PRIVACYCON

FEDERAL TRADE COMMISSION

DC // 1.14.16

Session 5:
Security and Usability

PRIVACYCON
FEDERAL TRADE COMMISSION

Sarthak Grover

Princeton University

The Internet of Unpatched Things

Co-author: Nick Feamster (Princeton University)

PRIVACYCON
FEDERAL TRADE COMMISSION

The Internet of Unpatched Things

Sarthak Grover and Nick Feamster
Princeton University

PrivacyCon '16

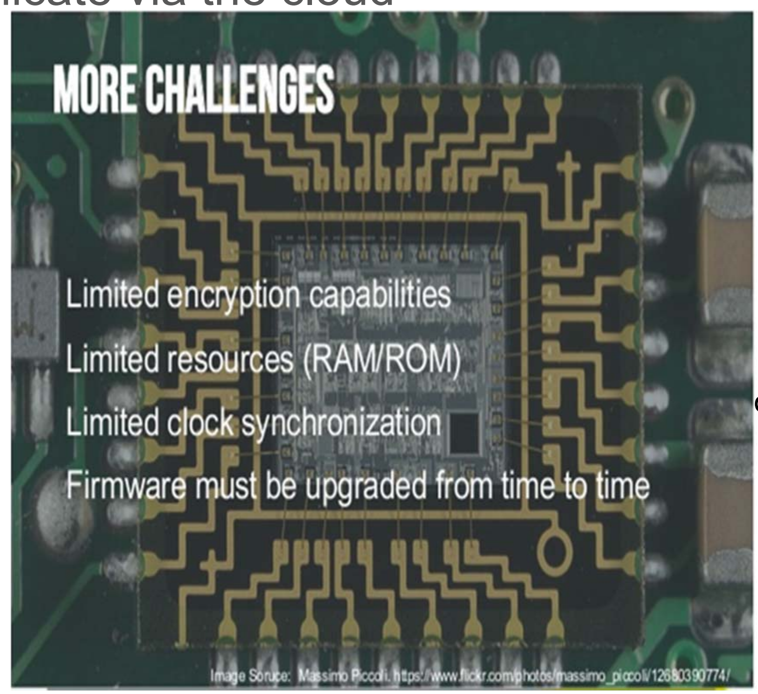
Current State of Consumer Smart Devices

Many different manufacturers, small startups, novice programmers

Low capability hardware, not enough for security protocols

Most data goes to an online server on the cloud

Even devices in the same home communicate via the cloud

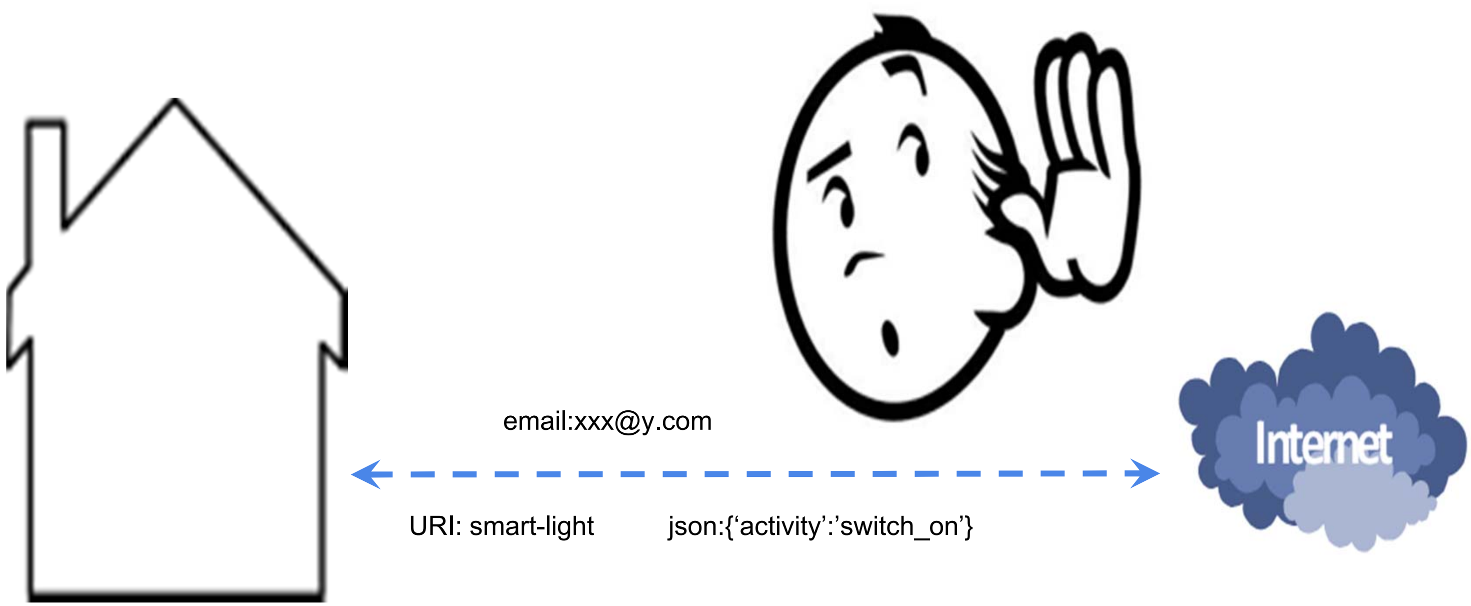


forgerock.com

Unpatched IoT Devices Put Our Privacy at Risk

IoT device network traffic:

- Leaks user information
- Identifies the device being used
- May also identify current user activity and behavior!



Case Study of Some Common Home IoTs

Digital Photoframe: Traffic Analysis

All traffic and feeds (RSS) cleartext over HTTP port 80

All actions sent to server in HTTP GET packet

Downloads radio streams in cleartext over different ports

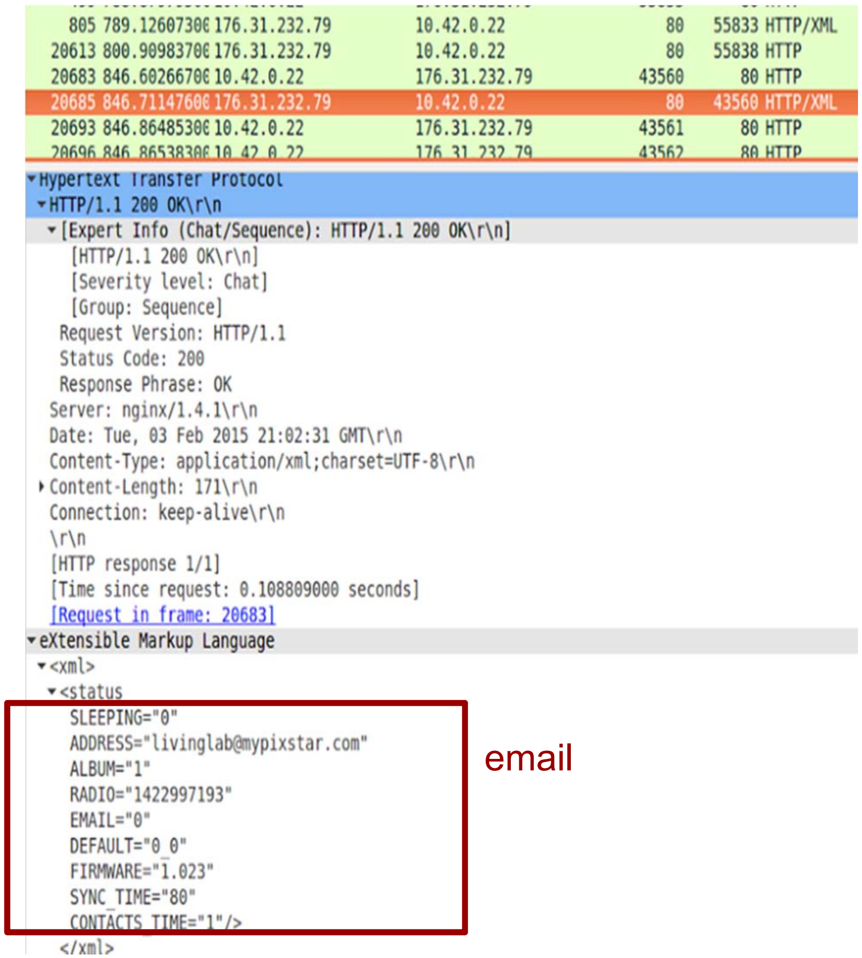
DNS queries: api.pix-star.com, iptime.pix-star.com

Photoframe: Privacy Issues

User **email ID is in clear** text when syncing account

Current **user activity** in clear text in HTTP GET

DNS queries and HTTP traffic identifies a pix-star photoframe



email

current activity

IP Camera: Traffic Analysis

All traffic over cleartext HTTP port 80, even though viewing the stream requires login password

Actions are sent as HTTP GET URI strings

Videos are sent as image/jpeg and image/gif in the clear

FTP requests also sent in clear over port 21, and FTP data is sent in clear text over many ports above 30,000

DNS query: www.sharxsecurity.com

IP Camera: Privacy Issues

Video can be recovered from FTP data traffic by network eavesdropper

DNS query, HTTP headers, and ports identify a Sharx security camera

private user data

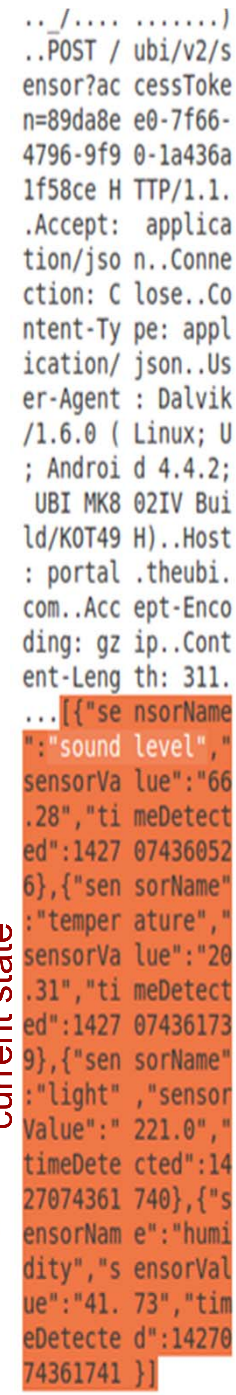
8	14.679939000	10.42.0.44	46.252.157.130	45962	21	FTP	74	Request: TYPE I
9	14.820736000	46.252.157.130	10.42.0.44	21	45962	FTP	96	Response: 200 TYPE is now 8-bit binary
10	14.821660000	10.42.0.44	46.252.157.130	45962	21	TCP	66	45962→21 [ACK] Seq=17 Ack=88 Win=8280 Len=0 TSval=1256532 TSe
11	14.823297000	10.42.0.44	46.252.157.130	45962	21	FTP	72	Request: PASV
12	14.957638000	46.252.157.130	10.42.0.44	21	45962	FTP	117	Response: 227 Entering Passive Mode (46,252,157,130,124,42)
13	14.959068000	10.42.0.44	46.252.157.130	60649	31786	TCP	74	60649→31786 [SYN] Seq=0 Win=5840 Len=0 MSS=1460 SACK_PERM=1 T
14	14.995413000	10.42.0.44	46.252.157.130	45962	21	TCP	66	45962→21 [ACK] Seq=23 Ack=139 Win=8280 Len=0 TSval=1256550 TS
15	15.092593000	46.252.157.130	10.42.0.44	31786	60649	TCP	74	31786→60649 [SYN, ACK] Seq=0 Ack=1 Win=14480 Len=0 MSS=1386 S
16	15.093262000	10.42.0.44	46.252.157.130	60649	31786	TCP	66	60649→31786 [ACK] Seq=1 Ack=1 Win=5840 Len=0 TSval=1256559 TS
17	15.096021000	10.42.0.44	46.252.157.130	45962	21	FTP	102	Request: STOR M_2015-03-17_17-37-23_348.jpg
18	15.230540000	46.252.157.130	10.42.0.44	21	45962	FTP	96	Response: 150 Accepted data connection
19	15.231793000	10.42.0.44	46.252.157.130	45962	21	TCP	66	45962→21 [ACK] Seq=59 Ack=169 Win=8280 Len=0 TSval=1256573 TS
20	15.233158000	10.42.0.44	46.252.157.130	60649	31786	FTP-DATA	1440	FTP Data: 1374 bytes
21	15.233544000	10.42.0.44	46.252.157.130	60649	31786	FTP-DATA	1440	FTP Data: 1374 bytes
22	15.233885000	10.42.0.44	46.252.157.130	60649	31786	FTP-DATA	1414	FTP Data: 1348 bytes
23	15.371483000	46.252.157.130	10.42.0.44	31786	60649	TCP	66	31786→60649 [ACK] Seq=1 Ack=1375 Win=17280 Len=0 TSval=258450
24	15.371922000	46.252.157.130	10.42.0.44	31786	60649	TCP	66	31786→60649 [ACK] Seq=1 Ack=2749 Win=20096 Len=0 TSval=258450
25	15.372409000	10.42.0.44	46.252.157.130	60649	31786	FTP-DATA	1440	FTP Data: 1374 bytes
26	15.372557000	10.42.0.44	46.252.157.130	60649	31786	FTP-DATA	1440	FTP Data: 1374 bytes
27	15.372976000	10.42.0.44	46.252.157.130	60649	31786	FTP-DATA	1440	FTP Data: 1374 bytes
28	15.373113000	10.42.0.44	46.252.157.130	60649	31786	FTP-DATA	1440	FTP Data: 1374 bytes

Ubi: Traffic Analysis

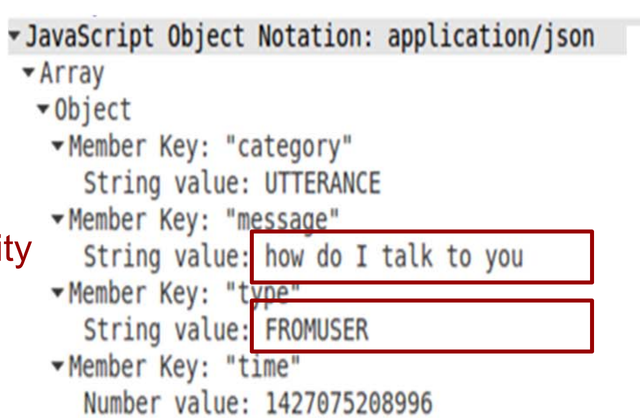
- All voice-to-text traffic sent in clear over port 80
- Activities sent in clear, and radio streamed over port 80
- Sensor readings are synced with server in the background over port 80
- Only communication with google API used HTTPS on port 443 and port 5228 (google talk)
- DNS query: portal.theubi.com, www.google.com, mtalk.google.com, api.grooveshark.com

Ubi: Privacy Issues

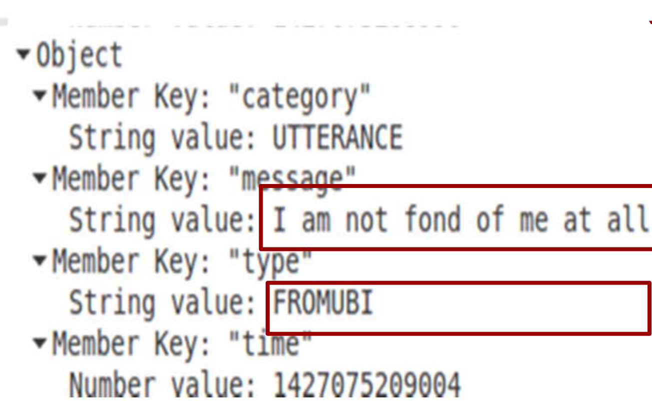
- Although HTTPS is clearly available, Ubi still uses HTTP to communicate to its portal. Eavesdropper can intercept **all voice chats and sensor readings** to Ubi's main portal
- Sensor values such as sound, temperature, light, humidity can identify if the user is home and currently active
- **Email in the clear** can identify the user
- DNS query, HTTP header (UA, Host) clearly identifies Ubi device



current state



current activity



Nest Thermostat: Traffic Analysis

All traffic to nest is HTTPS on port 443 and 9543

Uses TLSv1.2 and TLSv1.0 for all traffic

We found some incoming weather updates containing location information of the home and weather station in the clear. **Nest has fixed this bug after our report.**

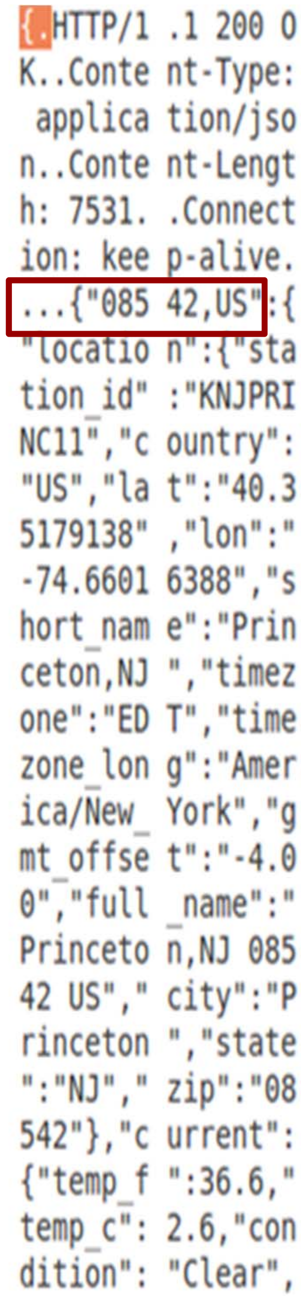
DNS query: time.nestlabs.com, frontdoor.nest.com, log-rts01-iad01.devices.nest.net. transport01-rts04-iad01.transport.home.nest.com

Nest: Privacy Issues

Fairly secure device: all outgoing personal traffic, including configuration settings and updates to the server, use HTTPS

*User zip code bug has been fixed

DNS query as well as the use of the unique port 9543 clearly identifies a Nest device.



Smartthings Hub: Traffic Analysis

All traffic over HTTPS on port 443 using TLS v1.2

No clear text port 80 traffic

Flows to an Amazon AWS instance running smartthings server

3-5 packets update every 10 sec in the background

DNS query: dc.connect.smartthings.com

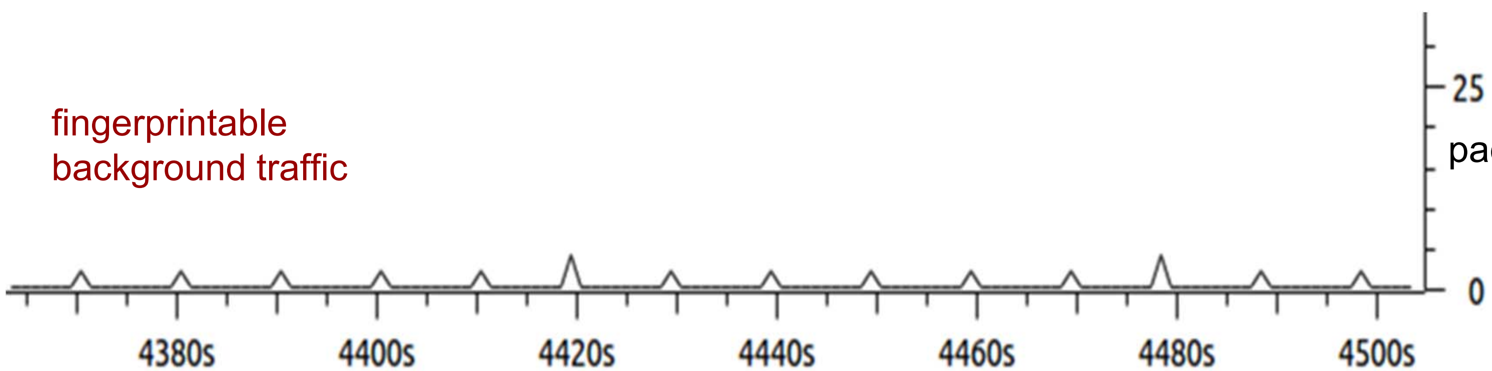
Smartthings: Privacy Issues

Very secure: No information about IoT devices attached to hub is leaked

Background updates every 10 seconds (over HTTPS) fingerprint the hub

DNS query identifies Smartthings hub, but not individual devices

fingerprintable
background traffic



packets/sec

Smartthings Traffic

Conclusion: Be Afraid!

Very difficult to enforce security standards

- Multiple manufacturers
- Low capability devices
- Use of non-standard protocols and ports

Difficult to maintain and patch due to low workforce and/or expertise

- Who is responsible? (ISPs? Consumers? Manufacturers?)
- Who is liable? Who should pay?

Conclusion: Be Afraid!

- Very difficult to enforce security standards
  - Multiple manufacturers
  - Low capability devices
  - Use of non-standard protocols and ports
- Difficult to maintain and patch due to low workforce and/or expertise
  - Who is responsible? (ISPs? Consumers? Manufacturers?)
  - Who is liable? Who should pay?

Can we solve this on the network? If so, how?

- How much information about user behavior do devices leak to the network?
- Can we offload device security to the home gateway or the cloud?

Thanks!

Vitaly Shmatikov

Cornell Tech

What Mobile Ads Know About Mobile Users

PRIVACYCON
FEDERAL TRADE COMMISSION

What Mobile Ads Know About Mobile Users

Vitaly Shmatikov

joint work with
Sooel Son and Daehyeok Kim

1.8 million

apps in Google Play Store

source: AppBrain

41% include at least one mobile advertising library

source: AppBrain

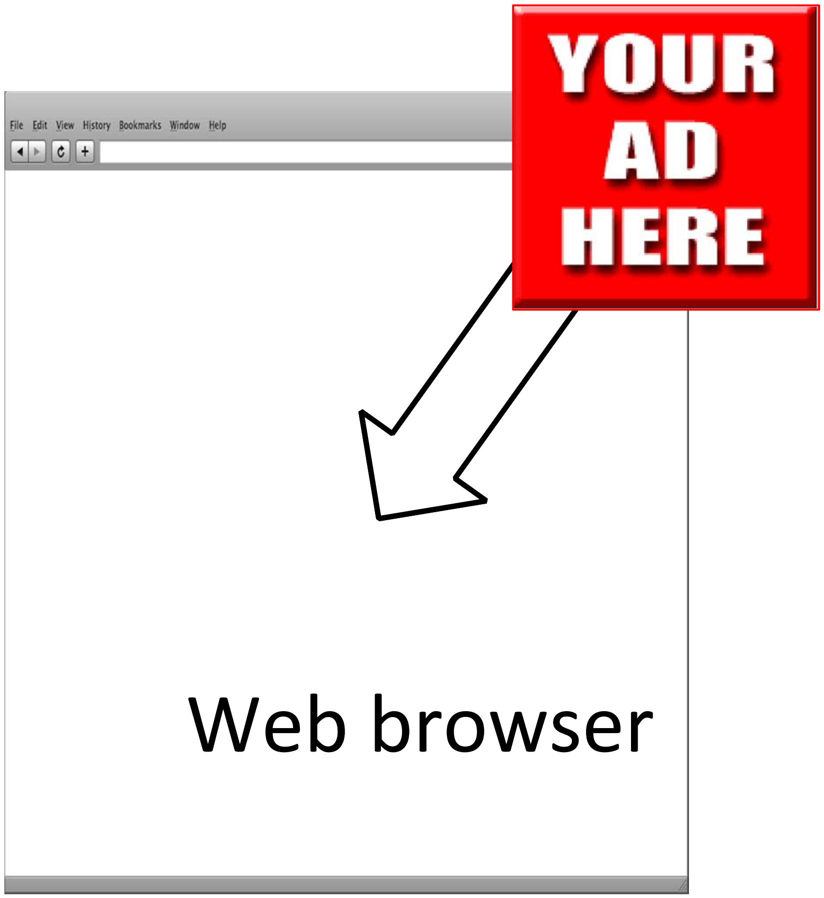
Every third

ad-supported app includes multiple advertising libraries

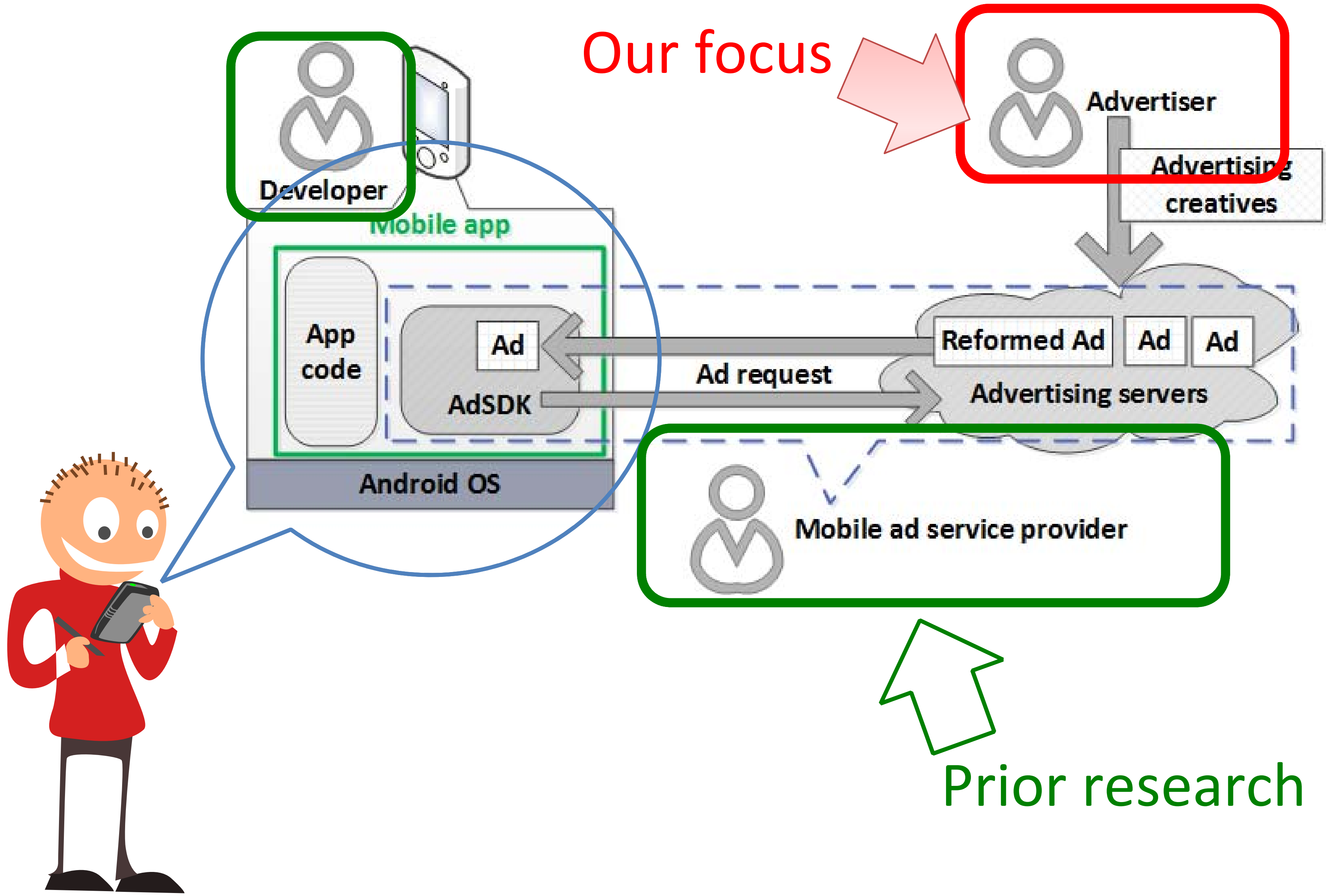
source: Shekhar et al. (USENIX Security 2012)

Web

Mobile


File Edit View History Bookmarks Window Help
YOUR AD HERE
Web browser


YOUR AD HERE
admob
AdMarvel
mopub
airpush
Ad library
Mobile app


Our focus
Advertiser
Advertising creatives
Developer
Mobile app
App code
Ad
AdSDK
Android OS
Ad request
Reformed Ad
Ad
Ad
Advertising servers
Mobile ad service provider
Prior research

Advertising services

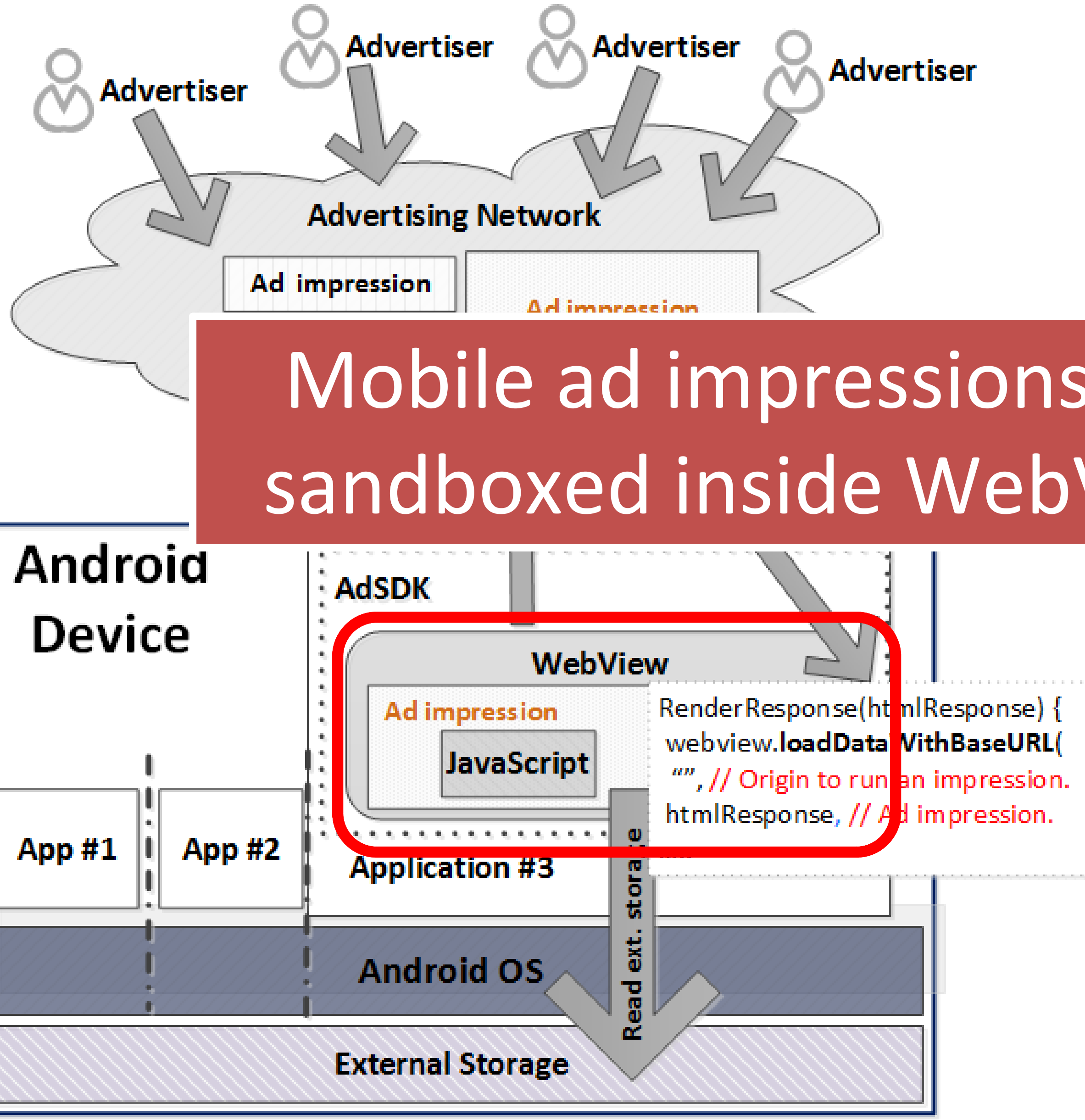
- Large businesses
  - AdMob (Google), Mopub (Twitter), AirPush, many others
- Provide **AdSDK libraries** to 100,000s of developers
- Millions of $ in revenue
- Reputation at stake

Advertisers

- Lots of fly-by-night operators
- Ads resold via auctions, brokers, exchanges
- No reputation at stake, no accountability
- Dynamic filtering and sanitization are hard

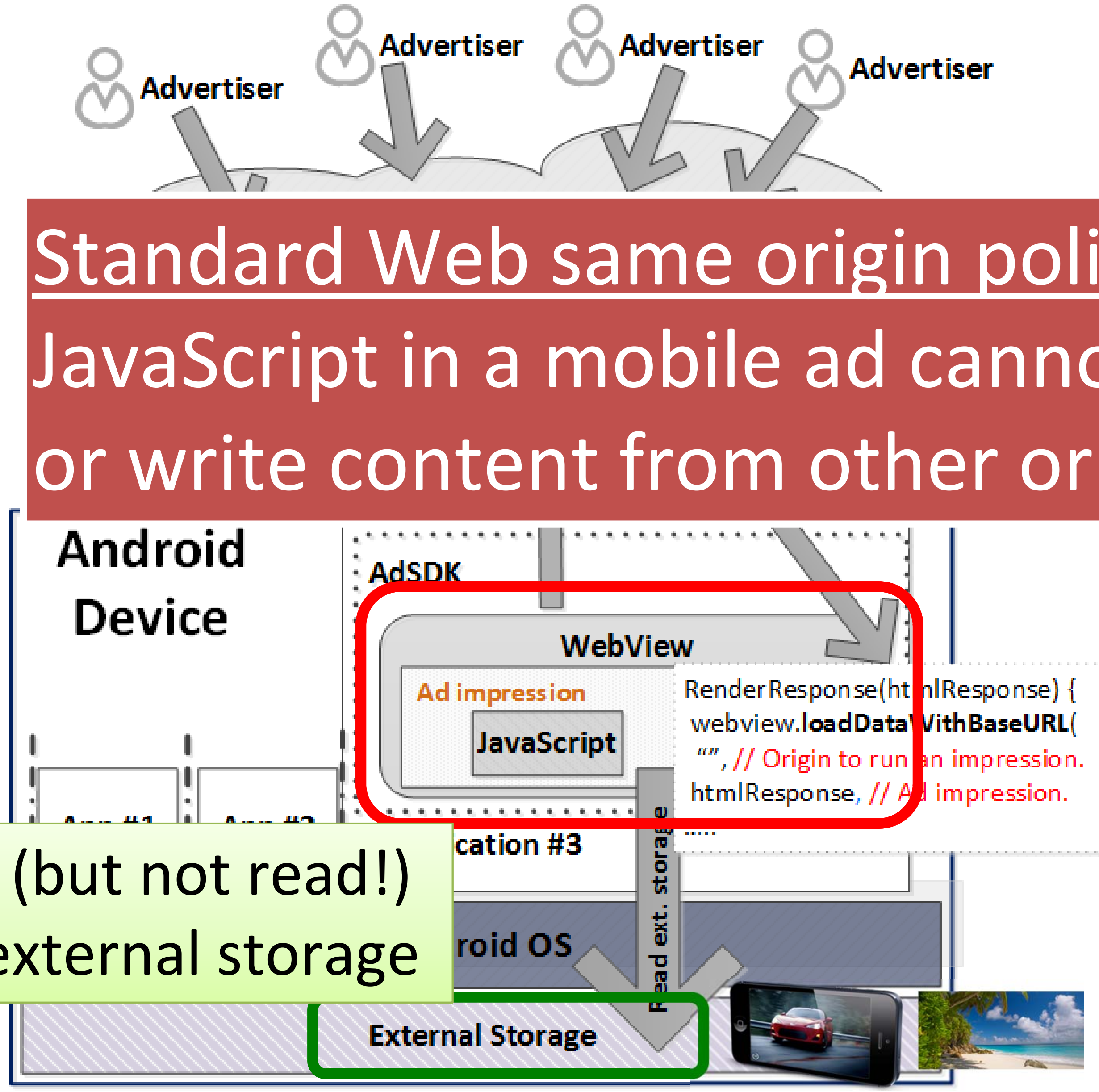
Ad libraries must protect users from malicious advertising

Mobile ad impressions are sandboxed inside WebView



Standard Web same origin policy:

JavaScript in a mobile ad cannot read or write content from other origins



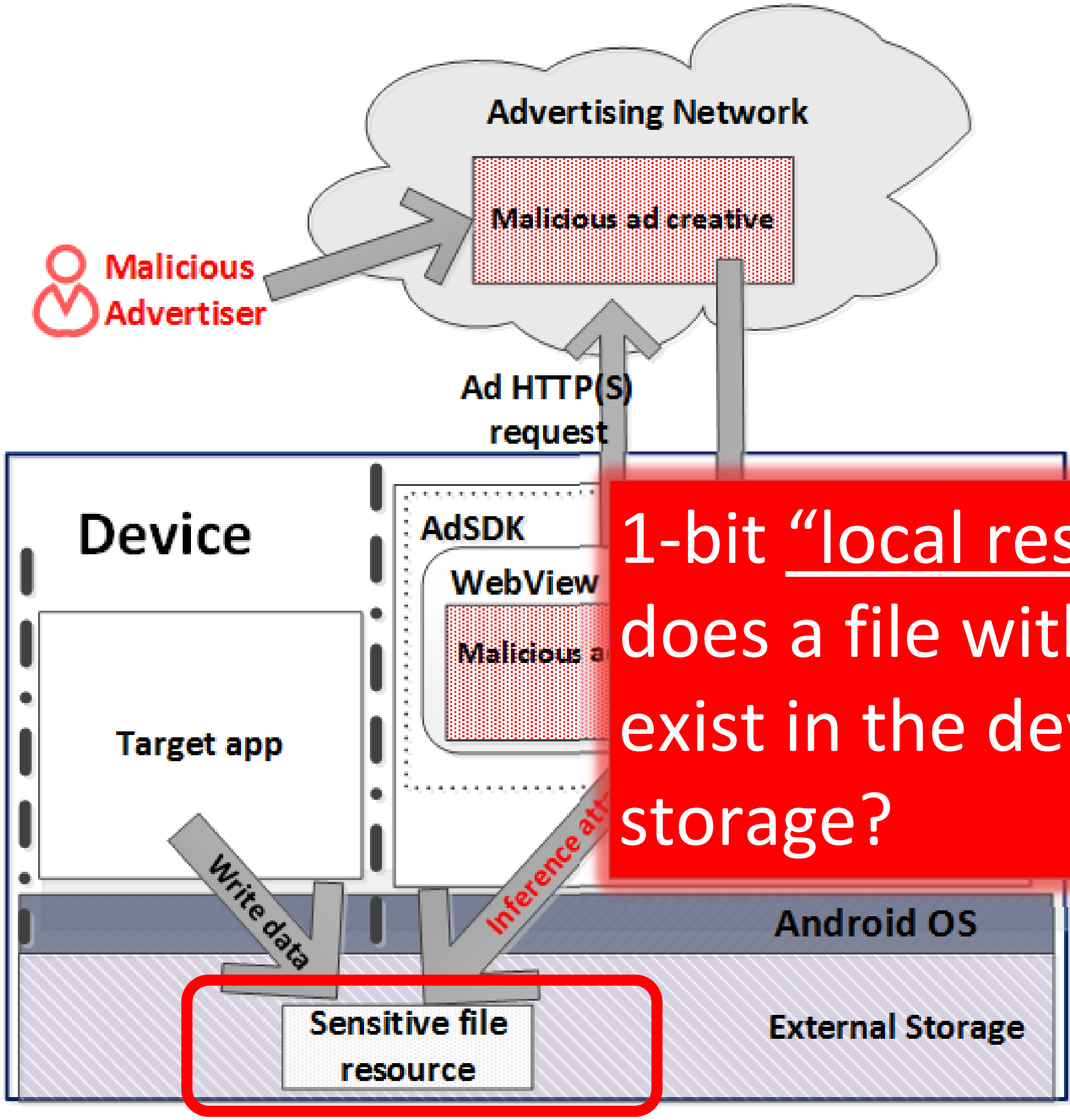
... can load (but not read!) files from external storage

Android External Storage

- Can be read or written by any app with appropriate permissions
- Media-rich mobile ads require access to external storage to cache images, video

iab MRAID 2.0

- Very weak access control for external storage
  - Any app can read any other app's files
  - But mobile ads are not apps. **Same origin policy = untrusted JavaScript cannot read ext-storage files … but can attempt to load them**



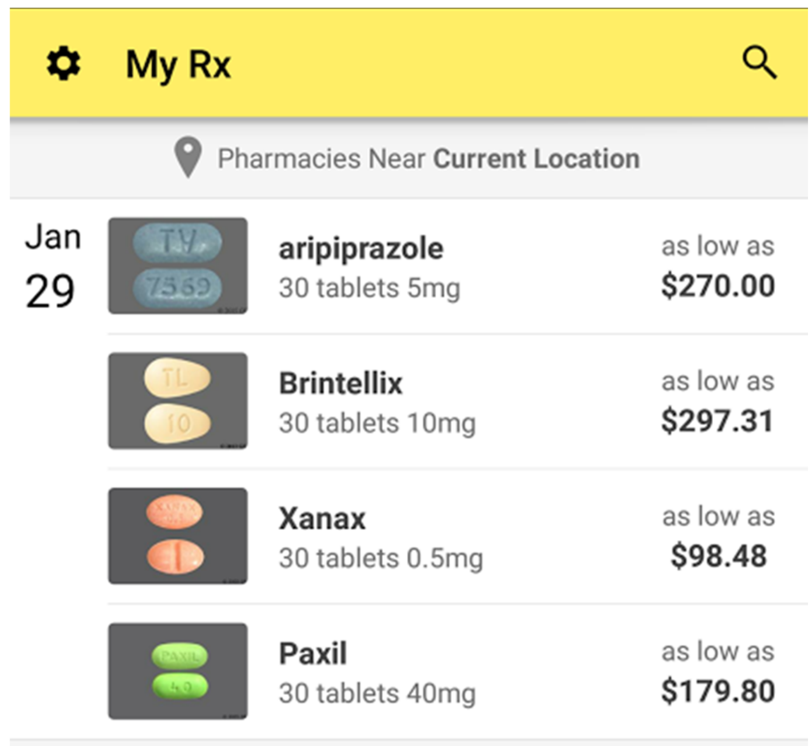
1-bit <u>“local resource oracle”:</u> does a file with a given name exist in the device’s external storage?

App for finding pharmacies, compare drug prices

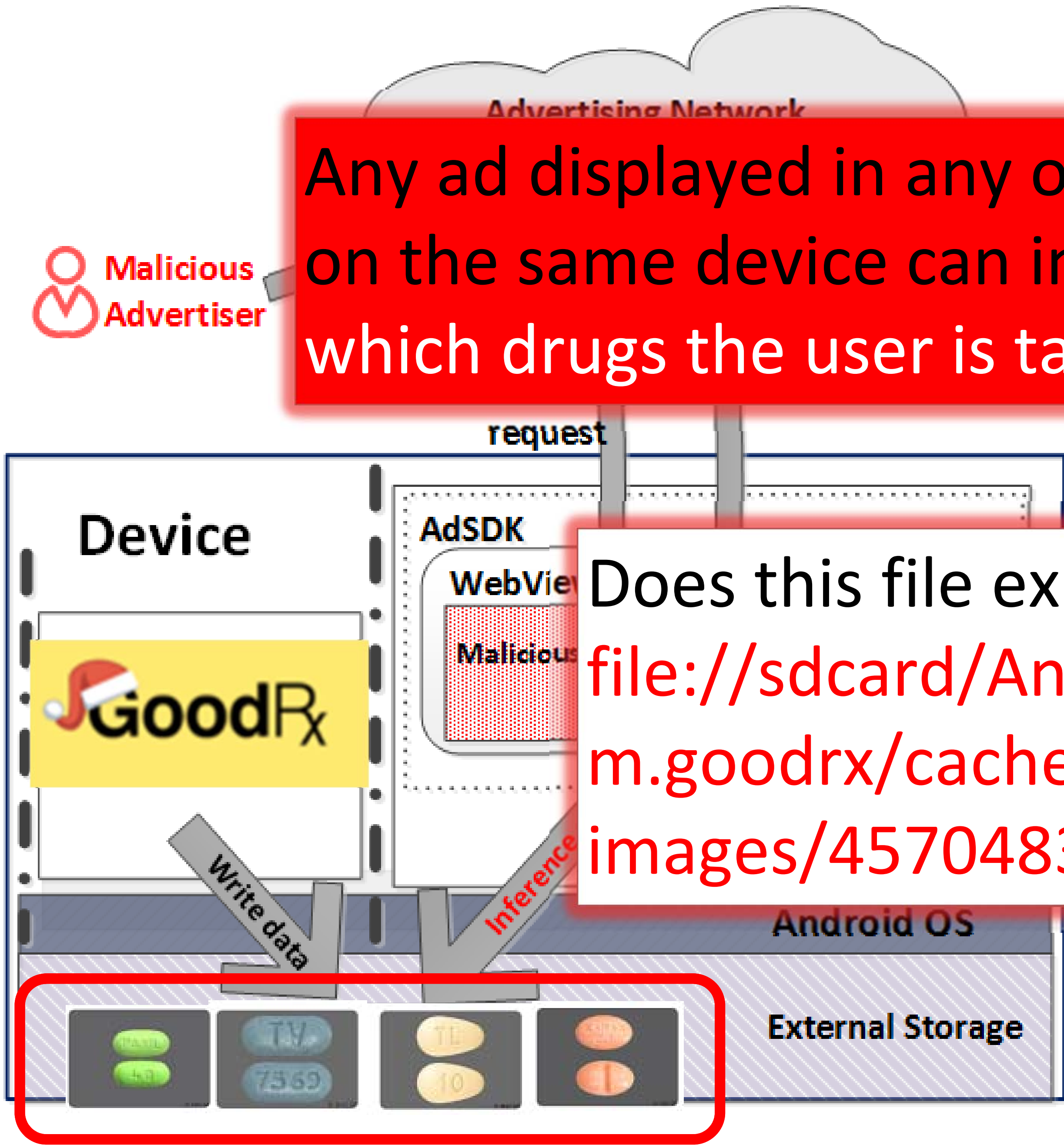
(1 to 5 million installs in Google Play Store)

Bookmark functionality

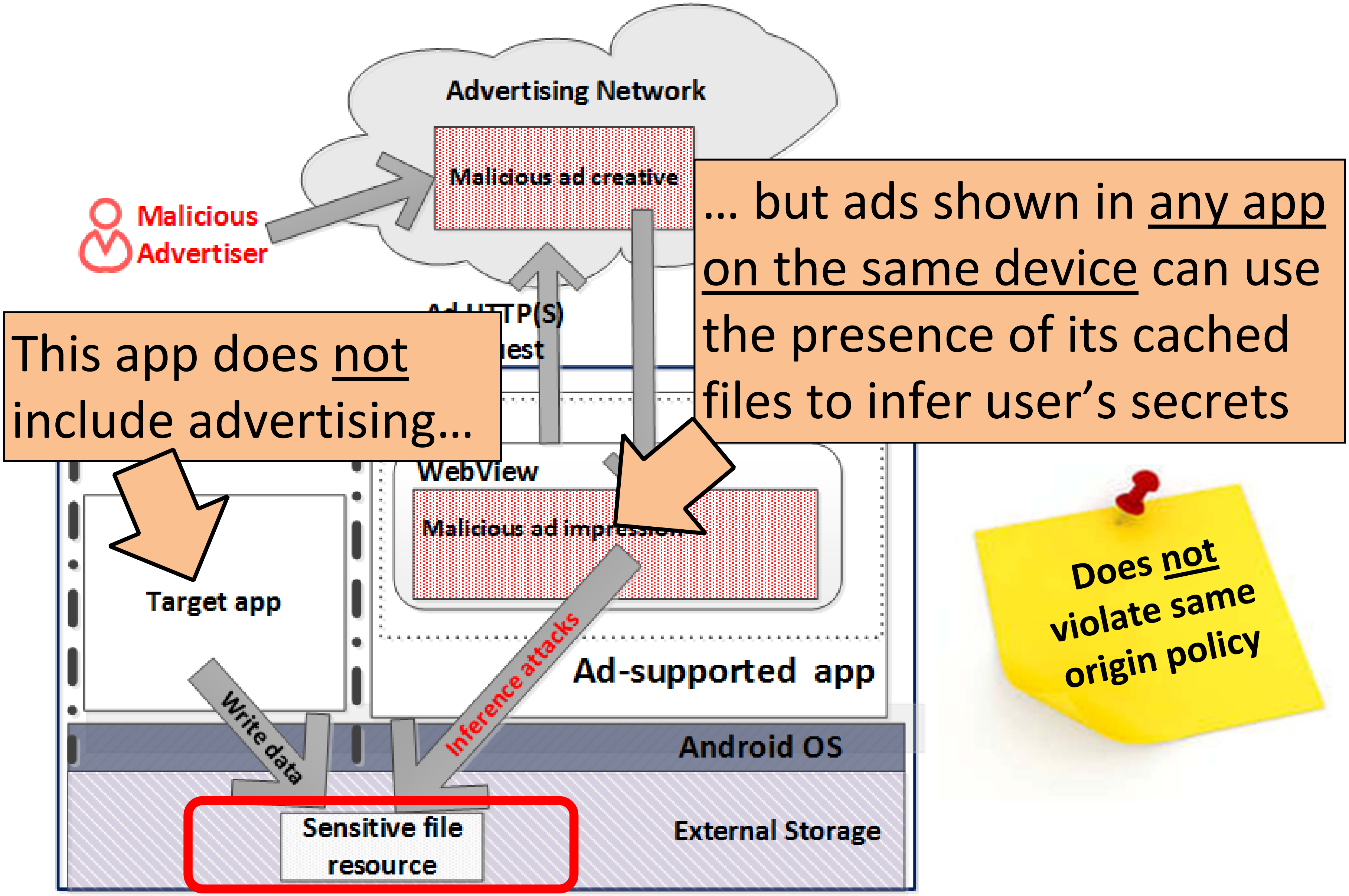
Thumbnail images of drugs that the user searched for cached in external storage



Any ad displayed in any other app on the same device can infer which drugs the user is taking



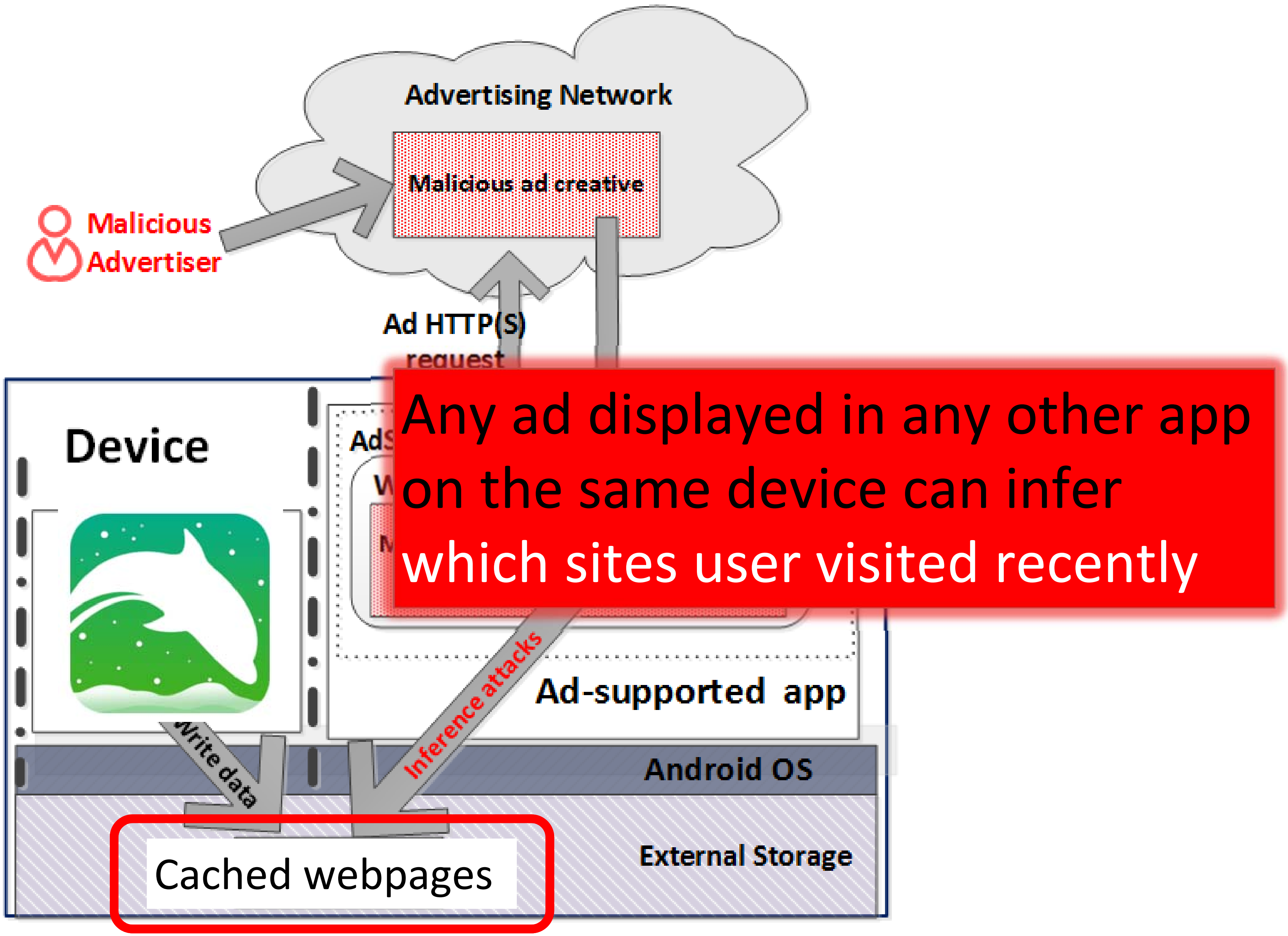
Does this file exist?
file://sdcard/Android/data/com.goodrx/cache/uil-images/45704837


Advertising Network
Malicious ad creative
Malicious Advertiser
... but ads shown in any app on the same device can use the presence of its cached files to infer user's secrets
This app does not include advertising...
WebView
Malicious ad impr
Target app
Ad-supported app
Inference attacks
Write data
Android OS
Sensitive file resource
External Storage
Does not violate same origin policy

Dolphin mobile browser

(50 to 100 million installs in Google Play Store)

To reduce bandwidth usage and response time, caches fetched images, HTML, and JavaScript in external storage



Advertising Network
Malicious ad creative
Malicious Advertiser
Ad HTTP(S) request
Device
Ad-supported app
Android OS
External Storage
Write data
Inference attacks
Cached webpages
Any ad displayed in any other app on the same device can infer which sites user visited recently

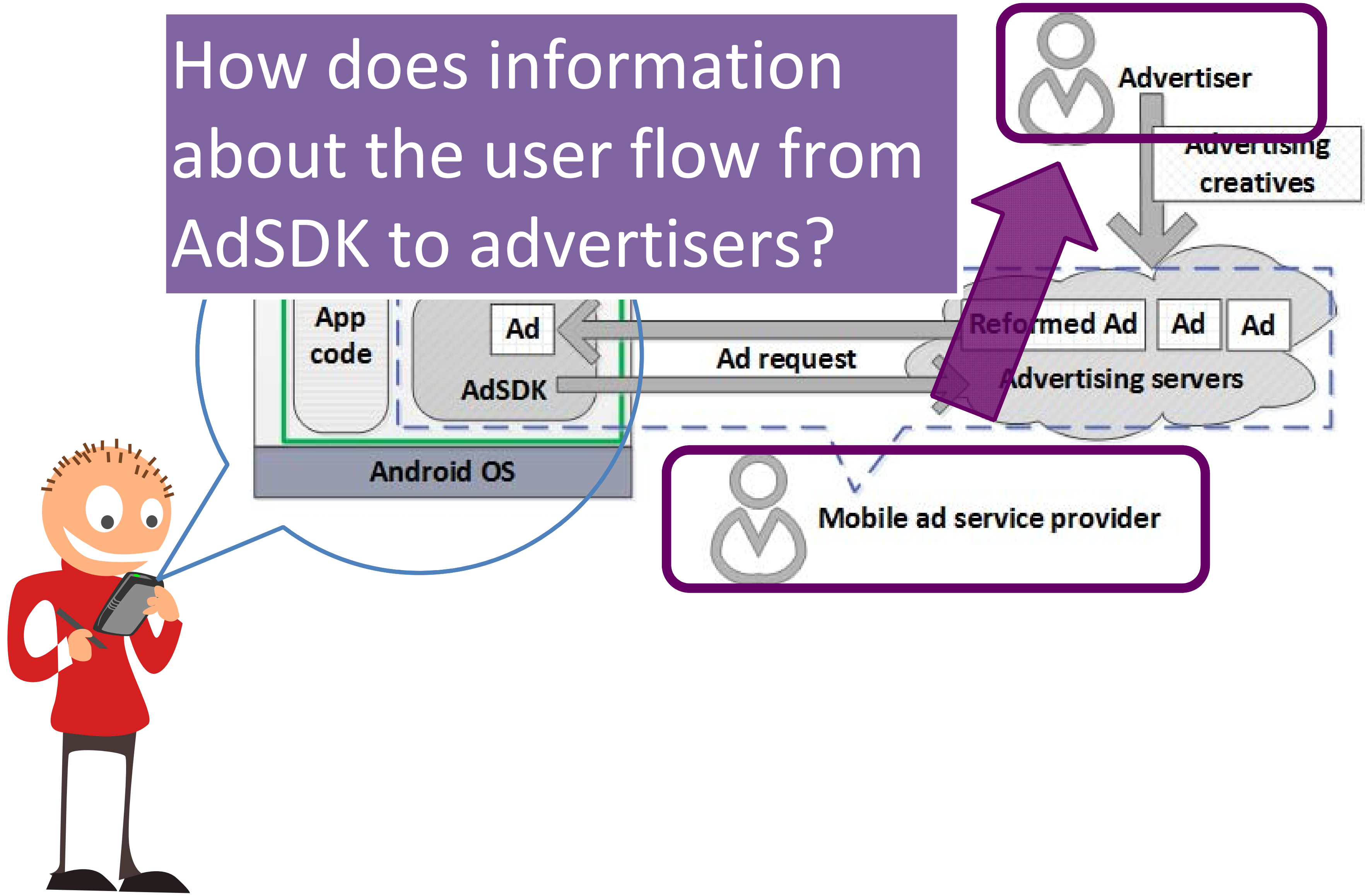
Our Study

- Several major Android advertising libraries

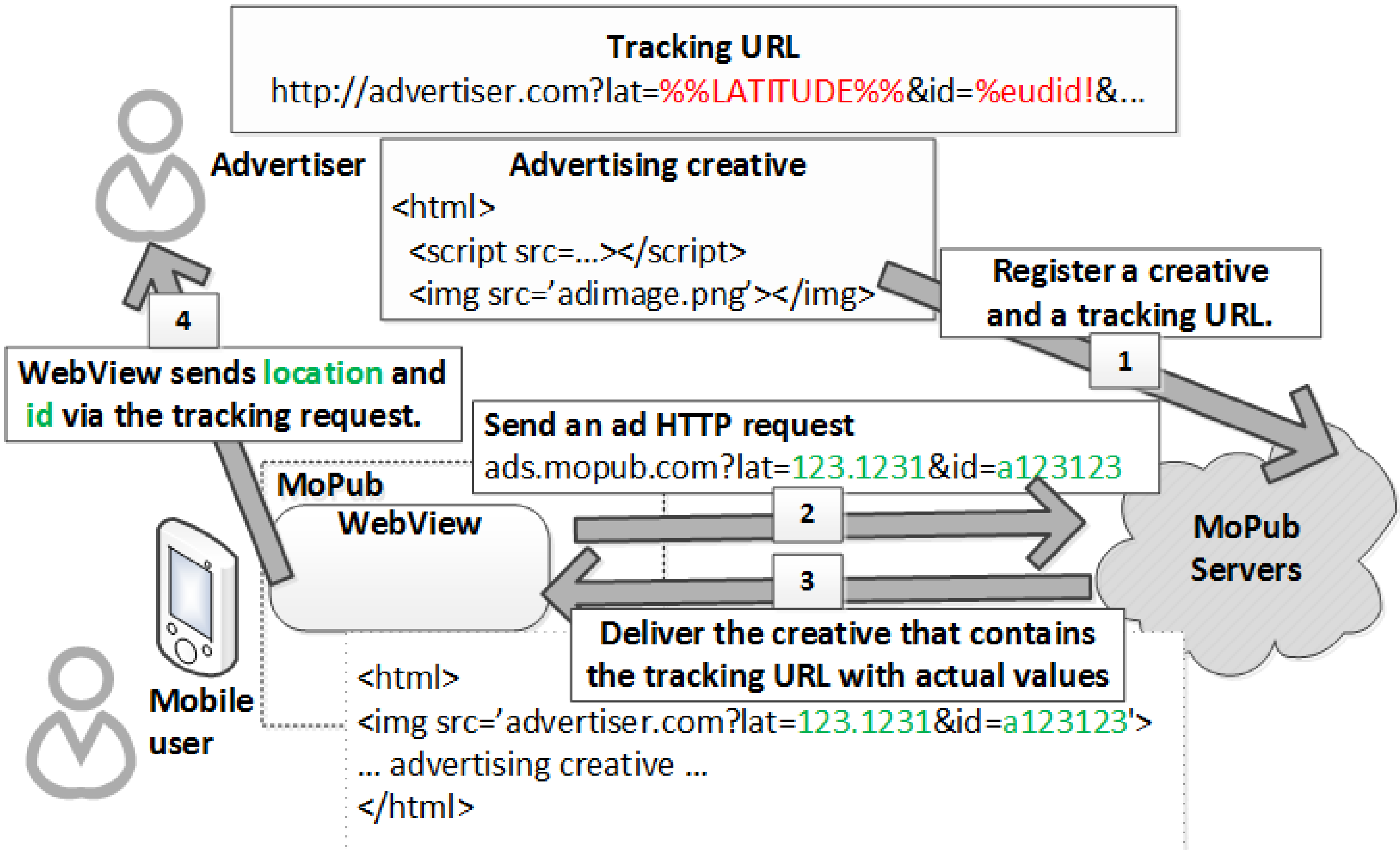
admob AdMarvel mopub airpush

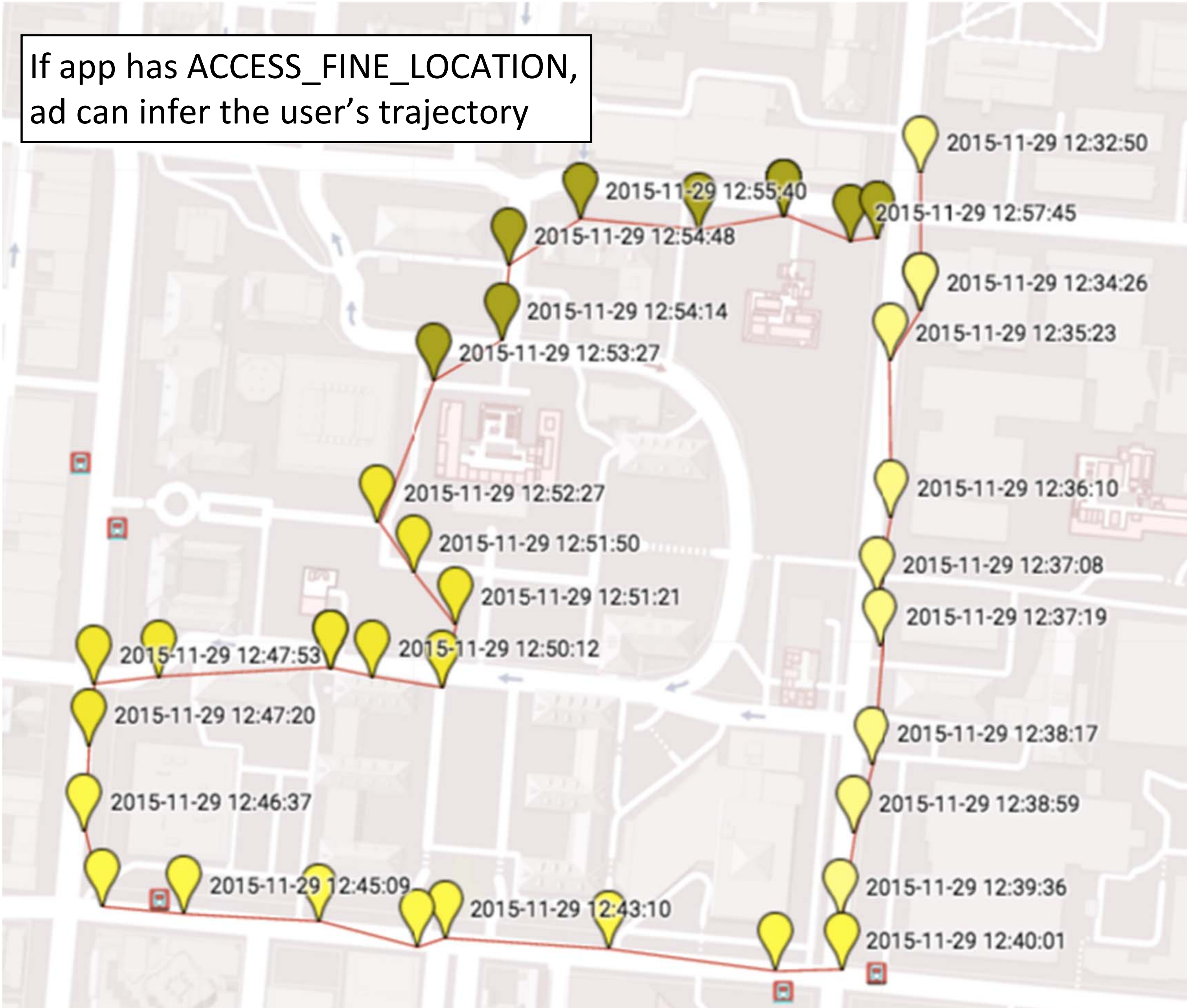
- "Local resource oracle" present in all of them
- All acknowledged the issue,
  several fixed in their latest AdSDK releases

How does information about the user flow from AdSDK to advertisers?

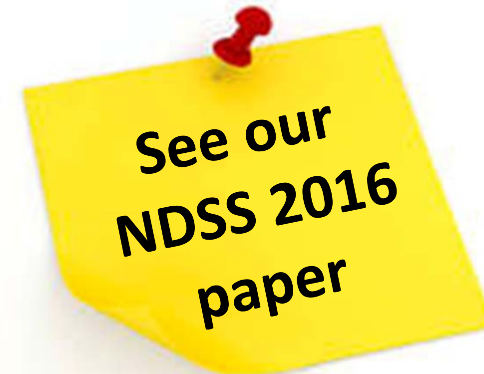


Flow of User's Location in MoPub




If app has ACCESS_FINE_LOCATION,
ad can infer the user's trajectory
2015-11-29 12:32:50
2015-11-29 12:55:40
2015-11-29 12:57:45
2015-11-29 12:54:48
2015-11-29 12:34:26
2015-11-29 12:54:14
2015-11-29 12:35:23
2015-11-29 12:53:27
2015-11-29 12:52:27
2015-11-29 12:36:10
2015-11-29 12:51:50
2015-11-29 12:37:08
2015-11-29 12:51:21
2015-11-29 12:37:19
2015-11-29 12:47:53
2015-11-29 12:50:12
2015-11-29 12:47:20
2015-11-29 12:38:17
2015-11-29 12:46:37
2015-11-29 12:38:59
2015-11-29 12:45:09
2015-11-29 12:43:10
2015-11-29 12:39:36
2015-11-29 12:40:01

Our Results



- First study of how Android advertising services protect users from malicious advertising
- Standard Web same origin policy is no longer secure in the mobile context
  - Mere existence of a certain file in external storage can reveal sensitive information about the user
  - Other security and privacy issues
- Proposed a defense; direct impact on the design of the mobile advertising software stack

Florian Schaub

Carnegie Mellon University

Towards Usable Privacy Policies: Semi-automatically Extracting Data Practices From Websites' Privacy Policies

Co-authors: Norman Sadeh, Alessandro Acquisti, Travis D. Breaux, Lorrie Faith Cranor, Noah A. Smith, Fei Liu, Shomir Wilson, James T. Graves, Pedro Giovanni Leon, Rohan Ramanath, Ashwini Rao (Carnegie Mellon University); Aleecia M. McDonald (Stanford University); Joel Reidenberg, N. Cameron Russell (Fordham University)

PRIVACY CON

FEDERAL TRADE COMMISSION

Towards Usable Privacy Policies

Semi-automatically Extracting
Data Practices from Privacy Policies

Florian Schaub
Carnegie Mellon University

Norman Sadeh | Lead Principal Investigator
Carnegie Mellon University

wwww.usableprivacy.org

USABLE PRIVACY.ORG
the usable privacy policy project

A NSF SaTC
Frontier project
(CNS-1330596)

Norman Sadeh, Alessandro Acquisti, Travis Breaux, Lorrie Cranor, Noah Smith

Jaspreet Bhatia, Aswarth Dara, Harishma Dayanidhi, James Graves, Bin Liu, Fei Liu, Alessandro Oltramari, Mads Schaarup Andersen, Florian Schaub, Shomir Wilson, Rohan Ramanath, Ashwini Rao, Kanthashree Sathyendra

Carnegie Mellon University

Joel Reidenberg

N. Cameron Russell, Thomas B. Norton, Antoine Bon, Samuel Borenzweig, Alexander Callen, Timothy Carter, Elle Davis, Amanda Grannis, Sophia Qasir, Stephanie Tallering

CLIP Center on Law and Information Policy AT FORDHAM LAW SCHOOL

Aleecia McDonald, Barbara van Schewick

Pedro Giovanni Leon, Margaret Hagan

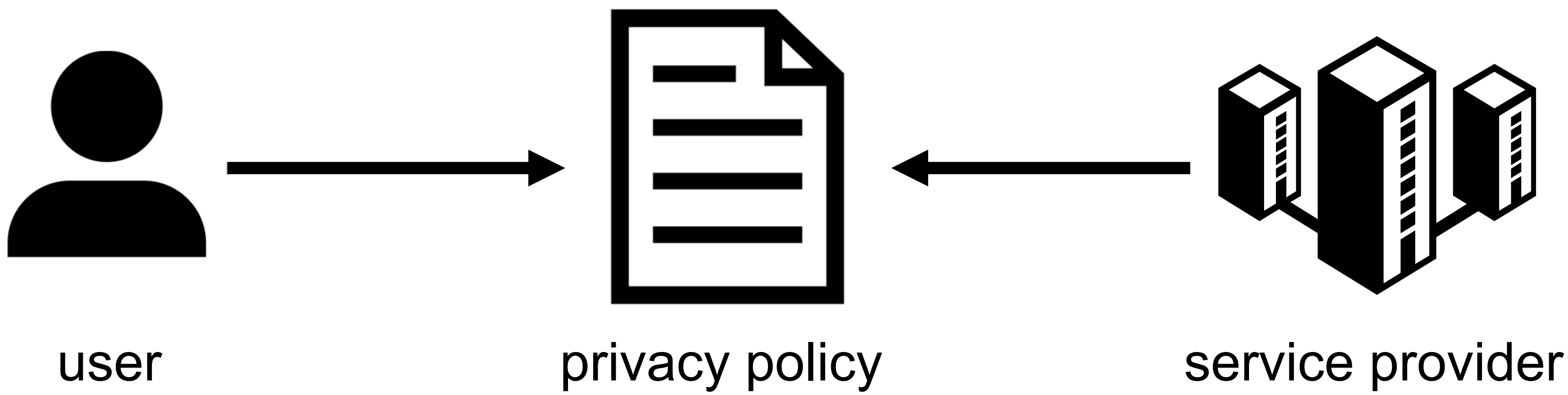
CIS The Center for Internet and Society

USABLE PRIVACY.ORG the usable privacy policy project

NSF

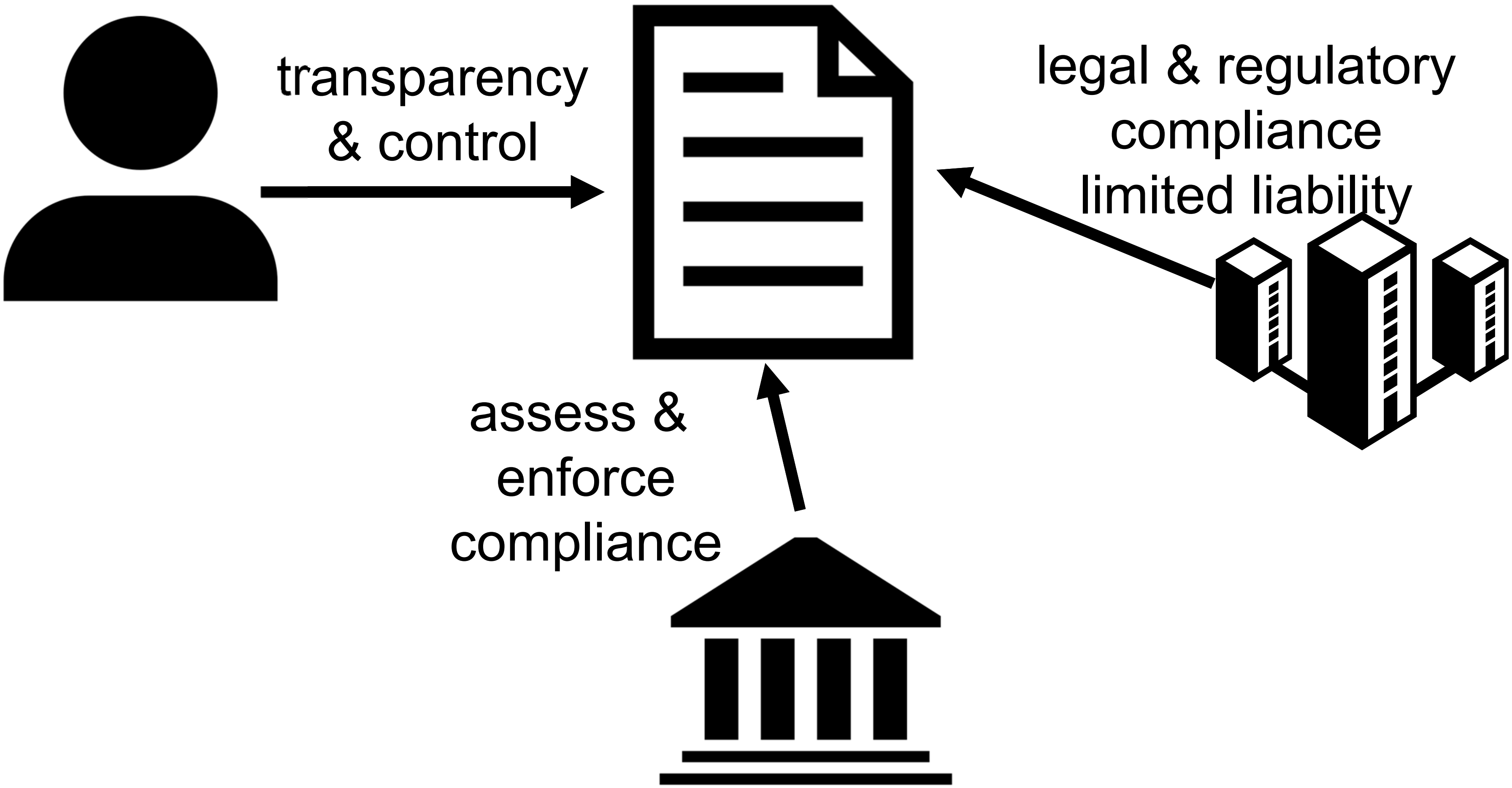
A NSF SaTC Frontier project (CNS-1330596)

Privacy notice & choice



USABLE PRIVACY.ORG

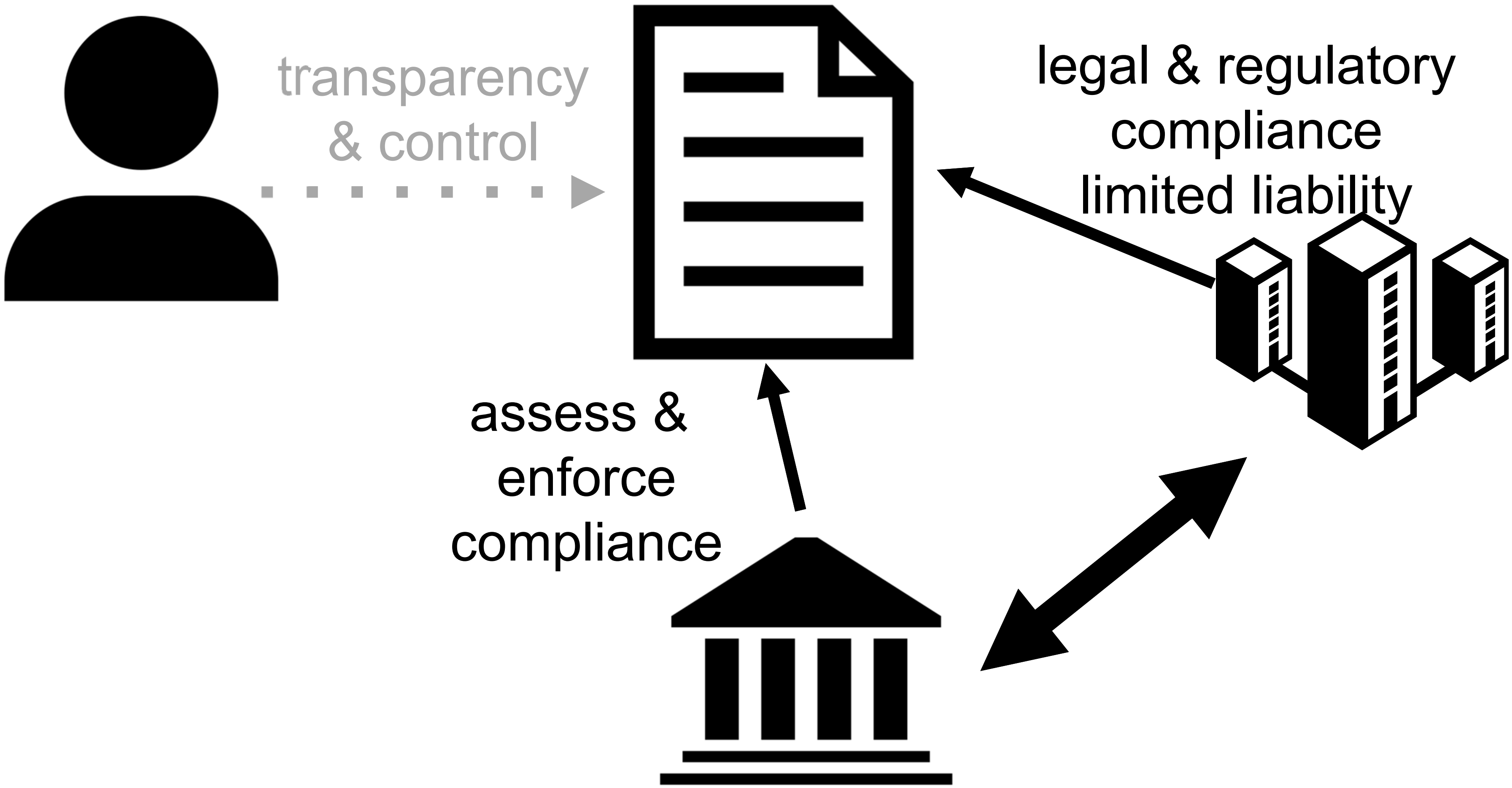
Privacy notice & choice



Schaub et al., *A Design Space for Effective Privacy Notices.* SOUPS'15: Symposium on Usable Privacy and Security, June 2015.

USABLE PRIVACY.ORG

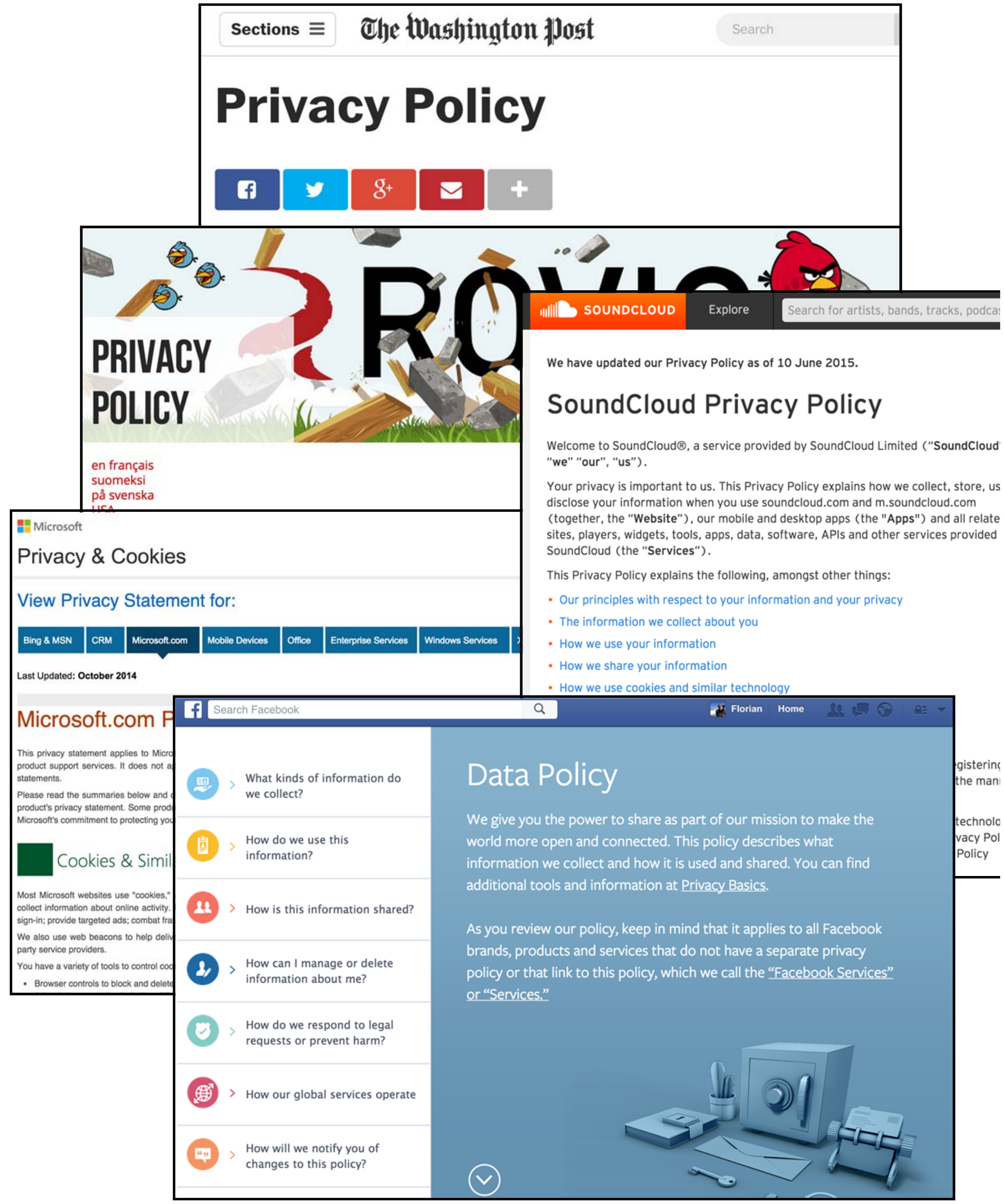
Privacy notice & choice



Schaub et al., *A Design Space for Effective Privacy Notices.* SOUPS'15: Symposium on Usable Privacy and Security, June 2015.

Privacy policies

- long & complex
- difficult to understand
- jargon & vagueness
- lack of choices

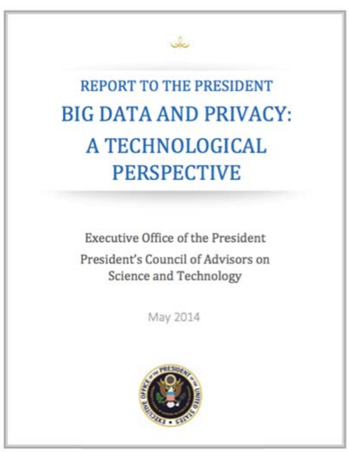


Schaub et al., *A Design Space for Effective Privacy Notices.* SOUPS'15: Symposium on Usable Privacy and Security, June 2015.

USABLE PRIVACY.ORG

Privacy policies

- long & complex
- difficult to understand
- jargon & vagueness
- lack of choices


REPORT TO THE PRESIDENT
BIG DATA AND PRIVACY: A TECHNOLOGICAL PERSPECTIVE

Executive Office of the President
President's Council of Advisors on Science and Technology

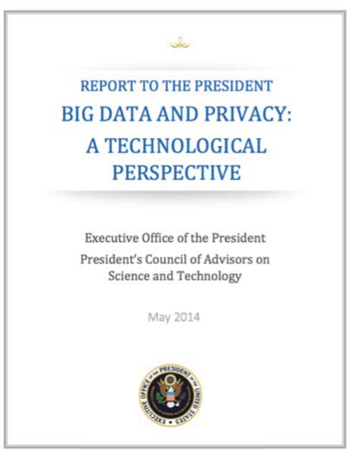
May 2014

"**Only in some fantasy world** do users actually **read these notices** and **understand their implications** before clicking to indicate their consent"

USABLE PRIVACY.ORG

Privacy policies

- long & complex
- difficult to understand
- jargon & vagueness
- lack of choices


REPORT TO THE PRESIDENT
BIG DATA AND PRIVACY: A TECHNOLOGICAL PERSPECTIVE

Executive Office of the President
President's Council of Advisors on Science and Technology

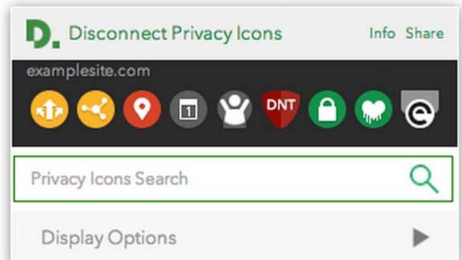
May 2014

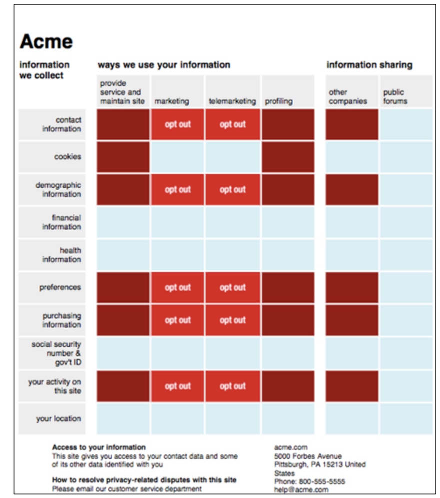
"**Only in some fantasy world** do users actually **read these notices** and **understand their implications** before clicking to indicate their consent"

USABLE PRIVACY.ORG

Overcoming the status quo

- Layered privacy notices
- Privacy nutrition labels
- Privacy icons
- Machine-readable policies (e.g. P3P or Do Not Track)
- …





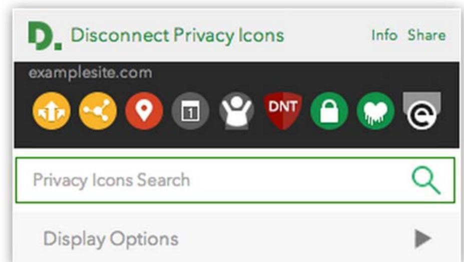
The Platform for Privacy Preferences 1.1 (P3P1.1) Specification

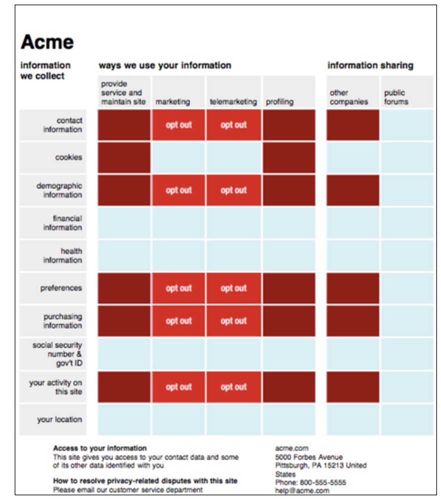
W3C Working Group Note 13 November 2006

```
<POLICY xmlns="http://www.w3.org/2000/P3Pv1"
        entity="TheCoolCatalog, 123 Main Street, Seattle, WA 98103, USA">
  <DISPUTES-GROUP>
    <DISPUTES service="http://www.PrivacySeal.org"
     resolution-type="independent"
     description="PrivacySeal, a third-party seal provider"
     image="http://www.PrivacySeal.org/Logo.gif"/>
    </DISPUTES-GROUP>
  <DISCLOSURE discuri="http://www.CoolCatalog.com/Practices.html" access="none"/>
  <STATEMENT>
    <CONSEQUENCE-GROUP>
      <CONSEQUENCE>a site with clothes you would  appreciate</CONSEQUENCE>
     </CONSEQUENCE-GROUP>
    <RECIPIENT><ours/></RECIPIENT>
     <RETENTION><indefinitely/></RETENTION>
    <PURPOSE><custom/><develop/></PURPOSE>
    <DATA-GROUP>
      <DATA name="dynamic.cookies" category="state"/>
      <DATA name="dynamic.miscdata" category="preference"/>
      <DATA name="user.gender"/>
      <DATA name="user.home." optional="yes"/>
    </DATA-GROUP>
  </STATEMENT>
  <STATEMENT>
    <RECIPIENT><ours/></RECIPIENT>
    <PURPOSE><admin/><develop/></PURPOSE>
     <RETENTION><indefinitely/></RETENTION>
    <DATA-GROUP>
      <DATA name="dynamic.clickstream.server"/>
      <DATA name="dynamic.http.useragent"/>
    </DATA-GROUP>
  </STATEMENT>
</POLICY>
```

Overcoming the status quo

- Layered privacy notices
- Privacy nutrition labels
- Privacy icons
- Machine-readable policies (e.g. P3P or Do Not Track)
- …





Acme

information we collect	ways we use your information: provide service and maintain site	marketing	telemarketing	profiling	information sharing: other companies	public forums
contact information		opt out	opt out			
cookies						
demographic information		opt out	opt out			
financial information						
health information						
preferences		opt out	opt out			
purchasing information		opt out	opt out			
social security number & gov't ID						
your activity on this site		opt out	opt out			
your location						

Access to your information
This site gives you access to your contact data and some of its other data identified with you

How to resolve privacy-related disputes with this site
Please email our customer service department

acme.com
5000 Forbes Avenue
Pittsburgh, PA 15213 United States
Phone: 800-555-5555
help@acme.com

The Platform for Privacy Preferences 1.1 (P3P1.1) Specification

W3C Working Group Note 13 November 2006

```
<POLICY xmlns="http://www.w3.org/2000/P3Pv1"
        entity="TheCoolCatalog, 123 Main Street, Seattle, WA 98103, USA">
  <DISPUTES-GROUP>
     <DISPUTES service="http://www.PrivacySeal.org"
      resolution-type="independent"
      description="PrivacySeal, a third-party seal provider"
      image="http://www.PrivacySeal.org/Logo.gif"/>
    </DISPUTES-GROUP>
  <DISCLOSURE discuri="http://www.CoolCatalog.com/Practices.html" access="none"/>
  <STATEMENT>
    <CONSEQUENCE-GROUP>
      <CONSEQUENCE>a site with clothes you would  appreciate</CONSEQUENCE>
    </CONSEQUENCE-GROUP>
    <RECIPIENT><ours/></RECIPIENT>
    <RETENTION><indefinitely/></RETENTION>
    <PURPOSE><custom/><develop/></PURPOSE>
    <DATA-GROUP>
      <DATA name="dynamic.cookies" category="state"/>
      <DATA name="dynamic.miscdata" category="preference"/>
      <DATA name="user.gender"/>
      <DATA name="user.home." optional="yes"/>
    </DATA-GROUP>
  </STATEMENT>
  <STATEMENT>
    <RECIPIENT><ours/></RECIPIENT>
    <PURPOSE><admin/><develop/></PURPOSE>
    <RETENTION><indefinitely/></RETENTION>
    <DATA-GROUP>
      <DATA name="dynamic.clickstream.server"/>
      <DATA name="dynamic.http.useragent"/>
    </DATA-GROUP>
  </STATEMENT>
</POLICY>
```

Lack of industry support & adoption incentives

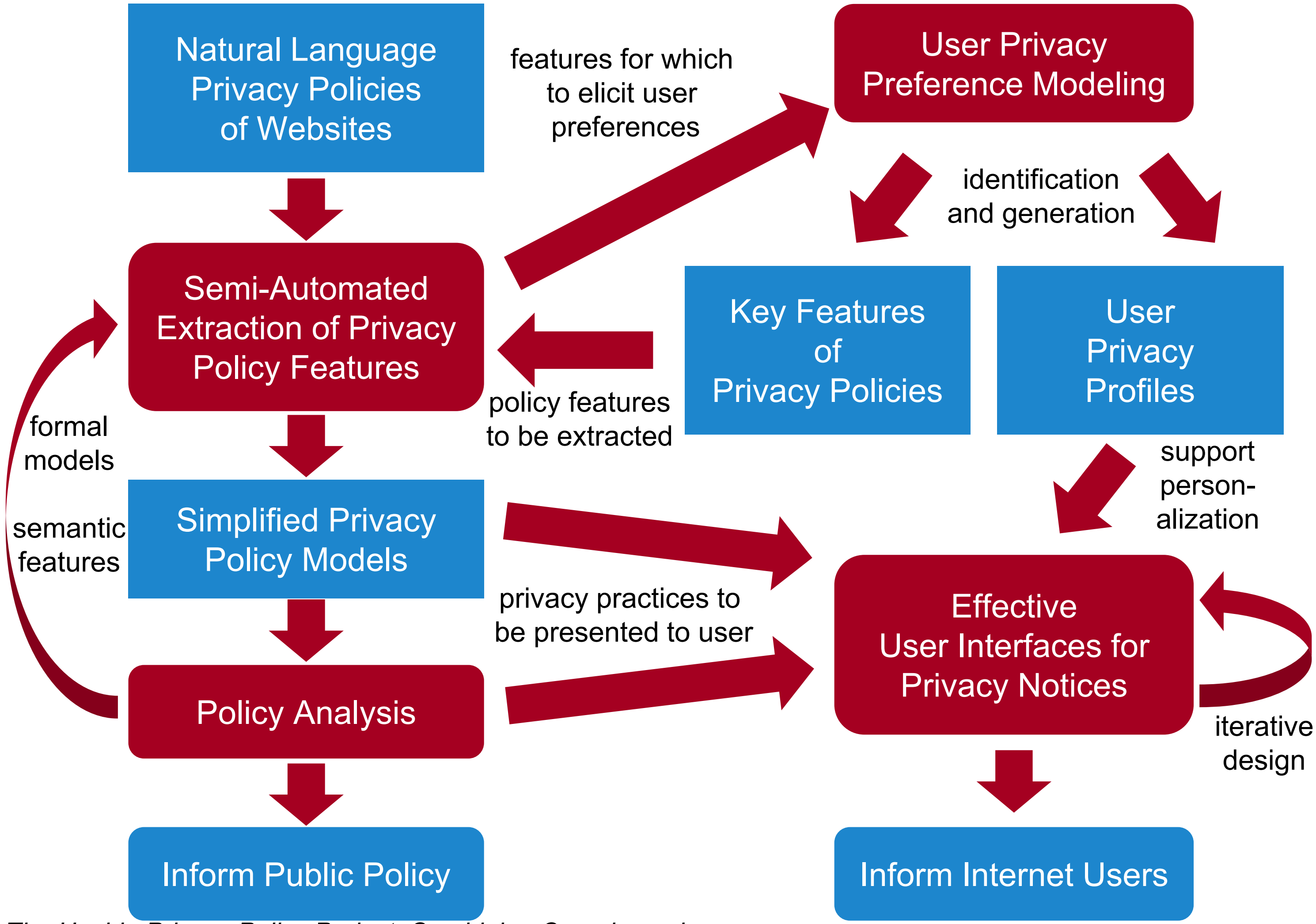
Project objectives

- **Semi-automatically analyze** natural language **privacy policies** to **extract key data practices**
- Combine **crowdsourcing, machine learning natural language processing** to enable **large-scale analysis** of privacy policies
- **Model users' privacy preferences** to focus on those practices they care about
- **Develop effective user interfaces** that convey relevant and actionable information to users

Sadeh et al., *The Usable Privacy Policy Project: Combining Crowdsourcing, Machine Learning and Natural Language Processing to Semi-Automatically Answer Those Privacy Questions Users Care About*, CMU Tech Report, 2013.

USABLE PRIVACY.ORG

Tightly interconnected threads



Sadeh et al., *The Usable Privacy Policy Project: Combining Crowdsourcing, Machine Learning and Natural Language Processing to Semi-Automatically Answer Those Privacy Questions Users Care About*, CMU Tech Report, 2013.

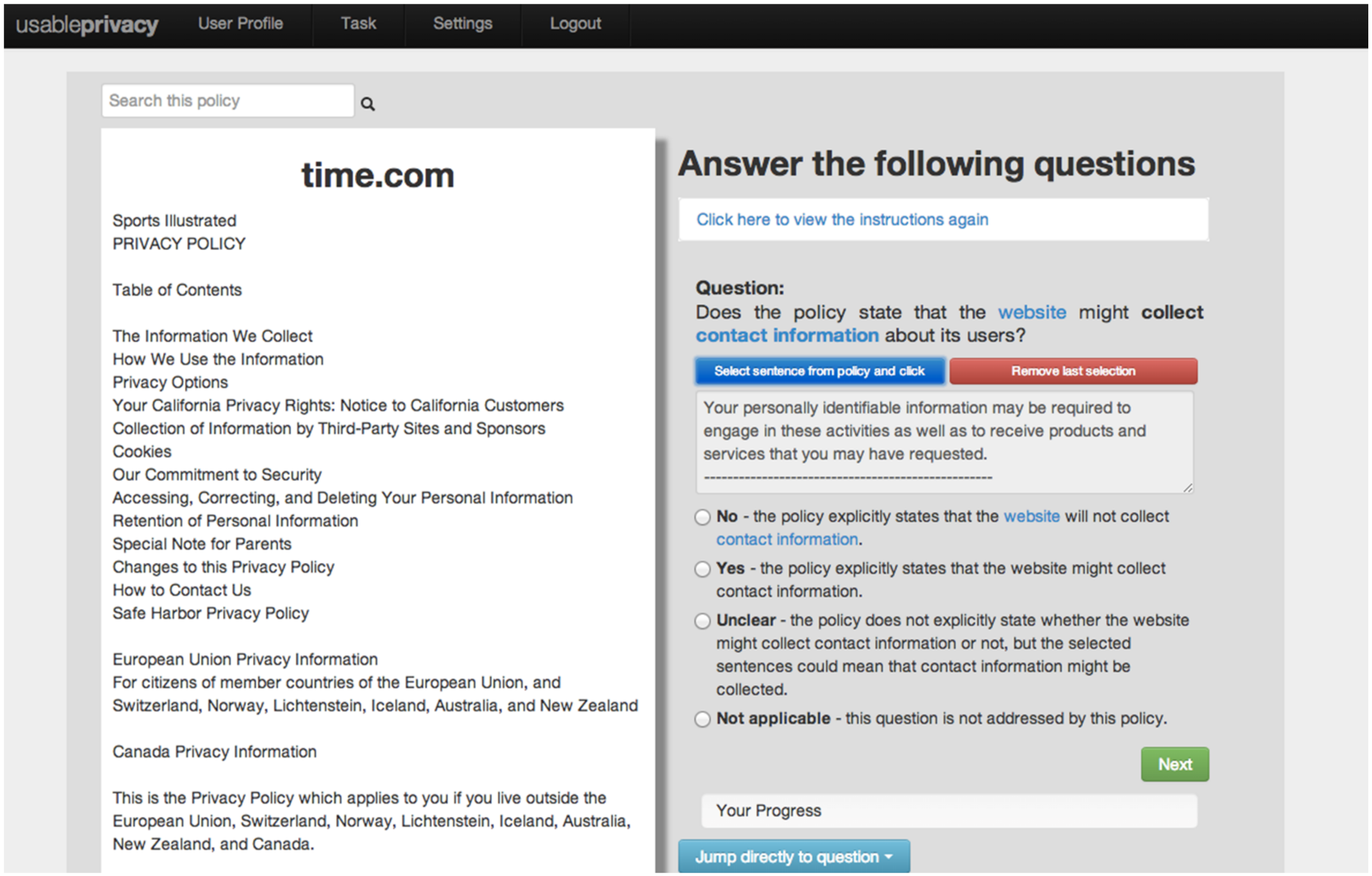
USABLEPRIVACY.ORG

Identifying data practices of interest

- **Legal analysis**
  - Analysis of privacy harms addressed through litigation
- **User modeling**
  - Studies on privacy preferences & concerns
- **Policy content analysis**
  - Analysis of how practices are described in privacy policies
  - Ambiguity and vagueness in privacy policies

Reidenberg et al., Privacy *Harms and the Effectiveness of the Notice and Choice Framework.* I/S Journal of Law & Policy for the Information Society. vol. 11, 2015.
Wilson et al., *Crowdsourcing Annotations for Websites' Privacy Policies: Can It Really work?* WWW'16: Intl. Worldwide Web Conference, April 2016.

Crowdsourcing policy annotations



Wilson et al., *Crowdsourcing Annotations for Websites' Privacy Policies: Can It Really work?* WWW'16: Intl. Worldwide Web Conference, April 2016.

USABLE PRIVACY.ORG

Crowdsourcing policy annotations

collection of contact information

2x **Yes:** The policy explicitly states that the website might collect contact information

6x **Unclear:** The policy does not explicitly state whether the website might collect contact information or not

The Information We Collect

At some Turner Network sites, you can order products, enter contests, vote in polls or otherwise express an opinion, subscribe to one of our services such as our online newsletters, or participate in one of our online forums or communities. In the course of these various offerings, we often seek to collect from you various forms of personal information. Examples of the types of personally identifiable information that may be collected at these pages include: name, address, e-mail address, telephone number, fax number, credit card information, and information about your interests in and use of various products, programs, and services.

At some Turner Network sites, you may also be able to submit information about other people. For example, you might submit a person's name and e-mail address to send an electronic greeting

How good are crowdworkers?

- **Studies to compare performance of**
  - privacy policy experts
  - grad students in law & public policy
  - MTurk crowdworkers

- **Annotation of 26 policies**
  - 26 policies annotated by crowdworkers & skilled annotators
  - 6 policies also annotated by experts

Reidenberg et al., *Disagreeable Privacy Policies: Mismatches between Meaning and Users' Understanding*. Berkeley Technology Law Journal, vol. 30, 1, pp.39-88, May 2015
Wilson et al., *Crowdsourcing Annotations for Websites' Privacy Policies: Can It Really work?* WWW'16: Intl. Worldwide Web Conference, April 2016.

USABLEPRIVACY.ORG

How good are crowdworkers?

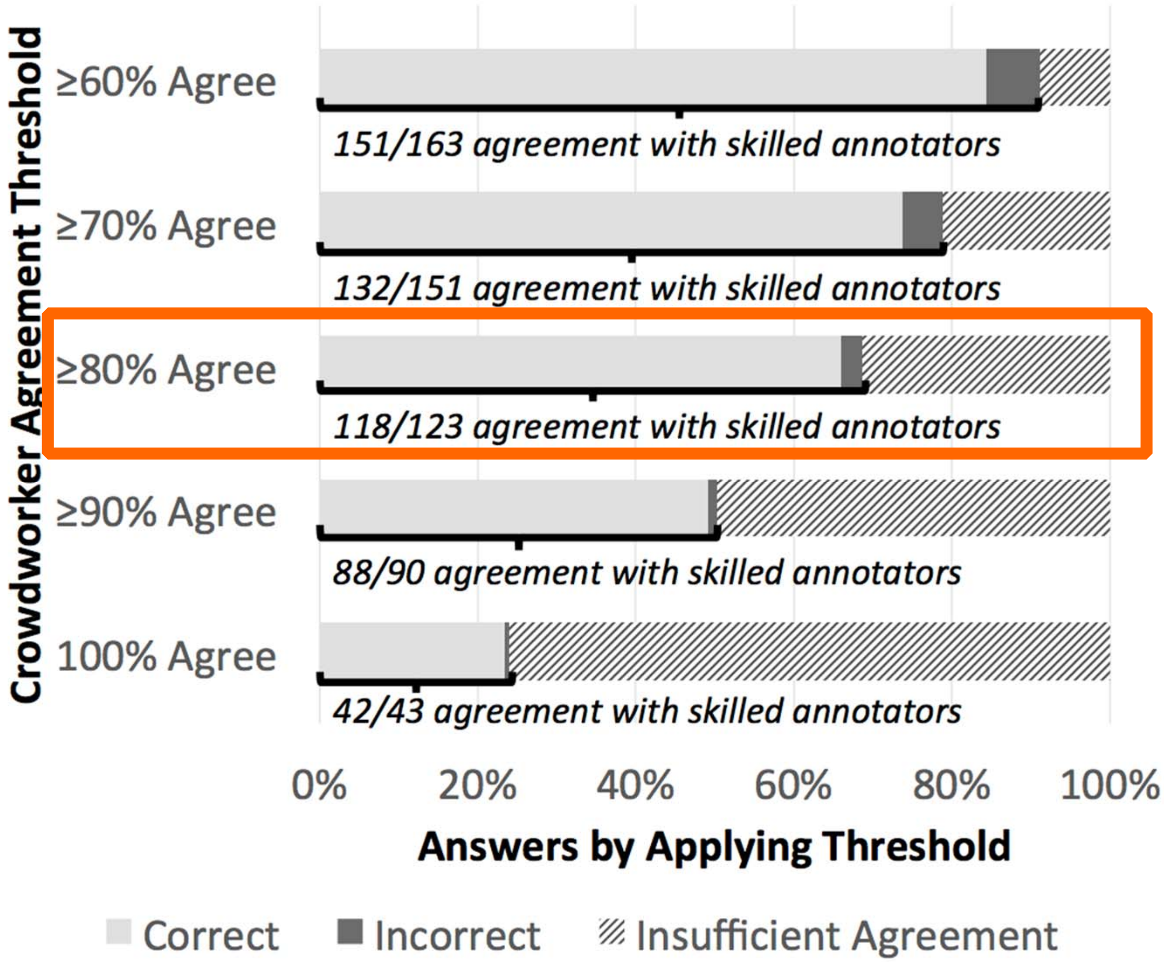
- **Results highlights**
  - Even experts do not always agree
  - Data collection relatively easy to identify
  - Data sharing practices more difficult
  - Finer nuances difficult to extract

Reidenberg et al., *Disagreeable Privacy Policies: Mismatches between Meaning and Users' Understanding*. Berkeley Technology Law Journal, vol. 30, 1, pp.39-88, May 2015

USABLE PRIVACY.ORG

Accuracy of crowdworker annotations

Compared to skilled annotators on 26 policies



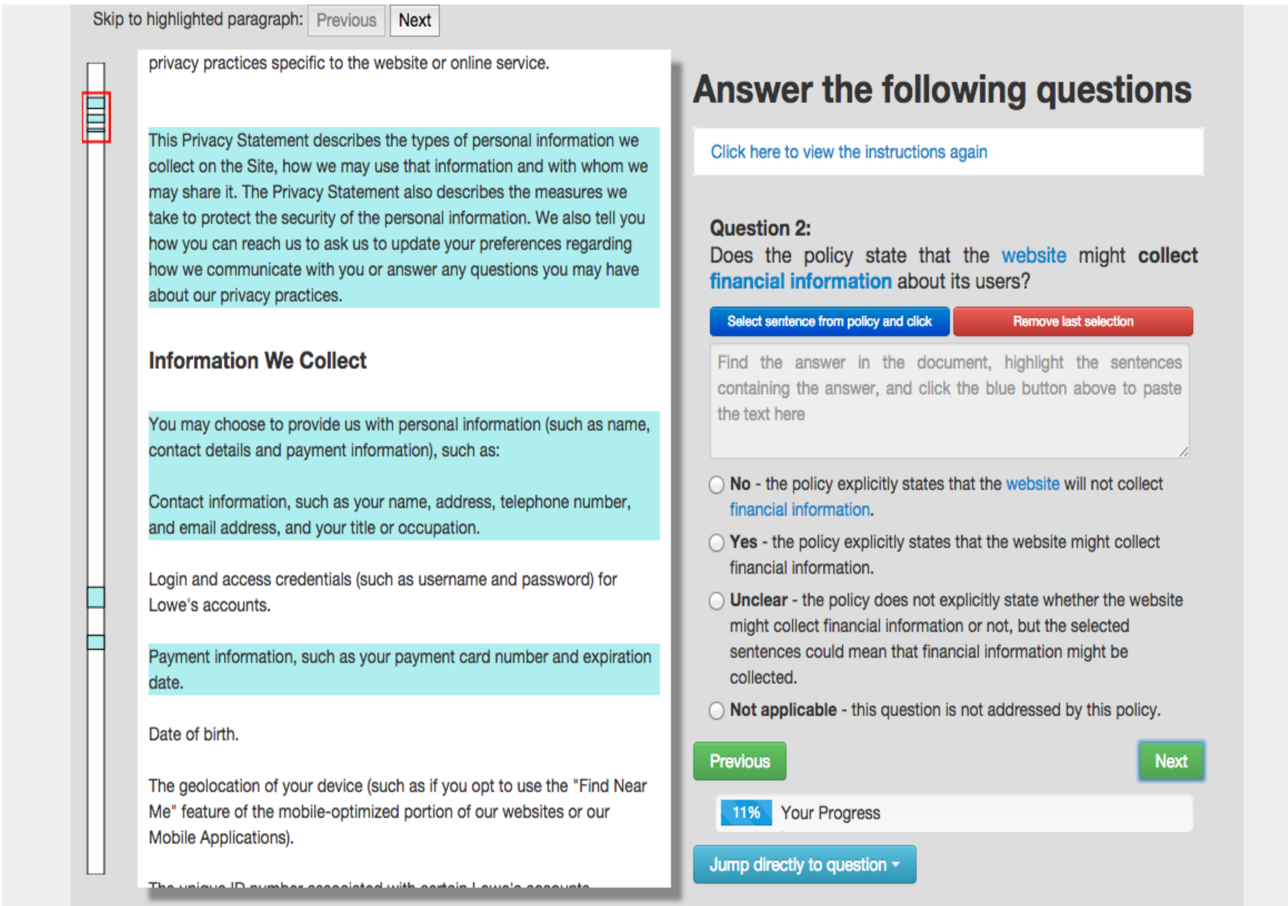
Wilson et al., *Crowdsourcing Annotations for Websites' Privacy Policies: Can It Really work?* WWW'16: Intl. Worldwide Web Conference, April 2016.

USABLE**PRIVACY**.ORG

Enhancing extraction tasks with Machine Learning and NLP

- Accurate crowdsourcing of policy annotations is feasible
- But privacy policies are still long and complex
- Goal: Help crowdworkers **read selectively** (thus working more rapidly) without loss of accuracy

Predicting & highlighting relevant paragraphs



Logistic regression based relevance models

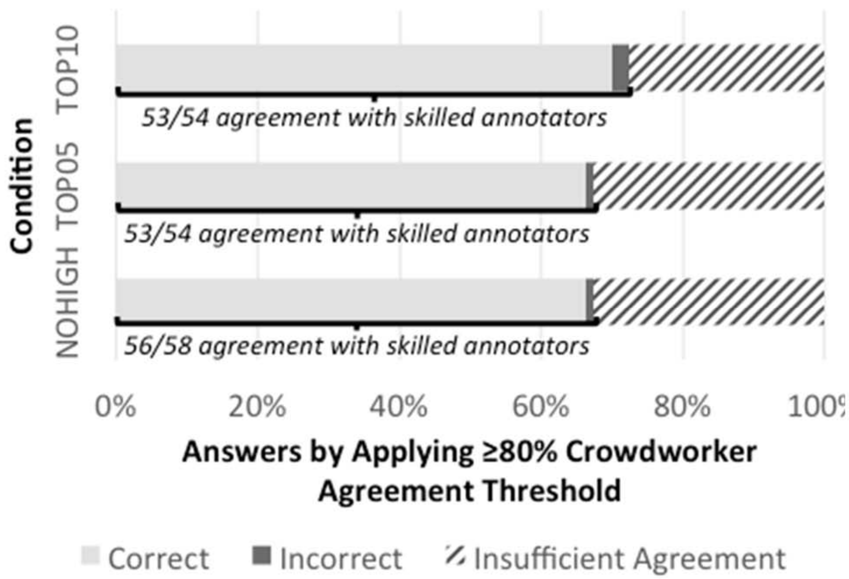
Highlight X paragraphs most relevant for current question

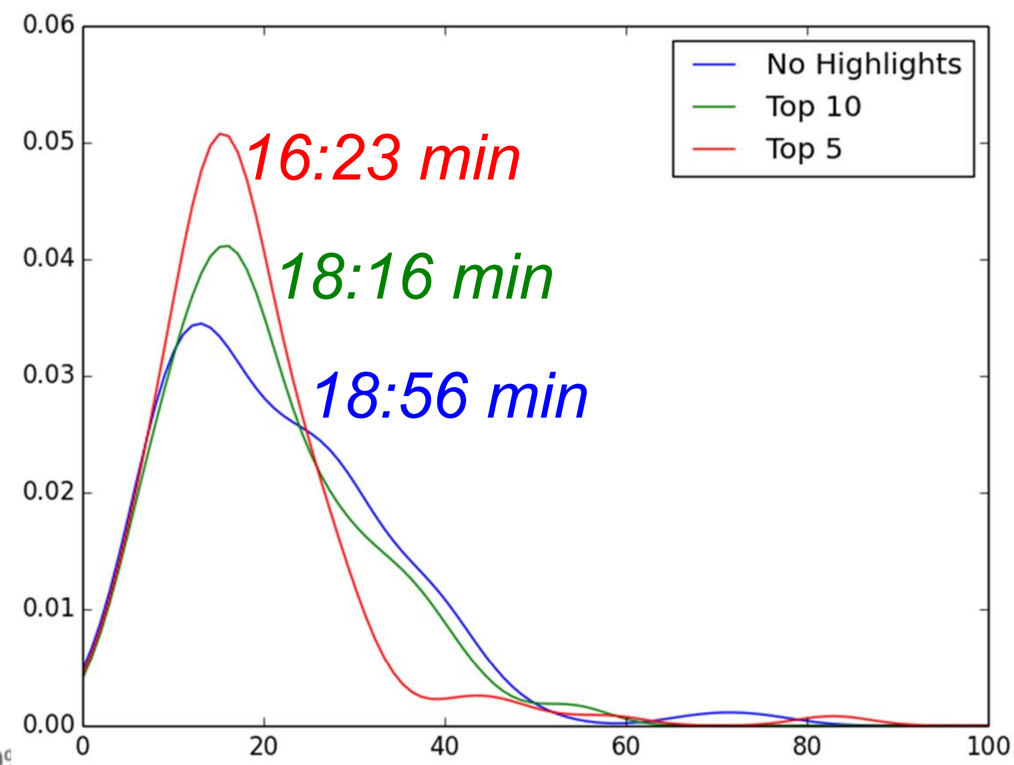
Wilson et al., *Crowdsourcing Annotations for Websites' Privacy Policies: Can It Really work?* WWW'16: Intl. Worldwide Web Conference, April 2016.

USABLEPRIVACY.ORG

Predicting & highlighting relevant paragraphs

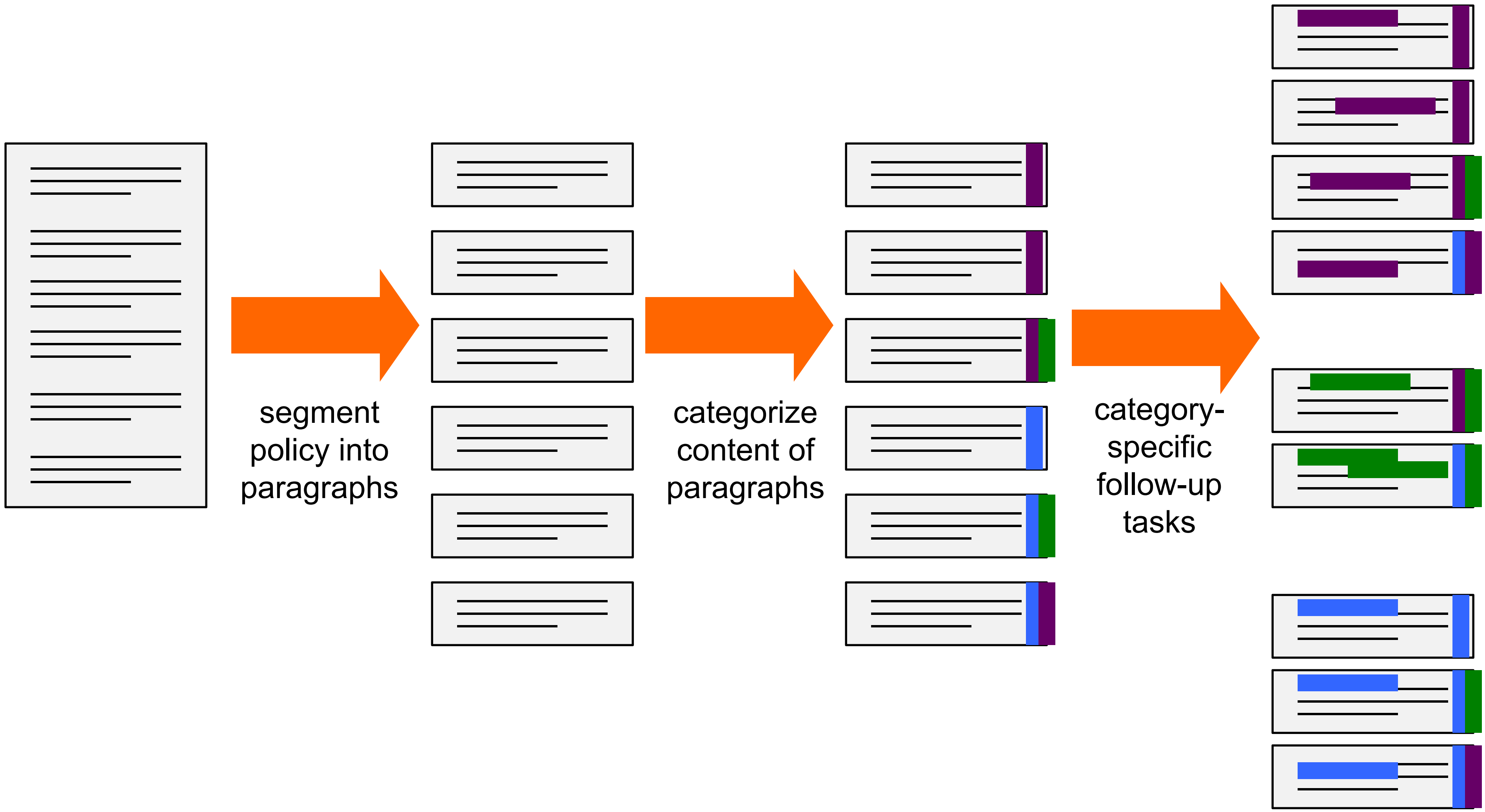
Crowdworkers can be induced to label privacy policies **faster without affecting accuracy.**





Wilson et al., *Crowdsourcing Annotations for Websites' Privacy Policies: Can It Really work?* WWW'16: Intl. Worldwide Web Conference, April 2016.

Multi-step annotation workflow



Breaux & Schaub, *Scaling Requirements Extraction to the Crowd: Experiments with Privacy Policies.* RE'14: Intl. Requirements Engineering Conf. 2014.

Simplified but fine-grained tasks

Response options for categorization

Select relevant words and press button

Click here to read the expanded instructions with an example.

Short Instructions: Select the action verbs with your mouse cursor and then press one of the following keys to indicate when the verb describes an act to:

- Press 'c' for collect - any act by Zynga to collect information from another party, including the user
- Press 'u' for use - any act by Zynga or another party to use or modify information for a particular purpose
- Press 't' for transfer - any act by Zynga to transfer or share information with another party, including the user
- Press 'r' for retain - any act by Zynga to retain, store or delete information

In the following paragraph, any pronouns "We" or "Us" refer to the game company Zynga, and "you" refers to the Zynga user.

Paragraph:

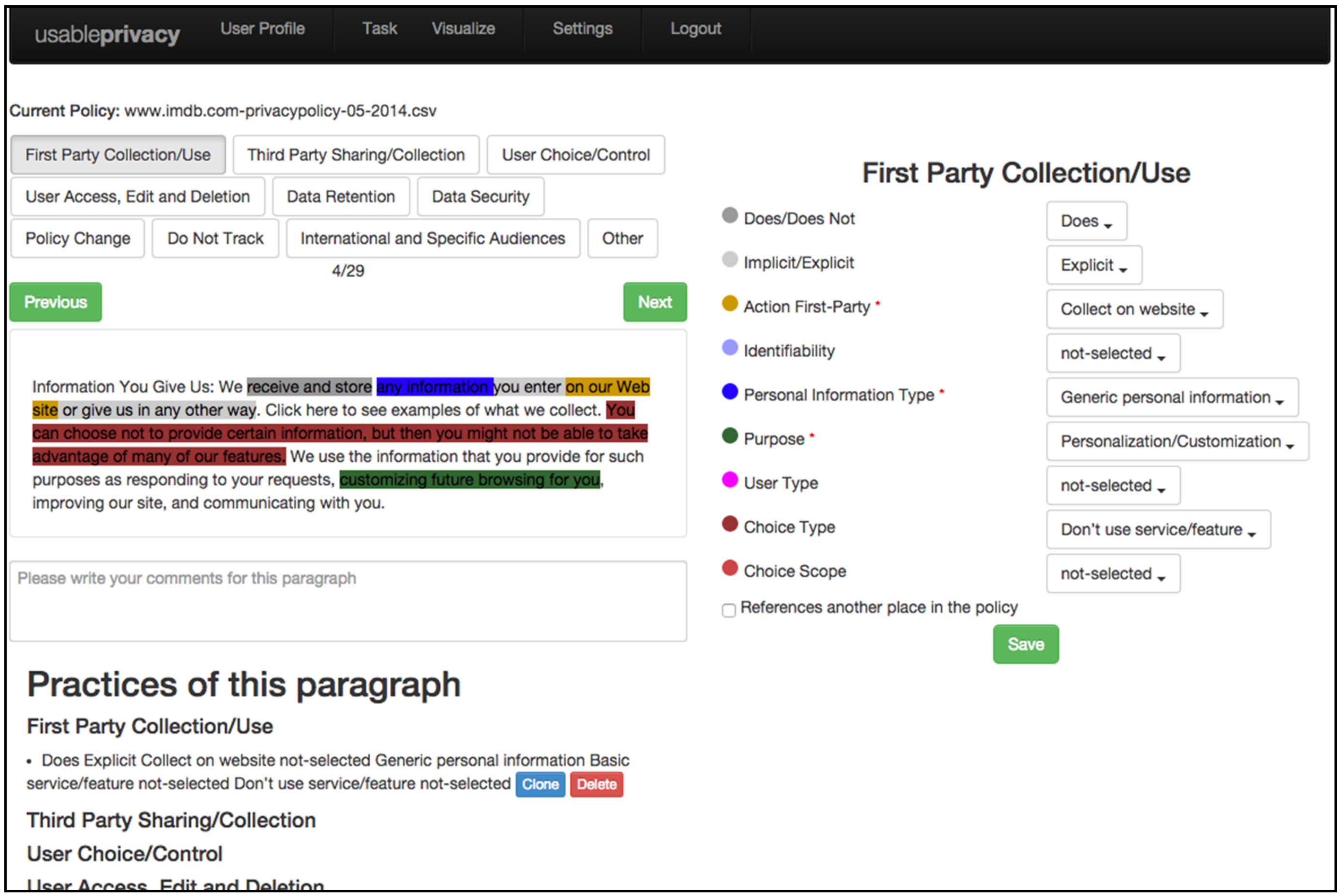
We may collect or receive information from other sources including (i) other Zynga users who choose to upload their email contacts; and (ii) third party information providers.

Submit Query

Clear Last | Clear All

Breaux & Schaub, *Scaling Requirements Extraction to the Crowd: Experiments with Privacy Policies.* RE'14: Intl. Requirements Engineering Conf. 2014.

Annotation dataset

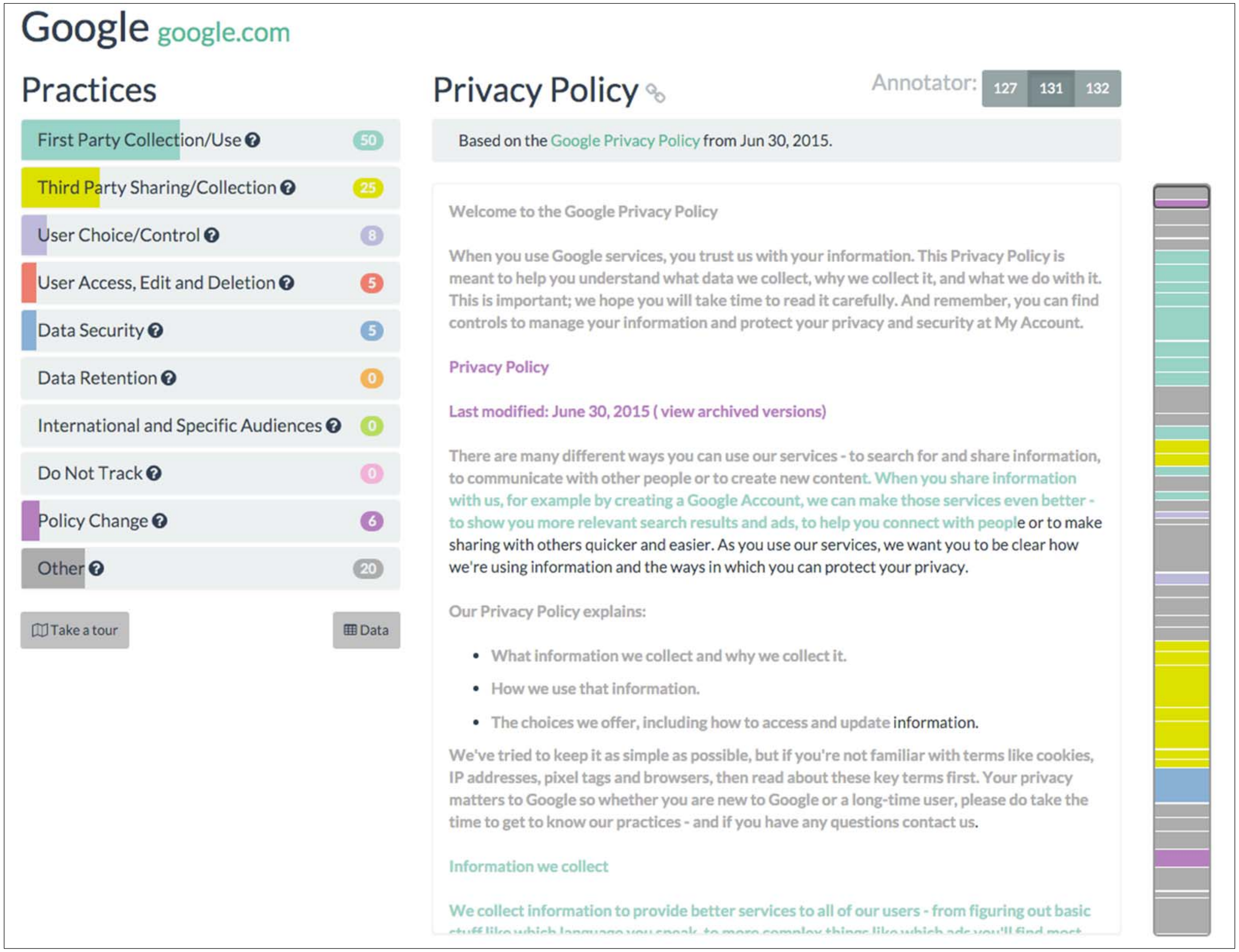


creating corpus of >100 privacy policies annotated by law students

gold standard data for ML/NLP research

USABLE PRIVACY.ORG

Annotation dataset



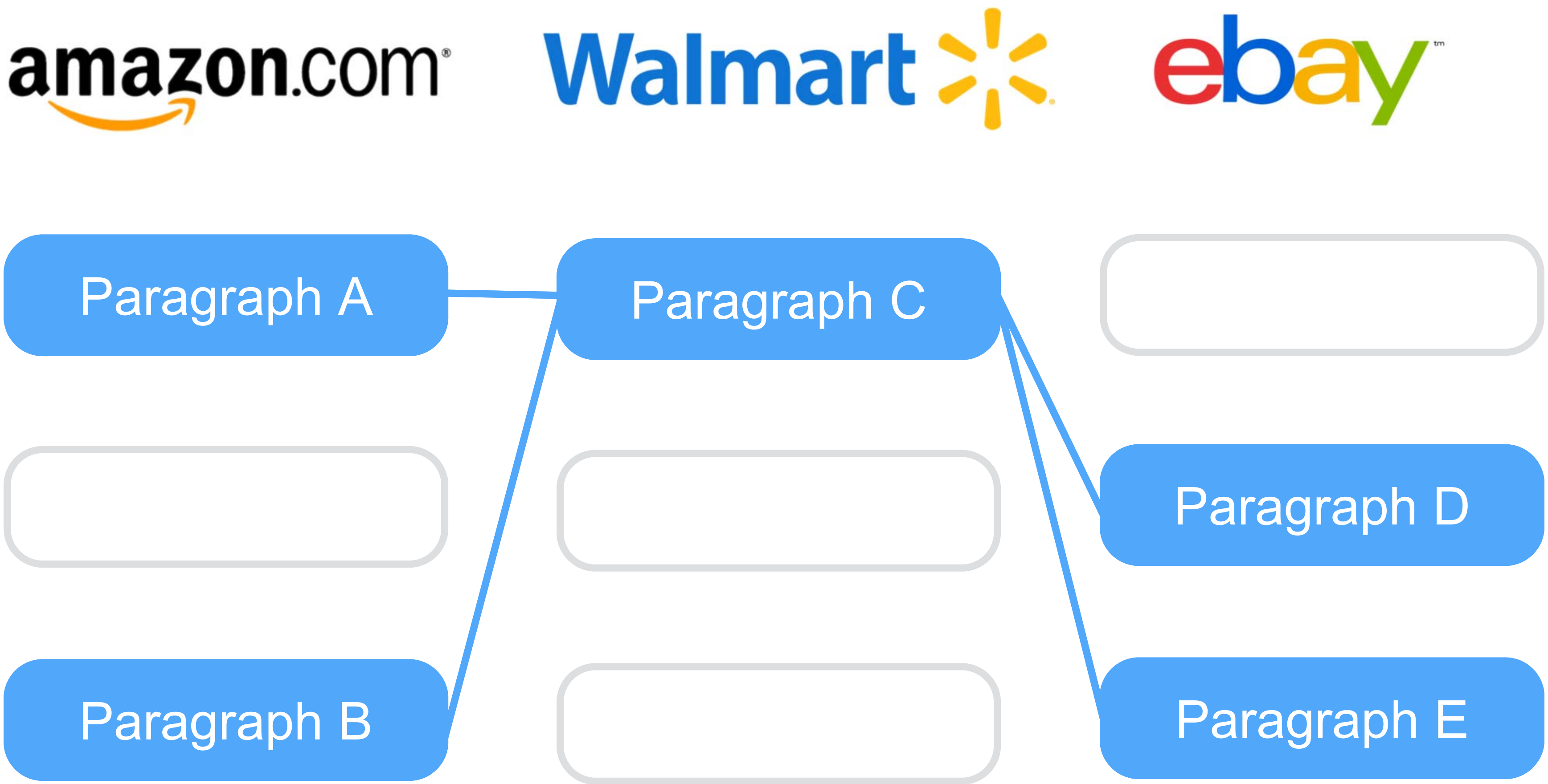
creating corpus of >100 privacy policies annotated by law students

gold standard data for ML/NLP research

USABLE PRIVACY.ORG

Towards automated extraction

Paragraph sequence alignment



Liu et al., *A Step Towards Usable Privacy Policy: Automatic Alignment of Privacy Statements.* COLING'14, 2014.

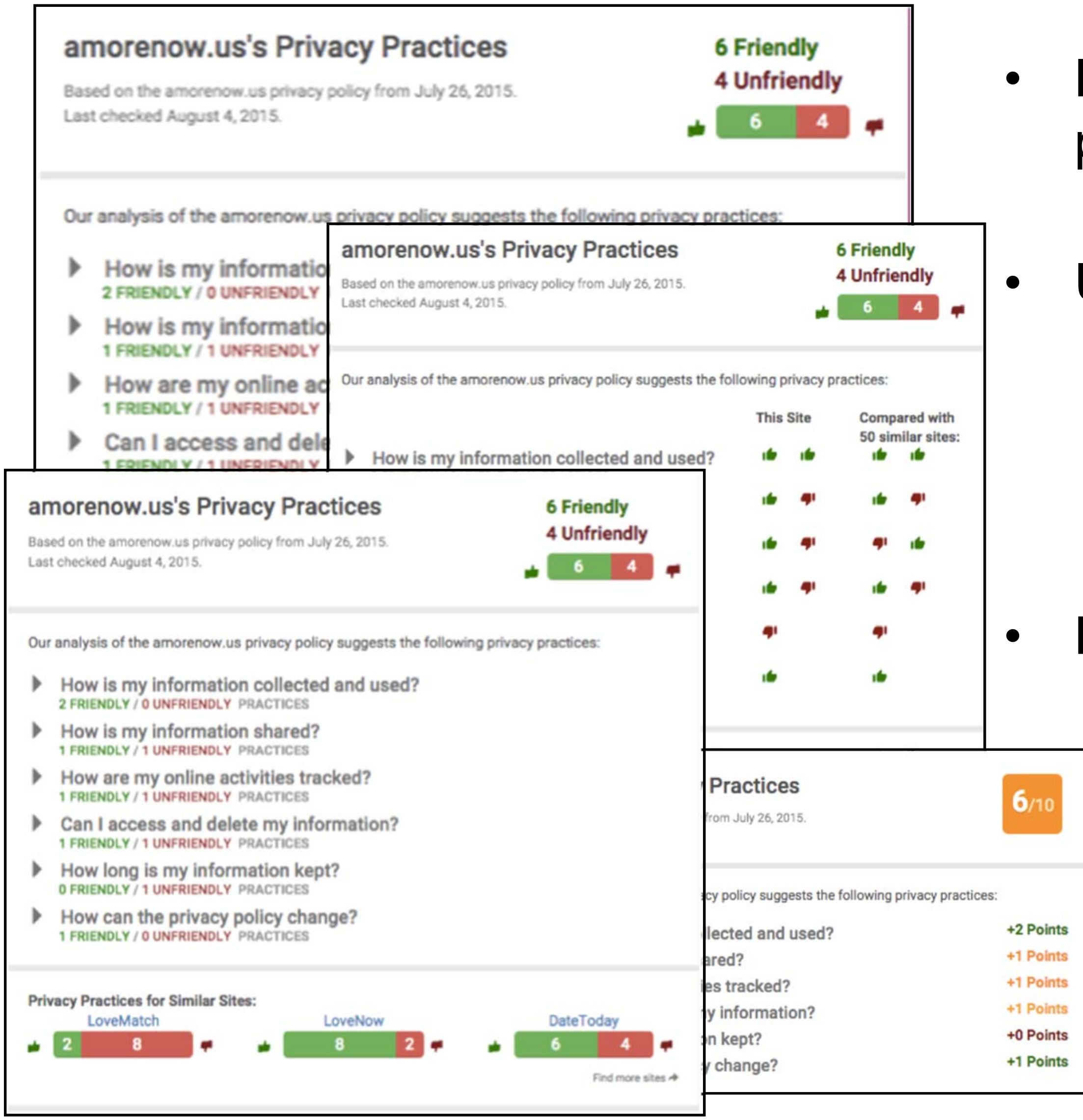
Providing notice to users

- **Relevant information**
  - highlight practices users care about
  - emphasize unexpected practices
  - usable and intuitive interface

- **Actionable information**
  - show available privacy choices
  - help users find privacy-friendly alternatives
  - enable users to express dislike of practices

- Development of **Privacy Browser Plugin**
  - provide information independent of website

Browser plugin design



- Display limited set of relevant practices
- User-centered iterative design
  - Focus groups
  - Online studies
  - Field studies
- Public release: Summer 2016

USABLEPRIVACY.ORG

Conclusions

- **Semi-automatic analysis** of **privacy policies** with crowdsourcing, natural language processing and machine learning
- Enable **large-scale analysis** of privacy policies
- **Modeling users' privacy preferences** to identify unexpected and relevant practices
- **Development of effective user interfaces** that convey relevant and actionable information to users

Florian Schaub
fschaub@cmu.edu

Norman Sadeh
Lead PI | sadeh@cmu.edu

USABLE PRIVACY.ORG

Norman Sadeh

Carnegie Mellon University

To Deny, or Not to Deny: A Personalized Privacy Assistant for Mobile App Permissions

Co-authors: Bin Liu, Mads Schaarup Andersen, Florian Schaub, Hazim Almuhimedi, Yuvraj Agarwal, Alessandro Acquisti (Carnegie Mellon University)

PRIVACY CON
FEDERAL TRADE COMMISSION

Personalized Privacy Assistants

From Android Apps to the Internet of Things

Norman Sadeh

Professor, School of Computer Science

Co-Director, MSIT Program in Privacy Engineering

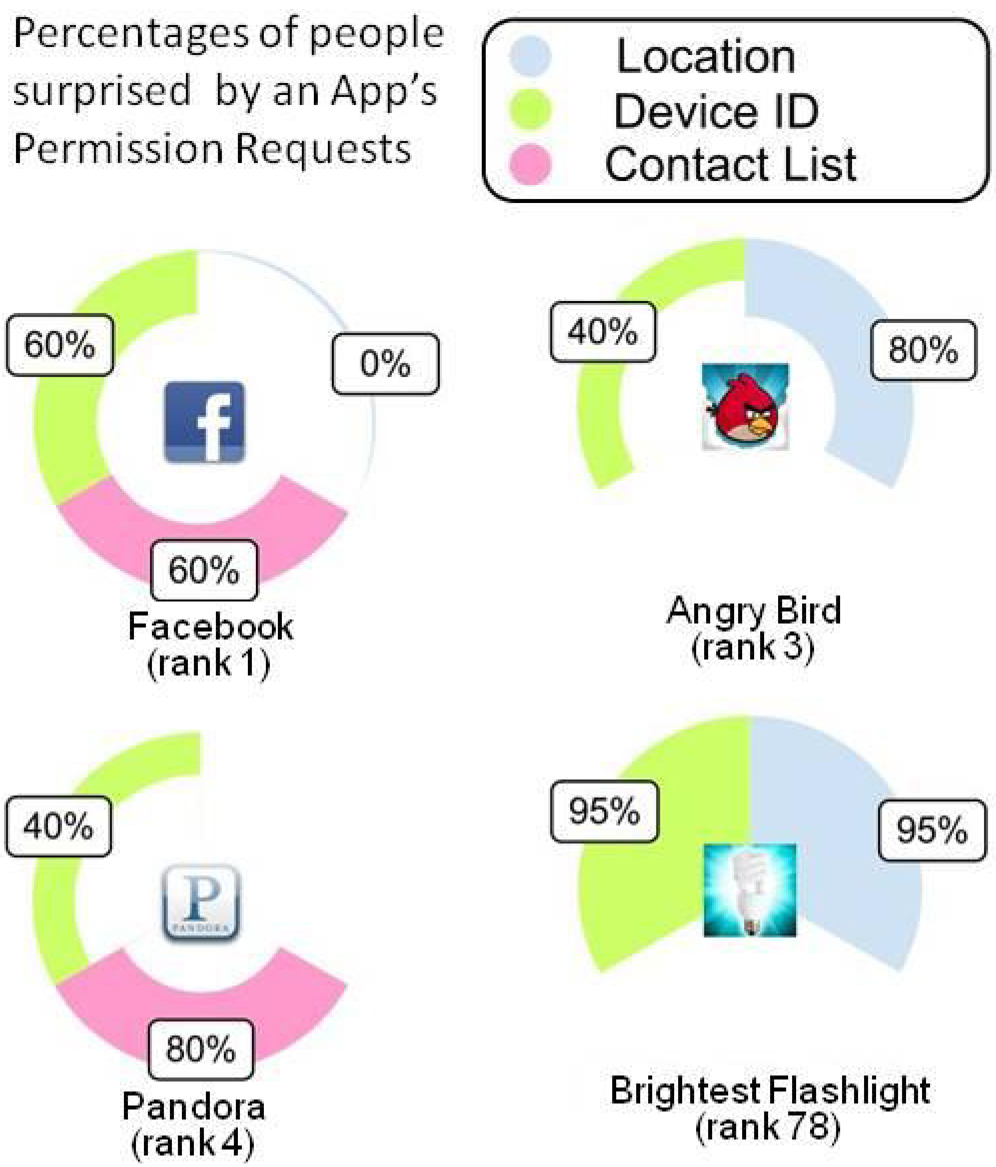
Carnegie Mellon University

www.normsadeh.org --- sadeh AT cs.cmu.edu

Carnegie Mellon University Master of Science in Privacy ENGINEERING

mobile commerce lab

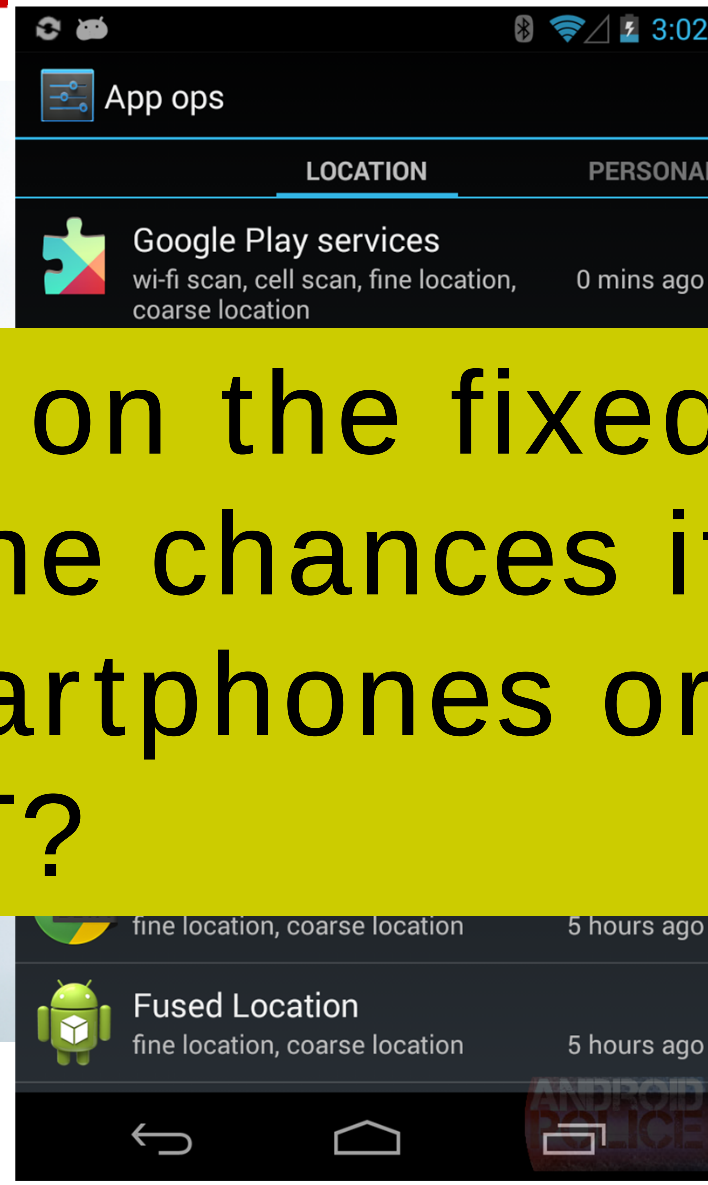
People Care About Privacy…



J. Lin, S. Amini, J. Hong, N. Sadeh, J. Lindqvist, J. Zhang, "Expectation and Purpose: Understanding Users' Mental Models of Mobile App Privacy through Crowdsourcing", Proc. of the 14th ACM International Conference on Ubiquitous Computing, Pittsburgh, USA, Sept. 2012

...But They Are Feeling Helpless...

- *Privacy policies are too long and too complex*

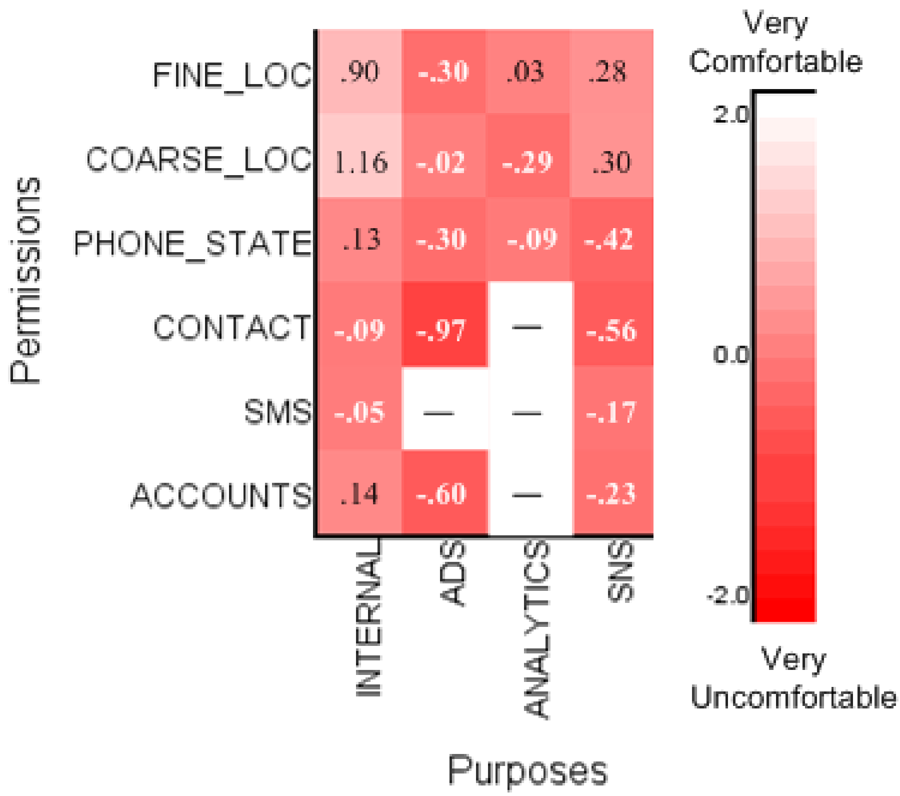


If this has failed on the fixed Web, what are the chances it will work on smartphones or in IoT?

Personalized Privacy Assistants

- **Selectively inform** us about privacy practices we may **not be expecting, yet care about**
- **Learn** many of our privacy preferences and **semi-automatically configure** many settings on our behalf
- **Motivate us** to occasionally revisit some of our preferences and decisions
- The assistants should ideally work across any number of environment and be **minimally disruptive**

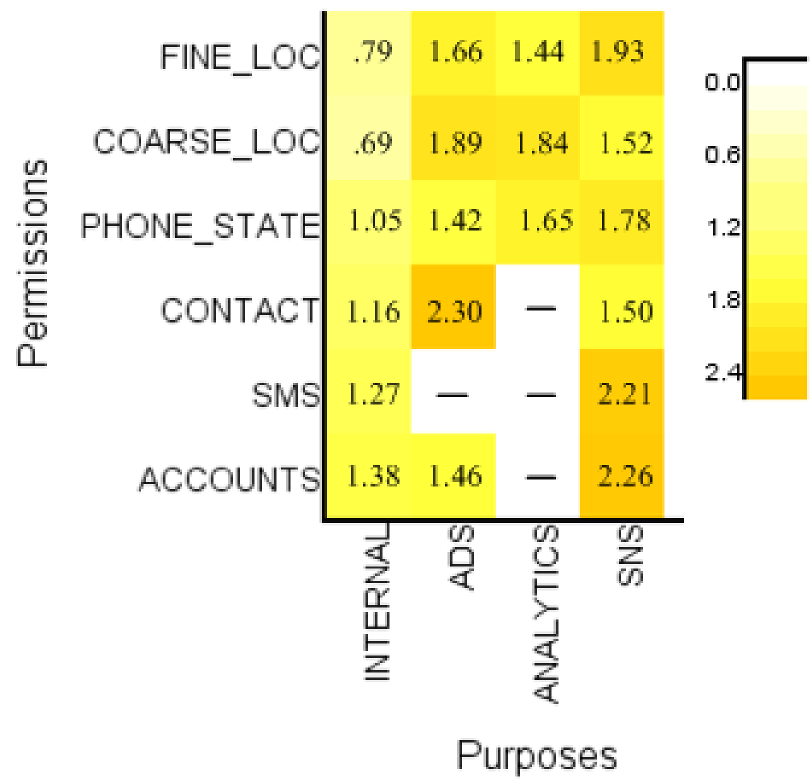
One Size-Fits-All Defaults Doesn't Work



Users' Average Preferences

White → comfortable

Red → uncomfortable



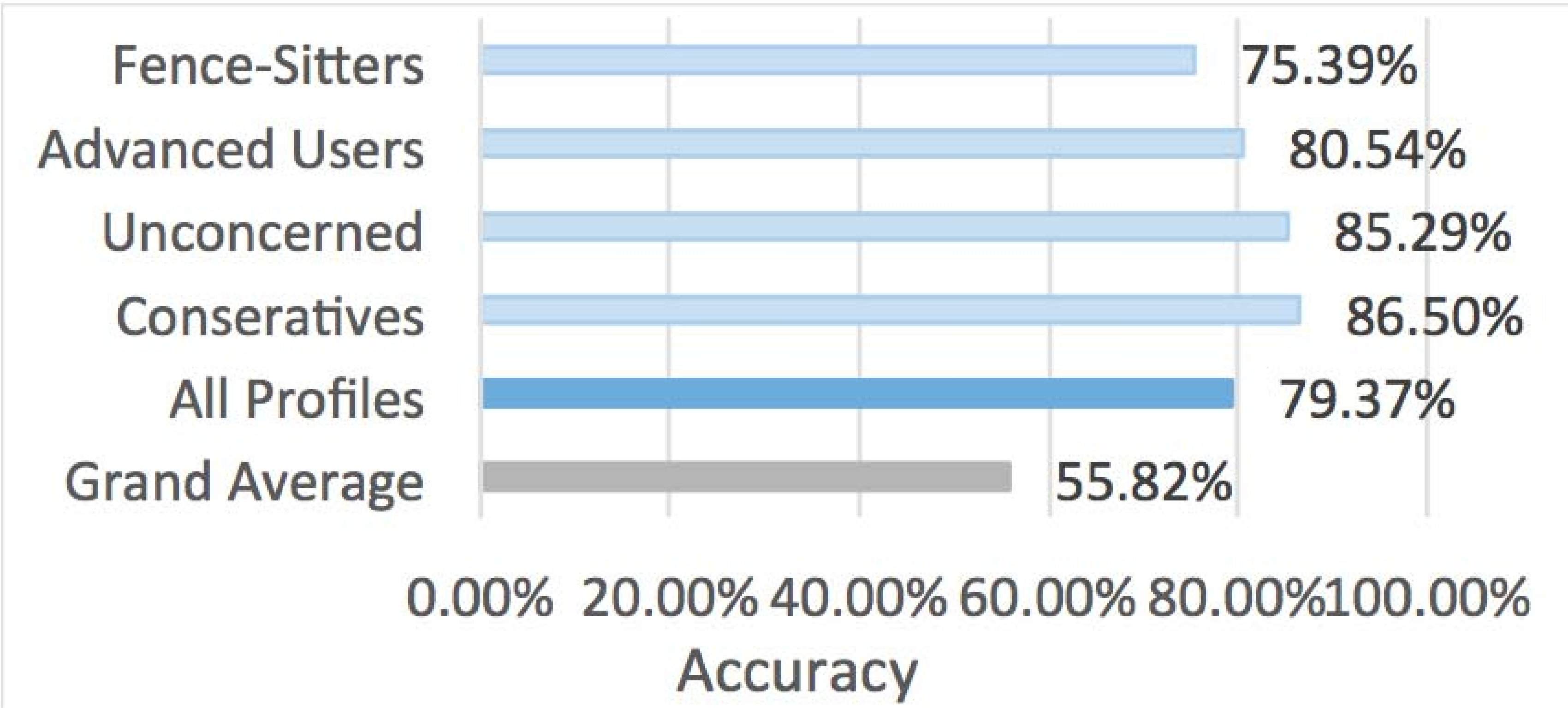
Variance among Users

Darker yellow → larger variance

Data based on 725 users and 837 apps (>21,000 HITs)

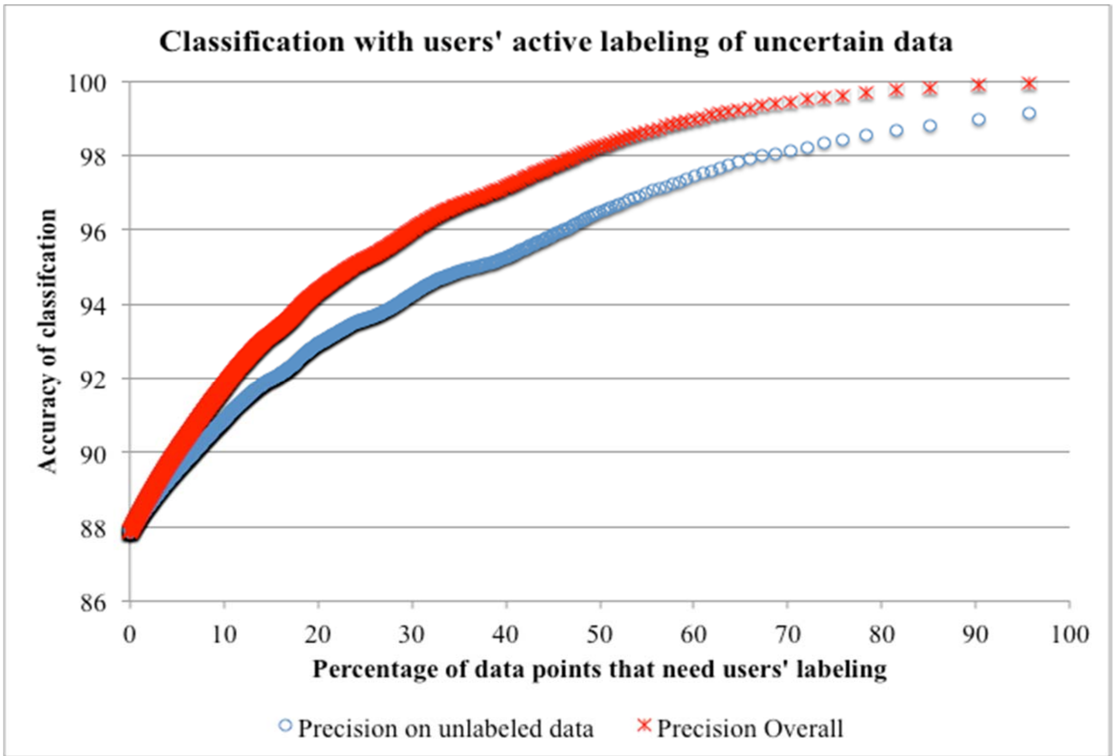
Mobile App Privacy Preferences

A small number of privacy profiles can go a long way



"Grand Average": Results obtained with "one-size-fits-all" profile

Pure Prediction vs. Interactive Model



Data from about 240,000 LBE users, 12,000 apps, 14.5M records

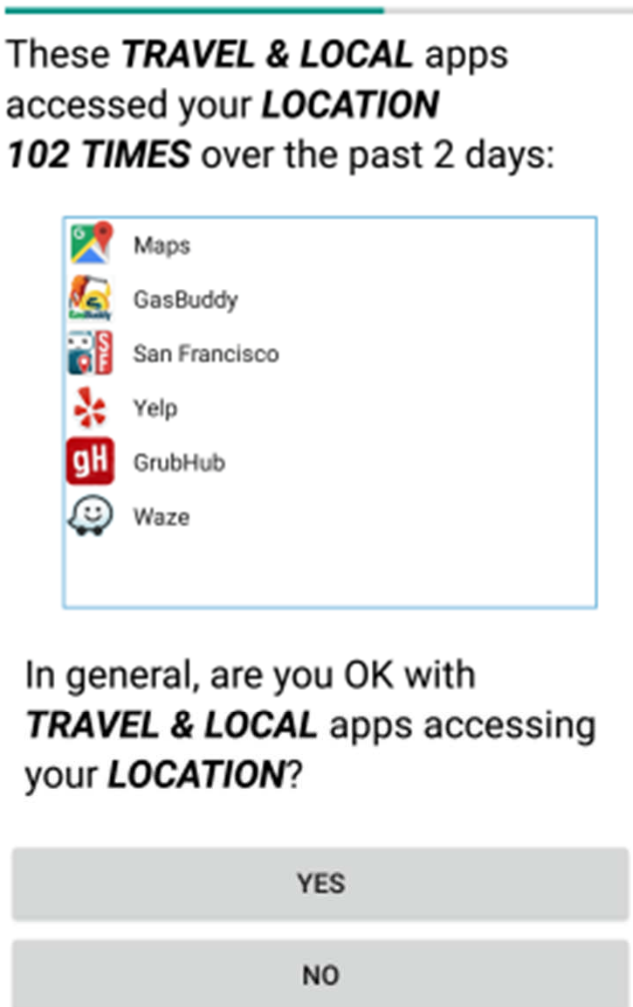
Learning personalized privacy preference models

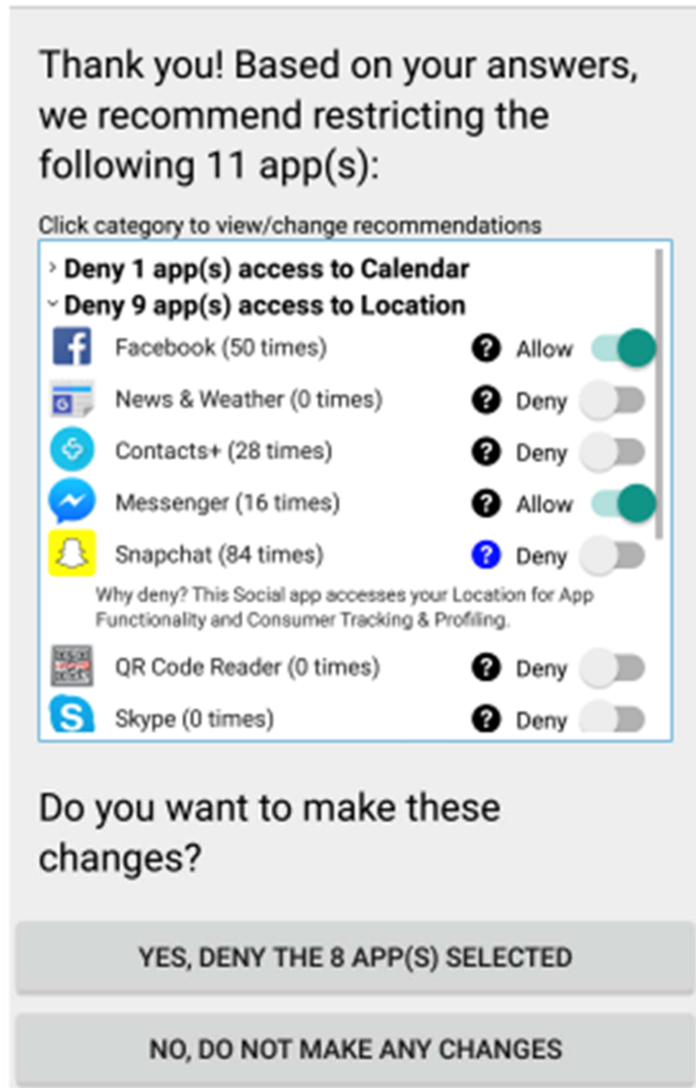
If users can label an additional 10% of their permission decisions, the **prediction accuracy will climb from 87.8% to 91.8%…and that's only 6 questions…**

At 20% (about 12 questions), accuracy climbs to 94%!

B. Liu, J. Lin, N. Sadeh, "Reconciling Mobile App Privacy and Usability on Smartphones: Could User Privacy Profiles Help?", WWW 2014. http://www.normsadeh.com/file_download/168

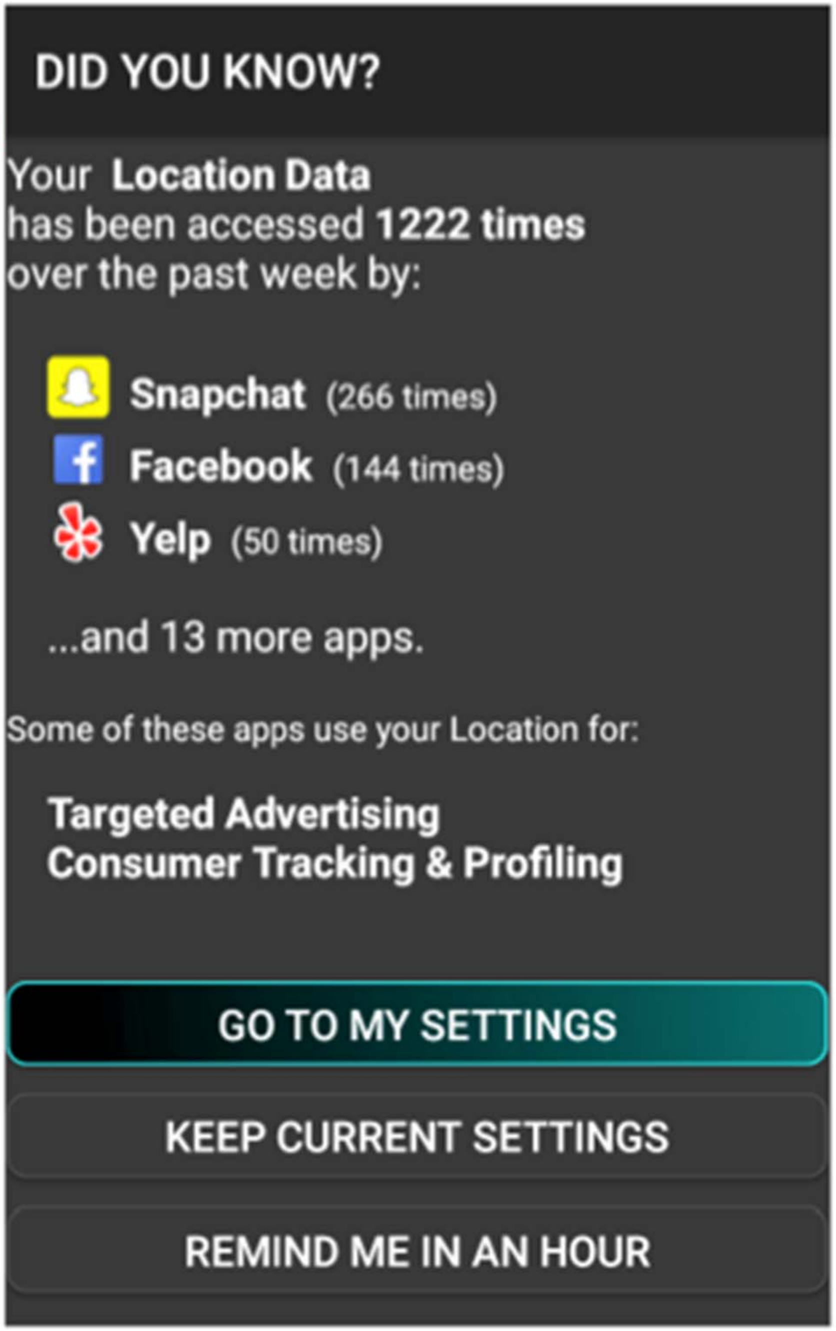
Personalized Privacy Assistant for Android Permissions





Nudging Users for 6 days

Are users just being nice or is this truly reflecting their preferences?

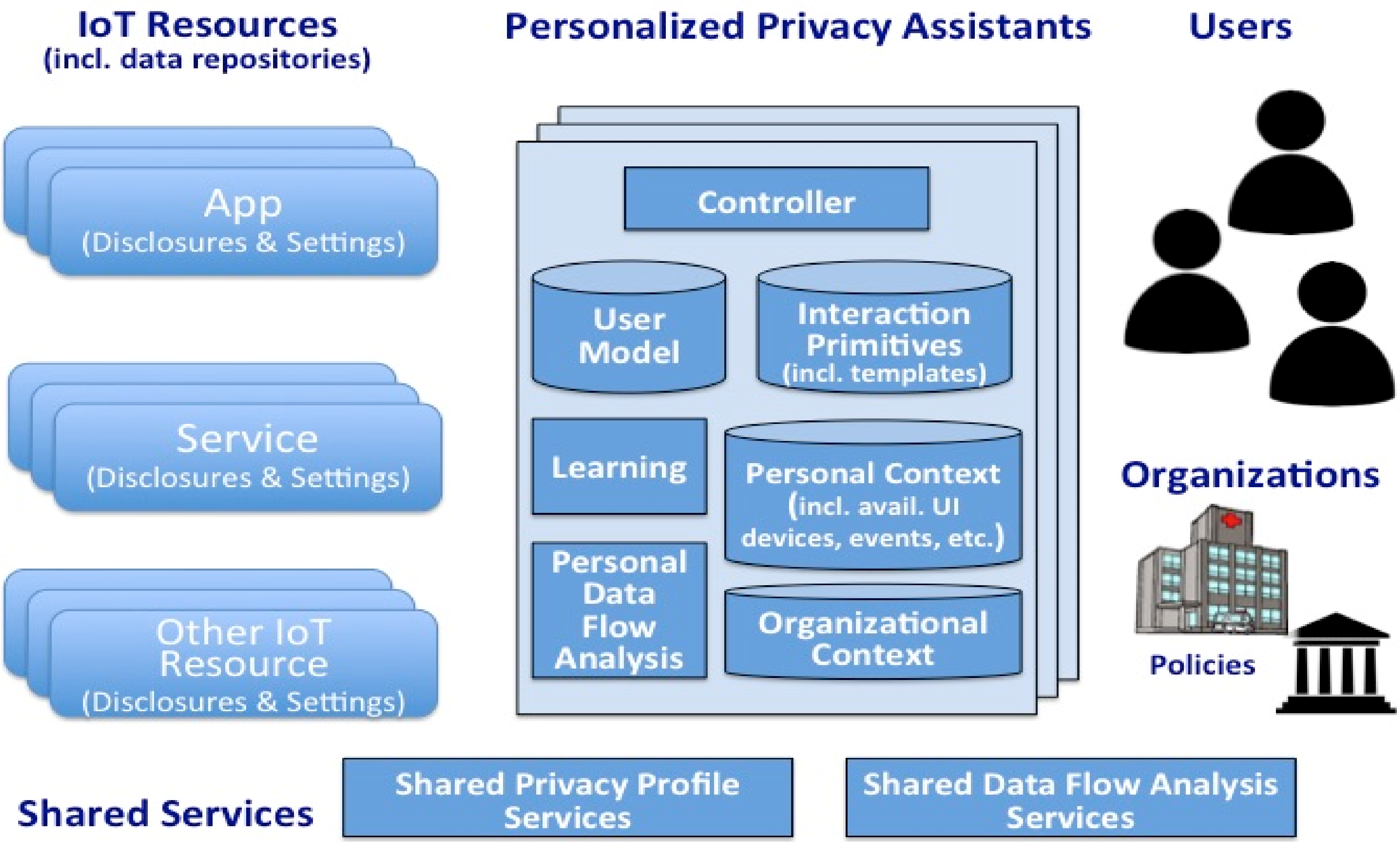


Successfully Piloted with Android Users

- Piloted with 29 Android users – 10 day study
- **Users accepted 73.7% of our recommendations**
- Only **5.6% of accepted recommendations were modified over the next 6 days**, despite nudges to revisit earlier decisions
  - Users showed great engagement, modifying many settings not covered in the recommendations
- Users are comfortable with the recommendations and see the value of the assistants

"To Deny, or Not to Deny: A Personalized Privacy Assistant for Mobile App Permissions," Bin Liu, Mads Schaarup Andersen, Florian Schaub, Norman Sadeh, Hazim Almuhimedi, Yuvraj Agarwal, Alessandro Acquisti - working paper, 2016

Extending this to IoT



Personalized Privacy Assistants for IoT

- Registries enable owners to **register their IoT resources**
  - Resources associated with locations/areas
  - Menus lead to automated **generation of machine-readable privacy policies**
- PPA's **discover relevant resources** by consulting registries & compare policies against user profiles (**expectations and preferences**)
  - Selective **alerts & semi-automated configuration** of available privacy settings

Concluding Remarks- I

- PPAs aim to provide a **pragmatic approach to notice and choice**
  - Leveraging machine learning and privacy profiles
  - Learning people's privacy preferences and expectations to **minimize user burden**, yet ensure that **users are informed** about those issues they care about and **retain sufficient control over their settings**

Concluding Remarks - II

- **Assumption**: Privacy profiles and learned preferences should **only be used for the purpose of managing user privacy**
- PPAs have to come with **strong privacy guarantees**
  - Could be offered by entities controlling specific ecosystems
  - Could be offered by **3rd parties dedicated to privacy management**
    - Opens the door to PPAs that cut across multiple ecosystems/environments but **requires open APIs**

Contact: *sadeh AT cs.cmu.edu*

Acknowledgement*: Funding provided under DARPA’s Brandeis intiative, NSF SaTC/SBE Program, Google Web of Things Expedition*

Contact: sadeh AT cs.cmu.edu

Collaborators*: Bin Liu, Jialiu Lin, Mads Scharup Andersen, Florian Schaub, Alessandro Acquisti, Yuvraj Agarwal, Lujo Bauer, Lorrie Cranor*

Carnegie Mellon University
Master of Science in
Privacy
ENGINEERING

Discussion of Session 5

Discussants:

- **Aaron Alva,** Federal Trade Commission
- **Geoffrey Manne,** International Center for Law and Economics
- **Davi Ottenheimer,** Institute for Applied Network Security

Presenters:

- **Sarthak Grover,** Princeton University
- **Vitaly Shmatikov,** Cornell Tech
- **Florian Schaub,** Carnegie Mellon University
- **Norman Sadeh,** Carnegie Mellon University

Closing Remarks

Lorrie Cranor, Chief Technologist

PRIVACYCON
FEDERAL TRADE COMMISSION