Esplainable Artificial Intelligence for Network Intrusion Detection in Industrial Control Systems

Léa Astrid Kenmogne*[1]

[1]Inria, CNRS, Grenoble INP, LIG – Université de Grenoble-Alpes – France

Résumé

The gradual increase in successful attacks against Industrial Control Systems (ICS) highlights the pressing need to implement defense mechanisms to accurately and quickly detect resulting process anomalies. Industrial systems are critical and very sensitive systems and an anomaly or a simple intrusion could have serious consequences. Considering their particularity and their criticality, it becomes urgent to set up Intrusion Detection Systems (IDS) in order to be able to detect process-oriented attacks. Signature-based IDSs that already exist do not meet our requirements because they usually identify abnormal behavior by matching it against pre-defined patterns of events that describe known attacks. As part of this work, we will set up an anomaly-based IDS which profiles normal behavior and attempts to identify anomaly patterns of activities that deviate from the defined profile in order to be able to detect new attacks. We will use artificial intelligence techniques for behavioral detection of anomalies and the objective is not only to have good performance of our model, but also and above all to understand how the model makes predictions. The model that we will train must be explainable so that operators can trust it. L'augmentation progressive des attaques réussies contre les systèmes de contrôle industriels (ICS) met en évidence le besoin pressant de mettre en œuvre des mécanismes de défense pour détecter avec précision et rapidité les anomalies de processus qui en résultent. Les systèmes industriels sont des systèmes critiques et très sensibles et une anomalie ou une simple intrusion peut avoir des conséquences graves. Compte tenu de leur particularité et de leur criticité, il devient urgent de mettre en place des systèmes de détection d'intrusion (IDS) afin de pouvoir détecter les attaques axées sur les processus. Les IDS basés sur des signatures qui existent déjà ne répondent pas à nos exigences car ils identifient généralement un comportement anormal en le comparant à des modèles d'événements prédéfinis qui décrivent des attaques connues. Dans le cadre de ce travail, nous mettrons en place un IDS basé sur les anomalies qui établit le profil d'un comportement normal et tente d'identifier les modèles d'activités anormales qui s'écartent du profil défini afin d'être en mesure de détecter de nouvelles attaques. Nous utiliserons des techniques d'intelligence artificielle pour la détection comportementale des anomalies et l'objectif n'est pas seulement d'obtenir de bonnes performances de notre modèle, mais aussi et surtout de comprendre comment le modèle fait des prédictions. Le modèle que nous allons entraîner doit être explicable afin que les opérateurs puissent lui faire confiance.

*Intervenant