Guest Editors’ Introduction: Special Issue on Data-Driven Dependability and Security

Domenico Cotroneo, Member, IEEE, Karthik Pattabiraman, Senior Member, IEEE, and Antonio Pecchia, Member, IEEE

Assessing dependability and security properties of computer systems is an important concern for engineers and practitioners. Analysis of data and log files produced under real workload conditions by applications, systems and networks, intrusion detection systems, and monitors plays a key role for dependability and security evaluation. Data analysis is crucial in a variety of engineering tasks, such as measuring availability and reliability of a system, characterizing failures, gaining insights into the progression of security attacks, designing mitigation means and countermeasures. Both academia and industry widely recognize the inherent potential of dependability and security data analysis. Nevertheless, in spite of decades of research and methodological advances, data analysis in this specific area still raises challenging research questions due to the heterogeneity, volume and velocity of the collected data, lack of systematic, end-to-end, analysis procedures, increasing diversity of analysis objectives and emerging application domains in critical areas.

This special issue aims to concentrate novel contributions addressing dependability and security of computer systems through data analysis, and to publish consolidated research results focusing on data-driven methodologies, measurements from production systems, and analysis of large datasets. In response to the call for papers, after rigorous peer review and careful revision, the following six papers were included in this special issue. The six papers in the special issue provide valuable contributions related to log-based measurements, operating systems dependability, and attack detection:


- The paper “Learning from the Ones that Got Away: Detecting New Forms of Phishing Attacks” by C. N. Gutierrez, T. Kim, R. Della Corte, J. Avery, S. Bagchi, D. Goldwasser, and M. Cinque proposes an approach to partition the code running in the Linux kernel based on runtime profiling, in order to minimize the performance overhead. The approach is demonstrated on two device drivers and a file system.

- The paper “Towards Automated Log Parsing for Large-Scale Log Data Analysis” presents a characterization of state-of-the-art log parsers. Starting from the limitations of existing solutions, the paper proposes a parallel log parser (POP) on top of a data processing platform to parse large-scale log data accurately and efficiently. Extensive experiments are done with both synthetic and real-world datasets.

- The paper “Towards Automated Log Parsing for Large-Scale Log Data Analysis” presents a characterization of state-of-the-art log parsers. Starting from the limitations of existing solutions, the paper proposes a parallel log parser (POP) on top of a data processing platform to parse large-scale log data accurately and efficiently. Extensive experiments are done with both synthetic and real-world datasets.

- The paper “A System Architecture for the Detection of Insider Attacks in Big Data Systems” by S. Aditham and N. Ranganathan presents a technique to detect security attacks by insiders in large cloud hosted systems. The main idea is to profile the control instruction sequences of each process in the system, and to match these sequences among the replica nodes. The approach is demonstrated on two big-data processing cloud systems, namely Hadoop and Spark.

- The paper “How to Fillet a Penguin: Runtime Data Driven Partitioning of Linux Code” by O. Schwahn, S. Winter, N. Coppik, and N. Suri proposes an automated approach to partition the code running in the Linux kernel based on runtime profiling, in order to minimize the performance overhead. The approach is demonstrated on two device drivers and a file system.

- The paper “How to Fillet a Penguin: Runtime Data Driven Partitioning of Linux Code” by O. Schwahn, S. Winter, N. Coppik, and N. Suri proposes an automated approach to partition the code running in the Linux kernel based on runtime profiling, in order to minimize the performance overhead. The approach is demonstrated on two device drivers and a file system.

Measurements show that the unsuccessful execution of recovery procedures, when additional failures occur during recovery, causes system-wide outages.

- The work by P. He, J. Zhu, S. He, J. Li, and M. R. Lyu entitled “Towards Automated Log Parsing for Large-Scale Log Data Analysis” presents a characterization of state-of-the-art log parsers. Starting from the limitations of existing solutions, the paper proposes a parallel log parser (POP) on top of a data processing platform to parse large-scale log data accurately and efficiently. Extensive experiments are done with both synthetic and real-world datasets.
can detect 70 percent of the phishing emails that are not detected by a commercial filter.

The Guest Editors would like to thank all of the authors who have submitted their valuable work to this special issue. We are grateful to the Editor-in-Chief and the editorial team of this journal for their support. Moreover, we would like to thank the many experts in this field who have participated in the review process and provided helpful suggestions to the authors for improving their work. We hope you enjoy the papers.

Domenico Cotroneo
Karthik Pattabiraman
Antonio Pecchia

Guest Editors

Domenico Cotroneo received the MSc degree in computer engineering from the University of Naples, in 1998, and the PhD degree from the Department of Computer Engineering and Systems, University of Naples, Italy, in 2001. He is currently an associate professor with the University of Naples. His main interests include software fault injection, dependability assessment and field-based measurements techniques. He is serving/served as Program Committee member in dependability conferences, including DSN, EDCC, ISSRE, SRDS, and LADC. He is a member of the IEEE.

Karthik Pattabiraman received the MS and PhD degrees from the University of Illinois at Urbana-Champaign (UIUC), in 2004 and 2009, respectively. After a post-doctoral stint at Microsoft Research, Redmond, he joined the University of British Columbia (UBC), in 2010, where he is now an associate professor of electrical and computer engineering. His research interests include building error-resilient software systems, and in software engineering and security. He regularly serves on the program/organizing committees of dependability conferences such as DSN and ISSRE. He is a senior member of the IEEE, and a member of the IFIP Working Group on Dependable Computing (10.4).

Antonio Pecchia received the BS, MS, and PhD degrees in computer engineering from the Federico II University of Naples, in 2005, 2008, and 2011, respectively, where he is now an assistant professor. He is a co-founder of the Critiware spin-off company (www.critiware.com). He was a post-doc with the National Interuniversity Consortium for Informatics (CINI) in European projects. He serves as TPC member and reviewer in conferences and workshops on software engineering and dependability. His research interests include data analytics, log analysis, empirical software engineering, dependable and secure distributed systems. He is a member of the IEEE.

For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.