DOI:10.11721/cqnuj20140317

A Joint Degree Distribution LT Codes Design*

MA Yao-guo
(Information Technology Center, Chongqing Normal University, Chongqing 401331, China)

Abstract: LT codes are a class of forward error correction channel codes which are developed to recover channel packet erasures. They have a multiple of uses in computer science, network transmission, media storage, large file download and so on. In the design of LT codes, the degree distribution is the key to successful decoding as well as fast operation. This paper presents a novel design with a weakened distribution for prior decoding and an improving distribution for posterior decoding. Since the weakened distribution has a low average degree, it substantially increases both the encoding and decoding speed, while the improving distribution has a high average degree that enhances the successful decoding probability. With a series of simulations, we observed that the overhead as well as the XOR operations is cut about 50% compared with the reference design of Robust Soliton distribution.

Key words: LT codes; degree distribution; release probability; overhead; XOR operation
中图分类号: TN911.22　**文献标志码:** A　**文章编号:** 1672-6693(2014)03-0076-06

1 Introduction

LT codes[1] have attracted great attention in such applications as peer-to-peer transmission, scalable video streaming, large file download and so on [2-4], due to their asynchronous coding-decoding characteristic between senders and receivers. The senders infinitely generate continuous encoding symbols from a finite set of source symbols. Each encoding symbol is independent of the others. Disregarding which portion of the encoding symbols is received by the receivers or in what order it is received, decoding is successful if only the encoding size is large enough.

However, LT codes have two primary drawbacks. One is that the overhead is asymptotical with the source size. For a sufficiently large source size it reaches a floor near zero, but for a small source size[5] it is significant and results in a great waste of channel capacity. The other is that LT codes rely on Robust Soliton degree distribution[1] to achieve high decoding probability. Each encoding symbol is generated by on average $O(\ln(K/\delta)$ XOR operations and overall decoding requires $O(K\ln(K/\delta))$ XOR operations.

There are several improvement for LT codes, e. g. Gaussian elimination algorithm[6] can suppress overhead; the concatenation code structure[7] can reduce decoding operations. In the various approaches, we have a heuristic idea of designing the encoding process with mixed types of degree distributions, which is named joint degree distribution LT (JDD-LT). Without changing the basic LT structure, two encoding stages are introduced, a prior stage and a posterior stage. In the prior stage, each encoding symbol is generated with the weakened distribution targeting recovering the majority of the source symbols. In the posterior stage, the distribution is replaced by the improving distribution targeting recovering the remained source symbols. The de-

* **Received:** 10-11-2013　修回日期:12-15-2013　**网络出版时间:** 2014-5-8 14:38
The first autor biography: Ma Yaoguo, male, assistant engineer, major in channel code, compressed video communication. E-mail: mayaoguo888@163.com
收稿日期: 2013-10-11　**修回日期:** 2013-12-15　**网络出版时间:** 2014-5-8 14:38
作者简介: 马耀国,男,助理工程师,研究方向为信道编码、视频传输,E-mail:mayaoguo888@163.com
网络出版地址: http://www.cnki.net/kcms/detail/50.1165.N.20140508.1438.018.html

sign has two advantages: first, it inherits the decoding algorithm of Belief-Propagation (BP) [1] with low decoding operations; second, it is accustomed to a wide range of source sizes without inner code[7] constraint. Simulations show that comparing the single Robust Soliton distribution LT design the JDD-LT design has lower overhead while reducing both the encoding and decoding operations.

The paper is organized as follows: Chap. 2 introduces the basic knowledge of LT codes; Chap. 3 analyzes the JDD-LT design; Chap. 4 presents the simulation results comparing the single Robust Soliton distribution LT design; Chap. 5 concludes the paper.

2 LT codes and characteristics

Before introducing the JDD-LT design, we have a brief overview of the basic knowledge of LT codes[1].

2.1 Encoding and decoding

An encoding symbol is generated as follows: firstly, choose an integer d from the set $\{1, \cdots, K\}$ according to a specially designed distribution $\Omega(x)=\sum_{i=1}^{K}\Omega_i x^i$ in which Ω_i represents the choosing probability on i; secondly, randomly generate a weight-d vector on F_2^K and choose d source symbols corresponding to the positions of 1's in the vector; finally, perform XOR operations on the d chosen symbols. d is the degree of the encoding symbol and $\Omega(x)$ is the degree distribution.

If receiving a sufficient portion of encoding symbols, the K source symbols can be recovered by the following iterations: on a bipartite graph, a degree-1 encoding symbol is copied to its neighbor, recovering one source symbol; the recovered source symbol is subsequently performed XOR operations on its neighbors, releasing more degree-1 encoding symbols; the released encoding symbols repeat this process until decoding finishes. Decoding is successful if all the source symbols are recovered, and decoding fails if any degree-1 encoding symbols disappear during the process.

2.2 Release probability

Time is defined as the number of the source symbols already recovered. An encoding symbol is called 'released' if it has $d-2$ recovered neighbors at time $K-(L+1)$, one neighbor is exactly recovered at time $K-L$ and the rest neighbor is among the L unrecovered ones. An example of the released $d=4$ encoding symbol is shown in Fig. 1 For a degree-d ($d\neq 1$) encoding symbol, the release probability at time $K-L$ ($1\leqslant L\leqslant K-d+1$) is

$$Q(K,d,L)=\frac{L\cdot d(d-1)\cdot\prod_{i=0}^{d-3}K-(L+1)-i}{\prod_{i=0}^{d-1}K-i} \tag{1}$$

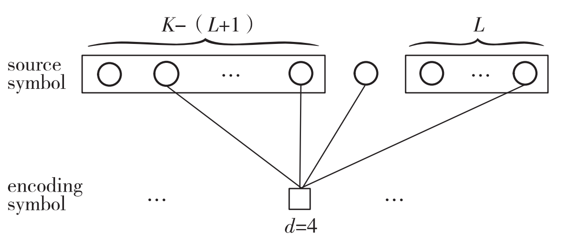

Fig. 1　An example of the released $d=4$ encoding symbol

Particularly, for $d=1$ the encoding symbol is released at time 0, $Q(K,1,K)=1$.

There are such characteristics: as L gradually decreases

1) the encoding symbols of lower degrees tend to release at earlier time, and those of higher degrees tend to release at later time;

2) the release probability of any encoding symbol in the overall process is 1;

3) for $3\leqslant d\leqslant\lceil K/2\rceil$, $Q(K,d,L)$ increases, reaches a peak, and then decreases, and for $\lceil K/2\rceil<d\leqslant K$, $Q(K,d,L)$ continuously increases until the peak at $L=1$;

4) the peak value of the release probability for a degree-d encoding symbol enlarges with d increasing from 3 to K.

2.3 Degree distribution

Assuggested in Ref. [1], good degree distributions should satisfy two requirements: the number of required encoding symbols for successful decoding is as small as possible in order to suppress overhead; the total degree of the encoding symbols is as small as possible in order to reduce XOR operations.

The optimal distribution satisfying the requirements is theoretically Ideal Soliton distribution, which keeps the degree-1 encoding size to be constantly 1 throughout the decoding process, i. e. each time a source symbol is recovered, exactly one encoding symbol is released. Complete recovery of the K source symbols requires exactly K encoding symbols, however, the solitary degree-1 encoding symbol is actually difficult to preserve during the process and decoding is commonly inclined to interrupt. Based on the concept of Ideal Soliton distribution, Robust Soliton distribution is introduced and generally adopted by many LT designs, which possess an enlarged degree-1 encoding size of $c\ln(K/\delta)\sqrt{K}$ while keeping a constant release rate. With this distribution, K source symbols can be recovered with additional $O(\ln^2(K/\delta)\sqrt{K})$ encoding symbols, and $O(K\ln(K/\delta)$ XOR operations. The performance depends on the parameter set (K,δ,c), which is studied by Ref. [8].

3 JDD-LT design

A message is divided into K equal length source symbols, and the encoding symbols are generated with the weakened distribution Ω_{WD}. The decoder collects and decodes them using BP algorithm. After $\gamma K(0<\gamma<1)$ source symbols are recovered, the improving distribution Ω_{ID} is applied and the remained source symbols are recovered.

3.1 Prior stage

This stage aims at partial recovery and the weakened distribution Ω_{WD} is suggested[7]

$$\Omega_{WD}(x)=0.008x+0.494x^2+0.166x^3+0.073x^4+0.083x^5+0.056x^8+ \\ 0.037x^9+0.056x^{19}+0.025x^{65}+0.003x^{66} \tag{2}$$

The average degree of the corresponding encoding symbols is constantly about 5.87. In order to evaluate the distribution performance, we define the cumulative probability

$$P_c(\gamma)=\frac{\sum_{j=1}^{N}\bar{n}(\gamma\leqslant\gamma_j\leqslant 1)}{K},0\leqslant\gamma\leqslant 1 \tag{3}$$

where N is the number of trials, γ_j is the recovery fraction in the j_{th} trial, and $\bar{n}$ is the number of the unrecovered symbols in each trial. Let $N=5000$, respectively $K=2000$, $K=3000$ and $K=4000$. $P_c(\gamma)$ for the constant overhead $\varepsilon=0.05$ is shown in Fig. 2.

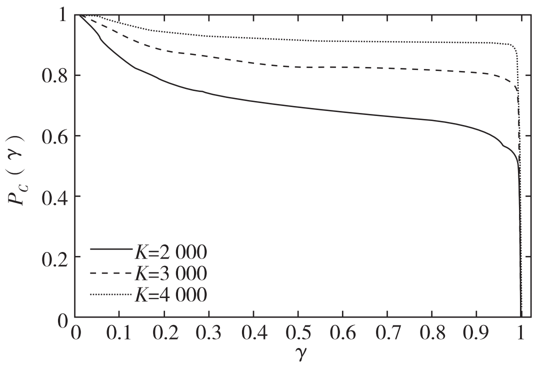

Fig. 2 Cumulative probability, $\varepsilon=0.05$

The three curves universally drop sharply in the final 1% fraction and cease before $\gamma=1$. It is implied that in 5000 trials for each source size the encoding symbols are stably recovered for a ratio of $\gamma=0.99$.

3.2 Posterior stage

Let Ω_{ISD} represent the ideal soliton distribution mentioned in Chap. 2.3

$$\Omega_{ISD}(i)=\begin{cases}1/K, & i=1\\ 1/i(i-1), & i=2,\cdots,K\end{cases} \tag{4}$$

Let $r(1\leqslant r<K)$ represent the number of the expected higher degree terms. By removing the first $K-r$ lower degree terms and making normalization, we obtain

$$\Omega_{ID}(i)=\frac{\Omega_{ISD}(i)}{\sum_{i=K-r+1}^{K}\Omega_{ISD}(i)}=\frac{(K-r)K}{ir(i-1)},\ K-r+1\leqslant i\leqslant K \tag{5}$$

The new distribution Ω_{ID} is the improving distribution. For lacking the lower degree terms, Ω_{ID} targets the posterior stage that requires large release rate as the majority of the source symbols are recovered.

As introduced in Chap. 2.2, a degree-d encoding symbol to be released at time $K-L$ has the probability $Q(K,d,L)$. The probability that an encoding symbol to be chosen as degree-d and released at time $K-L$ is

$$R(K,d,L)=\Omega_{ID}(d)Q(K,d,L) \tag{6}$$

Sinc eevery encoding symbol has an influence on release, the overall release probability at time $K-L$ is

$$R(K,L)=\sum_{i=1}^{K-L+1}\Omega_{ID}(i)Q(K,i,L) \tag{7}$$

Assume γK encoding symbols are recovered

$$R(K,L)=\sum_{i=K-r+1}^{K-L+1}\left[L\cdot\frac{(K-r)K}{r}\cdot\frac{\prod_{j=0}^{i-3}K-(L+1)-j}{\prod_{j=0}^{i-1}K-j}\right] \tag{8}$$

The release probability of the posterior stage is analyzed under several values of r and $\gamma=0.99$. Fig. 3 and Fig. 4 demonstrate the influence of K and r on the final 1% unrecovered source symbols. Particularly as $r=K-1$, i.e. only the degree-1 term is removed from the Ideal Soliton distribution, it is observed that the release probability keeps approximately constant (about $1/K$), which is almost correspondent to the Ideal Soliton distribution. By removing more of the lower degree terms, in the case of $r=K-50$, release probability universally increases for each case of the unrecovered size L. The enlarged fraction of the higher degree terms accelerates the posterior release rate. In the case of $r=K-100$, the posterior release probability is even higher, which is close to or even larger than 0.01. It is implied that each of the remained source symbols can be recovered with fewer encoding symbols compared with the case of a larger r. If r is reduced further, e.g. $r=K-200$ and $r=K-300$, the release rate inclines toward the lower part of L, and for the higher part it witnesses a slight decrease which suggests deterioration of the beginning rate in the posterior stage.

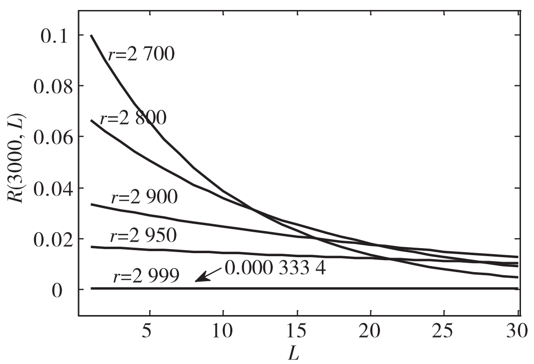

Fig. 3　Release probability, $K=3\,000$, $\gamma=0.99$

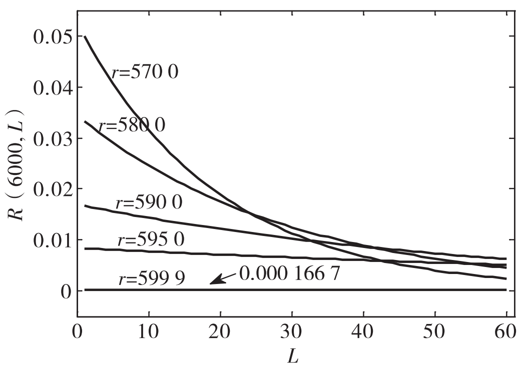

Fig. 4　Release probability, $K=6\,000$, $\gamma=0.99$

4 Simulation results

The performance is measured in terms of decoding inefficiency (the average value of $1+\varepsilon$), encoding speed (average degree per encoding symbol), and decoding speed (the number of XOR operations for overall decoding). Simulations are under a variety of source sizes ranging from $K=3\,000$ to $K=10\,000$. Each sample on a curve is computed by averaging the values of 1 000 trials. For JDD-LT design, $\gamma=0.99$, $r=K-100$. Two stages are respectively displayed, the result of the prior stage is denoted by JDD-1, and that of the posterior stage is denoted by JDD-2. For the reference design with Robust Soliton distribution (denoted by RSD), the optimized parameters $\delta=0.01$ and $c=0.02$ are selected based on Ref.[8].

The decoding inefficiency is shown in Fig 5. For RSD design, it starts at 1.098 at $K=3\,000$ and gradually drops to 1.058 at $K=10\,000$. The overhead ε approaches a floor near zero with increase of the source size. JDD-LT design has the similar characteristic as RSD design, but the inefficiency is substantially lower. The values are respectively 1.043 at $K=3\,000$ and 1.02 at $K=10\,000$ as for JDD-1, and respectively 1.05 and 1.

029 as for JDD-2. The average overhead for JDD-LT design is half of that for RSD design.

The average degree per symbol is shown in Fig. 6. The values of RSD design start at 15.91 ($K=3\ 000$) and slowly increase to 18.17 ($K=10\ 000$), and those of JDD-LT design are correspondingly 8.496 and 10.01. Note that for the prior stage (JDD-1) the average degree is constantly 5.87. In all, the average number of XOR operations for generating an encoding symbol by JDD-LT design is half of that by RSD design.

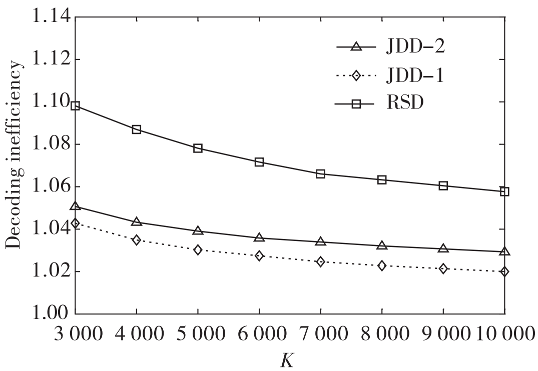

Fig. 5 Decoding inefficiency VS source size

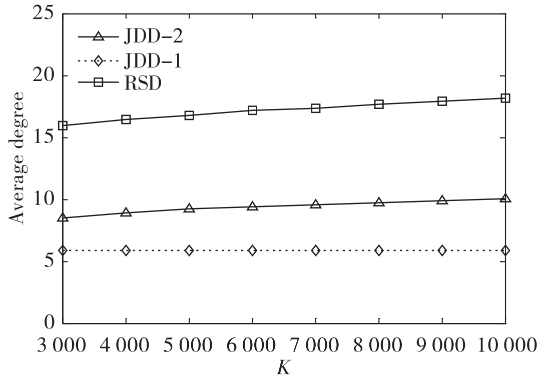

Fig. 6 Average degree VS source size

The number of XOR operations for decoding is shown in Fig. 7. It witnesses about a half decrease on the JDD-2 curve comparing the RSD curve concerning every source size. The lowest point of the JDD-2 curve is 0.268×10^5 ($K=3\ 000$) and the highest point is 1.03×10^5 ($K=10\ 000$), and those of the RSD curve are respectively 0.524×10^5 ($K=3\ 000$) and 1.922×10^5 ($K=10\ 000$). The improvement is mainly contributed by the JDD-1 stage with the weakened degree distribution reducing the overall degree of the encoding symbols. Shown on the JDD-1 curve, the lowest point is 0.184×10^5 ($K=3\ 000$) and the highest point is 0.599×10^5 ($K=10\ 000$). For each of the three curves, the number of XOR operations stably increases in approximate linearity. Note that the slopes are different. From the highest to the lowest are respectively RSD, JDD-2 and JDD-1. It is further implied that the number of XOR operations increases slower with the source size for JDD-LT design than for RSD design.

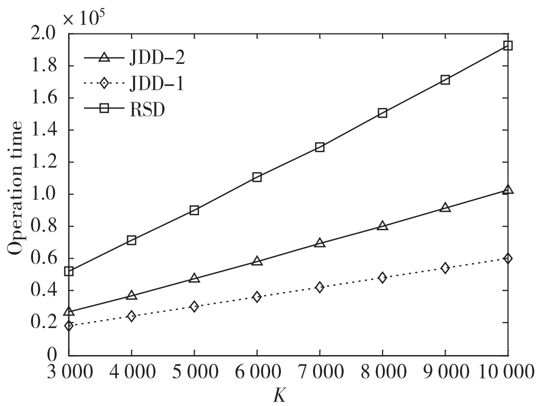

Fig. 7 Decoding speed VS source size

5 Conclusion

This paper proposes a joint degree distribution LT codes design with a weakened degree distribution for the prior encoding stage and an improving degree distribution for the posterior encoding stage. The combination of the two distributions suppresses the overhead while enhancing the encoding and decoding speed. Simulations demonstrate the improvement comparing the single Robust Soliton distribution design.

References:

[1] Luby M. LT codes[C]//Proceedings of the 43rd annual IEEE symposium on foundation of computer science (FOCS02). BC:Vancouver,Canada,2002: 271-280.

[2] Schierl T, Ganger K, Hellge C, et al. SVC-based multi-source streaming for robust video transmission in mobile ad hoc networks[J]. Wireless Communications, IEEE, 2006, 13(5): 96-103.

[3] Chuan W, Baochun L. rStream: Resilient and optimal peer-to-peer streaming with rateless codes[J]. Transactions on Parallel and Distributed Systems, 2008, 19(1): 77-92.

[4] Michael L, Mark W, Tiago G, et al. Raptor codes for reliable download delivery in wireless broadcast systems[C]// Consumer Communications and Networking Conference

2006, 3rd IEEE, Las Vegas: 3rd IEEE, USA, 2006: 192-197.

[5] Bodine E A, Cheng M K. Characterization of Luby transform codes with small message size for low-latency decoding[C]//2008 IEEE International Conference on Communications (ICC'08). Beijing: [s. n.], 2008: 1195-1199.

[6] Bioglio V, Grangetto M, Gaeta R, et al. On the fly Gaussian elimination for LT codes[J]. IEEE Communication Letters, 2009, 13(12): 953-955.

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

[8] Cataldi P, Shatarski M P, Grangetto M, et al. Implementation and performance evaluation of LT and raptor codes for multimedia applications[C]//2006 International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP'06). CA: Pasadena, 2006: 263-266.

一种联合度分布 LT 码设计

马 耀 国

（重庆师范大学 信息技术中心，重庆 401331）

摘要： LT 码是一类前向纠错信道编码，用于纠正信道分组删除（Packet erasure）。这类编码具有广泛的用途，包括计算机科学、网络传输、媒体存储、大文件下载等等。在 LT 码的设计中，度分布是成功解码和快速运算的关键。这篇文章展示了一种新的 LT 码设计，它将弱分布用于前期解码，再将增强型分布用于后期解码。由于弱分布具有低的平均度数，它可以显著地增加编码及解码的速度。同时，增强型度分布具有高的平均度数，能够提高成功解码的概率。通过一系列的仿真，笔者观察到这种设计的编码冗余度和编解码所需的异或运算量比使用 Robust soliton 分布的参考方案降低大约 50%。

关键词： LT 码；度分布；释放速度；冗余度；XOR 运算

（责任编辑　游中胜）