Appeared in the *Proc. of 14th National Conf. on Communications, NCC-2008*, IIT Bombay, 1-3 Feb 2008.

QoS in Ethernet Access Ring

G. Sathish Kumar,Timothy.A.Gonsalves
TeNet Group, Department of Computer Science and Engineering
Indian Institute of Technology Madras, India
Email:{gsatish,tag}@lantana.tenet.res.in

R.Thirumurthy
Midas Communication Technologies Pvt Ltd., Chennai
Email: rtm@midascomm.com

***Abstract*—Ethernet is being extended to Metro Area Networks (MAN) mainly due to its simplicity and low cost. However Ethernet protocols need extensions to succeed as MAN technology in terms of scalability, Quality of Service (QoS), resiliency and Operations, Administration and Maintenance. The QoS mechanism existing in Ethernet is localized to each node and affected by the spatial properties. Hence in order to enhance the QoS level in Ethernet, a distributed QoS mechanism is required which ensures fairness and bandwidth guarantees. We propose one such approach for Ethernet Access Ring in this paper. The mechanism ensures global fairness in the access ring that gives every node in the ring a fair share of the ring bandwidth. The QoS solution discussed here is implemented in a commercial product M2404, a Metro Ethernet Switch developed by Midas Communication Technologies and analysed in a ring network. The measurements show that the ring bandwidth is fairly shared irrespective of the spatial properties of the nodes.**

I. Introduction

Ethernet has proven itself in the local area networks (LAN) as a layer-2 transport technology. Ethernet is entering into the metropolitan area network [1]. The Ethernet when extended into access network is typically as shown in Fig 1. The multiple Ethernet access rings connecting the Remote Terminals (RT) are terminated in the Central Office Terminal (COT) switch and connected to the Network Service Provider's (NSP) Gateway router. The COT switches in turn can be interconnected by transport technologies like Resilient Packet Ring (RPR) [2], Multi Protocol Label Switching, Gigabit-Ethernet or Ethernet Over Sonet (EOS). The access rings can also be RPR, MPLS, Gigabit-Ethernet or Ethernet Over Sonet.

The RPR/MPLS/EOS solution in the access ring is expensive due to the additional hardware required and also, the overhead of conversions between different packet formats for these transport protocols consume a fair amount of bandwidth. Gigabit Ethernet in the access ring is a low cost and less complex solution for small-scale deployments. But the Gigabit Ethernet solution is weak in terms of providing SLA guarantees. VLAN Stacking is used to address the scalability issue with VLAN [3] and Rapid Spanning tree protocol to address the resiliency issue when Gigabit Ethernet is extended into access network. The Ring topology is chosen mainly because it is most prevalent in metro networks today and also for its protection and fault tolerance properties.

II. Overview

In Section III we discuss the QoS issue in Ethernet Access Ring and the proposed solution. Section IV gives the result of the analysis done with four node Ethernet ring without QoS and with the QoS mechanism applied. In Section V and VI we close with conclusion and the future work.

III. QOS Issues in ethernet Access ring

The QoS mechanism in Ethernet is localized within each node and also influenced by the topology. Currently native Ethernet supports QoS by means of Priority Marking [5]. Priority Marking uses the 3 Class of Service (CoS) bits in the VLAN header for preferential treatment of traffic. Priority marking does provide differentiation between different traffic types but it fails to give bandwidth guarantee and fairness.

A. Proposed Solution

In the Ethernet access ring shown in Fig 1. where multiple Remote Terminals are connected in a ring topology the customer packets are tunneled using VLAN Stacking [6]. The Outer tag (SVLAN) is used to identify the Ethernet Virtual Circuit (EVC) within the Ethernet access ring. The scope of the Quality of Service mentioned here is within the access ring. When the packet crosses the access network to core network there has to be a mapping of QoS parameters between the access and core network. In the Ethernet access ring the main challenge in providing QoS involves ensuring fairness. Ethernet is a broadcast medium, either the transit traffic or the injected traffic can unduly monopolize the available bandwidth.

In our bandwidth management scheme, to ensure QoS and fair usage of the ring bandwidth the parameters to the ingress policer and the egress scheduling are set in such a way that the free excess bandwidth in the committed class is distributed to the best effort class in a weighted way. In the spanning tree (RSTP,STP,MSTP) enabled ring the parameters (weights for egress) are exchanged using a proprietary Generic Attribute Registration Protocol (GARP) [4] application, which sets the QoS parameters in all nodes in the ring. The shaded blocks in Fig 2 requires changes in order to achieve this. The following are required operations in each building block of QoS in the proposed solution.

Classification is based on the Customer VLAN. Based on the classification the service tag (SVLAN) is inserted and the traffic contracts [10] (CIR, CBS, EIR, EBS) the flow will undergo is decided. In all the uplink ports connected in the ring the remapping of injected traffic group to transit traffic group is done in order to provide high priority to the transit traffic once it leaves each node and enters the ring. This will ensure that transit traffic is not overwhelmed by the injected traffic from other nodes. The weights among transit and injected group of queues are assigned in a fair manner.

Policing as per the traffic contract, this follows single rate two color marking or two rate three color marking algorithm. The policing will mark all the traffic within CIR to committed traffic class and exceeding CIR and below EIR to best effort traffic class.

Buffer management will ensure that congestion is avoided by setting a bit for all packets exceeding CIR and below EIR to mark high drop precedence at the time of congestion.

Scheduling/Shaping at the egress queue at all the nodes in the ring. Here two groups of traffic classes are created in the egress queues as shown in Table I. One for the injected (G-II) and the other for the transit traffic (G-I) to avoid unfairness. The weights for this traffic classes are set such that CIR is always assured and when CIR is unused the excess is taken by best effort traffic. This is achieved by having a minimum weight and maximum weight for each egress traffic class. The minimum weight given for a class of traffic is the one that is assured even under fully loaded condition whereas maximum is achievable only under lightly loaded condition. The ring bandwidth (R_{bw}) is first budgeted between two classes with the ratio Y:(100-Y).

$$CIR_{Total} \leq (Y/100)\ x\ (R_{bw})$$

$$EIR_{Total} >> ((100 - Y)/100)\ x\ (R_{bw}) \qquad (1)$$

CIR_{Total} ,EIR_{Total} are the total CIR and EIR provisioned in the ring.

Let $G\text{-}I_J$ be the minimum bandwidth allocated for Group - I set of queues at node *'J'* which will have high priority compared to Group -II set of queues at the same node *'J'*. All

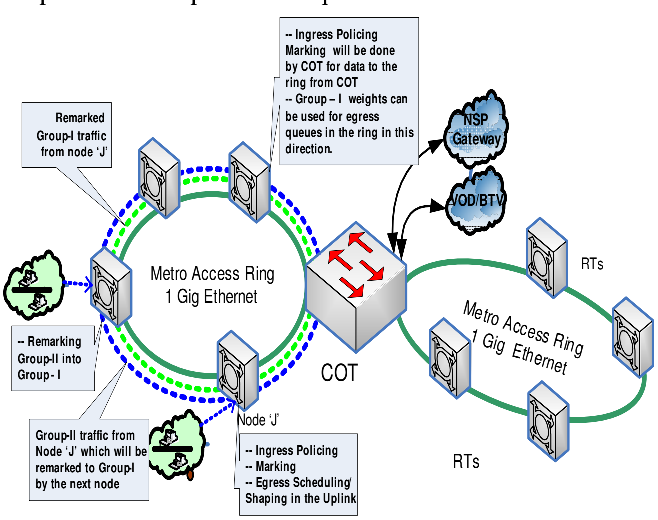

Fig. 1. Metro Ethernet Ring with Injected and Transit Traffic in different groups.

TABLE I
TRAFFIC CLASS MAPPING AND SCHEDULING SCHEME

Service Class	COS	Application Mapping	Scheduling Scheme
Premium (G-I,Committed)	6,7	Network Control	Strict Priority
Gold (G-I,Committed)	4,5	Voice/Video/EPL (Transit Traffic)	Strict Priority
Gold (G-II,Committed)	2,3	Voice/Video/EPL (Injected Traffic)	Strict Priority
Standard (G-I,Best Effort)	1	Best Effort (Transit Traffic)	Fair Queuing
Standard (G-II,Best Effort)	0	Best Effort (Injected Traffic)	Fair Queuing

the ring transit traffic will pass through Group-I set of queues.

$$G\text{-}I_J = R_{bw} - (CIR_J + (((100 - Y)/100)\ x\ R_{bw}\ x\ (EIR_J/ EIR_{Total})) \qquad (2)$$

Let $G\text{-}II_J$ be the minimum bandwidth allocated for Group-II set of queues at node *'J'* which is used by the injected traffic from the node *'J'*. Group-II set of queues are for the injected traffic.

$$G\text{-}II_J = CIR_J + (((100 - Y)/100)\ x\ R_{bw}\ x\ (EIR_J/ EIR_{Total})) \qquad (3)$$

CIR_J, EIR_J are the CIR and EIR provisioned in the node *'J'*.

There are two different classes of services within each group of queues. Premium and Gold class together falls in the committed bandwidth class. Premium class is for all control traffic in the ring. Standard class falls in the best effort bandwidth class. The minimum and maximum bandwidth allocated for each of the queues in the two groups at node *'J'* is as below. The maximum bandwidth for the standard class is limited by the total ring bandwidth R_{bw}.

Group – I Queues (Transit Traffic)

Gold Class Traffic (low latency)

Minimum,Maximum:

$$C^I_B = (CIR_{Total}) - (CIR_J)$$

Standard Class Traffic (high latency)

Minimum:

$$C^I_C = ((100-Y)/100)xR_{bw} - (((100-Y)/100)xR_{bw}x(EIR_J/EIR_{Total}))$$

Maximum:

$$R_{bw} \qquad (4)$$

Group – II Queues (Injected Traffic)

Premium/Gold Class Traffic (low latency)

Minimum,Maximum:

$$C^{II}_B = (CIR_J)$$

Standard Class Traffic (high latency)

Minimum:

$$C^{II}_C = (((100 - Y)/100)\ x\ R_{bw}\ x\ (EIR_J/ EIR_{Total}))$$

Maximum:

$$R_{bw} \qquad (5)$$

C_B, C_C are the minimum bandwidth committed to each class of traffic and the maximum bandwidth for standard class can go up to the link capacity under unutilized conditions.

Admission control is done having an application running over GARP. The main aim of this distributed application is to maintain the bandwidth availability information on the path to the COT switch. The proprietary GARP application exchanges the resource availability information consisting of attributes Node Id, Reservation Id, Traffic Class, Weight, EIR_{Total}.

Node Id : Mac address of each node in the ring.

Reservation Id: Identifier used to differentiate a GARP Update and Refresh message.

Traffic Class: Class of service

Weight : Bandwidth reserved for the particular traffic class.

EIR_{Total} : The total Best Effort Load provisioned in the ring.

At the edge as traffic enters the port, the traffic is policed at the edge with the Committed Information rate (CIR) and Excess Information rate (EIR) and classified into Committed and Best traffic class. The committed portion of the ring bandwidth is an assured portion of the total ring bandwidth which each allocated node can use whenever required, whereas the Best effort ring bandwidth can be used only to their minimum limit budgeted when the ring is fully loaded and can be fully utilized when the ring bandwidth is idle. The maximum amount of committed bandwidth is restricted by the total ring capacity, whereas the total best effort traffic provisioned can exceed the ring capacity.

At all the egress queues towards COT the transit traffic is given high priority compared to the injected traffic which will ensure that the transit traffic pass through without much delay. This is achieved by remapping to a traffic class of different group once the injected data leaves each node and enters the ring. For the data traffic from the COT there need not be a differentiation between the transit and injected traffic as only one node (COT) can inject traffic to the ring. So a single group will suffice in this direction. The scheduling schemes work by providing minimum bandwidth to all class and then distributing the remaining bandwidth up to the configured maximum based on priority. The egress parameters are set on the path to COT by the proprietary GARP application at the time of instantiation of new services.

The QoS GARP information propagation context will contain all uplink ports within the same ring to avoid propagating the parameters to other Ethernet rings connected by the COT switch. The metro access ring with the bandwidth management and prioritization scheme is as shown in Fig 1.

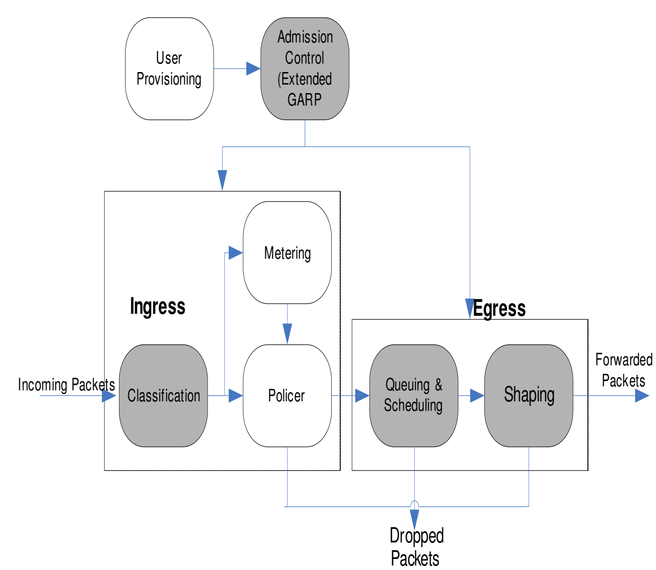

Fig. 2. QOS Building Blocks

IV. Evaluation of QoS in Ethernet Access Ring

The analysis was done in an experimental setup consisting of four metro Ethernet switches (refer Fig 3). The uplink ports of each switch are connected in a ring topology. The bandwidth of all the uplink ports in the switch is set to 100Mbps. Rapid Spanning Tree Protocol is enabled in all the uplink ports in the ring. The Switch-4 will act as the Root Bridge connecting to the Network Service Provider. The RSTP priorities are configured in such a way that the link between the Switch-1 and Switch –2 will be the alternate backup ports. The subscriber ports are connected to a tester to pump data traffic. The measurements are taken by capturing data from the Switch-4 back to the tester.

An Agilent N2X data traffic generator/tester [12] was used to study how the throughput and latency varies among different traffic streams. Traffic streams can be created in each port in the tester to pump data traffic. The size, data rate and data pattern for each of the traffic stream can be varied for the data traffic generated from the tester. Throughput and latency can also be measured for each of the traffic streams in the tester. Ethernet Line (E-Line) [9] and Best Effort traffic streams were created and analysis was done under two scenarios. In the first scenario no QoS policy is applied to the Ethernet Line (E-Line) stream in the Ethernet ring due to which the Ethernet Line traffic from each node is affected by the Best effort traffic. In the second scenario the QoS policy is applied for the Ethernet Line traffic and observations are made. Two traffic streams one of which is Ethernet Line and the other Best Effort stream is pumped from each node in the ring. The data traffic was pumped from the tester and the results for each traffic stream were measured from Switch -4. The size of the packet pumped is 64 bytes and the traffic pumped was at a constant rate.

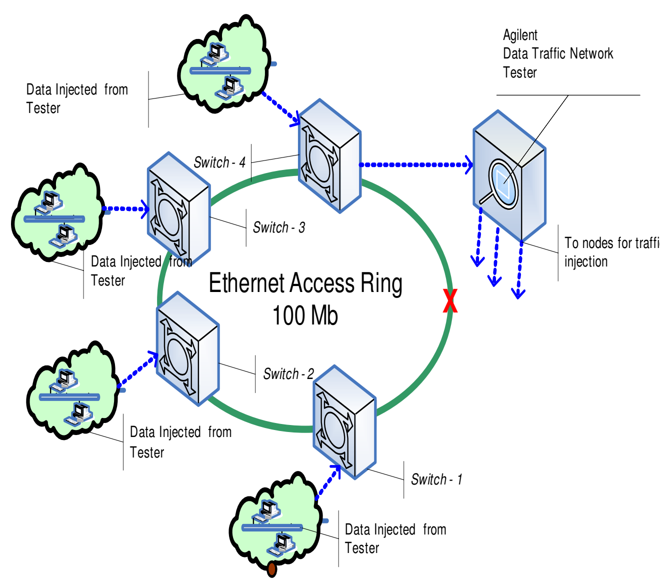

Fig. 3. QOS Test Bed.

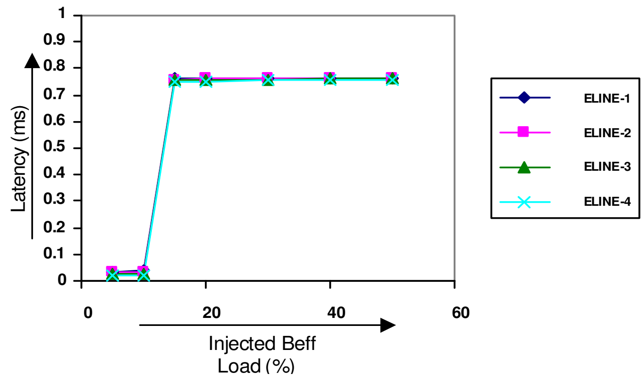

Fig. 4. Variation of Latency with total Injected Load in the ring, no QoS

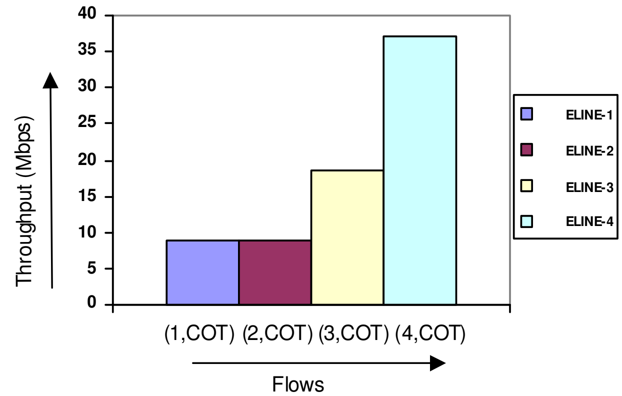

Fig. 6. Variation of throughput for different flows, no QoS.

A. *Without QoS applied in the Ethernet Access Ring*

As a baseline, we first measure performance without QoS. In this setup no QoS policy was applied to the Ethernet Line traffic stream. ELine-x is a stream from node 'x' in the Ethernet ring. Best Effort traffic stream's load was varied and the average end to end delay and receive throughput were measured for each of the traffic streams from the Switch -4.

As the Best effort load increases the latency also increases for all the traffic streams (refer Fig 4). The latency of all traffic streams shoots up when the ring is fully loaded. The Ethernet Line traffic stream's latency shoots up due to the Best Effort traffic load, as there is no prioritization between these streams when no QoS policy is applied. The latency is same for both Best Effort traffic and ELine traffic load. The latency for each stream is dependent on the total load in the Ethernet access ring.

Fig. 5 shows the impact of Best Effort traffic load on Ethernet Line traffic throughput. As the Best Effort load increases the throughput of the Ethernet Line traffic stream comes down which is not desired.

There are four different traffic streams towards the COT in the test setup, one each from all the nodes in the ring. The throughput received by each of the traffic streams towards COT is measured. The traffic streams from Switch-1 and Switch-2 receives only 1/8th of the ring bandwidth under full load. Traffic streams from Switch-3 receives 1/4 th of the ring bandwidth whereas Switch-4 receives 1/2 of the ring bandwidth due to its spatial proximity. This unfairness is depicted in Fig 6.

B. *With QoS applied in the Ethernet Access Ring*

In this setup a QoS policy containing the parameters for the ingress policer is applied to each of the Ethernet Line traffic streams. This satisfies the admission control rules discussed above. The admission control will ensure that the service fits before initiating it. The minimum and maximum weights for each of the egress queue are set as discussed by the proprietary GARP application.

The Ethernet Line traffic stream falls into the committed class traffic. ELINE-x is the traffic stream from node 'x' that falls into the committed traffic class. BEFF-x is the traffic stream from node 'x' that falls into the best effort traffic class. The average end to end delay and receive throughput were measured for each of the traffic stream.

Fig. 7 depicts the impact of the injected data load on latency. As the load increases the latency increases only for the all the Best Effort traffic stream from each node and Ethernet Line traffic streams latency does not change appreciably when QoS is applied to the Ethernet Line flow. The latency remains same for all traffic stream till the total load in the ring is less than the full ring bandwidth. Once the total load reaches the ring bandwidth the latency increases for the Best Effort traffic stream from each node. Fig. 8 shows how the Ethernet Line traffic streams throughput is unaffected even when Best Effort traffic load is increased. Fig. 9 show how the ring bandwidth is

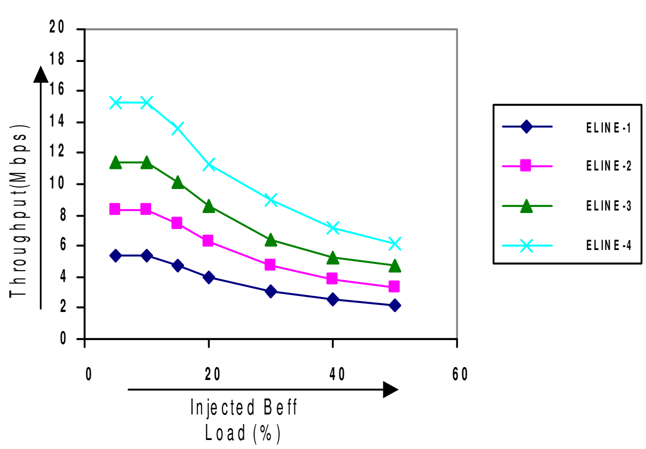

Fig. 5. Effect of injected Best Effort load on the throughput of ELINE traffic, no QoS

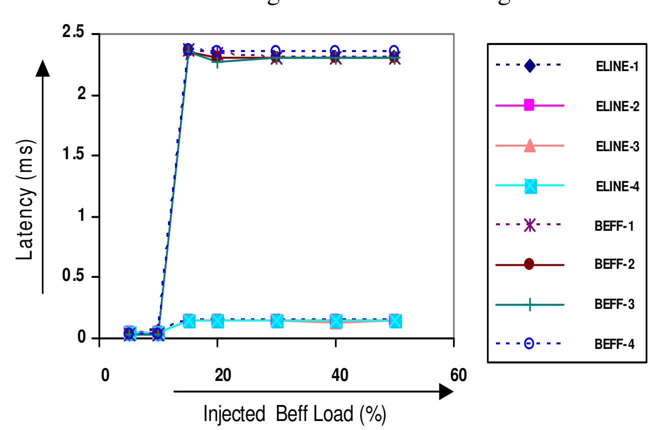

Fig. 7. Variation of latency with total injected load in the ring, with QoS applied.

shared in a weighted way by the different flows under full load when QoS is applied. The Best Effort throughput achieved is the minimum committed for each node in the ring.

Fig. 10 shows how the throughput of the Best effort traffic load comes down as the committed traffic class load increases. To prove that ring is fully utilized when there is no committed class load, the Best effort traffic was pumped at full load and total throughput received by the Best effort traffic adding the Ethernet overhead is 100 % of the ring bandwidth. Now when the committed class load was increased gradually, the Best effort throughput gradually comes down and stabilizes when the total throughput of Best effort class reaches the minimum guaranteed for the Best effort portion. The analysis results shown above will scale up to maximum diameter supported by STP standards.

V. Conclusion

We have proposed a QoS mechanism for Ethernet Access ring. Using the measurements we show that, without QoS the latency is very high and throughput is unfairly shared among streams and by applying QoS we achieve traffic guarantees (Guaranteed throughput (CIR) and bounded delay) for committed class of traffic. By applying QoS we achieve minimum guaranteed throughput for each node under full load for Best Effort class of traffic. Best Effort class utilizes full

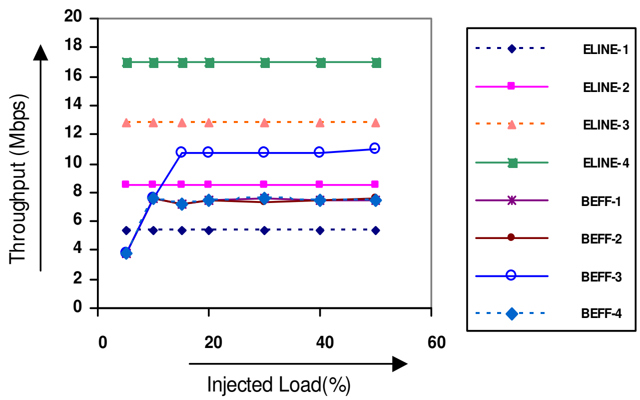

Fig. 8. Effect of total injected Best Effort Load on the throughput of ELINE traffic, with QoS applied.

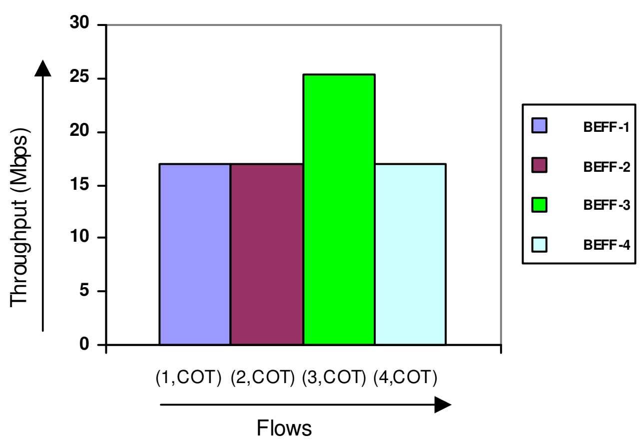

Fig. 9. Variation of throughput of different traffic streams at each node in the ring, with QoS applied.

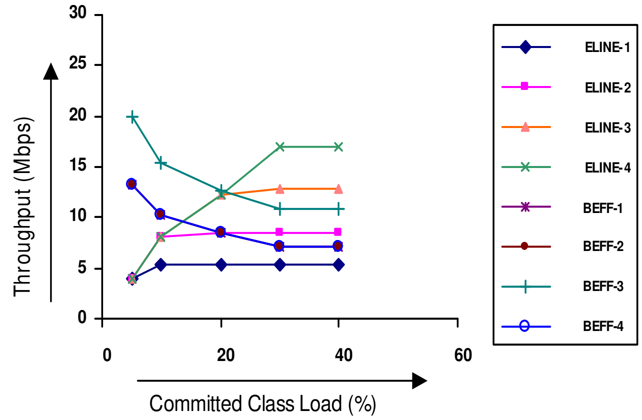

Fig. 10. Variation of Best Effort throughput as committed class load is increased, with QoS applied.

ring bandwidth when no committed class load is present. The fairness for the Best Effort traffic is maintained distributed across the network based on the weights and is not affected by the spatial properties.

VI. Future Work

The analysis can be extended to more number of nodes and also when a node/link fails the restoration time with respect to QoS parameters can be measured by varying the number of nodes.

References

[1] Girish Chiruvolu, An Ge, David Elie Dit Cosaque , “Issues and Approaches on Extending Ethernet beyond LAN’s”, IEEE Communications Magazine,pp.80-86, March 2004.
[2] Fredrik Davik, Mete Yilmaz , Steln Gjessing , Necdet Uzun , “IEEE 802.17 Resilient Packet ring Tutorial”, IEEE Communications, pp.112-118,March 2004.
[3] IEEE 802.1Q/P, “Virtual Bridged Local Area Networks”, 1998.
[4] IEEE 802.1D , “Media access control (MAC) bridges”, 1998.
[5] “Delivering Hard QoS in Carrier Ethernet Networks”, White Paper, www.atrica.com, 2005.
[6] “Encapsulation schemes to extend Ethernet to Metropolitan Area Networks”, White Paper, 2004.
[7] Ralph Santitoro, “Metro Ethernet Services: A Technical Overview”, www.metroethernetforum.org, 2004.
[8] Ethernet Services Model, Metro Ethernet Forum Technical Specification, 2004.
[9] Ethernet Services Definitions, Metro Ethernet Forum Technical Specification, 2004
[10] Traffic Management Specifications, Metro Ethernet Forum Technical Specification, 2004
[11] Metro Ethernet Forum, http://www.metroethernetforum.org.
[12] “Agilent N2x Tester”, http://www.agilent.com/comms/N2X.