SerNet

Distributed File Systems compared

SambaXP 2005

Göttingen

2.-4. Mai 2005

Volker Lendecke

SerNet - Service Network GmbH

Network Service in a Service Network

SerNet

Volker Lendecke

- Co-founder SerNet - Service Network GmbH
  - Free Software as a successful business model
  - Network Security for the industry and the public sector
  - Samba-Support/Development in Germany
- For 15 years concerned with Free Software
- First patches to Samba in 1994
- Consultant for industry in IT questions
- Co-founder emlix GmbH (Embedded Systems)

samba

Network Service in a Service Network

SerNet

Questions in a distributed file system

- How close is the coupling?
- Which semantics do the clients see?
- Who authenticates against whom, and how?
- Who controls access?
- Caching?
- Scalability?

Network Service in a Service Network

SerNet

Overview

- GFS, NFS, CIFS and AFS briefly introduced
- Authentication
- Access Control
- Locking
- Caching
- Scalability

Network Service in a Service Network

Global File System GFS

- Cluster-Dateisystem by RedHat (formerly Sistina)
- Coordinated access to a shared block device
- very close coupling
- Posix-Semantics closely followed
- Various protocols (SAN, TCP/UDP) for the different tasks
- GFS-nodes not protected against each other security-wise

SerNet

Network File System NFS

- Early Unix network file system
- Version 2: stateless, UDP-based
  - no caching/consistency
  - no access control, based on client IP address
- Version 3: TCP possible
  - Caching is being coordinated
- Version 4
  - Yet another ACL-definition
  - Caching, locking being done

Common Internet File System CIFS

- Neither common, nor Internet, and **not** a file system....
- started as DOS-File system calls on the net
- Many different dialects, beyond any attempt of documentation
- Various kinds of authentication
- Transport for several other protocols
- State-based protocol
- almost NTFS-Semantics on the net

SerNet

Andrew File System AFS

- Distributed file system from the Carnegie Mellon University
- Maybe the only file system really being „distributed“
- Kerberos-4 based authentication, 5 in the works
- aggressive disk-based caching
- Central mount point under /afs
- Data can be stored „anywhere“

samba

Network Service in a Service Network

Stateless / Stateful

- ... or: „Is there an open call“
- Stateless server is probably less code
- Knowing the client's state makes many optimizatinos possible
- Stateless protocols make authentication **very** hard
- Caching not possible

Authentication

- GFS, NFSv2: No authentication, client OS is trusted
- CIFS: User-based authentication per server, workstations cache passwords/tickets
- AFS: Kerberos 4, same ticket for all servers
- CIFS/Active Directory: Kerberos 5, fall back to NTLM
- NFSv4: Potentially Kerberos 5

Access control

- Not really protocol-bound, that's an implementation feature
- GFS, NFSv2/v3: Client-based, Posix
- NFSv4: Yet another ACL model (Windows, but not quite)
- CIFS/Windows: Complex ACLs
- CIFS/Samba3: Posix
- CIFS/Samba4: Posix, Optional Windows Semantics
- AFS: Own model, ACLs only per directory

Locking

- Locking is a means to get exclusive access to a resource
  - Complete file locks (Windows: „Share modes“)
  - Ranges of a file (byte range locks)
  - Advisory (Posix) vs Mandatory (Windows)
- Windows share modes have a set of rules to allow/deny based on existing locks and context, partly based on the file name
- Byte range locks are different between Posix and Windows (64-bit range, overlapping locks etc)

Locking

- GFS: Full posix locking: no per file locking, advisory byte range locks
- CIFS: Windows-compatible locking on the wire
  - cifs->Samba: Posix compatible locking
- NFSv2: No locking, protocols on their own
- NFSv4: Try to do both, but:
  - It is assumed that manipulating a lock is rare when compared to READ and WRITE operations.
- AFS: Locking on a file basis, no byte range locking

Caching

- GFS: block-based
- NFSv2: heuristics, no coordination
- CIFS, NFSv3/4 and AFS delegate the permission to cache (Oplocks, Leases, Delegations, Callbacks)
- AFS implemenations: disk-based cache based on version ID.

Local semantics vs scalability

- Local Semantics, for example:
  - Atomic operations (creat, fcntl locks)
  - Read/Write and mmap consistency?
- Scalability:
  - As little communication as possible
  - Defined inconsistencies
  - Latencies are **very** important

Scalability

- Ways to achieve scalability:
- Make single machines larger
- Distribute load over several machines
- IP-based server farm load balancing farm (Apache, Squid)
- Intelligently redirect clients

Scalability

- GFS: Whitepaper says up to 300 clients
- NFS/CIFS: Made for single servers, this gives limits
- Modern versions can redirect clients
  - CIFS: MS-DFS
  - Very static from a client's point of view
  - NFSv4 can redirect clients
- Stateless NFS can be load-balanced
- CIFS has too much state for real load balancing

AFS scalability

- Location-transparent file names under /afs
- Physical storage location in a database
- Replication of read-only data (Load Balancing)
- Disk-based caches for less load on servers
- Transparent data relocation while data is in use
- Written data only visible after close

What should I choose?

- Look **closely** at your environment and workloads
- Do you need the close coupling? High performance clusters might make use of GFS
- Do you have Windows clients around? **DON'T** touch the clients – Samba
- Do you have a WAN involved? GFS doesn't like that, something else on top, maybe AFS

Linux to Linux

- 5.000 Linux Clients – How do I share /home?
- Traditional: NFS, No File Security, Needs Fifty Sysadmins, etc...
- NFS relies on the client to do access checking, this might have worked for closely monitored multi-user machines
- AFS has per-user credentials, but is too complex to set up
- cifs/samba to the rescue?

SerNet

Questions/comments?

Volker Lendecke, VL@SerNet.DE

SerNet – Service Network GmbH
Bahnhofsallee 1b
37081 Göttingen

Tel: +49 551 370000 0
Fax: +49 551 370000 9
http://www.SerNet.DE

http://Samba.SerNet.DE

Network Service in a Service Network