ISSN No. 0976-5697

Volume 5, No. 3, March-April 2014

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Reliability, Security and Privacy of Data Storage in Cloud Computing

Anju Mishra
Department of Computer Application, IEC-CET, Greater Noida, India

Dr. Viresh Sharma
Department of Mathematics, N.A.S. (P G) College ,Meerut, India

Dr. Ashish Pandey
Sapient Consulting, Gurgaon, India

Abstract : In this article, we will present new challenges to the cloud computing, namely reliability, security and privacy. Cloud Computing is often marketed as an efficient and cheap solution that will replace the client-server paradigm. The paradigm shift involves/results in the loss of control over data as well as new security and privacy issues. For this reason caution is advised when deploying and using Cloud Computing in enterprises. With Cloud Computing rapidly gaining popularity, it is important to highlight the resulting risks. In this paper new security and privacy issues will be addressed and potential solutions will be discussed

Keywords: Reliability, security, privacy, cloud, computing.

I. INTRODUCTION

"Cloud Computing", a novel computing model for large-scale application, has attracted much attention from both academic and industry. Cloud computing provides scalable deployment and fine-grained management of services through efficient resource sharing mechanism. There seems to be no area of ICT that is not affected by Cloud Computing. As cloud computing continues to roll over the landscape, many enterprise IT organizations still struggle to resolve issues exist with reliability, security and privacy of Cloud Computing.

Cloud computing offers a compelling business model for information services. "Cloud Computing" is not just a computation model allowing us to break the technical restrictions of computation. It also establishes a new kind of business models and merges related technologies together into a single term [1][2][3].

Essential characteristics of cloud computing include on-demand self-service, broad network access, resource pooling, rapid elasticity and measured service. To achieve some of above features, a distributed storage system plays a vital role in cloud computing. Google said "The Data Center Is the Computer [4]." As we know, information must be stored in some sort of recording devices such as memory, hard disk, tape, and cache so that they can be reused or queried in future. Without reliable distributed storage system, cloud computing cannot provide reliable service to users.

Reliability of distributed data storage, including I/O performance [5][6] and security [7][8][9][10][11], had been studied in decades. Although people had developed much more concrete sense in security and hardware had become more powerful, many problems still lie in the reliability of cloud computing. Due to the new computation model, traditional solutions to achieving high reliability may not be appropriate for modern applications in cloud computing. In this paper, we focus on three important features of distributed storage: 1) capability/reliability of distributed storage, 2) information security in cloud computing and 3) privacy in cloud computing.

II. CLOUD SERVICE RELIABILITY

Cloud Service Reliability is defined as the probability that a cloud service under consideration can be successfully completed for a user in a specified period of time. Two type of failure will more or less affect this probability to provide a successful service.

(a). Request Stage Failures: Overflow and Timeout.

(b). Execution Stage Failures: Data resource missing, Computing resource missing, Software failure, Database failure, Hardware failure, and Network failure.

The reliability of individual element can be obtained from

$$R(element_n) = \exp\{- \lambda(element_n) . T_w(element_n)\}$$

Which is more realistic and practical than other conventional methods [12] assuming the reliabilities of elements (nodes and links) are constant, (e.g. a node is always 90% reliable, regardless of how long it works). In fact, the reliability of individual element is affected by various conditions such as failure rate, amount of data, bandwidth, operation time, etc.

a. New Evaluation Algorithm:

The conventional algorithms have one or some of the following assumptions that are not applicable to evaluate the reliability given the above new model: 1) the network topology is made up of physical nodes/links without considering the virtual nodes/links; 2) the operational probabilities (reliabilities) of nodes or links are constant; 3) only hardware failures of links and processors are considered without taking into account the software, data and resource failures. Therefore, we further present a new algorithm for evaluating the overall cloud service reliability considering all different factors during the execution stage given the new graph model. The new evaluation algorithm

based on Graph theory and Bayesian theorem is presented to derive the reliability, as follows.

A. Minimal Subtask Spanning Tree (*MSST*):

The set of all nodes and links involved in completing a specific subtask form a *Subtask Spanning Tree* (*SST*). This SST can be considered to be a combination of several minimal subtask spanning trees (*MSST*s), where each *MSST* represents a minimal possible combination of available elements (nodes and links) that guarantees the success to execute this specific subtask (i.e., failure of any element in *MSST* leads to the subtask failure). By this definition of *MSST*, we can see that each *MSST* contains exactly one set of data resources without any duplication, because any duplication could be reduced to another smaller SST. Therefore, for any *MSST*, the data resources and precedent subtasks that provide certain input for the subtask are also determined.

Some elements inside one *MSST* can still belong to several paths if they are involved in different communications tasks, such as data transmission or data resource access. Note that all elements in the execution stage are hot-standby although some elements/subtasks may be waiting for the output of some other subtasks. So during the waiting period, those elements are also possible to fail. Thus, we suppose that an MSST completes the entire service if all of its elements do not fail during the maximal time allowed to complete all subtasks in executing which they are involved. Therefore, when calculating the element reliability in a given MSST, one has to use the corresponding record with maximal time.

Assume there are a total of *K* elements in an *MSST*, and $element_i$ (i=1,2,…,*K*) denotes the *i*:th element in the *MSST*. Accordingly, the communication time of the *i*:th element is denoted by $T_w(element_i)$ and $\lambda(element_i)$ l element represents its failure rate. The reliability of this single *MSST* can be simply expressed as

$$R_{MSST} = \Pi_{i=1}^{k} \exp\{-\lambda(element_i) \cdot T_w(element_i)\}, \quad (1)$$

With this equation, the reliability of an *MSST* can be computed if the working times of all the elements are obtained. Hence, finding all the *MSST*s and determining the working time of their elements are the first step in deriving the execution reliability of a cloud service. To solve the graph traversal problem, several classical algorithms have been suggested, such as depth-first search, breadth-first search, etc. These algorithms can find all *MSST*s in an arbitrary graph. Here, we propose a depth-first search algorithm here, which is briefly described as follows:

Step 1. Given a program/subtask, say S_m, start from a node that contains this program, to search the required data resources and precedent subtasks/programs along the possible links, and record elements that compose the searching route and their communication times.

Step 2. Until all the required data resources and precedent subtasks/programs are reached, an *MSST* is found, and record this *MSST*.

Step 3. Then other routes are tried to search other *MSST*s until all the *MSST*s are searched.

Step 4. Change to another node that also contains the program S_m. Repeat the above three steps until all the nodes that have S_m are evaluated. Save all the *MSST*s found associated with S_m into the vector MSST (S_m) .

Step 5. Change to another program and repeat the above four steps until all the programs are explored. Then all the vectors of MSST (S_m) (m=1,2,…*M*) are generated.

B. Minimal Execution Spanning Tree (*MEST*):

Similar to the *MSST*, a Minimal Execution Spanning Tree (*MEST*) represents a minimal possible combination of available elements (nodes and links) that guarantees the success to execute the entire service. Thus, at least one *MSST* of each MEST (S_m) (m=1,2,…*M*) must be reliable, and then the subtask S_m (m=1,2,…*M*) can be connected to those remote resources and exchange data with them successfully through the network. If any set of the *M* subtasks are successful, then the execution is reliable for the cloud service to execute the required set of subtasks, so the *MEST* could be derived as the intersection of the above sets of *MSST*s as

$$MEST = \cap_{m=1}^{M} MSST(S_m) \quad (2)$$

In practice, all *MEST*s could be generated in the following steps:

Step 1: Select an *MSST* from each set of MSST (S_m) where (m=1,2,…*M*).

Step 2: *M MSST*s are obtained and put them together to generate the *MEST*. For each common element when intersecting trees together, record the greater working time as the final working time of this element in the *MEST*.

***Step 3*:** Repeat Step 1-2 until all combinations are tried to generate all *N MSST*s.

Similar to (1), the reliability of a single *MEST* can be calculated by

$$R_{MSST} = \Pi_{i\epsilon MEST} \exp\{-\lambda(element_i) \cdot T_w(element_i)\}, \quad (3)$$

C. Execution Reliability:

Having the list of *N MEST*s and the corresponding task completion time, one can determine the reliability of cloud service at the execution stage, as follows.

$$R_{execute} = Pr(\cup_{i=1}^{N} MSST_i) \quad (4)$$

Which means any one *MEST* out of the total *N MEST*s being succeeded will make the cloud service successfully executed in the execution stage. Denote event E_j the successful operation of the $MEST_j$ while $\bar{E}_j$ the failure of the $MEST_j$. Using the Bayesian theorem on conditional probability, we can derive (4) to a summation of conditional probabilities

$$\bar{R}_{execute} = Pr(\cup_{i=1}^{N} MSST_i) = \Sigma_{j=1}^{Ni} Pr(E_j).Pr(E_1, E_2, \ldots, E_{j-1} | \bar{E}_j) \quad (5)$$

The probability $Pr(E_j)$ can be directly obtained from (3) as R_{MESTj} and the probability, $Pr(E_1, E_2, \ldots, E_{j-1} | \bar{E}_j)$ can be computed by the following two-step algorithm.

Step 1 identifies the failures of all of the critical elements in a period of time during which they lead to the failures of any one MEST from previous j-1 MESTs, but do not affect $MEST_j$.

Step 2 generates all the possible combinations of the identified critical elements that lead to the event $E_1, E_2, \ldots, E_{j-1} | \bar{E}_j$ by a binary search, and computes the probabilities of those combinations. Their summation is $Pr\{E_1, E_2, \ldots, E_{j-1} | \bar{E}_j\}$

When calculating the failure probabilities of MESTs' elements the maximal time from the corresponding records in a list for the given MEST should be used.

Finally, if a cloud service needs to be successfully completed, both request stage and execution stage should be reliable. After we derive the reliability for both stages, we can hereby get the cloud service reliability $R_{Service}$ as

$$R_{Service} = R_{request}\, R_{execute} \quad (6)$$

Where $R_{request}$ can be derived from the reliability of request stage and $R_{execute}$ can be derived from the reliability of execute stage by (5) [13].

III. SECURITY AND PRIVACY OF CLOUD COMPUTING SERVICES

As Cloud Computing rapidly gaining popularity, it is important to highlight the resulting risks. As security and privacy issues are most important, they should be addressed before Cloud Computing establishes. Many IT and important research agencies are aware of these risks and have produced reports and analyses to document them [14], [15], [16] ,[17]. Two main issues exist with security and privacy aspects of Cloud Computing:

(a). loss of control over data and

(b). dependence on the Cloud Computing provider.

These two issues can lead to a number of legal and security concerns related to infrastructure, identity management, access control, risk management, regulatory and legislative compliance, auditing and logging, integrity control as well as Cloud Computing provider dependent risks.

A. Typical issues due to the loss of control over data are:

a. Most customers are aware of the danger of letting data control out of their hands and storing data with an outside Cloud Computing provider. Data could be compromised by the Cloud Computing provider itself or other competitive enterprises which are customers with the same Cloud Computing provider. There is a lack of transparency for customers on how, when, why and where their data is processed. This is in opposition to the data protection requirement that customers know what happens with their data.

b. Many Cloud Computing providers are technically able to perform data mining techniques to analyse user data. This is a very sensitive function and even more so, as users are often storing and processing sensitive data when using Cloud Computing services.

c. Mobile devices, in particular with their limited storage and computing capabilities are drivers for having services provided by Cloud Computing instead of using software on individual computers. Even data that are only to be transferred from one mobile device to another (local) device, are often transferred via the cloud, when cloud oriented applications on the mobile devices are involved. Therefore users often put themselves at risk without noticing this, as they assume that the data is transferred locally.

d. Since Cloud Computing is a service, it has to be accessed remotely. The connection between the Cloud Computing provider and customer is not always adequately protected. Security risks that threaten the transfer line include eavesdropping, DNS spoofing, and Denial-of-Service attacks.

e. The paradigm shift in Cloud computing makes the use of traditional risk management approaches hard or even impossible. Irrespective of the fact that control over data is transferred to the Cloud Computing provider, risk management and compliance issues are split between the Cloud Computing provider, Internet provider and customer. However, compliance can be seen as one of the important trust factors between the Cloud Computing provider and customer. Regulatory and legislative compliance is also problematic. Cloud data centers can be geographically dispersed. Therefore legislative compliance is not currently adequately defined.

f. As all technical control is given to the Cloud Computing provider, customers often want to have an external audit of this provider. Therefore logging and auditing information has to be stored and protected in order to enable verification. Appropriate logging could provide the possibility for forensic investigation in cases of incident.

g. Concerns also exist with regard to deletion of data: It is difficult to delete all copies of electronic material because it is difficult to find all copies. It is impossible to guarantee complete deletion of all copies of data. Therefore it is difficult to enforce mandatory deletion of data. However, mandatory deletion of data should be included into any forthcoming regulation of Cloud Computing services, but still it should not be relied on too much: the age of a "Guaranteed complete deletion of data", if it ever existed has passed. This needs to be considered, when data are gathered and stored.

h. Data Protection and Privacy legislation is not even similar in many countries around the globe yet Cloud Computing is a global service of the future. Consequently the problems and risks that affect data protection rules in Europe must be considered properly when Cloud Computing platforms are located on servers in non-European countries.

i. Cloud computing depends on a reliable and secure telecommunications network that assures and guarantees the operations of the terminal users of the services provided in the cloud by the cloud computing provider. Telecommunications networks are often provided separately from the Cloud computing services.

B. Typical issues with regard to the dependence on the Cloud Computing provider are:

a. A major concern regarding dependence on a specific Cloud Computing provider is availability. If the Cloud Computing provider were to go bankrupt and stopped providing services, the customer could experience problems in accessing data and therefore potentially in business continuity.

b. Some widely used Cloud Computing services (e.g. GoogleDocs) do not include any contract between the customer and Cloud Computing provider. Therefore a customer does not have anything to refer to if incidents occur or any problems arise.

c. Cloud Computing is a service similar to other more "traditional" services and utilities (e.g. telecommunication, transaction banking, electricity,

gas, water, etc.) Both Cloud Computing services and traditional services and utilities tend to be offered by large providers dealing with smaller customers. Therefore the customers usually depend on the providers because it is difficult to change providers if it is possible at all. Consequently traditional services (e.g. telecommunication, transaction banking, electricity, gas, water, etc.) are usually regulated with regard to the functionality range (e.g. mandatory functions, coverage), pricing, liability of the provider, and reliability.

Cloud Computing corroborates a trend that ICT security is no longer a purely technical issue but an issue between individuals and organizations and thus includes both human and organizational aspects such as management, contracting, and legal enforcement.

IV. CONCLUSION

In this article, we addressed the reliability, security and privacy issues of cloud computing. Cloud computing brings productive development and elastic resource management. In this we can well define the Risk management and (legal) compliance issues in the contract between Cloud Computing provider and customer. Therefore, information security plays a vital role in cloud services. Due to the new computing model, security issues had been studied by academia and hacker communities, such as Black Hat, in recent years.

Research on the basic concepts and issues in informatics, security, and privacy and their consequences and trade-off's with regard to Cloud Computing should be encouraged. Also issues concerning the possible impact of Cloud Computing platforms on the validity of certification of applications that are certified according to criteria (e.g. Common Criteria, European Privacy Seal, etc.) may need to be investigated.

V. REFERENCES

[1]. M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and others, "A view of cloud computing," Communications of the ACM, vol. 53, 2010, pp. 50–58.

[2]. I. Foster, Y. Zhao, I. Raicu, and S. Lu, "Cloud computing and grid computing 360-degree compared," Grid Computing Environments Workshop, 2008. GCE'08, 2009, pp. 1–10.

[3]. L.M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, A break in the clouds: towards a cloud definition," ACM SIGCOMM Computer Communication Review, vol. 39, 2008, pp. 50–55.

[4]. Urs Hoelzle , Luiz Andre Barroso, "The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines," Morgan and Claypool Publishers, 2009.

[5]. Y. Gu and R.L. Grossman, "Sector and Sphere: the design and implementation of a high-performance data cloud," Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 367, 2009, pp. 2429 -2445.

[6]. A.W. Leung, M. Shao, T. Bisson, S. Pasupathy, and E.L. Miller, "Spyglass: Fast, scalable metadata search for large-scale storage systems," Proccedings of the 7th conference on File and storage technologies, 2009, pp. 153–166..

[7]. Y. Chen, V. Paxson, and R. Katz, "What's New About Cloud Computing Security," University of California, Berkeley Report No. UCB/EECS-2010-5 January, vol. 20, 2010, pp. 2010–5.

[8]. L.M. Kaufman, "Data security in the world of cloud computing," Security & Privacy, IEEE, vol. 7, 2009, pp. 61–64.

[9]. Y.Y. Yumin, "Application of Cloud Computing on Network Security," Science, 2009, p 07.

[10]. M. Jensen, J. Schwenk, N. Gruschka, and L.L. Iacono, "On technical security issues in cloud computing," Cloud Computing, 2009. CLOUD'09. IEEE International Conference on, 2009, pp. 109–116.

[11]. L.M. Kaufman, "Data security in the world of cloud computing," Security & Privacy, IEEE, vol. 7, 2009, pp. 61–64.

[12]. M. Xie, Y.S. Dai, K.L. Poh, Computing Systems Reliability: Models and Analysis, (330 pages), Springer: New York, U.S.A., 2004. ISBN: 0-306-48496-X.

[13]. Yuan-Shun Dai, Bo Yang, Jack Dongarra, Gewei Zhang," Cloud Service Reliability: Modeling and Analysis

[14]. J. Brodkin, Gartner: Seven cloud-computing security risks,

[15]. Cloud Computing Security Considerations, A Microsoft Perspective, Microsoft Whitepaper, 2010,

[16]. Cloud Computing: Benefits, Risks and Recommendations for Information Security, ENISA Report, 2009,.

[17]. Security Guidance for Critical Areas of Focus in Cloud Computing V2.1, Cloud Security Alliance (CSA) Report, 2009,