Incentivized Data Sharing via Group-level Privacy-preservation

Stefano Bennati, Evangelos Pournaras (sbennati/epournaras@ethz.ch)

Professorship of Computational Social Science, ETH Zurich, Switzerland

keywords: privacy; Internet of Things; network; grouping; aggregation.

1 Extended Abstract

Big data collection practices are often privacy-intrusive and result in surveillance, profiling, and discriminatory actions over citizens [Medaglia and Serbanati, 2010]. Nonetheless, real-time data analytics and aggregate information open up tremendous opportunities for managing and regulating infrastructures of smart grids [Fang et al., 2012] and smart cities [Pellicer et al., 2013] in a more efficient and sustainable way.

In the scenario studied in this paper two actors are defined: users, who produce privacy-sensitive data, and the central aggregator, which runs a data analytics algorithm on the data provided by the users. The aggregator is assumed to be *honest but curious* [Goldreich, 2005] i.e. it may run privacy-intrusive algorithms on the data it receives. This scenario is an example of volunteer's dilemma [Diekmann, 1985]: the more users contribute to the public good the greater the benefit for all users, but those who contribute pay a privacy cost.
Users can vary their degree of contribution by reducing the quality of the shared data. A lower quality translates to higher privacy for the user and lower accuracy level of data analytics [Pournaras et al., 2016, Eibl and Engel, 2016].

The goal of this paper is to design a mechanism that improves the trade-off between privacy and accuracy. The mechanism should be bottom-up, i.e. implemented without requiring the collaboration of the aggregator, as opposed to top-down, i.e. a change at the central level is needed for the system to function.

A baseline scenario is considered in which the aggregator receives data from users individually. In the experimental scenario users are divided in groups and within each group they share their data with a local aggregator. The central aggregator receives aggregated data from each group instead of individual user data. The group-level aggregation step obfuscates the individual user contributions, thus making it computationally harder to infer privacy-sensitive information about individual users.
The proposed mechanism is empirically studied with computer simulations and validated with real-world data from the following application scenarios:

i. In order to *improve traffic congestion*, the aggregator computes the average speed across the road network from individual GPS traces. Real-world data is taken from the Regional Transportation Commission of Southern Nevada (RTCSNV) data set, containing GPS traces of cars [NREL, 2015].

ii. In order to *optimize the load of the network*, the aggregator computes the average power consumption of an electric grid. Real-world data is taken from the ECBT dataset, collecting electricity consumption profiles of households [ECBT, 2012].

Introducing groups in the social organization is empirically shown to improve individual privacy over the baseline, without an impact on accuracy. Furthermore, if groups are large enough, the mechanism improves the privacy independently of individual contributions (cf. Figure 1a). Inter-group effects such as the influence of individual contributions on the privacy of other group members are investigated. Finally, several grouping strategies are evaluated and compared (cf. Figure 1b), and the implications for the design of an incentive mechanism are discussed. The contributions of this paper are summarized as follows:

i. Findings that justify social organization as a new mean for enhancing privacy.

ii. The introduction of a bottom-up privacy-preserving mechanism that increases privacy without sacrificing on aggregation accuracy.

iii. The introduction of privacy and accuracy metrics for networks performing group-level data sharing.

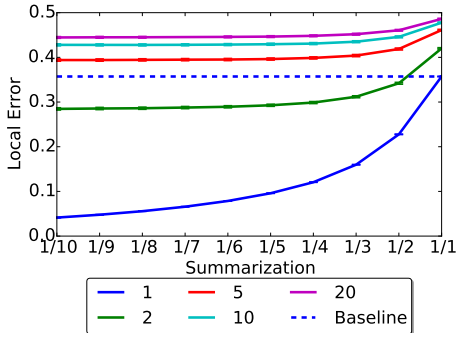
iv. The introduction of measurable trade-offs between privacy and accuracy for networks performing group-level data sharing.

v. The introduction and performance comparison of grouping strategies.

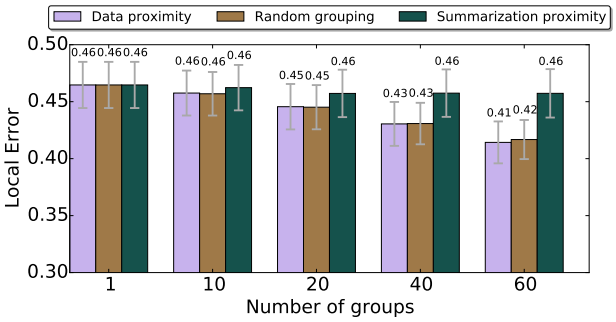
vi. The applicability and measurements of the proposed mechanism in two application scenarios in the domains of traffic and energy management.

The work presented in this paper is relevant to policy-making in regards to privacy of bottom-up crowd-sourcing platforms. Moreover, this paper can be of interest to practitioners, administrators and system operators of distributed networks with privacy-sensitive components, who are interested in improving the privacy of the users.

2 Figure(s)



(a) Privacy level for a given group size and a contribution level (higher is better).



(b) Comparison of privacy level for different grouping strategies (higher is better).

Figure 1: (a) The legend indicates the size of groups. Points (combinations of group size and contribution level) above the dashed line provide a higher privacy level than the highest privacy level without grouping. (b) Error bars represent the standard deviation across simulations.

3 References

[Diekmann, 1985] Diekmann, A. (1985). Volunteer's dilemma. *Journal of Conflict Resolution*.

[ECBT, 2012] ECBT (2012). [dataset] electricity customer behaviour trial.

[Eibl and Engel, 2016] Eibl, G. and Engel, D. (2016). Differential privacy for real smart metering data. *Computer Science - Research and Development*.

[Fang et al., 2012] Fang, X., Misra, S., Xue, G., and Yang, D. (2012). Smart grid; the new and improved power grid: A survey. *IEEE Communications Surveys & Tutorials*.

[Goldreich, 2005] Goldreich, O. (2005). Foundations of cryptography - a primer. *Foundations and Trends in Theoretical Computer Science*.

[Medaglia and Serbanati, 2010] Medaglia, C. M. and Serbanati, A. (2010). *An Overview of Privacy and Security Issues in the Internet of Things*.

[NREL, 2015] NREL (2015). [dataset] transportation secure data center.

[Pellicer et al., 2013] Pellicer, S., Santa, G., Bleda, A. L., Maestre, R., Jara, A. J., and Skarmeta, A. G. (2013). A global perspective of smart cities: A survey. In *Seventh International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*.

[Pournaras et al., 2016] Pournaras, E., Nikolic, J., Velásquez, P., Trovati, M., Bessis, N., and Helbing, D. (2016). Self-regulatory information sharing in participatory social sensing. *EPJ Data Science*.