Abstract

This thesis explores the intersection of AI and cybersecurity at a time of unprecedented technological acceleration. As AI reshapes expectations and capabilities, cyberattacks are also becoming faster, more sophisticated, and widespread. The risk of outpacing security experts is real and demands innovative solutions. In this context, modern AI models, with Large Language Models (LLMs) spearheading, offer a promising path forward.

The first part of this thesis focuses on automated analysis of cybersecurity logs. Although vast amounts of log data can be collected effortlessly, extracting actionable insights – such as identifying patterns across logs or detecting critical events – remains a significant challenge.

I start with SSH attack logs and introduce *LogPrécis*, a tool that utilizes pretrained models to identify attackers' intents by mapping SSH session data to MITRE attack classes. Collected from multiple honeypot deployments, this approach enables LogPrécis to significantly enhance security analysis. Specifically, it automatically reduces more than 400,000 unique SSH sessions to approximately 2,000 representative attack fingerprints, allowing security experts to efficiently track the evolution of the attack, compare incidents between deployments, and gain deeper insight into SSH threats.

Next, I analyse a real-world firewall deployment featuring a vast and highly diverse log dataset—approximately 2 million entries spanning 232 applications. To tackle this complexity, I developed a hybrid system that combines specialized small models with LLMs. The LLM synthesizes multiple data sources – including raw alerts, external threat intelligence, and predicted risk assessments – to deliver tailored support to security experts, particularly for logs associated with high-risk threats.

In the second part of the thesis, I explore key open challenges that still hinder the widespread adoption of LLMs in cybersecurity.

Specifically, in the framework of encrypted traffic classification, I highlight the risks of over-reliance on pre-trained models (LLM and Computer Vision style), showing how they can easily exploit spurious correlations in security data – such as recurring domain names or packet sequence numbers – leading to deceptively high performance (up to $+70\%$ accuracy). This critical flaw has even impacted studies published in top-tier conferences.

Finally, I tackle the challenge of constrained generation, *i.e.*, to ensure that the model generates valid answers with respect to given constraints (*e.g.*, output format must be a JSON format). This is a critical factor for the practical deployment of LLMs in real-world scenarios. I demonstrate that the performance of current standalone LLM solutions – regardless of model size or architecture – significantly declines as the number of constraints increases, with accuracy dropping up to -30%. In contrast, the proposed solution, that leverages a small and dedicated support-model, *FoCusNet*, maintains robust performance, achieving a +10% accuracy improvement over the best baseline and paving the way for more reliable and constraint-aware LLMs in real-world applications.