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.

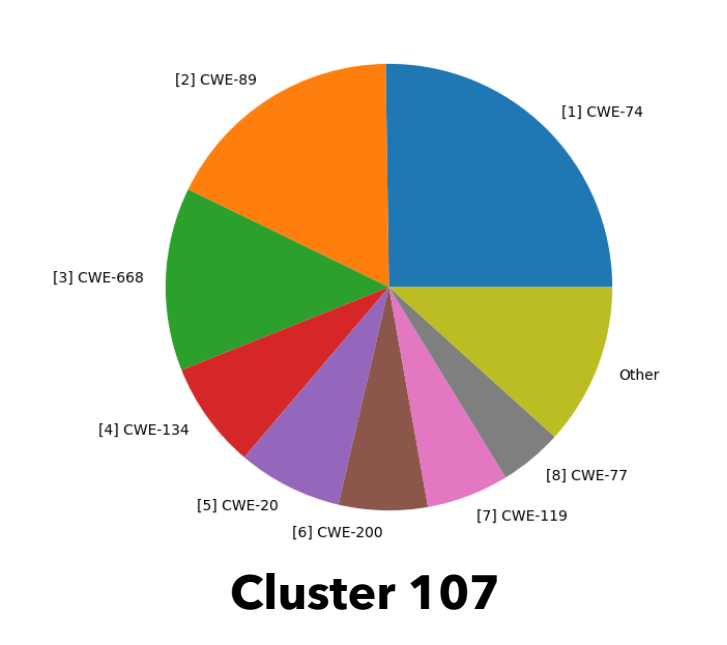
User Output from Cyber Advisor

- 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

Cluster #	1st Prediction	FP Count	2nd Prediction	SP Count	Total Count
107	CWE-74	483	CWE-89	336	~1914

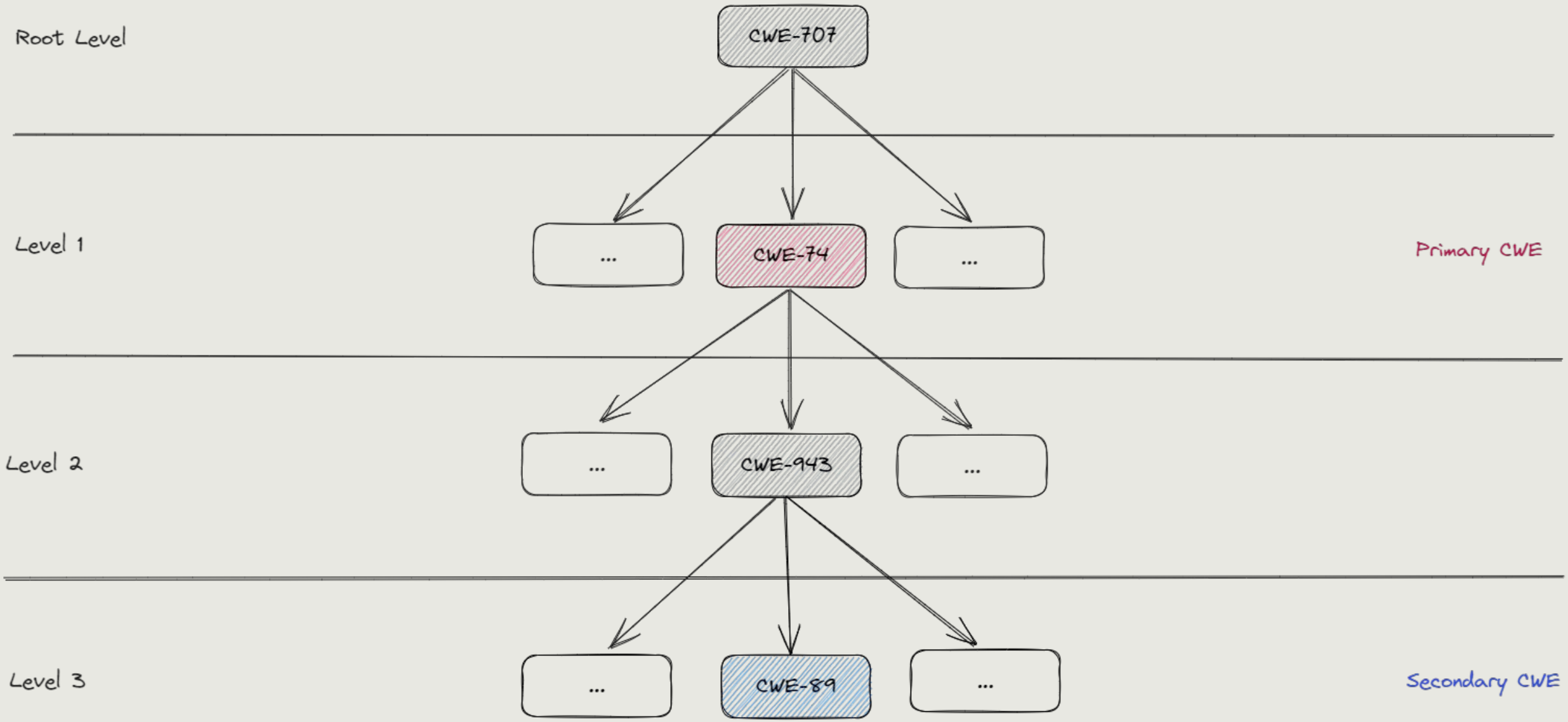


CWE-ID	Percentage
CWE-74	25.24%
CWE-89	17.56%
CWE-668	13.27%
CWE-134	7.73%
CWE-20	7.58%
CWE-200	6.43%
CWE-119	5.96%
CWE-77	4.55%
Other	11.70%

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

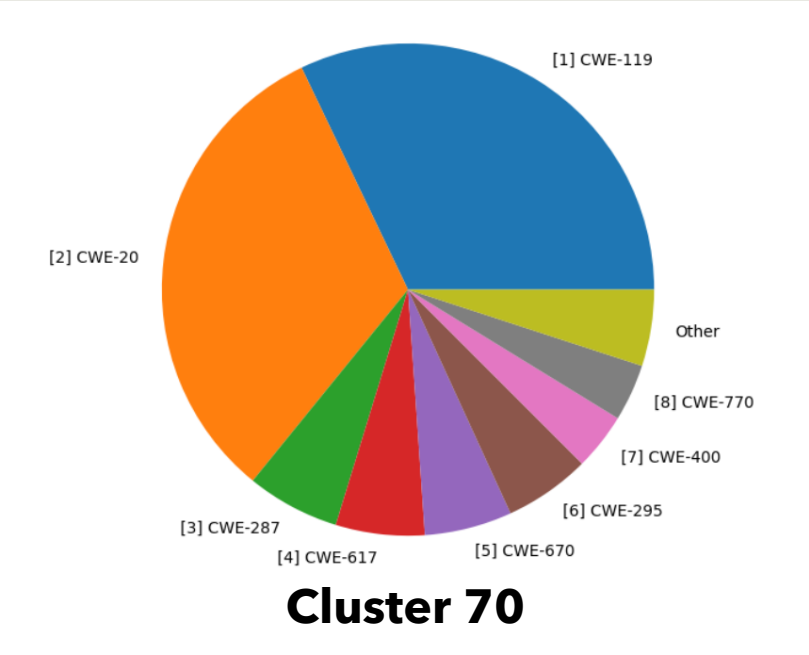
Cluster # 107

CWE-707 Tree



Cluster 70 Example

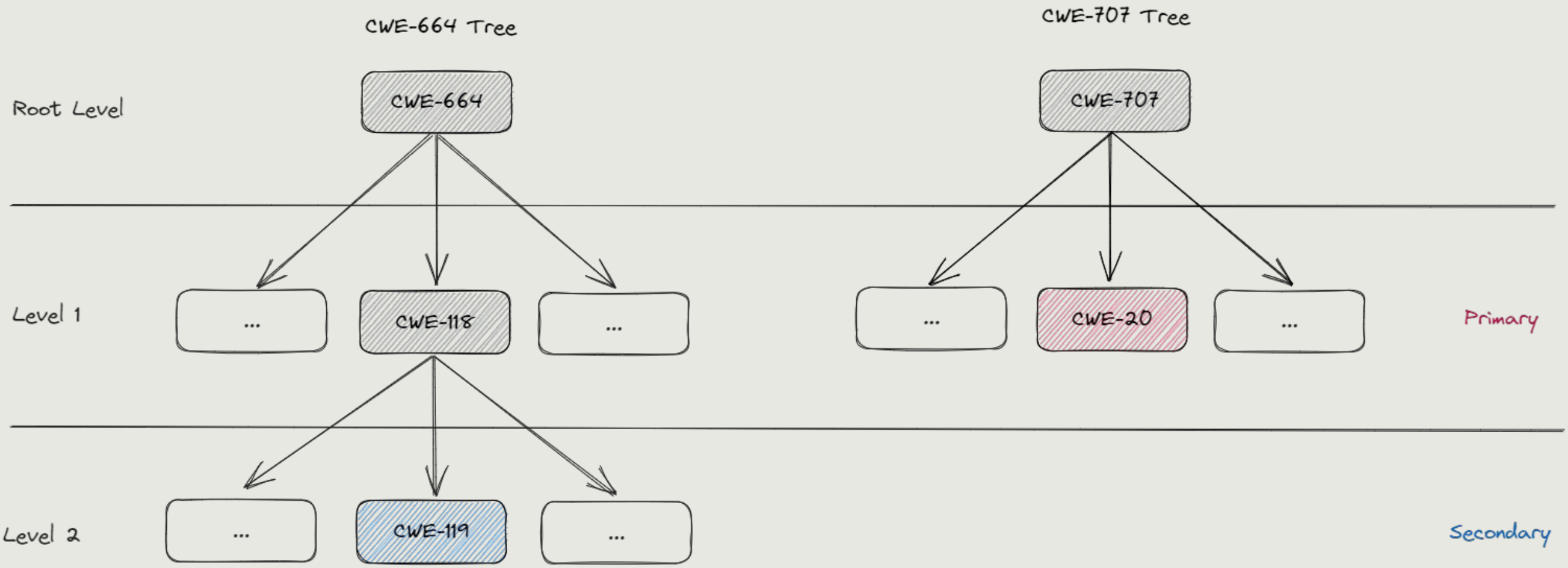
Cluster #	1st Prediction	FP Count	2nd Prediction	SP Count	Total Count
70	CWE-119	352	CWE-20	351	~1914



CWE-ID	Percentage
CWE-119	32.12%
CWE-20	32.03%
CWE-287	6.11%
CWE-617	5.84%
CWE-670	5.75%
CWE-295	5.66%
CWE-400	3.74%
CWE-770	3.74%
Other	5.02%

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

Cluster # 70


CWE-664 Tree
CWE-707 Tree
Root Level
CWE-664
CWE-707
Level 1
...
CWE-118
...
...
CWE-20
...
Primary
Level 2
...
CWE-119
...
Secondary

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 combinations 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?