Measuring the End User
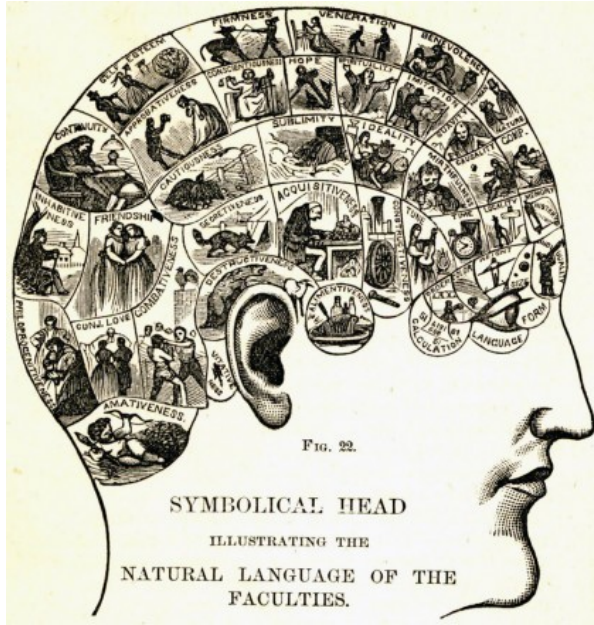
Geoff Huston
APNIC Labs

How to measure "the Internet"

What do we mean when we say
"we're measuring the Internet?"

SYMBOLICAL HEAD
ILLUSTRATING THE
NATURAL LANGUAGE OF THE
FACULTIES.

Measurement Self-Selection

When we first looked at measuring in the Internet, it was all about the network, and the distinction between network management and network measurement was not very clear

We ended up measuring what's *easy to measure* and often missed measuring what's *useful to understand*

User experience
- Responsiveness
- Sustained Throughput
- Application performance quality
- Consistency
- Availability

Network Behaviour
- Routing Stability
- Path characteristics

Element Behaviour
- Subnet characteristics
- Switch element behaviour
- Switch resource consumption

Network availability
- Element availability
- Transmission path availability
- Transmission element BER
- Network path availability

Path characteristics
- Latency
- Jitter characteristics
- Loss characteristics

Protocol behaviour
- Transport protocol behaviour
- IPv6 penetration levels
- Failover behaviour
- Connection failures

And so on...

User experience
- Responsiveness
- Sustained Throughput
- Application performance quality
- Consistency
- Availability

Network Behaviour
- Routing Stability
- Path characteristics

Element Behaviour
- Subnet characteristics
- Switch element behaviour
- Switch resource consumption

Network availability
- Element availability
- Transmission path availability
- Transmission element BER
- Network path availability

Path characteristics
- Latency characteristics
- Jitter characteristics
- Loss characteristics

Protocol behaviour
- Transport protocol behaviour
- IPv6 penetration levels
- Failover behaviour
- Connection failures

And so on...

What are you trying to measure?

Network Measurement Approaches

PING and related probe techniques
- Send an ICMP echo request to a target device and measure the time to respond
- Often used to interpret some indication of delay, loss and jitter
- BUT has little relationship to application performance, as the probe measurement is heavily impacted by the behaviour of the probe and the echo point
  - i.e. beyond being a remote device availability beacon, its of little practical use

Network Measurement Approaches

SNMP and device polling
- Per-element probe to poll various aspects of an element's current status
- Of little practical value in determining end-to-end network performance, as there is a distinct gap between end-to-end path performance and periodic polling of network element state

Network Measurement Approaches

Active Test Traffic

- – Perform a particular network transaction in a periodic fashion and correlate application performance across invocations
- – Often measures the performance limitations of the test gear and the target rather than the network
- – Tests only a small number of network transit paths
- – Provides only a weak correlation between measurement results and actual end-user experiences of application performance

What, Where and Why?

- Using a combination of active and passive measurement techniques there is a massive set of possible aspects of network behaviour that can be measured
- Few network measurements have any real bearing on the performance characteristics of applications that include some form of network interaction
  - i.e. there's a difference between measuring any old thing and measuring something relevant and useful

If you are going to measure something...
  - Know <u>why</u> you are measuring it
  - Understand the <u>limitations</u> of the measurement technique
  - Understand the limitations of any <u>interpretation</u> of the measurement
  - Understand who is the <u>consumer</u> of the measurement

IP Performance

The <u>end-to-end</u> architectural principle of IP:

- The network should not duplicate or mimic functionality that can or should be provided through end-to-end transport-level signalling
- Wired IP networks can be seen as lossy queue-controlled passive switching devices connected through fixed delay channels
- Wireless IP networks are worse!

IP parameters:

- Delay
- Delay stability
- Jitter
- Loss rate
- Loss burstiness

- In general the smaller the numbers the better, but ...

TCP Performance

- TCP performance is the interaction of concurrent end-to-end transport sessions performing a role of mutually enforced resource sharing
  - The network is not a mediator or controller of an application's resource requirements

- Its a lot like fluid dynamics:
  - Each network transport flow behaves in a fair greedy fashion, consuming as much of the network's resources as other concurrent network transport applications will permit

Observations

- Most networks are a collection of elephants and rats
    - And the elephants are uncontrollable
    - And the rats breed like crazy
- If all the traffic is long-held TCP and buffers are "right"-sized then the traffic should drive a network to oscillate around the point of packet loss
- If all the traffic is inelastic then congestion events can become catastrophic

Approaches to Measurement...

Measure the network

- – And claim that perfection is in the eye of the network management system
- – And everything else is a user problem!

This approach has its weaknesses of course

Maybe asking how to measure "network performance" is the wrong question

- How well your car operates is an **interaction** between the functions and characteristics of the car, the behavior of other cars, and the characteristics of the road

- How well an application operates across a network is also an interaction between the application and the local host and the interaction by its remote counterparts and their hosts as well as the interaction between the application's transport drivers and other concurrent applications that occur within the network

The Internet is all about users and use!

Measuring the End User

The Challenge:

How can we undertake meaningful public measurements that:

- quantify aspects of users' experiences
- drawn from across the entire Internet
- that don't rely on access to private data?

For example… IPv6

- It would be good to know how we are going with the transition to IPv6

- And it would be good everyone to know how everyone else is going with the transition to IPv6

For example… IPv6

- What **can** we measure?
  - IPv6 in the DNS – AAAA records in the Alexa top N
  - IPv6 in routing – IPv6 routing table
  - IPv6 traffic exchanges – traffic graphs
- What **should** we measure?
  - How many connected devices on today's Internet are capable of making IPv6 connections?

Approaches to Measurement...

From the inside looking out

- Set up a measurement station
    - Ping, traceroute and fetch routines
    - Measure the absolute outcomes and the variance
- Its not really a user metric
    - It's not <u>that</u> bad
    - But its not that useful either

Approaches to Measurement...

From the outside looking in:

- Set up a measurement station
  - Enrol end users to send traffic to it
  - Measure the absolute outcomes and the variance
- Or instrument your web server
- Its not a bad metric
  - But its a small sample set that is often nerd heavy
  - And nerds are "special"

Approaches to Measurement...

From the inside looking out

- RIPE Altas
  - Many thousands of end points installed in end user networks
  - Ping and traceroute to a small set of destinations
  - And report back
  - Analysing the data to produce relevant outcomes is a challenge
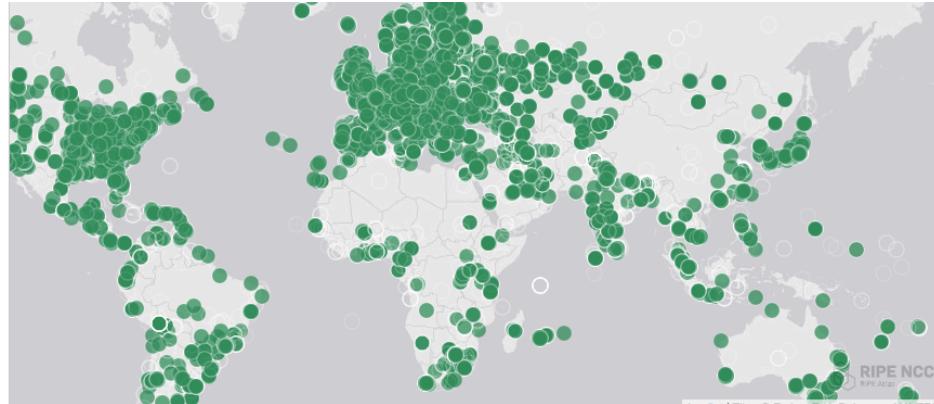
Approaches to Measurement...

From the inside looking out

- The "Sam Knows" approach
  - "be the user"
  - Use a known common platform (Open WRT DSL modem)
  - Use a common set of tests (short and large data transfer)
  - Take over the user's connection and perform the tests at regular intervals
  - Report results
- Originally developed to test DSL claims in the UK
- Used by the FCC and ISPs in the US
- Underway in Singapore and in Europe

performance bettered.

Home network | The internet

Device | Router | Last mile infrastructure | ISP network | Content servers

Internet performance
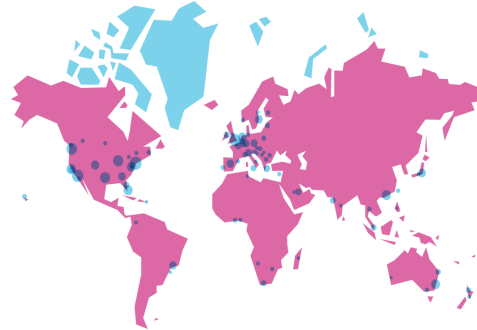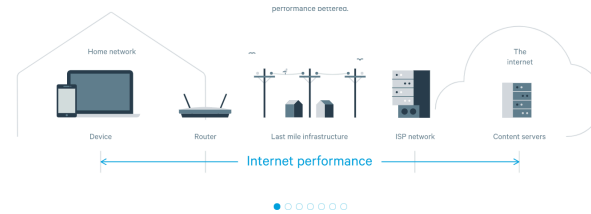
SamKnows One Platform

The SamKnows platform includes our full range of measurement agents for fixed and cellular internet. Also, our global test infrastructure and colourful, cloud-based platform: SamKnows One.

SamKnows One securely stores and visualises your performance data in real-time. We're forever developing and improving features to make complex analysis simple.

Observations

– IP performance measurement is <u>not</u> a well understood activity with mature tools and a coherent understanding of how to interpret various metrics that may be pulled out from hosts and networks

– The complex interaction of applications, host systems, protocols, network switches and transmission systems is at best only weakly understood

    – But there's a lot of slideware out there claiming to provide The Answer!
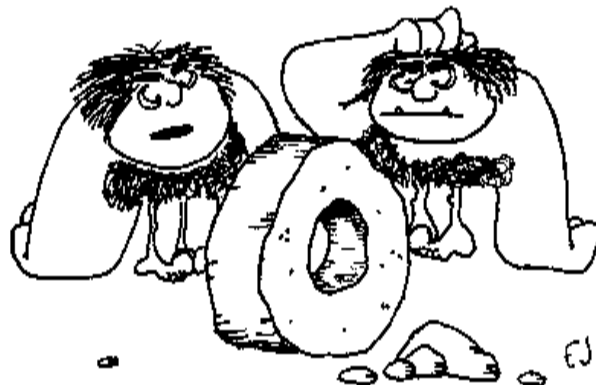
HOW DOES IT WORK AGAIN?

How to measure millions of end devices for their IPv6 capability?

How to measure millions of end devices for their IPv6 capability?

a) Be

How to measure millions of end devices for their IPv6 capability?

OR

Have your measurement code run on a million end devices

Approaches to Measurement

A case study: APNIC's approach

- we wanted to measure IPv6 deployment as seen by end users

- We wanted to say something about ALL users

- So we were looking at a way to sample end users in a random but statistically significant fashion

- We stumbled across the advertising networks...

…buy the user!

Ads are ubiquitous

Ads are ubiquitous

APNIC

Thank you for helping us measure the Internet.

APNIC

Thank you for helping us measure the Internet.

```html
<!DOCTYPE html>
<html><head data-gwd-animation-mode="quickMode"><meta name="GCD"
content="YTk3ODQ3ZWZhN2I4NzZmMzBkNTEwYjJl657daa7a9fa4c339ce298ace1f626e3e"/>

  <meta name="generator" content="Google Web Designer 1.2.1.0121">
  <meta http-equiv="Content-Type" content="text/html; charset=utf-8">
  <meta name="viewport" content="width=device-width, initial-scale=1.0">

  <script type="text/javascript" src="https://s0.2mdn.net/ads/studio/Enabler.js"></script>
  </script>
  <script type="text/javascript" src="html5ad.js">
 </script>
<body>
  <img src="apnic-logo.png" style="float:left"/>
  Thank you for helping us measure the Internet.

  <script type="text/javascript">
    runLabsTests();
  </script>
  <!-- This section contains metadata about the ad. Most importantly, the ad size. -->
  <script type="text/gwd-admetadata">
{"version":1,"type":"DoubleClick","format":"","template":"","politeload":true,"counters":[],"timers":[],"exits":
[],"creativeProperties":{"minWidth":468,"minHeight":60,"maxWidth":468,"maxHeight":60},"components":[]}</script>
</body></html>
```

The Ad Measurement Technique

Ad Server

Authoritative Name Server

End user

Web Server

The Ad Measurement Technique

Ad Server

1. Ad Impression

Authoritative Name Server

End user

Web Server

The Ad Measurement Technique

Ad Server

2. DNS resolution

DNS Resolvers

Authoritative Name Server
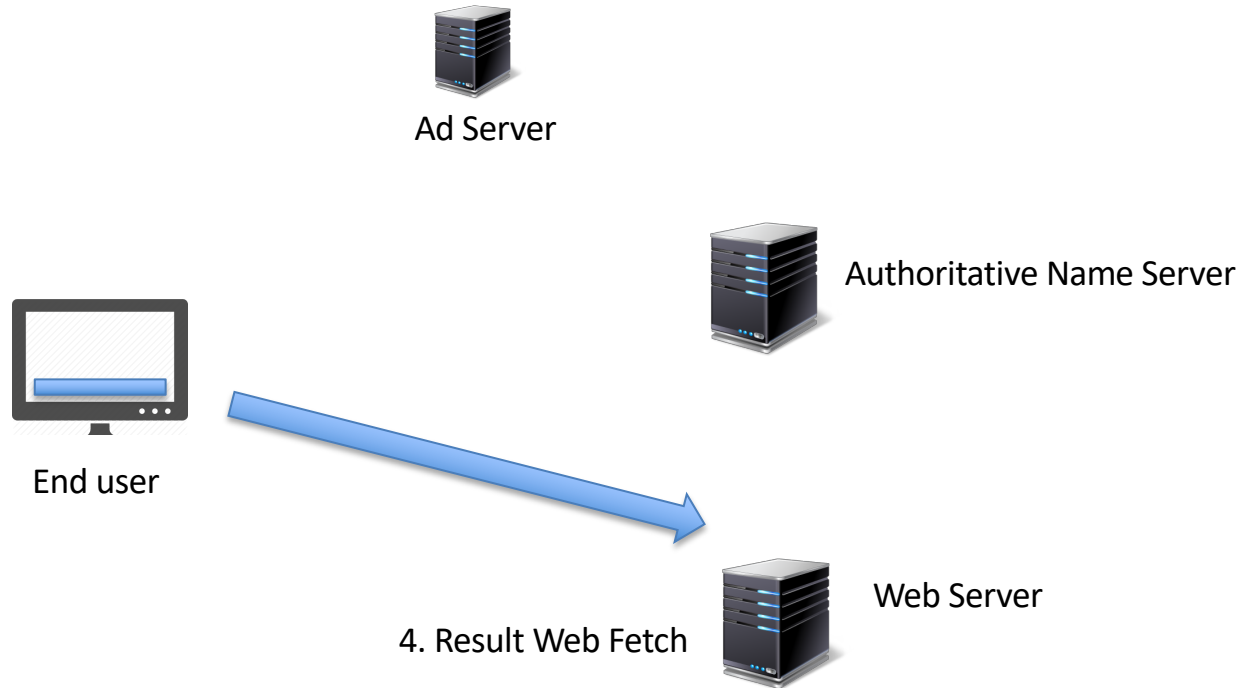
End user

Web Server

The Ad Measurement Technique

Ad Server

Authoritative Name Server

End user

Web Server

3. Web Fetch

The Ad Measurement Technique

Ad Server

Authoritative Name Server

End user

4. Result Web Fetch

Web Server
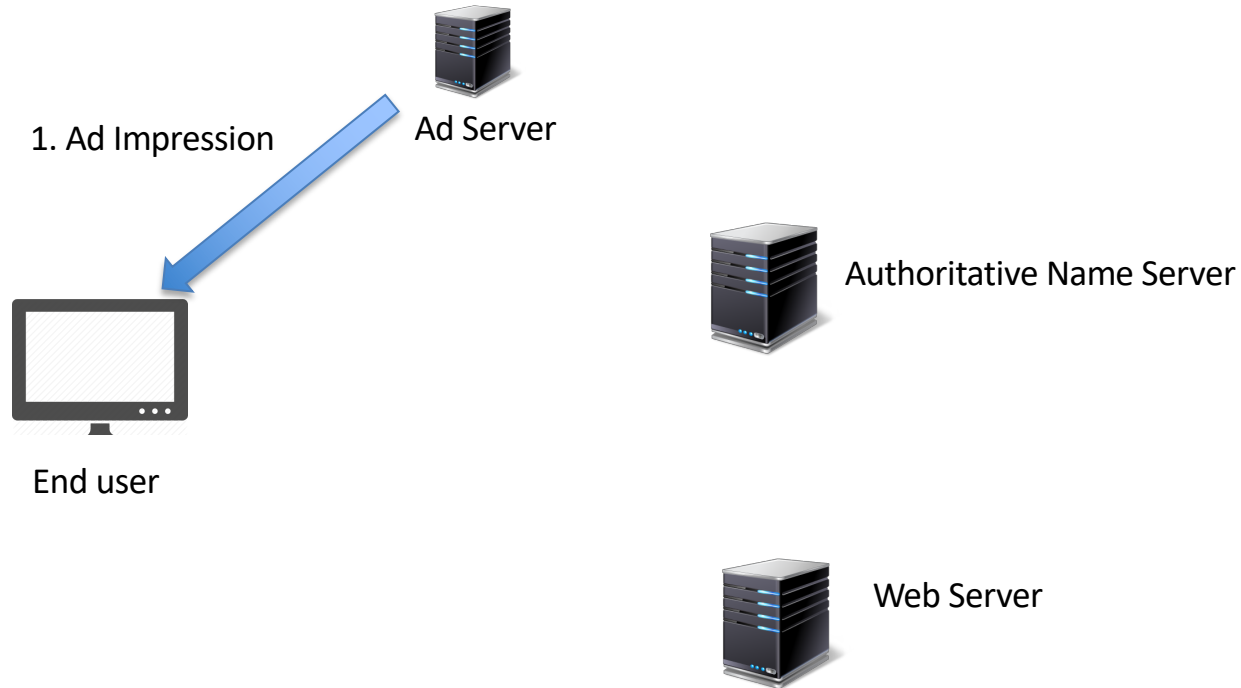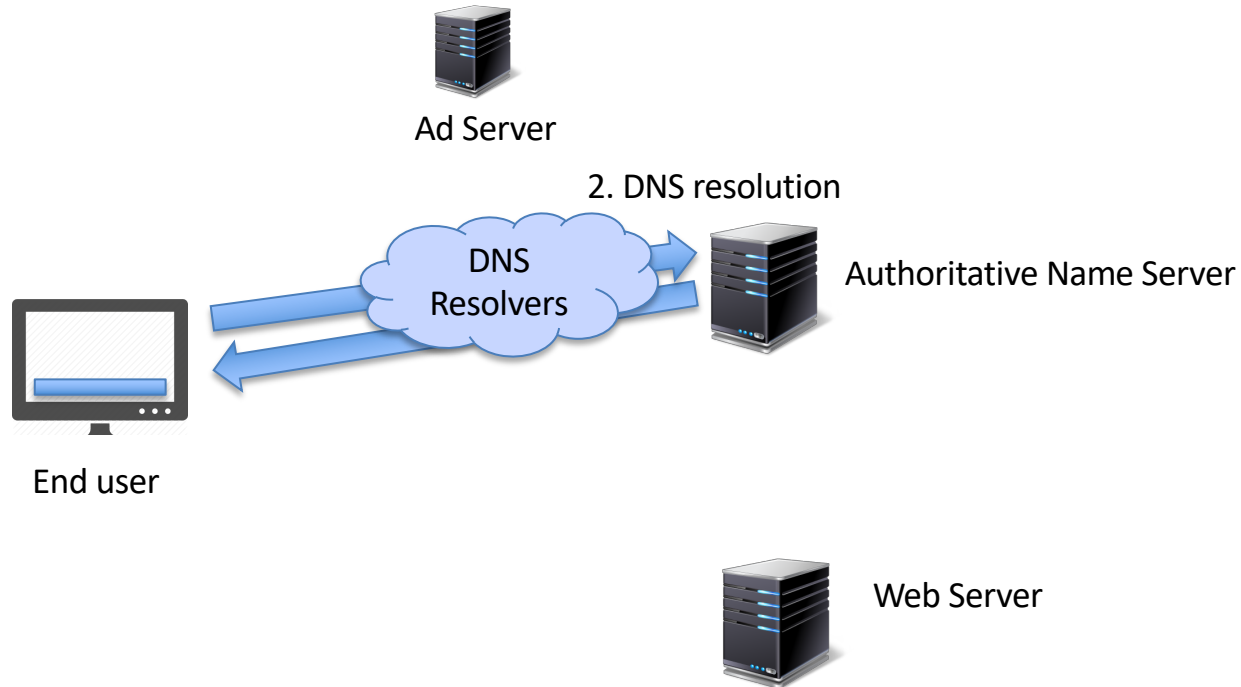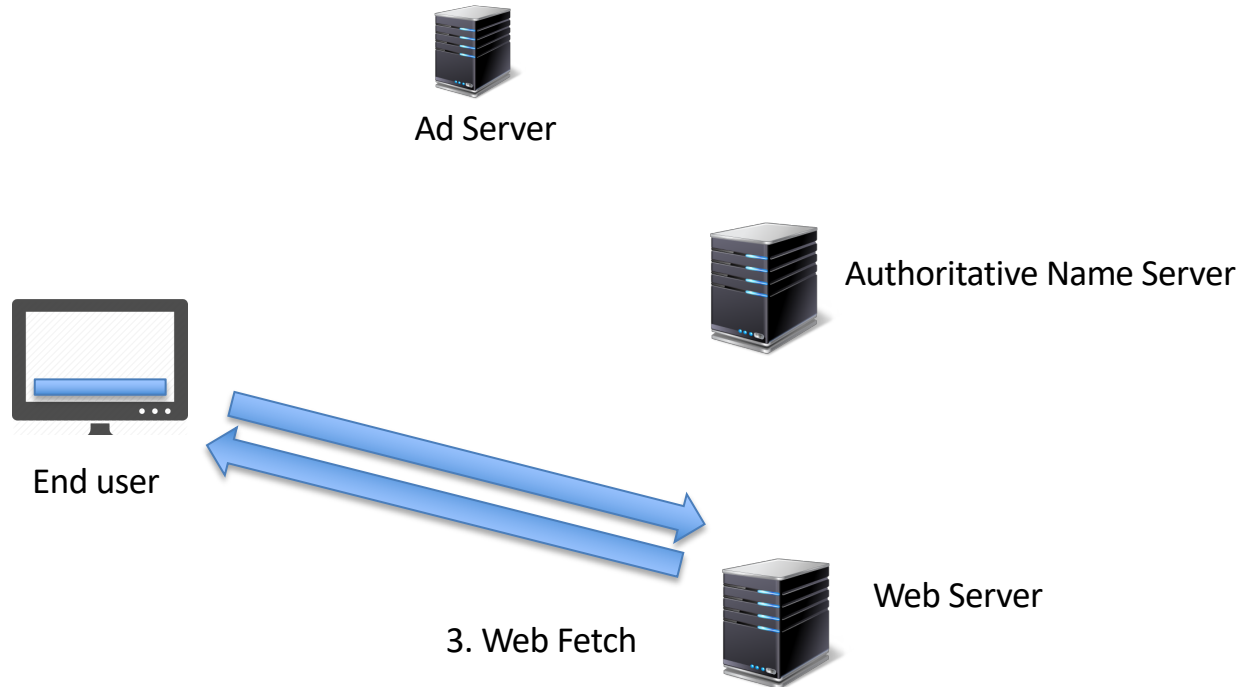
What can be scripted

- Not much:
  - http.FetchImg()
    - i.e. attempt to retrieve a URL

- But that's enough!
  - It's EXACTLY what users do!
  - A URL consists of a DNS question and an HTML question
  - What if we point both the DNS and the HTML to servers we run?
  - As long as each Ad execution uses unique names we can push the user query back to our servers

Tests

Think of a URL name as a microcoded instruction set directed to programmable DNS and HTTP servers …

http://06s-u69c5b052-c13-s1579128735-icb0a3c4c-0.ap.dotnxdomain.net/v61x1.png

Valid DNS

IPv6 access only

Valid DNSSEC signature available

User is located in Country 13 (Australia)

Time is 16 January 2020 9:52am

User's IPv4 address is 203.10.60.76

Ad Placement

At low CPM, the advertising network needs to present unique, new eyeballs to harvest impressions and take your money.
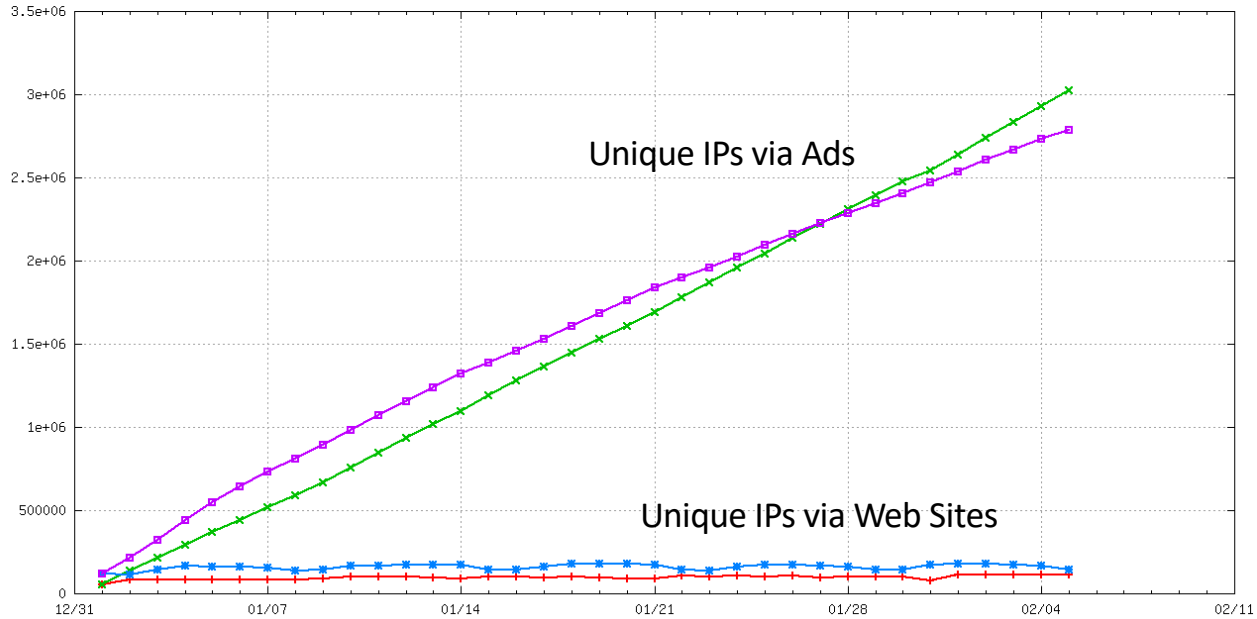
- Therefore, a 'good' advertising network provides fresh crop of unique clients per day
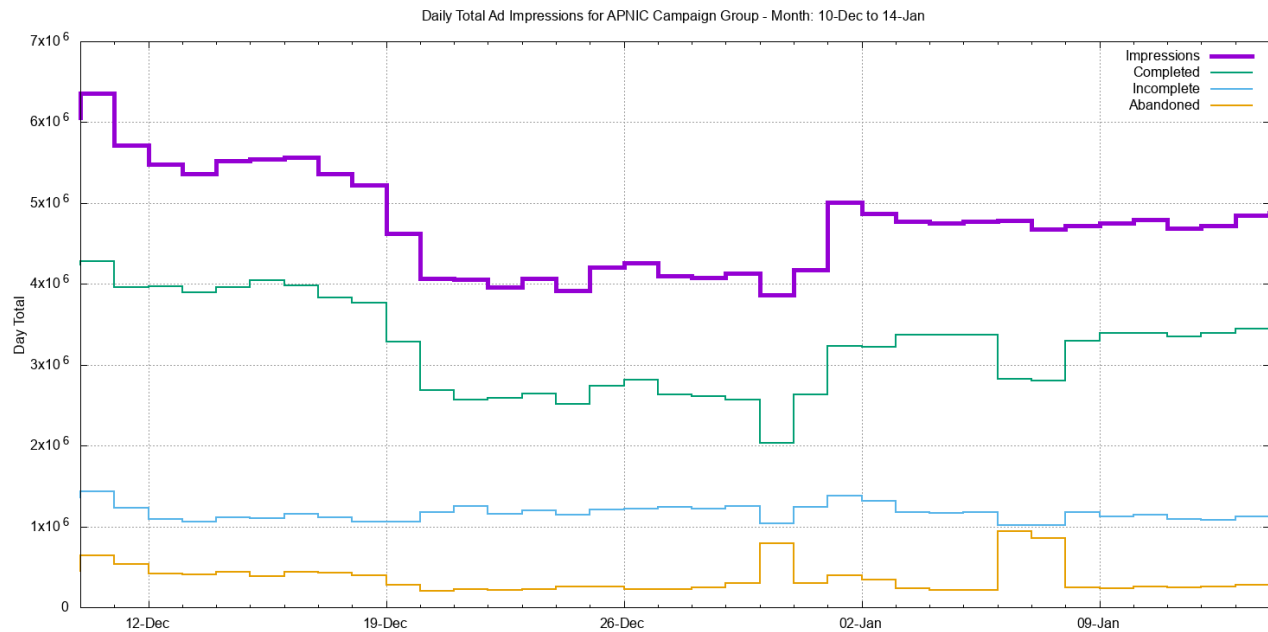
Unique IPS?

- Collect list of unique IP addresses seen
  - Per day
  - Since inception
- Plot to see behaviours of system
  - Do we see 'same eyeballs' all the time?
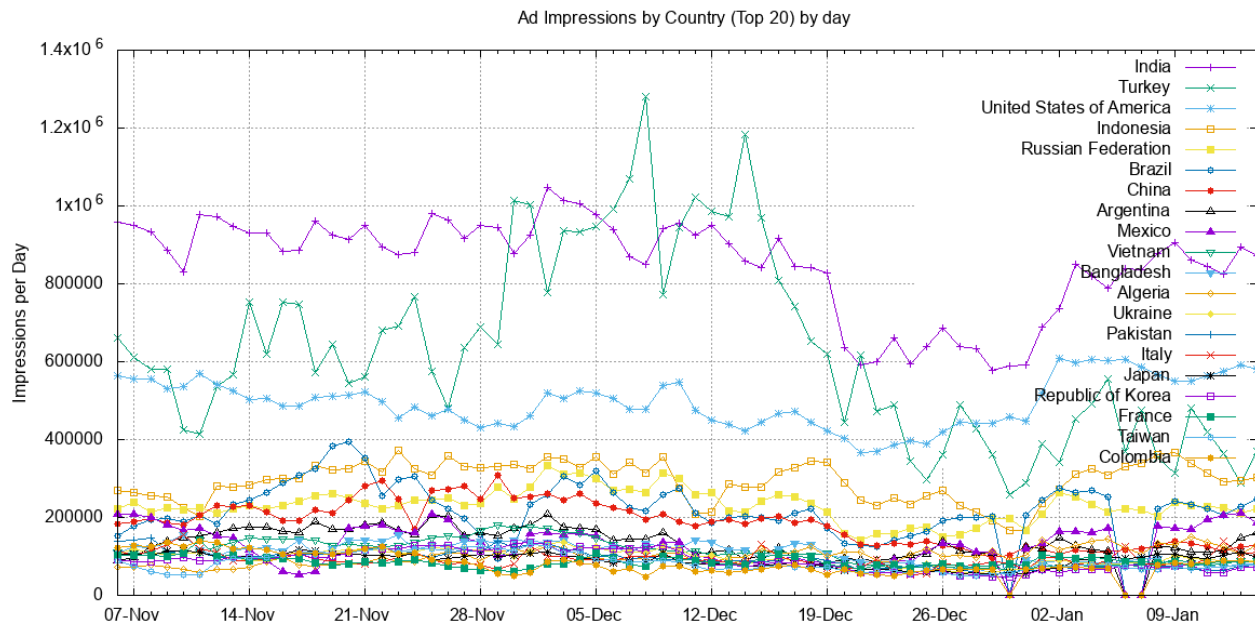
Lots of Unique IP'S

Unique IPs via Ads

Unique IPs via Web Sites

Ad Presentation Volumes

Daily Total Ad Impressions for APNIC Campaign Group - Month: 10-Dec to 14-Jan
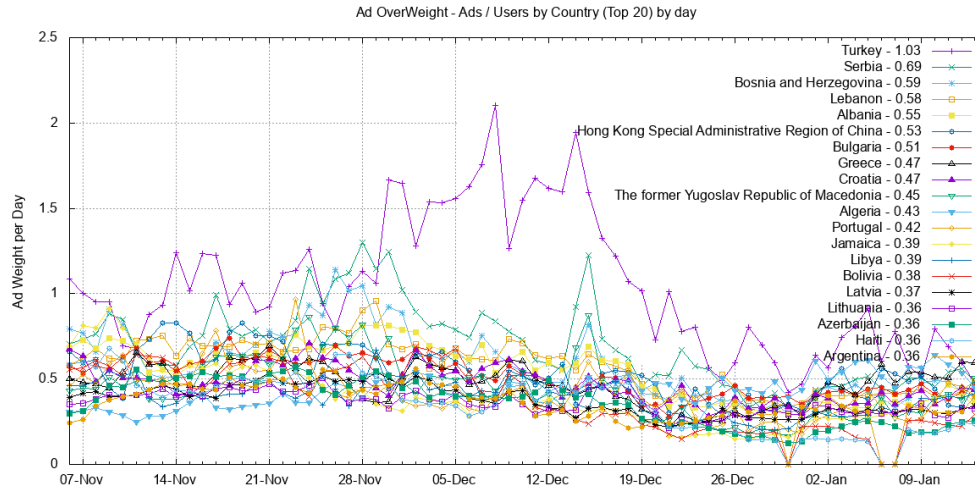
Ad Presentations: Countries


Ad Impressions by Country (Top 20) by day

Bias Compensation

- The ad presentation is NOT uniform across the Internet's user population
  - The ad machinery 'over-presents' in some countries:

Ad OverWeight - Ads / Users by Country (Top 20) by day

Bias Compensation

- The ad presentation is NOT uniform across the Internet's user population
  - The ad machinery 'under-presents' in some countries:
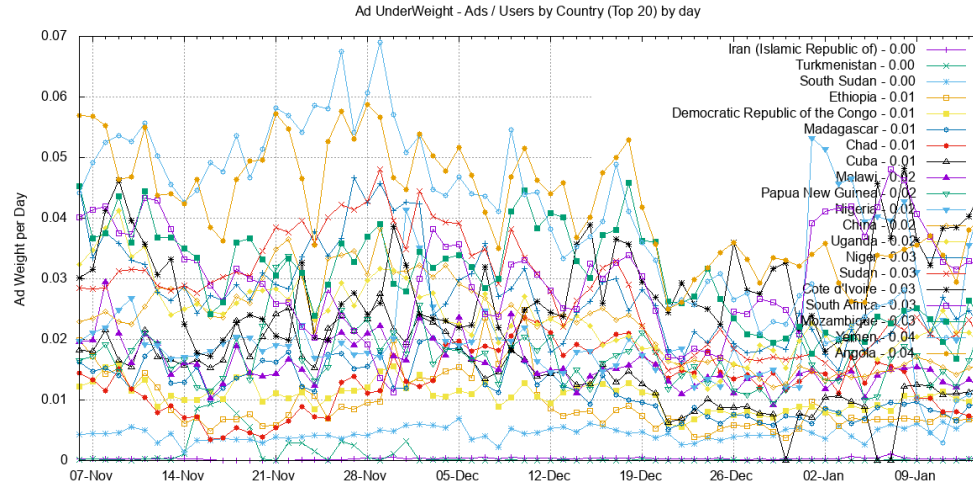
Bias Compensation

- Use ITU data on Internet users per country as the reference set, and weight the ad results to compensate for ad placement bias

| CC	Country	IPv6 Capable	IPv6 Preferred	Samples	Weight	Weighted Samples
IN	India, Southern Asia, Asia	63.64%	62.68%	1,030,515	0.97	1,003,783
BE	Belgium, Western Europe, Europe	58.89%	58.59%	13,139	1.64	21,592
US	United States of America, Northern America, Americas	55.76%	55.06%	782,992	0.78	614,124
MY	Malaysia, South-Eastern Asia, Asia	47.35%	46.80%	55,056	0.85	46,901
DE	Germany, Western Europe, Europe	46.12%	45.43%	78,484	1.95	153,196
GR	Greece, Southern Europe, Europe	45.69%	45.49%	56,917	0.27	15,228
TW	Taiwan, Eastern Asia, Asia	44.58%	43.06%	95,145	0.42	39,994

Dealing with the data

- Unified web logs, dns query logs, packet capture
- Map individual DNS and HTML transactions using a common experiment identifier
- For example:
  - DNSSEC validation implies:
    - DNS queries include EDNS(0) DNSSEC OK flag set
    - See DNS queries for DNSSEC signature records (DNSKEY / DS)
    - User fetches URL corresponding to a validly signed DNS name
    - User does not fetch URL corresponding to a in validly signed DNS name
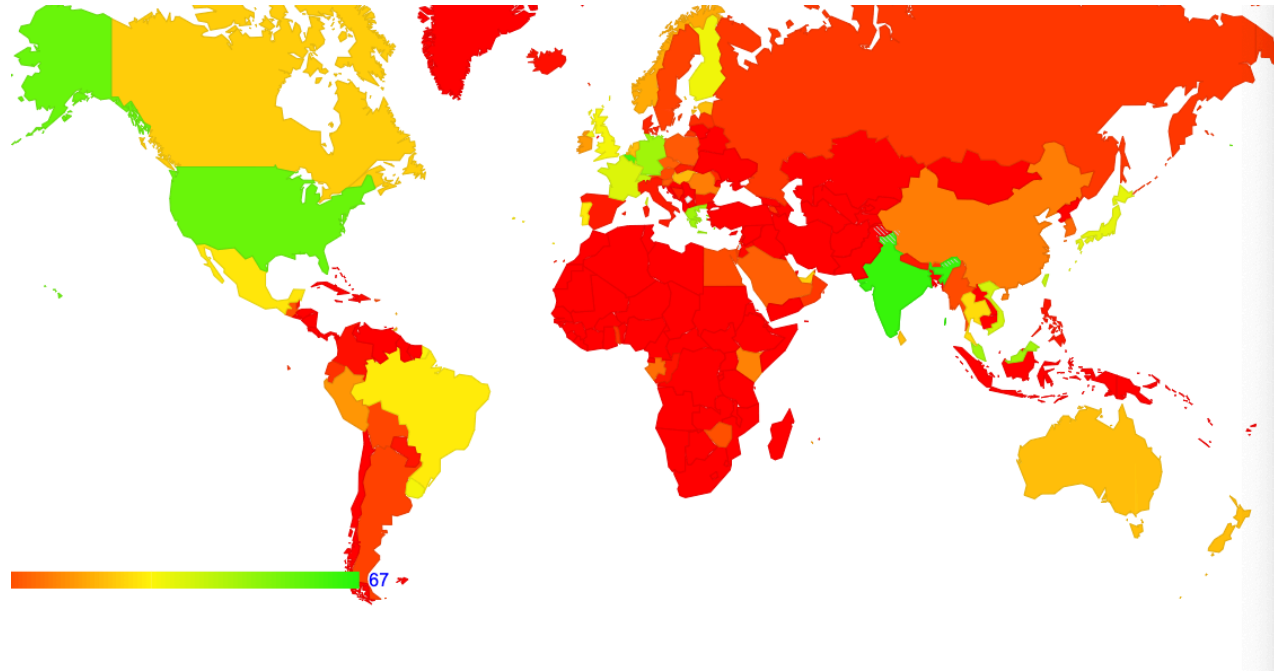
What are we measuring?

- IPv6 adoption
- IPv6 performance
- IPv6 Extension header fragility
- IPv4 and IPv4 Packet fragmentation resilience
- DNS resolver centrality
- DNS performance
- DNSSEC validation
- DNS over TCP
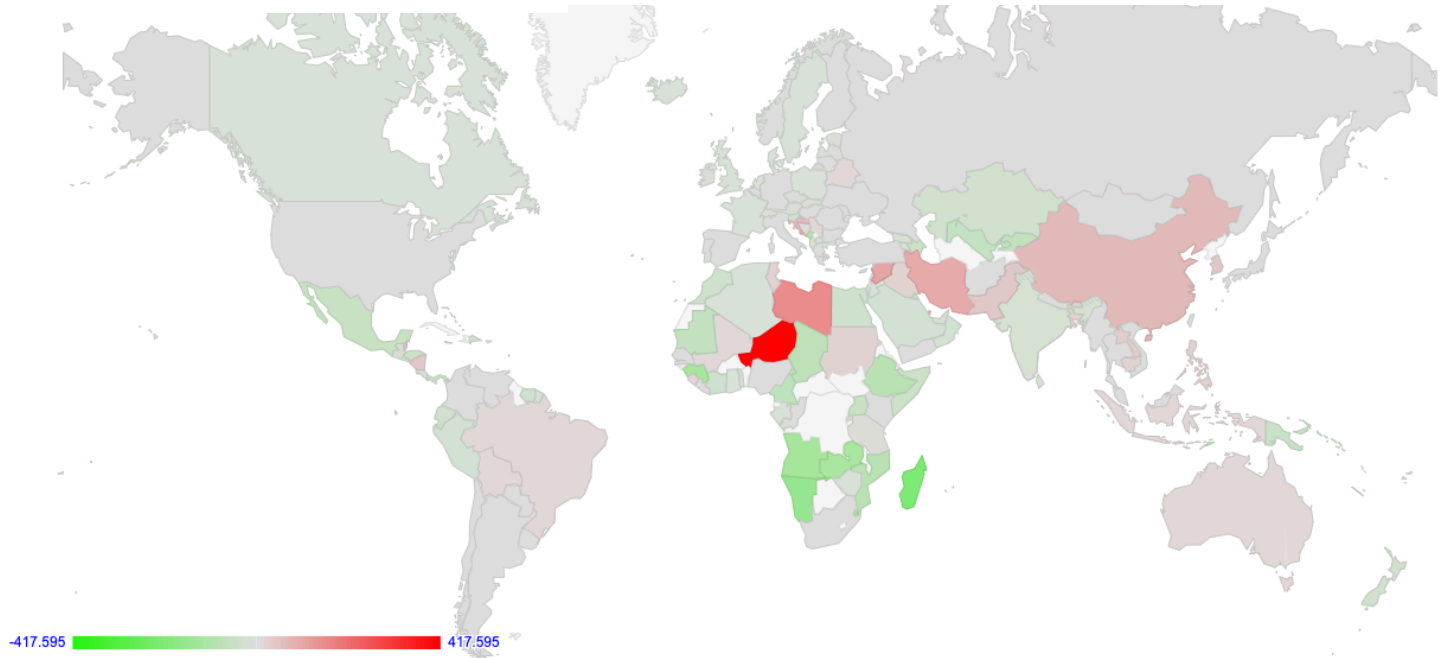- DNS NSEC caching

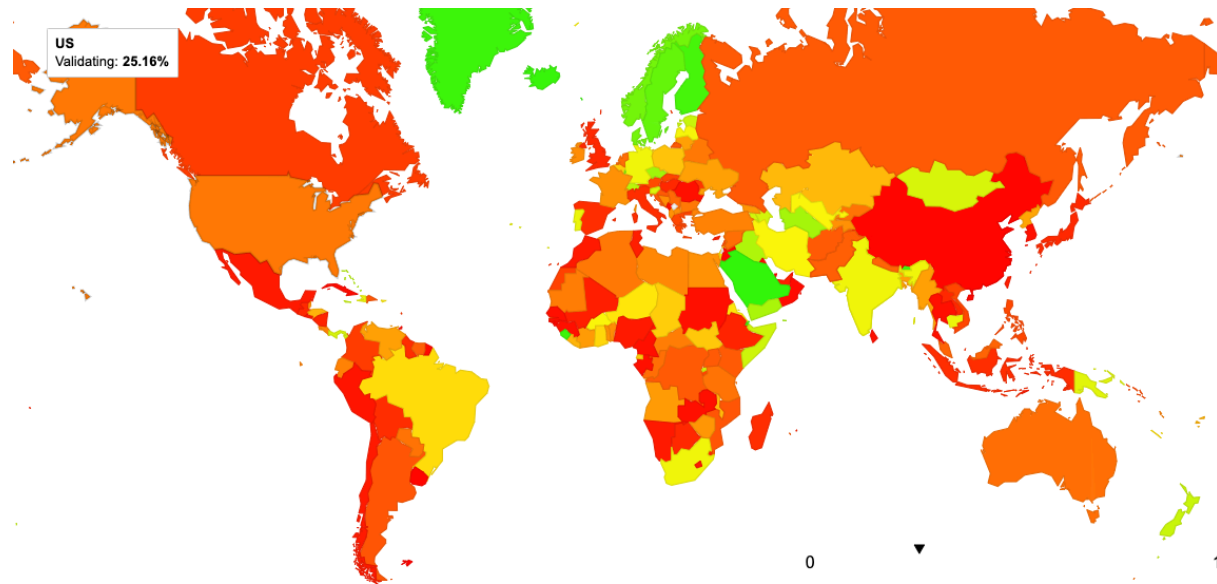# What are we seeing?
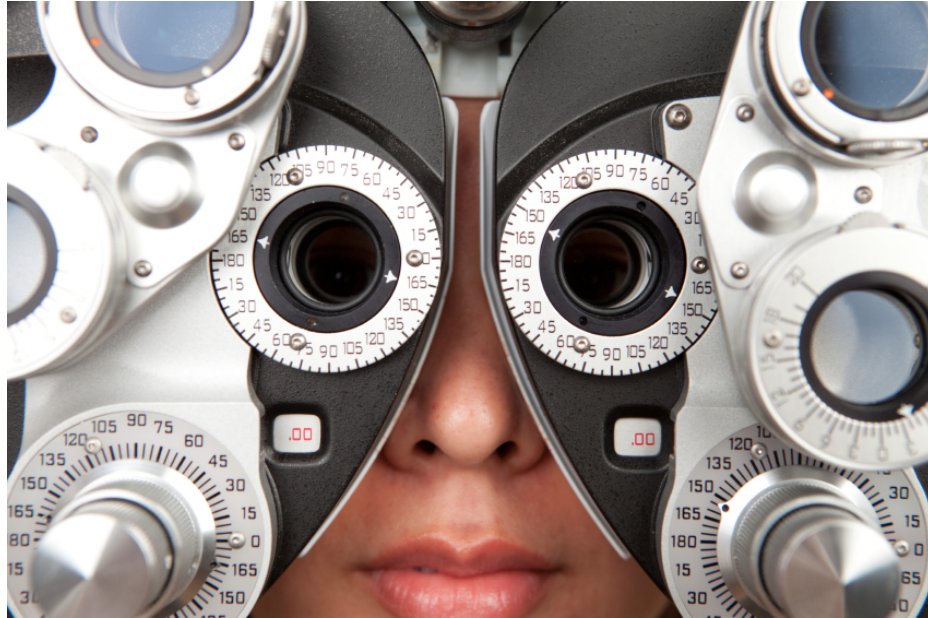
IPv6 Adoption by Country

IPv6 Performance

V6/V4 RTT Comparison by country (ms)

-417.595   417.595

DNSSEC validation by Country

US
Validating: **25.16%**

0 ▼ 1

Users per Network

Visible ASNs: Customer Populations (Est.)

| Rank	ASN	AS Name	CC	Users (est.)	% of country	% of Internet	Samples
1	AS1221	ASN-TELSTRA Telstra Corporation Ltd	AU	8,570,366	40.24	0.246	946,615
2	AS4804	MPX-AS Microplex PTY LTD	AU	4,028,750	18.92	0.116	444,984
3	AS7545	TPG-INTERNET-AP TPG Telecom Limited	AU	2,819,756	13.24	0.081	311,448
4	AS133612	VODAFONE-AS-AP Vodafone Australia Pty Ltd	AU	1,232,633	5.79	0.035	136,147
5	AS9443	INTERNETPRIMUS-AS-AP Primus Telecommunications	AU	879,476	4.13	0.025	97,140
6	AS4739	INTERNODE-AS Internode Pty Ltd	AU	528,971	2.48	0.015	58,426
7	AS135887	TELSTRA-BELONG-AP Telstra Corporation	AU	443,296	2.08	0.013	48,963
8	AS4764	WIDEBAND-AS-AP Aussie Broadband	AU	341,161	1.6	0.01	37,682
9	AS10143	EXETEL-AS-AP Exetel Pty Ltd	AU	240,294	1.13	0.007	26,541

DNS Resolver Centrality

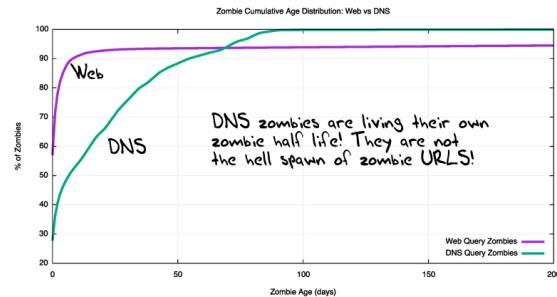

Use of DNS Resolvers for World (XA)
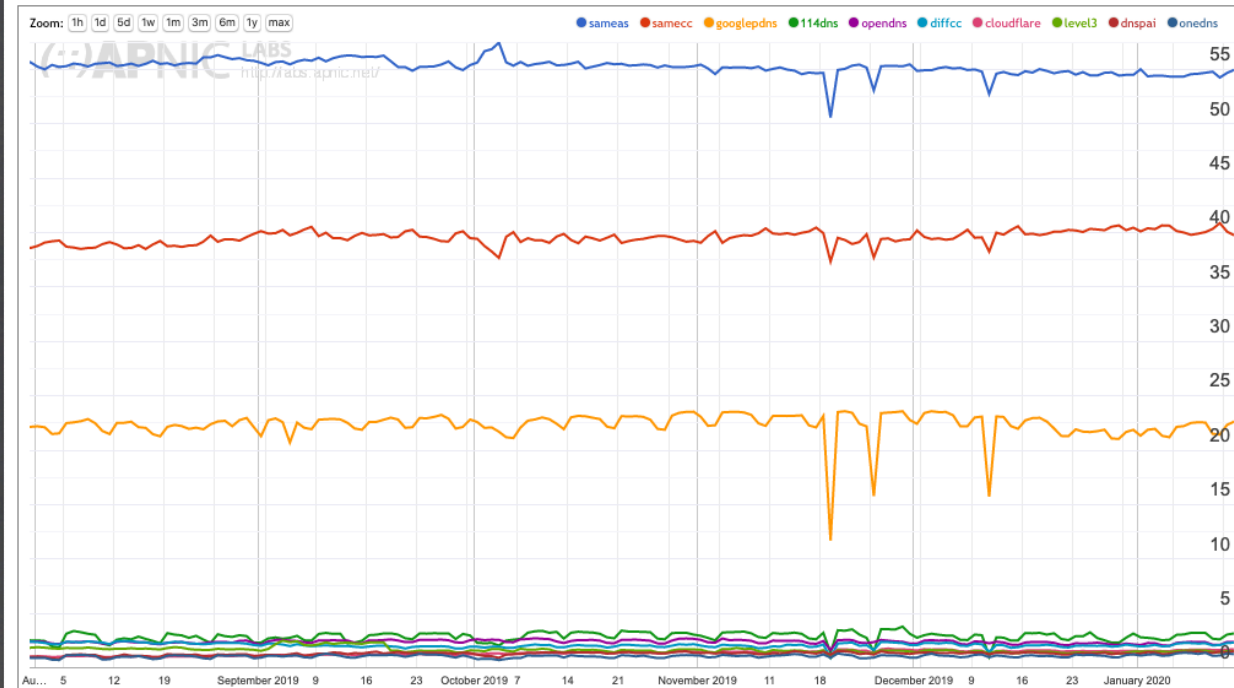
Zombies

- Remember the use of the current time in the URL?
- We can distinguish between "current" and "old" DNS and HTML queries

180 Days, All DNS Servers
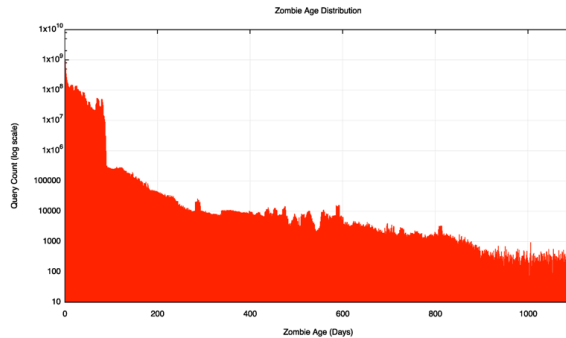
DNS vs URLs

44,733,946,408 DNS queries, of which 11,274,142,797 are zombies – a 25% zombie rating!

But...

It's not a general purpose compute platform, so it can't do many things
- Ping, traceroute, etc
- Send data to any destination
- Pull data from any destination
- Use different protocols

This is a "many-to-one" styled setup where the server instrumentation provides insight on the inferred behaviour of the edges

Measurement Ethics

- There is no user consent
- And cookies (even "don't measurement me!" cookies) are progressively being frowned upon
- Don't generate large data volumes
- Don't publish PII
- Don't use 'compromising' URL names

In Summary…

- Measuring what happens at the user level by measuring some artifact or behaviour in the infrastructure and inferring some form of user behaviour is always going to be a guess of some form

- If you really want to measure user behaviour then its useful to trigger the user to behave in the way you want to study or measure

- The technique of embedding simple test code behind ads is one way of achieving this objective
  - for certain kinds of behaviours relating to the DNS and to URL fetching

Thank You

Discussion?