ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

A Survey on: Secure File Handling on Cloud based on Hadoop using HDFS

Rohini Tambe[1], Pansare Tejashri[2], Hadawale Megha [3], Bhor Pooja [4], Prof. Salunkhe T. R. [5]

Student, Department of Computer Engineering, SVCET, Rajuri, Maharashtra, India[1,2,3,4]

Assistant Professor, Department of Computer Engineering, SVCET, Rajuri, Maharashtra, India [5]

ABSTRACT:Hadoop is an Apache open-source framework for storing and processing large amount of data across clusters of computers. But processing sensitive or personal data in hadoop framework requires security model. As hadoop was designed without any security model, in this paper we present self-destructing data system that meets this challenge through a novel integration of secure cryptography techniques with active storage techniques based on hadoop. In this system, we present Shamir's secret sharing algorithm against sniffing attacks by using the public key cryptosystem to protect from sniffing operations.

KEYWORDS: Cloud computing;Hadoop; MapReduce; HDFS; Self Destructing Data; Active Storage.

I. INTRODUCTION

In today's digital world, with high technologyeveryone prefer to store their personal data on the Cloud which may has account numbers, passwords and other important information that could be used and misused by a miscreant, a intruder. This data is retrieved, copied and achieved by Cloud Service Providers (CSPs), often without users permission and control. These problem present challenge to protect people's privacy from illegal actions. By taking this problem into consideration, we present self-destructing system based on active storage framework to protect people privacy.

In proposed system, we present self-destructingdata system that meets this challenge through a novel integration of secure cryptography techniques with active storage techniques based on hadoop. We implement a proof-of-concept SeDas prototype. By using functionality and security properties the SeDas prototype can be evaluated. SeDas is practically easy to use and achieve all the privacy preserving aims described above.Compared to the system without self-destructing data mechanism, performance of uploading and downloading acceptably decreases, while latency for uploading and downloading operations with self-destructing data mechanism increases.

II. LITERATURE REVIEW

In Literature survey is the most important step in software development process.

Lingfang Zeng, proposed improved Washington's Vanish system for self-destructing data under cloud computing, and it is open to "hopping attack" and "sniffer attack". In this paper working of Safe Vanish, to prevent hopping attacks by way of Increasing the length of the key shares to rise the attack cost did some more enhancement on the Shamir Secret Sharing algorithm implemented in the Original Vanish system. They present an improved approach to prevent sniffing attacks by using the public key cryptography system to protect from sniffing operations. In addition, they evaluate analytically the functionality of the proposed Safe Vanish system.

Yu Zhang, Introduced that paper we present a reconfigurable calculating solution that can provide high-performance, flexible processing capabilities for the storage nodes. The dynamic reconfiguration upturns the functional density; however, the configuration self results in extra overhead, which may make the overall performance be downgraded. In the future works, we will implement multiple Processing Elements in the reconfigurable accelerator, and design an efficient method for dispatching the Processing Elements to hide the reconfiguration latency to improve the performance.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

FU Xiao, Realized emails were being watched by the government. For the advantage that Big Data technologies such as large distributed storage and user behaviour analysis and so on emails became one of the highly popular Big Data that has been targeted at as a large source of intelligence by some organizations keeps eye on public accounts every hour every day. research work was just opposite to what the NSA has did: To design and implement a system which can store emails securely, and terminate them clearly when they expired. In another word, a self-destructing emails system. But in this system there is no parallel processing for multiuser access.

ShaofengZou, developed a novel information theoretic approach is proposed to solve the problem of secret sharing, in which a distributor distributes one or more secrets to participants in such a manner that for each secret only qualified sets of users can reconstruct this secret by combining their shares together while nonqualified sets of users does not obtain any information about the secret even they pool their shares together. While existing secret sharing systems assume that communications between the distributor and participants are noiseless, this paper takes a more practical assumption that the distributor delivers shares to the participants via a noisy broadcast channel. Thus, in contrast to the available solutions that are mainly based on number theoretic tools, an information theoretic approach is introduced, which exploits the channel randomness during delivery of shares as additional resources to fulfil secret sharing requirements. Secret sharing problems can be reformulated as same secure communication problems via wiretap channels, and can hence be resolved by employing the powerful information notional security techniques.

In the existing system there are multiple disadvantages are available. In this Hacker can attack the confidential data and gain all the information from the database. This is big disadvantages of this system.Because client want to security of the data which is confidential from other's. In this hacking process the sensitive data can be modified by anyone, or if anyone can do changes in this client data.

III. PROBLEM STATEMENT

The major problem while using Cloud and mobile computing is security of personal data stored on the cloud and handling the multiple client node efficiently without affecting the speed of data transferring from server. In case of security concern there is preference to store personal data on the Cloud. That data may contain account numbers, passwords and other personal information. The personal information may get misused by intruder, dark side hackers, etc. While handling multiple client, the server may slows down and results into less throughput.

IV. PROPOSED WORK

Initially,the client has to register at metadata server.After registration, client has to perform login operation. For performing operations, valid user has to enter into database with session. At the metadata server, MapReduce framework accepts multiple client request to register them on server. In which, clients requests are divided by MapReduce to decrease the load of server. To check the validation of user, divided part of session key for each client will be forward to client as well as to the storage node. To validate user, there is need to conquer this parts of session keys at storage node and metadata server. If entered user is valid then metadata server provides access to the database for file operations such as uploading and downloading. As we using Shamir's algorithm, security is also increased.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

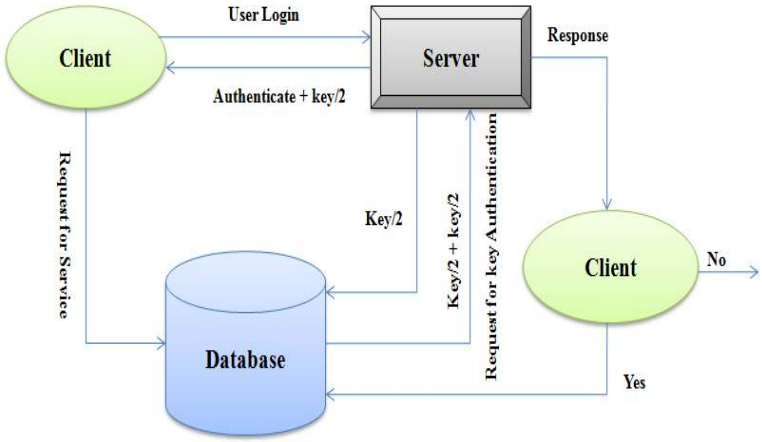

Fig.1.System_Architecture

A. ***MODULE:***

Our proposed system is divided into following modules:

1. Registration: In registration phase, we are taking the user details. If user was registered already by using attributes specified in the registration phase, then that user is discarded from registration. If user was not registered, then the user registration is processed and database is updated with generation of secrete key.

2. Login: In login phase, the user login details are taken from user. After taking user login details we are checking for user valid or not in our database.

3. Split: When user enters key, then this key is divided into 'n' shares from which one share is given to client and another is given to databases.

4. Encrypt: Before uploading file, it is convert from plaintext to cipher text using public key cryptography technique.

5. Upload: Here, user download the encrypted file.

6. Combine: As the key shares are distributed among all storage nodes and one share is distributed to client. To authenticate any file operation there is requirement to gather all required shares to reconstruct the key.

7. Decrypt: Before downloading file, it is convert from cipher text to plaintext using public key cryptography technique.

8. Download: Here, user download the decrypted file.

B. ***ADVANTAGES:***

- System can balances the load of server.
- System has ability to handle multiple clients.
- It also provides the security to sensitive data while transferring file.

V. MATHEMATICAL MODEL

System S:{ Q, ∑, O, δ, I, DD, NDD}

Where,

Q is set of states.

∑ is input given by user to select state(q)for its respective operation.

O is Output generated by state(Q).

δ is transition function which map Q*∑→ Q.

I is input value x given by user.

F is Functions to system.

DD is Deterministic data.

NDD is Non-deterministic data.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

S:Q(I)→ Q

I: Input value given by user.
Where,
Input={ Key | A - Z, a - z, 0 – 9}

Q :{ q_0,q_1,q_2,q_3,q_4,q_5}
Where,
q_0 = Initial state user registration.
q_1 = User Login.
q_2 = File uploading by taking encryption key from user.
q_3= Repeat step1.
q_4 = File downloading by taking decryption key.
q_5 = Exit.

∑: { Ek, eFL, ttl }
Where,
Ek ={ Ek | A - Z, a - z, 0 – 9}
eFL = { eFL | A - Z, a - z , 0 – 9 }
ttl ={ ttl | A-Z, a-z, 0 – 9 }
Ek ={ Ek | Sh1 + Sh2 + . . .+ Shn }
Sh1/Sh2/. . . /Shn = { Sh1/Sh2/ . . . /Shn | A - Z, a - z, 0 – 9 }

Where,
Ek is Encryption key,
eFL is Encrypted file,
ttl is Time-To-Live,
Sh1,Sh2,...Shn is Key shares.

O:{ dFL, Dk }
Where ,
dFL = { Df |Df(CT:Dk) }
Where,
dFL is decrypted file,
CT is cipher text,
Dk is Decryption key
Dk = {Dk | A - Z, a - z, 0 – 9}
Dk = {Dk | Sh1 + Sh2 + . . .+ Shn}
Shn is Share at threshold value.

δ: { Upload File(), Download File(), ttl() }

Upload File (f,s) =P :: takes the file and encryption key.
P = { h | h takes file, s | s takes encryption key }
Download File (d,k) = A:: outputs decrypted file and takes decryption key. A = d | d outputs decrypted file,k | k takes decryption key,
ttl (f,t) =B :: f takes encrypted file, t takes time to remove file.

DD: Set of Deterministic state data which gives required output.
NDD: Set of Non-deterministic state data which doesn't gives required output.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

VI. CONCLUSION AND FUTURE WORK

Hence, proposed system provides security to our personal data with help of hadoop. Hadoop has been efficient solution for companies dealing with the data in Peta bytes. According the above sections we can say that hadoop is one of the best ways to provide the security to sensitive data.

As a future work, we will extend new scheme to deal with other desirable features such as cloud based services and so on. We will also study the applicability of our newly proposed scheme to real world application.

ACKNOWLEDGEMENT

We express our sincere thanks to our project guide Prof. Salunkhe T. R. who always being with presence & constant, constructive criticism to made this paper. We would also like to thank all the staff of COMPUTER DEPARTMENT for their valuable guidance, suggestion and support through the project work, who has given co-operation for the project with personal attention. Above all we express our deepest gratitude to all of them for their kind-hearted support which helped us a lot during project work. At the last we thankful to our friends, colleagues for the inspirational help provided to us through a project work.

REFERENCES

1. Shaofeng Zou, Student Member IEEE, Yingbin Liang, "An Information Theoretic Approach to Secret Sharing", IEEE Transactions On Information Theory, VOL. 61, NO. 6, JUNE 2015.
2. FU Xiao, WANG Zhi-jian, WU Hao, YANG Jia-qi, WANG Zi-zhao, "How to send a Self-destructing Email", IEEE International Congress on Big Data, 978-1-4799-5057-7/14 2014.
3. Lingfang Zeng ,Shibin Chen , Qingsong Wei,"SeDas:A Self-Destructing Data System Based on Active Storage Framework", IEEE Transactions On Magnetics , VOL. 49, NO. 6, JUNE 2013.
4. L. Zeng, Z. Shi, S. Xu, and D. Feng, "Safevanish: An improved data self-destruction for protecting data privacy",IEEE,978-0-7695-4302-4/10,2010.
5. Yu ZHANG, Dan FENG, "An Active Storage System for High Performance Computing",IEEE 22nd International Conference on Advanced Information Networking and Applications,1550-445X/08, 2008.
6. Cong Wang,Qian Wang,Kui Ren, "Privacy- Preserving Public Auditing for Data Storage Security in Cloud Computing ",IEEE,978-1-4244-5837-0/10,2010.
7. Xukai Zou, Fabio Maino, Elisa Bertino,Yan Sui,Kai Wang and Feng Li,"A New Approach to Weighted Multi-Secret Sharing",IEEE, 978-1-4577-0638-7/11,2011.
8. Tina Miriam John, Anuradharthi Thiruvenkata Ramani,John A. Chandy,"Active Storage using Object-Based Devices",IEEE,978-1-4244-2640-9/08,2008.
9. Seung Woo Son, Samuel Lang, Philip Carns,Robert Ross, Rajeev Thakur,"Enabling Active Storage on Parallel I/O Software Stacks",IEEE ,978-1-4244-7153-9/10, 2010.
10. Mrudula Varade, Vimla Jethani, "Distributed Metadata Management Scheme in HDFS", International Journal of Scientific and Research Publications, VOL. 3, NO. 5, May 2013.
11. R.C.Dharmik,Hemlata Dakhore,Vaishali Jadhao,"Sedas: A Self Destructive Active Storage Framework for Data Privacy",International Journal of Scientific Engineering and Research,Volume 2,Issue 3,March 2014.
12. Sharifnawaj Y. Inamdar, Ajit H. Jadhav, Rohit B. Desai, "Data Security in Hadoop Distributed File System",International Research Journal of Engineering and Technology ,Volume 3,Issue 4,April 2016.