Research on the Selection Method of Output Degree Distributions for LT Codes

Shuang Wu[1(✉)], Chen Cui[2], and Qingyang Guan[1]

[1] College of Engineering, Xi'an International University, Xi'an 710070, China
[2] School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China
cuichen@hit.edu.cn

Abstract. LT codes, the first class of rateless codes, are designed to overcome the packet lossy problems. LT codes with the dynamic overhead to measure the coding efficiency. It is necessary to note that the overhead can be considered as the reciprocal of code rate. Although the overhead is dynamic, but the different overhead corresponding different coding efficiency, which means that in the practical transmission scenarios, the overhead is the lower the better. For this reason, one need to find some other parameters to make the LT codes provide suitable error correction capabilities. As the error performances of LT codes also can be determined by output degree distributions, it is necessary to select the suitable distributions to provide the LT codes can provide enough erasure correcting capability on a lower overhead. For this reason, we derive the symbol error rates and the redundant probabilities of LT codes. Based on the derive results, two constraint conditions of selection of output degree distributions are proposed. And the simulation results shown that these two conditions are valid.

Keywords: LT codes · Rateless codes · Symbol error rate · Redundancy probability · Overhead

1 Introduction

LT codes, the first class of practical fountain codes, which are designed to overcome the packet lossy problems [1]. As LT codes can generate output symbols as much as needed, which code also been known as rateless codes, and have the potential to suit for various channel states [2]. As LT codes can overcome the packet lossy problems, which codes are originally to be used to take place of the transmission protocols based on Automatic Repeat-reQuest (ARQ) mechanisms. And LT codes is also can be used as the inner codes in the concatenate coding schemes [3].

Different with the traditional error correcting codes, LT codes cannot provide bit error correcting capability, but instead of erasure correcting property [4]. For

Q. Guo et al. (Eds.): WiSATS 2021, LNICST 410, pp. 325–334, 2022.
https://doi.org/10.1007/978-3-030-93398-2_32

this reason, there is no need to added too much redundancy data to provide error correcting capability, and the structure of LT codes can generated by using a random manner. For this reason, the LT codes can be flexibility to be the basis of many LT-based coding schemes. For example, as the inner codes of raptor codes [3], to be concatenated with other codes to provide rateless property [5], to provide unequal error protection property [6], and be re-designed to suit for the network communication systems [7,8], etc.

As the structures of LT and LT-based rateless codes are very flexible, which codes also be used in many research fields, for example the access technologies [9], network coding schemes [10] and to provide differentiates to recognize the user on Non-Orthogonal Multiple Access (NOMA) ensembles [11].

Because of the rateless property, the coding efficiency of LT codes are not measured by using the constant termed code rate. Actually, as the output symbols of LT codes can be generated as much as needed, the coding efficiency of LT codes are measured by using a dynamic ratio termed overhead, which is the reciprocal of code rate. As the transmission efficiency in one of the most important performances of the communication systems, which means that the overhead would as lower as better when the decoding processes are finished. Here is a paradox exists, on the one hand the overhead can be as large as needed, on the other hand the overhead would as lower as better. As the decoding processes are finished if the symbol error rate is lower enough to suit for the reliability requirements, and the error performances of LT codes are mainly determined by the output degree distributions, the output degree distributions need to be optimized to suit for the transmission conditions.

As the output degree distribution is the most important parameter in the construction of a LT codes, which have attracted much attentions. Luby have originally provide the idea and robust degree distributions [1], and some practical output degree distributions are given in [3]. Besides these classical studies, there are many other researcher studied on this issue [12,13]. Although there are so many studies focus on the output degree distributions, but in practice the encoders are mainly adopting the pre-designed output degree distributions from [3], because of which degree distributions are all can provide lower overhead performances. Unfortunately, the distributions proposed by [3] are designed on the code length 10^5, but in the practical scenarios, the code length are much lower, for this reason, it is necessary to select the suitable output degree distributions for such scenarios.

To resolve the above problems, by derive the relation between symbol error rate and average degrees, and quantized the redundancy probabilities of output symbols, we proposed two propositions to select output degree distributions, and the simulation results shown that these propositions are worked.

This paper is organized as following. In Sect. 2, we review the LT codes, some necessary definitions are also given. We derived the symbol error rate of LT codes by using a iterative expression, and quantized the corresponding overheads of given symbol error rates in Sect. 3. And the proposed propositions are simulated in Sect. 4, in which the results hold that the propositions are worked. At the lats, the conclusion of this paper is drawn in Sect. 5.

2 Preliminaries

In a LT code, the original data are framed into k input symbols. An output symbol with degree d is generated by XoR d randomly selected input symbols. The degree d is determined by given the output degree distribution $\Omega(x) = \sum_{d=1}^{D} \Omega_d x^d$. The input symbols which are selected in encoding process are termed as been *covered*, and each output symbol with degree 1 is called as *released* symbol. An output symbol and the selected input symbols in whose generating process are *neighbors* for each other. An input symbol can be *recovered* only if one of its neighbors is released. As there are n output symbols have been generated, the *overhead* is defined as $\gamma = \frac{n}{k}$. It can be found that the overhead is the reciprocal of the code rate.

The number of recovered input symbols are usually less than the number of collected output symbols, as an output symbol have the potential to recover at most one input symbol, there is an redundancy probability exists. Actually, an output symbol is redundant is caused of two reasons, the one is which symbol cannot to be processed into an released symbol, the other one is all the neighbors of an output symbols are recovered by their other output neighbors.Let$y$ and $P_{rdu,d}$ denote the symbol error requirement of communication systems and redundancy probability of an output symbol with degree d, then the redundancy probability can be given by following equations.

$$\begin{aligned} P_{rdu,d} &\triangleq P_{reason\ 1} + P_{reason\ 2} \\ &= 1 - d(1-y)^{d-1}y - (1-y)^d + (1-y)^d \\ &= 1 - d(1-y)^{d-1}y. \end{aligned} \tag{1}$$

The input symbols of LT codes are barely recovered if $\gamma < 1$. The *recovery rate* of rateless codes with $\gamma < 1$ is dubbed the *intermediate performance*. To improve the intermediate performance of rateless codes, Growth codes are proposed that have a dynamic output degree distribution $\Omega(x)$ [14]. The design of $\Omega(x)$ for Growth codes is intuitive. The partitions of output symbols with lower degrees would decrease with the increase in the number of output symbols.

3 The Differences Between the Degree of Output Symbols

It is well-known that the decoding performances of LT codes depending on their output degree distributions, but as the decoding performances of LT codes can be measured through various facets, the output degree distributions would affected on LT codes through different aspects.

3.1 On the Erasure Correcting Capability of LT Codes

In most of the scenarios, the data have to be transmitted based on a given reliability requirements. And the symbol error rate of LT codes is the performance which is used to measure the reliability of data, then the erasure correcting capability is the most important performance of a LT code.

By consider a given LT code, the symbol error rate can be measured by using the expected error rate of each input symbol. As the LT codes are processed over the symbol lossy channels, which means the bit error problems can be neglected, and an input symbol can be recovered only if one of its neighbor output symbols can be processed into with degree 1, and the only neighbor of this output symbol is the above input symbol.

As the input and output degree distributions $\Lambda(x) = \sum_{\hat{d}=1}^{D_{int}} \Lambda_d x^d$ and $\Omega(x) = \sum_{d=1}^{D_{out}} \Omega_d x^d$ are given, then for a randomly selected input symbol, whose degree is d', then the expected error rate of this symbol is

$$P_{error,\hat{d}} \triangleq \prod_{i=1}^{\hat{d}} (1 - P_{rel}^{i}), \tag{2}$$

where P_{error} is the expected error rate of the input symbol, and $P_{rel,i}$ is the probability that the ith output neighbor of the input symbol can be processed to released.

And for each output neighbor with degree d, the released probability can be computed by

$$P_{rel,d} = \binom{d}{d-1}(1 - P_{error})^{d-1} = d(1 - P_{error})^{d-1}, \tag{3}$$

where P_{error} is the average error rate of input symbols, and which can be calculated by

$$P_{error} = \sum_{\hat{d}}^{D_{int}} \Lambda_{\hat{d}} P_{error,\hat{d}}. \tag{4}$$

It can be found from Eq. (3), the released probability of an output symbol would decreases with its degree growth.

Assuming the encoding process is strictly random, then Eq. (2) can be rewritten as follows.

$$P_{error,\hat{d}} = \hat{d}(1 - P_{rel})^{\hat{d}-1}, \tag{5}$$

where P_{rel} denotes the expected released probability of output symbols, which can be expressed as

$$\begin{aligned} P_{rel} &= \sum_{d=1}^{D_{out}} \Omega_d P_{rel,d} \\ &= \sum_{d=1}^{D_{out}} \Omega_d d(1 - P_{error})^{d-1} \\ &= \Omega'(1 - P_{error}). \end{aligned} \tag{6}$$

The expected symbol error rate which shown in Eq. (4) can be rewritten by following.

$$\begin{aligned} P_{error} &= \sum_{\hat{d}}^{D_{int}} \Lambda_{\hat{d}} P_{error,\hat{d}} \\ &= \sum_{\hat{d}}^{D_{int}} \Lambda_{\hat{d}} \hat{d}(1 - P_{rel})^{\hat{d}-1} \\ &= \sum_{\hat{d}}^{D_{int}} \Lambda_{\hat{d}} \hat{d}\big(1 - \Omega'(1 - P_{error})\big)^{\hat{d}-1} \\ &= \Lambda'\big(1 - \Omega'(1 - P_{error})\big). \end{aligned} \tag{7}$$

It no hard to seen that for an input symbol, whose error rate would dramatically decreases with its degree growth. Assuming that the overhead in enough large such that $\Omega'(1 - P_{error})$ tends to its maximum value, then P_{error} would also tends to its minimum value.

As $\Lambda(x)$ is generating in the encoding process, and the average degrees of input symbol and output symbols $\bar{\hat{d}}$ and $\bar{d}$ satisfying $\bar{\hat{d}} = \gamma\bar{d}$, we can arbitrary given the following proposition.

Proposition 1. *For LT codes with the same average degree, if the overhead is enough large, which code would share the same symbol error rate.*

3.2 On the Coding Efficiency for LT Codes

As the reliability and efficiency are the two most important target of communication systems, the coding efficiency is also an important performance of LT codes besides the symbol error rate.

For the traditional error correcting codes, the coding efficiency are measured by using the code rate. But for LT code, as its rateless property, the code rates are not fixed, then the coding efficiency of LT codes (also for other rateless codes) are measured by overheads. The overhead is the ratio of the number of output symbols versus the number of input symbols, which is the reciprocal of the code rate. And it is worth to not that a code rate is a pre-designed constant, but the overhead is a dynamic ratio.

As LT codes are proposed to overcome the symbol lossy problems, which codes no need to correcting the bit errors, then the redundant informations are not necessary. For this reason, to obtain high coding efficiency performance, the target overheads of LT codes are as approach to 1.

Then consider a LT code with k input symbols, in which the symbol error rate is given by P_{error}. Then based on Eq. (1), the redundant probability of an output symbol with degree d can be rewritten as

$$P_{rdu,d} = 1 - d(1 - P_{error})^{d-1} P_{error}. \tag{8}$$

Furthermore, the expected redundant probability of this code is

$$P_{rdu} = \sum_{d=1}^{D_{out}} \Omega_d P_{rdu,d}. \tag{9}$$

By substitute Eq. (8) into Eq. (9), the redundant probability of the given LT code is

$$\begin{aligned} P_{rdu} &= \sum_{d=1}^{D_{out}} \Omega_d P_{rdu,d} \\ &= \sum_{d=1}^{D_{out}} \Omega_d - \sum_{d=1}^{D_{out}} \Omega_d d(1-P_{error})^{d-1} P_{error} \\ &= 1 - P_{error}\Omega'(1-P_{error}). \end{aligned} \tag{10}$$

As symbol error rate is a function of overhead γ, the redundancy probability is the function of functions of overhead. And the redundant problems would accumulated in the decoding process. Based on this reason, the following proposition can be given.

Proposition 2. *For a LT code, as the redundancy probability is a function of overhead γ, and the redundancy problem would accumulated in the decoding process, the output degree distributions should satisfying that in the region $\gamma < 1$, the P_{rdu} would as lower as possible.*

4 Simulation Results

In this section, the decoding performances of LT codes with various output degree distributions under asymptotic conditions are compared.

Firstly, we choose the output degree distributions proposed in [3], which are given in following

$$\begin{aligned} \Omega(x) = {} & 0.007969x^1 + 0.493570x^2 + 0.166220x^3 \\ & + 0.072646x^4 + 0.082558x^5 + 0.056058x^8 \\ & + 0.037229x^9 + 0.055590x^{19} + 0.025023x^{64} \\ & + 0.003137x^{66}, \end{aligned} \tag{11}$$

$$\begin{aligned} \Omega(x) = {} & 0.007544x^1 + 0.493610x^2 + 0.166458x^3 \\ & + 0.067900x^4 + 0.089209x^5 + 0.041731x^8 \\ & + 0.050162x^9 + 0.038837x^{19} + 0.015537x^{20} \\ & + 0.016298x^{66} + 0.012677^{67} \end{aligned} \tag{12}$$

and

$$\Omega(x) = 0.0.004807x^1 + 0.496472x^2 + 0.166912x^3 + 0.073374x^4 + 0.082206x^5 + 0.057471x^8 + 0.035951x^9 + 0.001167x^{18} + 0.054305x^{19} + 0.018235x^{65} + 0.009100^{66}. \quad (13)$$

As shown in Fig. 1, the above degree distributions are with average output degree values $\bar{d_1} = 5.8454$, $\bar{d_2} = 5.9708$, $\bar{d_3} = 5.8250$, the code 1 and code 3 with the nearly same symbol error rates in the region $\gamma > 1.1$, and as the code 2 with the average degree value little larger than the other two, which code with the better symbol error rate performance than the others. Which results proof that the Proposition 1 is valid.

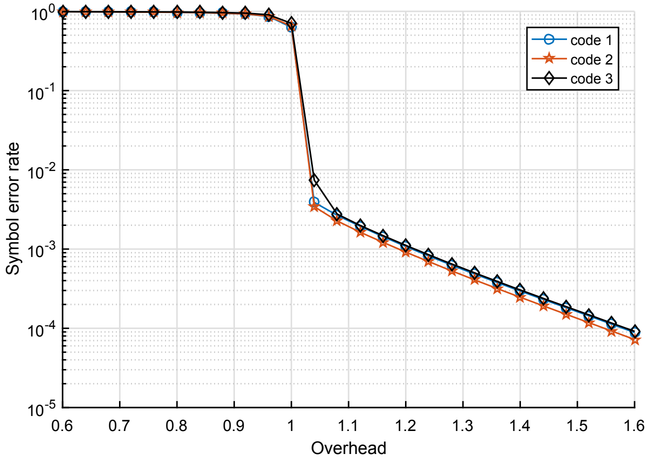

Fig. 1. The SER performances of LT codes with output degree distributions proposed in [3].

Then we consider the LT codes with robust degree distributions [1].As shown in Fig. 2, the codes are with the robust degree distributions with generating parameters $(k = 1000, c = 5, \delta = 0.5)$, $(k = 1000, c = 2, \delta = 0.2)$, $(k = 1000, c = 1, \delta = 0.1)$ and $(k = 1000, c = 0.2, \delta = 0.1)$. As the codes are with the average output degree values $\bar{d_1} = 5.9992$, $\bar{d_2} = 7.1500$, $\bar{d_3} = 7.9598$ $\bar{d_3} = 8.8662$, which values are increased with the sequence numbers growth, then symbol error rate performances of these codes are also as better as its sequence number goes. Which phenomenon also can be used to confirm the Proposition 1 in this paper.

By observing at Fig. 2, it can be found that although in the error flow region, the symbol error rate performances of the 4 codes are related to their average output degree values, but the error flow region are began from the different overhead which much larger than 1. By contrast, in the Fig. 1, the error flow

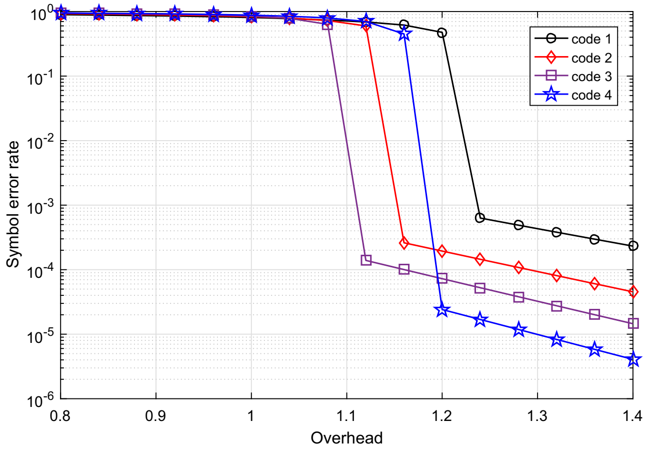

Fig. 2. The SER performances of LT codes with robust degree distributions.

region of the three given codes are began at the nearly same overhead which only a little larger than 1. Which because of the output degree distributions shown in Eqs. (11), (12) and (13) are carefully designed, but the above robust degree distributions are generated by random choosing the parameters. As the overheads are directly corresponding to the coding efficiency of LT codes, then we turned to focus on the redundancy probabilities of output symbols for LT codes.

Figure 3 illustrate the redundancy probabilities of output symbols for LT codes. In which the codes are with the output degree distributions shown in Eqs. (11) and (12), and the other codes are adopted the robust degree distributions with parameters $(k = 1000, c = 5, \delta = 0.5)$ and $(k = 1000, c = 1, \delta = 0.1)$. By observed at Fig. 1 and 2, it is can be found that the codes with output degree distributions proposed by [3] share the better overhead performances than which of the codes by using the robust degree distributions. And it can be found that for the codes share the same decoding overhead performances, whose redundancy probabilities are also the same, and it can be found that in the region $\gamma < 1$, the lower redundancy probabilities would leading to a lower decoding overhead, in other word leading to a higher coding efficiency. Which results illustrated that the redundant problems of output symbols would be accumulated in the decoding process, and the Proposition 2 can be holds.

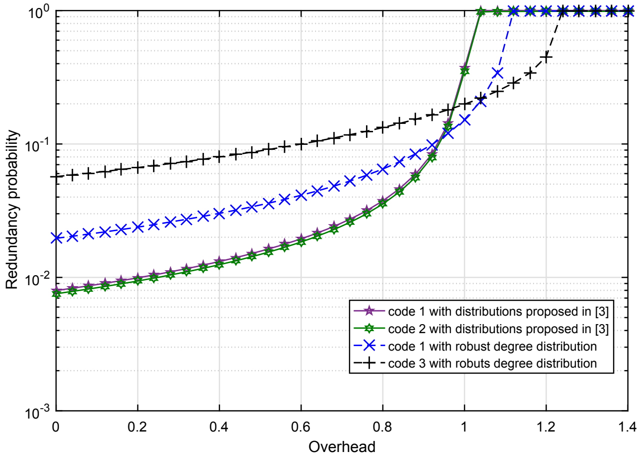

Fig. 3. The redundancy probabilities of output symbols for LT codes.

5 Conclusion

In this paper, we focus on the selection method of output degree distributions of LT codes. By derive the symbol error rate of input symbols and the redundancy of output symbols, two proposition are given. Firstly, the symbol error rate of input symbols would depending on the average degree values of output symbols. Secondly, the coding efficiency of LT codes would depending on the redundancy probabilities of output symbols in the decoding process, and the overall redundancy probability would be a accumulated value. The simulation results shown that the proposed proposition are valid, and which can be used to select an suitable output degree distributions before the encoding processes are beginning.

Acknowledgment. This work was supported by the Scientific Research Initiation Funds for the Doctoral Program of Xi'an International University (Grant No. XAIU2018070102 & XAIU2019002), Regional Innovation Capability Guidance Project (Grant No. 2021QFY01-08) and the General Project of science and Technology Department of Shaanxi Province (Grant No. 2020JM-638).

References

1. Luby, M.: LT codes. In: Proceedings of 43rd Annual IEEE Symposiumon Foundations of Computer Science, pp. 271–280 (2002)
2. Byers, J.W., Luby, M., Mitzenmacher, M.: A digital fountain approach to asynchronous reliable multicast. IEEE J. Sel. Areas Commun. **20**(8), 1528–1540 (2002)
3. Shokrollahi, A.: Raptor codes. IEEE Trans. Inf. Theory **52**, 2551–2567 (2006)

4. Mackay, D.J.C.: Fountain codes. IEE Proc. Commun. **152**(6), 1062–1068 (2005)
5. Li, B., Tse, D., Chen, K., Shen, H.: Capacity-achieving rateless polar codes. In: 2016 IEEE International Symposium on Information Theory (ISIT), pp. 46–50 (2016). https://doi.org/10.1109/ISIT.2016.7541258
6. Rahnavard, N., Vellambi, B., Fekri, F.: Rateless codes with unequal error protection property. IEEE Trans. Inf. Theory **53**(4), 1521–1532 (2007)
7. Talari, A., Rahnavard, N.: Distributed unequal error protection rateless codes over erasure channels: a two-source scenario. IEEE Trans. Commun. **60**(8), 2084–2090 (2012)
8. Gu, S., Yang, Y., Jian, J., Wei, X., Zhang, Q.: Distributed rateless coded collaboration for satellite relay networks. In: International Conference on Wireless Communications & Signal Processing (2015)
9. Shirvanimoghaddam, M., Li, Y., Dohler, M., Vucetic, B., Feng, S.: Probabilistic rateless multiple access for machine-to-machine communication. IEEE Trans. Wirel. Commun. **14**(12), 6815–6826 (2015). https://doi.org/10.1109/TWC.2015.2460254
10. Du, W., Li, Z., Liando, J.C., Li, M.: From rateless to distanceless: enabling sparse sensor network deployment in large areas. IEEE/ACM Trans. Netw. **24**(4), 2498–2511 (2016). https://doi.org/10.1109/TNET.2015.2476349
11. Chen, Y.M., Lai, W.M., Ueng, Y.L.: Rateless coded multiplexing for downlink transmission with two users: performance analysis and system design. IEEE Access **7**, 50440–50452 (2019). https://doi.org/10.1109/ACCESS.2019.2911214
12. Zhu, H., Zhang, G., Li, G.: A novel degree distribution algorithm of lt codes. In: 2008 11th IEEE International Conference on Communication Technology, pp. 221–224 (2008). https://doi.org/10.1109/ICCT.2008.4716207
13. Cheong, S.T., Fan, P.: Novel degree function over finite field for LT codes. In: 2016 International Conference on Electronics, Information, and Communications (ICEIC), pp. 1–5 (2016). https://doi.org/10.1109/ELINFOCOM.2016.7562925
14. Kamra, A., Misra, V.: Growth codes: maximizing sensor network data persistence. In: Proceedings of ACM Sigcomm, Pisa, Italy, pp. 255–266, September 2006