CM- Measurement Facets for Cloud Performance

Dr. Gurdev Singh
Associate Professor and Head
Computer science & Engineering
Eternal University
Baru Sahib (India)

Shanu Sood
Lecturer
Computer science & Engineering
Eternal University
Baru Sahib (India)

Amit Sharma
Lecturer
Computer science & Engineering
Eternal University
Baru Sahib (India)

ABSTRACT
Cloud Computing is the highly emerging area in the computer world now a days by which the web applications are delivering the computing services through internet on demand. Cloud computing making business applications more mobile, which is of a great concern and making the cloud computing more popular and necessity for the large business organization to grow more rapidly. So, it is necessary to know that how to choose a cloud by which the clients will get the better computing services than the others. There are many web application metrics which are available to calculate the performance of the web applications and their dynamic nature, but to know about the performance of clouds, cloud metric is required which makes the cloud performance measurement a reality . Cloud metric deals with the dynamic nature of the cloud. There is no metric available which can calculate the performance of cloud. A cloud metric is required for evaluating the cloud quality for recognizing the best cloud. In this paper there is an effort to develop metrics which can measure the performance of a cloud by keeping the dynamic nature of cloud in mind. This paper covers measures on which a cloud's performance can depend. Each of the measure depends on some factors and that factors are also mentioned in the paper below.

Keywords: Cloud Computing, Cloud Metric, Virtualization, Web Applications.

1. INTRODUCTION
Cloud computing at its core is a single element by which computing services are delivered over the Internet, on demand from a remote location, rather than residing on one's own desktop, laptop or even on an organization's server. Cloud computing provides a variable usage based fee or even possibly for free it would contract with a provider to deliver applications, computing power and storage via internet.

Many of services can be seen in a web browser running on a simple handheld device or inexpensive laptop, rather than requiring a powerful computer. For example Google's Chrome browser and many more were designed to meet this need. A new layer of software is needed so that online applications can draw on this new computing infrastructure and centrally stored data in such a way that a computer operating system control access to the processing power and memory on an individual PC. Companies which are building these online platforms include Amazon, Google and Microsoft. Cloud computing is an approach that involves the creation and deployment of services over the internet, supported by a centralized computing infrastructure [1].

Measurement is the process by which numbers or symbols are assigned to attributes of entities in the real world in such a way as to describe them according to clearly defined rules [2]. A metric is always used to measure and understand the behavior of software or a program. To evaluate the quality of cloud, there is a need to assess and analyze its design and implementation using appropriate metrics.
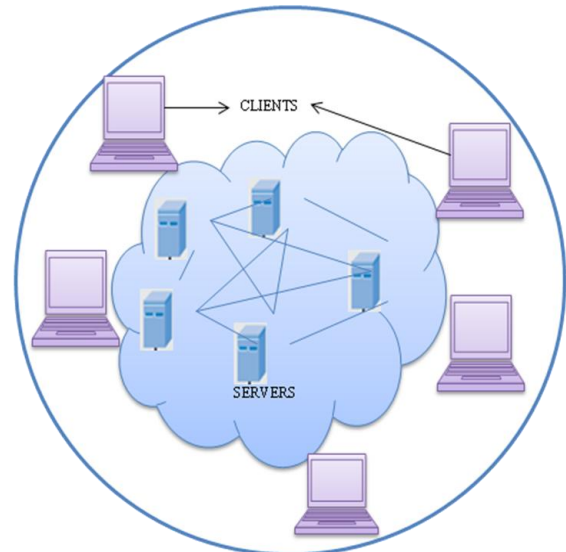

Figure 1: Cloud computing

Cloud metric (CM) can be used to measure the behavior of cloud which utilizes the resources from the computers as a collective virtual computer, where the applications can run independently from particular computer or server configurations. Cloud delivers its services through internet and provides the full user functionality of a software application by the web sites which provide Software as a Service (SaaS). Dynamic web sites provide regularly changing information to users about an organization, company etc. and utilize dynamically generated pages and maintain data for display in a database [3]. Cloud uses the dynamic web sites to deliver the web applications on demand. CM should follow some characteristics which help to evaluate cloud on each and every parameter which is necessary for a good quality cloud, so that a client can rely on it to choose the best cloud. In this paper that characteristics are listed and measurement factors are also considering them for evaluating a cloud performance. To

measure all facilities given by cloud there are some factors that have to be supported by the cloud. In this paper that factors on which a CM may reside are described. Each and every factor is examined and a particular set of metric measures are given in the paper which will help to develop a CM for a perfect measurement of cloud behavior.

2. CHARACTERISTICS OF CM

Characteristic describe the nature of a particular object. CM also has to follow some characteristics and these characteristics should be taken care of while considering the measurement factors for cloud. Characteristics which a CM will follow are described below:

2.1 Reliability

Reliability is the probability of success or the probability that the system will perform its intended function under specified design limits [4]. More specific, reliability is the probability that a product or part will operate properly for a specified period of time (design life) under the design operating conditions (such as temperature, volt, etc.) without failure [4]. The outcome of the measurement process is reproducible that is similar results over time for some different inputs and across many different situations. Cloud gets many requests simultaneously and will also give the similar results for some requests in a period of time so clouds have to be reliable.

2.2 Validity

The measurement process actually measures what it purports to measure [5]. Validity in a cloud should deliver the web application which clients really want and is valid for use by them. Cloud should understand what to deliver at what time and the delivered application should be the best among all choices.

2.3 Sensitivity

Measurement process is sensitive for the different inputs which occur at different times and in different situations. The measurement process shows variability in responses when it exists in the stimulus or situation. Cloud has to face many clients and their different requests at the same time, so it has to give response variably at that same time.

2.4 Robustness

Measurement process must tolerate the unexpected problems in some cases. Cloud is used by many clients at a single moment of time and every client tends to take the full advantage of the services, and the client must get that advantage too. So cloud has to be so robust so that any unexpected problem will not cause an interrupt in the service.

2.5 Simplicity

The process of measurement must be simple to use. The measurement of a cloud performance is so important that any client can easily choose a cloud according to its requirements.

2.6 Efficiency

Efficiency means the amount of computing resources and code required to perform a function. Cloud has to be so efficient that it should require minimum resources and code to perform any function and to serve any request. All the requirements of the users should meet for which cloud use minimum possible resources.

2.7 Flexibility

Measurement process is flexible for the modifications. Cloud may need modifications after some instance of time according to the change in technology and requirements. Flexibility will cause a measurement process to give the perfect results even after some change or modification.

CM has found good number of applications during measurement like reliability estimation, Source code coverage, test suite effectiveness etc. They provide meaningful and timely information, which may help us to take corrective actions as and when required. Effective implementation of metrics may improve the quality of the cloud performance and may help us to deliver the web services in time and within budget.

3. MEASURES FOR CLOUD

Cloud Computing is the implementation of engineering principals to obtain high quality applications through internet. Similar types of processes will be followed to make web applications of cloud as in traditional software's but with new ideas. Now days when the platform of programming has changed then it is difficult to develop the software only with traditional models. A cloud needs changes in traditional models which required for the development of online applications.

CM basically depends upon three categories of services of a cloud that is SaaS, PaaS (Platform as a Service), IaaS (Infrastructure as a Service).
The basic CMs are shown in Figure 2.

3.1 Availability

Cloud Services should be available maximum time [6]. The on-demand, elastic, scalable, and customizable nature of the cloud must be considered when deploying cloud architectures. Many different clients might be accessing the same back-end applications, and many provider are providing the cloud services has the expectation that only their application will be properly delivered to users. In cloud computing it is essentially required to gather the information instantly without making a user to wait and the gathered information should be related to each other.
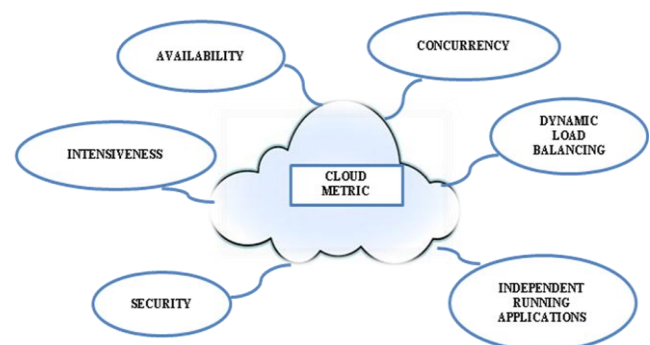

Figure 2: Measurement metrics of cloud

An Application Delivery Controller (ADC) [7] and Local Traffic Manager which virtualize the back-end systems were developed to ensure highest availability of requested resources.

Making sure that multiple instances of the same application are delivered in a scalable manner requires both load balancing and server virtualization techniques.

Applications are managed to define criteria relating to the number of concurrent users of an application or group of applications to maintain the availability of web applications in a cloud. The criteria will be based on:

3.1.1 ART

Application Response Time (ART) which tells the time in which a cloud application will give response to the clients for their requests. Cloud should response in the minimum possible time for a request so that the client will get the services as early as possible without waiting for any longer. Cloud which give response in minimum time is of good quality and which take long time to give response the quality of that cloud will be low.

3.1.2 CRT

Changes in response time (CRT) may vary according to the request in a cloud because for some requests the gathering of information will take more time than the others. The retrieval of data from the database of cloud may produce some change in the response time. The changes in response time should remain transparent to the users so that they don't know about that they have requested for information whose collection is taking time. Cloud which keeps this response time transparent to users is of good grade and its performance is better than those clouds which are not taking care of this transparency or whose response time is taking longer.

3.1.3 NSR

NSR (Number of session requests) is number of requests given at a time and to maintain these requests within the timing sessions. Session is the time within which a response will be given by cloud or a request will be given to the cloud. In cloud for every application's session timings have maintained and the request/response have to be given at that time limit only. How many number of requests are given at a time and how to maintain these requests within that timings is of a great concern, so cloud with maintained session requests/responses is the sign of a better performance and beneficial cloud.

3.1.4 TVS

It is the Amount of time visitors spend in the application(s). Clients may need a little time to work with an application or may need a large amount of time with others so cloud availability performance also depends on it because the clients spend slight time in application or the clients spend huge time in application all needs the full availability of cloud applications. A cloud keeps this in consideration is a high caliber cloud.
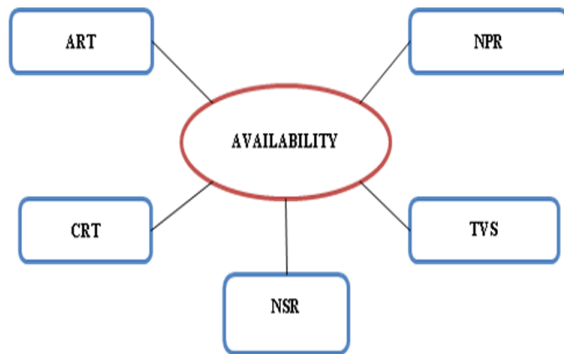
ART: Application response time
CRT: Changes in response time
NSR: Number of session requests
TVS: Amount of time visitors spend in the application(s)
NPR: Number of page requests for the application(s)

3.1.5 NPR

Within a cloud application there are many pages which correspond to different requests. So how many numbers of page requests (NPR) a client wants to serve with a single application depends on page requests by the client. Every page requests of application(s) must be satisfy with the complete availability even if it needs to make the multiple instances of a single page. Cloud concerns with this are a superior value cloud.

3.2 Concurrency

A Large no of users may assess the cloud web software at one time; every user wants an uninterrupted service through it. Concurrency in a system provides many features [8]. There are many instances of the same application running simultaneously and each is giving the full support of application. Each object lives in its own thread and deploy on different units. Threads define system interfaces to support the source portability of applications. Each thread works independently and can deploy at the same time or different on different units. Concurrency criteria in the cloud based on:

3.2.1 HME

It is the Hypermedia (HME) which allows accessing of information in a nonlinear manner so that a cloud may serve the different page requests from a client simultaneously. HME can be used to make web applications of cloud content driven (CDR) where the different application are running concurrently.

Content driven application brings users and information together. Access of the content gives the definition that how end user views the information. CDR applications in a cloud are implemented as a centralized or virtual client/server model.

A Content Management System [9] is implemented for a web application or set of applications, for easy creation of Web site and update and expansion. A cloud which provides the HME to run its applications concurrently and assure to serve the client's request by making the applications CDR is a moral cloud design.
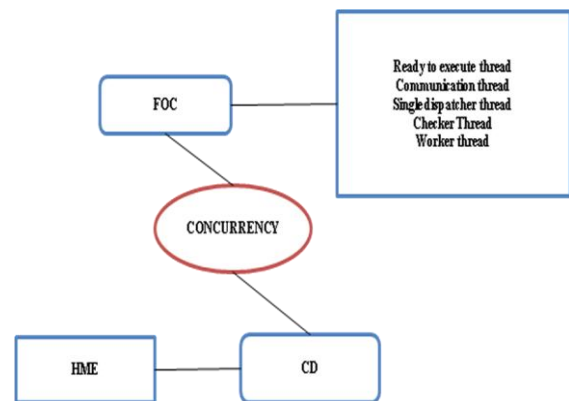
Figure 3: Availability criteria in cloud

Figure 4: Concurrency criteria in cloud

FOC: Flow of control
HME: Hypermedia
CDR: Content driven

3.2.2 FOC

It is the Flow of control (FOC) which is the creation, control, and termination of multiple FOC in the same cloud under the assumption of a common shared address space. Interfaces and functionality to support multiple flows of control, within a process called threads. The scope of a thread for the cloud is to define a sufficient set of functionality to support multiple threads of control within a process to run concurrent applications and to support the real time application [10] domain defining sufficient performance constraints and performance related functions to allow a real time application to achieve deterministic response from the cloud. The management of control flow provided by a cloud for a better performance of cloud is an essential feature. Cloud which supports this feature is healthier than the others.

Different threads which can be used for the cloud to support concurrency are ready to execute thread, communication thread [11], and single dispatcher thread, checker Thread, worker thread.

Ready to execute thread: In a cloud service requests coming from the clients which demands to make service instances are assigned to this thread.

Communication thread: To initiate the service instances in a cloud this thread is used.

Single dispatcher thread: It dispatches the serialized requests to the correct virtual server.

Checker Thread: To check whether the cloud server are running the applications properly or not.

Worker Thread: This thread handles accepting connections. It accepts a handle to the connection and handles processing.

3.3 Dynamic load balancing

In cloud multiple instances of the same application have to be delivered in time and it requires both load balancing and server virtualization. There can be an unpredictable load at any time which is handled in such a way so that it remains transparent to the user through virtualization servers. While there is a situation of unpredictable load the request will be send to a server or a virtual server having the minimum load.

Dynamic Load balancing and application delivery technology is used to increase the capacity and reliability of cloud. It forms a key value-add service that allows the clients to achieve improved service levels and gain detailed visibility of the performance and health of their services. Without load balancing, computing infrastructures cannot scale. The performance measurement considers this to evaluate the overall performance of cloud.

Two well-known load balancing algorithms, Round Robin (RR) [12] and Shortest Queue First (SQF) [12], were analyzed.

3.3.1 Dynamic load balancing for SaaS

SaaS services of cloud require dynamic load balancing for scalability and reliability reasons and use a shared load balancing platform that is completely concealed from end-users. When the scalability will be higher the cloud's application delivery will automatically be higher and measurement will give a good evaluation results for cloud.

3.3.2 Utility load balancing

Utility, load balancing is common in Cloud and hosting platforms. Clients use a 'self-care' [13] interface to provision virtual machines and can configure simple load balancing rules using this interface. The load balancing configuration is then automatically provisioned on a shared load balancing platform. A cloud provides this type of load balancing give all benefits of cloud to client.

3.3.3 Dedicated load balancing

High-value hosting services provide end-users with dedicated load balancers, giving them full control of how their services are delivered [13]. A subscription model provides usage-based pricing with minimal up-front costs. More sophisticated IaaS platforms offer users a range of pre-packaged services in the form of virtual machines. So a dedicated load balancing in cloud will make its performance high.

Application delivery services may be deployed using measurement of dynamic load balancing provided by a cloud in this fashion will give a client all of the benefits of owning their own, dedicated appliance, while running on the Service Provider's shared virtualized platform.

3.4 Intensiveness

In cloud software delivery is through internet which resides on a network and must serve the needs of a diverse community of clients. Web Applications are network dependents and so is cloud. Several parameters play a role in the request demand shown at every Web site [14]. Some of these parameters are:
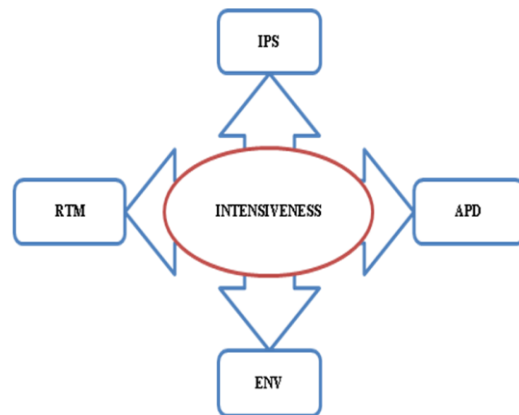

Figure 5: Intensiveness criteria in cloud

IPS: Information popularity and scope
RTM: Response time
APD: Appealing design

3.4.1 IPS

Intensiveness of a cloud depends upon the Information popularity and scope (IPS). The web applications with the popular information will be accessed most by the clients. Cloud considers this information and the number of time it is accessed and provides the maximum availability for this type of information is a moral cloud.

3.4.2 RTM

Response time (RTM) by for a request by a cloud is very important and it should be as minimum as possible. Cloud highly

depends on the response time for an application which has high demand, this is of a great concern and if provided gives a great performance.

3.4.3 APD
Appealing design (APD) of the applications must have to be so effective so that it will give a great impact on the client. Effective design of any application based on its graphical user interface and it should be interactive and easy to use. Cloud performance based on it so that it provides the applications which can be easily accessed. This facility provided by the cloud makes it so valued.

3.4.4 ENV
Navigate from one application to another have to be as easy as possible and should not take longer time which we call easy navigation (ENV). Switching from one application to another in the cloud has to be so friendly and simple which makes a cloud intelligible and increase its performance.

3.5 Security
Cloud is on world network, so there is need for securing the contents of web applications. Strong security measures are to be taken for protecting the information and data of web applications. The enterprise/personal that adopts the service of cloud computing, the necessary condition that both of the server and database on the front end must be trusted has to be satisfied [15]. With the increasing popularity of enterprise cloud computing and its public connectivity via the internet it is the next frontier for viruses, worms, hackers and cyber-terrorists to start probing and attacking [16].

From the IaaS on the bottom foundation layer to the PaaS on the middle layer and the SaaS on the top layer, cloud storage is always an important key factor for implement the application of cloud computing. Especially, the storage resource of the IaaS on the bottom foundation layer is the most important factor for supporting the regular operation of networking services. On-line storage and data access can be achieved via web-based application interface (API) [15] for all of the users of cloud computing in anytime at anywhere through wired or wireless networking devices. Cloud computing allow a client to upload their data on the client side. In order to assure the information security and data integrity, Hypertext Transfer Protocol Secure (HTTPS) [15] and Secure Shell (SSH) [15] are the most common adoption. However, data security on the cloud side is not only focused on the process of data transmission, but also the system security and data protection for those data stored on the storages of the cloud side. Cloud security can measure by taking some factors into consideration:

In the following, some consideration regarding the security issue must be focused for metrics:

3.5.1 SSP
The service provider on the cloud side must provide the assurance of storage and system protection(SSP) for those users that adopt the service of cloud computing to avoid the occurrence of storage damage and system failure that might probably induce the problem of data loss and generate the unnecessary arguments. Cloud performance depends on this issue to relax the clients so that they can keep their sensitive data too on cloud side.

3.5.2 DPR

Data protection (DPR) is a big issue regarding the security of data in a cloud. For those data stored on the service provider of the cloud side must not be accessed or spoofed by unauthorized user or intruders and even the employee of the service provider without the authorization and authentication of request [17] that presented by the user of the client side. In addition, the service provider must present the assurance of data integrity for the user of client side. A cloud provides better authorized access than others is more beneficial than others.
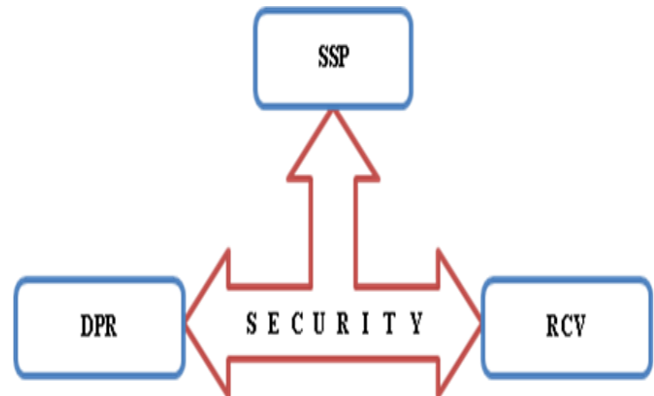

Figure 6: Security criteria in cloud

SSP: Storage and System Protection
DRP: Data protection
RCV: Recovery

3.5.3 RCV
Data in the cloud side may loss due to some unexpected situations, so Recovery (RCV) of that data is essentially required. Cloud should have a disaster recovery protocol [18] to protect user data. Due to some unwanted conditions or may be due to any natural disaster the data which lost on the cloud side have to be recovered. The recovery system has to be so strong so that a client will never face any problem for data access. Cloud providing a strong recovery system is more trustworthy.

3.6 Independent running applications
Application in the cloud runs simultaneously or parallel. These parallel running applications may be same or may not be same. A client may request for the application either at the same instant of time or after a gap in time. But when these applications will run they should not interfere with each other. In cloud the possibility of running the same application by many clients is very high, but this should be transparent to the client. Every client should get the delivery of the application individually; no application should depend on each other for running. Dependence between the applications can make them slow and the possibility for interrupted delivery of applications will increase.

Uninterrupted delivery of applications is achieved through server virtualization and scheduling according to the requests. Scheduling of the applications may depend upon time utility factor like algorithm Earliest Deadline first. [19]. Other algorithms like First in First Out and Round robin Scheduling are also used for this type of scheduling. Cloud providing the better scheduling algorithms provide better performance.

4. CONCLUSION

These days there are many companies who are developing the cloud to meet the requirements of clients and to satisfy their needs. Every cloud may develop with some shortcomings which should be known to the clients while they are about to use a cloud. For finding out those shortcomings cloud evaluation should be done. CM is a better solution to find out a cloud's performance which can be used to compare the performance of different cloud. Clients can then easily make the decision that which cloud they can use according to their requirements. The cloud performance evaluation can be done on some important factors which are required to make a cloud flawless and are considered in the paper above. So a CM will meet the factors to find out a cloud which will give the better performance than other clouds.

5. REFERENCES

[1] Ian Bott "How Cloud Computing Works" FT Graphic 2008.

[2] Norman E. Fenton, Thomson Publications, Fifth Edition 2005.

[3] David Cleary "Web Based Development and Functional Size Measurement" IFPUG Annual conference.

[4] Kareim M. Sobhe, Ahmed Sameh "Multi-Channel Clustered Web Application Server"

[5] Sheikh Umar Farooq, S. M. K. Quadri, and Nesar Ahmad. "Software Measurements And Metrics: Role In Effective Software Testing." International Journal of Engineering Science and Technology (IJEST) 3.1 (2011): 671-680.

[6] "Making Cloud Service Continuity a Reality" NetPrecept Software Ltd.

[7] Peter Silva "Availability and the Cloud" Cloud computing offers IT another tool to deliver applications. While enticing, challenges still exist in making sure the application is always available. F5's flexible, unified solutions ensure high availability for cloud deployments. F5 Networks, Inc.

[8] "Complete System and Enviourmental Management" www.streamlinecomputing.com/files/COMMAND_Web_V4 .pdf.

[9] P.M.Chawan, Suman Ninoriya "Towards the Quality Measurement of Web Content Management System" International Journal of Advanced Engineering & Application, Jan 2011 Issue.

[10] "Threads" The open group version 2 http://pubs.opengroup.org/onlinepubs/007908799/xsh/thread s.html

[11] Kareim M. Sobhe Ahmed Sameh "A Clustered Web Application Server" et al. / (IJAEST) INTERNATIONAL JOURNAL OF ADVANCED ENGINEERING SCIENCES AND TECHNOLOGIES Vol No. 4, Issue No. 1, 081 – 088

[12] Mário Meireles Teixeira Marcos José Santana "Analysis of Task Scheduling Algorithms in Distributed Web-server Systems"

[13] Zeus "Load balancing and application delivery for service providers."

[14] "Modeling Data Intensive Web Sites for Personalization, Integrity and Performance" Christos Bouras, Agisilaos Konidaris

[15] H. Pham, "System software reliability", Springer, 2006.

[16] Dr. Deepshikha Jamwal, Abhishek Singh Sambyal, and Prof.G.S Sambyal "Cloud Computing: Its security & Privacy Aspects". International Journal of Latest Trends in Computing Volume 2, Issue 1, March 2011

[17] M.Sudha, Dr.Bandaru Rama Krishna Rao "A Comprehensive Approach to Ensure Secure Data Communication in Cloud Environment",

[18] "Data Protection", http://www.bangor.ac.uk/ar/ro/records management/data protection/whatis.php.en.

[19] Shuo Liu Gang Quan "On-line Scheduling of Real-time Services for Cloud Computing", Electrical and Computer Engineering Department, Florida International University.