Title: Building a dataset of real-world cyber-attacks with Attack Flow (Ver 2)
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 Attack Flow project (https://medium.com/mitre-engenuity/attack-flow-beyond-atomic-behaviors-c646675cc793). The key idea is to model sophisticated attacks using models that capture the sequence of 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 cyberattacks. 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.

The project's primary objectives are as follows:

1. Document annotation support: The system users should be able to upload incident report documents (e.g., MS Word, PDF) and annotate them manually with multiple tags and codes (both pre-defined and user-defined). These documents with annotation details should be stored for future use. Document version controlling is essential to track the changes made by different users.

2. Standard dataset generation: 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 generate data file. The standardized format will be provided by the client in the initial stage of the project.

3. Attack flow visualization: The aim of this stage is to develop a UI to effectively visualize attack flows to enable users to study the attack flow in-detail. 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).

4. Validation: It is important to validate the system functionalities through creating at least “to be inserted” of attack flow files and get client approval.

This system can be implemented as a standalone or web platform. However, the system should be easy to use with a minimum number of clicks and navigations steps.