AIMS Workshop - CAIDA, February 2016

INTERNATIONAL COMPUTER SCIENCE INSTITUTE

The ICSI Haystack

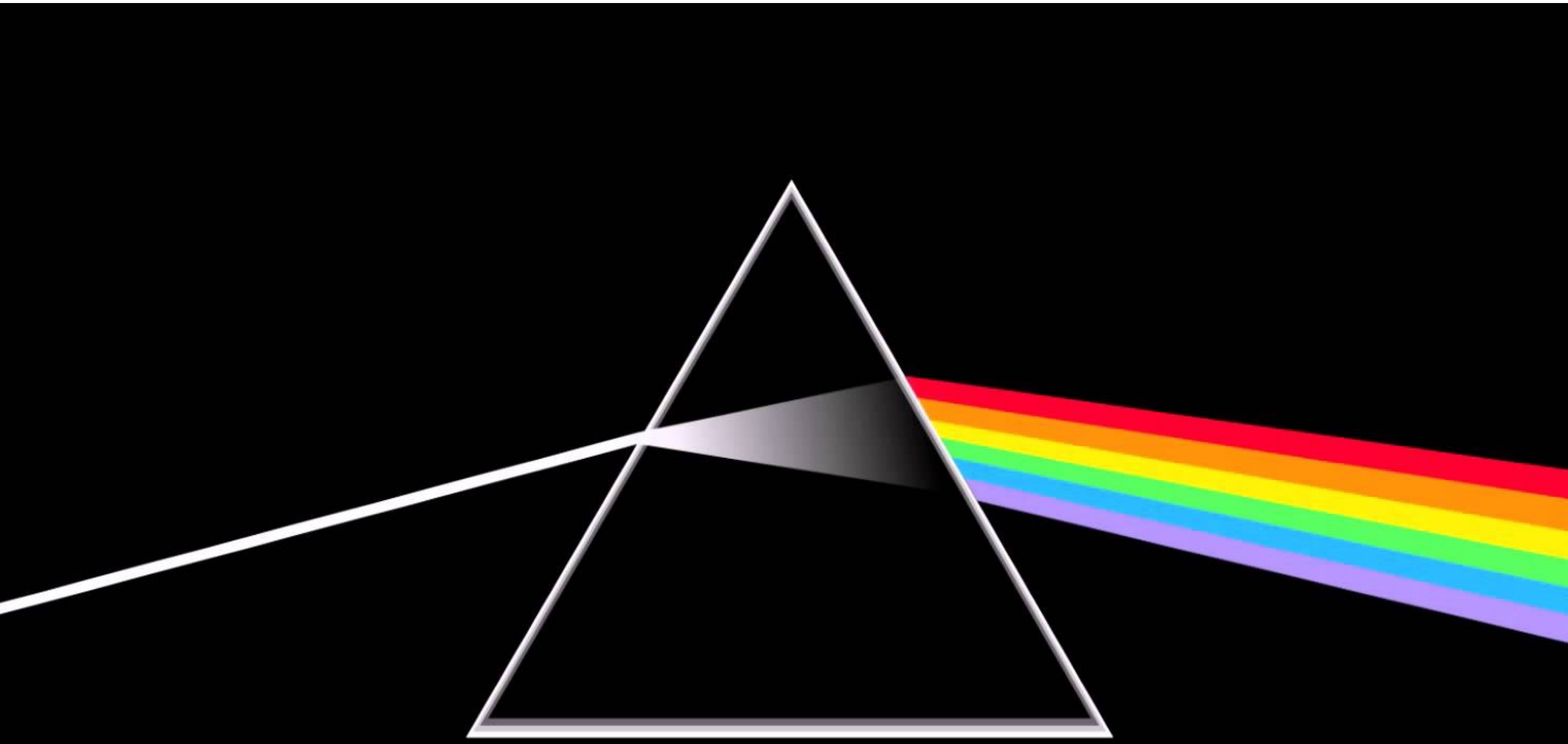
A Platform for Hybrid Mobile Measurements in the Wild

Narseo Vallina-Rodriguez

In collaboration with:

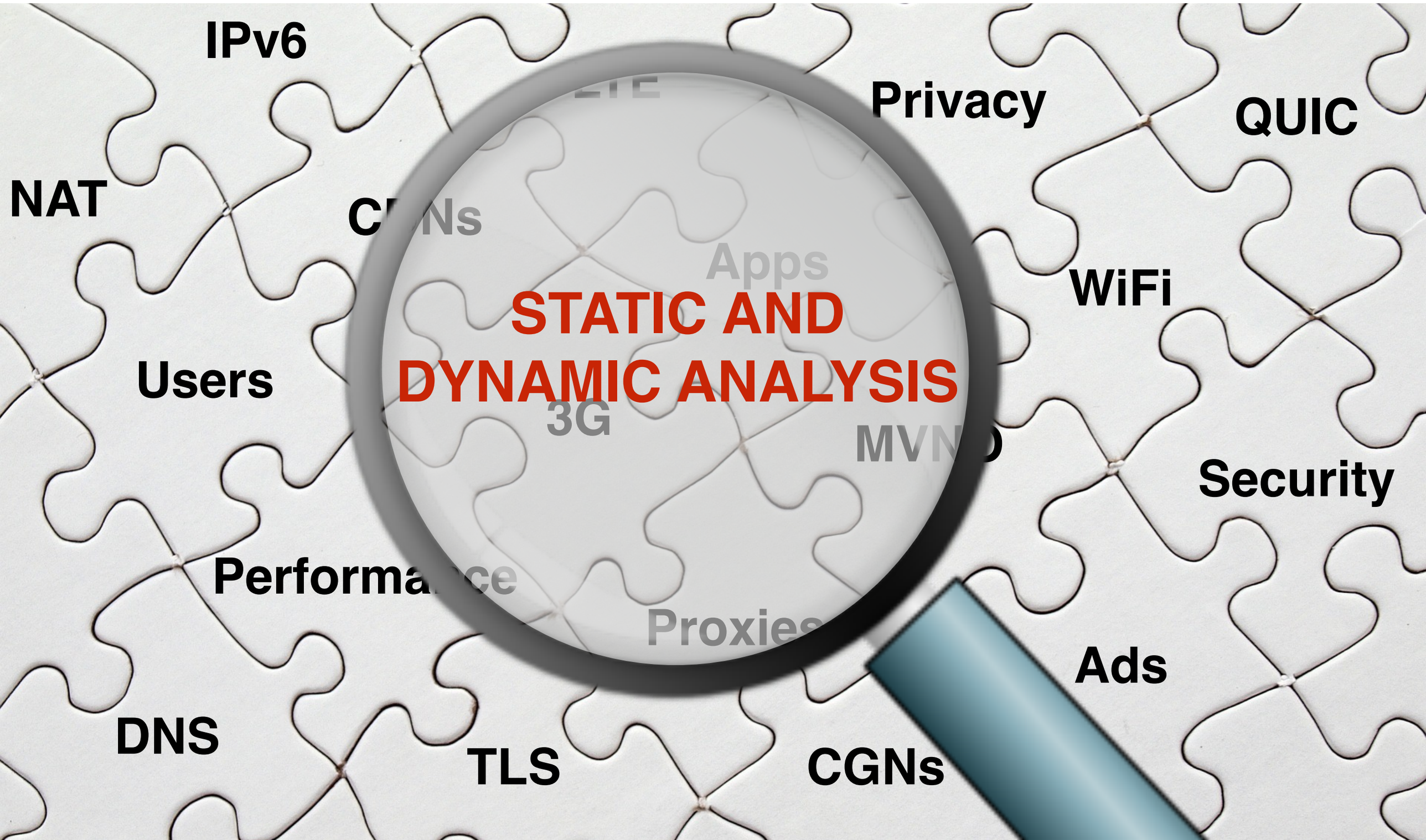
S. Sundaresan, C. Kreibich, M. Allman, V. Paxson (ICSI/UC Berkeley)
A. Razaghpanah, P. Gill (Stony Brook University)

How much do we know about the mobile ecosystem?



The mobile jigsaw

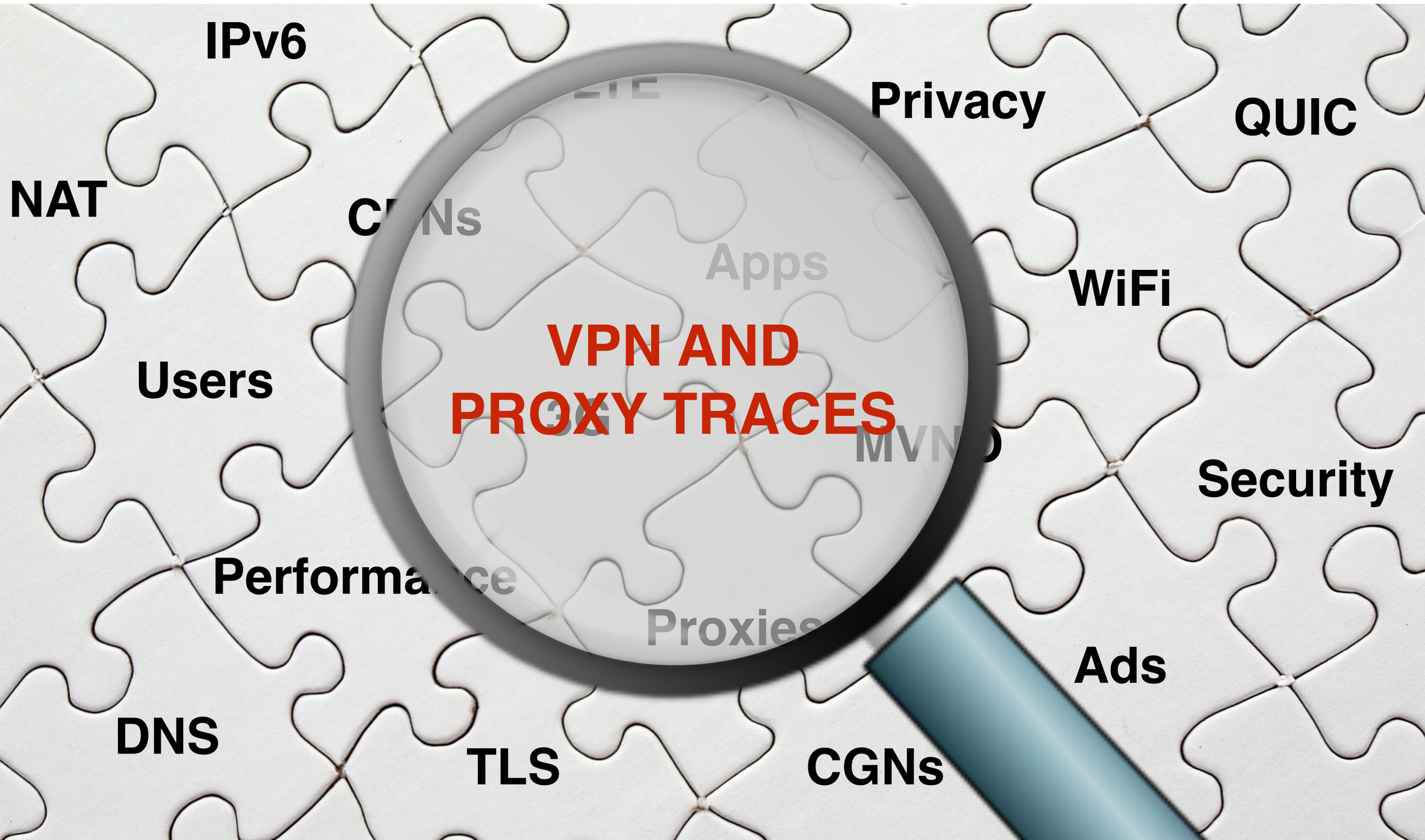
The mobile jigsaw



The mobile jigsaw

The mobile jigsaw

The mobile jigsaw



TRADE-OFFS!

The ideal mobile measurements platform:

Real-world operation

Comprehensiveness

Local operation

Large scale

The ICSI Haystack

A user-centric, and on-device measurements platform that intercepts and studies network traffic and app activity in **user space**

Schematic view of Haystack

Max throughput: ~55 Mbps
Extra latency < 1-4 ms
Battery overhead: 2-9 %

Contextualized traffic analysis

Traffic Analyzer (off-path)

Anonymized reports (IRB)

DB @ ICSI

Optional TLS interception

Java sockets! 😡
i.e., no-packet level traces

Forwarder

TLS Proxy

tun interface

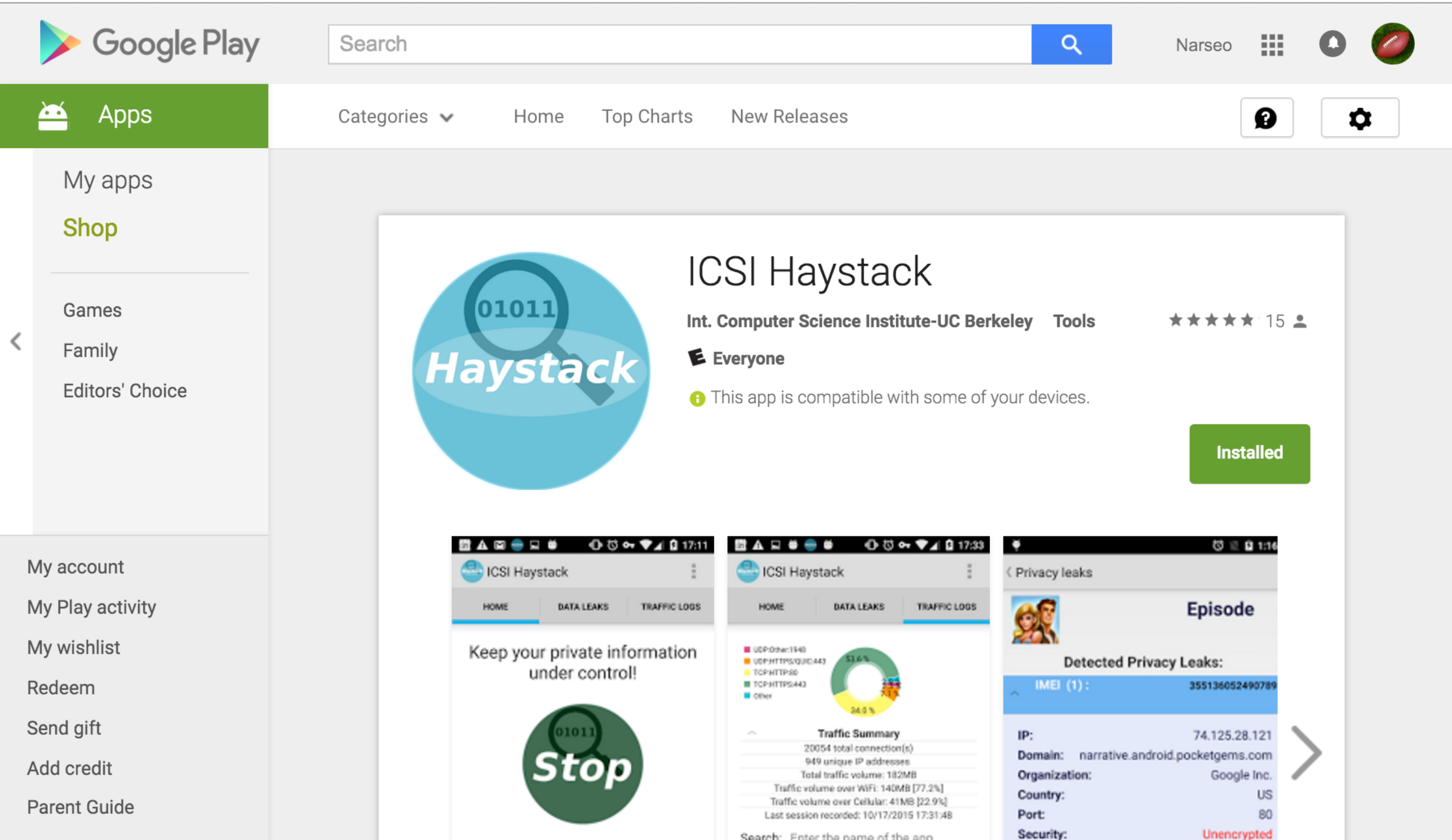
Default GW

App traffic

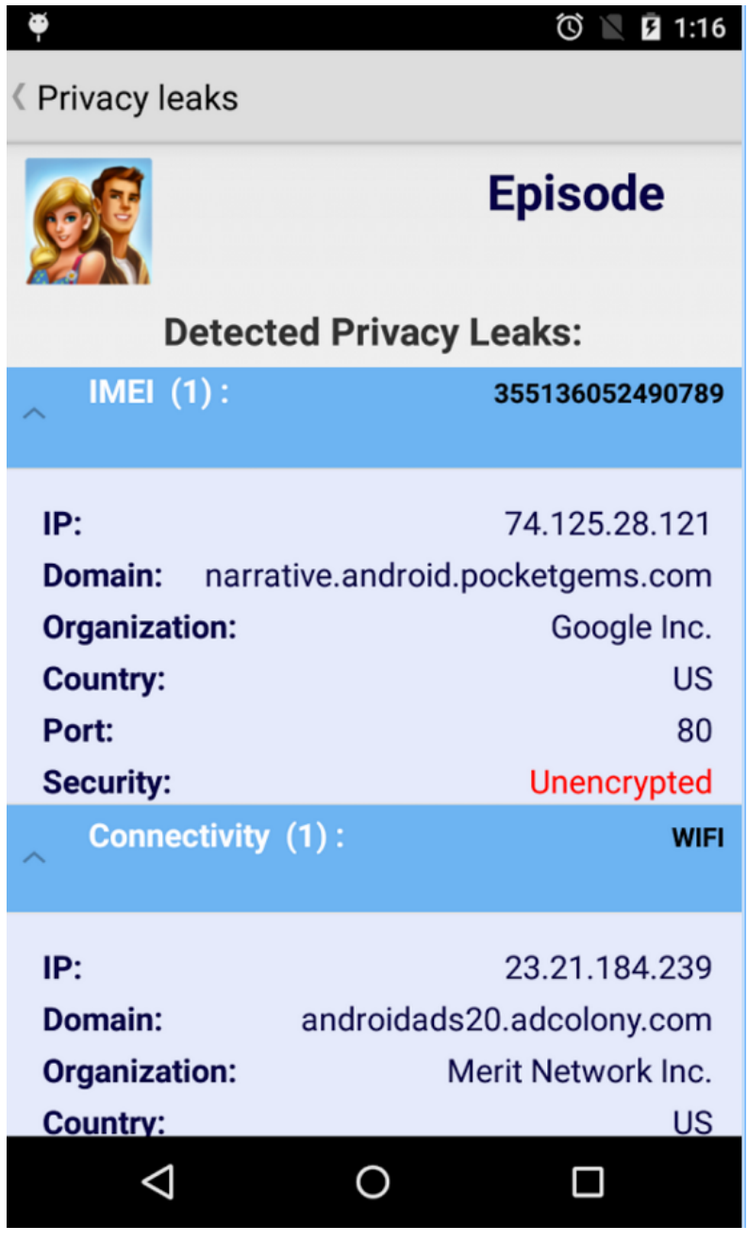
Raw packets

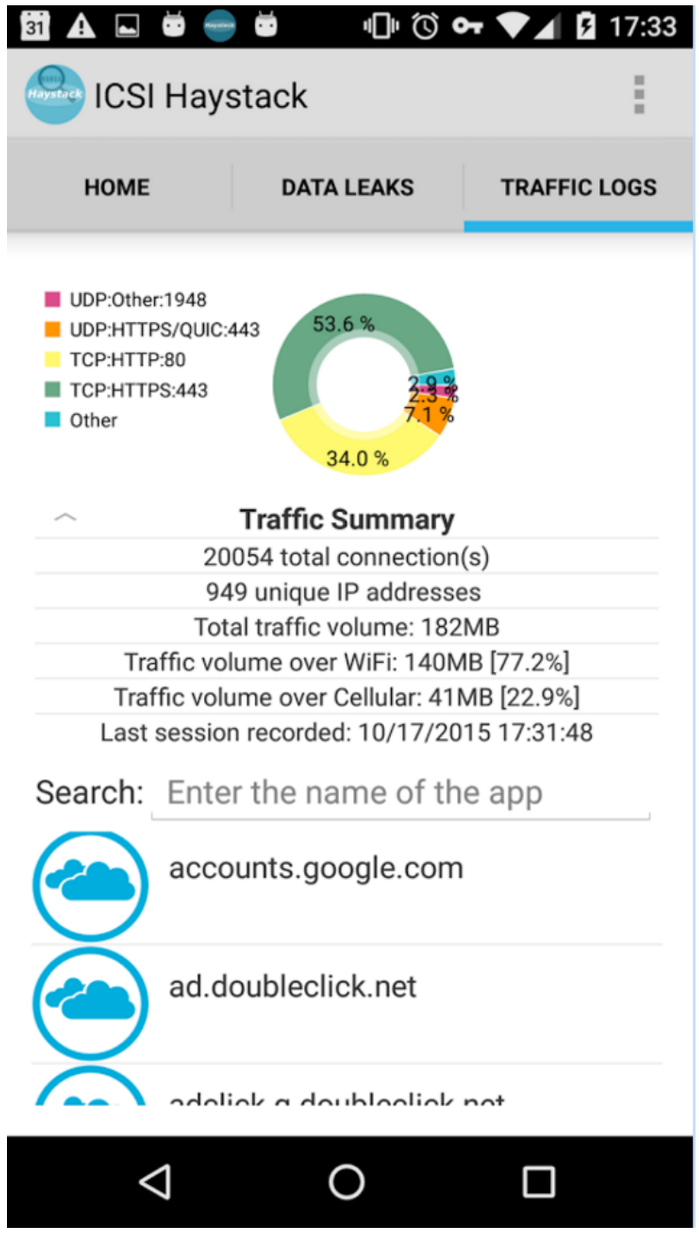
Internet

A easy-to-deploy tool for mobile users!



The user engagement challenge





Technical details and performance evaluation:

Computer Science > Networking and Internet Architecture

Haystack: In Situ Mobile Traffic Analysis in User Space

Abbas Razaghpanah, Narseo Vallina-Rodriguez, Srikanth Sundaresan, Christian Kreibich, Phillipa Gill, Mark Allman, Vern Paxson

(Submitted on 6 Oct 2015)

Despite our growing reliance on mobile phones for a wide range of daily tasks, we remain largely in the dark about the operation and performance of our devices, including how (or whether) they protect the information we entrust to them, and with whom they share it. The absence of easy, device-local access to the traffic of our mobile phones presents a fundamental impediment to improving this state of affairs. To develop detailed visibility, we devise Haystack, a system for unobtrusive and comprehensive monitoring of network communications on mobile phones, entirely from user-space. Haystack correlates disparate contextual information such as app identifiers and radio state with specific traffic flows destined to remote services, even if encrypted. Haystack facilitates user-friendly, large-scale deployment of mobile traffic measurements and services to illuminate mobile app performance, privacy and security. We discuss the design of Haystack and demonstrate its feasibility with an implementation that provides 26-55 Mbps throughput with less than 5% CPU overhead. Our system and results highlight the potential for client-side traffic analysis to help understand the mobile ecosystem at scale.

Comments: 13 pages incl. figures
Subjects: **Networking and Internet Architecture (cs.NI)**
Cite as: **arXiv:1510.01419 [cs.NI]**
(or **arXiv:1510.01419v1 [cs.NI]** for this version)

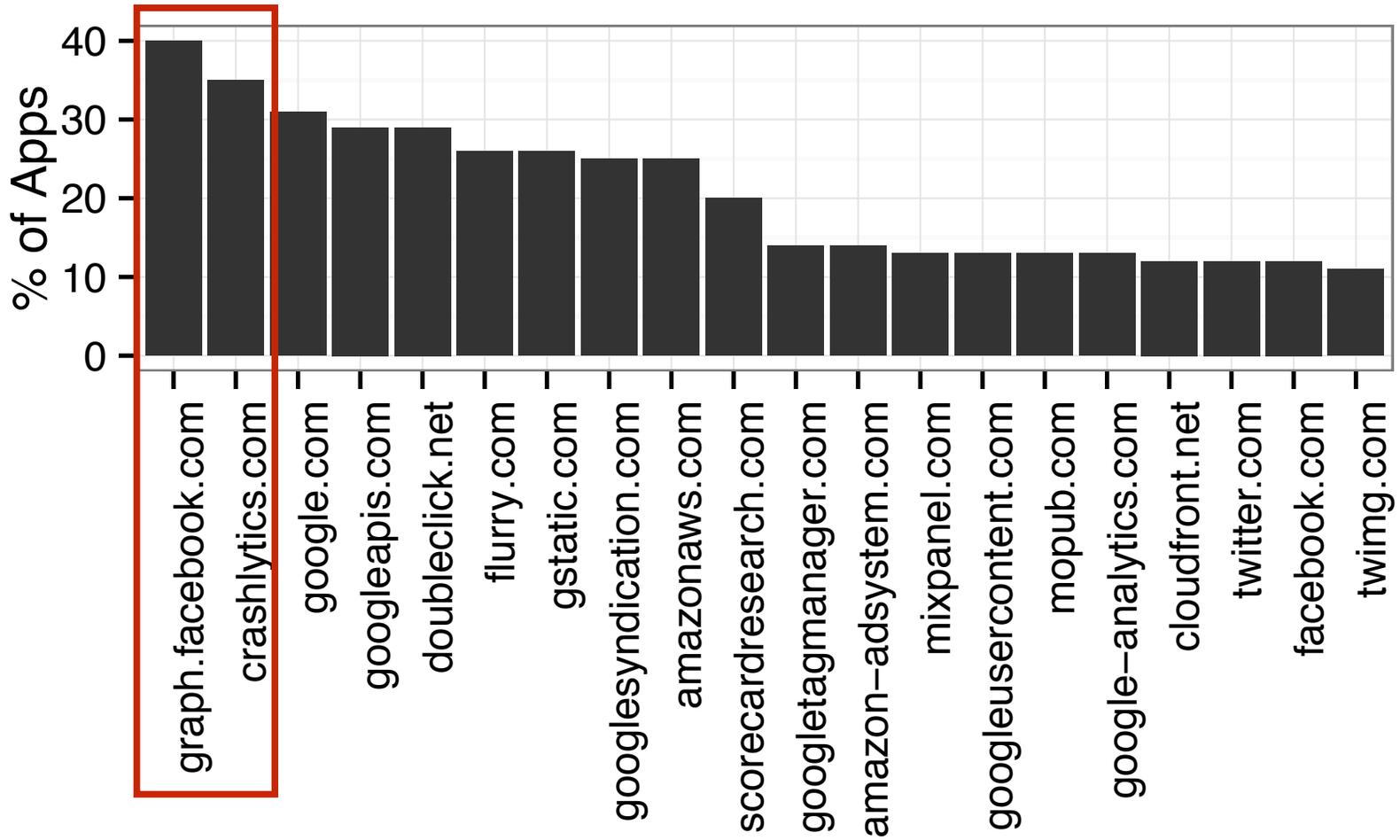
Ongoing and Future Research Directions

We are [mostly] in the dark about how mobile apps behave in **ANY** network!

"*I love working for the NSA, but if I'd wanted to snoop on people's most intimate information, I'd have become an app developer!*"

http://www.robcottingham.ca/

Who do apps talk to, what do they talk about, and how?



Provides DPI and generates accurate behavioral signatures

New-generation analytics and ad networks use TLS!

Allows users to stay in control of their traffic

Performance evaluation: Real-world DNS

App	Median $\Delta(t_{App}\text{-}t_{tcpdump})$ (μs)	StdDev $\Delta(t_{App}\text{-}t_{tcpdump})$ (μs)
JavaApp	1,254	658
Haystack	1,211	303

Can measure contextualized "**real-world**" traffic performance

Enables reactive measurements **[**Allman+Paxson, PAM 2008**]**

Community feedback:

- What are your reactions both as users and researchers?
- How can we improve app usability and mobile transparency?
- What are the most challenging, worrying and urging aspects of mobile systems?

Visit: www.haystack.com