UCF Stands For Opportunity

THE DEPARTMENT OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE, CS DIVISION

Presents the Fall 2013 EECS Seminar Series

Dr. Lok Yan

Air Force Research Lab

"Verifying Dynamic Taint Analyzers"

Tuesday, November 5, 2013 • 11:00 a.m. • HEC 450

Abstract

Taint analysis (also known as Dynamic Information Flow Tracking) is a program analysis technique that is designed to reveal data and control dependencies in programs. The central idea is simple. We label data of interest as tainted at the beginning of execution and propagate the taint to other data as execution continues until an event of interest occurs. While simple, the concept can be quite powerful. The technique has been used to identify stack smashing attacks ("is the return address being overwritten by user-input?"), malware unpacking behavior ("is the code being executed a function of the sample's binary executable?"), path exploration ("which data variables affect the outcome of this branch condition?"), and others. Unsurprisingly, the technique has been implemented in many different ways, each with their own advantages and disadvantages. Surprisingly though, there hasn't been much work in understanding the correctness of the implementations themselves.

In this talk, we will describe our approach to formally analyze taint analyzers. This talk will be given from an applied perspective and not a basic research/theory perspective. To put it differently, we will describe how we used currently available formal analysis tools to verify taint analyzers. We will first describe the basic motivations and need for a more fundamental understanding of taint analysis. We then briefly describe our formal model and what the fundamental tradeoffs between analysis efficiency and correctness are. Given these basic definitions, we then present our efforts in and the overall process of designing and implementing DECAF, a new taint analysis platform that is (mostly) correct by construction. DECAF is built upon the QEMU system emulator and propagates taint (one label per bit of data) through TCG-IR (Tiny Code Generator - Intermediate Representation). We will also describe how we verified DECAF's taint analysis implementation by comparing its taint analysis results with a reference implementation built on CMU's BAP (Binary Analysis Platform) and SMT (Satisfiability Modulo Theories) solvers.

Biography

Lok Yan received his Ph.D. in Computer and Information Science and Engineering in May 2013 from Syracuse University. He holds a B.S. in Computer Engineering and M.S. in Electrical Engineering both from Polytechnic University (now Polytechnic Institute of NYU/NYU-Poly). Lok is currently employed at the Air Force Research Laboratory, Information Directorate in Rome, NY as a Computer Engineer. He is also an adjunct faculty at NYU-Poly and teaches the online section of Information Security and Privacy, also known as Computer Security. He is also a contributor to the DECAF/DroidScope open source dynamic analysis platform (http://code.google.com/p/decaf-platform).