Contribution ID: **12** Type: **Oral Presentation**

(REMOTE) A Novel Fine-Grained Source Code Vulnerability Detection Model via Joint Token and Statement Representation Learning

Wednesday, 19 March 2025 15:00 (20 minutes)

Abstract— As software systems become increasingly complex, defects in source code pose significant security risks, such as user data leakage and malicious intrusions, making their detection crucial. Current approaches based on Graph Neural Networks (GNNs) can partially reveal defect information; however, they suffer from heavy graph construction costs and underutilization of heterogeneous edge information. In contrast, sequence model-based methods primarily capture token-level representations, failing to effectively learn statement-level features and their interrelations, which results in poor detection performance. Moreover, most existing methods only support coarse-grained vulnerability detection, lacking the capability for precise fine-grained analysis. To address these issues, this paper develops a novel sequence model that simultaneously learns both token and statement representations, thereby enhancing the detection of vulnerabilities at the statement level. The proposed approach is expected to achieve significant improvements in both accuracy and F1 score, offering a more refined and efficient solution for source code defect detection.

Thank you for your time and consideration.

Keywords—Code Vulnerabilities, Graph Neural Networks, Hierarchical Attention Mechanism

Primary author: SUN, ShengXuan (Insititute of High Energy Physics,Chinese Academy of Science)

Presenter: SUN, ShengXuan (Insititute of High Energy Physics,Chinese Academy of Science)

Session Classification: Network, Security, Infrastructure & Operations I

Track Classification: Track 7: Network, Security, Infrastructure & Operations