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, ISSTA,

CCS, PLAS, VMCAI, CISIM, and the IEEE Symposium on Security and Privacy. He has also been invited to lecture at several research institutions worldwide, including Harvard University, New York University, University of Maryland, Rutgers University, Virginia Tech, Stony Brook University, Fordham University, University of Texas at Austin and Stevens Institute of Technology in the United States, Tohoku University and the National Institute of Informatics in Japan, École Normale Supérieure in France, Dagstuhl School of Informatics and Saarland University in Germany, Eidgenössische Technische Hochschule (ETH) Zürich in Switzerland, La Sapienza University and Tor Vergata University in Italy, Tel Aviv University, Israel Institute of Technology (Technion) and Ben Gurion University of the Negev in Israel, University of Porto in Portugal, and Chalmers University of Technology in Sweden. He has been an Adjunct Professor of Computer Science at New York University, Polytechnic School of Engineering since 2000 and he is now an Adjunct Professor of Computer Science at Fordham University, New York. He was the General and Program Co-chair of PLAS 2008, and the Program Chair of the ACM Student Research Competition at PLDI 2009. Furthermore, he has served as Program Committee member on several conferences, including ICSE 2012, ICST 2012, ISSTA 2011, PLAS 2007, 2009, 2010, 2011 and 2012, 2014 and 2015, NDSS 2009, IEEE SSIRI 2009, 2010 and 2011, IEEE SERE 2012, ACSAC 2008 and 2009, and CISIM 2012. Dr. Pistoia received his Ph.D. in Mathematics from the New York University, Polytechnic School of Engineering in May 2005 with a thesis entitled *A Unified Mathematical Model for Stack- and Role-Based Authorization Systems*, and his Master of Science and Bachelor of Science degrees in Mathematics *summa cum laude* from Tor Vergata University, Rome, Italy in July 1995, with a thesis entitled *Theory of Reductive Algebraic Groups and Their Representations*. Dr. Pistoia has been the recipient of several awards, including three ACM SIGSOFT Distinguished Paper Awards (2007, 2011 and 2014), an IBM Research Pat Goldberg Memorial Best Paper Award (3 papers selected our of 130), and a European Community Erasmus Fellowship Award. In September 2007, the Italian Ministry of Education, University and Research, the National Committee of the Italian Presidents of Faculties of Sciences and Technologies, and Confindustria, Italy's leading organization representing all the Italian manufacturing and service companies, presented Pistoia as one of the 70 most successful Italian mathematicians who graduated from an Italian university between the years 1980 and 2000. His biography was published in the book *Matematici al Lavoro*.

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