Network Traffic Analyzer
And
Simulator for Cloud Ecosystem

By

Santosh Bhattarai
(214678)

A thesis submitted to the University of Stavanger
in partial fulfillment for the degree of

Master of Science in Computer Science

University of Stavanger
Department of Electrical and Computer Engineering

Abstract

Cloud computing is the computing service that is access through the Internet. The term cloud refers to the Internet and computing refers to the services that is provided through the Internet. The cloud provides the enormous services to the user providing the entire infrastructure in the cloud server with reducing the cost of infrastructure development. The cloud provides the service in request of customer to fulfill their business needs. Actually the performance of cloud depends on various factors which depends the reliability of customer towards the Internet. Among those various affecting factors, one of the major factors is the network traffic which makes delay in response of the request which ultimately increases the customer dissatisfaction. In this paper, we analyze the network traffic and sampling of network traffic in the cloud ecosystem. The data is migrated to cloud and the network traffic is analyzed and sampled during the process.

Our main aim is to analyze the network traffic and traffic sampling of the cloud ecosystem. For this process the simulator tool record the start and end time of data transfer to cloud environment. Then from this record, the total time taken in different network traffic scenario can be visualized. It helps to compare the traffic rate while varying the data size on migrating to cloud. Furthermore, the rate of transfer of data packet is calculated. It also calculates the average rate of transfer of data under different network condition. In simple the tool migrates the data to cloud platform, analyze the network traffic and also does sampling of traffic when there is heavy network traffic during the migrating process. Monitoring the performance of CPU and memory is another important feature which provides the information of the performance of the system. We hope this work helps the user or any cloud migration analyst to study the network traffic and to provide satisfaction over cloud services. They can estimate their total time taken to migrate their data in cloud on the basis of data size. It also provides information about other network parameter like the transfer rate of packet, average rate of flow of data in different network condition, relation of data size and time, the bytes of data transfer to and from the network in each time interval.

Acknowledgement

First and foremost, I would like to convey my warmest gratitude to my supervisor Professor Erdal Cayirci, who provide me the opportunity to conduct my study in his research group. The words are not only enough to describe his guidance, generous contribution of knowledge and experience. I always appreciate his valuable comments and encouragement from the beginning of my study till to the end. He is the hard-working professor and I believe his academic achievements will continue to increase.

I take this opportunity to express my sincere thanks to all the members of Department of Science and Technology for their help and support. I also like to thank to all my friends who directly or indirectly help in this work to make this success.

Table of Contents

List of Figures

List of Tables

1 Introduction

Cloud computing (Fershtman and Gandal 2012, p. 1) is defined as the internet based information technology where everything are done over the internet connection. The word cloud simply means an internet and computing means the services that can be accessed directly through the internet. In the world of Information Technology, the cloud computing services bring a great revolution to accomplish the task in the easy and simpler way. Technically taking about cloud computing, it is a collection of cloud data server or as a collection of computer servers maintained by the cloud provider to provide the cloud services to the users. The cloud computing service is very easy to use as it can be easily accessible from any devices like computer, laptop and mobile where there is the internet connection service. This service seems to be easy and portable as every infrastructure should not be available in the consumer premises.

1.1 Definition of Cloud Computing

There are many definition of cloud computing that the researchers have described in their own way. Barkley RAD defines Cloud Computing as,

"Cloud Computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the datacenters that provide those services. The datacenter hardware and software is what we will call a Cloud. Thus, Cloud Computing is the sum of SaaS and Utility Computing, but does not include Private Clouds." (Armbrust et al., 2009, p6)

1.2 History of Cloud Computing

The history of cloud computing is long evolving from mainframe computer finally to the cloud computing environment. There is medium level progress during transition from mainframe to the Internet. The concept of personal computer came from the mainframe computer and gradually to client server concept and then to the world of the Internet where every people are connecting.

After the Internet with the concept of technology and network there is rapid improvement to the Web Services then to Grid Computing similarly evolving to the Network Computing, Utility Computing and finally to the Cloud Computing (Dharmanto 2012).

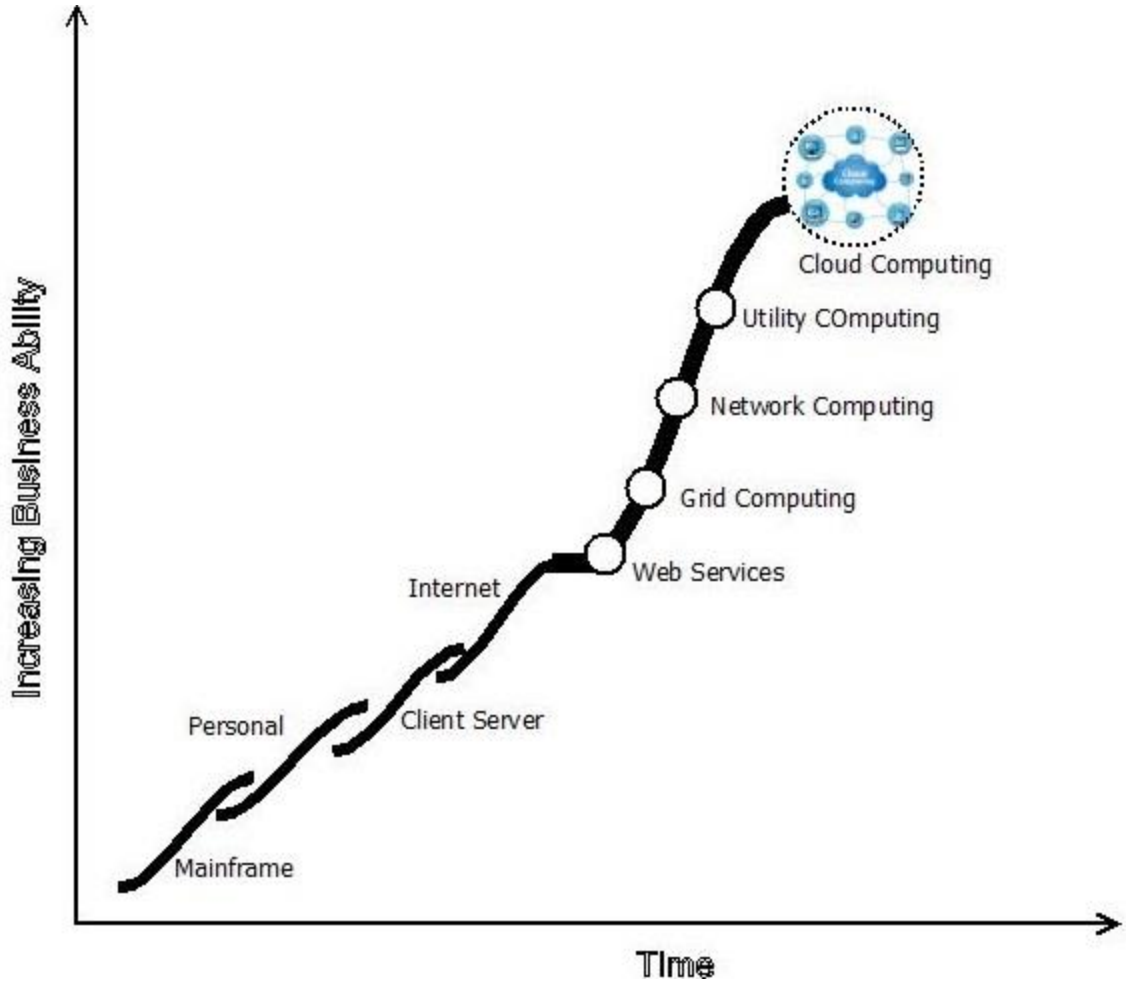

Figure 1: **History of Cloud Computing**

1.3 Importance of Network Effects in Cloud Computing

The application software is designed for the specific operating systems where the network effects play a vital role in determining the market equilibrium state (Fershtman and Gandal 2012, p. 5). There is a very strong effect of network in any on premises market where the customers prefer to select the operating system that is widely used by many computers which is a direct network effect. On the other hand some customers prefer the operating system that can support different application software which is an indirect network effect. Basically the developers prefer to develop the software for the operating system that is used by many users so there comes the network effects plays a importance role.

Now, while moving to the cloud ecosystem, the cloud will affected by the strength of the network effects of the virtual operating system. Here network effects comprise both direct and indirect effects. This completely changed the pattern of competition in the market compared to the tradition systems network effect theory. Here in the cloud environment the indirect effect is

seen to be weak in this system. This makes that the multiple platform exist in the equilibrium in the cloud. In fact, single platform is less likely in competition in cloud platform compared to the on-premise platform.

The importance and its benefits over the conventional practice is uncountable that always motivated me to study in more depth about the cloud ecosystem and want to analyze the effect of network traffic in the system. The network traffic and the cloud ecosystem have very close relationship. The effect in one side will have adverse effect on other side. So due to this factor, this topics always attract my interest.

Cloud computing models (Benson 2013) are very easy to operate by the end users with basic information. It does not require any sort of special training to operate the system. The end users can easily determine what is available to use simply by browsing the interface. The end user knows how much it is going to cost for the consumption of the desired resources and does not need to rely on somebodies expertise to purchase or to use the resources. So, it does not need any business to hire a personal staff to assist a customer to operate the system.

The cloud service goes beyond the network, computers and the supported browsers. The cloud solution can be accessed from any network and from any devices where the network connection is available. It is not limited with the particular web browsers and can be opened either by IE, Opera, Chrome and Safari etc. The technology is being advanced and users do not want to tolerate providing notification of not supporting operating system, browsers or any other compatibility alerts. Moreover, the cloud solutions are designed to work from any devices irrespective to the operating system.

The cloud computing also uses the content of resource pooling which not only increases the potential of the system by combining them but also minimize the risks and time taken by the single unit. It is the time that IT professionals need to broaden their mind abandoning the old thinking. It enforces the IT professionals to deal with the scalable solutions which can meet the needs of the business.

1.4 Motivation towards Network Traffic Analysis in Cloud Ecosystem

Cloud ecosystem is a rapid growing technology in the field of IT where every organization are slowly migrating their working environment to the cloud due to the useful features provides by cloud service. This days Microsoft and Google have already provided service application for their cloud platforms. Office 365 is the Microsoft most popular and powerful cloud version of Microsoft office software. Most of the people are using desktop version since the day but Office 365 is the cloud version which is gaining its popularity.

Among them one of the online services is the SharePoint online which also called as Office 365. It is the online Content Management System where any organization is interested to migrate their data in SharePoint due to the feature that the SharePoint have provided to the user. So, this always motivated me to do some research on cloud selecting SharePoint online as a cloud environment. Everybody is familiar to cloud services and here we have chosen SharePoint online

(Office 365) for our work. We want to analyze the network traffic while migrating data to the cloud platform.

In the process of migration to cloud, the simulator record the start time and end time from where the total time taken can be calculated. The time taken to transfer can vary with the data size so the tool finds the effect of network traffic under various data size. Furthermore, the size of data packet can be determined in certain time interval and more can be evaluated to find the average transfer rate of data in various network conditions.

1.5 Opportunities through Cloud Computing to the Market Players

The cloud computing is modifying the ICT ecosystem which come up to change the business roles in return providing the opportunities in enterprises (Dharmanto 2012) and providers level.

Some of the areas that the cloud computing provides opportunities are:-

1. Small and medium enterprises

The small and medium enterprises use the cloud computing as their opportunity **to improve flexibility** and also **to minimize their budget for IT Systems**. In addition, the ownership responsibility of hardware and software will also get reduced.

2. Hardware and Software Providers

The cloud computing promote the business growth of hardware and software providers because it requires more hardware and software resources which than provides the more opportunities for hardware and software providers.

3. Large ICT Enterprises

The cloud computing provides the good opportunity for business transformation into large ICT enterprises.

4. Other market players

The cloud computing provides opportunities f**or application developers, application integrators, application providers, content providers** etc. The operators of cloud provider serves as the intermediate between the above listed market players and operators.

The cloud service (Mishra 2014) helps to reduce the complexity of networks from user prospective and makes the network simple in user level. The user does not have to buy the software license which makes the process easier. The cloud provides an advance services e.g. e-mail which a single company can't afford or develop so cloud help to make the feature available. The system is scalable, reliable and efficient. Furthermore the information and data in the cloud

level is secure. So the various amazing feature of cloud always motivates to study its feature, measure and analysis the network traffic effect in the cloud ecosystem.

The cloud computing bring a revolution in the world of the Internet and a very important part in today's world of technology. The important factor that affects the service of cloud providers is the network traffic. So, it motivated to analysis the effect of network traffic in the cloud ecosystem.

1.6 The aim of Network Traffic Analyzer in Cloud Ecosystem

The delivery of the service can be said as computing service over the Internet and simply called as cloud computing. The popular cloud services are Gmail, Yahoo, Hotmail etc. Here instead of running e-mail program in our computer we simply access the remote e-mail account and utilize the feature. Actually the software and storage do not exit locally and will available only in service computer let's say cloud.

Although there exists some of the issues in cloud computing but the purposed work focuses on the network traffic analysis between the cloud users and the cloud service providers. The proposed work analyze the response time which is send by cloud providers in a specific size of data transfer in the cloud ecosystem. The simulator tool record the start and finish time of data transfer from where the total time can be calculated. Furthermore it analyzes the flow of packet of data in specific time interval and calculates the average transfer of data in different network conditions.

2 Related Work

2.1 Cloud Ecosystem

Cloud computing (Microsoftlearning, 2010) is the computing service that provides in the Internet. Cloud computing service consists of highly optimized data centers and it is the delivery of different computing service rather than the hardware, software and information. There are different types of services available namely SaaS, PaaS and IaaS which satisfies the user request.

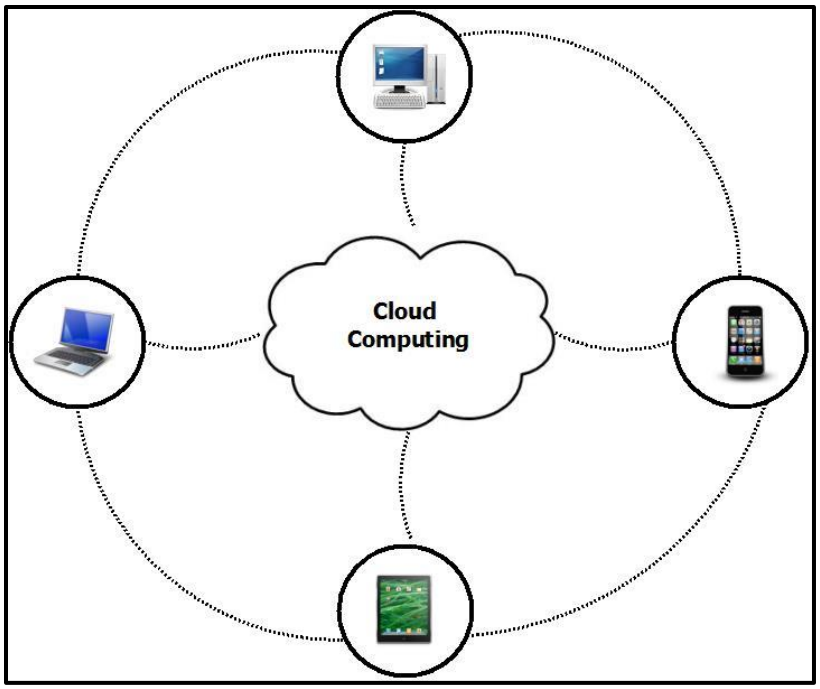

Figure 2: Cloud Computing

2.2 Cloud Components

Any system need successful implementation (Exforsys 2009) of its components similarly the cloud computing also requires the proper implementation of its components. Any component has its equal importance and without any one component the system can't be implemented. The system also can't be implemented by the single person.

The cloud computing requires manpower of various expertise, experience and knowledge so the cloud computing is expensive. Even though it is used by many company because the advantage and services the cloud provides is more than its initial spending so today many users are attracting towards the cloud environment.

It has three components. They are:-

1. Client Computers

Client is the devices through which the end users interact with the cloud and use the services. The client can be Computers, Laptops, Mobile, Tablets etc.

2. Data Center

Data Center is the collection of servers where application is placed so the user can access through the Internet.

3. Distributed Servers

The server is distributed in many places so when client request the services they got the prompt result as if the servers are next to them. The servers are distributed geographically and when the user requests the server next to them, will response and got the fast solution.

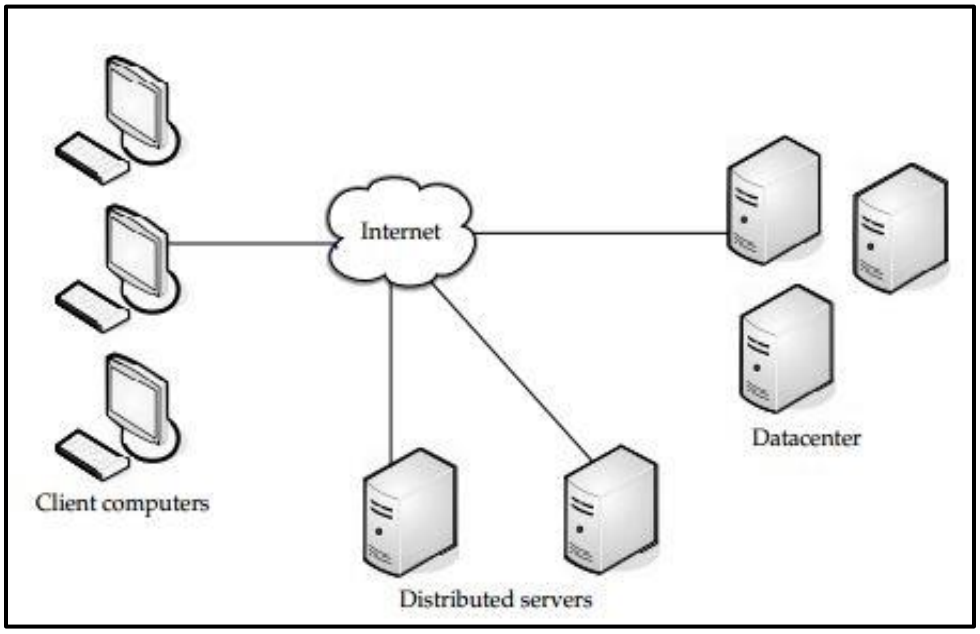

Figure 3: Cloud Components

2.2.1 Cloud Deployment Models

A cloud deployment model represents specific type of cloud environment where the model is distinguished in terms of its size, ownership and access.

There are commonly three types of deployment models:

1. Private Cloud

The private cloud (Arcitura Education Inc.) is owned by the single organization only. The private cloud uses the cloud computing as a centralized access of IT resources by the different department of the organization. The administration of the cloud service can be handled by the internal or the outsourced staff.

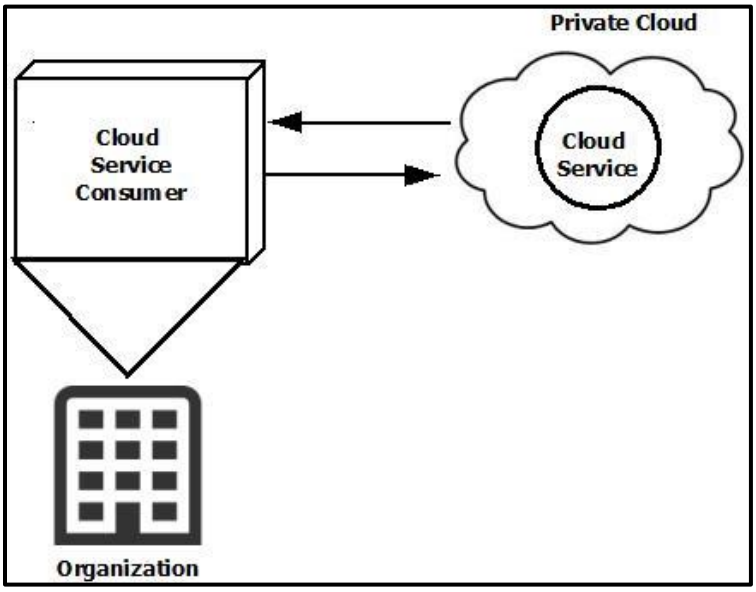

Figure 4: Private Cloud

2. Hybrid Cloud

A hybrid cloud model is the combination of two or more than two deployment model. The cloud service consumer can choose to access sensitive data through the private deployment model or the less sensitive data through the public deployment model. The combination of two deployment model is called a hybrid deployment model.

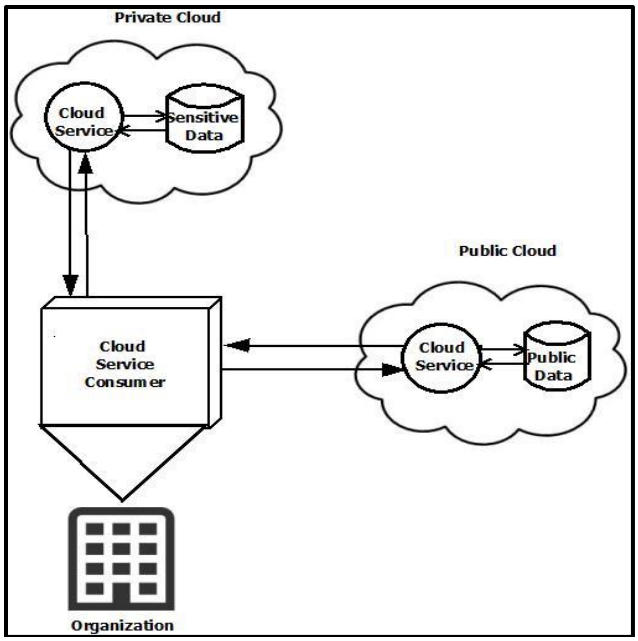

Figure 5: Hybrid Cloud

3. Public Cloud

A public cloud is an environment where the cloud services can access publically. The IT services offered to the consumers generally at no cost or commercially from the advertisement is public cloud. The cloud providers are responsible for the regular maintenance for the cloud services and the IT resources.

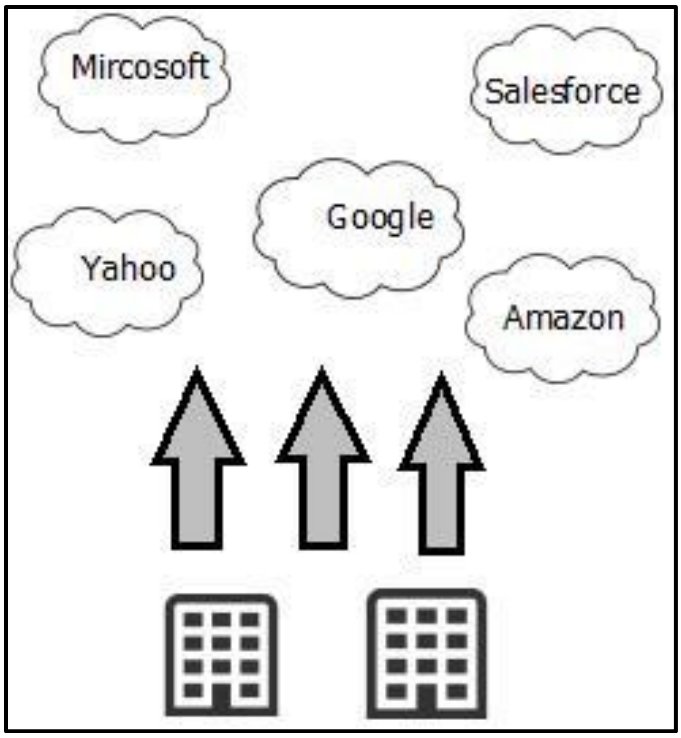

Figure 6: Public Cloud

2.2.2 Service Models

The Service Models in the cloud computing comprises of three different models namely, Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). An organization may obtain any combination of these available service models depending upon their organizational needs.

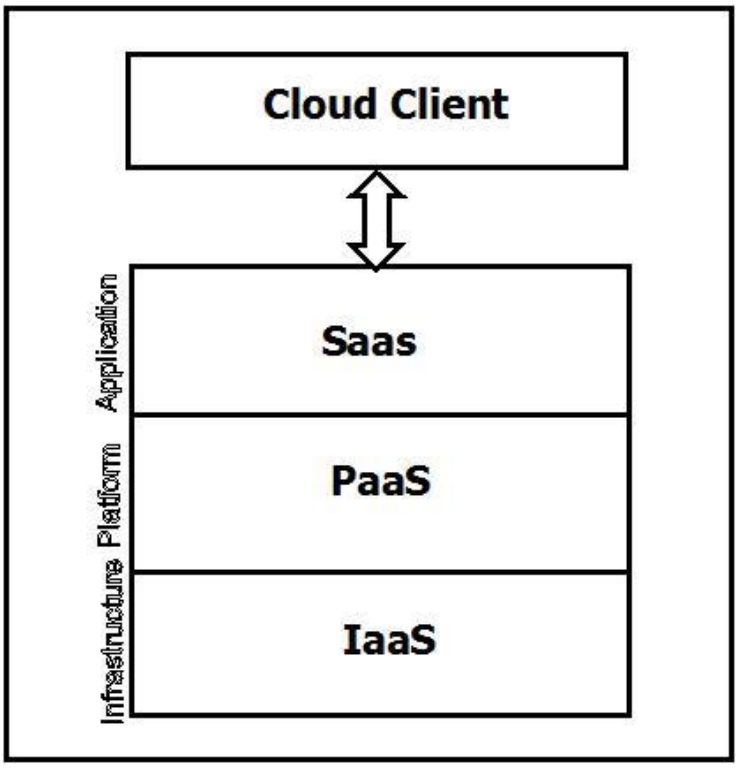

Figure 7: Service Model

1. Software-as-a-Service(SaaS)

Software-as-a-Service (cloud.cio.gov, 2014) is a service model where the software and its associated data are provided by the third party such as a Cloud Service Provider (CSP). Generally the consumers access the service through the web browser from their devices like computer or mobile device. Generally the organization does not have to buy the software but instead the CSP license the SaaS to the organization which then makes multiple users to access the software. Normally, service is used by the agency, staff or members. Some of the available software applications are email, word processing, office 365 and other management tools. The business task that is performed locally should be migrated to cloud to access by the authorized user from anywhere, anytime.

2. Platform-as-a-Service(PaaS)

Platform as a Service is a service model where the cloud service provider provides the application development platform to the organization as an agency. The agency developers can use the platform to develop, test, manage and host the application. The agency feels the relief of cost and management burden of application development. It mainly focuses for the developers

and the application managers which provide the common and consistent platform for application development.

3. Infrastructure-as-a-Service(IaaS)

Infrastructure as a Service is a service model where the cloud service provider provides the software and hardware where the customer can build their customized environment. The CSP provides the unmanaged resources and the customer have to managed installing the operating systems, software, storage etc. The customer itself is responsible and has full control for the management of the computing environment although the CSP is responsible for maintaining the physical equipment.

2.2.3 Cloud Essential Characteristics

There are five cloud essential characteristics of cloud computing is defined by National Institute of Standards and Technologies (NIST) but later many experts have refined its characteristics in different ways (Aidan, 2013).

1. On-demand self-service

The user can easily and quickly configure the computing resources whenever they needed by themself.

2. Broad network access

The user can access the computing services from any devices like laptop, mobile, computer in the network connection. They can enjoy the feature simply from the web browser.

3. Resource pooling

The cloud service provider creates a pool of all the virtual processor, network resources and storage of the cloud physical resources and allocates securely.

4. Rapid elasticity

The user can increase or decrease the capacity of processor, storage or the network very quickly in minutes or in hours. When the request is trigger in the service provider it will automatically provide the service as per the request.

5. Measured service

The measured service means the resources the user is using are measured and reported back to them. The user does not have to pay the extra processing power they are using as they have to buy the services from server in serve basis.

2.3 The traditional IT Ecosystem vs. the Cloud Ecosystem

Today the business environment is being complex and more competitive. With this the expectation of customers are also increasing day by day. The companies are trying to improve and enhance the system through IT. So comparing the tradition IT ecosystem requires more investment in IT resources but not able to optimize the utilization of resources. Furthermore in traditional system the organizations not only have to setup the on-premise computing environment but also have to manage the IT teams themselves to manage the resources which ultimately increase the cost.

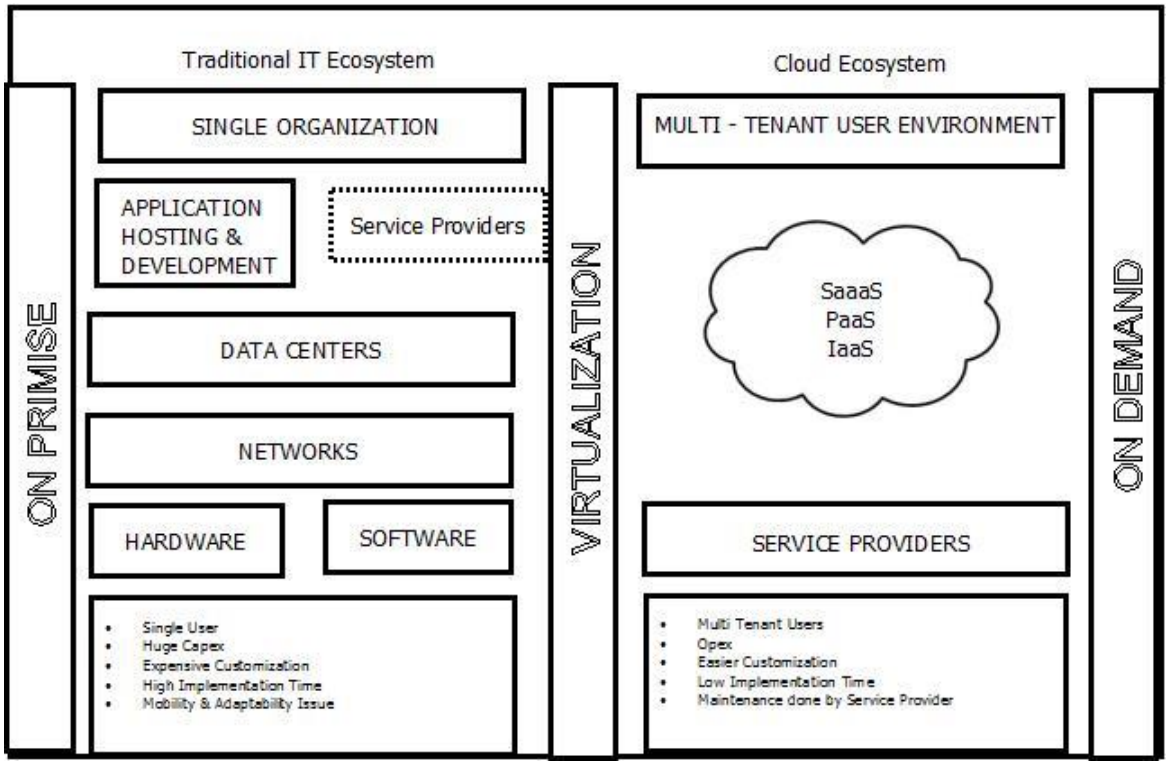

Figure 8: Tradition IT Ecosystem Vs Cloud Ecosystem

On the other hand the cloud computing signifies the complete transformation of IT setup. The computing process refers to the process of sharing of resources such as hardware, software and development platform over the Internet. The resources are mostly accessed based on pay per use or subscription basis. Virtualization is the first step for adopting the cloud environment. The cloud service are made available on the basis of virtualization and provided service to the user on the basis of used pricing model. The resources of the computing service are easily managed by the user without any support of the cloud server provider.

2.4 Advantage of Cloud Service for Business

The cloud computing provides services to the organization and enable the dynamic availability of IT resources regardless of any location. This enables the rapid delivery of the services and increases the throughput of any organization by customer retention, faster time to market expansion.

The development of the Internet and technology (Doug Thaler, 2013) has increased the productivity and performance in all types of business levels. The increased use of the Internet, improvement of data speeds with various kinds of electronic devices and bigger storage capacity has bring a new era in the technology of cloud computing. In turn, this technology has brought wide benefits and advantage in business.

The cloud is a platform that servers as a remote data center and provided service through the cloud provider. The members can access data easily through the Internet so it can be good for the individuals and the business organization who want to store and share the data and information to their colleague. So multi users can work together from different location which is good benefit of time and infrastructure. In facts, the network of cloud computing seems to be complex but everyone can use it without the stress. The person who uses the network do not have to understand the underlying complexity of the cloud computing instead they can simply log into the network and access the required files, work, share or perform any other task within a seconds.

In traditional method, the business owner need to create their own internal computing network which includes more cost and need to invest more time. Also they need to assign resources for regular maintenance of the server. The computing process helps to save all this process which is a milestone in the field of technology. Since a cloud computing provides an environment where the users can share their application, storage and networks where the people can share things and enjoy the feature with a minimum cost. The minimal effort that exists is the interaction with the cloud service provider. The ability that gives to store the data remotely gives wide verities of possibilities in the business. The members of business can access their files from laptop, mobiles or from any devices and carry on their work which makes easy to work from any places.

The cloud computing can also be used to store the offline copy of the important documents and also for the remote backup. This help in the case of data recovery and also for computer repairs. The cloud computing can be customized according to the needs of the company. As the cloud service provider gives the IT support so the business do not have to worry anything about the support related issues and can only focus on their task. This innovation helps the new generation of technology, products and services. Many business owners want to move in cloud because of the convenience feature provided by the computing service. It is good in terms of cost

perspective and also user friendly to share the documents and important files in the network. Now, analyzing all this benefits, seeking the assistance from cloud will be the great investment for development of business successfully.

2.5 Analysis of Network Traffic for Passive Measurement

Network traffic measurement is one of the most important means to establish accurate network models, to validate the new protocols and applications, to diagnosis the network failures, to enhance the network performance and quality of the service. In accordance with the traffic of the network, the network traffic can be divided into active and passive measurement. The passive measurement has no interference in the operation of the network. Moreover the measurement result reflects the network behavior and the measurement result more genuinely and correctly.

The main aim for solving the above address issues with cloud computing ecosystem on the basis of inducing the drawbacks of the traditional technical route, proposed the two different network analysis one of them is Analysis of Network traffic for passive measurement with a designed architecture of could pattern based network traffic platform. Here implemented architecture with a system prototype IP Trace Analysis System (Wang, Ding and Xia, p. 1-4) or in short called as IPTAS. Here, it describes and does the critical implementations of the system architecture to verify the system feasibility and the flexibility with an IPTAS application.

The elements of the passive measurement work contain traffic dataset, analysis algorithm, and measurement result. In addition to this also the storage and computing resource. The roles of the participants are divided into 1. Traffic collector, 2. Algorithm developer, 3. Measurement worker and 4. Computing Service Provider. In this pattern of analysis the traffic collector, algorithm developer and computing service provider makes the resources being scattered hence difficult to obtain and share. To overcome the drawbacks of the conventional pattern, here proposed a cloud pattern by deploying an intermediate cloud layer between measurement worker and the rest three roles as shown in Figure 9. The cloud layer centralizes the traffic datasets, reusable analysis algorithms, valuable measurement results, high-performance storage, computing infrastructure up-to-date.

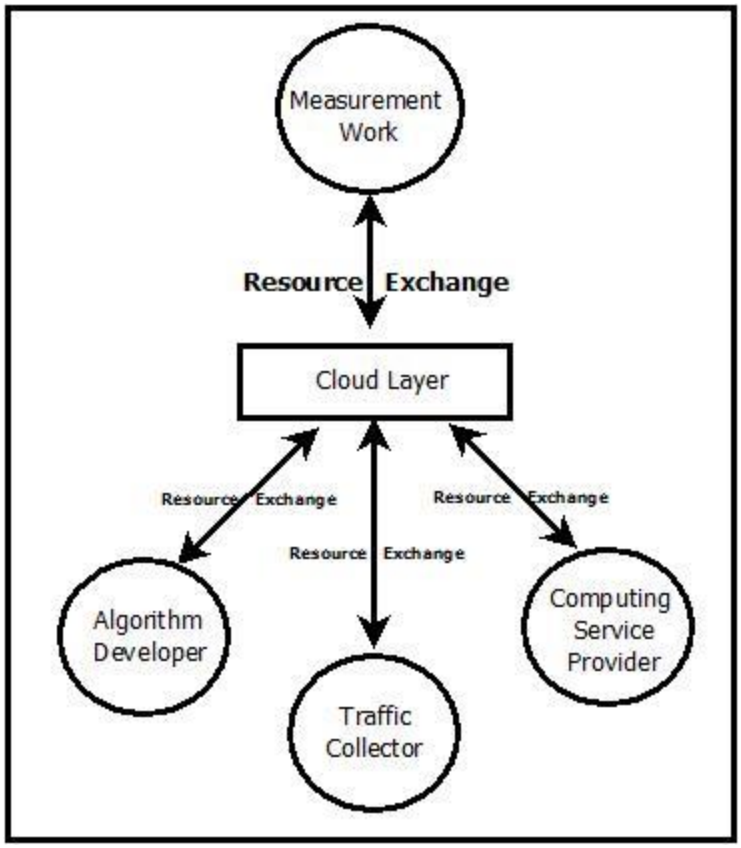

Figure 9: Cloud Pattern of passive measurement work

The cloud pattern of passive measurement work can be classified by the flowing features listed below:

1. Efficient access to massive traffic dataset
2. High efficiency analysis of traffic datasets
3. Reuse sharing of measurement achievement

The architecture of the cloud patterns based network traffic analysis platform consists of three layers. They are Resource Collection Layer, Resource Management Layer and Open Service Layer (shown in Figure 10). The three different layers have their own function and features.

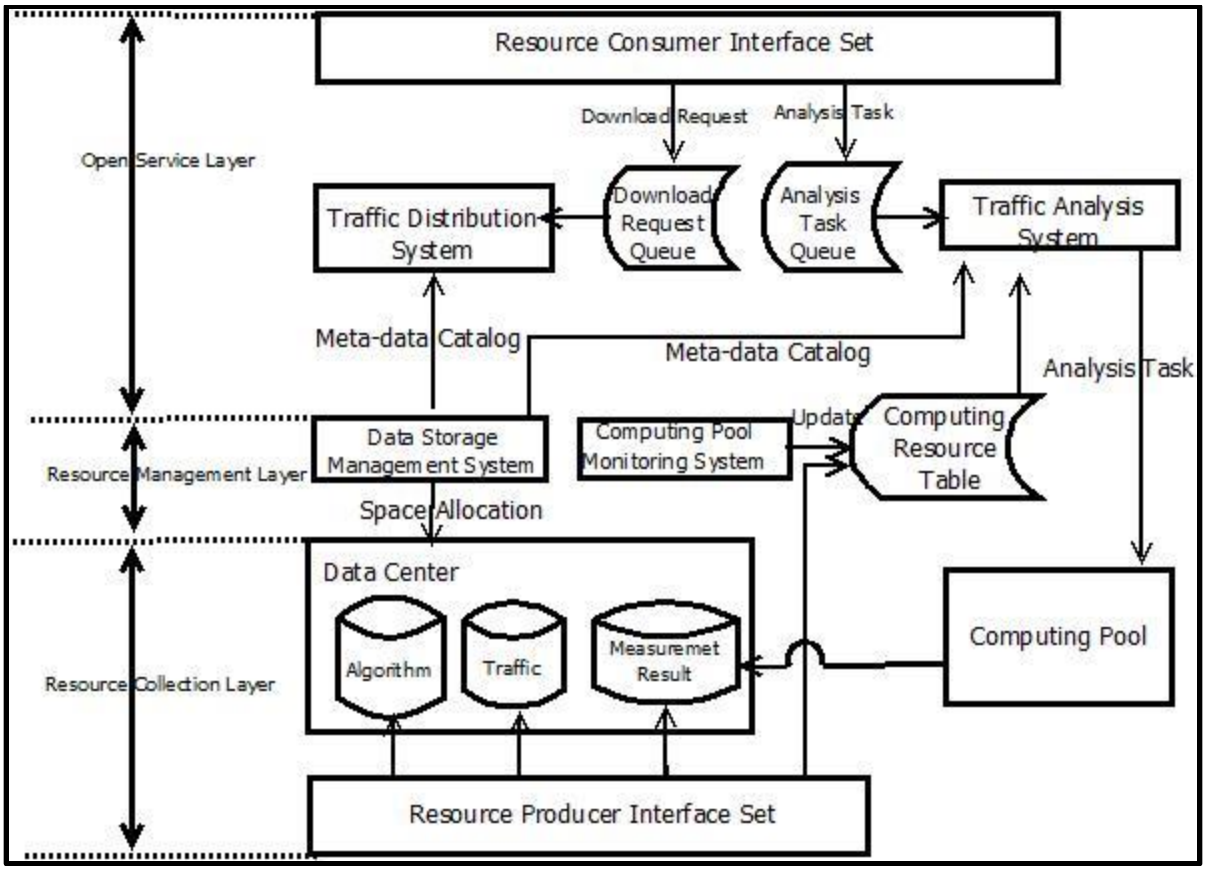

Figure 10: Architecture of cloud-pattern based network traffic analysis platform

The Resource Collection Layer which lies on the bottom of above architecture deals with algorithm, traffic, management result that are encapsulated in the data center of the layer. This layer basically provides traffic collector, algorithm developer and computing service that are correspondence to the resource producer interface set. The uploaded computing node creates computing pool to undertake traffic analysis computation. They register in a computing resource table so that the upper management layer can recognize the scaling of the computation pool.

The Resource Management Layer which lies in the middle of architecture takes care of computing pool monitoring and data storage management system. The computing pool monitoring system creates periodic heartbeat check and performance metrics collection for every note for pool. Then it updates the computing resource table which supports the upper layer to make the conclusion of the task assignment. In terms of data entry conduction and data description, the data storage management system acts as a centralized management system where the entry conduction focus on space allocation for each kind of data before entering the storage environment in order to improve the storage density.

The Open Service Layer which lies at the top of the architecture co-ordinates the traffic distribution service through the resource consumer interface set. The download request and analysis task is inserted into the download request queue and analysis task queue respectively. The traffic distribution system processes the download request in a queue. Then finally it notifies the user to set up a transmission based TCP or UDP connection.

The traffic analysis system is a centralized job scheduling and transferring system. The main role of the traffic analysis system is to schedule each waiting task in the task queue and transferring it to a suitable online node in the computing pool. When the node finishes the task, the measurement result is sent to the storage system whereas the task end information is returned to the system for updating the task queue and informing the user to download the measurement result.

2.6 Analysis of Network Traffic Measurement with MapReduce

This is an Internet flow analysis method of cloud computing platform. We present a MapReduce-based flow analysis method that can process huge files like Tera or Peta-byte collected from many monitoring servers. With the Hadoop data nodes, it achieved that flow statistics computation time for large flow files could intensely decrease when compared with the popular flow analysis tool that is run on the single host. Further it appears that the Map Reduced based flow analysis program finishes the work successfully if any single server stops working as well.

The map reduce program have been recently developed by Google, Yahoo, Amazon etc. in order to analyze the big data. But firstly the main aim to analysis the network terrific measurement with the Internet –scale flow analysis method with MapReduce.

Basically, the MapReduce (Lee, Kang and Son, p. 357-358) is a programming model for processing large amounts of data distributed over many machines. The map reduce programming model uses functional programming in two phases called map and reduce. Then the functional programming takes input as a set of key-value pairs and writes output a set of key-value pairs. It turns out that many data analytic problems can be formulated in the map reduce framework. Here the Hadoop is an open source implementation of Map Reduce. The Apache Hadroop (White, 3rd Edition) is a set of projects that are as linked for solving problems using the Map Reduce programming model. The Hadoop core project provides the Map Reduce implementation and Hadoop Distributed File System and in short called as HDFS. These both implementation can be used separately but typically Hadoop requires HDFS.

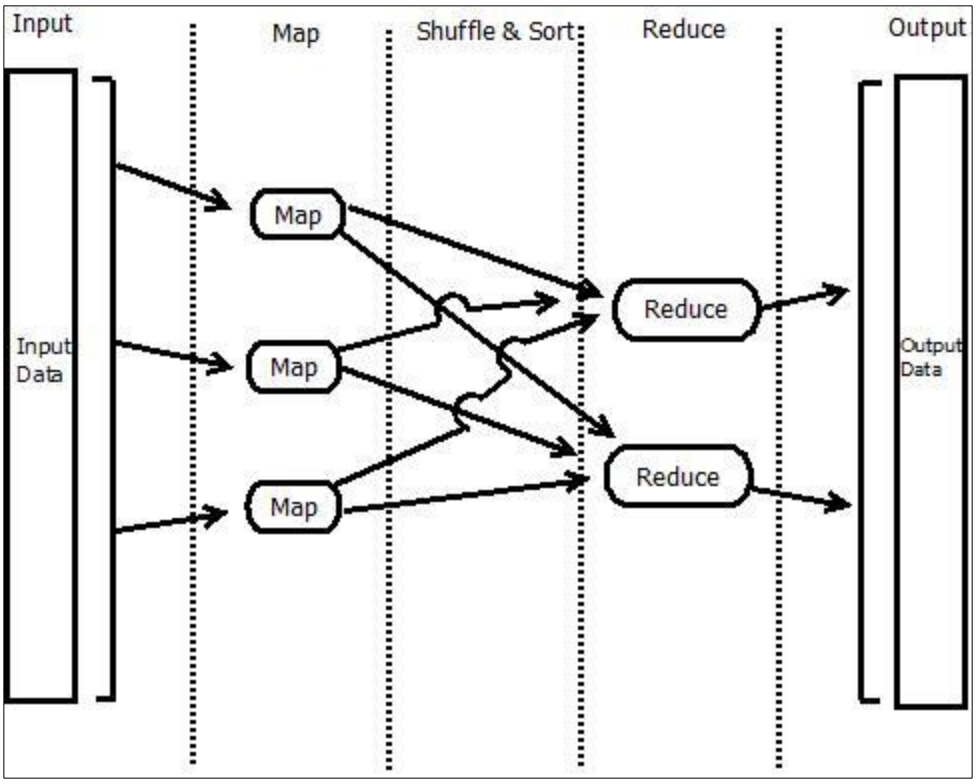

Figure 11: Map/Reduce Dataflow

2.6.1 Implementation of Mappers and Reducers in Hadoop programmatically

It is important that every map or reduce program must specify a mapper and typically a reducer. The mapper has a map method that transform input in the form of (key, value) pairs into any number of intermediate (Key', value') pairs. Similarly the reducer has a reduce method that transforms intermediate (key',value') that aggregates into any number of output in (key' ', value'') pairs.

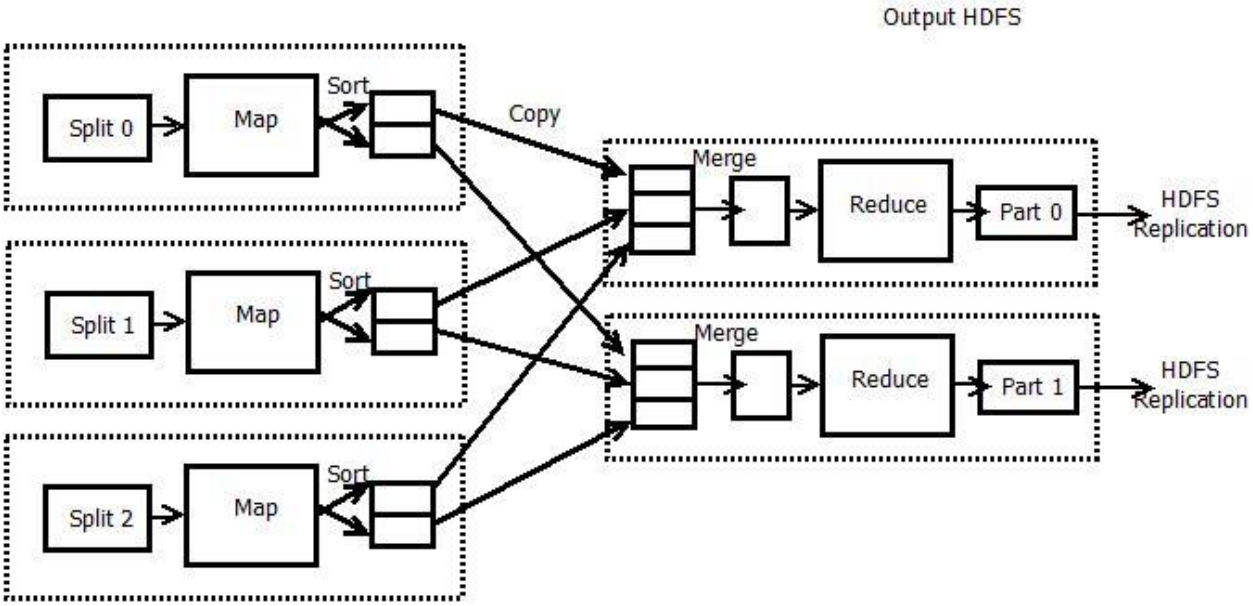

Figure 12: Hadoop Flow

In the Hadoop flow we have to provide a mapper and a reducer function to Hadoop and the number of instances of mapper is based on the number of input blocks that are provided. The number of instances of the reducer is specified by the person running the job. Sometimes no reducers are needed. It is observed that, one reducer will be bad if there are lots of data coming from mappers.

2.6.2 Flow Analysis based on Map Reduce

The architecture (Lee, Kang and Son, p. 357-358) of the flow measurement and analysis system is shown in Figure 13. The cloud platform provides the cluster file system and the cloud computing functions. The data sent from routers to the cluster through the means of unicasting or anycasting. In turn, the cluster nodes are operated by the master cluster node to save and process the flow data and to manage the cluster configuration. The MapReduce flow analysis program is run on the cloud platform when the flow data are archived on the cluster file system.

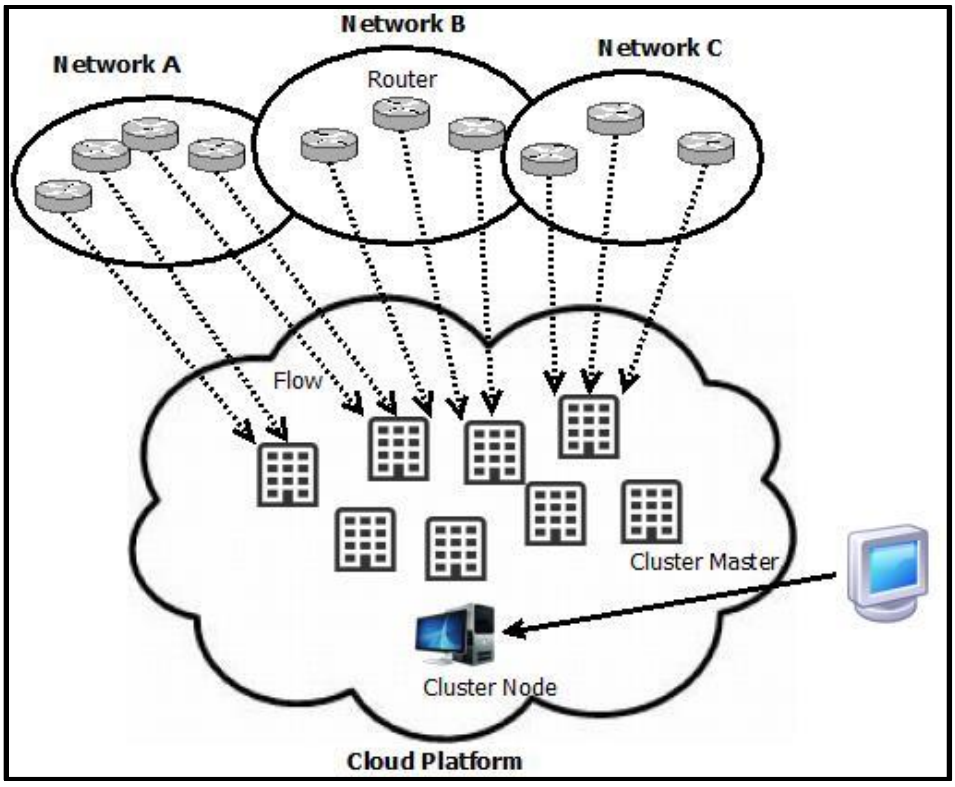

Figure 13: Architecture of the proposed flow measurement and analysis system

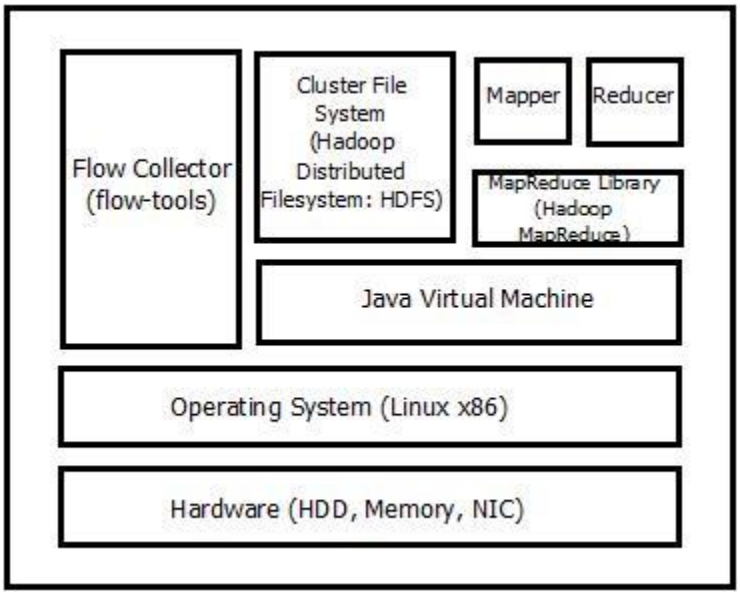

Figure 14: Functional components of a cluster node

The functional components of each cluster node is equipped with a flow collector or flow tools in the left, a distributed file system in the middle and MapReduce library in the right as shown in Figure 14. The flow collector receives the flow packets, store in the files and moves the files from local disk to the cluster file system. The packets of NetFlow coming from either routers or any monitoring servers are generally sent to cluster nodes in unicast. In NetFlow packets are sent in UDP which does not guarantee for reliability. Alternatively in Internet Protocol Flow Information Export (IPFIX), Stream Control Transmission Protocol (SCTP) is used instead of UDP for reliability. In the flow collector the flow tools are used for the NetFlow collecting and processing. Then the Mapper and Reducer will analyze the flow data with Hadoop MapReduce library. Here the distributed file system provides the easy management of very huge files and also acts as a fault tolerant service. For the cloud ecosystem platform, it is employed Hadoop that affords open MapReduce software framework and cluster file system in the Java Virtual Machine or in short as JVM. Specially, HDFS is suitable for handling huge files along with the streaming data access pattern. This pattern is called as write-once and ready-many times pattern.

2.7 Shortfalls of Existing Network Traffic Analysis Systems

1. Variation in Network Traffic Rate: The performance of cloud service depends on various factors among those network traffic is also one of them. The variation of network traffic may leads to the dissatisfaction over the services. There are thousands and thousands of users are requesting for the same services at a same time. It makes they have to wait for longer for their response. It leads to the problem and the users feel difficult to use the service.

2. Huge Data Traffic acquisition: There are heavy network traffic in cloud ecosystem and due to the limited local storage size, the limited download or upload bandwidth size, cost constraint and all other many factors it is difficult to download or upload the sufficient datasets to support the macro network model and to come in the conclusion of the analysis.

3. Huge Data Traffic Analysis: When there come to analyze the data let's say of tera bytes of data then definitely it requires high performance computing resource and also high speed I/O storage device to process the heavy traffic. There comes the problem of funds for managing the resources. It is different to carry out with individual and for research group as well. They have to invest extra funds or need to ask for computing service.

4. Information Share: In the process of network behavior analysis and studies by the research group they can find the valuable measurement results and algorithm from their long research and experiment. However due to the lack of resource and knowledge sharing mechanism among the research group or the other group, same work have to start from the beginning which eventually is the waste of time and money.

2.8 Why Cloud Network Traffic Analyzer?

The cloud provides various services to the cloud users. The major aim of cloud service is to provide the services instantly when the user request. Similarly the cloud user are motivated toward cloud services because they do not want unnecessary burden setting the infrastructure development which ultimately saves time and money. So, the demand of cloud service is being higher. Actually the performance of cloud service depends upon various factors but one of the major factor that needs to be focus is the consistency of the network traffic because the delay in responding the user request makes the customer dissatisfaction. The cloud service provider may provider uneven traffic rate or delay in response while handling or managing the peak demand of their request which may feel user difficult to use the service. So, to overcome such problems this Network Traffic Analyzer analyzes the traffic rate and provides the details information of flow of data to cloud and flow of data from cloud to the user computer. It also helping for sampling/scheduling the network traffic where the traffic is busy. It schedule on the basis of handling priority based traffic first and other consecutively. The other important feature of this tools is that it analyze the effect of Cloud Network Analyzer on various components of a device like Memory, CPU and shows the details information of Memory and CPU uses during traffic analyzing process. This tools helps for organization/co-operation to analyze the traffic rate from cloud or to cloud and can estimate the work plan accordingly.

It is a new concept of work from my analysis because not any system that analyzer the network traffic migrating data to cloud ecosystem and in the meantime it also does the traffic shaping work when there is heavy traffic in the network. It true that there are tools that will analyze the network traffic of the network, migrate data to the cloud ecosystem and some will do network sampling but I think this is the only tool which does network traffic analysis in data migration to cloud ecosystem and sampling of network traffic when the traffic is busy. I hope it provides the customer satisfaction over the cloud service provider. Obviously, the cloud service is being popular and it enhance more to attract the customer in the cloud service.

2.9 Available Network Traffic Analyzer and Simulation Tool

Today in this research era, it will be very costly to validate and verify the network protocol, specific network algorithm or to analyze network. In this scenario the network simulation saves lots of time and money to accomplish the goal. The network simulator is very important to analyze the network, to test new networking protocol or to study and research any network behavior.

Simulation is very important in this modern technology and it can be applied in different field like science, engineering and still more application fields with different objective. The computer simulation (Pan, 2008) help to model any hypothetical or the real objects on the computer, to

study and see how the system function. Moreover, the application of simulation into the networking field is the network traffic simulation.

More specifically, network simulation is a computer assisted simulation technology that applied a networking algorithm in the form of simulation. In network simulation we have given more focus for the network protocol or the algorithm rather than real time visibility features. It is used by people from different sector like academic researchers, industrial developers, Quality Assurance to analyze the performance of different networking protocols and its behavior. This is also used to study the effects of different protocol in network.

2.9.1 Network Simulation and Simulator

Generally, the network simulator models the real world networks. The main idea is that when the system is modeled, then the feature of the system can be changed it helps to analyze its corresponding output which will be our finding. This modeling of system is quite cheap compared to the real system which saves the cost and implementation time. Another important thing is that, it is not always feasible to experiment in the real scenario hence the simulation process will be a great milestone for that case. However, the network simulation will not be perfect always but when modeled correctly it help reaching the result close enough which gives the researchers or the tester the meaningful insight about how the system will affect its operation.

2.9.2 Available Network Simulator

There are different kinds of simulator (Pan, 2008) available based on their use and simplicity such as commercial or free Simulator. The simulator can also be categorized as simple or complex ones.

i. Commercial and open source simulators

Some of the simulators are commercial where they do not provide the source code for the user. The entire users have to pay to get the license and to use the simulator for their requirement. One of the examples of commercial simulator is OPNET. It has both advantage and disadvantage. The documentation is well organized so it is easy to understand whereas the open simulators are changed by many users and difficult to track as the documents are not well managed due to the lack of staff.

On the other hand open network source code is open so every organization can do their contribution and can fix the bug. The interface is also open for the improvement and the recent changes get reflected faster than the commercial simulator. Lack of complete documentation and

lack of version controls are some of the serious problems of open network simulator. Some of the examples of available open source simulators are NS2, NS2-Wiki and NS3.

	Network simulators name
Commercial	OPNET, QualNet
Open source	NS2, NS3, OMNeT++, SSFNet, J-Sim

Table 2.1: Network simulators

ii. Simple Vs Complex

Currently there are many varieties of network simulator available ranging from simple to complex ones. The network simulator enable user to know the network topology, specify the nodes of the network, links between the nodes, traffic between the nodes etc. Moreover in complex simulator system it allows the users to specify everything about the protocol that is used to process network traffic. In graphical application it helps the user easily visualize the working of the simulated environment whereas some are in text-based form which provides the less visual user interface. Some of the simulator are programming oriented that provides the framework where the user can create the application and simulates the network environment for testing.

2.10 Existing Network Simulator

Here discussed some of the network simulators some of them are of commercial purpose and rest are the open network simulator.

2.10.1 OPNET

OPNET's is basically specialized for network research and development. It is flexibly used for communication networks study, about devices, protocol and the applications. As this is commercial service provider it has a good graphical interface for user and the graphical interface is used to build the application of network topology and entities from the system application layer to the physical layer. Here, the object oriented programming language is used to create a mapping from the graphical interface for the real implementation. The image below shows the graphical representation of each network nodes and the graphical output. As it has a graphical view, the parameter can be changed and viewed the result repeatedly very easily without much

effort. This simulator is popular for industry and the network research for the development. The given GUI interface and programming tools are very useful to build the system in accordance of user requirement and to simulate the system.

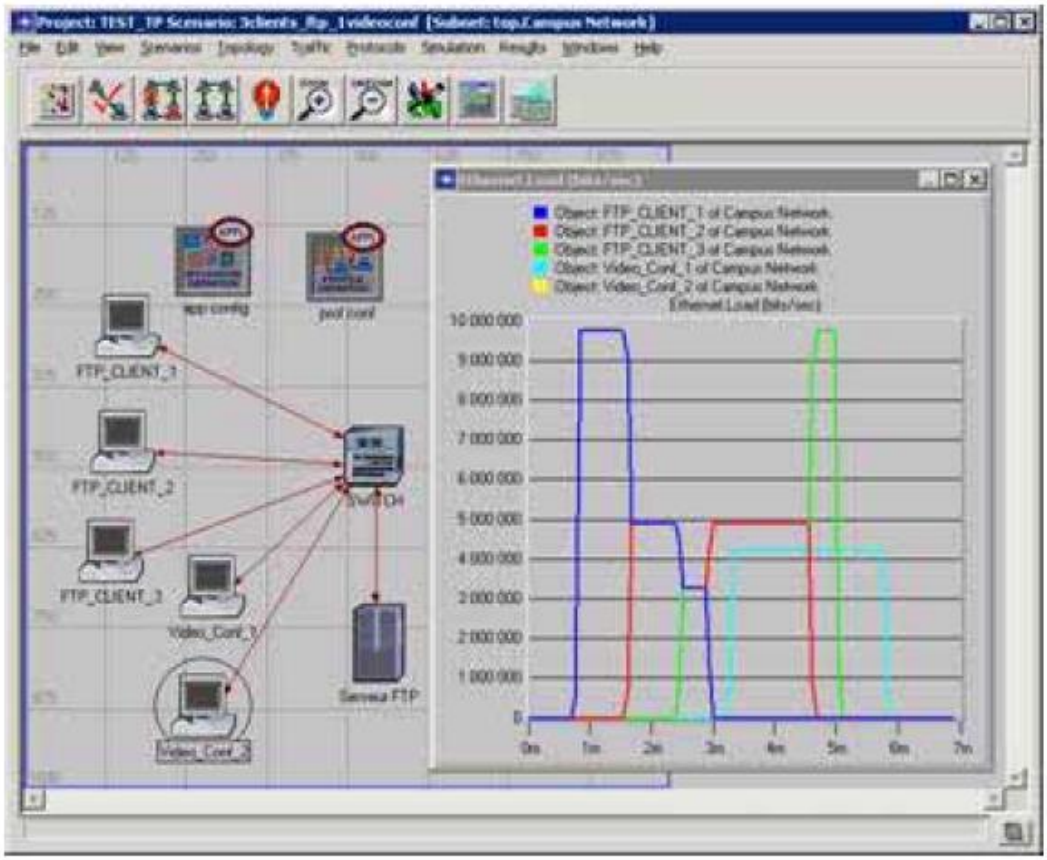

Figure 15: OPNET GUI

Main Features of OPNET

OPNET has three main functions as modeling, simulation and analysis. For modeling it provides nice graphical interface to define and create all kinds of model protocol. For simulation it used different form of advanced simulation technology to cover and address wide range of study purpose. For analysis the simulation results and data can be displayed graphically in user friendly forms of charts, graphs and in statistics form for user convenience. Moreover it is simpler to analyze the results, it's because the results are displayed in the graphical form. It is easy to

compare and analyze the results. Furthermore, the animation can be generated for the user convenience.

2.10.2 Network Simulator 2 (NS2)

NS2 is the most popular network simulators. This is a discrete event simulator mainly targeted for the network researchers. NS2 is the second version of NS (Network Simulator) and NS was developed in 1989. The current version of NS2 is widely used for academic research. Later lots of packages are contributed by many nonprofit groups to improve and make it much better.

Main features of Network Simulator 2 (NS2)

Network Simulator or in short NS2 is an object oriented discrete event driven network simulator. It was first developed at the University of California-Berkely. The programming language used is C++ and Tcl Script language with object-oriented extension (OTcl). There is reason using these two languages. C++ is very efficient to design but difficult for graphical and visual implementation. OTcl is used to fill the lap that the C++ lacks. So the combination of these two languages seems to be very effective. Basically the C++ is used to implement the detail protocol of simulation and OTcl is used for the user to control the simulation and schedule the events.

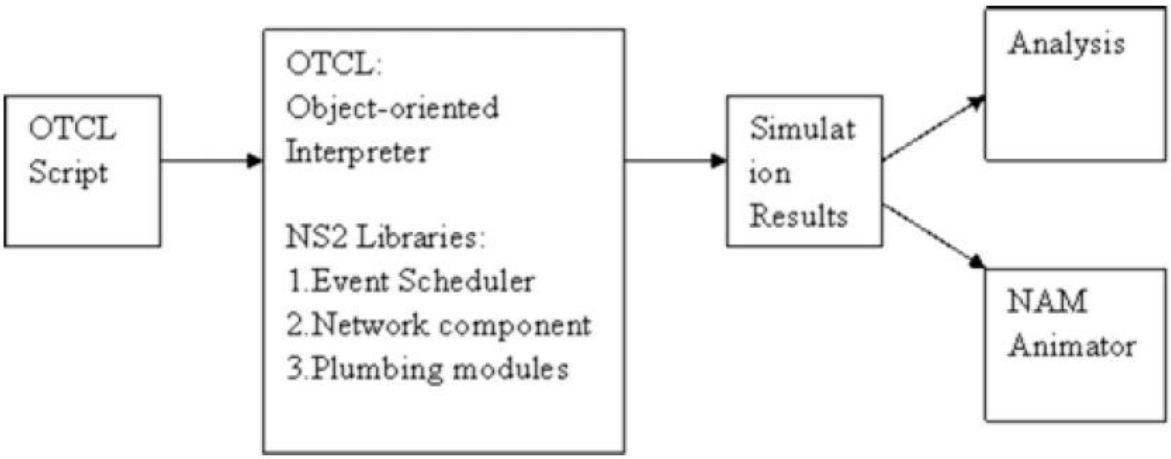

Figure 16: NS2

The OTcl script is used to initiate the event scheduler, to setup the network topology and to tell traffic source whether to send or stop sending the packet from event scheduler. The scene can

easily modify through the OTcl script. There is flexibility that when a user want a new network object they can simply write the code using the existing object library and also the plumb the data path from object. The important plumbing makes NS2 very powerful. Another important feature is that, the event scheduler which keeps tracks of simulation time. It plays a role to release the event in the event queue invoking an appropriate network components.

2.10.3 Network Simulator 3 (NS3)

Network Simulator 3 (NS3) is an open source discrete event network simulator which is mainly targeted for the research and education purpose. NS3 is not an updated version of NS2 and it is a tracing and statistics gathering tool. It is a new simulator so it is not backward compatible with NS2.

Main Feature of Network Simulator (NS3)

The main feature of Network Simulator (NS3) includes different software core which means NS3 is written in C++ and with python scripting interface. Here the advanced design pattern is used. The protocol entities are designed in such a way that it is closer to the real computers. It helps incorporation of other open source networking simulator so it helps to reduce the rewriting the module of simulation. Here for support of virtualization, the light weight virtual machine are used.

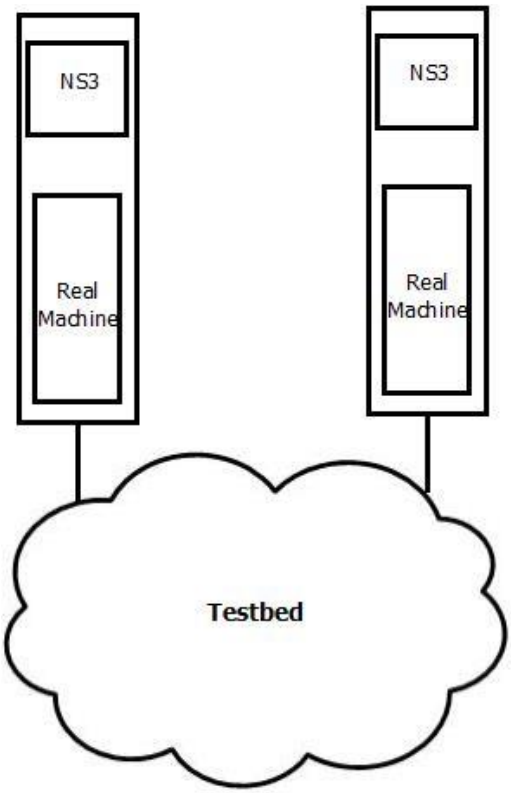

Figure 17: Testbeds interconnect NS3 stacks

2.10.4 OMNeT++

OMNeT++ is a public source, a discrete event simulator with GUI support of component based network simulator. The primary application area of this simulator is the communication networks along with its flexible architecture it has other areas like IT systems, hardware architecture, queuing network and also in business process. Here the components are called as modules and programmed in C++ language. Its working principal is similar to that of OTcl in NS2 and Python in NS3. The smaller components are assembled into larger components and models using high level language.

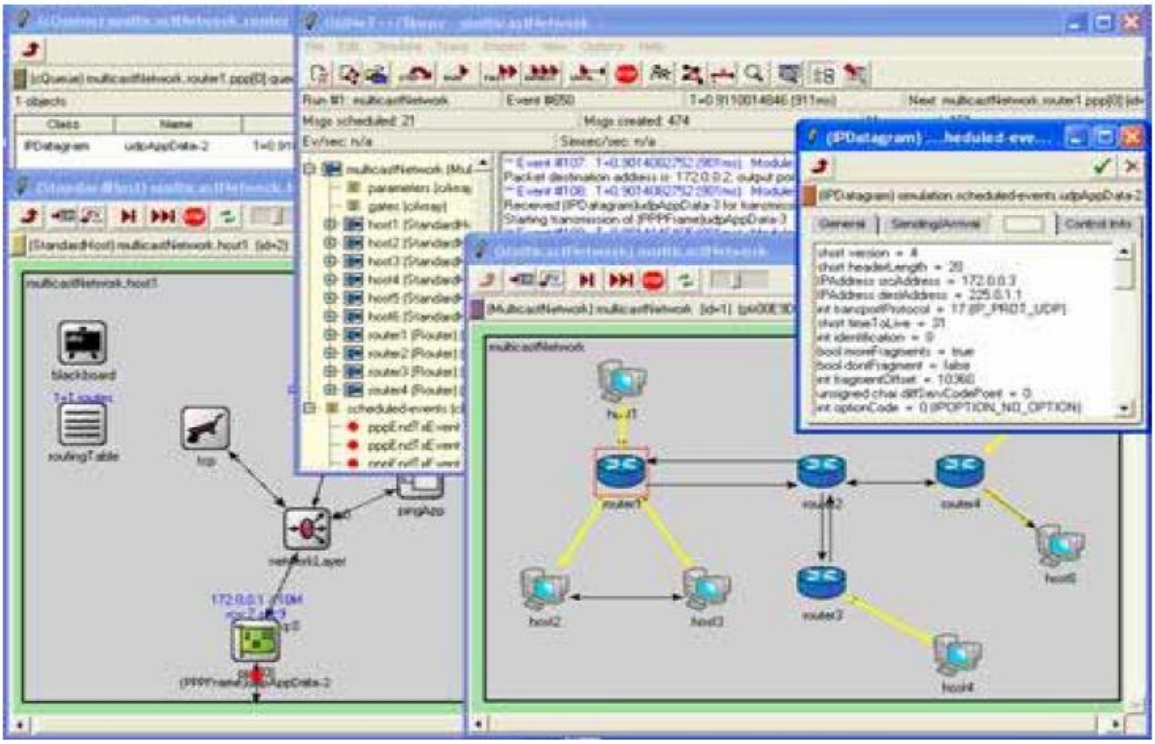

Figure 18: OMNeT++ GUI

Main features of OMNeT++

The OMNeT++ is designed especially for the complex based architecture. Generally the reusable components are assembled to form OMNeT++ module. The major features of OMNeT++ are the modules are reusable and the modules are combined in a various ways. The key feature is the simulation kernel C++ class library which consists of simulation kernel and utility class necessary for simulation components. It has runtime user interface or environments for simulation. OMNeT++ support multiple platform like it can run on Linux, other Unix-like systems and on Windows systems.

3 Cloud Network Traffic Analyzer and Simulator (CNETAS)

The Cloud Network Traffic Analyzer and Simulator is a simulation tool for analyzing cloud network traffic. It analyzes data while uploading data to cloud or downloading data from cloud. It provides details network analysis report to users which helps users to schedule there work. On the other hand it also schedule the network task to process the priority task earlier and to manage the task subsequently. It also checks the performance of devices such as Memory and CPU uses while running the simulation tool.

3.1 Data Transmission from User to Cloud and Vice versa in CNETAS

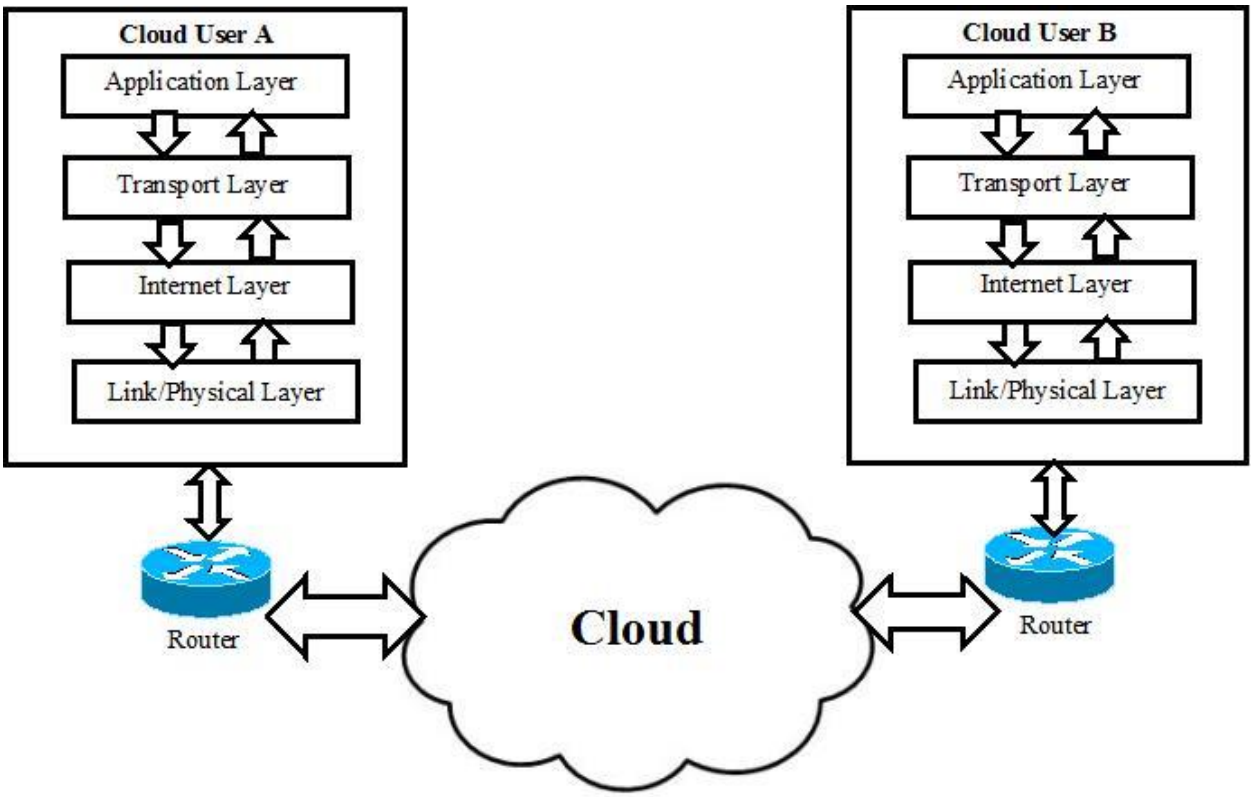

Figure 19: Data Transmission from User to Cloud and Vice Versa in CNETAS

3.1.1 Data Packets

Everything we do on the Internet involves the flow of Packets. When we upload data to cloud or download data from the cloud there is a series of flow of packets in the network. The huge data is broken down into parts of a certain size in bytes. These are the packets. Each data packets carries the information of its destination, the sender's IP address. The packets also have the information about in how many number of packets the data is broken into and the number of the particular data packet. Each packets contains the body of the message and follows the internet protocol i.e. Transmission Control Protocol/Internet Protocol (TCP/IP). Then the packet is delivered to the destination with the best available route. Depending on the type of networks the packets is termed as frame, block, cell or segment.

3.1.2 Routers

Router establish the links to connect computer to the Internet so the users can share the connection. It simply acts as a bridge between the private network to connect with the Internet and choose the best path to travel the information. So here the router connects the computer to cloud and to analyze the network traffic of cloud ecosystem. As the router works on IP address it operate in network layer.

3.1.3 Flow of Data through Different Layers

The data that we send need to convert into the electronic signal to flow in the network which is separately handled by the communication protocol. This protocol communicate between two or more layers. These layers are named as Application Layer, Transport Layer, Internet Layer and Link/Physical Layer respectively. The data we send from the Application Layer goes through these layers and broken down into small chunks called as Packets. Here, in our application the data we uploaded from the Cloud Traffic Analyzer and Simulator goes to the Transport Layer. The Transport Layer used the Transmission Control Protocol to encapsulate the data from Application Layer. Then the data transfer to the Internet Layer where the Internet Protocol is used to deliver the packets. Finally from the computer hardware the data flows to the cloud environment. Similarly the data flow from cloud to the computer in the same route from bottom Link/Physical Layer to the Application Layer.

3.2 Cloud Environment

In our architecture for the cloud environment, here chosen SharePoint Online to analyze the network traffic while moving data to cloud and from the cloud environment.

3.2.1 What is SharePoint?

SharePoint is the business collaboration platform for the enterprise. The demand of SharePoint is increasing due to its capabilities and wide feature available in the package. It has all the features that is essential for any organization to easily create and manage their contents. It helps to share the information across boundaries and enabled the better informed decisions. It has wide varieties of features that has attracted many organization and people. Initially the SharePoint has to setup inside the organization or any firm by themself maintaining the SharePoint server and its client computers. Now the SharePoint has upgraded on-premise to cloud environment which has attracted many customers because it is easily accessible regardless the cost and time of setting the SharePoint server inside the firm.

3.2.2 Uses of SharePoint in Organization

There are essential features that motivates the organization towards the SharePoint. In SharePoint (Dux Raymond Sy, 2009) the individual or the groups can have a collaborative web site and different access level can be maintained between the team members. The useful and relevant information for the organization can be centrally stored and access setting can be maintain in such a way that, it is available only for the concerned team members. It helps for maintaining the streamlined communication. The detail technical knowledge is not essential to use a SharePoint system. The basic knowledge of IT can use the system as it is based on familiar tools and technologies like Web, Windows, Microsoft Office etc.

3.2.3 Essential features of SharePoint

1. Easy to create a collaborative site collection

The site collection can be created easily with a minimal technical skill like the knowledge of Microsoft Windows, Microsoft Office, familiar with Web browsing etc. It is also easy to define the relevant access based on the communication needs. If the site are deployed appropriately then the IT does not have to deal with the contents updating, maintaining the document repository and defining the different account privileges.

2. Manage the information efficiently

The SharePoint provides wide varieties of features to effectively manage the information and documents. Some of them are:

- **Information Storage**

There is a wide verities of information storage facilities. The different list and libraries are available to store the information. Depending upon the type and nature of the content the information can be stored in different available list and libraries.

- **Check-in/Check-out**

There is a check-in/check-out feature to control the documents to edit from the selected users only. This options help the document can be modified single people at a time which helps to remove the chances of redundancy.

- **Version Control**

There is an option to control the version maintaining major version only or with major and minor version both. This option can be set according to the requirement how the version should be maintained. Every time the documents are edited the version will be upgraded. If minor version is set then unless and until the content is not approved it will not create major version. In this way the version can be tracked very easily.

- **Content approval**

This option is very important for the document management. The document status like draft, pending and approved are useful to denote the status of the documents. Unless the documents are not in approved status it can be assumed that is in processing stage and can work for the documents to get finished. One person can work on document and another person may have privilege to approval the documents hence the document correction and management can be done properly and in a managed way.

3. Makes easier for Team Collaboration

There is available a site template called document workspace which can be used jointly develop the requirement document, reports etc. Moreover the tools such as wiki which helps to document the learned lessons. There is a discussion boards which can be used for the offline communication to discuss about certain topic and can draw the conclusion. There is blog to post the blog article and post comments, and reply in it. To support meeting there is a meeting workspace which ultimately helps for team collaboration.

4. Enhance Communication

The SharePoint helps to deliver the right information to the right person in the right time. There is task list where the task can be schedule to notify the concerned user for the progress. Then the concerned members can check the task progress and can take necessary action to accomplish the task successfully. This helps to provide the task details on time. The access privilege can be set based on the information needs in different group level so that it helps maintaining the privacy level high as required in different content level.

5. Helps to Automate the Business Process

The important feature of SharePoint workflow helps to automate the process. The workflow of task can be designed according to the requirement of task to be executed. So, if the workflow is set it get executed automatically when the condition matches. This helps to automate the process which eventually helps to simply the task load in some extend. The custom workflows can be created which is very helpful for the processes. It helps to give the notification of different task and aware about the task deadlines and other important information .It also helps to execute the task sequentially and planned way which will be very helpful to accomplish the goal of the business organization.

6. Generate Reports

The SharePoint can be used to generate different useful reports like the summary of the task information, summary of the project and different automated alerts. It also help to generate different charts and use to trace the status of the task using different Key Performance Indicators (KPI) list. The report is very importance to get the overall progress of the organization and SharePoint provides different ways and easy process to generate the reports so it is increasing its demand in today's market.

3.2.4 Sharepoint Online

In simple the SharePoint (Microsoft, 2015) can be refer to one or more SharePoint products or technologies like SharePoint Online, SharePoint Foundation, SharePoint Server, SharePoint Designer, OneDrive for Business folder Sync. Among them here in our architecture used the SharePoint Online for cloud environment.

SharePoint Online is the cloud-based service hosted by Microsoft for the business for all sizes organization. It is easy to use as it is cloud based service instead of installing and deploying SharePoint Server on-premises any business can subscribe to an Office 365 plan. The SharePoint online in other words also called as Office 365. For our work we have created trial account and try migrating to SharePoint cloud environment and analyzing the network traffic.

3.3 Components Diagram of CNETAS

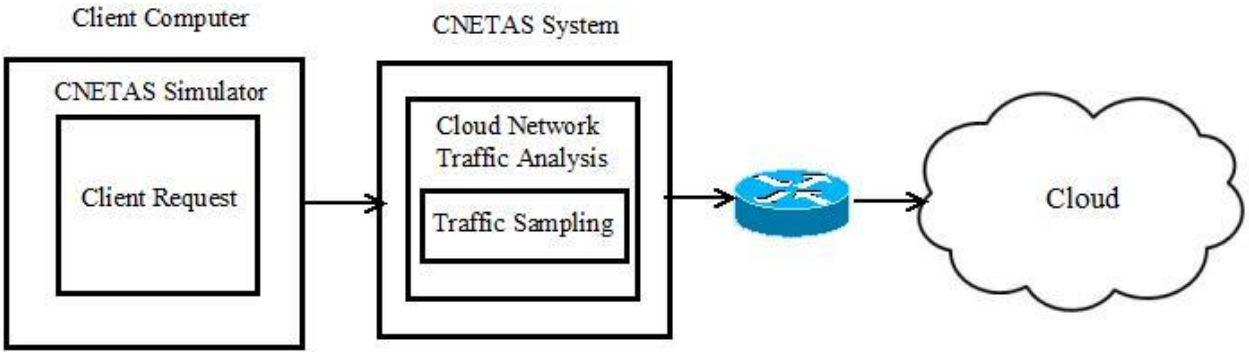

Figure 20: Components Diagram of CNETAS System

The component diagram of CNETAS have mainly three components. First the CNETAS Simulator which is running in the client computer to track the network process. Second is the CNETAS system which has two parts Cloud Network Traffic Analysis section and Traffic Sampling section. The traffic analysis is primarily handled by this section unless there is heavy flow in network. Once the network start busy then the Traffic Sampling section will schedule the priority based task first and proceed for analysis with highest priority. Third is the cloud system itself where the data is travelling from client computer to the cloud environment and from the cloud environment to the client computer.

3.4 Common Deployment Framework

The Figure 21 below describes the common deployment framework model of the cloud computing. Here, the cloud user request the cloud services using the application programming interface. The common deployment layer has the feature which helps for the efficient communication between the user and the cloud service provider. The Cloud Network Traffic Analyzer process the request that is send by the cloud user to the cloud server. The traffic sampling is mainly used to shape the traffic whenever there is fluctuation occurs in the network traffic which makes busy in the cloud communication. The traffic sampling main purpose is to provide the correct network path to the valuable users. When any organization is accessing cloud services for a longer time than of course the organization should have higher priority getting services rather than the normal user. The user dissatisfaction may increase if they do not get the cloud services on time because of busy network or of low bandwidth. The good solution for this problem might be giving constant range for all the customers.

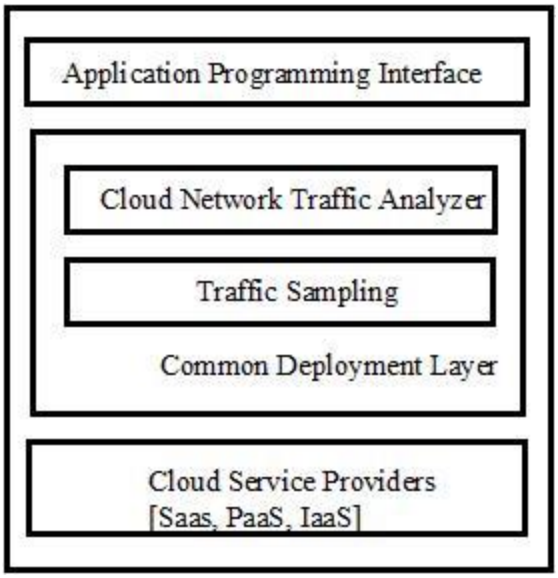

Figure 21: CNETAS Common Deployment Framework

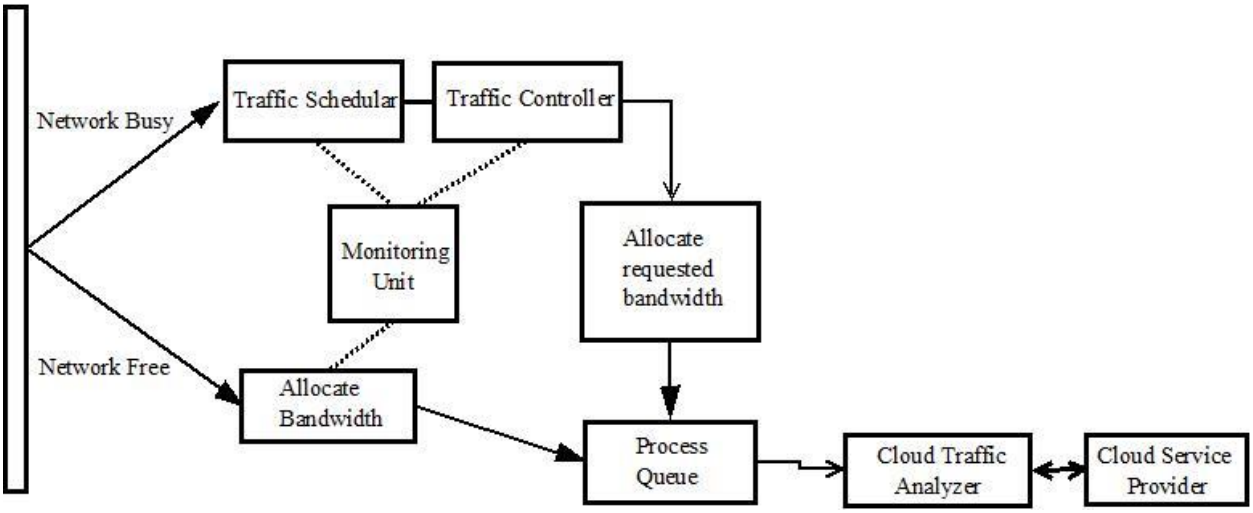

Figure 22: Traffic Sampling Implementation

Figure: 22 describes traffic flow of the Common Deployment Model. The user requests are sent to the traffic scheduler if there is busy in the network and traffic need to schedule. First the traffic scheduler need to identify causes creating a network busy and should handle respectively. The traffic controller check the network traffic during the transfer. If the traffic level exceeds

then the proper notification can be provided to make the user aware. The traffic controller control entire traffic and keep track of current processing task and the bandwidth using by the task. If the requested bandwidth is available it is directly send for analysis else the traffic controller and scheduler together find the correct route. The monitoring unit has to monitor and regular all the activities to run smoothly. When the task is allocated the requested bandwidth the process is analyzed by the Cloud Network Traffic Analyzer communicating with the cloud service provider.

3.5 CNETAS Architecture

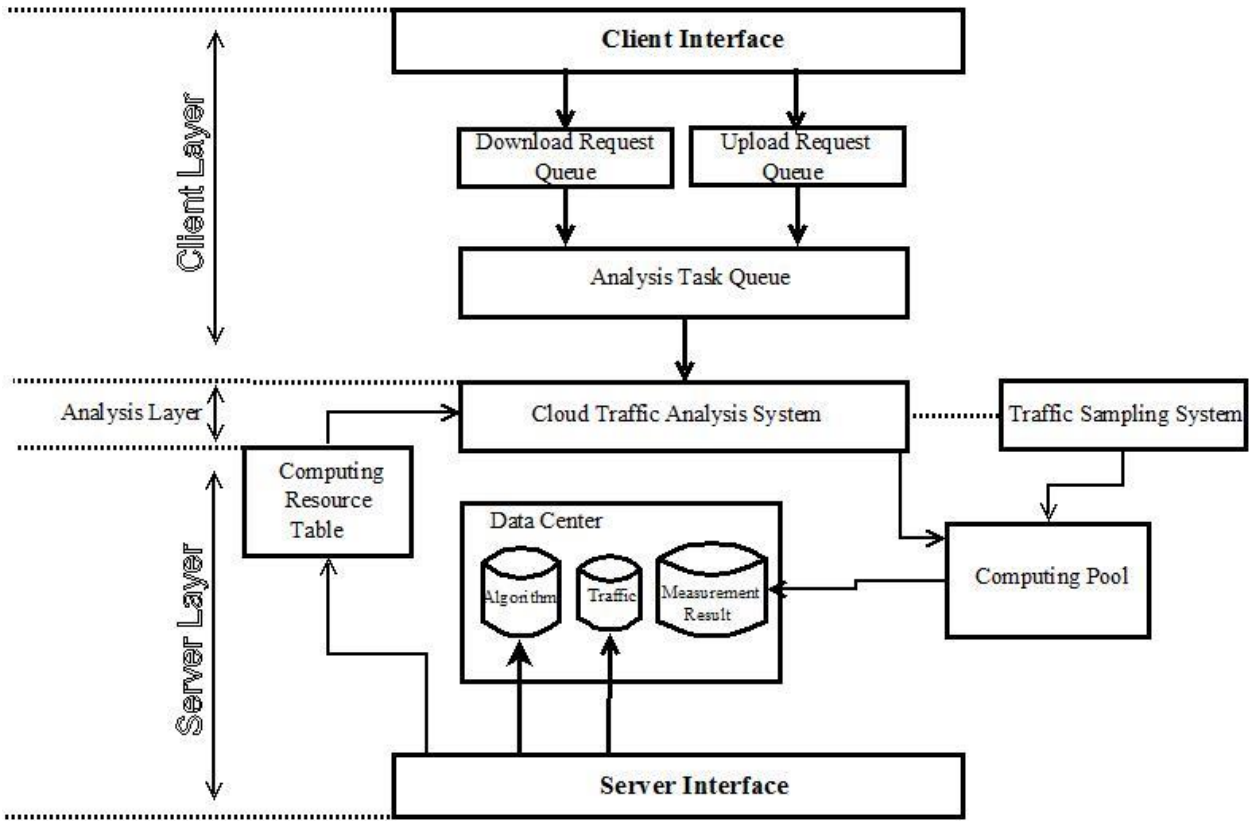

Figure 23: CNETAS Architecture

Here, we designed a practical architecture of Network Traffic Analyzer and Simulator for cloud ecosystem to analyze the traffic and improve the network traffic in cloud environment.

The overall architecture is divided into three layers: Client Layer, Analysis Layer and the Server Layer from top to bottom as shown in Figure 23. The function and composition of each layer is described below.

1. Client Layer

The client layer is also the interface or front end to the cloud user. The user send the task to cloud environment from the client Interface. On the basis of the task to be executed the upload request queue or the download request queue is formed. The both types of request are further analyzed by the analysis task queue and proceed forward for the processing. The priority of the task can be set by the user if the task is to be proceed earlier regardless of the other pending task in the queue.

2. Analysis Layer

This is the main body of the system where the analysis of the traffic is performed. The details information of data packet to and fro from the client system to cloud environment is analyzed. The download and upload task queue is handled sequentially for cloud processing and analysis unless there is busy network. In result, it provides the summary report for quick view and also details information of the packet flow track of every second from the start of task to the end of the task is provided. The analysis system also analyze the performance of the system while traffic analysis process is carried out. It track the information of load in memory and CPU which would be helpful for further analysis and estimation for the work. There is another section called traffic sampling system which has a role when there are number of request on the queue and in the meantime there is a request to handle with high priority. This block finds the requested bandwidth and schedule to provide the service faster for the best performance result. This helps providing the smooth service to the cloud user and helps to get the services faster and easier.

3. Server Layer

The server layer is the collection of cloud algorithm developer, traffic and the measurement work and computing pool for uploading the data in the server called the server Interface. The uploaded traffic dataset, analysis algorithm and measurement result are stored in a data center. The computing pool is used for cloud environment for execution services of the cloud environment. The computing resource tables have the information of the task that is in the queue which need to be handled and the task that is completed. So it handle the task sequentially unless the priority based task encounter in the system to process with the highest priority.

3.6 CNETAS Flowchart

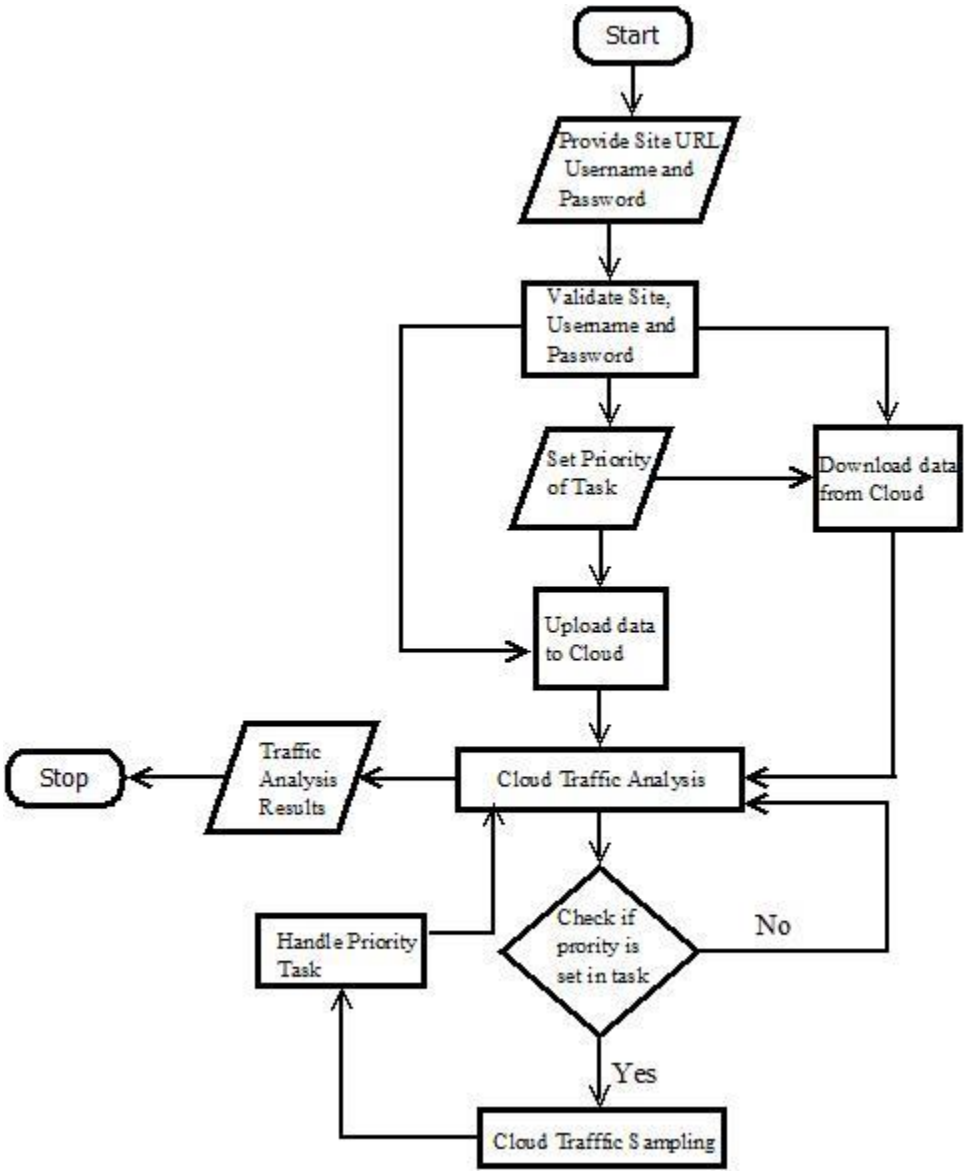

Figure 24: CNETAS Flowchart

The figure 24 shows the flow diagram of CNETAS processes. The process started submitting the site URL, username and password as the user input to communicate with the cloud service. Other extra input will be read from the setting file like the library name of the site where the documents will be uploaded or downloaded from the cloud services. Once the credentials has been passed to the CNETAS it will sent for the validation process. For the security purpose to make the

password more secure the password encryption and decryption technique can be used. This is useful when the credentials has been provided from the xml setting files instead from the input form. The both technique can be implemented considering the user simplicity. The priority can be set or it can't be set, as it is not the mandatory option. When the task need to be executed faster and the network is busy then this option is very important. Then download or upload operation is sent to the cloud traffic analysis block for the processing.

1. Cloud Traffic Analysis

Cloud Traffic Analysis in short can be called as CTA is to analyze the each packet flow in every instance of time interval during the upload or download operation. It keeps track of every packet flow to the cloud and from the cloud environment. With this network data collected from CTA other required information can be calculated like the speed of the packet received, sent, the size of data packet flow in upload or download, the content size, the total time taken for the whole task to be completed etc.

2. Cloud Traffic Sampling

Cloud Traffic Sampling in short can be called as CTS is the sampling block when the priority based task is to be handled earlier. This is for scheduling of the task queue in the busy network. First the priority checker block checks whether the priority has been set or not if it is set then the CTS blocks sends the task for processing and for traffic analysis with the highest priority.

With the references from the CTA data the various calculation can be done like

- **Cloud Service Cost estimation using the formula,**

  CSC = NRU * CSR + (PHS - PHE) / (PSR-OPSR)

  Whereas,
  CSC = Cloud Service Cost
  NRU = Number of Request per User
  CSR = Content Size per Request
  PHS = Peak Hour Start
  PHE = Peak Hour End
  PSR = Peak Service Rate
  OPSR = Off Peak Service Rate

- **Cloud Error Rate can be compared using the formula,**

  CER = (NResquest- NResponse) * 100

  Whereas,
  CER = Cloud Error Rate

NResquest = Number of Request
NResponse = Number of Response

- **The Cloud Response Rate can also be compared using the formula,**

  CRR = (CRB / 1024) * 1000

  Whereas,
  CRR = Cloud Response Rate
  CRB = Cloud Response Bandwidth

3. Traffic Analysis Result

The summary of Traffic Analysis Results can be provided in the graphical form. In addition to this the detail analysis report is provided in log format which has information of bytes of data packet flow to and fro from the cloud service provider. Also it gives the report of performance of CPU and memory in the time of simulation process. This reports helps for educational purpose, research purpose and also the migration purpose which gives the clear picture of data flow from cloud service to cloud user and vice versa.

3.7 CNETAS Algorithm for Cloud Network Traffic Analysis

3.7.1 Cloud Network Traffic Analysis in Upload/Download operation

1. Read Site URL, Username and Password from user. Also read the SharePoint Library name where the documents are going to uploaded. If it is a download operation provide the SharePoint Library name and documents to be downloaded.
   1.1 Here the Site URL, Username and Password read from user input but SharePoint Library name from reading XML file using XDocument class. The Library name can be defined as an attribute in the xml file.
2. Upload the documents using OpenFileDialog class.
3. Validate the Credentials using SharePointCredentials(string username, SecureString password)
4. If the password need to be handled securely then validate the Credentials using Cryptography class to encrypt and decrypt the password using the System. Security. Cryptography.SymmetricAlgorithm class. This is essential when the credentials are read from setting files and want the password to keep secure from others.
5. ElapsedEventHandler() function of Timer namespace track the network status in every 1 second. In the background thread the operation is repeated in every 1 second from the

start of Upload/Download operation till the end to track the packet flow to and from the network.

6. Start the Uplod/Download process.
7. Record the start time of upload.
8. Check the performance of CPU and Memory when the upload/download process has been started.
9. Get the SharePoint web URL and load using Microsoft.SharePoint.Client.ClientContext class.
10. Get the SharePoint List collection.
11. Read the SharePoint List properties.
12. Compare the Library name provided in XML file from the user with the SharePoint List that is read from SharePoint site collection.
13. When the target library is identified get the target server relative URL.
14. By combining server relative URL and file to upload get the final target location to upload the document or download the document.
15. Upload the documents to SharePoint library i.e the Cloud Library using Microsoft.SharePoint.Client.File.SaveBinaryDirect(context, targetLocation, fs, true).
16. When the documents are going to download from the SharePoint using Microsoft.SharePoint.Client.File.OpenBinaryDirect(context, sourceUrl) and store in FileInformation. Then reading every byte write the file in a document and save in the destination path provided by the user.
17. While executing from step 5 to 16 with the ElapsedEventHander() function track the each packets flow during the process calling function to calculate the bytes flow.
18. Record the task (upload/download) end time.
19. Calculate the file size and calculate the speed.
20. Also record the performance of RAM and CPU.
21. Provide the traffic analysis record as file uploaded, time stated to upload, time end to upload, total time to upload, content size etc.
22. If the operation is download record as file downloaded, time start to download, time end to download, total time to download, content size etc.
23. For more details, the log is written for packet flow in every 1 second from the network and to the network from cloud.
24. Record the byte sent during the upload/download operation from network.
25. Record the byte received during the upload/download operation from network.
26. Calculate the upload speed and download speed using parameter of byte received and byte sent.
27. Record the network speed during the upload/download operation.
28. The performance log report and network analysis of cloud traffic is saved in the log folder of the project working directory with the name of the file downloaded or uploaded and adding the current date value.

3.7.2 Traffic Sampling in Upload/Download Operation

1. Read Site URL, Username and Password from user. Also read the SharePoint Library name where the documents are going to uploaded/downloaded. If it is a download process provide the SharePoint Library name and documents to be downloaded.
   1.1 Here the Site URL, Username and Password read from user input but SharePoint Library from reading XML file using XDocument class. The Library name can be defined as attribute in the xml file.
2. Upload the documents using OpenFileDialog class.
3. Set the priority to proceed fast operation. Let say it as a task TP1.
4. Validate the Credentials using SharePointCredentials(string username, SecureString password)
5. If the password need to be handled securely then validate the Credentials using Cryptography class to encrypt and decrypt the password using the System. Security. Cryptography.SymmetricAlgorithm class. This is essential when the credentials are read from setting files and want the password to keep secure from others.
6. If the Network Traffic Analyzer is busy processing the task queue like T1, T2, T3…Tn and suppose it is processing task T3. Then after the task T3 is finished CENTAS processed the task TP1 with the first priority and does its network analysis.
7. After the task TP1 is completed then continue from last pending task T3, T4……..Tn.
8. Then rest of the analysis process is similar to that of upload/download operation in network traffic analysis to find the details of packets flow in every 1 seconds.

3.8 CNETAS Implementation Details

3.8.1 CNETAS Use Case Diagram

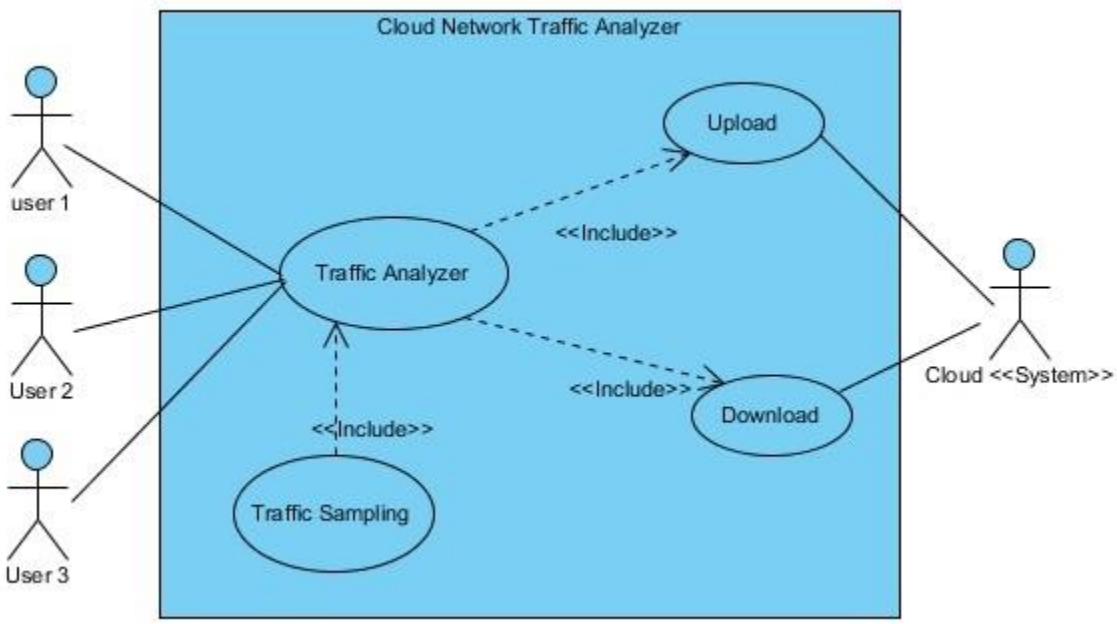

Figure 25: CNETAS Use Case Diagram

The figure 25 shows the CNETAS Use Case Diagram where the user is interacting with the cloud system while uploading or downloading data from cloud environment. It is showing the dependency relationship between the Traffic Analyzer with the Upload/Download operation. It shows the Upload/Download operation should be there for the Traffic Analysis to be carried out. It also shows the Traffic Sampling process depend upon the Traffic Analysis process. The Traffic sampling should only be considered if there is a Traffic Analysis process.

3.8.2 CNETAS Sequence Diagram

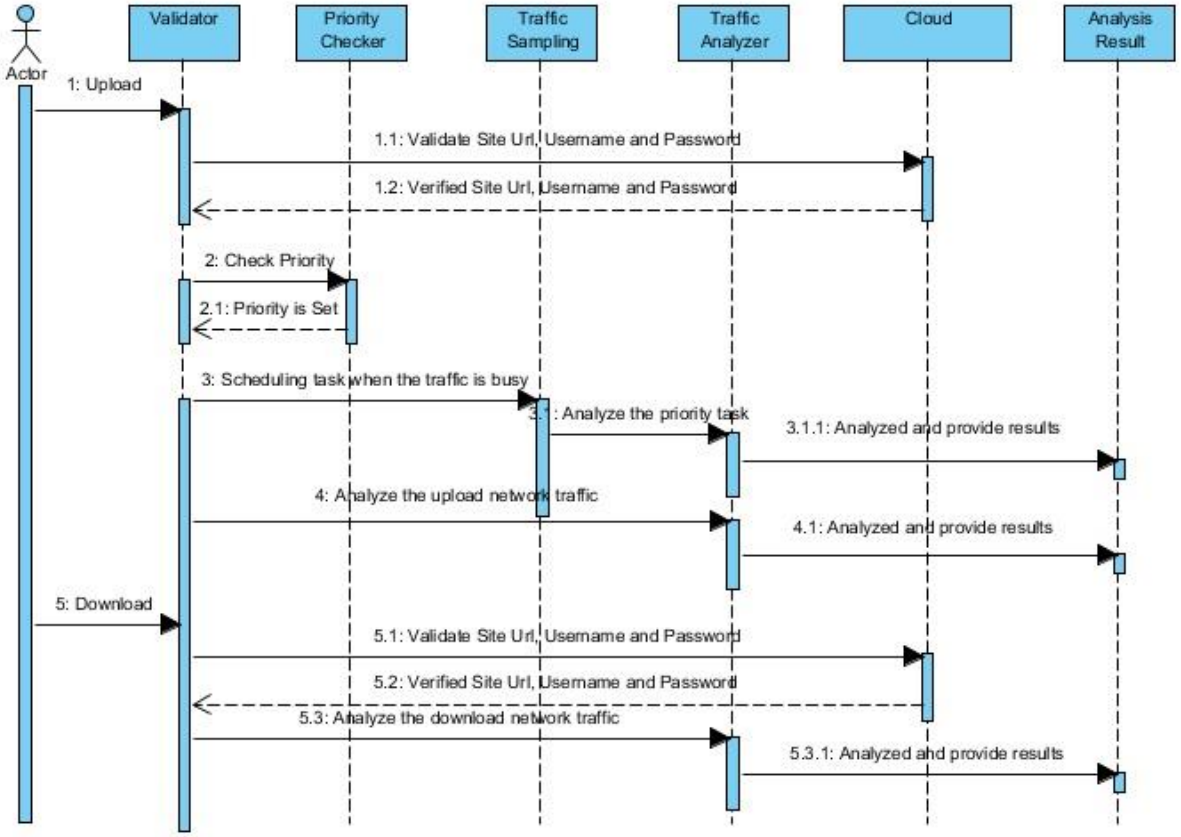

Figure 26: CNETAS Sequence Diagram

The CNETAS process involves the different steps for analysis and sampling of cloud traffic. The below are the section the system deals with,

1. Validator

The credentials that is passed either from the user interface or from the setting files will be validated by the validator and give the response of true or false. If it is not true further process can't be proceed ahead. The validator will verify the credentials from the cloud service provider.

2. Priority Checker

The Priority Checker block checks the priority of the task queue. If the priority is set by the user then the current task is processed and analyzed first and other task are handled sequentially.

3. Traffic Sampling

The Traffic sampling schedule the priority based task. If the priority is set for any task queue then the traffic sampling schedule to process the task with the high priority.

4. Traffic Analyzer

The traffic analyzer analyze the cloud network traffic when the operation of upload or download process is in progress to the cloud environment.

5. Cloud

In this system the cloud is the main environment when the user is accessing services for transferring data from the user system to the cloud environment. Here the SharePoint (office 365) is taken for the cloud environment. The SharePoint system is selected for the target because the demand of SharePoint is increasing rapidly due to the features and services that the SharePoint provides to the user.

6. Analysis Result

The Analysis Result section helps to provide the analysis result of cloud traffic to user. The result summary is provided in graphical form and details results will be provided in report format which the details information of packet flow in every seconds.

3.9 Tools used CNETAS

1. Development Tools(IDE)

Microsoft Visual Studio 2012

2. Language

C#.net

3. Documentation Tools

Dia

Visual Paradigm

Microsoft Office

3.10 Programmatically CNETAS components

1. Windows Forms

For the graphical point of view and to makes the user more user friendly the windows forms will be used. The form can be used for taking input for the site credentials, uploading the contents to the cloud environment, downloading content from cloud environment, providing various option for processing and to display the analysis report in graphical format.

2. Class Files

The class files will be used for definition and declaration of different function for making system performing different function as per the requirement of the system.

3. Xml Files

Xml files will be used to set different input parameter so that the program read the input parameter for the processing of cloud request. The input parameters that are not provided from the GUI interface can be provided from the configuration file with defining different parameters in the xml setting files.

4. Dll files

Microsoft.SharePoint.Client

Microsoft.SharePoint.Client.Publishing.dll

Microsoft.SharePoint.Client.Runtime.dll

The following dlls will be used as it helps to use the API of the SharePoint site. It makes easy to interact programmatically with the SharePoint site like authenticating with the site, reading site collection, reading list/library of the site, uploading or downloading files from SharePoint.

4 Simulation and Outcomes of Cloud Network Traffic Analyzer

The experiment is carried out simulating the different size data migrating to cloud and from the cloud. The outcome results are analyzed and represent in form of graphs to illustrate the results and the finding of the Cloud Network Traffic Analyzer.

4.1 Creation of Cloud Environment for Cloud Traffic Analysis

For the cloud environment we have created an account in SharePoint online where the documents are migrated and analyzed the cloud network traffic. The following are the steps to create a SharePoint Online account.

1. Create an Office 365 (SharePoint online) trail account as shown in the screenshot below,

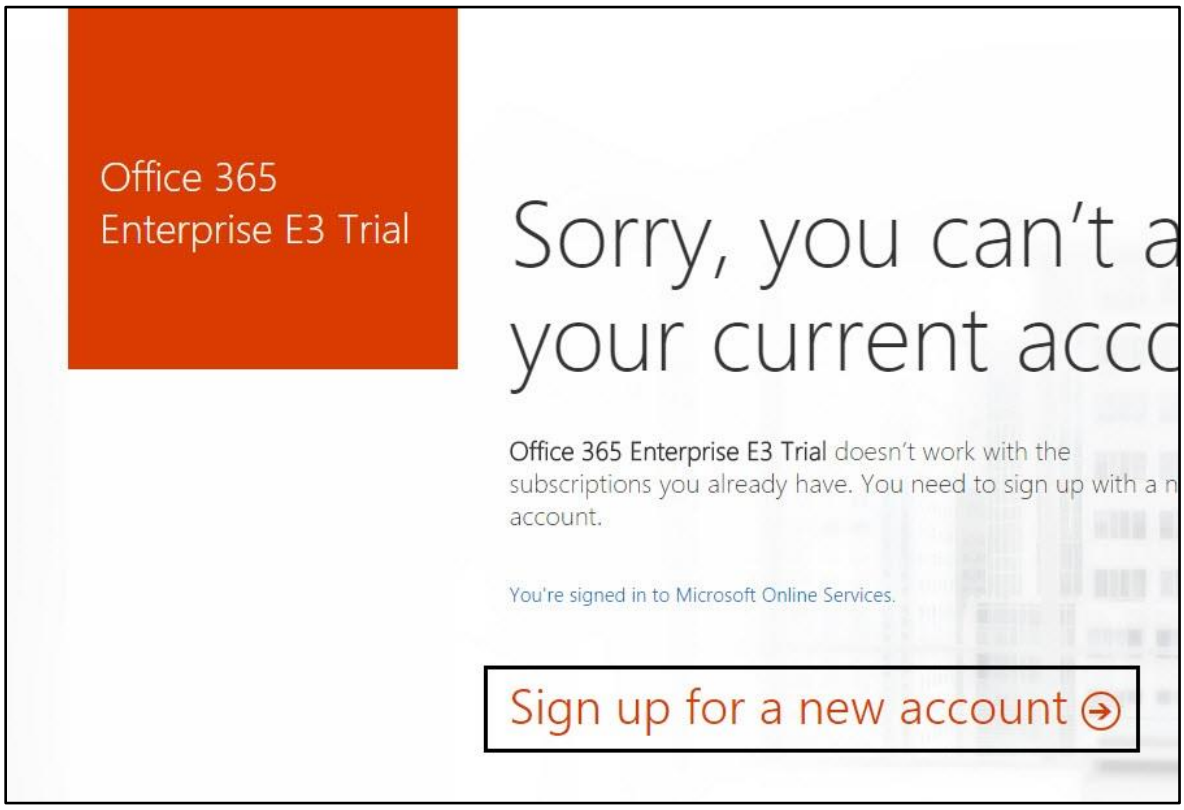

Figure 27: Showing the link to sign up the cloud account

2. Sign up for a new office 365 trail account filling all the required fields that is shown in the image below,

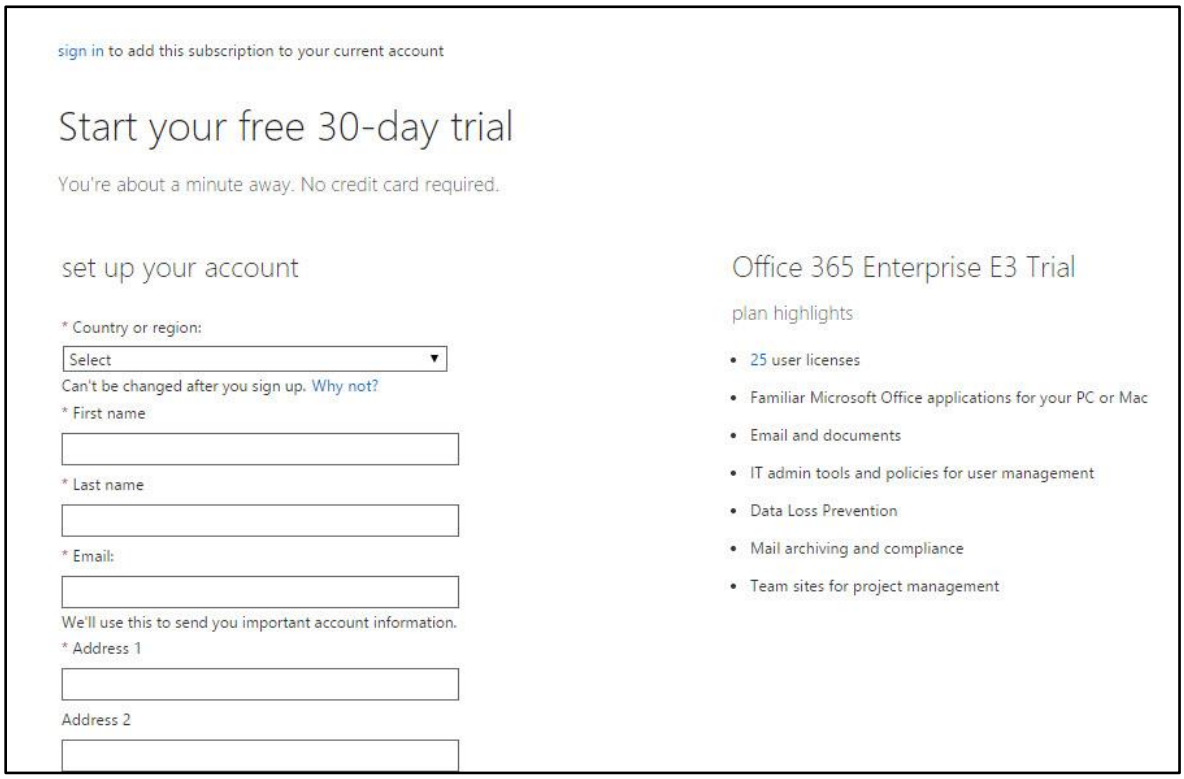

Figure 28: Showing the sign up page for cloud account

3. When the account is created sign in to the page from the link provided below,

   https://login.microsoftonline.com/login.srf?wa=wsignin1.0&rpsnv=4&ct=1431092071&rver=6.4.6456.0&wp=MCMBI&wreply=https:%2F%2Fportal.office.com%2Flanding.aspx%3Ftarget%3D%252fdefault.aspx%253flc%253d1033&lc=1033&did=501392

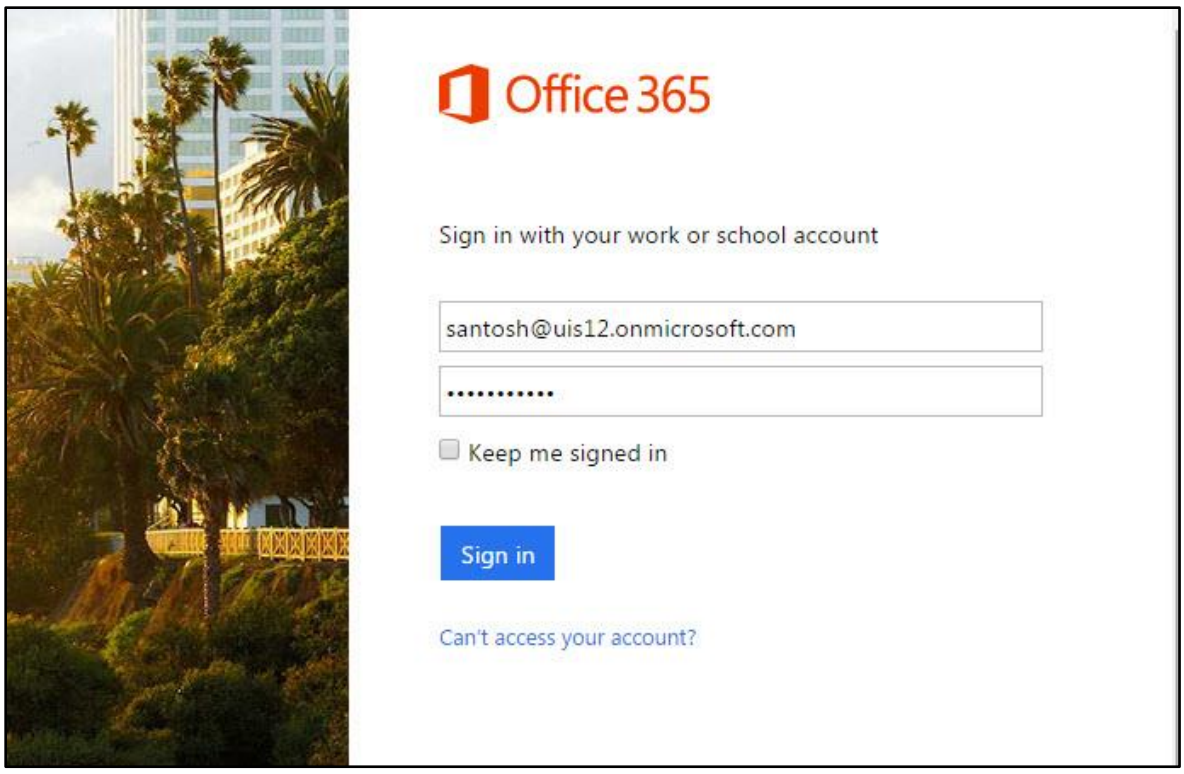

Figure 29: Page to sign in the cloud account

4. After sign in, go to SharePoint Admin Centre from the admin page and create a new site collection. There are different types of site templates but here we have selected a Team Site template to create a new site collection. New site is created providing all the required parameters.

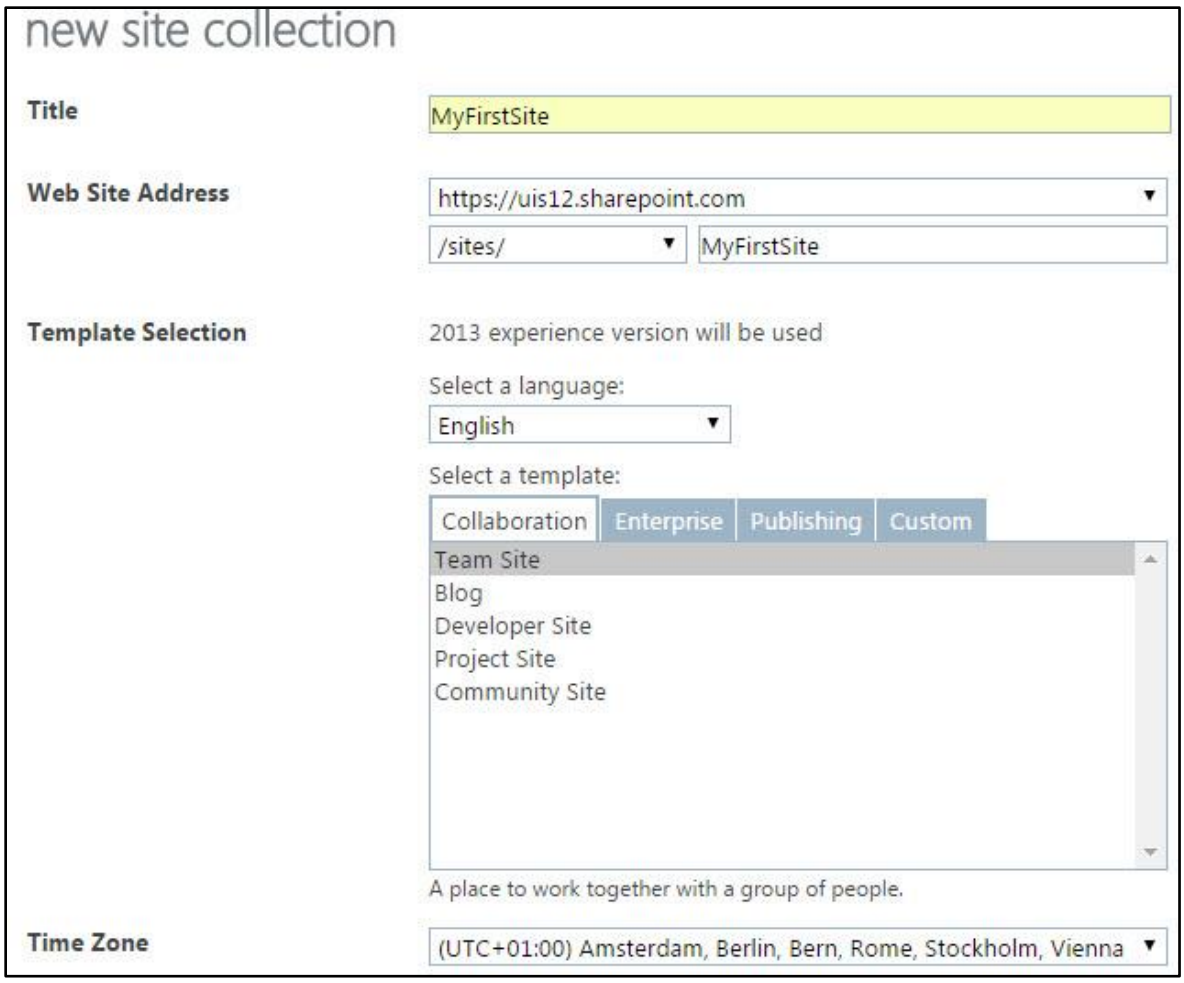

Figure 30: Creating new team site collection in cloud

5. The site collection is created with the following site properties as shown below,

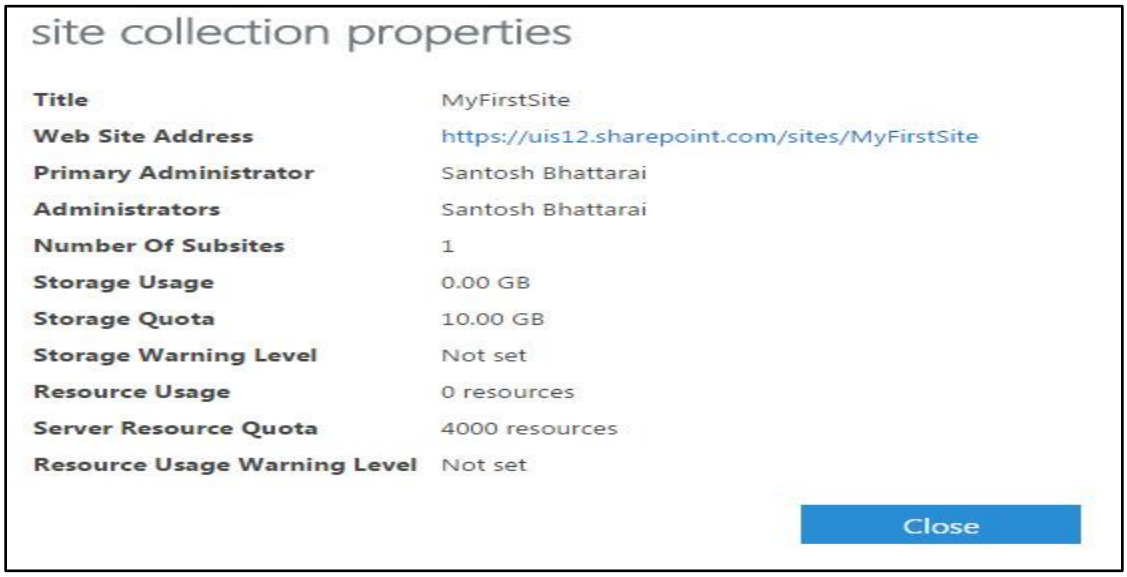

Figure 31: Site collection with its properties

6. Finally, the site collection look like,

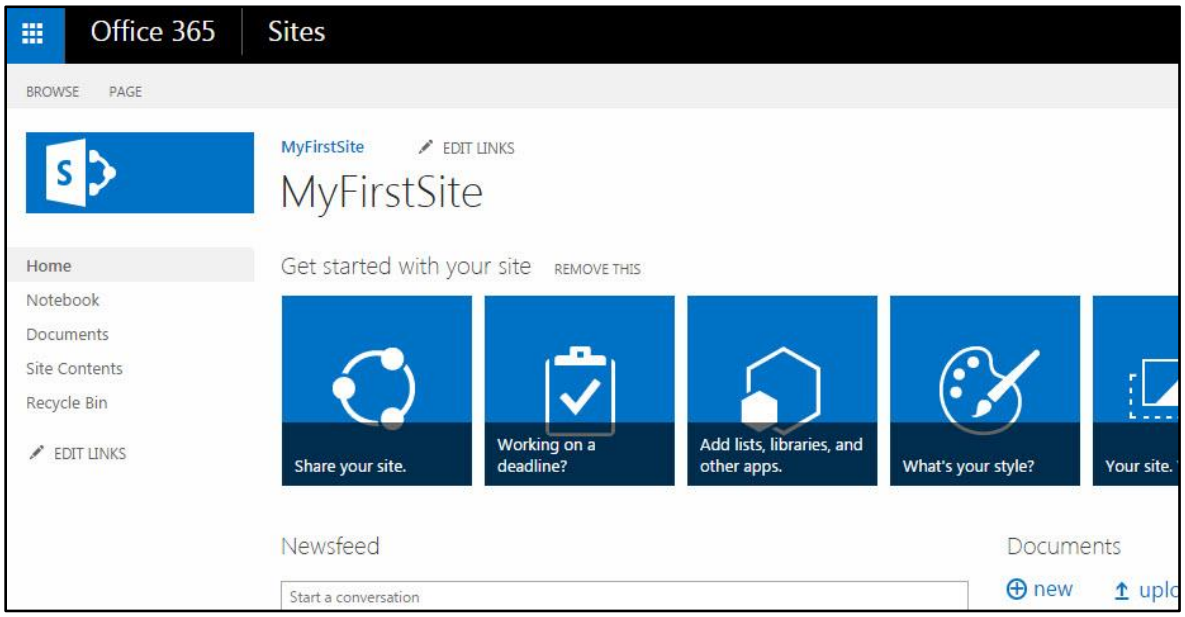

Figure 32: Layout of cloud site collection

7. The library with name “CloudLibrary” is created where the documents are uploaded and downloaded from the cloud. Here the new CloudLibrary is created as,

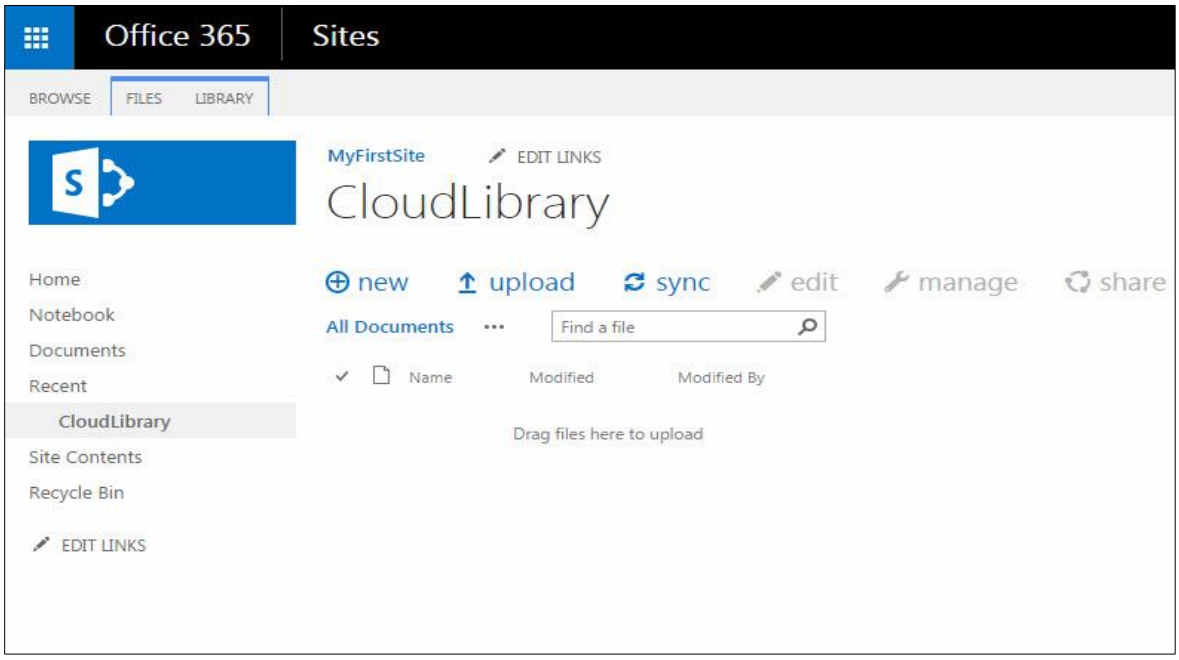

Figure 33: Showing CloudLibrary in cloud site collection

4.2 Simulation of Cloud Network Traffic Analysis

CNETAS uses a **GUI interface** to make it easy to use and understandable. It uses different graphical from to set the options for upload or download operation. In addition to this, the network analysis results are displayed in the GUI format and its details traffic analysis of each second network track results is written in log. So that, the network traffic analysis results are easy to read and interpret. The following as the output results of cloud network traffic analyzer.

1. The below is the first screen of cloud network traffic analyzer. It gives a general information showing the cloud network.

Figure 34: First screen of cloud network traffic analyzer

2. This is the option selection form where the option to upload or download is provided. There is also sub option in upload section to set the priority of the task if the task need to be handled with highest priority.

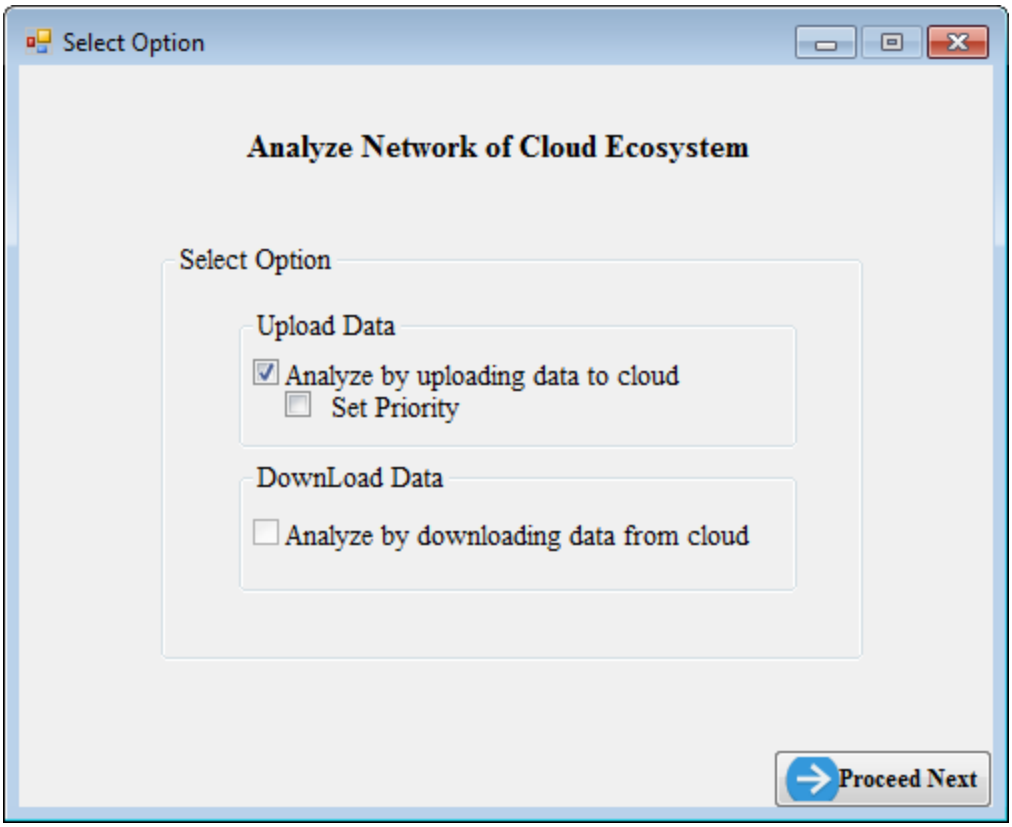

Figure 35: Section to select upload or download operation

3. This is the GUI form to provide site credentials to connect the cloud environment. In the left hand side there is the browse button to upload the document to cloud and analyze the network traffic in upload process. The screen below showing processing of 17 MB of document in cloud.

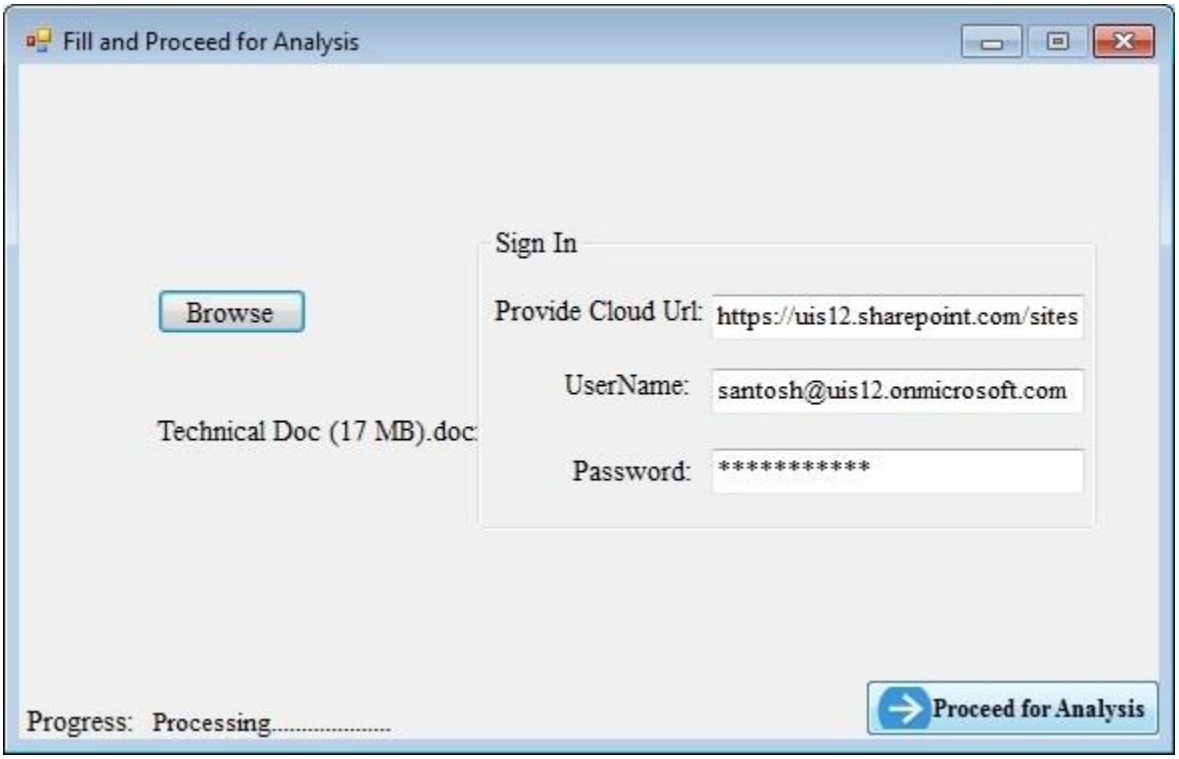

Figure 36: Uploading and analyzing the cloud network traffic

4. This form is to download the data from cloud environment and analyzing the network traffic during download process.

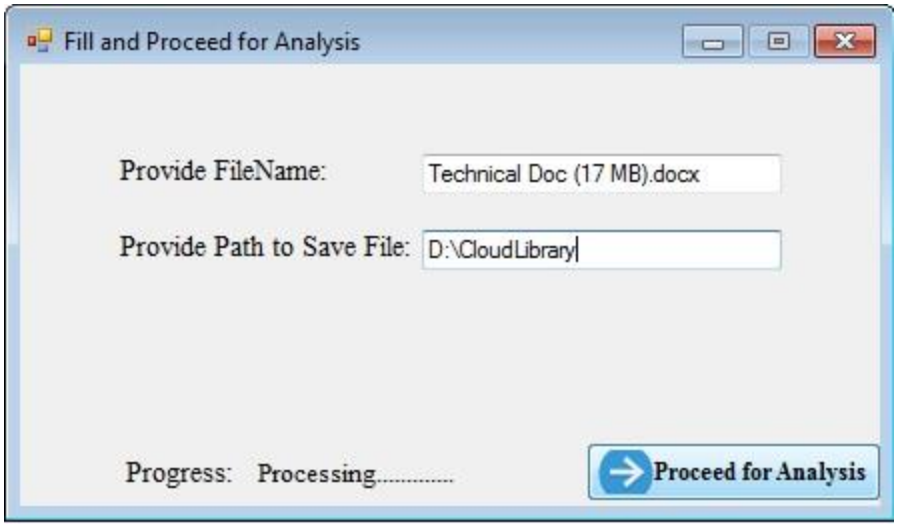

Figure 37: Downloading and analyzing the cloud network traffic

5. The below is the traffic analysis summary report showing the information of data migration to cloud and analysis summary report. For more details of traffic and performance analysis during the operation, it can be viewed by clicking the Traffic and Performance button. It shows the details analysis report of network traffic and performance.

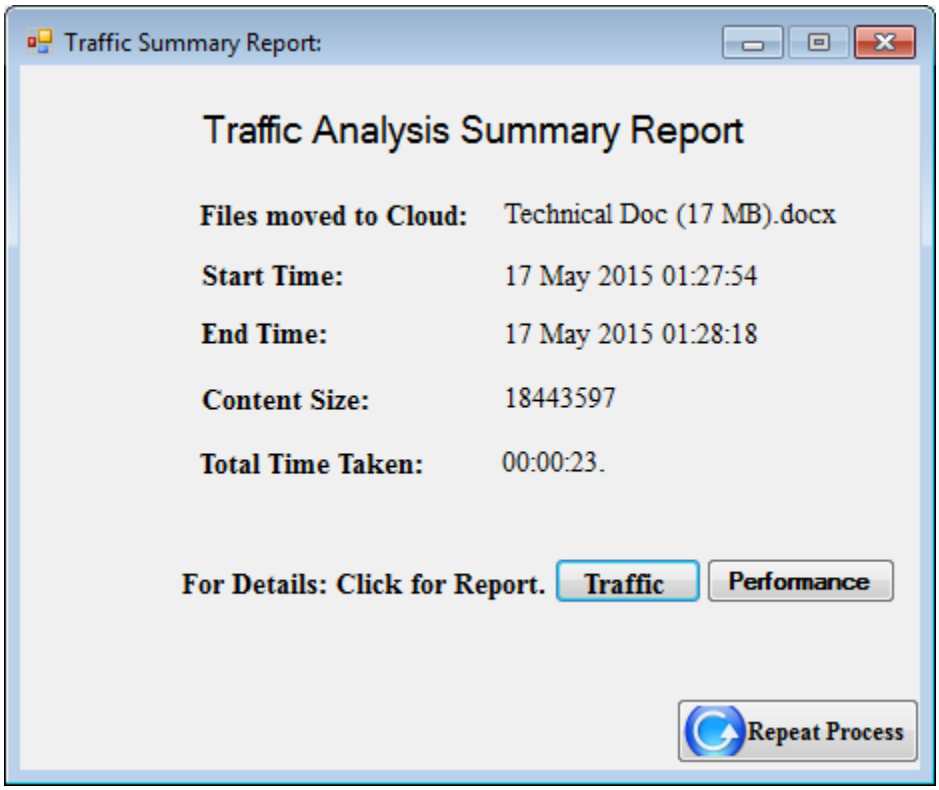

Figure 38: Traffic Analysis Summary Report

6. This is the traffic analysis report provided in log format. The log is written of every second while download or upload operation to cloud. The below highlighted is the trace of data flow to cloud in every second. This is the log of upload process so only the packet flow in this process shows value. Similarly for the download operation, the download section has data flow value. The log is saved with file name plus the current date and adding Tra in the beginning.

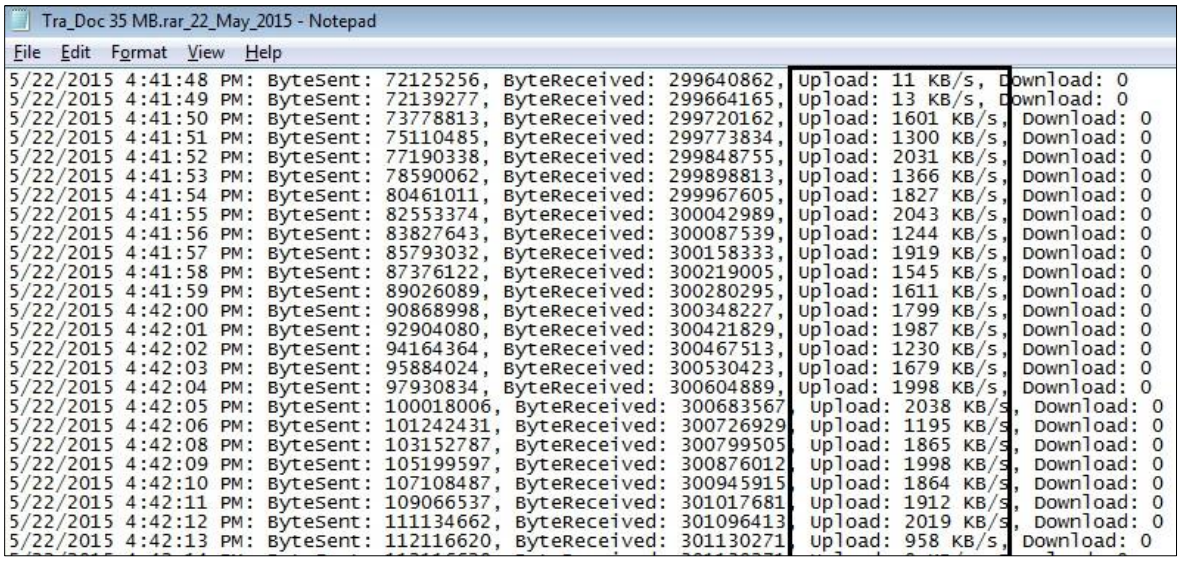

Figure 39: Showing data flow in every second in upload operation to cloud

7. This is the performance log result track during upload operation to cloud. The performance log is saved with the name of the file plus the current date and adding Per in the beginning of the name. The memory are defined as,

    - Memory - Private Working Set -> It is the subset of working set that specifically describes the amount of memory a process is using that can't be shared by other processes. We have named private memory uses to notify this memory uses.

    - Memory - Working Set -> It is the amount of memory in the private working set plus the amount of memory the process is using that can be shared by other processes. We have named total memory uses to notify this memory uses.

    - Memory - Peak Working Set -> it is the maximum amount of working set memory used by the process. We have named maximum memory uses to notify this memory uses.

    - CPU Usage -> It is the percentage of time that a process used the CPU since the last update.

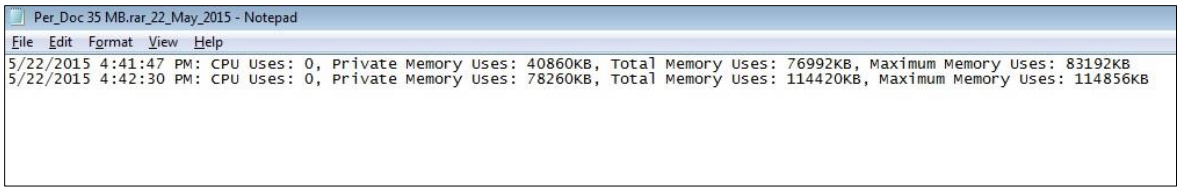

Figure 40: Showing performance trace in upload operation to cloud

4.3 Design of Experiments and Patterns

The experiments is perform with transferring different size documents to cloud and from cloud. The small size data taken as 2M, 4M, 6M, 8M and 10M. For experimenting for huge data size taken as 5M, 10MB, 15MB, 20MB and 25MB. The data is transferred to cloud in different time considering network peak hours and normal hours. The analysis of network traffic is performed uploading this data to cloud and downloading data from cloud. The observed results are noted and compared the results of different conditions.

4.4 Analysis of the Results of the Experiments

4.4.1 Transfer of small size data to cloud

The small size data of size 2MB, 4MB, 6MB, 8MB and 10MB is transferred to cloud and noted its total time taken to transfer the data to cloud environment. When we compare and analyze the data it shows that the more the size of data the time taken to move to cloud is in increasing trend. The graph plotted below shown the time taken is directly proportion to the size of data migrated to the cloud environment.

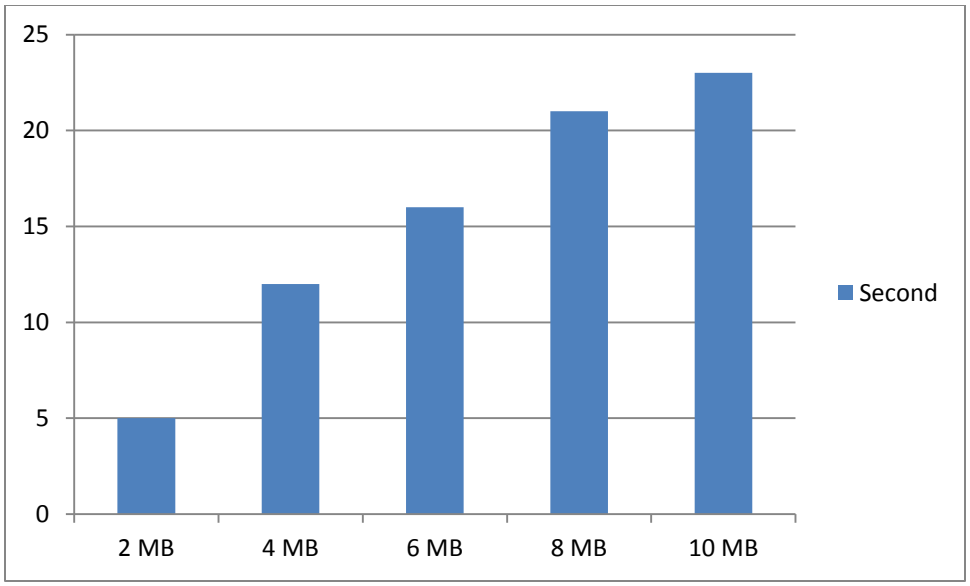

Figure 41: Moving data to cloud and comparing the results

4.4.2 Experiment with another combination of data moving to cloud

The data of size 15MB, 20MB, 25MB, 30MB and 35MB is moved to cloud and analyze the network traffic results. The below diagram shows the relationship between the size of data and time taken to move to the cloud.

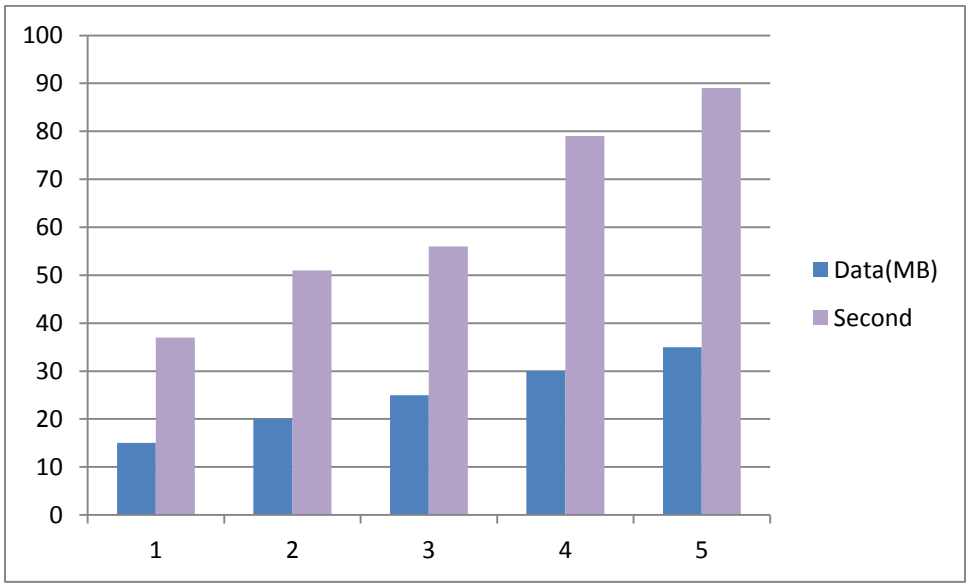

Figure 42: Comparing data and time relationship

4.4.3 Simulator performance with different size of data moving to cloud

The bar diagram below shows the performance of cloud network traffic analyzer during traffic analysis while moving data to cloud and performing the analysis. The data is moved to cloud and simulator is analyzing the network traffic simultaneously the simulator note the memory of the start and end of the processing. The results obtained while moving 15MB to 35MB data is noted and plotted a graph as shown below. It shown that, the memory consume by the process is also increasing with the size of data but the start and end memory is somehow irregular. In some cases the start memory is showing more and in some cases end memory is showing more as shown in diagram below.

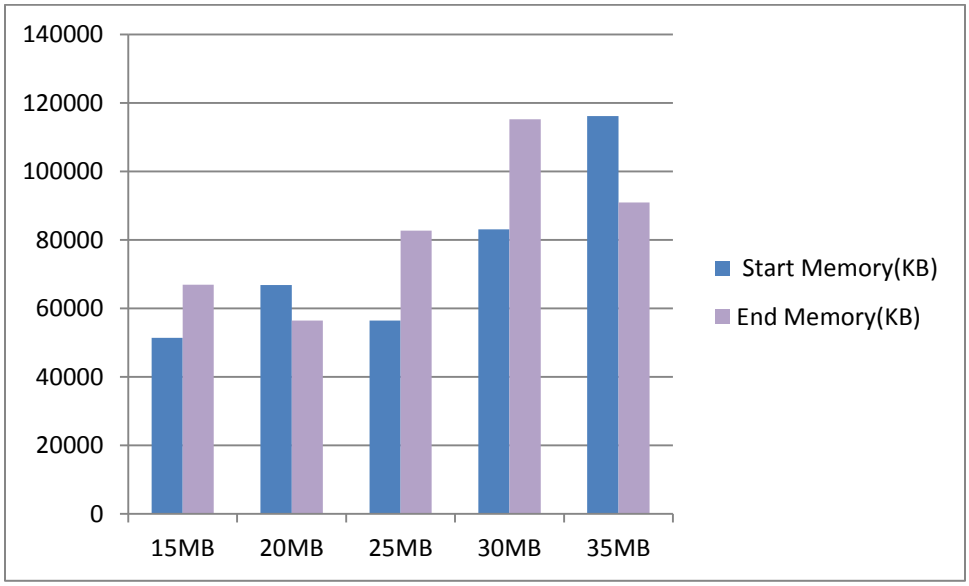

Figure 43: Comparison of Memory with Data transfer

4.4.4 Flow of data plotted in graph

The graph plotted below is showing the data transfer per second while transferring 4 MB data to cloud environment. The log result provided by the network traffic analysis simulator will have information of data transfer in every time interval (in seconds). The result is plotted in the graph as shown in figure below. The data transfer rate started from the minimum value reaches to the peak value and drop at the end of the completion of the operation which is the expected behavior.

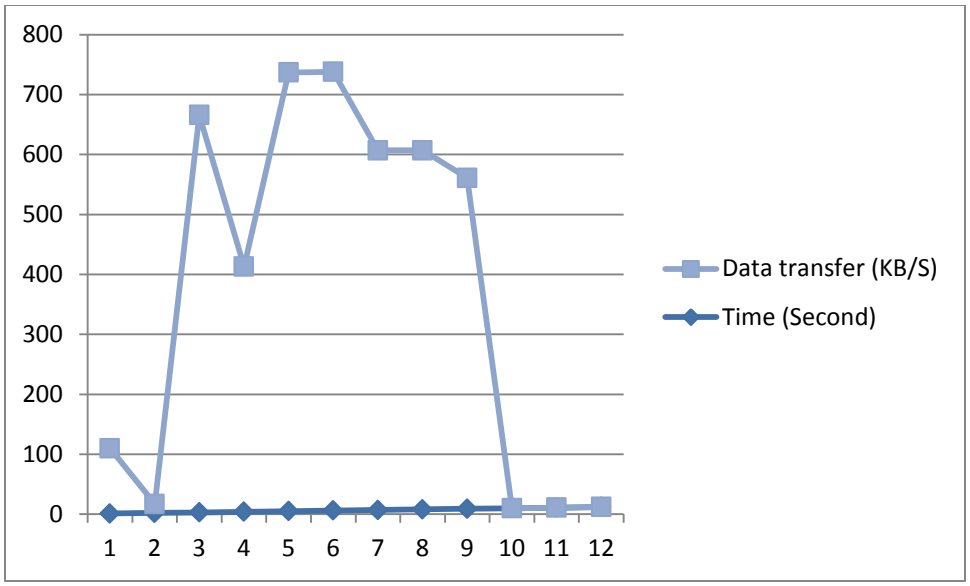

Figure 44: Graph plot of 4 MB data transfer per second

4.4.5 Downloading data from cloud and comparing the results

As the cloud network analyzer analyzes the result while uploading data and download data from cloud. The below bar diagram shows the time taken while downloading data from the cloud. The experiment is performed with downloading the different data size from cloud and plotted in a form a bar diagram. The diagram looks like as shown in figure below.

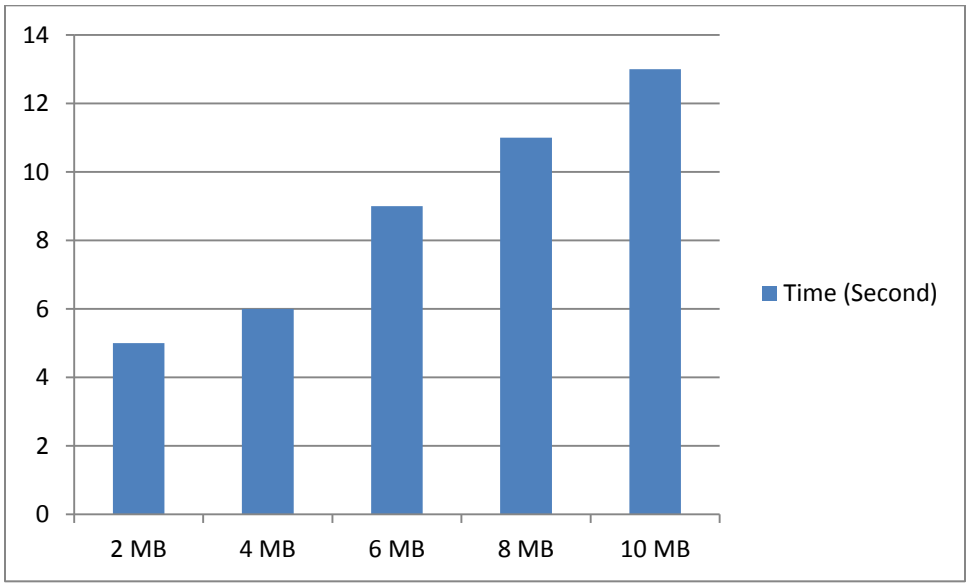

Figure 45: Comparison of downloading data with time

4.4.6 Data flow plotted in graph

The graph plotted below shows the flow of data in every second while analyzing the cloud network traffic while downloading data from cloud. The data is downloaded from cloud environment and the traffic analysis result of data flow in every second is plotted in form a graph.

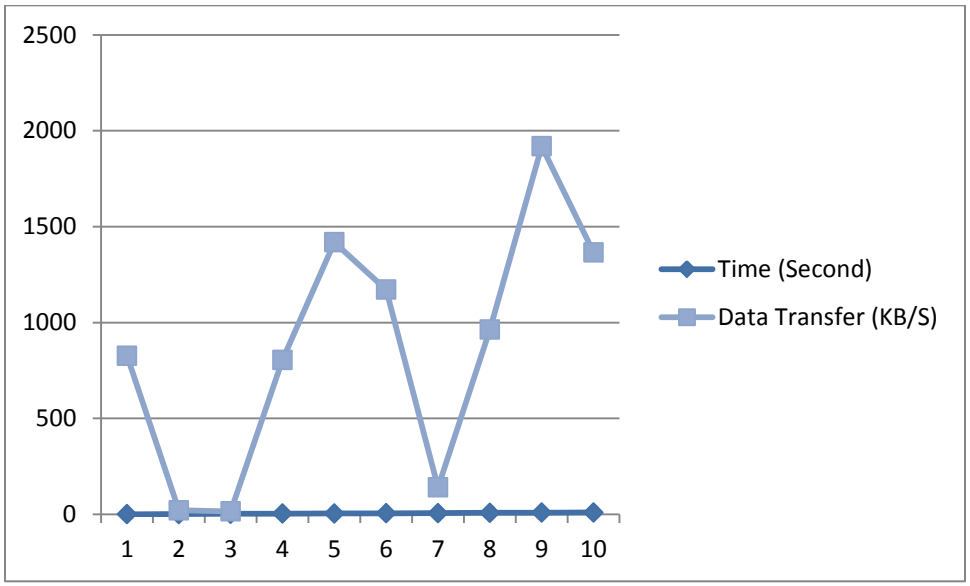

Figure 46: Download of 8MB data from cloud

4.4.7 QoS of the Cloud Service

Actually the Quality of the service is based the QoS parameters. The parameters like throughput, error rates, transit delay, security, priority, packet loss etc. The throughput is the simply a bandwidth which is used to send and receive the service request and the service response. The error rate is defined as the number of error that occur during the transfer. The transit delay is the time gap between the request service and waiting for service response. The simulation mainly focuses the network analysis for the smooth flow of information between cloud service request and the cloud service provider.

1. Cloud cost calculation

The cost of the cloud service can be calculated with the various data centers which the various data centers dependent parameters such a cost of the VM, Memory cost, Storage Cost, Data storage cost which helps calculating the total cost of the data centers. Finally it gives us the complete response of cloud service request. The graph illustrate below the total cost of the cloud service request is directly proportional to the cost of other parameters like the cost of VM, Memory, Storage, Data storage etc.

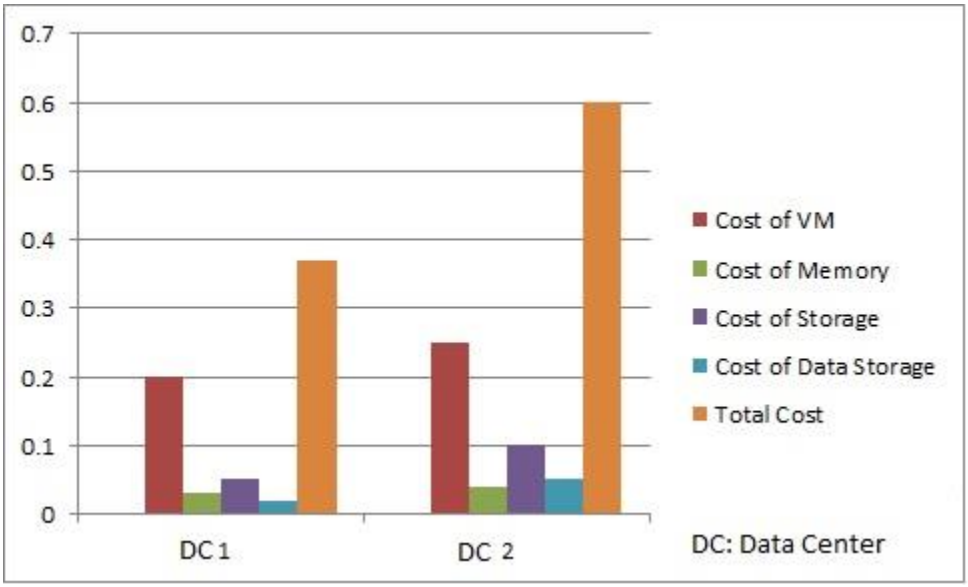

Figure 47: Comparison of Cloud Service Cost

2. Calculation of cloud error rate

The cloud error rate can be calculated with the difference value of number of request to the cloud server and the number of response from the cloud server multiplied by 100. The formula to calculate the cloud error rate can be summarized as,

$$CER = (NResquest- NResponse) * 100$$

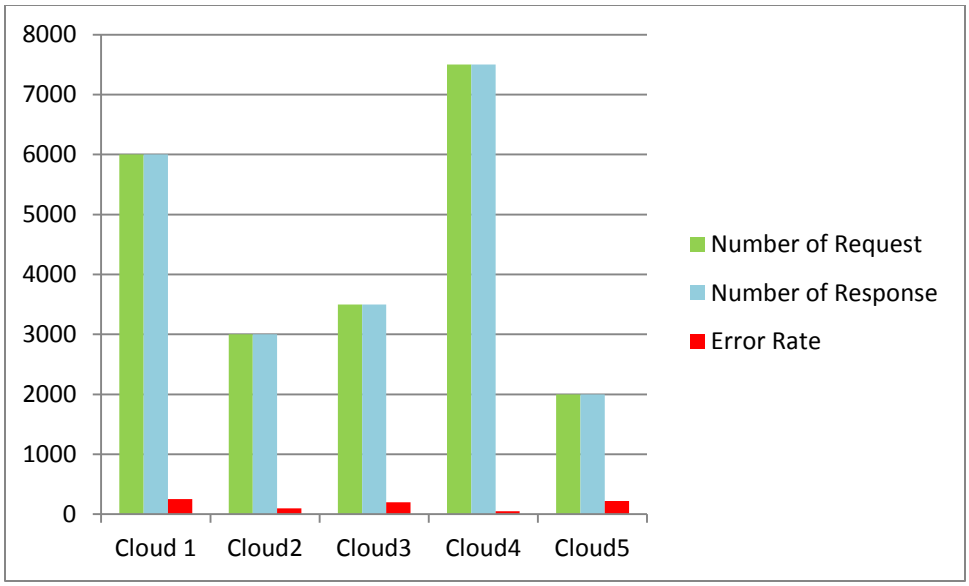

Figure 48: Calculation of cloud error rate

Based on the data received from cloud network traffic analyzer we are getting the response of the request from the cloud server with some exceptional case. This is because we are testing data with less number of request so getting response of all the request. But in real scenario the case is different as there are huge number of request coming from many user at a same time. The response are delivered from the number of request but there remain some error factor which is shown in the graph above considering the ideal case handing of large number of data by the cloud server.

3. Calculation of Cloud Response Rate

The cloud request is send to the cloud server to get the proper response. The cloud request for bandwidth and the response bandwidth is allocated on the basis of available bandwidth. Hence, the response rate can be calculated on the basis of response bandwidth divided by 1024 and the resulted value multiplied by 1000. The formula is,

CRR = (CRB/1024)* 1000

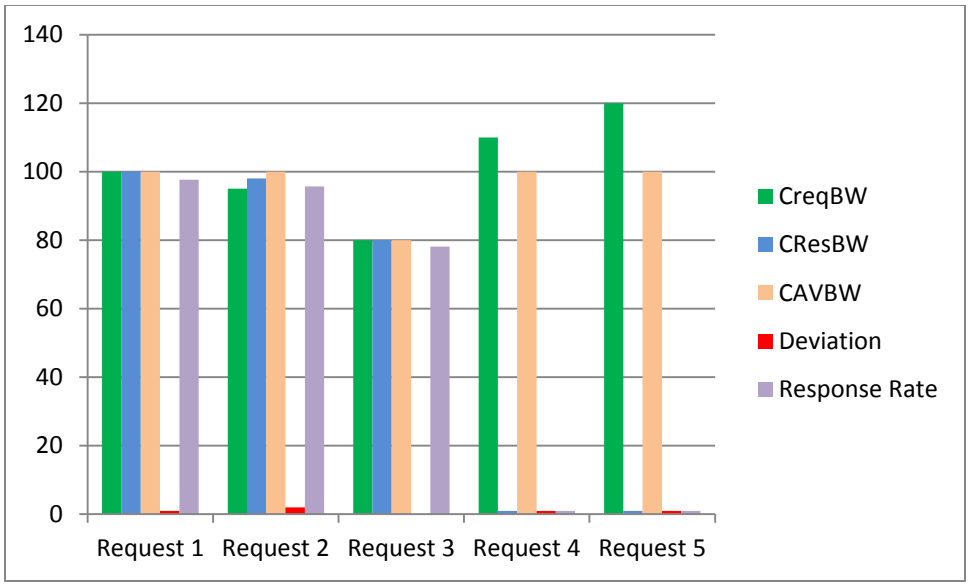

Figure 49: Calculating cloud response rate

The bar graph shows the comparison of cloud response rate considering the different scenario. It shows the response rate is depend open the requested bandwidth, available bandwidth and the response bandwidth itself. The graph predicts, it is good that the request bandwidth should be equal or less than the available bandwidth to get the good results.

5 Conclusion

There are large number of services are being exchanged between the cloud service and cloud user to meet the expectation from both ends. The cloud user request for the service and cloud handles it with high quality response but there are many limitation in which the cloud are not able to deliver the response in the correct time. This may lead the cloud user to increase the dissatisfaction over the services. So, in order to give the smooth service and to analyze network traffic of cloud ecosystem the cloud network analyzer and simulation will be a solution. It not only tells about the network traffic, data and time taken for the request and response of the service but also provides the clear picture about the current network scenario so that they can estimate the time and data for cloud service. It is equally helpful for researchers and students who want to gain a details information about cloud environment. This system will be the gateway for researchers and students to gain the information about cloud and its environment.

CNETAS helps to analyze the cloud data in terms of size, transfer rate, time, packet flow etc. Furthermore it helps for scheduling to allocate the necessary bandwidth in a busy network to handle urgent request faster. This will help to increase the customers believe and reliability towards the service. The other important area is that, these days the data migration to cloud environment is growing as the data stored in cloud is safe and easily accessible from anywhere irrespective to place and location. So this simulation result is equally useful to estimate their plan for migration in terms of data and time taken to migrate the data in cloud environment. As we know that, cloud is a very vague topic that has many sectors so there still remaining many topics to research and study where I want to explore more about in the future. Similarly, there are still more to work out on this topic so enhancing the network traffic analysis feature broadly is my future plan. Enhancement may be in terms of security, providing more network analysis results or adding more advanced feature so it can be used widely for commercial purpose as well.

Abbreviations

ICT	Information and Communications Technology
CSP	Cloud Service Provider
IPTAS	IP Trace Analysis System
HDFS	Hadoop Distributed File System
JVM	Java Virtual Machine
CNETAS	Cloud Network Traffic Analyzer and Simulator
CTA	Cloud Traffic Analysis
CTS	Cloud Traffic Sampling
KPI	Key Performance Indicators
CSC	Cloud Service Cost
NRU	Number of Request per User
CSR	Content Size per Request
PHS	Peak Hour Start
PHE	Peak Hour End
PSR	Peak Service Rate
OPSR	Off Peak Service Rate
CER	Cloud Error Rate
CRR	Cloud Request Rate
CRB	Cloud Request Bandwidth

REFERENCES

Fershtman, C. & Gandal, N. 2012, Migration to the Cloud Ecosystem: Ushering in a New Generation of Platform Competition

Armbrust, M., Fox, A., Griffith, R., Joseph, A., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I. and Zaharia, M.,2009, Above the Clouds: A Berkeley View of Cloud Computing,Technical Report. University of California at Berkeley.

Dharmanto, S. 2012, Cloud Ecosystem, IEEE Cloud computing One Day Course Institut Teknologi Telkom, Presentation PowerPoint Slides, http://www.slideshare.net/hamdani2/cloud-ecosystem

Benson, P. 2013, The Cloud Defined, Part 1 of 8: On-Demand Self Service, Available from: < http://www.pbenson.net/2013/04/the-cloud-defined-part-1-of-8-on-demand-self-service/>. [3 March 2015].

Benson, P. 2013, The Cloud Defined, Part 2 of 8: On-Demand Self Service, Available from: < http://www.pbenson.net/2013/05/the-cloud-defined-part-2-of-8-broad-network-access/>. [3 March 2015].

Benson, P. 2013, The Cloud Defined, Part 4 of 8: On-Demand Self Service, Available from: < http://www.pbenson.net/2013/05/the-cloud-defined-part-4-of-8-rapid-elasticity//>. [3 March 2015].

Mishra, R. 2014, Cloud Computing introduction ppt, B.tech Student at Kurukshetra University, Kurukshetra, Presentation PowerPoint Slides, http://www.slideshare.net/Rkmishra00/cloud-computing-32035613?related=1

Saravanakumar, C., Arun, C., 2012, Traffic Analysis and Shaping of the Cloud Services over

Common Deployment Model using Cloud Analyst, Volumne 43, Chennai, India, International Journal of Computer Applications.

Microsoftlearning, 2010. Cloud Computing and Microsoft Offerings.[video online] Available at:<www.youtube.com/watch?v=uYGQcmZUTaw>

Exforsys, 2009. Cloud Computing Basic Components. [online] Available at: <http://www.exforsys.com/tutorials/cloud-computing/cloud-computing-basic-components.html> [Accessed 3 March 2015]

Arcitura Education Inc., Private Clouds.[online] Available at: <http://whatiscloud.com/cloud_deployment_models/private_clouds> [Accessed 04 March 2015].

Arcitura Education Inc., Private Clouds.[online] Available at: < http://whatiscloud.com/cloud_deployment_models/hybrid_clouds> [Accessed 04 March 2015].

Arcitura Education Inc., Private Clouds.[online] Available at: < http://whatiscloud.com/cloud_deployment_models/public_clouds> [Accessed 04 March 2015].

cloud.cio.gov, 2014, Cloud Computing Service Models

Private Clouds.[online] Available at: < http://cloud.cio.gov/topics/cloud-computing-service-models > [Accessed 04 March 2015].

Aidan, 2013, Essential Characteristics of the Cloud Computing Model.[online] Available at: < http://news.dediserve.com/2013/03/29/essential-characteristics-of-the-cloud-computing-model-2/ > [Accessed 04 March 2015].

Doug Thaler, 2013, The Importance of Cloud Computing for Business.[online] Available at: < http://ezinearticles.com/?The-Importance-of-Cloud-Computing-for-Business&id=7475804 > [Accessed 04 March 2015].

Wang, H, Ding, W, Xia, Z, A Cloud-Pattern based Network Traffic Analysis Platform for Passive Measurement, School of Computer Science and Engineering, Southeast University, China, pp. 1-4.

Lee, Y, Kang, W, Son, H, An Internet Traffic Analysis Method with MapReduce, Chungnam National University,Daejeon, 305-764, Republic of Korea, pp. 357-358.

White, T, 3rd Edition, Hadoop: The Definitive Guide.

Pan, J. 2008, A Survey of Network Simulation Tools: Current Status and Future Developments, pp. 3-14.

FABIO RUINI, 2011. Available at: < http://www.fabioruini.eu/blog/wp-content/uploads/2010/12/Screen-shot-2010-12-06-at-00.26.46.jpg>.

Cellbio, 2013, chapter_1_-_internet_networks_and_tcp-ip » 1-2_the_tcp-ip_family_of_internet_protocols. [online] Available at:< http://www.cellbiol.com/bioinformatics_web_development/doku.php/chapter_1_-_internet_networks_and_tcp-ip/1-2_the_tcp-ip_family_of_internet_protocols> [Accessed 24 April 2015]

Academic Business Consultants, Inc., Internet protocol. [video online] Available at: < https://www.youtube.com/watch?v=zyL1Fud1Z1c&list=PLFD059A2C0DE2DD6B&index=1>

Dux Raymond Sy, PMP, 2009, What is SharePoint? , pp. 7-21.

Microsoft, 2015, What is SharePoint. [online] Available at: < https://support.office.com/en-us/article/What-is-SharePoint-97b915e6-651b-43b2-827d-fb25777f446f>[Accessed 24 April 2015]