Multi- Level Data Security Model for Big Data on Public Cloud: A New Model

Prof. Jay Dave
Department of Computer Engineering, RK University, Rajkot-
Email: jay.dave4u@gmail.com
Dr. Dhaval Vyas
Dean, M.Phil. Program, C.U. Shah University, Wadhvan City
Email: cudeanmphil@gmail.com

ABSTRACT

With the advent of cloud computing the big data has emerged as a very crucial technology. The certain type of cloud provides the consumers with the free services like storage, computational power etc. This paper is intended to make use of infrastructure as a service where the storage service from the public cloud providers is going to leveraged by an individual or organization. The paper will emphasize the model which can be used by anyone without any cost. They can store the confidential data without any type of security issue, as the data will be altered in such a way that it cannot be understood by the intruder if any. Not only that but the user can retrieve back the original data within no time. The proposed security model is going to effectively and efficiently provide a robust security while data is on cloud infrastructure as well as when data is getting migrated towards cloud infrastructure or vice versa.

Keywords - **Cloud, Big data, Data Security.**

Date of Submission: April 15, 2018 Date of Acceptance: May 18, 2018

I. INTRODUCTION

Cloud computing has emerged like anything in last decade and this has also increased the number of cloud service consumers. One of such service which has shown an immense level of growth in terms of usage is infrastructure as a service. People are using cloud storage to store their vital data like anything. From the end user's perspective, the biggest concern while using the cloud storage is the security as cloud's infrastructure service provides the storage on virtual basis. The biggest threat might arise when the data is getting migrated to the cloud or vice versa. During such type of transition if man in the middle attack is conducted or some intruder is trying to theft the data. Apart from this data can be hacked or leaked from the clod as well.in order to prevent such type of data loss or tempering the proposed model can be one of the most efficient and cost-effective solution.

1. Cloud Models

There exist mainly three cloud models private cloud [1], public cloud [1] and hybrid cloud [1] where in public cloud is the offering which might have less robust security mechanism as compared to private cloud and that is the solo reason why people act reluctant when it comes to use public loud to store the important information. The paper will be discussing couple of scenarios where either public or the private model could be leveraged to store the vital information of an individual or an organization.

2. Cloud Services

The cloud offers the consumers with mainly three types of services software as a service (SaaS), infrastructure as a service(IaaS), and plate form as a service(PaaS). In the paper area of concentration would be on Infrastructure as a service where the cloud consumers are going to make use of the storage utility provide by the various cloud service providers wither on the computational basis or totally free.

3. Issues with cloud computing

There are several issues with the cloud infrastructure like migration issues, heterogeneity, scalability, privacy, energy consumption and security [6]. In this paper, the security of data on the cloud and during the transition in discussed. The paper is highlighting a model which can provide significant security to the important data for the public cloud or private cloud.

There exist a very few models in the market which provide such type of services for consumers, but the concern is they are either very so or they are not commercialize. There are number of ways using which data can be made secure for the cloud storage like encryption, steganography and many more. There exist number of techniques in encryption [9] [10] like AES [12], DES [8] [12], RSA [12], blowfish [12] or hash function [12] which can be used to make the data more secure. The proposed model can work with the public cloud infrastructure as well as private cloud infrastructure. The first approach will cost absolutely nothing to the end user as it is totally providing at no cost because the public cloud obviously is made freely available to all the users with a little constraint of the space availability. Whereas the other approach is more scalable one where there is no constraint on the number of users as well as the amount of space that can be consumed by an individual or organization. The second approach is more suitable for the commercialization of the model where individuals must pay a little subscription amount which could be leveraged to prepare SLA and bear the expenditure. In either of the approach the consumers may use the model on the go using the portable devices like laptops or smart phones.

II. RELATED WORK

There exist a very few models in the market which provide such type of services for consumers, but the concern is they are either very so or they are not commercialize. There are number of ways using which data can be made secure [2] [3] [7] for the cloud storage like encryption, steganography and many more. There exist number of techniques in encryption like AES [8], DES [8] [10] [12], RSA [12], blowfish [12] or hash function [12] which can be used to make the data more secure. The proposed model can work with the public cloud infrastructure as well as private cloud infrastructure. The first approach will cost absolutely nothing to the end user as it is totally providing at no cost because the public cloud obviously is made freely available to all the users with a little constraint of the space availability. Whereas the other approach is more scalable one where there is no constraint on the number of users as well as the amount of space that can be consumed by an individual or organization. The second approach is more suitable for the commercialization of the model where individuals must pay a little subscription amount which could be leveraged to prepare SLA and bear the expenditure. In either of the approach the consumers may use the model on the go using the portable devices like laptops or smart phones.

One such model is V-GRT model [1] which relies upon RSA algorithm to make the data secure.in the model each of the potential users are provide with the unique user name and the password. The RSA algorithm is used to encrypt the password only. The cloud service providers need to decrypt the same and sends the login credentials along with the extra parameters required to the third party which is also termed as secured vendor. The trusted third party is also given with the credentials to access the cloud services. Now the user will be selecting the encryption algorithm and sends the data to the trusted third party which in turn will be managing the data on the cloud environment. The mentioned approach has the constant engagement of third party which will cause a lot of data transfer between all the parties and it may result in the significant amount of delay as far as the entire process is concern. There exists the system like Cumulus [4] and other [11] which is again a model for storing data in a secured way in cloud environment.

In this section the various techniques of enforcing encryption on the data has been discussed along with the time it takes to enforce such type of encryption technique. As it has been mentioned applying encryption is the crucial part of the model and it must be examined prior to selection of the appropriate algorithm.

For the experimental purpose with the sample data of chunk size of respectively 15KB, 30KB, 45 KB and 60 KB, each AES [10], DES [8] [10], And Blowfish [12] algorithms were applied to get an insight about the time it takes to enforce the encryption. Based on experimental results, it has been observed that the Blowfish and AES algorithms are outperforming. As it is clear from the results shown in the table that the mentioned algorithm takes little time compare to others.

Input (Chunk size)	**AES**	**DES**	**Blowfish**
15KB	2.2	4.9	2
30KB	2.9	6.1	2.6
45KB	3.3	8.25	3
60KB	3.5	9.75	3.2

Table1: Comparison of processing time of various algorithms in Milli-Seconds.

III. PROPOSED SYSTEM

The proposed system consists of the number of users where in each of them can leverage the multi-level data security model to store their credentials or data which is in the structured format. Users who are having the access to the proposed model can clean the data by eliminating the redundancy or the duplication which might be in existence. Once the first step is executed, the user will be in position to retrieve the unique size chunk of the data again which could be decided by the total size of the entire input file. The outcome of the second step would be number of chunks of the main input file. Now these chunks are ready to be encrypted [2] by the AES algorithm which will be creating the number of encrypted files. Observe that the number of encrypted files will be exactly equal to the number of obtained chunks of data. The next step is to use the steganography technique to hide the encrypted files behind the sample images [5]. The sample images could be in JPEG or PNG format. The only constraint is that the size of the sample image must be lesser then 1 MB, as it is going to incorporate the entire encrypted file behind it. Although doing the process of steganography [5] is going to enhance the size of the sample image slightly as the ample care is taken during the process that the size doesn't grow drastically. By doing this the intruder cannot detect that an image is containing the encrypted file behind it. In case a person with malicious intent cable to get the image and come to know that there might be some data behind it then too it is impossible for him/her to get the original data. As the data he/she might able to get from image is an encrypted file of one chunk. Therefore, to get the original data by attacker or intruder is next impossible task to achieve. Hence this model is effective that much it can enhance the security of the personal or organizational data to a level that data becomes safe during the transition towards the cloud storage, in the cloud storage or the transition back to local systems.

The diagram below best illustrates the tasks which are going to get executed in the proposed model which will make data safe. Not only that it can provide the feature to get back the original data intact from the cloud by performing the exact reverse scenario which is explained. The proposed model can perform its task on the go which might prove the model extremely mobile and effective to use.

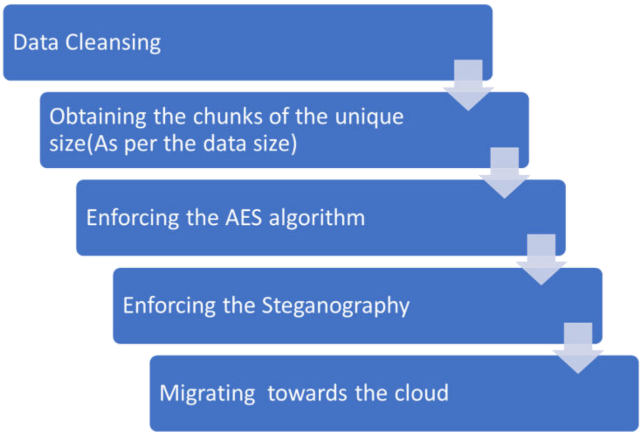
Data Cleansing
Obtaining the chunks of the unique size(As per the data size)
Enforcing the AES algorithm
Enforcing the Steganography
Migrating towards the cloud

Fig.1: Steps of proposed model.

IV. PROPOSED METHODOLOGY

Securing and transferring data to the public cloud:

- Step1: Start
- Step2: Take input File1.
- Step3: Remove the duplication.
- Step4: Obtain the equally sized segments of the data.
- Step5Take number of image as per the number of file
- Step6: Apply Encryption algorithm on the data obtain encrypted files.
- Step7: reduce the size of the encrypted files.
- Step8: Hide the encrypted files behind images using the steganography technique.
- Step9: Migrate the images towards the cloud.
- Step10: Stop.

Retrieving the original data from the public cloud:

- Step1: Start
- Step2: Download the images from the cloud.
- Step3: Get the hidden data in form of encrypted files from the image.
- Step4: Decrypt the files to obtain the original segmented files back.
- Step5: Merge the segments to get back the original data file.
- Step6: Stop.

V. IMPLEMENTATION

The entire coding of the model is done in the python where a user can use two major python files namely encode_file.py and Decode_file.py. These two files are coded in a way that accept the sample input data in this very example the sample data is retrieved form data.gov.in for research purpose, but in real life it could be any structured data which the belonging of might be an individual or organization.

Here the input data as mentioned is the sample data which has be taken form data.gov.in which is about the statistics from various states of India. The size of the sample data in this case is in Megabytes for the demonstration purpose, but the real data could be in giga bytes or tera bytes as well.

The encode_file.py is getting in the input data and checks for the redundancy by comparing each column of the respective rows with the remaining rows of the data. If any of the rows of data is found exactly like some other rows, it will be discarded immediately. The objective of the first step is to eliminate unnecessary data from the dataset Therefore, it will immediately shrink the data size. Now based on the size of the data user can have a provision to choose the chunk size or system will take care about generating the equal size chunks of the input file. The system will be generating three chunks in this case namely encoded_chunk_0.csv, encoded_chunk_1.csv and encoded_chunk_2.csv files. Which further will be encrypted using AES algorithm results in generation of three files again namely encoded_chunk_0.csv, encoded_chunk_1.csv, and encoded_chunk_2.csv which are basically encrypted files but not only that they also compressed so that the size of the files becomes minimal as it still need to be hidden in a way behind an image such that the size of an image should not be increased exponentially. As mentioned now three images will be taken as an input as there exists three encrypted files. In the example three images are lion.jpg, tiger.jpg and dog jpg. The mentioned three images will now go under the steganography process where they will be producing lion.enc.png, tiger.enc.png and dog.enc.png files respectively which are nothing but the images containing the encrypted files of the chunks of the original data.

The final step here is to store the data on the cloud and preferably the public cloud so here the drop box utility is leveraged for the experimental purpose, in the commercial version it has been planned to replace the drop box utility with the azure architecture or any other preferred cloud storage services. When so ever the user need to get back the data he/she has stored in the cloud, the second portion of the implemented algorithm takes over by downloading all images form the cloud, getting the data encrypted data from images, enforcing decryption algorithm, obtaining the chunks of the data and finally merging the.

VI. EXPERIMENTAL RESULTS

the results are obtained for different sized sample data and for two different scenarios. Senario-1 shows the process of securing the data and transferring it to the cloud environment. Whereas scenario 2 is the exact reciprocal to the scenario-1. The summary of the results is, it takes more time to get back the original data from the public cloud. The reason behind the caused latency is that time it takes to download the images in the local system is quite more as compared to upload those images on the cloud environment. The second vital observation from the experimental result is that as the data size grows the time it takes to prepare the data in the secured way and migrate the same data on the cloud environment is also getting increased.

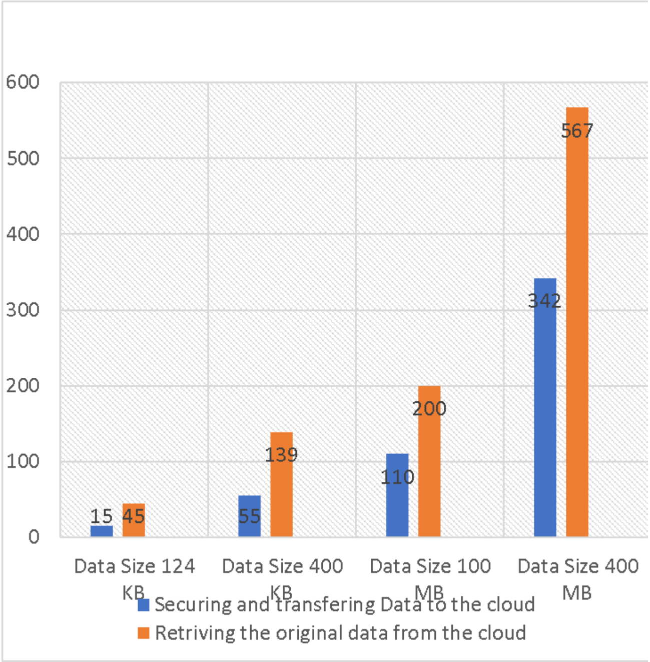

Fig.2: The time in seconds to make data secure and transfer to the cloud and retrieving back the original data from cloud.

VII. Conclusion and Future Enhancement

In this paper, the multi-level data security model is proposed and the implementation of the same is done with sample input data to show appropriate results. As indicated through the experimental results the model is effectively and efficiently securing the data, not only that it also takes care about the data during the transition phase.

In the future the plan is to design the system in such a way that it can take in huge amount of data which could be in unstructured format as well.

References

Journal Papers:

[1] M. Thamizhselvan, R. Raghuraman, S. Gershon Manoj, P. Victer Paul “DATA SECURITY MODEL FOR CLOUD COMPUTING USING V -GRT METHODOLOGY.” Published in IEEE.

[2] Mortada A.Aman and Egemen K¸Cetinkaya “Towards Cloud Security Improvement with Encryption Intensity Selection” Published in IEEE.

[3] Jingwei Li, Jin Li, Dongqing Xie and Zhang Cai “Secure Auditing and Deduplicating Data in Cloud.” Published in IEEE.

[4] Hend Gedawy, Sannan Tariq, Abderrahmen Mtibaa, Khaled Harras School of Computer Science, Carnegie Mellon University, “Cumulus: ADistributed and Flexible Computing Testbed for Edge Cloud Computational Offloading” published in IEEE-2016.

[5] Vinay kumar pant, Jyoti Prakash , Amit Asthana “Three Step Data Security Model for Cloud Computing based on RSA and Steganography Techniques ”, 2015 International Conference on Green Computing and Internet of Things (ICGCIoT).

[6] Meena Kumari , Rajender Nath “Security Concerns and Countermeasures in Cloud Computing Paradigm”, 2015 Fifth International Conference on Advanced Computing & Communication Technologies.

[7] Shakeeba S. Khan, Prof.R.R. Tuteja “Security in Cloud Computing using Cryptographic Algorithms” International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 3, Issue 1, January 2015.

[8] Neha Jain and Gurpreet Kaur ‘Implementing DES Algorithm in Cloud for Data Security” VSRD International Journal of CS & IT Vol. 2 Issue 4, pp. 316-321, 2012

[9] G. Devi , M. Pramod Kumar, “Cloud Computing: A CRM Service Based on a Separate Encryption and Decryption using Blowfish algorithm” International Journal Of Computer Trends And Technology Volume 3 Issue 4, ISSN: 2231-2803, pp. 592-596,2012

[10] Gurpreet Singh, Supriya Kinger”Integrating AES, DES, and 3-DES Encryption Algorithms for Enhanced Data Security “International Journal of Scientific & Engineering Research, Volume 4, Issue 7, July-2013

[11] Arthur Rahumed, Henry C. H. Chen, Yang Tang, Patrick P. C. Lee, and John C. S. Lui “A Secure Cloud Backup System with Assured Deletion and Version Control” 1530-2016/11 IEEE DOI 10.1109/ICPPW.2011.17

[12] Jay Dave, Ashwin Raiyani *“The security perusal of big data in cloud computing environment”, Proceedings of RK University’s First International Conference on Research & Entrepreneurship (Jan. 5th& Jan. 6th, 20*

Books:

[1] W. Diffie and M.E. Hellman. New directions in cryptography. IEEE Transactions on Information Theory, 1976.

BIOGRAPHIES AND PHOTOGRAPHS

Prof. Jay Dave is having B.E. Degree in Computer Engineering from Saurashtra University and Master’s Degree from Bharti Vidyapeeth University, Pune. He is research scholar from C.U. Shah University, Wadhvan City.

Dr. Dhaval Vyas working as dean at C.U. Shah University has more then 11 years of experience in the academics at C.U. Shah University, Wadhvan City, with expertise in various domains like embedded systems an data security in the cloud environment.