Kentish MAN QoS Report

Background

The link from JANET to Kentish MAN was a single 622Mbps connection prior to the SJ5 upgrade and throughout the duration of this trial. The University of Kent (UoK) connects directly to the Kentish MAN and currently hosts the Mirrors Service. A great concern existed with the Mirrors traffic as it was consuming a high percentage of bandwidth over the Regional Network's link to Janet and there was the potential for Mirrors to starve other traffic present. This network was seen as an ideal choice for participation in the QoS trials with UKERNA.

Aim

To configure (and confirm configuration with appropriate tests) QoS on the Kentish MAN network such that non-mirrors traffic would take precedence over mirrors traffic in the event of congestion and that during congestion the mirrors traffic would receive a guaranteed minimum bandwidth to prevent the service grinding to a halt.

QoS Model

The applied QoS will be based on the Diffserv model. Diffserv works by classifying packets into "classes" and marking the packets accordingly from the edge and then making decisions based on that class on a per hop behaviour (PHB).

The Classes associated with this trial are:
Premium (Expedited Forwarding) – strict priority service
Best Efforts – provides a delivery when possible
Less Best Efforts – moderate bandwidth guarantee

Premium	DCSP value 46 - Cos value 5
BE	DSCP value 0 - Cos value 0
LBE	DSCP value 8 - Cos value 1

Network Control (DSCP 48 and 56) were not re-written and will remain unchanged as this class was of no concern for this test.

Monitoring

Initially, netflow monitoring was configured on all core routers to determine the dscp values being generated throughout the network before remarking all traffic to best efforts.
The following DSCP values were detected-
Best Efforts
Near premium
Network Control
Others - Unclassified

No LBE or Premium DSCPs values were found.

The next step was to remark all traffic into the Kentish MAN core to BE and to honour all traffic from the UoK domain as the mirrors traffic was remarked to LBE and all other traffic to BE. Once traffic was remarked within the RN core, it was monitored for a short period to ensure there would be no repercussions.
To achieve the remarking of traffic within the RN, traffic was remarked at ingress. This was done using policy maps as seen below:

Remarking all traffic to Best Efforts

Global configuration

```
access-list 110 remark "QoS remarking"
access-list 110 permit ip any any

class-map match-all remark
  description "Remarking Traffic to Best Efforts"
  match access-group 110

policy-map QoS
  class remark
   set dscp default
```

Interface configuration

```
service-policy input QoS
```

Honouring traffic from UoK

Interface configuration

```
mls qos trust dscp
```

Queuing

The QoS domain consists of five 6500s, each with dual Supervisor Engine 720-3BXLs. The main line card associated with this trial is the WS-X6816-GBIC, a 16 port Gigabit Ethernet fibre LAN module. Within each port there are 3 transmit queues: 2 standard priority with 2 drop thresholds each and 1 strict priority.

Queue Id	Scheduling	Num of thresholds
1	WRR low	2
2	WRR high	2
3	Priority	1

Weighted Round Robin (WRR) was chosen as the scheduling algorithm on the hardware output queues as this guarantees a minimum amount of bandwidth to each class of service and allows the possibility for each class to use the entire link's available bandwidth if the other classes are not using it. The guaranteed percentage of bandwidth for LBE is intended to keep established TCP sessions going even during congestion periods so the hit on the traffic would not be as severe.

WRR was configured, mapping LBE to the low priority standard queue threshold 1 and BE to the high priority standard queue threshold 1. Premium traffic was mapped to the priority queue by default but as this class was not relevant to the test it was disregarded. Egress WRR scheduling was configured to assign appropriately bandwidth proportions to each class [this can only be done on standard transmit queues].

Queue	thresh	cos-map
1	1	1
1	2	2 3
2	1	0 4 6
2	2	7
3	1	5

Traffic Generation and Capture

Real-time UDP Data Emitter (RUDE) and Collector for RUDE (CRUDE) were used to generate, send and capture traffic streams. RUDE generates the traffic which is then received and logged by CRUDE. All traffic streams are UDP based.

Testing

11th July

Objective

To configure WRR ratios to an acceptable level for each class. Also to ensure the ratio values are honoured and to monitor the affect on each, for traffic crossing the 622Mbps link between Kentish MAN and Janet.

WRR Ratios:

10% was assigned for LBE and 85% to BE. The priority queue would be policed to 5% but as there was no priority traffic on the network this could be disregarded.

When the test began, approximately 200Mbps of background traffic was detected. The rude file was configured to generate the following over a 10 minute period:

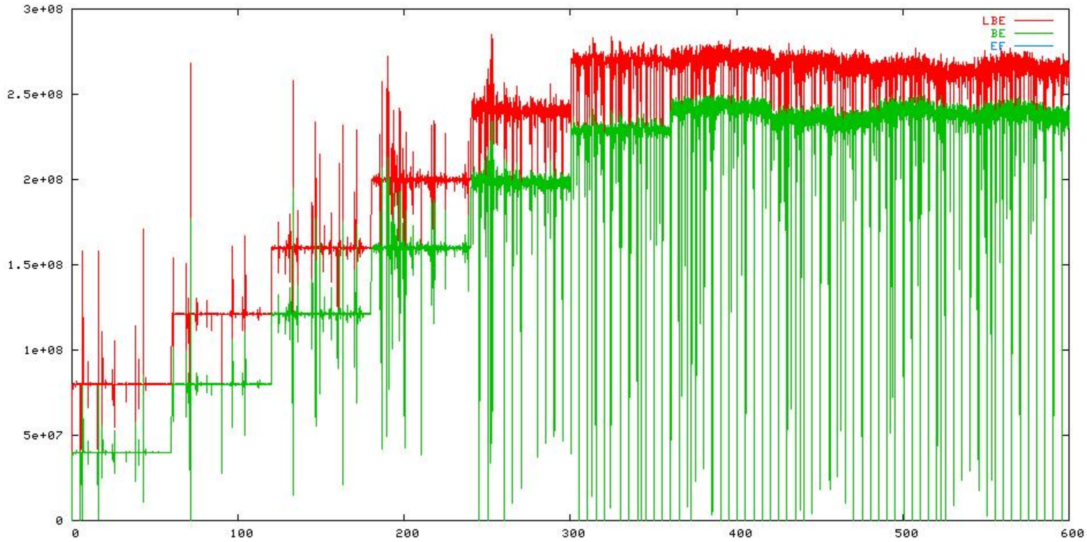
Time (seconds)	BE Bandwidth (Mbps)	LBE Bandwidth (Mbps)	Total Added Bandwidth (Mbps)
0	40	80	120
60	80	120	200
120	120	160	280
180	160	200	360
240	200	240	440
300	240	280	520
360	280	320	600
420	320	360	680

480	360	400	760
540	400	440	840
600	off	off	off

As this test would be generating more than 840Mbps of traffic and 200Mbps of background traffic this would be enough to saturate the 622Mbps link.

Results:

<u>Graph generated from CRUDE output</u>



The output from this graph suggests there was a problem as the rude file generates streams of up to 840Mbps but yet the graph shows the achievable throughput is less than 500Mbps. An explanation for this could be one of the following:
1) The traffic was produced but dropped during transit, either by the network failing to cope with the load, or because of QoS configuration not working correctly.
2) The Manchester server cannot capture the data fast enough
3) The required levels are not actually getting produced by the originating server.

The most likely answer would be the third option as the server might not be capable of generating the desired output. Although the servers at each end have gigabit interfaces there might have been issues with hardware, software, kernels or ip stacks. A recommendation drawn from this test for the next one was to divide the traffic load across two servers.

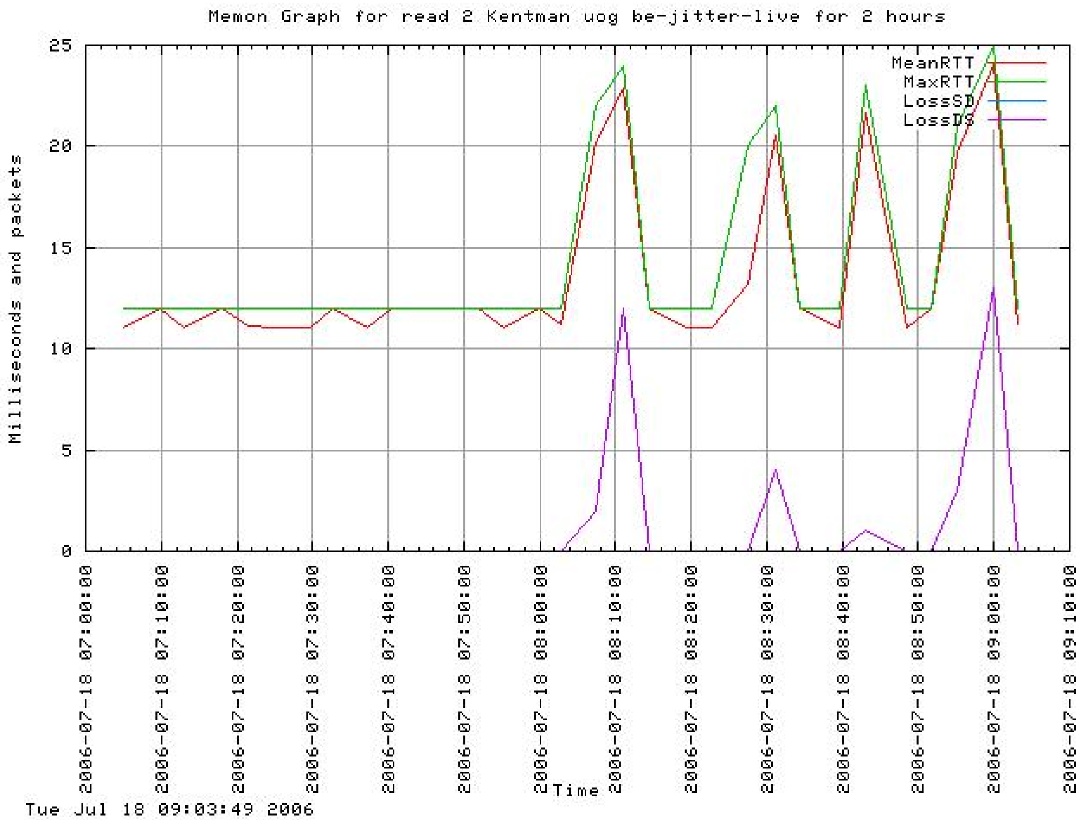
The QoS monitoring graphs have been lost, but from recollection the results from the graphs showed no distinction between any of the traffic classes which should not have been the case. This was investigated and a problem was discovered relating to a policy map statement on the interface to JANET which meant all traffic coming back into the Kent MAN QoS domain was being remarked to BE, this traffic should have been honoured, and as a result the test was inconclusive.

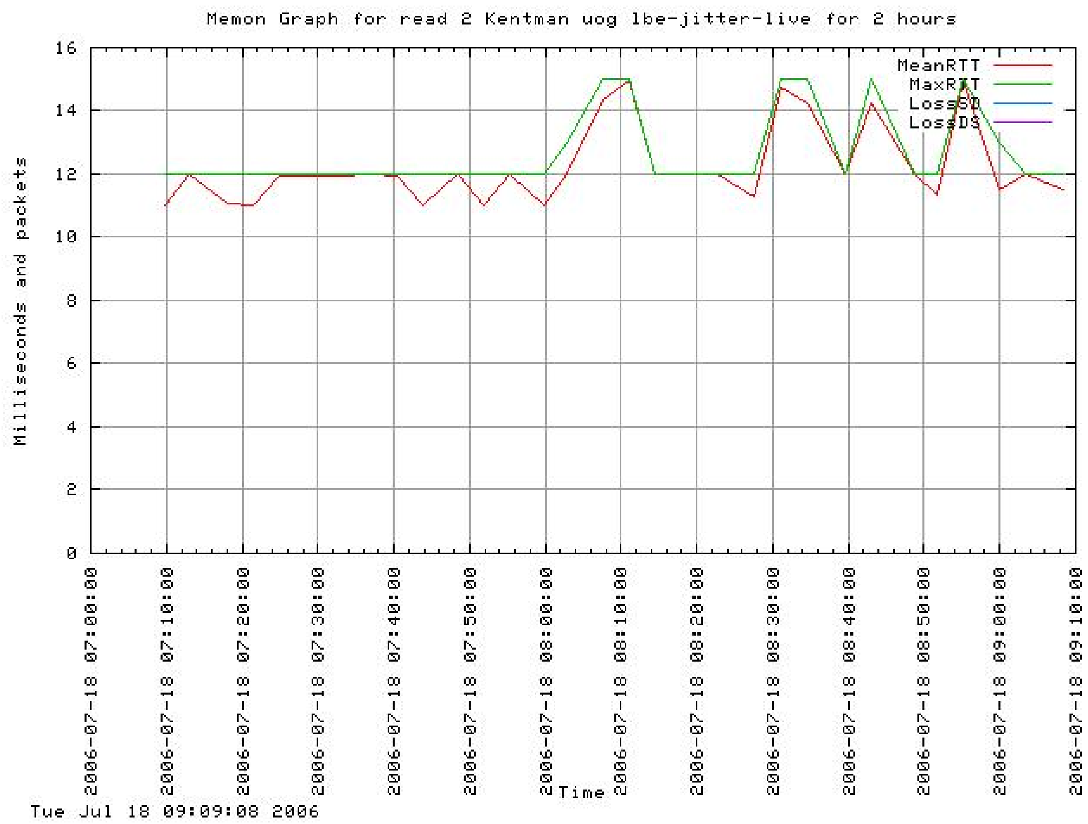
July 18th

This test was exactly the same as the previous week except for the policy map being removed from the interface to JANET, subsequently all traffic traversing this link would be honoured.

Results:

Unfortunately, the server's disk drive at Manchester was full to capacity and as no data was collected, no crude graphs were generated. However, the QoS monitoring graphs were generated and analysed.
Looking at the example of the two graphs below that were taken from the QoS monitoring site, it is clear that the affect seen on BE traffic is much greater than that of LBE. The RTT for BE doubles in time but yet there was only a marginal increase for LBE. No loss could be seen for the LBE traffic but yet there was a fairly significant loss for BE.
The assumption made from this graph was the values allocated to the ratio is inappropriate, 10% to LBE would appear to be excessive.





3rd October

<u>Objectives</u>

- To saturate the 622Mbps link by adding sufficient LBE to the background traffic, in an attempt to demonstrate that LBE can use all the available bandwidth.
- To test that, when the link is fully saturated, the throughput of LBE will drop as an increasing amount of BE is introduced. If successful, this would demonstrate that BE has a higher priority than LBE.
- To demonstrate that whilst the link is saturated with BE traffic, the minimum bandwidth guarantee configured for LBE is honoured.

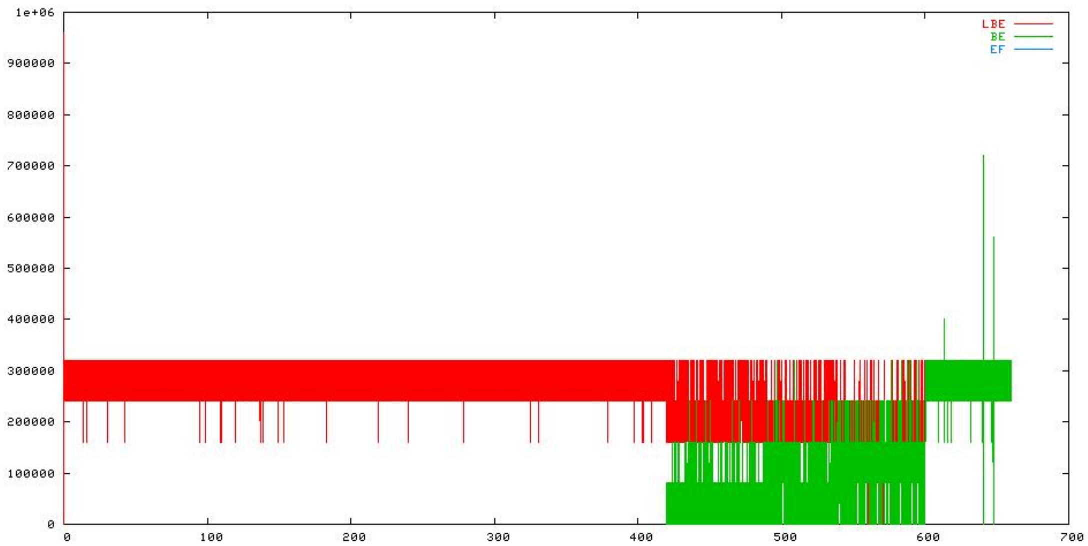
The first test ran for a period of 10 minutes starting with 250Mbps of LBE, incrementing to 500Mbps over a 5 minute interval and maintaining this traffic level for a further 5 minutes. BE was introduced on the 5th minute starting at a rate of 100Mbps and incrementing to 500Mbps for the remainder of the test.

The second test ran over a twenty minute period with the same throughput for LBE as the first test and starting the flow of BE traffic on the 10th minute incrementing from 0Mbps to 400Mbps.

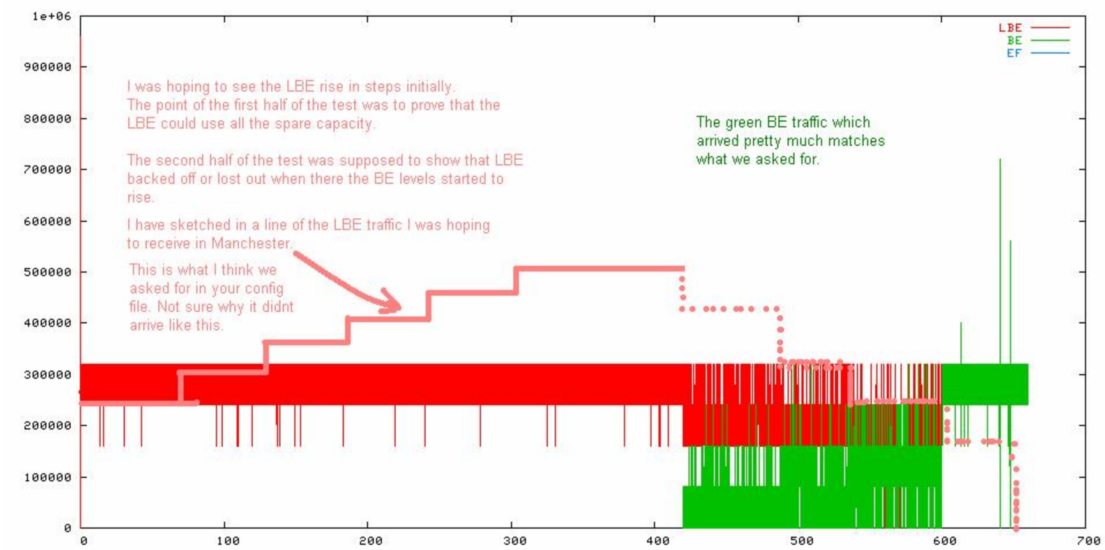
Results:

The graph below suggests that there may have been a problem as the maximum traffic received by CRUDE was only 300Mbps but should have achieved at least 900Mbps on the 20th minute. Both servers generating the streams were investigated by looking at the throughput on the monitoring (MRTG)graphs, these graphs reflected the output expected. Taking a closer look at the CRUDE server at Manchester, it was discovered that the firewall was limiting this server's ability to collect the RUDE data, which explains the behaviour below. The recommended action for the next test was to ensure that the firewall is disabled while the tests were undertaken.

<u>Graphs generated from CRUDE output</u>



This graph illustrates the expected output. If the firewall had been disabled then it is likely that QoS would have been working as anticipated.



10th October

<u>Objective</u>
The same as the previous tests on the 3rd October with the firewall on the server at Manchester being disabled.

The first test ran for a period of 10 minutes starting with 250Mbps of LBE, increasing to 500Mbps over a 5 minute interval and then maintaining this traffic level for a further 5 minutes. BE was introduced on the 5th minute starting with a 100Mbps and incrementing to 500Mbps for the remainder of the test.

The second test ran over a twenty minute interval with the same throughput for LBE as the first test and starting the flow of BE traffic on the 10th minute incrementing from 0Mbps to 400Mbps.

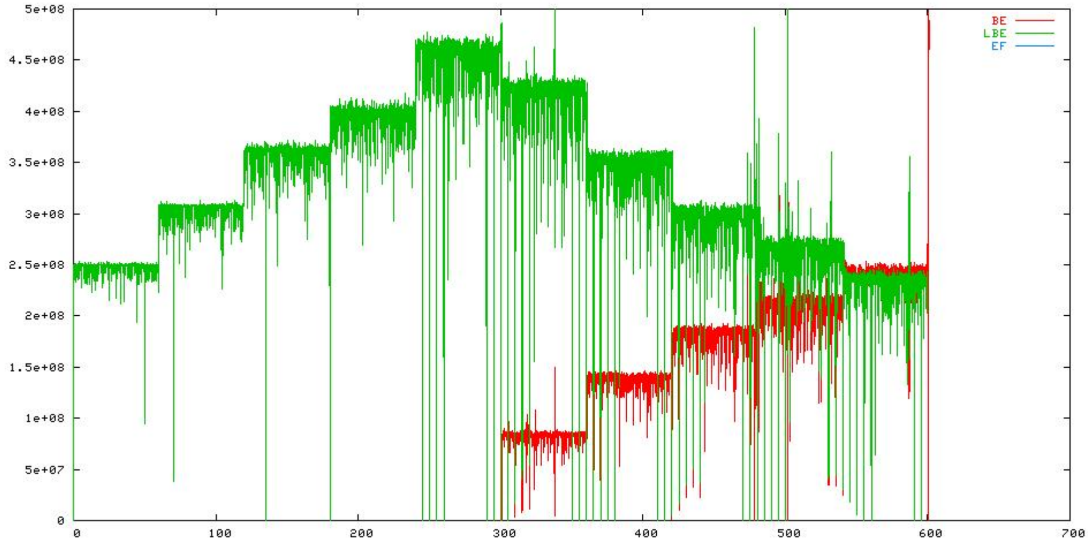
NOTE: During the second test, the maximum crude file size limit was exceeded and so was unable to capture the data for the entire 20 minutes.

Results:

The monitoring graphs in particular, UoG showed a severe affect on BE traffic. RTT for BE increased dramatically and packet loss was also visible. LBE RTT increased slightly at the beginning of the test but did not appear to have such a significant affect as BE. The other three graphs looked very similar and did not distinguish between the classes very much. The monitoring graphs for the second test and again in particular UoG showed that RTT did increase but it was not as jittery as the previous test. Packet loss seems to be lower in comparison to the first test. Once again the other graphs did not really show any differentiation between the classes but RTT and packet loss did increase.

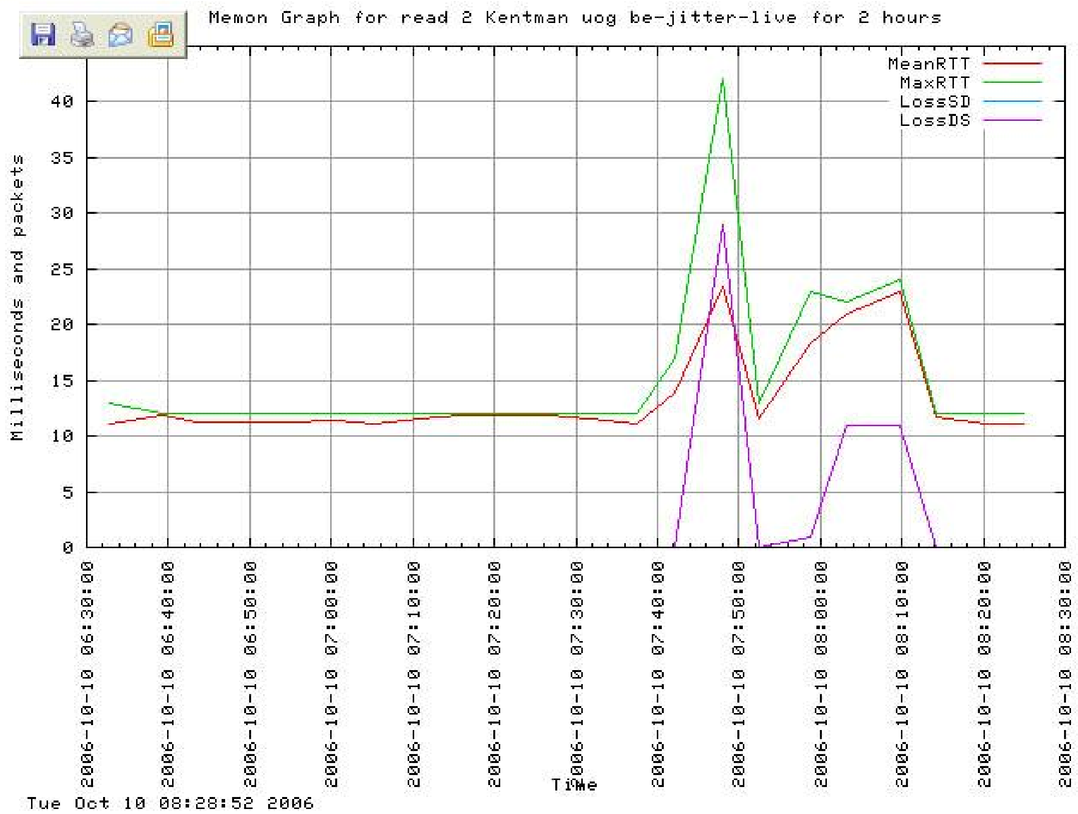
The results from the monitoring graphs are in conflict with the output from CRUDE which, as expected, shows downward spikes for LBE and when the link was fully congested it was LBE traffic that was being dropped and not BE. This was quite apparent for when BE traffic was increasing, it was able to use the bandwidth required but LBE was not.

Graph generated from CRUDE output



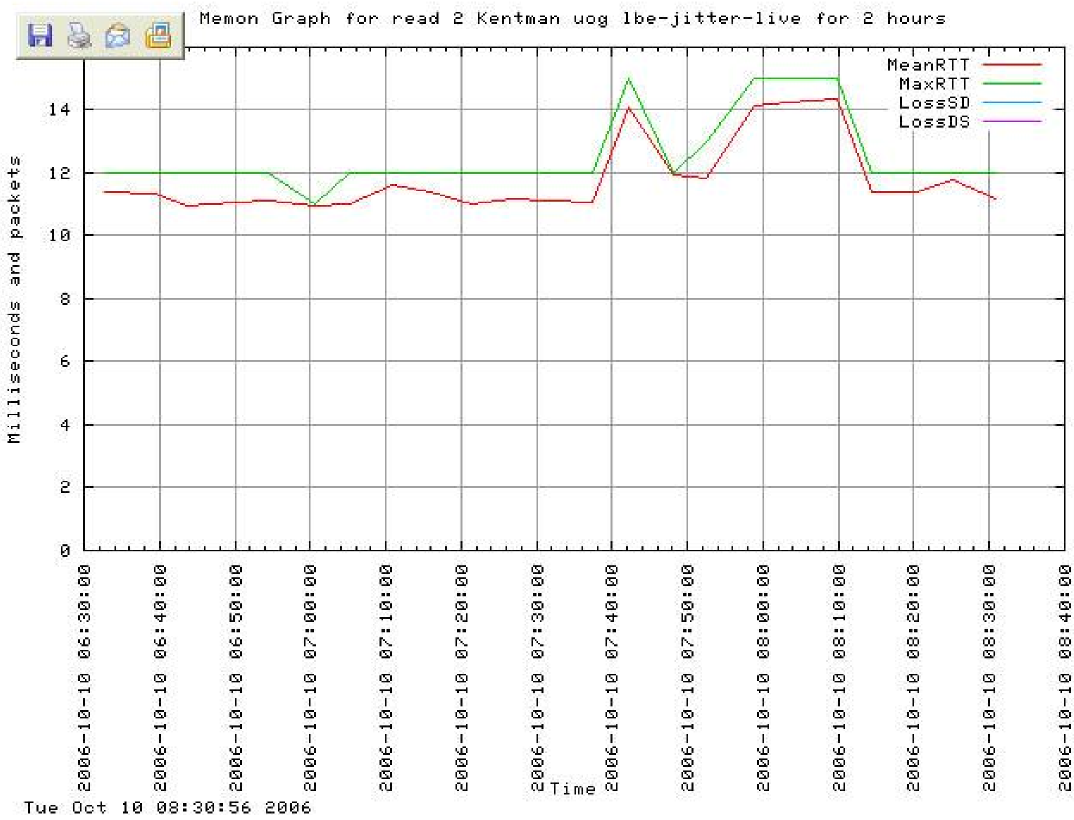
JANET Measurement

The dataset is: QoS Project
The duration is: 2 hours
Tablename is read_2_Kentman_uog_be
Graph type is Normal



JANET Measurement

The dataset is: QoS Project
The duration is: 2 hours
Tablename is read_2_Kentman_uog_lbe
Graph type is Normal



24th October

Objective

- To configure WRED and analyse what improvement (if any) this makes to a TCP streams' performance
- To show that LBE does not get starved out completely when BE levels increase

The first test was not carried out as there was not sufficient time within this test to start analysing WRED thresholds.

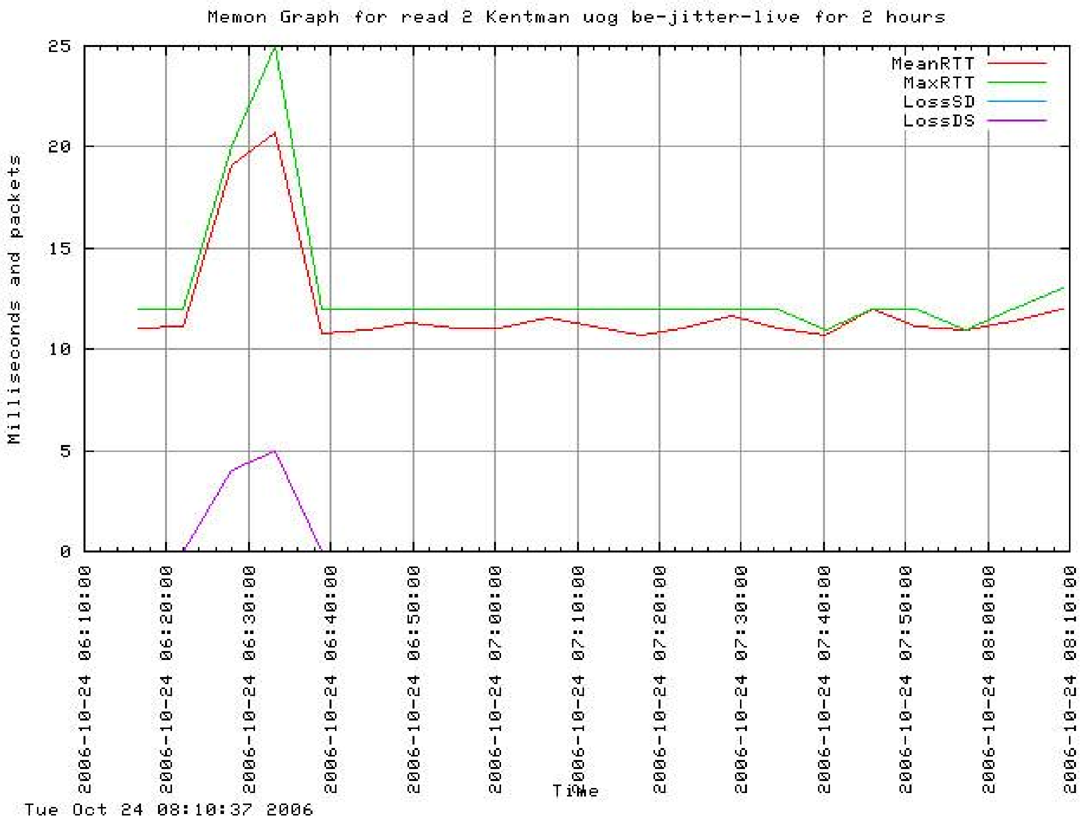
The next test was for a 20 minute period and started with BE traffic generating 250Mbps, increasing to 500Mbps over a 10 minute period, this amount was then sustained for a further 10 minutes. (This would show that BE was able to use all the available bandwidth). On the 10th minute, 100Mbps of LBE was introduced with this figure steadily increasing to 500Mbps over the remaining 10 minutes.

Results:

Unfortunately the server that collects the data for CRUDE had insufficient capacity and as a result there was no information available to analyse that data. The monitoring graphs were available and are shown below. They are quite similar to those generated previously and are difficult to interpret as they suggest that QoS was not working, although from the previous graphs generated by CRUDE it would suggest the opposite.

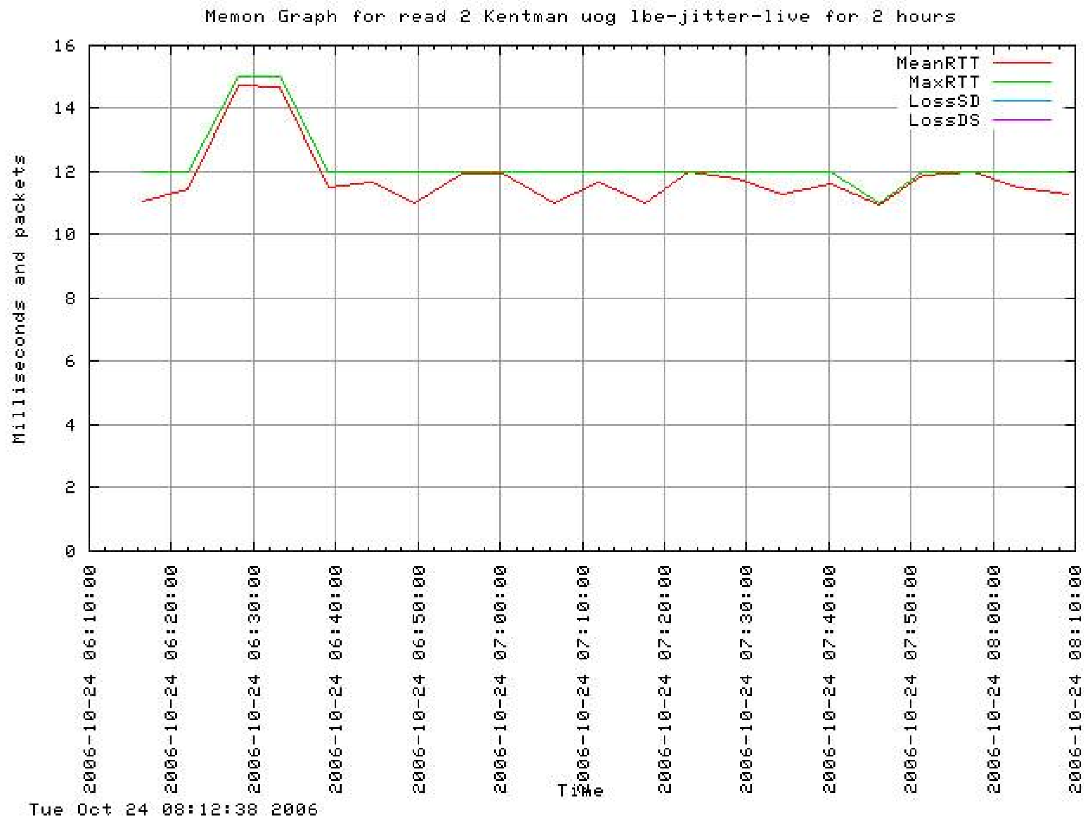
JANET Measurement

The dataset is: QoS Project
The duration is: 2 hours
Tablename is read_2_Kentman_uog_be
Graph type is Normal



JANET Measurement

The dataset is: QoS Project
The duration is: 2 hours
Tablename is read_2_Kentman_uog_lbe
Graph type is Normal

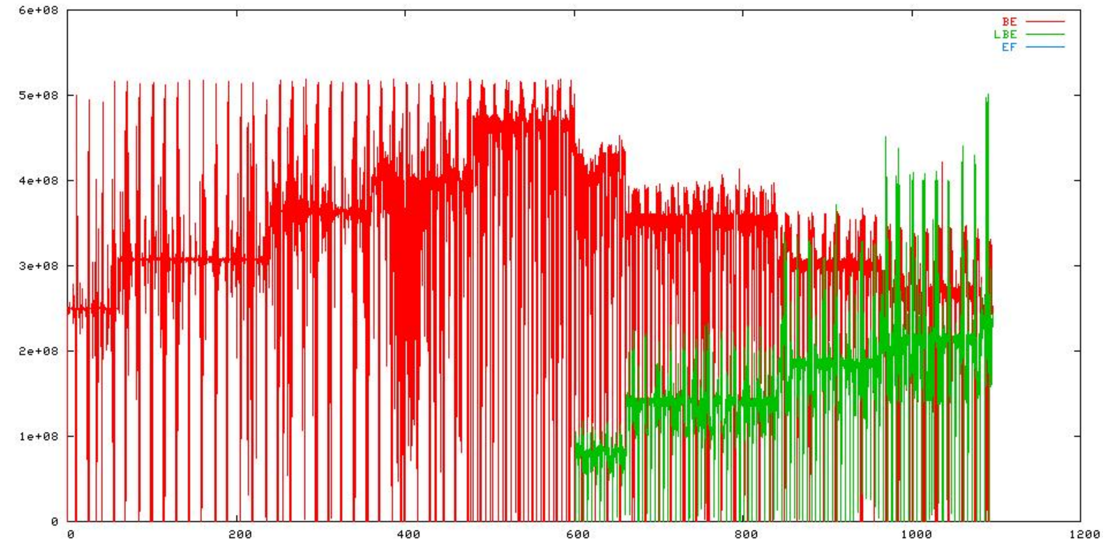


31st October

<u>Objective</u>
To shown that LBE does not get starved out completely when BE levels increase. (same as previous test on 24th)

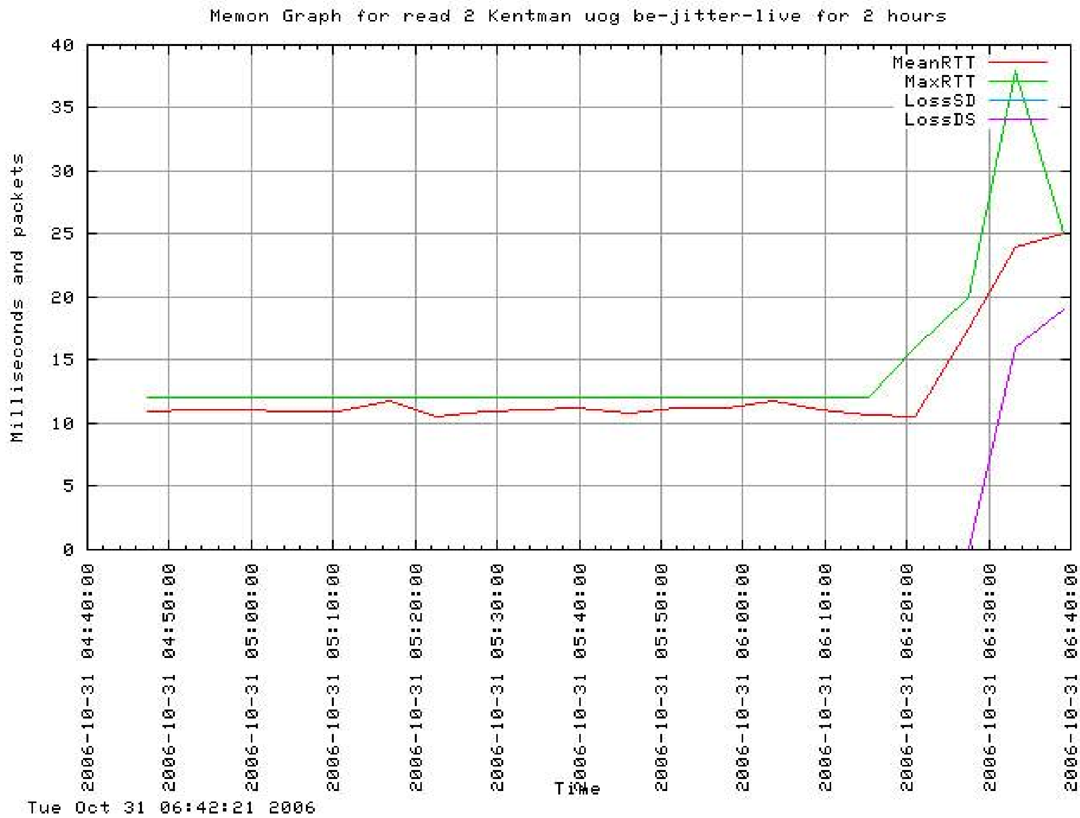
The test ran for a 20 minute period and started with 250Mbps of BE traffic, increasing to 500Mbps over a 10 minute period, this amount was then sustained for a further 10 minutes. (This would show that BE was able to use all the available bandwidth). On the 10th minute 100Mbps of LBE was introduced with this figure incrementing to 500Mbps over the reminder of the test (10 minutes).

Graph generated from CRUDE output



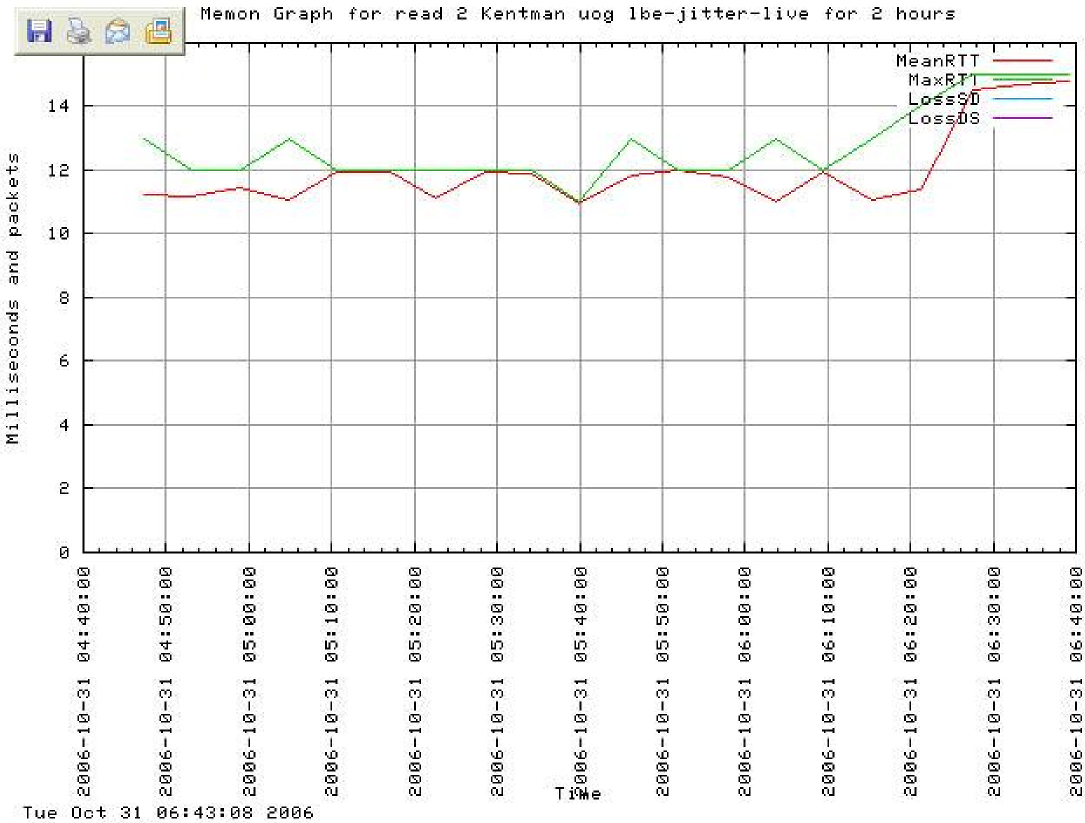
JANET Measurement

The dataset is: QoS Project
The duration is: 2 hours
Tablename is read_2_Kentman_uog_be
Graph type is Normal



JANET Measurement

The dataset is: QoS Project
The duration is: 2 hours
Tablename is read_2_Kentman_uog_lbe
Graph type is Normal



Results:

The QoS monitoring graphs indicate a loss of traffic but yet the CRUDE graphs suggest differently. No explanation can be given for this loss but the graphs generated from the CRUDE data do give hope that QoS does in fact work. One can see that BE is able to send 500Mbps and when LBE is introduced it is still guaranteed its minimum bandwidth but it does have a minor affect on the BE traffic as it begins to drop off slowly but this may be due to the problem with the server at Manchester being unable to cope with the traffic load.

Overall Conclusion:

Several factors affected our ability to perform the LBE tests and to achieve the expected results, notably the difficulty in generating streams of traffic to achieve a true saturation of the link; the inadvertent configuration of a policy map on the Kent MAN link out to JANET that, until noticed, rewrote DCSP values; as well as firewall and disk capacity issues on the CRUDE server.

Ultimately, the tests did not generate the expected results. We investigated the discrepancies and concluded that they were not due to problems within Kent MAN, but with QoS configuration on the Janet link under test, since the configuration used for these tests was identical with that of the one used by Manchester during the phase one trials, where the configuration proved to work and because tests were carried out to ensure the correct DSCP values were being preserved across the Kent MAN and JANET networks. During one of the tests we did see that one of the graphs showed that LBE traffic was affected when BE was introduced, as was the intention, but this was not consistent throughout the trial.

Due to the timing of the upgrade to SJ5, no further tests could be performed. However, if further, similar work was to be performed in the future, our experiences here have shown a need to first look into improving the measuring, monitoring and test traffic generation infrastructures available for use within the Janet and Regional Networks.