Objective

We are currently building a threat modelling tool that takes in information about a system by prompting the user to (1) fill in a questionnaire about the system details and (2) draw an architecture diagram showing the different components and how data flows through them. Using a public threat knowledge base (MITRE’s ATT&CK library), we can then determine which attack technique is applicable to the system (or in other words, which attack technique is a likely threat for this system).

To that end, the first phase of the internship is to build a microservice that maps each threat in the MITRE ATT&CK knowledge base to a threat category. There are multiple ways to categorize threats: STRIDE is one of the most commonly-used way. Since it is not scalable to ask a human expert to map the techniques in MITRE ATT&CK to STRIDE, we want to do the mapping automatically using unsupervised ML techniques. The first two weeks involve searching for tools or conference papers that do this mapping and testing out those solutions. The next two weeks involve using LLMs to perform the mapping instead and evaluating its accuracy. Finally, we will choose the better approach and make it adaptable for a given set of threat categories (i.e., replacing STRIDE with another threat model).

The second phase of the internship involves understanding cloud architecture design and translating it into a diagram that’s suited for threat modelling. For the first task, one of the crucial elements of threat modelling is identifying trust zones. Instead of relying on a human to mark out the trust zones, we can do this automatically by identifying different subnets as different areas of trust. The second task involves identifying design problems in architecture diagrams. For example, someone unfamiliar with cloud architecture may draw a VPC without placing it inside a cloud cluster box. We need to define such architectural constraints and automatically check that the constraints are met.

Timeline

Period: 15 Jan – 31 May 2024

W1 to W2 – briefing, Search for similar tools/papers, Download stride-to-att&ck mapping code, extract ground truth, and test

W3 to W4 – Use OpenAI’s API calls to map att&ck to stride, compare with ground truth and existing ML methods

W5 to W6 – Map att&ck to rapids and extract explanation and mapping, Make model customizable based on threat model

W7 to W10 – Identify trust zones

W11 to W16 – Codify architectural constraints and checker

W17 to end – Either IaC-to-Diagram or other architecture tasks

Tasks

**Phase 1:**

**Feature:** Map ATT&CK techniques to threat model automatically

Input: Description of ATT&CK technique

Output: Threat category and explanation

**Feature:** Customize threat model

Input: Description of new threat categories

Output: New mapping of ATT&CK technique to threat model

**Phase 2:**

**Feature:** Identify trust zones in architecture diagram automatically

Input: JSON representing architecture diagram

Output: Updated JSON with each node in architecture diagram having a trust level attribute

**Feature:** Identify problems with drawn architecture diagram

Input: JSON representing architecture diagram

Output: Design problems in architecture diagram

References

<https://github.com/sebyakuya/stride-classifier>

<https://ostering.com/blog/2022/03/07/capec-stride-mapping/>

<https://cheatsheetseries.owasp.org/cheatsheets/Secure_Cloud_Architecture_Cheat_Sheet.html>

<https://www.claranet.com/us/blog/2021-04-01-security-architecture-review-cloud-native-environment>