International Journal of Interdisciplinary Innovative Research &Development (IJIIRD)
ISSN: 2456-236X
Vol.04 Issue 01|2019

File Encryption Using Clustering Technique in Cloud Computing

Mst. Jahanara Akhtar[1], Md. Tahzib-Ul-Islam[2], Md. Nurul Islam Khan[3], Saiful Islam[4]

[1] *Associate Professor, Department of Computer Science & Engineering, Dhaka International University, Dhaka, Bangladesh.*
[2] *Assistant Professor, Department of Computer Science & Engineering, Dhaka International University, Dhaka, Bangladesh.*
[3] *Institute of Information and Communication Technology (IICT), Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh*
[4] *Department of Computer Science & Engineering, Dhaka University of Engineering and Technology (DUET), Gazipur, Bangladesh.*

ABSTRACT

Data security in web, cloud and computer system is very essential. Proper design of encryption and decryption process can improve the security feature of the system. In this work we have proposed a new idea of cryptography process. We have used clustering process for cryptography. We have selected BS, CHs and member files to create cluster then used standard encryption and decryption algorithm. We have analyzed theoretically and seen time consumption is very low as encryption process has been done in parallel. Time complexity is O(n).

Keyword - Cryptography, Clustering, BS, CH, Cloud Computing.

1. INTRODUCTION

Usually, to stock and calculate a vast quantity of data on cloud was hard because of inquiring, moving data etc. was inadequate to contract because they essential a software to track concurrently on lots of servers to become the wanted production. To become the production in fewer amount of time large data originated into being. Big-data is a term rummage-sale for big data sets that are inspected and totaled to stretch a productive outcome. Cloud is an intellectual term for internet that permits us to stock our data on the internet somewhat on our scheme. [8] In cloud, huge quantity of data is to be kept and in spread data hubs which brands data processing a multifaceted and a time intense process. The data and facilities that a operator needs to usage are kept at dissimilar physical parts. To deliver the data and facilities to the genuine operator is a challenge for the cloud service provider [5]. Since cloud computing suppleness and on request facilities, the growth of cloud is unavoidable. Hadoop is a service distributed all over the cloud and is used by the governments to supply cloud computing facilities. It supplies data by HDFS (Hadoop Distributed File System) which is founded on master-slave building. HDFS lets parallel computing through numerous nodes in a cluster. [10] K-means is one of the greatest basic algorithms used for unverified knowledge that is used to resolve the clustering problematic. Clustering is the process of segmenting a cluster of data points into a minor number of clusters. It is used to calculate distance in cloud to calculate distance between the centroid of a cluster and its data points. [17]

Hashing is a method to change a collection of data points into indexes. It supplies the data piece in such a way that assistances to localize it uniquely. A hash table is a group of matters which stores data in a method that will be informal to trace in upcoming. Encryption is a procedure to alter electric data into non readable method recognized as cipher text. Decryption is the conflicting procedure of encryption, it alters the cipher text into plain text that the end user can recite and comprehend well [3]. Symmetric key algorithms are cryptographic algorithms which usage similar secret key for both encryption and decryption. Symmetric key encryption can be used for any stream cipher or block cipher. [4] Stream cipher encrypts the digits of the message one at a time and block cipher takes a group of bits and encrypt it as a complete group.

2. LITERATURE REVIEW

Keji Hu and Wensheng Zhang [1], Escape of client's subtle data from cloud basis breaks to be an subject in spite of progression of cloud technology. One of the important details for such hovels to occur is the absenteeism of real encryption confirmation tactics. For a cloud computing framework, computational excellence of server necessity not be bore while encryption and confirmation is applied, which includes certain level of worry in structure up an well-organized confirmation procedure. In this paper an real activated

technique to authorize the info encryption on server side. This technique is greatest appropriate and earlier than the other encryption procedure. Additionally the competence, this plan can be applied to together the file info and changed info. The issue the paper focuses at is to give an arrangement that provide advantage on cloud server to save the information in encrypted way and clients can verify their data that is really presence in the form of encrypted at the servers storage. It is very difficult to keep the data secure if clients encrypt and decrypt the data by own and keep the keys. A cloud server has very effective processing power that each customer may long to use.

K.Brindha, N.Jeyanthi [2], Cloud computing perfect permits the customers to become the tenable data from cloud location deprived of the help of using the apparatus of scheme. For the productive use of the information from the cloud dealer, the individual who takes specialist on data, they encrypts the information and then deploy the information on cloud server. To safe the info in the cloud we have managed through the security issues prior.

B.Harikrishna, Dr.S.Kiran, R.Pradeep kumar Reddy [3] provided a review around the existing cryptography systems for subtle data on cloud. As we know that in today‟s age cloud is extremely required technology since it is similar an subcontracting of IT facilities and infrastructures. It assistances in easing the facilities and capitals which are mandatory at that case of time and is salaried for that only, that is cloud shadows the “pay on use” system. Occasionally cloud bombs to deliver the safety to the subtle information uploaded on the cloud, so to overawed this state the paper proposed an algorithm which is totally dissimilar from the extra current algorithms. Information is only said to be secured when it is intimate, combined and obtainable for the official and genuine user only. The paper deliberated about numerous cryptography algorithms in order to preserve the safety in the cloud. The paper also deliberated around Third Party Auditor (also known as TPA) which has privileges to square the truth of the files of the behalf of the operator and can announcement an assessment bang to the operator. The planned algorithm delivers the safety to the kept data and to the keys as well. Here is continuously a accidental for an insider assailant or an stranger assailant to spell the files on the cloud, so safety is actual vital and the keyless algorithms are showed to be healthier than key based algorithms since they do not go in contradiction of to the physiognomies or the topographies of the cloud, viz. verification, request admission, honesty facilities and so on.

Esteves, Rui Maximo, Rui Pais, and Chunming Rong [7], shows that clustering is a method to collection data substances on the foundation of coldness from additional data opinions. For scheming coldness amid data points we use Euclidean distance formula. K-means clustering is a extensively castoff idea in the world of clustering. This algorithm originates with a difficulty that it is not appropriate for big data as it does spring great presentation once functional to a minor data set and not to a big data set. Mahout is showed to be best algorithm because it is an cheap answer and also a talented one to the glitches created by big data, but there is a absence of study in this arena, so no one can potential that this exam will spring high recital. The Mahout scheme is motionless emerging and at the instant there is no talented outcome for clustering. On applying Mahout in cloud we saw that when the numbers of nodes are increased, the CPU usage falls and the network usage increases.

Authors of [19] shown tries to solve the problem of storing and managing big files over cloud by implementing hashing on Hadoop in big-data and ensure security while uploading and downloading files

3. CLUSTERING

Cluster analysis or clustering is the job of assemblage a set of matters in such a method that matters in the similar collection (named a cluster) stand extra alike (in certain logic) to each other than to those in other collections (clusters). It is a important work of searching data mining, and a joint technique for statistical data analysis, used in several arenas, with machine learning, pattern respect, image examination, information rescue, bioinformatics, data compression, and computer graphics [18].

Cluster study itself is not unique exact algorithm, then the overall job to be resolved. It can be realized by many algorithms that vary meaningfully in their considerate of what constitutes a cluster and how to professionally discovery them. General ideas of clusters include collections with minor distances among group memberships, thick parts of the data space, intervals or specific statistical distributions. Clustering can so be expressed as a multi-objective optimization problem. The suitable clustering algorithm and stricture locations (with limits such as the distance function to usage, a thickness verge or the amount of predictable groups) depend on the separate data set and envisioned usage of the results. Cluster study as such is not an involuntary job, then an iterative course of information finding or communicating multi-objective optimization that includes pilot and letdown. It is repeatedly required to adjust data preprocessing and perfect strictures till the effect attains the wanted goods.

Also the term clustering, here a amount of relations with alike senses, with involuntary ordering, arithmetical classification, bryology (from Greek βότρυς "grape"), typological study, and communal discovery.

The delicate changes are frequently in the usage of the results: though in data mining, the subsequent clusters are the stuff of attention, in involuntary organization the subsequent discriminative control is of interest

4. PROTOCOL DESCRIPTION

4.1 Over View

In file storage system of cloud computing, all files should be encrypted for security purpose. Cryptography can solve the security issue. Encryption can be done as like clustering system. At first, all files or chunks of data will be organized in some groups named clusters. Then one group leader or cluster head (CH) will be selected per cluster. Other files or chunks of data of the cluster will be considered as member of that CH. CH well perform encryption as algorithm to its members. All information of changing orientation of encryption of data will be stored in CH. One file or chunks of data will be act as base station (BS). BS will control all of CHs. All CHs files will be further encrypted and this information will be stored in BS file. Again, when decryption of files will be needed, then at first, BS will decrypt all CHs files according to standard decryption algorithm. Then CHs will perform process of decryption according to decryption algorithm. As all encryption or changing processes was stored in CHs files, so all member file will be retrieved according to decryption algorithm, so original files can be retrieved using this process.

4.2 Details

We will describe proposed protocol in step by step as below:

4.2.1 Creating Clustering: The process of creating clustering is shown in Fig. 1 . From this figure, it is shown that at first one BS file has been created. Then all other files has been organized into n groups. Then one CH has been selected form each group. All files other than CH of a group is considered as member file.

4.2.2 Encryption Process: The encryption process has been described in Fig. 2. From this figure, it is seen that, after creating a BS file, all other files has been organized into some group or cluster. Than on leader has been selected of each cluster to store encryption information of all member file. This file is known as cluster head (CH). Then all member files of each cluster has been encrypted according to a standard encryption algorithm. User can choice freely any suitable encryption algorithm. Then all encryption related information will be stored into corresponding CH file. After that, all CH files will be encrypted using that encryption algorithm and store that information into BS file.

4.2.3 Decryption Process: The decryption process has been described in Fig. 3. From this figure it is seen that, at first encryption related information of CHs files will be retrieved from BS file. According to that information, All CH file will be decrypted according to anti-encryption algorithm that was used to encrypt. After that, all encryption related information of member files will be retrieved from CH files. According to this information, all member files of each cluster will be decrypted.

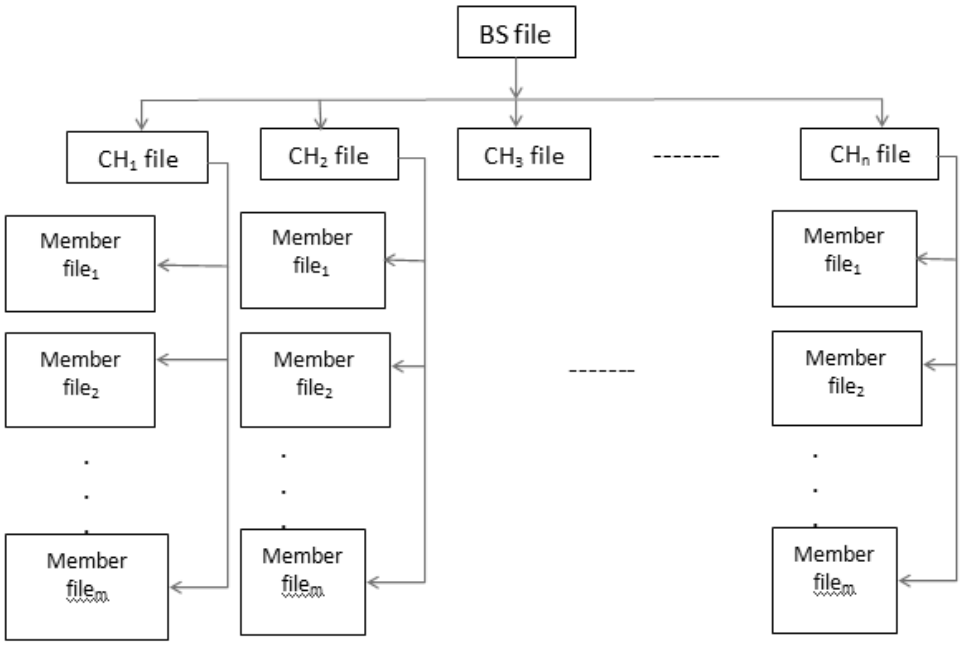

Figure 1. Clustering in cloud files

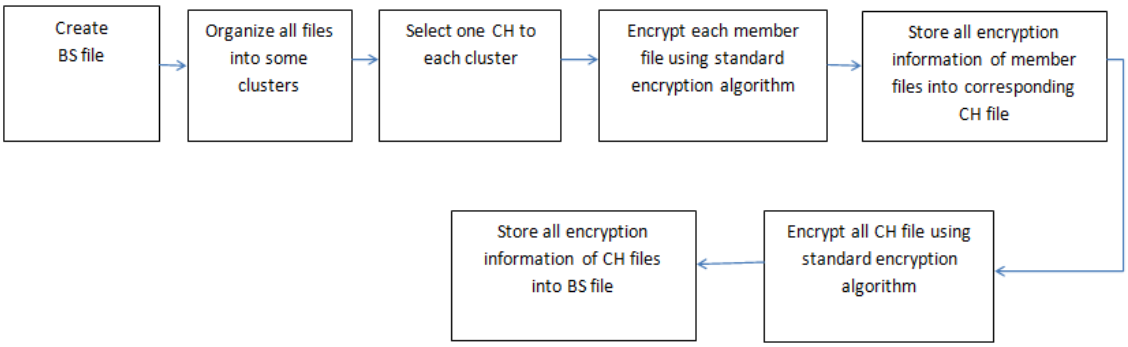

Figure 2. Block diagram of encryption process

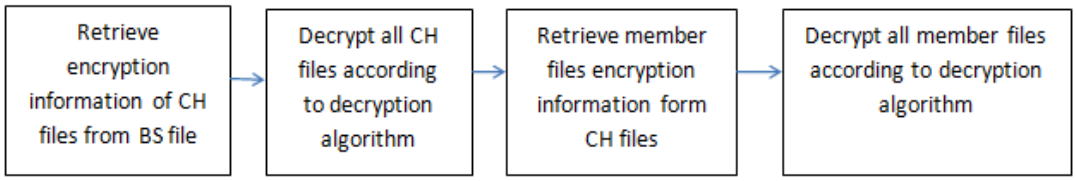

Figure 3. Block diagram of decryption process

5. ALGORITHMS

5.1 Encryption

Begin

Create a BS file.
Organize all files into K groups.

For each Group
Select a file as CH.
Consider all other files as member file.
Encrypt all member files using a standard encryption algorithm.
Store all encryption encryption related information to CH file.
End For

Encrypt all CH files and store information related information to BS file.
End.

5.2 Decryption

Begin

Retrieve encryption information of CH files form BS file.
Decryption all CH files using standard decryption algorithm

For each group of K
Retrieve encryption information of all member files from corresponding CH file.
Decrypt all member files using standard decryption algorithm.
End For

End.

6. ANALYSIS

As the best of our knowledge, this cryptography process is a new idea. We search in web related such type of process. But we get none. So we cannot compare our procedure to any other procedure. So we have analyzed only in theoretically. As clustering is a hierarchical process, so complexity of our procedure is very low. As encryption process of this technique will be done in every cluster parallel, so time consumption will be very low. We can see from encryption and decryption algorithm, we have used only one loop, so time complexity of our procedure is O(n)

6. CONCLUSIONS

We have tried to use clustering process in cryptography. We have used BS, CH and member concept in our procedure. This process is very simple and easy to understand and apply. Time consumption and time complexity O(n2) of the procedure is very low

7. REFERENCES

[1] Hu, Keji, and Wensheng Zhang 2014. "Efficient verification of data encryption on cloud servers." Privacy, Security and Trust (PST), 2014 Twelfth Annual International Conference on. IEEE.
[2] Brindha, K., and N. Jeyanthi 2015 "Securing cloud data using visual cryptography." Innovation Information in Computing Technologies (ICIICT), 2015 International Conference on. IEEE.
[3] Shynu, P. G., and K. John Singh.2016 "A Comprehensive Survey and Analysis on Access Control Schemes in Cloud Environment." Cybernetics and Information Technologies 16(1), pp.19-38.
[4] Jang, Miyoung, et al. "Clustering-Based Query Result Authentication for Encrypted Databases in Cloud,2014" High Performance Computing and Communications, 2014 IEEE 6th Intl Symp on Cyberspace Safety and Security, 2014 IEEE 11th Intl Conf on Embedded Software and Syst (HPCC, CSS, ICESS), 2014 IEEE Intl Conf on. IEEE.
[5] Maitri, Punam V., and Aruna Verma.2016 "Secure file storage in cloud computing using hybrid cryptography algorithm." Wireless Communications, Signal Processing and Networking (WiSPNET), International Conference on. IEEE.
[6] Zhu, Hongliang, et al.2016 "Based on the character of cloud storage string encryption and cipher text retrieval of string research." Cloud Computing and Intelligence Systems (CCIS), 2016 4th International Conference on. IEEE.
[7] Esteves, Rui Maximo, Rui Pais, and Chunming Rong,2011 "K-means clustering in the cloud—a Mahout test." Advanced Information Networking and Applications (WAINA), 2011 IEEE Workshops of International Conference on. IEEE.
[8] Kim, SungYe, et al. 2015 "Power efficient mapreduce workload acceleration using integratedgpu." Big Data Computing Service and Applications (BigDataService), 2015 IEEE First International Conference on. IEEE.
[9] Gugnani, Shashank, and Tamas Kiss.2015 "Extending Scientific Workflow Systems to Support MapReduce Based Applications in the Cloud." Science Gateways (IWSG), 2015 7th International Workshp on. IEEE.
[10] Saxena, Ankur, et al.2016"Implementation of cloud computing and big data with Java based web application." Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on. IEEE.
[11] Yetis, Yunus, et al.2016 "Application of Big Data Analytics via Cloud Computing." World Automation Congress (WAC), IEEE.
[12] Buyya, Rajkumar, et al.2015 "Big Data Analytics-Enhanced Cloud Computing: Challenges, Architectural Elements, and Future Directions." Parallel and Distributed Systems (ICPADS), 2015 IEEE 21st International Conference on. IEEE.
[13] Lu, Huang, Chen Hai-Shan, and Hu Ting-Ting 2012 "Research on hadoop cloud computing model and its applications." Networking and Distributed Computing (ICNDC), 2012 Third International Conference on. IEEE.
[14] Kanungo, Tapas, et al.2002 "An efficient k-means clustering algorithm: Analysis and implementation." IEEE transactions on pattern analysis and machine intelligence (24) 7,881-892.
[15] Adnan, Muhammad, et al.2014 "Minimizing big data problems using cloud computing based on Hadoop architecture." High-capacity Optical Networks and Emerging/Enabling Technologies (HONET), 2014 11th Annual. IEEE.
[16] Kuzu, Mehmet, Mohammad Saiful Islam, and Murat Kantarcioglu.2012 "Efficient similarity search over encrypted data." Data Engineering (ICDE), 2012 IEEE 28th International Conference on. IEEE.
[17] Iqjot singh, Prerna dwivedi, Taru gupta, Shynu P.G.2016 “An enhanced k-means clustering algorithm for big data in cloud.” International Journal of Pharmacy & Technology, (8)4.
[18] https://en.wikipedia.org/wiki/Cluster_analysis

[19] Iqjot singh, Prerna dwivedi, Taru gupta, Shynu P.G.2016 “Enhanced K-means clustering with encryption on cloud.” IOP Conf. Series: Materials Science and Engineering 263 (2017) 042057 doi:10.1088/1757-899X/263/4/042057

8. BIOGRAPHIES

	**Mst. Jahanara Akhtar** is now serving as an Associate Professor of the department of Computer Science and Engineering, Dhaka International University, Dhaka, Bangladesh. She is a Research Fellow (PhD) in the department of Computer Science and Engineering, Jahangirnagar University, Dhaka, Bangladesh. Jahanara completed B.Sc in Electronics & Computer Science and M.Sc. in Computer Science and Engineering from Jahangirnagar University. She has publications in national and international conference and journals. Her research interest includes Cryptography, Secure Wireless Sensor Network, Image Processing and Artificial Intelligence.
	**Md. Tahzib-Ul-Islam** is currently serving as an Assistant Professor of Computer Science and Engineering department, Dhaka International University, Dhaka, Bangladesh. He is a M.Sc. student in the Institute of Information and Technology, University of Dhaka, Dhaka, Bangladesh. Tahzib completed B.Sc in Computer Science and Engineering in Department of Computer Science and Engineering from University of Dhaka. He published research papers in national and international conference and journals. His research interest includes Cryptography, Network Security, Image Processing and Cloud Computing.
	**Saiful Islam** is currently pursuing his Ph.D. in Computer Science and Engineering degree in Dhaka University of Engineering and Technology (DUET), Gazipur, Bangladesh. Saiful has completed B.Sc. in CSE and M.Sc. in EEE from Dhaka University of Engineering and Technology (DUET), Gazipur, Bangladesh. He also has completed M.Sc. in ICT from Bangladesh University of Engineering and Technology (BUET), Bangladesh. He has publications in national and international conference and journals. His research interest includes Cryptography, Network Security, Wireless Sensor Networks, Cyber Physical System and Cloud Computing.
	**Md. Nurul Islam khan** is working in Health Service Division, Ministry of Health and Family Welfare as a programmer. He completed MSc. in ICT from Bangladesh University of Engineering and Technology. He completed BSc. in Computer Science from National University, Gazipur, Bangladesh. He has publication in international conference and journals. His research areas are Wireless Sensor Networks, Cognitive Radio Networks, Artificial Intelligence, Machine Learning, Cloud Computing.