Desperately Seeking Default

Geoff Huston
APNiC
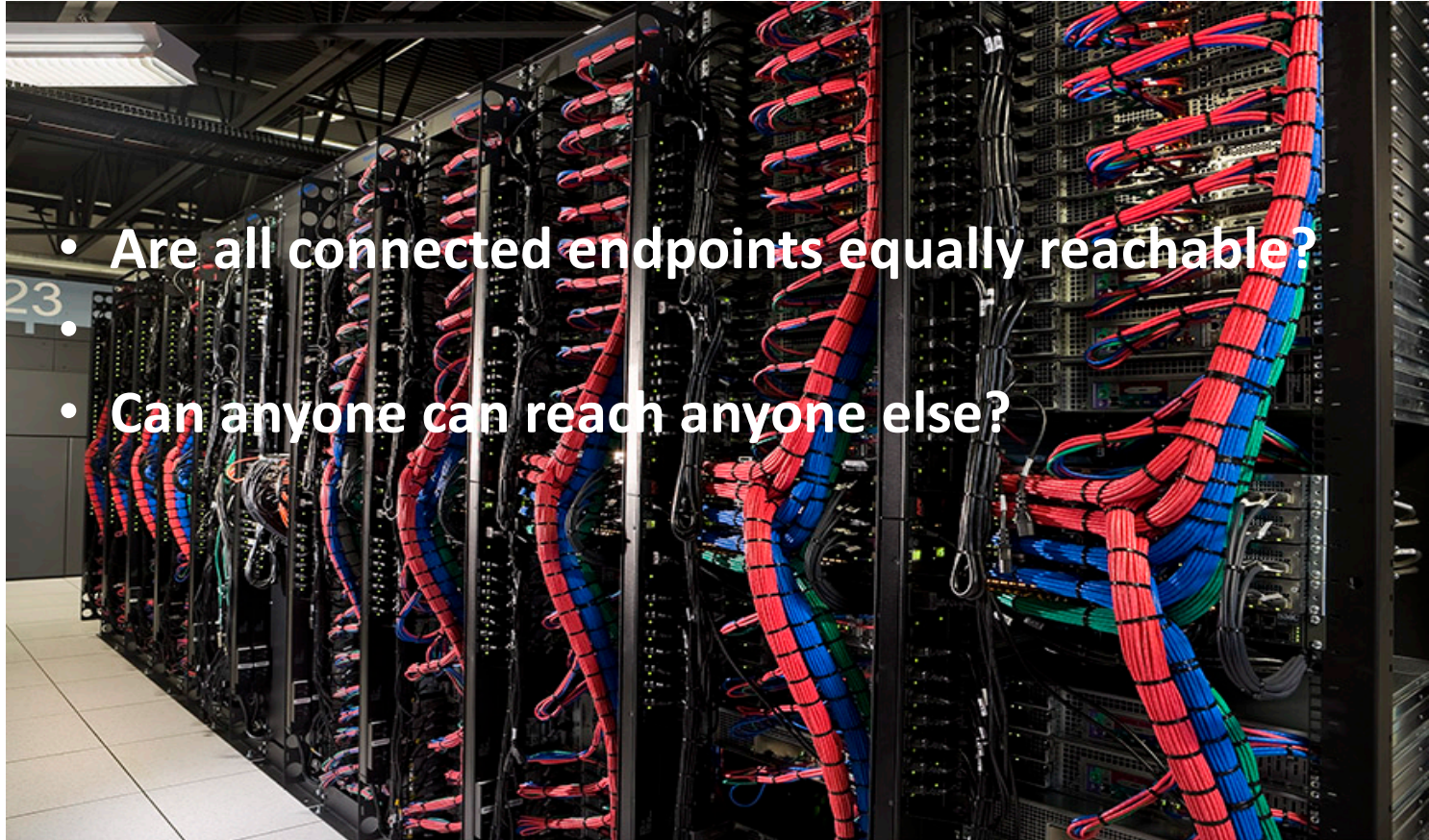
In the Telephone Network

- All telephones were equally reachable
- Anyone could dial anyone else

In the ~~Telephone~~ Network
internet

- **Are all connected endpoints equally reachable?**

- **Can anyone can reach anyone else?**

In the ~~Telephone~~ *internet* Network

- <u>Are all connected endpoints are equally reachable?</u>
- Can anyone can reach anyone else?

No!

In the ~~Telephone~~ *Internet* Network

- Are all connected endpoints are equally reachable?

No!

- Can anyone can reach anyone else?

NATs and Firewall Filters changed our conception of the internet from a peer-to-peer network to a limited client/server architecture.

Our current expectations are that if you stand up a public service on port 80 (or 443 for that matter) outside of a NAT, then maybe everyone can reach you, but otherwise, no.

In the ~~Telephone~~ *internet* Network

- All connected endpoints are equally reachable
- Can anyone can reach anyone else?

In the ~~Telephone~~ *internet* Network

- ~~All connected endpoints are equally reachable~~
- Can anyone can reach anyone else?

We might THINK this, but is it true ALL the time?

What do we see?

On the internet is everyone <u>really</u> connected to everyone else?

How We See
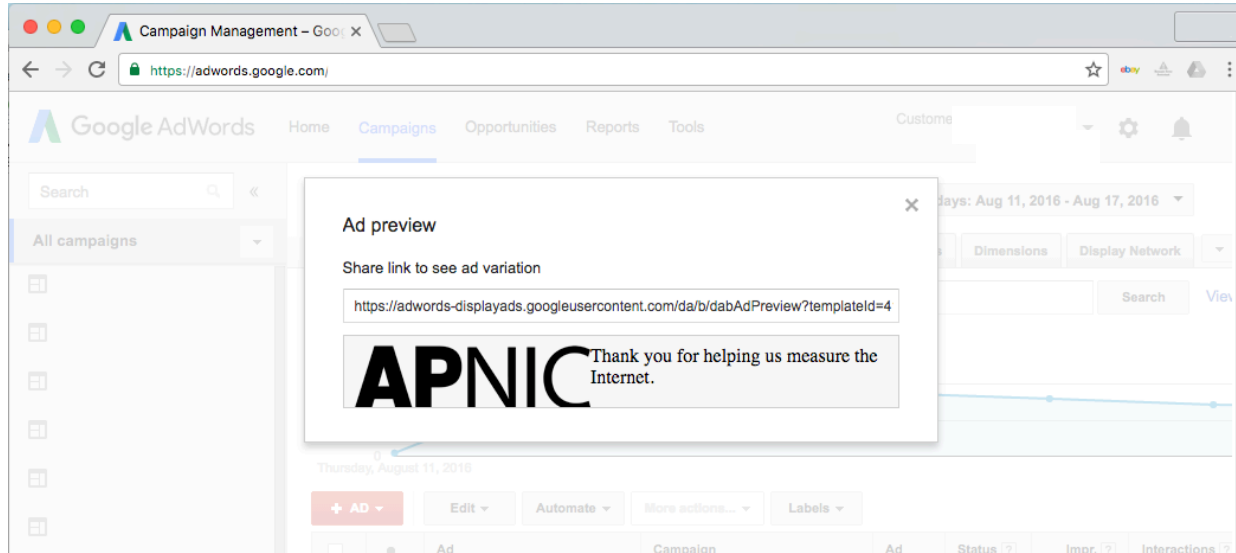
We use an online ad to present a sequence of small fetches to the user's browser

How We See
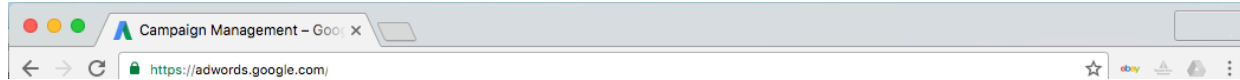
The sequence of tests is used to test a number of types of actions including fetches of IPv4, IPv6 and Dual stack
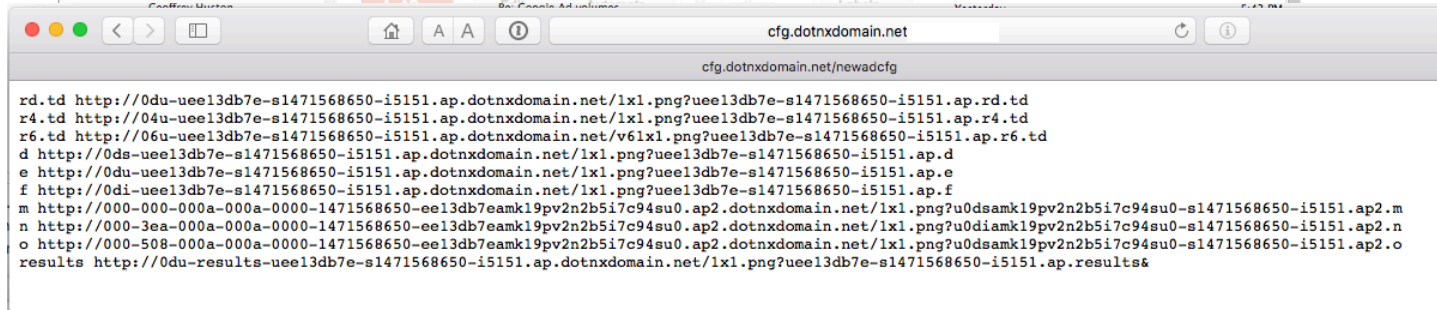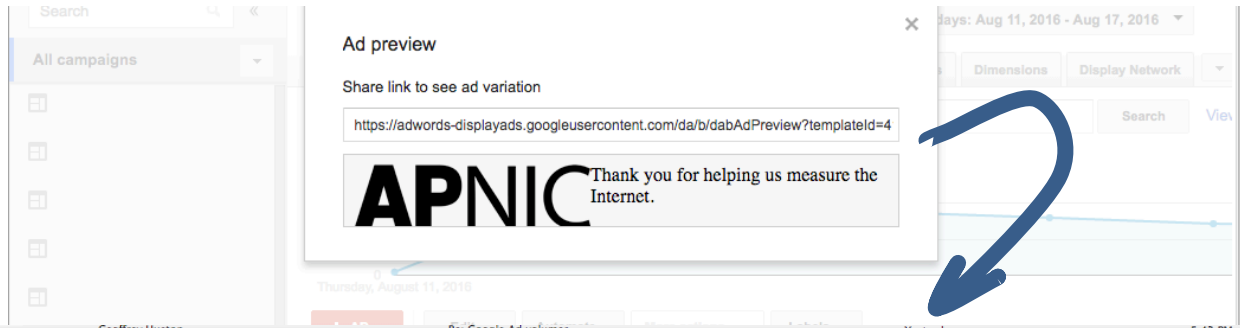
How We See

We use tcpdump to record all packet activity at the experiment's servers

What we see: Connection Failure

client

Outbound SYN

server

Busted SYN ACK.
Return path

What the server sees is an incoming SYN, but no matching incoming ACK
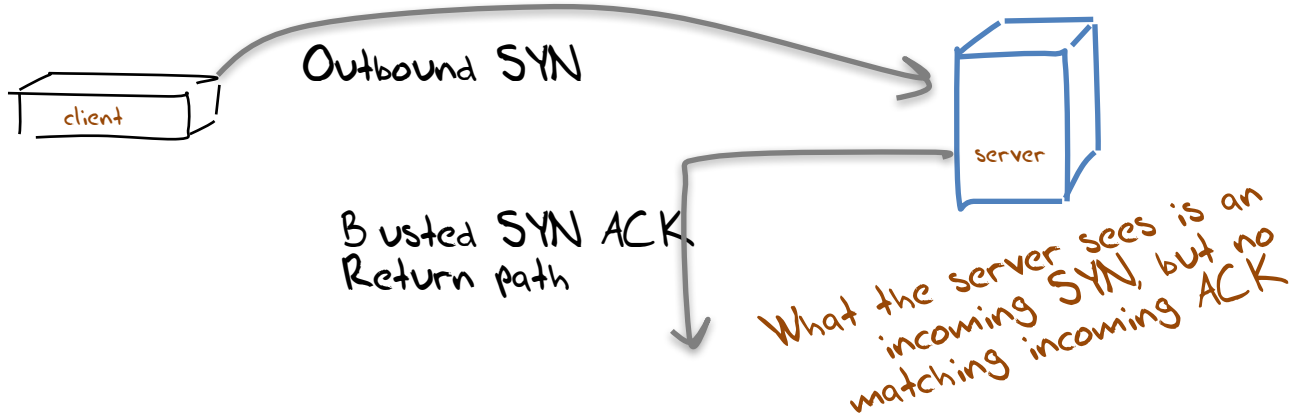
Daily IPv6 Failures

IPv6 Daily Connection Failure Rate
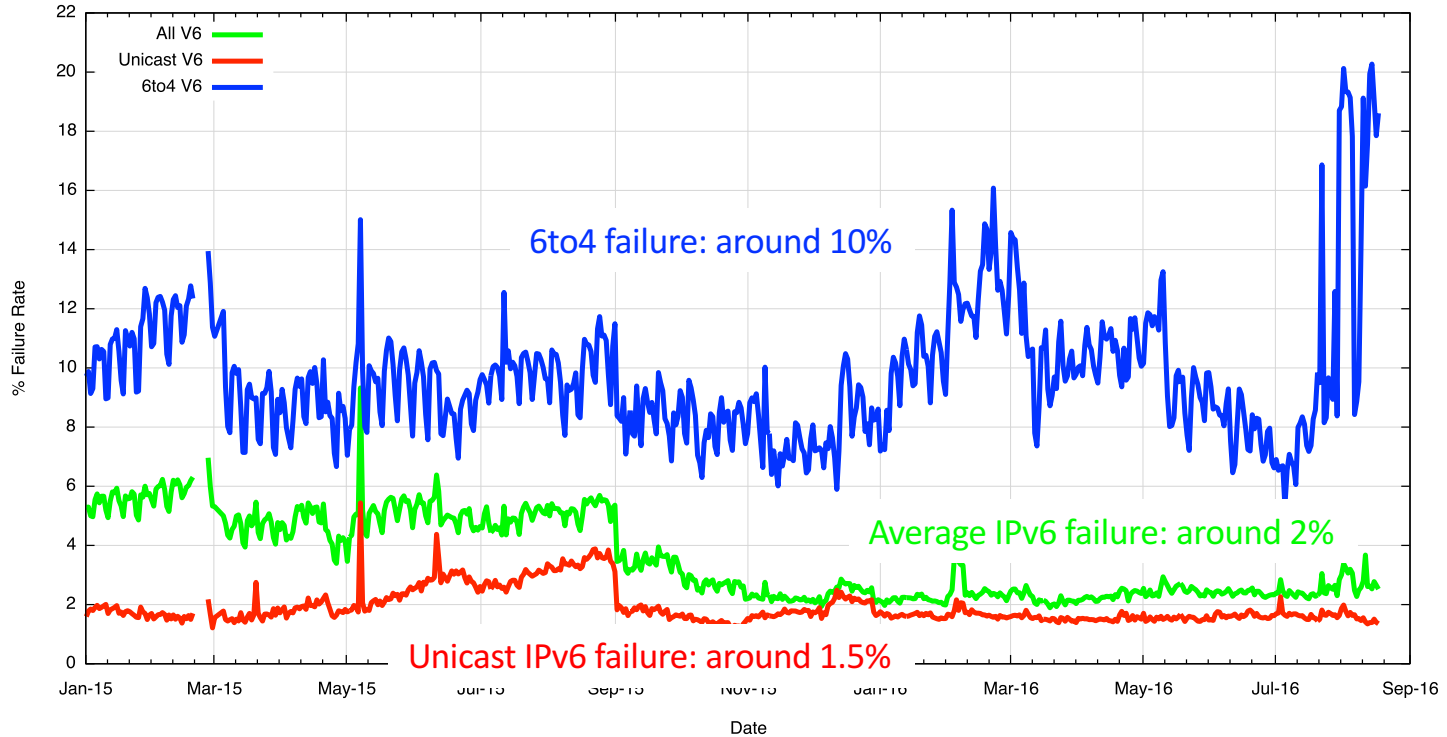
Daily IPv6 Unicast Address Failures

IPv6 Daily Connection Failure Rate
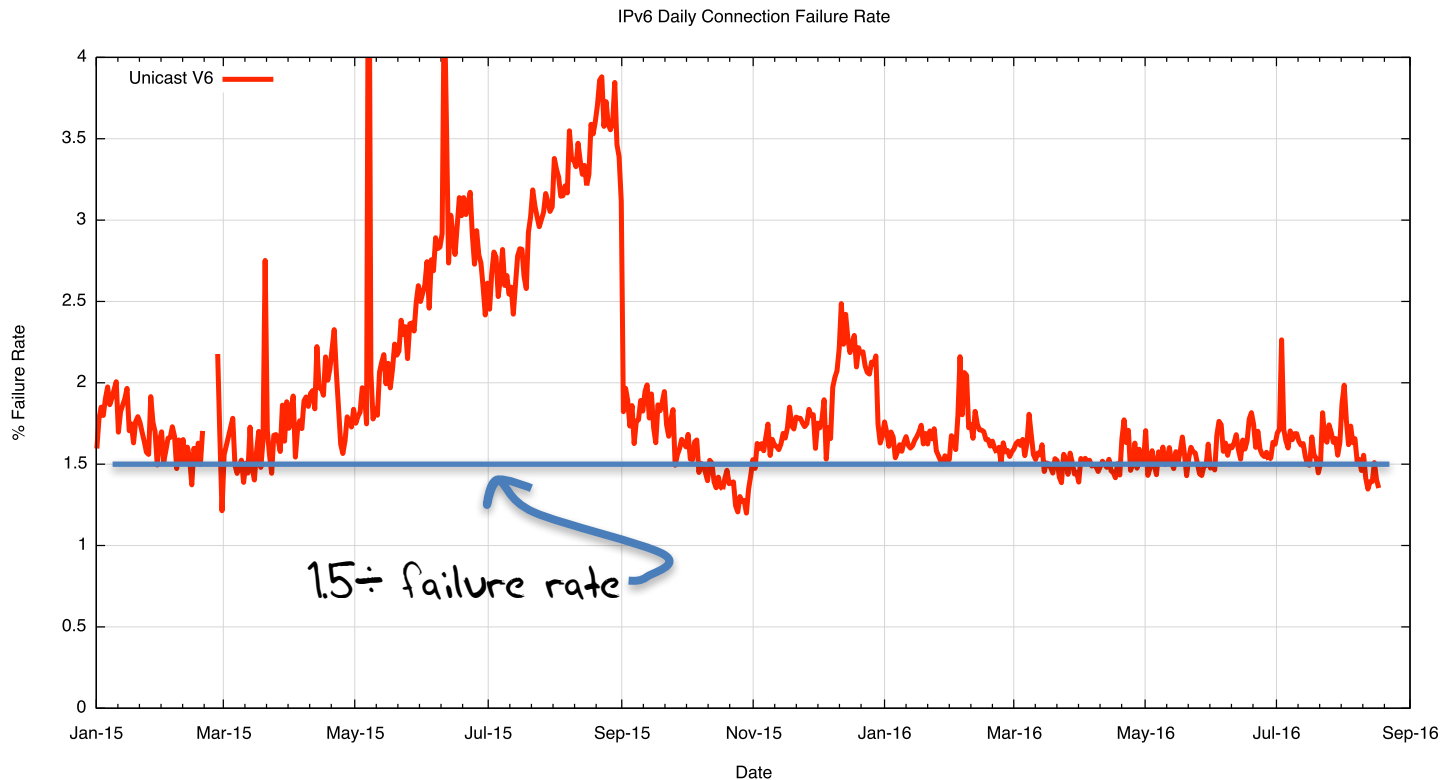
IPv6 Failures

- 1.5% failure for unicast V6 is unacceptable!
- Why is this happening
  - Auto-tunnelling?
  - Lousy CPE firmware?
  - Strange firewall filters?

- But is all of this due to local configuration / equipment?
- What is the comparable view in IPv4?

IPv4 Connection Failure

IPv4 Daily Connection Failure Rate
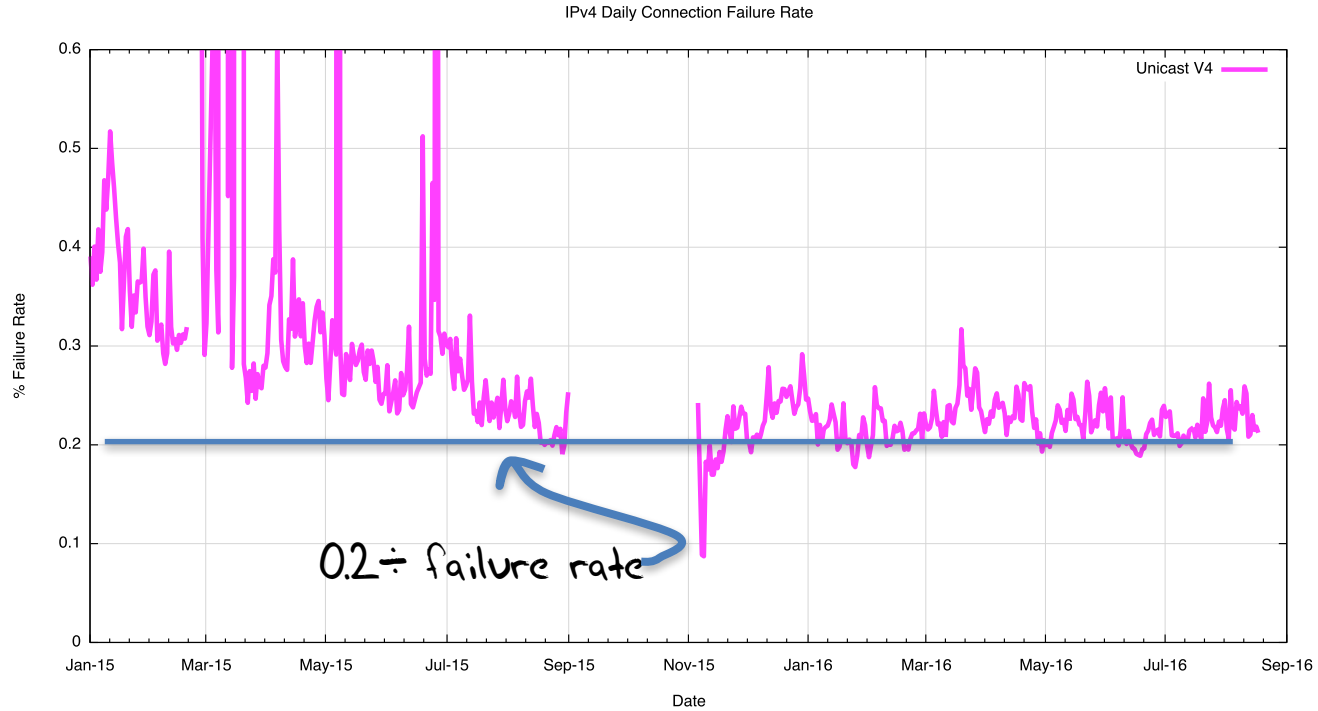
IPv4 Failures

- IPv4 failures are around 1 in 500
- And we are pretty sure its NOT:
  - Auto-tunnelling
  - Lousy CPE firmware
  - Strange firewall filters

- So what is the reason for this residual asymmetric failure rate?
- Is it asymmetric routing connectivity?

Route Views Routing Table

25 Years of Routing the Internet

This is a compound view pulled together from each of the iPv4 routing peers of Route Views and RiS
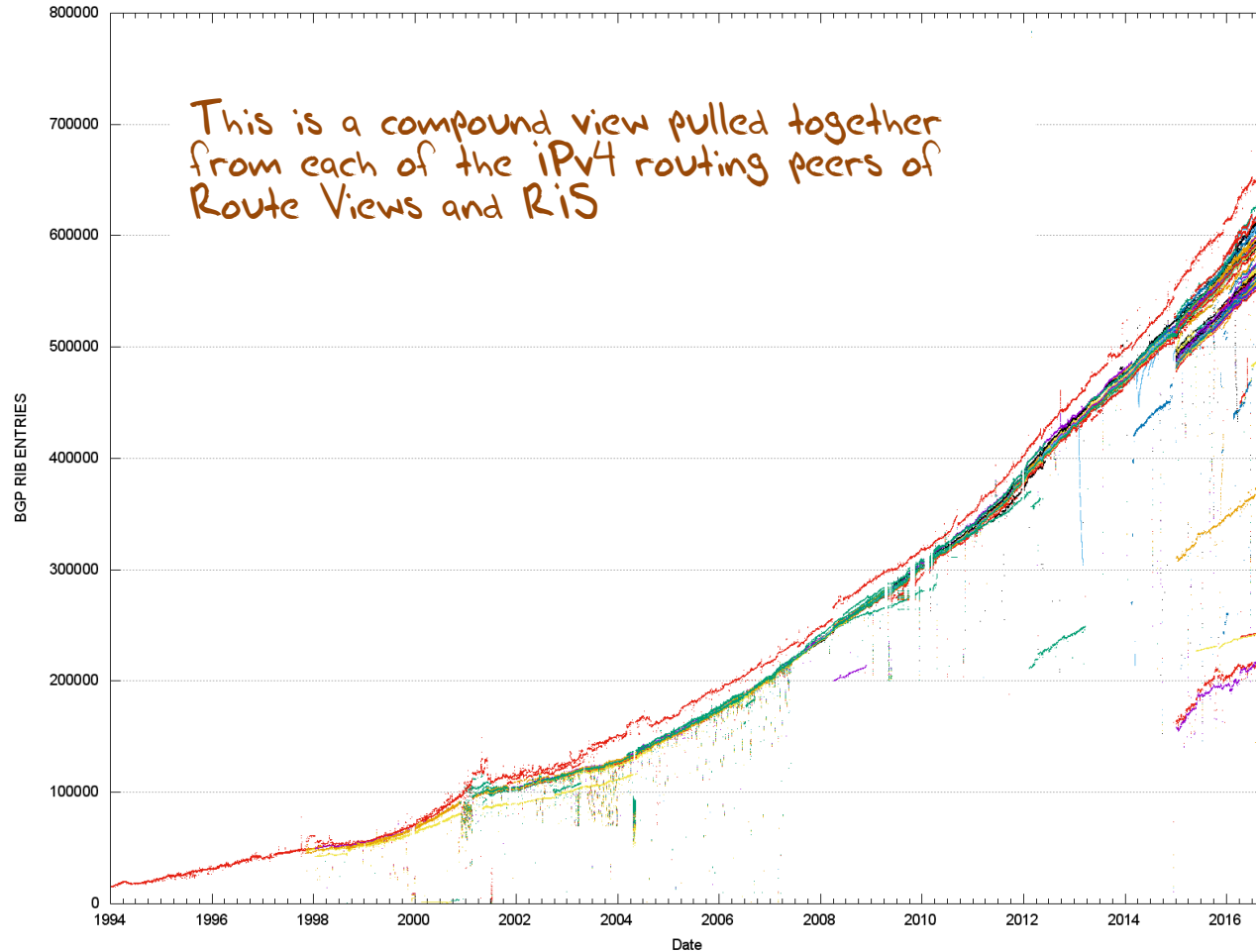
IPv4 - 2015/16 IPv4 Route Views + RIS

This is a view pulled together from each of the routing peers of Route Views and RIS

That's a range of 100,000 routes!
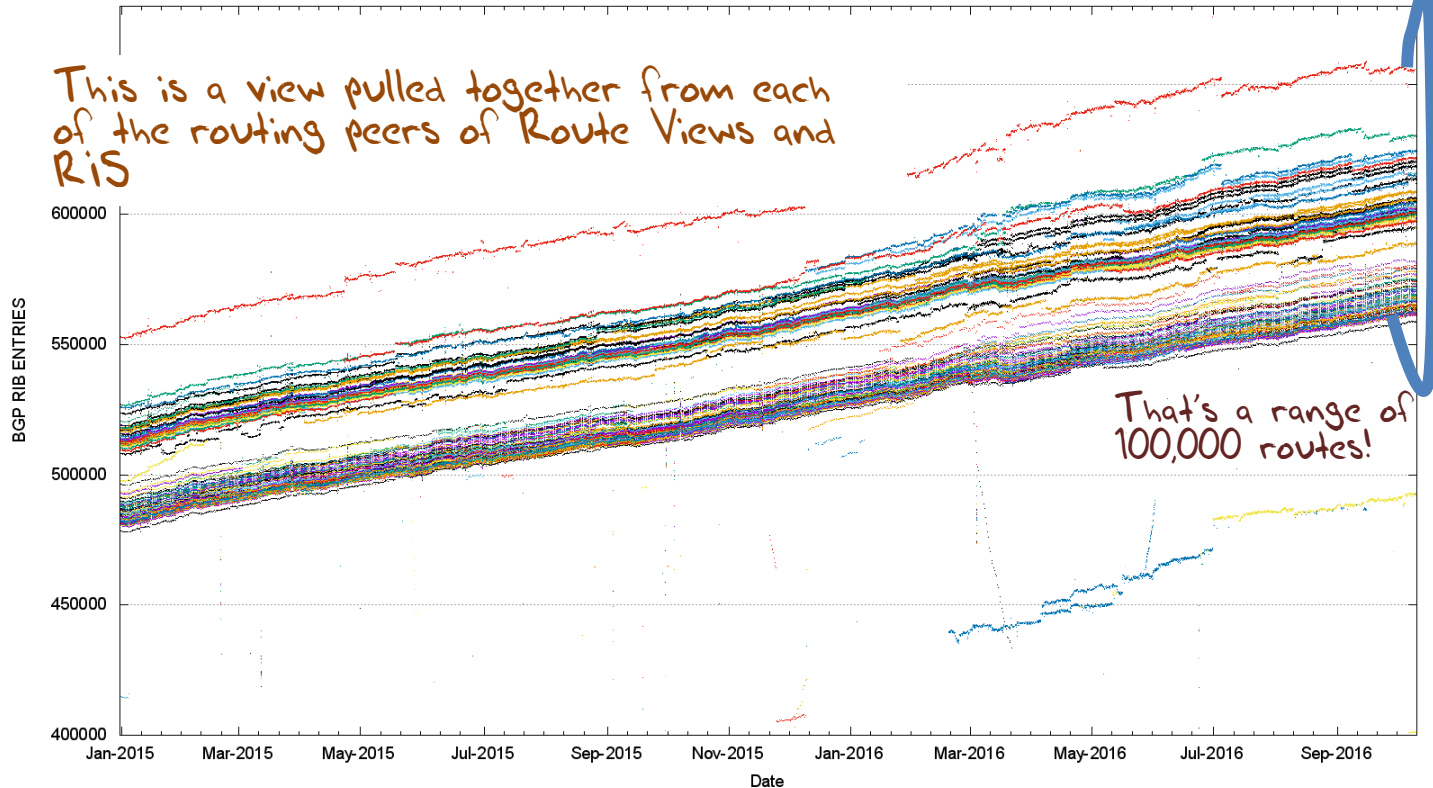
Different peers see a slightly different Internet

- But is this just traffic engineering more specifics?
- Or do different peers see a different set of reachable addresses in the routing table?

Address Span
(Route Views + RIS data sets)
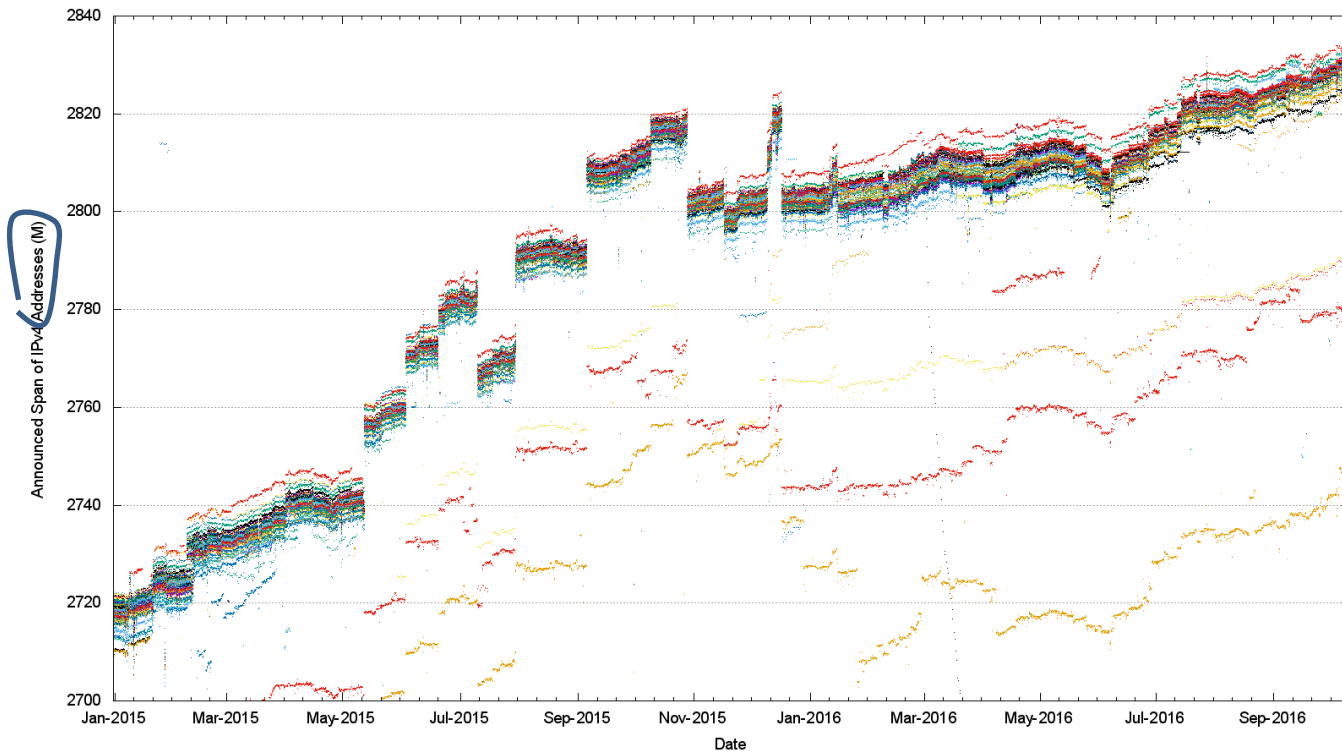
Address Span
(Route Views + RIS data sets)
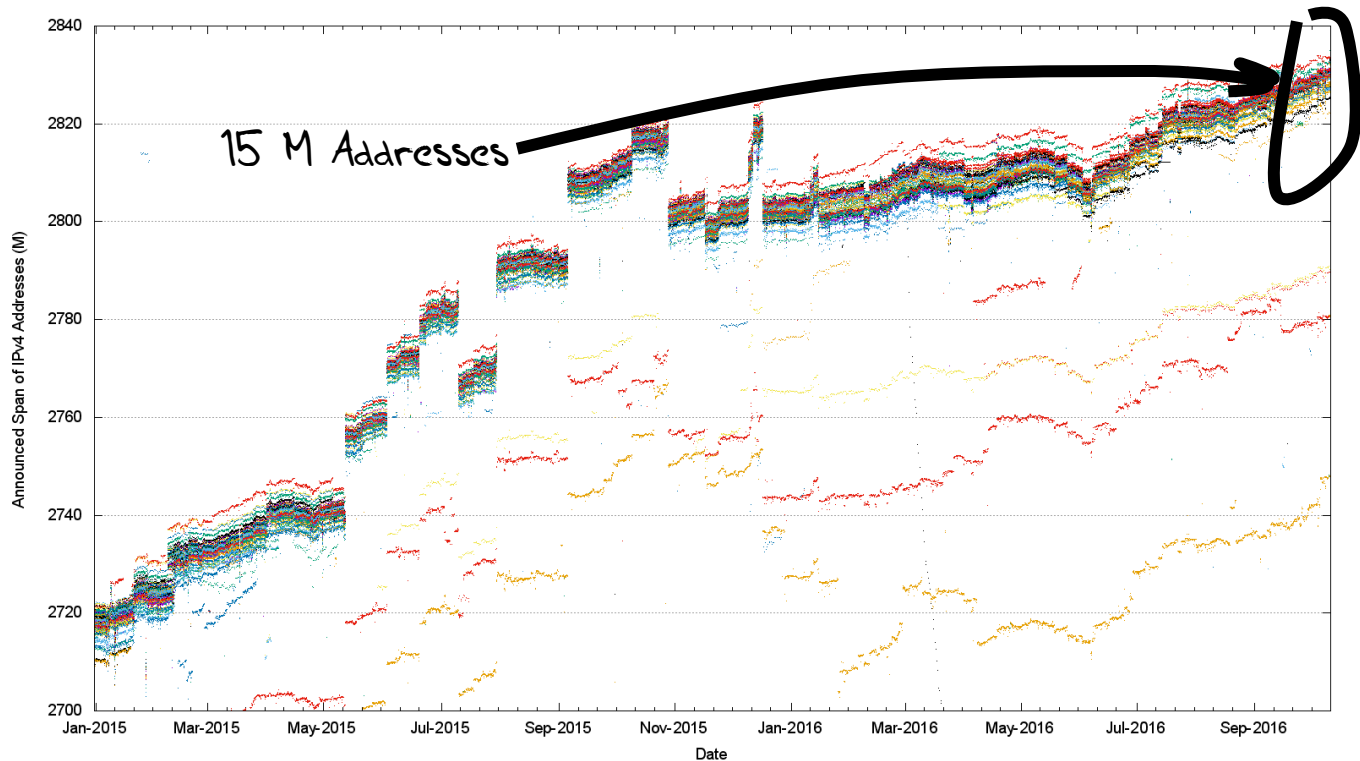
What does this mean?

- Each peer of RouteViews and RIS announces a span of addresses that appears to be a unique span.
- In total, these spans agree with other to within ~20M addresses, but this means that there are potentially some 20M uniquely addressed endpoints that cannot be reached from all other endpoints.
- This variation is stable over time for each peer, so its not transient routing that is generating this – the reasons for this difference in reachability are structural

What about IPv6?

The Route Views view of IPv6
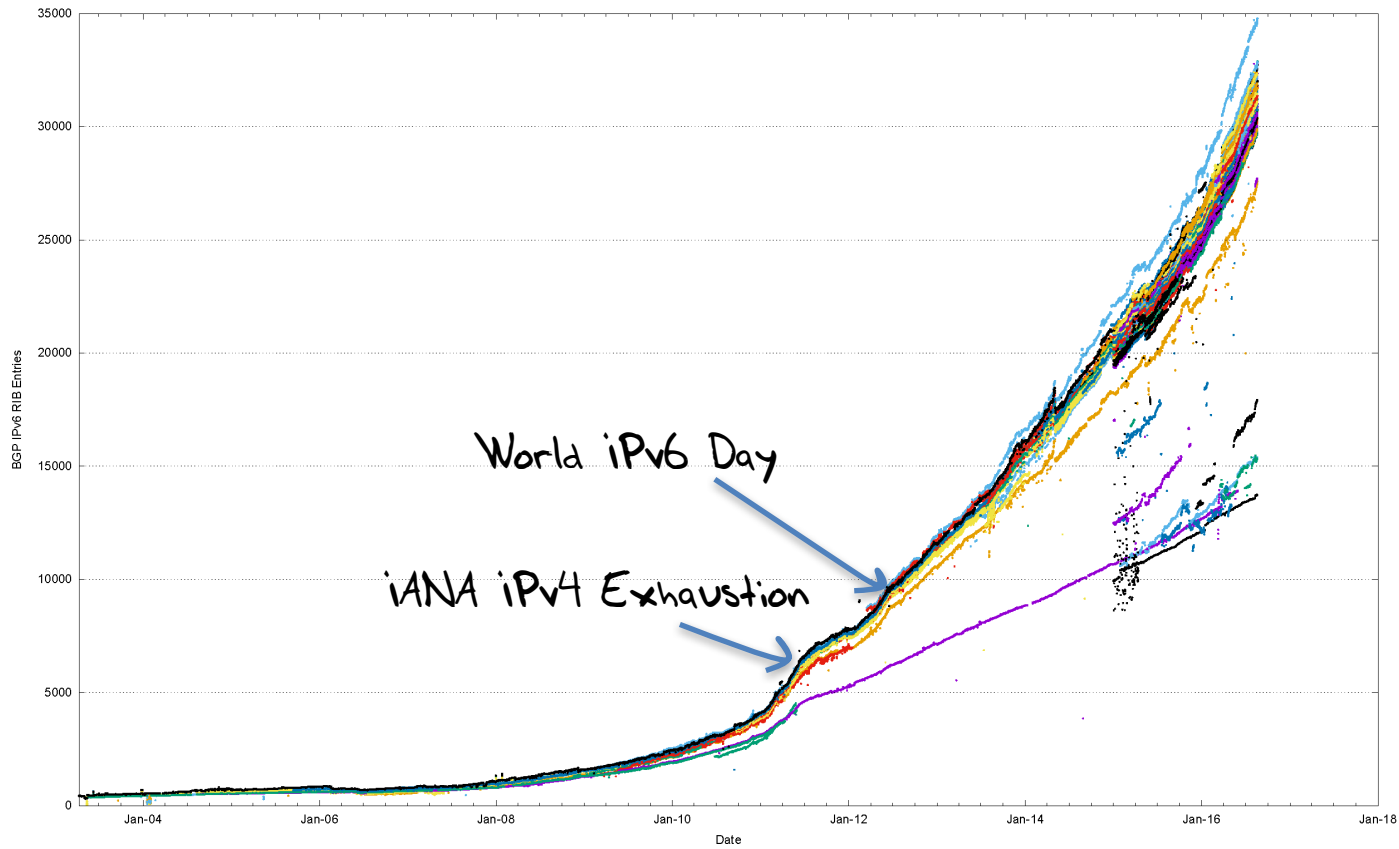
Number of IPv6 Routes in 2015/16
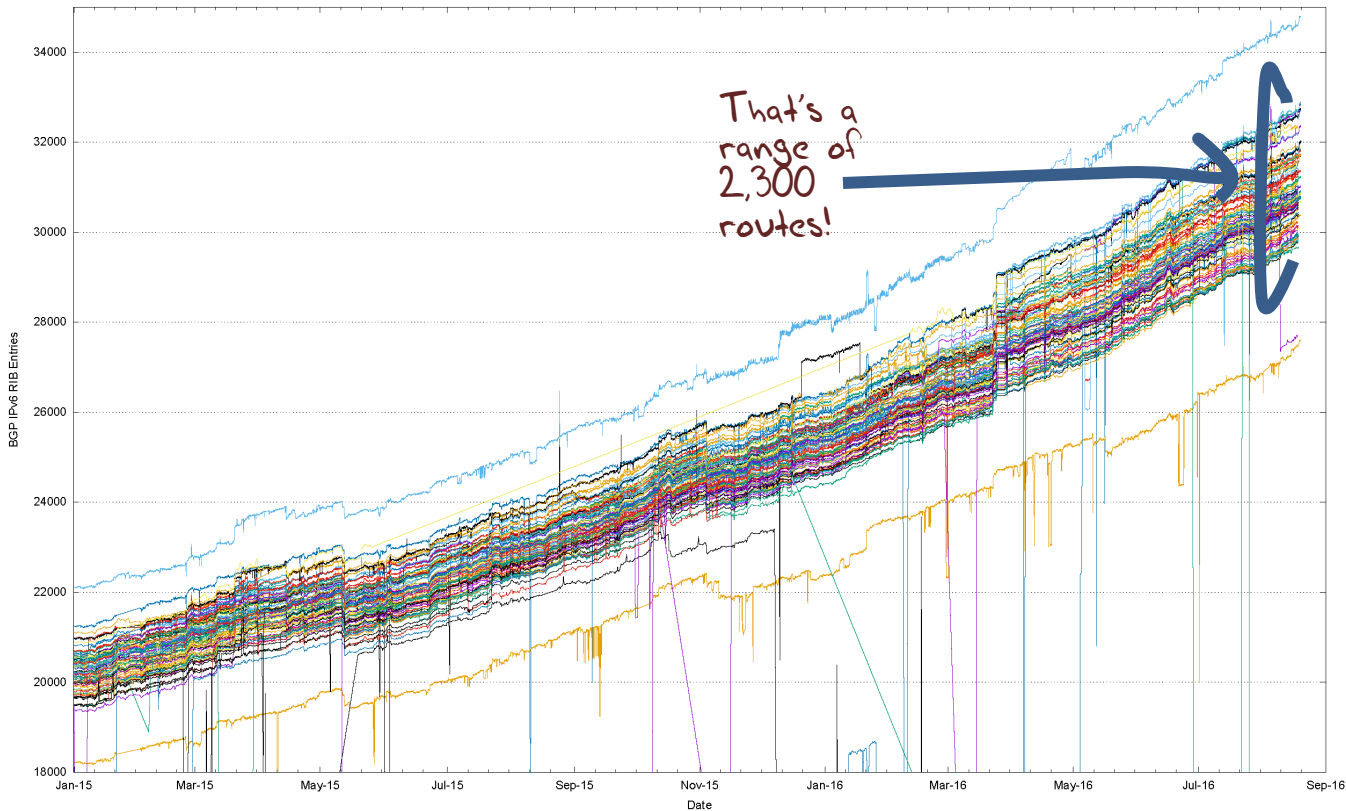
IPv6 Announced Address Span Variation (RV + RIS)
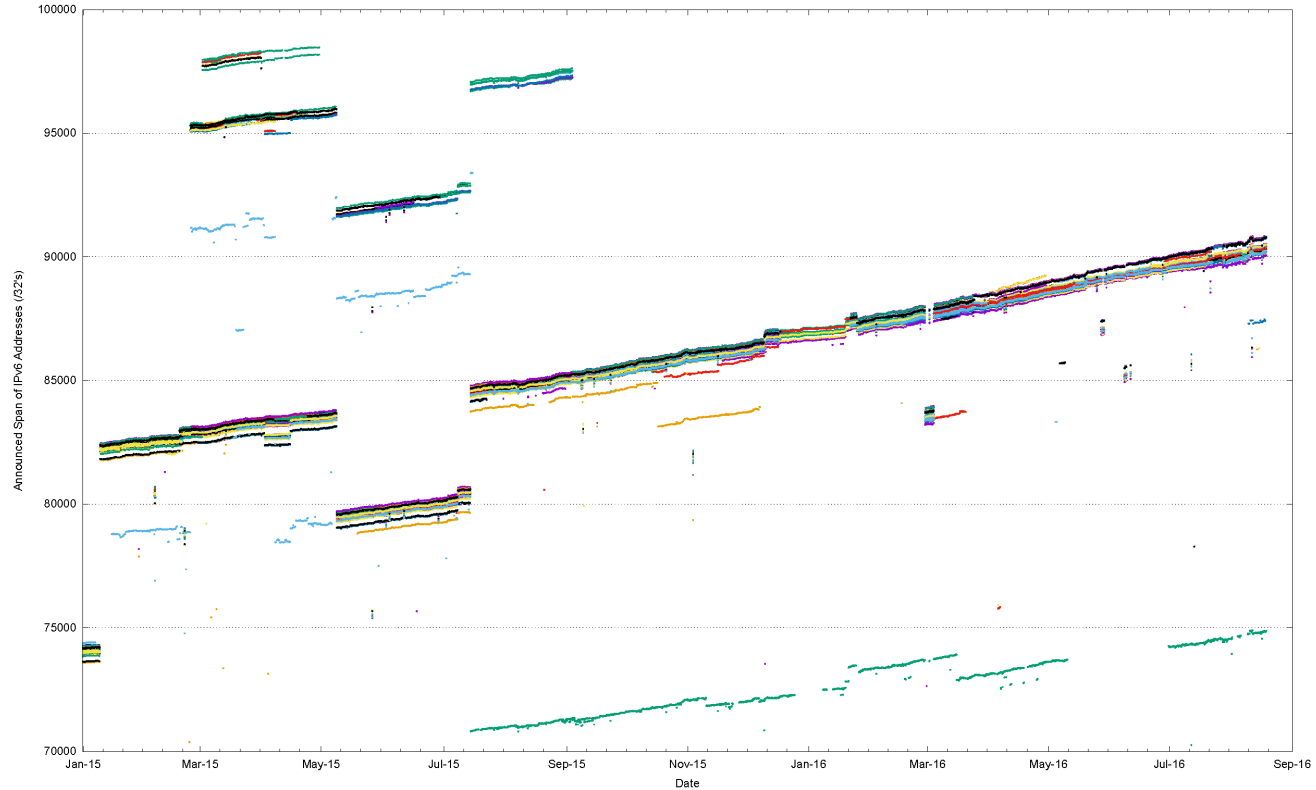
IPv6 Announced Address Span Variation (RV + RIS)
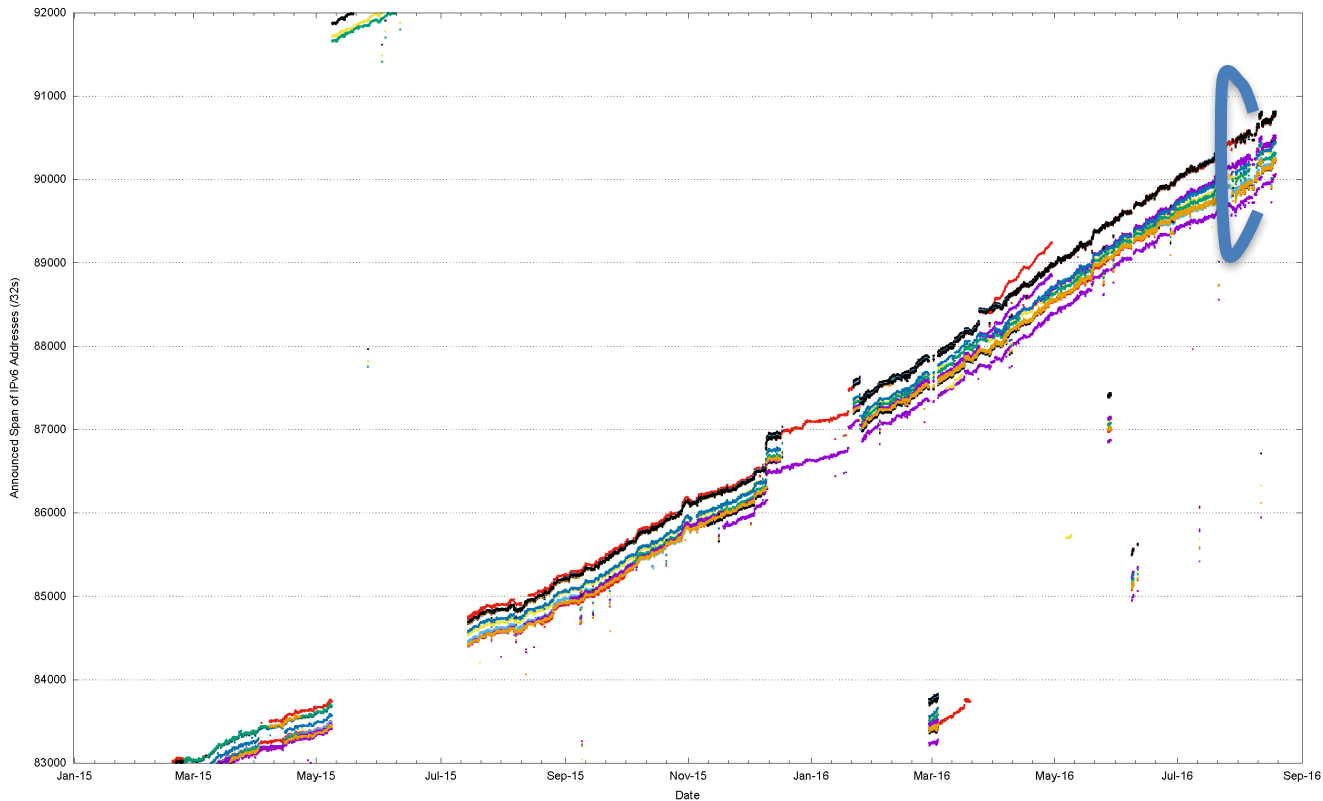
What is "default"?

We don't know!

There is no "default" route set that we can all agree on

What is "default"?

- At best "default" is an informal quorum
  - So lets define this quorum by arbitrarily setting the quorum threshold at 2/3
  - i.e. if 2/3 of the peers of a route collector advertise a route then it is part of the default quorum.
- Individual peer networks will contain route sets that differ from this quorum by having both additional prefixes and holes.
  - Lets look at the variance from the quorum

A "Quorum" deviation view of IPv4
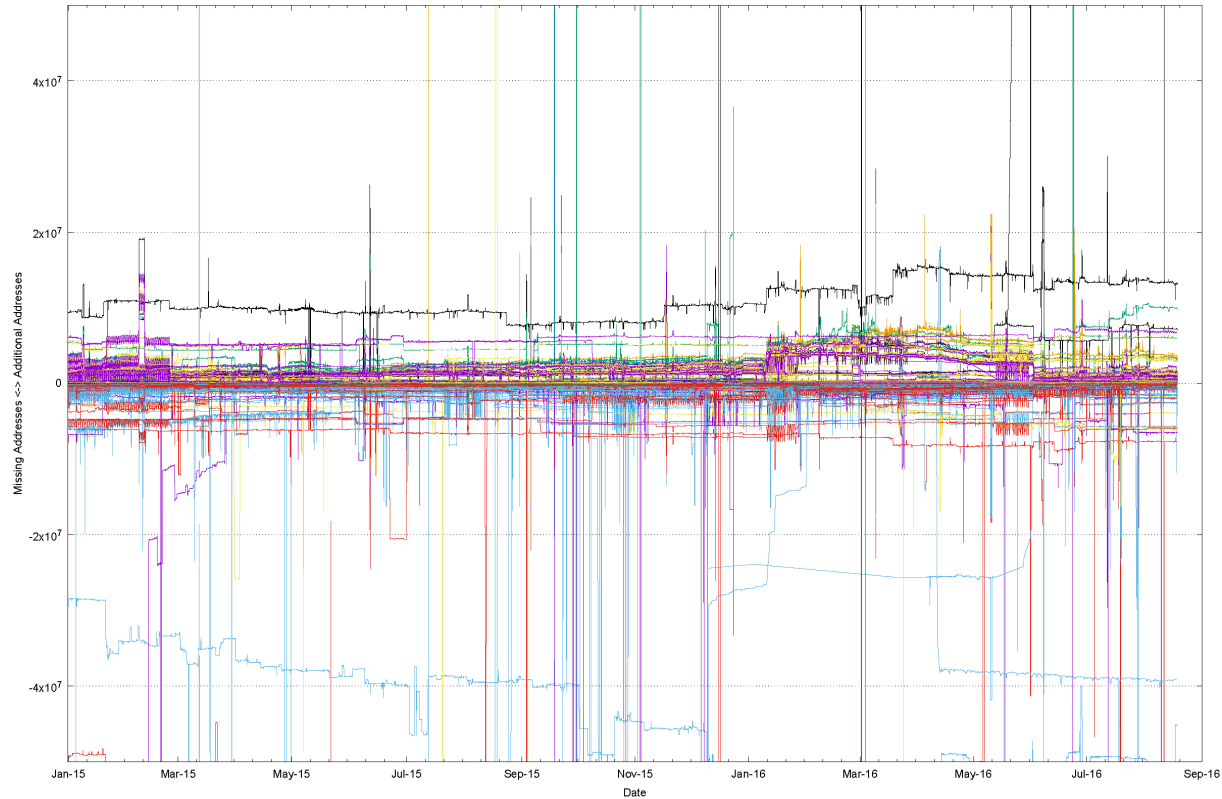
A magnified view
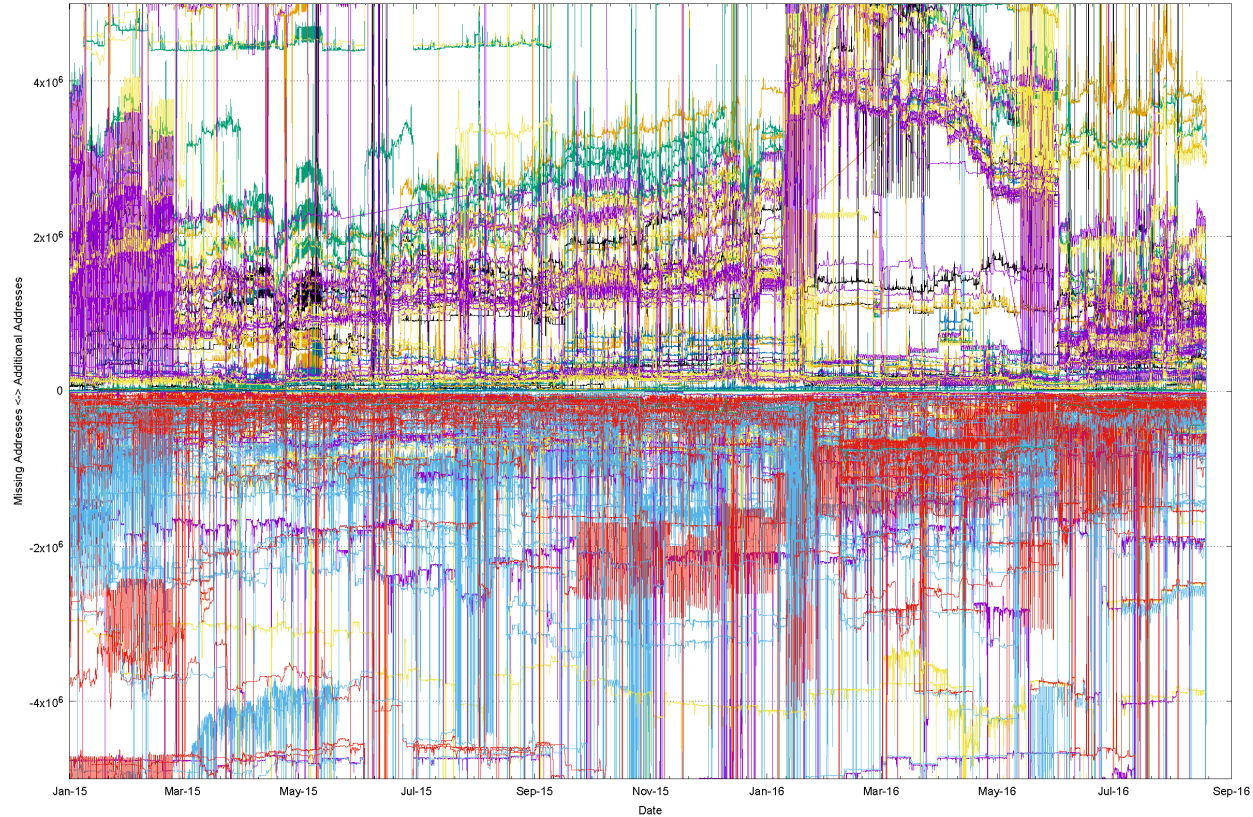
IPv6 "Quorum" Deviation
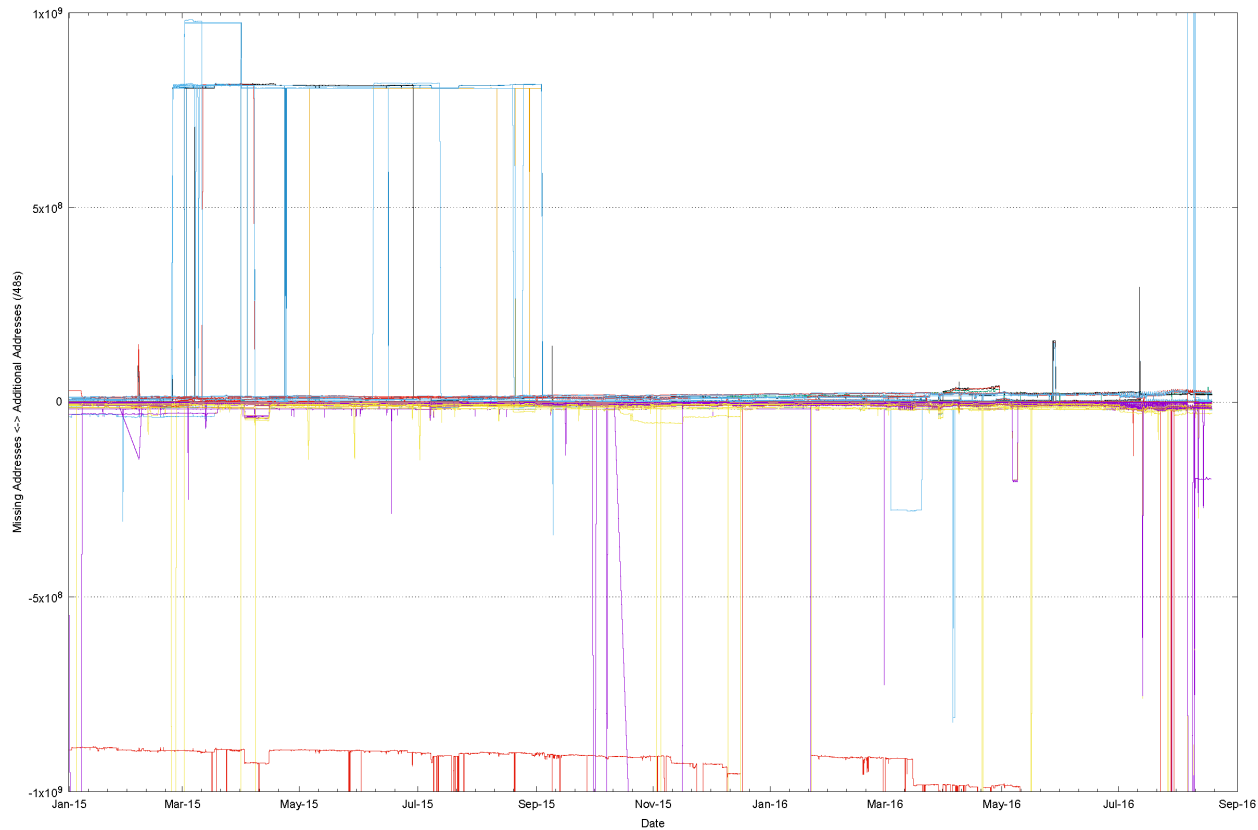
Zooming In
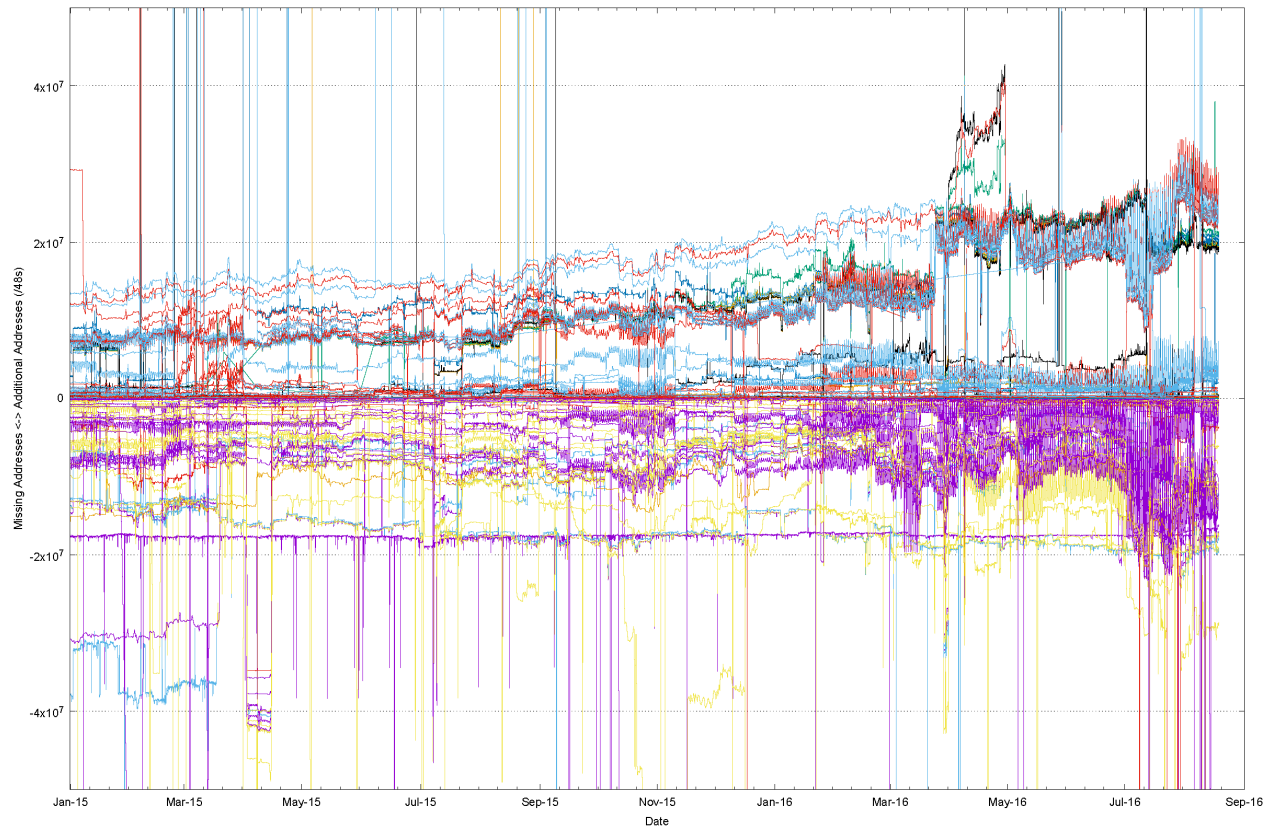
And Again
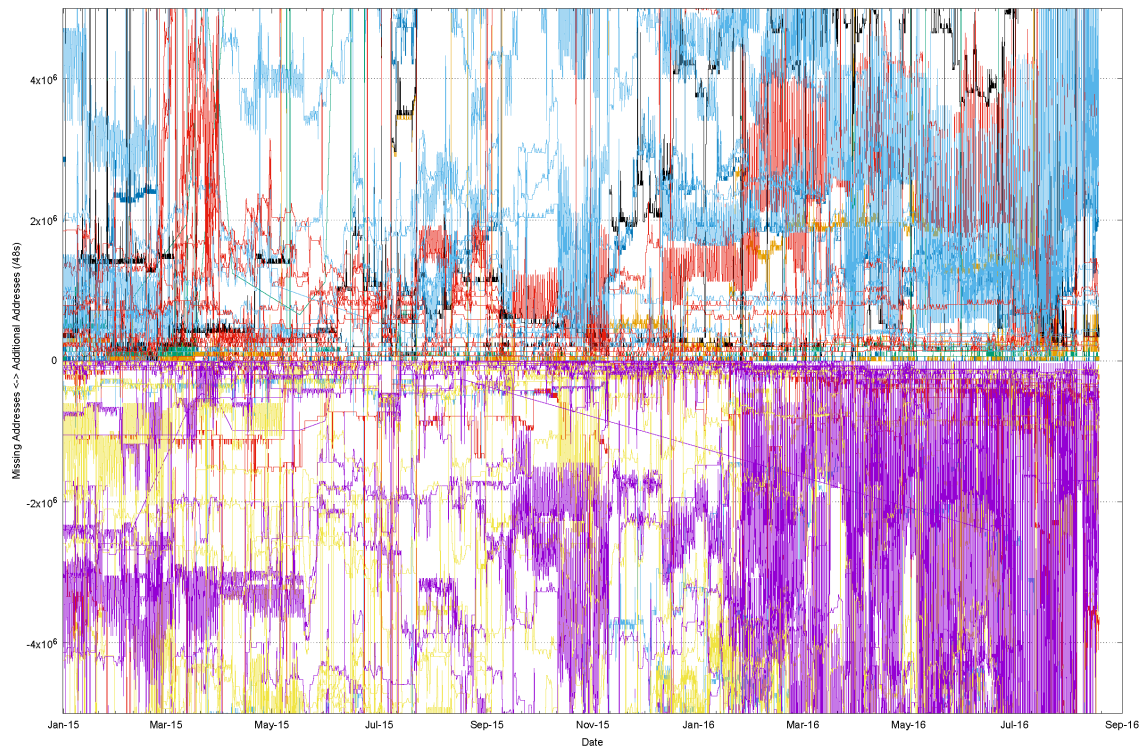
Quorum Deviation for RVA IPv4 Peers (18th August 2016)

AS	RIB	SPAN (/8s)	MISSING	EXTRA	NAME
AS57463	486,034	131.09	622,603,514	839,940	NETIX , BG
AS7660	581,864	162.96	88,091,728	969,448	APAN-JP Asia Pacific Advanced Network - Japan, JP
AS1299	583,170	165.45	45,216,000	1,280	TELIANET , SE
AS23673	629,607	168.47	7,667,968	13,087,744	ONLINE-AS Cogetel Online, Cambodia, ISP, KH
AS24441	609,412	168.35	6,401,536	9,763,840	CITYLINK-AS-KH CityLink Corporation, LTD, KH
AS34224	594,825	167.85	5,825,536	864,000	NETERRA-AS , BG
AS18106	608,422	168.23	5,740,544	7,048,080	VIEWQWEST-SG-AP Viewqwest Pte Ltd, SG
AS2497	602,117	168.17	5,626,368	5,985,792	IIJ Internet Initiative Japan Inc., JP
AS3303	595,422	167.91	4,301,312	338,736	SWISSCOM Swisscom (Switzerland) Ltd, CH
AS1668	592,954	167.93	3,856,640	150,528	AOL-ATDN - AOL Transit Data Network, US
AS3356	592,109	168.00	2,482,688	5,376	LEVEL3 - Level 3 Communications, Inc., US
AS3549	595,096	168.00	2,473,728	20,096	LVLT-3549 - Level 3 Communications, Inc., US
AS11686	599,478	168.13	1,922,560	1,566,464	ENA - Education Networks of America, US
AS1239	592,299	168.03	1,906,688	1,280	SPRINTLINK - Sprint, US
AS3130	591,111	168.05	1,699,584	13,824	RG-BIWA - RGnet, LLC, US
AS37100	598,975	168.14	1,534,976	1,385,472	SEACOM-AS, MU
AS293	615,277	168.26	1,419,776	3,337,310	ESNET - ESnet, US
AS20912	600,953	168.12	1,376,254	840,997	ASN-PANSERVICE , IT
AS6939	613,648	168.27	852,736	2,939,136	HURRICANE - Hurricane Electric, Inc., US
AS58511	616,638	168.29	806,144	3,163,136	CONNECTIVITYIT-AU Connectivity IT Pty Ltd, AU
AS47872	619,320	168.31	781,228	3,517,602	SOFIA-CONNECT-AS , BG
AS20771	617,937	168.31	777,984	3,525,121	CAUCASUS-CABLE-SYSTEM CCS Autonomous System, GE
AS6762	594,768	168.11	569,856	4,864	SEABONE-NET TELECOM ITALIA SPARKLE S.p.A., IT
AS3277	605,524	168.20	562,176	1,368,046	RUSNET-AS , RU
AS8492	607,998	168.19	541,696	1,313,536	OBIT-AS OBIT Ltd., RU
AS53364	593,422	168.13	512,512	157,952	AS-PRE2POST-2 - ZEROFAIL, US
AS3257	593,425	168.13	511,744	157,952	GTT-BACKBONE GTT, DE
AS852	594,937	168.16	499,456	325,376	ASN852 - TELUS Communications Inc., CA
AS3741	595,361	168.16	431,104	620,288	IS, ZA
AS31019	597,486	168.18	417,280	1,017,040	MEANIE Meanie # Transits and route servers:, NL
AS13030	591,282	168.18	385,024	853,762	INIT7 , CH
AS2152	596,033	168.16	290,048	477,184	CSUNET-NW - California State University, Office of the Chancellor, US
AS7018	593,164	168.14	279,040	157,696	ATT-INTERNET4 - AT&T Services, Inc., US
AS6539	598,322	168.14	222,720	16,128	GT-BELL - Bell Canada, CA
AS2914	593,425	168.14	219,904	9,984	NTT-COMMUNICATIONS-2914 - NTT America, Inc., US
AS202018	594,945	168.16	175,616	402,944	DIGITALOCEAN-ASN-3 , NL
AS3561	593,148	168.14	152,576	1,792	CENTURYLINK-LEGACY-SAVVIS - Savvis, US
AS1221	595,472	168.16	150,784	291,696	ASN-TELSTRA Telstra Pty Ltd, AU
AS5413	594,327	168.15	119,040	227,104	AS5413 , GB
AS22652	598,468	168.17	104,704	496,640	FIBRENOIRE-INTERNET - Fibrenoire Inc., CA
AS286	594,506	168.19	47,104	698,320	KPN , NL
AS40191	594,801	168.16	32,455	175,364	AS-PRE2POST-1 - ZEROFAIL, CA

Quorum Deviation for RIS IPv4 Peers

AS	RIB	SPAN (/8s)	MISSING	EXTRA	NAME
AS37989	243,051	165.84	39,676,160	947,712	DCS1-02-AS-AP DCS1 Pte Ltd, at DC02, SG
AS4777	567,778	168.19	5,953,536	6,699,364	APNIC-NSPIXP2-AS Asia Pacific Network Information Centre, JP
AS3549	556,838	168.00	2,651,904	209,536	LVLT-3549 - Level 3 Communications, Inc., US
AS50300	559,388	168.07	1,817,600	531,968	CUSTDC , GB
AS50763	369,369	168.07	1,742,848	428,288	MCKAYCOM , GB
AS3257	555,248	168.12	686,080	216,064	GTT-BACKBONE GTT, DE
AS4608	572,611	168.28	684,800	2,897,776	APNIC-SERVICES Asia Pacific Network Information Centre, AU
AS57821	556,704	168.12	590,848	253,696	NONATTACHED-AS , DE
AS7018	555,116	168.14	452,864	344,320	ATT-INTERNET4 - AT&T Services, Inc., US
AS6453	555,023	168.12	440,064	64,000	AS6453 - TATA COMMUNICATIONS (AMERICA) INC, US
AS2914	555,431	168.12	392,960	65,280	NTT-COMMUNICATIONS-2914 - NTT America, Inc., US
AS8758	559,880	168.19	389,376	1,112,320	IWAY , CH
AS1103	565,166	168.21	388,352	1,455,786	SURFNET-NL SURFnet, The Netherlands, NL
AS1836	558,296	168.17	238,080	712,448	GREEN green.ch AG Autonomous System, CH
AS13030	552,829	168.16	238,080	586,496	INIT7 , CH
AS29608	555,680	168.14	231,168	101,632	WAN2MANY-AS , FR
AS25160	558,803	168.15	231,168	354,048	VORBOSS_AS , GB
AS8283	560,112	168.25	212,224	2,007,082	COLOCLUE-AS Netwerkvereniging Coloclue, Amsterdam, Netherlands, NL
AS8455	556,536	168.24	211,456	1,902,848	ATOM86-AS Schuberg Philis B.V. trading as atom86, NL
AS57381	558,965	168.17	204,032	637,114	FNUTT , NO
AS50304	558,965	168.17	202,752	636,090	BLIX , NO
AS15435	556,571	168.17	198,656	573,504	KABELFOON CAIW Autonomous System, NL
AS8468	558,362	168.17	123,392	541,184	ENTANET ENTANET International Limited, GB
AS12859	558,716	168.19	123,392	879,360	NL-BIT BIT BV, NL
AS56730	560,989	168.17	114,176	541,184	WIREHIVE-AS , GB
AS22652	559,492	168.17	90,368	495,872	FIBRENOIRE-INTERNET - Fibrenoire Inc., CA
AS286	556,245	168.18	78,848	743,120	KPN , NL
AS8607	555,971	168.15	28,416	86,528	TIMICO United Kingdom, GB

Quorum Deviation: IPv6, RVA

AS	RIB	SPAN (/32s)	MISSING (/48)	EXTRA (/48)	NAME
AS30071	27,515	4,907,457,131	1,023,659,016	65,538	OCCAID - TowardEX Technologies International, Inc., US
AS1239	30,371	5,902,448,671	28,601,938	0	SPRINTLINK - Sprint, US
AS33437	30,784	5,931,199,283	19,099,133	19,247,807	HOTNIC - Hotnic LLC, US
AS6939	30,761	5,931,199,283	19,099,133	19,247,807	HURRICANE - Hurricane Electric, Inc., US
AS701	30,665	5,913,256,804	17,793,806	1	UUNET - MCI Communications Services, Inc. d/b/a Verizon Business,US
AS7018	30,731	5,913,545,260	17,734,736	229,387	ATT-INTERNET4 - AT&T Services, Inc., US
AS209	31,686	5,932,960,036	17,403,918	19,313,345	CENTURYLINK-US-LEGACY-QWEST - Qwest Communications Company, LLC, US
AS53364	31,047	5,913,912,082	17,138,527	0	AS-PRE2POST-2 - ZEROFAIL, US
AS3257	31,048	5,913,977,618	17,072,991	0	GTT-BACKBONE GTT, DE
AS2914	30,832	5,914,140,702	16,975,443	65,536	NTT-COMMUNICATIONS-2914 - NTT America, Inc., US
AS40191	31,715	5,914,960,678	16,155,467	65,536	AS-PRE2POST-1 - ZEROFAIL, CA
AS22652	32,161	5,918,991,499	12,190,205	131,095	FIBRENOIRE-INTERNET - Fibrenoire Inc., CA
AS13030	31,480	5,926,958,160	6,062,634	1,970,185	INIT7 , CH
AS37100	30,992	5,927,882,105	3,637,781	469,277	SEACOM-AS, MU
AS2497	31,207	5,949,252,620	950,892	19,152,903	IIJ Internet Initiative Japan Inc., JP
AS34224	32,905	5,949,544,751	819,206	19,313,348	NETERRA-AS , BG
AS31019	32,896	5,951,743,292	656,398	21,349,081	MEANIE Meanie # Transits and route servers:, NL
AS18106	33,106	5,950,559,547	525,323	20,034,261	VIEWQWEST-SG-AP Viewqwest Pte Ltd, SG
AS57463	31,598	5,949,808,939	425,989	19,184,319	NETIX , BG
AS3741	31,889	5,950,296,380	264,195	19,509,966	IS, ZA
AS47872	32,873	5,950,268,013	229,396	19,446,800	SOFIA-CONNECT-AS , BG
AS393406	31,586	5,950,200,121	229,381	19,378,893	DIGITALOCEAN-ASN-NY3 - Digital Ocean, Inc., US
AS202018	31,573	5,950,200,108	229,381	19,378,880	DIGITALOCEAN-ASN-3 , NL
AS62567	31,578	5,950,200,109	229,380	19,378,880	DIGITALOCEAN-ASN-NY2 - Digital Ocean, Inc., US
AS20912	31,716	5,950,396,919	229,379	19,575,689	ASN-PANSERVICE , IT
AS200130	31,578	5,950,200,110	229,379	19,378,880	DIGITALOCEAN-ASN-1 , EU
AS24441	32,262	5,951,248,894	229,379	20,427,664	CITYLINK-AS-KH CityLink Corporation, LTD, KH
AS3277	32,787	5,951,785,285	229,379	20,964,055	RUSNET-AS , RU

Quorum Deviation: IPv6, RIS

AS	RIB	SPAN (/32s)	MISSING (/48)	EXTRA (/48)	NAME
AS13030	31,480	5,929,887,998	19,516,932	25,900,240	HURRICANE - HurricaneElectric, Inc., US
AS57463	31,598	5,910,834,989	12,817,159	147,458	VORBOSS_AS , GB
AS209	31,686	5,912,954,870	10,927,192	377,372	ATT-INTERNET4 - AT&TServices, Inc., US
AS393406	31,586	5,913,254,942	10,463,283	213,535	CW Vodafone Ltd,GB
AS34224	32,905	5,913,616,121	10,101,561	212,992	NTT-COMMUNICATIONS-2914 - NTTAmerica, Inc., US
AS7018	30,731	5,917,651,397	7,741,835	1,888,542	WAN2MANY-AS , FR
AS53364	31,047	5,918,466,646	7,086,087	2,048,043	FIBRENOIRE-INTERNET - FibrenoireInc., CA
AS40191	31,715	5,917,811,232	6,496,295	802,837	CATALYST2-AS , IE
AS2497	31,207	5,923,059,225	2,434,081	1,988,616	PORTLANE www.portlane.com, SE
AS37100	30,992	5,947,841,507	1,694,497	26,031,314	TIMICO United Kingdom,GB
AS22652	32,161	5,927,156,547	1,122,317	4,774,174	RETN-AS , UA
AS3277	32,787	5,926,171,184	892,937	3,559,431	ATOM86-AS Schuber PhilisB.V. trading as atom86, NL
AS8758	29,668	5,951,708,764	706,730	28,910,804	IWAY , CH
AS18106	33,106	5,926,662,683	401,440	3,559,433	INIT7 , CH
AS3257	31,048	5,949,639,835	354,936	26,490,081	CUSTDC , GB
AS2914	30,832	5,927,514,683	335,892	4,345,885	GREEN green.ch AGAutonomous System, CH
AS47872	32,873	5,930,568,251	270,350	7,333,911	COLOCLUE-AS Netwerkvereniging Coloclue,Amsterdam, Netherlands, NL
AS3741	31,889	5,949,935,372	256,012	26,686,694	WIREHIVE-AS , GB
AS202018	31,573	5,949,347,588	253,969	26,096,867	NL-BIT BIT BV,NL
AS31019	32,896	5,949,345,529	190,477	26,031,316	DIGIWEB-AS , IE
AS24441	32,262	5,927,684,889	139,619	4,319,818	SURFNET-NL SURFnet, TheNetherlands, NL
AS20912	31,716	5,949,478,636	57,368	26,031,314	MCKAYCOM , GB
AS701	30,665	5,949,413,125	57,357	25,965,792	NONATTACHED-AS , DE
AS200130	31,578	5,949,413,111	57,357	25,965,778	KABELFOON CAIW AutonomousSystem, NL
AS1239	30,371	5,949,544,187	57,356	26,096,853	FNUTT , NO
AS33437	30,784	5,949,544,184	57,356	26,096,850	BLIX , NO
AS6939	30,761	5,949,544,184	57,356	26,096,850	NIXCZ NIX.CZ z.s.p.o.,CZ
AS62567	31,578	5,949,675,457	57,356	26,228,123	INTEROUTE 25 CanadaSquare, Canary Wharf, 31st Floor, GB

It's structural, not temporal

- There is a visible stability to this deviation from the quorum route set
  - The variation from the quorum is long–term stable, and does not rapidly self-correct – its not a transient routing state

- We appear to assume that all Tier 1 providers, and their Tier 2, 3, … resellers offer the same reachability set as each other – i.e. "default" is consistent everywhere
- But this is not necessarily the case all the time for every address in the routing system
- "Default" appears to vary by provider and by location
  - E.g.: 25 April, 1600 UTC:

    AS2914: RouteViews   2,808,560,896 addresses

    RIS: 2,807,358,208

"Default" is a market outcome

- There is no "global route arbiter"
- There is no way to enrol a route into a global Internet default route set
- There is no single common definition of "default"
- Instead "default" is a market outcome
  - You buy default from a transit
    - You hope your transit is offering you what it promised – but you just can't tell
  - You add your route to default via a transit
    - You hope that this will propagate reachability tp your network to all parts of the Internet – but you just can't tell

So What?

Surely all this is patched up by the widespread use of a routing default entry in addition to specific routes? (*)

* Internet Optometry: Assessing the Broken Glasses in Internet Reachability",
R. Bush, O. Maennel, M. Roughan, S Uhlig, ACM SIGCOMM IMC, 2009

So What?

Surely all this is patched up by the widespread use of a routing default entry in addition to specific routes?
  – Well, not really
  – Default points along upstream transits
  – It does not patch downstreams
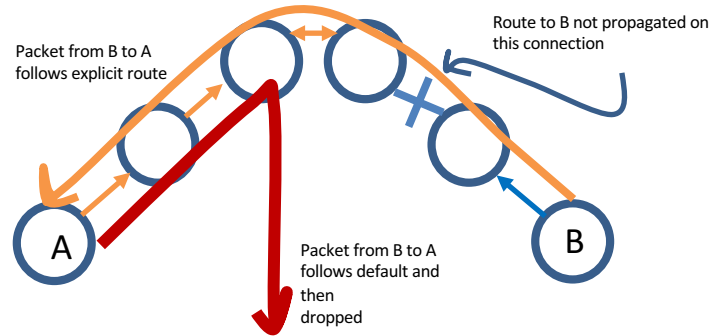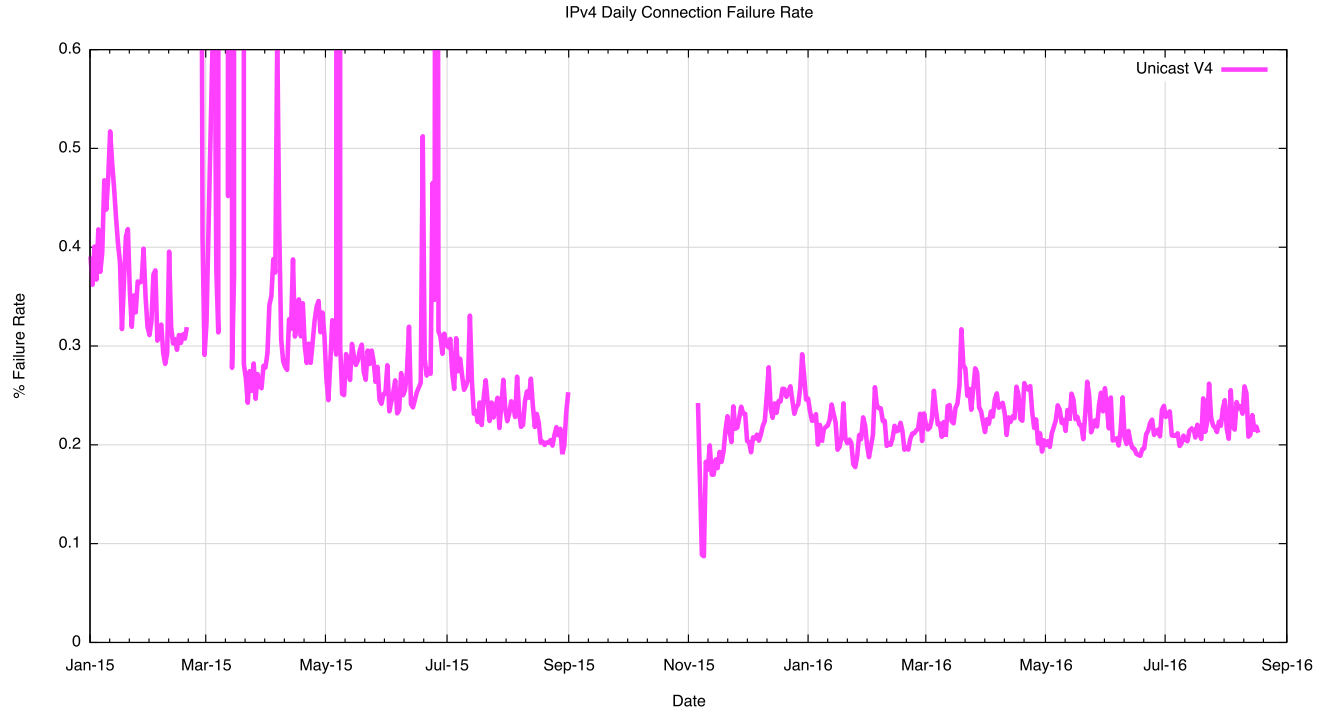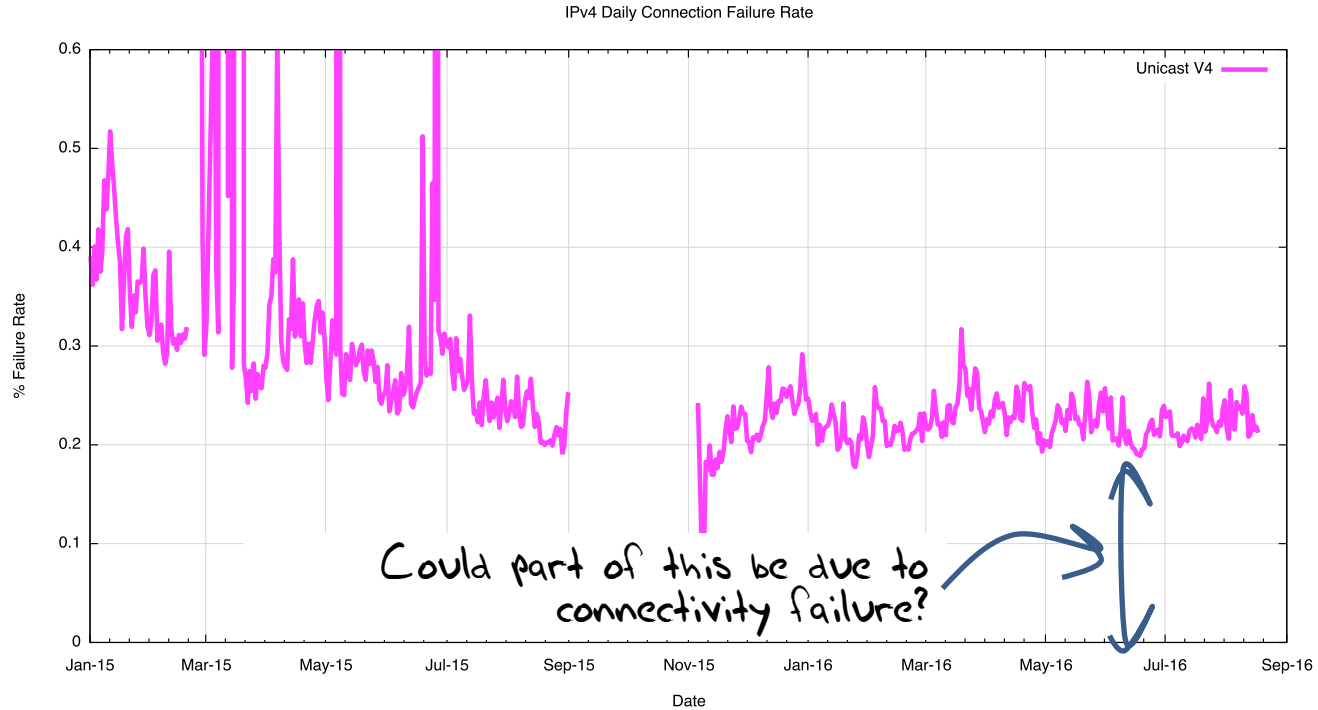
Packet from B to A follows explicit route

Route to B not propagated on this connection

Packet from B to A follows default and then dropped

A

B

To Recap: What is causing this?

IPv4 Daily Connection Failure Rate

To Recap: What is causing this?

IPv4 Daily Connection Failure Rate

Could part of this be due to connectivity failure?

Can we confirm this?

Outbound SYN

client

server

Can we confirm this?

client

Outbound SYN

server

SYN ACK
Return path

Can we confirm this?

client

Outbound SYN

router

server

Busted SYN ACK
Return path

Can we confirm this?

client

Outbound SYN

router

server

Busted SYN ACK
Return path

iCMP Dest Unreachable

Can we confirm this?

Outbound SYN

client

router

server

Busted SYN ACK
Return path

iCMP Dest Unreachable

What the server sees is an
incoming SYN, and then an
iCMP destination
unreachable
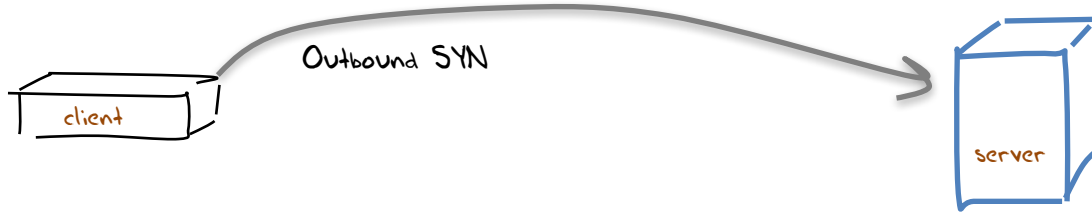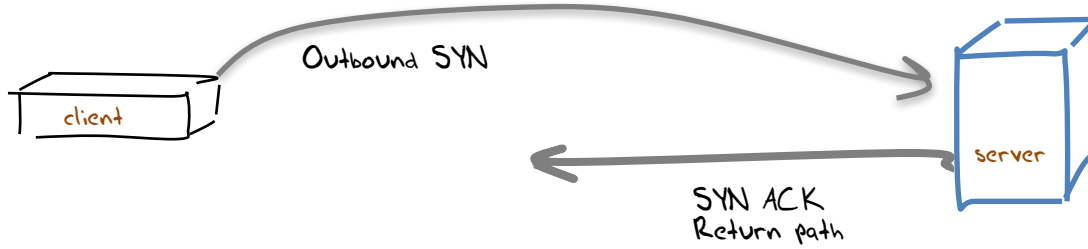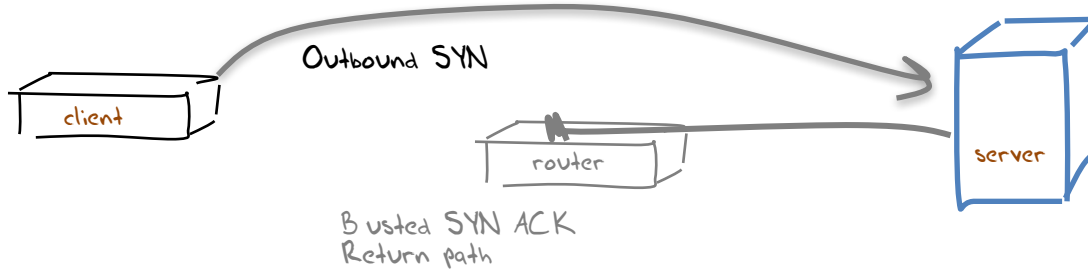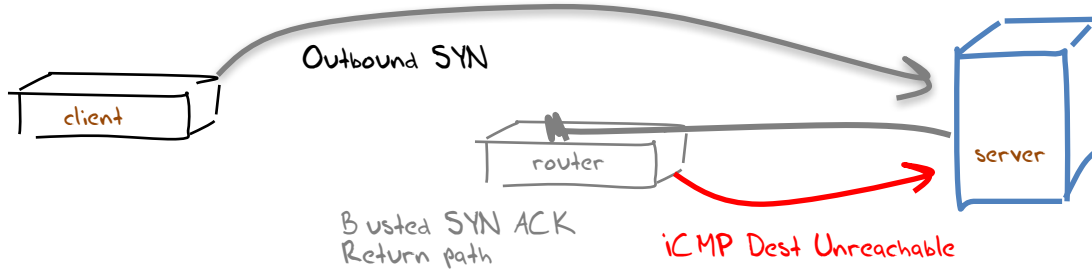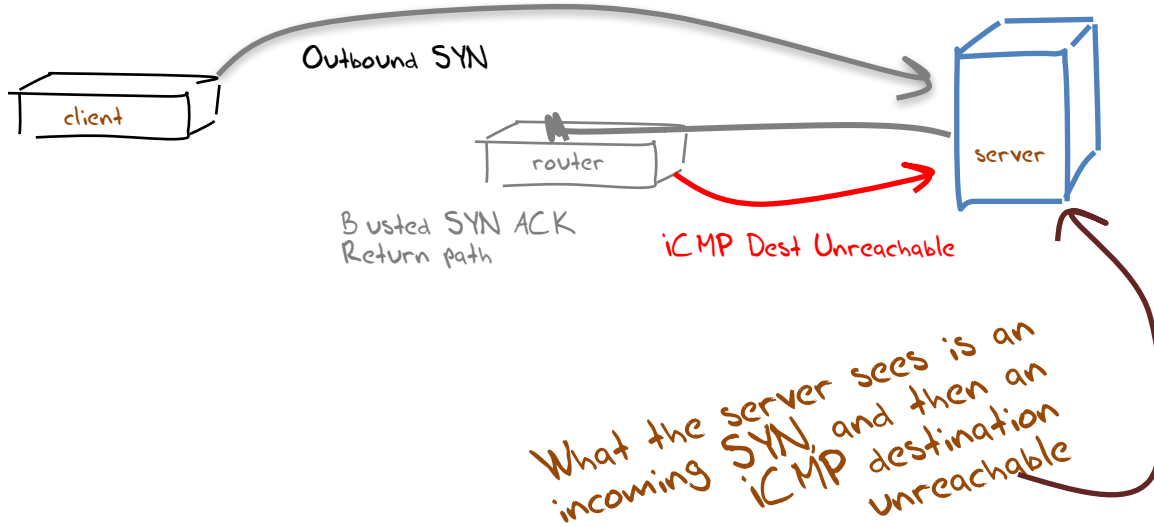
And we do see this …
here's an example

```
14:16:05.999497 IP (tos 0x0, ttl 55, id 31005, offset 0, flags [none], proto ICMP (1), length 80)
    84.41.108.74 > 139.162.146.97: ICMP host 46.163.63.47 unreachable, length 60
```

Outer packet is an ICMP Packet with a "destination unreachable" code sent to the server from the router at address 84.41.108.74

And we do see this …
here's an example

14:16:05.999497 IP (tos 0x0, ttl 55, id 31005, offset 0, flags [none], proto ICMP (1), length 80)
    84.41.108.74 > 139.162.146.97: ICMP host 46.163.63.4: unreachable, length 60

Outer packet is an iCMP Packet with a "destination unreachable" code
sent to the server from the router at address 84.41.108.74

    IP (tos 0x0, ttl 57, id 0, offset 0, flags [DF], proto TCP (6), length 52)
    139.162.146.97.443 > 46.163.63.xx.52087: Flags [S.] cksum 0x5130 (correct), seq 3917125220, ack
685287936, win 29200, options [mss 1460,nop,nop,sackOK,nop,wscale 7], length 0

Payload packet is a SYN+ACK packet

In the ~~Telephone~~ Internet Network

- All connected endpoints are equally reachable
- Anyone can reach anyone else

Almost

almost

most of the time!

In the ~~Telephone~~ internet Network

- All connected endpoints are equally reachable
- Anyone can reach anyone else

Almost

almost

most of the time!

As long as all the feeder access networks can connect to Facebook, Google, Netflix, Ebay, Amazon,... then nobody seems to care enough any more to be motivated to fix this

What we now have in 2016 is a tier 1 CDN feeder system. The internet is no longer a true ubiquitous fully connected peer network, and our collective care factor about that is pretty low

And that makes me sad!

Thanks!