Building a dataset of real-world cyber-attacks with Attack Flow (Ver 3)

Client: Associate Professor Hung Nguyen

Description

Attackers typically combine multiple techniques and procedures to compromise a system. Until recently, defenders track adversary behaviors individually, often focusing on only one specific action at a time. This mismatch between how attackers operate and how defenders try to track them has caused a significant gap in cyber defense. To address this problem, the MITRE Center for Threat-Informed Defense (Center) launched the AttackFlow project (https://medium.com/mitre-engenuity/attack-flow-beyond-atomic-behaviorsc646675cc793). The key idea is to model sophisticated attacks using models that capture the sequenceof attack steps, the context within those sequences, as well as the relationships among them. Such a model enables additional defensive capabilities that make defenders much more effective.

In this project, we will design and implement a system to facilitate building a corpus of real-world attacks using the open-source attack flow framework (https://github.com/center-for-threat-informed-defense/attack-flow). The output is a set of attack flow models that describe real-world cyber attacks. We will contribute these maps directly to the MITRE project. We will also develop algorithms that help defenders use the attack flow data to better defend their systems.

An incident report in cybersecurity is a detailed documentation of a security incident that has occurred within an organization's digital environment. A security incident refers to any adverse event or activity that potentially jeopardizes the confidentiality, integrity, or availability of an organization's information systems, data, or network resources. Incident reports play a crucial role in managing and responding to these incidents effectively.

The proposed system for creating a dataset of real-world cyber-attacks with Attack Flow caters to the aforementioned problem by providing a corpus of attack flows.

Quality control is of paramount importance for the success of this project. The project involves handling incident reports, which are composed in human language according to a predefined standardized format. Typically, these reports are authored by cybersecurity experts. The standardized format aims to faithfully represent the content of the original reports.

There’s no restriction on the project apart from using proprietary data or tools. Students are free to use any APIs, libraries, and data sets that are reliable and those should be documented.

Expected System Functionalities:

The basic system functionalities should include the following.

1. **Document annotation support**: As the first step, we need to scrape, download, store and provide metadata about the attacks and the incident reports. The system users

should be able to upload incident report documents (e.g., MS Word, PDF). The metadata of incident reports include date of the attack, target(s), attacker(s), authors of the report, and links to all related documents. Annotating an incident report could be a highly time-consuming activity. You could use recent advances in AI to help, for example using natural language processing model such as ChatGPT, LLAMA, etc, yet need to ensure the accuracy of annotated data.

Secondly, we need to facilitate the processing of the incident reports for each attack. This component helps attackflow builders to extract key information and annotate the incident reports, summarising key information that is required for actually building the attack flows. Information includes the techniques used, the assets that each technique unlocks, and the flow between techniques. The annotation could also be shared with the attackflow to provide a clearer understanding of the incident's details, analysis, and recommendations. This collaborative approach helps incident responders, management, and stakeholders make informed decisions and take appropriate actions in response to the incident. The documents with annotation details should be stored for future use. Document version control is essential to track the changes made by different users.

2. **Standard dataset generation**: The third functionality is to convert the incident reports (using the annotations) to attackflow per the format required by the MITRE framework. For this functionality, a clear understanding of the MITRE format and requirements is needed. The software solution should provide a clear template for the attackflow builders to follow. The actual task of encoding the attackflow could be done automatically via some smart algorithms or manually with the help of the software.The annotated data should be automatically mapped to a standardized format provided by the user. It is essential to keep track of the incident report document and the generated data file. The standardized format will be provided by the client in the initial stage of the project. Once the dataset is prepared, its usage extends to the entire cybersecurity industry. Initial uploaded incident reports should also be saved. This is necessary for maintaining data authenticity and governance.

3. **Attack flow visualisation**: We expect the visualisation of the attackflow and checking for the correctness of the produced attackflow under this objective. For visualisation, we could reuse the MITRE provided visualisation tool. Visualisation then is used for the attackflow builders to check whether all the techniques, the assets and the connections between them have been encoded correctly – both syntactically and semantically. Here, double checking with the annotations in functionality 2 is important. This is a key step to guarantee the quality of the outputs. Therefore, in this stage, the system should be able to visualize the attack flow using the generated data files. A sample of visualization can be found here (https://github.com/center-for-threat-informed-defense/attack-flow). Each diagram must be associated with a source document (incident report) to ensure clarity and traceability. This enables users to easily access both the source document and the corresponding attack flow diagram, facilitating comprehension.

4. **Validation**: It is important to validate the system functionalities through creating at least "to be inserted" attack flow files and getting client approval. Further, it needs to check whether the incident report documents are annotated correctly and need to validate the created attack flows.

This **system can be implemented as a standalone or web platform**, with web app the **preferred option** as it allows easy deployment and integration with MITRE framework. However, the system should be easy to use with a minimum number of clicks and navigation steps.

The software will be evaluated on two key metrics: (1) the improvement in efficiency for translating incident reports into attack flow – this is measured by the number of attackflows the group manage to produce using the software, (2) the quality of the attack flows – this is measured by the fidelity of the final attackflows in capturing key details of the attacks as described in the incident reports. We will be using the attackflows in the MITRE corpus for quality checking.

Resources:

- The DFIR reports: https://thedfirreport.com/
- Cyber peace institute: https://cyberconflicts.cyberpeaceinstitute.org/threats/attack-details
- Sample visualization: https://center-for-threat-informed-defense.github.io/attack-flow/visualization/#att-ck-navigator
- CISA resources: https://www.cisa.gov/resources-tools/resources
- Incident reports with predefined tags >> check MITRE dataset: https://github.com/center-for-threat-informed-defense/attack-flow/tree/main/corpus
- MITRE sample attack flows: https://center-for-threat-informed-defense.github.io/attack-flow/example_flows/#list-of-examples

- Code and example attackflows produced in previous student projects: https://github.com/CathHuo378/ML_model_for_attack_flow

Initial user stories:

Document annotation support:

- As a data annotator I want to upload a file, annotate it and save it so that users can use the file later for analysis.

Attack flow visualization:

- As a user I want to download the attack flows so that identify vulnerabilities and weaknesses in their systems and networks.