

## The Privacy Paradox and Optimal Bias-Variance Trade-offs in Data Acquisition

GUOCHENG LIAO\*, The Chinese University of Hong Kong & Shenzhen Institute of Artificial Intelligence and Robotics for Society, China

YU SU\*, California Institute of Technology, USA

JUBA ZIANI, University of Pennsylvania, USA

ADAM WIERMAN, California Institute of Technology, USA

JIANWEI HUANG, The Chinese University of Hong Kong, Shenzhen & Shenzhen Institute of Artificial Intelligence and Robotics for Society, China

While users claim to be concerned about privacy, often they do little to protect their privacy in their online actions. One prominent explanation for this “privacy paradox” is that when an individual shares her data, it is not just her privacy that is compromised; the privacy of other individuals with correlated data is also compromised. This information leakage encourages oversharing of data and significantly impacts the incentives of individuals in online platforms. In this paper, we study the mechanism design for data acquisition in settings with information leakage and verifiable data. We provide a closed-form solution of an incentive compatible mechanism that optimizes the worst-case trade-off between bias and variance of the estimation subject to a budget constraint, where the worst-case is over the unknown correlation between costs and data. The novelty of our mechanism is three-fold. First, we introduce the privacy cost to model impacts of data correlation and capture heterogeneous correlation strength. Second, in reality, not every agent always decides to contribute their data. Thus, we introduce the notion of participation rate as the ratio of the number of participants contributing their data to the number of total agents, and study its impact on mechanism design. Third, given that not every agent always participates, it is not always realistic to aim for an unbiased estimator. Thus, we present, for the first time, a trade-off between the bias and variance of the estimator when privacy cost is considered. We offer some interesting insights about mechanisms for data acquisition. First, an unbiased estimator is possible even if the budget is relatively small. Second, privacy cost due to information leakage makes it possible for the analyst to underpay the agents and encourage more agents to participate. This coincides with the “privacy paradox” phenomenon frequently observed in platforms today.

Link to full paper: <http://arxiv.org/abs/2105.14262>.

CCS Concepts: • **Theory of computation** → **Algorithmic mechanism design**; • **Security and privacy** → **Economics of security and privacy**.

Additional Key Words and Phrases: privacy paradox; mechanism design; data correlation; online platform

### ACM Reference Format:

Guocheng Liao, Yu Su, Juba Ziani, Adam Wierman, and Jianwei Huang. 2021. The Privacy Paradox and Optimal Bias-Variance Trade-offs in Data Acquisition. In *Proceedings of the 22nd ACM Conference on Economics and Computation (EC '21), July 18–23, 2021, Budapest, Hungary*. ACM, New York, NY, USA, 1 page. <https://doi.org/10.1145/3465456.3467595>

---

\*Both authors contributed equally.

---

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

EC '21, July 18–23, 2021, Budapest, Hungary

© 2021 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8554-1/21/07.

<https://doi.org/10.1145/3465456.3467595>