

# Experimental Security Panoramas for Critical System Protection Workshop, August 14

---

As of May 23, 2012

## Session 2

### Evaluating a ROP Defense Mechanism

**Professor Angelos Keromytis (Columbia)**

We recently developed "in-place code randomization", a practical mitigation technique against ROP attacks that can be applied directly on third-party software. Our method uses various narrow-scope code transformations that can be applied statically, without changing the location of basic blocks, allowing the safe randomization of stripped binaries even with partial disassembly coverage. These transformations effectively eliminate about 10%, and probabilistically break about 80% of the useful instruction sequences found in a large set of PE files. Since no additional code is inserted, in-place code randomization does not incur any measurable runtime overhead, enabling it to be easily used in tandem with existing exploit mitigations such as address space layout randomization. In this talk, I will discuss the challenges in evaluating this probabilistic software detection mechanism. We used both real ROP exploits and two ROP code generation toolkits. I will discuss the various metrics that are applicable, and why some of those were impossible or difficult to obtain.

### The Importance of Realistic Quantitative Studies of Malware Detection

**Dr. Mihai Christodorescu (IBM-T.J. Watson research Center)**

A running theme among existing detection techniques is the similar promises of high detection rates, in spite of the wildly different models of malicious activity used. In addition, the lack of a common testing methodology and the limited datasets used in the experiments make difficult to compare these models in order to determine which ones yield the best detection accuracy, especially when exposed to a diverse set of previously-unseen, real-world applications that operate on realistic inputs. This is particularly problematic as most previous work has used only a small set of programs to measure their technique's false positive rate. Moreover, these programs were run for a short time, often by the authors themselves.

I will describe a study of the diversity of system calls based on a large-scale data collection (compared to previous efforts) on hosts that run applications for regular users on actual inputs. The analysis of the data demonstrates that simple malware detectors, such as those based on system call sequences, face significant challenges in realistic environments. This suggests that commonly held beliefs about simple models are incorrect in how they relate changes in complexity to changes in detection accuracy. To address these limitations, I discuss an alternative detection model that characterizes the general interactions between benign programs and the operating system (OS). Our experiments enabled by the large-scale data set demonstrate that this approach captures well the behavior of benign programs and raises very few (even zero) false positives while being able to detect a significant fraction of today's malware.

### Coactive Emergence as a Sensemaking Strategy for Cyber Defense

**Dr. Jeffrey Bradshaw (Florida Institute for Human and Machine Cognition)**

In this talk, I describe how we are applying the concept of *coactive emergence* as an approach to the design of work methods for distributed sensemaking in cyber defense applications. Distributed sensemaking is a process whereby understanding and anticipation of complex situations is achieved through the collaboration of analysts and software

agents working together in tandem. As rationale for the principles used in our work design, we present a series of background studies. We show how coactive emergence as a strategy for threat understanding relates to Klein's Data/Frame theory of sensemaking. I will explain the similarities and differences of coactive emergence from the basic form of second-order emergence. I will then outline a set of considerations for resilient human-automation teamwork. These desiderata address the teamwork requirements of *observability*, *directability*, *interpredictability*, *adaptation*, and *multiplicity*. I will describe recent results and future plans for empirical studies addressing some of the issues raised in this talk.

## **Session 3**

### **Experimenting with Live Cyberattacks for Testing Deceptions**

**Professor Neil C. Rowe (Naval Postgraduate School)**

We have developed an experimental approach to finding deceptive tactics for computer system defense. This approach tries a variety of tactics against live Internet traffic and sees what responses it gets. These experiments are easiest to do on a honeypot, a computer system designed solely as an attack target. We report on three kinds of experiments with deceptive honeypots: one with packet modifications on packets provided by attackers using Snort Inline, and one on scripted responses to attacks using Honeyd, and one on a fake Web site. We found evidence of responses to our deceptions some in the form of increased session lengths, and sometimes by disappearance of attackers. Some benefit was obtained by varying the deceptions over time. These results are encouraging for developing more comprehensive automated deception strategies for defending computer systems, and provide a new experimentation methodology for systematically developing deception plans.

### **Empirical Analysis of System-level Vulnerability Metrics through Actual Attacks**

**Professor Mathias Ekstedt (KTH)**

The Common Vulnerability Scoring System (CVSS) is a widely used and well established standard for classifying the severity of security vulnerabilities. For instance, all vulnerabilities in the US National Vulnerability Database are scored according to this system. As systems typically have multiple vulnerabilities it is often desirable to aggregate the score of individual vulnerabilities to a system level. Several such metrics have been proposed, but their quality have not been studied. This presentation describes a statistical analysis of how a number of security estimation metrics using CVSS data relate to the time-to-compromise of 41 successful attacks. The empirical data originates from an international cyber defense exercise involving over 100 participants. Utilized data was collected through studying network traffic logs, attacker logs, observer logs and network vulnerabilities. Results suggest that security modeling through CVSS data alone doesn't accurately portray the security of a system. However, it also implies that the amount of CVSS information which is used by the metric is of relevance to its accuracy: a metric employing more CVSS data also explains time-to-compromise better.

### **Lessons learned and an experimental framework for access control biometric usability**

**Dr. Alex Kilpatrick (Tactical Information Systems)**

This talk will present lessons learned from the deployment of operational biometric systems for access control and security in Iraq and Afghanistan, and the special challenges faced with a non-English speaking population and high-threat environments. The talk will also present the results of a comprehensive study of biometric usability in DoD and commercial environments, as well as an experimental framework for future biometric usability experiments.

## Session 4

### Big Data for Security: Challenges, Opportunities, and Experiments

**Dr. Pratyusa Manadhata (HP labs)**

This is the age of big data. Big data for security, i.e., the analysis of very large data sets to identify actionable security relevant information, however, is a relatively unexplored area. Our research group has undertaken an initiative on big data for security in enterprise settings, i.e., our goal is to design algorithms and systems to analyze large data sets routinely collected by enterprises for compliance and other reasons. In this talk, we will highlight some of the challenges and opportunities in big data for security, and present a few experimental results.

### Testing the Edge: Cyber Security Testing in the Smart Grid

**Professor David Nicol (UIUC)**

This talk considers issues related to performing security testing of Smart Grid components and subsystems within the context of an evaluation testbed. We discuss methodologies for assessing a component, and for assessing ensembles of components, and metrics that help guide understanding of the testing outcome. We identify some hardware and software tools we have found useful for evaluation studies we've conducted. Our points are illustrated by stepping through a case study, of meters and their communication within an Advanced Metering Infrastructure.

### Trustworthy Transportation Networked Control Systems

**Professor Saurabh Amin (MIT)**

Networked control systems (NCS) are increasingly being deployed to facilitate monitoring and control of large-scale transportation systems, including vehicular traffic and public transit networks. In recent years, the deployment of advanced sensing technologies has caused a considerable increase in both heterogeneity and volume of real-time measurement data. Transportation agencies are now employing the public communication network (Internet) in addition to private (back-haul) networks for most of their routine tasks, such as monitoring, data processing, and distributed control. The significant drawback of this information technology (IT) modernization is lowered security of transportation NCS caused by the exposure to IT insecurities. The security threats primarily come from four channels: **(i)** off-the-shelf IT devices; **(ii)** open communication networks; **(iii)** multi-party data management; **(iv)** large number of field devices (sensors, displays, and actuators). This work focuses on experiment design for improving the trustworthiness of transportation NCS. Our goal is to improve NCS operational resilience against security failures (attacks) and reliability failures (faults).