Hardware-software co-designs for end-to-end security

A Data Management Plan created using DMPonline.be

Creator: Lesly-Ann Daniel https://orcid.org/0000-0002-2772-3722

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Grant number / URL: 12B2A24N

ID: 207015

Start date: 10-01-2023

End date: 10-01-2026

Project abstract:

Modern hardware heavily relies on optimizations to improve performance. Unfortunately, the last two decades of research on microarchitectural side-channels has shown that these optimizations often come at the expense of security. Hardware-software co-design are a promising solution to mitigate these attacks at an acceptable cost. However, research on hardware software co-design is challenging. Designing hardware defenses that provide provable end-to-end security, requires to (1) formalize the guarantees of these defenses with hardware-software security contracts, (2) implement software that are secure according to the contract, and (3) validate that the hardware implementation complies with the contract.

The general goal of this project is to support the development of hardware-software co-design for end-to-end security. First, we will develop compiler that transparently enforce software-level properties required by hardware defenses. Second, we will build a modular toolchain to validate that hardware defenses comply with their security contract. Finally, we will investigate new security conditions to enable end-to-end provable security for cryptographic primitives.

Last modified: 03-05-2024

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, .csv,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
Source code	Source code of the software developed as part of the project	Either new software, written from scratch, or by modifying existing open source projects (Binsec, angr, Jasmin, LLVM, Revizor, Proteus, etc.)	Digital	Software	C, C++, Ocaml, Python, Scala, etc.	< 1GB	
Binary code	Compiled from the aforementioned source code	Newly generated data	Digital	Compiled software	x86 or RISC-V assembly	< 10GB	
Experimental data	Performance data collected from program executions	Newly generated data	Digital	Experimental data	Compressed .txt or .csv files	< 50GB	
Manuscripts	Papers and technical reports	Newly generate data	Digital	Other	.pdf, .tex	< 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:

- LLVM: https://github.com/llvm/llvm-project
- Binsec: https://github.com/binsec/binsec
- angr: https://github.com/angr/angr
- jasmin: https://github.com/jasmin-lang/jasmin
- Revizor: https://github.com/microsoft/sca-fuzzer
- Proteus: https://github.com/proteus-core/proteus

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.

- No

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.

- No

All third-party software used in this project is distributed under an open-source license (Apache, LGPL-2.1, BSD-2, MIT), which permits modification and redistribution of the modified software (with the original license).

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).

- Source code will be documented according to best practices in software engineering (inline comments, function documentation, README file with installation and running instructions),
- Experimental data will be documented in published papers and corresponding technical reports,
- Docker containers will be set up for research artifacts, to ensure reproducibility and working build environments.

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

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

Where will the data be stored?

The infrastructure of KU Leuven includes communication and collaboration platforms and repository services with sufficient storage space for long-time storage and sharing of development artifacts (e.g., GitLab), which are managed by teams of full-time system administrators.
KU Leuven also operates the Lirias digital repository that stores, preserves, and provides open access to research papers and reports.

How will the data be backed up?

The project will rely on standard backup provided by KU Leuven ICTS; the data will be stored on KU Leuven servers with automatic daily backup procedures.

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

The size of data generated in this project is relatively small and well within the limits of the capacity.

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

Data will be protected by the standard access control mechanisms of the communication and collaboration platforms and repository services.

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

Data will be stored on the standard infrastructure offered by KU Leuven to researchers. No additional costs will be incurred.

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 research data will be retained for at least 5 years after the end of the project.

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

At the end of the project, the data will be archived and stored on the KU Leuven institutional repository (RDR). The repository guarantees a minimal retention period of 10 years during which the data is stored in RDR with the necessary backups and fixity checks to ensure continuity and perpetual availability of the data during those first 10 years.

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

Data will be stored on the standard infrastructure offered by KU Leuven to researchers. No additional costs will be incurred.

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

Open-source development artifacts will be made available in public Git repositories (on KU Leuven GitLab and/or GitHub).
Publications will be made available on the KU Leuven Lirias repository.

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

Question not answered.

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

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

Open-source development artifacts will be made available in public Git repositories (on KU Leuven GitLab and/or GitHub).
Publications will be made available on the KU Leuven Lirias repository.

When will the data be made available?

Upon publication of the research results.

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

Extensions to third-party software will be shared following the license of the original software.
New software developed for this project will be distributed under an open-source license (Apache, LGPL-2.1, BSD-2, MIT), allowing third-party modifications and redistribution.

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.

- Yes

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

Data will be stored on the standard infrastructure offered by KU Leuven to researchers or on free GitHub repositories. No additional costs will be incurred.

6. Responsibilities

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

Lesly-Ann Daniel

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

Lesly-Ann Daniel

Who will manage data preservation and sharing?

Lesly-Ann Daniel

Who will update and implement this DMP?

Lesly-Ann Daniel