Interpreting CVE/CWE Clusters

Professor John Miller
Professor Luis De La Torre
Jordan Liebe
Tim Cain
About me

- BS in Computer Science from WSU Tricities
- Software Engineer @ Pacific Northwest National Laboratory
- First Year Graduate Student @ WSU Tricities
Acronyms and Definitions

- ML: Machine Learning
- CWE: Common Weakness Enumeration
- CVE: Common Vulnerabilities and Exposures
- Cluster: A grouping of data points that have been collected by our ML Algorithms because of some similarity in their data.
User Output from Cyber Advisor

- Details about the CVE you provided
- The Cluster this CVE is associated with and other related CVEs
- Predictions of the Top 1 & 2 CWEs
- A listing of all CWE’s in the cluster and their percentages (> 5% only)

*Tim* will go into detail about the Cyber Advisor interface in his demo
Cluster 107 Example

<table>
<thead>
<tr>
<th>Cluster #</th>
<th>1st Prediction</th>
<th>FP Count</th>
<th>2nd Prediction</th>
<th>SP Count</th>
<th>Total Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>107</td>
<td>CWE-74</td>
<td>483</td>
<td>CWE-89</td>
<td>336</td>
<td>~1914</td>
</tr>
</tbody>
</table>

Now let's explore CWE-74 and CWE-89 a little deeper
Cluster 70 Example

<table>
<thead>
<tr>
<th>Cluster #</th>
<th>1st Prediction</th>
<th>FP Count</th>
<th>2nd Prediction</th>
<th>SP Count</th>
<th>Total Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>70</td>
<td>CWE-119</td>
<td>352</td>
<td>CWE-20</td>
<td>351</td>
<td>~1914</td>
</tr>
</tbody>
</table>

Now let's explore CWE-74 and CWE-89 a little deeper
Cluster # 70

CWE-664 Tree

Root Level

CWE-664

Level 1

... CWE-118 ...

Level 2

... CWE-119 ...

CWE-707 Tree

Root Level

CWE-707

Level 1

... CWE-20 ...

Secondary

Primary
Future Improvements

What’s next?
Enhanced Search

• Current Limitations:
  • Search is limited to only the CVE’s we have processed into our model at the time of request

• Future Work
  • Add an “Enhanced Search” that can be given a new vulnerability description and will process it through Language Processing, V2W-Bert, etc. to return cluster results and related CWE’s immediately.
Built-In Cluster Interpretation and Analysis

- Current Limitations:
  - Most of our analysis on the clusters is done by hand with a combination of spreadsheets and scripts.

- Future Work
  - Add in automatic processing between our Machine Learning model and our web interface to automatically process the hand calculations we have been doing to add more meaning to the results.
Thank you!

Any Questions?