Inter-Activity Sharing of Sensitive Data for Statistics via Secure Multi-party Computation: A Case Study at the US Department of Education

The US Federal Data Strategy (FDS) and the Foundations for Evidence-Based Policymaking Act of 2018 mandate inter-agency data sharing to promote informed decision making. However, current approaches to such sharing put the privacy of shared data at significant risk. Emerging privacy-preserving technologies (PPTs) such as *secure multi-party computation* (MPC) may foster practical inter-agency data sharing while eliminating this risk to privacy, yet this opportunity has not yet been broadly proven. To demonstrate the promise of PPTs for this purpose, we conducted a pilot project using MPC to reproduce a commonplace statistical application that requires such inter-agency sharing within the US Department of Education. The prototype application reproduces a portion of the annual 2015–16 National Postsecondary Student Aid Study (NPSAS:16) report, showing statistics on average federal Title IV aid received by undergraduate students. In this setting, input data comes from two agencies within the Department: the National Postsecondary Student Aid Study group (NPSAS) at the National Center for Education Statistics (NCES) and the National Student Loan Data System (NSLDS). Today, NSLDS must share sensitive student financial information with NPSAS, and NPSAS must share students' social security numbers with NSLDS. To avoid those disclosures while successfully and efficiently providing the same statistics, our prototype performs the same data linkage and statistical analysis without sharing that sensitive information either between the agencies or with a trusted third party. This "zero-trust" approach relies on computing the necessary statistics while the data remains encrypted, and then decrypting only the results of the analysis. Our experiments produced accurate results while at the same time providing strong cryptographic security and incurring total computation times and network traffic costs that were reasonable compared to those of the typical methods used to produce these statistics. We report on our methods and results from this recently completed project.

Keywords: Privacy; Inter-agency Data Sharing; Cryptography