New techniques for the cryptanalysis of hash functions

A Data Management Plan created using DMPonline.be

Creator: Tim Beyne

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Grant number / URL: 1274724N

ID: 206080

Project abstract:

Cryptographic hash functions map arbitrary length messages to fixed-length message digests. They are an essential building block in modern cryptography. For example, virtually all digital signature schemes produce a signature only on the message digest. To guarantee the security of this approach, it must be difficult to find two distinct messages with the same digest. Such a message pair is known as a collision and could be used to trick a victim into signing one message while being shown only the other.

The goal of this proposal is to develop new techniques for the cryptanalysis of hash functions, and collision attacks in particular. These techniques will be used to identify weaknesses in existing hash functions and to construct secure and efficient alternatives. The starting point for this proposal is a new development in differential cryptanalysis, the method that underlies existing collision attacks on hash functions. Specifically, in joint work with Vincent Rijmen, published at CRYPTO 2022, I introduced "quasidifferential trails" to compute the fixed-key probability of differentials in block ciphers. Previously, only key-averaged probabilities could be determined systematically, and only in an idealized model. Since hash functions are often constructed from block ciphers by repurposing the key as an additional input that can be controlled by the attacker, quasidifferential trails lead to new possibilities in the analysis of hash functions.

Last modified: 02-04-2024

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	
Simulated data	Results of numerical or statistical simulations/experiments.	New	Digital	Simulation data	.txt, .csv, binary	< 100GB	
Software for simulations	Source code of simulation/experimentation software.	New	Digital	Software	.c, .cpp, .py, .sage	< 1GB	
Automated cryptanalysis tools	Software to solve combinatorial optimization problems related to crytanalysis.	New	Digital	Software	.c, .cpp, .py, .sage	< 1GB	
Electronic notebooks	Python or Sage notebooks.	New	Digital	Other	.ipynb	< 1 GB	
LaTeX code papers/presentations	Source code of papers and presentations.	New	Digital	Other	.tex	< 1 GB	

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:

No data will be reused, except for publicly available software such as Sage (https://sagemath.org) and source code accompanying papers in the literature (various sources).

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

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 data produced in this project will be synthetic. The software used to generated the data will be kept, along with documentation (in text or markdown format) to explain the purpose of the software and its dependencies. Makefiles and installation instructions will be provided where applicable.

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?

All data will be stored on ESAT (department of electrical engineering) systems.

How will the data be backed up?

An on-site backup of the ESAT systems is made daily. Off-side backups are made monthly, and yearly backups are stored off-sited for the long term.

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

Sufficient storage is available on ESAT systems.

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

The data will be stored with appropriate usergroup permissions.

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

The costs of storing the data are external to the project (covered by the department).

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 data will be retained for at least five years.

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

The data will be archived on ESAT-managed systems.

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

The costs of storing the data are external to the project (covered by the department).

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 source code and data necessary to reproduce the results of publications will be made available, either as a git repository (hosted on the

ESAT GitLab instance), or as raw files that can be downloaded from the ESAT webservers.

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

Access will not be restricted.

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.

- Yes, Ethical aspects

Some of the software that will be developed may be used to generate collisions for widely used hash functions, which could be a concern for many security applications. Software will only be made available if it does not affect the security of widely used systems. If any vulnerabilities are found, a process of responsible disclosure will be followed.

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

On the ESAT GitLab instance, on through the ESAT webservers.

When will the data be made available?

At the time of publication of the corresponding scientific results.

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

Creative Commons CC BY, except in cases where this would conflict with existing licenses.

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?

None, the cost will be external to the project (the datasets are small).

6. Responsibilities

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

Tim Beyne

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

Saartje Verheyen (COSIC, i.e. within the research group), Frank Schoeters (ESAT)

Who will manage data preservation and sharing?

Saartje Verheyen (COSIC), Frank Schoeters (ESAT)

Who will update and implement this DMP?

Tim Beyne