PETs in Practice:

using secure multi-party computation to enable pay gap analysis in the Greater Boston area

January 2024

Open Data Institute

Contents

About

This report has been researched and produced by the Open Data Institute, and published in January 2024. Its lead author was Elea Himmelsbach, supported by Calum Inverarity, Jared Keller, Neil Majithia and Claudine Tinsman. If you want to share feedback by email or want to get in touch, contact the PETs programme team at pets@theodi.org.

To share feedback in the comments, highlight the relevant text and click the 'Add a comment' icon on the right-hand side of the page.

How can it be improved? We welcome suggestions from the community in the comments and by email.

PETs in Practice: using secure multi-party computation to enable pay gap analysis in the Greater Boston area

Secure multi-party computing (MPC) is a novel privacy-enhancing technology (PET) that enables collaborative data analysis and analysis of data from multiple sources whilst preserving the privacy and security of the underlying data. Despite its unique value proposition, there are still only relatively limited case studies that enable organisations to study its adoption and transferability to other contexts.

At the ODI, we are interested in understanding the requirements that need to be met to make PETs transferable to different use cases and increase their adoption by other organisations that may have more limited resources. This case study documents the use of secure MPC by the Boston Women's Workforce Council[1] (BWWC), which has been using the technology since 2016 to produce a bi-annual longitudinal study[2] to increase the visibility of the gender and racial pay gap in the Greater Boston area. The application concurrently enables participating organisations to securely and privately benchmark how they perform vis-a-vis these pay gaps relative to other businesses in the area.

There is value in increasing access to sensitive data for social benefit, and we believe that this example helps demonstrate this. We hope that the documentation of this example will prove useful for other non-profit organisations and researchers working on PETs to establish safe access to data to address societal challenges.

For this case study, insights are drawn from desk research and interviews with members and former members of the Rafik B. Hariri Institute for Computing and Computational Science & Engineering at Boston University (BU)[3] and BWWC

[1] BWWC (2022) 'About us' https://thebwwc.org/about-us
[2] For BWWC gender and racial wage gaps reports, see: BWWC (2022) 'Data' https://thebwwc.org/wage-gap-studies
[3] Boston University 'About the Hariri Institute' https://www.bu.edu/hic/about-hic/

between July and December 2023.

Background

The following case study documents the application of a web-based MPC protocol that was designed and implemented by the BU to help BWWC benchmark and report on the gender and racial pay gap in the Greater Boston area.

Boston Council[4] has been interested in producing a pay equity study to benchmark and address racial and gender pay inequalities across the Greater Boston area since 2013. Their ambition was to collect more granular data than the US Census[5] provides, relying as it does on self-reported data. The council commissioned BWWC to lead the initiative; however, the project was slow to get off the ground. The data-gathering effort stalled despite wide support from the local business community, which pledged to contribute aggregated salary information to BWWC on a bi-annual basis. The sensitive nature of the payroll data required to enable this analysis initially prevented the project from going ahead. While supportive in principle, businesses were concerned about inadvertently exposing themselves to legal and commercial risks when sharing payroll data. Potential third-party arbiters, who were approached to steward the data, were similarly concerned about exposing businesses and themselves to litigation risks. Learning about this dilemma, BU, who already worked with the Boston Council at that time, offered their services and proposed the use of secure MPC to overcome the deadlock.[6]

MPC is a cryptographic protocol that enables different stakeholders, each with private data, to carry out joint computations without the need to reveal their individual data input. The concept of MPC[7] is based on applying a secret sharing algorithm. The algorithm obfuscates and splits private data, such as businesses' payroll data, into shares amongst the participants. Once the data is obfuscated and split, individual shares cannot be revealed, and no insights can be gained from the data unless trusted parties act together collaboratively to pool the data.

BU presented a useful alternative to using a trusted arbiter, such as a data institution[8], which had been challenging to identify. It provided a solution to enable the safe collection and analysis of combined sensitive payroll data without exposing individual businesses to potential commercial or legal risks, such as costly third-party-arbiter equity lawsuits, because the underlying payroll data from individual businesses would never be shared in this model.

[4] Boston Gov 'Boston.Gov' https://www.boston.gov/
[5] United States Census Bureau 'Census.gov' https://www.census.gov/
[6] For more on this, see: Rich Barlow, Boston University (2015), 'Computational Thinking Breaks a Logjam' https://www.bu.edu/articles/2015/computational-thinking-breaks-a-logjam
[7] For more on MPC, see: Wikipedia (2023) 'Secure multi-party computation' https://en.wikipedia.org/wiki/Secure_multi-party_computation
[8] Jack Hardinges and Jared Robert Keller (2021) 'What are data institutions and why are they important?' https://theodi.org/news-and-events/blog/what-are-data-institutions-and-why-are-they-important/

Challenges in developing an MPC framework for benchmarking

Understanding and adapting a technological solution to fit the needs of a specific context and audience is vital to ensure adoption. This MPC framework, developed by BU, was tailored to the needs of this specific case study. It has been adapted to the financial and technical constraints of the stakeholders, BWWC, and businesses from the Greater Boston area, who pledged to contribute their data for analysis.

The project took two years from inception to deployment for two main reasons. First, BU invested a lot of time in socialising the MPC framework. This was necessary to help overcome the understandable apprehension that companies held towards making their sensitive payroll data available for analysis using a novel technology to protect private data input. Second, the participating businesses constituted a non-technical user group with limited computational capabilities, therefore efforts were made to develop a solution that accommodated these limitations and could be adopted by participants with relative ease. The aim of these efforts was to lower the barriers to participation for organisations, so as to increase the likelihood of their involvement.

The BU team outlined their key design principles as follows:

Accessibility

To lower the burden of adoption, BU did not introduce new application packages that businesses would not already be familiar with. Instead, BU employed commonly used infrastructure and software, web browsers, and Excel spreadsheets to deploy the MPC framework. This made it more straightforward to onboard organisations with limited infrastructure and/or those with strict security requirements.

In addition, BU chose to use a lesser-established client-server MPC model instead of a peer-to-peer model. The client-server model has the benefit of enabling one entity – BU in this instance – to act as a service provider to maintain the infrastructure and do the bulk of the computation. This effectively minimised the computing burdens for all other participants, limiting them to simple transactions relative to the peer-to-peer framework. In a peer-to-peer framework, all participating parties are connected and are required to install the MPC framework locally on their computing infrastructure to participate, thus not relying on a service provider.[9] For this reason, the peer-to-peer model was ruled out in this instance.

Usability

BU introduced two new features: **asynchronicity** and **idempotency,** to make the MPC model more accessible. Unlike traditional MPC models that require all participants to be simultaneously online, incorporating **asynchronicity** allowed each contributing party to join and leave a computation session at their convenience, thus expanding the submission window and reducing the need for coordination. In this design, BU, who acts as the service provider, is the only

[9] For more, see: Andrei Lapets et al. (2019) 'Role-Based Ecosystem for the Design, Development, and Deployment of Secure Multi-Party Data Analytics Applications' https://ieeexplore.ieee.org/document/8901614

actor required to remain online throughout an MPC session, which can last for a predetermined period, for instance, two weeks.

Idempotency is a concept that is used in mathematics and programming and defines the property of an operation so that something can be executed multiple times without changing the outcome. In this context, the concept of idempotency is used to enable the correction of errors post-submission within the duration of a computation session. For instance, if a business has erroneously submitted data, idempotency enables them to fix the error in their Excel file and resubmit the data. Idempotency provides a guardrail for input verification, which is critical to ensure correct data output.

Explainability

Choosing a semi-honest MPC protocol[10] that is easy to explain was a major success factor in inspiring trust from business leaders. A semi-honest protocol, sometimes called passive security, assumes that corrupt parties might merely cooperate to gather information out of the protocol but do not deviate from the protocol specification. In essence, a semi-honest protocol uses existing privacy laws and leverages incentives. In this use case, the semi-honest MPC protocol makes use of the fact that while BWWC and BU could theoretically collude to obtain private data inputs, they would never do so, as they rely on reputational integrity to provide their services. Moreover, any collusion between BU and BWWC would create liability risks for both. BWWC and BU are protected from the legal risks of processing sensitive data as long as they follow the protocol.[11]

In addition – and this is important both for the adoption of the technology as well as for security assurance – both the BU and BWWC have invested greatly into making the MPC concept accessible to business leaders through training and workshops that continue to this day.[12][13]

A comprehensive MPC protocol

Each of these elements was built into the design process to maximise the utility of the secure MPC protocol and minimise the obstacles for willing organisations to participate. While the design of this framework is highly context-specific, the principles behind the design carry wider relevance. They are useful elements to consider when developing new technological solutions for social enterprises with limited financial and infrastructural resources.

Deployment

Figure 1, below, provides a visual representation of part of the deployment of the web-based MPC framework. The fully annotated diagram[14] provides functionality

[10] For more on semi-honest MPC protocols, see: Wikipedia (2023) 'Secure multi-party computation' https://en.wikipedia.org/wiki/Secure_multi-party_computation

[11] For further details about potential attack surface of software application, see: Andrei Lapets et al (2019) Role-Based Ecosystem of the Design, Development, and Deployment of Secure Multi-Party Data Analytics Applications p.137 https://ieeexplore.ieee.org/document/8901614

[12] For further details on the efforts made by the BU team to maximise the utility of the MPC protocol, see: Lucy Qin et al. (2019) 'From Usability to Secure Computing and Back Again' https://eprint.iacr.org/2019/734.pdf

[13] For more on addressing accessibility of MPC, see: Andrei Lapets et al. (2018) 'Accessible Privacy-Preserving Web-Based Data Analysis for Assessing and Addressing Economic Inequalities' (p.3) https://dl.acm.org/doi/pdf/10.1145/3209811.3212701

[14] ODI (2024) 'PETs in Practice: using secure multi-party computation to enable pay gap analysis in the Greater Boston area' https://opendata.kumu.io/pets-in-practice-using-secure-multi-party-computation-to-enable-pay-gap-analysis-in-the-greater-boston-area

to zoom into detail and access additional information about the various elements of the framework, in addition to a narrative walkthrough of the MPC protocol design.

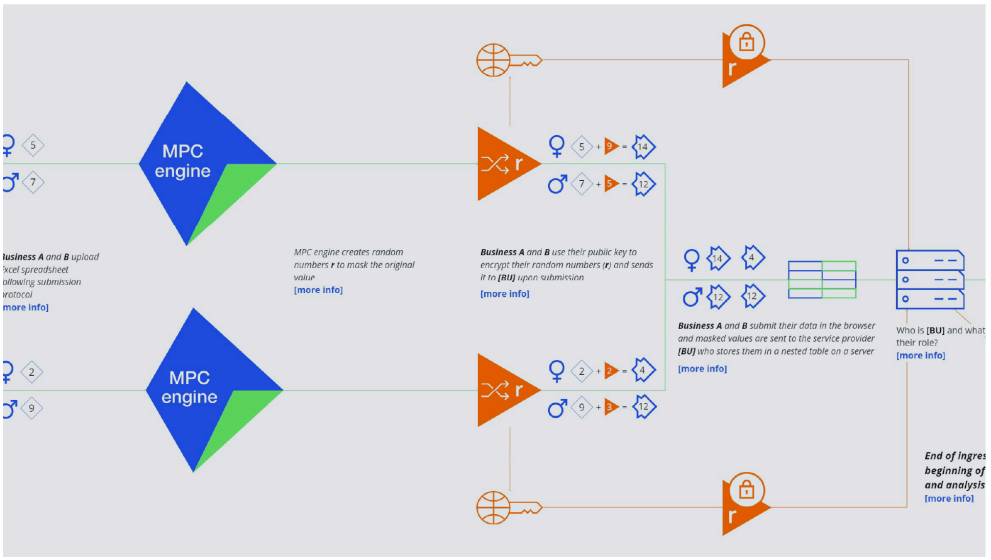

Figure 1: A snapshot of the deployment of the web-based MPC framework
Image source: Open Data Institute

Resourcing, implementation, and maintenance costs

This specific web-based MPC framework was developed to fit the needs and resources of Government agencies, non-profit organisations, and social scientists. The project benefited from in-kind support from the Rafik B. Hariri Institute for Computing and Computational Science & Engineering at Boston University,[15] which used resource constraints to inform their innovation and provided their otherwise expensive software engineering resources at a reduced cost.

From a computational perspective, the BWWC case study is not particularly resource-heavy, as many elements can be solved with basic statistics. The most resource-intensive component is not the computation nor technology and infrastructure itself, but the time needed to adopt the solution and engage with stakeholders.

Once the MPC protocol and infrastructure were set up, minimal effort has been required to maintain, upkeep, and add new features as the principle design has remained the same. As a result, technological maintenance costs have stayed low and largely constitute the upkeep of the infrastructure and protocol.

From a client perspective – BWWC – the main costs are incurred upfront with each new session. Hence, BU needs to write code for any new functionality upfront, such as when BWWC wishes to add a new indicator. This is a slight drawback for BWWC as there is no scope for changes later on in an MPC session once it has commenced. On average, BU has designed a session to last two weeks. As such,

[15] Boston University 'About the Hariri Institute' https://www.bu.edu/hic/about-hic/

queries need to be comprehensively considered and agreed before initiating a session of analysis to address a specific question, or a set of questions.

The developers of the BWWC MPC framework advise that when setting out a budget for a project like the BWWC pay gap study, it is important to understand everything necessary to make it a success and then take the time to listen to reasons for why this is the case.[16] They highlight the importance of starting small for ease of auditing and running small-scale pilots with either fictitious data or data from a smaller group of participants to build awareness of how to use the application before adding new features.[17]

Taking this approach can help in introducing new data technologies and governance measures to audiences on a step-by-step basis, which can build confidence and trust through iterative demonstrations of successful implementation and execution.

Value flows and tangible outcomes

BWWC was started as an organisation in 2013 to make Boston the best place for working women. As part of this mission, they were commissioned to develop a longitudinal pay gap study, which they have produced bi-annually since 2016. For the longitudinal study, BWWC looks at both the adjusted wage gap and raw wage gap.[18] The adjusted wage gap is measured by reviewing the salaries paid to men and women in the same position, adjusting for experience, education and other factors. The raw wage gap measures salaries paid to women across all positions in a business compared to those of all men. This allows BWWC to look at and reflect not only the salary for a given position but also where women stand in the labour hierarchy. The study is informed by more granular data than other wage gap measurements, typically based on self-reported census data. A drawback of using census data is that it is abstracted at higher levels through aggregation. This can prevent more granular analysis at a localised level, which can impede the development of more precise and targeted insights that can then be used to inform tailored interventions.

Insights from the reports created by BWWC have been used to inform academic research and equity policy by the Boston Mayor's office, as well as those from participating businesses and academics. Businesses participating in this study agree to examine their own data, look for wage equity problems and solve them, and report data anonymously every two years to measure community progress.

By contributing their data to the analysis, participating businesses can benchmark how successfully they are addressing gender and racial pay gaps compared to other businesses in the Greater Boston area. This information is only accessible to participating businesses and comes in the form of a ranking relative to the other participating businesses. Data about other businesses, such as their identities and their ranking, is not disclosed. As a result, participating in the longitudinal study

[16] For more on the lessons learned by the BU team from developing MPC solutions, see: Andrei Lapets et al. (2019) 'Role-Based Ecosystem for the Design, Development, and Deployment of Secure Multi-Party Data Analytics Applications' https://ieeexplore.ieee.org/document/8901614

[17] For more information about outcomes and lessons learned, see David Buckley's (2023) 'Boston Women's Workforce Council: Measuring salary disparity using secure multi-party computation' https://unstats.un.org/wiki/pages/viewpage.action?pageId=150012020

[18] For more information about Equal Pay for Equal Work by BWWC, see BWWC (2020) 'Interventions Report' https://www.boston.gov/sites/default/files/file/2020/12/BWWC%202020%20Interventions%20Report.pdf

can have additional reputational value for businesses if they perform well and choose to disclose this information.

Since designing and implementing the web-based MPC protocol for BWWC, BU has adapted the model to support the Pacesetters Initiative at the Boston Chamber of Commerce[19] to produce secure data analysis of business spending on local, minority-owned businesses to promote economic equity in 2018. They also applied a variation to produce the MPC protocol for Callisto[20], a platform that uses MPC to empower survivors of sexual violence, amongst other things.

Examples of fully deployed MPC protocols are still relatively rare, despite growing attention and curiosity towards experimenting with this technology. BU's thinking, in particular around the balancing of privacy and usability are useful insights to influence the successful adoption of new MPC projects.[21] However, a key challenge for developing MPC as a service model is that it is still difficult to customise a cryptographic protocol to a wide set of use cases due to an insufficient number of precedents. Researchers at BU, in collaboration with other organisations, are continuing to develop open-source libraries, frameworks, and systems[22] to enable the implementation and deployment of applications that employ secure multi-party computation and make it accessible and scalable for lesser-skilled or less-resourced teams in the future.[23]

MPC provides a valuable method to analyse data for the public good without disclosing individual data input. In theory, positive outcomes could lead to the creation of policies mandating businesses to share data more openly, similar to how businesses with more than 250 employees are mandated to report on gender pay equity in the UK.[24] While this may not be applicable or achievable for our case study, the implementation of the MPC to benchmark and analyse racial and pay gap equity in the Greater Boston area inspired policy changes including the Foundation for evidence-based policy[25] – or OPEN Government Data Act, Pub.L. 115–435 – a United States law that requires the federal government to modernise its data management practices.

The success of the BWWC case study has also been cited in further policy discussions, such as that on the Student Right to Know Before You Go Act[26], which showcased the potential for new privacy-protecting technologies to empower students as consumers in 2017.[27] While this Bill was not ultimately passed, it sought to address an informational gap, in that little or inaccurate data had been available to enable students to make informed choices about their prospects of employment after graduating from certain degrees from certain

---

[19] For more on the Pacesetters Program, see: Greater Boston Chamber of Commerce (2022) 'Pacesetters: Economic Inclusion through Supplier Diversity' https://bostonchamber.com/networks/pacesetters/

[20] For more information on the Callisto project, see: Callisto (2024) 'Callisto - Ending serial sexual assault… one match at a time' https://www.projectcallisto.org/

[21] See previous footnotes for further details.

[22] For resources to enable the implementation and deployment of MPC, see: GitHub (2024) 'multiparty' https://github.com/multiparty

[23] To find out more on these various efforts, see: multiparty.org https://multiparty.org/#:~:text=Researchers%20at%20Boston%20University%2C%20together,in%20accessible%20and%20scalable%20ways

[24] Government Equalities Office (2023) 'Gender pay gap reporting: guidance for employers' https://www.gov.uk/government/publications/gender-pay-gap-reporting-guidance-for-employers#:~:text=You%20must%20report%20and%20publish,employees%20on%20your%20snapshot%20date.

[25] Congress.Gov 'H.R.4174 - Foundations for Evidence-Based Policymaking Act of 2018' https://www.congress.gov/bill/115th-congress/house-bill/4174

[26] Congress.Gov 'S.3952 - Student Right to Know Before You Go Act of 2022' https://www.congress.gov/bill/117th-congress/senate-bill/3952

[27] For additional information, see Azer Bestavros presentation (2017) 'Sharing Knowledge without Sharing Data: Platforms for resolving the false dichotomy between privacy and utility of information' https://www.youtube.com/watch?v=P2MmO458xu4

institutions of higher learning. For instance, states and private websites tried to publish such information, but the data typically only examined first-time, full-time students or students remaining within a given state after college. Together, these instances demonstrate that successful interventions can spur broader consideration of alternative solutions in other cases in which data access is not addressable by traditional means or efforts.

Conclusion

This case study illustrates the deployment of MPC as a valid alternative to sharing sensitive data via a trusted arbiter. It provides contextual and practical guidance on how MPC enables shared insights based on sensitive data without disclosing individual contributions.

The design of this specific MPC framework is tailored to meet the needs and constraints of its stakeholders, government agencies, non-profit organisations, and social scientists. Its key features are:

- the deployment of the lesser-established client-server MPC model, whereby infrastructure, software development, and computation are offered as a service by BU and;
- the introduction of a semi-honest MPC protocol that balances security and usability, taking into account shared goals and individual incentives for participating in this longitudinal project.

We believe that there is value in breaking down real-world PET case studies into structures and components to illustrate how these novel technologies work in practice. Through doing so, this facilitates closer inspection and understanding of the practical impacts that they can have in enabling collaborative analysis without disclosing underlying data and putting data subjects at risk. It is our hope that explainers such as this and the corresponding annotated diagram can be useful to enable their wider adoption and contribute towards the documentation of case studies.

Get in touch

Please get in touch, if you want to share any feedback, or need help sharing sensitive data, we welcome all forms of input.

Acknowledgments

The writing of this report would not have been possible without the help and insights provided by Kim Borman of BWWC, William Tomlinson of Boston University, and Andrei Lapets of Magnite. We are incredibly grateful for the time they contributed to answering our questions on their experiences of both developing and deploying this use case.

We would also like to express our thanks to our colleagues at the ODI and Philpott Design who have contributed their time and assistance to the creation of this written report and our annotated diagram. This includes Charlotte Mitchell, Elena Simperl, Emma Thwaites, David Dinnage, Zoe Philpott, Adrian Philpott and Ian Dutnall.