



Dynamic Data Compression and Security Based Algorithm for Big Sensing Data

G. Muneeswari¹, M.Mahalakshmi², S.Lokeshwari³, R.Yashini⁴

¹Professor, ^{2,3,4}U.G.Student

Department of Computer Science and Engineering

Vel Tech Multitech Dr. Rangarajan Dr. Sakunthala Engineering College

*Corresponding Author Email: muneeswarig@veltechmultitech.org

Abstract

In this paper, we propose a data compression algorithm named data will be divided into chunks and similarity based compression for the efficient processing of sensing data on the cloud. In current technology, the big challenge lies in the efficient access to storage and evaluating sensing data as it is vital to consider the accuracy of the data. In the traditional compression algorithm, data is considered as a single unit but in our novel approach compression is applied over a partitioned data chunks. Once the sensing data is compressed, in order to regain the original data, we perform some kind of prediction method and restoration algorithm. MapReduce algorithm is incorporated in our approach for providing scalability over the network and similarity checking to verify the correctness of file. The results obtained shows that the technique predominantly increases the efficiency of the data compression with a very less percentage of data loss.

Keywords: Data similarity, sensing data, data modules, compression, mapreduce.

1. Introduction

Big data on cloud computing is evolved to provide a growing platform for large amount of data processing with its powerful computing capability, storage, scalability, resource reuse and low cost, and has attracted significant attention with big data. For example, in Amazon database on cloud data processing the actual data comes from the sensing area. Our paper is related to data mining and cloud computing where we could compress the big sensing data chunk and gain the space which increases the scalability on cloud. We use a novel compression technique over the partitioned data. Along with the Mapreduce algorithm, prediction and restoration functions improvement in scalability is provided through our project. The results obtained shows that the technique predominantly increases the efficiency of the data compression with a very less percentage of data loss.

The proposed data chunk similarity based compression approach is mainly used to compress the big sensing data based on chunk similarity and can predominantly increase the efficiency of the data compression with a very less percentage of data loss. Thus the performance evaluation of our project confirms that the proposed scheme is provably accurate and practically efficient.

In this paper, literature review has been discussed in section II, The working methodology is incorporated in section III and evaluation results are explained in section IV. Section V summarizes the results and concludes the paper.

2. Literature Survey

In the modern world, lower pieces of information is accompanied by the common transfer networks, huge amount of items of data obtained from the actual world should be processed. In cloud

computing, we expect large amount of invention in terms of analysis of large amount of data. Business and society are getting benefited out of the data. But the introduction of the two techniques, distributed data area and sophisticated data processing and workload allocation over the entire data processing, we need to have some modification over the existing scenario [1]. Most of the cloud service providers in the market directly performing big data reduction on cloud may arise serious security and privacy concerns, such as leakage of user privacy and business information, cyber attacks, and Inference threats [2]. Virtual Machines had been incorporated with a market based allocation of resources and a new architecture has been proposed. Typical forms of certain cloud representation platforms developed in industries and thus showing the difference between cloud computing workload and other web based workloads [4].

In the cloud storage area many community people can utilize real time resources on a public cloud [5]. A more complex survey of various approach and mechanism of deploying data representative applications to analyze the various design suitability to support certain classes of applications and the customers [6].

3. Chunk Generation and Similarity Based Compression

The server module is responsible for uploading and chunking the file. The user will upload the file and the server will split the file (chunk) using data chunks generation algorithm.



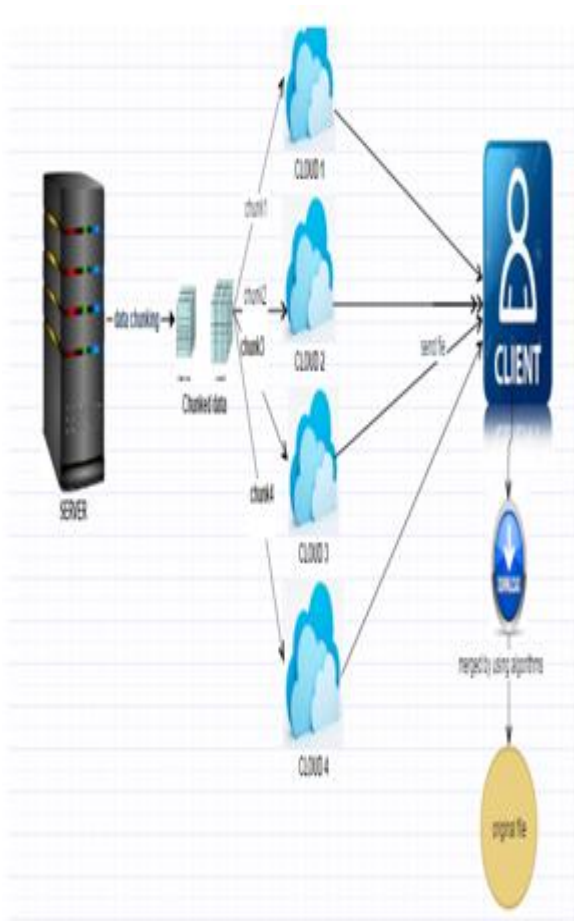


Fig. 1: Similarity based Compression and Map Reduction

The file converted as data chunks and send via server to various cloud nodes such as cloud 1,cloud 2 and cloud 3 by using data **Chunks Generation Algorithm**.After data chunking process completed we use a model for similarity that generates chunks of data. In this paper a novel idea of dynamic compression has been proposed.Instead of having single data compression part as in existing system, our idea uses complex partitioning of data during the entire process.There are many models available but the proposed work uses the model of geometric approach as it can measure the sensor data very accurately in the cloud environment.

4. Mapreduce and Similarity Checking

In MapReduce algorithm, the nearest node is found out and ultimately it reduces the data transmission. In the actual algorithm two cloud nodes are used: One is the master node and the other one is the worker node. The master node gets the file as the input, based on dynamic partitioning the smaller chunks are assigned to worker cloud nodes. All the low level nodes in turn partition into much more smaller levels thus forms a hierarchical structure. In the Reduce function actually the master gets all the information from the worker nodes and forms the final output.MapReduce and Compression Algorithm are used for compress the file at server and extract at client side.In this module, the available merged files are listed.

Finally, the client will choose the required file and download it. The file is received as packets (Data chunks) from cloud1,cloud2, cloud 3 ,cloud 4 and merged as one file at client. If the binary datas of the files are found hacked when we compare the files received at the client using the **similarity checking model** then the downloading of the files stops automatically intimating us that the files have been hacked.

5. Experimental Results

The overall simulation is carried over a public cloud. The sever contains information about all the cloud nodes. Fig.1. shows the cloud server.

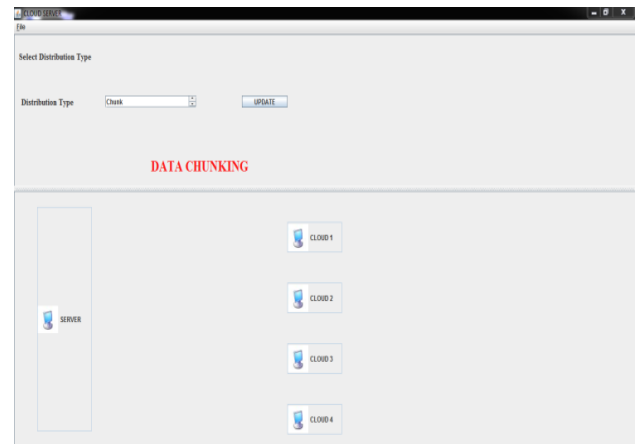


Fig. 2: Server Information

The public cloud clients for upload and download are shown in Fig.3.



Fig.3: Simulation of cloud for upload and download

The data chunking in cloud1 and cloud2 is shown in Fig.4.and Fig.5.

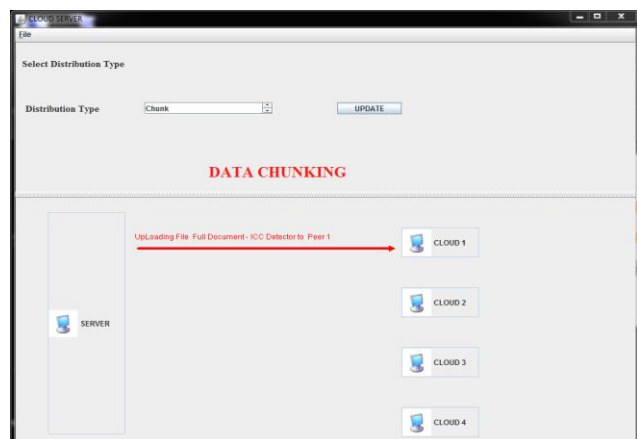


Fig.4: Data Chunking in cloud 1

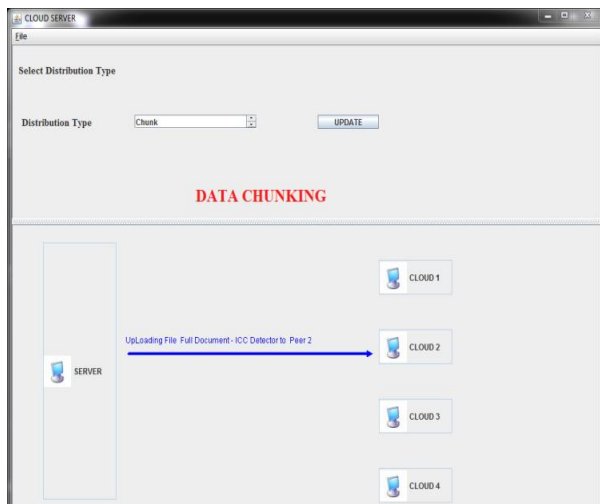


Fig.5: Data Chunking in cloud2

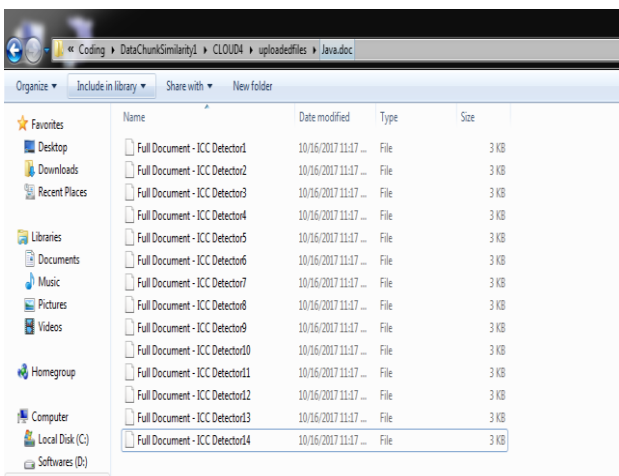


Fig.6: Chunk after upload

The file upload and download in cloud1 and cloud2 is shown in Fig.6.and Fig.7.

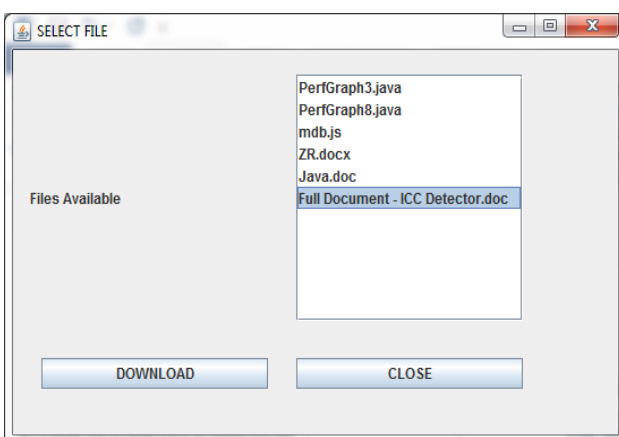


Fig.7: Client file download

The chunked file which is sent from the server, merged at the client side while downloading the file. MapReduce and Compression Algorithm are used for compress the file at server and extract at client side. In this module, the available merged files are listed. Finally, the client will choose the required file and download it. The file is received as packets (Data chunks) from cloud1,cloud2, cloud 3 ,cloud 4 and merged as one file at client. If the binary data of the files are found hacked when we compare the files received at the client using the **similarity checking model**

then the downloading of the files stops automatically intimating us that the files have been hacked.



Fig.8: File download details

6. Conclusion

In this paper, we have given a novel compression technique for partitioning data dynamically based on similarity checking model. This will be useful for data upload and data download in big data cloud storage which mainly incorporates sensing data. The proposed data chunk similarity based compression approach is mainly used to compress the big sensing data based on chunk similarity and can predominantly increases the efficiency of the data compression with a very less percentage of data loss. Thus the performance evaluation of our project confirms that the proposed scheme is provably accurate and practically efficient.

References

- [1] S. Tsuchiya, Y. Sakamoto, Y. Tsuchimoto, and V. Lee, "Big data processing in cloud environments," FUJITSU Sci. Technol. J., vol. 48, no. 2, pp. 159–168, 2012.
- [2] W. Dou, X. Zhang, J. Liu, and J. Chen, "HireSome-II: Towards privacy-aware cross-cloud service composition for big data applications," IEEE Trans. Parallel Distrib. Syst., vol. 26, no. 2, pp. 455–466, Feb. 2015.
- [3] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing," Commun. ACM, vol. 53, no. 4, pp. 50–58, 2010.
- [4] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, "Cloud computing and emerging it platforms: Vision, hype, and reality for delivering computing as the 5th utility," Future Gener. Comput.Syst., vol. 25, no. 6, pp. 599–616, 2009.
- [5] L. Wang, J. Zhan, W. Shi, and Y. Liang, "In cloud, can scientific communities benefit from the economies of scale?," IEEE Trans. Parallel Distrib. Syst. vol. 23, no. 2, pp. 296–303, Feb. 2012.
- [6] S. Sakr, A. Liu, D. Batista, and M. Alomari, "A survey of large scale data management approaches in cloud environments," IEEE Commun. Surveys Tuts., vol. 13, no. 3, pp. 311–336, Jul. - Sep. 2011.
- [7] T. Padmapriya and V.Saminadan, "Handoff Decision for Multi-user Multiclass Traffic in MIMO-LTE-A Networks", 2nd International Conference on Intelligent Computing, Communication & Convergence (ICCC-2016) – Elsevier - PROCEEDIA OF COMPUTER SCIENCE, vol. 92, pp: 410-417, August 2016
- [8] S.V.Manikanthan and D.Sugandhi "Interference Alignment Techniques For Mimo Multicell Based On Relay Interference Broadcast Channel" International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE) ISSN: 0976-1353 Volume- 7 .Issue 1 –MARCH 2014.
- [9] S.V. Manikanthan , T. Padmapriya "An enhanced distributed evolved node-b architecture in 5G tele-communications network" International Journal of Engineering & Technology (UAE), Vol 7 Issues No (2.8) (2018) 248-254.March2018