

USABLE PRIVACY POLICY AND PERSONALIZED PRIVACY ASSISTANT PROJECTS

Mobile App Privacy Compliance

Sebastian Zimmeck & Norman Sadeh

Carnegie Mellon University

www.sebastianzimmeck.de

www.normsadeh.org



Copyright © 2017 Sebastian Zimmeck & Norman Sadeh

Session II: Motivation

- Mobile apps collect e-mail addresses, locations, and a variety of other information from users
- Research has shown that many apps fail to comply with basic privacy requirements
- This session will provide an overview of **automatically analyzing mobile apps** for potential privacy requirement inconsistencies
 - **Machine learning** to analyze privacy policy text
 - **Code analysis** to evaluate what apps actually seem to do



Automated Analysis of Privacy Requirements for Mobile Apps, Sebastian Zimmeck, Ziqi Wang, Lieyong Zou, Roger Iyengar, Bin Liu, Florian Schaub, Shomir Wilson, Norman Sadeh, Steven M. Bellovin, and Joel Reidenberg, 24th Network & Distributed System Security Symposium (NDSS), San Diego, CA, USA, February 2017

Session II: Motivation

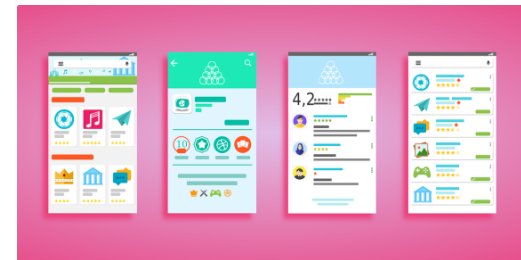
Snapchat's location data disclosures – an accidental find

- “We do not ask for, track, or access any location-specific information [...]”
- Snapchat's Android app transmitted Wi-Fi- and cell-based location data from users' devices to analytics service providers
- Accidental discovery by researcher who examined Snapchat's data deletion mechanism lead to involvement of Electronic Privacy Information Center and ultimately reached the Federal Trade Commission

Complaint In the Matter of Snapchat, Inc. (December 31, 2014).

Session II: Motivation

- **Help users, regulators, app stores, and software developers**
- Manual analysis of large amounts of apps with rapid update cycles is **challenging for regulators**
- **App stores want privacy requirement compliance** without imposing unrealistic requirements on app developers

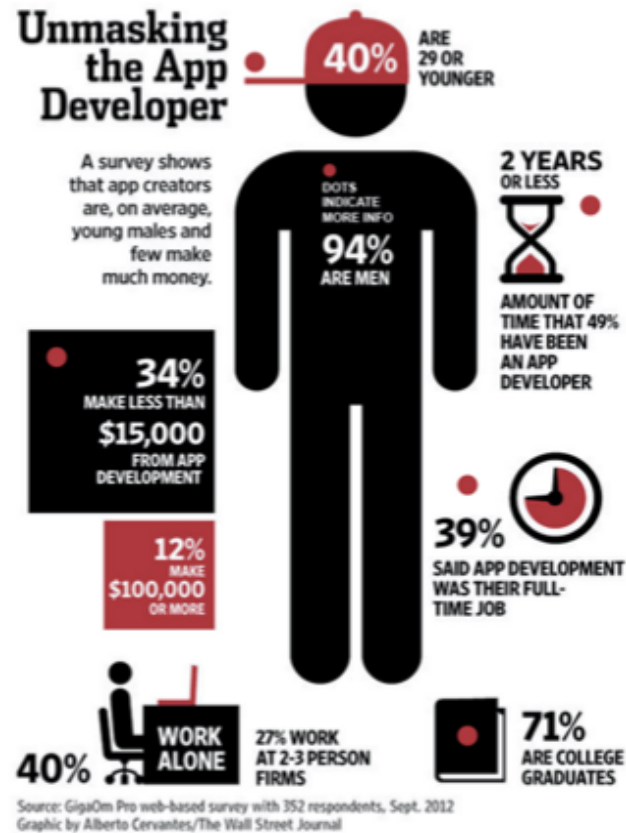


Session II: Outline

- Developers' and Regulators' Perspectives
- Challenges
- Customizable Privacy Requirements
- The Mobile App Compliance System
- Live Demo
- Preliminary Analysis Results
- Summary and Outlook



Developer's Perspective



Regulator's Perspective

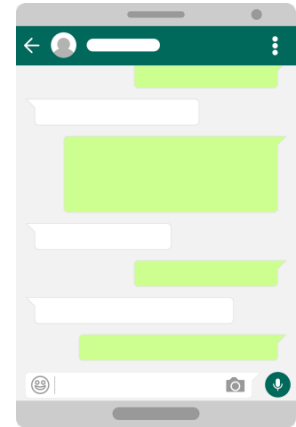
- **Regulatory sweeps are inefficient** (e.g., Global Privacy Enforcement Network: 26 agencies analyzed 1,211 apps in one week)
- **Regulators** are overwhelmed and lack the resources for systematic enforcement
 - Can we develop technology that helps regulators by automatically analyzing privacy policies and mobile app code at scale?
 - Identify and document potential violations
 - Support manual prioritization and vetting of flagged apps (e.g., children, education, car sharing, different jurisdictions, different interpretations and prioritization of compliance issues)



Big year for Global Privacy Enforcement Network: GPEN releases 2014 annual report, Global Privacy Enforcement Network, <https://www.privacyenforcement.net/node/513>

Challenges I

- **Apps' privacy practices are often opaque to users;** existing notifications, such as permissions, are coarse
- The functionality of **third party libraries inside apps** is often unclear to developers

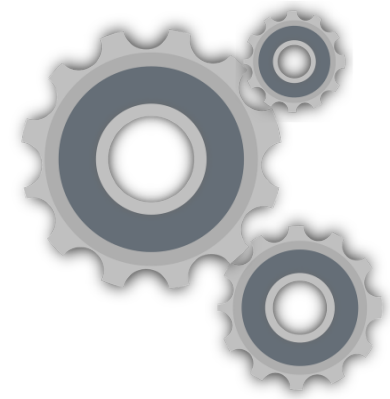


- Privacy policies are the instrument to **notify users of a service's privacy practices**; are they?
 - What is the **purpose of privacy policies**?
Notification of users? Hold services accountable?
 - What is the **legal nature of privacy policies**?

Challenges II



- Automating notice and choice is an **interdisciplinary task**: legal interpretation, security engineering, human computer interaction, ...
- Any automated system to provide legal analysis is necessarily based on **legal assumptions and interpretations**
- Make the system
 - **agnostic** (e.g., is location data PII?)
 - **modular** (app retrieval, app analysis, ...)
 - **fast** for large-scale analyses



Customizable Privacy Requirements

- **What are privacy requirements?**
 - Term of art: standards against which to measure data practices; do not measure directly against laws
 - Apps are subject to different laws (statutes, court decisions, regulations, self-regulations, ...)
- **Not every measured privacy requirement inconsistency is a privacy law violation**
 - Which laws are applicable/relevant?
 - Limited guidance on legal assumptions and how to interpret laws
 - Automated analysis is inherently subject to errors



Customizable Privacy Requirements

- The **California Online Privacy Protection Act** requires app publishers to have a privacy policy and transparently disclose data practices (California Business and Professions Code Sections (CalOPPA) 22575-22579)



- **Delaware Online Privacy Protection Act**
- **Children's Online Privacy Protection Act**
- **Federal Trade Commission actions (15 USC §45)**
- **State attorney general actions**

....

→ **Privacy policies**

Customizable Privacy Requirements

**How to identify a privacy policy?
Which policy is the right one?**

- Find policy links (store and/or app links)
- Detect whether a document is a policy
- Extract text from policies in various formats (html, pdf,)
- Compare documents to identify duplicate policies; merging policies
- Is the policy applicable to mobile apps?
- Keep track of policy changes over time





Customizable Privacy Requirements



- **How do we identify that an app does not have a policy?**
- Are **non-English policies** sufficient?
- What if a policy shows that the app developer or publisher is **not based in the United States?**
- What if an **app does not have a policy?**

Customizable Privacy Requirements

- Legal Interpretation: “The term ‘**personally identifiable information**’ [...] include[es] any of the following: [...] (6) An[] [...] **identifier that permits the physical or online contacting** of a specific individual. (7) Information concerning a user that the [...] online service collects online from the user and maintains in personally identifiable form **in combination with an identifier** described in this subdivision.” (CalOPPA 22577(a))
- Device identifiers? 
- Location data? 

Making your privacy practices public, State of California, Department of Justice, https://oag.ca.gov/sites/all/files/agweb/pdfs/cybersecurity/making_your_privacy_practices_public.pdf, Accessed, July 9, 2017

Customizable Privacy Requirements

- **What is considered “collection” and “sharing” of data?**
Is data ...
 - ... accessed on the device?
 - ... sent to the app developer or publisher?
 - ... sent to a third party?
 - ... stored? (If so, for how long?)
- **Secure transmission of sensitive data**
 - Which categories of data are sensitive?



Customizable Privacy Requirements

- There are many legal interpretations and possible assumptions
→ **a good system is one that lends itself to easy customization** (e.g., filtering results according to different criteria)
- A **customizable** baseline may assume that:
 - 1.If an app collects PII, it **must have a privacy policy**;
 - 2.A policy **has to describe data practices** occurring in the app it relates to (e.g., describe how location data is shared with third parties) and must **not omit any practice**; and
 - 3.An app **must follow the practices described in the privacy policy**
- Baseline can be made customizable using filters

Customizable Privacy Requirements

- **What qualifies as an inconsistency?**

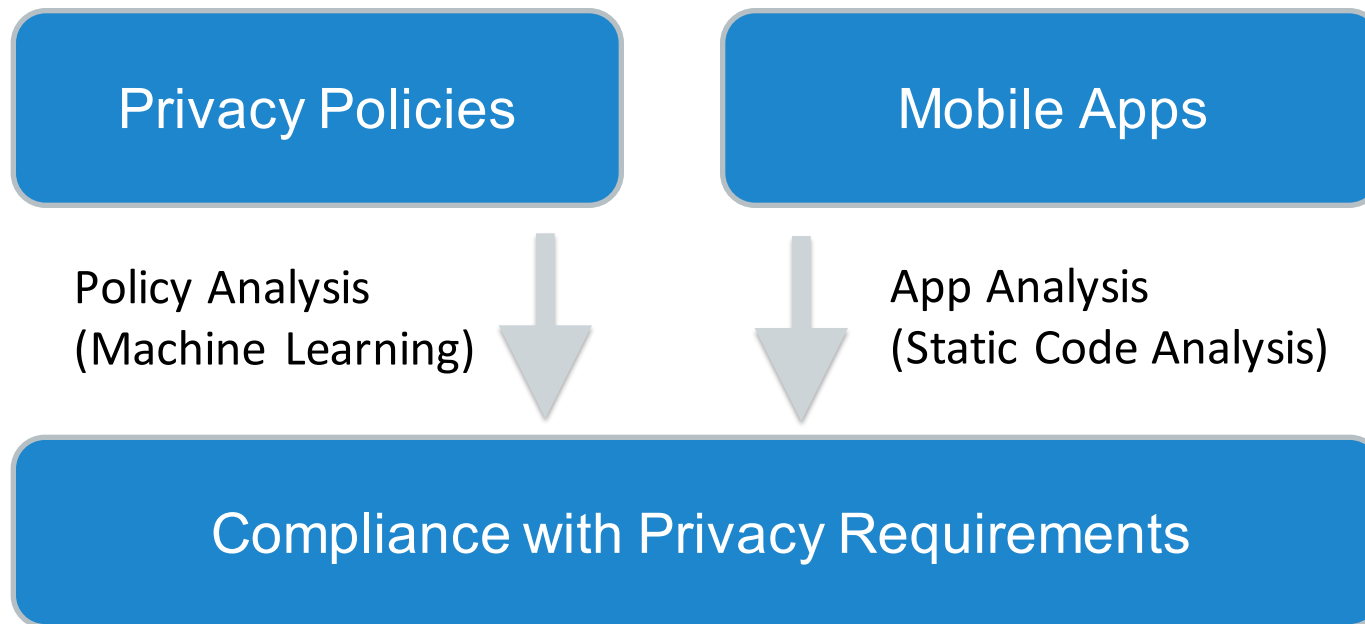
- Policy explicitly denies or omits a certain practice; **and**
- the app performs that practice



- **What if there are mistakes in the policy and/or app code analysis?**

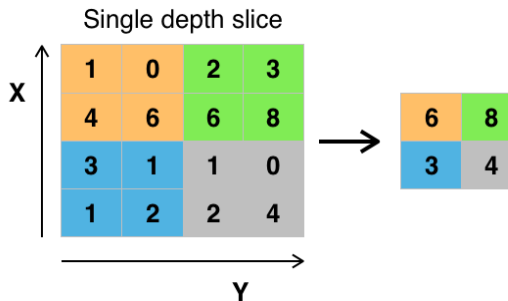
- Some matter (e.g., policy incorrectly classified as omitting a practice and app correctly identified as performing the practice)
- Others do not (e.g., policy correctly classified as disclosing a certain practice and app incorrectly identified as not performing the practice)
- Calculate F-1 and other scores based on whether a mistake matters

The Mobile App Compliance System



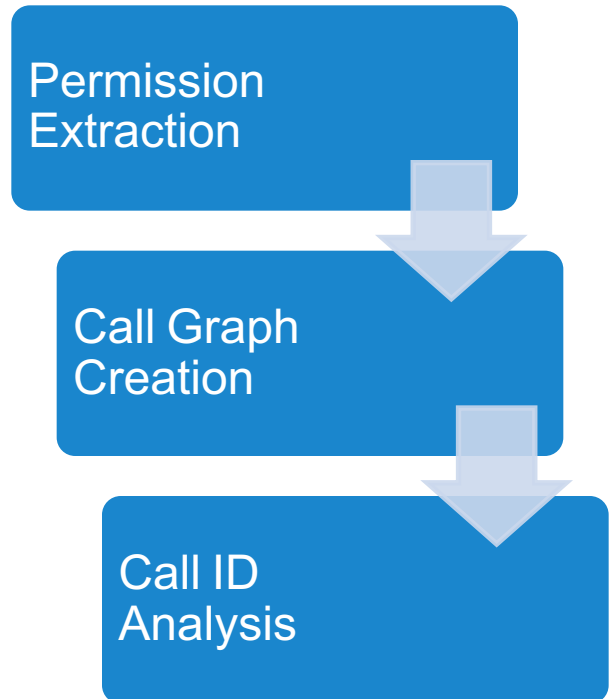
The Mobile App Compliance System

```
1 def location_feature_extraction(policy):
2
3     data_type_keywords = ['geo', 'gps']
4     action_keywords = ['share', 'partner']
5     relevant_sentences = ''
6     feature_vector = ''
7
8     for sentence in policy:
9         for keyword in data_type_keywords:
10            if (keyword in sentence):
11                relevant_sentences += sentence
12
13     words = tokenize(relevant_sentences)
14     bigrams = ngrams(words,2)
15
16     for bigram in bigrams:
17         for keyword in action_keywords:
18             if (keyword in bigram):
19                 feature_vector += bigram, bigram[0],
20                                     bigram[1]
21     return feature_vector
```



https://commons.wikimedia.org/wiki/File:Max_pooling.png, Accessed, July 9, 2017

Use logistic regression,
support vector machines,
or neural networks



Classifiers

Policy Analysis

App Analysis

The Mobile App Compliance System



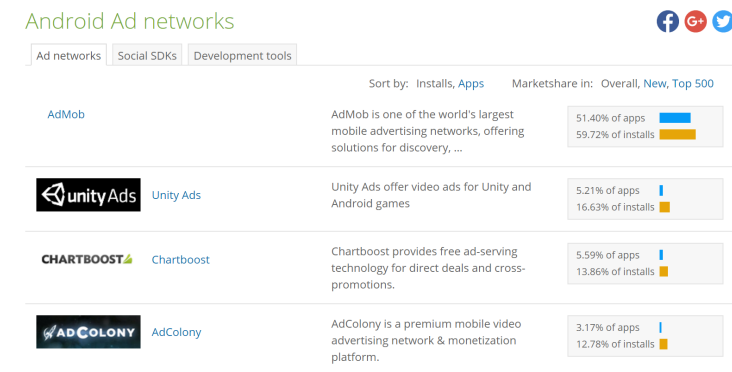
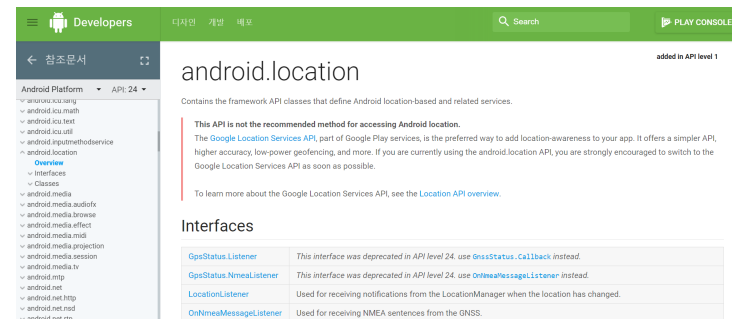
- **Supervised learning requires ground truth**
 - Policy annotations
 - Partly annotated dataset sufficient
 - Low agreement among annotators
 - Check for systematic disagreement

- **“We do not collect your GPS or other precise location information.”**
 - Scarcity of explicit negative instances

The Mobile App Compliance System

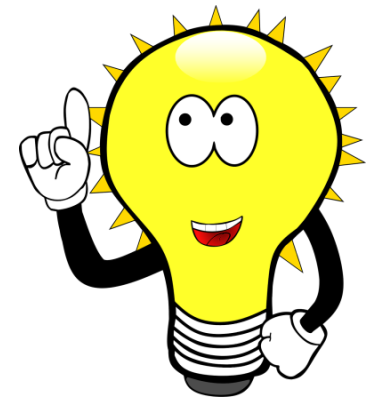
- **First party data practices**
 - Android system APIs
 - Android documentation

- **Third party data practices**
 - Libraries
 - AppBrain or other statistics



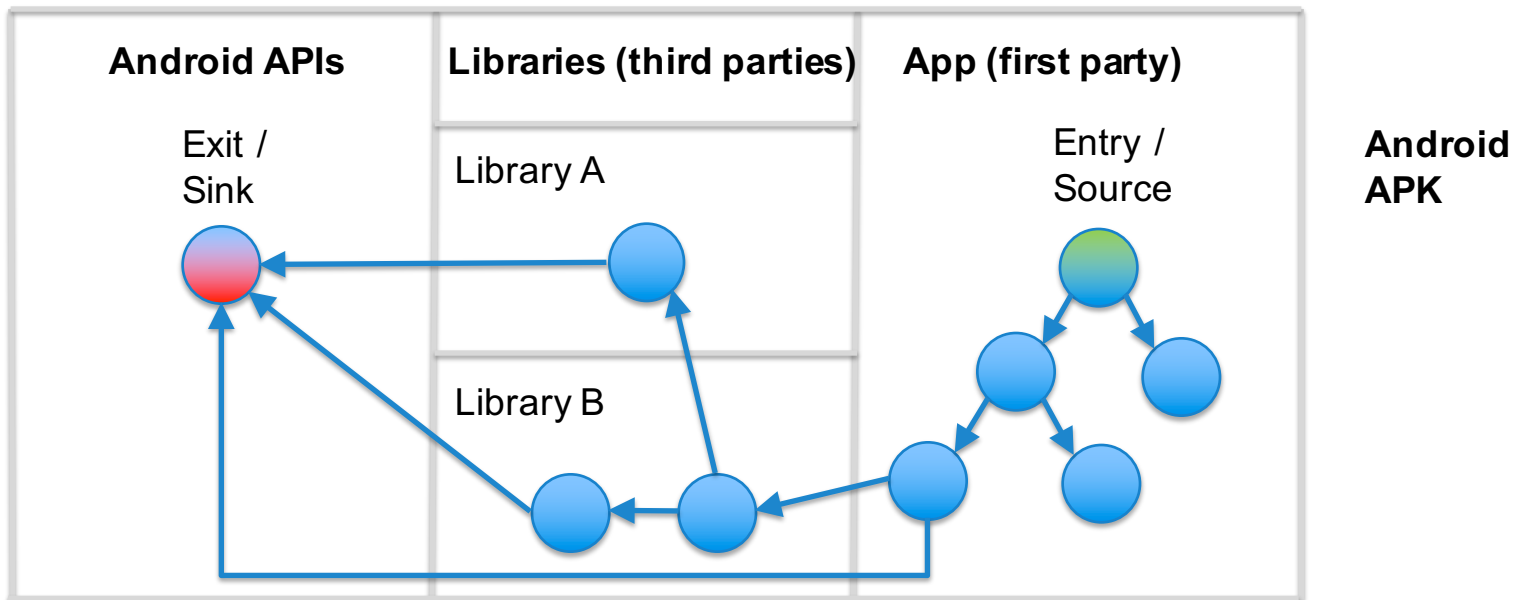
The Mobile App Compliance System

- **Static vs. dynamic code analysis**
 - Inconsistency w/ execution vs inconsistency w/ code
 - Scalability, detectability, accuracy
- **Some challenges for static code analysis**
 - Avoid downloading resource files
 - May need lots of accounts to download this many apps
 - Find characteristic APIs for privacy practices in question
 - Distinguish first party from third party code
 - Code obfuscation



The Mobile App Compliance System

Detecting first and/or third party functionality via static analysis



The Mobile App Compliance System

Example: SuperB Cleaner - Boost, Clean & APP LOCK (Hegogo)

- **Policy:**

“We do not collect Personal Information. [...] Personal Information includes your [...] email addresses, [...].


<http://www.enosisherмес.com/client/privacy-policy.html>, Accessed July 7, 2017

- **App:**

- `getAccountsByType`
- in `smali/com/hermes/superb/booster/feedback/b.smali`
- `INTERNET, GET_ACCOUNTS, MANAGE_ACCOUNTS, AUTHENTICATE_ACCOUNTS`

<https://play.google.com/store/apps/details?id=com.hermes.superb.booster&hl=en>, Accessed July 7, 2017



















Live Demo

 Reports **Analyses** Logout (root)

Reports

Analyses **Potential Compliance Issues** **Categories** **Developers** [Refresh](#) [more filters](#)

[Columns](#) [Export](#) Search:

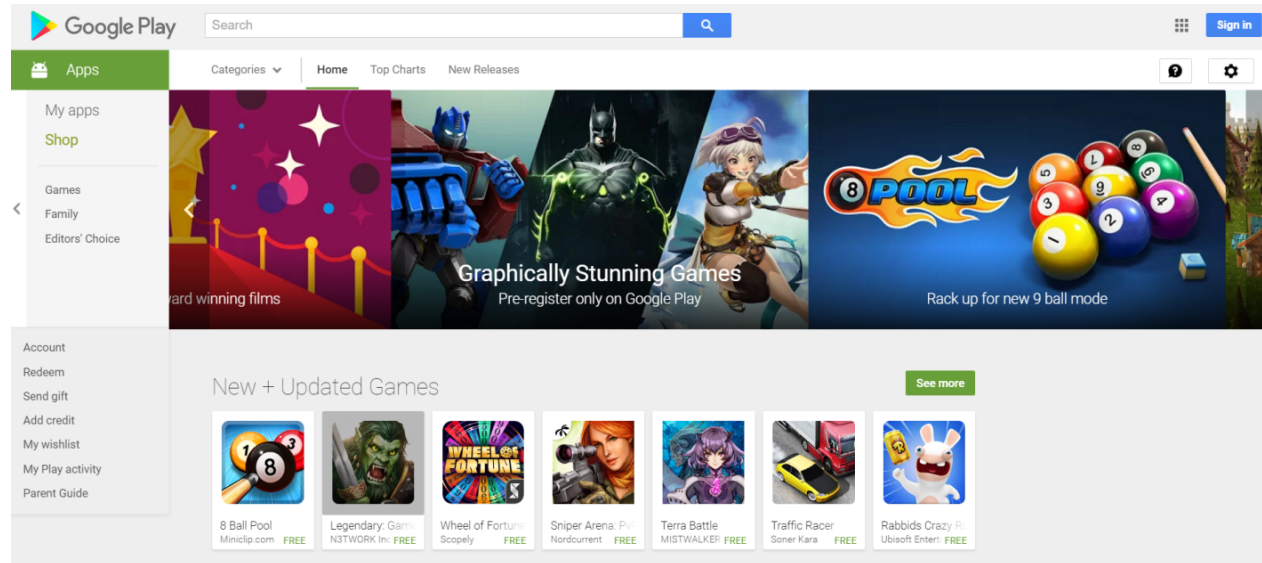
Application	Analysis Date	Static Analysis	Policy Analyses	Potential Compliance Issues
   360 Security -Free Antivirus,Booster,Space Cleaner	6/30/2017, 7:26:23 PM	27 practices performed	2 policies	2 potential issues
   3D Bowling	6/30/2017, 7:16:44 PM	3 practices performed	1 policy	0 potential issues
   8 Ball Pool	6/30/2017, 7:30:48 PM	11 practices performed	1 policy	0 potential issues
   aa	6/30/2017, 7:35:06 PM	1 practice performed	1 policy	0 potential issues
   AccuWeather	6/30/2017, 7:17:00 PM	14 practices performed	2 policies	2 potential issues
   Adobe Acrobat Reader	6/30/2017, 7:17:15 PM	4 practices performed	3 policies	2 potential issues

Preliminary Analysis Results

- Obviously, our system does not catch all potential privacy requirement inconsistencies, but ...



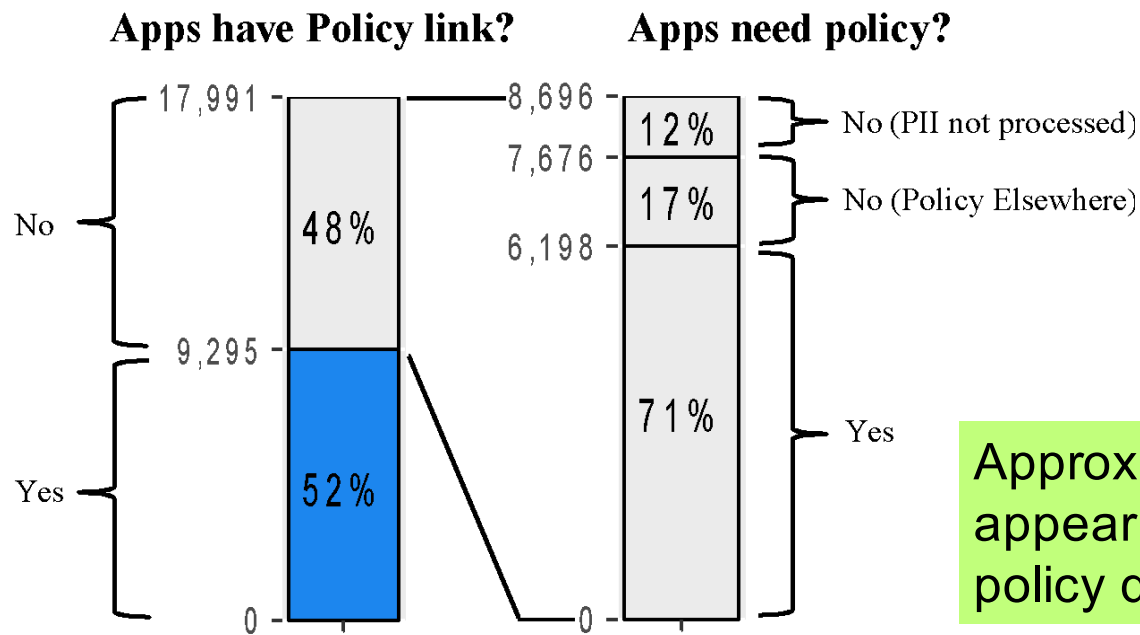
Preliminary Analysis Results



- 17,991 free apps from the Google Play Store and their metadata (e.g., whether an app has a policy link or the number of reviews)
- Started crawl from popular apps in each category and followed links to similar apps

Zimmeck et al, NDSS '17

Preliminary Analysis Results



Approximately 1/3 of apps appear to have no privacy policy despite processing PII

Potential Privacy Requirement Non-Compliance

Zimmeck et al, NDSS '17

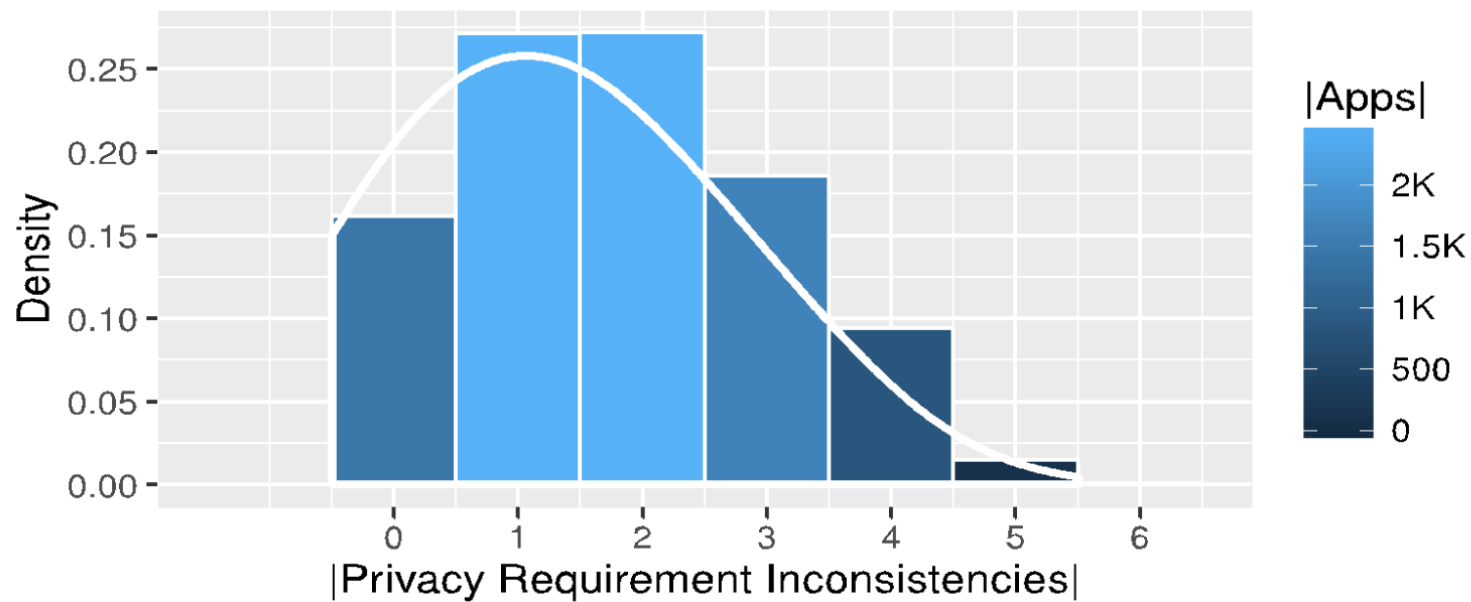
Preliminary Analysis Results

<i>Practice</i>	<i>Precision</i> (Test Set; n=40)	<i>Recall</i> (Test Set; n=40)	<i>F-1</i> (Test Set; n=40)	<i>% Potential Privacy Requirement Inconsistency (n=9K)</i>
Collection of Identifiers	0.75	1	0.86	50%
Sharing of Location	1	1	1	17%
Sharing of Contact	1	1	1	2%

→ Potential privacy requirement non-compliance is predictable reliably and at scale

Zimmeck et al, NDSS '17

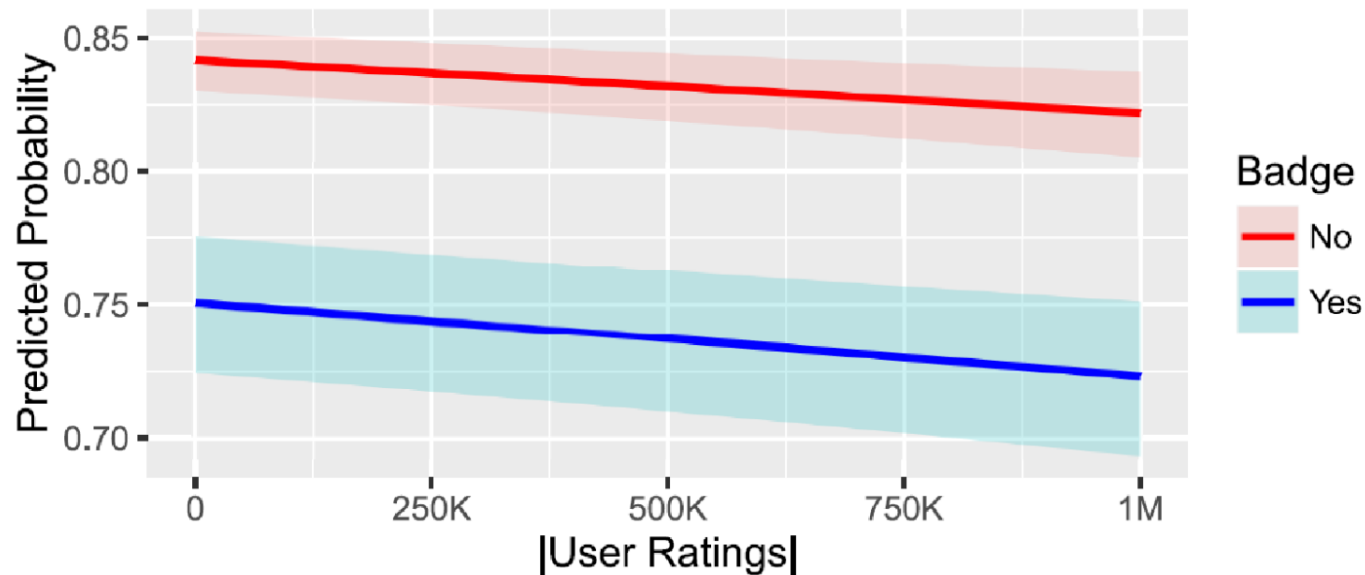
Preliminary Analysis Results



- Each app has mean of 1.83 instances of potential privacy requirement non-compliance
- Non-compliance does not necessarily mean that a law is violated; manual verification

Zimmeck et al, NDSS '17

Preliminary Analysis Results



→ Use app metadata to predict which app populations have increased probability of potential privacy requirement non-compliance

Zimmeck et al, NDSS '17

Summary and Outlook (I)

- Help developers, app store owners, and regulators check more systematically for potential mobile app compliance issues
- Current collaboration with the Office of the California Attorney General, Federal Trade Commission, and another regulator
- Additional collaborations with researchers and privacy non-profit organizations



Summary and Outlook (II)

The CalOPPA form is part of a multi-pronged approach to improve online privacy. The Attorney General's office is also partnering with the Usable Privacy Policy Project at Carnegie Mellon University to develop a tool that will identify mobile apps that may be in violation of CalOPPA. The tool is designed to look for discrepancies between disclosures in a given privacy policy and the mobile app's actual data collection and sharing practices (for example, a company might share personal information with third parties but doesn't disclose that in its privacy policies). This tool will help proactively identify and focus attention on policies that may require enforcement.

Excerpt from October 2016 press release from the Office of the California Attorney General

Attorney General Kamala D. Harris Launches New Tool to Help Consumers Report Violations of California Online Privacy Protection Act (CalOPPA), State of California, Department of Justice, <https://oag.ca.gov/news/press-releases/attorney-general-kamala-d-harris-launches-new-tool-help-consumers-report>, Accessed, July 9, 2017

Summary and Outlook (III)

- Prospective users: regulators, researchers, activists, developers, app stores, etc.
- Need to also identify a sustainable model for continuing to fund regular privacy sweeps



Summary and Outlook (IV)

Q & A



- The **Usable Privacy Policy Project** and the **Personalized Privacy Assistant Project** both involve collaborations with a number of organizations and individuals
- See **usableprivacy.org** and **[privacyassistant.org](https://www.privacyassistant.org)** for additional details incl. lists of collaborators, publications, sponsors and recent news
- Subscribe to our mailing lists to stay up to date:
<https://usableprivacy.org/contact> and
<https://www.privacyassistant.org/contact>

References

Automated Analysis of Privacy Requirements for Mobile Apps

Sebastian Zimmeck, Ziqi Wang, Lieyong Zou, Roger Iyengar, Bin Liu, Florian Schaub, Shomir Wilson, Norman Sadeh, Steven M. Bellovin, and Joel Reidenberg

24th Network & Distributed System Security Symposium (NDSS), San Diego, CA, USA, February 2017

Privee: An Architecture for Automatically Analyzing Web Privacy Policies

Sebastian Zimmeck and Steven M. Bellovin

23rd USENIX Security Symposium (USENIX Security), San Diego, CA, USA, August 2014

PrivOnto: A Semantic Framework for the Analysis of Privacy Policies

A. Oltramari, D. Piraviperumal, F. Schaub, S. Wilson, S. Cherivirala, T.B. Norton, N.C. Russell, P. Story, J. Reidenberg, N. Sadeh, Semantic Web Journal (SWJ), May 2017

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

N. Sadeh, A. Acquisti, T.D. Breaux, L.F. Cranor, A.M. McDonald, J. Reidenberg, N.A. Smith, F. Liu, N.C. Russell, F. Schaub, S. Wilson, J.T. Graves, P.G. Leon, R. Ramanath, A. Rao, Presentation at FTC PrivacyCon, Jan 2016

Automatic Extraction of Opt-Out Choices from Privacy Policies

K.M. Sathyendra, F. Schaub, S. Wilson, N. Sadeh

AAAI Fall Symposium on Privacy and Language Technologies, Nov 2016

References

The Creation and Analysis of a Website Privacy Policy Corpus

S. Wilson, F. Schaub, A. Dara, F. Liu, S. Cherivirala, P.G. Leon, M.S. Andersen, S. Zimmeck, K. Sathyendra, N.C. Russell, T.B. Norton, E. Hovy, J.R. Reidenberg, N. Sadeh
ACL '16: Annual Meeting of the Association for Computational Linguistics, Aug 2016

Attorney General Kamala D. Harris Launches New Tool to Help Consumers Report Violations of California Online Privacy Protection Act (CalOPPA)

State of California, Department of Justice

<https://oag.ca.gov/news/press-releases/attorney-general-kamala-d-harris-launches-new-tool-help-consumers-report>, Accessed, July 9, 2017

Big year for Global Privacy Enforcement Network: GPEN releases 2014 annual report

Global Privacy Enforcement Network

<https://www.privacyenforcement.net/node/513>, Accessed, July 9, 2017

Making your privacy practices public

State of California, Department of Justice

https://oag.ca.gov/sites/all/files/agweb/pdfs/cybersecurity/making_your_privacy_practices_public.pdf, Accessed, July 9, 2017

Complaint In the Matter of Snapchat, Inc. (December 31, 2014)