字喷泉码在大规模网络信息分发、合作传输、并行下载以及在大数据分布式备份等领域的实际应用。

1 数字喷泉码

1.1 LT 码

Byers 等人于 1998 年首次提出了将数字喷泉码的概念应用于大批数据的分发,但并未给出具体的码型设计^[3]。2002 年, Luby 给出不同于 RS 码及 Tornado 码的信道编码方式,称之为 Luby transform (LT) 码^[4],它是第一种真正意义上的随机编码,其编译码方法简单,复杂度相对较低,因此也常用作于数字喷泉码的理论分析。LT 码可以由原始数据符号长度 K 和度分布函数 $\Omega(x)$ 来确定,记为 $LT(K, \Omega(x))$ 。其中,度分布是定义在一个整数集 $\{1, 2, \dots, d_{\max}\}$ 上的概率分布 $\{\Omega_1, \Omega_2, \dots, \Omega_{d_{\max}}\}$ 。通常,度分布可以用生成多项式的形式进行表示,即 $\Omega(x) = \sum_{d=1}^{d_{\max}} \Omega_d x^d$ 。下面给出 LT 码编码的过程。(1) 根据度分布 $\Omega(x)$ 随机选取一个度数 d ; (2) 从 K 个原始数据符号中均匀随机地选取 d 个不相同的符号 $v_{i1}, v_{i2}, \dots, v_{id}$; (3) 将选取出的 d 个原始符号进行 $GF(2)$ 上的加法运算,得到一个编码符号,即

$$c_i = v_{i1} \oplus v_{i2} \oplus \dots \oplus v_{id} \quad (1)$$

重复上述步骤(1~3),即可以产生任意长度的 LT 码。原始符号和编码符号的关系可以用如图 1 所示的二分图或者 Tanner 图^[9]来表示。顶层的 K 个节点表示原始符号,称之为输入节点;底层的 N 个节点表示编码符号,称之为输出节点。每个输出节点所连接的输入节点的个数即是该节点的度,当输入和输出节点个数趋近无穷时,输出节点的度分布即可用 $\Omega(x)$ 来表示,故而也称作输出度分布。

LT 码的性能与译码方法密切相关^[6],目前译码方法主要分为置信传播(Belief propagation, BP)算法和高斯消除(Gaussian elimination, GE)算法。BP 译码算法是一种低复杂度的迭代译码算法,接收端根据收到的编码符号重构如图 1 所示的二分图,具体的译码过程如下^[4]:

(1) 在二分图中寻找度为 1 的输出节点 c_i , 并直接恢复与此相连的输入节点 v_j , 即 $v_j = c_i$; 如果找不到度为 1 的输出节点则此次译码失败;

(2) 将已经恢复的输入节点 v_j 和与之相连的所有输出节点的值分别进行异或,并将相应连接的边删除以更新二分图;

(3) 重复以上步骤,直到找不到度为 1 的输出节点时,译码停止。

需要指出的是,当译码停止时,如果并未恢复全部输入节点的值,则此次译码失败。数字喷泉码是一种无速率码,因此如果译码失败,接收端将接收更多的编码符号来继续译码直到译码成功。故而在数字喷泉码的研究中,定义一个接收端收到的编码符号数量与原始符号数量之间比值关系的物理量,称之为开销,即 $\epsilon = N/K - 1$ 。显然,开销越大,成功译码的可能性越高。通过对译码过程的分析可以看出,良好的度分布不仅可以提供度为 1 的节点开始译码过程,而且还能不断产生新的度为 1 节点使得译码

过程得以继续直至成功恢复原始符号。

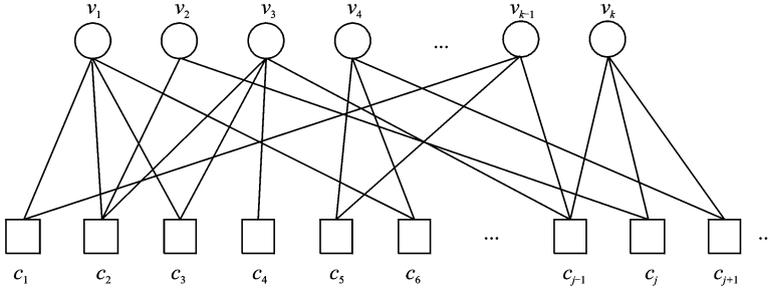


图1 LT码的二分图

Fig. 1 Bipartite graph of LT codes

GE译码类似于线性方程组求解的过程,虽然较BP译码有更好的误码率性能,但是译码复杂度较高,只适用于短码长喷泉码。国内外学者对GE译码算法进行了改进,减小了译码复杂度。Saejoon等^[10]提出了增量高斯删除译码算法,Bioglio等^[11]则提出运行中高斯删除译码算法。朱宏鹏等^[12]和朱宏杰等^[13]先后从混合译码的角度出发,将BP译码和GE译码相结合,先进行BP译码,如果译码失败再对残余的二分图进行GE译码,从而提高译码效率,减小了平均译码代价。

为了研究删除信道下LT码的渐进性能并为后续设计提供基础,Luby等^[14]提出与或树(And-Or Tree)分析方法。当采用BP译码时,经过 l 次迭代后,LT($K, \Omega(x)$)输入节点未正确恢复的概率由下面的递推公式确定

$$\begin{aligned} y_0 &= 1 \\ y_l &= \exp(-\alpha \omega(1 - y_{l-1})) \end{aligned} \quad (2)$$

式中: α 为图1所示二分图中输入节点的平均度数, $\omega(x)$ 为输出边分布且满足 $\omega(x) = \Omega'(x)/\Omega'(1)$ 。针对LT码的有限长分析,Karp等^[15]和Maneva等^[16]提出了计算有限长LT码误码率的方法,但是复杂度均较高,无法实现与仿真结果的比较。

1.2 Raptor码

虽然LT码构造简单且便于分析,但是由于编译码复杂度并非和码长呈线性关系且存在错误平底现象,因此LT码并不适合于长码。Shokrollahi^[5]提出了一种扩展LT码,其编译码复杂度与码长呈线性关系,称之为Raptor码。Raptor码是一种级联码,由纠错码,如低密度校验(Low-density parity-check, LDPC)码作为外码或者预编码,由平均度数较小的弱化LT码作为内码,其编码的二分图如图2所示。译码同编码相对,通常也分为两步,首先采用BP译码恢复出中间节点,再利用传统纠错码译码方法恢复输入节点信息。由于预编码具备一定的纠错能力,因此可以降低对LT码的性能要求。Shokrollahi在文献^[5]中给出了弱化的LT度分布。

1.3 无线信道数字喷泉码

数字喷泉码起初是针对二进制删除信道而提出的一种前向纠错码。然而,已有成果表明,将数字喷泉码推广到噪声信道中仍然能够提供较好的性能。这里所说的噪声信道包括加性高斯白噪声(Additive white Gaussian noise, AWGN)信道和衰落信道。Palanki等^[17]于2004年指出,同删除信道下的情况类似,LT码在噪声信道中存在明显的错误平底现象,而Raptor码则不存在明显的错误平底。Etesami等^[18]分析得出在删除信道中好数字喷泉码在AWGN信道下仍具有优异的性能表现,同时利用外信息转移(Extrinsic information transfer, EXIT)图方法对LT码进行了优化。Niesen等^[19]对高斯多址接入

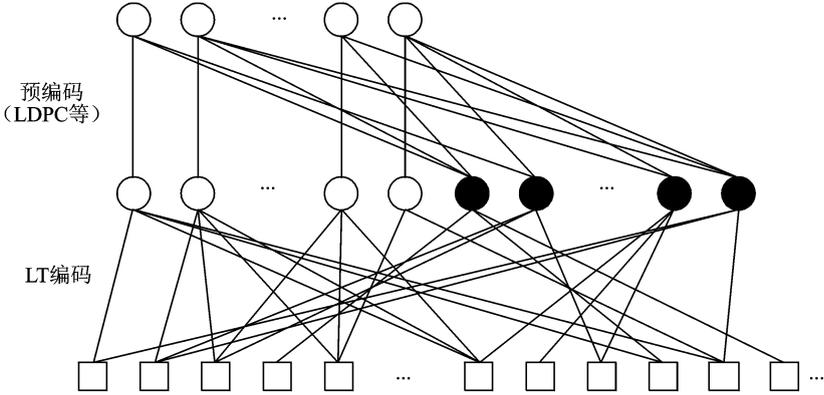


图 2 Raptor 码二分图

Fig. 2 Bipartite graph of Raptor codes

信道下的喷泉码性能进行了研究,提出了一种喷泉多址接入方案。Castura 等^[20]研究了衰落信道下数字喷泉码的传输方案,在接收端不知道信道状态信息时能较固定码率编码方案提供更好的性能。Venkiah 等^[21]对 Raptor 码的渐进性能进行了分析,并提供了度分布优化设计的方案。此外,研究人员针对噪声信道设计了一些无速率编译码方案。Erez 等^[22]设计了一种无速率码的编译码方案,能够逼近 AWGN 信道的信道容量,并且具有较低的译码复杂度。Bai 等^[23]提出了一种能够有效降低 AWGN 信道和衰落信道中错误平底的无速率码。Chen 等^[24]提出了一种适用于 AWGN 信道的非级联形式的喷泉 AR 码,有效地解决了 LT 码在 AWGN 信道下的错误平底问题。Kharel 等^[25]利用 Gauss-Jordan 消元思想提出了一种适用于 AWGN 信道下的短码长 LT 码译码算法。

在噪声信道中数字喷泉码的二分图和删除信道中一致,即如图 1 和图 2 所示。所不同的是,在噪声信道下,译码时输入与输出节点相互之间传递的是软译码信息。本文以 AWGN 信道中 LT 码为例进行阐述^[18]。在 LLR-BP 译码中,令 m_{ia}^l 表示第 l 次迭代时由输出节点 o 传递给输入节点 i 的对数似然比信息, m_{io}^l 表示第 l 次迭代时由输入节点 i 传递给输出节点 o 的对数似然比信息, Z_o 为每个输出节点对应的似然比信息初值,则第 l 次迭代时输出节点的似然比信息更新公式为

$$m_{ia}^l = 2 \operatorname{atanh} \left(\tanh \left(\frac{Z_o}{2} \right) \prod_{\substack{j \in S\{o\} \\ j \neq i}} m_{jo}^l \right) \quad (3)$$

输入节点似然比信息更新公式为

$$m_{io}^{l+1} = \sum_{p \in S\{i\}} m_{pi}^l \quad (4)$$

式中: $l=0$ 时的初值为 $m_{io}^0 = 0$, $\tanh(\cdot)$ 表示正切函数, $\operatorname{atanh}(\cdot)$ 表示反正切函数, $S\{o\}$ 表示与节点 o 相连的节点集合。

AWGN 信道下的 LT 码渐进性能分析可以利用高斯近似密度演化方法^[26]。高斯近似指的是将一次迭代中输入与输出节点之间传递的似然比信息当作是服从对称高斯分布的随机变量,因此只需要用随机变量的均值来衡量。令 μ^l 表示第 l 次迭代时输入节点传递给输出节点的似然比信息均值,定义^[18]

$$f_d(\mu^l) = 2E \left[\operatorname{atanh} \left(\tanh \left(\frac{Z_o}{2} \right) \prod_{i=1}^{d-1} \tanh \left(\frac{X_i}{2} \right) \right) \right] \quad (5)$$

式中 X_i 是满足均值为 μ^l 的相互独立的对称高斯随机变量。因此 $f_d(\mu^l)$ 计算的是度为 d 的输出节点给对应输入节点传递的似然比信息的均值,故而,第 $l+1$ 次迭代时输入节点似然比信息均值的更新公式

为

$$\mu^{l+1} = \alpha \sum_d \omega_d f_d(\mu^l) \quad (6)$$

1.4 广义分布式网络喷泉码

数字喷泉码和网络编码^[27-31]的结合能够有效地提高网络的传输效率,这也是其受到越来越多研究人员关注的主要原因。在网络中,根据信源的情况可以分为同构网络和异构网络两种模型。其中,各信源具有相同的度分布或者相同的接入信道容量的多信源接入网络称之为同构网络;与之相反的称为异构网络。可以看出,同构网络是异构网络的一种特例。因此,本节主要介绍异构网络中广义分布式喷泉码的度分布优化设计模型^[32-33]。图3给出分布式异构网络的一般模型。该分布式异构网络中含有 t 个独立信源,一个中继和一个目标节点。在该模型中,信源 S_i 的进行数字喷泉编码采用的度分布用 $\Phi_i(x_i)$ 来表示,中继节点的度分布则用 $\Gamma(x)$ 表示, q_i 表示信源 S_i 接入中继的概率。

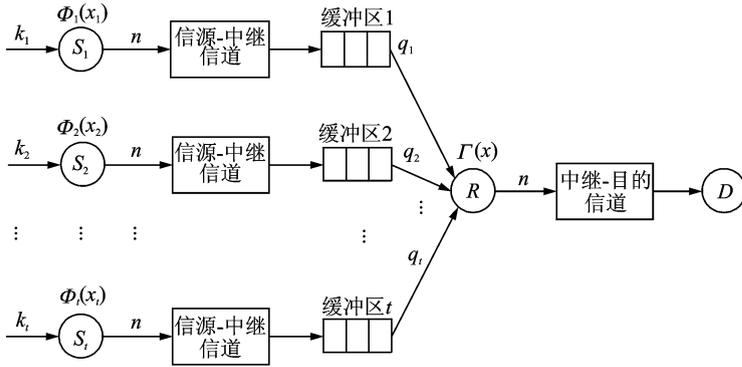


图3 分布式异构网络模型

Fig. 3 Distributed heterogeneous network model

异构网络中不同信源的译码失败概率是不同的,因此在这种模型中网络喷泉码的度分布函数将不能使用形如点对点通信中数字喷泉码的一元度分布函数来表示。文献[32]中提出了异构网络中多元度分布函数的概念,即

$$\Omega(x) = \Omega(x_1, x_2, \dots, x_t) = \sum_{l=1}^t \Gamma_l \left(\sum_{j=1}^l q_j \Phi_j(x_j) \right)^l = \Gamma \left(\sum_{j=1}^t q_j \Phi_j(x_j) \right) \quad (7)$$

当各信源的输出度分布和接入信道的概率均相同时,式(7)将简化成 $\Omega(x) = \Gamma(\Phi(x))$,也就意味着异构网络模型将退化成同构网络模型。在异构信源或者异构信道的情况下,不同信源的输出边分布可以由公式中的多元度分布函数的一组偏导数来定义。如图4所示的异构网络中输出边分布函数为

$$\omega_i(x) = \frac{\partial \Omega(x)}{\partial x_i} / \frac{\partial \Omega(x)}{\partial x_i} \Big|_{x=1} = \varphi_i(x_i) \gamma \left(\sum_{j=1}^t q_j \Phi_j(x_j) \right) \quad (8)$$

式中: $\varphi_i(x_i) = \Phi'_i(x_i) / \Phi'_i(1)$ 表示各信源的边分布函数, $\gamma(x) = \Gamma'(x) / \Gamma'(1)$ 表示中继的边分布函数。

基于上述分析可以得出异构网络模型中的多元密度演化方程组

$$y_{i,l} = \begin{cases} 1 & l=0 \\ \exp \left[-\alpha_i \varphi_i(1 - y_{i,l-1}) \gamma \left(\sum_{j=1}^t q_j \Phi_j(1 - y_{j,l-1}) \right) \right] & l \geq 1 \end{cases} \quad (9)$$

式中: α_i 是信源 S_i 的平均输入度, l 表示迭代的次数。各个信源的BP译码比特错误概率可以通过 $y_i = \lim_{l \rightarrow \infty} y_{i,l}$ 进行计算。

2 度分布

2.1 ISD 分布和 RSD 分布

LT 码的译码性能和二分图的结构有关,而二分图又是由度分布随机生成的,因此 LT 码的性能主要是由输出度分布决定的。Luby 在提出 LT 码的设计时给出了两种常见的度分布形式^[4],分别是理想孤波分布(Ideal soliton distribution, ISD)和在 ISD 分布基础上改进的鲁棒孤波分布(Robust soliton distribution, RSD)。关于两种度分布的定义如下

ISD 分布

$$\rho(i) = \begin{cases} 1/K & i=1 \\ 1/i(i-1) & i=2,3,\dots,K \end{cases} \quad (10)$$

RSD 分布

$$\mu(i) = (\rho(i) + \tau(i)) / \beta \quad (11)$$

其中, $\beta = \sum_{i=1}^K (\rho(i) + \tau(i))$, $\tau(i)$ 的定义如下

$$\tau(i) = \begin{cases} R/iK & 1 \leq i \leq \lceil K/R \rceil - 1 \\ R \ln(R/\delta) / K & i = \lceil K/R \rceil \\ 0 & \lceil K/R \rceil + 1 \leq i \leq K \end{cases} \quad (12)$$

式中: R 和 δ 是常数, $R \geq 1$, $\delta \in [0, 1]$ 且 $\lceil \cdot \rceil$ 表示向下取整。RSD 分布能够为 LT 码提供较优的比特误码率(Bit error ratio, BER)性能,但其仍不是最优。

2.2 删除信道度分布优化模型

Sejdinovic 等^[34]利用式(2)中的递推公式给出了删除信道下两种等价的度分布线性优化方案。预先设定一个期望的译码失败概率 ξ , x_i ($i=1, 2, \dots, m$) 是区间 $[0, 1-\xi]$ 上的 m 等分点。下面给出两种线性优化的具体模型,以式(2)中每次迭代误码率降低作为约束条件。

(a) 以最小复杂度为目标

$$\begin{aligned} \text{LPA: } & \min_{\Omega} \sum_{d=1}^{d_{\max}} d \Omega_d \\ \text{s. t. } & \begin{cases} \sum_{d=1}^{d_{\max}} d \Omega_d x_i^{d-1} \geq -\frac{\ln(1-x_i)}{1+\epsilon}, & i=1, 2, \dots, m \\ \sum_{d=1}^{d_{\max}} \Omega_d = 1 \end{cases} \end{aligned} \quad (13)$$

(b) 以最小开销为目标

$$\begin{aligned} \text{LPB: } & \min_{\omega} \sum_{d=1}^{d_{\max}} \frac{\omega_d}{d} \\ \text{s. t. } & \begin{cases} \alpha \sum_{d=1}^{d_{\max}} \omega_d x_i^{d-1} \geq -\ln(1-x_i), & i=1, 2, \dots, m \\ \sum_{d=1}^{d_{\max}} \omega_d = 1 \end{cases} \end{aligned} \quad (14)$$

LPA(LP B)算法能在一定程度上对 LT 码的度分布进行优化,图 4 给出了 LPA 算法优化度分布和 RSD 的比较结果。LPA 优化的参数设置为 $\xi=10^{-3}$, $d_{\max}=100$, $\epsilon=0.2$,可以看出经过 LPA 优化的度分布较 RSD 分布能够提供更好的 BER 性能。但是在预先设定的期望译码失败概率 ξ 减小时,为保证优化

的结果有效(即度为一的概率存在),就必须增大对应所取的 m 等分点的个数,这大大增加了算法的复杂度,且在 ξ 足够小时,可能出现有效解不存在的情况。为解决这些问题,本文对 LPA 算法进行改进,提出一种新的优化方法 LPC。LPA 算法无有效解的主要原因在于当译码失败概率 ξ 很小时所取的 m 值不足以与之匹配,为了兼顾性能和复杂度,本文在 LPC 中设置两个译码失败概率 ξ 和 y 。图 5 给出 LPA 和 LPC 两种 LT 码优化模型所得度分布的 BER 性能比较,参数设置同图 4 中 LPA 优化模型。

$$\begin{aligned}
 \text{LPC: } & \min_{\Omega} \sum_{d=1}^{d_{\max}} d\Omega_d \\
 \text{s. t. } & \begin{cases} \sum_{d=1}^{d_{\max}} d\Omega_d x_i^{d-1} \geq \frac{-\ln(1-x_i)}{1+\epsilon} & i \in 1, 2, \dots, m \\ \sum_{d=1}^{d_{\max}} \Omega_d d \geq -\frac{\ln y}{1+\epsilon} \\ \sum_{d=1}^{d_{\max}} \Omega_d = 1 \end{cases}
 \end{aligned} \tag{15}$$

式中: $x_i (i=1, 2, \dots, m)$ 是区间 $[0, 1-\xi]$ 上的 m 等分点, $\xi > y$ 。

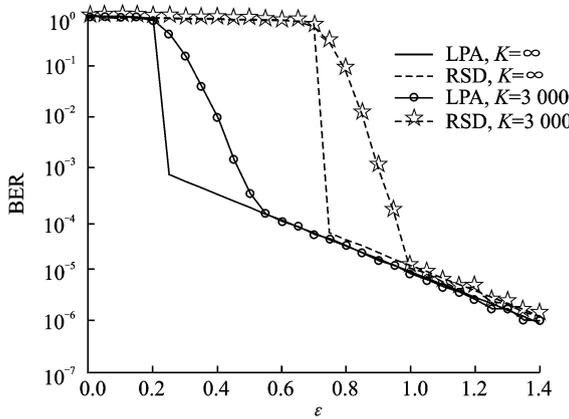


图 4 LPA 优化度分布与 RSD 分布的比特无码率性能比较

Fig. 4 Comparison of BER performances between the degree distribution of LPA and RSD

现有数字喷泉码度分布优化工作鲜有涉及系统形式的 gLT 码,而从纠删码已有结论容易知道,系统形式的码字能比非系统形式提供更好的性能。Nguyen 等^[35-36]提出了系统 LT(Systematic luby transform, SLT)码的概念并给出了一种截断度分布(Truncated degree distribution, TDD),该度分布能够在 SLT 码中提供与 RSD 分布相近的性能。然而,TDD 分布并不是最优的分布。需要注意的是,SLT 码的度分布仍然指的是用来生成 LT 编码符号的度分布。

利用与或树分析方法,可以得到删除信道下 SLTg 码的渐进性能递推公式为

$$\begin{aligned}
 y_0 &= p \\
 y_i &= p \cdot \exp[-\alpha\omega(1-y_{i-1})]
 \end{aligned} \tag{16}$$

式中 p 是删除信道的删除概率。同样,需要预先设定一个期望的译码失败概率 $\xi, z_i (i=1, 2, \dots, m)$ 是区间 $[1-p, 1-\xi]$ 上的 m 等分点。根据式(16)可以给出 SLT 码在删除信道下的两种等价度分布线性优化模型

$$\text{LPD: } \min_{\Omega} \sum_{d=1}^{d_{\max}} d\Omega_d$$

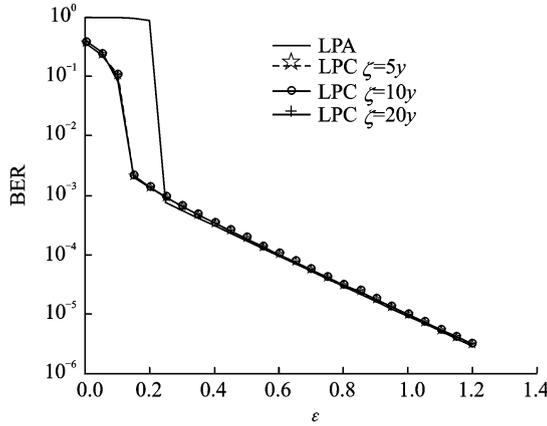


图 5 LPA 和 LPC 优化度分布的比特误码率性能比较

Fig. 5 Comparison of BER performances between degree distributions of LPA and LPC

$$\text{s. t. } \begin{cases} \sum_{d=1}^{d_{\max}} d \Omega_d x_i^{d-1} \geq -\ln \frac{(1-z_i)}{p(1+\epsilon)}, i=1, 2, \dots, m \\ \sum_{d=1}^{d_{\max}} \Omega_d = 1 \end{cases} \quad (17)$$

$$\text{LPE: } \min_{\omega} \alpha \sum_{d=1}^{d_{\max}} \frac{\omega_d}{d}$$

$$\text{s. t. } \begin{cases} \alpha \sum_{d=1}^{d_{\max}} \omega_d x_i^{d-1} \geq -\ln \frac{1-z_i}{p}, i=1, 2, \dots, m \\ \sum_{d=1}^{d_{\max}} \omega_d = 1 \end{cases} \quad (18)$$

图 6 给出 SLT 码通过 LPE 优化模型所得度分布同 RSD 分布以及 TDD 分布的 BER 性能比较。LPE 优化参数设置为 $d_{\max}=100, \epsilon=10^{-3}$ 。

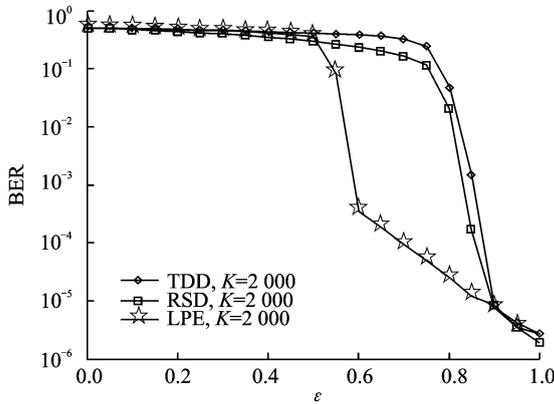


图 6 SLT 码中 LPE 优化度分布同 TDD 以及 RSD 的 BER 性能比较

Fig. 6 Comparison of BER performances among the degree distribution of LPE, TDD and RSD in SLT codes

除了上述较为经典的度分布线性优化模型外,国内外的学者对已有度分布也做出了一些改进。朱宏鹏等^[37]指出了LT码的最优度分布在实际应用中存在的问题,提出了一种实用的次优度分布。王新泽^[38]提出了基于二进制指数分布和RSD分布的开关度分布并给出了相应的改进方案。Zhu等^[39]和焦健等^[40]则致力于短码长LT码的度分布设计。Meng等^[41]研究了严重删除信道中的度分布设计问题,提出了一种以最大化平均恢复概率为目标的优化方法。Hyytiä等^[42-43]使用重要参数抽样理论构造参数迭代算法优化LT码度分布,并在短码长情况下,基于译码过程表现的马尔可夫性质提出优化方法。Hussain等^[44]对传统的随机编码方案进行修改,采用规则的输入度分布进行LT编码并能较好地降低错误平底。Hayajneh^[45-46]等提出编码时选取度数最高的输入节点的方案,使得输入度分布偏离泊松分布。Tsai等^[47]提出了一种新的密度演化方程用以分析LT码的性能并进行度分布优化设计。Wang等^[48]指出当译码器预先知道某些原始符号时能够较好地改善LT码的恢复概率,同时设计出满足这种框架的度分布。

2.3 无线信道度分布优化模型

Etesami等^[18]利用式(5,6)提出了噪声信道下数字喷泉码的度分布优化模型。预先设定一个输入节点似然比信息的均值 $\mu_0, \mu_i (i=1, 2, \dots, n)$ 是区间 $(0, \mu_0]$ 上的 n 等分点。具体的线性优化模型如下。

$$\begin{aligned} \text{LPF: } \min_{\omega} \alpha \sum_{d=1}^{d_{\max}} \frac{\omega_d}{d} \\ \text{s. t. } \begin{cases} \alpha \sum_{d=1}^{d_{\max}} \omega_d f_d(\mu_i) > \mu_i, & i=1, 2, \dots, n \\ \sum_{d=1}^{d_{\max}} \omega_d = 1 \end{cases} \end{aligned} \quad (19)$$

AWGN信道下,本文仍然利用高斯近似密度演化的方法来更新SLT码中输入、输出节点的似然比均值。令 $\mu_{R,j}^{(l)}$ 表示第 l 次迭代、度为 j 的输入节点传递给输出节点的似然比信息的均值, $\mu_{Q,i}^{(l)}$ 表示第 l 次迭代、度为 i 的输出节点传递给输入节点似然比信息的均值,则

$$\begin{aligned} \mu_{R,j}^{(l)} &= \varphi^{-1} \left(1 - \left(1 - \varphi \left(\frac{2}{\sigma_n^2} \right) \right) \left(1 - \sum_{i=1}^{d_i} \lambda_i \varphi(\mu_{Q,i}^{(l-1)}) \right)^{j-1} \right) \\ \mu_{Q,i}^{(l)} &= \frac{2}{\sigma_n^2} + (i-1) \mu_{R,i}^{(l)} \end{aligned} \quad (20)$$

式中: d_i 是输入节点的最大度数, σ_n^2 是信道中加性高斯白噪声的方差, $\mu_{R,i}^{(l)} = \sum_{j=1}^{d_{\max}} \omega_j \mu_{R,j}^{(l)}$, $\varphi(\cdot)$ 函数满足

$$\varphi(x) = \begin{cases} e^{-\langle 0.4527x^{0.88} + 0.0218 \rangle} & 0 < x < 10 \\ \sqrt{\frac{\pi}{x}} \left(1 - \frac{10}{7x} \right) e^{-\frac{x}{7}} & x \geq 10 \end{cases}$$

据此,可以给出AWGN信道下SLT码的度分布线性优化模型

$$\begin{aligned} \text{LPG: } \min_{\omega} \alpha \sum_{d=1}^{d_{\max}} \frac{\omega_d}{d} \\ \text{s. t. } \begin{cases} \alpha \sum_{d=1}^{d_{\max}} \omega_d f_d(\mu_i) + \frac{2}{\sigma_n^2} > \mu_i & i=1, 2, \dots, n \\ \sum_{d=1}^{d_{\max}} \omega_d = 1 \end{cases} \end{aligned} \quad (21)$$

式中 $f_d(\mu_i) = \varphi^{-1} \left(1 - \left(1 - \varphi \left(\frac{2}{\sigma_n^2} \right) \right) \left(1 - \varphi(\mu_i) \right)^{d-1} \right)$ 。图7给出AWGN信道下SLT码通过LPG优化模型所得度分布同TDD的BER性能比较。

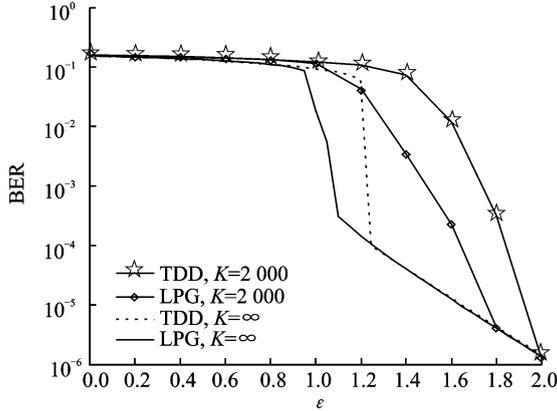


图7 AWGN信道下SLT码中LPG优化度分布同TDD比特误码率性能比较

Fig. 7 Comparison of BER performances between the distribution of LPE and TDD in SLT codes over AWGN channel

2.4 网络喷泉码度分布优化模型

与普通LT码类似,异构网络中广义分布式喷泉码的性能主要由其信源和中继度分布决定。以多元密度演化方程组作为主要约束条件,可以给出异构网络的两步联合优化模型^[32]。

(a) 中继度分布优化模型

$$\begin{aligned} & \min_{\gamma} \sum_{d=1}^t \frac{\gamma_d}{d} \\ \text{s. t. } & \begin{cases} -\sum_{d=1}^t \gamma_d \left(\sum_{j=1}^t q_j \Phi_j(z_{j,n}) \right)^{d-1} - \frac{1}{t} \sum_{j=1}^t \frac{\ln(1-z_{j,n})}{\Phi'_j(z_{j,n})} \leq 0, & n=1, 2, \dots, m \\ \gamma_d \geq 0, & d=1, 2, \dots, t \end{cases} \end{aligned} \quad (22)$$

$$\Gamma(x) = \frac{\int_0^x \gamma(u) du}{\int_0^1 \gamma(u) du} \quad (23)$$

(b) 信源度分布优化模型

$$\begin{aligned} & \min_{\Phi} \frac{\alpha_i k_i}{q_i \Phi'_i(1) \Gamma'(1) \sum_{j=1}^t k_j} \\ \text{s. t. } & \begin{cases} \frac{\Phi'_i(z_{i,n}) \Gamma' \left(\sum_{j=1}^t q_j \Phi_j(z_{j,n}) \right)}{\Phi'_i(1) \Gamma'(1)} - \frac{\ln(1-z_{i,n})}{\alpha_i} \leq 0, & n=1, 2, \dots, m \\ \sum_{d=1}^{d_{i,\max}} \Phi_{i,d} = 1 \\ 0 \leq \Phi_{i,d} \leq 1, & d=1, 2, \dots, d_{i,\max} \end{cases} \end{aligned} \quad (24)$$

两步联合优化算法主要包括两个步骤:首先信源的度分布采用文献[49]中的DSD,DDSD,TDSD等分布作为初始值,根据式(22)中的线性优化方程优化中继的输出度分布 $\Gamma(x)$;然后根据式(24)中的非线性优化方程来进一步优化信源的输出度分布 $\Phi_i(x_i)$ 。反复进行这两个优化步骤,可以不断地优化中继和信源的输出度分布,故而此过程称为迭代优化过程。

迭代优化过程虽然可以对异构网络的度分布进行整体优化,但是复杂度较高,而且信源初始度分布的选取会在很大程度上影响中继度分布优化的结果。虽然前文给出了信源初始度分布选取的建议,但不能保证中继得到最优的度分布。考虑两信源单中继的特殊情况,中继输出度分布 $\Gamma(x)$ 的最大度为 2,而信源为了考虑对原始信息符号的完整覆盖其输出度分布 $\Phi_i(x_i)$ 的最大度数一般远远大于 2,也就意味着 $\Phi_i(x_i)$ 中待优化的概率系数的数目要远多于 $\Gamma(x)$ 。在这种情况下,中继的度分布可以表示为 $\Gamma(x) = px + (1-p)x^2$ 且仅仅由一个参数 p 控制。因此,可以固定中继的度分布,直接优化信源的度分布如非线性优化方程所示。

$$\begin{aligned} & \min_{\Phi} \frac{\alpha_i k_i}{q_i(2-p)\Phi'_i(1)\sum_{j=1}^t k_j} \\ \text{s. t. } & \begin{cases} \frac{\Phi'_i(u_{i,n})\left[p+2(1-p)\sum_{j=1}^t q_j \Phi_j(u_{i,n})\right]}{\Phi'_i(1)(2-p)} - \frac{\ln(1-u_{i,n})}{\alpha_i} \leq 0, & n=1, \dots, m \\ \sum_d \Phi_{i,d} = 1 \\ 0 \leq \Phi_{i,d} \leq 1, & d=1, 2, \dots, d_{i,\max} \end{cases} \end{aligned} \quad (25)$$

3 结束语

从数字喷泉码的概念于 1998 年被首次提出至今,关于数字喷泉码的理论研究取得了丰硕的成果。从简单的二进制删除信道、无线噪声信道到复杂的网络模型,数字喷泉码的应用环境得到了很大的扩展,而且,数字喷泉码在不同的信道中均能够提供较好的性能,这也是其发展受到重视的主要原因。度分布在数字喷泉码的编译码过程中起到重要的作用,因此选择良好的度分布是设计的基础。目前对度分布的设计成果已经较为系统,对不同信道中的情况均有讨论研究,但是存在一些不足和欠缺,还有许多理论和技术问题亟待解决:(1)目前关于度分布的优化,主要是建立在密度演化的渐进性能分析基础上的,优化的结果并不能很好地适应有限长的码字。如何针对有限长数字喷泉码进行度分布优化设计,需要进行进一步研究;(2)目前存在的度分布线性优化方案虽然复杂度较低,但是存在无有效优化结果的弊端。如何在保持低复杂度的同时,提供可行的优化方案值得进一步研究,这将有效地丰富数字喷泉码的度分布优化设计体系;(3)目前有关度分布优化设计还主要局限于非系统形式的喷泉码,对系统形式的数字喷泉码进行研究更具有应用价值。本文提供了系统喷泉码度分布优化设计的思路,以期能够推动数字喷泉码的进一步研究和应用。度分布设计是数字喷泉码的基础研究,目前正处于不断完善的阶段。随着研究的进一步深入以及数字喷泉码研究体系的不断完善,研究成果必将推动数字喷泉码在其他领域的应用和发展。

参考文献:

- [1] Mitzenmacher M. Digital fountains: A survey and look forward[C]// 2004 IEEE Information Theory Workshop. San Antonio, USA: IEEE, 2004:271-276.
- [2] 慕建君,焦晓鹏,曹训志. 数字喷泉码及其应用的研究进展与展望[J]. 电子学报,2009,37(7):1571-1577.
Mu Jianjun, Jiao Xiaopeng, Cao Xunzhi. A survey of digital fountain codes and its application [J]. Acta Electronica Sinica, 2009,37(7):1571-1577.
- [3] Byers J W, Luby M, Mitzenmacher M, et al. A digital fountain approach to reliable distribution of bulk data [J]. ACM Sigcomm Computer Communication Review, 1998,28(4):56-57.
- [4] Luby M. LT codes [C]// Proceedings of 2002 IEEE Symposium Foundations of Computer Science (FOCS). Vancouver, Canada: IEEE, 2002:271-280.

- [5] Shokrollahi A. Raptor codes [J]. *IEEE Transactions on Information Theory*, 2006, 52(6):2551-2567.
- [6] Mackay D J C. Fountain codes [J]. *IEE Proceedings Communications*, 2005, 152(6):1062-1068.
- [7] 3rd Generation Partnership Project. Technical specification group services and system aspects; multimedia broadcast/multicast services(MBMS); protocols and codecs(Release 11) [S]. 3GPP TS 26.346 V11. 2.0-2012, Volbonne, France; 3GPP, 2012.
- [8] 石东新, 杨占昕, 张铨. 3GPP MBMS 中 Raptor 编解码研究 [J]. *数据采集与处理*, 2010, 25(S1):120-124.
Shi Dongxin, Yang Zhanxin, Zhang Quan. Key technologies and performance analysis of raptor codes in 3GPP MBMS [J]. *Journal of Data Acquisition and Processing*, 2010, 25(S1):120-124.
- [9] Tanner R M. A recursive approach to low complexity codes [J]. *IEEE Transactions on Information Theory*, 1981, 27(5):533-547.
- [10] Saejoon K, Karam K, Sae-Young C. Incremental Gaussian elimination decoding of raptor codes over BEC [J]. *IEEE Communication Letters*, 2008, 12(4):307-309.
- [11] Bioglio V, Grangetto M, Gaeta R, et al. On the fly Gaussian elimination for LT codes [J]. *IEEE Communications Letters*, 2009, 13(12):953-955.
- [12] 朱宏鹏, 李广侠, 冯少栋. LT 码的 BPML 译码算法 [J]. *计算机科学*, 2009, 36(10):77-81.
Zhu Hongpeng, Li Guangxia, Feng Shaodong. BPML decoding algorithm of LT codes [J]. *Computer Science*, 2009, 36(10):77-81.
- [13] 朱宏杰, 裴玉奎, 陆建华. 一种提高喷泉码译码成功率的算法 [J]. *清华大学学报:自然科学版*, 2010, 50(4):609-612.
Zhu Hongjie, Pei Yukui, Lu Jianhua. Algorithm improving the decoding performance of fountain codes [J]. *Journal of Tsinghua University: Science and Technology*, 2010, 50(4):609-612.
- [14] Luby M, Mitzenmacher M, Shokrollahi A. Analysis of random processes via And-Or tree evaluation [C]//*Proceedings of the Ninth Annual ACM-SI-AM Symposium on Discrete Algorithms*. San Francisco, USA; Society for Industrial and Applied Mathematics, 1998:364-373.
- [15] Karp R, Luby M, Shokrollahi A. Finite length analysis of LT codes [C]//*2004 IEEE International Symposium on Information Theory(ISIT)*. Chicago, USA; IEEE, 2004:39.
- [16] Maneva E, Shokrollahi A. New model for rigorous analysis of LT-codes [C] // *2006 IEEE International Symposium on Information Theory(ISIT)*. Seattle, USA; IEEE, 2006:2677-2679.
- [17] Palanki R, Yedidia J S. Rateless codes on noisy channels [C] // *2004 IEEE International Symposium on Information Theory (ISIT)*. Chicago, USA; IEEE, 2004:37.
- [18] Etesami O, Shokrollahi A. Raptor codes on binary memoryless symmetric channels [J]. *IEEE Transactions on Information Theory*, 2006, 52(5):2033-2051.
- [19] Niesen U, Erez U, Shah D, et al. Rateless codes for the Gaussian multiple access channel [C]//*2006 IEEE Global Telecommunications Conference (GLOBECOM)*. San Francisco, USA; IEEE, 2006:1-5.
- [20] Castura J, Mao Y. Rateless coding over fading channels [J]. *IEEE Communications Letters*, 2006, 10(1):46-48.
- [21] Venkiah A, Poulliat C, Declereq D. Jointly decoded raptor codes: Analysis and design for the BIAWGN channel [J]. *Eurasip Journal on Wireless Communications and Networking*, 2009, 10:1155.
- [22] Erez U, Trott M D, Wornell G W. Rateless coding for Gaussian channels [J]. *IEEE Transactions on Information Theory*, 2012, 58(2):530-547.
- [23] Bai B, Bai B, Ma X. Simple rateless error-correcting codes for fading channels [J]. *Science China Information Science*, 2012, 55(10):2194-2206.
- [24] Chen S, Zhang Z, Zhu L, et al. Accumulate rateless codes and their performance over additive white Gaussian noise channel [J]. *IET Communications*, 2013, 7(4):372-381.
- [25] Kharel A, Cao L. Decoding of short LT codes over BIAWGN channels with Gauss-Jordan elimination-assisted belief propagation method [C] // *2015 Wireless Telecommunications Symposium (WTS)*. New York, USA; IEEE, 2015: 1-6.
- [26] Wiberg N. Codes and decoding on general graphs [D]. Linkoping; Linkoping University, 1996.

- [27] Ahlswede R, Ning C, Li S Y R, et al. Networking information flow [J]. IEEE Transactions on Information Theory, 2000, 46(4):1204-1216.
- [28] Li S Y R, Yeung R W, Ning C. Linear network coding [J]. IEEE Transactions on Information Theory, 2003, 49(2):371-381.
- [29] Bassoli R, Marques H, Rodriguez J, et al. Network coding theory: A survey [J]. IEEE Communications Surveys & Tutorials, 2003, 15(4):1950-1978.
- [30] Matsuda T, Noguchi T, Takine T. Survey of network coding and its applications [J]. IEICE Transactions on Communications, 2011, E94-B(3):698-717.
- [31] Liew S C, Zhang S, Lu L. Physical-layer network coding: Tutorial, survey, and beyond [J]. Physical Communication, 2013, 6:4-42.
- [32] Shao H, Xu D, Zhang X. Asymptotic analysis and optimization for generalized distributed fountain codes [J]. IEEE Communications Letters, 2013, 17(5):988-991.
- [33] 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.
- [34] Sejdinovic D, Piechocki R J, Doufexi A. AND-OR tree analysis of distributed LT codes [C]// Proceedings 2009 IEEE Information Theory Workshop on Networking and Information Theory. Volos, Greece; IEEE, 2009:261-265.
- [35] Nguyen T D, Yang L L, Hanzo L. Systematic Luby transform codes and their soft decoding [C]//2007 IEEE Workshop on Signal Processing Systems. Shanghai, China; IEEE, 2007:67-72.
- [36] 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]//IEEE Wireless Communications and Networking Conference (WCNC 2008). Las Vegas, USA; IEEE, 2008:243-248.
- [37] 朱宏鹏, 张更新, 李广侠. 卫星数据广播分发系统中 LT 码的研究 [J]. 通信学报, 2010, 31(7):122-127.
Zhu Hongpeng, Zhang Gengxin, Li Guangxia. Research of LT code in satellite data broadcasting system [J]. Journal on Communications, 2010, 31(7):122-127.
- [38] 王新泽. LT 码的度分布设计及译码算法研究 [D]. 西安:西安电子科技大学, 2014.
Wang Xinze. Research on LT codes degree distribution optimization and decoding algorithms [M]. Xi'an: Xidian University, 2014.
- [39] Zhu H, Zhang C, Lu J. Designing of fountain codes with short code-length [C]//The Third International Workshop on Signal Design and Its Application in Communication (IWSDA 2007). Chengdu, China; IEEE, 2007: 65-68.
- [40] 焦健, 杨志华, 顾术实, 等. 基于随机置换展开与停止集的 LT 码联合编译码算法 [J]. 通信学报, 2013, 34(2):31-39.
Jiao Jian, Yang Zhihua, Gu Shushi, et al. Novel joint encoding/decoding algorithms of LT codes based on random permute edge-growth and stopping set [J]. Journal on Communications, 2013, 34(2):31-39.
- [41] Meng Z, Calderbank R, Shuguang C. On design of rateless codes over dying binary erasure channel [J]. IEEE Transactions on Communications, 2012, 60(4):889-894.
- [42] Hyytiä E, Tirronen T, Virtamo J. Optimizing the degree distribution of LT codes with an importance sampling approach [C]//Proceedings of the 6th International Workshop on Rare Event Simulation (RESIM 2006). Bamberg, Germany; [s. n.], 2006:6473.
- [43] Hyytiä E, Tirronen T, Virtamo J. Optimal degree distribution for LT codes with small message length [C]// 26th IEEE International Conference on Computer Communications (INFOCOM 2007). Alaska, USA; IEEE, 2007:2576-2580.
- [44] Hussain I, Xiao M, Rasmussen L K. Design of LT codes with equal and unequal erasure protection over binary erasure channels [J]. IEEE Communications Letters, 2013, 17(2):261-264.
- [45] Hayajneh K, F Yousefi S, Valipour M. Left degree distribution shaping for LT codes over the binary erasure channel [C]// Proceedings of IEEE 27th Biennial Symposium Communications (QBSC). Kingston, Canada; IEEE, 2014:198-202.
- [46] Hayajneh K F, Yousefi S, Valipour M. Improved finite-length Luby-transform codes in the binary erasure channel [J]. IET

Communications, 2015, 9(8):1122-1130.

- [47] Tsai P C, Chen C M, Chen Y P. A novel evaluation function for LT codes degree distribution optimization [C]//2014 IEEE Congress on Evolutionary Computation (CEC). Beijing, China: IEEE, 2014:3030-3035.
- [48] Wang Yan, Liu R, Duan R. Work in progress: Improving the intermediate recovery rate of LT codes using a novel degree distribution [C]//2014 9th International Conference on Communications and Networking in China (CHINACOM). Maoming, China: IEEE, 2014:252-255.
- [49] Puducheri S, Klierer J, Fuja T E. The design and performance of distributed LT codes [J]. IEEE Transactions on Information Theory, 2007, 53(10):3740-3754.

作者简介:



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。



华洁 (1991-) 女, 硕士研究生, 研究方向: 数字喷泉码, 网络编码。



程凯 (1990-) 男, 硕士研究生, 研究方向: 数字喷泉码, 网络编码。

