

# Towards Privacy-Preserving Mobile Apps: A Balancing Act

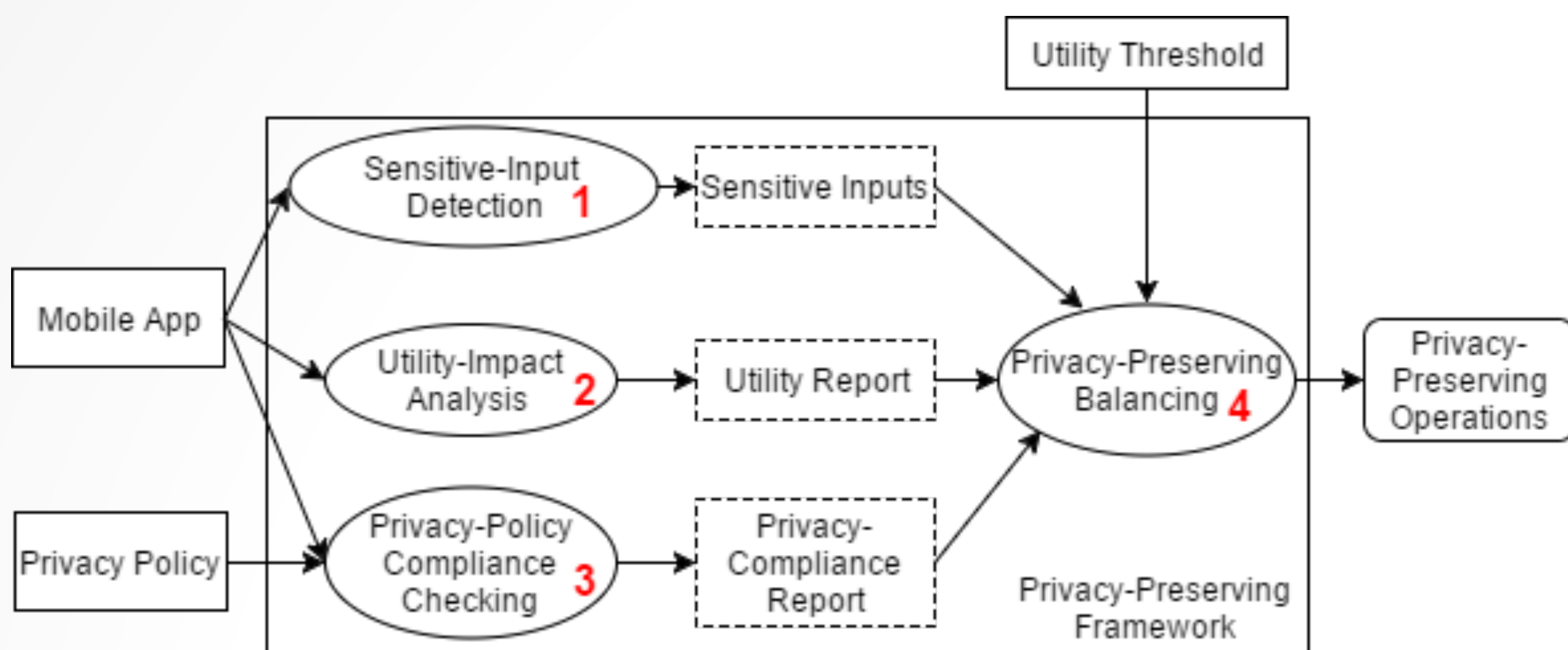
Dengfeng Li<sup>1</sup>, Wing Lam<sup>1</sup>, Wei Yang<sup>1</sup>, Zhengkai Wu<sup>1</sup>, Xusheng Xiao<sup>2</sup>, Tao Xie<sup>1</sup>

<sup>1</sup>(University of Illinois at Urbana-Champaign, email: taoxie@illinois.edu)

<sup>2</sup>(Case Western Reserve University, email: xusheng.xiao@case.edu)

## Objective

- Maximize utilities while minimizing the amount of sensitive information exposed to protect users' app usage data



Proposed framework

## 1. Sensitive-Input Detection

- Leverage UI rendering, geometrical layout analysis, and NLP to identify sensitive input fields
- Leverages static data flow analysis to detect sensitive information (such as a GPS location) obtained from the system

## 2. Utility-Impact Analysis

- Anonymize each input, and measure its impact on the utilities of an app and produce an utility report
- Provide measurement to show how each input contributes to an app's utilities

## 4. Privacy-Preserving Balancing

- Anonymize various sensitive information while assuring that the level of utility efficacy is above a user-predefined threshold

## Motivation

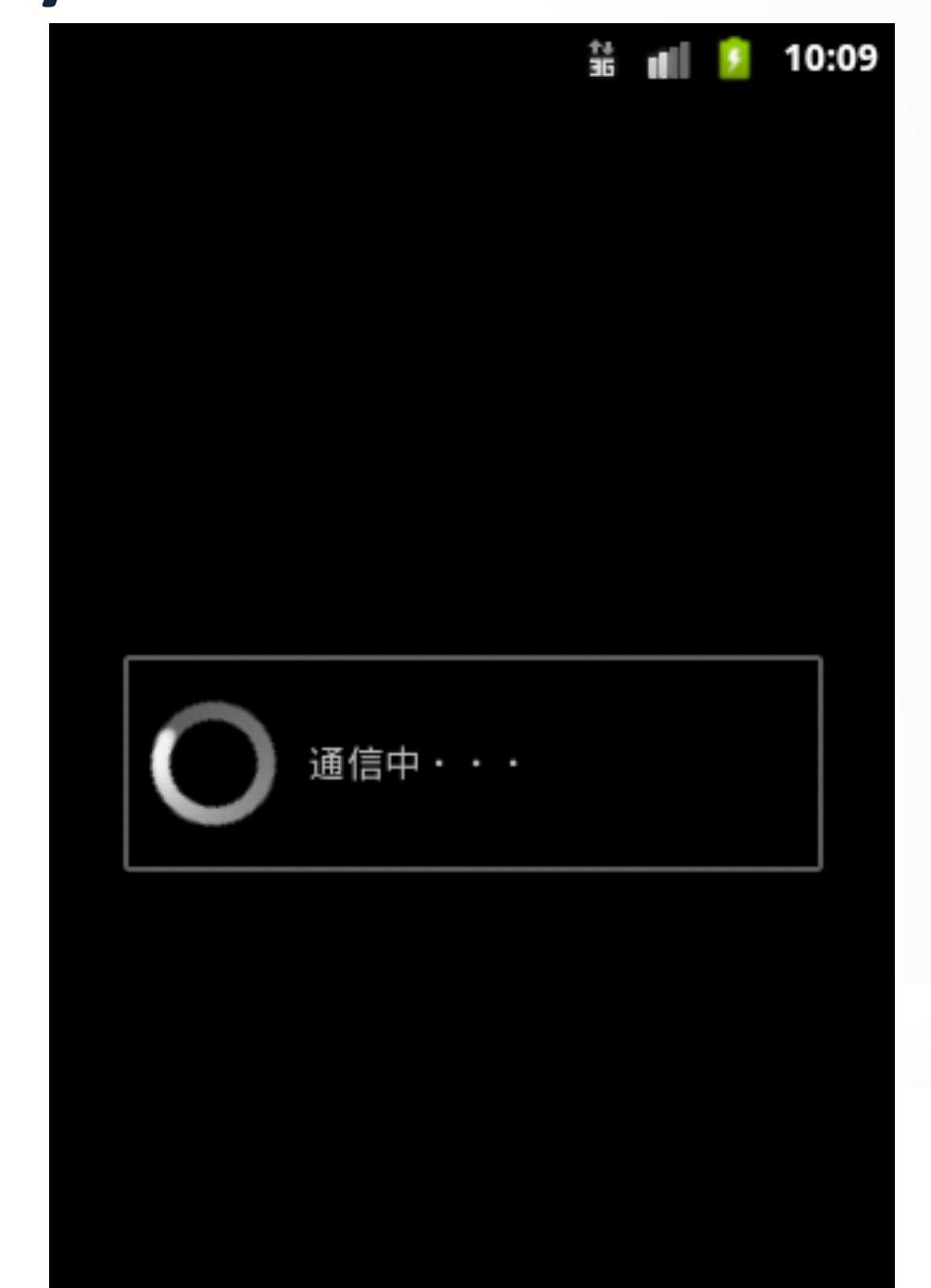
- Collecting some highly sensitive information provides little or no benefit towards delivering an app's utilities
- Existing techniques lack customized solutions to preserve user privacy at different levels while delivering user-desirable level of utility efficacy (e.g., the number of enabled features)

Example – App displays videos only if some sensitive information is previously sent to a remote server [1]



```
Content-Type: application/x-www-form-urlencoded
Host: d[redacted].jp
Connection: Keep-Alive
User-Agent: Apache-HttpClient/UNAVAILABLE (java 1.4)
```

```
id=12345&tel=15555215554&data=name%3AContact1%09tel%3A1111111111%09email%3Acontact1%40domain.com%0Aname%3AContact2%09tel%3A222-222-2222%09email%3Acontact2%40domain.com%0AHTTP/1.1 200 OK
```



## 3. Privacy-Policy Compliance Checking

- Check whether the sensitive information collected by an app is privacy preserving against the declared privacy policy
- Conduct static data flow analysis on the app and its backend server to generate a **usage summary** of the obtained sensitive information
- Leverage NLP to annotate declared privacy policy to extract **key features** related to sensitive-information usage
- Check generated **usage summary** with extracted **key features** for inconsistencies

[1] Android Malware Promises Video While Stealing Contacts: <https://securingtomorrow.mcafee.com/mcafee-labs/android-malware-promises-video-while-stealing-contacts/>

This material is based upon work supported by the Maryland Procurement Office under Contract No. H98230-14-C-0141

