APNIC 42

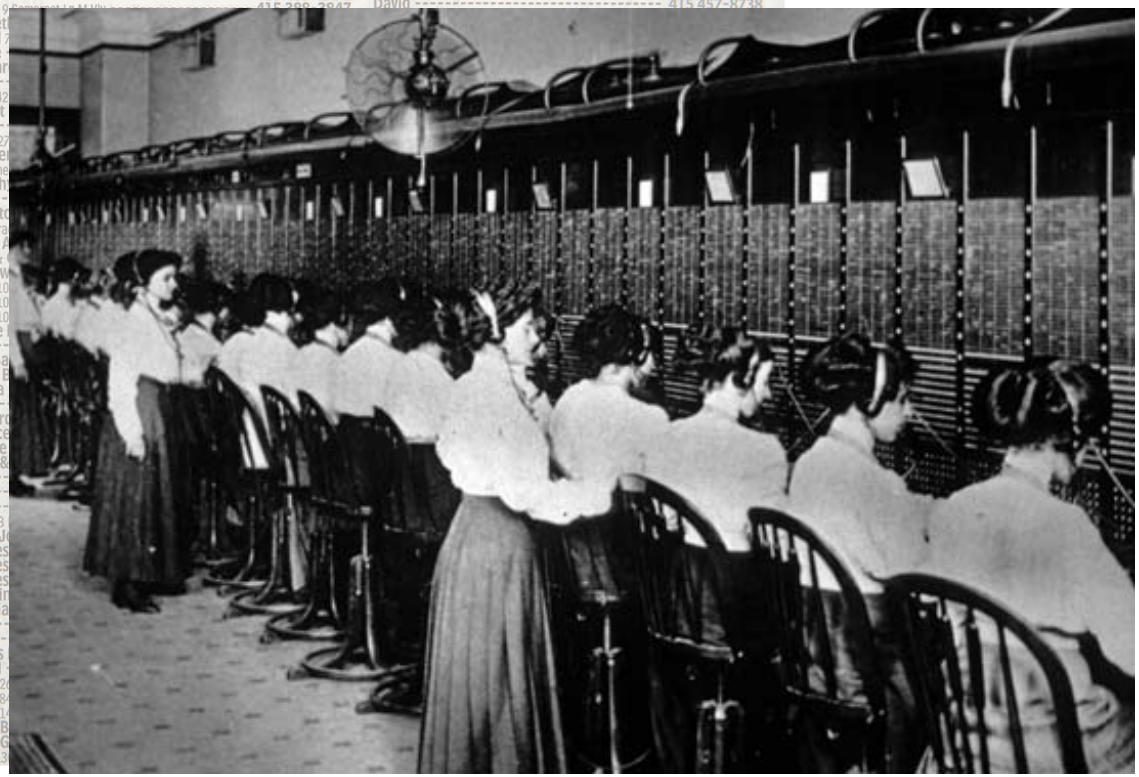
Desperately Seeking Default

Geoff Huston
APNIC

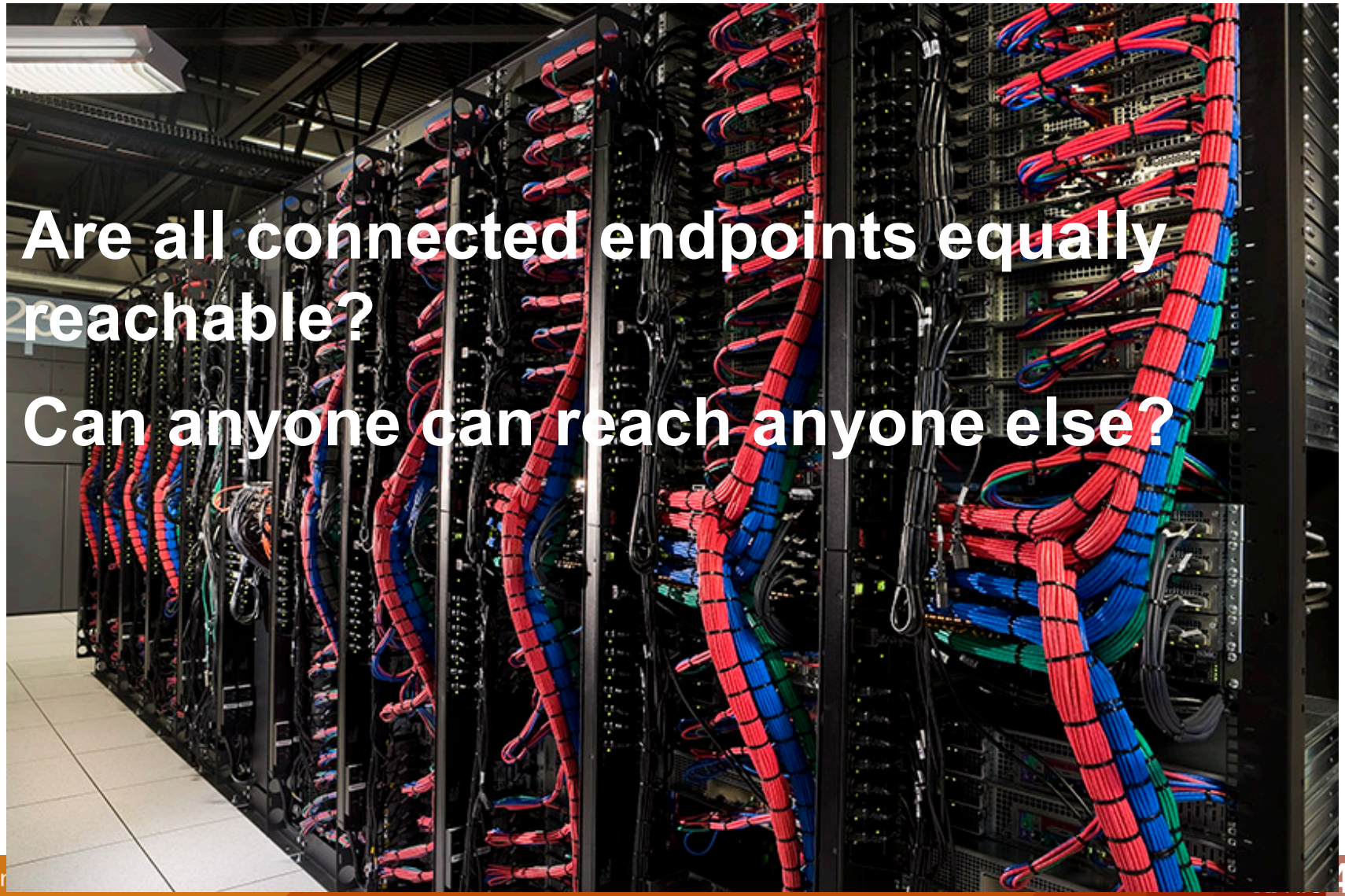
COLOMBO, SRI LANKA
28 September - 5 October 2016

In the Telephone Network

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



In the ~~Telephone Network~~ internet



In the ~~Telephone~~ internet ~~Network~~

- Are all connected endpoints are equally reachable? No!
- Can anyone can reach anyone else?

NATs and Firewall Filters changed all that – so our current expectations are that if you stand up a public service on port 80 (or 443 for that matter) 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

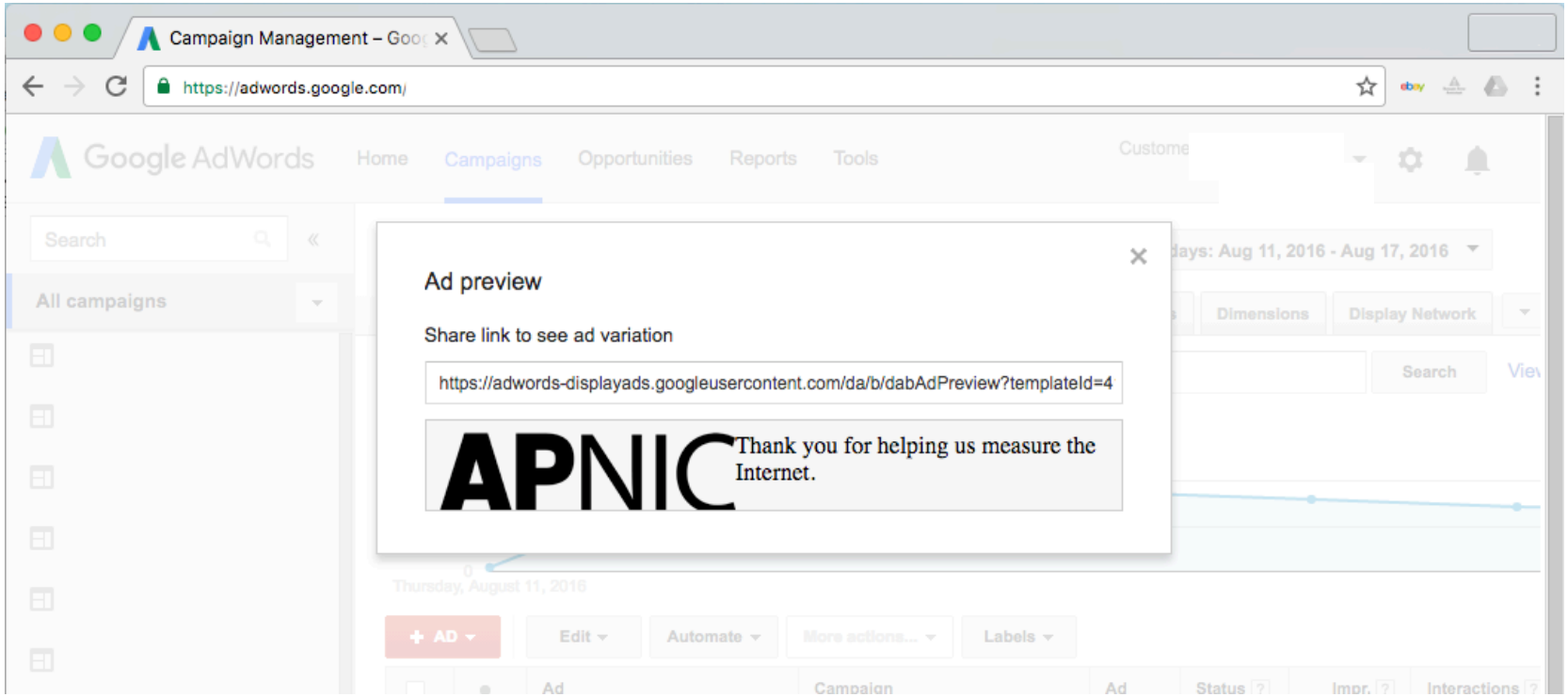
- ~~All connected endpoints are equally reachable~~
- Anyone can reach anyone else

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

What do we see?

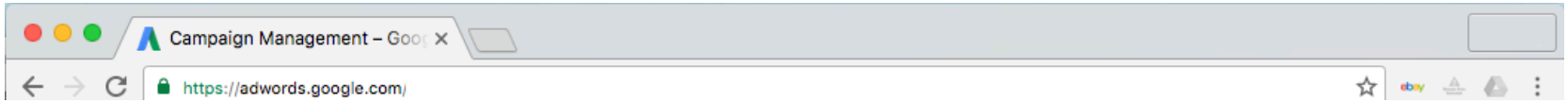
On the internet is everyone really connected to everyone else?

How We See

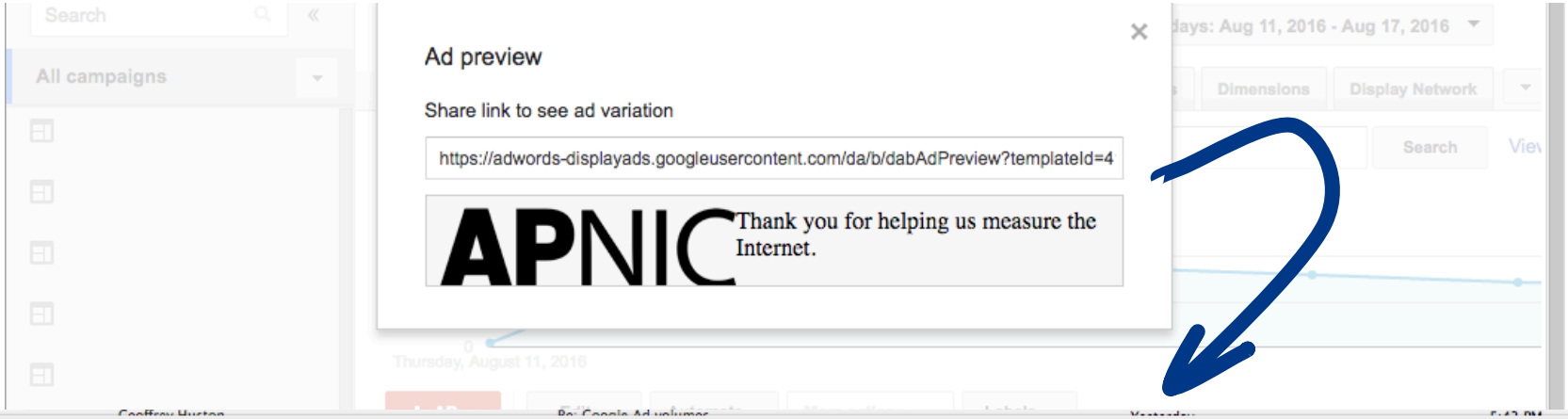


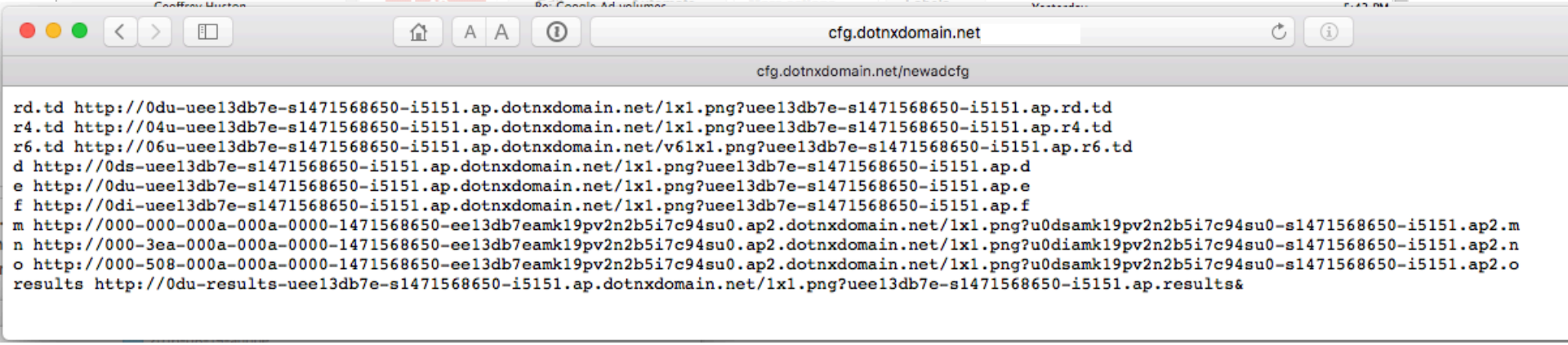
We use an online ad to present a sequence of small fetches to the user's browser

How We See



The sequence of tests is used to test a number of types of actions including fetches of IPv
IPv6 and Dual stack

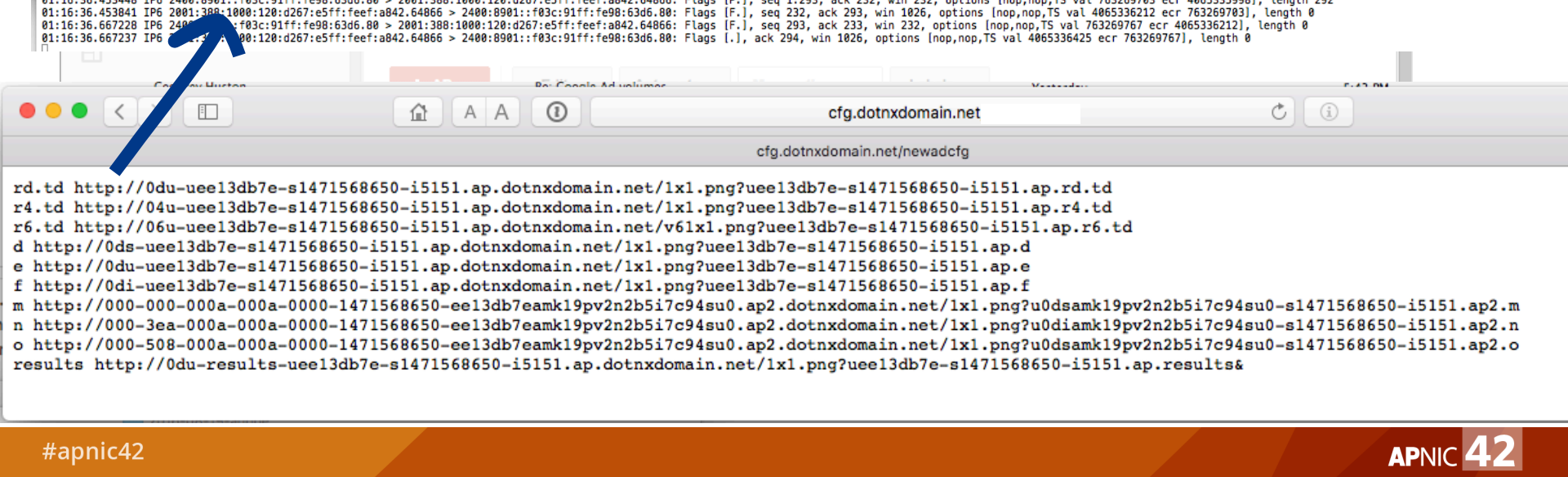




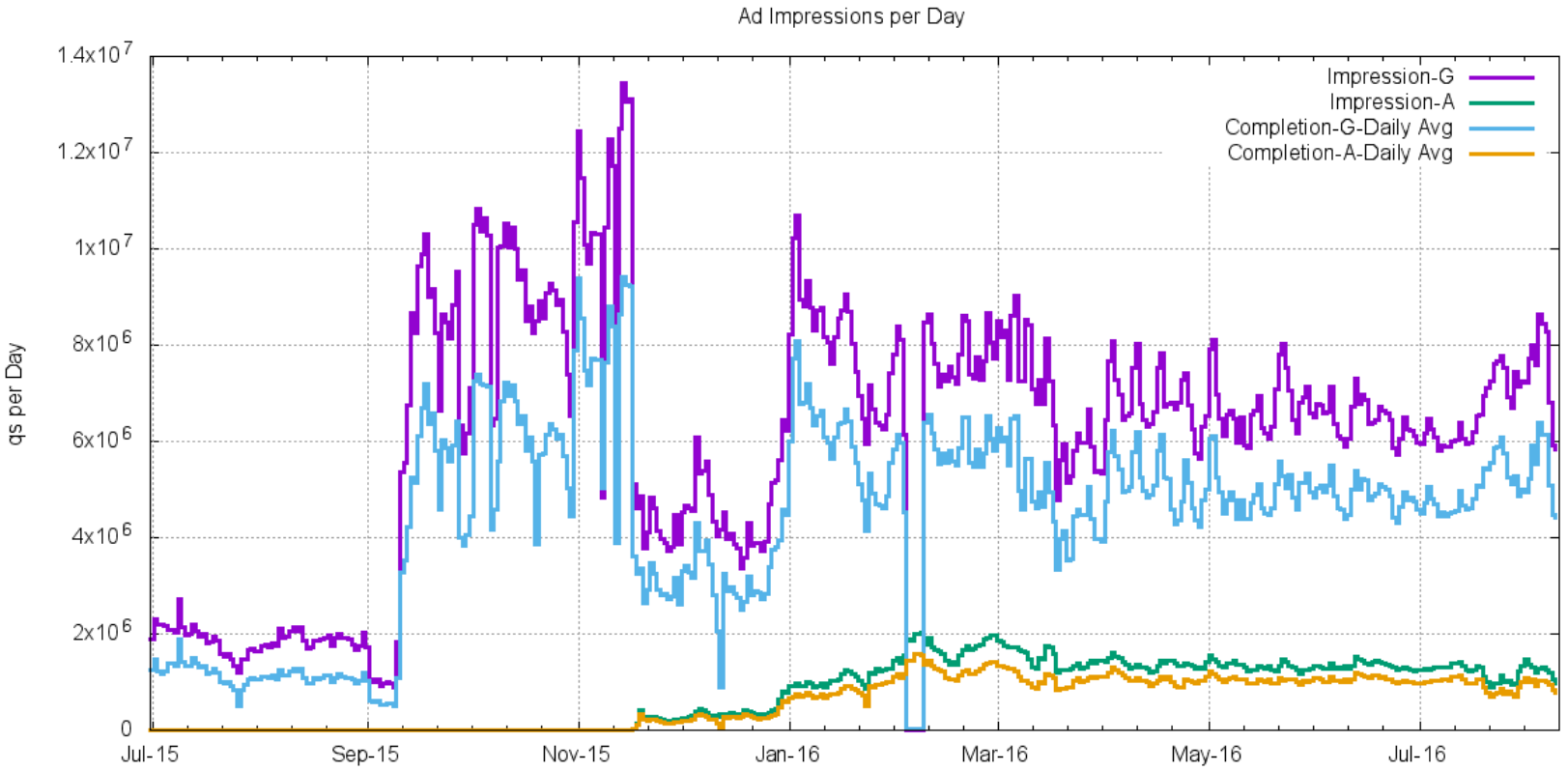
How We See

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

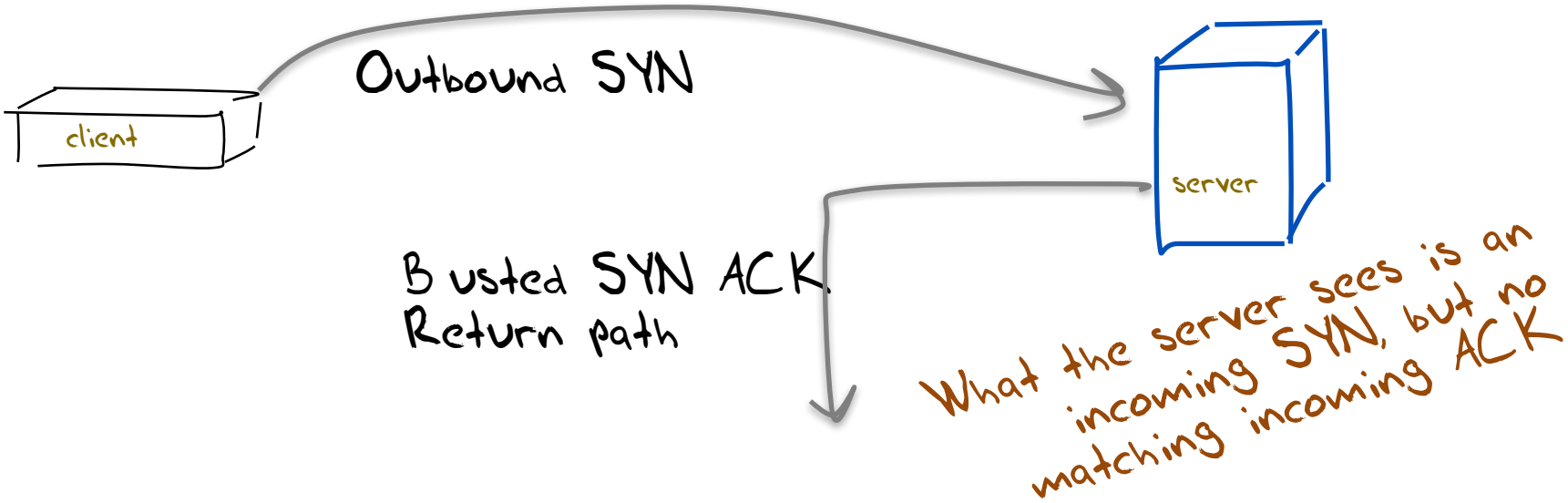
```
tcpdump: verbose output suppressed, use -v or -vv for full protocol decode
listening on eth0, link-type EN10MB (Ethernet), capture size 65535 bytes
01:15:56.160383 IP6 2001:388:1000:120:d267:e5ff:feef:a842.40836 > 2400:8901::f03c:91ff:fe98:63d6.80: Flags [S], seq 2414246339, win 65535, options [mss 1440,nop,wscale 6,sackOK,TS val 4065295918 ecr 0], length 0
01:15:56.373466 IP6 2400:8901::f03c:91ff:fe98:63d6.80 > 2001:388:1000:120:d267:e5ff:feef:a842.40836: Flags [S.], seq 2653147012, ack 2414246340, win 28560, options [mss 1440,sackOK,TS val 763257679 ecr 4065295918,nop,wscale 7], length 0
01:15:56.373486 IP6 2001:388:1000:120:d267:e5ff:feef:a842.40836 > 2400:8901::f03c:91ff:fe98:63d6.80: Flags [.], ack 1, win 1026, options [nop,nop,TS val 4065296132 ecr 763257679], length 0
01:15:56.373502 IP6 2001:388:1000:120:d267:e5ff:feef:a842.40836 > 2400:8901::f03c:91ff:fe98:63d6.80: Flags [P.], seq 1:222, ack 1, win 1026, options [nop,nop,TS val 4065296132 ecr 763257679], length 221
01:15:56.586494 IP6 2400:8901::f03c:91ff:fe98:63d6.80 > 2001:388:1000:120:d267:e5ff:feef:a842.40836: Flags [.], ack 222, win 232, options [nop,nop,TS val 763257743 ecr 4065296132], length 0
01:15:56.586644 IP6 2400:8901::f03c:91ff:fe98:63d6.80 > 2001:388:1000:120:d267:e5ff:feef:a842.40836: Flags [P.], seq 1:293, ack 222, win 232, options [nop,nop,TS val 763257743 ecr 4065296132], length 292
01:15:56.592169 IP6 2001:388:1000:120:d267:e5ff:feef:a842.40836 > 2400:8901::f03c:91ff:fe98:63d6.80: Flags [F.], seq 222, ack 293, win 1026, options [nop,nop,TS val 4065296350 ecr 763257743], length 0
01:15:56.805026 IP6 2400:8901::f03c:91ff:fe98:63d6.80 > 2001:388:1000:120:d267:e5ff:feef:a842.40836: Flags [F.], seq 293, ack 223, win 232, options [nop,nop,TS val 763257809 ecr 4065296350], length 0
01:15:56.805035 IP6 2001:388:1000:120:d267:e5ff:feef:a842.40836 > 2400:8901::f03c:91ff:fe98:63d6.80: Flags [.], ack 294, win 1026, options [nop,nop,TS val 4065296563 ecr 763257809], length 0
01:16:08.160813 IP 202.158.221.222.62158 > 139.162.2.194.80: Flags [S], seq 2110718859, win 65535, options [mss 1460,nop,wscale 6,sackOK,TS val 4065307919 ecr 0], length 0
01:16:08.484006 IP 139.162.2.194.80 > 202.158.221.222.62158: Flags [S.], seq 3450703102, ack 2110718860, win 28960, options [mss 1460,sackOK,TS val 763261314 ecr 4065307919,nop,wscale 7], length 0
01:16:08.484018 IP 202.158.221.222.62158 > 139.162.2.194.80: Flags [.], ack 1, win 1040, options [nop,nop,TS val 4065308242 ecr 763261314], length 0
01:16:08.484040 IP 202.158.221.222.62158 > 139.162.2.194.80: Flags [P.], seq 1:222, ack 1, win 1040, options [nop,nop,TS val 4065308242 ecr 763261314], length 221
01:16:08.807080 IP 139.162.2.194.80 > 202.158.221.222.62158: Flags [.], ack 222, win 235, options [nop,nop,TS val 763261411 ecr 4065308242], length 0
01:16:08.807231 IP 139.162.2.194.80 > 202.158.221.222.62158: Flags [P.], seq 1:293, ack 222, win 235, options [nop,nop,TS val 763261411 ecr 4065308242], length 292
01:16:08.807654 IP 202.158.221.222.62158 > 139.162.2.194.80: Flags [F.], seq 222, ack 293, win 1040, options [nop,nop,TS val 4065308566 ecr 763261411], length 0
01:16:09.130793 IP 139.162.2.194.80 > 202.158.221.222.62158: Flags [F.], seq 293, ack 223, win 235, options [nop,nop,TS val 763261508 ecr 4065308566], length 0
01:16:09.130801 IP 202.158.221.222.62158 > 139.162.2.194.80: Flags [.], ack 294, win 1040, options [nop,nop,TS val 4065308889 ecr 763261508], length 0
01:16:20.244088 IP6 2001:388:1000:120:d267:e5ff:feef:a842.31238 > 2400:8901::f03c:91ff:fe98:63d6.80: Flags [S], seq 1299030901, win 65535, options [mss 1440,nop,wscale 6,sackOK,TS val 4065320002 ecr 0], length 0
01:16:20.457379 IP6 2400:8901::f03c:91ff:fe98:63d6.80 > 2001:388:1000:120:d267:e5ff:feef:a842.31238: Flags [S.], seq 1748054555, ack 1299030902, win 28560, options [mss 1440,sackOK,TS val 763264905 ecr 4065320002,nop,wscale 7], length 0
01:16:20.457397 IP6 2001:388:1000:120:d267:e5ff:feef:a842.31238 > 2400:8901::f03c:91ff:fe98:63d6.80: Flags [.], ack 1, win 1026, options [nop,nop,TS val 4065320215 ecr 763264905], length 0
01:16:20.457413 IP6 2001:388:1000:120:d267:e5ff:feef:a842.31238 > 2400:8901::f03c:91ff:fe98:63d6.80: Flags [P.], seq 1:224, ack 1, win 1026, options [nop,nop,TS val 4065320215 ecr 763264905], length 223
01:16:20.670828 IP6 2400:8901::f03c:91ff:fe98:63d6.80 > 2001:388:1000:120:d267:e5ff:feef:a842.31238: Flags [.], ack 224, win 232, options [nop,nop,TS val 763264969 ecr 4065320215], length 0
01:16:20.670979 IP6 2400:8901::f03c:91ff:fe98:63d6.80 > 2001:388:1000:120:d267:e5ff:feef:a842.31238: Flags [P.], seq 1:293, ack 224, win 232, options [nop,nop,TS val 763264969 ecr 4065320215], length 292
01:16:20.671386 IP6 2001:388:1000:120:d267:e5ff:feef:a842.31238 > 2400:8901::f03c:91ff:fe98:63d6.80: Flags [F.], seq 224, ack 293, win 1026, options [nop,nop,TS val 4065320429 ecr 763264969], length 0
01:16:20.884786 IP6 2400:8901::f03c:91ff:fe98:63d6.80 > 2001:388:1000:120:d267:e5ff:feef:a842.31238: Flags [F.], seq 293, ack 225, win 232, options [nop,nop,TS val 763265033 ecr 4065320429], length 0
01:16:20.884796 IP6 2001:388:1000:120:d267:e5ff:feef:a842.31238 > 2400:8901::f03c:91ff:fe98:63d6.80: Flags [.], ack 294, win 1026, options [nop,nop,TS val 4065320643 ecr 763265033], length 0
01:16:36.025996 IP6 2001:388:1000:120:d267:e5ff:feef:a842.64866 > 2400:8901::f03c:91ff:fe98:63d6.80: Flags [S], seq 1648543162, win 65535, options [mss 1440,nop,wscale 6,sackOK,TS val 4065335784 ecr 0], length 0
01:16:36.239388 IP6 2400:8901::f03c:91ff:fe98:63d6.80 > 2001:388:1000:120:d267:e5ff:feef:a842.64866: Flags [S.], seq 2178829105, ack 1648543163, win 28560, options [mss 1440,sackOK,TS val 763269639 ecr 4065335784,nop,wscale 7], length 0
01:16:36.239407 IP6 2001:388:1000:120:d267:e5ff:feef:a842.64866 > 2400:8901::f03c:91ff:fe98:63d6.80: Flags [.], ack 1, win 1026, options [nop,nop,TS val 4065335998 ecr 763269639], length 0
01:16:36.239422 IP6 2001:388:1000:120:d267:e5ff:feef:a842.64866 > 2400:8901::f03c:91ff:fe98:63d6.80: Flags [P.], seq 1:232, ack 1, win 1026, options [nop,nop,TS val 4065335998 ecr 763269639], length 231
01:16:36.453147 IP6 2400:8901::f03c:91ff:fe98:63d6.80 > 2001:388:1000:120:d267:e5ff:feef:a842.64866: Flags [.], ack 232, win 232, options [nop,nop,TS val 763269703 ecr 4065335998], length 0
01:16:36.453448 IP6 2400:8901::f03c:91ff:fe98:63d6.80 > 2001:388:1000:120:d267:e5ff:feef:a842.64866: Flags [P.], seq 1:293, ack 232, win 232, options [nop,nop,TS val 763269703 ecr 4065335998], length 292
01:16:36.453841 IP6 2001:388:1000:120:d267:e5ff:feef:a842.64866 > 2400:8901::f03c:91ff:fe98:63d6.80: Flags [F.], seq 232, ack 293, win 1026, options [nop,nop,TS val 4065336212 ecr 763269703], length 0
01:16:36.667228 IP6 2400:8901::f03c:91ff:fe98:63d6.80 > 2001:388:1000:120:d267:e5ff:feef:a842.64866: Flags [F.], seq 293, ack 233, win 232, options [nop,nop,TS val 763269767 ecr 4065336212], length 0
01:16:36.667237 IP6 2001:388:1000:120:d267:e5ff:feef:a842.64866 > 2400:8901::f03c:91ff:fe98:63d6.80: Flags [.], ack 294, win 1026, options [nop,nop,TS val 4065336425 ecr 763269767], length 0
```



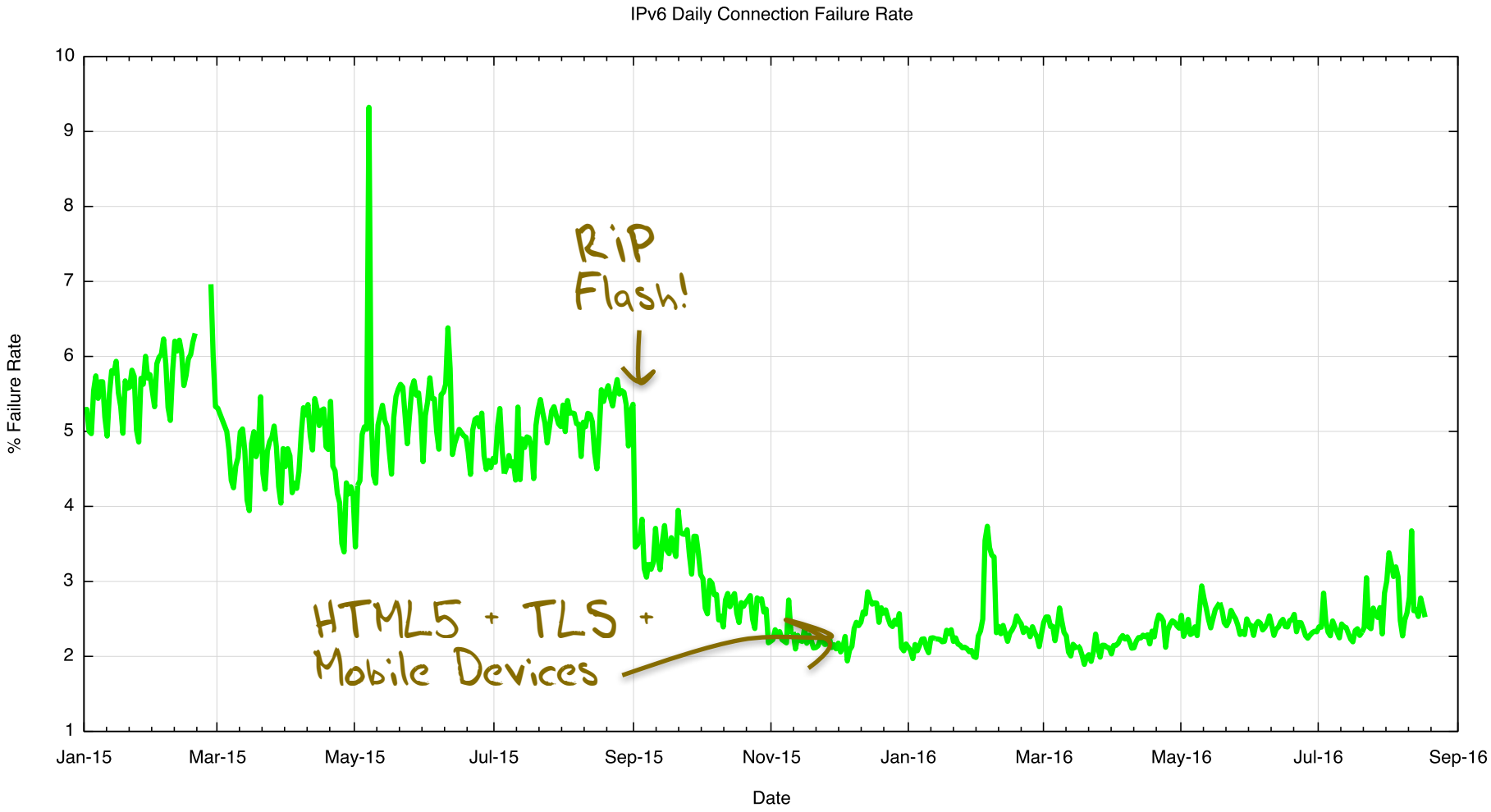
How Much do We See?



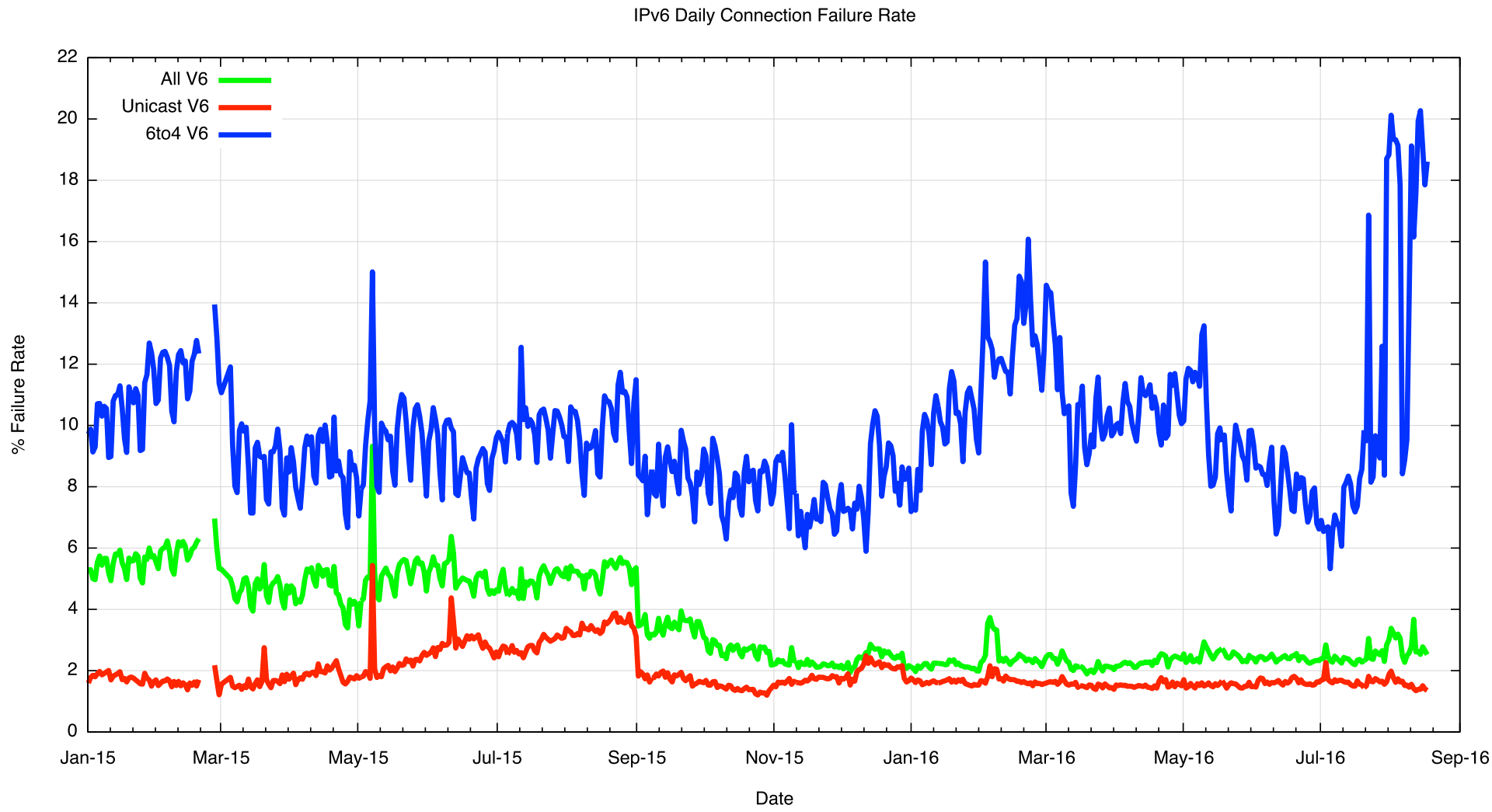
What we see: Connection Failure



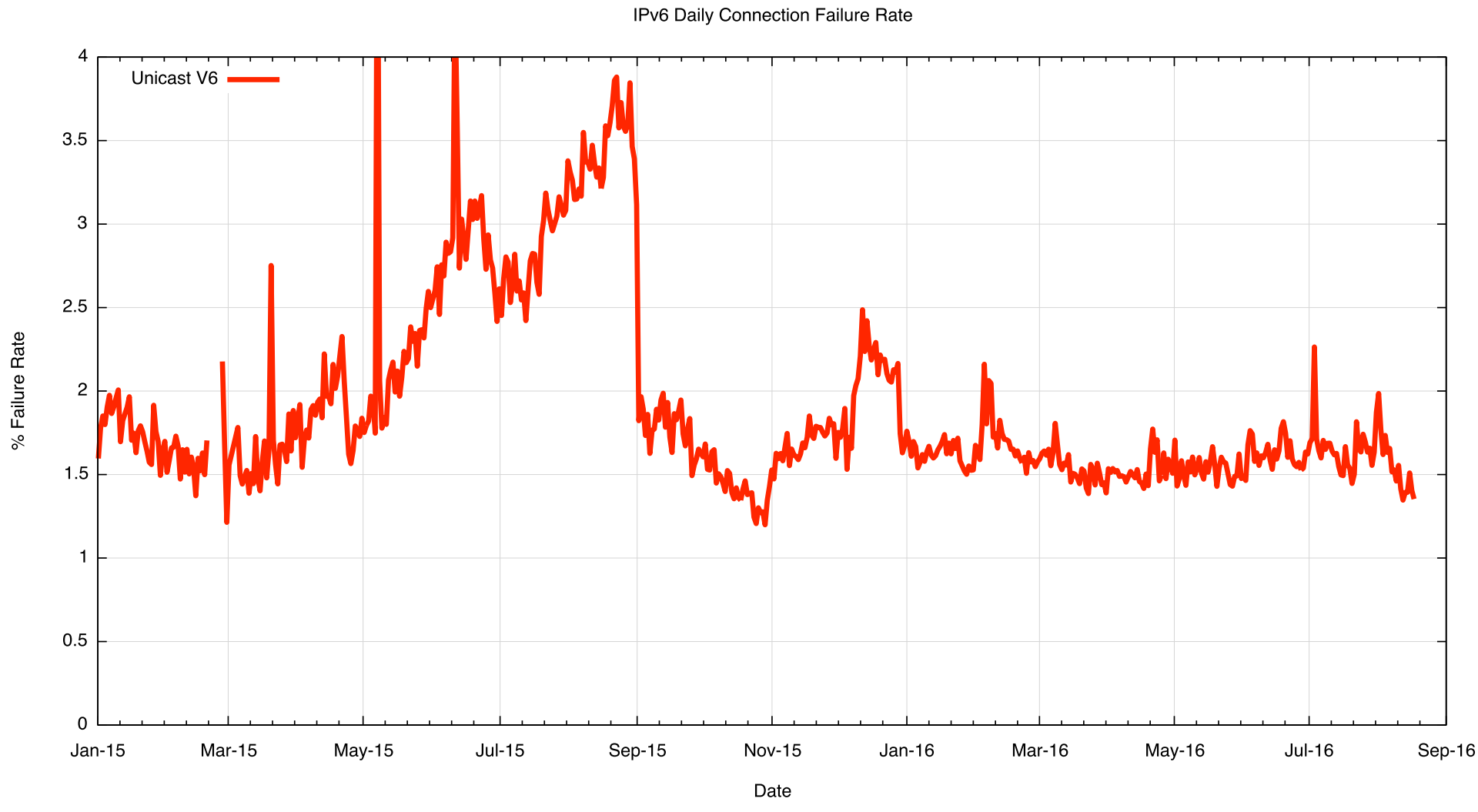
Daily IPv6 Failures



Daily IPv6 Failures



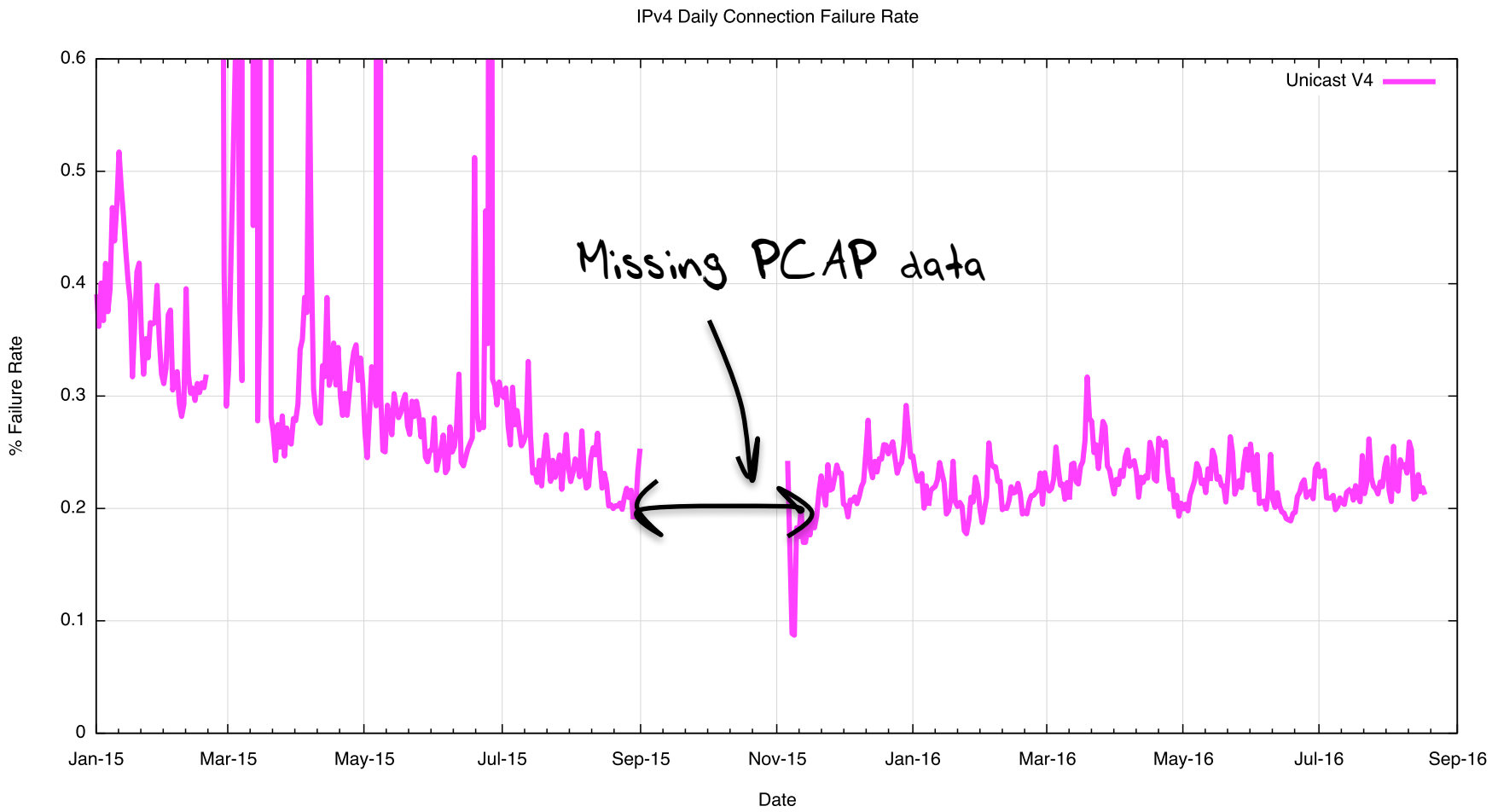
Daily IPv6 Failures



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



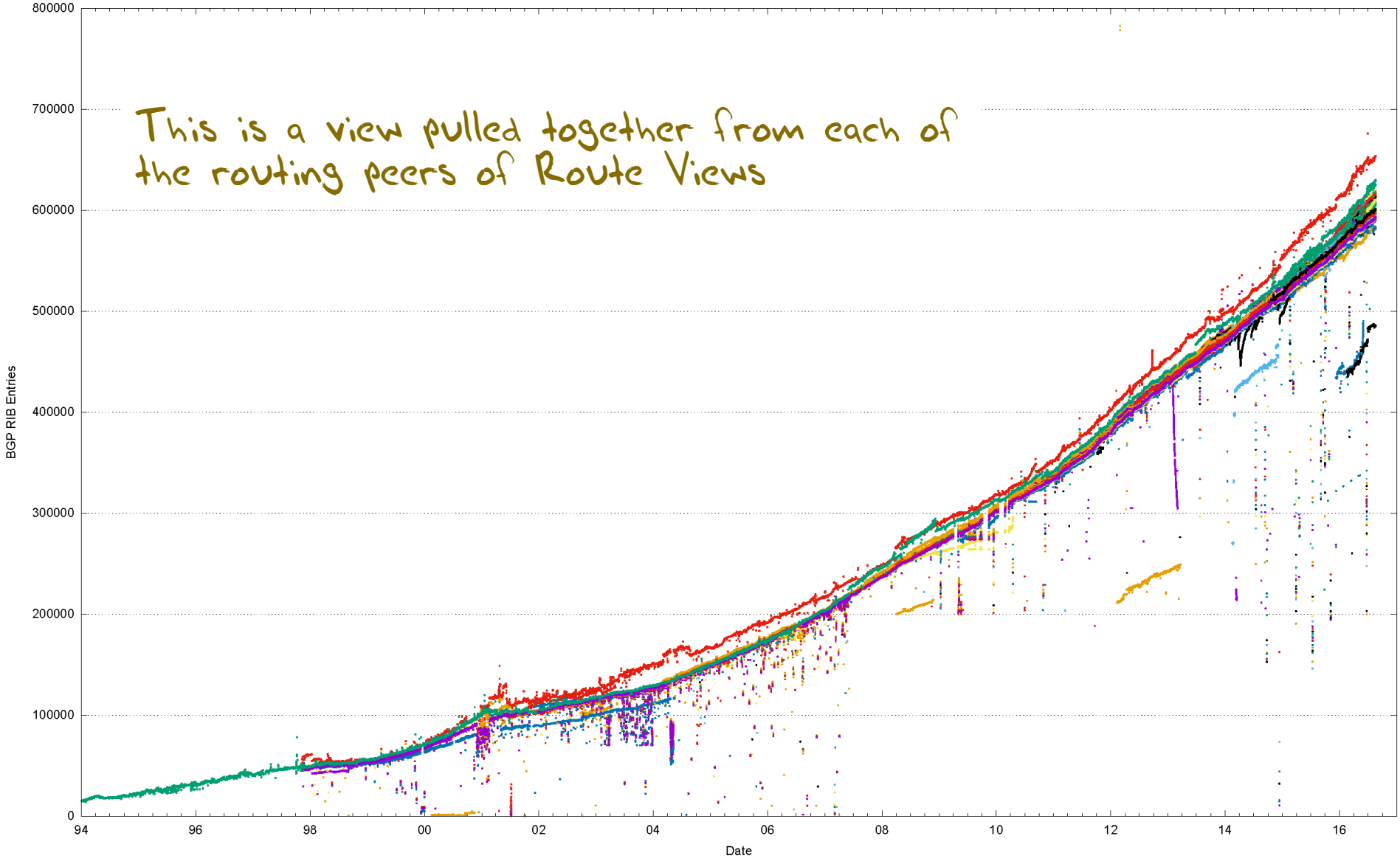
IPv4 Failures

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

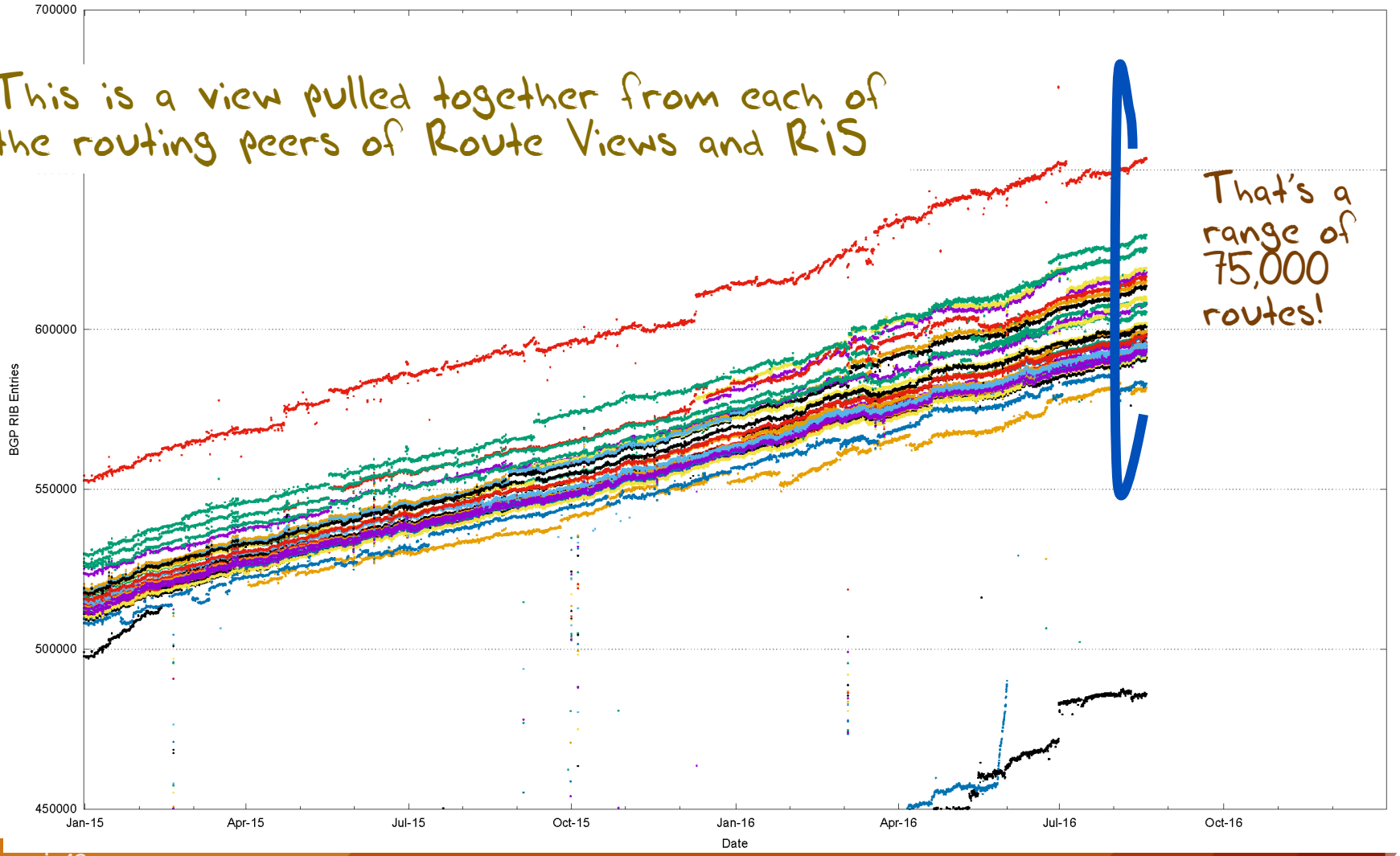
- So what is the reason for this residual asymmetric failure rate?
- Is it asymmetric routing connectivity?

Route Views Routing Table

25 Years of Routing the Internet



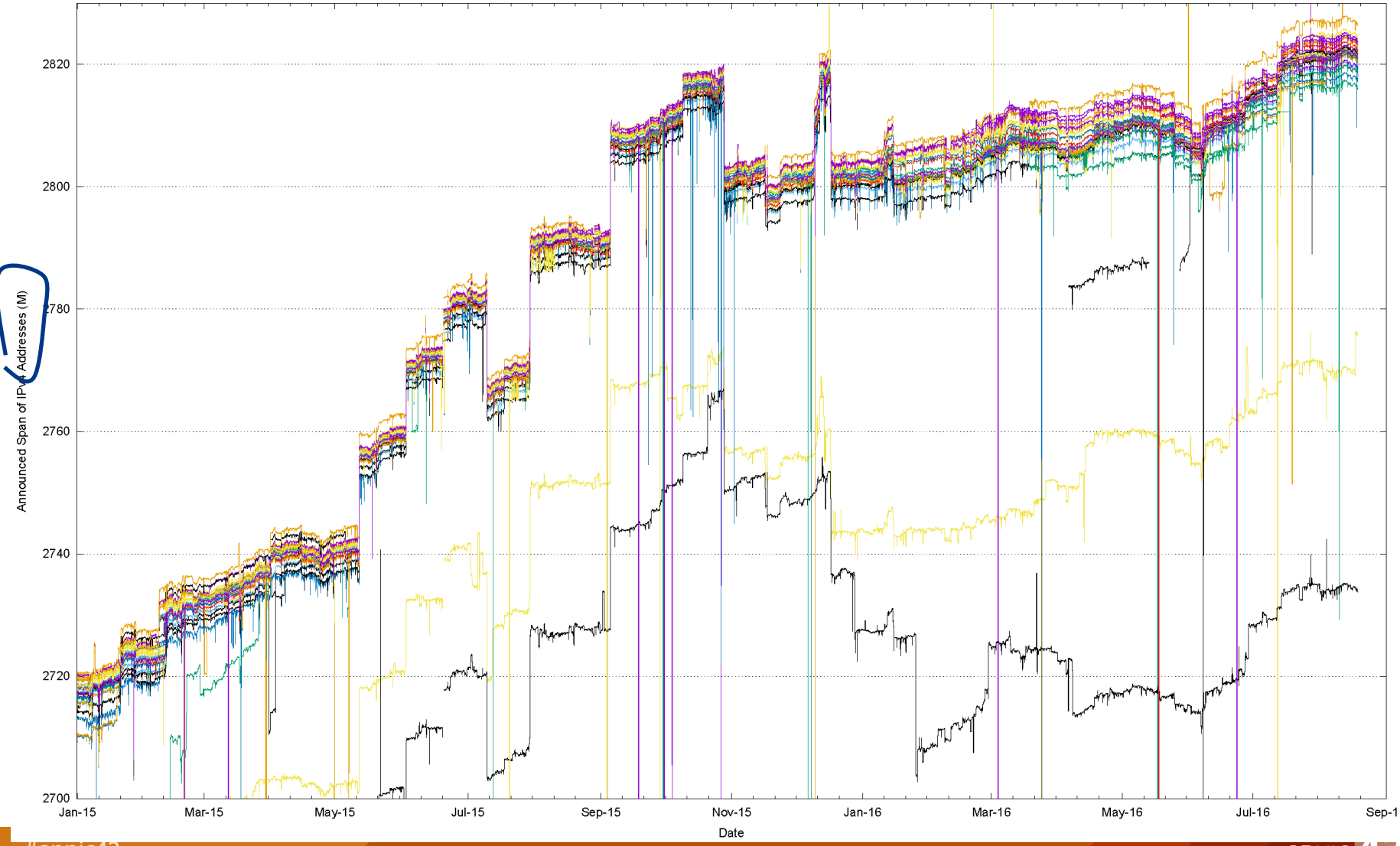
2015/16, Route Views + RIS



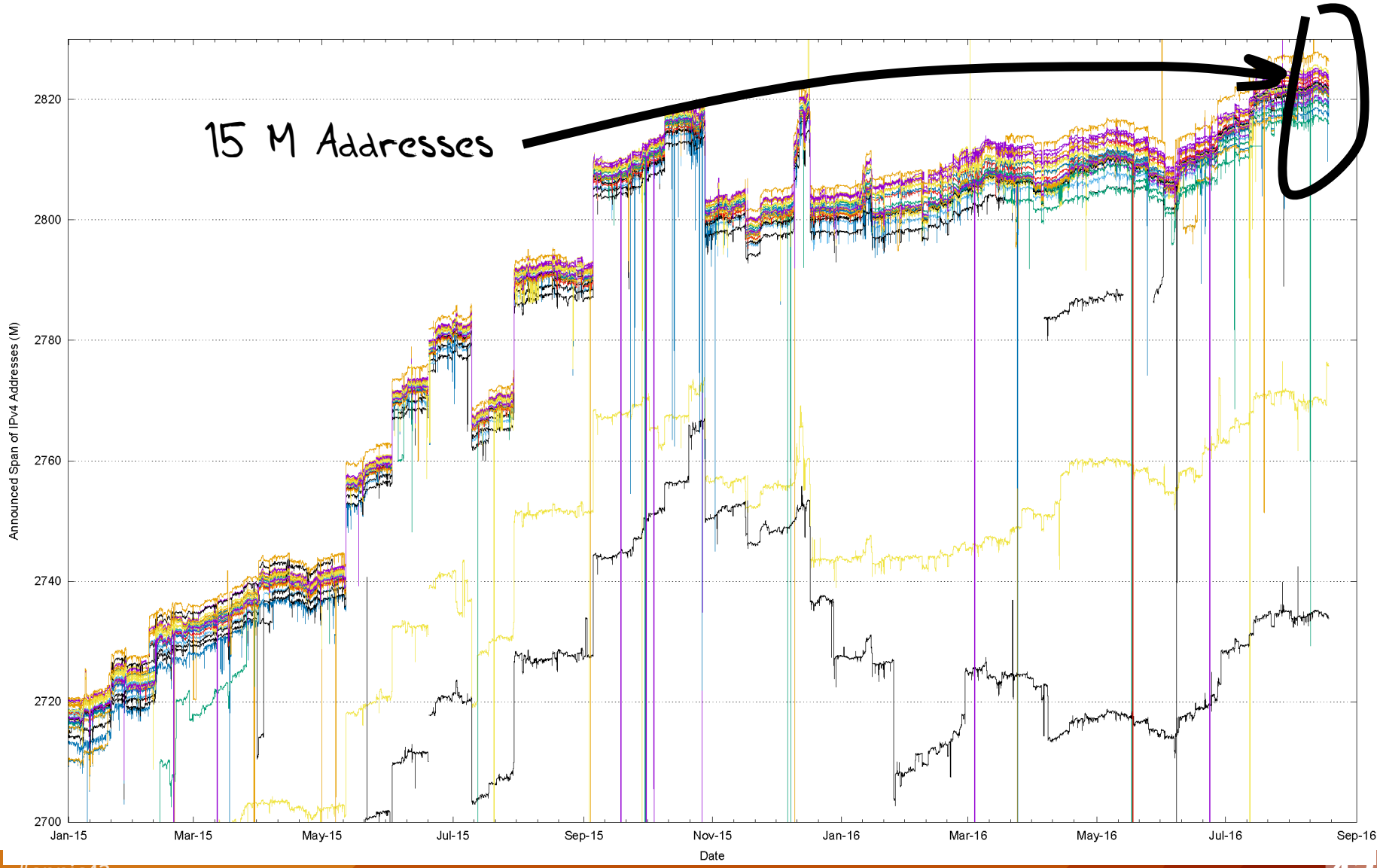
Different peers see a slightly different Internet

- 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)

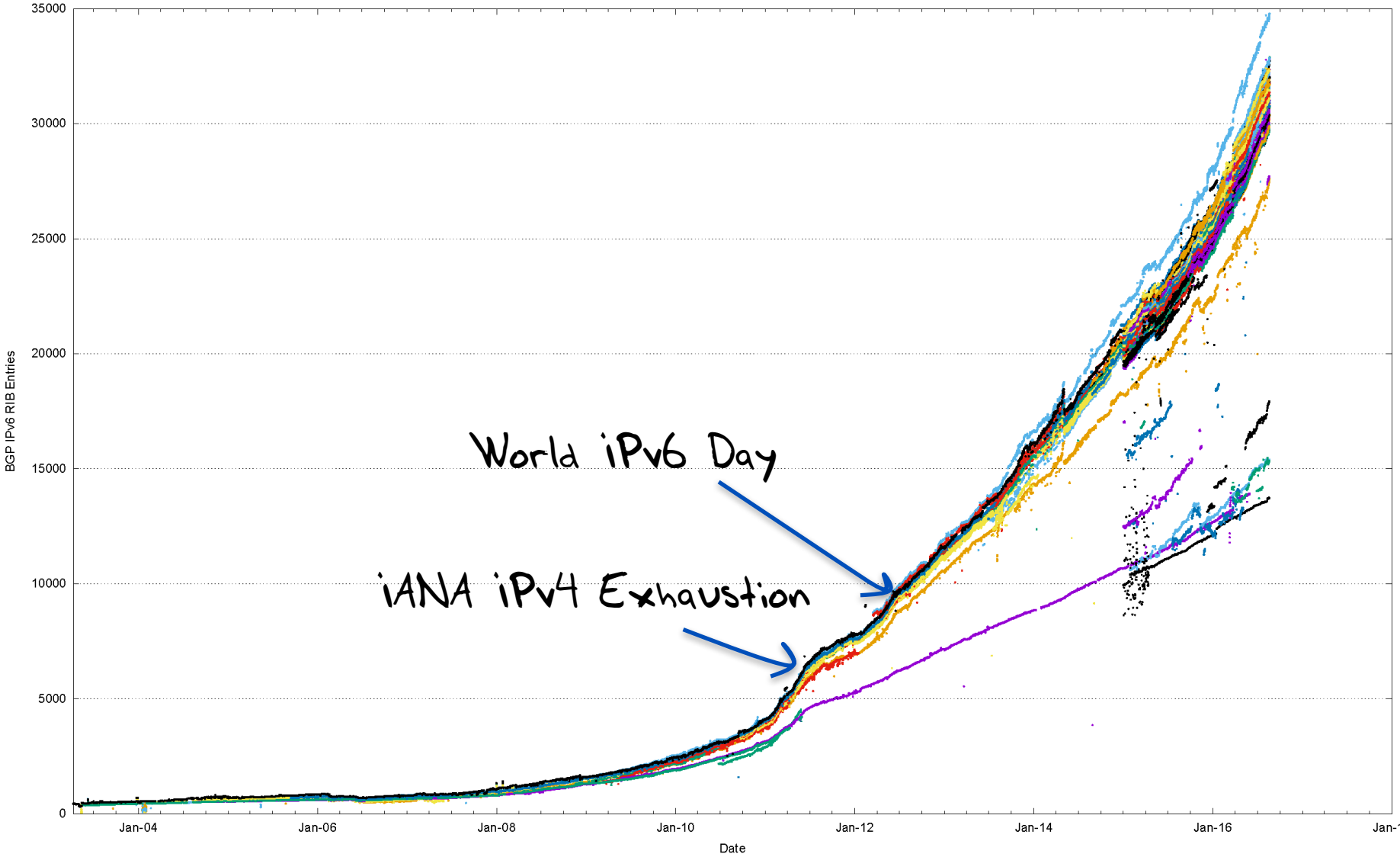


What does this mean?

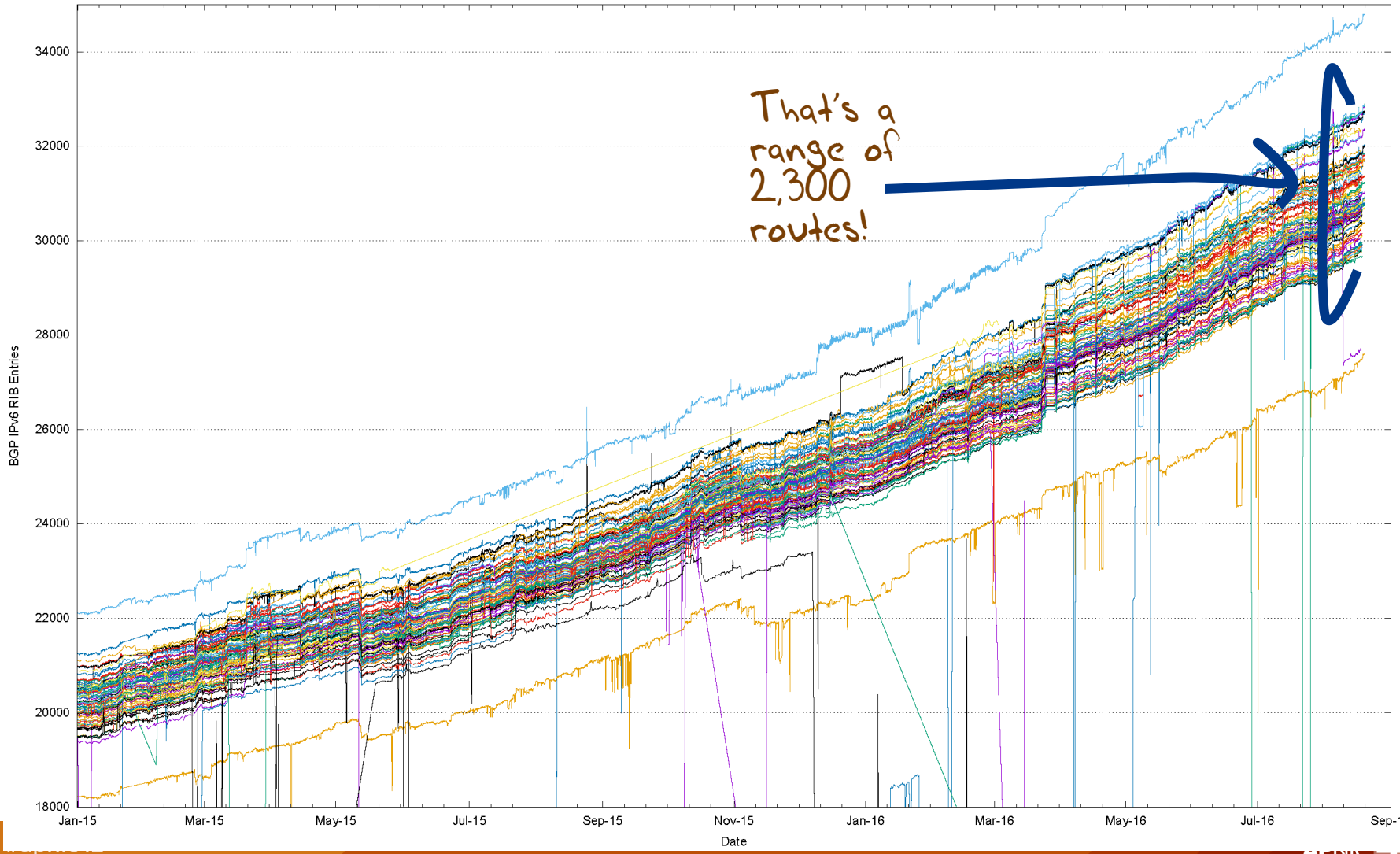
- Each peer of RouteViews and RIS announces a span of addresses that appears to be a unique span.
- In total, these spans agree with other to within ~20M addresses, but this means that there are potentially some 20M uniquely addressed endpoints that cannot be reached from all other endpoints.
- This 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?

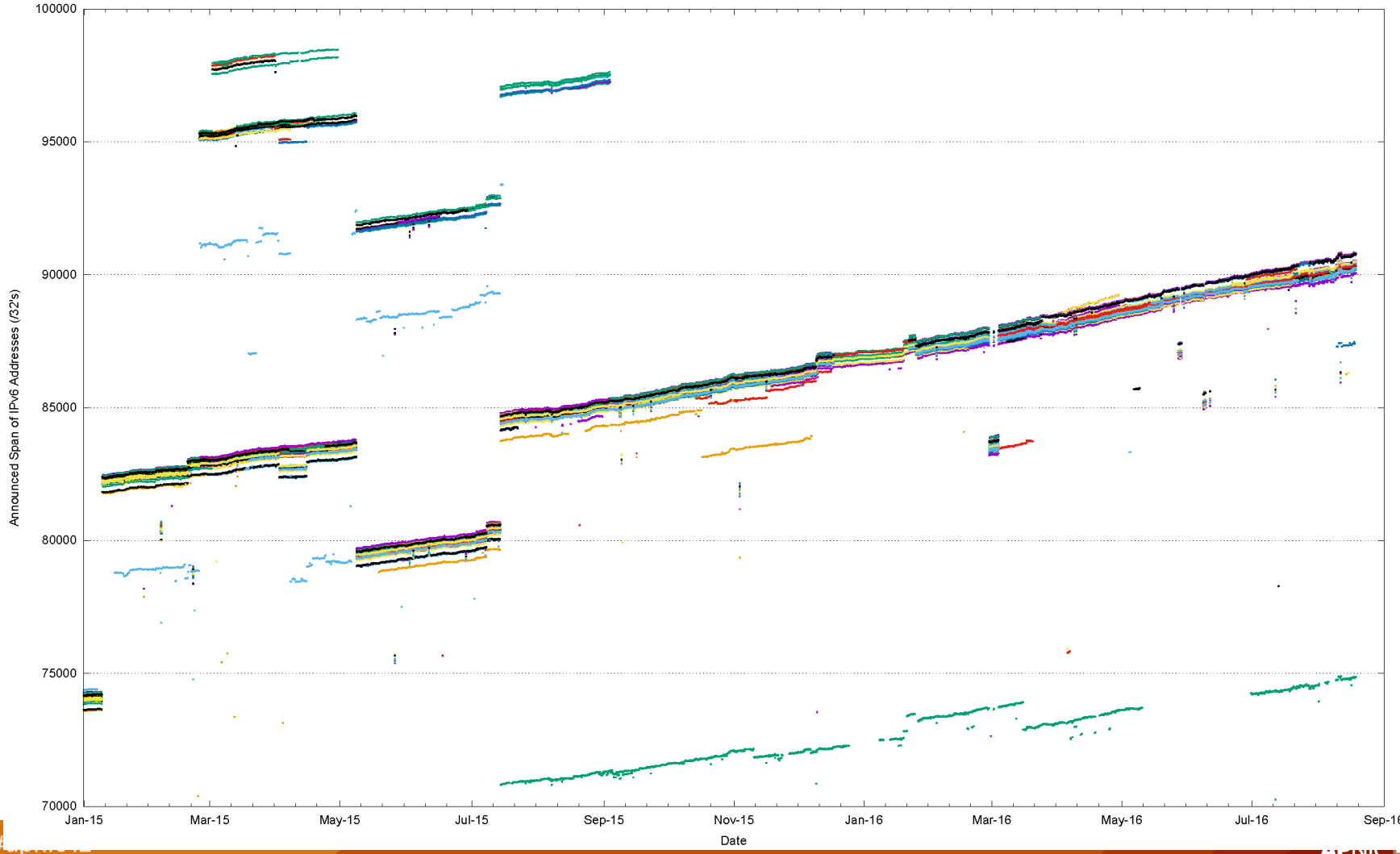
The Route Views view of IPv6



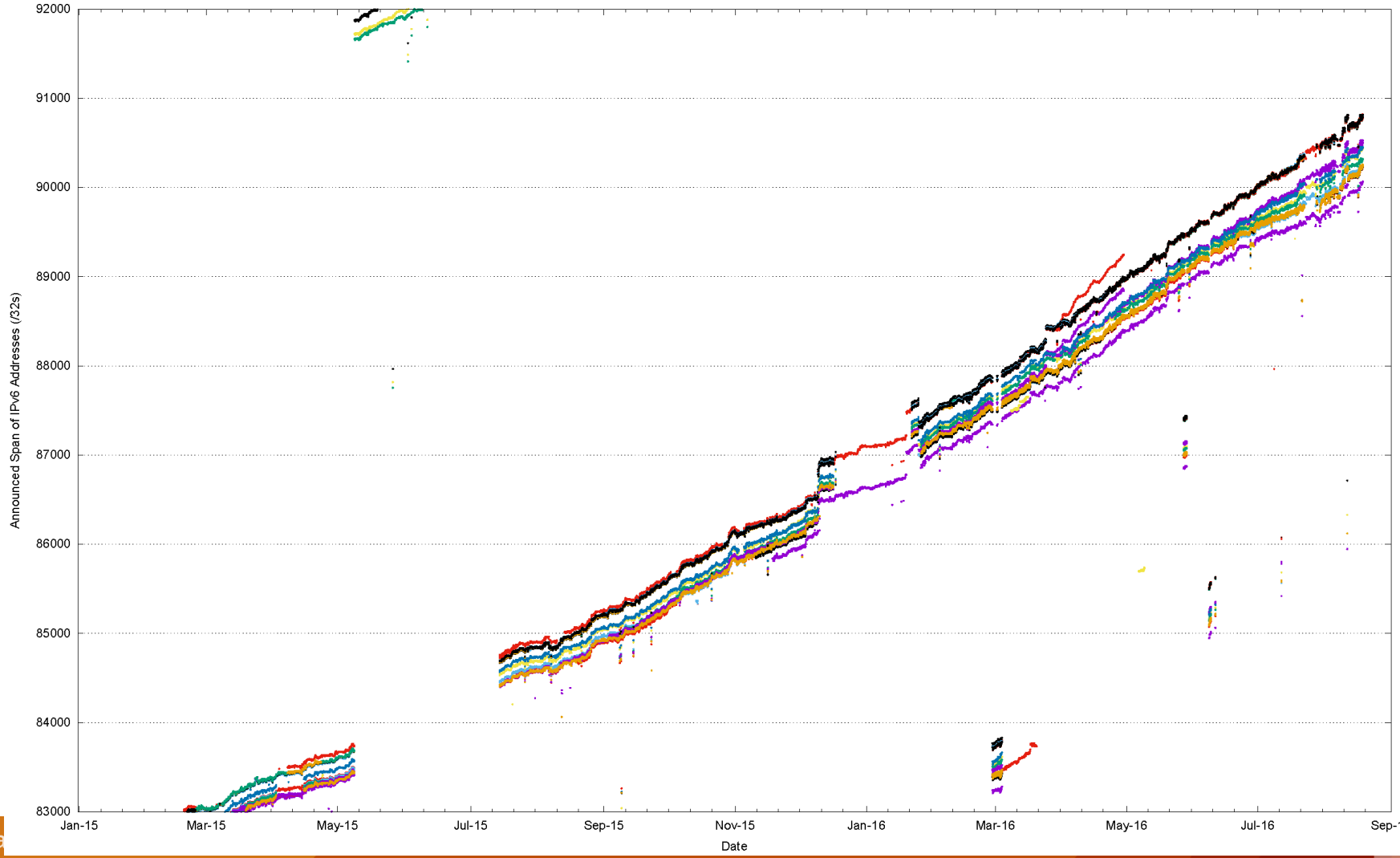
Number of IPv6 Routes in 2015/16



IPv6 Announced Address Span Variation (RV + RIS)



IPv6 Announced Address Span Variation (RV + RIS)



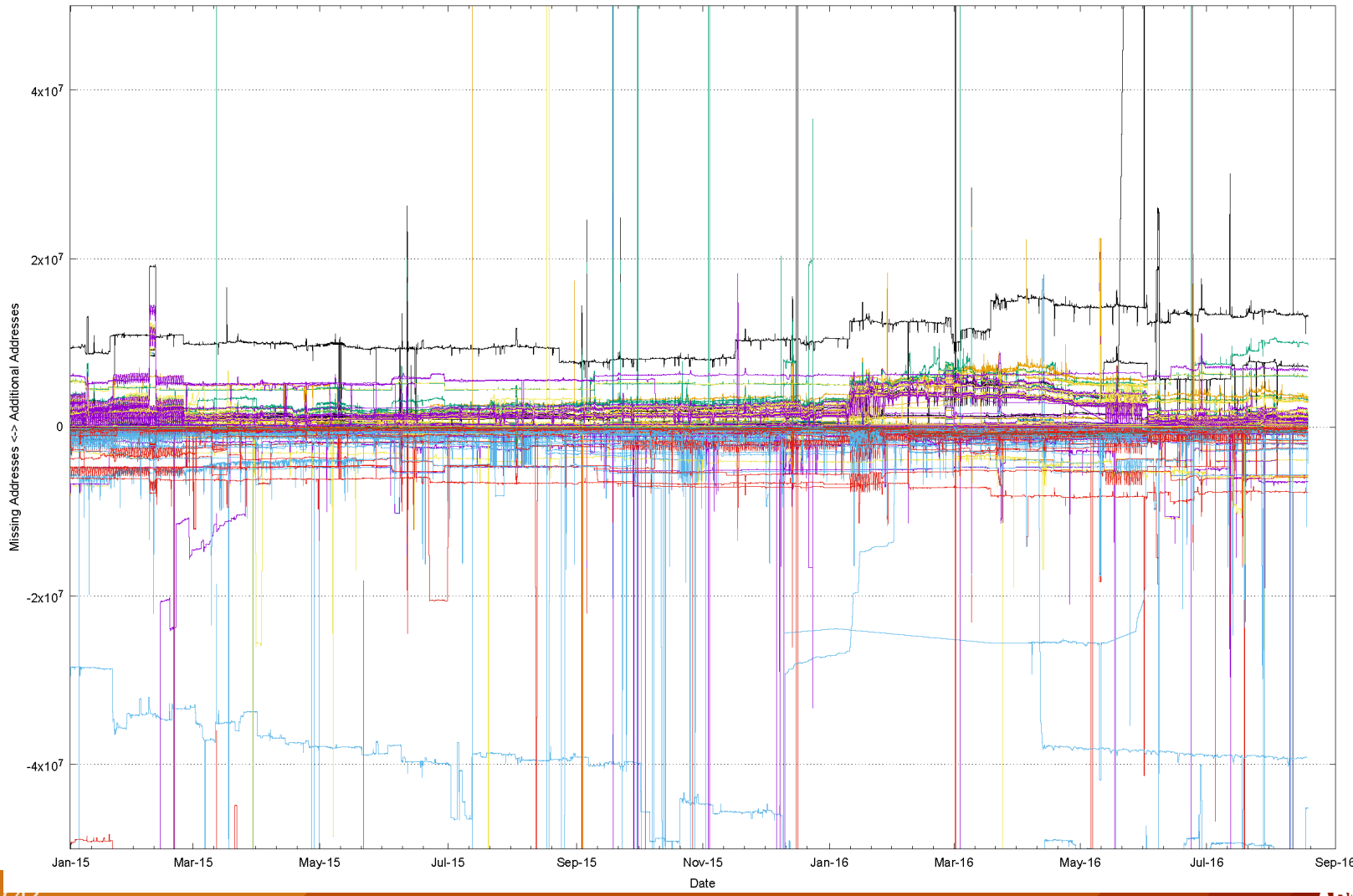
What is "default"?

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

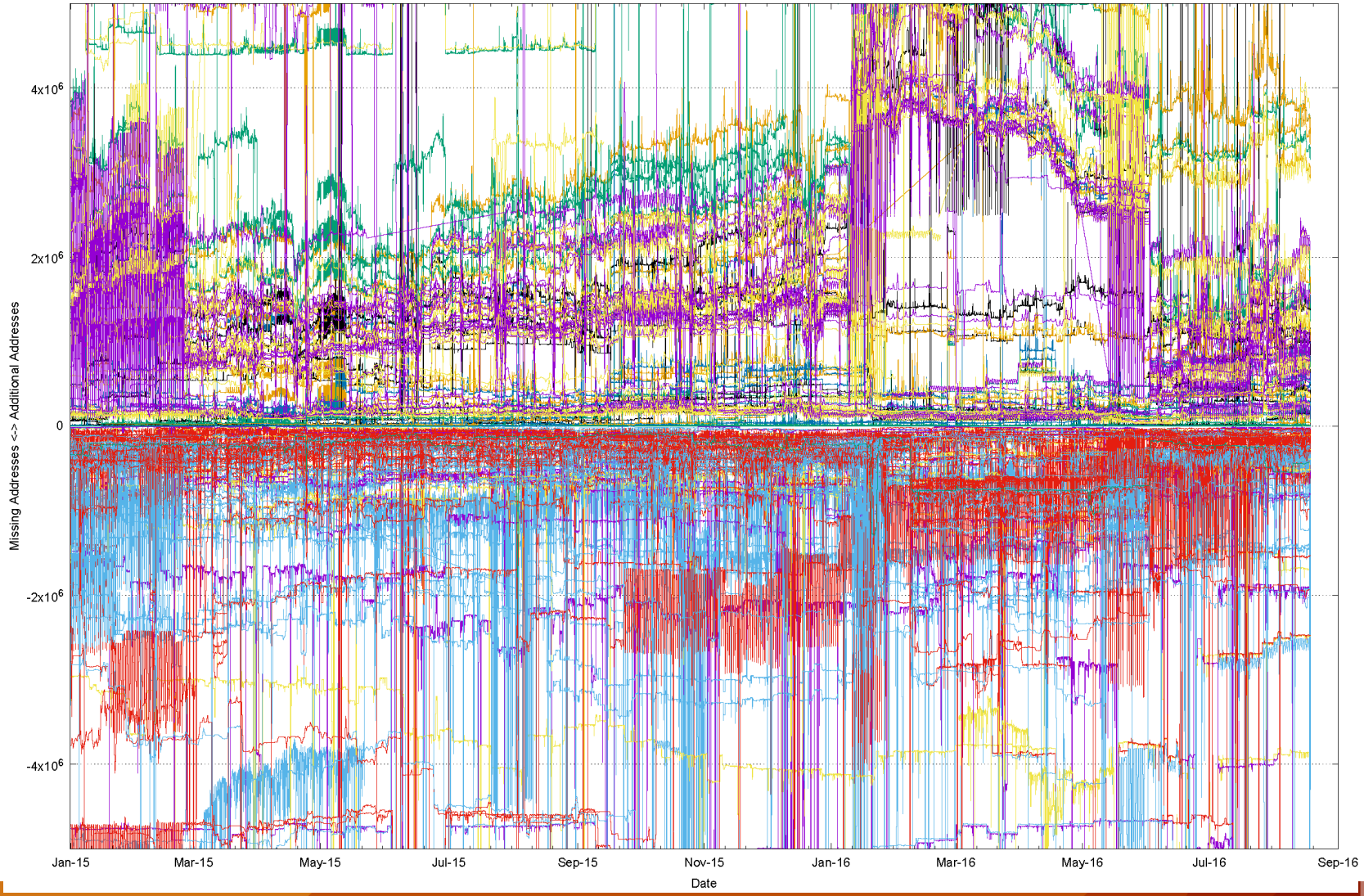
What is "default"?

- 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

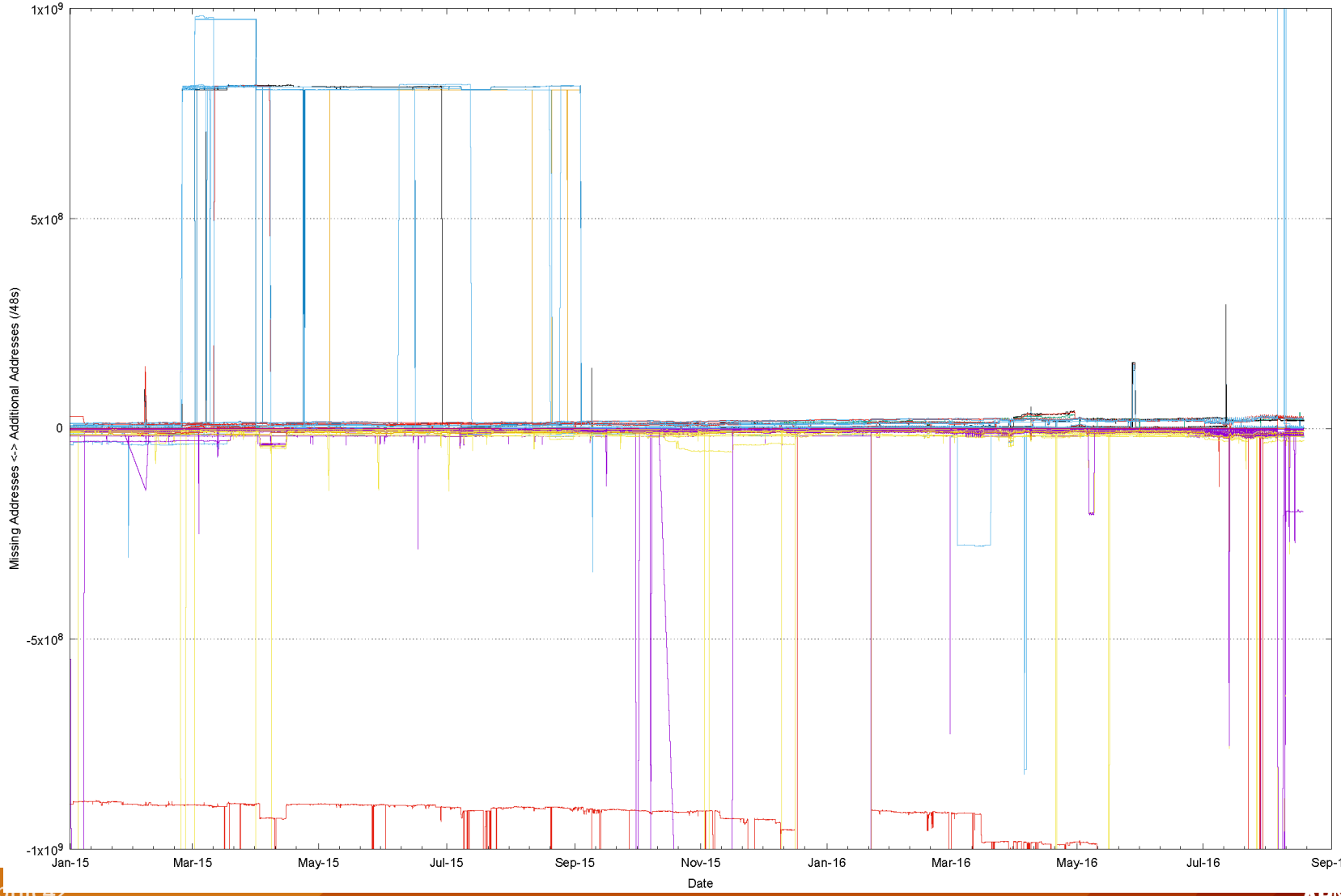
A "Quorum" deviation view of IPv4



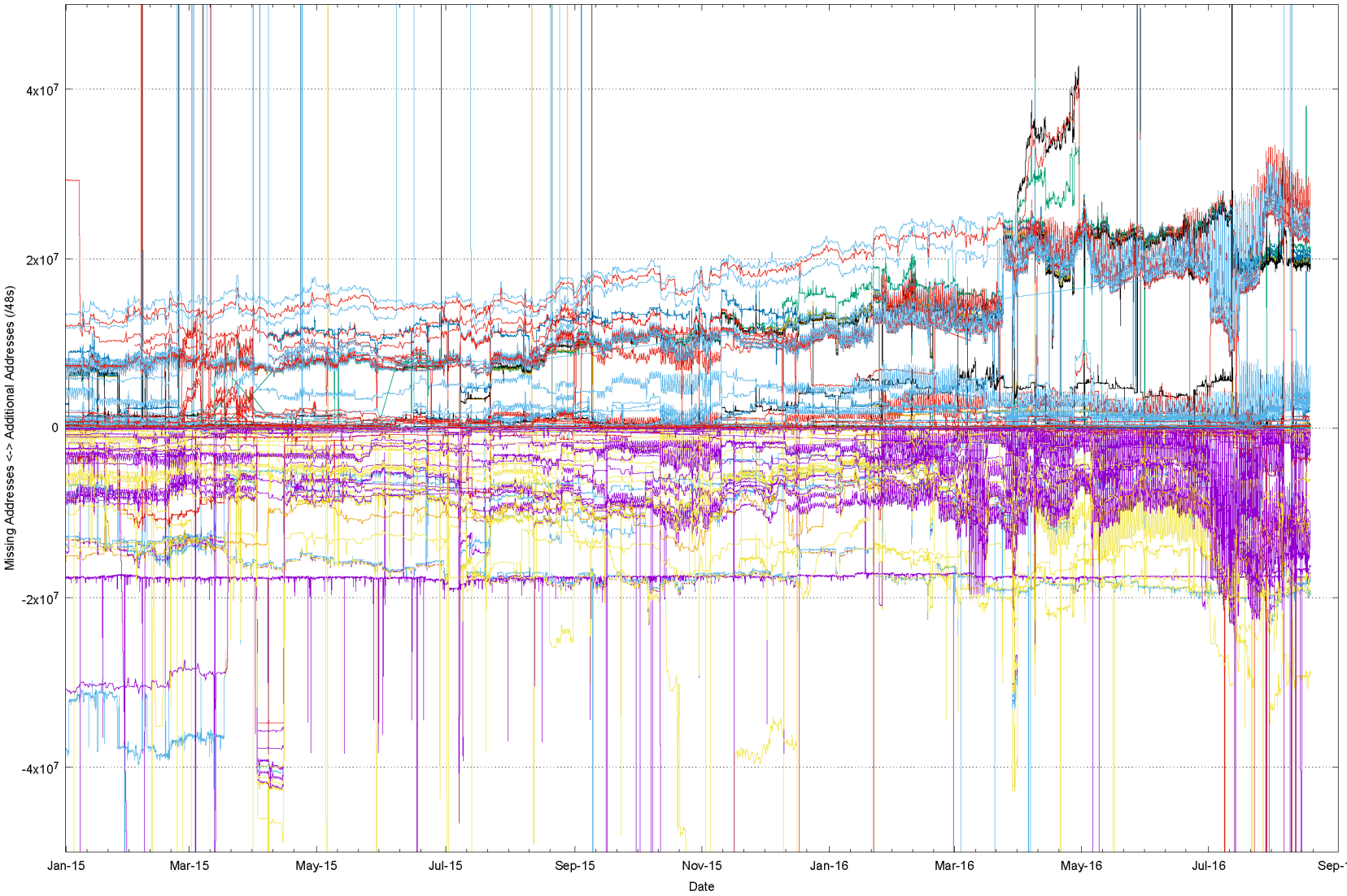
A magnified view



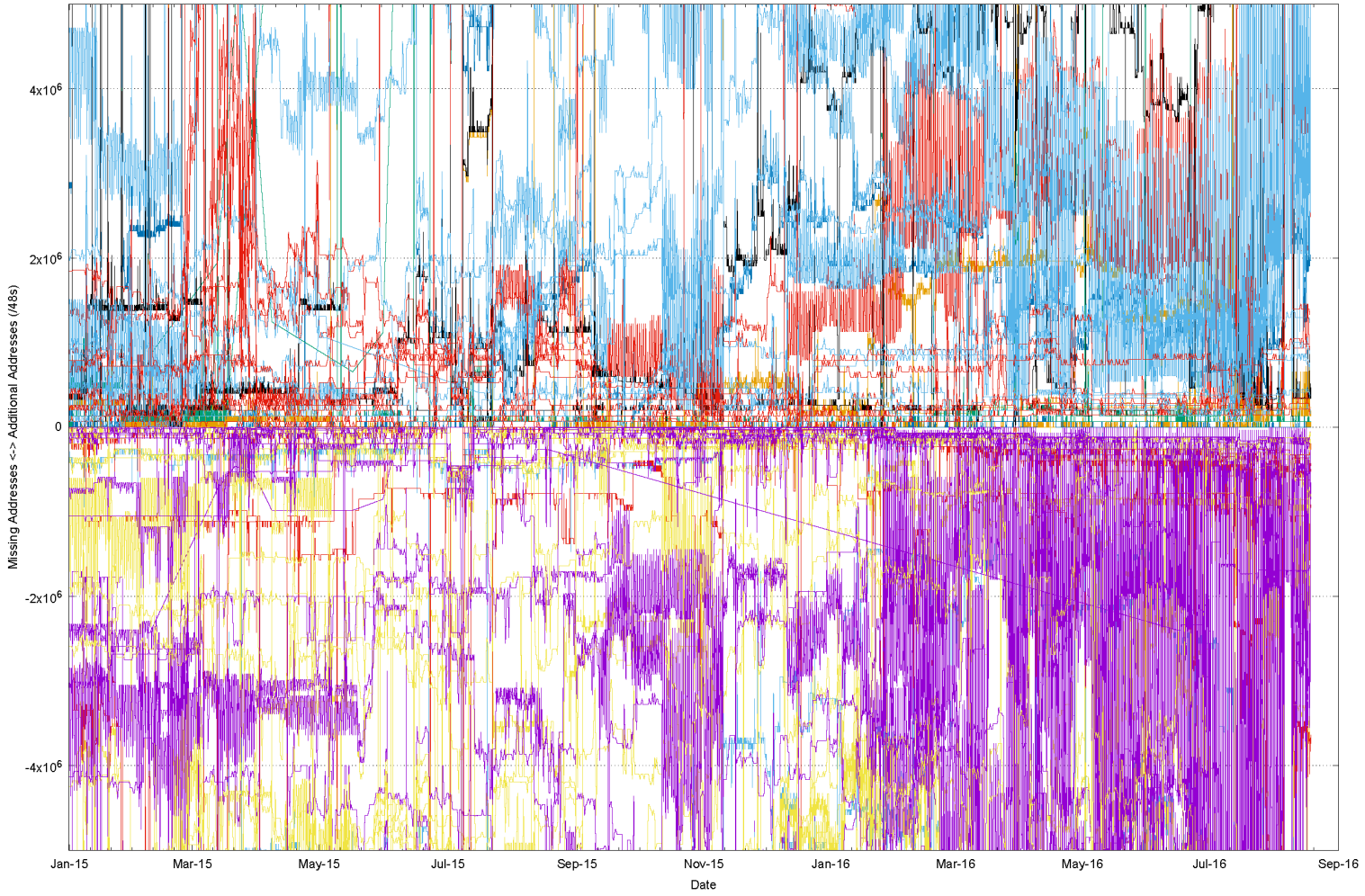
IPv6 "Quorum" Deviation



Zooming In


4x10^7
2x10^7
0
-2x10^7
-4x10^7
Missing Addresses <-> Additional Addresses (/48s)
Jan-15
Mar-15
May-15
Jul-15
Sep-15
Nov-15
Jan-16
Mar-16
May-16
Jul-16
Sep-16
Date

And Again


4x10^6
2x10^6
0
-2x10^6
-4x10^6
Missing Addresses <-> Additional Addresses (/48s)
Jan-15
Mar-15
May-15
Jul-15
Sep-15
Nov-15
Jan-16
Mar-16
May-16
Jul-16
Sep-16
Date

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	168.14	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 18th August 2016)

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 Schuberg 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

So What?

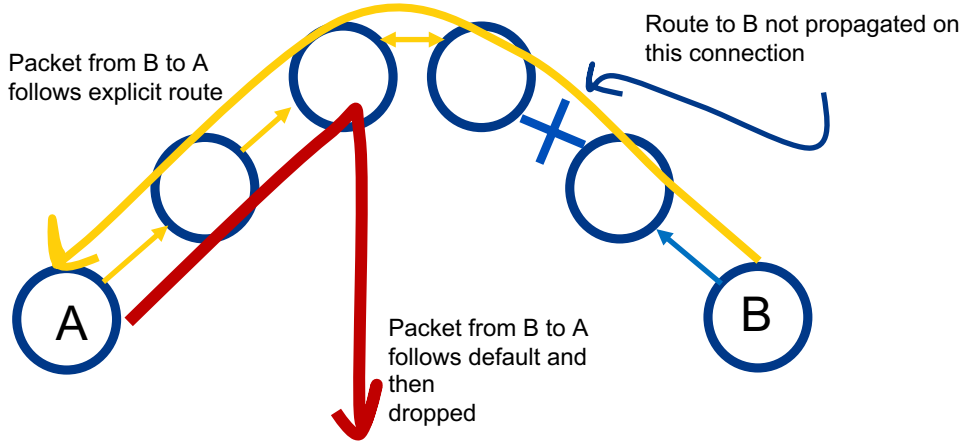
Surely all this is patched up by the widespread use of default in addition to specific routes? (*)

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

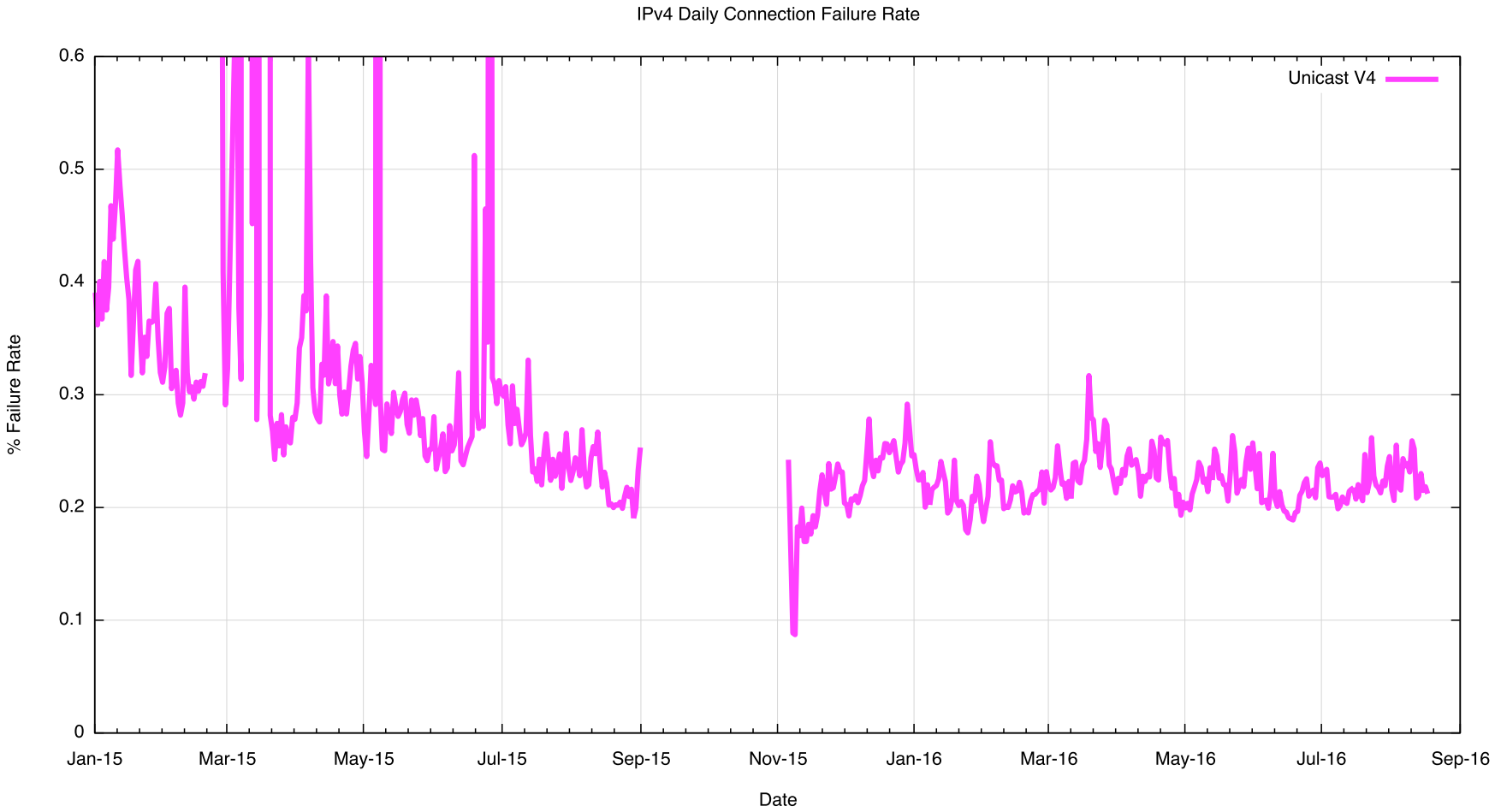
So What?

Surely all this is patched up by the widespread use of default in addition to specific routes?

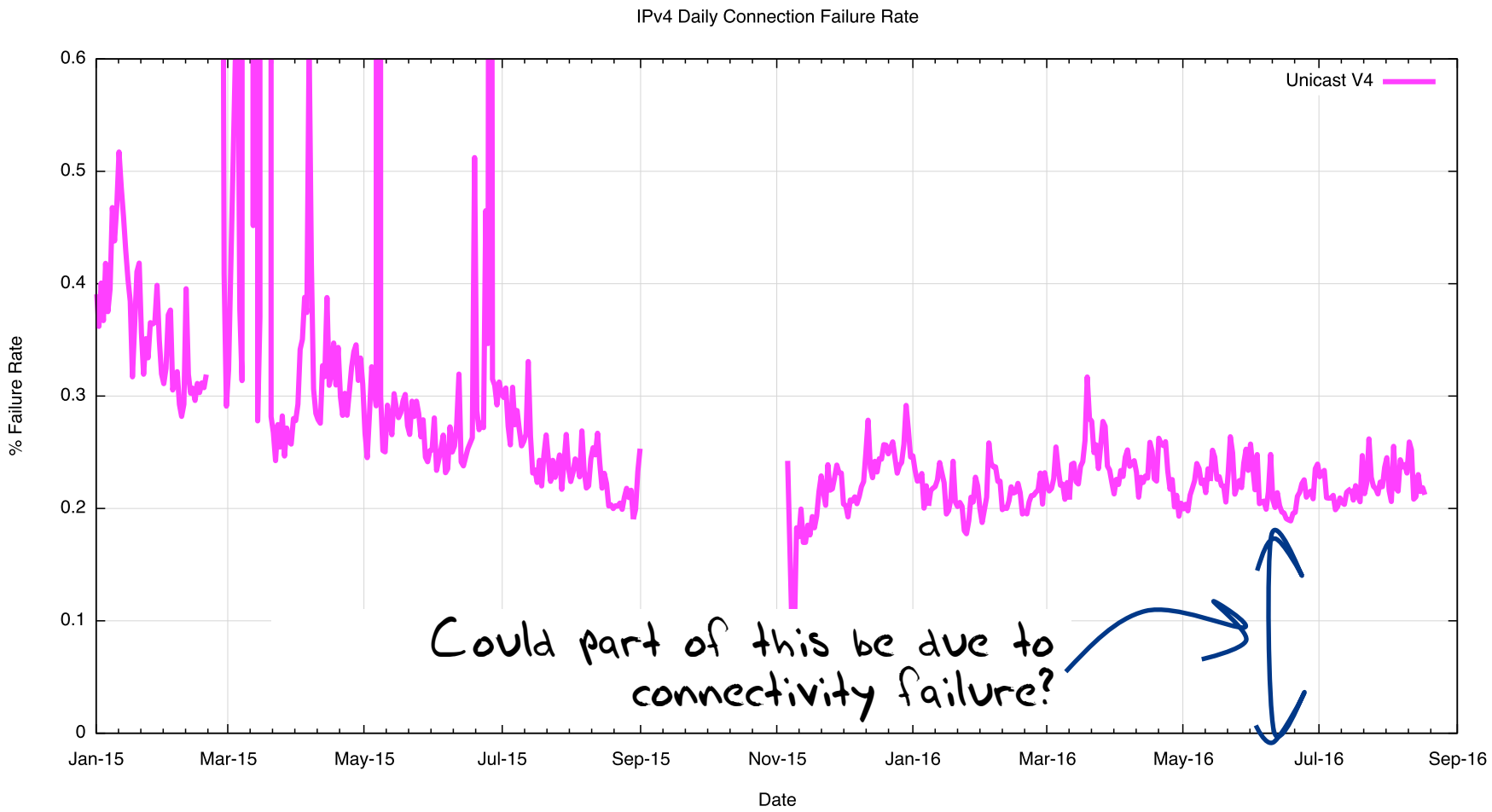
- Well, not really
- Default points along upstream transits
- It does not patch downstreams



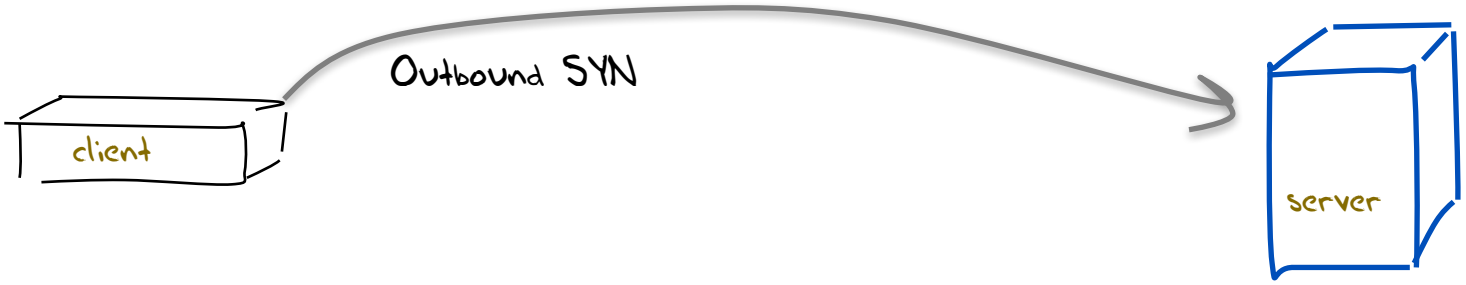
To Recap: What is causing this?



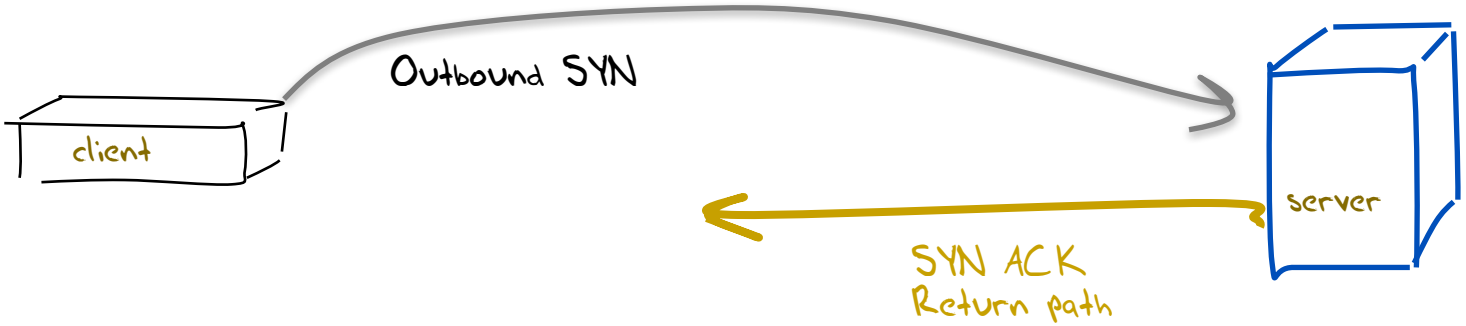
To Recap: What is causing this?



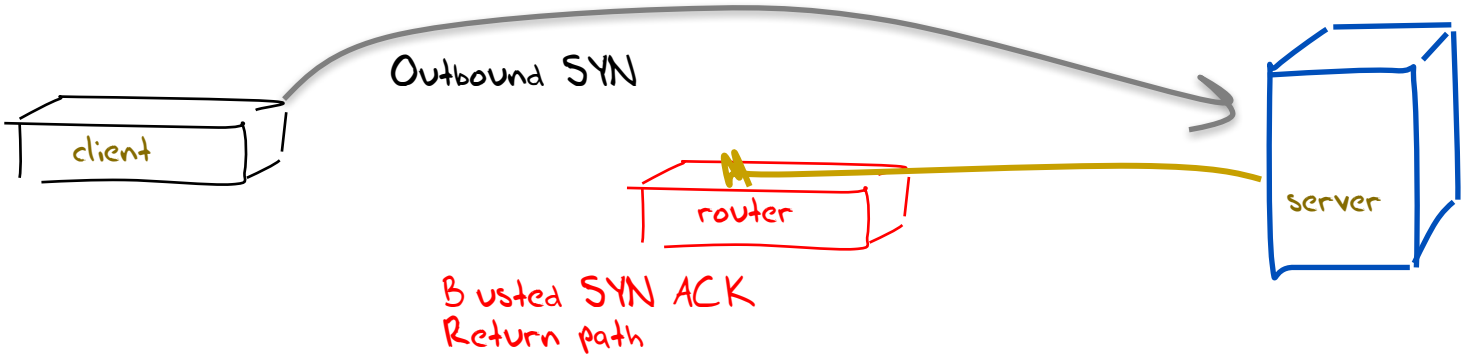
Connection Failure



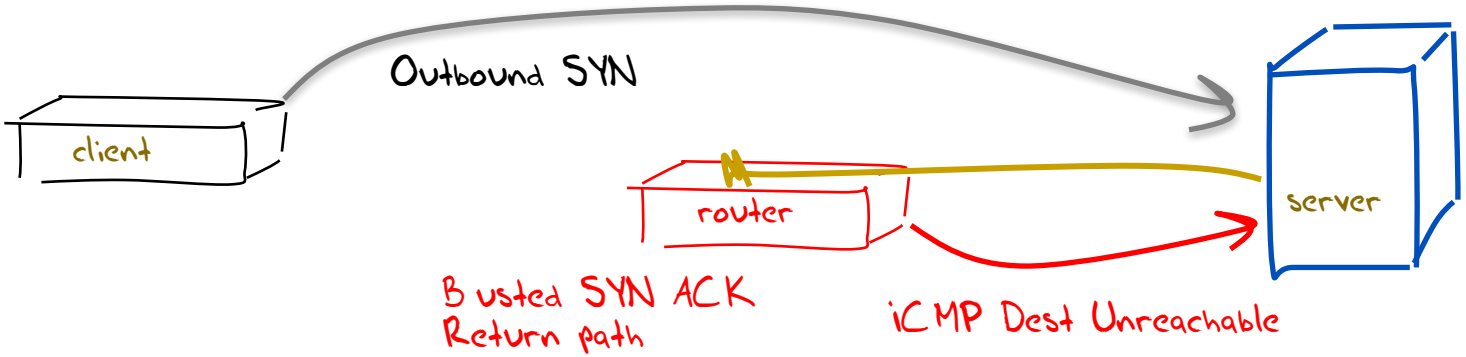
Connection Failure



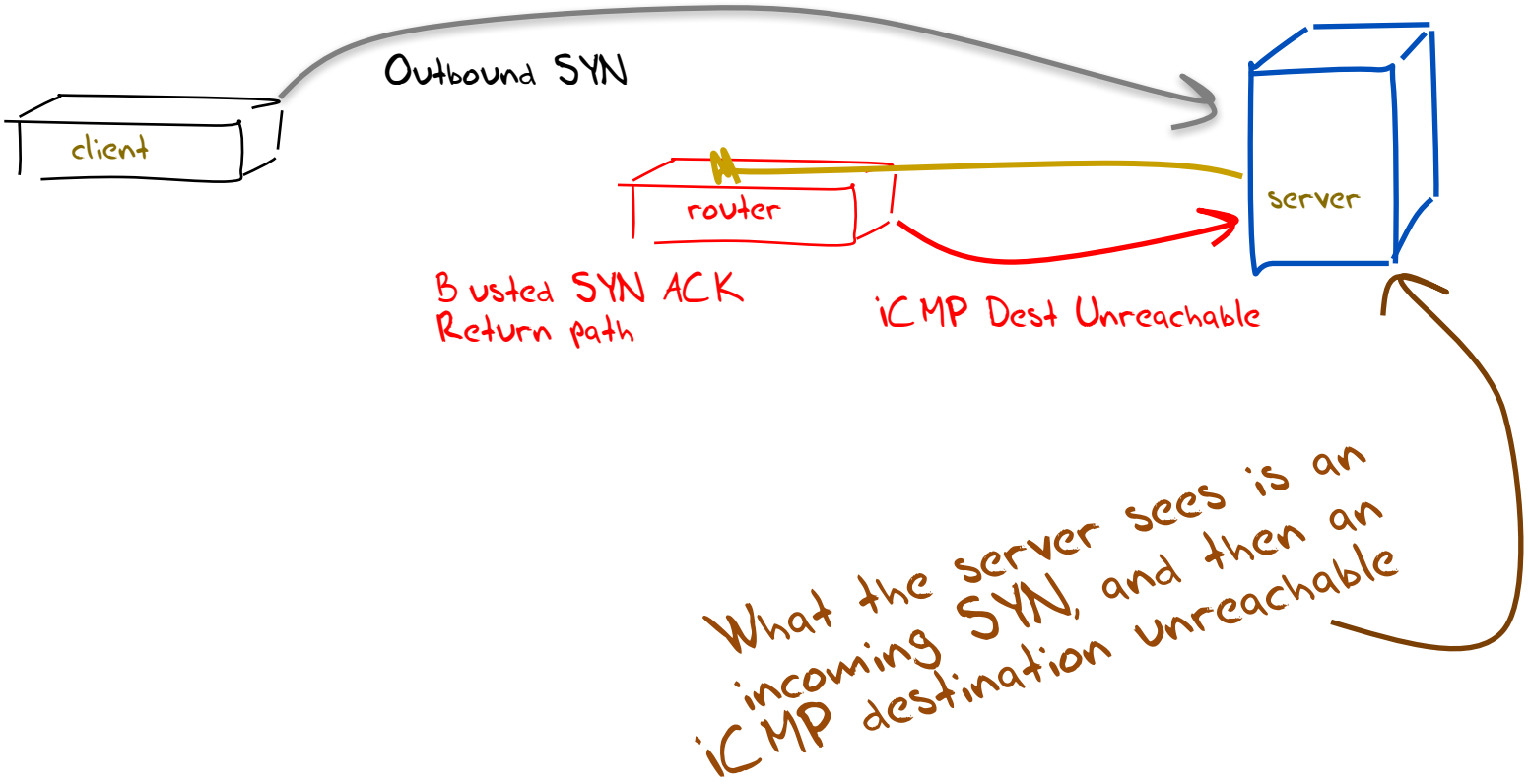
Connection Failure



Connection Failure



Connection Failure



And we 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
```

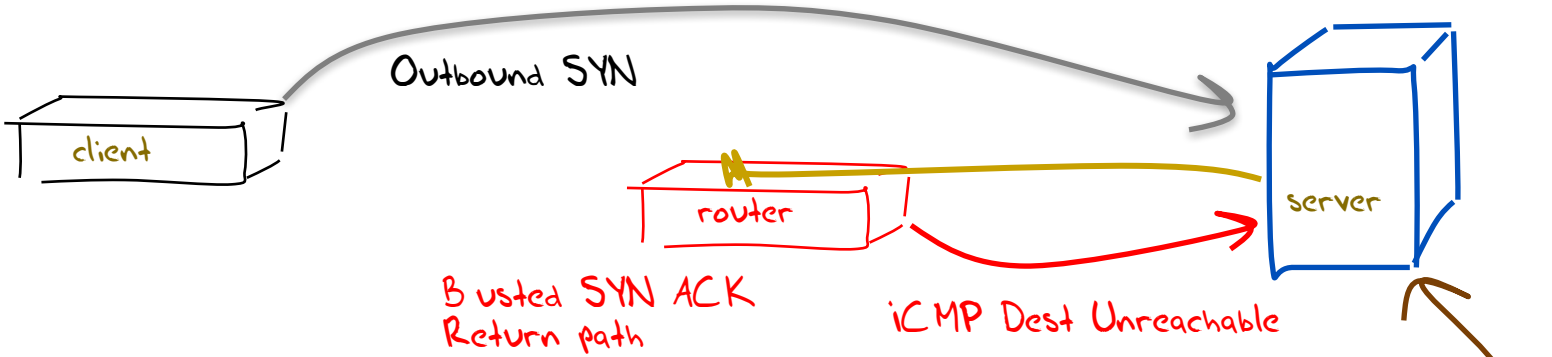
Outer packet is an ICMP Packet with a "destination unreachable" code sent to the server from the router at address 84.41.108.74

```
    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

Connection Failure



What the server sees is an incoming SYN, and then an ICMP destination unreachable

But this is not a reliable signal – routers do not always generate an ICMP Dest Unreachable, and firewalls often filter them. But they are still visible some of the time.

Of course it's not the only reason why connections fail in IPv4

NATs are also erratic:

```
14:16:37.959604 IP (tos 0x38, ttl 55, id 40091, offset 0, flags [none], proto ICMP (1), length 80)
    197.114.50.xx > 139.162.146.97: ICMP host 197.114.50.20 unreachable, length 60

        IP (tos 0x20, ttl 49, id 0, offset 0, flags [DF], proto TCP (6), length 52)
        139.162.146.97.443 > 197.114.50.xx.55648: Flags [S.], cksum 0xeb1e (correct), seq 3454637726, ack
3599361621, win 29200, options [mss 1400,nop,nop,sackOK,nop,wscale 7], length 0
```

Why is the destination telling us that it is unreachable?

Of course it's not the only reason why connections fail in IPv4

NATs are also erratic:

```
14:16:37.959604 IP (tos 0x38, ttl 55, id 40091, offset 0, flags [none], proto ICMP (1), length 80)
    197.114.50.xx > 139.162.146.97: ICMP host 197.114.50.20 unreachable, length 60

        IP (tos 0x20, ttl 49, id 0, offset 0, flags [DF], proto TCP (6), length 52)
        139.162.146.97.443 > 197.114.50.xx.55648: Flags [S.], cksum 0xeb1e (correct), seq 3454637726, ack
3599361621, win 29200, options [mss 1400,nop,nop,sackOK,nop,wscale 7], length 0
```

Why is the destination telling us that it is unreachable?

it's probably a CGN that has lost its binding state while the TCP session was being established!

Connectivity appears to have a 1 in 500 failure rate in IPv4

- Some of it is based on asymmetric views of connectivity from the routing system
- Some of it is based on anomalous NAT behaviours
- And some of it … we just can't tell!

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 it

What we now have in 2016 is a tier 1 feeder system, as it's no longer a true uniquitous fully connected peer network

And that makes me sad!

Thanks!

APNIC 42

COLOMBO, SRI LANKA

28 September – 5 October 2016

#apnic42