CS292F Project Proposal

Md Shafiuzzaman and Vinothini Gunasekaran

April, 2022

1 Title

Software Vulnerability Detection by Feeding Multiple Code Semantics to Graph Neural Network

2 Background and Research Questions

The exploitation of vulnerabilities threatens the security of the software systems. Despite the continued efforts of the software engineering community to improve software quality and security, an increasing trend in the number of vulnerabilities is reported through publicly disclosed information-security flaws and exposures (CVE). Actually, the effective identification and mitigation of software vulnerabilities never been an easy task and requires experience and specialized skills that go beyond the expertise of the typical developer. To improve usability and avoid the intense labor of human experts, recent works investigate the potential of deep neural networks on a more automated way of vulnerability detection. All of the existing works either treat the source code as a flat sequence, which is similar to natural languages, or represent it with a single aspect of representation such as Abstract Syntax Tree (AST), data flow, or control flow. However, source code is actually more structural and logical than natural languages and requires heterogeneous aspects of representation and multiple dimensions of semantics.
In this project, we argue to comprehend multiple aspects and dimensions of program semantics by combining several representations of programs such as AST, Control Flow Graph (CGF), and Data Flow Graph (DFG). Then use Graph Neural Network (GNN) to use all those semantics to identify vulnerable source codes.

Based on our arguments our research questions for this project are as follows:

- Is there an effective way to treat AST, CGF and DFG as subgraphs and combine them into one joint graph?
- Can the combined code semantics be orthogonal to the improvements of program vulnerability detection?

3 Background Papers

- Cheng, Xiao, Haoyu Wang, Jiayi Hua, Guoai Xu, and Yulei Sui. ”DeepWukong: Statically detecting software vulnerabilities using deep graph neural network.” ACM Transactions on Software Engineering and Methodology (TOSEM) 30, no. 3 (2021): 1-33.

- Li, Zhen, Deqing Zou, Shouhuai Xu, Zhaoxuan Chen, Yawei Zhu, and Hai Jin. ”Vuldeelocator: a deep learning-based fine-grained vulnerability detector.” IEEE Transactions on Dependable and Secure Computing (2021).

- Chakraborty, Saikat, Rahul Krishna, Yangruibo Ding, and Baishakhi Ray. ”Deep learning based vulnerability detection: Are we there yet.” IEEE Transactions on Software Engineering (2021).

- Xiao, Yang, Bihuan Chen, Chendong Yu, Zhengzi Xu, Zimu Yuan, Feng Li, Binghong Liu et al. ”MVP: Detecting Vulnerabilities using Patch-Enhanced Vulnerability Signatures.” In 29th USENIX Security Symposium (USENIX Security 20), pp. 1165-1182. 2020.

- Zhou, Yaqin, Shangqing Liu, Jingkai Siow, Xiaoning Du, and Yang Liu. ”Devign: Effective vulnerability identification by learning comprehensive program semantics via graph neural networks.” Advances in neural information processing systems 32 (2019).

4 Dataset

- CodeXGLUE

- CVEfixes

- FFmpeg