UMBC

A Dynamic Pipeline for Cybersecurity Information Extraction and Understanding

Priyanka Ranade, Aritran Piplai
University of Maryland Baltimore County

Acknowledgements

- Dr. Ahmad Ridley, NSA Mentor
- Dr. Anupam Joshi, Professor
- Dr. Tim Finin, Professor

Thank you to the NSA for their support!

Agenda

Building a Cybersecurity Knowledge Graph (CKG):

Approach for creating a CKG and using it to understand malware behavior

Understanding Security Risks for CKGs:

Showcasing a data poisoning attack to infiltrate the CKG and metods to defend against integrity attacks

UMBC

Building a Cybersecurity Knowledge Graph

After Action Report about Malware

The Naikon group used mostly spear-phished documents for the attacks, with CVE-2012-0158 exploits that dropped the group's signature backdoor.

While many of these attacks were successful, at least one of the targets didn't seem to like being hit, and instead of opening the documents, decided on a very different course of action.

The empire strikes back

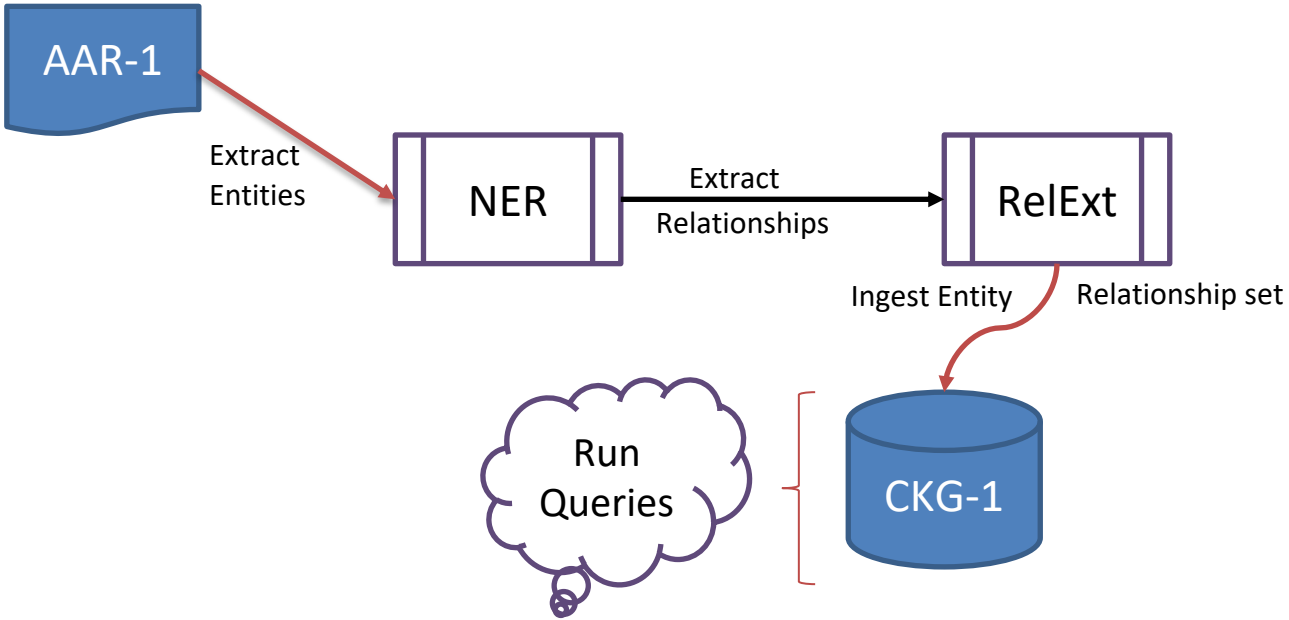
Here's a question - what should you do when you receiving a suspicious document from somebody you don't know, or know very little? Choose one:

- Open the document
- Don't open the document
- Open the document on a Mac (everybody knows Mac's don't get viruses)
- Open the document in a virtual machine with Linux

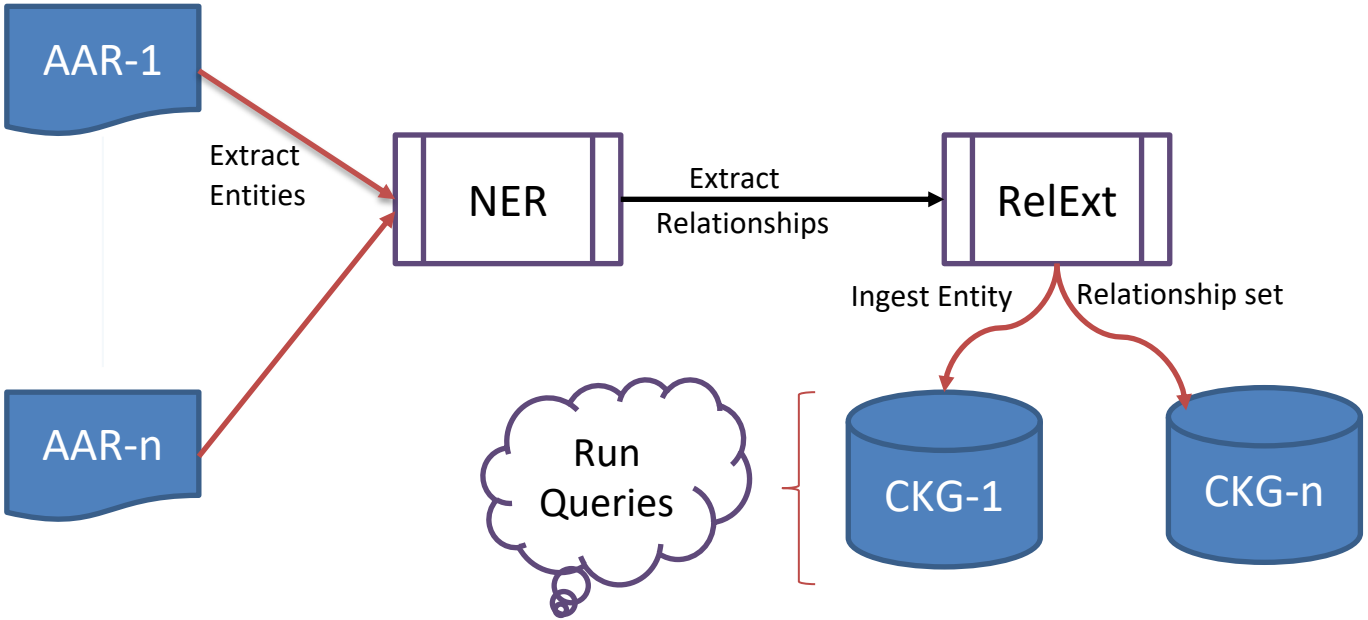
Based on our experience, most people would say 2, 3 or 4. Very few would open the document and even fewer would actually decide to test the attacker and verify its story.

But this is exactly what happened when one of the Naikon spear-phishing targets received a suspicious email. Instead of opening the document or choosing to open it on an exotic platform, they decided to check the story with the sender:

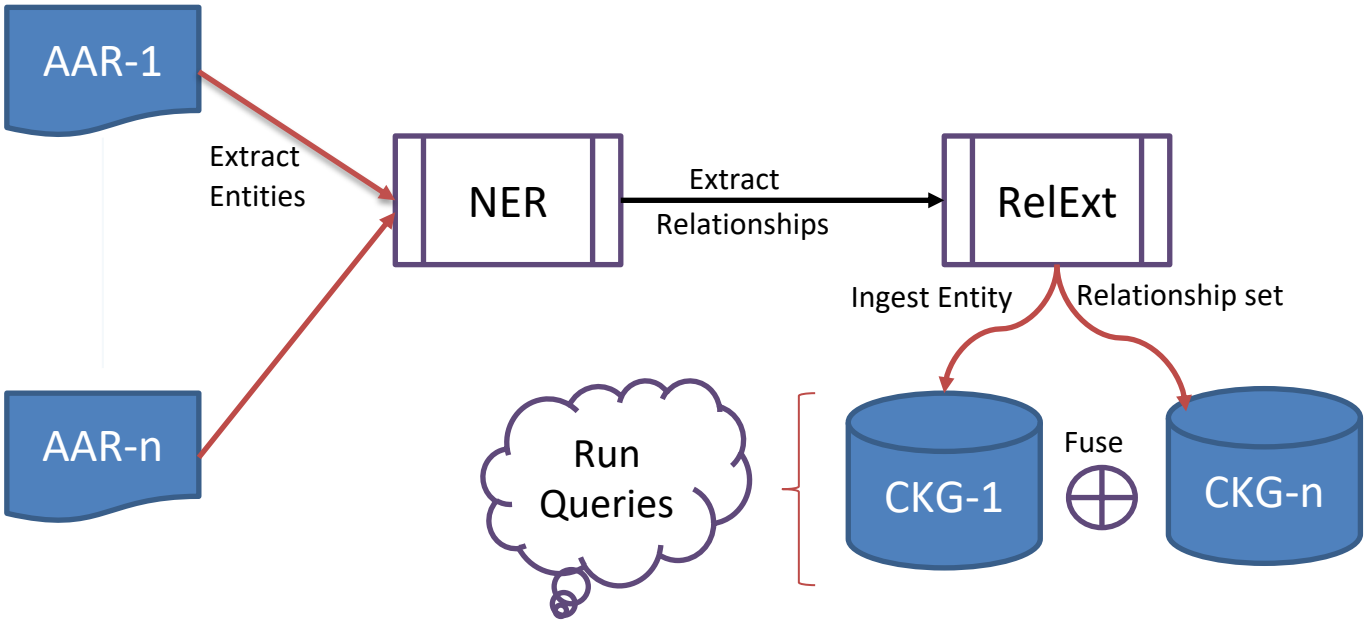
Architecture



Architecture



Architecture



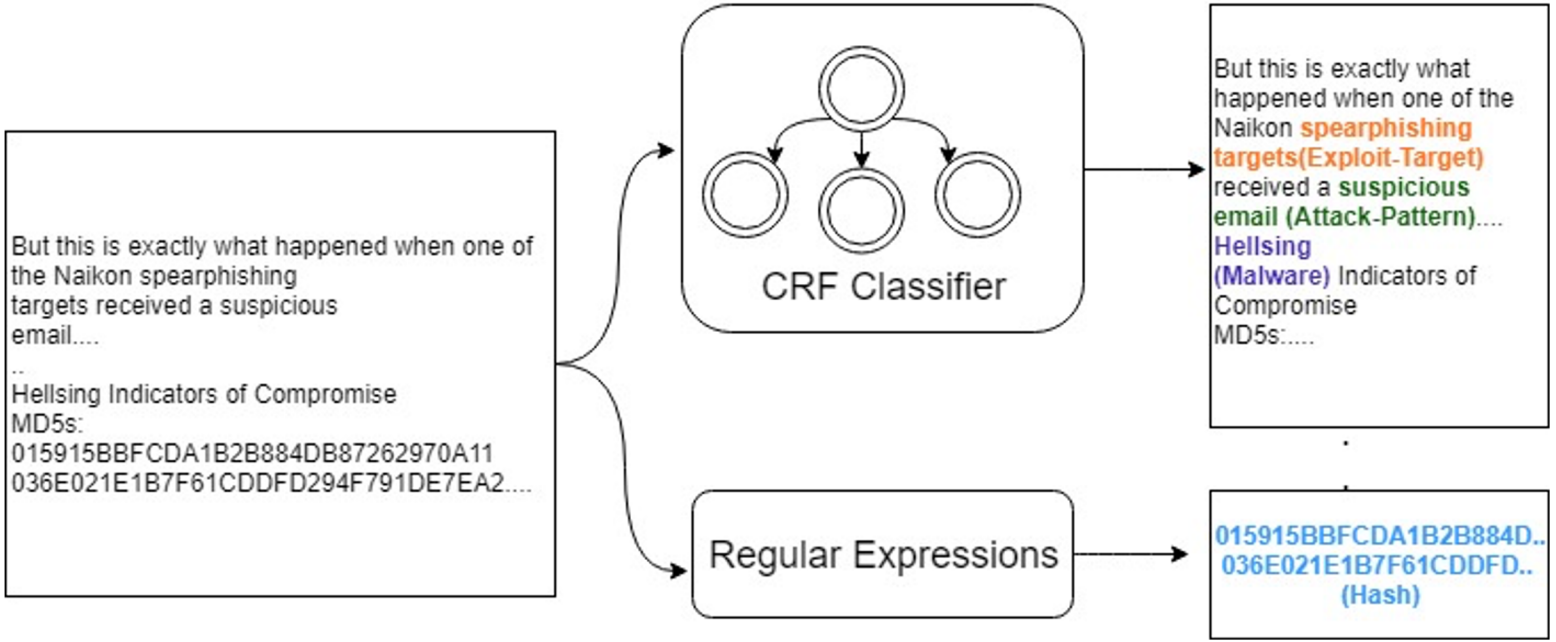
NER and Relationship Extractor

Entities and Relationships

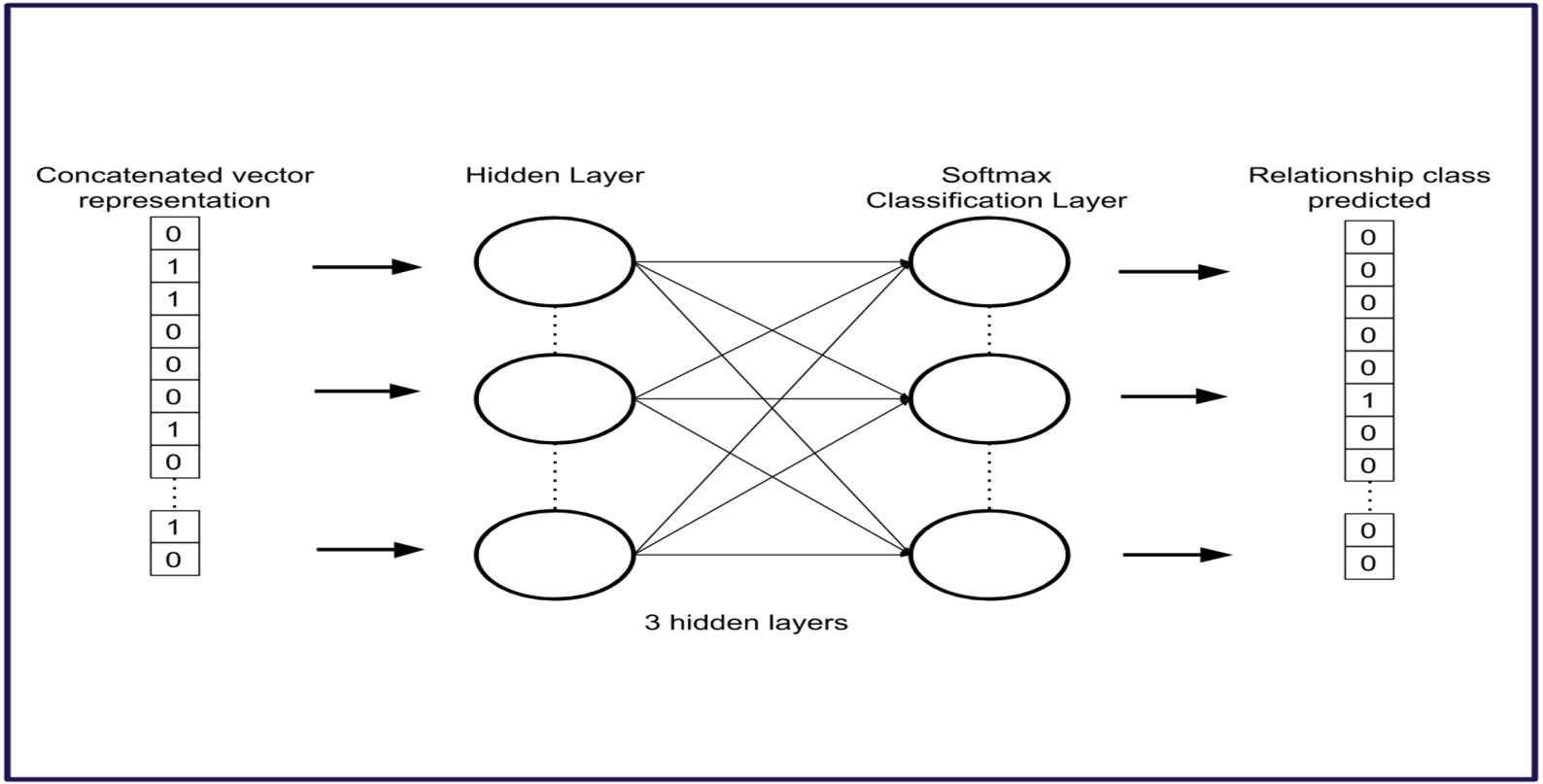
Software	Indicator
Attack-Pattern	IPAddresses
Course-of-Action	Malware
Exploit-Target	Campaign
Filename	Tool
Hashes	Vulnerability

hasProduct	targets
hasVulnerability	mitigates
uses	indicates

Example


But this is exactly what happened when one of the Naikon spearphishing targets received a suspicious email....
..
Hellsing Indicators of Compromise
MD5s:
015915BBFCDA1B2B884DB87262970A11
036E021E1B7F61CDDFD294F791DE7EA2....
CRF Classifier
But this is exactly what happened when one of the Naikon spearphishing targets(Exploit-Target) received a suspicious email (Attack-Pattern).... Hellsing (Malware) Indicators of Compromise MD5s:....
Regular Expressions
015915BBFCDA1B2B884D..
036E021E1B7F61CDDFD..
(Hash)

Relationship Extraction Neural Network



Relationship Extractor

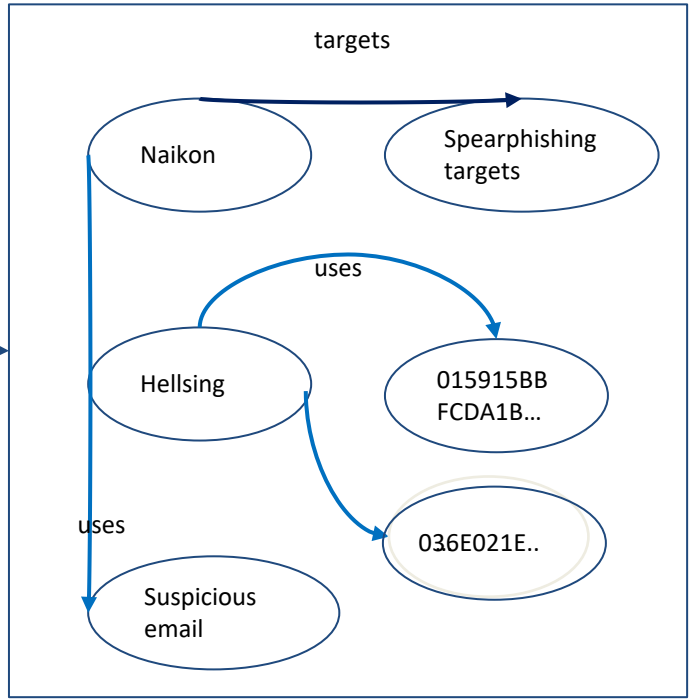
But this is what exactly happened when one of the Naikon spearphishing targets received a suspicious email

……..

Hellsing Indicators of Compromise

MD5s: 015915BBFCDA1B2B884DB87262970A11

036E021E1B7F61CDDFD294F791DE7EA2



Corpus

Prediction

Knowledge Graph Fusion

The Hellsing series chronicles the efforts of the mysterious and secret Hellsing Organization, as it combats vampires, ghouls, and other supernatural foes; which makes it perhaps an appropriate name for our group.

In addition to the Hellsing/msger malware, we've identified a second generation of Trojan samples which appear to be called "xweber" by the attackers:

"Xweber" seems to be the more recent Trojan, taking into account compilation timestamps. All the "msger" samples we have seen appear to have been compiled in 2012. The "Xweber" samples are from 2013 and from 2014, indicating that at some point during 2013 the "msger" malware project was renamed and/or integrated into "Xweber".

During our investigation we've observed the Hellsing APT using both the "Xweber" and "msger" backdoors in their attacks, as well as other tools named "xrat", "clare", "irene" and "xKat".

Hellsing Indicators of Compromise

MD5s:

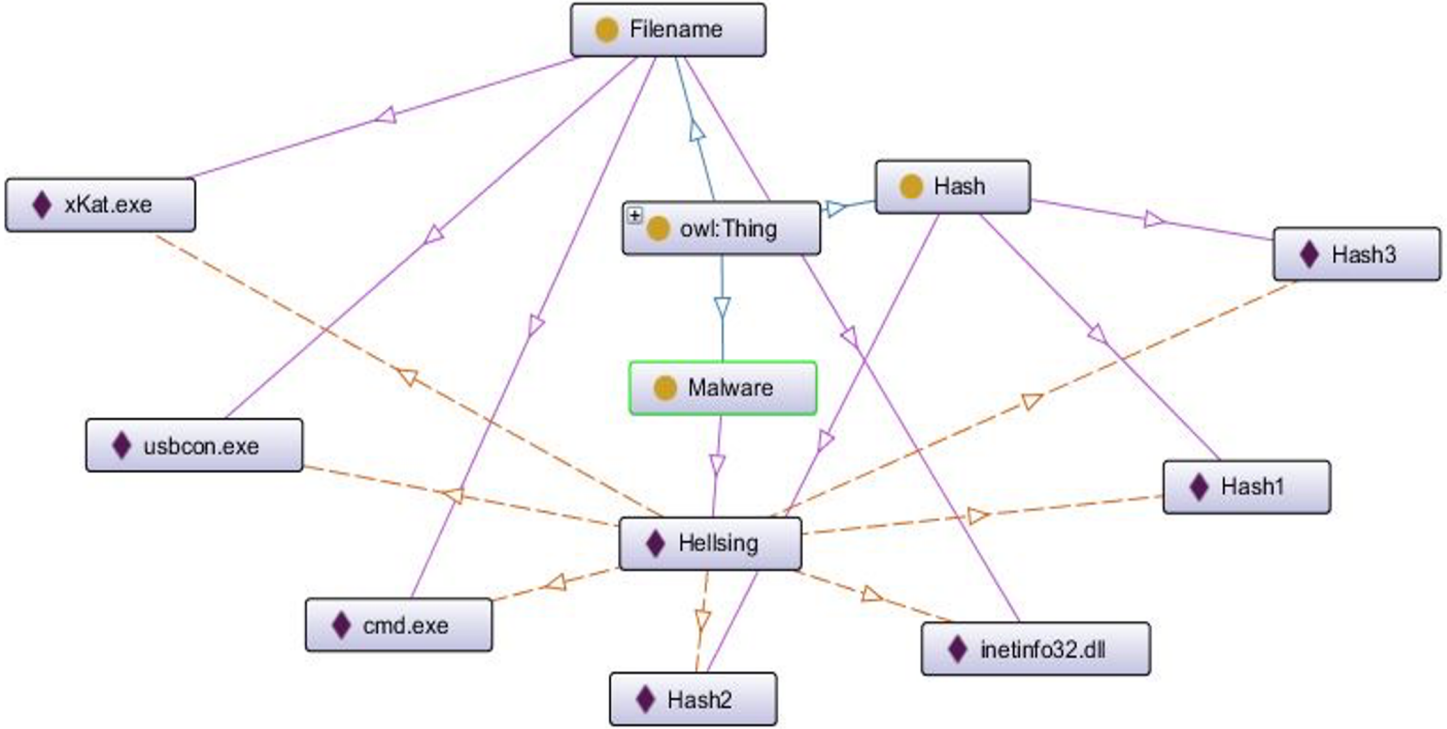
```
015915BBFCDA1B2B884DB87262970A11
036E021E1B7F61CDDFD294F791DE7EA2
04090aca47f5360b84f6a55033544863
055BC765A78DA9CC759D1BA7AC7AC05E
085FAAC21114C844529E11422EF684D1
0BA116AA1704A415812552A815FCD34B
0CBEFD8CD4B9A36C791D926F84F10B7B
0CC5918D426CD836C52207A8332296BC
0dfcbb858bd2d5fb1d33cd69dcd844ae
0F13DEAC7D2C1A971F98C9365B071DB9
0FFE80AF4461C68D6571BEDE9527CF74
13EF0DFE608440EE60449E4300AE9324
14309b52f5a3df8cb0eb5b6dae9ce4da
17EF094043761A917BA129280618C1D3
2682A1246199A18967C98CB32191230C
2CCE768DC3717E86C5D626ED7CE2E0B7
```

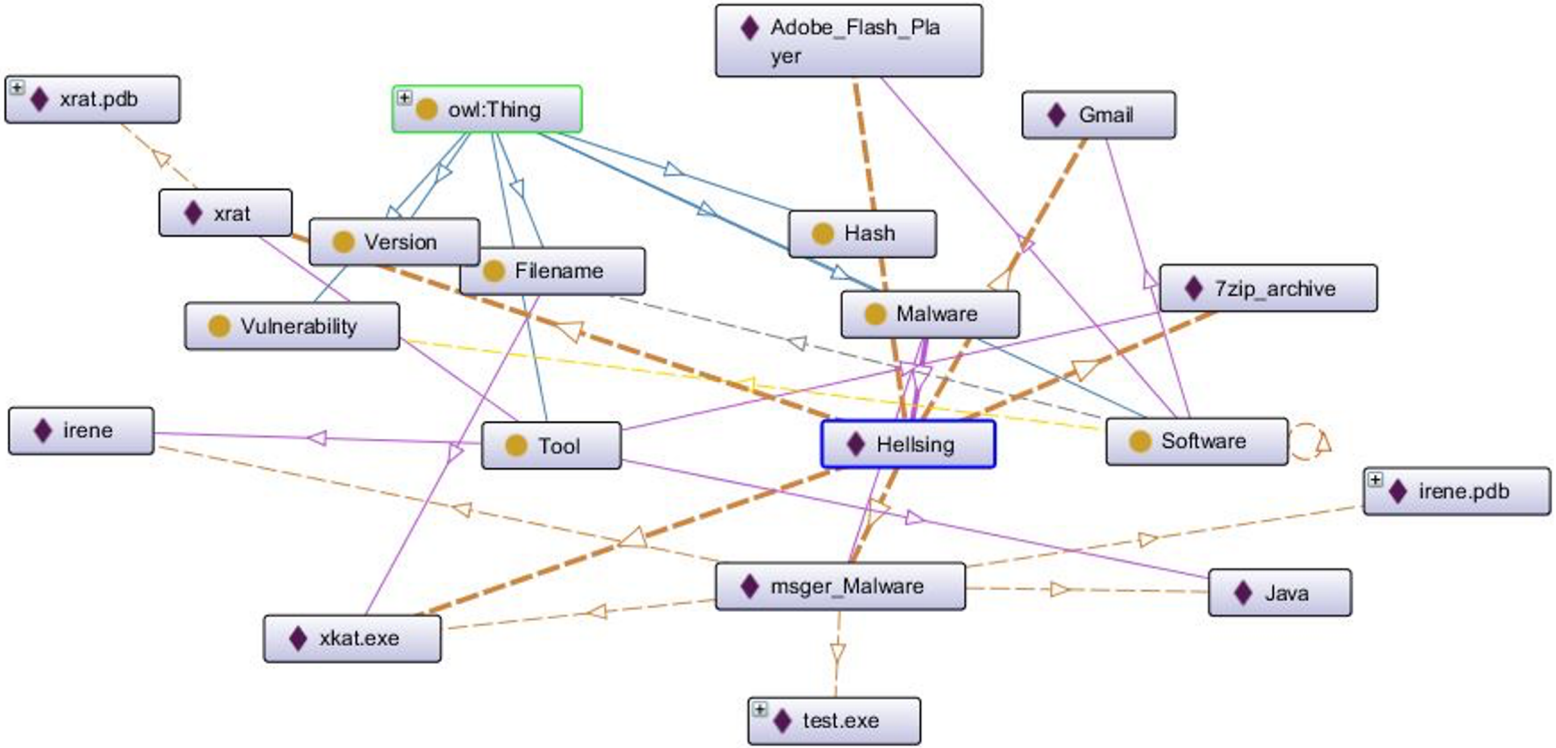
Domain registrations:

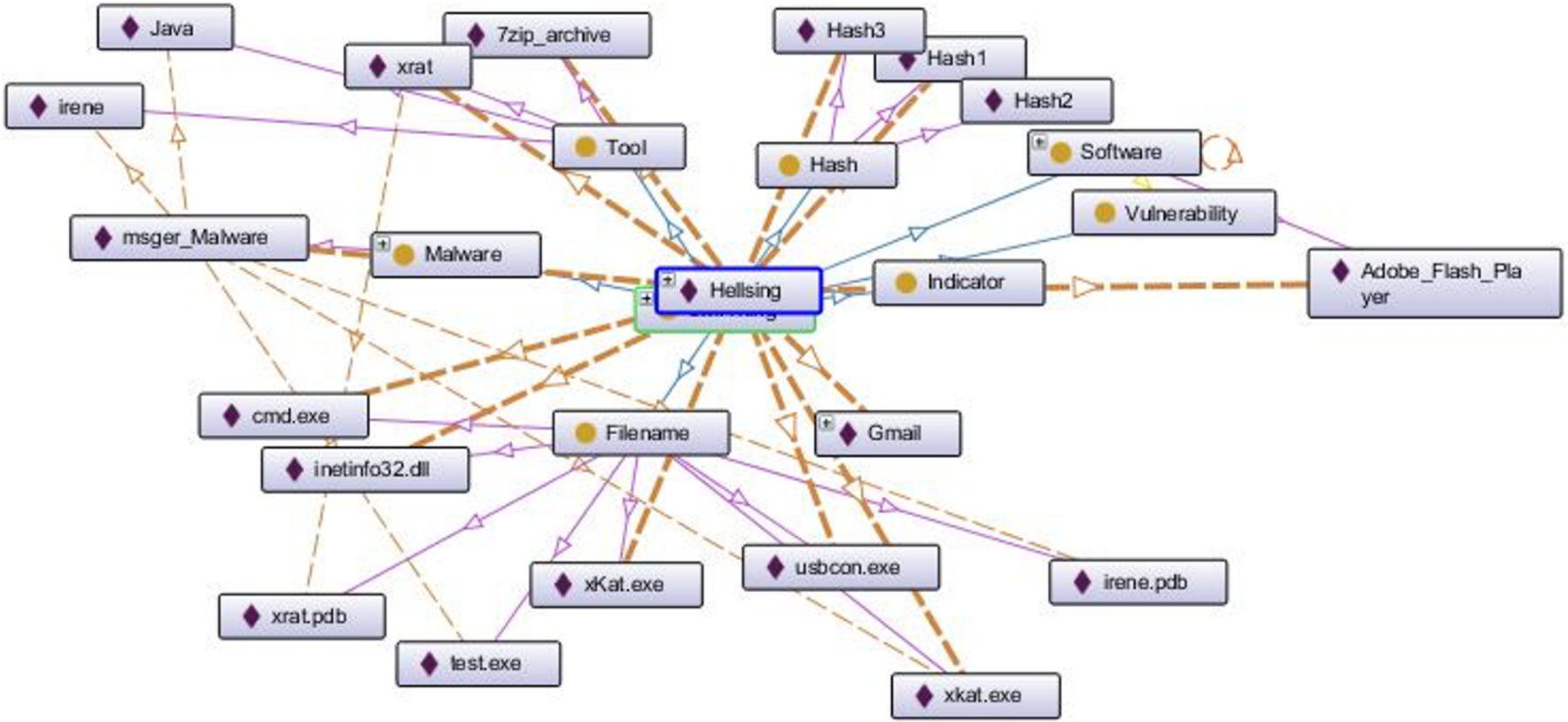
- huntingtomingalls[.]com - ssdfsddfs@qsdfsq.com
- philippinenewss[.]com - sambieber1990@yahoo.com
- philstarnotice[.]com - sambieber1990@yahoo.com

Filenames:

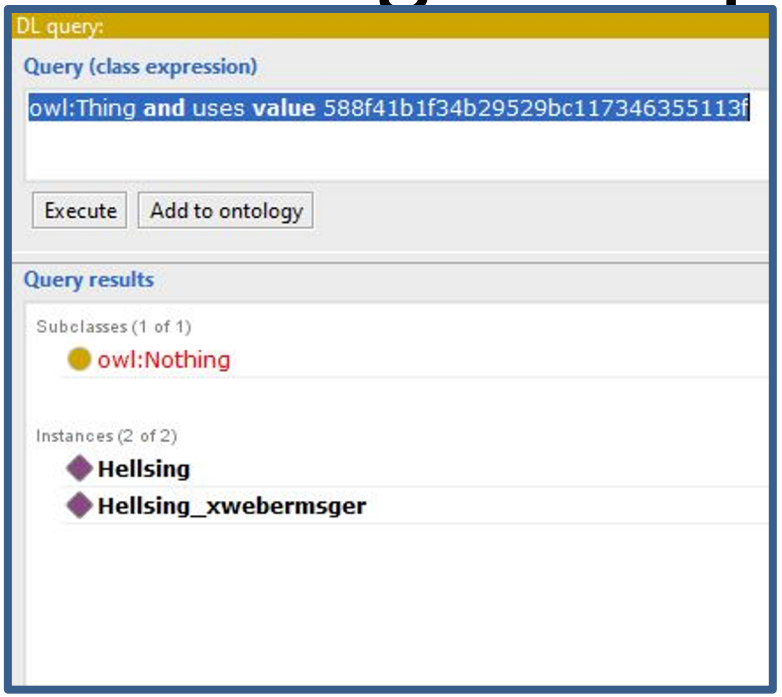
- %systemroot%\system32\irmon32.dll
- %systemroot%\system32\FastUserSwitchingCompatibilityex.dll
- %systemroot%\system32\inetinfo32.dll
- %systemroot%\system32\drivers\drivers\diskfilter.sys
- %systemroot%\system32\usbcon.exe
- %windir%\temp\xKat.exe
- %systemroot%\system32\drivers\drivers\usbmgr.sys
- %appdata%\Microsoft\MMC\mmc.exe
- %systemroot%\system32\lasex.dll
- %systemroot%\system32\Ipripex.dll


Filename
xKat.exe
owl:Thing
Hash
Hash3
Malware
usbcon.exe
Hash1
Hellsing
cmd.exe
inetinfo32.dll
Hash2

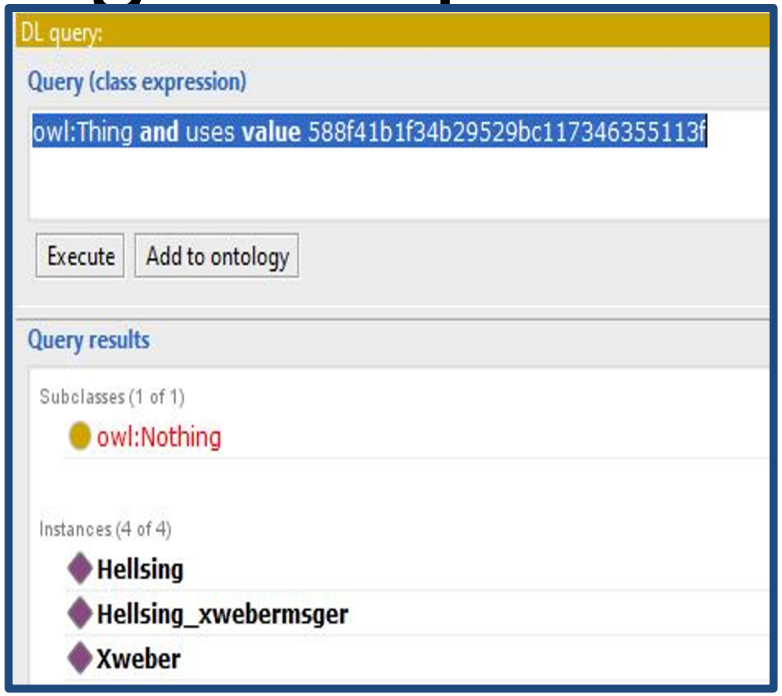

Adobe_Flash_Player
xrat.pdb
owl:Thing
Gmail
xrat
Version
Hash
Filename
7zip_archive
Vulnerability
Malware
irene
Tool
Hellsing
Software
irene.pdb
msger_Malware
Java
xkat.exe
test.exe


Java
7zip_archive
xrat
irene
Tool
Hash3
Hash1
Hash2
Hash
Software
Vulnerability
msger_Malware
Malware
Hellsing
Indicator
Adobe_Flash_Player
cmd.exe
Filename
Gmail
inetinfo32.dll
xKat.exe
usbcon.exe
irene.pdb
xrat.pdb
test.exe
xkat.exe

Reasoning examples



Reasoning examples after fusion



Data poisoning attack Risks and Defenses

Background and Motivation

Cyber Threat Intelligence and the Security Community

- Cyber Threat Intelligence (CTI)
  - Also referred to as *OSINT* (Open Source Intelligence Data)
- Information about cybersecurity vulnerabilities, exploits
- Valuable to the security community to stay informed about new vulnerabilities and exploits
- Used as training data for AI-based cyber defense systems

AI-Based Cyber Defense Systems

- CTI can be shared as text or as semi-structured data
- Text fields can use formats like Structured Threat Information Expression **(STIX)** and Malware Information Sharing Platform **(MISP)**
- Free text can be transformed to more structured formats and used as training data for machine learning systems aimed at detecting attack patterns.
- Used to populate cybersecurity knowledge graphs (CKG) for further reasoning.

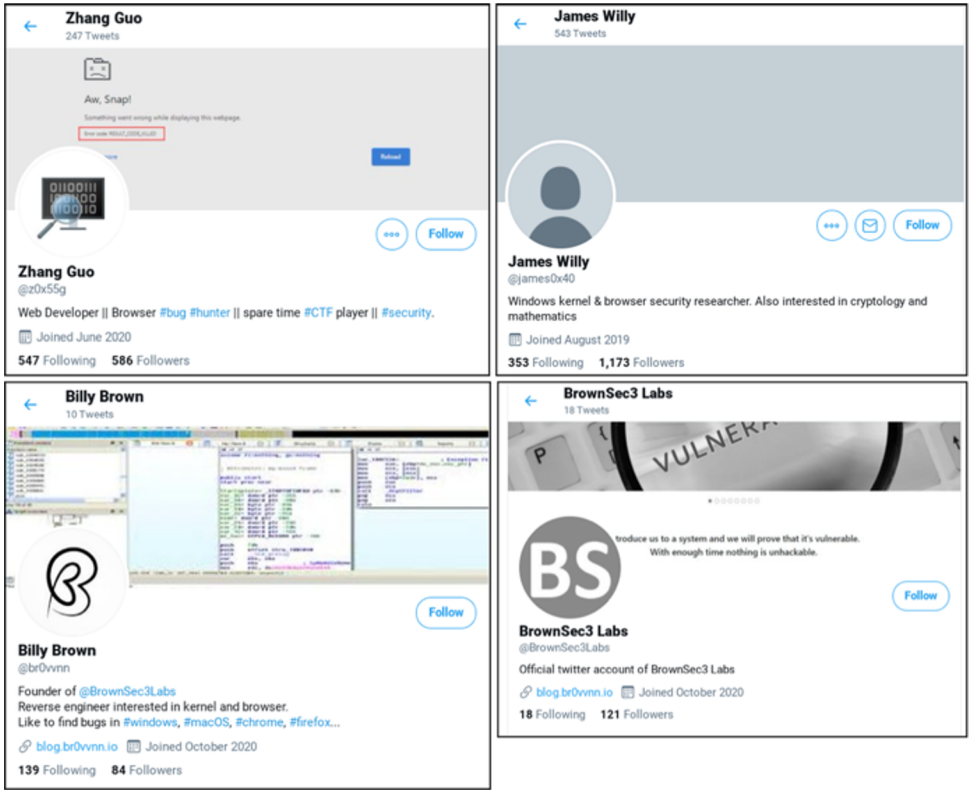
Related Work

- *Mittal et al.* created Cyber-All Intel and CyberTwitter which use CKG and other agents to aid cybersecurity analysts
- *Piplai et al.* use extracted malware after action report information as prior information in a reinforcement learning malware analysis environment
-

Cybersecurity Misinformation Risks

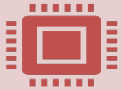
- The misinformation risk for the security community is the possible dissemination of false CTI by threat actors in an attempt to poison systems that ingest and use the information.
-

  January 2021- Google Threat Analysis Group discovered a nation-state backed group that faked security research accounts and blog posts



Google Threat Analysis Group

Contributions



Fine-tuned GPT-2 Model that generates fake CTI text

Poisoning pipeline for infiltrating a Cybersecurity Knowledge Graph

Evaluation and analysis of the fake and real CTI text

The Evolution of Natural Language Models

$$w_{x,y} = tf_{x,y} \times \log\left(\frac{N}{df_x}\right)$$

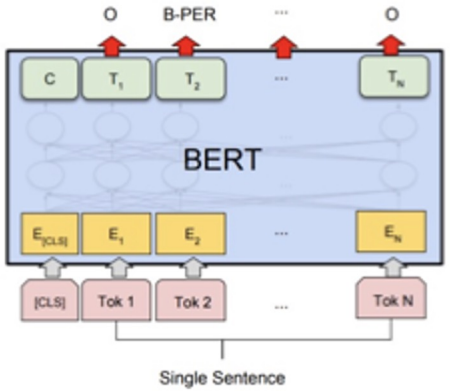
TF-IDF

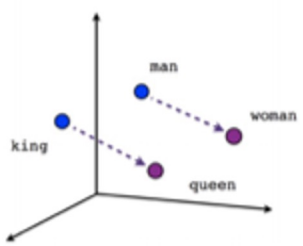
Term x within document y

$tf_{x,y}$ = frequency of x in y

df_x = number of documents containing x

N = total number of documents





Male-Female

Verb tense

Country-Capital

Transformer Models

- Inspired by encoder-decoder (RNN-based Sequence models) architectures
- Transformer encoders (BERT) map input sequences to high dimensional vector spaces
- Transformer decoders (GPT-2) map transforms vectors into output sequences.
- Rely solely on an upgraded attention mechanism rather than RNNs to generate sequences

$$\underbrace{\text{Attention}(Q, K, V)}_{\text{Queries, Keys, Values}} = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Related Work

- *Lee et al.* produced patent claims by fine-tuning the generic pretrained GPT-2 model with U.S. utility patents claims data
- *Feng et al.* finetuned GPT-2 on a small set of yelp review data-set
- *Vijjali et al.* utilize BERT-based transformers to detect false claims surrounding the COVID-19 pandemic.

Adversarial Machine Learning and Poisoning Attacks

- Adversarial machine learning is a technique used to subvert ML systems by providing deceptive inputs to the model.
- Poisoning attacks directly contaminate the training data-set
- Rely heavily on the use of synthesized and/or incorrect input data.

Related Work

- *VirusTotal* poisoning attack demonstrated by the McAfee Advanced Threat Research team
- *Kurana et al. created a reputation scoring mechanism to identify poisoned inputs.*

Methodology

Methodology

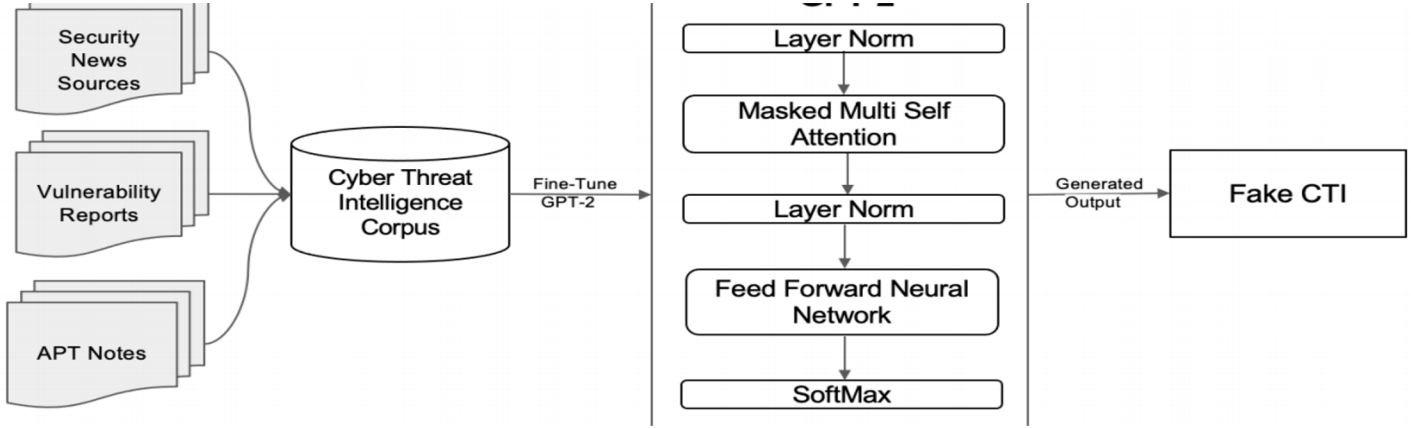
- Creating a Cybersecurity Corpus
- Fine-Tuning the GPT-2 with Cybersecurity Corpus
- Generating Fake Cyber Threat Intelligence Samples
- Evaluation

Creating a Cybersecurity Corpus

- Fine-Tune the public GPT-2 Model with a cybersecurity text to generate a large amount of fake, but plausible CTI
- Use the generated fake CTI text to perform a data poisoning attack on a CKG and cybersecurity corpus.

Fine-Tuning GPT-2 on Cyber Threat Intelligence Data

- Fine-Tune the public GPT-2 Model with a cybersecurity text to generate a large amount of fake, but plausible CTI
- Use the generated fake CTI text to perform a data poisoning attack on a CKG and cybersecurity corpus.



Evaluation

Samples of Generated Fake CTI using GPT-2

Research Question

Is the generated CTI plausible enough to bypass experienced Cyber Threat Hunters and Analysts?

Generation Input	Cyber Threat Intelligence	True/Fake
APT41 is a state-sponsored espionage group	APT41 is a state-sponsored espionage group which operates against higher education, travel services, and news/media firms which provide some indication that the group also tracks individuals and conducts surveillance.	**True CTI**
	APT41 is a state-sponsored espionage group. The Russian, Moscow-based group is responsible for capturing and stealing computer network communications and intellectual property. APT41 employs social engineering techniques to get around network defenses designed to last for years, and is capable of compromising and stealing dozens of companies	**Fake CTI (generated)**
Malicious Domain in SolarWinds Hack Turned into 'Killswitch'	Malicious Domain in SolarWinds Hack Turned into 'Killswitch'. A key malicious domain name used to control potentially thousands of computer systems compromised via the months-long breach at network monitoring software vendor SolarWinds was commandeered by security experts and used as a "killswitch" designed to turn the sprawling cybercrime operation against itself.	**True CTI**
	Malicious Domain in SolarWinds Hack Turned into Killswitch Service. This attack took place last year and affected hundreds of organizations. The attackers took advantage of the vulnerability in Win32k framework to attack their targets through crafted code.	**Fake CTI (generated)**
An issue was discovered in the Quiz and Survey Master plugin	An issue was discovered in the Quiz and Survey Master plugin before 7.0.1 for WordPress. It made it possible for unauthenticated attackers to upload arbitrary files and achieve remote code execution.	**True CTI**
	An issue was discovered in the Quiz and Survey Master plugin 1.2.1 for WordPress. By setting a CSRF token, an attacker can take advantage of a vulnerability in the survey.php admin panel.	**Fake CTI (generated)**

Evaluating the Generated Cyber Threat Intelligence

Evaluation

Ten threat hunters and analysts with 15+ years of experience evaluated 280 generated samples into true/false categories

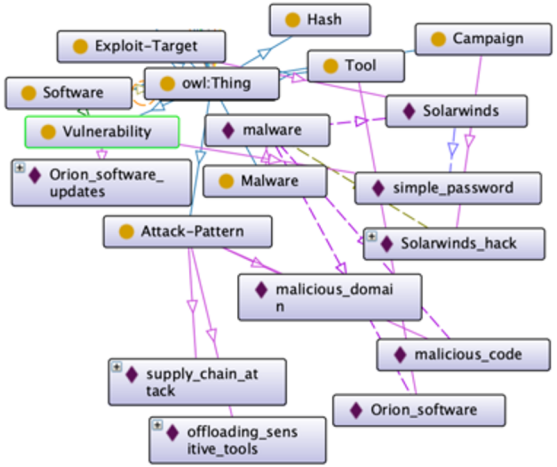
Results

Participants were only able to label 60/280 of the generated samples as fake and found the large majority (78.5%) of the fake samples as true.

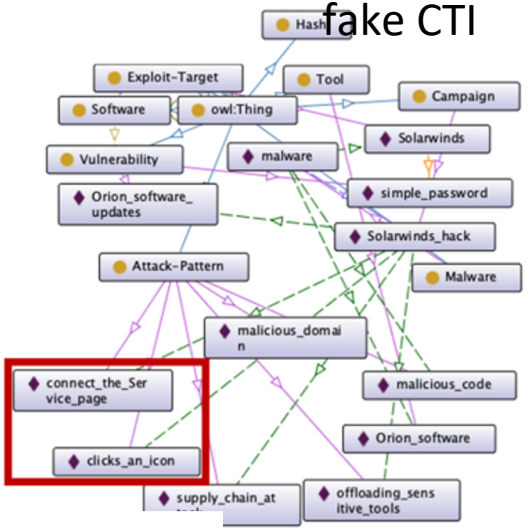
Actual Data	Participant Labels: True	Participant Labels: False	Total
True	206 Samples	74 Samples	280
False	220 Samples	60 Samples	280
Total	426	134	

Poisoning a Cybersecurity Knowledge Graph

CKG populated with data from legitimate CTI sources

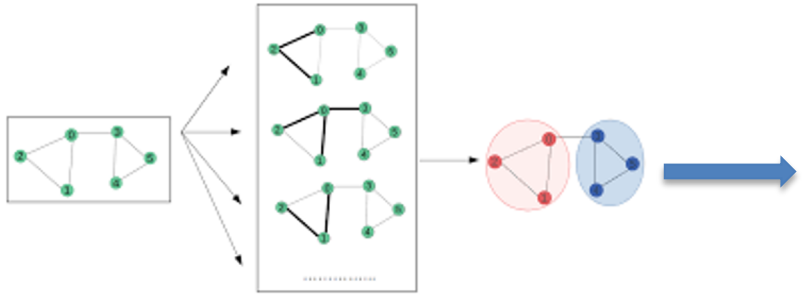


The poisoned CKG with additional data (red box) extracted from fake CTI




```
SELECT ?x WHERE {
  ?x a CKG:Attack-Pattern;
     ^CKG:uses CKG:Solarwinds-hack.}
```

Defenses against CTI poisoning (WIP)



Graph CNN

Trust scores for relationships

Future Work

- Defenses against fake cybersecurity information
- Develop systems that can detect linguistic errors and disfluencies that generative transformers commonly produce
- Detecting fake CTI text can use a combination of novelty, consistency, provenance, and trust.
- Fine-tuning the model with data from other specialized domains (like medicine).
- Change ontology to include classes that help us identify the source of the information

For questions please contact –
priyankaranade@umbc.edu
apiplai1@umbc.edu

THANK YOU!