Heuristic-based Task Selection and Allocation Framework in Dynamic Collaborative Cloud Service Platform

Biao Song, Mohammad Mehedi Hassan, Eui-Nam Huh

CloudCom 2010

Presented By:
Prof. Eui-Nam Huh

Dep. of Computer Eng.
Kyung Hee University

Kyung Hee University

Agenda

- **Background of Dynamic Cloud Collaboration**
- **Task Selection and Allocation Problem**
- **New Metric to Improve Resource Utilization**
- **Heuristic Algorithm**
- **Simulation Result**
- **Conclusion**

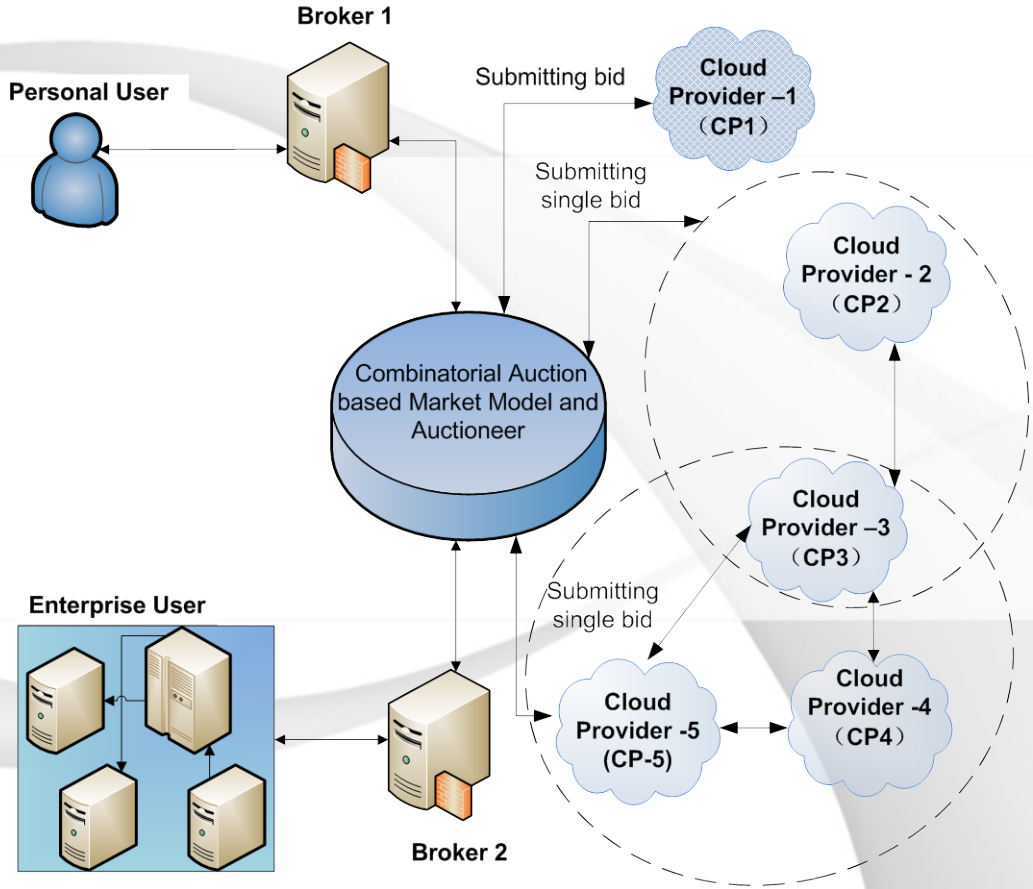
Agenda

- **Background of Dynamic Cloud Collaboration**
- **Task Selection and Allocation Problem**
- **New Metric to Improve Resource Utilization**
- **Heuristic Algorithm**
- **Simulation Result**
- **Conclusion**

Dynamic Collaborative Cloud Platform

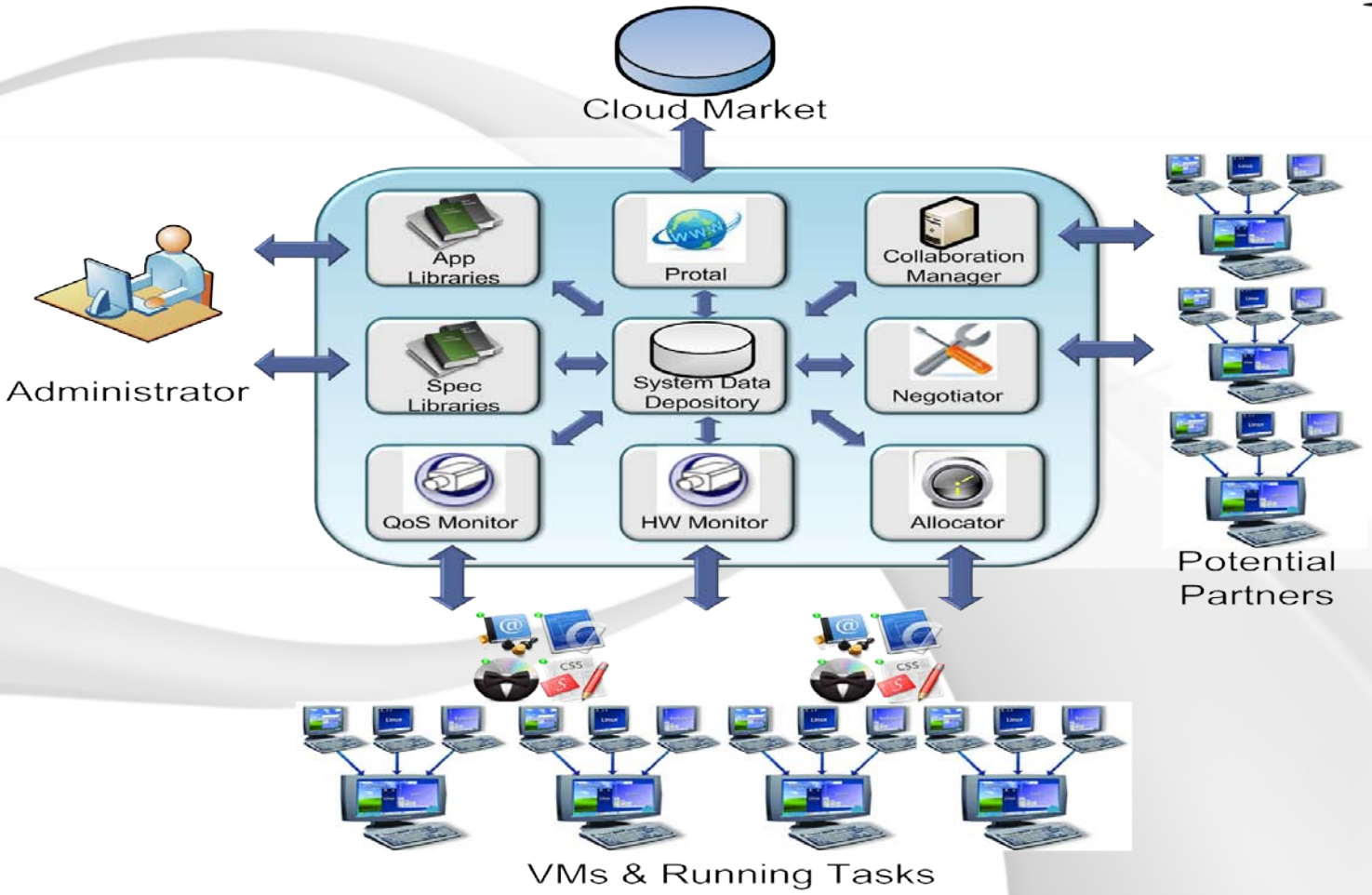
- To address scalability & interoperability, Dynamic collaboration among Cloud providers is a key issue
- Consumers want to simultaneously use multiple cloud services
- Cloud providers have specialization in their resource and service supplement
- Dynamic Collaborative Cloud platform facilitates offering collaborative or portable cloud services to consumers

Market Model



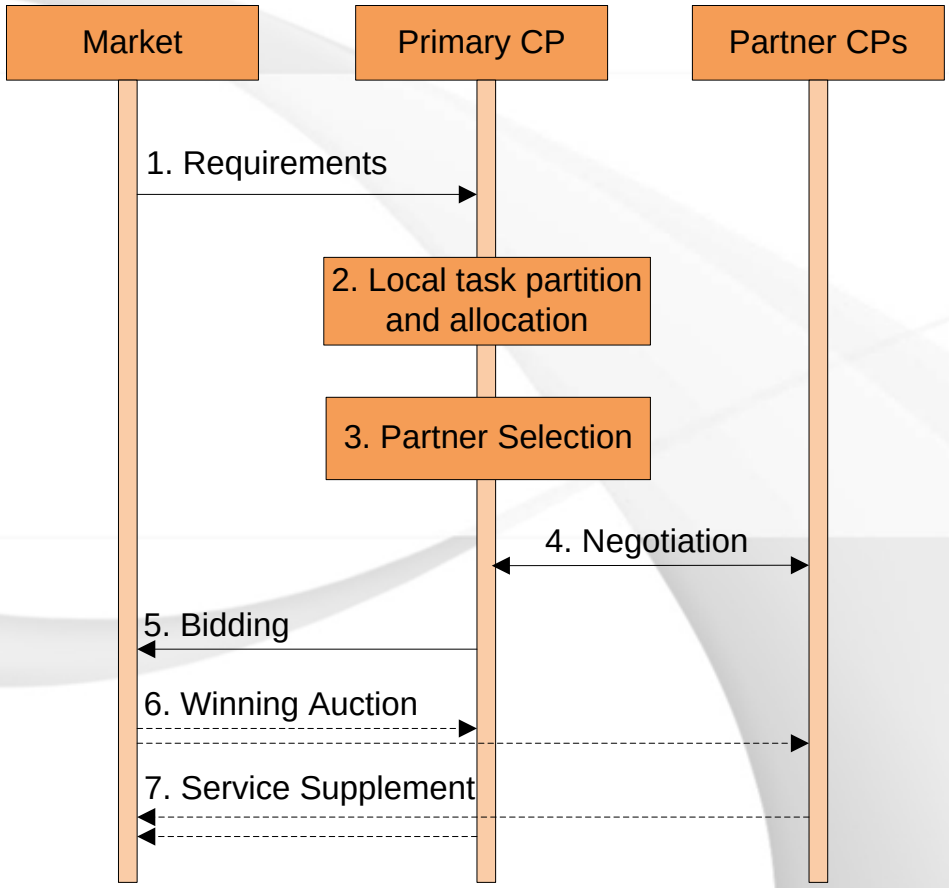
Dynamic Collaborative Cloud Market

Cloud Manager



Overall Architecture of Cloud Manager

Workflow of Dynamic Collaboration



Flowchart of dynamic collaboration

Task Selection and Allocation Problem

- **Cloud providers (CPs) are motivated by charging consumers for using their resources and services**
- **Key Issue**
  - Improving resource utilization
- **As CPs adopt virtualization technology, the task allocation problem becomes more challenging**
- **In collaborative environment, CPs also face task selection problem**

Agenda

- Background of Dynamic Cloud Collaboration
- Task Selection and Allocation Problem
- New Metric to Improve Resource Utilization
- Heuristic Algorithm
- Simulation Result
- Conclusion

Low Resource Utilization Scenario

Table.1 Resource of Primary Cloud Provider

Physical Machine	CPU	Memory	Hard Disk	Network Bandwidth
P1	1000	800	700	500
P2	900	700	650	500
P3	800	600	600	500

Table.2 Resource Requirement of Combinatorial Tasks

Task	CPU	Memory	Hard Disk	Network Bandwidth
T1	450	400	400	200
T2	400	200	400	200
T3	100	300	500	400
T4	200	200	300	300
T5	350	500	100	500
T6	250	450	450	250
T7	500	200	250	300
T8	400	300	100	100
T9	200	200	500	150
T10	350	300	200	450

Table.3 Allocation Process (using CPU-based Min-Min)

Allocated Task	Physical Machine	Remaining Resource on Physical Machine				
		CPU	Memory	Hard Disk	Bandwidth	
T3	P1	900	500	200	100	
T4	P2	700	500	350	200	Low Utilization
T9	P3	600	400	200	450	
T8	P1	500	200	100	0	Low Utilization
T10	P3	250	100	0	0	Full Utilization

Key Point: Resource BalancingTask

- **In each physical machine, CPU, memory, hard disk and network bandwidth resources should be uniformly utilized**
- **It helps to improve the overall resource utilization for Cloud providers**
- **It can be achieved by using new metric when Cloud providers select and allocate tasks**

Agenda

- Background of Dynamic Cloud Collaboration
- Task Selection and Allocation Problem
- New Metric to Improve Resource Utilization
- Heuristic Algorithm
- Simulation Result
- Conclusion

New Metric to Improve Resource Utilization

- **Standard deviation is a simple but quite effective metric to tackle the resource balancing issue:**

Averaging Utilization:

$$au = (fc + fm + fh + fb)/4$$

Balancing Condition:

$$bc = ((au - fc)^2 + (au - fm)^2 + (au - fh)^2 + (au - fb)^2)^{1/2}$$

fc, *fm*, *fh* and *fb* denote the utilization percentage of CPU, memory, hard-disk and network bandwidth respectively

Tradeoff-metric

- **Existing metrics also have many advantages, an existing metric could be:**
  - Suppose a task t_i allocated on p_j can get

$$c_{ij} + fc_j * (c_{ij} / (1- fc_j)) = c_{ij} / (1- fc_j)$$

    CPU capacity from p_j, where c_{ij} is the minimum CPU capacity required for running t_i on p_j
  - For t_i, the superiority of p_j is defined as

$$S_{ij} = c_{ij} /(c_{ij} / (1- fc_j))= 1- fc_j$$

- **Combine both metrics using trade-off method**

$$T(i, j)= \varphi_{ij} * S_{ij} + (1- \varphi_{ij})* bc_{ij}$$

  - φ_{ij} is related to the condition of remaining resources

Agenda

- Background of Dynamic Cloud Collaboration
- Task Selection and Allocation Problem
- New Metric to Improve Resource Utilization
- Heuristic Algorithm
- Simulation Result
- Conclusion

Heuristic Algorithms

- **The trade-off metric can be directly applicable for many existing static or dynamic heuristic allocation algorithm, such as Min-Min, Max-Min, UDA or Sufferage**
- **However, low resource utilization problem has not been fully solved since the task selection issue need to be addressed**

Threshold Method

- **By setting proper threshold value for *Balancing* metric, we can prevent improper task selection from happening**
- **Define ξ as the scalable threshold value, any single allocation which causes $bc_{ij} > \xi$ is not acceptable**

Choosing Threshold Value

- **In static environment, we assume the resource requirements of coming tasks are predictable.**
  - By adopting different threshold value and iteratively run the heuristic algorithms, we can find the optimal value to maximize the overall resource utilization

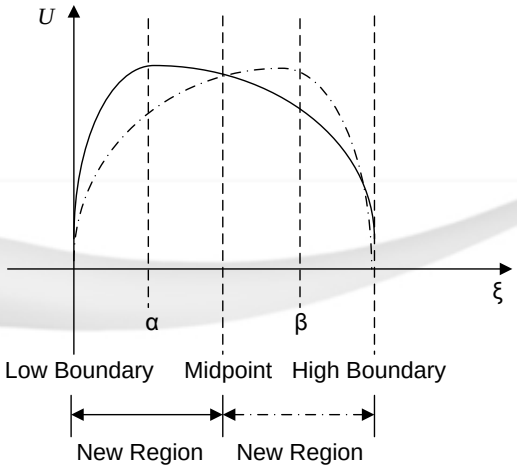


Fig.4(a) The peak locates in the side part

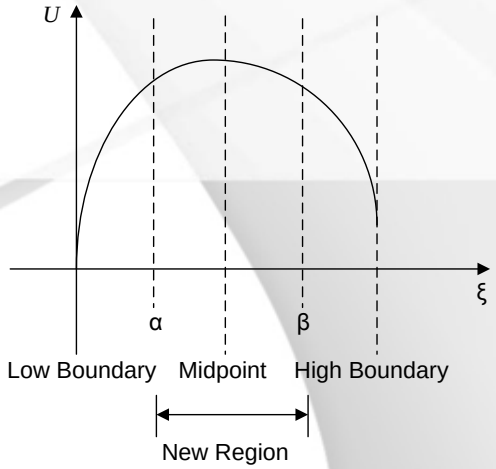


Fig.4(b) The peak locates in the middle part

  - In dynamic environment, the threshold value can be selected in a reactive way

Agenda

- Background of Dynamic Cloud Collaboration
- Task Selection and Allocation Problem
- New Metric to Improve Resource Utilization
- Heuristic Algorithm
- Simulation Result
- Conclusion

Impact of Proposed Approach

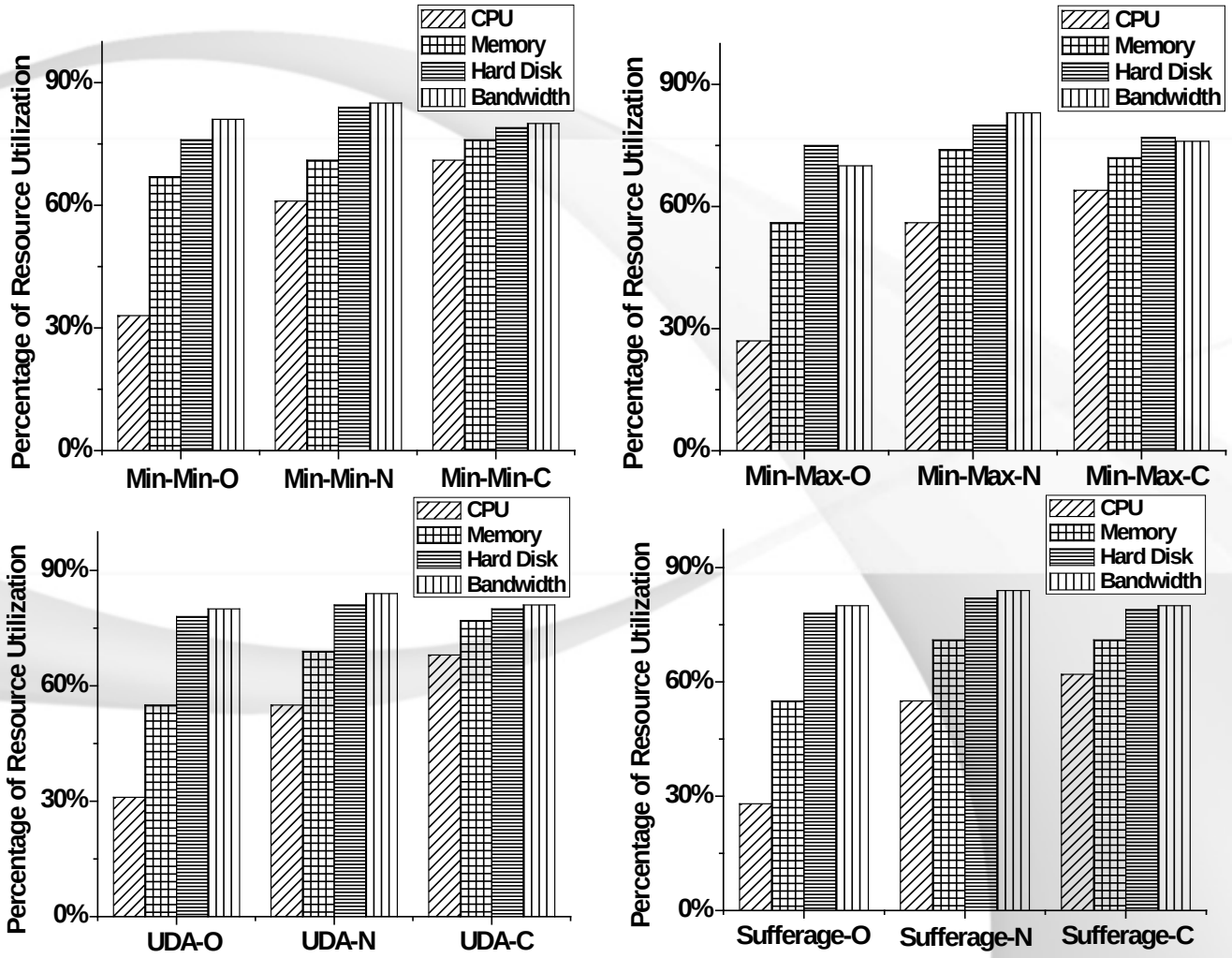
Table.3 Allocation Process (using CPU-based Min-Min)

Allocated Task	Physical Machine	Remaining Resource on Physical Machine			
		CPU	Memory	Hard Disk	Bandwidth
T3	P1	900	500	200	100
T4	P2	700	500	350	200
T9	P3	600	400	200	450
T8	P1	500	200	100	0
T10	P3	250	100	0	0

Table.4 Allocation Process (using proposed approach)

Allocated Task	Physical Machine	Remaining Resource on Physical Machine			
		CPU	Memory	Hard Disk	Bandwidth
T4	P1	800	600	400	200
T2	P3	400	400	200	300
T7	P2	300	400	350	200
T1	P1	350	200	0	0
T8	P3	0	100	100	200

Simulation Results



Overall Resource Utilization

Agenda

- Background of Dynamic Cloud Collaboration
- Task Selection and Allocation Problem
- New Metric to Improve Resource Utilization
- Heuristic Algorithm
- Simulation Result
- Conclusion

Conclusion

- To make dynamic collaborative Cloud service platform truly applicable, we propose a general heuristic-based task selection and allocation framework
- Our main target is to improve the overall resource utilization for Cloud providers
- Considering the specialty of Cloud environment, we address the balancing issue by developing a new metric and apply it into existing heuristic-based allocation algorithms
- By adding scalable threshold value, task selection is enabled to enhance heuristics
- Simulation results prove that our proposed approach achieves better resource utilization

Thank you

Email: bsong@khu.ac.kr