Translating Code to Privacy Statements

Vijayanta Jain
University of Maine
When was the last time you read a privacy policy of any of your mobile applications?
Application’s Privacy Behavior ≠ Privacy Policy
71% of Android apps have 1.83 inconsistencies on an average (Zimmeck et al. 2016)

341 violations among 477 top applications (Slavin et al. 2016)

10.5% of 68,501 share information without informing users (Okoyomon et al. 2019)

At least 19% of 5,855 of children’s apps are in violation with COPPA (Reyes et al. 2018)
Related Work

**MAPS** (Zimmeck et al. 2019)
- Identify inconsistency
- Helps regulators
- Does not resolve inconsistency
- Does not help developers

**AutoPPG** (Yu et al. 2015)
- Generate Privacy Policies
- Helps developers
- Uses a template
- Does not provide reason

**CLAP** (Liu et al. 2018)
- Permission Usage Description
- Helps Developers
- Not proper notices
- Analyzes Application Description
Our Solution

An end-to-end framework that identifies code segments processing personal information and translate it to privacy statements
Methodology

Phase One: Preparing Android Applications Dataset
- Comprehensive Mapping Algorithms & Static Analysis Tools
- NN Classifiers, Language Models and Tools
- Identification and Extraction of PICU Code Segments

Phase Two: Preparing Privacy Policies Dataset
- Privacy Policies
- HITL System, Annotation Tool, Classifiers & Language Models
- Extraction and Cleaning of Privacy Policies
- NN Classifiers & Tools for Sensitive/ non-Sensitive
- Multi-Labelled (ML) Privacy Statements

Phase Three: Mapping Code to Privacy Statements
- ML Privacy Statements
- Code-to-Statement Mapping Algorithms and Tools
- Privacy Generator Recommender System
- NMT Models, Algorithms and Tools for Translation

Phase Four: Empirical User Evaluation
- Dataset, Language Models, NMT Classifiers
- Pilot Study With Students
- User Study with Software Developers
- Modification of Recommender System
Contributions

- Novel approach to resolve inconsistency
- End-to-end framework for developers
- Translation dataset
- Software tools
- Language models
REFERENCES

Thank You!