MPC-based Post-Quantum Zero-Knowledge Proof Systems with Fast Verification

A Data Management Plan created using DMPonline.be

Creator: Cyprien Delpech de Saint Guilhem

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Grant number / **URL:** 1266123N

ID: 199061

Start date: 01-10-2022

End date: 30-09-2025

Project abstract:

To this present day, obtaining trust has required people to reveal private information. Think of sharing one's date of birth to prove one is over 18 or sharing one's financial records to prove that one is debt-free. For businesses that rely on private proprietary software, proving that they adhere to security and privacy regulations requires auditing by independent experts. The only way to protect the business's property is then to legally bound auditors to silence with the risk that the knowledge
might still be revealed one day. But what if this trust could be provided without the risks of giving up privacy?

Zero-knowledge proof systems are a cryptographic tool that provide just this: both trust and privacy. Invented in the 1980s, proof systems have served only within bigger cryptographic protocols, and never as applications of their own until very recently (e.g., in blockchains). To provide transformative services within our digital societies, these must be scaled up and optimized to work with much larger programs.

This project will use the modern technology of multiparty computation, optimised for large programs, to construct such proof systems. To addition to their scale, these systems will be provably post-quantum secure, to guarantee a life-time of security. Furthermore, they will focus on concrete efficiency and fast verification.

Last modified: 27-04-2023

MPC-based Post-Quantum Zero-Knowledge Proof Systems with Fast Verification
FWO DMP (Flemish Standard DMP)

1. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA	*Please choose from the following options:* • .por, .xml, .tab, .cvs,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
Code and scripts	Various software programs and scripts to either implement zero-knowledge proof systems or automatically generate parameters and estimate performance.	Generate new code and scripts.	Digital.	Software.	Source code (most like Python, C, C++ or Rust).	< 1GB.	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

During the project I may re-use existing software libraries, in accordance with the licenses under which they are published. As the software is not written yet, I cannot point to these libraries at this stage.

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- Yes

Should a suitable application be found, it could be possible for the software code generated during this project to be commercially valorized as software-as-a-service. However, the code itself would only be of an "academic" standard and would require significant improvement before being commercially viable.

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- No

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- Yes

The software libraries that this code would use may be published under certain restrictive licenses. Any software in this project that builds upon these libraries will be therefore also be subject to licenses that take into account the restrictions of the licenses therein.

2. Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g., in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).

Any software code or script produced during this project will be documented in-file, according to the practices of the relevant language. Also, readme files will be included alongside, describing how to compile and/or operate the software and how to interpret the output produced.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- No

Any software code made publicly available will be stored on the github.com website, which is an industry reference for open-source software, on the COSIC research group's github.com page (https://github.com/KULeuven-COSIC). It will be named according to the research publications that refer to it, and links to the source code's page will also be included in such publications. Any software code kept internally will be similarly named and referred to in private documentation.

3. Data storage & back-up during the research project

Where will the data be stored?

The software code and scripts will be primarily stored on my COSIC file system (which is managed and synchronized by the departmental ESAT IT team). A secondary storage will be the COSIC research group's GitLab infrastructure, which is private and also maintained by the ESAT department of the KU Leuven.

How will the data be backed up?

Both the data on my COSIC file system and the COSIC GitLab are automatically backed-up by the procedures put in place by the ESAT IT team and the KU Leuven ICTS. Projects that are completed or require archiving will have their GitLab ownership transferred to a permanent member of COSIC's administrative staff so that access can be granted for future use by other researchers. Finally, upon completion of the FWO research project, physical copies of the data will also be made for safekeeping.

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely.
If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.

- Yes

Several gigabytes of space are made available to members of staff, which is sufficient to store the < 1GB of source code and scripts that this project will produce.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

During the course of the project, software code and scripts will be protected by access rights and stored under my accounts on the ESAT network and within the GitLab

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

The costs of the GitLab maintenance are covered by the COSIC research group as part of the general IT costs. The use of GitHub for code released publicly is free of charge.

4. Data preservation after the end of the research project

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

All of the data (software code and scripts) will be maintained for at least 10 years according to the KU Leuven RDM policy.

Where will these data be archived (stored and curated for the long-term)?

These data will be archived in several locations. First, the ESAT GitLab will hold the code and scripts so that it can be easily retrieved for further use by researchers; this data is backed up. Second, a copy of the totality of the data will be stored in the KU Leuven RDR. Finally, another copy will be stored on disk as a backup solution.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

Given the small size (< 1GB) of the data, the storage cost will either be free of charge, or already covered by COSIC's or KU Leuven's IT costs.

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- Yes, in an Open Access repository
- Yes, in a restricted access repository (after approval, institutional access only, …)

Projects that are deemed advanced enough will be made publicly available on COSIC's GitHub page, with appropriate licenses.
All projects will be stored on COSIC's GitLab so that it can be shared (conditioned on internal approval) with other researchers.

If access is restricted, please specify who will be able to access the data and under what conditions.

For the projects stored internally on GitLab, approval will be granted to COSIC researchers, or affiliated, under the condition of legitimate use for further research. This acts as a sanity check so that the technology is not removed from COSIC's internal control without further approval.

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.

- No

Some open-source software libraries used during this project come with licenses that limit their commercial use, but no proprietary code will be used so that all projects can be made open-source without requiring further authorisation.

Where will the data be made available? If already known, please provide a repository per dataset or data type.

The data will be made publicly available on COSIC's GitHub account because this is a widely-used repository for open-source code. Alternative access may be made available via COSIC's webpages or other KU Leuven repositories.

When will the data be made available?

If software projects are in enough of a developed state, they will be made available upon publication of research results.

Which data usage licenses are you going to provide? If none, please explain why.

The data that is made publicly available will use appropriate licenses to permit the use of the code for further research, but prohibit commercial use, in addition to respecting any licenses of libraries used within.

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- No

What are the expected costs for data sharing? How will these costs be covered?

Given the small size (< 1 GB), data sharing is expected to be free of charge.

6. Responsibilities

Who will manage data documentation and metadata during the research project?

Myself (Cyprien Delpech de Saint Guilhem)

Who will manage data storage and backup during the research project?

Myself (Cyprien Delpech de Saint Guilhem)

Who will manage data preservation and sharing?

Myself (Cyprien Delpech de Saint Guilhem) during the project, and the COSIC permanent staff after the project's completion.

Who will update and implement this DMP?

Myself (Cyprien Delpech de Saint Guilhem)