Abstract

The field of Privacy Preserving Data Publication(PPDP) and its cousin Privacy Preserving Data Mining(PPDM) are concerned with the anonymization of sensitive information in data, whilst allowing its use for data analysis and mining. Therefore, PPDP and PPDM are faced with the often-conflicting tasks of reducing privacy risk and maximizing the utility of the published data. In this report, we analyse existing privacy-transformation techniques in the field of PPDP that anonymize datasets with Multiple Sensitive Attributes (MSA), as opposed to those with a Single Sensitive Attribute (SSA). Of these, we present an analysis of Decomposition, an algorithm which generates a dataset with distinct ℓ-diversity over Multiple Sensitive Attributes using a partitioning approach. We discuss some areas in which significant improvements can be made over Decomposition: in the realms of its running time, its data utility, and its applicability in the case of Multiple Release Publishing. To this effect, we describe an improved algorithm, *Decomposition+* that implements these improvements and is therefore more suited to use in real-life scenarios.