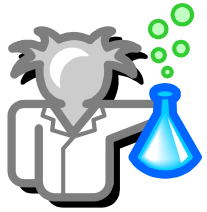
Operating the Internet's Largest Measurement System

Geoff Huston AM

APNIC Labs



Indonesian ISPs

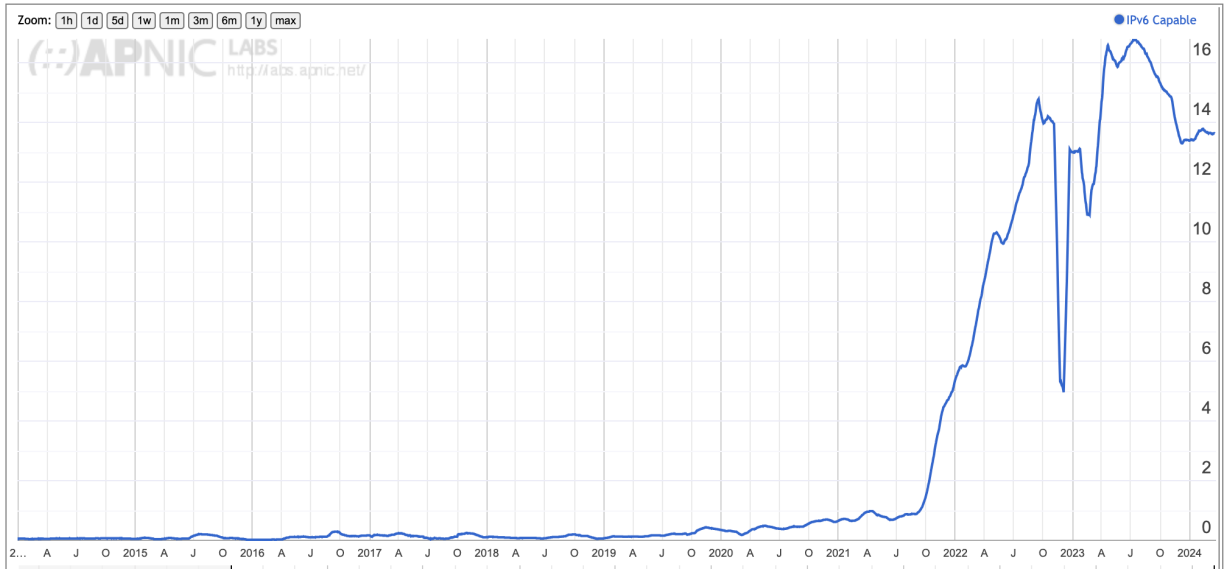
Visible ASNs: Customer Populations (Est.)

Date: 19/03/2024

Rank	ASN	AS Name	CC	Users (est.)	% of country	% of Internet	Samples
1	AS7713	TELKOMNET-AS-AP PT Telekomunikasi Indonesia	ID	31,984,039	27.94	0.761	10,095,262
2	AS23693	TELKOMSEL-ASN-ID PT. Telekomunikasi Selular	ID	21,240,149	18.56	0.505	6,704,121
3	AS4761	INDOSAT-INP-AP INDOSAT Internet Network Provider	ID	12,044,797	10.52	0.286	3,801,752
4	AS24203	NAPXLNET-AS-ID PT XL Axiata	ID	10,101,701	8.82	0.24	3,188,444
5	AS45727	THREE-AS-ID Hutchison CP Telecommunications, PT	ID	5,400,856	4.72	0.128	1,704,696
6	AS18004	WIRELESSNET-ID PT WIRELESS INDONESIA WIN	ID	4,685,551	4.09	0.111	1,478,921
7	AS9341	ICONPLN-ID-AP-ISP PT INDONESIA COMNETS PLUS	ID	2,799,752	2.45	0.067	883,698
8	AS63859	MYREPUBLIC-AS-ID PT. Eka Mas Republik	ID	2,125,548	1.86	0.051	670,896
9	AS23700	FASTNET-AS-ID Linknet-Fastnet ASN	ID	2,094,676	1.83	0.05	661,152
10	AS17451	BIZNET-AS-AP BIZNET NETWORKS	ID	1,570,342	1.37	0.037	495,654

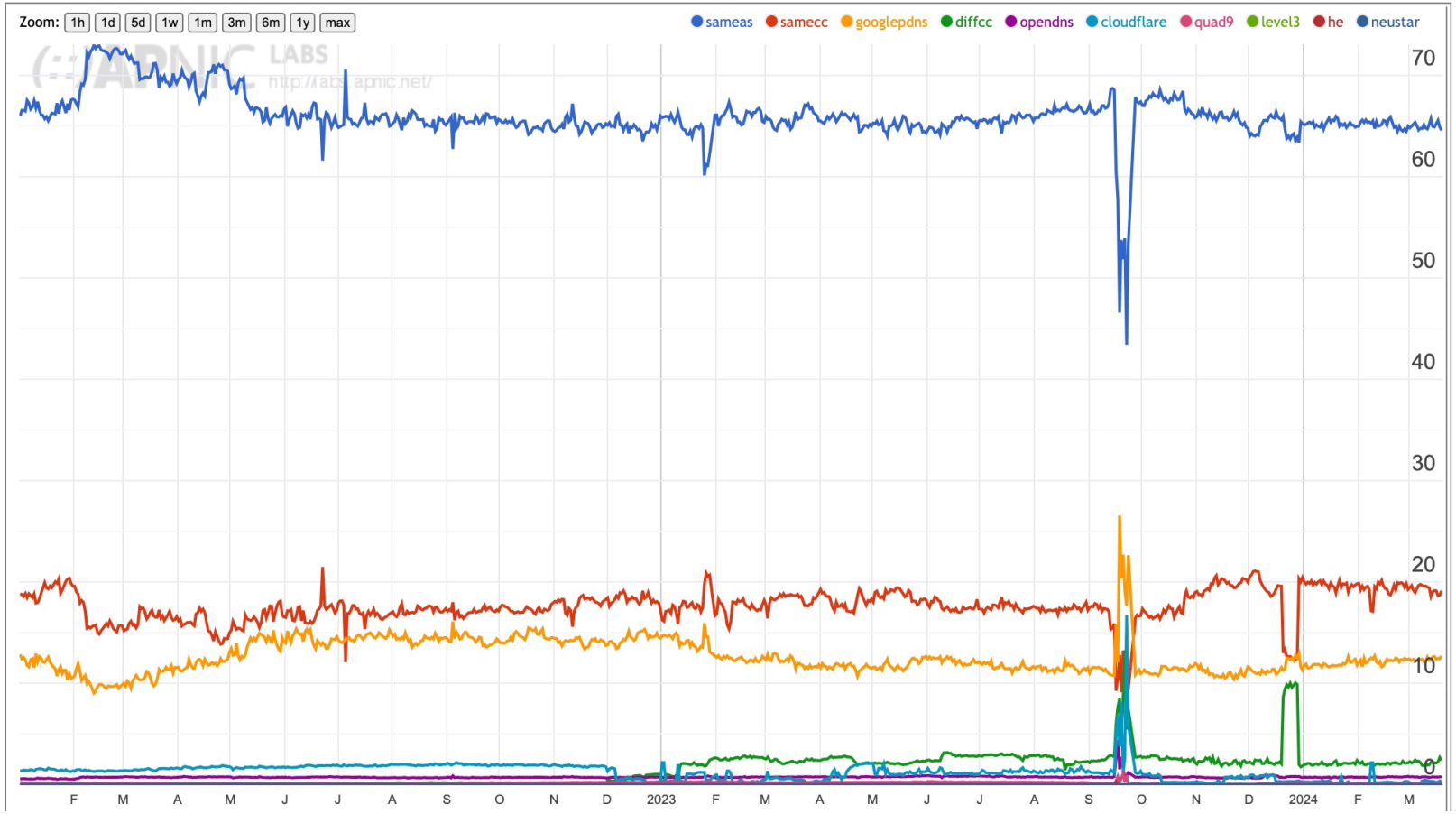
Use of IPv6 in Indonesia

Use of IPv6 for Indonesia (ID)



ASN	AS Name	IPv6 Capable	Samples
AS7713	TELKOMNET-AS-AP PT Telekomunikasi Indonesia	32.65%	5,346,904
AS23693	TELKOMSEL-ASN-ID PT. Telekomunikasi Selular	9.09%	3,507,070
AS4761	INDOSAT-INP-AP INDOSAT Internet Network Provider	0.42%	1,980,981
AS24203	NAPXLNET-AS-ID PT XL Axiata	21.82%	1,664,661
AS45727	THREE-AS-ID Hutchison CP Telecommunications, PT	0.03%	889,736
AS18004	WIRELESSNET-ID PT WIRELESS INDONESIA WIN	0.45%	769,580

DNS Resolver use in Indonesia



ISP resolver – 60%

Google’s resolver – 10%

So much to measure…

so little time!

Measurement at Scale

- We could measure the network using embedded measurement scripts in a web site – but we would really need to use a massively popular web service to conduct this experiment
  - But “massively popular web services” worry constantly about service resiliency and privacy of their data regarding users
  - They tend to be extremely suspicious of adding script elements to their service that performs third party dual stack tests with their clients (and I can’t blame them!)
- So, we need to rethink this approach…

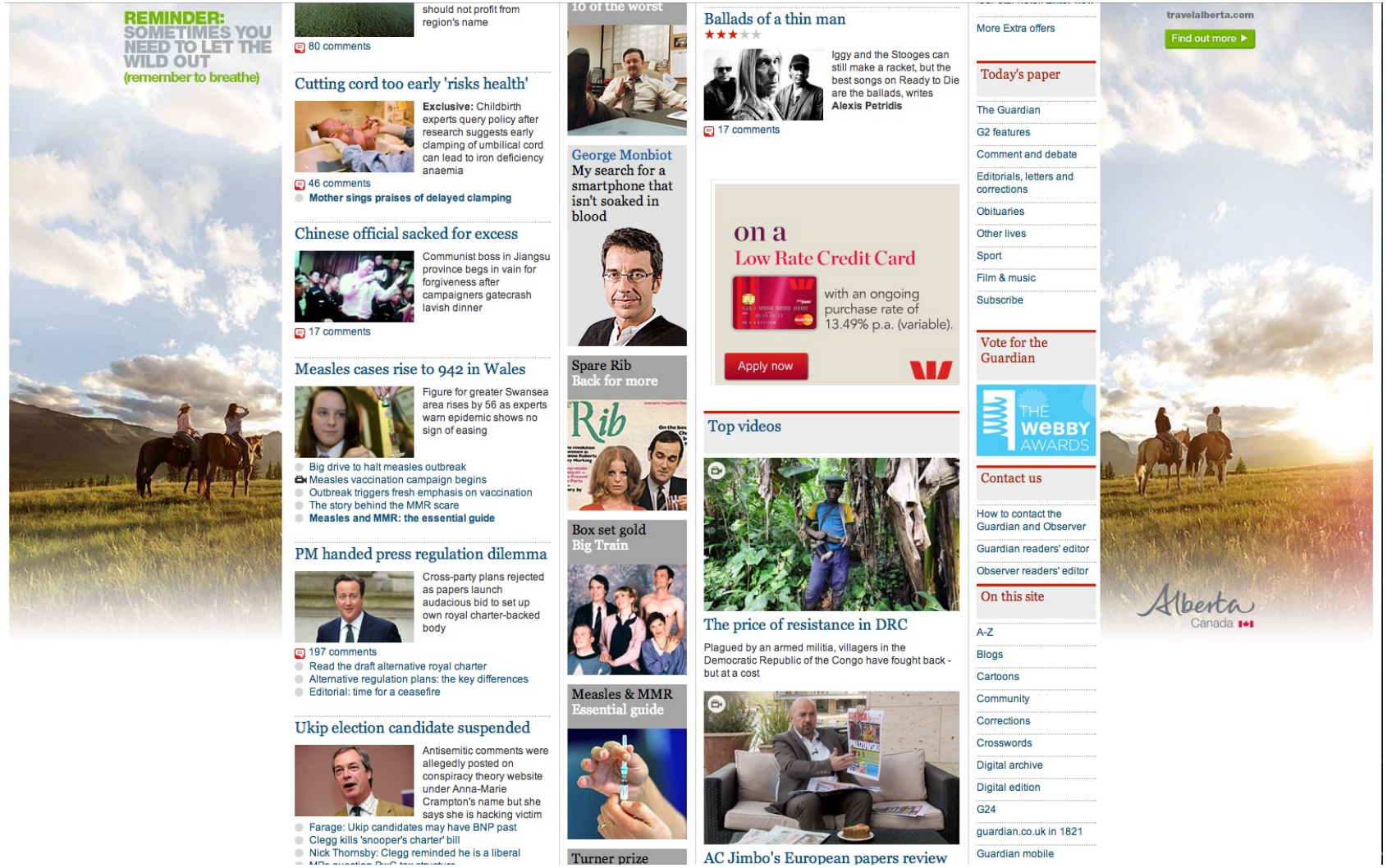
How to conduct measurements at scale

Use Google!

How to conduct measurements at scale

How?

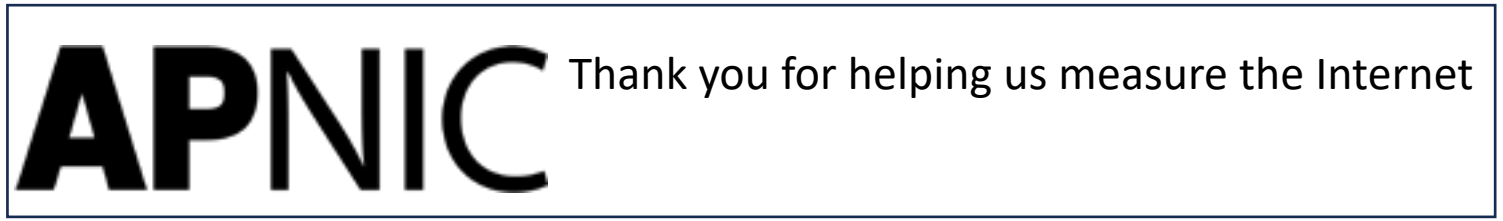
Online Ads


REMINDER:
SOMETIMES YOU NEED TO LET THE WILD OUT
(remember to breathe)
should not profit from region's name
80 comments
Cutting cord too early 'risks health'
Exclusive: Childbirth experts query policy after research suggests early clamping of umbilical cord can lead to iron deficiency anaemia
46 comments
Mother sings praises of delayed clamping
Chinese official sacked for excess
Communist boss in Jiangsu province begs in vain for forgiveness after campaigners gatecrash lavish dinner
17 comments
Measles cases rise to 942 in Wales
Figure for greater Swansea area rises by 56 as experts warn epidemic shows no sign of easing
Big drive to halt measles outbreak
Measles vaccination campaign begins
Outbreak triggers fresh emphasis on vaccination
The story behind the MMR scare
Measles and MMR: the essential guide
PM handed press regulation dilemma
Cross-party plans rejected as papers launch audacious bid to set up own royal charter-backed body
197 comments
Read the draft alternative royal charter
Alternative regulation plans: the key differences
Editorial: time for a ceasefire
Ukip election candidate suspended
Antisemitic comments were allegedly posted on conspiracy theory website under Anna-Marie Crampton's name but she says she is hacking victim
Farage: Ukip candidates may have BNP past
Clegg kills 'snooper's charter' bill
Nick Thornsby: Clegg reminded he is a liberal
10 of the worst
George Monbiot
My search for a smartphone that isn't soaked in blood
Spare Rib
Back for more
Rib
Box set gold
Big Train
Measles & MMR
Essential guide
Turner prize
Ballads of a thin man
Iggy and the Stooges can still make a racket, but the best songs on Ready to Die are the ballads, writes Alexis Petridis
17 comments
on a
Low Rate Credit Card
with an ongoing purchase rate of 13.49% p.a. (variable).
Apply now
Top videos
The price of resistance in DRC
Plagued by an armed militia, villagers in the Democratic Republic of the Congo have fought back - but at a cost
AC Jimbo's European papers review
More Extra offers
Today's paper
The Guardian
G2 features
Comment and debate
Editorials, letters and corrections
Obituaries
Other lives
Sport
Film & music
Subscribe
Vote for the Guardian
THE WEBBY AWARDS
Contact us
How to contact the Guardian and Observer
Guardian readers' editor
Observer readers' editor
On this site
A-Z
Blogs
Cartoons
Community
Corrections
Crosswords
Digital archive
Digital edition
G24
guardian.co.uk in 1821
Guardian mobile
travelalberta.com
Find out more
Alberta
Canada

Ads use scripts

- Each time an ad is loaded the ad server loads creative content and scripts on to the client's browser
- The script can include action items to fetch 'network assets'
  - Typically used to load alternate images, sequences
  - Its not a generalized network stack, subject to constraints such as limited to certain object loads, reduced run-time library
- There are on-Load, on-Hover and on-Click actions
  - We want to minimise interactions, so we use on-Load scripting

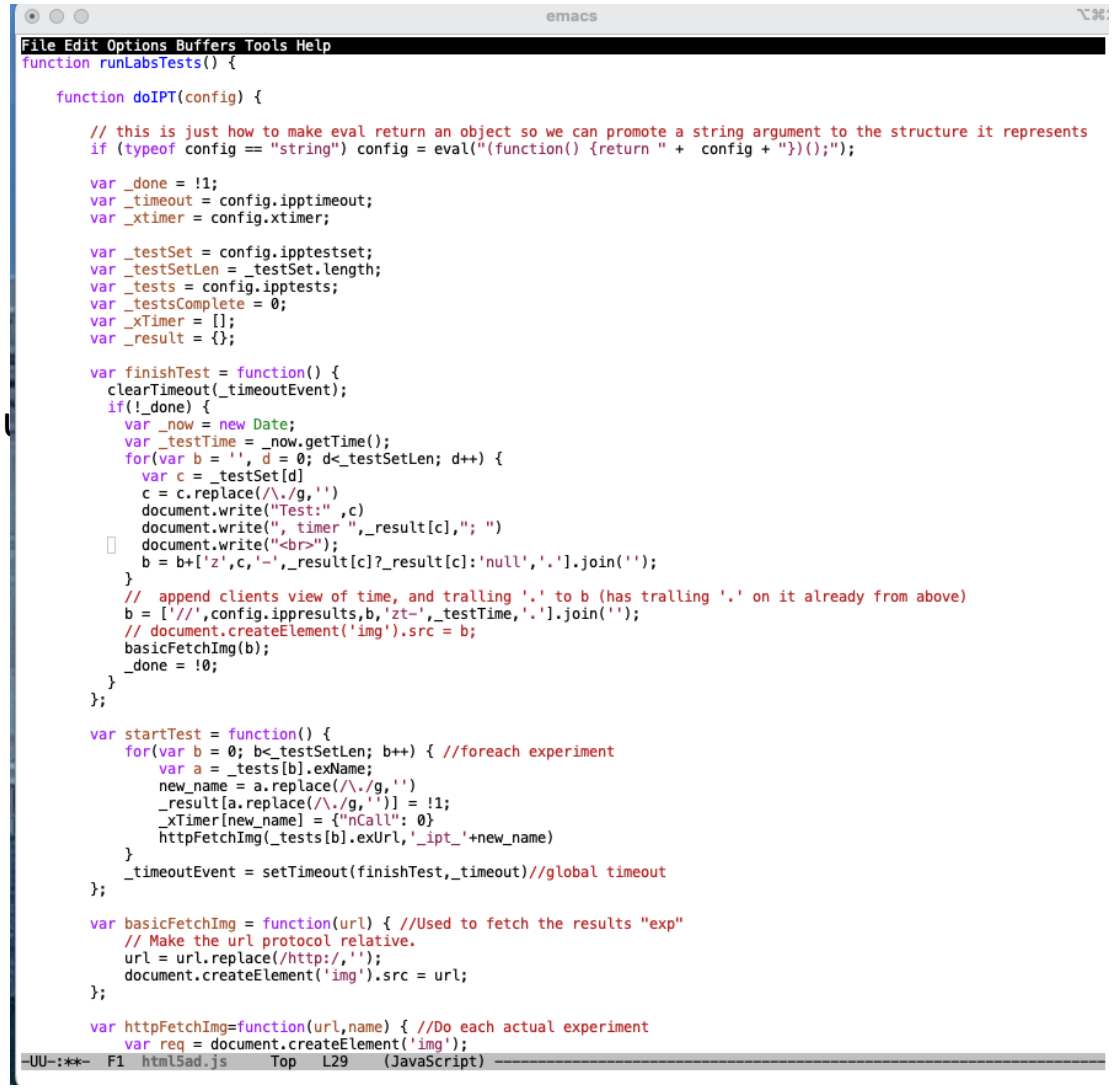
Our Ad



Our Ad Script

APNIC

Thank you



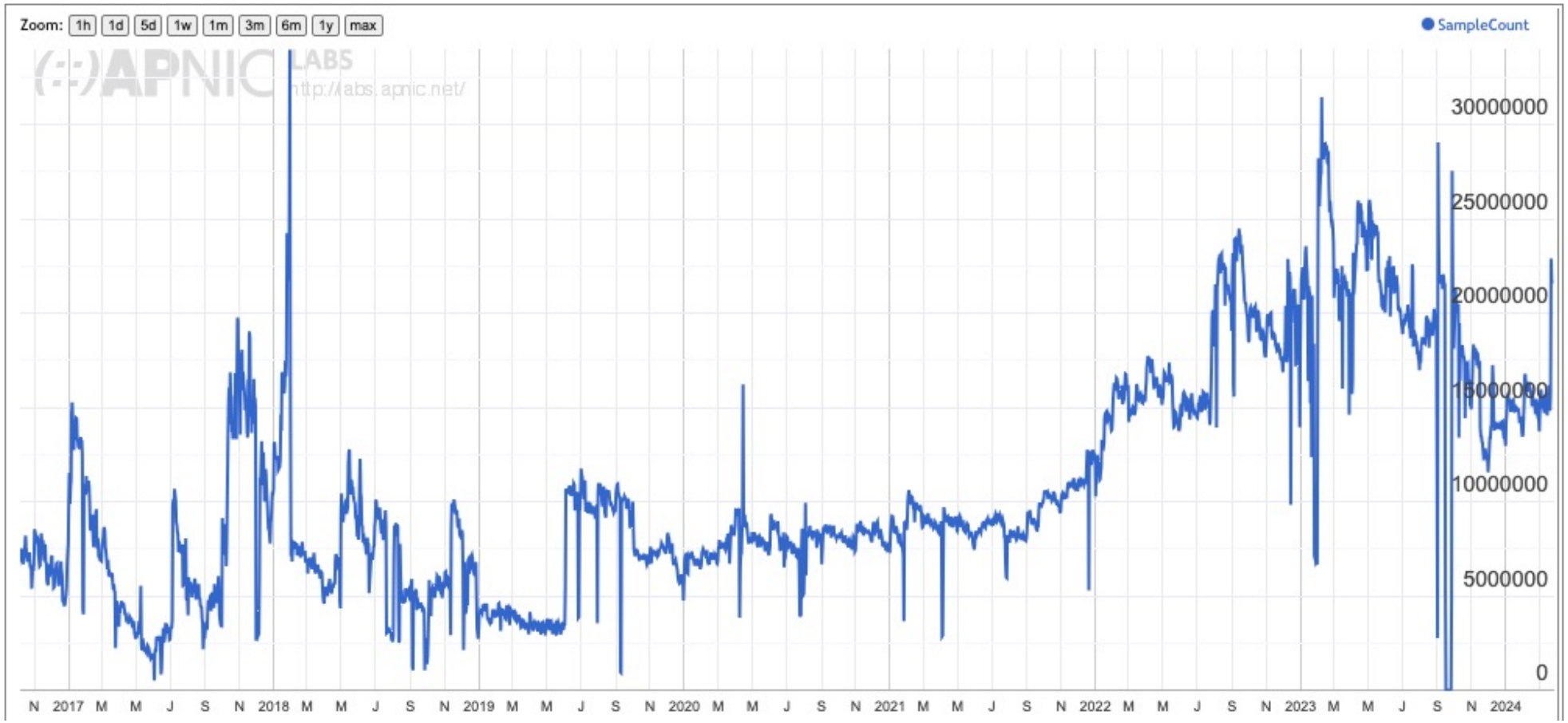
This can work

- We can instrument the target host via an ad script
  - we can constrain the ad script to talk **only** to our server(s)
  - And if we instrument these servers, then we can infer the host's properties
- Ads try to deliver to new users all of the time
  - We want to measure new sample points all of the time to avoid implicit repeat bias in the measurement set
- Ads are biased towards 'clicks'
  - We are not interested in clicks
  - We just want impressions
  - Impressions are far cheaper than clicks!

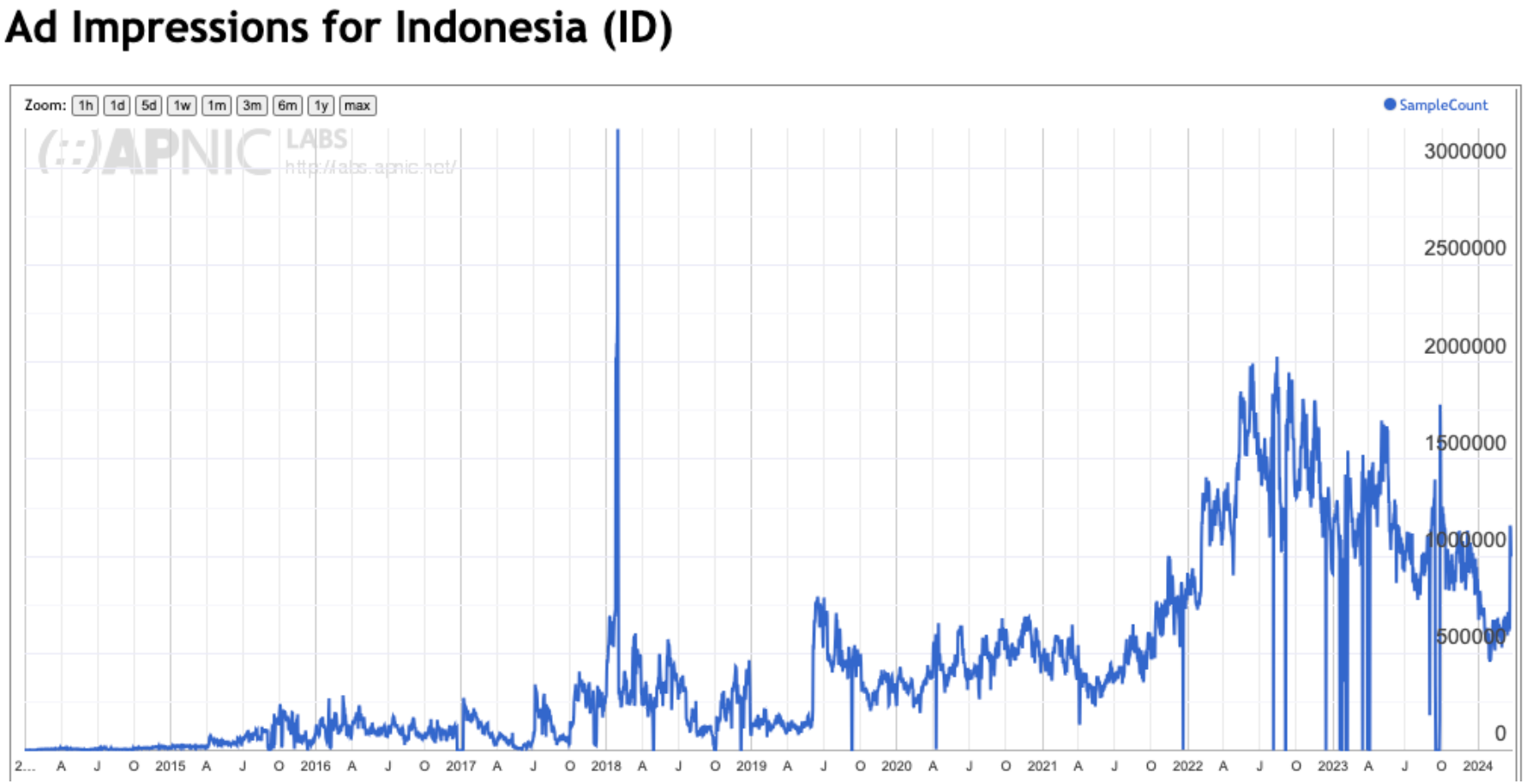
Advertising placement logic

- Fresh Eyeballs == Unique endpoints
  - We have good evidence the advertising channel is able to sustain a constant supply of unique endpoints
- Pay by click, or pay by impression
  - If you select a preference for impressions, then the channel tries hard to present your ad to as many unique endpoints as possible
- Time/Location/Context tuned
  - Can select for time of day, physical location or keyword contexts (for search-related ads)
  - But if you don't select, then placement is generalized
- Aim to fill budget
  - If you request $100 of ad placements per day, then inside 24h the algorithm tries hard to even placement but in the end the ad placement algorithm will ramp up your ad to achieve enough views, to bill you $100

Daily Ad Impression Count



Daily Ad Impression Count for Indonesia



Measuring IPv6 via Ads

- Use HTML5 code that is executed on ad impression
  - Client retrieves set of “tests” from an ad-controller
  - Client is given 10 URLs to load, including:
    - Dual Stack object
    - V4-only object
    - V6-only object
    - Result reporting URL (10 second timer)

    All DNS is dual stack

    All URLs use a unique DNS label

Experiment Server config

- There are five server sets, identically configured in VMs in DCs (Frankfurt, Singapore, Dallas, Sao Paulo, Mumbai)
- The experiment script directs the client to the “closest” server set (based on geolocation of the client IP address)
- Server set has dedicated DNS and web content server VMs

Collected Data

Per Server, Per Day:

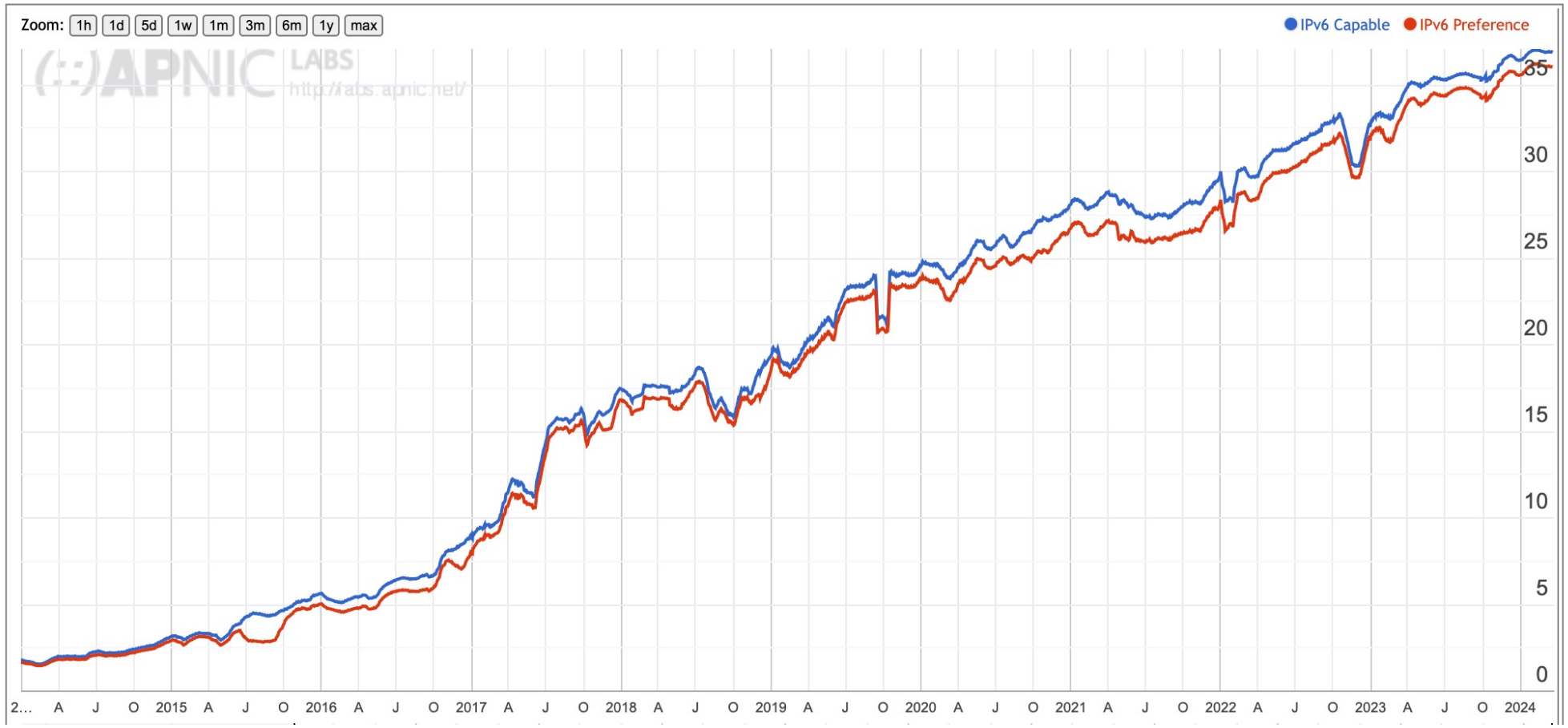
- HTTP access log
  (successfully completed fetches)
- DNS query log
  (incoming DNS queries)
- Packet capture
  All packets!

Data Analysis

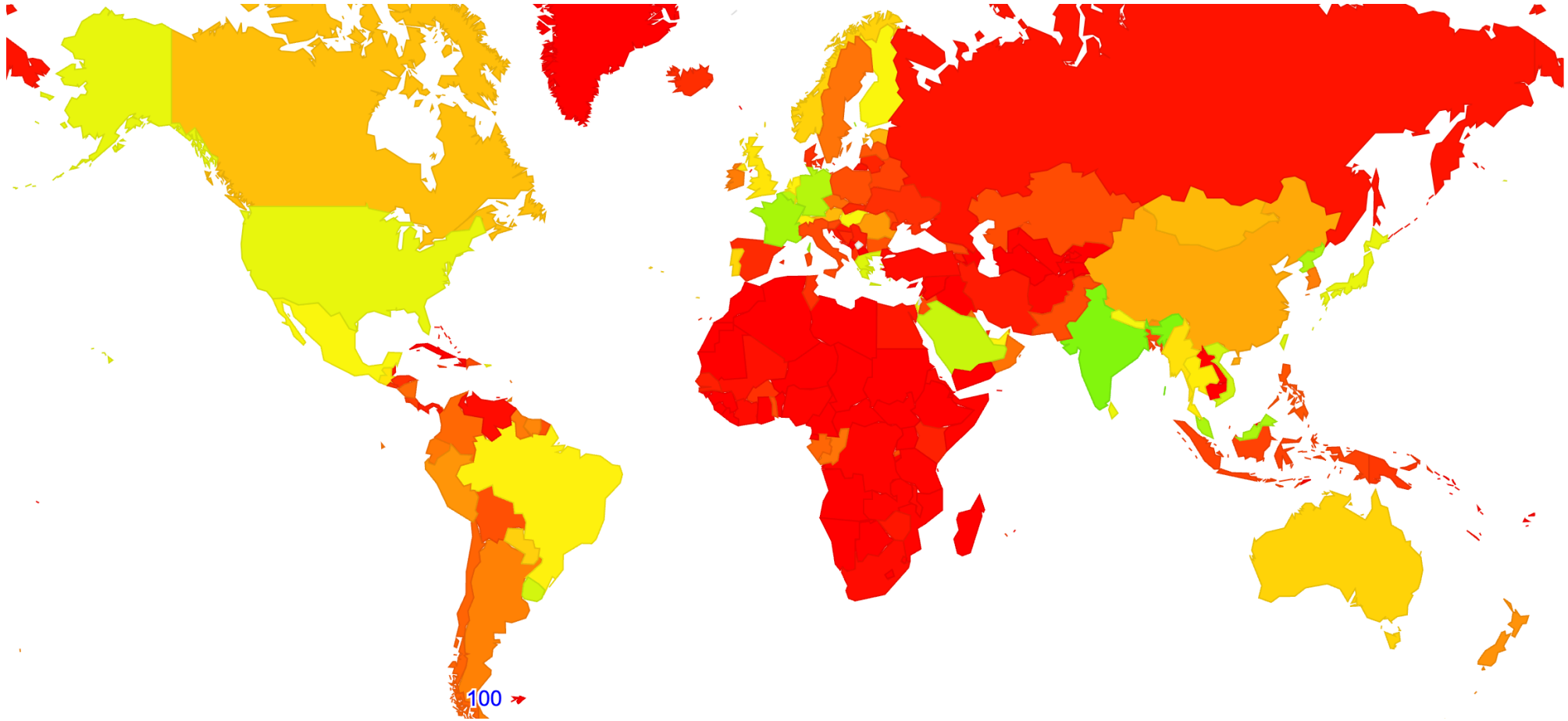
For example – IPv6 measurement

- IPv6 “capable” means that the client successfully fetched the URL target that is only accessible using IPv6
- IPv6 “preferred” means that the client used IPv6 to fetch the dual stack URL target
- Aggregate data by origin AS and by geolocation CC
- “Normalise” the country data against estimates of national user populations (to compensate for aD placement bias at a national level)
- Generate IPv6 daily report and data to data set

V6 Time Series for the entire Internet

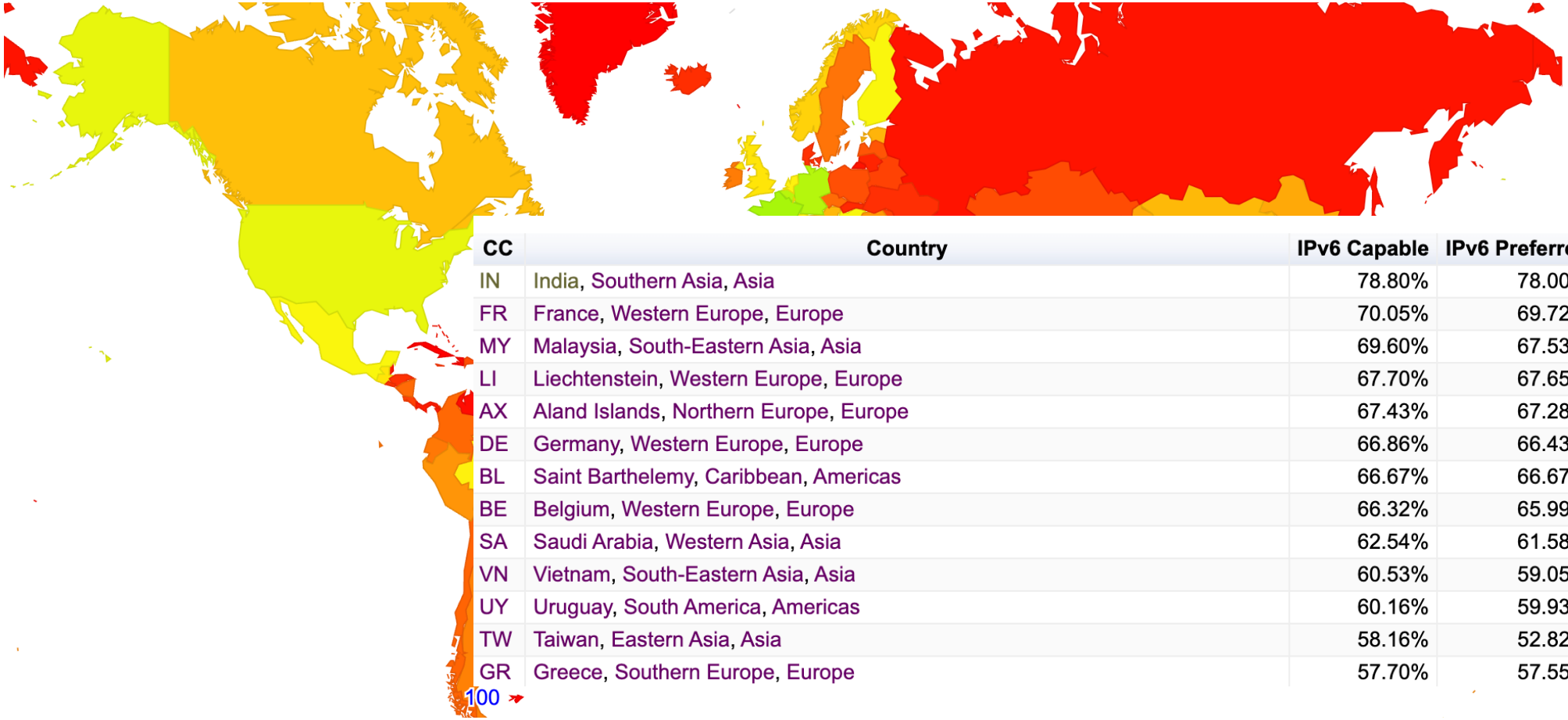


IPv6 Report



https://stats.labs.apnic.net/ipv6

IPv6 Report

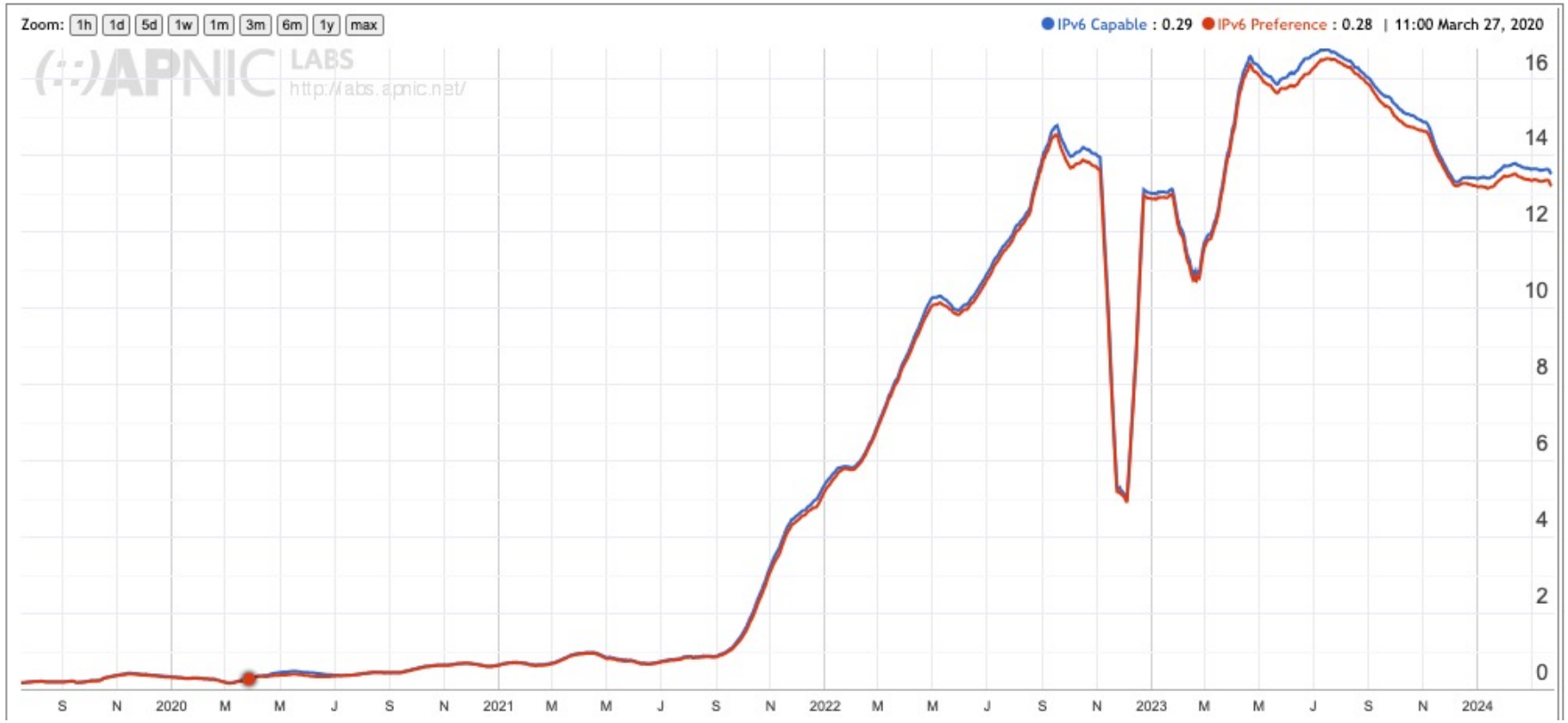


CC	Country	IPv6 Capable	IPv6 Preferred	Samples
IN	India, Southern Asia, Asia	78.80%	78.00%	71,711,972
FR	France, Western Europe, Europe	70.05%	69.72%	7,187,587
MY	Malaysia, South-Eastern Asia, Asia	69.60%	67.53%	4,227,730
LI	Liechtenstein, Western Europe, Europe	67.70%	67.65%	4,080
AX	Aland Islands, Northern Europe, Europe	67.43%	67.28%	6,494
DE	Germany, Western Europe, Europe	66.86%	66.43%	6,722,079
BL	Saint Barthelemy, Caribbean, Americas	66.67%	66.67%	126
BE	Belgium, Western Europe, Europe	66.32%	65.99%	1,702,605
SA	Saudi Arabia, Western Asia, Asia	62.54%	61.58%	4,009,647
VN	Vietnam, South-Eastern Asia, Asia	60.53%	59.05%	5,871,687
UY	Uruguay, South America, Americas	60.16%	59.93%	113,677
TW	Taiwan, Eastern Asia, Asia	58.16%	52.82%	4,932,292
GR	Greece, Southern Europe, Europe	57.70%	57.55%	1,704,160

https://stats.labs.apnic.net/ipv6

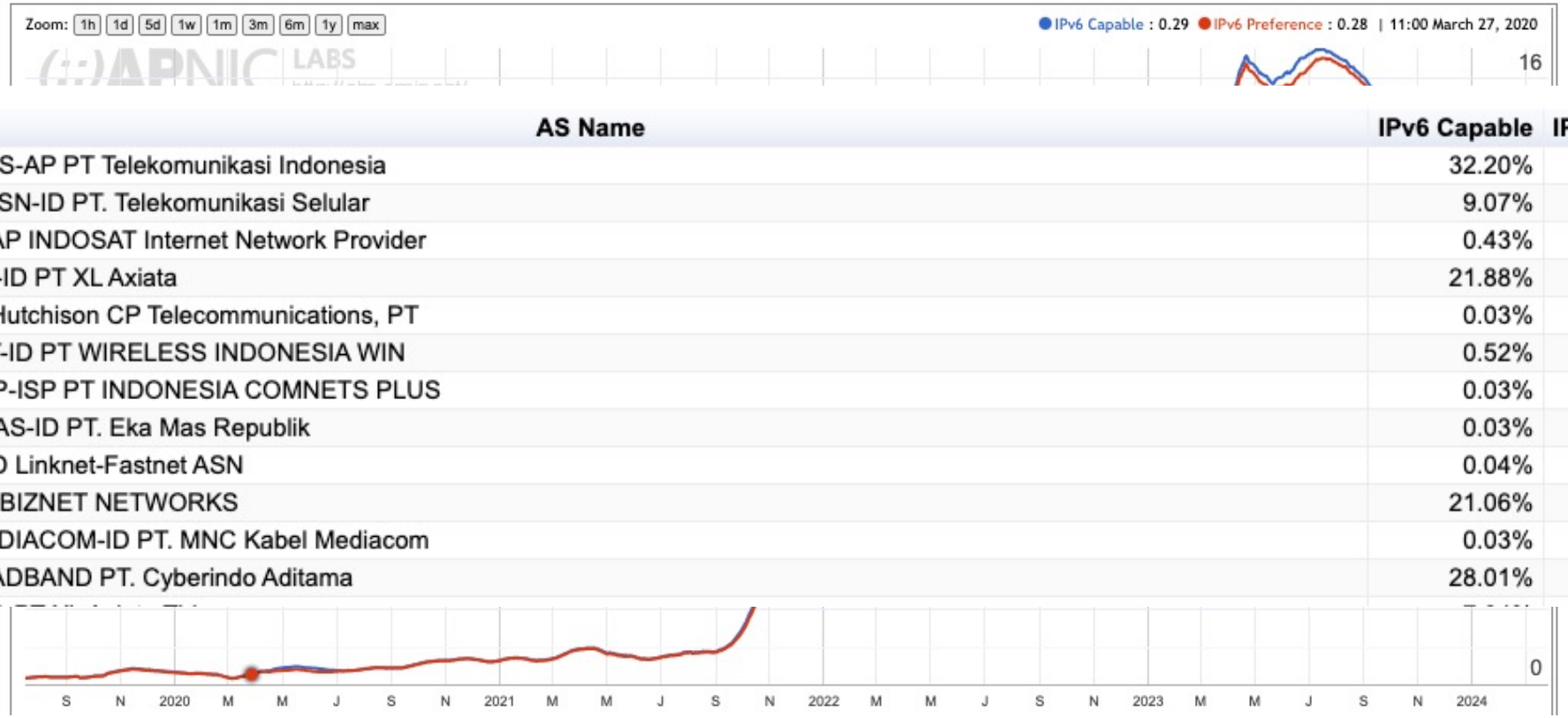
IPv6 - Indonesia

Use of IPv6 for Indonesia (ID)



IPv6 - Indonesia

Use of IPv6 for Indonesia (ID)



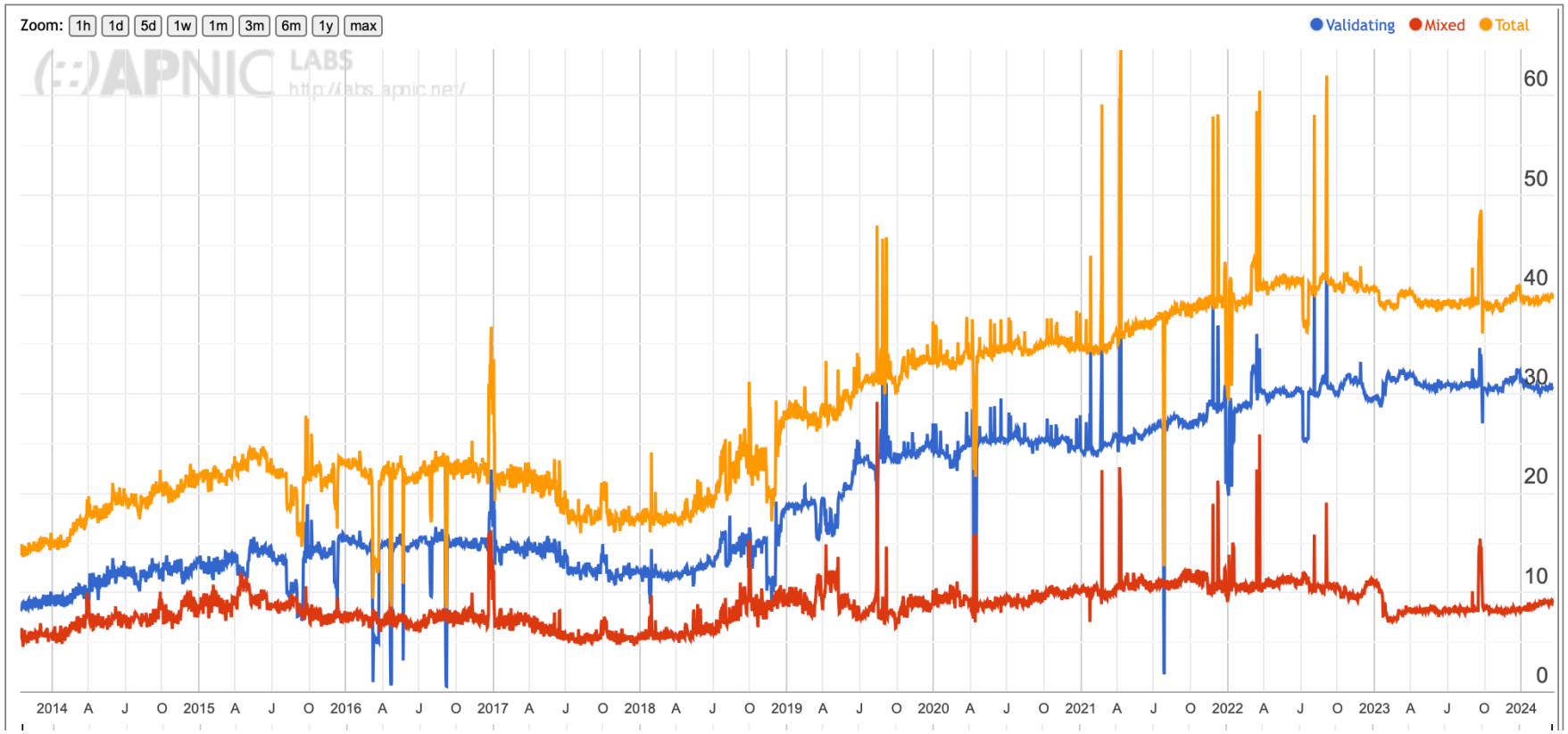
ASN	AS Name	IPv6 Capable	IPv6 Preferred	Samples
AS7713	TELKOMNET-AS-AP PT Telekomunikasi Indonesia	32.20%	31.40%	5,732,645
AS23693	TELKOMSEL-ASN-ID PT. Telekomunikasi Selular	9.07%	8.94%	3,717,007
AS4761	INDOSAT-INP-AP INDOSAT Internet Network Provider	0.43%	0.42%	2,114,387
AS24203	NAPXLNET-AS-ID PT XL Axiata	21.88%	21.37%	1,790,422
AS45727	THREE-AS-ID Hutchison CP Telecommunications, PT	0.03%	0.03%	948,171
AS18004	WIRELESSNET-ID PT WIRELESS INDONESIA WIN	0.52%	0.50%	822,047
AS9341	ICONPLN-ID-AP-ISP PT INDONESIA COMNETS PLUS	0.03%	0.02%	500,372
AS63859	MYREPUBLIC-AS-ID PT. Eka Mas Republik	0.03%	0.02%	376,249
AS23700	FASTNET-AS-ID Linknet-Fastnet ASN	0.04%	0.03%	365,514
AS17451	BIZNET-AS-AP BIZNET NETWORKS	21.06%	20.45%	272,378
AS17670	MNCKABELMEDIACOM-ID PT. MNC Kabel Mediacom	0.03%	0.03%	166,754
AS135478	ASN-CBNBROADBAND PT. Cyberindo Aditama	28.01%	27.51%	161,997

What about DNSSEC Use?

- Can we use the same platform to measure the proportion of users who sit behind DNS resolvers that perform DNSSEC validation?

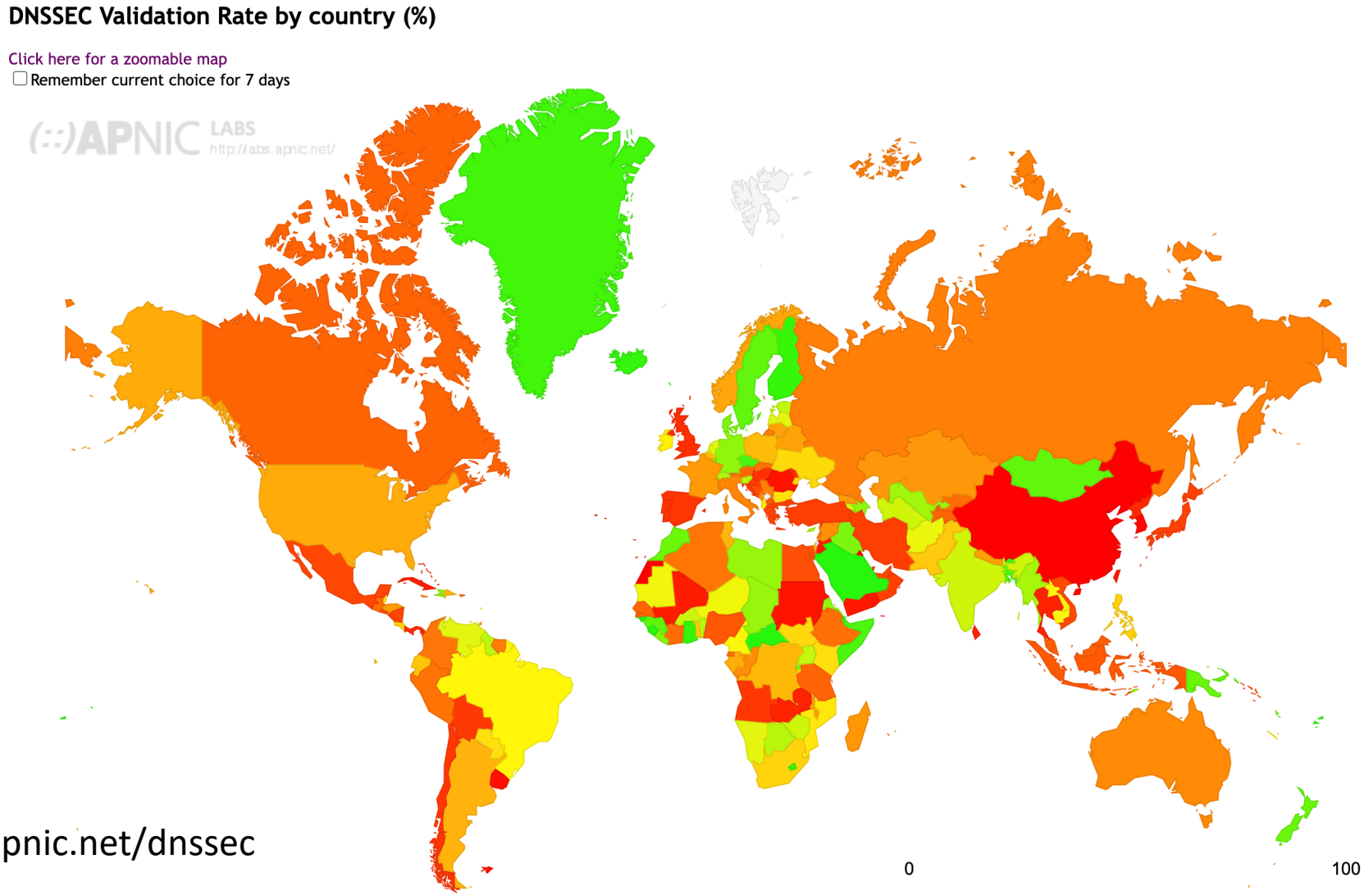
DNSSEC Results

Use of DNSSEC Validation for World (XA)



https://stats.labs.apnic.net/dnssec/XA

DNSSEC Results



https://stats.labs.apnic.net/dnssec

DNSSEC Validation in Indonesia

ASN	AS Name	DNSSEC Validates	Partial Validation	Samples
AS7713	TELKOMNET-AS-AP PT Telekomunikasi Indonesia	13.70%	2.30%	97,284
AS23693	TELKOMSEL-ASN-ID PT. Telekomunikasi Selular	4.49%	1.03%	63,363
AS4761	INDOSAT-INP-AP INDOSAT Internet Network Provider	4.47%	1.88%	35,604
AS24203	NAPXLNET-AS-ID PT XL Axiata	4.11%	1.26%	31,281
AS45727	THREE-AS-ID Hutchison CP Telecommunications, PT	21.90%	8.83%	16,097
AS18004	WIRELESSNET-ID PT WIRELESS INDONESIA WIN	0.40%	0.03%	14,245
AS9341	ICONPLN-ID-AP-ISP PT INDONESIA COMNETS PLUS	13.62%	3.96%	8,340
AS63859	MYREPUBLIC-AS-ID PT. Eka Mas Republik	8.86%	3.06%	6,403
AS23700	FASTNET-AS-ID Linknet-Fastnet ASN	53.76%	12.57%	5,586

https://stats.labs.apnic.net/dnssec/ID

Measurement Projects

- IPv6 Performance (connection reliability and relative speed)
- IPv6 Fragmentation
- IPv6 Extension Header loss Rates (HBH and DST)
- DNS: Use of ECDSA and EDDI DNSSEC signing algorithms
- DNS: Fragmentation Drop (and TCP support)
- DNS resolver use profile (use of open DNS resolvers)
- DNS KSK roll probes (RFC8509)
- Support for QUIC use (HTTP/3)
- Support for Route Origination Validation
- Zombies and tracking

https://stats.labs.apnic.net

Server Side Measurement

- This approach complements client side measurements (CAIDA's ARK, RIPE NCC's Atlas) and network-level internal measurements by using a large scale server side measurement platform
- In this form of server-side measurement the client does what clients always do - fetch URLs
- We can test particular client behaviours and network behaviours by deliberately altering the server-side behaviour and triggering the behaviour in a measured behaviour
- The benefit of this approach is that rather than measuring the effect and inferring the cause, in this approach we trigger a cause and then correlate the observed outcomes against the known cause.

Thanks!

Measurement Reports at APNIC Labs:
https://stats.labs.apnic.net