Service Level Agreement In Cloud Computing : *An Overview*

G. Justy Mirobi
Research Scholar,
Department of Computer Science
Bharathiar University
Coimbatore - 641 046, Tamil Nadu, India.
E-mail: georgejusjer@gmail.com
georgejustymirobi1@gmail.com

Dr. L. Arockiam
Associate Professor in Computer Science
Department of Computer Science,
St. Josephs College (Autonomous)
Tiruchirappalli, Tamil Nadu, India.
E-mail: larockiam@yahoo.co.in

Abstract **- Cloud computing is an internet based computing, providing the on demand services through the internet such as servers, storage disk, different platforms and applications to any business level company or organization. The cloud computing services are 'Pay based on usage' based on the agreement between Service Provider of the Cloud and customer. Service Level Agreement(SLA) is a contract, contracted between provider of the service and the third party such as purchaser of service, dealer(agent) or monitor(agent), where service is formally defined. Practically the term Service Level Agreement(SLA) is used to mention the delivery period of the contracted service, and to evaluate the performance of service. The cloud computing is a recent technology providing numerous services for critical commercial applications, trustworthy and flexible mechanisms for managing the contracts. Therefore SLA is very important for conducting the cloud business in a smooth way. This paper presents an overview of SLA, benefits of cloud computing, the necessity of cloud computing, the importance of SLA, classification of Service Level Agreement, SLA-based cloud framework of cloud computing.**

Keywords - Cloud computing, Remote Server, Service Level Agreement, Monitoring.

I. INTRODUCTION

The cloud computing is run-through of using remote servers that are hosted on the internet for storage purpose, manage and process data instead of a company's local server. From the specified infrastructure, the IT resources and services are provided with demand basis at a low price in a shared, multitenant and elastic environment, are known as transformational model for the enterprise. In recent trends, enterprises are expecting the cloud computing for helping them to provide a high performance than the existing as well as new, innovative services with a demand basis across network, computing and storage resources at lower price. When the services come to business, the "cloud" is completely different.

The "cloud" provides Software as a Service(SaaS), when a company subscribes for an application, access the service over the internet. A company can develop its own custom applications and its platform for the usage of all consumer, delivered by cloud is Platform as a Service(PaaS). Service providers Amazon, Google and Rackspace provide a backbone that can be rented out by other companies (for example, Netfix provides the services to other companies, but it accesses the services from Amazon). A Service Level Agreement(SLA) is an agreement between two or more parties, where one is the provider of the service and the other is consumer. The agreement can be legally binding as formal agreement or an informal agreement (for example, the relationships between the departments). SLA used for measuring, monitoring and reporting the performance of the cloud, based on the user's involvement or the capability to consume the resources. The A Service Level Agreement(SLA) will have the procedural definition with regard to Mean Time Between Failures(MTBF), Mean Time To Repair(MTTR) or of Mean Time To Recover(MTTR), the data rates, throughput, jitter and other measurable details. To ensure that SLAs are consistently met the specifications of the agreement, the agreement designed with specific lines of distinction, and the parties involved are required to meet regularly, so that an open forum is created for communication. The enforcement (rewards and penalties) of the contract should be enforced often, so it will be easy to do the modifications based on new information.

This paper presents a general view of Service Level Agreement (SLA) in cloud computing, arranged in sub sections as follows. Section II gives the service and deployment models, section III describes the necessity of cloud computing, section IV defines the importance of SLA, Section V defines the features of SLA, section VI classifies the SLA, Section VII shows the SLA based framework, Section VIII elaborates the literature, Section IX provides the issues and challenges and finally section X concludes the paper.

II. SERVICE AND DEPLOYMENT MODELS

There are three basic service models existing in cloud to provide the resources to the user. Recently the other service models also in cloud. A list of cloud service models are described as follows, Figure 1 shows an example of cloud service and deployment models.

A.Software as a Service (SaaS): The users can access the service provider's applications running on a cloud infrastructure on demand basis. The applications are delivered

to various client devices through the client interface such as a web browser (e.g.,Web-based email, Google Docs).

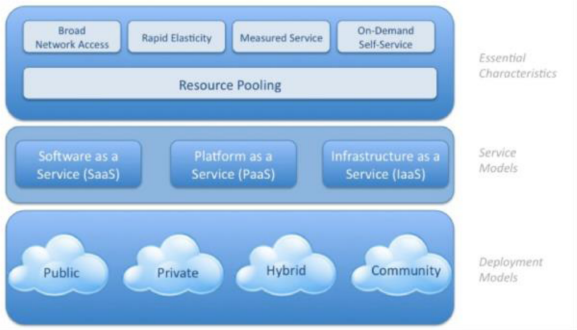

Figure1. Cloud Service & Deployment models [23]

B. Platform as a Service (PaaS): The user can access a platform which allows the user for developing, executing and managing the web based applications using the supported programming languages and tools without any difficulty of constructing and maintaining the infrastructure related to develop and install an application. (e.g., Google App Engine, Microsoft Azure).

C. Infrastructure as a Service (IaaS): The user can access the virtualized computing resources over the internet for processing and storage; more over the user can access other fundamental computing resources from the service providers (e.g.., Amazon Web Services).

D. Everything as a Service (XaaS): XaaS is a term refered to a rising number of services that are provided over the internet instead of providing locally. It is also known as "X-as-a-Service", "Anything-as-a-Service" or "Everything-as-as-Service". The examples of XaaS include Storage as a Service(SaaS), Communications as a Service(CaaS), Network as a Service(NaaS) and Monitoring as a Service(MaaS).

Cloud computing services can be organised in different methods depending on the organization and location. Four deployment models are used commonly, namely Private, Public, Community and Hybrid cloud.

E. Private Cloud: A private cloud is a deployment model of cloud computing that contains an individual and protected cloud based environment in which only the particular customer can use it. Comparing with other cloud models, private cloud will deliver computing power as a service in a virtualized. In private cloud model, the cloud(the resource) is only available for a specific single company providing that company with better control and confidentiality.

F. Public Cloud: The public cloud is a standard cloud computing model that referred as a multi-tenant background, where one customer demand a "server slice" in a cloud computing background that is accessed by a number of other users or tenants using the same public infrastructure.

G. Community Cloud: A community cloud is a collective effort in which infrastructure is accessed between several companies from a specific community with mutual concerns (security, agreement, authority, etc.), whether managed within the organizations or by a third-party and hosted within the organizations or externally.

H. Hybrid Cloud: A hybrid cloud is a combined cloud service consuming both private and public clouds to implement distinct purposes within the same organization. A company can increase their productivities by using public cloud services for all non-sensitive tasks, only depending on a private cloud based on requirement and certifying that all platforms are seamlessly combined.

III. THE NECESSITY OF CLOUD COMPUTING

1. **The Availability Of Resources :** There are several varieties of services provided by the cloud. The cloud services are accessed based on the IT needs of a company; instead of financing in a new infrastructure. In cloud computing, the hardware and software resources are available in CSP.
2. **The Flexibility Of Providing The IT Services In Cloud Computing :** While comparing with hosting services, the companies will be locked into contracts for a particular period(months or years), but in cloud computing, the services are provided on monthly basis or on consumption of resources. The companies can access the services quickly based on their IT needs.
3. **The Ability Of Refreshing An Aging Infrastructure With Less Price :** Nowadays the companies are virtualizing their critical applications. For virtualizing, the companies use the virtual machines associated with applications to execute on powerful servers. The cloud computing provides the facility without buying a new server.
4. **Economically Supporting For New IT Services And Many Users :** Cloud computing provides the facility to forward the applications to consumer's infrastructure and save the expenses for data center.
5. **Freeing The Staff Members And To Be Engaged In Other Projects :** Cloud computing provides infrastructure as well as management services, it allows the providers to offload the services; the cloud services are handled automatically based on customer request.So that the staff members will be engaged in other works; But this is a challenging one for the success of business.

IV. THE IMPORTANCE OF SLA

1. The consumer can get the information about the service providers.

2. SLA describes the complete information about the service and the type of services (SaaS, PaaS, IaaS) that are provided to a particular consumer.
3. SLA describes the purpose and objectives based on business level policies, which includes the part of the service provider and the customer.
4. The consumers will be able to identify the key security and management strategies of agreement.
5. SLA is used to monitor the quality of service, performance, response time from the service point of view.
6. The consumer can get the idea about the requirements for the management of the service in case of poor performance.

The applications forwarded from committed hardware into the cloud, need to reach the same level of service or extra as classical installations. SLA focuses on features of data center and the network to support the provider – to – customer. Usually the SLAs are based on "Output". The service, accessed by the customer is the main focus of the agreement. The service provider can demonstrate the value in an innovative way by developing with creativity, proficiency and the awareness of the service required to be delivered. And also, the organizations can specify the (service level) specification in which the service is to be delivered using the objectives those are related to the level of service. The specifications and objectives are known as "Input" of the SLA.

Managing the SLA focuses on two phases i. the negotiation of the agreement and ii. monitoring the performance in the real time. The SLA management focuses on basic schema with QoS (Quality of Service) constraints, negotiation, observing and implementation according to the policies mentioned in the agreement of SLA.

V. FEATURES OF SLA

1. Enhanced the satisfaction level of customer *:* A healthy defined SLA rises the customer satisfaction level, as it helps the provider to focus on the requirements of the customer and warrants that the effort is placed on the right direction.

2. Amended service quality*:* QoS requirements [19] are commonly identified in SLAs. SLA consists of pool of contractual agreement between service providers and customers that fact the nature, worth and objective of service to be provided. SLA may also identify the percentage of violations that can be accepted within a predefined time interval, before the service provider experiences a penalty.

3. Enhanced relationship between two parties*:* A healthy SLA specifies the compensation and penalty policies of the service. The customer can observe the services according to Service Level Objectives (SLO) identified in the SLA.

VI. CLASSIFICATION OF SERVICE LEVEL AGREEMENT(SLA)

1. **Service-Based SLA** : A contract between the service provider and the customer based on the required services to be delivered.
2. **Customer-Based SLA** : The customer-based SLA includes all the required services that are needed to a company and appropriate to an individual customer group. For example, the SLA between the provider and the finance department of a big company for the required services such as finance system, payroll system, billing system, procurement or purchase system etc.
3. **Multi Level SLA** : The SLA is categorized into different stages, each stage addresses different level of customers for the same services defined in the same SLA.

i) **Service Level SLA** : The service level SLA, includes all the relevant issues of the SLM (Service Level Management) for a specific service, with respect to the specific customer group.

ii) **Customer Level SLA** : The customer level SLA includes all the relevant issues of the SLM (Service Level Management) for a specific customer set, nevertheless of the facilities being used.

iii) **Corporate Level SLA** : The corporate level SLA contains all the basic issues of the SLM (Service Level Management) and suitable to the customers throughout the company. The defined issues are volatile, so that updates are not frequently required.

VII. SLA-BASED CLOUD FRAMEWORK

Cloud computing is one of the distributed system, and it consists of combinations of interconnected powerful servers and it delivers virtualized computing resources to consumers as service. In cloud computing the demand of resource and load will be varying by time. That is, the demand of resource in the late evening will be higher than the working hours. The physical resources are geographically distributed, the load of the provider can be shared based on the physical resources that are geographically distributed.

In cloud computing, the service providers provides the resources of IaaS, Paas, Saas as service. Before providing the service, the provider and the consumer make an agreement based on the service level resources, execution time, average response time, price and duration. But there may exist delay between the expected time(defined in the agreement) and the providing time. To reduce the delay, it would be negotiated based on defined requirement and regulation, so that mutually agreed level of requirements can be achieved.

In this cloud computing environment, there are number of service providers and consumers. Each service provider will be having many data centers that are geographically spread around the world. In this framework, the distributed data centers can be managed by using centric cloud management system by the service provider.

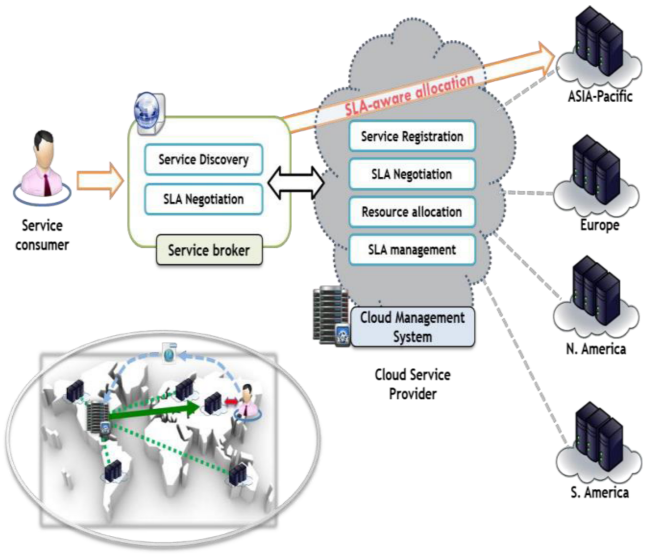

Figure 2. SLA Negotiation and Management Framework based Cloud Computing Environment[4]

Availability

Availability is a term used to ensure that an agreed level of functional performance for a greater than common period. Availability is stated as a percentage of uptime in a given period. The Table[1] displays the downtime that can be applicable for a certain percentage of availability, presuming that the system is needed to function continuously. Service level agreements mostly refer to monthly downtime or availability in order to compute service credits for matching monthly billing basis.

Availability %	Downtime per year	Downtime per month	Downtime per week	Downtime per day
90%	36.5 days	72 hrs	16.8 hrs	2.4 hrs
99%	3.65 days	7.20 hrs	1.68 hrs	14.4 min
99.9%	8.76 hrs	43.8 min	10.1 min	1.44 min
99.99%	52.56 min	4.38 min	1.01 min	8.66 sec
99.999%	5.26 min	25.9 sec	6.05 sec	864.3 msec

Downtime/Day =(1 – Availability%)*60Min*24Hrs

= 1 – (99.9/ 100) * 60 * 24 = 0.001 * 60 * 24 = 1.44min

VIII. LITERATURE REVIEW

Types of SLA violations

Linlin et al[1] mentioned the four main goals with respect to SLA –Oriented Utility Computing. i) The customer-driven management of service support regard to the customer's profiles and requested requirement of service. ii) identifying the computational risk strategy for defining and managing the risks involved during the application execution based on the requirements of service and the need of the customer. iii) The suitable strategies for managing the resources based on market level, around both customer-driven management of service and computation management of risk to maintain the allocation of resources based on SLA.

Yun et al[3] discussed the profit oriented issues and weak point in traditional technologies. i) The scheduling and dispatching in traditional technologies focus on the metrics such as database, average response time of the query and so on. ii) Though many business analysis tools exist for profit-oriented decisions, they are offline. To simplify these challenges, they developed a framework to support profit-oriented decisions efficiently in cloud database systems. The main idea of this framework is to look at the query buffer in front of the server for the profit information. A novel light weight data structure is SLA tree build based on the query buffered in front of the server with respect to its SLA.

Slavatore et al[5] presented that applications should act in support of their users and should be able to make the available cloud services in SLA negotiation and management. There will be practical applications that are aware about the status of the resources, services and interact with brokers, providers beyond the stateless SOA model and ultimately with other cloud actors with respect to persue the objectives of the user.

Rajkumar et al[7] discussed the hierarchical scheduling, and the need of job prioritization. To improve the performance in cloud environment, scheduling is followed in hierarchical order by scheduler which in both the Cloud Controller(CLC) and Cluster Controller(CC) Level. The job is executed one by one in the FCFS(First Come First Serve) method and RR(Round Robin) fashion, this may detect the deadline based jobs which are waiting in the queue for a long period. Surely, it leads to SLA violation and the consumer will not be satisfied. To avoid SLA violation, the hierarchical scheduling is applied using job prioritization in both CLC and CC Level.

Sherif et al[8] presented that in the fast growing and unpredictable environments, the approaches are based on capacity planning and workload forecasting techniques, that are in challenging conditions. When the allocated resource's capacity is defined for the average level of workload, there will be a low price incurred based on limited hardware usage but performance of the service will be a problem when it occurs a peak level of workload. The customers will not be satisfied because of poor performance and the income will be affected. Next, if the capacity of the allocated resource is defined for peak work load, most of the time, the resources remains idle position. In cloud computing environments, cost

management based on SLA overcomes these problems and gives an effective solution for controlling the cost of allocated resources based on the defined policies of the hosted application. The defined framework enables the provider to formally define the specific action rules of application to access scale out or scale in based on the status of the system which is monitored. The action rules are formally defined using a XML dialect that is executed by the Action Manager module, evaluates the conditions of the specified rules and triggers the execution with respect to the actions as necessary.

Pankesh et al[9] presented a WLSA framework with regard to XML and therefore in composition metrics, it restricts the ability of matching to syntactical. To highlight the descriptions and to improve the quality of matches, Semantic Web Technologies can be used.

Linlin et al[10] presented the customer-driven SLA based resource provisioning algorithms by limiting the resource, penalty to reduce the cost and improve the customer satisfaction by limiting SLA violations. To handle customer requests and infrastructure level heterogeneity, the provisioning algorithms focus on customer profiles and provides quality parameters, for an enterprise systems. The customer-side parameters and infrastructure-level parameters are considered to compare algorithms. The simulation results demonstrate that the algorithm minimize the total cost upto 54% and minimize the SLA violations upto 45% compared with other algorithms.

Due to dynamic nature of the cloud, continues monitoring on Quality of Service (QoS) attributes is necessary to ensure SLAs. Many other factors such as trust come in to considereation particularly for enterprice customers that may outsource its critical data.

Pankash et al [20] proposed a mechanism for managing SLA in cloud computing environment using the Web Service Level Agreement (WSLA) framework. It is developed for SLA monitoring and enforcement in a service oriented architecture. CSP are profit based companies that have incentive to cheat on the SLA. By providing the poor quality of resources to customers than specified in the SLA. The CSP can accommodate more users on the same hardware and increase their profit. The monitoring and verification of the SLA is typically performed on the CSP side, the user may not know the violations. To predict the violations Ryan et al [21] proposed auditing cloud Service Level Agreement on VM CPU speed. Third Party Auditor (TPA) is used in the framework to detect the CPU resource violations, CPU intensive calculation, transpose matrix multiplication to detect the violations done by CSP randomly.

The autonomic enforcement of a single business level objective is the maximisation of the revenue. Mario et al [22] proposed Rule-based SLA Management for Revenue Maximisation in Cloud Computing Markets. That revenue can be maximized by establishing a bidirectional data flow between market and resource layers: market brokers can perform negotiations that are more profitable if they use resource-level data, and the resource manager can help maximising the revenue if it manages the SLAs by considering this business level objective.

IX. ISSUES AND CHALLENGES

1. Computation management of risk to maintain the allocation of resources based on SLA [1].
2. Applications should act in care of their customers and should be capable to make the accessible cloud services in SLA negotiation [5].
3. User waiting in the queue for a long period to avail service [7].
4. Performance of the service affected because of peak level of workload. The customers will not be satisfied because of poor performance and the income will be affected [8].
5. Service cost related problems [8].
6. Not providing the specified amount of resources [21]

X. CONCLUSION AND FUTURE ENHANCEMENTS

SLA is a legal, formal and negotiated document that defines the service in terms of quantitative and qualitative metrics. The metrics which is involved in SLA should be proficient of being measured on a consistent basis, and the SLA should be evaluated by that metrics. SLA plays a role in the life cycle of the service. SLA cannot guarantee that the consumer can access the service as described in the SLA document. In future our work is focused on developing a approach to certify that the service is provided according the specified level of quality which is mentioned in the SLA.

REFERENCES

[1] Linlin Wu and Rajkumar Buyya, “Service Level Agreement(SLA) in Utility Computing Systems”, Technical Report CLOUDS-TR-2010-5, Cloud Computing and Distributed Systems Laboratory, The University of Melbourne, Australia, September 3, 2010.

[2] Jerry Gao, Xiaoying Bai and Wei-Tek Tsai, “Cloud Testing- Issues, Challenges, Needs and Practice” Software Engineering : An International Journal (SEIJ), SEPTEMBER 2011, Vol. 1, No. 1, pp. 9-23.

[3] Yun Chi, Hyun Jin Moon, Hakan Hacigumus and Junichi Tatemura, “SLA – Tree : A framework for Efficiently Supporting SLA-based Decisions in Cloud Computing”, Proceedings of the 14th International Conference on Extending Database Technology, 2011, ISBN: 978-1-4503-0528-0, Pages 129-140.

[4] “SLA-based Cloud Framework”, https://sites.google.com/site/gistcloudresearchgroup/services, Presented By GIST Cloud Computing Research Group.

[5] Salvatore Venticinque, Rocco Aversa, Beniamino di Martino Massimiliano Rak and Dana Petcu, “A cloud Agency for SLA Negotiation and Management”, Euro-Par 2010 Parallel Processing Workshops Lecture Notes in Computer Science Volume 6586, 2011, Online ISBN : 978-3-642-21878-1, Print ISBN : 978-3-642-21877-4, pp. 587-594.

[6] Service Level Agreement, http://en.wikipedia.org/wiki/Service-level_agreement.

[7] Rajkumar Rajavel and Mala, “Achieving Service Level Agreement in Cloud Environment using Job Prioritization in Hierarchical Scheduling”, Proceedings of the International Conference on Information Systems Design and Intelligent Applications 2012 (INDIA 2012) held in Visakhapatnam, India, January 2012, Pint ISBN : 978-3-642-27442-8, Online ISBN : 978-3-642-27443-5, Series ISSN : 1867-5662, pp 547-554.

[8] Sherif Sakr, Anna Liu, "SLA-Based and Consumer-centric Dynamic Provisioning for Cloud Databases", CLOUD, IEEE Sixth International

Conference on Cloud Computing, 2013, doi:10.1109/CLOUD.2012.11, ISBN: 978-1-4673-2892-0, pp. 360-367.

[9] Pankesh Patel, Ajith Ranabahu, and Amit Sheth, "Service Level Agreement in Cloud Computing", OOPSLA Cloud Computing workshop, http://corescholar.libraries.wright.edu/knoesis/78.

[10] Linlin Wu, Saurabh Kumar Garg, Steve Versteeg and Rajkumar Buyya, "SLA-based Resource Provisioning for Software-as-a-Service Applications in Cloud Computing Environments", 11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, 2011, ISBN 978-0-7695-4395-6/11, pp. 195-204.

[11] Dimension data white paper, "Comparing Public Cloud Service Level Agreements Questions to ask when evaluating public cloud SLAs", http://cloud.dimensiondata.com/sites/default/files/comparing_public_cloud_service_level_agreements_0_0.pdf, CS / DDMS-1249 / 02/13 © Copyright Dimension Data 2013.

[12] "Amazon EC2 Service Level Agreement", http://aws.amazon.com/ec2/sla/, Effective Date: June 1, 2013.

[13] CISCO, "Managing the Real Cost of On-Demand Enterprise Cloud Services with Chargeback Models", By Cisco, Aug 2010, http://www.cisco.com/en/US/services/ps2961/ps10364/ps10370/ps11104/Cloud_Services_Chargeback_Models_White_Paper.pdf

[14] Rajkumar Buyya, Saurabh Kumar Garg and Rodrigo N. Calheiros, "SLA-Oriented Resource Provisioning for Cloud Computing : Challenges, Architecture, and Solutions", CSC '11 Proceedings of the 2011 International Conference on Cloud and Service Computing, IEEE Computer Society Washington, DC, USA ©2011 ISBN: 978-1-4577-1635-5, pp. 1-10.

[15] Myerson, "Best Practices To Develop SLAs For Cloud Computing", Retrieved From http://www.ibm.com/developerworks/cloud/library/clslastandards/, January 2013.

[16] Linlin Wu, Saurabh Kumar and Rajkumar Buyya, "SLA-based Resource Allocation for Software as a Service Provider(SaaS) in Cloud Computing Environments", Proceeding CCGRID '11 Proceedings of the 2011 11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, IEEE Computer Society Washington, DC, USA ©2011, ISBN: 978-0-7695-4395-6, pp. 195-204.

[17] Mitchell Cochran and Paul D. Witman, "Governance and Service Level Agreement Issues in a Cloud Computing Environment" Journal of Information Technology Management, 2011, Volume XXII, Number 2, ISSN #1042-1319, pp. 41-55.

[18] Dr. Angel Luis Diaz, "Service Level Agreements in the Cloud: Who cares?" Retrieved From "http://www.wired.com/2011/12/service-level-agreements-in-the-cloud-who-cares/",14 December 2011.

[19] Stefano Ferretti, Vittorio Ghini, Fabio Panzieri, Michele Pellegrini and Elisa Turrini, "QoS – Aware Clouds", 2010 IEEE 3rd International Conference on Cloud Computing, 2010 IEEE, pp. 321-328.

[20] Pankesh Patel, Ajith Ranabahu and Amit Sheth, "Service Level Agreement in Cloud Computing", Report, Core Wright university Libraries, Kno.e.sis Publications, 2009.

[21] Ryan Houlihan, Xiaojiang Du, Chiu C. Tan, Jie Wu and Mohsen Guizani, "Auditing Cloud Service Level Agreement on VMCPU Speed", In proceedings of IEEE international Conference on Communications (ICC), 2014, INSPEC Accession Number:14545672, DOI:10.1109/ICC.2014.6883417, pp. 799-803.

[22] Mario Mac´ıas, J. Oriol Fit´o and Jordi Guitart, "Rule-based SLA Management for Revenue Maximisation in Cloud Computing Markets", International Conference on Network and Service Management – CNSM , 2010, ISBN 978-1-4244-8909-1 pp. 354-357.

[23] PCI DSS Compliance in a Cloud Computing Environment, http://zinatullin.com/2013/03.