Integrating Workload Specification and Extraction for Model-Based and Measurement-Based Performance Evaluation
An Approach for Session-Based Software Systems

André van Hoorn, Christian Vögele
Eike Schulz, Wilhelm Hasselbring, Helmut Krcmar
Problem Statement

Situation

• Workload specification and execution essential to evaluate performance properties of session-based application systems
  – Measurement-based approaches (e.g., Krishnamurthy et al. 2006, Menascé et al. 1999, Arlitt et al. 2001)
  – Model-based approaches (e.g., Becker et al. 2009, Kounev et al. 2014)

Complication

• Manual creation of representative workload specifications is difficult, time consuming, and error-prone (Shams et al. 2006)
• Nowadays, workload specifications for measurement- and model-based approaches are modelled separately of each other (M-by-N problem)

Resolution: WESSBAS

1) System- and tool-agnostic modeling of probabilistic workloads of session-based application systems
2) Automatic extraction of these specifications from running systems
3) Transformation of these specifications into
   1) load test scripts
   2) architecture-level workload model specifications
WESSBAS Approach
To appear at VALUETOOLS 14

To appear at LT 2015 @ ICPE 2015
Agenda

- Problem Statement and Overview of Approach
- WESSBAS Approach
  1. WESSBAS-DSL
  2. Extraction of WESSBAS-DSL Instances
  3. Clustering of Customer Groups
  4. Transformations
     1. Apache JMeter
     2. Palladio Component Model
- Evaluation
- Future Work
Background – Markov4JMeter

Example of an Application Model

[van Hoorn et al. 2008]
Background – Markov4JMeter

Example of an Application Model

Application Model

[n = number of items]

0.01 login

[van Hoorn et al. 2008]
WESSBAS-DSL

Application Model
Behavior Model + Behavior Mix

Background – Markov4JMeter

[van Hoorn et al. 2008]
WESSBAS-DSL

Behavior Mix and Behavior Model(s)
Behavior Model Extraction

Monitoring

Kicker

User Session Traces

TXT

Extraction

Java Code
Behavior Model Extractor

Probabilities + Think Times

CSV

Behavior Mix

TXT
Clustering of Customer Groups

X-means

- (Absolute) Transition count matrix per session
- Determine number of clusters and centroids
- Assigns instances to the nearest cluster centroid using Euclidean Distance (Non-normalized and normalized)
- Grouping based on minimum distance, Recalculation of centroids

- **Probabilities:** Each centroid represents the (relative) transition count matrix
- **Think Times:** Calculate based on resulting clusters (not part of clustering)
- **Behavior Model(s):**
- **Behavior Mix:** Proportion of the number of instances per cluster
Transformation into Apache JMeter Test Plans
Transformation into Palladio Component Models

Automatically generated (e.g., Brunnert et al. 2013, Brosig et al. 2011)

Missing possibility to reference usage models from usage models

Generation of large parts of the workload specification into the repository model

Becker et al. (2009)
Transformation into Palladio Component Models

WESSBAS-DSL
Transformation into Palladio Component Models

WEESBAS-DSL

Call to the modelled system operation
Agenda

• Problem Statement and Overview of Approach
• WESSBAS Approach
  1. WESSBAS-DSL
  2. Extraction of WESSBAS-DSL Instances
  3. Clustering of Customer Groups
  4. Transformations
     1. Apache JMeter
     2. Palladio Component Model

• Evaluation
• Future Work
Evaluation Goals and Methodology

Research Questions
1. How accurately do the clustering results match the input Behavior Mix?
2. What is the impact of the clustering results on the workload characteristics?
3. How well match the predicted workload characteristics the measured workload characteristics?

Methodology
– Instrumentation of SPECjEnterprise2010 using Kieker to obtain session logs
– Extraction of behavior models and behavior mix (includes clustering)
– Extraction of WESSBAS-DSL instances
– Scenario 1: Transformation to JMeter test plans
  • Generation of basic application model (only session layer)
  • No input data, no guards and actions
– Scenario 2: Transformation of WESSBAS-DSL instances to PCM
  • Execute simulation
  • Compare simulation results (request counts) with measured data
Probabilistic Representation of SPECj Workload
### Accuracy of Clustering

#### Research Question 1

**X-means (min 3 cluster, max 3 cluster)**

<table>
<thead>
<tr>
<th>TM</th>
<th>T</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>MC</th>
<th>NED</th>
<th>ED</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>B</td>
<td>0</td>
<td>0</td>
<td>31,060</td>
<td>0</td>
<td>31,060</td>
<td>2.91%</td>
<td>0</td>
<td>31,060</td>
<td>0</td>
</tr>
<tr>
<td>25</td>
<td>M</td>
<td>15,298</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15,298</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>P</td>
<td>1,789</td>
<td>13,353</td>
<td>0</td>
<td>0</td>
<td>15,142</td>
<td>24.62%</td>
<td>632</td>
<td>14,666</td>
<td>0</td>
</tr>
</tbody>
</table>

**X-means (min 2 cluster, max 20 cluster)**

<table>
<thead>
<tr>
<th>TM</th>
<th>T</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>MC</th>
<th>NED</th>
<th>ED</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>MC</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>B</td>
<td>0</td>
<td>0</td>
<td>31,060</td>
<td>0</td>
<td>31,060</td>
<td>0</td>
<td>0</td>
<td>31,060</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>M</td>
<td>15,298</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15,298</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>31,060</td>
<td>1.03%</td>
<td>61,500</td>
</tr>
</tbody>
</table>

| TM: Transaction Mix |
| T: Transaction |
| C<sub>N</sub>: Assigned Cluster |
| MC: Percentage of misclassified |
| N: Number of instances |

| ED: Euclidean Distance |
| NED: Normalized Euclidean Distance |
Accuracy of Extracted Workload Specifications

Research Question 2

<table>
<thead>
<tr>
<th>Request</th>
<th>Orig.</th>
<th>ED-2</th>
<th>NED-3</th>
<th>NED-4</th>
<th>Rel.</th>
</tr>
</thead>
<tbody>
<tr>
<td>add to cart</td>
<td>63,761</td>
<td>63,316</td>
<td>64,250</td>
<td>61,838</td>
<td>0.07</td>
</tr>
<tr>
<td>cancel order</td>
<td>632</td>
<td>607</td>
<td>634</td>
<td>591</td>
<td>0.00</td>
</tr>
<tr>
<td>clear cart</td>
<td>6,047</td>
<td>5,941</td>
<td>6,140</td>
<td>5,843</td>
<td>0.01</td>
</tr>
<tr>
<td>defer order</td>
<td>6,782</td>
<td>6,799</td>
<td>6,863</td>
<td>6,651</td>
<td>0.01</td>
</tr>
<tr>
<td>home</td>
<td>69,934</td>
<td>60,957</td>
<td>62,054</td>
<td>59,971</td>
<td>0.07</td>
</tr>
<tr>
<td>inventory</td>
<td>20,596</td>
<td>30,212</td>
<td>31,378</td>
<td>29,808</td>
<td>0.03</td>
</tr>
<tr>
<td>login</td>
<td>61,550</td>
<td>60,957</td>
<td>62,054</td>
<td>59,971</td>
<td>0.07</td>
</tr>
<tr>
<td>logout</td>
<td>59,934</td>
<td>60,957</td>
<td>62,054</td>
<td>59,971</td>
<td>0.07</td>
</tr>
<tr>
<td>purchase cart</td>
<td>8,360</td>
<td>8,328</td>
<td>8,351</td>
<td>8,139</td>
<td>0.01</td>
</tr>
<tr>
<td>remove</td>
<td>3,027</td>
<td>2,993</td>
<td>3,044</td>
<td>3,064</td>
<td>0.00</td>
</tr>
<tr>
<td>sell inventory</td>
<td>66,679</td>
<td>65,413</td>
<td>67,691</td>
<td>64,794</td>
<td>0.08</td>
</tr>
<tr>
<td>shopping cart</td>
<td>9,074</td>
<td>8,534</td>
<td>9,184</td>
<td>8,907</td>
<td>0.01</td>
</tr>
<tr>
<td>view items</td>
<td>498,601</td>
<td>492,675</td>
<td>499,983</td>
<td>485,611</td>
<td>0.57</td>
</tr>
<tr>
<td>Σ</td>
<td>874,927</td>
<td>868,089</td>
<td>883,680</td>
<td>855,159</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Violin plot (combination of box and density plot)
### Accuracy of PCM Workload Specification

#### Research Question 3

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>add to cart</td>
<td>63,761</td>
<td>64,943 1.82%</td>
<td>61,812 3.15%</td>
<td>60,986 4.55%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cancel order</td>
<td>632</td>
<td>609 3.78%</td>
<td>661 4.39%</td>
<td>625 1.12%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>clear cart</td>
<td>6,047</td>
<td>6,178 2.12%</td>
<td>5,927 2.02%</td>
<td>5,846 3.44%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>defer order</td>
<td>6,782</td>
<td>6,873 1.32%</td>
<td>6,524 3.95%</td>
<td>6,606 2.66%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>home</td>
<td>59,934</td>
<td>61,146 1.98%</td>
<td>58,747 2.02%</td>
<td>58,744 2.03%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>inventory</td>
<td>30,596</td>
<td>30,539 0.19%</td>
<td>29,574 3.46%</td>
<td>29,405 4.05%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>login</td>
<td>61,500</td>
<td>61,156 0.56%</td>
<td>58,747 4.69%</td>
<td>58,745 4.69%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>logout</td>
<td>59,934</td>
<td>61,146 1.98%</td>
<td>58,747 2.02%</td>
<td>58,744 2.03%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>purchase cart</td>
<td>8,360</td>
<td>8,388 0.33%</td>
<td>7,976 4.81%</td>
<td>7,836 6.69%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>remove</td>
<td>3,027</td>
<td>2,986 1.37%</td>
<td>2,876 5.25%</td>
<td>2,949 2.64%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sell inventory</td>
<td>66,679</td>
<td>66,131 0.83%</td>
<td>63,185 5.53%</td>
<td>63,914 4.33%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>shopping cart</td>
<td>9,074</td>
<td>9,164 0.98%</td>
<td>8,803 3.08%</td>
<td>8,795 3.17%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>view items</td>
<td>498,601</td>
<td>491,812 1.38%</td>
<td>470,392 6.00%</td>
<td>475,000 4.97%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>∑</strong></td>
<td>874,927</td>
<td>871,071 0.44%</td>
<td>833,971 4.91%</td>
<td>838,195 4.38%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MRC: Measured Request Count  
SRC: Simulated Request Count  
PE: Prediction Accuracy
Future Work

- Automatic generation of application model $\rightarrow$ Executable load tests
  - Automatic learning of guards and actions
  - Generation of protocol layer
  - Extraction and generation of input data
- Support for workload intensity $\rightarrow$ LIMBO (Kistowski et al. 2014)
- Additional transformation
  - to alternative workload generators
  - to other architecture-level performance models
    (e.g., Descartes Modeling Language) (Kounev et al. 2014)
  - from PCM to WESSBAS-DSL
- Online clustering to detect evolution of behavior mix
- Industrial case study
References

Related Work

• Several approaches for the extraction of workload specifications for session-based applications systems already exist (Arlitt et al. 2001, Krishnamurthy et al. 2006, Menascé et al. 1999) but formats are not envisaged for model-based approaches

• Approaches for the automatic extraction of performance models (Brunnert et al. 2013, Brosig et al. 2011) focus on the generation of system specific details of the SUT

• Several intermediate languages introduced to reduce the complexity of generating different kinds of performance models (Ciancone et al. 2011, Smith et al. 2004, Woodside et al. 2005) provide no support for complex workload specifications
WESSBAS Approach

Transformation: WESSBAS-DSL to Apache JMeter Test
WESSBAS Approach
Transformation into Palladio Component Models

WESSBAS-DSL

WEESBAS-DSL

app.usagemodel
Accuracy of Extracted Workload Specifications
Research Question 2

(a) Violin plot (combination of box and density plot)

(b) Summary statistics