Communications on Applied Electronics (CAE) – ISSN : 2394-4714
Foundation of Computer Science FCS, New York, USA
Volume 2 – No.9, September 2015 – www.caeaccess.org

The Huffman's Method of Secured Data Encoding and Error Correction using Residue Number System (RNS)

A. Alhassan
University for Development Studies (UDS)
Faculty of Mathematical Sciences (FMS)
Computer Science Department,
P. O. Box 24, Navrongo, Ghana. West Africa

I.Saeed
University for Development Studies (UDS)
Faculty of Mathematical Sciences (FMS)
Computer Science Department,
P. O. Box 24, Navrongo, Ghana. West Africa

P.A. Agbedemnab
University for Development Studies (UDS)
Faculty of Mathematical Sciences (FMS)
Computer Science Department,
P. O. Box 24, Navrongo, Ghana. West Africa

ABSTRACT
Over the centuries, information security has become a major issue. Encryption and decryption of data has recently been widely investigated and developed because there is a demand for a stronger encryption and decryption which is very hard for intrusion. Cryptography plays major roles in fulfilment of these demands. Many of researchers have proposed a lot of encryption and decryption algorithms. But most of the proposed algorithms encountered problems such as lack of reduced cost of data and error control mechanisms to maintain the security of data in the communication channel. In this paper, a highly secured data encryption and decryption scheme is proposed to enhance the Huffman's method. The Residue Number System (RNS) is employed with four moduli set $\{2^{n-1}, 2^n-1, 2^n+1, 2^{n+1}-1\}$ and two redundant moduli set $\{2^{2n}-3, 2^{2n}+1\}$ for error handling using the concept of the traditional Huffman's algorithm, where the frequency of occurrences of each character are used to generate binary codes. The proposed scheme allows for unreadable encrypted set of bits, except the intended recipient with the right moduli set can decrypt it, reduced cost of both data transmission and storage, and error detection and correction.

General Terms
Information Security, Number System, and Compression Algorithm.

Keywords
Encryption, Decryption, Residue Number System (RNS), Redundant Moduli Set, Moduli Set, Huffman's method, Data Security, Error Correction.

1. INTRODUCTION
Cryptography is the art and science of using cryptographic techniques and practicing the cryptographic techniques in designing a secure cryptosystem [5], [14]. The current network communication system requires exchange of information with highly secured data and reduction in both the space requirement and speed for data storage and transmission. In order to reduce the cost of both data storage and communication, there is a need to reduce the redundancy in the data representation, which is, compressing the data. Compression is a technology for minimizing the amount of data used to represent any content without excessively reducing the quality of the data [5]. The Residue Number System (RNS) that supports carry free, parallel, high speed, low power and secure computing is applied to enhance the Huffman coding technique with the moduli set $(2^{n-1}, 2^n-1, 2^n+1, 2^{n+1}-1)$ is proposed in this research. RNS has the advantage of fast computation over other number systems due to its inherent features of parallelism, fault tolerance, carry free operations, and modularity [8], [9], [10], [16]. The inherent error correction ability and fast intensive-arithmetic capability is of essence in this paper. Data encryption and decryption is certainly important to the security or integrity of information to be transmitted through a network communication channel, the need for fast and secured data transmission is therefore of great essence. RNS defined by a set of relatively prime integers called the moduli. The prime moduli set represented as $\{p_1, p_2, \ldots, p_n\}$, where p_i is the prime of *i*th modulus. Each integer X is represented as a set of smaller integers called residues. The residue set represented is $\{r_1, r_2, \ldots r_n\}$, Where r_i is the ith residue. The residue r_i is defined as the least positive remainder, when X is divided by modulo p_i. RNS is based on congruence relation, which is defined as follows.

$$|X|_{p_i} = r_i \qquad (1)$$

where p_i is the modulus, r_i is the residue and $|X|_{p_i} = X \bmod p_i = r_i$. The dynamic range is determined by the product of modulo set $\{p_1, p_2, \ldots p_n\}$ [1], [2] and represented by;

$$M = \prod_{i=1}^{n} p_i \qquad (2)$$

1.2 Arithmetic Operations in RNS
Let I and J be represented in RNS as $I = (i_1, i_2, i_3, \ldots, i_n)$ and $J = (j_1, j_2, j_3, \ldots, j_n)$ and $I, J \in H_M = (1,2,3, \ldots, M-1)$, M been the dynamic range.

Then $H = I \bowtie J$ where $\bowtie$ represents addition, subtraction or multiplication operation in RNS as follows;

$H = I \bowtie J \longrightarrow (|i_1 \bowtie j_1|_{m_1}, |i_2 \bowtie j_2|_{m_2}, \ldots, |i_n \bowtie j_n|_{m_n}$ (2.1) [19], [20].

For example, if $m_1 = 3$, $m_2 = 4$, $m_3 = 5$ then,

$\overset{+}{\underset{16}{\overset{5}{11}}}\ in\ RNS\ \overset{+}{\underset{(1,0,1)}{\overset{(2,1,0)}{(3,2,1)}}}$, $\overset{-}{\underset{-6}{\overset{5}{11}}}\ in\ RNS\ \overset{-}{\underset{(0,2,4)}{\overset{(2,1,0)}{(3,2,1)}}}$ and $\overset{\times}{\underset{55}{\overset{5}{11}}}\ in\ RNS\ \overset{\times}{\underset{(1,3,0)}{\overset{(2,1,0)}{(3,2,1)}}}$

There is a lot of literature on data compression and encryption which tend to view the subject from the perspective of data transmission only and forgotten the core value of compression in data storage. In this paper, the discussion will be centred on encryption and decryption of data for storage and transmission through network communication channels. Thus, the study is focus on an algorithm for data compression, encryption and decryption by adopting the traditional Huffman's method for data encoding but deal not with hardware component of the compression.

The whole idea of this form of data encryption started in [4], where the source message and their corresponding probabilities are arranged in order of non-increasing probabilities and then divided into two groups of nearly equal total possible probabilities. Each of the source messages in the first group is then assigned with '0' as the first codeword and those in the second group are assigned with '1' as their first codeword. The process is repeated until smallest possible groups are reached. The scheme in [6] proposed a modification for the method of coding in [4] by building an extended binary tree with a minimum weighted path length from a set given weights. It therfore takes a list of weights as inputs and constructs a full binary tree with every node having either zero or two children and the weights are assigned to the children leaves. The scheme in both [11], [15] enhanced the Huffman's method by proving that the upper bond on the redundancy of Huffman's code is the probability of the least source message $(n) + \frac{(log\ 2\ log\ e)}{e}$. The scheme in [3] also researched on enhancing the Huffman coding by applying RNS using the popular three moduli set $\{2^n - 1, 2n, 2n+1\}$.

The rest of the paper is organized as follows: section 2 presents the proposed method for encoding and decoding as well as the reverse conversion process to recover a sent message. Section 3 deals with the error handling aspect of the scheme; the performance of the proposed scheme is presented in Section 4, and Section 5 concludes the paper.

2. PROPOSED SCHEME

A highly secured data storage and transmission is being proposed in this application. One way of achieving this highly secured and reduced cost of data storage and transmission is the use of RNS However, RNS is not without challenges; arithmetic operations such as magnitude comparison, sign detection, overflow detection, moduli selection, reverse and forward conversions hinder the wide spread usage of RNS in general purpose computing [7]. For the success of any application of RNS, the conversion processes of binary (or decimal) to residue conversion and residue to binary conversion are required [13]. In order to achieve the objective of the study, we adopted the traditional Huffman's algorithm where frequencies of occurrences of each character in the source message (decimal numbers) are converted to residues and then to binary at the front end of the conversion. This paper also looks at the concept of Redundant Residue Number Systems (RRNS) as an error detection and correction mechanism for the proposed algorithm. RRNS are relatively pair prime moduli set that have been added to the chosen moduli set for error detection and correction. The reverse conversion process is achieved by the use of the CRT as follows;

For any integer set of residues $\{r_1, r_2, \ldots r_n\}$,

$$X = \left| \sum_{i=1}^{n} M_i \,\middle|\, (r_i)\, M_i^{-1} \,\middle|\, p_i \right|_M \qquad (3)$$

where, X is the resulting number, r_i is an element of the *i*th residue, p_i is the modulus, $M_i = \frac{M}{pi}$, M is the dynamic range of the system and M_i^{-1} is the multiplicative inverse of M_i over modulo P_i [2], [19].

The algorithm is made up of an efficient low complexity encoder and decoder to ensure a very high security level, reduced cost of data for storage and transmission, and error control. With the adoption of the Huffman's algorithm, the RNS moduli set is applied to the frequency of occurrence of each character in the source message (decimal number) X in the encryption process at the front end. A reversed process is applied to the encrypted bitstream in the decryption process at the back end.

2.1 Encoder

During encoding, the forward conversion process is employed with the four moduli set and the additional two redundant moduli set to encode the decimal number X (frequency of occurrence of each character) to residues as the first encryption. The resultant residues from the modulus operations are obtained simultaneously as the operations are executed in parallel by corresponding modulo circuits. The residues are converted to binary as the final encrypted bitstream for each particular character and then concatenated to be a RRNS codeword [10]. The encrypted bitstream are organized into special order before it is randomly transmitted or stored in memory cells.

2.2 Decoder

The decoding process is designed to detect and correct errors. During decoding, detection of errors is applied to any read codeword from the memory, a codeword is valid if its decoded value is within the legitimate range otherwise it is invalid, hence error correction is needed. However, the correction process involves an exhaustive systematic calculation to search for values that are within the legitimate range. From all recovered integer values, there will be ideally, at least one unique value falling within the legitimate range, which is the correct data. However, in a case where there are more than one recovered values falling within the legitimate range, such ambiguity is resolved by using maximum likelihood decoding scheme to determine the rightful data between the recovered values. The idea behind the scheme is the closest Hamming distance between and among the values falling within the legitimate range and the read codeword [10]. The correct encrypted bitstream is first converted from binary to residues, and then from the residues back to the decimal number X using the traditional Chinese Remainder Theorem (CRT) because it allows for larger modulo dynamic

range M operations. The schematic diagrammatic representation is shown in figure 1 and figure 2 respectively.

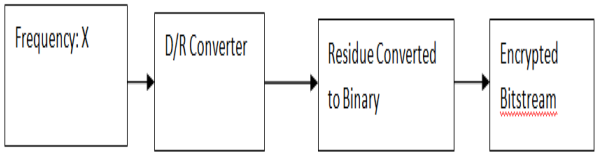

Figure 1 A Schematic Diagran of RNS Encoder

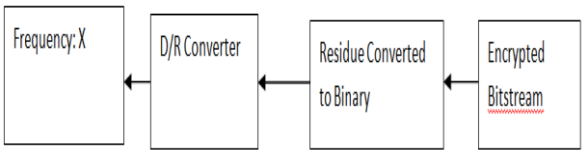

Figure 2: A Schematic Diagram of RNS Decoder

Table 1: Characters and their Frequencies showing some Examples for Decimal to Residue Conversion with the Moduli Set (2^{n-1}, 2^n—1, 2^n+1, 2^{n+1}—1) and (2^{2n}-3, 2^{2n}+1).

Character	Frequency	Moduli Set (2, 3, 5, 7, 13, 17)	Encrypted bitstream
A	25	1, 1, 0, 4, 12, 8	01. 01.00. 100. 1100. 1000
B	26	0, 0,1, 5, 0, 9	00. 00. 01. 101. 00. 1001
C	27	1, 1`, 2, 6, 1, 10	01. 01.10.110.01.1010
D	28	0, 2, 3, 0, 2, 11	00.10. 11.01. 10. 1011
E	44	0, 2, 4, 2, 5, 10	00. 10.100. 10. 101. 1010

2.3 Residue to Binary Conversion

The residue to binary conversion can be achieved by using the CRT [2]. For any integer X ≤ M, can be uniquely represented in RNS as $\{r_1, r_2, \ldots, r_n\}$ and (p_1, p_2) are pair relatively prime, and then X can be recovered by using equation (3).

2.3.1 Example 1

The following example shows the use of CRT to recover the number from its residue digits. The number 25 can be represented using the moduli (2, 3, 5, and 7) as (1, 1, 0, 4). To convert (1, 1, 0, 4) back to a decimal representation, the CRT is applied as follows: M = 2 x 3 x 5 x 7 = 210.

$M_i = \frac{210}{mi}$ = (105, 70, 42, 30)

The multiplicative inverses are obtained as follows:

$|105|_2 = 1$, $|1|_2 = 1$, therefore $M_1^{-1} = 1$

$|70|_3 = 1$, $|1|_3$ therefore $M_2^{-1} = 1$

$|42|_5 = 2$, $|(2 \times 3)|_5 = 1$, therefore $M_3^{-1} = 3$

$|30|_7 = 2$, $|(2 \times 4)|_7 = 1$, therefore $M_4^{-1} = 4$

Then using equation (4), X can be obtained as

$X = \{|105|(1 \times 1)|_2|_{210} + |70|(1 \times 1)|_3|_{210} + |42|(3 \times 0)|_5|_{210} + |30|(4 \times 4)|_7|_{210} = |105 + 70 + 0 + 480|_{210} = |655|_{210} = 25$

3. ERROR HANDLING

It is important to notice that all error correction methods in RNS employ the use of RRNS. By employing two redundant moduli set along the original chosen moduli set, the algorithm can be a better enhancement of the Huffman's coding technique in terms of bit error rate (BER) [9]. In RRNS, a number represented with an original moduli set (with *p* number of moduli) can still be represented with the original chosen moduli and redundant moduli (*q-p)* number of redundant moduli). The redundancy in the system allows for the reconstruction of that number by using any *v* combinations of the moduli at the receiver and such RRNS (*q*, *p*) code has capability of simultaneously detecting *s* residue digit errors and correcting *t* random residue digit errors, if and only if $t+s \leq (q-p)$ [9].

3.1 Example 2

This example illustrates how a single error in the received residue digits is detected and corrected by employing two redundant moduli. Let's consider two redundant moduli, namely 13 and 17 in addition to the original moduli (2, 3, 5, 7). Now, the mdouli set becomes (2, 3, 5, 7, 13, and 17). Let us consider the integer message x=25, which has residue digits (1, 1, 0, 4, 12, 8). Assume that the r_3 digit is in error, i.e. (1, 1, 2', 4, 12, 8). According to CRT, the integer X in the range (0, 210) can be recovered by invoking any four moduli and their corresponding residue digits, if no errors occurred in the received RNS representation.

Let us now attempt to recover the integer X represented as (1, 1, 2', 4, 12, 8) by considering all possible cases. Once all the possible combinations of four out of six residue digits retained, it results in:

$(r_1, r_2, r_3, r_4) = (1, 1, 2' 4) - X_{1234} = 67$

$(r_1, r_2, r_3, r_5) = (1, 1, 2', 12) - X_{1235} = 337$

$(r_1, r_2, r_3, r_6) = (1, 1, 2', 8) - X_{1236} = 127$

$(r_1, r_2, r_4, r_5) = (1, 1, 4, 12) - X_{1245} = 25$

$(r_1, r_2, r_4, r_6) = (1, 1, 4, 8) - X_{1246} = 25$

$(r_1, r_3, r_4, r_5) = (1, 2', 4, 12) - X_{1345} = 207$

$(r_1, r_3, r_4, r_6) = (1, 2', 4, 8) - X_{1346} = 7140$

$(r_1, r_4, r_5, r_6) = (1, 4, 12, 8) - X_{1456} = 25$

$(r_2, r_3, r_4, r_5) = (1, 2', 4, 12) - X_{2345} = 1117$

$(r_2, r_3, r_4, r_6) = (1, 2', 4, 8) - X_{2346} = 5735$

$(r_2, r_4, r_5, r_6) = (1, 4, 12, 8) - X_{2456} = 25$

$(r_3, r_4, r_5, r_6) = (2', 4, 12, 8) - X_{3456} = 1572$

Where X_{ijkl} represents the recovered result by using moduli p_i, p_j, p_k, p_l as well as their Corresponding residue digits r_i, r_j, r_k, r_l. From these results we observe that X_{1235}, X_{1346}, X_{2345}, X_{2346} and X_{3456} are all illegitimate numbers, since their values are out of the legitimate range [0,210]. In the remaining seven cases, all the results are same and equal to 25, except for X_{1234}, X_{1236} and X_{1345}. Moreover, all these results were recovered from four moduli without including p3, i.e. from X_{1245}, X_{1246}, X_{1456} and X_{2456} which are equal to 25. Hence, we can conclude that the correct result is 25 and there was an error in r_3, which can be corrected by computing $r_3 = |25|_5 = 0$.

4. PERFORMANCE ANALYSIS

The performance of the proposed RNS four set encryption algorithm is evaluated theoretically in terms of space and security. The proposed scheme provides high security than the traditional Huffman's coding presented in [6] and equivalent state of the art RNS three moduli set encryption presented in [3]. Also, the scheme provides better enhancement of Huffman's method than the one presented in [3]. Further, the storage capacity of the proposed scheme outweighs the state of the art RNS three moduli set encryption in [3], since the proposed scheme has one more channel.

5. CONCLUSION

In this paper, a new RNS four moduli set data encryption and decryption algorithm is proposed based on the traditional Huffman's encoding algorithm that utilizes the probability or frequency of occurrence of characters or strings in a particular set of data. Additionally, the proposed scheme allows for possible error detection and correction. On theoretical point of view, our proposed scheme outperforms equivalent known state of the art RNS three moduli set encryption in terms of security and storage capacity. For the intruder to be able to decrypt the message, then the moduli set should be known and as well as the encryption algorithm. The reduced size of compressed data is achieved using RNS which uses residue of numbers instead of the numbers themselves that have greater weights or magnitude. In feature, the proposed scheme can be enhanced by simulation.

5.1 Future Research

The research in future will focus on real experimental encoding and simulation. Matrix laboratory (MatLab) will be used to simulate the Huffman's method of encoding as well as the proposed RNS applied Huffman's algorithm using different file formats and sizes. The size and speed of execution will be determined for both algorithms to evaluate their performances.

6. REFERENCES

[1] Amamr A et al, 2001. A secured image coding scheme using Residue Number System. In: proceeding of the 18th national radio science conference, Egypt, pp 339-405.

[2] Omondi A and Premkuma B., 2007, Advances in Computer Science and Engineering: Texts – Vol. 2, Residue Number Systems Theory and Implementation. Imperial College Press 57 Shelton Street Covent Garden London WC2H 9HE, ISBN-13 978-1-86094-866-4.

[3] Weyori, B. A., et al, 2009. "Application of RNS to Huffman's Method of Secured Data Encryption Algorithm", International Journal of Soft Computing, 4(5):197-200, 2009, ISSN:1618-9503; Department of Computer Science, University for Development Studies, Navrongo, Ghana.

[4] Shannon, C. E., 1949. A Mathematical Theory of Communication. Bell Systems. Technical Journal. No. 27, PP: 379-423.

[5] Saravanan, C., and Surender, M.,2013. Enhancing Efficiency of Huffman Coding using Lempel Ziv Coding for Image Compression. International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-6, January 2013.

[6] Huffman, D. A., 1952. A method for the Construction of Minimum-redundancy Code. Proceedings of the Institute of Radio Engineers, 40(9): 1098-1101.

[7] Bankas, E. K., and Gbolagade, K. A., 2014. World Academy of Science, Engineering and Technology International Journal of Computer, Information, Systems and Control Engineering Vol:8 No:2, 2014.

[8] Gbolagade K. A., and Cotofana, S. D. 2009. Residue-to-Decimal Converters for Moduli Set with Common Factors, 52nd IEEE International Midwest Symposium on Circuits and Systems, pp 264-627.

[9] Yang, L. L., and Hanzo, L., 2001 "Redundant Residue Number System Based Error Correction Codes", Vehicular Technology Conference, VTC 2001 Fall, pp 1472 -1476, vol.3, 2001.

[10] Szabo, N., and Tanaka, R., 1967. Residue Arithmetic and its Application to Computer Technology. MC-Graw-Hill. New York. 1967.

[11] Gallager, R. G., 1987. Variations on a Theme by Huffman. IEEE. Trans. Inform. Theory, I.T. 24(6): 668-674.

[12] Wang, Y., 2000. "Residue-to-Binary Converters Based on new Chinese Remainder Theorems", IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing, pp197 -205, Vol.47, Issue 3, March 2000.

[13] Kulwar academic publisher (2002), "Residue Number Systems: Algorithms and Architecture.

[14] Ingels, F. M., 1971. Information and Coding Theory: Complete Edition, Intext, Scranton, Pennsylvania, USA, PP: 7-50. ISBN: 0631190724

[15] Connel, J. B., 1973. Huffman-Shannon-Fano Code. Proceedings of IEEE, 61(7): 1046-1047.

[16] Sun, J. D., and Krishna, H., 1992 "A Coding Theory Approach to Error Control in Redundant Residue Number System- Part II: Multiple Error Detection and Correction", IEEE Trans. on Circuits and Systems, pp. vol. 39, 18–34.

[17] Fano, R. M., 1949. Transmission of Information (Complete Edition), M.I.T., Cambrige University Press. Cambrige, England, PP: 593-597. ISBN: 978-1-60558-183-5.

[18] Cappellini, V., 1989. Data Compression and Error Control Technique with Application. 3rd Edn. Academic press, London. Pp: 9-37. ISBN: 0-8194-2427-7.

[19] Goh, V. T., and Siddiqi, M. U., 2008. "Multiple Error Detection and Correction Based on Redundant Residue Number Systems", IEEE Trans. On Communications, vol. 56, no. 3, pp. 325–330, March 2008.

[20] Mi Lu, 2004. "Arithmetic and Logic in Computer Systems", John Wiley & Sons, Inc., Hoboken, New Jersey, 2004.