Keynote Talk: A Probabilistic Look Ahead of Anonymization

Yücel Saygin
Faculty of Engineering and Natural Sciences
Sabanci University, Istanbul, Turkey
ysaygin@sabanciuniv.edu

ABSTRACT

Data anonymization is an expensive process, and sometimes the utility of the anonymized data may not justify the cost of anonymization. For example in a distributed setting where the data reside at different sites and needs to be anonymized without a trusted server, Secure Multiparty Computation (SMC) protocols need to be employed. However, the cost of SMC protocols could be prohibitive, and therefore the parties may want to look ahead of anonymization to decide if it is worth running the expensive SMC protocols. In this work, we describe a probabilistic fast look ahead of k-anonymization of horizontally partitioned data. The look ahead returns an upper bound on the probability that k-anonymity will be achieved at a certain utility where the utility is quantified by commonly used metrics from the anonymization literature. The look ahead process exploits prior information such as total data size, attribute distributions, or attribute correlations, all of which require simple SMC operations to compute. More specifically, given only statistics on the private dataset, we show how to calculate the probability that a mapping of values to generalizations will make a private dataset k-anonymous.

ACKNOWLEDGMENTS

This is a joint work with Ercan Nergiz, Ercument Çicek, and Thomas Pedersen from Sabanci University, Istanbul, Turkey.

Copyright is held by the author/owner(s).