Layouts Impeding File Recovery

IETF 114, NFSv4 WG

Tom Haynes July 26th, 2022

Problem Statement

- When a server reboots, clients have a grace period to reclaim open files
  - Clients know which files they have
  - Server trusts clients to reclaim the state they had
  - Clients open files by FH by using CLAIM_PREVIOUS
    - Typically only valid during reclaim period
- What happens if the client had encountered errors?

Depends on whether loosely vs strongly coupled

- Strongly coupled
  - MDS knows that the DS reported errors to the client
- Loosely coupled
  - MDS depends on the client to report the errors
    - LAYOUTRETURN
    - LAYOUTERROR

Can’t we just get a new layout after reclaim period?

Client side mirroring for the complication!

- Without client side mirroring, easy peasy
- With client side mirroring, no guarantee file is on the same DSes as before

Importance of Reporting the Errors

- Assume client side mirroring
- Assume client was writing
- If client does not reclaim the file, server has to assume there was an error
- Server then has to pick one of date file instances and resilver
  - Very costly
    - Number of open files
    - Size of open files
- If the client reports errors, server knows which instances are not corrupt

Potential Solutions

- Introduce a new operation/attribute for the server and client to agree that the server has the capability
- Use the special stateid with LAYOUTRETURN

Special stateid with LAYOUTRETURN

- If outside grace period, server returns NFS4ERR_BAD_STATEID
- If not special stateid, server returns NFS4ERR_GRACE
- If inside grace period and server does not support new feature, it will return NFS4ERR_BAD_STATEID
  - Client can thus know feature is not supported
- No need for LAYOUTERROR
  - Implies client can still use the layout
- No need to reclaim files

Where to put it?

- Could put it into delstid
  - It is about *Open* files
  - Only real reason to do this is to avoid a new document
- Could do a new document
  - Starts a new cycle
  - 2-3 pages