fortiss

Universität Stuttgart

TECHNISCHE UNIVERSITÄT MÜNCHEN

spec RG DevOps Performance

Austin, TX, USA, 2015-02-01

Automatic Extraction of Session-Based Workload Specifications for Architecture-Level Performance Models

Christian Vögele, **André van Hoorn**, *Helmut Krcmar*

Includes material created by Eike Schulz (Kiel University)

4th International Workshop on Large-Scale Testing (LT 2015)
co-located with 6th ACM/SPEC International Conference on Performance Engineering (ICPE 2015)

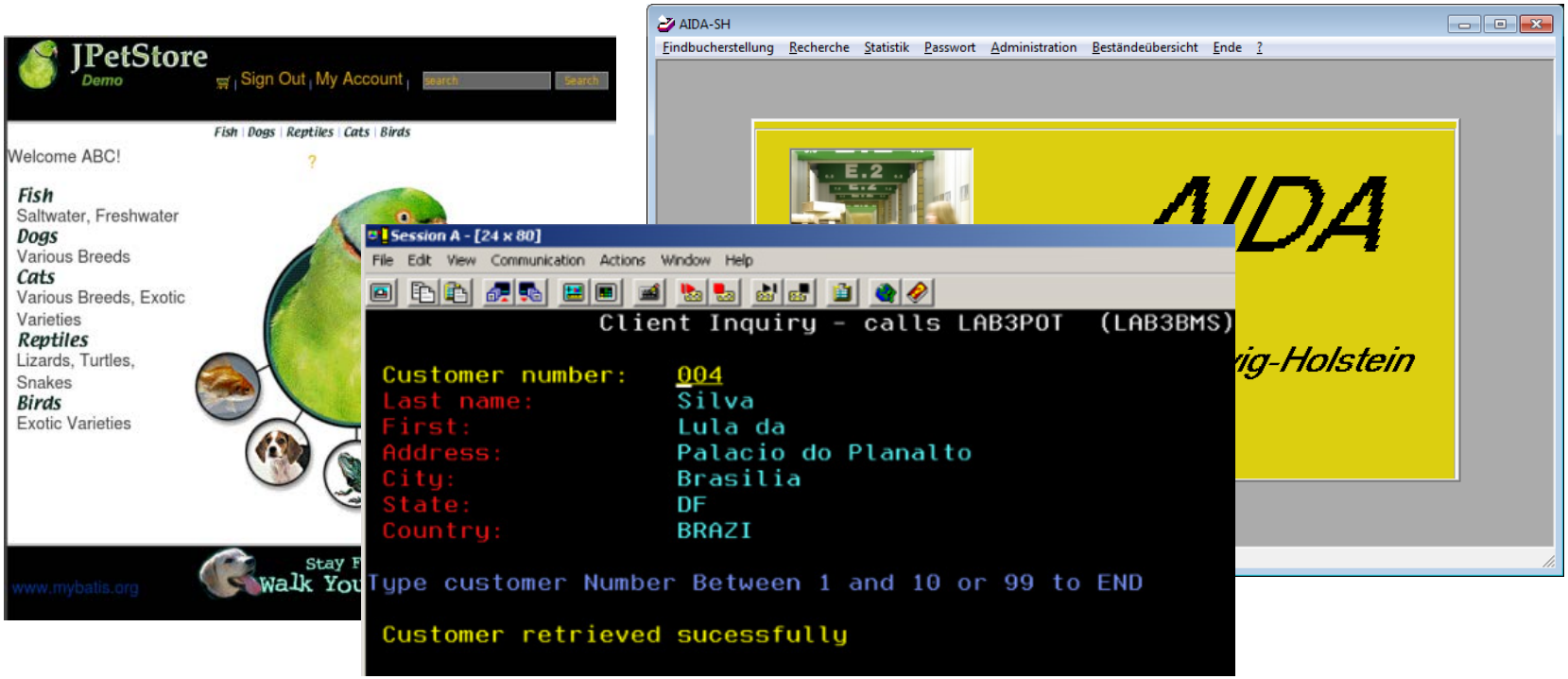
LT 2015 Large-Scale Testing

Domain

Background and Research Context

Interactive

business-critical software systems

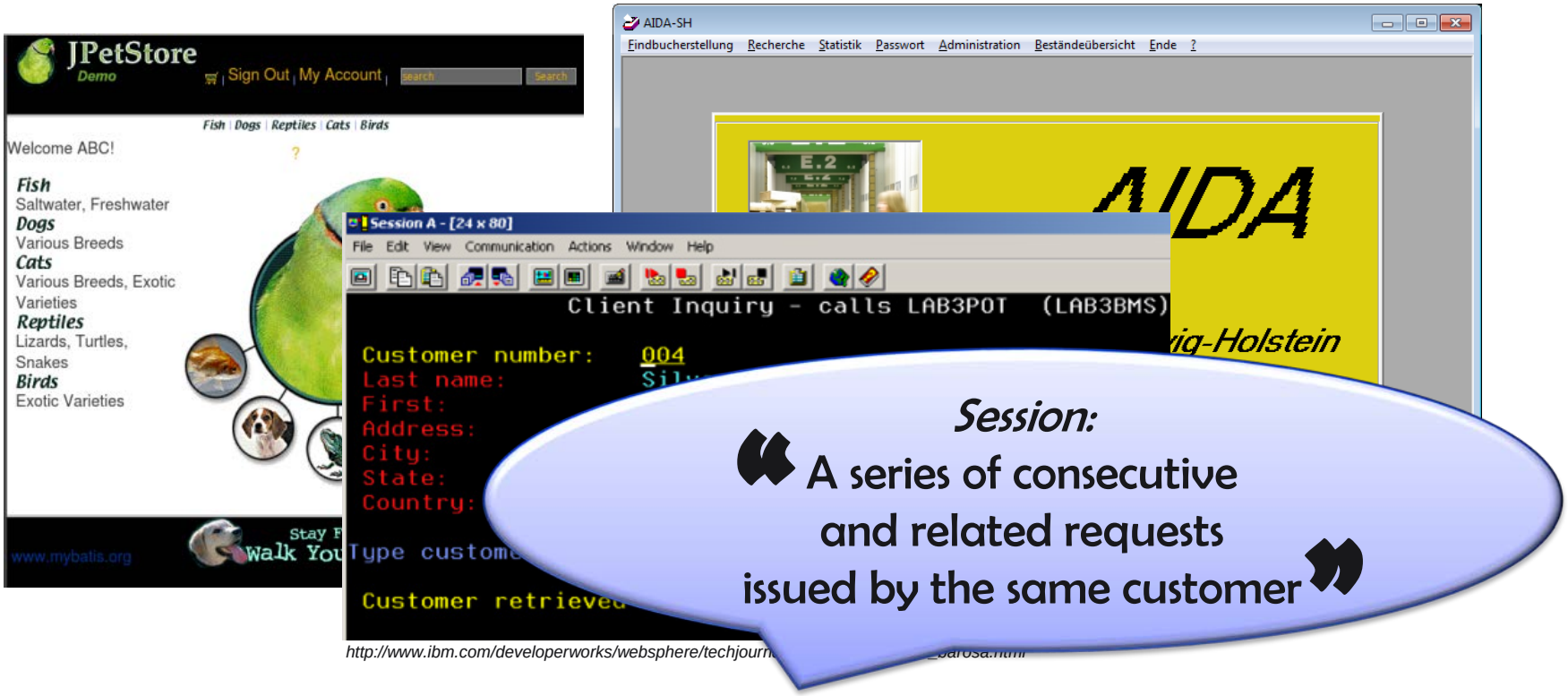


http://www.ibm.com/developerworks/websphere/techjournal/0909_barosa/0909_barosa.html

Domain

Background and Research Context

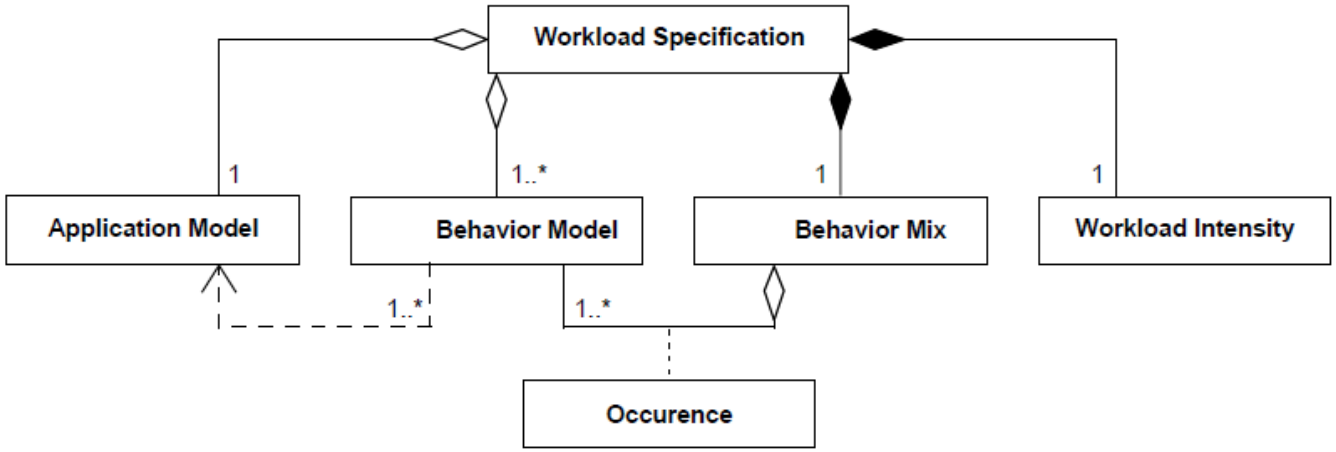
Interactive **session-based** business-critical software systems



(Menascé et al. 1999)

Specifying Session-Based Workloads

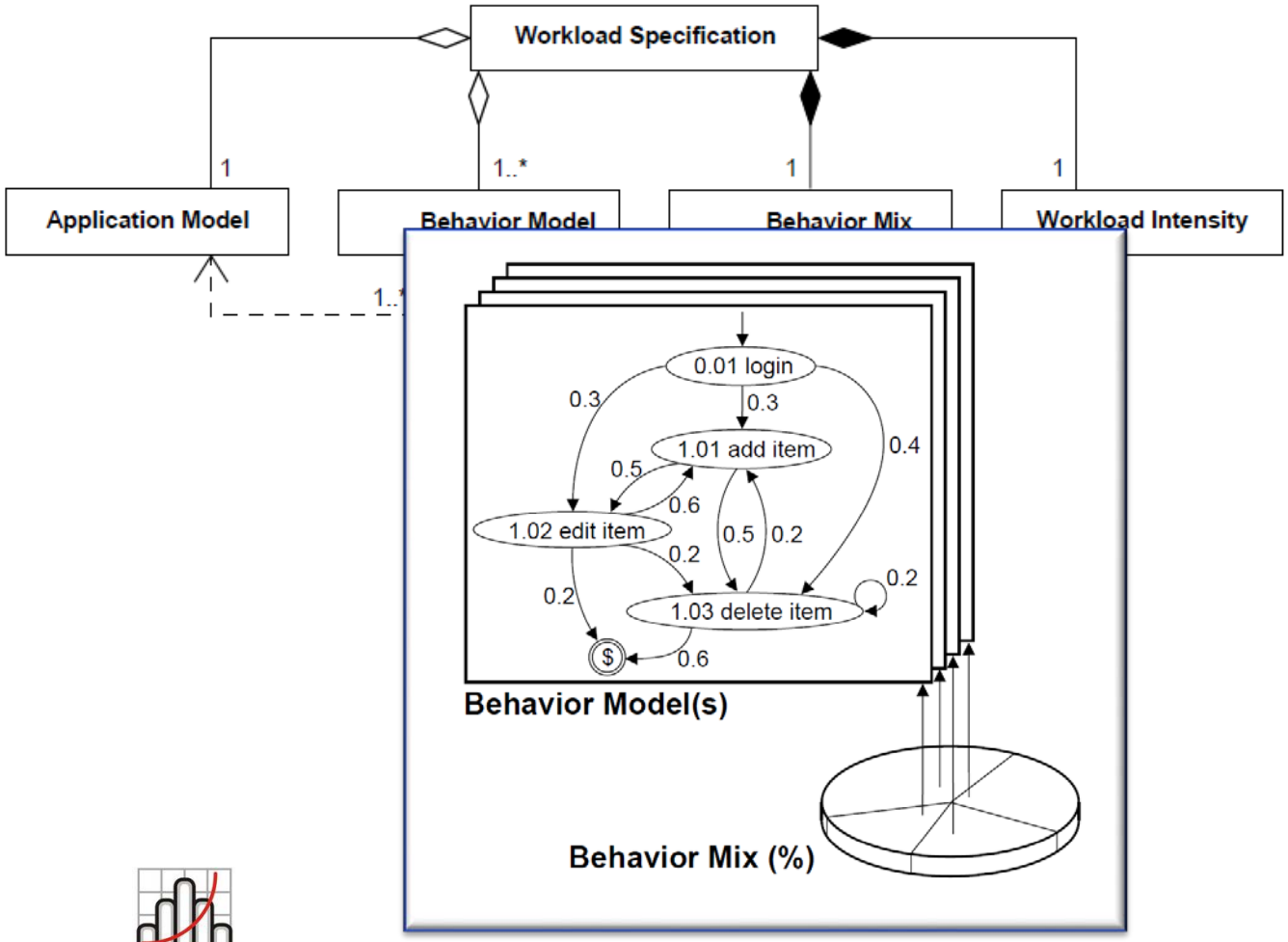
Background and Research Context



(van Hoorn et al., SIPEW 2008) – building on (Menascé et al. 1999) and (Krishnamurthy et al. 2006)

Specifying Session-Based Workloads

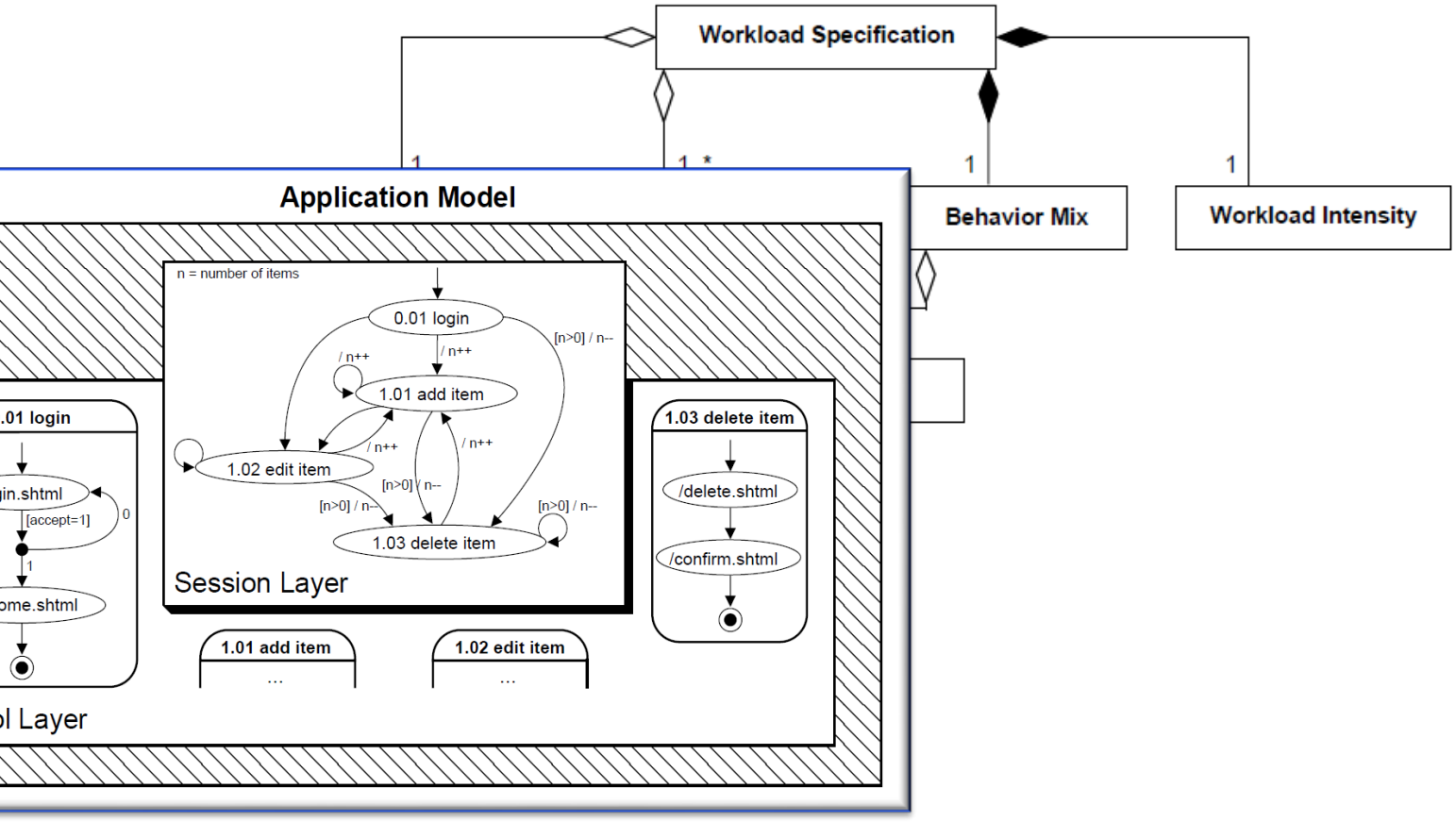
Background and Research Context



(van Hoorn et al., SIPEW 2008) – building on (Menascé et al. 1999) and (Krishnamurthy et al. 2006)

Specifying Session-Based Workloads

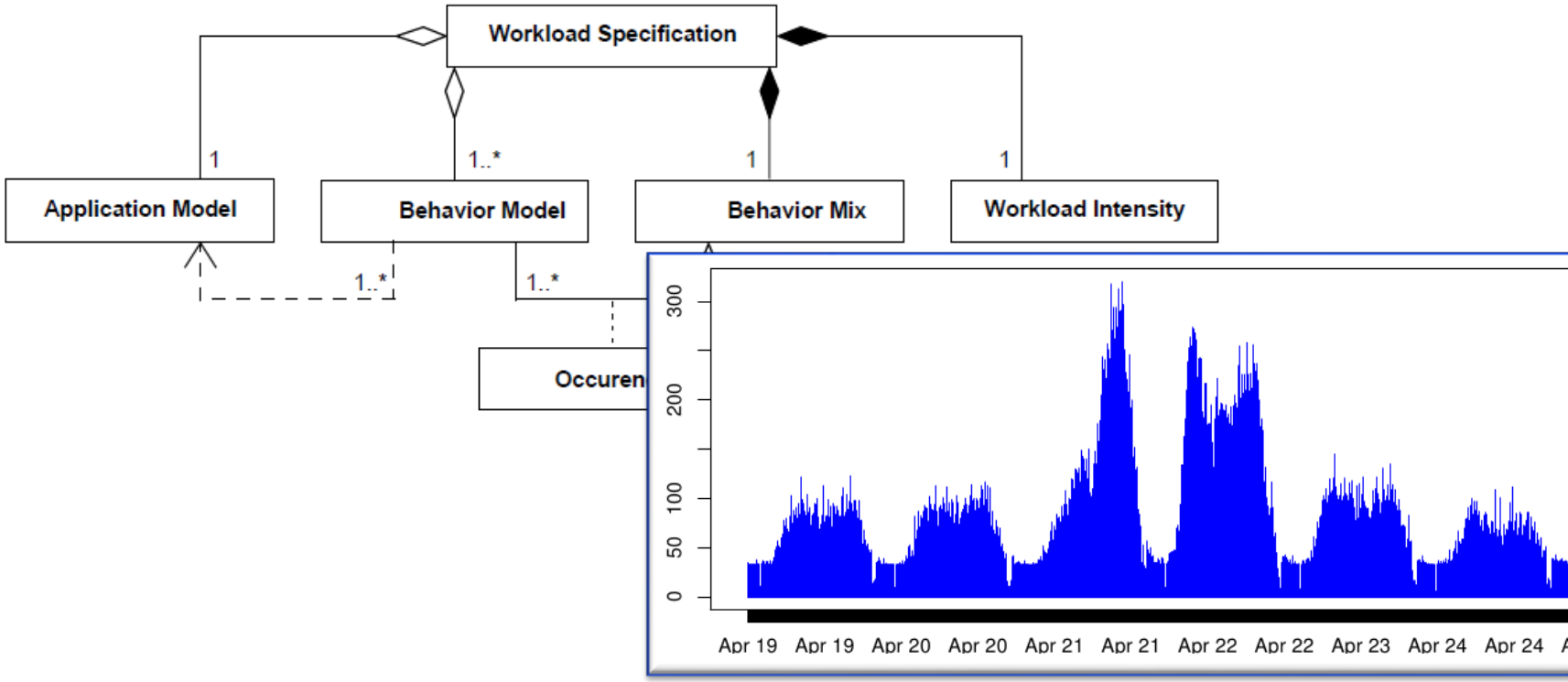
Background and Research Context



(van Hoorn et al., SIPEW 2008) – building on (Menascé et al. 1999) and (Krishnamurthy et al. 2006)

Specifying Session-Based Workloads

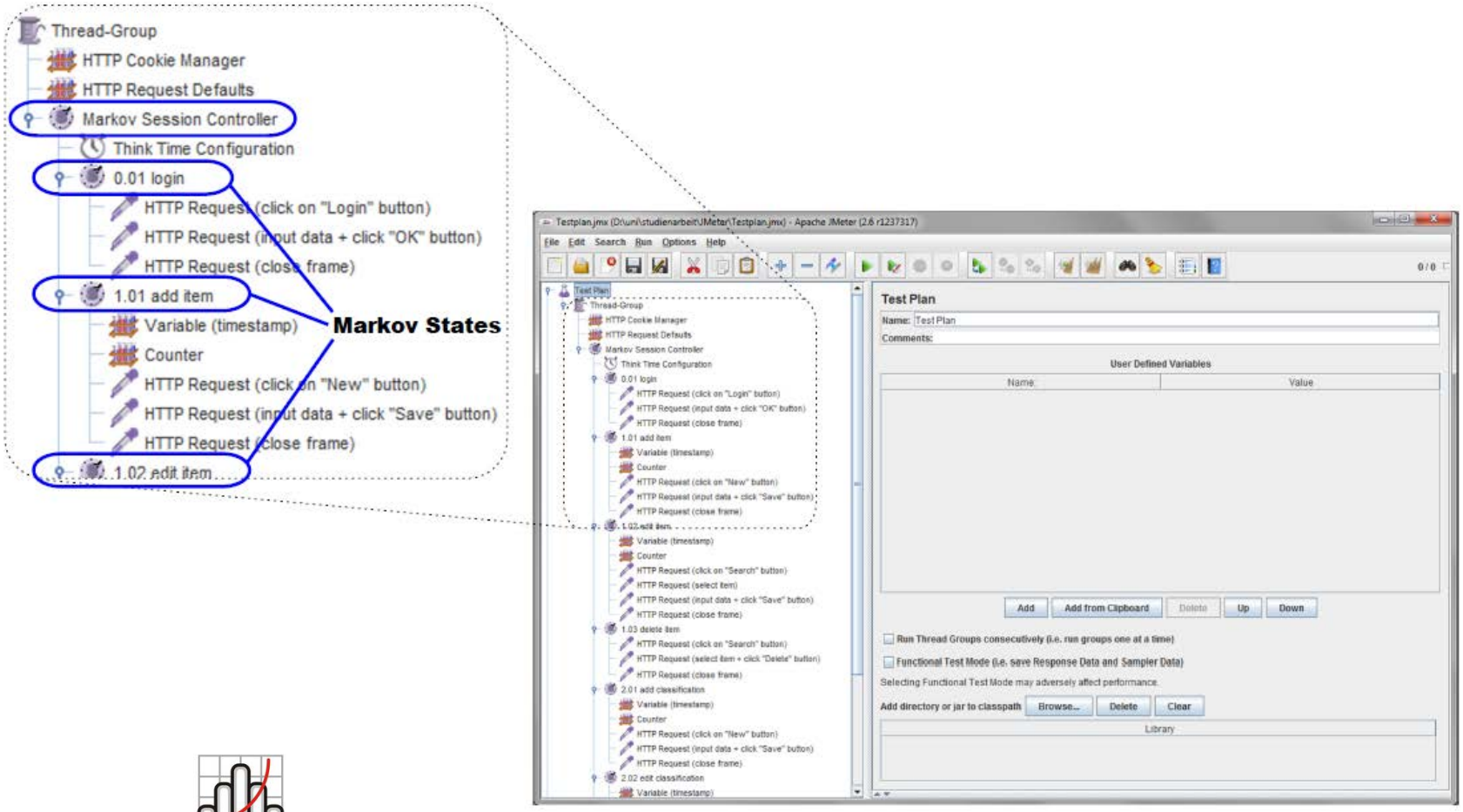
Background and Research Context



(van Hoorn et al., SIPEW 2008) – building on (Menascé et al. 1999) and (Krishnamurthy et al. 2006)

Markov4JMeter Reference Implementation

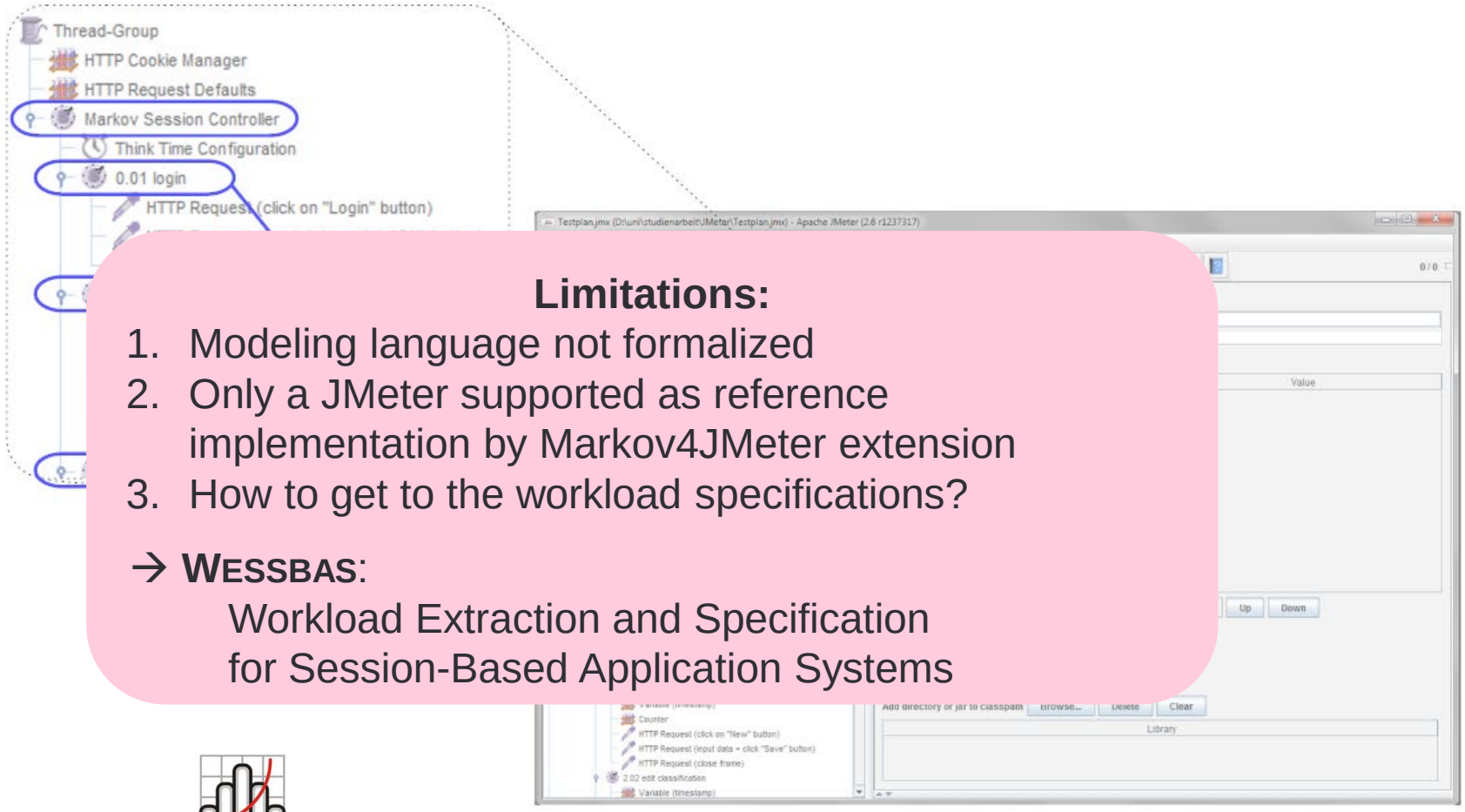
Background and Research Context



(van Hoorn et al., SIPEW 2008)

Markov4JMeter Reference Implementation

Background and Research Context



Limitations:

1. Modeling language not formalized
2. Only a JMeter supported as reference implementation by Markov4JMeter extension
3. How to get to the workload specifications?

→ **WESSBAS**:
Workload Extraction and Specification for Session-Based Application Systems

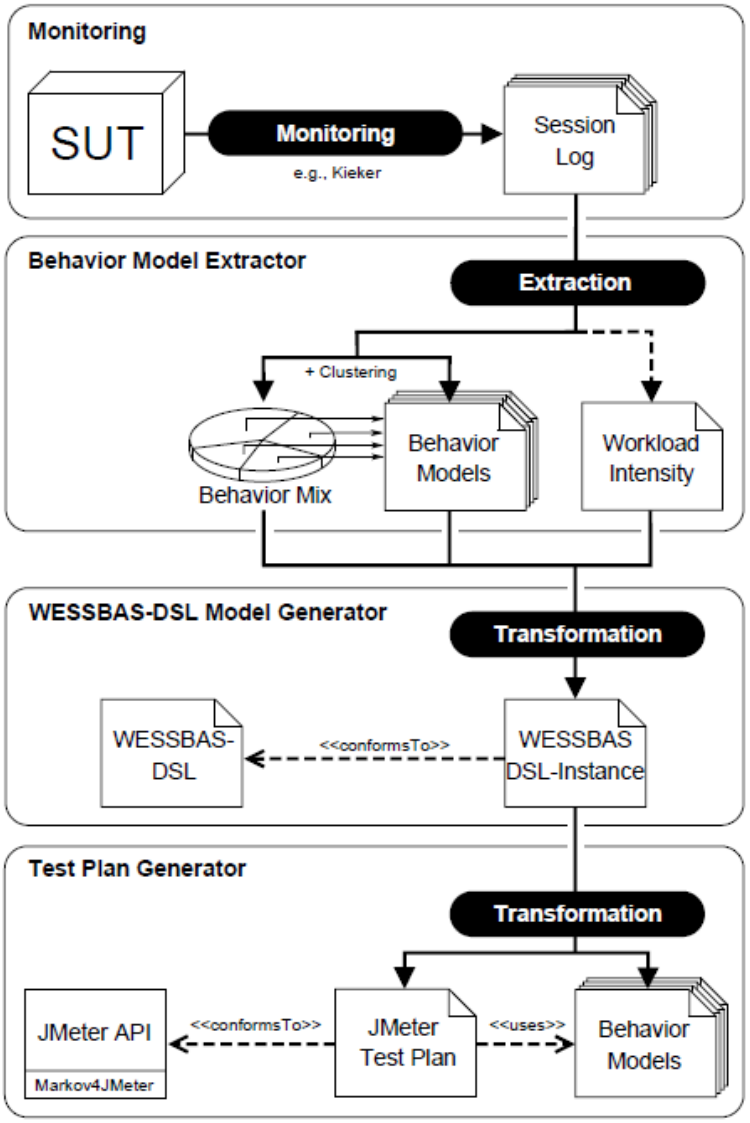
(van Hoorn et al., SIPEW 2008)

WESSBAS Approach **before** LT '15

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

Tooling

1. **WESSBAS-DSL** as tool- and system-agnostic (intermediate) modeling language
2. **Extraction of WESSBAS-DSL instances** from monitoring data (employing clustering)
3. **Transformation** from WESSBAS-DSL to JMeter/Markov4JMeter test plan
4. **Evaluation** using SPECjEnterprise

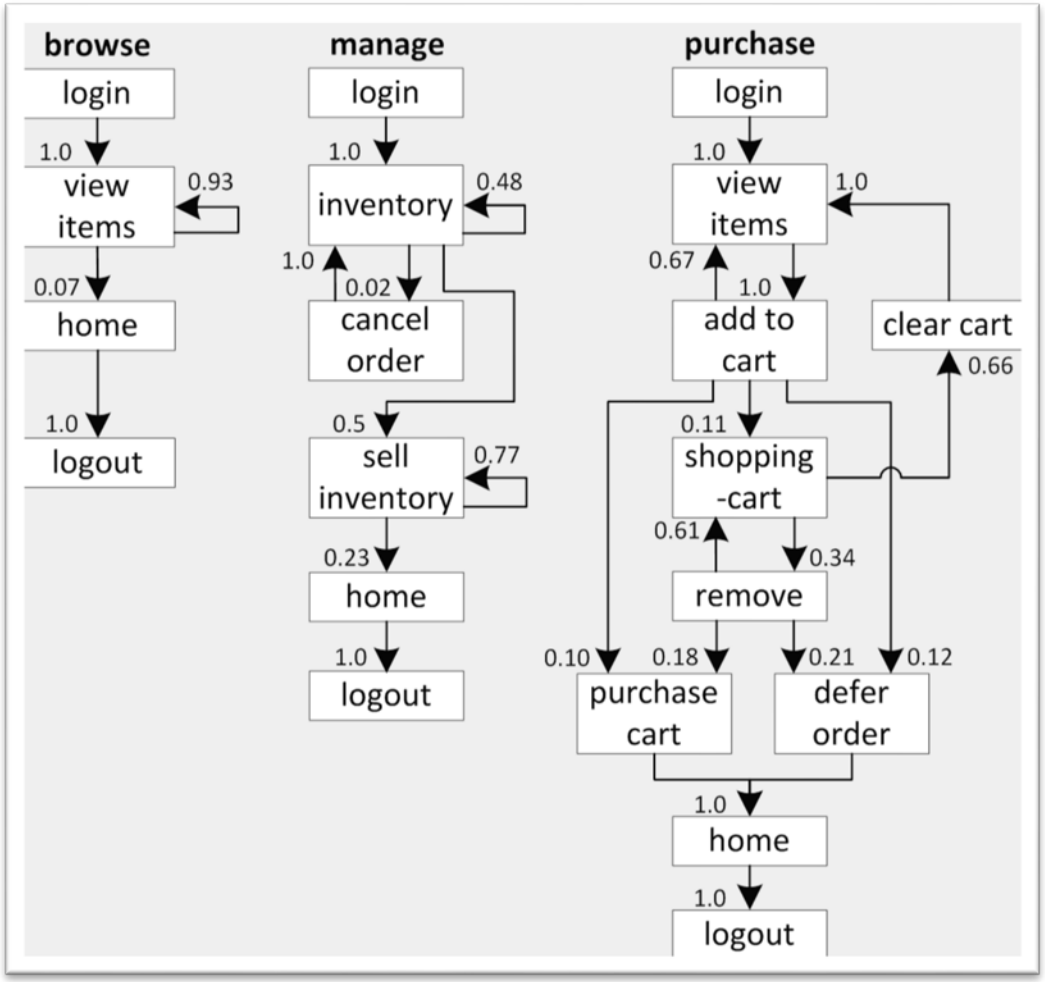


(van Hoorn et al., VALUETOOLS Performance Evaluation Methodologies and Tools 2014)

WESSBAS Approach **before** LT ‘15

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

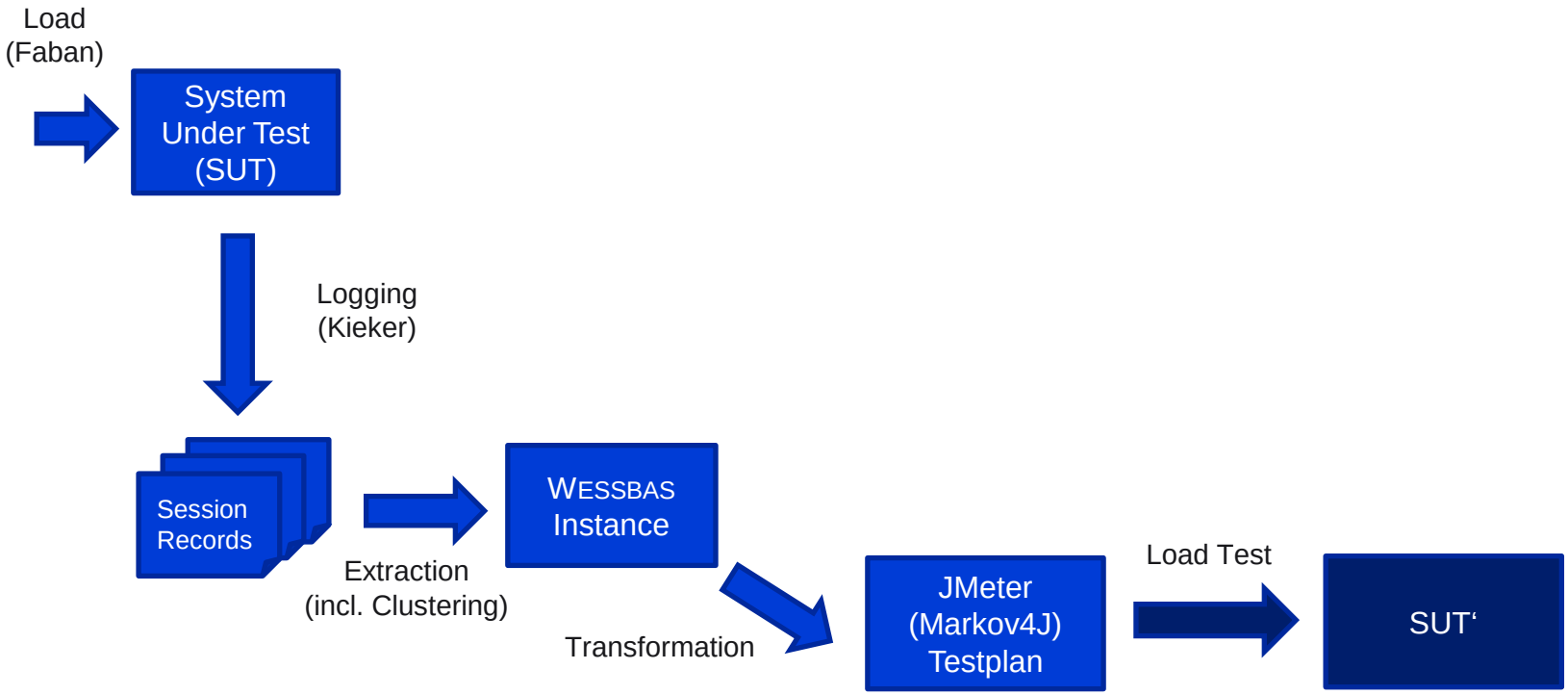
Selected Evaluation Results – SPECjEnterprise Experiments



WESSBAS Approach **before** LT '15

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

Selected Evaluation Results – SPECjEnterprise Experiments

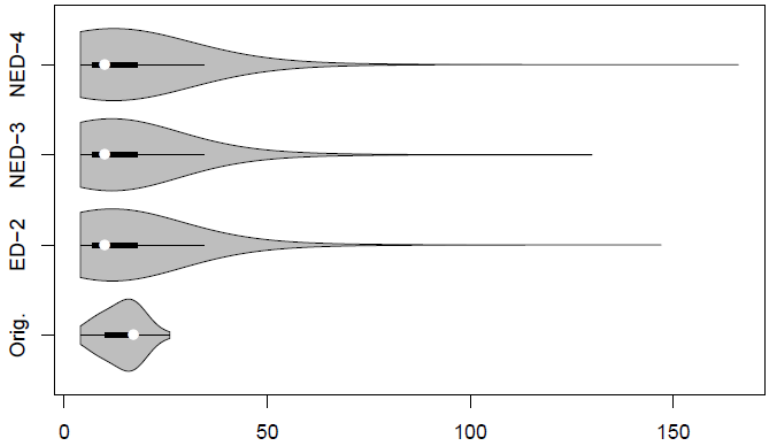


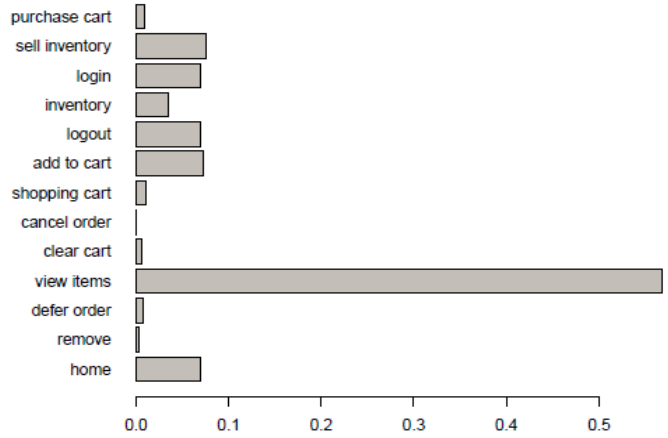
WESSBAS Approach **before** LT '15

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

Selected Evaluation Results – SPECjEnterprise Experiments

1. How accurately do the clustering results match the input Behavior Mix?
   - (Not surprisingly) Errors vary between clustering algorithms, workload mixes etc.
2. What is the impact of the clustering results on the workload characteristics?



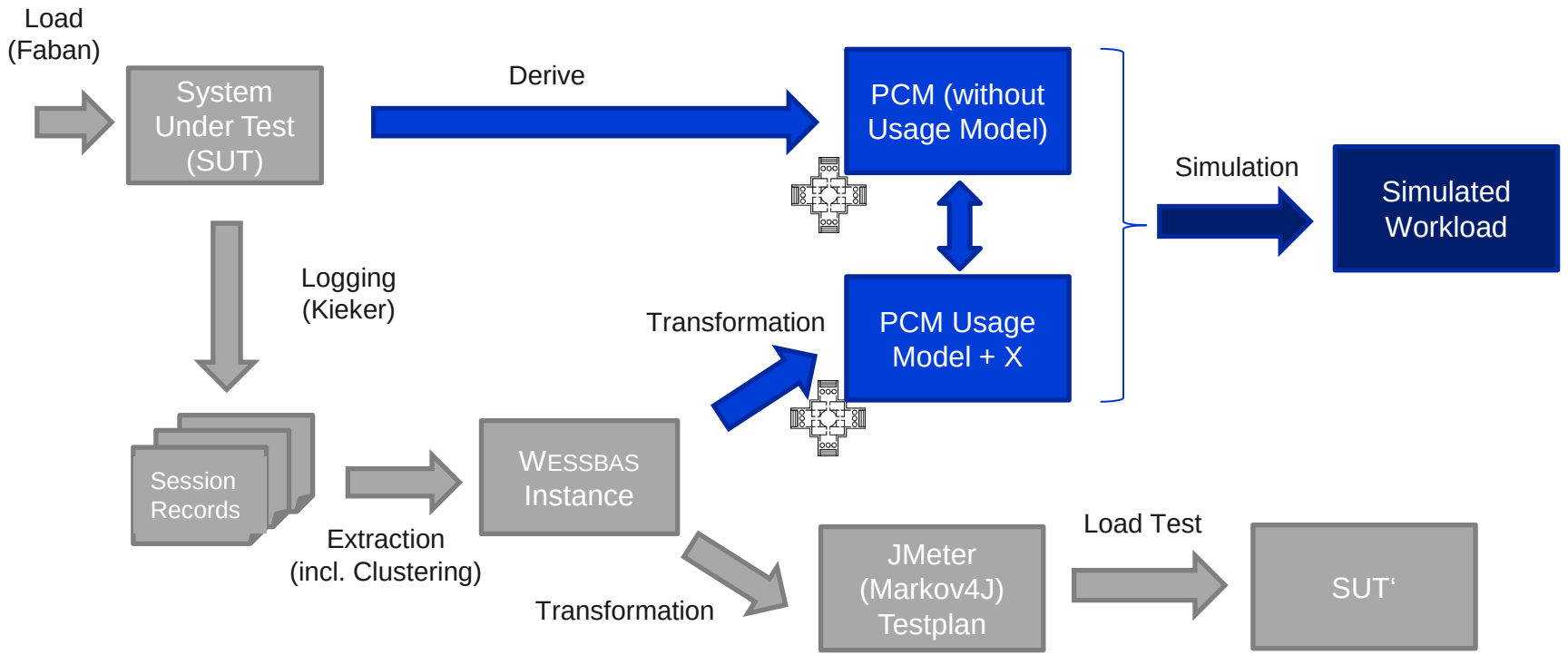


1. Workload characteristics **do not differ among each other** when using different clustering results
2. Session lengths and no. of distinct sessions **differ** from original characteristics

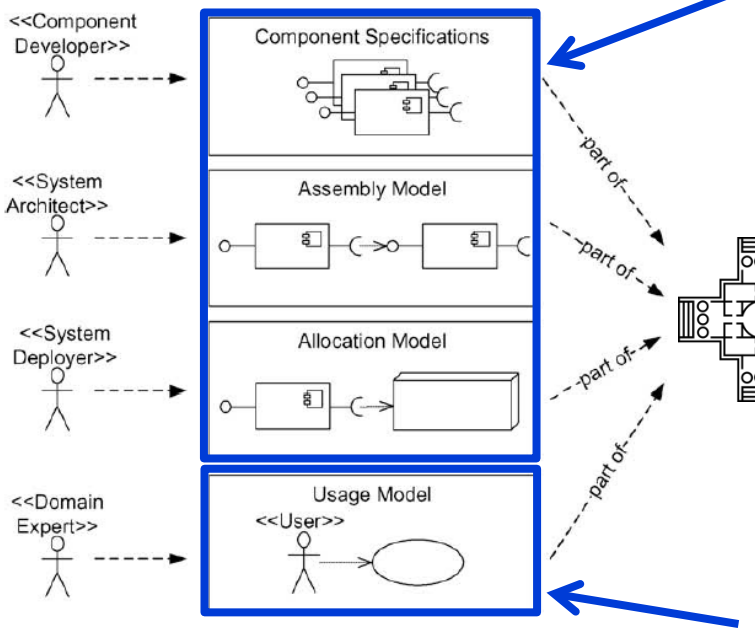
3. Server-side request counts exactly match the original characteristics

LT ‘15 Contribution: WESSBAS-DSL to PCM

- Longer term goal: Integration of workload modeling
- LT ‘15: Transformation of WESSBAS-DSL to Palladio Component Model (PCM)



Transformation into Palladio Component Models



Becker et al. (2009)

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

app.allocation
app.repository
app.resourceenvironment
app.system
app.usagemodel

Missing possibility to reference usage models from usage models

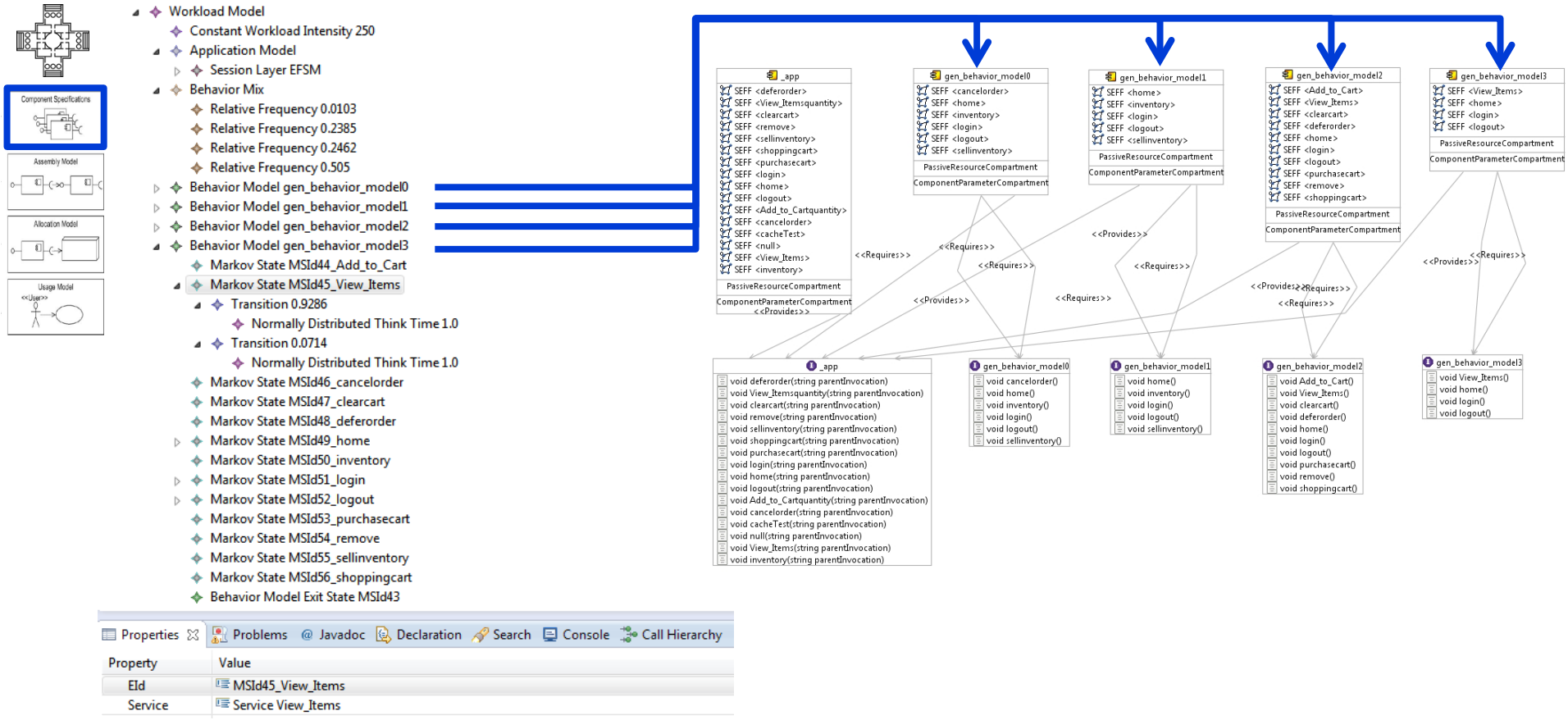
Generation of large parts of the workload specification into the repository model

WESSBAS-DSL	*PCM Model Elements*
Behavior Models	Repository Model (Basic Component, RDSEFF)
Session Layer FSMs	not required
Protocol Layer FSMs	not required
Workload Intensity	Usage Model (Closed Workload)
Behavior Mix	Usage Model (Branch)

Generation into PCM Repository Model

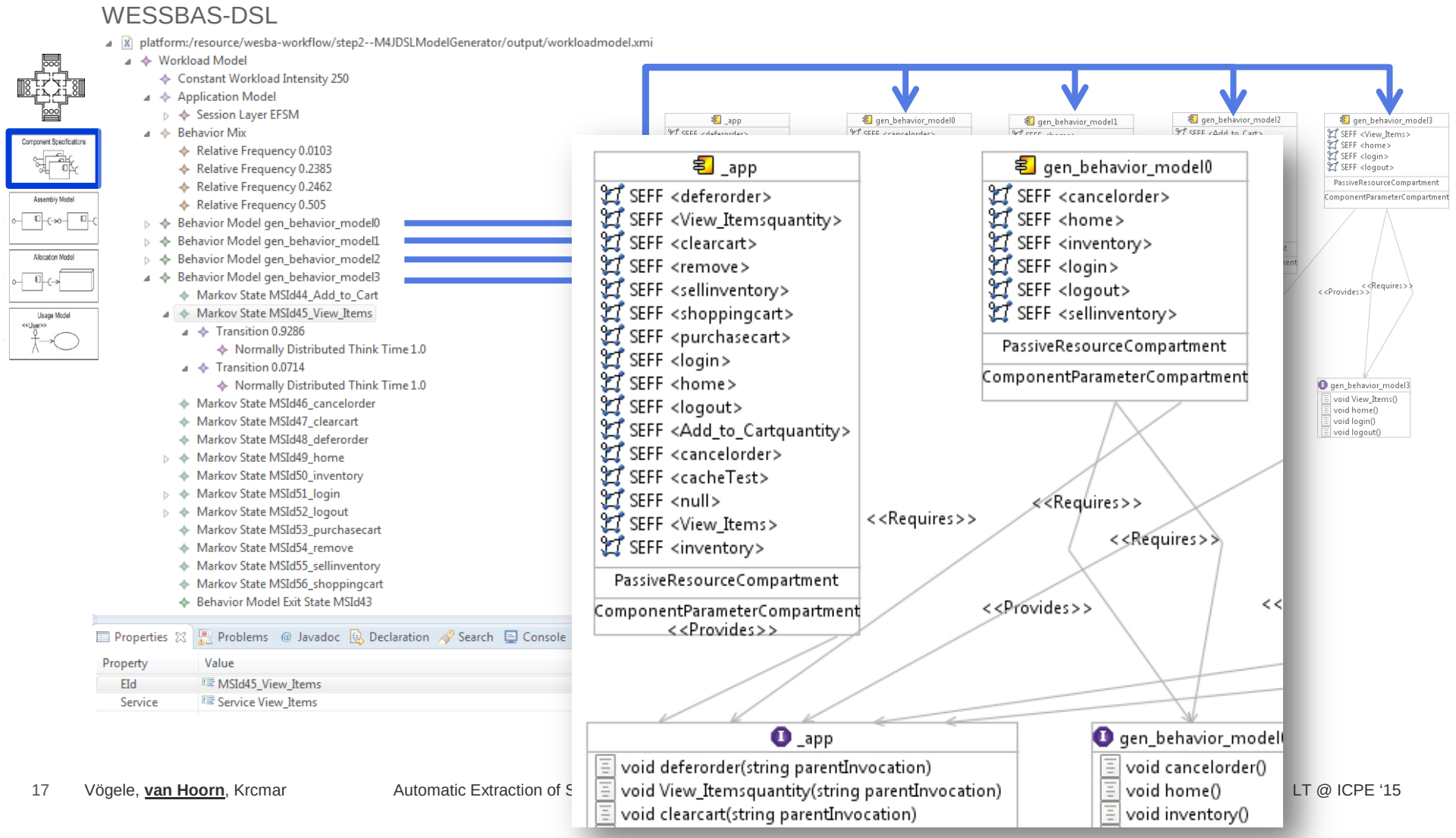
Transformation into Palladio Component Models

WESSBAS-DSL



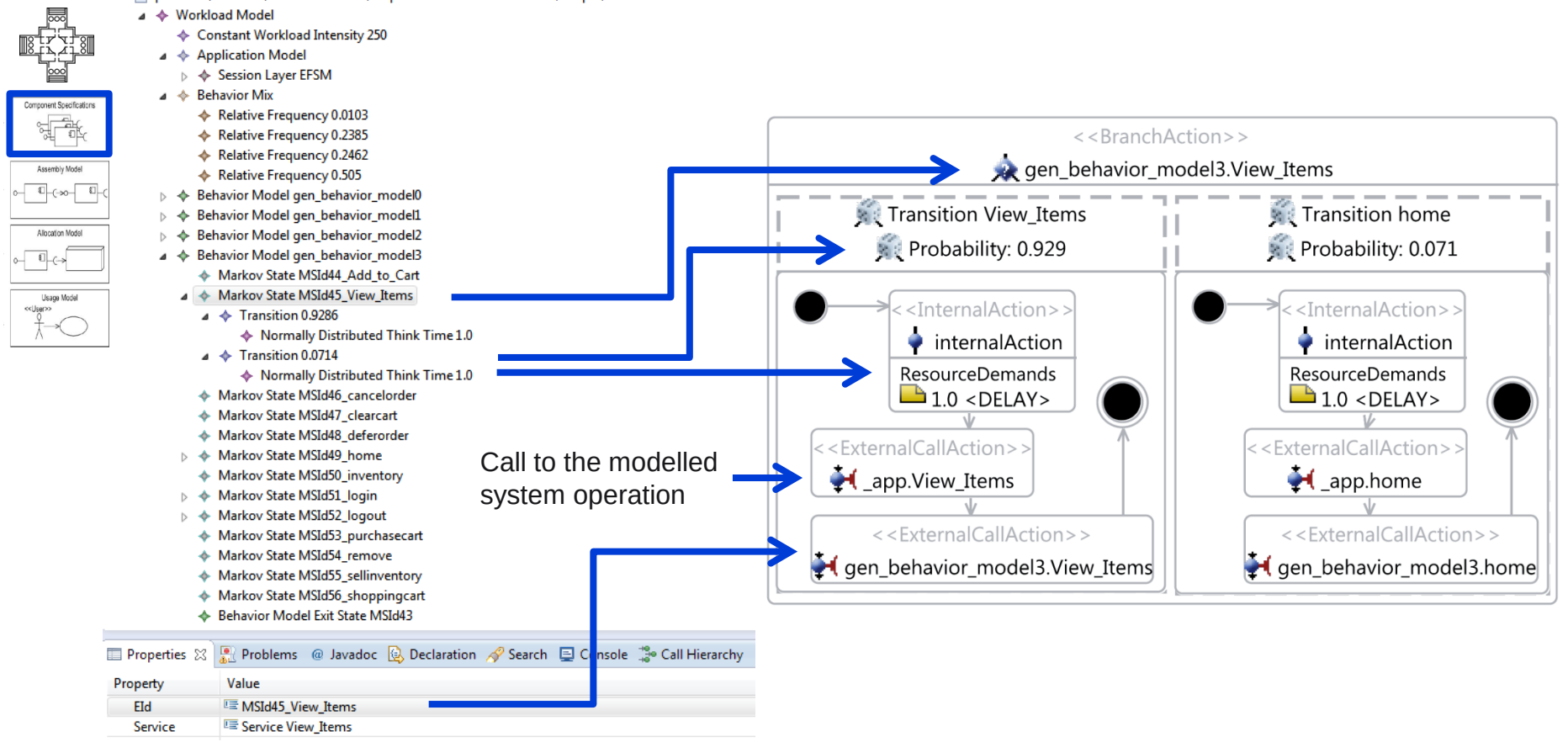
Generation into PCM Repository Model

Transformation into Palladio Component Models



Generation into PCM Repository Model

Transformation into Palladio Component Models



Accuracy of PCM Workload Specification?

Evaluation

- Question: How well match the predicted workload characteristics the measured workload characteristics?
- Experimental setting:
  - Transformation and simulation of our VALUETOOLS SPECjEnterprise instances
  - Comparison of measured and predicted request counts

		Orig.	2 Behavior Models		3 Behavior Models		4 Behavior Models	
	Request	MRC	SRC	PE	SRC	PE	SRC	PE
1	add to cart	63,761	64,943	1.82%	61,812	3.15%	60,986	4.55%
2	cancel order	632	609	3.78%	661	4.39%	625	1.12%
3	clear cart	6,047	6,178	2.12%	5,927	2.02%	5,846	3.44%
4	defer order	6,782	6,873	1.32%	6,524	3.95%	6,606	2.66%
5	home	59,934	61,146	1.98%	58,747	2.02%	58,744	2.03%
6	inventory	30,596	30,539	0.19%	29,574	3.46%	29,405	4.05%
7	login	61,500	61,156	0.56%	58,747	4.69%	58,745	4.69%
8	logout	59,934	61,146	1.98%	58,747	2.02%	58,744	2.03%
9	purchase cart	8,360	8,388	0.33%	7,976	4.81%	7,836	6.69%
10	remove	3,027	2,986	1.37%	2,876	5.25%	2,949	2.64%
11	sell inventory	66,679	66,131	0.83%	63,185	5.53%	63,914	4.33%
12	shopping cart	9,074	9,164	0.98%	8,803	3.08%	8,795	3.17%
13	view items	498,601	491,812	1.38%	470,392	6.00%	475,000	4.97%
	$\sum$	874,927	871,071	0.44%	833,971	4.91%	838,195	4.38%

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

Future Work

- Automatic generation of application model → Executable load tests
  - Automatic learning of guards and actions
  - Generation of protocol layer
  - Modeling, extraction and generation of parameters
- Support for workload intensity → LIMBO (Kistowski et al. 2014)
- 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
- Industrial case study

➢ Supplementary material (software, (meta-)models, data, scripts) publicly available online: http://markov4jmeter.sourceforge.net/lt15

Statements as Input for the LT ’15 Discussion

1. Thought-provoking statement or discussion question about the area
   (e.g., how could this work be validated?)

- What values/use cases do you see in integrating workload modeling and extraction for measurement- and model-based performance evaluation?

2. Thought-provoking statement or discussion question about the area
   (e.g., how can this work be of value to industry).

- How can we transfer state-of-the-art load modeling/testing approaches do industrial practice?
- Do you see a need to improve load testing practice?

References

- M. F. Arlitt, D. Krishnamurthy, and J. Rolia. Characterizing the scalability of a large web-based shopping system. ACM TOIT, 1(1):44-69, 2001.
- A. Ciancone, A. Filieri, M. Drago, R. Mirandola, and V. Grassi. Klapersuite: An integrated model-driven environment for reliability and performance analysis of component-based systems. In Objects, Models, Components, Patterns, volume 6705 of LNCS, pages 99-114. Springer Berlin Heidelberg, 2011.
- A. van Hoorn, M. Rohr, and W. Hasselbring. Generating probabilistic and intensity-varying workload for Web-based software systems. In Proc. SIPEW '08, pages 124-143, 2008.
- A. van Hoorn, C. Vögele, E. Schulz, W. Hasselbring, and H. Krcmar. Automatic extraction of probabilistic workload specifications for load testing session-based application systems. VALUETOOLS, 2014.
- A. van Hoorn, J. Waller, W. Hasselbring: Kieker: A Framework for Application Performance Monitoring and Dynamic Software Analysis. In Proc. ACM/SPEC ICPE '12, ACM, 2012, 247-248
- J. v. Kistowski, N. R. Herbst, and S. Kounev. Modeling variations in load intensity over time. In Proc. LT '14, pages 1-4. ACM, 2014.
- D. Krishnamurthy, J. A. Rolia, and S. Majumdar. A synthetic workload generation technique for stress testing session-based systems. IEEE TSE, 32(11):868-882, 2006.
- D. A. Menascé, V. A. F. Almeida, R. Fonseca, and M. A. Mendes. A methodology for workload characterization of e-commerce sites. In Proc. EC '99, pages 119-128, 1999.
- D. Pelleg, A. W. Moore. X-means: Extending K-means with ecient estimation of the number of clusters. In ICML '00, pages 727-734, 2000.
- M. Shams, D. Krishnamurthy, and B. Far. A model-based approach for testing the performance of web applications. In Proc. SOQUA 2006, pages 54-61. ACM, 2006.
- C. Smith and C. Llado. Performance model interchange format (pmif 2.0): Xml defintition and implementation. In Proc. QEST '04, pages 38-47, Sept 2004.
- 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
- M. Woodside, D. C. Petriu, D. B. Petriu, H. Shen, T. Israr, and J. Merseguer. Performance by unified model analysis (puma). In Proc. WOSP '05, pages 1-12, New York, NY, USA, 2005. ACM.