Developing Effective PET Descriptions for Ad Tracking & Analytics

Lu Xian, Song Mi Lee-Kan, Jane Im, Florian Schaub | University of Michigan School of Information

Motivation

How can we make Privacy Enhancing Technologies (PETs) comprehensible to users without a technical background?

PETs being adopted in ad tracking and analytics:
- Local differential privacy
- Federated learning
- Google's Topics

Research Questions

1. Does the **process**- and **implications**-focused approach to PET description work in the ad tracking and analytics context?
2. Which aspects of the descriptions do users understand accurately and inaccurately?

Study Design

- Developed 10 PETs descriptions: **process** vs. **process + [implications]**
- Online survey experiment of 10 conditions; randomly assigned 356 participants
- Context: a hypothetical social media platform

Key Insights

- Avoid noise, machine learning, and other jargon
- Help users pinpoint the source of privacy protection
- Provide more specificity about user data

QR code to poster abstract, survey instrument, PETs descriptions, and more results

Author contact information: Lu Xian, xianl@umich.edu

Our PET Descriptions (Examples)

Process

[Implications] +

Local differential privacy

To protect your information, the organization adds noise to your behavioral data (e.g., your interaction with the platform and with other apps/websites) before being sent to the organization for targeting ads. This means that your data is randomly modified, so that some of your actual data is used whereas some of it is random and not representative of your behavior. Your exact behavioral data is never sent to the organization, instead a subset of your noisy data is randomly selected and sent. The organization can still infer patterns from the noisy data across a large number of users.

Highlight data modification process

Underscore the sharing of modified user data

Federated learning

To protect your information, the organization uses machine learning on your device to infer interests from your behavioral data (e.g., your interaction with the platform and with other apps/websites) for targeting ads. Noise will be added to your behavioral data so that it is randomly modified before being used for training a machine learning model representing your inferred interests. This means that, for training the model, some of your actual data is used whereas some of it is random and not representative of your behavior. Your exact behavioral data is never sent to the organization and only the model representing your inferred interests will be sent. Then, to infer patterns across a large number of users, your model is merged with other users' models.

Highlight on-device training

Underscore the sharing & merging of models

Underscore the sharing of some of inferred topics

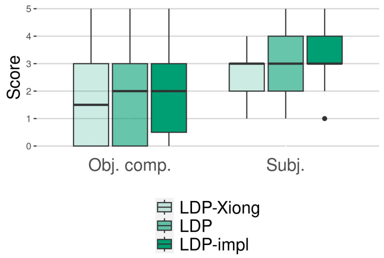
Google's Topics

To protect your information, the organization uses machine learning on your device to infer interests from your behavioral data (e.g., your interaction with the platform and with other apps/websites) for targeting ads. This means that the technology records inferred topics you may be interested in from your behavioral data only on your device. Your exact behavioral data is never sent to the organization, instead from your top topics of the last week, a small number are randomly selected and sent; there is also a small chance a random topic will be selected instead of one of yours.

[This way, the organization still learns aggregated interests across its users but not your exact behavior, which protects your privacy against the organization's employees or if the organization's database is compromised.]

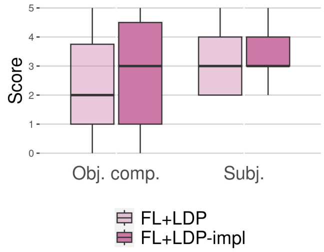
Key Findings

Local differential privacy



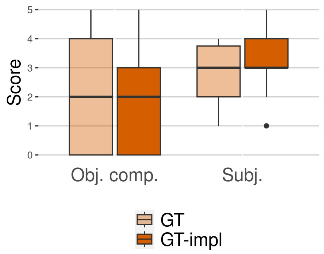
- Process+implications LDP description works well in ad tracking/analytics
- LDP implications statement has limited effect
- `Noise' jargon is confusing

Federated learning



- Process+implications approach can be adapted to federated learning
- Implications work better
- Mixed understanding of machine learning & model merging

Google's Topics



- Process+implications approach can be adapted to Google's Topics
- Adding implications doesn't hurt
- Mixed understanding of on-device data processing & data sharing

*Our descriptions are inspired by Xiong et al.'s (IEEE S&P 2020) approach of combining PET process and its privacy implications