THE ENTERPRISE COMPUTING CONFERENCE

EDOC2007 CONFERENCE

Annapolis, Maryland

An Economy Driven Resource Allocation Middleware for Grid Workflow

Pengcheng Xiong and Yushun Fan

Tsinghua University

11th International IEEE EDOC Conference (EDOC 2007)

Agenda

- Grid workflow background
- Performance analysis
- Resource allocation model and optimization algorithm
- Algorithm analysis and conclusion

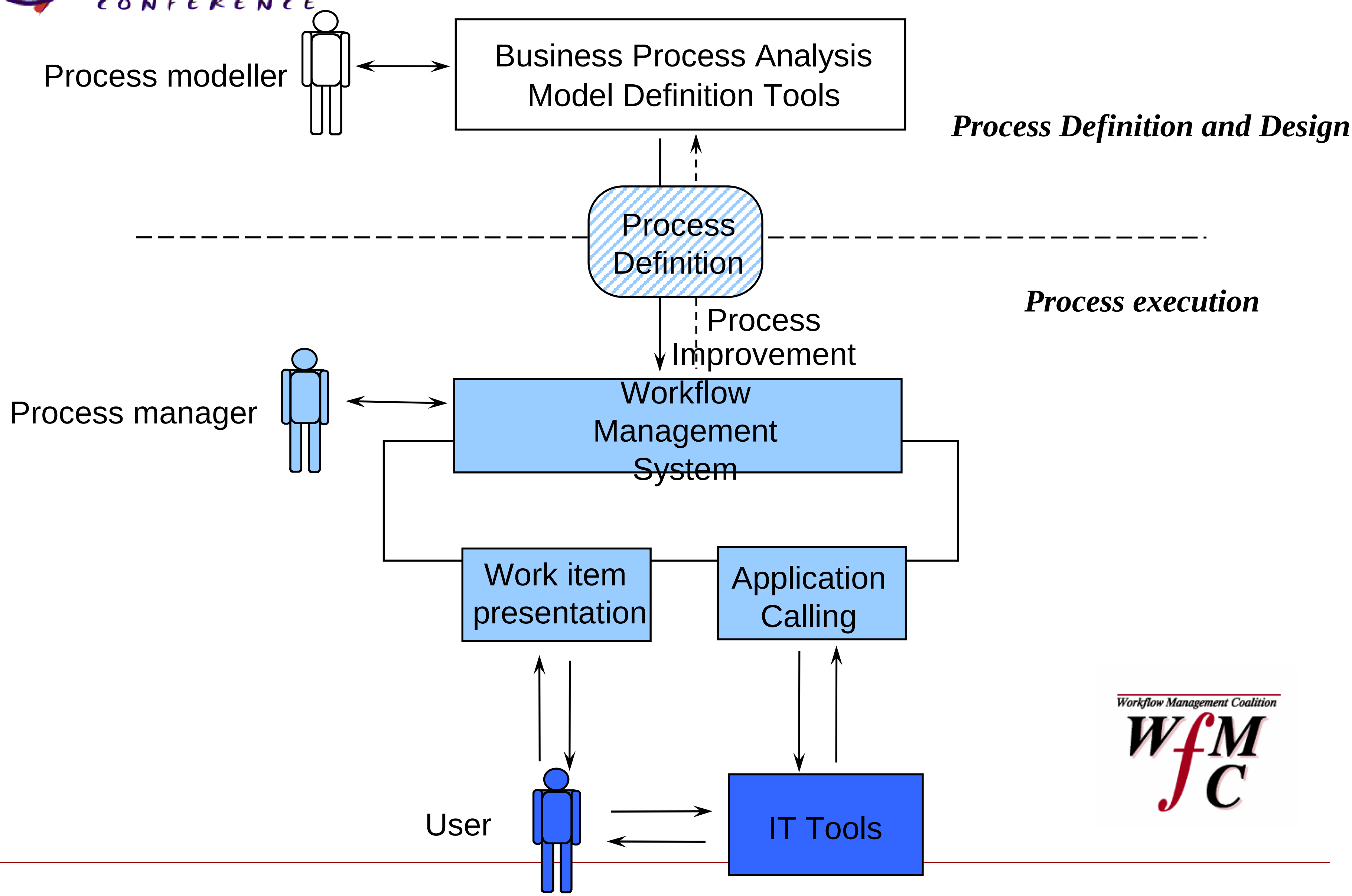
Agenda

- Grid workflow background
- Performance analysis
- Resource allocation model and optimization algorithm
- Algorithm analysis and conclusion

Introduction to grid workflow

- **Workflow**
  - **According to Workflow Management Coalition (WfMC), a workflow model is the formal representation of a business process in a form that supports automated manipulation.**
  - **It includes five relevant perspectives, i.e., process, resource, organization, information, and function perspectives, which are necessary for a workflow that can be automated by workflow management system (WfMS).**

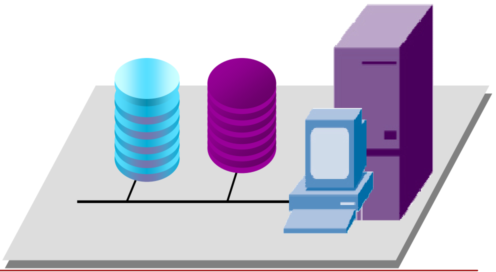
Workflow system



Introduction to grid workflow

One objective of Grid computing is to provide high performance computing power

- **Grid Workflow**
  - **A grid computing platform [I. Foster and C. Kesselman 1999] has evolved to become a promising technology for large-scale resource sharing and distributed system integration.**
  - ***Grid workflow* is defined as a composition of grid application services which execute on heterogeneous and distributed resources in a well-defined order to accomplish a specific goal.**



Agenda

- Grid workflow background
- Performance analysis
- Resource allocation model and optimization algorithm
- Algorithm analysis and conclusion

Performance analysis (1)

- **Grid workflow performance analysis focuses on evaluating the ability of the workflow to meet requirements with respect to some key performance indicators, such as:**
  - **Maximal parallelism**
  - **Throughput**
  - **Service levels**
  - **Sensitivity**
- **More specifically, analysis of *average turnaround time* is performed in this paper.**

Performance analysis (2)

- **Average turnaround time**
  - **Important**
    - **If a workflow processing turnaround time misses its deadline specified by users, exceptional actions such as compensating services which often introduce human intervention must be taken.**
  - **Complex**
    - **Related to the execution time of all the grid services involved**
    - **More resources-->shorter turnaround time**
    - **More resources-->more cost**
    - **Maximize return-on-investment !**

Literature review (1)

- **Grid computing point**
  - **Three alternative models, i.e., hierarchical, abstract owner, and market model for grid resource management architectures. [R. Buyyat, S. Chapin and D. DiNucci 2006]**
  - **Cheng *et al.* [C. Cheng, Z. Li, F. Huang and Z. Cui 2006] treat the resource cost and deadline as two separate optimization objectives.**
- **Workflow point**
  - **Li *et al.* [J. Li, Y. Fan and M. Zhou 2004] propose a computation algorithm for the lower bound of average turnaround time of workflow services through resource availability and workload analysis.**
  - **The optimization algorithm proposed by Li *et al.* has exponential complexity.**

Literature review (2)

- **Combination**
  - **Few of the above research have made substantial efforts in combining the internal execution process logic of gird workflow service nodes with a resource allocation algorithm.**
- **Efficiency**
  - **Catch up with the highly varying grid environment.**

- **Agenda**
  - **Grid workflow background**
  - **Performance analysis**
  - **Resource allocation model and optimization algorithm**
  - **Algorithm analysis and conclusion**

Extended market model (EMM) (1)

Extended market model (EMM) is a five tuple $(G, AO, R, F_{go}, F_{or})$ **, where:**

1) $G = \{a_1, a_2, ..., a_u\}$ is a set of grid workflow services, where a_i is a grid workflow service defined in grid workflow perspective;

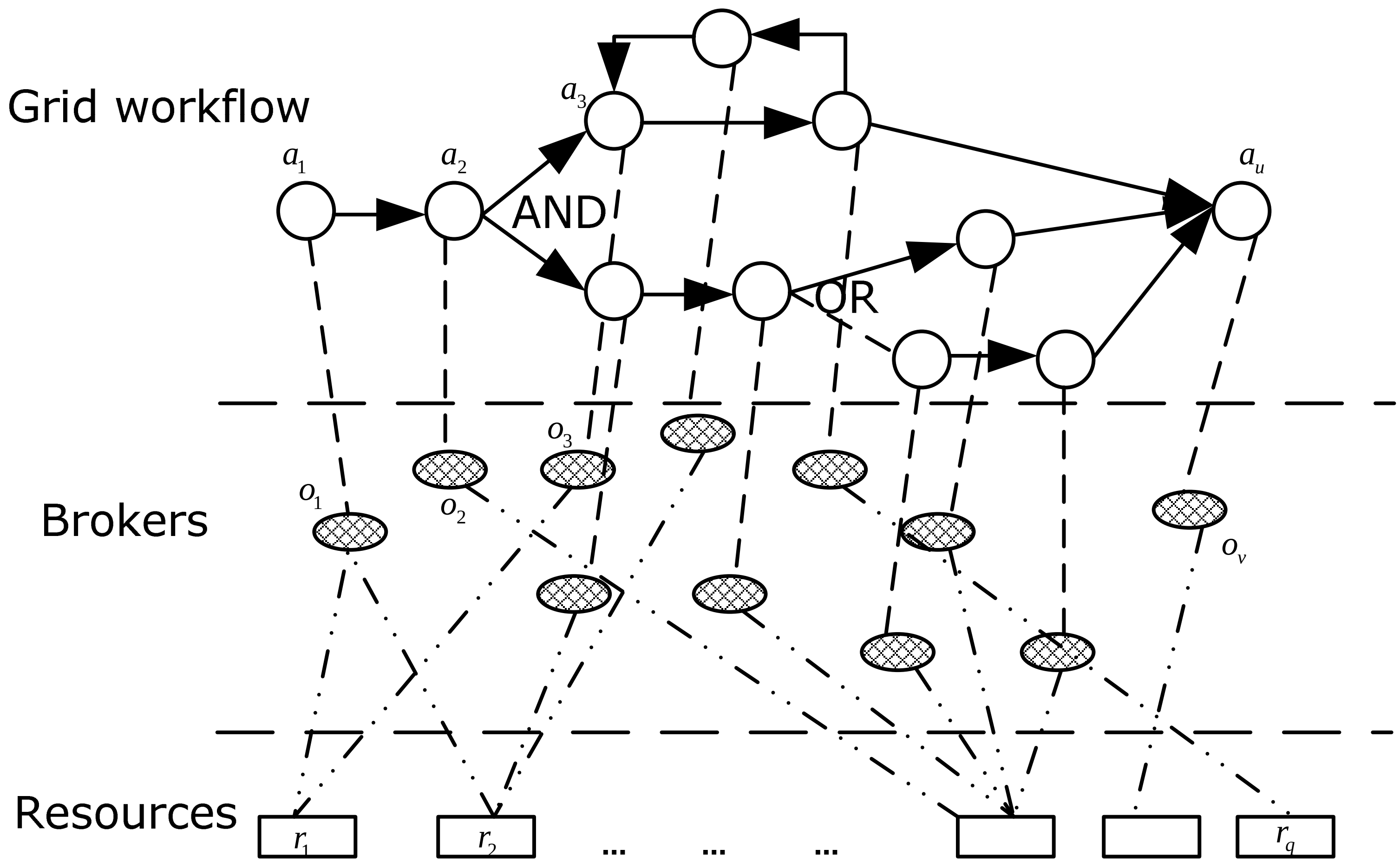
2) $AO = \{o_1, o_2, ..., o_v\}$ is a set of abstract owners, where o_i is an abstract owner defined in grid resource broker perspective;

3) $R = \{r_1, r_2, ..., r_q\}$ is a set of resource pools, where each r_i denotes a resource pool defined in the resource perspective;

4) $F_{go} \subseteq G \times AO$ denotes the mapping relation between grid workflow perspective and grid resource broker perspective;

5) $F_{or} \subseteq O \times R$ denotes the mapping relation between grid resource broker perspective and resource perspective;

Extended market model (EMM) (2)



Extended market model (EMM) (3)

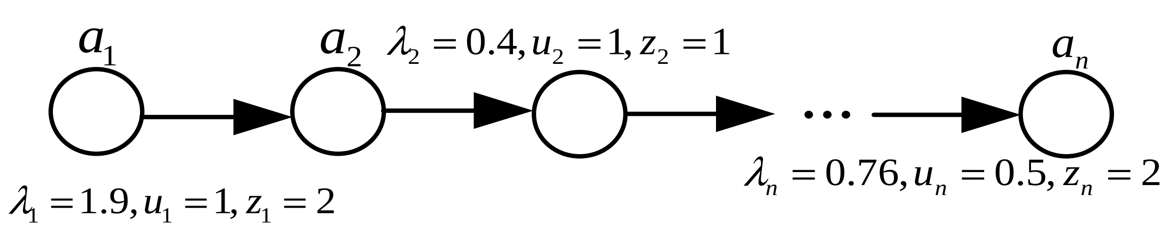
- **Assumption**
  - **suppose that the execution of grid workflow service *a* is appointed to abstract owner *o*.**
  - **abstract owner *o* is supported by n_1 individual resource agents of r_1, n_2 of r_2 ,…and n_k of r_k , which can be denoted as** $((r_1, n_1), (r_2, n_2), ..., (r_k, n_k))$
  - **the cost for an individual resource agent in resource pool can be computed as** $\sum_{i=1}^{k} n_i \cdot c_i$ **.**

Extended market model (EMM) (4)

- **Assumption**
  - **Each grid workflow service has exponential execution time**
  - **The arrival process of user's service requests is a Poisson process**
  - **The queue discipline is first come-first served**
- **We can therefore, model a grid workflow as an M/M/c queuing network, where each grid workflow service is an independent M/M/c queuing system.**

Critical path

- The critical path is a sequence of grid workflow services from the beginning to the end of a grid workflow that has the longest average execution time.
- Grid workflow control structures, i.e., sequential control structure (SCS), concurrent control structure (CCS), alternative control structure (ACS), and iterative control structure (ICS) affect critical path.
- We adopt the innermost control structure first (ICSF) method proposed by Son and Kim [J. H. Son and M. H. Kim 2001] to find the critical path of a grid workflow.



Problem model (1)

- Suppose the available budget as *C*, the problem is how to minimize the average execution time for the critical path of the grid workflow under the economic constraint.
- This problem can be modeled as a Multiple Choice Knapsack Problem (MCKP).

Given a set of items in several classes and a knapsack, where each item has a weight and profit, and the knapsack has a capacity, MCKP is to select one item from each class to be placed in the knapsack within the capacity yet to make the highest total profit.

Problem model (2)

Minimize the turnaround time of the critical path under the total cost constraint

Kinds of abstract owners → Classes in MCKP.

Number of extra abstract owners → Items in classes

Total cost of resource agents → Weight in MCKP

Zero minus the execution time → Profit of the items in MCKP

Heuristic algorithms

- **Genetic algorithm**
- **Tabu search**
- **Simulated annealing**
- **And so on…**
- **Disadvantages**
  - **Optimal solution? (not sure)**
  - **Complexity:** $O(\prod_{i=1}^{n}(M_i + 1))$

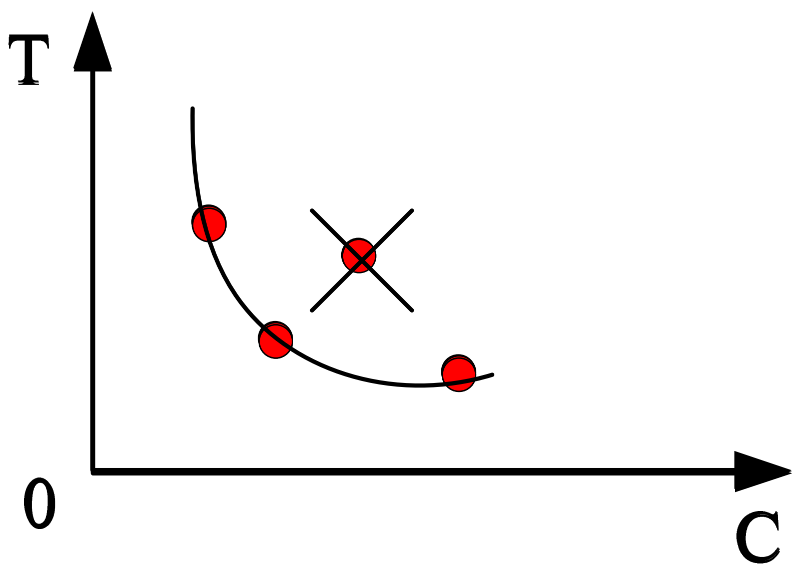
Pisinger algorithm (1)

- First solve the linear MCKP (LMCKP) by using a partitioning algorithm and derive an initial feasible solution (*initial core*) to MCKP.
- Then use dynamic programming to expand the initial core by adding new classes as needed.
- In this algorithm, a minimum number of classes are considered to solve MCKP and it uses the minimum effort for sorting and reduction.
- Advantages
  - Optimal solution? (sure)
  - Complexity: $O(\sum_{i=1}^{n}(M_i+1)+C\cdot\sum_{S_i\in Core}(M_i+1))$

Pisinger algorithm (2)

- **Step 1. Delete the LP-dominated solutions**

Definition. A feasible solution $j \in S_i$ is LP-dominated, iff $\exists j_1, j_2 \in S_i$, with $c_{ij_1} \le c_{ij} \le c_{ij_2}$ and $T_{ij_1} \ge T_{ij} \ge T_{ij_2}$ satisfy $\det(c_{ij} - c_{ij_1}, T_{ij_1} - T_{ij}, c_{ij_2} - c_{ij_1}, T_{ij_1} - T_{ij_2}) \le 0$



Pisinger algorithm (3)

- Step 2. Transform the MCKP to LMCKP
  - Relax the integrity constraint (4) to $0 \le x_{ij} \le 1$
  - Solve this LMCKP by using the algorithm proposed by Dyer [M.E. Dyer 1984]
  - Obtain the optimal solution b_i in each class S_i
  - If there exists and only exists a class S_a containing two fractional variables, we continue to the Step 3, otherwise end.

Pisinger algorithm (4)

- **Step 3. Choose and expand core.**
  - **Take the fractional class as the initial core.**
  - **Expand the core by alternatively including a new class.**

$$\lambda_i^{+} = Max_{j \in S_i, c_{ij} > c_{ib_i}} \frac{T_{ib_i} - T_{ij}}{c_{ij} - c_{ib_i}}, i = 1, 2, \ldots, k, i \neq a$$

$$\lambda_i^{-} = Max_{j \in S_i, c_{ij} < c_{ib_i}} \frac{T_{ij} - T_{ib_i}}{c_{ib_i} - c_{ij}}, i = 1, 2, \ldots, k, i \neq a$$

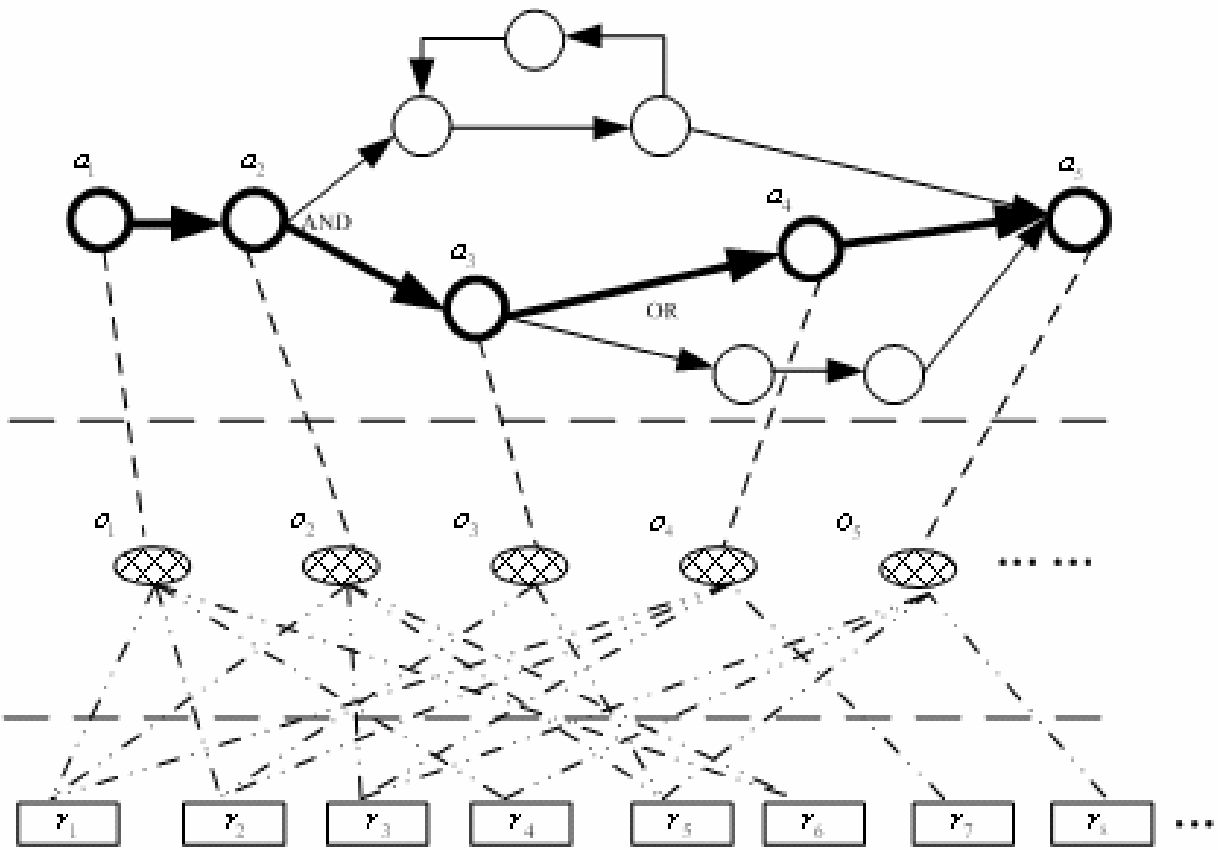
Pisinger algorithm (5)

- **Step 4. The result.**
  - **The optimal solution for MCKP is obtained after all the classes have been selected.**
  - **Mark the solution:**

$$u_i = \sum_{S_i \in Core} c_{iy_i} + \sum_{S_i \notin Core} c_{ib_i}$$

$$\pi_i = \sum_{S_i \in Core} T_{iy_i} + \sum_{S_i \notin Core} T_{ib_i}$$

Case study (1)



Graph representation of the EMM for the case study

Case study (2)

Table 1. Arrival rate λ_i, single abstract owner service rate u_i and minimal number of abstract owners z_i

Services on the critical path	Arrival rate λ_i	Single abstract owner service rate u_i	Minimal number of abstract owners z_i
a_1	2.00	1.10	2
a_2	0.55	1.00	1
a_3	1.43	0.80	2
a_4	1.97	1.15	2
a_5	0.76	0.50	2

Case study (3)

Table 2. Individual resource agent and price

Resource agent	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8
Price	1.20	2.00	1.50	0.80	4.00	5.00	3.00	0.5

Table 3. Mapping relationship F_{or}

Abstract owners	Individual resource agent	Total cost for abstract owners
o_1	$((r_1,2),(r_2,1),(r_4,3),(r_6,1))$	11.8
o_2	$((r_1,1),(r_3,2),(r_5,1),(r_6,2))$	18.2
o_3	$((r_2,2),(r_5,2))$	12
o_4	$((r_1,1),(r_2,4),(r_3,1),(r_7,2))$	16.7
o_5	$((r_3,2),(r_4,1),(r_5,2),(r_8,6))$	14.8

Case study (4)

Table 4. Total cost of abstract owners and the execution time for grid service $c_{ij}\,[T_{ij}]$

Grid Service	Add 0 abstract owner	Add 1 abstract owner	Add 2 abstract owners	Add 3 abstract owners	Add 4 abstract owners	Add 5 abstract owners	Add 6 abstract owners
a_1	23.6 [5.2381]	35.4 [1.1880]	47.2 [0.9642]	59.0 [0.9211]	70.8 [0.9116]	82.6 [0.9096]	94.4 [0.9092]
a_2	18.2 [2.2222]	36.4 [1.0818]	54.6 [1.0080]	72.8 [1.0007]			
a_3	24 [6.2124]	36 [1.6102]	48 [1.3212]	60 [1.2654]	72 [1.2532]	84 [1.2506]	
a_4	33.4 [3.2645]	50.1 [1.0847]	66.8 [0.9118]	83.5 [0.8785]	100.2[0.8714]		
a_5	29.6 [4.7348]	44.4 [2.3298]	59.2 [2.0626]	74.0 [2.0122]	88.8 [2.0022]		

Case study (5)

Table 4. Total cost of abstract owners and the execution time for grid service $c_{ij}[T_{ij}]$

Grid Service	Add 0 abstract owner	Add 1 abstract owner	Add 2 abstract owners	Add 3 abstract owners	Add 4 abstract owners	Add 5 abstract owners	Add 6 abstract owners
a_1	23.6 [5.2381]	35.4 [1.1880]	47.2 [0.9642]	59.0 [0.9211]	70.8 [0.9116]	82.6 [0.9096]	94.4 [0.9092]
a_2	18.2 [2.2222]	36.4 [1.0818]	54.6 [1.0080]	72.8 [1.0007]			
a_3	24 [6.2124]	36 [1.6102]	48 [1.3212]	60 [1.2654]	72 [1.2532]	84 [1.2506]	
a_4	33.4 [3.2645]	50.1 [1.0847]	66.8 [0.9118]	83.5 [0.8785]	100.2[0.8714]		
a_5	29.6 [4.7348]	44.4 [2.3298]	59.2 [2.0626]	74.0 [2.0122]	88.8 [2.0022]		

The budget constraint $C = 200$

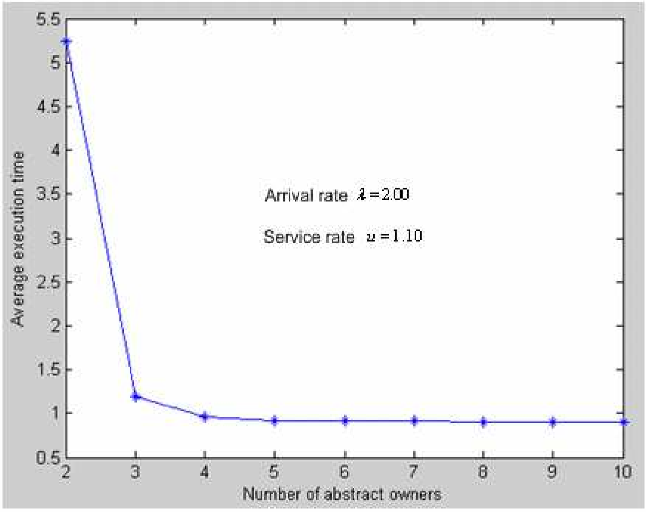
Result : $Y_{Core} = \{(196.1, 8.1459, 0)\}$ a_{1-5}=(3,1,4,3,3)

Agenda

- Grid workflow background
- Performance analysis
- Resource allocation model and optimization algorithm
- Algorithm analysis and conclusion

Algorithm Analysis

- **Characters**
  - **The average execution time will converge to** $1/u$
  - **Convex**
- **Results:**
  - **Non LP-dominated**
  - **Select** $b_i + 1$ **and 0 for calculating** λ_i^+ **and** λ_i^- **respectively.**



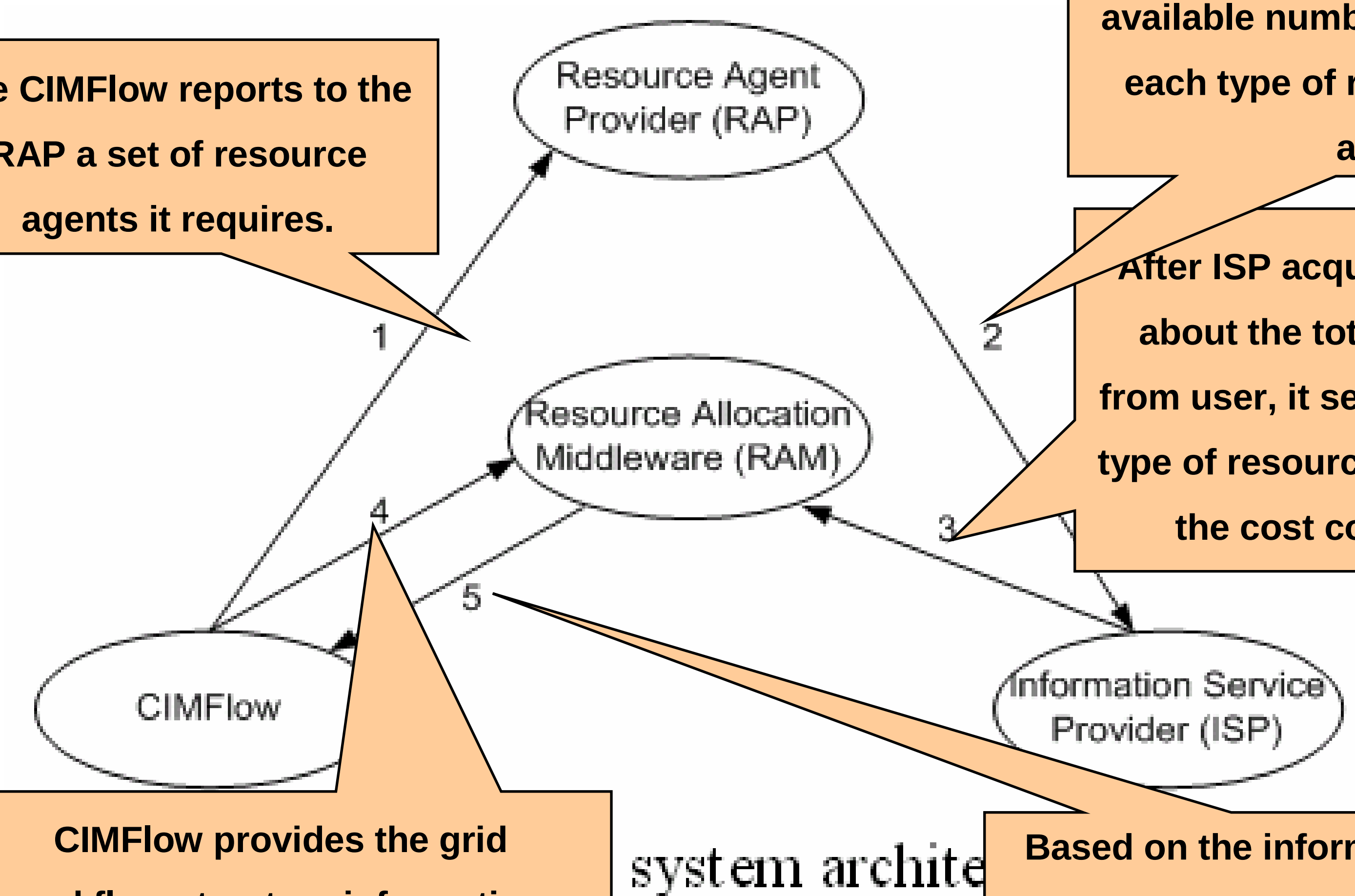
Average execution time for a_1 with different number of abstract owners

Prototype system

The CIMFlow reports to the RAP a set of resource agents it requires.

The RAP submits to the ISP the available number and the price of each type of required resource agent.

After ISP acquires the information about the total resource budget from user, it sends the price of each type of resource agent together with the cost constraint to RAM.



CIMFlow provides the grid workflow structure information, instance arrival rate and single abstract owner service rate to RAM.

Based on the information, the RAM builds the extended market model. Send back the results.

Conclusion

- This paper proposes extended market model, which connects the grid workflow performance with the economy factors.
- We model the minimum workflow turnaround time under cost constraint problem as MCKP and adopt efficient Pisinger's algorithm.
  - Compared with the performance of other algorithms, the Pisinger's algorithm is more suitable in the dynamic grid environment especially when the grid workflow is very complex and the abstract owner choice set is fairly large.
  - The problem's convexity characteristic provides additional convenience of using the algorithm.
- The proposed concepts and algorithms can be readily put into industrial applications.

Future works

- How to increase computational resource sharing rate since we assume that every grid workflow service consumes its resources by abstract owner exclusively in this paper.
- However, the remove of this constraint will damage the logical and temporal correctness of grid workflow, and make it more complex.

References

[I. Foster and C. Kesselman] I. Foster and C. Kesselman (editors), The Grid Blueprint for a Future Computing Infrastructure, Morgan Kaufmann Publishers, USA, 1999.

[R. Buyyat, S. Chapin and D. DiNucci 2006] R. Buyyat, S. Chapin and D. DiNucci, "Architectural Models for Resource Management in the Grid", Proc. of the First IEEE/ACM Internation Workshop on Grid Computing, Bangalore, India, 2000, pp. 1-13.

[C. Cheng, Z. Li, F. Huang and Z. Cui 2006] C. Cheng, Z. Li, F. Huang and Z. Cui, "Resource allocation based on multi-objective programming in economy grid: design and performance evaluation", Proc. of the 5th International Conference on Grid and Cooperative Computing, Changsha, China, pp. 38-41, Oct. 2006.

[J. Li, Y. Fan and M. Zhou 2004] J. Li, Y. Fan and M. Zhou, "Performance modeling and analysis of workflow", IEEE Transactions on Systems, Man and Cybernetics, Part A, Vol. 34, Iss. 2, Mar. 2004, pp. 229-242.

[J. H. Son and M. H. Kim 2001] J. H. Son and M. H. Kim, "Improving the performance of time-constrained workflow processing", J. Syst. Softw., Vol. 53, 2001, pp. 211-219.

[M.E. Dyer 1984] M.E. Dyer, "An O(n) algorithm for the multiple-choice knapsack linear program", Mathematical Programming, 29, pp. 57-63.

Thank you!
Questions and Answers