Table of Contents

Service Challenge Technical FAQ

This is a collection of frequently asked questions, and answers

Network Tuning

- MTU sizes (Mark van de Sanden, sanden@saraNOSPAMPLEASE.nl) During one of our first service challenges we had a mismatch in MTU size on the host sides. The oplapro nodes are default configured with a MTU of 1500. Our local nodes where configured with a MTU size of 9000 (jumbo frames). Network wise the configuration was correct because the first router our local nodes where connected to was also configured with jumbo frames. What we discovered that the single stream performance was very low but with multiple streams or (for example) iperf's we could fill the 1 GE bandwidth. We were very puzzled why the single stream performance was very low. One explanation is that the packages returned from our local nodes where jumbo frames which are fragmented on the router and send to the oplapro nodes. If fragmentation is done in software, as far as we know it was on this router, this is time consuming. With multiple streams you can fill the bandwidth and time needed for the fragmentation but with a single stream you can not.

  ```
  '''keep MTU sizes on the hosts the same to 1500'''
  ```

Host Tuning

- Network hangs due to memory starvation (Mark van de Sanden, sanden@saraNOSPAMPLEASE.nl) During one of our first service challenges we suffered from network hangs when transfering large amounts of data. We saw this problem not during our "iperf" tests but during our "globus-url-copy-tests". With "top" we could see that the memory free was dropping to about 10MB from the about 3GB which is available on the system. I assume that this memory used is for file buffer cache. After a while (about 10 - 20 minutes) the network on the node is hang state. From the dmesg command we could see that the kernel tries to swap but is not able. After sometime of a reboot avery works again. The solution in our case was to limit the file buffer cache and keep memory free for tcp buffers and user space memory. Our rule of thumb was to leave at least 10% of physical memory free. In the linux kernel 2.6 you have the "vm.min_free_kbytes" which forces the kernel to keep a minimum number of kilobytes free. How this is forced on a 2.4 kernel I do not know.

  ```
  Our current setting is:
  '''# sysctl vm.min_free_kbytes'''
  '''vm.min_free_kbytes = 409600'''
  ```

Cleaning up dcache disk-only files

Many sites will provide some TB of non-migratable disk space for the throughput phase. The FTS will place many files onto the disk (rather than the overwriting done in SC2). This is not cleaned by the FTS, and a local site cleanup is required. During SC2 we tried this model with RAL, and what they did was on the head node:

```
find /pnfs/rl.ac.uk/data/dteam/storage -amin +180 -type f | xargs -l10 rm -f
```

-- JamesCasey - 16 Jun 2005

or Ideas, requests, problems regarding TWiki? use Discourse or Send feedback