Impact Statement

David Mohaisen (mohaisen@ucf.edu)

My work in the domain of computer security looks into the design, analysis, integration, and evaluation of methods for decision automation and attribution of malicious actors, and has had various direct impacts on the practice, whether it is in the industry or academia. The core concepts of several of those works have been patented (see patents in the attached resume), and their implementations have been used in production to facilitate attribution and risk assessment of domain names of malicious software, domain name, or even systems. In the following, I elaborate on some of such impacts.

Behavior-based malware detection (pioneering) My work on behavior-based malware analysis, which is captured in my system AMAL, is one of the earliest works in this domain, initially published in 2013. AMAL, a tool to classify and cluster malware using autonomous feature extraction, sandboxes malicious binaries to collect fine-grained behavioral artifacts that characterize malware's usage of the file system, memory, network, and registry. A key contribution of AMAL is the extrapolation of labels on unlabeled malware samples. Since its inception (2013), AMAL was incorporated into iDefense (then a subsidary of Verisign) malware analysis toolset. Today, AMAL reduces the manual analysis effort by two orders of magnitude, while providing practical accuracies. Academically, and given that AMAL one of the earliest works in this domain, the work has been well-cited with more than 150 follow-up studies on the subject. Personally, this work was the start of a productive journey on ML/DL-based detection systems as well as adversarial machine learning techniques with 30+ publications published in top conferences, e.g., ACM WiSec, ICDCS, INFOCOM, DSN, etc.

Authorship attribution using deep learning (answering open questions) Code authorship attribution is an important problem where it is required to map a set of authors to their code snippets. The problem has witnessed a recent surge in interest, featured by various systems using hand-picked features for conducting this mapping, make the problem intractable for large-scale attribution. My work in this space, through DL-CAIS (CCS 2018) and Multi-X (PETS 2020) develops tools for automatically generating high-quality feature representations for large-scale authorship attribution. Compared the prior work, we were able to maintain a high accuracy for 10x the number of authors at a fraction of the overhead. Through Multi-X, we were able to attribute code segments in snippets with multiple authors. Our work in this domain closes key challenges in the area, by showing the feasibility of deep learning based code authorship attribution.

Securing blockchain systems (knowledge systemization) One of the key areas of effort that I have worked on during the past 5 years is blockchain systems security. My work has provided the underpinnings for foundations-driven guarantees of blockchain systems security through a deeper understanding of their attack surface. In this direction, my work explored various attack possibilities on widely deployed blockchain systems (e.g., Bitcoin and Ethereum), including partitioning attacks (ICDCS 2019), optimized mining attacks (CCS 2021), DDoS attacks (ICBC 2019), fair mining (TPDS 2021), among many others. My work has established the need for a systematic understanding of blockchain systems security by demonstrating various vulnerabilities. The impact of this work is not only to practice, but also in academia: our work systematically analyzing the attack surface of blockchains (IEEE Communications Surveys & Tutorials, 2020) has been the go-to resource to reference, as evident by more than 65 citations to date, in less than one year.

Mixing time of social graphs (paradigm shift) Until 2010, it was widely believed that social networks are "fast-mixing", and many Sybil defenses made crucial use of this property. An experimental verification of this property was lacking, and my work explored mathematical tools and used them to measure the mixing time of several social graphs. My findings show that the mixing time of social graphs is much larger than used in literature, which leads to several striking results. First, designs based on the fast-mixing property utilize a weaker property concerning the average mixing in the social graph (as opposed to the worst mixing, making most theoretical provable guarantees of these systems inaccurate). Second, current security systems based on fast-mixing properties have weaker guarantees and have to be less efficient in to compensate for the slower mixing graphs. The work presented a breakthrough, shifting interest in the community to other assumptions for system design, and is highly cited (200+ citations as of 2020).

Impact through standardization (beyond academia) Along with collaborators from TU Eindhoven, TU Darmstadt, genua GmbH, and Radboud University, I led the effort of specifying the implementation of a hash-based signature in the Internet Engineering Task Force (IETF) Request for Comments (RFC 8391). RFC 8391 provides a standard of the first hash-based signature algorithm, which is today incorporated into widely used software (OpenSSL), and is expected to be used on millions of devices. The impact of this work is rarely seen in an academic work, as such efforts are only done in industry.

Students and placement (biggest impact) At UCF (since 2017), I have advised 15 doctoral students to candidacy (+3 who switched advisors before passing candidacy). Out of those 15, ten have either graduated (7), or are to graduate within 2021 (3). Out of those 10, only one PhD student went to industry, and the rest went/will go to academia: five as tenure-track assistant professors (Wayne, Layola, Niagara, etc.) and four as postdoctoral researchers (Georgia Tech, TAMU, Northeastern). At Buffalo, I mentored two postdocs who went to academia as tenure-track assistant professors. At Buffalo and UCF, I mentored more than a dozen M.Sc. and B.Sc. students who joined government and industrial entities.

Recognitions To date, my work has received more than 3500 citations, with h-index of 31 and i10-index of 86, with a healthy trajectory. My work won the best paper award at IEEE Systems Journal 2020, IEEE DSC 2019, IEEE ICDCS 2017, WISA 2014, and IEEE CNS 2013, and was featured in news articles in various outlets, including MIT Technology Review, Science Daily, Scientific American, Financial Express, Slashdot, CBS news, The Verge, New Scientist, etc.