تعریف پروژه: (دانشجو می‌تواند با اضافه کردن فاصله لازم بر روی فایل قابل ویرایش این سند، توضیحات خود را در هر یک از قسمت‌های زیر تایپ کند.)

1- مقدمه (بیان مسئله کاربردی، ضرورت، انگیزه، اهداف، و چالش‌های انجام این پروژه):
پایگاه دانش CAPEC حاوی اطلاعات الگوهای حمله مورد استفاده مهاجمین در بهره‌برداری از آسیب‌پذیری‌ها و نفوذ به شبکه‌ها است. این پایگاه دانش برای افزایش درک رفتار مهاجمین و بهبود روش‌های دفاع در برابر حملات توسط تحلیل‌گران و متخصصین امنیت استفاده می‌شود.

به طور مشخص در سامانه های شکار تهدید برای ایجاد فرضیه های اولیه تهدید می توان از اطلاعات جمع‌آوری‌شده از سامانه‌های مدیریت رویدادهای امنیتی مانند SPLUNK استفاده نمود. بدین منظور، اطلاعات خروجی سامانه SPLUNK با پایگاه دانش CAPEC مطابقت داده می شود و الگوهای حملات انجام شده و یا در حال انجام روی شبکه تحت نظارت SPLUNK استخراج می‌شود. این الگوها می‌تواند برای ایجاد فرضیه تهدید و شروع فرایند شکار تهدید مورد استفاده قرار گیرد.

بر همین اساس، ارائه و پیاده‌سازی راه‌حل‌های خودکار برای نگاشت داده‌های خروجی سامانه‌های مدیریت رویدادهای امنیتی (مانند SPLUNK) به پایگاه دانش CAPEC ضرورت پیدا می‌کند. این راه‌حل‌ها می‌توانند از روش‌های یادگیری ماشین بهره ببرند. برای ارزیابی و آزمون روش‌های مبتنی بر یادگیری ماشین به مجموعه داده‌های برچسب‌گذاری‌شده نیاز است. به همین سبب هدف از انجام این پروژه در دو گام تعریف می شود. ۱- ایجاد یک مجموعه داده برچسب‌گذاری‌شده از داده‌های خروجی SPLUNK با استفاده از الگوهای حمله جمع‌آوری‌شده در پایگاه دانش CAPEC است. ۲- سپس پیاده‌سازی روش‌های یادگیری ماشین نظارت شده برای برچسب گذاری داده‌های خروجی SPLUNK

2- مروری بر پروژه‌ها و سامانه های مشابه و بیان نقاط قوتی که با انجام این پروژه حاصل می‌شود:
تحلیل گران امنیت از اطلاعات خروجی SPLUNK برای شناسایی حملات و تهدیدهای امنیتی و پیشگیری از آنها استفاده می کنند. مقالات گوناگون [1,2] از پایگاه دانش CAPEC برای برچسب‌گذاری لاگ های SPLUNK و توسعه مدل‌های مورد استفاده در شناسایی تهدیدها استفاده نمودند. برچسب گذاری اطلاعات خروجی SPLUNK به صورت دستی هزینه و زمان قابل توجهی نیاز دارد. در این پروژه قرار است با آماده سازی یک مجموعه داده برچسب گذاری شده و استفاده از مدل های یادگیری ماشین به خودکار سازی فرایند برچسب گذاری کمک کنیم.

3- روش انجام پروژه (روش، نمودار بلوکی اجزای سامانه‌ی مورد نظر پروژه، ورودی‌ها و خروجی‌ها):
برای انجام این پروژه در ابتدا لازم است که شناختی از پایگاه دانش CAPEC، پارامترهای خروجی SPLUNK، و مبانی حملات امنیتی فراهم شود. سپس برای انجام گام اول پروژه از روش‌های تحلیل داده‌های کیفی (Qualitative analysis) مانند inductive/deductive coding برای برچسب‌گذاری داده‌ها استفاده می‌شود. برای گام دوم الگوریتم های یادگیری ماشین نظارت شده برای ایجاد یک مدل برای پیش‌بینی برچسب‌های الگوهای حمله پیاده‌سازی می شود. لازم به ذکر است از آنجا که پایگاه دانش CAPEC شامل بیش از ۴۰۰ الگوی حمله می‌باشد برای اطمینان از دسترسی به داده کافی برای انجام پروژه؛ زیر مجموعه‌ای از الگوهای حملات که مرتبط با حملات APT است انتخاب می‌شود.

۴- روش ارزیابی:
برای اطمینان از اعتبار برچسب‌های اختصاص داده‌شده در فرایند برچسب گذاری از یک فرد خبره حوزه امنیت کمک گرفته می شود. برچسب گزاری در چند مرحله و با محاسبه شاخص reliability (IRR inter rater) در هر مرحله و با هدف دستیابی به شاخص با مقدار بالا انجام می‌شود.
مدل یادگیری ماشین بر روی مجموعه داده ایجاد شده در گام اول پروژه تست و ارزیابی می شود.

:

۵-مراجع و منابع :

[1] Scarabeo, Nicandro, Benjamin CM Fung, and Rashid H. Khokhar. "Mining known attack patterns from security-related events." ***PeerJ Computer Science*** **1 (2015): e25.**

[2] Navarro, Julio, et al. "Huma: A multi-layer framework for threat analysis in a heterogeneous log environment." ***Foundations and Practice of Security: 10th International Symposium, FPS 2017, Nancy, France, October 23-25, 2017, Revised Selected Papers 10*****. Springer International Publishing, 2018.**

https://journals.sagepub.com/doi/full/10.1177/15501329221084882
https://ieeexplore.ieee.org/abstract/document/9685977
https://link.springer.com/chapter/10.1007/978-1-4842-6276-4_1

۶- پیوست ها:

https://capec.mitre.org
https://www.splunk.com