SerNet

How to squeeze 700MB/sec out of SMB1

SNIA Storage Developer Conference

Santa Clara

September 2009

Volker Lendecke

SerNet

Samba Team

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 almost 20 years concerned with Free Software
- First patches to Samba in 1994
- Consultant for industry in IT questions
- Co-founder emlix GmbH (Embedded Systems)

Network Service in a Service Network

SerNet and Samba

- technological leadership of SerNet worldwide
  - involved in almost every big European Samba project
  - 5 out of 6 European developers work for SerNet
  - SerNet distributes up-to-date Samba packages
- samba eXPerience
  - *The* international Samba conference
  - > 150 developers & users from > 15 countries

SMB

- The old protocol we know for many years now
- SMB has seen many protocol extensions, but the basic header structure remained unchanged for many years
- Developed initially for very small machines
  - Limited memory
- Targeted at local area networks
  - Not very well suited to high-latency networks
    - Is that really true? :-)

CIFS

- Microsoft's reply to WebNFS
- With WebNFS SUN added some extensions to work around deficiencies
  - One lookup per path element
  - SMB never had this particular problem
- The only things MS changed:
  - *SMBSERVER
  - Get rid of the NetBIOS-level session setup
    - Client's don't know the server's names

Current networks

- Gigabit Ethernet best practice these days
- 10GigE available
  - I've still seen a lot of driver and performance issues until very recently
  - At most datacenter-only technology
- Latency has not improved as much as bandwidth:
  - A quick non-scientific test showed .5ms over 100MBit and .13ms over GigE. 10GigE not available here, but latency is **not** .013ms.

Windows use of SMB

- Tons of round-trips
- Directory listings open every single file
- Explorer extensions try to open filename::{GUID}
- Read and write sizes typically less than 64k
  - Signed r/w requests at 16k or less
- Writing a large block from Win32 apps leads to 1-byte writes at the end of that block to detect disk-full

expected_throughput.pl

- Little script by Tridge:
  - Based on server throughput, network bandwidth, network latency and buffer size calculates the maximum expected throughput
- At 5GB/sec server throuput (i.e. unlimited) and estimated .1 msec latency for 10GigE

	64k	16k	4k
100 MBit	9.27 MB/s	7.65 MB/s	4.5 MB/s
GigE	81.57 MB/s	54.23 MB/s	23.17 MB/s
10GigE	366.85 MB/s	136.92 MB/s	39.04 MB/s

expected_throughput.pl

- Windows uses 10GigE only to 36% at most
- Round-Trips kill performance
- That's what the TCP window (including TCP Window scaling) is supposed to solve
- Windows limits the effective TCP Window Size to at most 64k

SMB used right

- SMB is a multiplexed protocol
  - MID field identifies requests
  - Server is free to reply in the order it wants
- Samba 3.2 smbclient uses the MID field to fill the pipe with read and write requests
  - Criticial piece in the implementation: Send data before receive to keep the pipes busy
  - >700MB reading from RAM disk over 10GigE where a raw TCP test was able to ship 800MB

SMB2

- SMB2 has an elaborate credits system
- A server can allow a client a specified number of requests to keep open at a time
  - Some documents speak about this is done to throttle clients
- This sounds like TCP Window calculation in user space
- Question: WHY?? If a server is overloaded, just don't do the work and let the TCP stack throttle clients

SerNet

Demo

- Watch SMB1 and SMB2 on the wire

Network Service in a Service Network

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