

# Benchmarking QoS on Router Interfaces of Gigabit Speeds and Beyond

Javier Orellana  
[jo@hep.ucl.ac.uk](mailto:jo@hep.ucl.ac.uk)  
[www.hep.ucl.ac.uk/~jo](http://www.hep.ucl.ac.uk/~jo)

Andrea Di Donato  
[add@hep.ucl.ac.uk](mailto:add@hep.ucl.ac.uk)  
[www.hep.ucl.ac.uk/~add](http://www.hep.ucl.ac.uk/~add)

Frank Saka  
[fs@hep.ucl.ac.uk](mailto:fs@hep.ucl.ac.uk)  
[www.hep.ucl.ac.uk/~fs](http://www.hep.ucl.ac.uk/~fs)

Peter Clarke  
[clarke@hep.ucl.ac.uk](mailto:clarke@hep.ucl.ac.uk)  
[www.hep.ucl.ac.uk/~clarke](http://www.hep.ucl.ac.uk/~clarke)

University College London  
Department of Physics & Astronomy  
Gower Street  
London, WC1E 6BT  
United Kingdom  
July 14, 2003

**Abstract -- In this paper we present results of Quality of Service (QoS) performance evaluation of router interfaces. We use the equipment of two main manufacturers: Cisco and Juniper. The technologies used are of Gigabit per second and above, namely Gigabit Ethernet and 2.5 Gbit/s OC-48 POS. We examine the performance of the scheduler when we enable Differentiated Services and applying different policies to allocate bandwidth for two classes, one with a lower priority.**

**Keywords -- QoS, Performance, Cisco, Benchmarking, Juniper, DiffServ, Router Interfaces, Gigabit interfaces, Gigabit Ethernet, OC-48 POS.**

## I. INTRODUCTION

Extensive research and papers have been published regarding different solutions to implement QoS [ 1], [ 2], [ 3]. Some of these solutions have already been described in well-known standards, but there are few papers presenting results about the performance for different router interfaces and the implementation of Quality of Service (QoS).

In this report, we present the results obtained from the equipment of two manufacturers: Cisco [ 4] (the OC-48 interface on the 7609) and Juniper [ 5] (the Gigabit Ethernet interface on the M10). The aim is to provide the standalone performance measurements of these devices from the QoS point of view. The results presented here will ultimately be used to understand the composite behaviour when such devices are placed in wider network. Our goal is not to compare and contrast products of different manufacturers (which is not possible in this case since we are talking about different technologies) but to provide performance results for leading edge QoS capable off-the-shelf commodity products.

### A. Background

This paper is based on the current work being done for two projects: MB-NG [ 6] and DataTAG [ 7].

The Managed Bandwidth Next Generation (MB-NG) is a UK based “e-science” project. The aims are firstly to demonstrate end-to-end managed bandwidth services in a multi-domain environment, in the context of Grid project requirements. Secondly, to investigate and develop high performance data transport mechanisms for Grid data transfer across heterogeneous networks. Specific applications we have in mind, among others, are “RealityGrid” [ 8] and High Energy Physics experiments such as BaBar [ 9].

The goal of the DataTAG project is to create a large-scale intercontinental testbed for data-intensive Grids. The focus is mainly on the network research over a high-performance dedicated 2.5 Gbps circuit between CERN in Geneva (Switzerland) and Starlight in Chicago (USA).

### B. QoS

The deployment of QoS—the separation and unequal treatment of different traffic flows based on the application requirements and the agreement between different administrative domains—is a significant part of the MB-NG and DataTAG projects.

For the deployment of QoS and defining sensible Service Level Specification (SLS) and Service Level Agreement (SLA), it is important that the network behaviour is quantified and understood. QoS model we use is based on the Differentiated Services (DiffServ) [ 10] model for IP networks. The traffic entering the network device is marked using a single DS codepoint (DSCP). For each one of these codepoints there is assigned a different behaviour aggregate.

For the tests presented here, our objectives were to obtain the performance limits of the routers in standalone mode. We are interested in the maximum throughput for each class, paying special attention to the router’s scheduler, looking how it treats the different aggregates.

The rest of this paper contains: the setup we use for our performance evaluation, the results obtained for the Cisco and Juniper routers and finally some conclusions.

## II. TEST SETUP

Figure 1 shows the generic testbed we used to measure the performance of each single device and router interfaces.

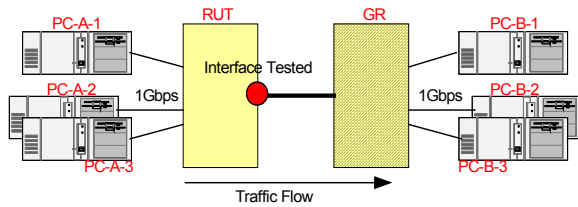


Figure 1 Generic Testbed

Three PCs (Supermicro 6022P-6 Dual Intel® Xeon [ 11]) were attached to the routers. Each PC had an Intel® PRO/1000 XT Server Gigethernet (Gigabit Ethernet) adapter (e1000 v4.4.12-k1 [ 12]). The PCs were running Linux kernel version 2.4.20. The two routers were connected back-to-back: “Router Under Test” (RUT) and “Generic Router” (GR). GR does not change between tests. The routers were connected using either POS OC-48 (2.5Gbps) or Gigethernet (1Gbps Ethernet) line cards.

The flows are sent from PC-A-1 to PC-B-1, PC-A-2 to PC-B-2 and PC-A-3 to PC-B-3. To baseline the performance of the PCs we connected two PCs back-to-back and we measured the throughput versus packet size. Figure 2 shows these results. To generate traffic from the PCs we use iperf [ 13] version 1.6.5, which generates a constant bit rate (CBR) pattern and the transport protocol being UDP.

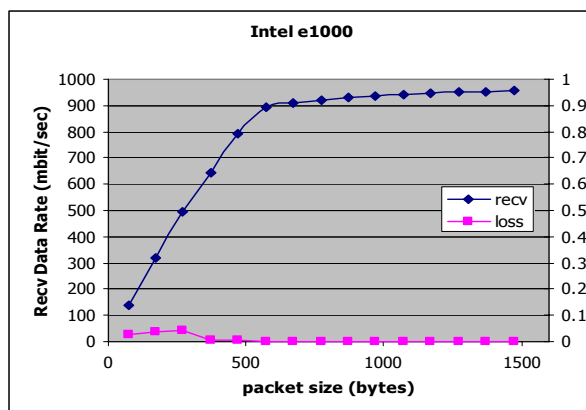


Figure 2: Performance results for Intel Gigethernet cards.

From the results in the Figure 2 we see that to achieve line rate from the PCs, we required a packet size quite close to the Ethernet MTU. We chose a packet size of 1470 bytes for our tests. The maximum achieved throughput at this packet size for the PCs plugged back-to-back is 955Mbps.

We use two classes in our tests. “Best effort” (BE) class with DSCP=0 and “Less than best effort” (LBE) class with DSCP=8 (001000). This consistent with the recommendation from Internet2 [ 14] group that used the same DSCP code. Note that for our tests the packets are marked with the DSCP code at the PCs before been transmitted.

## III. PERFORMANCE EVALUATION of the CISCO 7609 ROUTER.

As RUT we use a Cisco router 7609. This is a nine slots chassis mounting a Catalyst Supervisor 2 module and the Switching Fabric module. The IOS software is 12.1(19)E.

### A. OC-48 line card

We installed the OC-48 line card (OSM-1OC48-POS-SS+) in the Cisco 7609 chassis. This line card has a single OC-48 (2.5Gbps) interface and four 1Gigethernet catalyst ports. The PCs were connected via the catalyst ports. The OC-48 interface runs Packet over SONET (POS) and we used PPP encapsulation between the routers.

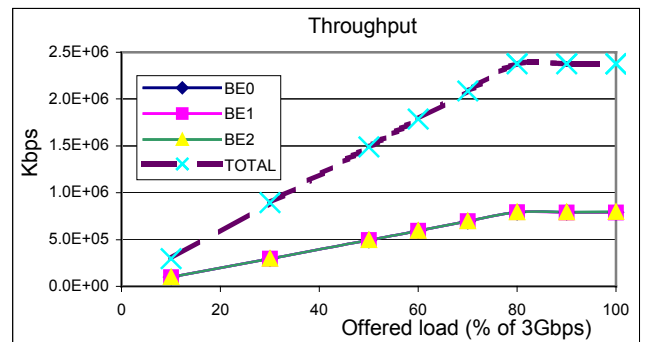


Figure 3: Maximum throughput achieved in the OC-48

We did some tests between PC-As and PC-Bs to double-check that the throughput remained the same as observed in Figure 2. We also tested the maximum achievable throughput when sending three flows from PC-As to PC-Bs.

In Figure 3 we were sending three BE flows increasing the ingress load in the RTU up to 100% of each ingress capacity, that is 1Gbps. The figure shows that the maximum achieved throughput is 2.37Gbps, the line rate of the OC-48.

```

policy-map ciscoqospolicy
 class BE
   bandwidth percent 89
 class LBE
   bandwidth percent 10
    
```

Figure 4: Policy applied for BE and LBE classes.

We defined the two classes mentioned earlier in the router: BE and LBE and configured a very simple policy called “ciscoqospolicy” shown in the Figure 4, and finally this policy is applied to the output interface in the RUT, connecting to the GR.

To allow control traffic going through in the link, Cisco IOS limits the user’s traffic to 99% of the total available bandwidth.

With this policy we configured the Class Based Weighted Fair Queuing (CBWFQ) algorithm and applied to the scheduler a bandwidth allocation of 89 percent of the 2.37Gbps for BE and 10 percent of the 2.37Gbps for LBE during congestion, that is 2.109Gbps for BE and 237Mbps for LBE.

The maximum throughput we can get for BE class using two PCs is 1.91Gbps so the remaining bandwidth<sup>1</sup> up to 2.109Gbps can be used by the LBE flow during congestion. That means that the theoretical bandwidth for the LBE class is 436Mbps.

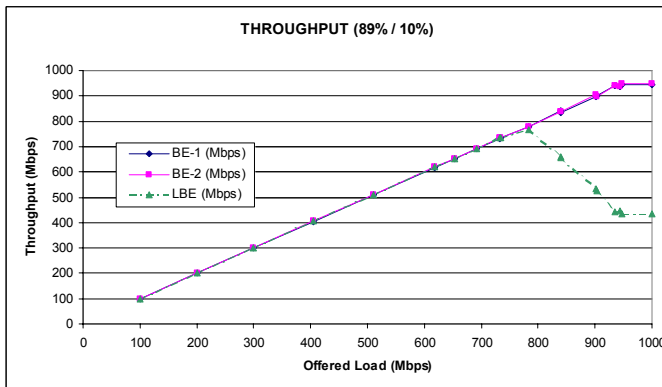


Figure 5: Throughput for BE and LBE classes with policy of 89% and 10%.

Figure 5 shows the results of the throughput for both classes when we increase the ingress load of the flows up to the congestion point. We see that LBE traffic is using 434Mbps in the maximum congestion point. The error between the theoretical value and the real value for LBE<sup>2</sup> is:

$$Err\_LBE(\%) = \frac{436 - 434}{436} * 100 = 0.45\%$$

Checking the figures in the congestion zone, the total utilization of the link for the two classes is 2.327Gbps, that is 98.2% of the total capacity of the POS OC-48 link.

We can stress the scheduler algorithm applying a more aggressive policy in the interface. The worst-case scenario following the suggestion from Internet2 [ 14]

is to apply a bandwidth allocation of 98% for BE and 1% for LBE<sup>3</sup>.

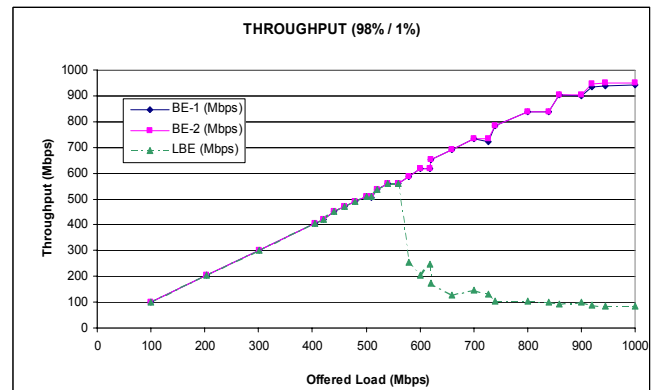


Figure 6: Throughput for BE and LBE classes with policy of 98% and 1%.

Figure 6 shows the performance of the scheduler for the most aggressive policy. In this case BE class would achieve 98% of the maximum capacity of the link, that is 2.322Gbps, but again using two PCs for BE, we are limited to 1.91Gbps. However, LBE class achieves a maximum throughput in congestion around 84.9Mbps and that means in this case the utilization of the link is 83% compare to the 98.2% in the previous case. In this case the error between the theoretical value and the real value for LBE is:

$$Err\_LBE(\%) = \frac{436 - 84.9}{436} * 100 = 80\%$$

Policy (BE/LBE)	LBE theo (Mbps)	LBE real (Mbps)	Err_LBE (%)
89/10	436	434	0.45
94/5	436	447	2.5
96/3	436	282	35
97/2	436	212	51
98/1	436	84.9	80

Table 1: Error allocation for LBE bandwidth for different policies.

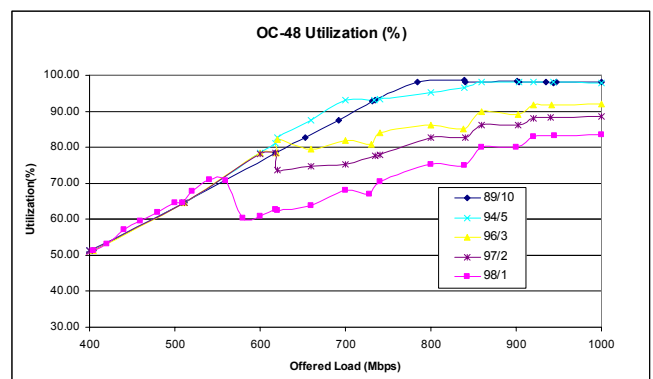


Figure 7: Utilization of the OC-48 link for different BE/LBE bandwidth allocations.

<sup>1</sup> Note that the scheduler does not enforce a hard limit on the maximum usable bandwidth by a flow when there is no congestion.

<sup>2</sup> In this case just the error for LBE class is estimated. For the BE class the PCs are not capable to achieve the required rate.

<sup>3</sup> Such an aggressive policy is not typical commercially and manufacturers do not recommend it. However it is interesting for the purpose of our research.

Table 1 shows the evolution of the error allocating LBE bandwidth for different policies from an unaggressive (89/10) up to an aggressive one (98/1).

Figure 7 shows the utilization of the OC-48 link expressed in percentage. The maximum achieved throughput, as seen, is 2.37Gbps equivalent to 100% utilization of the link. So, applying different bandwidth allocations in the scheduler for BE and LBE classes we can see that the utilization is almost 100% for a bandwidth allocation (BE/LBE) of 94/5 and higher.

#### IV. PERFORMANCE EVALUATION of the JUNIPER M10 ROUTER.

As RUT we use a Juniper router M10. It runs the software JUNOS OS [5.3R2.4].

##### A. GigEthernet card

We installed in this chassis a four ports 1Gigethernet card (rev 01, 750-005091). One of the interfaces was connected to the router GR, the other was not available for our use. The other two remaining interfaces were used to connect the PCs. So in this scenario we had two PCs (PC-A-1 and PC-A-2) sending traffic, and receiving in the other end by PC-B-1 and PC-B-2. In this case the maximum achievable throughput in the link between the two routers was measured to be 957Mbps.

Figure 8 shows the definition of the classifiers expressed in JUNOS fashion, where LBE class is defined as "cs1" or 001000.

```

classifiers {
  dscp UCL-classifier {
    forwarding-class LBE {
      loss-priority low code-points cs1;
    }
    forwarding-class best-effort {
      loss-priority low code-points 000000;
    }
  }
}

```

Figure 8 : Definition of Classifiers in Juniper.

```

schedulers {
  sch-BE {
    transmit-rate percent 90;
    buffer-size percent 90;
    priority high;
  }
  sch-LBE {
    transmit-rate percent 10;
    buffer-size percent 10;
    priority low;
  }
}

```

Figure 9: Definition of the Scheduler in JUNOS.

Figure 9 shows how to define the scheduler for the classes defined BE and LBE. We see in this case the bandwidth allocation we used is 90% for BE and 10% for LBE. It is worth mentioning that Juniper recommends but do not enforce reserving bandwidth for control traffic.

By applying this policy to the output interface in RUT, connecting to GR, we enabled WFQ. In this case the theoretical values are 90% of the total available bandwidth is 861Mbps for BE and 10% of the bandwidth is 95.7Mbps for LBE.

Figure 10 shows the results of the throughput for both classes when we increased the ingress load up and over the congestion point. In the results we see LBE uses 83Mbps and BE uses 873Mbps. The errors allocating bandwidth for both classes are:

$$Err\_BE(\%) = \left| \frac{861 - 873}{861} \right| * 100 = 1.39\%$$

$$Err\_LBE(\%) = \frac{95.7 - 83}{95.7} * 100 = 13.2\%$$

The total utilization of this 1Gigethernet link for this policy is 99.8%.

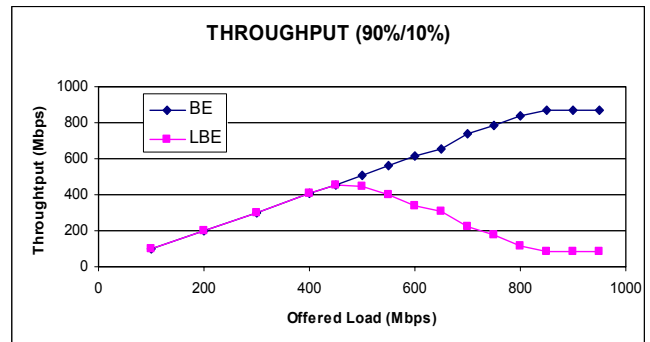


Figure 10: Throughput for BE and LBE classes with policy of 90% and 10%.

We configured the scheduler to the extreme case of 99% of the bandwidth for BE and 1% of the bandwidth for LBE.

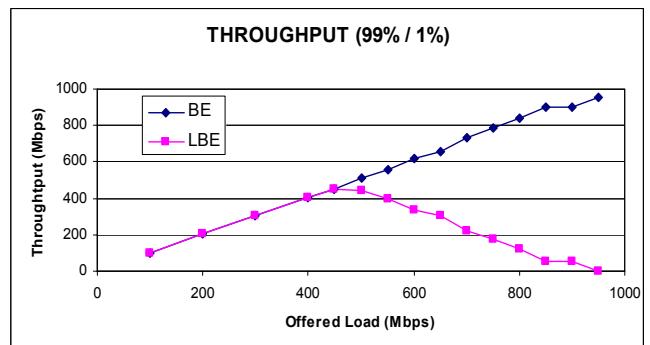


Figure 11: Throughput for BE and LBE classes with policy of 99% and 1%.

Figure 11 shows the performance of the scheduler for the policy 99% for BE over 1% for LBE. The theoretical values for the allocation of the bandwidth are 947Mbps for BE and 9.57Mbps for LBE. Checking the values from the tests we see BE class uses 954Mbps and LBE class uses around 1Mbps. The errors allocating bandwidth for both classes are:

$$Err\_BE(\%) = \left| \frac{947 - 954}{947} \right| * 100 = 0.74\%$$

$$Err\_LBE(\%) = \frac{9.57 - 1}{9.57} * 100 = 89.5\%$$

Table 2 shows the evolution of the error allocating LBE bandwidth for different policies from an unaggressive (90/10) up to an aggressive one (99/1). We see the error increasing when the difference between the bandwidth allocation for the two classes is getting bigger.

Policy (BE/LBE)	LBE theo (Mbps)	LBE real (Mbps)	Err_LBE (%)
90/10	95.7	83	13.2
93/7	66.99	54.8	18.2
95/5	47.58	34	28.5
97/3	28.71	15.2	47.0
98/2	19.14	6.4	66.5
99/1	9.57	1	89.5

Table 2 : Error allocation for LBE bandwidth for different policies.

The utilization of the 1Gigethernet link for this policy of 99/1, in Figure 12, remains close to 99.8%, so there is no difference comparing with the policy of 90/10.

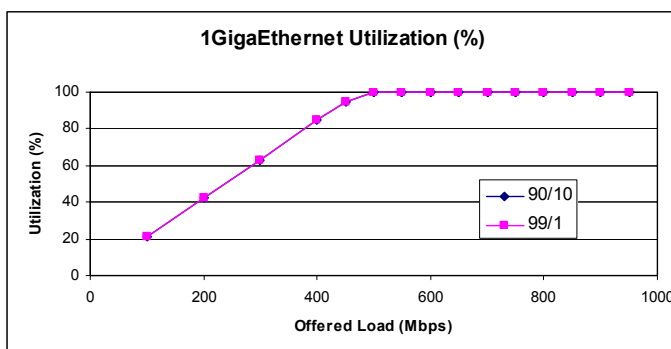


Figure 12: Utilization of the Gigethernet link for two policies of BE/LBE bandwidth allocations.

## V. CONCLUSION

We can see how the Cisco OC-48 line card performs very well for the policy where BE=89% and LBE=10%, having a total link utilization of around 98% during congestion (without counting the bandwidth allocated

to the control traffic). Also the error in allocating bandwidth for the class with low priority, LBE, is almost negligible (0.45%). Progressively moving to a more but commercially unlikely aggressive policy (BE=98% and LBE=1%), we see how the performance in the scheduler decreases down to a utilization of 83% of the total available bandwidth during congestion. In this case the resulting error allocating LBE bandwidth in the scheduler is around 80%.

As seen in the results provided, the performance of the Juniper Gigabit Ethernet card is very good from the point of view of utilization of the bandwidth. We see that regardless how aggressive is the policy in the scheduler is -- an unaggressive policy (BE=90% and LBE=10%) and an aggressive (BE=99% and LBE=1%) -- the utilization of the total available bandwidth during congestion is close to 100%. However we found quite a high relative error in allocating bandwidth to the lower priority class (LBE), from 13% for an unaggressive policy to 89% for an aggressive one.

Due to the current limitations of using PCs to generate and receive traffic, it will be interesting to benchmark these router interfaces with dedicated equipment. For example using Spirent's [ 15] Smartbits equipment, which is capable of providing line rate for all packet size. Also it will be interesting to compare the performance using realistic traffic profile with a variety of packet sizes, MTUs, traffic distributions and hundreds of flows associate with a given class.

We intend to test other QoS enabled router interfaces of gigabit per second rates and higher primarily for support of Grid projects with demands of high throughput and QoS features. We also intend to test these routers under more realistic conditions using a wider range of classes (apart from BE and LBE).

## Acknowledgement

We would like to thank everybody supporting this work and collaborating in both MB-NG and DataTAG projects. Special mention to people at UCL: Ian Bridge, Nicola Pezzi, Miguel Rio and Yee-Ting Li and at CERN: Edoardo Martelli, Paolo Moroni and Jean-Philippe Martin Flatin.

## References

- [ 1] Braden, R., Zhang, L., Berson, S., Herzog, S. and S. Jamin, "Resource Reservation Protocol (RSVP) Version 1 Functional Specification", RFC 2205, September 1997.
- [ 2] Y. Bernet, R. Yavatkar, P. Ford, F. Baker, L. Zhang, K. Nichols, and M. Speer, "A Framework for Use of RSVP with Diffserv Networks"
- [ 3] K. Nichols, V. Jacobson, and L. Zhang, "A Two-bit Differentiated Services Architecture for the Internet", <ftp://ftp.ee.lbl.gov/papers/dsarch.pdf>, November 1997.
- [ 4] [www.cisco.com](http://www.cisco.com)
- [ 5] [www.juniper.net](http://www.juniper.net)

- [ 6] [www.mb-ng.net](http://www.mb-ng.net)
- [ 7] [www.datatag.org](http://www.datatag.org)
- [ 8] <http://www.realitygrid.org/>
- [ 9] <http://www.slac.stanford.edu/BFROOT/>
- [ 10] D.Black et al., “An Architecture for Differentiated Services”, RFC 2475.
- [ 11] [www.supermicro.com/PRODUCT/SUPERServer/SuperServer6022P-6.htm](http://www.supermicro.com/PRODUCT/SUPERServer/SuperServer6022P-6.htm)
- [ 12] <http://support.intel.com/support/network/adapter/1000/index.htm>
- [ 13] <http://dast.nlanr.net/Projects/Iperf/>
- [ 14] Internet2 “Scavenger Service”  
<http://qbone.internet2.edu/qbss/>
- [ 15] <http://www.spirentcom.com>