Modeling Cyber Threat Intelligence

Cameron Noakes

January 2022

1 Sentence Overview

Proposition of a cyber threat intelligence (CTI) data model enabling large consumption of all relevant data, validation and analysis. This paper says its main focus is on the strictness of the data model, enabling automation of new knowledge.

An ACT data model is developed by the researchers to aid in the ease and simplification of challenges arose for CTI analysts when regarding consumption, normalizing and analyzing of data sources.

2 Research Motivation

Identification of data quality is a main metric for CTI is mentioned by the author researchers, when conduction automation for for finding such data it is crucial to enable the computer the skillset to be able to identify specific data types within various datasets with flexibility within the schema of a data model. It is mentioned that the format consistency should be enforced and can be done through a strict data model.

It is explained that the threat intelligence analysis (CTI analysts) often require a large amount of knowledge in order to carry out their work, adding knowledge into the strict data model of ACT will allow the knowledge to be available to more analysts.

3 Related Work

Acknowledgement of how determining the ways at structuring CTI sources and data can influence the outcome and end results depending on what the CTI is used for.

ATT&CK uses a data model with defined relationships for structuring their knowledge base and due to this, can be incorporated within the proposed solutions and research for the development of the ACT strict data model using Apache Cassandra 6 and Elasticsearch 7 on the back end.

4 Methodology

The development of the ACT model by the researchers was an iterative process based on the relevant threat intelligence data they had available to create a strict data model to help CTI analysts with their professional work. The model used Apache Cassandra 6 and Elasticsearch 7 on the back end for functionality. Prototyping and testing was said to have been developed including agile type development principles and standards.

5 Results

The results were developed for visualisation through some methods the researchers expand upon such as the graph engine which enables graph querying with the use of the graph query language Gremlin.

The resulting data from the proposed model stores the information as facts (one or more objects within the model) for the storage of data for the model and new facts can be used to overwrite old ones, where the old facts cannot be removed to be able to traverse the available data back and forth in time. The researchers explain further the reasoning behind why facts cannot be deleted since it allows the traversal of time to see what information was available at different instances to learn from mistakes.

6 Conclusion

The development of a proposed solution is implemented for a strict ACT data model based on available cyber threat intelligence (CTI) data with the usage of ATT&CK framework for the structuring of the development and the overall model.

This remediates the issue of CTI analysts requiring more information and professional education to complete the tasks for their work, creating an easier and more efficient workflow.

7 What problem does this solve?

The researchers explained how the consumption, normalizing and analyzing of cyber threat intelligence (CTI) from heterogeneous sources can become challenges for CTI analysts but the researchers didn't expand on how. The ACT data model was proposed to address these challenges. Successful defense against threats and the overall use of automation takes the available CTI data to another level to allow more insight into the threats.

8 How do they solve it?

The researchers developed a data model titled ACT in order to solve the challenges of how sources of data can be challenging for CTI analysts for consumption, normalizing and analysis of data. The proposed solution should fix this issue and is the main development solution by the researchers to adjacently contribute to their research.

9 How do they assess it?

They assess the proposed model through the use of data storage as facts to be able to compare the earliest data to the current to see how learning is available from previous data. Wrongful classifications or attributions may be easily found through such evaluations and therefore can create better datasets.