EE / CprE / SE 492 Bi-Weekly Report

Project title: Knowledge Graphs for Cybersecurity Reasoning

03/04/23 – 03/24/23

Group number: sdmay23-01

Client & Advisor: Benjamin Blakely

Team Members:

Brandon Richards - *Frontend Development Lead*

Micah Gwin - *Python/ML Development*

Alice Cheatum - *Programmer*

Nicklas Cahill - *Tester/Programmer*

Michael Watkins - *Python/ML Development*

Carter Kitelinger - *Client Interaction*

Summary

Frontend S3 bucket and CloudFront was added to Infrastructure as Code, as well as build steps in the pipeline. Constraints were added to Neo4J to ensure there are no duplicate nodes. The pipeline was improved by caching build steps if dependencies haven't changed. Changes were made to Lambda permissions to allow for communication with AWS S3. Articles were annotated and the process of firing a background job to fetch CVE metadata has been started. The parser now downloads the latest model from S3 and uses DynamoDB instead of MongoDB.

Past Two Week Accomplishments

- Frontend Infrastructure – Brandon
    - Add S3 and CloudFront resources to Terraform Infrastructure-as-code
    - Add frontend build step to repo CI/CD pipeline
- Neo4j Constraints – Brandon
    - Write script to add UNIQUE constraints to Neo4j database
    - Add try/catch to parser when inserting entities to not crash when failing to add duplicate entity
- GitHub Actions Pipeline improvements – Micah
    - Modified the pipeline to plan the Terraform configuration in a pull request and then add a comment to the PR detailing the steps and plan output. Then upon merge to develop the plan is applied sending the resources to AWS (before both were done on merge only).

    - Researched how to improve pipeline speed to conserve limited build minutes, found some solutions to implement next sprint such as caching jobs or only running the step when requirements/dependencies change.
- AWS IAM Updates – Micah
    - Modified parser and scraper Lambda policy actions in Terraform to allow them access to all actions in DynamoDB and S3.
    - Created administrator IAM users for team members to access the account.
- CVE data background job – Alice
    - Begin integrating the CVE data retrieval script into the parser
- Annotate articles for training – Alice
- Parser improvements – Michael
    - Implemented parser S3 download
    - Updated parser to use DynamoDB instead of Mongo
    - Experimenting with pipelining the parser for efficiency

Pending Issues

- After adding more dependencies to the parser Lambda, the pipeline step that packages the Lambda layer has been taking 8+ minutes to build. Given that GitHub actions only gives us 2000 build minutes per month, we need to find a way to reduce the time. However, a few potential fixes have been researched by the team and are pending implementation next sprint.

Individual Contributions

Name	Hours past two weeks	Hours cumulative
Brandon Richards	10.5	78.5
Micah Gwin	10	55
Alice Cheatum	5	47.83
Nicklas Cahill	1.5	33.33
Michael Watkins	6	31
Carter Kitelinger	1.5	40.2

Summary of weekly advisor meeting

This week's advisor meeting was skipped due to miniscule progress being made over spring break.

Plans for Coming Two Weeks

- Infrastructure Improvements – Micah, Brandon
    - Add EC2 instance to Terraform (with security groups) and setup Neo4J server.

    - Turn monolithic Terraform stack into an environment that contains develop/testing resources and production resources. Modify pipeline to deploy with the correct Terraform variables and environment based on the branch being merged into.
    - Modify Lambda scraper to run locally or in Lambda environment (make dynamic if possible)
    - Attempt to run parser code as a module in Lambda
- Additional Pipeline Improvements (Fix Pending Issue) - Micah
    - Modify the pipeline to convert from one monolithic job into a small job for each step. Then cache the steps in the pipeline that take a long time to build to save minutes. In addition (or alternatively) modify the pipeline to only run the time-consuming steps if the dependencies file changes, otherwise skip them.
- Annotated Articles and added new relation – Carter
- Grab S3 Bucket containing NER and RE Model – Nicklas
    - Modified parser to only get the new models if they don't exist or to update the current models to newer ones
- Continue working on CVE data background job - Alice
    - Test implementation
    - Ensure CVE nodes have data when inserted into knowledge graph
- Annotate articles for training data – Alice