September 2024

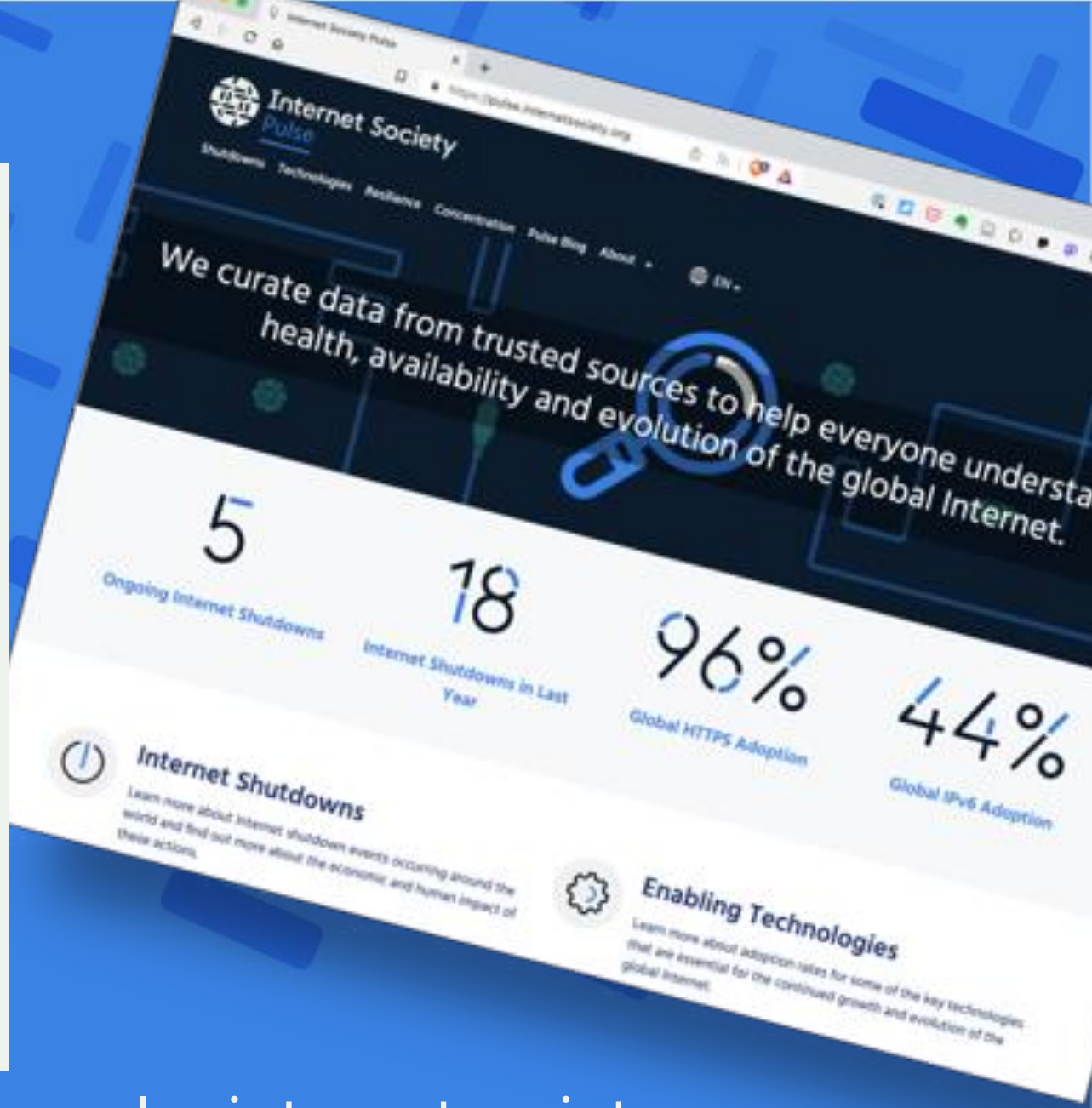
Let's Measure Content Locality in Pakistan

Aftab Siddiqui
siddiqui@isoc.org

- Launched December 2020.
- We curate Internet measurement data from trusted sources to help everyone gain deeper, data-driven insight into the Internet.

Trusted data from multiple sources:

- **Benefit**: Helps to assess whether efforts to ensure that the Internet remains open, globally connected, secure, and trustworthy are working.
- **Benefit:** Allows policymakers, researchers, journalists, network operators, civil society groups, and others to better understand the health, availability, and evolution of the Internet.



pulse.internetsociety.org

But..

- I'll be sharing insights that are not available through the Pulse platform (at the moment), as they derive from dataset we're examining as part of our research team.

- Data analysis is a fascinating process—it can range from uncovering trivial details to discovering invaluable information.

- Through our analysis, we aim to transform these complex datasets into meaningful information that can benefit all of us.

Presentation vs Interpretation Bias

- Bias exists in all types of data analysis and is difficult to eliminate.
- Bias can occur at any stage of the process, and it can impact the validity and reliability of your findings, leading to “misinterpretation” of data.

- **Interpretation Bias**: If the analyst is making a conclusion that are not directly supported by the dataset.

- **Presentation Bias**: If dataset are presented without interpretation, there is a high likelihood that readers may draw misleading conclusions.

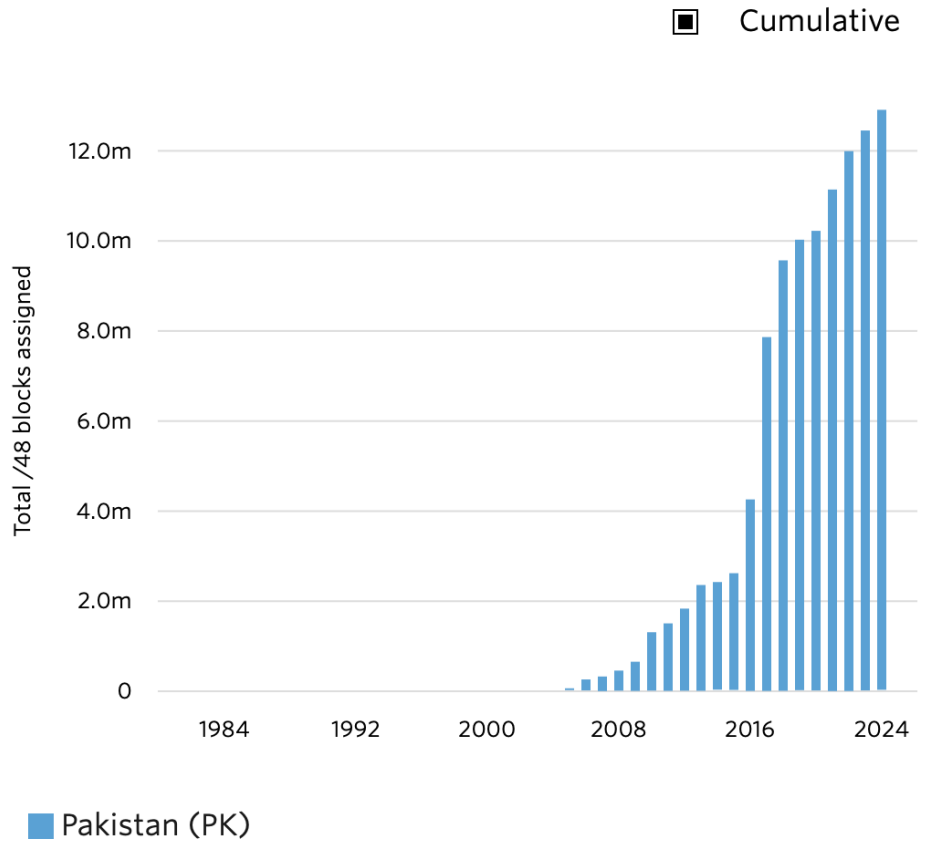
- *Let's address some presentation biases.*

Resource Delegation

https://rex.apnic.net

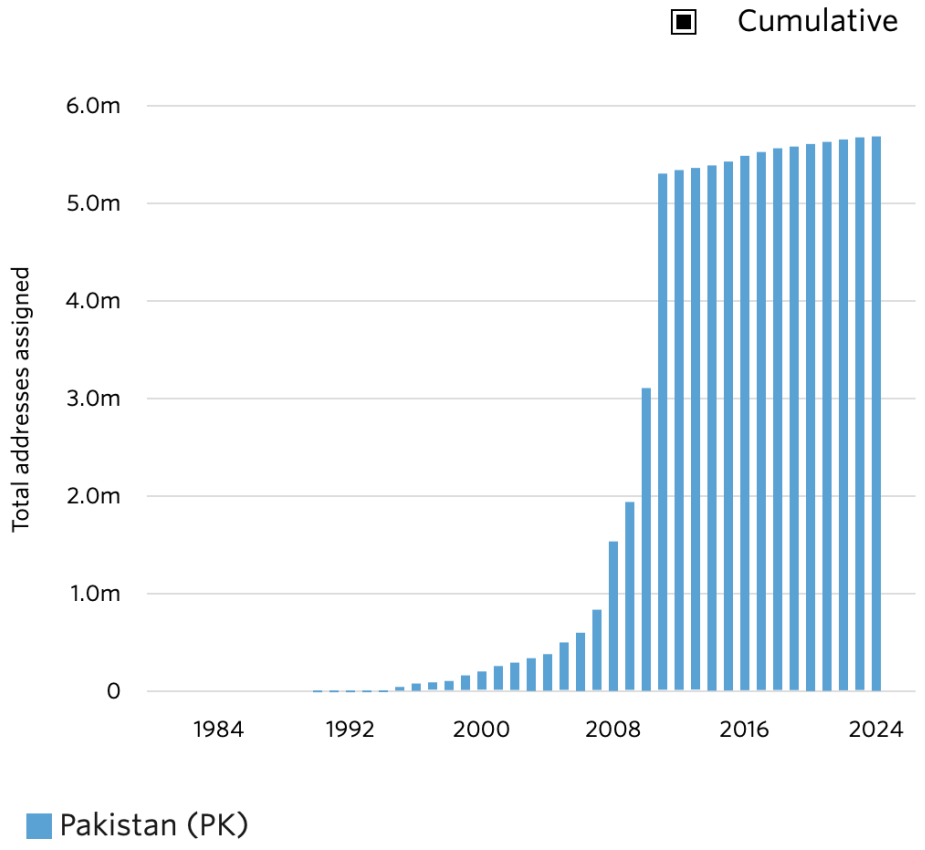
IPv6 for Pakistan PK

/48 prefixes	/32 prefixes	Delegations
12,910,750	197	320

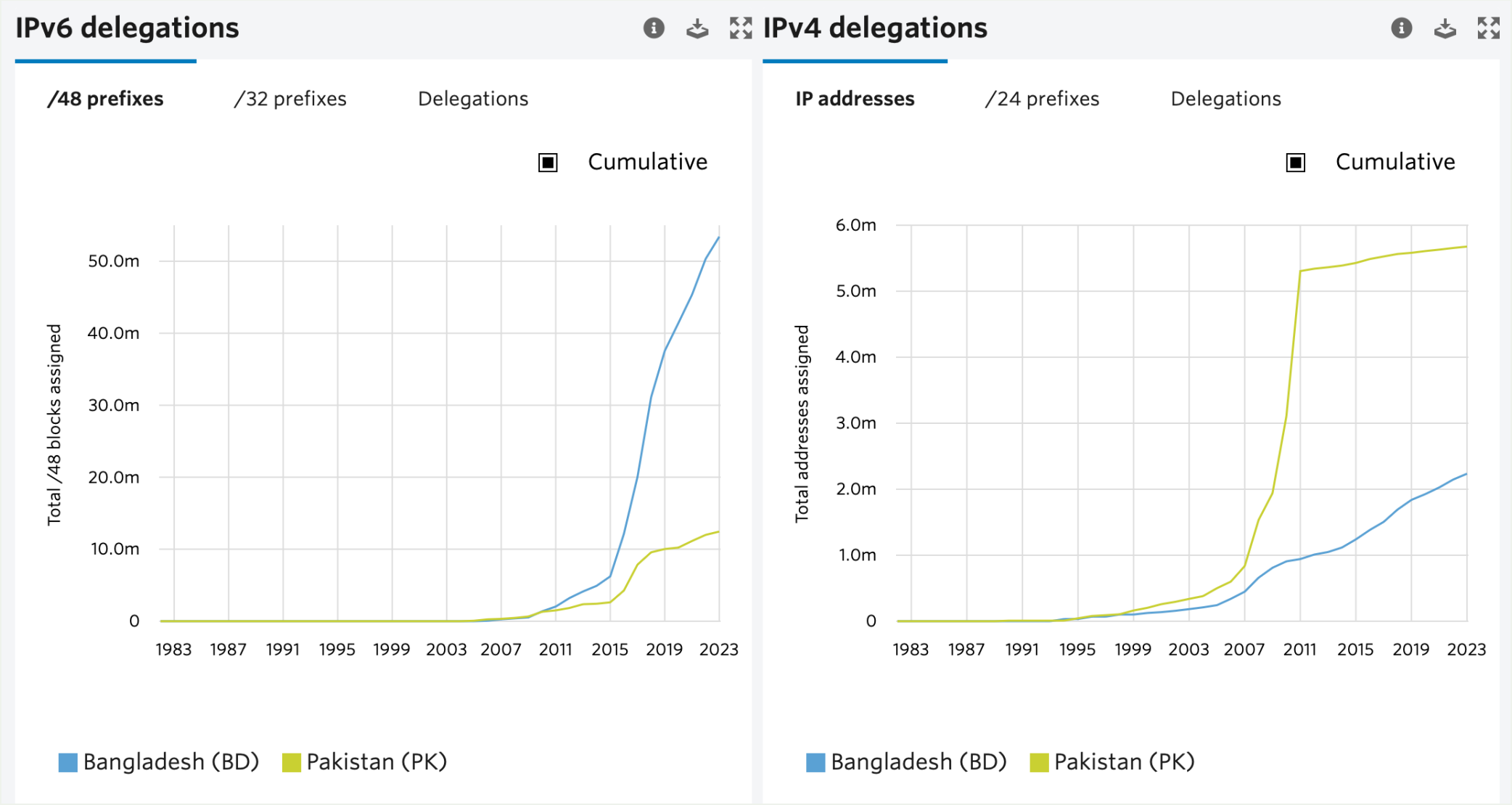


IPv4 for Pakistan PK

IP addresses	/24 prefixes	Delegations
5,686,272	22,212	776



Resource Delegation

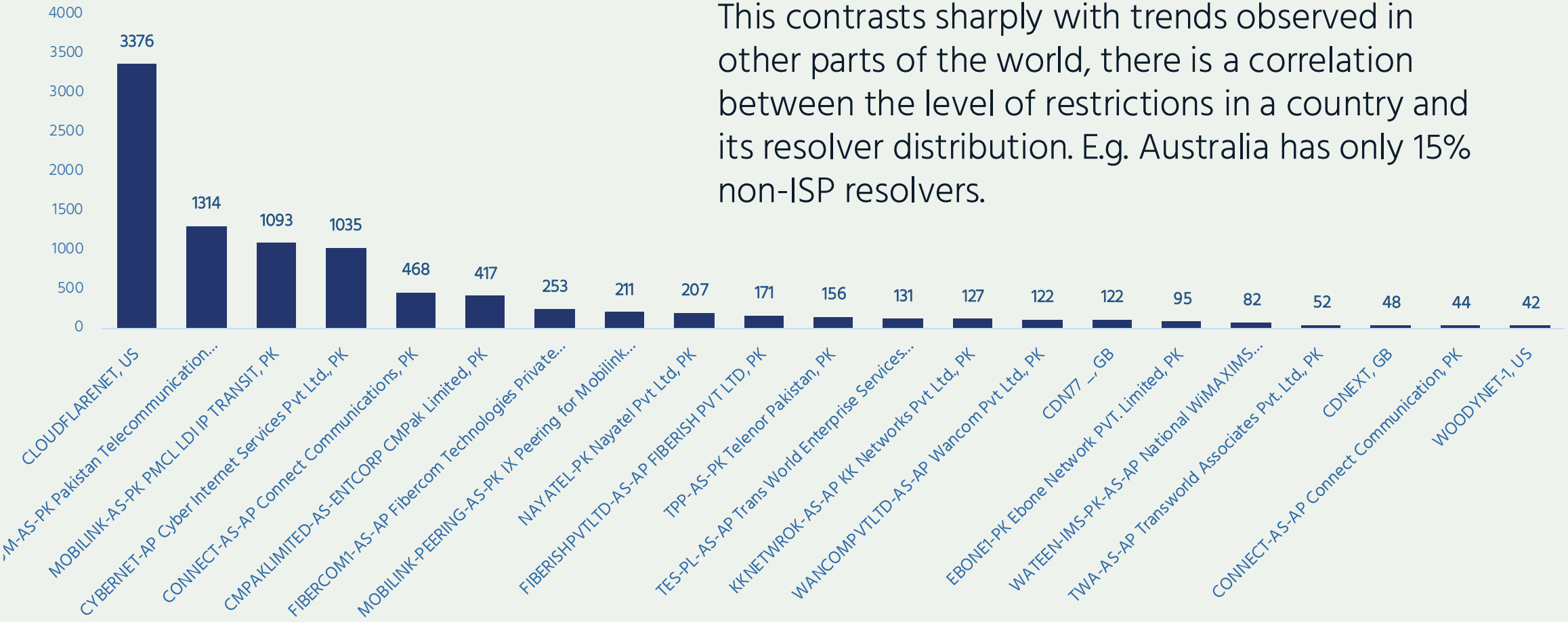


DNS Diversity

DNS Diversity

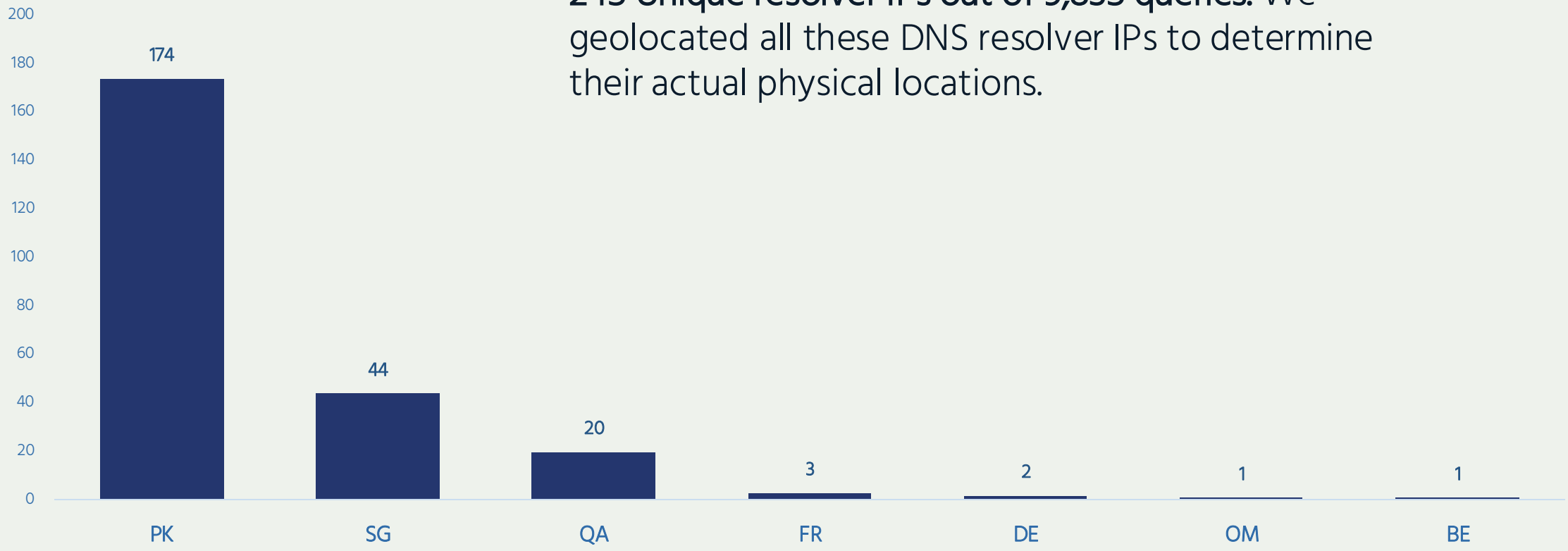
- DNS is fundamental to Internet functionality; nothing more on that!
- Proximity and Providers:
  - Critical to assess how far away the DNS service is located.
  - Important to identify who is providing the DNS service—whether it's a local service provider or a public DNS provider such as Google Public DNS, Cloudflare 1.1.1.1, Quad9, or OpenDNS, etc.
- Usage Patterns: Understanding what DNS services people are using is important.
- Contextual Relevance:
  - In open and democratic environments like Australia, New Zealand, the choice of DNS might not seem critical but still presents an informative landscape.
  - In other regions, the choice of DNS service can provide significant insights into local internet governance and access restrictions.

DNS Diversity



Out of 10k DNS queries, 3845 resolvers originated outside of the ISP network, which is almost 40%. This contrasts sharply with trends observed in other parts of the world, there is a correlation between the level of restrictions in a country and its resolver distribution. E.g. Australia has only 15% non-ISP resolvers.

DNS Diversity

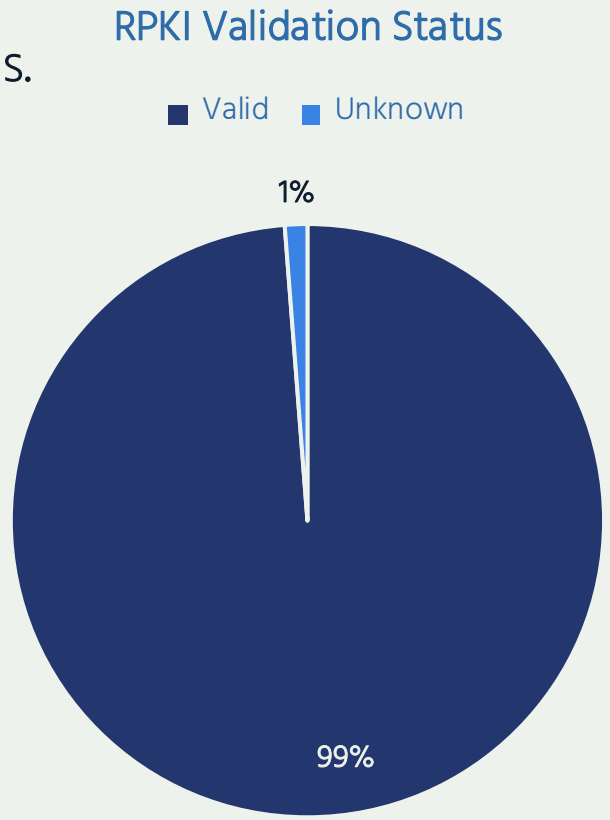


245 Unique resolver IPs out of 9,835 queries. We geolocated all these DNS resolver IPs to determine their actual physical locations.

DNS Diversity

245 Unique resolvers out of 9,835 queries.
We verified the RPKI validation status for these 245 IP addresses.
Only 3 resolvers with unknown RPKI status

149.102.250.14	CDNEXT, GB
119.73.122.254	TES-PL-AS-AP Trans World Enterprise Services Private Limited, PK
103.125.69.173	CONNECT2B-AS-PK Broadband ISP, FTTH and Cable Service Provider, PK



Lack of a valid ROA can expose DNS resolvers to BGP route hijacking.

Where is the content?

Content Locality

Having content hosted locally is important, especially for a country like Australia. As a vast island nation, having data hosted within the country is more efficient than relying on servers across multiple continents. Not every island is as lucky as we are!

We use Tranco List (lists of popular domains), which is a combination of Google CrUX, Cloudflare Radar, Farsight, Majestic and Cisco Umbrealla.

For this work we are only focusing on .au cctld. We picked 35,000 domains from the Full Tranco List (1 million domains).

Content Locality

Pakistan

Popular Content Locality

A measure of how much locally popular web content is hosted in-country or in-region

36%

Regional

Rank: 37

51%

Asia avg.

See details

Bangladesh

Popular Content Locality

A measure of how much locally popular web content is hosted in-country or in-region

69%

Regional

Rank: 13

51%

Asia avg.

See details

India

Popular Content Locality

A measure of how much locally popular web content is hosted in-country or in-region

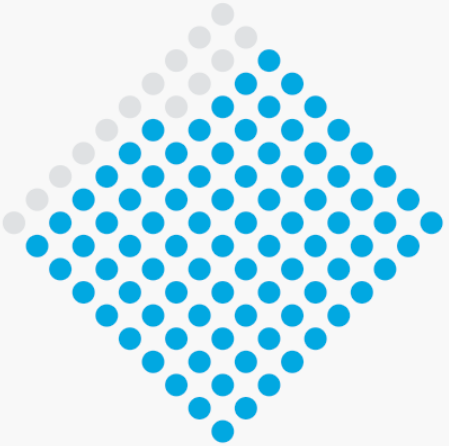
86%

Regional

Rank: 3

51%

Asia avg.

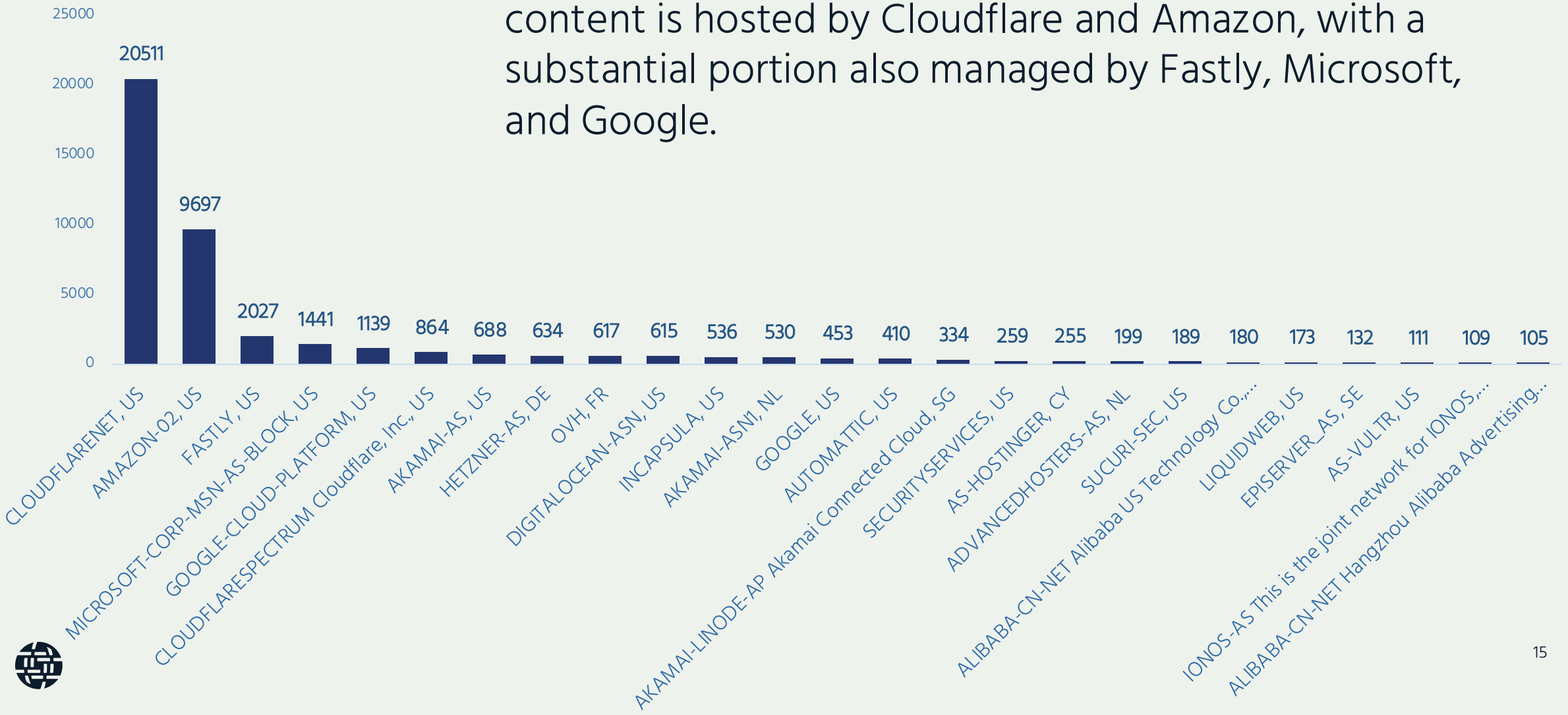


See details

https://pulse.internetsociety.org/reports

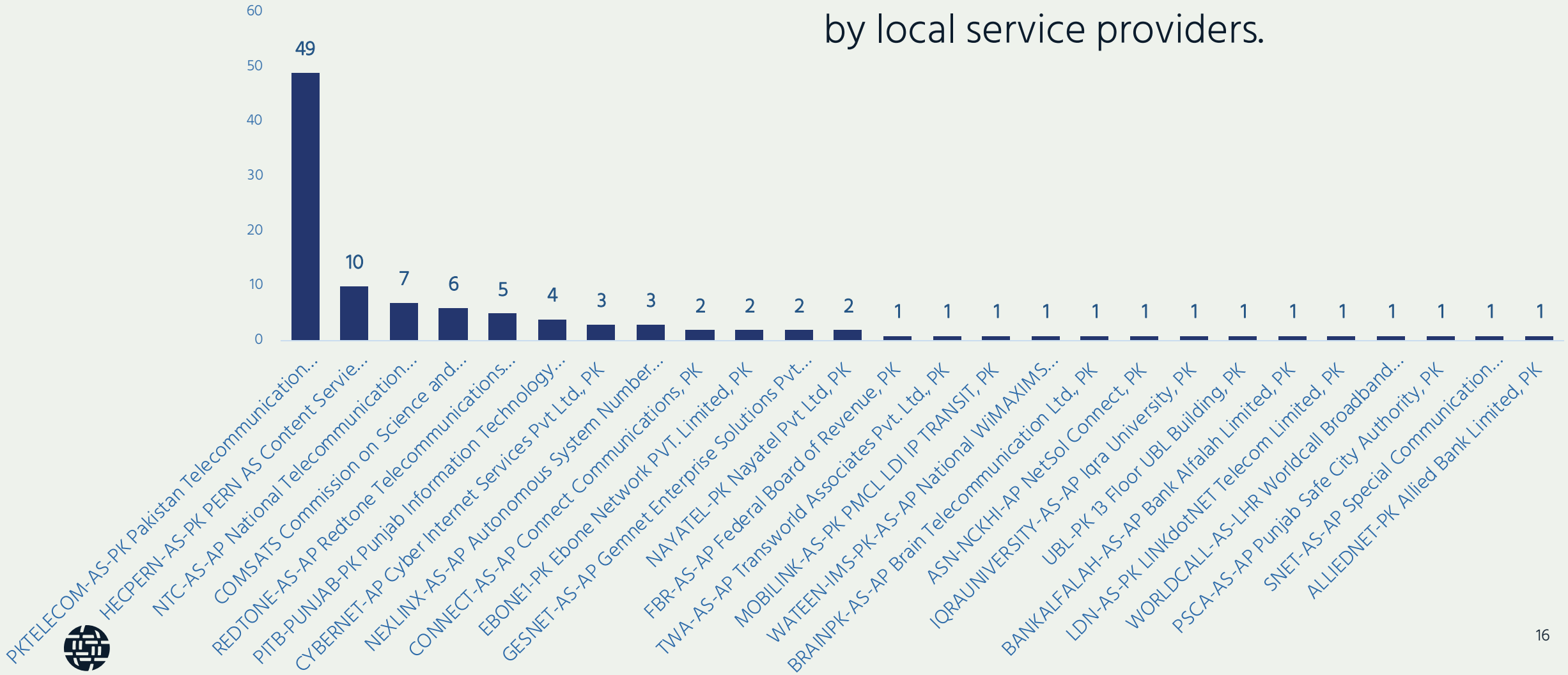
Content Locality

We checked top 50,000 websites and majority of content is hosted by Cloudflare and Amazon, with a substantial portion also managed by Fastly, Microsoft, and Google.



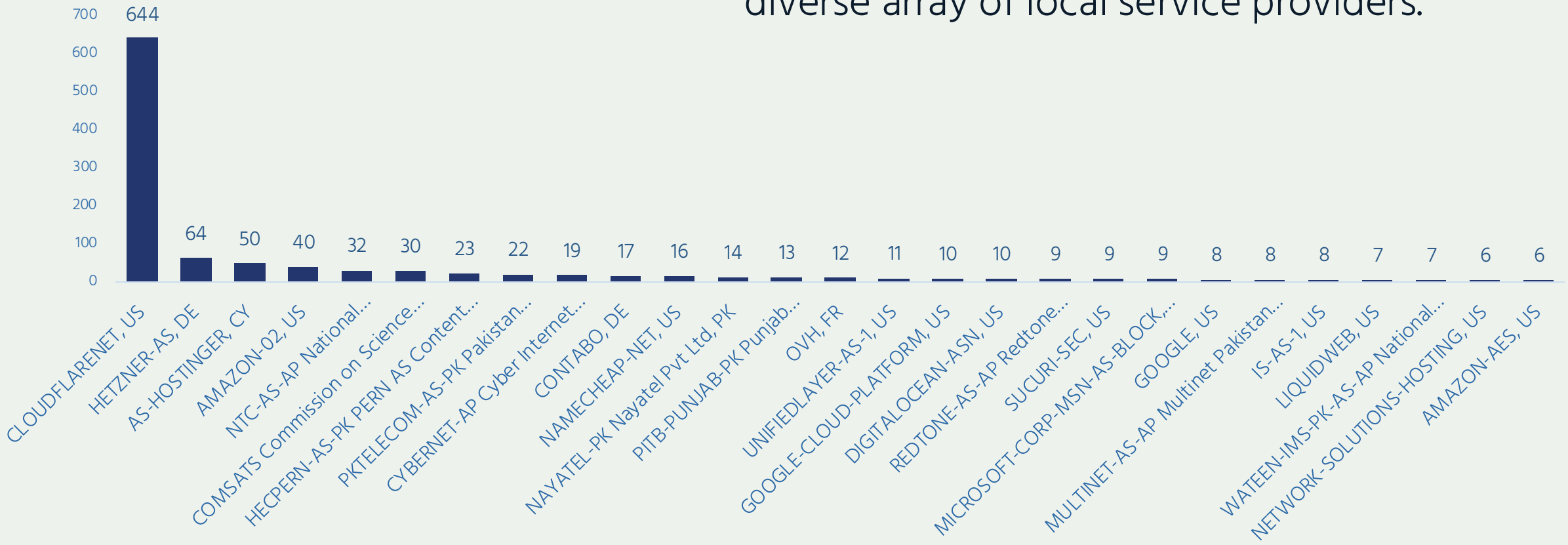
Content Locality – Local Hosting

Out of 50K websites, only 109 are hosted by local service providers.



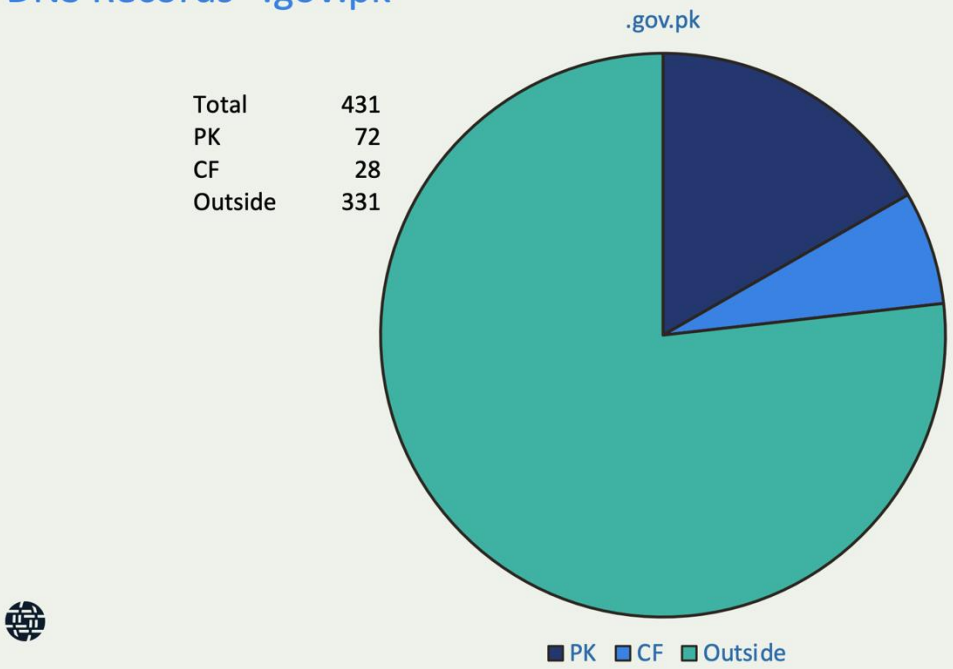
Content Locality – .pk

Out of 1289 highest visited .pk websites, most are hosted by major CDN providers, while the remainder are supported by a diverse array of local service providers.



SANOG 35 [2020] – Karachi, Pakistan

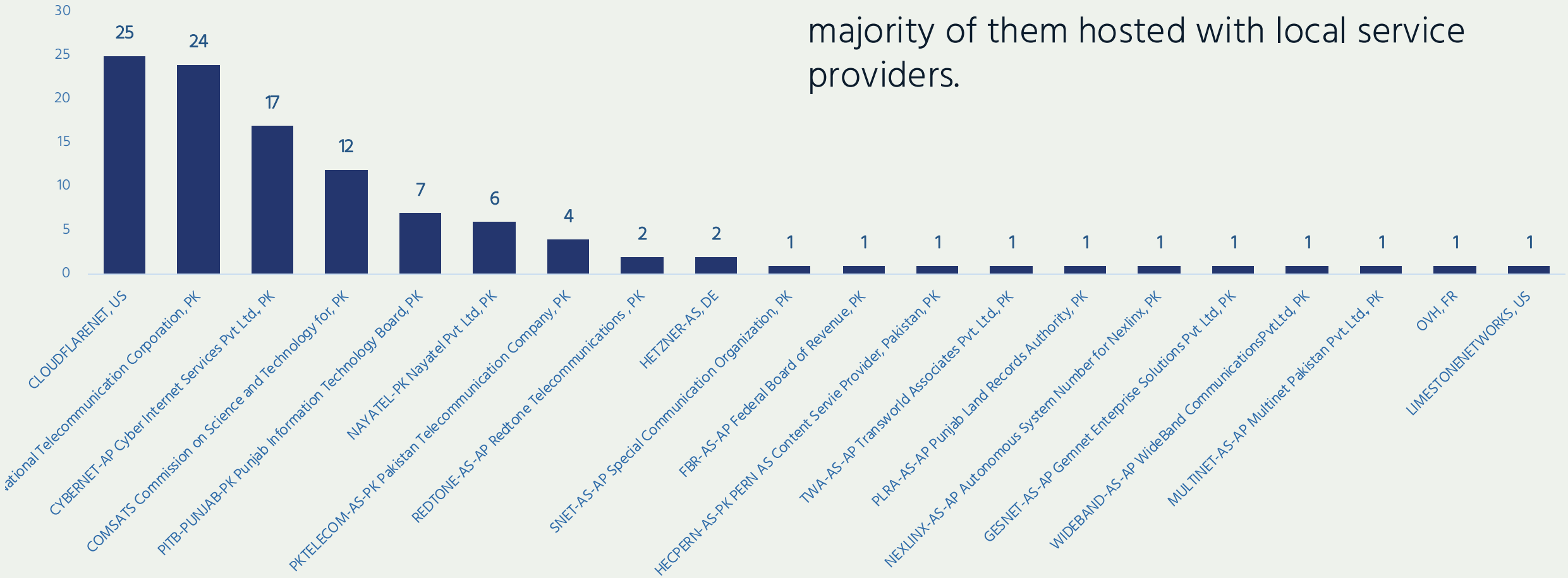
- Presented "Traffic Patterns in Pakistan" back in 2020
- Data was based on ccTLD records of .gov.pk.
- We added sub-domains at that time
- No geo-location data was checked at that time.
- Majority of .gov.pk had non-local A records.



https://sanog.org/resources/sanog35/SANOG35-Conference-Traffic-Patterns-PK-Aftab.pdf

Content Locality – gov.pk

Out of 110 .gov.pk websites, a big portion of them are behind Cloudflare CDN but majority of them hosted with local service providers.



Why lack of local hosting?

IXP Landscape

https://pulse.internetsociety.org/en/ixp-tracker/

India

The total number of IXPs in operation in India, as of October 2024.

45

Active IXPs

13.81 %

Proportion of the local Internet that can be reached through IXPs in this country.

Bangladesh

The total number of IXPs in operation in Bangladesh, as of October 2024.

11

Active IXPs

6.23 %

Proportion of the local Internet that can be reached through IXPs in this country.

Pakistan

The total number of IXPs in operation in Pakistan, as of October 2024.

2

Active IXPs

0.53 %

Proportion of the local Internet that can be reached through IXPs in this country.

Savings - Assumptions

Let's assume, If the total Internet usage of Pakistan is 5 Tbps (terabits per second) per month which is dependent on International bandwidth, and they are able to host 5% of the content locally:

Saved bandwidth=5% x 5 Tbps= 5/100 x 5 = 0.25 Tbps or 250 Gbps

Let’s assume, 1 Mbps costs $1

Saved Forex = 250,000 Mbps x 1 USD= 250,000 USD per month.

Above calculations have used ballpark values to estimate the bandwidth savings and overall internet usage.

Limitations

Limitations

- The data is pulled from external public sources, not always up-to-date.
- Without in-country measurements, it's difficult to validate the data.
  - RIPE Atlas and OONI are doing great work in this area, but more is needed.
- Some of the data undergoes processing, normalization, and weighing, we use a methodology that is reproducible.
  - You can see raw numbers via API. Email us for access pulse@isoc.org
- The idea is to help decision makers recognize gaps and weaknesses to conduct further study into validating these and work towards addressing them.

Take aways

- It's crucial to thoroughly understand the measurements to assess whether they convey the real picture or relevant or accurate picture
- Every data representation carries certain biases, which are manageable as long as they are acknowledged and addressed.
- *Final word: We are not doing good in local content hosting as a technical community!*

Subscribe, Review, Contribute

Subscribe to the Pulse newsletter

Contribute to Pulse

pulse@isoc.org

Review the Pulse IRI methodology

Thank you

Aftab Siddiqui
siddiqui@isoc.org