Journal of Privacy and Confidentiality (2010) **2**, Number 2, pp. 1–2

# In This Issue

Stephen E. Fienberg[*]
Editor-in-Chief, *Journal of Privacy and Confidentiality*

This issue of the *Journal of Privacy and Confidentiality* focuses on four technical contributions, three of which deal with different aspects of efforts to produce "sanitized" synthetic statistical databases or sanitized responses to queries. The first two of these papers focus on the utility of sanitized databases based on the technique of multiple imputation, the third deals with intruder-styled attacks on output perturbation sanitizers. The fourth addresses a different problem regarding access to, and the use of, web-based data via anonymous authentication.

Kinney, Reiter, and Berger adopt the approach of multiple imputation for the creation of sanitized databases and ask the important question of how analysts can use the output from a multiple imputation approach to carry out a valid Bayesian model selection strategy, via Bayes factors and model search algorithms. The technical issues they discuss include searches within the model used to generate the multiply imputed databases, $M*$, as opposed to the models of interest to the analyst, $M$, and possible conflicts between $M*$ and $M$.

Charest asks a very different type of question about the multiple imputation strategy. She recognizes that although the methodology produces replicate synthetic datasets, and that this heuristically provides confidentiality/privacy protection to the individuals in the original database, the method still offers no formal proof of protection. Thus she demonstrates one way to integrate the formal protection tool of differential privacy into the method and still produce valid synthetic data. This involves altering the method for drawing inferences from the multiply imputed data so that the inferential results are valid, in light of the added noise associated with the differential privacy mechanism. The paper concludes with a discussion of the extent to which the differential privacy mechanism impedes statistical inference as a consequence of the added variability inherent in it.

Merener revisits a landmark result from the privacy protection literature, due to Dinur and Nissim who provided a polynomial-time algorithm that reconstructs binary elements from a database from perturbed replies to queries. Protecting against such attacks has been the focus of much important literature, such as that on differential privacy. He uses a different approach to the problem which relies on new probabilistic proofs employing a well-known generalization of the central limit theorem. While the results in the paper can be found in other papers in the literature, the method of proof may prove useful in new problems.

Lindell addresses an important and difficult aspect of online privacy—that of anonymous authentication to protect against the damages that might arise from situations

[*] Department of Statistics, Carnegie Mellon University, Pittsburgh PA, mailto:fienberg@stat.cmu.edu

where the identification of the user engenders a loss of privacy and possible harm. He distinguishes between *user anonymity* and *unlinkability* and provides a formal description of how encryption techniques can be used to provide user protection. He concludes with a discussion of how the approach could actually be implemented in practice.