Secure MultiParty Computation (MPC) for Big Data Analytics:
technology readiness from an enterprise perspective

Paul Koster – November 2019

MPC can solve business problems in the area of (big) data analytics. But what does this mean for decision makers having to choose an approach for their analytics project? MPC has a lot of potential uses within the security and privacy domain as it lets participating parties get insights from their own and other parties' data without having to share or receive sensitive data. This may protect privacy (GDPR) of individuals and it can protect confidentiality of business assets like data and AI models.

Recent progress made privacy preserving (big) data analytics significantly more practical. This includes progress in core MPC protocols, MPC frameworks and MPC-based machine learning algorithms. Several machine learning algorithms such as ridge / logistic regression and decision trees / random forests can be applied today on thousands to millions of records.

However, MPC-based data analytics still has limitations. Despite recent progress, the performance of MPC-based machine learning is not sufficient for all applications. Deep learning is a classic example. Other major limitations stem from the availability of software frameworks and expertise.

MPC machine learning libraries are not readily available let alone libraries that come close to an MPC equivalent of scikit-learn or R with their mature and rich feature set. Consequently, an analytics project must plan for adapting or implementing algorithms themselves. For long term deployments one must also take into account that most MPC frameworks are not production-grade.

MPC expertise is critical to make MPC-based data analytics a success, but only available at a few universities, research centers, startups and enterprises. Yet, expertise is needed in an early phase to make estimates on feasibility and cost, software developers must have affinity and go through a learning curve, and use case specific algorithms or critical algorithm optimizations to make a solution scale typically involve deep MPC expertise or research.

Successful application of MPC-based analytics assumes certain criteria are met. This requires budget, skill, and stakeholder understanding. The budget should take into account development effort in the order of man months to implement a basic MPC-based algorithm, as well as operational aspects to manage the infrastructure and coordinate online execution. To onboard the necessary skill an organization has to develop the capability and / or bring in external expertise through consultancy.

Organizational stakeholders like ethical boards and privacy officers should be timely included to explain for example that MPC technology is an appropriate technical measure as required by the GDPR. Similarly, informed consent and end-user communication should be prepared to take into account that end-users care more about trust than they understand technology.

Conclusion. In business settings, MPC competes with alternatives, e.g. de-identification. MPC is practical for some big data analytics today, but with available technology and cost initial adoption mostly fits high value applications or applications where more cost effective alternatives are not feasible or sub-optimal. Those applications optimally benefit from the choice for an MPC-based approach.

Note: the field of secure multi-party computation is rapidly evolving. Therefore the statements made here should be re-assessed and updated not later than a year after above publication date.

The project SODA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731583.