

基于新型随机度分布的压缩喷泉码

陈月云* 刘伟

(北京科技大学计算机与通信工程学院 北京 100083)

摘要: 该文提出二进制指数随机度分布(Binary Exponential Random Degree Distribution, BERDD)函数, 用 BERDD 函数生成度时提高了度 1 生成的概率, 同时降低度 1 断层出现的概率, 提高了迭代效率。进一步, 为解决喷泉码需要传输带宽大的问题, 提出基于码本压缩的传输方式。仿真表明 BERDD 分布较二进制指数度分布及稳健孤子度分布更能有效提高译码效率, 在瑞利衰落信道条件下, 基于码本压缩喷泉码传输方法能有效降低编码占用信道带宽。

关键词: 喷泉码; 度分布; 码本压缩; 译码开销

中图分类号: TN911.22

文献标识码: A

文章编号: 1009-5896(2012)05-1185-06

DOI: 10.3724/SP.J.1146.2011.01360

Compressed Fountain Codes Based on New Random Degree Distribution

Chen Yue-yun Liu Wei

(Department of Computer & Communication Engineering, University of Science & Technology Beijing, Beijing 100083, China)

Abstract: A Binary Exponential Random Degree Distribution (BERDD) is proposed, the generated fountain codes using BERDD function improve the probability of generating degree “1” in encoding, decreases the probability of degree “1” chasm in decoding, so the decoding complexity is lowed and iteration efficiency is improved. Further, in order to solve the problem of wide transmission bandwidth, a new transmission scheme of compressing fountain coding is proposed based on codebook. It increases bandwidth utilization. Simulation results show that the proposed BERDD function improves decoding efficiency compared with robust soliton degree distribution and binary exponential degree distribution, and the compressing fountain codes reduce transmitting bandwidth effectively in the condition of Rayleigh fading channel.

Key words: Fountain codes; Degree distribution; Codebook compression; Decoding overhead

1 引言

喷泉码^[1]是一类新颖的基于图的信道编码技术, 无码率特性使其在发送编码之间无需等待反馈, 因其性能优越被逐步推广到无线广播多播服务中^[2]。

Luby 等人于 1998 年首次提出喷泉码概念, 并在 2002 年提出了第 1 种现实可行的喷泉码——LT 码^[3]。之后, 文献[4]在 LT 码的基础上进行改进, 使其译码时间有所提高, 这种新型的喷泉码后又被命名为 Raptor 码。

在文献[3]中给出两种度分布, 分别是理想孤子度分布和稳健孤子度分布。理想孤子度分布理论上使每一个编码符号在任何一次译码迭代中释放概率相同^[3], 保证在译码迭代过程中总存在仅有一个输入符号在预处理集中, 使每完成一次迭代都能译出一个输入符号。但在实际中, 由于度分布映射的随机

性使得极微小的波动误差都能造成预处理集消失并致使译码失败。稳健孤子度分布是在理想孤子度分布的基础上做了改进, 由两部分组成, 第 1 部分为上述的理想孤子度分布, 另一部分加入了两个调节参数 c 和 δ , 在结合了理想孤子度分布的优点基础上, 扩展了译码过程中的预处理集, 使其消失概率充分小, 提高译码成功概率, 但同时由于编码覆盖的冗余度增加, 造成译码效率降低。

文献[5]提出多目标进化算法, 以译码开销和计算复杂度作为目标参数, 采用不同度分布可获得不同指标的高译码性能。文献[6]提出喷泉码优先级的概念, 将高优先级的重要数据采用度 1 或 2 进行编码, 使重要数据能够被快速译码。文献[7]提出一种简单的度分布, 称为二进制指数分布(Binary Exponential Distribution, BED)。但是该方法可能会过分增加度 1 的产生概率, 降低了信源符号参与编码的概率, 增加度 1 断层的出现概率^[8]。

文献[8,9]在文献[7]的基础上对简单 BED 度分布加以改进, 进一步优化了喷泉码的性能。文献[9]

2011-12-21 收到, 2012-03-22 改回

北京市自然科学基金(4102041), 新一代宽带无线移动通信网科技重大专项(2011ZX03003-002-03, 2011ZX03004-002-02)资助课题

*通信作者: 陈月云 chenyy@ustb.edu.cn

提出基于二项分布函数的度分布函数，增加原始符号参与编码的概率；文献[10]提出的转换度分布是将BED度分布与稳健孤子度分布相结合，有效提高成功译码概率。但以上两种改进度分布的编码性能没能在最大程度上逼近理想孤子度分布，在译码效率方面仍有可提升空间。在文献[2]中对于喷泉码译码给出两种通用译码算法，即高斯消元译码算法及置信传播(BP)译码算法。文献[10]给出一种最大似然(ML)译码算法。高斯译码及ML译码算法较BP算法能够提高效率但译码复杂度较高，且随着信源符号数增加而迅速增加，但是由于BP算法受度1断层的制约，一旦出现度1断层则译码中断^[1]。

本文通过分析低的度值对编译码产生的影响，提出二进制指数随机度分布(BERDD)，具有理想孤子度分布及BED分布的性能折中，不仅降低译码过程中出现度1断层的概率而且译码复杂度低，译码开销小，迭代效率高。喷泉码的译码需要传输编码生成矩阵，且一行生成矩阵中元素个数等于信息包的符号个数，在发送过程中占用大量带宽，由此提出基于绝大多数生成矩阵的稀疏特性，采用码本方式对生成矩阵进行压缩的策略，有效压缩发送数据帧长，提高系统吞吐量。

2 系统模型

蜂窝小区内用户向基站发出多媒体业务请求，实现数据的实时传输，用户随机分布在小区内的不同位置，系统模型如图1所示。基站将请求相同多媒体数据的M个用户划分为一个多播组Group(U_1, U_2, \dots, U_M)，基站采用喷泉码对所发送数据包进行编码。基站以信息包为单位发送多媒体数据，当多播组中的所有用户都成功接收并完成一个信息包译码后，基站发送下一个信息包。由于多播组用户在实际地理位置上是随机分布的，因此每个用户与基站间的信道状况各不相同，所以信息的传输效率取决于多播组中信道条件最差的用户。

系统原理如图2所示。在信源端，将长为 K_{all} 个符号的信息分隔为 $n = K_{all} / K$ 个信息包(即每个信

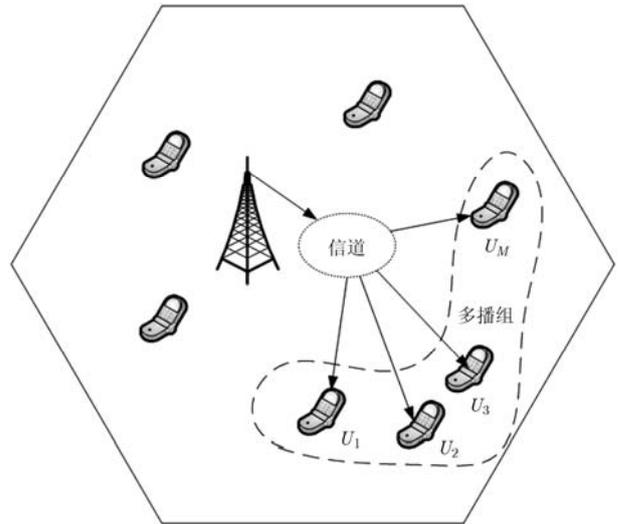


图1 系统模型

息包大小为K个符号)。

喷泉码的编码算法：

- (1)根据给定的度分布函数 $\rho(d)$ 随机产生度 d ；
- (2)在信息包中的K个符号中随机选取 d 个不同的输入符号；
- (3)编码后的输出符号为(2)中 d 个不同输入符号的异或和。

为避免因衰落信道产生传输差错导致译码失败，在发送端采用循环冗余校验(Cyclic Redundancy Check, CRC)，在接收端删除产生差错数据帧。由于喷泉码的无码率特性，因此删除错误帧不会对译码造成影响。在码本压缩模块中，数据帧以码本方式经过压缩处理，即用不同位数的标志位表示生成矩阵中1的位置，以达到缩短发送帧长的目的。在接收端，多播组内用户不断接收编码信息直至完成一个信息包的译码，当基站收到该多播组所有用户译码成功的确认信息后对下一信息包进行编码并发送。

3 二进制指数随机度分布(BERDD)

度分布的设计影响喷泉码的译码开销与计算复

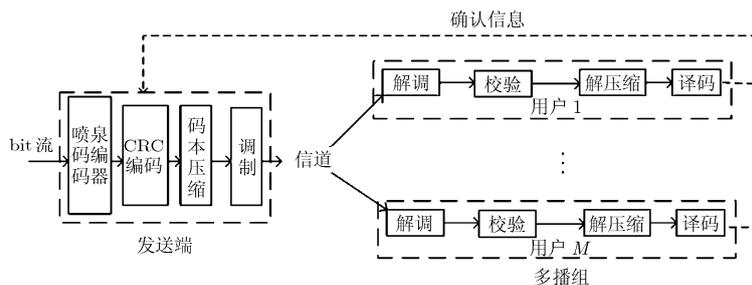


图2 系统原理框图

杂度。由于在译码过程中译码器首先寻找度为 1 的编码符号，即编码符号只和唯一的信源符号相连，并将其还原为信源符号，则度为 1 的编码可以最直接且最快速地还原信源符号。此后通过异或相加，删除他们在生成图中相连的边^[12]，降低未恢复编码的度值，从而使某些度为 2 的编码的度值降为 1，通过迭代实现译码。但是，如果在译码过程中生成矩阵不能收敛到度 1 时，译码过程中断，称之为度 1 断层。因此，度值较低的编码符号特别是度为 1 的编码符号个数多时，降低产生编码时输出符号间的关联性，则译码容易出现度 1 断层，需要接收更多的编码才能完成译码，导致译码开销增加。相反，若度值较低的编码符号特别是度为 1 的编码符号个数太少，译码所需模二加次数增加，导致译码复杂度增加，译码速度降低。

因此，度为 1 的编码个数应有合理取值，本文提出基于递缩等比数列的改进型随机度分布函数，叫做二进制指数随机度分布(Binary Exponential Random Degree Distribution, BERDD)，为

$$f(d) = \frac{\rho(d) + \tau(d)}{Z} \quad (1)$$

其中 $Z = \sum_d (\rho(d) + \tau(d))$ ， $\rho(d)$ 和 $\tau(d)$ 分别为

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

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

$\rho(d)$ 为二进制指数分布(BED)^[13]。BED分布是基于无穷递缩等比数列， $d = t(t = 1, \dots, K-2)$ 的产生概率是 $d = t+1$ 产生概率的 2 倍。 $\tau(d)$ 为理想孤子度分布。在模型式(1)中，将BED分布与理想孤子度分布进行归一化。BERDD分布既能弥补BED分布度 1 产生概率过高的缺陷，又能最大程度上保留理想孤子度分布的优良特性且有效避免度 1 断层的出现。

图3给出度 1 对译码过程产生的影响。图中以 4 个输出符号及 5 个编码符号为例，根据稳健孤子度分布、BED及BERDD的度分布函数得到度 1 的产生概率分别为 $\mu(1) = 0.1237$ ^[4]， $\rho(1) = 0.5$ ， $f(1) = 0.35$ ，四舍五入分别取度 1 编码符号个数为 1, 3, 2，如图 3(a), 3(b), 3(c)所示。图3(a)中，完成译码需 6 步迭代，8 次异或运算。图3(b)中，当译码进入状态 2 时，出现度 1 断层，导致译码中断。图3(c)所示译码过程，完成译码只需 3 步迭代，4 次异或运算，相比度 1 编码符号个数为 3 的情况，译码迭代次数减少。由此可见，BERDD 度分布能够在降低度 1 断层出现概率的同时，提高译码速度，降低译码复杂度。

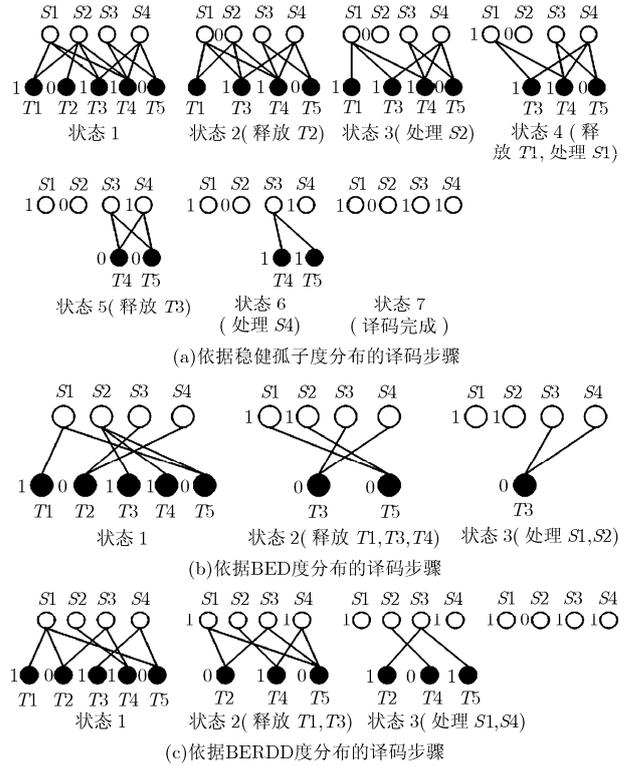


图 3 译码状态转移图

4 基于码本的压缩喷泉码设计

4.1 码本压缩编码原理

根据 BERDD 度分布函数可知，低度值的产生概率远高于高度值的产生概率，因此每进行一次编码生成矩阵大多数为 $K \times 1$ 的稀疏矩阵列，其中 K 为信息包大小。采用码本方式进行压缩，即用 M 个二进制比特来表示矩阵中 1 所在的位置，称此 M 个二进制比特为标志位。当矩阵中 1 的个数 $L < K$ 时，则存在 $M \times L < K$ 的情况，达到压缩生成矩阵的目的。根据 $D \times M < K$ 得到可实现压缩的最大度值 $D = \lfloor K/M \rfloor$ 。

成功译码所需接收的编码个数 $N = (1 + \epsilon)K$ ，其中 ϵ 为译码开销， $\epsilon < 1$ 。由于接收端需要编码产生的生成矩阵进行二部图重构以完成译码，因此，整个译码过程所需发送的符号总数为 $A = (1 + K)N$ 。采用码本压缩编码，当生成矩阵度值满足 $d < D$ 时，该生成矩阵可被压缩，且压缩后的符号数为 $M \times d$ 。有 BERDD 度分布函数 $f(d)$ 可知 N 个编码符号中度为 d 的个数是 $N \times f(d)$ ，则整个译码过程所需发送符号总数为

$$A' = N + N \times M \times [f(1) + 2f(2) + \dots + Df(D)] + N \times K \times [f(D+1) + \dots + f(K)] \quad (4)$$

得到压缩率为

$$R = A'/A = \{1 + M \times [f(1) + 2f(2) + \dots + Df(D)] + K \times [f(D+1) + \dots + f(K)]\} / (1 + K) \quad (5)$$

4.2 编码效率及占用带宽分析

定义 $E(K)$ 为信源符号数为 K 时的编码效率, 即信源符号数与发送符号数之比。则编码效率可表示为 $K/(1+K)N = 1/(1+K)(1+\epsilon)$ 。因此, 编码效率与信息包大小 K 近似成反比, 即信息包越大, 编码效率越低。

应用码本压缩喷泉码传输, 设发送端信源符号数为 K_{all} 个, 标志位的个数为 M , 则生成矩阵维数最大为 2^M , 即参与一次 LT 编码的信息包符号数最多为 $K = 2^M$ 。发送所有信源符号所需的信息包个数为 $K_{all}/2^M$, 则正确译出一个信息包所需的编码个数为 $N = K \times (1 + \epsilon) = 2^M \times (1 + \epsilon)$ 。由式(4)得, 正确译出一个信息包所需接收的比特数为

$$T = 2^M \times (1 + \epsilon) \times \{1 + M \times [f(1) + 2f(2) + \dots + Df(D)] + 2^M \times [f(D+1) + \dots + f(2^M)]\} \quad (6)$$

其中 $D = \lceil 2^M/M \rceil$, 正确译出 K_{all} 个信源符号的编码效率为

$$E = \frac{K}{K_{all}/2^M \times T} = 1/\{(1 + \epsilon) \times [1 + M \times [f(1) + \dots + Df(D)] + 2^M \times [f(D+1) + \dots + f(K)]]\} \quad (7)$$

由式(6)可以看出随着 M 的增大, 编码效率有所下降。

由于码本压缩的原理是用标志位比特来表示稀疏生成矩阵中 1 的位置, 因此产生编码的度值越小其产生的生成矩阵的可压缩度就越高。以 $K=32$ 为例, 其标志位个数 $M = \log_2 32 = 5$, 其可被压缩的度的最大值 $D = \lceil 2^M/M \rceil = \lceil 32/5 \rceil = 6$, 则当 $d > D$ 时, 发送一个编码所附带的生成矩阵列为 32 bit, 当 $d=1$ 时, 压缩后的生成矩阵列仅为 5 bit, 其压缩率为 15.6%。因此, 度 1 产生的个数对压缩效果起着至关重要的作用。

对编码产生的生成矩阵采用码本压缩后, 发送数据帧长缩短。设压缩前发送数据帧长为 L bit, 压缩后发送数据帧长为 $L_i (i=1, \dots, N)$, 且 $L_i < L$, 经

喷泉码编码后码率 R_B 不变。需要注意的是, 由于每次编码产生的生成矩阵的稀疏度不同, 因此压缩后 L_i 大小不定。由式(8)得 $R_{Bi}' < R_B (i=1, \dots, N)$, 即经数据压缩后码元速率降低, 则发送每帧所占用的信道带宽降低, 其中 R_{Bi}' 为压缩后的码元速率。

$$\frac{L}{R_B} = \frac{L_i}{R_{Bi}'} \quad (8)$$

综上所述, 对发送数据帧进行码本压缩后, 平均占用带宽降低, 节省了带宽资源, 提高了带宽利用率。

5 性能仿真及分析

在仿真中, 采用 LT 码作为喷泉码编码, 完成编码后为每个分组加上校验和, 不考虑校验和的漏检概率, 则近似将分组交换的逻辑信道等效为删除信道, 以 BPSK 作为调制方式, 接收端采用置信传播译码算法。图 4, 图 5 是单一用户条件下对性能参数的仿真结果。图 6, 图 7 仿真多播环境下考察译码开销, 为简化分析, 取多播组用户数为 5, 数据分组经瑞利衰落信道且信道条件相互独立。

表 1 给出不同度分布度值产生概率统计, 仿真中原始数据包为 100 个符号, 根据稳健孤子, BED 和 BERDD 3 种度分布分别产生 100×1000 个编码, 对不同度值进行统计, 表 1 给出度值分别为 1, 2, 3, 4, 5, 10, 50, 100 的产生概率。可以看出, BERDD 分布度值的产生概率介于稳健孤子度分布与 BED 分布之间, 与前文分析一致, 且随着度值的增大, 其产生概率逐渐降低。另外, 由表中数据知, 对于 3 种度分布度值在 5 以下的产生概率总和均接近或超过 50%, 说明编码产生的生成矩阵绝大多数为稀疏矩阵, 这为码本压缩提供了有利依据。

图 4, 图 5 分别给出根据 3 种度分布函数产生编码是对输出符号的覆盖情况曲线及译码复杂度曲线。如图 4 所示, 数据包大小 K 分别取 30, 50, 80, 100, 120, 计算在不同度分布函数下确保所有信息符号都参与编码的最少编码次数。BED 度分布在编码时要覆盖到所有信息符号所需编码次数较多, 因为其小

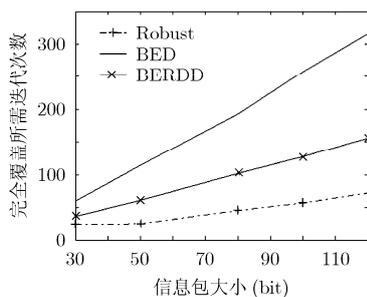


图 4 度分布覆盖情况

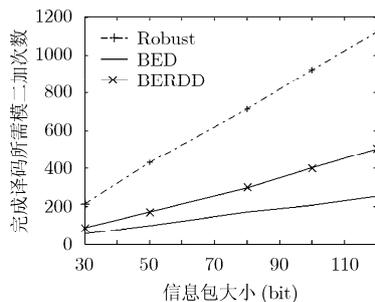


图 5 译码复杂度

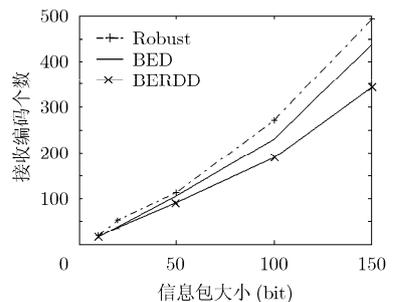


图 6 不同度分布完成译码所需编码个数

表1 度值产生概率统计

度值	1	2	3	4	5	10	50	100
Robust	0.026	0.453	0.154	0.078	0.046	1.20×10^{-2}	0.075×10^{-2}	0.01×10^{-2}
BED	0.498	0.249	0.126	0.063	0.031	0.10×10^{-2}	0	0
BERDD	0.257	0.373	0.146	0.074	0.041	0.76×10^{-2}	0.010×10^{-2}	0.01×10^{-2}

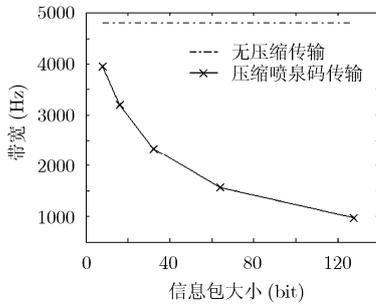


图7 采用码本压缩传输前后平均占用带

度值的产生概率过大，尤其是度 1 的产生概率，Robust 度分布的覆盖性最好。由图 5 可以看出，随着信息包大小 K 值增加，成功译码所需模二加运算次数增加，且在相同条件下，BED 度分布最优，BERDD 度分布次优。因此，编码时度分布对输出符号覆盖性越好，则译码复杂度越高。虽然 Robust 度分布的覆盖性最好，但在译码过程中译码复杂度高，译码速度较慢。BERDD 度分布取前两者的折中，在提高低度值产生概率的同时，兼顾对输出编码的覆盖。

图 6 为 3 种度分布函数译码开销性能曲线，通过统计成功译码所需编码符号个数来衡量度分布性能。由前文分析知， K 值越大译码复杂度越高，此处设定 K 为 150 以内数值。由图 6 可知，在信息包大小一定情况下，采用 BERDD 分布函数完成译码所需编码个数小于 Robust 度分布和 BED 度分布，且随着信息包大小的增加，其性能优势体现得更为明显。

图 7 为在不同信息包大小条件下采用码本压缩喷泉码传输前后平均占用信道带宽仿真曲线。其中，喷泉码的度分布采用本文提出的 BERDD 度分布，且由前文分析可知，在码本压缩过程中当标志位分布取 3, 4, 5, 6, 7 bit 时，信息包大小为 8, 16, 32, 64, 128 bit 的压缩率最高。因此，在仿真中取信息包大小分布为 8, 16, 32, 64, 128 bit。由图 6 可以看出，采用码本压缩喷泉码传输后，在一定信息包大小条件下平均占用信道带宽降低，节省了带宽资源。若喷泉码编码后码速率不变，随信息包大小的增加，采用码本压缩喷泉码传输，其带宽利用率性能优势更为明显。

6 结论

本文通过分析度分布对喷泉码编译码的影响，提出了 BERDD 分布，将理想孤子度分布和 BED 分布进行归一化，折中两者度值产生概率。仿真表明，对于符号数为 100 的信息包，采用 BERDD 度分布编码度 1 的产生概率较稳健孤子度分布提高了 23%，且译码复杂度较稳健孤子度分布大大降低。在多播传输环境下译码开销低于稳健孤子度分布及 BED 分布，随着信息包符号数的增多，其译码开销性能优势体现更为明显。采用码本压缩喷泉码传输方式，对编码产生的生成稀疏矩阵压缩，用标志位信息表示稀疏矩阵中 1 的位置，缩短发送数据帧长，节省了带宽资源，提高了带宽利用率，具有一定的实用价值。

参考文献

- [1] MacKay D. Fountain codes[J]. *IEEE Proceedings Communications*, 2005, 150(6): 1062-1068.
- [2] 3GPP TS 26.346 V6.12.0. Multimedia Broadcast/Multicast Service (MBMS)[S]. 2008.
- [3] Luby M. LT codes[C]. *Proceedings of the 43rd Annual IEEE Symposium on the Foundations of Computer Science, USA*, Nov. 19, 2002, 43, 271-280.
- [4] Shokrollahi A, et al. Raptor codes[J]. *IEEE Transactions on Information Theory*, 2007, 58(4): 2551-2567.
- [5] Chen Chi-ming and Chen Ying-ping. Optimizing degree distributions in LT codes by using the multiobjective evolutionary algorithm based on decomposition[C]. *Proceedings of Evolutionary Computation (CEC)*, Barcelona, Spain, July 18-23, 2010: 1-8.
- [6] Woo S and Cheng M. Prioritized LT codes[C]. *Proceedings of Information Sciences and Systems*, Princeton, USA, Mar. 19-21, 2008: 568-573.
- [7] Al Agha K and Kadi N. Fountain codes with XOR of encoded packets for broadcasting and source independent backbone in multi-hop networks using network coding[C]. *Proceedings of Vehicular Technology Conference*, Barcelona, Spain, Apr. 26-29, 2009: 1-5.
- [8] Zaman M S and RamaMurthy G. A new degree distribution for LT codes for broadcasting in Ad hoc network using network coding[C]. *First UK-India International Workshop on Cognitive Wireless Systems (UKIWCWS)*, Delhi, India,

- Dec. 10-12, 2009: 1-5.
- [9] Kadi N and Al Agha K. New degree distribution to improve LT-code in network coding for broadcasting in Ad hoc wireless networks[C]. Proceedings of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Istanbul, Turkey, Sept. 26-30, 2010: 1820-1825.
- [10] Schotsch B and Schepker H. The performance of short random linear fountain codes under maximum likelihood decoding[C]. Proceedings of IEEE International Conference on Computer Communications(ICC), Kyoto, Japan, June 5-9, 2011: 1-5.
- [11] Huang Wei-zheng and Li Huan-lin. Fountain codes with message passing and maximum likelihood decoding over erasure channels[C]. Proceedings of Wireless Telecommunications Symposium (WTS), New York, USA, Apr. 13-15, 2011: 1-5.
- [12] Lin Guang-rong. Finite length analysis of fountain codes based on stopping set[J]. *Journal of Electronics & Information Technology*, 2008, 30(11): 2634-2637.
- 陈月云: 女, 1966年生, 副教授, 研究方向为无线移动通信, 主要研究兴趣包括无线资源管理、宽带移动通信传输理论与技术、无线通信信号处理等.
- 刘伟: 男, 1987年生, 硕士生, 研究方向为无线移动通信.