Ph.D. in Information Technology
Thesis Defense

March 10th, 2026
at 10:00 am
"Beta" Room – building 24

Tommaso PALADINI – XXXVIII Cycle

Artificial Intelligence-based Cyberattack Mitigation Techniques

Supervisor: Prof. Michele Carminati

Abstract:

Cyberattacks are growing in sophistication and impact, compelling organizations to continuously evolve their defensive strategies.
As the actions of threat actors often exhibit recurring patterns, defenders have turned to Artificial Intelligence (AI)-based methodologies to automatically analyze data collected from targeted software systems, both to detect anomalous behavior in real time and to learn from previously recorded intrusions to inform future strategic decisions.
This thesis explores the challenges of two popular paradigms in AI-driven cyber defense: real-time detection systems, focusing on their application in the banking fraud domain, and proactive approaches based on Cyber Threat Intelligence (CTI). Accordingly, this thesis is structured in two parts.

In part one, we investigate the robustness of Machine Learning–based banking fraud detection systems against attackers attempting to subvert the intended behavior of the learning algorithm. First, we show how attackers may manipulate data to evade or poison fraud detectors and propose a corresponding mitigation tailored to realistic financial fraud scenarios. We evaluate its effectiveness under different levels of attacker knowledge. Then, we introduce an Online Learning-based detection framework leveraging the Multiplicative Weights Update algorithm to dynamically adapt to adversarial behavior, demonstrating its capacity to minimize economic loss and maintain robustness against evolving fraudulent strategies. We conduct our experimental evaluation on real-world banking data provided by an Italian financial institution.

In part two, we examine CTI-based proactive defense strategies, focusing on the quality of unstructured intelligence sources and the challenges inherent in the Natural Language Processing (NLP) methodologies required to extract actionable insights from them. First, we present a large-scale longitudinal analysis correlating discussions in underground forums (over 88 million posts) with CTI reports spanning two decades. Our findings reveal that hacker forums have historically contributed to malware later used in real-world attacks, and that systematic CTI monitoring has only caught up within the past decade. Finally, we provide a comprehensive systematization and empirical comparison of NLP-based Tactics, Techniques, and Procedures (TTP) extraction methods, uncovering limitations in current research and frameworks, and underscoring the need for future work on ontology refinement and large-scale dataset development.

As both paradigms continue to evolve, they should be employed in a complementary manner; future real-time detection systems could progress toward the identification of adversaries and automatic deployment of optimal defense strategies, possibly informed by CTI analysis. Conversely, CTI-based methodologies should broaden their data sources by taking into account the platforms that effectively anticipated cyberattacks and refine existing frameworks, which still exhibit ambiguities that hinder automatic interpretation.

PhD Committee

Prof. **Giacomo Boracchi, Politecnico di Milano**

Prof. **Giovanni Vigna, University of California**

Prof. **Daniel Arp, TU Wien**