SDA TAP Lab – Ground Infrastructure Security

Andrew Liu, Robert Drone, Sanjana Gadaginmath, Derek Marraudino, Elias Pike, Louis Vargas, Dylan Vigil,
Emily Johnson, Kyaw Paw, Terri Akse, Stephanie Villanueva, Aria Barbour, Dawson Roach

PROJECT SUMMARY

Problem Statement:
- Finding, identifying, and classifying vulnerabilities and threats to US Space Force Ground Infrastructure, with a concentration on optical telescope systems.

Research Outline:
- Identify supply chains and categorize known vulnerabilities
- Utilize web scraping for collecting threat data
- Classifying threat data using generalized threat parameters
- Outputting a final risk assessment for selected supply chain

Sub-Teams:
- Threat Matrix
- Web Scraping
- Embeddings

WEB SCRAPING TOOL

The web scraping tool takes an input of categories and keyword constraints and generates a report for the end user. The tool uses guidance from the threat matrix team via JSON and scrapes the web for data. The Tool performs ETL(Extract, Transform and Load the data) and hand over a cleaned JSON to the through the embeddings backend API to be used for training and testing, and later on to generate embeddings.

Functionality:

1.) Multi-Source Threat Intelligence Aggregation:
- Tool API Sources: Custom Google Search, Shodan API, CISA KEV, AlienVaultOTX, IBM X-Force Exchange, NVD API

2.) Custom Risk Assessment and Categorization
- Gathers CVSS scores and Defines key terms
- Maps severity and impact levels

3.) Structured JSON Output
- Output file is Stored in MongoDB
- Vulnerabilities

```
# Determine category
category = determine_otx_category(description, tags)

vulnerabilities.append({
    "name": name,
    "description": description,
    "category": category,
    "risk_assessment": risk_assessment,
    "affected_components": extract_affected_components(description),
    "patch_available": determine_patch_availability(description),
    "related_assets": extract_related_assets(description),
    "indicators": indicator_list,
    "tags": tags,
    "sources": [f"https://otx.alienvault.com/pulse/{pulse_id}"],
    "timestamp": datetime.utcnow().isoformat() + "Z",
    "keyword": search_term
})
```

The tool also gathers metadata such as publication dates and affected software versions. This structured information is then stored in a standardized format for easy access and analysis. Automating this process reduces the time required for manual research and enhances situational awareness for threat hunting teams.

EMBEDDINGS

The Embeddings Sub-team focused on designing and implementing an efficient **semantic search system** to retrieve appropriate cybersecurity-related events to permeate into the final threat matrix. The key **project phases** were:

Model Development: Fine-Tuning SBERT Mongo:
- Sentence-BERT (SBERT) ***all-MiniLM-L6-v2*** model
- training model on labeled cybersecurity incidents, allowing it to understand relationships between different attack types and make connections on vulnerability, threat, or asset descriptions
  - goal was to ensure model captures domain-specific patterns and return most contextually relevant

Vector Database & Indexing: Storing and Retrieving Embeddings:
- ***MongoDB*** vector database using ***MAtlas Vector Search*** and ***HNSW algorithm*** for optimized retrieval
  - Supports ANN and ENN search

Performance Evaluation: Testing & Refining Retrieval Quality:
- KPI for training and testing on different data formats and indexing methods
  - Memory usage, time, latency.

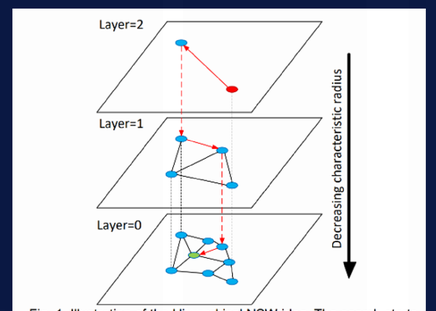

Fig. 1. Illustration of the Hierarchical NSW idea. The search starts from an element from the top layer (shown red). Red arrows show direction of the greedy algorithm from the entry point to the query (shown green).

THREAT MATRIX ANALYSIS

Threat Analysis allows us to prioritize common vulnerabilities and exploits (CVEs) relevant to critical infrastructure surrounding optical telescope technology.

Creating a Threat Matrix provides us with a visual aid from which to discern prioritized controls for high-priority assets.

Our team interacts with the other teams by:
1.) Define output matrix parameters that benefits all teams.
2.) Defines categories, like common vulnerabilities and exploits (CVEs), as labels that get passed to 'Embeddings.'
3.) Final threat matrix analysis for report output and action items.

TVA Chart		Asset 1	Asset 2	Asset 3	Asset 4	Asset 5	Asset 6	Asset 7	Asset 8
		Cloud Environment	External Perimeter of Site	Employees (Skilled Labor Force)	Maintenance Schedule (Outsouced Work)	Supply Chain + Vendors	Critical Softwares	Non-Critical Softwares	Oracle
Threat 1	Component Interference								
Threat 2	Cyberattacks								
Threat 3	Natural Disaster								
Threat 4	Geo-Political Posture								
Threat 5	Insider Threats								
Threat 6	Loss of Vendors								
Threat 7	Social Engineering								
Threat 8	Security Posture								
Priority of Controls (Unranked)		Control 1 Education, Training, and Fake attempts	Control 2 Badges, RFID Key cards, physical keys	Control 3 Bitlocker	Control 4 Password Updates/Requirments	Control 5 Access Audits	Control 6 Firewall	Control 7 Acceptable Use Policy/Privacy Policy	Control 8 Backups and Redundancies
		Control 9 Audit Log	Control 10 Cold Site Set Up and Plan	Control 11 Disaster Recovery Plan	Control 12 Backup Suppliers	Control 13 Vetting System for Employees	Control 14 Metal Detectors		

CONCLUSION AND NEXT STEPS

- Together, our team has developed a strong foundational application that will permeate its influence throughout the entities that interact with the SDA TAP Lab.
- This technology is scalable and able to fit the growing needs of the United States Space Force (USSF) relating to the evolving global cyber-threat climate, specifically targeting the widespread data available on the internet.

Next Steps:
1.) Validate data for meaningful output.
2.) Improve front-end of application for user-interaction.
3.) Custom application execution function for ease-of-use.
4.) Verify actual output from test-run for accuracy adjustments.
5.) Integration into existing and/or new technologies for validation.

REFERENCES

"Sentencetranbsformers Documentation." SentenceTransformers Documentation - Sentence Transformers Documentation, sbert.net/. Accessed 11 Apr. 2025.
"Atlas Vector Search Overview." Atlas Vector Search Overview - Atlas - MongoDB Docs, www.mongodb.com/docs/atlas/atlas-vector-search/vector-search-overview/. Accessed 11 Apr. 2025. CVE, cve.mitre.org/. Accessed 11 Apr. 2025.

ACKNOWLEDGMENTS

We'd like to take the chance to show our appreciation to all of our corporate sponsors and mentors who have helped guide us throughout this educational experience! The following individuals had a hand in making this project come to life:
➢ Megan Dison, Melissa Hills, and Ryan Morrell
Thank you so very much for your help and time! We grew as professionals & students thanks to you and your valuable input!