USABLE PRIVACY POLICY PROJECT

Privacy in the Age of IoT

Technologies to Help Users and Regulators

Norman Sadeh

Carnegie Mellon University

www.normsadeh.org

Carnegie Mellon University

CLIP Center on Law and Information Policy AT FORDHAM LAW SCHOOL

CIS The Center for Internet and Society

COLUMBIA UNIVERSITY

Outline

- “Notice and Choice”
- Privacy in the age of IoT
- The Usable Privacy Policy Project: Annotating Privacy Policies at Scale
- Technologies for Regulators and Developers
- Technologies for Users: Personalized Privacy Assistants

Privacy in the Age of IoT

- As we go about our daily lives, we interact with a number of devices, applications and services
- **Many of these devices applications and services may collect, share and mine data about us**
  - Many potential benefits
  - …but also many potential risks

Benefits …and Risks

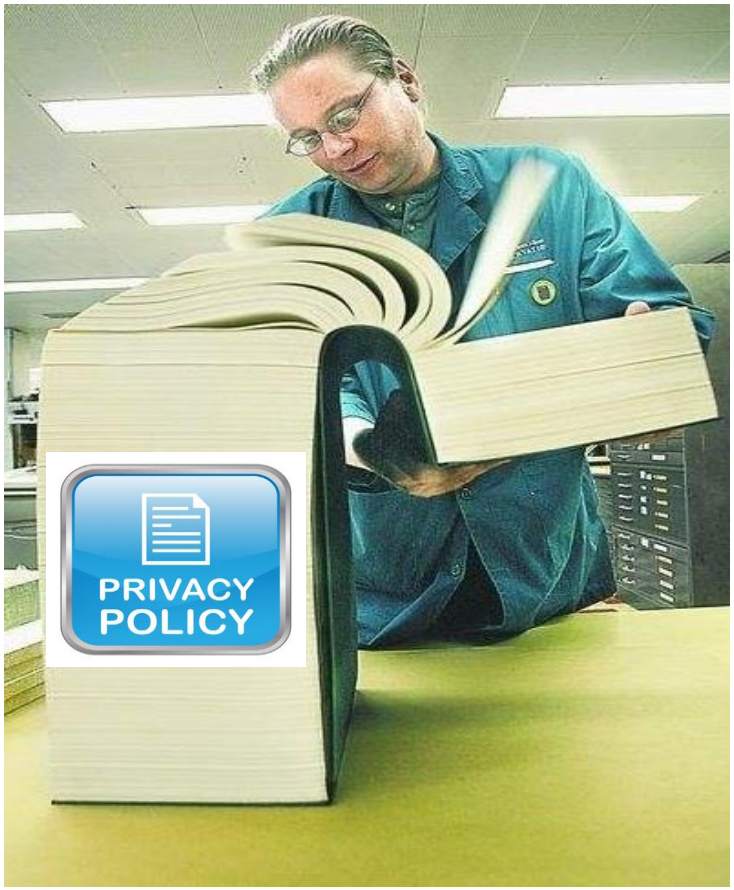
- Your home thermostat accesses your calendar to start the AC or furnace in time for when you return from work
- Your smartwatch might share your heart rate with your doctor
- …**But would you want…**
  - your phone to also report your driving habits to your car insurance provider?
  - your blood pressure to be sent to your health insurance provider?

“Notice and Choice”

- **Information Privacy:** People should have some control over what information about them is being collected and how it will be used
- **“Notice and Choice”** is intended to support **informed consent**
  - Different people have different privacy preferences
  - Enshrined in many legal documents
    - Including Hong Kong’s Personal Data (Privacy) Ordinance, EU GDRP, US COPPA, CalOPPA, etc.

People Are Feeling Helpless

- Reading a privacy policy takes about 10 minutes…or about 200 hours/year for an average Internet user…
  (McDonald & Cranor 2009)



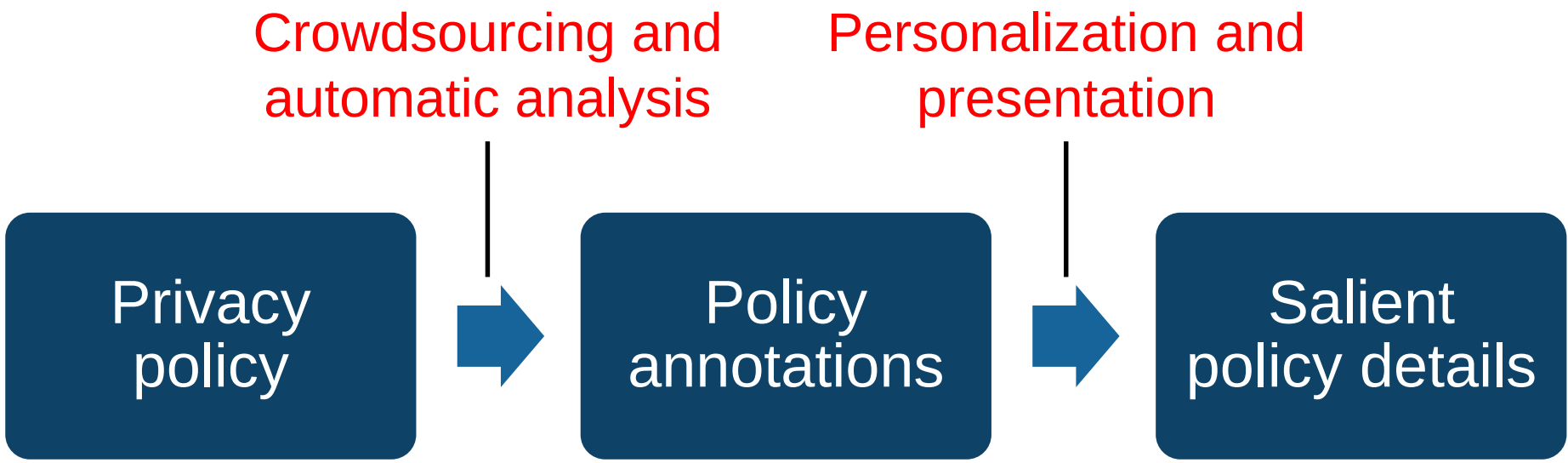
Mobile and IoT: A Number of Complicating Factors

- A typical mobile phone user with 50 mobile apps each requesting 3 permissions would have to **configure 150 settings**
- IoT: Technology is often **"invisible"**
- **Reading policies is even less practical**
- Explosion in the number of apps and devices: Developers often **lack the necessary sophistication**

"Modeling Users' Mobile App Privacy Preferences: Restoring Usablility in a Sea of Permission Settings", J. Lin, B. Liu, N. Sadeh, J. Hong, Proc. of the USENIX Symposium on Usable Privacy and Security, SOUPS 2014, Jul. 2014

The Usable Privacy Policy Project

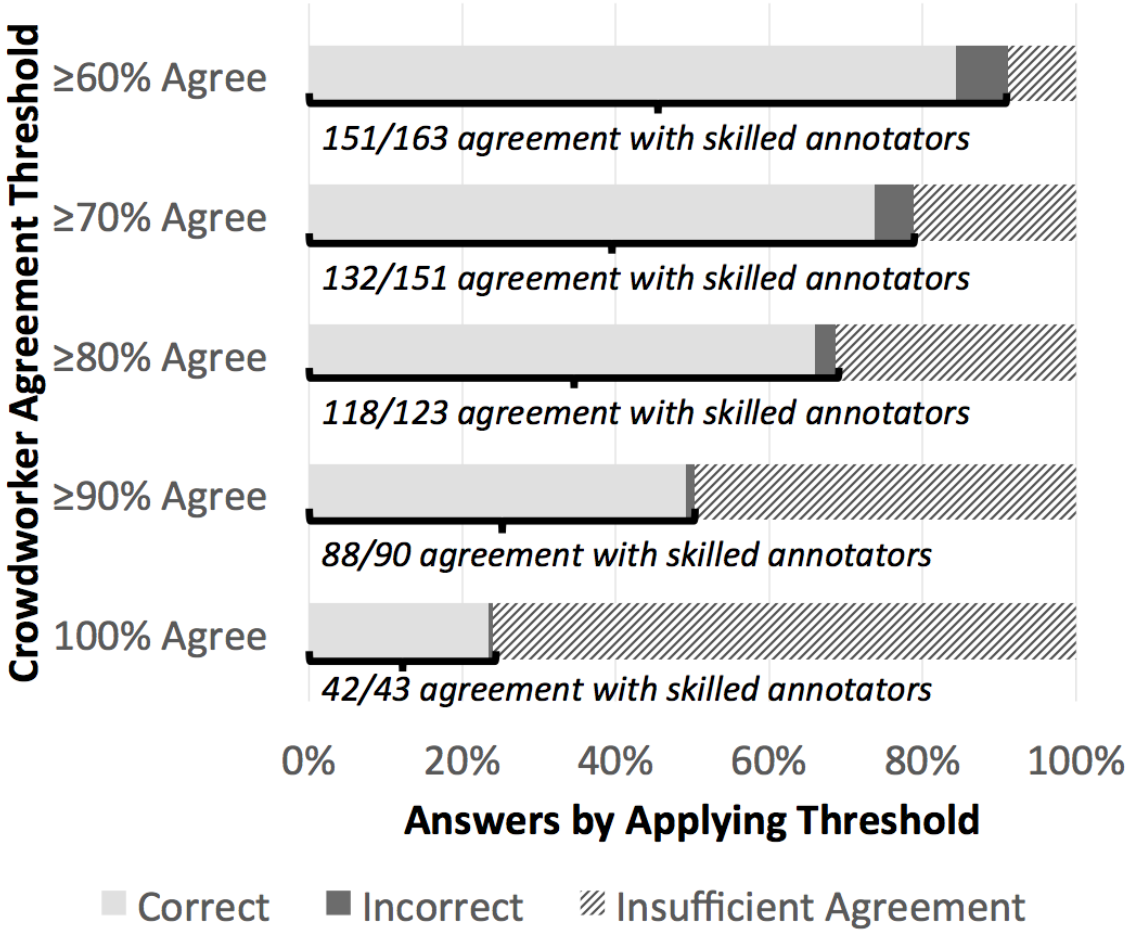
Approach: Use crowdsourcing, machine learning, and NLP techniques to automatically (or semi-automatically) extract salient details from privacy policies.



www.usableprivacy.org

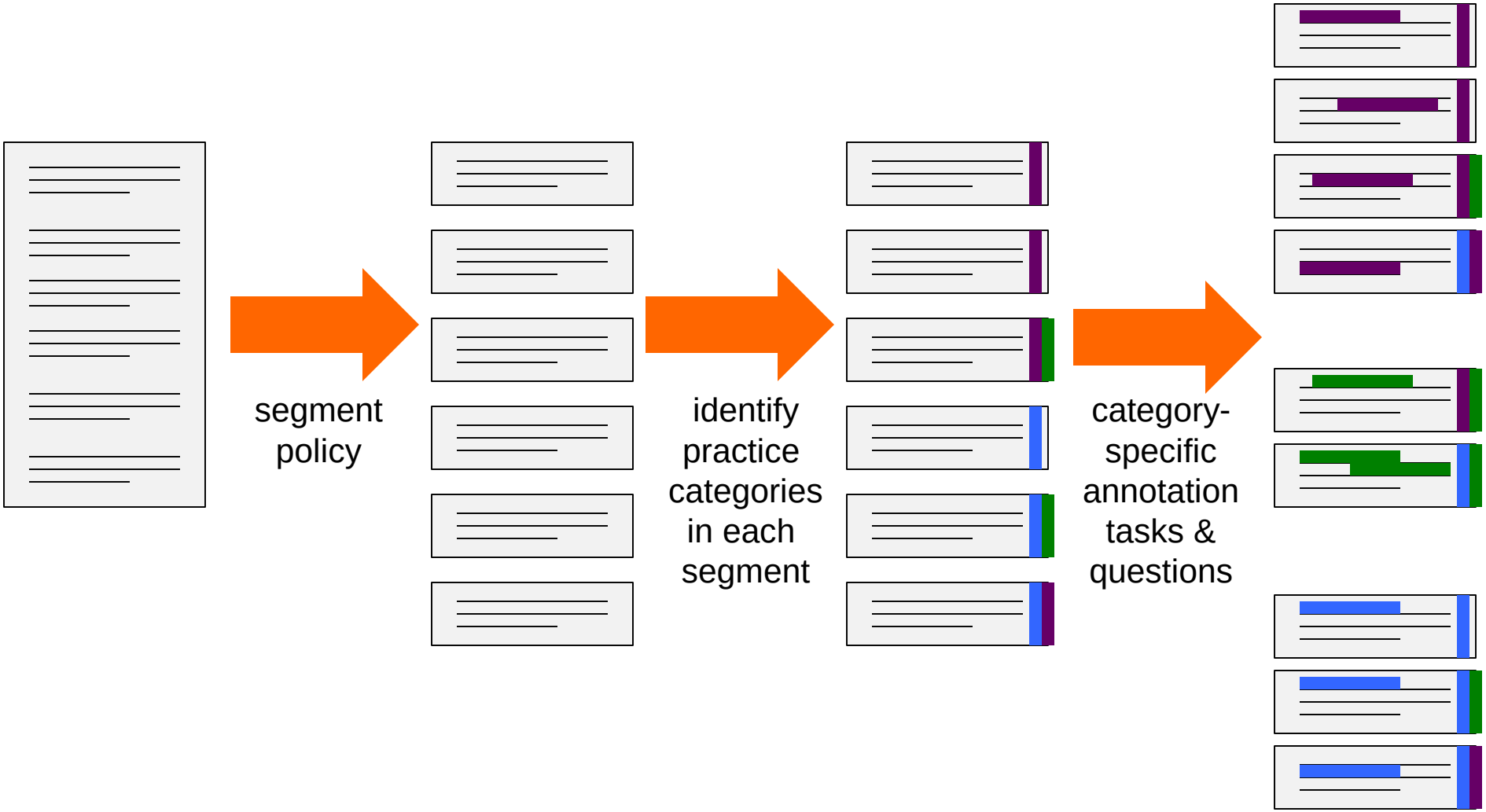
"The Usable Privacy Policy Project", N. Sadeh et al., CMU Technical Report, CMU-ISR-13-119, 2013

Crowdworkers Can Be Good at This(!)



Wilson, S., Schaub, F., Ramanath, R., Sadeh, N., Liu, F., Smith, N., and Liu, F. Crowdsourcing Annotations for Websites Privacy Policies: Can It Really Work? WWW Conference, May 2016

Multi-step annotations



Expert Annotation Tool

usableprivacy User Profile Task Visualize Settings Logout

Current Policy: www.imdb.com-privacypolicy-05-2014.csv

First Party Collection/Use | Third Party Sharing/Collection | User Choice/Control
User Access, Edit and Deletion | Data Retention | Data Security
Policy Change | Do Not Track | International and Specific Audiences | Other

4/29

Previous Next

Information You Give Us: We receive and store any information you enter on our Web site or give us in any other way. Click here to see examples of what we collect. You can choose not to provide certain information, but then you might not be able to take advantage of many of our features. We use the information that you provide for such purposes as responding to your requests, customizing future browsing for you, improving our site, and communicating with you.

Please write your comments for this paragraph

Practices of this paragraph

First Party Collection/Use

- Does Explicit Collect on website not-selected Generic personal information Basic service/feature not-selected Don't use service/feature not-selected Clone Delete

Third Party Sharing/Collection

User Choice/Control

User Access, Edit and Deletion

First Party Collection/Use

Does/Does Not	Does
Implicit/Explicit	Explicit
Action First-Party *	Collect on website
Identifiability	not-selected
Personal Information Type *	Generic personal information
Purpose *	Personalization/Customization
User Type	not-selected
Choice Type	Don't use service/feature
Choice Scope	not-selected

☐ References another place in the policy

Save

www.explore.usableprivacy.org

https://explore.usableprivacy.org/browse/category/

USABLEPRIVACY.ORG EXPLORE

About | Browse Privacy Policies | Search for a website

Browse

by Category | Readability | Popularity

- Arts 68
- Business 53
- Computers 42
- Games 26
- Health 35
- Home 37
- Kids and Teens 46
- News 32
- Recreation 42
- Reference 31

Arts 68

E! Online
Privacy policy from Jan 14, 2015 with 256 practice statements.

FOX Sports
Privacy policy from Jun 11, 2015 with 215 practice statements.

Racked
Privacy policy from May 1, 2014 with 204 practice statements.

See more

Business 53

Blogger
Privacy policy from Jun 30, 2015 with 241 practice statements.

AOL
Privacy policy from Jun 23, 2015 with 232 practice statements.

Allstate
Privacy policy from May 29, 2015 with 226 practice statements.

See more

USABLEPRIVACY.ORG EXPLORE

About | Browse Privacy Policies | Search for a website

Playstation playstation.com

Games | Kids and Teens | World

Take a tour

Privacy Practices

Click a category to filter practice statements.

- First Party Collection/Use 87
- Third Party Sharing/Collection 34
- User Choice/Control 5
- User Access, Edit and Deletion 5
- Data Retention 7
- Data Security 16
- Policy Change 5
- Do Not Track 0
- International and Specific Audiences 9

Privacy Policy

Playstation Privacy Policy from Apr 1, 2011. Reading Level: College Graduate (Grade 17)
168 privacy practice statements in total

Privacy Policy

Last Revised: April, 2011

Sony Computer Entertainment America LLC ("SCEA") is committed to respecting the privacy rights of all visitors to our websites. This privacy policy is intended to provide you with information on how we collect, use and store the information that you provide to us through our websites so that you can make appropriate choices for sharing information with us. If you have any questions, complaints or comments regarding our online or offline privacy policies, please contact SCEA's Consumer Services Hotline at 1-800-345-7669.

This Privacy Statement and the certification seal located to your right confirms that SCEA is a valid licensee and participating member in the Entertainment Software Rating Board's Privacy Online Program: ESRB Privacy Online. **To protect your privacy to the maximum extent possible, we have undertaken this privacy initiative and our websites have been reviewed and certified by ESRB Privacy Online to meet**

USABLEPRIVACY.ORG EXPLORE About Browse Privacy Policies Search for a website

Playstation playstation.com

Take a tour

Games | Kids and Teens | World

Privacy Practices

Clear Filters

Click a category to filter practice statements.

- First Party Collection/Use 87
- Third Party Sharing/Collection 34
- User Choice/Control 3

Choice Type

- All
- Opt-in (3)
- Don't use service (2)

Choice Scope

- All
- Unspecified (2)

- User Access, Edit and Deletion 5
- Data Retention 7
- Data Security 16

Privacy Policy

Playstation Privacy Policy from Apr 1, 2011. Reading Level: College Graduate (Grade 17)
168 privacy practice statements in total

HOW WE USE YOUR INFORMATION

Personally identifying information that we collect for a particular promotional purpose through one of our websites or to make a purchase from the PlayStation Shop is saved and used only for that purpose, unless the participant chooses to opt-in to one of our marketing programs. **Visitors to our websites may be given the opportunity to "opt-in" to two different programs. The first option is to receive marketing content from SCEA. The second is to have personal information shared with SCEA's third party partners so that they may send you marketing materials.**

Consumers who voluntarily provide personally identifying information via our website for purposes of receiving marketing materials or who opt-in to receiving marketing materials when they register a Network Adaptor (Ethernet/Modem)(for PlayStation2) or PlayStation2 with integrated network and line connectors for online gaming through the Online Start-Up Disc, become members of our marketing program.

By logging in as a Sony Entertainment Network account holder on our websites, consumers

The OPP-115 Corpus

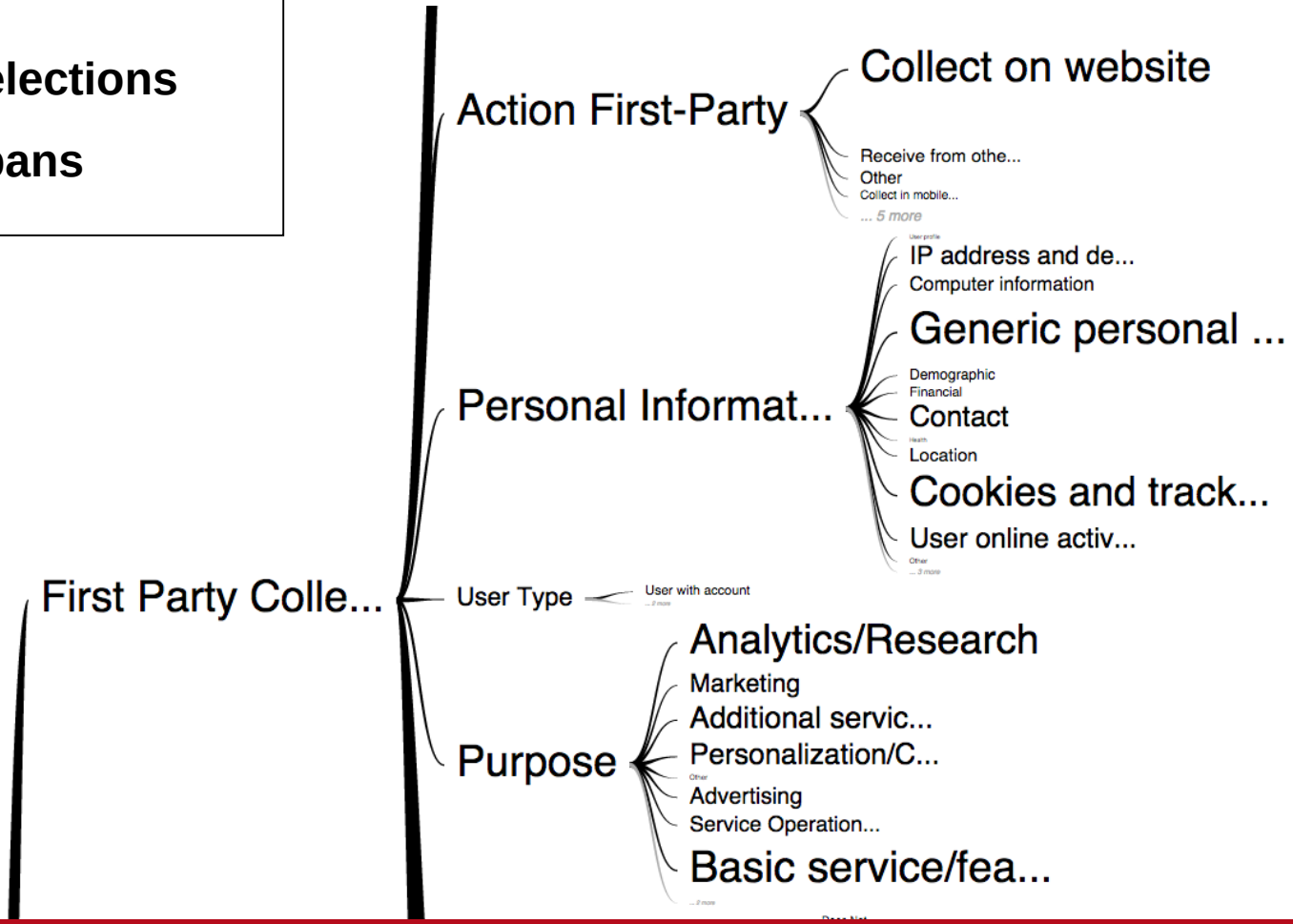
115 annotated privacy policies

267K words

23K data practices

128K attribute-value selections

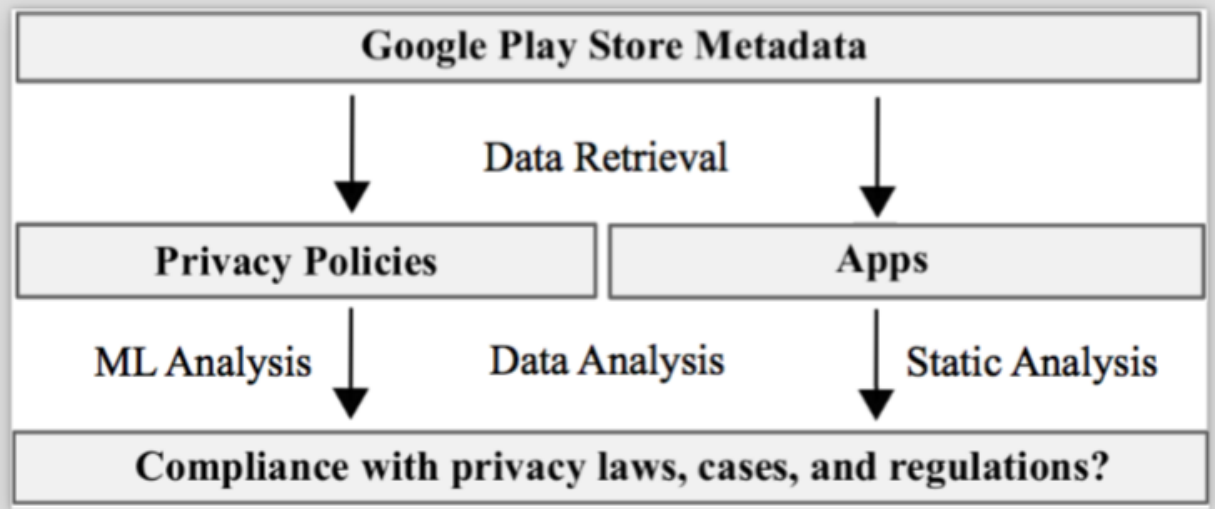
103K annotated text spans



Question

- Could we automatically analyze privacy policies and identify compliance issues?

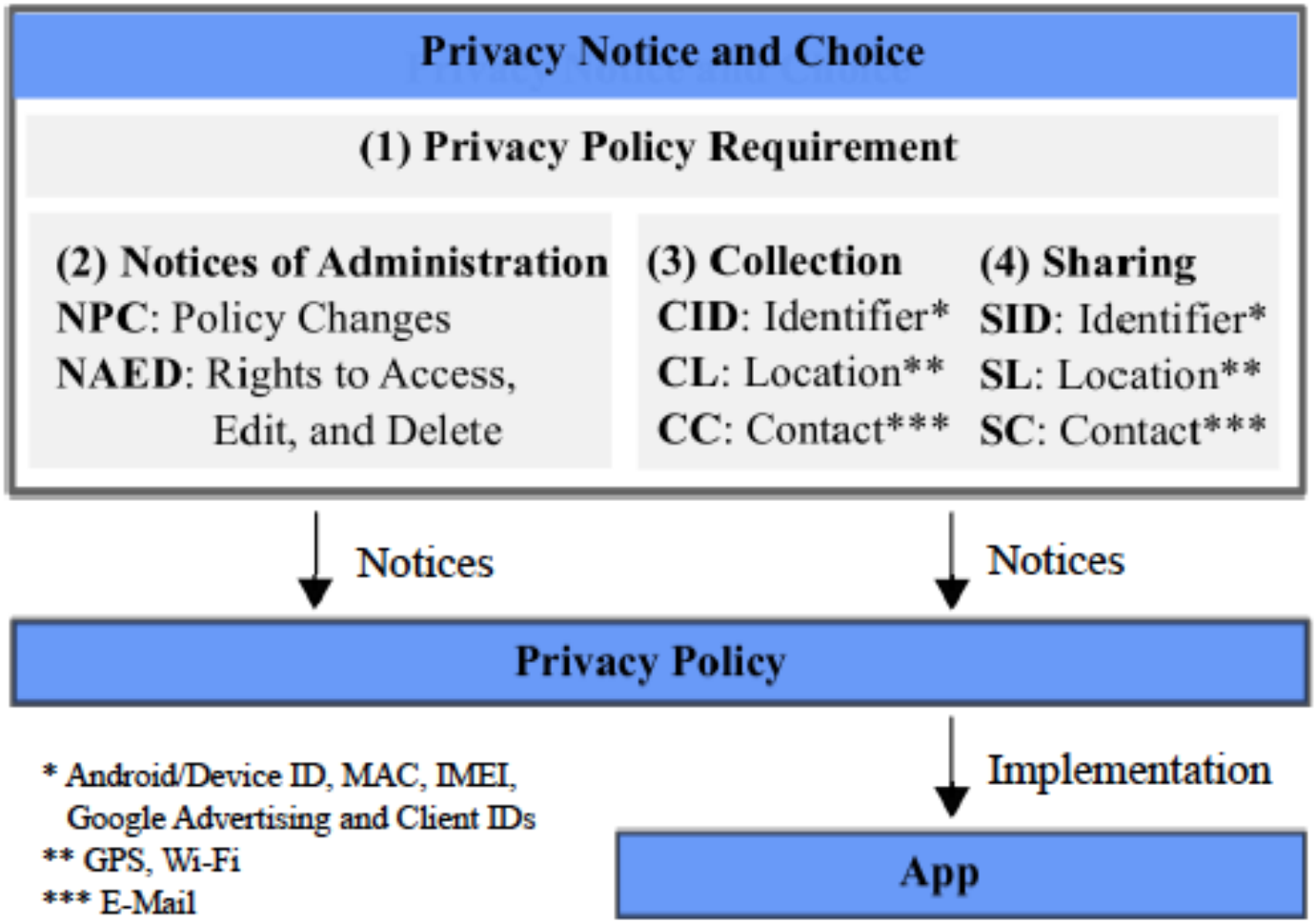
Approach: Mobile Apps



- Training **machine learning classifiers** to extract relevant policy statements
- Compare these statements against:
  - **Regulatory requirements**
  - What the software actually does
    - **Static and dynamic code analysis**

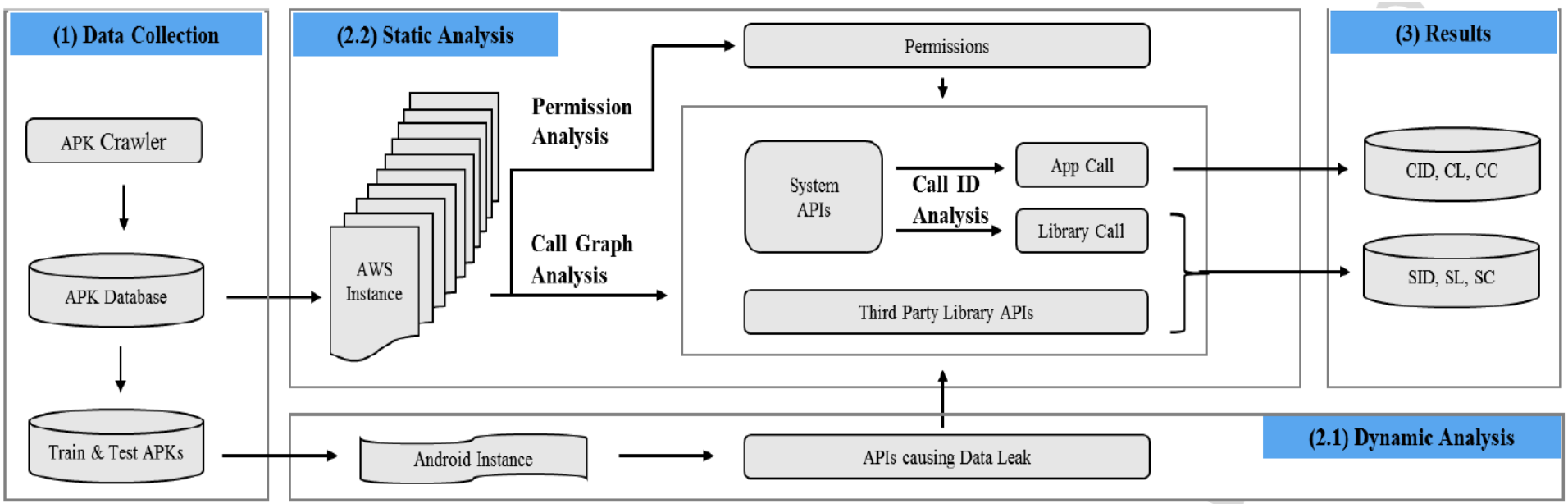
"Analyzing and Predicting Privacy Law Compliance of Mobile Apps", S. Zimmeck, Z.Wang, L. Zou, B. Liu, F. Schaub, S. Wilson, N. Sadeh, S. Bellovin, J. Reidenberg, paper under review, 2016

Formalizing the Problem



Note: In US, FTC FIPPS mandates notice before collection of PII; COPPA requires policies for apps directed to children; CalOPPA: policy required if PII collected; COPPA requires NAED; CID and CL require disclosureunder CalOPPA and COPPA and sharing requires consent; CalOPPA and DOPPA require description of notification process for policy change

Code Analysis – Mobile Apps



- Using Androguard
- Static analysis to identify use of sensitive data by 3rd party libraries
- Dynamic analysis to study the behavior of 3rd party libraries

Automatic Policy Analysis

- Looking for privacy practices **not** disclosed in the privacy policy
- One classifier built for each practice
- Classifiers trained on corpus of 115 privacy policies annotated by law students

Results - Policy Analysis

Practice	*Algorithm*	*Parameters*	*Base*	Acc_{pol}	*95% CI*	$Prec_{neg}$	Rec_{neg}	$F\text{-}1_{neg}$	$F\text{-}1_{pos}$	*Pos%*
NPC	SVM	RBF kernel, weight	0.7	0.9	0.76–0.97	0.79	0.92	**0.85**	0.93	46%
NAED	SVM	linear kernel	0.58	0.75	0.59–0.87	0.71	0.71	**0.71**	0.78	36%
CID	Log. Reg.	LIBLINEAR solver	0.65	0.83	0.67–0.93	0.77	0.71	**0.74**	0.87	46%
CL	SVM	linear kernel	0.53	0.88	0.73–0.96	0.83	0.95	**0.89**	0.86	34%
CC	Log. Reg.	LIBLINEAR, L2, weight	0.8	0.88	0.73–0.96	0.71	0.63	**0.67**	0.92	56%
SID	Log. Reg.	LBFGS solver, L2	0.88	0.88	0.73–0.96	0.94	0.91	**0.93**	**0.55**	10%
SL	SVM	linear kernel, weight	0.95	0.93	0.8–0.98	0.97	0.95	**0.96**	-	12%
SC	SVM	poly kernel (4 degrees)	0.73	0.78	0.62–0.89	0.79	0.93	**0.86**	**0.47**	6%

- Best results obtained with fairly simple classifiers: Logistic regressions and Support Vector Machines
- F-1: F score measures accuracy and recall
- **F-1neg**: measure focusing on negative condition (i.e. **absence of statement**), which is **what matters from a compliance perspective**

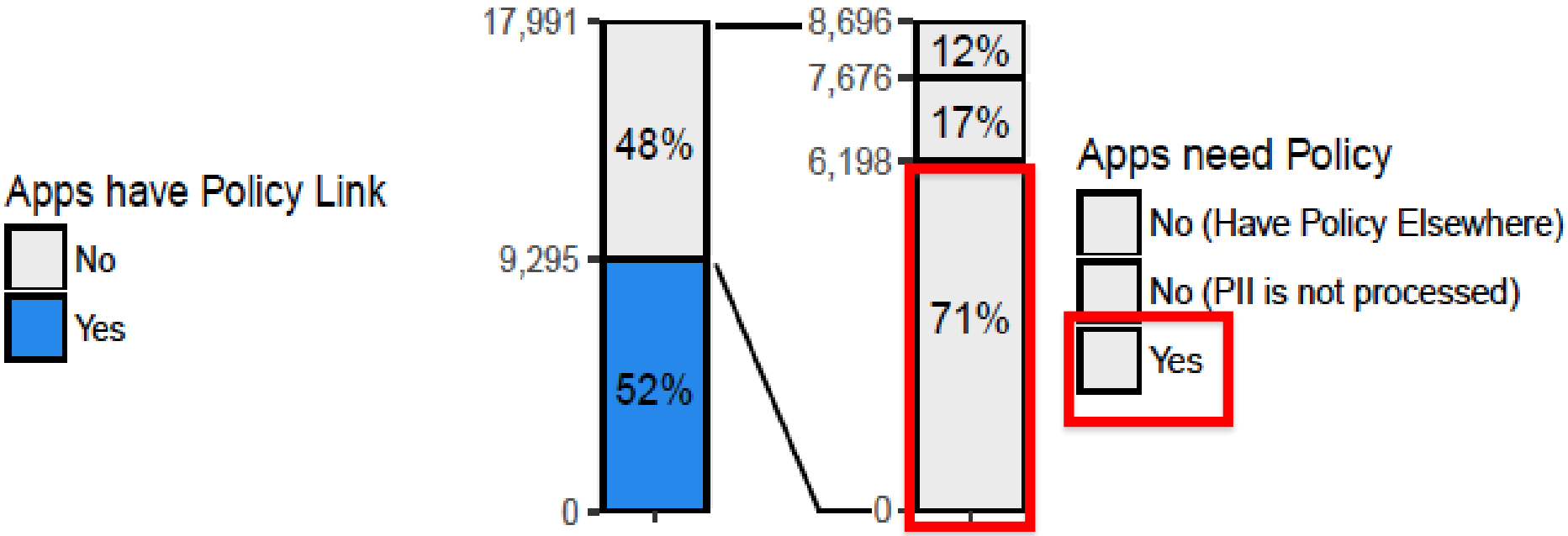
Results – Code Analysis

Prct	*Base*	Acc_{app}	*95% CI*	$Prec_{pos}$	Rec_{pos}	$F\text{-}1_{pos}$	$F\text{-}1_{neg}$	$Pos\%_{w/\ pol}$	$Pos\%_{w/o\ pol}$
CID	0.55	0.87	0.7–0.96	0.84	0.94	**0.89**	0.85	95%	87%
CL	0.52	0.84	0.66–0.95	0.92	0.73	**0.81**	0.86	66%	49%
CC	0.9	1	0.89–1	1	1	**1**	1	**25%**	12%
SID	0.71	0.94	0.79–0.99	0.95	0.95	**0.95**	0.89	**71%**	**62%**
SL	0.9	1	0.89–1	1	1	**1**	1	20%	16%
SC	0.97	1	0.89–1	1	1	**1**	1	**2%**	0%

- Automated Code analysis evaluated against manual analysis for 30 mobile apps
- F-1: F score measures accuracy and recall
- **F-1pos**: measure focusing on positive condition (i.e. **identification of the collection or sharing of sensitive data**), which is what matters here

Major Findings - I

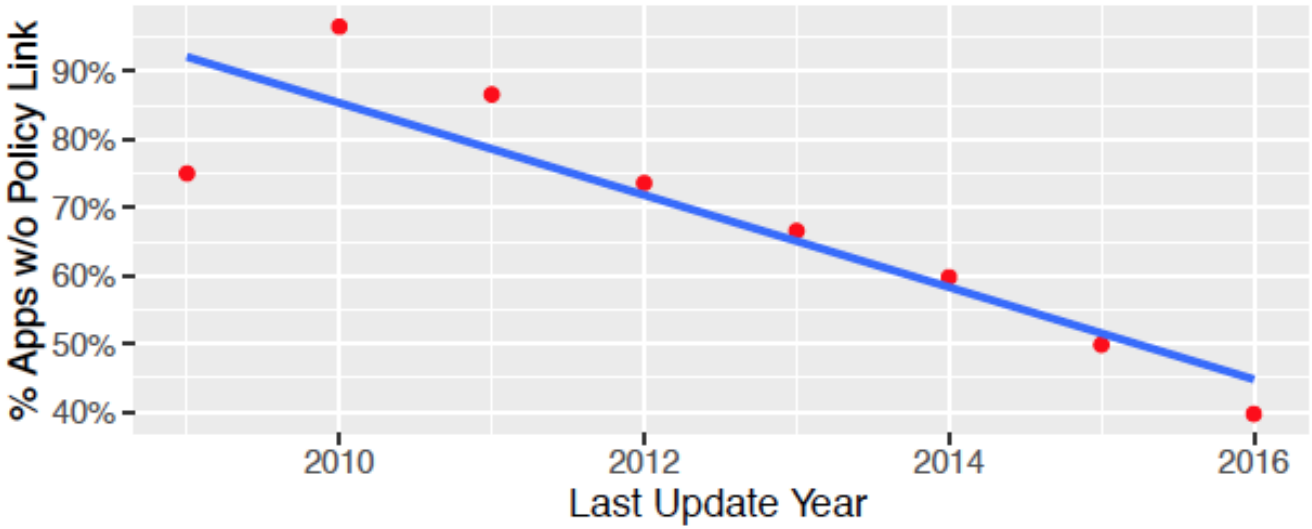
- No Policy Link

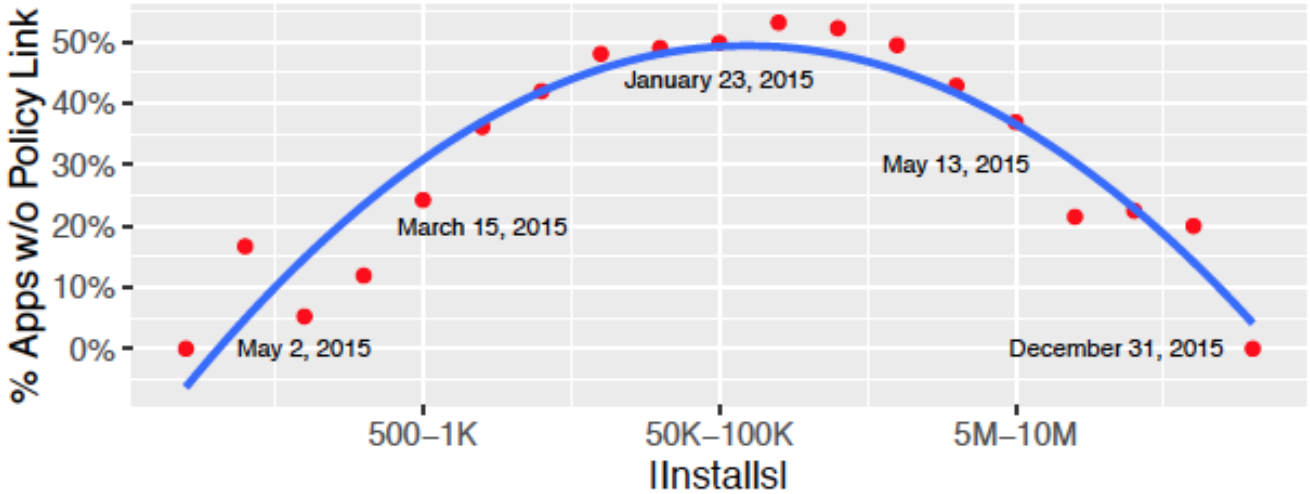


Analysis of 17,991 mobile apps

- 71% of apps with no policy seem to be in violation

Major Findings - II





Major Findings (Policy Analysis) - III

Practice	*Algorithm*	*Parameters*	*Base*	Acc_{pol}	*95% CI*	$Prec_{neg}$	Rec_{neg}	$F\text{-}1_{neg}$	$F\text{-}1_{pos}$	*Pos%*
NPC	SVM	RBF kernel, weight	0.7	0.9	0.76–0.97	0.79	0.92	**0.85**	0.93	46%
NAED	SVM	linear kernel	0.58	0.75	0.59–0.87	0.71	0.71	**0.71**	0.78	36%
CID	Log. Reg.	LIBLINEAR solver	0.65	0.83	0.67–0.93	0.77	0.71	**0.74**	0.87	46%
CL	SVM	linear kernel	0.53	0.88	0.73–0.96	0.83	0.95	**0.89**	0.86	34%
CC	Log. Reg.	LIBLINEAR, L2, weight	0.8	0.88	0.73–0.96	0.71	0.63	**0.67**	0.92	56%
SID	Log. Reg.	LBFGS solver, L2	0.88	0.88	0.73–0.96	0.94	0.91	**0.93**	**0.55**	10%
SL	SVM	linear kernel, weight	0.95	0.93	0.8–0.98	0.97	0.95	**0.96**	-	12%
SC	SVM	poly kernel (4 degrees)	0.73	0.78	0.62–0.89	0.79	0.93	**0.86**	**0.47**	6%

- Analysis of **9,050 mobile app privacy policies** (processing time about 30 minutes)
- **Only 46%** of apps seem to describe their **notification process for policy changes** – required under CalOPPA and DOPPA
- **Only 36%** seem to describe **user access, edit and deletion** rights (e.g. required by COPPA for children)
- **Sharing practices** (e.g. 12% location) **appear very low…more later**

Major Findings (Code Analysis)- IV

Prct	Base	Acc_{app}	95% CI	$Prec_{pos}$	Rec_{pos}	$F\text{-}1_{pos}$	$F\text{-}1_{neg}$	$Pos\%_{w/\ pol}$ (With policy)	$Pos\%_{w/o\ pol}$ (WO policy)
CID	0.55	0.87	0.7–0.96	0.84	0.94	**0.89**	0.85	95%	87%
CL	0.52	0.84	0.66–0.95	0.92	0.73	**0.81**	0.86	66%	49%
CC	0.9	1	0.89–1	1	1	**1**	1	**25%**	12%
SID	0.71	0.94	0.79–0.99	0.95	0.95	**0.95**	0.89	**71%**	**62%**
SL	0.9	1	0.89–1	1	1	**1**	1	20%	16%
SC	0.97	1	0.89–1	1	1	**1**	1	**2%**	0%

Code Analysis of 17,991 mobile apps & 6 practices

- *Apps with privacy policies (9,295)*: average of 2.79 positive practices (out of 6 possible practices)
  - 71% SID but only 10% disclose it (see previous slide)! **Suggests at least 61% are non-compliant**
- *Apps without privacy policies (8,696)*: average of 2.27 positive practices (out of 6 possible practices) – *reminder: These practices have to be disclosed* – ***indicative of likely violation in many of these 8,696 apps (71% of these apps)***

Question

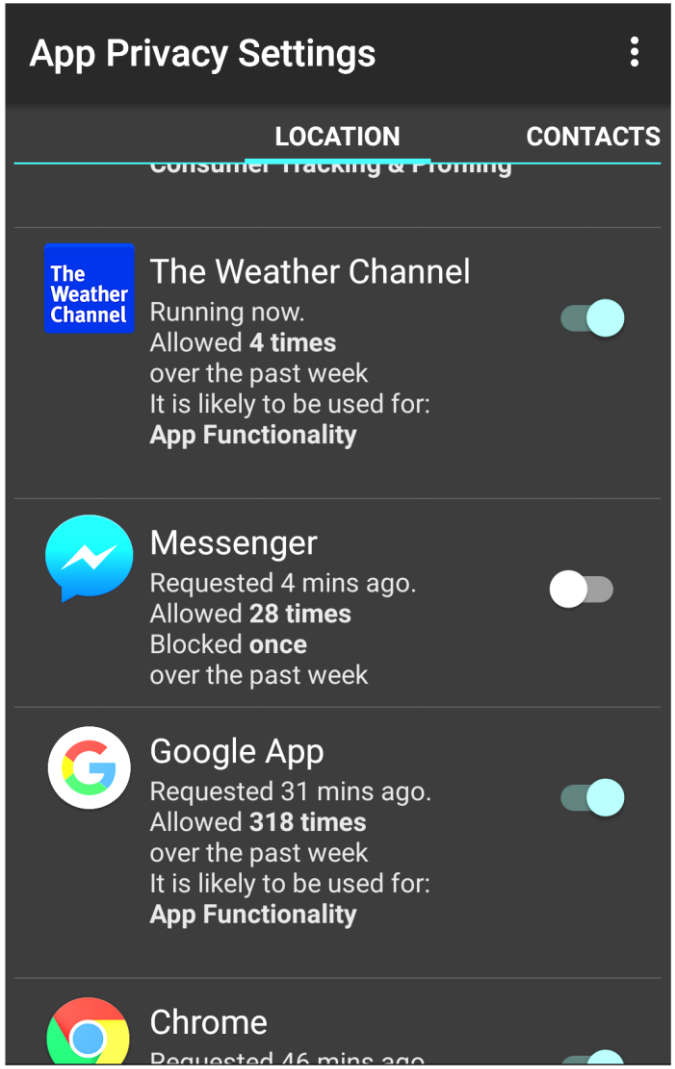
- Could similar technology also help users configure privacy settings?
- **Application**: Mobile App Permission Settings

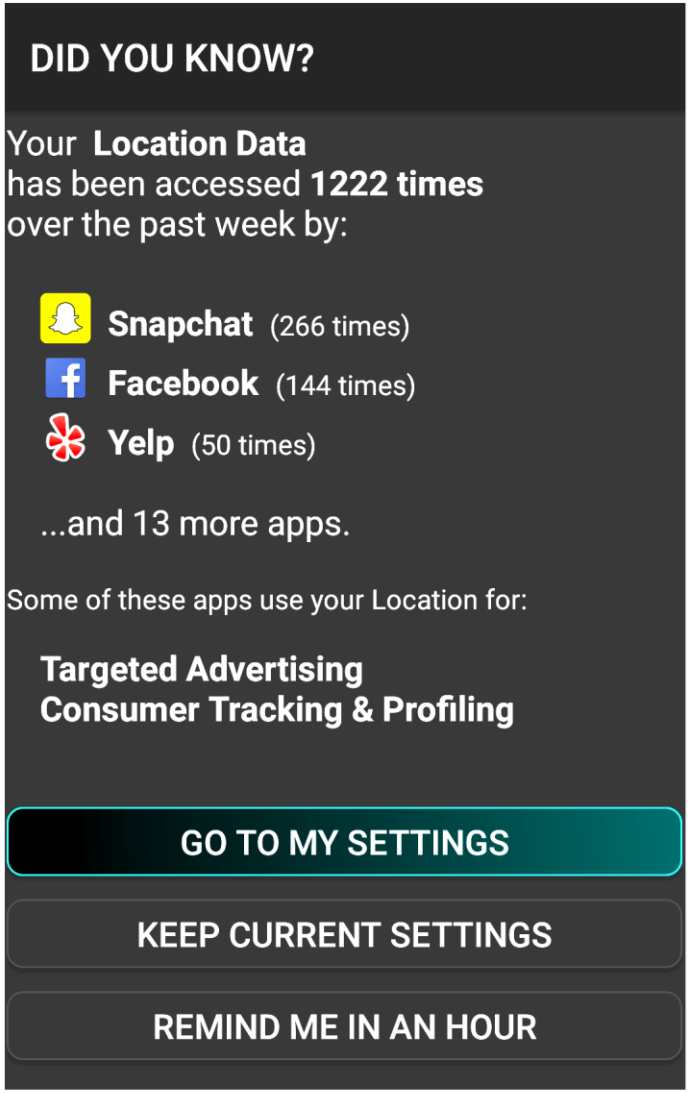
"Follow My Recommendations: A Personalized Privacy Assistant for Mobile App Permissions", B. Liu, M. Schaarups Andersen, F. Schaub, H. Almuhimedi, S. Zhang, N. Sadeh, A. Acquisti, Y. Agarwal, Proc. of the USENIX Symposium on Usable Privacy and Security, SOUPS 2016, June 2016

Approach

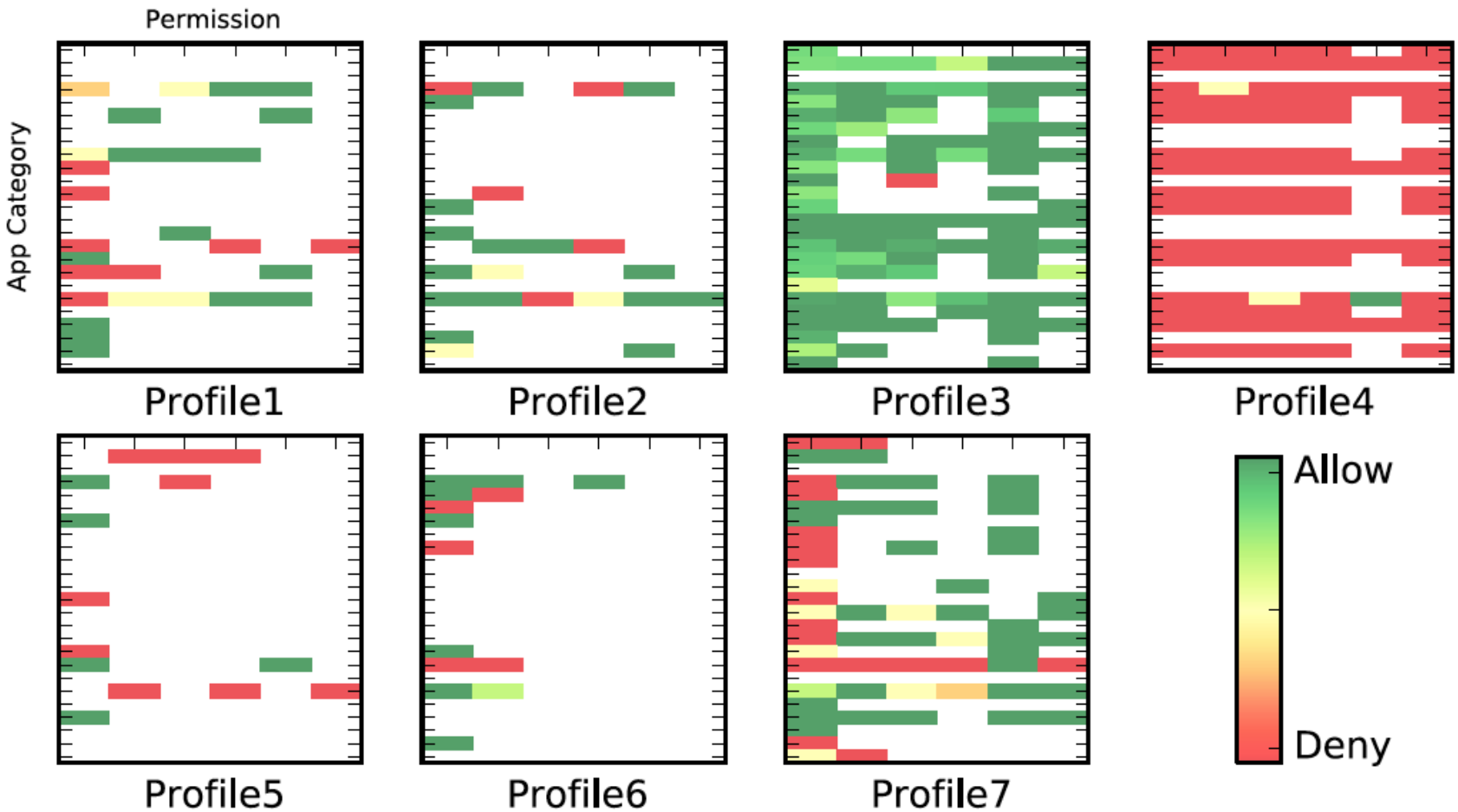
- Learn People's Mobile App Privacy Preferences
  - Including analysis of permission purpose, using code analysis
- Build Privacy Profiles (clusters of users)
- Ask each user a few questions to identify a profile that best matches their preferences
- Based on their profiles and the apps on their smartphones, recommend settings

Learning People's Privacy Preferences





Privacy Profiles – Hierarchical Clustering



- App categories along vertical axis; Permissions along horizontal axis
- Clustering based on triples for each user**: <app category, permission, purpose> - purpose can be obtained via static code analysis – similar to previous study**
- **Profile-based recommendations – using SVM**

Dialogue with Users: Profile Assignment & Setting Recommendations

These ***TRAVEL & LOCAL*** apps accessed your ***LOCATION*** ***102 TIMES*** over the past 2 days:

- Maps
- GasBuddy
- San Francisco
- Yelp
- GrubHub
- Waze

In general, are you OK with ***TRAVEL & LOCAL*** apps accessing your ***LOCATION***?

YES

NO

Thank you! Based on your answers, we recommend restricting the following 11 app(s):

Click category to view/change recommendations

› **Deny 1 app(s) access to Calendar**

˅ **Deny 9 app(s) access to Location**

Facebook (50 times)	Allow	ON
News & Weather (0 times)	Deny	OFF
Contacts+ (28 times)	Deny	OFF
Messenger (16 times)	Allow	ON
Snapchat (84 times)	Deny	OFF
Why deny? This Social app accesses your Location for App Functionality and Consumer Tracking & Profiling.		
QR Code Reader (0 times)	Deny	OFF
Skype (0 times)	Deny	OFF

Do you want to make these changes?

YES, DENY THE 8 APP(S) SELECTED

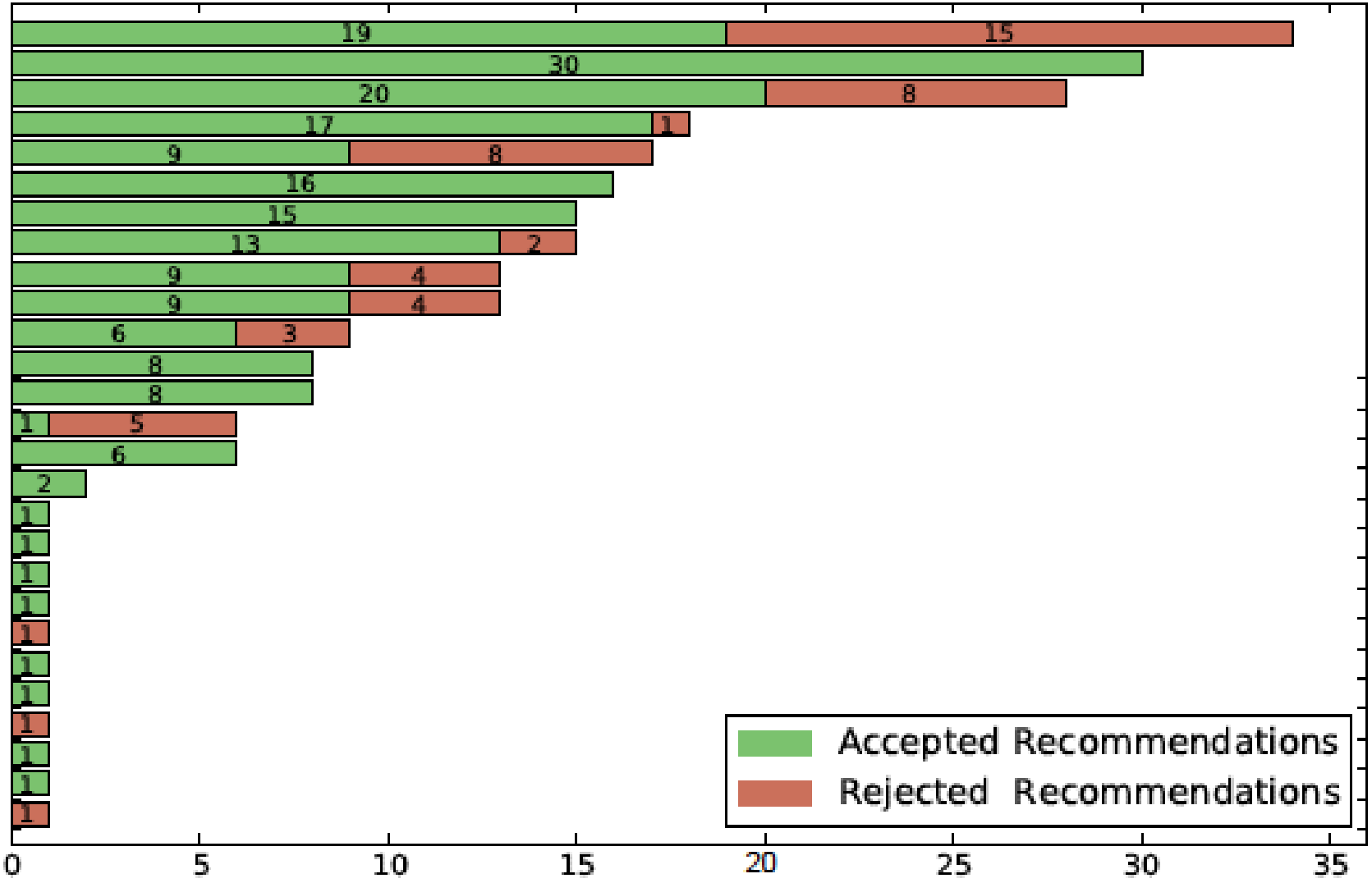
NO, DO NOT MAKE ANY CHANGES

Including explanation

Field Study: Evaluating the Recommendations

- **Recruited Android Users**: installed the privacy assistant on their actual Android phones; observed them as they used their phones and their apps as part of their regular activities
  - Day 1 and 2: collected usage data
  - Day 3: interaction with Privacy Assistant
- Starting on Day 4, participants were **subjected to nudges for an additional 6 days to see if they wanted to modify their settings**
- Total of 51 participants
  - **29 treatment condition – Privacy Assistant**
  - 22 control condition

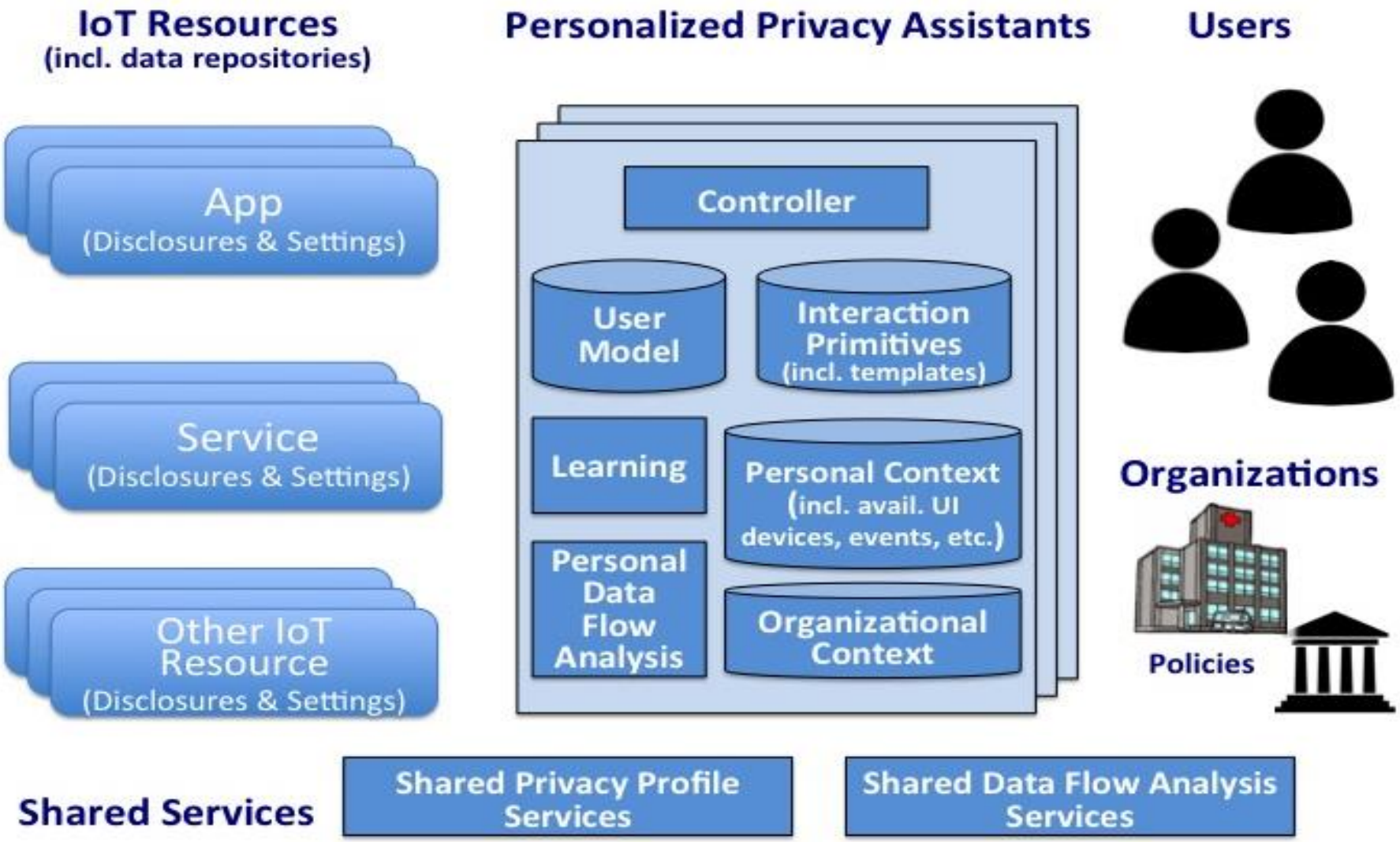
Breakdown by User



Results (Treatment condition)

- **Users accepted 78% of Privacy Assistant's recommendations**
  - Could probably do even better with larger training set & more personalized learning
- Users showed great engagement as they received nudges for 6 days following interaction with the recommendations
  - A number of settings not covered by the recommendations were modified
- **Only 5.1% of accepted recommendations were modified over the 6 days**

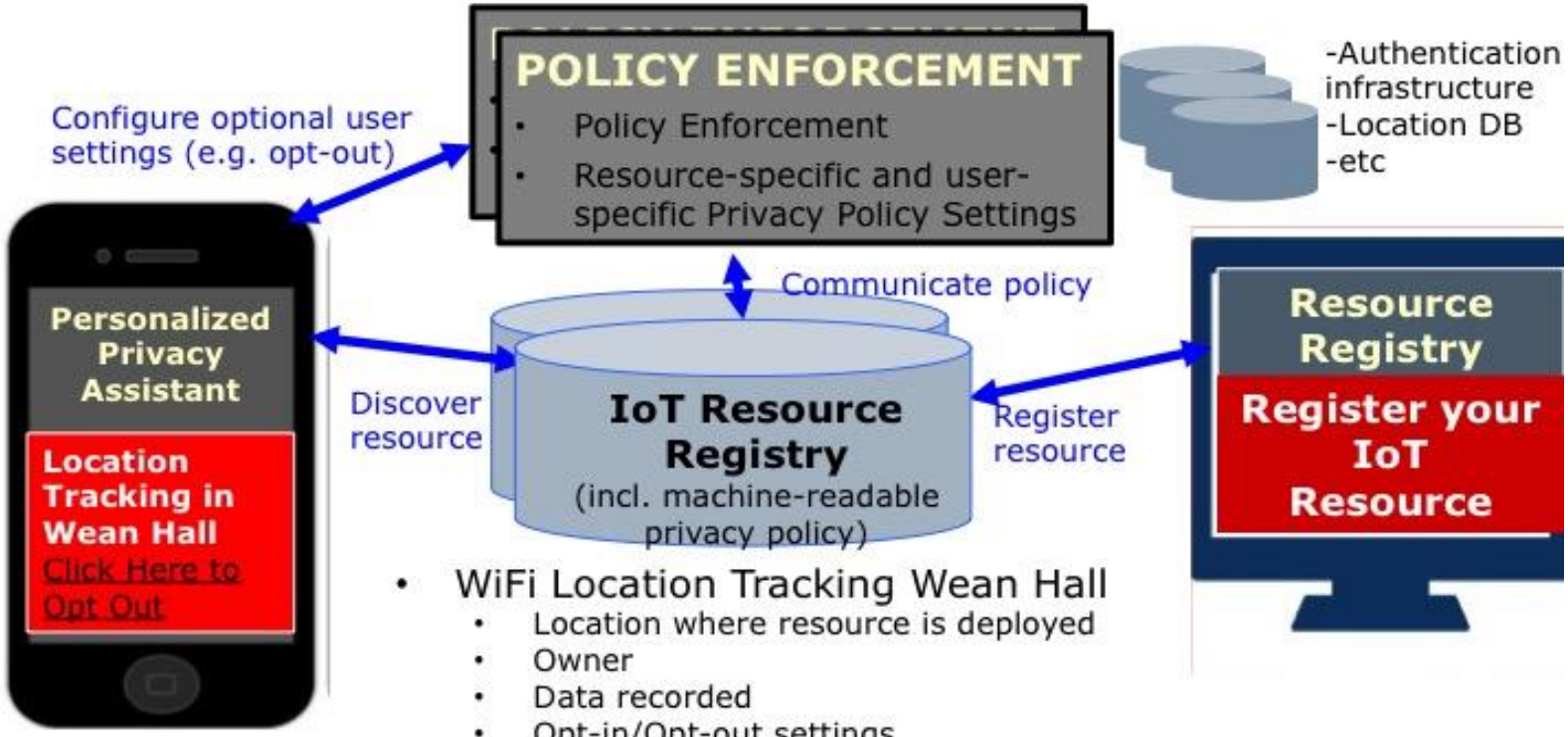
Extending this to IoT



Overall Vision: Personalized Privacy Assistants

- Help scale to interactions with a large number of apps and services
- Learn user preferences, learn models of what users already expect & what they want to be informed about & how to communicate with them (when, how often, how)
- Can **selectively enter in dialogues** with users and **nudge them** towards safer practices
- **Extend privacy profiles across many environments:** from your smartphone, to your browser, to your smart home to your social networking account, etc.

Privacy Assistant for IoT



- WiFi Location Tracking Wean Hall
  - Location where resource is deployed
  - Owner
  - Data recorded
  - Opt-in/Opt-out settings
  - etc.
- Camera Recording in Gates Center
  - Location where resource is deployed
  - Owner
  - Data recorded
  - Opt-in/Opt-out settings
  - etc.

Concluding Remarks - I

- “Notice and Choice” is the de facto approach to privacy on the Web
- Even on the fixed Web, this approach does not work
- On smartphones and with the emerging Internet of Things, this **framework (in its current form) simply does not scale**

Concluding Remarks - II

- Crowdsourcing, Machine Learning and Natural Language Processing offer the prospect of **semi-automatically annotate privacy policies to**:
  - Help users – through **succinct and personalized summaries**
  - Help **corporations** identify **potential compliance violations**
  - Help **regulators** understand **trends** and identify **potential violations**
  - Note: Presented results of fully automated analysis: current vision is to flag potential problems & rely on manual investigation
- **Compliance:** Lots of mobile apps seem to have compliance issues – needs manual verification
- **Learning people's privacy preferences** can be used to selectively inform users about what matters most to them and can also help them configure privacy settings
  - **Personalized Privacy Assistants successfully piloted**

Acknowledgements: Work funded by the National Science Foundation, DARPA and Google

The **Usable Privacy Policy Project** and the **Personalized Privacy Assistant Project both** involve a collaborations with a number of individuals.
See **usableprivacy.org** and **privacyassistant.org** for additional details

Q&A