Privacy in the Age of IoT
Technologies to Help Users and Regulators

Norman Sadeh
Carnegie Mellon University

www.normsadeh.org
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

The Usable Privacy Policy Project

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

Crowdsourcing and automatic analysis  
Personalization and presentation

Privacy policy → Policy annotations → Salient policy details

www.usableprivacy.org

Crowdworkers Can Be Good at This(!)

Multi-step annotations

1. Segment policy
2. Identify practice categories in each segment
3. Category-specific annotation tasks & questions
**Expert Annotation Tool**


First Party Collection/Use
- User Access, Edit and Deletion
- Data Retention
- Data Security
- Policy Change
- Do Not Track
- International and Specific Audiences

4/29

Information You Give Us: We receive and store the personal 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

Third Party Sharing/Collection
- User Choice/Control
- User Access, Edit and Deletion
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
### Privacy Policy

**Playstation Privacy Policy** from Apr 1, 2011.  
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**

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 disclosure under 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

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

- Automated Code analysis evaluated against manual analysis for 30 mobile apps
- F-1: F score measures accuracy and recall
- F-1\text{pos}: 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

- 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
**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

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

DID YOU KNOW?

Your Location Data has been accessed 1222 times over the past week by:

- Snapchat (266 times)
- Facebook (144 times)
- Yelp (50 times)

...and 13 more apps.

Some of these apps use your Location for:

- Targeted Advertising
- Consumer Tracking & Profiling

GO TO MY SETTINGS

KEEP CURRENT SETTINGS

REMINDE ME IN AN HOUR
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)  
Deny

News & Weather (0 times)  
Deny

Contacts+ (28 times)  
Deny

Messenger (16 times)  
Allow

Snapchat (84 times)  
Deny

Why deny? This Social app accesses your Location for App Functionality and Consumer Tracking & Profiling.

QR Code Reader (0 times)  
Deny

Skype (0 times)  
Deny

Do you want to make these changes?

**YES, DENY THE 8 APP(S) SELECTED**

**NO, DO NOT MAKE ANY CHANGES**
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

**POLICY ENFORCEMENT**
- Policy Enforcement
- Resource-specific and user-specific Privacy Policy Settings

**IoT Resource Registry**
(incl. machine-readable privacy policy)
- WiFi Location Tracking in 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.

Configure optional user settings (e.g. opt-out)

- Authentication infrastructure
- Location DB
- 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