BGP in 2013

Geoff Huston
APNIC

“Conventional “wisdom” about routing:

> “The rapid and sustained growth of the Internet over the past several decades has resulted in large state requirements for IP routers. In recent years, these requirements are continuing to worsen, due to increased deaggregation (advertising more specific routes) arising from load balancing and security concerns..”

Quoted from a 2012 research paper on routing

“Conventional “wisdom” about routing:

> “The rapid and sustained growth of the Internet over the past several decades has resulted in large state requirements for IP routers. In recent years, these requirements are continuing to worsen, due to increased deaggregation (advertising more specific routes) arising from load balancing and security concerns..”

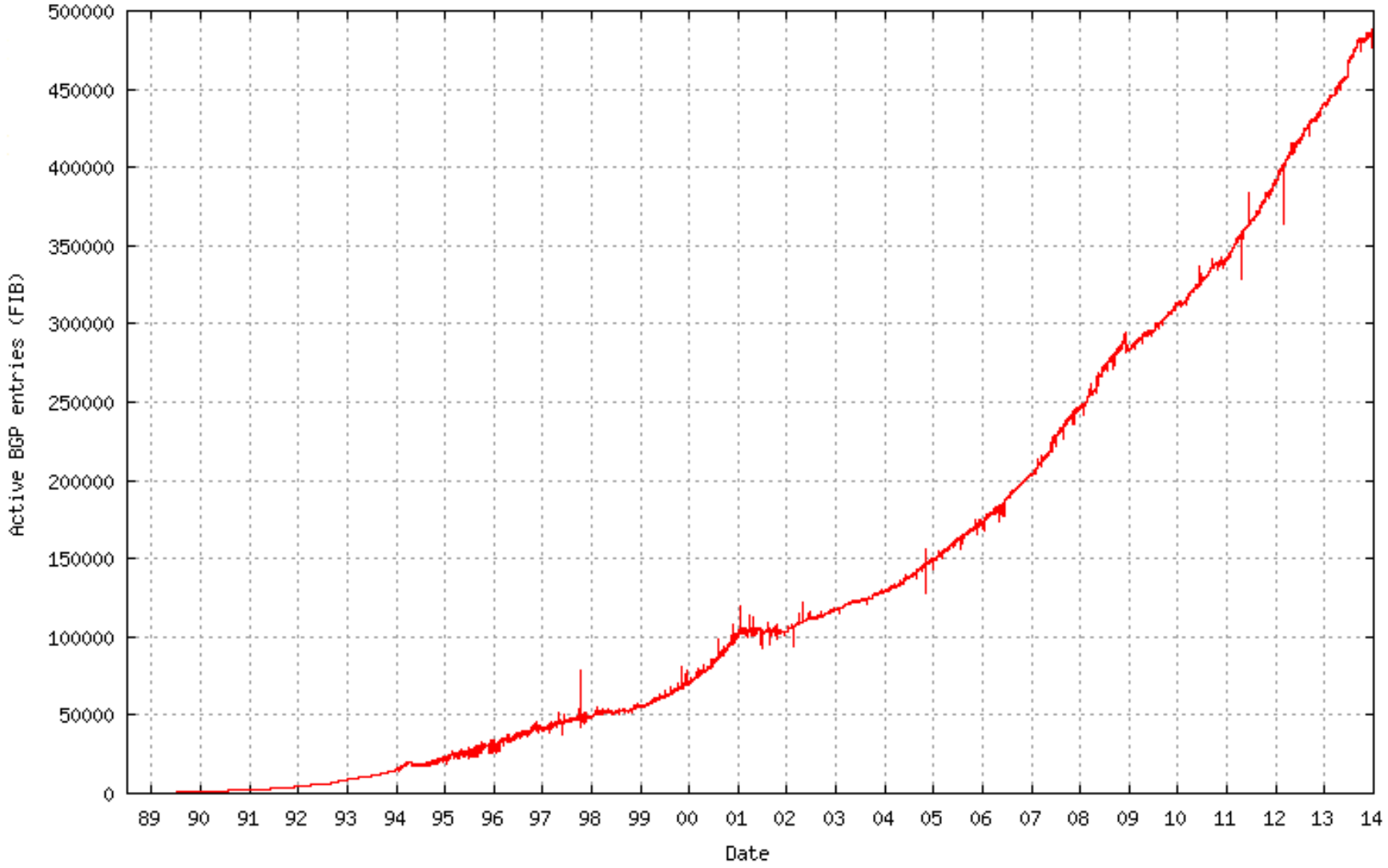
Quoted from ... 2012 ... routing

is this really true, or do we accept it as true without actually looking at the real behaviours of the Internet's routing system???

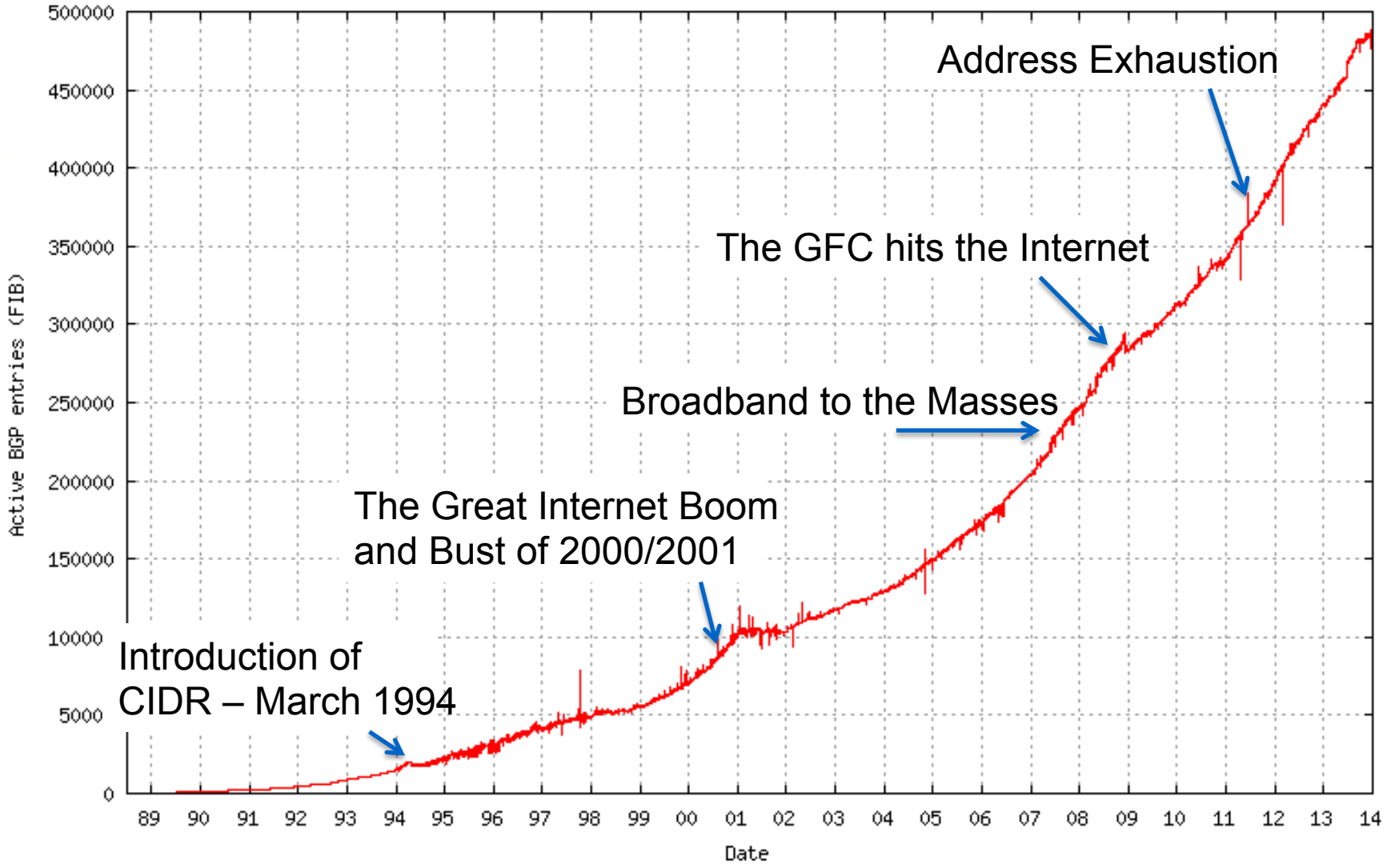
In this presentation we'll explore the space of inter-domain routing and look at

- the growth of the eBGP routing table over time and some projections for future growth
- the extent to which more specifics are dominating routing table growth ... or not

The Big Picture of the v4 Routing Table



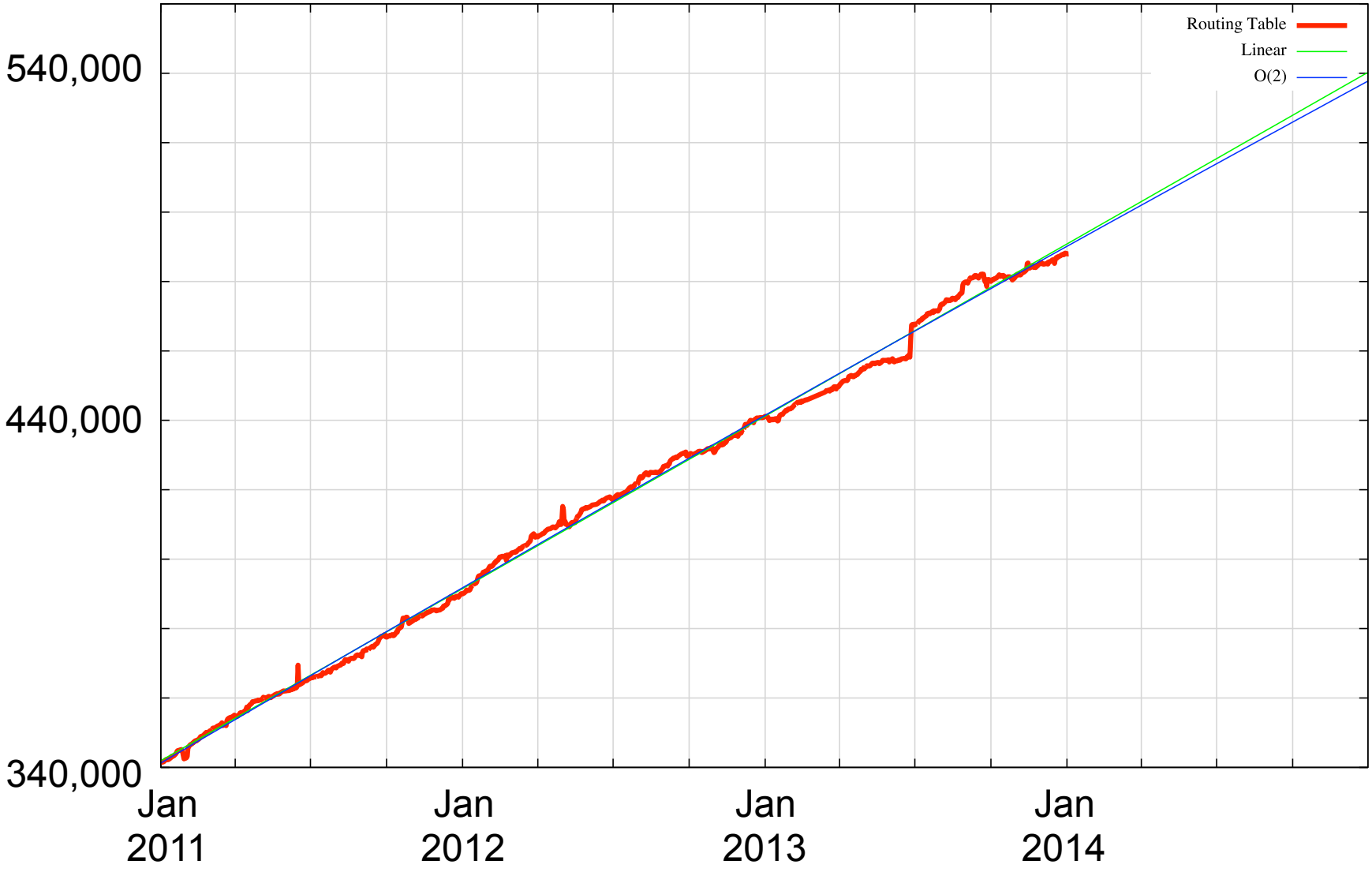
The Big Picture of the v4 Routing Table



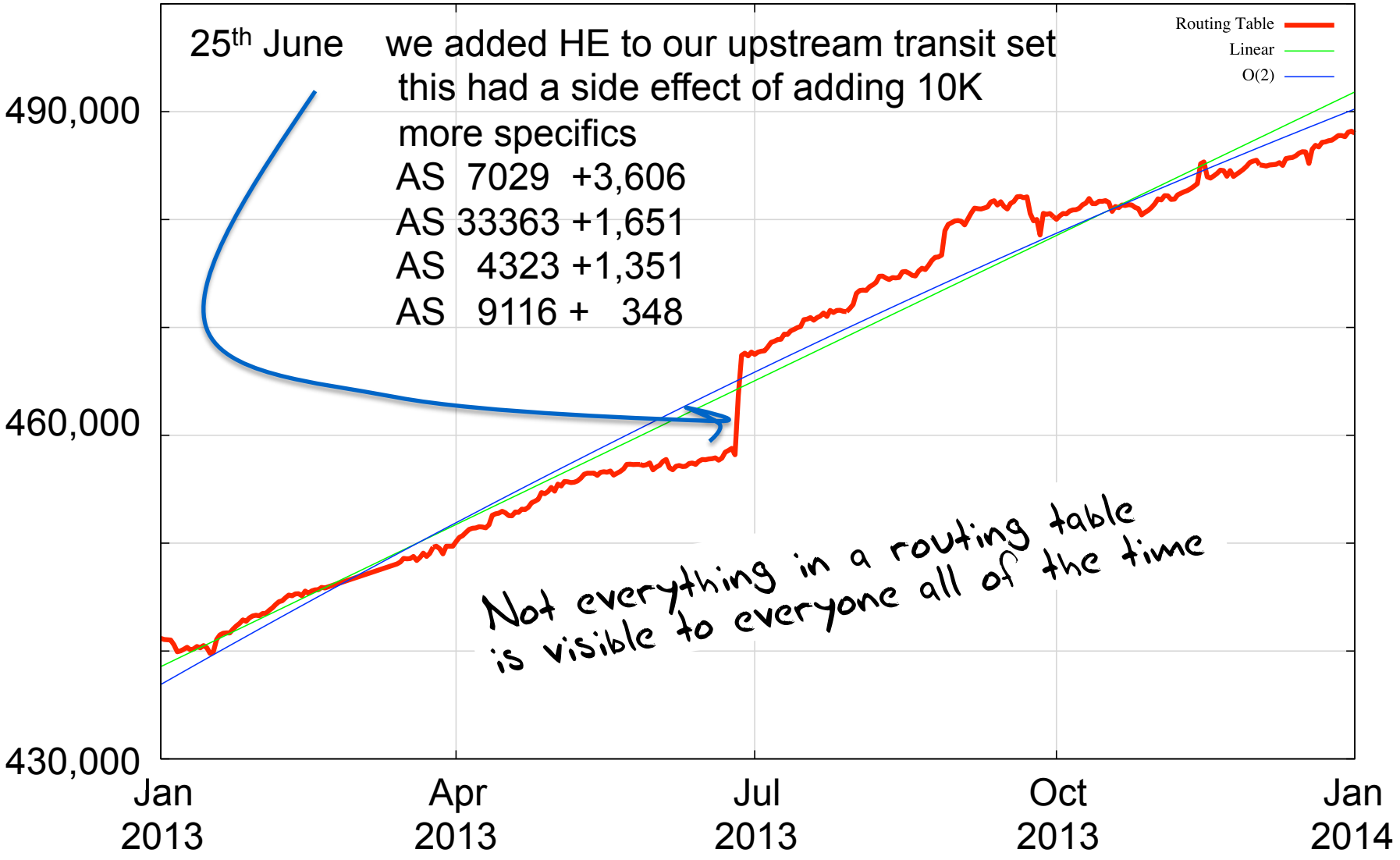
The Routing Table in 2012-2013

Lets look at the recent past in a little more detail...

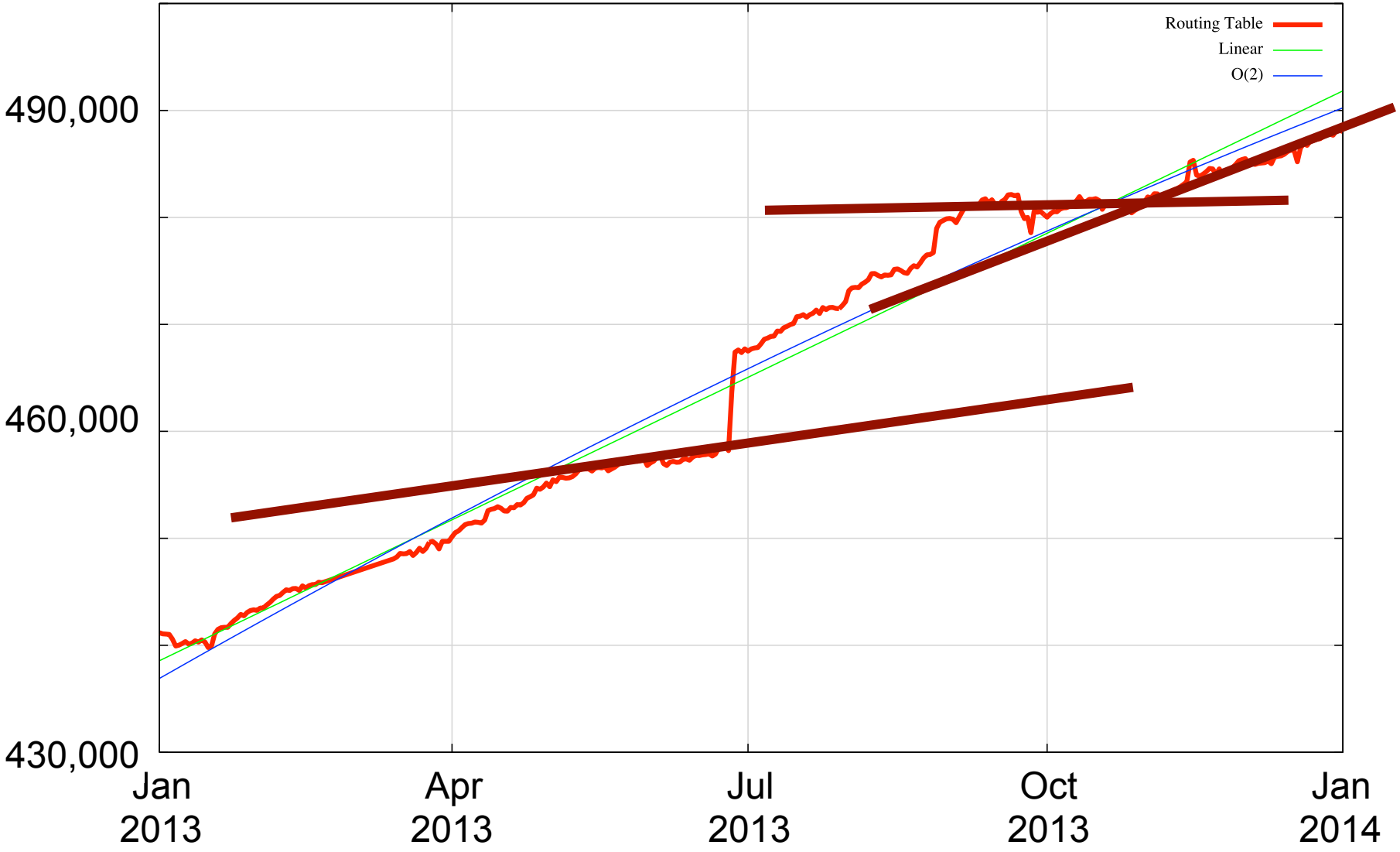
IPv4 BGP Prefix Count 2011 - 2013



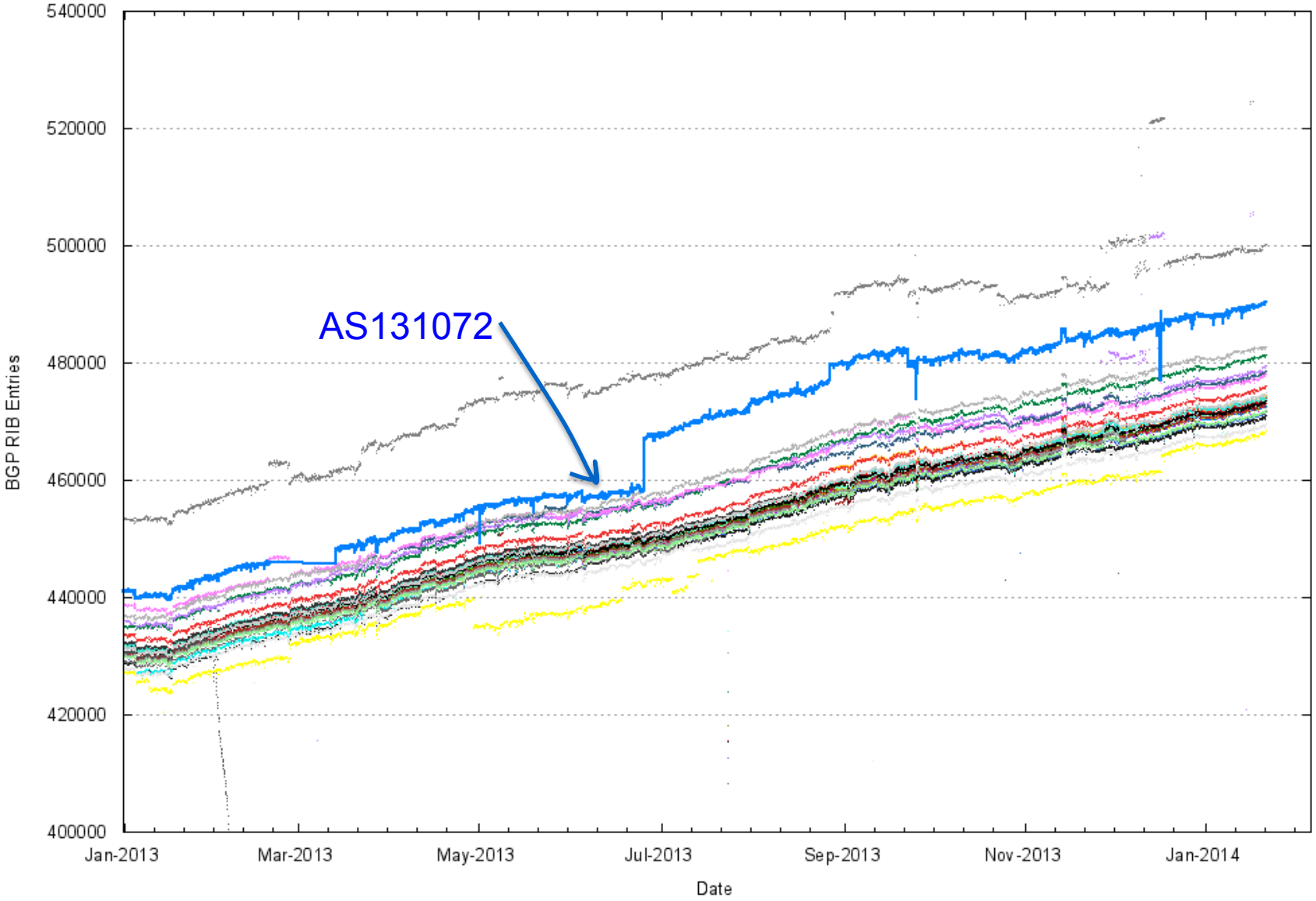
IPv4 BGP Prefix Count 2013



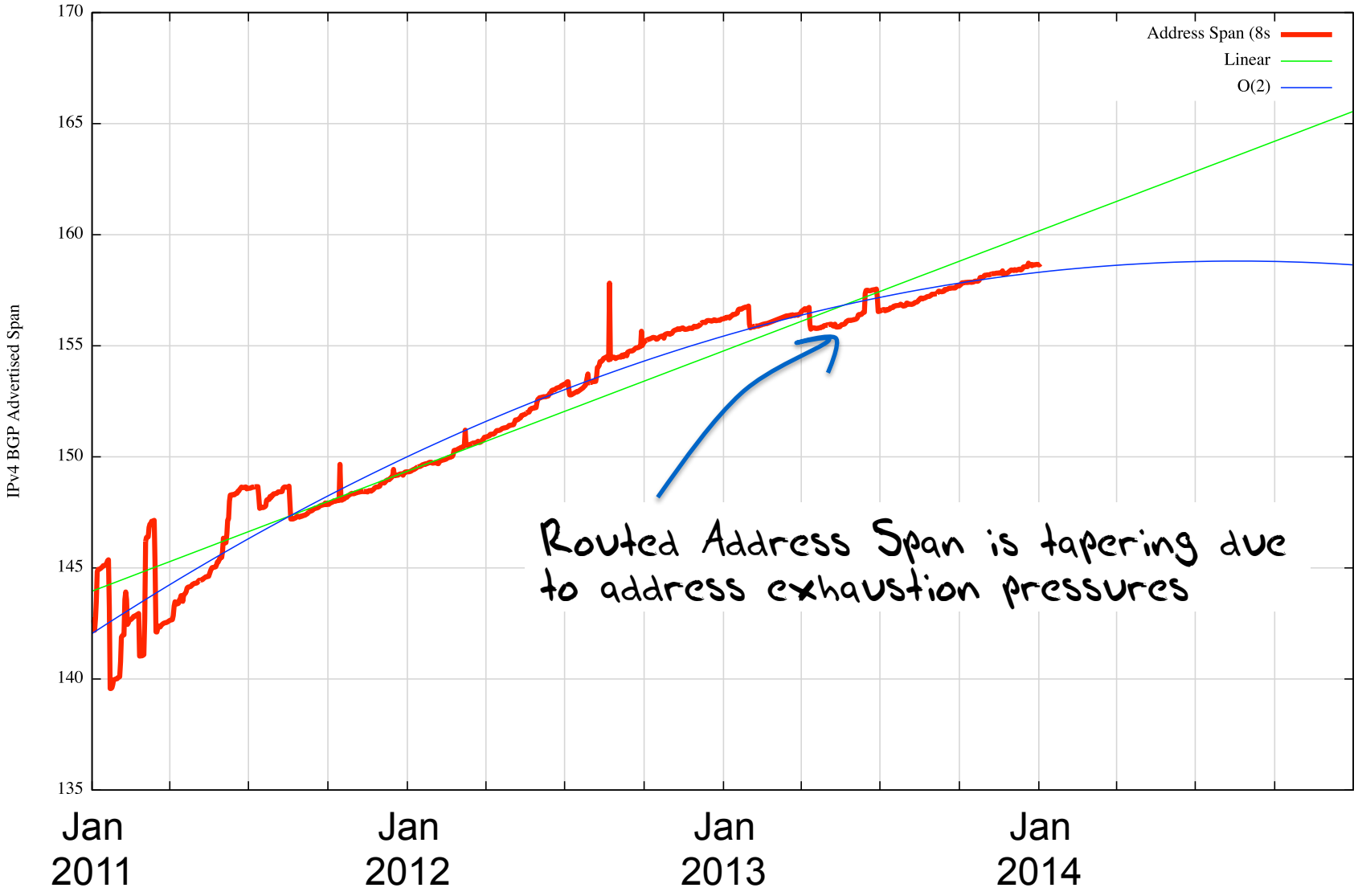
IPv4 BGP Prefix Count 2013



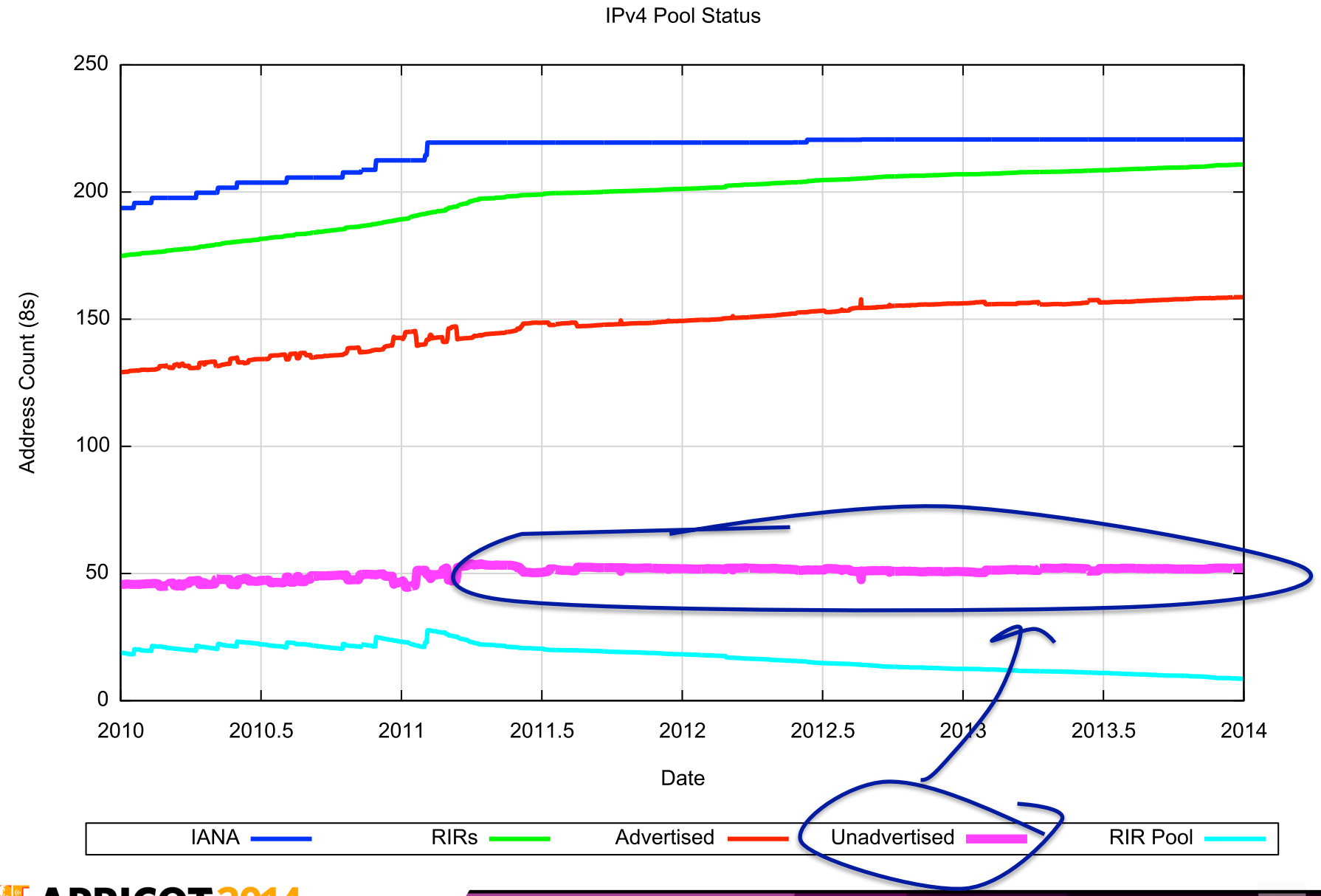
As seen by peers of Route Views



IPv4 Routed Address Span: 2011 - 2013



IPv4 Address Pool

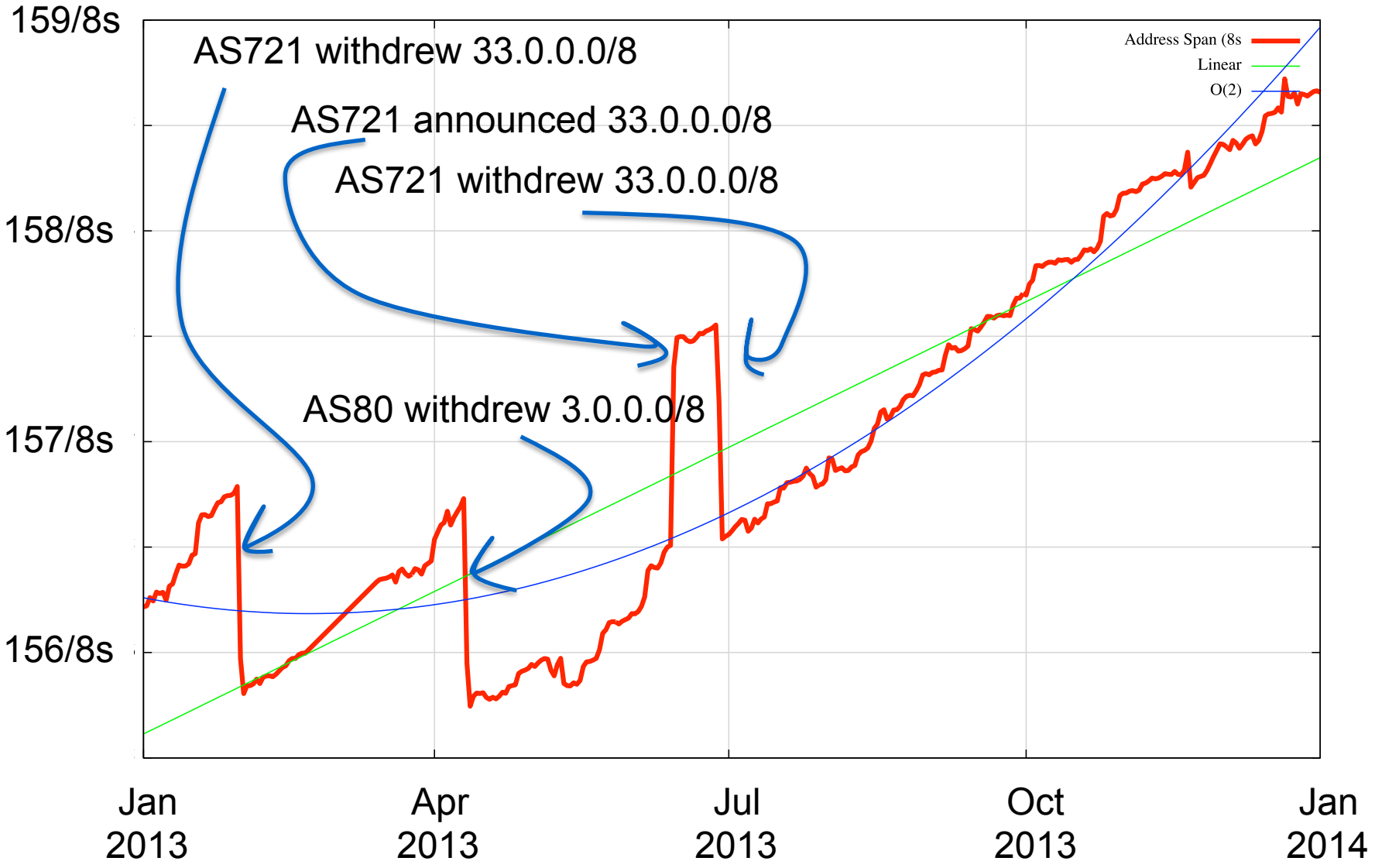


That Unadvertised Address Pool

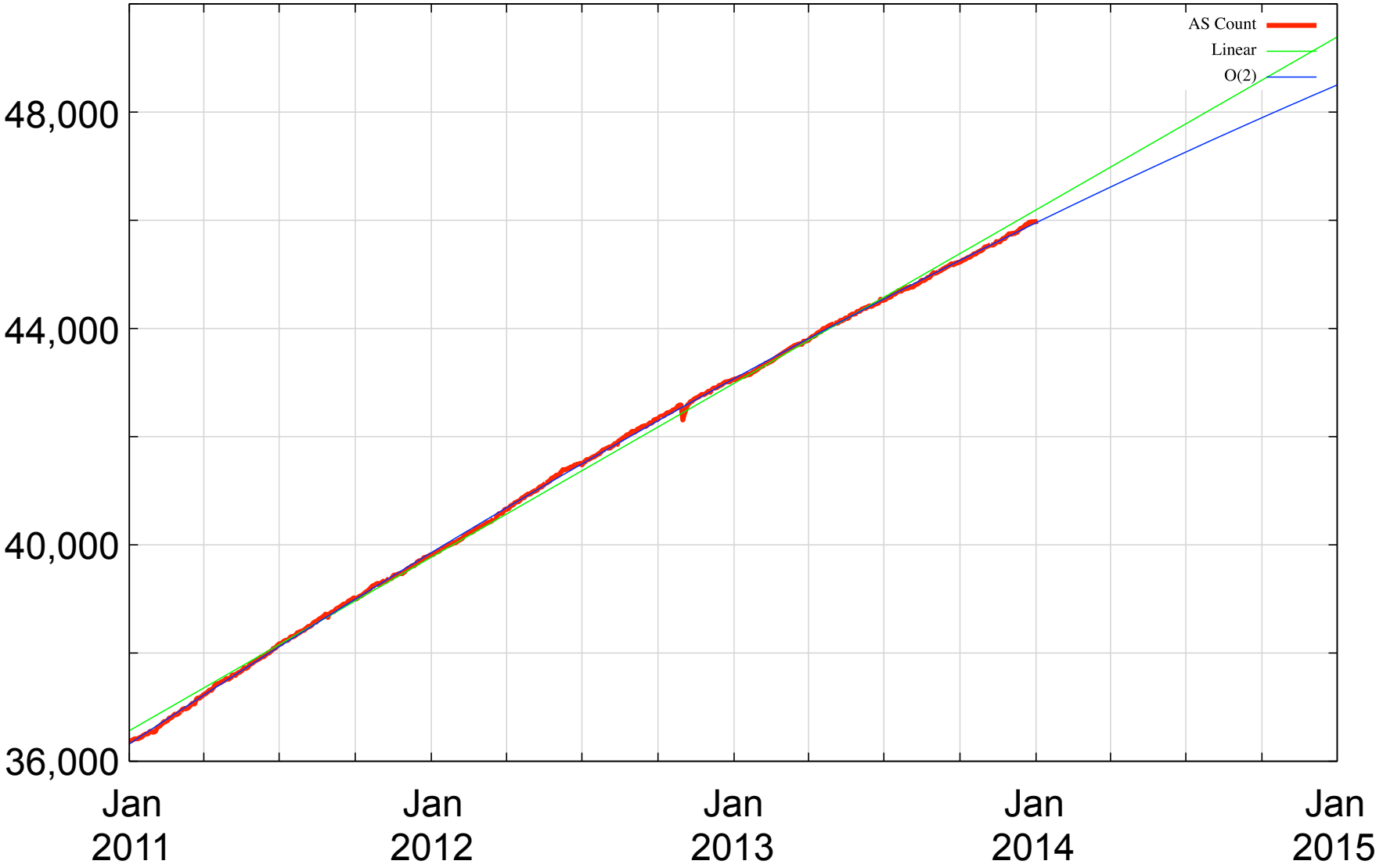
Appears to be relatively static in size since early 2011 at some 50 /8s, or 20% of the IPv4 global unicast space

At this stage its likely that ARIN and LACNIC will both hit their address pool exhaustion threshold levels at the end of 2014

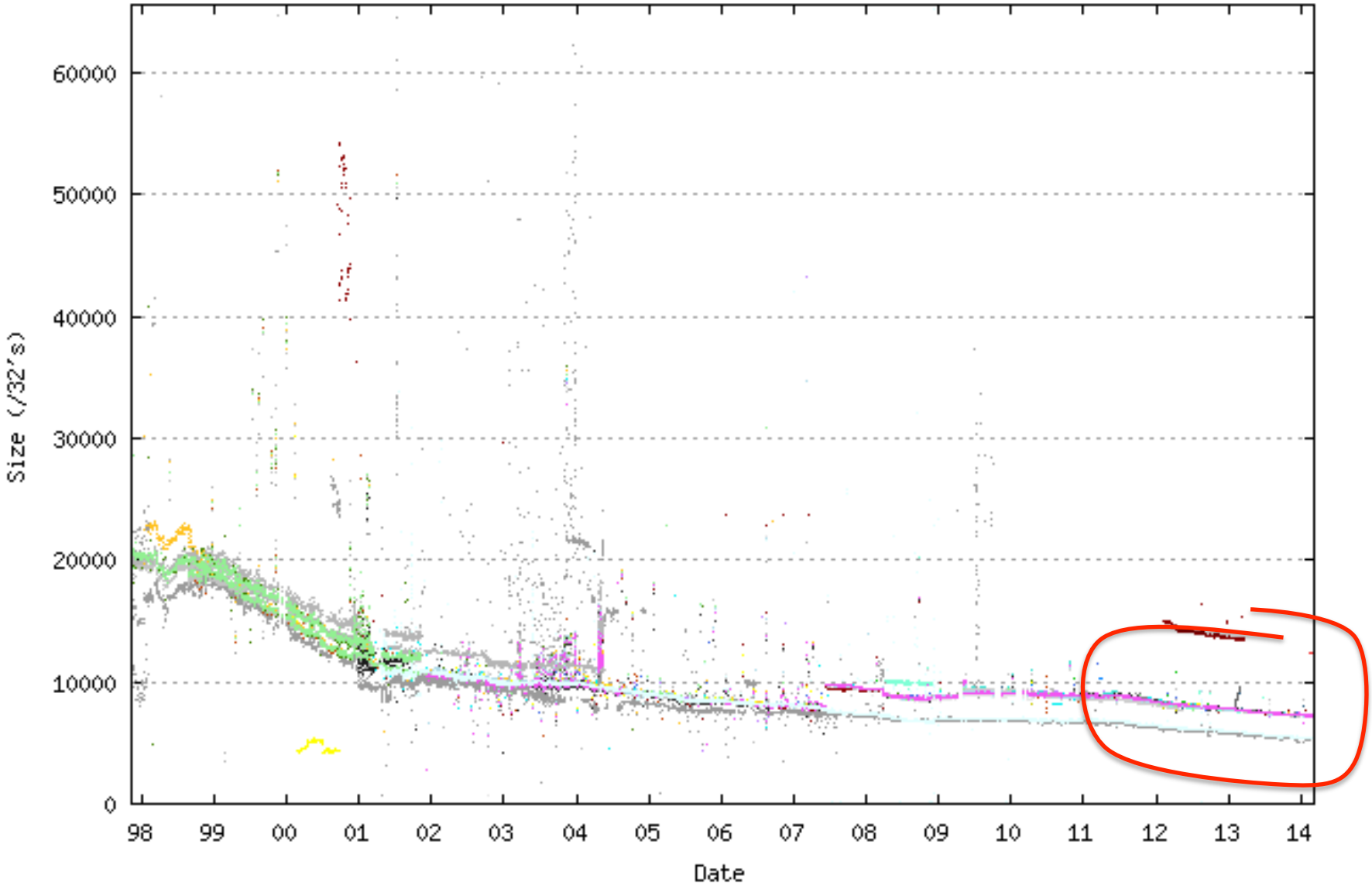
IPv4 Routed Address Span



IPv4 Routed AS Count



Average Announced Prefix Size



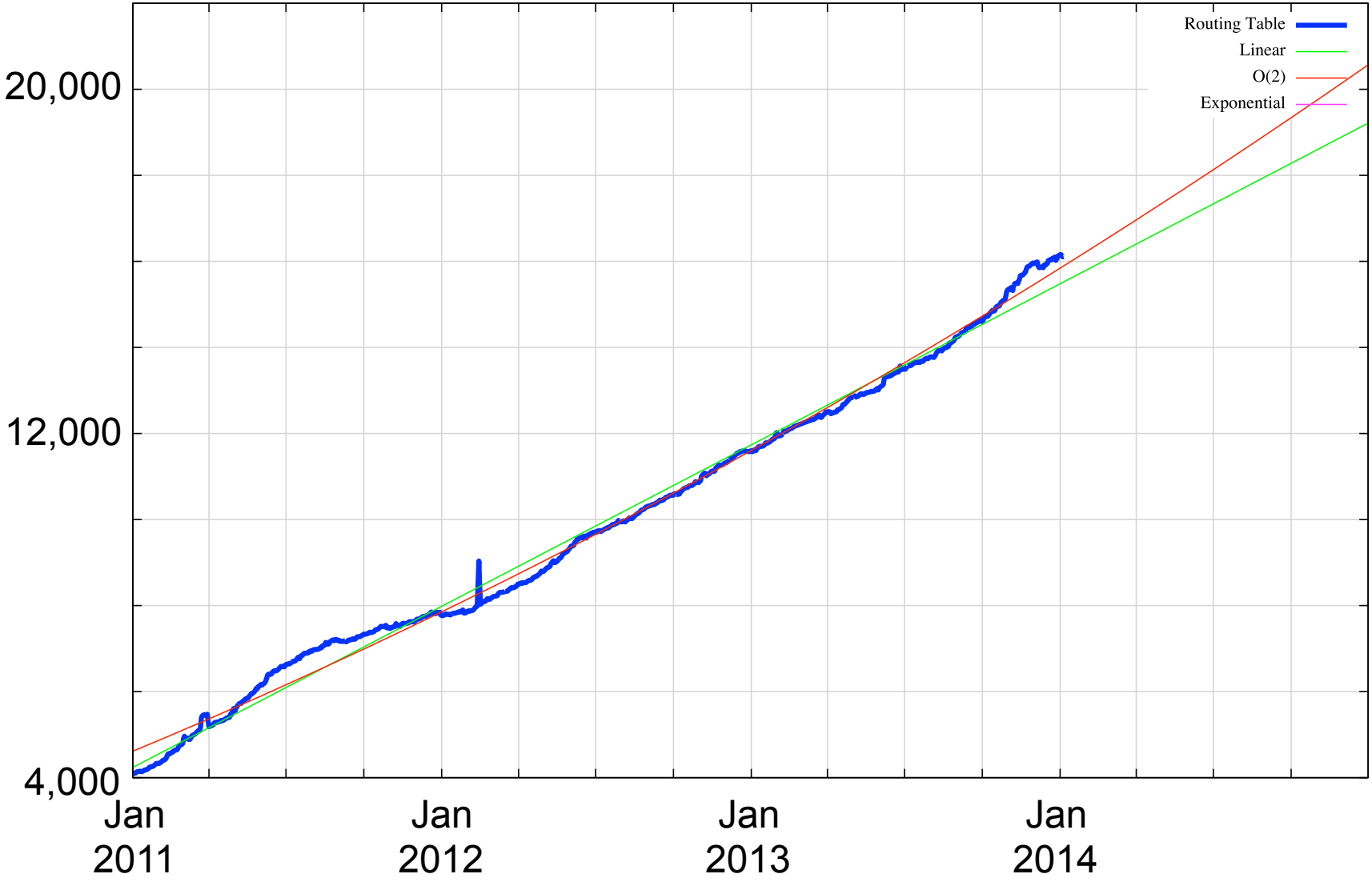
IPv4 2011 BGP Vital Statistics

	Jan-13	Jan-14	
Prefix Count	440,000	488,000	**+11%**
Roots	216,000	237,000	+10%
More Specifics	224,000	251,000	+12%
Address Span	156/8s	159/8s	**+ 2%**
AS Count	43,000	46,000	**+ 7%**
Transit	6,100	6,600	+ 8%
Stub	36,900	39,400	+ 7%

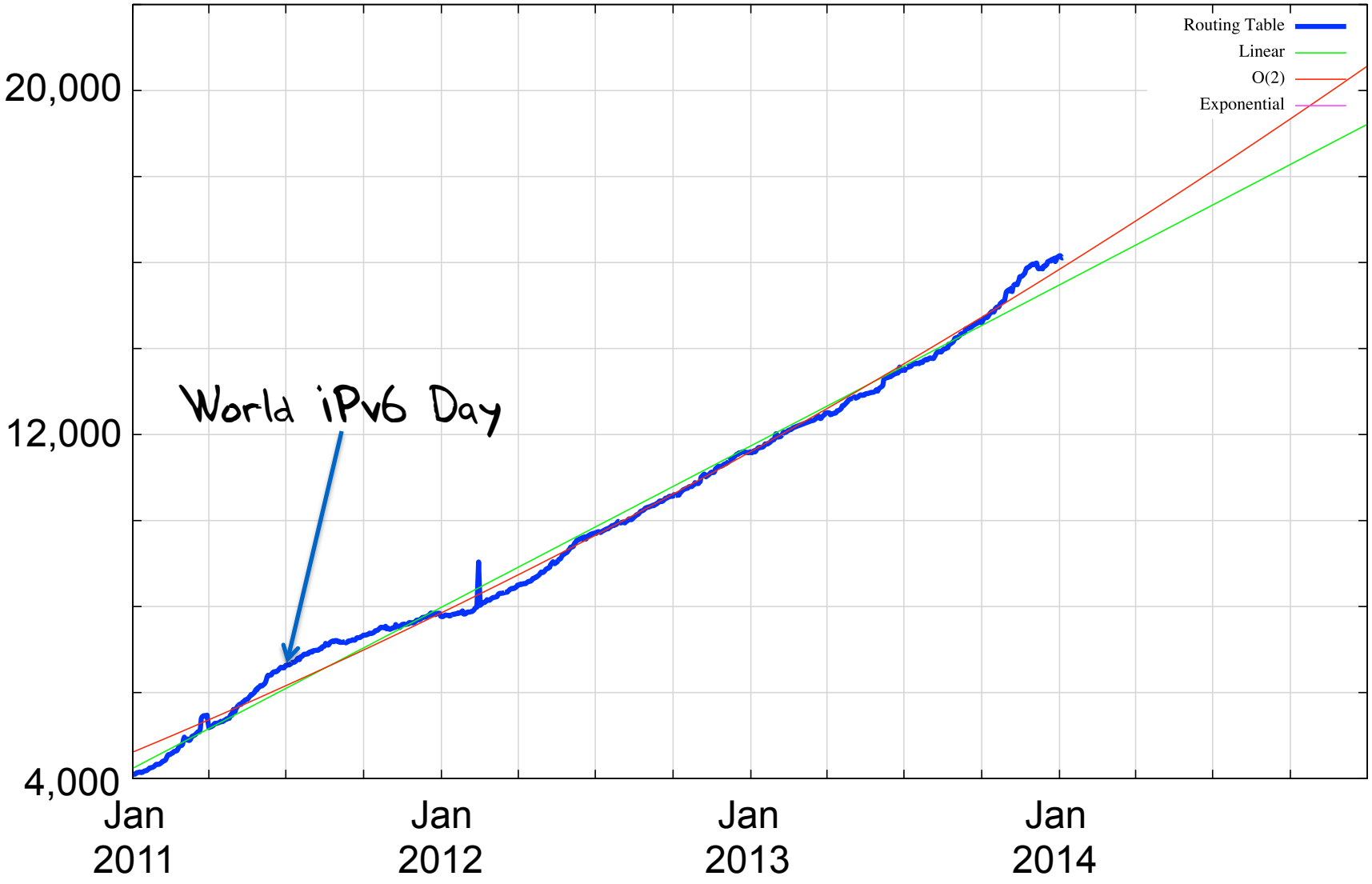
IPv4 in 2013 – Growth is Slowing

- Overall Internet growth in terms of BGP is at a rate of some **~8-10% p.a.**
  - This is down by 33% from 2010
- Address span growing far more slowly than the table size
- AS growth has persisted relatively uniformly
- The average announced prefix size has been falling since 2011

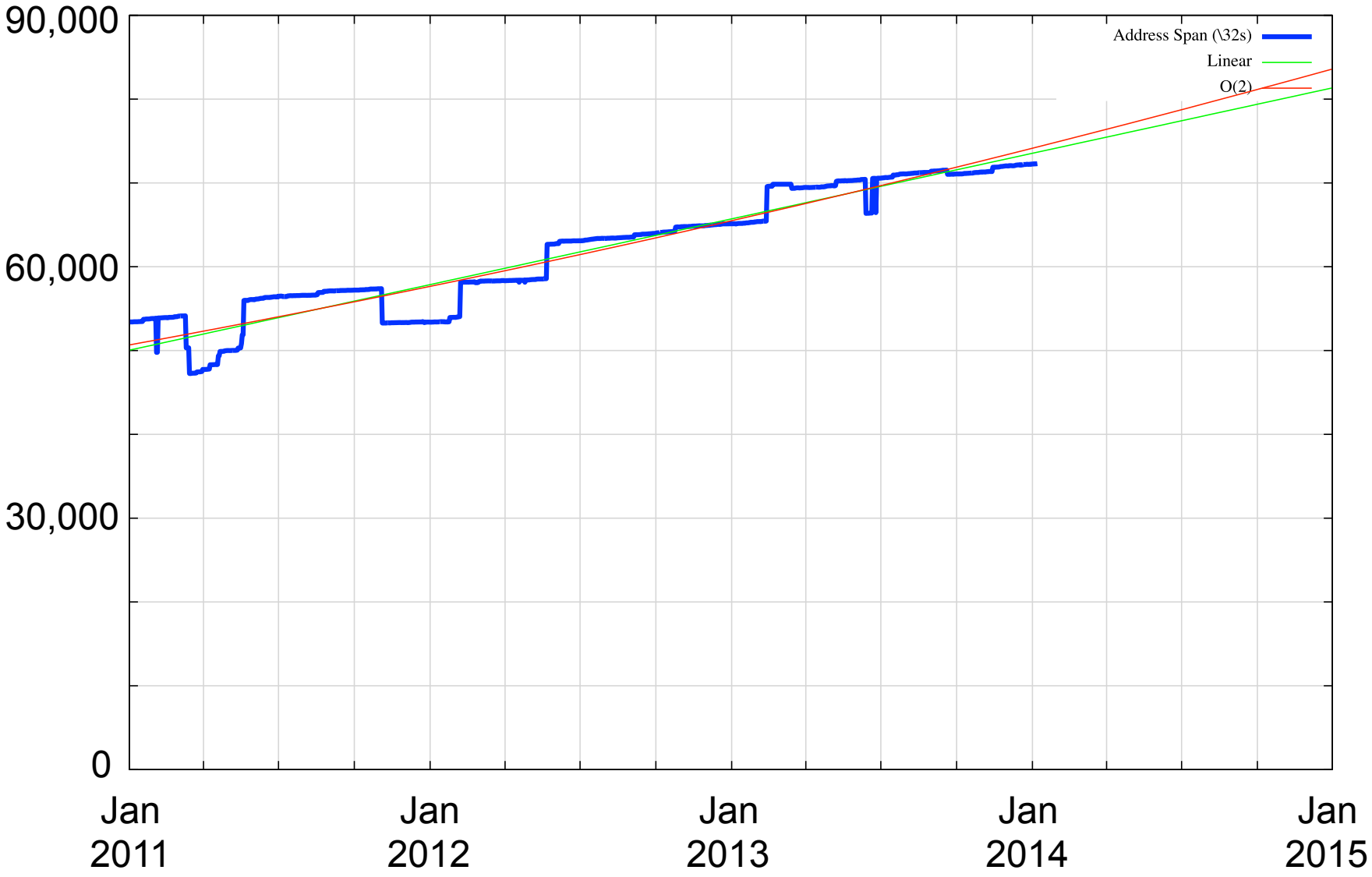
IPv6 BGP Prefix Count



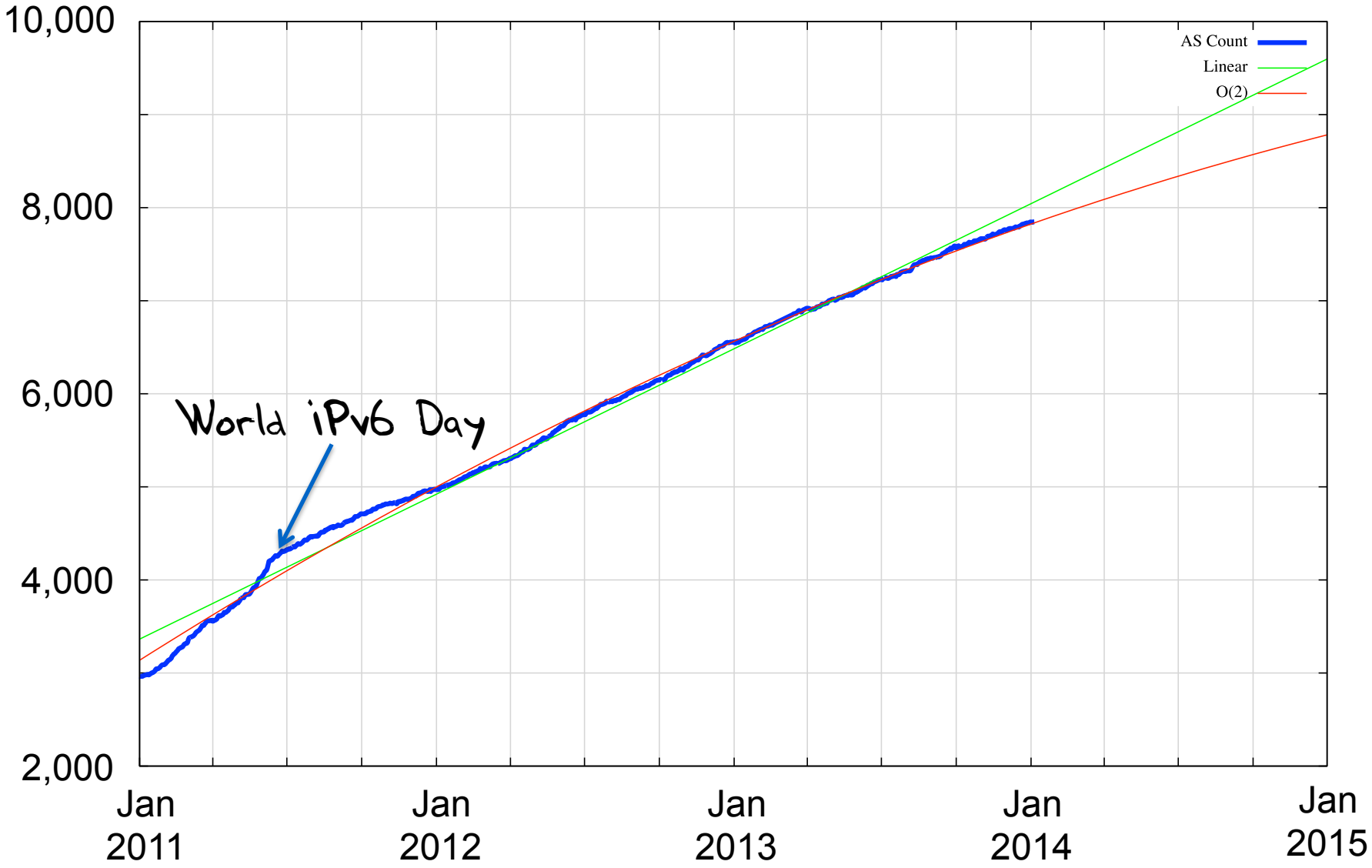
IPv6 BGP Prefix Count



IPv6 Routed Address Span



IPv6 Routed AS Count



IPv6 2011 BGP Vital Statistics

	Jan-13	Jan-14	p.a. rate
Prefix Count	11,500	16,100	+ 40%
Roots	8,451	11,301	+ 34%
More Specifics	3,049	4,799	+ 57%
Address Span (/32s)	65,127	72,245	+ **11%**
AS Count	6,560	7,845	+ **20%**
Transit	1,260	1,515	+ 20%
Stub	5,300	6,330	+ 19%

IPv6 in 2013

- Overall IPv6 Internet growth in terms of BGP is:

  20% - 40 % p.a.

  (2011 growth rate was ~ 90%)

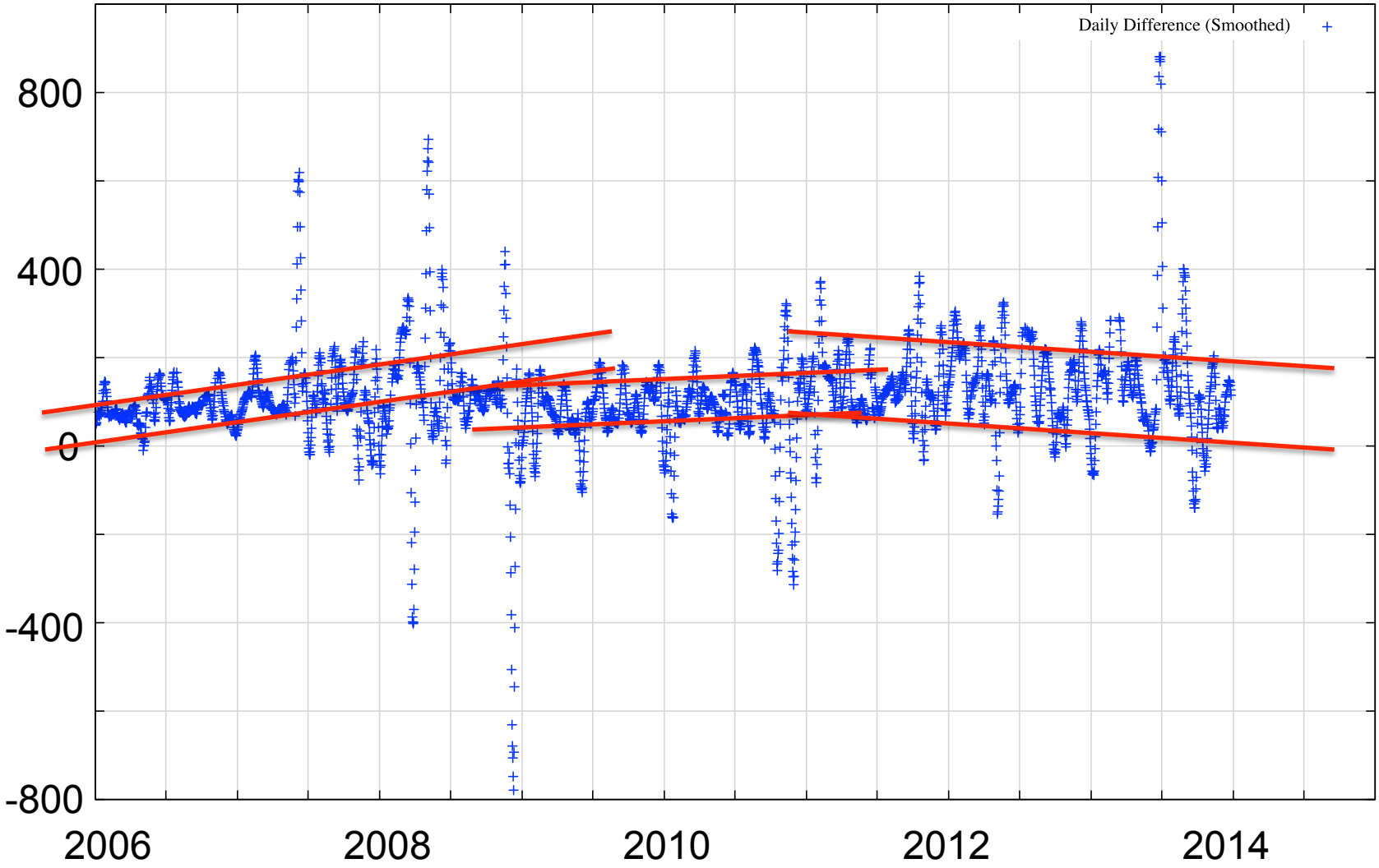
(Looking at the AS count, if these relative growth rates persist then the IPv6 network would span the same network domain as IPv4 in16 years time . That's by 2030!)

BGP Size Projections

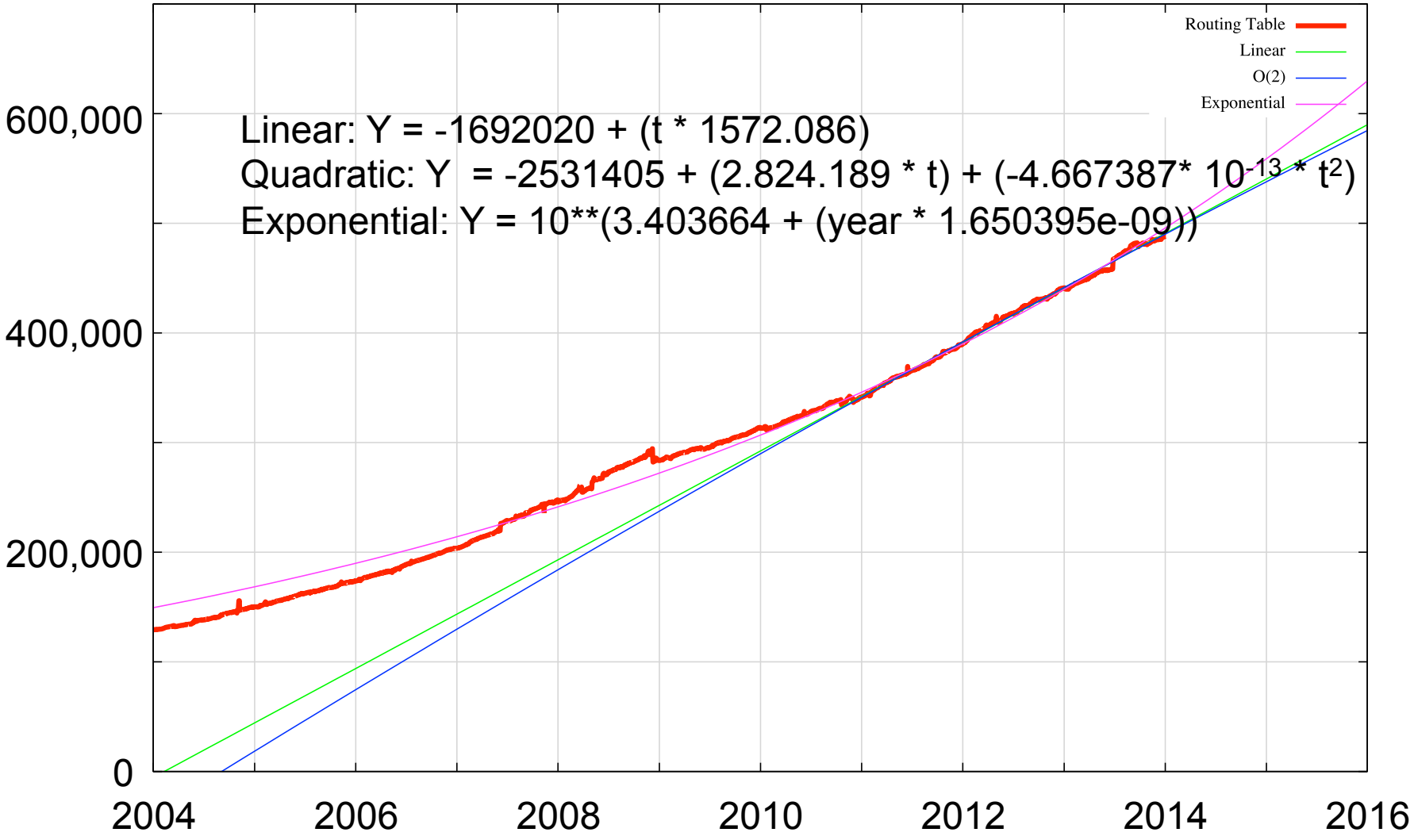
- How big does it get? How quickly?
  - For IPv4 this is a time of **extreme uncertainty**
    - Registry IPv4 address run out
    - Uncertainty over the impacts of any after-market in IPv4 on the routing table

      which makes this projection even more speculative than normal!

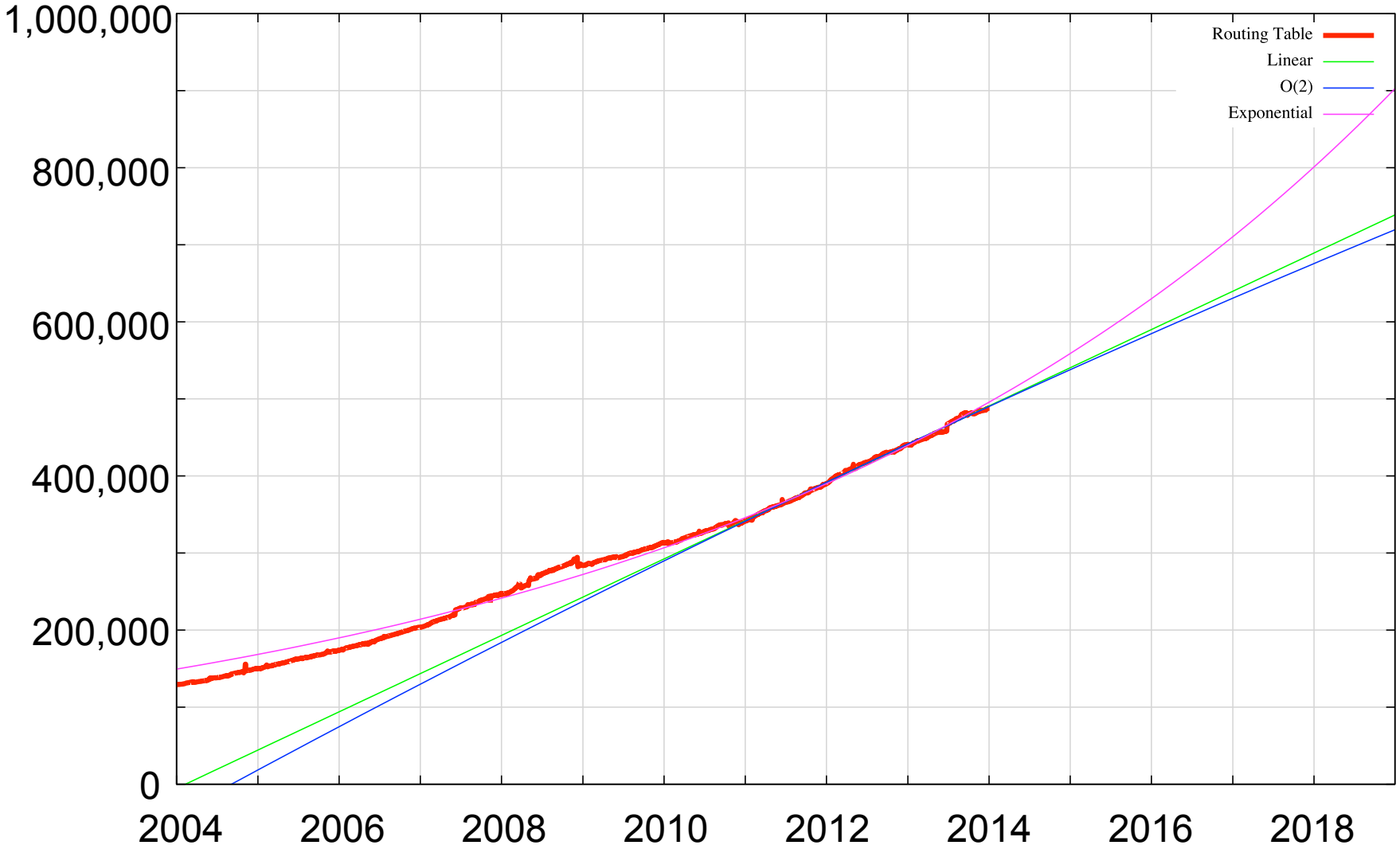
Daily Growth Rates for IPv4



IPv4 Table Size



IPv4 Table Size

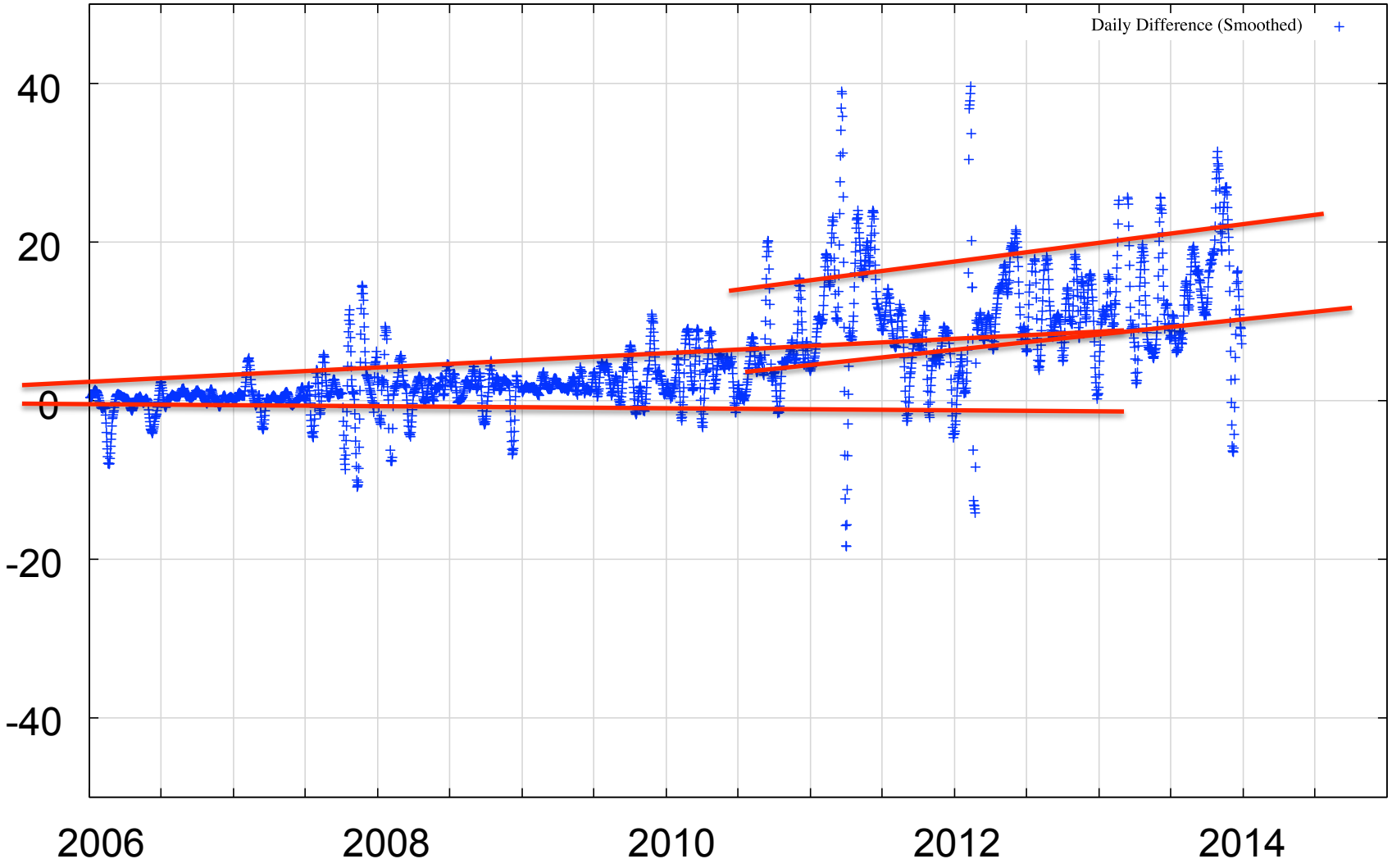


IPv4 BGP Table Size predictions

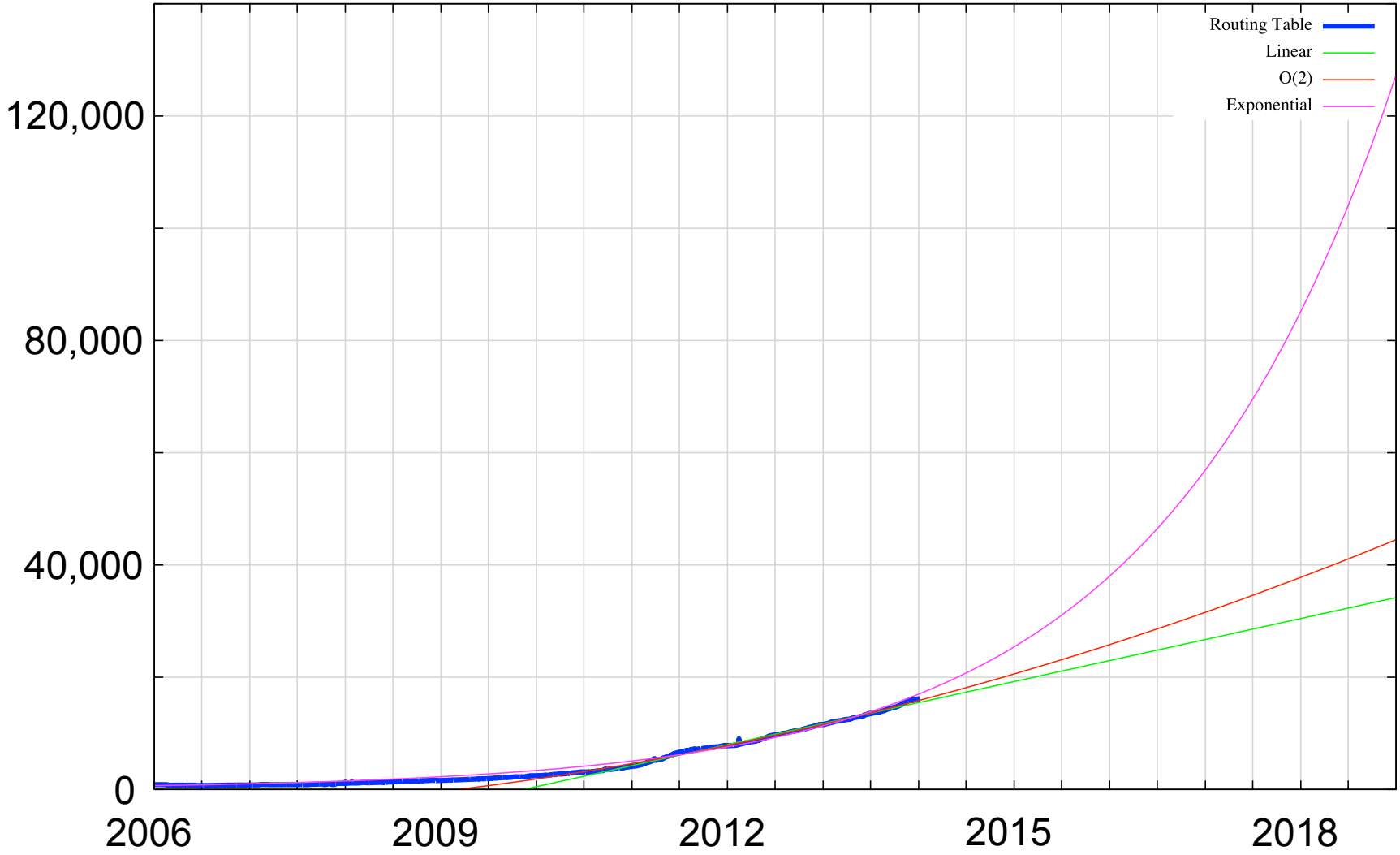
		Linear Model	Exponential Model
Jan	2013	441,172 entries	
	2014	488,011 entries	
	2015	*540,000 entries*	*559,000*
	2016	*590,000 entries*	*630,000*
	2017	*640,000 entries*	*710,000*
	2018	*690,000 entries*	*801,000*
	2019	*740,000 entries*	*902,000*

* *These numbers are dubious due to uncertainties introduced by IPv4 address exhaustion pressures.*

Daily Growth Rates for IPv6



IPv6 Table Projection



IPv6 BGP Table Size predictions

		Exponential Model	Linear Model
Jan	2013	11,600 entries	
	2014	16,200 entries	
	2015	*25,400 entries*	*19,000*
	2016	*38,000 entries*	*23,000*
	2017	*57,000 entries*	*27,000*
	2018	*85,000 entries*	*30,000*
	2019	*127,000 entries*	*35,000*

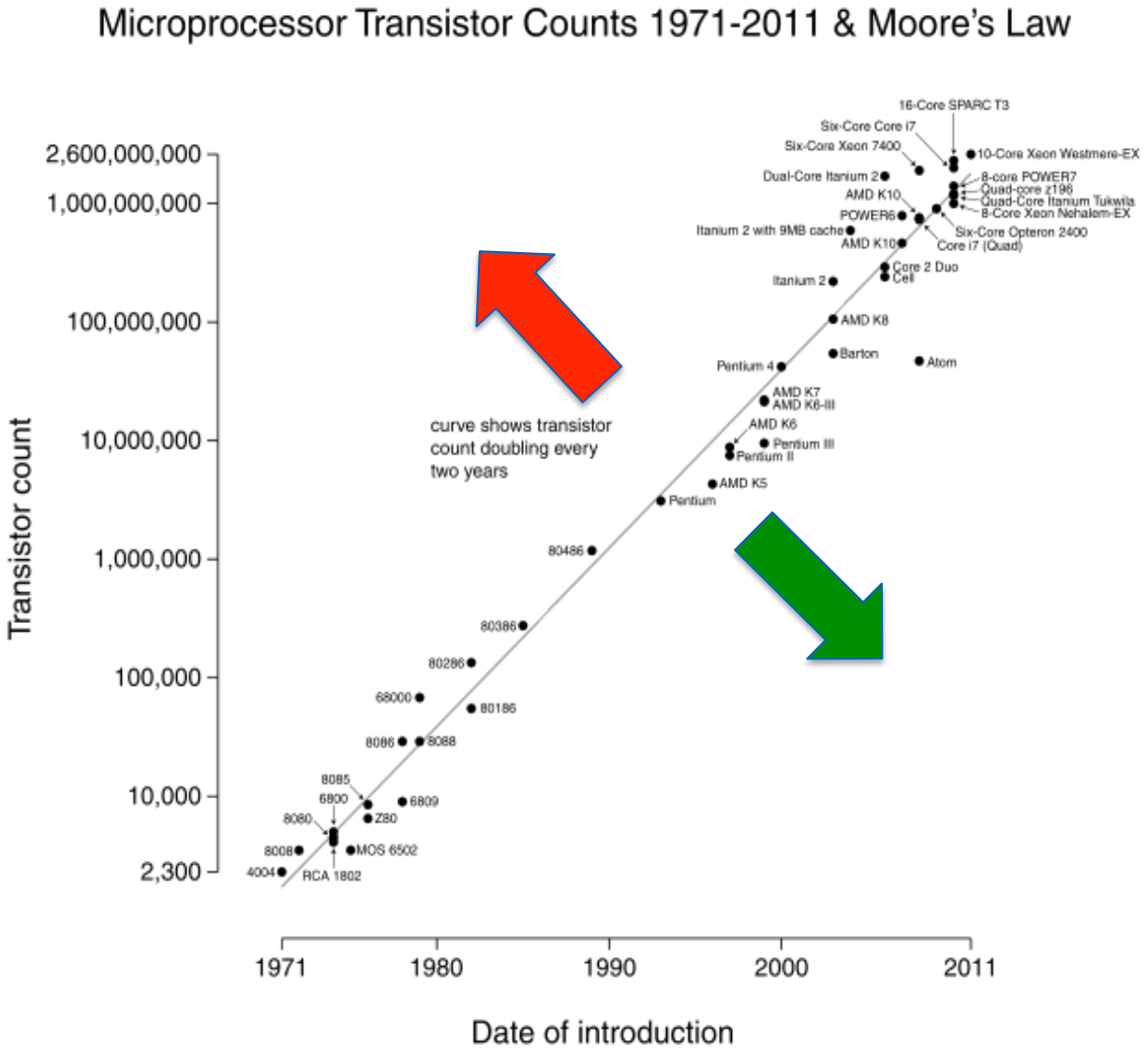
* *These numbers are dubious due to uncertainties introduced by IPv4 address exhaustion pressures.*

Up and to the Right == Pain?

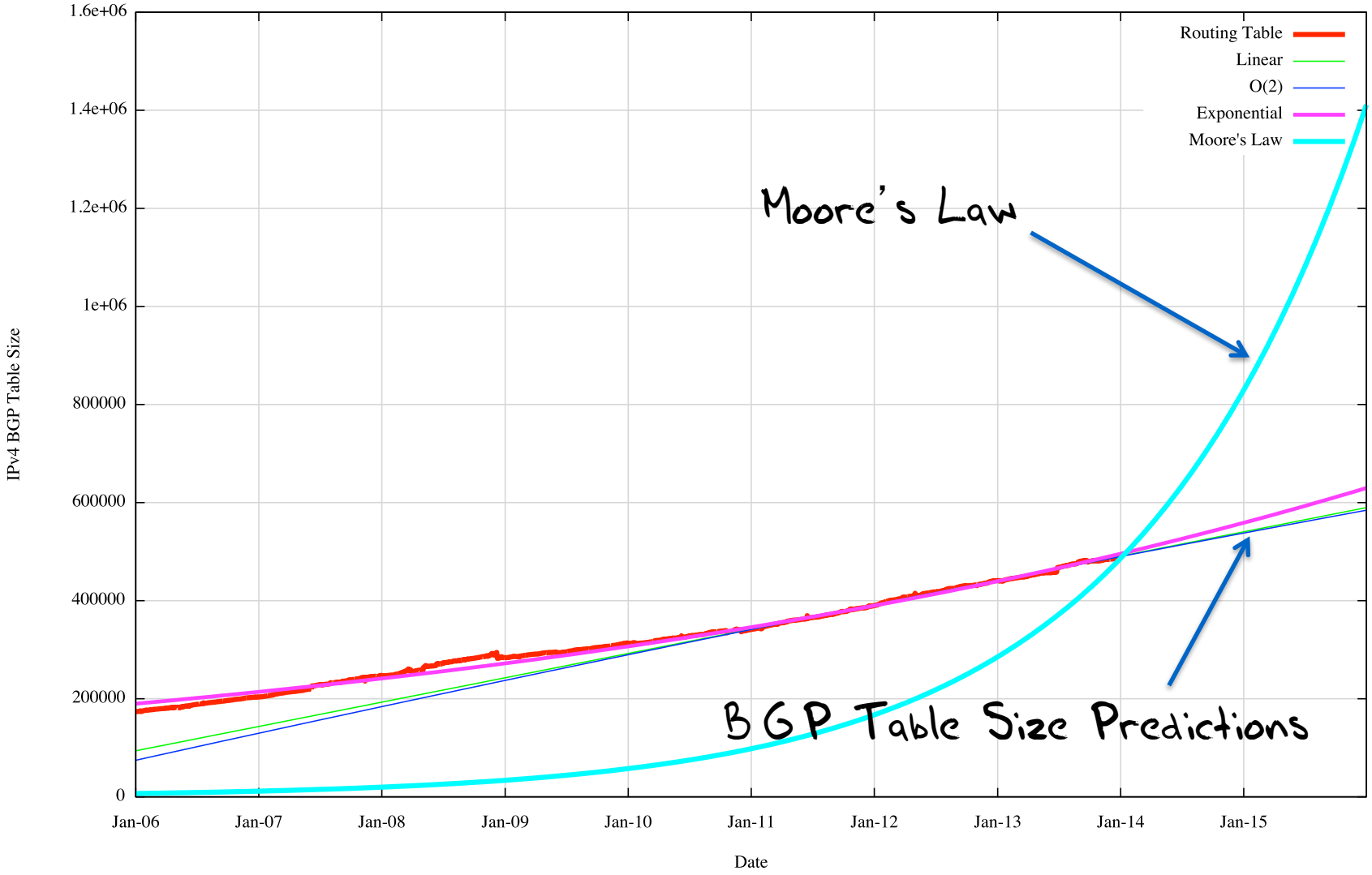
- Most Internet curves are “up and to the right”
- But what makes this curve painful?
  - The pain threshold is approximated by Moore’s Law

Moore's Law

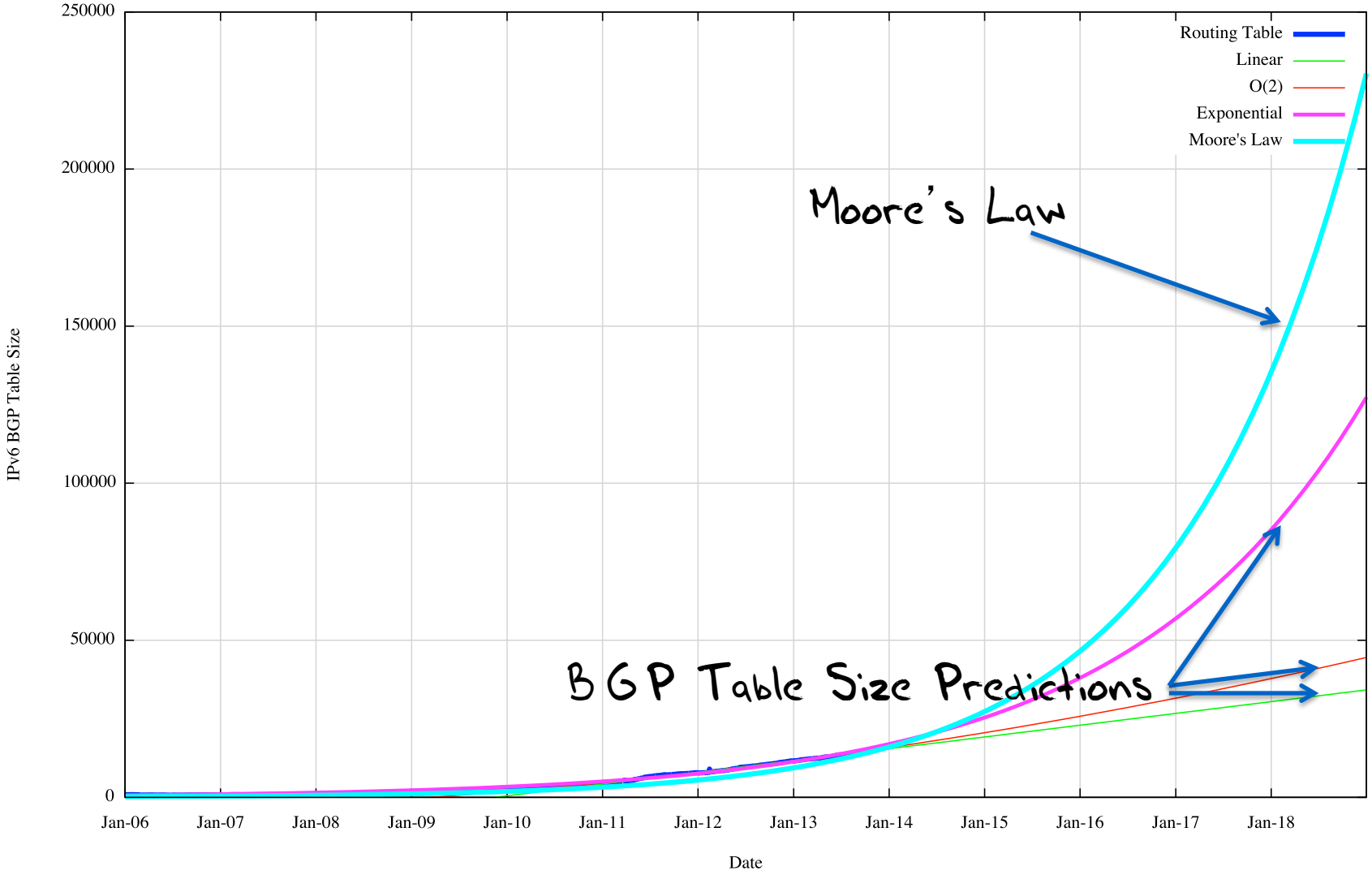
- As a rough rule of thumb, if the rate of growth of the table grows at a rate equal to, or less than Moore's Law, then the unit cost of storing the forwarding table should remain constant
  - Like all rough rules of thumb, there are many potential exceptions, and costs have many inputs as well as the raw cost of the the number of gates in a chip
  - Despite this, Moore's Law still a useful benchmark of a threshold of concern about routing growth


Microprocessor Transistor Counts 1971-2011 & Moore's Law
Transistor count
2,600,000,000
1,000,000,000
100,000,000
10,000,000
1,000,000
100,000
10,000
2,300
1971
1980
1990
2000
2011
Date of introduction
curve shows transistor count doubling every two years
16-Core SPARC T3
Six-Core Core i7
Six-Core Xeon 7400
10-Core Xeon Westmere-EX
Dual-Core Itanium 2
8-core POWER7
AMD K10
Quad-core z196
Quad-Core Itanium Tukwila
POWER6
8-Core Xeon Nehalem-EX
Itanium 2 with 9MB cache
Six-Core Opteron 2400
AMD K10
Core i7 (Quad)
Core 2 Duo
Cell
Itanium 2
AMD K8
Barton
Atom
Pentium 4
AMD K7
AMD K6-III
AMD K6
Pentium III
Pentium II
AMD K5
Pentium
80486
80386
80286
68000
80186
8086
8088
8085
6800
6809
8080
Z80
8008
MOS 6502
4004
RCA 1802

IPv4 BGP Table size and Moore's Law



IPv6 Projections and Moore's Law



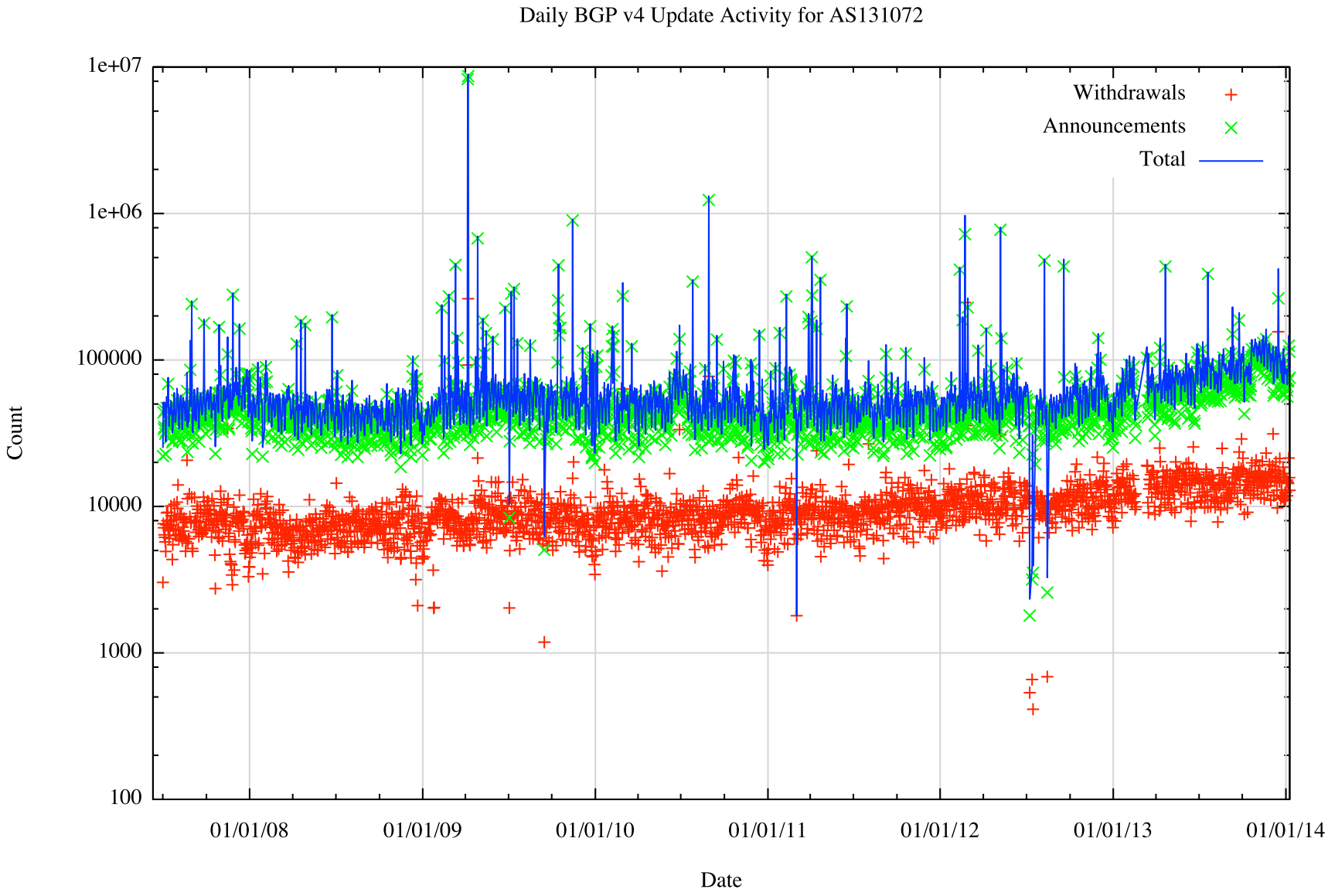
BGP Table Growth

- Nothing in these figures suggests that there is cause for urgent alarm -- at present
- The overall BGP growth rates for IPv4 are holding at a modest level, and the IPv6 table, although it is growing rapidly, is still relatively small in size in absolute terms
- As long as we are prepared to live within the technical constraints of the current routing paradigm the routing table size will continue to be viable for some time yet

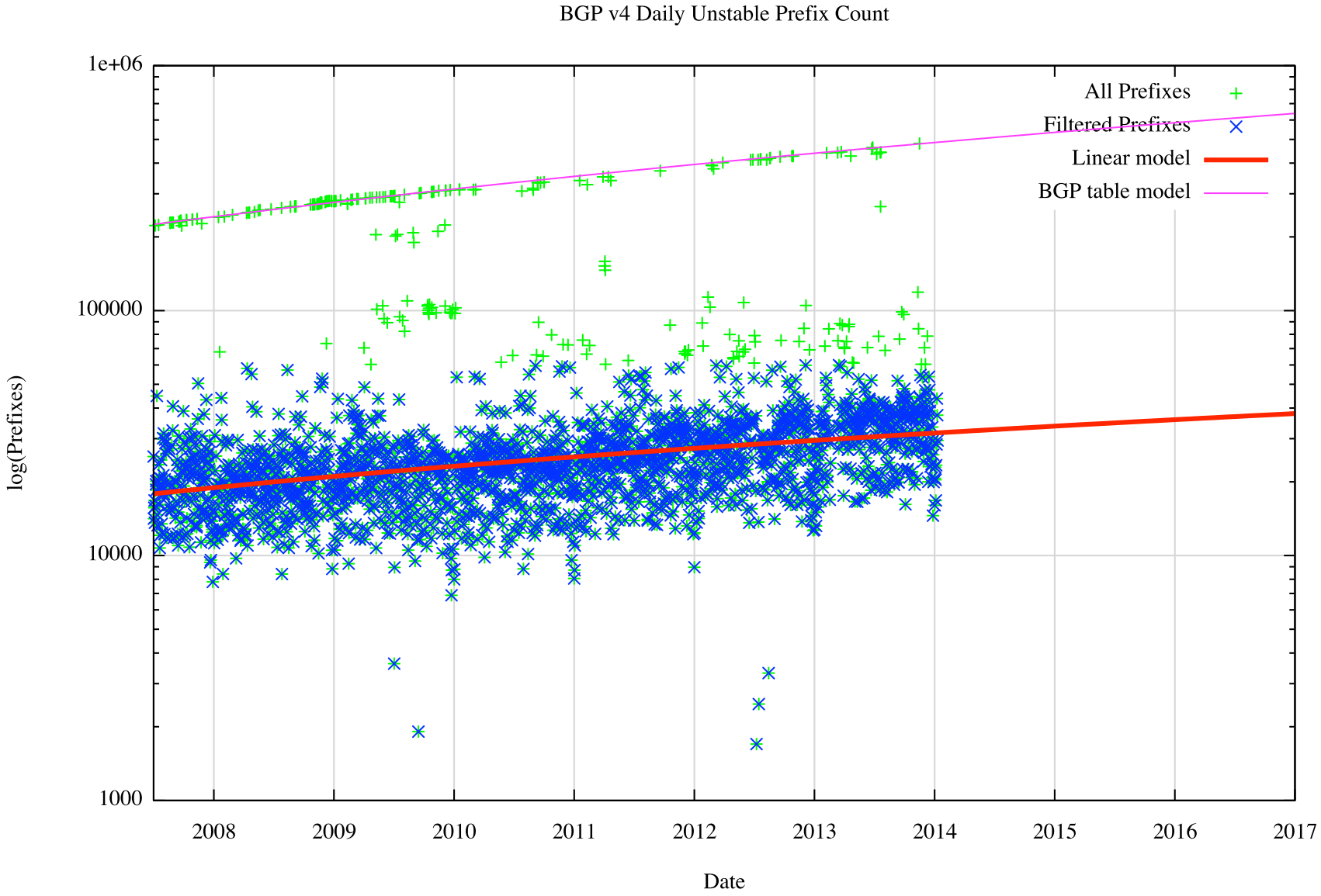
BGP Updates

- What about the level of updates in BGP?
- Let's look at the update load from a single eBGP feed in a DFZ context

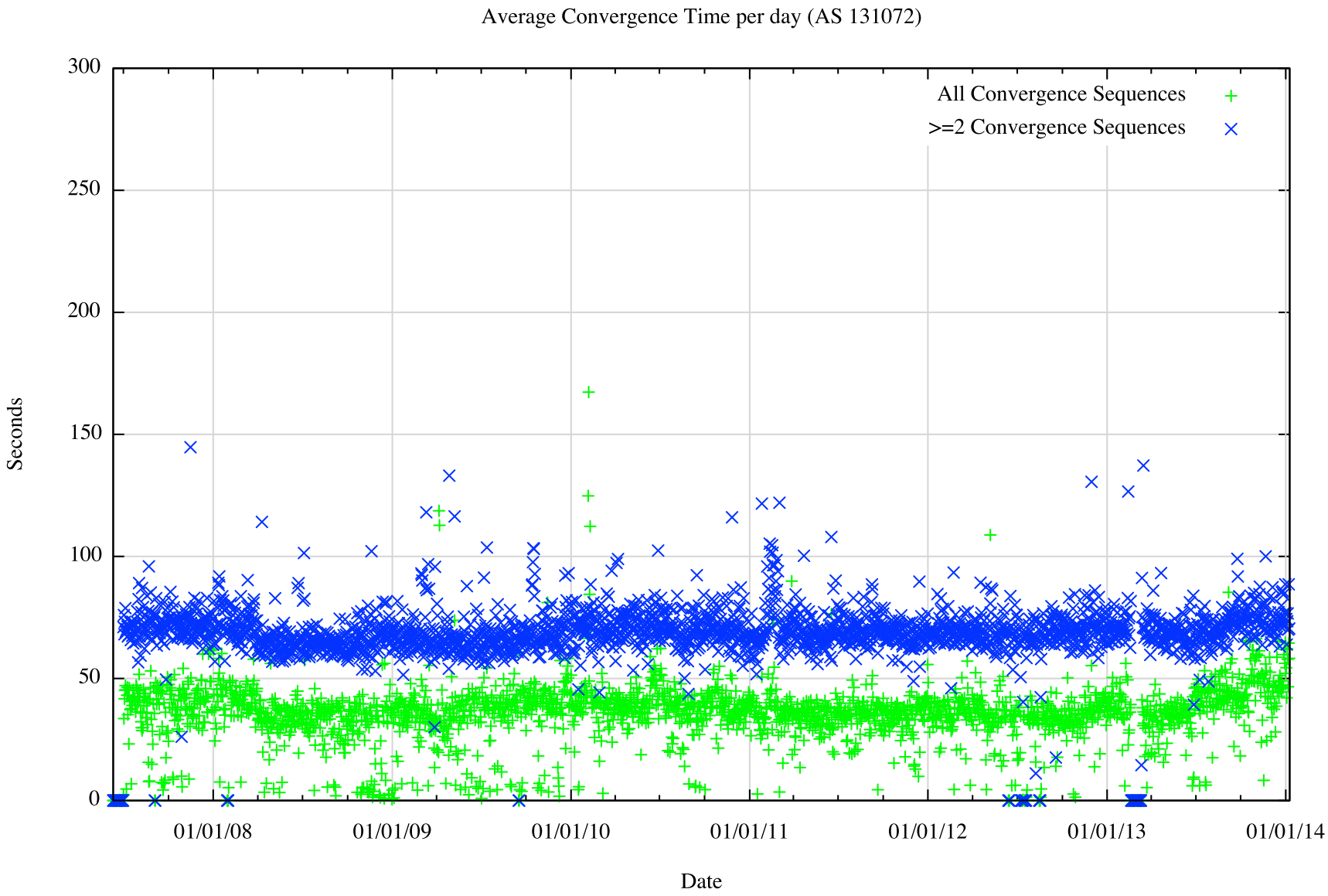
Announcements and Withdrawals



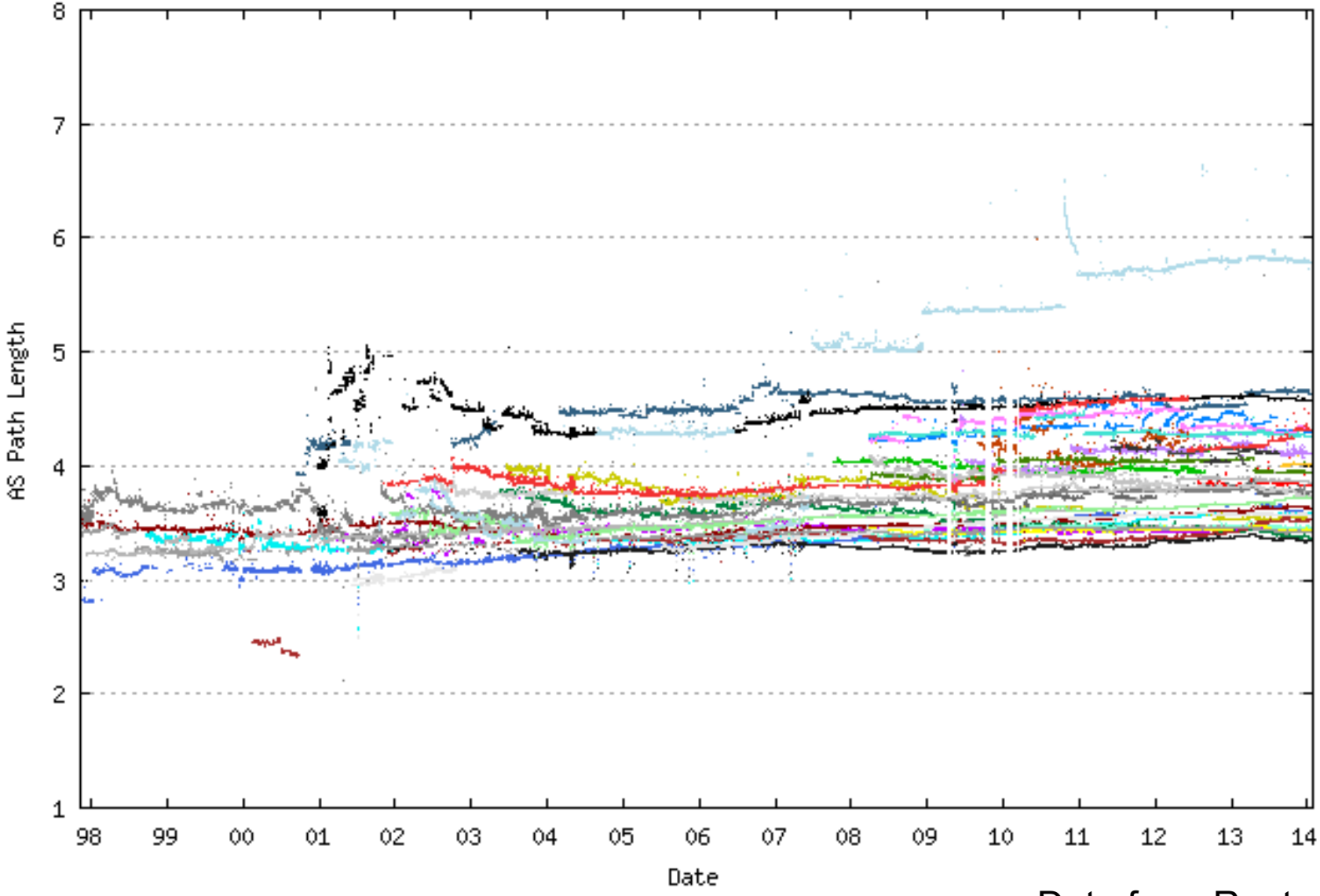
Unstable Prefixes



Convergence Performance



IPv4 Average AS Path Length



Data from Route Views

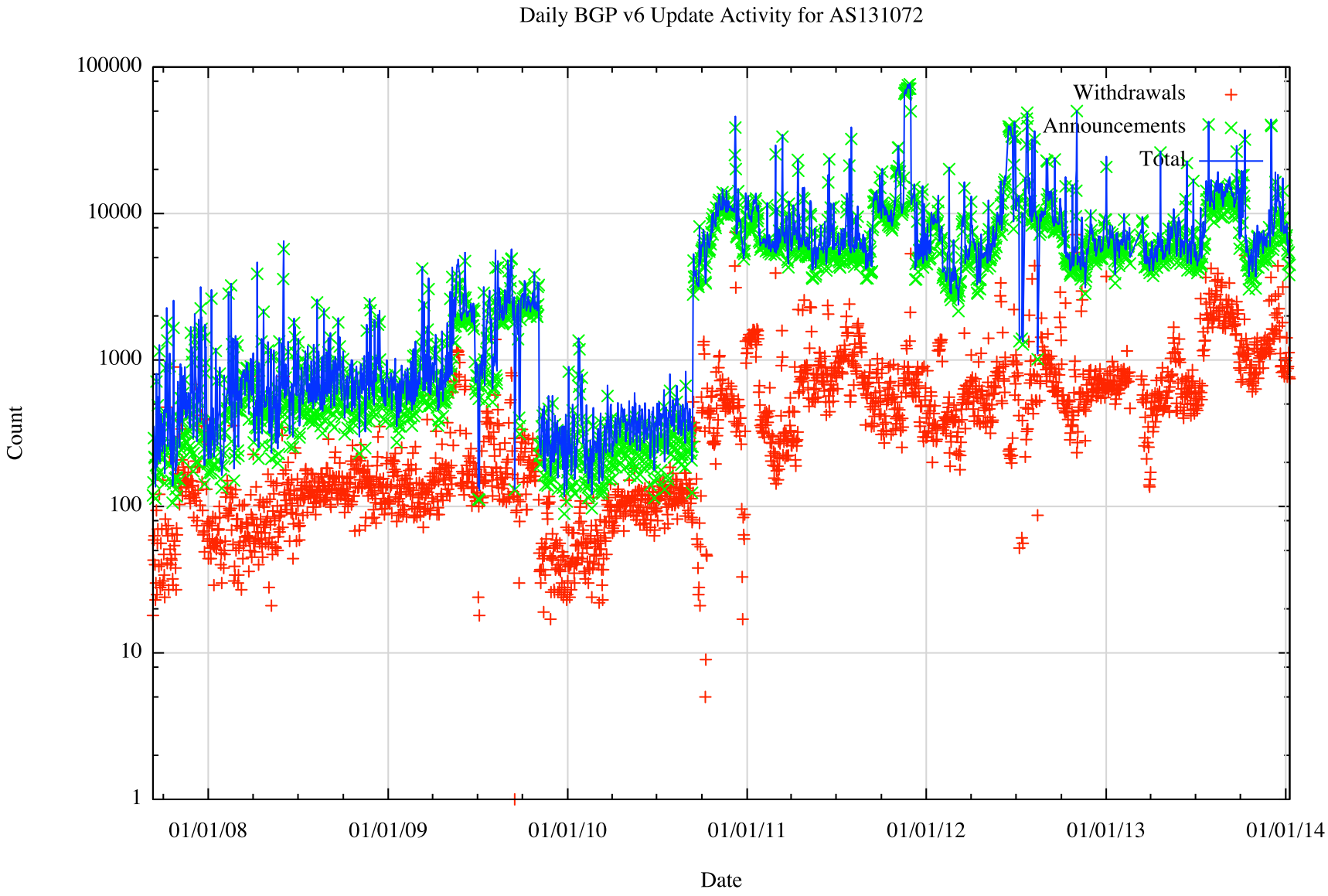
Updates in IPv4 BGP

Nothing in these figures is cause for any great level of concern …

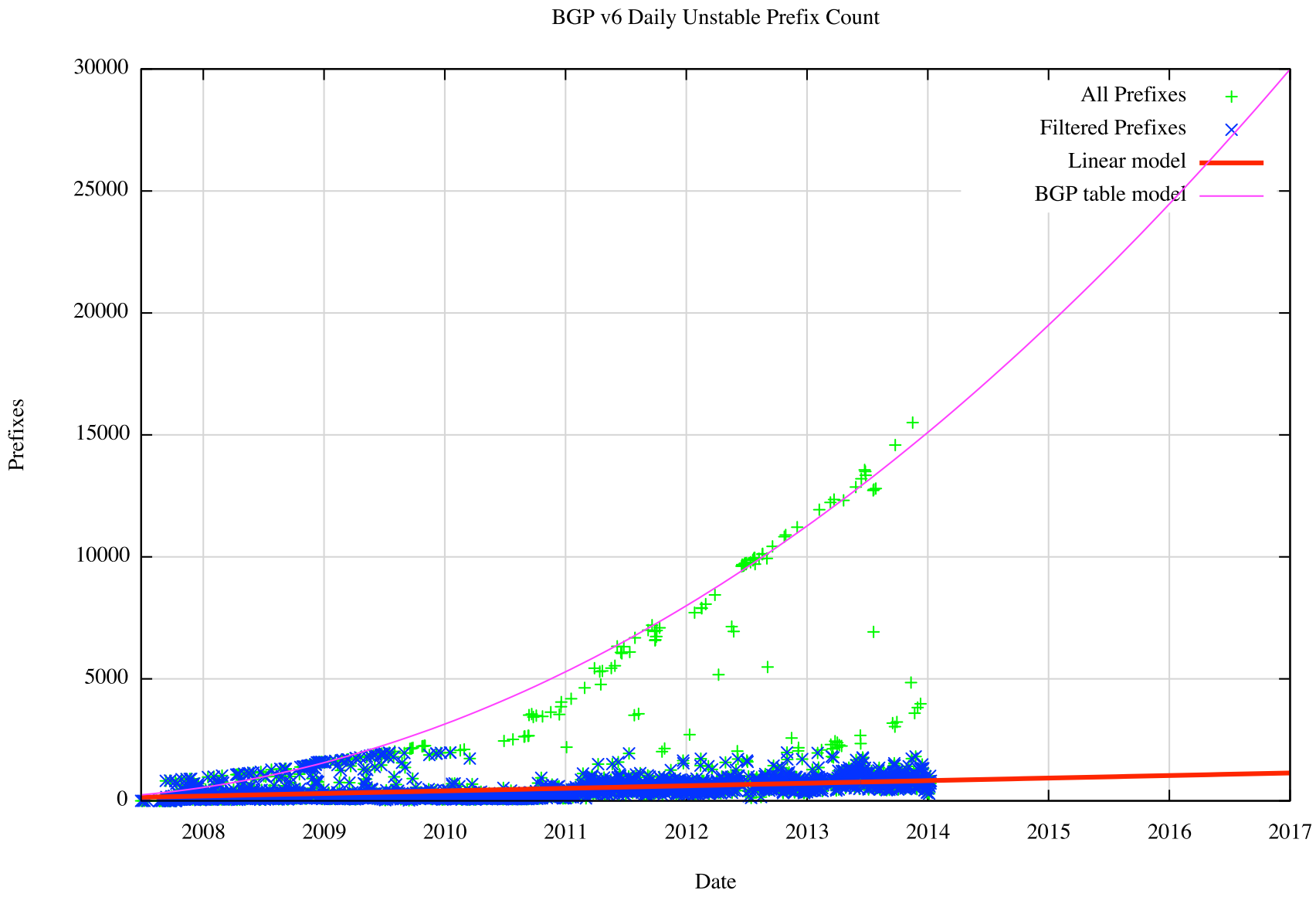
- The number of unstable prefixes per day is growing at a far lower rate than the number of announced prefixes
- The number of updates per instability event has been constant, due to the damping effect of the MRAI interval, and the relatively constant AS Path length over this interval
- As long as the average AS Path does not break out, BGP will continue to scale in terms of convergence properties irrespective of the number of announced objects

What about IPv6?

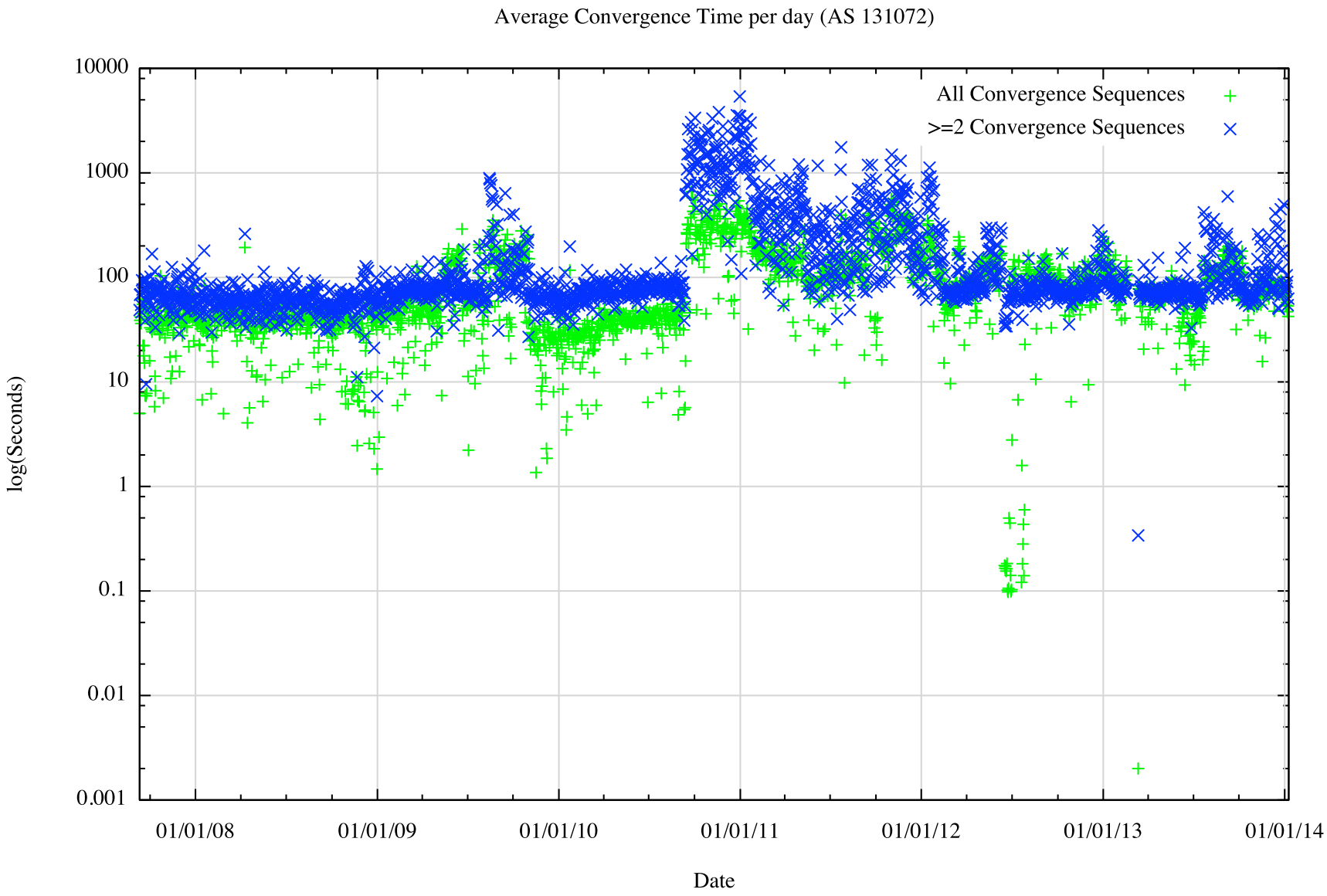
IPv6 Announcements and Withdrawals



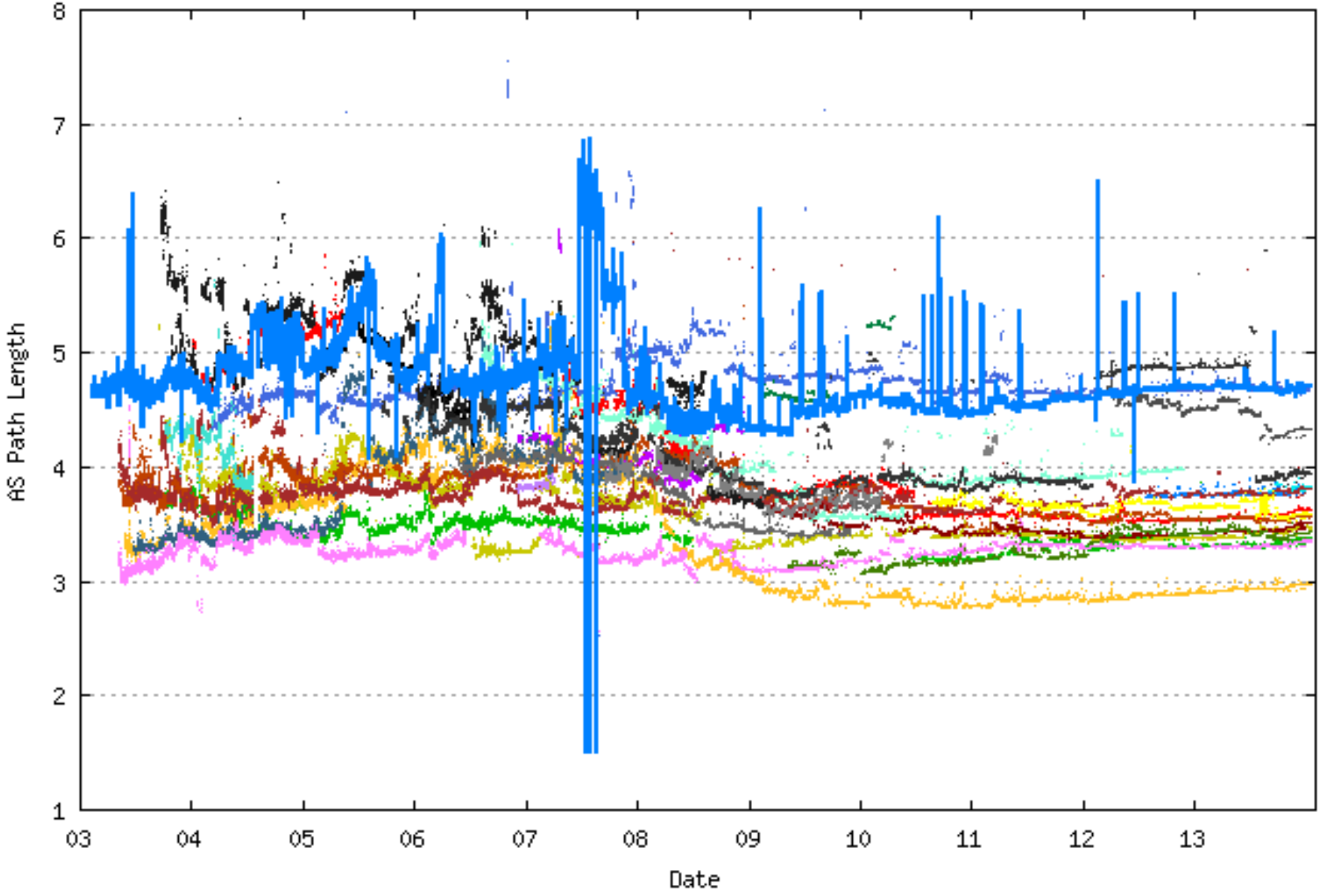
IPv6 Unstable Prefixes



IPv6 Convergence Performance



IPv6 Average AS Path Length



Data from Route Views

BGP Convergence

- The long term average convergence time for the IPv4 BGP network is some 70 seconds, or 2.3 updates given a 30 second MRAI timer
- The long term average convergence time for the IPv6 BGP network is some 90 seconds, or 3 updates
- The average AS Path appears to be stable, and the convergence performance is therefore stable

BGP Table Growth

However ... continued scalability of the routing system relies on continued conservatism in routing practices.

How good are we at “being conservative” in routing?

CIDR and BGP

- To what extent do we still practice “conservative” routing and refrain from announcing more specifics into the routing table?
- Are we getting better or worse at aggregation in routing?
- What is the distribution of advertising more specifics? Are we seeing a significant increase in the number of more specific /24s in the routing table?

An Example:

```
Prefix              AS Path
193.124.0.0/15      4608 1221 4637 3356 20485 2118 ?
193.124.0.0/24      4608 1221 4637 3356 20485 2118 ?
193.124.1.0/24      4608 1221 4637 3356 20485 2118 ?
193.124.2.0/24      4608 1221 4637 3356 20485 2118 ?
193.124.3.0/24      4608 1221 4637 3356 20485 2118 ?
193.124.4.0/24      4608 1221 4637 3356 20485 2118 ?
193.124.5.0/24      4608 1221 4637 3356 20485 2118 ?
193.124.6.0/24      4608 1221 4637 3356 20485 2118 ?
193.124.7.0/24      4608 1221 4637 3356 20485 2118 ?
193.124.8.0/24      4608 1221 4637 3356 20485 2118 ?
193.124.9.0/24      4608 1221 4637 3356 20485 2118 ?
193.124.10.0/24     4608 1221 4637 3356 20485 2118 ?
193.124.11.0/24     4608 1221 4637 3356 20485 2118 ?
193.124.12.0/24     4608 1221 4637 3356 20485 2118 ?
193.124.13.0/24     4608 1221 4637 3356 20485 2118 ?
193.124.14.0/24     4608 1221 4637 3356 20485 2118 ?
193.124.15.0/24     4608 1221 4637 3356 20485 2118 ?
```

Origin AS: AS 2118 RELCOM-AS OOO "NPO Relcom"

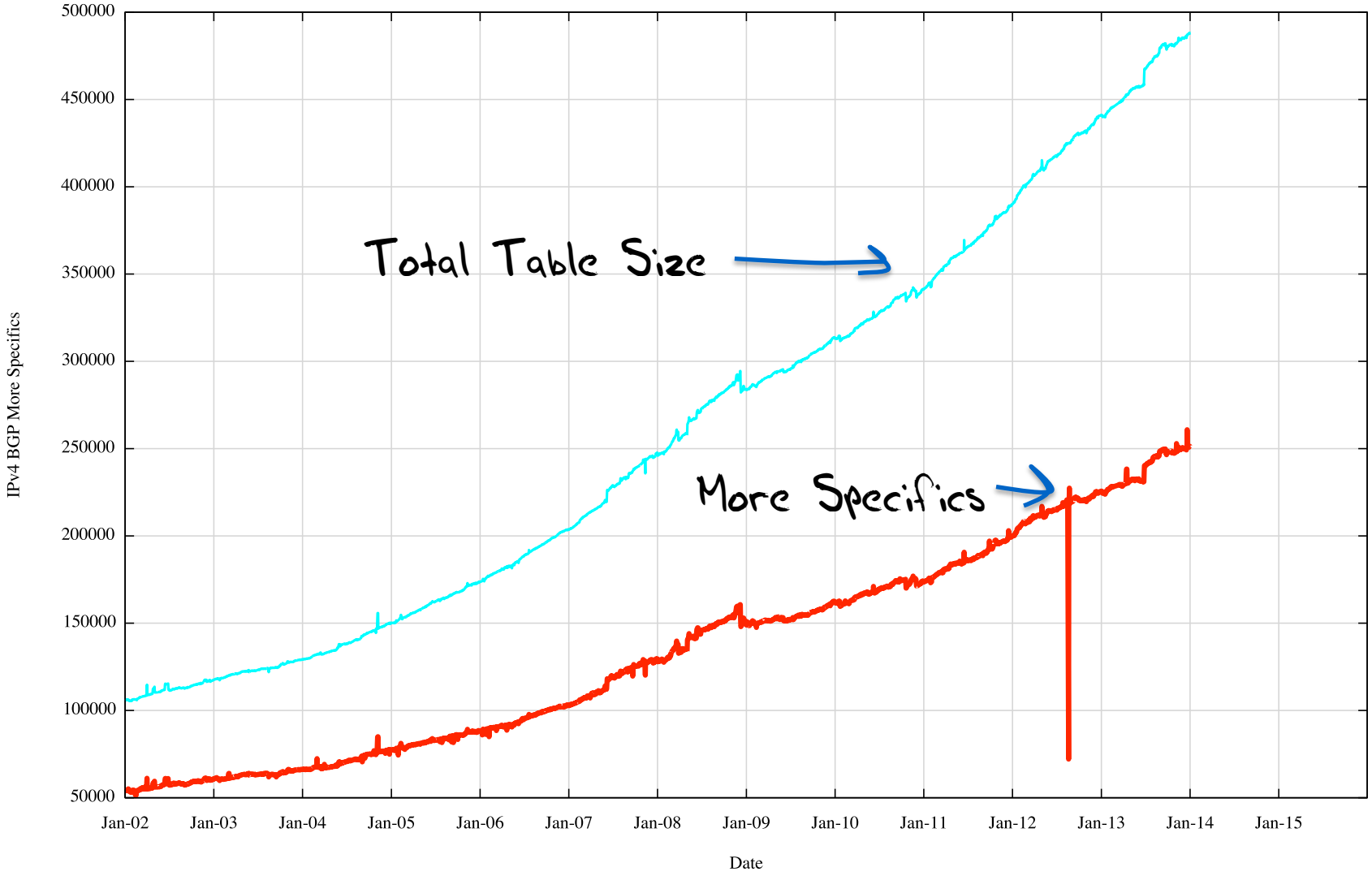
Who is doing this the most?

www.cidr-report.org

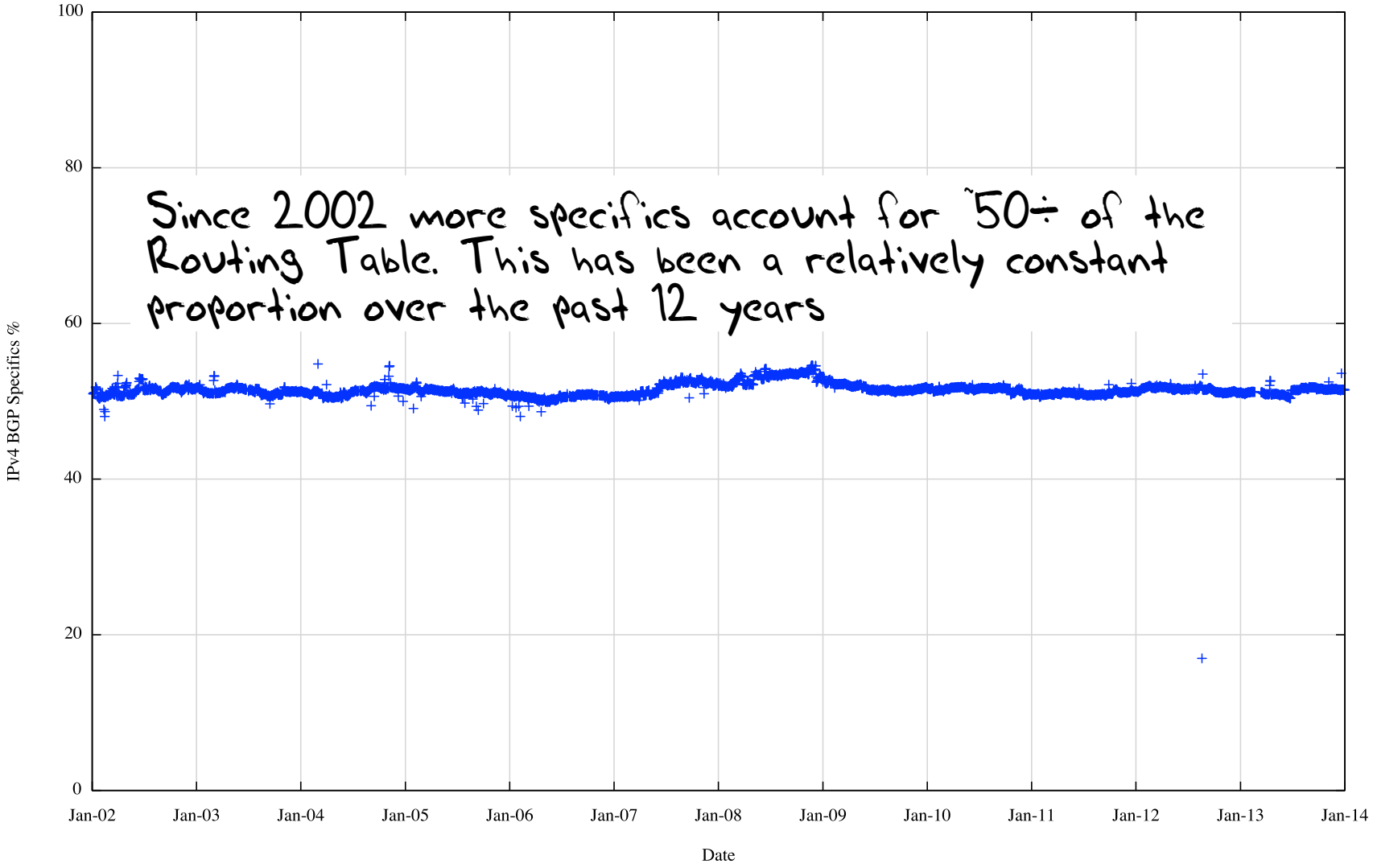
--- 14Jan14 ---

ASnum	NetsNow	NetsAggr	NetGain	% Gain	Description
Table	488946	273762	215184	44.0%	All ASes
AS28573	3447	91	3356	97.4%	NET Serviços de Comunicação S.A.
AS6389	3029	56	2973	98.2%	BELLSOUTH-NET-BLK - BellSouth.net Inc.
AS7029	4427	1657	2770	62.6%	WINDSTREAM - Windstream Communications Inc
AS17974	2735	184	2551	93.3%	TELKOMNET-AS2-AP PT Telekomunikasi Indonesia
AS22773	2326	160	2166	93.1%	ASN-CXA-ALL-CCI-22773-RDC - Cox Communications Inc.
AS4766	2944	962	1982	67.3%	KIXS-AS-KR Korea Telecom
AS18881	1796	32	1764	98.2%	Global Village Telecom
AS36998	1805	47	1758	97.4%	SDN-MOBITEL
AS1785	2149	392	1757	81.8%	AS-PAETEC-NET - PaeTec Communications, Inc.
AS10620	2696	1084	1612	59.8%	Telmex Colombia S.A.
AS18566	2048	565	1483	72.4%	MEGAPATH5-US - MegaPath Corporation
AS4323	2935	1509	1426	48.6%	TWTC - tw telecom holdings, inc.
AS7303	1744	451	1293	74.1%	Telecom Argentina S.A.
AS4755	1811	594	1217	67.2%	TATACOMM-AS TATA Communications formerly VSNL is Leading ISP
AS7552	1258	159	1099	87.4%	VIETEL-AS-AP Viettel Corporation
AS22561	1259	226	1033	82.0%	AS22561 - CenturyTel Internet Holdings, Inc.
AS9829	1567	691	876	55.9%	BSNL-NIB National Internet Backbone
AS7545	2135	1309	826	38.7%	TPG-INTERNET-AP TPG Telecom Limited
AS18101	987	184	803	81.4%	RELIANCE-COMMUNICATIONS-IN Reliance Communications Ltd.DAKC MUMBAI

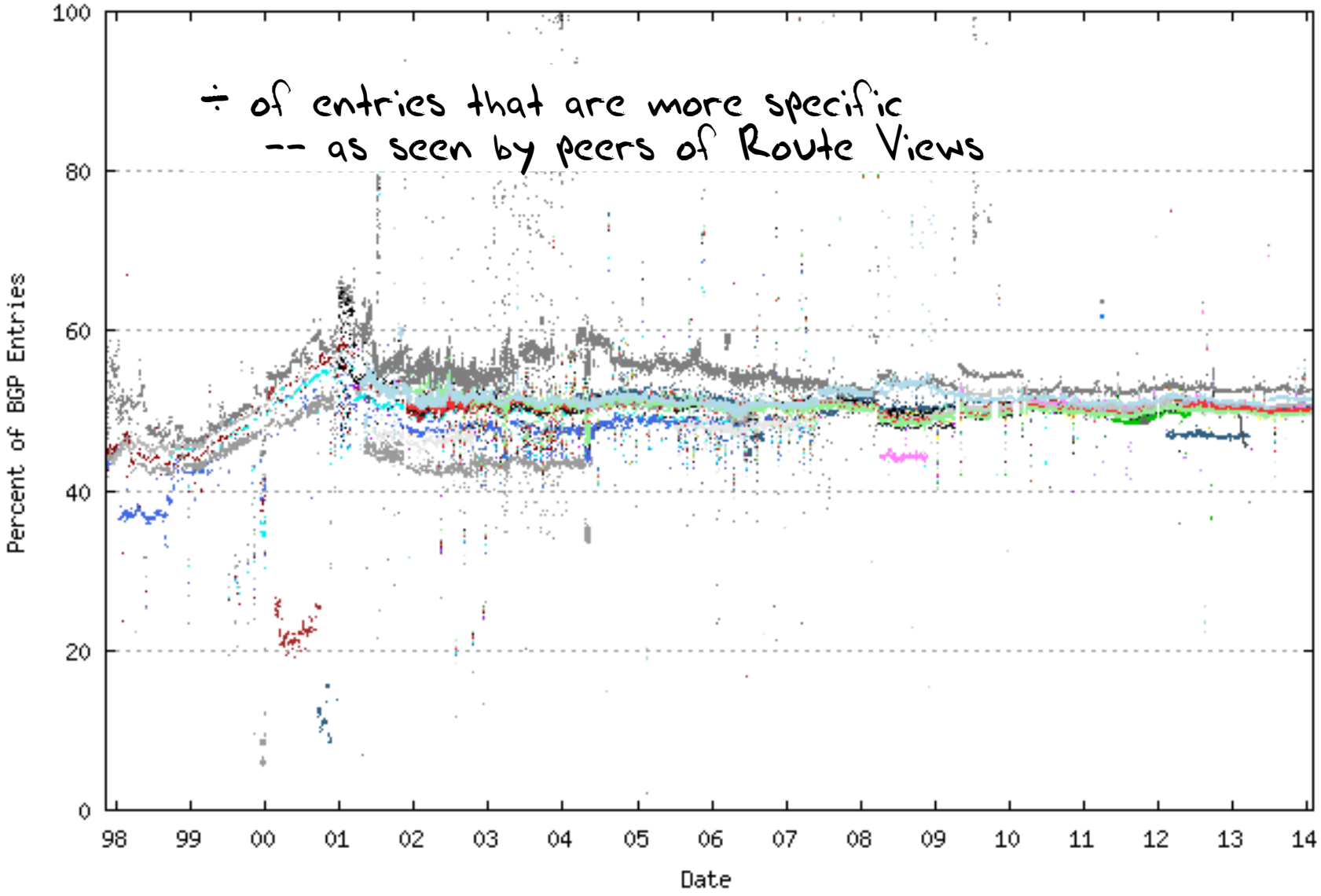
More specifics in the Routing Table



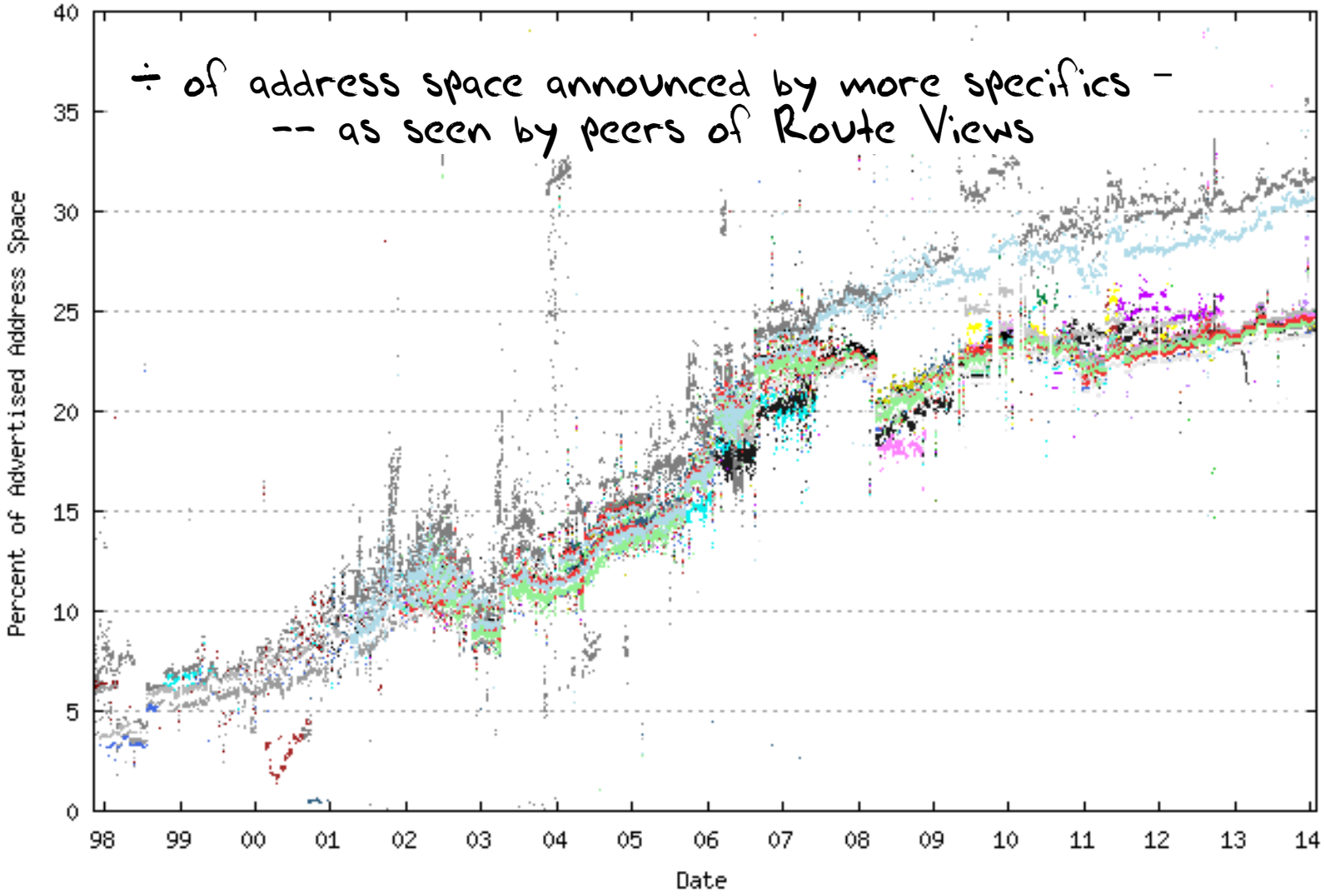
More specifics in the Routing Table



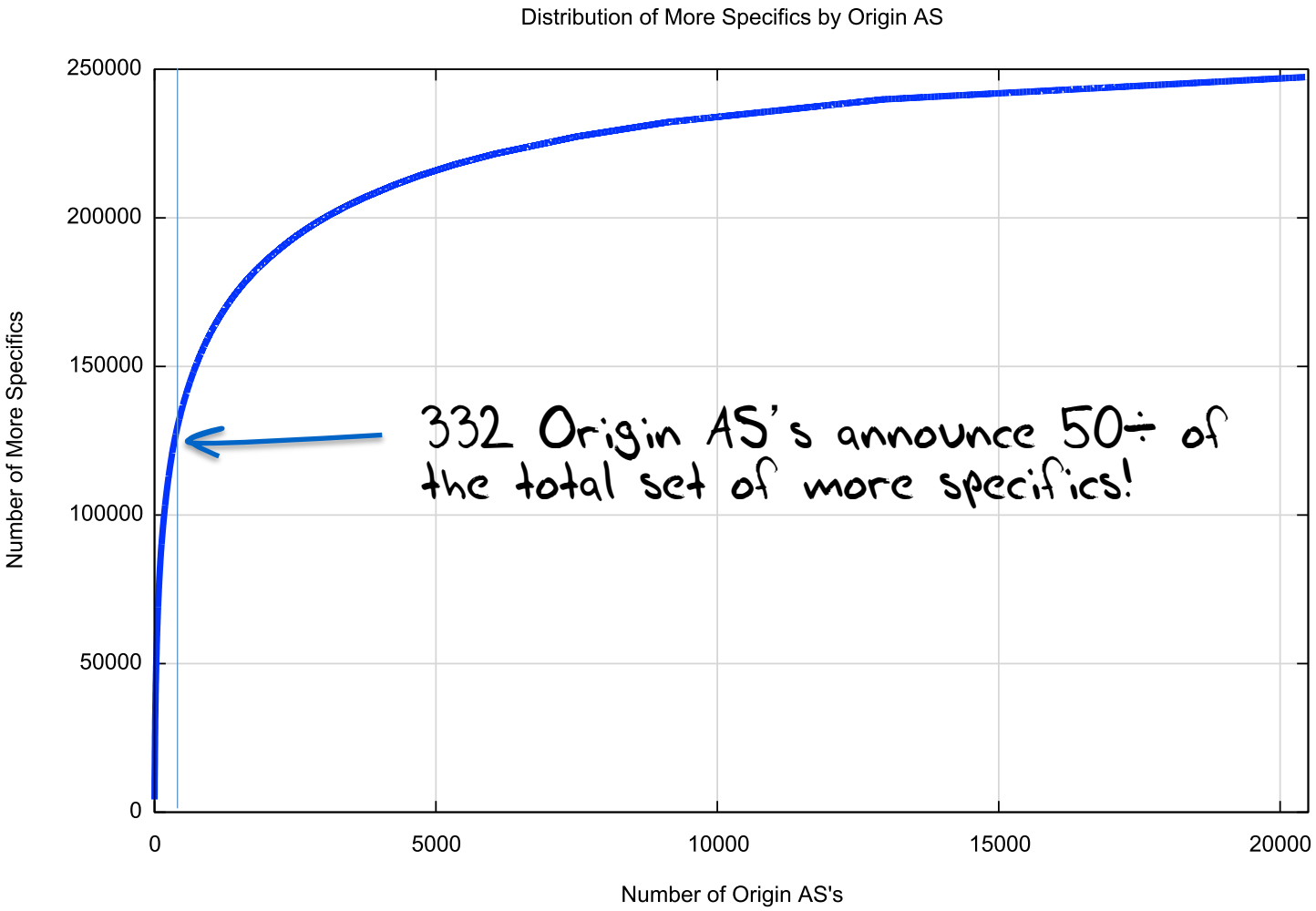
Does everyone see this?



How much address space is announced by more specifics?



Does everyone announce more specifics?



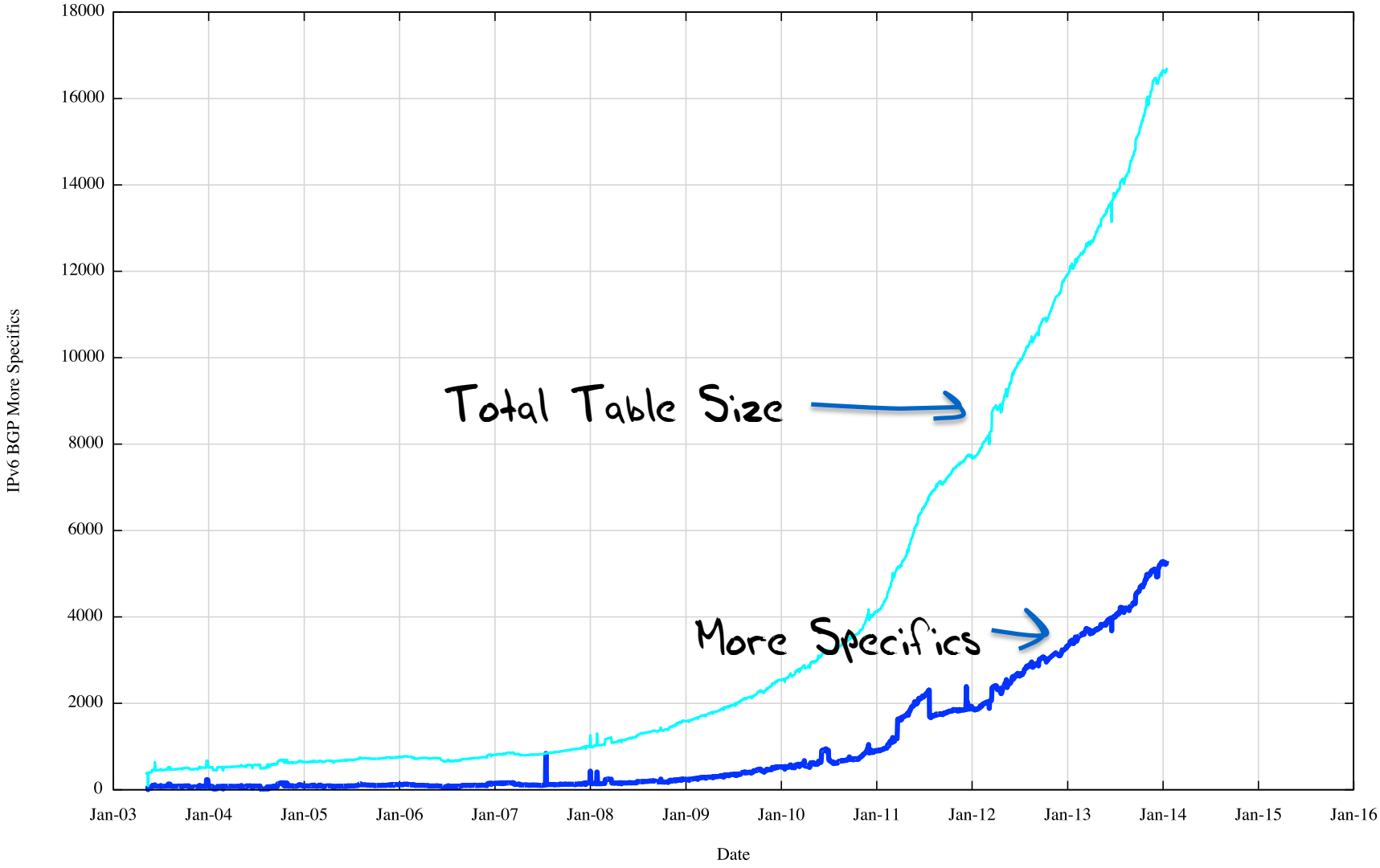
Is it Everyone?

- 1% of the ASes (458 ASes) announce 54% of the more specifics (133,688 announcements)
- 55% of the ASes announce **no** more specifics
- The top 20 ASes announce 40,404 more specifics

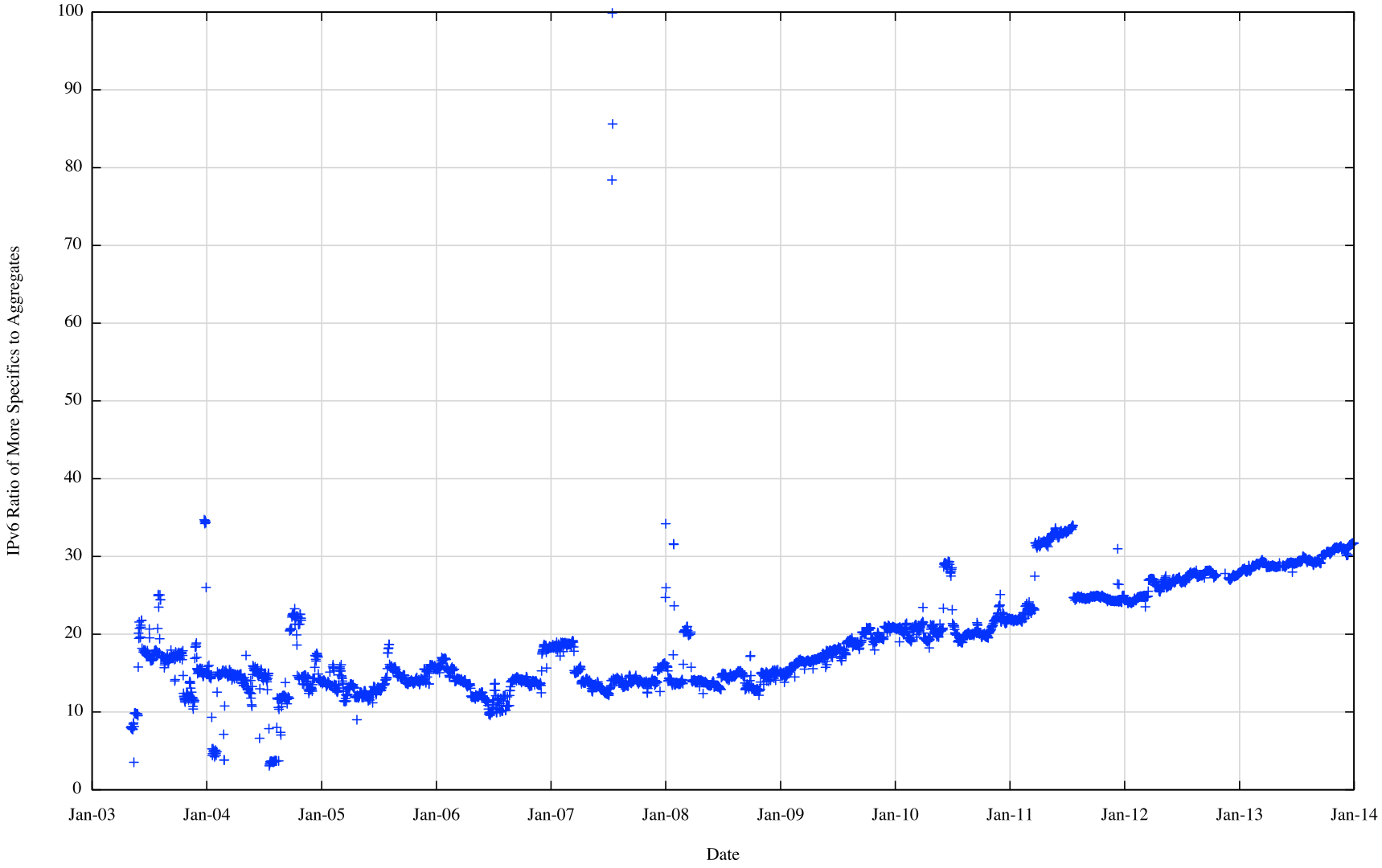
The Top 20 of V4 More Specifics

AS	Agg's	More Specifics	
7029	148	4,275	WINDSTREAM - Windstream Communications Inc US
6389	50	2,979	BELLSOUTH-NET-BLK - BellSouth.net Inc. US
28573	685	2,727	NET Servicos de Comunicatio S.A. BR
17974	224	2,511	TELKOMNET-AS2-AP PT Telekomunikasi Indonesia ID
4323	449	2,486	TWTC - tw telecom holdings, inc. US
22773	159	2,167	ASN-CXA-ALL-CCI-22773-RDC - Cox Communications Inc. US
1785	121	2,030	AS-PAETEC-NET - PaeTec Communications, Inc. US
18566	19	2,029	MEGAPATH5-US - MegaPath Corporation US
7545	113	2,023	TPG-INTERNET-AP TPG Telecom Limited AU
36998	5	1,800	SDN-MOBITEL SD
18881	22	1,774	Global Village Telecom BR
8402	14	1,726	CORBINA-AS OJSC "Vimpelcom" RU
10620	1,035	1,661	Telmex Colombia S.A. CO
4755	179	1,632	TATACOMM-AS TATA Communications formerly VSNL is Leading ISP IN
4766	564	1,591	KIXS-AS-KR Korea Telecom KR
7552	26	1,232	VIETEL-AS-AP Viettel Corporation VN
9829	371	1,189	BSNL-NIB National Internet Backbone IN
7011	8	1,164	FRONTIER-AND-CITIZENS - Frontier Communications of America, Inc. US
9498	53	1,159	BBIL-AP BHARTI Airtel Ltd. IN
5617	36	1,150	TPNET Telekomunikacja Polska S.A. PL
20940	119	1,099	AKAMAI-ASN1 Akamai International B.V. US

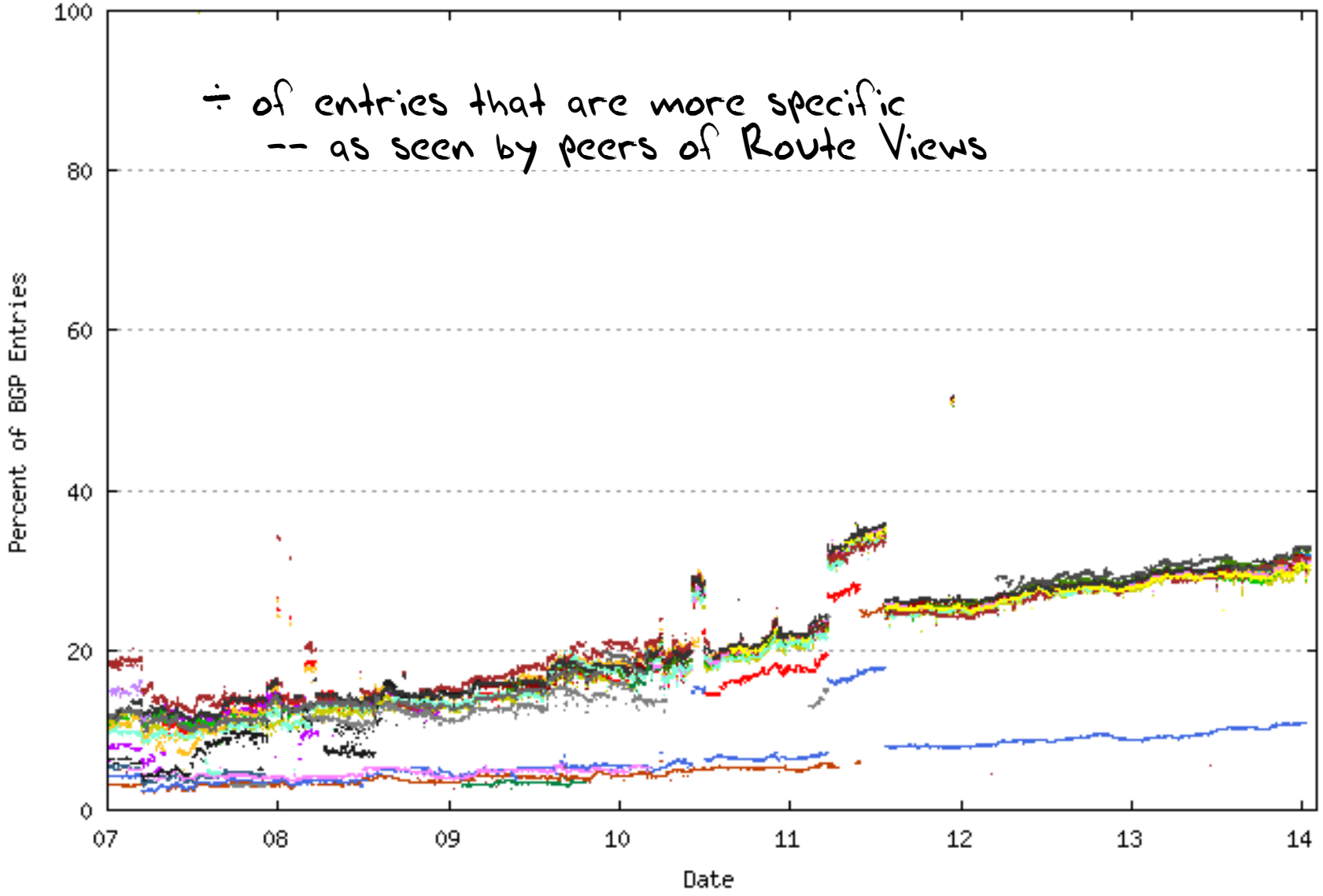
More specifics in the V6 Routing Table



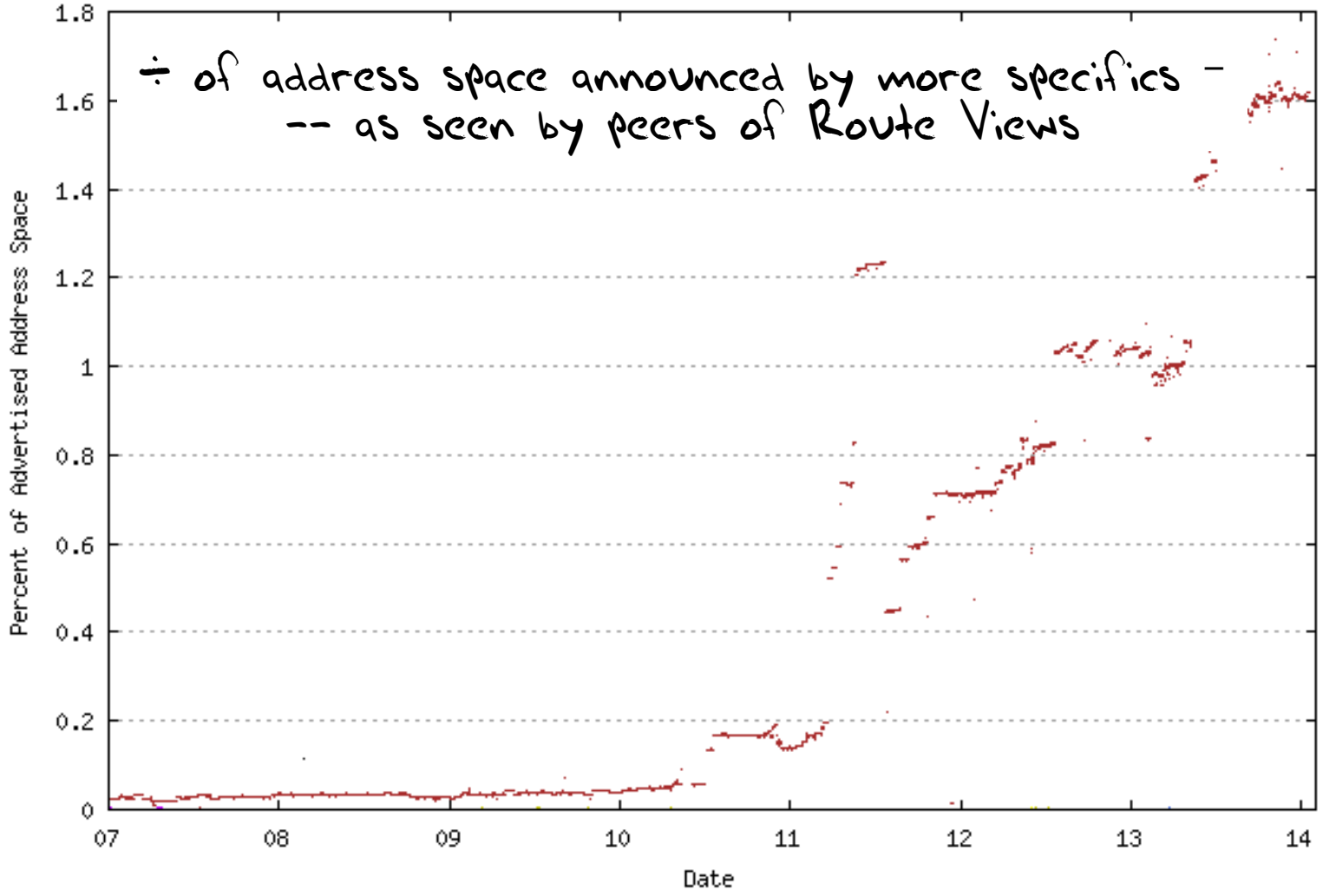
More specifics in the V6 Routing Table



Does everyone see this?



How much V6 address space is announced by more specifics?



Are We Getting Any Better?

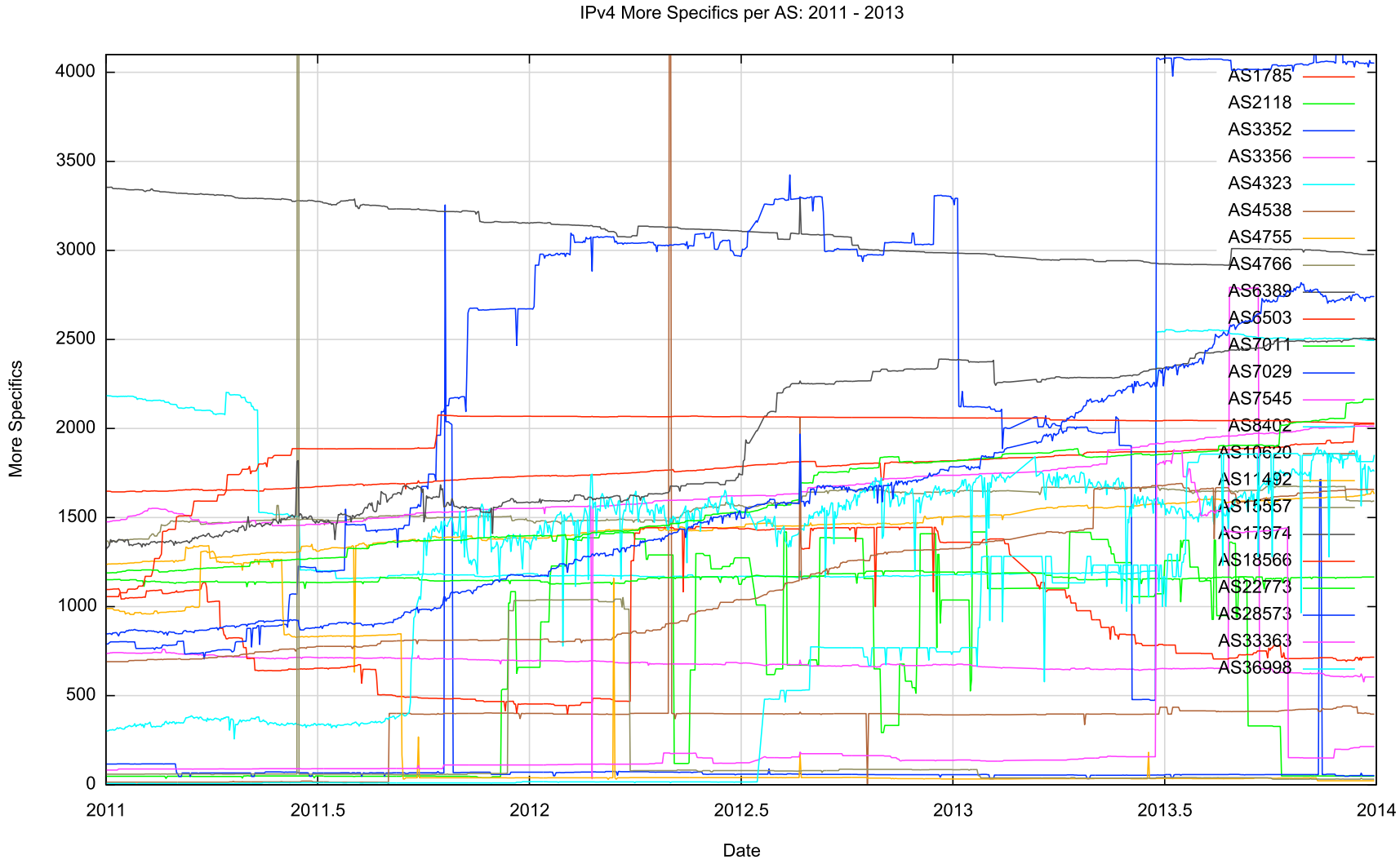
Take the daily top 10 ASes of advertisers of more specifics over the past 3 years and track the number of more specifics advertised by these ASes over the entire period

Are We Getting any Better?

AS's seen to be advertising the highest number of more specifics over the past 3 years:

1785	AS-PAETEC-NET - PaeTec Communications, Inc. US
2118	RELCOM-AS OOO "NPO Relcom" RU
3352	TELEFONICA-DATA-ESPANA TELEFONICA DE ESPANA ES
3356	LEVEL3 Level 3 Communications US
4323	TWTC - tw telecom holdings, inc. US
4538	ERX-CERNET-BKB China Education and Research Network Center CN
4755	TATACOMM-AS TATA Communications formerly VSNL is Leading ISP IN
4766	KIXS-AS-KR Korea Telecom KR
6389	BELLSOUTH-NET-BLK - BellSouth.net Inc. US
6503	Axtel, S.A.B. de C.V. MX
7011	FRONTIER-AND-CITIZENS - Frontier Communications of America, Inc. US
7029	WINDSTREAM - Windstream Communications Inc US
7545	TPG-INTERNET-AP TPG Telecom Limited AU
8402	CORBINA-AS OJSC "Vimpelcom" RU
10620	Telmex Colombia S.A. CO
11492	CABLEONE - CABLE ONE, INC. US
15557	LDCOMNET Societe Francaise du Radiotelephone S.A FR
17974	TELKOMNET-AS2-AP PT Telekomunikasi Indonesia ID
18566	MEGAPATH5-US - MegaPath Corporation US
22773	ASN-CXA-ALL-CCI-22773-RDC - Cox Communications Inc. US
28573	NET Servicos de Comunicatio S.A. BR
33363	BHN-TAMPA - BRIGHT HOUSE NETWORKS, LLC US
36998	SDN-MOBITEL SD

Yes ... and No


IPv4 More Specifics per AS: 2011 - 2013
More Specifics
Date
0
500
1000
1500
2000
2500
3000
3500
4000
2011
2011.5
2012
2012.5
2013
2013.5
2014
AS1785
AS2118
AS3352
AS3356
AS4323
AS4538
AS4755
AS4766
AS6389
AS6503
AS7011
AS7029
AS7545
AS8402
AS10620
AS11492
AS15557
AS17974
AS18566
AS22773
AS28573
AS33363
AS36998

Are We Getting Any Better?

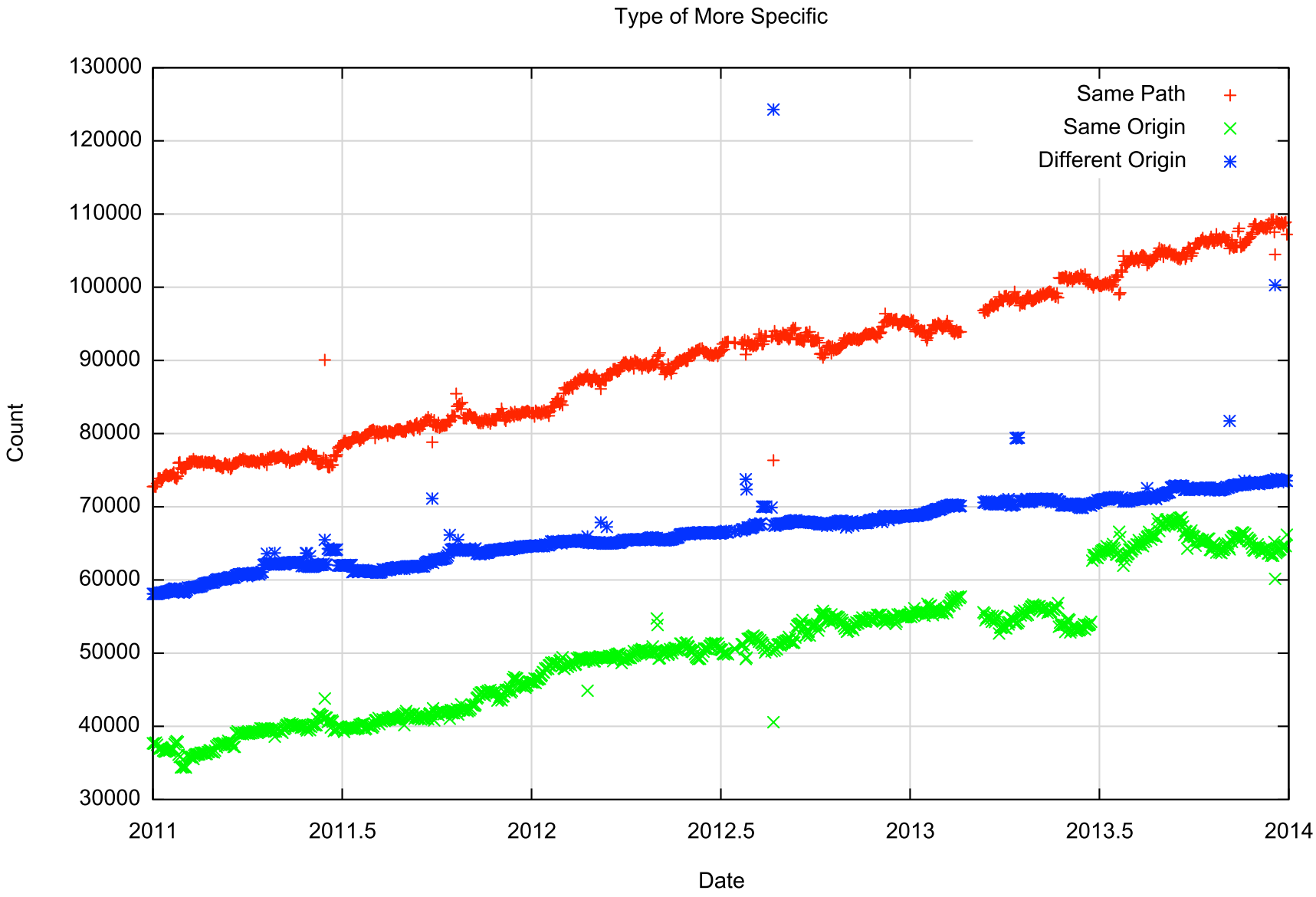
- Some ASes are effectively reducing the number of more specifics that are advertised into the global routing system
- Some ASes are increasing the number of more specifics
- And some are consistently advertising a significant number of more specifics
- There is no net change in the overall distribution and characteristics of more specifics in the routing system.

Why?

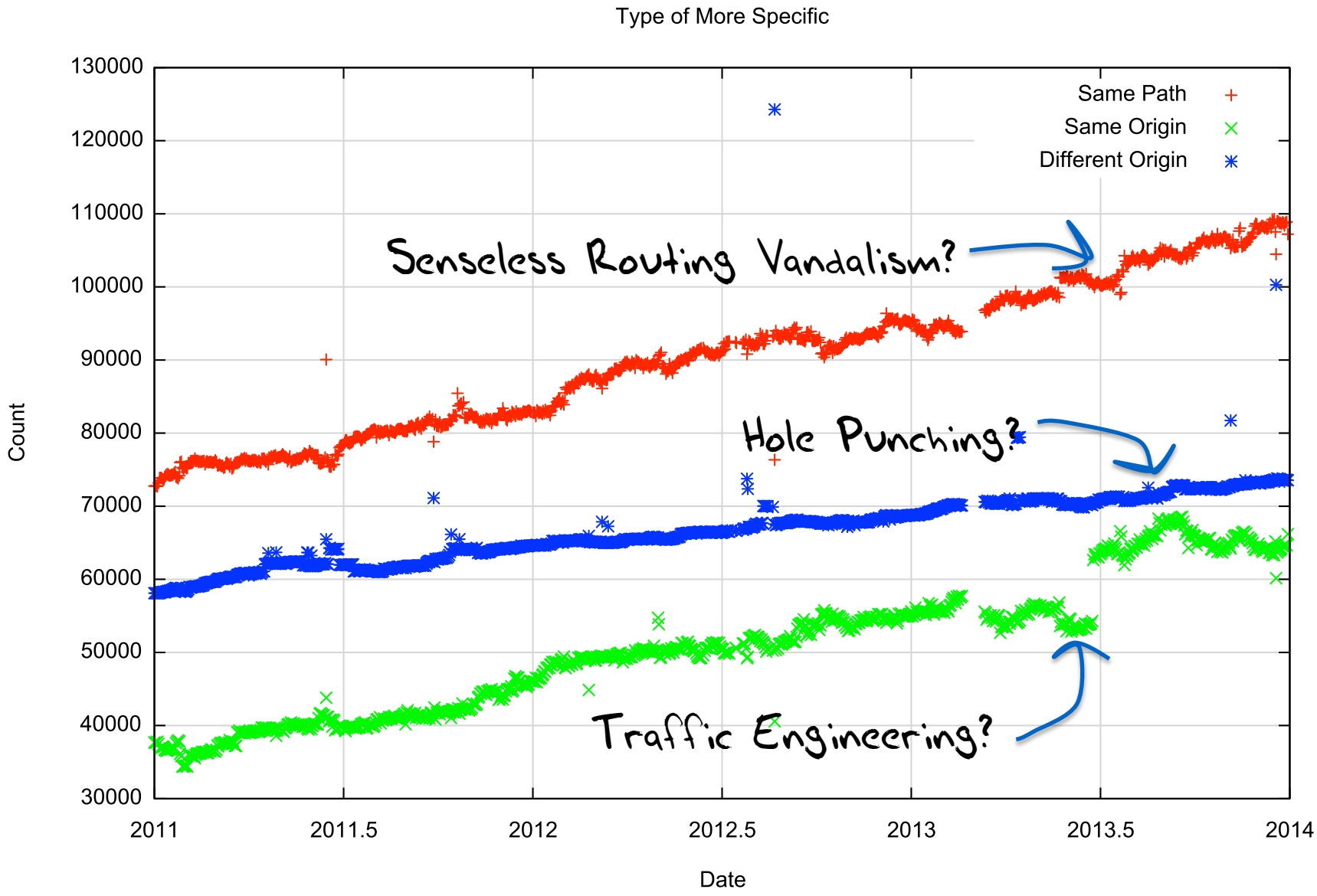
The reasons why we see more specifics in the routing system include:

- Different origination (“hole punching” in an aggregate)
- Traffic engineering of incoming traffic flows across multiple inter-AS paths
- “protection” against route hijacking by advertising more specifics
- Poor routing practices

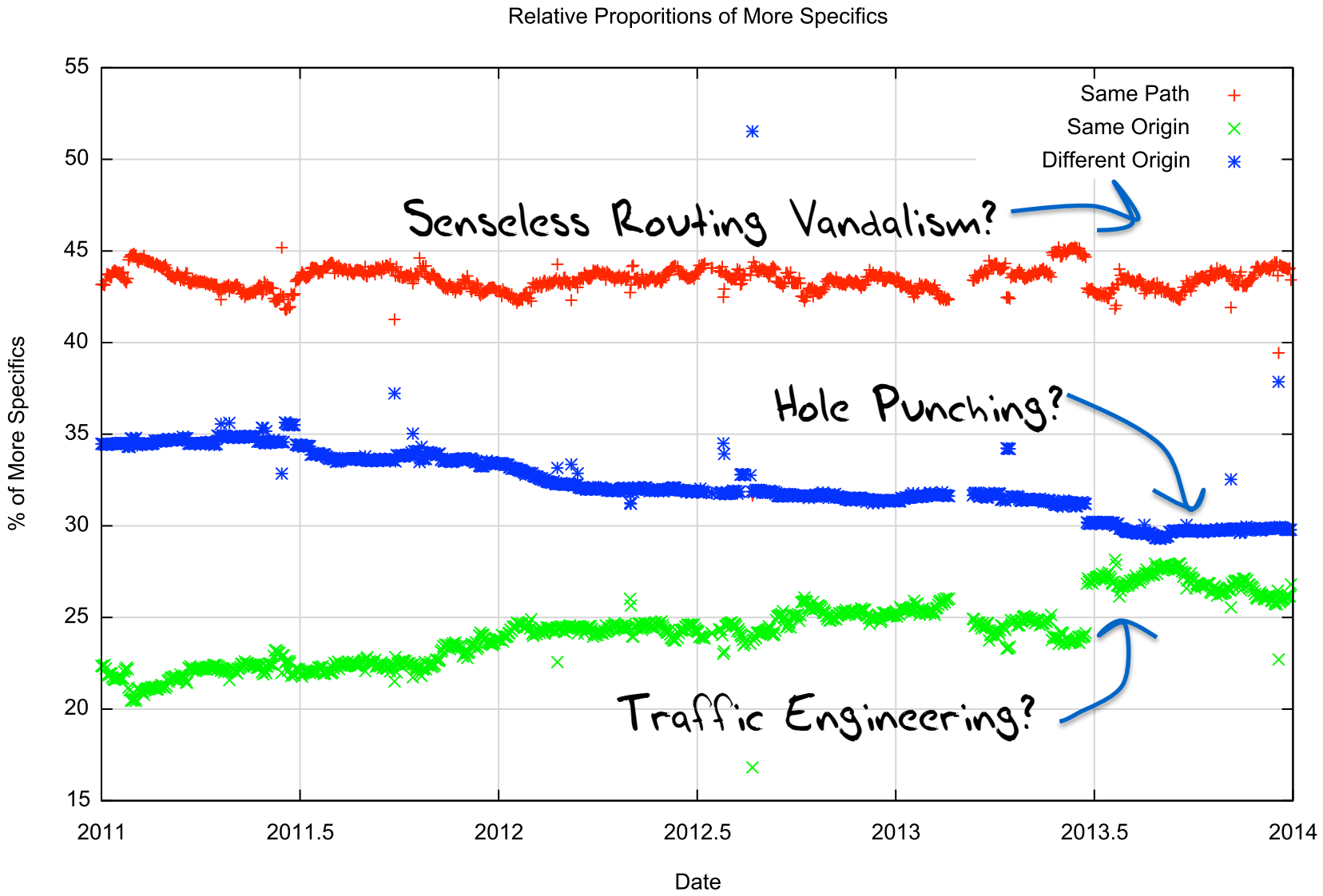
Types of More Specifics



Types of More Specifics



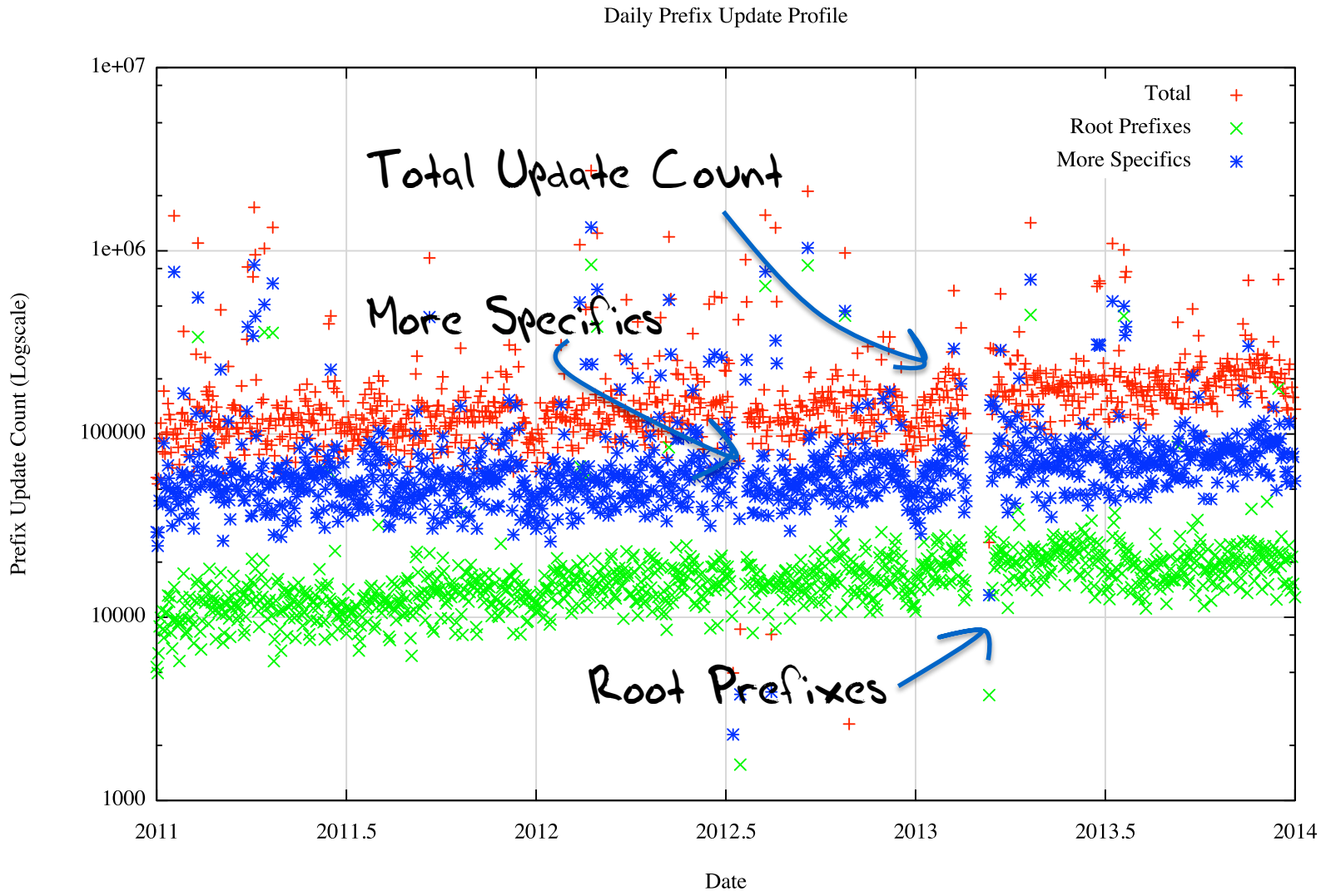
Types of More Specifics



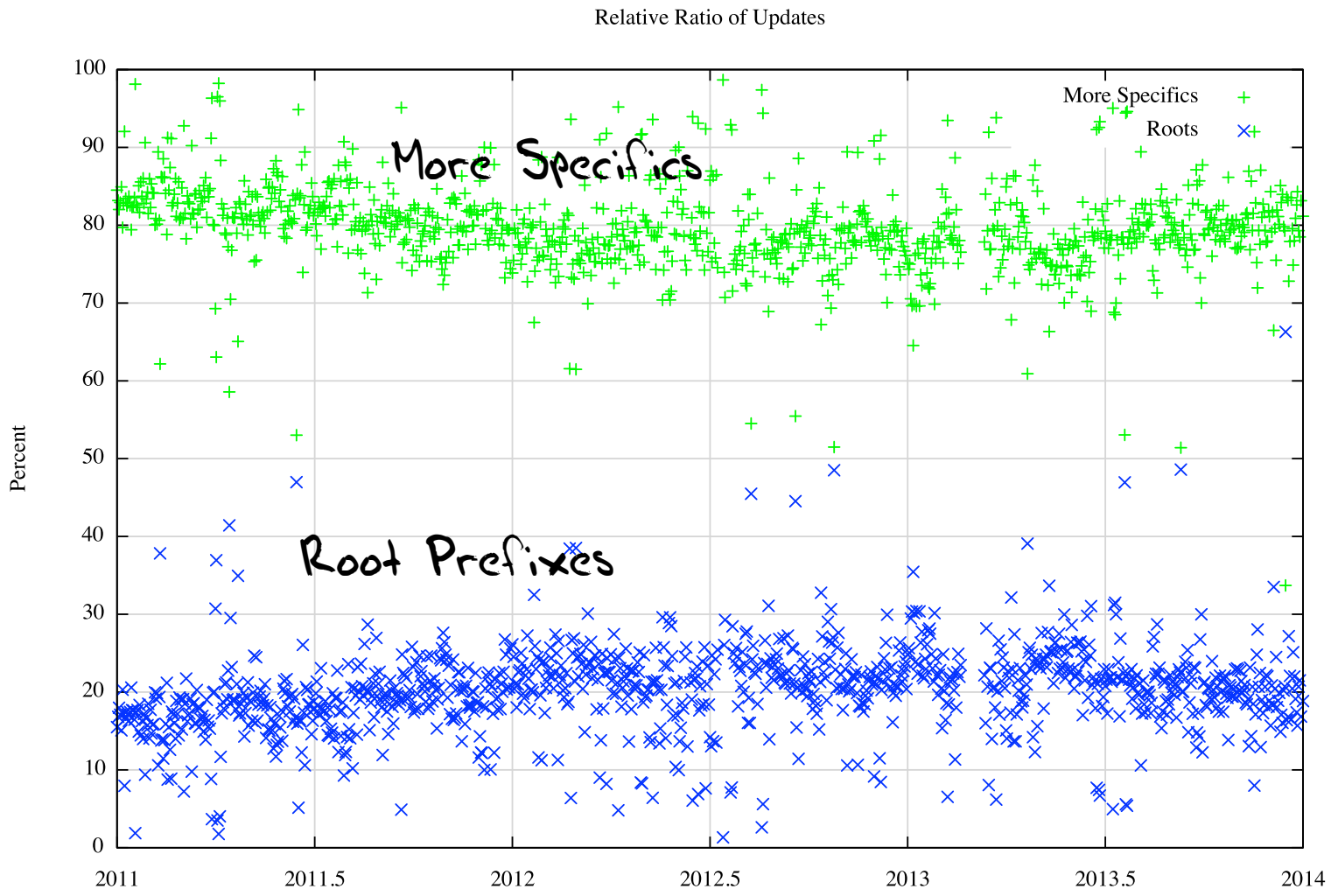
Daily Update Rates

- Do more specifics experience a higher update rate than aggregate advertisements?
- Lets examine the past 3 years of updates and examine the daily count of prefix updates for root aggregates and more specifics

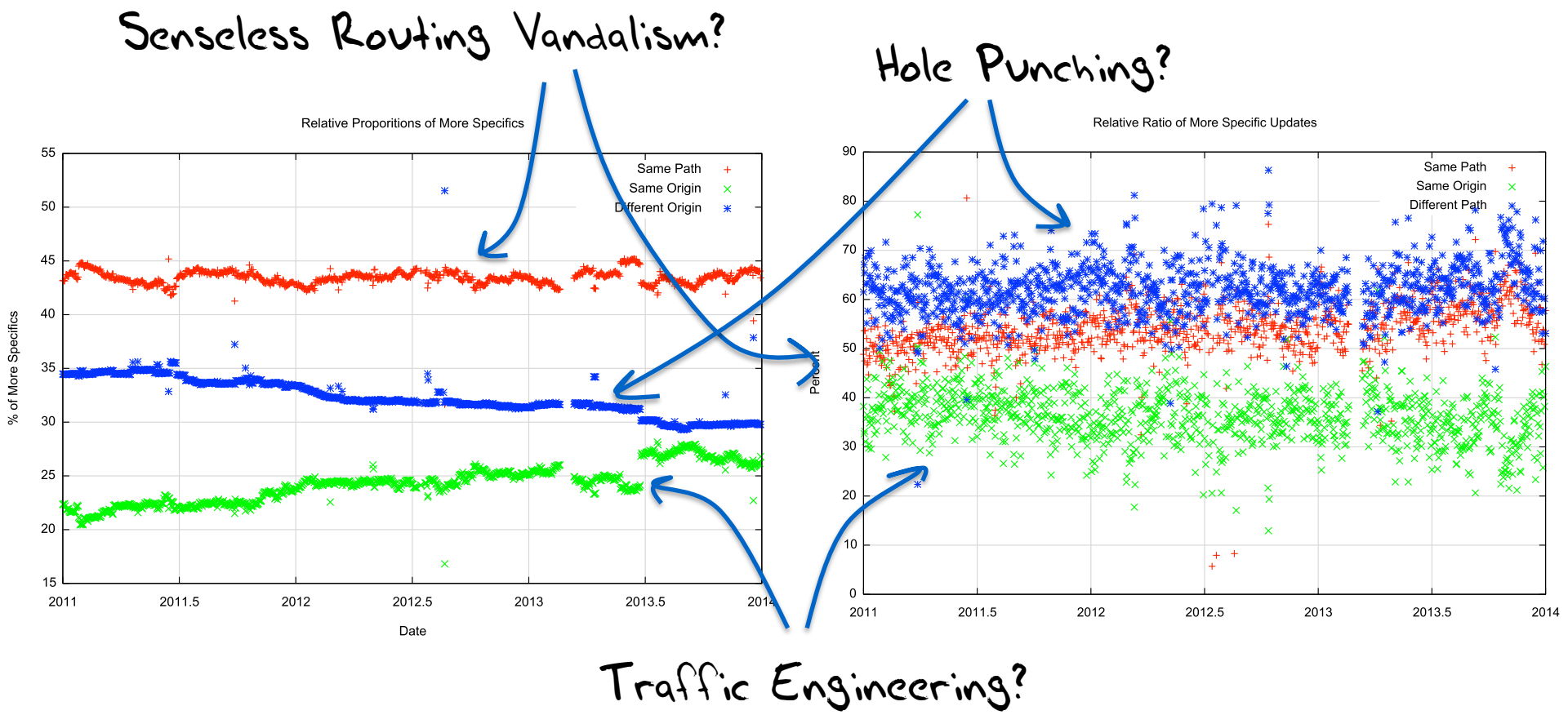
Daily BGP Updates



Relatively Speaking



More Specifics and Updates



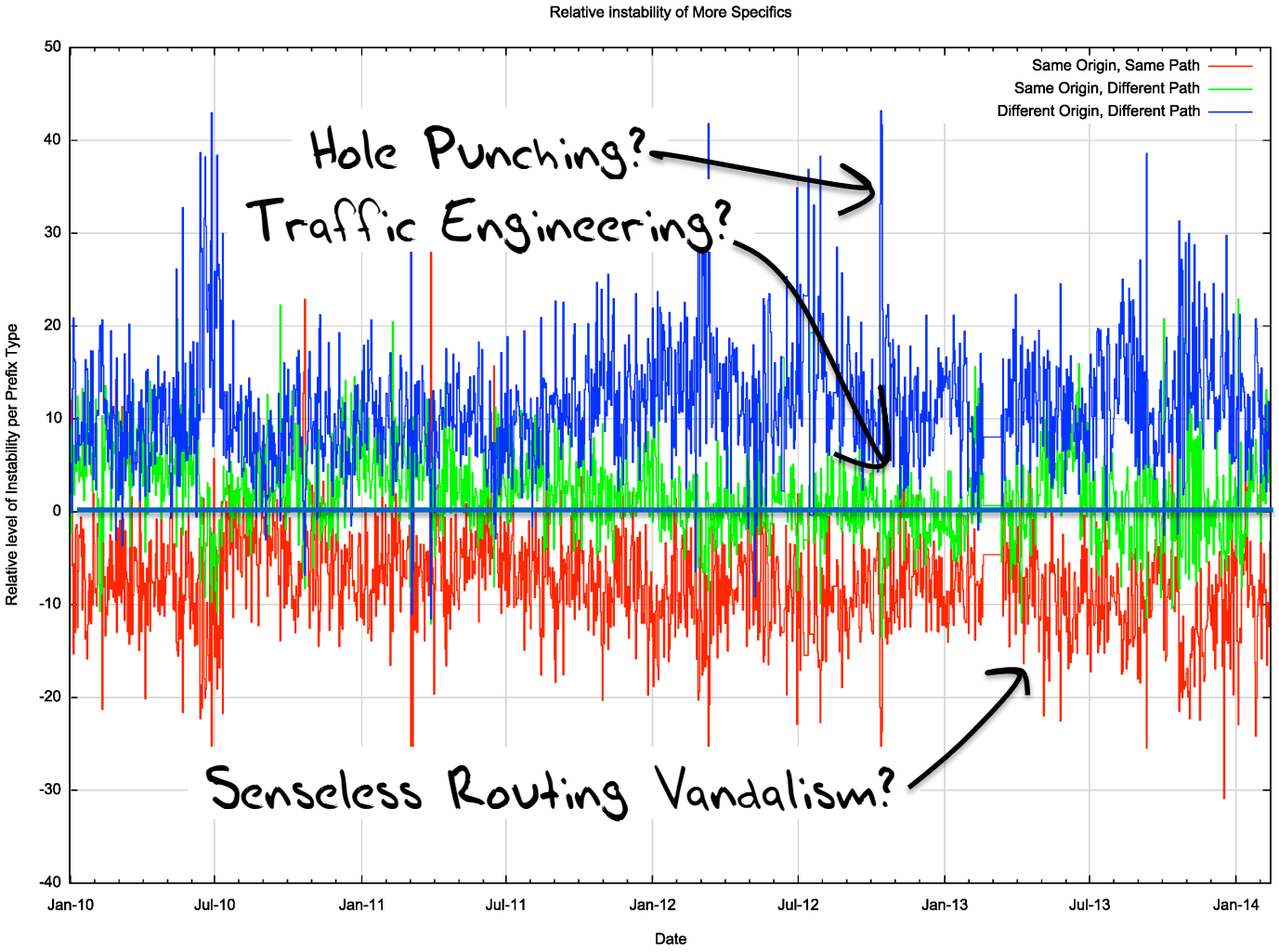
Daily Update Rates

- Do more specifics generate a higher update rate than aggregate advertisements?

  Yes – in terms of prefix updates, more specifics are some 4 times noisier than the aggregates in terms of update traffic totals

  More Specifics that “hole punch” (different origin AS) tend to be relatively noisier than other forms of more specifics. Is this because hole punching more specifics are less stable or are they “further away” and therefore noisier to converge?

Stability of More Specifics



Less Stable

More Stable

What are we seeing?

- The profile of updates in BGP is dominated by the instability of the more specific announcements, which are 4 x more likely to experience instability compared to aggregate advertisements
- With the set of more specifics, “hole punching” (different origin AS, different AS Path) is consistently less stable than the other two types of more specifics.

Problem? Not a Problem?

It’s evident that the global BGP routing environment suffers from a certain amount of neglect and inattention

Problem? Not a Problem?

It's evident that the global BGP routing environment suffers from a certain amount of neglect and inattention

Could we do better?

- Yes!
- A small number of networks originate the bulk of the more specifics and the bulk of the BGP update traffic.
- Rationalizing more specific advertisements will both reduce table size and also reduce the level of dynamic update in the inter-domain environment

Problem? Not a Problem?

It's evident that the global BGP routing environment suffers from a certain amount of neglect and inattention

Should we do better?

- It can be difficult to justify the effort and the cost: the current growth rates of the routing table lie within relatively modest parameters of growth and still sit within the broad parameters of constant unit cost of routing technology
- On the other hand, we need to recognize that we could do a lot better in terms of eliminating routing noise, and achieve this with with a relatively modest amount of effort

That's All!