Unstructured Threat Intelligence Processing using NLP
Enhancing Cyber Security Operations by Automating Threat Intelligence Extraction from Unstructured Sources

Human driven processes are inadequate to effectively utilize increasing amount of cyber threat intelligence

Challenges

- Threat intelligence provided in advisories, reports and other text formats require human analysts to parse and extract relevance.
- Difficult to scale human driven processes for increased amount of threat data that needs to analyzed.
- Multiple sources detail the same threat intelligence leading to wasted analysis effort and inefficiencies in general.
- Rely on human memory to spot connections among disparate advisories.
Desired Outcomes of Using Unstructured Threat Data

Solution that allows Threat Intelligence teams to utilize their human expertise by automating extraction of threat data from unstructured text sources

- Threat intelligence provided in advisories, reports and other text formats require human analysts to parse and extract relevance.
- Difficult to scale human driven processes for increased amount of threat data that needs to analyzed.
- Multiple sources detail the same threat intelligence leading to wasted analysis effort and inefficiencies in general.
- Rely on human memory to spot connections among disparate advisories.
- Conduct automated threat searches specific actors or TTPs
- Track threat data on indicators, actors and TTPs from multiple sources with minimal manual effort
Closing the Attacker’s Window of Opportunity

Cyber Kill Chain
- Reconnaissance
- Weaponize
- Deliver
- Exploit
- Install
- Command and Control
- Act

Attack State
- Pre-Attack
- Gain Foothold
- Ongoing Attack

Respond
- Reducing cost of containing and mitigating the attack

Structured Threat Intelligence Enabled
- Unstructured Threat Intelligence

Current Security Operations

Threat intelligence is disseminated quicker in unstructured text sources and is often more timely.
Solution Framework

Collect

Information Extraction

Standardize (STIX)

Filter & Prioritize

Reporting & Visualization

Natural Language Processing Module

Relational DB

Asset Categorization

Graph DB

Neo4j

Underground Websites

Social Media Sites

Analysis Reports

Advisories

Blogs

Flash Warnings

Websites

Advisories

Blogs

Analysis Reports

Underground

Social Media Sites

Relational DB

CMDB

Relevance Ratings

Asset Scoring

Indicator of Compromise

Course of Action (COA)

TTPS

Exploit Target

Campaigns

Actors

Incidents

Relational DB

Graph DB

Neo4j
Using the Unstructured Threat Intelligence Processing Tool
Upload a file for the parsing of threat data into STIX data constructs.

File Upload Tab

- To upload a file, click Choose Files. Select files from the pop-up menu, or drag files from your computer on to the box. To upload multiple files at once, hold down the Shift or Control key as you select files.
Copy and paste any text directly into the field below to extract threat data into STIX constructs.
Insert any URL into the field to retrieve threat data in text format of the given URL.
Automate threat data retrieval from various web sites via Scumblr to allow constant ingest of unstructured threat data.
The dashboard provides a view of extracted threat data from documents mapped to STIX data constructs. Each individual piece of threat data is counted to determine the overall threat's trend within the specified date range.
This page shows a prioritized list of threat documents that were uploaded based on their overall relevance score. Relevance is determined by the score of the ingested threat document. Additional actions can be taken on each ingested threat document, such as showing details or displaying visualizations of analytics on data from the documents.
FSISAC10.txt

10
affecting Adobe
affecting Java
Title: Microsoft Security Bulletin Advance Notification for December 2014
Tracking ID: 912536
Reported Date/Time: 05 Dec 2014 14:52:00 UTC
Priority: 3
Category: Upcoming Events
Audience: Analysts, Affiliates, Basic Members, Core Members, FSISAC Customer Service, Limited Observers, Premier Members, Standard Members, MSP Members
Description: This is an advance notification of security bulletins that Microsoft is intending to release on December 9, 2014.

Bulletin ID Maximum Severity Rating and Vulnerability Impact Restart Requirement
Affected Software
Drill down by STIX construct to see threat data extracted from the documents.

Up-vote or down-vote any element extracted to tune the NLP engine's accuracy.

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Asset Score</th>
<th>Element Extracted</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>80.0</td>
<td>ibm content collector</td>
<td>Two vulnerabilities in local users to gain control of the system.</td>
</tr>
<tr>
<td>0</td>
<td>80.0</td>
<td>oracle java se and jrookit</td>
<td>An unspecified vulnerability with partial confidentiality impact.</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>the way bind</td>
<td>A denial of service attack.</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>the vcac vmware remote console (vmrc ) function</td>
<td>VMware vCloud Access (VCA) Remote Console (VCAC) function vulnerability.</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>[Ports/Services Attacked:] Product(s):</td>
<td></td>
</tr>
</tbody>
</table>
A custom tree node visualization is generated for each document and shows the elements extracted, their categories under STIX and all other documents that contain a related element.