renesys

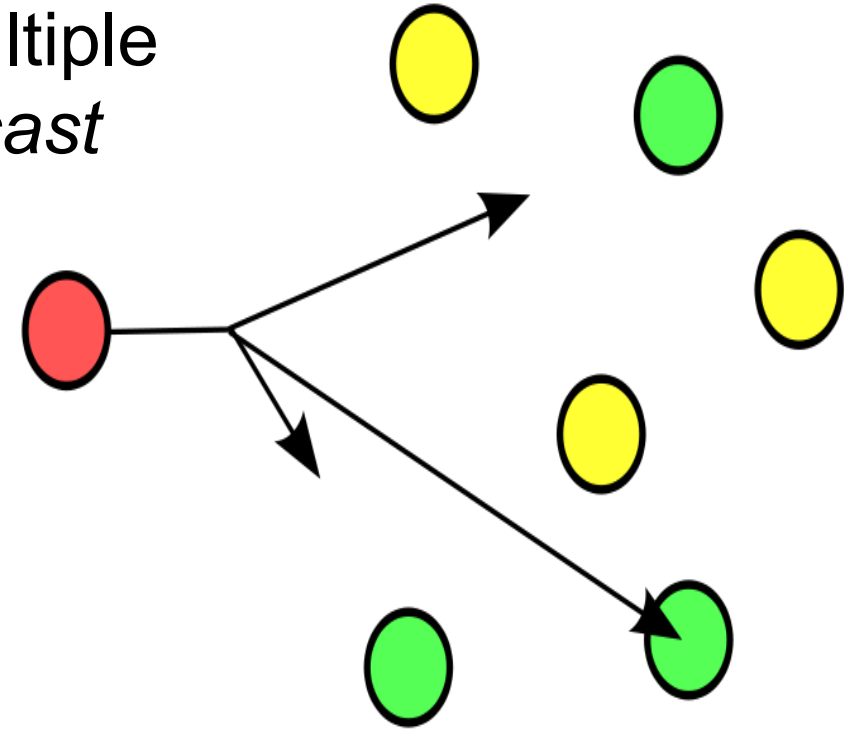
Who Are the Anycasters?

Doug Madory, Senior Analyst
Chris Cook, Intern (Bucknell)
Kevin Miao, Software Engineer

NANOG 59, Phoenix
9 October 2013

What is BGP Anycast?

- A prefix originated from multiple locations is said to be *anycast*
  - Goal: Improved reliability
  - Goal: Improved latencies
- Common Applications
  - Content Delivery
  - DDoS Protection
  - DNS Service



Two Techniques for Detecting Anycast

- We present two novel techniques for identifying anycast prefixes
  - Using global BGP routing data
  - Using global traceroute latency data
- Together they help us score prefixes for degree of *anycast-iness*
- High scoring prefixes are anycast

Why are We Doing This?

- Our work relies on *accurate* geolocation of IPs
- *Anycast* prefixes can be in *multiple locations*
- *Where is Google's 8.8.8.8? Who do you believe?*
  - United States (freegeoip.net)
  - New York, New York (Geobytes)
  - Mountain View, California (IP2Location)
  - Broomfield, Colorado (IPligence)
  - United States (Maxmind)
  - Mountain View, California (Quova)
  - United States (unlocktheinbox.com)

Besides, everyone is anycasting!
– *Anycast solves all problems!*

- "*At CloudFlare, we've done a significant amount of engineering to allow TCP to run across Anycast without flapping. This involves carefully adjusting routes in order to get optimal routing …*" – Cloudflare blog

- "*Anycast, which uses mirrored servers represented by a common IP address, can minimize transaction latency.*" – Verisign, Investor relations

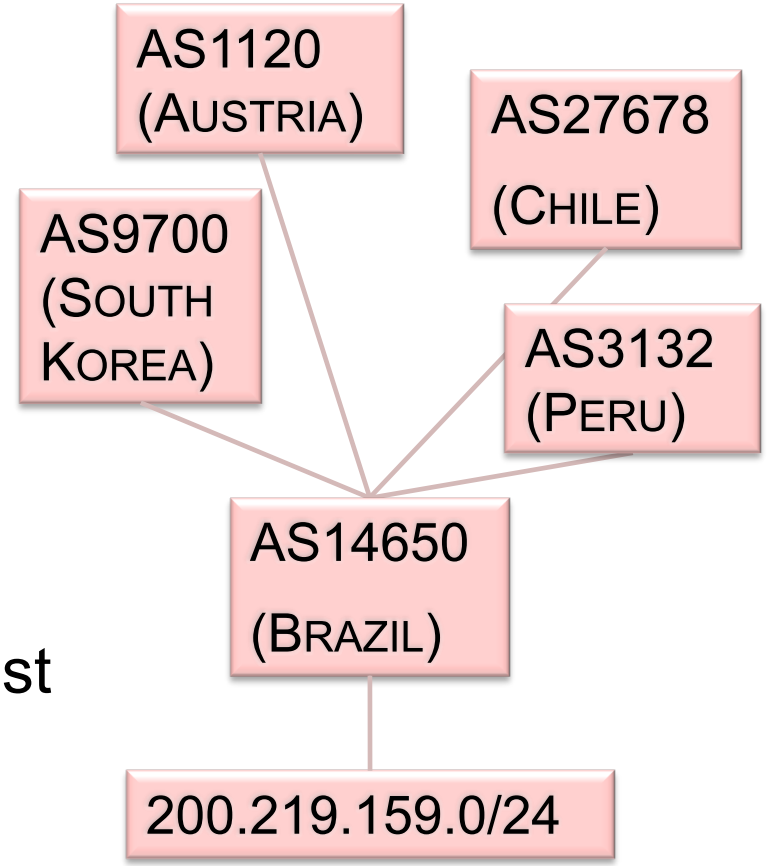
- "*Google Public DNS uses anycast routing to direct all packets to the closest DNS server.*" – Google Developers FAQ

A Brief Word about IP Geolocation Services

- We have found that commercial IP geolocation services tend to be *reasonably accurate* for eyeballs (i.e., where the ad revenue is located).
- But they can be *very inaccurate* for infrastructure IPs (e.g., routers, servers, etc.), multinational companies and 3rd world countries.
- We have spent considerable effort on fixing IP geolocation errors for *all IP addresses*, regardless of use.
- Anycast prefixes are but one corner case, but can be used to illustrate a few of our geolocation techniques.

BGP Routing Method

- Find all *domestic* providers (DPs) (i.e., those constrained in geographic scope)
- For each prefix …
  - Examine its transit tree (from business relationships)
  - Determine the DPs seen
  - Score prefix based on # of DPs, their adjacencies and geos
- Higher score → More likely anycast



Definition of a “Domestic Provider”

- What is a domestic provider?
  - 90% of transited prefixes in a single country
  - Arbitrary cut-off, but allows for some market penetration into neighboring countries and some misgeolocation
- Why domestic providers?
  - Multiple domestic providers in disparate countries/ continents transiting a given prefix imply the prefix is anycast.

Transit Tree Exceptions

- “Domestic” providers can provide transit to nearby countries
  - Russia and Kazakhstan
  - Australia and Papua New Guinea
- Satellite providers serve many countries
  - SatGate (AS 30721)
  - Eutelsat (AS 34444)
  - Many others
- Such edges are excluded from anycast scoring in the routing technique

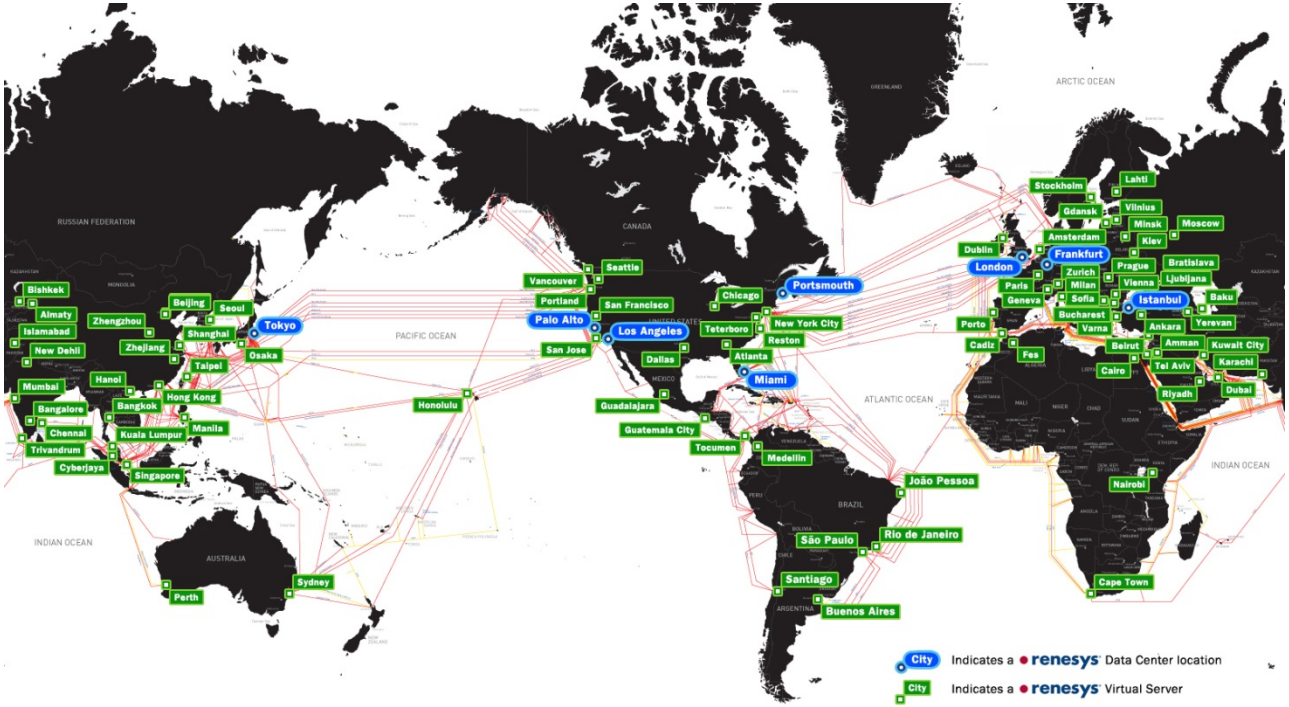
Example Anycast Prefixes from Routing

Prefix	Domestic Provider Countries	Org
178.249.136.0/21	CA DE DK ID NL RU SE US ZA	SoundCloud
93.174.200.0/21	DE DK ID RU SE US ZA	2AT
193.0.16.0/24	BT CZ DE GB RU US ZA	K root-server
194.14.219.0/24	BG DE IT NL PL US	Edis
178.250.72.0/21	BT DE GB RU US ZA	Exonetric Consulting
93.171.226.0/24	DE RU UA US	PE Zvonkova Tatyana Aleksandrovna
84.205.78.0/24	DE GB NL US	RIPE
200.1.121.0/24	CL CZ MX US	Chile NIC

- All of these prefixes score very highly – Many domestic providers in highly disparate locations

Latency Technique

- Uses recent latency measurements to all routed prefixes from locations around the world

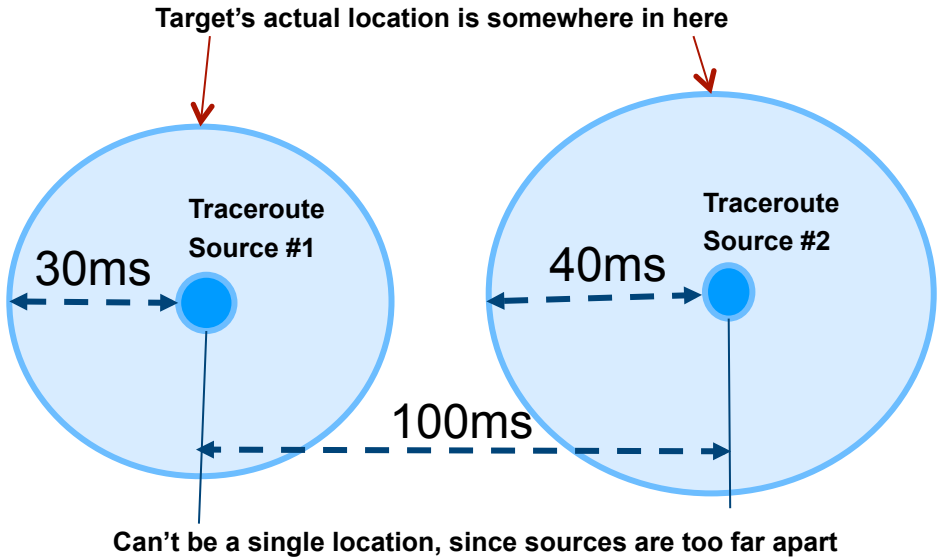


renesys **Active Measurement Infrastructure** - July 2013 (plus Global Submarine Cable Map)

Note: Some cities host multiple collectors. Cable map credit: Telegeography

Latency Technique

- Based on concept of “geo-inconsistency”
  - *If the sum of observed latencies to a single IP address from two traceroute sources is less than the minimum possible latency between the sources, then the IP must be in more than one location*



Let's Use One of Our Tools

$ min-latency --today 8.8.8.8

- Sydney 0ms
- Hong Kong 1ms
- Kuala Lumpur 2ms
- Taipei 2ms
- Frankfurt 2ms
- Chennai 3ms
- Amsterdam 5ms
- … (many more) …
- Cape Town 193ms
- São Paulo 200ms
- João Pessoa 210ms
- Nairobi 216ms
- Dubai 267ms

- Consider minimum observed latencies over some period
- Looks very geo-inconsistent, but not from everywhere
- *No one can be everywhere, peering with everyone*
- How can we measure degree of anycast-iness?
  - How many pairs of locations are geo-inconsistent?
  - How many possible pairs are there?
  - Take a ratio, higher is better: larger % of inconsistent pairs

Our tools do that too …

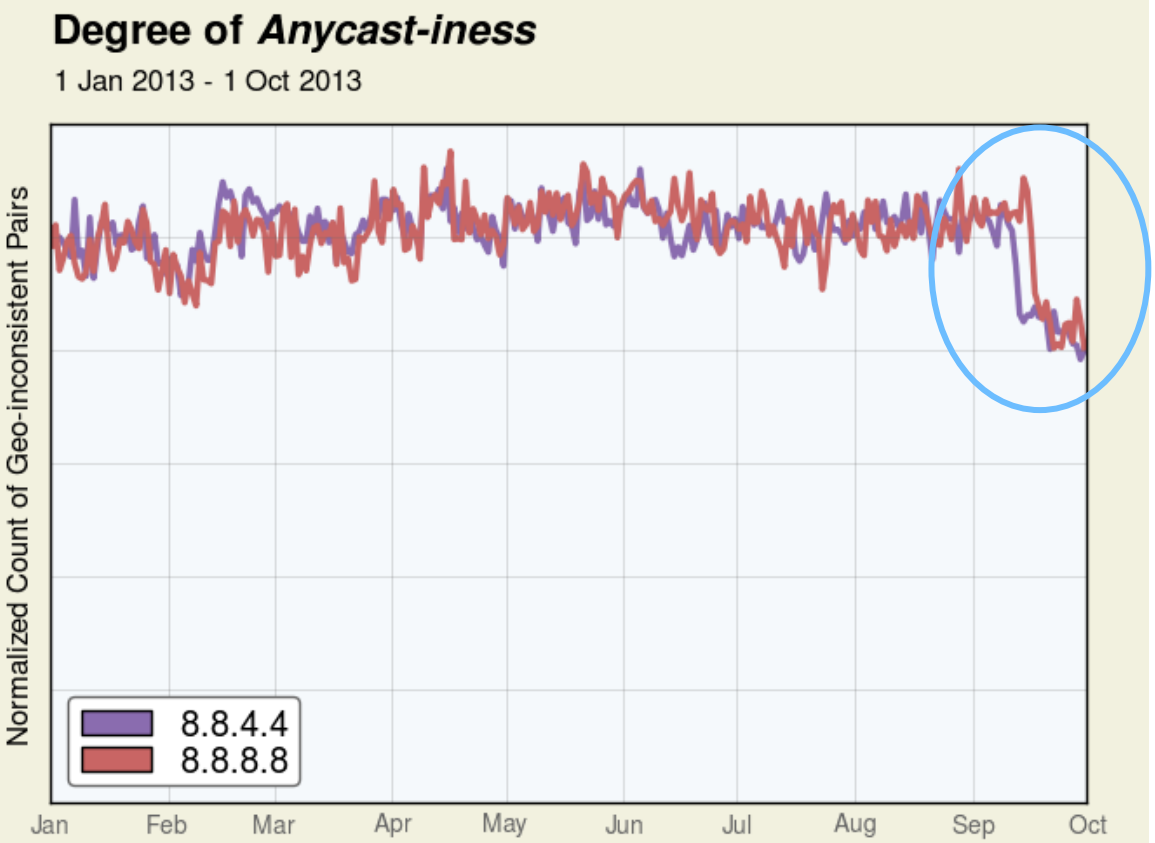
$ min-latency --geocheck --today 8.8.8.8

- 8.8.8.8 geo-inconsistent (2655)
- Hong Kong=1ms Sydney=0ms Min=72ms
- Kuala Lumpur=2ms Sydney=0ms Min=65ms
- … (many more) …
- Buenos Aires=157ms Taipei=2ms Min=187ms
- Buenos Aires=157ms Shanghai=30ms Min=190ms

- 2,655 geo-inconsistent pairs relative to this Google IP address on 4 October 2013
- Out of 6,216 possible pairs that day
- Google's DNS servers have very high anycast-iness!

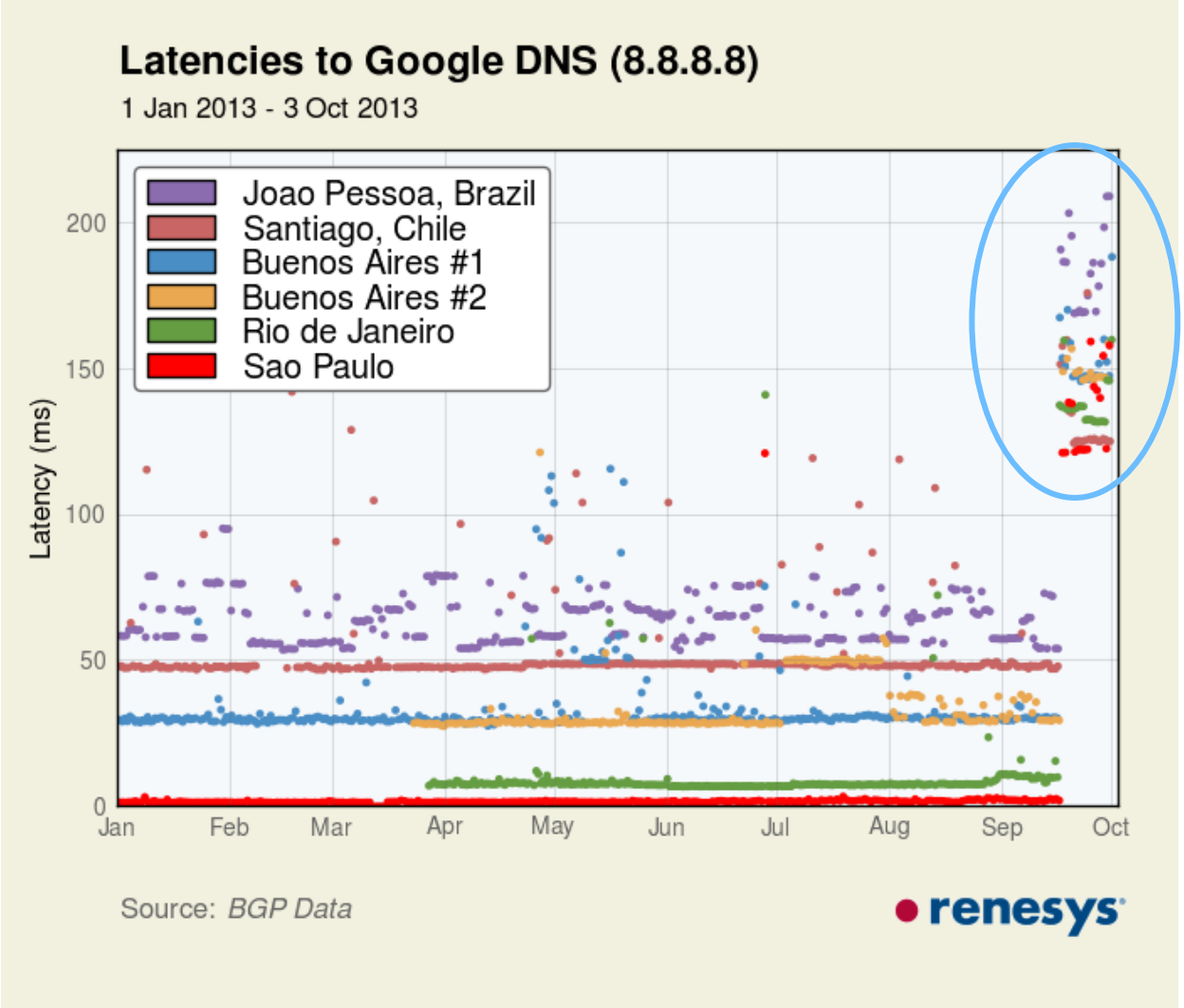
Anycast-iness of Google DNS over Time

- Consistently high anycast score over time for both of Google's public DNS server IPs
- Normal daily variance, depends on routing and provider options
- Something happened in September and is ongoing



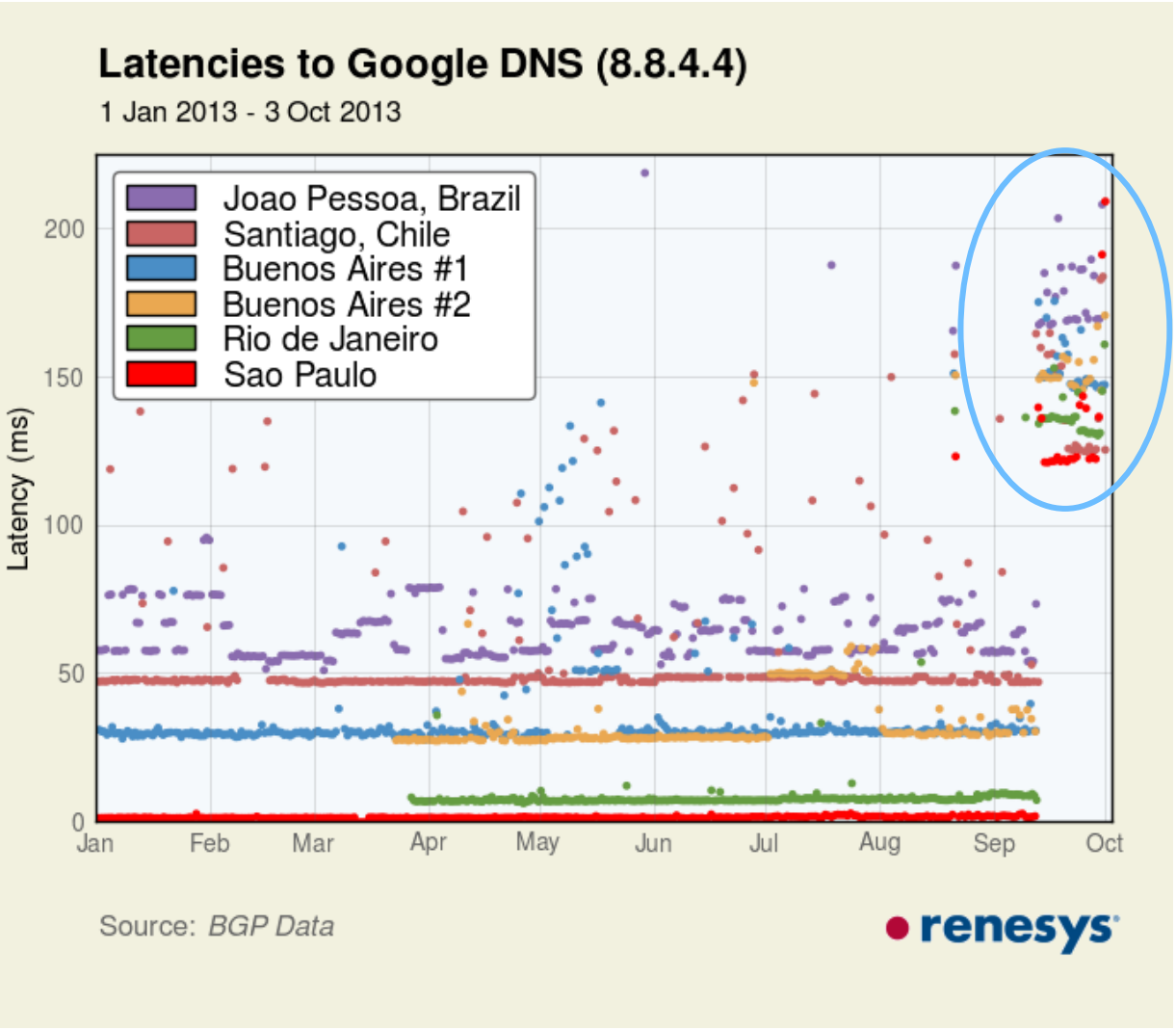
Google DNS Service via 8.8.8.8

- Considerable jump in traceroute latencies from South America to Google's 8.8.8.8
- Problem started September 16th
- Ongoing as of Oct 4th



Google DNS Service via 8.8.4.4

- Considerable jump in traceroute latencies from South America to Google's 8.8.4.4
- Problem started on September 12th (4 days before 8.8.8.8)
- Ongoing as of Oct 4th



What happened here?

- New Google IP hop makes an appearance on our traceroute measurements from this region
- Latencies jump considerably as a result
- Internal routing issue at Google? Google turns off South American data centers?
- We can only observe and alert on the change
- The problem, if any, must ultimately be resolved by the anycaster or its providers

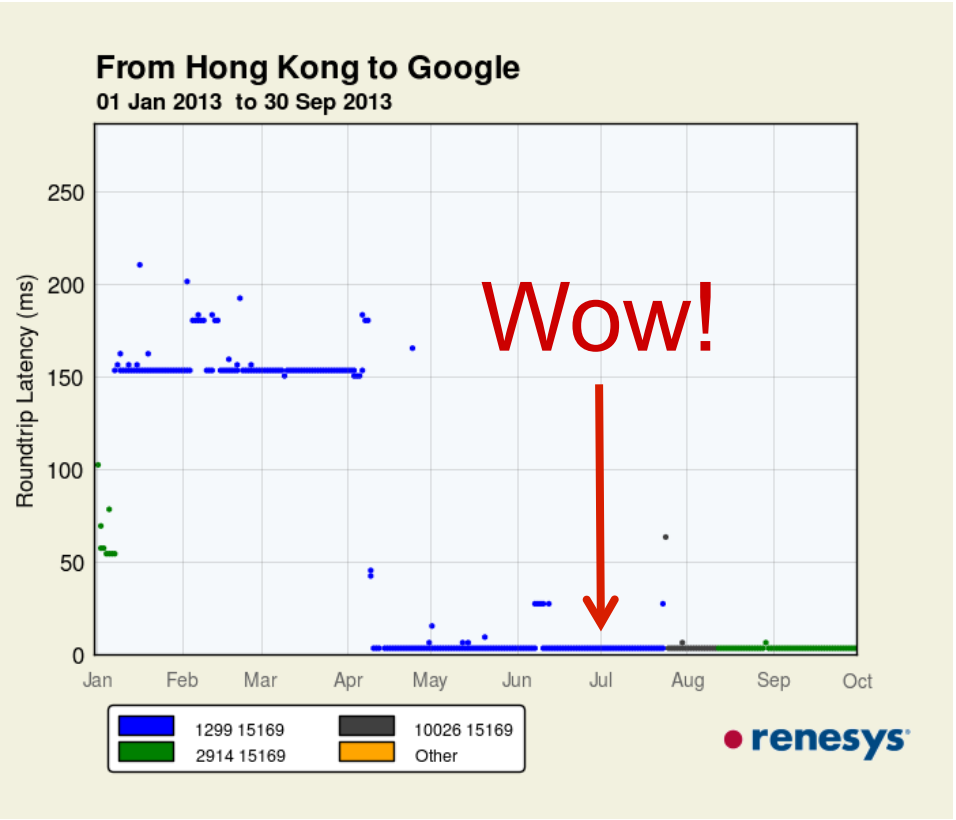
Example Trace Fragments (Before & After)

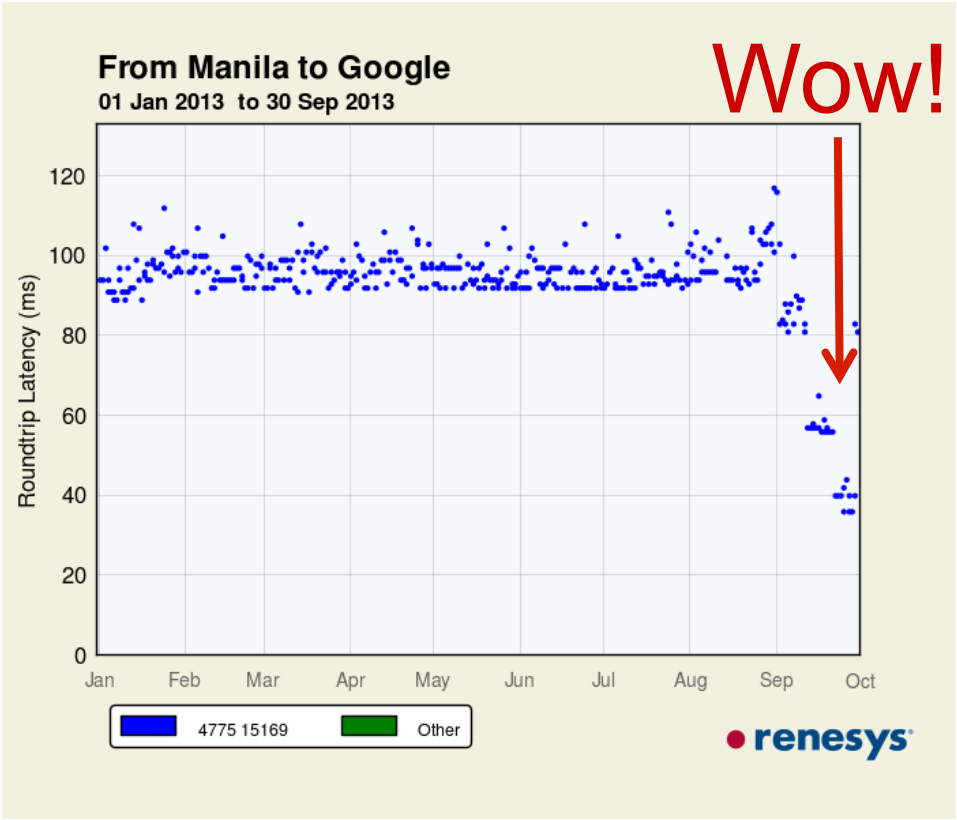
Hop	Before	2013/09/15 23:57
	…	
5	209.85.254.136	9.699ms
6	72.14.233.89	9.744ms
7	64.233.175.18	9.906ms
8	8.8.8.8	10.283ms

Hop	After	2013/09/16 13:59
	…	
5	209.85.254.136	9.703ms
6	72.14.236.174	16.650ms
7	72.14.235.106	137.974ms
8	209.85.252.96	137.974ms
9	209.85.248.29	135.957ms
10	*	*
11	8.8.8.8	136.594ms

Can't Anycast Reduce Latencies?
– *Absolutely, but no guarantees*

- Depends on available providers (both ends) & routing





What about Minimally Anycast Prefixes?
– *Same techniques find these as well*

$ min-latency --geocheck --today 37.209.240.0/24

sports.ru

- 37.209.240.1 geo-inconsistent (2)
- Kiev=1ms Moscow1=1ms Min=7ms
- Moscow2=3ms Kiev=1ms Min=7ms

$ min-latency --geocheck --today 180.76.9.0/24

baidu.hk

- 180.76.9.1 geo-inconsistent (2)
- Osaka=12ms Hong Kong1=2ms Min=26ms
- Osaka=12ms Hong Kong2=3ms Min=26ms

180.76.9.0/24 is also multi-originated (MOAS)

- Beijing Baidu Netcom Science and Technology Co. (AS 55967)
- BaiduJP (AS 38627)

Who Are the Anycasters?

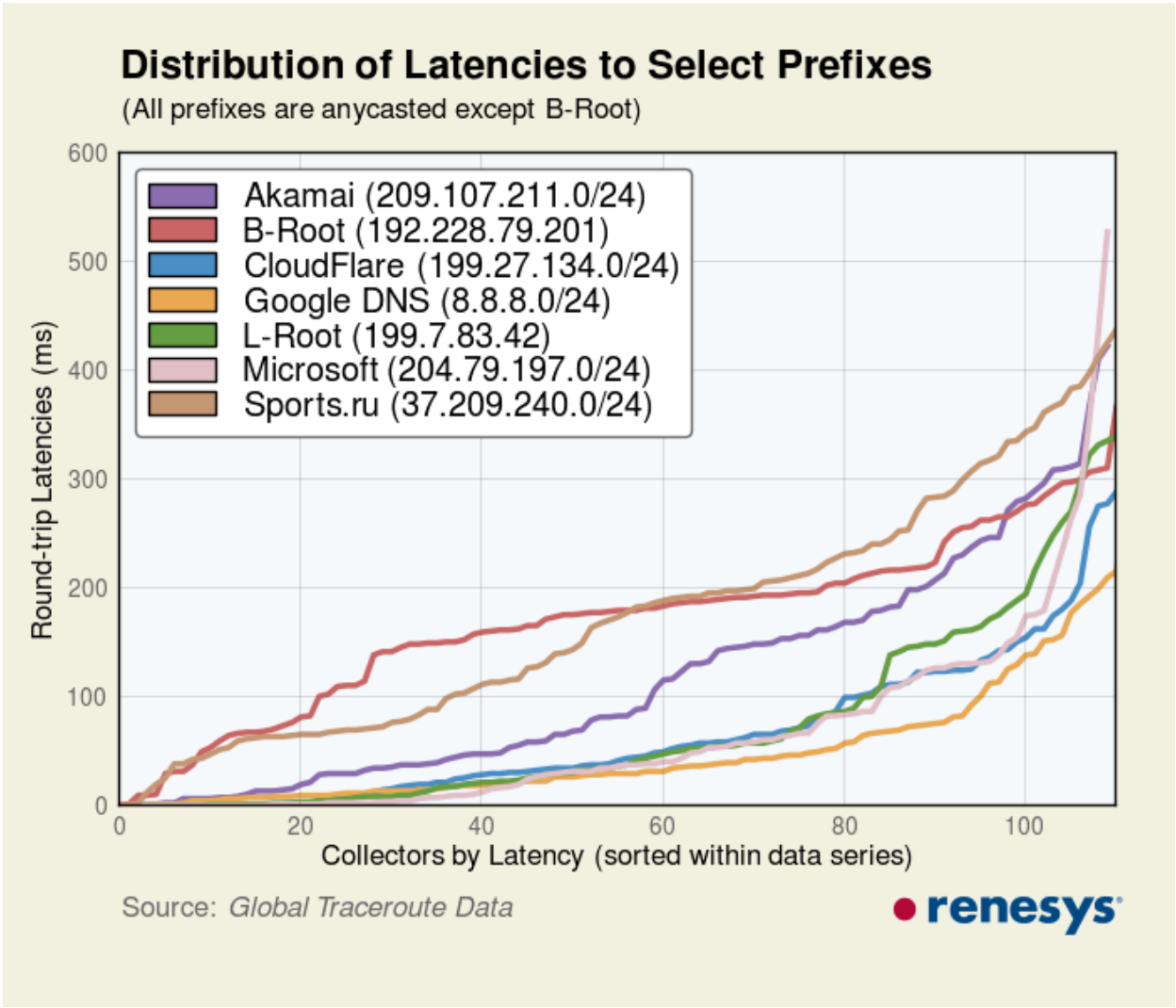
Top Dozen by Prefix Count

• Afilias Canada	69
• UltraDNS	32
• Pacnet	22
• VeriSign	20
• Edgecast	19
• CloudFlare	18
• Prolexic	15
• BitGravity	15
• AAPT	11
• Neustar Ultra Services	10
• China Internet Net Info Center	10
• AT&T	10

Other Notables

• Google	8
• Yahoo!	5
• Amazon	4
• Yandex	4
• Akamai	1

Anycasting is a Matter of Degree
– *Let's compare a few*



- Distinct differences in anycast coverage
- Google has best coverage

How Do These Prefixes Compare?

- Rootservers.org states
  - B root: 1 site
  - L root: 146 sites
- Similar latency distribution for Sports.ru (minimally anycast) and B root (not anycast).
- Google has lowest latencies → Best global coverage
- Big latency spikes imply inadequate coverage
- Microsoft spikes the worst at the tail of its distribution

Does having lots of sites help?
– *Seems to in this low latency example*

Traceroute from Los Angeles to L root:

1	sc-smv1717.servint.net	(206.214.212.114)	0.066	ms Los Angeles, CA
2	ge-9-0-2.er1.lax112.us.above.net	(64.125.195.217)	0.216	ms Los Angeles, CA
3	xe-5-3-0.cr2.lax112.us.above.net	(64.125.21.189)	0.489	ms Los Angeles, CA
4	xe-2-0-0.mpr1.lax12.us.above.net	(64.125.31.193)	0.984	ms Los Angeles, CA
5	ae7.edge2.LosAngles.Level3.net	(4.68.70.33)	0.413	ms Los Angeles, CA
6	ae-3-80.edge6.LosAngeles1.Level3.net	(4.69.144.142)	0.449	ms Los Angeles, CA
7	ae-17-17.car2.LosAngeles1.Level3.net	(4.69.201.14)	0.835	ms Los Angeles, CA
8	INTERNET-CO.car2.LosAngeles1.Level3.net	(4.59.60.170)	0.460	ms Los Angeles, CA
9	l.root-servers.net	(199.7.83.42)	0.470	ms Los Angeles, CA

Does having lots of sites help?
– *Not in this high latency example. Paths matter!*

Traceroute from Los Angeles(2) to L root:

1	vl221.mag02.lax04.atlas.cogentco.com	(38.122.147.49)	0.424	Ms Los Angeles, CA
2	te0-7-0-12.ccr22.lax04.atlas.cogentco.com	(154.54.87.6)	0.720	ms Los Angeles, CA
3	be2022.mpd22.lax01.atlas.cogentco.com	(154.54.88.1)	1.091	ms Los Angeles, CA
4	be2066.ccr22.iah01.atlas.cogentco.com	(154.54.7.53)	36.763	ms Houston, TX
5	te0-1-0-3.ccr21.atl01.atlas.cogentco.com	(154.54.2.81)	51.115	ms Atlanta, GA
6	te0-4-0-7.mpd22.dca01.atlas.cogentco.com	(154.54.27.93)	61.951	ms Washington, DC
7	te0-7-0-20.mpd21.jfk02.atlas.cogentco.com	(154.54.1.106)	68.533	ms New York City
8	te0-4-0-3.ccr21.bos01.atlas.cogentco.com	(154.54.44.18)	74.541	ms Boston, US
9	te0-3-0-1.ccr21.lpl01.atlas.cogentco.com	(154.54.31.230)	140.231	ms Liverpool, GB
10	te0-7-0-6.ccr21.man01.atlas.cogentco.com	(154.54.36.58)	141.216	ms Manchester, GB
11	te1-1.ccr01.gla01.atlas.cogentco.com	(130.117.3.173)	145.728	ms Glasgow, GB
12	te1-1.ccr01.edi01.atlas.cogentco.com	(130.117.3.181)	147.624	ms Edinburgh, GB
13	149.6.12.6	(149.6.12.6)	147.286	ms Edinburgh, GB
14	gi2.edge2.dun.scotland.net	(213.128.240.9)	151.869	ms Dundee, Scotland, GB
15	v242.g1-1.core2.dun.scotland.net	(213.128.240.34)	151.496	ms Dundee, Scotland, GB
16	**l.root-servers.net**	**(199.7.83.42)**	**151.820**	**ms Dundee, Scotland, GB**

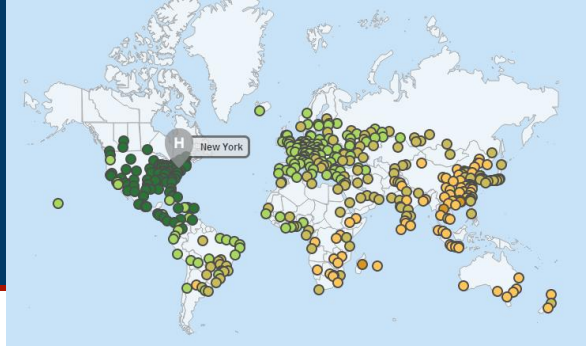
Anycast Prefix Totals & Sizes

- 593 Anycast Prefixes Detected Globally
  - 520 (88%) are /24s
  - 20 (3%) are /23s
  - 53 (9%) are larger
  - Larger prefixes may be anycast *in part* – they definitely contain some geo-inconsistent IP
- Some larger ones include …
  - 217.160.80.0/22 1&1 Internet
  - 114.114.112.0/21 China Unicom
  - 141.1.0.0/16 Cable & Wireless

Conclusions about Anycast Prefixes

- 593 routed prefixes are anycast
  - 0.13% of global IPv4 routing table
- We consistently see about 600 anycast prefixes over time
  - Surprisingly few and no real growth
- Mostly for CDN and DNS services
  - Also used for DDoS protection services
- Anycast prefixes are dominated by /24's

Overall Conclusions



- Unicast hosting from a single location
  - Local latencies might be low, distant ones will not be
- Anycast hosting from multiple locations
  - Latencies might be low from more locations
- No one can be everywhere, not even Google
  - Have to pick your markets & providers carefully then measure performance continuously
- Anycast is not a panacea. Traversed Paths Matter!

renesys

Who Are the Anycasters?

Doug Madory, Senior Analyst
Chris Cook, Intern (Bucknell)
Kevin Miao, Software Engineer

NANOG 59, Phoenix
9 October 2013