A Detailed Survey on Text Data Security and Compression Techniques Using Huffman Coding Compression

Dr.N.Kavitha

HOD, Computer Science Department, Nehru Arts and Science College, Coimbatore, Tamil Nadu, India.

Anu Mattatholi

Computer Science Department, Nehru Arts and Science College, Coimbatore, Tamil Nadu, India.

Abstract – Data security and data compression is much helpful for effective data management. Several applications concentrating on the multi level data security and data compression process. The compression process saves storage space and makes the transmission easier. Huffman coding is the well known and popular compression technique widely used for text data compression. Among the lossy and lossless data compression techniques, Huffman code treated as an optimal solution for secured data compression. This paper gives a comprehensive analysis and of existing data compression techniques, which are related to the Huffman code is presented. And this survey also provides direction to solve problems of such systems.

Index Terms – Data Compression, data Security, Huffman coding, lossless data compression.

1. INTRODUCTION

Digital Information flow became very huge and occupied more storage spaces due to wide range of internet applications. Due to the continuous growth of data size, it is difficult to handle and access. So, data compression [1] is a best method to achieve high security and reduce storage space. And it also eases the transaction time. The multi level security and compression can be performed on different file formats such as text, image, audio; video etc. This paper gives an analysis on text data security and compression techniques. Multi level security and data compression have several research directions. This paper flows on Huffman coding based text compression schemes. The figure 1.0 shows the process flow involved with the data security and compression.

The data compression techniques are categorized into two types such as Lossy compression and lossless compression. Several authors have described about these compression techniques. Among loss and lossless [2], lossy compression is an effective technique. However, this is effective when the compression made on images and audios. This paper gives the popular approaches and detailed survey about lossless compression techniques used for the data compression.
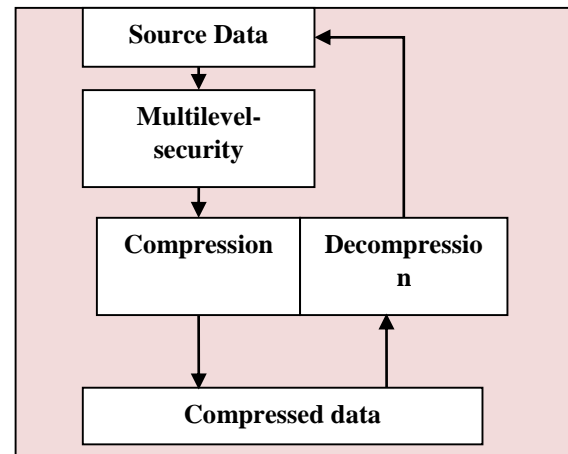

Fig 1.0 data compression and decompression

1.1 Data Compression Techniques:

Data compression and transmission consist of two steps such as modeling and coding, it takes the stream of symbols and transform into codes. The stream code size is determined the effectiveness of compression. If the stream of code is smaller than the original, then the compression is effective. There are two types of models can be usually used for lossless data compression. One is statistical and another one is dictionary oriented. The two models are unique with one another and the data can be embedded by every single symbol at a time in statistical model. This utilizes the character sequence selection in probabilistic manner to perform compression. Every single character will be replaced by symbols by single code method in the dictionary based compression method. However, when comparing with the statistical method, the dictionary modeling method is more significant and reduces the encoding problems.

The very popular methods for effective data compression is Huffman coding, Adaptive Huffman coding, Arithmetic encoding, Shannon entropy, Run-Length encoding and so on

[3]. However, there are numerous techniques available, currently a lot of researches commencing for better approach for securing and compressing the text data. This is very optimal when the data performed encryption, decryption and also to compression of the text data.

1.2 Huffman Compression Techniques:

In lossless data compression, Huffman code [4] is the popular and effective technique which follows a prefix code generation process. This technique creates a binary tree and generates different symbols with probability. In Huffman encoding an unique prefix code is assigned to each symbol The Huffman compression techniques are two types, one is static and another one is dynamic, where the static Huffman coding initially calculates the frequency and reads the content again to compress. So, the static Huffman code compression reads the data twice. Whereas the dynamic Huffman code initiated with the empty Huffman tree and modifies it as symbols. The compression and decompression will change the tree in a same way that used for the compression. Huffman decoding can start from any point as it is based on codeword for each symbol. Network related applications can use the Huffman compression technique.

2. LITERATURE SURVEY

Data security and compression of text data is studied together in recent approaches. In this section, some text data security using Huffman encoding and bit stuffing mechanisms are discussed. A significant amount of researches concentrated on research related process such as to encryption, decryption and compression of text data. Few works are reviewed in this paper.

In paper [5], authors Gulhane, suraj et al proposed a technique with the dynamic Huffman coding scheme for secure and speed data retrieval. Authors have implemented the concept of DDAS (A Distributed Data Aggregation Service) using Kerberos. This technique increases the security and it ensures the only authorized client is able to access distributed database and for compression and decompression of a data method. This improved the security and data retrieval using the adaptive Huffman coding.

In the paper [6] authors Subhra J. Sarkar et al, discussed about the data storage and security problems. The authors used Huffman Coding based data compression technique. The technique is improves the security and reduces the size of high dimensional data array.

In the paper [7], authors Hameed, Maan, performed an effective compression of text data by applying the lossless method of Huffman coding. This technique has achieved fast data compression and converts into confidential data array.

Authors in paper [8] developed a multilevel security and data compression technique by applying Huffman coding and bit stuffing algorithms. Authors have implemented and prove the

bit stuff and Huffman coding can provide high level security and high performance on compression processes. The data compression technique reduces the transmission time and bandwidth utilization.

In paper [9] authors developed an energy efficient data compression technique specially designed for wireless sensor network. The compression is performed using Huffman code and additionally a diffusion feedback mechanism is proposed. This variation is renamed as "LiHuffman"

Table 1.0 Summary of the previous data aggregation approaches with Huffman code.

Paper id	Summary
1	Introduction about data compression is studied.
2	The block Lossless compression is studied.
3	A review on various Lossless text data compression techniques are studied
4	Data compression with Huffman code is proposed. The technique follows a prefix code generation process for data compression.
5	A dynamic Huffman coding scheme is proposed. This is mainly concentrated on secure and speed data retrieval. A Distributed Data Aggregation Service is proposed to achieve high secure and fast data compression and retrieval.
6	Data compression technique with Huffman code for high dimensional data is proposed.
7	Lossless method of Huffman coding is used for compression and later authors converted the compressed data into confidential data array.
8	Multilevel security and data compression technique by applying Huffman coding and bit stuffing algorithms are proposed. This technique was successful in various aspects such as transmission time and bandwidth utilization.
9	LiHuffman compression technique is proposed with a diffusion feedback mechanism.

The table 1.0 provides the summary about the existing compression works based on Huffman code .

3. CONCLUSION

Data security and storage reduction tasks are more important in the current trend. The analysis of encoding techniques and tools for compression is discussed in this survey. In this paper, we specifically concentrated on the recent data compression techniques using Huffman coding related works and its drawbacks. The survey gives the technique of existing compression techniques for text data. However, the Huffman coding compression is popular, the execution issues still arises. This paper gives idea about such issues in brief. From this analysis, an optimal solution can be found in the future.

REFERENCES

[1]. Sayood, Khalid. *Introduction to data compression*. Morgan Kaufmann, 2017.

[2]. Chang, Weiling, Binxing Fang, Xiaochun Yun, and Shupeng Wang. "The block lossless data compression algorithm." *International Journal of Computer Science and Network Security (IJCSNS)* 9, no. 10 (2009): 116.

[3]. Sharma, Neha, Jasmeet Kaur, and Navmeet Kaur. "A review on various Lossless text data compression techniques." *International Journal of Engineering Sciences, Issue* 2 (2014).

[4]. Chau, Savio N., and Ridwan Rashid. "Data compression with Huffman code on multicore processors." U.S. Patent 9,258,013, issued February 9, 2016.

[5]. Gulhane, Suraj, and Sonali Bodkhe. "DDAS using Kerberos with Adaptive Huffman Coding to enhance data retrieval speed and security." In *Pervasive Computing (ICPC), 2015 International Conference on*, pp. 1-6. IEEE, 2015.

[6]. Sarkar, Subhra J., Nabendu Kr Sarkar, and Antra Banerjee. "A novel Huffman coding based approach to reduce the size of large data array." In *Circuit, Power and Computing Technologies (ICCPCT), 2016 International Conference on*, pp. 1-5. IEEE, 2016.

[7]. Hameed, Maan, Asem Khmag, Fakhrul Zaman, and Abd Rahman Ramli. "A New Lossless Method of Huffman Coding for Text Data Compression and Decompression Process with FPGA Implementation." *Journal of Engineering and Applied Sciences* 100, no. 3 (2016): 402-407.

[8]. Kodabagi, M. M., M. V. Jerabandi, and Nagaraj Gadagin. "Multilevel security and compression of text data using bit stuffing and huffman coding." In *Applied and Theoretical Computing and Communication Technology (iCATccT), 2015 International Conference on*, pp. 800-804. IEEE, 2015.

[9]. Karthikeyan, B., R. Kumar, and Srinivasa Rao. "Energy efficient data compression and aggregation technique forwireless sensor networks." *Nextgen Electronic Technologies: Silicon to Software (ICNETS2), 2017 International Conference on*. IEEE, 2017.