USENIX Security '24 Artifact Appendix: Towards Privacy and Security in Private Clouds: A Representative Survey on the Prevalence of Private Hosting and Administrator Characteristics

Lea Gröber [†‡] Simon Lenau[†] Rebecca Weil[†] Elena Groben[‡]
Michael Schilling[†] Katharina Krombholz[†]

[†]CISPA Helmholtz Center for Information Security
[‡]Saarland University

A Artifact Appendix

A.1 Abstract

The artifact contains materials on survey instruments, analysis (R scripts), and anonymized data. Further, we provide a dockerfile for convenient execution of analysis scripts.

A.2 Description & Requirements

The artifact contains a `Dockerfile` that provides an execution environment including the anonymized data and analysis scripts. All analyses can be run by simply executing `./main.R` from `/scr` in the container.

A.2.1 Security, privacy, and ethical concerns

The evaluation scripts or data pose no security risk to the evaluators' machine. There might be privacy implications due to third parties such as Docker and R dependencies needed to execute the scripts.

A.2.2 How to access

https://projects.cispa.saarland/lea.groeber/usenix24-sh-prevalence/-/tree/a82ca06181c1e922e70f638f8b422f17ec78f222

A.2.3 Hardware dependencies

None.

A.2.4 Software dependencies

We provide a dockerfile for easy access to a functional environment containing all dependencies. We recommend relying on it to execute the analyses. Docker is available for all operating systems. In case you do not want to rely on docker, you need to install the latest version of R. Refer to the `Dockerfile` in the artifact repository for a list of all dependencies.

A.2.5 Benchmarks

None.

A.3 Set-up

You need to install docker [https://www.docker.com/get-started/]. We tested the setup with docker `v4.25.0`.

A.3.1 Installation

Clone or download the repository from the link provided in A.2.2.

A.3.2 Basic Test

Navigate into the artifact repository. Use `docker build .` to create a docker image from the `Dockerfile` included in the artifact repository. On our test computer (12th Gen Intel(R) Core(TM) i5-12400F 2.50 GHz, 32.0 GB RAM) it took 900 seconds to create the docker image which was 2.95 GB big. The image will have no name, but the output will provide the ID of the image. To start the container execute `docker run -it <containerID> /bin/bash`. Please refer to the official documentation of the docker project for further information.[1]

A.4 Evaluation workflow

After having run `docker run -it <containerID> /bin/bash` the command line switches "inside" the container. Run `./main.R` from within `/src`. This executes all analysis

[1] https://docs.docker.com/guides/docker-concepts/building-images/build-tag-and-publish-an-image/

scripts. Results are stored in `/data` in folders following the naming convention of the analysis scripts. Log outputs are stored in `/data`, too. Refer to the `README` for a high-level overview of all scripts and output files. For convenient browsing and in-depth insights into all R scripts, refer to `src/R-code.html`.

A.4.1 Major Claims

The results of this paper are based on descriptive statistics, confidence intervals, Wald- for binary and t-Tests for continuous variables, and a regression analysis. Refer to Sections 5 and 8 in the paper. The results folders, obtained after running the analysis scripts, contain raw `.RData` files and `.csv` files for easy access. We also, generated `.tex` files for the tables in the paper. Below we explain how the key results mentioned in the papers' abstract can be reproduced.

(C1): We estimate an upper bound of 8.4% private self-hosters in the U.S. population. This is derived by the experiment (E1) described in Section 4.5 in the paper whose results are illustrated/reported in Table 8.

(C2): We find that self-hosters are not more privacy-, or security-sensitive than the general population. Instead, we find that IT administration skills, IT background, affinity for technology interaction, and "maker" self-identity positively correlate with self-hosting behavior. This is shown by experiment (E2) described in Section 7 in the paper whose results are reported in Table 5.

(C3): Websites are the most common use case for self-hosting, predominately running on home servers. All other use cases were equally frequent. Alt This is shown by experiment (E3) described in Section 5.2 in the paper whose results are reported in Table 2.

A.4.2 Experiments

(E1): [Prevalence] [20 human-hour + <1 compute-min]: Run `/src/[F]_SH_Distribution/001_SH-Prevalence.R` with all preceding `.R` files to obtain the prevalence (C1).

(E2): [Individual Characteristics] [10 human-hour + <1 compute-min]: Run `/src/[G]_SH_Prediction/002_Regression_Models.R` with all preceding `.R` files to obtain regression model underlying (C2).

(E3): [Use Case Distribution] [10 human-hour + <1 compute-min]: Run `src/[F]_SH_Distribution/005_SH-Categories-Tools.R` with all preceding `.R` files to obtain the use case distribution (C3).

A.5 Notes on Reusability

We explain data types and variable mappings of both surveys in `/data/Codebook_survey1_survey2`. Further, we include overviews of the survey flow and scales in folder `survey_instruments`. There, you also find details on use case and tool selection as part of the Survey 1 design.

A.6 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/.