Assessment of Service Level Agreements and QoS Functions for Web and Video Traffic Using OPNET Simulator

Binyam Shiferaw Heyi

Department of Computer Engineering Addis Ababa Science and Technology University (AASTU), Addis Ababa, Ethiopia

Email address:
bsh385@gmail.com

Abstract: Basically Internet it's a connection of small and large networks. From the beginning Internet used only for transferring some information between two computers but nowadays Internet are huge commercial communication network. As usual Internet used for communication between applications. The first Internet applications weren't so sensitive for packets loss and delays. But in the modern Internet and networking applications requires more high level of connection quality, application become more sensitive for delays and packets loss and this is a reason to provide levels of quality (Quality of Service).The goal of experiment is studying applying QoS on network router (Access Router) to conform SLA (Service Layer Agreement). SLA includes End-To-End delay and Packet Delay Variation for Video Demonstration traffic. The approach to this paper is to use QoS functions to fulfill the specifications of the performance parameters in SLA between the operator and the customer. From the study it is found that using RSVP (Resource Reservation Protocol) gives better delay and delay variation results at the expense of more bandwidth.

Keywords: Quality of Service, Web and Video Traffic, OPNET Simulator, Service Level Agreements

1. Introduction

Although in early time of internet development the concept of Quality of Service (QoS) was not noticed very much, after while by developing different kinds of application in internet world this concept become more noticeable. The concept of QoS is very important in point of costumer that pays for services in network and service provider who want have a successful business.

The first Internet applications weren't so sensitive for packets loss and delays. But in the modern Internet and networking applications requires more high level of connection quality, application become more sensitive for delays and packets loss and this is a reason to provide levels of quality (Quality of Service). In the first Internet protocols used principle of Best Effort this type of service works with three rules:

- No blocking for incoming traffic to the network
- The information will transmit with the same.
- Always try make operation with traffic as fast as its possible.

But principle of Best Effort used only when Internet traffic was small in the network because for modern applications Best Effort not suitable. There is no system of prioritization of packets but it is needed for streaming applications. In a case if the packet wasn't delivery only the source can make decision to repeat sending of the packet.

Modern applications more sensitive for delays and they require system that going to control of service quality. Every type of applications needs personal quality of service. There two main factors for different levels quality of service.

1. Integrated Services architecture (IntServ)
2. Differentiated Services architecture (DiffServ

This report will study effect of different QoS mechanisms on performance of network in case of meeting requirement for different kinds of application such as video conferencing.

2. Background

Although in early time of internet development the

concept of Quality of Service (QoS) was not noticed very much, after while by developing different kinds of application in internet world this concept become more noticeable.

Quality of service is an internetworking issue with goals of delivering predictable amount of packets in networks. In fact in today internet different application in network should treat differently based on their feature. As an example real-time data flow such as multimedia and VoIP should process sooner in comparison with FTP data flow.

By means of QoS concept, latency-sensitive applications can have sufficient resource to act with good quality in network and as conclusion the resource of network can be used more efficiently.

There are several techniques for improving the QoS in computer networks. Some of the important techniques are queuing disciplines, traffic shaping, traffic policing, admission control and resource reservation. (1)

The three main service models in QoS are as follow:

Best Effort

This is a simplest and early service model in internetworking. By using the best effort as QoS all the data packets are treated equally. In fact in this service there is no guarantee for delivery of any specific packets in the network.

Integrated Services

By using this method, before starting data transmission, applications use the Resource Reservation Protocol. By using this protocols applications can request and reserve resources through the network.

Differentiated Services

In differentiated Services model packets must be classified at the edge of network and then sent to the network. Inside the network the routers treat with packets differently based on their classification.

It is worth to mention different kinds of mechanisms are used to achiev all functions in QoS technology. Token bucket, leaky bucket and different queuing discipline like as PQ [1] and WFQ [2] are examples of these mechanisms.

Additionally, Reliability, Delay, Jitter and bandwidth are four characteristics in data flow that can be mentioned when we talk about QoS. (1)

2.1. Classification of Data in Network

One factor are used to classify the data flow is the information in IP headers. There is an IP header sub field named as Type of Service (ToS) that can be used for this aim.

As Figure 1 shows, ToS contains 8 bits in IP header. Routers by looking to this part of IP header can figure out which datagram were considered most important by the packet senders and based on that can apply its policy to the packet. (2)

Figure 1 also shows that these 8 bits are divided to, 3 bits

for Precedence, 1 bit for delay, 1 bit for throughput, and 1 bit for reliability. The rest 2 bits are reserved for future use.
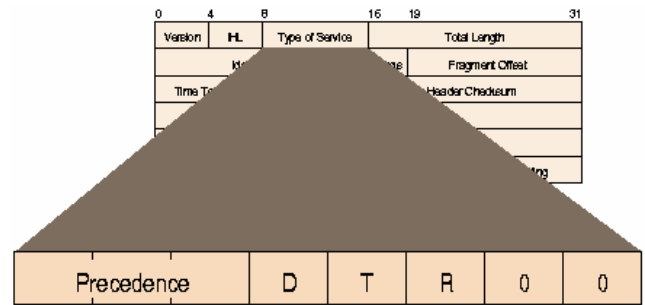

Figure 1. *IP header and Type of service sub field.*

2.2. Committed Access Rate (CAR)

Committed Access Rate is a QoS feature in a traffic policing. CAR introduces two QoS functions [3]

- Bandwidth management through rate limiting. By means of CAR bandwidth available for each of traffic type in the network can be limit. This function allows controlling maximum traffic received or transmitted on interfaces.
- Rate limits are able to specify which packets conform or exceed considering three parameters: average rate, normal burst size, excess burst size.
- Packet classification through IP precedence and QoS group [3] setting. This function allows to Partition networks into multiple priority levels or classes of service (CoS).

When traffic conforms to or exceeds the rate limit CAR provides re-actions such as: transmit, drop, set precedence, or set QoS group.

2.3. Resource Reservation Protocol (RSVP)

The Resource Reservation Protocol (RSVP) is a network control protocol like as ICMP IGMP. It is important to say clearly that the RSVP is not the routing protocol. The aim of this protocol is obtain differing qualities of service (QoS) for Internet applications.
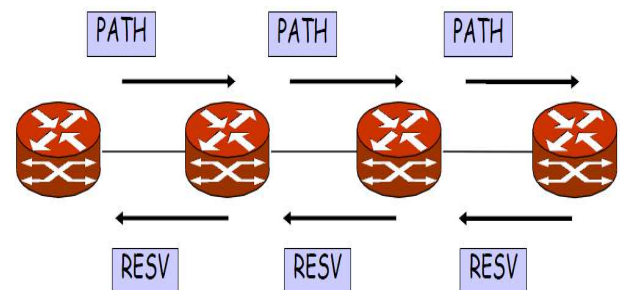

Figure 2. *RSVP basic messages.*

By means of these protocol resources reservation through the network for specific data flow will be done. In fact the

---

[1] Priority Queuing
[2] Weighted Fair Queuing

[3] A QoS group is a QoS class identifier internal to the router

sender firstly sends a path message to destination in order to set up a path between the sender and the receiver. By sending path request sender also could gather information about the available QoS feature for each router on the path. Destination will generate an RSVP message in response of path message. This message indicates the QoS and traffic information. When the RSVP message passing each of the routers on the path, if the router could support the transmission of the flow, it will reserve the resources for it; otherwise the router will generate an error message and send it back to the receiver [4]. Figure 2 has shown the basic messaging in RSVP.

2.4. What is SLA

SLA is an agreement between a service provider and a customer. The service level management (SLM) is the integrated management of all functionalities in the SLA. When a customer order a service from a service provider, first of all SLA is negotiated and then a contract will made. In SLA contract, QoS parameters that the service provider claim that can guarantee are included [5].

3. Simulation Schema

The schema and the topology used in this simulation are shown in Figure 3.
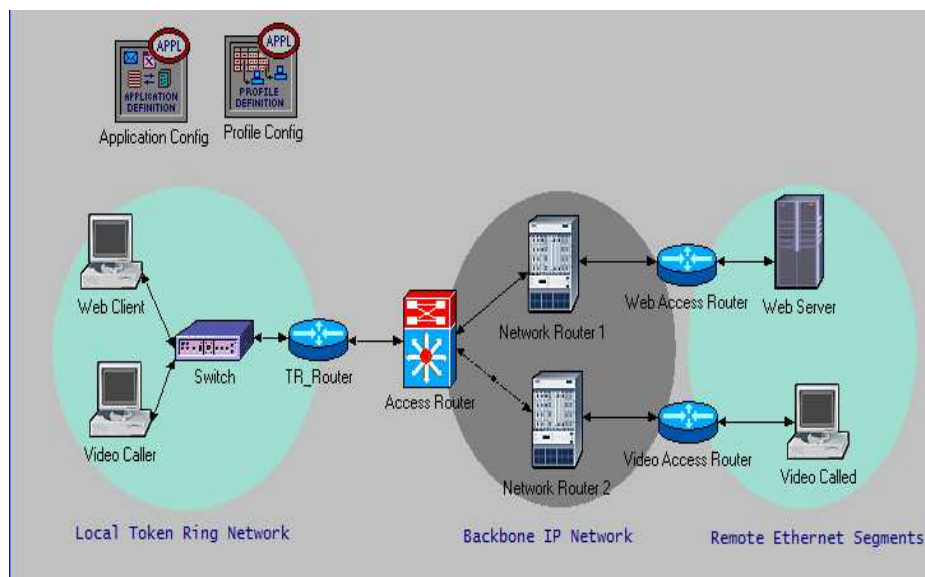

Figure 3. The topology used for simulated network.

In figure 3 the link between Access router and TR_router is a TR4 link with data rate 4Mbps.

Application and profile nodes are used for defining two used applications in the network.

One application is web page download by http with times between requests sent by client set to exponential10 (seconds). There are 101 objects contained in each page with one of them the size set to exponential100000 (bytes) and the rest exponential1000 (bytes).

The other application is a video demonstration application between Caller and Called.

Considering the setting for this application it can be implied that Caller sends frames with size of 1000byte each, every 1 second while Called has frames with same size but sends them every 0.002222 seconds.

Although the frame size is 1000bytes, the size of headers of protocols used to carry this frame should be considered while calculating the load as well. Different layers add RTP=12, UDP=8,IP=20 and Ethernet=14 bytes to this frame size.

The start time for both applications is set to 100seconds, and the offset is 10 seconds.

The configurations described till now are those common in all the four scenarios that are going to be simulated. These configurations describe a scenario as the default to show the problems that may occur in network performance without setting QoS policies.

The second scenario implemented adds the CAR functionalities as a parameter for providing QoS. CAR configuration is as follow:

A QoS node object is added to the topology, two rows are added under CAR profile settings for HTTP and Video Profile. For both profile a class of service as all traffic is defined to provide CAR policy based on traffic rate.

The all traffic service defines rate limit, conforming traffic and exceeding traffic policy.

Rate limit, has 3 attributes, average rate that specifies the amount under which all the traffic are conformed. Normal burst size defines the limit for traffic burst, excess burst size determines that the traffic between normal and excess is considered out of rate limit depending on a probability that goes up when the burst size increases[6].

In conforming traffic attribute an IP precedence is assigned to traffic and the action toward that traffic is

defined. The same is done for exceeding traffic policy.

Also car should be implemented on Access router interfaces with incoming traffics.

The third scenario uses both CAR and PQ queuing policy,

this new scenario is a duplicate of CAR scenario plus the queuing policy is implemented on the interface of the access router that passes the traffic to TR_router as figure 4.
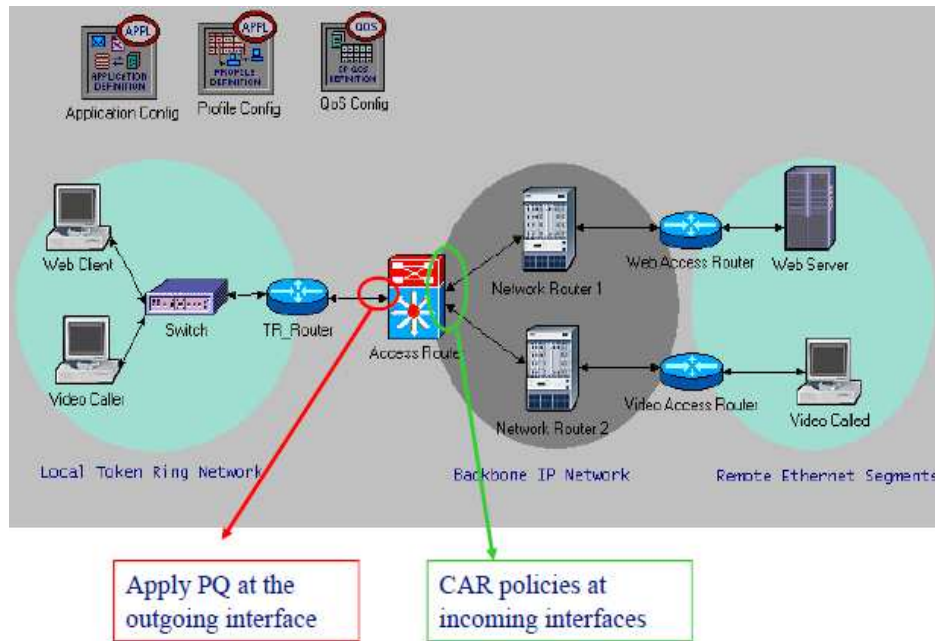

Figure 4. *the configuration for CAR and PQ.*

The forth scenario is the one with RSVP and the configuration is as figure 5.

Figure 5. *RSVP configuration.*

In QoS Config node the RSVP flow specification: Bandwidth is set to 490,000 bytes/sec

And buffer size to 50000 bytes

It is important to mention that SLA at these scenario are as follow:

The video traffic end-to-end delay must be less than the 40 millisec in more than 95% of running time.

The video traffic delay variation must be less than the 10 microsec in more than 95% of running time.

4. Analysis

4.1. The Default_network_scenario

In this part http and video application will be treated the same without setting any parameter for supporting QoS.

Traffic generated will arrive at access router and in the queue will be handled on FIFO policy.
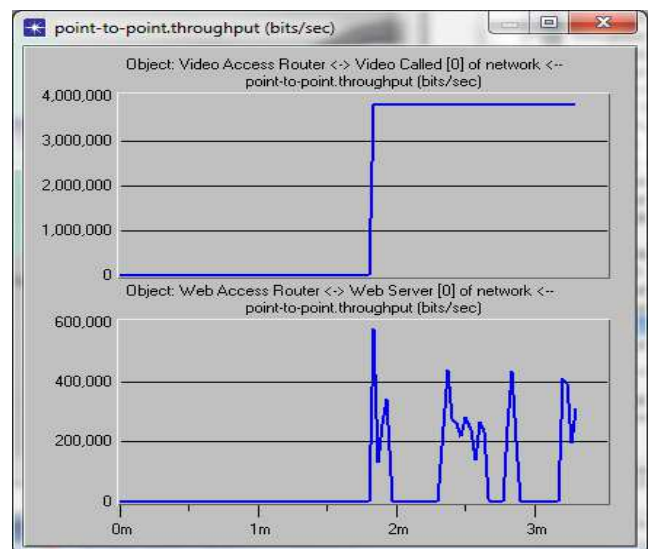
4.1.1. Throughput

Figure 6. *Throughput between Access routers and video called and web server.*

Figure 6 helps to explain the total throughput. The upper part of the figure is the graph for video traffic throughput the number is constant 3794400 bits/sec.

As explained in configuration part the frame size for video is 1000bytes also according to theory 54 bytes for overhead is added when it is being transmitted over network so the total frame size will be 1054, Video Called sends its packets every 0.002222 seconds meaning that every second it sends 450 packets therefore:

450*1054*8=3794400 bits is sent in one second which equals the result showed by graph.

Also the http traffic is sent on exponential basis both in size and time, which explains why the throughput has this fluctuated graph in lower part of figure 6 and why it goes to zero sometime (no traffic sent).
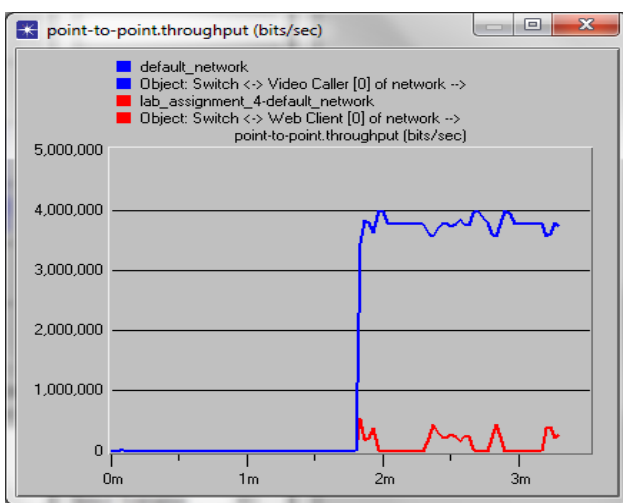

Figure 7. Throughput between Video caller and switch also web_client and switch.

Figure 7 explains that, throughput received at local network for video application is not the same as throughput on remote network at a constant rate, the reason is that when arriving at access router video packets may face http packets in the queue that have arrived earlier so they need to wait in queue before being served which causes delay.

Also another important note from figure 7 is that sum of throughput for http and video application never goes more than 4Mb which is the link capacity between Access router and Router_TR.It can be seen in the graph that http and throughput have reverse relation when web traffic decreases video traffic increases and vice versa.

Figure 8 shows that throughput between Access router and TR_router goes to more than 4Mbps,The logic behind this phenomena is that, video traffic that has the constant rate less than link between Access router and TR_router is able to use all its packets when there is not much http traffic, but at the time that web traffic rises total traffic rises more than 4Mbps, but all the traffic is not passed and some packets are kept in queue (queues are big and no drop packets), so again when web traffic decreases video packets are sent again, that is also the reason for delay difference.
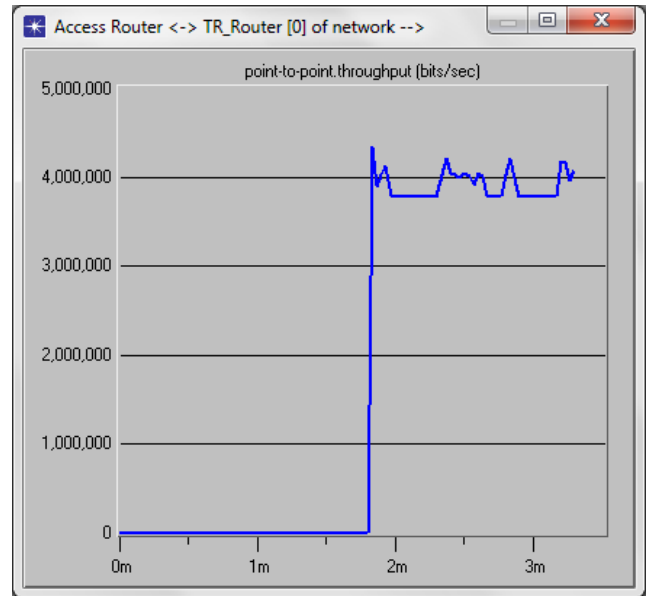

Figure 8. Throughputs between Access Router and TR_router.

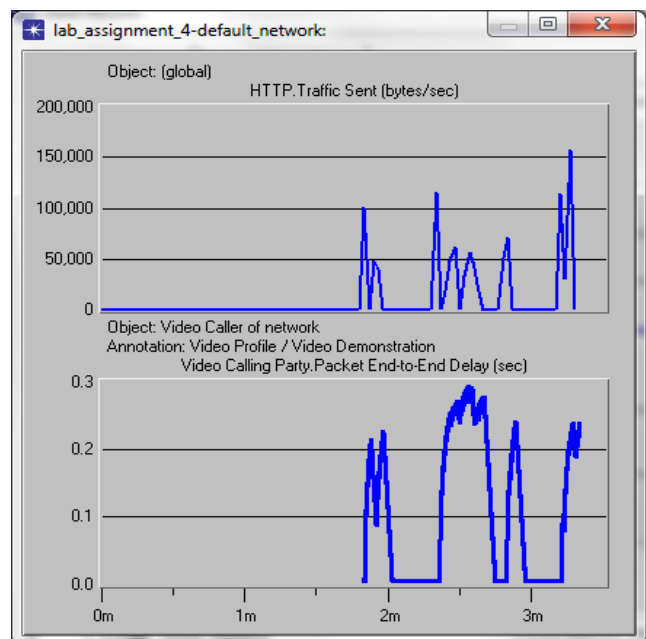### 4.1.2. End to End Delay and Delay Variation (jitter)

Figure 9. HTTP traffic sent and Video caller End to End Delay.

The lower part of Figure 9 is the End_to_End delay for video application, it can be seen that delay varies from 5ms to around 0.3 seconds, and a comparison of the delay graph with HTTP traffic graph shows that minimum delay relates to the time when http traffic equals zero that's because video packets are sent through the router immediately after they reach it. But when http traffic is generated as well, video packets have to wait in the FIFO queue of the router if they arrive later than http packets and that causes delay.

Figure 10 shows that delay variation increases for video packets that's because, the end to end delay changes for video packets according to the http traffic generated.
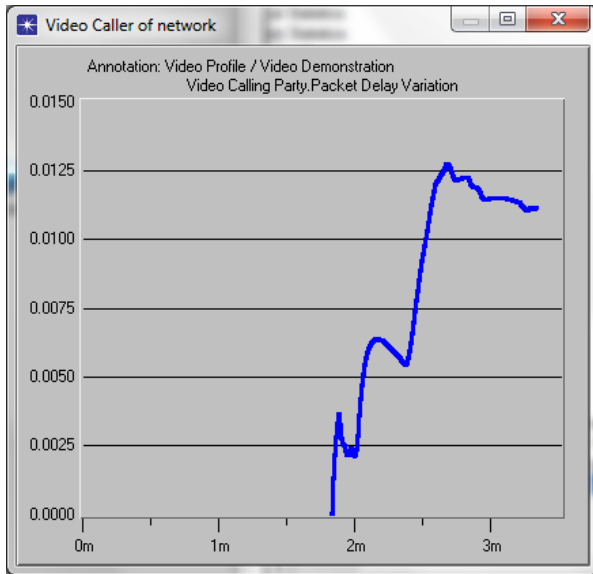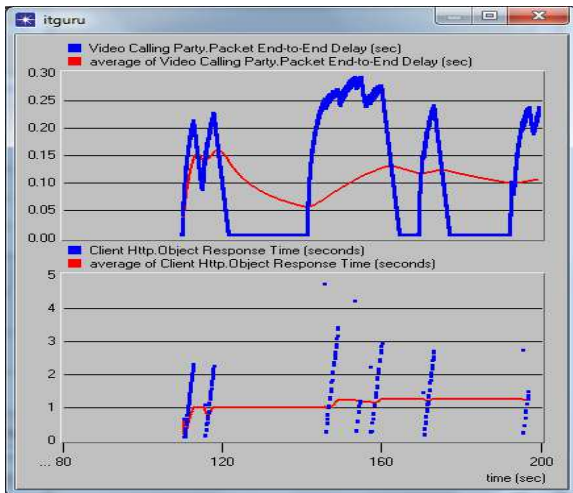
Figure 10. *Delay variation for Video Caller.*

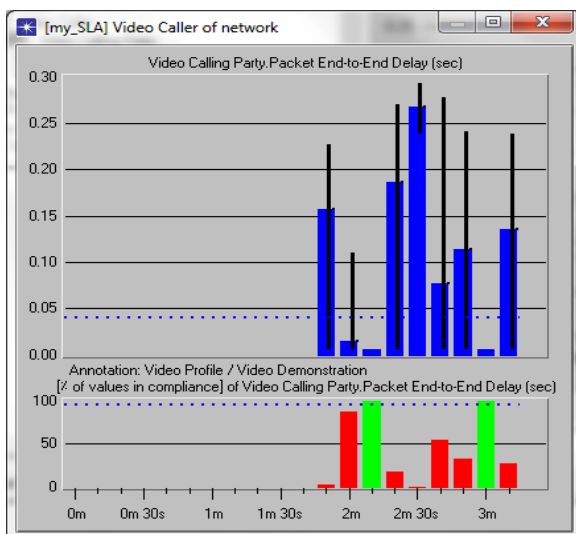**Figure 11.** *Video packet end_to_end delay and http object response time.*

Figure 12. *Video calling packet end-to-end delay.*

Figure 11 shows that http response time gets higher during the requests. Since there are 101 objects that are sent each time from the server with different large sizes and later sent objects face more time for being sent because it should wait for the first ones to be sent.
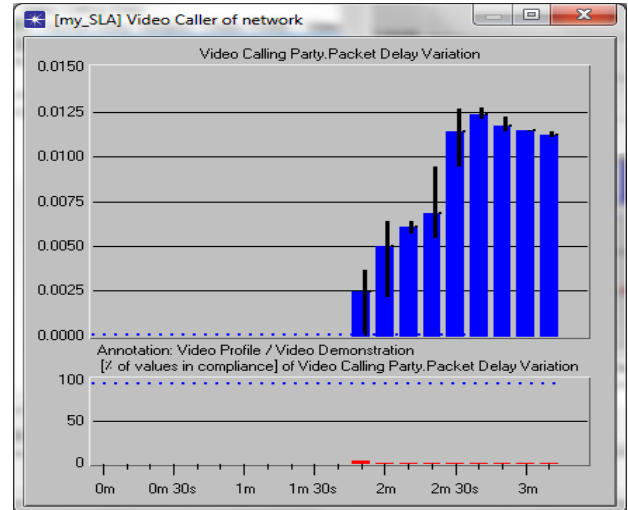

Figure 13. *Video calling packet delay variation.*

Figure 13 and Figure14 show the video packet end_to_end delay and delay variation (jitter), also the green and red bars at the lower part of the graphs show the recorded values that meet the SLA requirements. The width for each is 10 seconds and the black ledges show min and max value. It can be inferred from the figures that both delays vary a lot and delay variation does not meet SLA at all and only two bars for end_to_end delay meet the SLA which relates to the time that there is no http traffic.

What happens in access router during default scenario is that the FIFO queuing policy is implemented so the sooner a packet arrives (no matter it is web or video) the sooner it goes out.
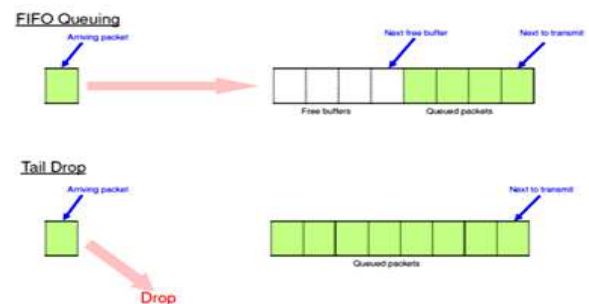

Figure 14. *FIFO policy [7].*

4.2. With_CAR_Scenario

In this scenario CAR functionality is examined as the QoS providing parameter. For providing the ability of bandwidth management in CAR, a mechanism is used called token bucket. As explained in configuration part, 3 attributes are set for traffic rate. Rate limit, conforming limit and exceeding limit.Rate limit includes average rate, normal brst and exccess

burst size. What committed access rate does is that it controls the traffic and its burst with these three values using the token bucket mechanism. What happens in this mechanism is that a token is defined with an arriving rate (average rate) to the bucket. The bucket has a size (normal burst size).The bucket will receive tokens until it's full. A traffic can pass (is conformed) if when arriving to the bucket founds enough tokens (according to the size of its packets). If traffic arrives at the bucket and see that there are not enough tokens, then a procedure as following is defined to bring some extending facility considering the extended burst size.

When traffic arrives needing number of tokens to be conformed, CAR makes a calculation of a value called compounded debt, which relates to Actual debt that is the number of token lent. Compounded dept is the summation of all actual debts for all the packets that did the borrowing procedure. If the Compounded debt is more than the extended burst size then the exceed action happens. The tokens that arrive are first used for clearing the borrowed tokens. [8]
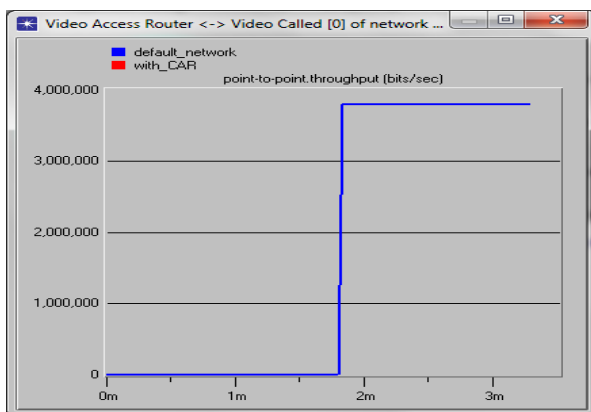
4.2.1. Throughput

Figure 15. *Throughput between Video Called and Video_Access_Router in default and with CAR scenario.*

Figure 15 compares throughput for Video application in default scenario and in with_CAR_scenario which shows that for both the throughput is same.
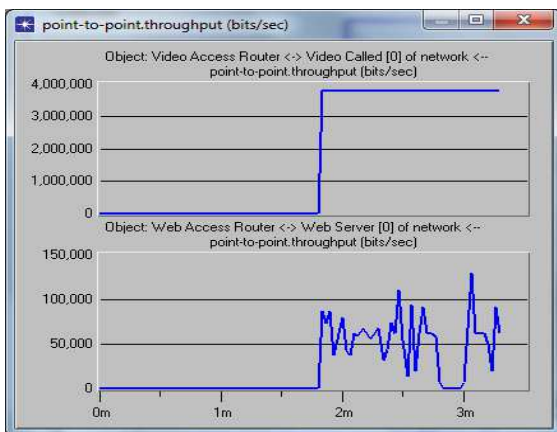

Figure 16. *Throughput between Web Server and Web_Access_Router and video Called with video access router in with CAR scenario.*

Figure 16 compares throughput for web application and video application. This graph shows that http throughput for with_CAR_scenario has dropped considerably also in default scenario http throughput goes down to zero continuously but this is not the case in with_CAR_scenario.

The explanation behind these phenomena is that, in default scenario video and http traffic are passed through the link when there is enough bandwidth or are kept in queue buffer so no packet loss happens and the zero throughput belongs to the time when http objects are not sent. But in with_CAR_scenario because the token buckets are used and as explained in configuration part they are set to size of 50kbits, if http traffic is more than 50kbit, packet drop will happen. Considering the fact that http uses TCP as transport protocol, after packet drop retransmission is needed. That's why throughput does not go to zero as often like default scenario and also according to TCP policy for changing the congestion window size the sent rate decreases and that causes the throughput to drop as well.

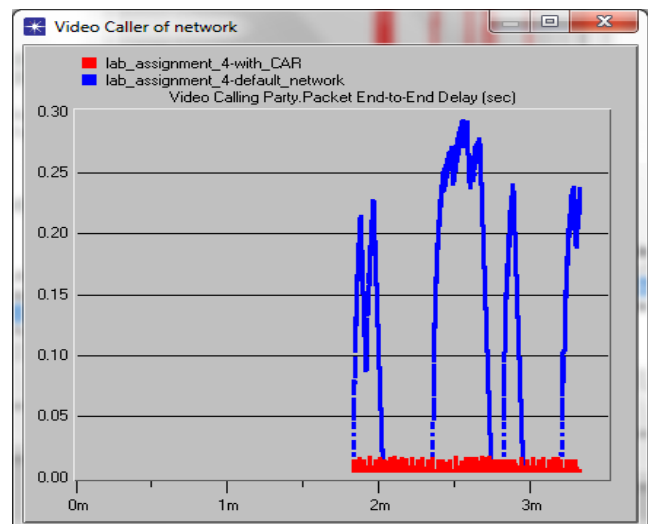### 4.2.2. End-to-End Delay Variation

Figure 17. *End to End video packet Delay in default scenario and with_CAR_scenario.*

Figure 17 shows that end to end delay for video packet has decreased a lot. The reason is that by implementing CAR functionality on access router, CAR implement limitations on http traffic and therefore more bandwidth is assigned to video traffic which causes less delay because the link does not act as a bottleneck anymore. But still some video packets may have to wait in queue of the router so still there is delay.

Figure 18 shows that as discussed before end to end delay (and its highest value) dropped a lot(around 10 times smaller for the max value), but on the other hand response time for http objects has increased a lot because of retransmissions that http traffic has and also the reduce in traffic according to TCP logic, the response time increases a lot the average has changed from 1 sec to the value of 7 and the maximum has reached up to 17 seconds.
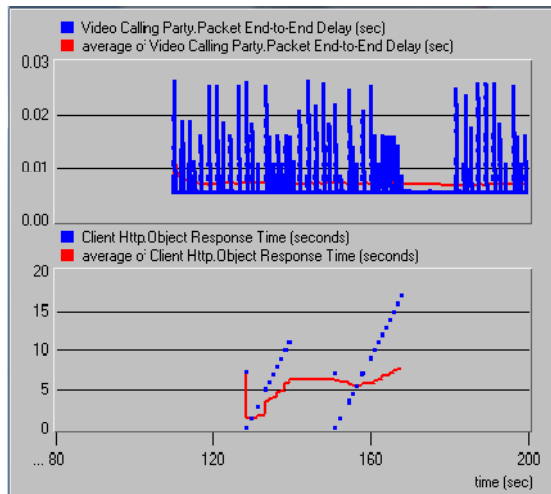
Figure 18. *Video packet end_to_end delay and Client http object response time.*
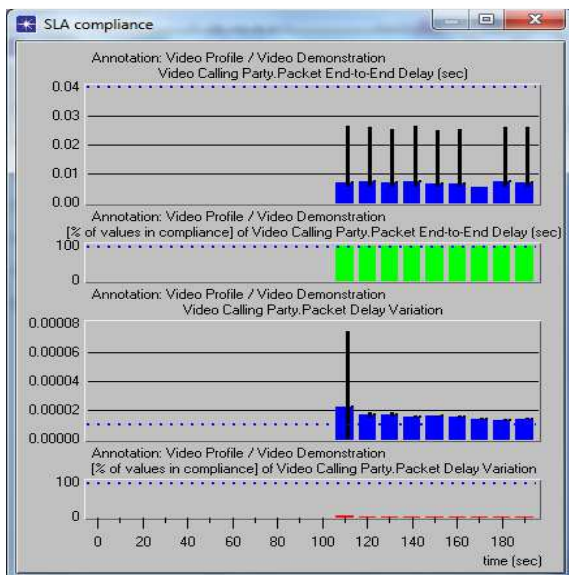

Figure 19. *Video packet end to end and variation delay*

Figure 19 shows that video end to end delay is small enough to meet SLA requirement due to CAR settings and the average is less than 0.04 sec of SLA. But the problem still exists with delay variation. Although it has been reduced but does not meet the SLA requirements because of the FIFO policy that is used at access router which causes some video packets to wait in the queue based on their arrival time.

4.3. With_CAR_and_PQ Scenario

As mentioned before CAR is able to assign IP precedence in packets. Other Devices in the network can uses assigned IP precedence to take decision about how should treated to traffic.

Based on HTTP CAR profile, http traffic that is conformed will be transformed with IP precedence equal to 2 and the Traffic that not conformed will be dropped.

In addition in video CAR profile traffic conformed is transported with IP precedence equal 5 also the Traffic that

doesn't conform is has an IP with precedence 4.

In this scenario a PQ is used in outgoing interface of Access Router. PQ is used IP precedence that CAR set at incoming interfaces. So in counter of previous scenario IP precedence is used in this scenario.

Now the effect of CAR and PQ will be analyzed.

4.3.1. Throughput of Http Service and Video Service

Figure 22 shows throughputs from web server to the web access router and from video called client to video access router.
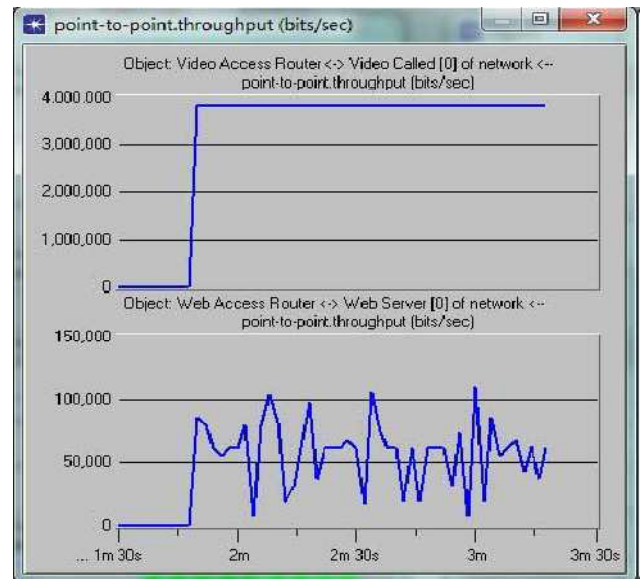

Figure 20. *Throughput of sent http traffic and video traffic in scenario of with_CAR_and_PQ.*

Figure 20 shows Throughput in the link between Web_A_Router and Web_Server. It can be seen the throughput of video is the same with the last scenario. BUT, there is change in throughput of http traffic. It worth to mention that In this scenario, PQ is enabled in outgoig interface of the accesrouter. Based on CAR configure, the video packets are assigned with Multimedia (5), and on exceeding rate limit packets are specified with StramingMultimedia (4) and both of them have the "transmit action" as the policy. In addition in this scenario http packets are assigned with Standard (2) and Background (1), as conforming rate with transmit and as exceedin rate with drop action. Consider to PQ in outgoing interfaces and mentioned policies for packets, there is no video dropped in any case due to high priority of these packets

In fact when PQ is enabled, the access router first transmit the video packets and if there is no waiting video packets, then router will send HTTP packets. As conclusion http packets will meet more packet loss and we have more retransmission in this case.

It worth to mention most of the time peaks value are below of 400 Kbit/s. This happen because the video packets will be seat in a queue with higher priority and only when we have no waiting video packet in queue http packets could be served.

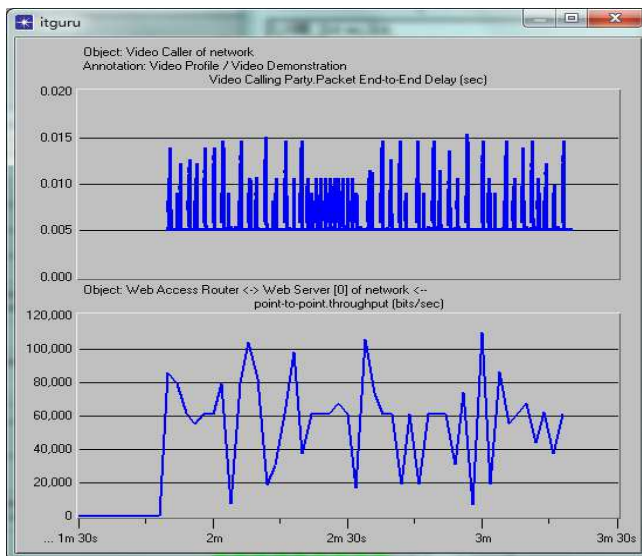### 4.3.2. Video traffic End-to-End Delay Analysis

Figure 21. *Video traffic end-to-end delay and http throughput.*

Figure 21 shows Video traffic end-to-end delay with_CAR_and_PQ scenario. This figure shows that the end delay for video traffic decrease again. The point is that the delay time no longer varies with the changes of http throughput. In fact when PQ has been used in outgoing interface, the router will send video packets when they are exist in queue and the router do not care about that how many http packets waiting in lower priority queue.

4.3.3. SLA Requirements Analysis

In this scenario with classified and remarked data traffic, other QoS feature like PQ can be added. In worth to mention that PQ provides delay guarantee for high priority data flow and it is enabled on outgoing interface of Access Router. Figure 22 and Figure 23 shows the video traffic end-to-end delay and delay variation in with_CAR_and_PQ scenario.
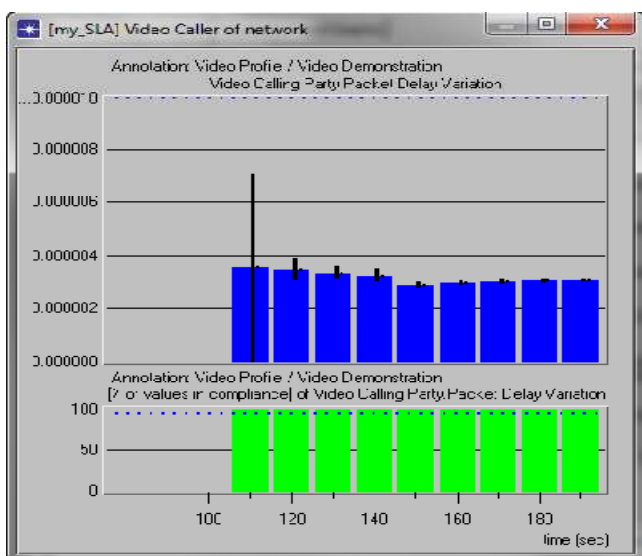

Figure 22. *video packet delay variation (jitter) for video calling in With_Car_PQ scenario.*
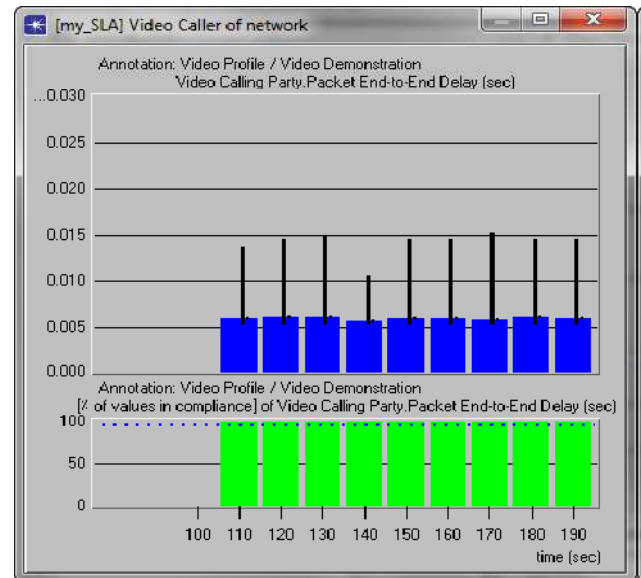

Figure 23. *Video End to End Delay for Video Caller in With_CAR_PQ Scenario.*

Figure 22 and Figure 23 shows, the end-to-end delay values are acceptable based on SLA requirement.

In fact by using CAR, http traffic is limited and more bandwidth is assigned for video traffic. In addition

By using PQ in outgoing interface of access router, delay time is stable and all of the variation values are meet the SLA requirement.

Figure 24 shows video packet delay and object response time, comparing to the result of CAR scenario a minor change (reduce in delay and increase in response time), can be seen. Because in this case video packets that have higher priority are transferred sooner still due to the small amount of http traffic the change comparing to previous scenario is not tremendous.
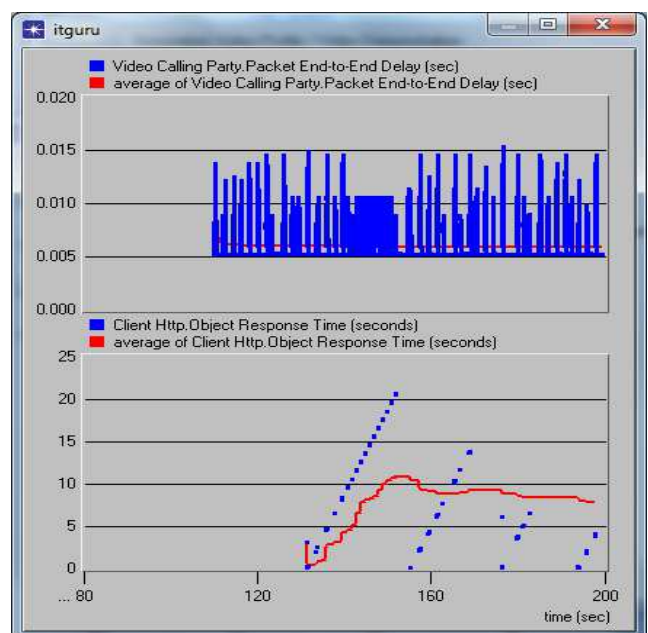

Figure 24. *Http object response time and video packet end to end delay.*

4.4. With RSVP Scenario

4.4.1. Video Traffic End-to-End Delay

Figure 25 show important values of delay for scenario with RSVP.
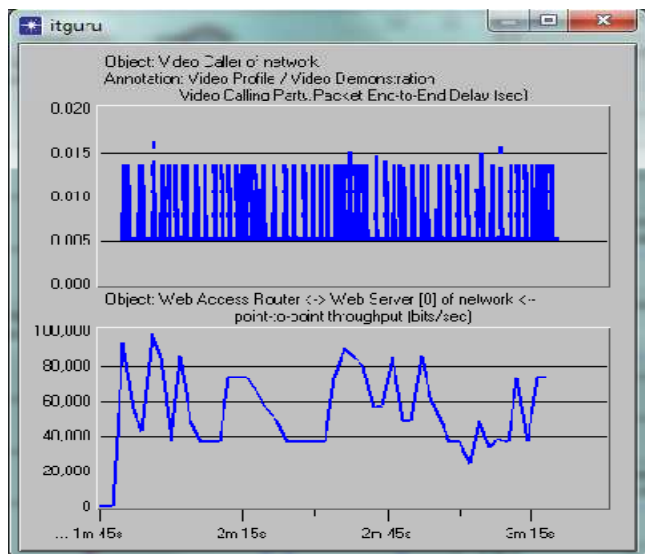

Figure 25. *Video Packet Delay and http throughput for With RSVP scenario.*

By study the results in Figure25, it can be find that video packet delay is low (with the average value of 6 (ms)).the interesting part in this scenario is packet delay is very stable. In RSVP profile, we reserved 3.92 Mbps bandwidth for video data flow. This amount of bandwidth is enough for the throughput nearly 3.8 Mbps. RSVP uses Weighted Fair Queuing (WFQ) method to guarantee this amount of bandwidth.
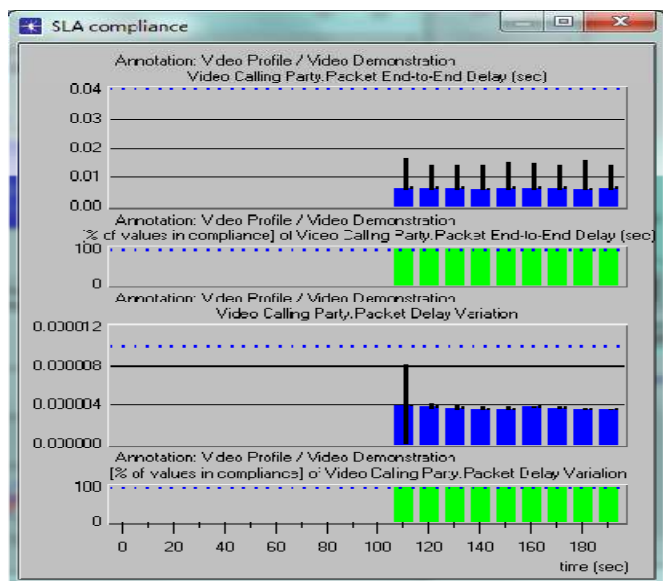
4.4.2. SLA Requirements

Figure 26. *Video Delay and Delay Variation with SLA for and RSVP.*

Figure 27 shows both video delay and video delay variation meet SLA requirements. In fact by using RSVP

profile, video data traffic is guaranteed to have sufficient amount of bandwidth. As result it is clear the delay and delay variation meet the SLA requirement.

The point in this scenario is, RSVP profile wastes a part of bandwidth. In fact RSVP reserve 3.92 Mbps for video data traffic, but the video throughput is only about 3.79 Mbps. As conclusion we have 130 kbps wasted bandwidth in this case.

5. Conclusion

In this report different mechanism of QoS on performance of networks is studied. This report shows Without QoS, the delay-sensitive applications will not be guaranteed with high service quality because there is no policy to serve the packet in router and both sensitive and non-sensitive traffic will treat in same manner.

In continue these study shows each QoS mechanism has its effect on network performance. By applying Committed Access Rate (CAR) on the router, video traffic (as delay sensitive traffic) guarantee its small packet delay. However, since FIFO queuing method is used on the interfaces, other kinds of traffic still causes the delay time varies. By applying PQ this problem also solved. in fact PQ on the interfaces helps video service guarantee its delay variation.

Finally this reports shows by using RSVP video service guarantee both delay and delay variation but this mechanisms wastes some bandwidth.

References

[1] A.Frouzan, Behrouz. TCP/IP Protocol Suite 3rd Ed. Mc. Graw Hill,2003

[2] J.Postel,"Internet Protocol :DARPA Internet Program Protocol Specification,"RFC 791 Sept,1981

[3] Semeria, Chuck. "Supporting Differentiated Service Classes: Queue Scheduling Disciplines". s.l. : Juniper Networks, December 2001.

[4] Quality of Service Solution Configuration Guide, Cisco IOS Release,Cisco Systems Inc,2010.

[5] IP QOS: traveling in first class on the Internet. 2, Mar/Apr 1999, Vol. 3, pp. 84 - 88.

[6] Opnet Simulater help tutarial.

[7] Understanding Delay in Packet Voice Networks. [Online] Cisco. [Cited: march 24, 2015.] http://www.cisco.com/en/US/tech/tk652/tk698/technologies_white_paper09186a00800a8993.shtml.

[8] Transmission Control Protocols (TCP). Behrouz A. Forouzan. TCP/IP protocol Suite [Fourth Edition]. s.l. : McGRAW.HILL, 2010, pp. 432-539.

[9] TRANSMISSION CONTROL PROTOCOL. The Internet Engineering Task Force (IETF). [Online] Information Sciences Institute, University of Southern California, September 1981. [Cited: 05 04, 201.] http://tools.ietf.org/html/rfc793.

[10] Adjusting IP MTU, TCP MSS, and PMTUD on Windows and Sun Systems. [Online] cisco. [Cited: 04 05, 2011.] http://www.cisco.com/en/US/tech/tk870/tk877/tk880/technolo gies_tech_note09186a008011a218.shtml#tcpmss_val.

[11] M.Allman," TCP congestion control.",RFC 2581, April 1999

[12] J.Postel,"Transmission Control Protocol :DARPA Internet Program Protocol Specification,"RFC 793 Sept,1981

[13] Lammle, Todd. Certified Network Associate Study Guide[Sixth Edition],Wiley Publishing, Inc,2007.

[14] Cisco System Inc White Paper accessed @ http://www.cisco.com/en/US/tech/tk652/tk698/technologies_ white_paper09 186a00800a8993.shtml Last Access May 16,2015

[15] Resrvation Protocol White Paper Accesd @ http://www.networksorcery.com/enp/protocol/rsvp.htm Last acces June 6,2015

[16] OPNETWORK Quality of Service for assuring SLA accessed @ http://opalsoft.net/qos/RSVP-12.htm Last access June 14,2015