ISSN: 0974-5823 Vol. 7 No. 2 February, 2022

International Journal of Mechanical Engineering

A bird's view on Deduplication and encryption technology – A secured data transaction in Cloud computing Network.

D. Arivazhagan[1] R. Kirubakaramoorthi[2] , G.Vennila[3]

AMET University, Kanathur, Chennai - 603112, Tamil Nadu, India;

Abstract

Cloud applications are getting rapid developments in world software market after the emergence of big data concept and data storage need for personal information to be more confident to store it in the cloud database. A technological breakthrough, challenges to provide the cloud security upon data in Cloud Computing. Basically, Cloud Computing allows sensor and temporary network for data access to computing resources with an effective security, low cost and low energy consumption on network service providers. Most of the applications in face book, Google applications and eBay etc., worked on big data. In this concept gives new dimensions for large amount of data to be maintained. In many cases the data can be repeated and those data are acquiring more spaces on the cloud, this vital purpose has become necessary to establish a new and advanced mechanism to properly treat those problems and to find the better solution for those problems. Through referred papers addressed following challenges in cloud computing are Key generation, accessing policy definition and enforcement, encryption-based keyword search process, user revocation for data sharing, data storage dimensionality for data engineering. This paper deals with introducing a solution for dynamic cloud storage system framework (DCSF) and policy based dedupe system which is provides advance security access control and quickest attribute key generations.

Key words: Cloud computing, Big Data, Security and Network

Introduction

Cloud computing becomes more demand in the field of big data-oriented applications. The clients can outsource their procedures and storage spaces through internet. This liberates clients from the issue of maintaining assets on local domain. The different type of information maintained in cloud environment is extremely sensitive, for instance health care data and shared network information. In cloud environment, securing and authenticating becomes challenging issue. At initial stage, before performing any transaction the user needs to validate itself and it must be assured that the cloud provider would not alter the records maintained. The user confidentiality should be enforced such that the cloud provider or any other cloud user cannot trace the credentials of the user. The cloud environment is responsible for the information maintained by users and also it is responsible for the services it offers. The authenticity of the cloud user who maintains the records also validated. Apart from the algorithm or procedures to assure confidentiality and security, the requirement for law administration is mandatory.

The lack of standardization, with regard to access cloud services, decreases interoperability and flexibility of switching among Cloud Service Providers(CSP). Hence, organizations may experience vendor lock-in, unless they are not willing to put a significant effort to accommodate their existing solutions to the new CSP. Access control systems should be reviewed and tailored for the suitability with the dynamic and distributed nature of the cloud environment. Wide variety of the use cases and imperfections of still developing cloud technologies and standards makes the aforementioned tasks even more complex and hard to realize in a unified way. Designing a secure data management system for a cloud environment is very essential. In particular, proposed the architecture of the system utilizing policy-based access control to the data. First, the research for the existing standards and technologies, and their potential adoption for the cloud environment were done. The results of prototype implementation in the proposed architecture, based on the findings will be discussed in this paper.

Literature Review

In general customers have to process the data which are to be stored in the cloud server. At the same time the cloud service provides has to minimize the storage space od their clients to optimize the storage space. Hence the service providers encrypt and decrypt the as per the requirement (Padhy *et al.*, 2011). The data security in cloud computing focusing two major areas in which the major focus is on random processing of encrypted and decrypted data without decryption and encryption is contradictory by Yuan and yu, (2013); The first one talks about how much storage space to be required and the second one talks about saving storage space by Beloglazov and Abawajy, (2012).

Data De-duplication

Eliminating duplicities of data is called as data deduplication which will leads to storage optimization. The aim of storage optimization is possible by removing the repeated data and multiple copies (Cui *et al*., 2019). Here the duplication ratio(DR) is expressed based on efficiency of deduplication It is a ratio between original size vs size od the duplication.

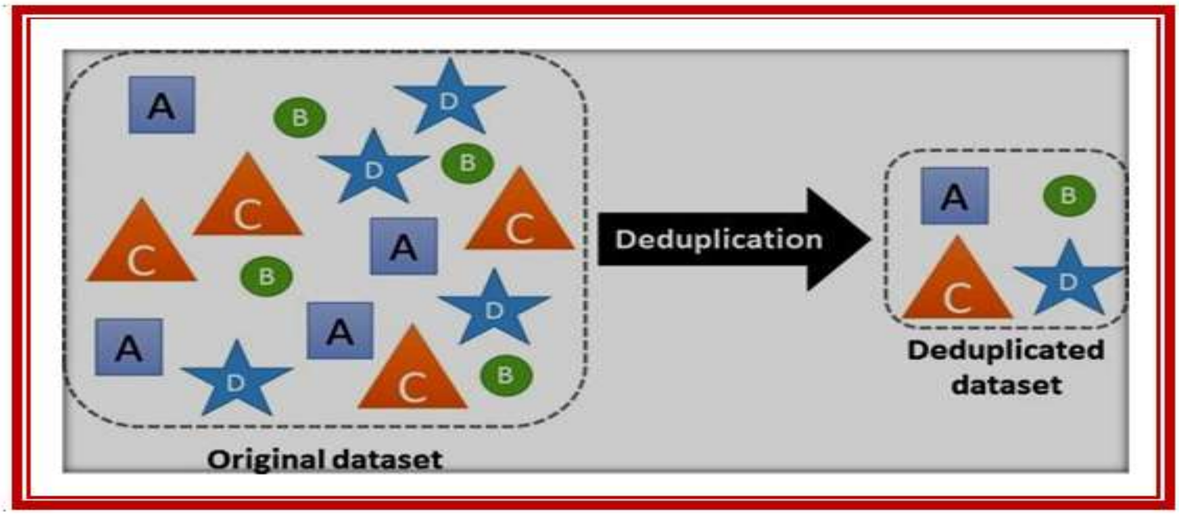


Diagram for data-deduplication-process

The above diagram shows that there are different kinds of files. The duplication ratio for the above file is 4:1. In general the idea of core duplication is more complex which comprises the issues like finding duplicate data, eliminating duplicate data and recording the details about the data removed in even method for easy retrieving.

For fining the duplicate data we have to classify the data set in to different pieces and then compared for equivalence. There are two methods for classifying data. One is file_based approach and another one is chunk_based approach. In the first method the file's data will be compared. If the data set has equal contents then only one set of data will be stored (Pollack et al., 2015). Chunk-based approach is more coarse-grained in which files are treated with rolling hash algorithm or fixed offset partitioning algorithm for splitting function and then the resultant chunk will be compared. The main advantage of first method is simplicity nad easy to implement. It takes lesser computation and quick yield result set. But the second method is better for elimination of duplication since it compare in-file changes. Both the methods need refined technology to remove duplicates.

The new algorithm or protocol developed based on MD5 is to eliminate the weakness of the old algorithm or protocol(Stallings, 1996). The new algorithm or protocol is highly secured and this system is fully depends on some hash value which is unique in nature. This also represents hard communication with summary. To achieve the goal the below facts to be taken into account:

- To find the negative aspects of MD5 protocol
- Explore all the possible technical solutions which will eliminate the negative components of MD5 by introducing new technology which supports original algorithm or protocol.
- Implementation of new Technology solution
- Use the original practical methods which was used in MD5 to test the new solution on MD5
- Test the new protocol and modify the algorithm based on outcome
- Check the new MD5 algorithm in real time applications
- Modify the existing algorithm using Asymmetric methods
- Check and validate the new algorithm or protocol.

Based on these attributes or factor of an entity recommendation cloud management servers will provide services. Cloud service mainly contains access control cloud servers and various unstructured cloud data storage . For an effective cloud services deduplication had 2 processing procedure there are classified in to Duplication with cipher text and Duplication without plain text.

Algorithm

The new algorithm developed with the aim of improving the old algorithm to be more effective and reliable but working methodology and structure should be maintained as per old algorithm. The main modification lies in the new hash algorithm in which calculation mechanism differs with compatibility. In general it occupies 1KB for identification and half KB for hash algorithm. The following simplified MDM protocol iteration shows the value o of the message.

iteration1

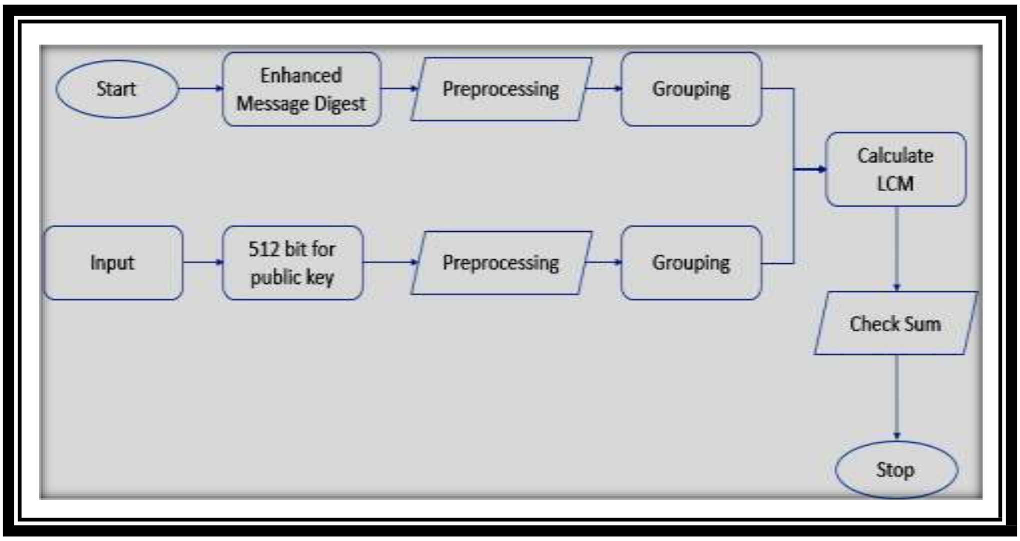
I$_{st}$ (message {0,1,2,3,4…}, x[i], Sol$_{n}$, Key_Genertation); /*
round$_{n}$*/

iteration2

II$_{nd}$ (message {0,1,2,3,4…}, x[i], Sol$_{n}$, Key_Genertation); /*
round$_{n}$*/

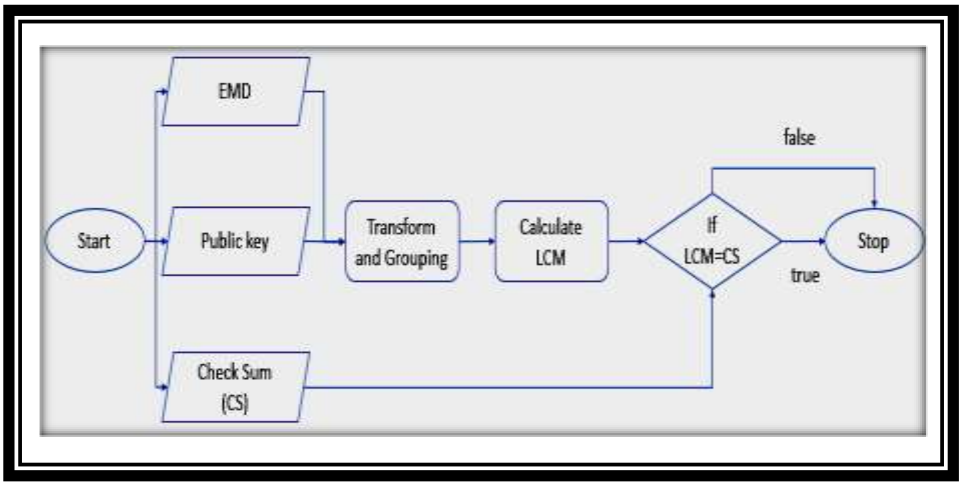
Create checksum

The matching key process diagram shows how the checksum calculated in MD5 algorithm. We have to generalize the values for comparison process like enhanced message digest has been sent for preprocessing where it has to be grouped before sending the same to calculate LCM in combination of half KB public key after processing. Once the Least Common Multiple (LCM) to found the value of division and sent the message digest value for checksum



Template and matching key process

Here, the files integrity has been checked and manipulated by issuing three values like public key, message and checksum. At the receiving end the message has been converted and the public key value from hexadecimal to decimal. Then the grouped value inserted in LCM equation for getting value and scaling comparison. The matched value gives a perfect file, otherwise file is not correct. The process has been described in the below diagram.

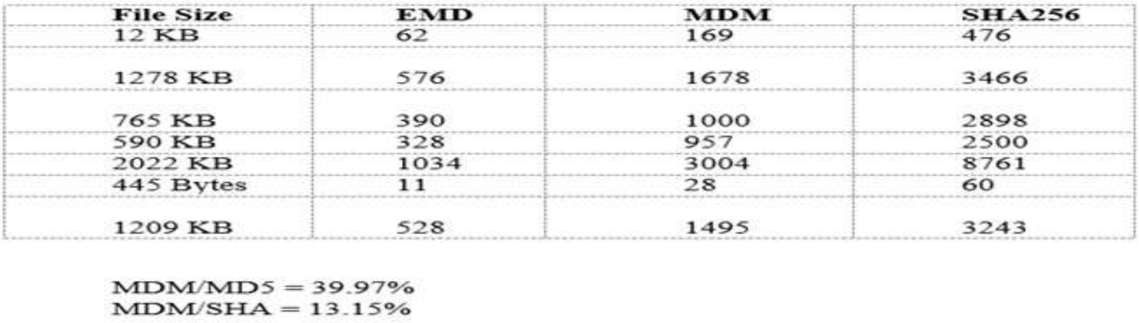


Password protection and permission access

Comparative study on MDM

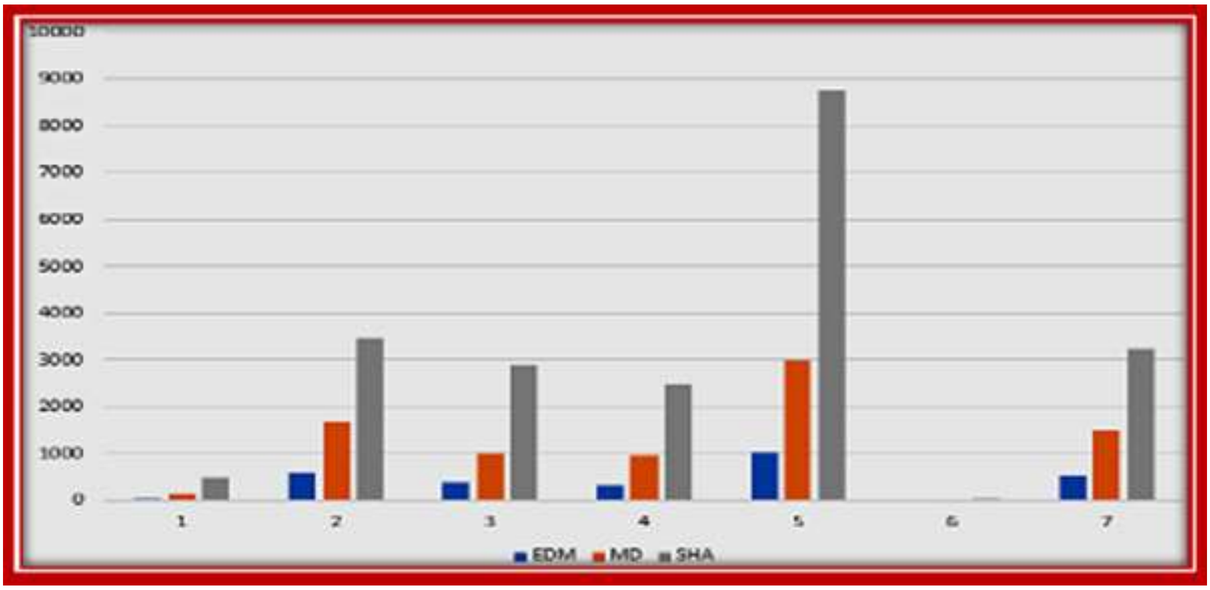
An illustrative table for new algorithm implementations and calculation of the time taken to implement, and comparing results with MD5 and SHA256, these results were selected through application in different companies for the purposes of checking the security of the algorithm and the possibility of dealing with them in the real environment, and the results were satisfactory to a large extent.

Comparative time of execution measured by (ms)



File Size	EMD	MDM	SHA256
12 KB	62	169	476
1278 KB	576	1678	3466
765 KB	390	1000	2898
590 KB	328	957	2500
2022 KB	1034	3004	8761
445 Bytes	11	28	60
1209 KB	528	1495	3243

MDM/MD5 = 39.97%
MDM/SHA = 13.15%

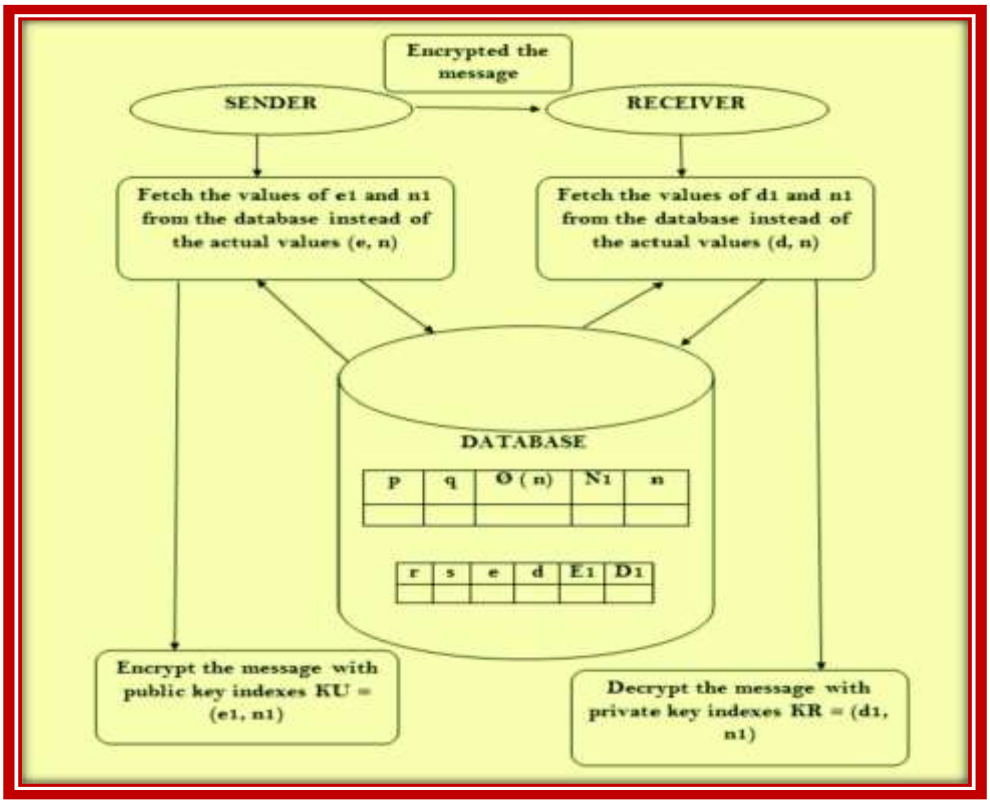


Comparative time of execution measured by (ms).

Secure deduplication attribute and policy based dedupe system

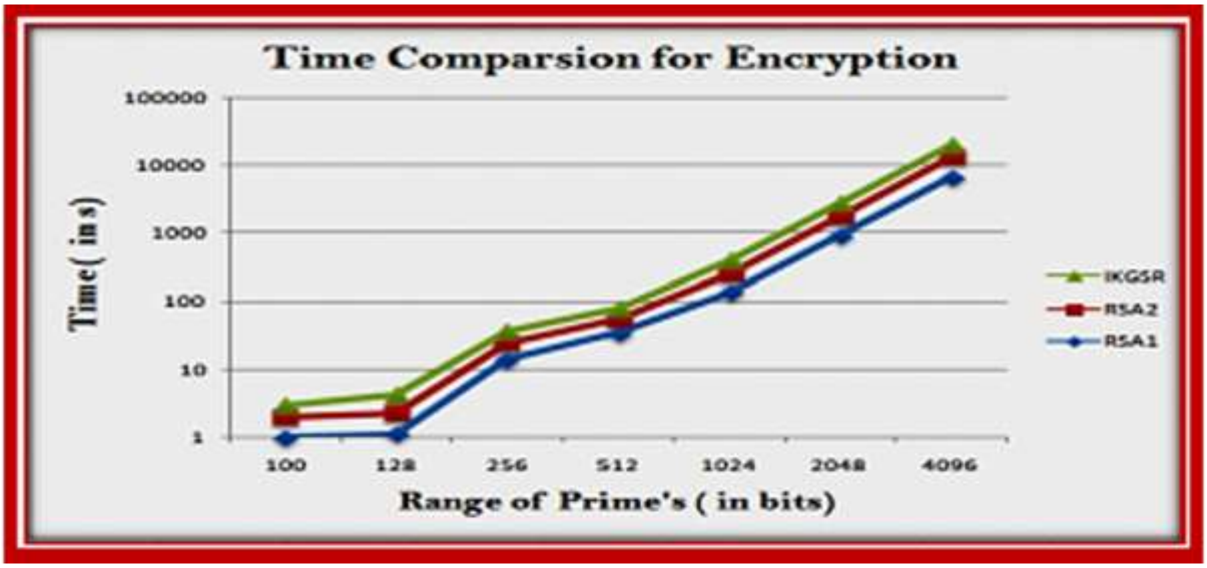
When dedupe checker receives signature and encrypted public key PK' of user for verification. First the dedupe checker will decrypt the public key using its secret key. Then the signature is decrypted by obtained user public key. Finally, a hash value will be produced. If that hash value matches with the data owner hash value then he/she is a valid data holder.

After performing wildcard-based keyword search, the cloud user gets the encrypted data from cloud server. Before doing decryption function, the following process is carried out between data owner and cloud user to enable the secret key sharing.

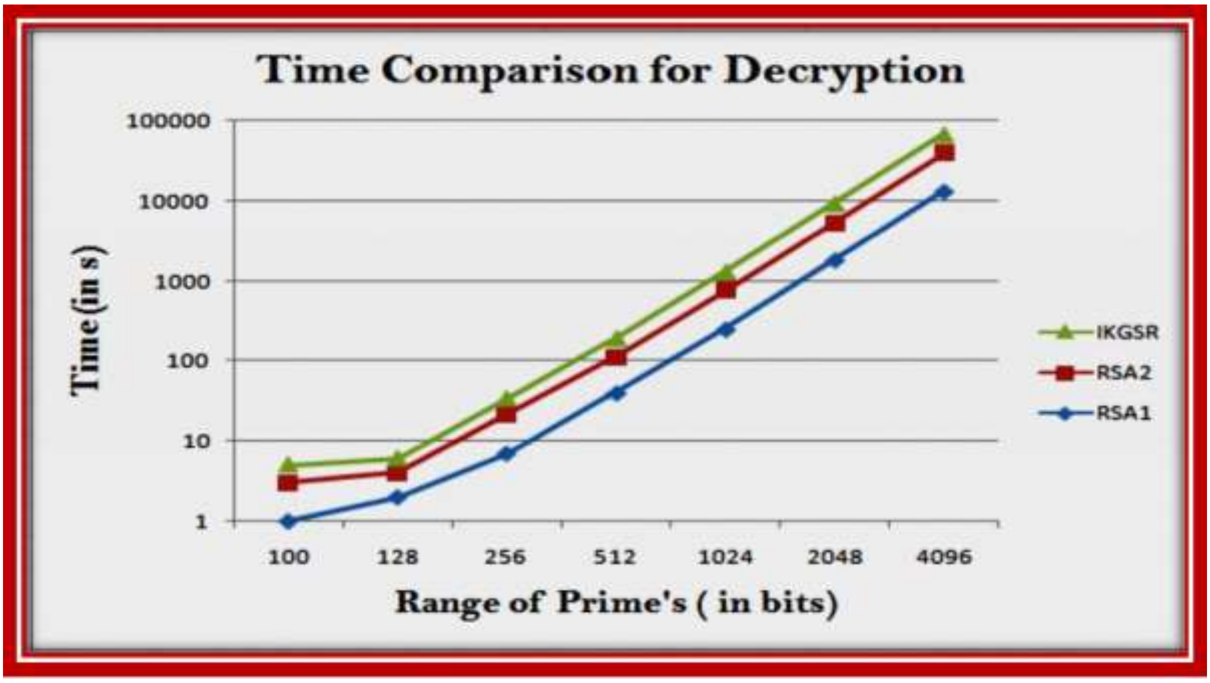

Encrypted the message
SENDER
RECEIVER
Fetch the values of e1 and n1 from the database instead of the actual values (e, n)
Fetch the values of d1 and n1 from the database instead of the actual values (d, n)
DATABASE
p
q
Ø (n)
N1
n
r
s
e
d
E1
D1
Encrypt the message with public key indexes KU = (e1, n1)
Decrypt the message with private key indexes KR = (d1, n1)

RESULTS

The performance of the proposed method was measured in terms of key formation time, encipher and decipher time by varying the length of input bits. The key formation time of the proposed technique is somewhat higher than that of both RSA1 and RSA2. The encryption and decryption time comparison of basic RSAs and proposed method are clearly shown in below figure.



Time Comparison for Encryption.



Time Comparison for Decryption.

SUMMARY AND CONCLUSION

A new method, namely, hybrid cryptographic access control, and secure retrieval of the health care cloud are proposed. Three research issues are overcome: secure cloud storage, key distribution on an untrusted network, and secure retrieval of encrypted data. The first issue is solved by introducing a hybrid cryptography method, which is a combination of the Blowfish and an enhanced RSA algorithm. Here, the existing method is applied to encrypt large data, and the secret key is encrypted using the enhanced RSA algorithm. The second issue, that of key distribution, is overcome by applying LSB-based steganography to hide the shared private key of the enhanced RSA algorithm in a cover image. The third issue is resolved by implementing a wildcard-fuzzy keyword search on encrypted health data, thus improving the efficiency of the proposed system and reducing medical errors. The performance of this contribution is compared with the traditional AES and Blowfish algorithms in terms of encipher time, decipher time, and index generation time. The storage space is also measured and compared with existing methods.

References:

1. Demirkan, H., & Delen, D. (2013). Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. *Decision Support Systems*, *55*(1), 412-421.
2. Ahuja, S. P., Mani, S., & Zambrano, J. (2012). A survey of the state of cloud computing in healthcare. *Network and Communication Technologies*, *1*(2), 12.
3. AlZain, M. A., Pardede, E., Soh, B., & Thom, J. A. (2012, January). Cloud computing security: from single to multi-clouds. In *2012 45th Hawaii International Conference on System Sciences* (pp. 5490-5499). IEEE.
4. Takabi, H., Joshi, J. B., & Ahn, G. J. (2010). Security and privacy challenges in cloud computing environments. *IEEE Security & Privacy*, *8*(6), 24-31.
5. **Dr. D. Arivazhagan, R Kirubakaramoorthi, (2020),**Develop Cloud Security In Cryptography Techniques Using DES-3L Algorithm Method In Cloud Computing, INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 9, ISSUE 01, ISSN 2277-8616