A Multi-Protocol Analysis of Android App's Traffic From The Edge

Abbas Razaghpanah*, **Julien Gamba**†‡, **Álvaro Feal**†‡, Narseo Vallina-Rodríguez†§, Mark Allman§, and Philipa Gill¶

*Stony Brook University †IMDEA Networks Institute ‡Universidad Carlos III § ICSI ¶ UMass-Amherst

Developing the Science of Networks

institute IMDEA networks

Motivation

Previous studies have shown some limitations when analyzing traffic for smart-phone applications:

- ISP traces analysis can not match network flows to specific apps.
- Some traffic, such as home network traffic, never goes trhough ISP routers.
- On-device solutions sacrifice scale due to the need of rooting the phone.
- Focusing only on HTTP and HTTPS leaves out a number of other protocols used by smartphone apps.

Research questions

These limitations on the state of the art leave questions unanswered:

1) How do mobile apps behave at the network level?
2) What are the numbers and types of domains that a phone contacts?
3) Which are the transport and application protocols used?
4) How do apps behave on the home network?
5) How has the adoption of protocols been longitudinally?

Methodology

Lumen [1] leverages the VPN permission to collect network flows from non-rooted devices.

- We use crowdsourcing techniques to record real-user traffic (without private information).
- Running on-device grants access to info which can be used to map flows to apps.
- Lumen has a TLS interception module used to study encrypted flows and apps resistance against MITM attacks.

Dataset

- Over 18K users from 142 countries.
- Over 23M real unique flows from over 67K apps.
- Over 168K Fully Qualified Domain Names and 507K destination IPs.
- 20.5% of apps with 1M plus downloads and 11.9% with 1K less downloads.
- 6.5% of flows in the home network.

Domain level analysis

- The presence of 3rd party libraries enlarges the number of domains contacted by apps.
- Multi-CDN strategies have the same effect, as apps combine several CDNs to increase resilience and availability [2].

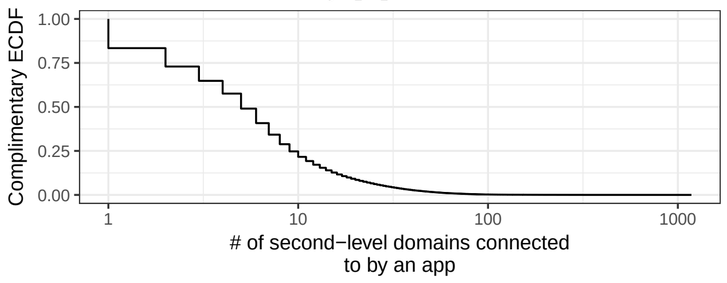


Figure 1: 75% of connect to at least 4 domains, and 25% connect to at least 9.

- **36% of traffic** in our dataset corresponds to **advertising and tracking services**, proving that mobile traffic is highly dependant of the 3rd party libraries used by developers.
- Though malicious domains are negligible in the scope of mobile traffic, their presence shows an effort by some developers to generate non-advertising revenue.

Malware family	% Domains	% Apps	Example Domain
Cryptocurrency mining [3]	8%	8%	coinhive.com
Phishing / Adware	70%	78%	rdldtrk.com
Malware	12%	8%	mobj.space
Possible misclassifications	10%	6%	sefh.es

Microscopic view of protocol usage

Previous studies have focused on HTTP and HTTPS analysis only, but our dataset shows that the amount of protocols present in mobile traffic is much more diverse.

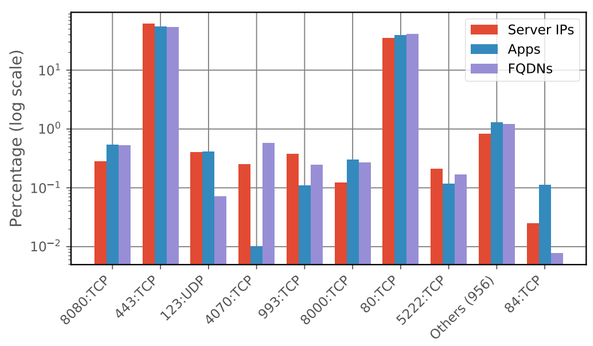


Figure 2: Protocol usage in the mobile ecosystem

QUIC

- UDP based transport protocol proposed by Google.
- We found only 32 (0.05%) apps which use QUIC. All of them but Snapchat are developed by Google.

NTP

- 337 apps (0.5 % of total) connect to 80 different NTP pools in order to get network synchronized time.
- The type of apps using NTP is very diverse (e.g,: games, IOT, VPNs, Social Media, etc).

P2P

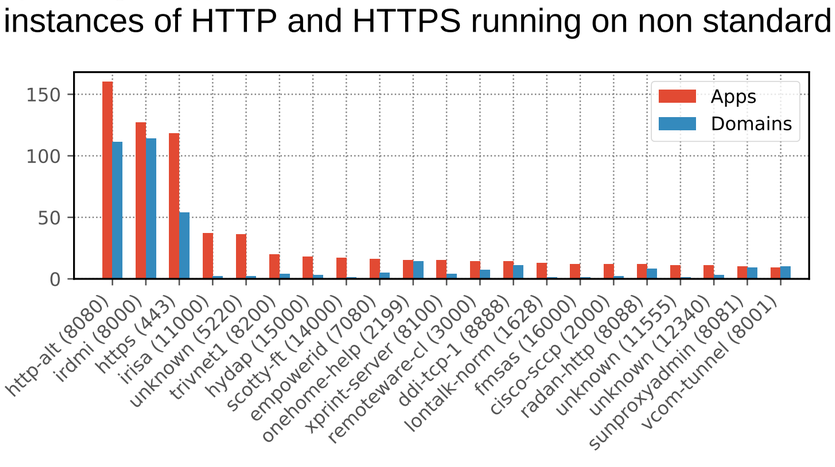
- 3.0 % of apps in our dataset connect to 1,408 port numbers beyond those used for HTTP(S), NTP and email services.
- These apps communications resemble that of P2P apps [4] since they communicate with a big number of ports.

Home network

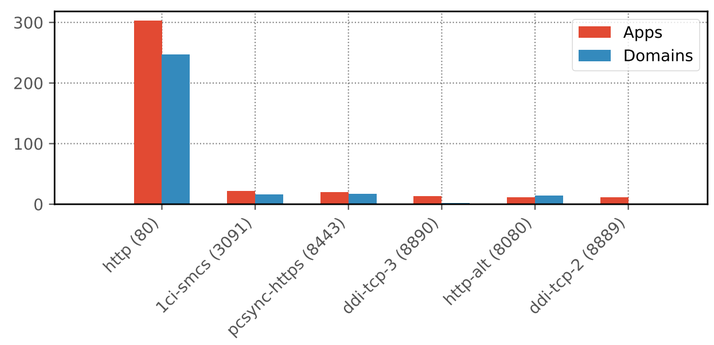
- We have identified 523 apps that communicate to the home network as well as with internet networks.
- 50% of these apps contact different ports depending on the type of network.

HTTP and HTTPS

- In total 37% of apps use HTTP and 70% use HTTPS.
- By actively probing the ports in our dataset for HTTP and HTTPS servers, we find many instances of HTTP and HTTPS running on non standard ports.



Non standard ports used for HTTP



Non standard ports used for HTTPS

Conclusions

- Mobile traffic is richer than what has been previously reported.
- The prevalence of third party services increases the number of domains contacted by apps.
- Despite being the most used protocols, we show that mobile apps protocol usage goes further than just HTTP and HTTPS.
- Non default ports are used for HTTP and HTTPS services, rendering classical protocol analysis incomplete.

References

[1] A. Razaghpanah, N. Vallina-Rodriguez, S. Sundaresan, C. Kreibich, P. Gill, M. Allman, and V. Paxson, "Haystack: In situ mobile traffic analysis in user space," *ArXiv e-prints*, 2015.

[2] F. Michclinakis, H. Doroud, A. Razaghpanah, A. Lutu, N. Vallina-Rodriguez, P. Gill, and J. Widmer, "The cloud that runs the mobile internet: A measurement study of mobile cloud services," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. IEEE, 2018, pp. 1619–1627.

[3] MalwareBytes, "Drive-by cryptomining campaign targets millions of Android users," https://blog.malwarebytes.com/threat-analysis/2018/02/drive-by-cryptomining-campaign-attracts-millions-of-android-users/.

[4] T. Karagiannis, K. Papagiannaki, and M. Faloutsos, "Blinc: multilevel traffic classification in the dark," in *ACM SIGCOMM Computer Communication Review*, vol. 35, no. 4. ACM, 2005, pp. 229–240.