Networking Kernel Summit

July 15-16
Beaverton Oregon

Attendees

David Miller

Stephen Hemminger

Arnaldo Carvalho de Melo

Harald Welte

James Morris

Rusty Russell

Yoshifuji Hideaki

Nivedita Singhvi

Soyoung Park

Chris Wright

Topics

- IPV6
- Netfilter
- Security
- TCP
- 2.7 plans

IPV6 routing

- Load balancing
  - multiple equivalent routes
  - should be hash based not round robin
- Multicast routing
  - don't want to clone existing Ipv4 mess
- NO NAT!

Mobile IPv6

- Needs additions to xfrm mechanism
- Mobile IP vs. router daemon table sharing
- PF_MOBILITY

IPV6 kernel issues

- Avoid unnecessary fragmentation
- Add recvmsg option with info about security
- Support u64 device counters
- More generic tunneling abstraction
- Long running kernel timers
- net/ip directory restructuring

Security

- CIPE is mess, forget it. Use OpenVPN
- Assembly optimized crypto
- Symmetric crypto
- Hardware acceleration
- Async crypto interface
  - needs to be able use from bh

Netfilter

- Need to look at performance
  - Simple fw has 250 rules (20 per pkt)
  - grand unified flow cache???
- packet tables generalization
- non-linear skb's (hurt)
- libqsearch -- pattern matching

802.11 stack(s)

- Prism card stack (HostAP)
- Originally BSD implementation
  - device driver passes packets to magic stack
  - only tested with Atheros,
- wlan-ng - handful of vendors slow update
- dave/acme - whole stack but half finished.

Random

- UDP frag id wrap issue– don't do NFS over UDP please
  - interesting intellectual exercise
- RDMA and off load
  - TOE bogus argument: network's getting faster need to handle more data
  - valid argument: how do we do RDMA?
- Receive Collation Offload (RCO)
- Network Async I/O

Cleanups

- Make all protocols use common code from TCP
- To many copies of same X.25 code

TCP

- Recent work
  - congestion control (Westwood, BIC, Vegas)
  - receive auto-tuning, scaling
  - memory management
- Working with research community (web100, etc)

Window Scaling

- Too many buggy devices out there
  - Cisco IOS can corrupt SACK block sequence numbers when scaling used
  - Older versions of tcp-window-tracking patch in netfilter buggy
  - Some ADSL devices corrupt the window scale TCP option when port forwarding?

TCP future

- silly window stall timer (RFC1122)
- BIC 1.1
  - SACK speedup
  - burst moderation
  - CUBIC

Performance

- IPV4 vs IPV6
- congestion control choices
- network modeling results
- RCU route overload
- metrics per socket (web100
- ipsec vs. freeswan
- measure overhead of qdisc, netfilter,IPSEC,...

Testing needs

- regression test (weekly)
  - TAHI – IPV6 validation suite
  - LTP
- more realistic network loads
- over capacity tests
  - DoS, Spam flood, ...

Testing infrastructure

- fault injection
  - packet loss
  - memory allocation failures
- out of kernel testbed?

2.7 ideas

- Skb scatter/gather DMA list
  - Driver API change
- IPSEC
  - ICMP matching
  - async and crypto hw support
- MPLS
- Unified flow cache??
- Heading away from route cache?