Homework 1

CSE 60770

Zanxiang Yin

Graduate students are expected to have evidence of *research synthesis* during the semester. To fulfill this requirement, please read the following paper and answer the questions below:  
[Beyond Metadata: Code-centric and Usage-based Analysis of Known Vulnerabilities in Open-source Software by Ponta et al (2018)](https://findtext.library.nd.edu/ndu_local?sid=google&auinit=SE&aulast=Ponta&atitle=Beyond+metadata:+Code-centric+and+usage-based+analysis+of+known+vulnerabilities+in+open-source+software&id=doi:10.1109/ICSME.2018.00054)

1. **What is the motivation and problem being tackled by this paper and how they solve this problem?**

**The motivation for this article is that the development of modern software is increasingly dependent on third-party open source libraries and frameworks, which often introduce vulnerabilities into projects. The authors of this paper realize that current methods of identifying vulnerabilities may not be able to capture all of them and mitigate their impact. Therefore, the author hopes to have a more accurate method to realize the identification of vulnerabilities in open source libraries and frameworks.**

**The problem that this paper hopes to solve is that the existing vulnerability identification methods have certain limitations. The author hopes that by updating the traditional identification methods, the vulnerability can be identified more effectively. The existence of vulnerabilities is not only related to the code itself, but also to the environment in which the code itself operates and how it is used.**

**This paper analyzes vulnerabilities in open source software by focusing on code and based on different usage cases. The authors create a comprehensive data set from the data obtained through various collection methods. Information is extracted from the source of the data, including the characteristics of the vulnerability and the impact of the vulnerability on the software, through self-soft language processing technology.**

1. **Given the ideas explored in the paper above, how could you change your vulnerable dependency finder to incorporate some ideas described in it?**

**Collect and analyze data:**

**Tracking mechanisms can be used during data collection to track the use of open-source software in a software project. With the data collected on each dependency interaction, a diagram is constructed to show which parts of the code base are governed by a particular dependency.**

**NLP technology is used to identify keywords in the text, including CVE identifiers, affected components, and vulnerability descriptions. There are many sources of data, from submitted information, problem descriptions, mailing lists, text documents, etc. NLP can efficiently extract key information from them and judge the severity of vulnerabilities by the occurrence frequency of key characters.**

**For vulnerability detection of DNS systems, code fingerprinting is usually used, using code fingerprinting technology to identify and track known vulnerabilities in the software. If we need to scan the codebase for known vulnerability patterns and issue warnings, we typically use static code analysis.**

**This article also mentions that the environment in which the code is used can also contribute to vulnerabilities. In this way, it is also possible to assess whether the path of the vulnerability code is being used. Impact analysis is also commonly used for context assessment of code, mainly to analyze the potential impact of vulnerabilities. For the impact of vulnerabilities, we can conduct risk scores, ranking the critical degree of vulnerabilities, so as to improve the efficiency of making up for vulnerabilities. CVSS can be used for reference here.**