Multi-layout pNFS server support

The current RFC5661 allows pNFS servers to implement support for multiple types of layouts by same server. But the protocol doesn’t say explicitly that same FS can be accessed by different clients at the same time. Although there is nothing to prevent this theoretically, in practice there can be some implementation challenges.

There are already several servers that support 2 types of layout but in 2 different MDS’s and FS. The issue is becoming more difficult when multiple pNFS layout servers MDS’s use same FS as there are possible lock issues for data access. Again in theory this is supported by the protocol but there will be some implementation challenges if multiple clients want to share access to same file but mounting 2 MDS’s with different pNFS layout type support. The problem is getting more complicated if the client mounts same MDS (and same fsid) with different pNFS layout support layout types (the RFC doesn’t deny this). Although the protocol is well defined for this it is the general impression that nobody really thought to which extent the implementations can go as most of the servers only supported one layout type at the time.

And the ultimate challenge is if multiple clients want to access same file on an MDS that support multiple layouts on same fsid to access same file. The locking between the 2 clients access is becoming difficult for implementations. The general opinion is that we need to add some, implementation recommendations, usecases and possible extension to all the 3 RFC5661, 5663 and 5664 in order to ensure there will be no data corruption opportunities. It was recommended that we should have a prototype implementation on Linux server/client to validate shared access to same file by different layout clients.

The major issue will be to define a way to CB the layouts from different types of accessing clients as the file layout server sends a single layout for the entire file while block and object send ranges and can recall ranges while the file client can do I/O’s to any DS and range and create a data corruption opportunity.

In summary the possible actions to be taken to address this issue are:

- Extend the RFC with additional usecases and implementation recommendations
- Extend the layout protocols to support same fsid for different layout clients
- Enhance the 4.1 to a more tight LAYOUTCOMMIT to match 5663 and 5664 (similar discussion at BAT)
- Implement a prototype in Linux to find the protocol gaps.
- Discuss this issue in the WG