BugSwarm

Enhancing an Infrastructure and Dataset to Support the Software Engineering Research Community

National Science Foundation Award
CNS-2016735 & CNS-1629976

PI: Cindy Rubio-González
Co-PI: Premkumar Devanbu
UC DAVIS
UNIVERSITY OF CALIFORNIA

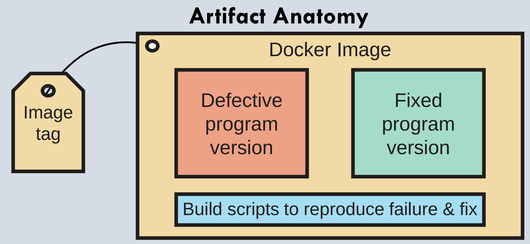
Overview

BugSwarm[1, 3] is a large-scale software defect dataset with its mining infrastructure. The BugSwarm dataset fulfils following desirable characteristics

Scale	Large enough to get statistical significance on tool evaluations
Realism	Reflects actual fixes made by real programmers to repair real bugs
Diversity	Enough variability in scale, maturity, domain, language, severity, etc.
Currency	Continuously keep up latest languages, platforms and libraries
Reproducibility	Failure & fix durably built, run and same results can be observed

BugSwarm Dataset

BugSwarm includes more than 3000 Java and Python software defect artifacts mined from open-source projects.



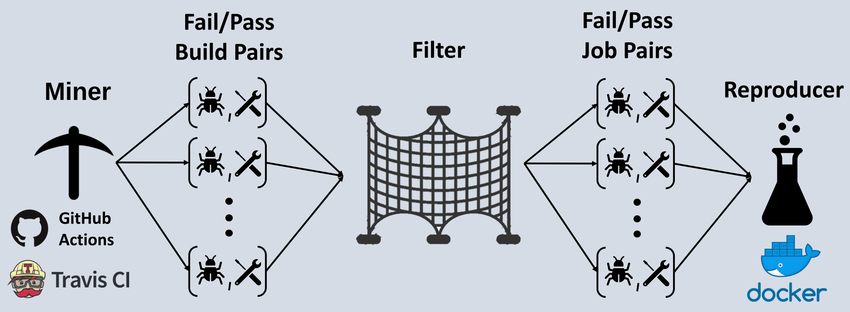
`https://www.bugswarm.org/dataset/`

BugSwarm Infrastructure

Pair Miner Mines consecutive fail-pass build pairs from CI services such as Travis-CI or GitHub Actions
- Scans pairs of builds in chronological order
- Extracts consecutive fail-pass build pairs

Pair Filter Applies filtering rules to discard unsuitable fail-pass pairs

Reproducer Reproduces the fail-pass pairs
- Checks out fail-pass commits to generate build scripts
- Runs build scripts to reproduce failure and fix
- Analyzes build logs to verify reproducibility



`https://github.com/bugswarm/bugswarm`

Artifact Metadata

Project	GitHub slug, primary language, build system, etc.
Reproducibility	Number of attempts/successful attempts to reproduce
Labels	Exceptions observed and patch location
Pull Request	Pull request ID and branch
CI Job	Build & job ID, trigger commit, failed tests (if any)
Image Tag	Unique identifier for an artifact

Supporting Tools

Dataset visualization
`http://www.bugswarm.org/dataset/`

BugSwarm API to query metadata
`https://www.bugswarm.org/docs/toolset/bugswarm-rest-api/`

BugSwarm CLI to download and run an artifact interactively
`https://www.bugswarm.org/docs/toolset/bugswarm-cli/`

Challenges for Growth & Sustainability

Maintaining Long-Term Reproducibility [2]
- Reproducibility of software defect datasets decays after creation
  - We conducted a study on 5 state-of-the-art Java defect datasets
  - Reproducibility of software defect datasets was as low as 30%
  - BugSwarm was not an exception, artifacts broke frequently
- Reproducibility is primarily affected due to missing dependencies. We created 10 patches to fix reproducibility, but they induce high maintenance overhead
- We proposed dependency caching and artifact isolation to ensure long-term reproducibility with minimal maintenance effort
  - BugSwarm's reproducibility has remained higher than 95%

Artifact Labeling
- Labels are critical not only for dataset usability but to guide the growth of the dataset
- Current labeling methodology is based on multiple criteria:
  - Exceptions raised in the failed build
  - Patch location: code files, test files and/or build files
  - Diff size, language, build system, number of tests run/failed
- Exploring other potential methodologies for labeling
  - Failure reproducibility is a big advantage over mining alone

Large-Scale Mining & Reproduction
- Big opportunity for dataset growth
- We mine 215 new reproducible artifacts in a month with single server
  - 1M GitHub Actions from 824 repositories in a month period
  - 25K initial fail-pass pairs ⇨ 12K fail-pass pairs after filtering
  - Attempted reproduction of 3K fail-pass pairs
  - Resulted in 467 reproducible artifacts ⇨ 215 isolated

`https://www.bugswarm.org/statistics/`

References
[1] D.A. Tomassi, N. Dmeiri, Y. Wang, A. Bhowmick, Y.-C. Liu, P.T. Devanbu, B. Vasilescu, C. Rubio-González, "BugSwarm: Mining and Continuously Growing a Dataset of Reproducible Failures and Fixes," in ICSE 2019.
[2] H.-N. Zhu, C. Rubio-González, "On the Reproducibility of Software Defect Datasets," in ICSE 2023.
[3] H.-N. Zhu, K.Z. Guan, R.M. Furth, C. Rubio-González, "ActionsRemaker: Reproducing GitHub Actions," in ICSE-Companion 2023.