Seminar in Computer & information Science

Date: Wednesday, Nov 11, 2015  
Time: 1:00 pm  
Place: JMH 312  
Title: Combining Static Analysis and Machine Learning for Industrial-quality Information-flow-security Enforcement  
Speaker: Dr. Marco Pistoia, IBM Thomas J. Watson Research Center in New York

Abstract: Security auditing of industry-scale software systems mandates automation. Static taint analysis enables deep and exhaustive tracking of suspicious data flows for detection of potential leakage and integrity violations, such as cross-site scripting (XSS), SQL injection (SQLi) and log forging. Research in this area has taken two directions: program slicing and type systems. Both of these approaches suffer from a high rate of false findings, which limits the usability of analysis tools based on these techniques. Attempts to reduce the number of false findings have resulted in analyses that are either (i) unsound, suffering from the dual problem of false negatives, or (ii) too expensive due to their high precision, thereby failing to scale to real-world applications. In this talk, we investigate a novel approach that combines static analysis and machine learning for improving the scalability of static taint analysis and reducing the number of false positives. From a static analysis perspective, the key observation informing our approach is that taint analysis is a demand-driven problem. This enables lazy computation of vulnerable information flows, instead of eagerly computing a complete data-flow solution, which is the reason for the traditional dichotomy between scalability and precision. With the analysis being scalable to large codebases, the user is still left to review hundreds, if not thousands, of potential warnings, and classify them as either actionable or spurious. This is both burdensome and error prone, leaving developers disenchanted by static security checkers. We address this challenge by introducing a general technique to refine the output of static security checkers. The key idea is to apply statistical learning to the warnings output by the analysis based on user feedback on a small set of warnings. This leads to an interactive solution, whereby the user classifies a small fragment of the issues reported by the analysis, and the learning algorithm then classifies the remaining warnings automatically. An important aspect of our solution is that it is user centric. The user can express different classification policies, ranging from strong bias toward elimination of false warnings to strong bias toward preservation of true warnings, which our filtering system then executes.

Biography: Marco Pistoia, Ph.D. has worked for IBM Corporation since January 1996 and is currently a Senior Manager and Principal Research Staff Member at the IBM Thomas J. Watson Research Center in New York, where he manages the Mobile Enterprise Software research group. In January 2010, he was one of 38 IBM employees worldwide to be bestowed the title of IBM Master Inventor. He is the inventor of 97 patents issued by the United States Patent and Trademark Office, and 164 patent applications. Dr. Pistoia has written ten books and published numerous papers and journal articles on various aspects of Program Analysis and Language-Based Security. He has published and presented at numerous conferences worldwide, including OOPSLA, ECOOP, PLDI, ICSE, ACSAC, ISSSTA,
Marco Pistoia, Ph.D. Senior Manager, Principal Research Staff Member, Master Inventor IBM T.J. Watson Research Center

http://www.research.ibm.com/people/p/pistoia