Defence Science Journal, Vol. 71, No. 6, November 2021, pp. 772-776, DOI : 10.14429/dsj.71.15047

Blind Recognition of Parameters of Reed Solomon Code from Intercepted Erroneous Codewords

Anand Sharma* and Praneesh Gupta

DRDO-Scientific Analysis Group, Delhi - 110 054, India
**E-mail: anandsharma2@rediffmail.com*

ABSTRACT

Error correcting codes are designed for reliable transmission of digital information over a noisy channel. Several papers have been published on blind identification of binary FEC codes but papers reported on the identification of non-binary error correcting codes are less. Due to its strong error correction capability, RS (Reed-Solomon) code is being used widely. So technique for blind recognition of RS code is required to analyse intercepted signal as well as for intelligent communication. This paper presents a technique for extraction of parameters of Reed-Solomon code from intercepted demodulated bitstream. The proposed algorithm is very simple and hence it is very practical for hardware implementation. Our approach has been verified using MATLAB simulation.

Keywords: Non-cooperative; Error correcting codes; CER; SER; GCD

1. INTRODUCTION

Figure 1 shows different blocks of digital communication system. Source output is fed to encoder to convert into bitstream that is encrypted through encryption block to achieve secure transmission. Channel coding block calculates redundancy that is appended to improve reliability of message transmitted over channel. In cooperative communication, receiver is aware of all the parameters of modulator, channel encoder and encryptor. Using these parameters, channel decoding is performed after demodulation to recover correct message bits. Minimum acceptable BER at demodulator output is 1%[1]. Finally, the message bits is decrypted and decoded to get transmitted message.

The technique presented in this paper is useful in non-cooperative context of military applications. In this context, adversary tries to access message exchanged between sender and receiver without knowing parameters of any blocks of transmitter. It is difficult to get error free bitstream from the signal intercepted over noisy channel. Bitstream alignment is another challenge in traffic analysis. Adversary has to align codewords and recognise FEC so as to retrieve error free message bits. As shown in Fig. 1, our approach will work on the erroneous bitstream obtained after demodulating intercepted noisy signal.

Several papers have been published on reverse engineering of different error correcting codes. A lot of results are published on the recognition of convolution code[2-4], while some results are published on recognition of block codes. A Valembois has published an approach for detection and recognition of binary linear block code[5]. His approach was based on recognition of dual codewords but this approach was not successful when dual code has large weight. Various authors[6-9] have proposed methodology for blind reconstruction of binary cyclic codes. Cluzeau[10] and Chabot[11] reconstructed parity check matrix of the code but restriction was that weight of the dual should be very low. Ref[12] explored probabilistic approach for separating message bits from parity bits and extracting codeword size and message size from intercepted noisy bitstream provided input message bits has biased probability. These techniques are useful for binary linear block code of smaller codeword size and can not be used for Reed-Solomon code. A method for blind recognition of RS code was first addressed by Liu[13], *et al.*. They followed Galois Field Fourier Transform (GFFT) based approach that was complex. An entropy based approach is published by Chen[14], *et al.* but its performance is poor for high code rate. Blind recognition of RS coding parameters is also proposed in[15] which work gracefully for

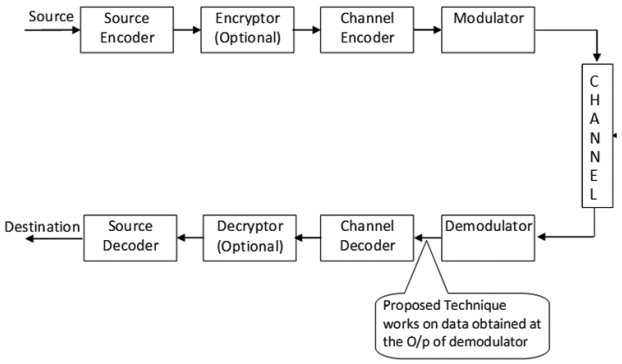

Figure 1. Applicability of work in communication.

Received : 05 September 2020, Revised : 10 March 2021
Accepted : 07 April 2021, Online published : 22 October 2021

error free RS encoded bitstream but it fails in recognising generator polynomial for RS encoded noisy bitstream. Similarly, recognition algorithm described Liu[16], *et al.* works for error free RS encoded bitstream only. Later, Li[17], *et al.* proposed Galois field columns Gaussian elimination based approach for recognition of RS code which works for code error rate of $3x10^{-3}$ with more than 90% probability. Due to its poor error performance, this approach may not be useful in practice. Swaminathan[18], *et al.* proposed algorithms for blind recognition of RS encoder based on rank reduction of data matrix and showed results for QAM modulated signal. Authors also published mechanism for joint reconstruction of RS encoder and interleaver[19] and for product code incorporating RS and BCH as component code[20], through same approach. But the limitation of rank reduction approach is that this can work for low code error rate as discussed above. Hui[21] proposed method using histogram statistic of Galois field spectra to detect RS codes from noisy encoded bitstreams. Author has claimed to achieve 90% success for RS(255,223) encoded bitstream corrupted with 0.5% Code Error Rate (CER).

We propose a method to reconstruct RS code. Here assumption is that we have RS coded binary strings transmitted over binary symmetric channel. The aim is to extract parameters of error correcting code so that intercepted demodulated bitstream can be correctly decoded to recover message. Our approach is effective in noisy environment for up to 90% CER.

2. MATHEMATICAL MODEL OF REED-SOLOMON CODE

Cyclic code is an important subclass of linear codes. These are attractive because of its rich algebraic structure and due to existence of various practical methods for decoding them efficiently[22-24].

Reed-Solomon (RS) codes come under non-binary class of cyclic codes, such that every symbol is of m-bits, where m is positive integer > 2. The codeword length n of an RS code is $q-1$, with $q = 2^m$ being the alphabet size of the symbols.

For $RS(n,\ k)$ code,

$$(n,k) = (2^m - 1, 2^m - 1 - 2t) \quad (1)$$

$$(n-k) = 2t$$

where t is the number of symbols that can be corrected by the code.

The number of symbols in which the two codewords of non-binary code differ is called the distance between codewords. Reed-Solomon code has the largest minimum distance for all linear codes with the same encoder input and output block length. The minimum distance for Reed-Solomon code is expressed by

$$d_{\min} = n - k + 1 \quad (2)$$

RS codes are used mainly for burst errors. The RS code can correct any combination of t or fewer errors, where t is given as

$$t = \frac{d_{\min} - 1}{2} = \frac{n-k}{2} \quad (3)$$

The generating polynomial $g(x)$ of a t symbol error correcting RS code with symbols from $GF(2^m)$ has $\alpha, \alpha^2, \cdots, \alpha^{2t}$ (α is the primitive element in $GF(2^m)$) as all its roots. The generating polynomial $g(x)$ can be written as:

$$g(x) = (x-\alpha)(x-\alpha^2)\cdots(x-\alpha^{2t}) \quad (4)$$

$$\text{or, } g(x) = g_0 + g_1 x + g_2 x^2 + \cdots + x^{2t} \quad (5)$$

where, $g_i \in GF(2^m)$

Generator polynomial of Reed-Solomon code has following properties:

i. Generator polynomial is monic.
ii. Generator polynomial has consecutive roots $\alpha, \alpha^2, \cdots, \alpha^{2t}$.
iii. Generator polynomial is a factor of $1 + x^n$.
iv. Degree of generator polynomial is $n - k = 2t$.

If $c(x)$ is encoded codeword polynomial of a message polynomial $M(x)$, using RS code (n,k) and generator polynomial $g(x)$, then

$$c(x) = x^{n-k} M(x) + \mathrm{Re}\, m[x^{n-k}, g(x)] \quad (6)$$

where, $M(x) = M_0 + M_1 x + M_2 x^2 + \cdots + M_{k-1} x^{k-1}$

and $c(x) = c_0 + c_1 x + c_2 x^2 + \cdots + c_{n-1} x^{n-1}$

Here all the coefficients are in the extension field as the code is non-binary.

Equation (6) can also be expressed as:

$$c(x) = M'(x) \times g(x) \quad (7)$$

where $M'(x)$ is some transformation of $M(x)$.

At receiver end, syndrome (S) of the received code word is calculated. If this syndrome comes out to be 0, it means that the received codeword is error free. But if the syndrome is non-zero, this means received codeword is erroneous and then this error is corrected.

3. PROBLEM DEFINITION

3.1 Assumptions

Our approach to estimate the parameters of FEC code assumes that Reed-Solomon code has been used and intercepted signal has been successfully demodulated to obtain erroneous RS coded bitstream.

3.2 Problem Statement

Proposed methodology assumes that the intercepted signal has been successfully demodulated to get erroneous encoded bitstream and that the FEC code used is Reed-Solomon code.

Subject to above assumptions, we present an approach to analyse given bitstream to extract following in non-cooperative context:

i. Identify symbol size (m)
ii. Identify code word size (n)
iii. Identify message size (k)
iv. Identify generator polynomial (g) used for encoding

These extracted parameters, then, can be used by the adversary to reconstruct corrected message intended for recipient.

4. PROPOSED METHOD

This section describes the proposed technique to identify various parameters of unknown systematic Reed-Solomon error correcting code. This approach exploits the algebraic structure of the code words.

From equation (7) mentioned in Section 2, it is clear that code polynomial is a multiple of the generator polynomial $g(x)$. So in majority of the cases, common factor of pair of valid code polynomials will be the generator polynomial of the Reed-Solomon code used for transmission,

i.e.

If $c_1(x) = a_1(x) \times g(x)$ and $c_2(x) = a_2(x) \times g(x)$

Then,

$$GCD(c_1(x), c_1(x)) = g(x) \times d(x) \quad (8)$$

where, $g(x)$ is generator polynomial of the code while $d(x)$ is residual polynomial whose multiplication with $g(x)$ gives Greatest Common Divisor (GCD) of code words. The GCD is calculated in $GF(2^m)$.

The flowchart of the proposed method is shown in Fig. 2. For range of expected symbol size, m (e.g. from 3 to 8), we obtain list of GCDs calculated over several pairs of noisy code polynomial (out of N codeword) and tabulate this for all m. Then, we filter out invalid values of generator polynomial by applying following conditions:

i. Discard the case when $deg(g(x)) < 2$
ii. Discard the case when $k = 1$
iii. Discard the case when $k \geq n$
iv. Discard the case when $deg(g(x))$ is odd

Then, applying majority logic over filtered GCD values, we obtain generator polynomial of the RS code as the one which occurs a maximum number of times i.e.

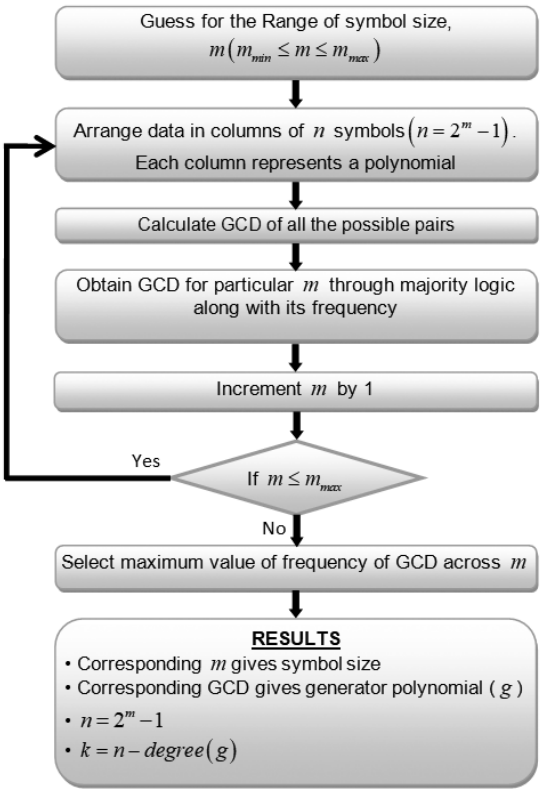

Figure 2. Flowchart of proposed technique.

$$i = N-1,\ j = N-1$$
$$g(x) = majority\ \{filter\ [GCD(c_i, c_j)]\}$$
$$i = 1,\ j = 1,\ i \neq j$$

Once the generator polynomial $g(x)$ is obtained, we calculate the value of k from the $g(x)$ as below:

$$k = n - \deg(g(x))$$

The above proposed technique takes aligned erroneous codewords for analysis but it can also handle unaligned as well as error-free codewords.

5. VALIDATION OF THE TECHNIQUE

To prove that the proposed technique works as per claim made here, MATLAB 7.9 has been used. A code was written to generate a sample file containing random symbols whose values ranges from 0 to 2^m. This file was then encoded with known $RS(n,k)$ code to get encoded file. Then, desired error (CER) is introduced to make it noisy. This is then fed as input to our technique to extract desired parameters i.e. symbol size (m), code word size (n), message size (k) and generator polynomial $(g(x))$. This exercise is repeated for 100 sample files to plot result for analysis as described in next section.

6. RESULTS AND ACHIEVEMENTS

Results of our technique based on simulation in MATLAB are plotted as shown in Figs. 3-6. These plots show results for *RS* (7,3), *RS* (15,7), and *RS* (63,55) encoded erroneous bitstream for different CER. Figure 3 shows plot for successful recognition of symbol size for *RS* (63,55). Here x-axis corresponds to expected value of symbol size (m) while y-axis corresponds to frequency of valid generator polynomial. This is plot for test case of *RS* (63,55) and it shows clearly that first peak (maximum frequency of generator polynomial) occurs when expected value of symbol size matches with the actual symbol size (i.e. m=6 in this case).

Figures 4, 5 and 6 show plots for success rate of our technique for identification of symbol size (m), code word size (n), message size (k) and generator polynomial $(g(x))$ for different CER for *RS* (7,3), *RS* (15,7), and *RS* (63,55). Here x-axis corresponds to code error rate in percentage while y-axis corresponds to the success rate.

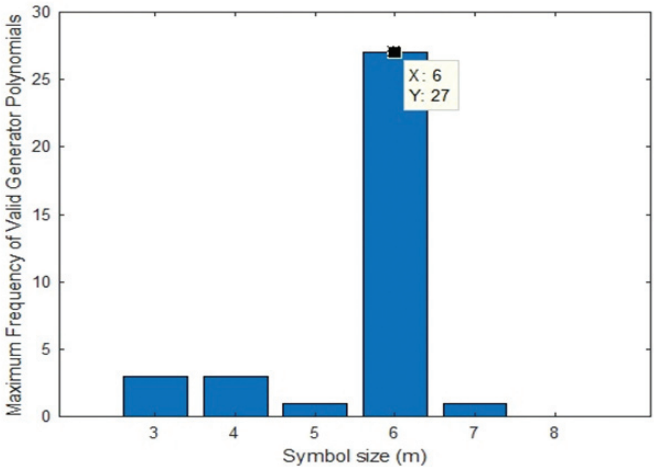

Figure 3. Plot for recognition of *m* for *RS*(63, 55).

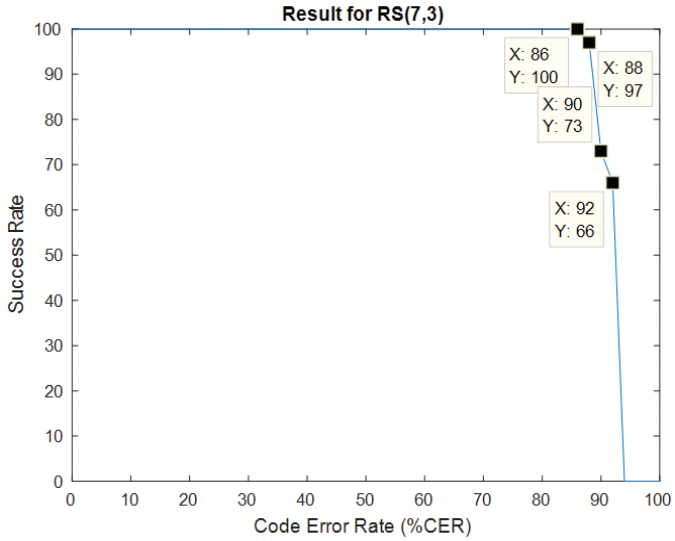

Figure 4. Plot of success of the proposed technique for *RS*(7, 3).

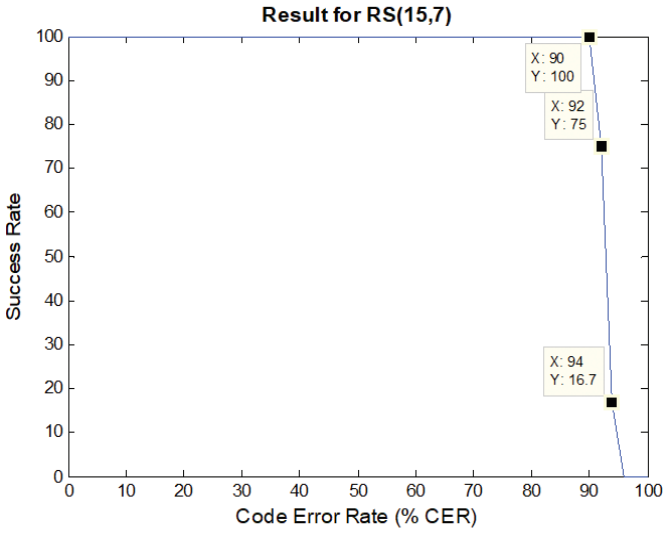

Figure 5. Plot of success of the proposed technique for *RS*(15, 7).

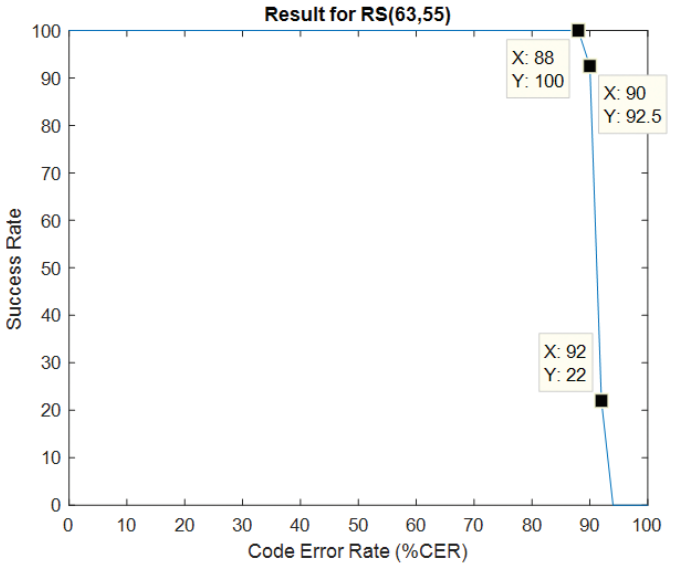

Figure 6. Plot of success of the proposed technique for *RS*(63, 55).

From Figs. 4, 5 and 6, it can be concluded that our approach works with 100% success rate for up to 80% CER. Success rate degrades beyond this CER.

We have also compared result of proposed technique with approach presented in[19,20], which are given in Table 1. For comparison and proving improvement over prior works, we have calculated success rate of our approach for various Symbol Error Rate (SER) and tabulated these alongwith the success rate of prior works. CER corresponding to each SER has also been calculated to compare our results with recent published works.

Table 1 indicates that proposed work in this paper performs better than approach presented earlier.

Table 1. Performance comparison of the proposed technique for *RS*(15, 7) code with prior works

SER	CER	Success rate of technique proposed in[19]	Success rate of technique proposed in[20]	Success rate of our approach
0.01	0.141	1.0	1.0	1.0
0.03	0.374	1.0	1.0	1.0
0.05	0.541	1.0	1.0	1.0
0.06	0.609	0.93	1.0	1.0
0.08	0.715	0.9	0.9	1.0
0.09	0.756	0.6	0.6	1.0
0.1	0.798	0.2	0.2	1.0

7. CONCLUSION

This paper presents a novel technique for reconstruction of RS code from intercepted demodulated erroneous RS coded bitstream. Our technique has been proved on different Reed-Solomon error correcting codes through simulation. This technique is simple to implement and works even at high error rate as is also clear from the plot given Figs. 4, 5 and 6. Our approach outperforms published techniques as explained in section 6. Presented approach is successful in recognition of RS code parameters from intercepted traffic in non-cooperative scenario.

REFERENCES

1. Haykin, S.S. *Digital communications*. Wiley, New York, 1988.
2. Wang, F.; Huang, Z. & Zhou, Y. A method for blind recognition of convolution code based on euclidean algorithm. *In* 2007 International Conference on Wireless Communications, Networking and Mobile Computing, IEEE, Shanghai, 2007, 1414-1417. doi: 10.1109/WICOM.2007.358
3. Dingel, J. & Hagenauer, J. Parameter estimation of a convolutional encoder from noisy observations. *In* 2007 IEEE International Symposium on Information Theory, IEEE, France, 2007, 1776-1780. doi: 10.1109/ISIT.2007.4557147
4. Marazin, M.; Gautier, R. & Burel, G. Dual code method for blind identification of convolutional encoder for cognitive radio receiver design. *In* 2009 IEEE Globecom Workshops, IEEE, Honolulu, HI, USA, 2009, 1-6. doi: 10.1109/GLOCOMW.2009.5360726

5. Valembois, A. Detection and recognition of a binary linear code. *Discrete Applied Mathematics*, 2001, **111**(1-2), 199-218.
   doi: 10.1016/S0166-218X(00)00353-X
6. Wang, J.; Yue, Y. & Yao, J. Statistical Recognition Method of Binary BCH Code. *Commun. Netw.*, 2011, **3**(1), 17-22.
   doi: 10.4236/cn.2011.31003
7. Jing, Z.; Zhiping, H.; Shaojing, S. & Shaowu, Y. Blind recognition of binary cyclic codes. *EURASIP Journal on Wireless Communications and Networking*, 2013, **2013**(1),1-17.
   doi: 10.1186/1687-1499-2013-218
8. Yardi, A.D.;Vijayakumaran, S. & Kumar, A. Blind reconstruction of binary cyclic codes. *In* European Wireless 2014; 20th European Wireless Conference, Barcelona, Spain,2014, 1-6.
9. Yardi, A.D.; Vijayakumaran, S. & Kumar, A. Blind reconstruction of binary cyclic codes from unsynchronized bitstream. *IEEE Trans. Commun.*, 2016, **64**(7), 2693-2706.
   doi: 10.1109/TCOMM.2016.2561931.
10. Cluzeau, M. Block code reconstruction using iterative decoding techniques. *In* 2006 IEEE International Symposium on Information Theory, IEEE, 2006, 2269-2273.
11. Chabot, C. Recognition of a code in a noisy environment. I*n* 2007 IEEE International Symposium on Information Theory, IEEE, USA, 2007, 2211-2215.
    doi: 10.1109/ISIT.2006.261971
12. Sharma, A. & Pillai, N.R. Blind recognition of parameters of linear block codes from intercepted bit stream. *In* 2016 International Conference on Computing, Communication and Automation (ICCCA), IEEE, India, 2016, 1262-1266.
    doi: 10.1109/CCAA.2016.7813910
13. Liu, J.; Xie, N. & Zhou, X.Y. Blind recognition method of RS coding. *J. University of Electron. Sci. Technol. China*, 2009, **38**(3), 363-367.
    doi: 10.3969/j.issn.1001-0548.2009.03.011
14. Chen, J.J. & Yang, J.A. Blind parameters identification approach for low code-rate linear block code based on code weight information entropy. *J. Circuits Syst.*, 2012, **17**(1), 41-46.
15. Chen, W.; Liu, J.; Zhou, X.; Wu, N.; Gao, X.; Wu, S.; Du, Y.; Shang, Q.; Zhou, L.; Jia, T.Y. & Zan, J.J. Blind identification method of coding parameters of RS code of error-tolerant code. Chinese patent CN101534168A, 16 September 2009.
16. Lin, Q.I.; Shiqi, H.A.O. & Jinshan, L.I. Recognition method of RS codes based on Euclidean algorithm in Galois field. *J. Detection Control*, 2011, **33**(2), 63-67.
17. Li, C.; Zhang, T.Q. & Liu, Y. Blind recognition of RS codes based on Galois field columns Gaussian elimination. *In* 2014 7th International Congress on Image and Signal Processing, IEEE, 2014, 836-841.
    doi: 10.1109/CISP.2014.7003893
18. Swaminathan, R.; Kumar, A.M.; Wang, G. & Ting, S.K. Parameter identification of Reed-Solomon codes over noisy environment. *In* 2017 IEEE 86th Vehicular Technology Conference (VTC-Fall), IEEE, 2017, 1-5.
    doi: 10.1109/VTCFall.2017.8287916
19. Swaminathan, R., Madhukumar, A.S.; Wang, G. & Kee, T.S. Blind reconstruction of Reed-Solomon encoder and interleavers over noisy environment. *IEEE Trans. Broadcasting*, 2018, **64**(4), 830-845.
    doi: 10.1109/TBC.2018.2795461
20. Swaminathan, R.; Madhukumar, A.S. & Guohua, W. Blind estimation of code parameters for product codes over noisy channel conditions. *IEEE Trans. Aerospace Electron. Syst.*, 2019, **56**(2), 1460-1473.
    doi: 10.1109/TAES.2019.2934308
21. Xie, H.; Wang, F.H. & Huang, Z.T. Blind recognition of reed-solomon codes based on histogram statistic of galois field spectra. *J. Adv. Mater. Res.*, 2013, **791-793**, 2088-2091.
    doi: 10.4028/www.scientific.net/AMR.791-793.2088
22. Sklar, B. Digital Communications: Fundamentals and Applications. Prentice-Hall, New Jersey, 2001.
23. Lin, S. & Costello, D.J. Error control coding. Prentice hall, Englewood Cliffs, USA, 2011.
24. MacWilliams, F.J. & Sloane, N.J.A., 1977. *T*he theory of error correcting codes. Elsevier, Amsterdam, Netherlands, 1977.

ACKNOWLEDGMENT

Authors wish to thank Director, Scientific Analysis Group, Defence R&D Organisation for her encouragement and support to carry out above work.

CONTRIBUTORS

Mr Anand Sharma did his BE (Electronics) from MNNIT, Allahabad and MTech (Communication System Engineering) from IIT, BHU. Currently he is a Scientist in DRDO-Scientific Analysis Group, Delhi. He is working in the area of information security. His research area includes cryptography, embedded system security and error correcting codes.

In the current study, he has proposed a technique for blind recovery of parameters of Reed-Solomon Code from intercepted signal. He has also carried out extensive literature survey to ensure novelty.

Mr Praneesh Gupta did his BTech (Electronics and Communication Engineering) from University of Allahabad, Allahabad and MTech (Signal Processing and Digital Design) from Delhi Technological University, Delhi. Currently, he is a Scientist in DRDO-Scientific Analysis Group, Delhi. He is working in the area of information security. His areas of interest are digital electronics, cryptography and error correcting codes.

The experimental study of the proposed technique using MATLAB and comparison with other techniques has been carried out by the author. He also studied several relevant papers to consolidate important results.