The eBGP DFZ in 2011

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..”

quote from a 2012 routing

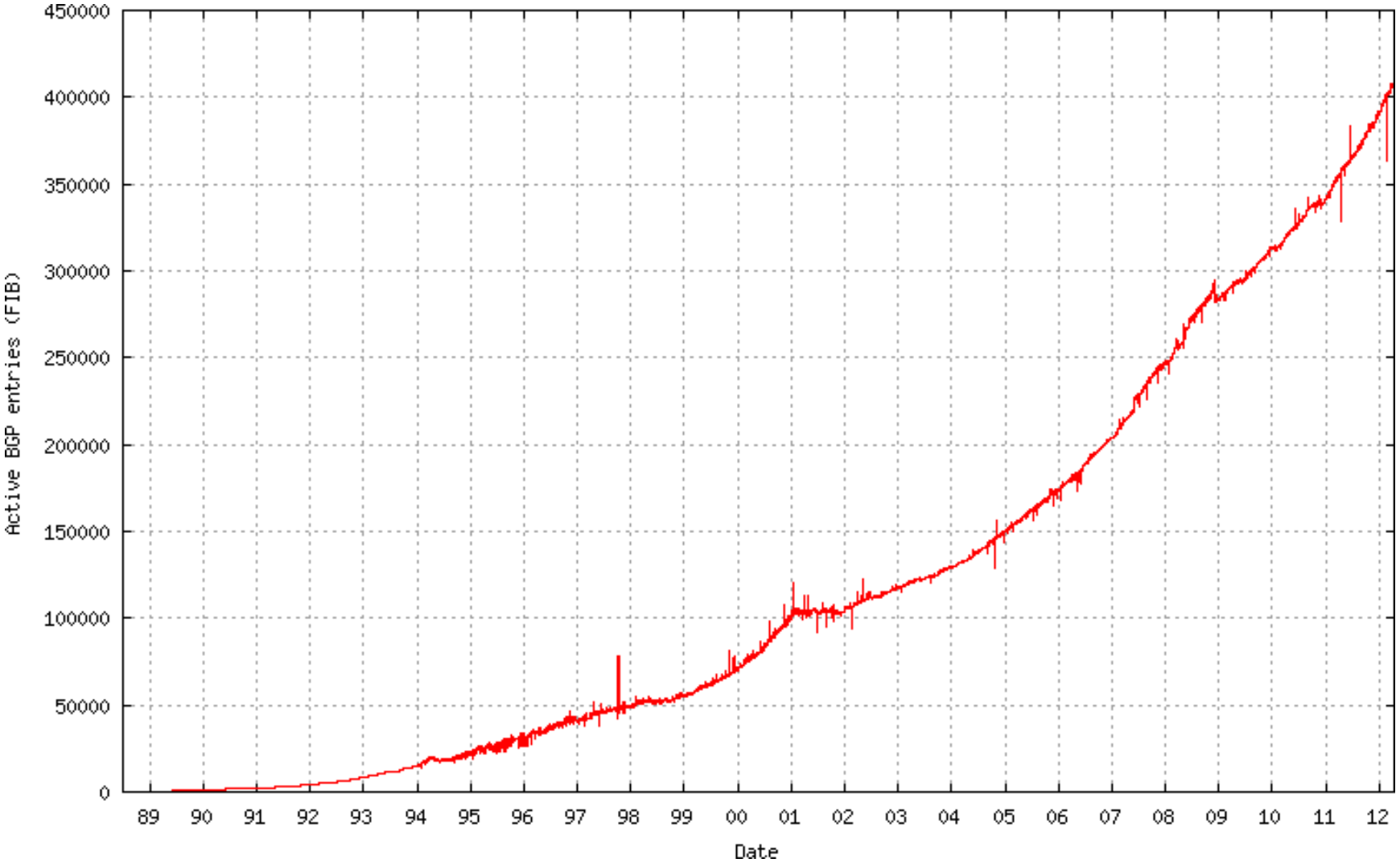
is this really true, or do we accept it as true without actually looking at the real behaviours of the Internet's routing system???

Agenda

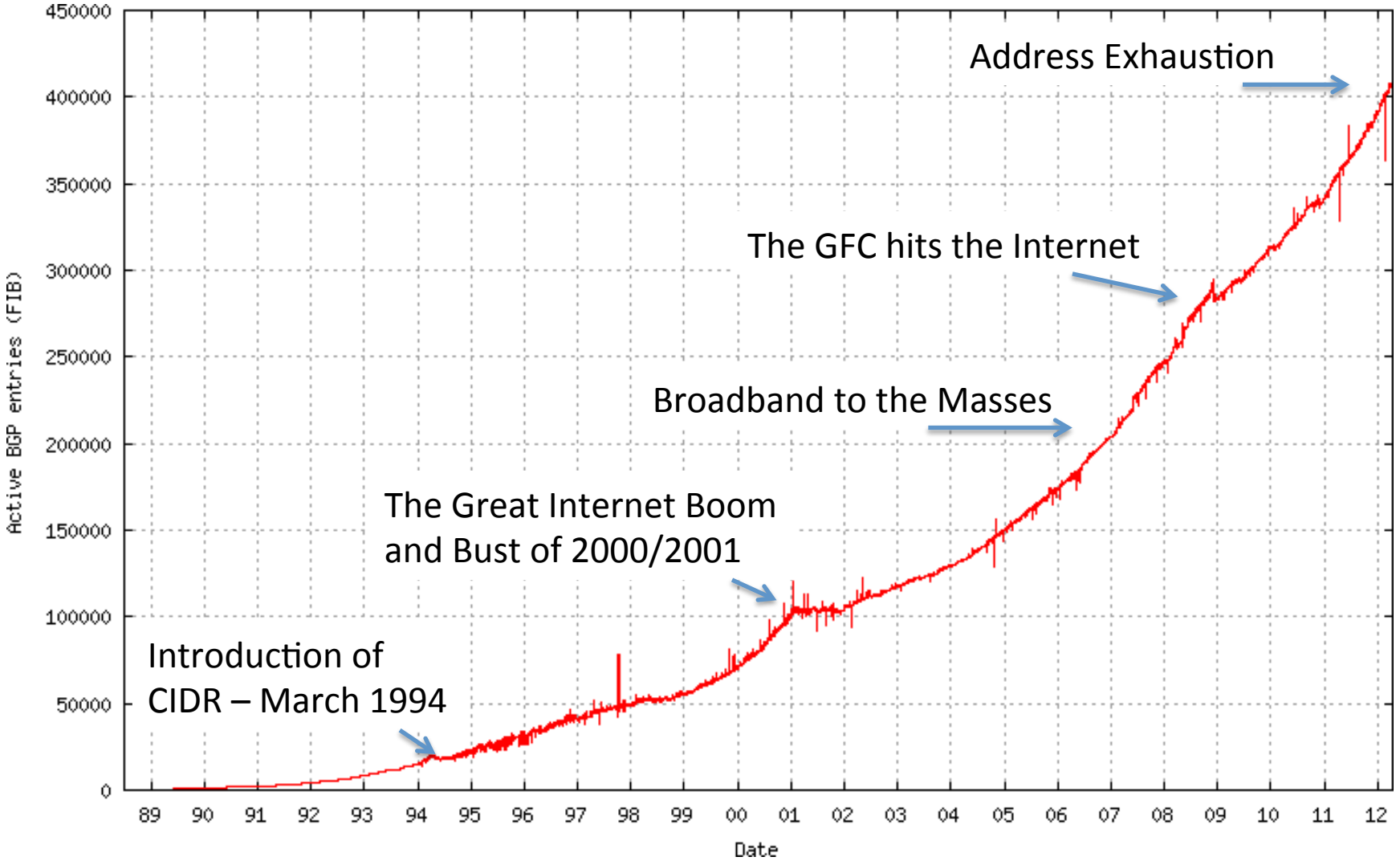
- In this presentation we will explore the space of inter-domain routing (the Border Gateway Protocol – BGP
  - We will look at the growth of the eBGP routing table over time and some projections for future growth
  - Then we'll look at the extent to which more specifics are dominating routing table growth ... or not

The Big Picture of BGP Growth

The Big Picture of the v4 Routing Table



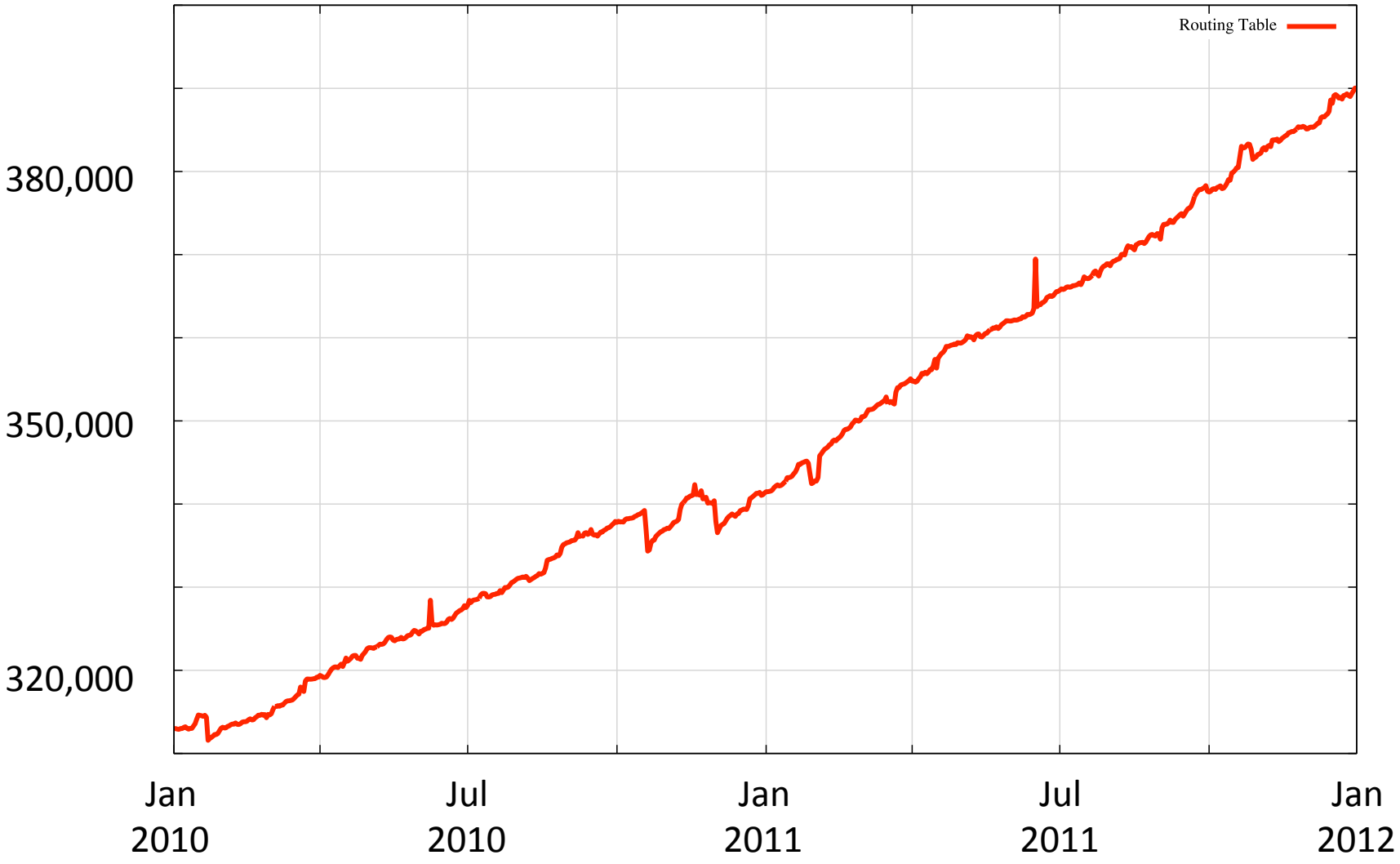
The Big Picture of the v4 Routing Table



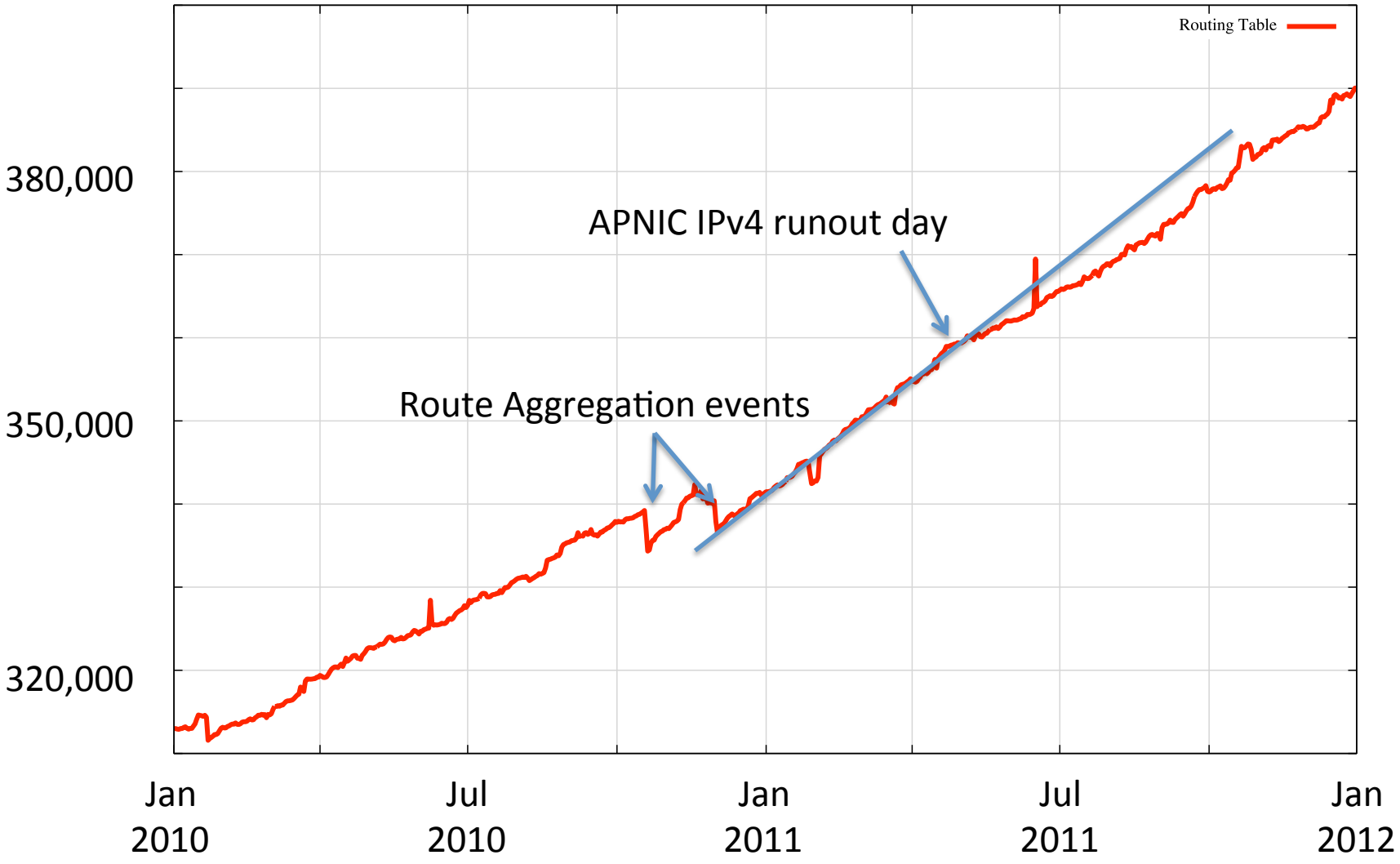
The Routing Table in 2010-2011

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

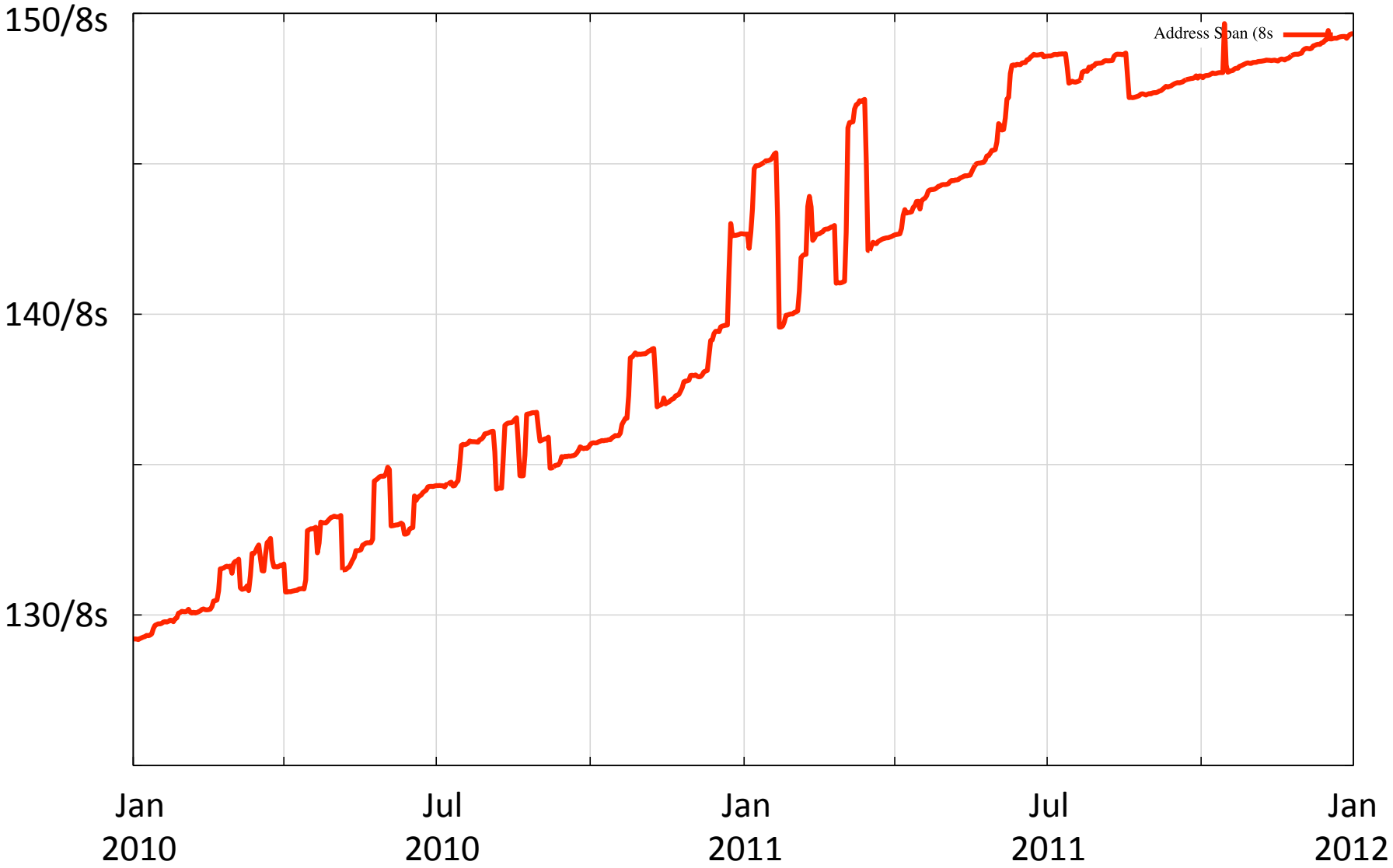
IPv4 BGP Prefix Count 2010 - 2011



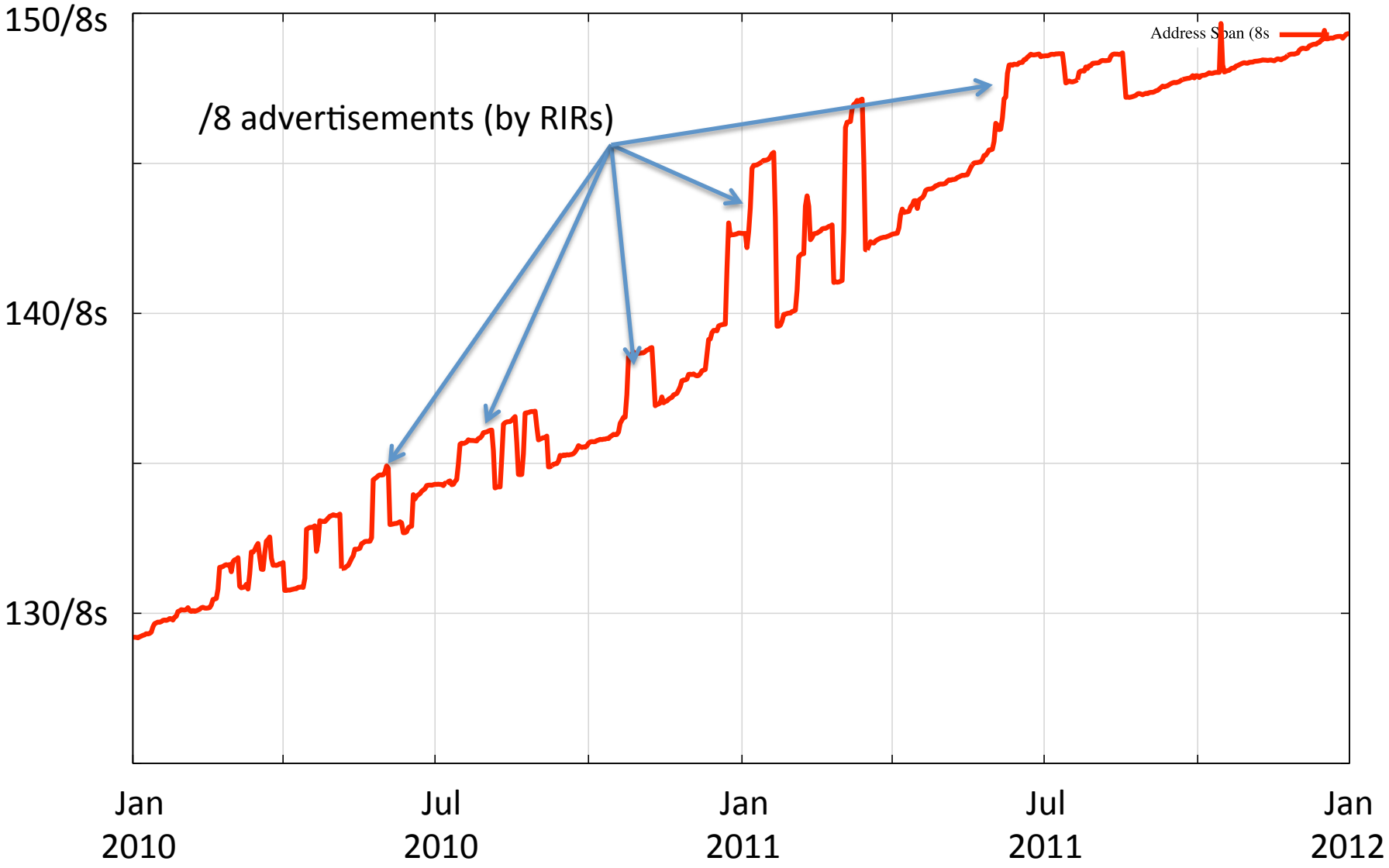
IPv4 BGP Prefix Count 2010 - 2011



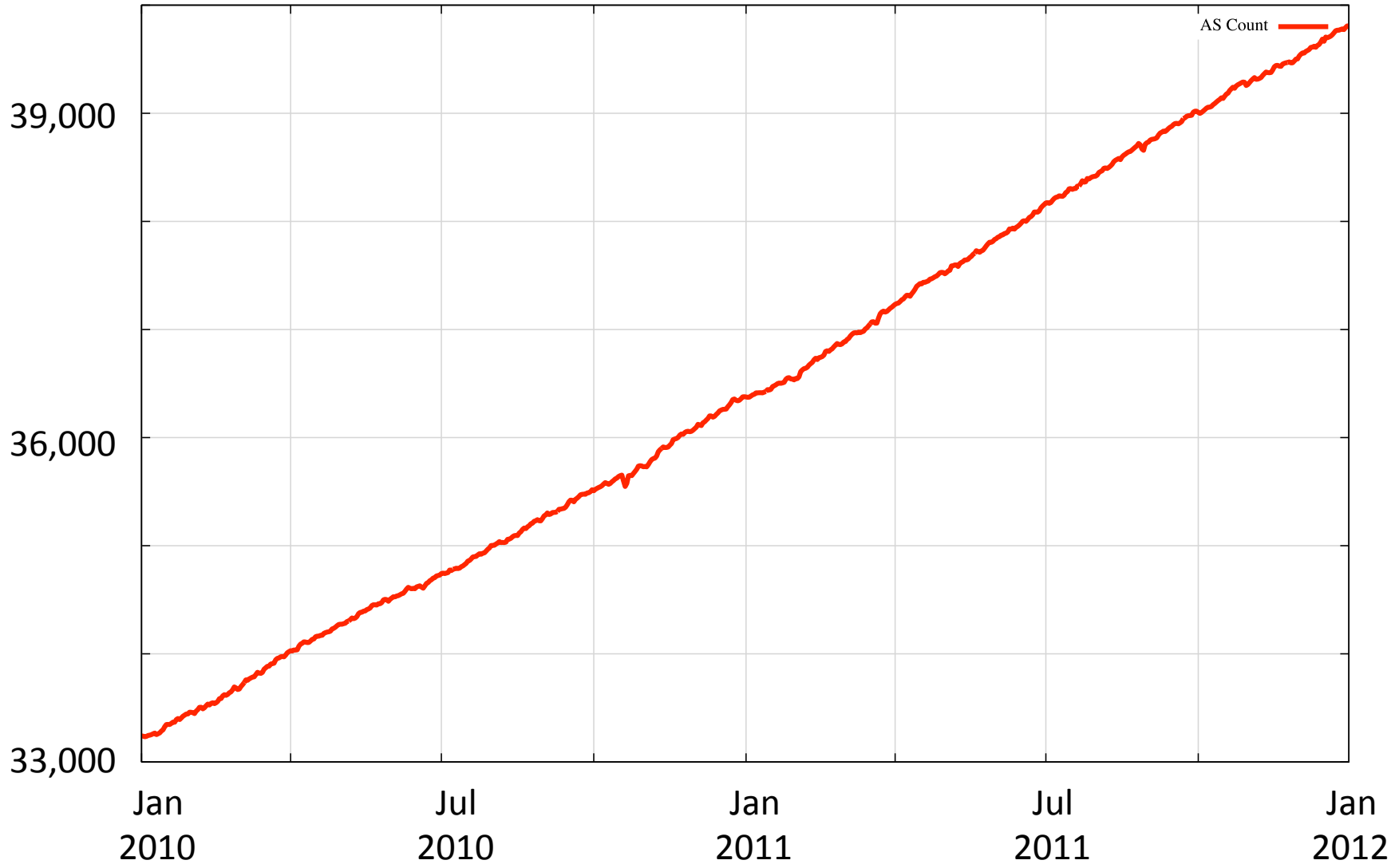
IPv4 Routed Address Span



IPv4 Routed Address Span



IPv4 Routed AS Count



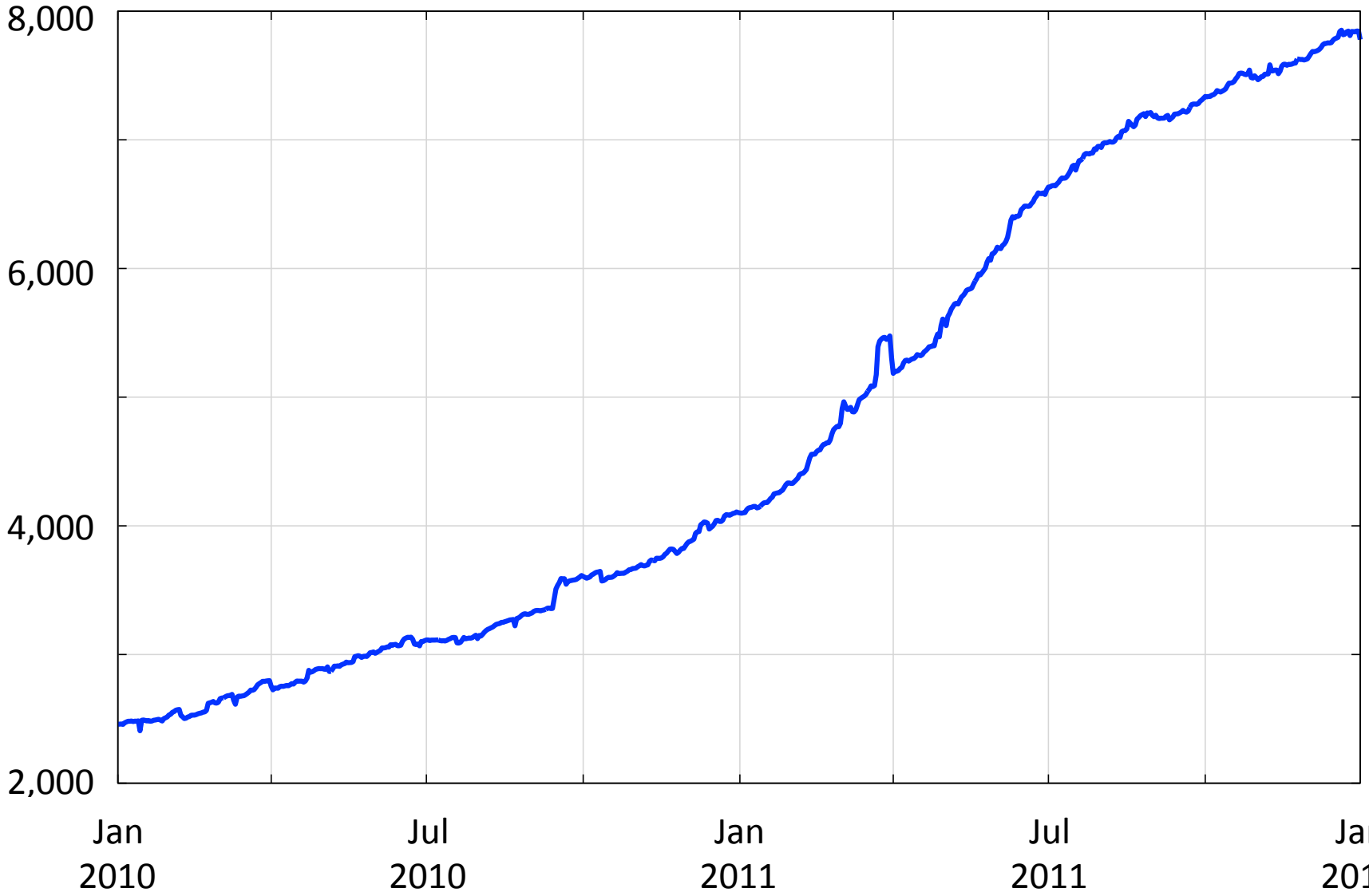
IPv4 2011 BGP Vital Statistics

	Jan-11	Jan-12	
Prefix Count	341,000	390,000	+14%
Roots	168,000	190,000	+13%
More Specifics	173,000	200,000	+15%
Address Span	140/8s	149/8s	+ 6%
AS Count	36,400	39,800	+ 9%
Transit	5,000	5,700	+14%
Stub	31,400	34,100	+ 9%

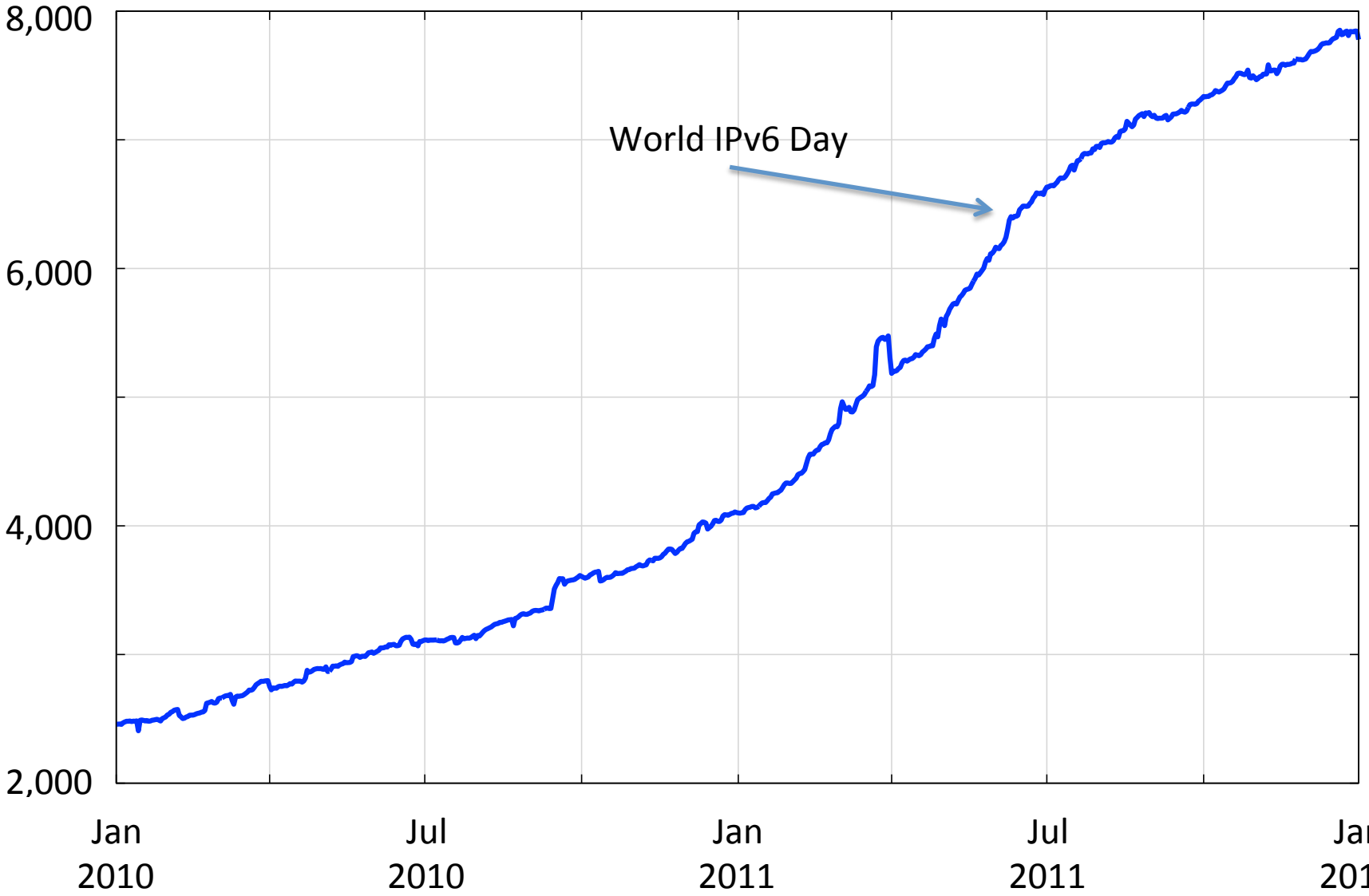
IPv4 in 2011

- Overall Internet growth in terms of BGP is at a rate of some ~12% p.a.
  - This is much the same as 2009 and 2010.
- Table growth has slowed since 20 April 2011, following APINC's IPv4 address run out
- Address span growing more slowly than the table size (address consumption pressures evident?)
  - But its picked up again in the first few months of 2012

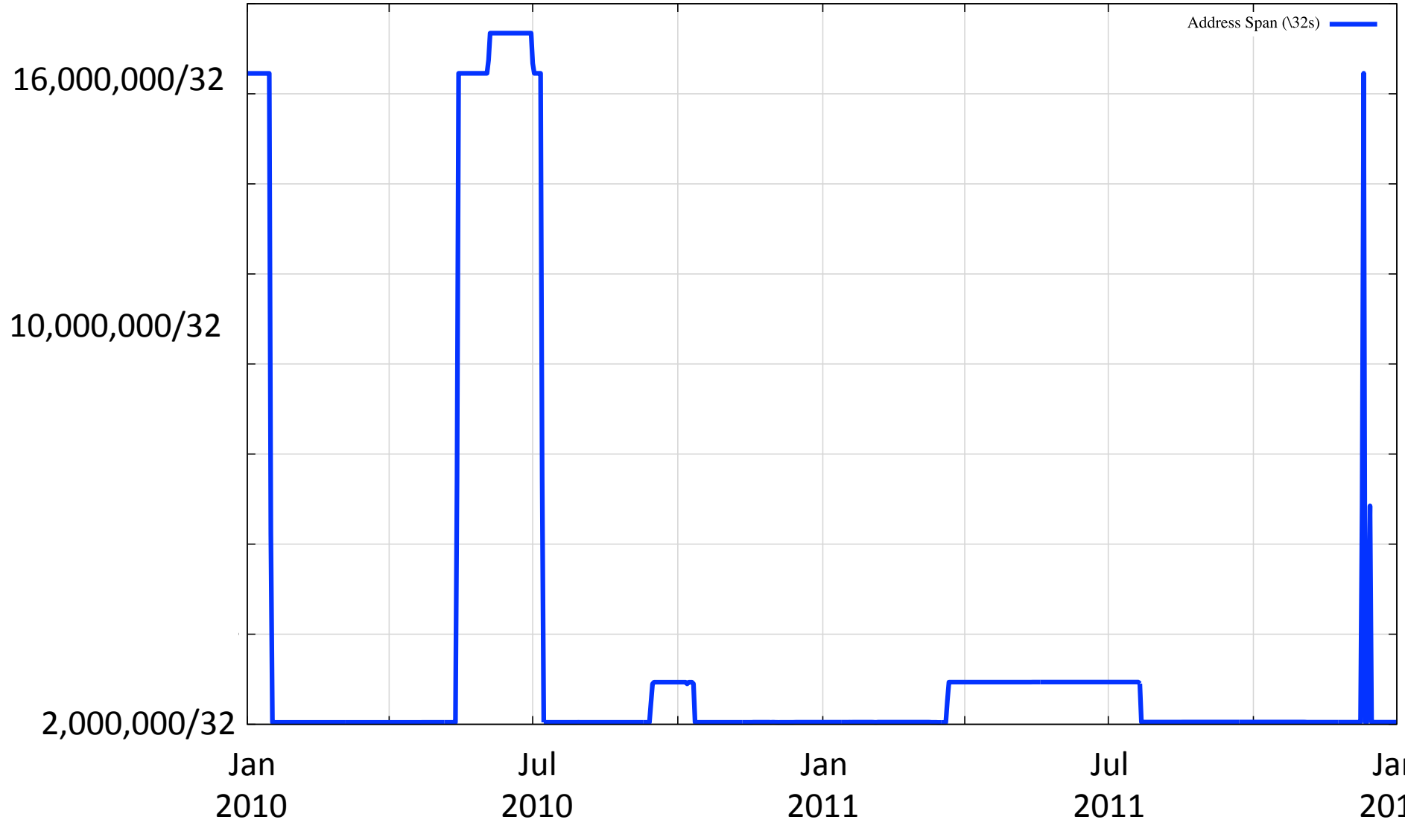
IPv6 BGP Prefix Count



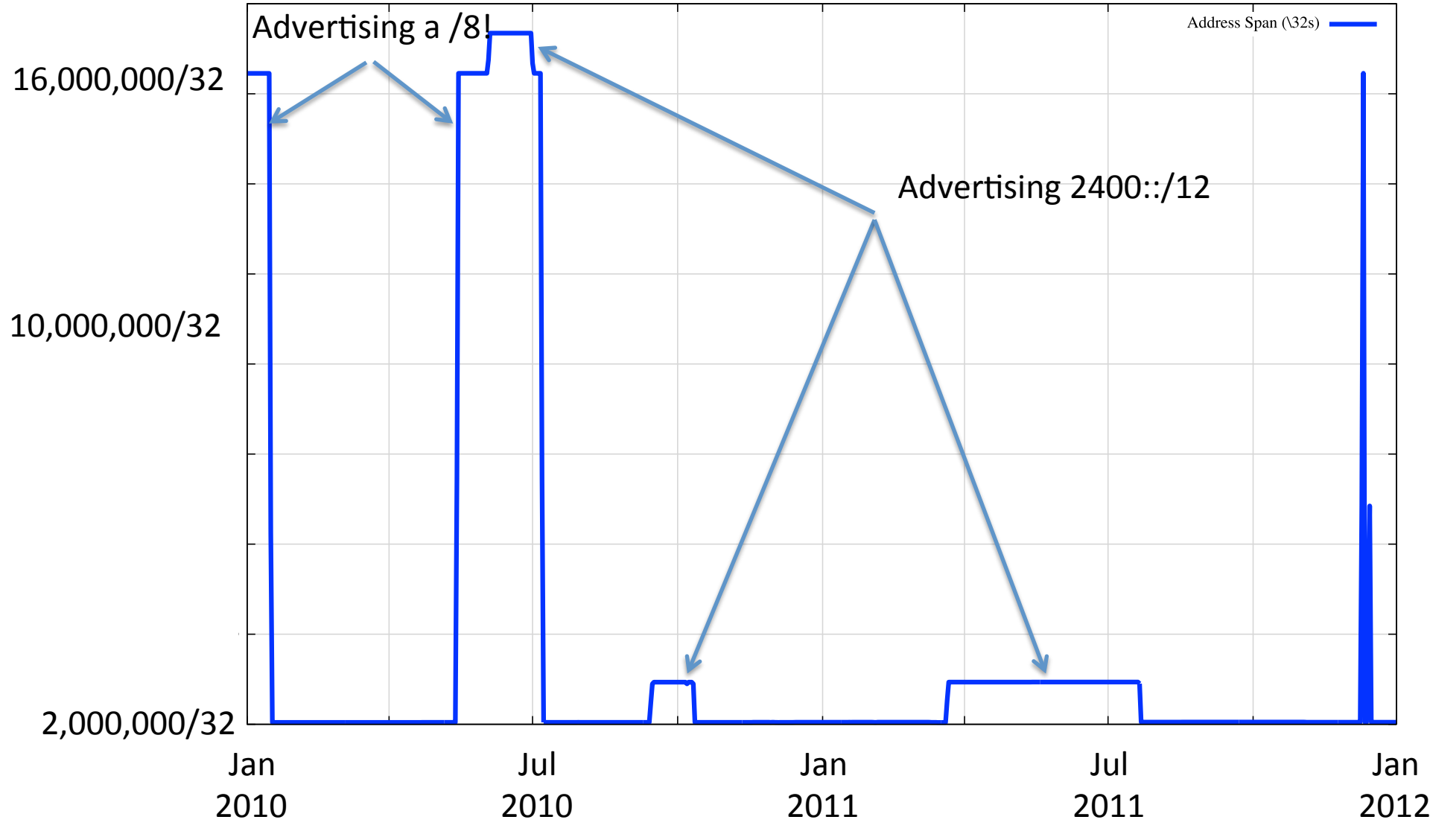
IPv6 BGP Prefix Count



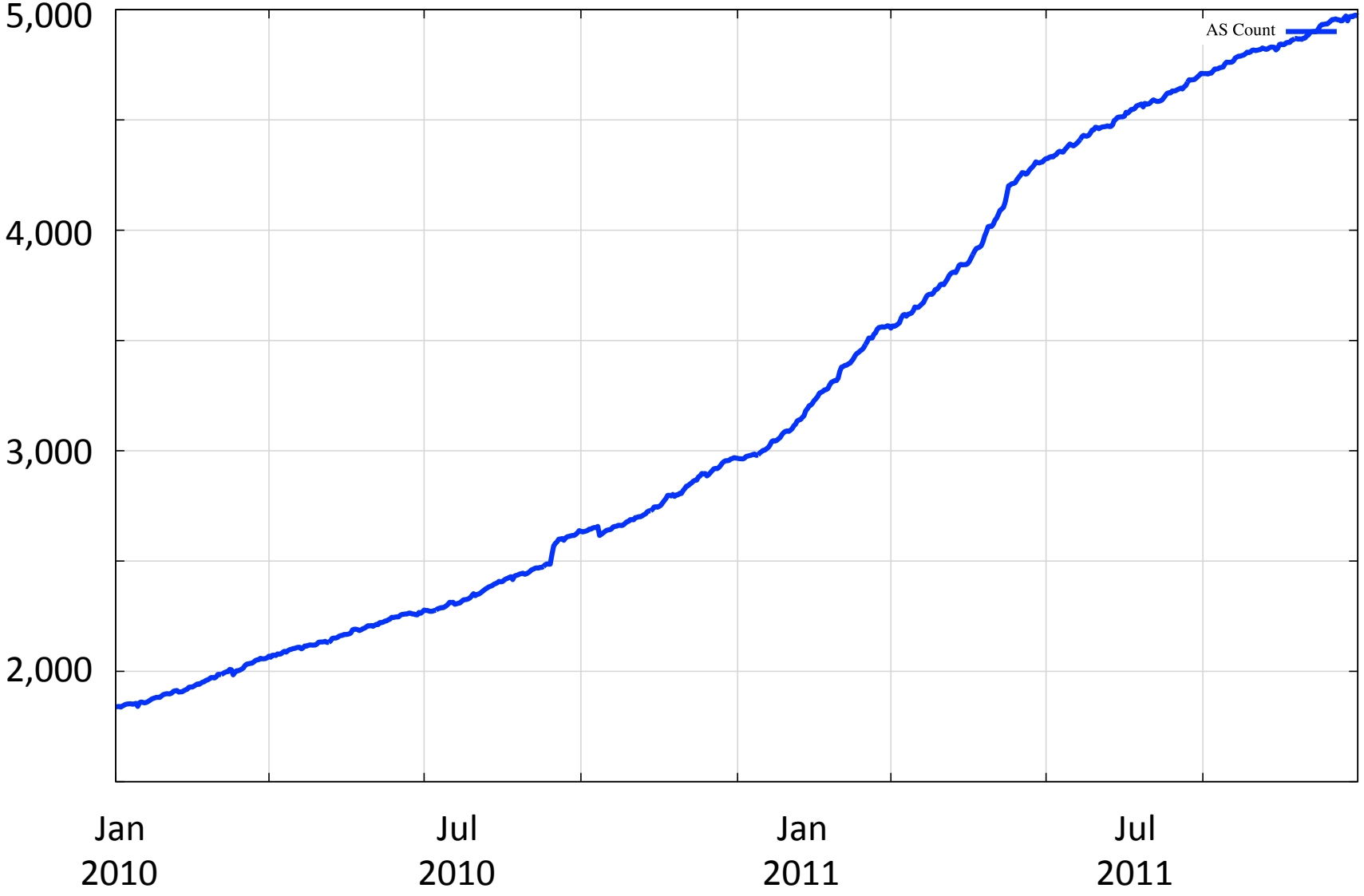
IPv6 Routed Address Span



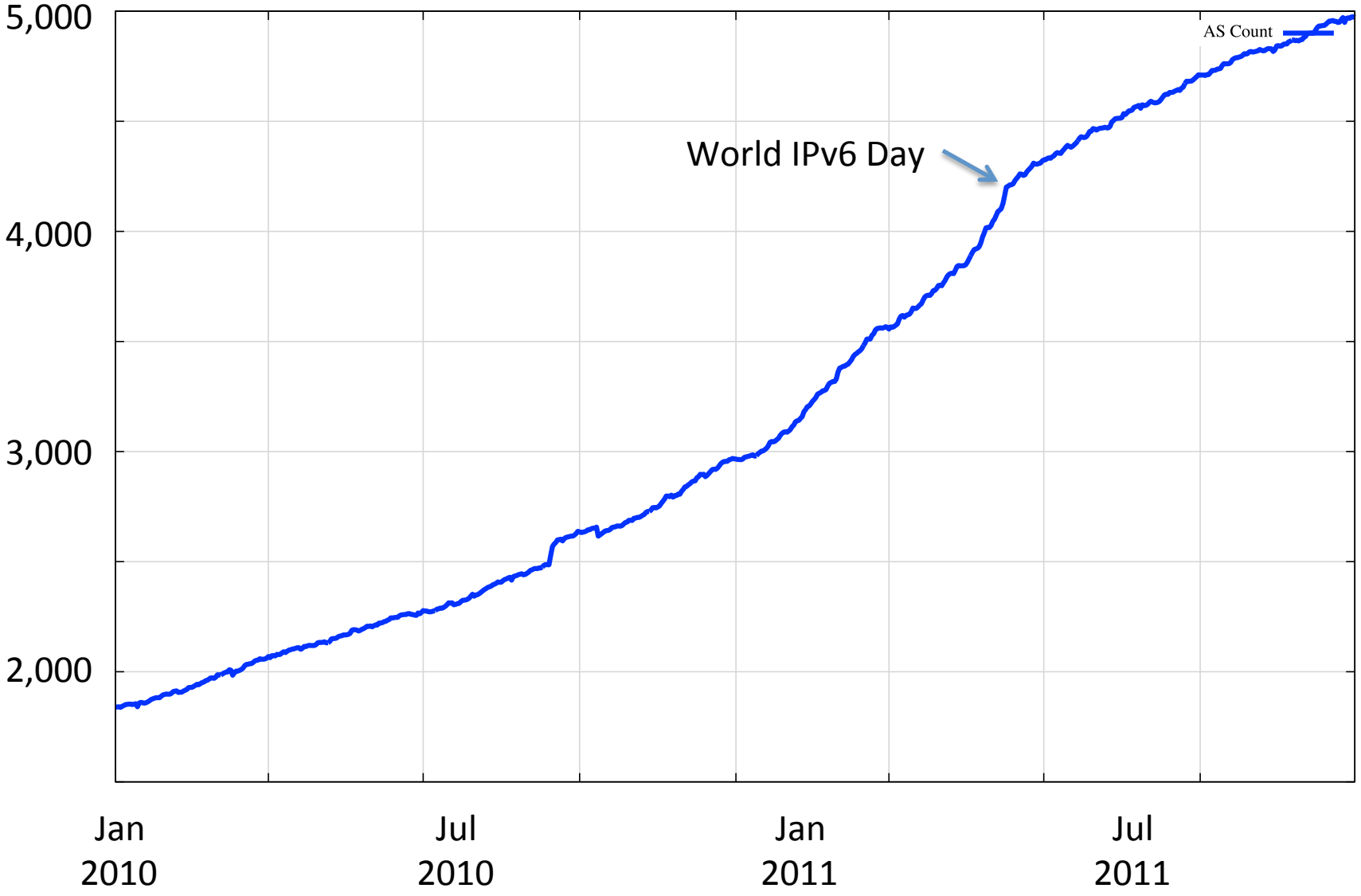
IPv6 Routed Address Span



IPv6 Routed AS Count



IPv6 Routed AS Count



IPv6 2011 BGP Vital Statistics

	Jan-11	Jan-12	p.a. rate
Prefix Count	4,100	7,759	+ 89%
Roots	3,178	5,751	+ 81%
More Specifics	922	2,008	+118%
Address Span (/32s)	53,415	53,387	+ 0%
AS Count	2,966	4,968	+ 67%
Transit	556	985	+ 77%
Stub	2,343	3,983	+ 70%

IPv6 in 2010 - 2011

- Overall IPv6 Internet growth in terms of BGP is 80% - 90 % p.a.
  - 2009 growth rate was ~ 50%.

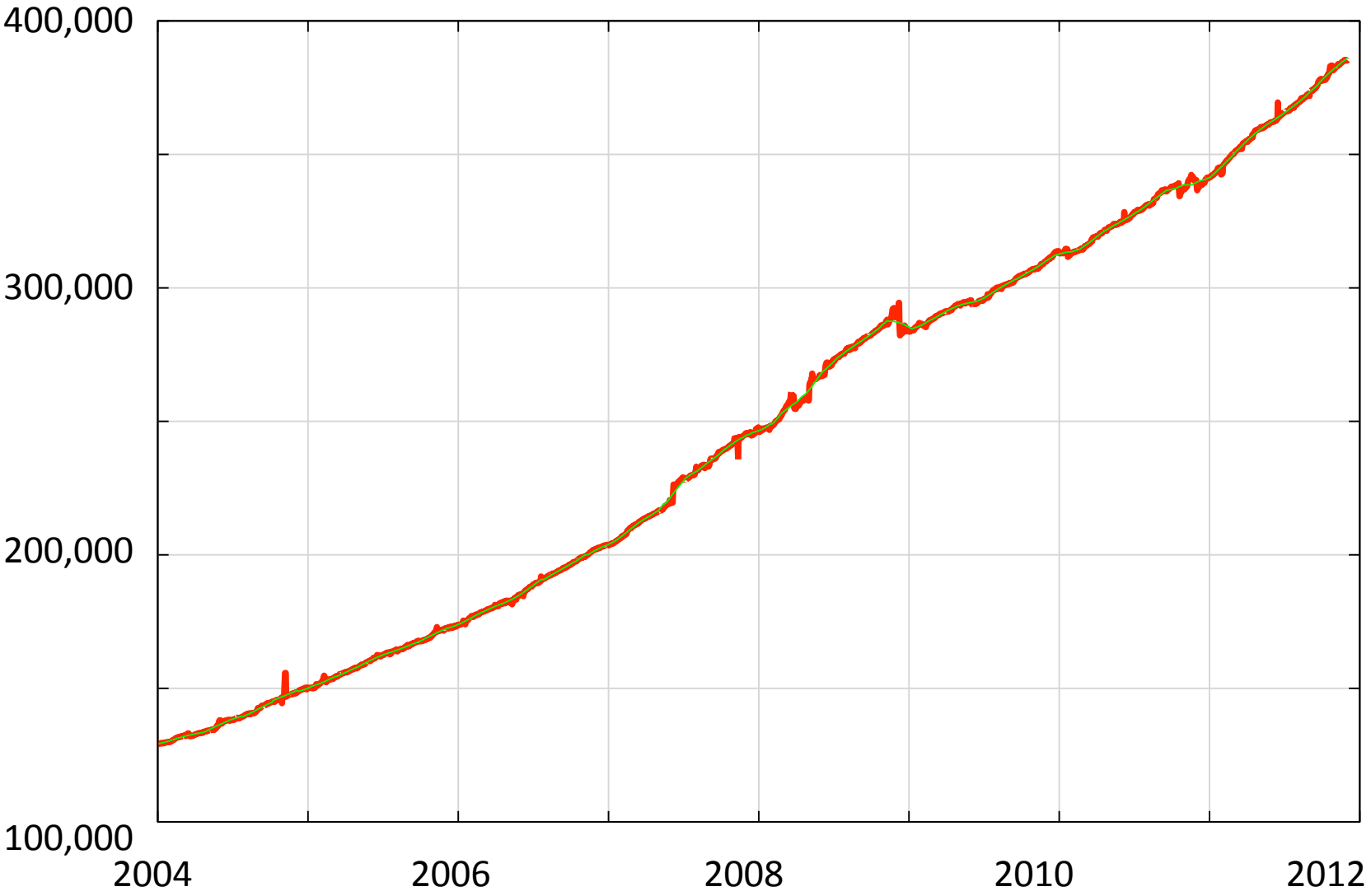
(Looking at the AS count, if these relative growth rates persist then the IPv6 network would span the same network domain as IPv4 in 4 years time -- mid/late 2016)

BGP Size Projections

- Generate a projection of the IPv4 routing table using a quadratic (O(2) polynomial) over the historic data
  - 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!

IPv4 Table Size



Daily Growth Rates

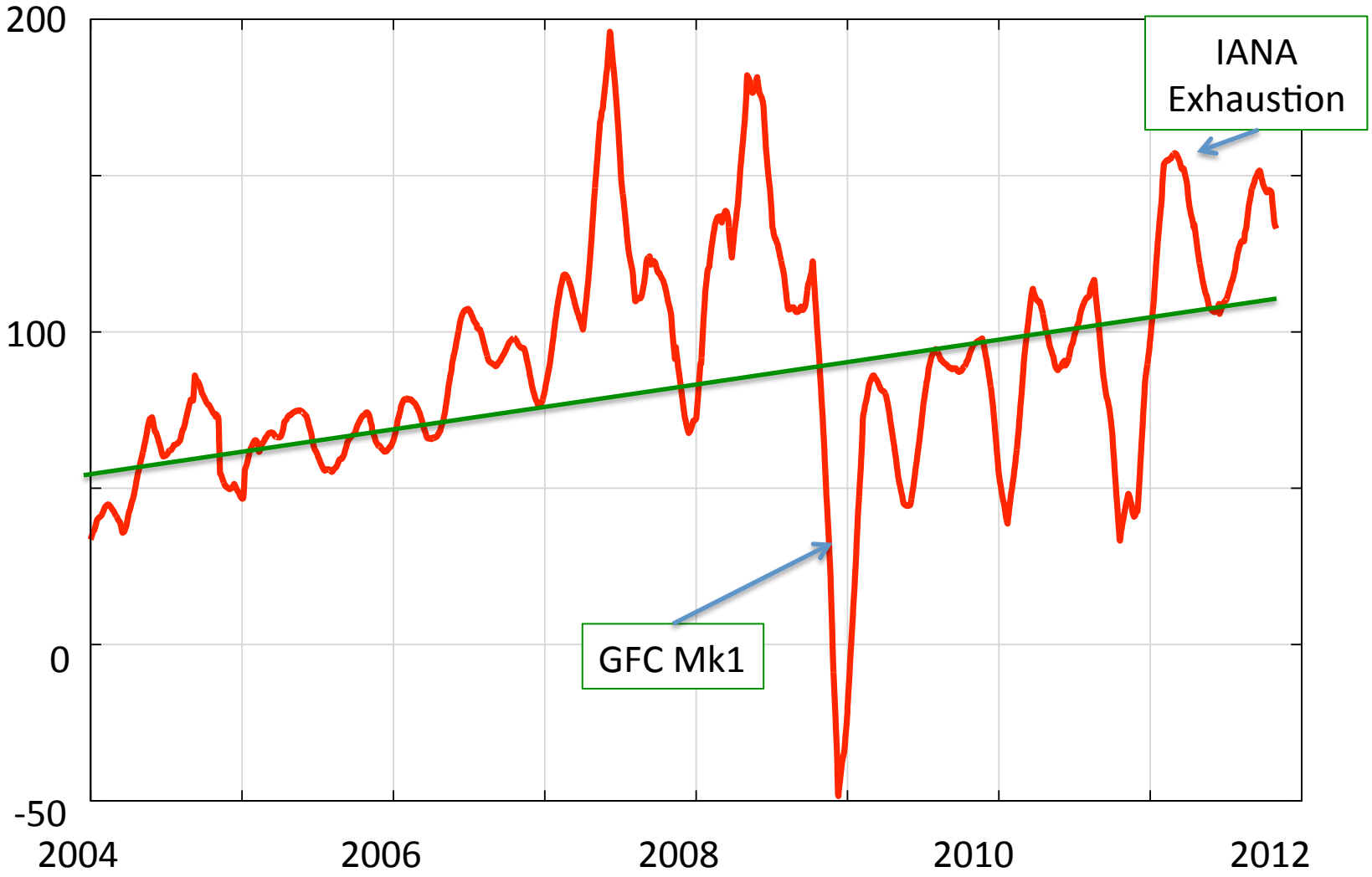
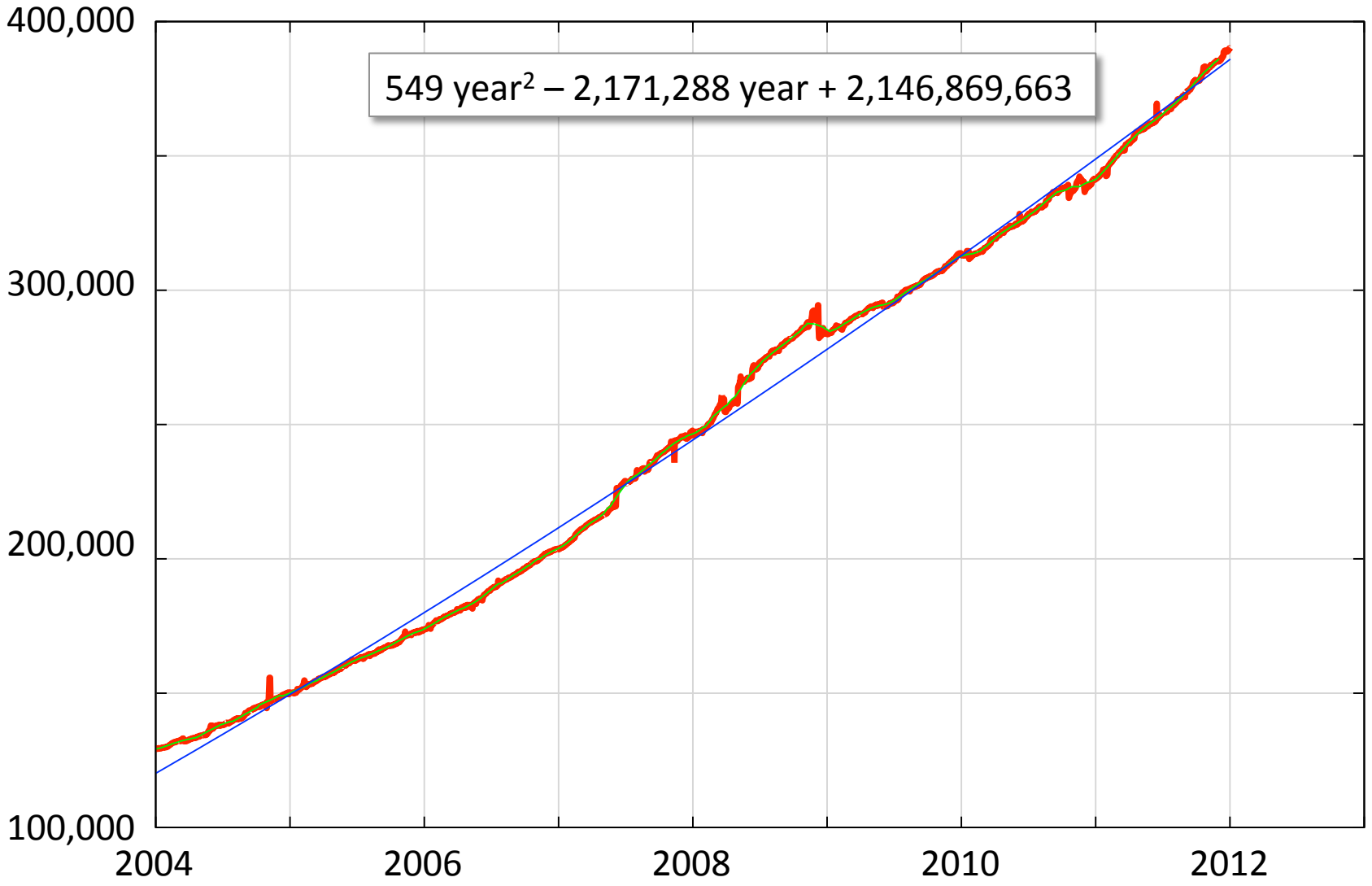
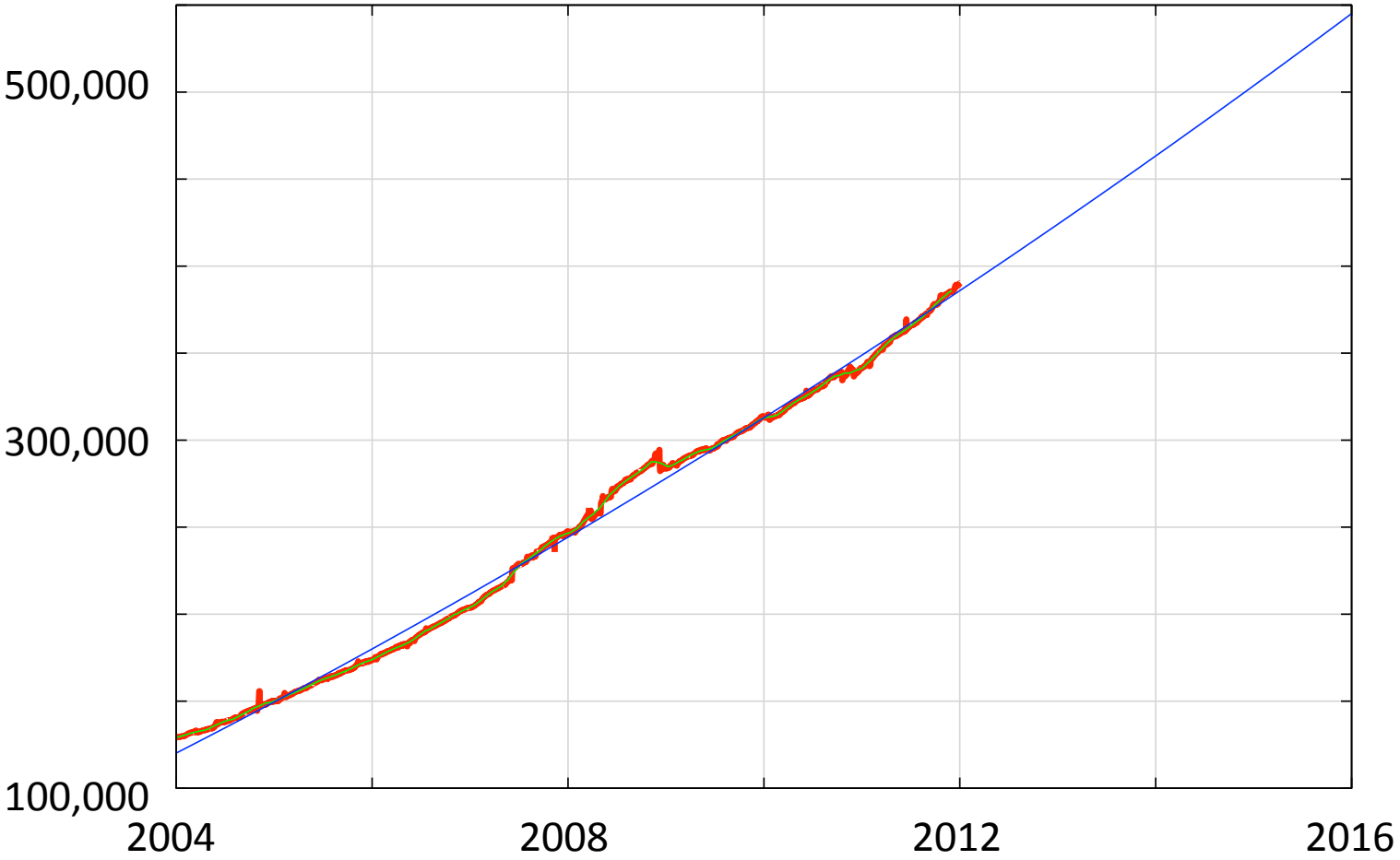


Table Growth Model



IPv4 Table Projection

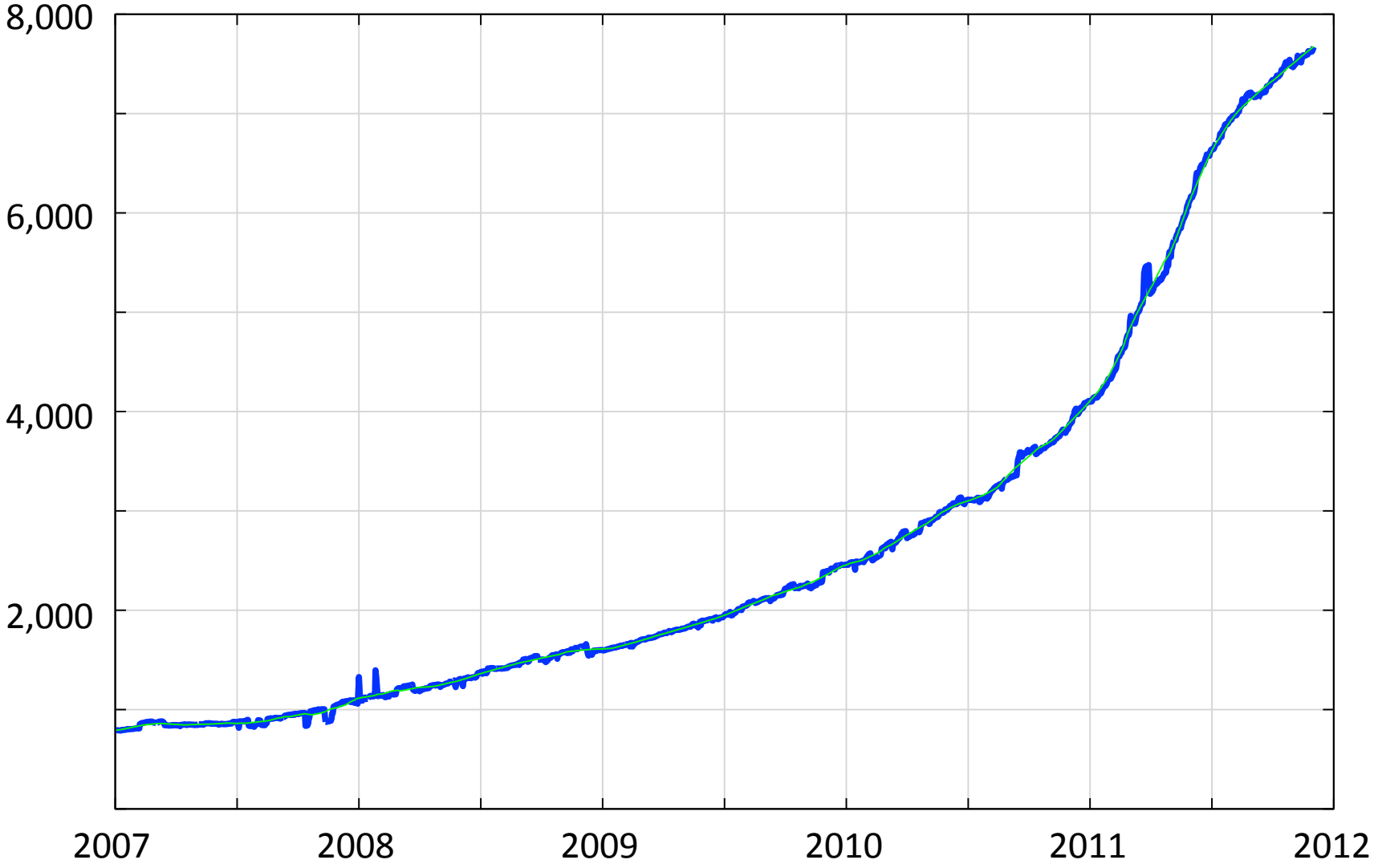


IPv4 BGP Table Size predictions

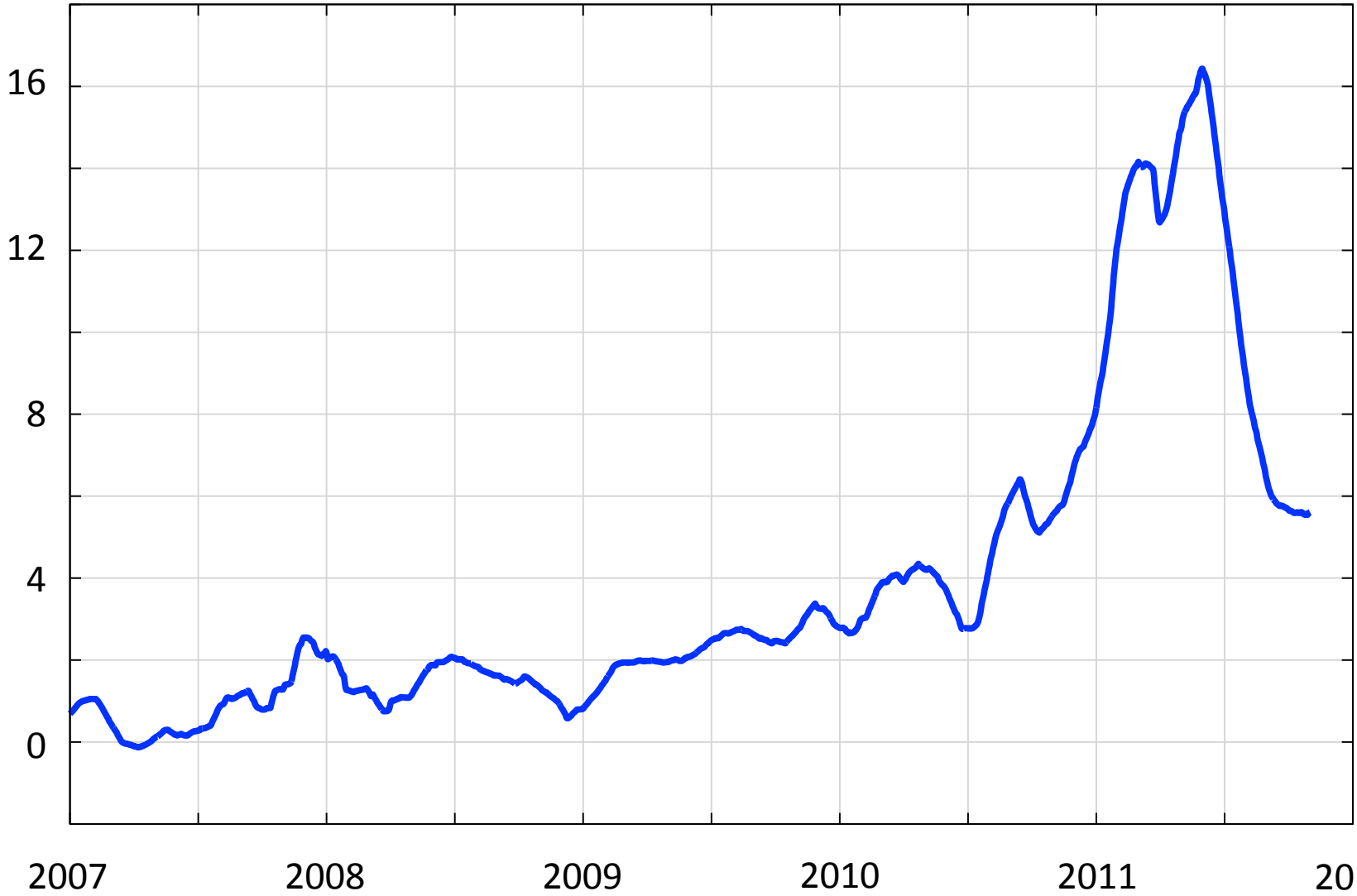
Jan 2011	347,000 entries
2012	390,000 entries
2013*	424,000 entries
2014*	463,000 entries
2015*	503,000 entries
2016*	545,000 entries

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

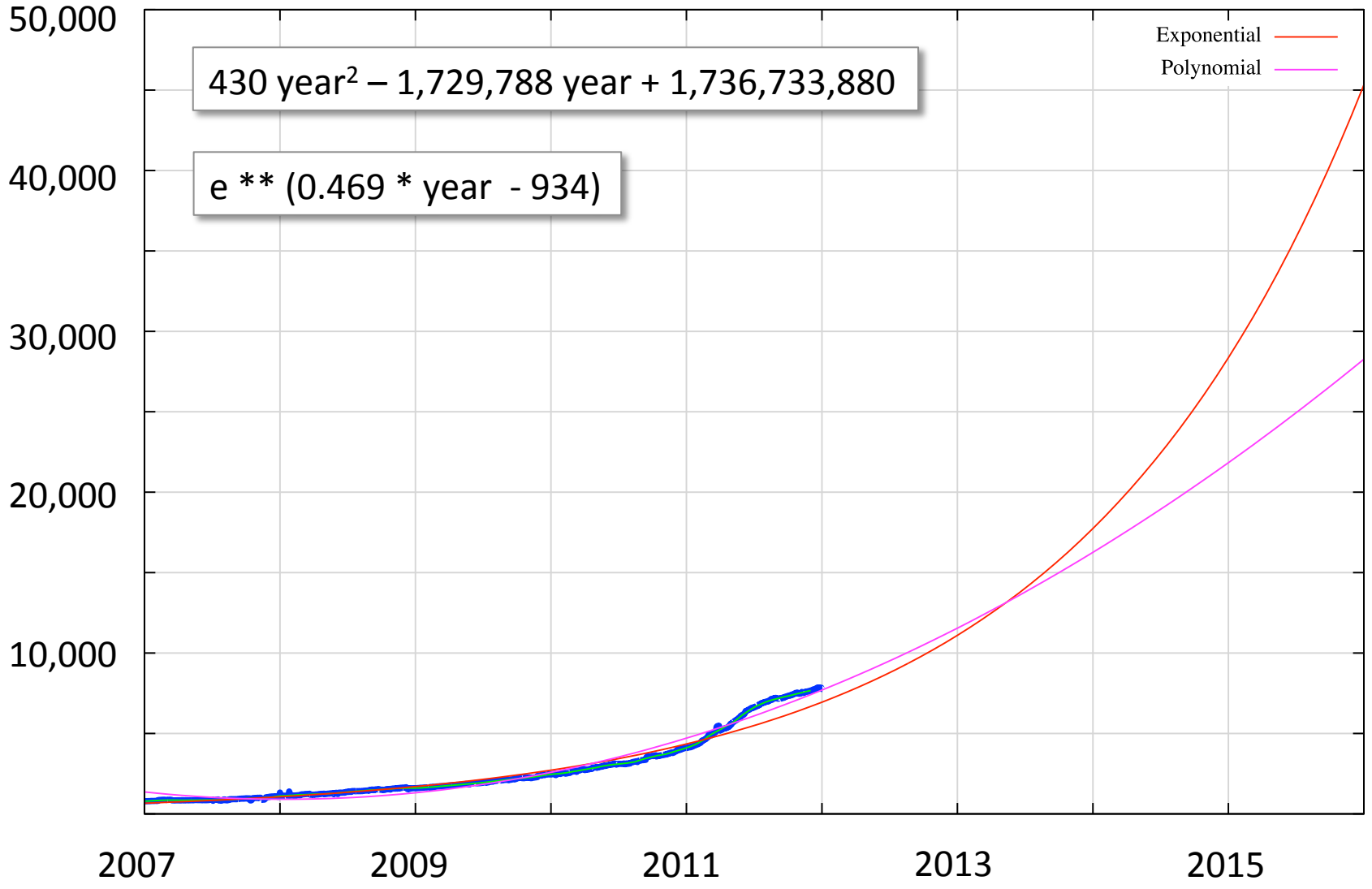
IPv6 Table Size



Daily Growth Rates



IPv6 Table Projection



IPv6 BGP Table Size predictions

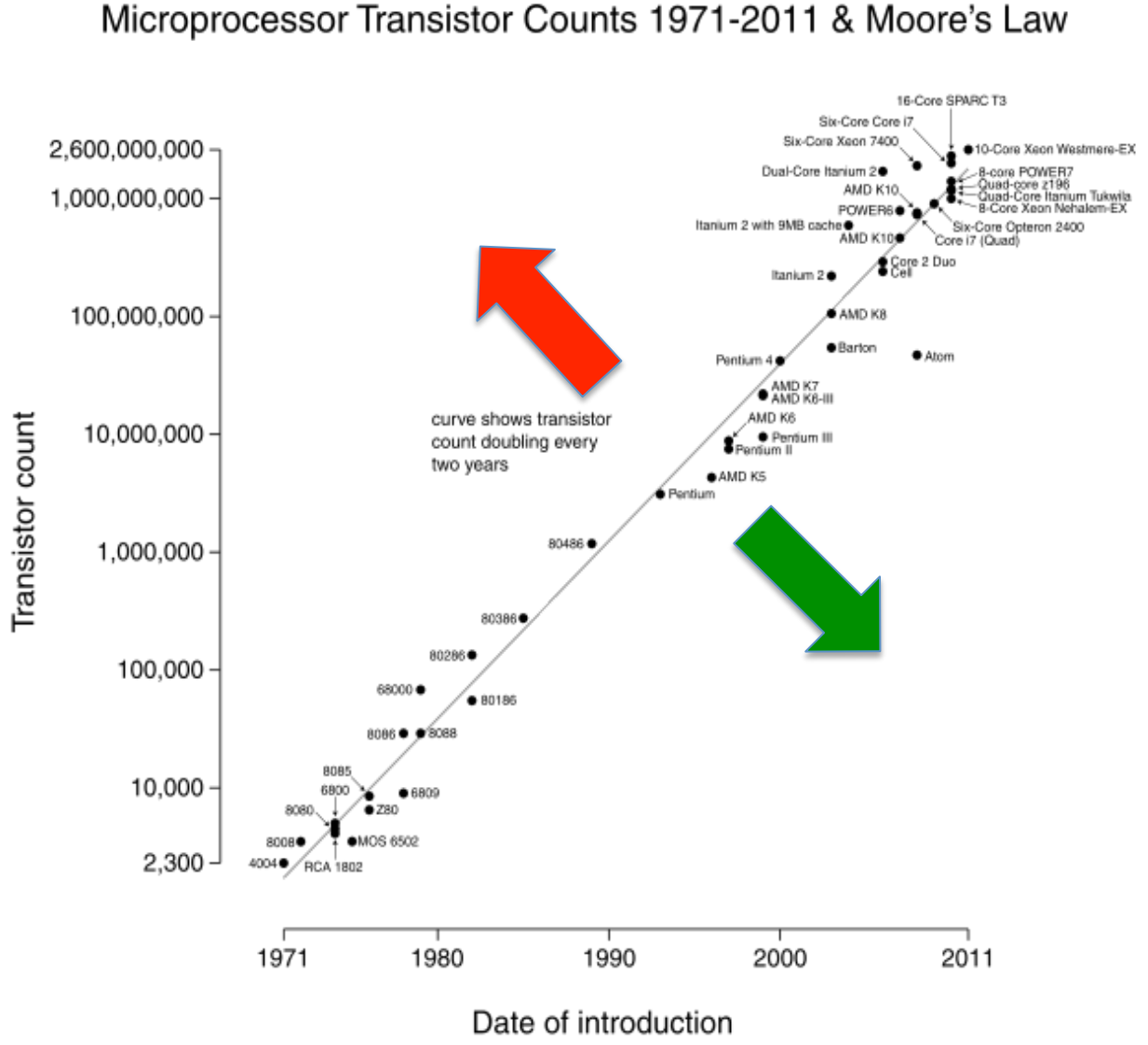
Jan 2011	4,000 entries
2012	8,000 entries
2013	11,500 entries
2014	16,300 entries
2015	21,800 entries
2016	28,300 entries

Up and to the Right

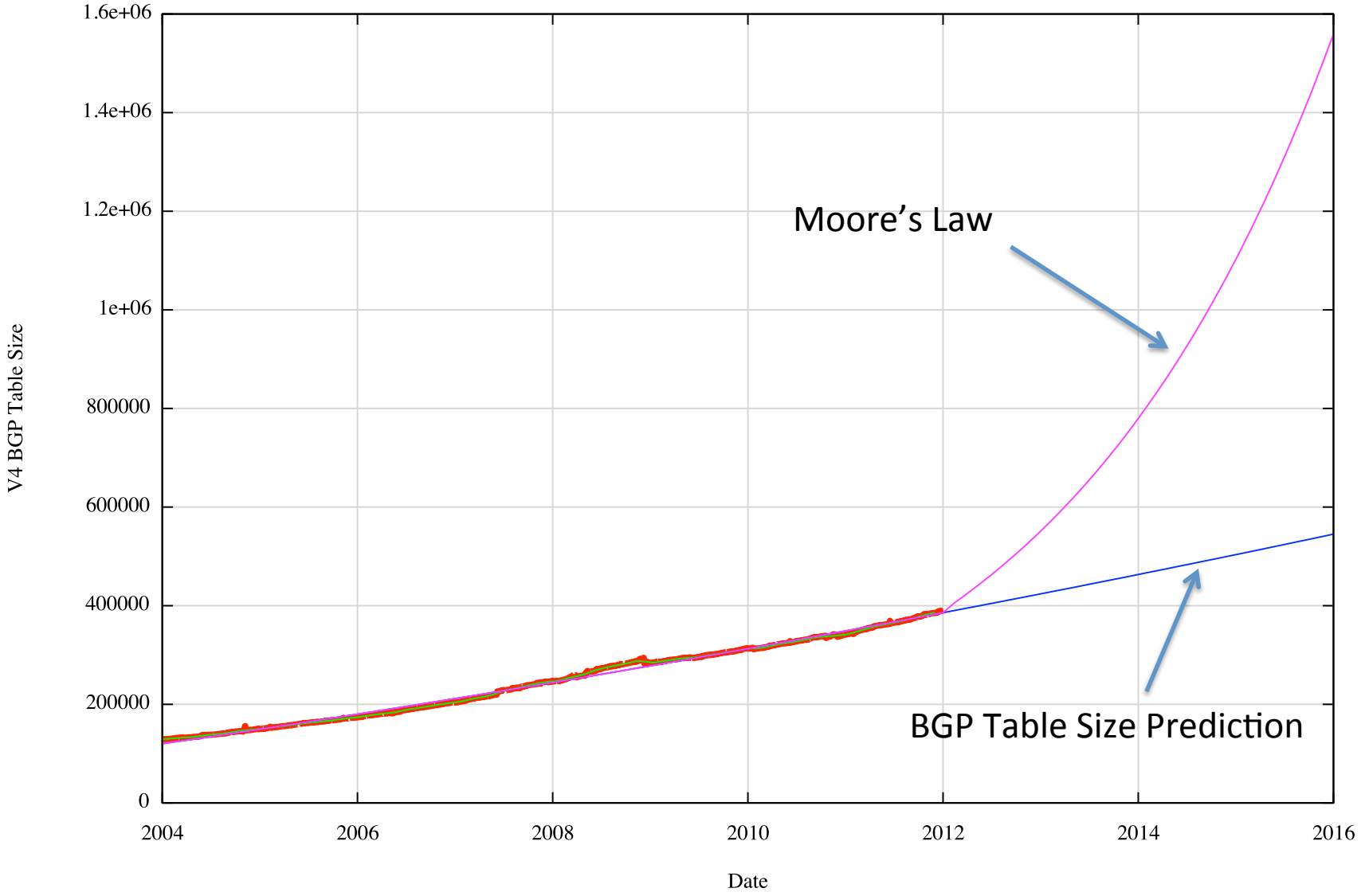
- Most Internet curves are “up and to the right”
- But what makes this curve painful?
  - The pain threshold is approximated by Moore’s Law

Moore's Law

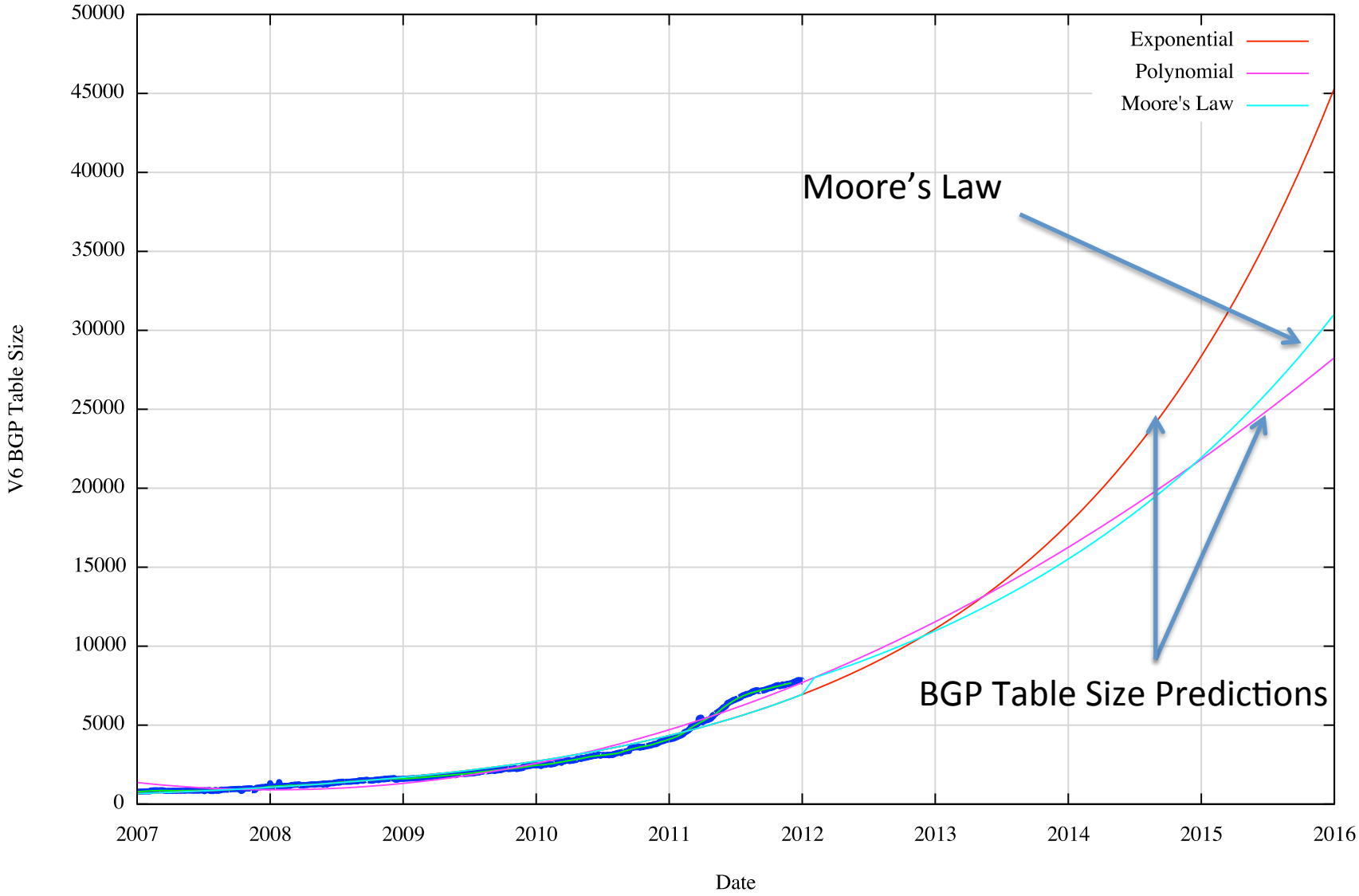
- As a rough rule of thumb, if the rate of growth of the table grows at a rate equal to, or less than Moore's Law, then the unit cost of storing the forwarding table should remain constant
  - Like all rough rules of thumb, there are many potential exceptions, and costs have many inputs as well as the 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
Quad-core z196
AMD K10
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 K8-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



eBGP Table Growth

- Nothing in these figures suggests that there is cause for urgent alarm -- at present
- The overall eBGP 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 it will continue to be viable for some time yet

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"

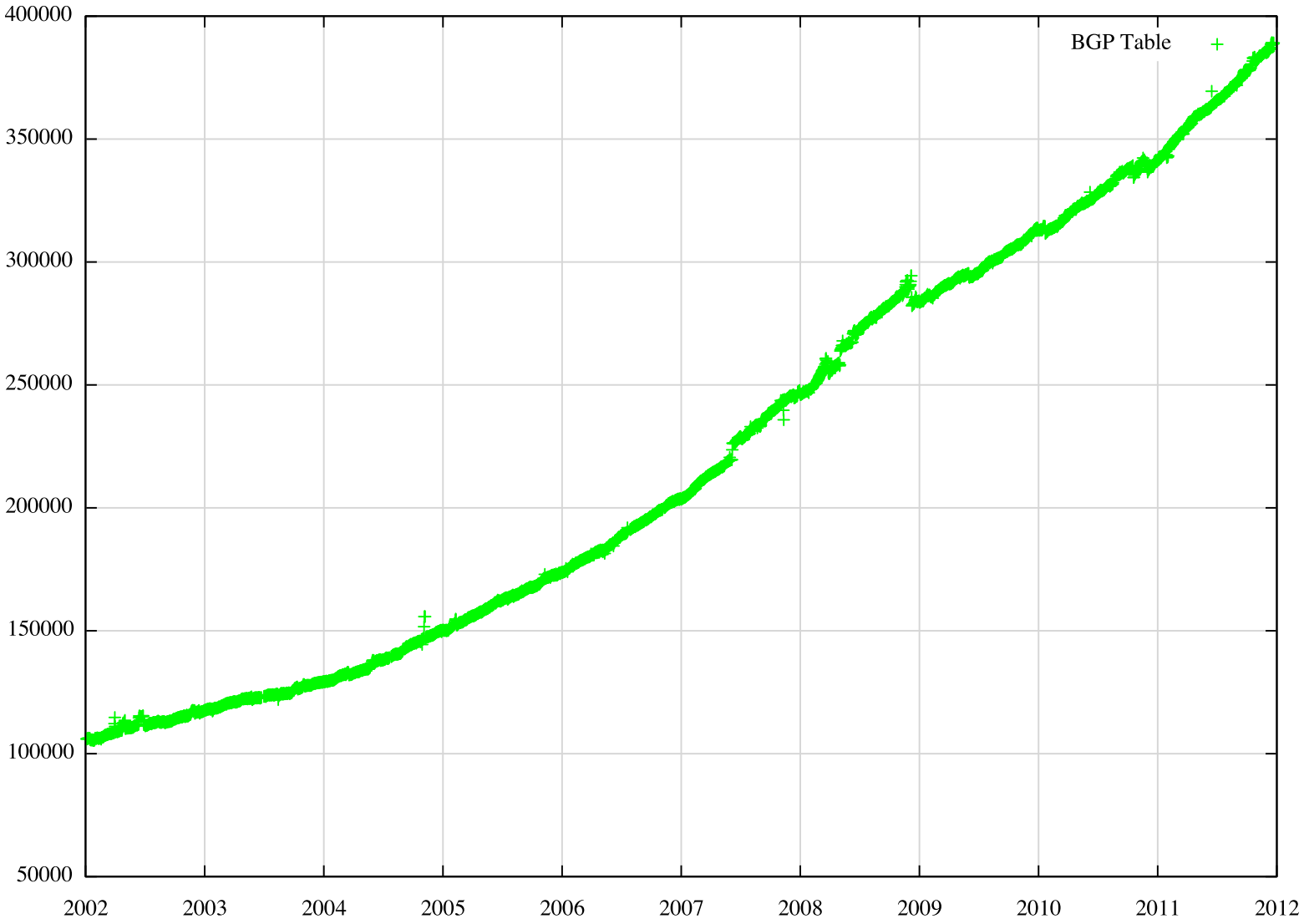
Who is doing this the most?

www.cidr-report.org

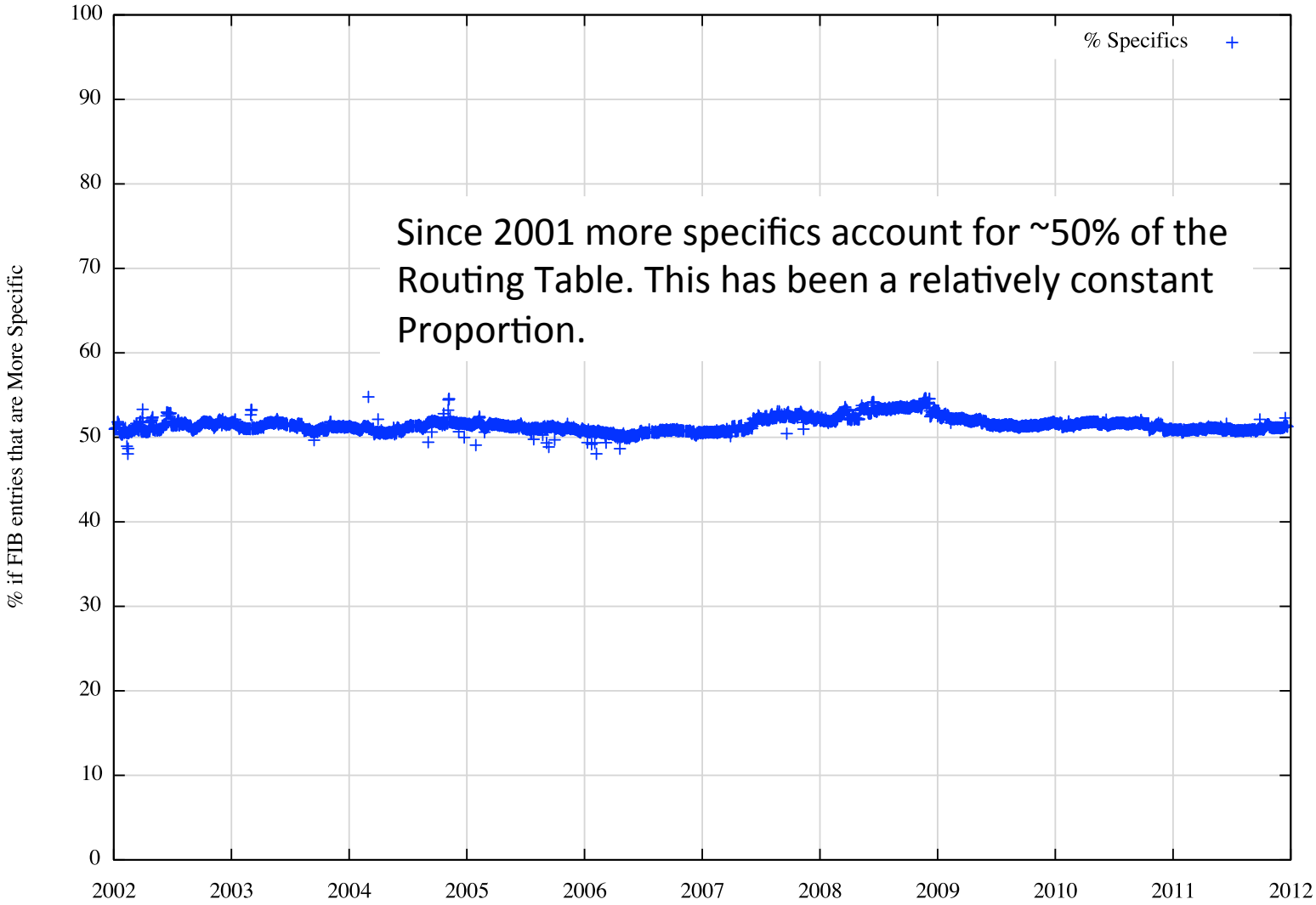
--- 23Dec11 ---

ASnum	NetsNow	NetsAggr	NetGain	% Gain	Description
Table	388637	227303	161334	41.5%	All ASes
AS6389	3473	223	**3250**	93.6%	BELLSOUTH-NET-BLK - BellSouth.net Inc.
AS18566	2093	412	**1681**	80.3%	COVAD - Covad Communications Co.
AS4766	2492	990	**1502**	60.3%	KIXS-AS-KR Korea Telecom
AS7029	2951	1521	**1430**	48.5%	WINDSTREAM - Windstream Communications Inc
AS22773	1515	116	**1399**	92.3%	Cox Communications Inc.
AS4755	1512	201	**1311**	86.7%	TATACOMM-AS TATA Communications
AS4323	1622	387	**1235**	76.1%	TWTC - tw telecom holdings, inc.
AS28573	1557	397	**1160**	74.5%	NET Servicos de Comunicao S.A.
AS10620	1719	641	**1078**	62.7%	Telmex Colombia S.A.
AS1785	1863	787	**1076**	57.8%	AS-PAETEC-NET - PaeTec Communications, Inc.

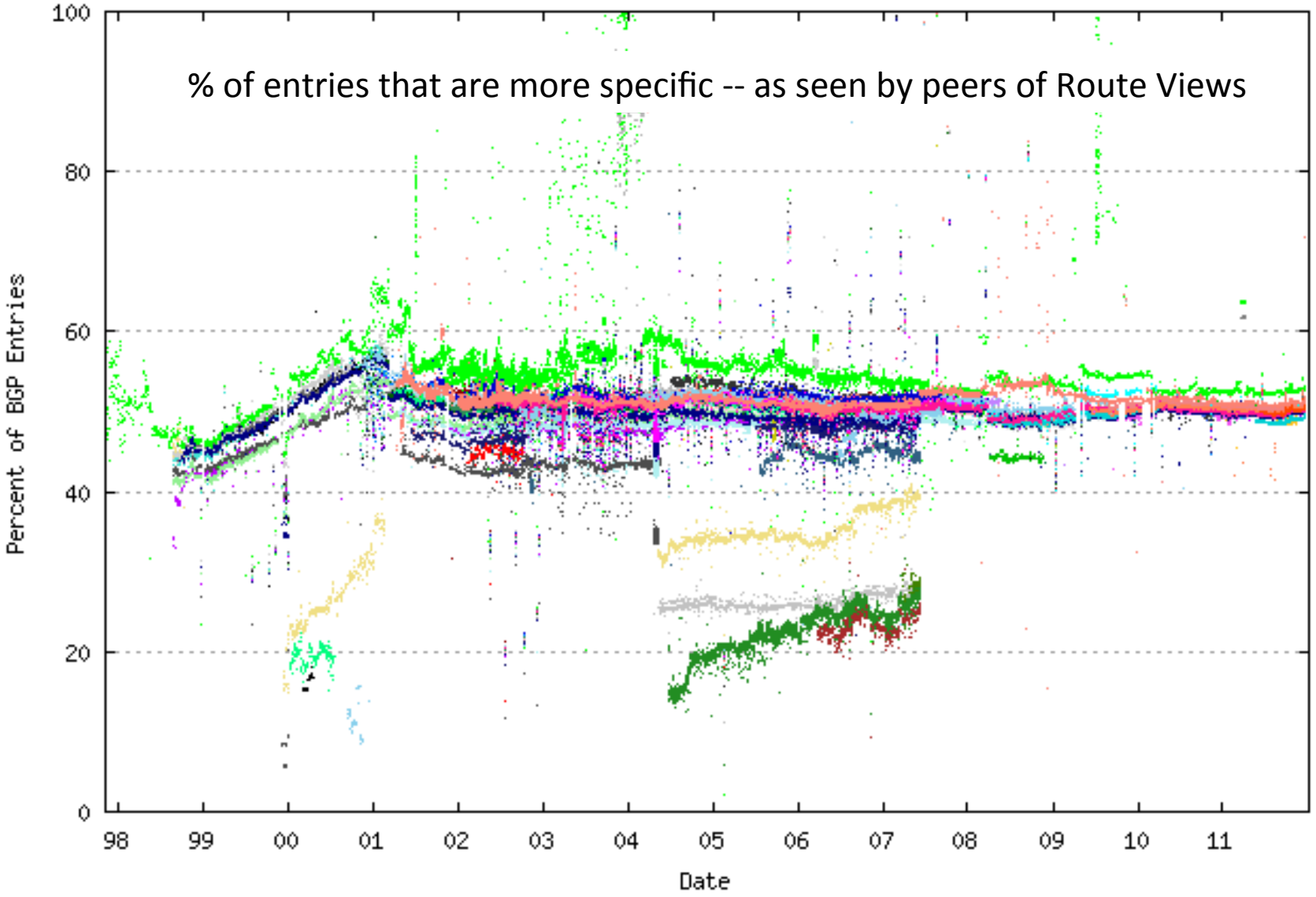
BGP Routing Table



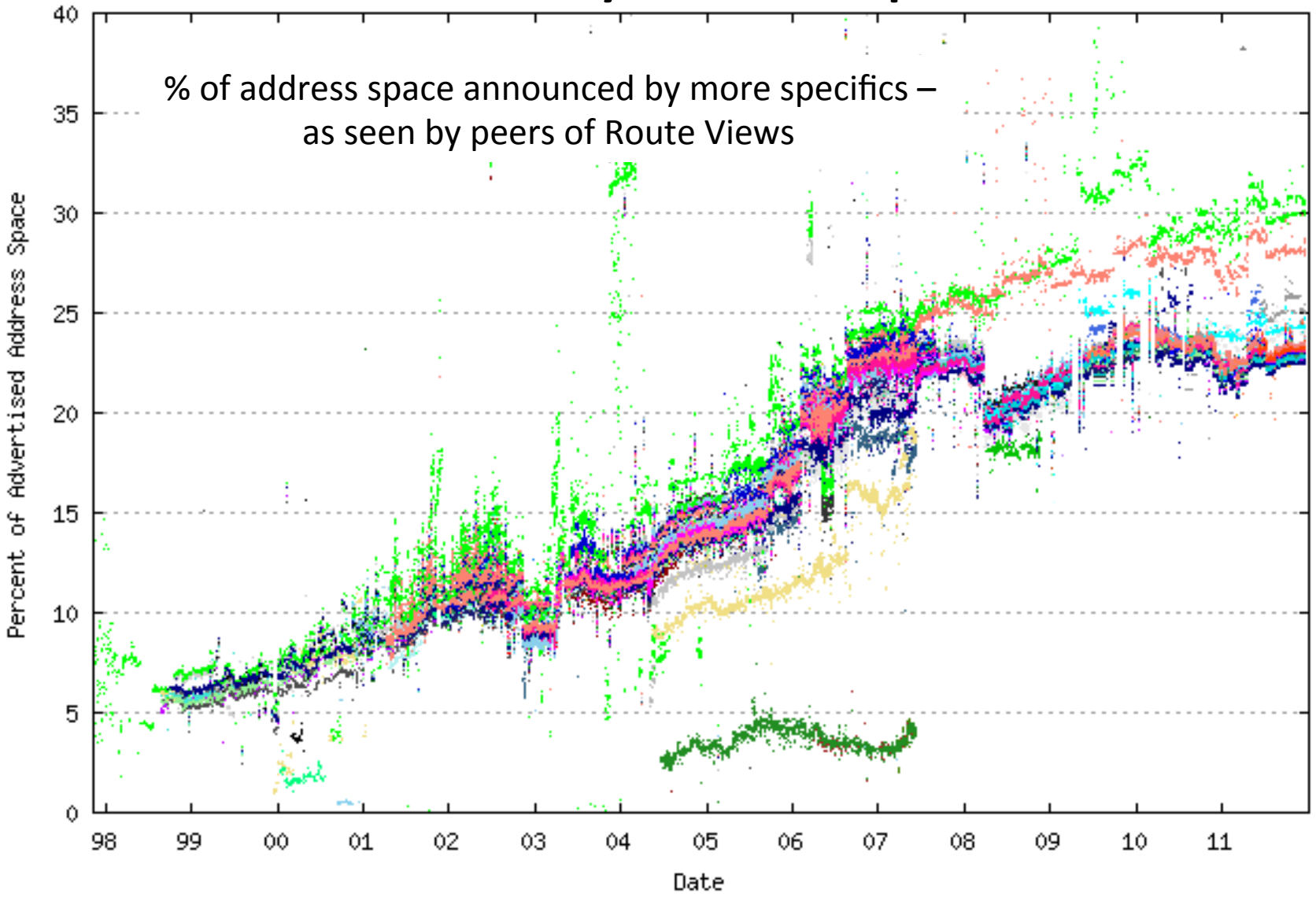
More specifics in the Routing Table



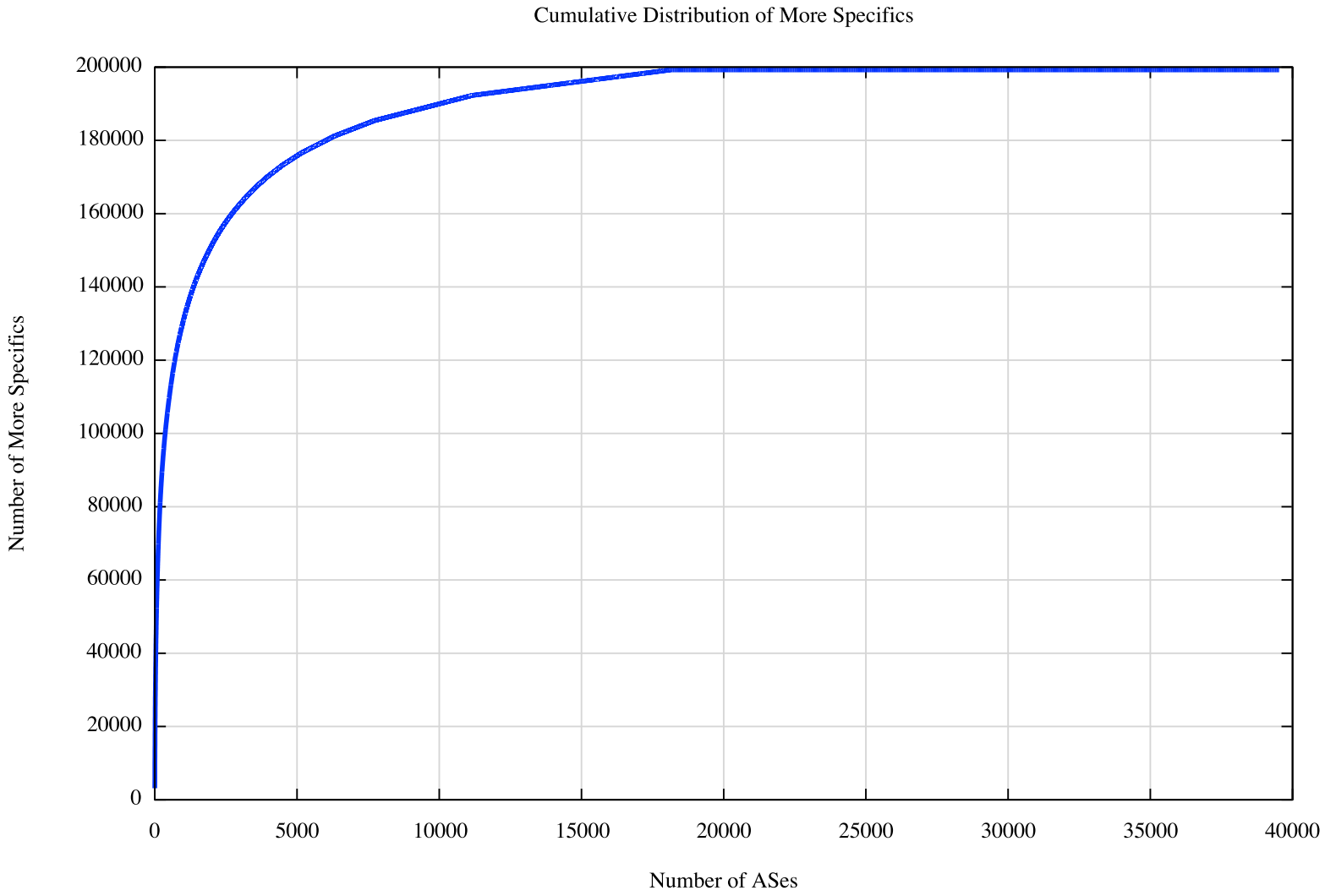
Does everyone see this?



How much address space is announced by more specifics?



Does everyone announce more specifics?



Is it Everyone?

- 3% of the ASes (1,186 ASes) announce 70% of the more specifics (136,023 announcements)
- 55% of the ASes announce no more specifics
- The top 10 ASes announce 19,163 more specifics

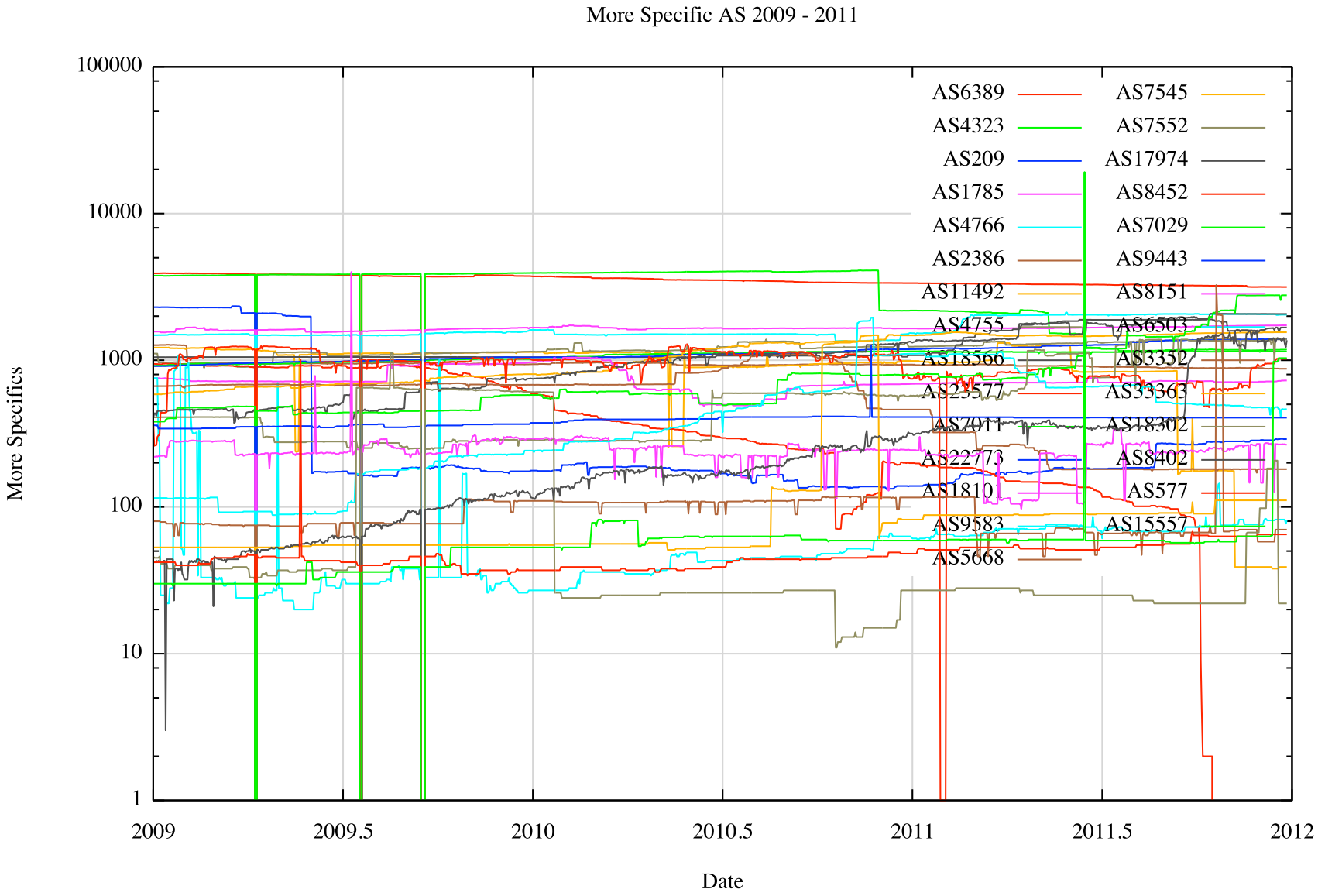
The Top 10 of More Specifics

AS	Aggregates	More Specifics	
6389	315	3155	BELLSOUTH-NET-BLK - BellSouth.net Inc.
7029	188	2770	WINDSTREAM - Windstream Communications
18566	25	2068	COVAD - Covad Communications Co.
4766	440	2043	KIX-AS-KR - Korea Telecom
1785	132	1731	AS-PAETEC-NET - PaeTec Communications
17974	44	1672	TELKOMNET-AS2-AP PT Telekomunikasi Indonesia
7545	78	1551	TPG-INTERNET-AP TPG Internet Pty Ltd
22773	118	1397	ASN-CXA-ALL-CCI-22773-RDC - Cox Communications
7552	31	1389	VIETEL-AS-AP Vietel Corporation
4755	127	1387	TATACOMM-AS TATA Communications

Are We Getting Any Better?

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

Yes ... and No


More Specific AS 2009 - 2011
More Specifics
100000
10000
1000
100
10
1
2009
2009.5
2010
2010.5
2011
2011.5
2012
Date
AS6389
AS4323
AS209
AS1785
AS4766
AS2386
AS11492
AS4755
AS18566
AS23577
AS7911
AS22773
AS1810
AS9583
AS5668
AS7545
AS7552
AS17974
AS8452
AS7029
AS9443
AS8151
AS6503
AS3352
AS33363
AS18302
AS8402
AS577
AS15557

Are We Getting Any Better?

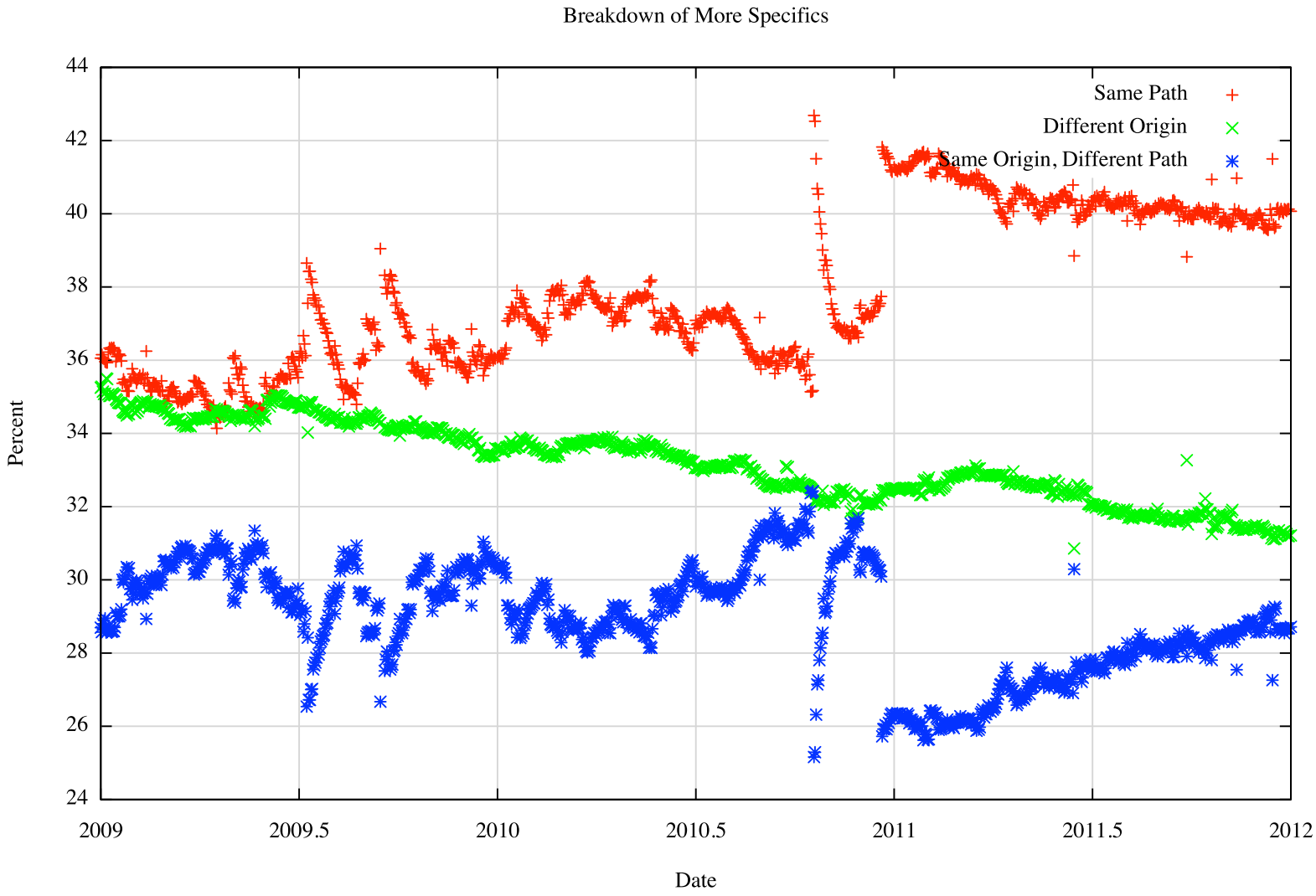
- Some ASes are effectively reducing the number of more specifics that are advertised into the global routing system
- Some ASes are increasing the number of more specifics
- And some are consistently advertising a significant number of more specifics
- There is no net change in the overall distribution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

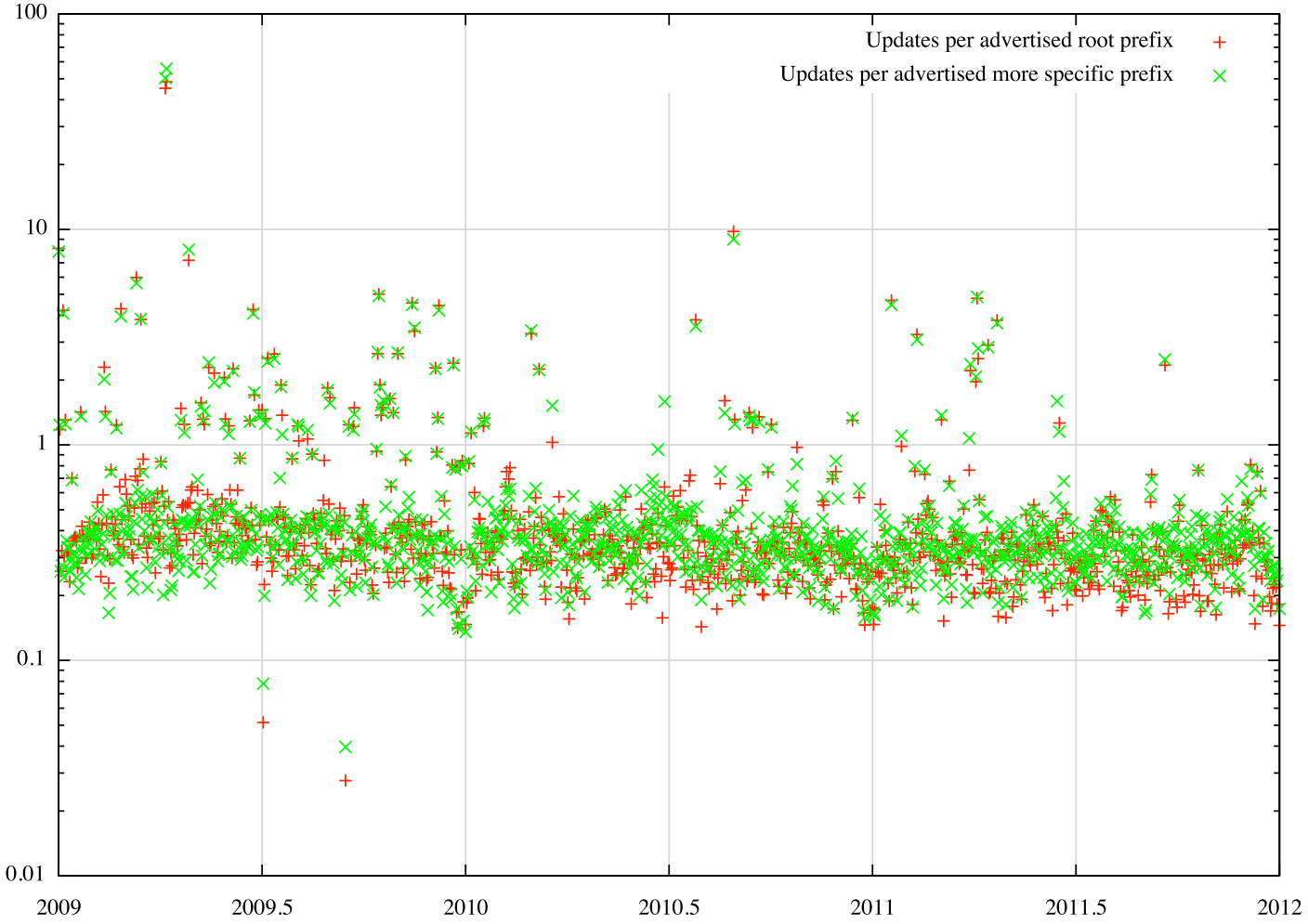
Types of More Specifics



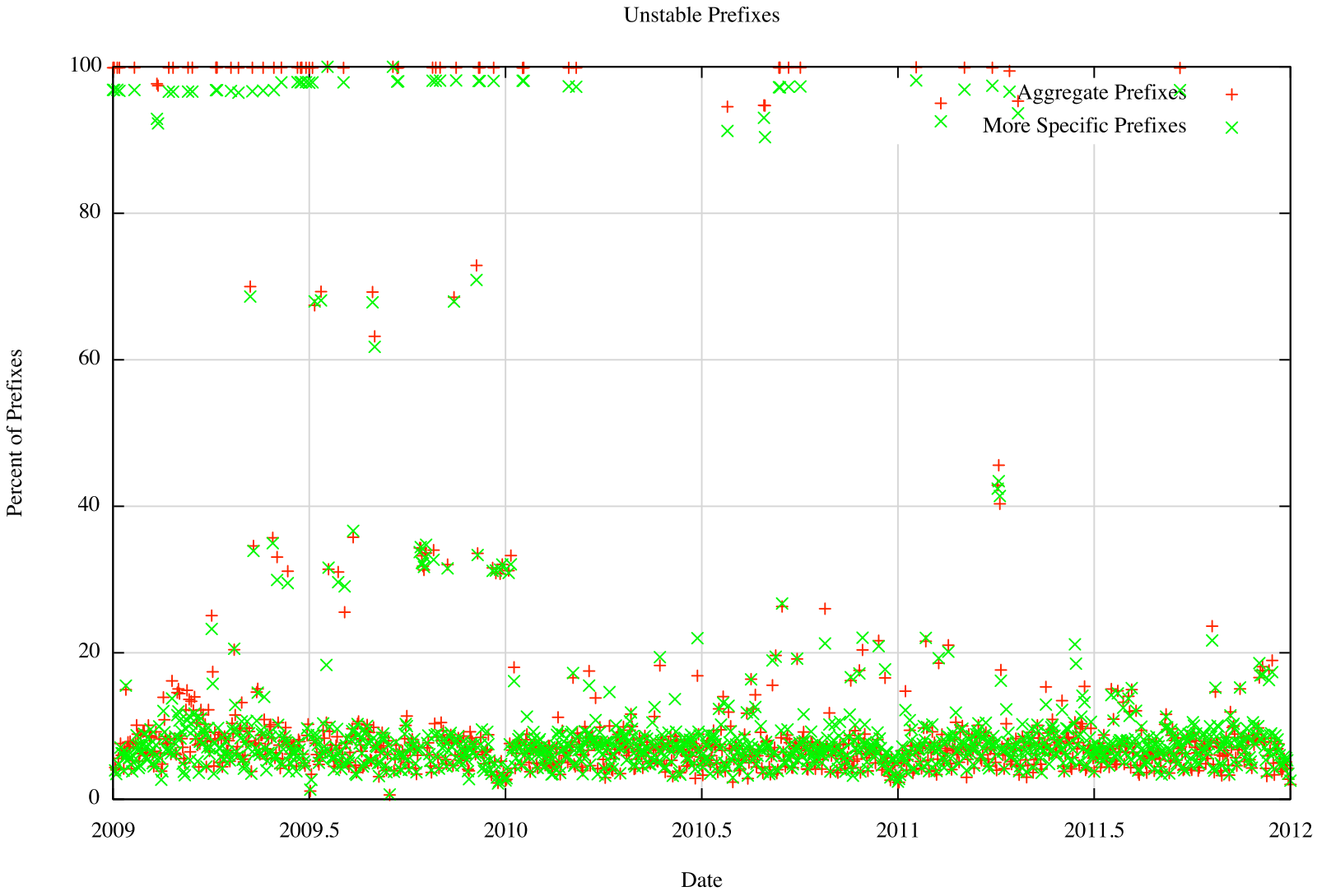
Daily Update Rates

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

Daily Update Rates



Prefix Instability Rates



Daily Update Rates

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

No!

This result is surprising – it was anticipated that more specifics would show a higher level of dynamic instability, particularly relating to TE more specifics. However nothing is visible in the data that supports this – advertised “root” prefixes are equally likely to be unstable as advertised more specific prefixes.

Problem? Not a Problem?

- Its evident that the global BGP routing environment suffers from a certain amount of neglect and inattention
- *Could* we do better?

  Yes!
- *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

Thank You

Questions?