An Empirical Study of Mobile Network Behavior and Application Performance in the Wild

S. Zhang, W. Li, D. Wu, B. Jin, RKC. Chang, D. Gao, Y. Wang, RKP. Mok

Southern University of Science and Technology, China

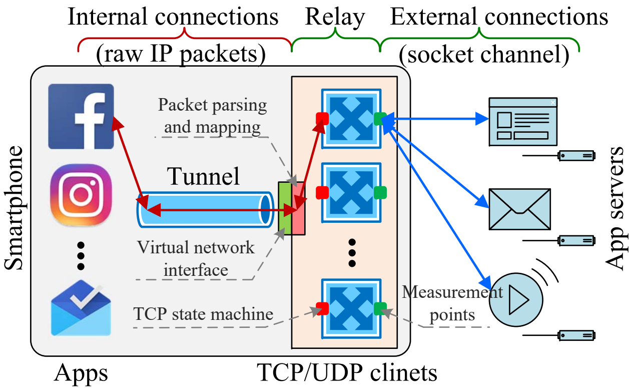
April 16, 2019

Introduction

- A two-year long dataset conducted by a mobile crowdsourcing app.
- Characterize the performance of different protocols, DNS deployments, IP anycast, etc. in the wild.
- An performance degradation detection method based on Apriori algorithm, tailored for imbalaced and sparse datasets.

Data Collection

- VPN-based
  - Real traffic
  - No “root” needed
- Crowdsourcing
- Per-app measurement

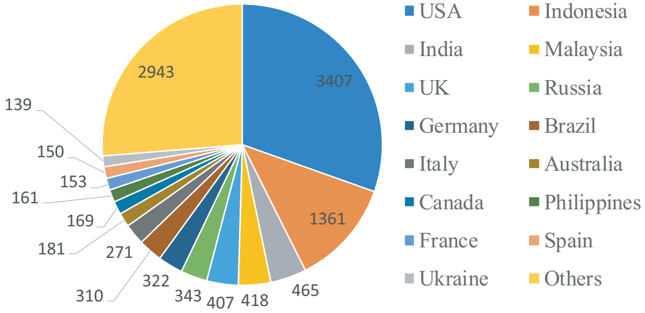


Data Features

- User Information
  - country, device model, android version, etc.
  - collects once per installation
- Network Infromation
  - type (WiFi or cellular), name (SSID or vendor name), geo-location etc.
  - collects each time on app enabled or network status changed
- Measurement
  - RTT, server IP and port, package name, the domain name etc.
  - measure each TCP connection or DNS query once.

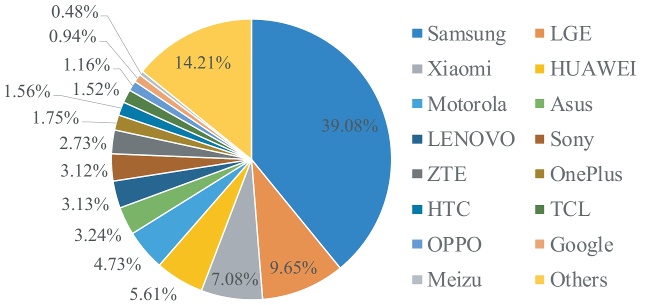
Basic Statistics

- Country Distribution: 11,200 users from 173 countries, mostly USA and Southeast Asia.



Basic Statistics

- Device Details: 1,615 different smartphone models from 226 manufacturers

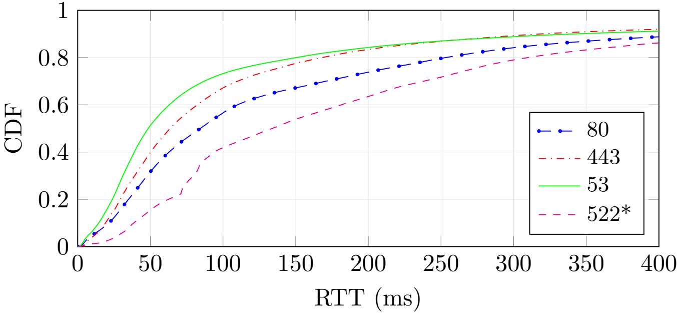


Basic Statistics

- Applications: 17,059 apps with 1,197 apps have >1k measurements
- Measurements: 13,204,649 TCP records and 6,489,646 DNS records, covering 286,404 destination IP addresses.
- Network types: 65.42% WiFi, 23.97% LTE, 10.61% other cellular networks.
    - only 5.94% of WiFi measurements were observed to have >300Mbps PHY rates.
    - more than one third of the ISPs (238 ISPs) have no 4G measurements observed, mainly in Africa and Asia.

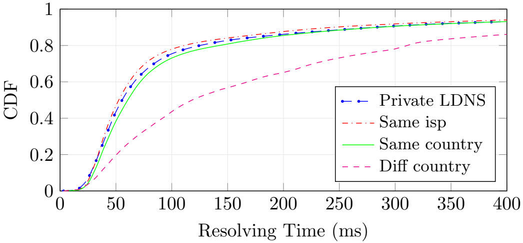
Protocols

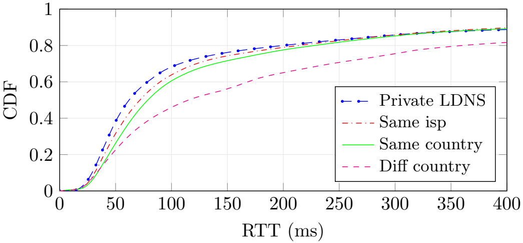
Our analysis shows that XMPP traffics experience longer latency than HTTP(s).



DNS Performance[1]

Users using DNS server that are located on different countries experience longer latency to app servers. This suggests the need for IP Anycasting.

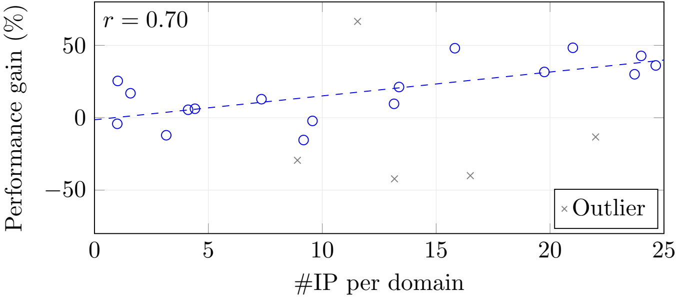




[1] servers deployed IP Anycast are considered "diff country" in this chapter

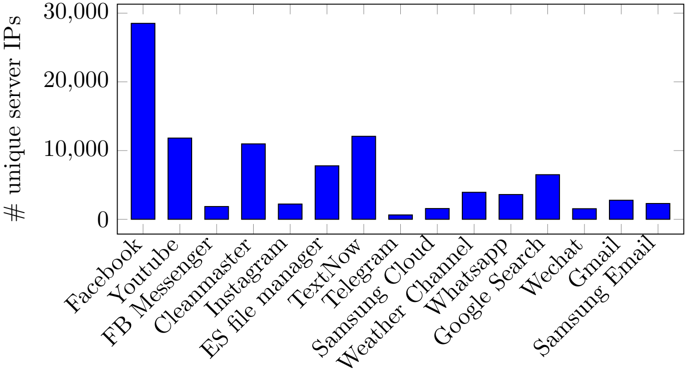
IP Anycast

We identify Anycast IP using the the list conducted by iGreedy.[2] We use rlm() from R package MASS with default parameters to perform robust regression.



[2] https://anycast.telecom-paristech.fr/dataset/

Application Servers



Application Servers

The Ad servers and trackers are identified by EasyList.[3]

App	Server role	Percentage	Median RTT (ms)
Facebook	Own server	53.3%	54
	3rd-party CDN	2.2%	86
	Ad	9.6%	71
	Tracker	3.1%	71
YouTube	Own server	81.5%	49
	3rd-party CDN	0%	n/a
	Ad	1.1%	56
	Tracker	0.9%	100
CleanMaster	Own server	13.0%	96
	3rd-party CDN	0.9%	50
	Ad	58.1%	96
	Tracker	4.2%	79
TextNow	Own server	4.0%	66
	3rd-party CDN	2.6%	80
	Ad	54.7%	73
	Tracker	10.4%	61

[3] https://easylist.to/

Performance Degradtion Detection

Challenges

- Imbalanced: For example, 83.5% of the 16,868 HSPAP measurements for ISP Mobilis are from one user. If those measurements are excluded, the median RTT can decrease from 332ms to 219ms.
  - normal association rules method bias to the performance of the dominating user.
- Sparse: Although the total number of observations is huge, records for each combination of features can be very small.
  - it's impossible to model the normal performance for all combinations of features separately.
- Large: We need a scalable method to process the increasingly large data.

Our Method[4]

1. Based on the famous association rules mining method, the Apriori algorithm.
2. We filter each candidate rule to ensure no more than half of the supporting records have the same feature.
3. We identify performance degradtion events by comparing the meadian RTT of the supporting records for one candidate rule and a subset of it.
   - For example, median RTT of LTE records is 73 in our data, while the RTT of the records that use LTE and linux kernel 3.10.49 has a median of 340.
4. Use Hypothesis test to verify that the supporting data cannot be split further.

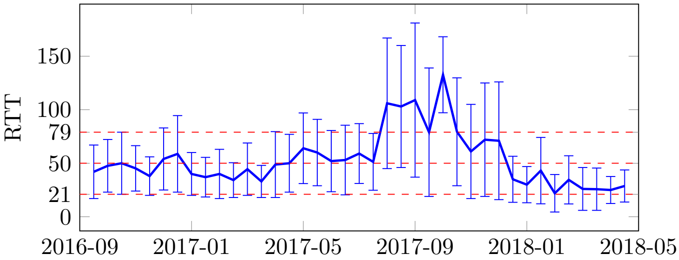
[4] For more detailed description of our method we refer interested readers to attend IWQoS on 24-25 June 2019, Phoenix, AZ, USA or read the proceedings.

Evaluation

1. Low false positive rate in random data
   - We randomly shuffle the RTT of the records.
   - We mathmatically proved that the probability of our methods thinking there is anomalies are very small in our configuration.

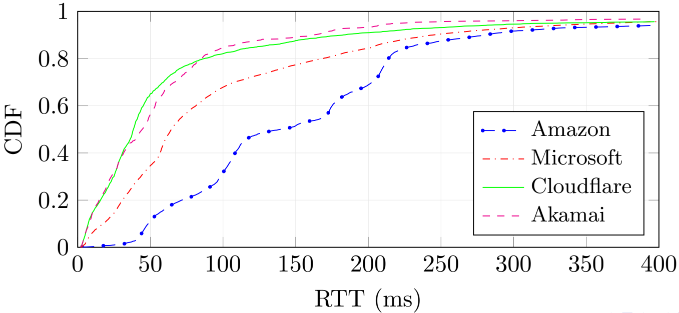
Evaluation

1. Low false positive rate in random data
2. Real world case of Google Germany



Evaluation

1. Low false positive rate in random data
2. Real world case of Google Germany
3. Real world case of Microsoft Office Mobile



Conclusion

- Though IEEE 802.11ac equipments has become the mainstream in the market, only a small portion (6%) of Wifi exceed PHY rates of 300Mbps.
- Still more than one third of the ISPs do not deploy 4G networks.
- There are many users use external DNS resolvers. IP Anycast may improve the mobile app performance in this case.
- Traffics using XMPP protocols experience longer RTT than HTTPS, which suggests that IM and VoIP services can be further improved.
- Advertisements servers often have longer latency than application servers.

Future Works

- 5G deployment and performance
- Actively measure the server when unexpected high RTTs are observed.

Any questions?

Thank you!