

Delegation of Stateids

IETF102

Tom Haynes
loghyr@hammerspace.com

Offline Files - The Use Case

- Some file managers look at the first part of a file to preview it
- Get filehandle and attributes: SEQUENCE, PUTFH, LOOKUP, GETFH, GETATTR
- Open the file: SEQUENCE, PUTFH, OPEN, GETATTR, LAYOUTGET
- Get the first part (to DS): NFSv3 READ

Offline Files - The Cost

- Offline now means off-prem, off-site, the cloud
- OPEN means we have to retrieve the file (or a copy) from the remote object store
- \$\$\$ and performance
- We are just going to throw it away or send it back to the remote object store

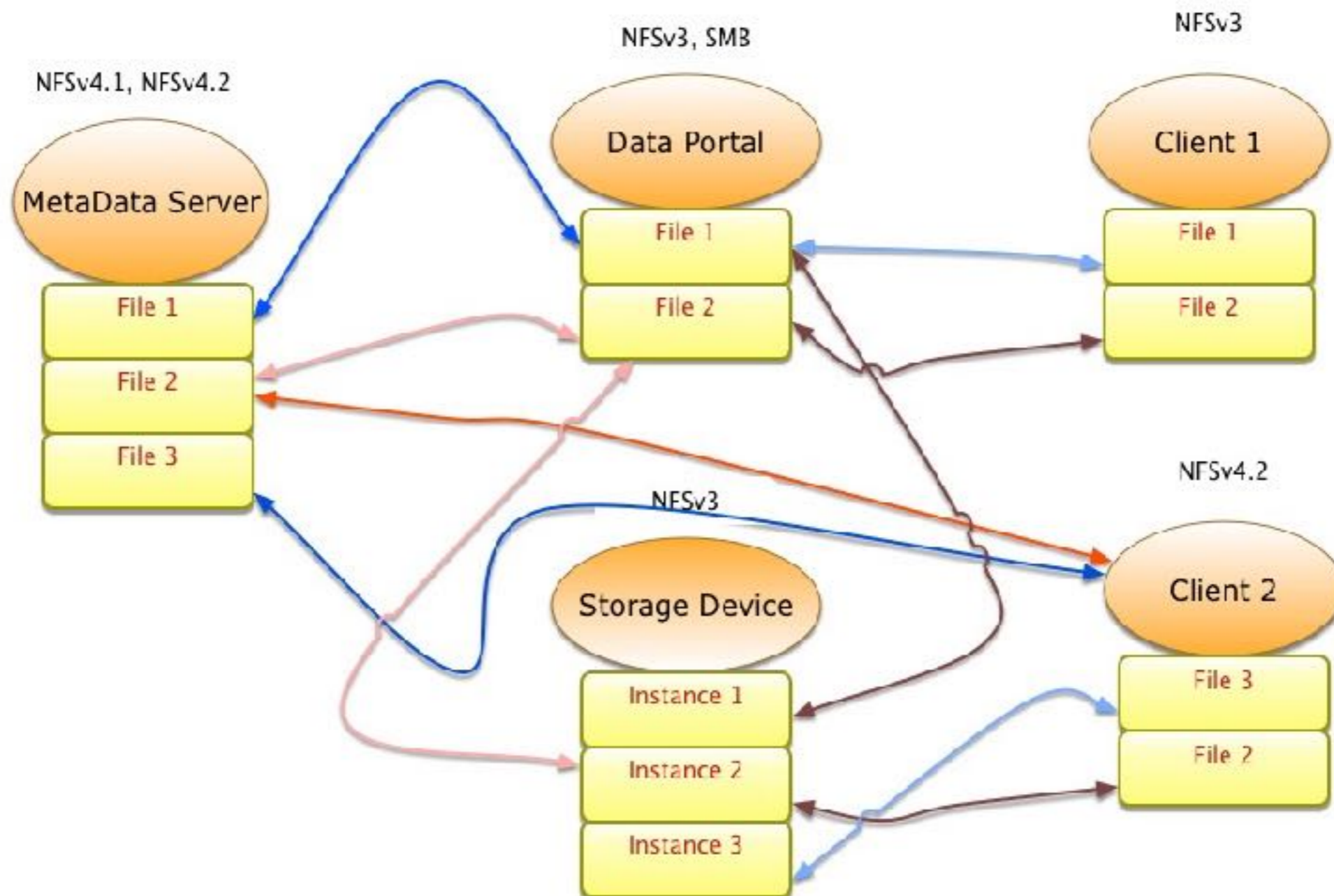
Offline Files - The Solution

- Get filehandle and attributes: SEQUENCE, PUTFH, LOOKUP, GETFH, GETATTR (including OFFLINE)
- Reply has OFFLINE attr set
- Client OS informs file manager the file is offline and it has the choice to force it online or not

Proxying of Times: The Use Case

- A data portal allows converting NFSv3, SMB, etc protocol traffic to NFSv4.2
 - Do not have to provide such capabilities on the metadata server
 - The data portal gets layouts from the MDS and accesses the DSes directly
 - Cluster tax applies

A FlexFile Server



Proxying of Times: The Cost

- NFSv3 and SMB clients are chatty
 - A lot of GETATTRs
 - Protocol translation might be at VFS layer
 - NFSv3 WRITES to DS can not be trusted
 - WCC could have information
 - DS is not authoritative on size, mtime, atime, and ctime
 - For an internal test workload copying 40k files we saw 72k NFSv4 GETATTRs

Proxying of Times - The Solution

- If the Data Portal gets a delegation, it is authoritative for the size
- Proposal is to add new delegation of attributes for proxying times
 - MDS, DS, and portal have to be in time sync
 - Meh, have that requirement for Kerberos anyway
- Data Portal can use WCC and bypass GETATTRs to the MDS
 - Has to allow callback to provide the information
 - Has to set the times when returning the delegation
 - MDS has to cap changes
- For same test workload, saw 3 NFSv4 GETATTRs

Open XOR Delegation: The Use Case

- A file is considered open by a client if it has either an OPEN or a DELEGATION stateid
- Lazy return model for LAYOUTS - the ods count goes to 0, the layout is returned

Open XOR Delegation - The Cost

- Server has to maintain an extra state
- Client has to send a compound to CLOSE the open stateid
 - Typically introduces another GETATTR to the mix
- Small costs, but they add up

Open XOR Delegation - The Solution

- The client asks for either an OPEN or a DELEGATION stateid
- The MDS tries to return only one
- The client has to be prepared to get both
- Note MDS has to determine the correct stateid for the current stateid in the compound: SEQUENCE, PUTFH, OPEN, GETATTR, LAYOUTGET

The draft status

- Version 0 was published
- 3 reviews - nothing majorly negative
- Review responses set to go out
- Want next draft version to be an official WG item