Preserving the privacy of healthcare data using cloud Computing

Pratiksha Bhalerao
Marathwada Mitra Mandal's College Of Engineering,
Karvenagar, Pune.
Email:pratikshabhalerao.it@mmcoe.edu.in

Priyanka Rathod
Marathwada Mitra Mandal's College Of Engineering,
Karvenagar, Pune.
Email:priyankarathod.it@mmcoe.edu.in

Akshata Patekar
Marathwada Mitra Mandal's College Of Engineering,
Karvenagar, Pune.
Email: akshatapatekar.it@mmcoe.edu.in

Rishabh Shukla
Marathwada Mitra Mandal's College Of Engineering,
Karvenagar, Pune.
Email:rishabhshukla.it@mmcoe.edu.in

Abstract - Now-a- days the medical organizations facing many challenges, cloud based electronic medical record services due to the cast of data contravention and the resulting deal of the patient data. Cloud computing technology is very popular on demand, where it can be used for storing purpose. There is some benefits of cloud computing which is cost, speed, performance, global scale, productivity and security. The main and important factor is the security of the cloud computing. If any user often to store sensitive medical information with cloud storage provider, but these providers are not trusted. The healthcare professional needs to access the patient electronic medical record (EMR), which contain big multimedia data for efficient access and support, EMR needs to be maintained the security of big data in the cloud. Now there is a requirement of developing a proper authorization rule of procedure for secure, safe, unbroken and convenient cloud based EHR managment.The data theft attack consider is the most serious contravention of the health care data security on the cloud. The main objective is protect private health care data in the cloud using fog computing. At the end agreement protocol was proposed which can be based on pairing based cryptography which can be use to generate a session key between user and admin to build communication secure with each other safely. Finally the private data can be accessed and stored securely with deduplication techniques.

Keywords-Medical Data, Fog Computing; Pairing-based cryptography, Security and Privacy.

I. INTRODUCTION

Cloud computing is an environment where the computing resources are outsource as a service through the internet [3]. Clouds typically involve service providers, resource providers, and service users. The cloud computing model allows access to data and computer resources from anywhere that a network connection are available. The cloud computing service models are Software as a Service, Platform as a Service and Infrastructure as a Service. In Software as a Service model, a pre-made application, along with any required software, operating system, hardware, and network are provided. In Platform as a service, an operating system, hardware, and network are provided, and the customer installs or develops its own software and applications [7]. Cloud computing specification include network-based access channel, resource pooling, multitenancy, automatic and elastic provisioning and metering of resource usage. Clients can use these resources to host applications and even client can use to store their data. Rapid changes of resources demand can help to deal with variable demand and proving optimum resource utilization. As more organizations are using cloud computing, cloud service providers are developing new technologies to progress the cloud capabilities. Now the term, mashups are a new trend. It means combining services from multiple clouds into a single service or application. This service composition makes cloud service providers to propose new functionalities to clients at lower development costs. Examples of cloud mashups and technologies are IBM's Mashups Center, Appirio Cloud Storage and Force.com for the Google App Engine. For example, cloudbased electronic medical record (EMR) management systems like Practice Fusion, Verizon Health Information Exchange, Medscribbler, and GE Healthcare Centricity Advance are emerging Cloud mashups want pre-established agreements among providers as well as the use of custom built, proprietary tools that combine services through low-level, tightly controlled and constraining integration techniques[6]. This approach to building new collaborative services does not support stability, flexibility, and openness. Realizing multicloud collaboration's full potential will require implicit, transparent, universal, and on-the-fly interaction involving different services reach across multiple clouds that lack pre-established agreements and proprietary collaboration tools. From a market perspective, it is doubtful that multiple CSP will agree on an easy and standardized way to access services, as this would give clients total freedom in changing providers, leading to increased open and direct competition with other providers [8].

Cloud-based computing also introduces new security concerns that affect collaboration across multi cloud applications they are, increase in the attack surface due to system complexity, loss of client control over resources, threats that target exposed interfaces due to data storage in public domains, and data privacy concerns due to multitenancy. So there is need of developing multi cloud system which provides trust, security and safety for applications and data. Keeping all these drawbacks our focus is to develop cloud collaboration allows clients and cloud applications to simultaneously use services from clouds [9]. There are restrictions in the current cloud computing model prevent direct collaboration among applications hosted by different clouds. A method that could remove these restrictions uses a network of proxies. A proxy is an edge-node-hosted software instance that a user or cloud service provider [CSP] can delegate to carry out operations on its behalf. Proxies can act as mediators for collaboration among services on different clouds. As an example of proxy-facilitated collaboration between clouds, consider a case in which a user or CSP wishes to simultaneously use a collection of services from multiple clouds. First, the requesting entity chooses proxies to act on its behalf and to interact with cloud applications. A user or CSP might employ multiple proxies to interact with multiple CSP.

The main objective of cloud computing is to support agility, flexibility and openness. Proprietary tools and services are integrated to control and monitor such a service. Different companies provide different services on cloud. Bringing these different services together and creating mashups require pre-established agreements among providers. Secure collaboration of such services is the need for tomorrow that motivated to take up this project [2].

An electronic health record (EHR) [1] is an electronic document that contains all relevant medical report details of person. application of cloud computing technology will be able to integrate high data for efficient data mining, which will greatly help to prevent chronic diseases make decision in diagnosis, treat diseases, and optimize drug inventory of a patient data in digital form, store and exchange securely and accessible by multiple authorize user it is important to note that proof of data integrity protocol checks the integrity of data that is if the data has been illegally modify or deleted.

Problem Statement:

In Cloud computing today, has played a crucial role in terms of data storage and reduction of overall costs for entrepreneurs. But most of them are worried about data security and storage management.

II. Related Work

M. Chen. et.al [1] streamline machine learning algorithms for effective prediction of chronic disease outbreak in disease-frequent communities. We experiment the modified prediction models over real-life hospital data collected from central China in 2013-2015. To overcome the difficulty of incomplete data, we use a latent factor model to reconstruct the missing data. We experiment on a regional chronic disease of cerebral infarction. We propose a new convolutional neural network based multimodal disease risk prediction (CNN-MDRP) algorithm using structured and unstructured data from hospital. To the best of our knowledge, none of the existing work focused on both data types in the area of medical big data analytics. Compared to several typical prediction algorithms, the prediction accuracy of our proposed algorithm reaches 94.8% with a convergence speed which is faster than that of the CNN-based unimodal disease risk prediction (CNN-UDRP) algorithm.

Hadeal Abdulaziz Al-Hamid et.al [2] gives main focus on, how to secure cloud data, particularly photos, by using fog computing facilities. Fog computing is an emerging model that provides storage, processing, and communication services closer to the end user. Also, fog computing is considered as a way to create decoy information and locate it beside the real information in the cloud to hide the true data of the user.

M. S. Hossain et.al [3] states that, the fast-growing healthcare big data plays an important role in healthcare service providing. Healthcare big data comprises data from different structured, semi-structured and unstructured sources. These data sources vary in terms of heterogeneity, volume, variety, velocity and value that traditional frameworks, algorithms, tools, and techniques are not fully capable of handling. Therefore, a framework is required that facilitates collection, extraction, storage, classification, processing, and modelling of this vast heterogeneous volume of data. The present paper proposes a healthcare big data framework using voice pathology assessment (VPA) as a case study. In the proposed VPA system, two robust features, MPEG-7 low-level audio and the interlaced derivative pattern (IDP), are used for processing the voice or speech signals. The machine learning algorithms in the form of a support vector machine (SVM), an extreme learning machine (ELM), and a Gaussian mixture model (GMM) are used as the classifier. In the experiments, the proposed VPA system shows its efficiency in terms of accuracy and time requirement.

M. Sriram et.al [4] state that, Cloud computing wherein third parties provide storage services. Insider attacks still continue to haunt cloud users as they tend to cause unprecedented damage, especially when privileged users who have access to sensitive information go rogue. Many proposals have been made to secure the Cloud from Insider threat Attacks and most of the standard approaches have been proven to fail from time to time. An implementation of a Hybrid protocol that uses Selective Encryption with data cleaning, Enhanced Neural Network based user profiling and decoy technology to combat the insider threat has been proposed. The proposed system gave unprecedented level of security.

Qinlong Huang et.al [5] propose this healthcare management application to an integrated algorithm technique for the jointly involved parties which is included privacy and security algorithm; it can protect data integrity also. This system provides the better efficiency and it is very useful to remote areas where hospitals are not easily accessible. This research to resolve wireless sensor network security relevant issues Using several patterns can reduce database load, and users to access data efficiently; the privacy control mechanism allow users to store data securely. The results of this research exhibit that the proposed system has better secured database access and maintain privacy for all patient data than the traditional database.

III. OPEN ISSUES

Lot of work has been done in this field because of its extensive usage and applications. In this section, some of the approaches which have been implemented to achieve the same purpose are mentioned. These works are majorly differentiated by the algorithm for Security.

As our point of view when we studied the papers, the issues are related to information Security. The challenge is to addressing medical data loss problem from healthcare community.

IV. PROPOSED APPROACHES:

As we studied about privacy of the healthcare data, so we want to propose Privacy of Medical Big Data in a Healthcare Cloud Using a Fog Computing Facility.

Hadeal Abdulaziz Al Hamid. Et.al propose a confidence scheme in the challenge of data ownership and cryptography to manage the storage of encrypted data with deduplication. Our goal is to solve the problem of deduplication in the situation where the data owner is not available. Meanwhile, the data size does not get affected by the performance of data deduplication in our schema. Hadeal Abdulaziz Al Hamid et.al has motivate, to save space in the cloud and to preserve the privacy of data owners by proposing a scheme to manage the storage of encrypted data with deduplication. Hadeal Abdulaziz Al Hamid et.al, test the safety and evaluate the performance of the proposed scheme through analysis and simulation. The result shows its efficiency, effectiveness and applicability.

Advantages:

- Increase the storage utilization using deduplication.
- Improve the reliability.
- Improve the security.

System Diagram:

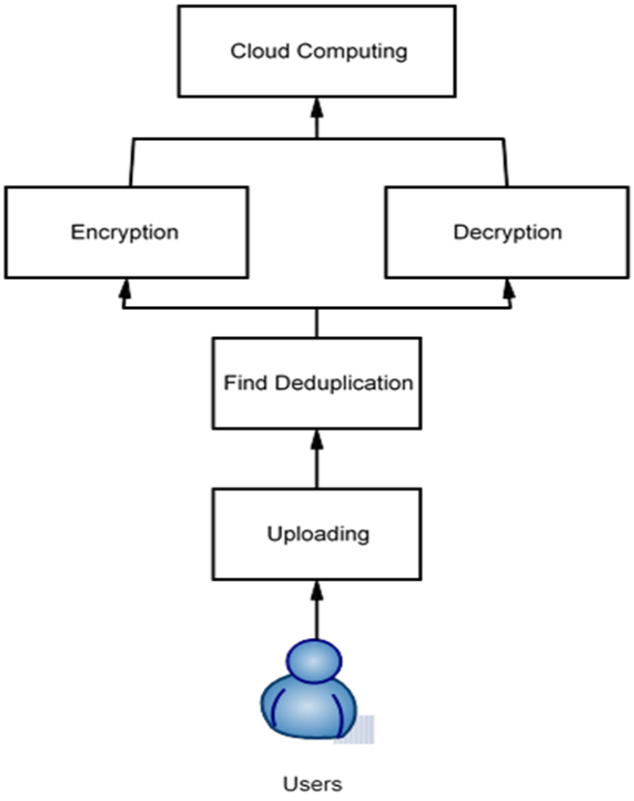


Fig. System Architecture

A. Algorithms

1. Paired Based cryptography for Encryption.

It used to convert plain text into cipher text .The need for coming with this algorithm is weakness in existing algorithm. The 56 bit key of des is no longer safe against attacks based on exhaustive key searches and 64-bit block also consider as weak. AES was to be used128-bit block with128-bit keys.

Input:

128_bit /192 bit/256 bit input (0, 1)

Secret key (128_bit) +plain text (128_bit).

Process:

10/12/14-rounds for-128_bit /192 bit/256 bit input

Xor state block (i/p)

Final round:10,12,14

Each round consists: sub byte, shift byte, mix columns, add round key.

Output:

cipher text(128 bit)

2. FRAGMENTATION ALGORITHM

Input: File

Output: Chunks

Step1: If file is to be split go to step 2 else merge the fragments of the file and go to step

Step2: Input source path, destination path

Step3: Size = size of source file

Step4: Fs = Fragment Size

Step5: NoF = number of fragments

Step6: Fs = Size/Nof

Step7: We get fragments with merge option

Step8: End

3. MD5 (Message-Digest Algorithm)

The MD5 message-digest algorithm is a widely used cryptographic hash function producing a 128-bit (16-byte) hash value, typically expressed in text format as a 32 digit hexadecimal number. MD5 has been utilized in a wide variety of cryptographic applications, and is also commonly used to verify data integrity.

Steps:

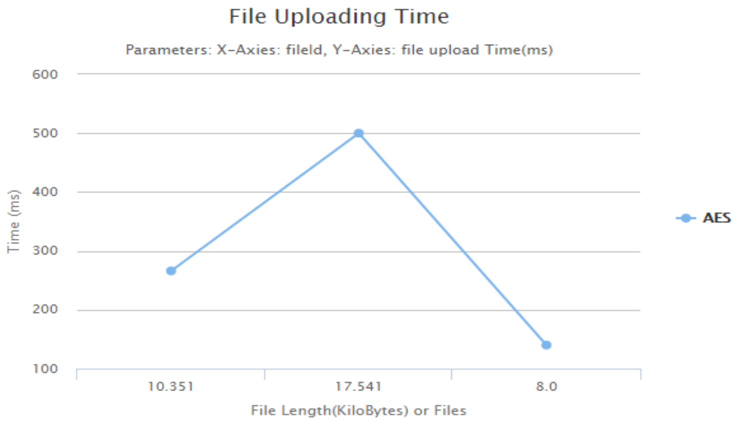
1. A message digest algorithm is a hash function that takes a bit sequence of any length and produces a bit sequence of a fixed small length.
2. The output of a message digest is considered as a digital signature of the input data.
3. MD5 is a message digest algorithm producing 128 bits of data.
4. It uses constants derived to trigonometric Sine function.
5. It loops through the original message in blocks of 512 bits, with 4 rounds of operations for each block, and 16 operations in each round.
6. Most modern programming languages provides MD5 algorithm as built-in functions

In our experimental setup, in table 1, find out different file upload and time required for time for uploading that file. In our experimental setup, in our system first is uploading file size and time for that file.

Sr.No	File Size(Kb)	Time(ms)
1	10351	226
2	17541	500
3	8500	140

Table1: File Uploading Time and Size

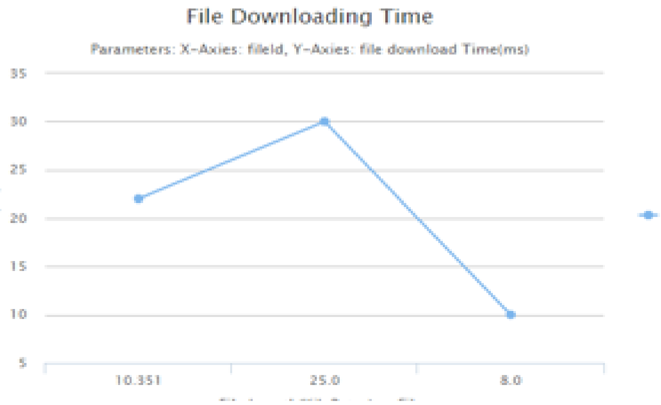
From above data, in graph 1, we can see file size of 1 is 10351 kb is required time uploading is 226 ms, and file size of 2 is 1751 kb is required time uploading is 500ms



In our experimental setup, in table 2, find out different file download and time required for time for uploading that file. In our experimental setup, in our system first is uploading file size and time for that file and so on.

Sr.No	File Size(Kb)	Time(Sec)
1	10351	22
2	25000	30
3	8000	10

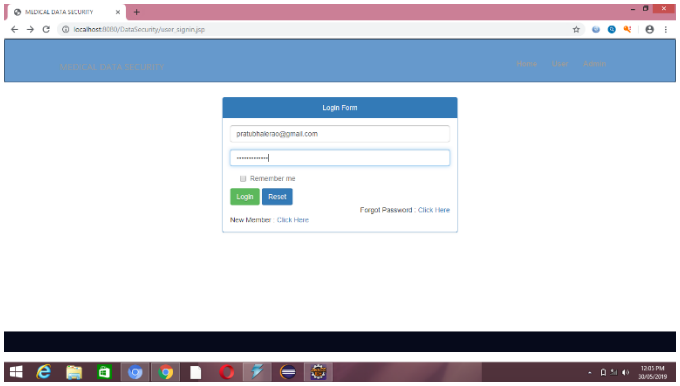
Table1: File downloading Time and Size



From above data, in graph 2, we can see file size of 1 is 10351 kb is required time uploading is 22 second, and file size of 2 is 25000 kb is required time uploading is 30 sec. and so on.

OUTPUT SCREEN:

HOME SCREEN

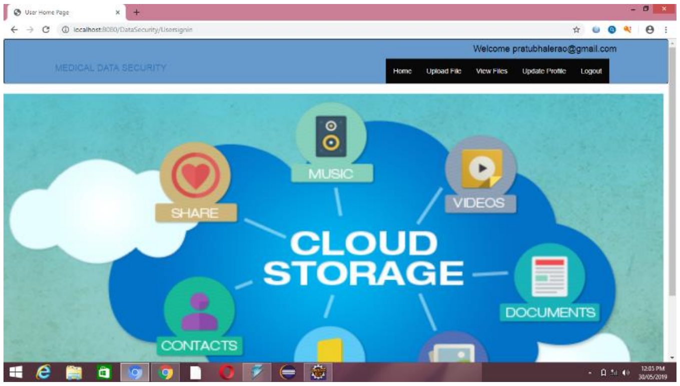

localhost:8080/DataSecurity/user_signin.jsp

Login Form

pratubhalerao@gmail.com

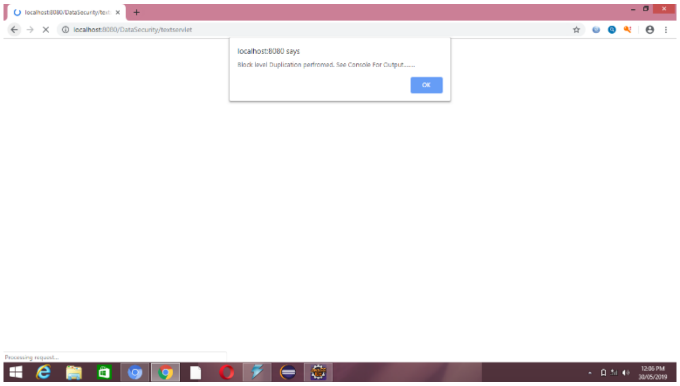
•••••••••••••

USER LOGIN


localhost:8080/DataSecurity/Usersignin

Welcome pratubhalerao@gmail.com

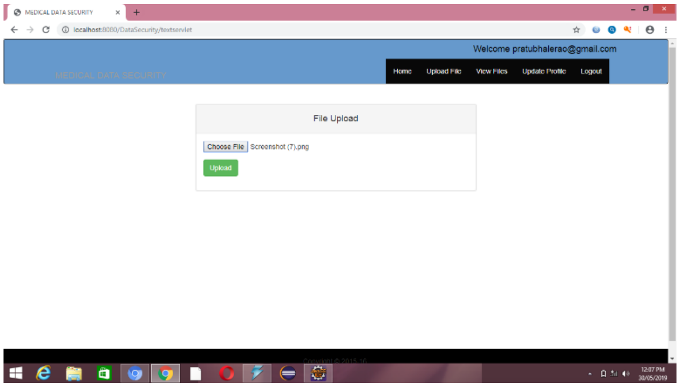
BLOCK DUPLICATION LEVEL


localhost:8080/DataSecurity/textservlet

localhost:8080 says

Block level Duplication perfromed. See Console For Output......

FILE UPLOADING

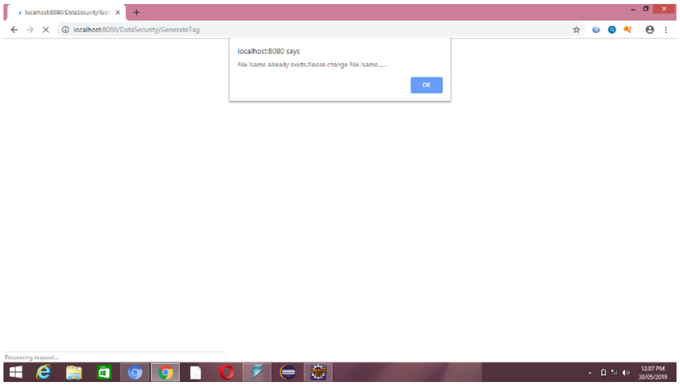

localhost:8080/DataSecurity/textservlet

Welcome pratubhalerao@gmail.com

File Upload

Screenshot (7).png

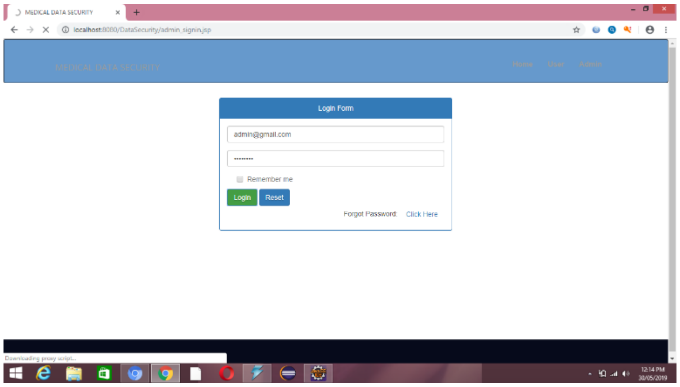
FILE ALREADY EXIST



localhost:8080 says

File Name Already exists,Please change File Name.....

ADMIN LOGIN



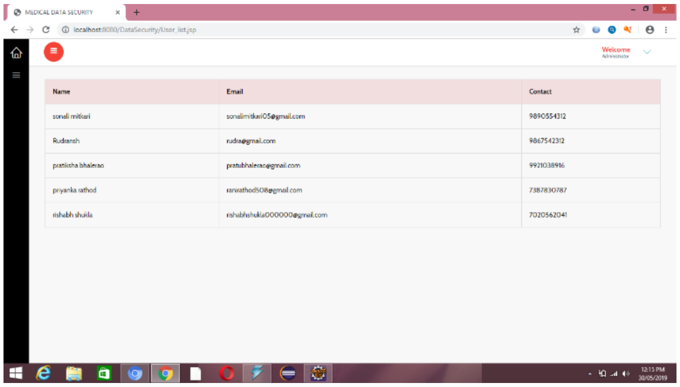
admin@gmail.com

****

□ Remember me

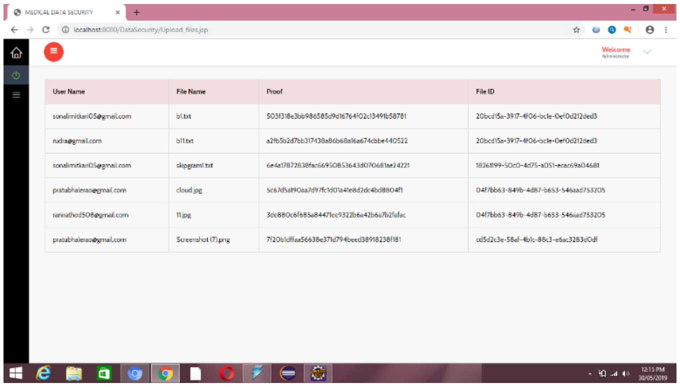
Forgot Password: Click Here

USER LIST



Name	Email	Contact
sonali mitkari	sonalimitkari05@gmail.com	9890554312
Rudransh	rudra@gmail.com	9867542312
pratiksha bhalerao	pratubhalerao@gmail.com	9921038916
priyanka rathod	ranirathod508@gmail.com	7387830787
rishabh shukla	rishabhshukla000000@gmail.com	7020562041

UPLOADED FILES



User Name	File Name	Proof	File ID
sonalimitkari05@gmail.com	b1.txt	503f318e3bb986585d9d16764f02c13491b58781	20bcd15a-3917-4f06-bc1e-0ef0d212ded3
rudra@gmail.com	b11.txt	a2fb5b2d7bb317438a86b68a16a674cbbe440522	20bcd15a-3917-4f06-bc1e-0ef0d212ded3
sonalimitkari05@gmail.com	skipgram1.txt	6e4a17872838fac66950853643d070681ae24221	18261199-50c0-4d75-a051-ecac69a04681
pratubhalerao@gmail.com	cloud.jpg	5c67d5a190aa7d97fc1d01a41e8d2dc4bd8804f1	04f7bb63-849b-4d87-b653-546aad753205
ranirathod508@gmail.com	11.jpg	3de880c6f685a84471ee9322b6a42b6a7b2fafac	04f7bb63-849b-4d87-b653-546aad753205
pratubhalerao@gmail.com	Screenshot (7).png	7f20b1dffaa56638e371d794beed38918238f181	cd5d2c3e-58af-4b1c-88c3-e6ac3283d0df

CONCLUSION

This Paper addresses a state of assisted monitoring in the cloud system of care with privacy preserving. The Cloud technology and the security process are used to provide an advantage to increase efficiency and accuracy with privacy using paired based cryptography and storage management using deduplication techniques.

ACKNOWLEDGMENT

Express our true sense of gratitude, sincere and sincere gratitude to my guide to the project Prof. Swapnil.S.Shinde for his precious collaboration and guidance that he gave us during my research, to inspire us and provide us with all the laboratory facilities, This it allowed us to carry out this research work in a very simple and practical way. We would also like to express my thanks and thanks to our coordinator, Prof. Nikhil Dhavase and all my friends who, knowingly or unknowingly, helped me during my hard work.

REFERENCES

[1] M. Chen, Y. Hao , K. Hwang, L. Wang, L. Wang, ”Disease Prediction by Machine Learning over Big Healthcare Data”, IEEE Access, Vol. 5, No. 1, pp. 8869-8879, 2017.

[2] Hadeal Abdulaziz Al-Hamid, Sk Md Mizanur Rahman, ”Securing Photos in the Cloud Using Decoy Photo Gallery”, 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, India, 22-24 March 2017

[3] M. S. Hossain, and G. Muhammad, ”Healthcare Big Data Voice Pathology Assessment Framework,” IEEE Access, vol. 4, no. 1, pp. 7806-7815,December 2016.

[4] M. Sriram, V. Patel, D. Harishma, and N Lakshmanan. A Hybrid Protocol to Secure the Cloud from Insider Threats. In Cloud Computing in Emerging Markets (CCEM), IEEE International Conference, Bangalore, 2014.

[5] Qinlong Huang,1,2 Licheng Wang,1,2 and Yixian Yang1,2 “Secure and Privacy-Preserving Data Sharing and Collaboration in Mobile Healthcare Social Networks of Smart Cities”

[6] D. Meister, J. Kaiser, and A. Brinkmann, “Block locality caching for data deduplication,” in Proc. 6th Int. Syst. Storage Conf., 2013, pp. 1–12.

[7] M. Lillibridge, K. Eshghi, and D. Bhagwat, “Improving restore speed for backup systems that use inline chunk-based deduplication,” in Proc. 11th

USENIX Conf. File Storage Technol, Feb. 2013, pp. 183–197.

[8] V. Tarasov, A. Mudrankit, W. Buik, P. Shilane, G. Kuenning, and E. Zadok, “Generating realistic datasets for deduplication analysis,” in Proc. USENIX Conf. Annu. Tech. Conf., Jun. 2012, pp. 261–272.

[9] D. T. Meyer and W. J. Bolosky, “A study of practical deduplication,” ACM Trans. Storage, vol. 7, no. 4, p. 14, 2012.

[10] G. Wallace, F. Douglis, H. Qian, P. Shilane, S. Smaldone, M. Chamness, and W. Hsu, “Characteristics of backup workloads in production systems,” in Proc. 10th USENIX Conf. File Storage Technol., Feb.2012,pp.33–48.

[11] El-Shimi, R. Kalach, A. Kumar, A. Ottean, J. Li, and S. Sengupta, “Primary data deduplication-large scale study and system design,” in Proc. Conf. USENIX Annu. Tech. Conf., Jun. 2012, pp.285–296.

[12] P. Shilane, M. Huang, G. Wallace, and W. Hsu, “WAN optimized replication of backup datasets using stream-informed delta compression,” in Proc. 10th USENIX Conf. File Storage Technol.,Feb.2012,pp.49–64.

[13] P. Kulkarni, F. Douglis, J. D. LaVoie, and J. M. Tracey, “Redundancy elimination within large collections of files,” in Proc.USENIXAnnu.Tech.Conf. Jun.2012, pp.59–72.

[14] Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou “A Hybrid Cloud Approach for Secure Authorized De-duplication” IEEE Transactions on Parallel and Distributed Systems: PP Year 2014.

[15] Shweta D. Pochhi, Prof. Pradnya V. Kasture “Encrypted Data Storage with De-duplication Approach on Twin Cloud “ International Journal of Innovative Research in Computer and Communication Engineering