A Programming Language for Data Privacy with Accuracy Estimations

Speaker: Elisabet Lobo Vesga
Chalmers University of Technology

Abstract

Differential privacy offers a formal framework for reasoning about privacy and accuracy of computations on private data. It also offers a rich set of building blocks for constructing private data analyses. When carefully calibrated, these analyses simultaneously guarantee the privacy of the individuals contributing their data, and the accuracy of the data analysis results, inferring useful properties about the population. The compositional nature of differential privacy has motivated the design and implementation of several programming languages aimed at helping a data analyst in programming differentially private analyses. However, most of the programming languages for differential privacy proposed so far provide support for reasoning about privacy but not for reasoning about the accuracy of data analyses. To overcome this limitation, in this work we present DPella, a programming framework providing data analysts with support for reasoning about privacy, accuracy and their trade-offs. The distinguishing feature of DPella is a novel component that statically tracks the accuracy of different data analyses. In order to make tighter accuracy estimations, this component leverages taint analysis for automatically inferring *statistical independence* of the different noise quantities added for guaranteeing privacy. We evaluate our approach by implementing several classical queries from the literature and showing how data analysts can figure out the best manner to calibrate privacy to meet the accuracy requirements.

Keywords: accuracy, concentration bounds, differential privacy, functional programming, databases, haskell