



# A Large-scale Investigation into Geodifferences in Mobile Apps

Renuka Kumar, Apurva Virkud, Ram Sundara Raman, Atul Prakash,  
and Roya Ensafi, *University of Michigan*

<https://www.usenix.org/conference/usenixsecurity22/presentation/kumar>

This artifact appendix is included in the Artifact Appendices to the Proceedings of the 31st USENIX Security Symposium and appends to the paper of the same name that appears in the Proceedings of the 31st USENIX Security Symposium.

August 10–12, 2022 • Boston, MA, USA

978-1-939133-31-1

Open access to the Artifact Appendices  
to the Proceedings of the 31st USENIX  
Security Symposium is sponsored  
by USENIX.



## A Artifact Appendix

### A.1 Abstract

We make the artifacts of our large scale geographical study on the geodifferences in mobile apps available to the research community. In this research, we crawl a set of 5,684 globally popular app binaries, their metadata, and privacy policies from Google Play store in 26 countries. We open-sourced the crawlers we used to download these apps, their metadata, and privacy policies, and provide documentation on how to setup these crawlers in a public *GitHub* repository. In addition, we provide a way for researchers to request access to the actual app binaries and privacy policies. The metadata of these apps and the error messages that we obtain during our crawls are both published and available for download.

### A.2 Artifact check-list (meta-information)

- **Run-time environment:** The code requires a Python 3 environment. The repository contains three crawlers (for metadata, APK, and privacy policies) and it is recommended to create separate environments for each crawler; we have provided Pipfiles to install Python dependencies to each environment. Instructions are located in the repository READMEs.
- **Output:** Depending on which code is run, the crawler will download the respective data (metadata, APK, privacy policy) to the output folder specified by the user. Note that the crawler is making requests to Google Play (or sites hosting individual privacy policies) and may have nondeterministic output depending on the time of the request.
- **Publicly available:** Our code is publicly available on GitHub and we provide a public Google Drive URL (link in the repository README) to download our app metadata and crawl error messages. The privacy policy and APK files are available upon request.
- **Code licenses:** The code is released under the GNU General Public License. Full licensing information is provided in the GitHub repository.
- **Archived:** A stable link to our GitHub repository can be found at the following link: <https://github.com/censoredplanet/geodiff-app/tree/9ae97196ee82e741e17126dfc6ad518a88ea2cac>. We have a public Google Drive folder containing a subset of our dataset for download: <https://drive.google.com/drive/folders/1-UGiOUEEge-DA53k9B7KbIOvMlXKfiYZ>.

### A.3 Description

#### A.3.1 How to access

The complete source code for the crawlers we customized for this research along with their setup instructions is available as a public *GitHub* repository here <https://github.com/censoredplanet/geodiff-app> (commit hash `9ae97196ee82e741e17126dfc6ad518a88ea2cac`). The app binaries and their privacy policies are available upon request (write

to us at [geodiff.app@umich.edu](mailto:geodiff.app@umich.edu) and we will provide a download link). We have provisioned it this way for us to keep track of who has access to our data at a time. The metadata for the apps, and Google's error messages we obtained at the time of app downloads (which we used to determine who is responsible for blocking) are available at <https://drive.google.com/drive/folders/1-UGiOUEEge-DA53k9B7KbIOvMlXKfiYZ>.

#### A.3.2 Data sets

Refer to [subsection A.3.1](#) on how to access our dataset.

### A.4 Installation

Our crawlers use the runtime environment provided by the Python interpreter. We use Pipfiles to manage Python dependencies. The installation instructions are provided as a README on our GitHub repository accessible at the URL: <https://github.com/censoredplanet/geodiff-app/tree/9ae97196ee82e741e17126dfc6ad518a88ea2cac>.

### A.5 Evaluation and expected results

Our setup involved ten Linux machines, ten Android phones, and VPN/S access to 26 countries. Once the measurement testbed is setup (as described in the paper), the crawlers can be used to download the apps from each respective country. However, note that reproducing the entire setup and experiment results depends on several factors, including the availability of reliable VPN/S access for downloads and the apps we tested on Google Play. That apart, the results may not be precisely reproducible given how these apps are updated regularly.