Enhancing Data Security in Cloud using Shuffling and Distribution Algorithm

S.Kaaviya
Department of IT
Vel High Tech Engg Coll

P.L.Revathi
Department of IT
Vel High Tech Engg Coll

R.Nithya
Department of IT
Velammal Engg Coll

ABSTRACT

Cloud Computing is a technology which is used to provide service over internet by means of computing resources(hardware and software).This paper describes conversion of data into images and the numbered images are shuffled using fisher Yates shuffling algorithm which randomly shuffles the image until nothing is left. However security is the main issue in cloud storage. It is then compressed using compression technique in order to reduce the storage space. Now the shuffled image has to be stored in various parts of the cloud server using the distribution algorithm. Then the key should be allotted to decrypt the data from the server. Finally we prove that using this technique can enhance the cloud storage security which increases the confidentiality and integrity.

Keywords

Cloud Storage, True Random number Generator Algorithm, Dispersion Algorithm.

1. INTRODUCTION

Nowadays Cloud has become the emerging trend to do all sorts of online activities. Cloud has the capability to satisfy various needs of the customer or user. Cloud computing uses the concept of virtualization. Cloud computing emerged from the topic grid computing. Cloud computing is the enhanced computing method of grid computing. During past days, Each and every node in the network had their own platform such linux platform,Windows platform and many other platforms .When some information is sent from one node to another node of different platforms through network this dint get supported which lead to the invention of distributed system where data can be exchanged to any nodes of any platform(windows,linux etc).Later this lead to the invention of grid computing which works behind the principle of pooled resources. This concept needs many server and it requires atleast one server to administer. To enchance and make the grid computing work in a better way cloud computing was introduced for the effective usage. Virtualization is the process of running multiple.

Virtual machines (VM) on one host using hypervisor. Server virtualization is the process of dividing the physical server into multiple server where each server can run on their own operating system. All the private data of the customers are stored in cloud. Customers cant manage their data directly or frequently control it. Inorder to maintain the data the cloud should provide the maximum security. So that the customers shall be confident about their data security. But there arises lot of security issues in cloud as the communication is via internet. The security issues in cloud are confidentiality, data integrity and managing the data.The major security challenge in cloud is that the user or owner of the data may not control his own data. Therefore safe guarding the data in cloud is the major issue Information Security is one of the most crucial aspects in computer networks. Moreover communication or storage of private data over internet need more security. Cryptographic technique is used to convert the plain text to image by splitting the text into blocks .An encryption method is used to encrypt the plain text to cipher text using a key. Decryption is the inverse technique of encryption where the cipher text is converted to plain text using the private key(known as symmetric or secret key) or public key(also known as asymmetric key) achieved by removing Shuffling is the act of randomizing the sequence of numbered images .Randomizing is the act of making some sequence in the form of random way. Eg. Picking up a random card from the deck of cards and randomizing the card. Shuffling is the process of picking up the numbered images until no more is left. Shuffling of the images will further increase the security. Compression is one of the spaces reducing technique which saves the space in hard disk. So that the complexity in the storage can be avoided. Image compression technique is the process of reducing the amount of data required to increase the effect of the digital image. Compression is mainly the redundancy. The two types of compression techniques are lossy and lossless compression technique.It mainly reduces the occurences of errors, reduces the storage requirements, reduces the cost of savings and increases the level of security. Distribution algorithms ia mainly used in allocation of resources where the resources are equally distributed to different nodes. Eg: Leader election.Leader election is the process of assigning an organizer as a single process of some task distributed among n number of nodes. Shuffling algorithm which is a combination of fisher-yates and Cloud security has been an important research focus to provide secure transmission of data from server to client through web browser. Consequently large amount of research work exist on the data security in cloud computing

2. RELATED WORK

In this paper we have proposed to provide the security in accessing the data from cloud server using encryption algorithms. We have applied both Shuffling and Distribution techniques. This work is mainly proposed to obtain maximum security to data stored in cloud and accessed from cloud. Here the text is converted to image format to avoid hacking of data from internet. In this method, the overall security for accessing the data from cloud server has been enhanced. The data accessing security have also achieved a great deal. This technique provides the maximum security in accessing the data from cloud server. It also assures the customers or users

of cloud to store and retrieve their personal data with ease and full security. It provides each customer to have direct contact with their data which has been in cloud server and they need not be worried about their lost or damage of their data from cloud server. It also provides integrity and confidentiality to the user about their data. To our knowledge the shuffling algorithms have been used to provide security for images or text. But we have used the Fisher Yates shuffling technique for shuffling images where these images were obtained from conversion of text to image technique. So we have used these shuffling techniques after converting a plain text to image. Some researches have increased security by implementing software in the data model architecture of cloud [1].

In paper[2] ,a third party audit is employed to look after the data security. But it is impossible that third party audit(TPA) is always trust worthy. In the above paper each and every block of text is assigned with the key and the data is retrieved by providing correct key .If the key given does not match then the data cannot be retrieved and it is tedious to remember the key of each every block. In one of the research papers [3] customers data is given security by the implementation of the digital signatures by using RSA algorithm. User uses internet as the source for the communication between the user and the cloud server so security level agreement has to be given for the protecting the data from stealers. As data security has become a big deal we have integrated the pseudo random number algorithm which randomize the numbered images and arrange it in a sequence. The random numbers can be picked out by the hardware used with the system.It can be generated based on the timers,processor speed etc and at last dispersion algorithm is used to disperse the data in the cloud for the security reason.However data stored in the server is in the encrypted form which is hard to retrieve as it is distributed.The files that are considered to be more important is provided with a private key so that the data can be retrieved by decrptying it.Thus security is provided to the data that is stored.

3. RANDOM NUMBER GENERATOR

3.1 FY Pseudo random number Generator Algorithm

This algorithm is derived from the algorithms Fisher Yates algorithm and pseudorandom number generator algorithm. This algorithm is the combination of Pseudorandom number generator algorithm and the Fisher Yates algorithm.At first the text which are divided into blocks is converted into images.Then the images are numbered according to the sequence.These sequence numbered images are randomized.

Step 1: Take all the sequentially marked images which the text block is converted into image

Step 2: A random numbered image is picked up by using the Pseudo random number generator in which a randomized numbered image is selected according to the hardware (processor speed, timers or clock used) used.

Step 3: Counting from the low end, strike out the kth number not yet struck out, and write it down elsewhere

Step 4: Repeat from the step 2 until all the numbers are picked out

Step 5: The numbers which are written in the third step is the permutation of the number given

3.2 Usage of Hardware Security Module

Hardware security module(HSM) can be used when the authentication by the server application is given in the form of digital signatures .It is also used to manage the transport data encryption.The main usage of this HSM is to make use of sensitive or important data in secured way and onboard secured storage.

4. SYSTEM ARCHITECTURE

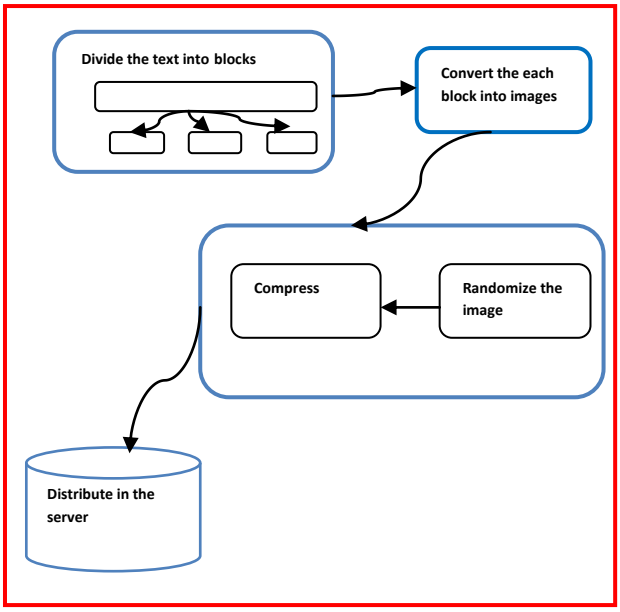

Figure 1.System architecture

At first the given test is divided into the blocks and it should be converted into the images.The text is transformed to image by using c# programming of .NET framework.Thus images of each text block is generated.The images are numbered in a sequence way such that they are randomized.The transformed data should be randomized using FY PseudoRandom number generator algorithm. A random numbered image is picked up by using the Pseudo random number generator in which a randomized numbered image is selected according to the hardware (processor speed, timers or clock used) used.This algorithm is used to randomize the images.Then the image is divided into 6 by 6 block pixels in which each pixel should be processed in a independent way such that luminance is more than chrominance.Then the compressed data has to be distributed in the server to enhance the security .Hardware security module is also used incase when the digital signatures are used.HSM system is used to take backup on keys that the users make use of .Symmetric keys are also handled by HSM.The physical and the logical protection is provided by the HSM.

5. COMPRESSION AND DISTRIBUTION ALGORITHM

5.1 Compression Algorithm

Compression technique is mainly used to reduce the space of storage and increases the capacity of the resources. The data or information which occupies more space is compressed using a compressing technique (i.e) Lossless compression technique. Then the compressed data can again be decompressed to obtain the original information inorder for future usage. This is mainly used to reduce the resources storage space and hence increase its productivity. There are two types of compression techniques. They are Lossy data compression algorithm and lossless data compression algorithm.In this section, we are going use the Lossless data compression technique where the data or information which is compressed to minimise its storage size does not undergo any loss of data or information. The lossless compression technique is highly secured. It works by finding the repeated patterns in a message and encoding those repeated patterns in an effective and efficient manner.It allows to construct the exact data or information which has been compressed.The algorithm here used for compression is Run-Length encoding(RLE).It is very simple Data Compression technique in which run length of data (i.e) in a sequence of data many data values are arranged in a consecutive order. It has a sequence which contains runs of repeated data in a consecutive manner. They are stored as a single data value instead of original run. This method is also used in many of the graphical file format inorder to compress many of the gray scale images.

For example:

Consider a sequence of characters displayed in a white screen.

AAAAABBBBAACCCCBBBDDDDAAAAABBBBB

If we apply Run-Length encoding data compression technique to the above sequence, we get the following:

5A4B2A4C3B4D5A5B

This is to interpret as five A's, four B's, two A's, four C's, three B's, four D's, five A's and five B's.

The Run-Length encoding represent the original 32 in only 16.This method can also be used in multiple ways. Hence the data compression is carried out in a secure and lossless manner using this run-length encoding compression technique.

5.2 Distribution of Data

The distribution algorithm is designed to run in the hardware of the computer within the processors interconnected. This is mainly used in the concept of distributed computing. It is mainly a derived type of parallel algorithm. In the parallel algorithm, the information's or data are executed simultaneously without any intervention. In this technique, more number of data can be processed parallels in less amount of time. Likewise in the distribution algorithm any number of data. This algorithms main aim is to maintain the reliability and resolve the failure and also coordinating the overall behaviour of the process. It also monitors the system when communication links becomes unreliable. data can be run in different processors simultaneously. The different parts of the algorithm can be executed concurrently using this distribution technique.

6. FUTURE WORK

The current system implementation is being evaluated for its enhancement of security in cloud architecture. Cloud computing is not fully secured and it still need to explored and developed. In this current work we are claiming that the security in cloud has been enhanced based on distribution and compression algorithms. Security is the most threat to all the vendors, researches and IT professionals. Different methods have been used to enhance and improve the security in cloud. But still the fruitful is not found. In our future works security will be enhanced in cloud computing.

7. CONCLUSION

This paper is proposed in order to increase the security by dividing the text into n no of blocks with the available data and convert those blocks into images .The images are randomized to enhance the security. As the images occupy more space in the memory which leads to the decrease in the performance of the system so it is compressed. Then the data is distributed in various server to increase the security.

8. ACKNOWLEDGMENTS

Our Sincere thanks to our Head of Department and the faculties of Information Technology to bring out this paper in a successful way.

9. REFERENCES

[1] Mladen A. Vouch, ―Cloud Computing Issues, Research and Implementations‖, Journal of Computing and Information Technology - CIT 16, 2008, 4, 235–246

[2]Jinpeng et al, ―Managing Security of Virtual Machine Images in a Cloud Environment‖, CCSW, 2009, Chicago, USA

[3] Miranda & Siani, ―A Client-Based Privacy Manager for Cloud computing, COMSWARE'09, 2009, Dublin, Ireland

[4]Center Of The Protection Of National Infrastructure CPNI by Deloitte*"Information Security Briefing 0112010 Cloud Computing"*, p.71Published March 2010.

[5] Ian Foster, Yong Zhao, loan Raicu, Shiyong Lu," *Cloud Computing and Grid Computing 360-Degree Compared* " Grid Computing Environments Workshop, 2008. GCE '08 p.l0, published 16 Nov 2008.

[6] Cloud Security Alliance Guidance, *"Security Guidance For Critical Areas of Focus In Cloud Computing Vl.0",*

www. Cloud security alliance. orglguidance/csaguide.v1.0.pdf, published April 2009.

[7] Cloud Security Alliance Guidance," *Security Guidance For Critical Areas of Focus In Cloud Computing V2.1"*, www .cloud security alliance.orglguidance/csaguide.v2.1.pdf, published Dec 2009.

[8]C.C. Tan, Q. Liu, and J. Wu. Secure locking for untrusted clouds. In *cloud Computing (CLOUD), 2011 IEEE International Conference on*, pages 131–138. IEEE, 2011.

[9] J. Vaidya. A survey of privacy-preserving methods acrossvertically partitioned data. *Privacy-Preserving Data Mining*,pages 337–358, 2008.

[10] M. Van Dijk and A. Juels. On the impossibility of cryptography alone for privacy-preserving cloud computing .IACR eprint,305,2010.

[11] K. Liu, C. Giannella, and H. Kargupta. A survey of attack techniques on privacy-preserving data perturbation methods. *Privacy-Preserving Data Mining*, pages 359–381, 2008.

[12] S. Pearson, Y. Shen, and M. Mowbray. A privacy manager for cloud computing. *Cloud Computing*, pages 90–106, 2009.

[13] Adi Shamir. How to share a secret. *Commun. ACM*, 22:612–613, November 1979.

[14] C.C. Tan, Q. Liu, and J. Wu. Secure locking for untrusted clouds. In *Cloud Computing (CLOUD), 2011 IEEE International Conference on*, pages 131–138. IEEE, 2011.

[15]L. Gu, A. Vaynberg, B. Ford, Z. Shao and D. Costanzo, "CertiKOS: A Certified Kernel for Secure Cloud Computing," in Proceedings of the Second Asia-Pacific Workshop on Systems APSys, 2011.

[16] P. Mell and T. Grance, “Draft nist working definition of cloud computing,”Referenced on Jan. 23rd, 2010 Online at http://csrc.nist.gov/groups/SNS/cloud-computing/index.html, 2010.

[17] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, “Above the clouds: A berkeley view of cloud computing,” University ofCalifornia, Berkeley, Tech. Rep. UCB-EECS-2009-28, Feb 2009.

[18] S. Kamara and K. Lauter, “Cryptographic cloud storage,” in *Proceedings of Financial Cryptography: Workshop on Real-Life CryptographicProtocols and Standardization 2010*, January 2010.

[19] Cloud Security Alliance, “Security guidance for critical areas of focus in cloud computing,” 2009, http://www.cloudsecurityalliance.org.

[20] I. H. Witten, A. Moffat, and T. C. Bell, “Managing gigabytes: Compressing and indexing documents and images,” Morgan Kaufmann Publishing, San Francisco, May 1999.