Trust-SEV: Hardware Root of Trust for Smart Electric Vehicles

A Data Management Plan created using DMPonline.be

Creator: Wouter Hellemans

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Grant number / URL: 1SH3824N

ID: 204500

Start date: 01-11-2023

End date: 31-10-2027

Project abstract:

Following the pledges for carbon-neutrality and the Paris Climate Agreement, countries have pushed for smart electric vehicles (SEVs). With over 100 million lines of code managing multiple safety-critical tasks like steering, braking, and self-driving, SEVs are becoming increasingly reliant on software. As software complexity grows, the number of issues affecting automotive safety-critical processes grows. Overwhelming SEV growth not only opens the door for better user satisfaction, but it also creates a fear of potential data theft and cyber-attacks like the Jeep-hack and the Charging station attack. With the increased availability of wireless interfaces and software integration, an attacker can get access to a vehicle's communication network and exploit vulnerabilities in Electronic Control Units (ECUs). Vehicle makers have logically separated in-vehicle networks into sub-networks to safeguard important systems against such attacks. However, attackers can still physically exploit numerous sub-networks and the CAN-bus. Thus, maintaining the sanity of the systems is an exigent task. During my PhD, I aim to develop intrusion detection systems and hardware-level protection mechanisms that take on the challenge of protecting in-vehicle networks.

Last modified: 24-04-2024

Trust-SEV: Hardware Root of Trust for Smart Electric Vehicles
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
IVN traffic datasets	Datasets containing (labeled) network traffic of an in-vehicle network (IVN)	Reuse existing data	Digital	Observational	.csv, .log, .hex, .txt	<1TB	NA
IVN testbed	Code that is responsible for replaying existing datasets on a realistic IVN test platform consisting of several ECUs. Furthermore, this code is able to inject attacks on existing datasets. This platform is used to evaluate the performance of the proposed solutions.	Generate new data	Digital	Software	.json, .py, .c, .cpp	<1GB	NA
ML codes	Machine learning (ML) toolflow that enables the training/evaluation/inference of a deep learning model (new or reused) on a given dataset. This toolflow is also responsible for generating the hardware representation of the chosen ML models.	Generate new data + Reuse existing data	Digital	Software	.py, .cpp, .c, .json, ipynb, txt	<1GB	NA
ML results	The generated ML models and the results of their evaluation. This includes the classification results as well as the hardware parameters (e.g., size, throughput) of the ML models.	Generate new data	Digital	Simulation data	.json, .csv, .pth, .onnx, .txt, .xlsx	<100GB	NA
RA codes	Code to implement and evaluate the proposed RA schemes on a realistic software/hardware platform.	Generate new data + Reuse existing data	Digital	Software	.py, .cpp, .c, .vhd, .v, .S	<100MB	NA
RA results	Performance results of the implemented RA schemes. These results include the timings (e.g., scalability, latency), but also the hardware parameters (e.g., size, hardware requirements)	Generate new data	Digital	Simulation data	.csv, xlsx	<1GB	NA
Stitching code	This code is responsible for establishing the interaction between proposed ML and the proposed RA solution. Furthermore, this code enables the solution to be integrated in a realistic test platform and as such is responsible for providing the solutions with data from the IVN.	Generate new data	Digital	Software	.py, .cpp, .c, .vhd, .v	<1GB	NA

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:

Throughout the research, several datasets containing in-vehicle network traffic will be used. Examples include:
Car-hacking dataset: https://ocslab.hksecurity.net/Datasets/car-hacking-dataset
ROAD dataset: https://0xsam.com/road/
Can-train-and-test dataset: https://data.dtu.dk/articles/dataset/can-train-and-test/24805533

Furthermore, several existing ML models and RA schemes will be evaluated. These models and schemes will be adopted from the state of the art and will be selected after a thorough literature analysis.

Please note that new datasets and models/schemes are proposed on a regular basis. Therefore, this list can be subject to changes.

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

The main valorization opportunity of this research lies in the proposed RA schemes and in the root of trust that combines ML with RA. Current RA schemes in literature do not take into account the physical architecture of a modern IVN. The development of a lightweight RA scheme that is enhanced with an ML model and takes into account the network architecture, might have direct applicability in modern vehicles.

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

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).

All source code will be accompanied by README files and other documentation (e.g., comments) that explain the structure and utilization of the code. The results of the ML and RA experiments will be accompanied by reports and dedicated documentation documents indicating how the data should be interpreted (e.g., the experimental settings that were used to obtain the data).

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

In the computer code, comments will be integrated to improve readability. README files will be included to explain the structure of the generated code, the compilation order and the overall architecture. Furthermore, the results of the ML experiments will be contained in the .json format, which is a standardized format for date exchange.

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

Where will the data be stored?

During the development process, the artifacts produced will be stored on (1) an offline local development copy, (2) a private Github repository, and (3) KU Leuven OneDrive for Business cloud storage. Furthermore, when needed, source code will be stored on Gitlab managed by KU Leuven ESAT, enabling efficient version control and sharing with people inside the organisation. Upon publication or immediately after the end of the project, the code and results will be made publicly available on Github with corresponding documentation. An additional archive of the Github repository will be created on Zenodo.

How will the data be backed up?

During the development, all digital artifacts will be stored on KU Leuven OneDrive for Business cloud storage. This storage option provides version control managed by Microsoft, ensures that the data is stored within Europe, and is additionally backed up by KU Leuven. Additionally, source code will be stored on ESAT-Gitlab/Github for efficient version control. With daily commits and local checkouts of the repositories, data loss should be limited.

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

For all the mentioned storage systems, the storage capacity provided by KU Leuven covers the estimated amounts of data.

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

All ICT solutions at KU Leuven are subject to the university-wide ICT information security standards. As such, the OneDrive for business cloud storage is suited for confidential data. Root access rights to the artifacts will only be given to the PIs and and the involved PhD fellow. However, other researchers participating in the project might be granted access rights to individual modules, as deemed relevant for the research.
Upon publication or immediately after the end of the project, the Github repositories (protected by Multi-Factor Authentication) with the artifacts and documentation will be open source. The PhD thesis and papers will not require any access control and will be openly available in KU Leuven Lirias.

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

All mentioned storage systems are for free.

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...).

According to KU Leuven RDM policy, all data described in the research data summary will be retained for at least a period of 10 years.

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

We will maintain our Github repository containing the digital artifacts with corresponding documentation.

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

No 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

All computer code, results and accompanying documentation will be stored in dedicated Github repositories under a relevant licence such as the Apache 2.0 license.

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

NA

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.

All computer code, results and accompanying documentation will be stored in dedicated open-source Github repositories under a relevant licence such as Apache 2.0 license.

When will the data be made available?

Upon publication of the research results or immediately after the project.

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

All computer code, results and accompanying documentation will be stored in dedicated open-source Github repositories under a relevant license such as Apache 2.0 license.

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

A permanent identifier is added to the data upon deposit in the repository.

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

No costs.

6. Responsibilities

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

Wouter Hellemans, Md Masoom Rabbani, Jo Vliegen, and Nele Mentens

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

Wouter Hellemans, Md Masoom Rabbani, Jo Vliegen, and Nele Mentens

Who will manage data preservation and sharing?

Wouter Hellemans, Md Masoom Rabbani, Jo Vliegen, and Nele Mentens

Who will update and implement this DMP?

Wouter Hellemans and Nele Mentens