Anonymity in the Personalized Web

Nuno Santos, Alan Mislove†, Marcel Dischinger, and Krishna P. Gummadi
MPI-SWS, †also Rice University

1. Problem
- Popular web sites are increasingly personalizing services
  - E.g., search engines, recommendation systems, social networks
- They are collecting huge amounts of data about users
  - E.g., search queries, browsing histories, and IP addresses
- To infer the likes and dislikes of individual users from the data
- The data collection raises severe privacy concerns
  - Many reported incidents of privacy violations by widely used sites
- Our goal: Alleviate user privacy concerns, while retaining the experience of personalized services

2. Challenge
- To personalize services, sites need info on user interests
- But, users don’t want to disclose too much personal data
- Can we prevent sites from characterizing individual users without adversely affecting user experience?
- Existing approaches to preserve anonymity degrade service personalization
  - E.g., TrackMeNot pollutes user profiles, Tor and Scroogle anonymize the source of requests

3. Anonymity-Preserving Personalization

Basic Idea
Leverage collaborations between groups of users with similar interests to obfuscate and anonymize user profiles

System Architecture
- Users communicate with Web sites via personal Privacy-Boxes (P-Boxes)
- P-Boxes act like client-side Web-proxies

- Provides k-anonymity: Anonymized profiles do not reflect service requests of individual users
- Results are still personalized because the anonymized profiles reflect the shared interests of group members

4. Current Status
- We are working on a prototype implementation
- Some open questions and unresolved issues:
  - How can users discover friends with similar tastes? Can we leverage user links in the existing online social networks for this purpose?
  - Should users participate in multiple interest groups and direct requests to specific groups? Or would one group be sufficient for all requests?
  - When users forward data from other users in a group, can we guarantee anonymity within the group members?
  - How can we evaluate the quality of personalized services? How effective are P-Boxes at preserving the quality of personalized services?