The Future of Congestion Control
Feedback from an Implementor’s Perspective

Lawrence Stewart
lastewart@swin.edu.au
Centre for Advanced Internet Architectures (CAIA)
Swinburne University of Technology
Outline

1 Recap

2 Experimental Tools

3 Some Results

4 General Thoughts for the RG

5 Wrapping Up
Who is this guy (and who let him past security)?

- Centre for Advanced Internet Architectures, Swinburne University (2003-2007)
  - Research assistant/engineer during/after studies

- Currently a PhD candidate in telecomms eng at CAIA (2007-)
  - Main focus on transport protocols & congestion control
1 Recap

2 Experimental Tools

3 Some Results

4 General Thoughts for the RG

5 Wrapping Up

1 Recap

- Where are we today
- Open issues
Where are we today

- Many incremental (partially implemented) improvements
- State of the CC union
  - NewReno is defacto standard with warts (LFN, wireless)
  - Many new proposals
  - BSD still uses NewReno
  - Linux uses CUBIC
  - Windows Vista uses Compound
- TCP/IP stack enhancements e.g.
  - CSO/TSO/LRO/TOE
  - Various locking/caching tricks
  - Socket buffer autotuning
Open issues

- High-speed CC algorithms\(^1\)
  - FAST, HS-TCP, H-TCP, CTCP, CUBIC, etc.
- Delay based CC algorithms
- How do we compare and evaluate TCPs?
- Multipath
- CSO/TSO/LRO/TOE obscure behaviours
- Testing/verification of TCP/IP stack behaviour

\(^1\) Nice summary:  
http://kb.pert.geant2.net/PERTKB/TcpHighSpeedVariants
Detailed outline (section 2 of 4)

1 Recap

2 Experimental Tools
   - FreeBSD As A Research Platform
   - TCP Testbed

3 Some Results

4 General Thoughts for the RG

5 Wrapping Up
FreeBSD As A Research Platform

- Modular congestion control
  - In svn project branch, coming to FreeBSD 7 and 8 soon
  - BSD licenced Newreno, HTCP & CUBIC implementations available
  - Sponsored by Cisco Systems

- Statistical Information for TCP Research (SIFTR)
  - FreeBSD kld to gather CSV in-kernel TCP endpoint connection data
  - Similar concept to Web100 with more variables
  - Sponsored by Cisco Systems and the FreeBSD Foundation

- Deterministic Packet Discard (DPD)
  - Adds 'pls' (packet loss set) option for dummynet pipes
  - e.g. ipfw pipe 1 config pls 1,5-10,30 would drop packets 1, 5-10 inclusive and 30

- Dummynet Forensic logging support
  - Log queue state on each packet event
Testbed

- Linux/FreeBSD hosts
- Modular congestion control
- Web100/SIFTR for Linux/FreeBSD testing
- Iperf/Tcpreplay for traffic generation
- FreeBSD dummynet router
- Endace DAG 3.7GF capture card
Detailed outline (section 3 of 4)

1 Recap

2 Experimental Tools

3 Some Results
   - Connection Dynamics
   - Collateral Damage
   - Subtle Queuing Implications

4 General Thoughts for the RG

5 Wrapping Up
Connection Dynamics

- 1 TCP flow, H-TCP, 100ms RTT, 1Mbps, 60000 byte queue
Collateral Damage

- Induced delay: 1 TCP vs 1 CBR UDP flow, 50ms RTT, 1Mbps, 60000 byte queue

CDF
- newreno
- htcp
- cubic

delay (ms)

0 100 300 500

0.0 0.4 0.8

CDF
Induced delay: 1 TCP vs 1 CBR UDP flow, 50ms RTT, 1.5Mbps/256Kbps, 20000 byte queue
Collateral Damage

- Retransmissions: n TCP vs 1 CBR UDP flow, 50ms RTT, 1Mbps, 60000 byte queue

![Graph showing average retransmits vs number of flows for newreno, htcp, and cubic protocols.](http://www.caia.swin.edu.au)
Subtle Queuing Implications

- Induced CBR loss: 1 TCP vs 1 CBR UDP flow, 100ms RTT, 1.5Mbps/256Kbps, NS
Detailed outline (section 4 of 4)

1 Recap

2 Experimental Tools

3 Some Results

4 General Thoughts for the RG

5 Wrapping Up
General Thoughts for the RG

■ I-Ds MUST:
  ■ Stipulate units of variables
  ■ Explicitly define default state for variables not discussed in I-D
  ■ Provide formulae for byte-based and pkt-based stacks

■ I-Ds SHOULD:
  ■ Provide more concrete help to implementors e.g. public domain code snippets

■ RG SHOULD:
  ■ Explore the design & impact of compatibility mechanisms more thoroughly
  ■ Solicit independent implementations more vigorously along with reviews
Detailed outline (section 5 of 4)

1 Recap
2 Experimental Tools
3 Some Results
4 General Thoughts for the RG
5 Wrapping Up

Further Information
Acknowledgements
Questions
Further Information

- **Papers**

- **Links**
  - [http://people.freebsd.org/~lstewart/](http://people.freebsd.org/~lstewart/)
Acknowledgements

- Cisco Systems

- The FreeBSD Foundation
tp->t_state = TCPS_QUESTIONS