COMSE6156 Project Progress Report

Tianzhi Huang

UNI: th2888

Overview

The goal of this project is to evaluate security threats and vulnerabilities of the software in three aspects: 1) potential threats and vulnerabilities in all external dependencies and Third-Party Libraries included in the project, 2) code errors in project source files, and 3) code similarity between source files in the project. The first aspect is detected and analyzed using dynamic analyzing tool (for example Infer) on the meta data files of the project; The second aspect is evaluated by statically analyzing all the source codes, so as the third aspect. This project is the combination and extension of three subsystems that are responsible for analyzing the project being tested. My project will extract needed information at first, such as meta data files (.xml), source files (.java), compress if required (to .jar file), and feed them to those three subsystems. Then the project will read the results produced by those three subsystems, build a evaluation matrix inside its evaluation module, and eventually output a concrete report about the security threats and vulnerabilities issue related to the project, including various information such as number of vulnerable external dependencies detected, the level of threat, number of code errors detected, number of source files that contain any code error, source files that have too high similarities, and so on.

The evaluation of this project will be done by testing 2-3 projects probably with different scale. The programming language of those testing project could be all Java, or vary, which depends on the availability of the open-source projects I could find, whether they are appropriate, and the compatibility of them against three subsystem (whether they support projects written in this type of programming language). Eventually, I will try to answer all the research questions below to assess different aspects of evaluation criteria of the project.

Value to User Community

The project attempts the identify and detect security threats and vulnerabilities of a software in different aspects within a single run of the project, and produces a detailed report containing detection results and evaluation.

By the papers that I have studied for the midterm paper, I am aware that software vulnerabilities could exists under various situations, and a single aspect of detection is not sufficient to identify all vulnerabilities thoroughly. In this project, the first aspect, described above in the overview section, is analyzed by Infer (a detection tool that does online and dynamically analysis of project included dependencies); the second aspect is done by a static code analyzing algorithm, and the third one is done by code scanning and comparing. In terms of tools, this project uses both static and dynamic way of vulnerabilities detection tools, trying to include different type of detection

tools when scanning and analyzing the target project. As for the aspects of evaluating, the first aspect is about security threats and vulnerabilities in all the external dependencies and Third-Party Libraries, which means vulnerabilities that are not necessarily produced by your code, but rather introduced by the libraries or dependencies that you choose to include in your project. The second one, which is security threats and vulnerabilities that are actually produced by your written code, includes code errors, misuse of certain APIs, and possibly insufficient error handling. The third one is the similarity of code portion within your source code, which introduce the possibility of having redundant code that does the identical functionality (possibly one is being used and the other is supposed to be deprecated but not, causing the vulnerabilities of inconsistent code usage).

To summarize, this project tries to cover three aspects of software security threats and vulnerabilities detection using both static and dynamically tools in one single run. It aims to give the user a more detailed and broader picture of software security threats related issues. Although the project can not cover more aspects and its analysis may be just a starting point of conducting a more thorough experiment, it gives the user a primary image of how their projects do in term of vulnerabilities protection, in a way that is fast and easy to do.

Research Questions

RQ1: Is the project efficient to run against different scale (size) of testing projects?
RQ2: Is the project effective in terms of detecting as many possible threats as possible?
RQ3: Is the project correctly output those security threads? (number of false positive or false negative cases)
RQ4: Is the final report readable and meaningful to the project developers (does it contain meaningful and useful information for the project developers to locate the problems)?

Demo Plan

The first 1 minute I will briefly introduce the idea of the project with any necessary background information, the structure of the project, and any tool it uses. Then the in the next 5 minutes I will try to run the project on one (or two if time permits) test project(s), and show the output result it produced. And if there is time remaining, I will briefly go over the evaluation result of my projects, as well as the answers to the research questions. If the running of my project can not be completed within 5 minutes, I will still run the project and show some intermediate results (scripts, messages printed in the cell...), and discuss the performance of the project using the results I got when doing the evaluation of the project on my own.

Code Deliver

The code of the project will be delivered via Github. I will make a public repo and put all the code in there.

Open-source Software

1. Infer: https://github.com/facebook/infer
2. dep-scan: https://github.com/AppThreat/dep-scan
3. Copydetect: https://github.com/blingenf/copydetect
4. SIM: https://github.com/andre-wojtowicz/sim
5. DeepSIM: https://github.com/parasol-aser/deepsim