A Necessary and Sufficient Condition for the Existence of Asymmetrical Reversible VLCs

Richa Gupta, Radhika Goel

Abstract: ***Affix-free codes are widely used in multimedia communications because of its error tolerance capbility. Reversible Variable Length Code (RVLC) is a type of affix-free code. In literature, there are many construction algorithms available for RVLCs. But unlike Variable Length Codes (VLCs), RVLCs lack in the area of its mathematical development in the form of lower bound or upper bound on average codeword length, bounds on existence, and related Theorems. Only few mathematicians have done some work on this. In 2014, Richa and Radhika have proposed and discussed the necessary and sufficient condition on the number of codewords for a particular (bit length vector) required for the existence of symmetrical RVLCs. This paper is an extension of the earlier published paper on the similar ground, but for asymmetrical RVLCs. This paper derives and discusses necessary and sufficient condition, on bit length vector (the number of codewords for a particular length), required for the existence of asymmetrical RVLCs over the given D-ary code alphabet.***

Index Terms: ***Affix-free codes, Symmetrical RVLC, asymmetrical RVLCs, mathematical bound on RVLC, bit length vector, Kraft inequality.***

I. INTRODUCTION

Variable length code is a type of code in which source symbols are mapped into variable number of bits according to their probabilities of occurrence. Huffman codes, also known as the optimum variable length code was introduced by Huffman in 1952. Huffman code gives the minimum average code length and thus enhances the transmission efficiency [1]. In literature, you will find Huffman code has been used to represent the class of VLC. Variable length code suffers with very poor error tolerance as it is not designed with that aim. But, we all know that all practical transmission channels are prone to noise and errors, thus resulting the encoded bit stream to suffer with noise and gets corrupted. For proper decoding of data, error detection and correction are required to be performed at receiver side. Error detection and correction process add cost of the hardware and also data size suffers. Therefore, there is a need of a VLC embossed with error detection/correction capability. VLCs are extremely susceptive to errors because they don't have any codeword boundaries and codewords are of variable length in nature. So, a single bit error can cause a complete loss of data from the position of occurrence of the noise. RVLC can be decoded in the forward and in the backward directions, because of being affix free in nature. Hence, RVLC has more error tolerance as compared to simple prefix-free VLCs, hence RVLC has replaced VLC in almost all multimedia applications [2][3][4]. Bounds play an important role in all physical phenomena involving quantitative studies. In some sense, they set a benchmarks of physical boundaries for feasible researches and ideal accomplishments. This has been an extensive area of research and study in 'Information and Coding Theory.' In this paper, we propose and derive an upper bound on the average codeword length of asymmetrical RVLCs. A brief literature review along with construction algorithms and basic properties of RVLCs are discussed in Section II. Section III discusses important bound on VLCs (Kraft Inequality) and the only available bound on symmetrical RVLCs. The statement and derivation of bound proposed for asymmetrical RVLCs is given in detail in Section IV. Section V gives a discussion on the derived bound of asymmetrical RVLCs with the help of example. The paper is concluded in Section VI.

Manuscript published on 28 February 2019.
*Correspondence Author(s)
Richa Gupta, Electronics and Communication Engineering, JIIT, Noida, India
Radhika Goel, Electronics and Communication Engineering Department, Krishna Engineering College, Ghaziabad, India.

II. REVERSIBLE VARIABLE LENGTH CODES

For variable length codes, designed for noiseless channels, the main interest is to have 'minimum' average codeword length so that maximum efficiency can be obtained. Similarly, for reliable communication over noisy channels, the approach is to introduce structured redundancy so that changes due to noise can be detected and corrected. However, redundancy reduces rate. Therefore, the central problem of efficient coding is to add 'minimum' redundancy to obtain the desired level of error correcting capability and the solution is Reversible Variable Length Codes.

RVLCs are divided in two categories:

1. Symmetric RVLCs
2. Asymmetric RVLCs

If all codewords of a RVLC are symmetric in nature, the code is called as symmetric RVLC, else asymmetric. For forward and backward decoding, symmetric RVLC uses same look-up table for decoding, while different decoding tables are required in the case of asymmetric RVLC. But, in general, average codeword length of symmetric code is less than that of asymmetric RVLC, but generation algorithms of symmetric RVLC is in general easier to implement and are less complex. The need of RVLCs was first explored and explained by Takishima in 1995 [5]. He also proposed an algorithm to generate RVLCs. The Takishima algorithm is based on the construction of RVLCs by first generating Huffman codes and then applying a top down approach to convert them into symmetrical RVLCs.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

However, a lot of variation was observed in constructing algorithm and many algorithms have been developed down the line. Some algorithms are based on Huffman Codes, while others are independent of Huffman Codes. On the basis of construction algorithms, RVLCs are broadly classified in four main categories, as given below:

1. Huffman dependent Symmetric RVLCs
2. Huffman independent Symmetric RVLCs
3. Huffman dependent Asymmetric RVLCs
4. Huffman independent Asymmetric RVLCs

The construction algorithms of RVLCs may or may not depend on the Huffman code for the given course source symbols. The algorithms given by Takishima's symmetrical RVLCs [5], Takishima's asymmetrical RVLCs [5], Tsai's symmetrical RVLCs [6], Jeong symmetrical RVLCs [7] and Richa-Radhika's asymmetrical RVLCs [8] are Huffman dependent algorithms, while algorithm given by Yan's symmetrical RVLCs [9] and Golomb - Rice's asymmetrical RVLCs [10] are Huffman independent algorithms.

The additional features of RVLCs, as compared with an ordinary VLCs, are:

1. It supports simultaneous decoding, of the received data at the decoder side, in both the forward and the reverse direction to speed up the process;
2. In case of the presence of the error, only a small segment containing the error will be damaged and the remaining portion of the binary file can be retrieved.

Some properties of RVLCs are well explained by [11][12][13]. Bit length vector (BLV) and Maximum Symmetric Suffix Length (MSSL) are two important parameters required in applying the algorithm for generating codewords. Dealing symmetric RVLC is an easier job in construction of codewords, in applying any condition and because codewords are symmetric in nature, checking affix free automatically ends with checking either prefix-free or suffix-free, not checking both simultaneously. Error resilience is one of the main reason of wide acceptance of RVLC in multimedia communications in place of simple source code, i.e. Huffman code. It is obvious that error tolerance is achieved at the cost of increased data rate, because of an increase in average codeword length.As discussed earlier, the main purpose of variable length code is to convert the processed data into binary bit-stream. Let us consider the case of single bit error introduced at the channel due to noise. If RVLCs are used in encoding, then due to their error resilience capability, we can avoid getting the whole bit-stream file corrupted due to the presence of single bit error. With the use of the forward decoding and the reverse decoding, we can obtain the decoded file at the receiver except for the portion (codeword) containing single bit of error. Many applications, like image coding standards like JPEG, EZW, SPIHT, JPEG 2000 and video coding standards like MPEG-4, H.263+ and H.263++ are extensively being explored by the researchers on using RVLCs in place of Huffman code nowadays [2-4][14].

III. IMPORTANT BOUNDS ON VLCS AND RVLCS

The "Fundamental Theorem of Information Theory", an all-important result in Information theory, is also a result in form of a bound – 'Capacity' of the channel being an upper bound on 'rate of reliable communication.' Similarly, another early result, namely "Source Coding Theorem", states that the average codeword length of a uniquely decodable variable length code cannot be less than entropy of the source, is a bound. This bound is also, in fact, an independent characterization of 'Shannon's Entropy.'

Differently speaking, these two theorems give answers of the following two questions:

1. What is the upper bound on transmission rate for reliable communication over unreliable (a noisy) channel?
2. What is the lower bound on irreducible complexity below which the source signal cannot be further compressed?

The answers are in terms of bounds, involving channel capacity and entropy. Kraft's and McMillian's inequalities are also bounds in their own way providing benchmarks on the pertinent measures [15]. And the value of this 'lower bound' may be properly decided using the bounds. Thus, it may be noted that in communications, bounds play the key role in deciding the boundaries of the code parameters required for an application.

The necessary and sufficient condition for the existence of instantaneous source codes (prefix-free variable length code) was developed by L. G. Kraft [15]. Richa and Radhika developed the similar condition for symmetrical RVLCs, also known as affix free codes [16]. This section describes both conditions in the required details.

a). Necessary and sufficient condition for Prefix-free VLCs

Kraft developed the mathematical condition for the existence of instantaneous codes. The condition is also known as Kraft inequality, and it is the necessary and sufficient condition that every instantaneous code satisfies. Let us discuss about the Kraft inequality in the next section.

Kraft inequality limits the set of codeword lengths for an instantaneous code in the form of an inequality, known as Kraft inequality. It states that

1) For the given source with symbols x_1, x_2,…x_n with probabilities p_1, p_2,….p_n respectively with the corresponding codewords as c_1, c_2…. c_n of length l_1, l_2,… l_n, then these lengths must satisfy Kraft's inequality

$$\sum_{n=1}^{N} \frac{1}{D^{l_n}} \leq 1$$

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

2) If the codewords lengths l_1, l_2,… l_n satisfy the above condition, then Kraft inequality ensures that there exists an instantaneous code with l_1, l_2,… l_n codeword length.

The poof of Kraft inequality with a detailed discussion is given by Kraft in [15].

b). Necessary and sufficient condition for Symmetrical Reversible VLCs

This section discusses a mathematical condition that is the necessary and sufficient condition for the codeword lengths of symmetrical reversible VLCs to exist. The inequality has been derived by Richa-Radhika in 2014 [16] on the analogy of the derivation of Kraft inequality done earlier.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

The statement of necessary and the sufficient condition for the existence of symmetrical affix-free codes (RVLCs) given by Richa-Radhika is as follows:

1) If C is a D-ary affix symmetrical code with bit length vector $w_1, w_2 \ldots\ldots w_\sigma$ where w_i denote the number of symmetrical reversible codewords of length i, then these length must satisfy this inequalty

$$w_\sigma \leq D^l - \sum_{i=1}^{l} w_i D^{(l-i)}$$

where $l = \lfloor \sigma/2 \rfloor$

2) If the numbers $w_1, w_2 \ldots\ldots w_\sigma$ and D satisfy the above inequality, then there must exist some symmetrical reversible D-ary code with bit length vector lengths $w_1, w_2 \ldots\ldots w_\sigma$.

The complete derivation with detailed discussion of the above inequality is given by Richa-Radhika [16].

IV. PROPOSED NECESSARY AND SUFFICIENT CONDITION FOR ASYMMETRICAL REVERSIBLE VLC

In case of symmetrical reversible codes, it was easier to obtain bound, reason being the codewords were symmetrical and hence checking prefix-free was equivalent to checking suffix-free condition. This section discusses necessary and sufficient condition of the lengths of codewords to satisfy for the existence of asymmetric reversible code. Let a source C is given in the form of source statistics, i.e. symbols and the corresponding probabilities. Our aim is to assign asymmetric reversible codewords, such that the codewords are affix-fee and asymmetric in nature.

We want to check that for the given bit length vector of the code, is it possible to assign asymmetric reversible codewords. Let wi represents the number of target codewords of length i. Then, obviously for a D-ary code, the possibilities for 1 length codeword would be

$$w_1 \leq D \qquad \ldots\ldots\ldots (1)$$

Similarly, for codewords who has length 2 , the condition for w_2 can be obtained by considering all possibilities for the first bit (obviously it will be the remaining cases from equation 1) and to avoid suffix condition we need to subtract the codewords which have been used earlier, i.e.

$$w_2 \leq (D\text{-}w_1).D\text{-}w_1 \qquad \ldots. (2)$$

or

$$w_2 \leq (D^2\text{-}w_1 D\text{-}w_1)$$

Next, for the number of codewords of length 3, w_3. The first two characters can be chosen in *((D-w₁).D-w₂)* ways, thus the upper bound would be

$$w_3 \leq \big((D\text{-}w_1).D\text{-}w_2\big).D\text{-}w_2 \qquad \ldots. (3)$$

$$or \quad w_3 \leq (D^3\text{-}w_1 D^2\text{-}w_2\, D\text{-}\, w_2)$$

Similarly, for four bit codewords, we can compute bound as

$$w_4 \leq \Big(\big((D\text{-}w_1).D\text{-}w_2\big).D\text{-}w_3\Big).D\text{-}w_3 \ldots. (4)$$

$$or \quad w_4 \leq D^4\text{-}w_1 D^3\text{-}w_2\, D^2\text{-}w_3 D\text{-}w_3$$

Since the first two characters can be chosen in $((D\text{-}w_1)\ D\text{-}w_2)$ ways and the first three character can be chosen in $((D\text{-}w_1)\ D\text{-}w_2)D\text{-}w_3)$ way and the fourth character can be chosen in $(D\text{-}w_1)$.

Now,

$$w_5 \leq \Big(\big(((D\text{-}w_1).D\text{-}w_2\big).D\text{-}w_3)D\text{-}w_4\Big).D\text{-}w_4$$

$$or \quad w_5 \leq D^5\text{-}w_1 D^4\text{-}w_2 D^3\text{-}\, w_3 D^2\text{-}w_4 D\text{-}\, w_4$$

Thus, w_σ satisfies the inequality

$$w_\sigma \leq D^\sigma\text{-}w_1 D^{\sigma\text{-}1}\text{-}w_2 D^{\sigma\text{-}2} - w_3 D^{\sigma\text{-}3} \ldots\ldots - w_{\sigma\text{-}1} D - w_{\sigma\text{-}1}$$

where σ is the number of code characters in the longest word.

Rearranging the terms

$$w_\sigma \leq D^\sigma\text{-}w_1 D^{\sigma\text{-}1}\text{-}w_2 D^{\sigma\text{-}2} - w_3 D^{\sigma\text{-}3} \ldots\ldots - w_{\sigma\text{-}1} D - w_{\sigma\text{-}1} \qquad \ldots\ldots. (5)$$

Equation (5) represents the inequality obtained for an asymmetric reversible code with number of codewords of length l_σ as w_σ. It can be noted that this condition signifies that for the given source, if bit length vector $w_1, w_2, \ldots w_\sigma$ is satisfying the inequality mentioned in equation (5), it is possible to construct codewords with that number of codewords of particular length. It doesn't indicate that any code which satisfies the above condition is a valid asymmetric reversible code. Let us consider an example for better understanding of the inequality.

V. RESULT AND DISCUSSION

The necessary and sufficient condition on the existence of asymmetric RVLCs is derived in Section IV in terms of the bound on number of available codewords for a particular length. Let us consider an example for better understanding of the derived expression. Example: Let us consider that $w_1 = 1$ and $w_2 = 1$ are given for an asymmetric code C. We want to check how many codewords of length 3 are available which can be assigned to code C.

Solution: The given values are:

$$w_1 = 1,\ w_2 = 1,$$

$$w_3 \leq \big((D\text{-}w_1).D\text{-}w_2\big).D\text{-}w_1$$

on substituting values of w_1 and w_2, we get

$$w_3 \leq ((2-1).2-1)2-1$$

$$w_3 \leq 1$$

which indicates, there is only one available codeword which can be assigned of length 3, after considering 1 codeword of length 1 and 1 codeword of length 2. Let us assume "0" is assigned as codeword of length 1, then we can have "11" as a 2 length codeword such that affix free condition is satisfied.

After considering "0" and "11", we can have only "101" as codeword of length 3, all other ("000". "001", "010", "011", "100", "110" and "111") don't satisfy the affix-free condition.

Similarly, the inequality can be checked and justified on any example of asymmetric RVLCs. This bound is of great importance to decide number of asymmetric reversible variable length codewords that can be designed for a particular length.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

VI. CONCLUSION

A detailed proof of the inequality which is the necessary and sufficient condition for the existence of asymmetrical RVLCs over a D-ary code alphabet for a given bit length vector is presented in the paper. The proof of the above said inequality is based on well-known Kraft's inequality for prefix free codes and inequality for the symmetrical Reversible Variable Length Codes.

REFERENCES

1. D. Huffman, "A method for the Construction of Minimum Redundancy Codes", Proceedings of IRE, 40, 1962, pp. 1098-1101.
2. ISO/IEC JTC1/SC29/WG11/N3908, "MPEG-4 video verification model," Vers. 18.0, Jan. 2001.
3. ITU-T Recommendation H.263, "Video coding for low bit rate communication," Annex D, Feb. 1998.
4. H. Wang, S. N. Koh, and W.W. Chang, "Application of reversible variable-length codes in robust speech coding," IEEE Proc. Commun., vol. 152, no. 3, June 2005, pp. 272-276.
5. Y. Takishima, M. Wada, and H. Murakami, "Reversible variable length codes," IEEE Trans. Commun., vol. 43, Feb.-Apr. 1995, pp. 158–162.
6. C. W. Tsai and J. L. Wu, "Modified symmetrical reversible variable-length code and its theoretical bounds," IEEE Trans. Inform. Theory, vol. 47, Sept. 2001, pp. 2543-2548.
7. W. H. Jeong and Y. S. Ho, "Design of Symmetrical Reversible Variable Length Codes from the Huffman Code," Picture Coding Symposium, 2003, pp. 135-138.
8. R.Goel and R. Gupta. "Redesigning of the Construction of Symmetrical RVLCS Based On Graph Model.", International Journal of Information & Computation Technology, vol. 4, no. 11, 2014, pp. 1063-1068.
9. H. J. Yan, C. Y. Lin, L. Zhong , "On constructing symmetrical reversible variable-length codes independent of the Huffman code, ''The National Key Laboratory on Integrated Service Networks, Xidian University, Xi'an 710071, China, accepted Feb. 22, 2006.
10. S. Golomb, "Run Length Encodings," IEEE Transactions on Information Theory, vol. 12, no. 3, 1966, pp. 399-401.
11. Abedini, S. P. Khatri, and S. A. Savari, "A SAT-based scheme to determine optimal fix-free codes," Proc. of the 2010 IEEE Data Compression Conference, Snowbird, Utah, March 2010, pp. 169-178.
12. S. M. Hossein, T. Yazdi and S. A. Savari, "On the Relationships among Optimal Symmetric Fix-Free Codes," IEEE Data Compression Conference, 2013, pp. 391-400.
13. A. Savari, "On optimal reversible-variable-length codes," Proc. Information Theory and Applications Workshop, La Jolla, CA, February 10, 2009, pp. 311-317.
14. K. Sayood, Introduction to data compression, New Delhi: Elsevier, 2011.
15. L.G. Kraft, A device for quantizing, grouping and coding amplitude modulated pulses, Master's thesis, Dept. of Electrical Engineering, M.I.T., Cambridge, Mass., 1949.
16. Goel, R., and Gupta, R. Necessary and sufficient condition for the existence of symmetrical Reversible Variable Length Codes, based on Kraft's inequality. In IEEE Conference publication Recent Advances and Innovations in Engineering (ICRAIE), May, 2014, pp. 1-3.

AUTHORS PROFILE

Richa Gupta received her Ph.D. Degree from JIIT, Noida, India in the year 2013 and M.Tech in Information Systems from IIT, Kanpur, India in 2005. Since August 2007, she has been working as Assistant Professor in the Department of Electronics and Communications Engineering, JIIT Noida. Prior to this, she has worked with a USA based research company 'ATC Labs, Noida' for 2 years. She has published several research papers in International Journals and presented many in International Conferences. She is a Fellow member of IETE, a Life member of IEI and a Life member of ISTE. She has been associated with research and development in the field of Digital Signal Processing, Communications, Digital Electronics, Digital Image Processing, Multimedia Communications, Information Theory and Applications. She has recently published a book entitled "Introduction to Digital Systems", with Satyam Publishers.

Ms Radhika Goel received her B.Tech and M.Tech degrees in Electronics and Communication Engineering in 2011 and 2014, from ABES-IT and Jaypee Institute of Information Technology respectively. Her topic of M.Tech Thesis was Mathematical Analysis of Reversible Variable Length Code and Application, she was able to derive Necessary and Sufficient Condition for the existence of Reversible Variable length codes based on the analogy of Kraft's Inequality. She has been awarded with the Gold Medal for Academic excellence in M.Tech. Her Interest includes Information Theory and Coding, Variable length Codes, Audio Processing and Video Processing, Image Processing and Wavelet Theory. She is active member of the International Association of Engineers and Universal Association of Computer and Electronics Engineers

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)