SecKG: Leveraging attack detection and prediction using knowledge graphs

Cameron Noakes

January 2022

1 Sentence Overview

In this paper, the researchers carry out research to propose an overall approach that uses a combination of both knowledge graphs and machine learning techniques to detect and predict cyber attacks to help better defend against threat actors.

Using Cyber Threat Intelligence, helped to build a knowledge graph that processes event logs in order to detect attack techniques and learn how to predict them. The events in the logs are in general used as datasets to provide specific information without needing all the data.

2 Introduction

The research theoretic tool of graphing was discussed to be the most convenient approach to catch the different coordinated steps of an attack and to be used in relation to the event logs.

The detection module finds evidence for each specific attack technique, while the prediction module allows, using machine learning techniques, their prediction at early stages. The proposed approach provides additionally two extensions for attack detection and prediction.

A knowledge graph was built using open and standard CTI sources and mainly the ATT&CK framework to gather knowledge on attacks from different sources.

3 Related Work

Discussion of how attack trees and attack graphs are too simplistic for complex, comprehensive attacks to be captured and analysed therefore requiring the machine learning and the self-made knowledge graph by the researchers for the proposed approach.

The literature discusses the development of a similar proposed solution from another paper, where the researchers of the related paper support the detection of multi-step attacks from events and use external CTI sources and propose an automated method for processing security events to build a knowledge graph. This related paper is similar to the literature of what the researchers developed. Comparing both of these related papers can give insight for the reader into understanding more of the underlying features for the proposed solution that was developed and the research accompanied by it.

4 Background and ATT&CK usage

Cyber Threat Intelligence (CTI) is the collecting of information on threats and threat actors, this information can then be used for analytical conclusions or to proposed new models.

The researchers built a schema that was titled SecKG (where the KG stands for knowledge graph), using different information sources, mainly the ATT&CK framework and related open source projects. This knowledge graph defines the interest that is wanted when capturing event logs and the relationships between these concepts. The researchers define a set of inference rules that query the knowledge graph for evidence of specific attacks.

For the prediction model, the researchers created a "Knowledge Graph Convolutional Network" (KGCN) to learn about how to predict each attack technique.

The researchers use algorithmic pseudocode to show how the processes within the knowledge graph.

5 Metamodel: the SecKG schema

To build the proposed schema, the researchers identified that basing it on an already widely used model is advantageous, the researchers based the schema on the CAR data model which can be monitored from a host-based or network-based perspective.

Classes that represent the abstract entities are titled as nodes in the KG knowledge graph by the researchers when developing the SecKG schema to help capture the necessary information embedded into security logs.

6 Conclusion

In this paper a proposition is made to create a way to identify cyber attacks as they arise or just after they are initiated, since most security solutions identify these risks closer to the end of the attack rather than the initial stages. Development of a SecKG schema was conducted to use data in event logs to help detect and prevent more appropriately whilst using a number of attack techniques from the ATTCK framework to accurately develop the schema Metamodel.

7 What problem does this solve?

Acknowledgement of how most security solutions detect attacks either at the end of the attack, when the attacker has reached there goal, or after the attack is fully complete, this is a security issue as it is often too late to prevent any of the cyber attacks.

8 How do they solve it?

They proposed building a knowledge graph based on the dataset of event logs to be able to predict cyber attacks before the attacks are almost or fully complete, which tends to be in most cases. This will help prevent attackers in real time and stop ongoing cyber attacks from adversarial attackers.

This model is proposed to use machine learning and knowledge graphs to detect and prevent cyber attacks as they arise or close to the initial attacking stage.

9 How do they assess it?

The researchers have expressed how they are experimenting with using larger datasets to identify more advanced attacks when initiated and therefore a higher success rate on the machine learning algorithm due to having more data points to learn from.