
ORACLE®

ORACLE


# NFSv4 Migration Challenges

Chuck Lever
Consulting Member of Technical Staff

# Outline

- Background
- Technical issues
- Impact on existing draft updates and WG charter

ORACLE

# Background

- First, there was RFC 3530 section 8.14
- A few years ago, Solaris NFS team attempted to implement client- and server-side migration
  - Discovered that parts of RFC 3530 were problematic
  - Attempted some creative workarounds
- In mid-2010, Linux team was approached to implement client-side migration
  - Concerns about undocumented “workarounds”

ORACLE

# Background

- Solaris and Linux migration implementations introduced at Connectathon 2011
  - Presented some of the issues
- Informal discussion of how to fix the NFSv4.0 specification began during IETF 81
  - We want our migration to interoperate, therefore WG should be involved
- Created an informational draft to allow 3530bis to be completed while we continue work on migration issues

ORACLE

# Current Practice

- “Non-uniform client string”
  - Client embeds server identifier (IP address) in nfs_client_id4
  - RFC 3530 section 8.1.1 makes this a “should”
  - One client can have more than one lease on a server
- This is harmless...
  - ...*until we want to perform Transparent State Migration*

# Transparent State Migration

- TSM minimizes risk of losing state during migration recovery, thus it really ought to be reliable
  - Use TSM whenever possible
  - Perform state recovery only as a last resort

# Transparent State Migration

- Should servers merge leases after transparent state migration?
  - No: State can get unmanageably complex
    - RFC 3530 assumed migration would be rare, but we expect it to occur frequently in practice
  - Yes: How does a server match a migrated lease with an existing lease it may already have?
    - One client uses unique nfs_client_id4 strings for each server, so server can't know state is eligible to be merged

ORACLE

# Transparent State Migration

- Can a callback update put existing state on the destination server at risk?
- What happens when a migrated client reboots?
  - Old nfs_client_id4 used on destination server
  - nfs_client_id4 changes, server won't recognize it
  - Client's old state is reaped after lease expiry
- How can we make LEASE_MOVED recovery scalable?

ORACLE

# Proposed Practice

- "Uniform client string"
  - Client MUST use 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
- It was difficult to continue working with non-UCS
  - Client would have to help server bind nfs_client_id4 and clientid4
  - UCS is more compatible with NFSv4.1
  - Traditionally have been told UCS is not workable
  - Finally decided change was required for clients to support migration

ORACLE

# Proposed Practice

- Server trunking detection
  - To keep to one lease per client, client must determine “clientid4 to server” IP address mapping
  - Use SETCLIENTID_CONFIRM
    - { clientid4, boot_verf } should be recognized by just one server, but maybe through several IP addresses
  - Is it possible for two unique servers to have the same boot_verf and pass out the same clientid4?

ORACLE

# Additional Recommendations

- Clarify that original intent was single lease per client
- Clarify that callback update cannot cause server to purge state
- Detect absent FSIDs asynchronously and in parallel
- Use a guard operation when retrieving fs_locations data
  - Server uses GETATTR(fs_locations) to clear the LEMO flag for this client

ORACLE

# Current Exploration

- Solaris IP-based failover is a problem
  - Taken-over server combines all resources of both servers
  - Give-back relies on non-UCS clients to sort out what clients are handed back to secondary
  - Is it helpful to think of IP-based take-over as a trunking relationship change?
  - Strictly a backwards-compatibility problem
- Otherwise, we foresee no issues with UCS

ORACLE

# What About NFSv4.1?

- Originally, migration draft was to focus on only NFSv4.0
  - Named “NFSv4.0 migration: Implementation experience and spec issues to resolve”
- Study of NFSv4.1 issues is not complete
  - Has EXCHANGE_ID addressed all open TSM issues?
  - Is NFSv4.1 definition of trunking robust?
  - Should sessions be migrated or not?
  - What does a pNFS migration look like?

ORACLE

ORACLE IS THE INFORMATION COMPANY