ORACLE®

# Linux NFSv4 Migration Implementation Update

Chuck Lever
Consulting Member of Technical Staff

## Starting Assumptions

- Use Transparent State Migration whenever possible to minimize risk of losing open and lock state during NFSv4 migration recovery
- To cap resource consumption, server wants single lease per client
- nfs_client_id4 is opaque to server
  - Server can compare two nfs_client_id4 strings for equality only

## Current Practice

- “Non-uniform client string”
  - Client embeds server identifier (IP address) in nfs_client_id4
    - Recommended by 3530bis
- When migrated lease arrives, how does server determine if this lease can be merged with an existing lease?
  - Transparent State Migration becomes problematic
  - Some reboot recovery scenarios result in abandoned leases

ORACLE

## Proposed Practice

- "Uniform client string"
  - Client uses same nfs_client_id4 for all servers
  - Server can immediately recognize when migrated lease matches an existing one, and can merge state into a single lease
- Server trunking detection
  - To keep to one lease per client, client must determine "clientid4 to server" IP address mapping
  - Use SETCLIENTID_CONFIRM
    - With two separate servers, one will remain in the unconfirmed state

ORACLE

## Additional Draft Recommendations

- Detecting absent FSIDs asynchronously and in parallel
- Using a guard operation when retrieving fs_locations data
  - Server uses GETATTR(fs_locations) to clear the LEMO flag for this client

## Implementation Status

- Server trunking detection has been prototyped
  - Added second mechanism for establishing a clientid4
    - In addition to existing mechanism, used during state recovery
  - Invoked only when encountering a server IP address client has not seen before
  - Operation
    - Second walk through nfs_client_list
    - SETCLIENTID_CONFIRM done if clientid4 matches
  - Have not implemented this for NFSv4.1 yet

## Implementation Status
*continued*

- Single nfs_client_id4 string has been prototyped
  - String now contains “Linux NFSv4.x <nodename>”
    - Establishes a separate lease for NFSv4.0 and NFSv4.1 state
    - IP addresses no longer appear in this string
    - To ensure uniqueness, can replace “nodename” with something else (say, a UUID)
  - Same logic now performs NFS4ERR_CLID_INUSE recovery for all minor versions
    - NFS4ERR_CLID_INUSE means client used inconsistent auth flavor
    - Client retries SETCLIENTID with all flavors it knows

## Next Steps

- Port non-UCS migration implementation to UCS prototype client
  - See previous talks for details
- Flesh out NFSv4.1 implementation
- Implement asynchronous FSID presence test
- Implement guard operation when retrieving fs_locations
- TBD: serialization when updating client's data structures
- Testing

ORACLE

ORACLE IS THE INFORMATION COMPANY