International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

A Providing Security on Cloud Data Hadoop using Symmetric Encryption and Identity Authentication

Kale Rohitkumar Bhausaheb[1], R.L.Paikrao[2]

[1]Savitribai Phule Pune University, Amrutvahini College of Engineering, Sangamner, Maharashtra, India.

[2] Professor, Head of Department, Amrutvahini College of Engineering, Sangamner, Maharashtra, India.

Abstract*: Cloud computing is evolving as a much known data interactive paradigm to realize users data remotely stored in an online cloud server. - Security of distributed network is very important part in today's era. In order to solve the existing security problem of the distributed network cloud disk; such as transmission, storage security problems, etc., a network cloud disk safety storage scheme based on Hadoop is proposed. Based on the different confidential level of user data, it gives selective encryption scheme, which gives full consideration to the following security issues, such as the security of the user data transmission in the network, user data no verification, the user data privacy might be leaked, etc. Combined with the rapid encryption of symmetric encryption algorithm like Blowfish and identity authentication of RSA, overtime checking and the performance of Hadoop, the distributed network cloud data security storage disk can supply secured, effective, stable effect.*

Keywords: HDFS, Cloud Computing, Blowfish, AES, RSA 3way handshake

1. Introduction

The concept of Clouds is not new, it is undisputable that they have proven a major commercial success over recent years and the cloud computing is an extension of grid computing and distributed computing. It works through variety of technologies such as software technologies, integration, management, and the use of various hardware resources.

Cloud storage is an important part of cloud computing which is used to achieve the target of storing data in the cloud. The network cloud disk, which is popular in recent years, is the Hadoop. Hadoop is a Distributed framework for analyzing huge quantity of data or Big data. It is work on Hadoop Distributed File System (HDFS).But there is no attempt to verify the identity and group membership of users who interact with Hadoop Distributed File System (HDFS) and Logged users can store data in the Hadoop in a browser without any additional storage media, and the user can obtain the data wherever they are by ordinary computers, mobile phones, laptop, iPad, etc. But there is not having a effective security for the confidential data which is not access to unauthorized user. Also in Hadoop Kerberos authentication is used but The Kerberos authentication is not very effective to securing confidential data on Hadoop.

Cloud computing is an extension of grid computing and distributed computing, which is a software concept indeed [1], it works through variety of technologies such as software technologies, integration, management, and the use of various hardware resources. Cloud computing is realized mainly through the virtual technology. The virtual technology can be divided into single virtualization and multiple virtualizations; the single virtualization uses the virtual technology on a machine to work as many machines working together, such as VMware, while multiple machine virtualizations links the machines through the control center and makes them work like one machine. Hadoop is the representative of the relevant technology; obviously, its storage structure is distributed. The distributed storage system stores the data in different devices which are independent of each other. Cloud storage is an important part of cloud computing, which is used to achieve the target of storing data in the cloud. The network cloud disk, which is popular in recent years, is one of the ways to realize the target. Logged users can store data in the cloud in a browser without any additional storage ordinary computers, mobile phones, laptop, iPad, etc.But there is not having a effective security for the confidential data which is not access to unauthorized user. Also in hadoop Kerberos authentication is used but The Kerberos authentication is not very effective to securing confidential data on hadoop.There is main focus for providing security onto following aspect :

1) **Provide security over data transmission**: Data in transmission process may be intercepted, but the data transmission is not working with the strong encryption protection measures
2) **Provide security over Access control**: Access control authority is weak, the user data stored in the clouds without setting access authority, the user lost absolute right to monitor.
3) **Provide security over Data storage**: User upload data after the clouds, it is likely to be distributive stored, users do not know the specific position where the data is stored. And the confidential data and non-confidential data stored is not classified, which may cause the leakage of data.
4) **Provide security on Data verification**: The cloud makes no verification and inspection on the data uploaded. It can't guarantee that the uploaded data is corresponding to the right user's data or the original data from the user.

To solve the existing security problems, we propose a cloud disk storage based on Hadoop, the program draw lessons from Kerberos' authorization process. We utilize the classic algorithms such as AES, RSA, Blowfish to realize encryption, and authentication, and we also check the time to

inspect if it can complete the encryption and transmission in an acceptable period.

2. Literature Survey

[1] Hadoop is an Apache open source project which consists of HDFS,MapRedcue,HBase,Hive,ZooKeeper and other projects. [2] Its main parts are HDFS and Map Reduce. Map Reduce aims at paralleling and dealing with tasks on a large scale, which would make the Map Reduce scheduler become particularly important. It has a high fault tolerance and certain data access control. We mainly use Hadoop's HDFS (Hadoop Distributed File System). Kerberose: As [3]Hadoop also lacks safety measures, Kerberos was integrated into the Hadoop in 2009 by yahoo. The user have to obtain access certification from the third party center for key issues before access to Hadoop cluster first, and it greatly reduced the risk of users data leakages caused by identification. A lot of researchers proposed many different methods to increase the security of cloud storage. proposed the use of the HDFS to build a private enterprise cloud, which combine the Hadoops fault tolerance and suitable for big data attribute.

Attribute Encryption: In[4] SSL secure connection and secure virtual machine monitor are evaluated encrypted the cloud data using attribute encryption. Encrypted the cloud data using attribute encryption (ABE) scheme, using the property as a public key to encrypt the data before it is uploaded, this limits the data access user to have K attributes to decrypt the data, in which the K is the number of threshold to decrypt the data. [5] This scheme ensured the safety of data storage, and at the same time, the server has no need to keep a public key for each user, the users attributes are used to be the user's public key, but they could decrypt the data. completely when different users hold their attributes together and get all attributes[7],identifies the users with image processing methods, such as face recognition, fingerprint recognition, etc.[6]And made the transmission, encryption transmission and storage of the key files by Symmetric encryption and asymmetric encryption features complementary. [9] HDFS is an open source project of Google distributed file system(GFS).

But all these schemes mentioned just encrypt the data or identify the user from one perspective for a single demand, but no comprehensive data protection are taken.

3. Implementation Strategy

In our designed system contains mainly two parts. First Client module and Second is Server module. Client module contains Data Transmission Module ,Data Encryption Module. Where Server module contains Secret Key Production Module Data Decryption Module Data Authentication Module. Use of cloud storage for data encryption storage must shake hands three times to establish safety connection and storage between client and server cloud. The three times handshake generated between client and server is the main part of our designed system.

Type: 0 means: The uploaded file needs not to be Encrypted.

Type: 1 means: The uploaded file needs to be Encrypted.

3.1 3 Way Handshaking

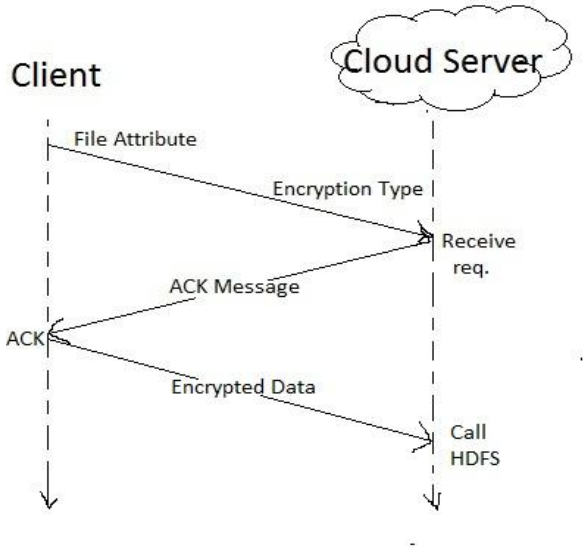

Figure 1: 3 Way Handshaking between Client and Server

3.1.1 First Handshake

1) Confirm the file that needs to be uploaded.
2) Choose the encryption Type.
3) Generate data structure according to the size of the file and the value of type. Type's value stands for the confidentiality level of the file. If the value of type is 0, the attribution of file is unsecured and the file can be uploaded to the cloud server directly.

{TYPE: 0 or 1, SIZE: file Size, FILENAME : filename };

4) Send the data generated in step3 to the server and wait for response.

3.1.2 Second Handshake

1) Receive request sent from the client.
2) Analyses the request and get the confidentiality level.
3) According to the confidentiality level call the secret key production and distribution module to generate symmetric key and signature key.
4) We have generated the signature key in step 3, in this step we will generate the data used for signature. Firstly, generate a random number 'rand' and get the current system time 'response_Current_Time', then we could use 'rand' plus 'response_Current_Time' to get the signature number 'Rand'. We also need a time factor 'T' which is used to check whether the encryption and transmission of the data is completed within a valid period. In our experiment ,The calculation of T is given below:

T=S*C+TS+D

Where S represents the size of the file, C represents the complexity of algorithm, TS represents the transmission time and D represents the acceptable delay time.

5) After step d, we could begin to form a data structure named ACK which is shown below.
{{Rand, symmetric key, signature key, filename} encrypted using user's master key};
6) Send ACK to client and wait for response.

3.1.3 Third Handshake

There are two main issues needed to be finished in this step. One issue is the encryption and signature of the data which will be finished at the client; another issue is receiving the data and inspecting its effectiveness which will be finished on the server. Obviously, the core of this thesis is the three times handshake, and this step turns out to be the core of the operation.

1) Receive request sent from the client.
2) Decrypt ACK using mast key and then get the symmetric key, signature key and 'Rand'.
3) We need to encrypt the data needed to be uploaded.
   Firstly, we would call the data encryption module to encrypt the uploading data using symmetric key, then data signature module also would be called to encrypt 'Rand' using signature key. Now it is time for forming the data structure which would be sent to the server. Details are as follows:
   {{user data} encrypt using symmetric key,{Rand} encrypt using signature key , filename: filename}
4) Send the data structure to client and wait for the response.
5) When the data is received by the server, it needs to be verified and validated before being stored in the HDFS. Data authentication module would be called to verify the signed data. Only the client has Signing Keys with no repudiation. If signature verification is not successful, as a result, the data may be distorted, and the package is discarded. However, if it is successful, the server can make sure that the data received is packaged by right user at a time. After signature verification, it also needs to check that the whole time, encryption time plus the transmission time, is carried out in a valid period. It's calculated as follows:

(System Time-(Rand-rand))-T

If the value of this expression is greater than 0, the entire process is completed within a suitable period, we believe that the security of the data is reliable.

6) If all the authentications are successful in step 5, we will believe that the entire encryption and transmission process is safe, and user's data could be stored in HDFS. Then, the data storage module will be called for storing user's data

4. Experiment

In our experiment performance optimization brought by the distributed encryption. Many cloud disk do not encrypt user data nor does it in the clouds, however, what we adopted is distribute encryption. Our experiment compares the distributed encryption with data encryption on the clouds, and the comparative parameter is encryption spent time. Our system provides 2 kinds of encryption for user data, Blowfish encryption and AES encryption. There are two reasons for providing the two encryption, one is the differences of encryption simulation study is carried out, degree and the other is the two encryption techniques are more mature and stable. So we designed two groups of experiment, compared the time consuming difference between distributed encryption and clouds encryption. The consumption includes the whole process spent time from data encryption to data transfer and till the end of data storage. Due to the limitation of experimental equipment and test site collecting the consumption time of 5 people,10 people,15 people,…,30 people upload data at the same time. We are also set the size of uploaded data. The results of the experiment are checked after completion of experiment.

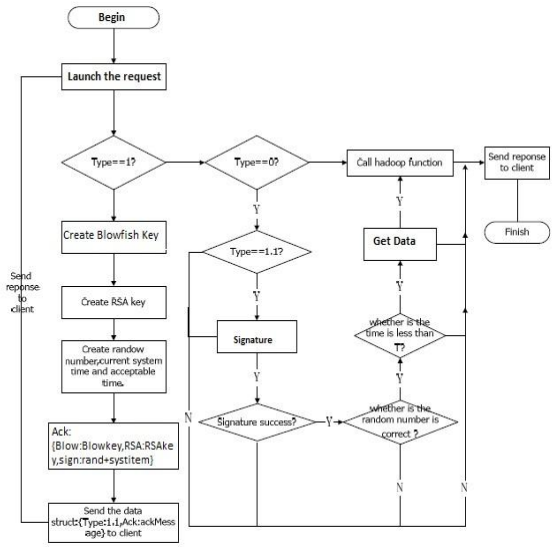

Figure 2: flowchart for Server

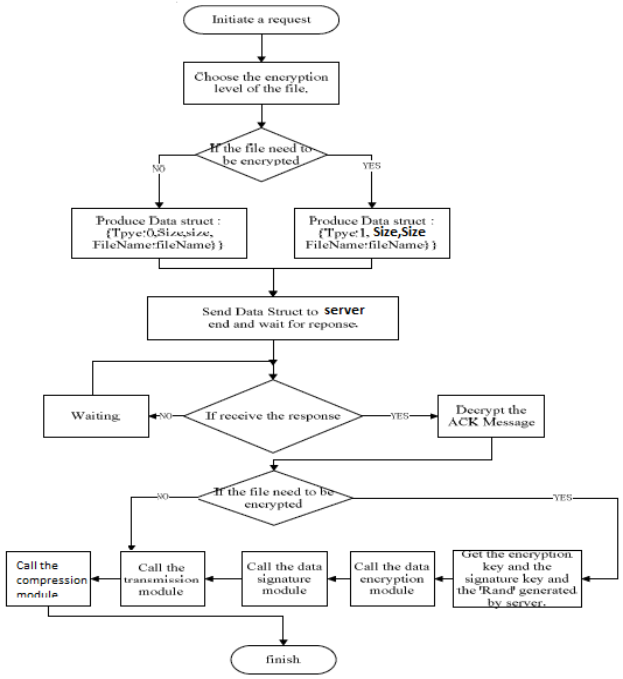

Figure 3: flowchart for Client

5. Conclusion

We put forward a security encryption schemes based on Hadoop, aiming at the existing popular cloud disc security weakness & which satisfy the data transmission and storage security and satisfy the server executes digital signature for client data at the same time. It is a distributed encryption system that could reduce the burden on the server, and finally achieve security, stability, efficient and effective storage. Our current system does not have enough sophistication. In the

future versions of our system, we plan to implement a more sophisticated technique for encryption and authentication.

References

[1] A. Huang Jing,B. LI Renfa,C. Tang Zhuo, The Research of the Data Security for Cloud Disk Based on the Hadoop Framework . in 4th ICICIP Beijing, China ,2013)
[2] HongBo Zhou. Cloud computing: technology, application, standard Electronic Industry Press.2011.
[3] Hou Qinhua,Wu Yongwei,Zheng Weimin "A Method on Protection of User Data Privacy in Cloud Storage Platform" .in Journal of Computer Research and Development, pages 1146-1154 2011.
[4] MUNIER M."Self-Protecting Documents for Cloud Storage Security[C]."Trust, Security and Privacy in Computing and Commu. Liverpool,2012: 1231-1238.
[5] SHAIKH F B." Security threats in cloud computing [C]".Internet Technology and Secured Transactions (ICIT. Abu Dhabi,2011: 214-219).
[6] V. Winkler, "Securing the Cloud Computer: Security Techniques and Tactics," Elsevier Inc., ISBN: 978-1-59749-592-9, 2011.
[7] ZHANG Da-wei. Research on hadoop-based enterprise file cloud storage system[C]".Awareness Science and Technology (iCAST), 2011 3rd. Dalian, 2011: 434-437
[8] BAO Rong-chang. Access Security on Cloud Computing Implemented in Hadoop System[C].Genetic and Evolutionary Computing (ICGEC),2011.
[9] LIU K.The Security Analysis on Otway-Rees Protocol Based on BANLogic[C].Computational and Infor - mation Sciences (ICCIS), Chongqing, 2012: 341-344
[10] http://en.wikipedia.org/wiki/Virtualization
[11] http://www.cse.wustl.edu/jain/cse567-6/ftp/encryption-perf/.

Author Profile

Kale Rohitkumar Bhausaheb received the B.E in InfoTech Engineering 2011 and M.E appeared degree in Computer Engineering from Amrutvahini College of Engineering, Sangamner 2015.