Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

Descartes research

Trade-offs in Different Modeling Approaches for Performance Prediction of Software Systems

Samuel Kounev, Fabian Brosig, Philipp Meier, Steffen Becker, Anne Koziolek, Heiko Koziolek and Piotr Rygielski

Slides available at **http://descartes.tools**

Software Engineering 2016, Vienna, February 24, 2016

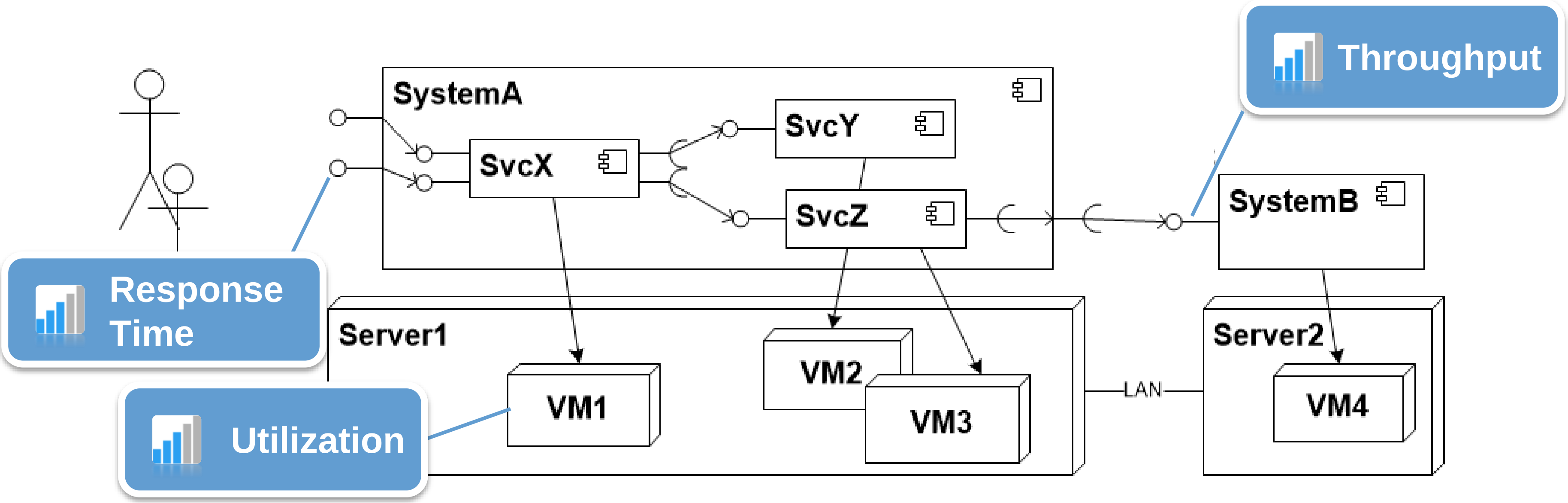
Main References

- F. Brosig, P. Meier, S. Becker, A. Koziolek, H. Koziolek, and S. Kounev. **Quantitative Evaluation of Model-Driven Performance Analysis and Simulation of Component-based Architectures**. *IEEE Transactions on Software Engineering (TSE)*, 41(2):157-175, February 2015, IEEE. [DOI | http | .pdf]

- P. Rygielski, S. Kounev, P. Tran-Gia. **Flexible Performance Prediction of Data Center Networks using Automatically Generated Simulation Models**. Proceedings of the Eighth EAI International Conference on Simulation Tools and Techniques (SIMUTools 2015), August 2015.

Research Context
Software Performance Engineering

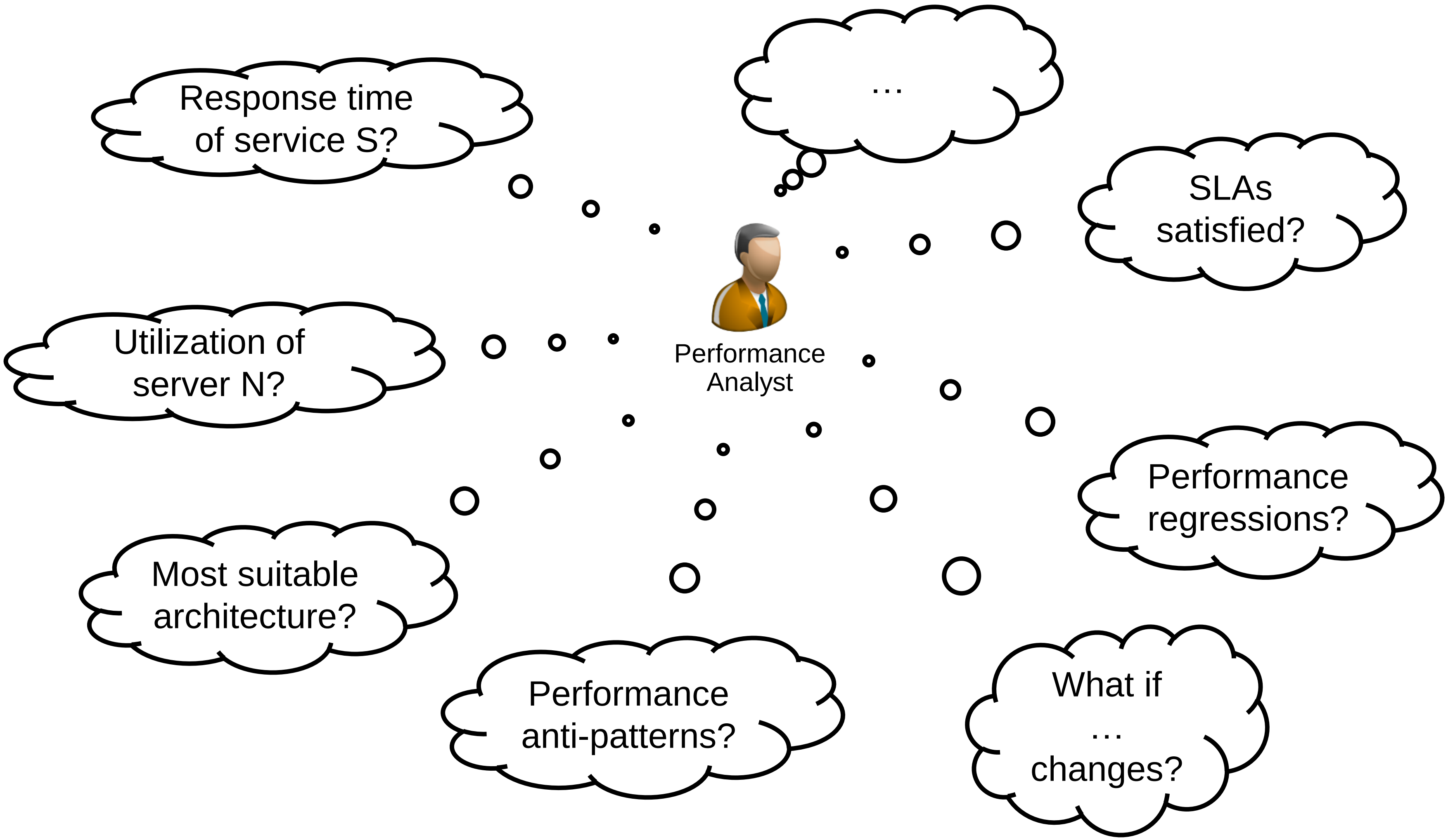
Performance = timing behavior + resource usage



Software Performance Engineering (SPE)

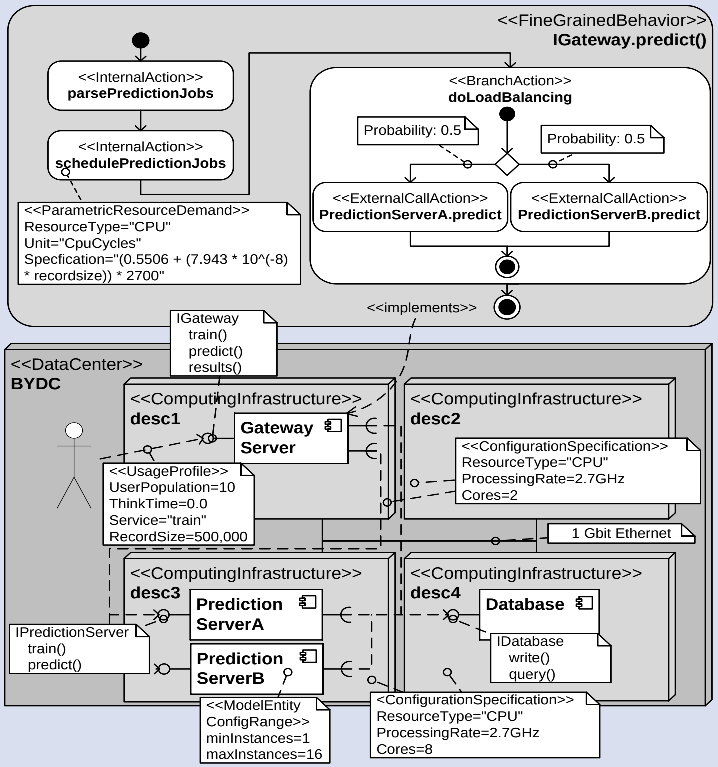
"the entire collection of **software engineering activities and related analyses** used throughout the **software development cycle**, which are directed to meeting **performance requirements**."

Performance-Relevant Concerns Spanning the Software Lifecycle



Performance Prediction

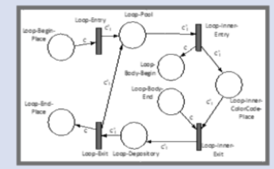
Performance-annotated Software Architecture Model



Transformation to Stochastic Performance Model

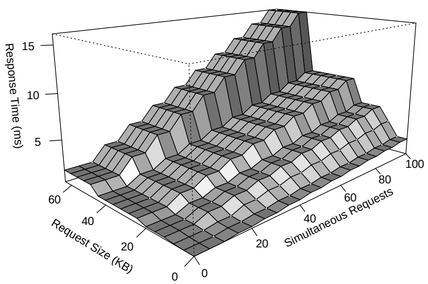
$$\overline{X} \le min\left\{\frac{N}{\sum_{i=0}^{n} D_i^{sync}}, min_{1\le i\le n}\left\{\frac{1}{D_i}\right\}\right\}$$

$$\overline{R} = \frac{N}{\overline{X}} \ge max\left\{\sum_{i=0}^{n} D_i^{sync}, N * max_{1\le i\le n}\{D_i\}\right\}$$

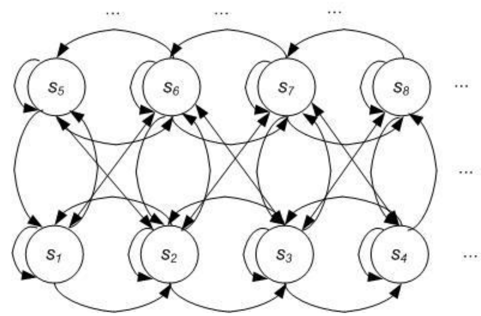


Predicted Performance Metrics

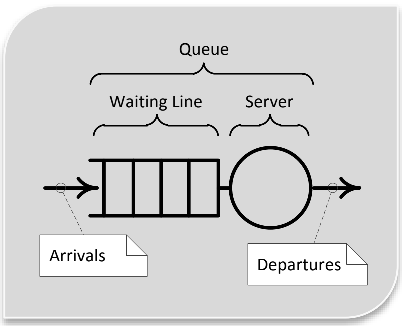
Stochastic Performance Models



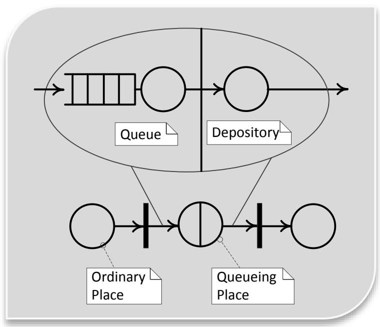
Statistical Regression



Markov Model



Queueing Network (QN)

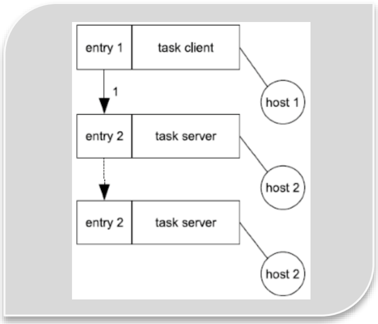


Queueing Petri net

$$R \geq \max\left[N \times \max\{D_i\}, \sum_{i=1}^{K} D_i\right] \qquad X_0 \leq \min\left[\frac{1}{\max\{D_i\}}, \frac{N}{\sum_{i=1}^{K} D_i}\right]$$

$$\frac{N}{\max\{D_i\}[K+N-1]} \leq X_0 \leq \frac{N}{avg\{D_i\}[K+N-1]}$$

Analytical Bounds Analysis



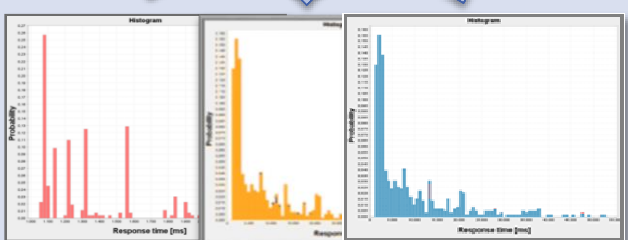
Layered Queueing Network (LQN)

Model Solution

Analytical Analysis

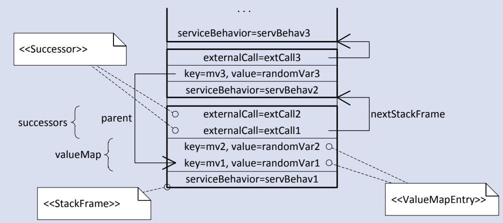
$$R \geq \max\left[N \times \max\{D_i\}, \sum_{i=1}^{K} D_i\right] \qquad X_0 \leq \min\left[\frac{1}{\max\{D_i\}}, \frac{N}{\sum_{i=1}^{K} D_i}\right]$$

$$\frac{N}{\max\{D_i\}[K+N-1]} \leq X_0 \leq \frac{N}{avg\{D_i\}[K+N-1]}$$



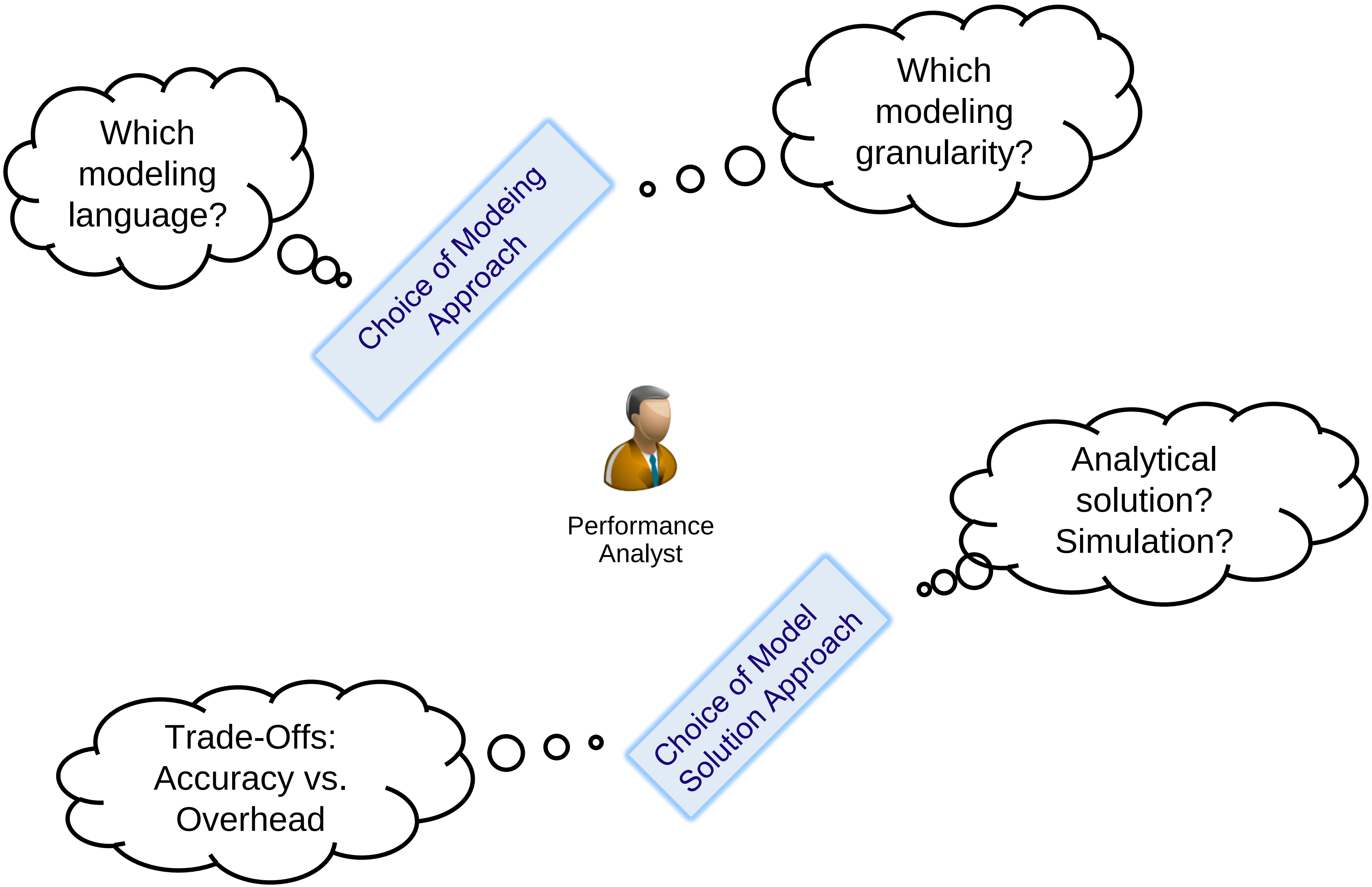
Analysis Results

Simulative Analysis

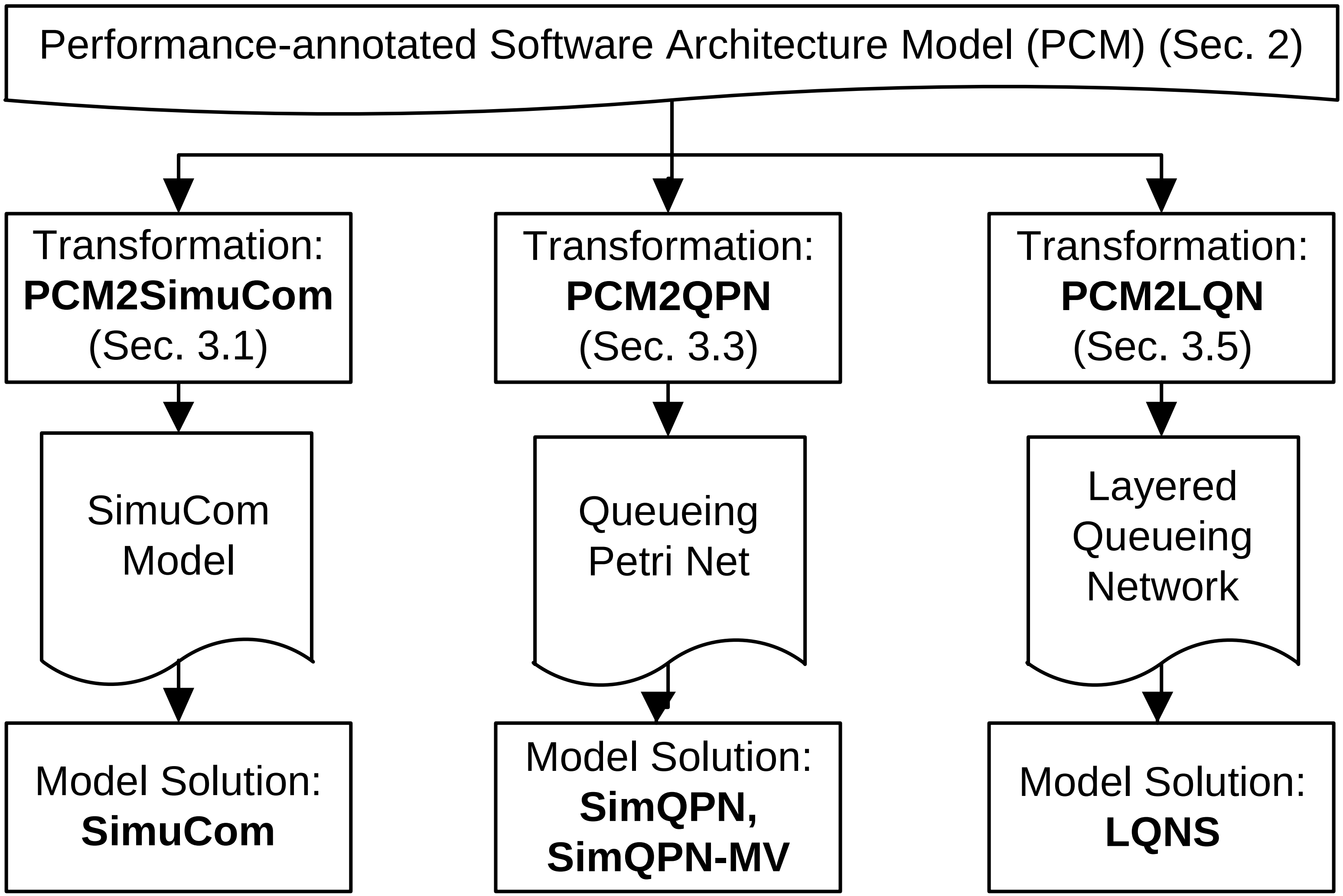


Analysis Results

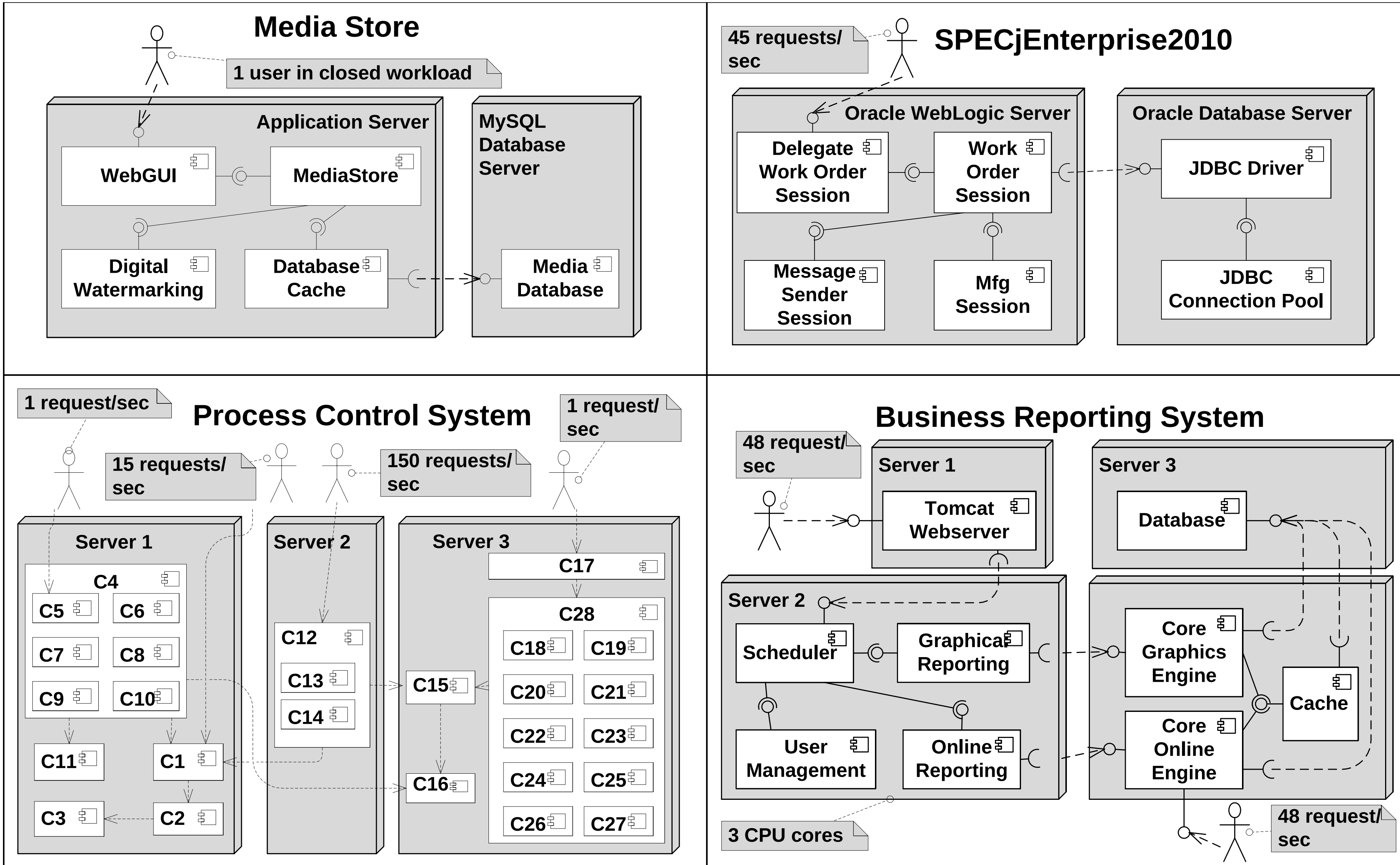
Modeling Challenges



Considered Prediction Approaches



Case Studies



Semantic Gaps

Modeling feature	SimuCom	SimQPN	SimQPN-MV	LQNS
Loops	X	(X)	(X)	(X)
Forks with synchronization barrier	X	(X)	(X)	(X)
Parameter dependencies	X	(X)	(X)	(X)
Response time distributions	X	X	-	-
Flexible parameter characterizations	X	X	X	-
Blocking behavior	X	X	X	(X)
X support	(X) partial support	- no support		

TABLE 1
Semantic gaps.

Quantitative Evaluation: Accuracy

	SimuCom (reference)	SimQPN	SimQPN (relDiff)	SimQPN-MV	SimQPN-MV (relDiff)	LQNS	LQNS (relDiff)
Media Store							
RT(Scenario1)	1.332	1.331	0.0%	1.324	-0.6%	1.288	-3.3%
TP(Scenario1)	0.751	0.751	0.0%	0.755	0.6%	0.753	0.3%
U(AppServer_CPU)	0.341	0.345	1.2%	0.340	-0.3%	0.342	0.2%
U(DBServer_CPU)	1.4%	1.4%	1.6%	1.4%	2.6%	1.4%	2.2%
U(DBServer_HDD)	64.4%	64.0%	-0.6%	64.5%	0.2%	64.4%	-0.1%
SPECjEnterprise2010							
RT(Scenario1)	1.060	1.058	-0.2%	1.065	0.4%	1.065	0.4%
TP(Scenario1)	24.567	24.970	1.6%	24.792	0.9%	25.000	1.8%
U(Oracle_CPU)	29.1%	29.6%	1.7%	29.5%	1.3%	29.6%	1.8%
U(WLS_CPU)	51.2%	52.0%	1.5%	51.9%	1.3%	52.2%	2.0%
Process Control System							
RT(Scenario1)	0.00787	0.00764	-2.9%	0.00763	-3.0%	0.00790	0.4%
RT(Scenario2)	0.06690	0.06636	-0.8%	0.06903	3.2%	0.06710	0.3%
TP(Scenario1)	149.258	149.254	0.0%	149.254	0.0%	149.254	0.0%
TP(Scenario2)	15.001	15.000	0.0%	15.000	0.0%	14.999	0.0%
U(Server1_CPU)	6.5%	6.4%	-2.2%	6.4%	-2.0%	6.5%	-0.5%
U(Server1_HDD)	45.0%	45.0%	0.0%	45.0%	0.0%	45.0%	0.0%
U(Server2_CPU)	1.1%	1.1%	-4.0%	1.1%	-4.0%	1.1%	-4.0%
U(Server3_CPU)	55.0%	54.1%	-1.6%	54.0%	-1.8%	55.1%	0.2%
Business Reporting System							
RT(Scenario1 - cons1)	5.223	5.918	13.3%	5.921	13.4%	7.391	41.5%
TP(Scenario1 - cons1)	1.000	1.000	0.0%	1.000	0.0%	1.000	0.0%
U(Server1_CPU) Scen. 1	45.0%	44.8%	-0.5%	44.9%	-0.2%	45.0%	0.0%
U(Server2_CPU) Scen. 1	68.8%	68.6%	-0.3%	68.8%	-0.1%	68.9%	0.1%
U(Server3_CPU) Scen. 1	74.3%	74.8%	0.7%	74.5%	0.3%	74.3%	0.0%
U(Server4_CPU) Scen. 1	23.5%	23.5%	-0.2%	23.5%	0.1%	23.5%	0.2%
RT(Scenario2 - Exp(1))	7.302	7.256	-0.6%	7.500	2.7%	7.391	1.2%
TP(Scenario2 - Exp(1))	1.002	1.000	-0.2%	0.996	-0.6%	1.000	-0.2%

KEY: RT = Response Time (sec), TP = Throughput (requests / sec), U = Utilization (%), relDiff = relative difference

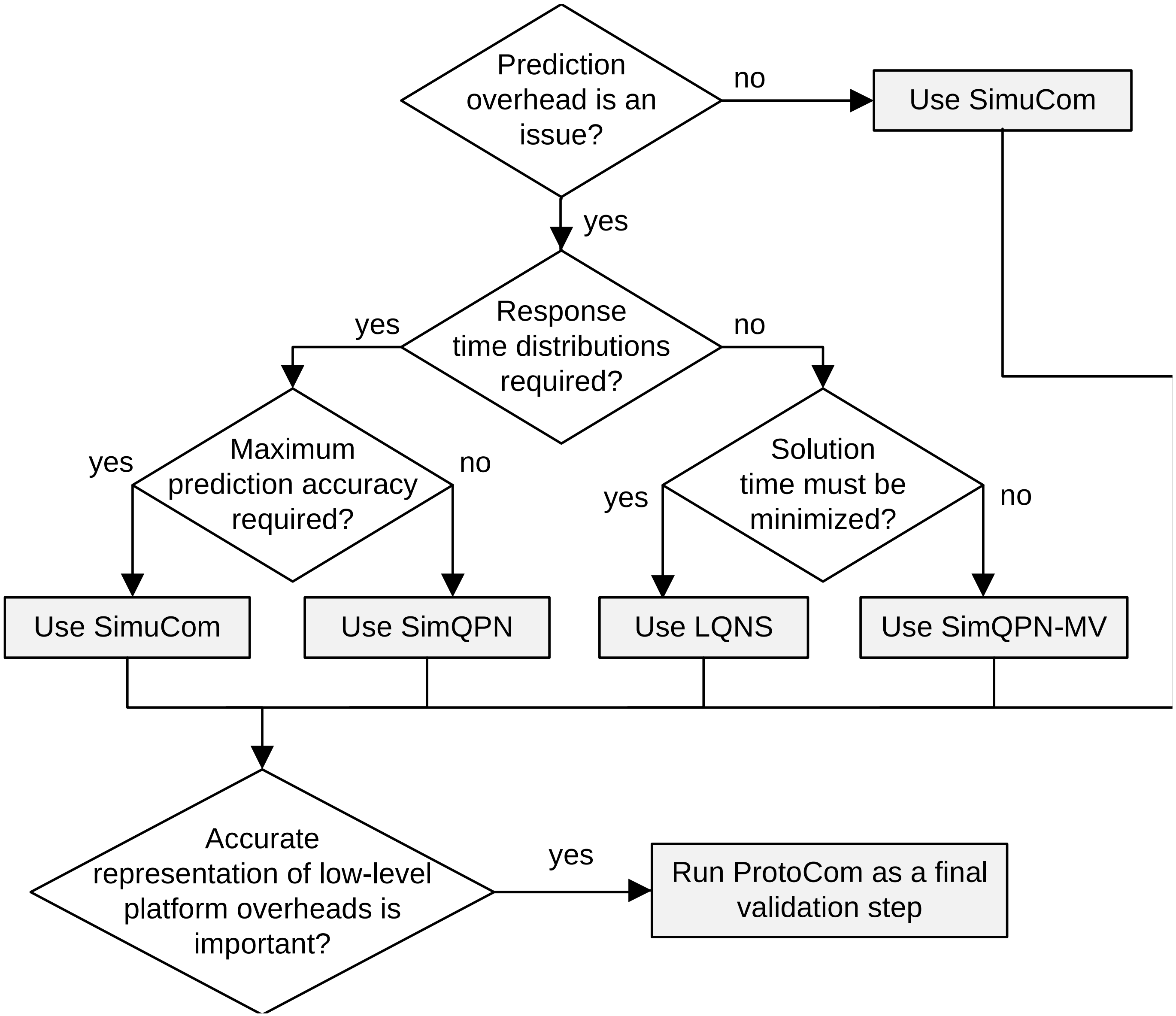
Quantitative Evaluation: Overhead

	SimuCom (reference)	SimQPN	SimQPN (relDiff)	SimQPN-MV	SimQPN-MV (relDiff)	LQNS	LQNS (relDiff)
Media Store							
Logical Simulation Time	25000	25000	0.0%	25000	0.0%	n/a	n/a
Number of Measurements	18773	18775	0.0%	18775	0.0%	n/a	n/a
Total Execution Time	31.4	6.4	-79.6%	3.6	-88.5%	1.3	-96.0%
Wall-clock Sim./Ana. Time	25.6	1.7	-93.4%	0.9	-96.5%	0.2	-99.1%
SPECjEnterprise2010							
Logical Simulation Time	500	500	0.0%	500	0.0%	n/a	n/a
Number of Measurements	12257	12485	1.9%	12485	1.9%	n/a	n/a
Total Execution Time	72.6	7.9	-89.1%	5.6	-92.3%	1.8	-97.6%
Wall-clock Sim./Ana. Time	55.6	3.8	-93.3%	2.1	-96.3%	0.5	-99.2%
Process Control System							
Logical Simulation Time	500	500	0.0%	500	0.0%	n/a	n/a
Number of Measurements	74627	74627	0.0%	74627	0.0%	n/a	n/a
Total Execution Time	96.7	11.9	-87.7%	7.2	-92.6%	3.9	-95.9%
Wall-clock Sim./Ana. Time	65.4	1.6	-97.6%	2.6	-96.0%	2.0	-96.9%
Business Reporting System							
Logical Simulation Time	9994	10000	0.1%	10000	0.1%	n/a	n/a
Number of Measurements	10016	9997	-0.2%	9997	-0.2%	n/a	n/a
Total Execution Time	591.5	74.5	-87.4%	34.1	-94.2%	4.0	-99.3%
Wall-clock Sim./Ana. Time	587.6	35.4	-94.0%	26.6	-95.5%	1.8	-99.7%
All times in seconds							

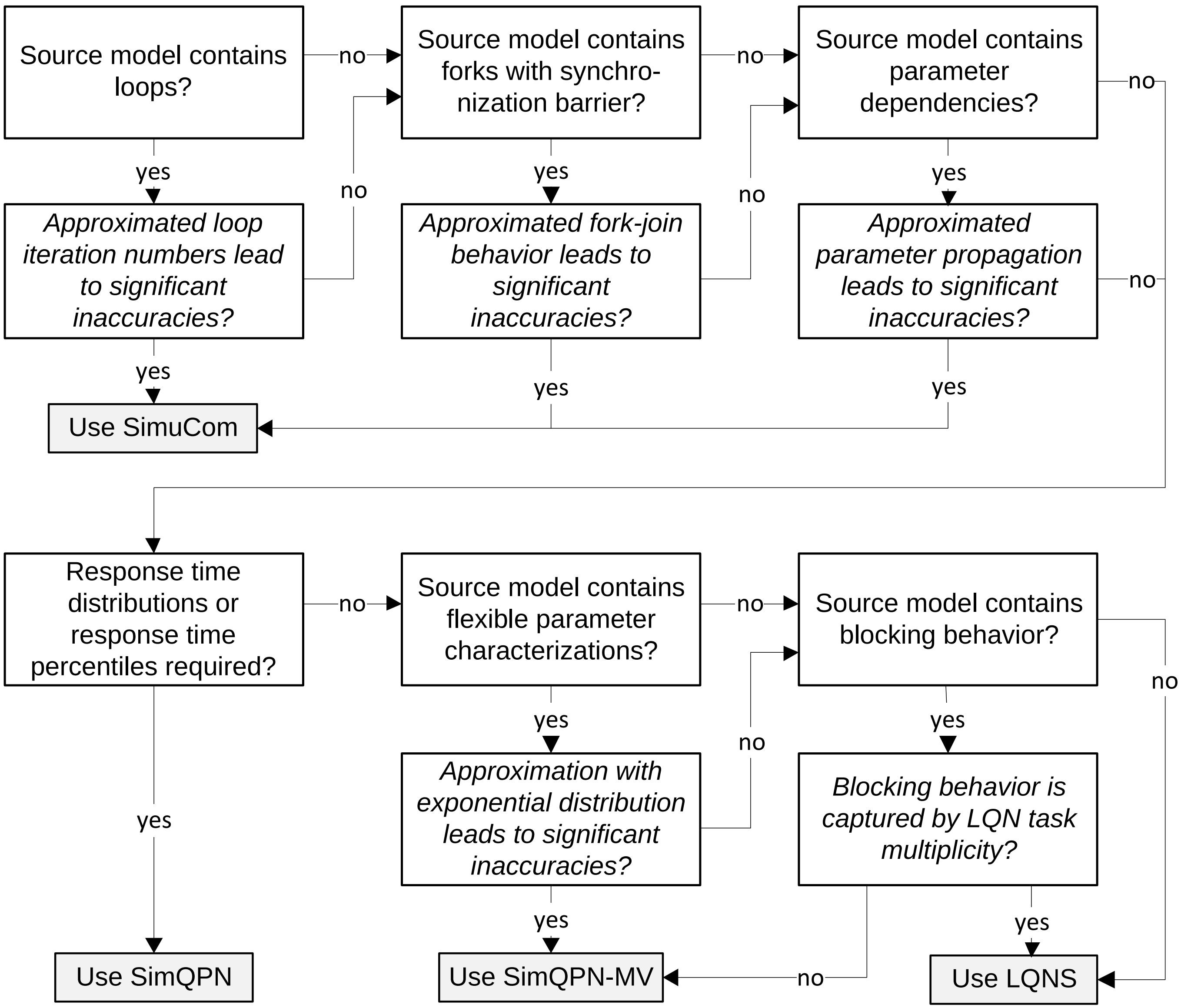
Quantitative Evaluation: Impact of Load

	SimuCom	SimQPN	SimQPN-MV	LQNS
Media Store (low load)				
Total Execution Time	21.5	5.7	3.4	1.3
Wall-clock Sim./Ana. Time	16.0	1.1	0.6	0.2
Media Store (medium load)				
Total Execution Time	31.4	6.4	3.6	1.3
Wall-clock Sim./Ana. Time	25.6	1.7	0.9	0.2
Media Store (high load)				
Total Execution Time	40.7	7.3	4.0	1.4
Wall-clock Sim./Ana. Time	34.1	2.4	1.2	0.2
All times in seconds				

Decision Tree (simplified)



Decision Tree



DECLARE Project
Declarative Performance Engineering

DFG Priority Programme 1593

Project PIs
Dr.-Ing. **André van Hoorn** (Prof.-Vertr.), University of Stuttgart
Prof. Dr.-Ing. **Samuel Kounev**, University of Würzburg

Members
Dr.-Ing. **Dušan Okanović**, University of Stuttgart
Dipl.-Inform. **Jürgen Walter**, University of Würzburg

Associated Partners
Capgemini Deutschland GmbH, Stuttgart, Germany

University of Stuttgart
Germany

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

Declare

Declarative Performance Engineering

Performance Concerns

Performance Analyst

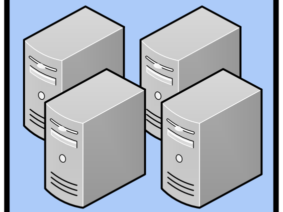
Software Performance Engineering

Established Methods, Techniques, and Tools

Asking “What?“,
Automating the “How?“

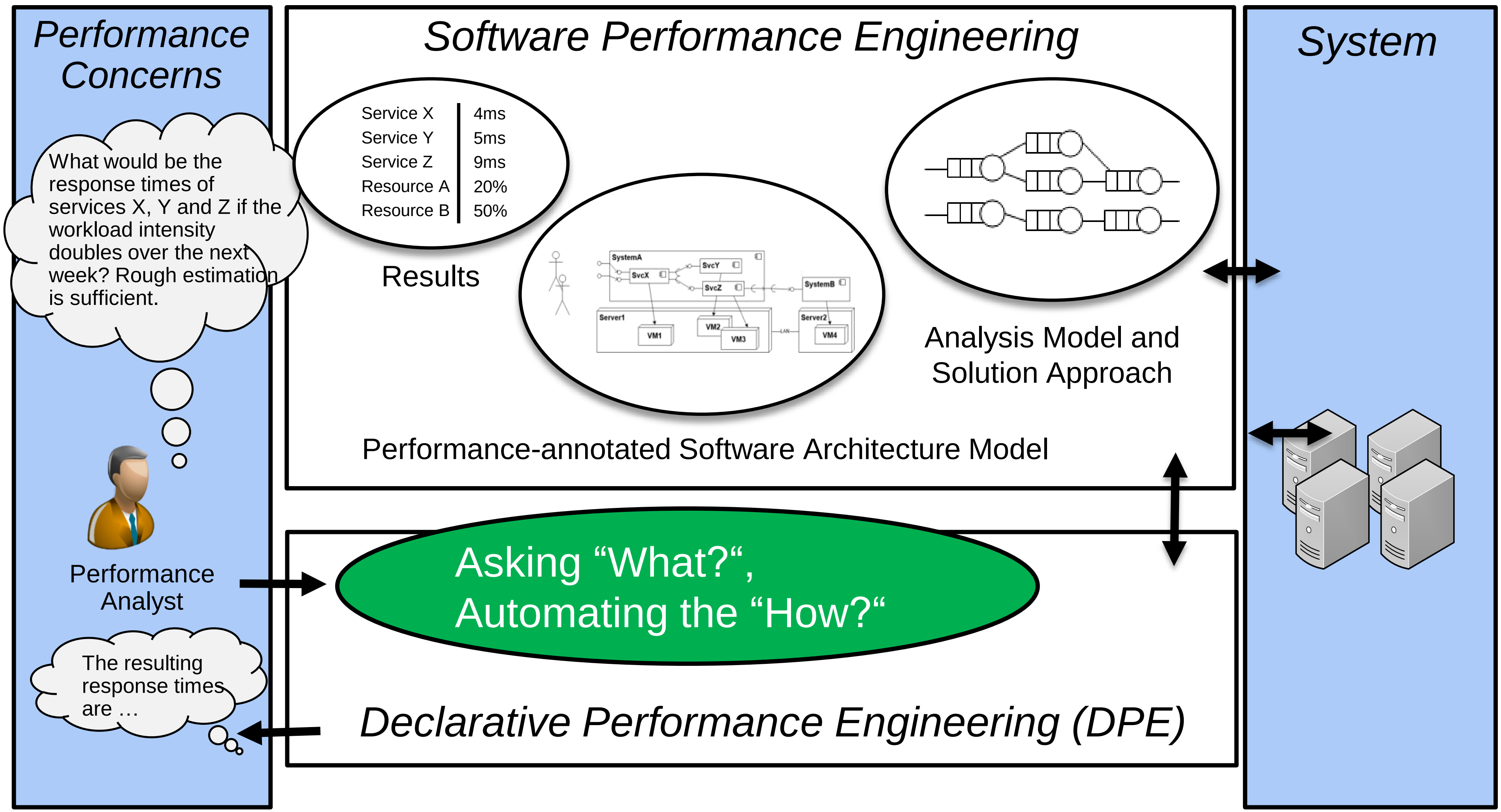
Declarative Performance Engineering (DPE)

System



Jürgen Walter, Andre van Hoorn, Heiko Koziolek, Dusan Okanovic, and Samuel Kounev. Asking “What?”, Automating the “How?”:The Vision of Declarative Performance Engineering. In *Proceedings of the 7th ACM/SPEC International Conference on Performance Engineering (ICPE 2016)*, Delft, the Netherlands, March 12, 2016. [.pdf]

Declarative Performance Engineering



van Hoorn/Kounev, Declare, Renewal Kickoff of SPP1593, Hannover, Jan. 14 – 15, 2016

References

- S. Kounev, N. Huber, F. Brosig, and X. Zhu. **Model-Based Approach to Designing Self-Aware IT Systems and Infrastructures**. *IEEE Computer Magazine*, 2016, IEEE. Accepted for Publication, To appear in 2016. *Available on request.*
- S. Kounev, F. Brosig, and N. Huber. **The Descartes Modeling Language**. Technical report, Department of Computer Science, University of Wuerzburg, October 2014. [http | http | .pdf]
- F. Brosig, N. Huber, and S. Kounev. **Architecture-Level Software Performance Abstractions for Online Performance Prediction**. *Elsevier Science of Computer Programming Journal (SciCo)*, Vol. 90, Part B:71-92, 2014, Elsevier. [DOI | http | .pdf]
- N. Huber, A. van Hoorn, A. Koziolek, F. Brosig, and S. Kounev. **Modeling Run-Time Adaptation at the System Architecture Level in Dynamic Service-Oriented Environments**. *Service Oriented Computing and Applications Journal (SOCA)*, 8(1):73-89, 2014, Springer-Verlag. [DOI | .pdf]
- F. Brosig, P. Meier, S. Becker, A. Koziolek, H. Koziolek, and S. Kounev. **Quantitative Evaluation of Model-Driven Performance Analysis and Simulation of Component-based Architectures**. *IEEE Transactions on Software Engineering (TSE)*, 41(2):157-175, February 2015, IEEE. [DOI | http | .pdf]
- S. Spinner, G. Casale, F. Brosig, and S. Kounev. **Evaluating Approaches to Resource Demand Estimation**. *Performance Evaluation*, 92:51-71, October 2015, Elsevier B.V. [DOI | http | .pdf]
- N. Herbst, N. Huber, S. Kounev and E. Amrehn. **Self-Adaptive Workload Classification and Forecasting for Proactive Resource Provisioning**. *Concurrency and Computation - Practice and Experience, 26(12):2053-2078, John Wiley and Sons, Ltd.*, 2014. [DOI | http | .pdf]
- J. v. Kistowski, N. Herbst, D. Zoller, S. Kounev, and A. Hotho. **Modeling and Extracting Load Intensity Profiles**. In *Proceedings of the 10th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS 2015)*, Firenze, Italy, May 18-19, 2015. [DOI | slides | .pdf]
- N. Herbst, S. Kounev and R. Reussner. ***Elasticity: What it is, and What it is Not.*** *In 10th Intl. Conference on Autonomic Computing (ICAC 2013), San Jose, CA, June 24-28, 2013.* [slides | http | .pdf]
- N. Herbst, S. Kounev, A. Weber, and H. Groenda. **BUNGEE: An Elasticity Benchmark for Self-Adaptive IaaS Cloud Environments**. In *Proceedings of the 10th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS 2015)*, Firenze, Italy, May 18-19, 2015. [slides | .pdf]
- A. Milenkoski, M. Vieira, S. Kounev, A. Avrtizer, and B. Payne. **Evaluating Computer Intrusion Detection Systems: A Survey of Common Practices**. *ACM Computing Surveys*, 48(1):12:1-12:41, September 2015, ACM, New York, NY, USA.
- A. Milenkoski, B. Payne, N. Antunes, M. Vieira, S. Kounev, A. Avritzer, and M. Luft. **Evaluation of Intrusion Detection Systems in Virtualized Environments Using Attack Injection**. In The 18th International Symposium on Research in Attacks, Intrusions, and Defenses (RAID 2015), Kyoto, Japan, November 2015. Springer. November 2015.

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

Descartes
research

Questions?

skounev@acm.org
http://descartes.tools