RESEARCH ON DATA SECURITY ISSUES OF CLOUD COMPUTING

Chaoqun Yu[1],Lin Yang[2],Yuan Liu[1],Xiangyang Luo[1,2]

[1]Zhengzhou Information Science and Technology Institute, Zhengzhou 450002, China;
[2]China Institute of Electronic Equipment System Engineering, Beijing 100141, China.
Email: xiangyangluo@126.com

Abstract

Data security issue is a key bottleneck restricting the application of cloud computing promoting and applications. In this paper, states of the art of the techniques on cloud computing data security issues, such as data encryption, access control, integrity authentication and other issues is surveyed, on this basis, some important technical issues of the cloud computing data security should concern about and focus on are indicated.

Keywords: Cloud computing; data security; Risks assessment; State of the art.

1 Introduction

Service-based cloud computing is an important form of the information infrastructure in the Internet era, which adopts new business model to provide high-performance, low-cost computing and data services, supporting all kinds of informatization application. Along with the rapid development and application of this new network technology, new security problems appear constantly, and become an important factor in restricting industrial development. Coupled with the popularization of cloud computing, and the deepening understanding of cloud computing, the security issue has become the biggest concern in the using cloud computing and the migration to cloud computing. If the bottleneck problem of cloud computing security cannot be resolved, cloud computing technology is difficult to carry out the industrial upgrading and application promotion. Following cloud computing security research in the academic community, some researchers have begun to pay attention to cloud computing security issues, but the vast majority of literature still remain in the research of cloud computing deployments, services, applications and other related issues, and depth research on cloud computing security issues has not yet commenced, and the key security issues like data privacy protection related to cloud computing are still lack of support of basic theory and effective technology[1].

Currently, cloud computing security issues have obtained increasing attention. The International Conference RSA on information security listed cloud computing security as focal issues. Many research institutions, enterprises and standardization organizations have launched related research. Security vendors are concerned about and developing various types of security cloud computing products. Berkeley Cloud Computing White Paper[2] sets out 10 issues and opportunities facing Cloud Computing, in which, related security issues include data loss, the security and auditability of data, and the virtualization security. These show the important status of the data security in a series of security issues of cloud computing.

In terms of data loss, the software sets have already been improved a lot in the cross-platform, but essentially, API (Application Programming Interface) of cloud computing is still private, or an uniform standard is not established currently. So, it's difficult for users to migrate their data and programs from one station to another, and this is also the reasons that a lot of users do not want to use cloud computing. For the auditability of data, current cloud offers the public network essentially, and would be suffered more attacks. So it's difficult for users to put sensitive data into the cloud.

In terms of research in the data security of cloud computing, the article [3] also believes that data security is an important security issues of cloud computing. Handing the data originally stored in the local control to an external service center of cloud computing is not easy. However, as money has long been accustomed saving in the bank, in the future, data Bank will certainly appear, sooner or later. Technique may not be the main stumbling block, and institutions, laws, Integrities, habits and ideas, these non-technical factors will directly influence the popularity of cloud computing. the article [4] investigates the risks of cloud computing, and puts forward a system design scheme can be used to capture information flow in the cloud for whether users of cloud computing could use their own information at any time, and how to prevent their information to be illegally obtained. Providing safe and effective access in large-scale off-site data is an important component of cloud computing, and the article [5] gives a different key to encrypt the data block to provide a mechanism for the access control based on the elastic encryption, against the security issues in this mode of data owner "Write" - user "read". The article [6] discussed the cloud storage, raised cloud storage architecture assumptions and involved issues, including storage security, but did not give a corresponding solution. The article [7] and [8] analyzed

privacy, security and the issue of credibility of cloud computing, and discussed a number of methods which can enhance the credibility and security.

From the existing literatures, as the applications and services model of cloud computing is different from the traditional end-to-end implementation of the encrypted communication to ensure data security, an untrusted third party will participate the process of virtualization storage and processing of massive data, this brings new data security issues. Certainly, the encryption technology is still a powerful measure to ensure data security of cloud computing. But how to implement the highly efficient encryption, how to quickly search on the data encrypted, how to carry on disaster recovery and fast recovery, how to proceed the access control of data and so on, all of above are a series of key and difficult problems that data security of cloud computing must resolve.

2. Research Progress of Related Technologies on Data Security Issues of Cloud Computing

In cloud computing environments, the usual data transmission and storage mode is shown in Figure 1:

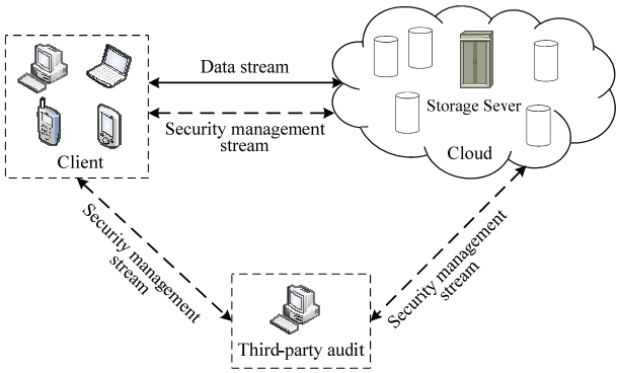

Figure 1. Data storage structure of cloud computing

The effective protection of user privacy data is the primary problem of cloud security. From the above figure, we can see that cloud computing stores a large number of data files on the distant servers, and users can reduce the burden of storage and computation, then enjoy the flexible and efficient service brought by cloud computing. But the characteristics of cloud storage make users' data faced with many security risks, includes: (1) the traditional security region partition is invalid. Because of the cloud storage service must be scalable, security boundaries and protection equipment cannot be clearly defined, which increases some difficulty for the implementation of specific protection measures; (2) the cloud storage transmits data through the network. The service interruptions, data destruction, information stolen and tampered caused by the malicious attacks in the network pose a severe challenge to the security of data communications, access authentication and confidentiality; (3) from the user's view, the cloud storage of data makes cloud computing service provider obtains the data access control, and the user's data is faced with privacy security threats. People worry about that the sensitive personal data will be disclosure, misuse or missing by putting the data in cloud environment. To solve the above issues, in recent years, researchers made a lot of research work in the data security access control mechanisms, data integrity, authentication, ciphertext to retrieve and data encryption technique of cloud computing environment.

2.1 Data security access control mechanisms of cloud computing environment

As the cloud service requires secure cross-domain collaboration, and needs to provide protection for being mistaken of personnel and equipment identity, users need to have the convenient and comprehensive ability to control their own data. If not to access all resources like an interoperable stand-alone system, it will affect the usability of cloud. Therefore, cloud service providers should provide appropriate mechanisms to support users with regulation, authorization and access control on their own data, and let the user know whether there are other users to access or copy data stored in the cloud.

Secure identity management and control are essential for any network environment, but it will become more complex in the cloud computing environment. Cloud Security Alliance (CSA) believe that the management of authentication and access control for company is still the biggest challenge facing IT department, however, the existing cloud service providers themselves didn't comprehensively place authentication service in their cloud computing platform. The implementation of sound data classification mechanism can determine data sets involved in the things, and determine the control mechanism applied to the data set in a particular case. Related research progress can be divided into two respects.

(1) Access Control based on Virtualization Technology

As virtualization technology is becoming the core technology of cloud computing, security control for cloud computing is concentrated on virtual machine-based access control. Currently it mainly uses strong access control mechanisms to achieve isolation in the communication of virtual machines. sHype implemented strong access control module in the Xen Virtual Machine Manager, and using this module can control communications between multiple virtual machines on a single physical node. IBM proposed Trusted Virtual Domain (TVD) [9], and the security control for TVD is the control for the inter-domain communication. Payne et al. proposed a hierarchical access control model[10], and for the complex problem of the virtual machine-based access control strategy, they proposed a hierarchical classification framework to simplify it.

(2)Cross-domain access control

When the users cross-domain access resources they need to set up certification services in the domain boundaries, and make a unified identity management for accessing to shared resources. Since each trusted domain has its own access policy, so it needs to support the synthesis of strategies. It is first proposed by Mclean in mandatory access control framework, and synthesizes two security grids to a new grid structure. The synthesis of strategies while also guarantees the security of the new strategy, and

the new synthetic strategies must not be contrary to the original access control policies in various domains. In this aspect, Bonatti[11] proposed an algorithm for synthesis of access control policies, it use synthesis operator to synthesize the security policy based on set theory. Wijesekera et al.[12] proposed a synthesis algebraic framework of a strategy based on the license status changes.

2.2 Data integrity certification of cloud computing environment

In cloud computing environment, users do not have to reserve the data locally, but they must be convinced that their own data in the cloud can be well preserved and maintained. So users need a safe way to regularly verify the correctness of their own stored data. Users have not enough time, ability or resources to manage their data, thus they put this task entrusted to a trust TPA. This brought a lot of security challenges unresolved, in which the most concerned is the validation issues of the integrity of data stored on untrusted servers. The existing research achievements include: Provable Data Possession (PDP)[13][14] and Proof of Retrievability (POR)[15] model.

Atenises et al. define the Provable Data Possession (PDP) model to confirm the existence of the file on the server untrusted. Their programs audit data with RSA-based homomorphic tags. So they can provide public verifiability. But Atenises did not consider the dynamic data storage, and the using may reveal the users' data information. Since then, Atenises et al. also proposed a dynamic version of the PDP[14]. But this system has confined of inquired number, and cannot support dynamic data manipulation, for example, it does not support insertion of data blocks.

Juels et al. define the Proof of Retrievability (PoR) model, using spot-checking and error-correcting codes to ensure persistence and recoverability of data. Specifically, some special data blocks called sentinels will be embedded into the data file F for the purpose of detection, and then file F will be encrypted to hide the position of these special blocks. However, this method also only supports a limited inquiry, and special precomputation hinder the dynamic data updating. Shacham[16] designed an improvement scheme of PoR, and gives a complete proof of security in the security model defined in paper [17], but still only considered the static data files. Gellman[16] gives an exploratory construct for dynamic provable data possession, and they improved PDP model mentioned in paper [13] to make it supports provable data possession. Particular, in order to support the update of the data, they did not use the method of calculation of the index information in the original model, but before certification process, use the authenticated skip list data structure to authenticate the tag information to updating data blocks.

2.3 Ciphertext retrieval technology

A common method to solve the problem of data protection is users encrypt data, and then put ciphertext into the server. When the encrypted data stored in the cloud formed the scale, retrieval of encrypted data become an urgent problem which needs to be solved. Under normal circumstances, the ciphertext is not available for retrieval semantic and statistical properties. So retrieval of ciphertext is a more difficult problem. The existing ciphertext retrieval methods mainly include Linear searching method, Public key based on keyword searching method, Security index searching method, Order preserving encrypted searching method and so on.

(1) Linear searching method

Song et al.[18] first proposed a linear search algorithm for ciphertext data. They first use the symmetric encryption algorithm to encrypt the plaintext messages. For each ciphertext information corresponding with each keyword, it generates a bunch of pseudo-random sequence, and generates a check sequence determined by the pseudo-random sequence and the ciphertext. Sum of the Pseudo-random sequence length and the test sequence length is equal to the length of the ciphertext information. Pseudo-random sequence and test sequence encrypted the ciphertext again. When searching, the user submits the plaintext information to be searched, and then searching method calculates the corresponding ciphertext sequence. On the server side, the ciphertext message sequence is linear model 2 plus with each sequence in the searching range. If the obtained results meet the parity relations, it shows that the ciphertext information sequence matches successfully,. Linear searching method is a one-time pad, extremely resistant to statistical analysis. However, such methodology has an obvious drawback., which needs to match ciphertext information successively, and the time complexity is very high, then it's difficult to apply into large data sets searching.

(2) Public key based on keyword searching method

Boneh et al.[19] proposed the public key based on keyword searching method. This method can access the remote database to obtain data with the lack of client storage and computing resources. The encrypted data has a number of different sources, and it needs to make a search for such encrypted data. A viable idea is that data is encrypted with public key encrypt. It first generates a public key and a private key, and then encrypts plaintext keywords which will be stored with the public key to generate the ciphertext can be used to search. During the searching, it encrypts the plaintext sequence provided by the user intended to search with the public key, and then carries out ciphertext keyword matching.

(3) Security index searching method

Boneh et al. proposed the security index searching method, to solve the disadvantage that simple index is vulnerable to statistical attacks. The mechanism is that the key used for per encryption is a set of pre-generated inverse Hash sequence, and encrypted index is being put in Bloom filter. When searching, it first uses the inverse hash sequence key to generate a trapdoor, and then does a Bloom detection. The returned ciphertext is the document required. The shortcoming of this method is the need to generate a large number of key sequences. With the increase of the number of searches, the computing complexity degree increases linearly. In the above searching methods for encrypted information, searching models are all the Boolean model, and thus it cannot do a sorting operation based on the relevance of the retrieved documents. In the actual situation, especially in the cloud

storage applications with larger data size, there may be many documents containing a same query keyword. How to identify the most relevant one or several documents in a number of possible document needs to be addressed.

(4) Order preserving encryption searching method

Swaminathan et al.[20] proposed the order preserving encrypted searching method. In this method, the term frequency of each keyword in the document is encrypted by order preserving encryption algorithm. After the encrypted documents that users need to query are submitted to the server, the server first searches encrypted document containing keywords ciphertext, and then makes a sort treatment for corresponding ciphertext of frequency encrypted with order preserving algorithm. Finally, the encrypted document with a high evaluation will be returned to users. This method can sort encrypted document in the case of given many relational documents, and thus returns the most relational document to users. However, this method does not apply to the query that contains a number of query terms, because sorting method does not know which query terms according to.

2. 4 Study of data encryption technology in cloud computing environment

IBM Fellow C. Gentry, at ACM International Symposium on Theory of Computing (STOC), published a papers entitled "Fully homomorphic encryption using the ideal lattice"[21], solved the fully homomorphic encryption problem proposed by well-known cryptographer R. the Rivest and L. Adlemanand 30 years ago. The paper's publishing not only generates a great sensation in academia, but also has a major impact on industry. Fully homomorphic encryption technology was known as the title of "the holy grail of cryptography". After Gentry's paper published, fully homomorphic encryption once again has become the hot issue in the field of cryptography nearly two years. In the top three cryptology annual meetings (CRYPTO, EUROCRYPT, ASIACRYPT) in the field of cryptology, the cryptology scientists have researched fully homomorphic encryption on the basis of the work done by Gentry and made a lot of new results[22-32]. Now large companies such as IBM and Google are pushing fully homomorphic encryption technology to the practical application, and apply it to respective system.

The basic principle of fully homomorphic encryption algorithm is as follows, note the encryption operation as E, plaintext as m, and get e after Encryption. That is, $e = E(m)$, $m = E'(e)$. Known the plaintext operation f, we can construct F for E, satisfying $F(e) = E(f(m))$. So E is a homomorphic encryption algorithm for f. It assumes that f is a very complex operation. With homomorphic encryption, the sender can send the e encrypted to a third party. The third party does an operation F. The sender gets back the $F(e)$, get $f(m)$ after decryption. The third party completes the work on behalf of the sender, and still knows nothing about the m. This is a surprising result! However, looking for such an E is not easy. Purely from the view of mathematic, $E(x) = x$, is homomorphic. But unfortunately there is no cryptographic effectiveness. The RSA algorithm is homomorphic for the multiplication operation. The corresponding operation F is also a multiplication. Others such as the addition will not be able to construct the corresponding F. But the Paillier algorithm is homomorphic for addition.

If an encryption algorithm that multiplications and additions all can find the corresponding operation, this encryption algorithm will be fully homomorphic. Fully homomorphic encryption technology gets an output with the processing for encrypted data. The result of decrypting the output is same to the output using the same approach for the original, un-encrypted data. In other words, fully homomorphic encryption technology can construct the corresponding cryptographic operations for arbitrarily complex expressly operation. Special significance of the homomorphic encryption is that it can construct the corresponding F for any f. In this way, you can get some incredible applications. I can solve your problem, even though I do not know your problems-this is a less proper metaphor.

We can completely believe that if we really achieve a mature fully homomorphic encryption, and the loss of encryption efficiency compared with classical encryption algorithms is not too much, it will be perfectly used in cloud computing environments, and the range of applications will be very extensive.

3. Problems should be concerned about

(1) Cloud computing data security system

For the data security problems cloud computing brought, how to implement total life cycle management to cloud computing data security is an important research direction. How to give out a more complete cloud computing data security technology system is worthy study, and it is also a key step to clear the problem needed to be resolved in cloud computing data security. In addition to study data security system of data life cycle, there is an important research idea at this stage, which is to weaken or transfer the outstanding existing security risks. For example, the most worried thing for the existing cloud computing users is whether system administrators could spy on users' data.

(2) Authentication and access control

Authentication and access control is a common network security measures. Compared with traditional network service model, cloud computing is more flexible, and it has higher requirements on authentication and access control. In particular, cloud service providers should focus on the following security technologies in authentication and access control: 1) SaaS Account Management; 2) SaaS Collection and analysis of safety records; 3) PaaS application access policy; 4) Control technology of privileged user in the IaaS; 5) Monitoring and auditing techniques.

(3) Data integrity and availability certification

Huge communication cost caused by the large-scale data makes it impossible for the user to verify the integrity and availability after the data is completely downloaded from the cloud. Thus, cloud users must judge with high confidence probability whether the cloud data is complete through some kind of knowledge proof agreement or probability analysis means under the case of getting few data. In terms of cloud storage protocol design, the main

considerations currently are data integrity authentication, data error positioning, dynamic updating of data, etc. But the design is not perfect, for example, the dynamic data operation of the protocol is only at data block level.

(4)Practical homomorphic encryption technology

The homomorphic encryption technology is of revolutionary significance for solving the problem of data security of cloud computing. Once the practical full-homomorphic encryption technology appears, data security issues plagued cloud computing applications will be fundamentally resolved. Therefore, it is foreseeable that the practical full-homomorphic encryption technology has been one of the focus issues of cloud computing security research, until it be completely resolved.

4. Conclusions

The security issue has become a bottleneck restricting cloud computing industrial applications. This article focuses on the important issue of cloud computing, data security. The research progress of issues of data encryption, access control, integrity authentication and so on with respect to cloud computing data security is surveyed. On this basis, we point out the key technologies and key issues the cloud computing data security issues should be concerned about. Overall, the research of cloud computing security is at the initial stage of development. In terms of cloud computing data security, there are still a large number of key issues to be studied in depth.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (No. 61379151, 61373020), and the Excellent Youth Foundation of Henan Province of China (No. 144100510001), and the Foundation of Science and Technology on Information Assurance Laboratory (No. KJ-14-108).

References

[1] D. Feng, M. Zhang, Y. Zhang, Z. Xu. "Cloud Computing Security Research", *Chinese Journal of Software*, 22(1): 71-83, 2011.

[2] Above the Clouds: A Berkeley View of Cloud Computing. http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28. pdf

[3] J. Yao. "The Future Needs Cloud Computing", *Development and Application of High Performance Computing*, 2009, 26(1): 7-9.

[4] L. Sumter. "Cloud Computing: Security Risk", Proceedings of *ACM Southeast Conference*, pp. 1-4, 2010.

[5] W. Wang, Z. Li, R. Owens, B. Bhargava. "Secure and Efficient Access to Outsourced Data", Proceedings of *ACM Cloud Computing Security Workshop*, pp. 55-65, 2009.

[6] W. Zeng, Y. Zhao, K. Ou, Wei Song. "Research on Cloud Storage Architecture and Key Technologies", Proceedings of *ACM International Conference on Information System*, pp. 1044-1048, 2009.

[7] H. Takabi, B. D. James, A. Gail-Joon. "Security and Privacy Challenges in Cloud Computing Environments", *IEEE Security & Privacy*, 10(6): 24-31, 2010.

[8] S. Pearson, A. Benameur. "Privacy, Security and Trust Issues Arising from Cloud Computing", Proceedings of IEEE International Conference on *Cloud Computing Technology and Science*, pp. 693-702, 2010.

[9] IBM Research. http://domino.research.ibm.com/comm/research_projects.nsf/pages/ssd_tvd.index.html

[10] B. D. Payne. "Improving Host-Based Computer Security Using Secure Active Monitoring and Memory Analysis", A Thesis Presented to The Academic Faculty, *Georgia Institute of Technology*, 2010.

[11] P. Bonatti, S. Vimercati, P. Samarati. "An Algebra for Composing Access Control Policies", *ACM Transactions on Information and System Security*, 5(1): 1−35, 2002.

[12] D. Wijesekera, S. Jajodia. "A Propositional Policy Algebra for Access Control", *ACM Transactions on Information and System Security*, 6(2): 286−325, 2003.

[13] G. Ateniese, R. Burns, R. Curtmola, et al. "Provable Data Possession at Untrusted Stores", Proceedings of 14th *ACM Conference on Computer and Communications Security*, pp. 598–609, 2007.

[14] G. Ateniese, R. D. Pietro, L. V. Mancini, and G. Tsudik. "Scalable and efficient Provable Data Possession", Proceedings of the Conference on *Security and Privacy in Communication Networks*. Doi: 10.1145/1460877.1460889, 2008.

[15] A. Juels, B. S. Kaliski. PORs: "Proofs of Retrievability for Large Files", Proceedings of 14th *ACM Conference on Computer and Communications Security*, pp. 584–597, 2007.

[16] H. Shacham, B. Waters. "Compact Proofs of Retrievability", Proceedings of *ASIACRYPT*'08, pp. 90-107, 2008.

[17] R. Gellman. "Privacy in the Clouds: Risks to Privacy and Confidentiality from Cloud Computing", *Report*, February 23, 2009.

[18] D. Song, D. Wagner, A. Perrig. "Practical Techniques for Searches on Encrypted Data", Proceedings of *the IEEE Symposium on Security and Privacy*, pp. 44-55, 2000.

[19] D. Boneh, G. Crescenzo, R. Ostrovsky, et al. "Public Key Encryption with Keyword Search", Proceedings of the 23rd Annual International Conference on the Theory and Applications of *Cryptographic Techniques (EUROCRYPT'04)*, Lecture Notes in Computer Science, vol. 3027, pp. 506-522, 2004.

[20] Ashwin Swaminathan, Yinian Mao, Guan-Ming Su, et al. "System and method for confidentiality-preserving rank-ordered search", US patent: 20100146299 A1.

[21] C. Gentry. "Fully Homomorphic Encryption Using Ideal Lattices", Proceedings of the *ACM International Symposium on Theory of Computing*, pp. 169−178. 2009.

[22] D. Stehle, R. Steinfeld. "Faster Fully Homomorphic

Encryption", Proceedings of *ASIACRYPT 2010*, *Lecture Notes in Computer Science*, vol. 6477, pp. 377-394, 2010.

[23] M. Dijk, C. Gentry, S. Halevi, V. Vaikuntanathan. "Fully Homomorphic Encryption over the Integers", Proceedings of EUROCRYPT 2010, Lecture Notes in *Computer Science*, vol. 6110, pp. 24-43, 2010.

[24] C. Gentry. "Toward Basing Fully Homomorphic Encryption on Worst-Case Hardness", Proceedings of CRYPTO 2010, Lecture Notes in *Computer Science*, vol. 6223, pp. 116-137, 2010.

[25] C. Aguilar, P. Gaborit, J. Herranz. "Additively Homomorphic Encryption with d-Operand Multiplications", Proceedings of CRYPTO 2010, Lecture Notes in *Computer Science*, vol. 6223, pp. 138-154, 2010.

[26] C. Gentry, S. Halevi, V. Vaikuntanathan. "i-Hop Homomorphic Encryption and Rerandomizable Yao Circuits", Proceedings of CRYPTO 2010, Lecture Notes in *Computer Science*, vol. 6223, pp. 155-172, 2010.

[27] K. Chung, Y. Kalai, S. Vadhan. "Improved Delegation of Computation Using Fully Homomorphic Encryption", Proceedings of CRYPTO 2010, Lecture Notes in *Computer Science*, vol. 6223, pp. 483-501, 2010.

[28] C. Centry, S. Halevi. "Implementing Gentry's Fully-Homomorphic Encryption Scheme", Proceedings of EUROCRYPT 2011, Lecture Notes in *Computer Science*, vol. 6632, pp. 129-148, 2011.

[29] D. Boneh, D. Freeman. "Homomorphic Signatures for Polynomial Functions", Proceedings of EUROCRYPT 2011, Lecture Notes in *Computer Science*, vol. 6632, pp. 149-168, 2011.

[30] R. Bendlin, I. Damgard, C. Orlandi, S. Zakarias. "Semi-Homomorphic Encryption and Multiparty Computation", Proceedings of EUROCRYPT 2011, Lecture Notes in *Computer Science*, vol. 6632, pp. 169-188, 2011.

[31] Z. Brakerski, V. Vaikuntanathan. "Fully Homomorphic Encryption from Ring- LWE and Security for Key Dependent Messages", Proceedings of CRYPTO 2011, Lecture Notes in *Computer Science*, vol. 6841, pp. 505-524, 2011.

[32] J. Coron, A. Mandal, D. Naccache, M. Tibouchi. "Fully Homomorphic Encryption over the Integers with Shorter Public-Keys", Proc. of CRYPTO 2011, Lecture Notes in *Computer Science*, vol. 6841, pp. 487-504, 2011.