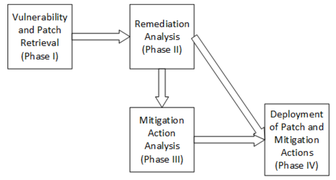
Towards Automated Asset-Vulnerability Mapping for Vulnerability Management

Philip Huff*, Kylie McClanahan*, Thao Le-Vasicek$, Qinghua Li*
* University of Arkansas
$ Bastazo, Inc.

Background

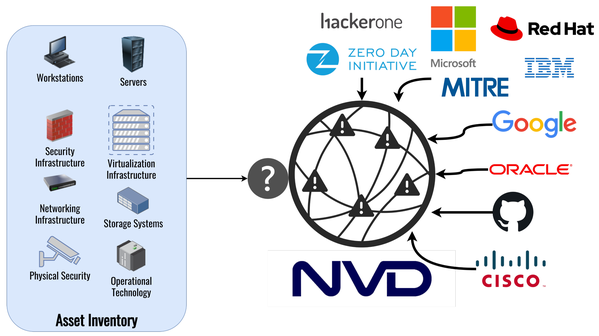
- Vulnerability and patch management (VPM) is one core component of security in energy companies



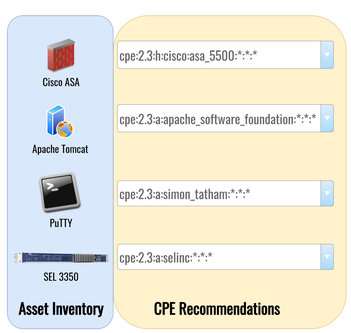
- Challenges
  - Many vulnerabilities with assets emerge each month
  - The National Electric Regulatory Commission (NERC) Critical Infrastructure Protection (CIP) compliance regulation CIP-007-6 R2 requires flawless vulnerability remediation
  - Energy companies need to remediate each and every vulnerability through patching or mitigation plans

Asset-Vulnerability Mapping

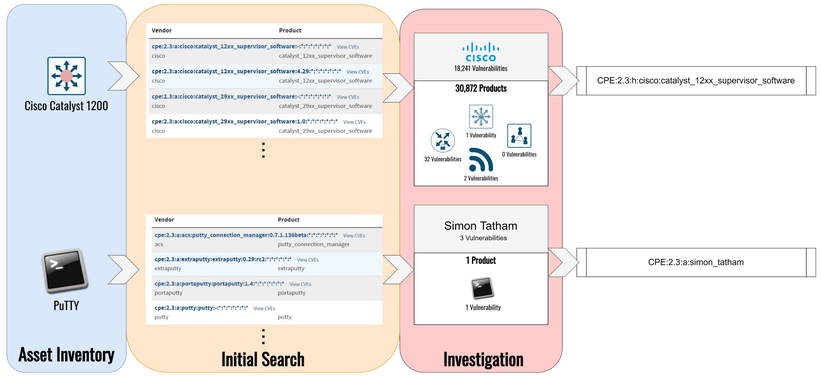
- Problem in identifying applicable vulnerabilities
  - Software/hardware names in an organization's asset inventory usually do not follow the same naming convention used by public vulnerability databases



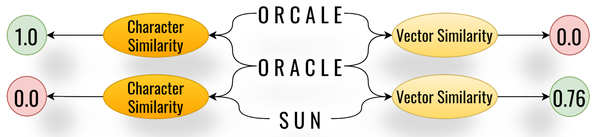
- Goal: automatically mapping asset to a small set of Common Platform Enumerations (CPEs)



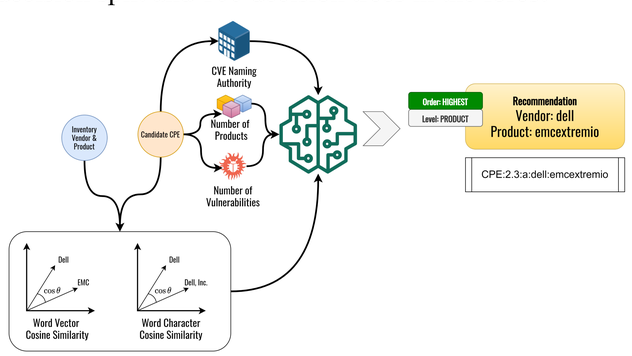
CPE Recommender



- Common Platform Enumeration (CPE)
  - Type: Hardware, Operating System, Applications
  - Vendor: Cisco, Microsoft, Adobe
  - Product: Catalyst_3560cg-8tc-s, Windows Server 2016, Acrobat_reader_dc
- Measuring similarity
  - word2vec vector representation to capture semantic meaning
  - Synthetic data used to create vectors for many Out-of-Vocabulary vectors in the CPE dictionary
  - Simple character cosine similarity meant to account for variance in an entity's asset inventory



- Fuzzy mapping pipeline
  - Output the order of recommendation and whether to match on the vendor only or the vendor-product
  - Random Forest Classifier using Gini impurity measure for the decision split and 100 decision trees in the forest



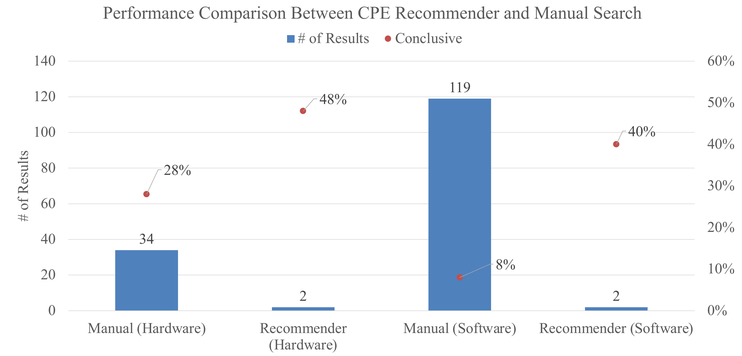
Evaluation Results

- Training dataset of 23,048 annotated samples generated from Wikipedia scrape for software and hardware products (available at https://github.com/pdhuff/cpe_recommender)
- We used an 80/20 train/test split.

Table 3: Machine Learning 'Order' Classification Results for CPE Matching

Recommended Order	Precision	Recall	F-Score	Support
HIGHEST	100%	98%	99%	937
HIGH	80%	66%	72%	232
MEDIUM	72%	46%	56%	211
LOW	89%	87%	88%	1,076
LOWEST	98%	99%	98%	5,854
REJECT	99%	100%	100%	14,738
Weighted Average	**98%**	**98%**	**98%**	**23,048**

- We tested this methodology on 50 software products and 50 hardware products
- Comparison with a human analyst
- Time savings: over 7 hours



Acknowledgment

- Philip Huff, Kylie McClanahan, Thao Le-Vasicek, and Qinghua Li, "A Recommender System for Tracking Vulnerabilities," the 3rd International Workshop on Next Generation Security Operations Center (NG-SOC), 2021
- This work was supported in part by the NSF under award 1751255