Accomplishments in Research

Dr. Kyu Hyung Lee
May 2019

Dr. Kyu Hyung Lee's research interest lies in combining analyses at the program level and the system level to develop synergetic solutions for problems in cybersecurity, software engineering, and computer systems. His main accomplishments are in the area of cyber forensics using program analysis and instrumentation techniques. Dr. Lee has been recently focusing on developing a framework for advanced persistent threats (APT) detection and investigation in heterogeneous environments including mobile devices, Internet-of-things (IoT) devices and enterprise servers.

Publications

Dr. Lee has published 22 scientific articles in Peer-Reviewed Scientific Journals and Academic Conferences. Thirteen of them have been published since the completion of his doctoral studies. Many of his recent publications appeared in the most prestigious computer security, software engineering, and operating system venues, such as the Usenix Security, the Network and Distributed System Security Symposium (NDSS), Usenix Annual Technical Conference (ATC), the ACM Conference on Computer and Communications Security (CCS), the ACM SIGPLAN Programming Language Design and Implementation (PLDI), and the ACM Symposium on Operating Systems Principles (SOSP).
Several of Dr. Lee's publications are highly cited. According to Google Scholar, his publications have cited more than 570 times including 113 citations received in 2018.

Funded Grants

Dr. Lee has obtained extramural funding from the United States Air Force and Defense Advanced Research Agency (DARPA). The total amount for the award is $5,352,881 and Dr. Lee's portion is $360,000 for 4 years beginning at August 2015. He also has received $389,902 from US Department of Army as co-PI with Prof. Prashant Doshi (PI). Additionally, Dr. Lee was a recipient of 2015 UGA Faculty Research Grants Program (FRG).
Dr. Lee's recent proposal to NSF has recommended for an award and currently being processed by NSF. This is a 4-year collaborative project with junior faculty members in Georgia Tech and the University of Virginia. The total amount for the award will be $1,200,000 and Dr. Lee's portion is $361,601.

Presentation of Research Works

Dr. Lee has been invited to present his research and share his expertise to eleven international conferences and institutes including National Security Research Institute in Korea, Ulsan National Institute of Science and Technology (UNIST), and NEC Laboratories America.

Research Expertise

Cyber Forensics. Cyberattacks against personal machines and enterprises are on the rise. Recent attacks have become increasingly stealthy and sophisticated involving social engineering or spear-phishing of a specific individual. They often apply deception techniques to hide the presence of malware in the system and erase evidence of the attacks afterwards. Highly accurate cyber forensics becomes more and more critical to investigating todays cybercrimes. Cyber forensics aims to understand the provenance of suspicious events, disclosing the root cause and ramifications of cyberattacks by collecting and analyzing evidence from cybercrime scenes such as hard disks, memory images, and log files on the suspects or victims' machines. Furthermore, mobile and smart devices are becoming increasingly popular but at the same time, also constantly attract cyber criminals. For example, a recent stagefright attack exploits a vulnerability in Android core component, which potentially infects 950 million Android devices.

Dr. Lee has been studying effective and efficient systems for cyber forensics that are able to reconstruct sophisticated attacks in servers, desktops and mobile devices. Particularly, Dr. Lee and his colleagues have developed effective and efficient audit logging techniques for Linux server, called KCal, that is in-kernel cache-based online log-reduction system. KCal is designed to reduce the runtime overhead caused by transferring, processing, and writing logs, as well as the space overhead caused by storing them on disk. This work has published in the Usenix Annual Conference (ATC) 2018, the most prestigious computer systems conference.
He also has developed two novel causality inference techniques for accurate cyberattack investigation. MCI is a model-based causality inference technique to determine dependencies between system events and allow investigators to determine the origin of a cyberattack. The evaluation on a set of real-world programs shows that MCI is highly effective, and it can accurately infer causalities without any false positives. This work has published in the 2018 Network and Distributed System Security Symposium (NDSS'18), the top-tier conference in cybersecurity field.
Dr. Lee also proposed MPI, a multiple perspective attack investigation technique. These perspectives reflect high-level semantic partitioning of the target program, such as individual tabs or individual web pages for Web-browser applications. MPI automatically analyzes and instruments the target program to proactively report the execution context of the program. MPI can provide accurate, multiple perspective and semantics rich information about the execution. MPI has appeared in the USENIX Security Symposium 2017 and received the Distinguished Paper Award.

To enable forensics analysis on mobile devices, Dr. Lee has developed DroidForensics, a multi-layer forensic logging technique for Android. The goal of this approach is to provide

the user with detailed information about attack behaviors that can enable accurate post-mortem investigation of Android attacks. DroidForensics consists of three logging modules. API logger captures Android API calls that contain high-level semantics of application. Binder logger records interactions between applications to identify causal relations between processes, and system call logger efficiently monitors low-level system events. It also provides the user interface that the user can easily compose SQL-like queries to inspect an attack. DroidForensics has low runtime overhead (2.9% on average) and low space overhead (only 105~169 MByte data generated in 24 hours) on real Android devices. It is effective in the reconstruction of real-world Android attacks we have studied. DroidForensics has appeared in the ACM Asia Conference on Computer and Communications Security (ASIACCS) 2017.

Dr. Lee has authored several other publications in this area as well. In particular, the recent joint works with Dr. Perdisci have published in NDSS 2017 and NDSS 2018. In the work presented in NDSS 2017, we have developed ChromePic, a web browser equipped with a novel forensic engine that aims to greatly enhance the browsers logging capabilities. We have demonstrated that ChromePic can successfully capture and aid the reconstruction of attacks on users. We also have developed JSGraph, a forensic engine that is able to efficiently record fine-grained details pertaining to the execution of JavaScript (JS) programs within the browser, with focus on JS-driven DOM modifications. This work has presented in NDSS 2018. Dr. Lee also has developed several effective cyber forensics techniques for Linux and Windows systems, and published the results in NDSS'13, CCS'13, and ACSAC'15 conferences.

Software Attack Protection. Attacks which exploit software vulnerabilities are among the most prevalent cyber-security threats to date. This is due, in part, to many complex combinations of potential attack vectors: Buffer overflow attacks, Return-to-libc attacks, ROP, Jump-oriented programming (JOP), and Heap spraying. Unfortunately, variety of exploit attack vectors has led to a constant “cat and mouse” game of building defenses as each new attack is released. Dr. Lee and his colleagues have developed A2C, a system that provides protection against payload injection attacks. It is based on the observation that payloads are highly fragile and thus any mutation would likely break their functionalities. A2C mutates inputs from untrusted sources. Malicious payloads that reside in these inputs are hence mutated and broken. This work was published in NDSS 2017.

Recently, Dr. Lee and his colleagues in Georgia Tech have developed Fuzzification, to help program developers protect the released, binary-only software from adversarial fuzzing techniques. Fuzzing is popular and effective techniques to malicious attackers to find zero-day vulnerabilities. In this work, we propose a new direction of binary protection technique that can effectively hinder attackers from discovering zero-day vulnerabilities. Particularly, we proposed three techniques: 1) SpeedBump, which amplifies the slowdown in normal executions by hundreds of times to the fuzzed execution, 2) BranchTrap, interfering with feedback logic by hiding paths and polluting coverage maps, and 3) Anti-Hybrid, hindering taint-analysis and symbolic execution. Each technique is designed with defensive measures that attempt to hinder adversaries from bypassing Fuzzification. Our evaluation on popular fuzzers and real-world applications shows that Fuzzification effectively reduces the number of discovered paths by 70.3% and decreases the number of identified crashes by 93.0% from real-world binaries. Fuzzification will be appeared in the USENIX Security Symposium 2019.

Software Debugging. Software bugs are the root cause of most cyberattacks. They also substantially limit the productivity of developers. They are however inevitable given the size and complexity of modern software applications. Record-and-replay is an important technique for reproducing, understanding and even tolerating software failures in practice. The technique records the interactions between a program and its environment during execution by logging system calls and signals. A recorded execution can be replayed as many times as necessary for various purposes such as diagnosis of software failures and state recovery from failures. However, there are several challenges to record-and-replay real-world applications in commodity systems. First, long running programs such as server programs and user interactive (UI) programs may produce large replay logs which entail long replay times, hours or even days. Repeated replays of such long executions are hardly affordable. Second, concurrent program executions are difficult to replay due to the non-deterministic resource schedules (e.g., threading). Many existing solutions for these challenges lie solely in the area of system research or program analysis.

Dr. Lee has been looking for synergetic integration of ideas and techniques from both areas to develop novel capabilities. Particularly, Dr. Lee has focused on developing practical record-and-replay solutions to address above mentioned challenges. Dr. Lee has designed and implemented reducible execution replay that allows analytically reducing a replay log while retaining its ability to reproduce failures. In the upcoming "Big Data" era, the accuracy and space-efficiency provided by these techniques will be highly valuable for log analytics. The result of this work has published in the ACM SIGPLAN conference on Programming Language Design and Implementation (PLDI), which is considered as the most prestigious programming language and software engineering conference.

Additionally, Dr. Lee was one of the contributors in a pioneer project PRES that studies the effects of using various kinds of runtime information, such as the executed basic blocks and the synchronization events, to facilitate replay. The work has published in the ACM Symposium on Operating Systems Principles (SOSP) conference. It is widely cited (224 times) and has substantial influence on the follow-up work in the area. Dr. Lee also has developed a self-contained replay engine that addresses the limitations of PRES. It does not require any infrastructure support and features very low logging overhead as it does not log any synchronization operations or shared memory accesses. Replay is efficient as it is an incremental and demand-driven process. This work was appeared in the European Conference on Object-Oriented Programming (ECOOP) conference.

In addition to the failure replay, Dr. Lee has done research on diagnosis of performance problems leveraging program analysis techniques. Performance problems is an essential part of software development and maintenance. This is a challenging problem to be solved in the production environment where only program binaries are available with limited or zero knowledge of the source code. This problem is compounded by the integration with a significant number of third-party software in most large-scale applications. To address these challenges Dr. Lee has proposed an automated approach to analyze application binaries and instrument the binary code transparently to inject and apply performance assertions on application transactions. This work was appeared in the ACM SIGSOFT International

Symposium on the Foundations of Software Engineering (FSE), which is considered as one of the top-tier conferences in software engineering area.

Research Advising

Dr. Lee is currently a research advisor of five PhD, two master's, and two undergraduate students. Three master's students have graduated under his supervision. He is also an advisory committee member of nine graduate students at UGA. In addition, Dr. Lee is advising other graduate and undergraduate students through directed studies.