Improve Software Security and Performance through Data Analytics

Linhai Song

Linhai Song is a staff research scientist at FireEye Labs. Linhai received his Ph.D. from the Department of Computer Sciences at the University of Wisconsin-Madison in 2015. Linhai's research interests span software engineering, security, programming languages, and operating systems, with a focus on applying data analytics to solve both software security and software performance problems. Linhai’s work won Micro’2014 best paper runner-up, and ACM-SIGPLAN Research Highlight Award in 2011.

Abstract:

The big data era brings us many opportunities in large-scale analysis and learning of real-world data. Song’s work focuses on two such opportunities: leveraging data analytics to improve both software security and performance.

In the first part of his talk, he will present his recent work in using big data to study malware and anti-virus engines. Specifically, he will discuss the study of VirusTotal, the largest malware repository with hundreds of millions of malware samples and analysis results from state-of-the-art anti-virus engines. In this study, they focused on a new angle that has not been explored before in the security community, the relationship of anti-virus vendors. They proposed a graph-based approach to quantify the influence across anti-virus vendors. They found that there do exist significant influence across vendors and this influence can be predicted with statistical models we built.

In the second part of his talk, he will introduce a new way to use statistical debugging for diagnosing performance issues in software execution. Statistical debugging is a technique that uses statistical models to analyze a vast amount of data collected from millions of program executions in production. Statistical debugging was originally proposed for functional bugs, where it is used to identify program behaviors that result in program failures. They explored how to extend statistical debugging to diagnose performance failures. To further improve results from statistical debugging, they also designed program analysis tools to provide more detailed debugging information.