A Unified Model-Driven Approach for Extracting and Generating Workload Specifications for Load Testing and Performance Prediction of Application Systems

Domain – Session-based Application Systems

Session: A series of consecutive and related requests issued by the same customer

(Menascé et al. 1999)
Workload specification and execution essential to evaluate performance properties of (session-based) application systems

- Measurement-based approaches (e.g., load testing)
- Model-based approaches (e.g., performance prediction)
Problem Statement

Workload specification and execution essential to evaluate performance properties of (session-based) application systems

- Measurement-based approaches (e.g., load testing)
- Model-based approaches (e.g., performance prediction)

Problems

- Manual creation (and maintenance) of **representative workload specifications** is difficult, time consuming, and error-prone
- **Workload specifications for measurement- and model-based approaches** are modelled separately of each other (M-by-N problem)
WESSBAS* Approach – Overview

Automatic Extraction of WESSBAS-DSL instances

Analysis of request logs

Model transformations

- Load tests scripts
- Workload models for performance prediction

WESSBAS-DSL

A tool- and system-agnostic (intermediate) modeling language

*WESSBAS is an acronym for Workload Extraction and Specification for Session-Based Application Systems
WESSBAS Approach – Overview

Monitoring

Production System / SUT

Monitoring

e.g., Kieker

Request Logs

- manual process

WESSBAS Activity

- automatic process

External Activity
WESSBAS Approach – Overview

Monitoring

Production System / SUT → Monitoring (e.g., Kieker) → Request Logs

Session Log Generator

Transformation → Session Log
WESSBAS Approach – Overview

Monitoring
Production System / SUT
- Monitoring
  e.g., Kieker
  Request Logs

Session Log Generator
- Transformation
  Session Log

Behavior Mix Extractor
- Extraction
  + Clustering
  Behavior Models

Workload Intensity Extractor
- Extraction
  Workload Intensity

WESSBAS-DSL Model Generator
- WESSBAS-DSL
  <<conformsTo>>
  WESSBAS DSL-Instance
  + Guards and Actions
  + Conditional Probabilities

WESSBAS Activity
- manual process
- automatic process

External Activity
Specifying Session-Based Workloads

Application Model

Session Layer

Protocol Layer

n = number of items
G = Guard
A = Action
A: n++
A: n++
G: n > 0
A: n--
G: n > 0
A: n--

login
view items
add2cart
remove

$
Specifying Session-Based Workloads
Specifying Session-Based Workloads
Specifying Session-Based Workloads
Probabilistic Representation of SPECjEnterprise2010 Workload
Transformation into Apache JMeter Test Plans
Transformation into Apache JMeter Test Plans

- WESSBAS-DSL
  - Ecore
  - M4J Workload Model
    - XML / Ecore
    - Validation
  - Transformation
    - Java Code
    - Test Plan Generator
      - JMeter Test Plan
        - XML
        - JMeter API
          - Markov4JMeter
  - CSV
    - + Think Times
  - <<conformsTo>>
  - <<uses>>
  - <<uses>>
  - <<uses>>
Transformation into Palladio Component Models

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

Generated by WESSBAS

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

WESSBAS-DSL instance

Call to the modelled system operation
RQ1: How accurately do the clustering results match the input Behavior Mix?

RQ2: What is the impact of the clustering results on the workload characteristics of the executed and predicted workload?

RQ3: How accurately do the performance characteristics of the production system/SUT match the performance characteristics using the generated and predicted workload?

RQ4: How accurately do the workload and performance characteristics match when applying different workload settings to the extracted workload?

RQ5: What is the impact of GaAs on the workload and performance characteristics?
Evaluation – Case Studies

1. SPECjEnterprise2010
   - Scenario 0: Generate benchmark load with the Faban harness (+ monitoring)
   - Scenario 1: Workload generation (+ monitoring)
   - Scenario 2: Performance prediction
   - Monitored (and predicted) measures:
     - request and session statistics
     - response times, CPU and memory utilization
SPECjEnterprise2010 Case Study Setting

- **Load (Faban)**
  - System Under Test (SUT)
    - Logging (Kieker)
      - Session Records
        - Extraction (incl. Clustering)
          - WEBSBAS Instance
            - Transformation
              - JMeter (Markov4J) Testplan
                - Load Test
                  - SUT'
        - Transformation
          - PCM Usage Model + X
            - PCM (without Usage Model)
              - Simulation
                - Simulated Workload
Evaluation – Case Studies

1. SPECjEnterprise2010
   - Scenario 0: Generate benchmark load with the Faban harness (+ monitoring)
   - Scenario 1: Workload generation (+ monitoring)
   - Scenario 2: Performance prediction
   - Monitored (and predicted) measures:
     - request and session statistics
     - response times, CPU and memory utilization

2. FIFA World Cup 1998 web server access logs
   - Scenario 1: Workload generation (+ monitoring)
   - Scenario 2: Performance Prediction
   - Measures: request and session statistics
RQ2: What is the impact of the clustering results on the workload characteristics of the executed and predicted workload?

### Selected Results for SPECjEnterprise2010

#### (a) Absolute and relative (Rel.) counts (JMeter)

<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>20,625</td>
<td>21,474</td>
<td>21,129</td>
<td>21,217</td>
<td>0.07</td>
</tr>
<tr>
<td>Cancel order</td>
<td>191</td>
<td>198</td>
<td>176</td>
<td>168</td>
<td>0.00</td>
</tr>
<tr>
<td>Clear cart</td>
<td>1932</td>
<td>2129</td>
<td>2011</td>
<td>1976</td>
<td>0.01</td>
</tr>
<tr>
<td>Defer order</td>
<td>2236</td>
<td>2228</td>
<td>2218</td>
<td>2212</td>
<td>0.01</td>
</tr>
<tr>
<td>Home</td>
<td>19,371</td>
<td>20,119</td>
<td>20,358</td>
<td>20,299</td>
<td>0.07</td>
</tr>
<tr>
<td>Inventory</td>
<td>10,034</td>
<td>10,273</td>
<td>10,136</td>
<td>10,064</td>
<td>0.04</td>
</tr>
<tr>
<td>Login</td>
<td>19,890</td>
<td>20,119</td>
<td>20,358</td>
<td>20,299</td>
<td>0.07</td>
</tr>
<tr>
<td>Logout</td>
<td>19,372</td>
<td>20,119</td>
<td>20,358</td>
<td>20,299</td>
<td>0.07</td>
</tr>
<tr>
<td>Purchase cart</td>
<td>2682</td>
<td>2780</td>
<td>2673</td>
<td>2795</td>
<td>0.01</td>
</tr>
<tr>
<td>Remove</td>
<td>923</td>
<td>660</td>
<td>675</td>
<td>723</td>
<td>0.00</td>
</tr>
<tr>
<td>Sell inventory</td>
<td>21,949</td>
<td>22,703</td>
<td>21,854</td>
<td>21,653</td>
<td>0.08</td>
</tr>
<tr>
<td>Shopping cart</td>
<td>2855</td>
<td>2789</td>
<td>2686</td>
<td>2699</td>
<td>0.01</td>
</tr>
<tr>
<td>View items</td>
<td>139,370</td>
<td>133,766</td>
<td>136,529</td>
<td>137,723</td>
<td>0.49</td>
</tr>
<tr>
<td>View items quantity</td>
<td>20,625</td>
<td>21,474</td>
<td>21,129</td>
<td>21,217</td>
<td>0.07</td>
</tr>
</tbody>
</table>

#### (b) Absolute and relative (Rel.) counts (PCM)

<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>20,625</td>
<td>22,416</td>
<td>22,466</td>
<td>21,936</td>
<td>0.07</td>
</tr>
<tr>
<td>Cancel order</td>
<td>191</td>
<td>217</td>
<td>165</td>
<td>208</td>
<td>0.00</td>
</tr>
<tr>
<td>Clear cart</td>
<td>1932</td>
<td>2094</td>
<td>2222</td>
<td>2062</td>
<td>0.01</td>
</tr>
<tr>
<td>Defer order</td>
<td>2236</td>
<td>2425</td>
<td>2379</td>
<td>2275</td>
<td>0.01</td>
</tr>
<tr>
<td>Home</td>
<td>19,371</td>
<td>21,131</td>
<td>21,190</td>
<td>20,990</td>
<td>0.07</td>
</tr>
<tr>
<td>Inventory</td>
<td>10,034</td>
<td>10,703</td>
<td>10,656</td>
<td>10,932</td>
<td>0.04</td>
</tr>
<tr>
<td>Login</td>
<td>19,890</td>
<td>21,128</td>
<td>21,190</td>
<td>20,997</td>
<td>0.07</td>
</tr>
<tr>
<td>Logout</td>
<td>19,372</td>
<td>21,128</td>
<td>21,190</td>
<td>20,997</td>
<td>0.07</td>
</tr>
<tr>
<td>Purchase cart</td>
<td>2682</td>
<td>2806</td>
<td>2919</td>
<td>2840</td>
<td>0.01</td>
</tr>
<tr>
<td>Remove</td>
<td>923</td>
<td>711</td>
<td>713</td>
<td>692</td>
<td>0.00</td>
</tr>
<tr>
<td>Sell inventory</td>
<td>21,949</td>
<td>23,867</td>
<td>23,552</td>
<td>23,807</td>
<td>0.08</td>
</tr>
<tr>
<td>Shopping cart</td>
<td>2855</td>
<td>2808</td>
<td>2939</td>
<td>2755</td>
<td>0.01</td>
</tr>
<tr>
<td>View items</td>
<td>139,370</td>
<td>146,637</td>
<td>146,903</td>
<td>148,698</td>
<td>0.49</td>
</tr>
<tr>
<td>View items quantity</td>
<td>20,625</td>
<td>22,425</td>
<td>22,472</td>
<td>21,930</td>
<td>0.07</td>
</tr>
</tbody>
</table>

#### Session length (number of requests)

<table>
<thead>
<tr>
<th>Min.</th>
<th>Q1</th>
<th>Med.</th>
<th>Mean</th>
<th>CI_{0.95}</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>14.18</td>
<td>17</td>
<td>26</td>
<td>19,890</td>
</tr>
<tr>
<td>ED-2</td>
<td>4</td>
<td>7</td>
<td>10</td>
<td>13.96</td>
<td>17</td>
<td>112</td>
<td>20,119</td>
</tr>
<tr>
<td>NED-3</td>
<td>4</td>
<td>7</td>
<td>10</td>
<td>13.88</td>
<td>17</td>
<td>107</td>
<td>20,358</td>
</tr>
<tr>
<td>NED-4</td>
<td>4</td>
<td>7</td>
<td>10</td>
<td>13.96</td>
<td>17</td>
<td>129</td>
<td>20,299</td>
</tr>
</tbody>
</table>
Selected Results for SPECjEnterprise2010

RQ3: How accurately do the performance characteristics of the production system/SUT match the performance characteristics using the generated and predicted workload?

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>$\pm$ CI_{0.95}</th>
<th>Std. dev.</th>
<th>Median</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faban (1-idle)</td>
<td>33.67</td>
<td>$\pm$ 0.22</td>
<td>0.92</td>
<td>33.62</td>
<td>72</td>
</tr>
<tr>
<td>Faban (user)</td>
<td>31.06</td>
<td>$\pm$ 0.21</td>
<td>0.89</td>
<td>31.02</td>
<td>72</td>
</tr>
<tr>
<td>JMeter (1-idle)</td>
<td>33.99</td>
<td>$\pm$ 0.38</td>
<td>1.63</td>
<td>33.66</td>
<td>72</td>
</tr>
<tr>
<td>JMeter (user)</td>
<td>31.36</td>
<td>$\pm$ 0.37</td>
<td>1.60</td>
<td>31.01</td>
<td>72</td>
</tr>
<tr>
<td>PCM</td>
<td>29.84</td>
<td>$\pm$ 0.10</td>
<td>0.41</td>
<td>29.80</td>
<td>72</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>$\pm$ CI_{0.95}</th>
<th>Std. dev.</th>
<th>Median</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faban</td>
<td>2.35</td>
<td>$\pm$ 0.15</td>
<td>0.64</td>
<td>2.35</td>
<td>72</td>
</tr>
<tr>
<td>JMeter</td>
<td>2.23</td>
<td>$\pm$ 0.16</td>
<td>0.66</td>
<td>2.23</td>
<td>72</td>
</tr>
</tbody>
</table>
Summary of Results

RQ2: What is the impact of the clustering results on the workload characteristics of the executed and predicted workload?

- The session-based characteristics, like session length and the number of distinct sessions, deviate from the measured logs in case of SPECjEnterprise2010.
- The invocation frequencies for requests match with almost 100%.

RQ3: How accurately do the performance characteristics of the production system/SUT match the performance characteristics using the generated and predicted workload?

- Performance characteristics in terms of CPU utilization, response times and heap usage are, with a few minor exceptions similar to the original executed workload.
SoSyM Article for this Talk

Christian Vögele, André van Hoorn, Eike Schulz, Wilhelm Hasselbring, and Helmut Krcmar: WESSBAS*: extraction of probabilistic workload specifications for load testing and performance prediction—a model-driven approach for session-based application systems.


Data: http://dx.doi.org/10.5281/zenodo.54859

Software: http://wessbas.github.io/

*WESSBAS is an acronym for Workload Extraction and Specification for Session-Based Application Systems
Future Work

- Extensions
  - Support for workload intensity
  - Inclusion of input data
  - Additional transformations
    - to alternative workload generators
    - to other architecture-level performance models
    - from PCM to WESSBAS-DSL
- Online clustering to detect evolution of behavior mix
- Co-evolution of manually created and generated parts in the specification
- Applications
  - Load test selection and prioritization (in continuous SE)
  - Performance regression testing and diagnosis (in continuous SE)
  - Model-driven software modernization/evolution
- Industrial case study with Sonatype (Nexus)
References

- C. Vögele, A. Brunnert, A. Danciu, D. Tertilt, H. Krcmar: Using performance models to support load testing in a large SOA environment. In Proc. LT '15, pages 5-6, 2014