Automatic Mapping of Vulnerability Information to Adversary Techniques

Cameron Noakes

February 2022

1 Sentence Overview

A mathematical paper detailing machine learning algorithms to accurately map popular vulnerabilities (CVEs and alike) to ATT&CK tactics and techniques where 8,077 examples from public, open datasets prepared by ENISA were used.

The researchers propose a method to automatically map software vulnerability and cyber threat intelligence reports using a multi-label classification approach to ATT&CK tactics and techniques.

They took the vector of the vulnerability textual description and classified using various multi-label classification methods to identify and evaluate different measures. This helped to identify that the LabelPowerset method with Multilayer Perceptron performs best in their proposed experiment.

Explored and experimented with various multi-label classification methods to compare the performance and efficiency of th proposed system.

2 Introduction

The research and study enabled a method to automatically map the vulnerability information from textual descriptions to adversary techniques and tactics in the ATT&CK framework matrix. Developing a multi-label classification model can be adequate considering how a specific vulnerability can be mapped into various adversarial techniques.

The researchers identify that with this research and literature a practical element to the paper can be transposed and used using the classic multi-label classification algorithm. They continued to explain how they experimented with various multi-label classification methods to evaluate the best approach to automatically map the vector representations to ATT&CK tactics and techniques.

This paper is extremely similar to the BIBTEX 'lakhdhar2021machine' which outlines almost the exact same and using the same methods, one was heavily influenced by the other.

3 Background

The researchers identify an elaboration on the 'software vulnerabilities' they mentioned in the abstract detailing CVEs of focus and interest, which entails a direct mapping from CVEs to ATT&CK tactics and techniques, this identifies the practical aspect to the work conducted due to red and blue team operators being able to find proper mitigation strategies for remediation or potential mitigation checking for offensive teams. There is also a focus on CAPEC.

Identification of definitions are given to clarify what each terminology refers to, in order to not leave knowledge gaps or misinterpretation when it comes to reading and reviewing the literature such as adversary group and mitigations.

Acknowledgement of how the ATT&CK framework has a wide range of categories and sub categories in the matrix used for threat modelling, it is exclaimed that there are 266 techniques/sub-techniques of 12 tactics.

Since multi-label classifications are used to develop the proposed system by the researchers, this enables the examples to be associated with a set of labels which can be given by algorithmic mathematical representations which the researchers included.

4 Multi-label classification

Multi-label classification methods are described by the researchers to help improve the readers knowledge and identify key aspects of the multi-label classifi-

cations to better understand how it was used to build and develop the proposed solution of the mapping to ATT&CK techniques and tactics.

The methods are given in the literature and can be seen below:

• Algorithm adaptation methods - Pre-existing ML learning algorithms that are adapted, extended, and customized for multi-label classifications.

• Problem transformation methods - Transforms the multi-label classification into one or more single-label classifications.

• Ensemble classification - Developed on top of existing problem transformation or algorithm adaptation methods.

The literature uses mathematical formulas to calculate variables such as accuracy (the subset accuracy/ Exact Match Ratio) in respect to the h instance and the F1 scores with Precision and Recall also being calculated for the mapping using multi-classification methods.

5 Conclusion

Within this literature, the research conducted and the method proposed to solve the problem at hand is an effective one which provides accurate machine learning and a multi-classification approach to mapping known CVE vulnerabilities to tactics and techniques in the ATT&CK framework matrix where 8,077 examples from public, open datasets prepared by ENISA were used.

6 What problem does this solve?

There is not many proper mitigation strategies given in the textual or numerical descriptions of popular software vulnerabilities/vulnerability databases which can make it more difficult to accurately remediate the vulnerabilities.

7 How do they solve it?

The development of the research and proposed solution is through the use of mathematics detailing calculations and machine learning algorithms to accurately map popular vulnerabilities (CVEs and alike) to ATT&CK tactics and techniques.

8 How do they assess it?

The researchers used the best performing neural LabelPowerset model as analysis on the e remaining 200 examples as a prediction only task the experimental result could not be fully tested in real-life due to the experimental nature of the research and development conducted.