Migration-Issues-xx
Where it's been and might be going

David Noveck

IETF99 in Prague

July 20, 2017

Contents

- History (up to -12)
- Recent changes
- Work being done now; Future documents
- Next steps
- The end of migration-issues-xx

History

- Started as WG document in 4/2012
  - Co-authored with C. Lever, P. Shivam, B. Baker
- Initial Focus was on NFSv4.0
  - Proposed work eventually done as a standards-track document
    - Draft-ietf-nfsv4-rfc3530-migration-update (11/2012-2/2016)
    - Published as RFC7931 (in 7/2016)
- After publication of RFC, document needed to shift
  - The v4.0 treatment shifted from proposing to describing
  - Needed to shift to focus more on v4.1

Recent Changes

Overview

- Started to get serious about v4.1 State Migration
  - Had been assuming it would be simple/trivial
  - Oracle experiments showed otherwise
- Had to confront trunking discovery issues
  - Needed different sorts of handling for v4.0 and v4.1
- Also needed to deal with interaction of migration and trunking.
- With -13, document has a new title
  - NFSv4 Migration and Trunking: Implementation and Specification Issues

Recent Changes
New Focus on v4.1 State Migration

- Issues discovered by Oracle experiments:
  - Confirmation Status of transferred client IDs
  - Handling of SEQ4_STATUS flags
    - Effect of LEASE_MOVED not being an error
    - Use of status bits to reflect possible lock state loss during migration
- Started to look more closely at migration in RFC5661
  - Found lots of troublesome stuff
  - Many issues that had been ignored/deferred:
    - Session migration
    - Relationship of pNFS and Migration

Recent Changes (v4.1 State Migration)
Issues discovered by Oracle experiments

- Confirmation status of transferred client IDs
  - Needed to be considered "confirmed"
  - Requires correction of EXCHANGE_ID description.
- Handling of SEQ4_STATUS flags
  - Effect of LEASE_MOVED not being an error (LEASE_MOVED slide)
    - More extensive treatment of client recovery issues needed
  - Use of status bits to reflect possible lock state loss during migration
    - Allows lock loss across migration to be handled like other sort of lock loss

Issues not previously dealt with

- Session migration
  - Provides a framework to allow it
  - Needed to clarify transfer issues
    - Harder than for transferring lock state since ops that return NFS4ERR_{DELAY,MOVED} do affect the session.
- Relationship of pNFS and Migration in order to allow:
  - Migration of an MDS
  - Migration of a file system
  - Migration between file system that differ as to their pNFS support.

Recent Changes
Changing Handling of Trunking

- In v4.0:
  - Initially, no means of trunking detection/discovery made available.
  - In RFCs 3530/7530 trunking is treated as an obnoxious problem to be worked around.
  - In RFC 7931, an OPTIONAL means of trunking detection was provided, moving trunking from bug to feature.
- In v4.1:
  - Trunking detection provided for in RFC5661.
  - But multi-server namespace treated two paths to same fs as if they were two replicas
  - Made trunking discovery using location attributes hard to address
- In both, needed to treat issues of changes in trunking patterns

Work being Done Now
Migration in RFC5661

- Treating two addresses for same replica like two replicas
  - Two replicas which could be used simultaneously and share state
  - Confusion with other cases of simultaneous use.
  - Need to distinguish:
    - Trunking from using two replicas simultaneously
    - Switching network addresses from Migration
- Issues with existing treatment of server scope
  - Unrealistic expectations about id sharing
  - Need to more explicit about Transparent State Migration

Future Documents

Overview

- Two I-Ds being worked on:
  - draft-adamson-nfsv4-mv0-trunking-update will address v4.0 issues
  - draft-dnoveck-nfsv4-mv1-msns-update will address v4.1 issues
- Will discuss these in later slides
- These I-Ds will not be submitted until after IETF99
- Could be basis for WG standards-track documents
  - Working group needs to discuss and decide

Future Documents
draft-adamson-nfsv4-mv0-trunking-update

- Co-authored with D. Noveck, C. Lever

- Parallels draft-dnoveck-nfsv4-mv1-msns-update
  - Adds trunking as an fs_locations attribute feature
  - Migration issues already handled in RFC7931

- Clarifies means of trunking detection/discovery

- Describes interactions of trunking addresses with the migration and replication fs_locations features

Future Documents: draft-adamson-nfsv4-mv0-trunking-update
Rework of Multi-server Namespace

- All proposed changes are to Section 8 of RFC7530

- Updated existing sub-sections to include trunking

- New "Trunking Discovery and Detection" sub-section
  - Uniform client ID not in use: RFC7931 trunking detection is problematic
  - Use server host names that resolve via DNS into multiple IP addresses

- New "Interaction of Trunking, Migration, and Replication" sub-section
  - Includes use of NFS4ERR_MOVED to indicate a shift in network addresses used to access a file system with no migration event

- Some new Security Considerations text

Future Documents
draft-dnoveck-nfsv4-mv1-msns-update

- Co-authored with C. Lever, A. Adamson
- Will address:
  - Lack of State Migration section in RFC5661
  - Need to deal with the possibility of session migration in v4.1
  - Differences in recovery model due to SEQ4_STATUS bits
  - Questions of relation to pNFS and of Session Migration
  - Mis-handling of trunking in RFC5661
    - Fixing this winds up providing trunking discovery for NFSv4.1
  - Relationship of Trunking and Migration

Future Documents: draft-dnoveck-nfsv4-mv1-msns-update

Rework of Multi-server Namespace

- Treats trunking as one of the functions of the location attributes
- Shift to network addresses separated from migration
    - Discussed in connection with changes in trunking discovery.
    - Always transparent because state is shared.
- More limited expectations regarding simultaneous use
- Server scope treatment omits discussion of state sharing
- Added discussion of pNFS options
- Explicit discussion of migration of lock and session state
    - New sections to clarify server responsibilities and client recovery path

Future Documents: draft-dnoveck-nfsv4-mv1-msns-update
New Approach to LEASE_MOVED

- V4.1 has SEQ4_STATUS_LEASE_MOVED indication rather than NFS4ERR_LEASE_MOVED.

- Allows operations to continue on non-migrated fs's while migrated fs's is dealt with

- Since it is not an error in v4.1, needs special handling
  - Does not to be dealt with immediately, but it should be dealt with to find migrated fs's
  - So that they are referenced, in time to avoid lease expiration
  - Goes on in parallel with normal operation and migration recovery
  - Needs to be ignored as long as the discovery process is going on.

Future Documents: draft-dnoveck-nfsv4-mv1-msns-update
Changes in Other Sections

- New section in Introduction about multi-server namespace

- New treatment of server scope

  - To match actual use.

- New description of NFSV4ERR_MOVED

  - Made necessary because switching between two addresses is not switching between two replicas.

  - Still covers both, but now these are treated as two different cases.

- Revised description of EXCHANGE_ID

  - Now CREATE_SESSION is not the *only* way to confirm a client ID.

  - Details in [Next Slide](#).

Future Documents: draft-dnoveck-nfsv4-mv1-msns-update

Revision of EXCHANGE_ID

- In RFC5661, EXCHANGE_ID must be confirmed by CREATE_SESSION
- Doesn't fit Client ID's transferred by Transparent State Migration:
  - They were confirmed on source server
  - But the client still needs a slot sequence value (unlike other confirmed client IDs)
- Revised description of EXCHANGE_ID:
  - Client ID often is confirmed by EXCHANGE_ID, but can be confirmed otherwise (without a CREATE_SESSION on destination server)
  - Slot sequence value is of no use when confirmed by CREATE_SESSION
    - But avoid previous RFC2119-inappropriate "MUST"

Next Steps

- WG needs to consider what the right update model is to be:
  - Current approach is one document for each of v4.0/v4.1.
  - Considered doing trunking separately and having it cover all minor versions
    - It didn't work out because v4.0 and v4.1 are so different
- If the decision is to stay with documents for v4.0 and v4.1,
  - Need to decide on WG standards-track documents.
  - Should decide on appropriate milestones
- Otherwise, need to come up with an alternate plan.

The End of migration-issues-xx

- Probably no need to publish as an RFC
  - Once the v4.0 and v4.1 issues have been dealt with, there would be no point.
- Am anticipating a final cycle of Working Group review
  - To make sure multi-server namespace is being correctly dealt with
  - Have been calling it WGLC but that might not be right.
- Need to clarify timing
  - Should be after there are WG documents addressing issues.
  - Should not need to wait for WGLC on those documents