Introduction

The MITRE ATT&CK Framework (MAF) is a comprehensive knowledge base of adversary Tactics and Techniques based on real-world observations.[5] There are 14 Tactics in the framework all of which describe the adversary's goal or reason for performing the action. There are over 200 techniques that are further broken down into sub-techniques within the framework. These techniques describe how the adversary achieved the goal. When a cybersecurity analyst obtains Open Source Intelligence (OSINT)[6] they must read the entire piece of unstructured data and classify parts of it within the MAF for the overall goal that our organization has detection mechanisms in place that will detect the given technique. This takes an enormous amount of time from a very expensive analyst who could better use their time on higher-level tasks such as constructing campaign analysis reports.

Proposal

I plan to follow the work of Y. -T. Huang et al. However, I would like to deviate some by comparing the accuracy of the DeBERTa model fine-tuned using the dataset (see link below) to that of zero-shot LLM such as ChatGTP. The reason for this is that I have seen cybersecurity analysts use ChatGTP to obtain a MAF mapping. When I tried using a single record in the dataset I got the following results:

Prompt:
Using the MITRE ATT&CK framework, identify the technique and tactic in the following statement: "During Operation Wocoa, threat actors encrypted IP addresses used for 'Agent' proxy hops with RC4."

Response:
Technique: T1027 Obfuscated Files or Information
Tactic: Defense TA0005 Defense Evasion

At first glance, this seems like a reasonable mapping but upon further analysis encrypted IPs of the Wocoa operation are classified as T1001 Data Obfuscation under TA0011 Command and Control. I am up for the challenge of creating a more accurate classification model and reporting those results to my team.

Relevant Literature

1. Y. -T. Huang et al., "MITREtrieval: Retrieving MITRE Techniques From Unstructured Threat Reports by Fusion of Deep Learning and Ontology," in IEEE Transactions on Network and Service Management, doi: 10.1109/TNSM.2024.3401200.
2. R. Marinho and R. Holanda, "Automated Emerging Cyber Threat Identification and Profiling Based on Natural Language Processing," in *IEEE Access*, vol. 11, pp. 58915-58936, 2023, doi: 10.1109/ACCESS.2023.3260020.
3. Z. Liu, "A Review of Advancements and Applications of Pre-Trained Language Models in Cybersecurity," *2024 12th International Symposium on Digital Forensics and Security (ISDFS)*, San Antonio, TX, USA, 2024, pp. 1-10, doi: 10.1109/ISDFS60797.2024.10527236.
4. arXiv:2311.08807 - Used BERT for classification of 'CTI reports'

References

5. MITRE ATT&CK Framework: https://attack.mitre.org/
6. SANS: https://www.sans.org/blog/what-is-open-source-intelligence/

Dataset

https://media.githubusercontent.com/media/dessertlab/cti-to-mitre-with-nlp/main/data/dataset.csv