APNIC 44

How we measure IPv6

Geoff Huston, Joao Damas George Michalson
APNiC

George Michaelson

Geoff Huston

Joao Damas

APNIC Labs

TAICHUNG, TAIWAN
7–14 September 2017

#apnic44

Background

- Measurement is a big topic in today’s Internet
- Reliable, unbiased, open measurements in today’s Internet are tough to find, and a number of groups are trying to fill this gap
- Much effort has been invested in techniques using “active probes” where a pool of special purpose devices can be collectively programmed to perform custom measurements (Atlas, PlanetLabs, Archipelago, …)
  - All these systems use a relatively small number of relatively constant measurement points, but they are programmable to probe and measure against any target or targets
  - **“A small number of highly agile measurement agents”**

Background

- At APNIC we’ve used precisely the opposite approach:
  - **“An extremely large number of relatively inflexible co-opted measurement agents”**

Objectives

- Enlist a massive (100’s millions) number of measurement endpoints, and operate the system at a scale of millions of individual experiment sets every day, continually enrolling new measurement endpoints
- Perform measurements by doing exactly what users do all the time:
  - Resolve DNS names
  - Fetch objects references by URLs
- And measure how well users perform in these tasks

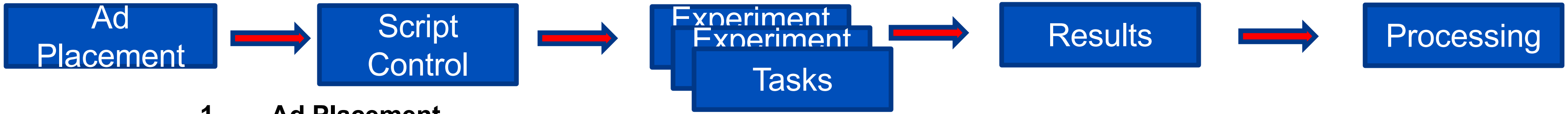
How?

By using online ad campaigns

- Ads are almost ubiquitous across the “human use” Internet
- Ads have script elements that are executed when the ad is passed into the display device
- Ad scripts have limited functionality, but if we send all of the generated DNS and Web requests to servers that we operate then we can control how the service is managed, and we can measure the agent’s performance on the server’s side of the interaction

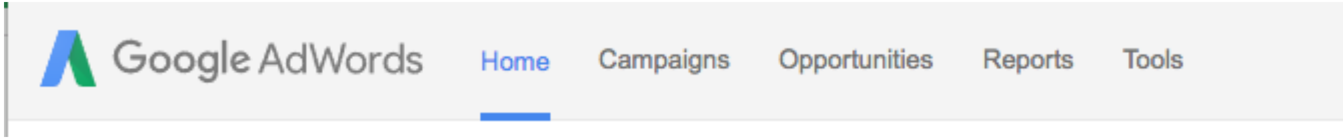
Lets look in detail at some aspects of the design of this measurement system

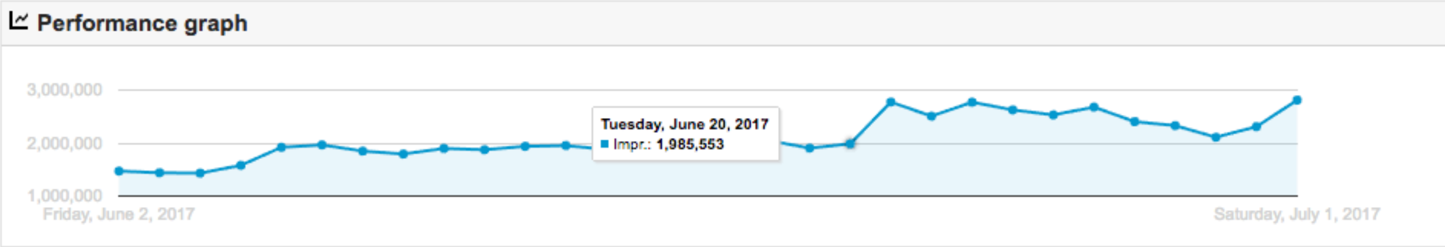
The phases of a measurement experiment



1. **Ad Placement**
   - Controlled by Google, under a campaign model
   - Client (APNIC) preferencing of impressions over clicks
   - Client (APNIC) placement directive of all device types (includes mobile/cellular)
2. **Script Control**
   - APNIC response supplies list of tasks, assigns unique ID to the experiment
3. **Tasks**
   - Name Resolution and Web retrieval of the task
   - Order not guaranteed
   - May complete before experiment timer, or after
4. **Results**
   - Summary of task results seen before 10 second experiment timer has elapsed
   - We still see task completion events after timer expiration
5. **Processing**
   - Gathering of task results and server-side packet captures, and analysis of data

1. Advertisement Placement



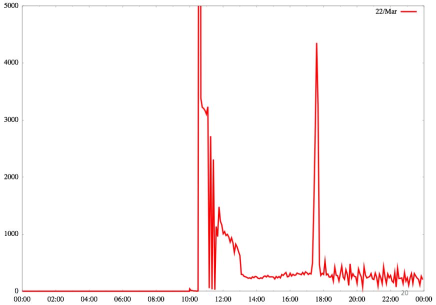


Ad Placement “Learning”

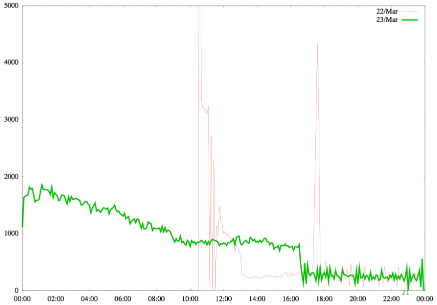
- Advertisement placement is based on a default 24 hour cycle, but is not uniformly distributed over that cycle
- The placement system attempts to soak up the advertiser’s available budget within 24 hours by performing the bulk of the placements within 20 hours, leaving 4 hours to ensure that the budget is met
- The placement ‘learns’ over a period of weeks to adapt the ad placement rate the the impression and click rate targets

Ad Placement “Learning”

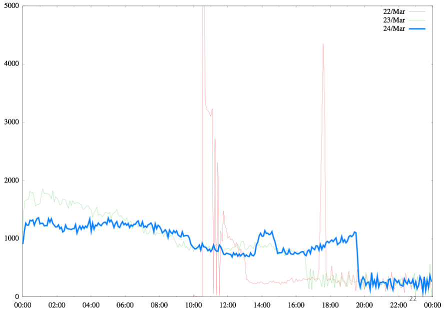
Ad Placement Training – Day 1



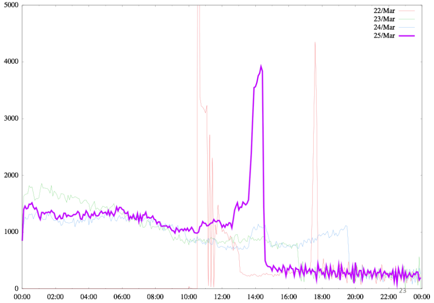
Ad Placement Training – Day 2



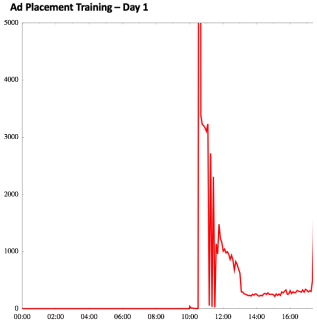
Ad Placement Training – Day 3

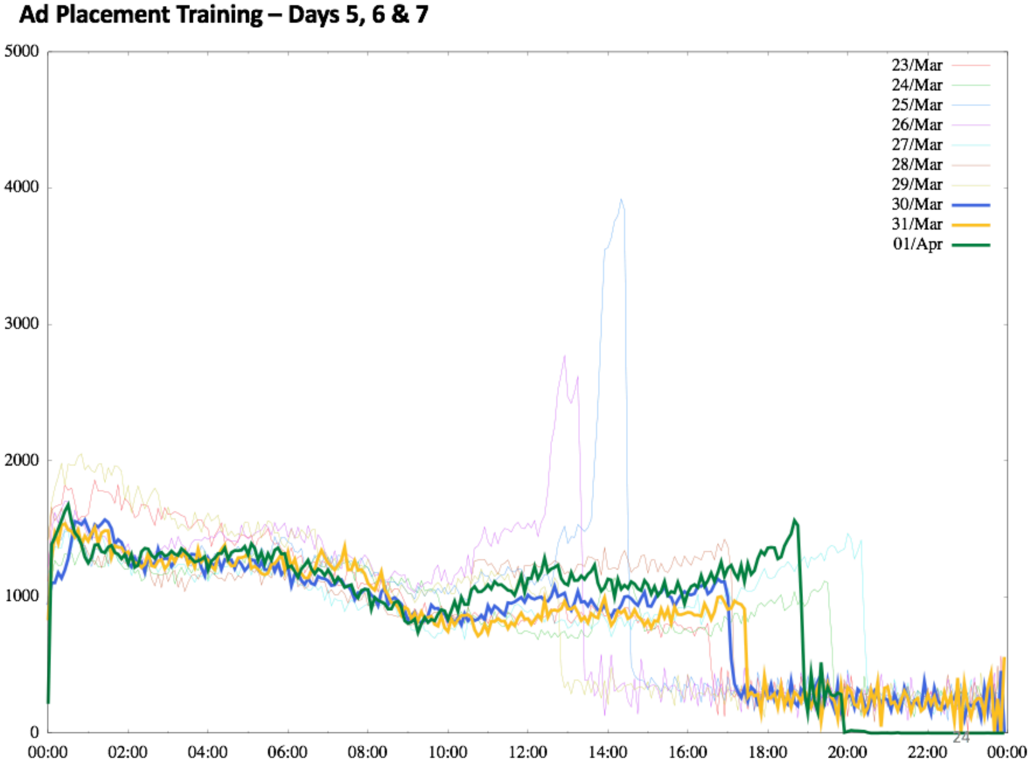


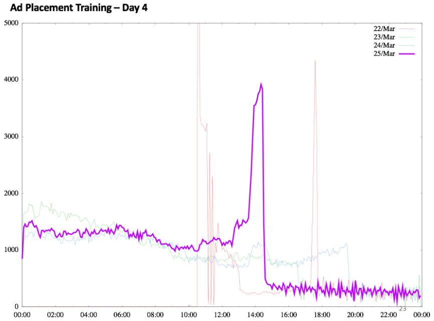
Ad Placement Training – Day 4



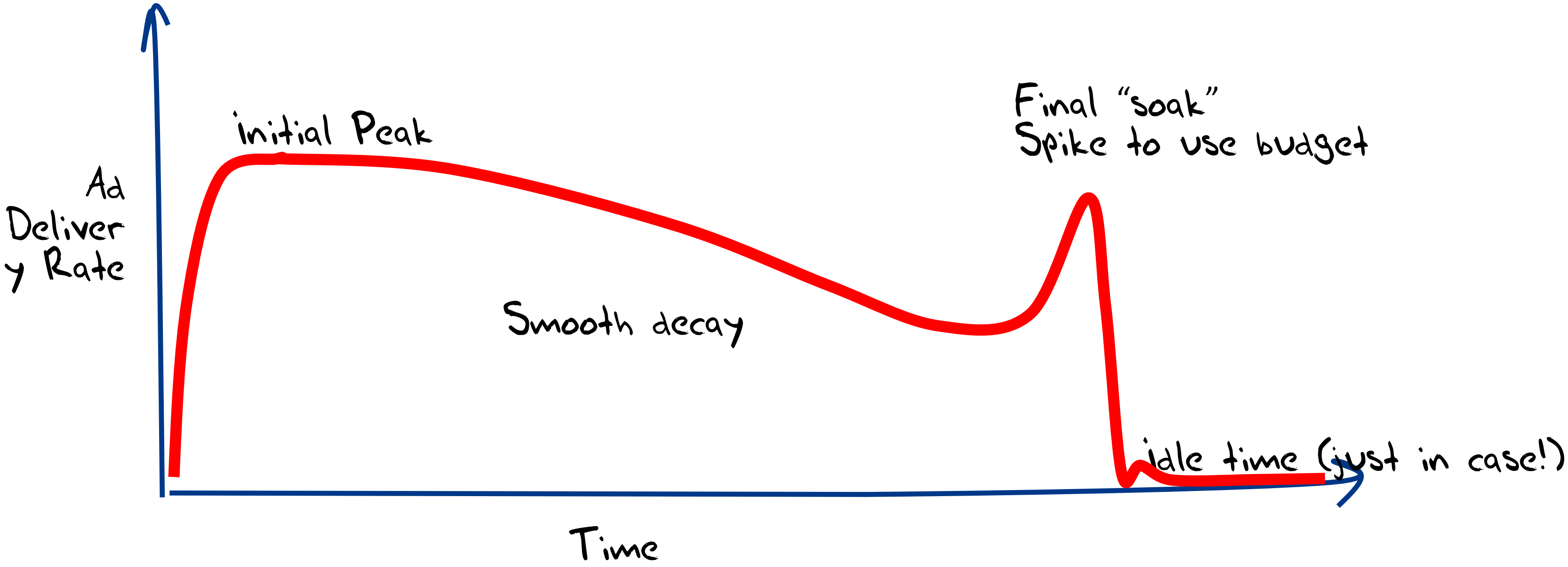
Ad Placement “Learning”


Ad Placement Training – Day 1


Ad Placement Training – Days 5, 6 & 7


Ad Placement Training – Day 4

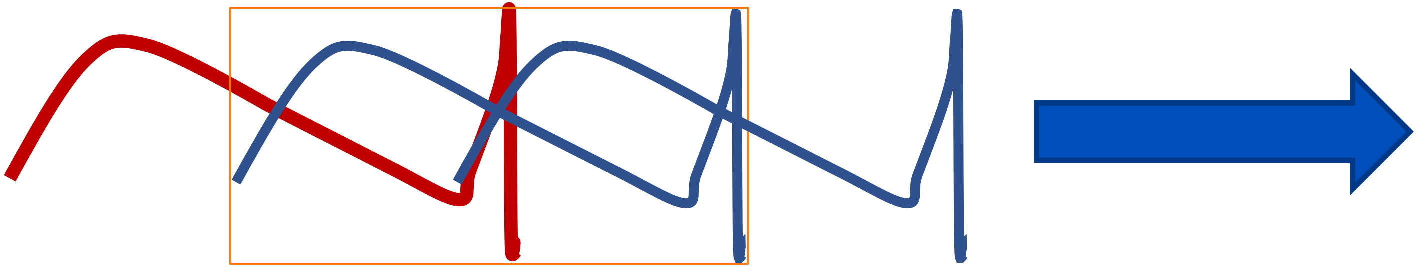
Ad Campaign 24 hour life-cycle



Advertisement Placement Controls

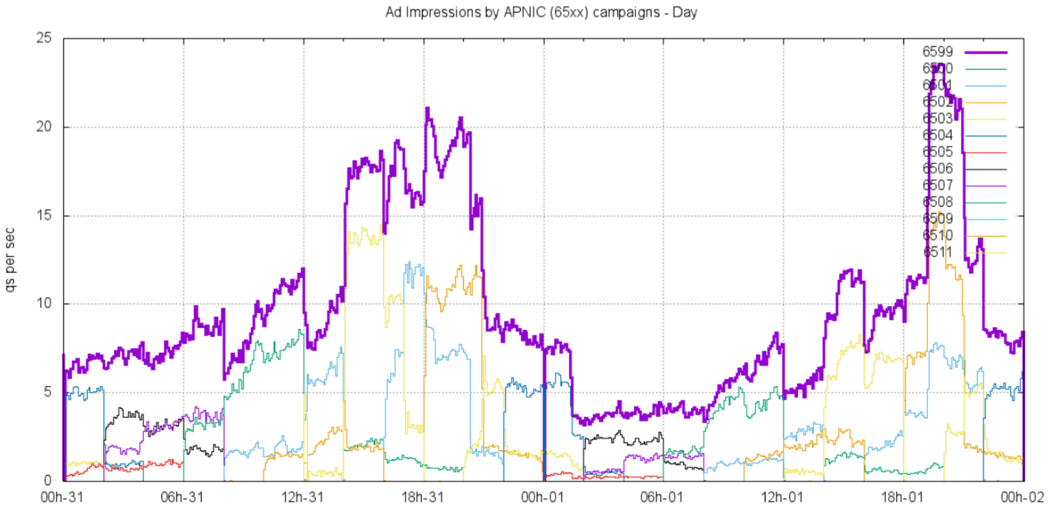
- Google groups adverts as ‘campaigns’
- APNIC runs 12 ‘campaigns’ per day, separated by time
  - All campaigns preference impressions over clicks
  - All campaigns ask for desktop, mobile and tablet device presentations
  - Solely image based adverts, based on generic keywords
  - Campaigns overlap in time, to smooth out presentation rate
    - One at ‘start’ phase, one at ‘mid’ phase, one at ‘end’ phase 24/7

Overlapping campaigns



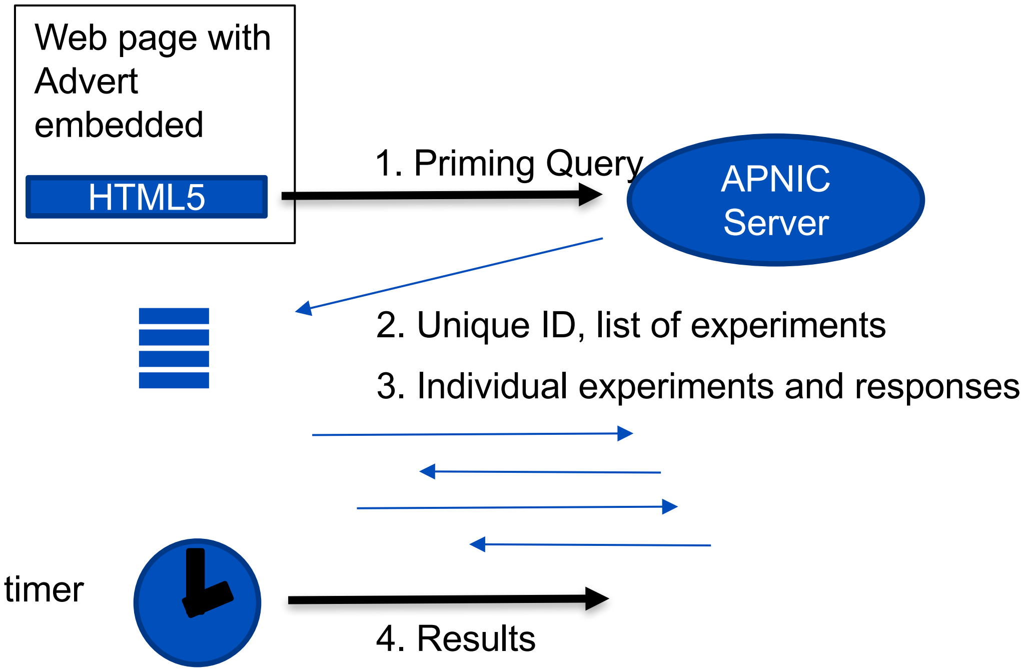
Continuous placement, shifted in time, to ensure there is always a mix of initial, mid and endpoint ad placement behaviour in the day for every user no matter in which timezone they are located

Overlapping campaigns



Continuous placement, shifted in time, to ensure there is always a mix of initial, mid and endpoint ad placement behaviour in the day for every user no matter in which timezone they are located

2. Script Control



Web page with Advert embedded
HTML5
1. Priming Query
APNIC Server
2. Unique ID, list of experiments
3. Individual experiments and responses
timer
4. Results

The “Priming Query”

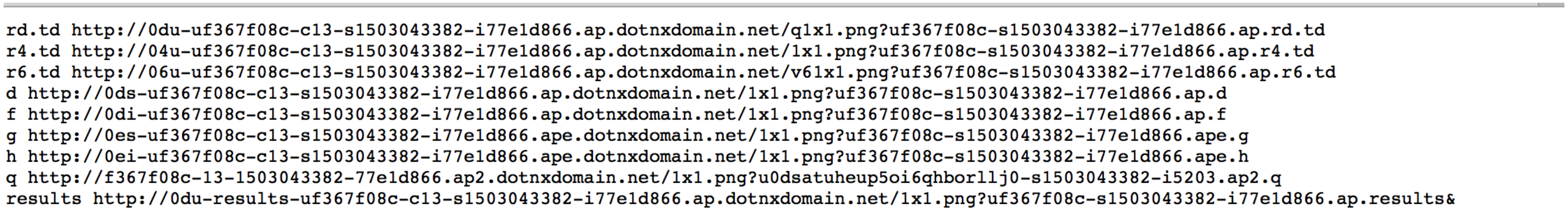
- There is a code segment in an Ad that is executed by the host at the time the Ad is loaded into the endpoint (on “impression”)
- This code segment does not require the user to click on the ad or interact with it in any way.
- The ad script is programmed to fetch an experiment task list from an APNIC server, then execute the list of tasks
- We call this initial fetch the “Priming Query”

Priming Actions

- When a server receives a priming query, it will:
  - Locate the client into a geo region
  - Pass the client a set of specific tasks that direct the tests against the APNIC server that serves that client’s locale

- The task list is controlled by the APNIC server, not the ad code. This implies that we can change the individual tasks without interruption to the ad placement systems, and we can add (or remove) task servers without reconfiguring the ad itself

An Example



```
rd.td http://0du-uf367f08c-c13-s1503043382-i77e1d866.ap.dotnxdomain.net/q1x1.png?uf367f08c-s1503043382-i77e1d866.ap.rd.td
r4.td http://04u-uf367f08c-c13-s1503043382-i77e1d866.ap.dotnxdomain.net/1x1.png?uf367f08c-s1503043382-i77e1d866.ap.r4.td
r6.td http://06u-uf367f08c-c13-s1503043382-i77e1d866.ap.dotnxdomain.net/v61x1.png?uf367f08c-s1503043382-i77e1d866.ap.r6.td
d http://0ds-uf367f08c-c13-s1503043382-i77e1d866.ap.dotnxdomain.net/1x1.png?uf367f08c-s1503043382-i77e1d866.ap.d
f http://0di-uf367f08c-c13-s1503043382-i77e1d866.ap.dotnxdomain.net/1x1.png?uf367f08c-s1503043382-i77e1d866.ap.f
g http://0es-uf367f08c-c13-s1503043382-i77e1d866.ape.dotnxdomain.net/1x1.png?uf367f08c-s1503043382-i77e1d866.ape.g
h http://0ei-uf367f08c-c13-s1503043382-i77e1d866.ape.dotnxdomain.net/1x1.png?uf367f08c-s1503043382-i77e1d866.ape.h
q http://f367f08c-13-1503043382-77e1d866.ap2.dotnxdomain.net/1x1.png?u0dsatuheup5oi6qhborllj0-s1503043382-i5203.ap2.q
results http://0du-results-uf367f08c-c13-s1503043382-i77e1d866.ap.dotnxdomain.net/1x1.png?uf367f08c-s1503043382-i77e1d866.ap.results&
```

Query Details

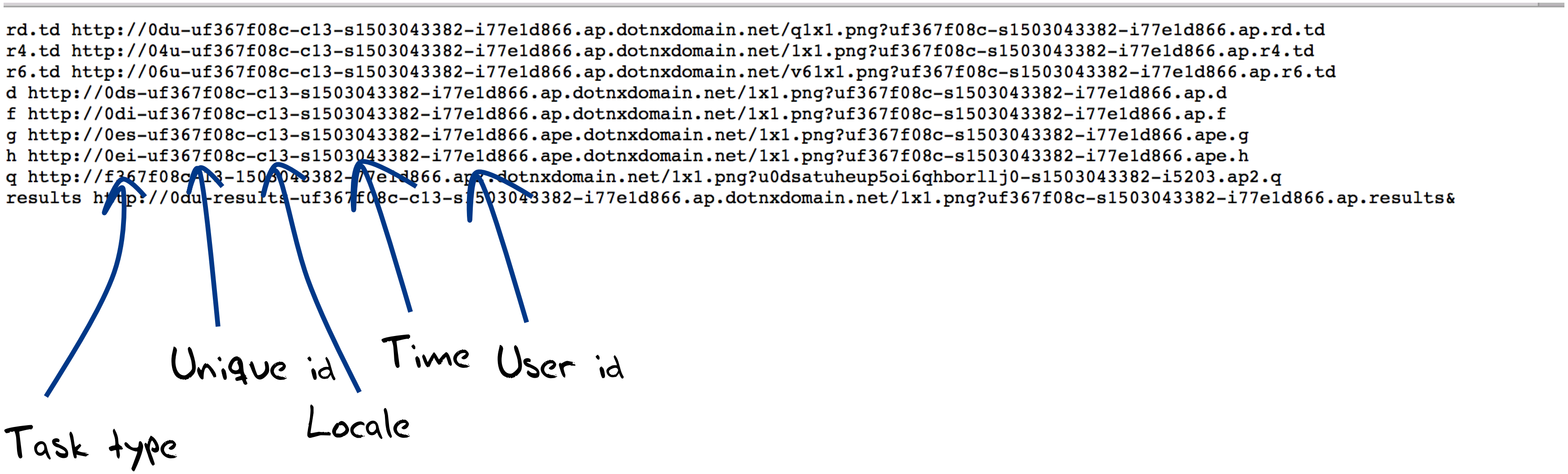
1. Connects to an APNIC server via an anycast address
2. The connection is via a HTTP(S) “GET” request
   - Server assigns the experiment a unique identity string
   - Server assigns a locale to the experiment based on the assumed location of the client IP source address
   - Server sends list of specific tasks against a nominated measurement server as the response to the web request
3. Each task is a URL
   - a component of the URL directs the test to be performed against the server operating in the assigned geographic locale

Task List

Each Task is a URL:

- The domain name component of the URL is a unique string (to bypass various forms of caches)
- The domain name label is a multi-part string constructed by the priming query server. The parts include:
  - Task name
  - Unique experiment identifier
  - Locale (country code)
  - Time of label creation
- The web component is a 1x1 pixel gif blot
- The arguments to the web object include the same parts as the DNS label

An Example



```
rd.td http://0du-uf367f08c-c13-s1503043382-i77e1d866.ap.dotnxdomain.net/q1x1.png?uf367f08c-s1503043382-i77e1d866.ap.rd.td
r4.td http://04u-uf367f08c-c13-s1503043382-i77e1d866.ap.dotnxdomain.net/1x1.png?uf367f08c-s1503043382-i77e1d866.ap.r4.td
r6.td http://06u-uf367f08c-c13-s1503043382-i77e1d866.ap.dotnxdomain.net/v61x1.png?uf367f08c-s1503043382-i77e1d866.ap.r6.td
d http://0ds-uf367f08c-c13-s1503043382-i77e1d866.ap.dotnxdomain.net/1x1.png?uf367f08c-s1503043382-i77e1d866.ap.d
f http://0di-uf367f08c-c13-s1503043382-i77e1d866.ap.dotnxdomain.net/1x1.png?uf367f08c-s1503043382-i77e1d866.ap.f
g http://0es-uf367f08c-c13-s1503043382-i77e1d866.ape.dotnxdomain.net/1x1.png?uf367f08c-s1503043382-i77e1d866.ape.g
h http://0ei-uf367f08c-c13-s1503043382-i77e1d866.ape.dotnxdomain.net/1x1.png?uf367f08c-s1503043382-i77e1d866.ape.h
q http://f367f08c-13-1503043382-i77e1d866.ape.dotnxdomain.net/1x1.png?u0dsatuheup5oi6qhborllj0-s1503043382-i5203.ap2.q
results http://0du-results-uf367f08c-c13-s1503043382-i77e1d866.ap.dotnxdomain.net/1x1.png?uf367f08c-s1503043382-i77e1d866.ap.results&
```

3. Tasks

- The tasks are processed by the script interpreter running on the endpoint system
- Each tasks involves resolution to the DNS name into an IP address
  - This lookup is essentially a gethostbyname() call performed by the endpoint, as it attempts to map the DNS name to an IP address.
  - Because the terminal label of the DNS name is unique, we will see queries from the recursive resolver used by the endpoint against an APNIC name server as it resolves this name
- Successful completion of the DNS name lookup then triggers an attempt to fetch the web object
  - Because the URL name is unique, the web fetch will be performed against an APNIC web server

Task Sequencing

- Each task is ‘independent’. There is no assumption relating to sequencing, synchronicity or fate sharing
  - Order of experiments depends on the threading model in the browser. From our perspective it’s essentially random
- Some tasks act as ‘control’ to other experiments, illustrating base capability as a comparison to a test of the support of some extension mechanism
  - Again, the order of the tasks should not be critical to the test

Task Variability

We can determine what aspect of end user capabilities we are probing by changing the behaviour of the DNS server or the HTTP server

- IPv6-only web objects and Dual Stack web objects
- DNSSEC signing of DNS names
- Packet size variation
- QUIC capability

4. Results

- The Ad script running in the endpoint maintains its own record of task execution.
- This local record for all assigned tasks is passed back to the APNIC servers through the arguments to an HTTP(S) “GET” request as a “ Result” fetch
- The “Result” fetch is performed when all tasks have been completed, or when the tasks have been running for 10 seconds.
  - This way we are provided a record of “normal” termination of the experiment

Why a “Result” fetch?

- Some measurement rely on the absence of a seen user behaviour
  - (for example, the endpoint did NOT fetch an object that was invalidly signed with DNSSEC)
- Endpoints are able to “leave” an ad at any time
  - When a user skips an Ad then the script is halted immediately
- How can we tell the difference between a task that the user could not complete, as compared to a skipped task?
- The “Result” fetch allows us to differentiate between these two cases

5. Processing

Each server maintains:

- A log from the NGINX web server of all HTTP(S) “GET” fetch requests
- A log of all DNS queries and responses sent to the authoritative name server running on the server
- A packet capture dump of the headers of all IP packets seen by the server

Processing Logic

- Ignore task records which are ‘out of time’
  - The Unique ID includes a time component. We ignore records where the time is too far back in the past.
  - Actually we don’t ignore them, as these “echo” queries provide some insight into other behaviours on the Internet relating to cache behaviour and user tracking, but we don’t use these ‘old’ tasks in the primary reports
- Ignore repeat presentations
  - Each GET component of a task should only be seen once.
  - Always take the first and ignore any following duplicates
  - Again, these duplicates are in themselves a dedicated topic of investigation

Processing Logic

- Believe the client’s fetch record, not the server’s delivery record
  - Because middleware
- Believe the server’s clock not the client’s clock
  - Because my time is always better than your time
- We collate by unique ID to combine tasks records into an “experiment” set
  - This IPv4 address with that IPv6 address
  - This IPv4 address or IPv6 address with that resolver address
- We add Origin-AS (BGP) and geolocation economy and RIR
  - Daily BGP dumps, daily cross-RIR delegation stats data

Report “Weighting”

- The Ad presentation is NOT a uniformly random sample
- We would like to report on overall Internet data, so we need to compensate for this bias in the raw AD presentation data
- We assume that the Ad presentation within an individual economy is uniform, and that the Ad presentation bias is between countries
  - This assumption is probably incorrect, but we cannot compensate at a level of granularity finer than countries, as we have no reference data

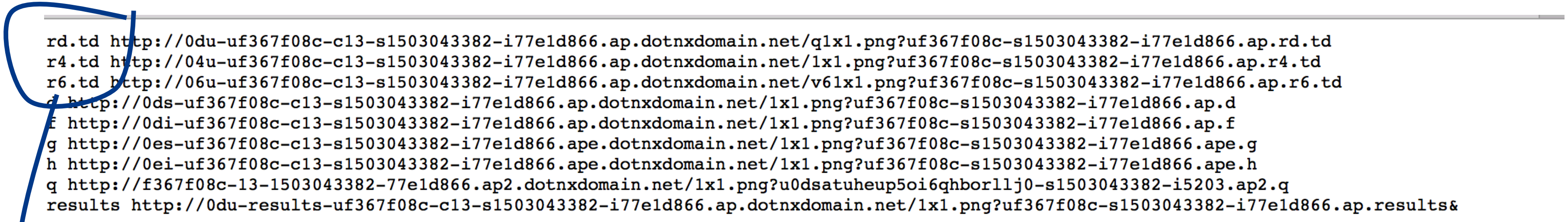
Report “Weighting”

- We use the current ITU statistics on internet users as the correcting data
  - Update the per-country user population data to today using UN population data and current population estimates
  - Generate a set of relativity numbers per economy of user populations

- Re-weight sample totals by adjustment factor, to calculate economy contribution in proportion to its estimated relative user population
  - Used to derive worldwide, regional totals
  - Does not alter counts of origin-AS seen, within one economy

IPv6 Measurement

An Example



```
rd.td http://0du-uf367f08c-c13-s1503043382-i77e1d866.ap.dotnxdomain.net/q1x1.png?uf367f08c-s1503043382-i77e1d866.ap.rd.td
r4.td http://04u-uf367f08c-c13-s1503043382-i77e1d866.ap.dotnxdomain.net/1x1.png?uf367f08c-s1503043382-i77e1d866.ap.r4.td
r6.td http://06u-uf367f08c-c13-s1503043382-i77e1d866.ap.dotnxdomain.net/v61x1.png?uf367f08c-s1503043382-i77e1d866.ap.r6.td
d http://0ds-uf367f08c-c13-s1503043382-i77e1d866.ap.dotnxdomain.net/1x1.png?uf367f08c-s1503043382-i77e1d866.ap.d
f http://0di-uf367f08c-c13-s1503043382-i77e1d866.ap.dotnxdomain.net/1x1.png?uf367f08c-s1503043382-i77e1d866.ap.f
g http://0es-uf367f08c-c13-s1503043382-i77e1d866.ape.dotnxdomain.net/1x1.png?uf367f08c-s1503043382-i77e1d866.ape.g
h http://0ei-uf367f08c-c13-s1503043382-i77e1d866.ape.dotnxdomain.net/1x1.png?uf367f08c-s1503043382-i77e1d866.ape.h
q http://f367f08c-13-1503043382-77e1d866.ap2.dotnxdomain.net/1x1.png?u0dsatuheup5oi6qhborllj0-s1503043382-i5203.ap2.q
results http://0du-results-uf367f08c-c13-s1503043382-i77e1d866.ap.dotnxdomain.net/1x1.png?uf367f08c-s1503043382-i77e1d866.ap.results&
```

1. Dual Stack object

2. iPv4-only object

3. iPv6-only object

IPv6 Capability measurement

- Number of experiments that are able to retrieve the IPv6-only web object
- Number of experiments that are able to retrieve the IPv4-only object
- The protocol choice made when processing the dual stack object

“IPv6 Capable” and “IPv6 Preferred”

- *Capable* means “can fetch IPv6”
  - The endpoint was observed to fetch a web object when the only address binding to the DNS label was an IPv6 address (IPv6 only)

- *Preferred* means “given dual-stack options, selected IPv6”
  - The endpoint was given a DNS name that had bindings to both IPv4 and IPv6 addresses
  - The observed web fetch was over IPv6. “Happy Eyeballs” or some other local protocol selection logic at the endpoint preferred to use IPv6 over IPv4.

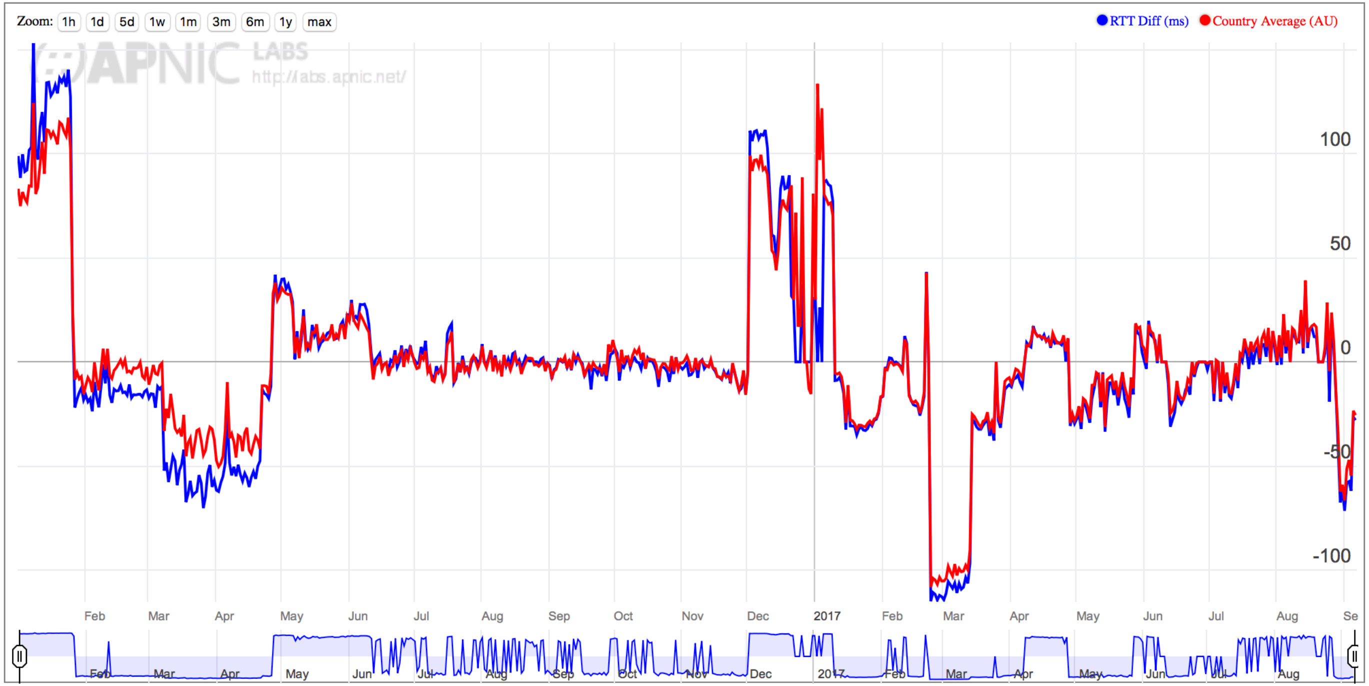
IPv6 “Performance” Measurement

- We can look at the time between the receipt of the TCP SYN and the TCP ACK of the initial TCP handshake
  - The difference is a reasonable measurement of the RTT between the server and the client
- Where we have IPv4 and IPv6 measurements for the same endpoint we can compare the two RTT values

IPv6 “Performance” Measurement

- We car SYN ar
  - The di server
- Where endpoi

V6 Performance for AS1221: ASN-TELSTRA Telstra Pty Ltd, Australia (AU)



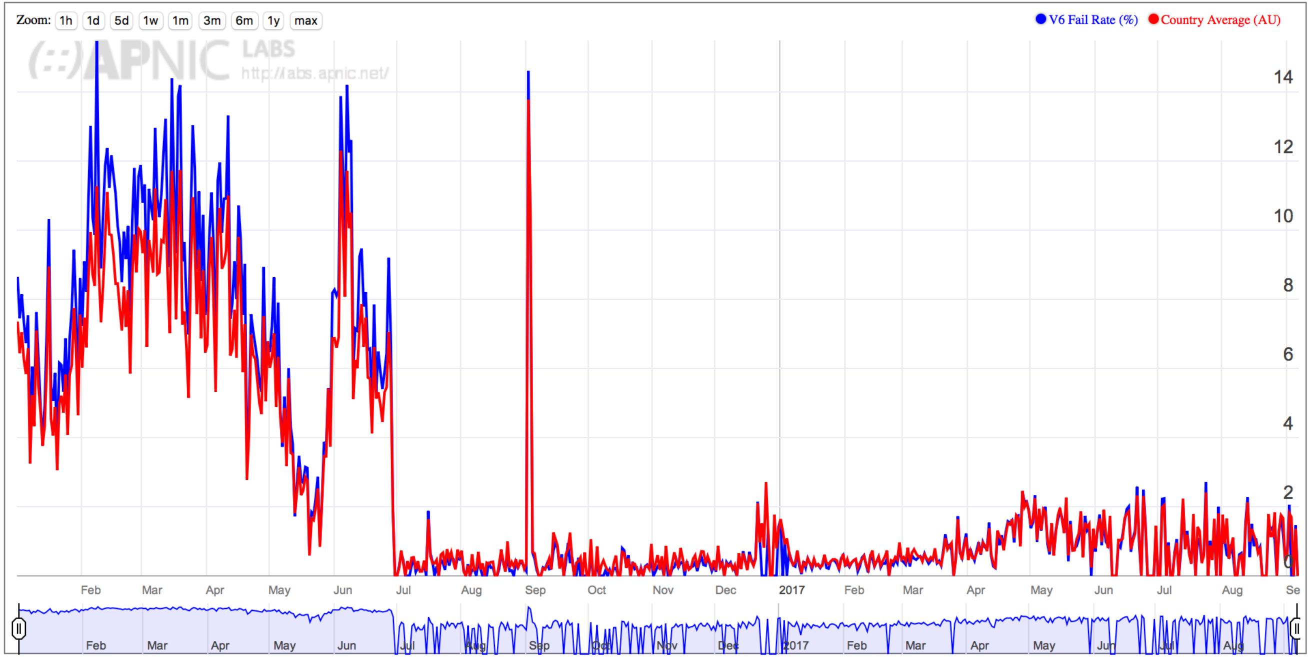
IPv6 “Robustness” Measurement

- As well as looking at completed TCP handshakes, we can look at the incomplete handshakes
- Here the server receives the initial TCP SYN packet, and responds with a TCP SYN/ACK
- Most of the time the session completes with a received ACK
- But what about the others? How many sessions remain “hanging” in an incomplete state?

IPv6 “Robustness” Measurement

- As well as lc
look at the ir
- Here the se
responds wi
- Most of the
- But what ab
“hanging” in

V6 Connection Failure Rate for AS1221: ASN-TELSTRA Telstra Pty Ltd, Aust



Reports

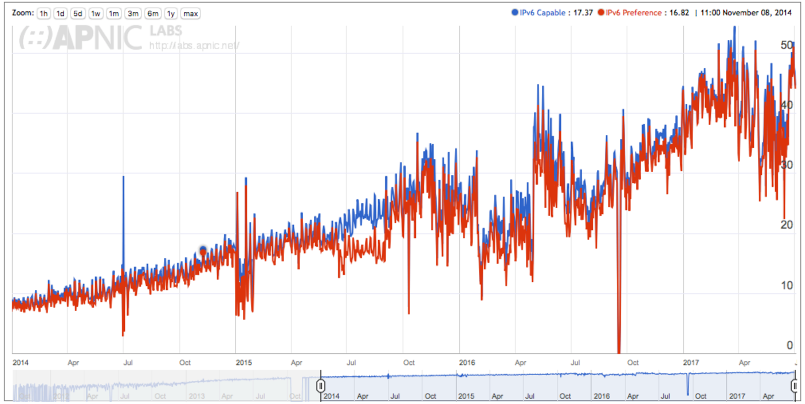
‘Standing’ Reports - We are performing these reports every day to establish a long baseline of data

1. IPv6 Deployment – a measure of the proportion of endpoints that can successfully perform Internet transactions using the IPv6 protocol
2. IPv6 Performance – a measure of the relative RTT delay between the endpoints and an APNIC server in IPv4 and IPv6, and a measure of the IPv6 TCP session drop probability
3. DNSSEC – a measure of the proportion of endpoints that exclusively use DNSSEC-validating DNS resolvers
4. DNSSEC using ECDSA – a refinement of the DNSSEC test comparing the resolvers to their ability to validate using the ECDSA crypto algorithm as compared the RSA

– These reports are by geographic classification (economy, region, global) and by network (origin AS totals)

https://stats.labs.apnic.net

Use of IPv6 for Germany (DE)

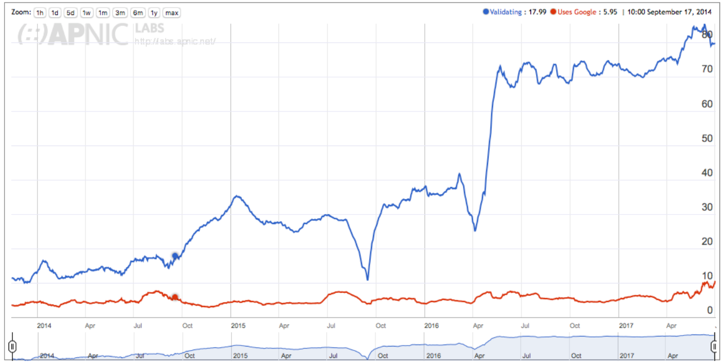


Average Interval (days) 1

☐ Hide country ASN List
☐ Hide IPv6 Preference
☐ Hide Regional Use

Redraw

Use of DNSSEC Validation for Norway (NO)

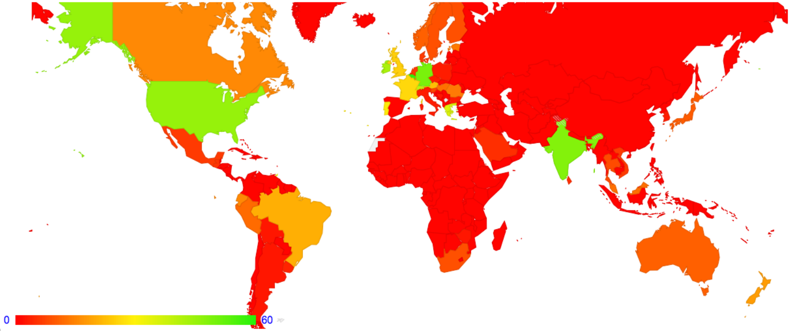


Average Interval (days) 30

☐ Hide country ASN List
☐ Hide Google PDNS Use
☐ Hide Regional Use

Redraw

IPv6 Capable Rate by country (%)



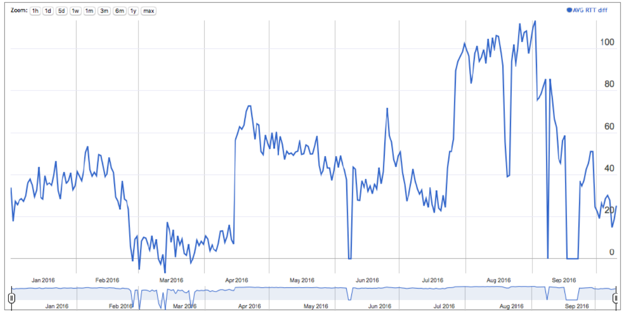
7 day average (26/06/2017 - 02/07/2017)

Custom Measurements

- We have used this platform to measure other user behaviours. For example:
  - What is the extent of user ’shadowing’? This was looking at the distribution and extent of second time presentations of the same unique tasks
  - How do names ‘decay’ in the DNS? This looks at the presentation patterns of ‘old’ DNS names over time
  - What proportion of DNS resolvers cannot receive a fragmented DNS UDP response using IPv6?
  - What proportion of users cannot receive a fragmented IPv6 packet?

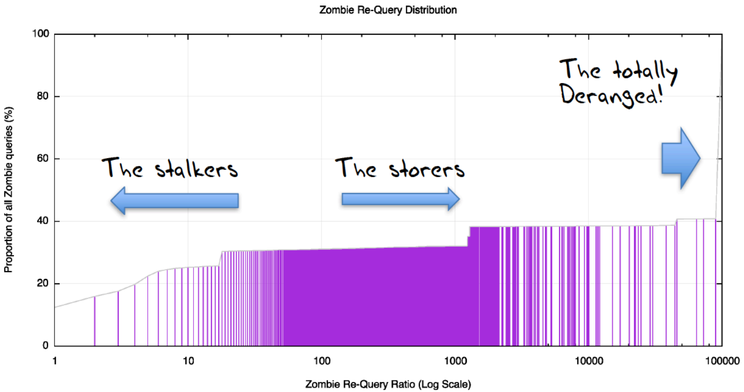
Global Results

Average RTT Difference (ms) (V6 - V4) for World (XA)



On average iPv6 is showing 20ms – 40ms slower that iPv4

Three Zombie Factories

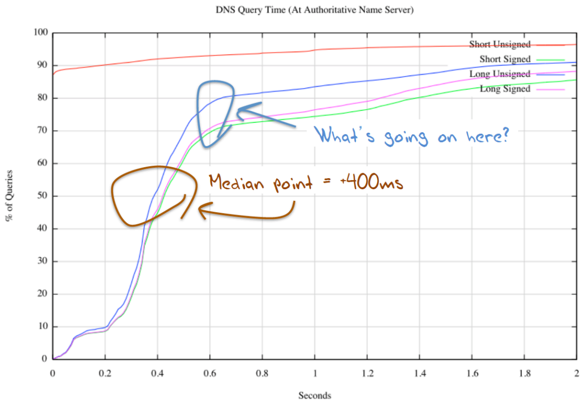


The stalkers

The storers

The totally Deranged!

Time to resolve a name



What's going on here?

Median point = +400ms

Some common questions…

Why does my economy see more or less samples than other economies?

- Google choses how to display the advertisement to maximize advertisement revenue to Google
  - The ad placement algorithm is obviously a secret, but the algorithm is not only based on the advertiser's preferences, but on other campaigns from other advertisers that are running at the same time
  - From time time time, google changes its display placement algorithms
  - At APNIC we designed the Ad placement directives to maximise impressions, so we bid low for clicks. This influences where Google choose to place our measurement Ad.
- We have no control over this placement process to any useful level of detail
  - We can change the campaign budget, but the finer control over the ad placement is not directly available to us as an advertiser
- We have seen significant peaks and troughs for some economies
  - But we have always remained above a threshold of statistical validity for the major economies worldwide
  - It is possible some smaller internet economies are less reliably measured

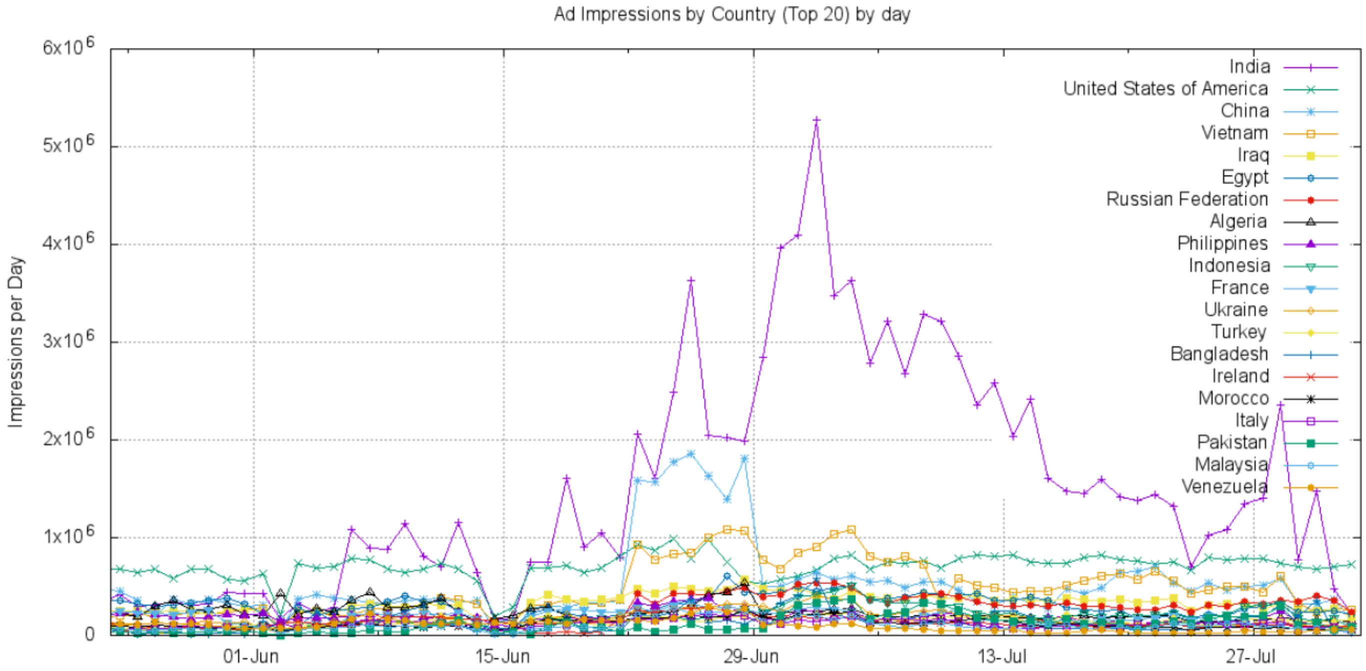
Why does my network see more or less samples than other networks?

- See previous slide

And…

- Not all networks are the same size, and smaller networks get fewer ads than larger networks, as Google is trying to reach “eyeballs” not networks
- Some networks have a higher prevalence of ad blocker use than others
- Some markets do not use Google Ads as intensively as others

Why does my economy see more or less samples than other economies?



Why is your estimate of user population in my network different than mine?

- It’s just an estimate
- It’s rough
- It relies on a number of sources, including world population and the ITU’s user population estimates
- It has a hard time coping with multiple devices per user
- It’s not sure if it is measuring devices or users
- Ad placement is a dark art known only to the ad placer, not the advertiser
- And all it can see are devices that get Ads!
- So it’s a very rough estimate
- It’s probably wrong in the detail, but as a distinguisher between large, small and somewhere in the middle its not completely bad!

Personal Privacy

- We take personal privacy seriously at APNIC
- We do not divulge individual IP addresses gathered in this measurement
- We report only aggregate data using totals grouped by Origin AS (network) and Economy

Can I stop APNIC measuring me?

- Not at this point
- We had a “Don’t measure me” Cookie, but these days Google Ads are not permitted to read or write cookies in the user’s browser
- If you are running an Ad blocker then the measurement will not be presented to you

Acknowledgements

- Google has supported our efforts to conduct this measurement since its inception, and continues to support us.
- ICANN provide support for this measurement activity, particularly relating to DNS and DNSSEC capabilities
- ISC supports us with a server located at PAIX
- And Ray Bellis - development of the dynamic DNS server code

Thanks!