Router Scaling Trends

John Scudder
jgs@juniper.net
RIPE-54, May 10, 2007

Agenda

- Problem Statement
- Router Implementation Approaches
- Architectural Approaches

Problem Statement

Problem Statement

- **Fundamental: concern in Internet community about growth of Internet routing table.**
  - "Up and to the right"
  - We must have an answer now and in near future!
  - Focused, immediate concern => router implementation approach.

Problem Statement

- **Fundamental: concern in Internet community about growth of Internet routing table.**
  - "Up and to the right"
  - We must have an answer now and in near future!
  - Focused, immediate concern => router implementation approach.
- **Details: ongoing dialogue.**
  - Multihoming
  - Traffic engineering
  - Poor deployment practice
  - Complicated problem space => architectural approach.

Notable Scaling Attributes

Notable Scaling Attributes

- **Related to Internet routing table size**
  - FIB size
  - RIB size
  - RIB-FIB download speed
  - Interdomain convergence

Notable Scaling Attributes

- **Related to Internet routing table size**
  - FIB size
  - RIB size
  - RIB-FIB download speed
  - Interdomain convergence
- **Orthogonal to Internet routing table size**
  - Intradomain convergence
  - Forwarding speed
  - Port density
  - Power/heat

Review: What's a routing table? RIB vs. FIB

Review: What's a routing table? RIB vs. FIB

- **RIB (Routing Information Base)**
  - Various names – "BGP table", "Adj-RIB-In", "Loc-RIB", etc. In combination, these are the RIB.
  - Stores all routes/paths – large storage demands
  - Control plane only – just on control processor
  - Modest performance demands (compared to FIB)
  - Scales like general-purpose computers

Review: What's a routing table? RIB vs. FIB

- **RIB (Routing Information Base)**
  - Various names – "BGP table", "Adj-RIB-In", "Loc-RIB", etc. In combination, these are the RIB.
  - Stores all routes/paths – large storage demands
  - Control plane only – just on control processor
  - Modest performance demands (compared to FIB)
  - Scales like general-purpose computers
- **FIB (Forwarding Information Base)**
  - Stores only routes selected as "best" from RIB – more modest storage demands
  - Forwarding plane – all forwarding hardware must store
  - High performance demands – performance of FIB limits packet forwarding rate

FIB Problem Space

FIB Problem Space

- **FIB must be**
  - Fast (lookup per packet)
  - Big (all your routes go there, routing table up-and-right)

FIB Problem Space

- **FIB must be**
  - Fast (lookup per packet)
  - Big (all your routes go there, routing 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

FIB Problem Space

- **FIB must be**
  - Fast (lookup per packet)
  - Big (all your routes go there, routing 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, TCAM is exotic, expensive, and low-density
  - DRAM (many flavors) is commodity, denser, tends to follow Moore's Law

Commodity Memory (xDRAM) for FIBs

Commodity Memory (xDRAM) for FIBs

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

Commodity Memory (xDRAM) for FIBs

- **Pros**
  - Cost
  - Memory density
  - Historically, scales like Moore's Law
- **Cons**
  - Slower than SRAM

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**

Power/Heat

Power/Heat

- FIB memory represents small percentage of total forwarding budget (~10%)

Power/Heat

- FIB memory represents small percentage of total forwarding budget (~10%)
- RLDRAM much more power-efficient than SRAM

Power/Heat

- **FIB memory represents small percentage of total forwarding budget (~10%)**
- **RLDRAM much more power-efficient than SRAM**
- **Packet rate, features primary power/heat drivers**
  - Some cause for optimism from recent Intel, IBM process announcements

Current State of the Art

Current State of the Art

- **Shipping routers with RLDRAM FIBs**
  - M120, MX960, other vendors
  - Millions of entries in FIB (~2M)

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

FIB Scaling Expectations — Near Future

- **xDRAM density continues to increase**
  - Moore's Law

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

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

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 will be increasingly relegated to uses with less scary scaling properties (e.g., caches)

Other Tricks Available

Other Tricks Available

- **FIB compression**
  - Don’t bother installing redundant more-specifics in FIB
  - Behavior identical to non-compressed FIB
  - Aligns with arbitrary de-aggregation (as long as aggregate is also advertised)
  - Some vendors shipping

Other Challenges

Other Challenges

- **Control Plane**
  - RIB->FIB download rate
  - RIB size
  - Intradomain convergence
  - Interdomain convergence

Other Challenges

- **Control Plane**
  - RIB->FIB download rate
  - RIB size
  - Intradomain convergence
  - Interdomain convergence
- **Forwarding Plane**
  - Packet rate
  - Features (packet inspection, etc)
  - Port density
  - Orthogonal to FIB size
  - Some of these features do use TCAMs, SRAM

Routing/Addressing Approaches

Routing/Addressing Approaches

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

—David Wheeler

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]

Routing/Addressing Approaches [2]

- **Wouldn't it be great if we didn't have to throw hardware at FIB?**
  - Sure! But…

Routing/Addressing Approaches [2]

- **Wouldn't it be great if we didn't have to throw hardware at FIB?**
  - Sure! But…
- **TANSTAAFL (There Ain't No Such Thing As A Free Lunch)**
  - State in Internet routing table 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
  - If something is too good to be true… it probably isn't.

Routing/Addressing Approaches [2]

- **Wouldn't it be great if we didn't have to throw hardware at FIB?**
  - Sure! But…
- **TANSTAAFL (There Ain't No Such Thing As A Free Lunch)**
  - State in Internet routing table 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
  - If something is too good to be true… it probably isn't.
- **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

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.

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
  - Works today

Good News

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

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

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**
  - Works today

Juniper your Net™