ISSN: 0975-766X
CODEN: IJPTFI
Available Online through *Research Article*
www.ijptonline.com

APPROACH ON BIG DATA USING HYBRID CLOUD

E.Vijayan*[1], Senthil Kumar. K[2], Aishwarya Singh[3], Yasmeen Khatoon[4]
School of Information Technology and Engineering [site], Vit University.
Email: vijayvsi81@gmail.com

Received on 09-08-2016 *Accepted on 05-09-2016*

Abstract

The quantity of data being generated is huge with rapid increase in technology and electronic communication , which results in difficulty in management of big data cost effectively. Cloud computing along with big data provides reduction in cost and storage. But there is concern regarding privacy of data stored in public cloud as the data stored in cloud could be sensitive and the owner would not want any other to scan the data .as the volume of data is huge due to increase in usage of mobile devices, the conventional cryptographic approach is not suitable. Hence in this paper we propose a methodology focusing on image data which has huge volume rather than text data.

Keywords: Cloud computing, Big data, Image, Data security and Privacy.

I. Introduction

The data generated by social sites , medical and surveillance systems have grown in an unexpected manner due to rapid development of electronic and communication technology , thus it is difficult for large organizations to store and manage big data.Cloud computing is considered cost effective and thus efficient model for storage of big data in an efficient manner. Cloud computing emphasizes on pay-per-use model. But considering certain cases there is a growing concern regarding privacy of data which results in confusion on whether the adoption of cloud computing for big data is safe or not ? Social networking sites and medical systems contain sensitive data in form of image or text . Cloud Service Providers (CSP's) are the people whom own the infrastructure on which client's data is stored, have full access on the data. Therefore sometimes the data stored in Public Cloud can be scanned. The data stored in cloud can be vulnerable to attackers, if the cloud is not secure.

Currently many people implement conventional cryptographic algorithms, like AES. The data is encrypted and thus stored in public cloud. As for Image data the size is more than text data. So, by implementing conventional cryptographic algorithms on image data leads to heavy computation overhead. In mobile devices heavy battery

consumption and slow performance due to constraints regarding resources. Hence use of traditional cryptographic techniques is not advisable. Recently many image encryption algorithms have been invented to enhance the speed of process which uses substitution-diffusion process. In substitution step the pixels of the image are shifted and therefore in diffusion step the pixels are shuffled. Another aspect is to utilize hybrid cloud by dividing sensitive data and non-sensitive data and thus storing them in trusted private cloud and un-trusted private cloud. But if we implement this methodology most of the sensitive data will be stored in private cloud only and therefore their will be loads of data in private cloud. The need is to minimize the storage and processing in private cloud and also leading to public cloud do most of the processing. Therefore one needs to analyse aspects like how to achieve big data privacy and security using hybrid cloud ? Using Hybrid Cloud would lead to communication stress between the private and public cloud. Factors to focus on is not only on data privacy but also to reduce storage and processing between the private and public cloud. A factor that should be kept in mind is that the delay should be small between the public and private cloud.

In this paper we propose a methodology that can help in achieving Image data privacy in hybrid cloud. We use one – to- one mapping function for image encryption, which results in high speed process of substitution and diffusion. Therefore private cloud stores only the parameters of mapping function.

II. Related Work

A. Cloud and data security and privacy

[3] For security and privacy of data we use combination of attributed based encryption(ABE), proxy re-encryption to gain access over encrypted data. The access control mechanism is initiated by secure private cloud to check user's authorization. [5] proposes anonymous cloud access and control over privileges by using multiple authorities in cloud computing system. Above stated papers do specify the methods these days are implemented to provide data secrecy or privacy, data owner will have to encrypt data using traditional cryptographic techniques before sending to public cloud.

B. Image Encryption Methods

In [6], the image is defined by pixels and thus each pixel is permutated using technique called nearest-neighboring coupled map lattices(NCML) . A random sequences are generated which are XOR's with multi-scroll chaos system. The pixel values are further changed. But using chaos system results in heavy computation as the data is iterated many times therefore this method is very difficult to implement. [10] Implements an efficient Image Encryption algorithm using Hybrid Cloud to secure data stored in public cloud. In this technique also pixels of the image are

shuffled by random noise before being shuffled .However, pixels within one pixel are modified by the common noise, leading to attacks based on gradient analysis .

III. Approach On Big Data Using Cloud

In hybrid cloud the data comes from private cloud which has to be already processed from the servers.If the data is insensitive then it can be sent to public cloud. After processing of data it is sent to public cloud whereas only small amount of data is sent to private cloud. When user will ask for the data then both the public as well as private cloud are connected and then send to the user. Public cloud has full access over the user's hardware, software and network. Therefore our emphasis should be to protect the data in public cloud using hybrid cloud. Most importantly the focus should be to remove the sensitive data and store it in trusted private cloud and thus to store the in-sensitive data in un-trusted public cloud. Complete storage of sensitive data in private cloud will require too much storage. Therefore our aim is to manage the amount of data stored in private cloud, manage communication between the public and private cloud , and also manage the delay due to communication between the public and private cloud.

IV. Image Data Privacy

A. Dividing image into blocks

In this step we divide the large image into n number of blocks, in which every block has the same size . for example ,an image has the size of 256*256 , and the size of each block is 32*32 , therefore the image is divided into n=256/32*256/32 =64 pieces.

B. Mapping Function

In this step we perform a one-to-one mapping function which maps, the original pixel value to a unpredictable value. The value of the original pixel value is mapped to p' which is any value.

C. Reverse of mapping function

We perform reverse of one to one mapping function and thus determine from where p' pieces came from and also determine the line which is the line where point locates and also location within the line.

D. Recovering Images

When the user asks for the image data, the request is sent to both the private cloud and public cloud at the same time. The data required to recover image is taken from private cloud. To do that first we take the shuffle order of the image, then the blocks of the image are re-ordered of the shuffled image from the public cloud , which thus gives the modified image.

V. Performance Evaluation

Here evaluation of the image privacy scheme will be done.

A. Security analysis of image data

As we are dividing the image into pieces, and therefore when we are shuffling the pieces, like the jigsaw puzzle. We propose a technique where each pixel of a particular color dimension is mapped to another value using one-to-one mapping. The jig saw puzzle performs NP completeness; therefore it cannot be used in polynomial time.

B. Efficiency of Image Encryption

For understanding the efficiency of the technique we will compare our technique with AES algorithm. Thus we find that the time required for processing an image increases as the size of the image increases. AES algorithm has iteration which requires much processing time, whereas our technique has no iteration therefore it results in faster processing. AES as we know has four rounds of iterations, which takes much long to process.

VI. Issues In Big Data & Cloud

There are many issues regarding big data in cloud computing, therefore it is difficult to say that which solution would be best to optimize the features of cloud. Initially and still many organizations use large database systems or large data warehouse systems. Cloud is also called Elastic Cloud Computing-EC2, which depicts the characteristic of cloud which is elasticity, it can have immense of big data. Using cloud computing provides reduced cost, reliability, elasticity. one very prominent feature of cloud computing is that it provides pay per use model i.e. instead of developing and deploying an entire database in organization ,we use cloud on service on rent. Therefore rent has to be paid for only used services .cloud computing is good for medium or small organization whereas for big organizations we require large database only because as the data becomes huge simultaneously the cost of using cloud also increases. Sometimes for big organizations the data might become unreachable, therefore they need a robust environment wherein the data is accessible as and when required. The rent of using cloud is more than developing an database system. Therefore these issues have to be focused on to get a better cloud computing environment. Big Data consists of structured and semi-structured or unstructured data. Therefore goal should be to provide scalable and distributed databases to manage the workload and intensive queries. Parallel databases do help in intensive workload handling whereas distributed databases are not very helpful in handling such workload. To handle excessive workload we use new kind of database which is key-value database and document database etc.

In industries and academics we use Map-Reduce and Hadoop, which uses Key-value store database. MapReduce is very efficiently scalable but difficult to use. To work in mapReduce we need very efficient programming skills.

Document, graph, column oriented databases are other popular databases to use. Size, Complexity, Design, Query language and Optimization are the aspects which have to be focused on to provide a platform of big data using cloud computing. Size in the database becomes very huge when many entities are included in the table. Hence conversion of the table using key-value is required. Another aspect to focus on is *complexity*, big data comprises of various kinds of data therefore to reduce the complexity and thus differentiate the structured and semi-structured data a standard approach has to be established. Hence another aspect is design, the design of the database is dependent on time, speed and occurrence of the data. These three components prescribe the design of the database will be, Time refers to the time required to process the data, speed determine the frequency at which database is refreshed for new data updation. Query language is used for the optimization of the database so that it can fully functionally depend and formulate all the processing done. For big data storage is one of the most prominent aspects. The data has to be managed and processed effectively.

The transaction process is managed by ACID ,which is conventional processing and BASE which is used by BIG data, ACID supports only structured data whereas BASE supports all kinds of data like structured ,semi-structured . Developer according to its suitability chose the kind according to the application and the advantages and disadvantages required.

VII. Data Management

Big data and cloud both have their own features though cloud should always focus on aspects like scalability, elasticity, fault tolerance. Traditional database systems are meant for only database storage frequent updating are not very well supported by them. Therefore here cloud wins as cloud provides all the features like scalability, elasticity and thus the data is frequently updated anytime. As the data size keeps on increasing there has to be an efficient mechanism to handle Data Management efficiently. Therefore following are some procedures which help in cloud computing of Big Data.

A. Transaction management

Partition Tolerance: when there is network partition the system should behave effectively and thus tolerate the partition or exchange of data. *Consistency*: Since the network is partitioned consistency of data should be there so that while communication or transaction the data does not gets distorted. *Availability*: as the network is partitioned, the data has to be consistent to be available. Hence data has to be protected until the transaction between nodes is done. Therefore high availability is not possible as the data has to beconsistent. When the data is partitioned it is very difficult to maintain atomicity and consistency.

B. NoSQL

Recently there has been enormous increase in data, be it structured or semi structured but semi-structured data is found more because this kind of data is easily created .Not Only SQL (NoSQL) follows non-relational data storage system.

Relational data storage systems require join operations which are costly to create and manage whereas NoSQL does not follow join operation of fixed table schema. NoSQL works on CAP theorem.

VIII. Impact of Big Data on Hybrid Cloud

Cloud and big data both co-relate with each other .Cloud based data model helps where the data is Scalable .Public and Private cloud, combination of both help in storage of different kind of data .Cloud computing is efficient for small and medium organizations .public cloud is mostly used when there is no major security issue whereas private cloud is used when the data is sensitive and thus the privacy of data has to be emphasized on. When the data is non-relational, users have to use techniques like for key-value databases we use Map Reduce .when Map Reduce is to be used then we go for Hadoop. When Document based databases are there then we use MongoDB. For Node based data we use Neo4j or OrientDB. Therefore according to the need of the application or organization we use the best suited Cloud Data model, also thus implementing these structures requires one to check on parameters like privacy, integrity and security support. Deployment of best suited Data model for the application to check risk, requirement. Analysis of the kind of data is very important which results in suitable selection and division of data in private or public cloud. so that the sensitive data can be categorized accordingly and effectively

IX. Conclusion

To demonstrate the effectiveness of using cloud computing for Big Data an efficient technique to emphasize the fact of data or image privacy in cloud for big data. In the technique we divide the image in certain number of blocks and then shuffle the blocks to form a “jig-saw puzzle” kind of structure. Instead of shuffling and working on pixels we focus more on blocks, which is efficient also and speeds up the process. Further one-to-one mapping is done. This further makes the image complex. This technique results in secure image further by reversing the process we get the image back which is stored in private cloud. Therefore we analyse the aspects of cloud computing and big data, which is highly needed these days to store huge amount of data. though cloud computing is costly for big organizations as it follows pay-per-use model still cloud computing is a long way to go and thus diminishing negative factors of security and privacy we can handle huge amounts of data.

References

1. L. Zhang, C. Wu, Z. Li, C. Guo, M. Chen, and F. C. Lau, "Moving big data to the cloud: An online cost-minimizing approach," IEEE Journal on Selected Areas in Communications, 2013.
2. D. Chen and H. Zhao, "Data security and privacy protection issues in cloud computing," in Computer Science and Electronics Engineering (ICCSEE), 2012 International Conference on, 2012.
3. S. Yu, C. Wang, K. Ren, and W. Lou, "Achieving secure, scalable, and fine-grained data access control in cloud computing," in INFOCOM, 2010 Proceedings IEEE, 2010.
4. J. Li, C. Jia, J. Li, and Z. Liu, "Novel framework for outsourcing and sharing searchable encrypted data on hybrid cloud," in Intelligent Networking and Collaborative Systems (INCoS), 2012 4th International Conference on. Springer, 2012.
5. T. Jung, X.-Y. Li, Z. Wan, and M. Wan, "Privacy preserving cloud data access with multi-authorities," in IEEE INFOCOM, 2013.
6. Y. Wang, K.-W. Wong, X. Liao, and G. Chen, "A new chaos-based fast image encryption algorithm," Applied soft computing, 2011.
7. F. Sufi, F. Han, I. Khalil, and J. Hu, "A chaos-based encryption technique to protect ecg packets for time critical telecardiology applications," Security and Communication Networks, 2011.
8. X. Wang and L. Teng, "An image blocks encryption algorithm based on spatiotemporal chaos," Nonlinear Dynamics, 2012.
9. G. Zhang and Q. Liu, "A novel image encryption method based on total shuffling scheme," Optics Communications, 2011.
10. X. Huang and X. Du, "Ensuring data privacy by hybrid cloud," in IEEE ICC, 2013.
11. E. Demaine and M. Demaine, "Jigsaw puzzles, edge matching, and polyomino packing: Connections and complexity," Graphs and Combinatorics, vol. 23, 2007.
12. T.Cho, S.Avidan, and W.Freeman, "Aprobabilisticimagejigsaw puzzle solver," in Computer Visionand Pattern Recognition (CVPR), 2010IEEE Conference on, 2010.

Corresponding Author:

E.Vijayan*,

Email: *vijayvsi81@gmail.com*