Lock Recovery Delimitation

or The Revenge of Section 8.6.3

Dave Noveck

60th IETF: August 3, 2004

What’s the problem here?

- Sec 8.6.3 describes two “edge conditions”
  - Says there may be more
  - Turns out there are ☹
  - Problems with interrupted lock reclamation
- Need to have a marker for end of reclaims
  - How to deal with this in v4.0
  - How to deal with this in v4.1

New Edge Condition

Network

- **A gets some locks**
- **Server reboots**
- **A reclaims some (but not yet all of its locks)**
- **Network partition**
- **Grace period ends**
- **B gets a lock which conflicts with one of A's unreclaimed locks**

8.6.3

- **Svr. boot tstamp is T**
- **A’s lock tstamp is T+1**
- **Svr. boot tstamp is T+2**
- **A’s lock tstamp is updated to T+2.5**

New Edge Condition (cont.)

Network (cont.)

- B releases that lock
- Server reboots (again)
- Partition heals
- A gets stale client id error
- A reclaims locks incl. one for which granting of lock to B was in conflict. *Oops!!*

8.6.3 (cont.)

- Svr. boot tstamp is T+3
- Prev. tstamp is T+2
- Reclaims succeed since T+2 < T+2.5 < T+3. *Oops!!*

So what’s going on?

- Is this just a problem with the algorithm?
  - Well, yes. It doesn’t work.
  - But no, it is deeper than that.
  - Server has to know when reclaims finished.
- If he doesn’t, then algorithm is not fixable
  - Unless server stores locks persistently,
  - He doesn’t know whether to grant reclaims
  - A reclaim not done, is like a revoked lock

Other Possibilities?

- Client handles incomplete reclaims
  - If I don’t get to reclaim a lock,
  - Then if the server reboots, client doesn’t reclaim that lock again
- Problems:
  - For v4.0, spec doesn’t say that
  - Reclaim may be OK given server capability
    - If server keeps locks persistently, it is OK.
    - Spec allows more capable & less capable servers

The Situation in V4.0

- How to delimit recovery period
  - Many opinions, much dispute
- What does spec say on issue?
  - Not much
- What does it imply?
  - Opinions differ

V4.0 Needs Cue for Reclaim Done

- Attempt at non-reclaim OPEN or LOCK?
  - Would mean future reclaims get NOGRACE
  - But spec doesn’t say I can’t do reclaims after attempting non-reclaim open’s
  - Spec also doesn’t say server has to accept it
  - Server even free to give NOGRACE always
- What about client not doing open
  - Need some heuristic (Lease time after last reclaim?)

Interoperability Issues

- Servers have flexibility on NOGRACE
  - But if they use it with abandon,
  - Reclaim won’t work very well
- Spec’s position is:
  - “So what, reclaim doesn’t have to work”
- Servers need to explain what they will do
  - The clients can adjust
  - If server is reasonable, and they should be.

What to do in V4.1?

- If server needs to know when reclaims are done, tell him
- DONE_RECOVERY op
- Fixes 8.6.3 problem
- Supports early grace termination
  - Big help
  - Many times you want long lease period
  - But don’t want to wait for long grace period

Early Grace Termination

- Servers with persistent locking can do it already
- What about servers that just store lists of client persistently?
  - If all clients have completed reclaims, then there can’t be any more
  - If there can’t be any more, you don’t need to return GRACE
  - Algorithm is simple. Life is good