Differential Privacy with Arbitrary Metrics

Kostas Chatzikokolakis

CNRS, INRIA and LIX, Ecole Polytechnique

Abstract. Differential privacy is a well-known notion of privacy from the field of statistical databases. It requires that a randomized query, when applied to *adjacent* databases – i.e. those differing in at most one row – should produce outcomes with similar probability. The notion of adjacency is crucial here; it implies that the *distance* between databases is measured by the number of rows in which they differ, i.e. by the hamming metric. Nevertheless, the definition can be adapted to use other notions of distance. Indeed, a few works in the literature do use other metrics, but in general this direction is not widely explored.

In this work, we study a generalized definition of differential privacy using arbitrary metrics. We show several results for the extended definition, many of which are direct generalizations of corresponding results for the standard one. Moreover, we develop a characterization of differential privacy, which provides an intuitive explanation of the privacy guarantees offered by the extended definition. We also study a generalized family of Laplace mechanisms, shown to be differential private for any metric, which can be instantiated to several known mechanisms, both continuous and discrete. Then, we study optimality results in the context of generic metrics. We show that, although it is known that counting queries are the only family for which a universally optimal mechanism exists in the standard case, optimality can be achieved for other families when different metrics are employed.

Finally, we provide examples illustrating different goals achieved by different metrics. First, a metric is used as a relaxation of differential privacy, in the case we interested in protecting user's value within a certain range. In this case, a mechanism can achieve differential privacy using less noise, thus providing better utility. Furthermore, generic metrics can be used to apply differential privacy in scenarios where secrets cannot be naturally viewed as databases, and the adjacency relation is not canonical. A good example of such an application are location-based systems.