Haystack

A multi-purpose mobile vantage point in user space

Christian Kreibich

Abbas Razaghpanah, Narseo Vallina, Srikanth Sundaresan, Phillipa Gill, Mark Allman, Vern Paxson

International Computer Science Institute
Stony Brook University
Part I

Background
Smartphones are everywhere
...but can we trust them?

- Privacy violations
- Malicious apps (ransomware, spyware, ...)
- App permission overuse
- Insecure operation
Investigation implies trade-offs
Tradeoffs: ISP traces

- Large scale
- Real-world traffic

- No context
- Encryption a problem

[IMC'12]
Tradeoffs: instrumented devices

- Device context
- Real-world activity
- Comprehensive analysis

- Small scale
- Tricky setup

[OSDI'10, IMC'13, CoNEXT'13]
Tradeoffs: static analysis

- Large scale
- Sufficient for some analyses
- No organic user activity

[NDSS'11, CCS'12-13, MobiSys'15]
Tradeoffs: proxy MITM

- Real-world traffic
- Comprehensive analysis

- No device context
- Detoured routes
- Higher trust hurdle

[CoNEXT'12, C2BID'15]
Tradeoffs: crowdsourced active measurement

- Large scale
- Comprehensive analysis
- No organic user activity

[CoNEXT'14, MobiSys'15, HotMiddlebox'15]
Can we do better?
Part II
The Haystack app
A few observations

• We want to run on the device
  • Best access to device context and the user

• We do not want to root the device
  • Drastically limits and skews deployment

• We like crowd-sourced measurement
  • Demonstrated large scale in Netalyzr
Mhmm...
Android's VPN API

- VPN apps don't require rooting
- Access to all packets sent by the device
- Nobody says you have to tunnel them!
- Instead inspect & rewrite, and interact directly with intended destination
The Haystack app

A user-centric, on-device measurement platform that intercepts and inspects network traffic and app activity in user-space.
Architecture overview
Architecture overview

- Non-encrypted Traffic
- Encrypted Traffic
- Off-path channels

Apps → Forwarder
  - raw_packet
  - Java Sockets

Default Gateway → Internet
Architecture overview

- Aho-Corasick Parsers
  - Location
  - Contacts

- Intelligence Service
  - Off-path traffic Analysis

- Forwarder
  - $\text{flow}_{\text{tcp}}$
  - $\text{flow}_{\text{udp}}$

- Apps
  - raw_packet

- Internet
  - java Sockets
  - Default Gateway

- Non-encrypted Traffic
- Encrypted Traffic
- Off-path channels
Architecture overview

Aho-Corasick Parsers
- Location
- Contacts

Intelligence Service
- Off-path traffic Analysis

Forwarder
- raw_packet

Tun

Default Gateway

Internet

SSL Sockets

TLS Proxy
- TLS stream
- TLS interception

Non-encrypted Traffic
- Encrypted Traffic
- Off-path channels

flow_{tcp}
flow_{udp}
Polling state machine

- Sleep
- tun read
- nio read
Polling state machine

- cpu active
- cpu inactive
- incoming packet
- outgoing packet
- packet 1
- packet 2
- t_{buff} Packet Buffering Time
- t_{proc} Packet Processing Time
- idle cycles
- idle sleep (ms)
Part III
Evaluation
CPU and power overhead
**CPU and power overhead**

Interactive: 10ms  
Idle: 100ms

![Graph showing CPU and power overhead](image)

**max_idle_cycles**
- 1
- 10
- 100
- 200

idle_sleep (ms)
CPU and power overhead

Interactive: 10ms

Idle: 100ms

Power: +3.1% when idle, +9.1% when busy
Latency overhead

![Graph showing latency overhead with varying idle_sleep (ms) and max_idle_cycles (1, 10, 100, 200).]
Latency overhead

Interactive: 3.4ms
Idle: 60ms

TCP Connection Time (ms)

idle_sleep (ms)

max_idle_cycles 1 10 100 200
Throughput

![Bar chart showing throughput comparison between downlink and uplink with TA enabled and disabled.]
Part IV
Use cases
Pilot study

450 users, 1340 apps, 6 months, app-focused data collection
Traffic properties

- Less than 20% of apps only send cleartext
- 22% of flows are encrypted
- 59% of TLS-using apps allow MITM
- 40 apps generate local IoT traffic
Privacy-related leakages
App properties

- 15% of apps do not come from Google Play
  - Pre-installed or from other stores
  - They create 22% of the observed traffic
- 78% use third-party trackers
  - Advertising, analytics, social net interactions, ...
https://haystack.mobi/panopticon
Future work

• More direct user involvement
  • Notify of leakages as they happen
  • Highlight third-party footprint
• Alter / block traffic
  • Suppress third-party trackers
• Reactive measurement
  • Active measurement can give context or inform traffic alterations
Summary
The Haystack app

A user-centric, on-device measurement platform based on the Android VPN API

- Access to organic user activity
- Optionally inspects TLS
- Has full device context
- Enables user interaction
- No rooting required, thus scalable
- (Modest) performance overheads
- Subject to crowdsourcing biases
Thanks!

https://haystack.mobi

christian@icir.org

@ckreibich