

Efficient private record linkage of very large datasets

Rainer Schnell

University of Duisburg-Essen
German Record Linkage Center
rainer.schnell@uni-due.de

Abstract

Increasingly, administrative data is being used for statistical purposes, for example registry based census taking. In practice, this usually requires linking separate files containing information on the same unit, without revealing the identity of the unit. If the linkage has to be done without a unique identification number, the comparison of keys derived from unit identifiers and assumed to be similar is necessary. When dealing with large files like census data or population registries, comparing each possible pair of keys of two files is impossible. Therefore, special algorithms (blocking methods) have to be used to reduce the number of comparisons needed. The presentation will discuss the most widely used blocking algorithms for encrypted data, describe a recently introduced algorithm and compare the performance of the blocking methods currently thought to be the most effective for very large files.

Keywords: Privacy Preserving Record-Linkage, Blocking, Indexing, Bloom-Filter