pNFS block layout server user guide

The Linux NFS server now supports the pNFS block layout extension. In this case the NFS server acts as Metadata Server (MDS) for pNFS, which in addition to handling all the metadata access to the NFS export also hands out layouts to the clients to directly access the underlying block devices that are shared with the client.

To use pNFS block layouts with the Linux NFS server the exported file system needs to support the pNFS block layouts (currently just XFS), and the file system must sit on shared storage (typically iSCSI) that is accessible to the clients in addition to the MDS. As of now the file system needs to sit directly on the exported volume, striping or concatenation of volumes on the MDS and clients is not supported yet.

On the server, pNFS block volume support is automatically if the file system support it. On the client make sure the kernel has the CONFIG_PNFS_BLOCK option enabled, the blkmapd daemon from nfs-utils is running, and the file system is mounted using the NFSv4.1 protocol version (mount -o vers=4.1).

If the nfsd server needs to fence a non-responding client it calls /sbin/nfsd-recall-failed with the first argument set to the IP address of the client, and the second argument set to the device node without the /dev prefix for the file system to be fenced. Below is an example file that shows how to translate the device into a serial number from SCSI EVPD 0x80:

```
cat > /sbin/nfsd-recall-failed << EOF
```

```
#!/bin/sh

CLIENT="$1"
DEV="/dev/$2"
EVPD=`sg_inq --page=0x80 ${DEV} | \
        grep "Unit serial number:" | \
        awk -F ': ' '{print $2}'`

echo "fencing client ${CLIENT} serial ${EVPD}" >> /var/log/pnfsd-fence.log
EOF
```