Demonstrating Praxi
SOFTWARE DISCOVERY THAT LEARNS FROM PRACTICE

BY ANTHONY BYRNE¹, SADIE L. ALLEN¹, SHRIPAD NADGOWDA², AND AYSE K. COSKUN¹

¹Boston University; ²IBM T.J. Watson Research Center

20th ACM/IFIP International Middleware Conference
December 11, 2019
Motivation

- Cloud systems (bare metal, VMs, containers) evolve rapidly over time
- New software and vulnerabilities announced every day
- Without constant visibility, cloud software quickly ages/becomes insecure
  - How do we keep track of what software is installed on a cloud system?

![Diagram showing the process of a security bug being discovered and patched]

- Cloud user deploys software with unknown security bug to some servers
- Security bug discovered. Cloud operators scramble to patch affected systems
- Administrator uses Praxi to generate fingerprint of affected software
- Praxi uses detection history to identify systems that need to be patched
Previous Solution: Statistical Analysis

Columbus: Practice-Based Discovery Method

- Exploit software naming conventions to build modified trie
- Trie then analyzed via freq. counts to pull out significant tags
- Tags hold useful information like app name, version, etc.
- Upside: corpus-less, lightweight
- Downside: tags not consistent/machine-readable

S. Nadgowda et al., "Columbus: Filesystem Tree Introspection for Software Discovery" (IEEE IC2E 2017)
Key Insight: Discovery By Example

- Machine Learning by Experience
  - Automatic
  - Incremental
  - Generic
  - Distortion resistant

- So how do we apply this to software discovery?

H. Chen et al., "Automated system change discovery and management in the cloud" (IBM Journ. of R. & D. 2016)
A. Byrne et al., “Praxi: Cloud Software Discovery That Learns From Practice” (*IEEE TCC*, submitted Nov. ’18, revised Jul. ’19)

**Praxi: Learning From Practice**

COMBINING THE BEST ELEMENTS OF LEARNING- AND PRACTICE-BASED METHODS

Also seen in: *IEEE IC2E 2019 Tutorials*
Accuracy

(higher F1 scores are better)

Single-label Classification
- Installed one application per recording period
- Average F1 > 0.99

Multi-label Classification
- Installed multiple applications per recording period
- Average F1 = 0.967
Praxi Overhead Compared to Previous Work

<table>
<thead>
<tr>
<th>Phase</th>
<th>Operation</th>
<th>Praxi Time (min)</th>
<th>Praxi Disk (MB)</th>
<th>DeltaSherlock Operation</th>
<th>DeltaSherlock Time (min)</th>
<th>DeltaSherlock Disk (MB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature Reduction</td>
<td>Columbus Tag Extraction</td>
<td>3.7</td>
<td>55</td>
<td>w2v Dictionary Generation</td>
<td>13.1</td>
<td>370</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Fingerprinting</td>
<td>55</td>
<td>24</td>
</tr>
<tr>
<td>Discovery By Example</td>
<td>VW Model Training</td>
<td>1.5</td>
<td>59</td>
<td>RBF Model Training</td>
<td>11</td>
<td>489</td>
</tr>
<tr>
<td></td>
<td>VW Model Evaluation</td>
<td>0.2</td>
<td>-</td>
<td>RBF Model Evaluation</td>
<td>0.7</td>
<td>-</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td>5.4</td>
<td>114</td>
<td>Overall</td>
<td>79.8</td>
<td>883</td>
</tr>
</tbody>
</table>

- Main savings come from...
  - Lack of dictionary generation step
  - Faster machine learning system
  - Smaller machine learning models
Iterative Training

- Initial ML model training is costly
- Non-iterative models will quickly become "stale," requiring full retraining
- Iterative models can be updated several times, minimizing training costs
Live Demo Overview

1) Initial Training

- Initial Training Tagsets:
  - django
  - gitweb
  - ... subversion

- Testing Tagsets:
  - django
  - gitweb

- train_first_model

- Initial Model

- Mislabeled Tagsets:
  - label: pandas prediction: django
  - label: pandas prediction: django
  - label: pandas prediction: django
  - label: pandas prediction: django
  - F1 Score: 0.977

- show_results_1

2) Recording a Changeset and Converting it into a Tagset

- New Pandas Changeset: [empty]
- record_pandas
- install_pandas
- generate_tagset

- New Pandas Tagset:
  - pandas/io/clipboard/clipboards.py
  - pandas/tests/io/test_gcs.py
  - pandas/tests/frame/test_rank.py
  - ... subversion
- pandas

3) Iteratively Training the Model

- Initial Model
- retrain_model

- Iterative Model

- Mislabeled Tagsets:
  - [none]
  - F1 Score: 1.00

- show_results_2
- Software discovery key to any cloud integrity solution
- Praxi discovers software accurately and automatically with low overhead

**Concluding Remarks**

More info at bu.edu/peaclab

Please send feedback to abyrne19@bu.edu