

Server Tuning

or

"How I learnt to dance with Elephants"

Jamie Baddeley

REANNZ Workshop

2006



Introduction

➤ Councillor for InternetNZ

- Network with like-minded people dedicated to development of the Internet both here and abroad.
- Not-for-profit membership organisation advocating that the Internet be available to everyone, everywhere, all the time.
- Democratically run and active participation is welcome.
- Supports and encourages the coordinated and cooperative growth of the Internet in New Zealand.
- We are New Zealand's Internet representatives at a global level through our role in holding the delegation of the .nz country code top level domain.
- Independent source of policy that is of benefit to the *local* Internet Community

Introduction

- Also General Manager – Internet Services – FX Networks, which also operates an 10Gbps Network, similar to Karen, but is commercial not experimental.
- Real Life practical experience with Gigabit services and host issues
- Thanks to Perry Lorier and the good folks at WAND

Elephant?

- LFN – Long Fat Network (e.g Karen/FX etc)
- Going for the easy win....
- Bandwidth Delay Product
- How can we fix it?
- Host Issues
- Other Issues
- I speak english. I am not from Planet Acronym.
Ask me questions.

GoogleWords (tm)

- BDP, TCP receive window, LFN, Window Scaling, RFC1323, SACK

Going for the easy win..

- The options below are presented in the order that they should be checked and adjusted.
 - **Maximum TCP Buffer (Memory) space**
 - Socket Buffer Sizes
 - **TCP Large Window Extensions (RFC1323)**
 - TCP Selective Acknowledgments Option (SACK, RFC2018)
 - Path MTU

Bandwidth Delay Product

- Public Enemy #1
- Affects TCP flows – not UDP
- Your ability to fill the pipe depends on how much data TCP allows to have “in flight” before it waits for an acknowledgement back.
- Waiting around is waste of time and bps.
- Some host operating systems are not optimised for networks that are LFN.
- $BDP = \text{Max_BW} \times \text{RTT}$

Beating BDP

- Why is BDP an issue?
- For many systems you have deployed, Default TCP Receive Window is too small for Elephants, but *mostly* OK for the Interweb
- Affects a single TCP flow. Doesn't affect firewalls, does affect proxies (PEP) doesn't affect routers, does affect end systems
- On default Win2K..
 - FS DSL to the USA – 1.17 Mbps @ 120ms
 - LFN to the USA – 1.17 Mbps @120ms
 - Christchurch to Hamilton via LFN – 7 Mbps @ 20ms

Beating BDP

- Change your default TCP receive Window. (TCP receive buffer)
- Be aware of the problem you're trying to solve – 1Gbps to the other end of the country or really fast to the other end of the world?
- Oversized buffers can be bad – especially during loss.
- Remember Memory is a finite resource
- 16MB is good for East Coast US. Now multiply that by the number of long held flows.

Host Systems

- Default Settings on a range of hosts:
 - Windows 98 8192 bytes
 - Windows 2000 17520 bytes
 - Windows XP 17520 bytes
 - Windows Vista - TCP Receive Window Auto-Tuning
 - Linux 64 KB (auto tuning used since 2.4.27/2.6.7)
 - FreeBSD 256 KB
 - Netware 6 - 1GBytes
 - Links to more later

Other Issues

- JumboFrames (9000 byte MTU)
- Good to have, but not essential
- Really good in the event of packet loss
 - 1Gig flow @ .1% loss @ 1500 bytes = 28Mbps
 - 1Gig flow @ .1% loss @ 9000 bytes = 162Mbps
- Update core/distribution/access/desktop – in that order to avoid pmtu hell.
- Possibly requires a total refit of your network, including all NICs on your desktops. Ugh.

Linkage

- <http://noc.fx.net.nz/calculator/>
- <http://www.psc.edu/networking/projects/tcptune/>
- <http://www.ietf.org/rfc/rfc1323.txt>
- <http://www.abilene.iu.edu/JumboMTU.html>

Thank You

Jamie Baddeley

Councillor

jamie.baddeley@internetnz.net.nz

General Manager – Internet Services

jamie.baddeley@fx.net.nz