Journal of Science, Computing and Engineering Research (JSCER)
Volume-6, Issue- 9, October 2023.
DOI: *https://doi.org/10.46379/jscer.2023.060903*

An Approach to Protect the Data in Cloud Data Storage by Using Cloud Data Compression Techniques

R. krishnan, R. Mananda

Dept of CSE, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore (Tamil Nadu), India.

Article Information

Received : 06 Sep 2023
Revised : 14 Sep 2023
Accepted : 10 Sep 2023
Published : 16 Oct 2023

Corresponding Author:
R. krishnan

Abstract—: Cloud computing significantly plays a role in the aspect of effective resource utilization and service consumption. Irrespective of the type of clouds (ex. Private, public, hybrid or inter-cloud), every service providers concentrates on the data residing in cloud servers. Each and every moment, the researchers and scholars are proposing multiplicity of security algorithms to secure cloud data during the transactions. Most of the cloud data secure algorithms are focusing on the way to secure to cloud data in a single direction by using cryptographic algorithms. In this research paper focuses on a new direction to combine the features of data compression with the cloud data in order to secure the cloud data storage.

Keywords: ***Cloud, data, storage and compression.***

Citation: Ramlal, Mohit banawar "An Approach to Protect the Data in Cloud Data Storage by Using Cloud Data Compression Techniques", Journal of Science, Computing and Engineering Research, 6(9), 13-17, October 2023.

I. INTRODUCTION

Either to consider a Profitable or non-profitable organizations, dedicated resource utilization brings more elevation in the economic impact and makes huge loss. In order to overcome this defect, every clients hunt for a new technology to solve their demand with minimum effort. In this aspect, the cloud computing provides an excellent environment for the resource seekers over the network. Most of the categories, the cloud service providers are not consider the secure way of data transaction under the public cloud. But, in the same time, private cloud give more attention to secure the data resides in its cloud servers as well as to maintain the security for confidential cloud data. The general cloud storage mechanism is comprised with two major components such as data and its applications. Both data and applications are always handling with the help of cloud data owner and cloud service provider [5]. The most challenging task in the cloud data servers are focusing on the handling of residing data under the category of private and confidential sector classification. In order to ensure the secure data in cloud storage by using a cryptographic mechanism apply encryption on the storage sector and decryption on the authenticated receiving sector [1]. In the aspect of applying cryptographic algorithms such as RSA (Rivest, Shamir and Aldimer) and AES (Advanced Encryption Standard) regarding to secure cloud data using its own way of number theoretical information's and key exchange values [2]. The data or information is required to store in cloud service provider to keep as it is not guarantee for secure data. Because of this reason, the researchers concentrate on cryptographic mechanism. It provides only the crypto mechanism as the solution provider in order to secure the cloud data residing in Cloud Service providers. Most of the cloud security algorithms are focusing on the above said algorithms with performance analysis and comparisons of its required amount of storage [3]. The following block diagram (figure 1) depicted in the existing security mechanism in cloud data. In this research paper to frame, a challenging architectural framework as well as to implement the secure cloud data with data compression mechanism.

II. RELATED WORK

Huffman Coding Huffman coding [7] is an entropy encoding algorithm used for lossless data compression. It uses a specific method for choosing the representation for each symbol, resulting in a prefix-free code that expresses the most common characters using shorter strings of bits than are used for less common source symbols. Huffman coding is optimal when the probability of each input symbol is a negative power of two. Prefix-free codes tend to have slight inefficiency on small alphabets, where probabilities often fall between these optimal points.

"Blocking", or expanding the alphabet size by coalescing multiple symbols into "words" of fixed or variable-length before Huffman coding, usually helps, especially when adjacent symbols are correlated. Prediction by Partial Matching (PPM) [8,9] is an adaptive statistical data compression technique based on context modeling and prediction. In general, PPM predicts the probability of a given character based on a given number of characters that immediately precede it. Predictions are usually reduced to symbol rankings. The number of previous symbols, n, determines the order of the PPM model which is denoted as PPM(n). Unbounded variants where the context has no length limitations also exist and are denoted as PPM*. If no prediction can be made based on all n context symbols a prediction is attempted with just n-1 symbols. This process is repeated until a match is found or no more symbols remain in context. At that point a fixed prediction is made.

PPM is conceptually simple, but often computationally expensive. Much of the work in optimizing a PPM model is handling inputs that have not already occurred in the input stream[7]. The obvious way to handle them is to create a "neverseen" symbol which triggers the escape sequence. But what probability should be assigned to a symbol that has never been seen. This is called the zero-frequency problem. PPM compression implementations vary greatly in other details. The actual symbol selection is usually recorded using arithmetic coding, though it is also possible to use Huffman encoding or even some type of dictionary coding technique. The underlying model used in most PPM algorithms can also be extended to predict multiple symbols. The symbol size is usually static, typically a single byte, which makes generic handling of any file format easy. LZ77 LZ77 algorithms achieve compression by replacing repeated occurrences of data with references to a single copy of that data existing earlier in the input (uncompressed) data stream. A match is encoded by a pair of numbers called a length distance pair, which is equivalent to the statement "each of the next length characters is characters behind it in the uncompressed stream". To spot matches, the encoder must keep track of some amount of the most recent data, such as the last 2 kB, 4 kB, or 32 kB. The structure in which this data is held is called a sliding window, which is why LZ77 is sometimes called sliding window compression. The encoder needs to keep this data to look for matches, and the decoder needs to keep this data to interpret the matches the encoder refers to. The larger the sliding window is, the longer back the encoder may search for creating references. It is not only acceptable but frequently useful to allow length-distance pairs to specify a length that actually exceeds the distance. As a copy command, this is puzzling: "Go back four characters and copy e10characters from that position into the current position"[7]. How can ten characters be copied over when only four of them are actually in the buffer? Tackling one byte at a time, there is no problem serving this request, because as a byte is copied over, it may be fed again as input to the copy command. When the copy-from position makes it to the initial destination position, it is consequently fed data that was pasted from the beginning of the copy-from position. The operation is thus equivalent to the statement "copy the data you were given and repetitively paste it until it fits". LZ78 LZ78algorithms achieve compression by replacing repeated occurrences of data with references to a dictionary that is built based on the input data stream. Each dictionary entry is of the form dictionary[...] = {index, character}, where index is the index to a previous dictionary entry, and character is appended to the string represented by dictionary[index]. For example, "abc" would be stored (in reverse order) as follows: dictionary[k] = {j, 'c'}, dictionary[j] = {i, 'b'} , dictionary[i] = {0, 'a'}, where an index of 0 implies the end of a string. The algorithm initializes last matching index = 0 and next available index = 1. For each character of the input stream, the dictionary is searched for a match: {last matching index, character}. If a match is found, then last matching index is set to the index of the matching entry, and nothing is output. If a match is not found, then a new dictionary entry is created: dictionary[next available index] = {last matching index, character}, and the algorithm outputs last matching index, followed by character, then resets last matching index = 0 and increments next available index. Once the dictionary is full, no more entries are added. When the end in th of the input stream is reached, the algorithm outputs last matching index. It is very important to know that the strings stored in the dictionary is in the reversed order[11-13]. LZW is an LZ78 based algorithm that uses a dictionary pre-initialized with all possible symbols. The main improvement of LZW is that when a match is not found, the current input stream character is assumed that it will be the first character of an existing string in the dictionary (since the dictionary is initialized with all possible characters), so only the last matching index is output (which may be the pre-initiali

In our research work, we have proposed a novel model that combines compression algorithm and crypto mechanism. After analyzing the available compression algorithms like Arithmetic coding, Huffman coding, LZ78 and LZW, we have suggested a new compression algorithm, named as Parallelized Sparse Data Compression Algorithm (PSDCA). It offers significant advantages in saving more storage space by avoiding the wastage of storage locations by eliminating the memory entries with values of “0”. Here, as a unique provision, if "1" exceeds the threshold limit, then "0" will be stored. Measuring Compression Performances Typically, the performance measure proves the efficiency of a compression technique or otherwise i.e. whether it is compatible with any particular criteria. Based on the nature of the application, the criteria is selected to measure the performance of compression algorithms. In fact, time complexity and space complexity are usually regarded as the most significant criteria, used interchangeably. The compression behavior largely is determined by the category of compression algorithm chosen, whether lossy or lossless. A major advantage of the proposed framework is that it has the potential to reduce the accessing time between cloud clients via cloud servers, apart from less cloud storage on account of wastage by utilizing “NULL” values. However, a likely drawback could be seen in the unusually longer execution time while matching the missing storage values in the reproduction section due to slow connections in the communication channel. The comparative performance analysis is factored on the cloud storage efficiency, quality of data in retrieval efficiency and time of execution efficiency. The following chart depicts the comparative analysis between various compression algorithms such as Arithmetic Coding, Huffman Coding, LZ78, LZW and PSDCA

III. PROPOSED WORK

The data for the transmission is used for the encryption by using Secured Adaptive Block Based Data Encryption is

combined with data compression algorithm before it will store into cloud storage. In Adaptive Block Based Data Encryption each block of data is encrypted using a random number. i.e use only one random number at beginning of encryption and that can used recursively to all the data blocks using one-way hash function. This random number must be shared securely between the participants before the encryption by a one pass method. In order to eliminate the threat for data loss is managed by the implementation of an exact link between the segmented storage element in the cloud server as well as the service providers. In most of the cases, the data retrieval or decoding information face a problem at the situation of inactive intermediate cloud servers. Such demerit or drawback is handled with the help of multiple copy of (replication) compressed as well as encrypted data to make it available in redundancy mechanism.

The cloud data which one is ready for transmission is encrypted along with compression in default manner and it will transfer towards the cloud server storage location. The number of key attributes and the file sizes are the major factors in encryption and decryption process. The time requirement is more for large file size and even number of attributes increased. Let us discuss the performance analysis of time, security performance and speed for each method

Table 2: Security performance analysis for maximum number of attributes

No. of Attributes	Security Performance (%)				
	KP-ABE	CP-ABE	CP-ABE-WP	EPPDR	SABBDE
2	59.27	58.97	58.50	62.99	68.99
4	66.75	66.52	67.00	63.74	69.74
6	70.90	73.17	73.25	69.92	78.92
8	78.44	76.62	78.45	76.53	82.53
10	75.19	76.91	76.97	79.14	87.14
12	83.93	83.19	83.41	83.42	92.42
14	83.96	83.26	83.18	85.91	92.91

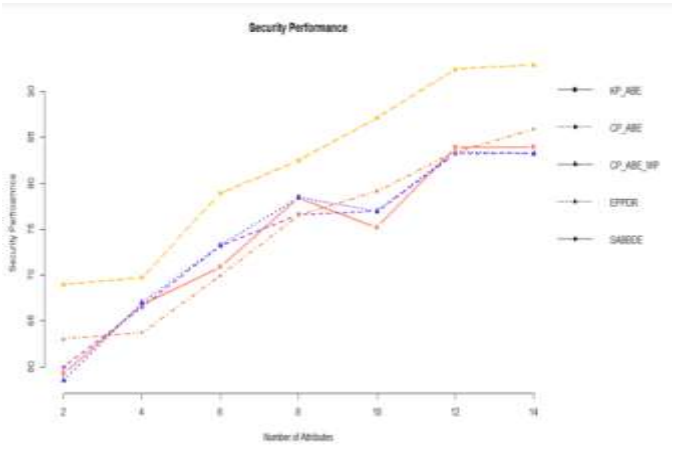

Figure 6: Analysis of different security algorithms

Table 3: Encryption time of different files sizes and different algorithms

File Size	Encryption Time (ms)				
	KP-ABE	CP-ABE	CP-ABE-WP	EPPDR	SABBDE
10MB	93.67	92.45	90.23	89.32	65.89
100MB	189.23	183.38	178.83	172.35	142.44
1GB	484.26	436.45	418.28	389.24	338.49
2GB	784.44	764.87	755.88	684.43	656.29
3GB	1024.56	1113.34	1023.56	968.75	924.56
5GB	2346.58	2289.52	2212.87	1759.35	1649.70

.

IV. RESULTS AND DISCUSSIONS

In most cases, the performance of security algorithms are assessed with the help of its run-time efficiency and it will not focus on its storage. In this implementation, it gives an equal heaviness for both run time as well as storage efficiency for the cloud data residing in the cloud data server. The above graph (figure 6) clearly depicted the detailed performance evaluation

V. CONCLUSION

Security is mostly concerned in each and every data transactions over the internet among the different users. In the same aspect, the researchers are struggling to propose a standard architectural framework in order to save the cloud data residing in cloud service provider with different mechanism. In this research paper provide an excellent roadmap to ensure the cloud data security with the help of compression and encryption mechanism.

References

[1]. P. Nirmala, T. Manimegalai, J. R. Arunkumar, S. Vimala, G. Vinoth Rajkumar, Raja Raju, "A Mechanism for Detecting the Intruder in the Network through a Stacking Dilated CNN Model", Wireless Communications and Mobile Computing, vol. 2022, Article ID 1955009, 13 pages, 2022. https://doi.org/10.1155/2022/1955009.

[2]. J.R.Arunkumar, Dr.E.Muthukumar,"A Novel Method to Improve AODV Protocol for WSN" Journal of Engineering Sciences" Volume 3, Issue 1, Jul 2012. ISSN NO: 0377-9254

[3]. J. R. Arunkumar, S. Velmurugan, B. Chinnaiah, G. Charulatha, M. Ramkumar Prabhu et al., "Logistic regression with elliptical curve cryptography to establish secure iot," Computer Systems Science and Engineering, vol. 45, no.3, pp. 2635–2645, 2023.

[4]. P. K. Devi, D. Arulanantham, C. Kalaivanan, N. Gomathi, J. R. Arunkumar and G. Ramkumar, "An Secure and Low Energy Consumption based Intelligent Street Light Managing System using LoRa Network," 2022 6th International Conference on Electronics, Communication and Aerospace Technology, Coimbatore, India, 2022, pp. 638-645, doi: 10.1109/ICECA55336.2022.10009408.

[5]. Prathima Chilukuri , J.R. Arun Kumar , R. Anusuya , M. Ramkumar Prabhu. "Auto Encoders and Decoders Techniques

of Convolutional Neural Network Approach for Image Denoising In Deep Learning" Journal of Pharmaceutical Negative Results, 13(4), 1036–1040. https://doi.org/10.47750/pnr.2022.13.04.142 ,November 4, 2022.

[6]. R. Yugha, V. Vinodhini, J. R. Arunkumar, K. Varalakshmi, G. Karthikeyan and G. Ramkumar, "An Automated Glaucoma Detection from Fundus Images based on Deep Learning Network," 2022 Sixth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Dharan, Nepal, 2022, pp. 757-763, doi: 10.1109/I-SMAC55078.2022.9987254.

[7]. E. Thenmozhi, A. Karunakaran, J. R. Arunkumar, V. Chinnammal, C. Kalaivanan and G. Anitha, "An Efficient Object Detection and Classification from Restored Thermal Images based on Mask RCNN," 2022 Sixth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Dharan, Nepal, 2022, pp. 639-645, doi: 10.1109/I-SMAC55078.2022.9987422.

[8]. L. Saravanan, W. Nancy, K. P. Chandran, D. Vijayanandh, J. R. Arunkumar and R. T. Prabhu, "A Novel Approach for a Smart Early Flood Detection and Awareness System using IoT," 2022 8th International Conference on Smart Structures and Systems (ICSSS), Chennai, India, 2022, pp. 1-4, doi: 10.1109/ICSSS54381.2022.9782286.

[9]. S. Bharathi, A. Balaji, D. Irene. J, C. Kalaivanan and R. Anusuya, "An Efficient Liver Disease Prediction based on Deep Convolutional Neural Network using Biopsy Images," 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2022, pp. 1141-1147, doi: 10.1109/ICOSEC54921.2022.9951870.

[10]. Prathima, C. H., Anusuya, R., & Prabhu, M. R. K. (2022). Comprehensive Design Analysis of Digital Marketing in Agriculture Sector. International Journal of Early Childhood Special Education, 14(5), 2022.

[11]. Atul Kumar Dwivedi, Deepali Virmani, Anusuya Ramasamy,Purnendu Bikash Acharjee, Mohit Tiwari" Modelling And Analysis Of Artificial Intelligence Approaches In Enhancing The Speech Recognition For Effective Multi-Functional Machine Learning Platform – A Multi Regression Modelling Approach " Journal of Engineering Research - ICMET Special Issue, 2022-04-06.

[12]. M.Ramkumar Prabhu, A.Rajalingam, J.R.Arunkumar, Dr.R.Anusuya" Microstrip Patch Antenna Using Combined Slots for Bandwidth Enhancement and Size", Journal of Engineering Sciences, Vol 11, Issue 1, Jan, 2020, ISSN NO: 0377-9254.

[13]. Anusuya Ramasamy, Abel Adane Changare" Hybrid Fuzzy Knowledge Based Prediction Model for the Software Development and Maintenance Quality in Software Engineering Approach" International Journal of Innovative Technology and Exploring Engineering (IJITEE), ISSN: 2278-3075, Volume-9 Issue-10, August 2020.

[14]. J.R.Arunkumar," Chaotic African Buffalo Optimization Based Efficient Key Mechanism in Categorized Sensor Networks:, International Journal of Engineering and Advanced Technology (IJEAT), ISSN: 2249 – 8958, Volume-9 Issue-3, February, 2020.

[15]. R. Anusuya, M. Ramkumar Prabhu, Ch. Prathima, J. R. Arun Kumar" Detection of TCP, UDP and ICMP DDOS attacks in SDN Using Machine Learning approach" Journal of Survey in Fisheries Sciences, Vol. 10 No. 4S (2023): Special Issue 4.

[16]. M.Ramkumar Prabhu, A.Rajalingam, J.R.Arunkumar, R.Anusuya, "Microstrip Patch Antenna Using Combined Slots For Bandwidth Enhancement And Size" Journal of Engineering Sciences (JES),Vol 11, Issue 1, Jan / 2020, ISSN NO: 0377-9254.

[17]. M.Ramkumar Prabhu, J.R.Arunkumar, A.Rajalingam, R.Anusuya "A Modified Square Patch Antenna with Rhombus slot for High bandwidth" International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-9, July 2019.

[18]. Revanesh, M., Gundal, S.S., Arunkumar, J.R.Arunkumar et al. Artificial neural networks-based improved Levenberg–Marquardt neural network for energy efficiency and anomaly detection in WSN. Wireless Netw (2023). https://doi.org/10.1007/s11276-023-03297-6.

[19]. I. Chandra, K. V. Karthikeyan, R. V, S. K, M. Tamilselvi and J. R. Arunkumar, "A Robust and Efficient Computational Offloading and Task Scheduling Model in Mobile Cloud Computing," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1-8, doi: 10.1109/ICECONF57129.2023.10084293.

[20]. Jangam Raghunath, S Kiran, G Siva Nageswara Rao, JR Arun Kumar, R Anasuya, C Siva Kumar," A MACHINE LEARNING TECHNIQUE TO DETECT BEHAVIOR BASED MALWARE", Semiconductor Optoelectronics, Vol. 42 No. 1 (2023), 1268-1278

[21]. Dr.J.R.Arunkumar. "Enhanced Dynamic Authorized Secured Protocol for Wireless Sensor Networks, "Journal of Science, Computing and Engineering Research, 1(1), 07-11, Mar-Apr 2020.

[22]. Anitha Gopalan,O. Vignesh,R. Anusuya,K. P. Senthilkumar,V. S. Nishok,T. Helan Vidhya, Florin Wilfred," Reconstructing the Photoacoustic Image with High Quality using the Deep Neural Network Model", Contrast Media & Molecular Imaging, Volume 2023 | Article ID 1172473 | https://doi.org/10.1155/2023/1172473.

[23]. R. Anusuya, N. Anusha, V. Sujatha, R. Radhika and S. Iniyan, "Machine Learning based Landslide Detection System," 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2023, pp. 319-323, doi: 10.1109/ICCMC56507.2023.10084226.

[24]. S. Sivakumar, R. Anusuya, V. Nagaraju, L. P. Narendruni and R. Thamizhamuthu, "QoS Based Efficient Link and Consistent Routing in Wireless Sensor Network," 2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE), Bengaluru, India, 2023, pp. 1241-1246, doi: 10.1109/IITCEE57236.2023.10091080.

[25]. I. Chandra, G. Sowmiya, G. Charulatha, S. D, S. Gomathi and R. Anusuya, "An efficient Intelligent Systems for Low-Power Consumption Zigbee-Based Wearable Device for Voice Data Transmission," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1-7, doi: 10.1109/ICECONF57129.2023.10083856.

[26]. R. Meena, T. Kavitha, A. K. S, D. M. Mathew, R. Anusuya and G. Karthik, "Extracting Behavioral Characteristics of College Students Using Data Mining on Big Data," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1-7, doi: 10.1109/ICECONF57129.2023.10084276.

[27]. G. Karthikeyan, D. T. G, R. Anusuya, K. K. G, J. T and R. T. Prabu, "Real-Time Sidewalk Crack Identification and Classification based on Convolutional Neural Network using Thermal Images," 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, 2022, pp. 1266-1274, doi: 10.1109/ICACRS55517.2022.10029202.

[28]. S. Bharathi, A. Balaji, D. Irene. J, C. Kalaivanan and R. Anusuya, "An Efficient Liver Disease Prediction based on Deep Convolutional Neural Network using Biopsy Images," 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2022, pp. 1141-1147, doi: 10.1109/ICOSEC54921.2022.9951870.

[29]. Dr. R.Anusuya. ―Stacking Dilated CNN Authorized Secured Protocol for IoT Security, ―Journal of Science, Computing and Engineering Research, 1(1), 01-07, May- June 2022.