Ubuntu 15.04 - Multipath Problem

Contributed by Ozan Uzun

Currently I am
working on a PLinux/Ubuntu- IBM Flashsystems/820 proof of concept.

I already knew that multipath was broken on version 14.04

I started with Ubuntu v15.04. Bad decision:)

p { margin-bottom: 0.1in; line-height: 120%; }

Power 822L server
has 4 FC Adapters installed, and I need to push the system to the
limit.

FlashSystem supports
max 500K random read iops. I would like to see about ~400K with
multipath,LVM and FS overhead.

I need all of the 4
FC paths to achieve my target.

Actually it is a
long story but:

My first testbed was
Ubuntu 15.04, using the recommended multipath.conf multibus conf.
(official documentation)

- Multipath-deamon
only uses half of the FC (2 of 4)

If I unplug 2
active FC cables paths-it fails

If I unplug 2
in-active FC cables, multipath uses 1 of 2 (what the!!?)

- I can not set
queue_depth value

- Forced multipath
over 4 Paths with LVM stripe and mdadm stripe

Seq. read is ok
(3.2GB/s limit), random iops is limited at ~200K.

No matter what I
tried, I could not get any satisfying results with multipath.

- Created 4
different file systems on 4 LV Volumes, played with LVM conf.

random read iops
max ~260K

- Disabled
multipath, used 4 different LUNS, over 4 different cables (1 LUN
access from each cable-4 total)

Not used LVM.
Magic happens- 450K read random iops.

Last two options are
just for testing, and is not acceptable.

Upgraded to 15.10,
relieved:)

- Created a VG,
consists 4 disks, 4 way striped LVM (64K),one file system.

Random 4K IOPS
~390K, @0.6 ms - Seq. Read 3.2GB/s

- Now I can set
the queue_depth higher (I was hitting the limit), changed the I/O
scheduler to deadline.

A quick note or
two:

ext4 performs
slightly better than the XFS. mount options really helps.

I could not find a
sensible way to monitor FC adapter throughput.

The Linux tools such
as iostat, sar, etc. are block device based and not scsi address
(h:c:t:l) nor storage type nor storage topology aware.

You can check the
queues on Linux;*

```
cat  $(echo
/proc/scsi/sg/device{_hdr,s}
```

```
host   chan    id     lun     type    opens   qdepth  busy   online
3              0         0      12         0           1           32          5               1
*********
****
***
```

Ref*:

http://www.faqs.org/docs/Linux-HOWTO/SCSI-Generic-HOWTO.html

host host
number (indexes 'hosts' table, origin 0)

chan
channel number of device

id SCSI
id of device

lun
Logical Unit number of device

type SCSI
type (e.g. 0->disk, 5->cdrom, 6->scanner)

opens
number of opens (by sd, sr, sr and sg) at this time

depth
maximum queue depth supported by device

busy
number of commands being processed by host for this device

online 1
indicates device is in normal online state, 0->offline