NFSv4 In Depth

Spencer Shepler
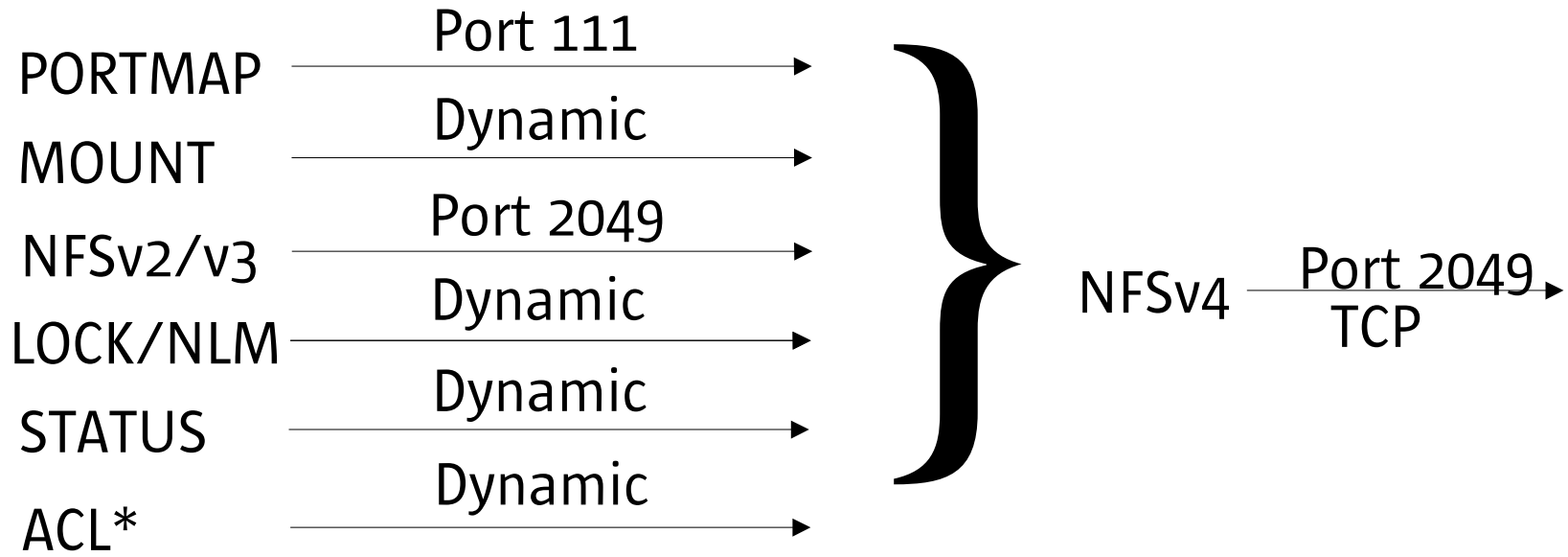
Senior Staff Engineer

Sun Microsystems

spencer.shepler@sun.com

NFSv4

- Brief review of operations
- Solaris client structure and its use of the NFSv4 operations
- Interesting errors

Many to One

PORTMAP → Port 111

MOUNT → Dynamic

NFSv2/v3 → Port 2049

LOCK/NLM → Dynamic

STATUS → Dynamic

ACL* → Dynamic

} NFSv4 → Port 2049 TCP

NFSv4 Operations

- COMPOUND allows for flexibility

- Client implementations are just starting to take advantage of this

- Expect more behavioral changes as time passes and minor version work is done

NFSv4 Operations

- ACCESS
- CLOSE
- COMMIT
- CREATE
- DELEGPURGE
- DELEGRETURN
- GETATTR
- GETFH
- LINK
- LOCK
- LOCKT
- LOCKU
- LOOKUP
- LOOKUPP
- NVERIFY
- OPEN
- OPENATTR
- OPEN_CONFIRM
- OPEN_DOWNGRADE

- PUTFH
- PUTPUBFH
- PUTROOTFH
- READ
- READDIR
- READLINK
- RENAME
- RESTOREFH
- SAVEFH
- SECINFO
- SETATTR
- SETCLIENTID
- SETCLIENTID_CONFIRM
- VERIFY
- WRITE
- RELEASE_LOCKOWNER

Ex: mount server:/test/dir

- Client generates this COMPOUND

  - PUTROOFH
  - GETFH
  - LOOKUP(test)
  - GETFH
  - GETATTR
  - SECINFO(dir)
  - LOOKUP(dir)
  - GETFH
  - GETATTR

Ex: mount server:/test/dir

- GETATTR is for these attributes

  - SYMLINK_SUPPORT
  - LINK_SUPPORT
  - FH_EXPIRE_TYPE
  - TYPE
  - SUPPORTED_ATTRS
  - FSID

Ex: mount server:/test/dir

- ## SETCLIENTID

  - Contains callback info (addr/program #)

  - Returns clientid used for OPEN

- ## SETCLIENTID_CONFIRM

Ex: ls file

- ls generates this COMPOUND

  - PUTFH
  - SAVEFH
  - LOOKUP (file)
  - GETFH
  - GETATTR
  - RESTOREFH
  - NVERIFY
  - GETATTR

Ex: Is directory

- Is of directory generates these COMPOUNDS

  - Getattr (of containing directory/cache chk)

  - Lookup (directory)

  - Access

  - Getattr (of containing directory – open)

  - Readdir (with full attributes if first readdir)

Ex: ls directory

- Readdir COMPOUND

    - PUTFH

    - READDIR

        - (first time with attrs)
        - SIZE, CHANGE, TYPE, RDATTR_ERROR, FSID, FILEID,
        - FILEHANDLE, OWNER_GROUP, OWNER, NUMLINKS, MODE, TIME_ACCESS, SPACE_USED, RAWDEV, MOUNTED_ON_FILEID, TIME_MODIFY, TIME_METADATA

    - LOOKUPP

    - GETFH

    - GETATTR

October 12-14, 2004

user/group mapping

- user@domain / group@domain
- Use of user-level nfsmapid
  - /etc/default/nfs NFSMAPID_DOMAIN
  - _nfsv4idmapdomain TXT RR
  - Local "domain"

Page 12 of 21

2004 NAS Industry Conference

Ex: open(2)

- Open COMPOUND
  - PUTFH
  - OPEN
  - GETFH
  - GETATTR

- Server may ask for OPEN_CONFIRM

- Stateid is returned

- Delegation may be returned
  - Type and stateid

Ex: read/write

- read/write COMPOUND

  - PUTFH
  - READ

- The operation carries a stateid

  - Delegation Stateid
  - Lock Stateid
  - Open Stateid
  - Special Stateid

  -

Ex: close (2)

- Flush of modified data

- Close COMPOUND

    - PUTFH
    - GETATTR
    - CLOSE

- If applicable, delegreturn most likely to come later

Recovery events

- Server reboot

  - NFS4ERR_STALE_CLIENTID
  - NFS4ERR_STALE_STATEID

- Network partition

  - NFS4ERR_EXPIRED

Interrupted open/close

- Client must ensure that the operation occurs at the server and clean up if applicable

Server Namespace

- Shares are managed the same as before

- Client must traverse namespace from root to target directory

- With MOUNT protocol, some intermediate permission checking avoided

- Clients may not traverse entired namespace with single mount

- NFS4ERR_WRONGSEC

RENEW

- NFS4ERR_CB_PATH_DOWN
- NFS4ERR_ADMIN_REVOKED
- NFS4ERR_LEASE_MOVED

Errors that are interesting

- NFS4ERR_GRACE
- NFS4ERR_NO_GRACE
- NFS4ERR_DELAY
- NFS4ERR_FHEXPIRED
- NFS4ERR_STALE
- NFS4ERR_MOVED
- NFS4ERR_WRONGSEC

NFSv4 In Depth

Spencer Shepler

Senior Staff Engineer

Sun Microsystems

spencer.shepler@sun.com