NFSv4 Migration Implementation Status

Chuck Lever <chuck.lever@oracle.com>
Consulting Member of Technical Staff
Implementation Experience

• Uniform Client String approach
• Linux client UCS
• Recent Solaris team findings
Reliable NFSv4 State Migration

- Fileservers must be able to merge migrated open and lock state with existing open and lock state for the same client. Therefore:
  - Clients must identify themselves identically to all servers
  - Each server-client pair must establish no more than one lease
The Uniform Client String Approach

- Clients must use the same nfs_client_id4/client_owner4 with all servers.

  - RFC 3530bis RECOMMENDS use of distinct client ID strings for each server (section 8.1.1).

  - RFC 5661 RECOMMENDS use of same client ID string for all servers, except when upgrading from NFSv4.0 to NFSv4.1 (sections 2.4 and 2.4.1).
Server Trunking Detection

• To prevent the formation of more than one lease between them:

  • NFSv4.0 servers can now detect UCS clients using multiple IP addresses

  • NFSv4.0 UCS clients can detect multi-homed servers by spotting familiar client_id4 values
Implementation Experience

UCS on Linux

- Replace per-mount boot verifier
- Replace id string containing security flavor and IP addresses
- Deal with NFS4ERR_CLID_INUSE
- Establish lease immediately in order to perform server trunking detection
- Split \{ setclid_cfm; putrootfh; getattr(lease_time) \}
Implementation Experience

Linux NFSv4 Migration Next Steps

- Continue watching for issues with UCS
  - More security improvements coming soon

- Forward-port last year's prototype
  - Main challenge is serialization between recovery and user processes
  - Linux wants NFSv4.1 migration support, too
Implementation Experience
Solaris Server Migration Prototype

• If filesystem state has been frozen for migration, how should a server stall client progress?
  • NFS4ERRRESOURCE can be treated by clients as a permanent error
  • NFS4ERR_DELAY mutates sequence IDs, thus would alter state server is trying to freeze
  • NFS4ERR_GRACE may cause other client side-effects

• Possible solutions
  • Return NFS4ERR_DELAY on FH-bearing op in same compound (for example, PUTFH)
  • Drop the request and connection
Implementation Experience
Solaris Server Migration Prototype

• After a migration, how long must the source server recognize a moved file handle?
  
  • Permanently: NFS4ERR_MOVED always returned, FH never re-used
  
  • For a fixed period: NFS4ERR_MOVED returned for a time, then NFS4ERR_STALE, and FH reuse is permitted
Implementation Experience
Solaris Server Migration Prototype

• A client may verify the identity of a trunked server by testing a state_id it is using on one of the server’s other IP addresses

• NFSv4.0 has no TEST_STATEID operation

• Suggested replacement is a zero-length READ
  • Client might not have an open file to use
  • Client might not have a file with a read state_id
Implementation Experience
Solaris Server Migration Prototype

- Cluster giveback fails for UCS clients

  - ZFSSA cluster implementation relies on separate leases to allow merging state during giveback

  - Linux has a mount option to disable UCS for NFSv4.0

  - Other server implementations may be affected
Questions/Discussion