TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol.12, No.7, July 2014, pp. 5469 ~ 5475
DOI: 10.11591/telkomnika.v12i7.4038

Rateless Codes Design Scheme Based on Two Stages

Yunji Li[1]*, Xiaolong Yang[2], Keping Long[2]
[1]School of Communication & Information Engineering, Univ. of Electronic Sci. & Tech. of China
Qingshuihe Campus, No. 2006, Xiyuan Ave, West Hi-Tech Zone, 611731, Chengdu, Sichuan, P. R.China
[2]School of Computer & Communication Engineering, Univ. of Sci. & Tech. of Beijing
No. 30 Xueyuan Road, Haidian District, Beijing, 100083, P. R.China
*Corresponding author, e-mail: yunjili_uestc@sohu.com

Abstract

Current rateless codes coding schemes ignored the order recovery of packets. To cope with this problem, average delay and maximum memory consumption were proposed as performance indices to characterize the order recovery performance, then a coding scheme based on two stages coding was proposed to improve the order recovery performance. Encoding of the front k coded symbols is the first stage, which the i^{th} coded symbol is compose of the i^{th} packet and other d_i-1 packets which are chosen from the front i-1 packets with equal probability. Encoding of the remaining infinite coded symbols is the second stage which the packets are chosen from all packets equal probably. The simulation results show that the rateless codes from the proposed scheme have better order recovery performance, meanwhile have high bandwidth efficiency and good uniformity recovery than LT codes.

Keywords: *rateless codes, order recovery, two stages coding, LT codes, systematic LT codes*

1. Introduction

In wireless networks, automatic retransmission request (ARQ) was commonly adopted to combat channel fading, and enhance communication reliability. However, communication systems with a large number of feedbacks will cause “feedback storm”. Hybrid ARQ was developed with assistance of forward error correction (FEC) code to significantly reduce the frequent feedbacks. More often than not, efficient FEC code needs to match with the channel. However, it is difficult to know the accurate priori knowledge of channel in wireless networks. What's more, it is intractable in network multicast even though the source knows the condition of all channels if they are not identical because the source cannot change the coding scheme according to the conditions of one or some (not all) channels. To overcome the drawback of FEC code, rateless codes were proposed, which can remarkably promote transmission efficiency without any priori channel knowledge.

Rateless codes over lossy channels do not require any priori knowledge of the channels. Therefore, rateless codes are very suitable candidates for lossy multicast channels, nonuniform channels, and time-varying channels. The first realization of rateless codes is Luby Transform (LT) codes [1]. LT codes can be precoded with a block code to yield Raptor codes [2]. LDPC-LT (Low Density Parity Check) and RS-LT (Reed-Solomon) are two common Raptor codes. Therefore, the perfromance of LT codes has important effect on Raptor codes. LT codes with robust soliton distribution (RSD) have good capacity-achievability [1]. The bigger *k* is, the better capacity-achievability performance LT codes have. But LT codes have several drawbacks because the purpose of LT codes is only to improve the transmission efficiency. First drawback is the poor intermediate performance which means that it only recovers few packets when the receiver cannot receive sufficient coded symbols. The poor intermediate performance will affect the real-time application, especially for bigger *k*. The second is the longer average delay, which is also disadvantageous to real-time applications. The third is the serious disorder recovery, which the i^{th} packet is often be recovered before the $(i\text{-}1)^{th}$ packet. The disorder recovery is not conducive to real-time applications. Besides, sometimes, the source cannot get the accurate channel knowledge, but it can get the worst conditions from statistical estimation, history data or other research results. Consideration of worst conditions is applied to network multicast for many nonuniform channels.

Received July 28, 2013; Revised January 27, 2014; Accepted February 16, 2014

In this paper, we propose a rateless codes design scheme to reduce the average delay, disorder recovery, and improve the uniformity recovery and transmission efficiency. Uniformity recovery means that the number of recovered packets from any *i* coded symbols is close to *i*. The uniformity recovery can be explained by the following example. If *k* balls will be distributed to $N \geq k$ boxes while ensure the sum number of balls of the front *i* boxes does not exceed *i* and the N^{th} box must has at least one ball. The uniform distribution of ball is similar to the uniformity recovery of rateless codes. The order recovery means that the i^{th} packet is recovered before the $(i+1)^{th}$ packet. The proposed scheme has two prominent advantages. The first is the better performance of order recovery. The second is the higher transmission efficiency.

This paper is organized as follows. Some related works are described in section 2. Section 3 describes the system model. Maximum memory consumption, average delay and uniformity recovery entropy are defined in section 4. Section 5 proposes a coding scheme to achieve the good performance of above indices. The simulation results are shown and analyzed in section 6. Finally, we draw some conclusions.

2. Related Work

The first drawback of LT codes was studied in [3], where the authors got the outer bound which the fraction of packets that can be recovered for any degree distribution, and designed a degree distribution to get close to this bound. The fraction of packets can be recovered as a performance index has limitation because part of disorderly packets may be unusable in many applications. A scheme has both good intermediate performance and capacity-achievability with smaller number of feedbacks was studied in [4]. Shifted LT (SLT) codes modified the robust soliton distribution of LT codes at the source, based on the number of input symbols already decoded at the receivers [5]. The use of feedbacks is contradictory to the essence of rateless codes and it is infeasible in some applications such as deep-space communication and network multicast. Ali Talari et al optimized the rateless codes with genetic algorithm under the assumption that the source knows the channel erasure rate [6]. The assumption may not applicable in many scenarios and the algorithm is very complex. Ali Talari et al proposed an efficient packets sorting algorithm for high intermediate recovery rate of LT codes [6, 7], which is only to improve the intermediate performance regardless of the order recovery. A family of systematic rateless codes that are universally capacity-approaching on BEC regardless of the channel erasure rate were studied in [8], which only proposed an approach to design systematic LT codes.

Distributed LT codes were proposed in [9, 10], which are to improve the transmission efficiency of rateless codes by design a degree distribution such that the coded symbols received by the destination follow the RSD in degree. The essence of distribution LT codes is multi-sources fusion. Unequal error protection is important in video and image applications. An coding scheme based on expanding window for unequal error protection was proposed in [11, 12], which classified *k* input packets into *Q* different important classes. The scheme guarantees the recovery of the most important packets regardless of order and uniformity recovery of the same important packets. Classes of optimal rateless codes were proposed in [13], which is to design rateless codes based on the tranditional linear block systematic code. In this paper, we use the definition of systematic rateless codes in [2, 8]. The systematic rateless codes of [13] are the special case of our scheme. The rateless codes of [13] have good intermediate performance, order recovery and high transmission efficiency when the erasure rate is low. However, the performance advantage of above is not obvious when the erasure rate is high. Beyond that, the major disadvantage of [13] is that the scheme is not suit to unequal error protection.

3. System Model

We assume that parts of disorderly packets are unavailable because all the packets must be sent to next step or processed in order. In general, only intact data can be used. This assumption is tenable in streaming media, file distribution and some other applications.

If there are *k* packets to be transmitted, the ideal rateless codes are that the receivers can recover all packets with high probability from a slightly greater number of coded symbols, and the *k* packets are recovered evenly and orderly.

We use an example to explain the proposed model. If there are {x_1, x_2, …, x_{10}} packets will be encoded into {y_1, y_2, y_3, …} infinite coded symbols. As shown in Table 1. n_r is the number of received coded symbols. p_r is the recovered packets. We can see that scheme A has lower bandwidth efficiency than scheme B, C and D. Scheme C has poor order recovery performance, and it will consumes more memory to store the disorderly packets. Scheme D has poor uniformity recovery performance, which will causes more average delay. In summary, scheme B has better performance of bandwidth utilization, uniformity and order recovery than other schemes from intuition.

Table 1. Example for Order, Uniformity Recovery and Bandwidth Efficiency

n_r	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A-p_r				x_4			x_7		x_2			x_1,x_8	x_5	x_3,x_6,x_9,x_{10}
B-p_r	x_1		x_2	x_3		x_4	x_5,x_6	x_7	x_8	x_9	x_{10}			
C-p_r	x_{10}	x_9			x_8	x_7	x_6	x_5	x_4	x_2	x_3,x_1			
D-p_r							x_9,x_3	x_8	x_2,	x_4	x_1, ,x_5,x_6,x_7,x_{10}			

The details of RSD can be found in [1]. The probability of degree-1 and degree-2 generated from RSD is very high in a wide range of c and δ, where δ is allowable failure probability of the decoder to recover the data from a given number k of coded symbols. Constant c is almost set as $0<c<1$. It is suitable to choose c and δ as c=0.2, δ=0.05 [14] in the following simulations studies.

4. Three Novel Performance Indices

Based on the proposed ideal rateless codes model, we propose three performance indices to estimate the rateless codes of different coding schemes. Maximum memory consumption and average delay characterize the disorder. Uniformity recovery entropy (URE) characterizes the uniformity recovery performance. The three novel indices and overhead can comprehensively describe the performance of rateless codes. The overhead of rateless codes is defined as the number of output coded symbols that the receiver needs to collect in order to recover the input messages with high probability, and it is measured as a multiple of the number k of input symbols [1].

4.1. Maximum Memory Consumption

It is difficult to know which and how many packets are recovered when the receiver receives a coded symbol because the rateless codes are random codes. The randomness will result in disorder recovery of packets in the receiver. The number of disorderly packets can be used as disorder measurement. The rateless codes communication is a dynamic process because the receiver simultaneously receives and decodes the coded symbols. The disorderly packets will be stored and waiting until some of them are order, which cause the number of disorderly packets is changing. Here, we only consider the memory consumption for the recovered packets. Maximum memory consumption is the accumulation of the number of disorderly packets of decoding process, which can reflect the disorder degree. We explain the maximum memory consumption with scheme A in Table 1. When the receiver receives 5 symbols, packet x_4 is recovered and stored, and waiting for the packet x_1, x_2 and x_3. So x_4 is a disorderly packet. The number of disorderly packets is 2 when the receiver receives the eighth coded symbol. When the receiver receives 13th symbol, packet x_1 and x_8 are recovered. The memory consumption is 5 because the new recovered packets will be stored for sequencing. After the sequencing, the packet x_1 and x_2 will be sent to next step, and the memory consumption becomes 3. The number of disorderly packets is 4 when the receiver receives the 14th symbol, and then it becomes 8 when the receiver receives the 15th symbol. Therefore, the maximum memory consumption of scheme A is 8. And so on, the maximum memory consumption of scheme B, C and D in Table 1 are 2, 10, 10 respectively. Therefore, scheme B has less maximum memory consumption than others.

4.2. Average Delay

Real-time is important in many applications. Delay is a good parameter to characterize the real-time. We define delay of the i^{th} packet as the follows:

$$dl_i = j - i \tag{1}$$

Where j is the recovery time of the i^{th} packet, which the receiver cannot recover the i^{th} packet until it receives the j^{th} coded symbol. If $i > j$, the i^{th} packet will wait for the recovery of all former i-1 packets. Average delay is defined as formula (2):

$$dl = \frac{1}{k}\sum_{i=1}^{k} dl_i \tag{2}$$

The average delay of scheme A, B, C and D in Table 1 are 6.8, 2.1, 4.1 and 5.2 respectively. So, scheme B has the least average delay among the four schemes.

4.3. Uniformity Recovery Entropy

Entropy is a physics terminology which used to estimate the confusion degree of system. In information, entropy is used to estimate the amount of information. In this paper, we use uniformity recovery entropy (URE) to estimate the uniformity of the distribution of recovered packets. We assume that the receiver recovers k_i packets when it receives i coded symbols, and recovers k_{i+1} packets when it receives i+1 coded symbols. The k_{i+1}-k_i packets are the new recovered when the receiver receives the $(i+1)^{th}$ coded symbols. All k packets are recovered when r coded symbols are received. URE is defined as follows:

$$URE = -\sum_{i=0}^{r-1} \frac{k_{i+1} - k_i}{k} \log \frac{k_{i+1} - k_i}{k}, 0\log 0 = 0, k_0 = 0 \tag{3}$$

From intuition, the scheme B and C has better uniformity recovery in Table 1. The URE of schemes A, B, C and D are 1.6094, 2.1640, 2.1640 and 1.3592 respectively. The scheme B and C have better uniformity recovery performance than scheme A and D. The calculations and intuition are consistent. Therefore, URE is a suitable index to estimate the uniformity of the recovered packets.

Form above, we can see that scheme B has the best performance among the four schemes. The computation results of three novel performance indices are similar to the intuition judgement. Our objective is to design a coding scheme has the characters as scheme B.

5. Enoding Scheme with Maximum Channel Erasure Rate

The encoding is divided into two stages. The first stage is the encoding of the front k coded symbols. Encoding of the remaining infinite coded symbols is the second stage. At first, the source obtains degree probability distribution $u(i)$ from RSD, generates k degrees according to $u(i)$, and arranges them from the smallest to the largest order, then computes the average degree d_{av} of the k degrees. In the first stage, the source gets the i^{th} ($1 \le i \le k$) degree d_i and chooses d_i-1 packets from $\{x_1, x_2, \ldots, x_{i-1}\}$ equiprobably, then encodes the d_i-1 packets and the i^{th} packet into the i^{th} coded symbol with XOR operation. In the second stage, the source chooses an integer d_j ($k<j$) equiprobably from ceil[$d_{av}(1-p_m)$] to ceil[$p_m d_{av}+k/2(1-p_m)$] as the degree, then chooses d_j packets equiprobably from k packets and encodes them into the j^{th} coded symbol with XOR operation. The source repeats the second stage until all receivers receive sufficient coded symbols to recover all packets. Here, p_m is the maximum channel erasure rate. Ceil[A] is a mathematical function which rounds the elements of A to the nearest integer which greater than or equal to A. The details of the coding scheme are shown as Figure 1.

The use of RSD is in order to take advantage of the capacity-achievability of RSD. The ratio of small degrees of the front *k* coded symbols is high. We arrange the small degrees in front because the coded symbols which have small degree are easy to be decoded. The recovered packets will be used in the subsequent decoding. The packets of the i^{th} coded symbol are from the first to the i^{th} packet, which may improve the order recovery. After step 3, the receiver has received $n \le k$ coded symbols and recovered $n_r \le n$ packets. If it chooses a small degree d_j in step 4 and chooses d_j packets, the probability that d_j chosed packets have been recoverd may be so high that the coded symbol comprose of the d_j packets is redundancy. On the contrary, if d_j is very large, it may increase the decoding difficulty for $d_s - n_r \gg 1$. So we choose the degree from ceil[$d_{av}(1-p_m)$] to ceil[$p_m d_{av}+k/2(1-p_m)$]. The lower bound and upper bound of the degree in the second stage are inversely proportional to p_m. The degree range is from 0 to d_{av} when p_m=1,and from d_{av} to k/2 when p_m=0. The average upper bound for many times of computation of degree is about k/2 according to RSD. So we chose d_{av} and k/2 as the base lower bound and upper bound respectively, and then adjust the bound according to p_m.

Algorithm Two stages Encoding Algorithm for Rateless Codes

Input: k, c, δ.

1. Generate degree probability distribution $u(i)$, i=1, 2 ,... , k according to RSD with given δ and c.

2. Generate k degrees according to $u(i)$, and arrange them from the smallest to the largest order, get $D=\{d_1, d_2, \ldots, d_k\}$, where $d_{i-1} \le d_i$, for $\forall$ i=2, 3, …, k, i=1.

3. **while** $i \le k$

   x_i must be chosen. Then choose d_i-1 packets from $\{x_1, x_2, \cdots, x_{i-1}\}$ equal probably, and encode the d_i packets into the ith coded symbol with XOR operation. $i=i+1$.

   end

4. Choose an integer d_j randomly from ceil[$d_{av}(1-p_m)$] to ceil[$p_m d_{av}+k/2(1-p_m)$] as the degree, then choose d_j packets from $\{x_1, x_2, \cdots, x_k\}$ equal probably. And then encode the d_j packets into the jth coded symbol with XOR operation.

5. Repeat step 4 until all receivers receive sufficient coded symbols to recover all packets with high probability 1-δ.

Figure 1. Encoding Algorithm

6. Simulation

The rateless codes from the proposed scheme are systematic rateless codes with very high probability if it chooses suitable parameters in RSD because the probability of degree-1 always is so high that it can generate many degree-1 even though the generation process is random. Another advantage of the proposed scheme is that the proposed scheme can recover all *k* packets orderly when the erasure rate is zero.

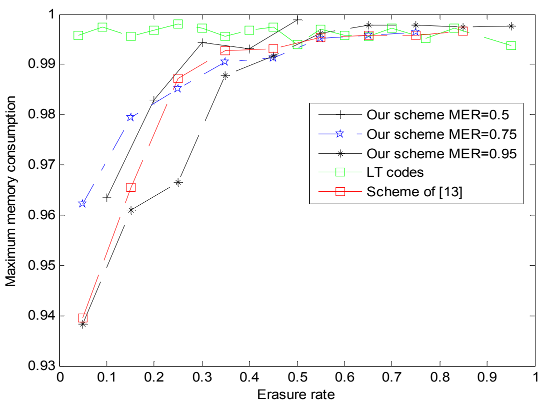

Figure 2. Maximum Memory Consumption

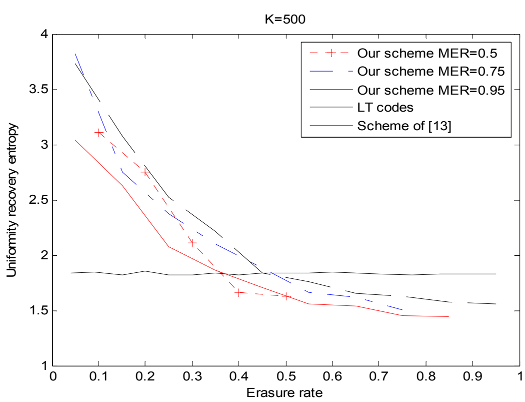

Figure 3. Uniformity Recovery Entropy

We simulated the maximum memory consumption, overhead, average delay and URE under k=500 and p_m=0.5, 0.75, 0.95, c=0.2, δ=0.05 [14]. In the simulation, the decoding algorithm is Gaussian elimination which can decode all packets thoroughly more than BP (belief propagation) algorithm. The results of simulation are illustrated in Figure 2-5. For convenience, the maximum memory consumption is normalized. From Figure 2, we can see that the proposed scheme is superior in maximum memory consumption than LT codes, and is approximate to the scheme of [13]. The less maximum memory consumption it has, the more packets are recovered orderly. The advantage is very prominent when the erasure rate is low. The maximum memory consumption of LT codes is always very much in a very wide range of erasure rate. The maximum memory consumption of the proposed scheme increases with the increase of erasure rate. The advantage of maximum memory consumption of the proposed scheme is due to the degree arrangement and packets choice mechanism in the i^{th} ($1 \le i \le k$) coded symbol. The small degrees are arranged in front, which the coded symbols with small degree are very easy to be decoded. The recovered packets will contribute to the subsequent decoding. The degree arrangement mechanism will be beneficial to the reduction of disorder. The packets of the i^{th} coded symbol is from the front i packets, which make the latter packets appear in the latter coded symbols. So, the packets choice mechanism also is in favor of the reduction of disorder. Although the scheme of [13] has better good decoding performance in the front k coded symbols, the proposed scheme and the scheme of [13] has similar performance because of the advantage of the proposed scheme and channel erasure.

The UREs for LT codes, the proposed scheme and the scheme of [13] are shown in Figure 3. The URE of the proposed scheme is descended with the increase of erasure rate and bigger than that of LT codes and the scheme of [13] under different erasure rates. The bigger URE is, the better performance of uniformity recovery and intermediate performance the scheme has. But the LT codes has poor uniformity recovery performance under different erasure rates. The improvement of uniformity recovery of our scheme is due to the arrangement of the front k degrees for that the coded symbols with small degree are easy to be decoded and the previous recovered packets will help to the subsequent decoding. So, there are more recovered packets in any front i coded symbols.

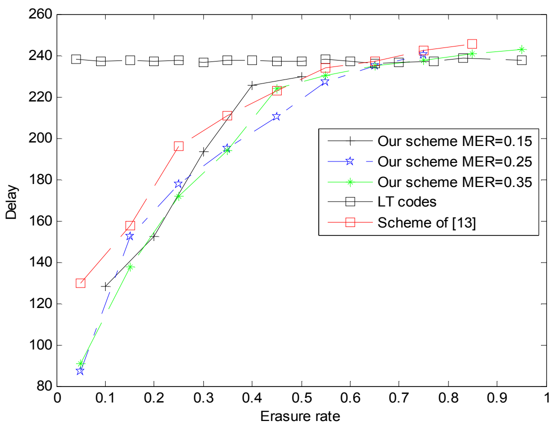

Figure 4. Average Delay

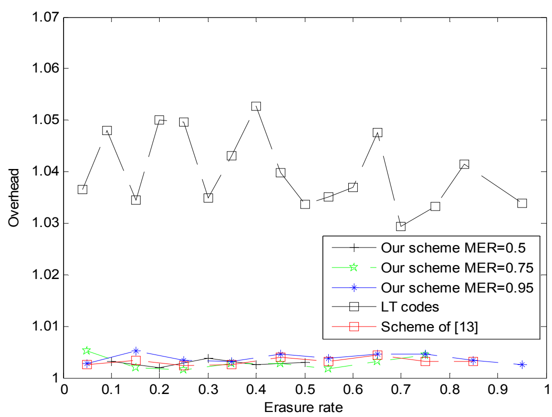

Figure 5. Overhead

The average delay is shown in Figure 4. The proposed scheme has less average delay than that of LT codes and the scheme of [13] under different erasure rates. The average delay of LT codes is very long in a wide range of erasure rates. The advantage of the proposed scheme is very prominent when erasure rate is low, and the average delay is increased with the increase of erasure rate. The average delay gain profits from the packets selection of the i^{th} ($1 \le i \le k$) coded symbol and the degree arrangement mechanism of the front k coded symbols, which is similar to the reduction of maximum memory consumption.

Figure 5 shows the overheads of the rateless codes from the three coding schemes. The overhead of our scheme is decreased with the increase of k and is lower than that of LT codes, and is similar to that of [13] under different erasure rates. The overhead gain of the proposed scheme and the scheme of [13] is due to the linearly independence of the front *k* coded symbols, which means that the front *k* coded symbols have no redundancy information. But the LT codes cannot ensure the linear independence of the front *k* coded symbols.

From above discussion, we can see that the proposed scheme has better performance of uniformity recovery, memory consumption, average delay and overhead than LT codes in a wide range of erasure rates. Beyond that, our scheme can generate systematic code with high probability. Our scheme also is superior to or similar to the scheme of [13].

7. Conclusion

In this paper, we propose a novel rateless codes design scheme based on two stages coding and forward equal probability. The encoding of the front *k* coded symbols is the first stage which the i^{th} coded symbol is compose of the i^{th} packet and d_i-1 packets which are chosen from the first to the $(i\text{-}1)^{th}$ packet with equal probability. The second stage chooses a degree d_j randomly from ceil[$d_{av}(1\text{-}p_m)$] to ceil[$p_m d_{av}+k/2(1\text{-}p_m)$] and then chooses d_j packets from the *k* packets. The source repeats the second stage until all receivers have decoded all *k* packets. The simulation results show that the proposed coding scheme has less average delay, memory consumption, overhead and better uniformity recovery than LT codes without feedbacks. The proposed scheme has better uniformity recovery performance than the scheme of [13]. The performance of overhead, average delay and maximum memory consumption of the proposed scheme is superior to or similar to the scheme of [13]. Besides, all packets of the rateless codes of the proposed scheme are recovered orderly when the erasure rate is zero.

Acknowledgements

This work is supported by Chang Jiang Scholars Program of the MoE of China, 973 Program (No. 2012CB315905), NSFC (Nos. 61201089, 61071101, 61172050, 61172048, 61173149, 61100184), NCET Program, Sichuan Youth Sci. & Tech. Found (No. 09ZQ026-032), Fundamental Research Funds for the Central Universities (No. ZYGX2009J005).

References

[1] M Luby. *LT-codes.* The 43rd Annu. IEEE Symp. Foundations of Computer Science. 2002: 271-280.
[2] A Shokrollahi. Raptor Codes. *IEEE Transactions on Information Theory*. 2006; 52(6): 2551-2567.
[3] Sujay Sanghavi. *Intermediate Performance of Rateless Codes.* Information theory workshop 2007. Tahoe City, CA. 2007: 478-482.
[4] Saejoon Kim, Seunghyuk Lee. Improved Intermediate Performance of Rateless Codes. 11th International Conference on Advanced Communication Technology. Phoenix Park; 2009; 3: 1682-1686.
[5] Andrew Hagedorn, Agarwal S, Starobinski D, et al. Rateless coding with Feedback. *IEEE INFOCOM* Rio de Janeiro. 2009: 1791-1799.
[6] Ali Talari, Behzad Shahrasbi, Nazanin Rahnavard. *Efficient Symbol Sorting for High Intermediate Recovery Rate of LT Codes*. Information Theory Proceeding 2010. Austin, TX. 2010: 2443-2447.
[7] Ali Talari, Nazanin Rahnavard. On the Intermediate Symbol Recovery Rate of Rateless Codes. *IEEE Transactions on Communications*. 2012; 60(5): 1237-1242.
[8] Xiaojun Yuan, Li Ping. On Systematic LT Codes. *IEEE Communications letters*. 2008; 12(6): 681-683.
[9] Srinath Puducheri, J Kliewer, Thomas E Fuja. The Design and Performance of Distributed LT Codes. *IEEE Transactions on Information Theory.* 2007; 53(10): 3740-3754.
[10] Srinath Puducheri, J Kliewer, Thomas E Fuja. *Distributed LT Codes. Information Theory.* IEEE International Symposium on. Seattle, WA. 2005: 987 - 991.
[11] Dino Sejdinovi'c, Dejan Vukobratovi'c, et al. Expanding Window Fountain Codes for Unequal Error Protection. *IEEE Transactions on Communications.* 2009; 57(9): 2510-2516.
[12] Dejan Vukobratovic, Vladimir Stankovic, Dino Sejdinovic., et al. Scalable video multicast using expanding window fountain codes. *IEEE Trans. Multimedia.* 2009; 11(6): 1094-1104.
[13] Vijay G Subramanian, Douglas J Leith. *On a class of optimal rateless codes.* 46th Annual Allerton Conference on Communication, Control, and Computing. Urbana-Champaign, IL. 2008: 418-425.
[14] Pingyi Fan. Network information theory. Beijing: Tsinghua University Press. 2009: 151-168.