A Study of Smartphone User Privacy from the Advertiser's Perspective

Jie Yang

Department of Computer Science
Florida State University

Motivation

- "Free" apps are not entirely free: users pay the price of their privacy.
- The consequences of privacy leakages, especially when an advertiser gathers such private data across many apps, are not well studied.
- Deriving social and community information by using private data from multiple apps across users is possible, and needs more attention.

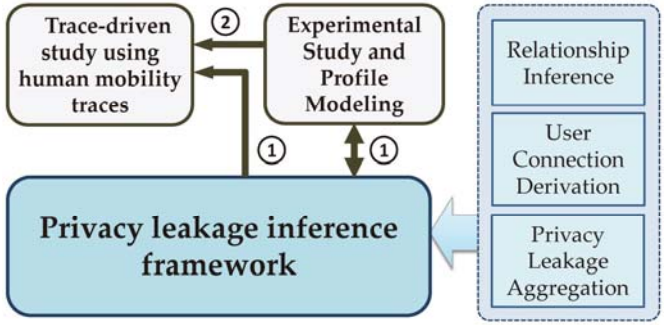
Contributions

- Modeling the *relationship inference process* in a three-layer framework by using the concept of *connection*, which is exemplified by two users sharing *similar patterns in their leaked data*.
- Conducting experiments with 10 participants and their families for over *one-month time period* to study the privacy leakages on using smartphones in their *daily lives*.
- Proposing an *Activeness Based Profile* as users' temporal privacy leakage profiles based on the experimental study.
- Verifying the generality of our findings from the real experiments based on trace-driven study using *human mobility traces*.

Concept of Connections

- The connection bridges the gap between the privacy leakage information and the users' relationship inference.
- Definition: a *connection* between two users exists if the same type of privacy leakage from the two users share certain spatial, temporal or content similarities.
  - *Contact list*: share common contacts or in each other's contacts.
  - *Wi-Fi Access Point list*: share common access points.
  - *GPS/network-based location*: leaked GPS or network-based locations are close.

Study the Consequences of Privacy Leakages



- Two types of social relationship:
  - Fact-based Relationship: carry similar, regular and repetitive spatial-temporal connection patterns as dictated by the relationship (e.g., colleagues, classmates, roommates, and families).
  - Intelligence-based Relationship: does not necessarily carry regular patterns (e.g., friends).
- Utilizing the *temporal and spatial patterns of the connections* to classify the type of relationship, for example:
  - The connections of colleagues occur in work hours of weekdays.
  - The connections of families in early morning and late night.
  - The connections of friends after working time and in weekends.

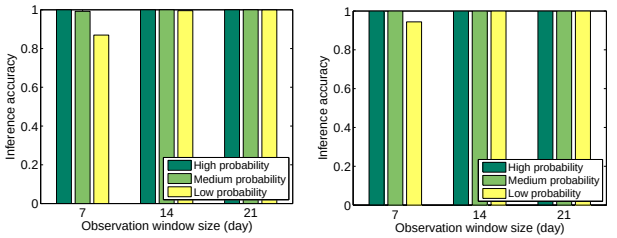
Real Experimental Study

- Involves 10 volunteer students and their family members over one month period including *five types relationships* (i.e., colleague, collaborator, classmate, friend, and family).
- Developed a tool to capture the privacy leakage *information in real-time* leveraging TaintDroid.
- Participants use their experimental smartphones *at least three times a day* with *no restriction* of how and when to use apps.

Activeness Based Profile

- Generating Profiles: derive the *probability* of each type of leakages happening in particular time windows for every participant.
- Categorize profiles with three representative user categories based on the hours that the particular user has leakages (i.e., *active user*, *regular user*, and *inactive user*).

Simulation Results



- Applying the activeness based profiles to the Foursquare trace.
- Observations:
  - An advertiser can achieve over 80% inference accuracy for both fact-based and intelligence-based relationship.
  - The inference accuracy decreases for less active users, and longer observation windows help improve the accuracy.

Conclusions

- This work serves as the *first step* towards a comprehensive understanding of the advertiser's perspective.
- We seek to discover what an advertiser can infer about users' social relationships by combining different private data from many apps.
- We propose a privacy leakage inference framework that describes a general method for inferring users' social relationships, which can achieve high accuracy.