Towards Modeling Singling Out*†

Aloni Cohen‡ Kobbi Nissim§

FORC 2020

Abstract

There is a significant conceptual gap between legal and mathematical thinking around data privacy. The effect is uncertainty as to which technical offerings meet legal standards. This uncertainty is exacerbated by a litany of successful privacy attacks demonstrating that traditional statistical disclosure limitation techniques often fall short of the privacy envisioned by regulators. This work demonstrates the role that principled analysis supported by mathematical argument can and should play in articulating and informing public policy at the interface between law and technology.

The article focuses on *singling out*, which is a concept appearing in the GDPR. We draw on the legislation, regulatory guidance, and mathematical reasoning to define a new technical concept—*predicate singling-out*—aimed at capturing a core part of GDPR's intent. An adversary predicate singles-out a dataset x using the output of a data-release mechanism $M(x)$ if it finds a predicate p matching exactly one row in x with probability much better than a statistical baseline. A data-release mechanism that precludes such attacks is *secure against predicate singling-out* (*PSO secure*).

We argue that PSO security is a mathematical concept with legal consequences. Any data-release mechanism that purports to "render anonymous" personal data under the GDPR must prevent singling out and, hence, must be PSO secure.

We analyze the properties of PSO security, showing that it fails to compose. Namely, a composition of super-logarithmically many exact counts, each individually PSO secure, facilitates predicate singling-out. Furthermore, there exist two PSO secure mechanisms whose composition is not PSO secure.

Finally, we ask whether differential privacy and k-anonymity are PSO secure. Leveraging a connection to statistical generalization, we show that differential privacy implies PSO security. However, and in contrast with current legal guidance, k-anonymity does not: There exists a simple predicate singling-out attack under mild assumptions on the k-anonymizer and the data distribution.

*For the full version of the paper see: Aloni Cohen and Kobbi Nissim. Towards formalizing the gdpr's notion of singling out. *Proceedings of the National Academy of Sciences*, 117(15):8344–8352, 2020.

†This material is based upon work supported by the U.S. Census Bureau under cooperative agreement No. CB16ADR0160001 and by the National Science Foundation under Grant No. CNS-1413920. Aloni Cohen was additionally supported by the 2018 Facebook Fellowship and MIT's RSA Professorship and Fintech Initiative. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of our funders.

‡School of Law and Rafik B. Hariri Institute for Computing and Computational Science & Engineering, Boston University, aloni@bu.edu.

§Department of Computer Science, Georgetown University, kobbi.nissim@georgetown.edu.