PUSHING THE FRONTIER:
Exploring the African Web Ecosystem

Rodérick Fanou[1], Gareth Tyson[2],
Pierre Francois[3], and Arjuna Sathiaseelan[4]

[1]IMDEA Networks Institute & UC3M;
[2]Queen Mary University of London;
[3]CISCO Systems; [4]University of Cambridge

roderick.fanou@imdea.org;
gareth.tyson@qmul.ac.uk

The Paper

Pushing the Frontier: Exploring the African Web Ecosystem

Rodérick Fanou,[1,2] Gareth Tyson,[3] Pierre Francois[4] and Arjuna Sathiaseelan[5]
[1]IMDEA Networks Institute, [2]Universidad Carlos III de Madrid,
[3]Queen Mary University of London, [4]Cisco Systems, Inc. [5]University of Cambridge
roderick.fanou@imdea.org, gareth.tyson@qmul.ac.uk,
pifranco@cisco.com, arjuna.sathiaseelan@cl.cam.ac.uk

ABSTRACT

It is well known that Africa's mobile and fixed Internet infrastructure is progressing at a rapid pace. A flurry of recent research has quantified this, highlighting the expansion of its underlying connectivity network. However, improving the infrastructure is not useful without appropriately provisioned services to utilise it. This paper measures the availability of *web content infrastructure* in Africa. Whereas others have explored web infrastructure in developed regions, we shed light on practices in developing regions. To achieve this, we apply a comprehensive measurement methodology to collect data from a variety of sources. We focus on a large content delivery network to reveal that Africa's content infrastructure is, indeed, expanding. However, we find much web content is still served from the US and Europe. We discover that many of the problems faced are actually caused by significant inter-AS delays in Africa, which contribute to local ISPs not sharing their cache capacity. We discover that a related problem is the poor DNS configuration used by some ISPs, which confounds the attempts of providers to optimise their delivery. We then explore a number of other websites to show that large web infrastructure deployments are a rarity in Africa and that even regional websites host their services abroad. We conclude by making suggestions

project [2], as well as on underlying edge connectivity, with projects such as Liquid Telecom, Project Loon and Google Link deploying both wireless and wireline connectivity [22, 13, 1],

Despite these positive steps, challenges still remain. Most prominently, several studies have revealed content as the dominant component of network traffic [11]. Yet the lack of service infrastructure (*e.g.*, web servers) in Africa means that both mobile and wireline users must often fetch content from the other side of the world [20]. Hence, we argue that researchers and engineers should begin to place more focus on both underlying connectivity and content infrastructure (*e.g.*, web servers, caches) in the region. With this in mind, many large companies have begun to deploy content infrastructure in Africa. Google is perhaps the most famous in this regard [1]. As of yet, however, there is little evidence of how these companies approach the challenge.

Several recent efforts have shed light on web infrastructures worldwide [23, 27, 40, 18, 14, 17, 42, 20]. But, they have not (*i*) focussed on underdeveloped countries/regions; or (*ii*) explored if worldwide results apply to underdeveloped regions. This leaves critical questions unanswered, largely driven by the unusual make-up of African Internet and web infrastructures when compared to more developed regions.

R. Fanou, G. Tyson, P. Francois, and A. Sathiaseelan (2016) *Pushing the Frontier: Exploring the African Web Ecosystem.* In: The 25th International World Wide Web Conference (WWW 2016), 11-15 April 2016, Montreal, Canada.

Background: A brief intro to Web CDNs

Let's suppose that she wants to access a CDN …

NOW

A user (in Africa) wants to access to www.website.com

1- She types www.website.com it in the browser

2- Her Computer sends a DNS lookup to the closest DNS server

3- The address www.website.com is translated into a set of IPs, those of her closest caches which are sent back to the computer

4- Depending on the content of the webpage, the computer may query one or more IP caches

5- They respond with parts of/all the webpage content

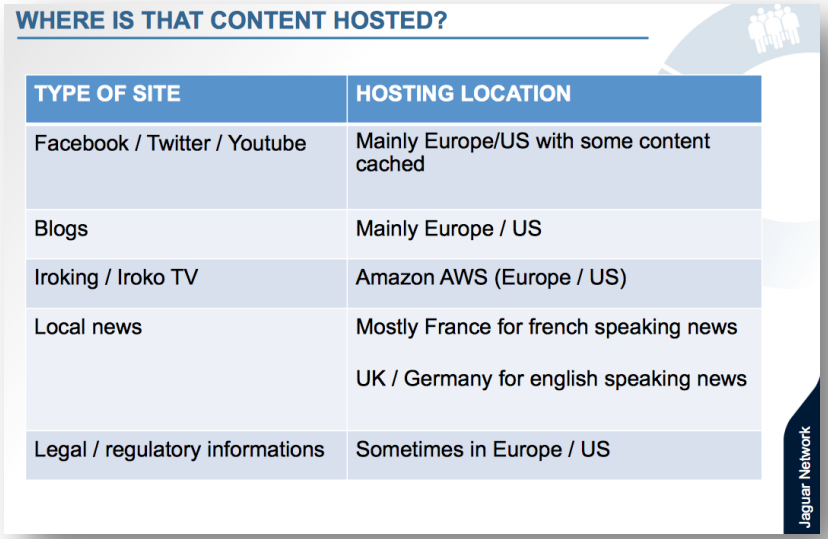
6- The user can open to the webpage

Conclusion: The closer (in terms of delay, i.e. geographically) are the caches the user are redirected to, the better is her QoS while accessing the website

Related Work

- **IXPs deployments of particular interest in the African region [1, 2]**
- **Zaki et *al.* [3] investigated web performance in Ghana**
  - Key bottlenecks include slow DNS resolution
  - Lack of content caching
- **Calder et *al.* [4] studied Google infrastructure**
  - Enumerated caches IPs & found their locations
  - Studied its growth & matched users to clusters
- **Otto et *al.* [5] examined the role of DNS in the redirection process**
- **Su et *al.* [6] found Akamai redirects clients are based on active networks conditions**

Background: Objective



WHERE IS THAT CONTENT HOSTED?

TYPE OF SITE	HOSTING LOCATION
Facebook / Twitter / Youtube	Mainly Europe/US with some content cached
Blogs	Mainly Europe / US
Iroking / Iroko TV	Amazon AWS (Europe / US)
Local news	Mostly France for french speaking news UK / Germany for english speaking news
Legal / regulatory informations	Sometimes in Europe / US

Jaguar Network

[7] Mathieu Paonessa (Jaguar Networks), Future of Content, http://isoc-ny.org/afpif2014/AfPIF2014_Future_of_Content.pdf, AFPIF2014

Top 20	Website	Server Location
1. IGIHE	www.igihe.com	USA
2. Umuseke	www.umuseke.rw	USA
3. Kigali Today	www.kigalitoday.com	USA
4. Umuryango	www.umuryango.com	USA
5. Inyarwanda	www.inyarwanda.com	USA
6. Tohoza	www.tohoza.com	Switzerland
7. The New Times	www.newtimes.co.rw	USA
8. Imali	www.imali.biz	USA
9. Rwanda Directorate General of Immigration and Emigration	www.migration.gov.rw	Rwanda
10. University of Rwanda	www.nur.ac.rw	Rwanda
11. College of Science and Technology, University of Rwanda	www.kist.ac.rw	Rwanda
12. Rwanda Broadcasting Agency	www.orinfor.gov.rw	Rwanda
13. Living in Kigali	www.livinginkigali.com	USA
14. Ubugingo	www.ubugingo.com	USA
15. Rumalex	www.rumalex.net	Germany
16. Zion Temple	www.ziontemple.rbm.tv	USA

[8] M. Kende and K. Rose, Promoting Local Content Hosting to Develop the Internet Ecosystem. ISOC Report, 2015

Investigate the way users in Africa access the web (African Web Ecosystem), since the outcome has serious impacts on both mobile and wireline performance.

Possible Methodology Approaches

- **Straightforward method:** record in a public repository, data from all (mobile and wireline) users in Africa for websites they access (quite difficult to achieve as of today)
- **Alternative:** Ask all local ASes to randomly launched from their servers, scripts that locally collect this data (quite difficult to achieve as of today – Some reasons: Networks' trust, frequent current cuts)
- **Another one:** Use an existing measurement infrastructure
  - Planet Lab (in 2015, 8 monitors)
  - Archipelago (in 2015, 5 monitors)
  - RIPE Atlas (in 2015, 379 probes in 170 ASes across 45 countries /58) : **the best option**

Our choice for the measurements infrastructure: RIPE Atlas

- **RIPE Atlas probes:**
  - The largest measurement platform in Africa
  - DNS measurements, traceroutes, pings, HTTP requests, etc
  - Make our measurements publicly available

RIPE Atlas probes

RIPE Atlas probes in African networks

4 Main Tasks

- **Task1:** Mapping a large Content Provider from African ASes
- **Task2:** Compare DNS queries results to EDNS0 results
- **Task3:** Investigate DNS in African ASes
- **Task4:** Comparing top Global Providers Infrastructures serving Africa to those of top Regional ones

Methodology Overview: Summary (1)

- **Collect all IP prefixes allocated by AFRINIC (African prefixes)**
  - 3,082 prefixes of various length allocated by AFRINIC
- **Discover all content servers/caches that serve them**
  - EDNS0 Client-Subnet Probes for a week
  - 28,387,226 RIPE Atlas DNS Probes
  - Targets: top regional, top global Alexa websites
- **Geolocate caches** (Improved version of geoloc method in [1])
  - Cross-checking 10 data sources: OIM, MM, TC, RDNS, WHOIS & RP, AR, AF, LAC, AP assignments databases
  - 5 sets of pings measurements to tie break between multiple CCs
  - Checking Speed of light violation (Extended version - Ongoing): 100 random RIPE Atlas probes worldwide per cache IP

Methodology Overview: Summary (2)

- **IP Geoloc Summary**

DB	3,105 GGCs IPs Coverage	Trust	144 DNS resolvers Coverage	Trust
OIM	0.45%	100%	0%	N/A
RDNS	8.27%	93.77%	0%	N/A
MM	98.29%	89.54%	100%	98.61%
RP	10.04%	75.32%	12.5%	88.89%
AF	35.81%	93.07%	81.25%	94.02%
AP	2.58%	100%	0.69%	100%
AR	10.66%	98.49%	22.91%	87.88%
LAC	0%	N/A	0%	N/A
TC	98.97%	90.34%	100%	95.13%
WHOIS	97.93%	47.41%	94.44%	8.82%

Table 1: *Comparison of Geolocation DBs for both GGCs' and DNS resolvers' IPs. N/A stands for Not Applicable.*

- **Measure paths characteristics**
  - 1,309,151 UDP Paris-traceroutes (February 18 to May, 22 2015)
  - randomly launched from RIPE Atlas probes to GGCs IPs
- **Measure Web performance**
  - HTTP queries (from 242 probes in Europe) with RIPE Atlas probes towards www.google.com & HTTP queries (from 225 RIPE Atlas probes in Africa) towards top global and regional Alexa websites

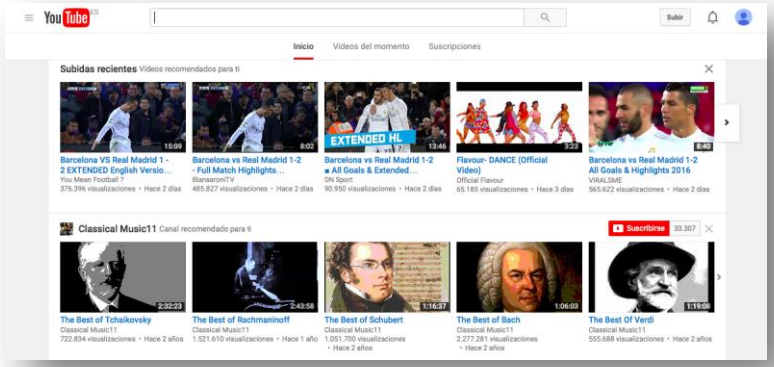
SOME TOP GLOBAL WEBSITES (ALEXA RANKING)

amazon.com (E-commerce)

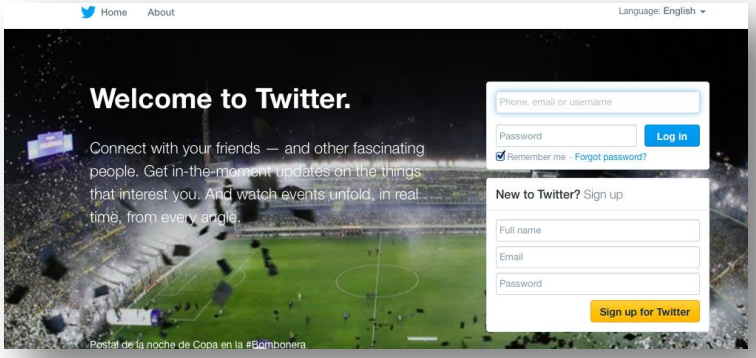
tabao.com (E-commerce)

wikipedia.com (Encyclopedia)

youtube.com (Videos)



twitter.com (Social network)

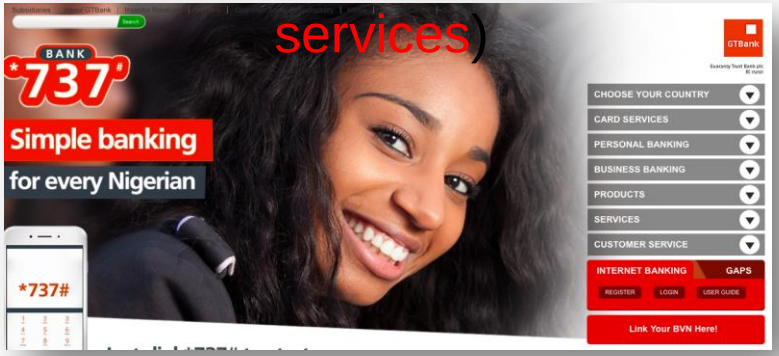


facebook.com (Social network)

SOME TOP REGIONAL WEBSITES (ALEXA & AFRODIGIT RANKING)

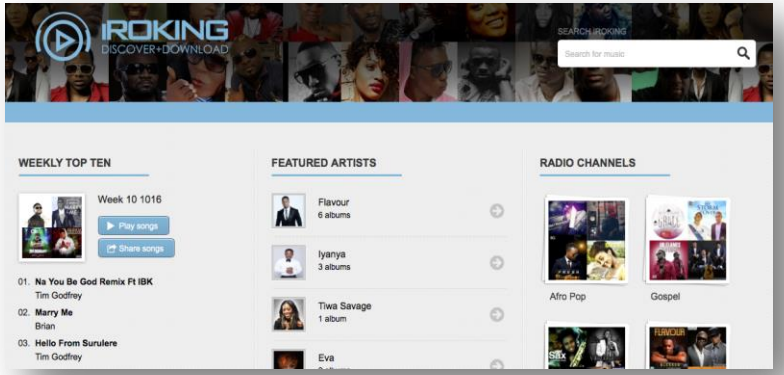
jumia.com (E-commerce)

gtbank.com (Financial services)

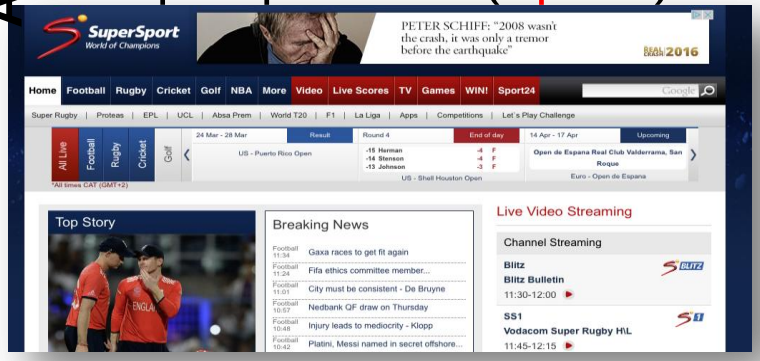


news24.com (News/media)

iroking.com



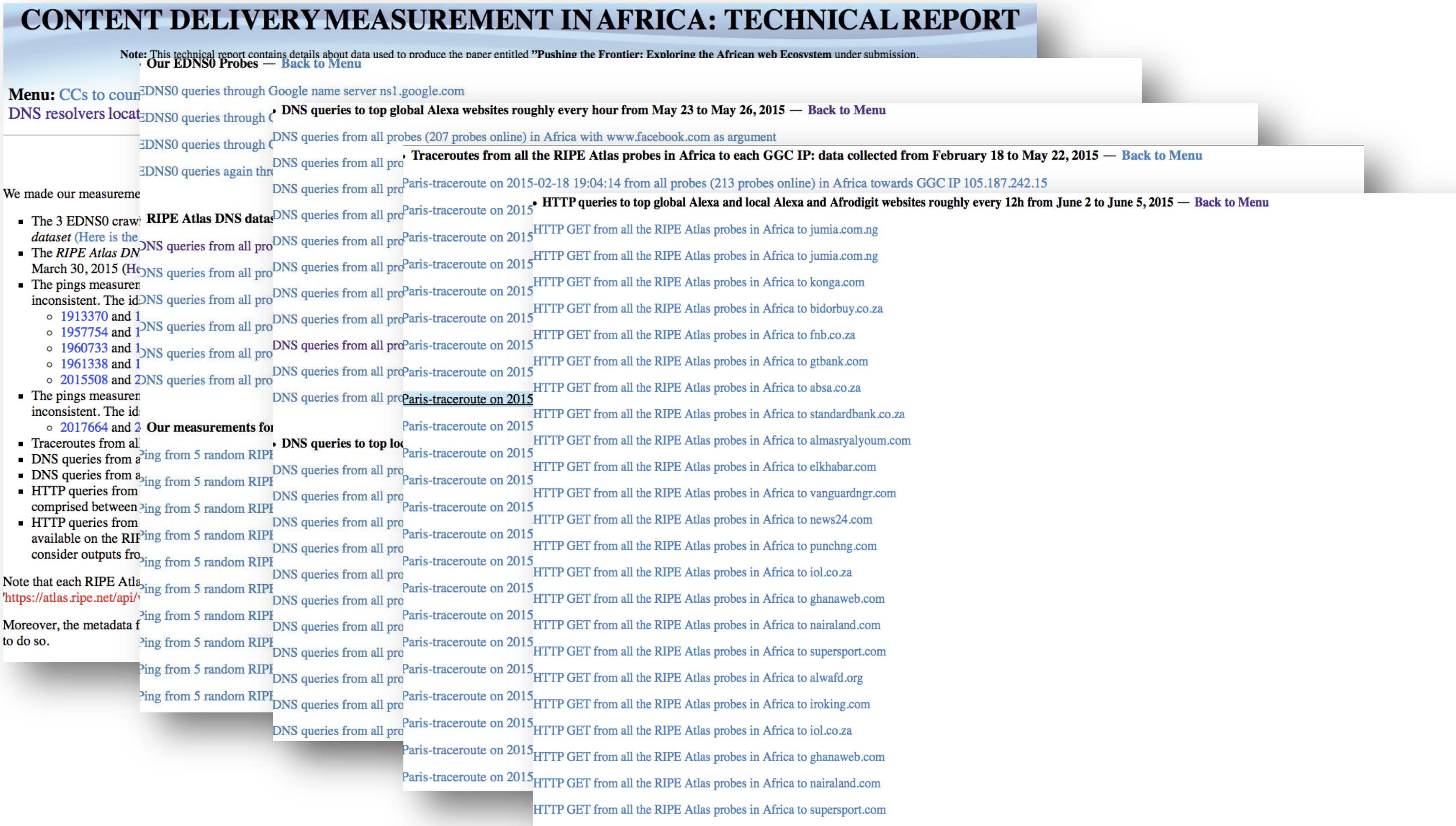
supersport.com (Sports)



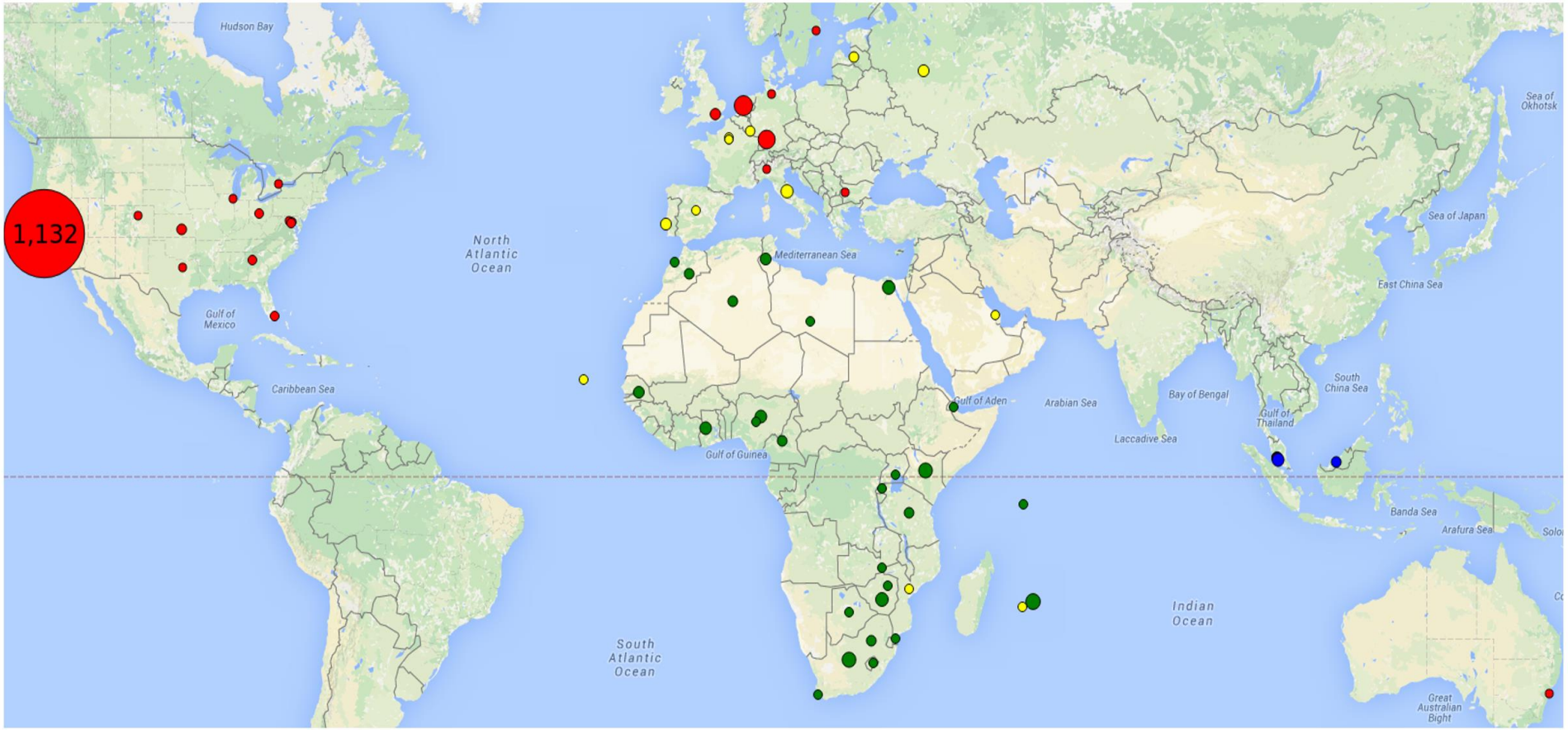
nairaland.com (Online Community)

Data publicly available

R. Fanou, G. Tyson, P. Francois, A. Sathiaseelan, **Technical report: African Content Measurement Campaign**, https://techrep_cdma:PDQ7Rjkj@fourier.networks.imdea.org/external/techrep_cdma/index

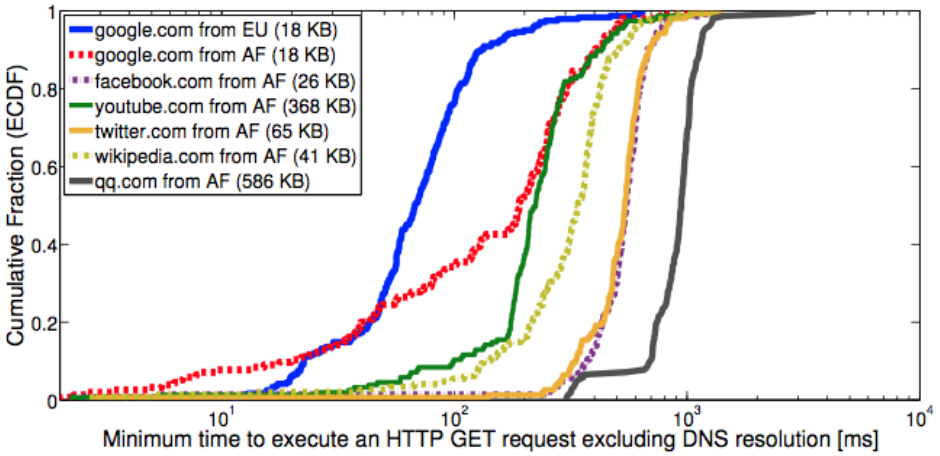


How are African prefixes served by Google?

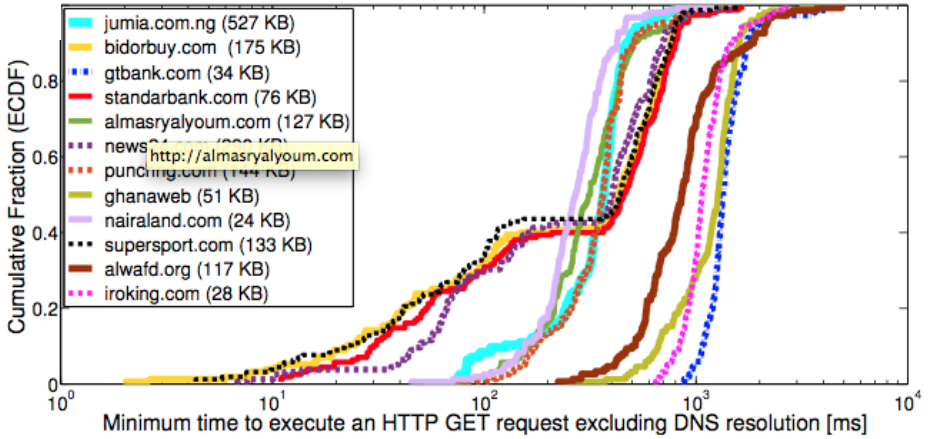


Spread of the 3,120 Caches IPs discovered. Each dot represents a set of IP caches. Its size is proportional to the number IP caches at its location

Websites performance



(a) *Distribution of minimum time to execute an HTTP GET request per probe (ms) from Europe (EU) and Africa (AF) to top global Alexa websites.*



(b) *Distribution of minimum time to execute an HTTP GET request per probe from Africa to selected top local Alexa & Afrodigit websites.*

Any Feedbacks on the Methodology?

- **Any comments on the methodology?**
  - Any idea on how we can agree on where were located 3,120 IP caches and the 157 DNS resolvers?
  - May be Groundtruth on discovered IPs geoloc from CDNs for comparison?
- **Possible Improvements? => frequent assessment of the web African ecosystem**
- **Any idea to improve African measurements infrastruc-ture for enabling deeper studies**
- **Possible to obtain from local operators similar/related datasets to better the analysis?**

References

[1] R. Fanou, P. Francois, E. Aben, On the Diversity of Interdomain in Africa, In PAM 2015

[2] A. Gupta, M. Calder, N. Feamster, M. Chetty, E. Calandro ,and E. Katz- Bassett. Peering at the Internet's Frontier: A First Look at ISP interconnectivity in Africa. In *PAM*, 2014

[3] C. Matt, F. Xun, H. Zi, E. Katz-Basset, H. John, and G. Ramesh. Mapping the Expansion of Google's Serving Infrastructure. In *IMC*, 2013

[4] J. S. Otto, M. A. Sanchez, J. P. Rula, and F. E. Bustamante. Content Delivery and The Natural Evolution of DNS: Remote DNS Trends, Performance Issues and Alternative Solutions. In *ACM,* 2012

[5] A.-J. Su, D. R. Choffnes, A. Kuzmanovic, and F. E. Bustamante. Draft-ing Behind Akamai (Travelocity-Based Detouring). In *ACM SIGCOMM '06*, 2006.

[6] Y. Zaki, J. Chen, T. Pötsch, and T. A. Lakshminarayanan Subramanian. Dissecting Web Latency in Ghana. In *IMC*, 2014.

Thank you!
Questions or Comments?

BACKUP SLIDES

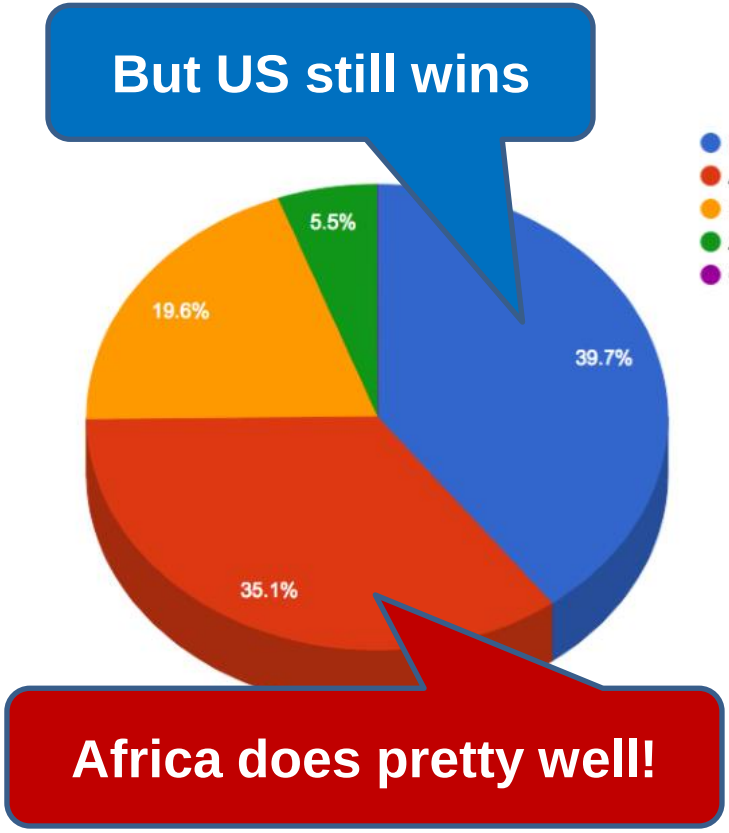
Extended Version: Geolocation results Rechecks (Ongoing)

- 100 RIPE atlas probes randomly selected in the world
- Around 18,751 pings measurements towards 1,870 IPs among
- Computation of the delay
  - XXX of IPs whose geoloc are expected to be wrong

Striking Results

Who's serving whom (DNS+EDNS0 results)?

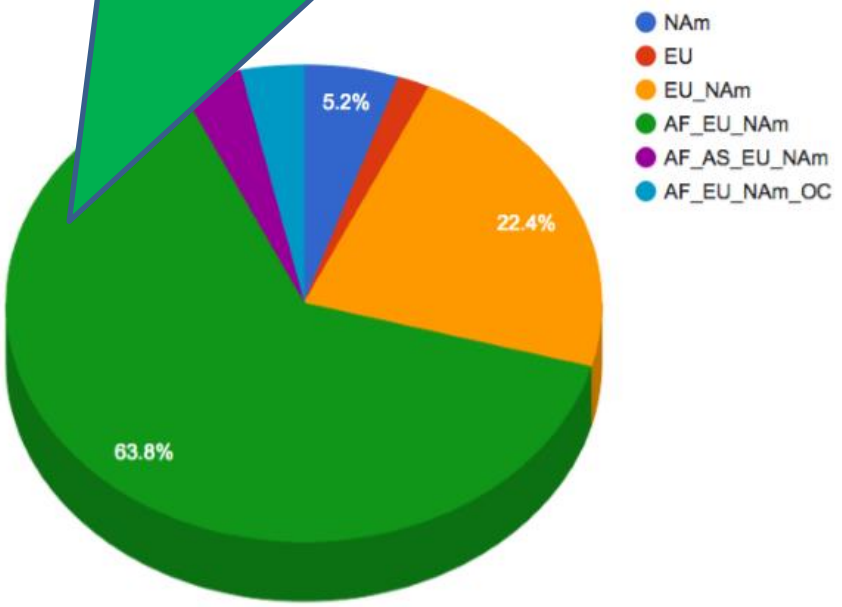
But US still wins



Africa does pretty well!

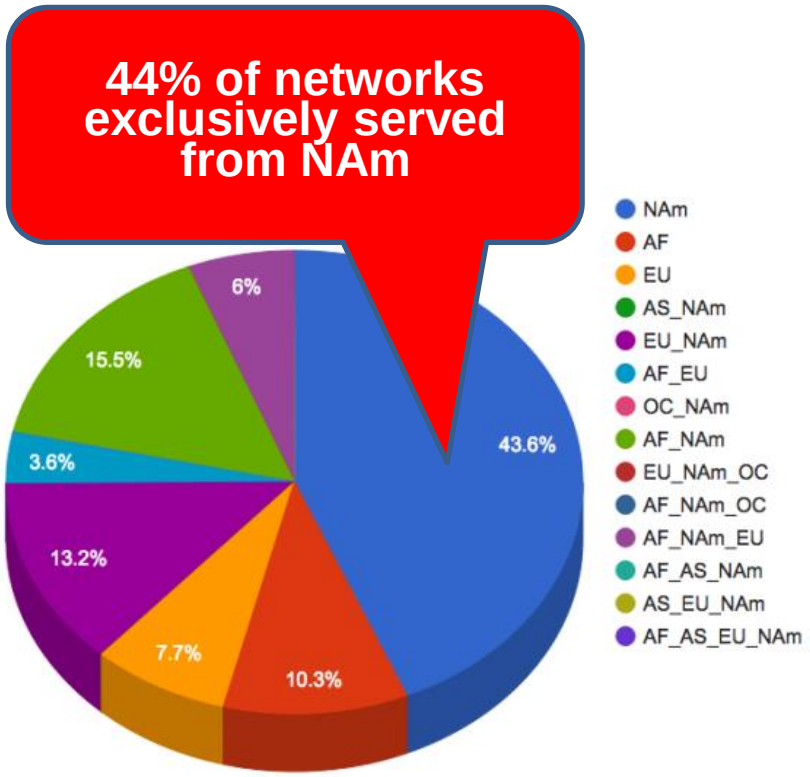
Location of caches

Most countries are a combo



Percentage of countries served by each continent

But it's not just about the countries

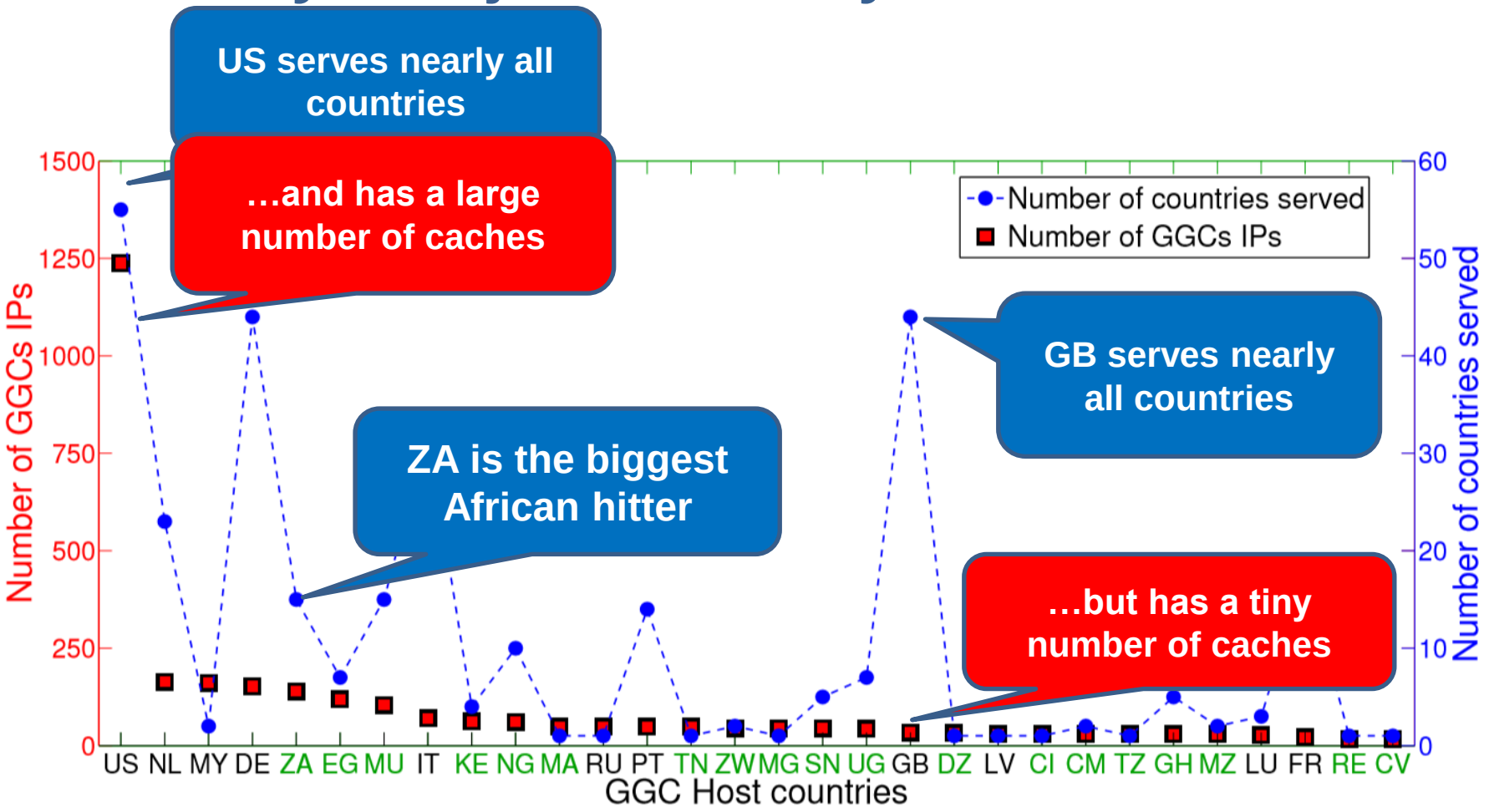


Percentage of prefixes served by each continent

- Only 5% of **countries** exclusively served by US
- But 44% of **networks!**
- No **country** is exclusively served by Africa
- But 10% of **networks** are

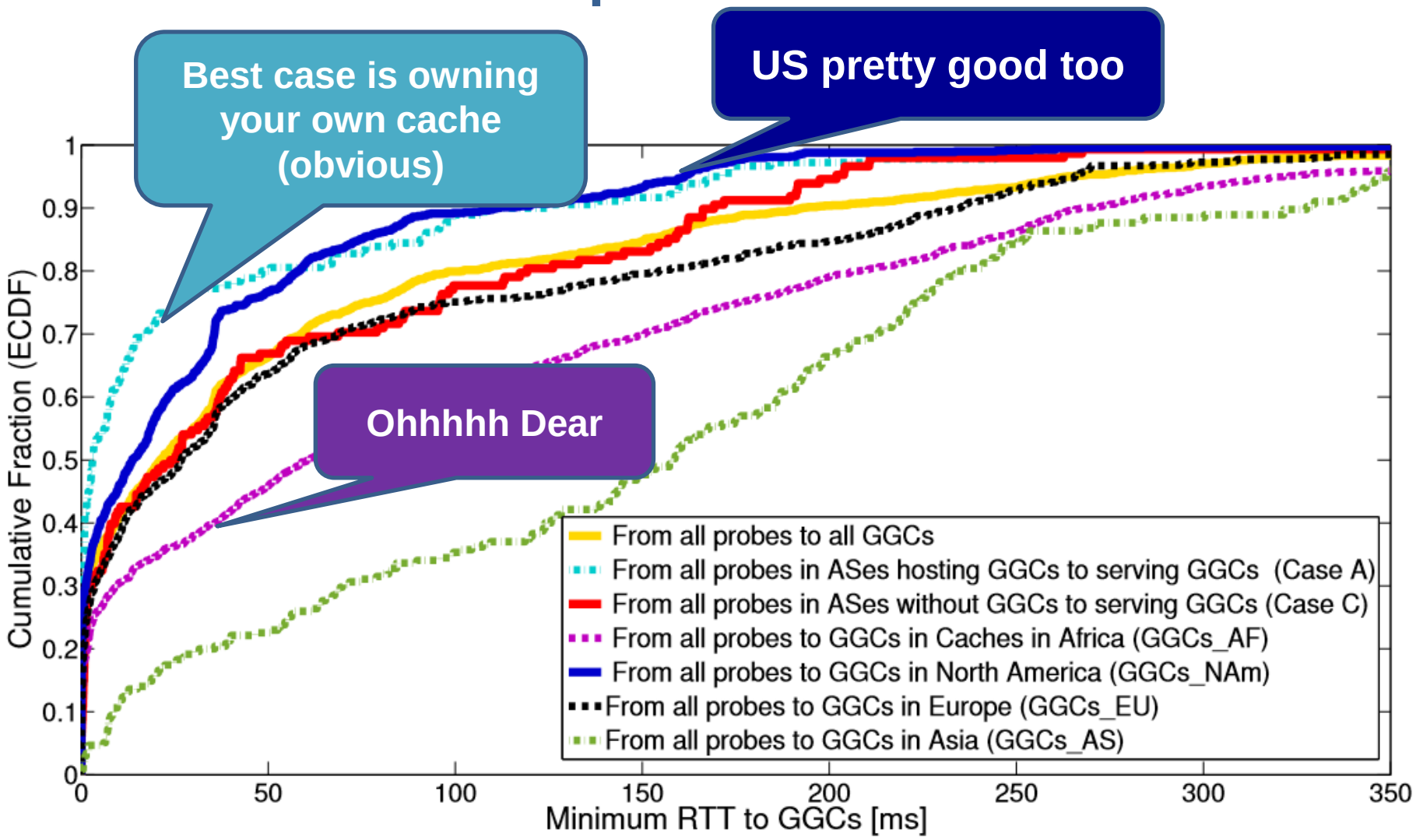
	AS
1	GOOGLE, US (37.21%)
2	TMNET-AS-AP, MY (5.13%)
3	YOUTUBE GOOGLE, US (4.74%)
4	LEVEL3, US (2.56%)
5	MEO-INTERNACIONAL, PT (2.05%)
6	RETN-AS, UA (1.98%)
7	ROSTELECOM-AS, RU (1.53%)
8	ETISALAT-MISR, EG (1.51%)
9	TELECOM ITALIA, IT (1.5%)
10	MTNNS-AS, ZA (1.47%)

Why not just share your cache?!



Distribution of GCCs serving African prefixes across countries

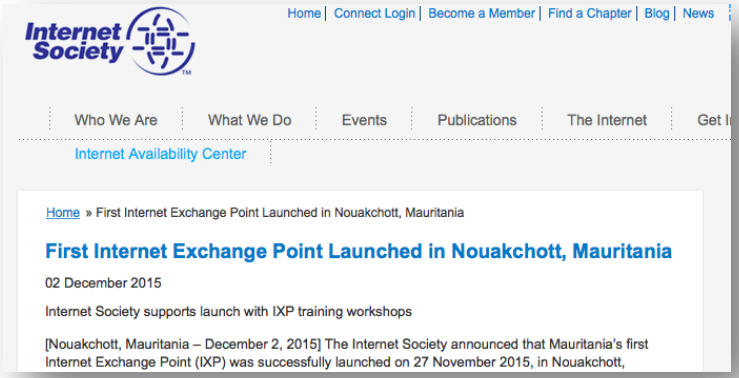
What's the performance hit?



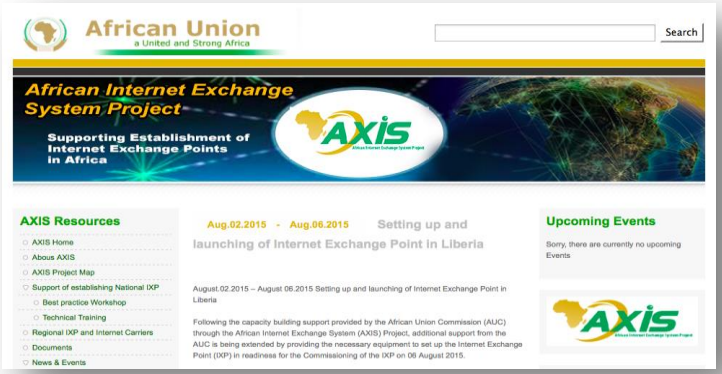
CDF of the minimum RTT distribution

Peering is the key to low delays

BFIX IXP launched (www.bfix.bf)



Liberia-IXP launched in LR



Windoek-IXP launch in NA

RIMIX IXP launched in MR

Peering is the key to low delays

- Launch of some of them proved [2]
  - RTTs among peers decrease
  - Financial costs decrease
- But many networks still remain **disconnected from IXPs**
  - Currently, on average 16 members at an African IXP

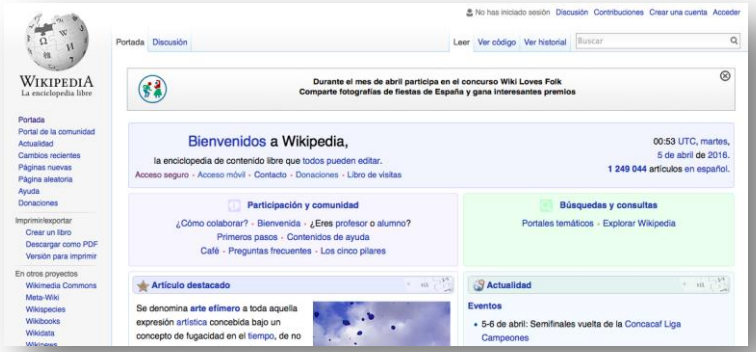
[2] Fanou et al. On the Inderdomain Routing in Africa, in PAM 2015

SOME TOP GLOBAL WEBSITES (ALEXA RANKING)

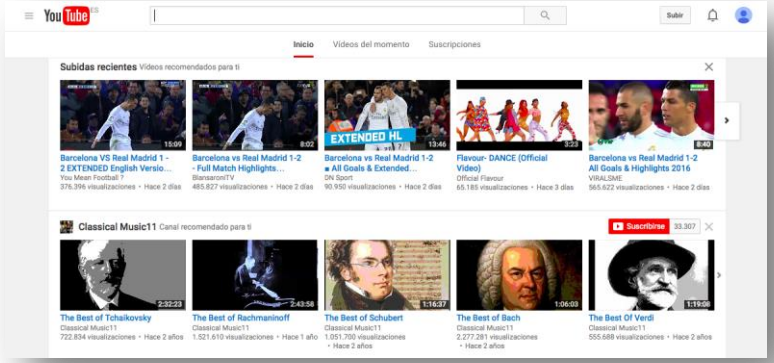
amazon.com (E-commerce)

tabao.com (E-commerce)

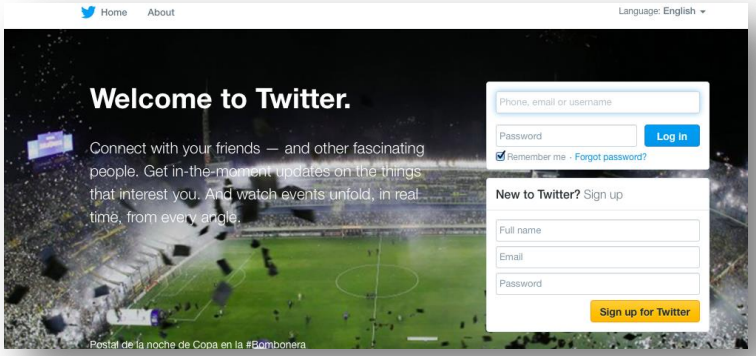
wikipedia.com (Encyclopedia)



youtube.com (Videos)



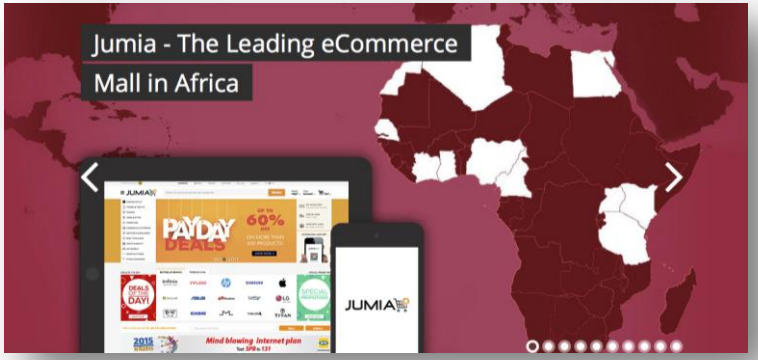
twitter.com (Social network)



facebook.com (Social network)

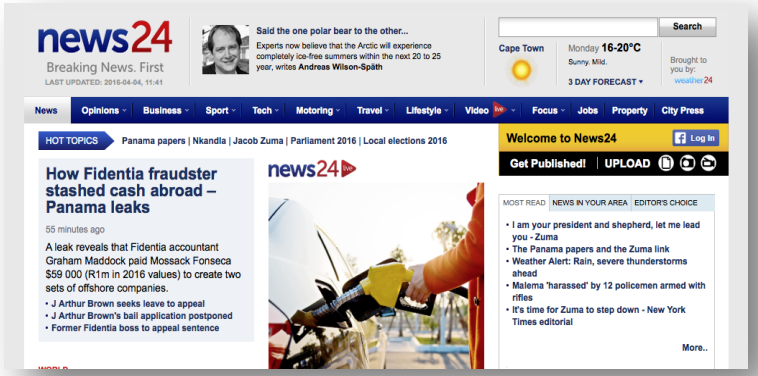
SOME TOP REGIONAL WEBSITES (ALEXA & AFRODIGIT RANKING)

jumia.com (E-commerce)

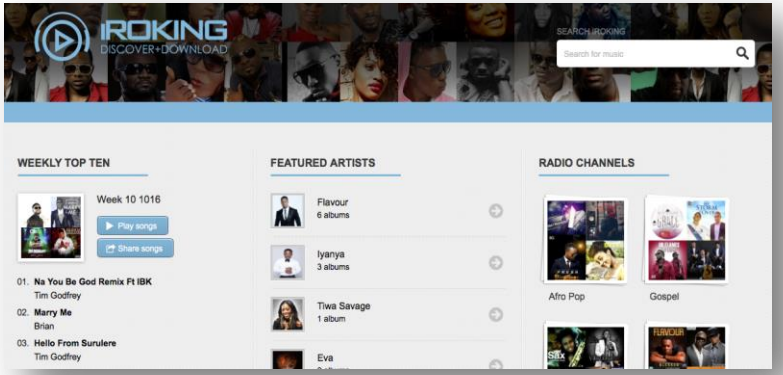


gtbank.com (Financial services)

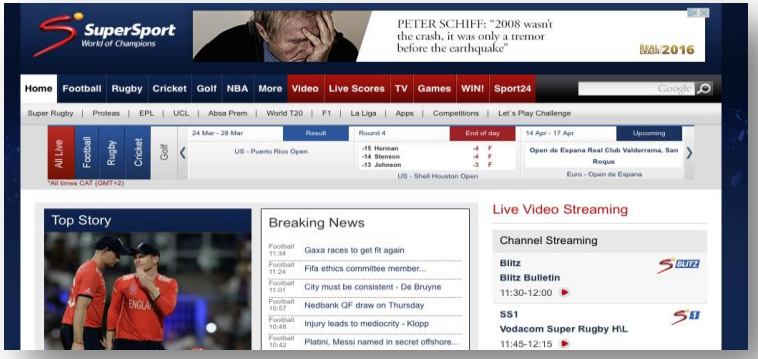
news24.com (News/media)



iroking.com (Videos)



supersport.com (Sports)



nairaland.com (Online Community)

Expanding to other providers

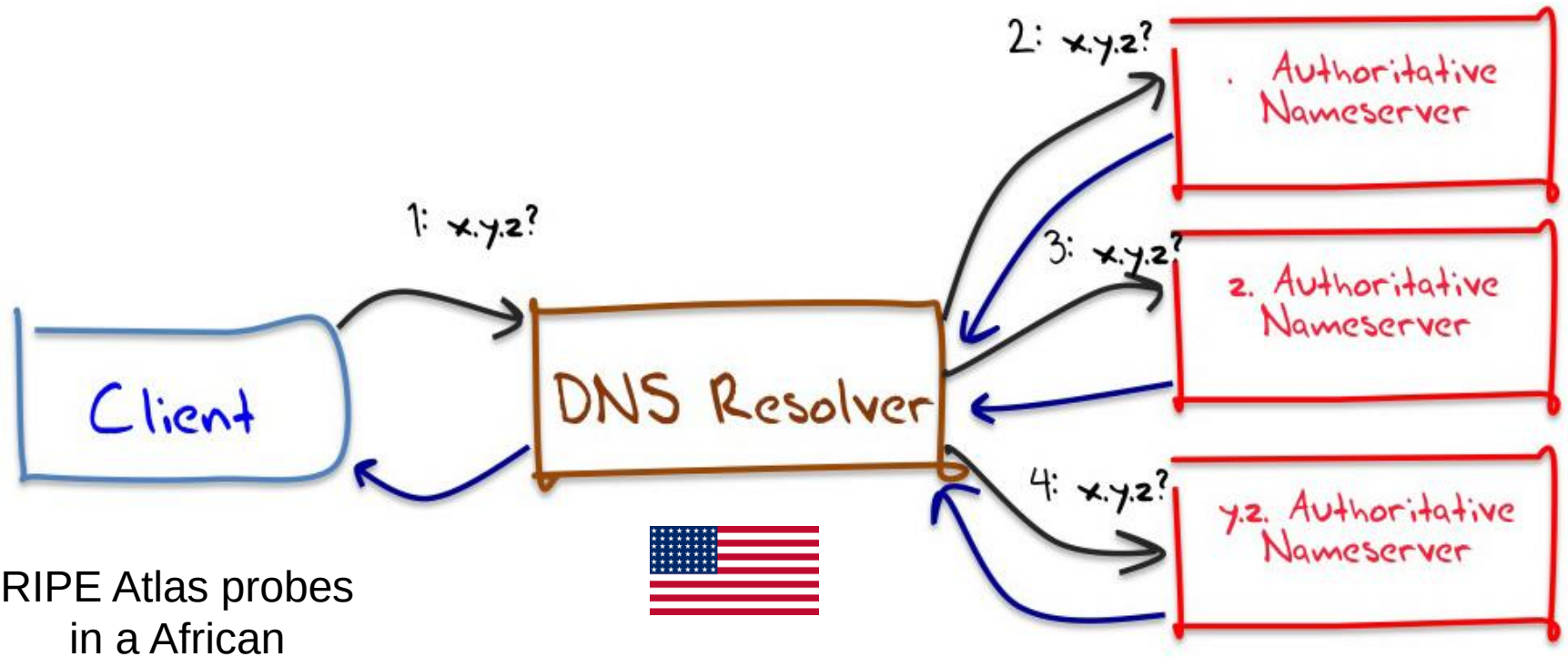
Top 15 sites in Africa (by Alexa & Afrodigit)	Type	#IPs caches	CCs host caches	ASes
jumia.com.ng	E-commerce	1	DE	20546
konga.com	E-commerce	1	US	15169
bidorbuy.co.za	E-commerce	1	ZA	3741
fnb.co.za	Financial services	1	ZA	17148
gtbank.com	Financial services	1	US	26496
absa.co.za	Financial services	1	ZA	3741
standardbank.co.za	Financial services	1	ZA	10798
almasryalyoum.com	News/media	5	NL, CR	13335
elkhabar.com	News/media	2	US	13335
vanguardngr.com	News/media	1	US	14618
news24.com	News/media	1	ZA	10474
punchng.com	News/media	1	IE	16509
iol.co.za	News/media	2	IE	16509
ghanaweb.com	News/media	1	US	7859
nairaland.com	Online community	5	US	13335
supersport.com	Sports	1	ZA	10474
alwafd.org	Politics	2	NL	13335
iroking.com	Videos	2	IE	16509

Top 10 global web-sites (by Alexa)	Type	#IPs caches	CCs host caches	#ASes
amazon.com	E-commerce	4	US	2
tabao.com	E-commerce			
Qq.com	Internet services	2	CN	2
google.com	Search engine	924	18 (§ 4.1)	26
yahoo.com	Search engine	4	US, UK	2
baidu.com	Search engine	1	HK	1
wikipedia.com	encyclopedia	2	NL, US	2
facebook.com	Social network	5	US, DE, NL	1
twitter.com	Social network	7	US	2
youtube.com	Videos	41	SN, MU, US	3

DNS resolvers used by Atlas probes

Each dot represents a set of DNS resolvers. Its size is proportional to the number ASes that used them for the resolution

Outsourcing the operation of DNS



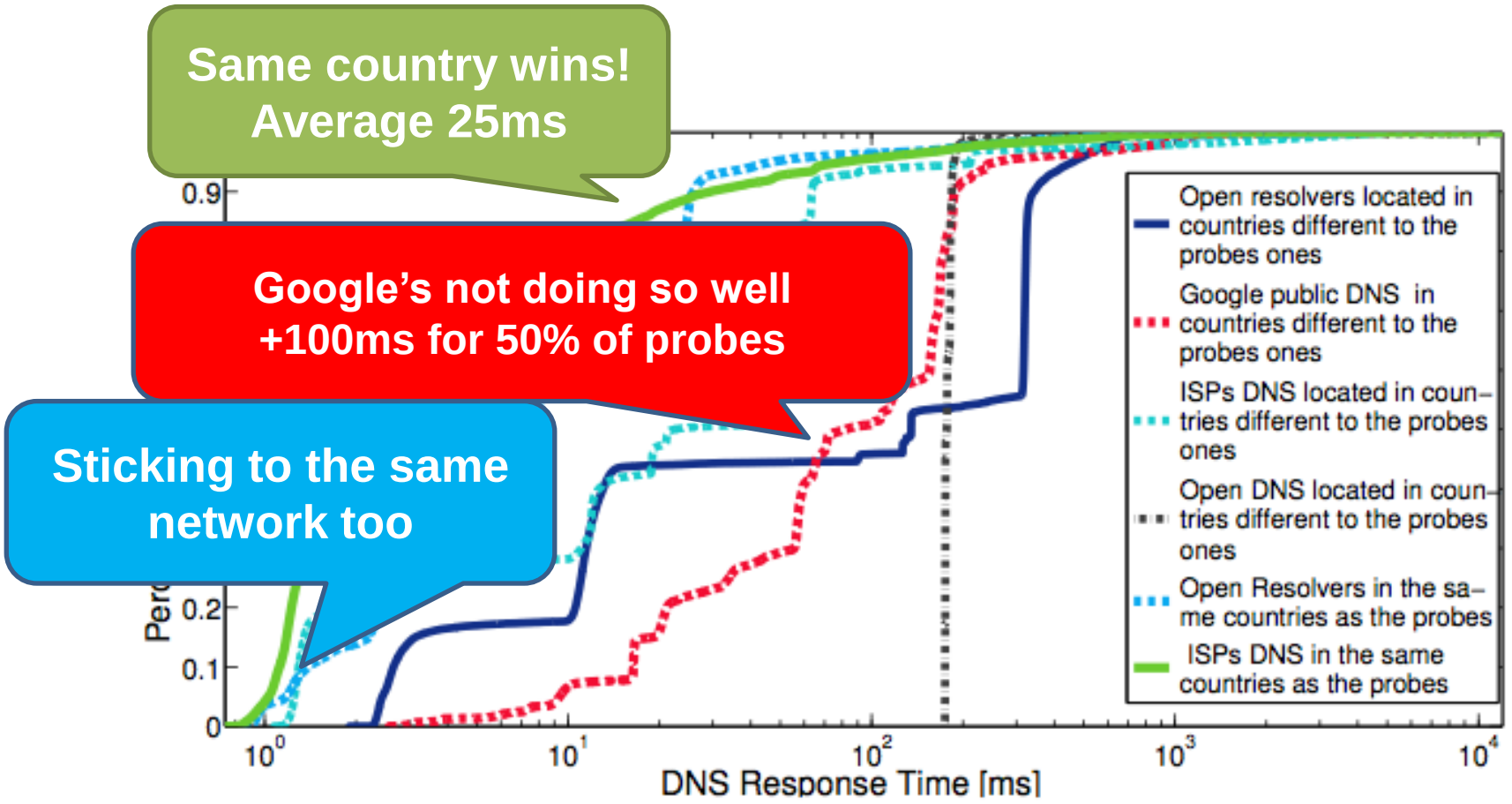
RIPE Atlas probes in a African Networks

13.6% of DNS resolvers outside Africa & 4.7% in US

Google

Analyzing DNS in Africa

This translates into big delays !



Cumulative Distribution of DNS resolution delays

Conclusions

- Superior connectivity is only part of the equation
- Africa is not self sufficient in hosting infrastructures
  - Most probed networks reliant on US infrastructure
  - Few third party networks share their GGC caches
- Google is more developed in Africa than other Content providers
- Even local and regional websites hosted outside
  - Higher HTTP load times from non-Google websites hosted outside
- Poor DNS configuration by African operators
  - High reliance on Open resolvers and Open DNS vs. ISPs DNS

Our suggestions

- CDNs should begin to improve their presence in Africa
  - Popular regional providers as front-runners
- Operators should improve peering
- CDNs must concurrently host caches at existing African IXPs (37 functional/ 39 on PCH website - https://prefix.pch.net/applications/ixpdir/summary/growth-region/)
- Network operators should correct their DNS settings
- More public DNS resolvers should be placed in Africa
- Technical report available:

R. Fanou, G. Tyson, P. Francois, A. Sathiaseelan, **Technical report: African Content Measurement Campaign**, https://techrep_cdma:PDQ7Rjkj@fourier.networks.imdea.org/external/techrep_cdma/index

A few ideas to finish on …

- Deploying CDN caches/replication points is
  - Let's do more of it!
- Network operators should improve peering
  - … and stop using third party DNS (or place resolvers in Africa)
- Need better hosting infrastructure in Africa to bring in smaller players
  - E.g. cheaper, more reliable etc.
- Pick the right ISP!
- Users can adopt smarter protocols
  - E.g. SPDY/HTTP 2.0

Acknowledgements

- We are grateful to all RIPE Atlas probes hosts and the RIPE Atlas team
- Arjuna Sathiaseelan is funded by H2020 RIFE Project

References

[1] A. Gupta

[2] R. Fanou, P. Francois, E. Aben, On the Diversity of Interdomain in Africa, PAM2015, USA, April 2015

[3]

[4]

[5]

[6]

Thank you! Any questions?