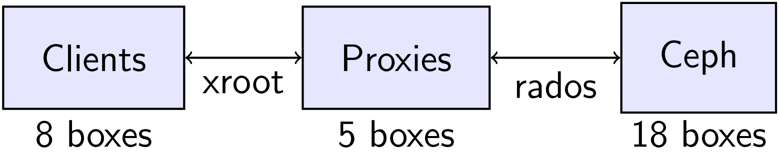
CERN

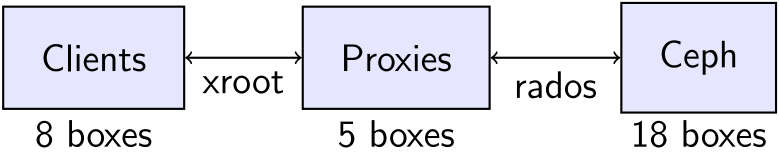
(Stress)Testing Ceph for CASTOR

Sébastien Ponce
sebastien.ponce@cern.ch

The setup



The setup



Some details

- client and proxy machines are batch nodes
- all machines have 10 Gb/s connection
- ceph machines have 540 disks in total
- ceph cluster has 2 PB of effective space

The importance of buffer size

Situation up to xrootd 4.2

- maximum size of xrootd buffers is 2MB
- relation between buffer size and transfer speed
  - single stream, single box
  - with recompiled version of xrootd

Buffer size (MB)	2	32	64
Speed (MB/s)X	65	300	>500

The importance of buffer size

Situation from xrootd 4.3 on

- big buffers have been added
- activate with

```
xrd.buffers maxbsz <bsz>
```

- now max buffer size is 1 GB

The importance of buffer size

Situation from xrootd 4.3 on

- big buffers have been added
- activate with

  ```
  xrd.buffers maxbsz <bsz>
  ```
- now max buffer size is 1 GB
- but async reading from ceph is broken in 4.3...

The importance of parallelization

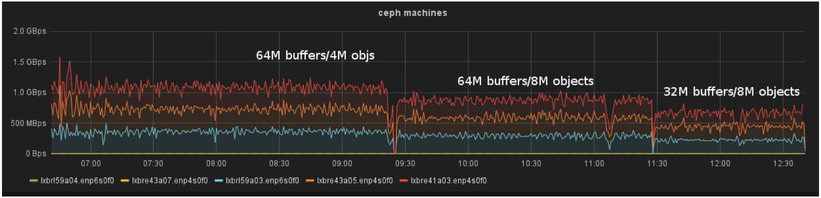
Ceph has latency

- so async transfers are fundamental
- but they need to be sufficiently numerous

The importance of parallelization

Ceph has latency

- so async transfers are fundamental
- but they need to be sufficiently numerous

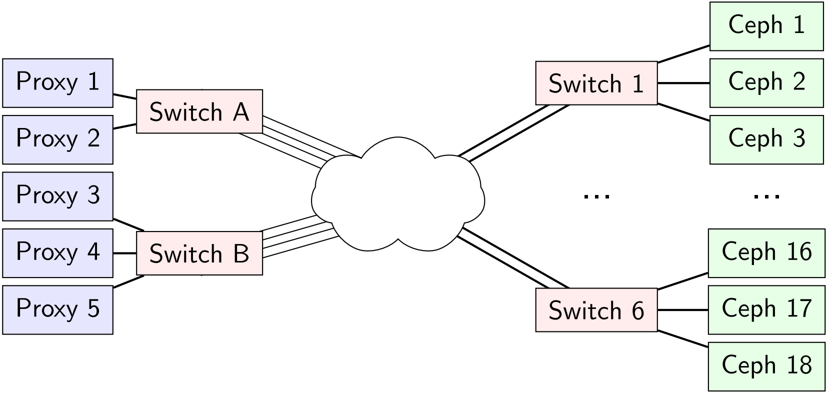


The importance of network layout

Why can I not get more than 2.2GB/s ?

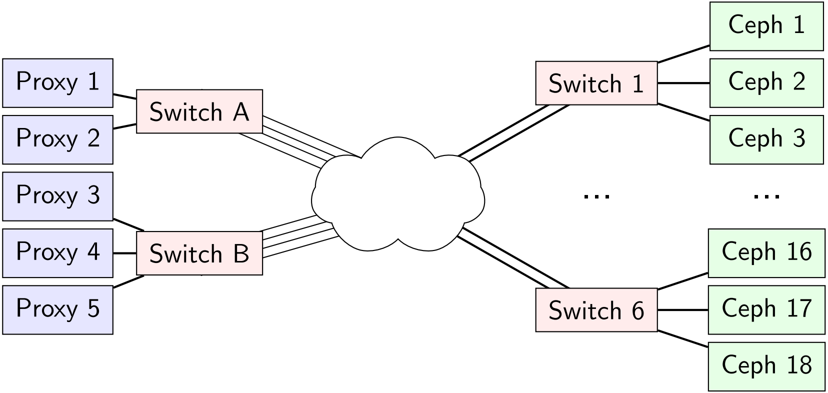
The importance of network layout

Why can I not get more than 2.2GB/s ?



The importance of network layout

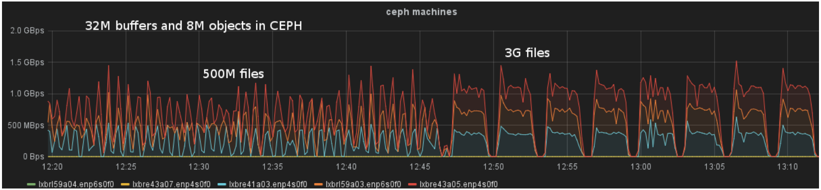
Why can I not get more than 2.2GB/s ?

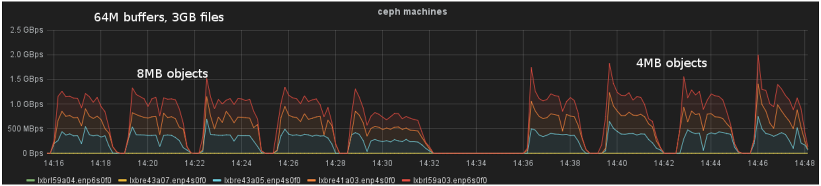


Uplinks of switches 1-6 are saturating ! no clue...

Back to ceph latency

We got weird patterns





Back to ceph latency

Ceph slowness to ack async writes

- Ceph ack can come 20s after end of write
- ack tend to be delayed until end of activity
- file writing synchonize and files all end together
- at that moment, we wait

Back to ceph latency

Some ideas of the origin

- no SSD for the ceph journal
- we are using 8+3 erasure coding
- 1GB/s = 200 objs/s = 2200 sync/s

www.cern.ch