Bratislava, Slovakia, 2014-12-10

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

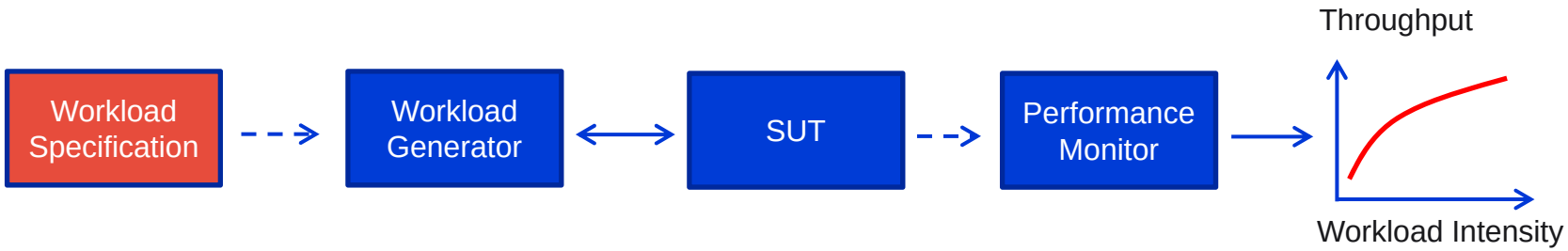
Problem Statement

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)

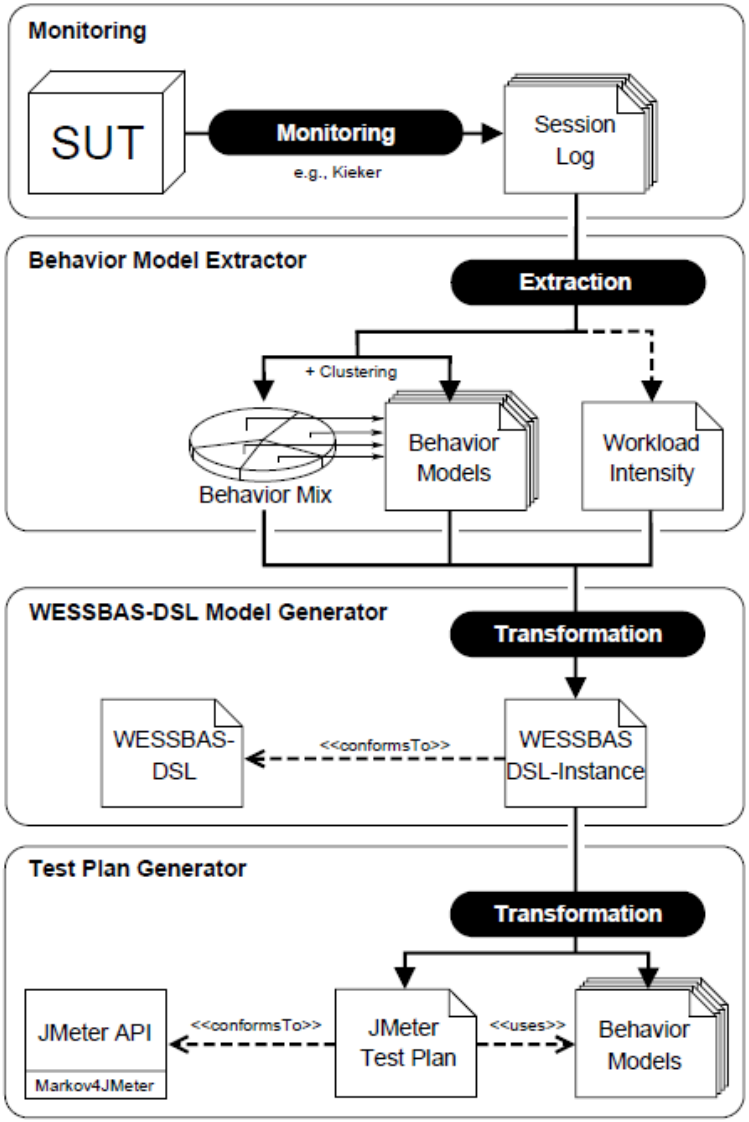
Complication

- 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

WESSBAS Approach



Related Work

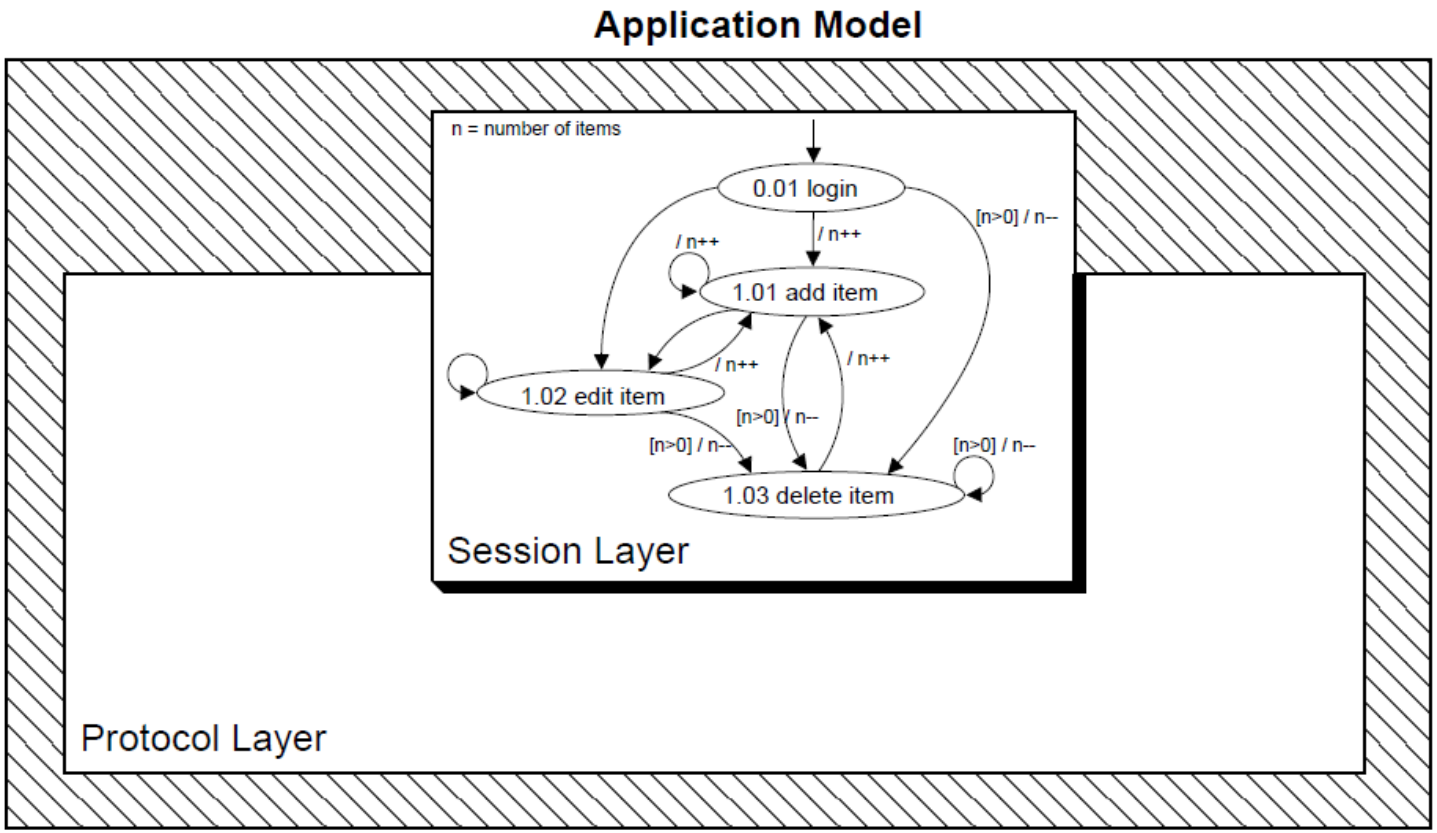
- Extraction/characterization of workload specifications for session-based applications systems (Arlitt et al. 2001, Krishnamurthy et al. 2006)
- Extraction of Customer Behavior Model Graphs (CBMGs) from HTTP server logs, including K-means clustering to identify CBMGs for similar types of users (Menascé 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

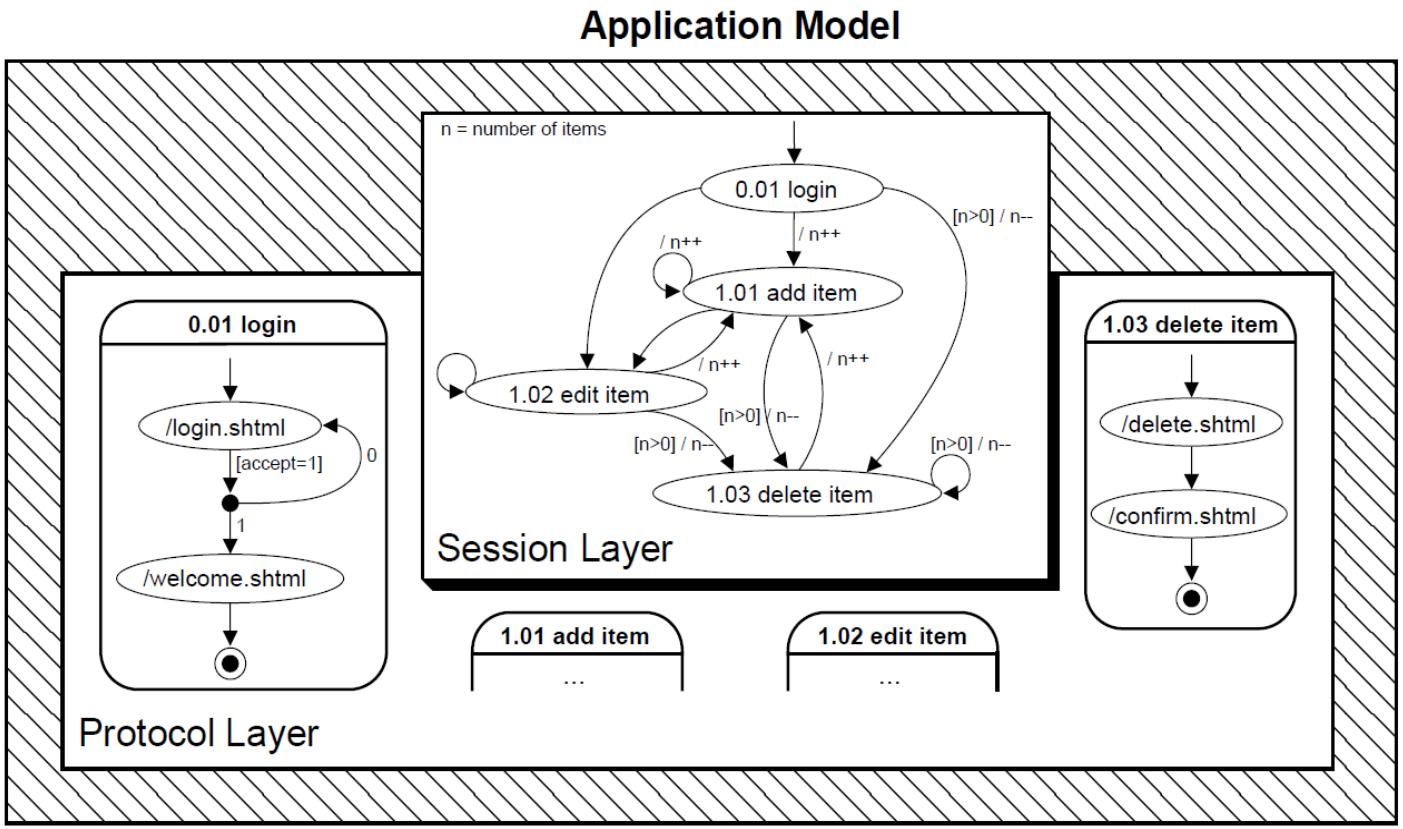
Example of an Application Model



[van Hoorn et al. 2008]

Background – Markov4JMeter

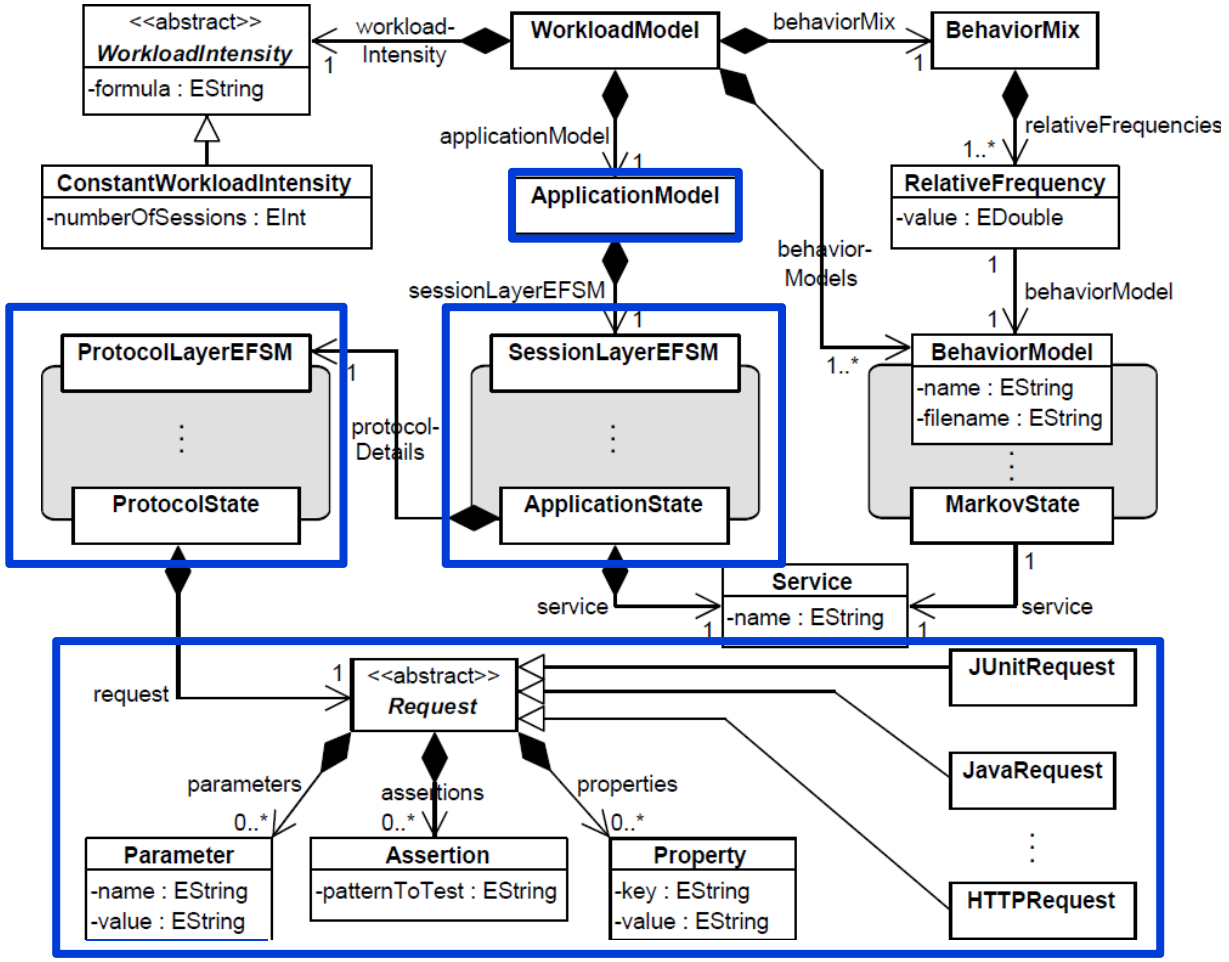
Example of an Application Model



[van Hoorn et al. 2008]

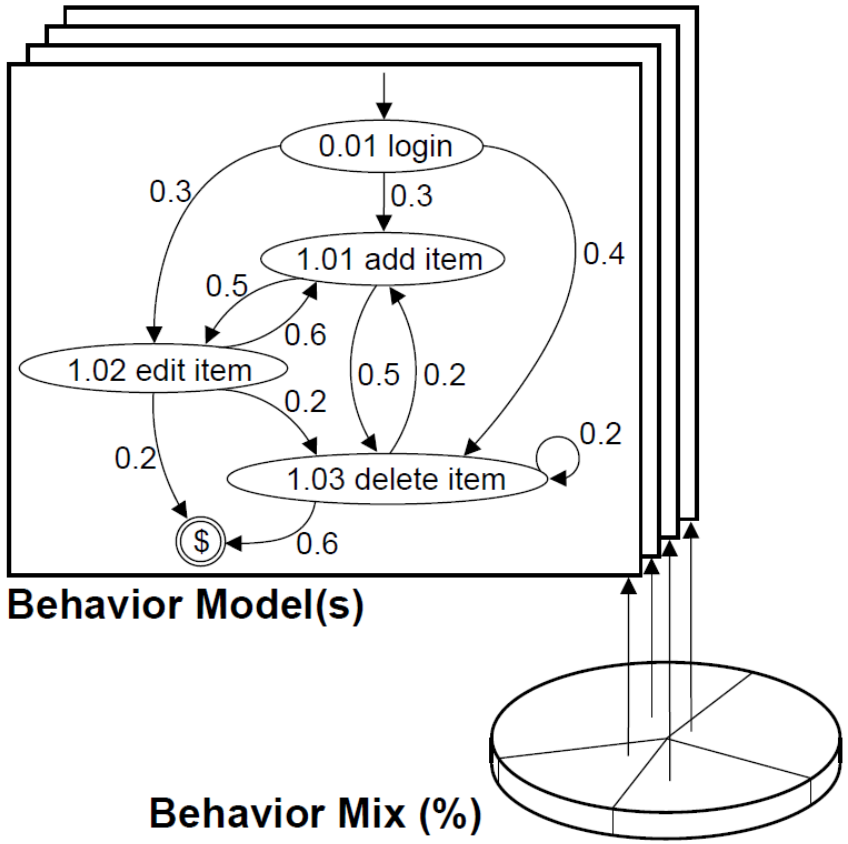
WESSBAS-DSL

Application Model



Behavior Model + Behavior Mix

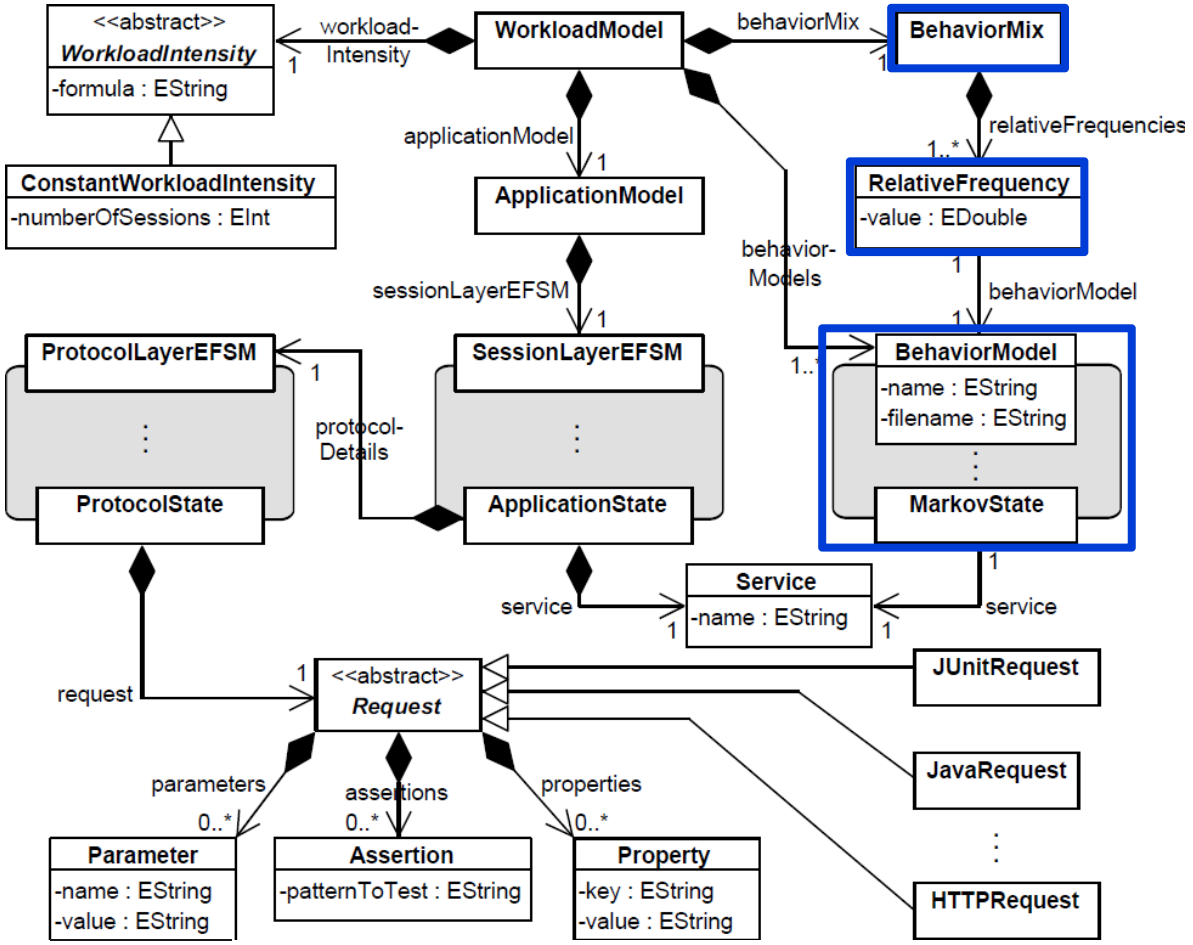
Background – Markov4JMeter



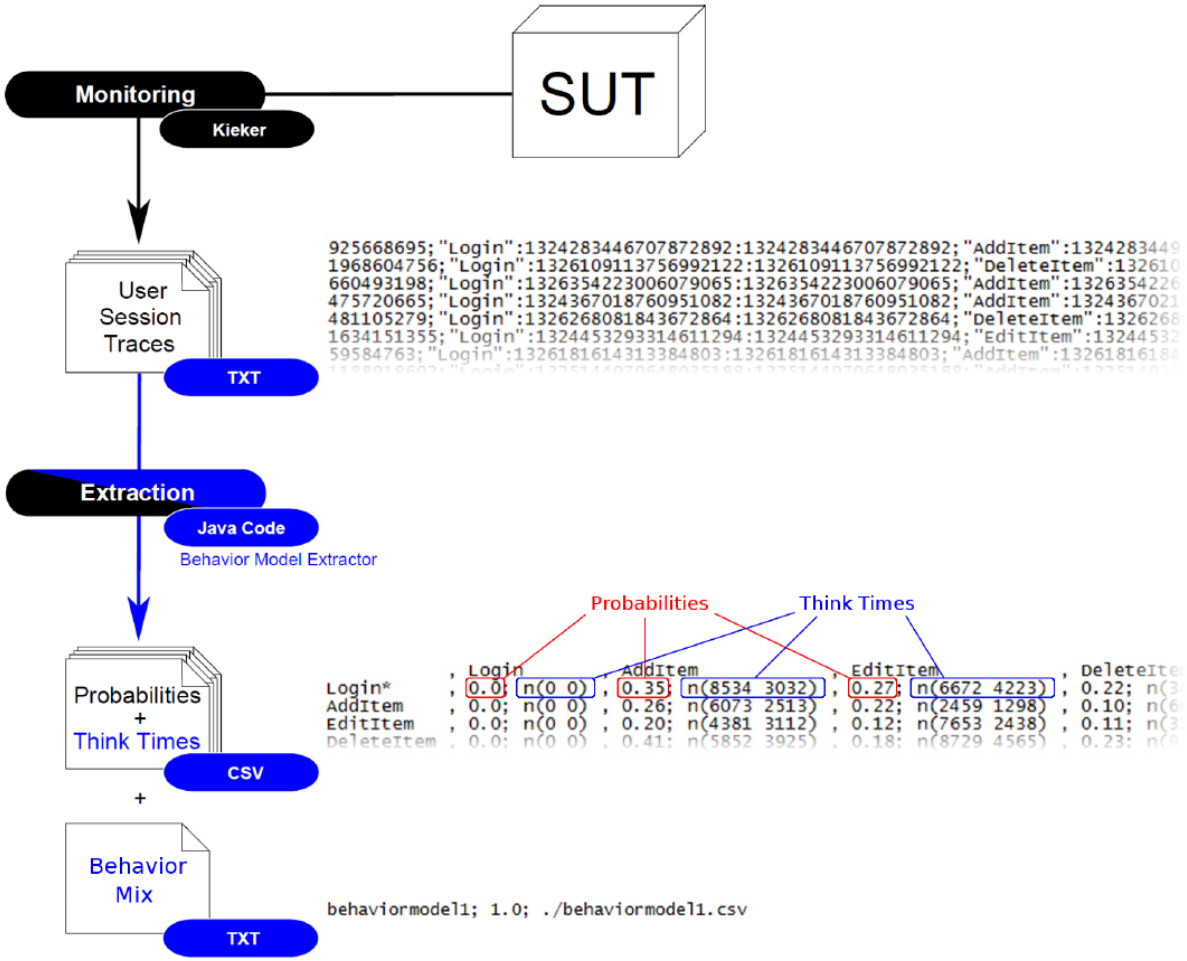
[van Hoorn et al. 2008]

WESSBAS-DSL

Behavior Mix and Behavior Model(s)

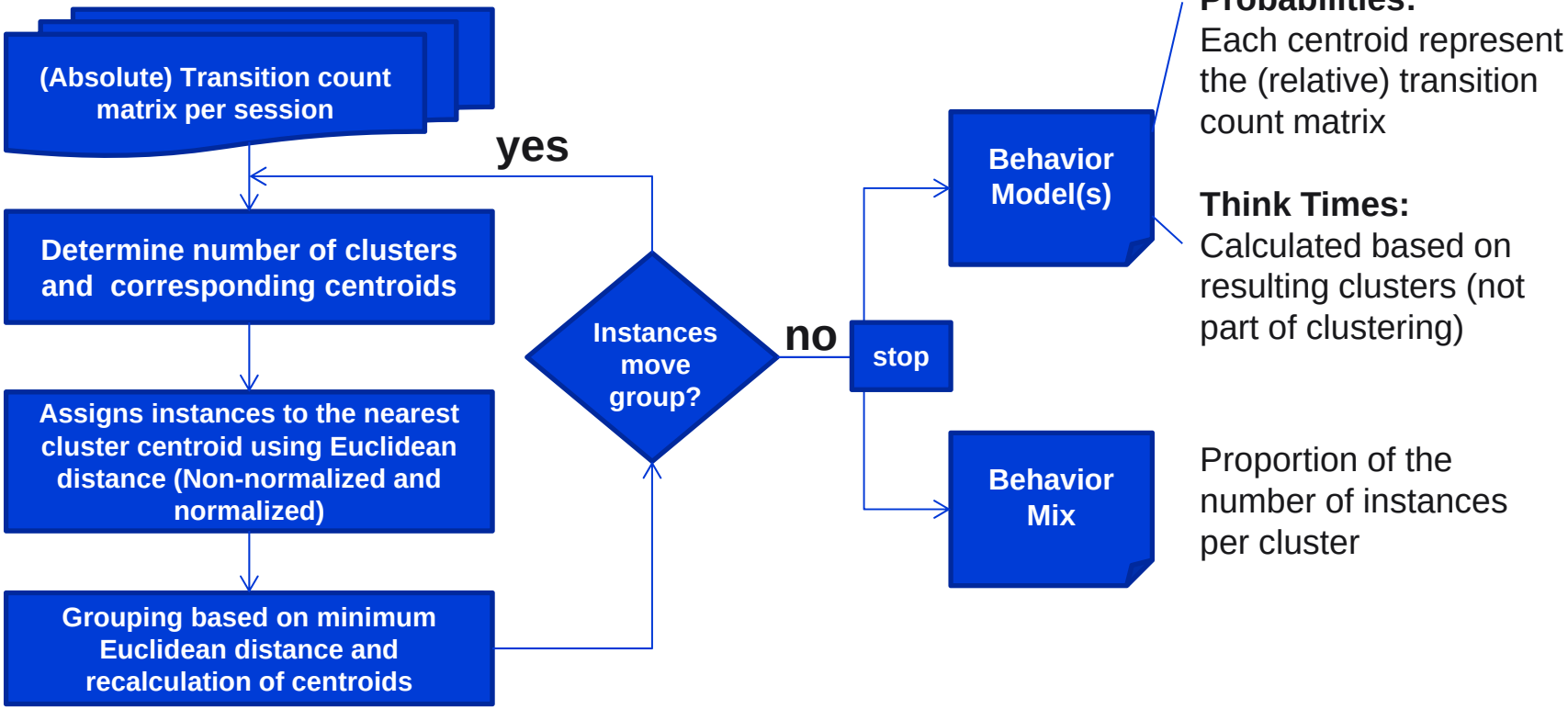


Behavior Model Extraction

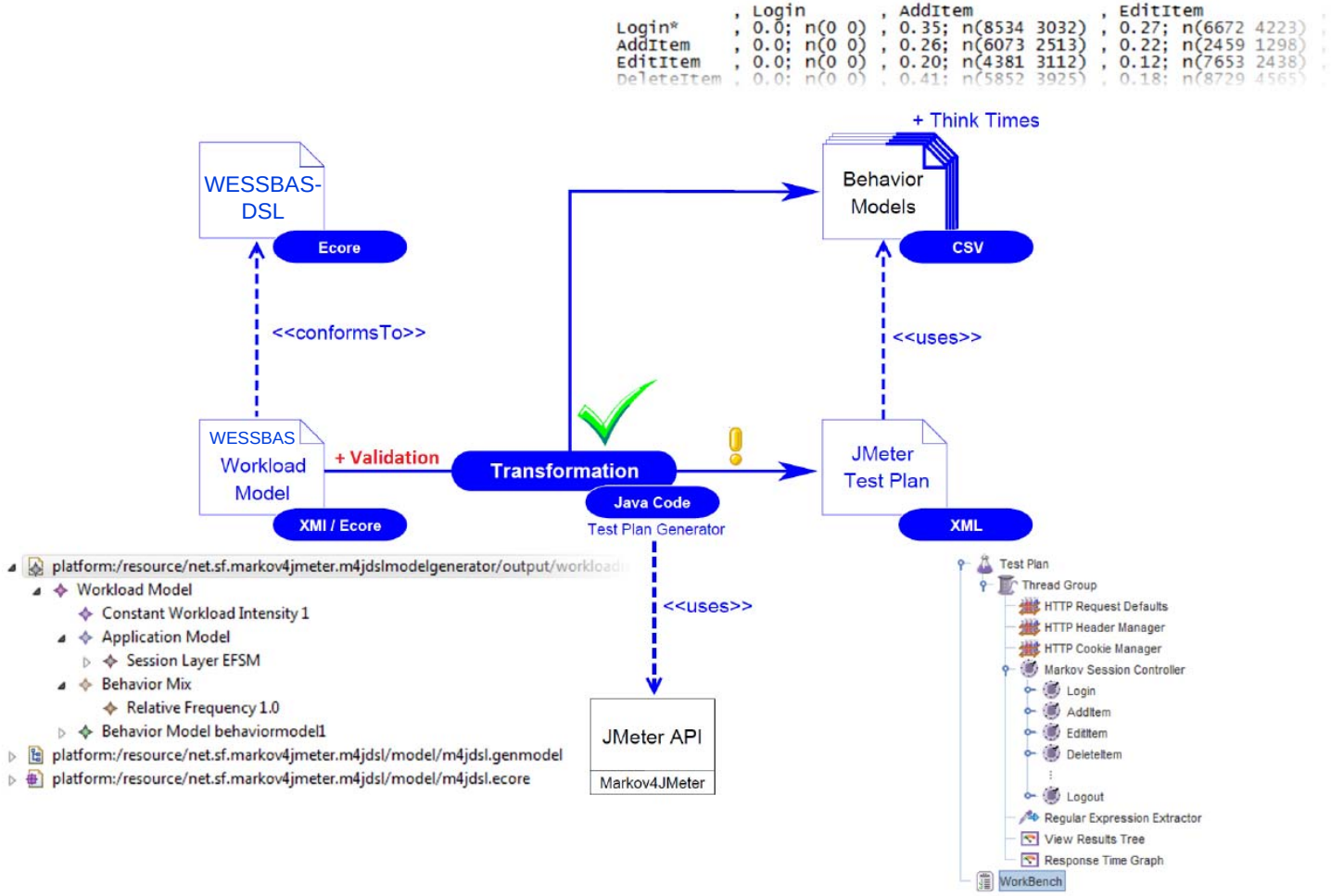


Clustering of Customer Groups

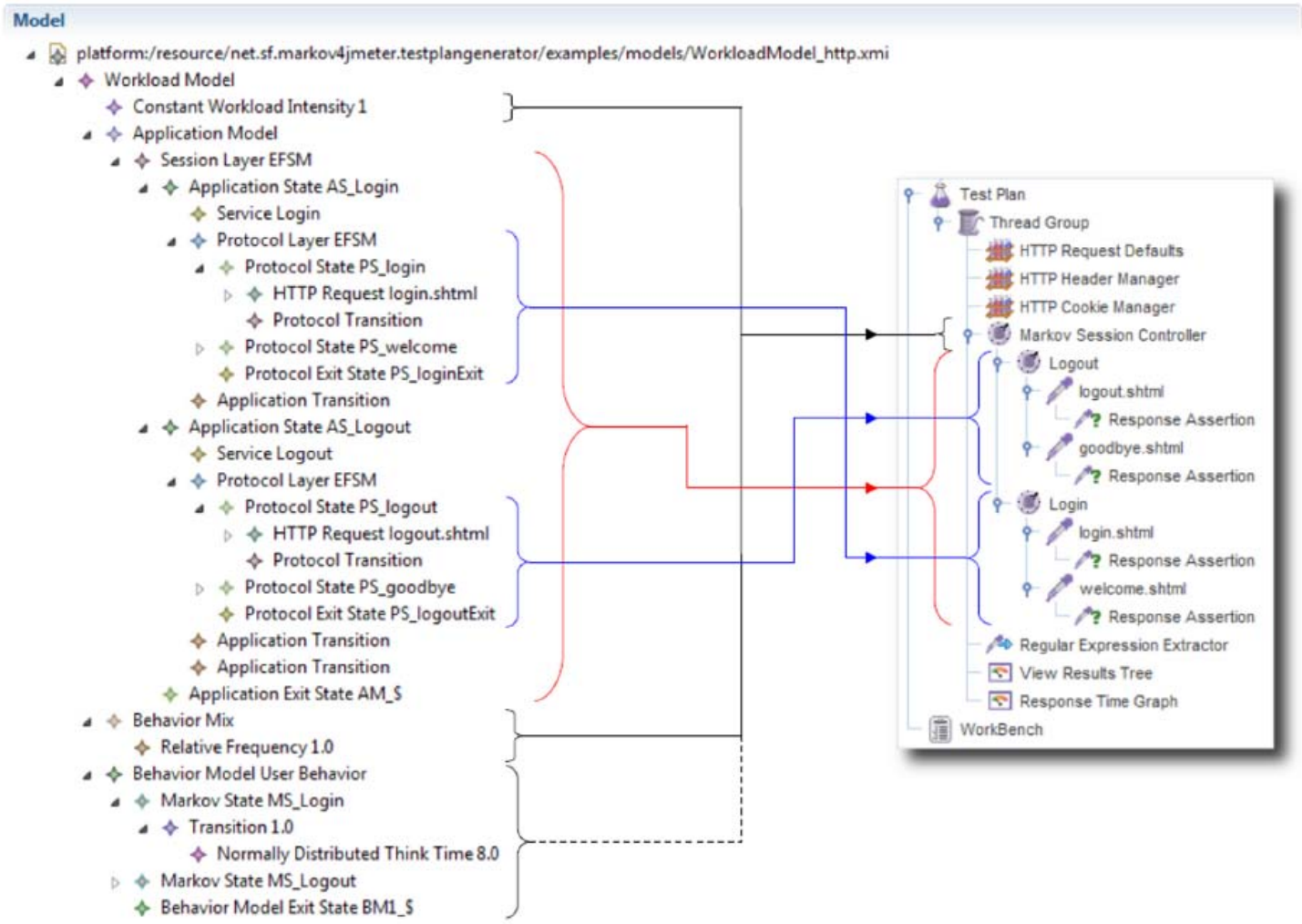
X-means (Pelleg, Moore 2000)



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

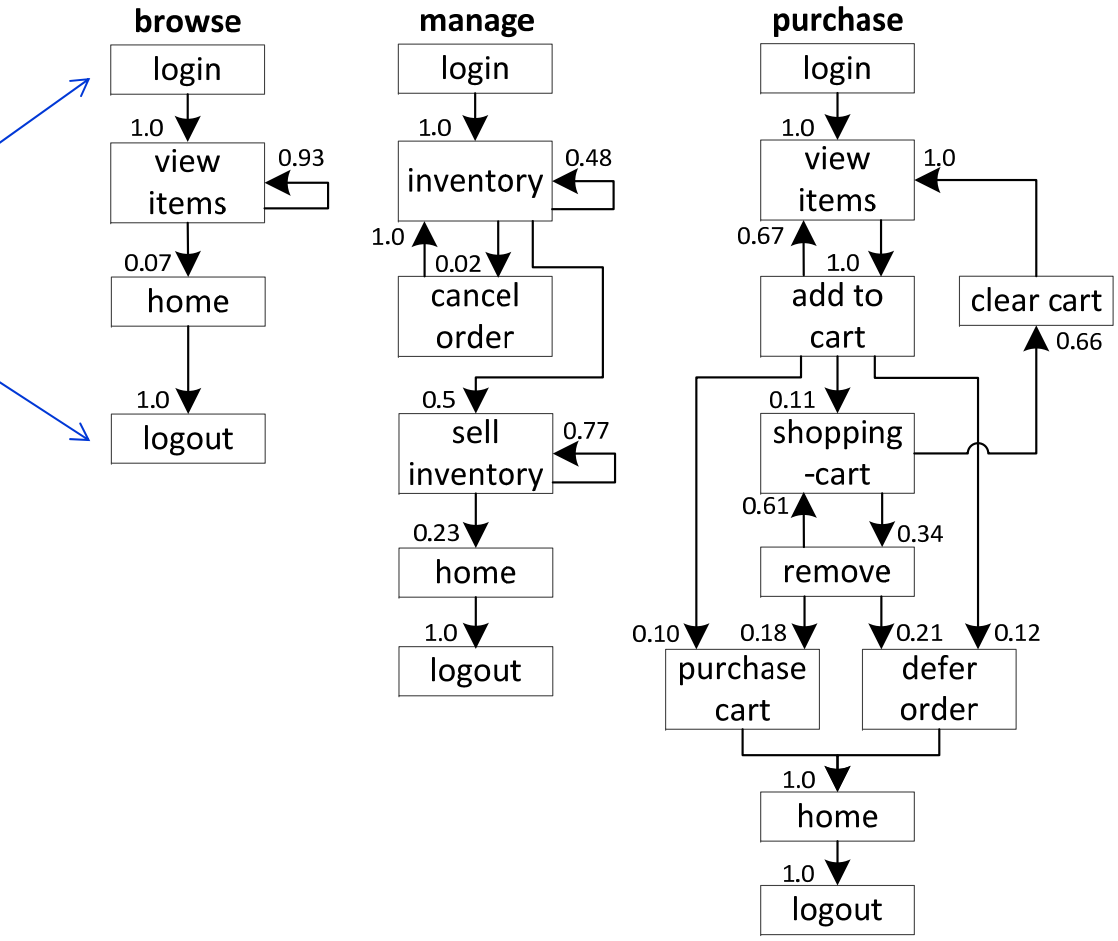
1. How accurately do the clustering results match the input Behavior Mix?
2. What is the impact of the clustering results on the workload characteristics?

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

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



Transaction Mix = Behavior Mix

Accuracy of Clustering

Research Question 1

		X-means (min 3 cluster, max 3 cluster)								X-means (min 2 cluster, max 20 cluster)								
		ED				NED				ED			NED					
TM	T	C1	C2	C3	MC	C1	C2	C3	MC	C1	C2	MC	C1	C2	C3	C4	MC	N
50	B	0	0	31,060		0	31,060	0		0	31,060		0	0	0	31,060		
25	M	15,298	0	0	2.91%	15,298	0	0	0%	15,298	0	24.62%	632	14,666	0	0	1.03%	61,500
25	P	1,789	13,353	0		0	0	15,142		15,142	0		0	0	15,142	0		
25	B	15,091	0	0		15,091	0	0		0	15,091		0	15,091	0	0		
25	M	0	0	15,000	15.98%	0	15,000	0	0%	15,000	0	24.96%	0	0	707	14,293	15.30%	60,089
50	P	0	20,397	9,601		0	0	29,998		29,998	0		21,513	8,485	0	0		
25	B	0	15,231	0		15,231	0	0		0	15,231		0	0	0	15,231		
50	M	30,510	0	0	2.99%	0	30,510	0	0%	30,510	0	25.16%	29,375	1,135	0	0	1.86%	61,118
25	P	1,824	0	13,553		0	0	15,377		15,377	0		0	0	15,377	0		

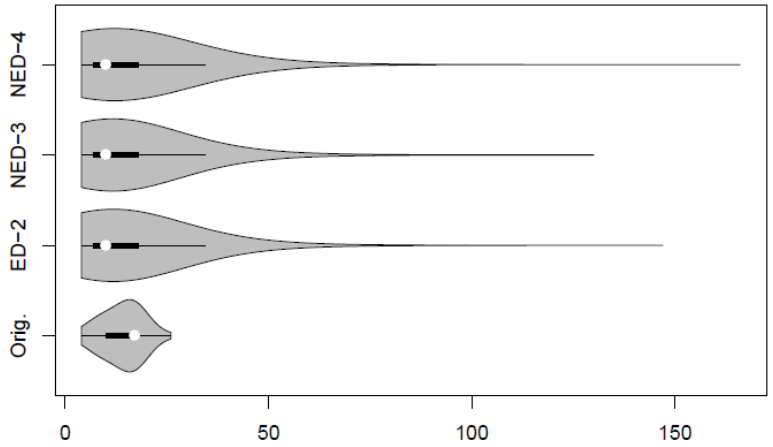
TM: Transaction Mix
T: Transaction
C_N: Assigned Cluster
MC: Percentage of misclassified
N: Number of instances

ED: Euclidean Distance
NED: Normalized Euclidean Distance

Accuracy of Extracted Workload Specifications

Research Question 2

Session-based metrics

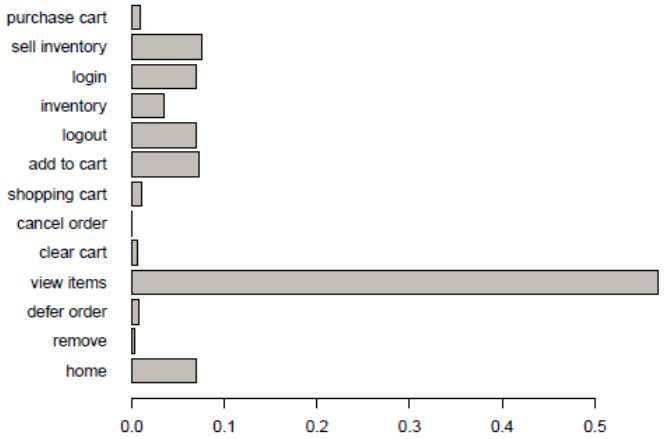


Session length

	Min.	Q_1	*Med.*	*Mean*	$CI_{0.95}$	Q_3	*Max.*	*N*
Orig.	4	10	17	14.23	[14.19,14.26]	17	26	61,500
ED-2	4	7	10	14.24	[14.15,14.33]	18	147	60,957
NED-3	4	7	10	14.24	[14.15,14.33]	18	130	62,054
NED-4	4	7	10	14.26	[14.17,14.35]	18	166	59,971

Statistics about session-based metrics

Request-based metrics



Relative counts (common to all workloads)

	Request	*Orig.*	*ED-2*	*NED-3*	*NED-4*	*Rel.*
1	add to cart	63,761	63,316	64,250	61,838	0.07
2	cancel order	632	607	634	591	0.00
3	clear cart	6,047	5,941	6,140	5,843	0.01
4	defer order	6,782	6,799	6,863	6,651	0.01
5	home	59,934	60,957	62,054	59,971	0.07
6	inventory	30,596	30,212	31,378	29,808	0.03
7	login	61,500	60,957	62,054	59,971	0.07
8	logout	59,934	60,957	62,054	59,971	0.07
9	purchase cart	8,360	8,328	8,351	8,139	0.01
10	remove	3,027	2,993	3,044	3,064	0.00
11	sell inventory	66,679	65,413	67,691	64,794	0.08
12	shopping cart	9,074	8,934	9,184	8,907	0.01
13	view items	498,601	492,675	499,983	485,611	0.57
	$\sum$	874,927	868,089	883,680	855,159	1.00

Absolute and relative counts

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

- 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. 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.
- 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.
- 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.
- 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. Vögele, A. van Hoorn, H. Krcmar. Automatic Extraction of Session-Based Workload Specifications for Architecture-Level Performance Models. In Proc. LT@ICPE 2015 (to appear). ACM. 2015
- 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.
- 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.
- 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.