Route Table Size and Its Implications for Routers: A Vendor Perspective

Ananda Rajagopal

Product Line Manager, Service Provider Solutions

March 26, 2007

Agenda

- About Foundry
- FIB limits: The Issue
- Inside The World Of A High End Router
  - State of technology available today and on the horizon
  - Organization of routes in the FIB
- Opportunities for Optimization
- “Ever Increasing Router Costs” … Really?
- Conclusion

© 2007 Foundry Networks, Inc.

About Foundry

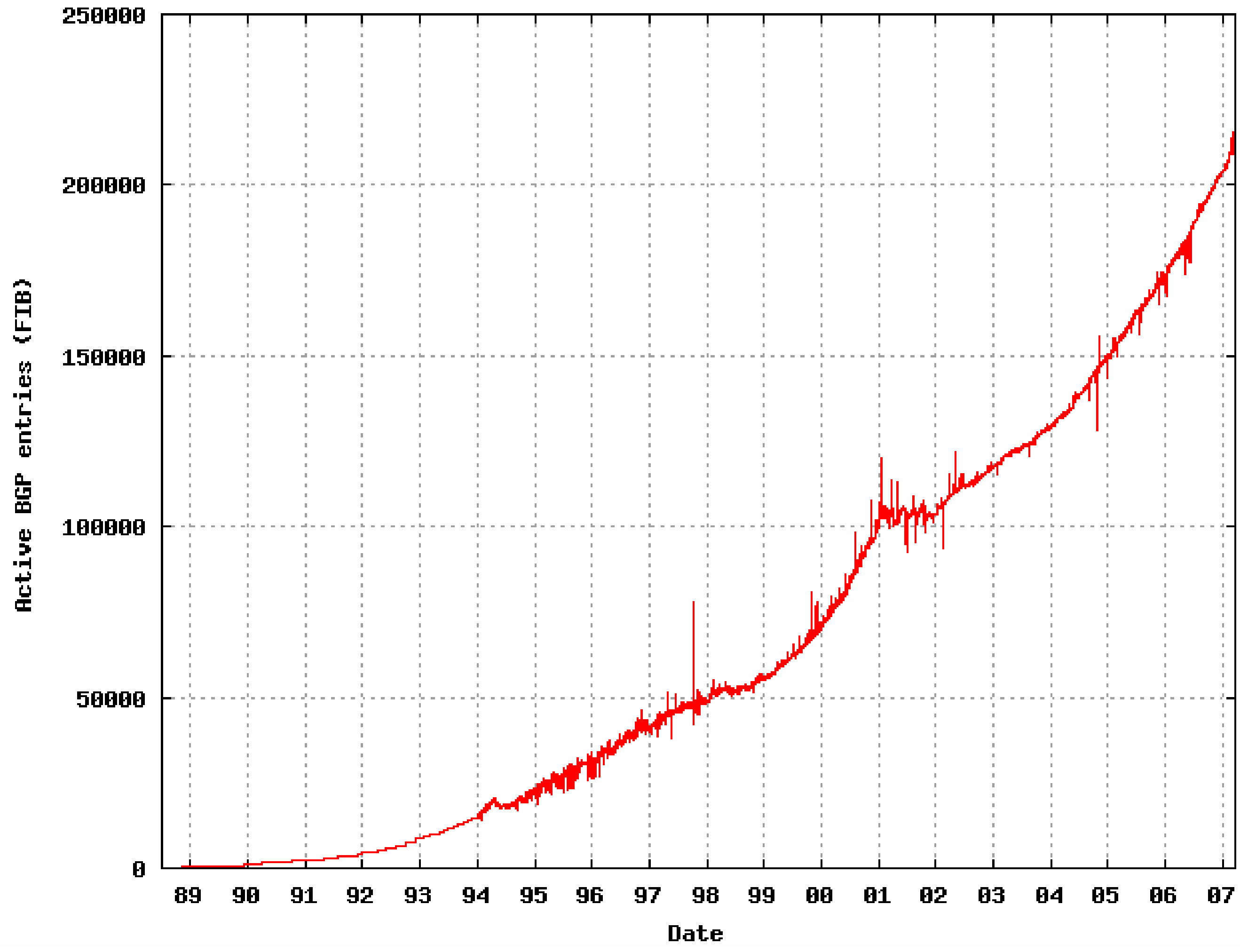

World Headquarters, Santa Clara, California


Manufacturing Operations, San Jose, California

- 11^{th} year of operations
- > 8 consecutive years of net profitability
- $886 million in cash with NO DEBT
  - Pretty useful for sponsoring events such as GPF ☺
- Over 10,400 customers worldwide
- Our routers are more popular today, though we also offer switches
  - NetIron XMR/MLX Series of Internet and metro routers now the top-revenue product for Foundry

© 2007 Foundry Networks, Inc.

FIB Limits: The Issue Today



- Total IPv4 routes @ ~213,000
- Growth rate @ ~17% for the last 4 years
- Or ~400-500 incremental routes added per week currently

Source: CIDR Report, March 23 2007

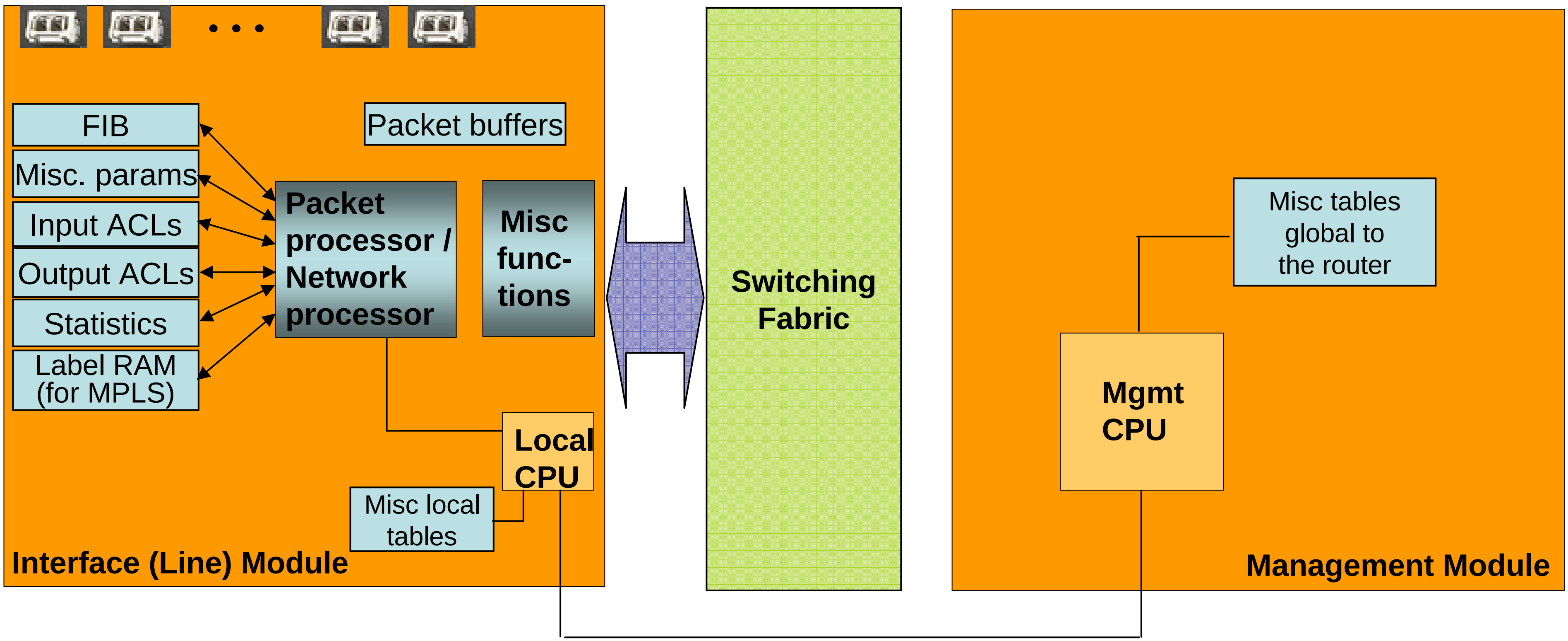
© 2007 Foundry Networks, Inc.

How Serious Is the Problem?

- It depends:
  - If you are using routers with known FIB limits that are being approached, then this is a serious problem that needs to be solved **NOW!**
  - If you are using routers that have headroom for growth
    - And your internal routes are significant, you should be justifiably worried
    - And your internal routes are “not many”, have a drink!
- Factors influencing growth:
  - Multi-homing
  - Excessive (indiscriminate?) de-aggregation
  - Traffic engineering policies
  - M&A
  - …

© 2007 Foundry Networks, Inc.

Inside The World Of A High-End Router



© 2007 Foundry Networks, Inc.

So, What is Used to Store the FIB (1)?

- Multiple options available:
  - CAM, SRAM, DRAM
  - CAM
    - The most deterministic way of doing a longest prefix match
    - The most expensive of the three options
    - Most dense device in mass production today is 18 Megabit CAMs
      - Assuming 1 IPv4 entry consumes 32 bits, 512k IPv4 entries can easily fit into the densest CAM available today
    - Multiple CAMs can be cascaded to give a larger logical capacity
    - Multiple CAMs may be used for parallel lookups if a single CAM cannot hold everything
    - 36 Megabit CAMs expected to hit mass production soon (2007)
      - Can easily pack 1M IPv4 entries in a single CAM
    - Conclusion: CAM technology exists today to easily scale beyond 1M IPv4 entries if the market demands it

FOUNDRY NETWORKS
© 2007 Foundry Networks, Inc.

So, What Is Used To Store The FIB (2)?

- Multiple options available (continued):
  - SRAM:
    - Overall retrieval time is longer because of multiple operations involved
    - Less expensive than CAMs but more than DRAMs
    - Less deterministic than CAMs but much better than DRAMs
    - A storage device; so lookup mechanism has to be implemented
    - 72 Megabit SRAMs available today in mass production
    - Innovations in SRAM continue to be made to bring more determinism into lookups
    - Conclusion: SRAM technology exists today to easily scale beyond 1M IPv4 entries if the market demands it
  - DRAM:
    - Depending on the type of DRAM used, price can vary
    - Can be non-deterministic when it comes to doing a lookup
    - The D in DRAM means dynamic; which means refreshes are needed periodically
    - Although feasible, less frequently used because of non-determinism
      - Difficult to scale to high density configurations as a result

© 2007 Foundry Networks, Inc.

Storing An Ever Increasing Number Of Routes

- Growing route table sizes would affect all of the following:
  - FIB used by the packet processor / network processor / classification logic
    - Stored using CAM or SRAM technology typically
  - FIB copy on a central management module / control module (or any other favorite name you may have)
    - Stored using DRAM typically
  - FIB copy on the local line module
    - Stored using DRAM typically
  - RIB table maintained on the module running the routing protocol
    - Stored using DRAM typically
    - Transient memory also required to handle the UPDATE messages received from peers

© 2007 Foundry Networks, Inc.

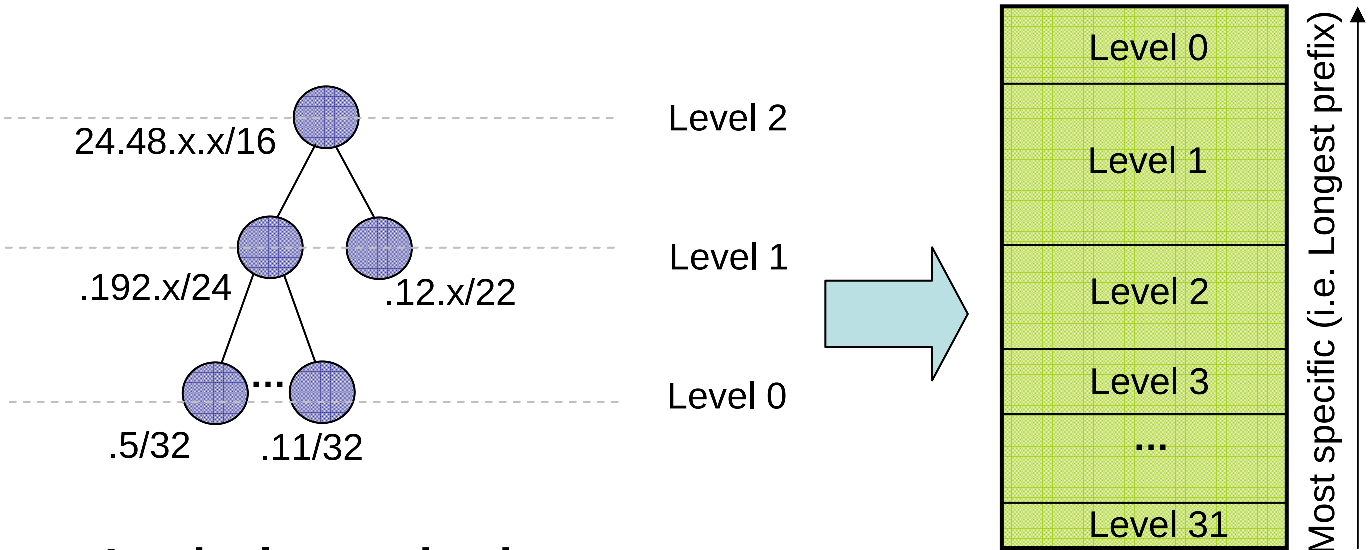
Organization Of Routes In The FIB (1)

- Beyond size of memory, an appropriate data structure needs to be chosen to organize the routes
  - For SRAM-based FIBs, this is mandatory because SRAM is just a storage device
  - For CAMs, the order of entries is important.
    - For a longest prefix match, the longer prefixes should show up in the earlier portions of the CAM
- Generally accepted today that a trie-approach to storing the FIB is the most optimal data structure for re**trie**val
  - Easy to quantify the worst-case lookup time (depends on key length)
  - Each node in a trie structure represents a portion of the prefix
  - 2 factors:
    - Stride: Number of bits inspected at a time
    - Depth of the tree: Path from the root to any node represents the prefix

© 2007 Foundry Networks, Inc.

Organization Of Routes In The FIB (2)

- Some types of tries:
  - M-level trie
    - E.g. a 3-level trie structure with strides of 16, 8, 8 means:
      - At first level, /16 is maintained (or total of 64k entries at this level)
      - At second level an additional /8 is maintained (or total of 256 entries per parent)
      - At third level, an additional /8 is maintained (or total of 256 entries per parent)
  - Dynamic prefix trie (DP-trie)
    - Number of levels may vary but greatly optimizes the amount of memory used



Logical organization

Physical organization

FOUNDRY NETWORKS
© 2007 Foundry Networks, Inc.

Organization Of Routes In The FIB (3)

- What is nice about the trie approach:
  - Typically, the first 4 levels capture a very high majority (> 99%) of routes
  - When doing a route update, route updates within the same level are trivial
    - Read that as if the length of the longest prefix for the new route is the same as the length of a longest prefix for an existing route, then you can just add to an existing level
  - Doing a route update that involves addition of a new level takes relatively more time but practically un-noticeable
  - As long as the size of each level is dynamic, it gives a lot of room for flexibility
- Innovations in new forms of organizing longest prefixes (e.g. new trie structures) affect the software only
- Could potentially speed updates that need to be made to the underlying storage device (CAM / SRAM etc)

FOUNDRY NETWORKS
© 2007 Foundry Networks, Inc.

Optimization Opportunities

- For any problem to be optimized, the constraints have to be well understood
  - Classical time vs. space tradeoff
- As prior slides show, technology is not the near-term constraint
  - Both CAM and SRAM technologies are well ahead of what the Internet table size demands
- However, an implementation could impose an artificial constraint
- Opportunities for optimization include:
  - RIB space for maintaining concurrent route views
  - Updates to RIB
  - Updates to FIB
  - Distribution of FIB
  - Other optimization opportunities
- Each of these will be considered in some detail

© 2007 Foundry Networks, Inc.

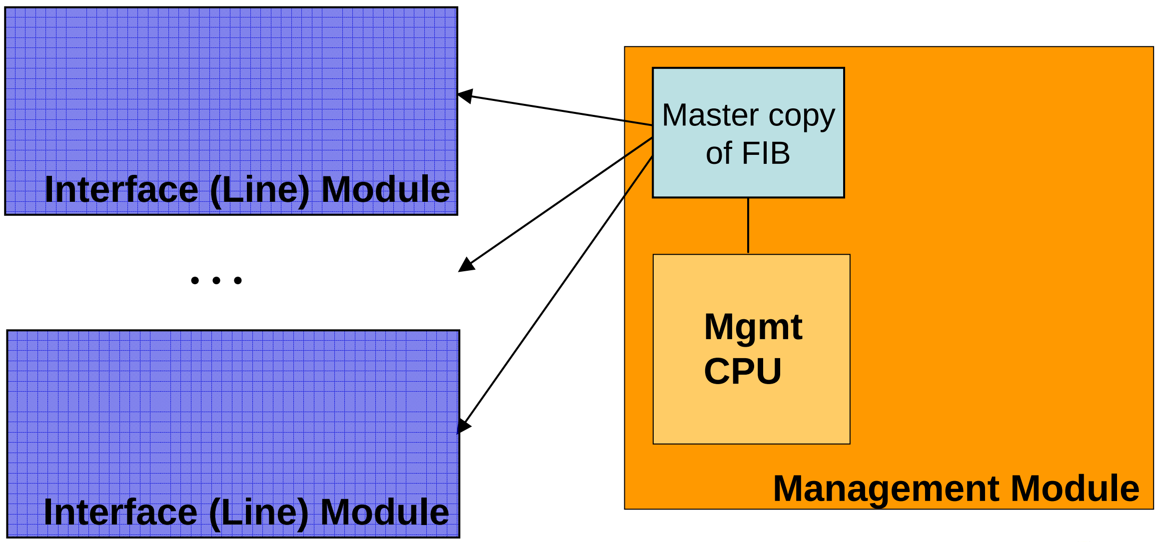
RIB Space And Updates To RIB

- Optimization of route map processing
- In terms of updates to RIB, the time it takes to do this is also dependent on amount of CPU power left
  - Overhead of the underlying operating system kernel
  - How distributed is the operating system?
  - Can line card CPUs sufficiently offload the typical processing done by the central “management” CPU?
  - Can hardware assists be used to offload repetitive CPU tasks?
- On the CPU technology side,
  - Dual core CPUs are now commercially available
  - CPU speeds in excess of 1.5 GHz for embedded applications now commercially available

© 2007 Foundry Networks, Inc.

Distribution Of FIB

- On multi-processor systems, the main (management) module maintains a central copy of the FIB
- Local copies kept on the line cards
- Distribution of main FIB copy to the interface modules needs to be efficient. Possible optimizations:
  - Have an internal high-speed network for distribution of copies
  - Compression of updates sent by main management module to the interface module



FOUNDRY NETWORKS
© 2007 Foundry Networks, Inc.

Other Optimization Possibilities

- Net aggregation
  - Attempt to aggregate those entries that have the same next-hop into a single “aggregate” entry
  - Apply a specific-length prefix to each aggregate entry
  - Net effect is to reverse the effect of de-aggregation (assuming the next hop is still the same)
  - Good for getting more mileage out of your existing product (if your product supports net aggregation)
  - May become more important in future if de-aggregation continues to increase
- Downsides of net aggregation:
  - Due to flapping and dynamic nature of Internet, what was aggregated may later no longer be valid causing the aggregate to be split again
  - FIB space may be conserved but trade-off is higher burden on CPU

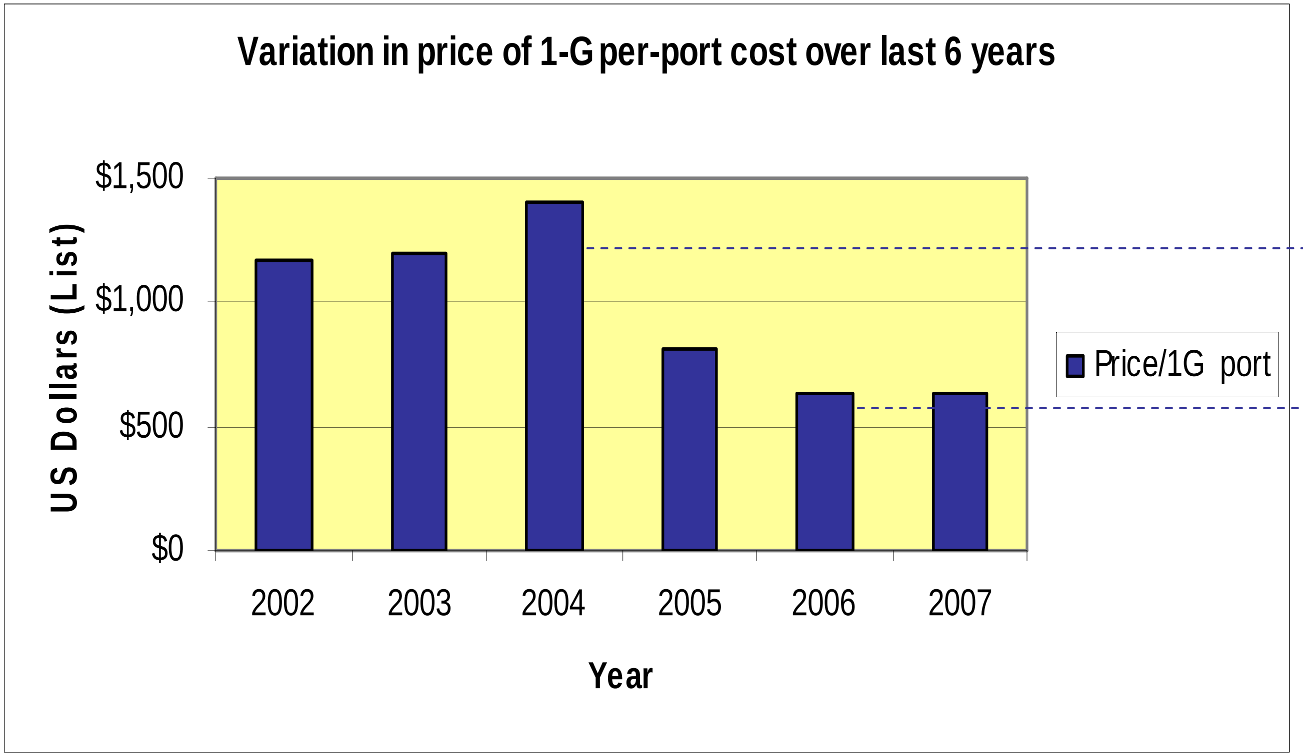
FOUNDRY NETWORKS
© 2007 Foundry Networks, Inc.

"Ever Increasing Router Costs" … Fact or Fiction?

- To determine this, 6 years of data was plotted for:
  - Price per Gigabit of routing
  - Price per 10 Gigabits of routing
  - Price per management module (since it is used to keep track of growing demands on RIB)
- While past performance is no guarantee for the future, history is a good guide to the future
- Method used:
  - If you had to pick the highest density router configuration available in the first quarter of each year from the same vendor, would there be any variation in price?
  - Fully burdened cost (excluding optics) used for comparison

© 2007 Foundry Networks, Inc.

Price / Gigabit Of A Router Over The Last 6 Years



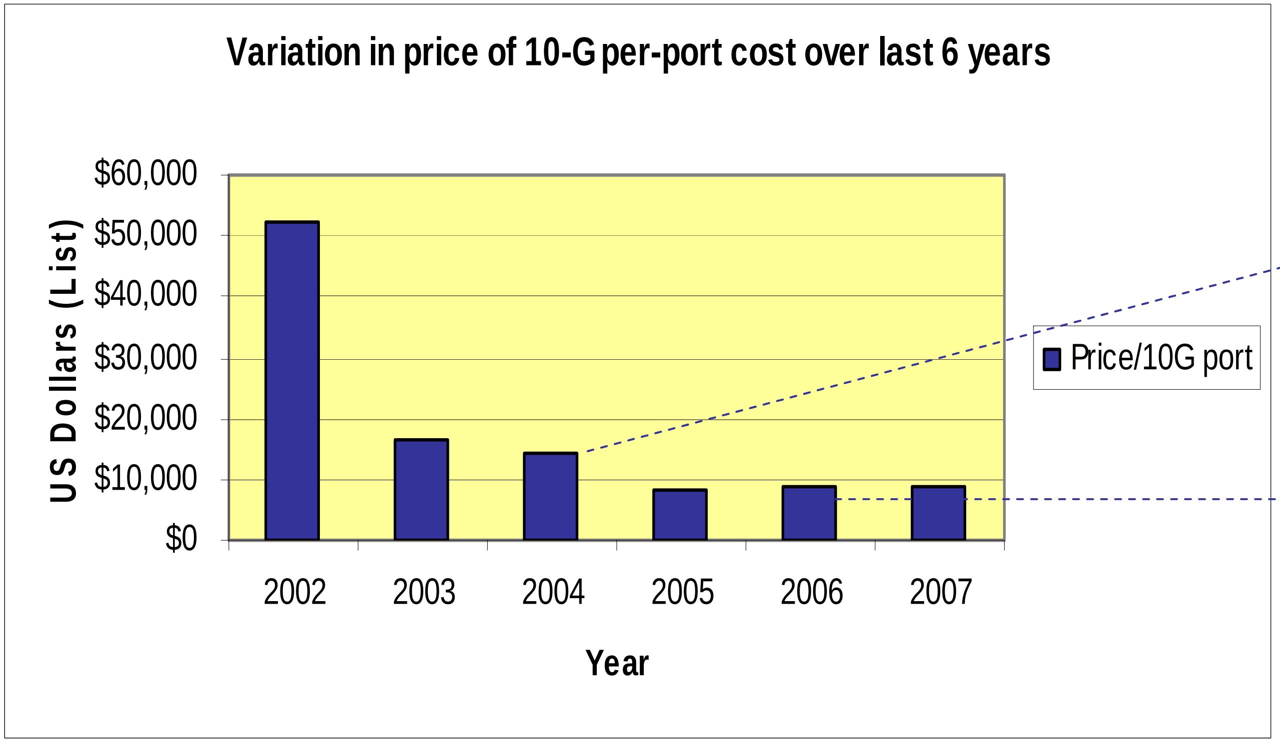
IPv6 capability added in hardware

IPv4/v6/MPLS in same hardware; 1M IPv4 FIB entries; more ACLs

- Graphed using the highest density product from a vendor available in Q1 of that year
- Fully loaded, redundant configuration assumed
- Optics not included
- Vendor data from Foundry used above

© 2007 Foundry Networks, Inc.

Price / 10 Gigabits Of A Router Over The Last 6 Years



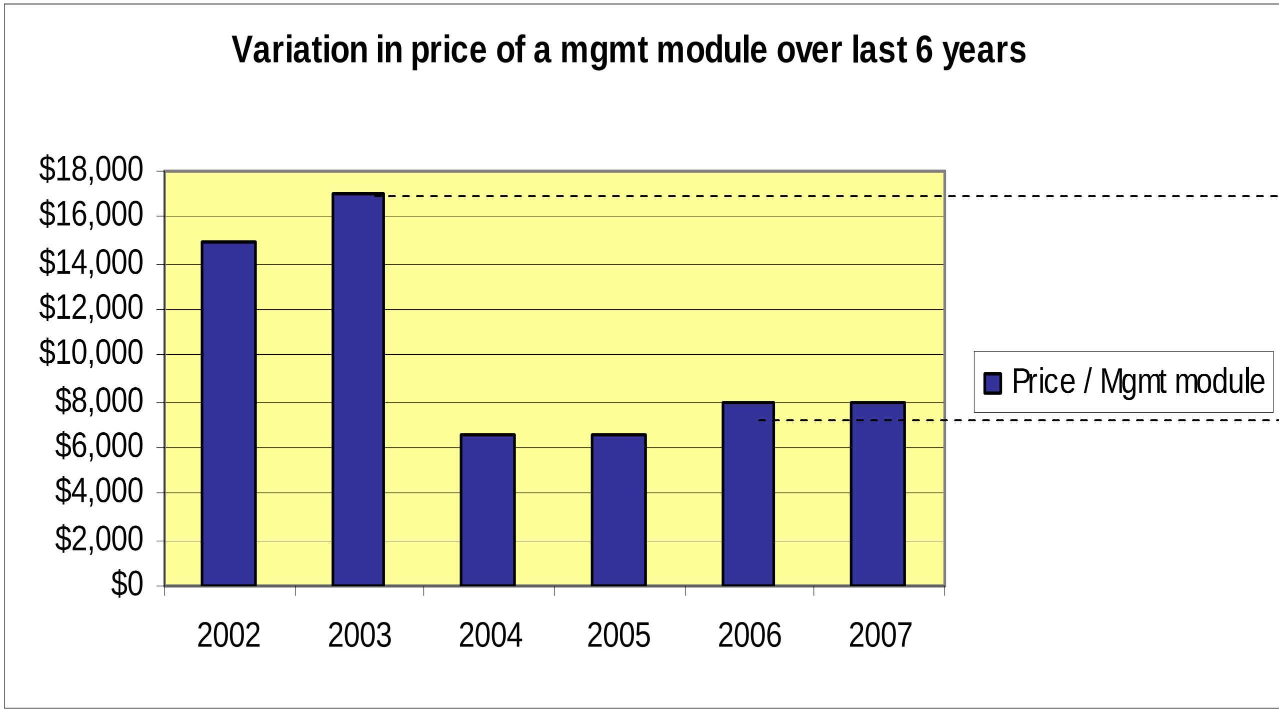
IPv6 capability added in hardware

IPv4/v6/MPLS in same hardware; 1M IPv4 FIB entries; more ACLs

- Graphed using the highest density product from a vendor available in Q1 of that year
- Fully loaded, redundant configuration assumed
- Optics not included
- Vendor data from Foundry used above

© 2007 Foundry Networks, Inc.

Price / Management Module Of A Router Over The Last 6 Years



With 512 MB DRAM, “x” MHz CPU

With 2 GB DRAM, “> 2x” MHz CPU

- Graphed using the highest density product from a vendor available in Q1 of that year
- Vendor data from Foundry used above

© 2007 Foundry Networks, Inc.

Conclusions

- Growth in route table size is a problem to be acknowledged and a problem that needs to be worked on separately
- BUT, technology is able to keep up with the problem and is not falling behind
- Choosing a router with anything less than 512k routes in FIB today would be unwise
  - Unless you are willing to live with partial route views and default routes
- The idea that routers will continue to become pricier to keep pace with the growth in the Internet routing tables is a fallacy
  - Cutting-edge router engineering can dramatically reduce cost
- Doomsday is not around as some people would have you believe
  - So, enjoy your cruise!

© 2007 Foundry Networks, Inc.

FOUNDRY® NETWORKS

Questions?

THE POWER OF PERFORMANCE™

FOUNDRY®
NETWORKS

Thank You!

THE POWER OF PERFORMANCE™