Automated Known Problem Diagnosis with Event Traces

Motivation

- Problem diagnosis
  - Labor intensive diagnosis process
  - Manual Inspection of Solutions
  - Inefficient due to too much human involvement
- Automating diagnosis process for known problems
  - Novel trace based problem diagnosis
Solution Approach

- Known problems are annotated with relevant system behavior
- New behavior -> classify to some known problem

High Level System Design
Tracer

- What events to collect?
  - System calls

- What attributes to collect?
  - Process / thread Id
  - Process / thread name
  - System call name, parameters, return value

Trace Example

<table>
<thead>
<tr>
<th>#</th>
<th>process</th>
<th>thread</th>
<th>syscall</th>
<th>parameters &amp; return value</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18419</td>
<td>iexplore.exe</td>
<td>3892</td>
<td>CreateThread</td>
<td>Process: 3888, Thread: 3896 SUCCESS</td>
</tr>
<tr>
<td>18420</td>
<td>iexplore.exe</td>
<td>3892</td>
<td>PostMessageWM_USER+0x300</td>
<td>1</td>
</tr>
<tr>
<td>18421</td>
<td>iexplore.exe</td>
<td>3892</td>
<td>OpenKey</td>
<td>HKCU\SOFTWARE \Microsoft\Internet Explorer\Main SUCCESS</td>
</tr>
<tr>
<td>18422</td>
<td>iexplore.exe</td>
<td>3892</td>
<td>QueryValue</td>
<td>HKCU\SOFTWARE\Microsoft \Internet Explorer\Main\Enable Browser Extensions NOTFOUND</td>
</tr>
<tr>
<td>18423</td>
<td>iexplore.exe</td>
<td>3892</td>
<td>OpenKey</td>
<td>HKLM\SOFTWARE\Microsoft \Windows\CurrentVersion\Internet Settings SUCCESS</td>
</tr>
</tbody>
</table>
**Classifier**

- **Classification:**
  - Training: learn a model from annotated training data
  - Testing: predict the class a new one belongs to

- **Accuracy:**
  - Percentage of true positive data

- **Cross validation**
- **N-gram**
  - Any N successive elements in a sequence
- **Support Vector Mechanism**

**System Call Variation**

- **Noise filtering**
  - Patterns occurring at less than a threshold % times are discarded
- **Object name canonicalization**
  - File path is discarded
- **Cross machine comparison**
Evaluation

- 4 target problems:
  - IE display
  - Firefox display
  - Outlook Express Open
  - Shared Folder

- Data Collection
  - Machine > Round > Problem
    - Inject fault
    - Start tracer → Reproduce → Stop tracer
    - Remove the fault

Canonicalization

[Bar chart showing accuracy before and after for different problem instances]
Higher N-grams

- No thread name, parameters and return values

Attributes
- No thread name, parameters and return values
Summary

- Canonicalization has no effect
- Longer patterns only helpful when fewer attributes are available

Cross machine evaluation

- Canonicalization

![Accuracy Chart](chart.png)
No thread name, parameter and return value
Impact of number of Training machines

![Graph showing accuracy convergence with number of training machines for different datasets.]

Summary

- Canonicalization is good
- 1-gram is good enough, 2-gram useful when smaller number of attributes
- Accuracy converges more quickly with larger number of machines
Questions?

Thank You