ACSAC Artifact Documentation

Artifact Configuration

Our artifacts consist of the followings:

1) README.md: specifying the code execution environment and method
2) requirements.txt: specifying the dependent library
3) shell scripts and python codes for BinShot execution
4) some samples of our full binary dataset
5) pre-trained and fine-tuned models (in Section 6-2).

Additionally, preprocessed data were uploaded to the "https://bit.ly/3xov03n" due to capacity problems. Note that the original binaries were not uploaded due to the SPEC benchmark's copyright issue.

References to Paper Sections

All of the experiments in Sections 6.1-6.4 can be conducted with our artifacts. The dataset preprocessing in Section 6.1 runs with the codes: 'gen_ida.sh', 'gen_norm.sh', and 'corpusgen.py'. With preprocessed data, the results of Sections 6.2-6.4 can be obtained by performing pre-training (bert_mlm.py), fine-tuning (binshot.py), and inference (binshot.py). The final output for the whole dataset is presented as accuracy, precision, recall, and F1 score, which are the metrics used in our paper. In order to obtain the results according to compiler configurations, you should run another code (result.py). Also, in case of the practicality evaluation, another code (result_cve.py) needs to be run to obtain metrics that correspond to our realistic scenario in Section 6.4. The codes for visualization in Section 6.5 and runtime efficiency experiments in Section 6.6 are not included, but they are easy to conduct.

Environments

In addition to the python packages mentioned in requirements.txt that can be installed using pip, python3, IDA Pro, and PyTorch are required for code execution. Note that our codes are tested on python3.8, IDA Pro 7.6, and PyTorch 1.11. Also, we recommend using GPU to run our codes at a practical speed. Pre-training, fine-tuning, and inference phases each take about an hour and a half, 10 hours, and 8 minutes per epoch on the server equipped with two

Intel Xeon Silver 4210R CPUs (with 20 cores in total) running at 2.40 GHz, 576 GB RAM, and NVIDIA TITAN RTX GPU card. Each phase(pre-training, fine-tuning, and inference) uses 15,863/4,536, 8,193/23,247, 5,169/5,103 MB of RAM/GPU memory, respectively.