Improving Transparency and Control in Mobile Internet Systems

David Choffnes

Northeastern University

with INRIA, U. Washington, Berkeley, MSR & Stony Brook

A little about my work

Primary focus is reliability and performance

  ‣ Ono, NEWS, LIFEGUARD [SIGCOMM '06, '08, '10, '12, '13]

Privacy

  ‣ Web logins sent in the clear [W-MUST '11]

  ‣ Traffic analysis resistance [SIGCOMM '13]

Measurement

  ‣ ISP characterization [NSDI '13]

  ‣ View from the edge [CCR '10, SIGCOMM '11]

Mobile systems

  ‣ Mobile network characterization [PAM '14 (x2)]

  ‣ Improving transparency and control using software middleboxes

A little about my work

Primary focus is reliability and performance

- Ono, NEWS, LIFEGUARD [SIGCOMM '06, '08, '10, '12, '13]

Privacy

- Web logins sent in the clear [W-MUST '11]
- Traffic analysis resistance [SIGCOMM '13]

Measurement

- ISP characterization [NSDI '13]
- View from the edge [CCR '10, SIGCOMM '11]

Mobile systems

- Mobile network characterization [PAM '14 (x2)]
- Improving transparency and control using software middleboxes

A little about my work

Primary focus is reliability and performance

- Ono, NEWS, LIFEGUARD [SIGCOMM '06, '08, '10, '12, '13]

Privacy

- Web logins sent in the clear [W-MUST '11]
- Traffic analysis resistance [SIGCOMM '13]

Measurement

- ISP characterization [NSDI '13]
- View from the edge [CCR '10, SIGCOMM '11]

Mobile systems

- Mobile network characterization [PAM '14 (x2)]
- Improving transparency and control using software middleboxes

Improving Transparency and Control in Mobile Internet Systems

A little about my work

Primary focus is reliability and performance
- Ono, NEWS, LIFEGUARD [SIGCOMM '06, '08, '10, '12, '13]

Privacy
- Web logins sent in the clear [W-MUST '11]
- Traffic analysis resistance [SIGCOMM '13]

Measurement
- ISP characterization [NSDI '13]
- View from the edge [CCR '10, SIGCOMM '11]

Mobile systems
- Mobile network characterization [PAM '14 (x2)]
- Improving transparency and control using software middleboxes

I'm looking for grad students!

Background and Motivation

▶ **Smartphones are amazing**

  ▶ Tiny powerful computers *in our pockets* (or purses)

  ▶ All kinds of cool apps

  ▶ Potentially fast bandwidth

  ▶ Keep us connected all the time

**Improving Transparency and Control in
Mobile Internet Systems**

Background and Motivation

▸ **Smartphones are amazing**

   ▸ Tiny powerful computers *in our pockets* (or purses)

   ▸ All kinds of cool apps

      ▸ That are doing who knows what with our data/network

   ▸ Potentially fast bandwidth

   ▸ Keep us connected all the time

**Improving Transparency and Control in
Mobile Internet Systems**

Background and Motivation

▸ **Smartphones are amazing**

  ▸ Tiny powerful computers *in our pockets* (or purses)

  ▸ All kinds of cool apps

    ▸ That are doing who knows what with our data/network

  ▸ Potentially fast bandwidth

    ▸ But with data caps and spendy plans

  ▸ Keep us connected all the time

**Improving Transparency and Control in
Mobile Internet Systems**

Background and Motivation

▸ **Smartphones are amazing**
  ▸ Tiny powerful computers *in our pockets* (or purses)

  ▸ All kinds of cool apps
    ▸ That are doing who knows what with our data/network
  ▸ Potentially fast bandwidth
    ▸ But with data caps and spendy plans
  ▸ Keep us connected all the time
    ▸ Until you take 3 steps to the left
    ▸ Unless your provider decides to block services

Why do these issues persist?

Researchers, operators and end-users lack good tools for **understanding** and **controlling** *network* activity from their mobile systems

‣ Visibility
  ‣ How do we understand mobile network performance?
  ‣ What are apps doing with our data?
  ‣ What are carriers doing to our traffic?

‣ Control
  ‣ What can we do about any of these problems?
  ‣ How do we deploy a solution?

Outline

▸ Intro

▸ Goals and Challenges

▸ Meddle: Software Middleboxes for Mobile

▸ Mobilyzer: Controllable Mobile Network Measurements

▸ Wrap up

Improving Transparency and Control in Mobile Internet Systems

Focus of this talk

- ▸ **Network interactions from mobile systems**
  - ▸ Most of what mobile devices provide uses the network
  - ▸ These translate into user-facing problems
    - ▸ Has implications for privacy, battery life, service contract cost
  - ▸ Still poorly understood

Technology
THE BUSINESS AND CULTURE OF OUR DIGITAL LIVES, FROM THE L.A. TIMES

Pandora's Android app gathers personal info, report says

TECHNOLOGY

T-Mobile Investigates Data-Network Outage
By THOMAS GRYTA
July 3, 2012 1:58 p.m. ET

OCTOBER 24, 2013, 8:03 PM | 💬 10 Comments

LinkedIn's New Mobile App Called 'a Dream for Attackers'

Goals

- **Network Characterization**
  - What **is**
  - What **if**
  - End-to-end (capture view from end systems)
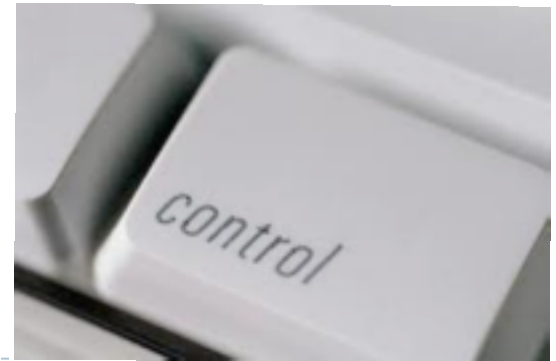 - Across apps, devices, carriers, over time, over space

Improving Transparency and Control in Mobile Internet Systems

Goals

- **Network Characterization**
  - What **is**
  - What **if**
  - End-to-end (capture view from end systems)
  - Across apps, devices, carriers, over time, over space

- **Network Control**
  - Inject
    - Controlled experiments with active measurements
  - Interpose
    - Modify, block, shape existing traffic

**Improving Transparency and Control in
Mobile Internet Systems**

Challenges
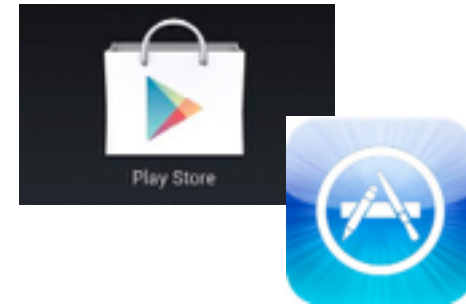
▸ **Measurement cost**
  ▸ Limited quota, battery life

▸ **Deployability / Portability**
  ▸ Limited view from apps
  ▸ Constrained by app store policy

▸ **Broad coverage**
  ▸ Over time, space, across carriers

Improving Transparency and Control in Mobile Internet Systems

Challenges

- ## Measurement cost
  - Limited quota, battery life

- ## Deployability / Portability
  - Limited view from apps
  - Constrained by app store policy

- ## Broad coverage
  - Over time, space, across carriers

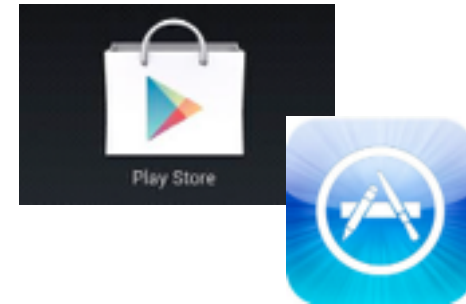**Improving Transparency and Control in Mobile Internet Systems**

Challenges

- ## Measurement cost
  - Limited quota, battery life

- ## Deployability / Portability
  - Limited view from apps
  - Constrained by app store policy

- ## Broad coverage
  - Over time, space, across carriers

Improving Transparency and Control in Mobile Internet Systems

Other approaches

Study	Measurement Type	Deployment model	Portability	Coverage
AT&T/ Telefonica	Passive	Instrument carrier network	Any OS	Single carrier
UConn WiFi	Passive	Instrument campus network	Any OS	Single campus WiFi
PhoneLab / TaintDroid	Active / Passive	Custom OS	Android	All traffic
Mobiperf / Speedtest	Active	App	Android / iOS	Generated traffic

Improving Transparency and Control in Mobile Internet Systems

Other approaches

Study	Measurement Type	Deployment model	Portability	Coverage
AT&T/ Telefonica	Passive	Instrument carrier network	Any OS	Single carrier
UConn WiFi	Passive	Instrument campus network	Any OS	Single campus WiFi
PhoneLab / TaintDroid	Active / Passive	Custom OS	Android	All traffic
Mobiperf / Speedtest	Active	App	Android / iOS	Generated traffic

Improving Transparency and Control in Mobile Internet Systems

Other approaches

Study	Measurement Type	Deployment model	Portability	Coverage
AT&T/ Telefonica	Passive	Instrument carrier network	Any OS	Single carrier
UConn WiFi	Passive	Instrument campus network	Any OS	Single campus WiFi
PhoneLab / TaintDroid	Active / Passive	Custom OS	Android	All traffic
Mobiperf / Speedtest	Active	App	Android / iOS	Generated traffic

Improving Transparency and Control in Mobile Internet Systems

Improving Visibility and Control (x2)

Meddle

▸ Passive, pervasive, portable system for measurement and control
Visibility: App traffic, across devices and carriers
Control: Software middleboxes to interpose on traffic

Mobilyzer

▸ Platform for controllable active measurements in mobile networks
Visibility: Active measurements independent of device activity
Control: Experiments that are easy to manage and deploy

Outline

▸ Intro

▸ Goals and Challenges

▸ **Meddle: Software Middleboxes for Mobile**

▸ Mobilyzer: Controllable Mobile Network Measurements

▸ **Wrap up**

Improving Transparency and Control in Mobile Internet Systems

Using the Middle to Meddle with Mobile

▸ **Challenge: Devices are locked down by carriers and OSes**

    ▸ How do we understand network performance?

    ▸ What are apps doing with our data?

    ▸ What can we do about it?

Using the Middle to Meddle with Mobile

- Challenge: Devices are locked down by carriers and OSes
  - How do we understand network performance?
  - What are apps doing with our data?
  - What can we do about it?

- Opportunity: (almost) all devices support VPNs
  - Tunnel traffic to a server we control (the middle)
  - Measure, modify, shape or block traffic with user opt-in

- Currently building *Meddle*
  - IRB-approved study with real subjects using their own phones
  - User incentives (e.g., custom network filtering)

Improving Transparency and Control in Mobile Internet Systems

Using the Middle to Meddle with Mobile

- Challenge: Devices are locked down by carriers and OSes
  - How do we understand network performance?
  - What are apps doing with our data?
  - What can we do about it?

- Opportunity: (almost) all devices support VPNs
  - Tunnel traffic to a server we control (the middle)
  - Measure, modify, shape or block traffic with user opt-in

- Currently building *Meddle*
  - IRB-approved study with real subjects using their own phones
  - User incentives (e.g., custom network filtering)

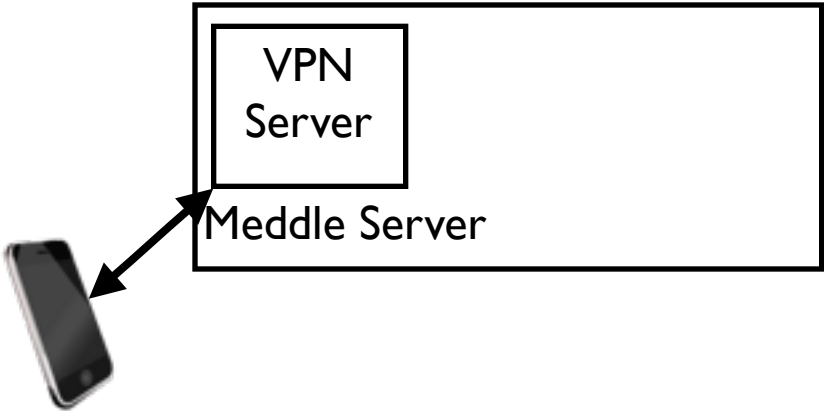# Using the Middle to Meddle with Mobile

- ▸ **Challenge: Devices are locked down by carriers and OSes**
  - ▸ How do we understand network performance?
  - ▸ What are apps doing with our data?
  - ▸ What can we do about it?

- ▸ **Opportunity: (almost) all devices support VPNs**
  - ▸ Tunnel traffic to a server we control (the middle)
  - ▸ Measure, modify, shape or block traffic with user opt-in

- ▸ **Currently building *Meddle***
  - ▸ IRB-approved study with real subjects using their own phones
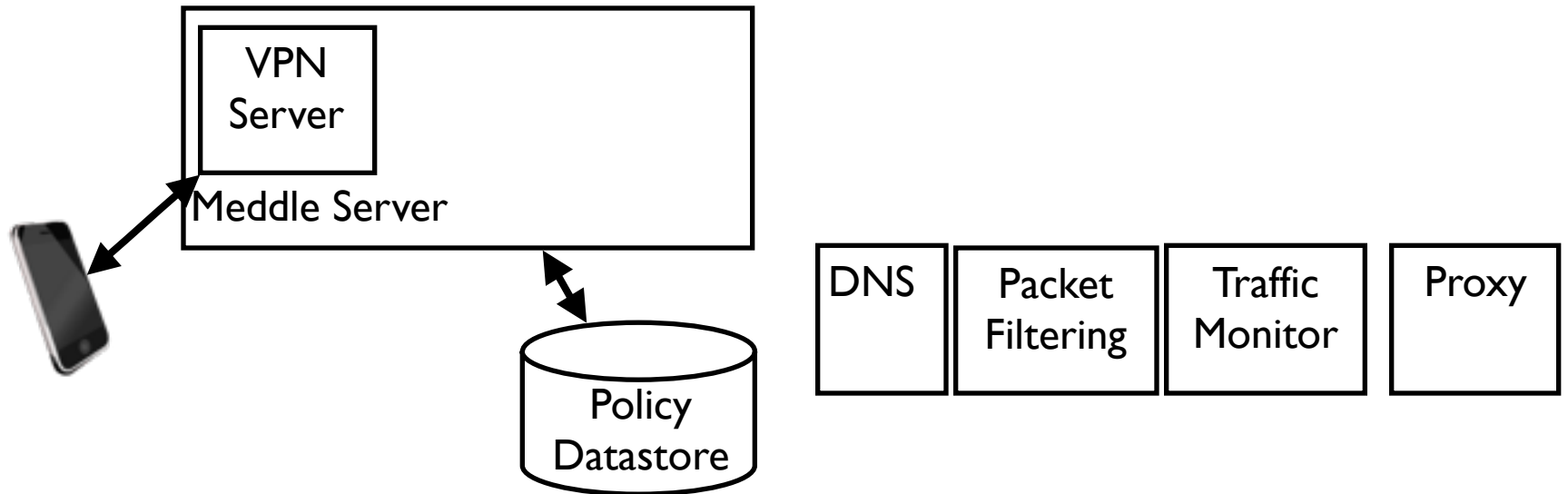 - ▸ User incentives (e.g., custom network filtering)

Improving Transparency and Control in Mobile Internet Systems

Meddle Research Challenges

▸ System design

▸ Network measurement and characterization

▸ Building useful Meddleboxes

**Improving Transparency and Control in
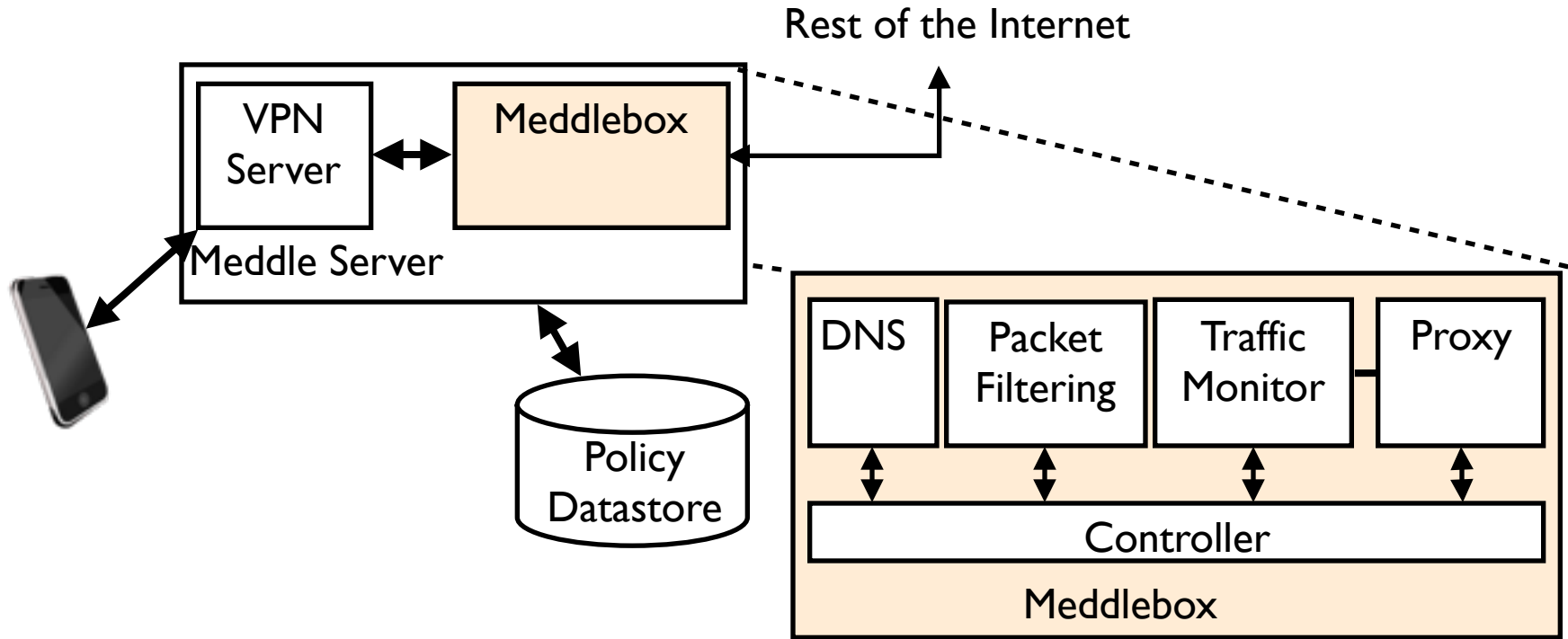Mobile Internet Systems**

Meddle Architecture

**Improving Transparency and Control in
Mobile Internet Systems**

Meddle Architecture

VPN
Server

Meddle Server

**Improving Transparency and Control in
Mobile Internet Systems**

Meddle Architecture

Improving Transparency and Control in Mobile Internet Systems

Meddle Architecture

**Improving Transparency and Control in
Mobile Internet Systems**

Meddle Architecture

**Improving Transparency and Control in
Mobile Internet Systems**

Meddle Architecture

Rest of the Internet

VPN Server ↔ Meddlebox

Meddle Server

Policy Datastore

| DNS | Packet Filtering | Traffic Monitor | Proxy |

Controller

Meddlebox

Celebrate the arrival of Surface.

Meddle Architecture

Improving Transparency and Control in Mobile Internet Systems

Implementation goals

Goal: Make it simple to add functionality that allows users to customize meddleboxes

Challenges:
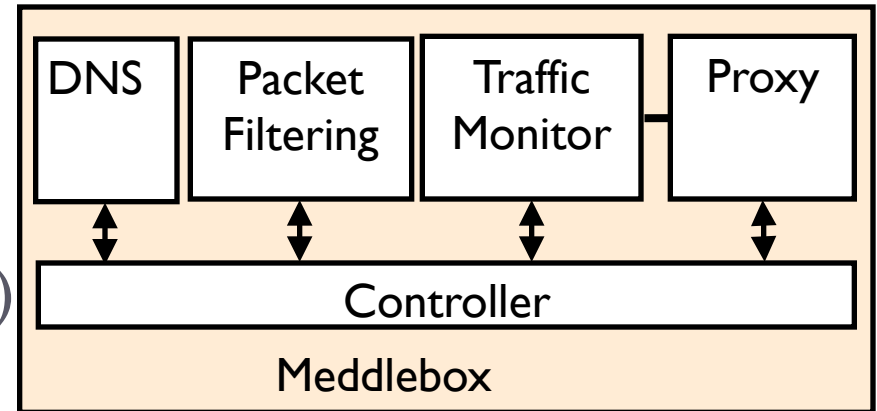‣ Interface (input/output streams)
‣ Composability (correctness)

**Improving Transparency and Control in
Mobile Internet Systems**

Implementation goals

Goal: Make it simple to add functionality that allows users to customize meddleboxes

Challenges:

‣ Interface (input/output streams)

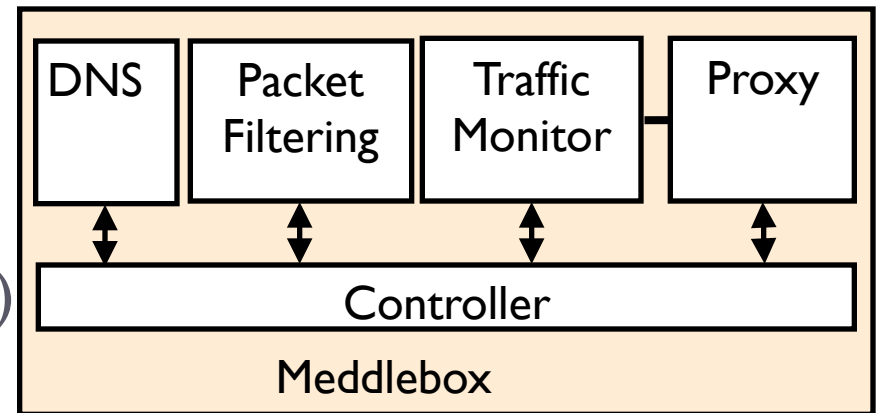‣ Composability (correctness)

‣ Scalability (support users globally)

Implementation goals

Goal: Make it simple to add functionality that allows users to customize meddleboxes

Challenges:
‣ Interface (input/output streams)
‣ Composability (correctness)
‣ Scalability (support users globally)
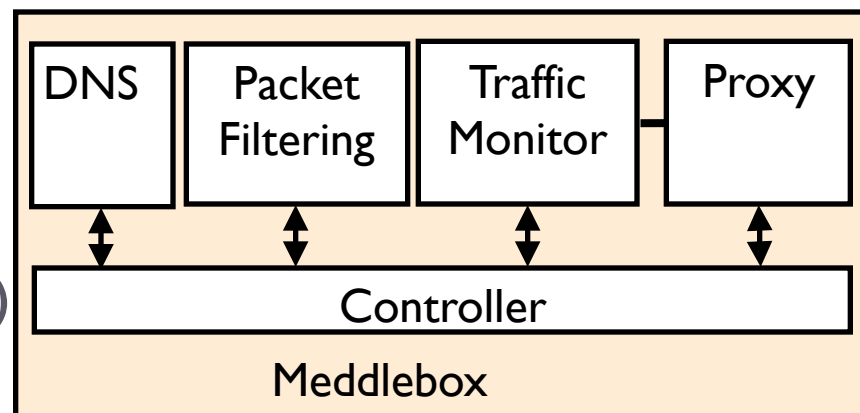‣ Privacy (Do not reduce protection of data)

Implementation goals

Goal: Make it simple to add functionality that allows users to customize meddleboxes

Challenges:
▸ Interface (input/output streams)
▸ Composability (correctness)
▸ Scalability (support users globally)
▸ Privacy (Do not reduce protection of data)
▸ Performance (Low cost)

DNS	Packet Filtering	Traffic Monitor	Proxy

Controller

Meddlebox

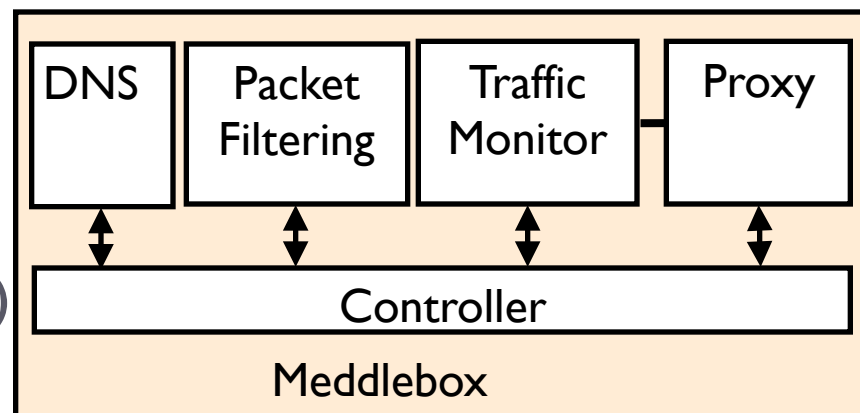**Improving Transparency and Control in Mobile Internet Systems**

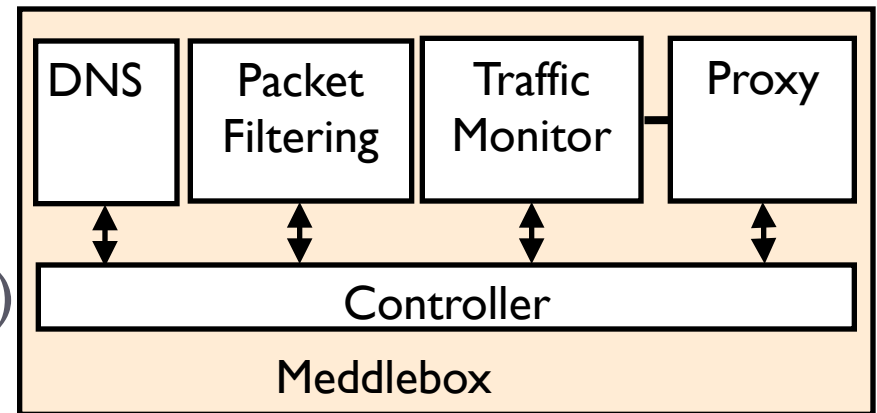# Implementation goals

Goal: Make it simple to add functionality that allows users to customize meddleboxes

Challenges:
‣ Interface (input/output streams)
‣ Composability (correctness)
‣ Scalability (support users globally)
‣ Privacy (Do not reduce protection of data)
‣ Performance (Low cost)
‣ Incentives

Improving Transparency and Control in Mobile Internet Systems

Current status

- ## Scalability/Correctness
  - Cloud! Also looking at SDN-based approaches
- ## Performance
  - Connection establishment: 0.6 to 2 seconds
  - Indirection delay: <10ms (within US)
  - Data/power overhead: 10%
    - Could be less with NULL IPSec encryption

Current status

- ## Scalability/Correctness
  - Cloud! Also looking at SDN-based approaches
- ## Performance
  - Connection establishment: 0.6 to 2 seconds
  - Indirection delay: <10ms (within US)
  - Data/power overhead: 10%
    - Could be less with NULL IPSec encryption
- ## Privacy
  - IRB-approved study
  - Phase I: Full packet traces, Phase II: Packet headers

Current status

- **Scalability/Correctness**
  - Cloud! Also looking at SDN-based approaches
- **Performance**
  - Connection establishment: 0.6 to 2 seconds
  - Indirection delay: <10ms (within US)
  - Data/power overhead: 10%
    - Could be less with NULL IPSec encryption
- **Privacy**
  - IRB-approved study
  - Phase I: Full packet traces, Phase II: Packet headers
- **Incentives**
  - Ad blocking, tracking the trackers, Web proxying

Improving Transparency and Control in Mobile Internet Systems

Meddle Research Challenges

‣ System design

‣ Network measurement and characterization

‣ Building useful Meddleboxes

Improving Transparency and Control in Mobile Internet Systems

Current deployment

- ## 25 devices, mainly in US and France
    - Recently added 80+ in China (h/t Kaigui Bian, Wenjun Hu)

- ## 15 iOS, 10 Android

- ## 202 days of data

Improving Transparency and Control in Mobile Internet Systems

Aggregate observations

▶ **How much privacy is there?**

  ▶ Majority of TCP flows (70%) encrypted

  ▶ These *cannot* be compressed by traditional middleboxes

**Improving Transparency and Control in
Mobile Internet Systems**

Aggregate observations

- **How much privacy is there?**
  - Majority of TCP flows (70%) encrypted
  - These *cannot* be compressed by traditional middleboxes

- **Do we lose much by monitoring only WiFi or only cell?**
  - Yes: 60/40 split between the two, different usage models

Aggregate observations

- How much privacy is there?
  - Majority of TCP flows (70%) encrypted
  - These *cannot* be compressed by traditional middleboxes

- Do we lose much by monitoring only WiFi or only cell?
  - Yes: 60/40 split between the two, different usage models

- Should we always offload to WiFi?
  - WiFi and LTE latencies are nearly identical
  - Several devices consistently have better performance over cell

Aggregate observations

▸ **How much privacy is there?**
  ▸ Majority of TCP flows (70%) encrypted
  ▸ These *cannot* be compressed by traditional middleboxes

▸ **Do we lose much by monitoring only WiFi or only cell?**
  ▸ Yes: 60/40 split between the two, different usage models

▸ **Should we always offload to WiFi?**
  ▸ WiFi and LTE latencies are nearly identical
  ▸ Several devices consistently have better performance over cell

Improving Transparency and Control in Mobile Internet Systems

Aggregate observations

‣ How much privacy is there?
  ‣ Majority of TCP flows (70%) encrypted
  ‣ These *cannot* be compressed by traditional middleboxes

‣ Do we lose much by monitoring only WiFi or only cell?
  ‣ Yes: 60/40 split between the two, different usage models

‣ Should we always offload to WiFi?
  ‣ WiFi and LTE latencies are nearly identical
  ‣ Several devices consistently have better performance over cell

Improving Transparency and Control in Mobile Internet Systems

Aggregate observations

▸ **How much privacy is there?**
  ▸ Majority of TCP flows (70%) encrypted
  ▸ These *cannot* be compressed by traditional middleboxes

▸ **Do we lose much by monitoring only WiFi or only cell?**
  ▸ Yes: 60/40 split between the two, different usage models

▸ **Should we always offload to WiFi?**
  ▸ WiFi and LTE latencies are nearly identical
  ▸ Several devices consistently have better performance over cell

Improving Transparency and Control in Mobile Internet Systems

Aggregate observations

‣ **How much privacy is there?**
  ‣ Majority of TCP flows (70%) encrypted
  ‣ These *cannot* be compressed by traditional middleboxes

‣ **Do we lose much by monitoring only WiFi or only cell?**
  ‣ Yes: 60/40 split between the two, different usage models

‣ **Should we always offload to WiFi?**
  ‣ WiFi and LTE latencies are nearly identical
  ‣ Several devices consistently have better performance over cell

Improving Transparency and Control in Mobile Internet Systems

Aggregate observations (2)

- ## Do we need coverage across OSes, regions?
  - Yes: Apps, and how you identify them, differ substantially

iOS in Western countries

Android in Western countries

Popular user agents differ across OSes

Aggregate observations (2)

- ## Do we need coverage across OSes, regions?
  - Yes: Apps, and how you identify them, differ substantially

iOS in Western countries

Android in Western countries

iOS in China

Android in China

Improving Transparency and Control in Mobile Internet Systems

Aggregate observations (2)

▸ ## Do we need coverage across OSes, regions?

   ▸ Yes: Apps, and how you identify them, differ substantially

iOS in Western countries

Android in Western countries

iOS in China

Android in China

Improving Transparency and Control in Mobile Internet Systems

Aggregate observations (2)

- Do we need coverage across OSes, regions?
  - Yes: Apps, and how you identify them, differ substantially

iOS in Western countries

Android in Western countries

iOS in China

Android in China

Improving Transparency and Control in Mobile Internet Systems

Meddle Research Challenges

‣ System design

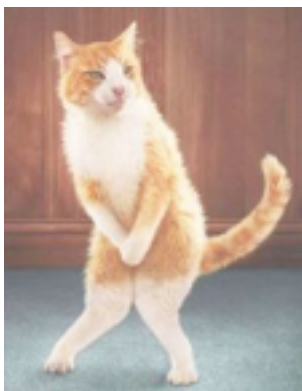‣ Network measurement and characterization

‣ **Building useful Meddleboxes**

**Improving Transparency and Control in
Mobile Internet Systems**

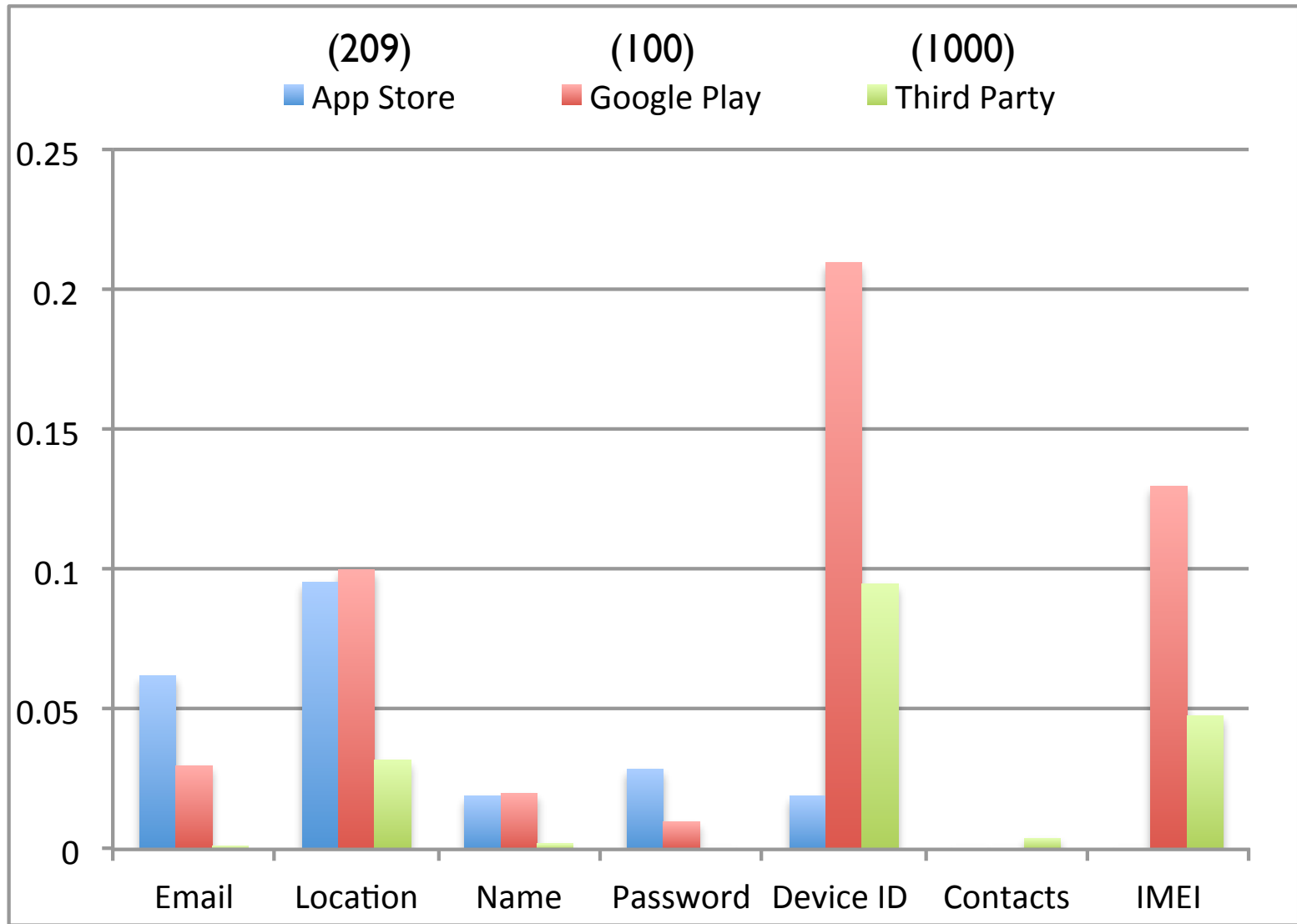# Meddleboxes under development

▸ **Improving performance**

  ▸ Optimizing Web pages for mobile networks

  ▸ Moving expensive network communication to the cloud

Improving Transparency and Control in Mobile Internet Systems
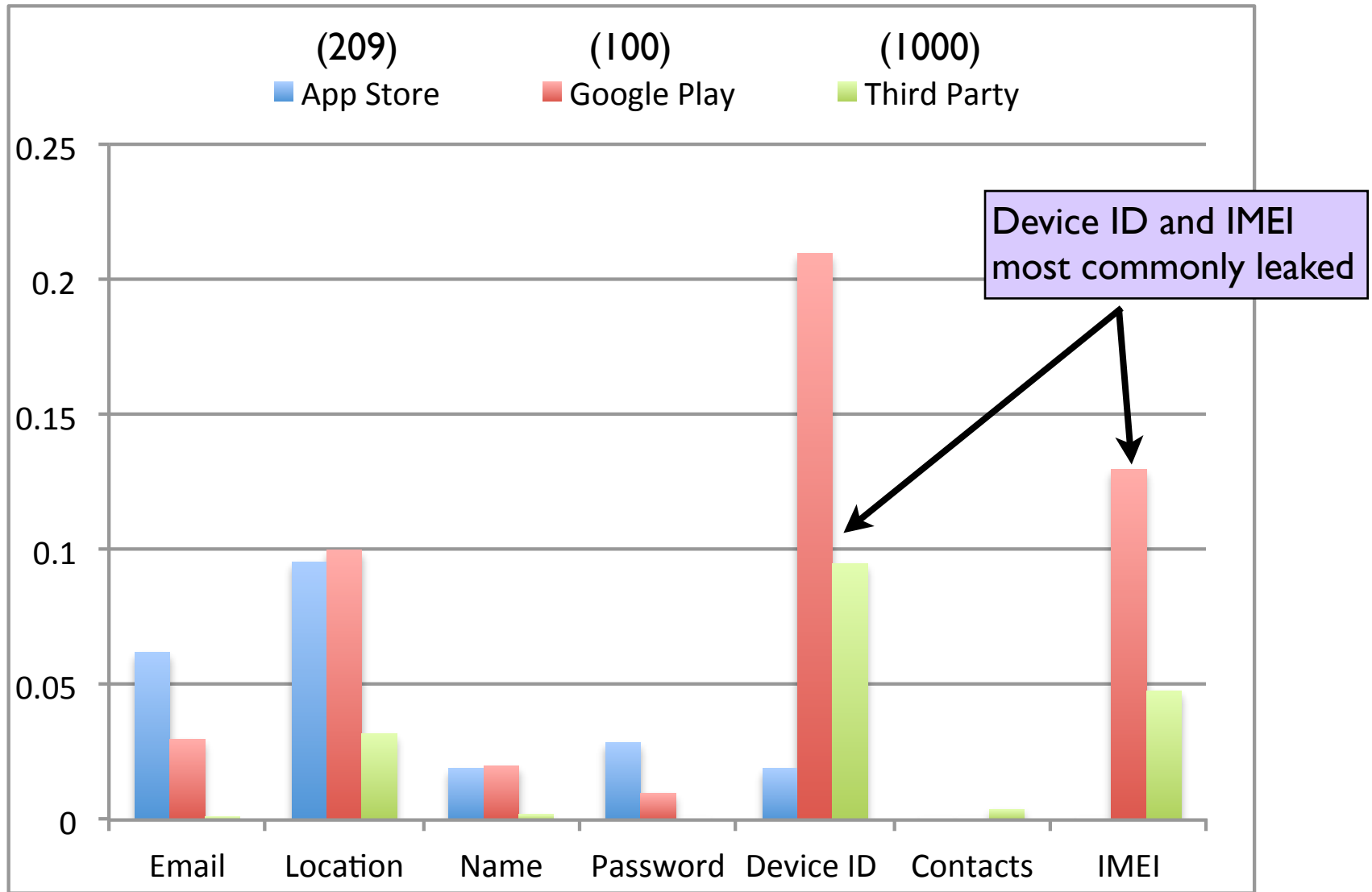
Meddleboxes under development

- **Improving performance**
  - Optimizing Web pages for mobile networks
  - Moving expensive network communication to the cloud

- **Revealing privacy policies**
  - Searches sent in clear (iOS < 6, Android pre JellyBean)
  - Extensive IMEI, AndriodID and location leakage (5-20% of apps)
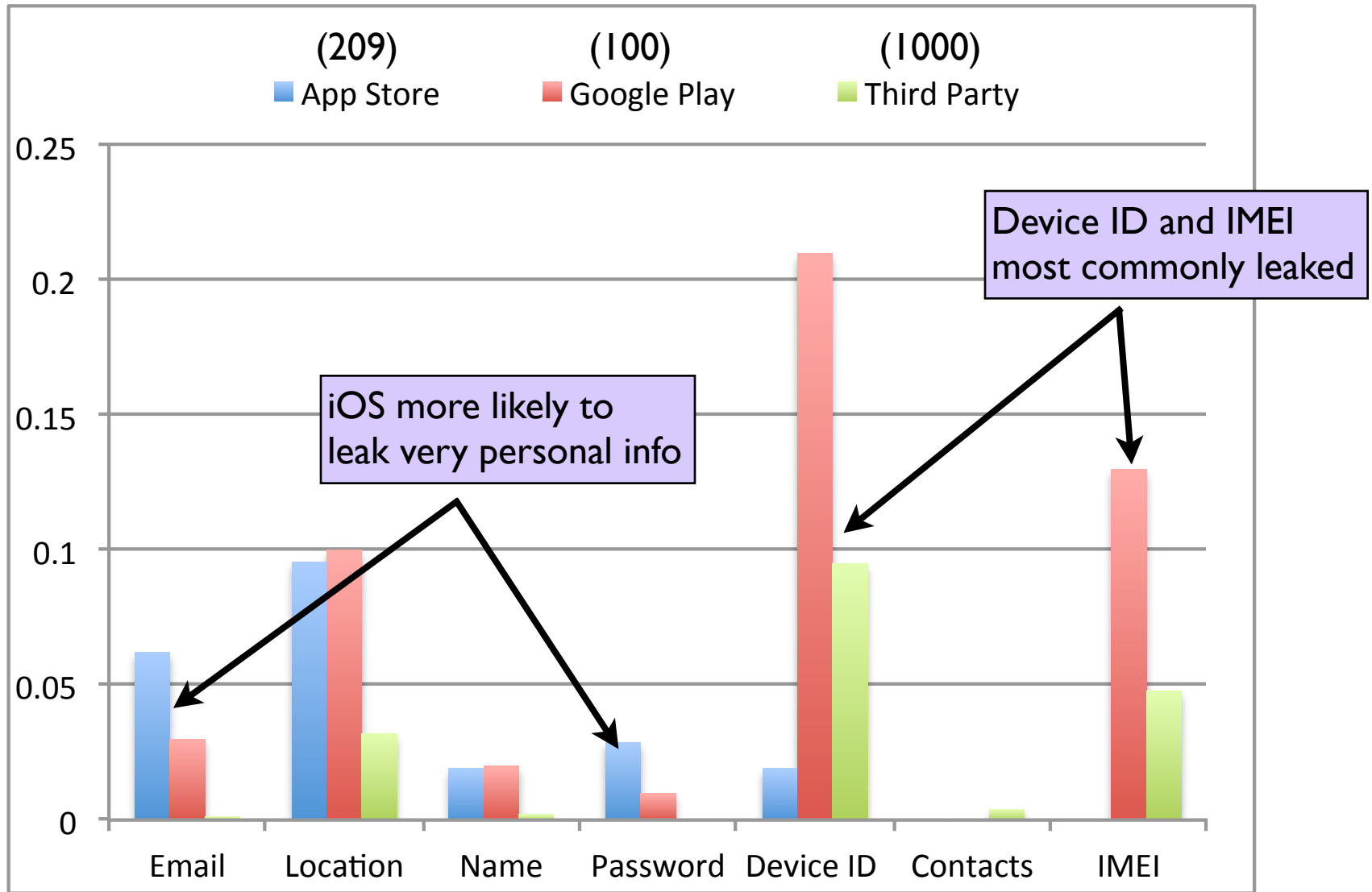 - E-mail, name, contacts, and even password sent in the clear!

Improving Transparency and Control in Mobile Internet Systems

Fraction of apps leaking PII

Improving Transparency and Control in Mobile Internet Systems

Fraction of apps leaking PII

Improving Transparency and Control in Mobile Internet Systems

Fraction of apps leaking PII

Improving Transparency and Control in Mobile Internet Systems

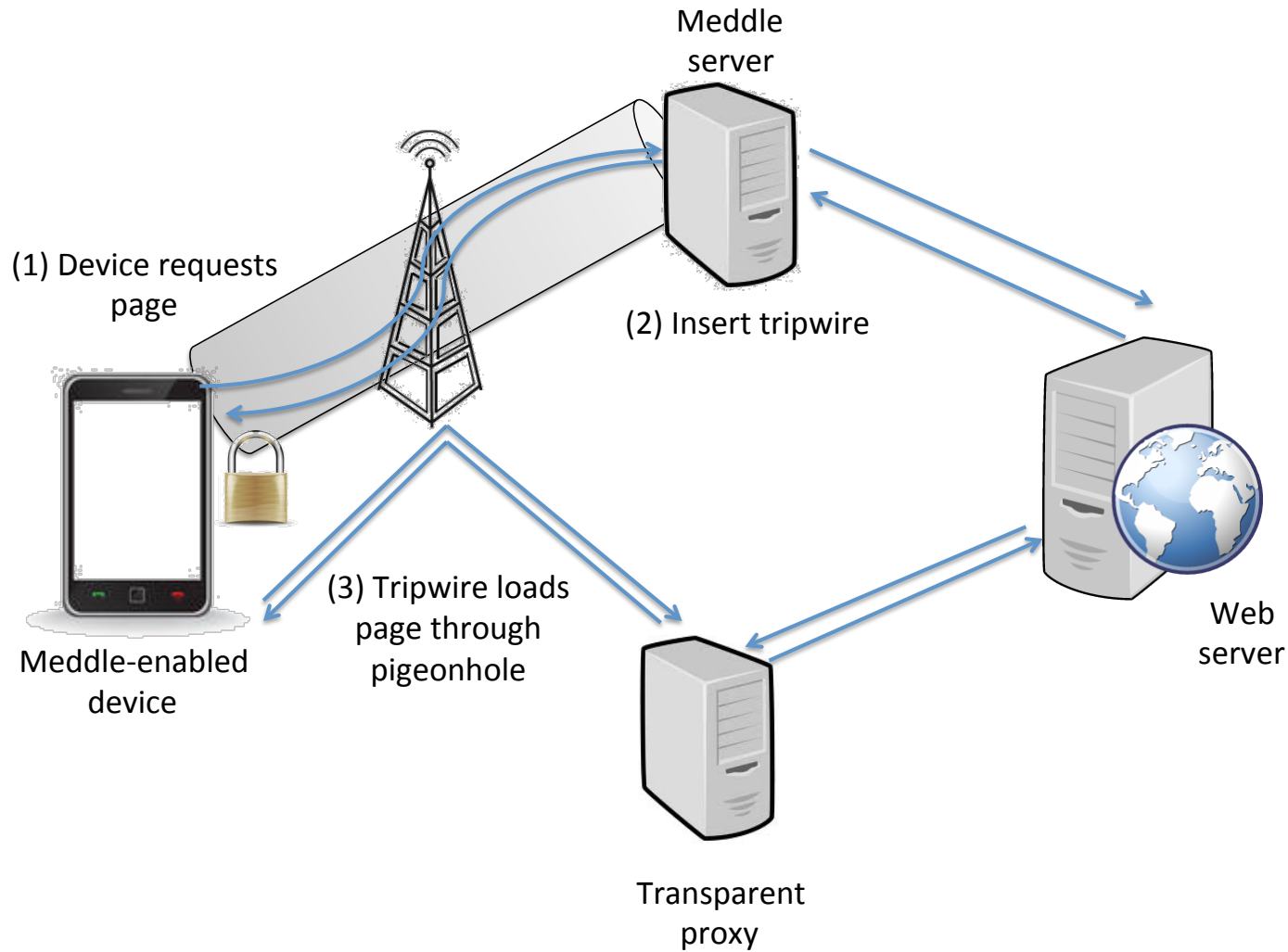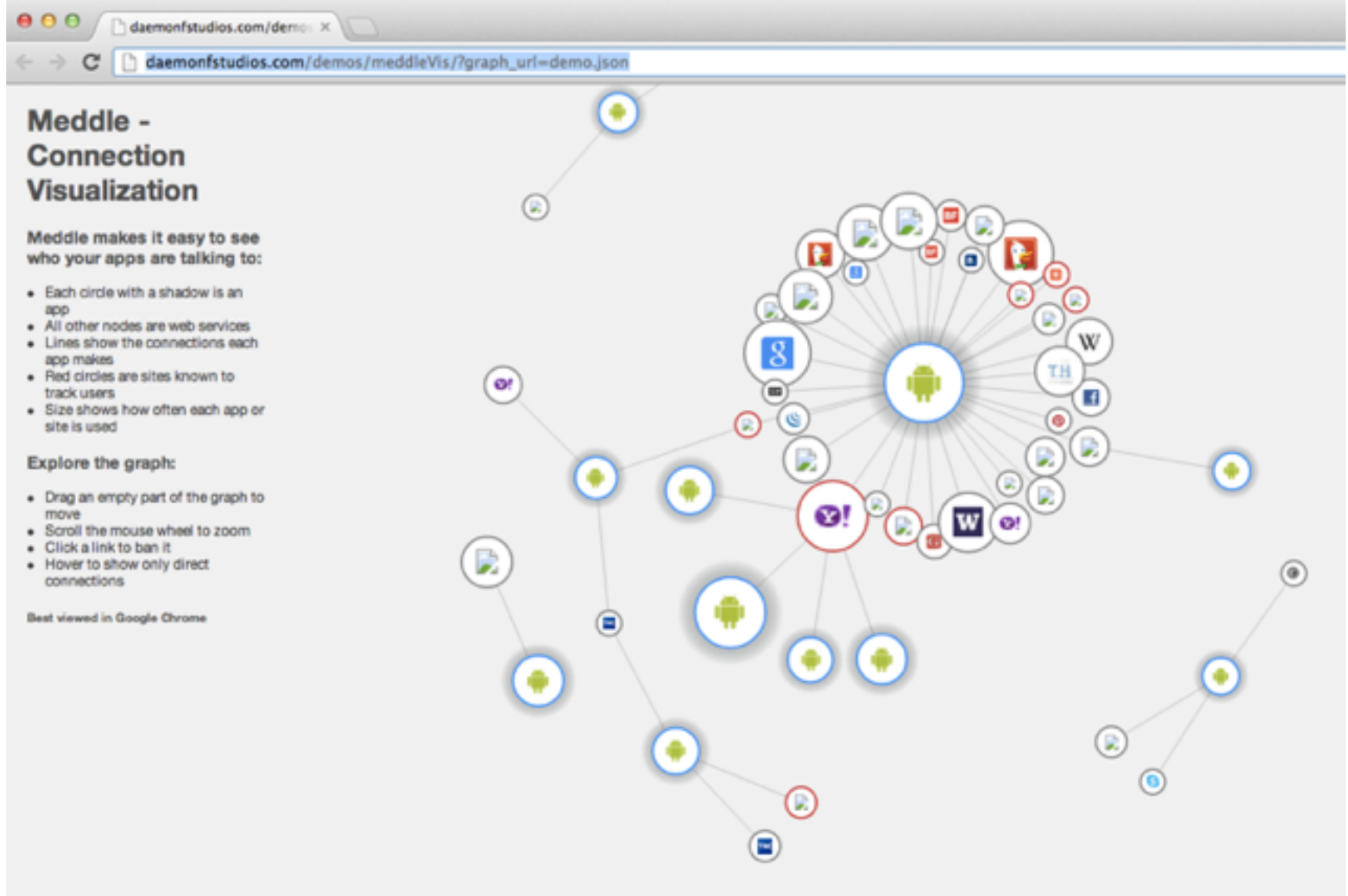# Meddleboxes under development

- Informing public policy
  - Detect differential treatment for traffic
  - Identify ISP interference, injection of traffic (Web Tripnets)
 - Understand differences across carriers, countries (e.g., China)

Improving Transparency and Control in Mobile Internet Systems

Detecting ISP interference with TripNets

Meddle server

(1) Device requests page

(2) Insert tripwire

Meddle-enabled device

(3) Tripwire loads page through pigeonhole

Transparent proxy

Web server

Improving Transparency and Control in Mobile Internet Systems

Meddle ConVis Reveals and Blocks Tracking

**Improving Transparency and Control in
Mobile Internet Systems**

Meddle Wrap up

▸ *Meddle* offers a new opportunity to experiment with middleboxes for mobile systems

  ▸ Currently expanding our set of users
  ▸ Building out several meddlebox systems

**Improving Transparency and Control in
Mobile Internet Systems**
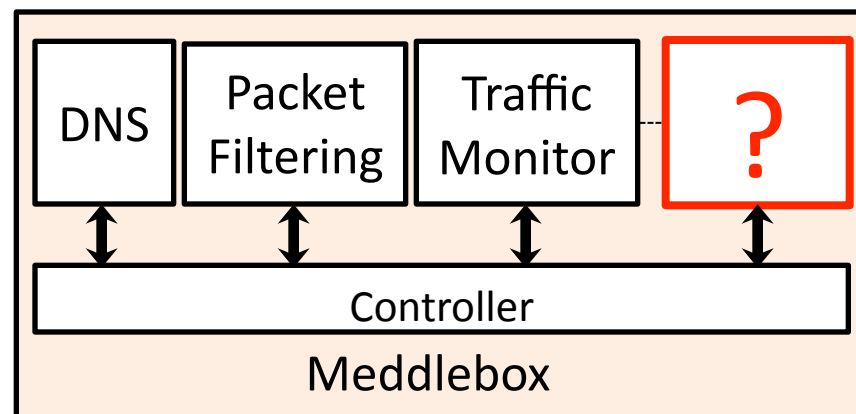
Concluding remarks

Most (mobile) network systems problems relate to lack of **visibility** and **control**

Visibility:

▸ Gather measurements from the end system perspective

▸ Build systems with incentives, low barrier to entry

Control:

▸ Meddle: Indirection + software middleboxes

▸ Mobilyzer: Coordinated active measurements at scale

Acknowledgements

▶ Meddle

  ▶ Ashwin Rao, Arnaud Legout (INRIA), Justine Sherry, Amy Tang (Berkeley), Wenjun Hu (MSRA), Arvind Krishnamurthy (UW), Arash Kakhki, Alan Mislove (NEU), Abbas Razzaghpanah, Phillipa Gill (Stony Brook)

▶ Mobilyzer

  ▶ Morley Mao (UMich), Kyriakos Zarifis, Tobi Flach, Srikanth Nori, Ethan Katz-Bassett, Ramesh Govindan (USC), Matt Welsh (Google), Dominic Hamon (M-Lab), Nick Feamster (GaTech)

▶ Support

  ▶ Google, M-Lab, NSF, CI Fellowship