JANET QoS Experiments

Results from University of Southampton trials

Experiments with Less than Best Effort

Editor: Dr T J Chown

School of Electronics and Computer Science,
University of Southampton

Version 1.04

10th October 2004

Contents

Figures

1 Introduction

This document reports the results observed by the School of Electronics and Computer Science (ECS) at the University of Southampton in conducting tests within the JANET QoS testing activities of Q1/Q2 2004 in partnership with Imperial College London.

The tests were to focus on use of Less than Best Effort (LBE) in a "real" environment, using both simulated traffic flows (using off-the-shelf generators and collectors/measurement tools) and real applications (in particular, the AccessGrid multiparty, high quality conferencing and collaboration system). The aim was to show that LBE flows could be serviced by the network without disrupting the "regular" applications in an everyday network environment (outside of a more clinical testbed environment).

General details of the test plans for the JANET QoS tests can be found in the contemporary QoS Test Plan document published by UKERNA.

Three test slots were booked. This document describes the results of the two sets of tests carried out on 9th and 16th March 2004, between Southampton and Imperial College, London. The testing on 2nd March was postponed due to delays in deploying (moving) the AccessGrid node at Southampton, although this slot was used for some basic calibration tests.

The results show that for the class of router used at the edge (Cisco 7206 series), the LBE and BE traffic behaved as expected (hoped) in the tests.

Useful experience was also gained with traffic measurement tools, and we believe there is useful further work to do in the measurement area.

However, the AccessGrid tests were inconclusive, because, we suspect, of the different handling of multicast traffic by the QoS-enabled routers in the tests. Further work is required in this area. Our report also suggests other potential interesting areas for further work in the next phase of JANET-QoS testing.

2 Components for Testing

There were a considerable number of factors involved in the deployment of the systems for the LBE testing. These included:

- Determining the network architecture and topology for the testing.
- Configuring the QoS Cisco (7206, running Geshi-II image) accordingly.
- Deploying the traffic generation, collection and monitoring tools. Deciding which tools were relevant, and which may be interesting for further (later) trials. Only software tools were used; no hardware test equipment was used (or available).
- Deciding how the LBE and BE traffic would be tested, i.e. the methodology of flow variation over time to give meaningful results for inspection.

- Ensuring the AccessGrid application was available. This was complicated by the Southampton AG node being moved just before the planned test slots. In the end, it was only used in the final slot.

- Configuring the VoIP handset and SAA device (Cisco 805) for Premium IP tests. This was not used in the LBE tests, and ultimately not in the Premium IP tests as LeNSE (the regional MAN connecting Southampton to JANET) was not Premium IP enabled.

- Handling firewall issues as they arose (enabling ports for the tests, checking performance problems in the Southampton campus firewall, which were resolved in time for the first test). The Southampton ECS firewall was bypassed for the tests.

The amount of effort involved in the generation, capture and interpretation of the measurements should not be underestimated.

3 Methodology

The JANET QoS tests encompassed a range of different QoS goals and technologies.

3.1 Scope

In the context of the LBE tests, we wished to experiment and demonstrate the following properties for LBE:

- LBE traffic is dropped in preference to BE when congestion occurs

- LBE traffic receives a minimum percentage of the available bandwidth to avoid session starvation (particularly important for TCP applications, though we focused on UDP tests).

In principle we would also ideally test that LBE is also dropped in deference to Premium IP traffic on congestion, and that LBE maintains a minimum bandwidth (when LBE traffic flows) in the presence of Premium IP and BE together, but the Southampton and Imperial regional networks did not support Premium IP, thus the tests were not feasible.

The minimum allocation for LBE is typically set to anything from 1% to 5% of available bandwidth. The precise value is left as a decision for the operator. A lower value allows more BE (regular) traffic to flow while not completely killing LBE flows, while a 5% value has the advantage of encouraging use of LBE for applications (in particular more heavy, but non real-time bulk data transfer operations). Of course LBE may occupy more than 5% of the available bandwidth if the BE and other traffic does not consume the rest of the path.

We also wished to test how regular IP multicast operated in the presence of LBE background, in particular testing with the AccessGrid, which is a popular tool in the Grid and other communities for high quality conferencing. It is quite probable that the very sites using AccessGrid may also be those interested in bulk (Grid) data transfers. Other regular BE traffic may also be expected in such networks, of course.

In due course one would like to test Premium IP multicast; this may be on the horizon for JANET tests, but is not in scope yet. It is also a distinctly non-trivial service to provide.

All our tests used IPv4. It is expected that future tests should also encompass IPv6, now that the JANET backbone, and some regional MANs (e.g. LeNSE) offer a dual-stack IPv4-IPv6 service.

3.2 LBE deployment

LBE is more commonly applied at the edge of a network rather than in the core. Indeed LBE has the advantage that it need only be deployed incrementally, since there is nothing for users to gain by running an application at lower than normal priority. This makes deployment much easier than Premium IP, since no end-to-end SLS/SLA needs to be agreed and provisioned. Also, no policing on inbound aggregates is required, nor is any authentication scheme needed to enable a site/user to be allowed to send LBE traffic.

LBE applications do not have strict timing requirements – they cannot because there is no guaranteed service, beyond the minimum aggregate of bandwidth that may pass any one LBE-enabled router at any point.

The only requirement on the network is that it does not alter the Differentiated Services Code Point (DSCP) value in transit between source and destination. LBE has a “well-known” (at least agreed between many national academic networks) value of 8.

It should be noted that where LBE is not implemented, LBE traffic is usually trated just like BE traffic. It is thus in the interests of an operator to enable LBE where they know that LBE traffic is flowing (this can be determined by observing DSCP values in transit).

LBE can be applied at just one ingress or egress point. It may for example be used in a campus environment as a form of campus bandwidth management, e.g. by tagging student hall/dormitory traffic as LBE.

3.3 Measurements

The typical QoS characteristics that are measured or monitored include:

- One way delay (OWD)
- Round trip time (RTT)
- Jitter (variation in delay)
- Packet loss
- Throughput over time
- Packet misordering/reordering

Because LBE assumes no traffic properties, beyond there being a minimum pipe available to avoid complete session starvation, these characteristics are not important for the LBE per se. However, they are important for the BE traffic that the LBE may be completing with, since BE must perform well (utilise available bandwidth) in the presence of LBE.

OWD is not trivial to measure accurately because it generally requires accurate time synchronisation. Also, OWD and RTT are generally only of keen interest for Premium IP (or better than BE) traffic. Likewise reordering is also more important for Premium (time sensitive applications) because reordering can adversely affect TCP throughput and performance that is not critical for BE and LBE applications.

Thus in these tests, we focus on measuring the packet loss and jitter for BE traffic in the presence of LBE. We also measure LBE loss and jitter.

3.4 Network

One of the features of these JANET QoS tests are that they are applied on a real network (JANET, the MANs and campus networks) with heterogeneous hardware, rather than being constrained to homogeneous testbed networks.

For our tests, the path between the two test sites – Southampton and Imperial – is the one of interest.

Figure 3-1 is taken from UKERNA documentation of the JANET network as of Q4 2003, and was verified by running a network trace between the sites (the path traversing LeNSE, Cosham, London, ULCC and LMN. The trace showed 11 hops just to get from the testbed systems in Southampton to Cosham. (A full trace was not possible since Imperial filters ICMP, presumably for security reasons).

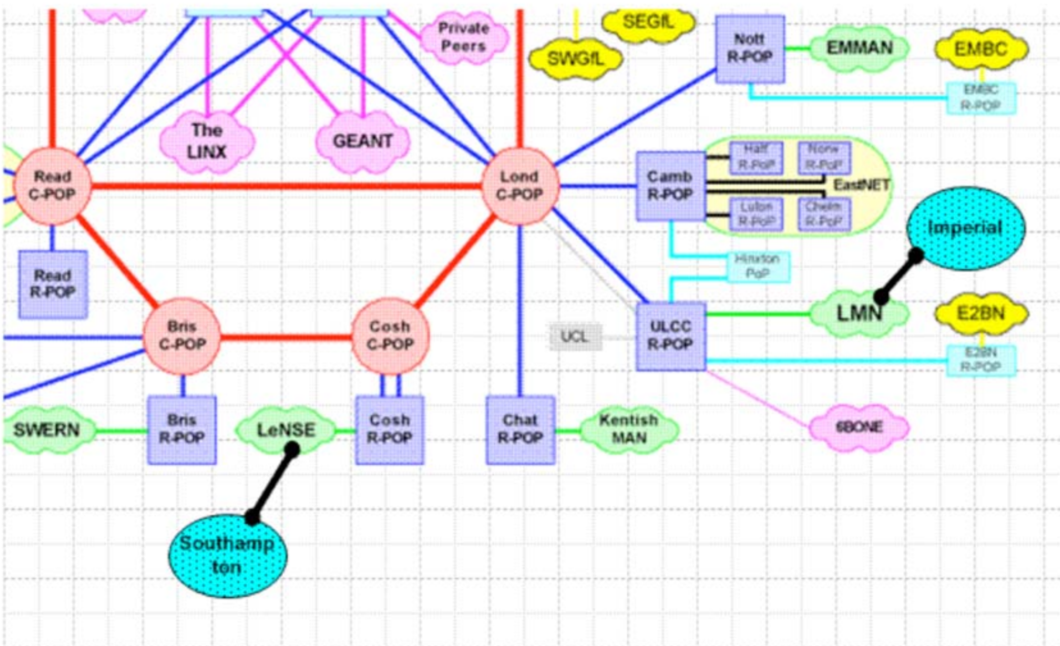


Figure 3-1: Network path Southampton - Imperial

The Southampton site (ECS) is connected to the Southampton campus network via a Gigabit Ethernet link. The campus has a switched Extreme network. The link from ECS to the campus egress was provisioned via a dedicated VLAN. The campus-LeNSE border includes a Nokia IP700 series hardware-assisted firewall. LeNSE has a 2.5Gbit link to JANET at Cosham.

The link for the QoS tests was provisioned using a Cisco 7206 router running the Geshi-II image (an experimental, but robust, image). This Cisco is not the ingress/egress router for regular ECS traffic, but was used by a testbed network including the School's AccessGrid system (the AccessGrid is quite heavily used in ECS by a large community of users) and any devices connecting in the AccessGrid room.

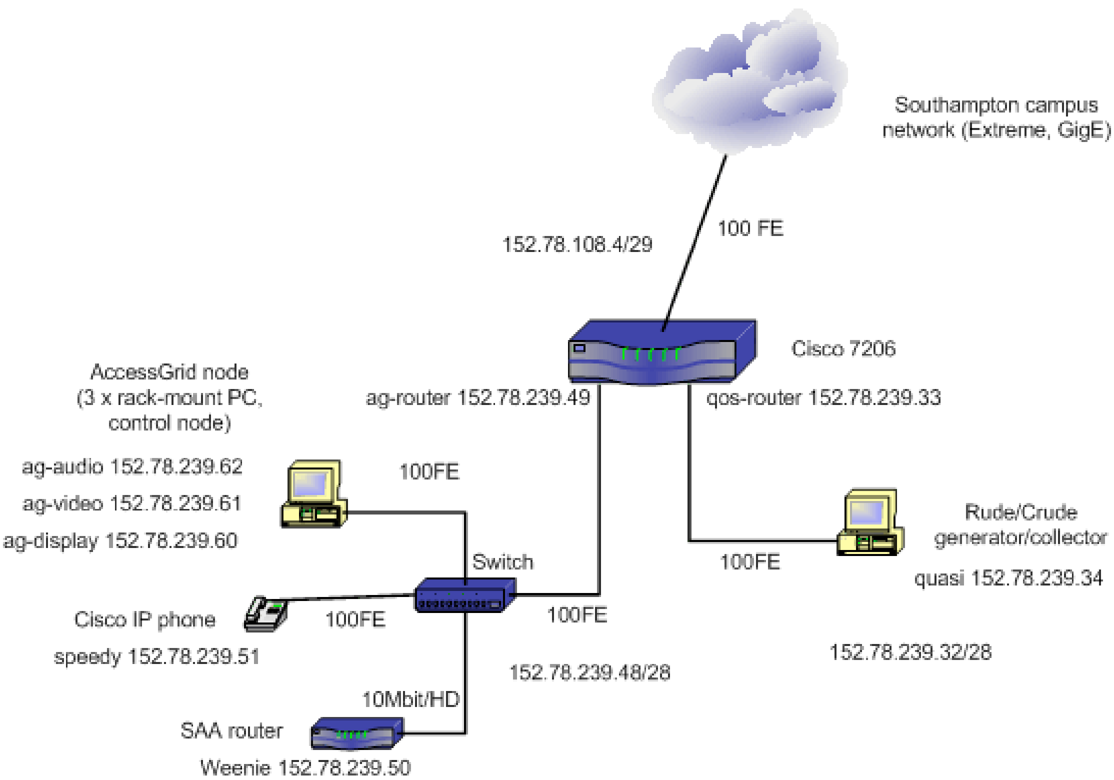


Figure 3-2: Southampton-ECS QoS deployment

The deployment saw one interface of the Cisco router used for traffic generation and collection (LBE background) and one for the application (BE). A laptop device was used on the application interface where BE traffic was generated/received during specific tests.

The SAA router and VoIP equipment were deployed in support of potential Premium IP work that was not ultimately possible due to the lack of Premium IP deployment on LeNSE. However, the SAA system would still be available to measure network characteristics on the application side while LBE background traffic was generated.

The SAA router interface is limited to 10Mbit/s half duplex.

The main router interfaces were designed to create congestion at the uplink from the Cisco QoS router, with the uplink forced down to 100Mbit/s (it is a Gigabit Ethernet interface) and two incoming 100Mbit/s interfaces (on a dual 100FE card).

By avoiding Gigabit Ethernet, the tests also avoided the issue of how PC-based systems handle Gigabit Ethernet transmission/reception. This means that the tests were based purely on the router performance, rather than PC performance.

The Imperial site network was described in their results documentation, but is repeated below:

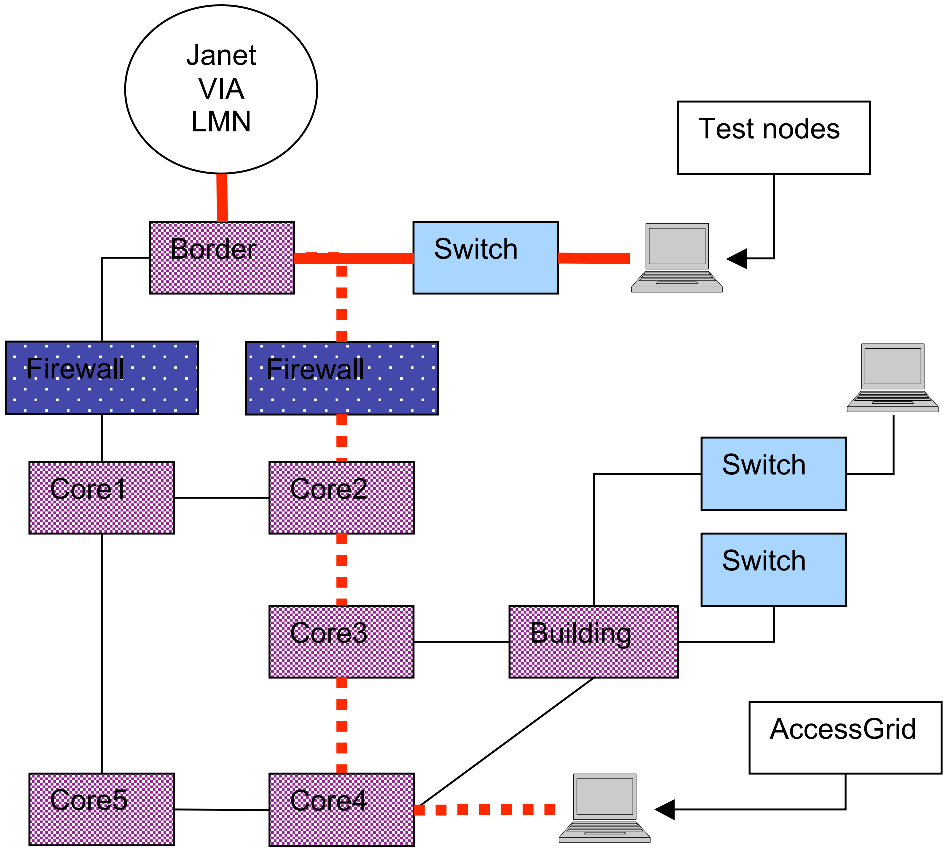


Figure 3-3: Imperial site QoS network

The transparency of the DSCP field in the packet headers was checked using the traceroute utility with the –t option (specify TOS byte). A TOS value of 32 is the equivalent of a DSCP value of 8. Using traceroute –t, any hop on the path that resets or alters the DSCP is clearly highlighted.

3.5 Router configuration

In this section, we cite the router configuration.

3.5.1 Main Cisco router (7206)

The Cisco 7206 system was running the Geshi-II IOS image.

An excerpt of the configuration:

```
class-map match-any BE
 match ip dscp default
class-map match-any LBE
```

```
 match ip dscp 8
!
policy-map TEST
 class BE
  bandwidth percent 95
 class LBE
  bandwidth percent 5
!
interface GigabitEthernet0/1
 ip address 152.78.108.4 255.255.255.248
 ip pim sparse-mode
 max-reserved-bandwidth 100
 service-policy output TEST
 duplex auto
 speed 100
 media-type rj45
 no negotiation auto
```

No specific WRED was configured on the interface(s).

3.5.2 SAA router

The Cisco 805 router used for SAA monitoring ran 12.2.13T11 (IPPlus), and was configured with the assistance of Swansea.

3.6 Firewall considerations

The firewall at the Southampton campus border was set to default deny inbound connections, and also to block some outbound connections (based on specific security concerns). This impacted the tests, and required some firewall configuration changes.

- The SAA router needed TCP ports 21400, 21401 and 21405 opened inbound, and UDP 161/162 for SNMP.

- The IP phone needed 69/udp outbound (tftp) and 2000/tcp outbound to the call manager and 2000/tcp inbound. The IP phone also used “random” UDP ports above 1024. TFTP is often blocked outbound at sites because of blaster and other worms that use it heavily to infect other systems.

- All userland ports (>1024) were opened in/outbound for the rude/crude system.

At a time up until the first test slot, there were problems on the campus firewall (between Southampton and LeNSE). The firewall became CPU bound, such that a flow of 100Mbit/s actually caused the regular incoming accepted traffic to drop to near zero. This fault was correct before the second test slot.

3.7 Applications

A number of applications were considered, including:

- iperf

- AccessGrid
- AWM
- VoIP
- FTP

AccessGrid was chosen, for the third test slot, due to its use of IP multicast in a high performance way.

3.8 Measurement tools

An interesting aspect of the QoS measurement is which tools to use, and, where multiple tools are available, how their measurements compare (are they consistant?).

In addition to specific objective tests, subjective tests on audio and video are also possible (in such cases the test subjects should not know when background LBE traffic is being used, and when it isn't).

3.8.1 iperf

The iperf package is a good tool for measuring TCP or UDP throughput between two nodes. For example we tested iperf with UKERNA on Feb 25th 2004:

```
$quasi iperf-1.7.0]$ ./iperf -c 212.219.210.6 -u
------------------------------------------------------------
Client connecting to 212.219.210.6, UDP port 5001
Sending 1470 byte datagrams
UDP buffer size: 64.0 KByte (default)
------------------------------------------------------------
[  3] local 152.78.239.34 port 32800 connected with 212.219.210.6 port 5001
[ ID] Interval       Transfer     Bandwidth
[  3]  0.0-10.0 sec  1.25 MBytes  1.05 Mbits/sec
[  3] Server Report:
[  3]  0.0-10.0 sec  1.25 MBytes  1.05 Mbits/sec  1.316 ms    3/  893 (0.34%)
[  3] Sent 893 datagrams
```

One problem with iperf is that it isn't amenable to ramping test bandwidth up and down in such a way that per-period (e.g. per second) QoS characteristics can be measured. However, it has use as a means of measuring the overall throughput in the presence of contending traffic.

3.8.2 Rude/Crude

In our tests we chose to use rude/crude for their flexibility. The rude program is a data emitter (packet generator) while crude is its corresponding collector. With rude/crude, it is easy to configure one flow ID per period (e.g. as the BE is ramped up and down per period). It is also possible to specify the TOS value in a rude script in the form of the command '*TOS <id> <value>*' where ID is the flow ID to be measured.

We choose to use rude/crude port for sending/collection of data for our tests, measuring both ways (Southampton to Imperial and Imperial to Southampton for both BE and LBE flows). We used version 0.62 of the software.

The rude traffic generation scripts were written by Southampton, e.g. to ramp up and down an LBE flow you could use a script like the following:

```
START NOW
002000 30 ON 5000 152.78.239.34:20002 CONSTANT 4000 1472
TOS 30 32
011999 30 OFF
```

```
012000 31 ON 5001 152.78.239.34:20002 CONSTANT 4100 1472
TOS 31 32
017999 31 OFF
018000 32 ON 5002 152.78.239.34:20002 CONSTANT 4200 1472
TOS 32 32
023999 32 OFF

<snip>

366000 90 ON 5060 152.78.239.34:20002 CONSTANT 10000 1472
TOS 90 32
371999 90 OFF
372000 91 ON 5061 152.78.239.34:20002 CONSTANT 10100 1472
TOS 91 32
377999 91 OFF
378000 92 ON 5062 152.78.239.34:20002 CONSTANT 10200 1472
TOS 92 32
497999 92 OFF
498000 93 ON 5063 152.78.239.34:20002 CONSTANT 10100 1472
TOS 93 32
503999 93 OFF
504000 94 ON 5064 152.78.239.34:20002 CONSTANT 10000 1472
TOS 94 32
509999 94 OFF

<snip>

792000 127 ON 5097 152.78.239.34:20002 CONSTANT 6700 1472
TOS 127 32
797999 127 OFF
798000 128 ON 5098 152.78.239.34:20002 CONSTANT 6600 1472
TOS 128 32
803999 128 OFF
804000 129 ON 5099 152.78.239.34:20002 CONSTANT 6500 1472
TOS 129 32
813999 129 OFF
```

And then collect the data with a listening crude command, e.g.:

```
$ crude -p 20002 –s
30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,
51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,
72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,
93,94,95,96,97,98,99,100,101,102,103,104,105,106,107,108,109,11
0,111,112,113,114,115,116,117,118,119,120,121,122,123,124,125,1
26,127,128,129
```

Each transmitted flow at a specific bandwidth needs to be matched by listening for each flow ID on the receiver end. If you use a different flow ID for each increment then the collector can report stats for each flow (increment). But the collector must listen for the flows (-s option). Alternatively crude will just log all packets. The per flow summaries are very useful in the context of the JANET QoS tests.

An example of the collection output for a flow ID is:

```
Flow_ID=35
Packets: received=807   out-of-seq=1   lost(est)=0
Total bytes received=1151589
Sequence numbers: first=0   last=806
Delay: average = 69.428769   jitter=0.000397   seconds
Absolute maximum jitter=0.014761   seconds
Throughput=1.16145e+06   Bps  (from first to last packet received)
```

Rather than process the resulting long collection log file by hand, Imperial wrote a Python script to parse the collection file to a CSV for import to Excel for chart plotting. This was used for the plots in the sections describing the tests for slot 2 and slot 3 later in this document.

3.8.3 SAA

Cisco SAA measurement systems were deployed at each JANET QoS partner site by the University of Swansea. The details of the SAA deployment are described in the Swansea QoS report. SAA allows measurement of statistics including RTT (min and max), OWD (source to destination and vice versa) and packet loss (source to destination and vice versa).

Note that as configured for the JANET QoS tests, OWD statistics were not available.

For the LBE tests between Southampton and Imperial, the Reading-to-Southampton and Imperial-to-Reading links were monitored, and reported characteristics for LBE, BE and Premium (EF) marked packets. Specifically the SAA monitoring router was deployed on the applications subnet of the QoS network, thus seeing the same network characteristics as the applications used (e.g. the AccessGrid traffic). If there were congestion there, this would be highlighted by the graphs.

In the reporting that follows, we focus on the LBE SAA results for the Reading-to-Southampton data.

One limitation of SAA as deployed for the JANET QoS sites was that it only measured statistics once per minute. Where some tests only lasted a few minutes, better results were achieved via rude/crude which can report characteristics on much shorter frequencies (e.g. once per second).

3.8.4 RTCP

Another option to measure the QoS characteristics is to use the RTCP information gathered by the applications, where they support such a mechanism. The MICE tools used by the AccessGrid support this, but it is not trivial to extract the data.

In our third test slot we recorded the raw RTCP data flowing “on the wire” with Ethereal, but it was not possible to easily correlate the data where multiple RTP streams were present. Nor did the application have any apparent mechanism to log the data. This is an interesting area worthy of further investigation, but we did not use the RTCP data in our assessments.

Some tools that may be of use for RTCP and Multicast measurement include:

- UCL RTP quality matrix: http://www-mice.cs.ucl.ac.uk/multimedia/software/rqm/
- SM: http://carmen.cselt.it/ipmc/sm/
- RTPmon: http://bmrc.berkeley.edu/~drbacher/projects/mm96-demo/demo2.html
- The new AG Multicast beacon based on RTCP reporting: http://dast.nlanr.net/Projects/Beacon/

These tools could be tested in any Phase 2 JANET QoS work looking at QoS and Multicast.

3.8.5 Netsight

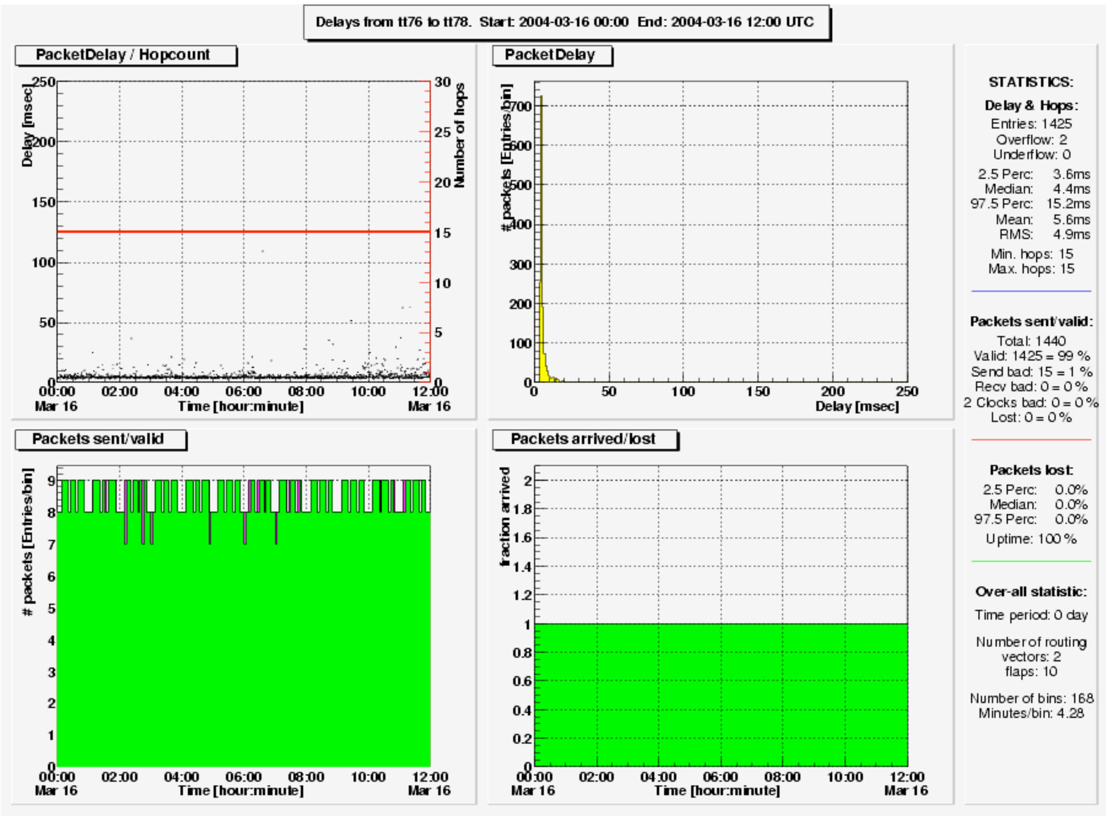
Netsight is the monitoring and measurement tool used on the JANET network. LeNSE made the bandwidth utilisation statistics available to ECS during the test slots, though no figures for loss or delays were available, nor were the measurements end-to-end (they showed just aggregate bandwidth used by the whole Southampton campus). Thus the statistics were not specific to the tests or nodes being used, reducing the value for the assessment, but the figures were still useful for correlation purposes.

The granularity of the Netsight readings (once per 5 minutes average using MRTG) meant that the true bandwidth peaks were not visible.

3.8.6 RIPE TTM server

Southampton is one of around 100 sites that have a RIPE-NCC Test Traffic Measurement server deployed. The TTM architecture allows the gathering of statistics including OWD, hop count, loss and jitter between any two such servers. Now open to all on signing of a legal declaration. The data is gathered at a central server, and any period can be viewed historically.

While the statistics are useful, the TTM server is not deployed directly on the QoS network in Southampton, nor was there a TTM server at Imperial, the other LBE tests site. Like SAA, statistics are only gathered at relatively coarse intervals.



RIPE data, 16th March 2004-10-09

Figure 3-4: Sample RIPE TTM measurement output

Note that RIPE NCC now requires a T&C Agreement to be signed before gathered data may be published, but the results are now more open for access than they were in 2003 (when results were available only to TTM sites).

3.8.7 MRTG

MRTG is a traffic-graphing tool that typically polls SNMP MIB data to produce charts of QoS characteristics over time. As an example, the ingress/egress QoS router at a site could be polled for statistics during the tests. However, one needs to be careful that the polling isn't affected by the congestion or loading that is occurring as part of the test itself.

3.8.8 Discussion

There is clearly a good selection of measurement tools that could be utilised.

We decided to focus on rude/crude because it has the flexibility to ramp streams up and down easily with scripting, and gather the required characteristic measurements over quite fine-grained periods of time. The statistics crude records can be converted to a format that is readily visualised by appropriate script(s). Imperial authored such a script for the QoS tests, using the Python language. This was used to produce the result graphs in the sections below (for slot 2 and slot 3).

UDP data flows were chosen for the tests, to avoid any potential issues with TCP backoff algorithms impacting the observed statistics of raw network response. Future tests on practical results with TCP in a real environment should be considered however, for the LBE case.

The test scripts used fixed size data packets. It would be useful to run future tests that would involve the use of typical packet size distributions as seen in "real" networks. (The use of fixed size packets may have been the cause of an interesting observation in the results below whereby increasing congestion caused average jitter to fall.)

While the focus of the measurements was chosen to be UDP rude/crude, the statistics from SAA and Netsight would also be viewed for correlation (sanity) purposes. The further work section at the end of this report recommends that more formal comparisons of the various measurement tools be conducted in any JANET QoS Phase 2 tests.

4 Test Slot I: March 2nd

The first test slot was the "at risk" period on the morning of March 2nd 2004.

4.1 Goals

The initial plan for the three slots was to use the first slot to test the basic sanity of the rude/crude tools in conjunction with AccessGrid, then use the second two slots for detailed AccessGrid and other application tests.

However, delays in ECS moving its AccessGrid to a new location (for over a month the equipment was "boxed up" due to internal logistics issues) meant that the first test turned out to only be an opportunity to test the basic rude/crude operation (transmission and reception).

It also allowed us to benchmark the potential end-to-end performance with iperf, in this case to ensure that the 100Mbit/s capacity – deliberately limited by local end-site links – was really available.

Finally, we tested that the DSCP field was passed transparently end-to-end on the Imperial – Southampton path.

4.2 Configuration

The Cisco 7206 was configured with the BE/LBE classes as described above in Section 3.

The rude/crude scripts were set to run with a growing BE bandwidth up to and above the point of congestion, with no generated competing LBE traffic, to check that

a) the correct capacity was being reached (100Mbit/s)

b) the packet loss statistics were correctly recorded

The use of these basic tools also allowed the sites to determine the required firewall "holes" that needed to be punched to allow the tools to be run successfully. These have been noted above in Section 3.

.

4.3 Notes

The rude generation script was as follows, for Imperial to Southampton:

```
START NOW
000000 30 ON 5000 quasi.ecs.soton.ac.uk:20001 CONSTANT 7000 1472
009999 30 OFF
010000 31 ON 5001 quasi.ecs.soton.ac.uk:20001 CONSTANT 8000 1472
019999 31 OFF
020000 32 ON 5002 quasi.ecs.soton.ac.uk:20001 CONSTANT 8400 1472
029999 32 OFF
030000 33 ON 5003 quasi.ecs.soton.ac.uk:20001 CONSTANT 8800 1472
039999 33 OFF
040000 34 ON 5004 quasi.ecs.soton.ac.uk:20001 CONSTANT 9200 1472
049999 34 OFF
050000 35 ON 5005 quasi.ecs.soton.ac.uk:20001 CONSTANT 9600 1472
059999 35 OFF
060000 36 ON 5006 quasi.ecs.soton.ac.uk:20001 CONSTANT 10000 1472
069999 36 OFF
```

The gathered crude results for this test stream were as follows:

```
Flow_ID=30
Packets: received=70391   out-of-seq=1   lost(est)=25
Total bytes received=103615552
Sequence numbers: first=0   last=70415
Delay: average = 0.005634   jitter=0.000203   seconds
Absolute maximum jitter=0.013730   seconds
Throughput=1.03746e+07   Bps  (from first to last packet received)

Flow_ID=31
Packets: received=79932   out-of-seq=1   lost(est)=61
Total bytes received=117659904
Sequence numbers: first=0   last=79992
```

```
Delay: average = 0.007659   jitter=0.000180   seconds
Absolute maximum jitter=0.013124   seconds
Throughput=1.17654e+07   Bps  (from first to last packet received)

Flow_ID=32
Packets: received=81236   out-of-seq=1   lost(est)=2790
Total bytes received=119579392
Sequence numbers: first=0   last=84025
Delay: average = 0.024667   jitter=0.000177   seconds
Absolute maximum jitter=0.003924   seconds
Throughput=1.19365e+07   Bps  (from first to last packet received)

Flow_ID=33
Packets: received=81110   out-of-seq=1   lost(est)=7377
Total bytes received=119393920
Sequence numbers: first=0   last=88486
Delay: average = 0.025306   jitter=0.000178   seconds
Absolute maximum jitter=0.001466   seconds
Throughput=1.19399e+07   Bps  (from first to last packet received)

Flow_ID=34
Packets: received=81093   out-of-seq=1   lost(est)=11491
Total bytes received=119368896
Sequence numbers: first=0   last=92583
Delay: average = 0.025403   jitter=0.000179   seconds
Absolute maximum jitter=0.008157   seconds
Throughput=1.19366e+07   Bps  (from first to last packet received)

Flow_ID=35
Packets: received=81135   out-of-seq=1   lost(est)=15010
Total bytes received=119430720
Sequence numbers: first=0   last=96144
Delay: average = 0.025453   jitter=0.000180   seconds
Absolute maximum jitter=0.003741   seconds
Throughput=1.19436e+07   Bps  (from first to last packet received)

Flow_ID=36
Packets: received=81110   out-of-seq=1   lost(est)=18881
Total bytes received=119393920
Sequence numbers: first=0   last=99990
Delay: average = 0.025432   jitter=0.000180   seconds
Absolute maximum jitter=0.008226   seconds
Throughput=1.19402e+07   Bps  (from first to last packet received)

crude: captured/processed 556007 packets
```

The tests for end-to-end DSCP transparency were successful (using '*traceroute –t* 32' for the LBE TOS, which is DSCP 8).

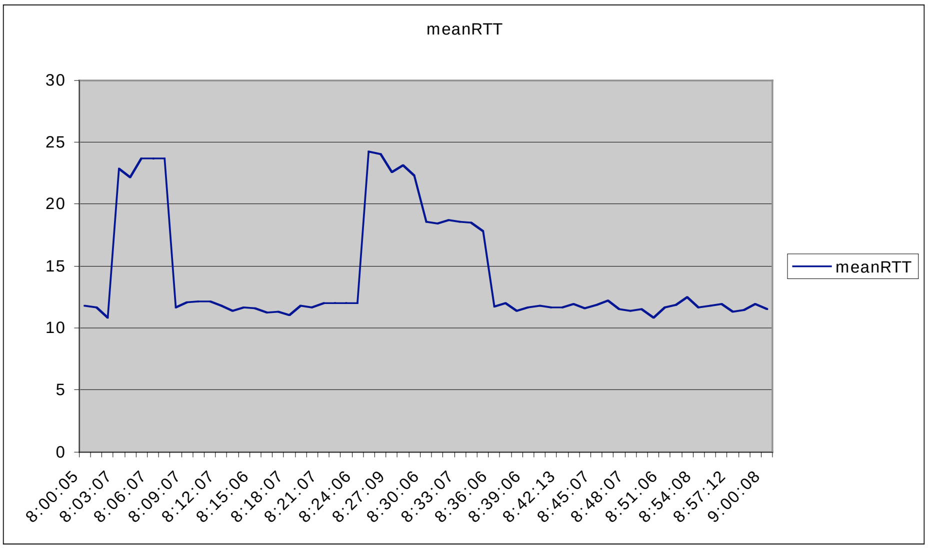
In the case that the traceroute output shows failure, e.g. the TOS value being reset to zero, an error stating *'(TOS=0!)'* would be shown, e.g. as below for a tested path to a node on Abilene:

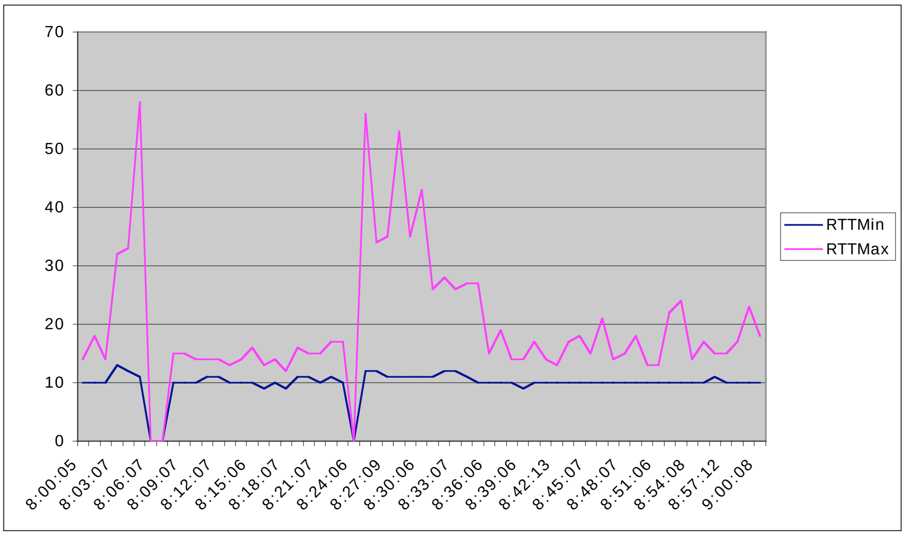
```
$>dtraceroute -t 32 www.internet2.edu

…<snip>…

19 chinng-nycmng.abilene.ucaid.edu (198.32.8.82) 112 ms  107 ms  109 ms
20 iplsng-chinng.abilene.ucaid.edu (198.32.8.77) 129 ms (TOS=0!) 110 ms  122 ms
21 so-0-2-0x1.aa1.mich.net (192.122.183.9)  125 ms  142 ms  122 ms
```

Such resets would impact LBE usage where LBE is used for traffic management on ingress to the target (destination) network.

The end-to-end property of course is not relevant if LBE is being used as a local egress traffic management tool (an internal campus bandwidth management tool).

4.3.1 SAA LBE data





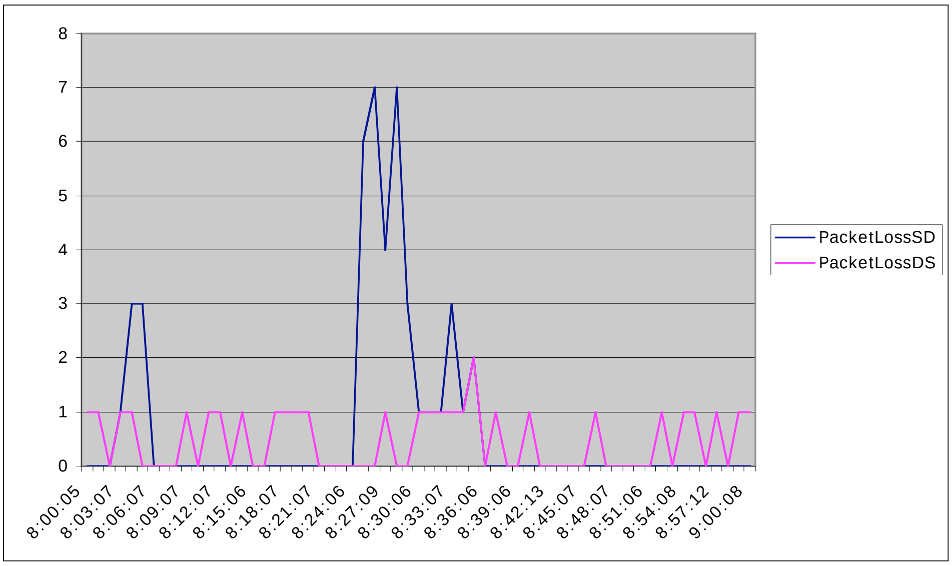


Figure 4-1: SAA LBE data for Test Slot 1

The SAA LBE data shows that the test LBE data is being dropped in contention with BE, but since we were not gathering LBE data with rude/crude, we left the more detailed experiment for the second test slot.

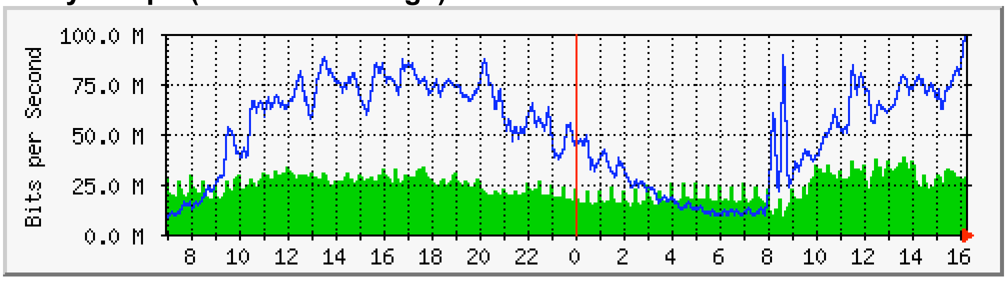
The SAA LBE measurement shows the impact on RTT where LBE congestion is occurring.

4.3.2 Netsight data

The Netsight data for the morning's tests was made available by LeNSE, as shown in the Figure below.

The statistics were last updated **Tuesday, 2 March 2004 at 16:10**, at which time **'SNUR701'** had been up for **177 days, 19:57:50**.

`Daily' Graph (5 Minute Average)



Max In:	39.1 Mb/s (3.9%)	Average In:	24.3 Mb/s (2.4%)	Current In:	28.2 Mb/s (2.8%)
Max Out:	99.8 Mb/s (10.0%)	Average Out:	51.1 Mb/s (5.1%)	Current Out:	99.8 Mb/s (10.0%)

Figure 4-2: Netsight data, Test Slot 1

The Netsight data shows traffic peaks at the expected times, that also correlate correctly with the SAA data.

4.4 Conclusions

The first test slot allowed us to successfully test and verify the operation of the traffic measurement tools and to ensure the tools would be available for future test slot activities with the applications.

The only real weakness of rude/crude is that results must be collected at the receiver; the ability to collect data client side would be a useful enhancement to the tool.

5 Test Slot II: March 9th

In this section we document the tests run in the second test slot of 9th March 2004.

5.1 Goals

The general goal of this test slot, in the continued absence of the AccessGrid node being available at Southampton, was to test best effort (BE) UDP throughput and behaviour as BE is ramped up and down, this time in the presence of a constant UDP LBE background stream.

5.2 Configuration

We used a Linux laptop on the AG network at ECS to be the BE rude source.

We ran two rude streams in parallel, measuring each rude-crude pair side by side, the BE stream coming from the laptop on one internal Cisco QoS interface, and the rude/crude host on the other Cisco QoS interface generating the LBE stream. These two streams are then in contention for the uplink interface towards the campus (and thence to LeNSE, JANET and Imperial).

We planned to ramp BE right up to and beyond the available bandwidth limit (though it would be rate limited by the incoming interface – an alternative experiment would see both the internal interfaces running at 1 Gbit/s to force the contention at the uplink 100Mbit/s interface).

With a constant but significant LBE stream in contention, one expects the LBE to defer to BE, but for LBE to keep its guaranteed minimum bandwidth allocation (5% in the ECS configuration).

After an initial burst for a few seconds we stepped up the BE rate at one second intervals to exceed 100Mbit/s (the congestion point), while keeping the LBE rate constant at around 10Mbit/s. We then have delay, throughput, jitter and out-of-sequence information recorded by crude for each second.

There is a delay in the rude script before the first packets are sent otherwise rude has socket problems (we believe rude needs a second or two to create all the sockets for the separate ID flows). An excerpt of the rude script for LBE generation (TOS=32 / DSCP=8) was as follows:

```
START NOW
002000 30 ON 5100 quasi.ecs.soton.ac.uk:20002 CONSTANT 848 1427
TOS 30 32
009999 30 OFF
010000 31 ON 5101 quasi.ecs.soton.ac.uk:20002 CONSTANT 848 1427
TOS 31 32
010999 31 OFF
011000 32 ON 5102 quasi.ecs.soton.ac.uk:20002 CONSTANT 848 1427
TOS 32 32
011999 32 OFF
012000 33 ON 5103 quasi.ecs.soton.ac.uk:20002 CONSTANT 848 1427
```

Even though the offered load is the same per second, we use one flow per second to allow per second statistics to be gathered at the receiving crude collector.

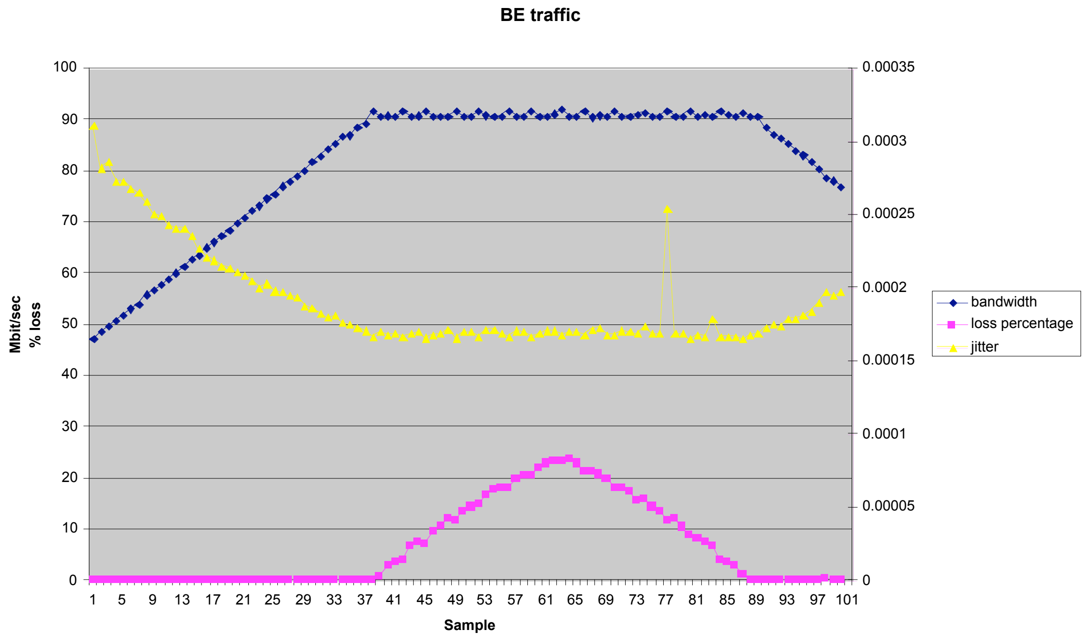
5.3 Notes

The tests were run and the various measurements taken for the tools used.

5.3.1 Rude/Crude data

The following Figure shows the traffic results from Southampton to Imperial, as produced by processing the crude collection data with the Python scripts written by Imperial, and then plotting the resultant extracted data.

BE traffic was 1472 byte UDP payloads at rates of 4000-10200 packets/second increasing at 100 packets/second/second. LBE traffic was a constant background of around 10Mbit/sec (848 packets/second, with fixed size packets of 1427 bytes).



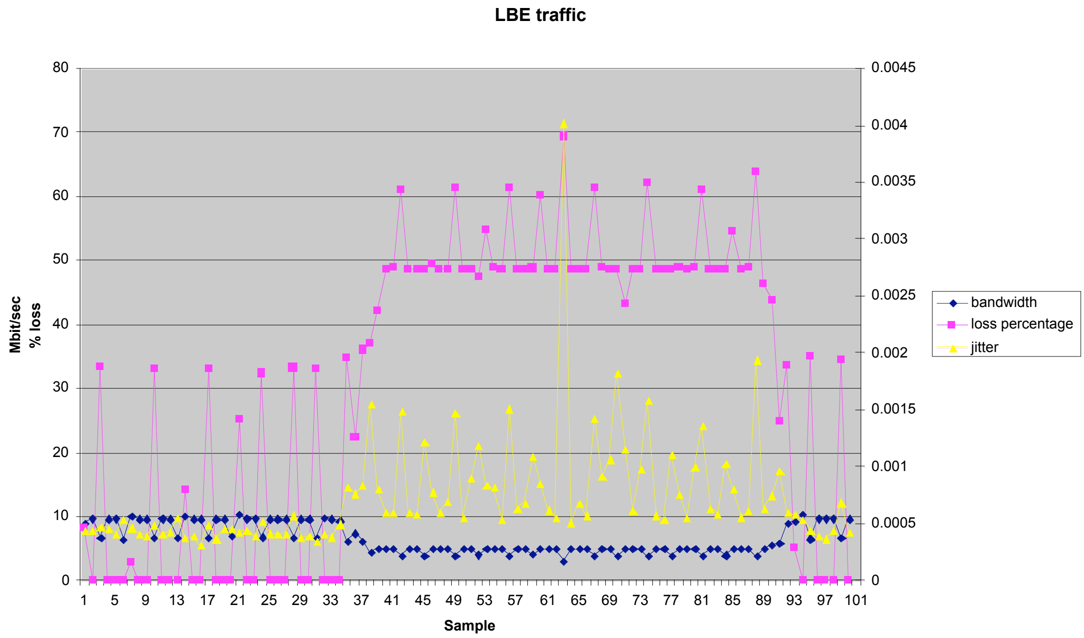
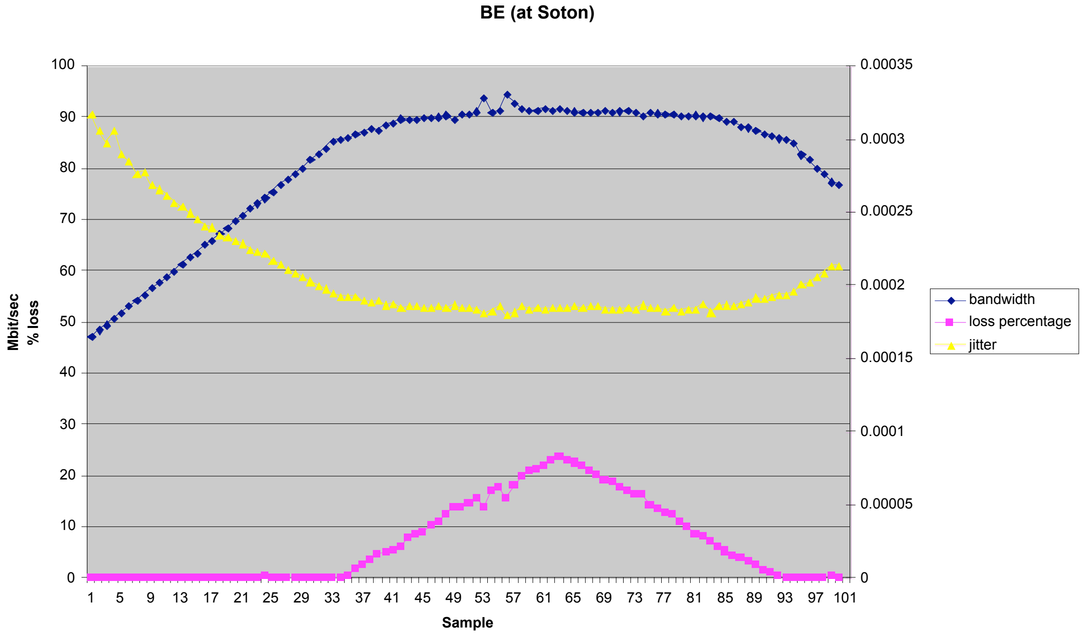


Figure 5-1: Test slot 2, crude data for Southampton to Imperial

The BE traffic rises as expected until congestion occurs on the link, and the LBE drop off correlates well with the congestion of the link being approached, with LBE maintaining the expected reserved bandwidth.

In both cases the LBE traffic drops off at the point of contention (90Mbit/s of BE, against the 10Mbit/s LBE “background” flow), but the LBE maintains its minimum bandwidth allocation (5% of available capacity). This is the desired behaviour.

The following Figure shows the results run from Imperial to Southampton.



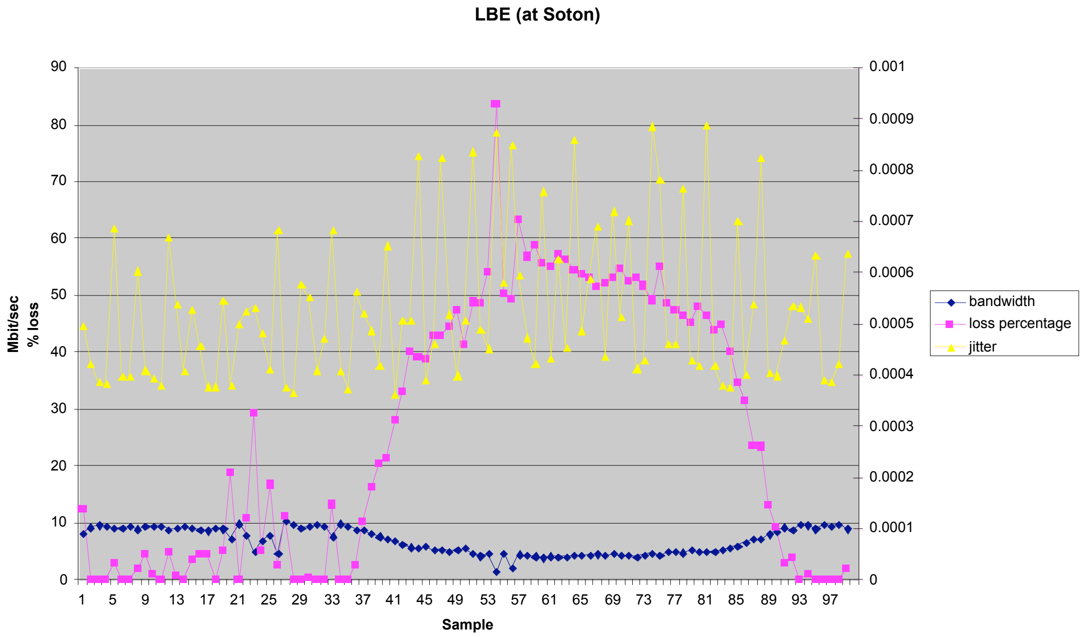


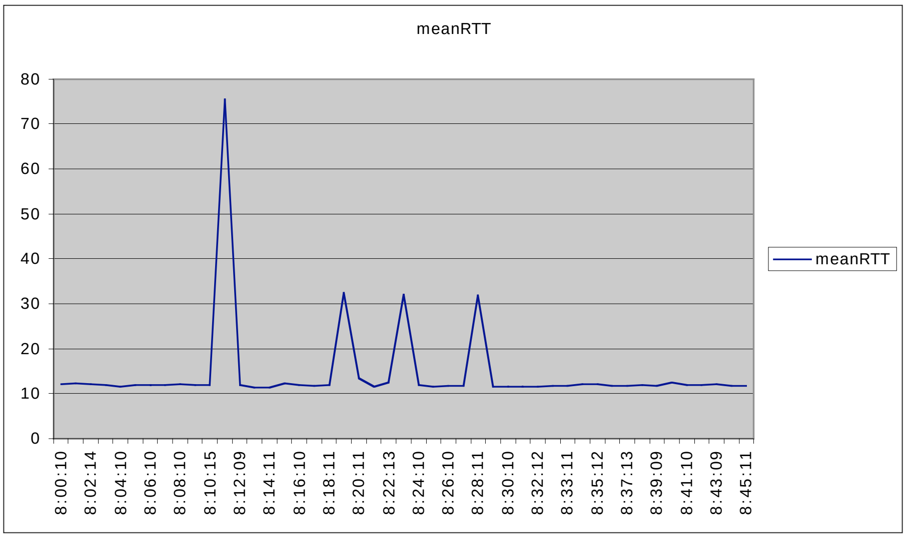
Figure 5-2: Test slot 2, crude data for Imperial to Southampton

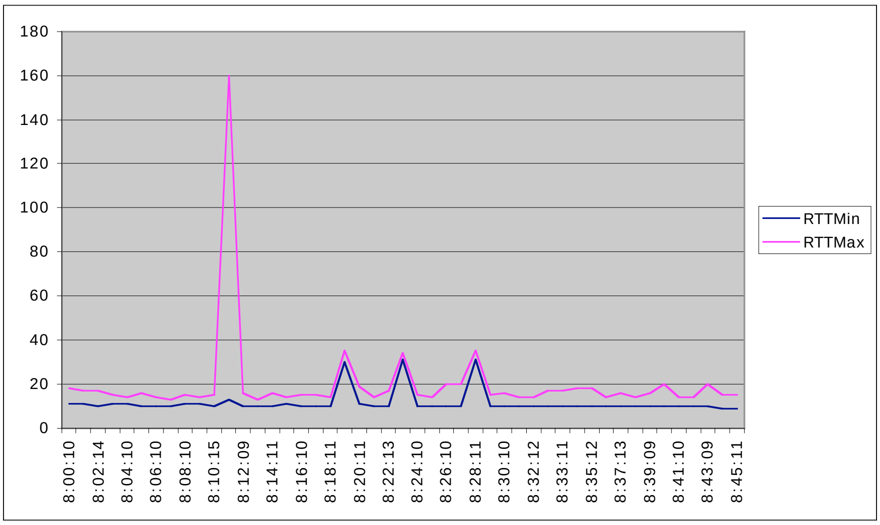
The results are similar to those for Southampton – Imperial, validating the desired behaviour, though the loss for LBE is a little more predictable.

The variation in the detailed LBE behaviour is likely explained by the different hardware being used at each site; the Imperial LBE egress control gives smoother degradation of LBE, while Southampton's egress is more abrupt, with more variation of LBE characteristics under load (where the LBE minimum allocation is being honoured).

This may be worthy of further investigation.

5.3.2 SAA LBE data





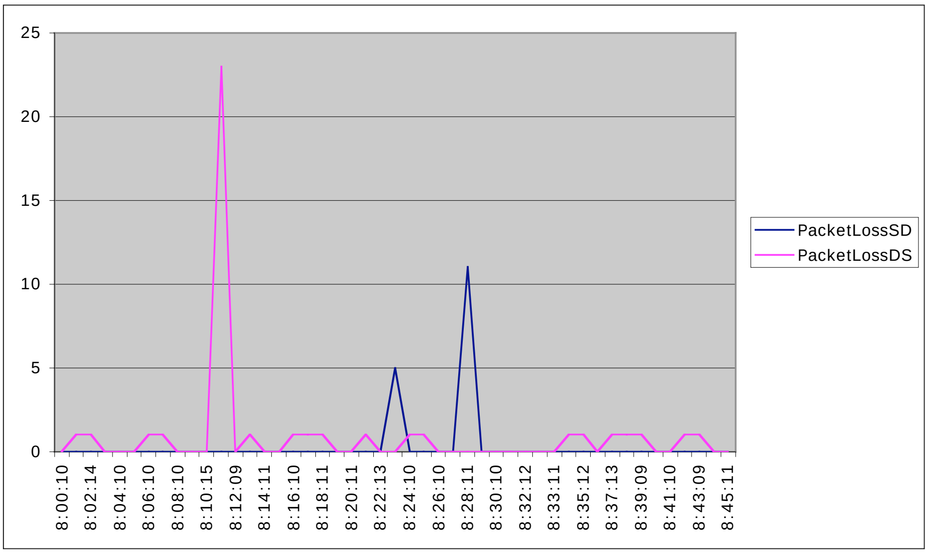


Figure 5-3: SAA LBE data for Test Slot 2

The SAA data correlates with the BE loading test being run in each direction. The second test was interrupted and rerun twice due to an error in the crude collecting script (two flow Ids were accidentally repeated).

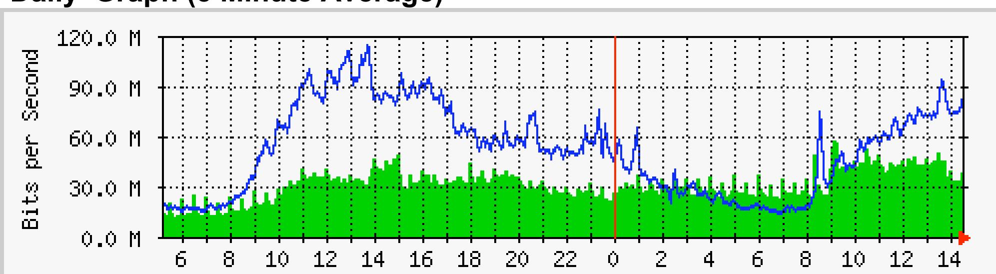
It is interesting to note the lesser LBE packet loss for the Imperial to Southampton path for LBE, particularly given the way the AccessGrid test behaves in test slot 3 for Imperial to Southampton (as described in the next main section). That difference may be because this specific result (for SAA monitoring) may be misleading because the SAA LBE traffic is contending on a 100Mbit/s physical link with – at times – more than 100Mbit/s of BE traffic (because the BE generator was on the application link). Thus LBE does not get its guaranteed minimum allocation; it merely contends on an equal layer 2 footing with BE. In hindsight, this should have been avoided, but there is a useful lesson to be learnt for real-world QoS deployment(!).

5.3.3 Netsight data

Again, LeNSE made the Netsight bandwidth data available for the test period.

The statistics were last updated **Tuesday, 9 March 2004 at 14:30**, at which time **'SNUR701'** had been up for **184 days, 18:17:39**.

`Daily' Graph (5 Minute Average)



Max In:	58.1 Mb/s (5.8%)	Average In:	31.4 Mb/s (3.1%)	Current In:	38.9 Mb/s (3.9%)
Max Out:	114.6 Mb/s (11.5%)	Average Out:	53.1 Mb/s (5.3%)	Current Out:	85.4 Mb/s (8.5%)

Figure 5-4: Netsight data for Test Slot 2

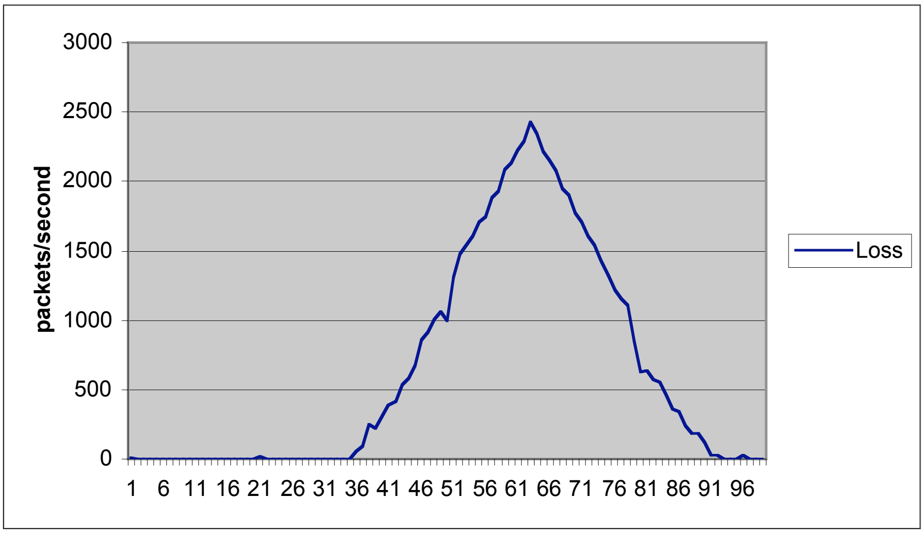
The campus firewall maintainers reported that the Nokia IP700 series firewall was performing correctly during this (and the third) test slot. It had been running with 0% idle CPU during the first test due to a bug in the resiliency code between the primary and secondary firewalls that are actually deployed.

5.3.4 Observation on load and jitter

While gathering the rude/crude data we noted an interesting property that as load increased to the point of congestion, the average jitter fell.

This is illustrated specifically in the next Figure, but is also discernable from the rude/crude charts above.

The project partners discussed the observation of reduced jitter under congestion. The most plausible explanation is that as congestion is reached, most of the packets will be test packets (whereas before some variation would be seen from other traffic on the link), and thus of the same size (because the test traffic packet size was not varied), therefore the same transmission delay for each packet is less variable. However, it may simply be that crude just ignores jitter values for dropped packets, thus skewing the reported figure.



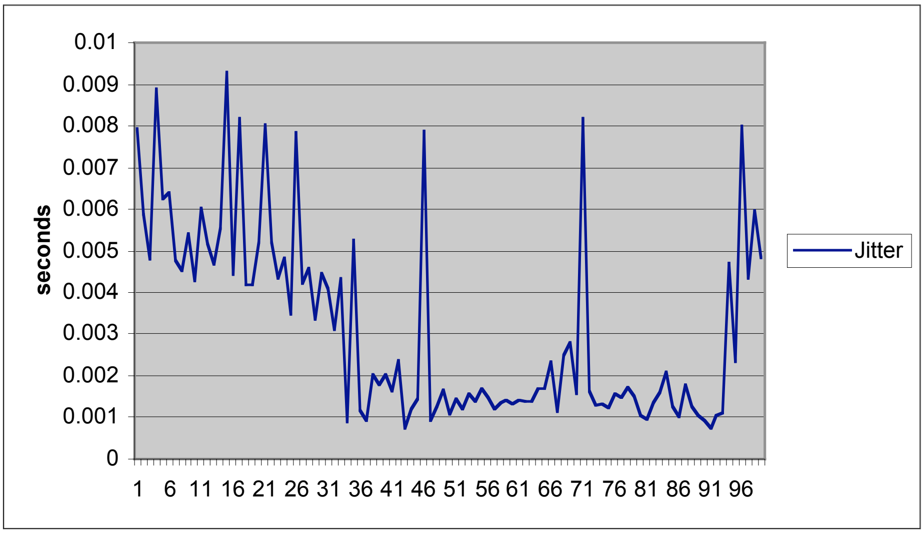


Figure 5-5: Jitter falling under congestion (Test Slot 2)

This may be an issue worth further investigation, by performing tests with varying packet sizes on the test streams.

5.4 Conclusions

The results from BE loading against a constant LBE background were as desired. We were thus confident in the measurement processes, and the BE versus LBE behaviour, going into the final test slot where AccessGrid would be used, and the Multicast "complexity" added.

6 Test Slot III: March 16th

The final test slot, for AccessGrid and LBE background testing, was run on 16th March 2004 in the JANET "at risk" period.

6.1 Goals

The main goal of test slot 3 was to test the AccessGrid (AG) application (as best effort Multicast traffic) with LBE background traffic ramping up and down, using the higher level of LBE background to cause congestion (because the AccessGrid traffic while sizeable, only aggregates to low 10's of Mbit/s on a multiparty conference).

The AG testing would also run for longer than the test slot 2 tests, with a 30 minute AG session and a 10-15 minute LBE flow in competition in the middle of the session.

The usual crude data collection would be used for LBE traffic characteristic measurement.

For AG, when measuring the characteristics, we could

a) Note subjective quality, from the operators

b) Attempt to note objective quality, e.g. using the Multicast AG Beacon, or the reception quality matrix built into RAT, or the RTCP available in the AG MICE tools

Both the reception quality matrix built into RAT and the Multicast AG Beacon appear to have no logging function; this would be worth further exploration for any JANET QoS Phase 2 tests that involve Multicast.

6.2 Configuration

The configuration of the network is as per the descriptions in Section 3.

The AG traffic level averaged around 6-10Mbit/s to one multicast group, based on an aggregate of UDP audio and video RTP streams.

We used the rude generator for LBE, ramping the offered load up and down, with the peak saturating the 100Mbit/s link, the load running from 45Mbit/s to 125Mbit/s.

We observed the values reported on screen for the RAT reception quality matrix as the AG sessions ran.

6.3 Notes

The observed data was as described in the following sections. The first test had LBE background from Imperial to Southampton. The second test had LBE background from Southampton to Imperial.

6.3.1 Rude/crude data

The LBE charts for the LBE data are shown in the following two Figures.

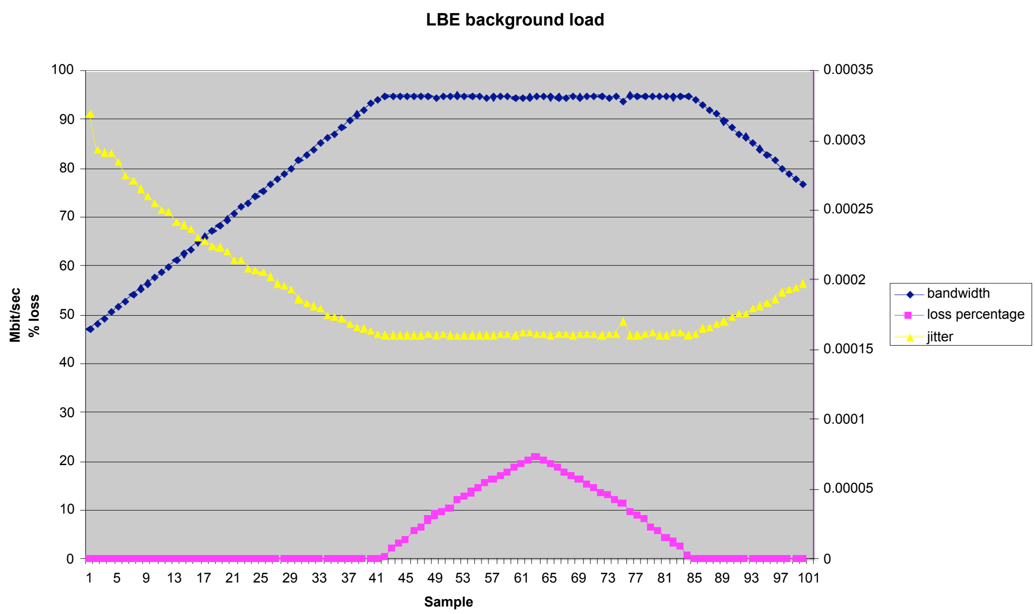


Figure 6-1: Test slot 3, LBE data from Imperial to Southampton

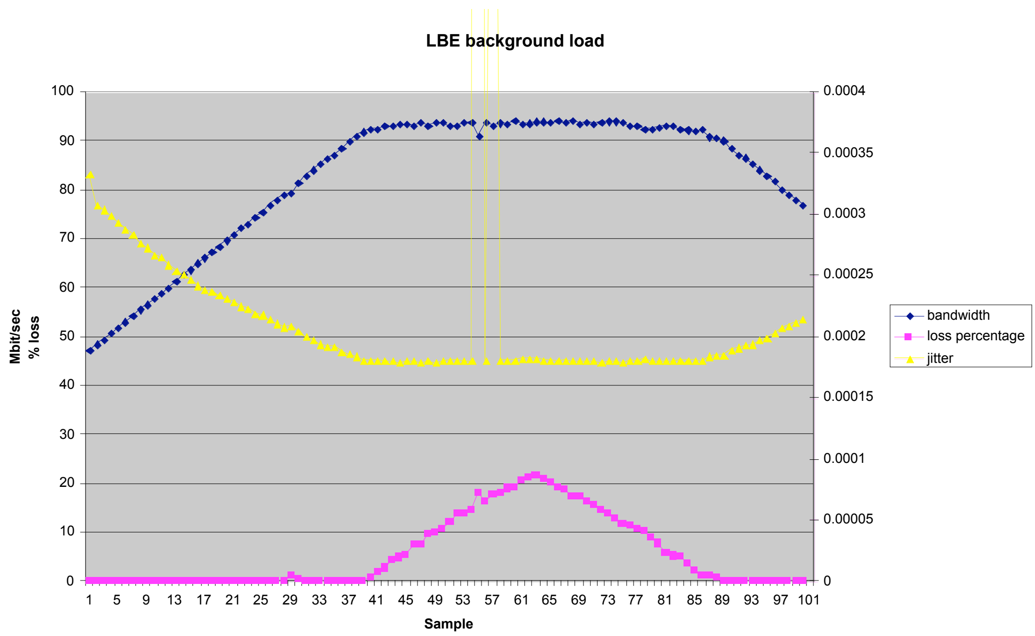
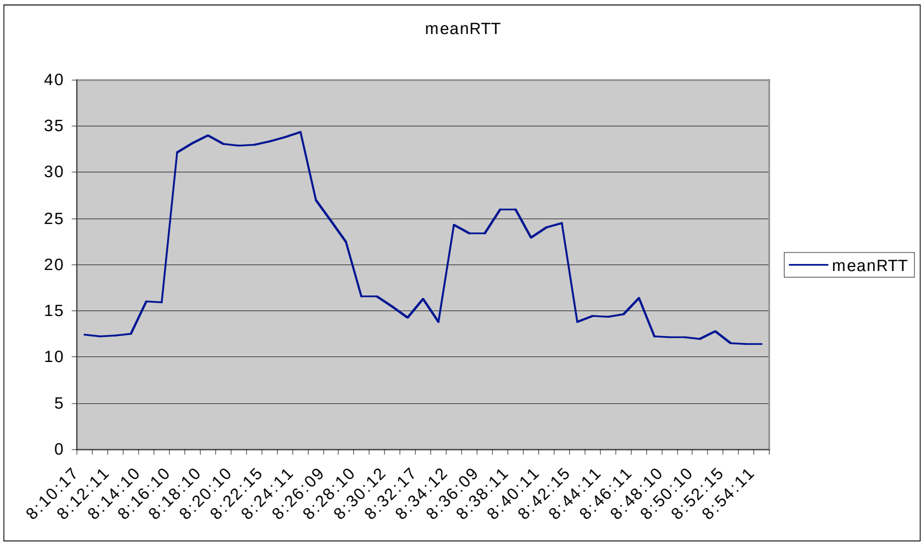


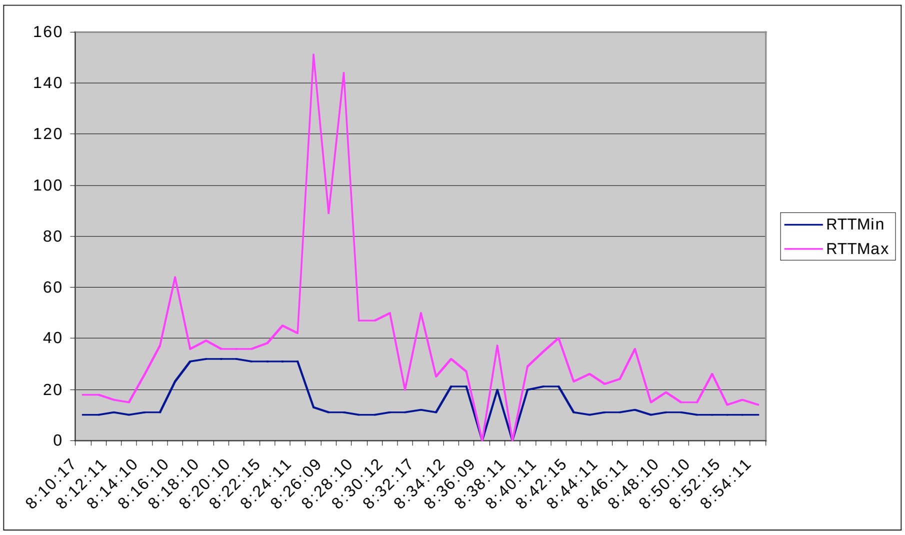
Figure 6-2: Test slot 3, LBE data for Southampton to Imperial

The Figures suggest that in the tests LBE behaved like BE relative to the AccessGrid data. The AG session with the LBE stream from Imperial to Southampton caused significant disruption (indeed complete failure), while the test with LBE sourced at Southampton ran quite acceptably (only minor degradation in the middle.

A clue may lie in the nature of the maximum LBE bandwidth reported above – for the path from Imperial the level is higher and more static, suggesting the link is near-saturated with LBE, with no LBE drop-off. The chart from Southampton suggests a lower level, with more variation as LBE defers to the Multicast traffic. The specifics of how Multicast is handled in a QoS queuing scenario will depend on the hardware used (which is different at Imperial and Southampton). It is clear that further work is required to test IP Multicast in QoS environments, whether LBE, BE or Premium.

6.3.2 SAA LBE data





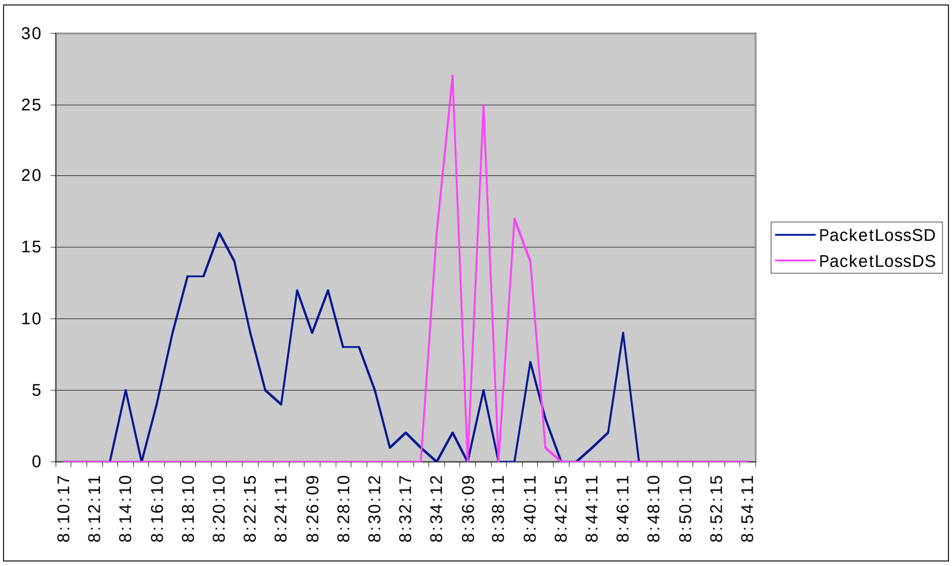


Figure 6-3: SAA LBE data for Test Slot 3

The SAA charts show the Multicast traffic impact on the LBE traffic being generated, through the two AG sessions. In both cases the RTT for (unicast) LBE SAA test traffic rises, and in both cases significant packet loss is observed during the tests.

In this case, unlike test slot 2, the SAA LBE data is not contending at congestion with the application-oriented network traffic.

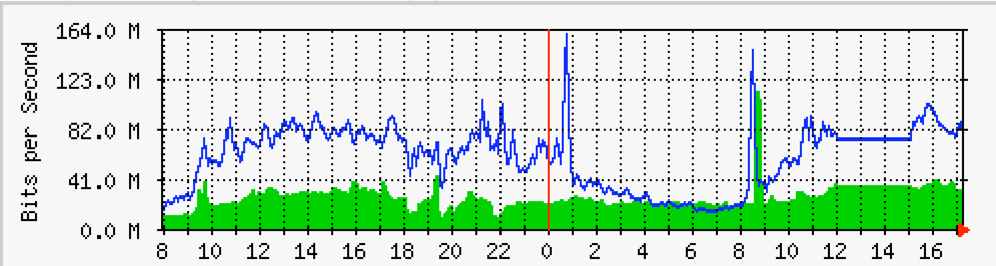
It is difficult to draw firm conclusions from the SAA data. When the contending LBE data is source from Imperial to Southampton, the impact on LBE RTT is greater, suggesting LBE is more adversely affected for the first test. The marked, repeated rise and fall in LBE packet loss on the second test is not readily explained. These measurements are inconclusive is explaining the (very different) AG behaviour and performance during the two tests (contending LBE flowing in each direction). Further tests are required.

6.3.3 Netsight data

Again, LeNSE made bandwidth data available:

The statistics were last updated **Tuesday, 16 March 2004 at 17:16**, at which time **'SNUR701'** had been up for **6 days, 7:20:36**.

`Daily' Graph (5 Minute Average)



Max In:	114.0 Mb/s (11.4%)	Average In:	27.1 Mb/s (2.7%)	Current In:	33.8 Mb/s (3.4%)
Max Out:	160.1 Mb/s (16.0%)	Average Out:	60.3 Mb/s (6.0%)	Current Out:	79.3 Mb/s (7.9%)

Figure 6-4: Netsight data for Test Slot 3

Interestingly the Netsight data shows a traffic load (over a 5 minute average) exceeding 100Mbit/s during the test slot.

This may be due to coincident traffic elsewhere on campus, though such traffic before 9am is unusual.

6.3.4 Subjective assessment

In the first test we generated LBE at the Imperial end. Here, the AG session degraded to (and beyond) the point of collapse.

The test with LBE from Southampton to Imperial was more successful, in that the AG session was still very usable. While the quality fell a little in the middle, it was still good, and certainly did not break up completely.

6.3.5 RTCP

We captured RTCP traffic on the link. This proved difficult to analyse, although the five sources showing on the Southampton receiving end tallied up with the VIC tool being run multiple times on the video system for reception.

For future Multicast AG QoS experiments, logging of RTCP in the application (VIC or RAT) would be highly preferred, and should be investigated.

6.3.6 RAT reception matrix

We observed the RAT reception matrix visually during the AG sessions (because no logging feature was found).

The following Figure shows the matrix over time, with the higher values representing lower quality (a value of 0 is perfect reception).

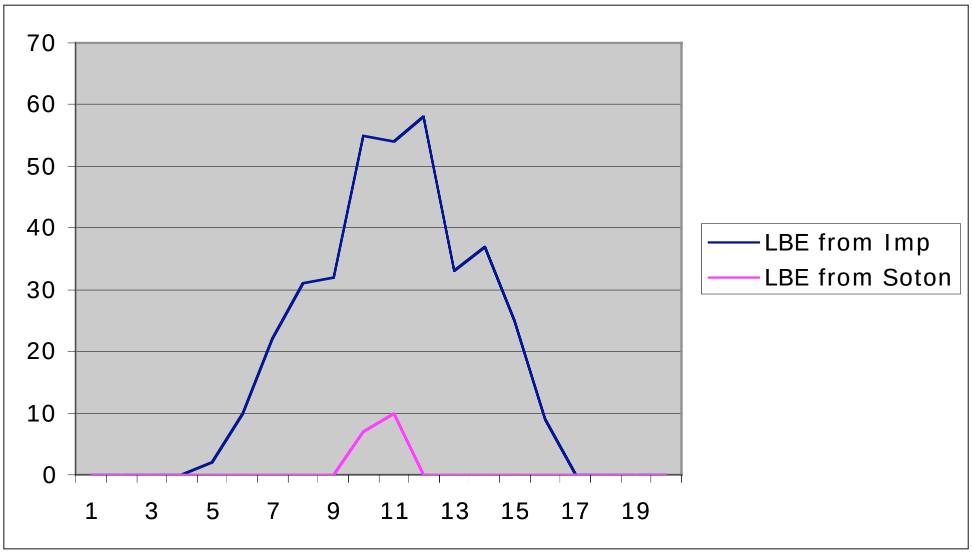


Figure 6-5: RAT reception matrix over time

The chart shows that the LBE sourced at Imperial caused significantly higher degradation than the second session where LBE was generated from Southampton.

6.4 Conclusions

It is clear that the behaviour of IP Multicast needs further investigation in a QoS environment. It does not appear safe to assume that it will be queued and treated the same as unicast BE traffic.

The SAA LBE measurements did not really help the explanation of the AG performance.

The performance of AG traffic in the presence of LBE background depends on the hardware used; the Southampton egress router (Cisco 7206) appeared to handle the congestion in a more desirable fashion.

7 Conclusions

The JANET QoS work has produced many interesting results, and also raised some issues that either caused problems in the scheduled tests, or that would be appropriate for further work in any Phase 2 tests.

Overall we felt the experiments were very valuable, and we have identified a number of follow-up areas (see next section).

7.1 Results obtained

The basic conclusion of the testing is that, for the edge-oriented tests we ran over the real networks used, LBE performed as desired, i.e. LBE deferred to (unicast) BE under congestion yet maintained its minimum aggregate (set to 5% in our tests) under congestion.

However, the results were not as desired when IP Multicast was tested against an LBE background stream. Here the performance is dependent on the specific hardware being used (because of possible differences in the way IP Multicast may be queued in comparison to how unicast BE traffic is queued).

The monitoring tools used appeared to give sane and generally corroborating results for the first two test slots, though the SAA LBE daa for test slot 3 appeared less useful in fully understanding the observed actual AG behaviour. In test slot 2, the competition at congestion between BE and LBE at Layer 2 (the 100Mbit/s switch) is likely to have "skewed" the SAA collected data a little.

For the first two tests this included rude/crude, SAA and Netsight. For the final AccessGrid test, we also used the RAT reception quality matrix, which was useful, even if no built-in logging appeared to be available.

The effort in deploying the monitoring tools should not be underestimated.

7.2 Specific items not tested or investigated

There were many things we could have tested, given more slots and resources (person effort). These included:

- Using variable packet sizes in the test streams
- Passing traffic through the ECS firewall as well as the campus firewall
- Using Gigabit Ethernet internal interfaces to force congestion only at the egress router (of the Cisco 7206). This would mirror more the typical Further Education college deployment of FE internally facing and a 2-8Mbit/s uplink.
- Related to the above, ideally the SAA device would have been on a separate Cisco interface, to force queue behaviour analysis on the Cisco (rather than a 10Mbit/s HD port on a 100Mbit/s capacity Layer 2 switch). Though the reality of course is that Layer 2 *is* an issue for QoS.
- We did not check the DSCP values observed on our production network to understand where LBE may already be in use by applications.

- There were other monitoring tools and test suites we would have liked to have used, e.g. IBM's AWM.

Further suggestions for future work appear in the final section below.

7.3 Problems encountered

We encountered a number of problems, some logistical, others more technical.

In pre-tests to the first test slot there were some unusual packet delays. This turned out to be a problem on the campus firewall (that was certainly resolved before Test Slot 2).

We did not realise that the Cisco 805 SAA device could only run 10Mbit/s half duplex. This caused some confusion.

Southampton's AG node was not deployed until just before the second test slot; due to confusion with Imperial, inter-site AG testing was then only possible for the final (third) test slot.

8 Further Work

There are many pieces of further work that could be undertaken in any subsequent JANET QoS tests Phase 2.

Some of these items arose from general discussion in the QoS project, rather than directly from our own LBE-oriented tests.

These work items include (in no particular order of importance):

- More studies of monitoring tools and their comparison – rude/crude, 6QM, ripe-ttm, iperf, SAA, RTCP, netsight, etc. This implies the tools are deployed in a meaningful location for the tests, which may be a problem in some cases.

- More studies of multicast and QoS – this is useful for AccessGrid and other advanced collaborative environments.

- Testing of IPv6 QoS, including deployment of IPv4 and IPv6 QoS together (e.g. Premium IPv4 alongside Premium IPv6). It may be possible the the IPv6 Flow Label field may be beneficial here.

- Better understanding of the handling of QoS by firewalls, particularly when the firewall is under load from regular traffic.

- Modifying rude/crude to report statistics to the generating client, rather than having to collect at the receiver (amongst other things, making the issue of guest access to remote machines go away).

- Testing on low bandwidth links and lower-end routers, e.g. Cisco 2600 series routers on 1-8Mbit/s links. This is important for FE college adoption of QoS. We understand the FE colleges will be supplied with Cisco 2691's by UKERNA, for the next three years.

- Understanding the relationship of QoS methods with other campus bandwidth management services (perhaps liaising with the JANET CBMS team at Manchester).

- Application-oriented work for LBE, demonstrating how applications can be modified to mark their packets as LBE.

- Running TCP-based LBE tests.

- Using variable sized packets in the test traffic.

- Building logging functionality into Multicast/AG measurement tools like the RAT reception quality matrix, and the AG Multicast beacon statistics.

- Repeating the tests using 1Gbit/s on internal egress interfaces with lower (100Mbit/s) interfaces on egress to the campus/MAN.

- Studying how a Premium IP Multicast service could be deployed.

- Studies into the DSCP values observed on production networks, and which applications are marking the packets (including LBE).

- Better understanding Layer 2 QoS issues, within the campus (i.e. not just QoS applied at Layer 3 ingress/egress).