Data privacy: state of the art

Vicenç Torra*, **Josep Domingo-Ferrer****

*Department of Computing Science, Umeå University, Sweden.

** Universitat Rovira i Virgili, Department of Computer Engineering and Mathematics, CYBERCAT-Center for Cybersecurity Research of Catalonia, UNESCO Chair in Data Privacy, Tarragona, Catalonia.

E-mail: vtorra@cs.umu.se, josep.domingo@urv.cat

Received 31 January 2023; accepted 31 January 2023

Abstract. In 2008, we published the first issue of the *Transactions on Data Privacy.* Now, 15 years later, we publish this special issue to celebrate the journal's anniversary. The issue includes papers that give an overview of key privacy technologies, and it discusses important challenges and future research directions.

Keywords. Data privacy, privacy technologies, k-anonymity, differential privacy, human mobility data, tabular data, synthetic data.

1 Introduction

In 2008, we started this journal to provide an international forum for researchers on all topics related to data privacy technologies. Julia Lane *et al.* [7] inaugurated the issue with a position paper on data access and data privacy.

The field has changed in the last 15 years. The relevance and importance of data privacy have increased. Machine and statistical learning, combined with big data [1], can lead to the disclosure of sensitive information. Rules and regulations to enforce data protection and privacy, such as the General Data Protection Regulation (GDPR) in Europe, are nowadays in force in most countries. Identity disclosure, attribute disclosure, membership inference attacks, re-identification, k-anonymity [8], differential privacy [4], privacy by design are examples of terms that are currently known beyond the specialists of data privacy. Data privacy and statistical disclosure control are now topics regularly taught in computer engineering and statistics schools. Several books have been published [5, 6, 9] that describe privacy models and technologies for researchers, students, practitioners, and the general public.

To celebrate the fifteenth anniversary of our journal, we publish this special issue that includes four papers.

The first paper by Jordi Castro is on privacy technologies for statistical tables. The author discusses the research done in the field in the last 30 years, focusing on the approaches that are of practical use. The discussion includes their pros and cons compared to recent techniques that are not based on optimization methods.

The second paper by Jerome P. Reiter is about synthetic data. Synthetic data have gained popularity in the last years as a tool to provide privacy. The author reviews the progression of data synthesis, describes some unresolved challenges, and speculates about its future.

The third paper by Sabrina De Capitani di Vimercati, Sara Foresti, Giovanni Livraga, and Pierangela Samarati is on k-anonymity. The concepts of k-anonymity and re-identification have been cornerstones of data privacy. The current limitations [9, 3, 2] of differential privacy to provide strong privacy guarantees in practical applications are a reason to supplement it with other privacy models in the data privacy toolkit. In this paper, the authors review k-anonymity and its main extensions, and discuss some advanced application scenarios.

The last paper is by Anna Monreale and Roberto Pellungrini. Complex data such as, for example, sets of records that have time dependencies among them, are difficult to protect. Human mobility data are an example of complex big data whose analysis is in great demand. Monreale and Pellungrini survey the advancements on privacy-preserving mobility data publishing. They discuss attacks and privacy models, as well as tools for data protection.

The celebration of this 15th anniversary would not have been possible without the help of the authors that have contributed to the journal over that period with their interesting works, and the reviewers that have examined the submissions. To all of them, as well as to the members of the editorial board, thanks!

References

[1] D'Acquisto, G., Domingo-Ferrer, J., Kikiras, P., Torra, V., de Montjoye, Y.-A., and Bourka, A. (2015) Privacy by design in big data: An overview of privacy enhancing technologies in the era of big data analytics. ENISA Report.

[2] Blanco-Justicia, A., Sánchez, D., Domingo-Ferrer, J., and Muralidhar, K. (2023) A critical review on the use (and misuse) of differential privacy in machine learning. *ACM Computing Surveys* 55(8):1-16.

[3] Domingo-Ferrer, J., Sánchez, D., and Blanco-Justicia, A. (2021) The limits of differential privacy (and its misuse in data release and machine learning). *Communications of the ACM* 64(7):33-35.

[4] Dwork, C. (2006) Differential privacy. In: *Proc. of ICALP 2006*, pp. 1-12. LNCS 4052, Springer.

[5] Duncan, G. T., Elliot, M., and Salazar, J. J. (2011) *Statistical Confidentiality*. Springer.

[6] Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Nordholt, E. S., Spicer, K., and De Wolf, P.-P. (2012) *Statistical Disclosure Control*. Wiley.

[7] Lane, J., Heus, P., and Mulcahy, T. (2008) Data access in a cyber world: making use of cyberinfrastructure. *Transactions on Data Privacy* 1(1):2-16.

[8] Samarati, P. (2001) Protecting respondents' identities in microdata release. *IEEE Trans. on Knowledge and Data Engineering*, 13(6):1010-1027.

[9] Torra, V. (2022) *Guide to Data Privacy: Models, Technologies, Solutions*. Springer.