Asian Journal of Information Technology 15 (16): 2937-2942, 2016
ISSN: 1682-3915
© Medwell Journals, 2016

Proxy Server for Secure Document Retrieval in Cloud Computing

[1]R. Kalaiselvi, [2]K. Kousalya and [3]S. Chandramathi
[1]Department of Computer Science and Engineering,
Kumaraguru College of Technology, Coimbatore, India
[2]Department of Computer Science and Engineering,
Kongu Engineering College, Perundurai, Tamil Nadu
[3]Department of Computer Science and Engineering,
Sri Krishna Institute of Technology, Karnataka, 560090 Bengaluru, India

Abstract: Cloud computing is one of the expeditious developing internet qwbased technology that aid users to resort to services by utilizing large pool of resources without installing any software and services available in a pay as-you-go manner. Adopting this technology increases rapidly because of its scalability, elasticity and low cost. As cloud encourages the users to experience its pay per use services, there have been wide privacy concerns in case of data usage. To ensure the security, cloud paradigm allows encrypted format for data storage. To achieve proper encryption and secured data forwarding, cloud can employ proxy server concept. This paper deals with creating a proxy server to view a replicated data document after proper authentication such that the original data in the main server is not accessible that avoids accidental changes and malicious user attacks. Additional security level is incorporated by data re-encryption technique before transmitting encrypted data which resides in server and it also facilitates erasure code for authentication and secured data forwarding. Only after authentication the user can view the desired information completely by receiving key for decrypting the received file. As the user can only get the re-encrypted copy of encrypted data stored by data owner, user cannot modify the original data in the server. The user has only the permission to read the data after the decryption by proper key. This experimental analysis proves that the idea of proxy server provides efficient secure data storage and retrieval.

Key words: Cloud computing, proxy server, proxy re-encryption, secured data forwarding decryption, India

INTRODUCTION

Cloud computing is a technology that has come to light recently. It enables the enterprise to reduce the physical storage, the rent required for the physical storage, the cost of the investment in software and/or software licenses for every employee and relocating each tasks at hand from local computers that has resorted to cloud computing providers such as IBM, Amazon, Yahoo, Google, Microsoft, etc. (Almulla and Yeun, 2010; Shaikh and Haider, 2011). Cloud Service Providers (CSP) are responsible for making the data available and usable and provides protected virtual environment for running applications. Cloud storage enables users to remotely save their data and experience on demand high quality cloud applications without the intervention of management of local hardware and software. Although, cloud computing is a blooming skill it has its own defects. Utilization of cloud computing is increasing vigorously because of its alluring characteristics. Regardless of its advantages there exist some of the greatest limitations like its security. To defeat security problems, several researchers proposed their algorithms and proved security levels of those techniques. Currently every organization uses the information shielding system as hospitals maintain the patients' medical details, banking system protects customers' details. For preserving technical and sensitive information and to prevent data loss and data hacking by the intruder, DLP (Data Loss Protection) system is used (Soofi and Khan, 2014). Data confidentiality is a security concern in cloud computing. Though many of the methods have been developed to overcome security issues, encryption is the most suitable method to attain data confidentiality in cloud environment (Amin *et al*., 2014). Insiders pose a high level of threats along with the rapid growth of the cloud ecosystem. Many insider technophiles can use their vast extent of knowledge on company security's weaknesses to breach

Corresponding Author: R. Kalaiselvi, Department of Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore, India

clearance and access privileged information. The holistic way to defend against a malicious insider working for the cloud provider is to identify the likely problem for the scenario and propose appropriate countermeasures to mitigate the problem (Kandias *et al*., 2011). Since, data are scattered in various machines and storage devices such as PCs, servers and various mobile devices, data security becomes notably severe in the cloud computing environment. As security plays a censorious role in the present era of computing, the utility, data confidentiality, safety mechanisms, avoiding malicious insiders' illegal operations and service hijacking, cloud server monitoring or tracing have become sub categories of data security (Sun *et al*., 2014). Protecting the outsourced sensitive data is a major challenge in cloud computing. Encryption is the widely used technique for data security in distributed environments. As encryption consumes more processor overhead many CSP offers only primitive encryption on vital databases like password. Encrypting the user's entire database is more expensive than physical storage of data in local. Few applications like searching data can execute operations on encrypted data but shows performance degradation as a result after the work done without decrypting the data (Tebaa *et al*., 2012). Usually, the ciphertext is converted into a plaintext using a decryption technique with a key after getting proper authentication and authorization. Authorization is an act of giving permission to access the data. In the other hand, authentication is checking whether requested user is genuine. Both authorization and authentication decide the user's roles in the cloud and promoting accounting, controlling resource usage and isolation. The obstacles and solutions are explored to provide an authentic cloud computing environment (Bhisikar and Sahu, 2013). An AES algorithm based technique is used to encrypt sensitive data before outsourcing, it to cloud using symmetric key with rotation (Prakash *et al*., 2014).

Key generation is a cryptographic feature which combines the current time stamp with the key itself ensuring that the cloud server contains the authorized data set from the data owner. Instead of one time encryption using single key, double encryption may protect sensitive data effectively where two keys are used for encryption and re-encryption. Random functions may also be used to generate a key. A key can be of 16, 32, 64, and 128-bits and so on. Key management deals with the creating, use, exchange, replacement and storage of keys. The identification of the cryptographic key management leads to a challenge related to architectural solutions that are used for cryptographic operations. Proxy re-encryption is one where the ciphertext is again encrypted and another key is provided to decrypt that ciphertext into plaintext. Two level security system can ensure a secure distributed system where as first level is design of the encryption scheme which supports the data owner for encoding operations over plain messages and second level for forwarding data after re-encryption that is integrated with a secure decentralized erasure code (Priyadharshini *et al*., 2013). Before outsourcing as data are partitioned into fragments and scattered geographically over servers to ensure data protection. A pipelined coding strategy is discussed for data partitioning and it leads to fast archival of data where neither data reliability nor storage overheads are compromised. As data retrieval involves data transmission over network, it is fortunate to have secure data transmission, otherwise secured data storage is meaningless. Collaborative Cloud Computing (CCC) is an efficient method in cloud that ensures secure data transmission. Additional level security can be incorporated by an advanced technique called harmony in collaborative cloud computing to promote the security in distributed environment and moreover data can be traced securely from different online trading platforms like flip kart, Amazon, etc. (Bhisikar and Sahu, 2013).

Existing system: Platform providers such as Google, Amazon and Microsoft, etc., attract organizations and individuals to outsource their data from physical to remote servers. Data confidentiality may be upturned as a serious concern when storing data in a third party's cloud system where common encryption schemes can protect data it may also limit the working of the storage system. Inbarani *et al.* (2013) formulates a distributed storage system that supports secure and robust data storage and its retrieval and also lets a user forward his data to the storage servers. Anuchart and Gong (2011) articulated a scheme which provides end-to-end encryption and ABE-based tokens to establish authorization so that the owners can take control over their data when it is stored in semi-untrusted cloud storage. The data contents are not only one referred when talking about data privacy. Since, the computation outsourcing is the most attractive part of the cloud computing, it is more likely that the users want to control the rights of data handling over cloud servers or users. This is because when sensitive data is shared with another user, risks would rise radically because the user might illegally inspect the data and get access to sensitive data. Hence, the access along with the operation should be controlled. Jung *et al.* (2015) present a semi-anonymous privilege control scheme anony control that deals with the data privacy and user identity privacy to limit the identity leakage achieving semi-anonymity.

People tend to be more concerned about their identity privacy as our identity also needs to be protected. Preferably, any client's personal information should not be known to authority or any server. Currently digital identity is provided for the users to access their services and this might bring some troublesomeness private and/or public clouds. To provide data security and mutual authentication asymmetric and traditional public key cryptography are being used. Identity-based cryptography has some attraction characteristics that seem to fit well the requirements of cloud computing. The key distribution and mutual authentication can be abridged by adopting Hierarchical Identity-Based Cryptography (HIBC) incorporated with federated identity management as proposed by Yan *et al*. (2009).

Hong proposed a Shared Authority based Privacy preserving Authentication protocol (SAPA) that does not compromise the data owner's information by checking the user's authentication and authorization (Purohit and Singh, 2013). Both authorization and authentication decide the user's roles in the cloud, promoting accounting, controlling resource usage and isolation. Pradnyesh *et al*. (2013) explored the obstacles and solutions to provide an authentic cloud computing environment. Establishing new privacy hindrances in cloud storage and addressing an issue where in the request itself cannot reveal the user's privacy even if it can obtain the access authority is one of the main contributions. Anonymous access request matching mechanism is used to achieve shared access authority and an authentication protocol is proposed to enhance a user's access request related privacy. Authentication process delivers login and password. But on the password can be hacked easily by the hacker on a public network so that there are techniques used such as key infrastructure, both symmetric and asymmetric algorithm. RSA is popular and a very strong algorithm that uses Kerberos authentication protocol but a disadvantage of this method is generating keys over head of keys. Kerberos is an authentication protocol used in many real life systems and it is security mechanism based threshold cryptography and Kerberos protocol (Santosh and Dudhani, 2015). Proxy re-encryption and cipher text-policy attribute based access control are some of the major concepts used. The advantage of this existing system is that the data owner can decide whether the user can access the system or not. Proxy re-encryption is where the ciphertext is again encrypted and another key is provided to decrypt that ciphertext to plaintext. Pothumani and Gosh (2015) introduces a secure distributed system which is designed when the re-encryption scheme that supports the encoding operations over encrypted messages as well as forwarding operations is integrated with a secure decentralized erasure code (Almulla and Yeun, 2013). Ciphertext-policy attribute-based access control where a user's private-key is associated with a set of attributes and an access policy over a defined attributes is specified by the ciphertext. As attributes satisfy the policy of the respective ciphertext, a user will be able to decrypt a ciphertext. Bethencourt *et al*. (2007) proposed a system for realizing complex access control on encrypted data where the encrypted data can be kept a secret even if the storage server is untrusted also being secure against collusion attacks (Inbarani *et al*., 2013).

Even though, data outsourced in encrypted format once the encrypted data been accessed by malicious user it can be decrypted by trial and error method. As many techniques used for secure data storage none worry about access of encrypted data by malicious users. Our proposed approach permits the user to access their data and can view the accessed data only for a specific time span. Proxy server which was created to access the data automatically gets destroyed after the specific time span in turn leads to more secure data storage and retrieval. As malicious user access the encrypted data they cannot try with trial and error method as time span is too short.

MATERIALS AND METHODS

Proposed system

System model: Data ownership; both proprietary of and responsibility for information, power as well as control, refers to data ownership. Data ownership does not just mean the creation, access, derive benefit from, modify, sell or remove data or package, it is having legal rights and overall control over a their data elements. Here data owners can create a login and they can upload their own files in the cloud storage.

Encryption: The process of encoding information or messages in a way such that only authorized users can be able to read, it is called encryption in the context of cryptography. Encryption does not hinder interception but withholds the content to the interceptor. The plaintext, the intended information to communicate is encrypted using any encryption algorithm, a ciphertext is generated that when only decrypted can be read. Encryption is mainly used to overthrow unauthorized access and theft protection. When the file is uploaded in the server, it is stored in the encrypted format and when it reaches the proxy server it is again re-encrypted. Key generation: A unique identification key is being generated for the file or a document which the other user request to view. The key (random number of letters, numbers, symbols, etc) is being provided to the user on his request if he is authenticated user in cloud server. Only after the unique key is given to the user he is able to view the file in the proxy server and cannot make any further modifications in the remote server. Once the key has been generated it is being (i.e., the key is copied) copied and pasted in the proxy server and the permission is given for the user to view the file which he requested.

Proxy re-encryption: A process where already encrypted data is re-encrypted using a public key for an unlimited number of times where in the private key of the data is not exposed. This is one of the advanced encryption model that works on virtual as well as the real systems and more efficiently on cloud systems. It is similar to the traditional symmetric or asymmetric (bi-directional and unidirectional varieties) encryption schemes. Proxy re-encryption schemes should not be confused with proxy signatures which is a whole other concept. The key holder generates a re-encryption key based on his secret key. The proxy uses this re-encryption key as input to the re-encryption function that is executed to translate ciphertext into plaintext.

Data forwarding: When the data sources are distributed in the cloud, constructing an erasure code for storage over a network poses as a problem. It is assumed that there are k <n sources generating the data and n storage nodes with finite memory. The decentralized erasure codes are optimally used and result in reduced storage, communication and computation cost.

A protocol based proxy re-encryption technique is being blended with a decentralized erasure code such that it leads to secure data storage distribution. With the use of the proxy re-encryption technique, a secure cloud storage system is ensured that provides secure data forwarding too in a decentralized structure. Storing the data in third party's cloud system may cause serious concern on data confidentiality. In order to provide full confidentiality for messages stored in remote server, the user can encrypt messages by any cryptographic method. As the data owner uploads the encrypted data that is partitioned and scatted over several servers by the cloud service provider. As many fragmentations like horizontal, vertical or hybrid fragmentation available, this analysis uses horizontal fragmentation for simplicity. Whenever the data owner or user requires the data, after receiving the request followed by proper authentication, a proxy server is created by the provider for concern user. User can access the proxy server only after proper authentication. Provider provides the data in encrypted format to the owner or user through proxy server. User should receive the decryption key from the data owner after owning proper membership. Data owner or user can view the data only for a specific amount of time that the proxy server itself gets destroyed automatically after the specified time span. Time span can be chosen by the owner or user during proxy server creation. This automation destruction of proxy server is to avoid the unwanted accidental access.

For authentication, a key of size 32-bit with a mixture of numbers, letters, special characters, etc., is generated by using a random function without repetition of any of those letters. For specification of time span duration to view the document a certain time format is followed. As the readability of persons varies it is required to offer the data for certain duration. Offering the proxy server for a constant duration may lead to unnecessary risk and time wastage. To avoid this, three option scheme: first, one time based which is for seeing the data only for few seconds, secondly option allows the user to view the data for certain period of time, third, date based where data can be viewed for a day. Once the time limit exceeds the data will be deleted automatically. Architecture and algorithm are given below

Algorithm; partitioning:

```
Start
 Input src, dest, sof
 src <- source path//Get source document path
 dest <- destination path//Set path to store documents after fragmentation
sof <- size of file//Analyze the size of document
 partition(sof)//split the document into two based on size
Store fragments
Stop
```

Figure 1 shows the concept of loading and retrieving the encrypted data securely on cloud server. After the storage of encrypted data by data owner, the documents are partitioned using the above mentioned algorithm and stored in different locations. For experimental purpose and simplicity the encrypted document is portioned into only two halves. When the user/owner raises request to access the document after appropriate authentication the partitioned encrypted documents are retrieved from different locations and documents are merged temporarily before transmitted to user. After receiving the decryption key from the data owner, authenticated user can view the document only for a specific amount of time. The partitioning and merging processes which ensure security and responsibility of service provider are hidden from data owner or user. As encrypted documents are loaded into cloud server by data owner the documents are partitioned and stored in different locations.

When the user/owner raises request to access the document after appropriate authentication the partitions of the documents are merged temporarily and transmitted

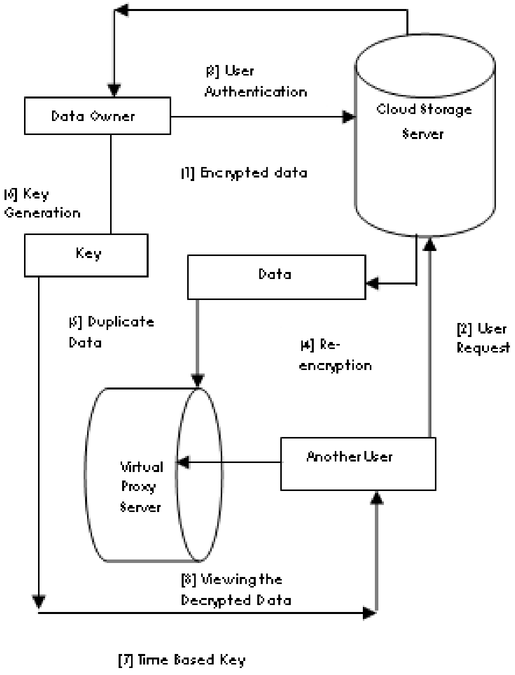

Fig. 1: Secure retrieval of encrypted data

to user. After receiving the decryption key user can view the document only for a specific amount of time.

RESULTS AND DISCUSSION

Table 1 shows the retrieval cost of documents in terms of time complexity on an ordinary as well as in proxy server. Experiments reveal that proxy server concept is a time consuming one where a significant amount of time is spent on creation of proxy server itself. Figure 2 conforms that time complexity increases as the retrieval document size increases. This can be overcome by higher capacity systems, but here the experiments were carried out on an i3 processor of RAM size 8 GB, HDD 120 GB.

Analysis of malicious attacks using cloudsim 3.0.3: Cloudsim was employed to analysis the attacks of malicious users where different number of documents with different size were taken for study purpose. Cloudsim is a simlation tool used in cloud which is predominantly for analyzing the performance and allows the users to have a proper insight into cloud scenarios without worrying about the low level implementation details. Document loss details are tabulated in Table 2.

Figure 3 conforms that proxy server takes more time relatively for encryption and creating itself comparing to traditional server. Even though, the graph indicates that without proxy server requires low time, the security issues are not resolved to a significant level. Thus, it is clear that the proxy server requires more time but the data is highly secured as that the number of attacks in the proxy server is low when compared to the normal servers. Graph indicates a negligible level of security threats that is due to the key transferred to the unauthorized persons but the hinder against the attack is very high in the proxy server than any of the servers being used. Thus, experimental analysis proves that proxy server can efficiently provide secure data storage and retrieval.

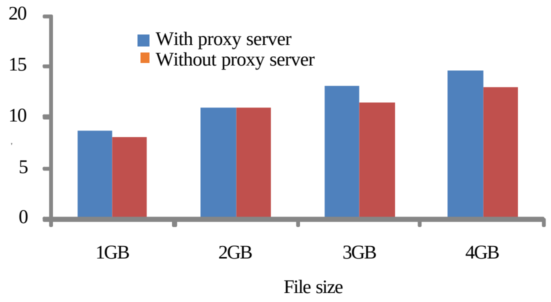

Fig. 2: Comparison of retrieval on traditional server and server with proxy

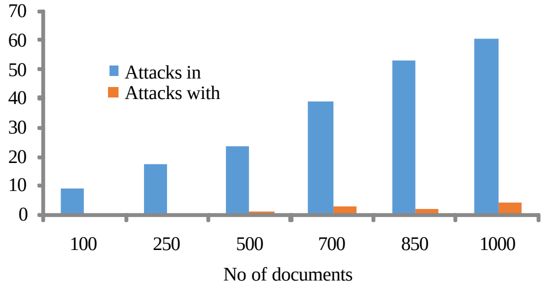

Fig. 3: Document loss comparison on normal and proxy servers

Table 1: Retrieval cost of encrypted data

File size (GB)	With proxy server (ms)	Without proxy server (ms)
1	8.60	8.00
2	10.9	10.89
3	13.0	11.40
4	14.5	12.90

Table 2: Comparison of loss of documents on normal and proxy servers

No. of documents	Size (GB)	Loss of documents in a server without proxy concept	Loss of documents due to attacks in the proxy server
100	1	9	0
250	2	17	0
500	2	23	1
700	3	39	3
850	3	53	2
1000	4	60	4

CONCLUSION

Experiments reveal that proxy server creation for document access takes time but ensures the access of document for a specific time span only, this leads to significant level of security. Results of simulation tool supports to conclude that document loss have been reduced around 97% by using proxy server. It shows that as size of documents increases time complexity also increases but as the capacity of the processor or remote server increases, the time complexity can be reduced and security can be promoted to a significant level in order to provide secure sensitive data storage and retrieval.

RECOMMENDATIONS

This preliminary work can be extended to multimedia and stream nature voluminous data with different encryption technique and fragmentation techniques in distributed environments.

REFERENCES

Almulla, S.A. and C.Y. Yeun, 2010. Cloud computing security management. Proceedings of the 2nd International Conference on Engineering Systems Management and its Applications, March 30-April 1, Sharjah, pp: 1-7.

Anuchart, T. and G. Gong, 2011. ABE based authorization in semi-trusted cloud computing proceeding data cloud. Proceedings of the Second International Workshop on Data Intensive Computing in the Clouds, November 12-18, 2011, ACM, Seattle, Washington, USA., ISBN:978-1-4503-1144-1, pp: 41-50.

Bethencourt, J., A. Sahai and B. Waters, 2007. Ciphertext-policy attribute-based encryption. Proceedings of the IEEE Symposium on Security and Privacy, May 20-23, 2007, Berkeley, CA., USA., pp: 321-334.

Bhisikar, P. and A. Sahu, 2013. Security in data storage and transmission in cloud computing. Int. J. Adv. Res. Comput. Sci. Software Eng., 3: 410-415.

Inbarani, W.S., G.S. Moorthy and C.K.C. Paul, 2013. An approach for storage security in cloud computing-a survey. Int. J. Adv. Res. Comput. Eng. Technol. (IJARCET.), 2: 174-179.

Jung, T., X.Y. Li, Z. Wan and M. Wan, 2015. Control cloud data access privilege and anonymity with fully anonymous attribute-based encryption. IEEE. Trans. Inf. Forensics Security, 10: 190-199.

Kandias, M., N. Virvilis and D. Gritzalis, 2011. The Insider Threat in Cloud Computing. In: Critical Information Infrastructure Security. Sandro, B., B. Hammerli, D. Gritzalis and S. Wolthusen (Eds.). Springer, Berlin, Germany, ISBN:978-3-642-41475-6, pp: 93-103.

Pothumani, B. and A. Gosh, 2015. Literature survey on collaborative cloud computing for sharing resource in trustworthy manner. Int. J. Innovative Res. Comput. Commun. Eng., 3: 2391-2397.

Prakash, G.L., M. Prateek and I. Singh, 2014. Data encryption and decryption algorithms using key rotations for data security in cloud system. Proceedings of the 2014 International Conference on Signal Propagation and Computer Technology (ICSPCT.), July 12-13, 2014, IEEE, New York, USA., ISBN:978-1-4799-3140-8, pp: 624-629.

Priyadharshini, B., C.M. Belinda and M.R. Kumar, 2013. A secure code based cloud storage system using proxy re-encryption scheme in cloud computing. IOSR. J. Comput. Eng. (IOSR-JCE.), 9: 22-27.

Purohit, B. and P.P. Singh, 2013. Data leakage analysis on cloud computing. Int. J. Eng. Res. Appl. (IJERA.), 3: 1311-1316.

Santosh, L. and S. Dudhani, 2015. Secure key for authentication and secret sharing in cloud computing. Int. J. Adv. Res. Comput. Sci. Software Eng., 5: 1008-1011.

Shaikh, F.B. and S. Haider, 2011. Security threats in cloud computing. Proceedings of the International Conference on Internet Technology and Secured Transactions, December 11-14, 2011, Abu Dhabi, pp: 214-219.

Soofi, A.A. and M.I. Khan, 2014. Encryption techniques for cloud data confidentiality. Int. J. Grid Distrib. Comput., 7: 11-20.

Sun, Y., J. Zhang, Y. Xiong and G. Zhu, 2014. Data security and privacy in cloud computing. Int. J. Distrib. Sensor Networks, 2014: 1-9.

Tebaa, M., S. El Hajji and A. El Ghazi, 2012. Homomorphic encryption applied to the cloud computing security. Proceedings of the World Congress on Engineering, Volume 1, July 4-6, 2012, London, UK., pp: 1-4.

Yan, L., C. Rong and G. Zhao, 2009. Strengthen cloud computing security with federal identity management using hierarchical identity-based cryptography. Proceedings of the IEEE International Conference on Cloud Computing, December 1-4, 2009, Springer, Berlin, Germany, ISBN:978-3-642-10664-4, pp: 167-177.