Interaction Traces Mining for Efficient System Responses Generation
Interaction Traces Mining for Efficient System Responses Generation

Agenda

Introduction

Background & Motivation

Approach

Evaluation

Future Work

Conclusion
Introduction

- Realistic interaction trace collection is informative
  - Interactive data
  - Basis for system/service emulation technique

- Challenge: Searching the entire interaction collection during runtime
  - Time-consuming
  - Impractical

- Our approach:
  - Using mining techniques to accelerate the searching process
Background & Motivation

- Modern software systems
  - Large
  - Often rely on third party services/systems
  - Complex interactions among systems
- Quality assurance becomes a problem
- Provision of an emulation environment, which is
  -Executable
  -Interactive
  -Less resource requirement
Background & Motivation

- Configuration of the Emulation Environment

Diagram:

Enterprise System → Modeled → Executable Interaction Model → Emulator

System Under Test
Background & Motivation

- Executable interaction models
  - Objective:
    - Efficiently generate required responses for the SUT
  - Common approaches:
    - Manual
    - Automatic
- Prior work
  - An automatic method
  - Iterate the entire interaction trace collection
Proposed Framework

Pre-processing Stage:
Analysis Function

Run-time Stage:
Distance Function

Translation Function

Send Request “Req”
Receive Response “Res”

Req_{in}  Req_{sim}, Res_{sim}

Res_{out}
### Scenario Example

- **Incoming request**
  - Message ID: 18
  - ProtocolOp: `searchRequest`
  - `ObjectName: cn=Mal BAIL, ou=Administration, ou=Corporate, o=DEMOCORP, c=AU`

- **Synthesized response**
  - Message ID: 18
  - ProtocolOp: `searchResEntry`
  - `ObjectName: cn=Mal BAIL, ou=Administration, ou=Corporate, o=DEMOCORP, c=AU`
  - ProtocolOp: `searchResDone`
  - `resultCode: success`

- **Recorded request**
  - Message ID: 37
  - ProtocolOp: `searchRequest`
  - `ObjectName: cn=Miao DU, ou=Research, ou=Project, o=DEMOCORP, c=AU`

- **Recorded response**
  - Message ID: 37
  - ProtocolOp: `searchResEntry`
  - `ObjectName: cn=Miao DU, ou=Research, ou=Project, o=DEMOCORP, c=AU`
  - ProtocolOp: `searchResDone`
  - `resultCode: success`
Proposed Framework

Pre-processing Stage:
- Analysis Function
- Interaction Traces
  - Extracted Information

Run-time Stage:
- Distance Function
- Translation Function
  - \( R_{\text{in}} \), \( R_{\text{out}} \)
  - \( R_{\text{in}} \), \( R_{\text{sim}}, R_{\text{Sim}} \)

Send Request “Req”
Receive Response “Res”
Design of Analysis Function

Step 1:

Transforming Network Data

Index
1: addRequest(36) "cn=Miao DU, ou=administration, ou=Corporate, o=DEMOCORP, c=AU"
2: searchRequest(147) "cn=Alfred FITZGERALD, ou=Legal, ou=Corporate, o=DEMOCORP, c=AU" baseObject
... 
5: addRequest(171) "cn=Debbie DALLY, ou=Finance, ou=Corporate, o=DEMOCORP, c=AU"
... 
8: searchRequest(159) "cn=Barbara HARTLEY, ou=Management, ou=Corporate, o=DEMOCORP, c=AU" baseObject

Translate raw network data to corresponding text format
Design of Analysis Function

Step 2:

<table>
<thead>
<tr>
<th>Building Messages Distance Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0000  0.2168  Ö  0.1227  0.2111</td>
</tr>
<tr>
<td>0.2168  0.0000  Ö  0.1860  0.1410</td>
</tr>
<tr>
<td>Ö    Ö     Ö     Ö     Ö    Ö</td>
</tr>
<tr>
<td>0.1227  0.1860  Ö  0.0000  0.1806</td>
</tr>
<tr>
<td>0.2111  0.1410  Ö  0.1806  0.0000</td>
</tr>
</tbody>
</table>

Distance Matrix Image
Design of Analysis Function

Step 3:

<table>
<thead>
<tr>
<th>Clustering Messages</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0000   0.2168   0.1860   0.1410</td>
</tr>
<tr>
<td>0.2168   0.0000   0.1227   0.2111</td>
</tr>
<tr>
<td>0.1860   0.1227   0.0000   0.1806</td>
</tr>
<tr>
<td>0.1410   0.2111   0.1806   0.0000</td>
</tr>
</tbody>
</table>

Distance Matrix Image
Design of Analysis Function

Step 4:

Exporting Clusters

0.0000 0.2168 0 0 0 0
0.2168 0.0000 0 0 0.0000 0.1806
0 0 ... 0 0.1806 0.0000

Manually select clusters to export
Interaction Traces Mining for Efficient System Responses Generation

Design of Analysis Function

Step 5:

Clustersí Centres Selection

Centres:

- addRequest ...
- searchRequest ...

SCIENCE | TECHNOLOGY | INNOVATION
Evaluation

- Test Cases
  - Simple Object Access Protocol (SOAP)
  - Lightweight Directory Access Protocol (LDAP)

- Trace size: 1000 interactions for each test case

- Experiment
  - No Cluster – Not use analysis function
  - Whole Cluster – Require further searching in the selected cluster
  - Cluster Only – No further searching in the selected cluster
Evaluation

- Selected Clustering Methods
  - VAT : Visual Assessment of (cluster) Tendency
  - BEA : Bond Energy Algorithm

- Criteria for evaluating the effectiveness of our proposed approach
  - Identical
    - Protocol Conformant Valid
  - Well-Formed (Well formed message, but out of sequence) Invalid
  - Ill-Formed
Interaction Traces Mining for Efficient System Responses Generation

Effectiveness Evaluation Results

- **SOAP**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No Cluster</td>
<td>None</td>
<td>1,000</td>
<td>1,000</td>
<td>85</td>
<td>915</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Whole Cluster</td>
<td>VAT</td>
<td>1,000</td>
<td>1,000</td>
<td>90</td>
<td>910</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>BEA</td>
<td>1,000</td>
<td>1,000</td>
<td>82</td>
<td>918</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Center Only</td>
<td>VAT</td>
<td>1,000</td>
<td>1,000</td>
<td>87</td>
<td>913</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>BEA</td>
<td>1,000</td>
<td>1,000</td>
<td>92</td>
<td>908</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

- **LDAP**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No Cluster</td>
<td>None</td>
<td>1,000</td>
<td>908</td>
<td>451</td>
<td>457</td>
<td>89</td>
<td>3</td>
</tr>
<tr>
<td>Whole Cluster</td>
<td>VAT</td>
<td>1,000</td>
<td>751</td>
<td>360</td>
<td>391</td>
<td>246</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>BEA</td>
<td>1,000</td>
<td>753</td>
<td>383</td>
<td>370</td>
<td>241</td>
<td>6</td>
</tr>
<tr>
<td>Center Only</td>
<td>VAT</td>
<td>1,000</td>
<td>751</td>
<td>296</td>
<td>455</td>
<td>235</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>BEA</td>
<td>1,000</td>
<td>761</td>
<td>330</td>
<td>431</td>
<td>232</td>
<td>7</td>
</tr>
</tbody>
</table>
Efficiency Evaluation Results

- SOAP

![Graph showing Response Generation Time vs Incoming Request Length](chart.png)
Evaluation Results

- LDAP

Response Generation Time

Summary of Evaluation Results for LDAP
## Result Summary

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Cluster Method (VAT)</th>
<th>Response Time (ms)</th>
<th>Dist. Calcs.</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SOAP</strong></td>
<td>No Cluster</td>
<td>128.6</td>
<td>899</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>Whole</td>
<td>22.92</td>
<td>156</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>Center</td>
<td>0.90</td>
<td>6</td>
<td>100%</td>
</tr>
<tr>
<td><strong>LDAP</strong></td>
<td>No Cluster</td>
<td>53.28</td>
<td>899</td>
<td>90.8%</td>
</tr>
<tr>
<td></td>
<td>Whole</td>
<td>5.46</td>
<td>96</td>
<td>75.1%</td>
</tr>
<tr>
<td></td>
<td>Center</td>
<td>0.79</td>
<td>11</td>
<td>75.1%</td>
</tr>
</tbody>
</table>
Future Work

- Relief the reliance on formation translation function of the network traffic capture tool
- Devise better cluster centre selection method
- Further improve the efficiency of distance calculation approach
Conclusion

- Using mining techniques before runtime
  - Greatly improve response generation time
  - Facilitate the mimicking of software interaction behaviours
  - Eliminate the required human effort
Interaction Traces Mining for Efficient System Responses Generation

Many thanks!

Contact: miaodu@swin.edu.au