Privacy-Preserving Cipher text Multi-Sharing Control for Big Data Storage

C.P Nijitha Mahalakshmi
Research scholar, School of computing science, Vels University

Dr.Y.Kalpana
Associate Professor, School of computing science, Vels University

Abstract- The need of secure big data storage service is more desirable than ever to date. The basic requirement of the service is to guarantee the confidentiality of the data. However, the anonymity of the service clients, one of the most essential aspects of privacy, should be considered simultaneously. Moreover, the service also should provide practical and fine-grained encrypted data sharing such that a data owner is allowed to share a cipher text of data among others under some specified conditions. This paper, for the first time, proposes a privacy-preserving cipher text multi-sharing mechanism to achieve the above properties. It combines the merits of proxy re-encryption with anonymous technique in which a cipher text can be securely and conditionally shared multiple times without leaking both the knowledge of underlying message and the identity information of cipher text senders/recipients. Furthermore, this paper shows that the new primitive is secure against chosen-cipher text attacks in the standard model.

Key Terms: Encryption, Decryption, Privacy-Anonymity.

I. INTRODUCTION

Security is the most important concern for any type of services which provides storage for data. Due to its efficient data processing capability cloud play a vital role in keeping big data.

Many individuals and organizations can view, modify and update their data stored in the cloud through remote accessing. During remote accessing there is an possibility for some common issues like privacy, security, data integrity, dynamic updates etc… every time it is not possible to check the data for consistency, as trillions of individual and organizations 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.

The data of an individual user should be kept confidential and it should be accessed only by the authenticated users. While providing security, the most important aspect to be considered before storing the data is that, the anonymity of the service providers. The services which are used for data storage should provide a high quality encrypted data sharing. These services provides 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 features mentioned above are commonly required to maintain secure processing, and these features are achieved by employing a new technique called cipher text multi sharing mechanism.

In this mechanism a 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 receivers is not leaked and it also ensures it is not vulnerable to cipher text attacks.

II. LITERATURE SURVEY

Magoulas Roger, et al., paper of “Introduction of Big data” helps to gain knowledge about the basics of big data.

Hassan, et al., paper of “Demystifying cloud computing” helps to gain knowledge of cloud computing basics.

Bellare, et al., paper of “Introduction to modern cryptography” helps to gain knowledge of cryptography and its types.

G.Ateniese, et al., proposed proxy re-encryption scheme which provides security in cloud based data sharing.

D.Boneh and X.Boyen proposed ID secure identity-based encryption scheme which helps to achieve Anonymity by hiding the sensitive information’s from the hacker.

C.K.Chu and W, .G.Tzeng proposed a new paper for “Identity-based proxy re-encryption” provides the information about the combination of proxy based re-encryption and identity-based encryption algorithms.

Let us discuss the basics of technologies that are used like big data, cloud and cryptography.

III. BIG DATA

Big data is a concept which is used to describe a huge amount of both structured and unstructured data that is so large. It becomes very difficult to process such data using traditional database models like (DBMS, RDMS) and software methodologies. A most important concern is that, if the volume of data is too big or it moves too fast or it exceeds current processing capacity, then it becomes a risky one.

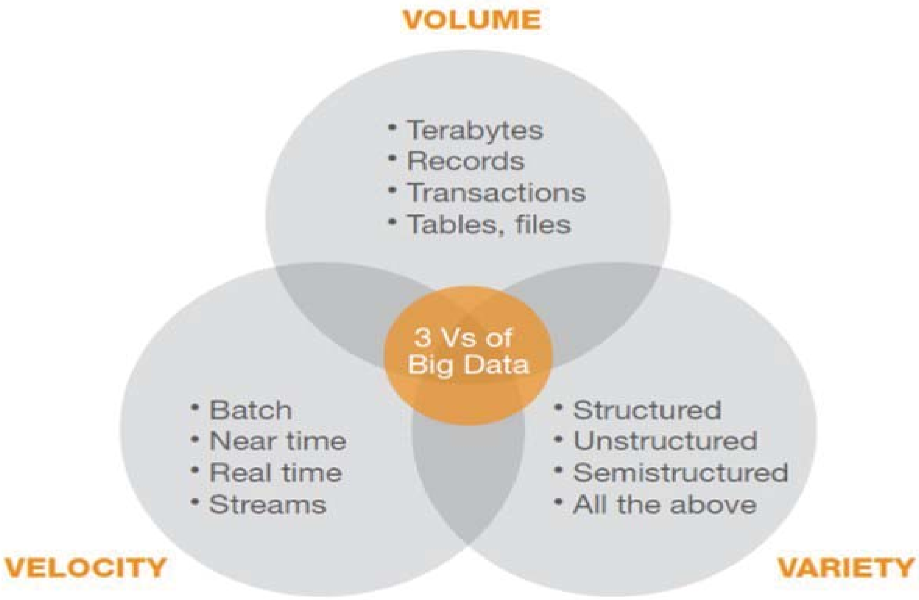
Big data has the ability to provide, improve operations and it makes process faster, and take more intelligent decisions for the organizations. It gets origin from Web search companies who had the problem of querying very large distributed aggregations of loosely-structured data (XML, XHTML, and web based document).

A. Characteristics:

Big data can be characterized as 3v’s:

- Volume: big data doesn't sample. It just observes and tracks what happens
- Velocity: big data is often available in real-time
- Variety: big data draws from text, images, audio, video; plus it completes missing pieces through data fusion

But, the challenge of keeping those huge amounts of structured and unstructured data leads to the change in 3v’s. As a result of increase in number of data sharing devices, it alternates the traditional 3v’s definition.



B. Importance of Big Data:

When big data is effectively captured and analyzed efficiently, it can lead to efficiency improvements, increased sales, lower costs, better customer service, and improved products service. Companies are able to gain a more complete understanding of their business, and their customers.

1.Effective use of big data exists in the following areas:

- Using information technology (IT) logs to improve IT troubleshooting and security breach detection, speed, effectiveness, and future occurrence prevention.
- Use of voluminous historical calls centre information more quickly, in order to improve customer interaction and satisfaction.
- Use of social media content in order to better and more quickly understand customer sentiment about you/your customers, and improve products, services, and customer interaction.
- Fraud detection and prevention in any industry that processes financial transactions on-line, such as shopping, banking, investing, insurance and health care claims.
- Use of financial market transaction information to more quickly assess risk and take corrective action.

C. Evaluation of Big Data

1. Column Oriented Databases:

Row oriented databases are the databases which are excellent for online transaction processing, online shopping system etc… with high efficiency, but, if the amount of data volume increases or more unstructured data are available, then the efficiency of query processing becomes degraded.

In order to solve this efficiency problem, column oriented databases came to existence which stores data by focusing on columns instead of storing it as rows. It enables huge data compression and processes the queries efficiently.

These databases are more efficient for customer relationship management, data ware housing etc…

2. Schema-less databases:

There are number of database types available in big data technology, such as document storage, key-value pair storage which is used for storing and retrieving huge amount of both structured and unstructured data.

These databases achieve high grade performances by avoiding part or all of the restrictions associated with traditional databases, such as read-write consistency, read-only consistency etc…

3. Map Reduce:

This is a programming paradigm that allows for massive job execution scalability against thousands of servers or clusters of servers. Any Map Reduce implementation consists of two tasks: The "Map" task, where an input dataset is converted into a different set of key/value pairs. The "Reduce" task, where several of the outputs of the "Map" task are combined to form a reduced set of tuples.

IV.CLOUD COMPUTING

Cloud computing is a technology to access the resources available in the servers through Internet. Cloud computing technology becomes popular in the recent years due to its several advantages over traditional methods, like flexibility, scalability, agility, elasticity, energy efficiency, transparency, and cost saving. Cloud resources are shared resources which can be accessed by any one, anytime and anywhere. It is accessible through any devices like mobile, desktops, laptops, tablets etc… The resources and information are provided for the users based on on-demand services. It allows the users to pay only for the resources and workloads they use.

Cloud is nothing but a server and a number of servers interconnected through it. Cloud providers are the one who own large data centers with massive computation and storage capacities. They sell these capacities on-demand to the cloud users who can be software, service, or content providers for the users over the internet. In the recent years the major cloud providers are Google, Microsoft, and Amazon etc...

*A. Architects to Be Factor in Cloud computing Designs***:**

1.1 Infrastructure as a Service:

Infrastructure as a Service is a form of cloud computing service which provides virtualized resources which are required over the Internet. Among many services it is an important one because, it provides, server spaces, bandwidth requirement, internet connections, load balancing etc…

1.2 Platform as a Service:

Platform as a service is a form of cloud computing services which provides a platform which allows customers to develop, run, and manage their web applications without the necessity of developing and maintaining the infrastructure which is required for developing and launching an application.

1.3 Software as a Service:

Software as a Service is a form of cloud computing services which provides the software's in which the developed applications are hosted by the service provider. Further, a service provider gives access for those applications to the customers through Internet by terms of pay per use.

1.4 Network as a Service:

Network as a Service is a type of business model which allows us to access the network functionalities directly and securely. Service providers allow us to access the Internet virtually by terms of pay per use or for monthly basis.

B. Virtualization:

Virtualization is the key concept in sharing the resources. It allows the single instance of resources to share among multiple customers or among different organizations. Creating a virtual machine over existing operating system and hardware is referred as Hardware Virtualization. Virtual Machines provide an environment that is logically separated from the existing hardware.

C. Big Data in cloud:

Most of the technologies are closely associated with the cloud. The products and platforms mentioned are either entirely cloud-based or have cloud versions themselves. Big Data and cloud computing go hand-in-hand. Cloud computing allows organizations of all sizes to get more value for their data than ever before, by enabling fast analytics at a minute of previous costs. This, in turn drives companies to acquire and store even more data, creating more need for processing power and driving a virtuous circle.

V. SECURITY TECHNIQUE

A. Cryptography:

Cryptography is the study and practice of techniques which is used for storing and retrieving information securely and privately, by protecting the data from third parties. To provide security, it involves the processes of encryption and de-encryption.

It serves wide range of applications such as online banking, ticket reservation, logging in to Facebook, Gmail, Twitter etc… where user personal identities are protected confidentially.

1.1 Terminologies:

- *Encryption:* It is the process of converting the original data into some unreadable form to protect the data while transferring from sender to receiver.
- *Authentication:* It ensures that the message was sent from the sender by verifying sender information associated with the message.
- *Integrity:* It ensures that, information received by the receiver is not modified anywhere during transfer.

B. Types of cryptography:

1.1 Secret key cryptography:

It is a type of cryptography which uses a single key for encryption and de-encryption. A key used by the sender for encryption is used by the receiver for de-encryption.

1.2 Public key cryptography:

It is a type of cryptography which uses two keys to provide security. In this sender and receiver has a separate private key as a secret key and a public key is shared between them for communication.

1.3 Hash function:

It is a type of cryptography which does not involve any key. Instead of using keys, it uses hash values of fixed length. Hash values are calculated depending on the text message.

VII. TRADITIONAL METHOD

Traditional Cryptography encryption techniques such as identity based encryption, public key encryption etc… are used to provide security to the data from third party hackers.

By employing traditional mechanisms it is not possible to protect some confidential sensitive information being leaked to the public and also to the cloud server. This is because traditional mechanisms do not consider the anonymity of a cipher text sender or receiver. Accordingly anyone with the knowledge of obtaining a cipher text can obtain the public key of the text, which means hacker will know the owner of the text.

Public key encryption (PKE) is the more frequently used encryption mechanism which allows a data sender to encrypt data by using the public key of the receiver such that, only the valid recipient can access gain to those data.

Public key type of encryption does not support anonymity, update of cipher text receiver which is required to maintain consistency and efficiency.

There are some traditional mechanisms such as anonymous IBE which consider anonymity of cipher text sender and receiver, but it does not support the update of cipher text recipients.

Traditional encryption mechanisms are applicable only for small amount of data. If the encrypted data is large, encryption and de-encryption process might be a time consuming and a costlier one.

Drawbacks:

- Encryption and de-encryption process is time consuming and cost effective.
- Security is less.
- Update of cipher text recipient is not possible.
- Anonymity is not considered.

A. Anonymity and Multi-Hop (AMH):

To solve these problems a new technique known as AMH-IBCPRE is proposed. It is a unidirectional approach which achieves multiple receiver cipher text updates; consider anonymity, enables conditional data sharing etc…

Anonymity: Anonymous communication is required for the users to send messages to other users without revealing their identities.

Multi-Hop: Receiver can be updated multiple times for a given cipher text.

Proxy Re-encryption (PRE) technique is proposed which allows a semi trusted party known as proxy, which converts an encrypted cipher text of particular key into an encryption of the same message by using another separate key, which leads to decrease the workload of the data owner.

Identity-Based Cryptography (IBC) is a form of public key cryptography in which recipient identities such as e-mail address, id, and name can be used to evaluate a public key.

Identity-based conditional proxy re-encryption (IBCPRE) technique is a form of PRE method. It focuses on two aspects.

1. To extend the proxy re-encryption method to identity-based encryption method.
2. To extend the proxy re-encryption method to support conditional proxy re-encryption method.

Advantages:

- Enable anonymity.
- Multiple receiver update.
- Conditional data sharing.
- Less workload.
- Free from chosen cipher text attacks (CCA).
- Free from collusion attacks.

VII. CONCLUSION

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

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.

VIII .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 (ANO-PRE) 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.

REFERENCES

[1] G. Ateniese, K. Benson, and S. Hohenberger, 2009 ,“Key-private proxy Re-encryption,” in *Topics in Cryptology–CT-RSA* (Lecture Notesin Computer Science), vol. 5473. Berlin, Germany: Springer-Verlag, pp. 279-294.

[2] J. Shao, 2012, “Anonymous ID-based proxy re-encryption,” in *Information Security and Privacy* (Lecture Notes in Computer Science), vol. 7372. Berlin, Germany: Springer-Verlag, pp. 364–375.

[3] Sun Microsystems, 2009, “Introduction to Cloud Computing Architecture”, Sun Microsystems Inc., white paper, pp. 1-17.

[4] MELL, P. and GRANCE, T, 2009. “Definition of Cloud Computing”, Draft NIST working, vol.5, pp. 7-19.

[5] Magoulas, Roger; Lorica, Ben (February 2009), “Introduction to big data”, vol. *Release 2.0* (Sebastopol CA: O'Reilly Media), pp.1-7.

[6] Bellare, Mihir; Rogaway, Phillip (21 September 2005), “Introduction *to Modern Cryptography*, by random grids, vol.1, pp.10-21.

[7] D.Boneh and X.Boyen, 2007, “Introduction”. “ID secures identity-based encryption”, Berlin, Germany: Springer-Verlag, 2007, vol.3027, pp. 223–238.

[8] C.K.Chu and W, .G.Tzeng (August 2006), “Identity-based proxy re-encryption”, (Lecture Notes in Computer Science), Berlin, Germany: Springer-Verlag, 2006, vol. 4779, pp. 189–202.