Anycast Peering and Sinkholes

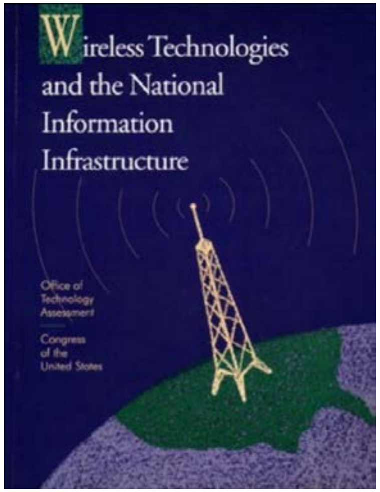
Greg Wallace

ICANN - 63, Barcelona
ccNSO Tech Day
Monday 21 October, 2018

Agenda

- Introduction
- Some anycast best practices
- Sinkhole examples

Intro: Whois Greg Wallace



Wireless Technologies and the National Information Infrastructure

Office of Technology Assessment

Congress of the United States

1995

2001

INLET TECHNOLOGIES

2008

SONY CREATIVE SOFTWARE ACQUISITION OF NETBLENDER BOOSTS BLU-RAY AUTHORING

2011

RootBSD

2015

THE LINUX FOUNDATION

node JS

2017

netactuate.com @netactuate

Intro: Who is NetActuate

- Global infrastructure provider and integrator: connectivity, colocation, cloud, IaaS, and managed services
- HQ in Raleigh, NC
- 7th largest global network by number of peers (source: https://bgp.he.net/report/peers)

2,100+
Clients

33
Datacenters

112
Expansion PoPs

2400+
BGP Peers

25 billion
Transactions Processed Per Day

7th
Generation Cloud Platform

25
Domestic & International Markets

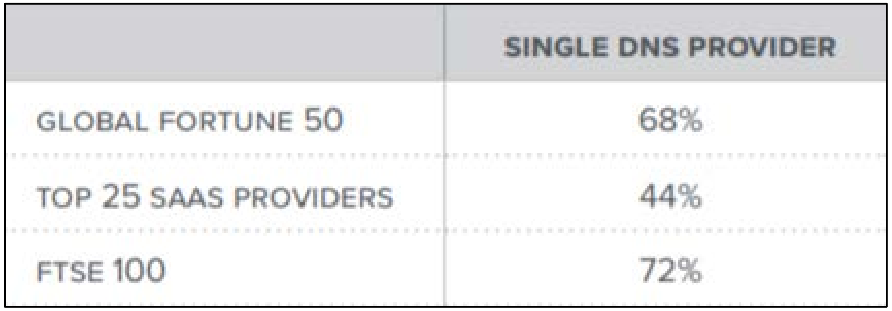
20
Internet Exchanges

FreeBSD FOUNDATION

Anycast best practices

1. Avoid SPOFs (networks/vendors)
2. Global monitoring
3. DDoS mitigation plan
4. Announce with even AS Paths
5. Make use of BGP communities
6. Consistent transit providers

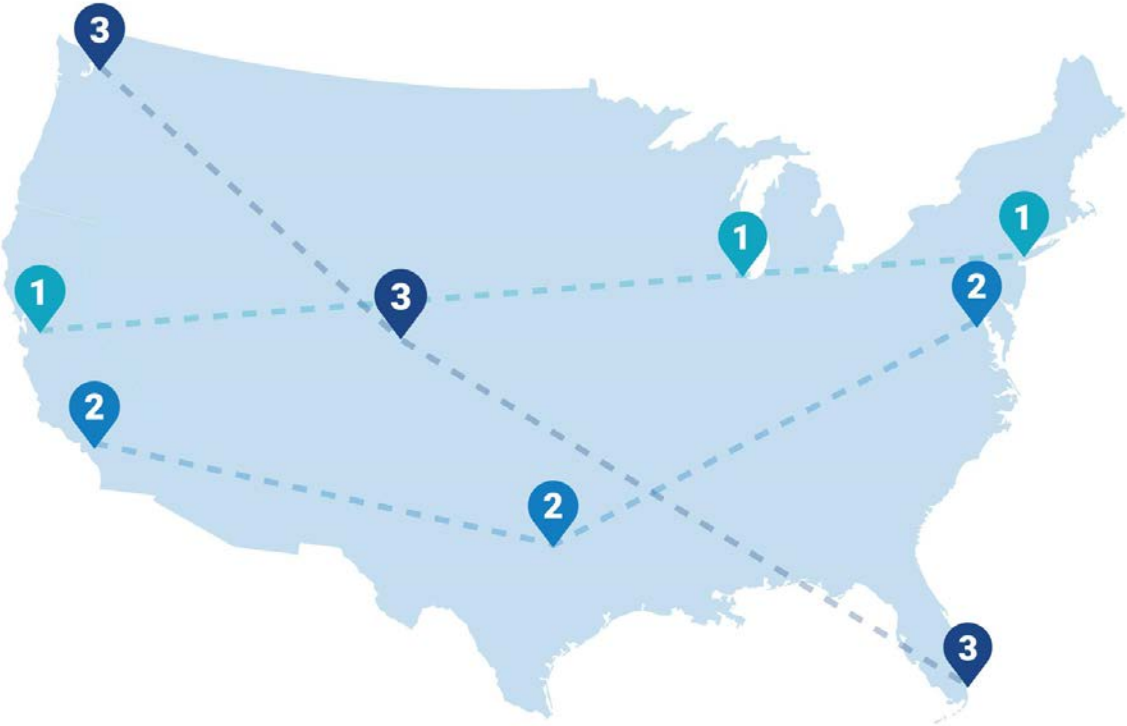
Avoid single network or vendor dependencies



	SINGLE DNS PROVIDER
GLOBAL FORTUNE 50	68%
TOP 25 SAAS PROVIDERS	44%
FTSE 100	72%

According to Thousand Eyes Global DNS performance report https://www.thousandeyes.com/resources/2018-global-dns-performance-benchmark-report

Sample anycast groups



Anycast Group #1
San Jose
Chicago
New York

Anycast Group #2
Los Angeles
Dallas
Ashburn

Anycast Group #3
Seattle
Denver
Miami

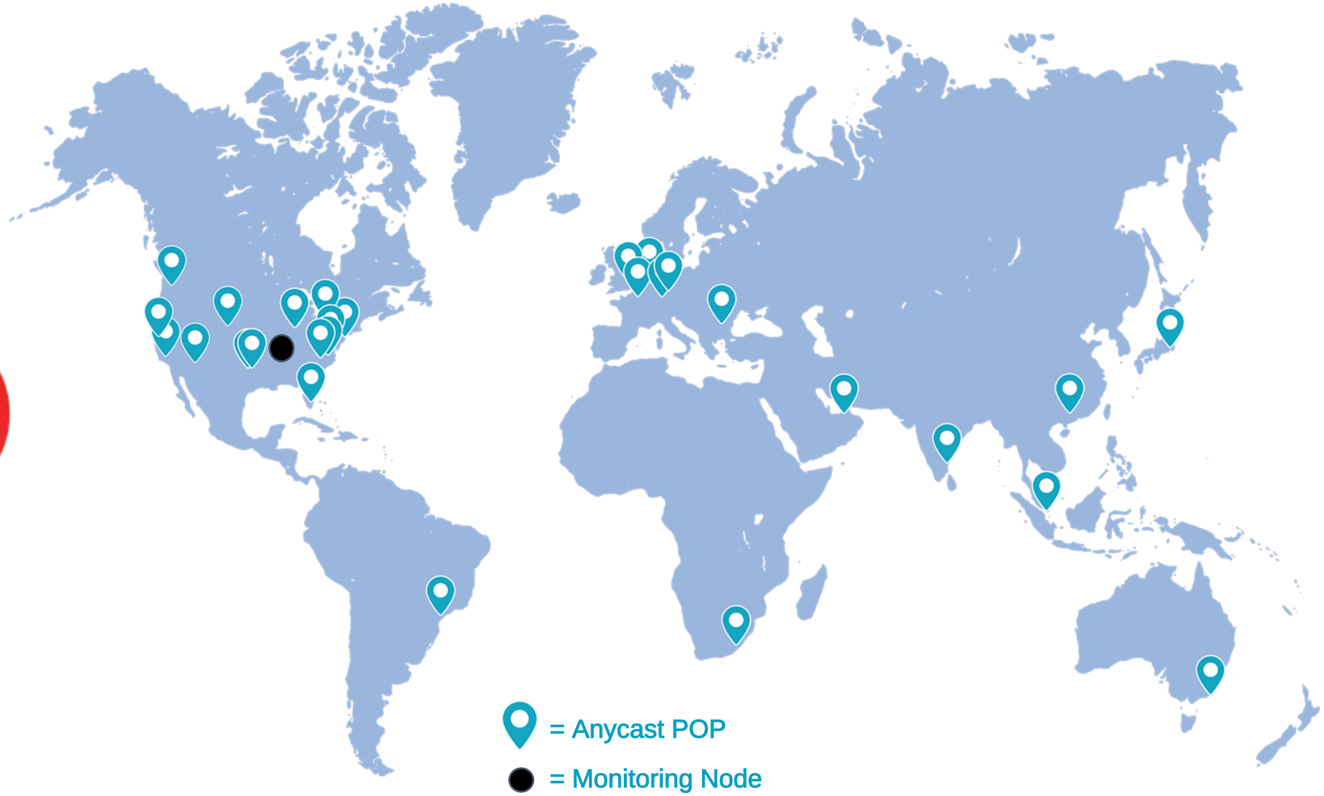
DDoS mitigation

- Have detection tools in place and automated response plan
  - NetFlow/sFlow sampling
    - Open source tools to visualize and alert
      - NfSen
      - FastNetMon
    - Commercial tools
      - Kentik
      - SolarWinds
- DDoS mitigation plan
  - Make it as automated as possible
    - E.g. pre-programmed routing rules to mitigation POPs for scrubbing
    - Run drills regularly to stress test your response

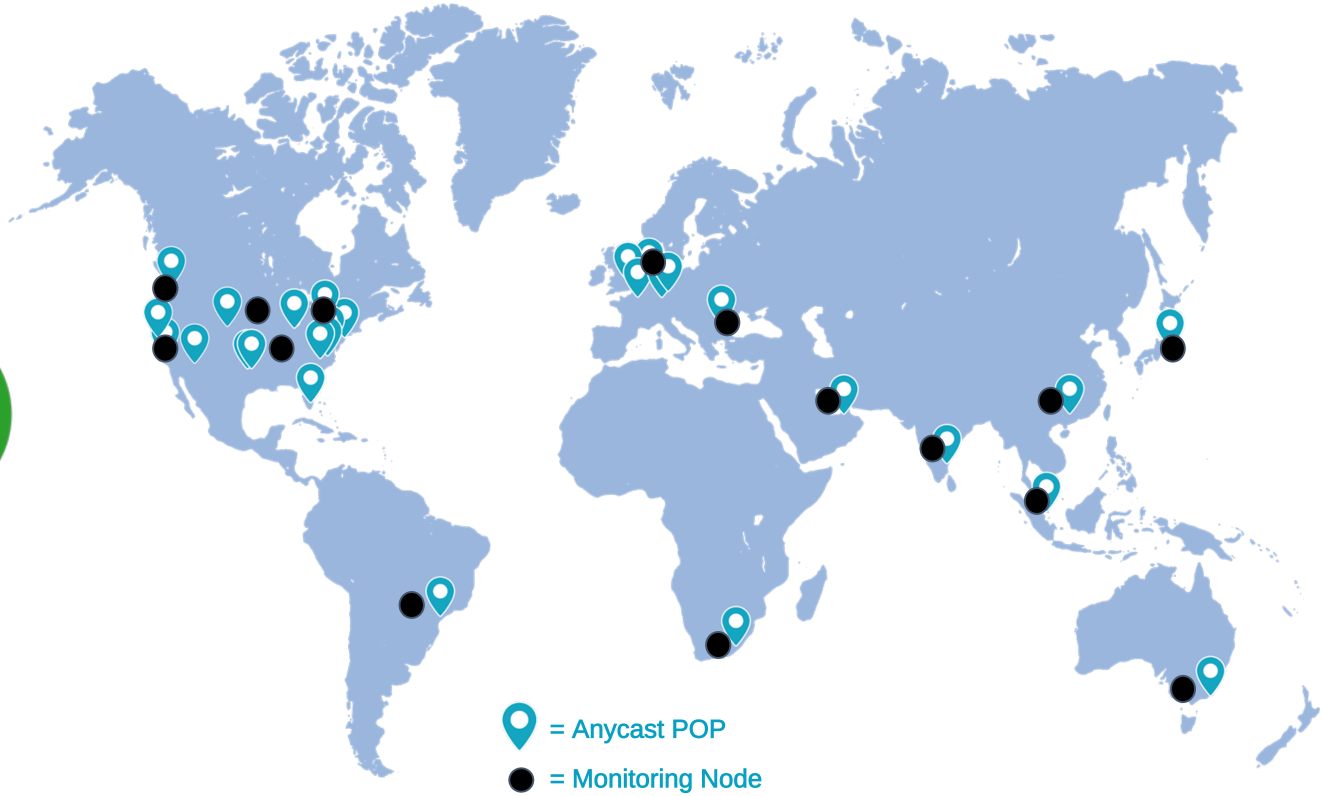
Monitoring

- Open source and commercial options
  - Commercial
    - Catchpoint, Grafana worldPing, Thousand Eyes
  - Roll your own + open source
    - RIPE Atlas probes
      - (article: https://labs.ripe.net/Members/kenneth_finnegan/measuring-anycast-dns-services-using-ripe-atlas)
    - Public cloud and VPS providers
      - Nagios, Icinga
- Monitoring probes need to be distributed to show you what end users are seeing
  - Put probes on diverse networks and on eyeball networks (RIPE Atlas is best for this)
  - Avoid putting probes on inferior networks/infrastructure (this can trigger false alerts)
  - Authoritative DNS providers should be probing popular resolvers globally (Google 8.8.8.8, Cloudflare 1.1.1.1, etc)

General network monitoring

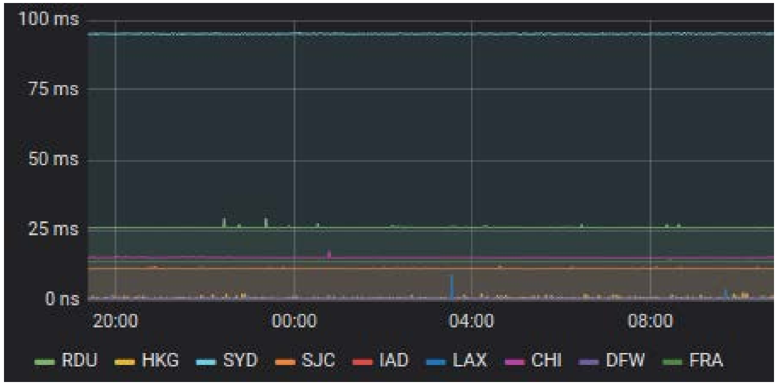


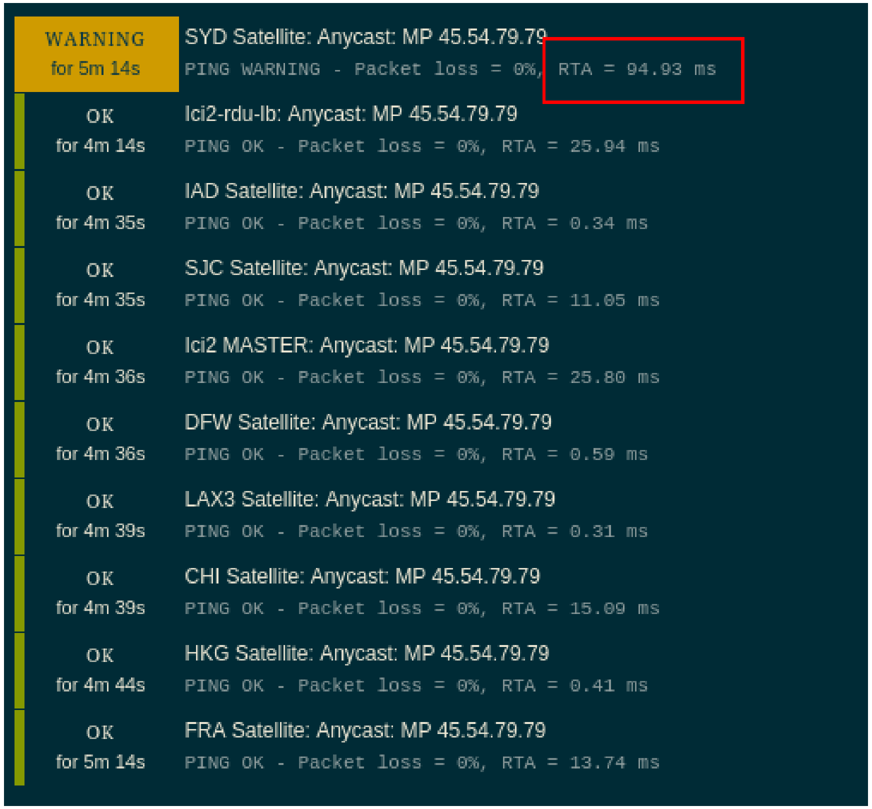
General network monitoring



Monitoring example: Icinga + satellites

Icinga is an open source distributed monitoring toolkit, example pinging an anycast IP from multiple regions





WARNING
for 5m 14s
SYD Satellite: Anycast: MP 45.54.79.79
PING WARNING - Packet loss = 0%, RTA = 94.93 ms

OK
for 4m 14s
Ici2-rdu-lb: Anycast: MP 45.54.79.79
PING OK - Packet loss = 0%, RTA = 25.94 ms

OK
for 4m 35s
IAD Satellite: Anycast: MP 45.54.79.79
PING OK - Packet loss = 0%, RTA = 0.34 ms

OK
for 4m 35s
SJC Satellite: Anycast: MP 45.54.79.79
PING OK - Packet loss = 0%, RTA = 11.05 ms

OK
for 4m 36s
Ici2 MASTER: Anycast: MP 45.54.79.79
PING OK - Packet loss = 0%, RTA = 25.80 ms

OK
for 4m 36s
DFW Satellite: Anycast: MP 45.54.79.79
PING OK - Packet loss = 0%, RTA = 0.59 ms

OK
for 4m 39s
LAX3 Satellite: Anycast: MP 45.54.79.79
PING OK - Packet loss = 0%, RTA = 0.31 ms

OK
for 4m 39s
CHI Satellite: Anycast: MP 45.54.79.79
PING OK - Packet loss = 0%, RTA = 15.09 ms

OK
for 4m 44s
HKG Satellite: Anycast: MP 45.54.79.79
PING OK - Packet loss = 0%, RTA = 0.41 ms

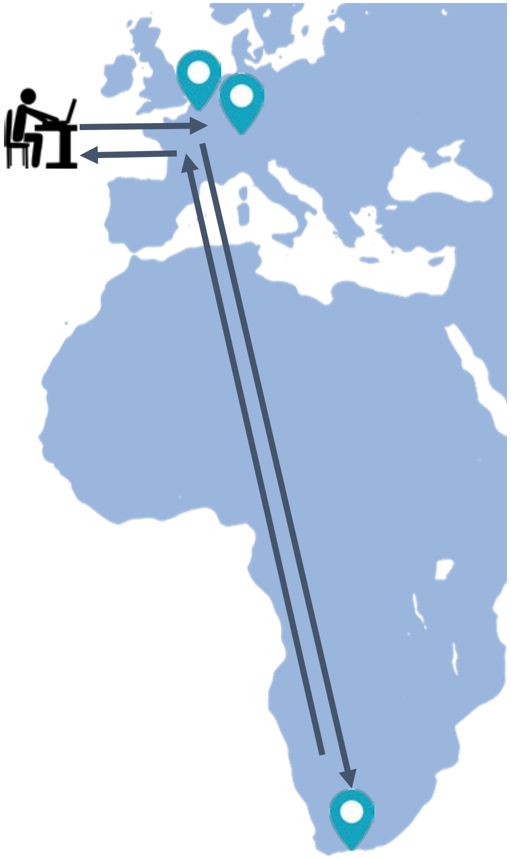
OK
for 5m 14s
FRA Satellite: Anycast: MP 45.54.79.79
PING OK - Packet loss = 0%, RTA = 13.74 ms

What’s a sinkhole? Why are they bad?

- Suboptimal routing path that can happen unintentionally when deploying Anycast across multiple geographic regions
- We often see sinkholes happening with IXes
- More peering, more problems (sometimes)

Sinkhole example

1. Users of DNSFilter.com in Belgium go on the Web



2. Users' DNS requests should be handled from DNSFilter servers in EU, they are deployed in Amsterdam, London and Frankfurt

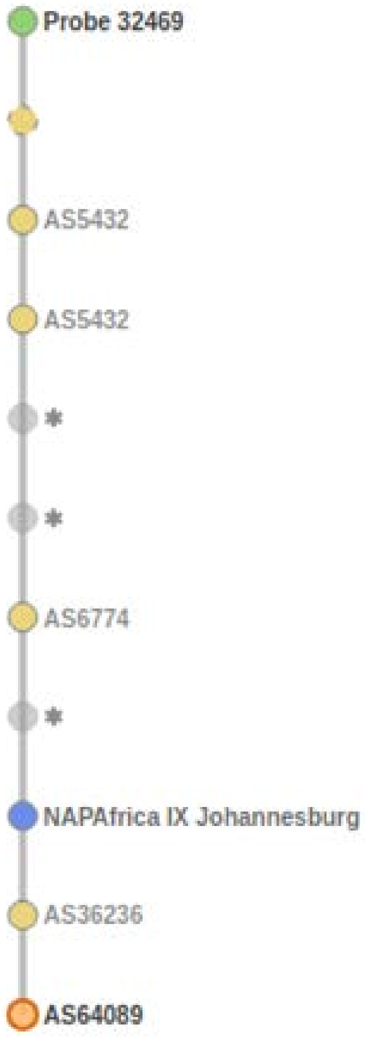
3. But, no. The traffic is sent to our Johannesburg POP

What are the facts

1. DNSFilter recently deployed to Johannesburg (JNB) for providing lower latency to users in South Africa
2. DNSFilter announced their anycast prefixes to the Internet Exchange, NAPAfrica in Johannesburg
3. Analyzed client request IPs on the JNB DNS servers and found some out-of-region client IPs
4. Testing confirmed users from Belgium were landing in JNB

AS Path: BGP is not latency or geographically aware

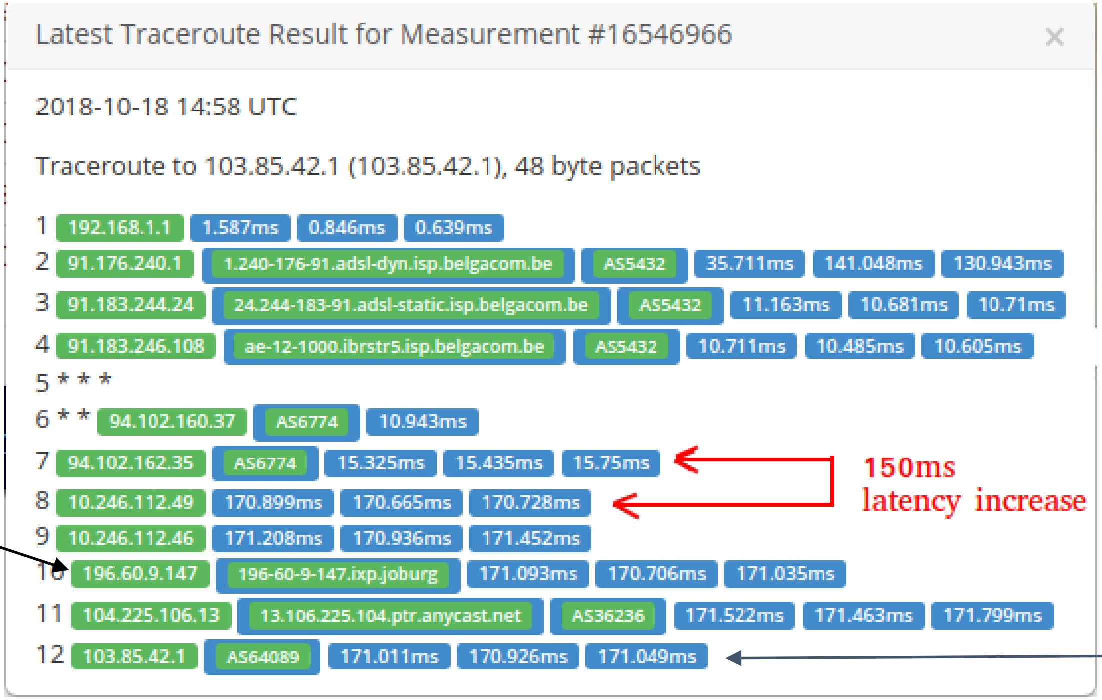
Test from RIPE Atlas using a probe in Belgium. The graph is from the TraceMON tool which shows AS hops, relatively short path of only 4 total AS numbers from client to server



Traceroutes to **103.85.42.1** from 1 of 5 probes [select] at **October 18th 2018, 15:00:00 UTC**

● Source ○ Target ● Host ● IXP ● Private IP ● No response — Connected ¦¦¦¦¦ Disconnected

Traffic from EU going to NAPAfrica IX



Latest Traceroute Result for Measurement #16546966 ×

2018-10-18 14:58 UTC

Traceroute to 103.85.42.1 (103.85.42.1), 48 byte packets

```
1 192.168.1.1 1.587ms 0.846ms 0.639ms
2 91.176.240.1 1.240-176-91.adsl-dyn.isp.belgacom.be AS5432 35.711ms 141.048ms 130.943ms
3 91.183.244.24 24.244-183-91.adsl-static.isp.belgacom.be AS5432 11.163ms 10.681ms 10.71ms
4 91.183.246.108 ae-12-1000.ibrstr5.isp.belgacom.be AS5432 10.711ms 10.485ms 10.605ms
5 * * *
6 * * 94.102.160.37 AS6774 10.943ms
7 94.102.162.35 AS6774 15.325ms 15.435ms 15.75ms
8 10.246.112.49 170.899ms 170.665ms 170.728ms
9 10.246.112.46 171.208ms 170.936ms 171.452ms
10 196.60.9.147 196-60-9-147.ixp.joburg 171.093ms 170.706ms 171.035ms
11 104.225.106.13 13.106.225.104.ptr.anycast.net AS36236 171.522ms 171.463ms 171.799ms
12 103.85.42.1 AS64089 171.011ms 170.926ms 171.049ms
```

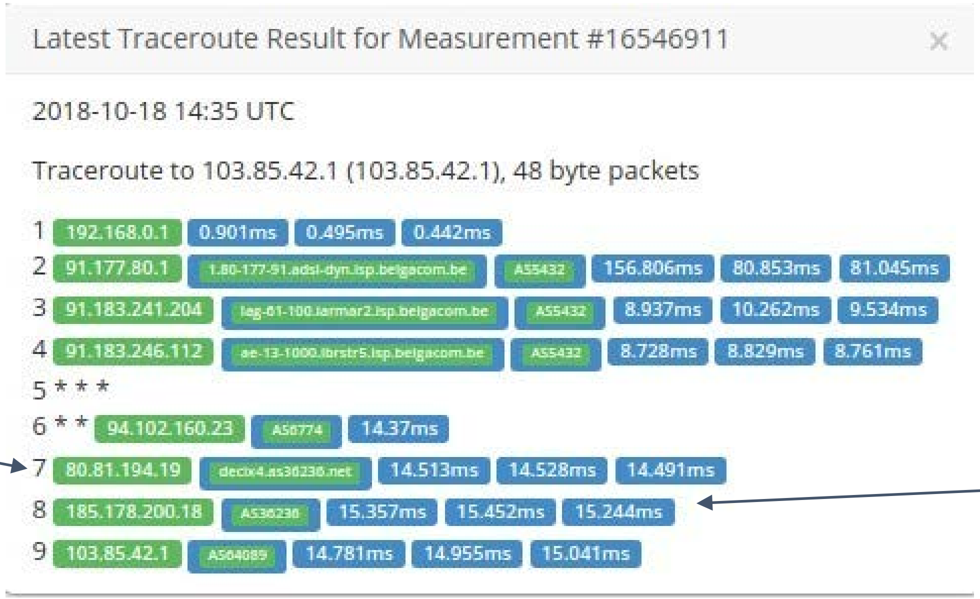
NAPAfrica peering IP

150ms latency increase

171ms RTT

Sinkhole identified and fixed.

Why? One network in EU was peering with out-of-region IX Route server but not peering with in-region IX route servers.
Traceroute looks better now after adding direct peering sessions in EU:



Latest Traceroute Result for Measurement #16546911 ×

2018-10-18 14:35 UTC

Traceroute to 103.85.42.1 (103.85.42.1), 48 byte packets

```
1 192.168.0.1 0.901ms 0.495ms 0.442ms
2 91.177.80.1 1.80-177-91.adsl-dyn.isp.belgacom.be AS5432 156.806ms 80.853ms 81.045ms
3 91.183.241.204 lag-61-100.iarmar2.isp.belgacom.be AS5432 8.937ms 10.262ms 9.534ms
4 91.183.246.112 ae-13-1000.ibrstr5.isp.belgacom.be AS5432 8.728ms 8.829ms 8.761ms
5 * * *
6 * * 94.102.160.23 AS6774 14.37ms
7 80.81.194.19 decix4.as36236.net 14.513ms 14.528ms 14.491ms
8 185.178.200.18 AS36236 15.357ms 15.452ms 15.244ms
9 103.85.42.1 AS64089 14.781ms 14.955ms 15.041ms
```

DE-CIX Frankfurt
Peer IP

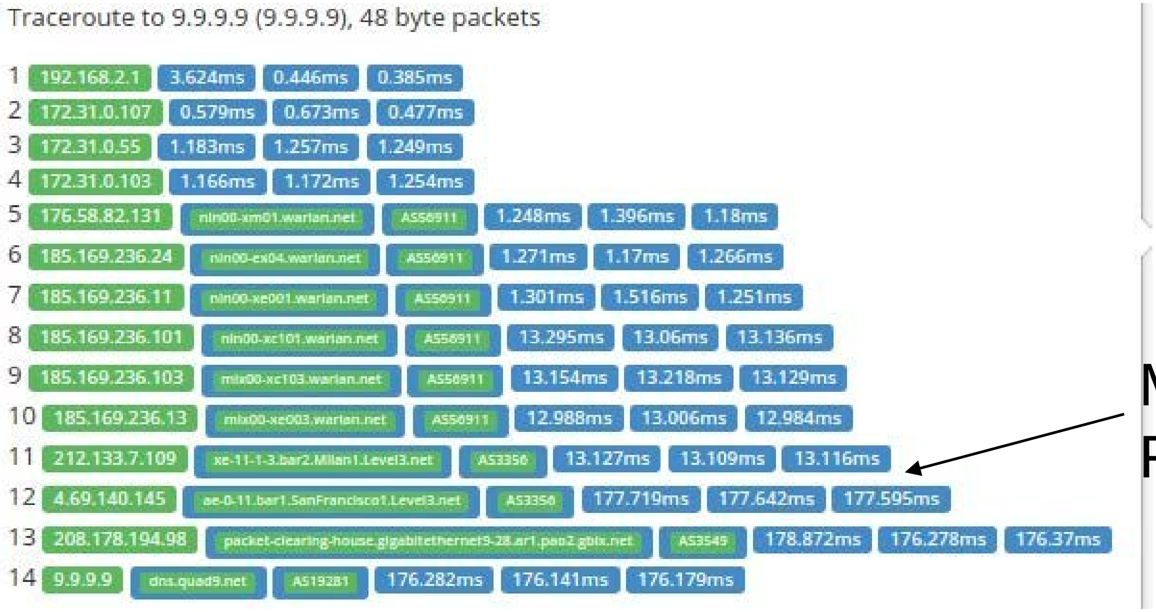
15ms
RTT

Sinkhole identification

- Perform pings from your anycast nodes back to source IPs
  - If latency is high, add to list to investigate
- For source IPs that do not respond to ping:
- Maxmind GeoLite database (free) can be used to identify likely problems to investigate further

Sinkhole Example #2: non-consistent transit

- Quad 9 (9.9.9.9) is a free recursive DNS service
- Sinkhole can happen from end-user clients to 9.9.9.9:
- They are announcing to Level3 transit in the US, but not in EU. This results in traffic hitting Level3 in EU and carried to west coast US:



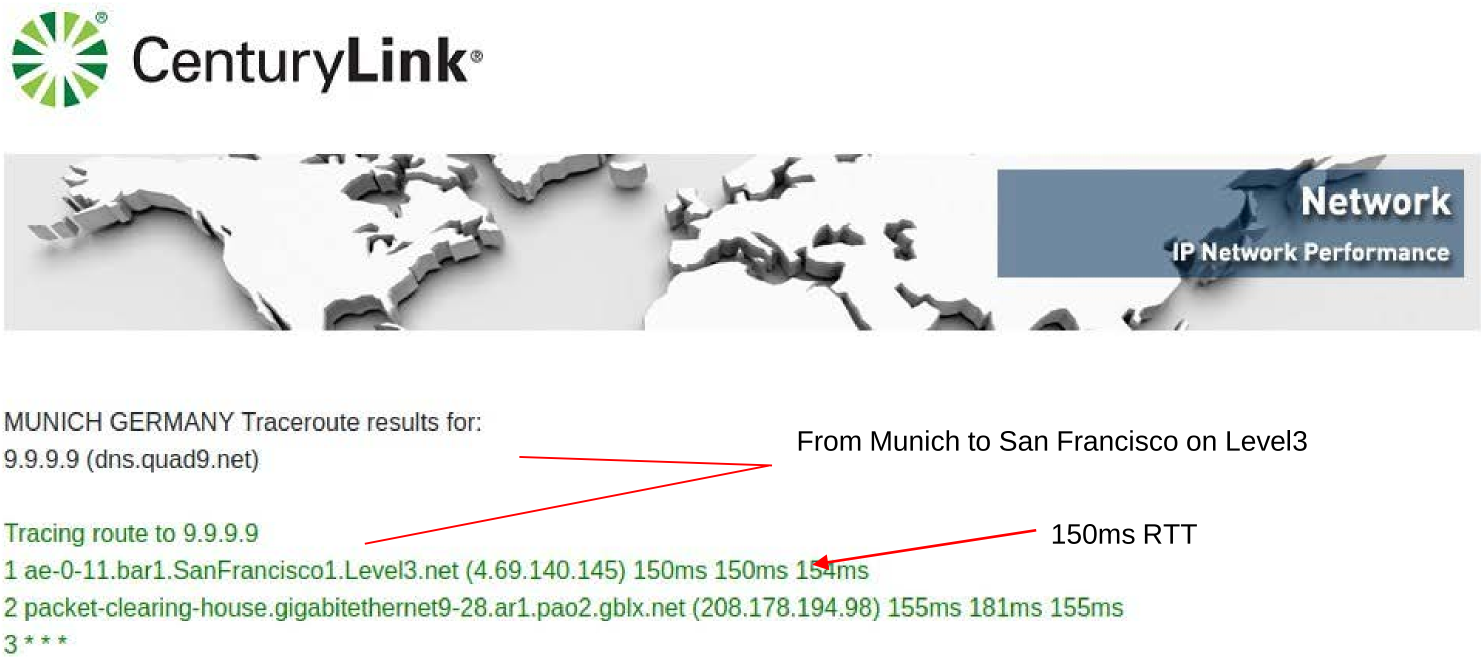
```
Traceroute to 9.9.9.9 (9.9.9.9), 48 byte packets

1  192.168.2.1      3.624ms  0.446ms  0.385ms
2  172.31.0.107     0.579ms  0.673ms  0.477ms
3  172.31.0.55      1.183ms  1.257ms  1.249ms
4  172.31.0.103     1.166ms  1.172ms  1.254ms
5  176.58.82.131    nin00-xm01.warian.net  AS56911  1.248ms  1.396ms  1.18ms
6  185.169.236.24   nin00-ex04.warian.net  AS56911  1.271ms  1.17ms  1.266ms
7  185.169.236.11   nin00-xe001.warian.net  AS56911  1.301ms  1.516ms  1.251ms
8  185.169.236.101  nin00-xc101.warian.net  AS56911  13.295ms  13.06ms  13.136ms
9  185.169.236.103  mix00-xc103.warian.net  AS56911  13.154ms  13.218ms  13.129ms
10 185.169.236.13   mix00-xe003.warian.net  AS56911  12.988ms  13.006ms  12.984ms
11 212.133.7.109    xe-11-1-3.bar2.Milan1.Level3.net  AS3356  13.127ms  13.109ms  13.116ms
12 4.69.140.145     ae-0-11.bar1.SanFrancisco1.Level3.net  AS3356  177.719ms  177.642ms  177.595ms
13 208.178.194.98   packet-clearing-house.gigabitethernet9-28.ar1.pao2.gblx.net  AS3549  178.872ms  176.278ms  176.37ms
14 9.9.9.9          dns.quad9.net  AS19281  176.282ms  176.141ms  176.179ms
```

Milan to San Francisco

Sinkhole Example #2: non-consistent transit

- Level 3 Looking Glass view



MUNICH GERMANY Traceroute results for:
9.9.9.9 (dns.quad9.net)

```
Tracing route to 9.9.9.9
1 ae-0-11.bar1.SanFrancisco1.Level3.net (4.69.140.145) 150ms 150ms 154ms
2 packet-clearing-house.gigabitethernet9-28.ar1.pao2.gblx.net (208.178.194.98) 155ms 181ms 155ms
3 * * *
```

From Munich to San Francisco on Level3

150ms RTT

Thank you!

WWW.netactuate.com

@netactuate

gwallace@netactuate.com