NFS version 4
LISA `05

Mike Eisler

Network Appliance, Inc.

email2mre-lisa@yahoo.com

▶ **Top 5 things to you need to know about NFSv4**

▶ **Acronyms**

▶ **Basic concepts**

▶ **Futures**

▶ **Pointers**

▶ **Questions**

▸ **Comparison of NFSv3 and NFSv4**

▸ **Benefits**

▸ **Drawbacks**

▸ **Misconceptions**

▸ **Who has it?**

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

▸ **Mandates strong security**

– **Every NFSv4 implementation has Kerberos V5**

– **You can use weak authentication if you want (but why would you?)**

▸ **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**

▸ **Byte range locking protocol is much more robust**

– **Recovery algorithms are simpler, hence more reliable**

Drawbacks

- **Fewer implementations than NFSv3**
  - But unlike with NFSv3, Linux is not the laggard
- **Not all features uniformly implemented right now**

▶ **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**

▶ **I need NFSv4 in order to use Kerberos V5**

- – **No, Kerberos V5 works on NFSv[23] and has for years on EMC, Hummingbird, NetApp, Solaris**

▶ **IBM (AIX 5.3)**

▶ **EMC**

▶ **Hummingbird**

▶ **Network Appliance (best is 7.0.x)**

▶ **FreeBSD 5.3**

▶ **Linux 2.6 (Fedora Core)**

▶ **Solaris 10**

▶ **More to come …**

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**

▸ **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**

▸ **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**

▸ **Directory delegations**

▸ **Implementation identifiers**

   – **What types of clients are talking to my server?**

   – **What types of servers are talking to my clients?**

▸ **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 server0**

      • **offset 1024 served by server1**

      • **offset 2048 served by server2**

      • **offset 3072 served by server0**

      • **…**

▸ **www.nfsv4.org**

▸ **ietf.org/html.charters/nfsv4-charter.html – NFSv4 working group page at IETF**

▸ **www.ietf.org/rfc/rfc3530.txt - The protocol specification for NFSv4**

▸ **Blogs**
  - **Some co-authors of NFSv4:**
    • **Eisler: nfsworld.blogspot.com**
    • **Shepler:blogs.sun.com/roller/page/shepler/Weblog?catname=%2FNFS**

▸ **Linux NFSv4 client:**
  - **wiki.linux-nfs.org/index.php/Main_Page**
  - **linux-nfs.org/cgi-bin/mailman/listinfo/nfsv4**

Questions?