Automatic Extraction of Probabilistic Workload Specifications for Load Testing Session-Based Application Systems

André van Hoorn, Christian Vögele

Eike Schulz, Wilhelm Hasselbring, Helmut Krcmar

VALUETOOLS 2014
Problem Statement

Situation

- Workload specification and execution essential to evaluate performance properties of session-based application systems (e.g., Krishnamurthy et al. 2006, Menascé et al. 1999, Arlitt et al. 2001)

Complication

- Manual creation of representative workload specifications is difficult, time consuming, and error-prone (Shams et al. 2006)
- Extraction and specification of workloads strongly depends on the used workload generation tool
Problem Statement

Situation

- Workload specification and execution essential to evaluate performance properties of session-based application systems (e.g., Krishnamurthy et al. 2006, Menascé et al. 1999, Arlitt et al. 2001)

Complication

- Manual creation of representative workload specifications is difficult, time consuming, and error-prone (Shams et al. 2006)
- Extraction and specification of workloads strongly depends on the used workload generation tool

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 including the clustering of navigational patterns
3) Transformation of these specifications into load test scripts
4) Tool support for this approach
WEESBAS Approach

Monitoring

SUT

Monitoring

Session Log

Behavior Model Extractor

Clustering

Behavior Mix

Behavior Models

Workload Intensity

WEESBAS-DSL Model Generator

Transformation

WEESBAS-DSL

WEESBAS-DSL-Instance

Test Plan Generator

Transformation

JMeter API

Markov4JMeter

JMeter Test Plan

Behavior Models
Related Work

- **Extraction/characterization of workload specifications** for session-based applications systems (Arlitt et al. 2001, Krishnamurthy et al. 2006)

- Extraction of **Customer Behavior Model Graphs (CBMGs)** from HTTP server logs, including **K-means clustering** to identify CBMGs for similar types of users (Menascé et al. 1999)

- A DSL-based definition of **variable and dynamic load profiles** and workload scenarios over time (e.g., Kistowski et al. 2014)

- **Intermediate languages** to reduce the complexity of generating different kinds of **performance models** (Ciancone et al. 2011, Smith et al. 2004, Woodside et al. 2005)
Agenda

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

Example of an Application Model

[van Hoorn et al. 2008]
Background – Markov4JMeter

Example of an Application Model
**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

User Session Traces

Extract

Java Code

Behavior Model Extractor

Probabilities + Think Times

Probabilities

Think Times

CSV

Behavior Mix

TXT

SUT
Clustering of Customer Groups

X-means (Pelleg, Moore 2000)

(int) Transition count matrix per session

Determine number of clusters and corresponding centroids

Assigns instances to the nearest cluster centroid using Euclidean distance (Non-normalized and normalized)

Grouping based on minimum Euclidean distance and recalculation of centroids

Instances move group?

yes

Behavior Model(s)

no

Behavior Mix

stop

Probabilities:
Each centroid represents the (relative) transition count matrix

Think Times:
Calculated based on resulting clusters (not part of clustering)

Proportion of the number of instances per cluster

Think Times:
Calculated based on resulting clusters (not part of clustering)
Transformation into Apache JMeter Test Plans
Transformation into Apache JMeter Test Plans
Agenda

- Problem Statement and Overview of Approach
- Related Work
- WESSBAS Approach
  1. WESSBAS-DSL
  2. Extraction of WESSBAS-DSL Instances
  3. Clustering of Customer Groups
  4. Transformations into Apache JMeter
- 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?

Methodology
- Instrumentation of SPECjEnterprise2010 using Kieker to obtain session logs
- Extraction of behavior models and behavior mix (includes clustering)
- Extraction of WESSBAS-DSL instances
- Transformation to JMeter test plans
  - Generation of basic application model (only session layer)
  - No input data, no guards and actions
  - Generation of dummy HTTP requests, e.g.,
    http://localhost:8080/ActionServlet?type=Add_to_Cart
- Create dummy web application with actionServlet
- Execute workload on dummy web application and measure workload with Kieker
SPECjEnterprise2010 Transactions

Transaction Mix = Behavior Mix

Adapt SPECj dealer driver: Login / logout at beginning / end of each transaction
## Accuracy of Clustering

### Research Question 1

<table>
<thead>
<tr>
<th></th>
<th>X-means (min 3 cluster, max 3 cluster)</th>
<th>X-means (min 2 cluster, max 20 cluster)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ED</td>
<td>NED</td>
</tr>
<tr>
<td>TM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 P</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>X-means (min 3 cluster, max 3 cluster)</th>
<th>X-means (min 2 cluster, max 20 cluster)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ED</td>
<td>NED</td>
</tr>
<tr>
<td>TM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 P</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TM:** Transaction Mix  
**T:** Transaction  
**Cₙ:** 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

Session-based metrics

![Session length distribution](image)

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Q1</th>
<th>Med</th>
<th>Mean</th>
<th>Close</th>
<th>Q3</th>
<th>Max</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orig</td>
<td>4</td>
<td>10</td>
<td>17</td>
<td>14.23</td>
<td>14.19</td>
<td>14.96</td>
<td>17</td>
<td>26</td>
</tr>
<tr>
<td>ED-2</td>
<td>4</td>
<td>7</td>
<td>10</td>
<td>14.24</td>
<td>14.15</td>
<td>14.43</td>
<td>18</td>
<td>147</td>
</tr>
<tr>
<td>NED-3</td>
<td>4</td>
<td>7</td>
<td>10</td>
<td>14.24</td>
<td>14.15</td>
<td>14.43</td>
<td>18</td>
<td>130</td>
</tr>
<tr>
<td>NED-4</td>
<td>4</td>
<td>7</td>
<td>10</td>
<td>14.26</td>
<td>14.17</td>
<td>14.35</td>
<td>18</td>
<td>166</td>
</tr>
</tbody>
</table>

Request-based metrics

Statistics about session-based metrics

<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>1 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>2 cancel order</td>
<td>632</td>
<td>607</td>
<td>634</td>
<td>591</td>
<td>0.00</td>
</tr>
<tr>
<td>3 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>4 deliver order</td>
<td>6,782</td>
<td>6,759</td>
<td>6,883</td>
<td>6,551</td>
<td>0.01</td>
</tr>
<tr>
<td>5 home</td>
<td>59,934</td>
<td>60,957</td>
<td>62,054</td>
<td>50,971</td>
<td>0.07</td>
</tr>
<tr>
<td>6 inventory</td>
<td>30,596</td>
<td>30,212</td>
<td>31,378</td>
<td>29,808</td>
<td>0.03</td>
</tr>
<tr>
<td>7 login</td>
<td>61,500</td>
<td>60,957</td>
<td>62,054</td>
<td>50,971</td>
<td>0.07</td>
</tr>
<tr>
<td>8 logout</td>
<td>59,934</td>
<td>60,957</td>
<td>62,054</td>
<td>50,971</td>
<td>0.07</td>
</tr>
<tr>
<td>9 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>10 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>11 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>12 shopping cart</td>
<td>9,074</td>
<td>8,934</td>
<td>9,184</td>
<td>8,907</td>
<td>0.01</td>
</tr>
<tr>
<td>13 view items</td>
<td>498,661</td>
<td>492,675</td>
<td>499,983</td>
<td>485,611</td>
<td>0.57</td>
</tr>
<tr>
<td>∑</td>
<td>874,927</td>
<td>868,089</td>
<td>883,680</td>
<td>855,159</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Future Work

- Extension for model-based performance evaluation (Vögele et. 2014)
- Automatic generation of application model → Executable load tests
  - Automatic learning of guards and actions
  - Generation of protocol layer
  - Extraction and generation of input data
- Support for workload intensity → LIMBO (Kistowski et al. 2014)
- Additional transformation to alternative workload generators
- Online clustering to detect evolution of behavior mix
- Industrial case study

Supplementary material (software, (meta-)models, data, scripts) publicly available online: http://markov4jmeter.sourceforge.net/valuetools14
References