Proactive Digital Defenses: Next Generation Systems for Emerging Threats in Data Science & Engineering

Abstract: Data-driven science and engineering are having a transformative effect on the world, from public health, to basic science, public policy, education, and even criminal investigations. While this unprecedented increase in data has opened new possibilities, the exponential growth of data has outpaced exponential growth in computing resources. In 2015 the digital universe grew to 8.6 zettabytes, representing increase of over 300% from 2012. By 2020 the digital universe is expected to contain a staggering 44 zettabytes. Our ability to create data has exceeded our ability to store, analyze, and utilize it. These challenges are joined by emerging threats to data privacy and security with the rise of adversarial behavior in data science. For these reasons we need systems that are not only capable of storing large quantities of raw data, but doing so in a trustworthy manner, while enabling state of the art analytics.

In this talk we will explore emerging threats and problems in data science applications to medicine, climate science, natural history, and law enforcement, and outline the privacy, security, reliability, and analytics challenges to data in these domains. We will detail the emerging threats in data-driven fields, outlining novel attacks, evidence of the emergence of adversarial data science, and extant threats to privacy and data integrity. We will present novel, intelligent, systems designed to combat these threats, introduce a new formal logic for privacy preserving data operations, and demonstrate their performability and correctness, along with metrics for their improved privacy and suitability for high-assurance areas of data science. We will examine real case studies in derbarment and corruption in international procurement with the World Bank and the FBI. We will show how malicious actors manipulated the data collection and data analytics process to damage the integrity of data used by machine learning and predictive analytic processes, or the outcomes derived from these processes, to avoid regulatory oversight, sanctions, and investigations launched by national and multi-national authorities. We will present evidence for the growth of these attacks, and discuss the counter measures we are developing to enable proactive digital defenses. Finally, we will conclude by discussing future work in systems engineering for Big Data, outline current challenges, and future pitfalls of next generation systems for data science.

Dr. Eric Rozier is an Assistant Professor of Electrical Engineering and Computing Systems and head of the Trustworthy Data Engineering Laboratory at the University of Cincinnati. His research focuses primarily on the intersection of data with privacy, security, reliability, and analytics. He has received over $4M in grants for his research, and been recognized by numerous awards including his selection as one of the inaugural speakers for the USENIX Enigma conference on emerging threats, a two time Eric and Wendy Schmidt Data Science for Social Good Faculty Fellow at the University of Chicago, a National Academy of Engineering Frontiers of Engineering Education Faculty member, and an IBM Research Fellow. Before joining the University of Cincinnati, Dr. Rozier was the founding director of the Fortinet Cybersecurity Laboratory at the University of Miami where he worked to develop and commercialize new technologies in homomorphic encryption for cloud-based systems. He earned his Ph.D. from the University of Illinois at Urbana-Champaign where he worked on applications in fault-tolerance and security with the National Center for Supercomputing Applications, and the Information Trust Institute.