Vulnerability Detection based on Differential Analysis Dataset

Shuo Liu, sl4921@columbia.edu

Mar. 25, 2022

1 Project Idea

"Vulnerabilities are flaws in a computer system that weakens the overall security of the system". Some of them can cause performance degradation, functionality failure, or security issues. In my final project, I plan to use deep learning techniques such as fully connected networks, convolutional neural networks, etc. to detect the bugs in the program. The experiment will be based on the advanced Differential Analysis Dataset (D2A), which includes the trace associated with the identified vulnerabilities. Experiment results should include the detection accuracy and an interpretive explanation of how the model works in reality.

2 Course Relevance

Vulnerability discovery has always been a hot topic in software engineering. As one of the most important topics covered in this course, the project idea is also aligned with many presentations in this course, so when conducting the experiment of this project, I can refer to some classmates' talk.

3 Motivation

The most important reason for me to select this topic as my final project is that I made a presentation, which reviews the recent works that use deep

learning techniques to discover software vulnerabilities. During that presentation, I listed some possible future works, including creating a benchmark dataset for evaluating and comparing the effectiveness of different approaches on the same scale. In the discussion, the professor shared with us a standard IBM dataset in this area D2A. Hence, I plan to migrate some previous valuable works based on this dataset and to see how it performs. By analyzing the results, probably I will try to enhance these works. If things go well, I will also try to post my work to the official submission board or to compare my results with theirs.