Linux TCP

Stephen Hemminger
shemminger@linux-foundation.org

Sr. Staff Engineer
The Linux Foundation

Linux Networking Features

- Receive aggregation (NAPI)

- Hardware Checksum Offload

- TCP Segmentation Offload

- Advanced Congestion Control

  – BIC, H-TCP, Westwood

- Modularization of Congestion Control

- Appropriate Byte Count (ABC)

Linux process

- Collective Intelligence
  - Input from research community
  - Integration/Review by network developers
  - continual release
- Trust
  - Certificate of Origin
  - User testing

Issues with TCP research

- "Brand name" TCP
  - put multiple changes into one version
  - hard to isolate
- Code quality/relevance
  - Each conversion step can introduce problems
- Infrastructure changes affect timing
  - Researchers avoid NAPI, TSO effects
- Testing
  - Need common testbeds

Deployment problems

- Applications can do anything

  - Java debugger 6 writes per RPC

- Middle boxes suck

  - improper TCP state tracking

- Need to be self tuning

  - Mobile Linux vs. 2048 CPU SGI box

Multiple congestion controls are OK

- Like filesystems:
  - supporting multiple not a problem
  - need one or more robust moder versions (ext3)
  - keep one legacy version (ext2)
- There is no "one answer"
  - but most distro's will follow default
- Supporting & growing the community is critical

Linux – research ideas

- Project support

- Summer of Code?

- Cross pollination?

Reporting problems

- Do
    - ask questions
    - provide complete information
    - mail to me and netdev@vger.kernel.org
- Don't
    - Cry Wolf
    - Wait until you have a full paper