An Inconvenient Prefix: Is Routing Table Pollution Leading To Global Datacenter Warming?

Richard A Steenbergen <ras@nlayer.net> nLayer Communications

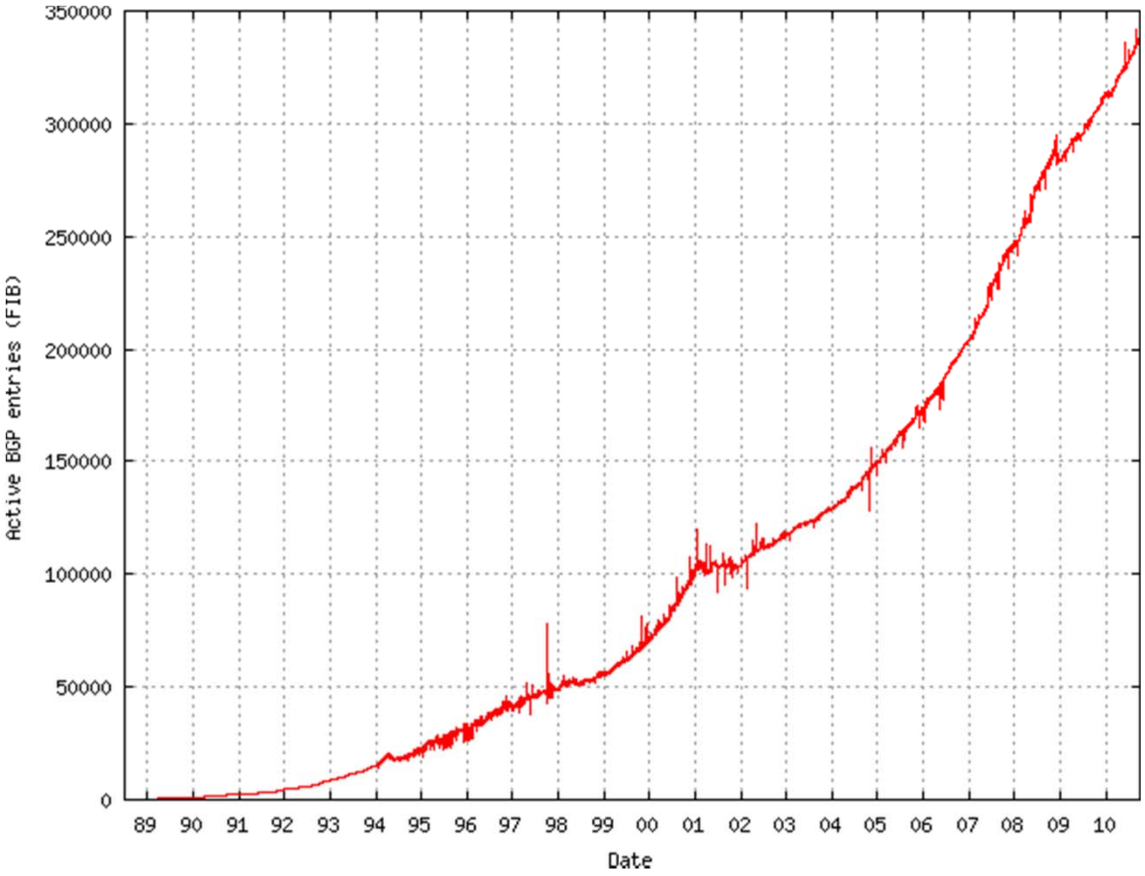
Rob Mosher <rmosher@he.net> Hurricane Electric

NANOG 50 – Atlanta GA

October 4 2010

Global Routing Table Size Over Time

- Oh My God! It’s up and to the right! We’re all going to die!!!
- Look at that curve! It looks exponential! The Internet is doomed!

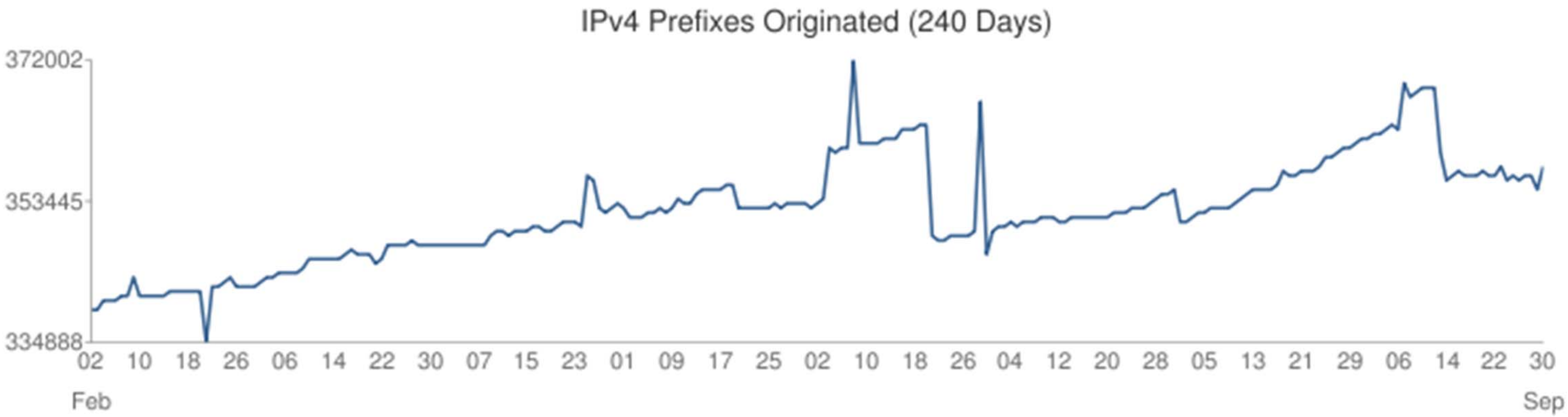


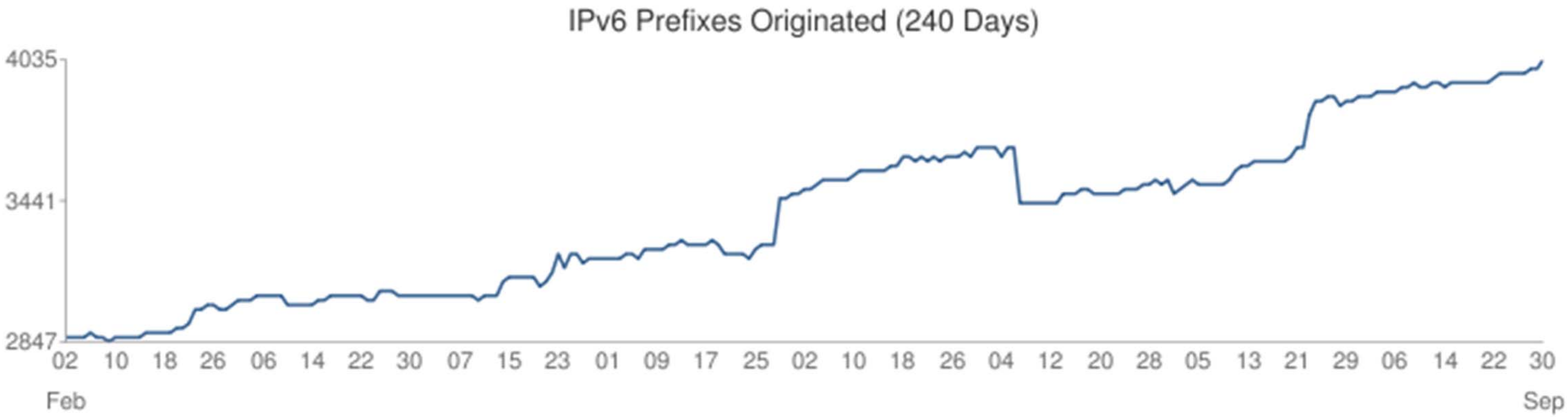
- Just kidding. Sorry, had to get that out of the way up front.

Why Does Routing Table Size Matter?

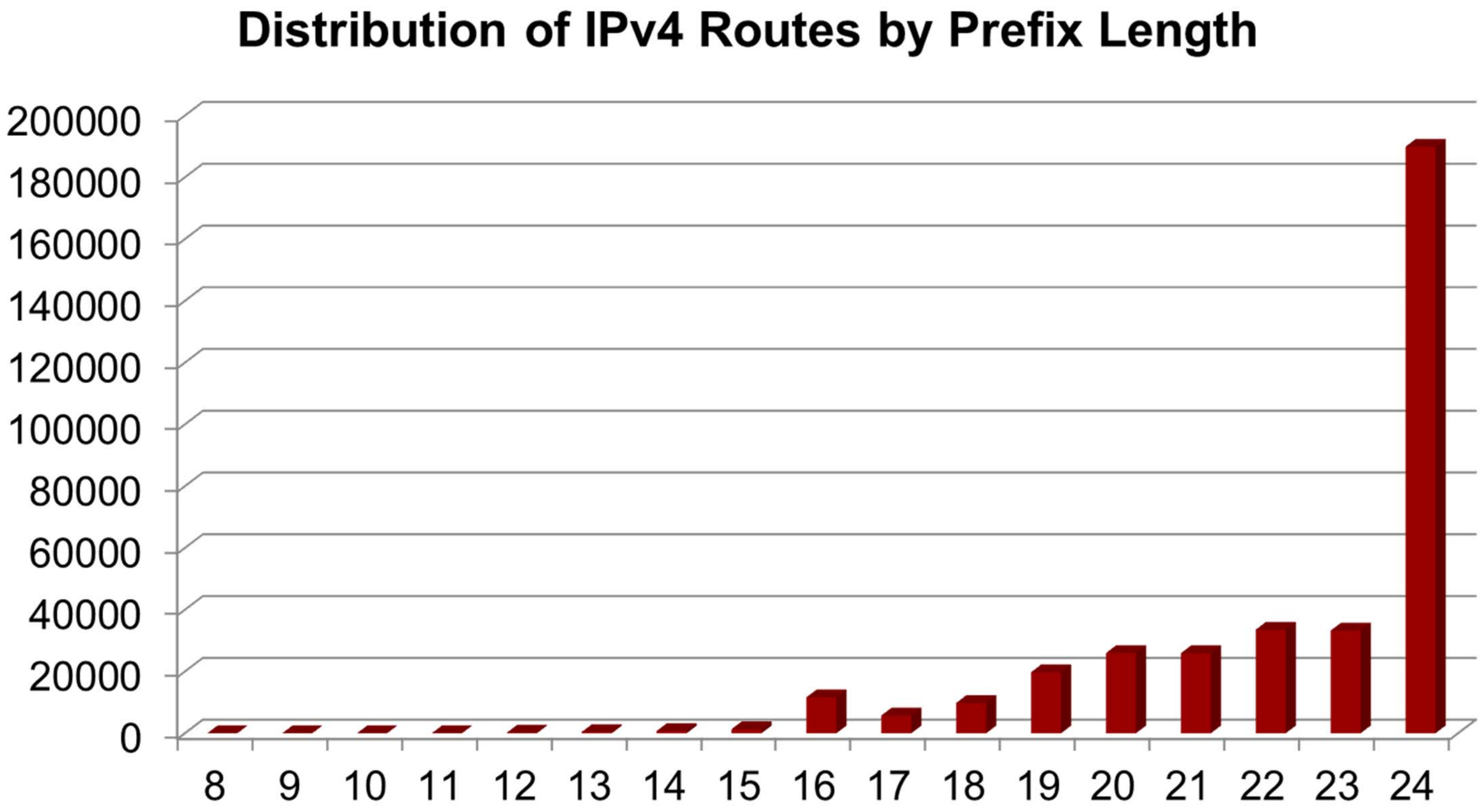
- Because everything you announce into the global table is heard by every other BGP speaking router on the planet.
- Larger routing tables use more RAM, FIB space, and CPU.
  - And it’s not just about “does the most common low end router have enough RAM and FIB to hold a full table”.
  - Most of the Internet is multi-homed at some level, so networks with extensive peering will easily see millions of possible BGP paths.
  - Networks with many POPs will see large numbers of routes in their IBGP core, slowing convergence after a BGP flap or router reload.
  - Even top of the line core routers with the maximum amount of CPU and RAM available for purchase today are becoming stressed.
- And more routes means more potential for BGP churn.
  - Further increasing CPU use and degrading performance.

Global Routing Table Size Over 240 Days





So Where Are All These Routes?



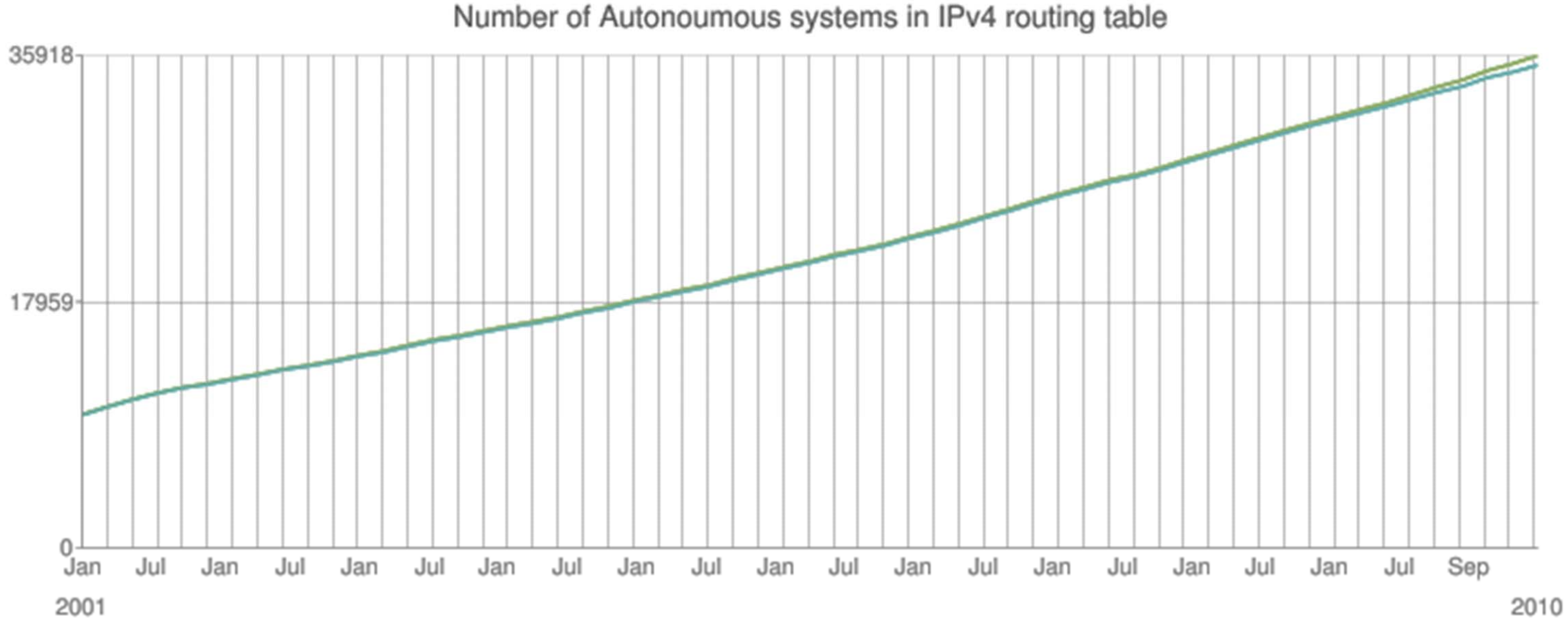
Drivers Behind Routing Table Growth

Theories Behind Routing Table Growth

- What is behind the ever-increasing size of the routing table?
- Many theories have been suggested.
- But let’s examine the 4 most common:
  - “More networks are multi-homing, putting more routes into BGP”.
  - “Slow growth allocation methods cause fragmentation”.
  - “It’s all being done for traffic engineering purposes”.
  - “Large numbers of networks are redistributing routes into BGP”.
  - “People are just being stupid with their configurations”.

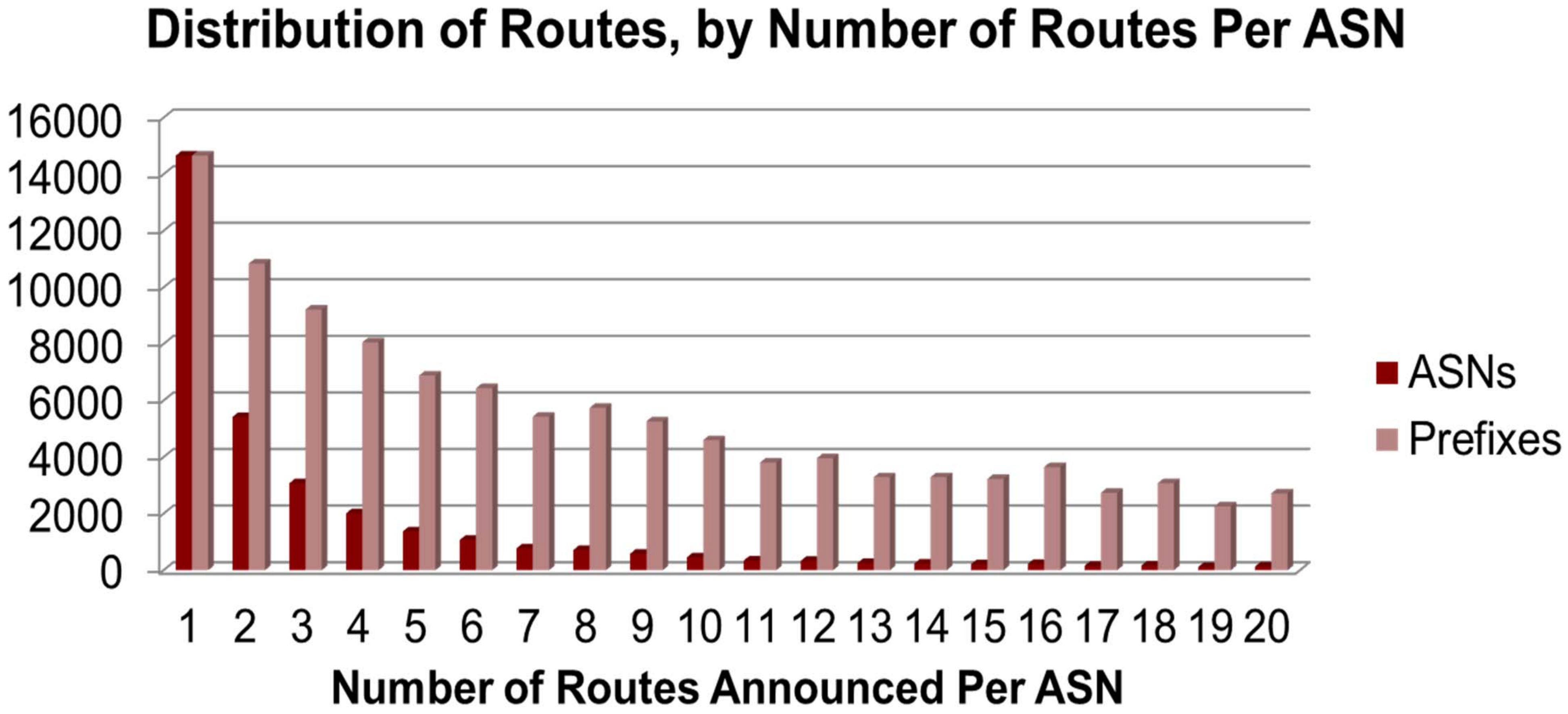
Theory: More Networks Are Multihoming

- True. But there are still only around 35K active ASNs, or around 1/10th the number of routes in the global table.
- Growth is also very linear.



Distribution of Routes by ASN Size

- Small ASNs (under 20 routes each) are:
  - 86.5% of the total active ASNs (those which announce any routes)
  - But less than 33% of the routes in the global routing table.



Theory: Slow Growth Causes Fragmentation

- True. If not for fragmentation, every ASN would need only one route, and the routing table would only be ~35K.
- Remember, this occurs at multiple levels:
  - An ISP gets slow growth allocations from a RIR.
  - The ISP's customer gets slow growth allocations from the ISP.
  - Their customers may get slow growth allocations from them…
- And not every network manages long term growth well.
  - Large, smart, efficient networks with proper documentation and a clear pattern of growth can easily justify a /11 at a time from a RIR.
  - But poorly managed networks may find it much "easier" to get a /24 at a time from their providers, once a month, for the next 10 years.
    - How many people here have customers who ask for "20 Class C's"?
    - Unfortunately this doesn't just harm that network, it harms everyone.

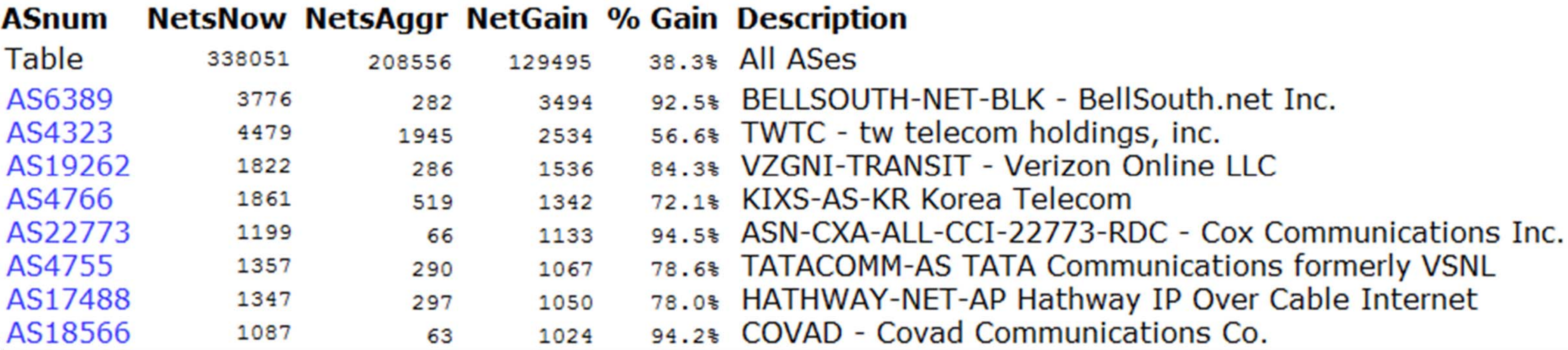
Theory: Slow Growth Causes Fragmentation

- It’s difficult to calculate exactly how much bloat this causes.
- But it sure is easy to find examples in the routing table.
  - This particular example is a hosting company announcing 129 /24s, all with the same AS-PATH, and all from their provider’s aggregates.
- As IPv4 runs out, efficient allocation will become even harder.

A Real Life Fragmentation Example (Octets Changed to Protect the Guilty)			
xxx.62.137.0/24	xxx.62.196.0/24	xxx.82.4.0/24	xxx.82.35.0/24
xxx.62.140.0/24	xxx.62.201.0/24	xxx.82.6.0/24	xxx.82.43.0/24
xxx.62.144.0/24	xxx.62.253.0/24	xxx.82.7.0/24	xxx.82.44.0/24
xxx.62.159.0/24	xxx.71.167.0/24	xxx.82.8.0/24	xxx.82.55.0/24
xxx.62.160.0/24	xxx.71.174.0/24	xxx.82.10.0/24	xxx.82.57.0/24
xxx.62.175.0/24	xxx.71.185.0/24	xxx.82.11.0/24	xxx.115.2.0/24
xxx.62.191.0/24	xxx.71.193.0/24	xxx.82.24.0/24	xxx.115.4.0/24…

Theory: It's All Traffic Engineering

- A lot of it is, particularly for inbound-heavy networks.
  - An ISP may get a /11, but often carves it up into ~/19s per market.
  - And they usually want their transit provider to haul it to the right POP.
- It can also be difficult to detect from an outsiders' view.
  - When each market is originated by its own ASN, it's easy.
  - But you can't see differing BGP nexthop attributes from the outside.
- It's difficult to know exactly how much bloat is caused by TE
  - But it's clearly responsible for the top offenders on the CIDR Report.



ASnum	NetsNow	NetsAggr	NetGain	% Gain	Description
Table	338051	208556	129495	38.3%	All ASes
AS6389	3776	282	3494	92.5%	BELLSOUTH-NET-BLK - BellSouth.net Inc.
AS4323	4479	1945	2534	56.6%	TWTC - tw telecom holdings, inc.
AS19262	1822	286	1536	84.3%	VZGNI-TRANSIT - Verizon Online LLC
AS4766	1861	519	1342	72.1%	KIXS-AS-KR Korea Telecom
AS22773	1199	66	1133	94.5%	ASN-CXA-ALL-CCI-22773-RDC - Cox Communications Inc.
AS4755	1357	290	1067	78.6%	TATACOMM-AS TATA Communications formerly VSNL
AS17488	1347	297	1050	78.0%	HATHWAY-NET-AP Hathway IP Over Cable Internet
AS18566	1087	63	1024	94.2%	COVAD - Covad Communications Co.

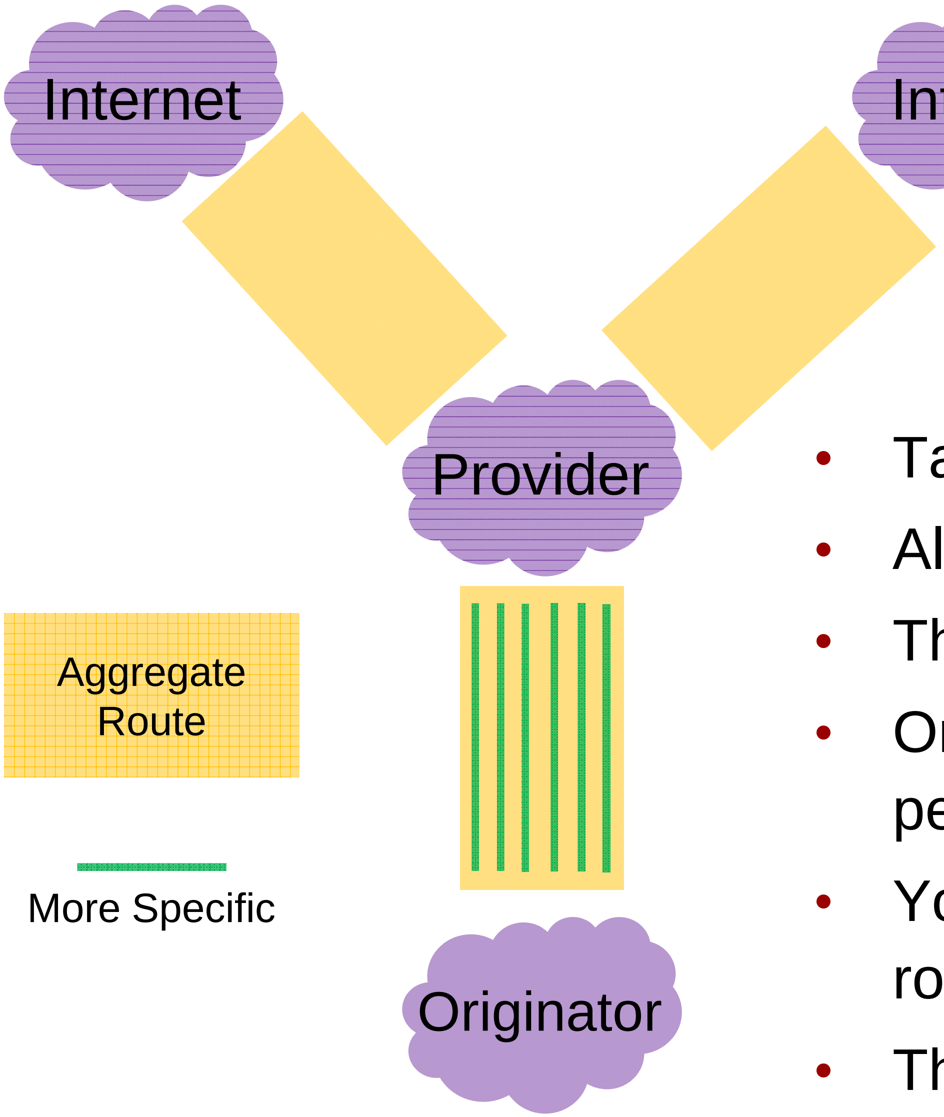
Traffic Engineering: Bellsouth

Aggregate Prefixes	# of More Specific Prefixes
65.0.0.0/12	302
65.80.0.0/14	165
66.156.0.0/15	21
66.20.0.0/15	88
67.32.0.0/14	69
68.152.0.0/13	256
68.16.0.0/14	117
68.208.0.0/12	329
70.144.0.0/12	373
72.144.0.0/12	195
74.160.0.0/11	272
74.224.0.0/11	345
98.64.0.0/11	94
184.32.0.0/12	16
216.75.0.0/14	164
Total	**2806**

Traffic Engineering: Time Warner Telecom

Aggregate Prefixes	# of More Specific Prefixes
64.132.0.0/16	59
66.192.0.0/14	659
97.65.0.0/16	47
173.226.0.0/15	126
174.46.0.0/15	66
206.169.0.0/16	52
207.67.0.0/17	79
207.235.0.0/17	62
207.250.0.0/16	168
209.12.0.0/16	50
209.136.0.0/16	39
209.163.128.0/17	67
209.234.128.0/17	75
216.54.128.0/17	98
216.136.0.0/16	39
Total	**1686**

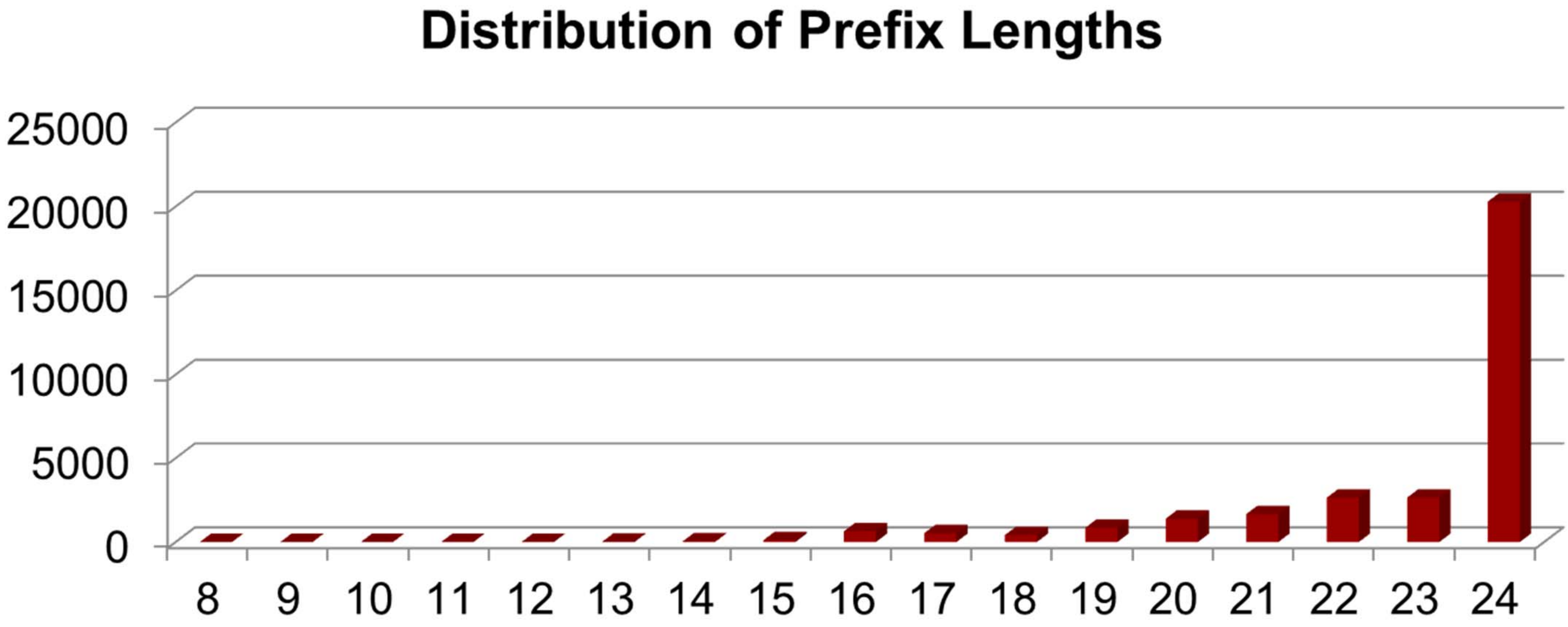
A Technique to do TE Without Pollution



- Tag your more-specifics with No-Advertise.
- Allow the aggregates to propagate normally.
- The aggregate draws traffic to your provider.
- Once there, the more-specifics kick in and perform their traffic engineering function.
- Your provider still deals with the increased routes, but the rest of the Internet is spared.
- This can also help reduce BGP route churn!

Theory: Lots of Redistribution

- Looking at routes with an Unknown BGP Origin Code:
  - These account for 31K (or around 9%) of the global table.
  - A bit higher % of /24s, but not wildly different from the global view.

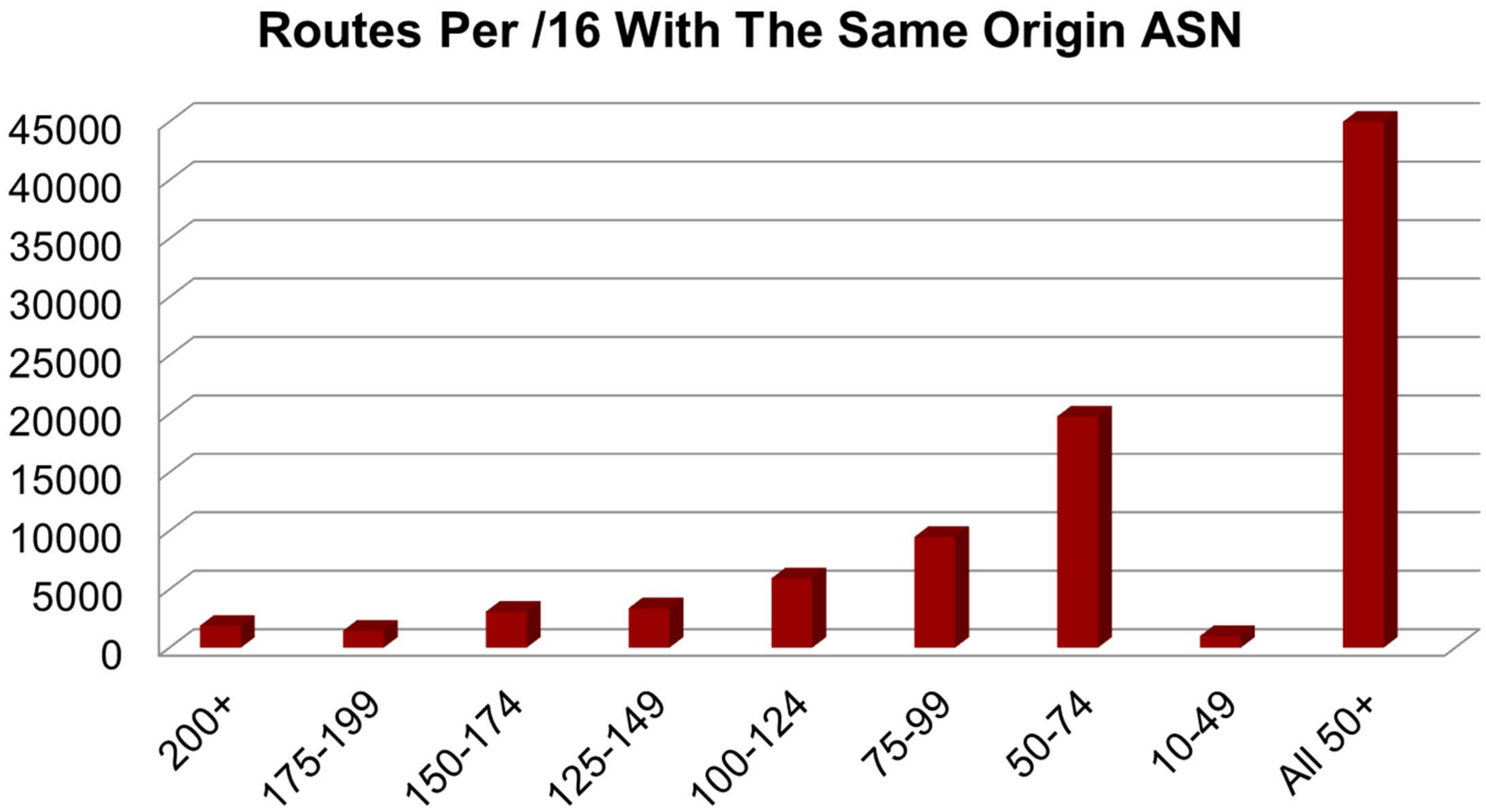


Theory: People Are Just Being Stupid

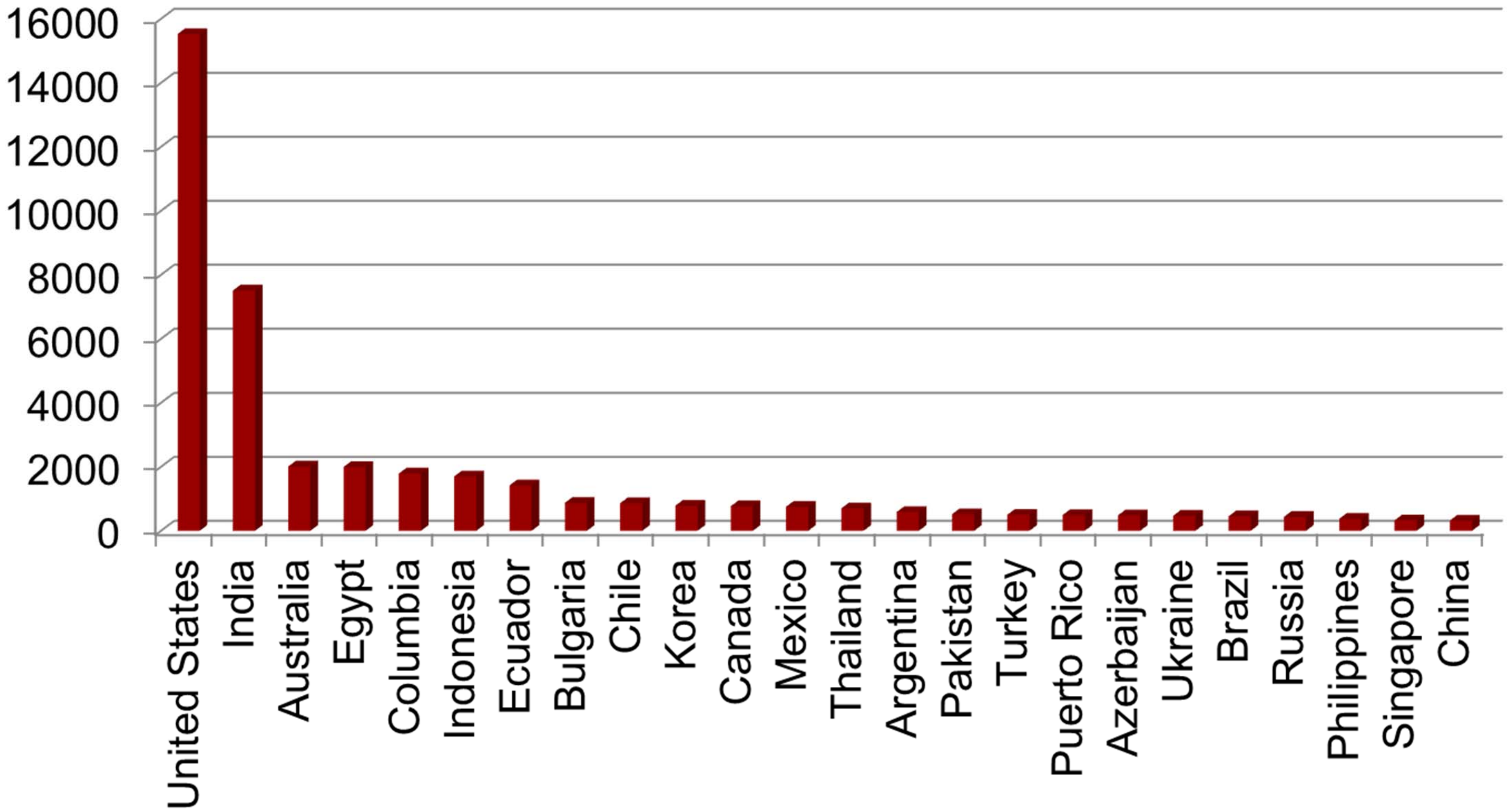
Worst offenders: Routes with the same origin ASN, by count per /16

/16 Block	Route Count	Origin ASN	Country
186.42.0.0/16	226	14420	Ecuador
72.27.0.0/16	219	10292	Jamaica
94.20.0.0/16	215	29049	Azerbaijan
125.99.0.0/16	213	17488	India
60.243.0.0/16	208	17488	India
116.72.0.0/16	205	17488	India
220.227.0.0/16	204	18101	India
190.152.0.0/16	204	14420	Ecuador
116.74.0.0/16	202	17488	India
190.131.0.0/16	192	27738	Ecuador
41.235.0.0/16	183	8452	Egypt
66.192.0.0/16	182	4323	United States

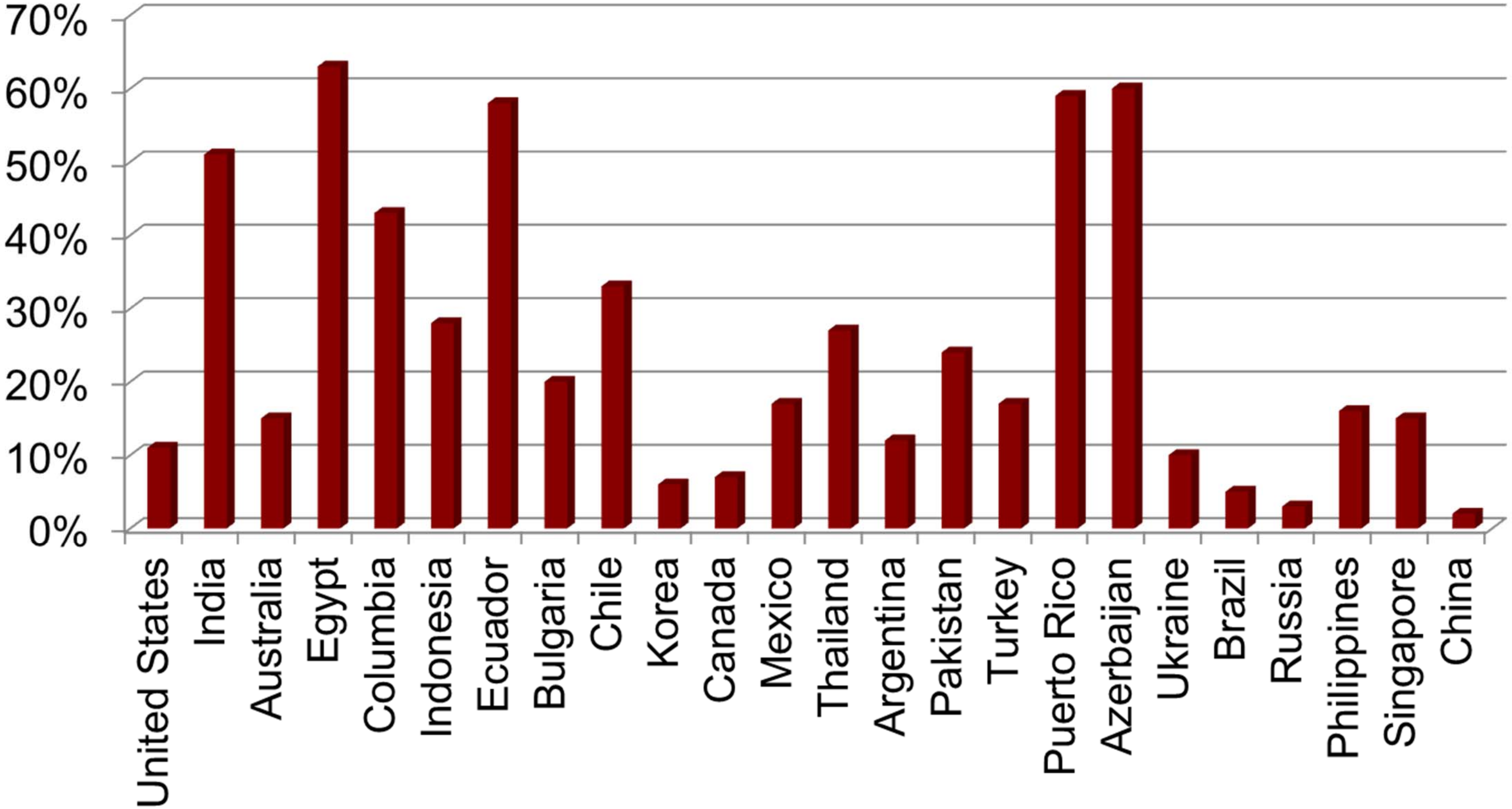
Can We Identify Deaggregates Automatically?



Breakdown of Deaggregates By Country



Deaggregates as Percentage of Total Routes

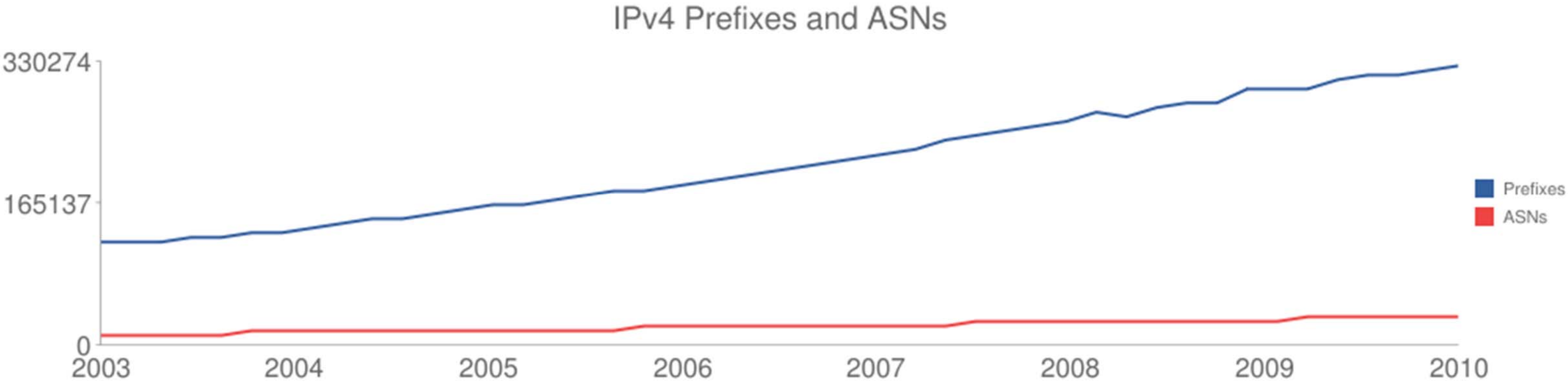


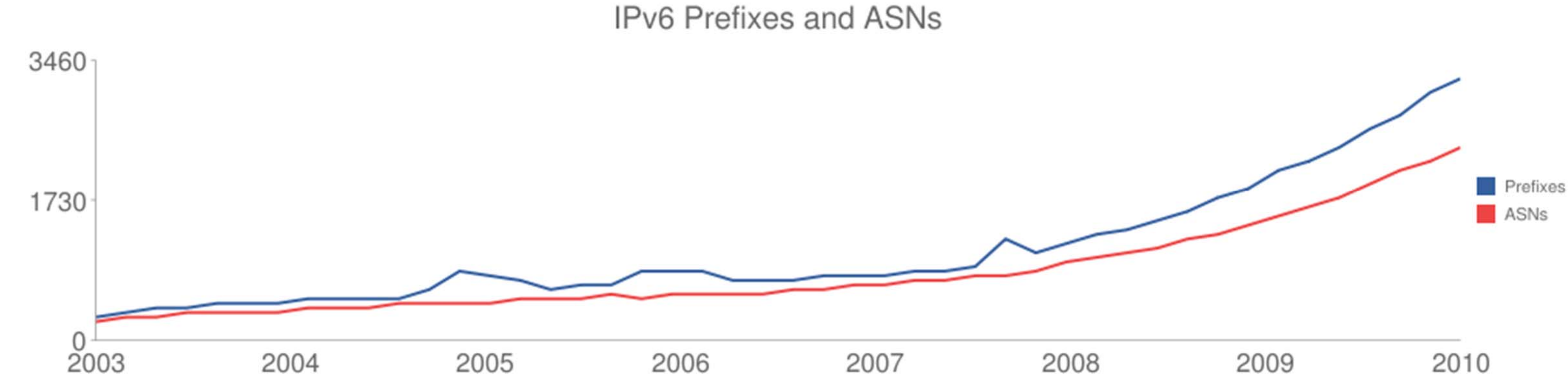
Some Random Funny Bad Routes

7018 Originating Starbucks’ 98.96.0.0/14, One /24 At A Time			
98.96.41.0/24	98.97.114.0/24	98.97.142.0/24	98.97.155.0/24
98.96.74.0/24	98.97.116.0/24	98.97.143.0/24	98.97.156.0/24
98.96.86.0/24	98.97.117.0/24	98.97.144.0/24	98.97.160.0/24
98.96.100.0/24	98.97.118.0/24	98.97.149.0/24	98.97.161.0/24
98.96.108.0/24	98.97.131.0/24	98.97.150.0/24	98.97.162.0/24
98.96.149.0/24	98.97.140.0/24	98.97.152.0/24	98.97.164.0/24
98.96.247.0/24	98.97.141.0/24	98.97.154.0/24	98.97.168.0/24

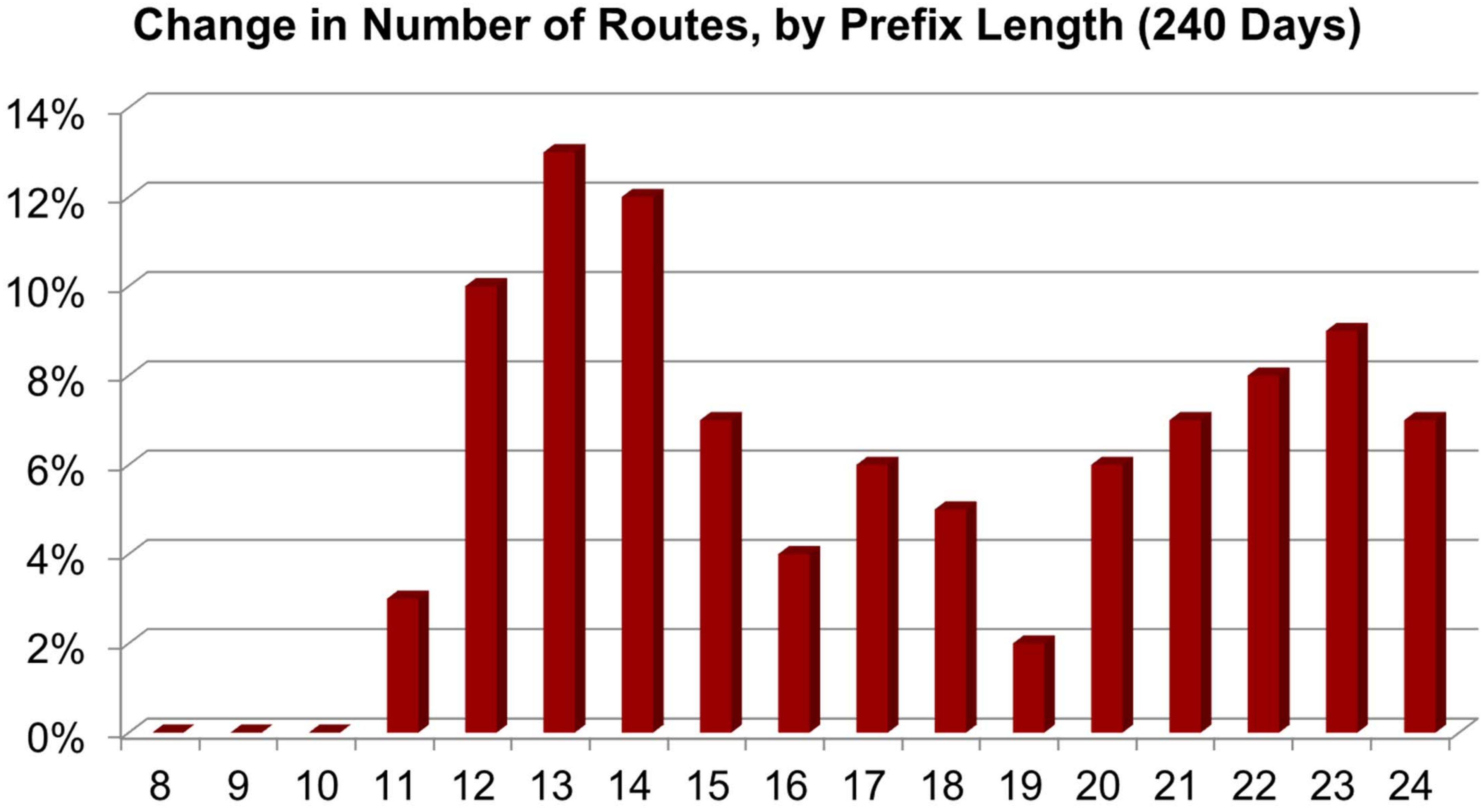
The Impact of IPv6 On The Routing Table

Routes vs. ASNs, v4 and v6 Comparison


IPv4 Prefixes and ASNs


IPv6 Prefixes and ASNs

Question: Is Deaggregation Increasing?



Send questions, comments, complaints to:

Richard A Steenbergen ras@nlayer.net