BGP 2010-2011

Geoff Huston
APNIC R&D

Conventional (Historical) BGP Wisdom

IAB Workshop on Inter-Domain routing in October 2006 – RFC 4984:

> **"routing scalability is the most important problem facing the Internet today and must be solved"**

Measuring BGP

There are a number of ways to "measure" BGP:

1. Assemble a large set of BGP peering sessions and record everything
   - RIPE NCC's RIS service
   - Route Views
2. Perform carefully controlled injections of route information and observe the propagation of information
   - Beacons
   - AS Set manipulation
   - Bogon Detection and Triangulation
3. Take a single BGP perspective and perform continuous recording of a number of BGP metrics over a long baseline

Measuring BGP

There are a number of ways to "measure" BGP:

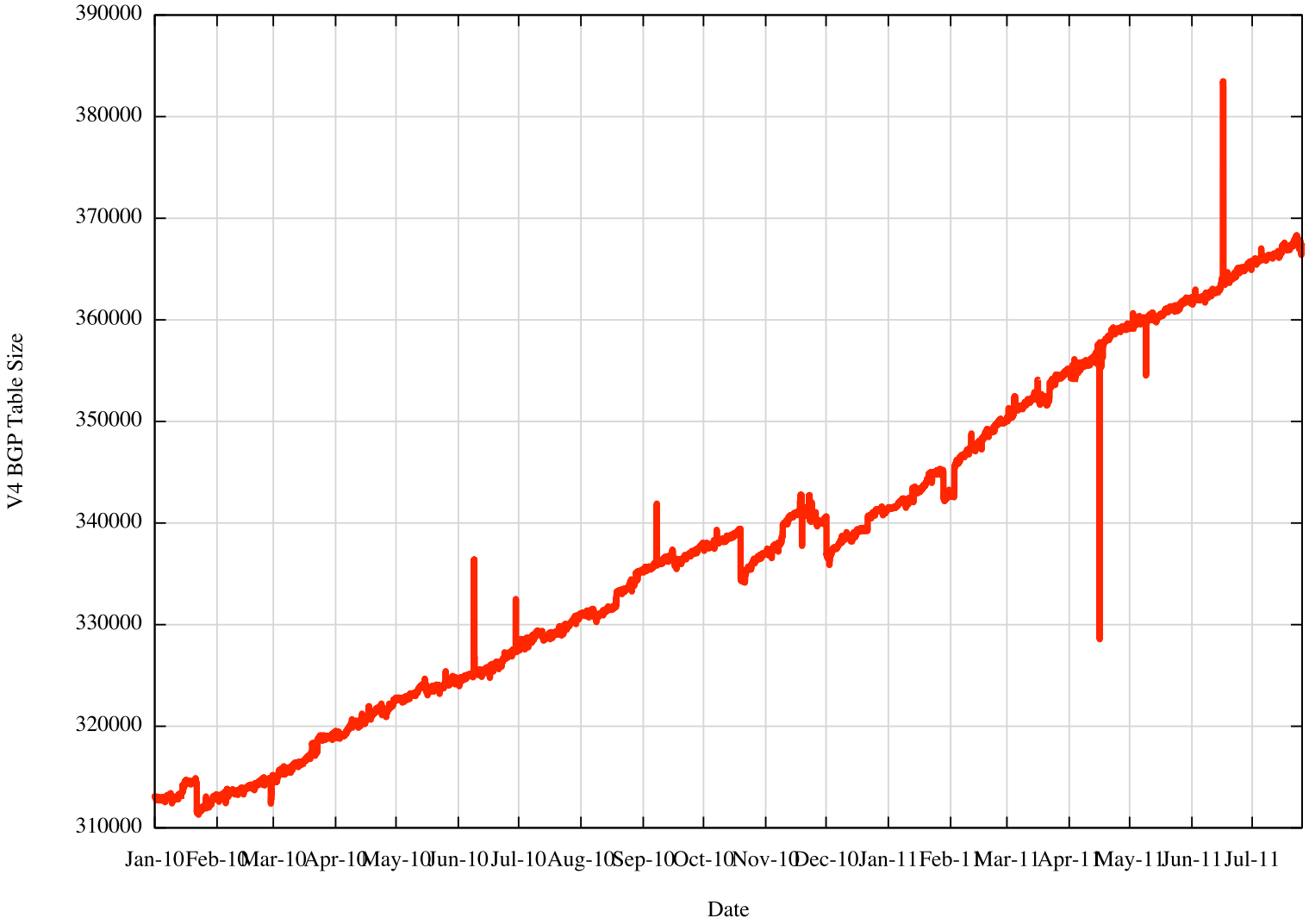
1. Assemble a large set of BGP peering sessions and record everything
   - RIPE NCC's RIS service
   - Route Views
2. Perform carefully controlled injections of route information and observe the propagation of information
   - Beacons
   - AS Set manipulation
   - Bogon Detection and Triangulation
3. Take a single BGP perspective and perform continuous recording of a number of BGP metrics over a long baseline

AS131072 BGP Measurements

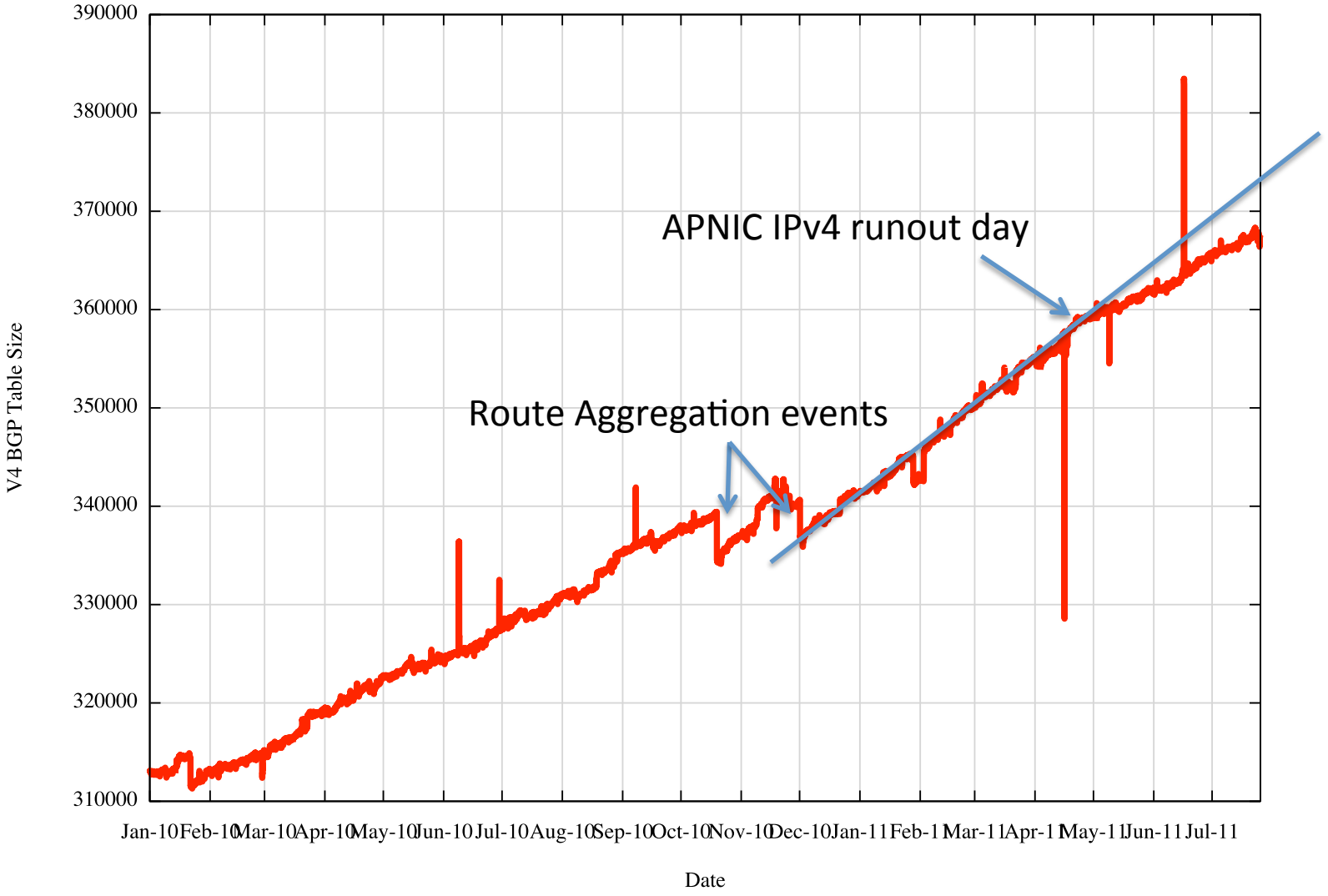
- Data collection since 1 July 2007
- Passive data measurement technique (no advertisements or probes)
- Quagga platform, connected to AS4608
- Dual Stack BGP
- Archive of all BGP updates and daily RIB dumps
- Data and reports are continuously updated and published: http://bgp.potaroo.net

BGP in 2010-2011

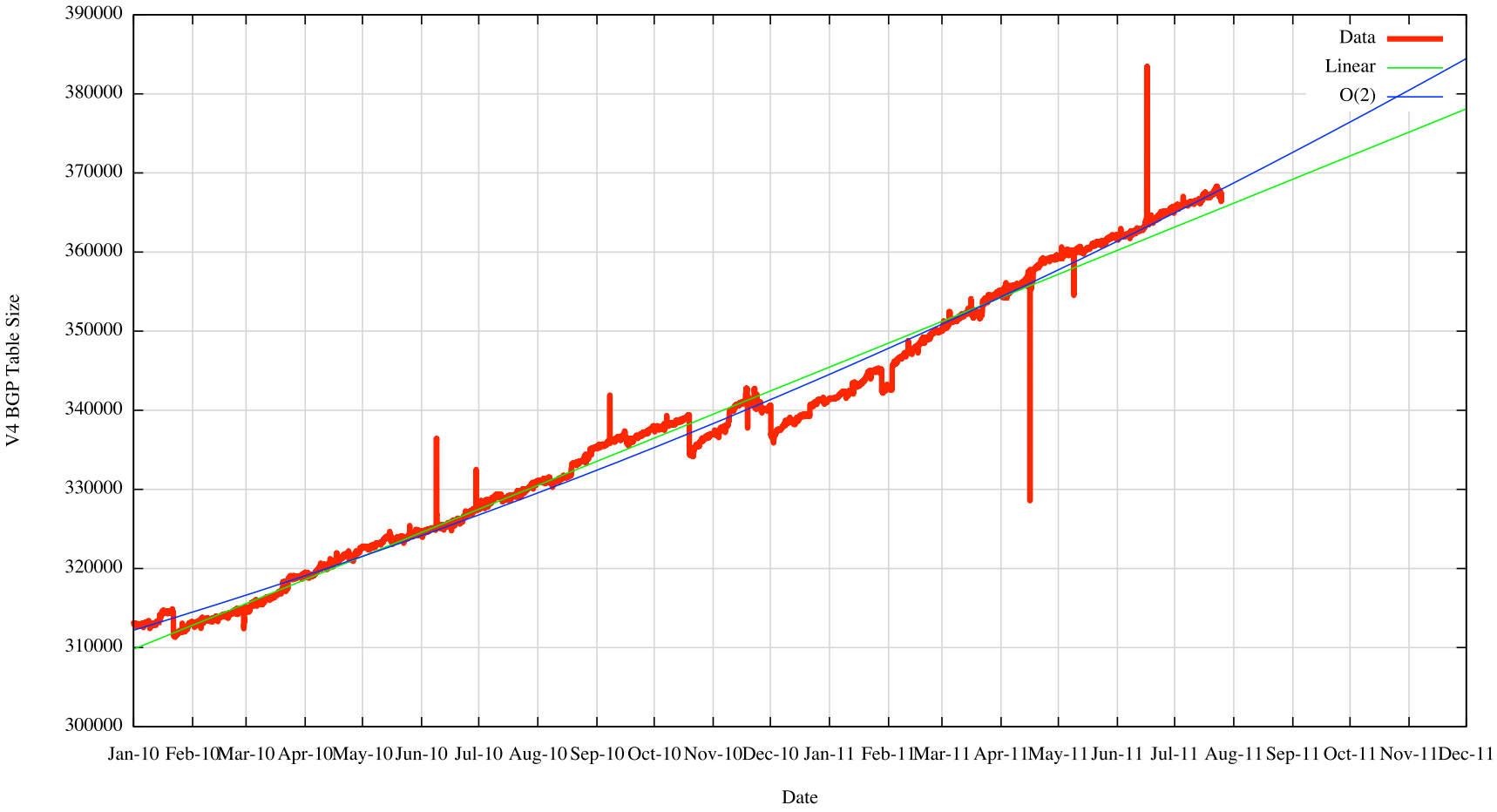
IPv4 BGP Prefix Count



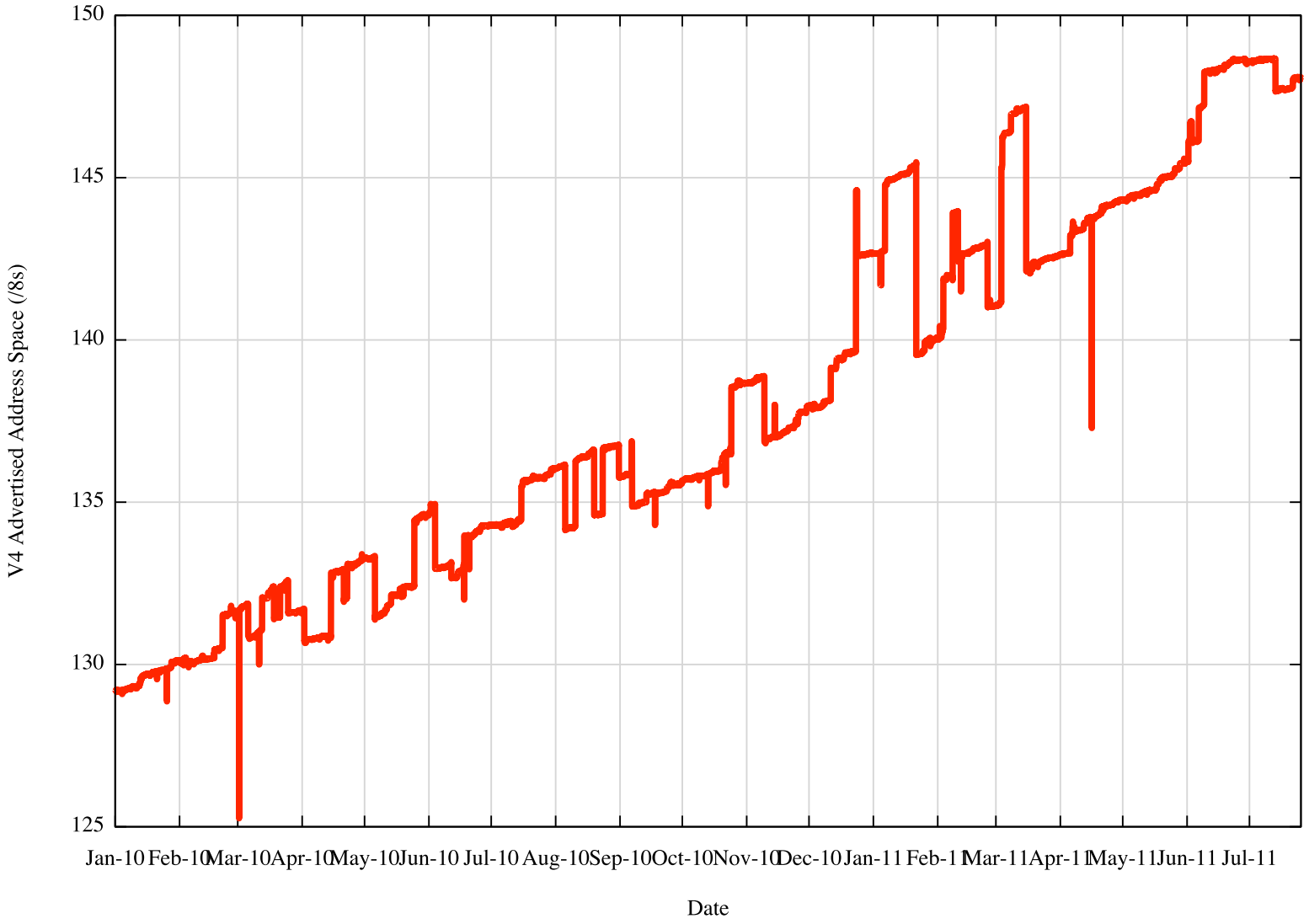
IPv4 BGP Prefix Count



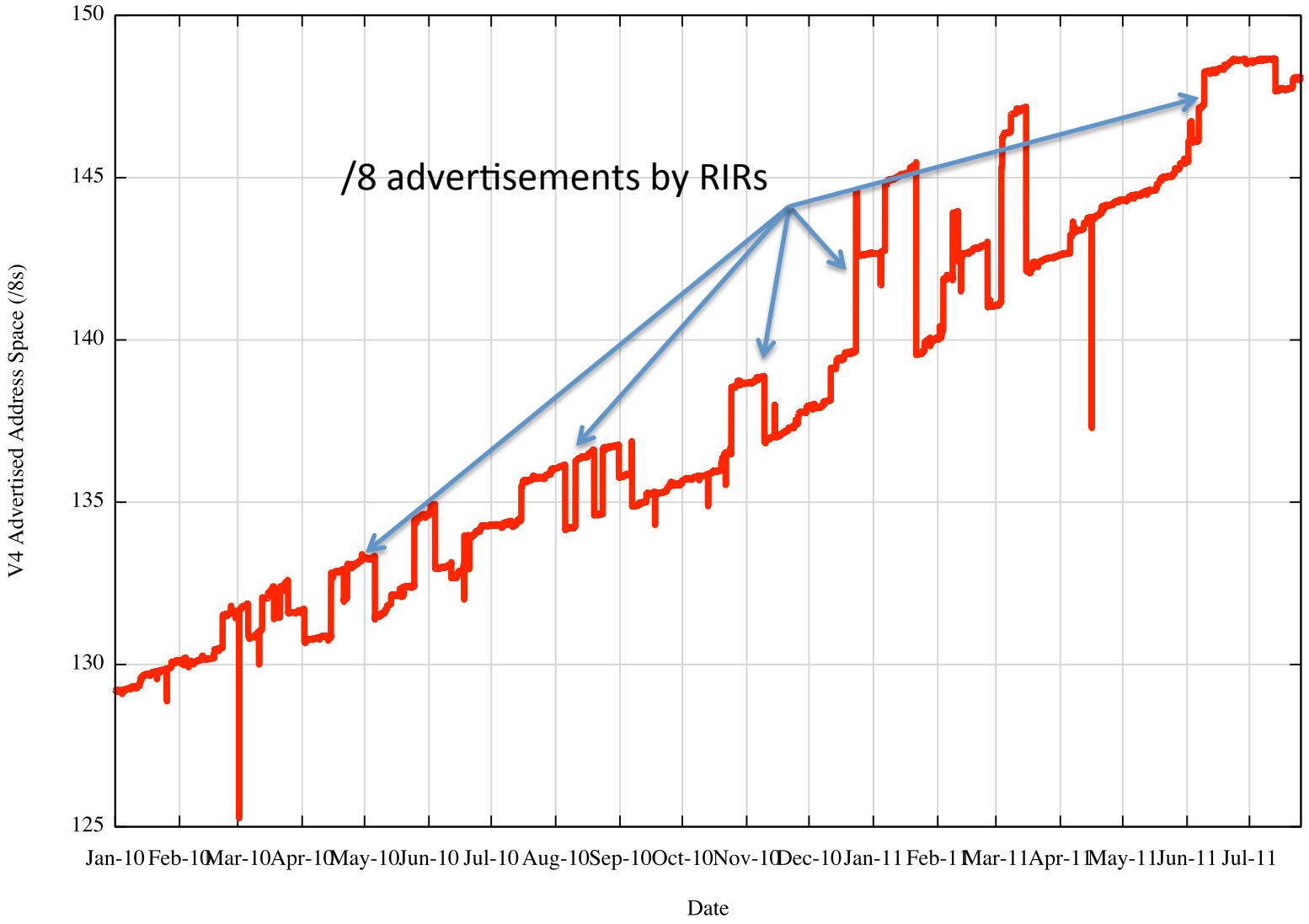
IPv4 BGP Prefix Count



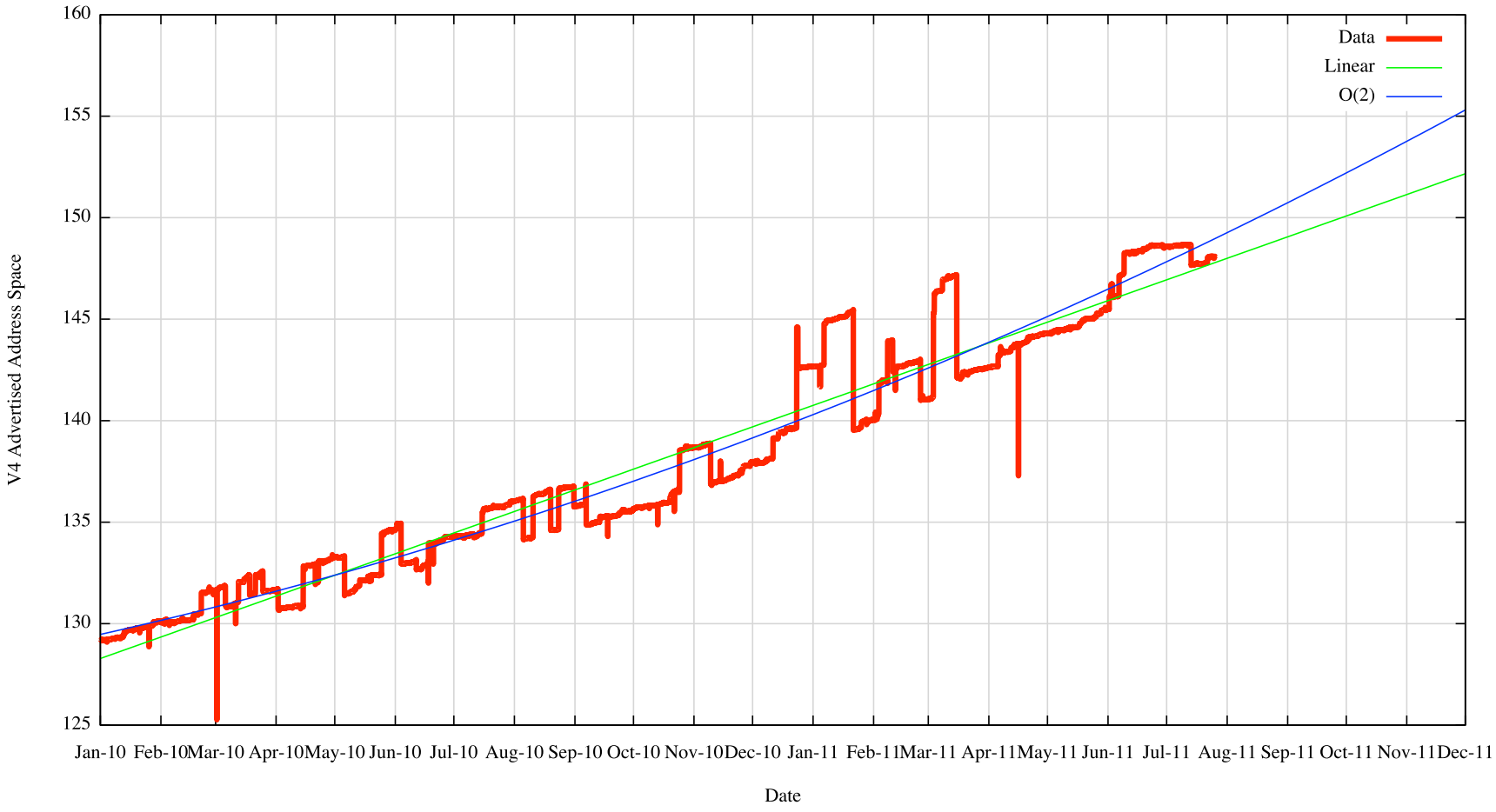
IPv4 Routed Address Span



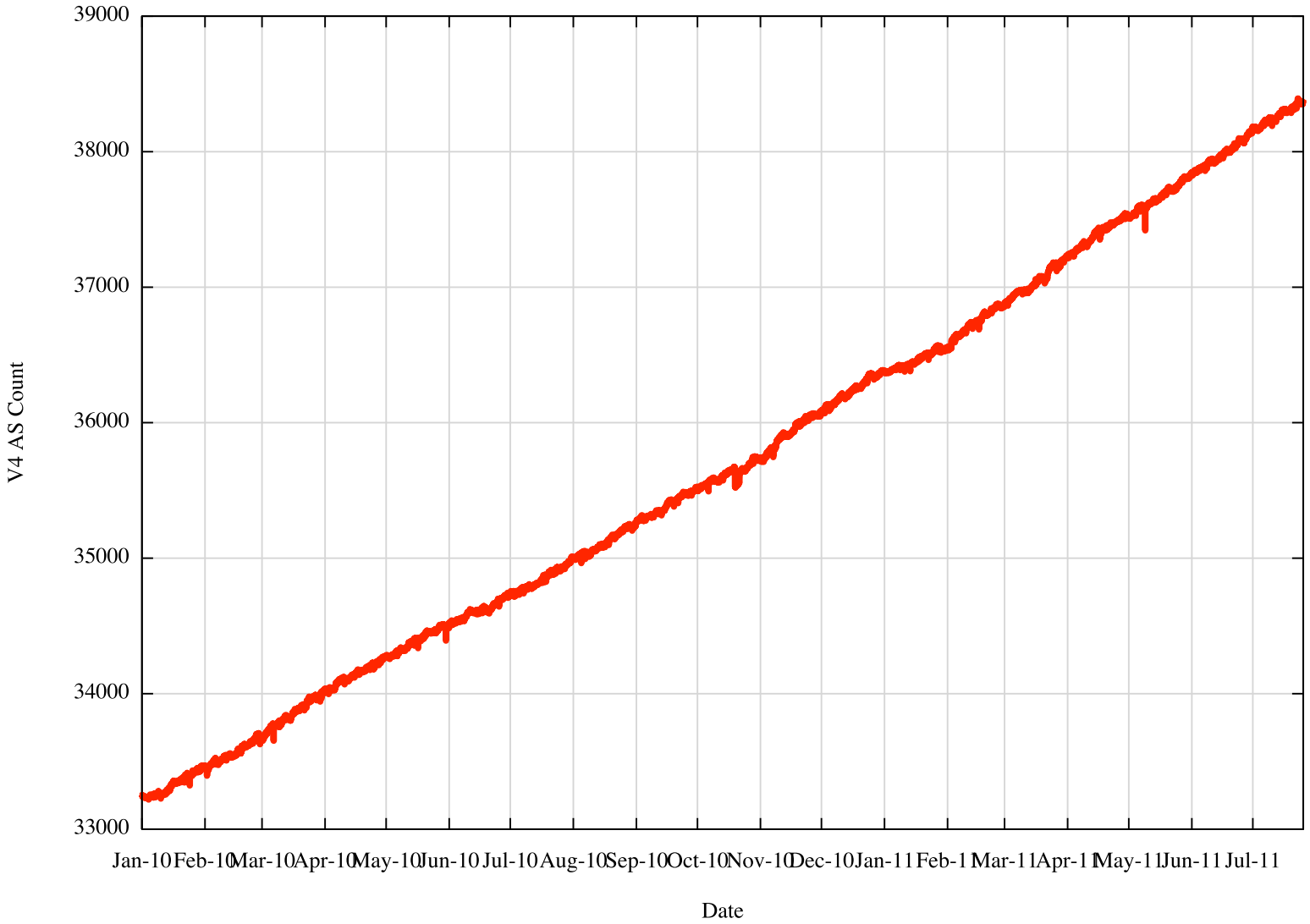
IPv4 Routed Address Span



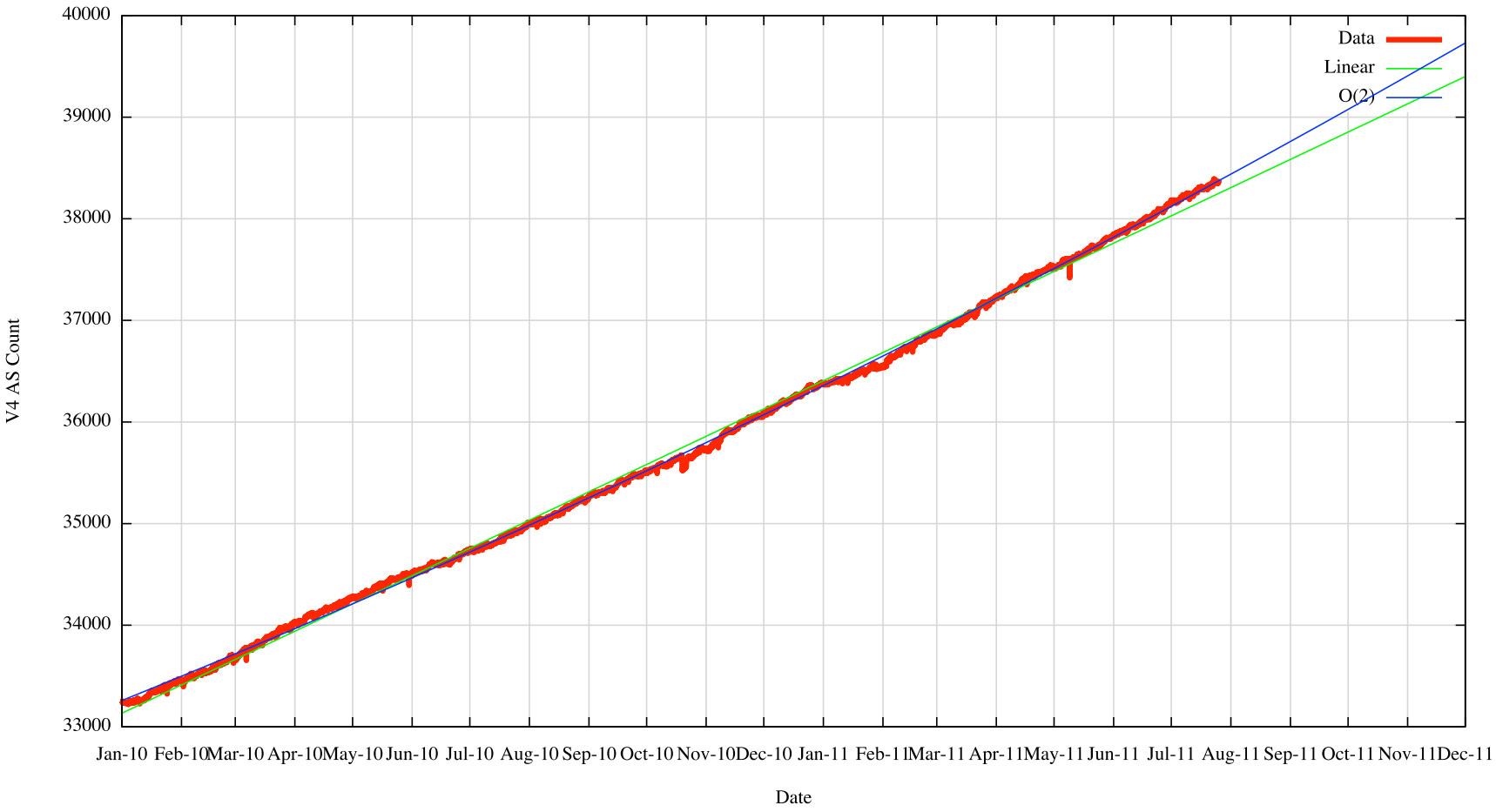
IPv4 Routed Address Span



IPv4 Routed AS Count



IPv4 Routed AS Count



IPv4 2010 BGP Vital Statistics

	Jan-10	Jan-11	
Prefix Count	313,000	341,000	+ **9%**
Roots	151,000	168,000	+11%
More Specifics	162,000	174,000	+ 7%
Address Span	129/8s	143/8s	+ **11%**
AS Count	33,200	36,400	+**10%**
Transit	4,400	5,200	+18%
Stub	28,800	31,200	+ 8%

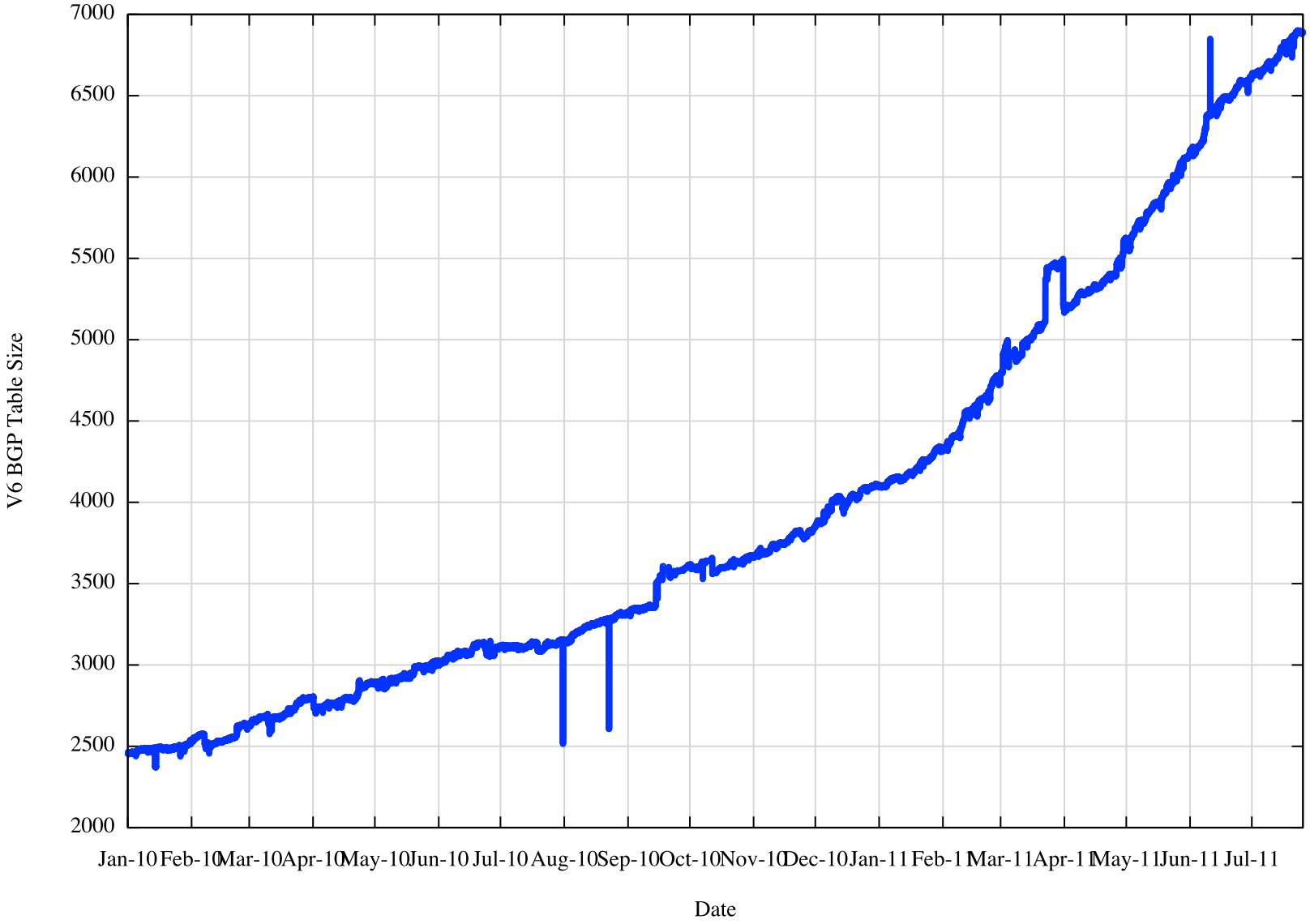
IPv4 2011 BGP Vital Statistics (to July)

	Jan-10	Jan-11	Jul-11	p.a. rate
Prefix Count	313,000	341,000	368,000	**+14%**
Roots	151,000	168,000	181,000	+13%
More Specifics	162,000	174,000	186,000	+13%
Address Span	129/8s	143/8s	148/8s	**+11%**
AS Count	33,200	36,400	38,400	**+10%**
Transit	4,400	5,200	5,300	+18%
Stub	28,800	31,200	32,900	+ 9%

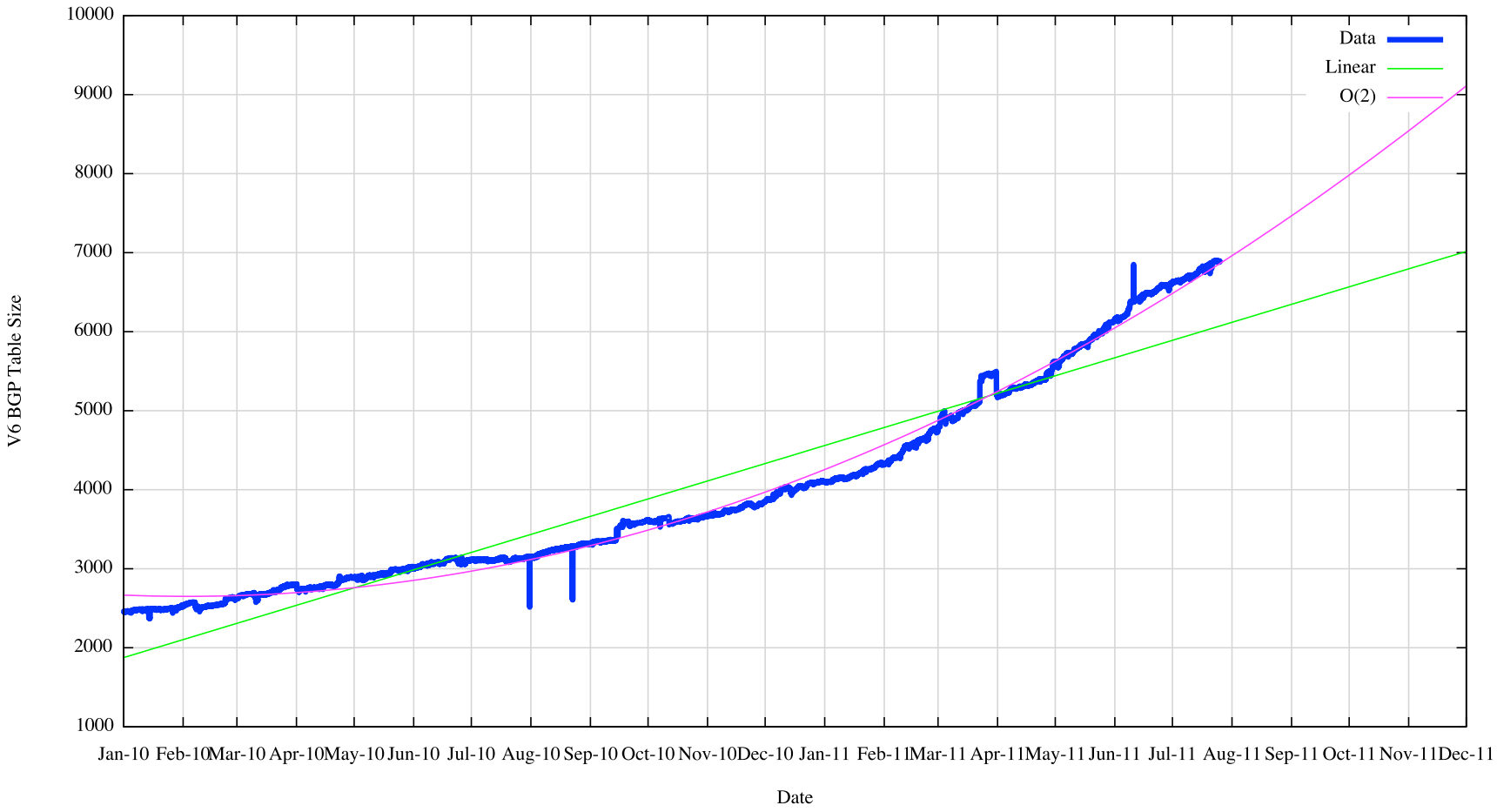
IPv4 in 2010 / 2011

- Overall Internet growth in terms of BGP is at a rate of some 10% p.a.
  - This is the same as 2009.
- Growth has slowed since 20 April 2011, following APINC's IPv4 address run out
- Address fragmentation is slowing
- Transit ASs growing faster than Stub AS count

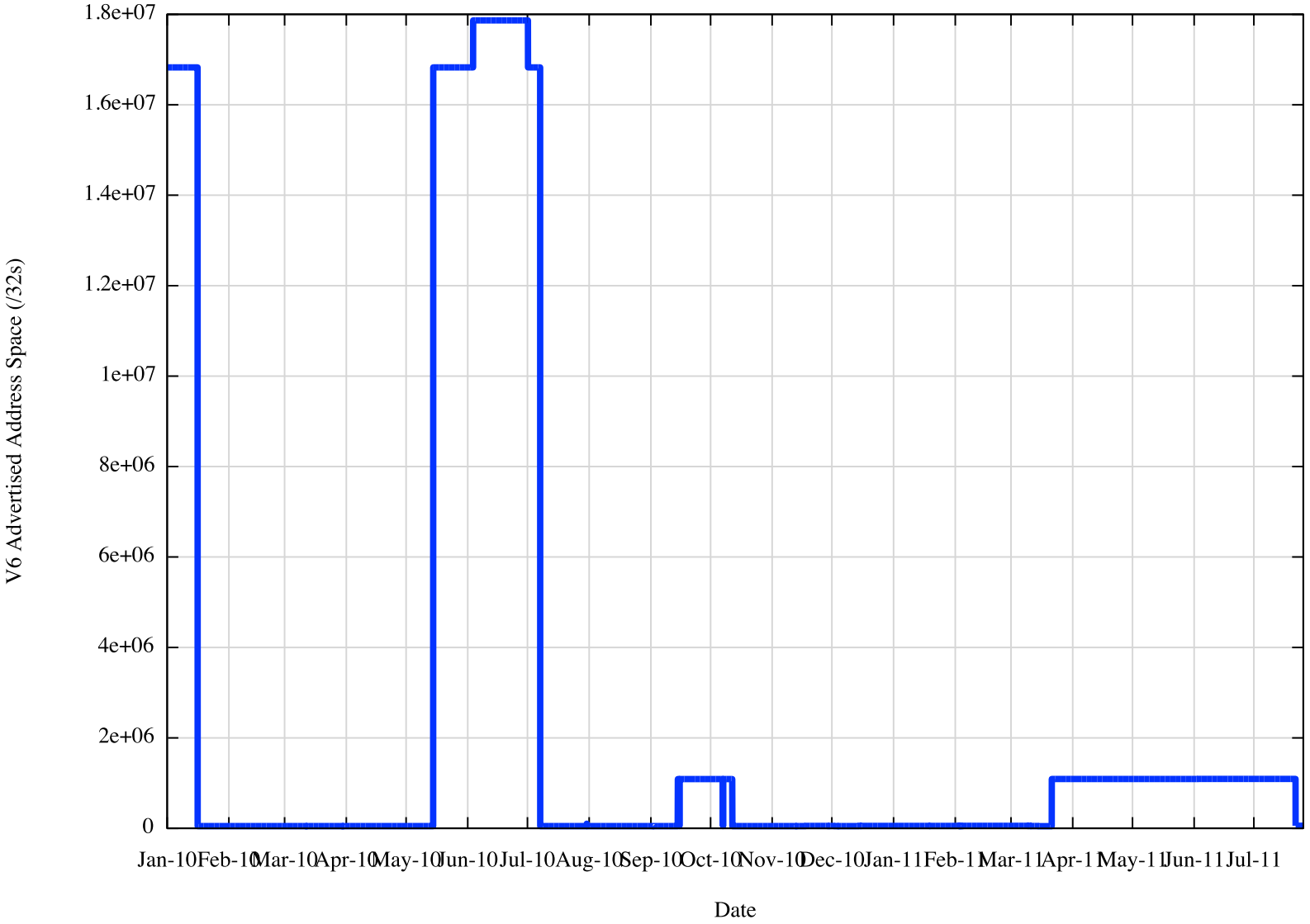
IPv6 BGP Prefix Count



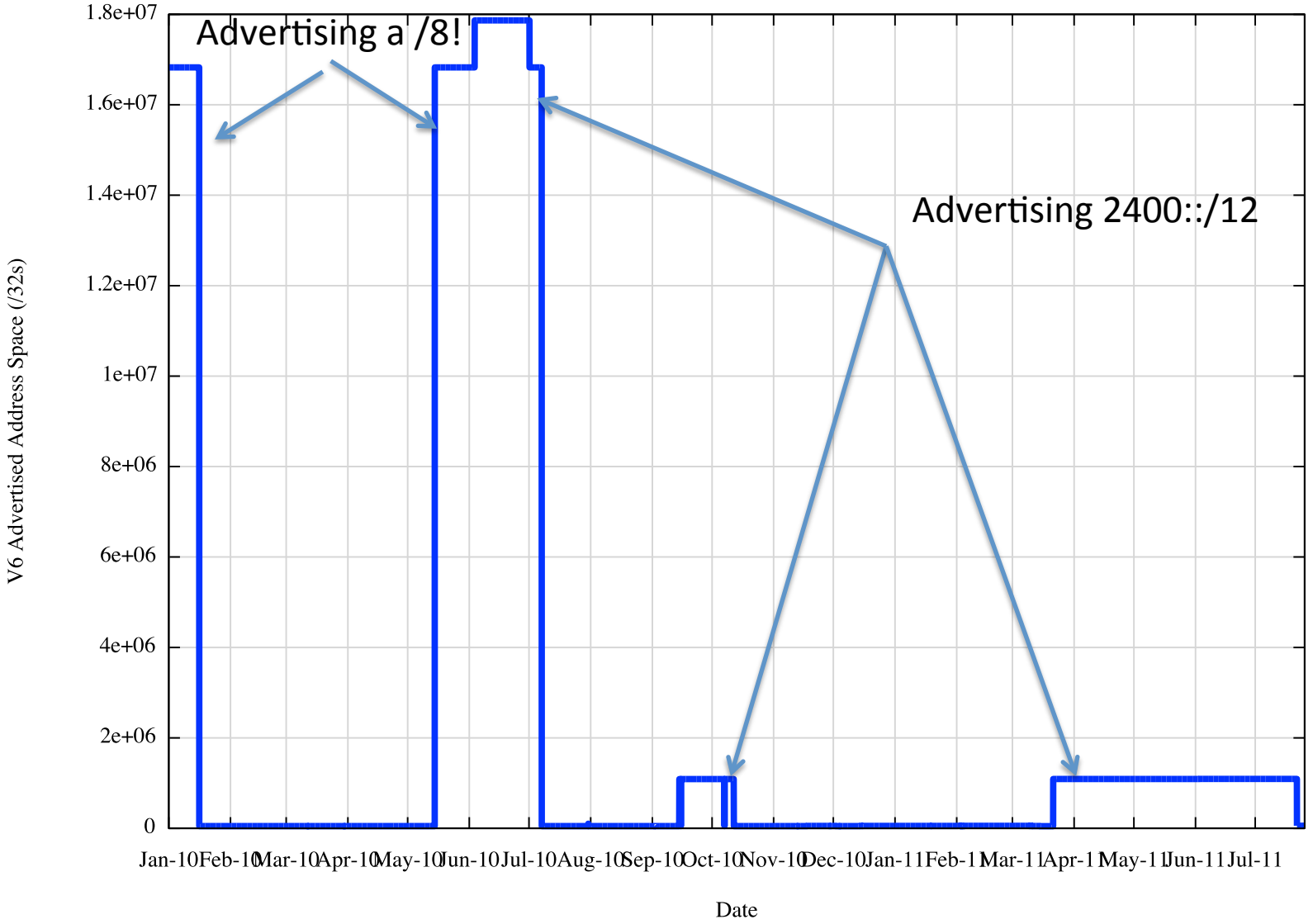
IPv6 BGP Prefix Count



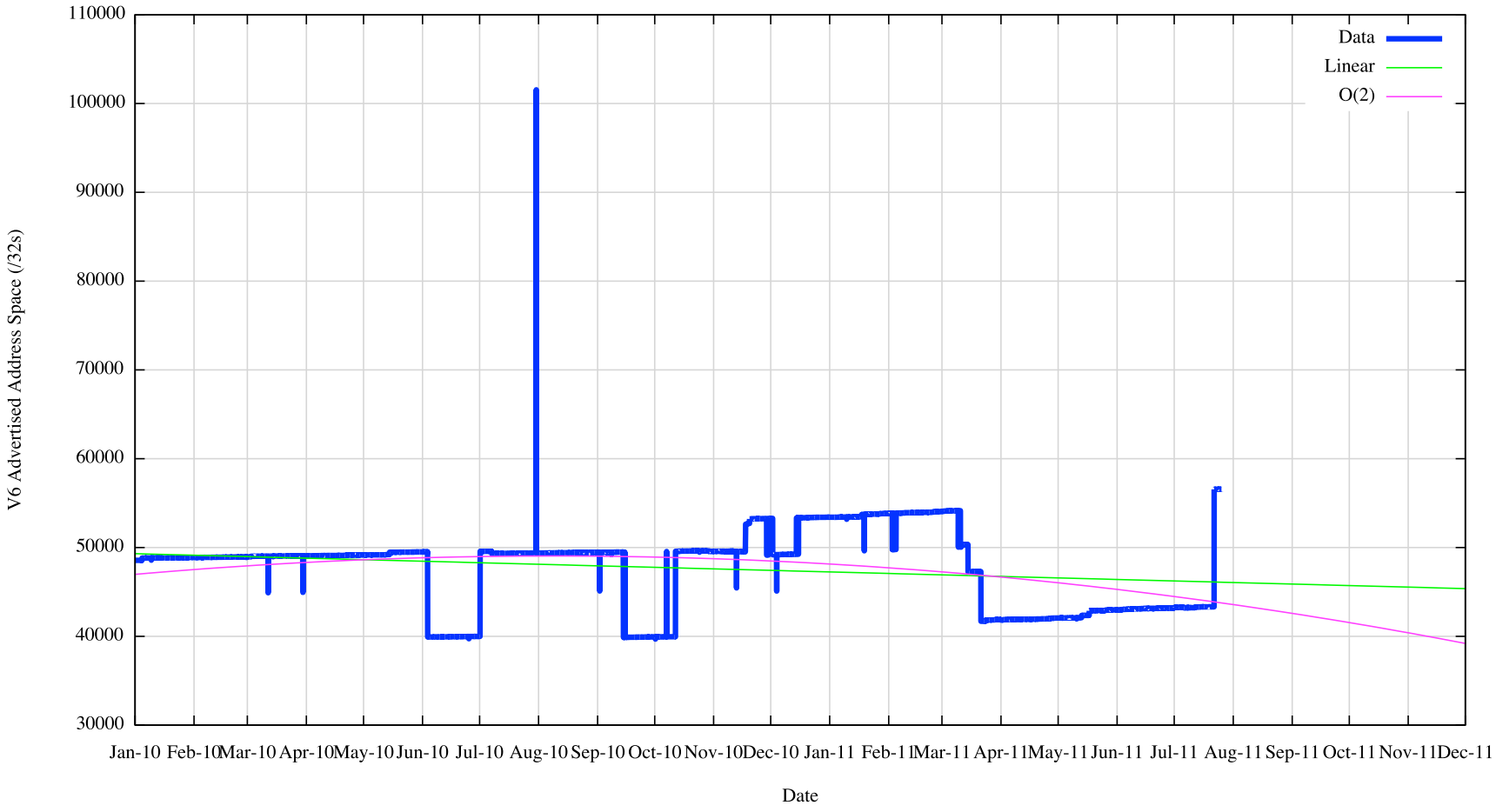
IPv6 Routed Address Span



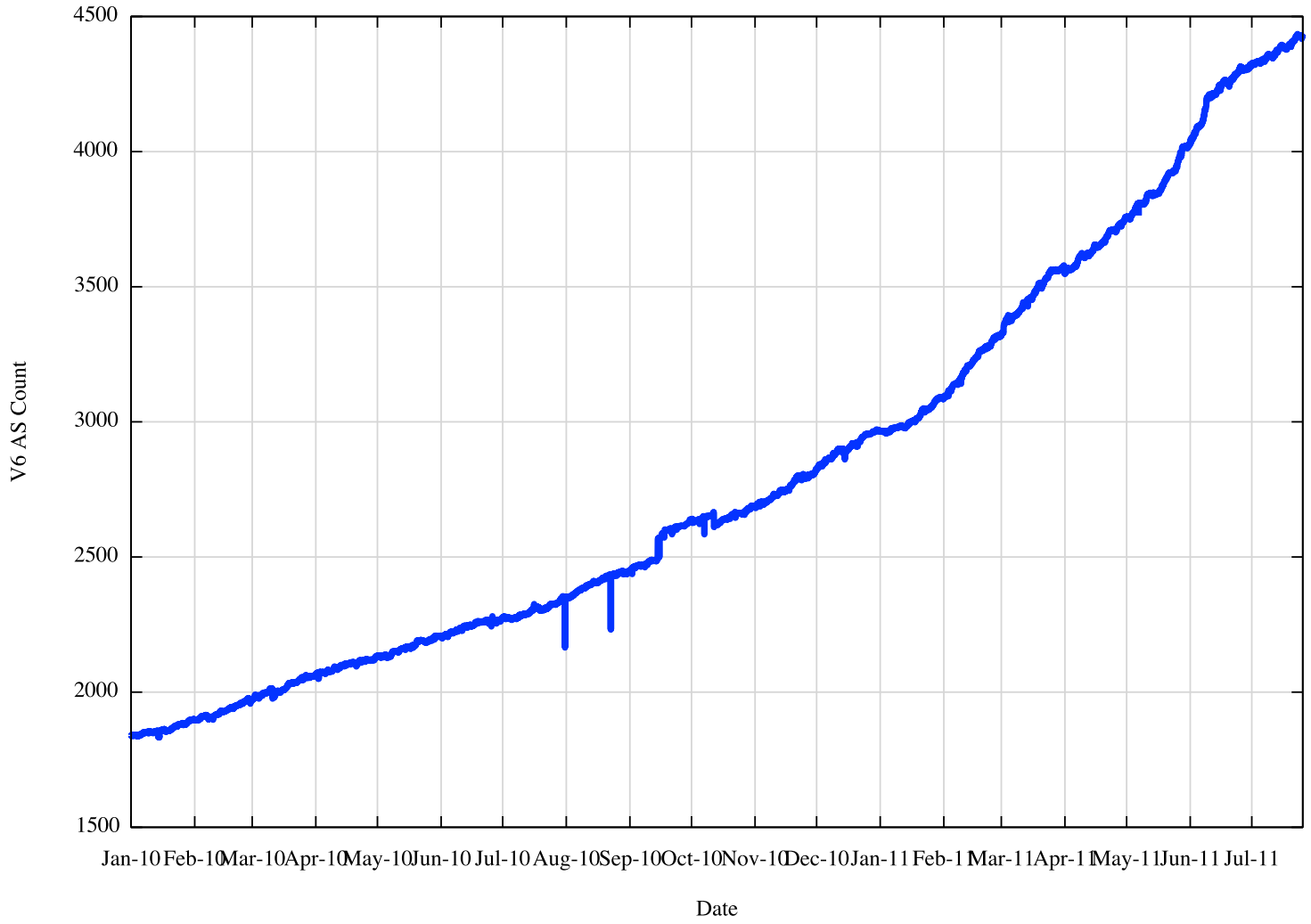
IPv6 Routed Address Span



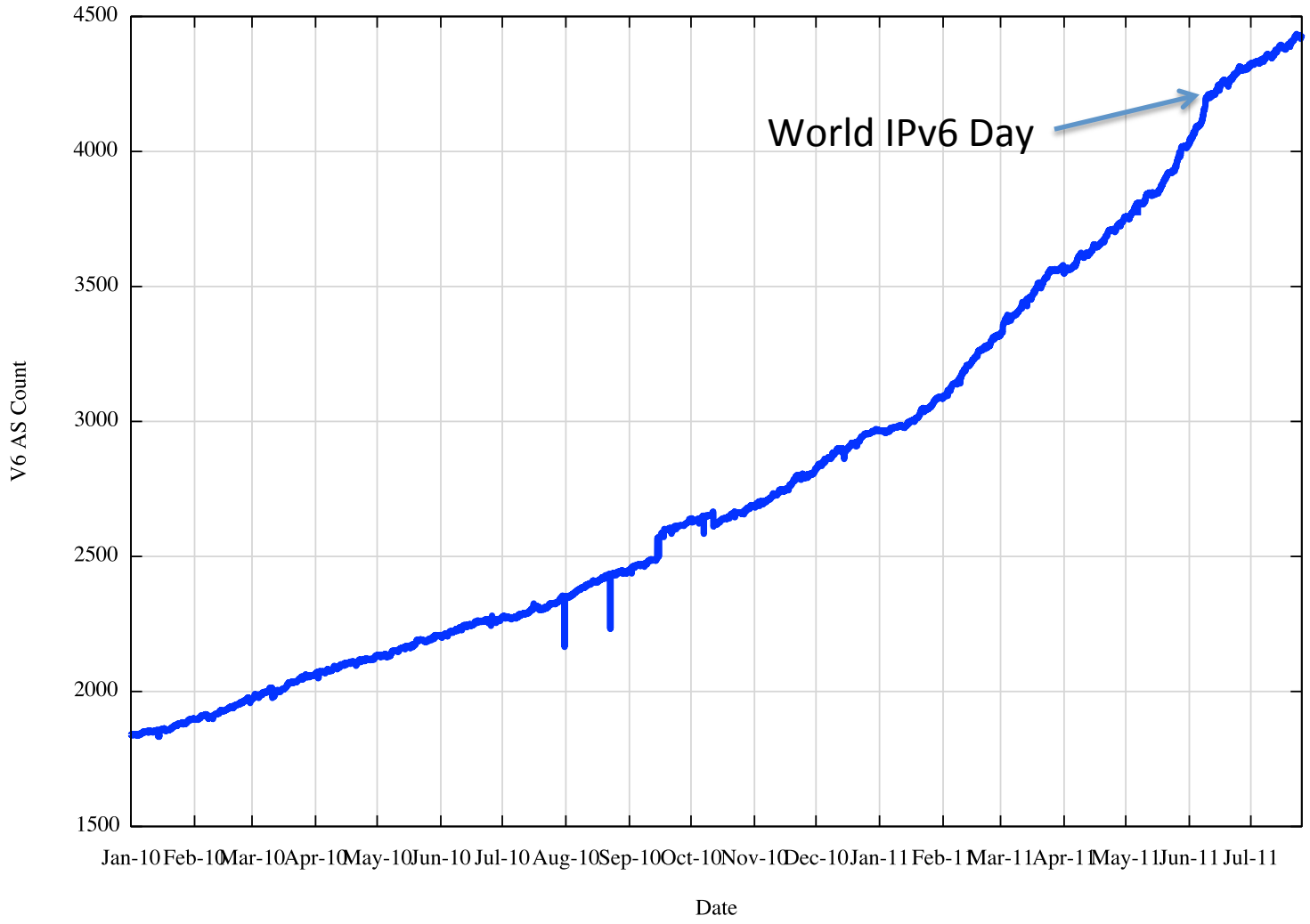
IPv6 Routed Address Span



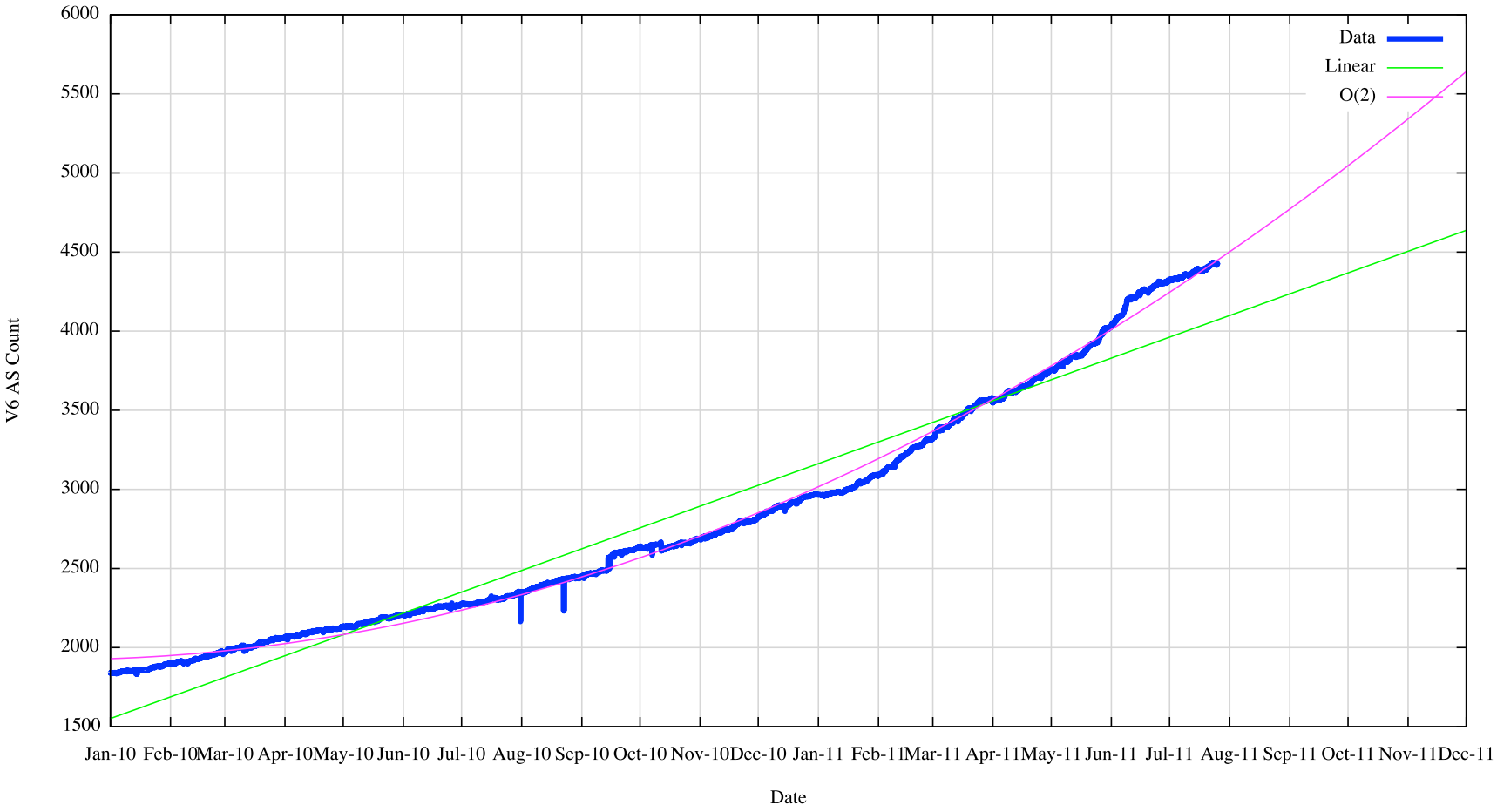
IPv6 Routed AS Count



IPv6 Routed AS Count



IPv6 Routed AS Count



IPv6 2010 BGP Vital Statistics

	Jan-10	Jan-11	
Prefix Count	2,458	4,100	+ 67%
Roots	1,965	3,178	+ 61%
More Specifics	494	922	+ 86%
Address Span	48,559/32s	53,415/32s	+ **10%**
AS Count	1,839	2,966	+ **61%**
Transit	348	556	+ 60%
Stub	1,437	2,343	+ 63%

IPv6 2011 BGP Vital Statistics

	Jan-10	Jan-11	Jul-11	p.a. rate
Prefix Count	2,458	4,100	6,889	+ 117%
Roots	1,965	3,178	5,090	+ 103%
More Specifics	494	922	1,799	+ 163%
Address Span (/32s)	48,559	53,415	56,561	+ **10%**
AS Count	1,839	2,966	4,424	+ **84%**
Transit	348	556	808	+ 78%
Stub	1,437	2,343	3,549	+ 88%

IPv6 in 2010 - 2011

- Overall IPv6 Internet growth in terms of BGP is 60% - 100 % p.a.
  - 2009 growth rate was ~ 50%.
- Looking at the Transit AS count, if these relative growth rates persist then the IPv6 transit AS count will equal the IPv4 transit AS count in 5 years time (mid/late 2016)

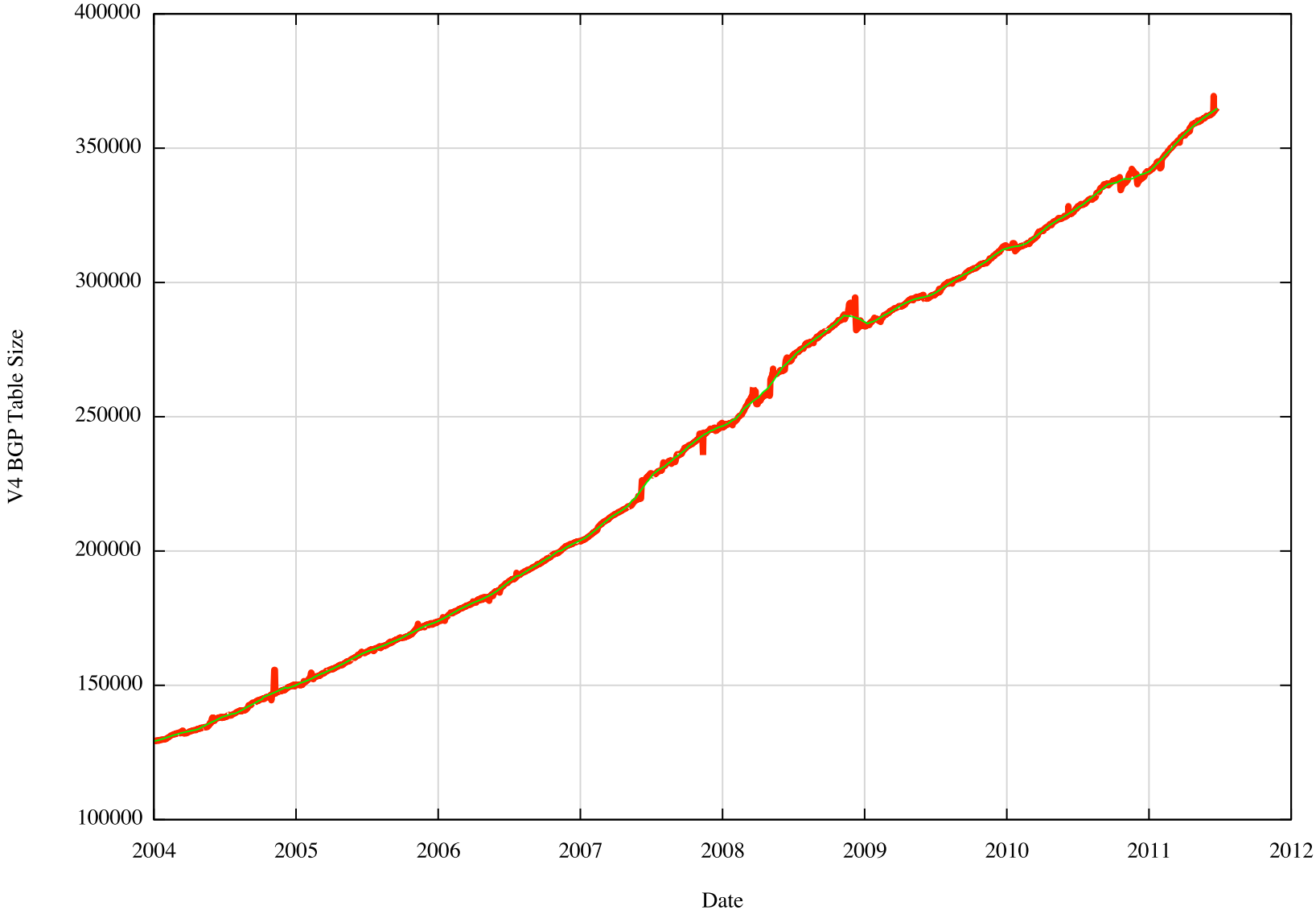
Where is this heading?

BGP Size Projections

- Generate a 4 year projection of the IPv4 routing table using a quadratic (O(2) polynomial) over the 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

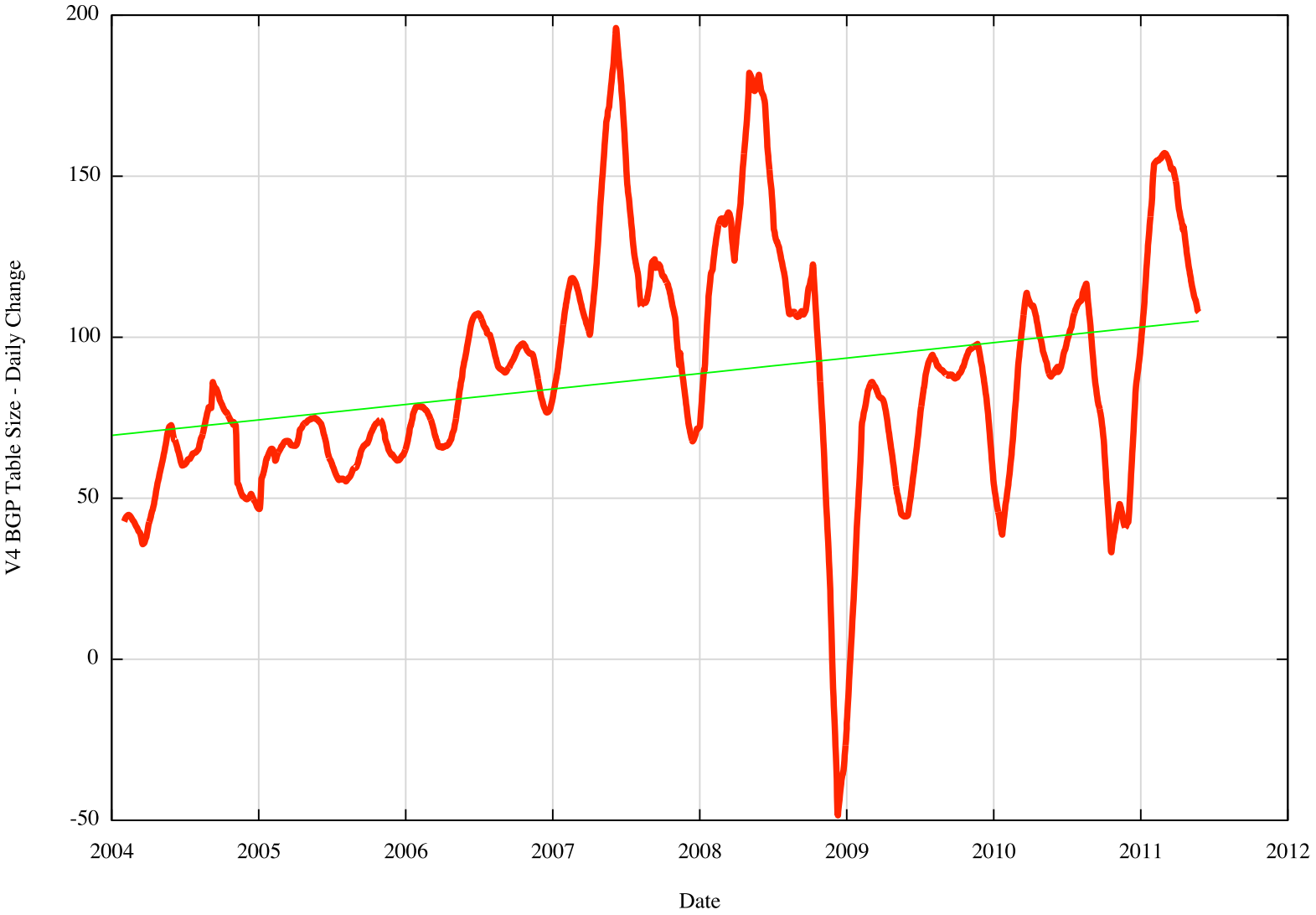
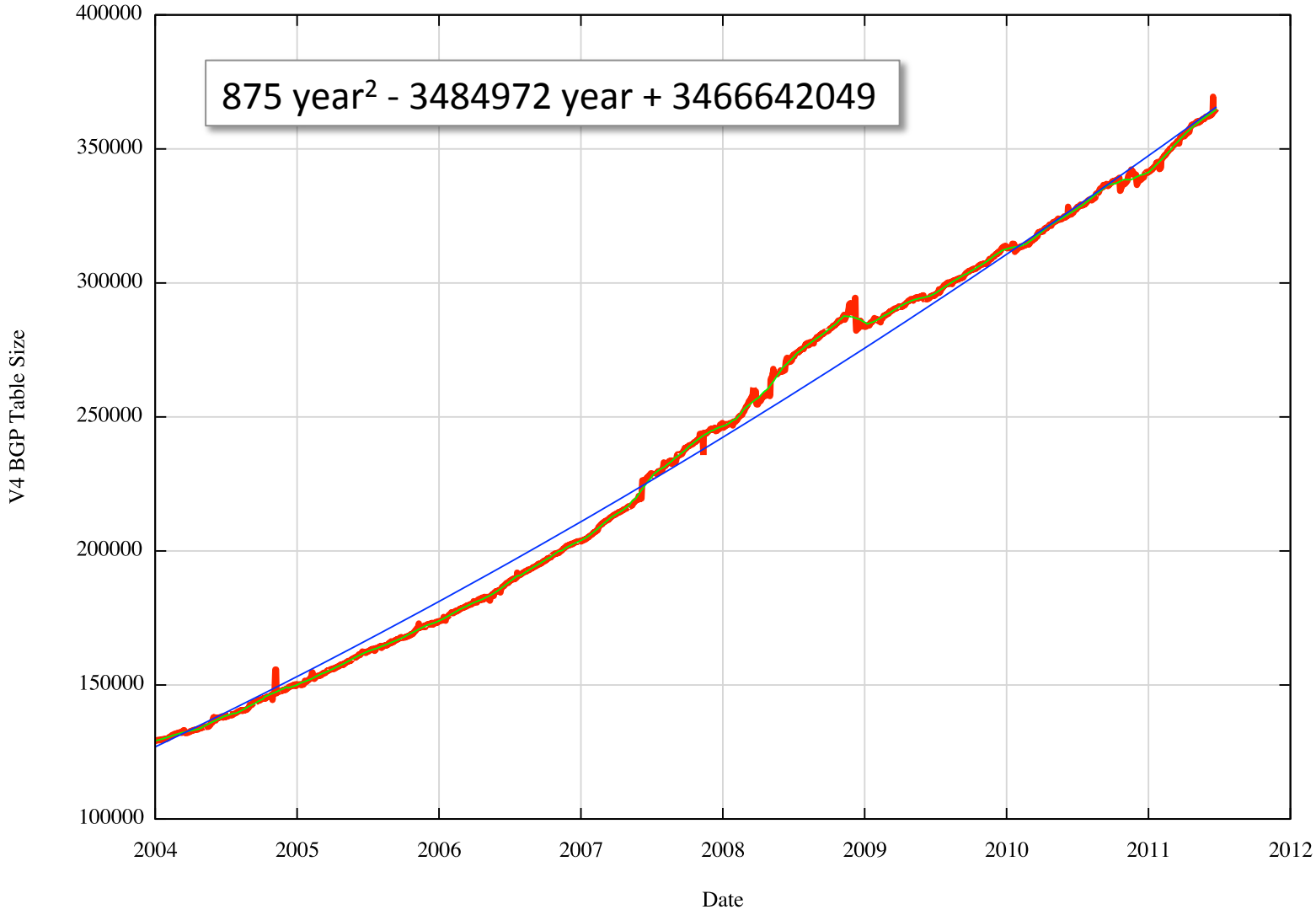
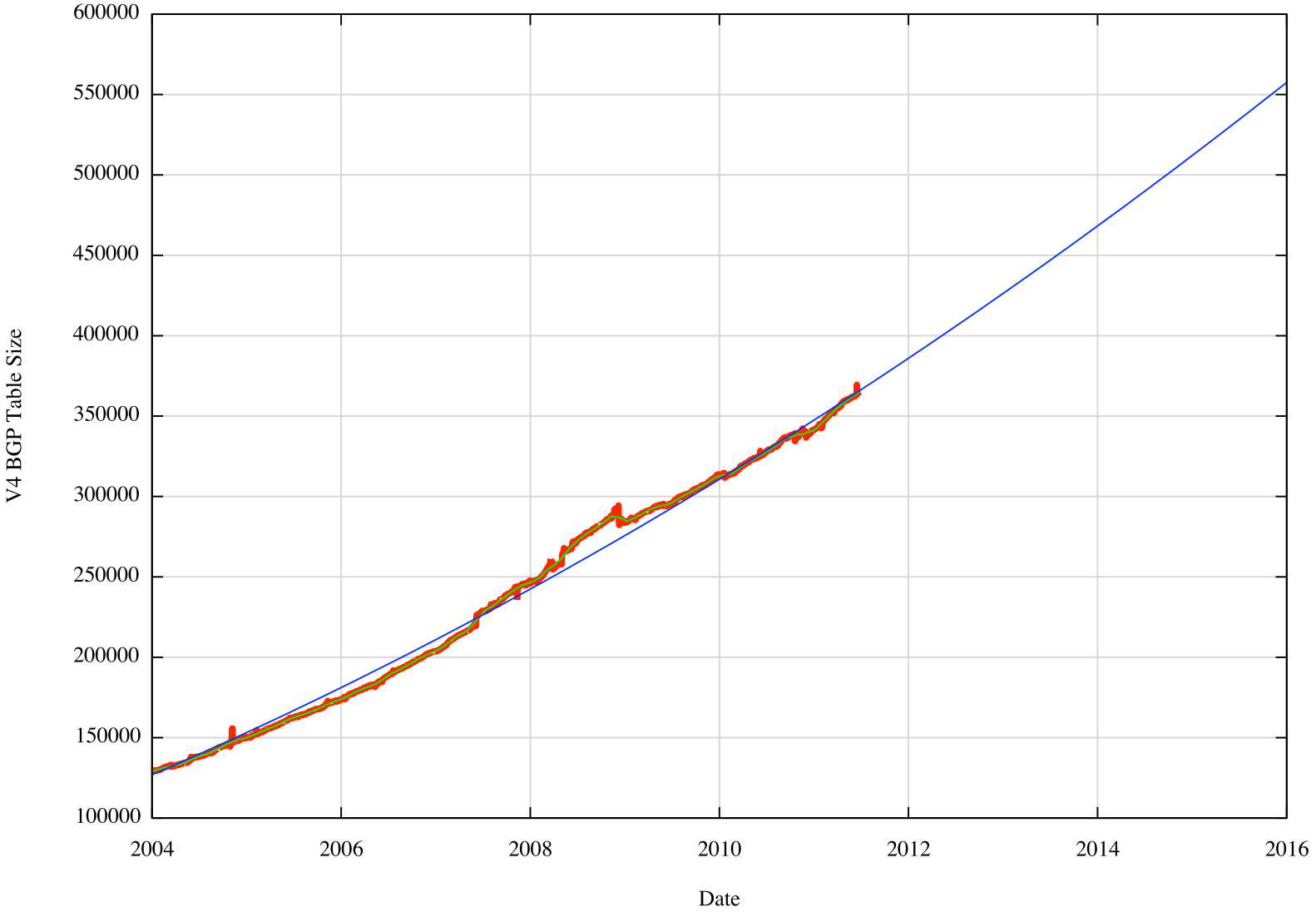


Table Growth Model



IPv4 Table Projection

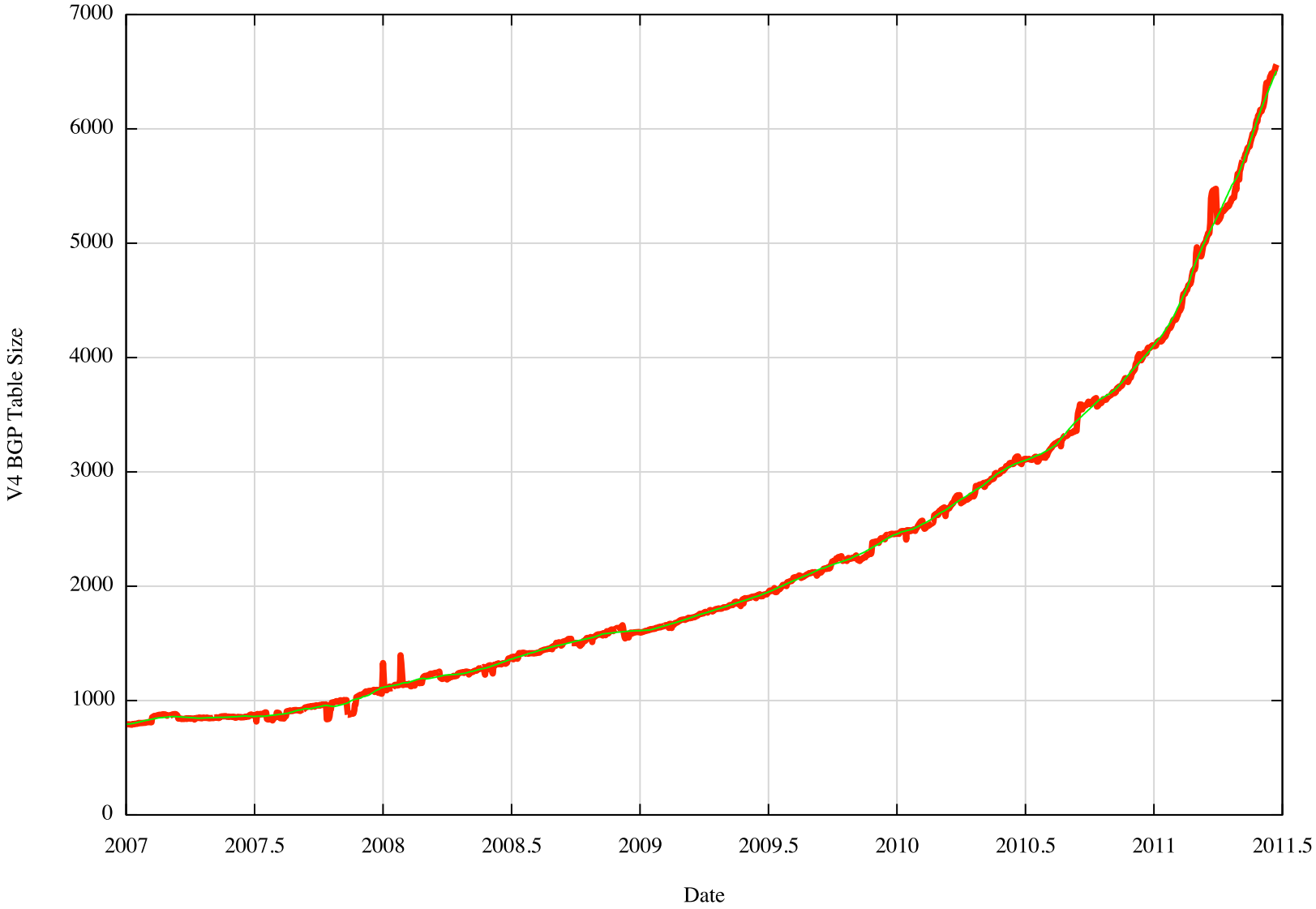


IPv4 BGP Table Size predictions

Jan 2011	347,000 entries
2012*	385,000 entries
2013*	426,000 entries
2014*	468,000 entries
2015*	512,000 entries
2016*	557,000 entries

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

IPv6 Table Size



Daily Growth Rates

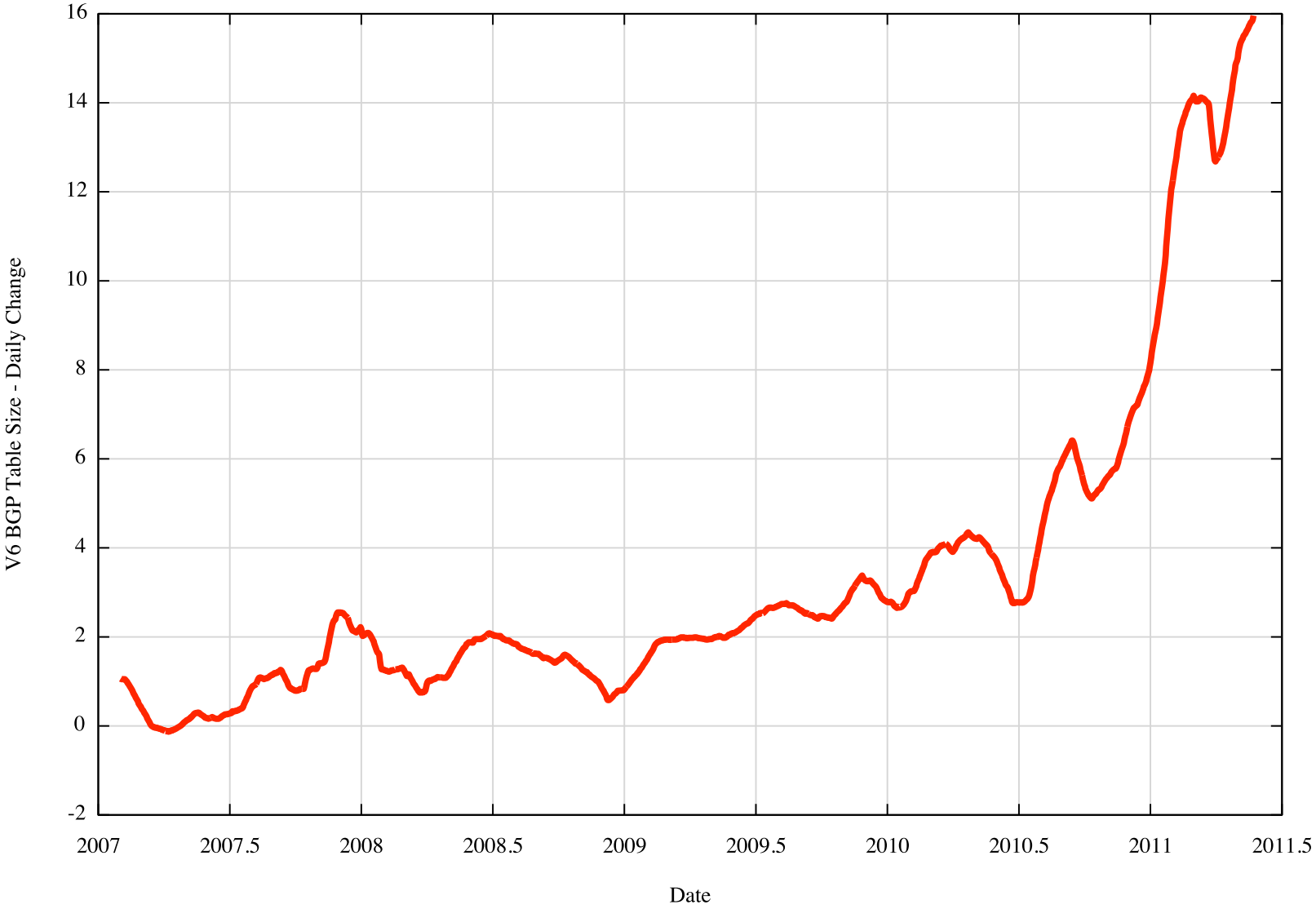
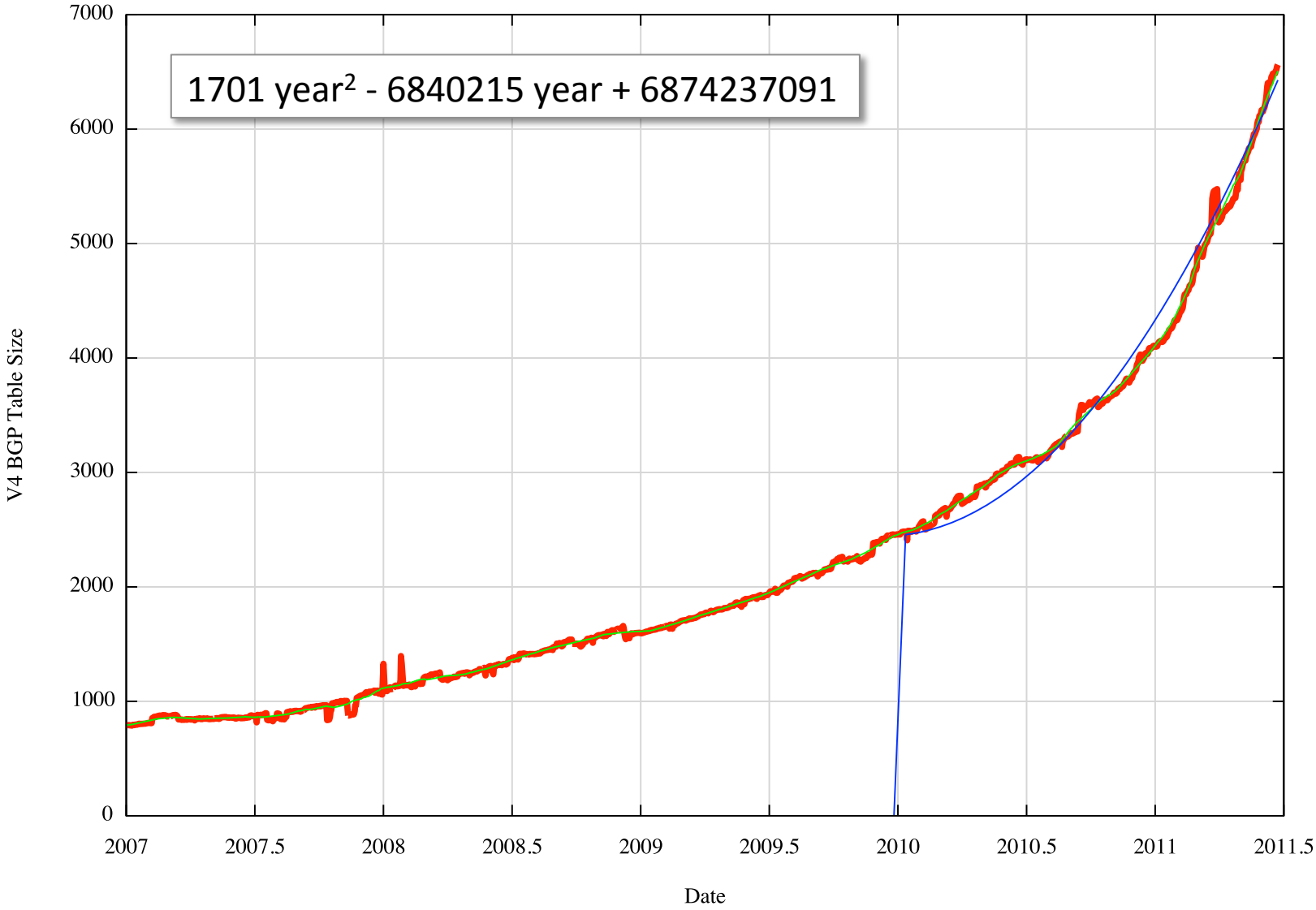
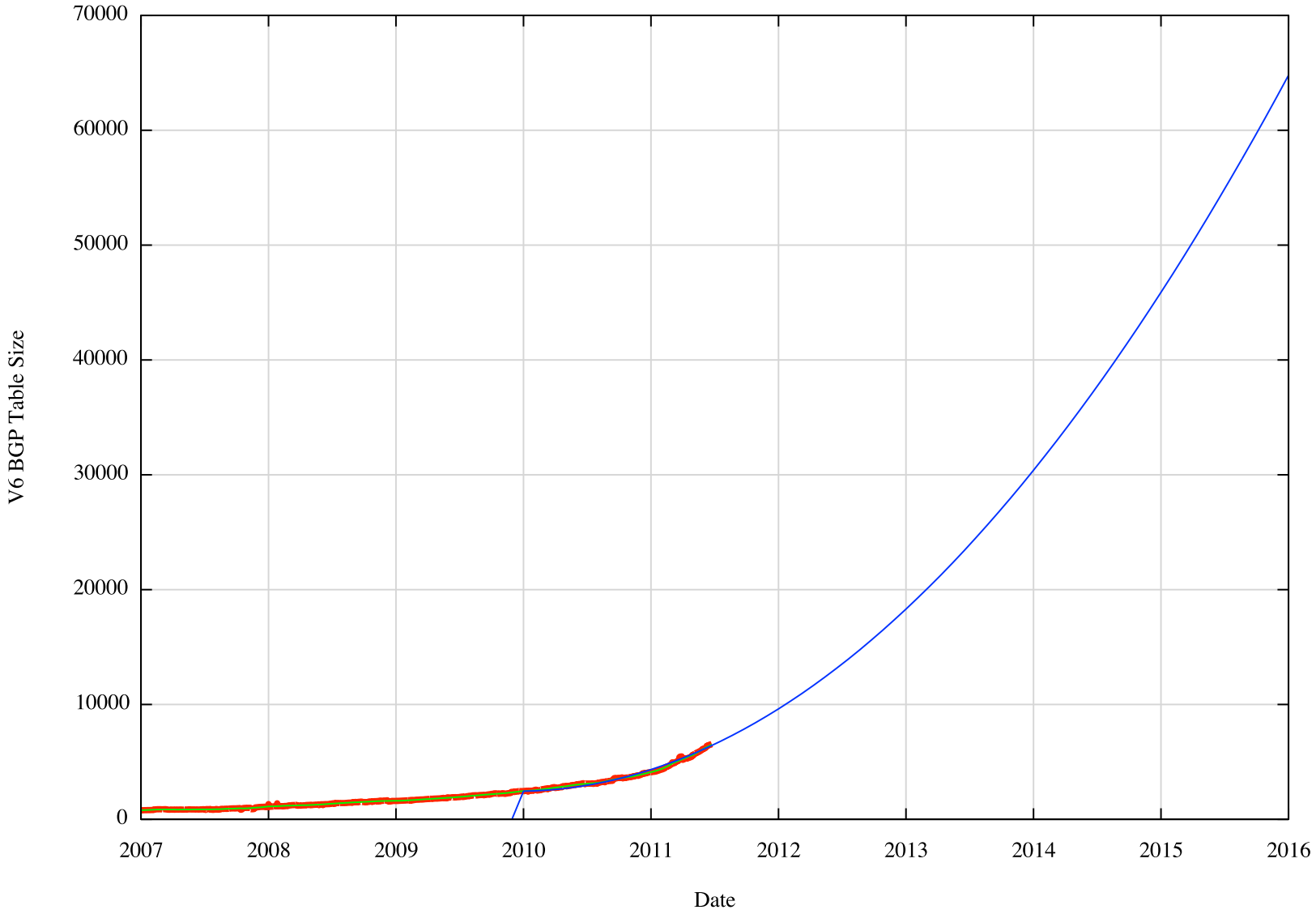


Table Growth Model



IPv6 Table Projection



IPv6 BGP Table Size predictions

Jan 2011	4,000 entries
2012	10,000 entries
2013	18,000 entries
2014	30,000 entries
2015	46,000 entries
2016	65,000 entries

BGP Table Size Predictions

Jan 2011	$347{,}000_4$ + $4{,}000_6$ entries
2012	$385{,}000_4$ + $10{,}000_6$ entries
2013*	$426{,}000_4$ + $18{,}000_6$ entries
2014*	$468{,}000_4$ + $30{,}000_6$ entries
2015*	$512{,}000_4$ + $46{,}000_6$ entries
2016*	$557{,}000_4$ + $65{,}000_6$ entries

** These numbers are dubious due to IPv4 address exhaustion pressures. It is possible that the number will be larger than the values predicted by this model.*

Divination Uncertainty

- As we get further into the IPv6 transition we may see:
  - accelerated IPv4 routing fragmentation as an outcome from the operation of a V4 address trading market that starts to slice up the V4 space into smaller routed units
  - parallel V6 deployment that picks up pace
- These projections of FIB size are going to be *low*.
- Just how low it will be is far harder to estimate.

Is This a Problem?

Is This a Problem?

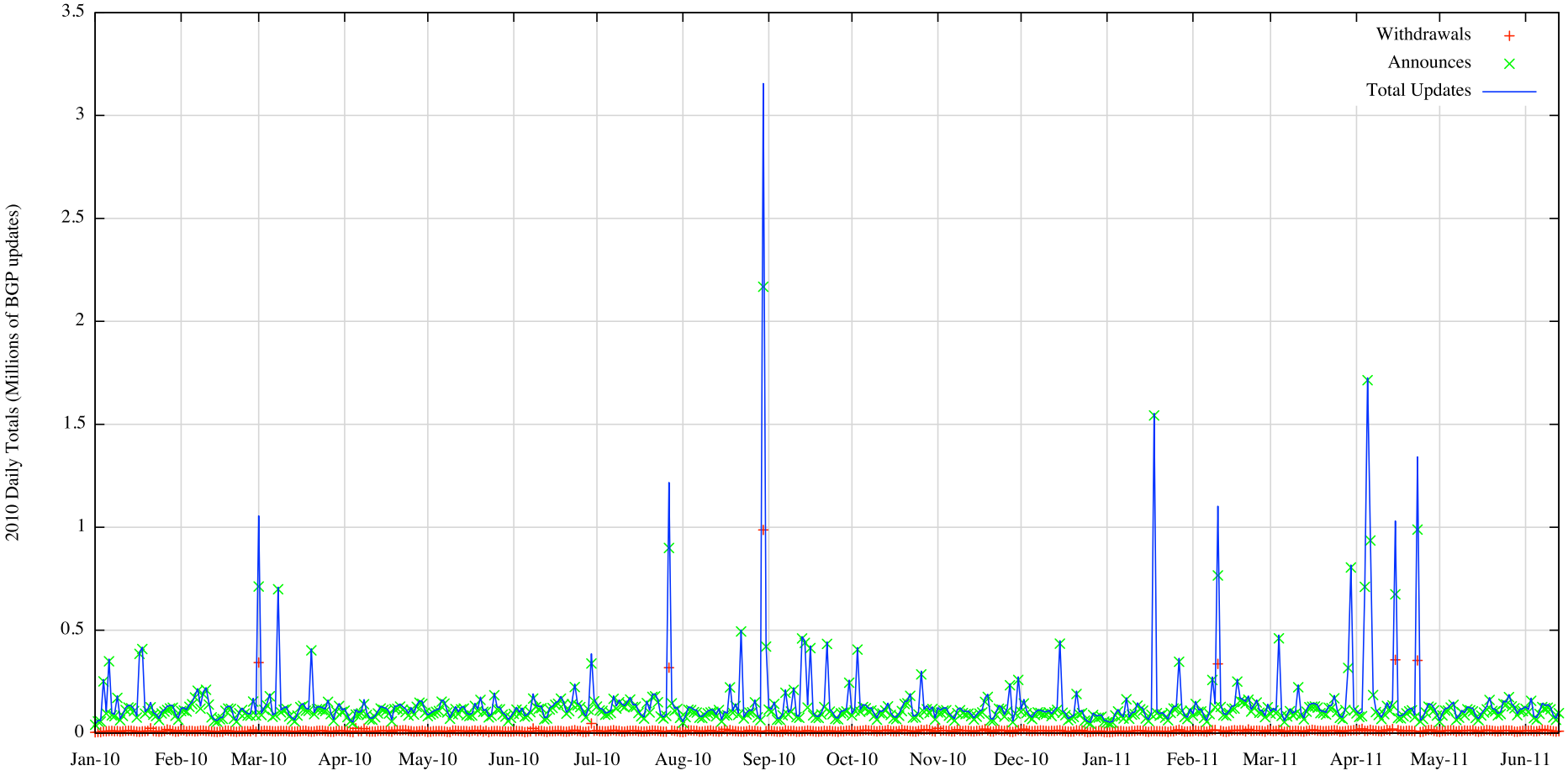
- What is the anticipated end of service life of your core routers?
- What's the price/performance curve for forwarding engine ASICS?
- What's a sustainable growth factor in FIB size that will allow for continued improvement in unit costs of routing?
- What is a reasonable margin of uncertainty in these projections?

Does Size REALLY matter?

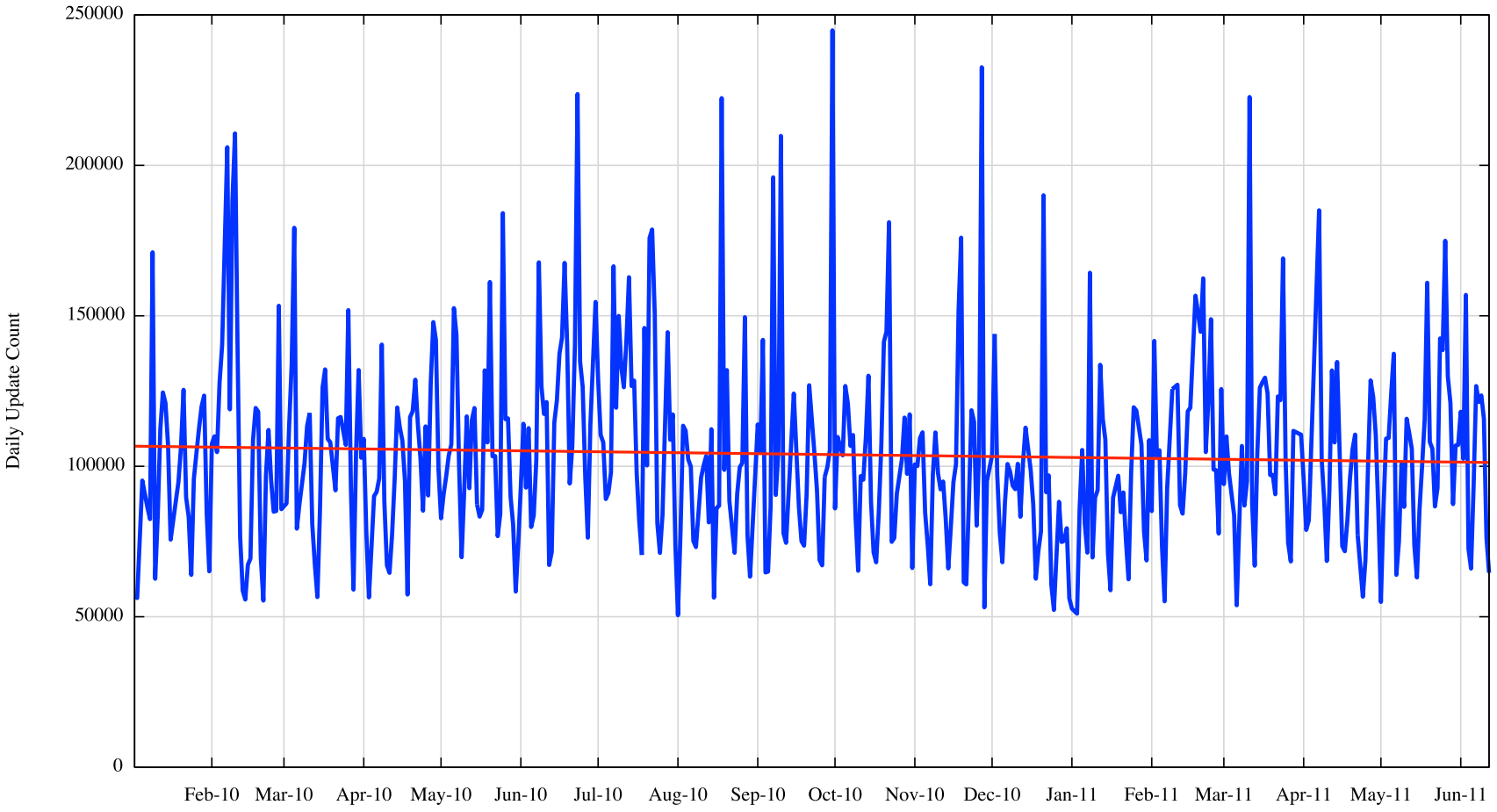
Is it the **size of the RIB** or the **level of dynamic update and routing stability** that is the concern here?

So lets look at update trends in BGP…

Daily Announce and Withdrawal Rates

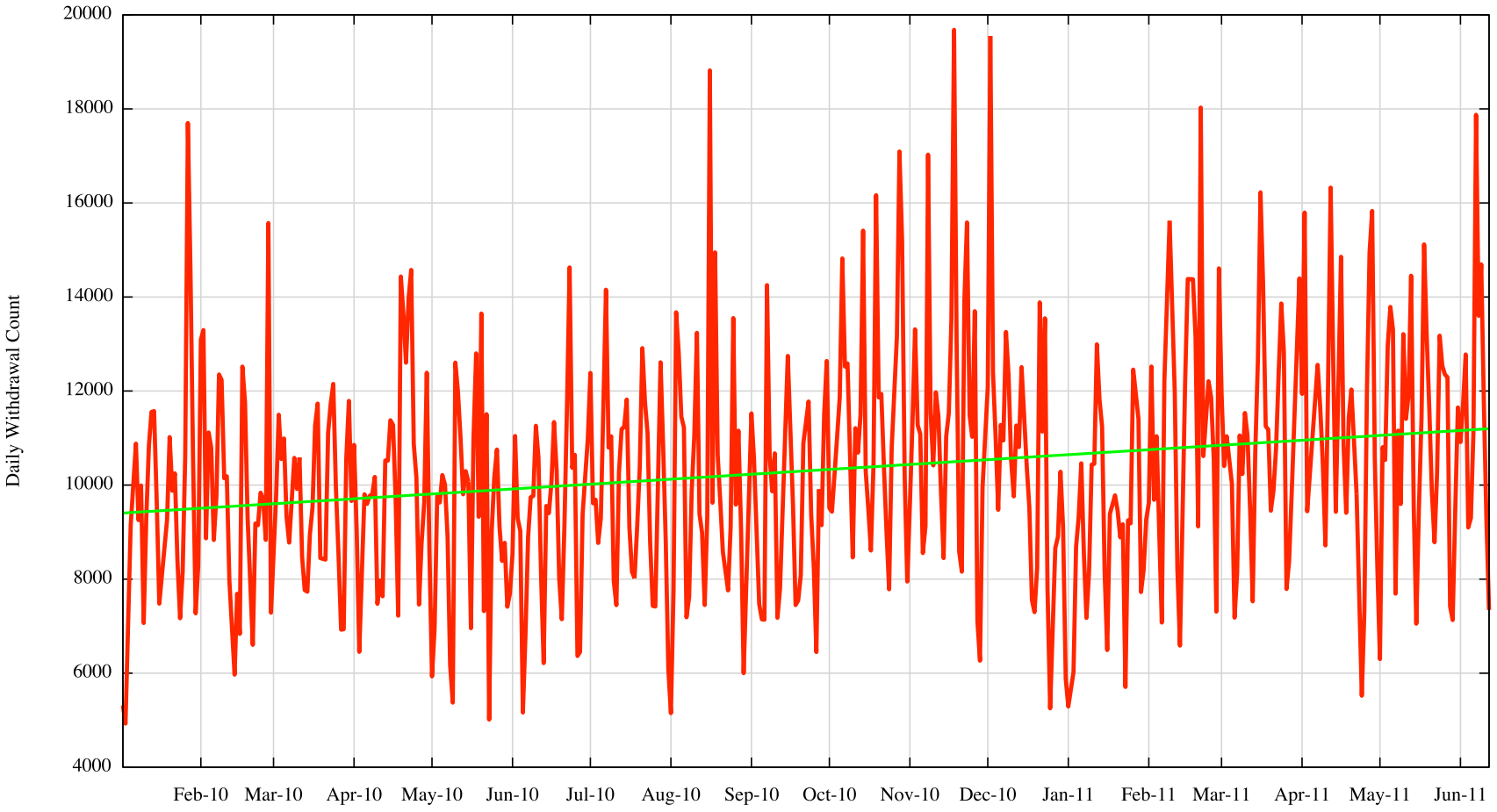


Daily Updates (clipped)

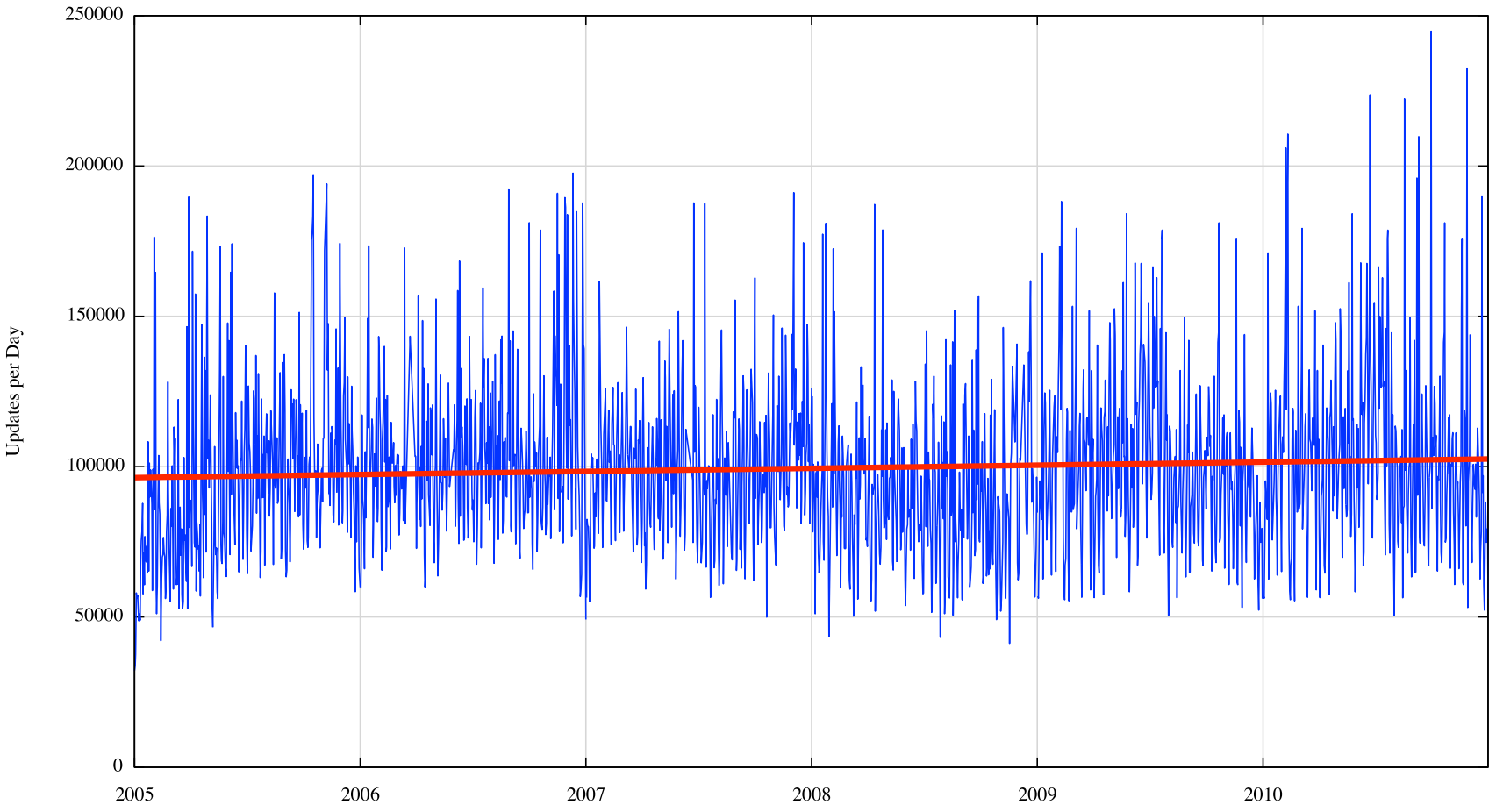


The threshold

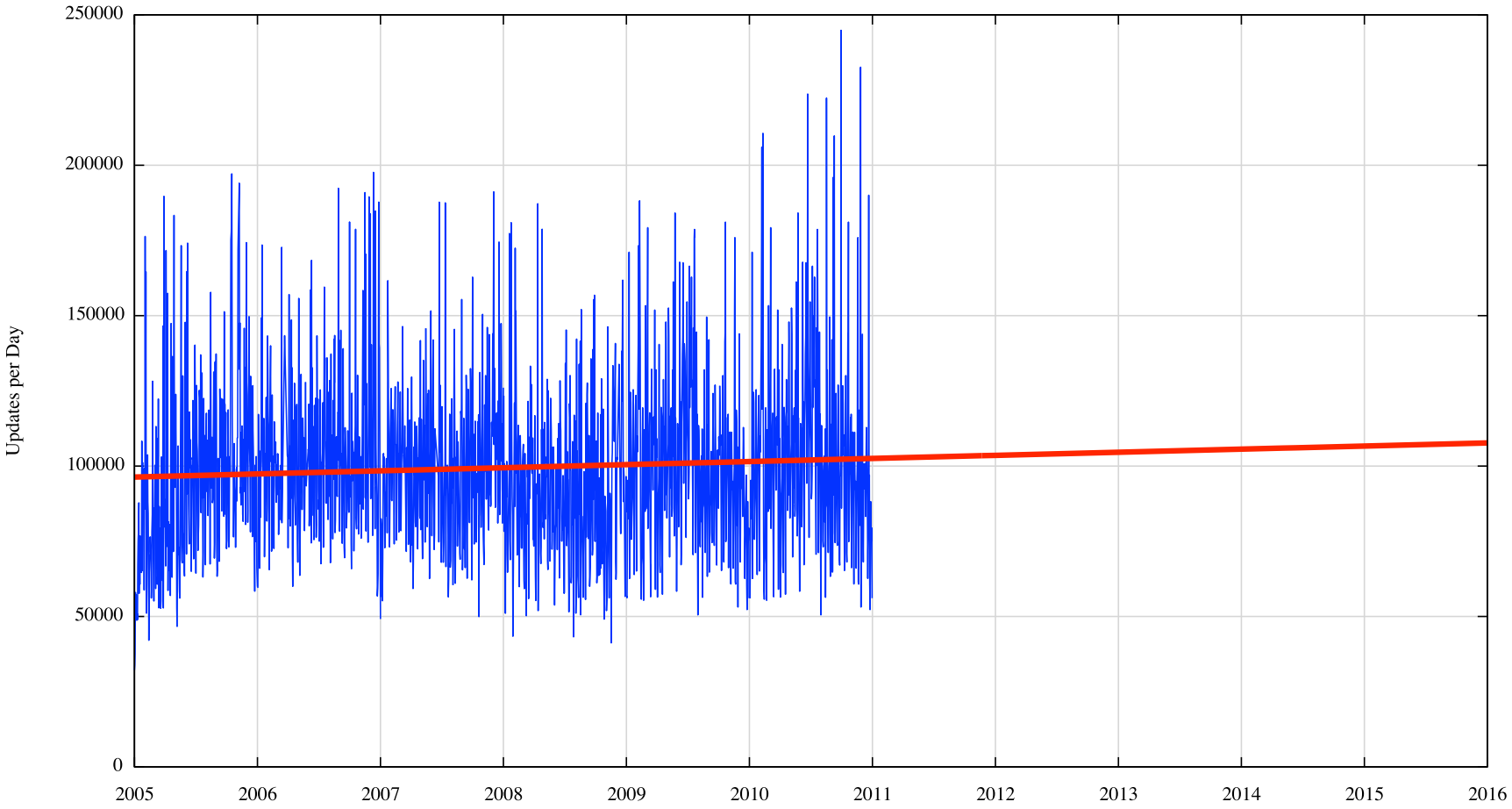
Daily Withdrawals



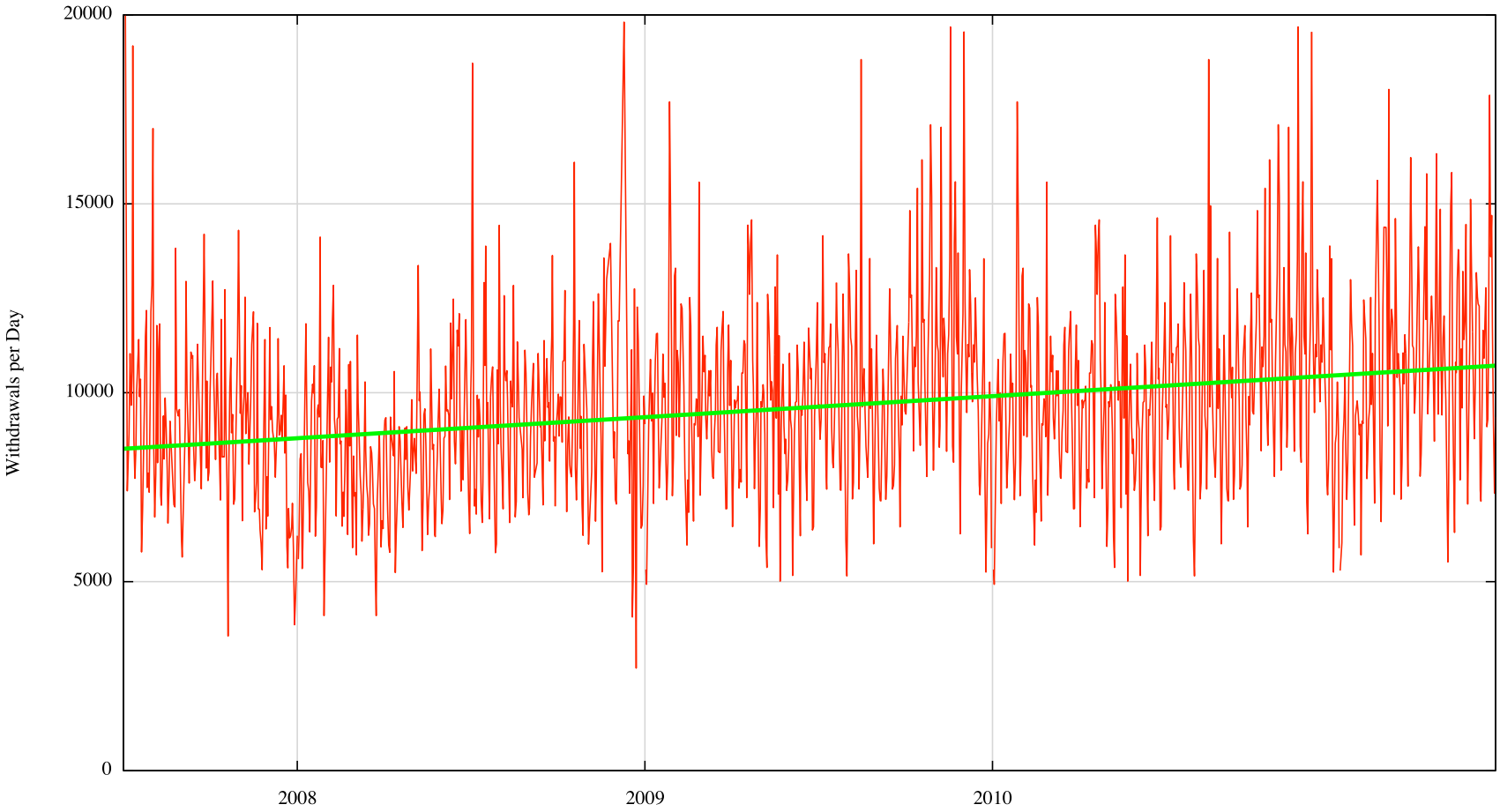
Updates – Extended Data Set



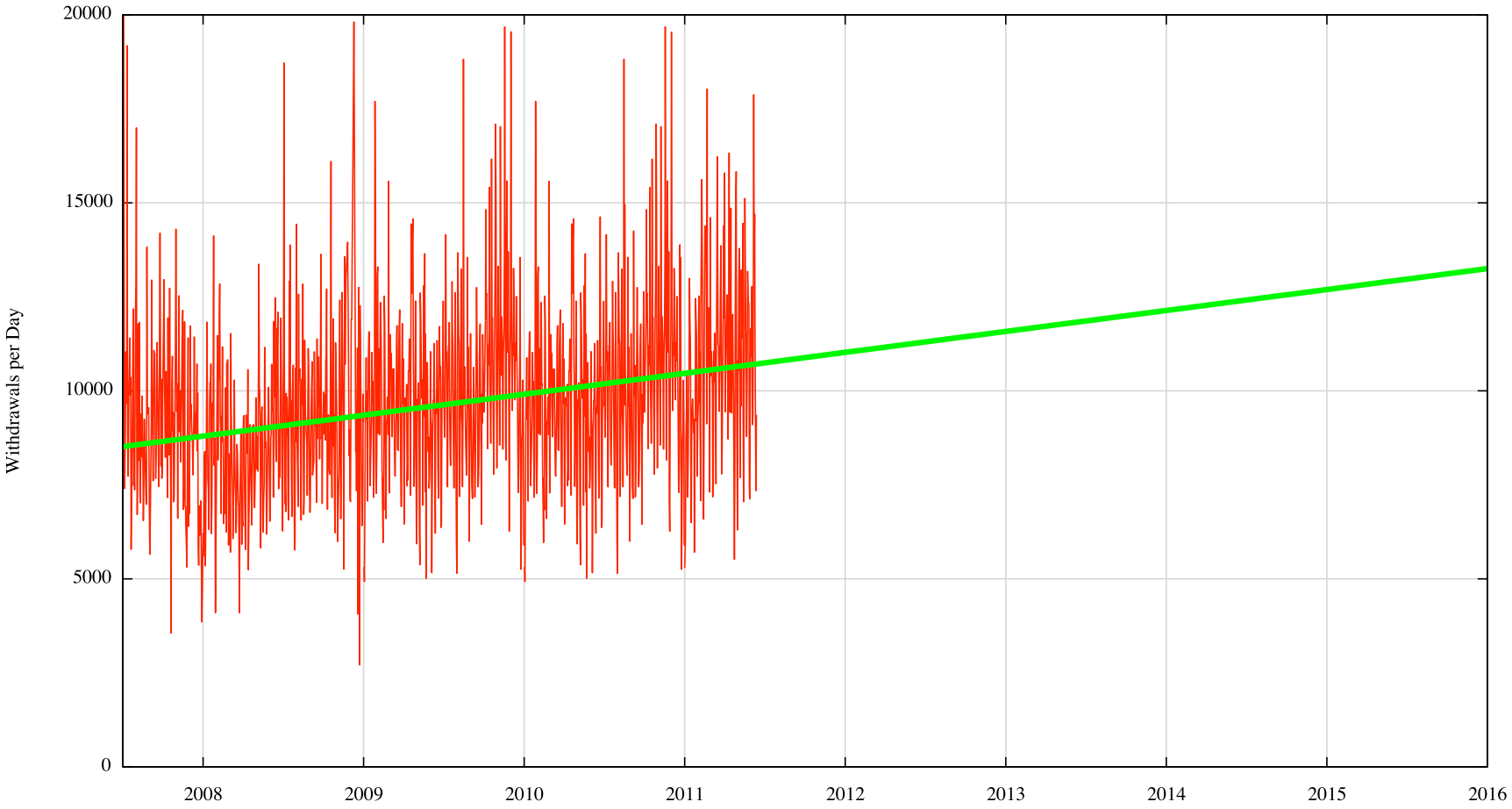
Daily Update Rate – Linear Projection



Withdrawals – Extended Data Set



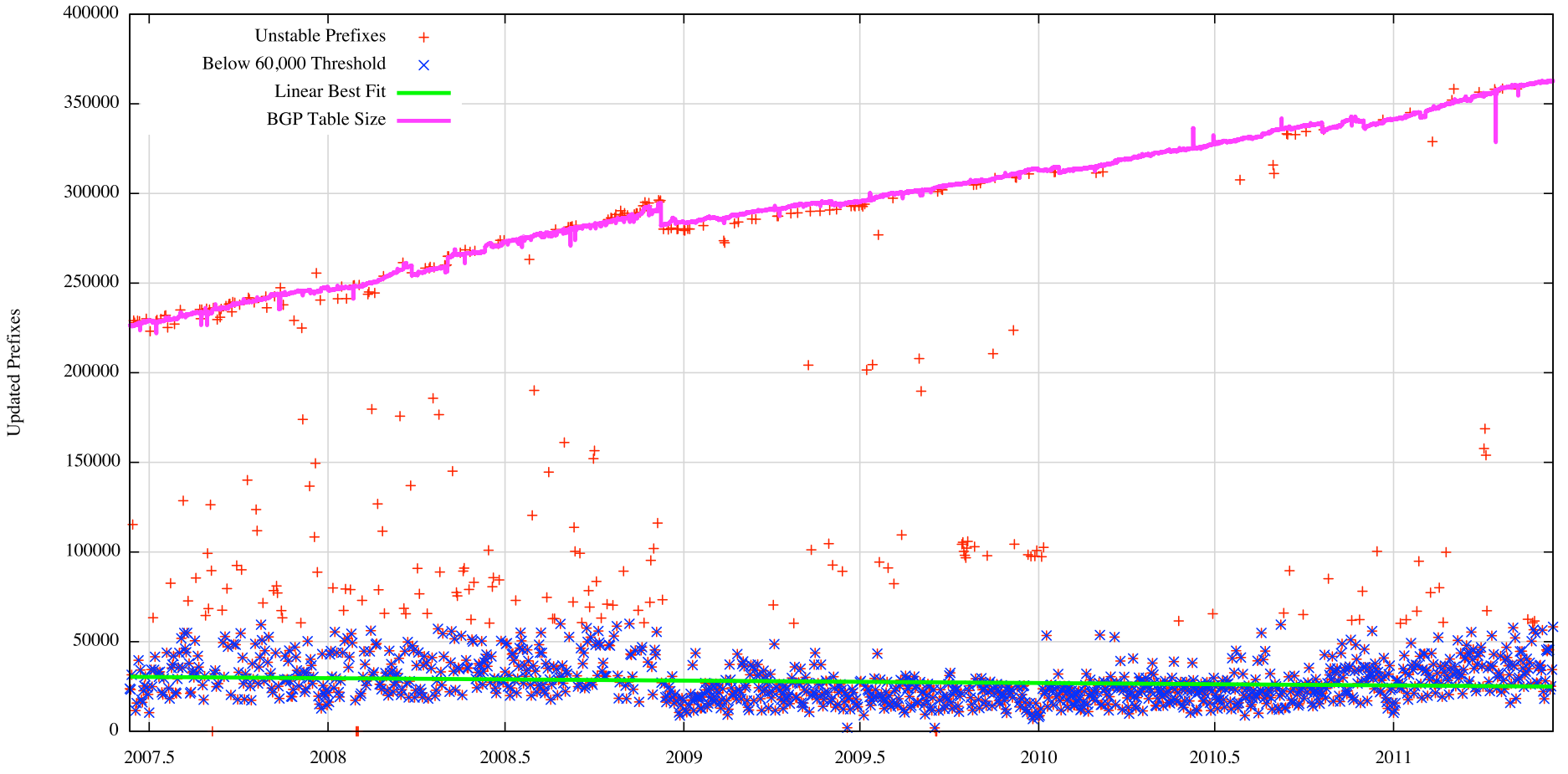
Daily Withdrawal Rate – Linear Projection



Why is the world so flat?

- Given that the BGP routing table has been growing at a compound growth rate across this entire period, why is the rate of growth of updates and withdrawals so flat?
- An intuitive model of BGP updated would see instability as being related to the number of entries and the density of interconnectivity
  - This is obviously not the model we see here
  - So why is this particular part of the Internet's BGP behaviour so anomalous?

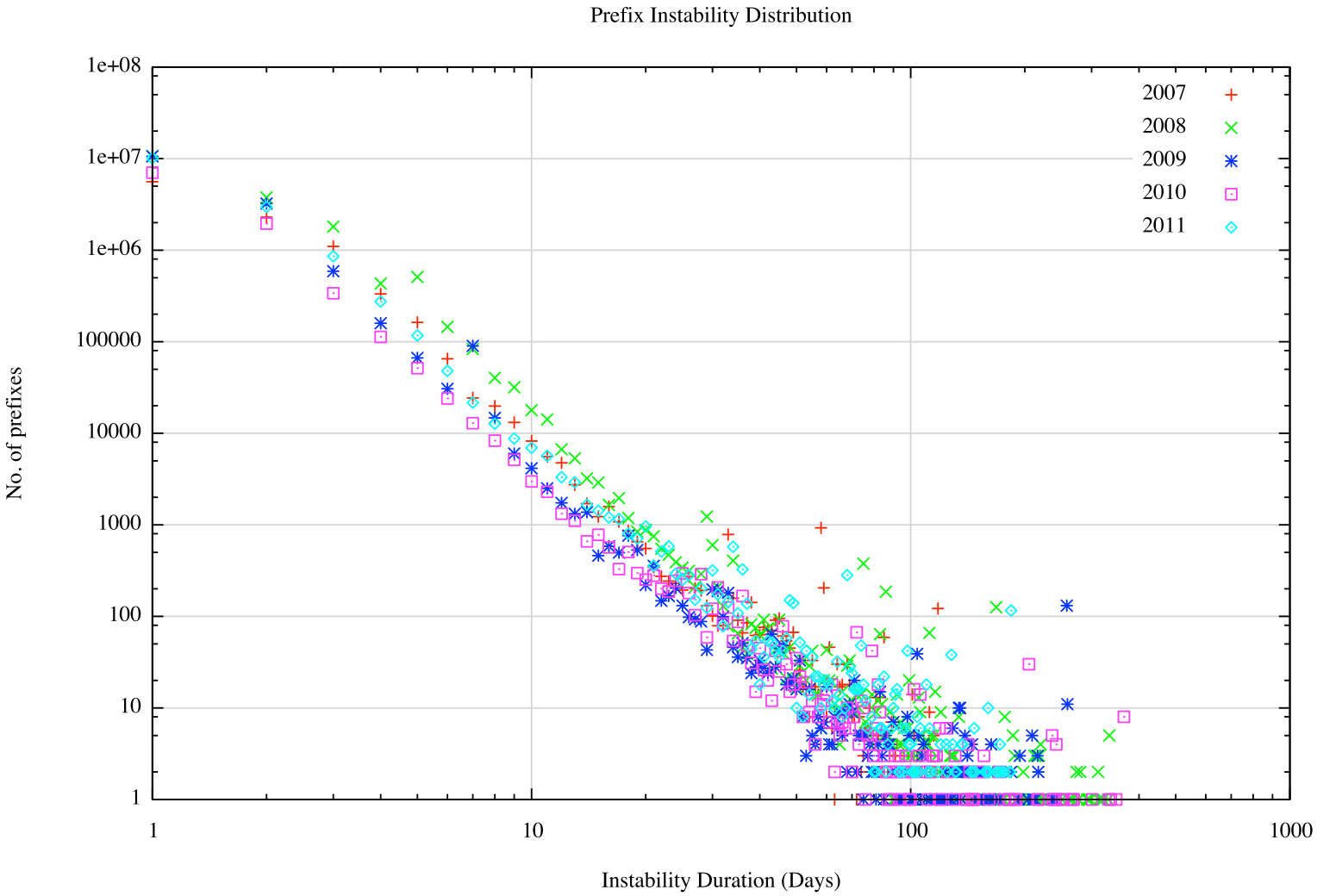
Unstable Prefixes



Unstable Prefixes

- Over the past 4 years the number of unstable prefixes lies between 20,000 – 50,000 prefixes per day
- How “stable” is this set of unstable prefixes?
  - Are they the same prefixes?
  - Are they equally noisy?
  - What are the characteristics of this “noise”?

Prefix Instability Duration



Prefix Instability

- Prefix Instability is generally short lived
  - 90% of all prefixes are unstable for 2 days or less
  - 6 prefixes are persistently unstable – these are beacon prefixes.
- The distribution of the duration of prefix instability at a coarse level (per day) appears to be a power law distribution (see Zipfs'Law)

Convergence

- BGP is a distance vector protocol
- This implies that BGP may send a number of updates in a tight “cluster” before converging to the “best” path
- This is clearly evident in withdrawals and convergence to (longer) secondary paths

For Example

Withdrawal at source at 08:00:00 03-Apr of 84.205.77.0/24 at MSK-IX, as observed at AS 2.0

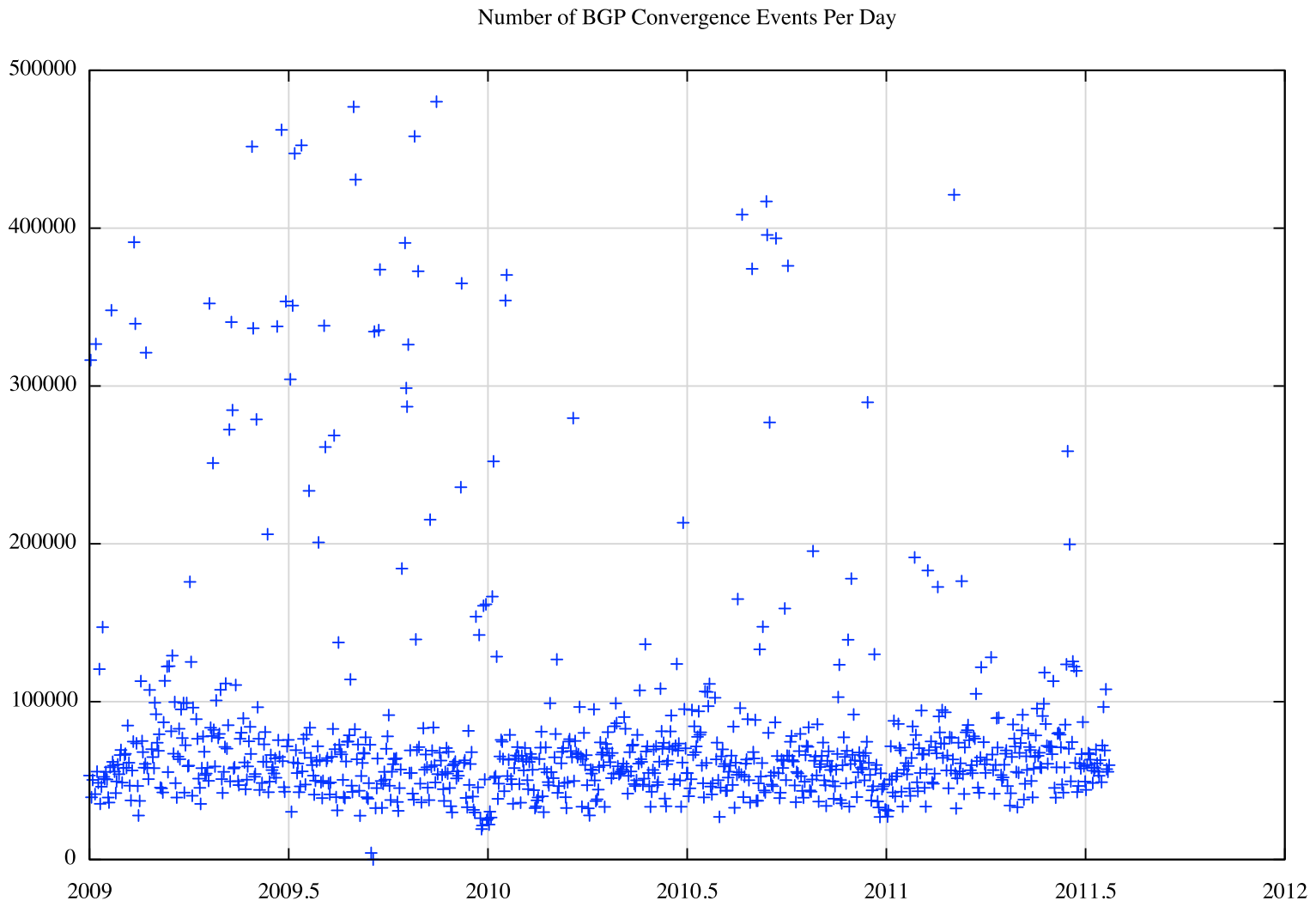
Announced AS Path: <4777 2497 9002 12654>

Received update sequence:

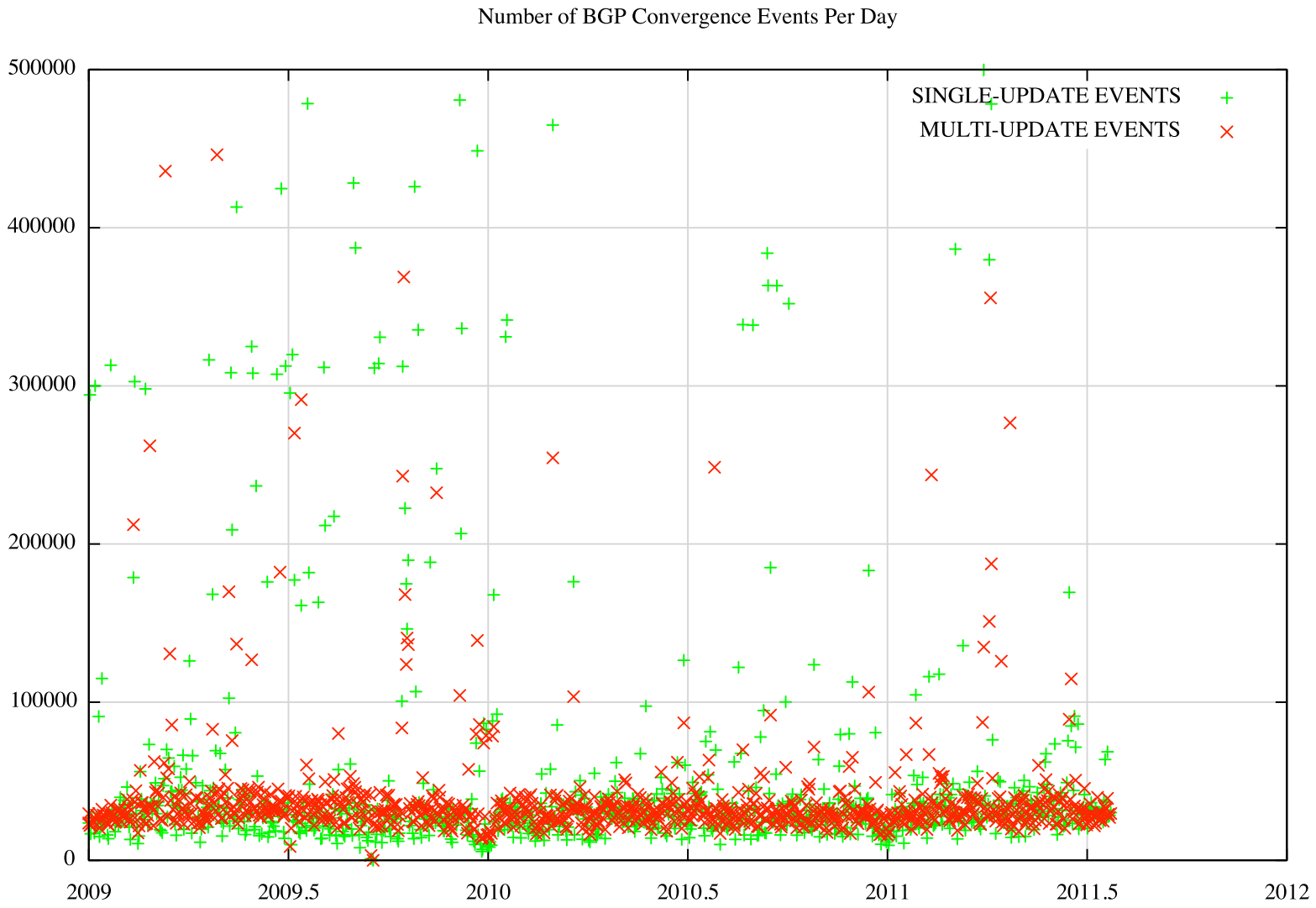
08:02:22 03-Apr + <4777 2516 3549 3327 12976 20483 31323 12654>
08:02:51 03-Apr + <4777 2497 3549 3327 12976 20483 39792 8359 12654>
08:03:52 03-Apr + <4777 2516 3549 3327 12976 20483 39792 6939 16150 8359 12654>
08:04:28 03-Apr + <4777 2516 1239 3549 3327 12976 20483 39792 6939 16150 8359 12654>
08:04:52 03-Apr - <4777 2516 1239 3549 3327 12976 20483 39792 6939 16150 8359 12654>

1 withdrawal at source generated a convergence sequence of 5 events, spanning 150 seconds

BGP “Convergence Events” Per Day



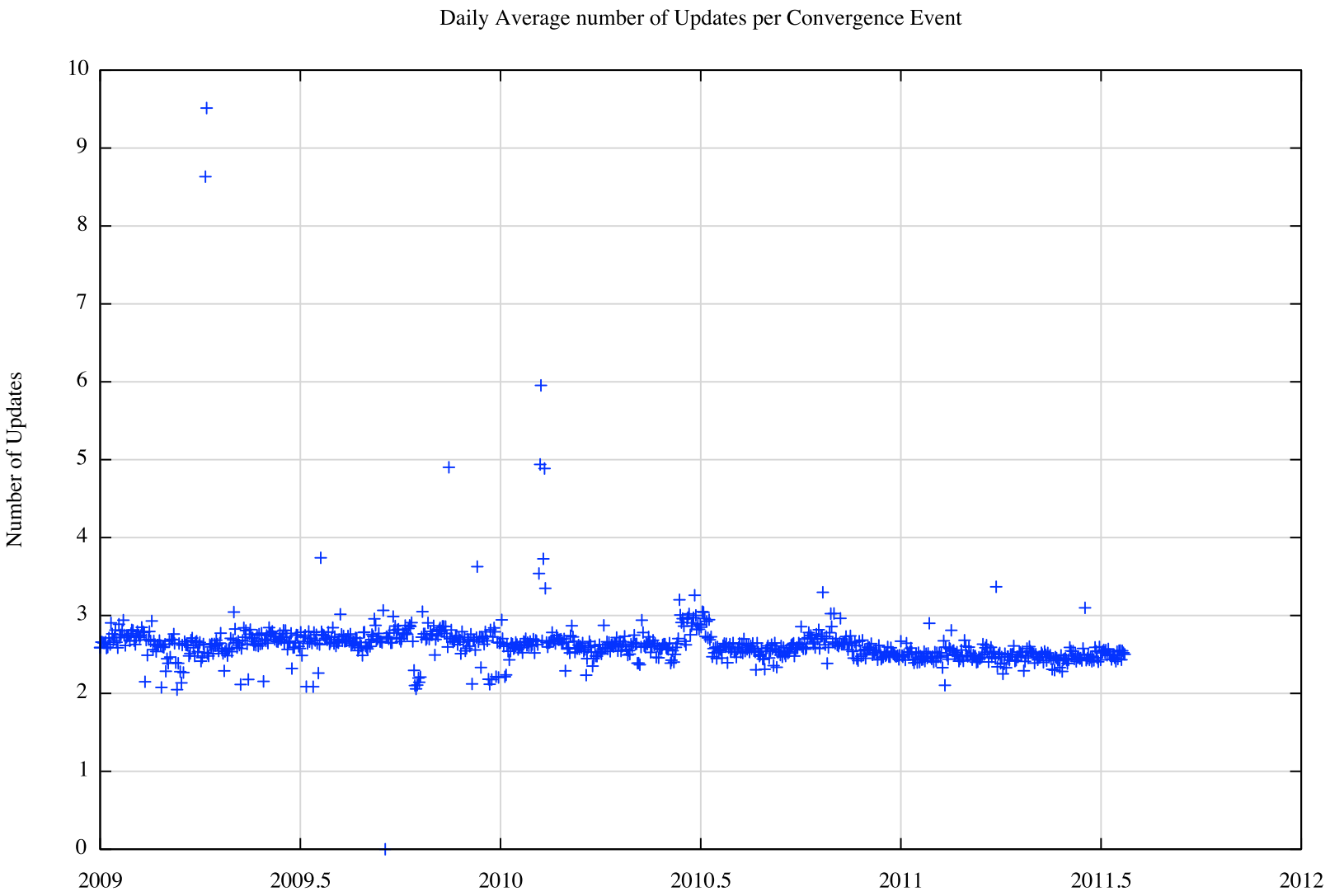
BGP Single/Multi-Update Convergence Events Per Day



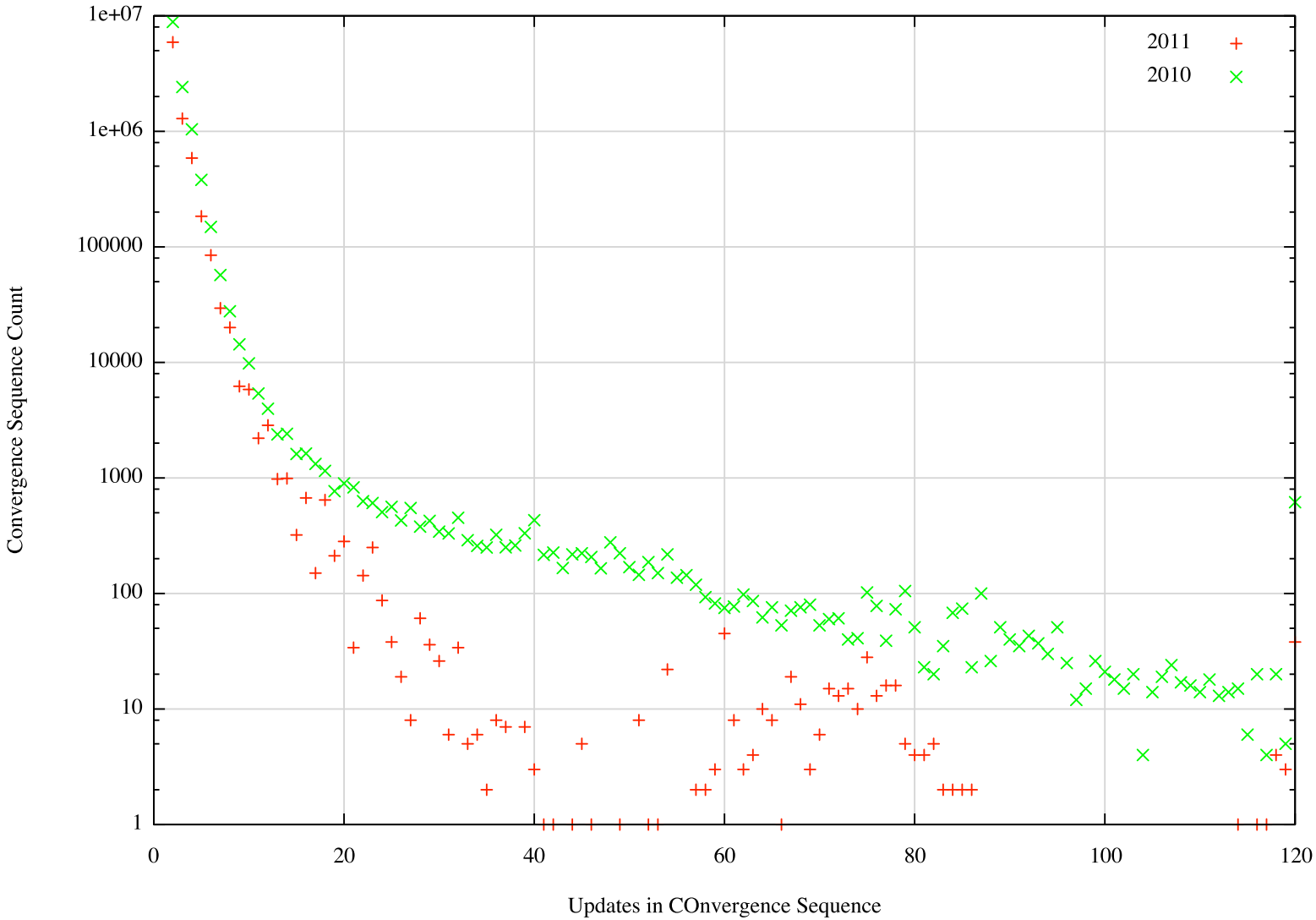
Flat Worlds

- The number of convergence events and the number of multi-update convergence events appears to be constant for the past couple of years
- Routing instability is following a different scaling path to the size of the BGP routing table

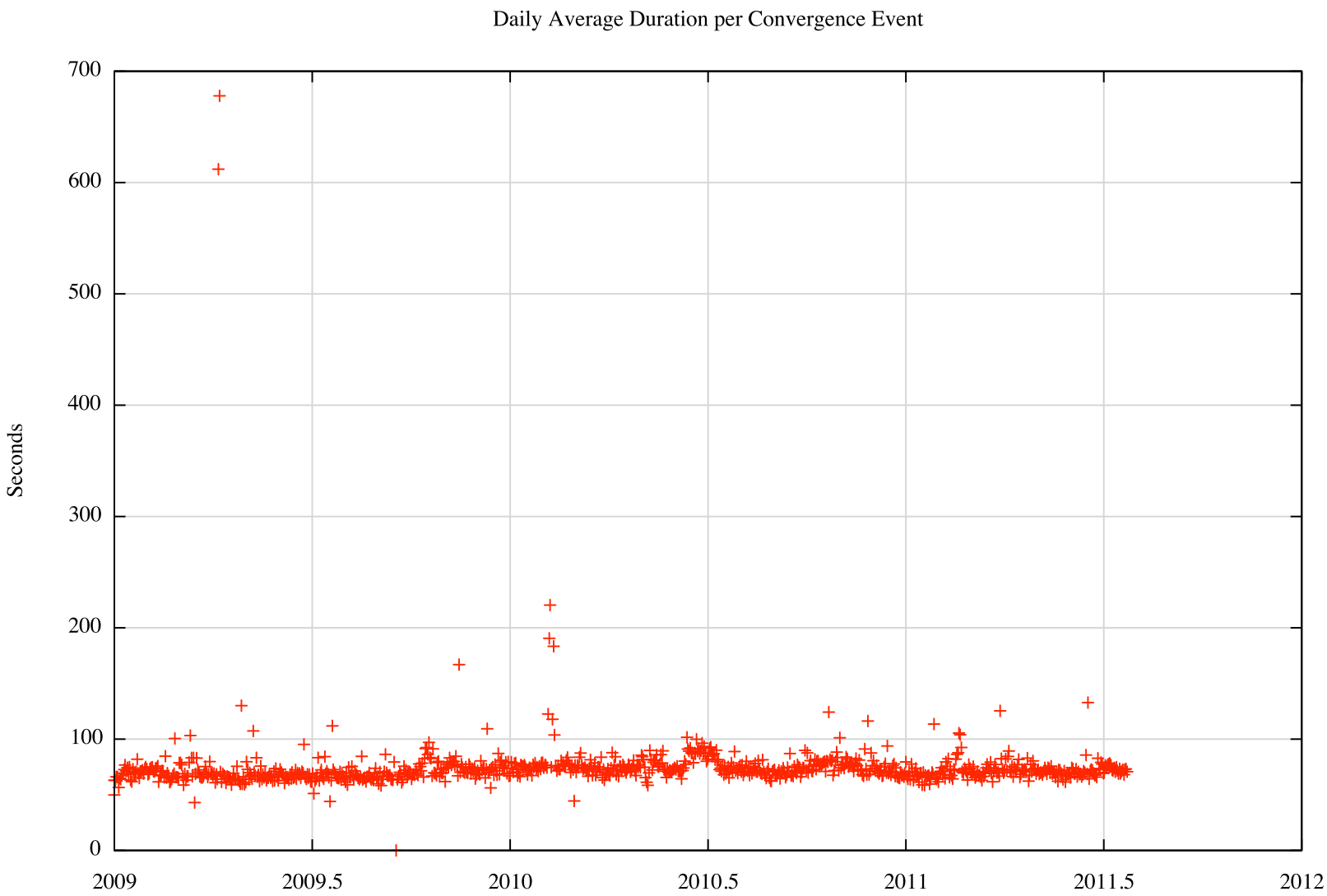
Convergence Updates



Distribution of Convergence Updates



Convergence Time



Convergence Time Distribution