Back ground information on CDF progress using UKLight

Valeria Bartsch, Barney Garrett, Mark Lancaster, Nicola Pezzi

1 status

- gridftp tests from cdf00-uklight at the UCL site to dmzmon0.deemz.net at the Fermilab site which are both directly connected to the UK light link show a max. transfer rate of about 250 Mbit/sec, although the disks on both sites are capable of higher read/write rates (550Mbit/sec at dzeromon measured with dd and 700MBit/sec at cdf00-uklight measured with dd and Richard Hughes-Jones tool):

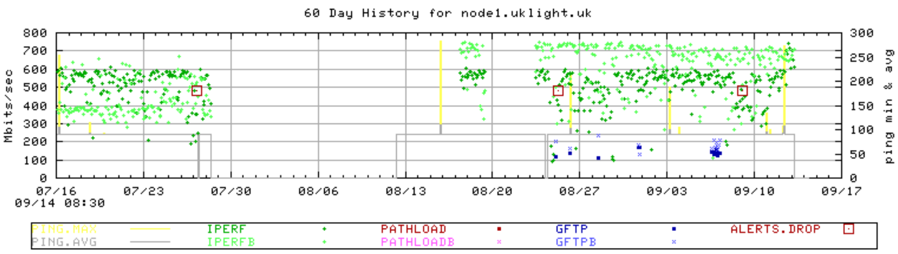

Figure 1: The data transfer between dmzmon0.deemz.net and cdf00-uklight.hep.ucl.ac.uk with several data transfer methods (iperf and gridftp) is shown. Figure from: http://dmzmon0.deemz.net/ wanbanmon/html/wan_bw_tests2_CDF.html which shows the transfers between dmzmon0 and cdf00-uklight

- gridftp test via storage resource managers from the CDF Dcache system to UCL shows data transfer rates of max. 250 Mbit/sec. The parameters of these transfers have been tuned already. Please note that we are here connecting to the DCache system, that means not a system directly connected to the Starlight link.

- iperf tests which are memory to memory transfers only show data transfer rates which are consistent with the nominal data transfer rates of the UK light link of 600 MBit/sec. This shows that there is no obvious bottleneck in terms of switches/network as far as dmzmon0.deemz.net is concerned. However it is still lower than the expected 1 Gbit/sec.

- Compared with other connection the iperf rates compare well with other connections. E.g. Italy has 250Mbit/sec, Canada 400Mbit/sec and Taiwan 300Mbit/sec. Gridftp cannot support transfer rates in excess of

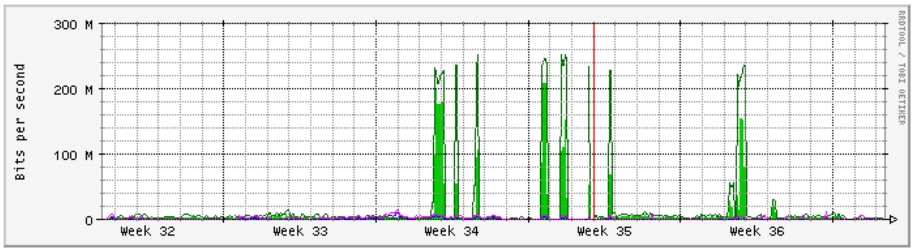

Figure 2: The graph shows the traffic on the UK light link, the last tests were done with the storage resource managers interfacing directly to the DCache data pools.Figure from: http://m-s-fcc-mrtg.fnal.gov/ netadmin/mrtg/mrtg-rrd.cgi/r-s-starlight-fnal/r-s-starlight-fnal_gi1_11.html

250 Mbit/sec; Canada the has 200Mbit/sec and Italy 60Mbit/sec; so it is still the highest but far from ideal. (Information obtained from http://dmzmon0.deemz.net/ wanbanmon/html/wan_bw_tests2_CDF.html which shows the transfers between dmzmon0 and cdf00-uklight).

2 summary of the status

Obvious bottlenecks like disk IO have been circumvented by tuning the disks at cdf00-uklight and installing storage resource managers on the Fermilab site.

3 possiblitiy of further bottlenecks

With the help of SRM (storage resource managers) data is transferred directly between the dCache pools and cdf00-uklight as shown in Fig.3. The files of one dataset are distributed over several dCache pools, this is automatically done when the files are staged from tape. To connect directly to the dCache pools is an advantage compared to transferring directly with gridftp, otherwise the data would need to go through a dCache door and this is very likely a bottle neck for high data transfer rates. The SAM station can spawn 10 eworkers, in other words 10 parallel data transfers. This means that the data could be transferred with 10 times an average single disk IO, because it is connecting to ten different disks. It has been checked with netstat that it one usually connects to different disks. However there are on average about 10 parallel transfers at each disk running at one time from different users. (The information can be found at: http://cdfdca.fnal.gov:2288/queueInfo .) Therefore the increase of data transfer rate is levelled out by the other users using the pools. The advantage gained by using the SRM model is lost again. Increasing the number of parallel transfers would rise other issues.

4 possible solution

First of all it would be nice to see that a higher data transfer rate is achievable even when transferring data from dCache. This has been shown only from dmzmon0.deemz.net which is located directly behind the link so far.

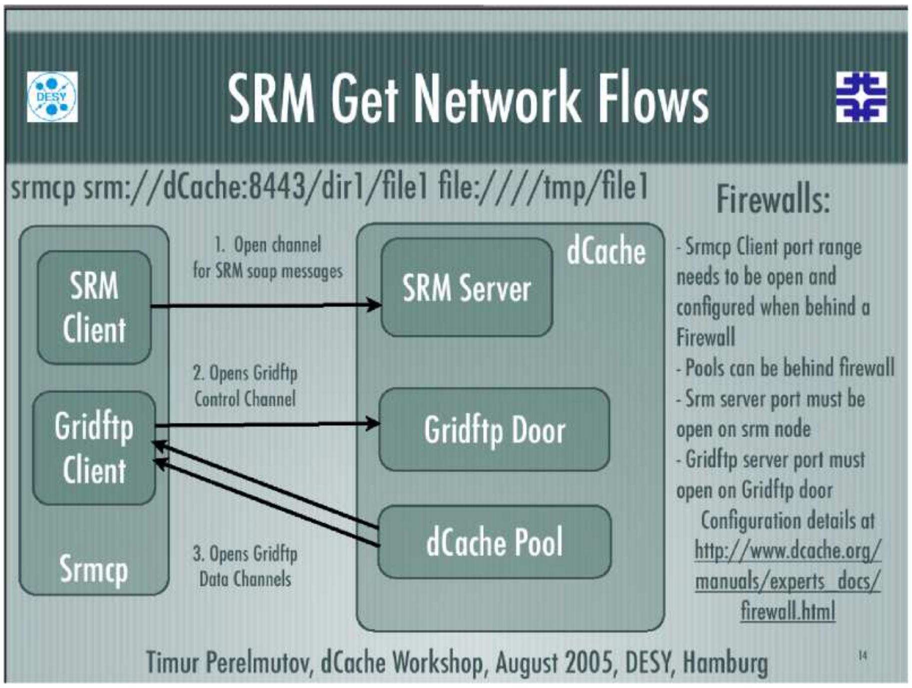

Figure 3: Storage Resource Managers copy directly from the DCache pool, from dCache Workshop, Aug. 2005 DESY, Hamburg by Timur Perelmutov.

If these tests prove that a higher data transfer rate is at all possible, the only way to increase the total number of parallel transfers is to use more than one machine at UCL for the transfers. Up to now we do not have another machine close to the UK light link, which is not behind firewalls therefore it is not possible to check this idea. A different setup is therefore needed.