SerNet

Clustered SMB with Samba (4) and CTDB (2)

UK GPFS User Group Meeting, 2013-04-24

Michael Adam

SerNet GmbH, Göttingen / Berlin, Germany

/me

SerNet

- Michael Adam (with**out** trailing “s” ...;-)
- Background: Mathematics
- Developer of Samba and CTDB
- Samba Team at SerNet
- Nick “obnox” on freenode

SerNet GmbH

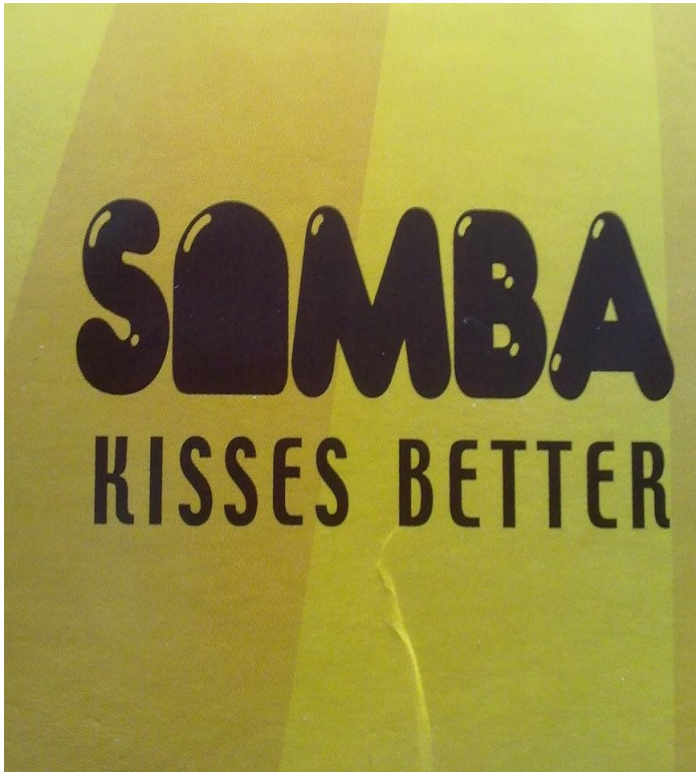
SerNet

- founded 1997, offices in Göttingen and Berlin (Germany)
- topics:
  - Informations security and data protection
  - Open Source, firewalls, VPN, mail, group ware, ...
  - certification, audits, verinice. (Open Source ISMS tool)
- Samba:
  - Support, Consulting, Development as a service
  - 5 Samba team members, release-management
  - sambaXP: annual conference

Samba

SerNet

- inter-op: Unix/Linux ↔ Windows
- *The* Open Source SMB file server (Samba 3.X)
- *New*: Active Directory (Compatible) Server (Samba 4.0)
- http://www.samba.org/
- Started in 1992
- some 15-20 active developers



Samba 3.X

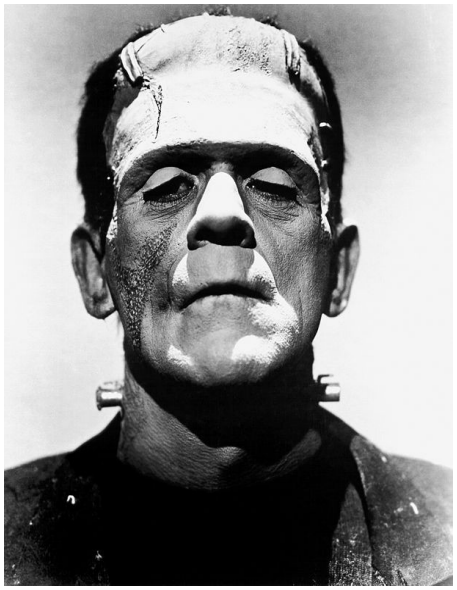
SerNet

- Stable: Linux distributions, products…
- Bug-fix mode. latest: 3.6.13
- Components:
  - SMB server (smbd)
  - all-active SMB Clustering with CTDB (ctdbd)
  - Active Directory domain member (winbindd)
  - NetBIOS name service / network neighbourhood (nmbd)
  - Windows NT like domain controller

Samba 4.0

SerNet

- 4.0.0: released December 11, 2012
- → almost a cool date... ;-)
- Direct continuation of the 3.X file server releases (with new features)
- *Great new feature*:
  Active Directory-Compatible Server
  (After 10 years of development...)
- Latest: 4.0.5



4.0 – also a great file server release!

SerNet

- SMB 2 (Windows Vista)
  - durable handles
- SMB 2.1 (Windows 7)
  - Basics
  - Multi-credit / large MTU
  - Dynamic reauth
- SMB 3 (Windows 8 / Server 2012)
  - Basics, crypto, secure negotiation
  - durable handles v2

CTDB: clustering Samba (challenges)

SerNet

- Prerequisite: clustered file system (like GPFS), POSIX
- Goal: all-active SMB cluster with Samba on top
  (available, scalable, ...)
- → Samba instances need to appear as *one* SMB server.
  No client changes (pre-windows-2012)
- Samba stores Windows-level open file info in TDB Databases
  for IPC (due to multi-process architecture!)
  - `locking.tdb` (open files/share modes/oplocks)
  - `brlock.tdb` (byte range locks)
- → Need clustered TDB database (CTDB)

CTDB: clustering Samba (ideas)

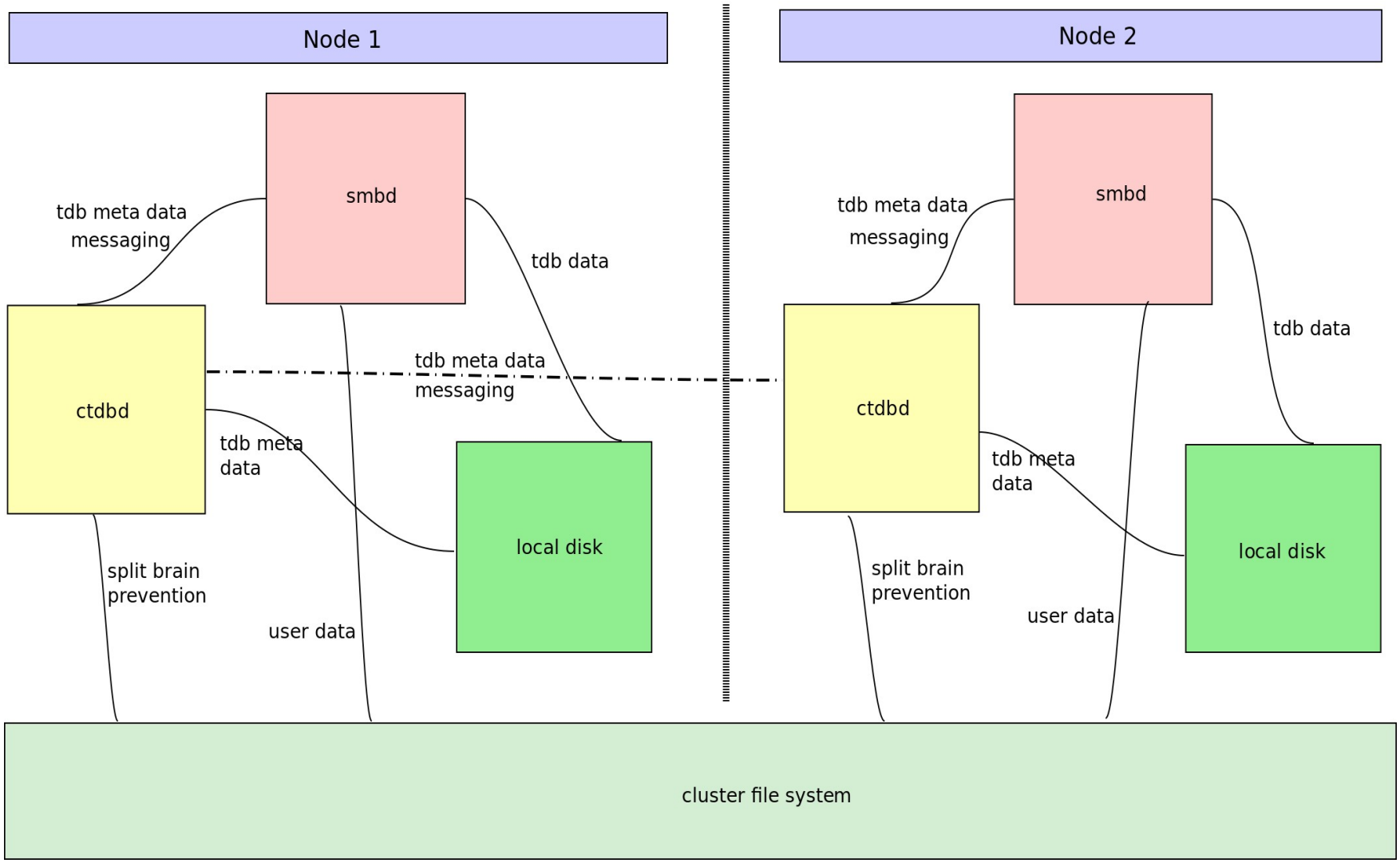
SerNet

- Clustered TDB database and messaging engine
- Very, very special clustered database:
  - may loose certain records!
  - Nodes associated to a record:
    - DMASTER (data master, moving)
    - LMASTER (location, calculated)
  - Special node in Cluster:
    - recovery master (elected)

- Samba accesses record content directly (→ fast!)

CTDB: clustering Samba (layout)

SerNet



CTDB: why does it work, really?

SerNet

- locking.tdb:
  - One record per file, entries for each open handle:
    - → by opening server
    - share mode
    - oplock
- brlock.tdb:
  - One record per file, entries for byte ranges:
    - → by server holding byte range
- Recovery + automatic clean-up of record entries
  → makes node failure (relatively) safe

CTDB: good and bad workloads

SerNet

- *Good*: a file is only open on a single node
  - Records only on the opening node
- *Bad*: files open on multiple nodes
  - Record ping-pong (migrations)
- *Ugly*: reading a large file in parallel
  - → strict locking (brlock)
  - → read only record copies (1.2+)

CTDB: hard to get right

SerNet

- Vacuuming
- Determinism (record chasing)
- Robustness in failure situations
- Scaling
- Persistent Databases

CTDB: history and community

SerNet

- Started 2006 (Volker Lendecke, Andrew Tridgell)
- First usable version 2007
- Maintainer today: Amitay Isaacs
- http://ctdb.samba.org/ (also http://wiki.samba.org/)
- Mailing list: samba-technical@lists.samba.org
- IRC: `#ctdb` (and `#samba-technical`) on freenode
- Code: `git://git.samba.org/ctdb.git`
- Versions(somewhat): 1.0.114(.x), 1.2.40+, 2.X (master)

Demo

SerNet

SMB 3 – what's next?

SerNet

- Leases (2.1)
- Replay / Retry
- Multi channel
- Persistent handles
- SMB over RDMA
- Witness protocol
- Transparent fail over
- Scale-Out shares
- Storage features / snapshots
- ...

Coming next...

SerNet

- on http://www.enterpriseamba.com/(by SerNet):
  - Samba 4.0 packages
    - Debian/Ubuntu
    - RHEL/CentOS
    - SLES/openSUSE
- sambaXP conference:
  - http://www.sambaxp.org/
  - May 14—17, 2013
  - In Göttingen

Contact

SerNet

Michael Adam / ma@sernet.de

SerNet GmbH
Bahnhofsallee 1b
37081 Göttingen
Germany

tel +49 551 370000-0
fax +49 551 370000-9

Schützenstr. 18
10117 Berlin
Germany

+49 30 5 779 779 0
+49 30 5 779 779 9

http://www.sernet.de/