Defending the Cloud–Edge Continuum: LLM-Based Approaches to DDoS Detection and Mitigation (A Presentation)

Yinuo Zhang

Department of Computing Science, Umeå University, Sweden

yinuo.zhang@cs.umu.se

Abstract: The cloud–edge continuum is emerging as a paradigm for low-latency, real-time IoT and edge computing applications. This architecture distributes services across cloud, fog, and edge layers to meet stringent latency requirements. However, this expanded landscape increases the attack surface and enables sophisticated multi-vector cyber threats. In particular, distributed denial-of-service (DDoS) attacks can leverage botnets of edge devices to generate massive traffic volumes that overwhelm servers and networks. Such threats to availability demand advanced, distributed detection strategies.

To address these challenges, we explore Large Language Models (LLMs) such as BERT for anomaly detection in network traffic. Recent work shows that traffic flows can be tokenised as pseudo-natural-language sequences, allowing pre-trained BERT models to capture rich contextual patterns in packet data. We conduct literature work and do experiments comparing on cloud benchmark datasets to confirm that the BERT model "significantly outperforms traditional ML and DL methods" in identifying complex attack patterns.

However, these accuracy gains come at a cost, large transformer models have high inference latency and memory use. Deploying BERT on edge devices requires careful optimisation. We analyse performance–efficiency trade-offs and apply model compression (quantisation, pruning, distillation) to reduce LLM size and speed up inference. This presentation is framed as both a survey of cloud–edge cybersecurity research and a report on recent technical experiments. It highlights the promise of LLM-based detection for DDoS and other cyberattacks in the cloud–edge continuum, while noting the critical importance of compression and efficient inference for practical deployment.

Keywords: Cloud-Edge Continuum, LLM, Cyberattack, DDoS, Model comparison