BENJAMIN SANTOS, Statistics Canada

Multi-Party Privacy Preserving Record Linkage based on Circuit Private Set Intersection

Record Linkage (RL) is the process of combining information about entities in multiple data sources into a single linked dataset. In some linkages, the desired output is not the linked data itself, but a set of aggregates based on the cross-linked dataset, such as, aggregated tables. In our previous work [1], we designed and implemented a protocol for Privacy-Preserving RL (PPRL) with aggregation based on Oblivious Programmable Pseudo-Random Functions (OPPRFs) and Secure Multi-Party Computation (SMPC). This protocol allows two parties with datasets, e.g., a National Statistical Office (NSO) and a Government Agency (GA), to obtain weighted aggregates based on values present in the intersection of both datasets while ensuring privacy in a semi-honest scenario. The goal is to extend it to more than two parties, i.e., Multi-Party PPRL (MP-PPRL). This is a natural extension since parties could be playing the role of an NSO, GAs, regional and/or private partners. We based our work on Chandran et al. [2], that implements Relaxed Batch OPPRFs and SMPC to build a protocol for Circuit Private Set Intersection, which we extended to MP-PPRL. We found that the multi-party extension to PPRL is more complex and stiffer, meaning the solution must be tailored to the problem of study: datasets and aggregations.

 References

 [1] Dugdale, et al. Practical Privacy-Aware Data Linkage and Statistical Aggregation based on Privacy Enhancing Techniques. CROSS-NTTS 2023.

 [2] Chandran, et al. Efficient Linear Multiparty PSI and Extensions to Circuit/Quorum PSI. Proceedings of the 2021 ACM SIGSAC CCCS.