Secured Data Outsourcing in Cloud Computing

Rajendra N. Kankrale[*], **Mahendra B. Gawali**

IT Dept, SRES COE, SPPU, Pune.

Abstract- Cloud computing has gained a lot of publicity in the current IT world. After the internet, Cloud computing is the next big thing in the computer world. Cloud computing is the use of the Internet for the tasks performed on the computer and it is the next- generation architecture of IT Industry. Cloud computing is related to different technologies and the convergence of various technologies has emerged to be called cloud computing. Cloud Computing moves the application software and databases to the large data centers, where the management of the data and services may not be fully trustworthy. This unique attribute, however, poses many new security challenges which have not been well understood. In this paper, we discuss the design mechanisms that not only protect sensitive data with encrypted data, but also protect customers from malicious behaviors by enabling the validation of the computation result.

Index Terms- Cloud Computing, SaaS, PaaS, IaaS,EC2

I. Introduction

Cloud computing is the use of computing resources (hardware and software) that are delivered as a service over a network (typically the Internet). The name comes from the use of a cloud-shaped symbol as an abstraction for the complex infrastructure it contains in system diagrams. Cloud computing entrusts remote services with a user's data, software and computation. Cloud computing is a general term for anything that involves delivering hosted services over the Internet. A Cloud service has three distinct characteristics that differentiate it from traditional hosting. It is sold on demand, typically by the minute or the hour; it is elastic - a user can have as much or as little of a service as we want at any given time; and the service is fully managed by the provider. A Cloud can be private or public. A public cloud sells services to anyone on the Internet. Currently, Amazon Web Services is the largest public cloud provider [6].

From the perspective of data security, which has always been an important aspect of quality of service, Cloud Computing inevitably poses new challenging security threats for number of reasons. First security threat is the traditional cryptographic primitives for the purpose of data security protection cannot be directly adopted due to the users' loss control of data under Cloud Computing[2]. Therefore, verification of correct data storage in the cloud must be conducted without explicit knowledge of the whole data. Considering various kinds of data for each user stored in the cloud and the demand of long term continuous assurance of their data safety, the problem of verifying correctness of data storage in the cloud becomes even more challenging. Another security threat is the data stored in the cloud may be frequently updated by the users, including insertion, deletion, modification, appending, reordering, etc. To ensure storage correctness under dynamic data update is hence of paramount importance[1]. Lastly, the deployment of Cloud Computing is powered by data centers running in a simultaneous, cooperated and distributed manner. Individual user's data is redundantly stored in multiple physical locations to further reduce the data integrity threats. Therefore, distributed protocols for storage correctness assurance will be of most importance in achieving a robust and secure cloud data storage system in the real world. However, such important area remains to be fully explored in the literature. The cloud computing have various service models which are given below.

In this paper, we only focus on the Software as a service. In this model, cloud providers install and operate application software in the cloud and cloud users access the software from cloud clients. The cloud users do not manage the cloud infrastructure and platform on which the application is running. This eliminates the need to install and run the application on the cloud user's own computers simplifying maintenance and support. What makes a cloud application different from other applications is its elasticity. This can be achieved by cloning tasks onto multiple virtual machines at run-time to meet the changing work demand. Load balancers distribute the work over the set of virtual machines. This process is inconspicuous to the cloud user who sees only a single access point. To accommodate a large number of cloud users, cloud applications can be multitenant, that is, any machine serves more than one cloud user organization. It is common to refer to special types of cloud based application software with a similar naming convention: desktop as a service, business process as a service, test environment as a service, communication as a service.

Several trends are opening up the era of Cloud Computing, which is an Internet-based development and use of computer technology. The ever cheaper and more powerful processors, together with the software as a service (SaaS) computing architecture, are transforming data centers into pools of computing service on a huge scale. The increasing network bandwidth and reliable yet flexible network connections make it even possible that users can now subscribe high quality services from data and software that reside solely on remote data centers.

Moving data into the cloud offers great convenience to users since they don't have to care about the complexities of direct hardware management. The pioneer of Cloud Computing vendors, Amazon Simple Storage Service (S3) and Amazon Elastic Compute Cloud (EC2) [5] are both well known examples. While these internet-based online services do provide huge amounts of storage space and customizable computing resources, this computing platform shift, however, is eliminating the responsibility of local machines for data maintenance at the same time. As a result, users are at the mercy of their cloud service providers for the availability and integrity of their data. Recent downtime of Amazon's S3 is such an example.

From the perspective of data security, which has always been an important aspect of quality of service, In Cloud Computing inevitably poses new challenging security threats for number of reasons. Firstly, traditional cryptographic primitives for the purpose of data security protection cannot be directly adopted due to the users' loss control of data under Cloud Computing. Therefore, verification of correct data storage in the cloud must be conducted without explicit knowledge of the whole data [7]. Considering various kinds of data for each user stored in the cloud and the demand of long term continuous assurance of their data safety, the problem of verifying correctness of data storage in the cloud becomes even more challenging. Secondly, Cloud Computing is not just a third party data warehouse. The users may be frequently update their data which is stored on the cloud. The updating includes insertion, deletion, modification, appending, reordering, etc. To ensure storage correctness under dynamic data update is more important. However, this dynamic feature also makes traditional integrity insurance techniques futile and entails new solutions. Last but not the least, the deployment of Cloud Computing is powered by data centers running in a simultaneous, cooperated and distributed manner. Individual user's data is redundantly stored in multiple physical locations to further reduce the data integrity threats. Therefore, distributed protocols for storage correctness assurance will be of most importance in achieving a robust and secure cloud data storage system in the real world.

The tremendous benefits, outsourcing computation to the commercial public cloud is also depriving customers direct control over the systems that consume and produce their data during the computation, which inevitably brings in new security concerns and challenges towards this promising computing model. On the one hand, the outsourced computation on the cloud server often contain sensitive information, such as the business financial records, proprietary research data, or personally identifiable health information etc[1].

To take action against unauthorized information leakage, sensitive data have to be encrypted before outsourcing. So as to provide end-to-end data confidentiality assurance in the cloud and beyond. However, ordinary data encryption techniques in essence prevent cloud from performing any meaningful operation of the underlying plaintext data, making the computation over encrypted data a very hard problem[1]. On the other hand, the operational details which are done inside the cloud are not transparent to customers. As a result, there do exist various motivations for cloud server to behave unfaithfully and to return incorrect results, i.e., they may behave beyond the classical semi honest model. For example, for the computations that require a large amount of computing resources, there are huge financial incentives for the cloud to be "lazy" if the customers cannot tell the correctness of the output [1]. Besides, possible software bugs, hardware failures, or even outsider attacks might also affect the quality of the computed results.

Thus, we can say that, with respect to the customers point of view the cloud is not secure. Without providing a mechanism for secure computation outsourcing, i.e., to protect the sensitive input and output information on the cloud servers and to validate the integrity of the computation result, it would be hard to expect cloud customers to turn over control of their cloud servers from local machines to cloud solely based on its economic savings and resource flexibility. For practical consideration, such a design should further ensure that customers perform fewer amounts of operations following the mechanism than completing the computations by themselves directly[1]. Otherwise, there is no point for customers to seek help from cloud. Recent researches in both the cryptography and the theoretical computer science communities have made steady advances in "secure outsourcing expensive computations".

II. LITERATURE SURVEY

According to "Non-interactive verifiable computing: Outsourcing computation to entrusted workers" the author R. Gennaro [4] said that, the work is based on the crucial (and somewhat surprising) observation that Yao's Garbled Circuit Construction, in addition to providing secure two-party computation, also provides a "one-time" verifiable computation. In other words, we can adapted Yao's construction to allow a client to outsource the computation of a function on a single input. More specifically, in the preprocessing stage the client garbles the circuit C according to Yao's construction. Then in the "input preparation" stage, the client reveals the random labels associated with the input bits of x in the garbling. This allows the worker to compute the random labels associated with the output bits, and from them the client will reconstruct F(x). If the output bit labels are sufficiently long and random, the worker would not be able to guess the labels for an incorrect output, and therefore the client is assured that F(x) is the correct output.

According to "Secure outsourcing of sequence comparisons", the author M. J. Atallah said that, we now more precisely stateted the edit distance problem, in which the cost of an insertion or deletion or substitution is a symbol-dependent non-negative weight, and the edited distance then the least-cost set of insertions, deletions, and substitutions required to transform one string into the other. More formally, if we let λ be a string of length n, λ = λ1 . . . λn and μ be a string of lengthm, μ = μ1 . . . μm, both over some alphabet ∑. There are three types of allowed edit operations to be done on λ: insertion of a symbol, deletion of a symbol, and substitution of one symbol by another [5]. Each operation has a cost associated with it, namely I(a) denotes the cost of inserting the symbol a, D(a) denotes the cost of deleting a, and S(a, b) denotes the cost of substituting a with b. Each sequence of operations that transforms λ into μ has a cost associated with it and the least-cost of such sequence is the edit-distance. The edit path is the actual sequence of operations that corresponds to the edit distance. According to “Secure outsourcing of scientific computation”, the authors M. J. Atallah et. al. said that [9], they produced the first investigation of secure outsourcing of numerical and scientific computation. A set of problem dependent disguising techniques are proposed for different scientific applications like linear algebra, sorting, string pattern matching, etc. However, these disguise techniques explicitly allow information disclosure to certain degree. Atallah et al. discuss in the paper [10] and [11], produced two protocol designs for both secured sequence comparison outsourcing and secured algebraic computation outsourcing. However, both protocols used heavy cryptographic primitive such as homomorphic encryptions and/or oblivious transfer and do not scale well for large problem set.

In addition, both designs are built upon the assumption of two non-colluding servers and thus vulnerable to colluding attacks. Based on the same assumption, the authors, Hohenberger et al. [3] provide protocols for secure outsourcing of modular exponentiation, which is considered as prohibitively expensive in most public-key cryptography operations. Very recently, Atallah et al. [5] given a provably secure protocol for secure outsourcing matrix multiplications based on secret sharing. While this work outperforms their previous work [11] in the sense of single server assumption and computation efficiency, the main drawback is the large communication overhead. Namely, due to secret sharing technique, all scalar operations in original matrix multiplication are expanded to polynomials, introducing significant amount of overhead. Considering the case of the result verification, the communication overhead must be further doubled, due to the introducing of additional pre-computed "random noise" matrices.

Another large existing list of work that relates to (but is also significantly different from) Secure Multi-party Computation (SMC), first introduced by Yao [11] and later extended by Goldreich et al. [1] and and many others. SMC allows two or more parties to jointly compute some general function while hiding their inputs to each other. As general SMC can be very inefficient, Du and Atallah et. al. have proposed a series of customized solutions under the SMC context to a spectrum of special computation problems, such as privacy-preserving cooperative statistical analysis, scientific computation, geometric computations, sequence comparisons, etc. [14]. However, directly applying these approaches to the cloud computing model for secure computation outsourcing would still be problematic. With the major reason is that they did not address the asymmetry among the computational powers possessed by cloud and the customers, i.e., all these schemes in the context of SMC impose each involved parties comparable computation burdens, which we specifically avoid in the mechanism design by shifting as much as possible computation burden to cloud only. Another reason is the asymmetric security requirement. In SMC no single involved party knows all the problem input information, making result verification a very difficult task. But in our model, we can explicitly exploit the fact that the customer knows all input information and thus design efficient result verification mechanism.

Author	Method	Publication Year	Remark
R. Gennaro, C. Gentry, B. Parno[4]	Yao's Grabled Circuit Construction	2010	This technique is not practical due to its huge computation complexity
M. J. Atallah, Jiangtao Li[16]	Simplex Algorithm	2006	Series of interactive cryptographic protocols collaboratively exevuted in each iteration step.
S. Hohenberger, A. Lysyanskaya[3]	CCA2	2005	The general idea given in this paper is securing outsourcing of modular exponetiation, it is very expensive in public key cryptographic operation.
M. J. Atallah J. Li [10]	Edit Distance Protocol	2005	This protocol efficient for a customer to securely outsource sequence comparisions to two remote agent.
W. Du, J. Jia, M. Mangal, M. Murugesan[17]	Commitment Based Sampling Technique, Merkle Tree based commitment technique	2004	A method of cheating detection for general computation outsourcing in grid computing.
M. J. Atallah, K. N. Pantazopoulos, J. R. Rice, E. H. Spafford[9]	Framework for Disguising scientific computation	2001	The general senario for this method is to design local preprocessing the problem or data before sending to agent and local postprocessing to return exact or true answer.
W. Du, M. J. Atallah[14]	Secure Multi-party Computation (SMC)	2001	Customer knows all input information and thus design efficient result verification mechanism
P. Golle I. Mironov [15]	Generic Distributed Computation	2001	This scheme is very efficient in computation as well as in communication overhead for the participants.

Recently, Li and Atallah [16] given a study for secure and collaborative computation of linear programming under the SMC framework. Their solution is based on the additive split of the constraint matrix between two involved parties, followed by a series of interactive (and arguably heavy) cryptographic protocols collaboratively executed in each iteration step of the Simplex Algorithm. This solution has the computation asymmetry issue mentioned previously. Besides, they only consider honest-but-curious model and thus do not guarantee that the final solution is optimal.

Detecting the unfaithful behaviors for computation outsourcing is not an easy task, even without consideration of input/output privacy. Verifiable computation delegation, where a computationally weak customer can verify the correctness of the delegated computation results from a powerful but untrusted server without investing too many resources, has found great interests in theoretical computer science community. Some recent general result can be found in Goldwasser et al. In distributed computing and targeting the specific computation delegation of one-way function inversion, Golle et al. [15] proposed to insert some pre-computed results (images of "ringers") along with the computation workload to defeat untrusted (or lazy) workers. In Du. et al. [14] proposed a method of cheating detection for general computation outsourcing in grid computing. The server is required to provide a commitment via a Merkle tree based on the results it computed. The customer can then use the commitment combined with a sampling approach to carry out the result verification (without re-doing much of the outsourced work.) However, all above schemes allow server actually see the data and result it is computing with, which is strictly prohibited in the cloud computing model for data privacy. Thus, the problem of result verification essentially becomes more difficult, when both input/output privacy is demanded. So the duality theory of LP problem and effectively bundles the result verification within the mechanism design, with little extra overhead on both customer and cloud server.

III. CONCLUSION

Here, we study different method for securing the data outsourcing in cloud computing. In secure outsourcing of scientific computation we studied about secure outsourcing of numerical and scientific computation. In secure outsourcing of sequence comparison we studied that two protocol designs for both secure sequence comparison outsourcing and secure algebraic computation outsourcing. In the securly multi-party computation we study the method of cheating detection for general computation outsourcing in grid computing. So all the method is not fully securing the data outsourcing in cloud computing.

REFERENCES

[1] C. Wang, K. Ren and J. Wang, “Secure and practical outsourcing of linear programming in Cloud computing”, IEEE Transition on cloud computing, pp.820-828, April 2011

[2] C. Wang, Q. Wang, K. Ren and W. Lou, “Ensuring data storage security in Cloud Computing”, in Proc. Of IWQoS’09, July 2009

[3] S. Hohenberger and A. Lysyanskaya, “How to securely outsource cryptographic computations”, in Proc. of TCC, 2005, pp 264-282.

[4] R. Gennaro, C. Gentry and B. Parno, “Non-interactive verifiable computing: Outsourcing computation to entrusted workers”, in Proc. of CRYTO’10, Aug 2010.

[5] M. Atallah and K. Frikken, “Securely outsourcing linear algebra computations”, in Proc of ASIACCS, 2010, pp. 48-59.

[6] N. Gohring, “Amazon’s S3 down for several hours”, Online at http://www.pcworld.com/businesscenter/article/142549/amazons_s3_down_for_several_hours.html, 2008.

[7] Amazon.com, “Amazon Web Services (AWS)”, Online at http://aws.amazon.com, 2008.

[8] Sun Microsystems, Inc., “Building customer trust in cloud computing with transparent security,” 2009, online at https://www.sun.com/offers/details/sun_trans parency.xml.

[9] M. J. Atallah, K. N. Pantazopoulos, J. R. Rice, and E. H. Spafford, “Secure outsourcing of scientific computations,” Advances in Computers, vol. 54, pp. 216–272, 2001

[10] M. J. Atallah and J. Li, “Secure outsourcing of sequence comparisons,” Int. J. Inf. Sec., vol. 4, no. 4, pp. 277–287, 2005.

[11] D. Benjamin and M. J. Atallah, “Private and cheating-free outsourcing of algebraic computations,” in Proc. of 6th Conf. on Privacy, Security, and Trust (PST), 2008, pp. 240–245.

[12] A. C.-C. Yao, “Protocols for secure computations (extended abstract),” in Proc. of FOCS’82, 1982, pp. 160–164..

[13] O. Goldreich, S. Micali, and A. Wigderson, “How to play any mental game or a completeness theorem for protocols with honest majority,” in Proc. of STOC’87, 1987, pp. 218–229.

[14] W. Du and M. J. Atallah, “Secure multi-party computation problems and their applications: a review and open problems,” in Proc. of New Security Paradigms Workshop (NSPW), 2001, pp. 13–22.

[15] P. Golle and I. Mironov, “Uncheatable distibuted computations”, in Proc. of CT-RSA, 2001, pp. 425-440.

[16] J. Li and M. J. Atallah, “Secure and private collaborative linear programming,” in Proc. of CollaborateCom, Nov. 2006.

[17] W. Du, J. Jia, M. Mangal and M. Murugesan, “Uncheatable grid computing”, in Proc. Of ICDCS, 2004, pp. 4-11.

AUTHORS

First Author – Rajendra N. Kankrale, IT Dept, SRES COE, SPPU, Pune., rkankrale@gmail.com
Second Author – Mahendra B. Gawali, IT Dept, SRES COE, SPPU, Pune., gawali_mahen@yahoo.com