Haystack: A Tool For In Situ Mobile Traffic Analysis

Narseo Vallina-Rodriguez, Srikanth Sundaresan, Christian Kreibich, Mark Allman, Vern Paxson (ICSI)
Abbas Razaghpanah, Phillipa Gill (Stony Brook University)

INTERNATIONAL COMPUTER SCIENCE INSTITUTE

How Do Mobile Apps Operate Behind The Scenes?

Mobile apps access a wide range of resources and sensitive data on smartphones

Users and researchers remain in the dark about the operation and performance of their apps

- Which information do apps extract from their phones?
- Who do they share this information with?
- What is the importance of user's private data for the mobile ecosystem?

There are no **tools** to understand the mobile ecosystem **at scale** and **in the wild** yet

Our Measurement Platform: Haystack

Haystack is a **handset-**, **traffic-**, and **user-centric** platform that provides high-fidelity insight about security and privacy aspects of mobile apps **in the wild**

Uses Android's *VPN permission* to capture and forward all app's traffic to a **user space process:**

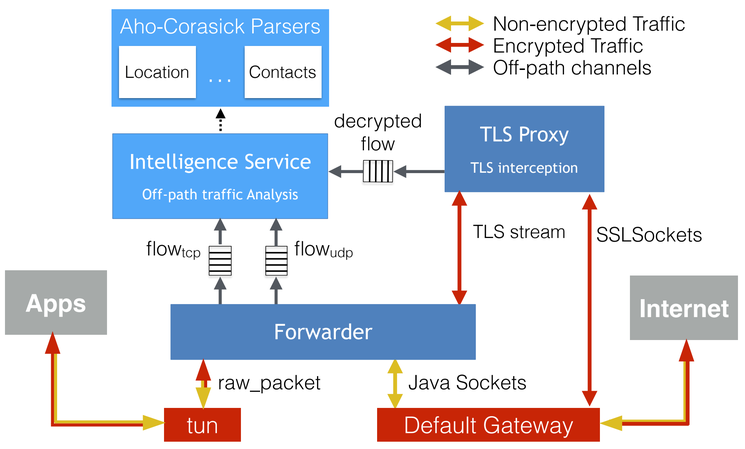
- Does not require rooted phones or custom firmware
- Eases user-friendly deployments through **Google Play** at scale

Runs **locally** on the phone:

- Processes user's traffic completely on the phone
- Allows the observation and correlation of crucial information with traffic:
  - ✦ App and OS context (system logs, app generating traffic)
  - ✦ Device and network status (location, screen state, connectivity)
  - ✦ Device-specific information (MAC addresses, IMEI, IMSI)
  - ✦ User-related information (accounts, messages, calls, contacts)

Performs **TLS interception**:

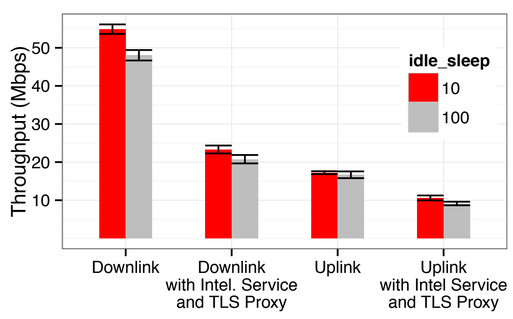
- Identifies privacy leaks even in encrypted traffic
- Flags potential TLS interception practices in the network
- Ensures correct use of TLS by apps



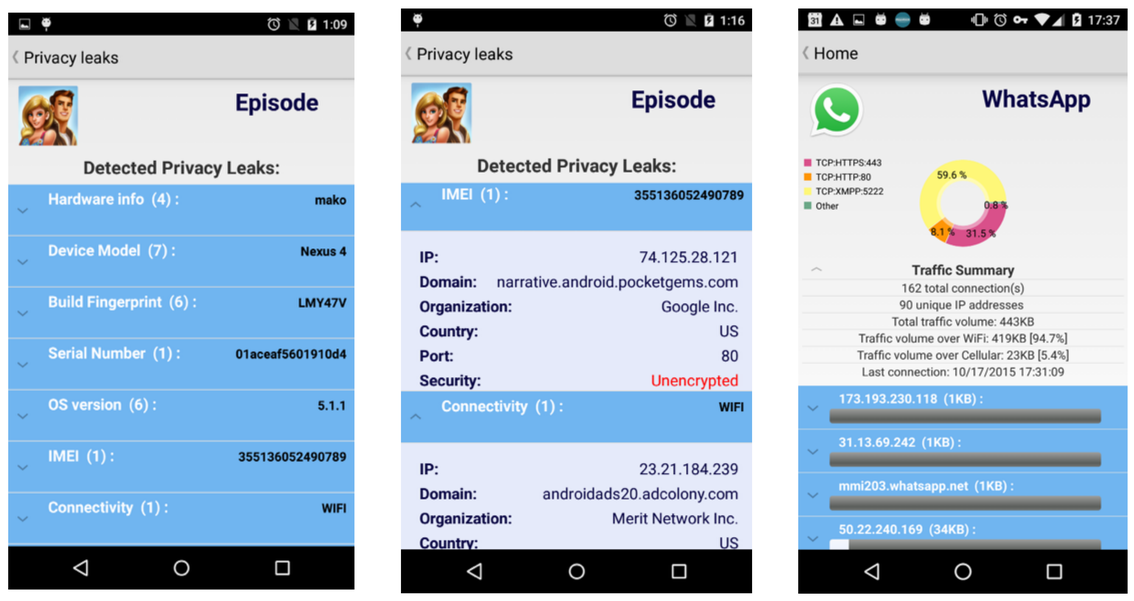
Performance Evaluation

Haystack adapts its behavior to traffic demands to reduce overhead

CPU Load	**IDLE**	<1%
	HD Video	<45%
Battery Overhead	**Idle**	<3%
	HD Video	<9%
Latency Overhead		4ms
Maximum Throughput		52Mbps



Haystack App: A Tool for the User



Available for **free** in **Google Play** app store

Data collection process approved by ICSI/UC Berkeley's IRB

Preliminary Results: Unique Identifiers

Tracker	App	Category	Destination	Protocol	# of Installs
MAC	AdVenture Capitalist!	Game	api.swrve.com	HTTP	1M-5M
	O.S. WiFi Mapper	Tools	api.staircase3.com	HTTP	100K-500K
Serial #	Just Eat	Lifestyle	public.je-apis.com	HTTPS	1M-5M
	Basis Peak	Health	api.mybasis.com	HTTPS	10K-50K
	Saavn	Music	www.saavn.com	HTTP	10M-50M
	DH5	Game	gdid.gameloft.com	HTTP	1M-5M
	ParkWhiz	Transport	crashlytics.com	HTTPS	100K-500K
	Slack	Business	slack-msgs.com	HTTPS	1M-5M
	Hootsuite	Social	api.hootsuite.com	HTTPS	1M-5M
IMEI	Saavn	Music	s.saavn.com	HTTP	10M-50M
	Gallery Dr. Cleaner	Tools	dws.flayvr.com	HTTP	100K-500K
	Yi Sports Kamera	Photog.	log.xiaoyi.com	HTTP	100K-500K
	AdVenture Capitalist!	Game	api.swrve.com	HTTP	1M-5M
	Paytm	Shopping	54.230.190.136	HTTP	10M-50M

Future Research Efforts

Haystack platform enables a number of studies beyond traffic characterization and privacy leak detection.

- User empowerment and awareness: Help users to stay in control of their traffic and personal information
- Censorship circumvention: Selective traffic forwarding through trusted VPNs
- Privacy and security firewall: Traffic anonymization and TLS validation
- Traffic performance enhancement, measurements and troubleshooting

Visit our website:
https://www.haystack.mobi

Google Play:
https://play.google.com/store/apps/details?id=edu.berkeley.icsi.haystack