Bachelor Project on Local Differential Privacy

Johan Ringmann Fagerberg (jofag17) - 1996-09-24

Motivation

Differential privacy is a recently proposed measure of the level of privacy provided by random functions. It provides a formal measure for how much information a potential attacker can achieve from a perturbed data set, when the data perturbation consists of random noise.

In global differential privacy, a global aggregator has access to the real data prior to perturbing it. This differs from local differential privacy, where users perturb their data before sending it to the aggregator.

The goal of this bachelor project is to gain an understanding of differential privacy in general, with a particular focus on local differential privacy, and attempt to use that to produce an architecture for a trust-less and privacy-respecting web analytics solution.

Plan

The project will consist in 3 phases ad detailed below

- Survey of the state of the art (tentatively February-March). The survey will start by consulting [Des19], i.e., a reading list of key papers in differential privacy presented as a semi-technical blog series on differential privacy. Moreover, we will also read [Yan+20], i.e., a recent survey on local differential privacy methods for various use cases.
- Development and analysis of a novel locally differentially private analytics solution (tentatively March-April). A web analytics system à la [Akk+12] will be developed that uses local differential privacy to provide privacy guarantees for users. The analysis of our system will primarily focus on the strength of the privacy guarantee provided by our system, as well as its impact on queries that are classically of interest in web analytics.
- Development of a prototype (tentatively end of March). A prototype of our system will be implemented. The prototype will be tested to evaluate computational overhead on both client and server, as well as bandwidth overhead in terms of communicating the perturbed data from client to server.
- Report writing (tentatively April-May). The final report will be written following the academic conventions.

Risks

There is a possibility that it would be impossible to generate an actually useful trust-less web analytics solution (e.g. due to increased noise from local differential privacy).

Outcomes

At the end of the project, the outcomes of the project will be

- A report written in English and following the standard academic writing conventions. The report will contain the description of the state-of-the-art, the description of the prototype and its preliminary evaluation.
- The code of the prototype and the data for the experiments performed

References

[Akk+12] Istemi Ekin Akkus et al. "Non-tracking web analytics". In: *Proceedings of the 2012 ACM conference on Computer and communications security*. CCS '12. Raleigh, North Carolina, USA: Association for Computing Machinery, Oct. 2012, pp. 687–698. ISBN: 9781450316514. DOI: 10.1145/2382196.2382268. URL: https://doi.org/10.1145/2382196.2382268 (visited on 11/25/2020).

[Des19] Damien Desfontaines. *A reading list on differential privacy*. 2019. URL: https://desfontain.es/privacy/differential-privacy-reading-list.html (visited on 11/25/2020).

[Yan+20] Mengmeng Yang et al. "Local Differential Privacy and Its Applications: A Comprehensive Survey". In: *arXiv:2008.03686 [cs]* (Aug. 2020). arXiv: 2008.03686. URL: http://arxiv.org/abs/2008.03686 (visited on 11/25/2020).