Notes From 10 May 2006 Meeting of the Network Stack Cabal

11 May 2006

Robert N. M. Watson

Security Research
Computer Laboratory
University of Cambridge

UNIVERSITY OF CAMBRIDGE

Introduction

- Assumed this slot would be to report on results
  - Long boring network session involving only network people
  - Instead everyone had to listen
- Summary of discussion and conclusions

IPSEC/IPv6/KAME

- We now own the netinet6 code
  - It contains significant quantities of improperly or under-locked code
  - We should do what we need to do to make it FreeBSD
- gnn is working on IPv6 support for FAST_IPSEC, should be mergeable to 6.x
  - Suggests future is FAST_IPSEC

netisr.pcbinfolock

- Problem
  - Want parallelism in in-bound network stack
- Challenge
  - Lock granularity: no potential parallelism in in-bound path
- Solution
  - sockref and pcb referencing will allow parallel processing of connections in different threads
- Owners: rwatson, mohans

netisr.parallelism

- Problem
  - Want parallelism in in-bound network stack
- Challenge
  - Need multiple threads to perform work in parallel, and must assign work to threads
- Solution
  - Ncpu netisrs, assign work by source
  - Pin ifnets to netisrs, possibly by CPU
- Owners: rwatson

netisr.interrupts

- Problem
  - Want to maximize cache locality for processing in-bound packets, avoid bad scheduler behavior
- Challenge
  - Scheduler has no idea
- Solution
  - pin interrupts, ithreads to processors
  - Requires manual configuration
- Owners: jhb

netisr.wakeup

- Problem
  - For loopback traffic, netisr preempts sending thread resulting in performance regression
- Challenge
  - Splx unwinding in 4.x and non-preemption in 5.x avoided premature/excess context switching
- Solution
  - Explore deferred wakeup, possibly queued
  - Also good for amortization of wakeups, contention
  - Owner: jhb, rwatson

mbuf.leak

- Problem
  - Reports of significant mbuf cluster leaks on 7.x
- Challenge
  - Mbuma suffers from a number of bugs relating to secondary zones and cache size; complicated by lock order between zone locks
- Solution
  - Impose resource limit on secondary Packet zone to prevent over-caching of Packets.
- Owner: andre, rwatson

mbuf.moreclustersizes

- Problem
  - SCTP wastes moderate quantities of persisting memory because of poor match between allocation requirements and available sizes
- Challenge
  - Balance overhead/complexity of variable sizes
- Solution
  - Experiment with local mbuma zones in SCTP to measure real-world performance
  - Owner: rrs

Virtualization

- Extensive discussion of varying techniques
  - Andre presents proposal for multi-IP jail with vifs
  - Andre presents routing table virtualization
  - Marko presents full network stack virtualization
- Possible concensus
  - Room for a variety of approaches
  - Concerns about multi-IP jail approach
  - Open questions about full virtualization and klds
- Owners: andre, zec

Link Layer Rationalization

- Problem
  - Ethernet code split over modules, undesirable administrative side effects due to separation
- Solution
  - Combine LLC and VLAN support into if_ethersubr.c
- Owners: rwatson

ip6fw

- Problem
  - 4 firewalls
- Challenge
  - IPv6 support in ipfw
- Solution
  - Brooks
- Owner
  - Brooks