New Directions in Anonymization: Permutation Paradigm, Verifiability and Transparency

Josep Domingo-Ferrer
Universitat Rovira i Virgili, Tarragona, Catalonia - josep.domingo@urv.cat

Krishnamurty Muralidhar
University of Oklahoma, Norman OK, U.S.A. - krishm@ou.edu

Abstract

Current approaches to anonymization of microdata sets are either utility-first (use an anonymization method with suitable utility features, then evaluate the disclosure risk and, if needed, reduce the risk by possibly sacrificing some utility) or privacy-first (enforce a target privacy level via a privacy model, *e.g.* k-anonymity or ε-differential privacy, without regard to utility). The second approach is the only one that offers formal privacy guarantees, but it is seldom used in practice because it produces data releases with no utility guarantees. We address the previous conflict between utility and privacy, by showing how to get privacy guarantees without destroying more utility than necessary. Furthermore, we tackle the following unresolved issues: how to make anonymization verifiable by the data subject (so that she can verify how safe is the record she has contributed), how to get rid of background knowledge assumptions when defining the intruder, and what does transparency of anonymization to the user mean. We present a permutation paradigm of anonymization, whereby any microdata anonymization method is functionally equivalent to permutation plus a residual amount of noise addition. Thus, the privacy offered by a method is the amount of permutation it achieves and this amount can be verified not only by the data protector, but also by the subject contributing each record (subject-verifiability). Furthermore, we define an intruder model that makes no assumption on background knowledge and we show how to determine the right amount of permutation to withstand such an intruder without losing more utility than necessary. Finally, we show that an anonymization method safe against such an intruder can also be safely transparent to any user, which increases the analytical utility of anonymized data.

Keywords: data anonymization; statistical disclosure control; subject verifiability; intruder model; transparency to users.