International Journal of Soft Computing 4 (5): 197-200, 2009
ISSN: 1816-9503

Application of RNS to Huffman's Method of Secured Data Encryption Algorithm

B.A. Weyori, S. Akobre and G.K. Armah
Department of Applied Mathematics and Computer Science,
University of Development Studies, Navrongo, Ghana

Abstract: Data encryption is very important to the security of information or valuable data. The need for highly secured communication through a transmission medium is of paramount interest, recognizing the fact that most of our business and personal matters are involved. A new highly secured data encryption technique is proposed in this study. The residue number system has been effectively applied to data encryption with $(2^n-1, 2^n, 2^n+1)$ moduli set, which is an adaption of the Traditional Huffman's algorithm of encryption of data, where the frequency of occurrences are used to generate binary codes. This proposed algorithm will lead to an unreadable encrypted set of bits; except for the decoder with the moduli set can decrypt it.

Key words: Data encryption, Huffman's method, RNS, moduli set, encoder, decoder

INTRODUCTION

Data compression is often referred to as encoding, where a very general term is encompassing any special representation of data, which satisfies a given need. Information theory is defined to be study of efficient coding and its consequences in the form of speed of transmission and probability or error (Ingels, 1971).

Data encryption and compression may be viewed as a branch of information theory in which the primary objective is to minimize the amount of data to be transmitted or stored. A simple characterization of data compression is that it involves transforming a string of character in some representation (such as ASCII) in to a new string (of bits, for example), which contains the same information, but whose length is as small as possible. Data encryption and compression has important application in the area of data transmission and data storage. One important method of transmitting messages is to transmit the symbols in their appropriate places and sequences. If there are messages, which might be sent that have the same kinds of symbols, then some of the messages must use >1 symbol. If it is assumed that each symbol requires the same time for transmission, then the time for transmission of the message is directly proportional to the number of symbols associated with it (Connell, 1973; Faller, 1973).

Data encryption and compression is certainly important to the security or integrity of information to be transmitted through a network. The need for secured communication is more profound than ever, recognizing the fact that the conduct of almost all the business and personal matters are being carried out today by computer networks (Ammar *et al*., 2001). Hence, an efficient and low-complexity encryption and compression algorithm that can offer security for fast transmission and storage applications is of paramount importance to our information which must be secured against intrusion and threats that are continually increasing in frequency and sophistication.

Much of the available literature on data compression approaches the topic from the point of view of data transmission. As noted earlier, data compression is of value in data storage as well. Although, this discussion will be framed in the terminology of data transmission, encryption and decryption of data files for storage is essentially the same task as sending and receiving compressed data over a communication channel. The focus of this study is on algorithms for data compression; it does not deal with hardware aspects of data compression and transmission. Cappellini (1989) research focuses on a discussion of the techniques with natural hardware implementation.

MATERIALS AND METHODS

Shannon-Fano coding is a technique, which is constructed as follows, the source message a(i) and their probability p(a(i)) are listed in order of non-increasing probability. This is then divided in such a way as to form two groups of as nearly equal total probabilities as possible. Each in the first group receives a0 as the first digit of its codeword. The messages in the in the second half have codeword beginning with 1. Each of these

Corresponding Author: B.A. Weyori, Department of Applied Mathematics and Computer Science,
University of Development Studies, Navrongo, Ghana

groups is then divided according to the same criterion and additional code digits are appended as presented by Fano (1949) and Shannon (1949).

Huffman (1952) modified this Shannon-Fano coding and proposed a minimum redundancy coding technique which is expressed graphically, it takes as input a list of non-negative weights (w(1), ... w(n)) and construct a full binary tree a binary tree is full if every node has either zero or two children whose children leaves are labeled with the weights.

Gallager (1978) researched on the improvement of the Huffman coding technique which proved that an upper bound on the redundancy of Huffman codes of (p(n) + (log(2log e))/e, which is approximately p(n) + 0.086, where p(n) is the probability of the least source message.

A secured transmission of data through the computer network needs to be considered in some applications. One method of achieve a secured and high speed processing is to use the residue number system. The residue number system has been applied to enhance the Huffman's coding technique in this study.

A residue number system is defined in terms of relatively prime moduli set $(m_i)_{i=1,...n}$ such that GCD(m_i, m_j) for $i \neq j$, where, GCD means Greatest Common Divisor of m_i and m_j, while $M = \Pi^n_{i=1}\ m_i$ is the dynamic range. The residues of a decimal number can be obtained as $x_i = |X|_{mi}$ thus X can be represent in RNS as $X = (x_1, x_2, x_3,...x_n)$, $0 \leq x_i < m_i$.

This representation is unique for any integer in RNS, $X \in (0, M-1)$. For simplicity sake X mod m_i will be represented as $|X|_{mi}$ in this study is presented by Ammar *et al.* (2001) and Parhami (2000).

RNS is a carry-free system for addition, subtraction and multiplication operation. Given a two integer numbers K and L, RNS represented by $K = (k_1, k_2, k_3,...k_n)$ and $L = (l_1, l_2, l_3,...l_n)$, respectively. We note here that in this study, for simplicity sake we use the operator Θ for addition, subtraction and multiplication.

$W = K\Theta L$ can be calculated as $W = (w_1, w_2, w_3,...w_n)$, where $w_i = |k_i \Theta l_i|m_i$, for i = 1, n. This means that the complexity of the calculation of the operation Θ is determined by the number of bits required to represent the residue and not by the one required to represent the input operands (Gbolagade and Cotofana, 2008; Parhami, 2000; Wang *et al.*, 2002).

RNS system achieves high speed computation because of the parallel computing nature of the system. In order to convert numbers from binary to residues numbers a residue-to-binary converter is required at the front end and to convert back from residue to binary a residue-to-binary converter is required at the back end. The residue-to-binary, converter usually consists of a lot of modulo operation which is very tedious. The reverse converter (residue-to-binary) is a crucial part of the RNS system. To perform the conversion of residue-to-binary that is convert the residue number $(x_1, x_2, x_3,...x_n)$ into the binary number X, the traditional CRT is used (Gbolagade and Cotofana, 2008; Wang *et al.*, 2003). The traditional CRT is shown in Eq. 1:

$$X = \left| \sum_{i=1}^{n} M_i \left| M_i^{-1} x_i \right|_{m_i} \right|_M \quad (1)$$

Where:

$$M = \prod_{i=1}^{n} m_i,\ M_i = \frac{M}{m_i} \text{ and } M_i^{-1}$$

is the multiplicative inverse of M_i with respect to m_i.

RESULTS AND DISCUSSION

This algorithm consists of a simplified encoder with a very high security level and a decoder pair. The RNS is applied to the decimal number X, which is the frequency of each character which is used in the encryption process. The method of using the frequency is an adaption from the traditional Huffman's method of data coding/encryption.

Encoder: The moduli set $(2^n-1, 2^n, 2^n+1)$ is used in the forward conversion process to encode the decimal number to residues, which is the process of converting the frequency of occurrence of each particular character to residues. When the frequencies are converted to residues using the three moduli set $(2^n-1, 2^n, 2^n+1)$, the residues are then converted to binary as the encrypted bitstreams for each particular character.

The method of converting the residues into binary is an adaption of the traditional Huffman's method of data coding, where the end result of the coding process using the tree are zeros and ones (Fig. 1).

Decoder: The outputs of the RNS encoder are received as a small wordlength and are arranged in a certain order. The bitstream is first converted from the binary forms to residues. The residues are also converted back to the decimal number X, by the use of the modified CRT.

The modified CRT is used because the Traditional CRT has a large modulo M (dynamic range) operation and so it is not so efficient for the implementation as compared to the modified CRT. We apply the modified CRT that reduces the large modulo M presented by Wang *et al.* (1999, 2004) (Fig. 2).

$$X = x_1 + P_1 \left| \sum_{i=1}^{n} w_i x_i \right|_{p_2 \cdots p_n} \quad (2)$$

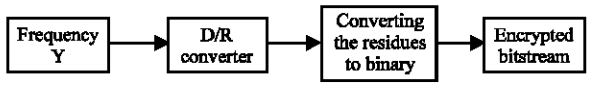

Fig. 1: A schematic diagram of the RNS encoder

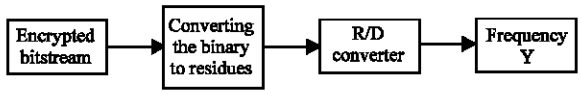

Fig. 2: A schematic diagram of the RNS decoder

Table 1: Show the characters and their frequencies

Characters	Frequency of occurrence
A	15
B	16
C	17
D	18
E	34

Table 2: Process of encoding using the new rns model of huffman's coding method

Characters	Frequency	Moduli set (3,4,5)	Encrypted bitstream binary form
A	15	0,3,0	0.11.0
B	16	1,0,1	1.0.1
C	17	2,1,2	10.1.10
D	18	3,2,3	11.10.11
E	34	1,2,4	1.10.100

Where:

$$m > 1, w_1 = \frac{N_1\left|N_1^{-1}\right|_{P_1} - 1}{P_1},\ w_i = \frac{N_i}{P_i} \text{ for } i = 2, 3, ..., m$$

$$x_1^{`} = x_1 \text{ and } x_i^{`} = \left|N_i^{-1}x_i\right|_{P_i},\ \text{for } i = 2, 3, ..., m$$

Security: The encrypted bitstream are send through a transmission line. These bitstream are arranged in a certain order. The intruder or the unauthorized person who breaks through the network does not understand the coding scheme used and does not also know the moduli set used in the conversion process.

Example 1: application of RNS to Huffman's method of secured data encryption algorithm: Table 1 and 2, are extracted from the study presented by Huffman (1952) and then modified for the process of the coding theory using RNS as a tool for the method of generating codes. The moduli set (2^n-1, 2^n, 2^n+1), the number of bit required is 2. Dynamic range, M = 3×4×5 = 60

CONCLUSION

In this study, a data encryption scheme is using RNS proposed and tested for its security levels. This scheme proposed adapts the traditional Huffman's method of data encryption using the frequency of occurrences of each particular character in the data or information and applying the residue number system with the moduli set (2^n-1, 2^n, 2^n+1) to encrypt the data.

This proposed scheme achieves a highly level of security and also speed of transmission of the bit through a computer network as compared to the traditional Huffman's method of data encryption.

REFERENCES

Ammar, A., A. Al-Kabbany, M. Youssef and A. Emam, 2001. A secure image coding scheme using residue number system. In: Proceedings of the 18th National Radio Science Conference, Egypt, pp: 339-405.

Cappellini, V., 1989. Data Compression and Error Control Techniques with Applications. 3rd Edn. Academic Press, London, pp: 9-37. ISBN: 0-8194-2427-7.

Connell, J.B., 1973. Huffman-Shannon-Fano code. Proceedings of IEEE, 61 (7): 1046-1047.

Faller, N., 1973. An adaptive system for data compression. Record of the 7th Asilomar Conference on Circuits, Systems and Computers: Parcific Grove, CA., pp: 593-597.

Fano, R.M., 1949. Transmission of Information (Complete Edition), M.I.T., Cambridge University Press, Cambridge, Englang, pp: 593-597. ISBN: 978-1-60558-183-5.

Gallager, R.G., 1978. Variations on a theme by Huffman. IEEE. Trans. Inform. Theory, IT-24 (6): 668-674.

Gbolagade, K.A. and S.D. Cotofana, 2008. Residue Number System operands to decimal conversion for three-moduli set. 51st IEEE Midwest Symposium on Circuits and Systems (MWSCAS), Knoxville, USA, pp: 791-794.

Huffman, D.A., 1952. A method for the construction of minimum-redundancy codes. Proceedings of the Institute of Radio Engineers, 40 (9): 1098-1101.

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

Parhami, B., 2000. Computer Architecture: Algorithms and Hardware Designs: Complete Edition, Oxford: University Press, New York, USA, pp: 5-40. ISBN: 0-19-512583-5.

Shannon, C.E., 1949. A mathematical theory of communication. Bell. Syst. Technical J., No. 27, pp: 379-423.

Wang, Y., X. Song, M. Aboulhamid and H. Shem, 2002. Adder based residue to binary numbers converters for (2n-1, 2n, 2n+1). IEEE. Trans. Signal Proc., 5 (7): 1772-1779.

Wang, W., M.N.S. Swamy and M.O. Ahmad, 2004. RNS application for digital image processing. 4th IEEE International Workshop on System-on-Chip for Real-time Application, pp: 77-80.

Wang, W., M.N.S. Swamy, M. Omair Ahmad and Yuke Wang, 2003. A study of the three-moduli set. IEEE. Trans. Circuits and Systems I: Fundamental Theory Appl., 50 (2): 235-245.

Wang, W., M.N.S. Swamy, M.O. Ahmad and Y. Wang, 1999. A comprehensive study of three moduli sets for Residue Arithmetic. Proceeding of IEEE Canadian Conference on Electrical and Computer Engineering, Alberta, Canada, pp: 513-518.