Creating a Knowledge Graph of Cybersecurity Resources

Codebase available at: github.com/STEELISI/CKIDS

Jelena Mirkovic
Daniel Garijo
Information Sciences Institute
Department of Computer Science
University of Southern California

Alex Zihuan Ran
Hardik Mahipal Surana
Tianxin Zhou
Department of Computer Science
University of Southern California

Project Motivation

- It is difficult to search for cybersecurity related papers, code and dataset in general-purpose repositories
- Cybersecurity resources are often disconnected. For example, it is difficult to retrieve datasets and software used by different publications

Goal, Impact, Challenges

Goal:
To **improve the accessibility and reuse of cybersecurity** publications, methodologies, tools and datasets using a Knowledge Graph.

Expected Impact:
1. Increase available expertise (rather than just lab resources)
2. Enable vertical development, improving quality, maximizing efficiency, reducing time and effort to adopt existing methods, datasets and software
3. Improve data and knowledge sharing across organizations

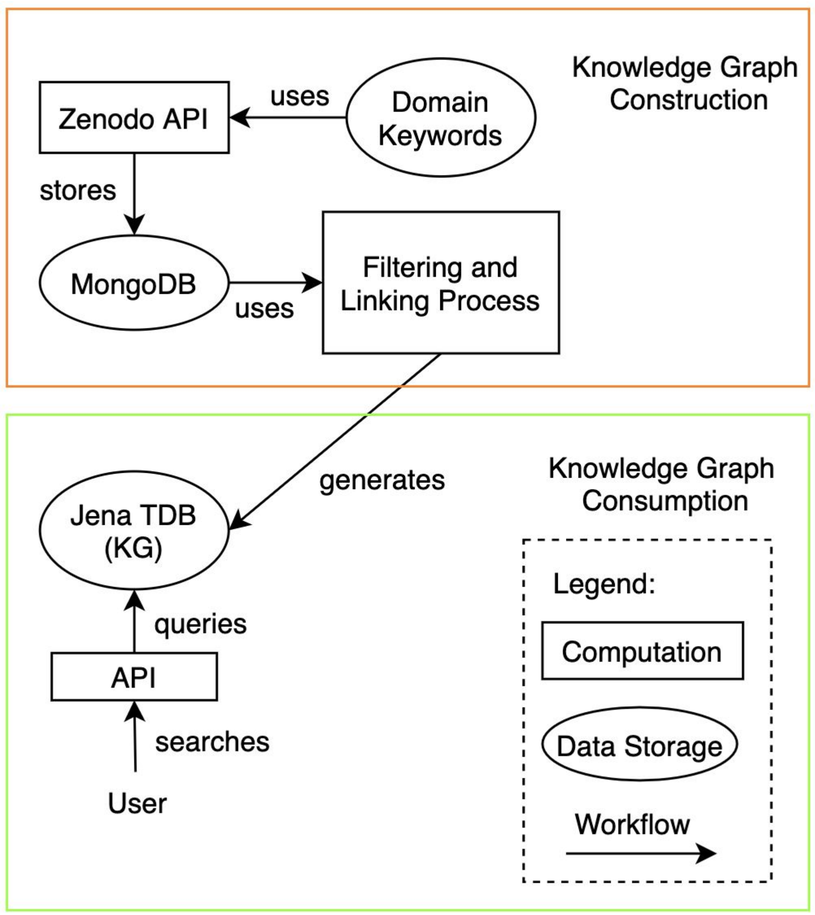
Challenges:
Establishing relationships among artifacts based on keywords or dependencies

Data/Resources Available

- Cybersecurity vocabulary from **NICCS** (National Initiative for Cybersecurity Careers and Studies)
- **Zenodo**: general-purpose, open access repository where researchers upload artifacts

https://zenodo.org/

Approach



Tech Stack:
Python, Apache Jena TDB, AWS EC2, MongoDB

Amazon EC2

Current Deliverables and Next Steps

1. A comprehensive keyword list relevant to cybersecurity research (Complete)

```
(DDoS)
(IRC)
0-RTT mechanism
Access control
address space layout randomization
admission control
code-reuse attacks
adversarial perturbations
AES
```

2. A scraping script for collecting artifact data using Zenodo API (Complete)

Execution Examples:

```
cd <PROJECT_ROOT>

# query with a single keyword, store results in JSON file
python3 src/scrape.py -k cybersecurity -s 100

# query with multiple keywords, store results in JSON file
python3 src/scrape.py -k cybersecurity vulnerability -s 100

# store results in MongoDB in the AWS EC2 instance
python3 src/scrape.py -k cybersecurity -s 100 -db
```

3. Cloud database on AWS EC2 that stores collected data (Complete)

id: Zenodo URL/DOI.

name: Title of the artifact

type: whether it's a dataset, code or paper.

abstract: abstract of the artifact

author: Authors

keywords: keywords used for describing the artifact.

relatedTo: Relationship used to link relevant resources together (datasets and papers, etc).

4. To filter collected data based on relevance to cybersecurity (In Progress)

5. To develop an understanding of knowledge graphs, represent relationships and build the graph (In Progress)

6. To build a Search API which queries the knowledge graph (To Do)

7. Testing (To Do)