INFORMATICS INSTITUTE OF TECHNOLOGY

UNIVERSITY OF WESTMINSTER

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

SHERLOCK: A Deep Learning Approach To Detect Software Vulnerabilities

A Thesis by

Mr. Saadh Jawwadh

Supervised by

Mr. Guhanathan Poravi

Submitted in partial fulfilment of the requirements for the BSc in Computer Science degree at the University of Westminster.

May 2023

Abstract

The increasing reliance on software in various applications has made the problem of software vulnerability detection more critical. Software vulnerabilities can lead to security breaches, data theft, and other negative outcomes. Traditional software vulnerability detection techniques, such as static and dynamic analysis, have been shown to be ineffective at detecting multiple vulnerabilities.

To address this issue, this study employed a deep learning approach, specifically Convolutional Neural Networks (CNN), to solve the software vulnerability detection problem. A 5-split cross-validation approach was used to train and evaluate the CNN model, which takes tokenized source code as input.

The findings indicated that Sherlock successfully detected multiple vulnerabilities at the function level, and its performance was particularly strong for CWE-199, CWE-120, and CWE-Other, with an overall high accuracy rate and significant true positive and true negative values. However, the performance was less reliable for some vulnerabilities due to the lack of a standardized dataset which will be a future research direction. The results suggest that compared to current techniques, the proposed deep learning approach has the potential to substantially enhance the accuracy of software vulnerability detection.

Keywords: Software Vulnerability Detection, AI, Deep Learning, Convolutional Neural Network, Gaussian Noise.

Subject Descriptors:

- Security and privacy → Systems security → Vulnerability management → Vulnerability scanners
- Computing methodologies → Artificial intelligence
- Computing methodologies → Machine learning → Machine learning approaches → Neural networks