Title: Improving Software Security with Automated Vulnerability Analysis and Inference with Multi-Case Based Reasoning

Date: 5/11/2018
PI(s): Shaoen Wu
University: Ball State University

Long Term Goal(s)

The long-term goal of this project is to design and develop a GPU cloud-based intelligent automated software vulnerability analysis and inference system by using a machine learning method called Multi-Case Based Reasoning (MCBR) to parse the confirmed software vulnerabilities in databases such as the National Vulnerability Database (NVD). This system consists of two major components. First, by using a deep learning method such as convolutional neural network (CNN), this system will automatically classify the reported and confirmed vulnerabilities into various categories (cases). Then, the MCBR method will be used to compute the similarities between a new input and the established cases, as to predict how likely an input is to multiple established cases.

Background for Long Term Goals

CBR and MCBR has been widely used in software engineering to predict costs and quality, but rarely on software vulnerability [1-3]. A number of software vulnerability analysis solutions based machine learning have been proposed [4-8], but none of them uses CBR/MCBR approach. This project will contribute to the community in two aspects: (1) it uses a deep learning based data mining approach to learn features of each vulnerability report and classify them into cases based on their features, which will build a database of vulnerability cases, and (2) it adopts Multi-Case Based Reasoning (MCBR) in software vulnerability analysis and detection, which can predict a potential software vulnerability and its similarity to the confirmed cases within the database. We will develop a MCBR model for software vulnerability called MCBR-SV, which expects to improve both the prediction accuracy and the probability to identify new vulnerability types by evaluating on multiple relevant existing vulnerability types, whereas the conventional CBR can only decide on one type. Currently we have gained rich experience of using CBR in software engineering by have developed a software bug automation tool in a master thesis project.

Intermediate Term Objectives

This project has the following intermediate-term objectives: (1) development of software vulnerability case database by using deep learning based data mining algorithms, and (2) development of a MCBR model for automated software vulnerability prediction.
Schedule of Major Steps

The follow major milestones are planned: (1) we will use five months to establish the software vulnerability case database including features/attributes formulation and classification, (2) we will use another six months to develop the MCBR based software vulnerability prediction system, and (3) we will one month to evaluate the system performance.

 Dependencies

In the project, a few open source tools will be used an integrated, including Apache Hadoop, Ant, Nutch, Solr and myCBR.

 Major Risks

na

 Budget

The total budget is $43,799 including the administrative indirect cost.

This budget includes the stipends: a summer month of the PI and a full year support of a graduate research assistant (GRA) including the stipend and the tuition.

The budget also supports travels of $4,500 to showcases for status and final reports as well as outcome publication at conferences.

 Staffing

PI: Shaoen Wu; A undergraduate research assistant.

 Category of Current Stage

The project is to start on May 14th, the beginning of 2018 summer session at Ball State University.

 Contacts with Affiliates

None.

 Publications and Other Research Products (actual or potential)

The expected deliverables include: (1) a case database of software vulnerabilities, (2) a set of software tools, (3) an automated system, and (4) at least one publication

 References


