International Multidisciplinary Research Journal 2012, 2(7):49-53
ISSN: 2231-6302
Available Online: http://irjs.info/

INTERNATIONAL MULTIDISCIPLINARY RESEARCH JOURNAL

Recoganisation and security guidance of data integrity in cloud storage

Ramalingam Sugumar[1], Tamilenthi.S[2]* and K.Gopinath[1]

[1]Dept .of Computer Science, Christhuraj College, Panjappur,Trichy- 620012, India
[2]Department of Earth Science, Tamil University, Thanjavur- 613 010, India

Abstract
Cloud computing has been envisioned as the de-facto solution to the rising storage costs of IT Enterprises. With the high costs of data storage devices as well as the rapid rate at which data is being generated it proves costly for enterprises or individual users to frequently update their hardware. Storage outsourcing of data to cloud storage helps such firms by reducing the costs of storage, maintenance and personnel. It can also assure a reliable storage of important data by keeping multiple copies of the data thereby reducing the chance of losing data by hardware failures. The study deal with the problem of implementing a protocol for obtaining a proof of data possession in the cloud sometimes referred to as Proof of irretrievability (POR). The problem tries to obtain and verify a proof that the data that is stored by a user at remote data storage in the cloud (called cloud storage archives or simply archives) is not modified by the archive and thereby the integrity of the data is assured. The verification systems prevent the cloud storage archives from misrepresenting or modifying the data stored at it without the consent of the data owner by using frequent checks on the storage archives.

Keywords: Cloud storage, Sentinels, Simply Archives and Meta-Data.

INTRODUCTION

Cloud computing has been envisioned as the de-facto solution to the rising storage costs of IT Enterprises. With the high costs of data storage devices as well as the rapid rate at which data is being generated it proves costly for enterprises or individual users to frequently update their hardware. Apart from reduction in storage costs data outsourcing to the cloud also helps in reducing the maintenance. Cloud storage moves the user's data to large data centers, which are remotely located, on which user does not have any control. However, the unique feature of the cloud poses many new security challenges which need to be clearly understood and resolved. It provides a scheme which gives a proof of data integrity in the cloud which the customer can employ to check the correctness of his data in the cloud. The proof can be agreed upon by both the cloud and the customer and can be incorporated in the Service level agreement (SLA).

Cloud computing provides huge computing services to the business for improving the organizational growth. Basic requirement needed for this technology is Internet but provides higher capability when compared to the Internet. Cloud computing is a combination of computation, software, data access and also provides storage services. In Cloud, storage of data and the location of stored data are not known to the user. Cloud computing adopts the concept of virtualization, service oriented architecture, autonomic, and utility computing. The cloud has more advantages and easy to implement with any business logics. Cloud delivers services from different data sources and servers located on different geographical location but the user gets single point of view from the cloud service.

Data outsourcing to cloud storage servers is raising trend among many firms and users owing to its economic advantages. The essentially means that the owner (client) of the data moves its data to a third party cloud storage server which is supposed to - presumably for a fee - faithfully store the data with it and provide it back to the owner whenever required.

As data generation is far outpacing data storage it proves costly for small firms to frequently update their hardware whenever additional data is created. Also maintaining the storages can be a difficult task. Storage outsourcing of data to cloud storage helps such firms by reducing the costs of storage, maintenance and personnel. It can also assure a reliable storage of important data by keeping multiple copies of the data thereby reducing the chance of losing data by hardware failures.

Storing of user data in the cloud despite its advantages has many interesting security concerns which need to be extensively investigated for making it a reliable solution to the problem of avoiding local storage of data. In the paper deal with the problem of implementing a protocol for obtaining a proof of data possession in the cloud sometimes referred to as Proof of irretrievability (POR).The problem tries to obtain and verify a proof that the data that is stored by a user at a remote data storage in the cloud (called cloud storage archives or simply archives) is Not modified by the archive and thereby the integrity of the data is assured.

Such verification systems prevent the cloud storage archives from misrepresenting or modifying the data stored at it without the consent of the data owner by using frequent checks on the storage archives. Such checks must allow the data owner to efficiently, frequently, quickly and securely verify that the cloud archive is not cheating the owner. Cheating, in the context, means that the storage archive might delete some of the data or may modify some of the data.

Received: May 12, 2012; Revised: June 18, 2012; Accepted: July 20, 2012.

*Corresponding Author

S.Tamilenthi
Department of Earth Science, Tamil University, Thanjavur, India

Email: rst_geo2011@yahoo.com

LITERATURATURE OVERVIEW

Storing of user data in the cloud despite its advantages has many interesting security concerns which need to be extensively investigated for making it a reliable solution to the problem of avoiding local storage of data. Many problems like data authentication and integrity (i.e., how to efficiently and securely ensure that the cloud storage server returns correct and complete results in response to its clients' queries E. Mykletun et al(2006), outsourcing encrypted data and associated difficult problems dealing with querying over encrypted domain (D. X. Song et al(2000) were discussed in research literature.

Ari Juels and Burton S. Kaliski Jr(2007) proposed a scheme called Proof of retrievability for large files using "sentinels" In this scheme, unlike in the key-hash approach scheme, only a single key can be used irrespective of the size of the file or the number of files whose retrievability it wants to verify.

At the client we only store two functions, the bit generator function g, and the function h which is used for encrypting the data. Hence the storage at the client is very much minimal compared to all other schemes acoording to G. Ateniese et al(2007) that were developed. Hence this scheme proves advantageous to thin clients like PDAs and mobile phones.

It is not so efficient by using a few technologies such as flexible distributed scheme, Reed Solomon technique, and BLS algorithm , Qian Wang et al(2011) to give more integrity to the operation like change, removal, and add of data from cloud server.

Security is not a new issue and now it is recognized as one of the most complex problems. Due to the importance of security importance, it has been an issue in an increasingly growing network connectivity, size and implementation of new information technologies (Anderson *et al.,* 2001; McClure *et al.,* 2003). Several attempts have been made to provide security using a software agent systems approach. In these systems, the main focus was on providing a solution for specific security issues, such as authentication and authorization.

The ultimate challenge in cloud computing is data-level security, and sensitive data is the domain of the enterprise, not the Cloud Service Providers. Security will need to move to the data level so that enterprises can be sure their data is protected wherever it goes. For example, with data-level security, the enterprise can specify that this data is not allowed to go outside of the specific cloud server. It can also force encryption of certain types of data, and permit only specified users to access the data (John *et al.,* 2010).

In a data possession work (Ateniese *et al.,* 2007) defined the "provable data possession" (PDP) model for ensuring possession of file on untrusted storages. Their scheme utilized public key based homomorphic tags for auditing the data file, thus providing public verifiability.

In another data possession work (Curtmola *et al.,* 2008) aimed to ensure data possession of multiple replicas across the distributed storage system. They extended the PDP scheme to cover multiple replicas without encoding each replica separately, providing guarantees that multiple copies of data are actually maintained. Our DER Agent will do the same job of (Curtmola *et al.,* 2008).

The work (Schwarz. T. S. J, & Miller. E. L., 2006) proposed to ensure file integrity across multiple distributed servers using erasure-coding and block-level file integrity checks.

MATERIALS

Hardware Requirements

- System : Pentium IV 2.4 GHz.
- Hard Disk : 40 GB.
- Floppy Drive : 1.44 Mb.
- Monitor : 15 VGA Colour.
- Mouse : Logitech.
- Ram : 512 Mb.

Software Requirements:

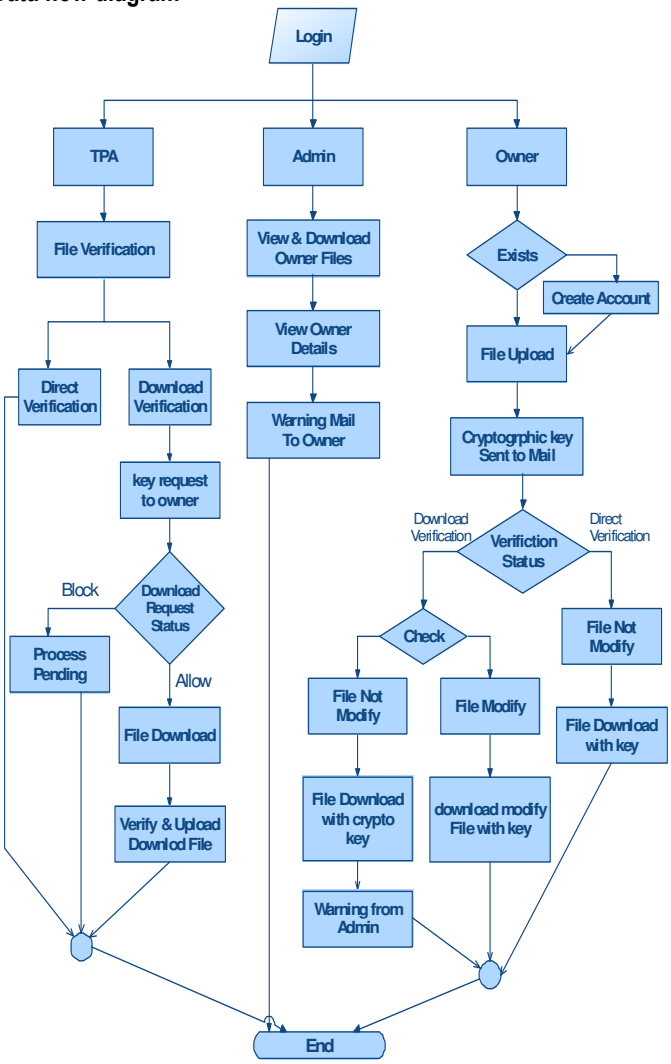
- Operating system : Windows XP.
- Coding Language : ASP.Net with C#
- Data Base : SQL Server 20**05**

ANALYSIS

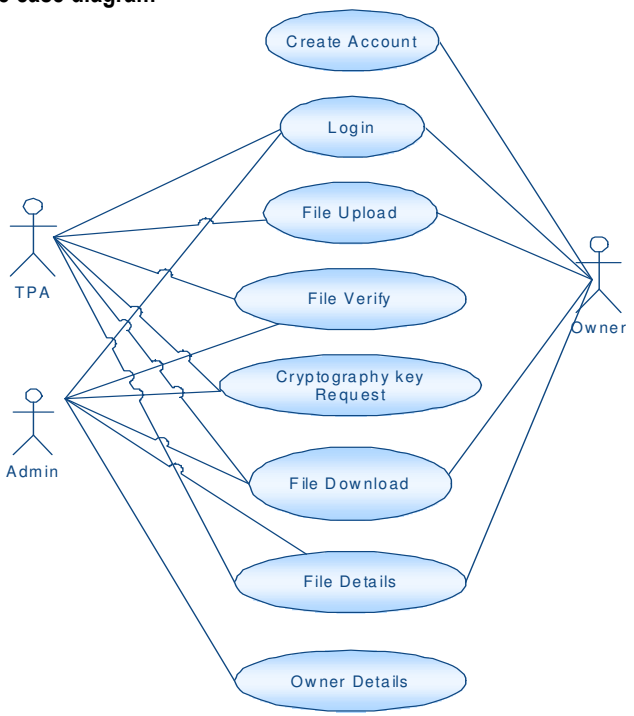
System design

This is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of the input data to the system, various processing carried out on these data, and the output data is generated by the system.

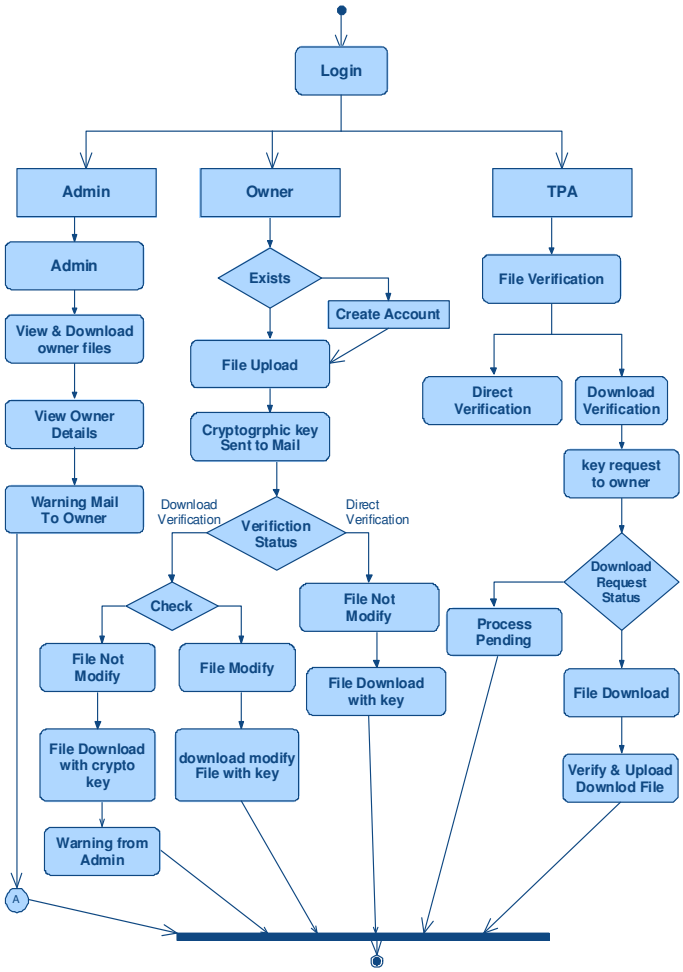
Data flow diagram



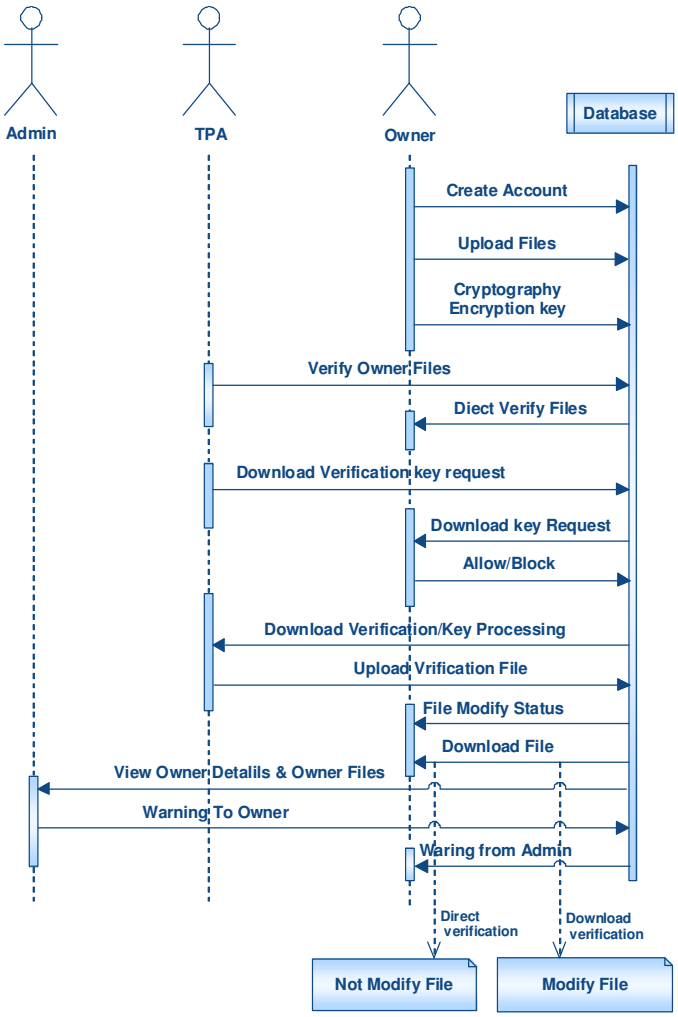
Use case diagram



Activity diagram



Sequence diagram



In Existing System, As data generation is far outpacing data storage it proves costly for small firms to frequently update their hardware whenever additional data is created. Also maintaining the storages can be a difficult task. It transmitting the file across the network to the client can consume heavy bandwidths. The problem is further complicated by the fact that the owner of the data may be a small device, like a PDA (personal digital assist) or a mobile phone, which have limited CPU power, battery power and communication bandwidth.

It has some disadvantages

- The main drawback of this scheme is the high resource costs it requires for the implementation.
- Also computing hash value for even a moderately large data files can be computationally burdensome for some clients (PDAs, mobile phones, etc).
- Data encryption is large so the disadvantage is small users with limited computational power (PDAs, mobile phones etc.).

In the Proposed System, One of the important concerns that need to be addressed is to assure the customer of the integrity i.e. correctness of his data in the cloud. As the data is physically not accessible to the user the cloud should provide a way for the user to check if the integrity of his data is maintained

or is compromised. In this paper we provide a scheme which gives a proof of data integrity in the cloud which the customer can employ to check the correctness of his data in the cloud. This proof can be agreed upon by both the cloud and the customer and can be incorporated in the Service level agreement (SLA). It is important to note that our proof of data integrity protocol just checks the integrity of data i.e. if the data has been illegally modified or deleted. The advantages are

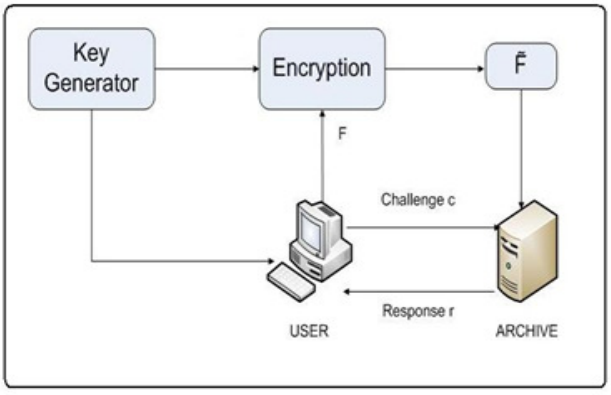
- Apart from reduction in storage costs data outsourcing to the cloud also helps in reducing the maintenance.
- Avoiding local storage of data.
- By reducing the costs of storage, maintenance and personnel.
- It reduces the chance of losing data by hardware failures.
- Not cheating the owner.

Let the verifier V wishes to the store the file F with the archive. Let this file F consist of n file blocks. We initially preprocess the file and create metadata to be appended to the file. Let each of the n data blocks have m bits in them. A typical data file F which the client wishes to store in the cloud.

Each of the Meta data from the data blocks m_i is encrypted by using a suitable algorithm to give a new modified Meta data M_i. Without loss of generality we show this process by using a simple XOR operation. The encryption method can be improvised to provide still stronger protection for verifier's data. All the Meta data bit blocks that are generated using the above procedure are to be concatenated together. This concatenated Meta data should be appended to the file F before storing it at the cloud server. The file F along with the appended Meta data e F is archived with the cloud.

Data outsourcing to cloud storage servers is raising trend among many firms and users owing to its economic advantages. The essentially means that the owner (client) of the data moves its data to a third party cloud storage server which is supposed to - presumably for a fee - faithfully store the data with it and provide it back to the owner whenever required.

As data generation is far outpacing data storage it proves costly for small firms to frequently update their hardware whenever additional data is created. Also maintaining the storages can be a difficult task. Storage outsourcing of data to cloud storage helps such firms by reducing the costs of storage, maintenance and personnel. It can also assure a reliable storage of important data by keeping multiple copies of the data thereby reducing the chance of losing data by hardware failures.



Algorithm
Meta-Data Generation

Let the verifier V wishes to the store the file F with the archive. Let this file F consist of n file blocks. We initially preprocess the file and create metadata to be appended to the file. Let each of the n data blocks have m bits in them. A typical data file F which the client wishes to store in the cloud.

Each of the Meta data from the data blocks m_i is encrypted by using a suitable algorithm to give a new modified Meta data M_i. Without loss of generality show this process by using a simple XOR operation. The encryption method can be improvised to provide still stronger protection for verifier's data. All the Meta data bit blocks that are generated using the above procedure are to be concatenated together. This concatenated Meta data should be appended to the file F before storing it at the cloud server. The file F along with the appended Meta data e F is archived with the cloud.

Modules
Cloud Storage

Data outsourcing to cloud storage servers is raising trend among many firms and users owing to its economic advantages. This essentially means that the owner (client) of the data moves its data to a third party cloud storage server which is supposed to - presumably for a fee - faithfully store the data with it and provide it back to the owner whenever required.

Simply Archives

The problem tries to obtain and verify a proof that the data that is stored by a user at remote data storage in the cloud (called cloud storage archives or simply archives) is not modified by the archive and thereby the integrity of the data is assured. Cloud archive is not cheating the owner, if cheating, in the context, means that the storage archive might delete some of the data or may modify some of the data. While developing proofs for data possession at untrusted cloud storage servers we are often limited by the resources at the cloud server as well as at the client.

Sentinels

In this scheme, unlike in the key-hash approach scheme, only a single key can be used irrespective of the size of the file or the number of files whose retrievability it wants to verify. Also the archive needs to access only a small portion of the file F unlike in the key-has scheme which required the archive to process the entire file F for each protocol verification. If the prover has modified or deleted a substantial portion of F, then with high probability it will also have suppressed a number of sentinels.

Verification Phase

The verifier before storing the file at the archive preprocesses the file and appends some Meta data to the file and stores at the archive. At the time of verification the verifier uses this Meta data to verify the integrity of the data. It is important to note that our proof of data integrity protocol just checks the integrity of data i.e. if the data has been illegally modified or deleted. It does not prevent the archive from modifying the data.

CONCLUSION

Findings

Cloud computing provides huge computing services to the business for improving the organizational growth. Basic requirement needed for this technology is Internet but provides higher capability when compared to the Internet. Cloud computing is a combination of computation, software, data access and also provides storage services. In Cloud, storage of data and the location of stored data are not known to the user. One of the important concerns that need to be addressed is to assure the customer of the integrity i.e. correctness of his data in the cloud. As the data is physically not accessible to the user the cloud should provide a way for the user to check if the integrity of his data is maintained or is compromised. In this paper we provide a scheme which gives a proof of data integrity in the cloud which the customer can employ to check the correctness of his data in the cloud. This proof can be agreed upon by both the cloud and the customer and can be incorporated in the Service level agreement (SLA). It is important to note that our proof of data integrity protocol just checks the integrity of data i.e. if the data has been illegally modified or deleted. The advantages are

- Apart from reduction in storage costs data outsourcing to the cloud also helps in reducing the maintenance.
- Avoiding local storage of data.
- By reducing the costs of storage, maintenance and personnel.
- It reduces the chance of losing data by hardware failures.
- Not cheating the owner.

Cloud computing has been envisioned as the de-facto solution to the rising storage costs of IT Enterprises. With the high costs of data storage devices as well as the rapid rate at which data is being generated it proves costly for enterprises or individual users to frequently update their hardware. Apart from reduction in storage costs data outsourcing to the cloud also helps in reducing the maintenance.

In this study is to facilitate the client in getting a proof of integrity of the data which he wishes to store in the cloud storage servers with bare minimum costs and efforts. Our scheme was developed to reduce the computational and storage overhead of the client as well as to minimize the computational overhead of the cloud storage server. We also minimized the size of the proof of data integrity so as to reduce the network bandwidth consumption. Many of the schemes proposed earlier require the archive to perform tasks that need a lot of computational power to generate the proof of data integrity. But in our scheme the archive just need to fetch and send few bits of data to the client.

REFERENCE

[1] E. Mykletun, M. Narasimha, and G. Tsudik, 2006."Authentication and integrity in outsourced databases," *Trans. Storage*, vol. 2, no. 2, pp. 107–138.

[2] D. X. Song, D. Wagner, and A. Perrig, 2000."Practical techniques for searches on encrypted data," in SP '00: Proceedings of the 2000 IEEE Symposium on Security and Privacy. Washington, DC, USA: *IEEE Computer Society*, p. 44.

[3] A. Juels and B. S. Kaliski, Jr., 2007."Pors: proofs of retrievability for large files," in CCS '07: *Proceedings of the 14th ACM conference on Computer and communications security. New York, NY, USA: ACM*, pp. 584–597.

[4] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, 2007. "Provable data possession at untrusted stores," in CCS '07:*Proceedings of the 14th ACM conference on Computer and communications security. New York, NY, USA: ACM,* pp. 598–609.

[5] Towards Secure and Dependable Storage Services in Cloud Computing Cong Wang, Student Member, IEEE, Qian Wang, Student Member, IEEE, Kui Ren, Member, IEEE, Ning Cao, Student Member, IEEE, and Wenjing Lou, Senior Member, IEEE-2011.

[6] Schwarz. T. S. J, & Miller. E. L. 2006. "Store, Forget, and Check: Using Algebraic Signatures to Check Remotely Administered Storage," *Proc. of ICDCS* '06, pp. 12–12.

[7] Curtmola. R, Khan. O, Burns. R, & Ateniese. G. 2008. "MR-PDP: Multiple- Replica Provable Data Possession," *Proc. of ICDCS* '08, pp. 411–420.

[8] Ateniese. G, Burns. R, Curtmola. R, Herring. J, Kissner. L, Peterson. Z, & Song. D, 2007. "Provable Data Possession at Untrusted Stores," *Proc. Of CCS* '07, pp. 598–609

[9] John, W. Rittinghouse Jame, F. & Ransome. 2010. "Cloud Computing Implementation, Management and Security", CRC Press, p. 153.

[10] Anderson R. 2001. In: Security engineering: a guide to building dependable distributed systems. New York: John Wiley & Sons Inc

[11] McClure S, Scambray J, Kurtz G. 2003. In: "Hacking exposed: network security secrets and solutions". McGraw-Hill Osborne Media.