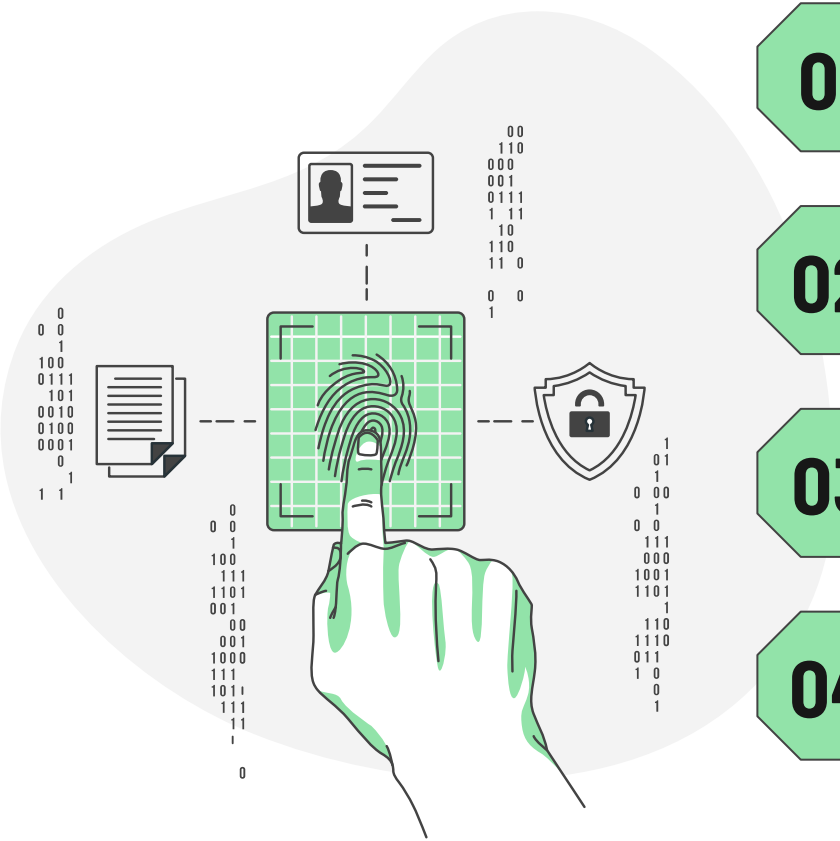
Graph Database for Threat Actors

A Solution for BAE Systems Problem 1

Nhung Nguyen | Hanoi, April 2023

AGENDA



01 **PROBLEM** STATEMENT
Short Description of BAE Systems Problem 1

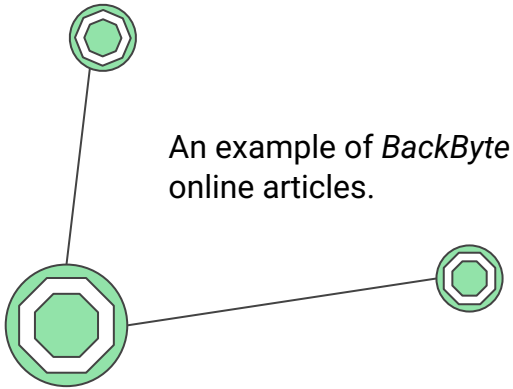
02 **TECHNICAL** SOLUTION
Brief of the Technical Solution

03 **SOLUTION** DEMOSTRATION
How would the Solution Solve the Problem?

04 **FURTHER** WORKS
Further Improvements

PROBLEM STATEMENT

- Research articles on cyber threat actors coming from different sources
- Single threat actor is described under different names
- Commonalities missed
- Relationships between threat actors/ targets aren't obvious
- Data inconsistency



An example of *BackByte Ransomware* written in online articles.

CROWDSTRIKE | BLOG

Featured

All too often, I hear from executives of large and small organizations that aside from targeted attacks, impact from ransomware is their number one concern. That impact was starkly highlighted earlier this month when Los Angeles-based Hollywood Presbyterian Medical Center hospital was hit by ransomware which encrypted data in the electronic medical records system

two weeks.

microsoft.com/en-us/security/blog/2022/06/13/the-many-lives-of-blackcat-ransomware/

The BlackCat ransomware, also known as ALPHV, is a prevalent threat and a prime example of the growing ransomware as a service (RaaS) gig economy. It's noteworthy due to its unconventional programming language (Rust), multiple target devices and possible entry points, and affiliation with prolific threat activity groups. While BlackCat's arrival and execution vary based on the actors deploying it, the outcome is the same—target data is encrypted, exfiltrated, and used for "double extortion," where attackers threaten to release the stolen data to the public if the ransom isn't paid.

ghly

ystems.

prevalent

nd individuals

First observed in N

ransomware familie

its payload, this ran

Exbyte: BlackByte Ransomware Attackers Deploy New Exfiltration Tool

Exbyte is the latest tool developed by ransomware attackers to expedite data theft from victims.

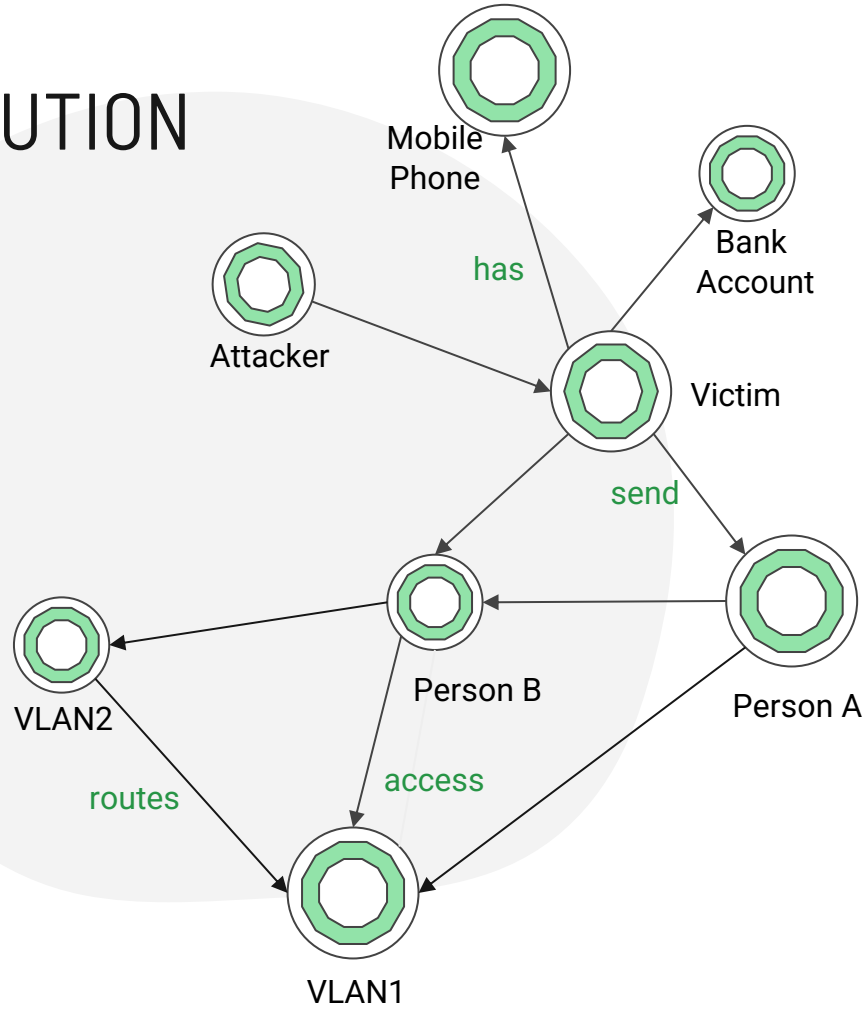
Symantec's Threat Hunter Team has discovered that at least one affiliate of the BlackByte ransomware (Ransom.Blackbyte) operation has begun using a custom data exfiltration tool during their attacks. The malware (Infostealer.Exbyte) is designed to expedite the theft of data from the victim's network and upload it to an external server.

BlackByte is a ransomware-as-a-service operation that is run by a cyber-crime group Symantec calls Hecamede. The group sprang to public attention in February 2022 when the U.S. Federal

VALUE PROPOSITION/ SOLUTION

Build a **graph database** named "**threat-actor-database**" that:

- Represents threat actors/ targets/ organizations as nodes
- Represents relationships between them as edges
- Automatically generate/ update the graph that can support further research and maintain



SOLUTION PIPELINE

4 main URLs

Extract articles' URLs

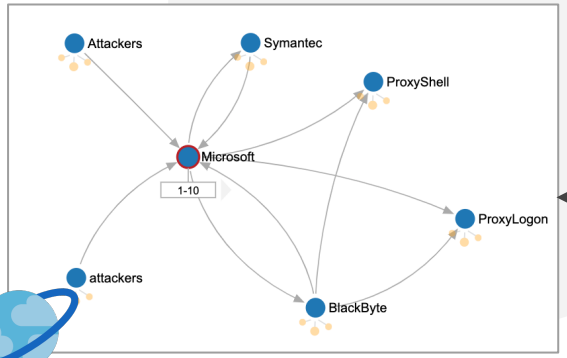
Web pages

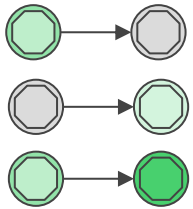
BeautifulSoup

Plain Text

Language service

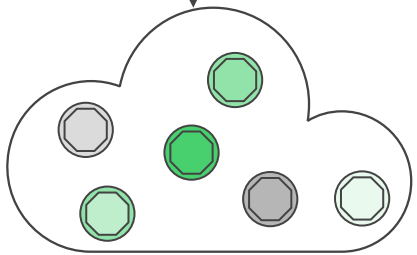
Knowledge Graph Database





Relation Extraction

Rules-based

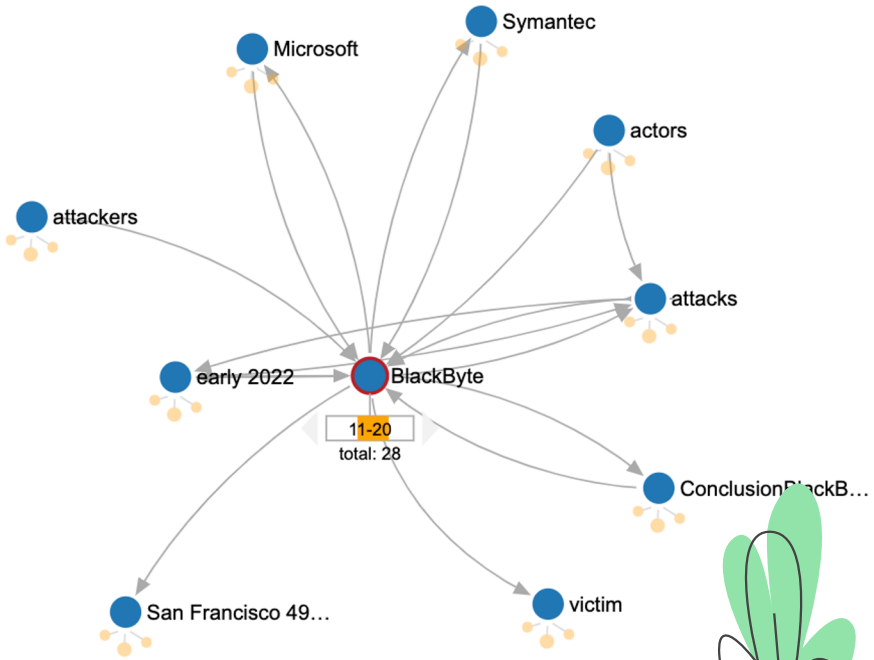


Extracted Entities

SOLUTION DEMONSTRATION

MY CONTRIBUTION

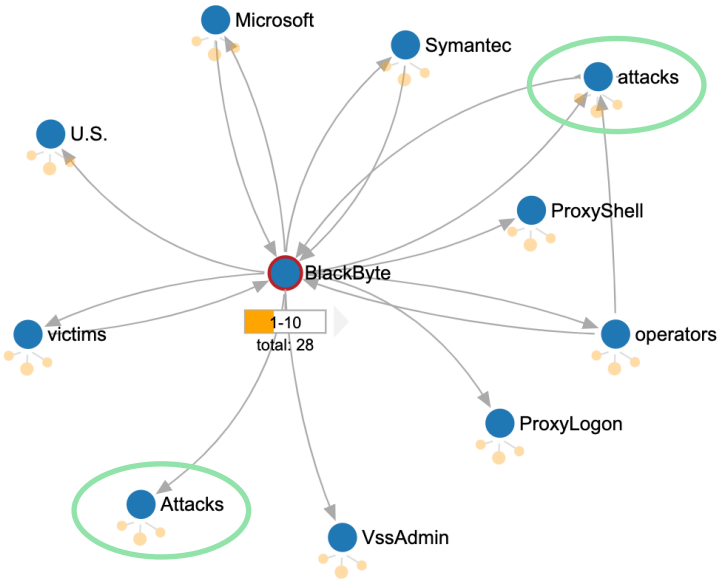
- Whole pipeline implemented
- Leveraged Azure services to speed up implementing
- Defined a threshold control to map extracted entities from Azure language service with relations

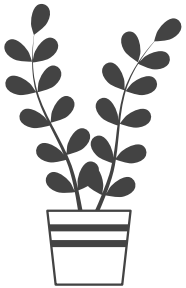


Vertices: People, Places, Events, DateTime, Organizations, Products, PersonType
Edges: Targets, affected by, are protected from, blocks…

FUTURE WORKS

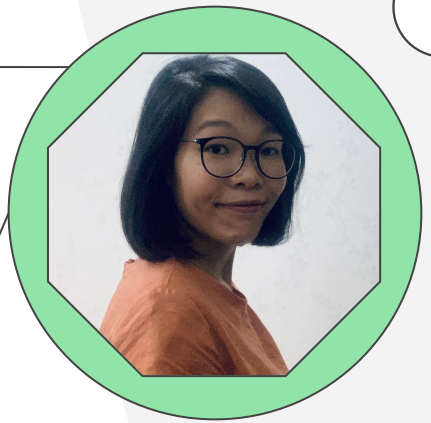
- Increase the number of relation extractions by using some relation extraction deep learning models
- Add post-processing to refine entities (e.g., unify entities)
- Find a mechanism to avoid being detected as Autobot by the webpage
- Add a crontab or schedule to run the script daily





```
2023-04-28 09:44:52,761: INFO: Number of vertices to insert into the graph: 361
INFO:BAE1_data_extraction:Number of vertices to insert into the graph: 361
2023-04-28 09:44:54,178: INFO: Number of edges to insert into the graph: 76
INFO:BAE1_data_extraction:Number of edges to insert into the graph: 76
2023-04-28 09:44:54,178: INFO: Count number of vertices in the graph
INFO:BAE1_data_extraction:Count number of vertices in the graph
        Count of vertices: [299]
```

ABOUT ME



Nhung Nguyen
Data Scientist/ Project Manager
softpast@gmail.com

- Passionate about AI and would love to apply AI to real-world problems
- Certified Azure Engineer Associate, Certified ScrumMaster
- Graduated from Keio University, Japan
- Experience in designing and implementing enterprise systems