Researchers develop automated approach to extract security policies from software

January 31 2022

Credit: CC0 Public Domain

A team of UTSA researchers is exploring how a new automated approach could prevent software security vulnerabilities.

The team—made up of Ram Krishnan, associate professor in the UTSA Department of Electrical and Computer Engineering; Yufei Huang, professor in Electrical and Computer Engineering; Jianwei Niu, professor in Computer Science; Ravi Sandhu, professor and Lutcher Brown Distinguished Chair in Cyber Security; and John Heaps, a postdoctoral researcher in the UTSA Institute for Cyber Security—sought to develop a deep learning model that could teach software how to extract security policies automatically.

Unlike traditional software models, the agile software development process is meant to produce software at a faster pace, eliminating the need to spend time on comprehensive documents and changing software requirements. User stories, the specifications that define the software's requirements, are the only required documentation. However, the practices innate to this process, such as constant changes in code, limit the ability to conduct security assurance reviews.

"The basic idea of addressing this disconnect between security policies and agile software development came from happenstance conversation with software leaders in the industry," Krishnan said. "We were able to assemble a team of faculty and students with expertise in cybersecurity, software engineering and machine learning to start investigating this problem and develop a practical solution."

The researchers looked at different machine learning approaches before settling on a deep learning approach, which can handle several formats of user stories. The model consists of three pieces to perform the prediction: Access control classifications, named entity recognition and access type classification. Access control classification helps the software decide if user stories contain access control information. Named entity identifies the actors and data objects in the story. The access type classification determines the relationship between the two.

The team took a data set of 21 web applications, each consisting of 50-130 user stories, or 1,600 total, to test their approach.

"With a dataset of 1,600 user stories, we developed a learning model based on transformers, a powerful machine learning technique," Krishnan said. "We were able to extract security policies with good accuracy and visualize the results to help stakeholders better refine user stories and maintain an overview of the system's access control."

This innovative new approach will serve as a valuable tool in the modern agile software development life cycle, Krishnan said.

"Since agile software development focuses on incremental changes to code, a manual process of extracting security policies would be error-prone and burdensome," he added. "This is yet another area where machine learning/artificial intelligence shows to be a powerful approach."

Krishnan said the team still has several directions they would like to take the project.

"We recognize that there is little additional information about access control that can be extracted or determined directly from user stories in a fully automated approach," Krishnan said. "That means it is difficult, or impossible, to determine a software's exact access control from user stories without human involvement. We plan to extend our approach to make it interactive with stakeholders so that they can help refine the access control information."

Provided by University of Texas at San Antonio