ELECTRICAL, ELECTRONICS AND COMMUNICATIONS ENGINEERING

Comitato ICT - Machine learning based solutions to monitor real-time communications

Funded By	COMITATO PER LA GESTIONE DEL FONDO PER LO SVILUPPO DELLA RICERCA E DELLA FORMAZIONE NEL SETTORE DELL [P.iva/CF:97623280019]
Supervisor	MELLIA MARCO - marco.mellia@polito.it
Contact	LEONARDI EMILIO - emilio.leonardi@polito.it
Context of the research activity	This thesis will focus on anomaly detection based on machine learning on complex multivariate time-series collected from communication networks in general. It will tackle the following research questions: • How to effectively collect data to get a comprehensive picture of operational networks? • How to efficiently collect and process the time-series to spot anomalies on a timely fashion? • How to identify anomalies considering a complete view of the system using graph neural networks?
Objectives	Unsupervised Machine Learning aims at finding unexpected patterns in data. It has been used in several problems in computer networks, from the detection of port scans to the monitoring of time series collected from Internet systems. Developments in AI bring new possibilities for anomaly detection too. Indeed, data-driven approaches and machine learning have seen widespread application in anomaly detection, and new AI algorithms, such as auto-encoders, adversarial networks, and graph neural networks have accelerated this trend. Classic anomaly detection approaches are far from appropriate for complex multi-dimensional problems. They either produce an unbearable high number of irrelevant anomalies or miss complex cases characterized by events only noticeable when taking multiple sensors into account simultaneously. Multi-dimensional time series are the norm for some security applications. This is the case for analysing the data collected by network monitoring systems. Finding relevant anomalies in these complex multivariate data is essential for security analysts, which must (i) pinpoint cases where attackers have succeeded in breaching the system – thus producing relevant traces;

	and (ii) identify when attackers may have taken control of the system altogether – thus representing a threat for the network environment. A second approach is to monitor the system as a graph, where nodes represent the devices and edges represent the communications they establish. Graph neural networks (GNN) would allow us to model the communication patterns and then identify anomalies in edges or nodes. The thesis will focus on the development of time-evolving GNN-based solutions to process data and detect anomalies.

Skills and competencies for the development of the activity	• Knowledge of network security • Knowledge about operation of internet • Knowledge of basic machine learning algorithms • Knowledge of classic anomaly detection algorithms • Python programming