OTTAWA 2023
64TH WORLD STATISTICS CONGRESS

IPS Paper

Multi-Party Secure Private Computing-as-a-service for the production of future official statistics: motivations, requirements and challenges

Author: Mr Albrecht Wirthmann

Coauthors: Fabio Ricciato

Submission ID: 1328

Reference Number: 1328

Brief Description

The availability of unprecedented amounts of digital data, combined with increasing awareness by citizens and companies about the value of their data, call statistical institutions to adopt stronger mechanisms to ensure privacy and transparency of the statistical production process in order to preserve public acceptance and trust.

Some novel technologies that have emerged recently at the intersection between the fields of cryptography, distributed systems and computer science, are instrumental to achieve this goal.

The family of so-called Privacy Enhancing Technologies (e.g.

Secure Multiparty Computation, Homomorphic Encryption and Trusted Execution Environment) open new possibilities in terms of how “data” and “computation” are handled and controlled within and across organisations.

With these technologies, governance policies can be enforced technologically without necessarily relying on a single point of trust.

These technologies have reached maturity and are already moving from laboratories to real-world applications.

An increasing number of statistical institutions have started to pioneer possible applications of these technologies in explorative or pilot projects around different use-cases.

The goal of this session is to provide a stage for such pioneering activities, share early experiences and lessons learned, discuss what we can (or can not) expect from these technologies and exchange views as to how the adoption of these technologies impacts (or is impacted by) organisational aspects, business processes and legislation.

The session will consist of 4 speakers reporting on experimental and pilot projects leveraging PET for statistics, including one speaker from the UN PET Lab.

The session will be divided in two parts, with a series of brief presentations by the speakers followed by a panel discussion between the speakers.

Abstract

In the traditional model of statistical production a single organization, i.e., the statistical authority, collects the whole input data and from there computes the desired output statistics. Whenever the desired output statistics requires the integration/combination of different input data sets held by different organizations, the traditional solution is to exchange (a copy of) the input data, either directly between the concerned institutions or with a Trusted Third Party. This approach amplifies the risks, as it multiplies the number of data copies and actors that can access them. Exchanging the input data is just a means towards the goal of computing the desired output – and is not the only means available today. Alternative solutions based on Privacy Enhancing Technologies, and specifically Multi-Party Secure Private Computing (MPSPC), allow nowadays to computing the desired output statistics without disclosing the input data to any entities other than their respective data holders.
At the core of MPSPC lies the requirement that no single entity should be technically capable of taking control of the process (no single-point-of-trust). Process control is split among K>1 “processing parties” that are to be trusted collectively, not individually. MPSPC operates according to policies based on the principle that no computation may be executed on the data without preliminary explicit approval by all K processing parties, and is engineered based on state-of-the-art technologies that enforce these policies. Hence, the robustness of MPSPC depends jointly (1) on the choice of processing parties; (2) on the strength of the policies that define their roles and powers; and (3) on the strength of the technologies that enforce these policies at hardware and/or software level.
Setting up a robust MPSPC solution requires investments, capacity and also specialized skills on the side of potential adopters. Not all statistical institutions may have the internal resources and/or the necessary knowledge to develop, deploy and maintain their own solutions, and anyway the costs might be disproportionate compared to the expected benefit. The cost factor may discourage adoption wholly or drive towards adoption of sub-optimal solutions with less-than-maximum levels of security and robustness. Furthermore, interoperability may not be guaranteed among

solutions developed independently by different institutions.
The concerns about costs, robustness and interoperability led Eurostat, in the framework of the UNECE HLG-MOS project on Input Privacy-Preservation, to elaborate the vision of a shared MPSPC infrastructure, developed and operated by a network of statistical institutions and made available on demand to execute computation based on the MPSPC paradigm. As in other areas of Information Technologies, the basic idea is to decouple the development (and maintenance) of the infrastructure from its utilization. This allows to pool together resources and expert knowledge during the development phase, increasing cost-effectiveness and ultimately enabling the achievement of very high levels of robustness and security guarantees, based on state-of-the-art technologies and design criteria.
The shared MPSPC infrastructure developed in this way could then be used on demand by statistical institutions and by their partners (e.g., external data providers). This model was named MPSPC-as-a-service (MPSPCaaS).

The talk will provide a high-level overview of the motivations, challenges, and key requirements underlying the MPSPCaaS concept as initially elaborated in the context of the UNECE HLG-MOS Input Privacy-Preservation project. Furthermore, it will provide a detailed account of the steps being taken by Eurostat towards the specification and demonstration of a MPSPCaaS prototype serving the (future) needs of the European Statistical System.