A NOVEL APPROACH OF DATA ENCRYPTION AND COMPRESSION METHOD USING STRANGE NUMBER SYSTEM

[1]Debasis Das, [2]Nitin Goje and [3]Ujjwal Lanjewar

[1]Assistant Professor, Department of Computer Science, Prerna College of Commerce, Nagpur, India

[2]Professor, Webster University, Tashkent, Uzbekistan.

[3]Professor & Principal, Prerna College of Commerce, Nagpur, India

E-mail: [1]debasisdas22583@gmail.com, [2]Nitin.goje@tiu.edu.iq, [3]ualanjewar@gmail.com

Abstract: *In the digital world the way of communication and transmission of data has dramatically changed. The rapid growth in digital era, explosive advancement of Information Technology and familiarization of Internet create a massive amount of data at each and every second from organizations, business and users by computer, mobile devices, cloud computing and Internet of Things. But with compare to the advancement of digital communication and massive volume of data, a less number of compression and cryptography techniques are proposed. Since the beginning, number system (i.e., decimal, binary, octal and hexadecimal) plays an important role in the computing. In this paper, we propose a new and better data encryption and compression technique using strange number system for general data.*

Keywords: **Strange Number System, Cryptography, Data Encryption, Compression, Bit reduction**

1. INTRODUCTION

With the rapid development of Information Technology and populization of Internet change our daily activities. Today, the way we communicate, the way we transmit data has dramatically changed. We can send any digital file to any digital devices at anytime from anywhere in the planet over the internet. In other words, we create a massive amount of data at each and every second using various electronic devices. As data is the life wire of every organization, data encryption is a key element to secure the data in the digital world.

In digital era, Data Encryption and Compression is the technique to representing the information in encrypts and compact form rather than its original or uncompressed form in data storage or transmission. Today the security is the vital issue to sending information from sender to receiver in online data transmission [1].

In cryptography various number system are used for the encryption and decryption from the beginning. Decimal, binary, octal and hexadecimal are some common number system and others are the strange number system (SNS). Strange number system is used to develop and implements various system architectures and computing.

Since last few decades, the research is continuing in the field of data encryption and compression. And various techniques and algorithms are already developed to encrypt and compress different data formats. In the meantime a variety of techniques and algorithms are available for data encryption and compression. Selecting a proper technique is always a compromise that tries to reconcile contrary characteristics.

But, the algorithm of data encryption and compression using strange number system will provide real physical security towards all possible ways of attacks while data transmission and to compact the data in some special format, such that the data occupies less memory or data can be transmitted in less time. In this paper, we propose a better data encryption and compression technique which is free from time complexity.

2. BACKGROUND

2.1 CRYPTOGRAPHY

Cryptography is one of the oldest techniques start to used at least 4,000 years back. Some scientist argues that cryptography appeared after writing was invented. The first cryptography concept in the documented use was in 1900 BC [2]. And after that, developments are growing in this field more, more and more… In past, cryptography was only used for military and diplomatic circles. From the 20th century, in the field of cryptography a rapid change was begun. Today's cryptography is heavily based on mathematical theory and computer science practice.

However, today the way we communicate with others, the way we transmit the data over internet is changed. Today's cryptography is more important in many aspects: from message to email, from e-Banking to e-Transaction, from e-Shopping to e-Business. There are so many cryptography techniques or algorithms are already developed like, 3DES, AES, RC4 etc. Cryptography using Strange Number System is a novel technique to encrypt and decrypt the data [3].

2.2 DATA COMPRESSION

In today's aspect, data compression is one of the important techniques to compact the large volume of data. There are two types of data compression 1) lossless compression and 2) lossy compression. In lossless compression, the compressed data can be recovered without any data loss and in lossy compression, the compressed photographs and videos can be recovered with some loss of image quality which is unnoticeable [4][5].

Some examples of lossless data compression include entropy encoding, Burrows-Wheeler Transform, Prediction by Partial Matching (also known as PPM), Dictionary Coders (LZ77 & LZ78 and LZW), Dynamic Markov Compression (DMC), Run-length encoding and context mixing. Examples of lossy data compression include vector quantization, A-law Compander, Mu-law Compander, Distributed Source Coding Using Syndromes (for correlated data), Discrete Cosine Transform, Fractal compression, Wavelet compression, Modulo-N code for correlated data and linear predictive coding. This is a new data compression algorithm using pentaoctagesimal strange number system [6] [7].

2.3 STRANGE NUMBER SYSTEM

Number system is used in everywhere in computing. There are two types of number system in computing – 1) Traditional Number System (i.e., decimal, binary, octal and hexadecimal number system) and 2) Strange Number System [8] (i.e., all other number system except traditional number system). Some of the strange number systems are unary, ternary … nonary, unodecimal … vigesimal, etc. In the digital era, binary number system is used at each and everywhere. But, the deficiencies of binary number system are steadily increasing in the field of computing. However, the strange number system is using to avoid the deficiencies of traditional number system. For the potential advantages of strange number system like, greater speed, greater density, better usage plays a significant role in computing than the traditional number system. Strange number system is also used to develop some system architecture like, ternary computer was developed at Masko University using the ternary number system. In this manuscript, we use strange number system in the field of cryptography and data compression.

2.4 DUOCENTIHEXAPENTADECIMAL NUMBER SYSTEM

The number system with base two hundred and fifty six is known as the duocentihexapentadecimal number system. In this system two hundred and fifty six symbols are used to represent numbers, these symbols are described in Appendix II. It is also a positional number system that each bit position corresponds to a power of 256. It has two parts, the integral part and the fractional part, set apart by radix point. For example (4z8.l3)256.

In duocentihexapentadecimal number system the leftmost bit is known as most significant bit (MSB) and the right most bit is known as least significant bit (LSB). The following expression shows the position and the powers of the base 256:

$$\ldots..256^{3}256^{2}256^{1}256^{0}.256^{-1}256^{-2}256^{-3}\ldots..$$

The arithmetic operations like addition, subtraction, multiplication and division operations of decimal numbers can be also performed on duocentihexapentadecimal numbers.

3. METHODOLOGY

First we need to read all the words in the input file and count the occurrence of each word. Then sort the words according to the order of their occurrence in descending order and store them in a file. If any word appears more than twice, replace the same using a special character and then modify the original file by replacing those words with their respective assigned special characters. Now we need to count the number of distinct characters from the modified file and assigned numerical codes to each distinct character and concatenate them to obtain binary output [9]. Then, add number of 0's in the most significant bit of binary output until it is divisible by 8 and find one's complement for every 8 bits binary data. After that, convert every 8 bits of binary data to equivalent decimal value and now convert each decimal value into Duocentihexapentadecimal Number System (Base – 256) and concatenate them to get final cipher and compressed file.

4. ALGORITHMS

4.1 STEPS FOR COMPRESSION

1. Begin
2. Input the text data to be compressed in the input file.
3. Find the number of occurrences of each word and sort them in descending order.
4. If a word appears more than twice, replace it with a special character (ASCII range 128-254) and maintain a dictionary or replaced words in an array.
5. Again read all the strings from this modified input file.
6. Find the number of distinct characters from the modified file.
7. Assign the numeric code to the distinct characters found in the step 5.
8. Starting from first character in the modified file; find the binary code of each distinct characters from assigned numerical codes and concatenate them to obtain binary output.
9. Add number of 0's in MSB of binary output until it is divisible by 8.
10. Find one's complement for every 8 bits of the previous binary data.
11. Convert every 8 bits of binary data to equivalent 3 digit decimal number.
12. Now convert each decimal value into Duocentihexapentadecimal Number System (Base – 256) and concatenate them.

13. Display the final result obtained in step 11. Result file is the final cipher and compressed file.
14. End

4.2 STEPS FOR DECOMPRESSION

1. Begin
2. Input the final output from compressed file.
3. Convert each character from Duocentihexapentadecimal Number System (Base-256) to decimal number and concatenate them.
4. Convert each three digit decimal value to its 8 bit binary equivalent.
5. Find one's complement for every 8 bits of the previous binary data.
6. Remove the extra bits from the binary output added in the compression phase.
7. Calculate the numeric code for every 8 bits obtained in the Step VI.
8. For every numeric value obtained in the step VII, find the corresponding set of distinct character.
9. Now map each distinct character appearing in the file extract from the dictionary mentioned in the compression routine and replace them to retrieve the final decompressed file.
10. Display the final result as an original text.
11. End

5. ALGORITHM ILLUSTRATION

For example we consider a text written: "My name is Debam. Debam is a good boy. Debam lives in Mumbai." Here 'is' and 'Debam' appears two times and three times, now these words should be replaced by special characters say '$' and '#'. Then the modified text look like this: "My name $ #. # $ a good boy. # lives in Mumbai." Here, the total number of characters in this modified text is 47 but the distinct character set is [M, y, n, a, m, e, $, #, ., g, o, d, b, l, v, s, u, i, ,].

Now, assign the numeric code to the unique symbols as: {M=1, y=2, n=3, a=4, m=5, e=6, $=7, #=8, .=9, g=10, o=11, d=12, b=13, l=14, i=15, v=16, s=17, u=18 and ' '=19}.

Find the binary value of each numeric code starting from first symbol as:

1=1; 2=10; 3=11; 4=100; 5=101; 6=110; 7=111; 8=1000; 9=1001; 10=1010; 11=1011; 12=1100; 13=1101; 14=1110; 15=1111; 16=10000; 17=10001; 18=10010; 19=10011

Concatenate all binary value and add number of 0's in MSB of final output until it is divisible by 8 as:

00011011100101110111100010011010111100110111101111100001000110010100 11

Now, find one's complement for every 8 bits of the previous binary data.

11100100011010001000011101100101010000110010000100000111011100111 0101100

Then, convert each eight bit binary value to decimal value as:

228104120154188222248070083

Finally, each three digit decimal value into Duocentihexapentadecimal Number System (Base – 256) and concatenate them as:

Ϫ π ϑ Є ц ґ ѫ+}

So, by representing in this way, the total size of the compressed file will be (9 × 8) bits= 72 bits whereas the original size of the text file was (61 × 8) bits=488 bits. Hence, a compression percentage of around 85.25% is achieved. In this way, the complete file can be compressed and hence, achieving a better compression ratio.

6. CONCLUSION

Data Encryption and Compression using strange number system, perhaps the next, largest step in computing, also provides the newest hopes for Data Encryption and Compression, creating the potential for new Data Encryption and Compression methods algorithms, obsolescing modern applications and algorithms at the same time. In this proposed work, a different bit reduction algorithm is developed to compress and encrypt the text data using Duocentihexapentadecimal Number System (Base – 256). This algorithm is applicable for different datasets such as Random, Alphanumeric, Numeral and Special Characters dataset. From the algorithm illustration, it is shown that the compression results by the proposed system are better than the existing systems.

REFERENCES

[1] R.S. Brar and B. Singh, “A survey on different compression techniques and bit reduction Algorithm for compression of text data” International Journal of Advanced Research In Computer Science and Software Engineering (IJARCSSE) Volume 3, Issue 3, March 2013.

[2] S. Porwal, Y. Chaudhary, J. Joshi and M. Jain, “Data Compression Methodologies for Lossless Data and Comparison between Algorithms” International Journal of Engineering Science and Innovative Technology (IJESIT) Volume 2, Issue 2, March 2013.

[3] Debasis Das, U. A. Lanjewar, S. J. Sharma3” Design an Algorithm for Data Encryption and Decryption Using Pentaoctagesimal SNS” International Journal of Computer Trends and Technology (IJCTT) – volume 6 number2 – Dec 2013.

[4] U. Khurana and A. Koul, “Text Compression and Superfast Searching” Thapar Institute of Engineering and Technology, Patiala, Punjab, India-147004.

[5] M. Kaur and U.Garg “ Lossless Text Data Compression Algorithm Using Modified Huffman Algorithm”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 7, July 2015, ISSN: 2277 128X.

[6] Anupama Mishra “Enhancing Security of Caesar Cipher Using Different Methods” IJRET: International Journal of Research in Engineering and Technology ISSN: 2319-1163, ISSN: 2321-7308.

[7] J. Ziv and A. Lempel, “Compression of individual sequences via variable length coding”, IEEE Transaction on InformationTheory ,Vol 24: pp. 530 – 536, 1978.

[8] Debasis Das, Dr. U A Lanjewar, “Realistic Approach of Strange Number System from Unary to Decimal,” International Journal of Computer Technology and Applications, vol. 3(1), pp. 235–241, January 2012.

[9] Debashis Chakraborty, Sandipan Bera, Anil Kumar Gupta and Soujit Mondal,, “Efficient Data Compression using Character Replacement through Generated Code”, IEEE NCETACS 2011, Shillong, India, March 4-5,2011.