



Transferring large files to remote sites: an update

C.M. Maynard, S.P. Booth

*EPCC, The University of Edinburgh, James Clerk Maxwell Building,
Mayfield Road, Edinburgh, EH9 3JZ, UK*

May 2, 2007

Abstract

Transferring large files to remote sites is revisited in the light of two new developments. The UK academic network has now been upgraded to SuperJANET5. The successor service to HPCx, the HECToR machine, will be located at the ACF, University of Edinburgh. Transferring data to HECToR from HPCx will be an important issue for the new service.

This is a Technical Report from the HPCx Consortium

© HPCx UoE Ltd 2007

Neither HPCx UoE Ltd nor its members separately accept any responsibility for loss or damage from the use of information contained in any of their reports or in any communication about their tests or investigations.

A year ago a HPCx Technical Report [1] explored transferring large files between various remote machines. The aim of this report was to compare two different methods, scp and gridftp. The latter was clearly better, for two fairly obvious reasons, *viz.* the different encryption practice of the two methods, and the ability of gridftp to invoke many parallel data streams.

The use of parallel data streams is significant because of the way the TCP/IP protocol is implemented. As the internet does not guarantee to deliver all packets TCP/IP includes packet acknowledgement to inform the sending computer when a packet needs to be re-transmitted. This often results in the achievable network bandwidth for a single socket connection being less than the available hardware bandwidth because the sending computer keeps stopping sending data to wait for acknowledge packets. If multiple data streams are used at the same time it is very unlikely that all data streams will be stalled at the same time. It is possible for system administrators to reduce the impact of this problem by tuning the TCP/IP network parameters for their hosts. However even when these parameters have been tuned the use of multiple data-streams can still be beneficial.

Since the original report the UK academic network service has been upgraded to superJANET5 [2] (SJ5). Whilst the ultimate bandwidth for end-to-end transfers is not likely to saturate the backbone 10 Gbit/s nor “collector arc” 2.5 Gbit/s due to many local throttle points, the topology and speed of SJ5 is relevant for connecting services such as HPCx and HECToR [3] and considering how to plan local network configurations to eliminate or reduce likely bandwidth throttle points. This report examines how the transfers of large files is likely to fair between HPCx and HECToR.

Shown in figure 1 is a network map of superJANET5. HPCx connects to SJ5 via the NNW Regional Network, and HECToR will connect via the EASTMAN Regional Network.

As the HECToR machine is not yet in service, it is therefore not possible to produce a reliable indication of the achievable file transfer rate. However, it is possible to look at the network performance between HPCx and a machine at a similar network location. In order to factor out the impact of the file-system performance at the end-points (Any machine currently available will have a significantly different IO capability to the HECToR service machine) we use the `iperf`[4] tool to measure only the network bandwidth. Like Grid-FTP this tool supports performance measurement using multiple data streams.

The machine chosen to represent HECToR in these tests is `edqcdgrid` on the University of Edinburgh 1 SRIF network. This machine is essentially located on the correct network. Though it is physically located a couple of miles away from the eventual HECToR location the two sites are connected by a 8Gb/s network so this difference should not be significant in performance terms. All tests were run for 30 seconds and the total amount of data transferred and the aggregate bandwidth achieved recorded.

From this data it is clear that although it is difficult to achieve good performance with a single TCP/IP stream, it is possible to obtain a good fraction of the expected 1000Mb/sec peak network performance if multiple streams are used. The number of streams required is remarkably high but hopefully by tuning the networking parameters on the HECToR service when it becomes available this can be reduced.

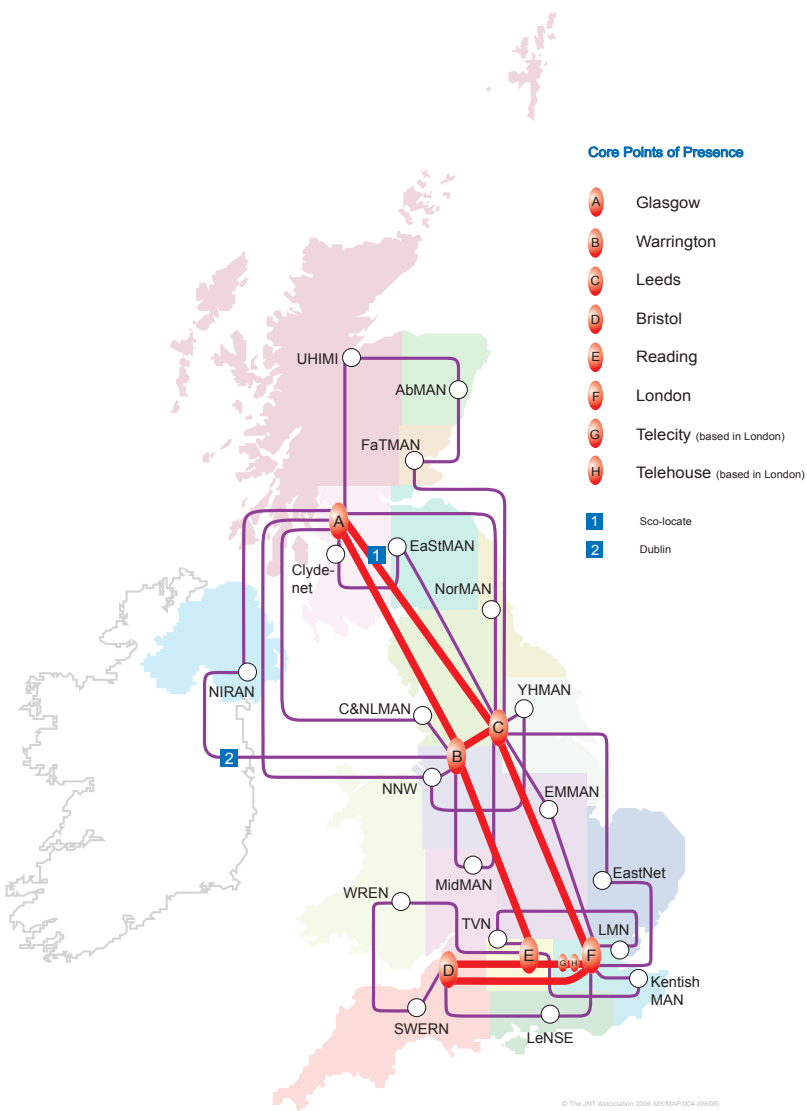


Figure 1: Network map of superJANET5, taken from [2]

Edinburgh to HPCx		
N	Data	Bandwidth
1	198 MBytes	55.3 Mbit/s
2	383 MBytes	107 Mbit/s
3	558 MBytes	156 Mbit/s
4	729 MBytes	204 Mbit/s
5	926 MBytes	259 Mbit/s
6	1.03 GBytes	295 Mbit/s
7	1.06 GBytes	303 Mbit/s
8	1.35 GBytes	385 Mbit/s
9	1.52 GBytes	438 Mbit/s
10	1.75 GBytes	501 Mbit/s
11	1.84 GBytes	528 Mbit/s
12	2.1 GBytes	602 Mbit/s
13	2.37 GBytes	677 Mbit/s
14	2.46 GBytes	704 Mbit/s
15	2.55 GBytes	731 Mbit/s
16	2.57 GBytes	729 Mbit/s
17	2.53 GBytes	724 Mbit/s
18	2.58 GBytes	737 Mbit/s
19	2.59 GBytes	741 Mbit/s
20	2.66 GBytes	760 Mbit/s
HPCx to Edinburgh		
N	Data	Bandwidth
1	132 MBytes	36.9 Mbit/s
2	259 MBytes	72.4 Mbit/s
3	387 MBytes	108 Mbit/s
4	523 MBytes	146 Mbit/s
5	654 MBytes	183 Mbit/s
6	781 MBytes	218 Mbit/s
7	913 MBytes	255 Mbit/s
8	548 MBytes	152 Mbit/s
9	1.13 GBytes	324 Mbit/s
10	1.26 GBytes	362 Mbit/s
11	1.39 GBytes	397 Mbit/s
12	1.51 GBytes	432 Mbit/s
13	1.63 GBytes	467 Mbit/s
14	1.75 GBytes	501 Mbit/s
15	1.86 GBytes	533 Mbit/s
16	1.96 GBytes	561 Mbit/s
17	2.08 GBytes	595 Mbit/s
18	2.17 GBytes	622 Mbit/s
19	2.29 GBytes	655 Mbit/s
20	2.17 GBytes	621 Mbit/s

Table 1: Transfer performance for 30 second transfers N denotes the number of parallel streams.

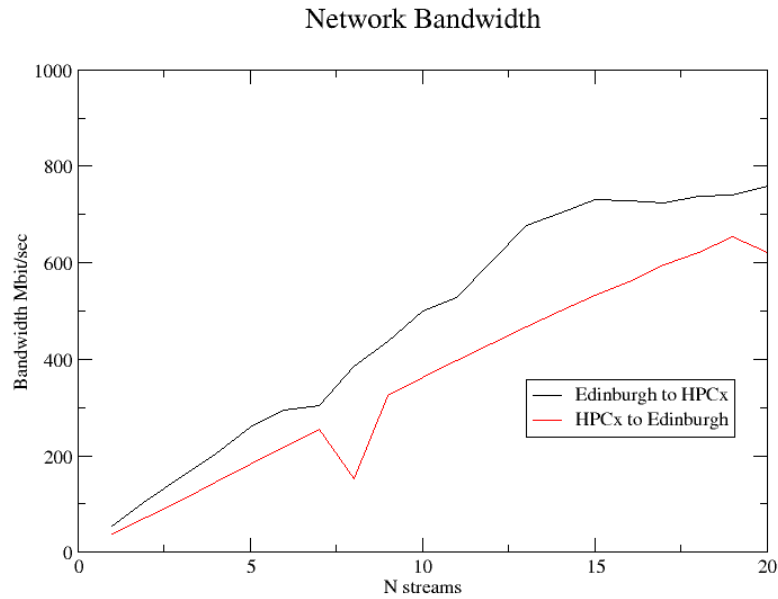


Figure 2: Measures network bandwidth between HPCx and Edinburgh as a function of number of streams.

References

- [1] C. Johnson and S. Booth HPCx TR603;
http://www.hpcx.ac.uk/research/hpc/technical_reports/HPCxTR0603.pdf
- [2] <http://www.ja.net/sj5/index.html>
- [3] <http://www.hector.ac.uk/>
- [4] <http://dast.nlanr.net/Projects/Iperf/>