

Scaling issues with routing+multihoming

Vince Fuller, Cisco Systems

17 March 2007

Agenda

- **Look at the state growth of routing and addressing on the Internet**
- **Examine current trends and project how the future might look if nothing changes**
- **Explore an alternative approach that might better serve the Internet community**

Acknowledgements

This is not original work and credit is due:

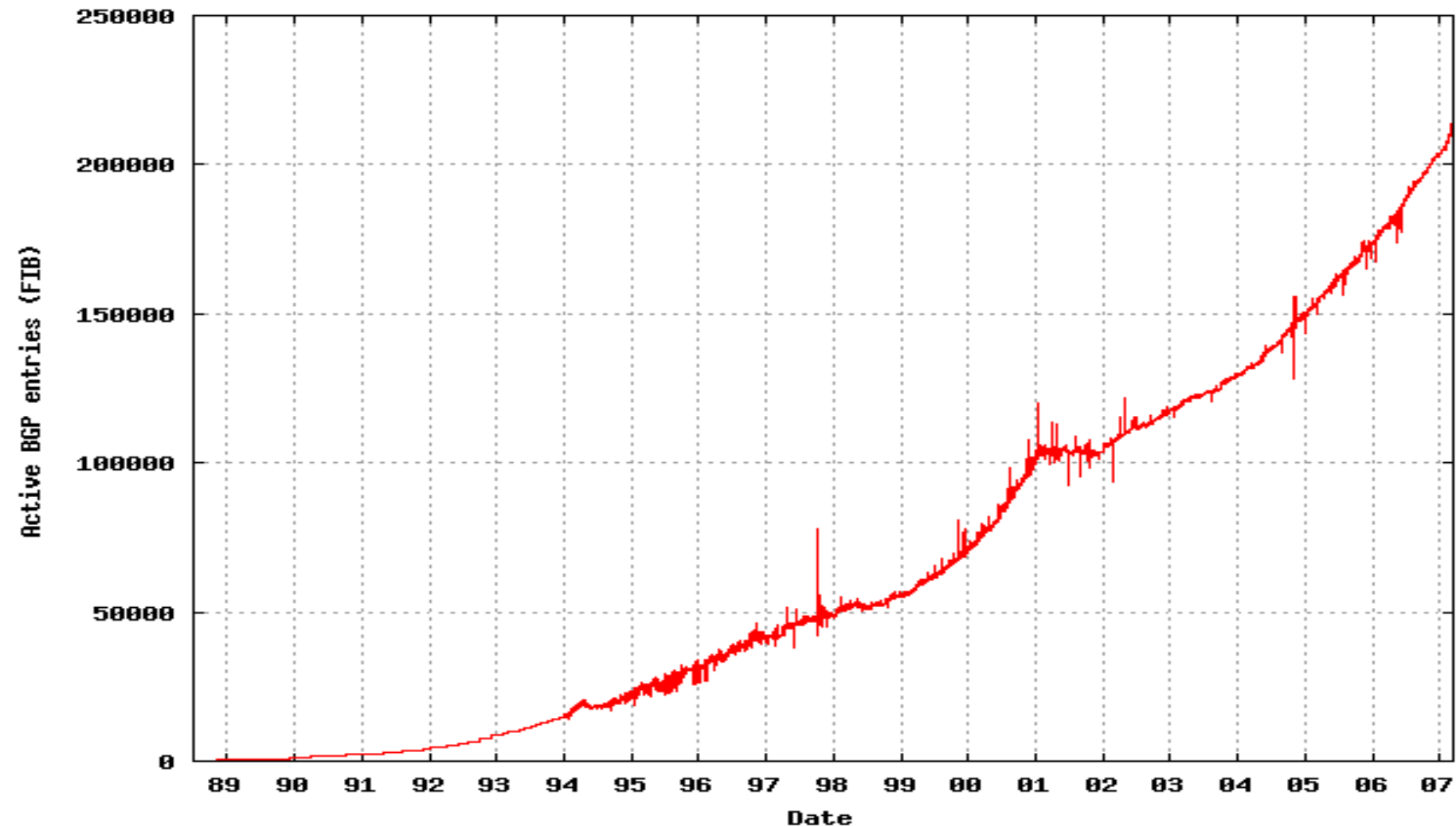
- **Noel Chiappa for his extensive writings over the years on ID/Locator separation**
- **Mike O'Dell for developing GSE/8+8**
- **Geoff Huston for his ongoing global routing system analysis work (CIDR report, BGP report, etc.)**
- **Jason Schiller and Sven Maduschke for the growth projection section (and Jason for tag-teaming to present this at NANOG)**
- **Tony Li for the information on hardware scaling**
- **Marshall Eubanks for finding and projecting the number of businesses (potential multi-homers) in the U.S. and the world**

Problem statement

- **There are reasons to believe that current trends in the growth of routing and addressing state on the global Internet may cause difficulty in the long term**
- **The Internet needs an easier, more scalable mechanism for multi-homing with traffic engineering**
- **An Internet-wide replacement of IPv4 with ipv6 represents a one-in-a-generation opportunity to either continue current trends or to deploy something truly innovative and sustainable**
- **As currently specified, routing and addressing with ipv6 is not significantly different than with IPv4 – it shares many of**

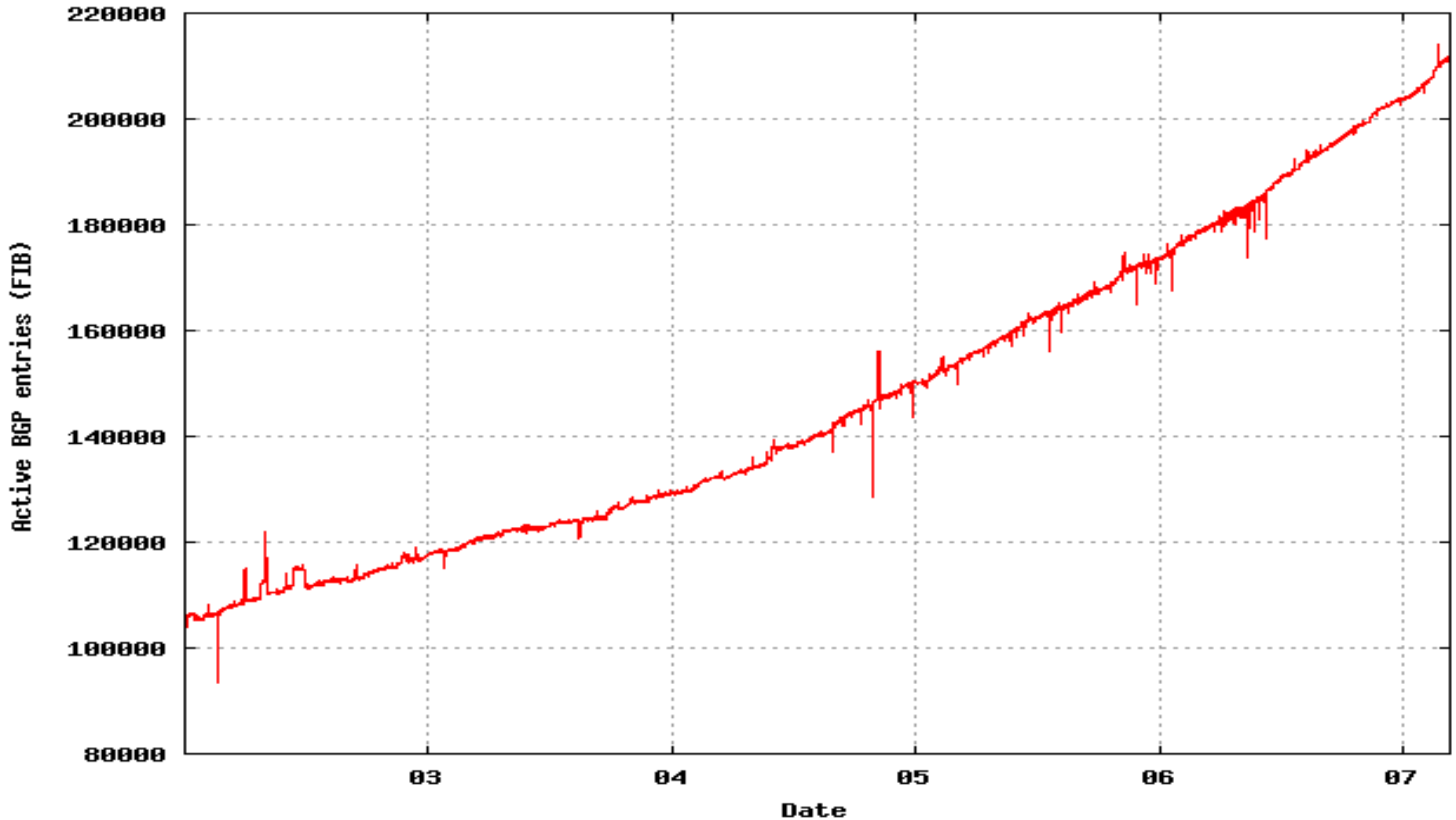
A view of routing state growth: 1988 to now

From bgp.potaroo.net/cidr/



A close-up of the post-bubble view

From bgp.potaroo.net/cidr/

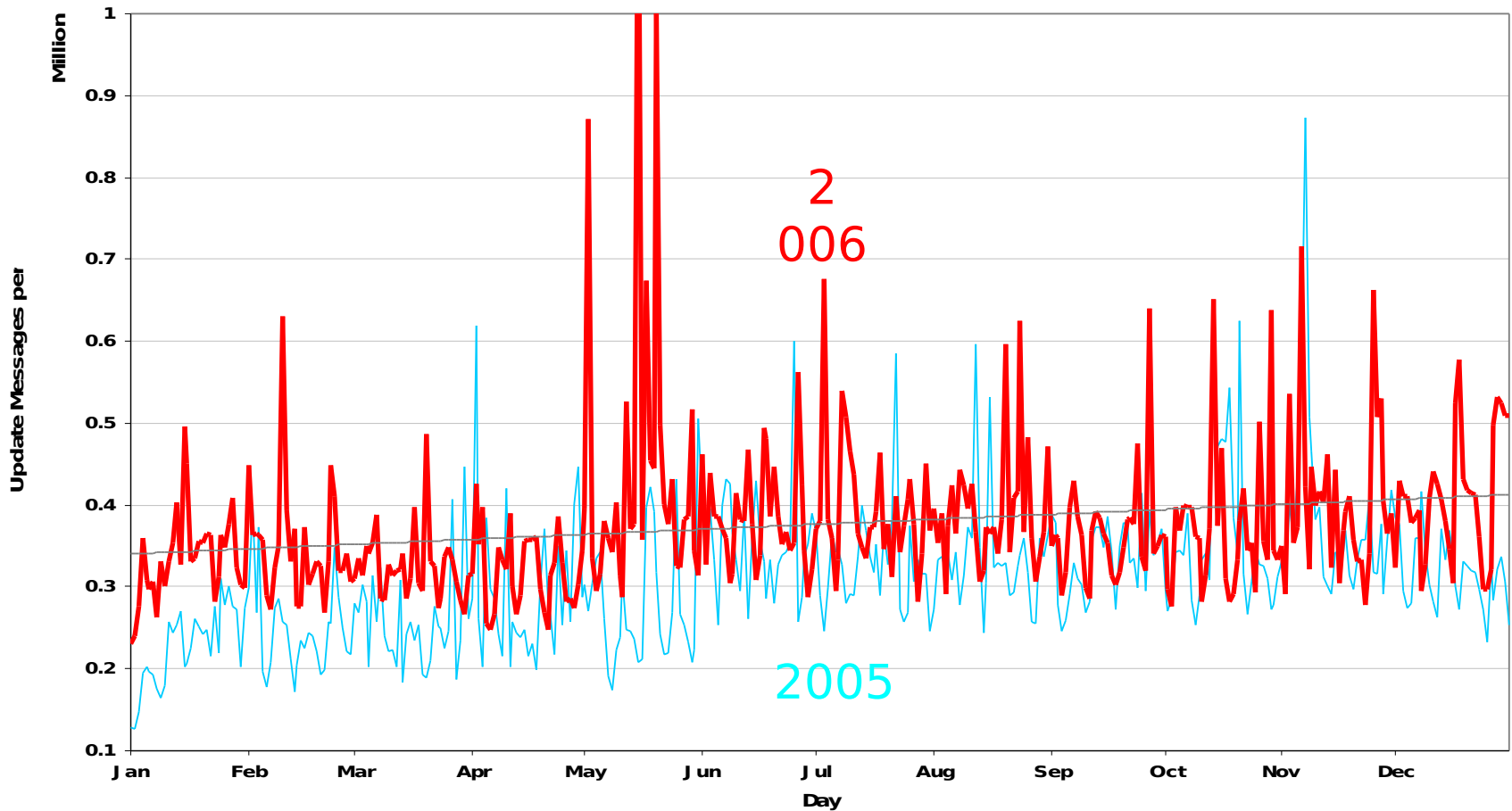


IPv4 Current/near-term view - Geoff's BGP report

- **How bad are the growth trends? Geoff's BGP reports show:**
 - **Prefixes: 130K to 170K (+30%) in 2005, to 204K in 2006 (+20%)**
 - **projected increase to approx 450K by March, 2012 (5 years)**
 - **global routes only – each SP has additional internal routes**
 - **Churn rate: 0.9M updates/day and 0.45M withdrawals/day now**
 - **projected increase to 2.25M/day and 2.65M/day by 2012**
 - **suggests need for much faster CPUs and much faster memories**
- **These are guesses based on a limited view of the routing system and on low-confidence projections (cloudy crystal ball); the truth could be worse, especially for peak demands**

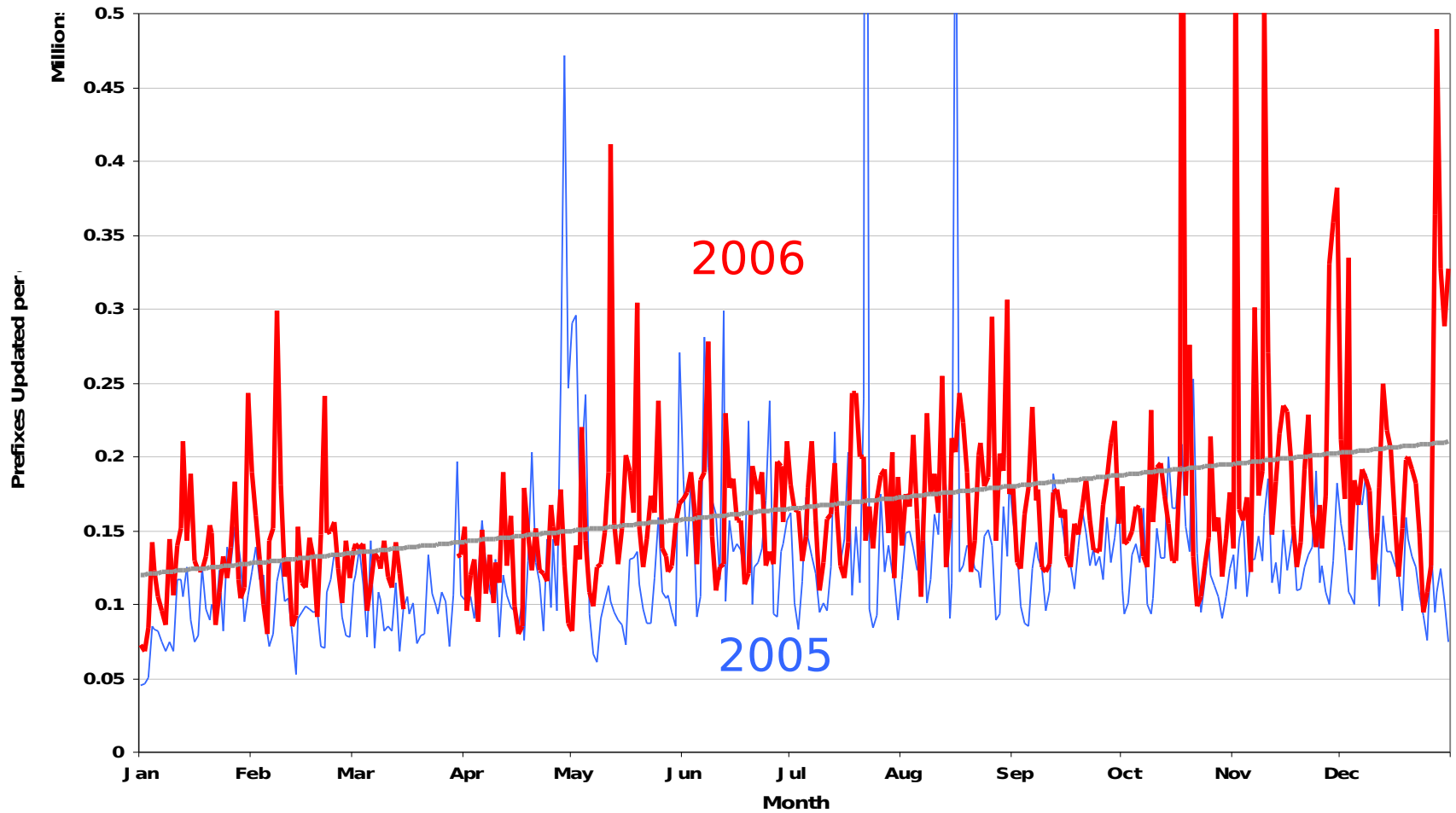
Update Message Rate

BGP Update Messages per Day



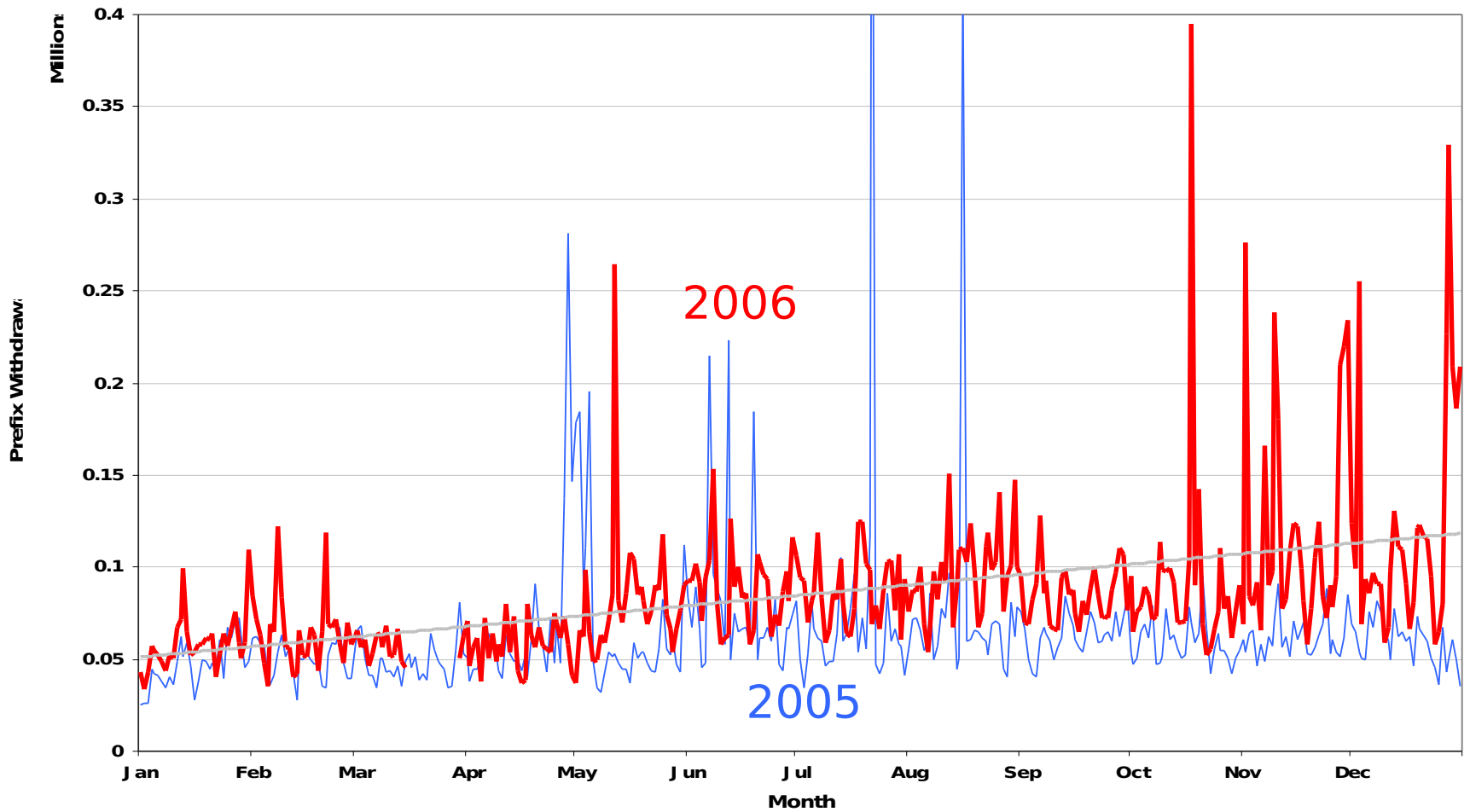
Prefix Update Rates

Prefix Updates per Day



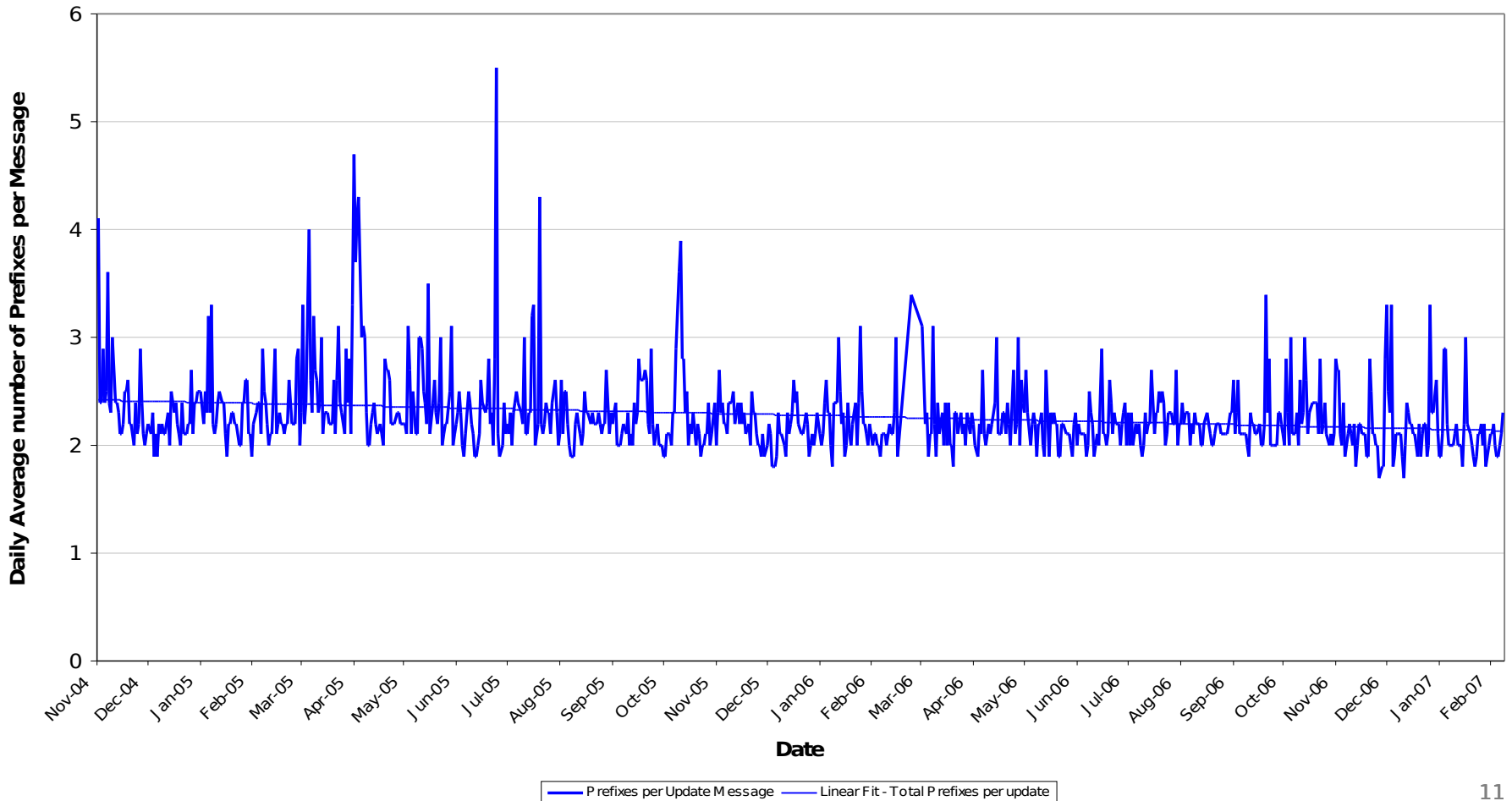
Withdrawal Rates

Prefix Withdrawals Per Day



Average Prefixes per BGP Update

Prefixes per BGP Update Message



Things are getting uglier... in many places

- **Philip Smith's NANOG-39 "lightening talk":**
<http://www.nanog.org/mtg-0702/presentations/smith-lightning>
- **Summary: de-aggregation is getting worse**
 - **De-aggregation factor: size of routing table/aggregated size**
- **For "original Internet", global de-agg factor is 1.85**
 - **North America: 1.69**
 - **EMEA: 1.53**
- **Faster-growing/developing regions are much higher:**
 - **Asia/Pacific: 2.48**
 - **Africa: 2.58**
 - **Latin/Caribbean: 3.40**
- **Trend may be additional pressure on table sizes, cause for concern**

What if we do nothing? Assume & project

- **Assume ipv6 widely deployed in parallel with IPv4**
 - **Need to carry global state for both indefinitely**
- **Multihoming trends continue unchanged (valid?)**
- **ipv6 does IPv4-like multihoming/traffic engineering**
 - **“PI” prefixes, no significant uptake of shim6**
- **Infer ipv6 table size from existing IPv4 deployment**
 - **One ipv6 prefix per ASN**
 - **One ipv6 more-specific per observed IPv4 more-specific**

Estimated IPv4+ipv6 Routing Table (Jason, 03/07)

Assume that everyone does dual-stack tomorrow...

**Current IPv4 Internet routing table: 211K
routes**

**New ipv6 routes (based on 1 prefix per AS): + 24K
routes**

**Intentional ipv6 de-aggregates for TE: + 74K
routes**

**These numbers exceed the FIB size of some deployed equipment
Combined global IP-routing table 309 K routes**

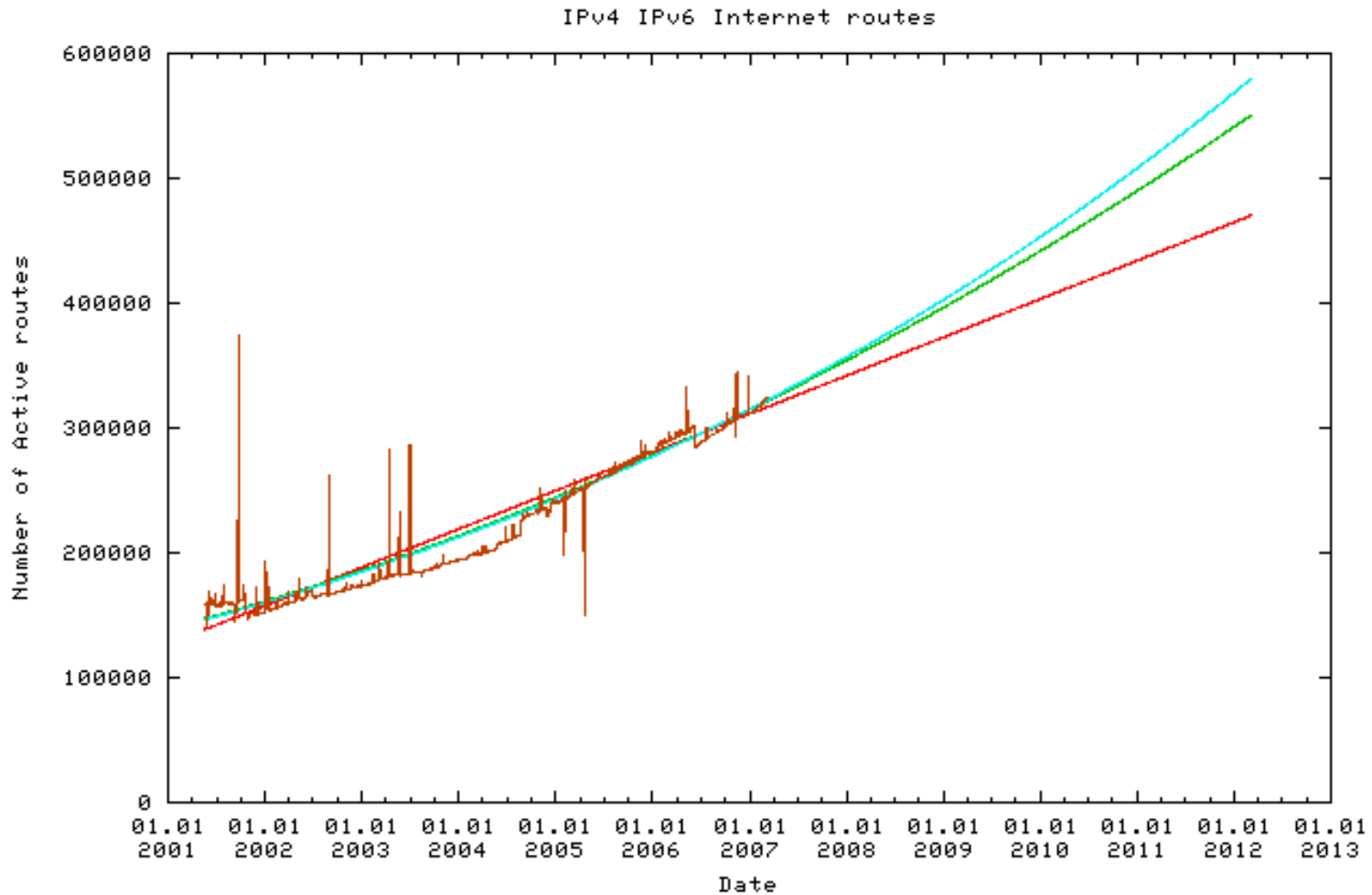
Of course, ipv6 will not be ubiquitous overnight

- but if/when it is, state growth will approach projections**

This is only looking at the global table

We'll consider the reality of "tier-1" routers next

Plot. projection of combined IPv4 + IPv6 global routing state (03/12/2007)



Legend			
projected linear	—	projected expo	—
projected poly	—	projected power	—
projected quad	—	IPv4 IPv6 Internet routes	—

“tier-1” internal routing estimate (3/07)

Current IPv4 Internet routing table:
211K routes

New ipv6 routes (based on 1 prefix per AS): +
24K routes

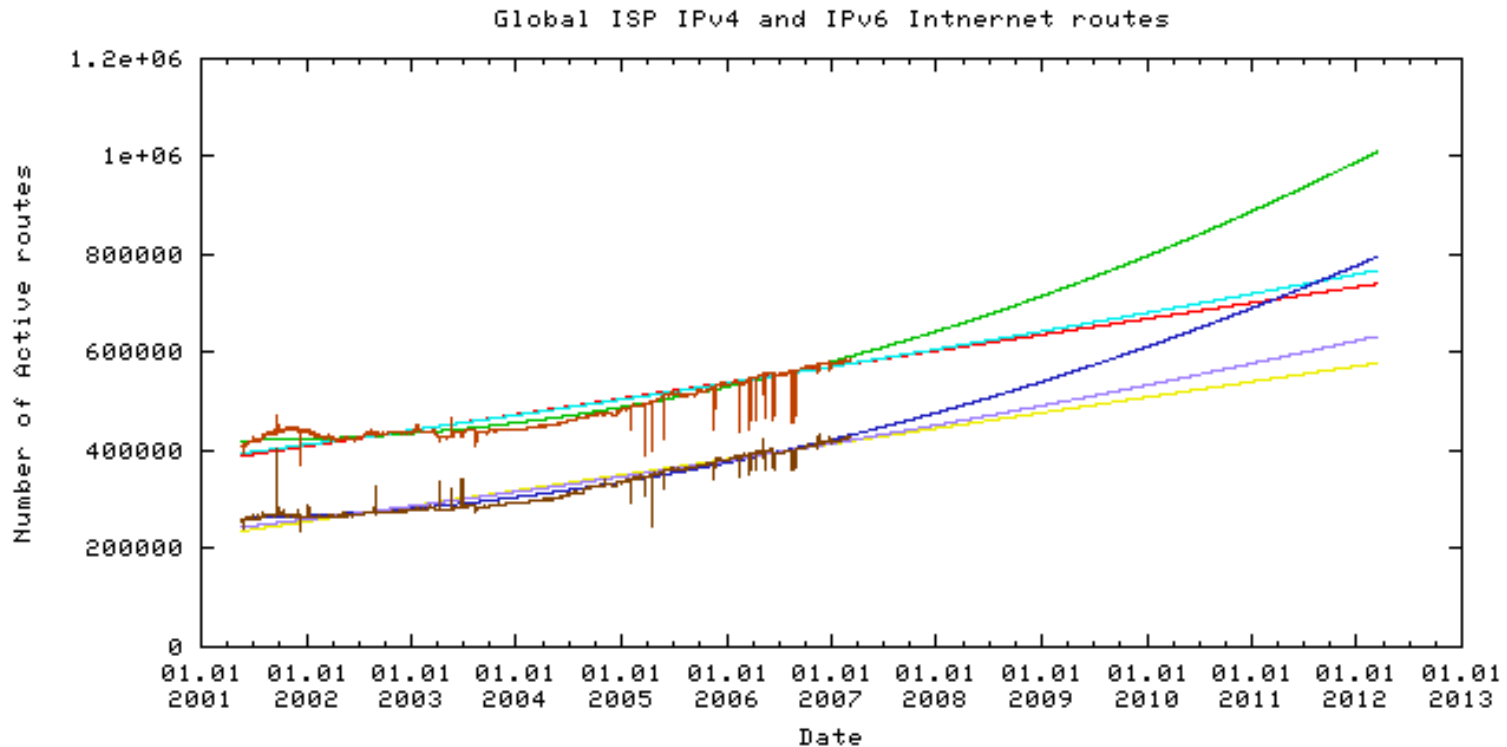
Intentional de-aggregates for IPv4-style TE: +
74K routes

Internal IPv4 customer de-aggregates + 50K to
150K routes

Internal ipv6 customer de-aggregates + 40K to
120K routes

These numbers exceed the FIB limits of a lot of
(projected from number of IPv4 customers)
currently-deployed equipment... and this
Total size of tier-1 ISP routing table **doesn't include routes used for VPNs/VRFS** 200K to
579K routes **(estimated at 200K to 500K for a large ISP**
today)

Plot: global routing state + "tier1" internals (03/12/2007)



Legend	
projected linear	—
projected poly	—
projected quad	—
projected expo	—
projected power	—
Global ISP IPv4 and IPv6 Internet routes	—
projected linear	—
projected poly	—
projected quad	—
projected expo	—
projected power	—
Global ISP IPv4 and IPv6 Internet routes	—

Plot: global routing state + "tier1" internals (03/12/2007)

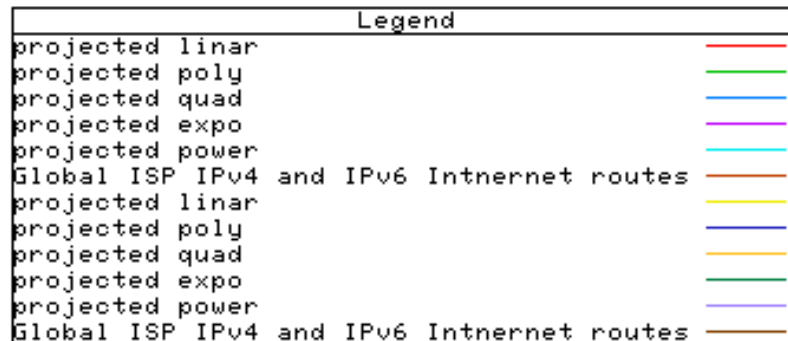
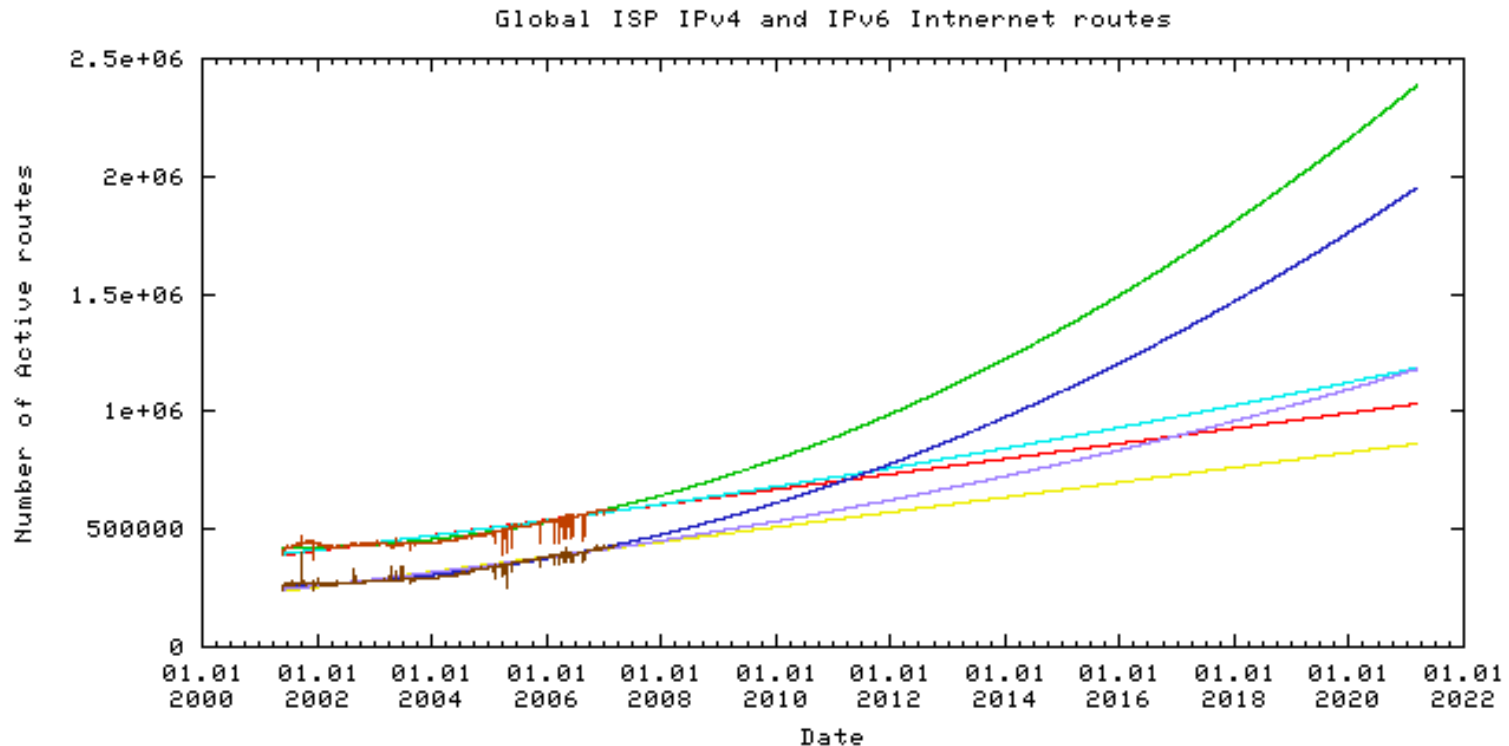


Table of big numbers (03/12/2007)

Route type	now	5yr (2012)	7yr (2014)	10y (2017)	14y (2021)
IPv4 Internet routes	220785	370943	447659	580706	794116
IPv4 CIDR Aggregates	119114				
IPv4 intentional de-aggregates	77811	131452	155889	195858	255259
Active Ases	25551	38003	42960	50401	60320
Projected IPv6 Internet routes	103362	163844	189853	230641	288857
Total IPv4/IPv6 Internet routes	324147	550502	662464	852459	1146608
Internal IPv4 low number	52944	73651	85822	108643	147554
Internal IPv4 high number	133465	185662	216343	273871	371960
Projected internal IPv6 (low)	50850	67486	78513	100184	138621
Projected internal IPv6 (high)	128186	170121	197918	252547	349441
Total IPv4/IPv6 routes (low)	427943	794607	995493	1357112	1951187
Total IPv4/IPv6 routes (high)	585799	1009246	1245412	1674695	2386401

Are these numbers insane?

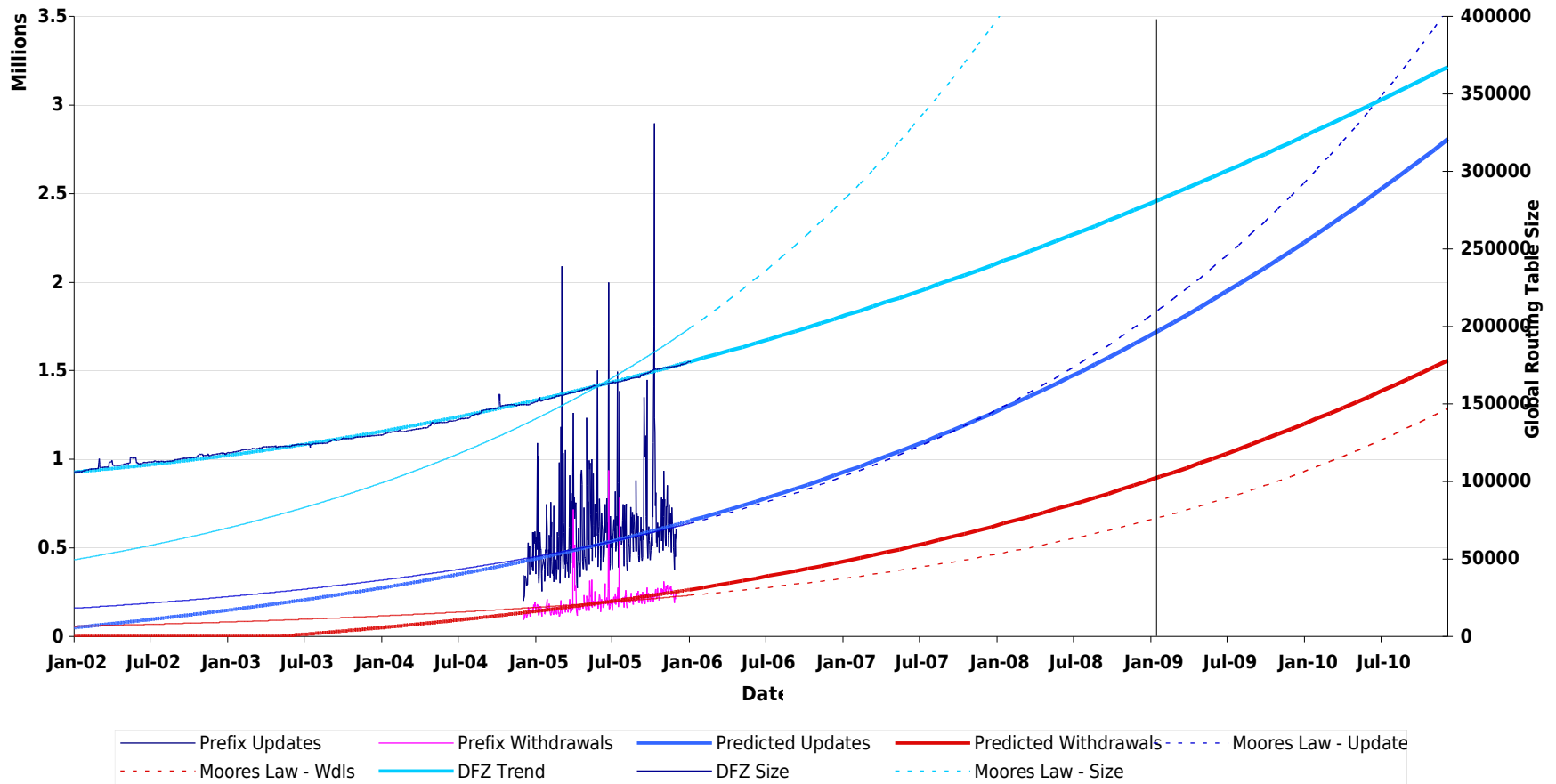
- **Marshall Eubanks did some analysis during discussion on the ARIN policy mailing list (PPML):**
- **How many multi-homed sites could there really be? Consider as an upper-bound the number of small-to-medium businesses worldwide**
- **1,237,198 U.S. companies with ≥ 10 employees**
 - (from http://www.sba.gov/advo/research/us_03ss.pdf)
- **U.S. is approximately 1/5 of global economy**
- **Suggests up to 6 million businesses that might want to multi-home someday... would be 6 million routes if multi-homing is done with “provider independent” address space**
- **Of course, this is just a WAG... and doesn't consider other factors that may or may not increase/decrease a demand for multi-homing (mobility? individuals' personal networks, ...?)**

Won't "Moore's Law" save us? Maybe

- **DRAM-based RIB/FIB should be able to ride growth curve, so raw size may not be a problem**
 - **Designers says no problem building 10M-entry RIB/FIB)**
 - **But with what tradeoffs? Power/chip space are real issues**
- **TCAM/SRAM are low-volume and have much lower growth rates; platforms that using those will have issues**
- **Forwarding ASICs already push limits of tech.**
- **"Moore's Law" tracks component density, not speed**
 - **Memory speeds improve at only about 10% per year**
- **BGP and RIB/FIB update rates are bounded by memory/CPU speeds and seem to be growing non-linearly; "meshiness" of topology is an issue**

Plot of growth trends vs. "Moore's Law"

Update and Withdrawal Rate Predictive



Current direction doesn't seem to be helping

- **Original ipv6 strict hierarchical assignments**
 - **Fails in the face of large numbers of multi-homed sites**
 - **RIRs already moving away**
- **“PI for all” – see the earlier growth projections**
- **“geographic/metro/exchange” – constrains topology, requires new regulatory regime**
 - *“Addressing can follow topology or topology can follow addressing; choose one” – Y. Rekhter*
- **Shim6 – maybe workable for SOHO but nobody (SPs, hosting providers, end-sites)**

So, why doesn't IP routing scale?

- **It's all about the schizophrenic nature of addresses**
 - they need to provide location information for routing
 - but also identify the endpoints for sessions
- **For routing to scale, locators need to be assigned according to topology and change as topology changes (*"Addressing can follow topology or topology can follow addressing; choose one"* – Y. Rekhter)**
- **But as identifiers, assignment is along organizational hierarchy and stability is needed – users and applications don't want renumbering when network attachment points change**
- **A single numbering space cannot serve both of these needs in a scalable way (see "further reading" section for a more in depth discussion of this)**
- **The really scary thing is that the scaling problem won't become obvious until (and if) ipv6 becomes widely-deployed**

Maybe we something other than “addresses”?

- **What if instead of addresses there were “endpoint identifiers” associated with sites and “locators” used by the routing system?**
 - **Identifiers are hierarchically assigned to sites along administrative lines (like DNS hostnames) and do not change on devices that remain associated with the site; think “provider-independent” numbering but not routable**
 - **Locators are assigned according to the network topology; think “provider-based” CIDR block address assignments**
 - **Locators are aggregated/abstracted at topological boundaries to keep routing state scalable**
 - **When site’s connection to network topology changes, so do the locators – aggregation is**

A new approach - continued

- **This is not a new idea – see the “additional reading” section for more discussion about the concepts of endpoint naming and topological locators**
- **October IAB-sponsored workshop found fairly good consensus among a group of ISPs, vendors, IESG, and IAB that the problem exists and needs to be solved... ID/LOC separation seems likely part of the solution**
- **More recent email list discussions suggest that we are far from good consensus (and ugly politics/egos in the IETF may be muddling things a bit)**

ID/LOC separation – a little bit of why and how

- **Common concepts:**
 - **Topologically-assigned locators (think “PA”)**
 - **Organizationally-assigned identifiers (think “PI”)**
- **Two different dimensions of approaches/trade-offs:**
 - **Host-based vs. network/router-based (which devices change?)**
 - **New name space vs. re-use/re-purpose of existing name space**
- **For more on this, see Dave Meyer’s APRICOT2007 presentation:**

<http://www.1-4->

[E.net/~dmm/talks/apricot2007/locid/](http://www.1-4-)

Again, this isn't new

- **Several past and present approaches:**
 - **8+8/GSE – ipv6 address format (split into two parts), router changes, limited host changes**
 - **shim6/HIP/SCTP – new name space, major host changes**
 - **LISP – IPv4/ipv6 address format (different roles for prefixes), no host changes, some router changes**
 - **NIMROD – new name space, new routing architecture, no host changes (maybe)**
 - **Some others being discussed on mailing lists... (see “what’s next” below)**
- **See “additional reading” section for full references**

Conclusions and recommendations

- **Projected growth trends of routing state may exceed the cost-effectiveness of hardware improvements.**
- **Vendors can and will build products to accommodate growth (routers buyable today can handle worst case we project here) but there will be costs and tradeoffs... and there may be pain for service providers (remember the 1990s?)**
- **An Internet-wide replacement of IPv4 with ipv6 represents a unique opportunity to either continue current trends or to pursue a new direction toward long-term scalability**
- **ipv6, as currently defined, doesn't help – its routing and addressing is much the same as IPv4, with similar properties and scaling characteristics**
- **Perhaps a new approach, based on identifier/locator separation, would be a better path**

What's next?

- **Is there a real problem here?**
- **Is the Internet operations community interested in looking at this problem and working on a solution? Where could/should the work be done?**
 - **Recent IAB workshop was good – problem recognized, www.ietf.org/internet-drafts/draft-iab-raws-report-00.txt**
 - **Follow-up discussions in IETF/IESG/IAB less encouraging**
 - **NANOG/RIPE/APRICOT? That's why we're here...**
 - **ITU? Vendors? Research community? Other forums?**
- **Current discussion occurring at:**
 - architecture-discuss@ietf.org**
 - ram@iab.org**
 - rrg@psg.com**
- **Stay tuned... more to come**

Recommended Reading - historic

“The Long and Winding ROAD”, a brief history of Internet routing and address evolution,
<http://rms46.vlsm.org/1/42.html>

“Endpoints and Endpoint names: A Proposed Enhancement to the Internet Architecture”, J. Noel Chiappa, 1999, <http://ana.lcs.mit.edu/~jnc//tech/endpoints.txt>

“On the Naming and Binding of Network Destinations”, J. Saltzer, August, 1993, published as RFC1498,
<http://www.ietf.org/rfc/rfc1498.txt?number=1498>

“The NIMROD Routing Architecture”, I. Castineyra, N. Chiappa, M. Steenstrup. February 2006, published as RFC1992,
<http://www.ietf.org/rfc/rfc1992.txt?number=1992>

“GSE - An Alternative Addressing Architecture for IPv6”, M. O’Dell.

Recommended Reading - recent work

- “Routing the Internet in 2006”, G. Huston,**
<http://www.potaroo.net/presentations/2007-03-18-routi>
- “Projecting Future IPv4 Router Requirements from Trends in Dynamic BGP Behavior”, G. Huston and G. Armitage,**
<http://www.potaroo.net/papers/phd/atnac-2006/bgp-atn>
- “Report from the IAB Workshop on Routing and Addressing”, Meyer, D., Zhang, L., and Fall, K. (editors),**
<http://www.ietf.org/internet-drafts/draft-iab-raws-repor>
- “Locator/ID Separation Protocol”, Farinacci, D., Fuller, V., and D. Oran,**
<http://www.ietf.org/internet-drafts/draft-farinacci-lisp-0>
- “The Locator/ID split, its implications for the IP Architecture, and a few current approaches”, D. Meyer,**
ARRICOT 2007