

QMUL – Experiences with the Cern Virtual Machine File System (CVMFS)

Christopher J. Walker

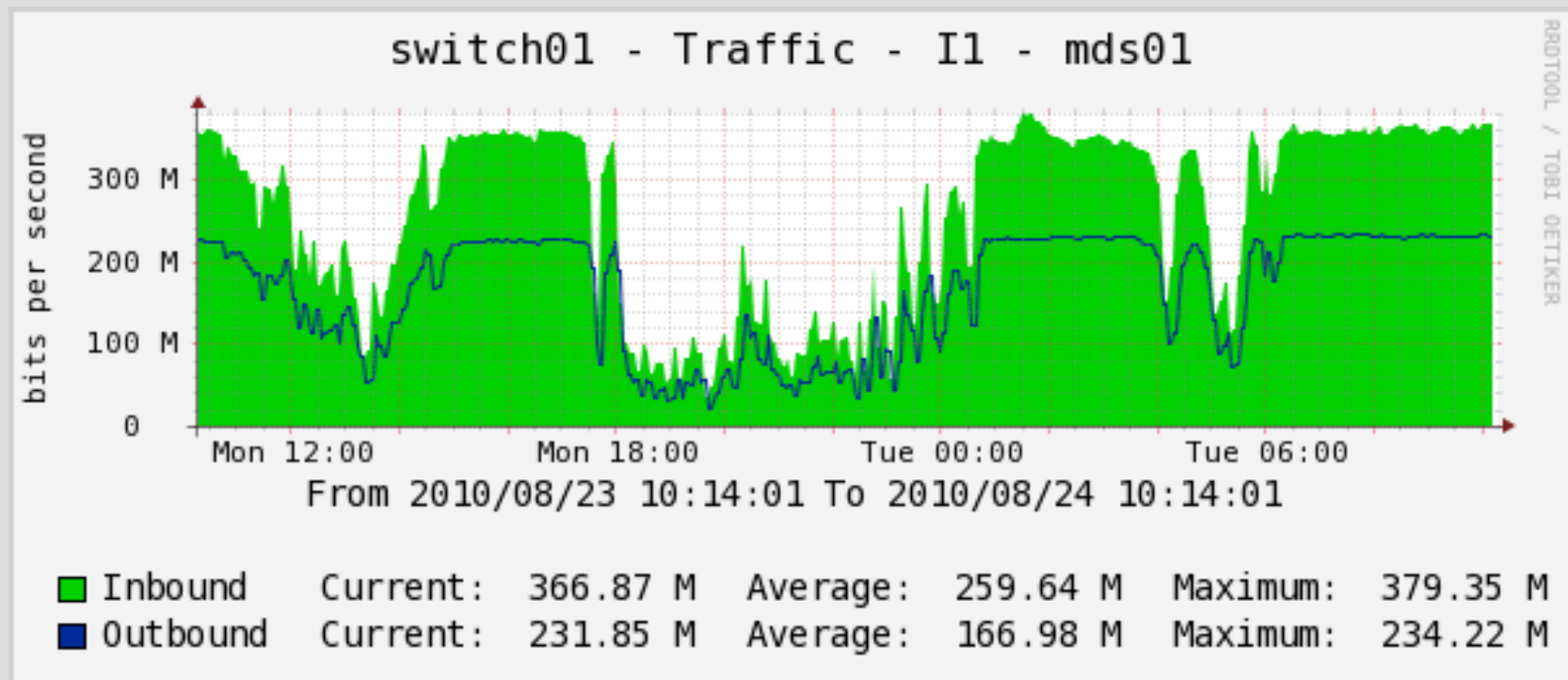
Queen Mary, University of London

Overview

- Motivation
 - NFS locking issues
 - Lustre - MDS load
 - Software install hassles
- What is CVMFS
 - Not just for virtual machines
- Installation
- Performance
- Conclusions

Motivation

- NFS: locking issues
- 4 900 000 files in atlas software area
- Lustre: High MDS Load (>100)
- Which releases to provide



What is CVMFS

- Developed for the CERN VM
 - Standalone software – VM not needed
- FUSE module
 - Web directory tree looks like a filesystem
 - Local cache
 - Deduplication
 - SQUID Proxy

How do I install it?

Installation procedure

- *yum install fuse cvmfs cvmfs-init-scripts*
- Configuration:

```
cat /etc/cvmfs/local.d/default.conf
CVMFS_REPOSITORIES=atlas
CVMFS_CACHE_DIR=/scratch/lcg/cvmfs2
CVMFS_HTTP_PROXY="frontiercache.esc.q
mul.ac.uk:3128"
```
- Run
 - `/sbin/chkconfig cvmfs on`
 - `/sbin/service cvmfs start`
- <https://twiki.cern.ch/twiki/bin/view/Atlas/Tier3CVMFS2SLC5>

\$VO_ATLAS_SW_DIR

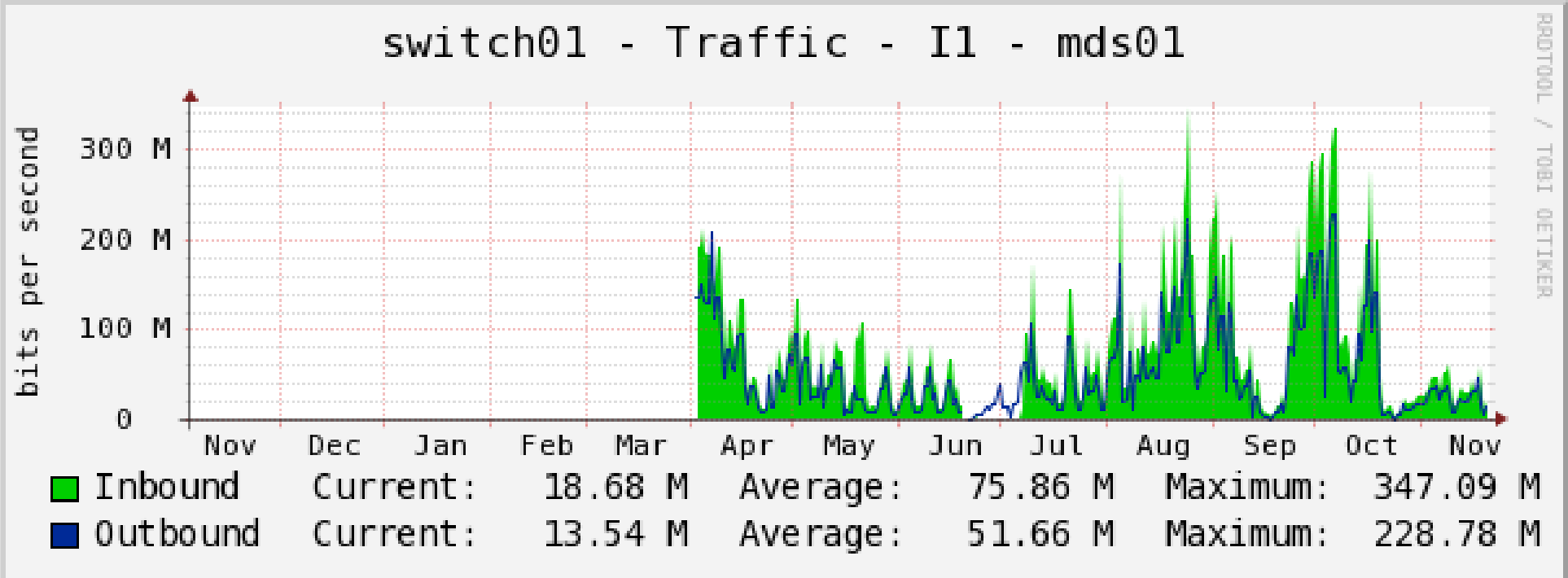
```
$ ls -l /mnt/lustre_0/software/lcg_experimental_sw/sl5/atlas
total 44
drwxr-xr-x  3 atlassgm atlas 4096 Jan 25  2010 atlas-gcc
-rw-r--r--  1 atlassgm atlas  178 Feb  1  2010 AtlasSiteConfig.sh
-rw-r--r--  1 atlassgm atlas   94 Feb  1  2010 AtlasSiteConfig.sh.orig
drwxr-xr-x  4 atlassgm atlas 4096 Jul 15 12:56 cctools
lrwxrwxrwx  1 root    root   20 Oct 20 10:11 database ->
/opt/atlas/database/
drwxr-xr-x  6 atlassgm atlas 4096 Apr 23  2010 ddm
drwxr-xr-x  5 atlassgm atlas 4096 Nov 21 00:14 local
drwxr-xr-x  3 atlassgm atlas 4096 Jan 21  2010 prod
lrwxrwxrwx  1 root    root   39 Oct 20 10:40 software ->
/opt/atlas/software/i686-slc5-gcc43-opt
drwxr-xr-x 27 atlassgm atlas 4096 Oct 19 19:19 software.old
-rw-r--r--  1 atlassgm atlas 5314 Nov 21 01:47 tags
```

Job Changes

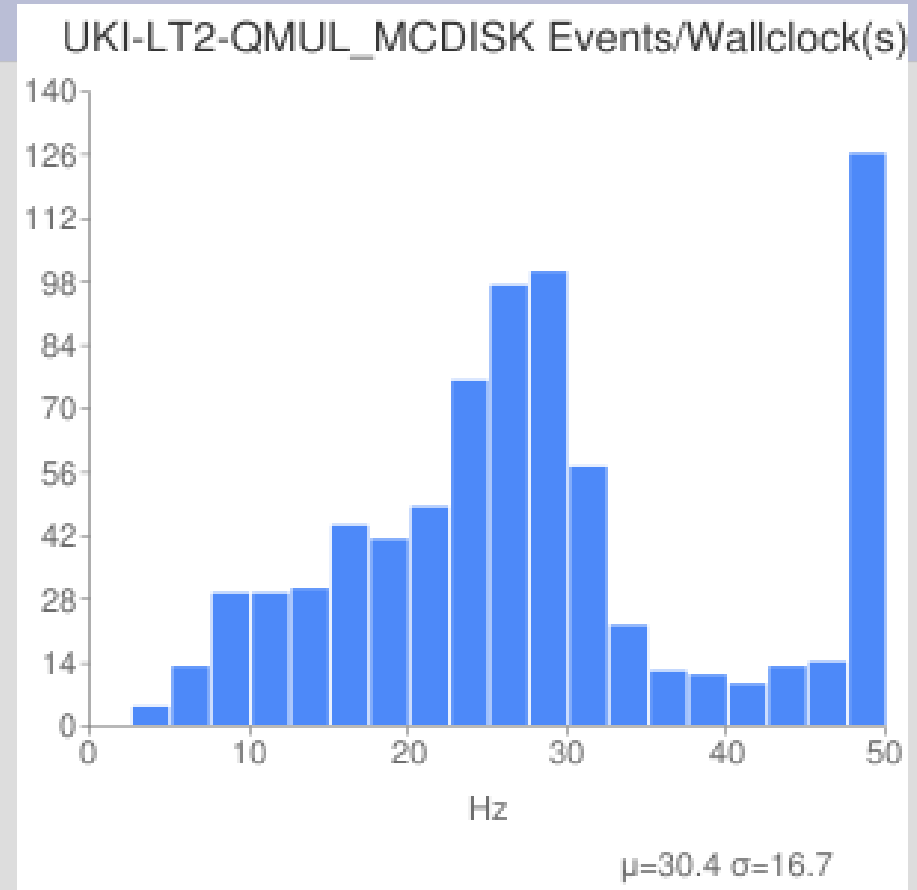
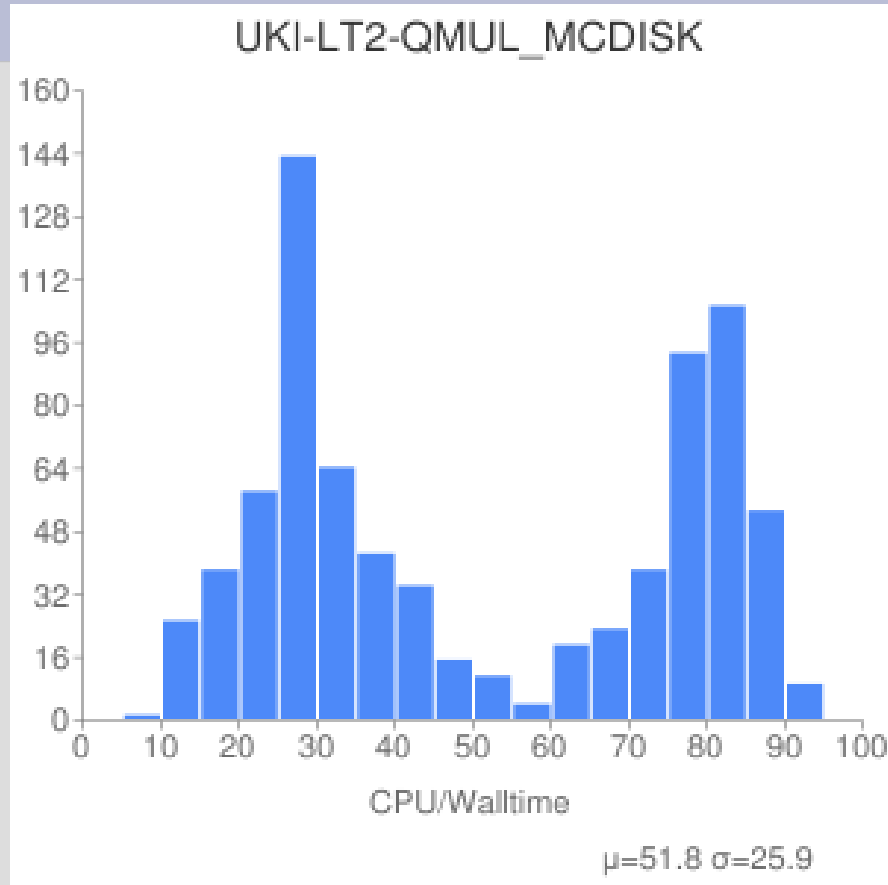
- Old:
 - source \$VO_ATLAS_SW_DIR/software/
\$RELEASE/setup.sh
- New (<16.0.0):
 - source \$VO_ATLAS_SW_DIR/software/
\$RELEASE/cmtsite/setup.sh -tag=AtlasOffline,
\$RELEASE
- New (>16.0.0)
 - source \$VO_ATLAS_SW_DIR/software/
\$RELEASE/cmtsite/asetup.sh AtlasOffline
\$RELEASE

Results

- Installed 21 Oct 2010
- Metadata traffic reduced
- Job failures reduced

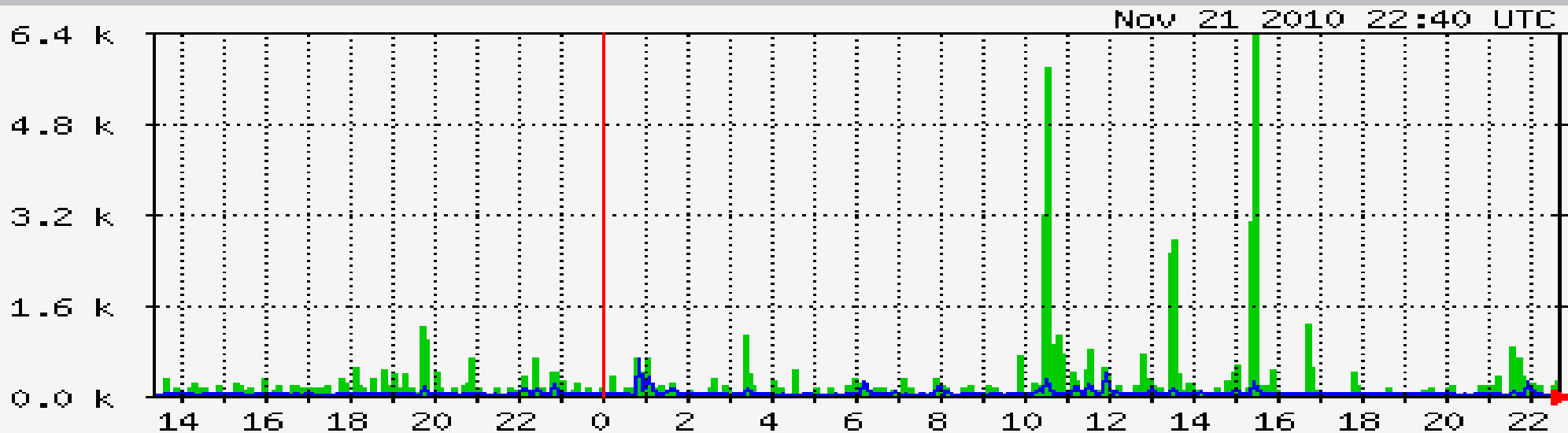


Results: Hammercloud: 10001578

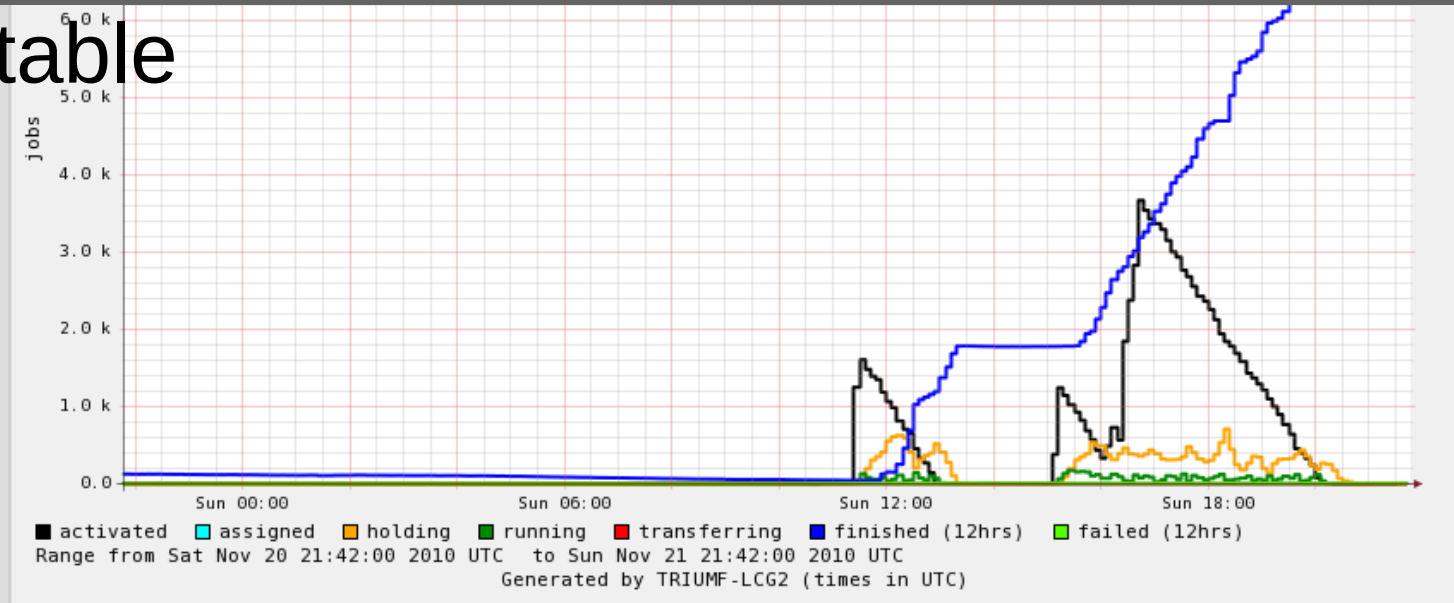


- 71 000 000 events (8h)
- Rack uplinks saturated – need bonded 10GigE.

Squid: load

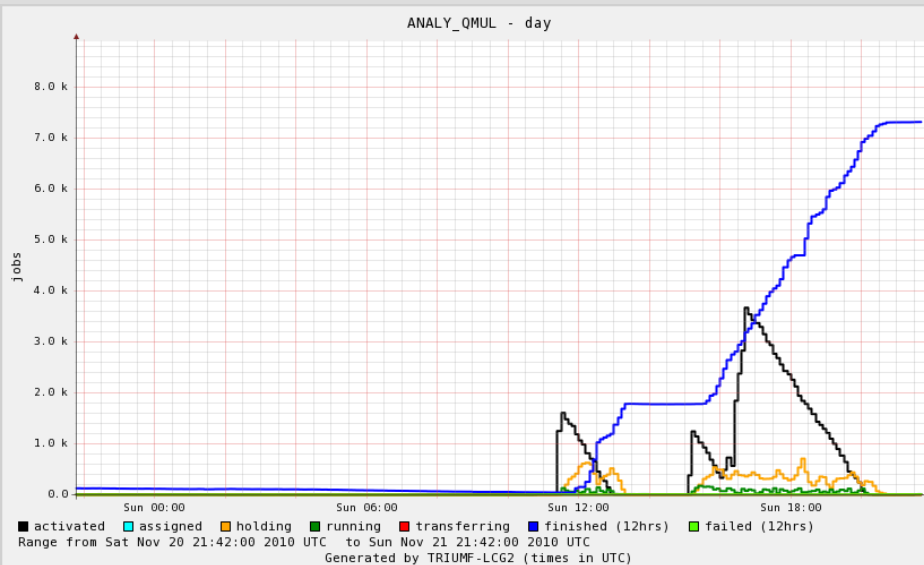
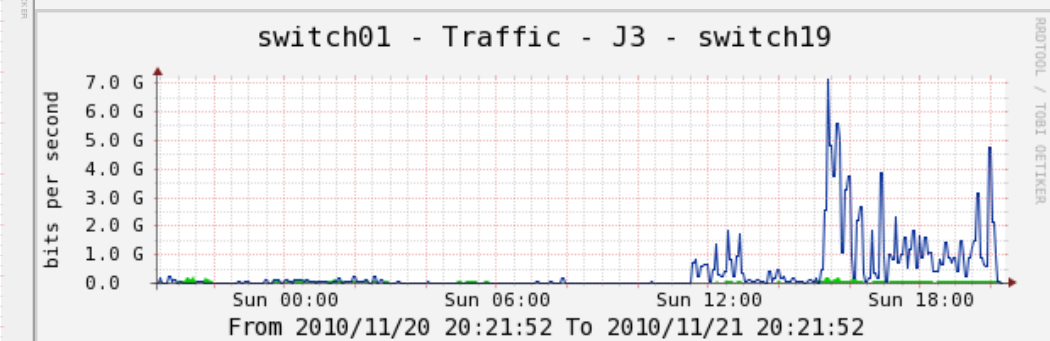
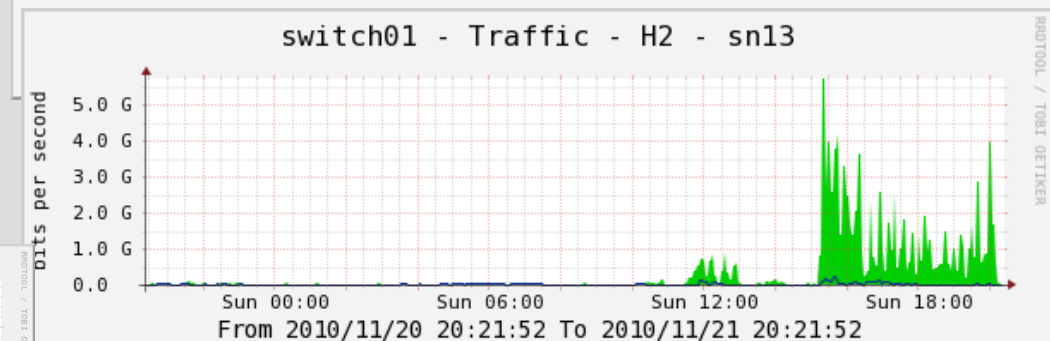
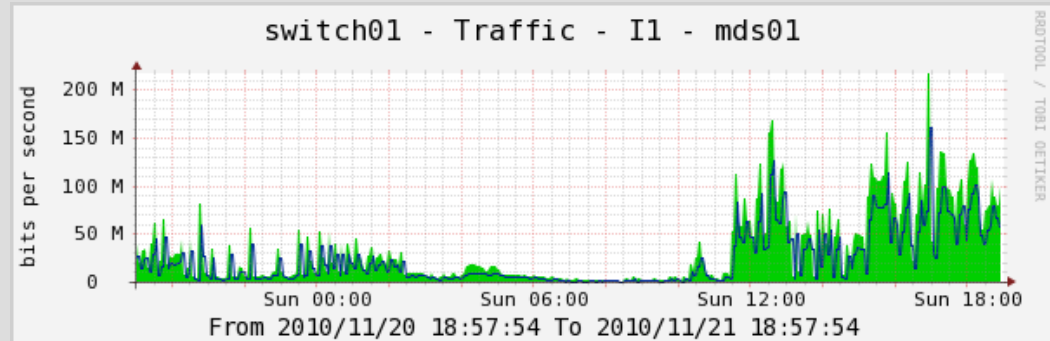


- Load acceptable
- Proxy:
 - 5 years old
 - 2 Gig RAM
 - 2 cores



Network Performance

- MDS load acceptable
- 10 GigE Storage node peaks at 5 Gig
- 10 Gig top of rack - bottleneck

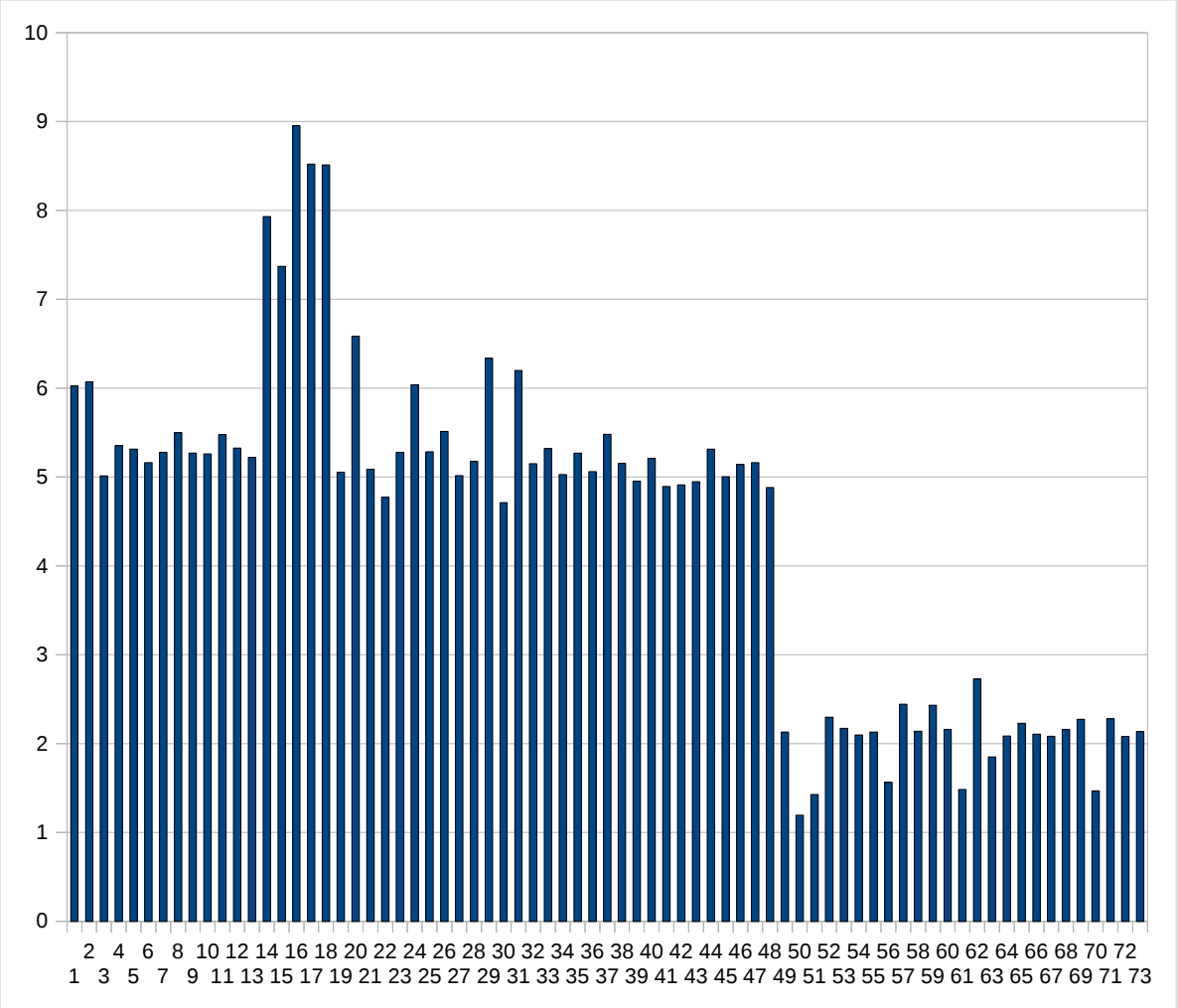


Inbound	Current: 16.16 M	Average: 21.04 M	Maximum: 203.44 M
Outbound	Current: 80.32 M	Average: 400.62 M	Maximum: 7.13 G

Issues

- No logging of download failures
 - “What happens is that the file system returns IO error for the open call. Since this can happen with any file system in any situation, I thought that it doesn't require additional exception handling such as writing to syslog.”
 - “ But perhaps you're right and it makes debugging easier, I'll think about it.”
- Cache size monitoring

Cache Size (GB)



Conclusions

- Easy to install
- “Instant” software availability (Tier-3)
 - Install job needed for Tier-2
- Performant
 - Dramatically improved QMUL's throughput
- Small disk cache (9Gig)
- Requires reliable network connectivity
 - Better logging useful
- Should other sites upgrade?
 - Yes if they have a software area performance problem