International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 - 5161

Available at http://inpressco.com/category/ijcet

Research Article

Security Framework to Ensure the Confidentiality of Outsourced Data in Public Cloud Storage

L. Arockiam[Á] and S. Monikandan[B*]

[Á]St. Joseph's College, Trichy, Tamilnadu, India
[B]M S University, Tirunelveli, Tamilnadu, India

Accepted 01 May 2014, Available online 01 June 2014, **Vol.4, No.3 (June 2014)**

Abstract

Cloud makes data storage very smart. It provides enormous amount of space to store the user's data. These data are instantly available for the users at any time and it reduces the cost with respect to hardware support, technical experts available and license of the database. It leverages the small and medium scale enterprise straight away to start their business. Nevertheless cloud has many advantages; enterprises are undecided to deploy their data in the cloud storage, because data security issue is the top concern in Cloud Storage (CS). Security issues arose from the attack of cloud data by the hackers. Hackers are either outsiders from the enterprises or insiders from cloud service providers. This paper considers the issue that the CS is attacked by insiders from the cloud service provider. To protect this attack, this paper proposes the data confidentiality framework called AROMO. As per the proposed framework, user's data are protected in the Cloud Storage. The proposed framework has a mechanism which uses two techniques to protect the data that are encryption and obfuscation. The data are encrypted and obfuscated before it is uploaded to CS. A metadata is maintained in the client. It contains the details of encryption and obfuscation applied on the data. To retrieve data from the CS, query is generated to run on encrypted and obfuscated data on the server side. Based on the query from the client, required data is fetched out from the CS; it could be decrypted and deobfuscated in client side based on the metadata details.

Keywords: *Cloud Storage; Security; Confidentiality; Data Outsourcing; Metadata*

1. Introduction

Storage and maintenance of data are the most important tasks in all kinds of enterprises. Data storage in an enterprise involves high costs because it requires physical resources like hard disks, servers, network etc., and frequent management procedures like backup, tuning, etc. and skilled administrative personnel for monitoring the whole database administration process. This would increase the enterprise's headache. The cheaper solution for data storage and maintenance is data outsourcing (Pierangela *et al*., 2010). Once the data are outsourced to a specialized provider, then they will take care of the data storage within its own specialized structure to offer high availability and disaster protection (Francesco, 2011).

Cloud offers increased amount of storage and processing power to run users' applications (Mell and Grance, 2009), (Ross, 2010). It enables new ways to access information, process and analyze data. It also connects people and resources from any location all over the world (Arockiam *et al*., 2011). Even though, the users have gained more advantages in cloud, there are certain limitations faced by the users when it is implemented. Data protection, operational integrity, vulnerability management, business continuity (BC), disaster recovery (DR), and identity management (IM) are top concerns of security issues for Cloud Computing. Among the above, data protection is the most important key concern. Security of Cloud storage becomes a key factor for users to adapt it.

Cloud data are managed by skilled administrators. They have full control over the data stored in CS. If any data loss or crash occurs, they are responsible for data recovery, restoring the data and tuning of the database. This will lead to security issues from the insiders in the cloud providers. These types of threats are not easy to be trashed (Miller, 2010). The proposed framework should avoid occurrence of security issues from the insider from the cloud providers.

This paper mainly focuses on the data security issues in Cloud Storage**.** Cloud security is a major issue for the cloud users. Cloud users always expect high level protection for their sensitive data. Violation of protection leads to user's dissatisfaction. For example, consider that an organization maintains its sensitive data in the cloud; unfortunately the data may be stolen. This will impair the organizational growth and it may also leverage the competitor to come up.

2. Data Outsourcing

*Corresponding author **S. Monikandan** is a Research scholar and **Dr. L. Arockiam** is working as Associate Professor

Data outsourcing in public cloud is becoming increasingly popular and introducing a new paradigm, called *database-as-a-service*, (Damiani *et al*., 2005) where a client's data is stored at an external service provider. This scenario presents new research challenges on which the usability of the system is based.

The main advantage of outsourcing is related to the costs of in-house versus outsourced hosting; outsourcing provides

i) Significant cost savings and service benefits and

ii) Promises higher availability and more effective disaster protection than in-house operations.

Users can outsource their data to cloud and retrieve them when needed. Cloud database providers should store the user's data in the database server and provide maximum availability of data and effective disaster recovery. The data outsourcing scenario in public cloud is represented in Figure 1. Cloud users may be an enterprise or a single user (Fatima *et al*., 2011).

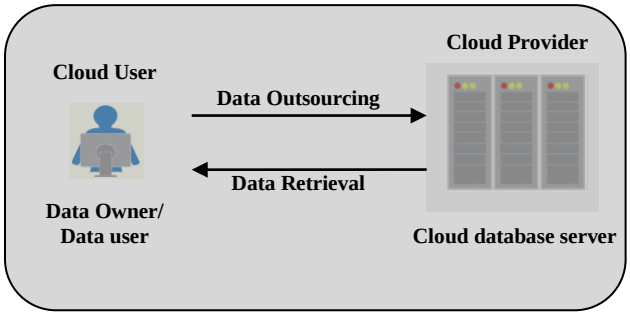

Figure 1: Data outsourcing in public cloud storage - owners and users are the same

Outsourced data can be accessed in two schemes. In the first scheme data owner and data user are the same. In the second data owner and data user are different. Figure 1 represents the first scenario and figure 2 represents the second scenario.

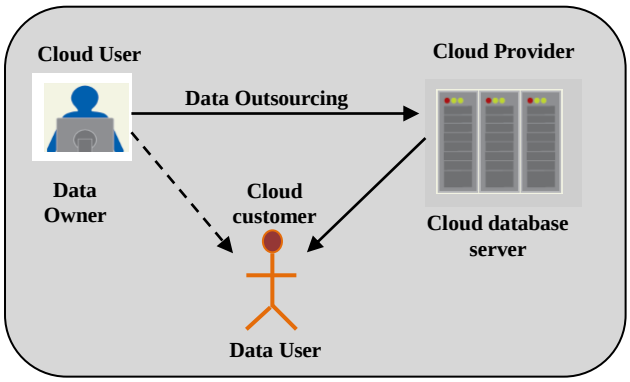

Figure 2: Data outsourcing in public cloud storage – owners and users are different

As a consequence of this trend toward outsourcing, highly sensitive data is now stored on systems run in locations that are not under the data owner's control. Therefore, data confidentiality and integrity can be put at risk. There is a possibility of potential improper use of database information can be achieved by the provider itself. The traditional access control techniques may prevent data access by external users, but not by internal administrators.

3. Confidentiality of Outsourced Data

Data sent to the cloud are not stored in single cloud storage server. It will be replicated to different cloud data centers located in different places in the world. Data centers are controlled and maintained by different experts. The data may be hacked from any data center. In cloud storage, maintaining the confidentiality of the data is the main issue. To address this issue, the confidentiality ensuring technique namely encryption is used. The problem is that encryption limits data access from the cloud server. Because, accessing the cloud data needs decryption in cloud server, but if the cloud data is decrypted in the cloud server then the privileged user of service providers can look into the data (Rehman and Hussain, 2011). In particular, searching and indexing the data becomes problematic with encrypted data (Chow *et al*., 2009). For example, if data is stored in clear-text, one can efficiently search for a document by specifying a keyword. This is impossible to do with traditional, randomized encryption schemes. State-of-the-art of cryptography may offer new tools to solve these problems. Cryptographers have recently invented versatile encryption schemes that allow operation and computation on the ciphertext. For example, searchable encryption (Yao *et al*., 2012)(Kamara *et al*., 2006, 2011) (Kamara *et al*., 2012)(Curtmola el al., 2006) allows the data owners to compute a capability to search on encrypted data from their secret key. This encodes a search query, and the cloud can use this capability to decide which documents match the search query, without learning any additional information. Other cryptographic primitives such as homomorphic encryption (Ogburn *et al*., 2013) (Lauter *et al*., 2011) and Private Information Retrieval (PIR) (Stainslaw *et al*., 2013)(Huang *et al*., 2013a, 2013b) perform computations on encrypted data without decrypting.

Apart from this, encryption alone is not sufficient for protecting the outsourced data in the cloud storage. Data obfuscation is a technique used for data security. This research work integrates the data obfuscation (Maheswari *et al*., 2012) with data encryption to enhance confidentiality of the data stored in the cloud storage.

4. Issues in Cloud Storage

User's data are outside the user's control and could potentially be granted to untrusted parties. Cloud storage provider does learn information about customer data. Followings are the issues corresponding to cloud storage.

- Data sent to the cloud is stored in the public cloud database.
- Outsourced data should be controlled and maintained by the providers.
- Users don't have any control on the data in cloud storage.
- Data may mingle with other user's data.
- Data may not be encrypted by the Cloud providers(Ex: Amazon S3)

- Due to key management, same key is used to encrypt all consumer data.
- Lead to the data protection issue: Security, Confidentiality and Integrity can be put at risk.
- Potential improper use of database information can be achieved by the provider itself.
- Confidentiality of cloud data is broken and which will cause loss of data to the Enterprises.

Ensuring the confidentiality of the data in cloud is a vital part in cloud storage. There are several questions from the users, which could not be answered by the cloud service providers.

- Where is the data stored?
- Is the data is encrypted or not?
- Which encryption is used for data security?
- How to maintain the key for different user's data?
- How are the data retrieved from cloud storage?
- Who are all responsible to access the data in cloud provider side?

5. Related Works

S kamara *et al*. (Kamara *et al*., 2010) considered the problem to build a secure cloud storage service on top of a public cloud infrastructure where the service provider was not completely trusted by the customer. They proposed a possible architecture for cloud storage that combines recent and non-standard cryptographic primitives. The architecture consists of three components: a data processor (DP) that processes data before it is sent to the cloud; a data verifier (DV) that checks whether the data in the cloud has been tampered with and a token generator (TG) that generates tokens that enable the cloud storage provider to retrieve segments of customer data and a credential generator that implements an access control policy by issuing credentials to the various parties in the system (these credentials will enable the parties to decrypt the encrypted files according to the policy). This architecture is designed for both consumer and enterprise scenarios.

Atiq ur Rehman *et al*. (Rehman and Hussain, 2011) proposed a model to preserve confidentiality of data stored in DaaS model. The proposed model stores sensitive data with a combination of encryption and obfuscation. The sensitive character based attributes are encrypted and numeric sensitive columns are obfuscated based on the metadata repository. Moreover, query will be executed over encrypted and obfuscated data. Once the required data is filtered on provider end, then it is transferred to client where it is decrypted and deobfuscated before the presentation to the user.

Miranda *et al*. (Miranda *et al*., 2009) have proposed user-centric trust model to achieve data confidentiality of cloud paradigm. According to proposed model, data owner controls the sensitivity of data through privacy manager which is available on client side. This privacy manager uses the technique called obfuscation to achieve confidentiality rather than the traditional encryption technique. According to the given approach, the data is first disguised with mathematical functions or by using any logic on client side and then it is sent to service provider end for storage. Authors have used the database of salesforce.com where enterprises upload their business critical information. According to the given scenario the security threat to data is from service provider end. The presented framework has five modules namely Preference, Data Access, Feedback, Personae and obfuscation. Preference module allows data owner to set confidentiality related inclinations. Data access module stores log whenever confidentiality related violations are observed. Feedback module informs about the use of information. Personae module allows the data owner to select different security levels during interaction with service provider. Obfuscation module defends confidentiality of data when service provider is malevolent. Although this technique is efficient, however confidentiality of data can be compromised through brute force attack.

6. Motivation

Nowadays Cloud storage is widely popular and it is used by millions of people in the world. The users are eager to adopt the cloud storage by outsourcing their IT requisite. But, due to the risk related to security of data in the cloud storage, the users are reluctant to store their data in cloud. Securing the outsourced data is the most important issue in cloud storage. By ensuring the confidentiality of the outsourced data in cloud, data security can be achieved in the cloud storage. Hence, there is a need to ensure the confidentiality of the data in cloud storage. Motivated by this fact, this paper aims at ensuring the confidentiality of the outsourced data by achieving the following goals.

- To ensure that once the data is stored in the cloud, it can be accessed or read only by the data owner.
- To encrypt non-numerical data and obfuscate numeric data before they are uploaded to cloud storage.
- To maintain metadata regarding the encrypted and obfuscated data.
- To define a concrete confidentiality framework for cloud users to secure data storage in cloud.

7. Proposed Framework: AROMO

The overall proposed framework AROMO is given in figure 3. In order to store data into the cloud storage, first required data are analyzed for knowing their types. Based on the data analysis the non-numeric data are encrypted and numeric data are obfuscated on client side according to the requirement through respective module. The encryption and obfuscation related information is stored on metadata repository on client side. After the completion of encryption and obfuscation module on the client side, data is transferred to the cloud storage. To achieve confidentiality, data is not decrypted or deobfuscated on server side. When the user queries from the database available on cloud server, client query is first transformed on client side to run on encrypted and obfuscated data. After the transformation, the query is transferred on server to execute and to fetch the required data from the database engine. When required data is fetched on server side then

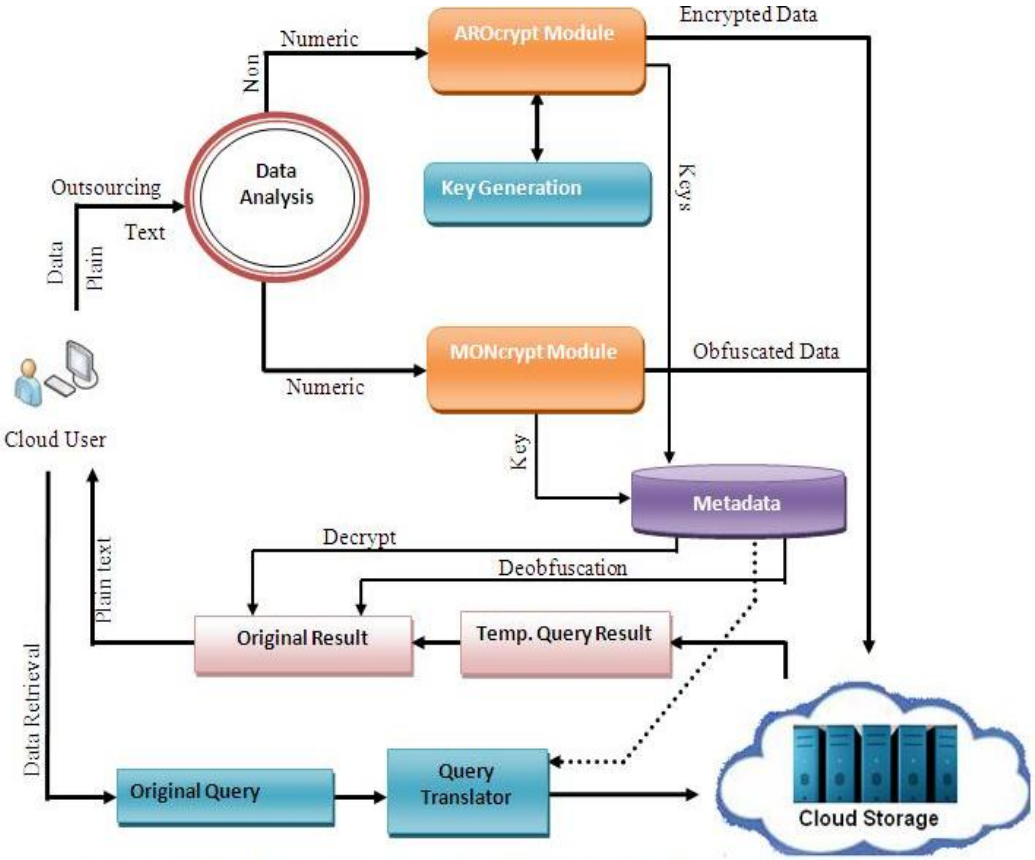

Figure 3: AROMO Overall proposed framework for ensuring the confidentiality of cloud data

it is transferred to client side where it is first decrypted and deobfuscated before it is presented to user terminal.

7.1 Encryption module

Encryption technique is used to mask the data. This module is applied on the non-numerical data. This module uses the encryption technique proposed in (Arockiam *et al*., 2013). The proposed encryption technique in (Arockiam *et al*., 2013) gives maximum protection to data in CS. The technique ensures the confidentiality of data stored in the CS. It uses the ASCII values of each character in the plaintext data. Four keys are used for proposed encryption technique in (Arockiam *et al*., 2013).

Keys used for encryption are to be got from the key generation module and they are stored in the client side metadata. Key generation module generates the key based on the procedure of random number generation. Keys are generated each time before the data are encrypted. Keys used for encryption are not communicated to the cloud providers. Cloud providers don't have the knowledge of keys and encryption algorithm used of data confidentiality.

7.2 Obfuscation module

Obfuscation is a process of masking the original text into irrelevant text without using any key unlike encryption. This module uses the obfuscation technique proposed in (Arockiam *et al*., 2014). The proposed obfuscation technique in (Arockiam *et al*., 2014) uses mathematical functions and program logic to obfuscate the data. Generally, obfuscation technique is not using any key. The proposed technique in (Arockiam *et al*., 2014) generates a value during the process of obfuscation. This value is considered as key for deobfuscation. This key value is stored in the client side metadata and it is not communicated to cloud provider. Only numerical data are considered for obfuscation.

This obfuscation technique is not only ensures the security of the data but also reduces the size of the data sent to the cloud storage. For example, the obfuscated data size is reduced to 1 byte from its corresponding plaintext size of 4 bytes.

7.3 Storage procedure

In order to understand the proposed algorithms in AROMO Framework, let us consider a sample transactional table where the data is stored in the cloud storage as shown in the Table 1. The data are encrypted and obfuscated by the proposed algorithms (Arockiam *et al*., 2013, 2014).

Table 1 Transactional table with plain text

Trans_Id	Cust_Id	Item_Name	Quantity	Total_P
TId_1003	A230kum	Lux	20	200
TId_923	B301sus	Himalaya	17	1500
TId_2304	C100mon	Bovonto	3	145
TId_9087	B002lav	Laxmi	9	100
TId_0012	G123aro	Medicine	12	60
TId_9999	X987ren	Chocolate	50	15

Table 1 presents the plaintext data of a shopping mall which consists of Trans_Id, Cust_Id, Item_Name, Quantity and Total_P for six transactions. This given data are encrypted and obfuscated as shown in the table 2 below.

Table 2 Transactional table with cipher text using encryption and obfuscation

]zUdbju|p	L!pvL|xh	RymhQnwhy	Zz!dljx}&	]m\wb}ru{
]k<g3mL:?	J<y6x<5t>	U,{{kx|	2	@
]k?g5mLBU	K= 3x96|?	Qm&pdvlum	!	0
]k<g6mL;@	L<z3r94v=	K{{yw rw{	)	!
]kDg3mLBC	K>#3d93u<	Uy%{drdry	Q	0
]k=g3mL9>	P?{5u;4j=	Vuzglmhlq	2	0
]kEg<mLBE	aCz;hA<{E	Lomronkx!	F	C

It can be noted in table 2 that non-numeric data are encrypted and numeric data are obfuscated. Encrypted and obfuscated data in table 2 consumes less memory in comparison with the table 1. This scenario deduces the space consumption in the cloud storage and also increases the data security level in the cloud storage.

7.4 Metadata Repository

Metadata repository is created and maintained at data owner side. The information of encryption and decryption keys and obfuscation and deobfuscation keys are stored in metadata repository. Let us consider the table 1, the keys for encryption and decryption for non-numeric values and keys for obfuscation and deobfuscation for numeric values are shown in table 3.

Table 3: Metadata details in client side

Column Name	Transformed Column Name	Type of Confidentiality	Key to lock and unlock			
			K_1	K_2	K_3	K_4
Trans_Id	00	Encrypt	58	87	34	lKi
Cust_Id	01	Encrypt	43	67	23	NVu
Item_Name	02	Encrypt	23	34	67	AtR
Quantity	03	Obfuscate	1 1 0 0 0 9			
Total_P	04	Obfuscate	156 8789 82 39 14 0			

As shown in above table 3, encryption algorithm uses four keys; K_1, K_2, K_3 and K_4. Among these four keys K_1, K_2, K_3 are integer values and K_4 is a string. A key is generated during the process of obfuscation which is stored in metadata and it is used for deobfuscation. The metadata repository is particularly used for transformation of queries in order to preserve confidentiality. It is created in any text file. However, it is recommended to store in small DBMS like MS-Access to achieve better performance.

7.5 Query Processing

When users query to the cloud storage running on cloud provider end, their query is first transformed to execute on encrypted and obfuscated data. This transformation is performed on client side based on the information stored in metadata repository. After this transformation, query is executed on encrypted and obfuscated data without decryption or deobfuscation on server side. Once the required data is fetched, it will be transferred to client side for deobfuscation as well as for decryption. Decryption and deobfuscation is performed based on the information stored in metadata repository. Some queries may require further processing or filtering, and in that case original query will be executed again on the temporary result available at this stage before sending to user terminal

The proposed framework AROMO protects the data in cloud storage from the unauthorized privileged users from the cloud provider's side. The framework ensures the confidentiality of the data stored in the cloud storage.

8. Conclusion

Cloud provides reliable storage of data through maintaining multiple copies of data in different cloud data centers. Data are controlled and monitored by different untrusted privileged providers. Data in the cloud is hacked by internal or external attack. Internal attack is very difficult to protect. This paper has presented a framework to preserve confidentiality of data stored in CS. The proposed framework uses encryption and obfuscation technique to ensure the confidentiality. As far as data storage is concerned, a framework named AROMO is proposed to encrypt non-numerical data through an encryption algorithm on client side before sending to CS. Similarly, it also describes to obfuscate numeric type data through obfuscation technique on client side before sending to CS. Moreover the proposed AROMO framework focuses on query over encrypted and obfuscated data. All user queries are transformed on client side to execute over encrypted and obfuscated data stored in CS. Once the required data is filtered out, it is sent to the client side where it is decrypted and deobfuscated. Integration of obfuscation with encryption ensures the confidentiality and reduces the size of the data sent to Cloud Storage. Thus, AROMO will definitely improve the usage of cloud storage by data outsourcing.

References

Pierangela Samarati and Sabrina De Capitani di Vimercati (2010), Data protection in outsourcing scenarios: issues and directions, *In Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security*, pp. 1-14.

Francesco Pagano (2011), A Distributed Approach to Privacy on the Cloud, University of Milan - 26013 Crema, Italy.

Mell, P., & Grance, T. (2009), The NIST Definition of Cloud Computing, from NIST Information Technology Laboratory, http://www.nist.gov/itl/cloud/upload/cloud-def-v15.pdf, retrieved on may 2011.

Ross A. Lumley (2010), Cyber Security and Privacy in Cloud Computing: Multidisciplinary Research Problems in Business, the George Washington University, Report GW-CSPRI-2010-4, December, pp. 1-10.

L. Arockiam, S. Monikandan, G.Parthasarathy (2011), Cloud Computing: A Survey, *International Journal of Internet Computing*, ISSN No: 2231 – 6965, Volume-1, Issue-2, pp. 26-33.

Michael Miller (2010), Cloud Computing: Web-Based Applications That Change the Way You Work and Collaborate Online, *QueKnowledge and Grids*, pp. 105-112.

Minqi Zhou *et al*. (2008), Security and Privacy in Cloud Computing: A Survey, In *Proceedings of Sixth International Conference on Semantics, Publications,* First Printing, Augutst, pp. 149-150.

Damiani E., De Capitani di Vimercati S., Foresti S., Jajodia S., Paraboschi S., and Samarati P. (2005), Metadata management in outsourced encrypted databases, Springer-Verlag, Lecture Notes in Computer Science, *In Proceedings of the 2nd VLDB Workshop on Secure Data Management*, Trondheim, Norway, September, pp. 1-17.

Fatima Trindade Neves, Fernando Cruz Marta, Ana Maria Ramalho Correia, Miguel de Castro Neto (2011), The Adoption of Cloud Computing by SMEs: Identifying and Coping with External Factors, *11ª Conferência da Associação Portuguesa de Sistemas de Informação*, October, pp. 1-11.

Atiq ur Rehman, M.Hussain (2011), Efficient Cloud Data Confidentiality for DaaS, *International Journal of Advanced Science and Technology*, Vol. 35, October, pp. 1-10.

Richard Chow, Philippe Golle, Markus Jakobsson, Ryusuke Masuoka, Jesus Molina ,Elaine Shi, Jessica Staddon (2009), Controlling Data in the Cloud: Outsourcing Computation without Outsourcing Control, *In Proceedings of the ACM workshop on Cloud computing security*, November, pp. 85-90.

Jyun-Yao Huang, I-En Liao (2012), A Searchable Encryption Scheme for Outsourcing Cloud Storage, *COMNETSAT '12, IEEE*, pp. 142-146.

Kamara S. and Papamanthou P. (2006), Parallel and Dynamic Searchable Symmetric Encryption, *In Proceedings of the 13th ACM conference on Computer and communications security*, ACM, NY, USA, October, pp. 79 – 88.

Kamara S., Papamanthou C. and Roeder T. (2012), Dynamic Searchable Symmetric Encryption, *In Proceedings of the 2012 ACM conference on Computer and communications security*, pp. 965-976.

Kamara S., Papamanthou C. (2011), CS2: A Searchable Cryptographic Cloud Storage System, *T Roeder Microsoft Research*, Tech Report MSR-TR-2011-58, pp. 1-25.

Reza Curtmola, Juan Garay, Seny Kamara, Rafail Ostrovsky (2006), Searchable symmetric encryption: improved definitions and efficient constructions, *In Proceedings of the 13th ACM conference on Computer and communications security*, ACM, pp. 79-88.

Monique Ogburn, Claude Turner, Pushkar Dahal (2013), Homomorphic Encryption, *Elsevier journal, Procedia Computer Science*, Volume 20, pp. 502 – 509.

Kristin Lauter, Michael Naehrig, Vinod Vaikuntanathan (2011), Can Homomorphic Encryption be Practical?, *In Proceedings of the 3rd ACM workshop on Cloud computing security workshop*, pp. 113-124.

Stanislaw Jarecki, Charanjit Jutla, Hugo Krawczyk, Marcel Rosu, Michael Steiner (2013), Outsourced Symmetric Private Information Retrieval, *In Proceedings of the ACM SIGSAC conference on Computer & communications security*, pp. 875-888.

Yizhou Huang, Ian Goldberg (2013a), Outsourced Private Information Retrieval, *In Proceedings of the 12th ACM workshop on Workshop on privacy in the electronic society*, , pp. 119-130.

Huang Y. and Goldberg I. (2013b), Outsourced Private Information Retrieval with Pricing and Access Control, *Technical Report 2013-11*, CACR, 2013, http://cacr.uwaterloo.ca/techreports /2013/cacr2013-11.pdf.

Varun Maheshwari, Arash Nourian and Muthucumaru Maheswaran (2012), Character-based Search with Data Confidentiality in the Clouds, *IEEE 4th International Conference on Cloud Computing Technology and Science*, pp. 895-899.

Kamara S., Lauter K. (2010), Cryptographic Cloud Storage, *IFCA/ LNCS 6054, Springer-verlag, Berlin Heidelberg*, pp. 136-149.

Miranda Mowbray and Siani Person (2009), A client based privacy manager for Cloud Computing *ICST COMSWARE*.

L. Arockiam, S. Monikandan (2013), Data Security and Privacy in Cloud Storage using Hybrid Symmetric Encryption Algorithm, *International Journal of Advanced Research in Computer and Communication Engineering*, Vol. 2, Issue 8, August, pp 3064-3070.

L. Arockiam, S. Monikandan, P. D. Sheba K Malarchelvi (2014), Obfuscrypt: A Novel Confidentiality Technique for Cloud Storage, *International Journal of Computer Applications (0975 – 8887)*, Volume 88, No.1, pp. 17-21.