"These results must be false": A usability evaluation of constant-time analysis tools

Marcel Fourné, *Paderborn University and MPI-SP;* Daniel De Almeida Braga, *Rennes University, CNRS, IRISA;* Jan Jancar, *Masaryk University;* Mohamed Sabt, *Rennes University, CNRS, IRISA;* Peter Schwabe, *MPI-SP and Radboud University;* Gilles Barthe, *MPI-SP and IMDEA Software Institute;* Pierre-Alain Fouque, *Rennes University, CNRS, IRISA;* Yasemin Acar, *Paderborn University and George Washington University*

https://www.usenix.org/conference/usenixsecurity24/presentation/fourne

This artifact appendix is included in the Artifact Appendices to the Proceedings of the 33rd USENIX Security Symposium and appends to the paper of the same name that appears in the Proceedings of the 33rd USENIX Security Symposium.

August 14–16, 2024 • Philadelphia, PA, USA

978-1-939133-44-1

Open access to the Artifact Appendices to the Proceedings of the 33rd USENIX Security Symposium is sponsored by USENIX.

USENIX Security '24 Artifact Appendix: "These results must be false": A usability evaluation of constant-time analysis tools

Marcel Fourné[1,4], Daniel De Almeida Braga[2], Jan Jancar[3], Mohamed Sabt[2], Peter Schwabe[4,5], Gilles Barthe[4,6], Pierre-Alain Fouque[2], and Yasemin Acar[1,7]

[1]*Paderborn University, Paderborn, Germany*
[2]*Rennes University, CNRS, IRISA, Rennes, France*
[3]*Masaryk University, Brno, Czechia*
[4]*MPI-SP, Bochum, Germany*
[5]*Radboud University, Nijmegen, The Netherlands*
[6]*IMDEA Software Institute, Madrid, Spain*
[7]*George Washington University, Washington D.C., United States of America*

A Artifact Appendix

A.1 Abstract

This artifact contains installation scripts, documentation and content for recreating surveys necessary to replicate and/or extend the study done for the paper.

A.2 Description & Requirements

In subfolder "audit_tasks" and "repair_tasks" one may find the task sets used in the study as well as solutions for the tools used in the study. Subfolder "surveys" contains the *LimeSurvey* files and PDF renderings of the surveys used during the study. "tutorials" contains documentation for each tool, "page" the web page and automation for creating user accounts for the study. "README.md" files are located in some subfolders.

A.2.1 Security, privacy, and ethical concerns

Security risk is in installation of third-party code, also when running the installation scripts. The privacy risk is due to network connections when downloading the code for installation and third parties monitoring the download activity. Ethically, there are no special concerns of note.

A.2.2 How to access

The artifact is hosted on Zenodo, at https://zenodo.org/records/11143400. Additional software dependencies pulled in by the installation scripts are hosted by third parties.

A.2.3 Hardware dependencies

We used a set of 7 Ubuntu 20.04 virtual machines in our study, together with a web server and a LimeSurvey server. The hardware architecture of the VMs needs to be x86_64 and x86_32 code execution needed to be available for one tool. This architecture requirement may change depending on CT analysis tools and versions used in recreating and/or adapting the study.

A.2.4 Software dependencies

LimeSurvey for the surveys, *bash*, *apt*, and *patch* on a Ubuntu 20.04 are the basic requirements for the installation scripts for the tools. The web page in subdirectory "page" requires *Flask* and therefore *Python*. The installation scripts pull in dependencies based on requirements of each tool.

A.2.5 Benchmarks

None.

A.3 Set-up

Artifact can be downloaded with *GNU wget* and unpacked with *GNU tar* and *gzip*. Installation scripts require *GNU bash*.

A.3.1 Installation

Artifact contains installation scripts in folder "installation_scripts". Running each of those scripts should install the corresponding tool into the current working directory.

A.3.2 Basic Test

Some CT analysis tool installation scripts run basic functionality tests themselves, others provide necessary steps to run tests where appropriate, while the rest requires solving the tasks in the "repair_tasks" folder. Also in that folder is a subfolder "solutions" which contains tool subdirectories which each contain a "run_tests.sh" script that can be used to check the provided sample solutions with the installed tool.

A.4 Notes on Reusability

To reproduce the study, a hardware and software setup is not sufficient, but participants fulfilling the requirements described in the paper need to be recruited. Their number depends on how many tools should be compared, but for each possible combination of tools used, at least one participant should test that combination to avoid unseen interactions between testing different tools. When extending the survey for another CT analysis tool, installation scripts and tutorials should be provided. The set of repair tasks can be easily extended for interesting problems, for example when checking for Spectre-style vulnerabilities with other tools. Time on the audit tasks should not be too generous, since participants should be able to find something after a few hours and the study is only interested in them finding something with a tool at all, not every possible CT violation.

A.5 Version

Based on the LaTeX template for Artifact Evaluation V20231005. Submission, reviewing and badging methodology followed for the evaluation of this artifact can be found at https://secartifacts.github.io/usenixsec2024/.