MINISTRY OF EDUCATION AND RESEARCH

TECHNICAL UNIVERSITY
OF CLUJ-NAPOCA, ROMANIA

Fundamental field: Engineering
Specialisation: Computer Science

HABILITATION THESIS

– ABSTRACT –

Cybersecurity Engineering in the Age of the Artificial Intelligence

Assoc. Prof. Eng. Ciprian-Pavel OPRIȘA, PhD
Faculty of Automation and Computer Science
Technical University of Cluj-Napoca

– Cluj-Napoca –
2019

Cybersecurity and artificial intelligence are two of the most important research fields today, both having a deep impact on society. While cybersecurity requires precise results and determinism, artificial intelligence introduces probabilistic reasoning, approximations, and randomness. Nevertheless, modern cybersecurity problems require modern tools, and the artificial intelligence is undoubtedly an important component of these tools.

This habilitation thesis presents my main scientific, professional, and academic achievements since obtaining my Ph.D. in 2016. I have pursued both an academic career at the Technical University of Cluj-Napoca and an industry career at Bitdefender. This dual perspective has helped me acquire skills that are valuable in both academic research and in industrial research and development.

My career, both before and after defending my Ph.D., has been centered on cybersecurity, developing tools and solutions based on artificial intelligence that either ensure cybersecurity defense or facilitate cybersecurity analysis.

The first subfield to which I contributed was malware analysis and classification, continuing the research I conducted during my Ph.D. I authored multiple articles describing efficient solutions for indexing massive file collections for efficient retrieval of related or similar programs. I also highlighted the importance of a deep understanding of detected threats, proposing solutions for automatically associating attacks with MITRE techniques and for mitigating and reducing false positive detections. Malware detection can also be performed on homomorphic encrypted data, where the Shannon entropy, an important feature for detecting packed code can be efficiently computed. Malicious programs are present on multiple platforms, including Internet of Things (IoT) devices. A recent paper presents a technique for semantic representation of binary code, agnostic to the processor type for which it was compiled.

The second research topic is networks security and the security of IoT devices, which can also be addressed at the network level. I proposed a framework for detonating suspicious programs, focusing on simulating and monitoring the network activity. The studied attacks were presented in another paper, in which I created a taxonomy for web attacks against IoT devices. I proposed multiple defense techniques against these attacks, based on anomaly detection. A major challenge for anomaly detection is reducing the number of false positives. This objective can be achieved by analyzing the detected URLs. To support this analysis, I proposed a clustering algorithm optimized for processing large collections of URLs detected by anomalies. I also proposed a solution for detecting Domain Generation Algorithms, a technique used by malware to hide its command-and-control server. The proposed method, based on machine learning, uses a lightweight classifier that can run even on low-powered routers.

The third subfield focuses on fraud detection through semantic text analysis. My research goal was to find a balance between modern techniques, which are more accurate but slower, and classic techniques, which are faster but sometimes less accurate. An important step in fake news detection is identifying the original source, which requires a similarity metric for news articles. I proposed two solutions: one based on n-grams extracted from the text, and other one semantic, based on the BERT model. Although the second solution is more accurate, the textual solution achieves reasonable results with a significantly lower running time. Another result in the area of news articles proposes a solution for clickbait detection by analyzing both the title elements and the match between the title and the article content. Text analysis is also useful for detecting phishing e-mails. Large Language Models can analyze the message text and answer questions regarding the authority, the urgency, the familiarity, the reciprocity or the consistency of a text.

The last subfield addressed is the offensive security. Often, to analyze the security of a computer system, we must think as an attacker, even if the ultimate goal is defense. One of my published papers proposes a machine learning technique to automate the pentesting process by selecting the appropriate exploit. Another paper presents an offensive security tool for red teams, providing a complete infrastructure for deploying malware on Kubernetes clusters.

Since obtaining my Ph.D., I have published:

- one book, one lecture support, and two lab guides;
- two journal papers, both indexed in WoS in the Q1/Q2 category;
- 18 papers in international conferences, of which 12 are indexed in WoS and 6 in international databases.

I also participated in six research projects, having a leading role as partner coordinator in three of them.

My professional career started in June 2010, when I was hired as a Malware Researcher at Bitdefender. Currently, I am a Principal Security Researcher Lead in the Cyber Threats Intelligence Lab at the same company. In February 2017, after obtaining my Ph.D., I became a Lecturer at the Technical University of Cluj-Napoca, where I am currently an Associate Professor in the Computer Science Department. I started teaching at the Technical University of Cluj-Napoca as a lab assistant in October 2011.

Over the years, I have taught lectures, seminaries and labs for six courses at the Bachelor level and four courses at the Master level. Currently, I teach "Operating Systems", "Fundamental Algorithms" and "Operating Systems Administration" at the Bachelor level, „Mobile Security" and „Big Data and Machine Learning for Cybersecurity" for the „Cybersecurity Engineering" Master, and „Introduction to Big Data" and „Algorithms" for the „Data Science" Master program. For each subject that I have taught, I contributed by creating new lecture or lab materials or by improving the existing ones.

Besides my teaching activities, I also supervise students for their Bachelor and Master theses, with more than 100 successful theses to date. I am also involved in organizing the ACM ICPC competition, serve as a member of the Bachelor theses evaluation committee, and act as a secretary for the Master graduation committee.

My academic development plans include adapting my evaluation methods to account for generative AI, continuously updating my lectures and labs with modern and relevant materials, building cross competences by collaborating with colleagues who teach related topics, and writing books for the lectures I teach that don't rely on well-known existing books. I also wish to continue my involvement in the ACM ICPC competition and increase the student interest in it.

For my scientific development, I plan to expand my research group, CYDAR, by supervising Ph.D. students in the cybersecurity field and participating in research projects. I will broaden my research interests to include agentic AI and the synergy between large and small AI models, which I consider important both from a fundamental research perspective and for developing industrial solutions. I also plan to employ semantic representation of code to improve my existing plagiarism detection solution.

By pursuing these objectives, I will be able to make significant research contributions, offering tools and solutions for cybersecurity and preparing a new generation of researchers with whom I will explore the cybersecurity challenges of the future. Academically, I will transition to the new paradigm in which the AI is part of the context, leveraging its advantages and mitigating its disadvantages, while maintaining high quality standards.