Improving the data recovery for short length LT codes

Nadhir Ibrahim Abdulkhaleq[1], Faeza Abbas Abed[2], Ihsan Jabbar Hasan[3], Falah Hasan Mahdi[4]
[1]Department of Mobile Computing and Communication Engineering-College of Engineering, University of Information Technology and Communication, Baghdad Al Mansour-Iraq
[2,4]Department of Electronic Techniques, Institute of Technology-Middle Technical University, Iraq
[3]Department of Electronic, Medical Technical Institute/Al Mansour-Middle Technical University, Iraq

Article Info	ABSTRACT

Luby Transform (LT) code is considered as an efficient erasure fountain code. The construction of the coded symbols is based on the formation of the degree distribution which played a significant role in ensuring a smooth decoding process. In this paper, we propose a new encoding scheme for LT code generation. This encoding presents a deterministic degree generation (DDG) with time hoping pattern which found to be suitable for the case of short data length where the well-known Robust Soliton Distribution (RSD) witnessed a severe performance degradation. It is shown via computer simulations that the proposed (DDG) has the lowest records for unrecovered data packets when compared to that using random degree distribution like RSD and non-uniform data selection (NUDS). The success interpreted in decreasing the overhead required for data recovery to the order of 25% for a data length of 32 packets.

Corresponding Author:

Nadhir Ibrahim Abdulkhaleq,
Department of Electronic Techniques
Middle Technical University,
Al Zaafaraniyah, Baghdad-Iraq.
Email: Nadir66raouf@gmail.com

1. INTRODUCTION

The word fountain describes the nature of these codes. An infinite stream of coded packets is sent via the channel without information about the channel conditions. With such way of coding the rate of the code is not fixed and is set while processing the received packets. Two codes represent the first practical fountain codes, Luby transform (LT) [1], and Raptor codes [2]. These codes are designed for erasure channels which fit the nature of internet communication systems [3]. fountain like codes can decode all source packets when sufficient number of encoded packets are received [1]. For the encoding of LT codes first, a degree d is randomly chosen from a known degree distribution. Then, d source packets are uniformly chosen at random to be XORed to form the coded packet. This procedure continues until an acknowledgement is received which declared complete recovery of the information packets.

The universal shape of the coded packets helps the receiver to gather any N coded packets. With each coded packet, the receiver has the knowledge about the dgree,indicies of the data packets that forming the coded packet. The means for sending such data are explained in [1]. In Figure 1, an illustration for a bipartite graphical representation of LT encoding and decoding operations. In this prototype code, the data length of $k = 5$, and code-word length of $N = 6$ with packet size of one-bitlpacket is represented. Figure 1-a is devoted to the encoding operation whereas Figure 1-b up to Figure 1-g are used to illustrate the decoding operation. For the decoding operation to be successful, the existence of degree-one coded packet is vital. At every stage of the decoding operation this degree-one coded packet has to be found. Once a degree-one coded packet is found then its connecting information packet is recovered. Secondly, the recovered value updates the values of the neighbor coded packets. Finally, the connections of

the recovered information packet were released. Moreover, this procedure is repeated until all the information packets are recovered.

It is obvious from understanding the decoding process, that the generation of degrees is the core for successful data recovery. Many attempts are presented to modify the conventional scheme of degree generations. The first design for degree distribution used was the ideal soliton distribution (ISD) and robust soliton distribution (RSD) [1]. The features of the RSD are suitable for bulk data files [1, 2, 4]. Esa Hyytiä et al [5] used a sampling approach to determine the optimized degree distribution. They found that for large information sequences, i.e., thousands of packets, the RSD shows good performance. In [6] the authors investigated the optimization criteria of degree distributions for small file sizes. They concluded that, in a well-design distribution just a few parameters require to be controlled to insure almost maximal performance. Chen Z and Zhou Q [7], suggested an LT code with a revised RSD used for degree generation based on using Kent chaotic map and pseudo-random number generator. Their idea based on joining both of the probability functions $\rho(i), \tau(i)$ but with no attention for the δ parameter. The result of this new degree generation is shown in having more small degree values. C.M. Chen et al in [8] used a covariance matrix adaptation as evolution strategy (CMA-ES) to enhance the degree values for better performance.

Their approach for the new distribution generates the degrees adaptively without changing the main parameters. In [9] the author presents another attempt to improve the ability of degree generation to get better encoding-decoding frame. They enforce the distribution to have an expected number of degree-one coded packets as a division of data length. A state of the art work on fountain codes was done in [10] where a memory-based rateless encoding method is proposed. In this new data allocation method, the degree one coded symbol is isolated and connected to the data packets with the highest connection. The previous literature review gives a bright vision for the concern to modify the RSD as an optimal degree distribution. As a new attempt in this paper, a deterministic manner for both, the generation of the degrees and the choice of the data packets which form the coded packets is introduced.
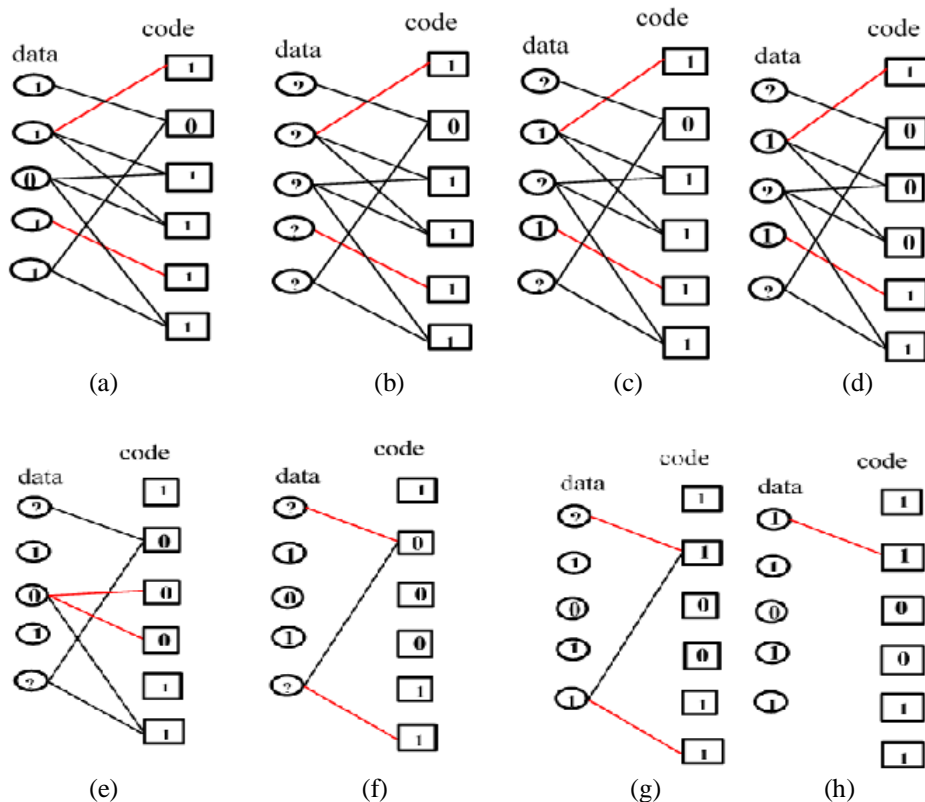

Figure 1. Encoding and decoding operations for lt code ($k = 5, N = 6$)

In the next section a review for degree generations is presented. Section 3 is dedicated for our deterministic degree generation. Testing the ability of this new degree generation to enhance the code performance is achieved in section 4. The last section is used to conclude the whole idea in few condensed lines.

2.   DEGREE GENERATION APPROACHES

LT code design mainly depend on the quality of the degree generation method. The design has two features, the degree distribution and the way of data selection. The degree generation in all conventional approaches for LT code design are using random degree generation and also the data selection is applied randomly. After the introduction of ideal soliton distribution and robust soliton distribution by Luby in [1] many other studies [3, 5-9] on RSD were performed. And all these studies focused on improving RSD, i.e., on improving the right-hand side distribution. While in [10] Khaled et al introduced a memory based rateless encoding operation yielding non-uniform left hand side distribution which shows better performance in terms of BER. In this paper, we propose a new approach for rateless encoding operation. The proposed approach is based on deterministic construction of the generator matrix with non-uniform left hand and right hand side distributions. Below we discuss, the degree generation methods that been used as a comparing tool for our proposed one.

2.1. Ideal soliton distribution

This distribution offers an ideal distribution but not practical, because at certain point in the decoding process a degree one coded symbol could be lost with high probability. It is defined as:

$$\rho(d) = \begin{cases} \frac{1}{k}, d = 1 \\ \frac{1}{d(d-1)}, d = 2, \dots \dots, k \end{cases} \tag{1}$$

where d is an integer represent the degree and k is the length of information message.

2.2. Robust soliton distribution

In this modified degree distribution, the average number of degree one coded packets has been increased [1]. The robust soliton distribution used the $\rho(d)$ in (1) to construct its new distribution as:

$$\mu(d) = \frac{\tau(d) + \rho(d)}{\beta} \tag{2}$$

where the normalization factor $\beta = \sum_{d=1}^{k}(\rho(d) + \tau(d))$ and $\tau(d)$ is given as:

$$\tau(d) = \begin{cases} \frac{R}{dk}, d = 1,2,..,\frac{k}{R} - 1 \\ \frac{R}{k} \ln\frac{R}{\delta}, d = \frac{k}{R} \\ 0, d = \frac{k}{R} + 1, \dots., k. \end{cases} \tag{3}$$

where $R = c. \ln(k/\delta)\sqrt{k}$) is the average number of coded packets with degree one. The constant c has the value between 0 and 1, while δ represents the probability of decoder failure.

2.3. Non-uniform data selection (NUDS)

Approximately all the previous researches worked on improving the right-hand distribution. However, in [10] a memory based encoding approach, resulting in better performance with the help of memory tracking for the data packets degrees. Below, the algorithm proposed in [10] is outlined.

Algorithm 1 (NUDS)
1: Repeat
2: Generate a degree d using RSD.
3: if $d = 1$
4: Select the data symbol with
   largest degree, i.e., select the data
   symbol with the most number of
   connections.
5: else
6: Select d data packets uniformly.
7: XOR the selected d data symbols to
 form the code symbol
8: end if
9: Until (Acknowledgement signal is received destination).

3.   DETERMINISTIC DEGREE GENERATION (DDG)

RSD is considered as an optimal degree distribution for large file sizes, but it may suffer from certain drawbacks when dealing with short size data files [11]. The reason for that is the amount of extra coded packets that have to be collected to insure successful decoding. This extra needed packet for LT codes has been approximated to be $5\% \, for \, k = 10000$, while it could be enlarged for small message sizes to reach $30\% \, for \, k = 100$,[12]. Sevral attempts have been presented in the field of improving the communication systems using channel coding [13-18]. In addition, motivating from the decoding treatment which has done by the pattern recognition (PR) [19-25] to resume decoding when the decoder halt, and the work introduced by [10] for the NUDS, we propose a deterministic encoding (DDG) approach for LT like codes. The proposed approach is designed to be used especially for short frame lengths. In our deterministic degree generation, in case of having k data packets, the degrees are generated following this formula:

$$degree(C_n) = \begin{cases} n, n = 1,2, \dots R_p \\ n - m \cdot R_p, n = R_p + 1, \dots N \end{cases} \quad (4)$$

Where R_p is the repetition period value chosen to be between $(3 \leq R_p \leq (k/2))$, m is an integer take the values ($1,2,3, \dots k/R_p$) and N is the size of the code length. For the information sequence consisting of k symbols, i.e., for $(u_1 \, u_2 \, u_3 \cdots u_{k/2} \, u_{k/2+1} \cdots u_k)$, with $\left(R_p = \frac{k}{2} \, and \, m = 1,2\right)$ the coded symbols $(c_1 \, c_2 \, c_3 \cdots c_k)$ are formed using:

$$\begin{aligned} c_1 &= u_1, \\ c_2 &= u_1 \oplus u_2, \\ c_3 &= u_1 \oplus u_2 \oplus u_3, \\ c_{k/2} &= u_1 \oplus u_2 \oplus u_3 \dots \oplus u_{k/2} \\ c_{k/2+1} &= u_{k/2+1}, \\ &\vdots \\ c_k &= u_{k/2+1} \oplus u_{k/2+2} \oplus u_{k/2+3} \dots \oplus u_k \end{aligned} \quad (5)$$

It is clear from (5) that the generator matrix is $(k \times k)$, however the extra needed coded packets are generated using (4) with repetition of the same allocation of the data symbols illustrated in (5). More immunity for such encoding scheme in case of bad channel conditions and in order to prevent losing the same data packet, a Time-Hopping-Encoding scheme is used to produce a Time-Hopping-encoding scheme. In this scheme, the encoding matrix is altered between two shapes, one is the original deterministic (G_{od}) and the other is the random shuffle of it (G_{rd}), as shown in Figure 2:
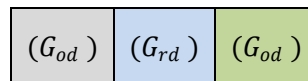
$$(G_{od}) \quad (G_{rd}) \quad (G_{od})$$

Figure 2. Deterministic degree generation frame

Where $T_1 , T_2 \, and \, T_3$ can be adjusted as required by the transmitter. The proposed approach is explained in Algorithm 2:

Algorithm 2 Deterministic Degree generation (DDG)
1: Repeat
2: for $d = 1: k/2$
3: if $d = 1$
4: $c_1 = u_1$
5: else
6: $c_d = u_1 \oplus u_2 \cdots \oplus u_d$
7: end
8: end
9: for $d = k/2 + 1: k$
10: if $d = k/2 + 1$
11: $c_{k/2+1} = u_{k/2+1}$
12: else

13: $c_d = u_{k/2+1} \oplus u_{k/2+2} \cdots \oplus u_d$
14: end
15: end
16: Until (Acknowledgement is obtained from receiver)

4. DECODING IN THE ABSENCE OF DEGREE-ONE CODE PACKET

LT-like codes suffer from the absence of degree-one coded packets during the decoding operation especially for short length messages. If degree-one coded packet is not found at any stage of the decoding operation, decoding failure is declared. Our work in [19], succeeded in breaking the decoder failure in a way that even there is no degree-one coded symbol decoding operation can be continued by fetching for certain configuration patterns inside the remaining coded symbols matrix. For this Pattern Recognition (PR) proposed approach, assume that there is no degree-one coded symbol at any stage of the decoding operation but there are two code-words c_i and c_j which are formed as:

$$c_i = u_2 \oplus u_4 \oplus u_6$$
$$c_j = u_2 \oplus u_4 \oplus u_6 \oplus u_{10} \tag{6}$$

where it is obvious that the degree of the code-words c_j can be reduced to '1' using:

$$c_j = c_i \oplus c_j \tag{7}$$

which result is $\deg(c_j) = 1$. Getting motivation from this approach, we added further fetching steps by trying to reduce the degree of the coded symbols which are contained in each other, algorithm 3 illustrate this approach, [19]:

Algorithm 3 Decoding Stuck Removal Approach
1: for iteration number=1: last % last may be any number (2, 3, 4)
2: Repeat
3: $i = 1$; $flag \leftarrow check$
4: $c_i \leftarrow set\ as\ reference\ coded\ symbol$
5: for any other code-words c_j other than c_i
6: if (deg ($c_j \oplus c_i$) equal 1, then
7: $c_j \leftarrow c_i \oplus c_j$; flag $\leftarrow success$
8: break; resume BP decoding
9: end if
10:end-for
11: $i = + +$,return to step 4
12: if flag $\neq$ success repeat steps (4-10) to decrease the degrees and replace
step 6 to (deg($c_j \oplus c_i$)<deg (c_j))
13: end if
14: until (All degrees are tried.)
15: end-for

5. RESULTS AND ANALYSIS

The LT code performance using BP algorithm is compared in an erasure channel environment using three types of degree distributions which are DDG, RSD and NUDS. For the RSD the parameters are chosen as $c = 0.02$, $\delta = 0.1$. The message length is chosen as 32 packets (the packet length could be with any length, in our simulation it is represented by 1 bit). For the binary erasure channel erasure probability $\alpha = 0.08$ is employed. Computer simulations are performed such that for each rate point on the graph (($R = k/N$) is checked until 100 erroneous frames are received. It is seen in Figure 3 that the proposed method is better than that of the LT-RSD and NUDS in terms of the number of unrecovered data packets for all code rates.

The performance improvement for LT code is obvious for all distribution types under study with the proposed decoding stuck removal algorithm BP-PR as shown in Figure 4. It is clear from Figure 4 that the performance improvement for LT-RSD and NUDS is more compared to that on DDG. However, DDG still gets the best performance. Another performance comparison graph is given in Figure 5 where success rate is defined as:

$$Success\ Rate = \frac{successful\ recovered\ frames}{Total\ number\ of\ the\ file\ frames} \qquad (8)$$

It is clear from Figure 5 that DDG approach recovers all the information frames at a rate of $R = \frac{k}{n} \rightarrow \frac{1}{2}$ while the other distributons have approximately the same performance and they do need to collect more overhead to reach the achievement of our DDG.
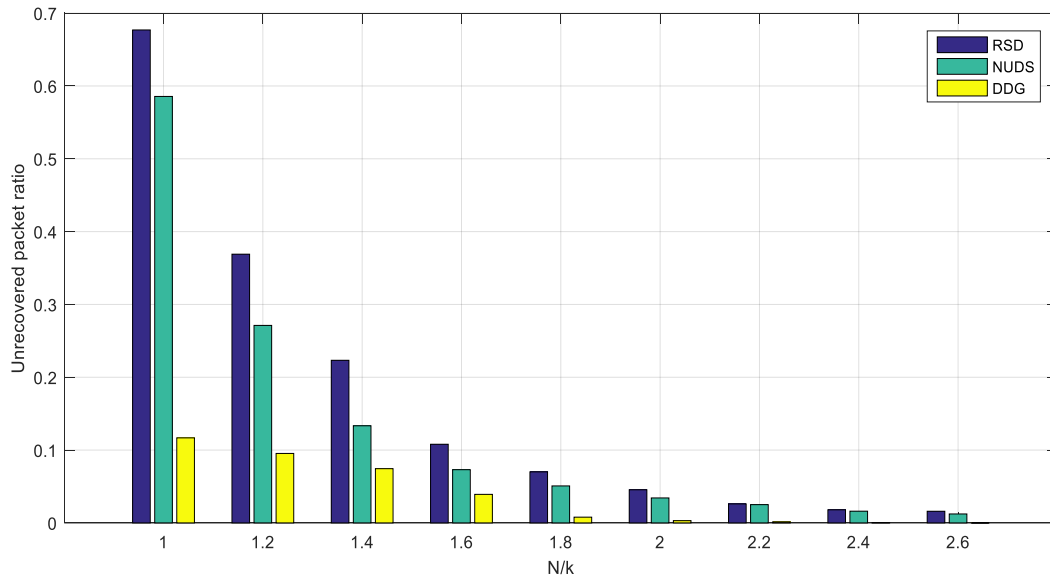

Figure 3. Unrecovered packets results, using BP, RSD of $(c\ = 0.02\ and\ \delta\ =\ 0.1), k\ =\ 32, \alpha\ =\ 0.08)$
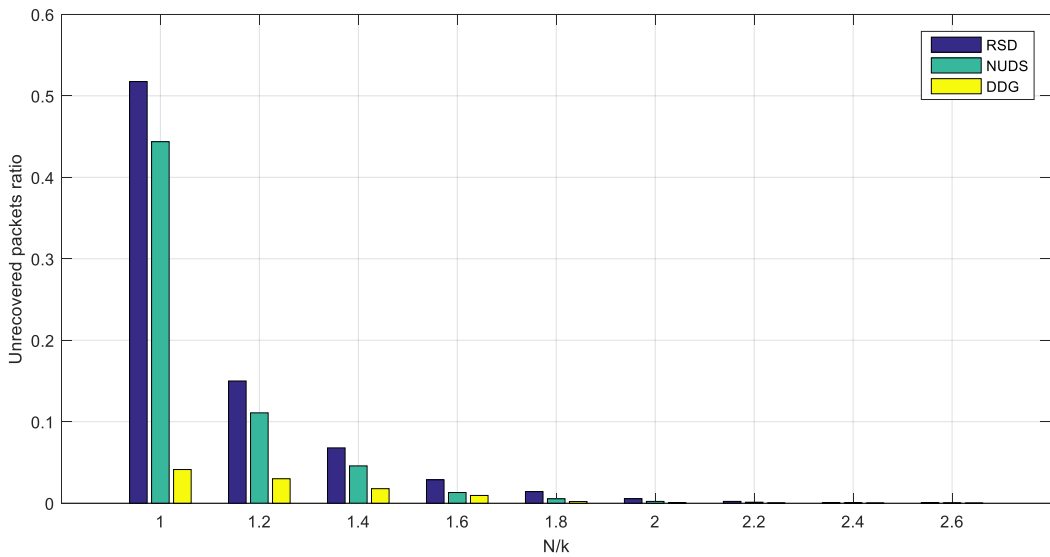

Figure 4. Unrecovered packets results, using BP-PR, RSD of $(c\ = 0.02\ and\ \delta\ =\ 0.1), k = 32, \alpha\ =\ 0.08)$
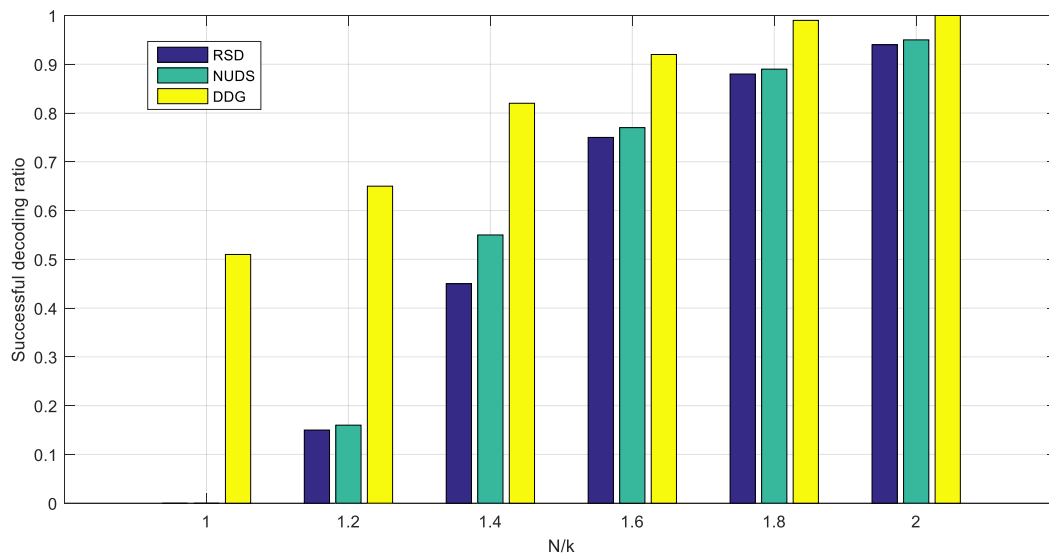
Figure 5. Successful decoding rates, using BP-PR, RSD of ($c = 0.02\ and\ \delta = 0.1$),$k = 32, \alpha = 0.08$)

6.   CONCLUSION

In this paper we proposed a time hopping deterministic degree generation encoding approach for the forming of LT-like codes especially for small data files. It is shown via simulation results that the proposed approach is better in terms of minimum number of unrecovered packets and overhead amount required for full recovery of data packets than that of the classical LT-RSD and NUDS. The flexibility features of the proposed approach support the design to overcome the problem of extra needed overhead for the case of short length LT codes. For the decoding operation BP algorithm is employed supported by the enhancement algorithm of pattern recognition (BP-PR). It has been approved that BP-PR has little improvement to the proposed DDG-LT code compared to the clear improvement shown for the classical RSD and NUDS-LT codes.

REFERENCES

[1]   Luby M., "LT codes," *In The 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002, Proceedings*, *IEEE*, pp. 271-280, 2002.
[2]   Shokrollahi A., "Raptor codes," *IEEE/ACM Transactions on Networking (TON)*. vol. 1(14), pp. 2551-2267.
[3]   Jafarizadeh S, & Jamalipour A., *"An exact solution to degree distribution optimization in LT codes*," In IEEE Wireless Communications and Networking Conference (WCNC), IEEE, pp. 764-768, 2014.
[4]   MacKay DJ, "Information theory, inference and learning algorithms," *Cambridge University Press*, 2003.
[5]   Hyytiä E, Tirronen T, Virtamo J., *"Optimizing the degree distribution of LT codes with an importance sampling approach,"* In 6th International Workshop on Rare Event Simulation, pp. 56-66, 2006.
[6]   Hyytia E, Tirronen T, Virtamo J., *"Optimal degree distribution for LT codes with small message length*," In 26th IEEE International Conference on Computer Communications, IEEE, pp. 2576-2580, 2007.
[7]   Chen Z. & Zhou Q., *"Implementation of it codes with a revised robust soliton distribution by using kent chaotic map*," In 2010 International Workshop on Chaos-Fractal Theories and Applications, IEEE, pp. 149-153, 2010.
[8]   Chen C. M., Chen Y. P., Shen T. C., Zao J. K., "On the optimization of degree distributions in LT code with covariance matrix adaptation evolution strategy," In *IEEE Congress on Evolutionary Computation*, *IEEE,* pp. 1-8, 2010.
[9]   Yen K. K., Liao Y. C., Chen C. L, Chang H. C.*, "Adjusted robust soliton distribution (ARSD) with reshaped ripple size for LT codes*," In 2013 International Conference on Wireless Communications and Signal Processing, IEEE, pp. 1-6, 2013.
[10]   Hayajneh KF, Yousefi S, Valipour M., "Improved finite-length Luby-transform codes in the binary erasure channel," *IET Communications*, vol. 9(8), pp. 1122-1130, 2015.
[11]   Zao J. K, Hornansky M, Diao P. L., *"Design of optimal short-length LT codes using evolution strategies*," In 2012 IEEE Congress on Evolutionary Computation, IEEE, pp. 1-9, 2012.
[12]   Wang X, Willig A, Woodward G., *"Improving fountain codes for short message lengths by adding memory*," In 2013 IEEE Eighth International Conference on Intelligent Sensors, IEEE, pp. 189-194, 2013.

[13] Sukar M. & Pal M., "SC-FDMA & OFDMA in LTE physical layer," *International Journal of Engineering Trends and Technology (IJETT)*, vol. 12(2), pp. 74-85, 2014.

[14] Sanghvi A, Mishra N, Waghmode R, Talele K., "Performance of reed-solomon codes in AWGN channel," *International Journal of Electronics and Communication Engineering*, vol. 4(3), pp. 259-266, 2011.

[15] Hussain G. A. & Audah L., "RS codes for downlink LTE system over LTE-MIMO channel," *TELKOMNIKA Indonesian Journal of Electrical Engineering*, vol. 16(6), pp. 2563-2569, 2018.

[16] Adiono T, Ramdani A, Putra R., "Reversed-trellis tail-biting convolutional code (RT-TBCC) decoder architecture design for LTE," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8(1), pp. 198-209, 2018.

[17] Wang J., Li J., Cai C., "Two novel decoding algorithms for turbo codes based on taylor series in 3GPP LTE system," *TELKOMNIKA Indonesian Journal of Electrical Engineering*, vol. 12(5), pp. 3467-3475, 2014.

[18] Sirait F., Dani A., Pangaribowo T., "Capacity improvement and protection of LTE network on ethernet based technique," *TELKOMNIKA Telecommunication Computing Electronics and Control*, vol. 16(1), pp. 200-209, 2018.

[19] Abdulkhaleq N. I., Gazi O., "Decoding of lt-like codes in the absence of degree-one code symbols," *ETRI Journal*, vol. 38(5), pp.896-902, 2016.

[20] Abdulkhaleq N. I., & Gazi O., "*A sequential coding approach for short length LT codes over AWGN channel,*" in 4th International Conference on Electrical and Electronic Engineering (ICEEE), IEEE, pp. 267-271, 2017.

[21] Abdulkhaleq N. I., Hasan I. J., Salih N. A., "*Mitigation of packet erasure using new fountain code design,*" In 2018 IEEE International Multidisciplinary Conference on Engineering Technology (IMCET), IEEE, pp. 1-7, 2018.

[22] Hasan I. J., Salih N. A., Abdulkhaleq N. I. "*An Android smart application for an Arduino based local meteorological data recording*" 2019 IOP Conf. Ser.: Mater. Sci. Eng. Vol. 518, 2019.

[23] Salih N. A., Hasan I. J., Abdulkhaleq N. I., "Design and implementation of a smart monitoring system for water quality of fish farms," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, Vol. 14, No. 1, pp. 45-52, April 2019.

[24] Leftah H. A., Alminshid H. N.., "Channel capacity and performance evaluation of precoded MIMO-OFDM system with large-size constellation," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, Vol. 9, No. 6, pp. 5024~5030, December 2019.

[25] Leftah H. A., Boussakta, S*., "Efficient modulation scheme for OFDM system with ZP and MMSE equalizer,*" In 2013 IEEE International Conference on Communications (ICC), IEEE, pp. 4703-4707, 2013.