Title: VulDeePecker: A Deep Learning-Based System for Automatic Vulnerability Detection

Readme:

Code Gadget Database (CGD) focuses on two types of vulnerabilities in C/C++ programs, buﬀer error vulnerability (CWE-119) and resource management vulnerability (CWE-399). Each *code gadget* is composed of a number of program statements (i.e., lines of code), which are related to each other according to the data ﬂow associated to the arguments of some library/API function calls.

Based on the National Vulnerability Database (NVD) and the NIST Software Assurance Reference Dataset (SARD) project, we collect 520 open source software program files with corresponding diff files and 8,122 test cases for the buﬀer error vulnerabilities, and 320 open source software program files with corresponding diff files and 1,729 test cases for the resource management error vulnerabilities.

In total, the CGD database contains 49,665 code gadgets, including 14,110 code gadgets that are vulnerable and 35,555 code gadgets that are not vulnerable. Among the 14,110 code gadgets that vulnerable, 8,398 corresponds to buﬀer error vulnerabilities and the rest 5,712 corresponds to resource management error vulnerabilities.