A Novel Approach For Multi Sharing Authenticated Filtered Data For Big Data Storage

TAMMIREDDY SUPRIYA
Mtech Student, Dept Of CSE
Priyadarshini Institute Of Technology
Nellore, Andhra Pradesh, India.

PARICHERLA MUTYALAIAH
Associate Professor, Dept Of CSE
Priyadarshini Institute Of Technology
Nellore, Andhra Pradesh, India.

CHALLA BHASKAR RAO
Associate Professor, Dept Of CSE
Priyadarshini Institute Of Technology
Nellore, Andhra Pradesh, India.

Abstract: **The requirement of unharmed big data storage is greater helpful than ever to date. The prime concern of the service is to assurance the privacy of the data. Nevertheless, the anonymity of the service providers, one of the greatest crucial appearances of privacy, should be deliberate concurrently. Furthermore, the service also should contribute realistic and delicate encrypted data sharing like that a data owner is allowed to share a cipher text of data between others under some specified conditions. In this mechanism the advantage of proxy re-encryption technique are employed in which only the cipher text to be shared securely and conditionally over multiple times. It also ensures that, original message and information identity of cipher text senders and it is not vulnerable to cipher text attacks. Furthermore, this paper shows that the new primitive is secure against chosen-cipher text attacks in the standard model.**

Keywords: **Privacy; Anonymity; Proxy Re-Encryption; Big Data;**

I. INTRODUCTION

Security is the greatest essential mission for any kind of services which produces storage for data. Due to its well-organized efficient data processing adequacy cloud play a vital role in preserving big data.

Due to the discovery of smart phones and new technology, the number of people using internet has increased as a result of this, data production is also increased millions of data is produced per day. For example over the world billions of people use face book as a result amount of data production is more so the need of big data storage is more desirable.

Many individuals and companies can view, modify and update their data stored in the cloud through remote accessing. In the middle of remote accessing there is an feasibility for some frequent affairs like privacy, security, data integrity, dynamic updates etc., Whenever it is not possible to analysis the data for consistency, as trillions of individual and companies data are flooding over the internet. As increase in number of individual users and public and private organizations choose to upload their data in cloud force us to keep the data more securable from being hacked.

For example in social networking site like face book there are personal information that is flooding over internet and the user does not want these information to be seen by other people so providing security for big data is more important than ever. Even companies upload their personal information in cloud and they expect this information to be secure. A basic need for big data is security for the data.

The data of a singular user should be maintained confidential and it should be accessed only by the authenticated user. While providing security, the most substantial appearances to be deliberate before storing the data is that, the anonymity of the service providers. The services which are used for data storage should provide high quality encrypted data sharing.

The services produce the way that, only the cipher text of the data is shared to the authorized individuals by the data owners under some restricted and specified conditions.

The characteristics noticed above are usually necessary to secure processing and these characteristics are obtained by employing a new approach called cipher text multi sharing mechanism.

In this mechanism a proxy re-encryption technique are employed in which the cipher text to be shared securely and conditionally over multiple times. It also ensures that, original message and information identity of the cipher text senders and receivers is not leaked and it also not vulnerable to cipher text attacks.

II. RELATED WORK

Following the concept of delegation of decryption rights introduced by Mambo and Okamoto, Blaze et al. [5] formalized the concept of proxy re-encryption and proposed a seminal bidirectional

PRE scheme. Afterwards, many PRE schemes have been proposed, such as [2], [3], [11].

Employing traditional PRE in the IBE setting, Green and Ateniese first defined the notion of IBPRE and proposed two unidirectional IBPRE schemes in the random oracle model: one is CPA secure and the other holds against CCA. Later, two CPA-secure IBE-PRE schemes (in the types of PKE-IBE and IBE-IBE) have been proposed. In 2008, Tang et al. proposed an IBPRE scheme with CPA security in the random oracle model, in which the delegator and the delegate can come from different domains. Afterwards, Wang et al. proposed two IBPRE schemes that are both collusion-safe and non-transferable in the random oracle model. However, of the PKG. With the same technique, a CPA-secure IBPRE scheme (in the type of IBE-PKE) without random oracles was proposed by Minzuno and Doi. Recently, an IBPRE with revocability and hierarchical confidentialities was proposed by Wang et al. Despite the scheme combines the IBPRE with proxy decryption, it still requires an interaction for re-encryption key generation.

In the multiple cipher text receiver update scenario, Green and Ateniese proposed the first MH-IBPRE scheme which is CPA secure in 2007. Later, a RCCA-secure MH-IBPRE scheme without random oracles was proposed by Chu and Tzeng. In 2010, Wang et al. proposed the first CCA secure MH-IBPRE with random oracles. These three schemes, however, are not collusion-safe. To solve the problem, Shao and Cao proposed the first CCA-secure MH-IBPRE in the standard model with collusion-safe property.

To preserve the anonymity, the following cryptosystems are proposed in the literature. To hide the information revealed by the re-encryption key, Ateniese et al. [1] first define the notion of key-privacy (i.e. an adversary cannot identify delegator and delegate even given re-encryption key) and proposed a CPA-secure scheme in the standard model. Later, Shao et al. revised the security model of key privacy defined in [1] and proposed a single-hop unidirectional PRE scheme with CCA security in the standard model. To prevent from being traceable, Emura et al. proposed a unidirectional IBPRE scheme for multi-hop setting, in which an adversary cannot identify the source from the destination cipher text. The scheme is proven to be CCA secure in the random oracle model. To protect the privacy of both delegator and delegate, Shao et al. proposed the first Anonymous PRE (ANO-PRE) scheme which is CCA secure with collusion safe in the random oracle model. The scheme ensures that an adversary cannot identify the recipient of original and re-encrypted cipher text even given the re-encryption key. In 2012, Shao also proposed the first anonymous IBPRE with CCA security in the random oracle model.

In the identity-based and attribute-based encryption settings, some well-known systems supporting anonymity that have been proposed. However, we will focus on the combination of anonymity and cipher text update properties. While multiple cipher text receiver update (denoting as M.U.), conditional share, collusion resistance (denoting as C.R.), anonymity, and without random oracle (denoting as W.R.O.), have all five been partially achieved by previous schemes, there is no efficient CCA-secure proposal that achieves all properties simultaneously in the standard model. This paper, for the first time, fills the gap.

III. PROPOSED SYSTEM

The proposed system architecture is as follows.

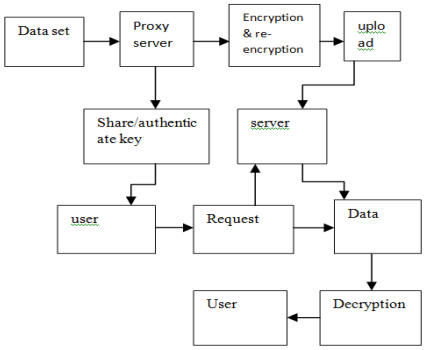

Fig: 3.1 System Architecture

First we have to collect the dataset to the proxy server. After the data set is encrypted and re-encrypted. After the data is upload to the server and decrypted to the data. Otherwise, the user wants to particular information and the proxy server share the authenticate key to user. If the user is authenticated, and request to the server and the server is decrypted to the data to the receiver.

Modules

We can implement this work using the following modules.

- Dataset selection
- Data Encryption and re-encryption
- Anonymity
- Conditional sharing
- Data Access

Dataset Selection

In this module, first we have to collect the dataset. Patient Dataset is selected for further process, which contains attributes such as name, age, gender, month, location and symptoms of the patients. After the dataset has been chosen, pre-process is done on the dataset. Pre-processing is nothing but a data cleaning. It is the process of

eliminating unwanted data, unwanted noise values etc., Data Pre-processing includes cleaning, normalization, transformation, feature extraction and selection, etc.

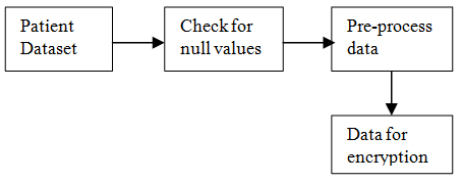

Fig: 3.2 Data set selection

Algorithm for pre-processing dataset:

Input: Dataset D

Output: Tokenisation (D)

Tokenize (d, delimit (,));

For (int i=0;i<=n; i++) //where n is the number of records in the dataset

{

if(i=”null” or i=”unwanted symbols”)

Remove (i);

}

Data Encryption and re-encryption:

Data is encrypted using encryption algorithm called Password Based Encryption with MD5 and DES. In this Module patient data is encrypted using PBEwithMD5andDES. Then, upload the data into the HDFS location. Encrypted and re-encrypted data is stored into HDFS Location. Hadoop File System was developed using distributed file system design. HDFS holds very large amount of data and provides easier access. To store such huge data, the files are stored across multiple machines. After Encrypt the dataset, the generated cipher text is stored in the HDFS.

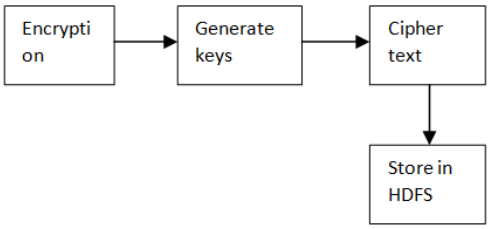

Fig: 3.3 Data encryption and re-encryption

Anonymity and multiple receivers communication:

After encryption the data stored in the HDFS. Stored data from HDFS can be share with multiple receivers. For a given cipher text, no one knows the identity information of sender and receiver. Here for hiding the receiver information's anonymization of particular receiver is implemented. For Example, if the receiver enters the user name that name will be anonymized and after that it will send to the particular user.

Receivers send their profiles such as id, password, and category to the senders. Category defines the data which is the receiver going to receive from sender and also receiver defines the category of what receiver category. For each and every individual receiver we generate the unique ids. Because, receiver’s information has been anonymized and after that only it will be forwarded to sender.

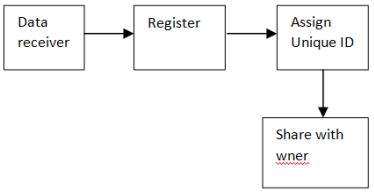

Fig: 3.4 Anonymity

Algorithm of Multiple Receivers Communication

For (int i=0; i<=n; i++)

{

Register (cloud);

Update (profile);

Anonymize (sensitive Attribute);

Generate (unique ID);

Generate random number=unique id;

Update Profile;

Send to Owner;

}

Conditional sharing:

After the Multiple receiver update process is completed, conditional sharing is going to be processed. In our implementation conditional sharing is takes place based on the user’s category and their receiver receiving data category, here we check the condition. If both the categories are similar means, receiver will be authenticated and he/she can receive their data in cipher text format only. Otherwise they can’t access the data and receiver will be excluded from the process.

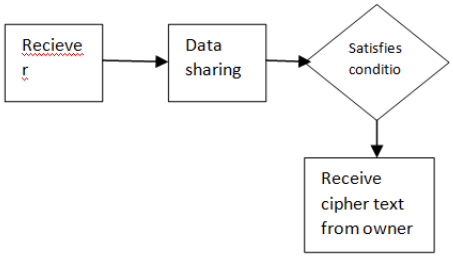

Fig: 3.5 Conditional sharing

Algorithm for conditional sharing

For (int i=0; i<=n; i++)

{
Receive (Multiple receiver details);
Share data ();
Share data ()
{
If (receiver category satisfies condition or matched with patient category)
{
Receive cipher text from owner;
}

}

Data Access

After user authentication, data has been accessed in encrypted format. User will get the key from sender and decrypt the data and they can view their original data. Original Data has been accessed by the receiver if they are authenticated only. Algorithm efficiency has been calculated by using the time efficiency of the algorithm. Here we calculated the encryption time and decryption time of the algorithm and finally we calculate the total time for executing the algorithm. By using these types of parameters we evaluate the performance of the existing system and proposed system.

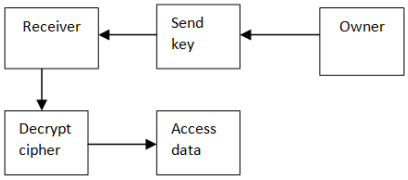

Fig: 3.6 Data Access

Algorithm for data access and evaluation:

For (int i=0; i<=n; i++)

{

Get keys from owner;

Decrypt ();

Access data;

}

IV. EXPERIMENTAL ANALYSIS

The fig.4.1 shows the experimental analysis of how the dataset has been collected and displayed.

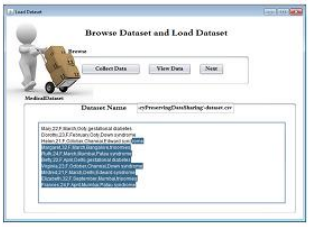

Fig: 4.1 Experimental result for data collection

The fig 4.2, 4.3 and 4.4 shows the experimental analysis of how data has been encrypted using MD5 and DES and re-encrypt using AES algorithm for all the attributes.

Fig: 4.2 Data Encryption

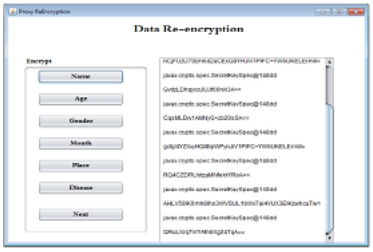

Fig:4.3 Data Re-Encryption

Fig:4.4 Encryption and Re-encryption for all attributes

The fig 4.5 shows the experimental analysis of how the multiple receiver details are received.

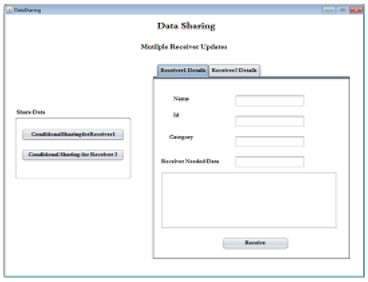

Fig: 4.5 Multiple Receiver update

The fig 4.6 shows the experimental analysis of conditional sharing mechanism and how the user has been authenticated.

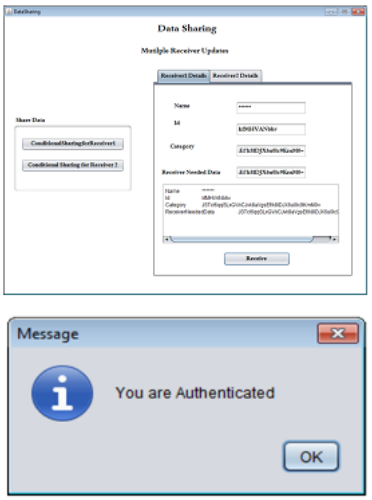

Fig: 4.6 Conditional sharing mechanism

The fig 4.7 shows the experimental analysis for how the data is decrypted and accessed in receiver side.

Fig: 4.7 Data accessing in Receiver side

V. PERFORMANCE ANALYSIS

The performance analysis shows that how the proposed system works better when compared to existing system. The fig 5.1 clearly shows that proposed system algorithm reduces the time taken for processing the data for both receivers ,which shown in the graph.

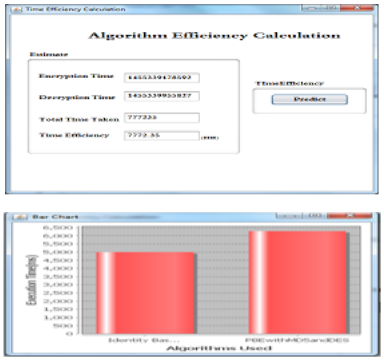

Fig: 5.1 Performance analysis of proposed algorithm based on execution time (receiver).

VI. CONCLUSION

This paper has introduced a new mechanism known as Anonymity Multi Hop – Identity Based Conditional Proxy Re-Encryption for secure data sharing in Big data.

This work specially focused on anonymity of the recipient and multiple cipher text of recipient which is required for protecting some sensitive confidential information while transferring the information.

This mechanism also ensures consistency and efficiency of data sharing in a time consuming way and in a cheaper way. It is the first time this new mechanism is approached to ensure security against chosen cipher text attack primitives.

VII. FUTURE ENHANCEMENT

The new mechanism proposed in this paper called AMH-IBCPRE has a problem that, it provides security against some of the chosen cipher text attacks because of its unidirectional property. This unidirectional IBCPRE scheme in which a hacker is not able to identify the source properties from the encrypted destination cipher text.

To safeguard the information of both sender and the receiver, a new scheme called, Anonymous-PRE (ANOPRE) was developed. This scheme guarantees that the hacker cannot identify the sender of original and re-encrypted cipher text even the re-encryption is provided. This scheme also ensures security from most of the chosen cipher text attacks.

Even there are lots of models proposed for providing security, this is the only scheme that achieves all the properties, even it combine some important features of standard models.

VIII. REFERENCES

[1] G. Ateniese, K. Benson, and S. Hohenberger, "Key-private proxy re-

encryption," in Topics in Cryptology–CT-RSA (Lecture Notes in Computer Science), vol. 5473. Berlin, Germany: Springer-Verlag, 2009, pp. 279–294.

[2] G. Ateniese, K. Fu, M. Green, and S. Hohenberger, "Improved proxy re-encryption schemes with applications to secure distributed storage," in Network and Distributed System Security. Berlin, Germany: Springer-Verlag, 2005, pp. 29–43.

[3] G. Ateniese, K. Fu, M. Green, and S. Hohenberger, "Improved proxy re-encryption schemes with applications to secure distributed storage," ACM Trans. Inf. Syst. Secure., vol. 9, no. 1, pp. 1–30, 2006.

[4] M. Bellare and S. Shoup, "Two-tier signatures, strongly unforgeable signatures, and Fiat–Shamir without random oracles," in Public Key Cryptography (Lecture Notes in Computer Science), vol. 4450. Berlin, Germany: Springer-Verlag, 2007, pp. 201–216.

[5] M. Blaze, G. Bleumer, and M. Strauss, "Divertible protocols and atomic proxy cryptography," in Advances in Cryptology. Berlin, Germany: Springer-Verlag, 1998, pp. 127–144.

[6] D. Boneh and X. Boyen, "Efficient selective-ID secure identity-based encryption without random oracles," in Advances in Cryptology– EUROCRYPT (Lecture Notes in Computer Science), vol. 3027. Berlin, Germany: Springer-Verlag, 2004, pp. 223–238.

[7] D. Boneh, X. Boyen, and E.-J. Goh, "Hierarchical identity based encryption with constant size ciphertext," in Advances in Cryptology– EUROCRYPT (Lecture Notes in Computer Science), vol. 3494. Berlin, Germany: Springer-Verlag, 2005, pp. 440–456.

[8] X. Boyen and B. Waters, "Anonymous hierarchical identity-based encryption (without random oracles)," in Advances in Cryptology– CRYPTO (Lecture Notes in Computer Science), vol. 4117. Berlin, Germany: Springer-Verlag, Aug. 2006, pp. 290–307.

[9] M. Green and G. Ateniese, "Identity-based proxy re-encryption," in *Applied Cryptography and Network Security* (Lecture Notes in Computer Science), vol. 4521. Berlin, Germany: Springer-Verlag, 2007,pp. 288–306.

[10] J. Camenisch, M. Kohlweiss, A. Rial, and C. Sheedy, "Blind and anonymous identity-based encryption and authorized private searches on public key encrypted data," in Public Key Cryptography (Lecture Notes in Computer Science), vol. 5443. Berlin, Germany: Springer-Verlag, 2009, pp. 196–214.

[11] R. Canetti, S. Halevi, and J. Katz, "Chosen-cipher text security from identity-based encryption," in Advances in Cryptology–EUROCRYPT (Lecture Notes in Computer Science), vol. 3027. Berlin, Germany: Springer-Verlag, 2004, pp. 207–222.

AUTHOR's PROFILE

TAMMIREDDY SUPRIYA has received her B.Tech degree in Information Technology from JNTU, Anantapur in 2013. She is now pursuing the M.Tech degree in Computer Science & Engineering at Priyadarshini Institute of Technology, Nellore, Andhra Pradesh, India. Her areas of research include Big Data and Information Forensics.

Paricherla Mutyalaiah has received his B.Tech degree in Computer Science & Engineering from Sree kalahasteeswara Institute of Technology at Srikalahasti in 2003 and M.Tech degree in Computer Science & Engineering from Rajeev Gandhi memorial college of Engineering (RGMCET) at Nandyala, Kurnool in 2009. He is dedicated to teaching field from the last 10 years. His research areas included "Data Mining and Big Data". At present he is working as an Associate Professor in Priyadarshini Inst of Technology, Nellore, Andhra Pradesh, India

CHALLA BHASKAR RAO has received his B.Tech degree in information technology from Narayana Engineering College, Nellore in 2006 and M.Tech degree in computer science and Engineering from SVU college of Engineering, Tirupati in 2008. He has 9 of years Teaching and Industrial Experience. His areas of research include "Data mining and Natural Language Processing". At present he is working as a Associate professor and HOD of CSE in Priyadarshini Institute of Technology Nellore, Andhra Pradesh, India.

