A novel embedding-based framework improving the User and Entity Behavior Analysis

Thomas Anglade[1] and Christophe Denis[2] and Thierry Berthier[3]

[1]Data scientist, iTrust, tanglade@itrust.fr
[2]Sorbonne University, LIP6, Paris, France. christophe.denis@lip6.fr
[3]Limoges University, Limoges, France. thierry.berthier@unilim.fr

Abstract.

In the recent years, the number and the variety of cyber attacks has been constantly growing. The landscape of cyber-attacks has become extremely large (DoS, DDoS, phishing, C&C, botnets, malwares, ransomwares, etc.). UEBA (User and Entity Behavior Analysis) is today the best solution that companies need to use to adapt to these changes. Using UEBA, companies do not track security events or monitor devices. Instead, they track all the users and entities in the system. They use machine learning algorithms and statistical analyses to know when there is a deviation from established patterns.
This paper propose a novel embedding-based framework that facilitate UEBA by projecting sparse and unstructured log data into a new mathematical space in which numerous behavior trends and changes can be analyzed in a simpler and more visual way than using typical deep learning algorithms. The last part of the paper deals with the validation and the explanation of prediction obtained by black box Machine Learning methods. Indeed, the he operational benefit of using Machine Learning methods is recognized but is hampered by the lack of understanding of their mechanisms, at the origin of operational, legal and ethical operational problems. This is largely dependent on the ability of engineers, decision-makers and users to understand the meaning and the properties of the results produced by these tools

Keywords:
cybersecurity, UEBA, Machine Learning, Explainable AI, node2vec