Survey on Privacy Preservation of Big Data via Hybrid Cloud

Sarika S. Chaudhari, Apurva Jagdhan, Mrunal S. Jagtap, Tejasweeni T. Jawalkar
Dept.computer engineering

Abstract: Nowadays the amount of data is being produced exponentially with the rapid development of electronic technology and communication, which makes it hard to cost-effectively store and manage these big data. Cloud computing, a new business model, is considered as one of most attractive solutions for big data, and provides the advantage of reduced cost through sharing of computing and storage resources. However, the growing concerns in term of the privacy of data stored in public cloud have slowed down the adoption of cloud computing for big data because sensitive information may be contained among the big data or the data owner themselves do not want any other people to scan their data. Since the data volume is huge and mobile devices are widely used, the traditional cryptographic approaches are not suitable for big data.

Keywords: Big data, hybrid cloud, big data privacy, encryption & decryption algorithm.

I. Introduction

Existing work in cloud computing has considered the storage on a private cloud in order to make the data storage secure. In this system we use the hybrid cloud where concept of public cloud is used for large data storage and private cloud is used for the small information storage. Existing work has proposed hybrid cloud for image data, but we propose the algorithms for the image data as well as we extend our work for the text data with existing image data.

In this project work, our aim is to achieve the image data privacy using hybrid cloud. But the drawback of existing system is, it takes much amount of time for the communication between the private and public cloud. So our first aim is to reduce the communication delay between the private and public cloud. Secondly to reduce the overhead caused by the creation and shuffling of images block using complex algorithm with implementation of simple algorithm,. Third this is to reduce the amount of data stored on the private cloud. Also we are extending our work for the encryption of data stored on the private cloud. This is the major contribution we added to the project.

II. Literature Survey

[1] In this paper, we propose an efficient scheme for image data, which has much more volume than text data. We evaluate our scheme in real networks (including Amazon EC2), and our experimental results on image show that our scheme achieves privacy. [2] This work studies timely, cost-minimizing upload of massive, dynamically-generated, geo-dispersed data into the cloud, for processing using a Map Reduce-like framework. Targeting at a cloud encompassing disparate data centers, we model a cost-minimizing data migration problem, and propose two online algorithms: an online lazy migration (OLM) algorithm and a randomizedxed horizon control (RFHC) algorithm, for optimizing at any given time the choice of the data center for data aggregation and processing, as well as the routes for transmitting data there.[3] This paper addresses this challenging open issue by, on one hand, defining and enforcing access policies based on data attributes, and, on the other hand, allowing the data owner to delegate most of the computation tasks involved in fine grained data access control to untrusted cloud servers without disclosing the underlying data contents. We achieve this goal by exploiting and uniquely combining techniques of attribute-based encryption (ABE), proxy re-encryption, and lazy re-encryption. Our proposed scheme also has salient properties of user access privilege confidentiality and user secret key accountability. Extensive analysis shows that our proposed scheme is highly efficient and provably secure under existing security models.[4] The experimental results indicate that the proposed algorithm has a satisfactory security level with a low computational complexity, compared to the two round encryption schemes, which renders it a good candidate for real-time secure image transmission applications. [5] In this paper, we analyse the security weakness of the proposal. The main problem of the original scheme is that the generated key stream remains unchanged for encrypting every image. Based on the aws, we demonstrate a chosen plaintext attack for revealing the equivalent keys with only 6 pairs of plaintext/ cipher text used. Finally, experimental results show the validity of our attack. [6] In this paper, a new scheduling model is proposed for workflow applications, which uses the hybrid cloud architecture in order to maintain privacy of the users' sensitive and private tasks and information. The proposed method schedules the sensitive tasks on the private cloud, which is the property and under control of the organization. If a user defined deadline is required to be

met, the scheduler is able to schedule other tasks of the workflow on pay-per-use resources of the public cloud. In addition, the scheduler tries to find the schedule map with minimum cost. Experimental results on both scientific and randomly generated workflows show higher performance in the cost and success rate than the comparing methods. [7] In this paper, we design and implement a privacy-aware framework to address data privacy challenges by supporting sensitive data segregation on hybrid clouds. Specifically, we model data sensitivity in a comprehensive and dynamic manner using a set of tagging mechanisms, which include a coarse-grained _le level tagging, a _ne-grained line level tagging, temporal and spatial tagging. The framework can also process data dynamically generated on the- y. We demonstrate the effectiveness of this framework using a big data application, and the experimental results show that the privacy-aware framework successfully enables data sensitivity protection while providing good performance. [8] The introduction of numerous clouds based services and geographically dispersed cloud service providers, sensitive information of different entities are normally stored in remote servers and locations with the possibilities of being exposed to unwanted parties in situations where the cloud servers storing those information are compromised. If security is not robust and consistent, the flexibility and advantages that cloud computing has to offer will have little credibility. This paper presents a review on the cloud computing concepts as well as security issues inherent within the context of cloud computing and cloud infrastructure. [9] we propose a new advanced ensemble pruning method, Hybrid Consensus Pruning(HCP), which is the first pruning algorithm that employs a fast consensus function to combine several classier classes into one scheme. To test the effectiveness of the HCP method, we conducted experiments comparing its performance with Ensemble Pruning via Individual Contribution ordering (EPIC), Directed Hill Climbing Ensemble Pruning (DHCEP) and K-Means Pruning approaches for pruning very large ensemble classifiers for malware detection. The results of the experiments show that HCP achieved better results by producing better ensemble classifiers as compared to those created by EPIC, DHCEP and K-Means Pruning. [10] It is able to make decisions about scheduling sensitive tasks on private cloud and uses public cloud's resources for non-sensitive tasks, such that the make span is minimized, while the budget limitation imposed by the user is satised. Experimental results show that the proposed method guarantees the execution of sensitive tasks on private cloud while achieving at least 7 percent lower make span and higher success rate in comparison to similar existing techniques.

III. Major Constraints

A. Private cloud

A private cloud is established for a specific organization and limits the access to it.It offers increased security because of its private nature.

B. Public cloud

It is the traditional cloud computing. As the name indicates it can be accessed by any subscriber who has an internet connection and access to the cloud space. The public cloud may be less secure because of its openness.

C. Hybrid cloud

It is a combination of at least two clouds, where the clouds can be public, private or community. When there is a public and another private cloud combination, then the critical activities are performed using private cloud and non-critical activities are performed using public cloud. Hybrid cloud for image and text data storage. Also we are enhancing file data.

IV. Equations

Fme: functions used
1. Upload ()
2. Download ()
3. Getinfo ()
4. Decryptimage ()
5. Encryptimage ()

Formulae:
OPix= (EPix * nBlocks) / (M * nBlocks2+g)
Success Conditions: we can achieve the image data privacy using hybrid cloud. Failure

Conditions: If the delay between the private and public cloud is more than the system fails to achieve the image data

V. Methodologies of Problem Solving and Efficiency Issues

A. Encryption

It means convert the original form of data into the cipher text form cipher text is the unreadable form of data.

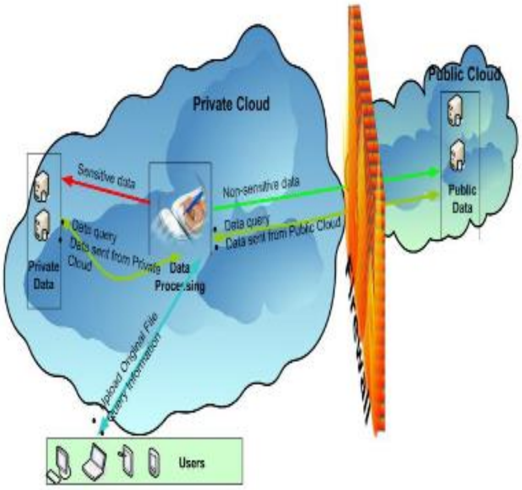
B. Decryption

It means convert the cipher text into the original form of text.

C. Efficiency Issues

All functions on totally depends on the system configuration.

VI. Architectural Design of Proposed System



The original data come from private cloud, and are processed on servers within private cloud. If there are no sensitive data, the original data may be sent to public cloud directly. Otherwise, the original data will be processed to make no sensitive data leaked out. After being processed, most data are sent to public cloud, and small amounts of sensitive data are kept in private cloud. When a user queries the data, both private cloud and public cloud will be contacted to provide the complete query result. Consider an untrusted public cloud who are curious and may intend to browse users' data. The public cloud has full control of its hardware, software, and network

VII. Conclusion

We can conclude with points such as our project is reduced time for communication between public and private a cloud. We decreased the data load on private cloud. and try to balance the data on public cloud. We provide the security for the image data stored on public cloud. This project is useful for image, text, file data security and storage

VIII. Refrences

[1]. Author Name "Xueli Huang and Xiaojiang Du" 2014 IEEE INFOCOM Workshops: 2014 IEEE INFOCOM Workshop on Security and Privacy in Big Data. "Achieving Big Data Privacy via Hybrid Cloud"

[2]. Author Name: "L. Zhang, C. Wu, Z. Li, C. Guo, M. Chen, and F. C. Lau" 2013 ,IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS "Moving big data to the cloud: An online cost-minimizing

[3]. Author Name:"S. Yu, C. Wang, K. Ren, and W. Lou"INFOCOM, 2010 Proceedings IEEE, 2010 "Achieving secure, scalable, and ne-grained data access control in cloud computing"

[4]. Author Name:"F. Su , F. Han, I. Khalil, and J. Hu" Security and Communication Networks,2011 "A chaos-based encryption technique to protect ecg packets for time critical telecardiology applications"

[5]. Author Name:"X. Wang and L. Teng" 2012 "An image blocks encryption algorithm based on spatiotemporal chaos"

[6]. Author Name:"G. Zhang and Q. Liu" 2011 "A novel image encryption method based on total shuing scheme"

[7]. Author Name:"J. Li, C. Jia, J. Li, and Z. Liu" 2012 4th International Conference "Novel framework for outsourcing and sharing searchable encrypted data on hybrid cloud"

[8]. Author Name:"D. Chen and H. Zhao" 2012 "Data security and privacy protection issues in cloud computing"

[9]. Author Name:"Jemal Abawajy, Senior Member, IEEE, Morshed Chowdhury and Andrei Kelarev" OCTOBER 2015 "Hybrid Consensus Pruning of Ensemble Classifiers for Big Data Malware Detection"

[10]. Author Name:"Amin Rezaeian, Hamid Abrishami, Saeid Abrishami and Mahmoud Naghibzadeh" 2016 IEEE International Conference on Cloud Engineering "A Budget Constrained Scheduling Algorithm for Hybrid Cloud Computing Systems Under Data Privacy"