E-ISSN: 2708-3977
P-ISSN: 2708-3969
IJEDC 2022; 3(2): 24-28

www.datacomjournal.com
Received: 04-05-2022
Accepted: 07-06-2022

Angham Khalid Hussein
Assistant Lecturer,
Department of Computer Technique Engineering, Al. Turath University Collage, Baghdad, Iraq

Correspondence
Angham Khalid Hussein
Assistant Lecturer,
Department of Computer Technique Engineering, Al. Turath University Collage, Baghdad, Iraq

Effect of data compression on computer networks performance

Angham Khalid Hussein

Abstract
In this paper a study and analysis of data compression schemes effect on network transmission and performance are applied using mat lab programming tool and Cisco packet tracer network simulator software is used for design of a network and measure the different performance metrics before and after compression. A two data compression method Huffman and RSA used on the transmitted data file and the effect of these method on file size and the consumed time to transmit this file taking into account the time required to compress the file.

Keywords: Huffman and RSA, consumed time

1. Introduction
Data compression has a very wide application in different life aspects like medical [1], multimedia and image processing, [2-4], wireless sensor network (WSN) which is a very interesting networking technology [5] as well as computer network, the mathematical computational methods had historically outperformed improvement of the topologies and interconnects of the network. A score counts per compute node had increased, the network bandwidth amount that is available per the compute node had risen, however, not sufficiently fast for keeping up with the node's increasing computational performance. Such gap has been expected to be increased when entering the computing area, with the super-computers that contain hundreds of the computational cores in each computational node. In this new field, interconnect band-width between the nodes and a variety of the file systems will beat premium. The compression of the data prior to sending it to Input/Output nodes results in reducing load upon network. Even in the case where Input/Output nodes require decompressing data prior to sending it on external network, the compression may still be beneficial due to the fact that that several of the compression approaches are asymmetrical according to the throughput the de-compression is usually a magnitude order faster compared to the compression, ensuring the fact that the de-compression of the data utilizing limited amount of the Input/Output nodes doesn't tur into bottle-neck [6]. Two compression techniques have been defined, which include either lossless or lossy. The lossless compression results in the reduction of the bits through the identification and elimination of the statistical redundancy and nothing has been lost in the information. The 2nd approach is the lossy compression that decreases the bits through the removal of the less important or unnecessary information [7-9]. Usually, any device applying the data compression has been known as encoder, which can be use many different techniques even neural network [10] in one or more level [11]. Compression is good due to the fact that it reduces resources required for storing and transmitting data. Some of the computational sources are consumed in the compression and decompression processes which are considered one of tradeoff of compression. Data compression system design has some trade-offs amongst a variety of the factors, which include the compression degree, the distortion amount that has been introduced, particularly in the case of utilizing the lossy compression of data, and computational sources that are needed required for the compression and decompression of data [12]. Compression and coding of the data are considered a good tool in enhancing network performance because it implies sending or storing a smaller number of bits which will require less time to send data [13].

2. Proposed Method

In this section the effect of data compression in network performance are analyzed, first step a data file are compressed using DCT and Huffman coding and a computer programs are written using mat lab program. A two programs are written to perform the steps of compression of each methods separately and the compression result obtained**.** Also provide information and details about main idea compression methods lossy and lossless and its details and mathematical formula for Huffman as a lossless method and DCT lossy and also illustrates the main metrics used to measure the network performance.

2.1 Data Compression

2.1.1 Lossy and lossless data compression

Compression was extensively utilized in the wireless networks, like the 3-G networks, and for the optimizations of the web-sites for the reduction of end-to-end time of the transmission [14-20]. In those situations, the compression had resulted in the reduction of the latency of the transfer and improvement of the response times. The compression results in the improvement of the network end-to-end time of the transmission, band-width of the network, and subsequently throughput of the file Input/Output. A wide range of the researches were carried out on compression effects for the minimization of the consumption of energy. There are 2 kay compression classes [21-25].

2.1.2. Lossless compression

It depend on reducing the number of bits through the identification and elimination of the statical redundancy. The number of information bit is the same before and after compression, Lempel-Ziv, Huffman coding and run length encoding are lossless compression techniques [26].

Huffman Coding Method

Huffman code has been mapped to symbols of fixed length to codes with different lengths. The Huffman algorithm has been based upon list of all data or symbols that have been arranged in the descending order of the likelihoods. Based on those likelihoods, code words have been assigned. Longer code words have been assigned for smaller probabilities and shorter code words for the higher ones. After that, it creates binary tree, by a bottom up method with symbol at each one of the leaves. Which has a few of the steps, in every one of the steps, 2 symbols with minimal frequency values have been selected, then added to the partial tree's top. The chose minimal frequency symbols have been eliminated from the list, substituted with secondary symbol that denotes the 2 original symbols. Which is why, the list has been decreased to a single secondary symbol, denoting the fact that the tree is complete. Finally, a code word is assigned for every one of the leaves or symbols according to the path from root node to the symbols in that list.

The following are the steps of the Huffman logarithm used in compress the data file that contain an image:

1. Arranging possibilities of character frequencies for downloading form and considering it as final series in the tree form.
2. Repeating this process in the case where there's an additional one note in the diagram of the tree.

A**.** every 2 notes with lower frequency repetition likelihood which equals result note repetitions.

B**.** Coding every pair of the branches of the tree diagram in the dual formula. The benefit of that approach includes code length stability besides the fact that there aren't any intervals amongst the codes, which results in causing the ambiguity throughout the code opening and returning a file to the original state, figure1 shows the steps of Huffman coding [27, 28].

2.1.3 Lossy compression:

In lossy compression it reduces the number bit by identified the required information and eliminating it, this method of the compression require some loss of data that can't recovered to original data. Which indicates that data distortion is quite low. Lower distraction results in better outcomes and amongst numerous logarithms is the Discrete Cosine Transform (DCT), which expresses finite sequence of the data points based on the summation of the cosine functions that oscillate at a variety of the frequencies. DCT, has been defined as one of the commonly utilized techniques of the transformation in the data compression and signal processing. It's utilized in the majority of the digital media, which includes the digital images (like the HEIF and JPEG**,** in which the little components of high-frequency may be disregarded), digital audio (e.g. MP3, Dolby Digital and AAC), digital videos (e.g. H.26x and MPEG), digital radio (e.g. DAB+ and AAC+), digital TV (e.g. HDTV, VOD and SDTV,), and speech coding (e.g. Siren, Opus and AACLD). DCT is similar to DFT: as it performs the transformation of an image or signal from spatial to frequency domains, as can be seen from Fig. 1.

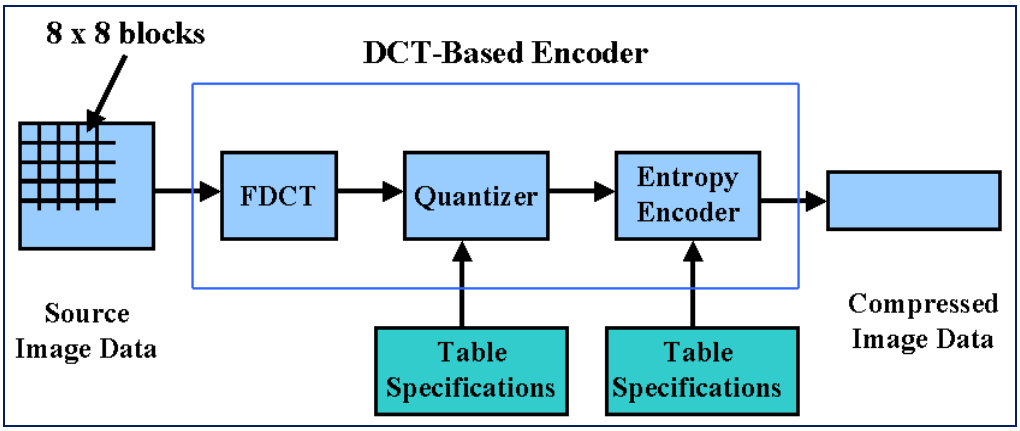

Fig 1: Transformation DCT function

DCT formula

The DCT can be mathematically represented in equation (1) where p (x, y) represents x, y[th] element of an image that is represented by matrix p. N represents the block size which DCT is performed on.

$$D(i,j) = \frac{1}{\sqrt{2n}} C(i)C(j) \sum_{x=0}^{n-1} p(x,y) cos\left[\frac{(2x+1)i\pi}{2n}\right] \cos\left[\frac{(2y+1)j\pi}{2n}\right] \quad (1)$$

$$C(u) = \left\{ \begin{matrix} \frac{1}{\sqrt{2}} \; if \; u = 0 \\ 1 \; if \; u > 0 \end{matrix} \right\} \quad (2)$$

The eq. performs the calculation of 1 entry (i, j [th]) of a transformed image from the original image matrix's pixel values. Due to the fact that DCT utilizes the cos functions, the resultant matrix is dependent upon diagonal, horizontal and vertical frequency values. Which is why, an image black with much frequency change has a highly random looking resulting matrix, whereas image matrix of only 1 color, has resulting matrix of a high value for 1[st] element and 0's for the rest of the elements [27-30].

2.2 Network Performance Parameters

The network performance notion is associated with the service which is provided by it to the users. For example, a content delivery network (CDN) provides services for the delivery of the content to the users instead of the content provider. The CDN performance notion is dependent upon the content type which it delivers. For the CDN which delivers e-commerce site to the users, usually, the efficiency is measured based on page download times as well as page availability. The availability is a measure of the time percent the user is capable of downloading a website page with no failure, and the time of the download is a measure the speed by which a page has been downloaded and rendered by a browser [31, 32]. The performance can be measured by many matrices like the delay (latency), bandwidth, jitter, packet loss, and throughput.

2.2.1 Bandwidth

The band-width (also referred to as the bit rate can be defined as the network's capability in moving the data amount alternatively over a time unit, it represents the rate at which the network carries an application traffic.

2.2.2 Latency

Latency can be defined as time that is needed for the transmission of a packet from a source device to a destination one. It can be defined as well, as time delay to travel via network, which includes all intermediate nodes from a source to destination.

2.2.3 Jitter

For packet stream, the jitter can be defined as measure of the packets' latency variation. In some of the cases, it's also referred to as the variation in the packet delay.

2.2.4 Packet Loss

Packet loss can be defined as percentage of the packets arriving with the errors or failing to get to destination. It results from network congestion in intermediate nodes or links which don't have the needed band-width.

2.2.5 Throughput

Throughput is the actual measure of the speed of data transmission via a network. Even though, initially, the band-width in the bits per second and the throughput appear to be identical, however, they're not. Band-width can be defined as potential measurement of link and throughput represents actual measurement of the speed of data transmission [30, 31].

3. Results and Analysis

In the present study, the effect of data compression in network performance are analyzed, first step a data file are compressed using DCT and Huffman coding and a computer programs are written using matlab program. a two programs are written to perform the steps of compression algorithm explained previously of Huffman coding are applied on image and the compression result obtained are shown in Table1 with compression ratio 2 and Figure 2 shows the image after applying this method as we can see there is no visible different between the images before and after compression. DCT compression are used by applying DCT eq. 1 and eq.2 on the same image this result compressed image as shown in figure.2 with compression ratio 1.7 Table2 show the time required for compression in each technique. To steady the effect of data compression on computer networks performance Cisco packet tracer network simulator is used to create a local area network consist of server and computer as shown in Figure 3 and simulate the operation of sending data file and observing the time and byte rate (byte/second) for the original file before and after compression using both Huffman coding and DCT compression and the result are shown Table 2. Other Network performance factors are calculated as shown in Table 3.

Table 1: Compression ratio for each technique

File size	DCT lossy compression ratio	Huffman lossless compression
72 KB	1.7	2

Fig 2: Results from DCT and Huffman coding

Fig 3: Computer network used to measure performance connected using Cisco packet tracer

Table 2: Time for each compression method

Compression technique	Compression time(sec)
DCT	2
Huffman coding	29

Table 3: Network performance results

File	Throughput	Latency	Bandwidth
Original	34.816	2.26	100MB/SEC
DCT file	37.974	1.094	100MB/SEC
Huffman coding file	35.416	1.016	100MB/SEC

4. Conclusion

From the result above in table 1.we can see that the original file is compressed with compression ratio of 1.7 and 2 for DCT and Huffman coding which mean successful redaction of original file size. This reduction will improve network performance because as shown in table 3. Will reduce latency required for the transmission of data over network. the throughput for the three cases are close to each other because the throughput depend on the number of the packets that the network can transmit each second and the bandwidth used is 100Mb/sec for the three cases and because we use simulation for measuring network performance metric there is no packet loss or jitter. So it very effective to use data compression methods for reducing time and improve performance for computer network but some limitation in this process is the time required for compression which depend on the speed of the computer processor that applies compression technique to reduce this problem it's possible to use high speed computer.

5. Acknowledgements

I'm deeply thankful to my family especially my mother and my husband which they are very supportive and helpful during the writing of this paper.

6. Reference

1. Mustafa Emad Hameed, Masrullizam Mat Ibrahim, Nurul fajar Abd Manap. Compression and encryption for ECG biomedical signal in healthcare system, TELKOMNIKA. Accredited First Grade by Kemenristekdikti. 2019 Dec;17(6):2826-2833. ISSN: 1693-6930. Decree No: 21/E/KPT/2018 DOI: 10.12928/TELKOMNIKA.v17i6.13240.
2. Bo Wang, Yubin Gao. An Image Compression Scheme Based on Fuzzy Neural Network, TELKOMNIKA, 2015 Mar;13(1):137-145. ISSN:1693-6930. Accredited A by DIKTI, Decree No: 58/DIKTI/Kep/2013 DOI: 10.12928/TELKOMNIKA.v13i1.1270.
3. Mohsen Nasri, Abdel Hamid Helali, Halim Sghaier, Hassen Maaref. Efficient JPEG 2000 Image Compression Scheme for Multihop Wireless Networks. TELKOMNIKA. 2011 Aug;9(2):311-318. ISSN: 1693-6930. Accredited by DGHE (DIKTI), Decree No: 51/Dikti/Kep/2010.
4. Walaa M. Abd-Elhafiez, Wajeb Gharibi, Mohamed Heshmat. An efficient color image compression technique, TELKOMNIKA Telecommunication, Computing, Electronics and Control. 2020 Oct;18(4): 2371-2377. ISSN:1693-6930. Accredited First Grade by Kemenristekdikti, Decree No: 21/E/KPT/2018 DOI: 10.12928/TELKOMNIKA.v18i5.8632
5. Osama F Abdel Wahab, Aziza I Hussein, Hesham FA Hamed, Hamdy M Kelash, Ashraf AM Khalaf, Hanafy M Ali. Hiding data in images using steganography techniques with compression algorithms, TELKOMNIKA. 2019 Jun;17(3):1168-1175. ISSN:1693-6930. Accredited First Grade by Kemenristekdikti, Decree No: 21/E/KPT/2018 DOI: 10.12928/TELKOMNIKA.v17i3.12230.
6. Benjamin Welt, Dries Kimpe, Jason Cope, Christina M Patrick, KamilIskra Robert Ross. Improving I/O Forwarding Throughput with Data Compression, IEEE international conference on cluster computing, 2011.
7. Wade Graham. Signal coding and processing (2ed.), Cambridge University Press, 34, ISBN 0-521-42336-8.
8. Mahdi OA, Mohammed MA, Mohamed AJ. Implementing a Novel Approach a Convert Audio Compression to Text Coding via Hybrid Technique, International Journal of Computer Science Issues. Retrieved. 2013 Mar 6;9(6-3):53-59.
9. Pujar JH, Kadlaskar LM. A New Lossless Method of Image Compression and Decompression Using Huffman Coding Techniques, Journal of Theoretical and Applied Information Technology. 2010 May;15(1):18-23.
10. Hussein AK. Encryption System by Using BAM Artificial Neural network, Journal of AL-Turath University College, 2014, 53-61.
11. Hussein AK. Multistage encryption system using bidirectional associated memory neural network, International Journal of Science and Engineering Invention. 2019;5(7):1-4.
12. Salomon David. A Concise Introduction to Data Compression, Berlin, Springer, ISBN 9781848000728(2008).
13. Chen X, Wang W, Wei G. Impact of HTTP compression on web response time in asymmetrical wireless network in NSWCTC '09: Proceedings of the International Conference on Networks Security, Wireless Communications and Trusted Computing, 2009, 679-682.
14. Mogul JC, Douglis F, Feldmann A, Krishnamurthy. Potential benefits of delta encoding and data compression for HTTP, in SIGCOMM 97: Proceedings of the ACMSIGCOMM '97 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication, 1997, 181-194
15. Dong W, Chen X, Xu S, Wang W, Wei G. Proxy based object packaging and compression: A web acceleration scheme for UMTS, in WiCOM'09: Proceedings of the 5th International Conference on Wireless Communications, Networking and Mobile Computing, 2009, 4965-4969
16. Ma KJ, Bartos R. Analysis of transport optimization techniques, in ICWS '06: Proceedings of the IEEE International Conference on Web Services, 2006, 611-620
17. King AB. Website Optimization, O'Reilly, 2008 Jul, ISBN: 978-00596-51508-9, first edition.
18. R RL Mattson, Ghosh S. HTTP-MPLEX: An enhanced hypertext transfer protocol and its performance evaluation, J Network. Computer. Appl. 2009;32(4):925-939.
19. Deshpande S, Zeng W. Scalable streaming of JPEG200 images using hypertext transfer protocol, Proceedings of the ninth ACM international conference.

20. Kothiyal R, Tarasov V, Sehgal P, Zadok E. Energy and performance evaluation of lossless file data compression on server systems, in SYSTOR '09: Proceedings of SYSTOR2009: The Israeli Experimental Systems Conference, 2009, 1-12.
21. Alameldeen AR, Wood DA. Adaptive cache compression for high-performance processors, in ISCA '04**:** A ZX Proceedings of the 31st Annual International Symposium on Computer Architecture, 2004, 212. on Multimedia, 2001, 372-381.
22. Canal R, Gonzalez A, Smith JE. Very low power ́pipelines using significance compression, in MICRO 33: Proceedings of the 33rd annual ACM/IEEE International Symposium on Micro architecture, 2000, 181-190.
23. Barr KC, Asanovic K. Energy-aware lossless data compression, ACM Trans. Computer. System. 2006;24(3):250-291.
24. Das R, Mishra A, Nicopoulos C, Park D, Narayanan V, Iyer R, *et al*. Performance and power optimization through data compression in network-on-chip architectures, in IEEE 14th International Symposium on High Performance Computer Architecture, 2008, 215-225.
25. Tank MK. Implementation of Lempel-Ziv algorithm for lossless compression using VHDL. Implementation of Lempel-Ziv algorithm for lossless compression using VHDL. Think quest 2010: Proceedings of the First International Conference on Contours of Computing Technology. Berlin: Springer, 2011, 275-283. Doi: 10.1007/978-81-8489-989-4_51. ISBN 978-81-8489-988-7.
26. Abhishek Kaushik, Maneesha Gupta. Analysis of Image Compression Algorithms International Journal of Engineering Research and Applications (IJERA). 2012 Mar-Apr;2(2):773-779. ISSN: 2248-9622 www.ijera.com.
27. Hervey Allen, Carlos Vicente. Network Startup Resource Center Algorithms 2nd, edi. 18 terhttp://www.cs.princeton.edu/introalgsds/65compression.
28. Davide salmon. Data compression: the complete reference, springer science and business media, third edition, 2004, 195-329.
29. A Study on Network Performance Metrics and their Composition Andreas Hanemann1, Athanassios Liakopoulos2, Maurizio Molina3, D. Martin Swany.
30. Khalid Sayood. Introduction to Data Compression, 3rd edition, university of Nebraska, Lincoln, Elsevier science (USA), 2006.
31. John Dilley, Bruce M Maggs, Jay Parikh, Harald Prokop, Ramesh K Sitaraman, William E Weihl. Globally distributed content delivery. IEEE Internet Computing. 2002;6(5):50-58.
32. Nygren E, Sitaraman RK, Sun J. The Akamai Network: A platform for high-performance Internet applications. ACM SIGOPS Operating Systems Review. 2010;44(3):2-19.
33. Ze-Nianli Mark S, Draw Jiangchuan Liu. Fundamentals of multimedia, 2nd edition, springer, 2017.
34. Behrouz A Forouzan. Data communication and networking, 4th edition, ISBN 0072967757, huge media, 2007.