

NFS version 4 and Beyond LISA 2006

Mike Eisler
Network Appliance, Inc.
email2mre-lisa@yahoo.com

NetApp[®] Outline

- Top 5 things to you need to know about NFSv4
 - Comparison of NFSv3 and NFSv4
 - Benefits
 - Misconceptions
 - Who has it?
 - Drawbacks
- Basic concepts
- Futures
- Pointers
- Questions



Comparison of NFSv3 and NFSv4

NFSv3

- A collection of protocols (file, mount, lock, status)
- Stateless
- UNIX-centric, but seen in Windows too
- Deployed with weak authentication
- ▶ 32 bit numeric uids/gids
- Ad-hoc caching
- UNIX permissions
- Works over UDP, TCP
- Needs a-priori agreement on character sets

NFSv4

- One protocol to a single port (2049)
- Lease-based state
- Supports UNIX and Windows file semantics
- Mandates strong authentication
- String-based identities
- Real caching handshake
- Windows-like access
- Bans UDP
- Uses a universal character set for file names

NetApp[®] Benefits

- Mandates strong security
 - Every NFSv4 implementation has Kerberos V5
 - You can use weak authentication if you want
- Finer grained access control
 - Go beyond UNIX owner, group, mode
- Read-only, read-mostly, or single writer workloads can benefit from formal caching extensions
- Multi-protocol (NFS, CIFS) access experience is cleaner
 - NFSv4 has an OPEN operation; thus CIFS clients can't disrupt NFSv4 clients
- Byte range locking protocol is much more robust
 - Recovery algorithms are simpler, hence more reliable



Misconceptions

- NFSv4 is a new protocol, so I can use more than 16 supplemental gids?
 - No, the 16 gid limit is a property of the weak authentication flavor of the remote procedure call
 - Use Kerberos V5, and you can go beyond 16 gids
 - Limited primarily by server's operating system and server's local file system
- ▶ I need NFSv4 in order to use Kerberos V5, right?
 - No, Kerberos V5 works on NFSv[23] too and has for years on AIX (IBM), EMC, Hummingbird, Linux, NetApp, Solaris

Lec NetApp°

Who Has NFSv4?

- ▶ IBM (AIX 5.3)
- **EMC**
- Hummingbird
- Network Appliance (best is 7.x)
- FreeBSD 5.3
- Linux 2.6 (Fedora Core)
- OSX (Rick Macklem, not Apple)
- Solaris 10
- 2 others tested at Connectathon 2006



Basic Concept: Delegation

- A delegation is a grant from an NFSv4 server to a client for rights to perform read-only or read/modifying operations on a particular file
- With a read-only delegation, multiple NFSv4 clients can cache a file with impunity
 - With NFSv3, a client that caches a file would periodically send GETATTRs to re-validate its cache
 - Some workloads are absolutely hammered with GETATTRs even after the customer carefully tunes his clients to cache the workload's working set
- With a write delegation, a single NFSv4 client can cache and modify a file with impunity
 - Useful for applications like home directories where the data set owner tends to be the only reader and writer



Basic Concept: Referrals

- NFSv4 has hooks for data migration
- When a file system moves from one server to another, the NFSv4 client receives an NFS4ERR_MOVED error from the original server
- The NFSv4 client issues a GETATTR for the "fs_locations" attribute to tell the client which server has the file system, and the location within the new server
- Removes NFS mount/server IP address straitjacket



- ► Fewer implementations than NFSv3
 - OSDL has publicly pronounced NFSv4 (kernel.org) as "ready"
 - Enterprise Editions of major Linux distributions don't fully support NFSv4 or Kerberized NFS
- Not all features uniformly implemented right now
 - NFSv4 referrals turned out to be the most compelling to customers, but are the least completely implemented of all NFSv4 features

NetApp°

Futures: NFSv4.1

- Sessions and Exactly Once Semantics
- Directory delegations
- ► RDMA
 - Origins in Direct Access File System (DAFS)
 - Early access (Linux) for NFSv[34] available now
- Parallel NFS
 - Single File I/O can be served by multiple data servers
 - E.g. a file blocked at 1024 bytes, striped over 3 servers, might have
 - offset 0 served by data server0
 - offset 1024 served by data server1
 - offset 2048 served by data server2
 - offset 3072 served by data server0
 - ...
 - 3 styles of data servers: blocks, files, objects
 - Linear scaling is possible

NetApp[®] Pointers

- www.nfsv4.org
- <u>ietf.org/html.charters/nfsv4-charter.html</u> NFSv4 working group page at IETF
- <u>www.ietf.org/rfc/rfc3530.txt</u> The protocol specification for NFSv4
- Blogs
 - Some co-authors of NFSv4:
 - Eisler: nfsworld.blogspot.com
 - Shepler: blogs.sun.com/roller/page/shepler/Weblog?catname=%2 FNFS
- Linux NFSv4 client:
 - wiki.linux-nfs.org/index.php/Main_Page
 - linux-nfs.org/cgi-bin/mailman/listinfo/nfsv4
- OS X client:
 - <u>ftp.cis.uoguelph.ca:/pub/nfsv4/darwin-port/xnu-client.tar.gz</u>
- Linux NFS/RDMA client and server: http://sourceforge.net/projects/nfs-rdma/



Questions?



Backup Slides

NetApp[®] Acronyms

- ONC RPC Open Network Computing Remote Procedure Call: used by NFS
- GSS Generic Security Services: allows security mechanisms like Kerberos V5 to plug into a common programming interface for security
- AUTH_SYS UNIX System Authentication: weak authentication for ONC RPC and NFS
- RPCSEC_GSS GSS-based security flavor for ONC RPC and NFS
- ACE Access Control Entry: consisting of a uid or gid, permissions, deny/allow
- ACL Access Control List: a list of ACEs for a file
- GETATTR NFS Get Attribute operation
- UTF8 (8-bit Unicode Transformation Format) is a variable-length encoding for Unicode. US-ASCII characters go out in 8 bits; other locale character sets require 16 bits or more