SORMA

IISM Institute of Information Systems and Management

Ping Symposium, Universität Karlsruhe (TH), 16. März 2007

Auction-based Scheduling for Interactive Grid Applications

Jochen Stößer

European Commission

Information Society

Universität Karlsruhe (TH)

Agenda

Motivation

Auction-based Scheduling

GREEDEX – A Truthful and Scalable Auction Mechanism

Preliminary Simulation Results

Conclusion & Outlook

Agenda

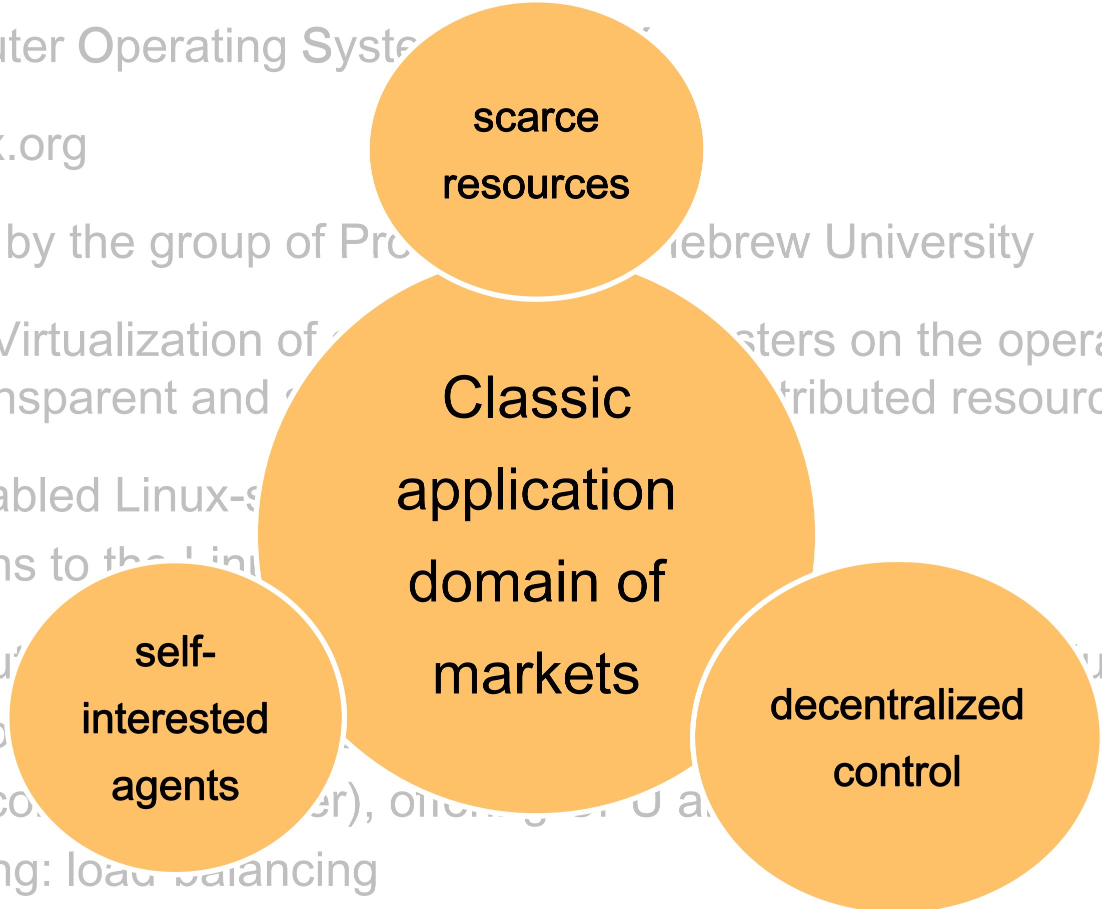
- Motivation
- Auction-based Scheduling
- GREEDEX – A Truthful and Scalable Auction Mechanism
- Preliminary Simulation Results
- Conclusion & Outlook

Motivation: MOSIX

- Multicomputer Operating System for UnIX (www.mosix.org)
- Developed by the group of Prof. Barak at Hebrew University
- Objective: Virtualization of a cluster/grid of clusters on the operation system level → transparent and seamless access to distributed resources
- Cluster-enabled Linux-system
  - Extensions to the Linux kernel
- Jobs are automatically being migrated to faster nodes in the cluster/grid
  - Job: set of processes → CPU and memory
  - Node: a computer (server), offering CPU and memory
  - Scheduling: load-balancing

Motivation: MOSIX

- Multicomputer Operating Syste…
- www.mosix.org
- Developed by the group of Pro… …ebrew University
- Objective: Virtualization of … …sters on the operation system level → transparent and … …tributed resources
- Cluster-enabled Linux-…
  - Extensions to the Linu…
- Jobs are au… …uster/grid
  - Job: set o…
  - Node: a co… …er), offer… …U a…
  - Scheduling: load balancing



Agenda

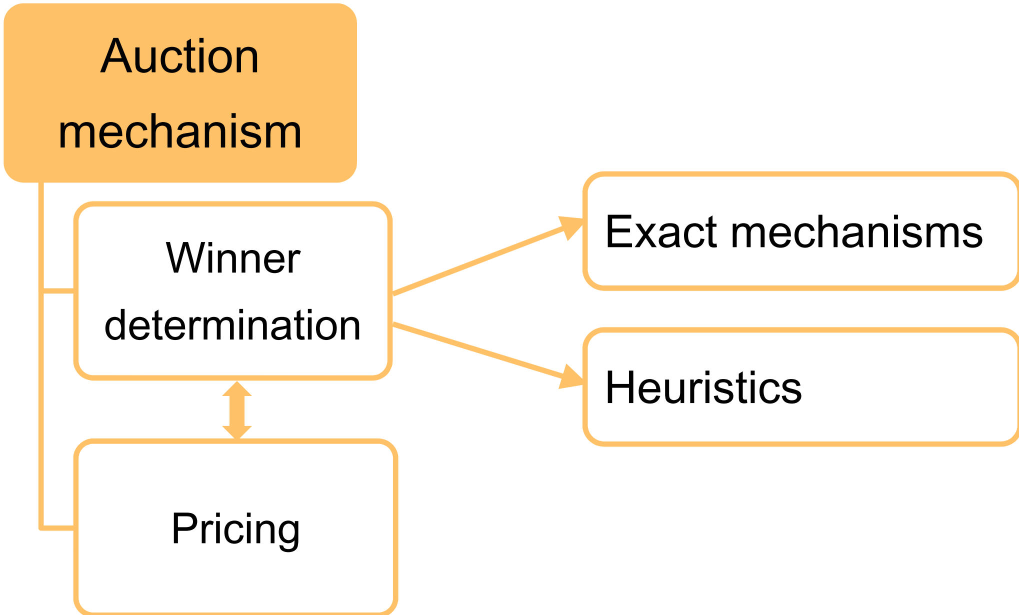
- Motivation
- **Auction-based Scheduling**
- GreedEx – A Truthful and Scalable Auction Mechanism
- Preliminary Simulation Results
- Conclusion & Outlook

Economic Requirements

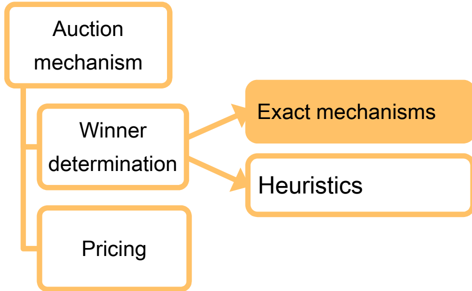
- **Allocative efficiency:** maximize overall “happiness”, no waste
- **Budget balance:** no outside payments
- **Computational tractability:** compute market outcome in polynomial time
- **Individual rationality:** no individual loss from participating
- **Truthfulness:** report true characteristics to the system

Auction mechanisms

Job j	b_j [$]	$\underline{c}_j$	$\underline{m}_j$	s_j	e_j	Node n	r_n [$]	$\bar{c}_n$	$\bar{m}_n$	ε_n	λ_n
J1	11	60	50	2	3	N1	5	150	100	1	5
J2	10	80	50	1	3	N2	7	100	120	2	6
J3	9	90	100	2	5	N3	8	140	100	1	5
J4	8	130	100	4	5						
J5	7	80	90	1	3						
J6	6	70	50	2	4						
J7	5	30	40	1	3						



Exact Mechanisms – Scheduling Problem



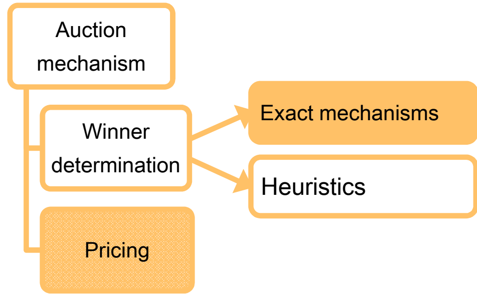
$$\max_X V := \sum_{j \in J} c_j \sum_{t \in T^b(j)} \sum_{n \in N(j)} x_{jnt} (b_j - r_n)$$

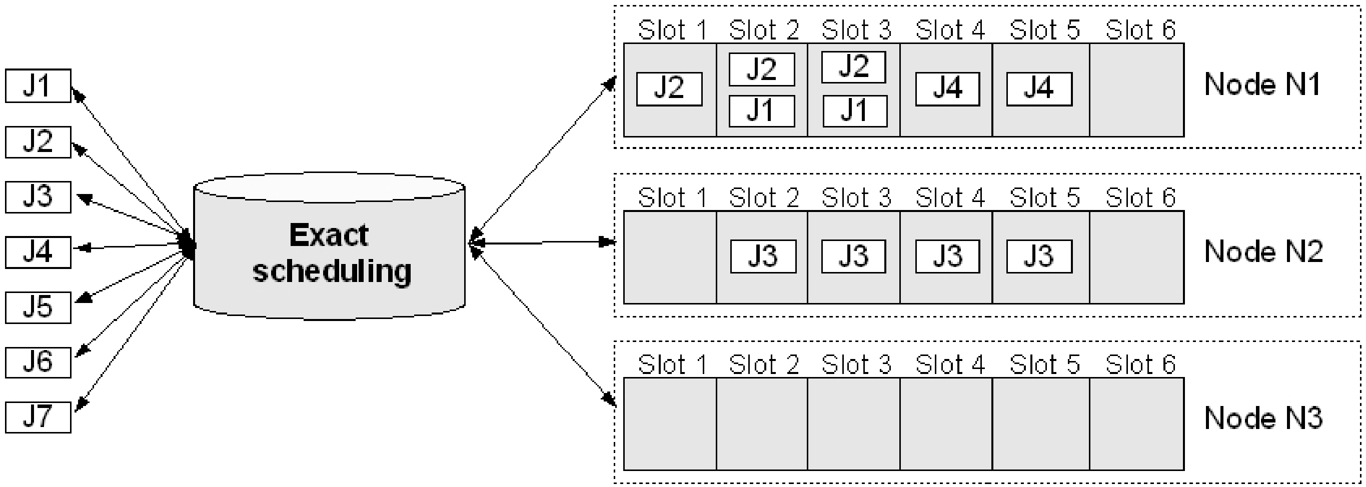
maximize welfare

$$s.t. \quad feasible(X)$$

constraints encode use case

Exact Mechanisms – Allocation





$\Rightarrow$ (optimal) welfare V* = $3,420

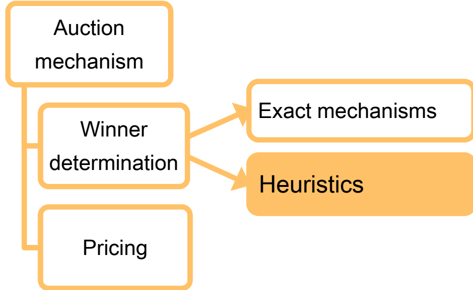
- Vickrey-Clarke-Groves mechanism: not budget-balanced
- Parkes: budget-balanced, but only approximately truthful

(theoretically) computationally intractable

Agenda

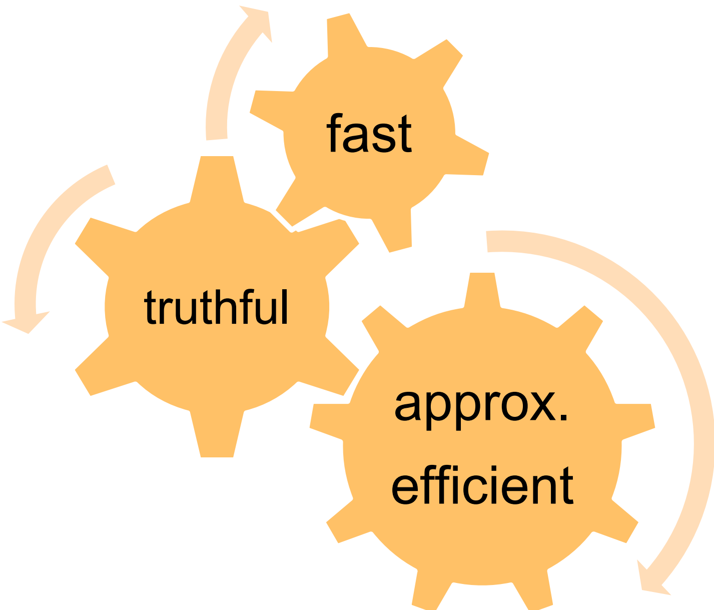
- Motivation
- Auction-based Scheduling
- GREEDEX – A Truthful and Scalable Auction Mechanism
- Preliminary Simulation Results
- Conclusion & Outlook

GreedEx – Motivation

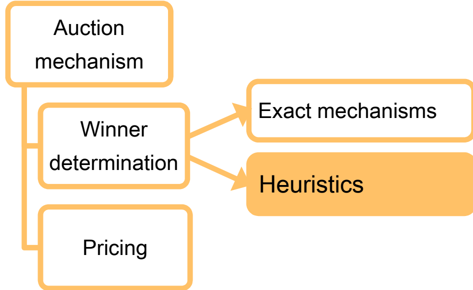


Is it possible to

- exploit the restrictions of the MOSIX scenario (computing power and memory only) so as to
- design a market mechanism which is…



GREEDEX – Allocation



A greedy heuristic:

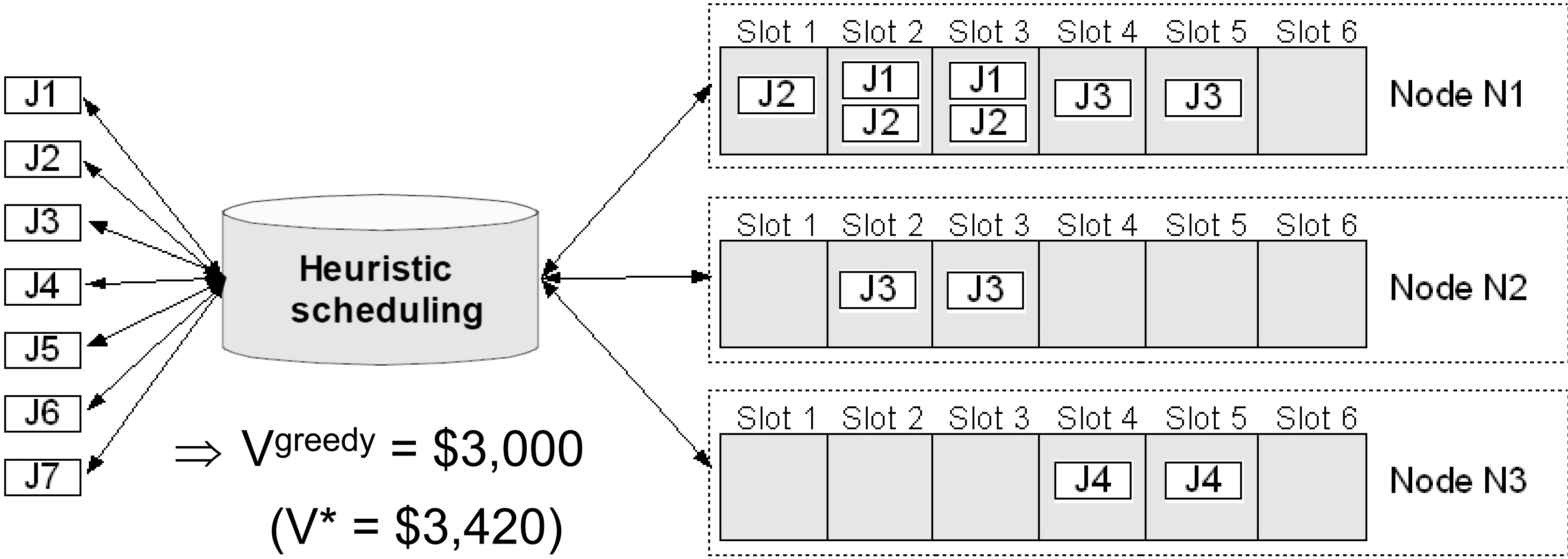
- Rank jobs and nodes wrt to some norm, e.g. b_j, r_n
- Sequentially run through the job ranking and for each job j
  - Sequentially run through the node ranking
  - Assign j to the nodes with the highest ranking which can together accommodate j

$\Rightarrow$ maximize $b_j - r_n$: $\max_X V := \sum_{j \in J} \underline{c}_j \sum_{t \in T^b(j)} \sum_{n \in N(j)} x_{jnt} (b_j - r_n)$

GREEDEX – Allocation (cont'd)

Job j	b_j [\$]	c_j	m_j	s_j	e_j
J1	11	60	50	2	3
J2	10	80	50	1	3
J3	9	90	100	2	5
J4	8	130	100	4	5
J5	7	80	90	1	3
J6	6	70	50	2	4
J7	5	30	40	1	3

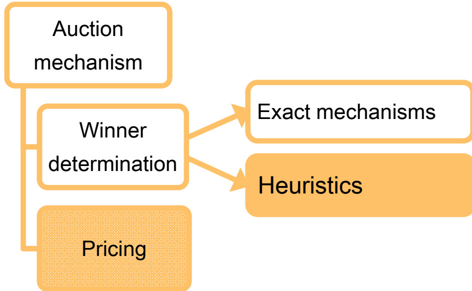
Node n	r_n [\$]	$\bar{c}_n$	$\bar{m}_n$	ε_n	λ_n
N1	5	150	100	1	5
N2	7	100	120	2	6
N3	8	140	100	1	5



$\Rightarrow V^{greedy} = \$3{,}000$

$(V^* = \$3{,}420)$

GREEDEX – Pricing



Basic idea for the pricing:

- The payment of each job depends on its *critical value*, i.e. the minimum valuation it would have needed to report to the mechanism in order to still remain in the allocation.
- The possible surplus (total revenue minus reservation prices) is distributed to the nodes so as to induce approximately truthful reports of reservation prices.

In the example:

$$\Phi_{J1}(A, b_{-J1}) = \$5 \rightarrow p_{J1} = \$600$$

$$\Phi_{J2}(A, b_{-J2}) = \$5 \rightarrow p_{J2} = \$1{,}200$$

$$\Phi_{J3}(A, b_{-J3}) = \$7 \rightarrow p_{J3} = \$2{,}520$$

$$\Phi_{J4}(A, b_{-J4}) = \$8 \rightarrow p_{J4} = \$2{,}080$$

$$p_{J5} = p_{J6} = p_{J7} = 0$$

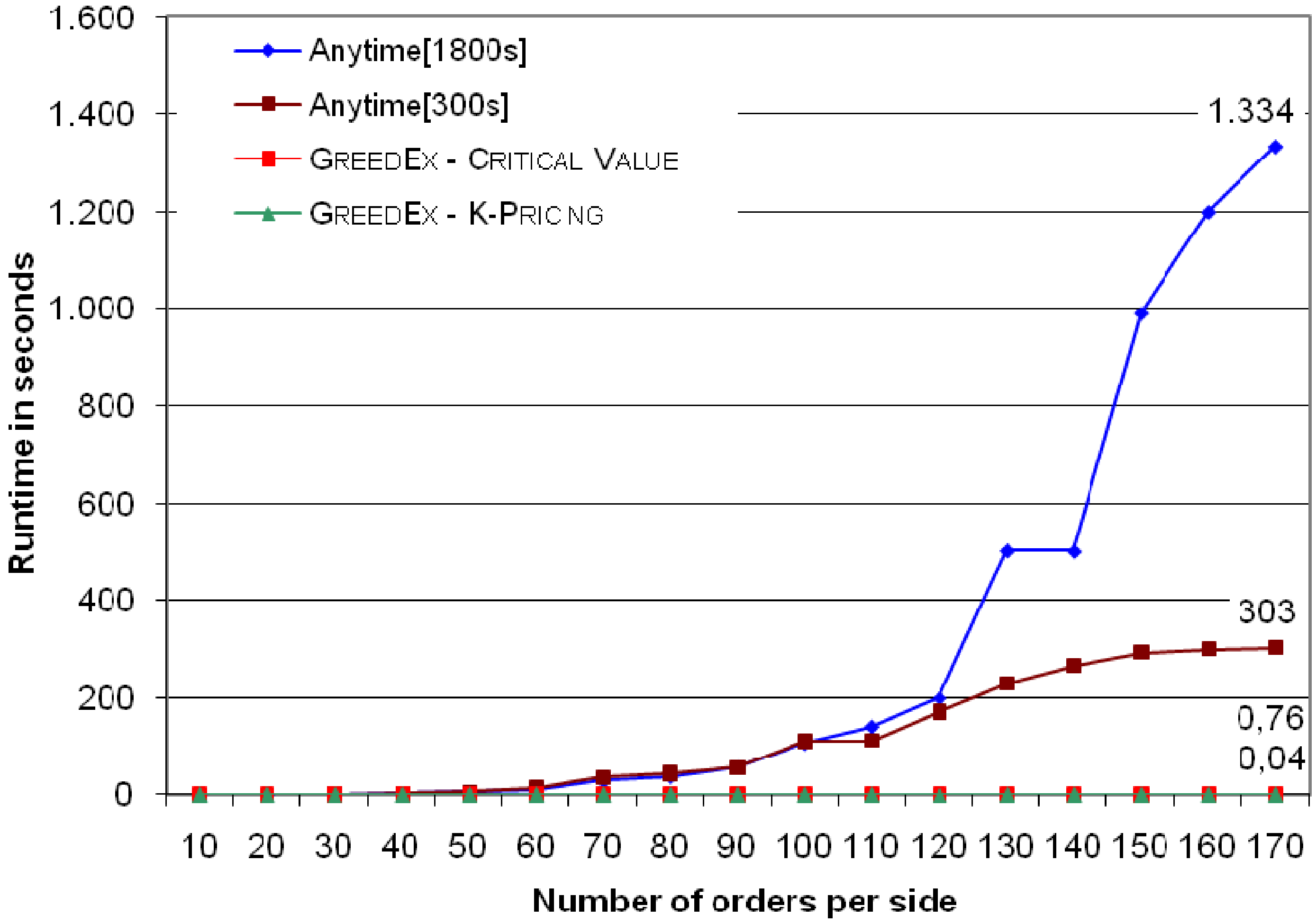
Agenda

- Motivation
- Auction-based Scheduling
- GREEDEX – A Truthful and Scalable Auction Mechanism
- **Preliminary Simulation Results**
- Conclusion & Outlook

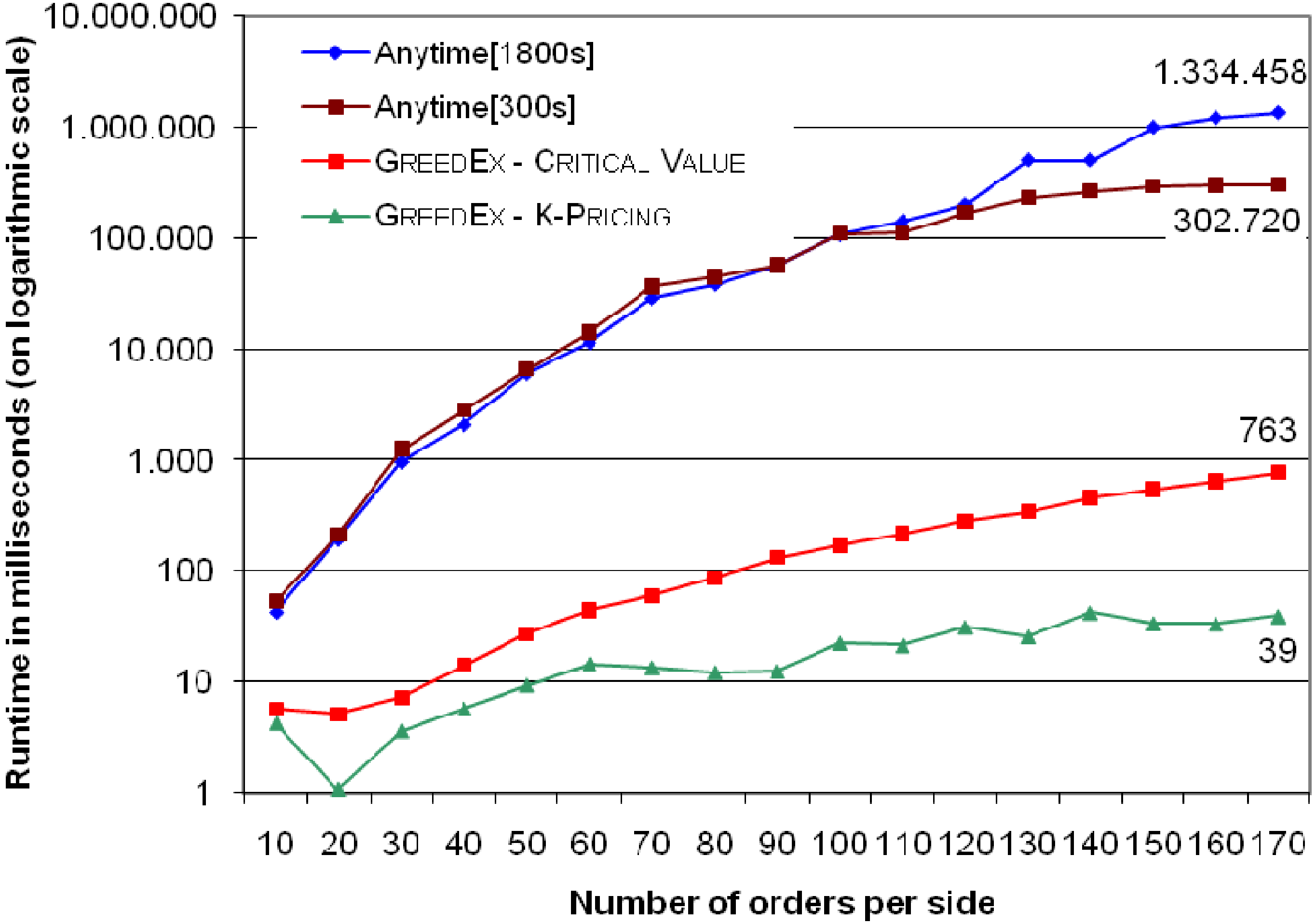
First simulation setting

- feed both mechanisms with identical order streams
- successively increase number of orders on both sides from 10, 20, 30, … up to 170 orders per side
- run each setting 30 times
- exact mechanism is run with CPLEX 9.1 and **stopped after a predefined time limit** → „anytime“ behavior
- underlying distributions:
  - attributes cpu and memory: lognormal distribution (cf. Feitelson)
  - time slots: normal distribution
  - valuations: uniform

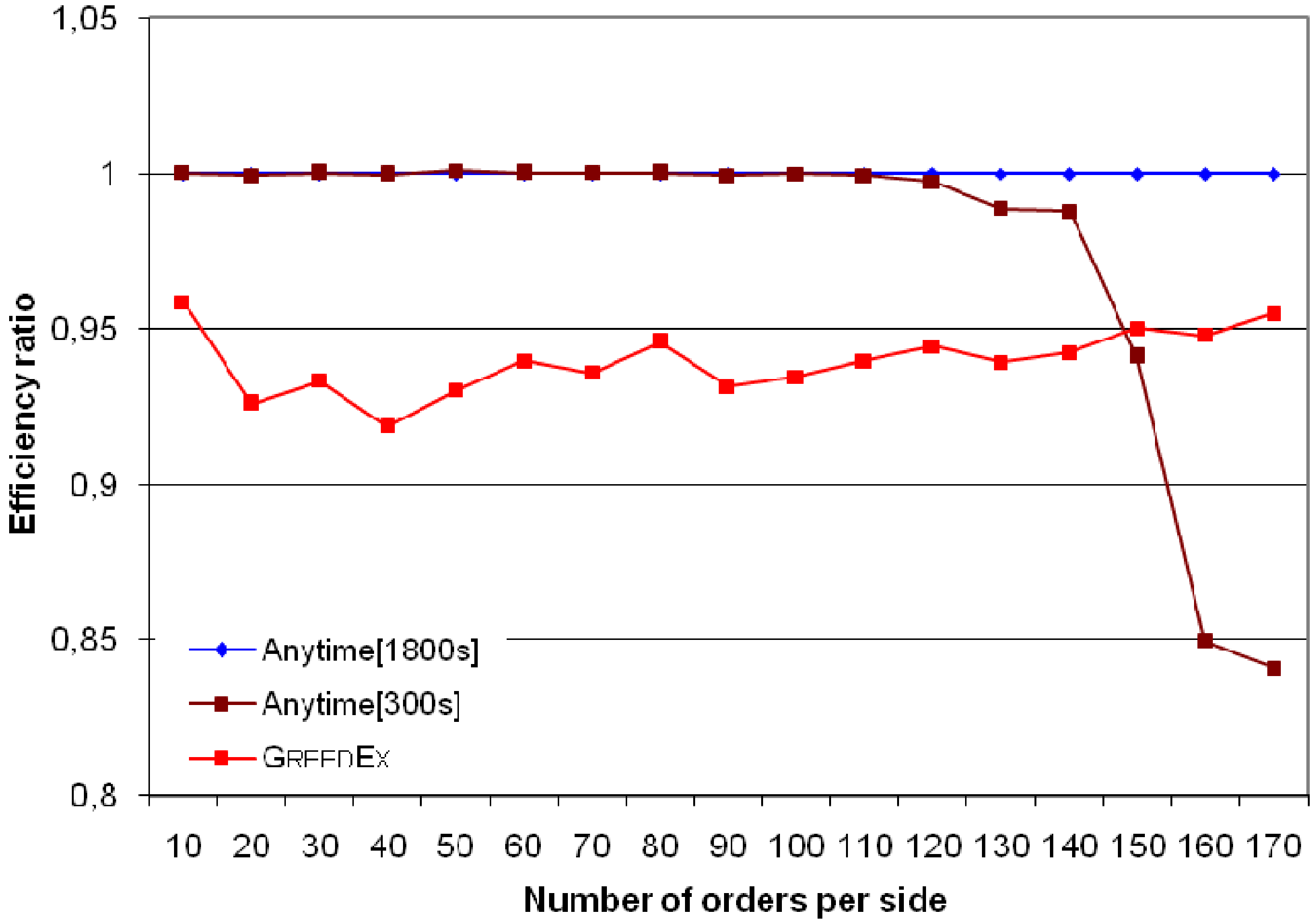
Preliminary results – runtime behavior



Preliminary results – runtime behavior



Preliminary results – efficiency ratio



Agenda

- Motivation
- Auction-based Scheduling
- GREEDEX – A Truthful and Scalable Auction Mechanism
- Preliminary Simulation Results
- Conclusion & Outlook

Conclusion

positive results	negative results
• preliminary simulation results showcase the speedy allocation scheme and high efficiency ratio • truthfulness of critical-value-based pricing is analytically proven	• there is no payment scheme which can complement the greedy allocation algorithm so as to generate truthful payments to resource providers • worst-case efficiency ratio can be made arbitrarily bad

Outlook

- analyze
  - alternative norms, pricing schemes
  - lookahead extensions, random-based ranking
  - strategic behavior
  - …
- run simulations with real workload traces from MOSIX
- design a true online mechanism

Thanks for your attention!

Contact Information

Jochen Stößer

mailto:stoesser@iism.uni-karlsruhe.de

IISM – Information & Market Engineering

Universität Karlsruhe (TH)
Englerstr. 14, 76 131 Karlsruhe, Germany

www.iism.uni-karlsruhe.de/stoesser
www.sorma-project.eu