Volume 5, Issue 6, June 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Analysis of Service Level Agreements & Performance Dimension in Cloud Computing Environments

[1]Mamta Dadhich, [2]Pramod Kumar Pandey, [3]Prof.(Dr.) A K Tiwari
[1]Research Scholar, Department of Computer Science, IIS University, Jaipur, Rajasthan, India
[2]Research Scholar, Department of Computer Science, Jagarnath University, Jaipur, Rajasthan, India
[3](Director) Wilfred's Institute of Information Technology, Jaipur, Rajasthan, India

Abstract: The Cloud computing has changed the approach used for provided that circulated services to various business, government agents. The Cloud computing delivers scalable and on-demand services to a large amount users in different domains. The Cloud computing has been the focus of active and widespread research since late 2007. Before the term 'cloud' was coined, there was grid technology. at this moment, the hot topic of research is cloud and more proposed frameworks and models of various solutions for the new technology have started to be applied to the cloud architecture on the other hand, this new technology has also created many challenges for service providers and customers, especially for those users who already own complicated legacy systems. This paper reviews the challenges related to the concepts of trust, SLA management, and cloud computing. We begin with a study of cloud computing architecture. Then, we converse regarding existing frameworks of cloud SLA (service level agreements) in different domains such as cloud web services and grid computing. In the last section of the paper, we have discussed the some advantages, limitations of present performance measurement models for SOA and distributed systems, grid computing, and cloud services. Finally, we go over the main points and conclude our works.

Keywords:

I. INTRODUCTION CLOUD COMPUTING

Cloud computing has been the focus of active and extensive research since late 2007. Before the term 'cloud' was coined, there was grid technology. Now, the hot topic of research is cloud and more proposed frameworks and models of various solutions for the new technology have ongoing to be applied to the cloud architecture. In this section, we analysis the literature in order to find out the most appropriate definition of "cloud computing". Also, we review the different architectural frameworks and the frequent challenges that may present major problems for providers and customers who are interested in understanding this type of distributed computing.

The Google trends report shows that cloud computing had surpassed grid computing

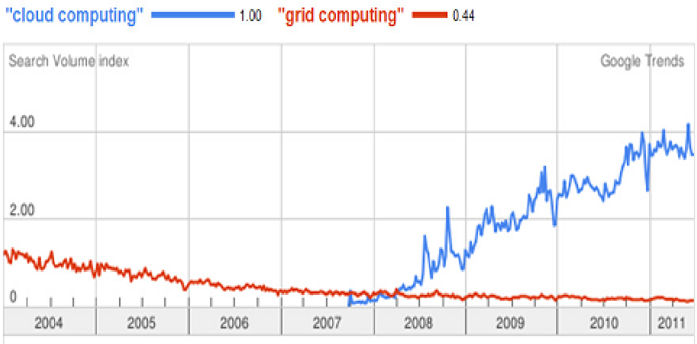

Figure 1 Showing Cloud computing trend, source and Google search engine

The term "cloud" is a metaphor for the common internet concept. The cloud symbol is also used to present the meaning of network connection and the way that the cloud technology is provided by internet infrastructure. "Computing" in the context of the cloud domain refers to the technology and applications that are implemented in the cloud data centers.

In common definition of cloud computing. They state that developers and business decision makers puzzle the understanding of the technology with the features of cloud data centers. So, large budgets may be allocated to implement private or even public cloud data centers.

However, these data centers face several problems when users or public customers want to connect the interfaces of their legacy systems with the new technology of cloud architecture. the challenge of maximizing the revenue of building cloud technology to professionals who are involved in distributed services. Because they come from a traditional computing domain, they are confused about the other concepts of distributed services such as grid and web services.

The definition used in is as follows: "Clouds are a large pool of easily usable and accessible virtualized resources (such the same as hardware, development platforms or services). These resources can be dynamically reconfigured to adjust to a variable load scale, allowing also for an optimum resource utilization. This pool of resources is typically exploited by a pay per use model in which guarantees are offered by the Infrastructure Provider by means of customized SLAs". Although, this definition presents the main features of cloud computing, it does not encompass other important components of cloud architecture which include the method of establishing and managing network, applications, and supporting services.
The cloud computing architecture takes advantage of the way that different distributed services (mainly web services and SOA) are implemented. Two types of cloud services are presented along with this definition they define SaaS and PaaS. Despite the importance of IaaS as a main component of cloud architecture, they do not adequately discuss this type of cloud delivery model.
In this paper, we adopted and considered the definition provided by U.S. NIST (National Institute of Standards and Technology) , according to which "Cloud computing is a model for enabling convenient, on demand network access to a share pool of configurable computing resources (Example- networks, servers, storage, applications, and services) to facilitate can be rapidly provisioned and released with minimal management afford or service provider interaction" .

II. TAXONOMY OF CLOUD COMPUTING

According to Buyya presents more than fifteen characteristics to differentiate cloud computing from other distributed systems. Buyya uses scalability, automatic adaption, virtualization, and dynamic model of billing as major concepts to construct the architecture of cloud computing. In addition, he explains how cloud services can be delivered to different types of users. For example, users who want to develop small size applications may connect to one of the PaaS such as Microsoft without the need to install any of the development tools. identifies a clear taxonomy framework for presented categories of cloud services. The class of cloud services is described in a tree-structured taxonomy, and the unique characteristics of each model of service are used to identify each node of the proposed tree structure.
According to Hoefer's classification provides a understandable comparison of cloud services at high level on the tree structure. on the other hand, at the base of the structure, the taxonomy of cloud services is not enough to distinguish services in more detail. The taxonomy presented by Laird in defines cloud technology from the perspective of service providers.

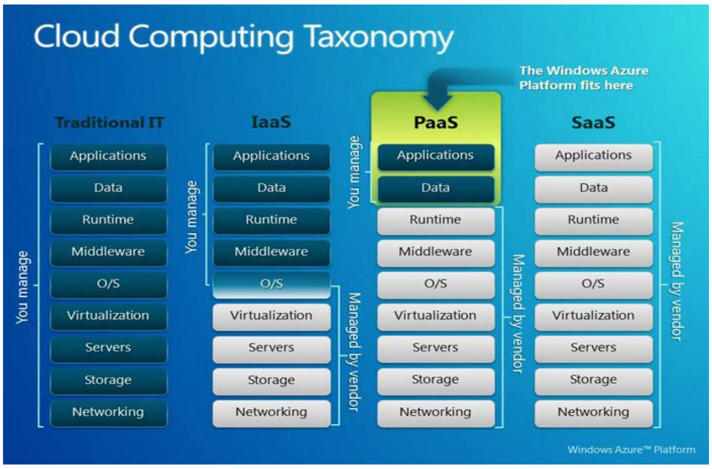

Figure 2 Showing Taxonomy Cloud Computing Services

The proposed taxonomy presents the ordinary vendors of cloud services. According to Laird presents two classifications of services. The first class defines the infrastructure of cloud services, and the second one defines the services based on cloud features such as security, billing, and applications which are built into the system. According to Rimal presents a comprehensive framework for the architecture of cloud computing. He describes the taxonomy of cloud services with more focus on the management domain of cloud contents. The concepts of management, business, billing, and support of cloud services are defined in great depth in order to present the cloud architecture as a new business model. The main advantage of the proposed work by Rimal is that relationships between security features and cloud components are provided as a part of the comparison of service models in cloud computing. The taxonomy proposed by Oliveira classifies the concepts of cloud computing according to dimensions of cloud architecture, business model, technology infrastructure, pricing, privacy, and standards.
The proposed taxonomy is provided in a hierarchical tree with parent and child relationships. Oliveira uses SaaS, PaaS, IaaS, and DaaS as sub-taxonomy for the business model. This classification is used in the literature of cloud computing to distinguish the service delivery for end users of cloud services. These sub-taxonomy terms may cause confusion in understanding the way that various business models are constructed for cloud services. The taxonomy proposed by Oliveira describes the concepts of cloud architecture from the point of view of e-science. Therefore, many of the technical aspects of cloud computing are missing from the proposed taxonomy.

III. SERVICE LEVEL AGREEMENTS

A service level agreement is a document that includes a explanation of the agreed service, service level parameters, guarantees, and actions for all cases of violation. The SLA is very important as a contract between consumer and provider. The main idea of SLAs is to give a clear definition of the formal agreements about service terms like performance, accessibility and billing. It is important that the SLA include the obligations and the actions that will be taken in the event of any violation, with clearly expressed and shared semantics between each party involved in the online contract.

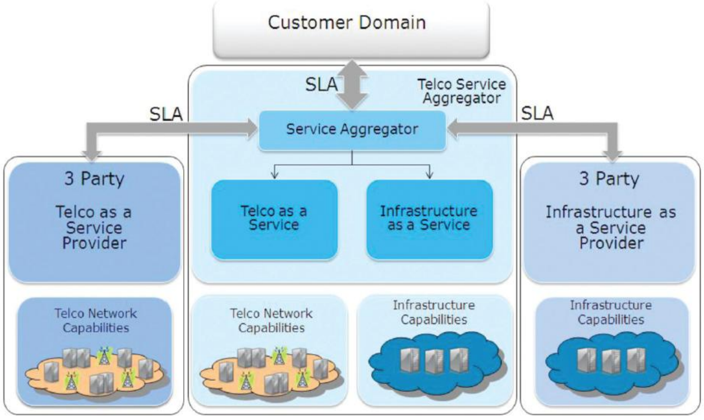

Figure 3 Showing Taxonomy Cloud Computing Services

In This section discusses works associated to SLAs in three domains of distributed services. Firstly, we discuss the proposed SLAs structure for web services. Secondly, the frameworks of SLAs designed to grid computing are reviewed; thirdly, we discuss the main works that specifically focus on cloud computing. Finally, we include in this section the main shortcomings of these SLA frameworks.

3.1 SLAs for Web Services

Several specifications for defining SLAs have been proposed for web services. WSLA language introduces a mechanism to help users of web services to configure and control their resources in order to meet the service level. Also, the service users can monitor SLA parameters at run time and report any violation of the service.

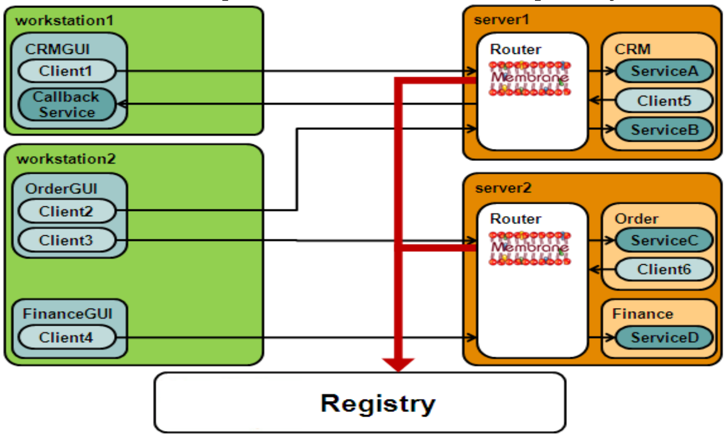

Figure 4 WSLA Web Services

WSLA was developed to describe services under three categories:

- Parties: in this section, information about service consumers, service providers, and agents are described.
- SLA parameters: in this section the main parameters which are measurable parameters are presented in two types of metrics. The first is resource metrics, a type of metrics used to describe a service provider's resources as row information. The second one is composite metrics. This metrics is used to calculate the combination of information about a service provider's resources. The final section of the WSAL specification is Service Level Objective (SLO). This section is used to specify the obligations and all actions when service consumers or service providers do not comply with the guarantees of services.
- The WSLA provides an adequate level of online monitoring and contracting, but does not clearly specify when and how a level of service can be considered a violation. WSOL is a service level specification designed mainly to specify different objectives of web services.

The information about service consumer and service provider, domain of service, and other specifications of service are presented in the context section. Terms of services and guarantees are described in greater detail in the terms section. These types of online agreements were developed for use with general services. For cloud computing,

service consumers need more specific solutions for SLAs in order to reflect the main parameters of the visualization environment; at the same time, these SLA solutions should be dynamically integrated with the business rules of cloud consumers.
The primary shortcomings of these approaches is that they do not provide for dynamic negotiation, and various types of cloud consumers need a different structure for the implementation of SLAs to integrate their own business rules with the guarantees that are presented in the targeted SLA.

3.2 SLAs for Grid Computing
In the context of grid computing there are a number of proposed specifications which have been developed especially to improve security and trust for grid services. In an SLA based knowledge domain has been proposed by Sahai to represent the measurable metrics for business relationships between all parties involved in the transaction of grid services. Also the author proposed a framework to evaluate the management proprieties of grid services in the lifecycle.

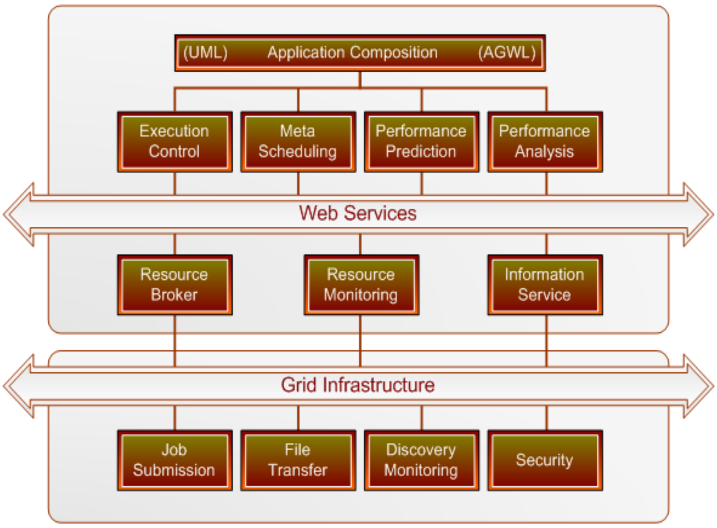

Figure 5 Showing SLA for Grid Computing Services

In this work, business metrics and a management evaluation framework are combined to Produce an estimated cost model for grid services. In our research, we extend this approach in order to build a general costing model based on the technical and business metrics of the cloud domain. The framework proposed in this work lacks a dynamic monitoring technique to help service customers know who takes responsibility when a service level is not provided as specified in SLA documents. Leff conducted a study of the main requirements to define and implement SLAs for the grid community.
The source code of information and hardware modeling are used to predict the value of performance metrics for grid services. The model proposed by Keung can be used in other types of distributed computing. But in the cloud environment, this model cannot be integrated with a dynamic price model of cloud services. It needs to be improved by using different metrics for cost parameters to reflect the actual price of cloud services.
The system proposed by Padget in considers the response time of applications in the grid systems. The main advantage of the proposed system is that it can predict the CPU time for any node in the grid network before conducting the execution. When Padget tested the adaptation SLA model using a real experiment on the grid, the prediction system produced values for response time close to the values obtained when users executed the same application on the grid. Noticing the delay recorded for the large size of executed files, the author claims that the reason for this delay is the external infrastructure such as internet connections. The author also discusses the impact of the time delay caused by external parties to the reputation of service providers when using SLA management systems. Although the author provides a good method for calculating the response time for grid resources, other
Metrics such as security and management metrics, are absent in this work.

3.3 SLAs for Cloud Computing
The context of this research is the management of service level agreements in cloud communities. In the sections above, we presented the frameworks and models in the current literature that are designed mainly for managing SLAs in traditional distributed systems. In this section, SLAs and approaches to agreement negotiations in the cloud community are presented.
According to Valdimir describes the quality of services related to cloud services and different approaches applied to map SLA to the QoS. Services ontology for cloud computing is presented in order to define service capabilities and the cost of service for building a general SLAs framework. The proposed framework does not consider all types of cloud services; it is general and was tested on the Amazone EC2 only. It also needs to consider other types of cloud providers such as PaaS, DaaS, and SaaS. Our framework in this research considers this issue in the validation phase of the research. The framework developed by Hsien focuses on software as a service model of delivery in cloud computing.

More details are provided on how the services can be integrated to support the concept of stability of cloud community especially for SaaS.

The Shortcomings of the projected Works for SLAs in the framework of Distributed Services The frameworks and structures that were discussed in previous sections have the following problems:

1. The accessible frameworks focus more on the technical attributes than on the security and management aspects of services.
2. The proposed structures of SLAs in the above domains do not consist of a clear definition of the relationship between levels of violation and the cost of services.
3. Most of the above studies do not assimilate a framework of trust management of the service provider with the collected data from monitoring systems of SLAs.
4. The concepts, definitions of service objectives and service descriptions included in SLAs are not easy to understand, especially for business decision makers.
5. The proposed works for cloud environments focus more on the evaluation of virtualization machines on local servers than on existing cloud service providers.
6. Most of the proposed structures of SLAs are defined by technical experts.

IV. PERFORMANCE MEASUREMENTS MODELS

Cloud providers have been increased to deliver different models of services. These services are provided at different levels of quality of services. Cloud customers need to have a reliable mechanism to measure the trust level of a given service provider. Trust models can be implemented with various measurement models of services. As a part of this research, we investigate the use of a measurement approach in order to develop a general trust model for cloud community. In this section, the measurement model of SOA, distributed, and grid services will be reviewed.

4.1 SOA Performance Models

According to Kounev propose an analytical approach to modelling performance problems in SOA-based applications. The authors discuss the different realistic J2EE applications for large systems of SOA architecture. A validated approach has been weathered for capacity planning of the organizations that use distributed services as an outsourcing infrastructure. The advantage of the proposed method is its ability to predict the number of application servers based on the collected information of SLA metrics.

The main focus of Rud’s method is on the management aspects of services. However, this approach does not consider performance issues of response time, data storage, and other metrics of technical infrastructure. For the optimization of total execution time and minimization of business processes cost, Menasce in provides an optimized methodology based on the comparison of performance metrics of SOA-based services. In this study, Menasce developed the proposed method to estimate the cost level of all services which are registered in the SOA directory under medium sized organizations.

Then, the cost metric is compared to the real performance of services. The parameters of the performance metrics can be selected by service customers. So, the proposed model can be used for different types of services. Although, the proposed method produces a high level of reliability and usability, issues such as risk management, and trust mechanisms of the relationship between service providers and service customers are not discussed in more details.

4.2 Distributed Systems Performance Models

According to Kalepu propose a QoS-based attribute model to define the non-functional metrics of distributed services. Availability, reliability, throughput, and cost attributes are used in their work to define performance of resources of a given service provider.

Two approaches of resources are used to calculate the final value of reputation. The first resource is the local rating record. Ratings of services which are invoked by local customers are stored in this record. In the second resource, global ratings of all services that are executed on resources of a given service provider are stored.

According to Kalepu discuss the need to use SLA parameters to calculate the value of performance metrics, they do not explain how these parameters can be linked to the local global resources of a rating system. According to Yeom provide a monitoring methodology of the performance parameters of service. The proposed methodology uses the broker monitoring systems to evaluate the performance of resources of a service provider. Collected data of performance metrics are not maintained on the service consumer database. This method incurs low cost in terms of implementing measurement architecture but more risk in terms of privacy, availability of data, and security. Such risks are not easy to control, especially in the case of multi tenant distributed systems.

According to Guster provides an evaluation methodology for distributed parallel processing. In the proposed method, authors use a parallel virtual machine (PVM) and real hosting servers to compare the results of their experiments. The efficiency of the evaluation method performed better in PVM for the processing time. In the real server environment, the experiments presented better performance in terms of communication time. The evaluation of this work does not include the implementation processes and the experiment resultsare not clearly explained.

4.3 Cloud Computing Performance Models

Several studies already exist on the scalability of virtual machines. Most of these studies considered the measurement of performance metrics on the local machines. The background loads of tested machines are

controlled to compare the results of performance with a different scale of loads. Evangelinos and Hill evaluated the performance of Amazon EC2 to host High Performance Computing (HPC). They use a 32-bit architecture for only two types of Amazon instances. In our study, we run various experiments on most types of Amazon EC2 instances. These instances are: small, large, extra large, high CPU, medium, and high CPU extra large instance.
According to Jureta, and Herssens propose a model called QVDP which has three functions: specifying the quality level, determining the dependency value, and ranking the quality priority. These functions consider the quality of services from the customers' perspective. However, the performance issues related to cloud resources are not discussed and details are missing regarding the correlation of the quality model with the costing model of services.
However, our proposed work profiles the performance of virtualization resources that are already running on the infrastructure of existing cloud providers.
The Shortcomings of the Proposed Works for Above Performance models

1. The above proposed models for evaluating the virtualization services focus on how to measure the performance of virtual machines using local experiments. However, the techniques used for measuring the actual resources of cloud providers need further refinement in order to ensure some level of trust between service providers and the customers.
2. Most of the proposed works on performance evaluation do not allow service customers to specify the parameters of performance metrics. In cloud computing, service customers need a more flexible and dynamic approach to modify the parameters of performance metrics in order to solve the problem of dynamic changes of service requirements and business models of customers.
3. The experiments using the above proposed models do not specify the benchmarks for the performance evaluation.
4. In cloud computing architecture, the relationship between performance monitoring and costing metric is very important. The proposed models do not link the results of performance monitoring with the actual cost metric of services.

V. CONCLUSIONS

A service level agreement is a document that's includes a explanation of the agreed service & service level parameters, guarantees, actions for all cases of violation.

The SLA is most important as a contract among consumer and Services provider. The most important idea of SLAs is to give a clear definition of the formal agreements about the Quality of service and terms like performance, accessibility and billing. It is most important that the SLA include the obligations and the actions that Effective SLAs are extremely important to ensure effective outsourcing engagements. The SLA commitments are the heart of a successful agreement and are a critical factor in long-term success. Lack of experience in the use and implementation of performance metrics can cause problems for many organizations as they attempt to formulate their SLA strategies and select and set the metrics needed to support those strategies.

The **cloud provider** uses service level management to make decisions about its infrastructure; for example, if throughput isn't always meeting a customer's requirements, the provider can reallocate bandwidth or add more hardware. SLAs make use of the knowledge of enterprise capacity demands, peak periods, and standard usage baselines to compose the enforceable and measurable outsourcing agreement between vendor and clients.

REFERENCES

[1] D. C. Wyld, Moving to the cloud: An introduction to cloud computing in government: IBM Center for the Business of Government, 2009.

[2] L. Wang, "Cloud computing: a perspective study," New Generation Computing, vol. 28, pp. 137-146, 2010.

[3] P. Mell and T. Grance, "Draft nist working definition of cloud computing," Referenced on June. 3rd, 2009.

[4] R. Buyya, "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility," Future Generation Computer Systems, vol. 25, pp. 599-616, 2009.

[5] R. Jennings, Cloud Computing with the Windows Azure Platform: Wrox, 2010.

[6] C. Hoefer and G. Karagiannis, "Taxonomy of cloud computing services," 2010.

[7] P. Laird, "Different strokes for different folks: a taxonomy of cloud offerings," Enterprise cloud submit, INTEROP, 2009.

[8] B. P. Rimal, "A Taxonomy, Survey, and Issues of Cloud Computing Ecosystems," Cloud Computing, pp. 21-46, 2010.

[9] D. Oliveira, "Towards Taxonomy for Cloud Computing from an e-Science Perspective," Cloud Computing, pp. 47-62, 2010.

[10] H. Ludwig, "Web service level agreement (WSLA) language specification," IBM Corporation, 2013.

[11] A. Andrieux, "Web services agreement specification (WS-Agreement)," 2004.

[12] A. Sahai, "Specifying and monitoring guarantees in commercial grids through SLA," 2003.

[13] A. Leff, et al., "Service-level agreements and commercial grids," IEEE Internet Computing, vol. 7, pp. 44-50, 2013.

[14] H. N. L. C. Keung, et al., "Self-adaptive and self-optimising resource monitoring for dynamic grid environments," 2004, pp. 689-693.

[15] J. Padgett, "Predictive adaptation for service level agreements on the grid," International Journal of Simulation: Systems, Science and Technology, vol. 7, pp. 29–42, 2006.

[16] V. Stantchev and C. Schröpfer, "Negotiating and enforcing qos and slas in grid and cloud computing," Advances in Grid and Pervasive Computing, pp. 25-35, 2009.

[17] C. H. Wen, "A SLA-based dynamically integrating services Saas framework," pp. 306-311.

[18] S. Kounev and A. Buchmann, "Performance modeling and evaluation of large-scale J2EE applications," 2003, pp. 273-284.

[19] A. Walter and D. Potter, "COMPOSITION, PERFORMANCE ANALYSIS AND SIMULATION OF WEB SERVICES," 2007.

[20] C. Evangelinos and C. Hill, "Cloud Computing for parallel Scientific HPC Applications: Feasibility of running Coupled Atmosphere-Ocean Climate Models on Amazon's EC2," ratio, vol. 2, p. 2.34, 2008.

