

Pushing the FIB limits, yada, yada

John Scudder

jgs@juniper.net

NANOG 39, February 5, 2007

Problem Space

- **FIB must be**
 - Fast (lookup per packet)
 - Big (all your routes go there, DFZ table up-and-right)
- **Different ways to be fast**
 - Just go fast (exceedingly kewl silicon)
 - Parallelism (go less fast, but in parallel)
 - Computing industry is choosing parallelism
- **Just one way to be big: lots of memory**
 - SRAM is exotic, expensive, and low-density
 - DRAM (many flavors) is commodity, denser, tends to follow Moore's Law

Commodity Memory (xDRAM) for FIBs

■ Pros

- Cost
- Memory density
- Historically, scales like Moore's Law

■ Cons

- Slower than SRAM

■ **Speed limitations absorbed using parallelism, cunning search algorithms**

Current State of the Art

- **Shipping routers with RLDRAM FIBs**
 - M120, MX960, other vendors
 - Millions of entries in FIB (~2M)
- **Actual number of routes depends on**
 - size of routes (e.g., IPv6 is bigger than IPv4)
 - other demands on memory (e.g., filtering rules, uRPF, policers, etc)

FIB Scaling Expectations — Near Future

- **xDRAM density continues to increase**
 - Moore's Law
- **Current forwarding ASICs capable of addressing much larger memories**
 - As larger parts become available
- **Reasonable to expect (IPv4) FIBs ~10M within a few years if demand exists**
 - With current shipping architectures – no new R&D
- **SRAM, TCAMs still useful**
 - But relegated to uses with less scary scaling properties (e.g., caches)

Other Challenges

- **Packet rate**
- **Features (packet inspection, etc)**

- **Orthogonal to FIB size**
- **Some of these features do use TCAMs, SRAM**

Routing/Addressing Approaches

“Any problem in computer science can be solved with another layer of indirection.”

—David Wheeler

“But that usually will create another problem.”

—rest of the quote

Routing/Addressing Approaches [2]

- **Wouldn't it be great if we didn't have to throw hardware at FIB?**
 - Sure! I'll take two magic bullets please, in red!
- **TANSTAAFL (There Ain't No Such Thing As A Free Lunch)**
 - State in DFZ is (mostly) there for a reason =>
 - State will need to exist in some form in any system that provides as much functionality as present system!
 - ...unless we are willing to throw away some functionality
 - Magic bullets usually have feet of clay (and mixed metaphors)
- **Absolutely worth investigating... but don't bet the farm**
 - Routing/addressing research could bear fruit for something other than raw scaling, e.g. better operational characteristics
 - Long-term effort, so good thing we have a hardware solution medium-term.

Tunnel-Based Approaches

■ Promising line of research

- EID/RLOC split, etc — various proposals
- Some proposals use control processor for small fraction of traffic
 - This should worry you. Relying on caches should too.

■ BGP-free core

- Protects core routers from FIB growth
- Limits need for big-FIB deployment to edge
- No additional load on forwarding or control
- Deployed, works today

Good News

- **We can throw hardware at FIB scaling for at least the next decade or so, with existing technology**
 - Several big-FIB boxes shipping now
- **This provides time to research routing/addressing architectures**
 - Really don't want to build Internet on a R/A architecture that was hacked up quick under deadline pressure
- **BGP-free core can protect core (“P”) router FIBs**
 - Deployed today

Juniper
your
Net™

