Towards Makespan Minimization Task Allocation in Data Centers

Kangkang Li, Ziqi Wan, Jie Wu and Adam Blaisse

Temple University, Philadelphia, PA, USA

Outline

1. Introduction
2. Problem Formulation
3. A One-layer Cluster Study
4. Multi-layer Cluster Study
5. Evaluations
6. Conclusions and Future work

Introduction

- Data Centers:

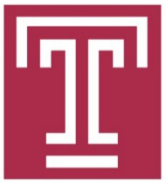
  ----Servers with identical computing capabilities

  ----Links connected by binary tree-topology structure

- Tasks：

  ----computation workloads

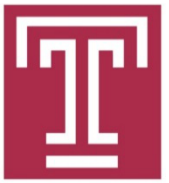
  ----communication traffic



Task Model

Tasks: 3 steps

Notice that communication between two tasks cannot start before both tasks finish their 1st step, generating waiting time between 1st step and 2nd step



Communication and Computation Model

- Communication Model:

  ---The bandwidth of the link is evenly shared by tasks running under that link

  ---the more tasks communicating through the same link, the lower the bandwidth each task pair can be allocated.

- Computation Model:

  ---The computing capability of a server is evenly shared by tasks running on it.

  ---the more tasks, the lower the computing capability each task can be allocated.

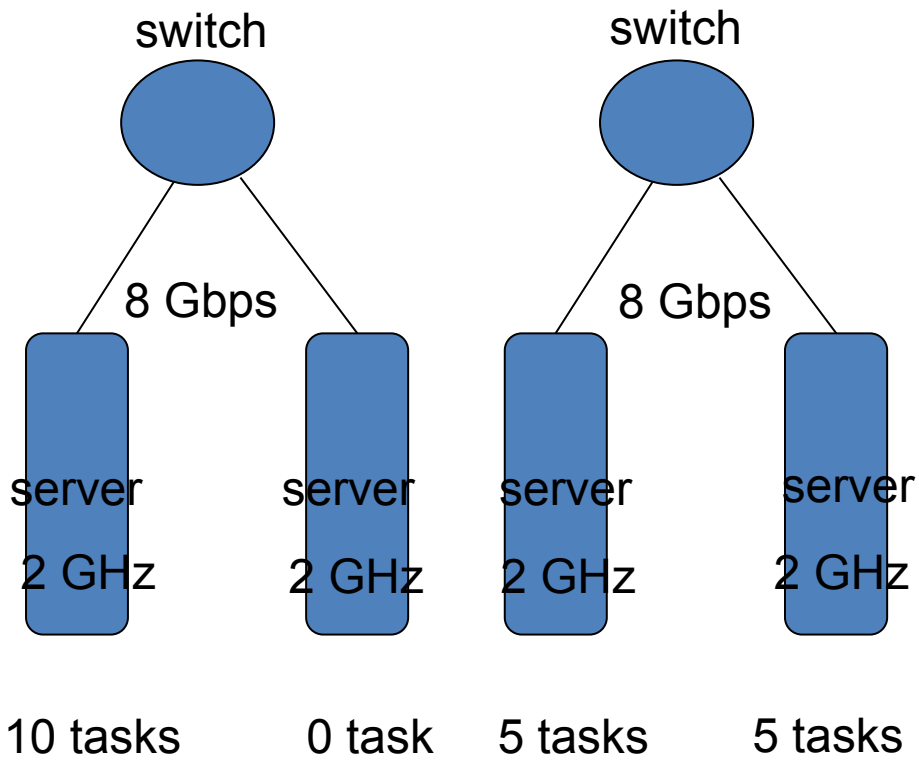
Motivational Example

- Objective:

  find am allocation scheme to minimize the avargage makespan of input tasks
- Computation dominant

  communicaton could be ignored, load-balancing is the best allocation scheme
- Communication dominant

  computation workloads could be ignored, locality is more important



Tradeoff between locality and load-balancing!

Outline

1. Introduction
2. Problem Formulation
3. A One-layer Cluster Study
4. Multi-layer Cluster Study
5. Evaluations
6. Conclusions and Future work

Problem Formulation

- K-layer bianry data center with M servers at the bottom
- semi-homogeneous configuration
- Each server has indential computing capability

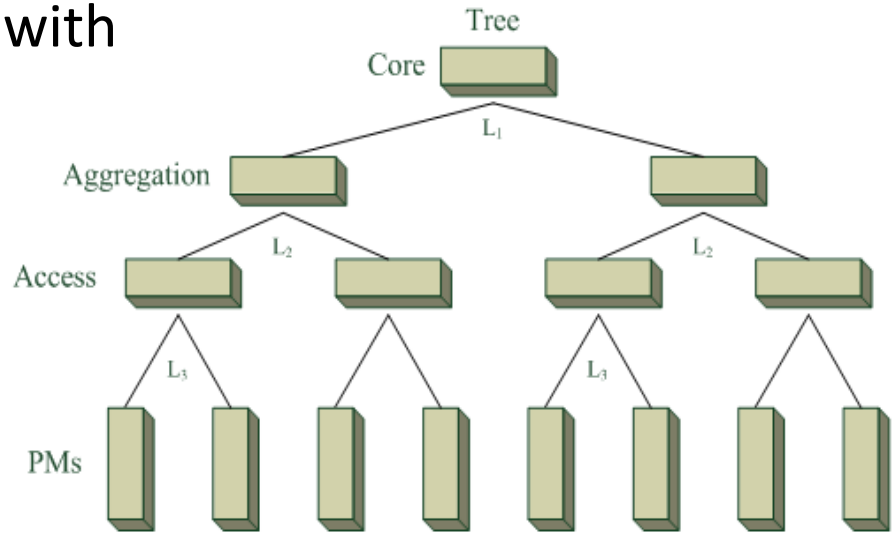


Fig. 2. Tree-based network topology

- Each link of the same layer has the same bandwidth capacity
- Upper layer link capacity is twice as the connected lower layer link to reduce upper-layer congestion.

Problem Formulation

- Objective:

- Find the best task allocation to minimize the average makespan of tasks

- makespan(i) = pre(i)+wait(i)+commun(i)+post(i)

- Average makespan is

$$\overline{makespan} = \frac{\sum_i^N makespan(i)}{N}$$

Outline

1. Introduction
2. Problem Formulation
3. A One-layer Cluster Study
4. Multi-layer Cluster Study
5. Evaluations
6. Conclusions and Future work

A One-layer Cluster Study

•By enumerating all possible allocation choices to get optimal allocation

- Enumerate from 0 to N/2
  - N is the number of tasks

•One-layer time complexity: O(N)

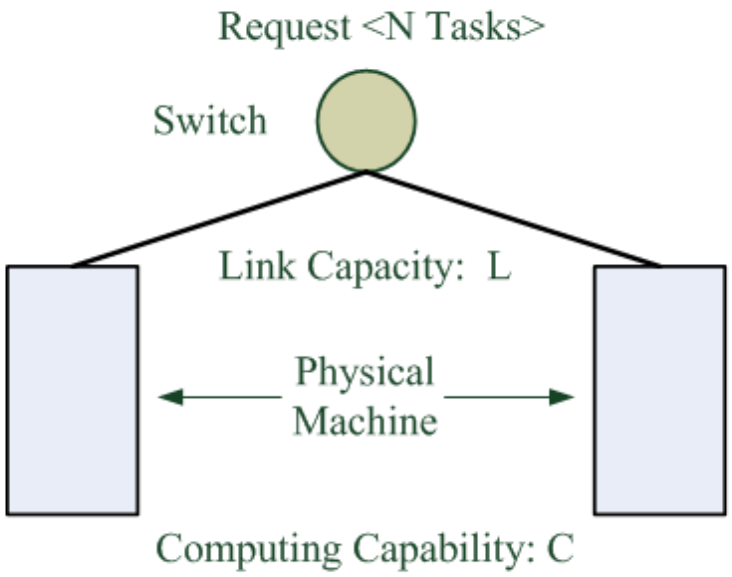


Fig. 3: One-layer cluster

Outline

1. Introduction
2. Problem Formulation
3. A One-layer Cluster Study
4. Multi-layer Cluster Study
5. Evaluations
6. Conclusions and Future work

Multi-layer Cluster Study

- Multi-layer Cluster
- ----Resursive abstraction (bottom to top) : O(M)
- ----Hierarchical task allocation (top to bottom): O(KN)

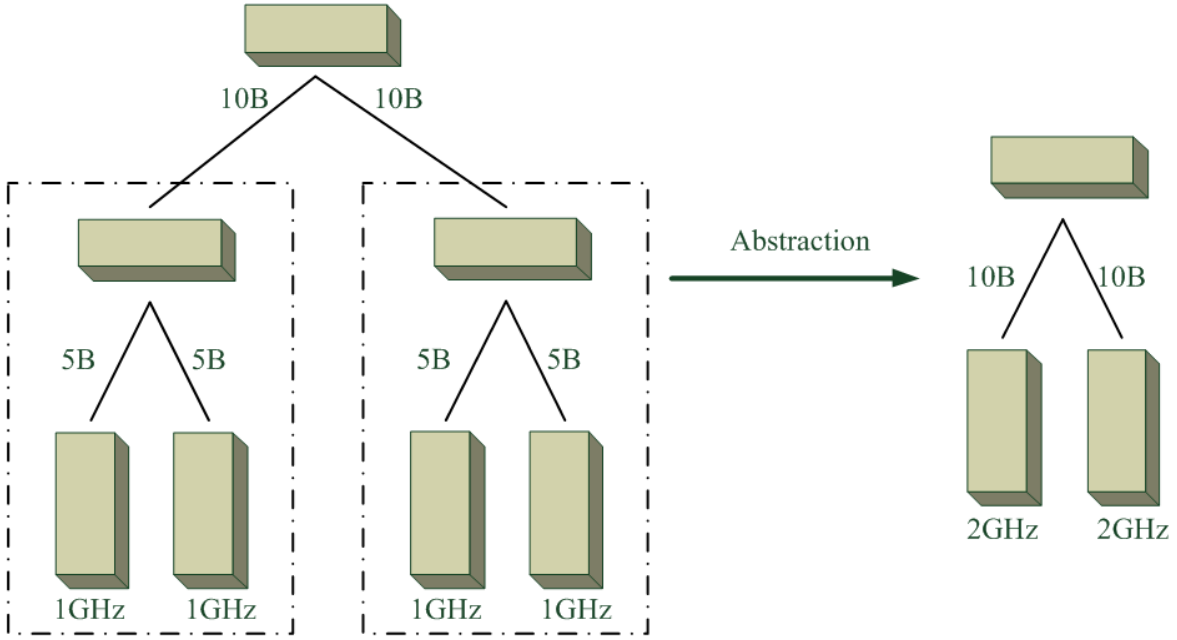


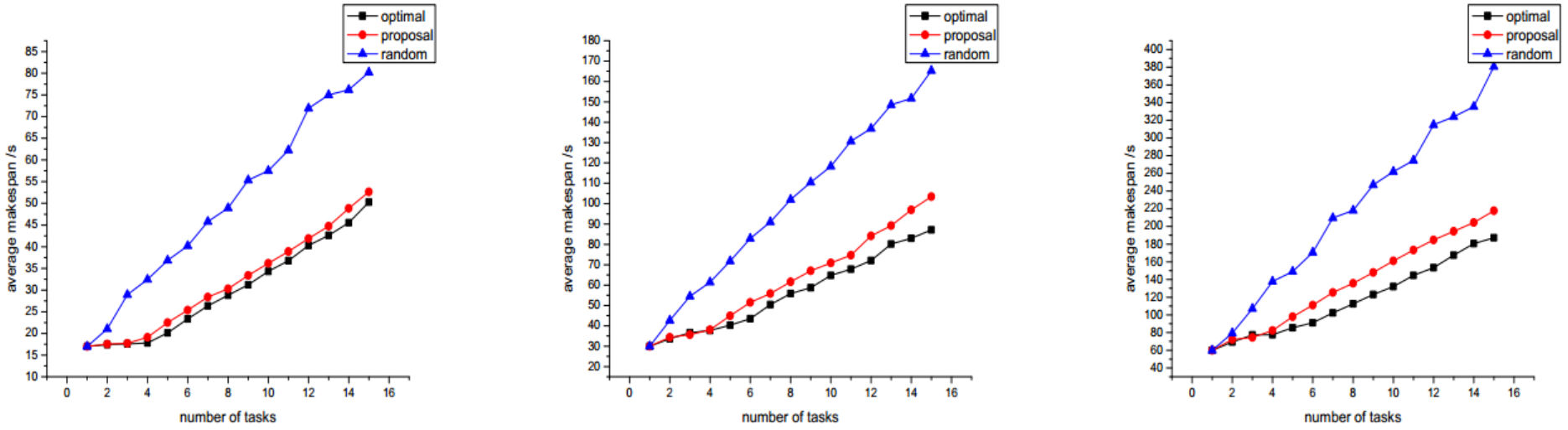
Fig. 4: Abstraction process

- The time complexity is O(KN + M)

Outline

1. Introduction
2. Problem Formulation
3. A One-layer Cluster Study
4. Multi-layer Cluster Study
5. Simulations
6. Conclusions and Future work

Simulations

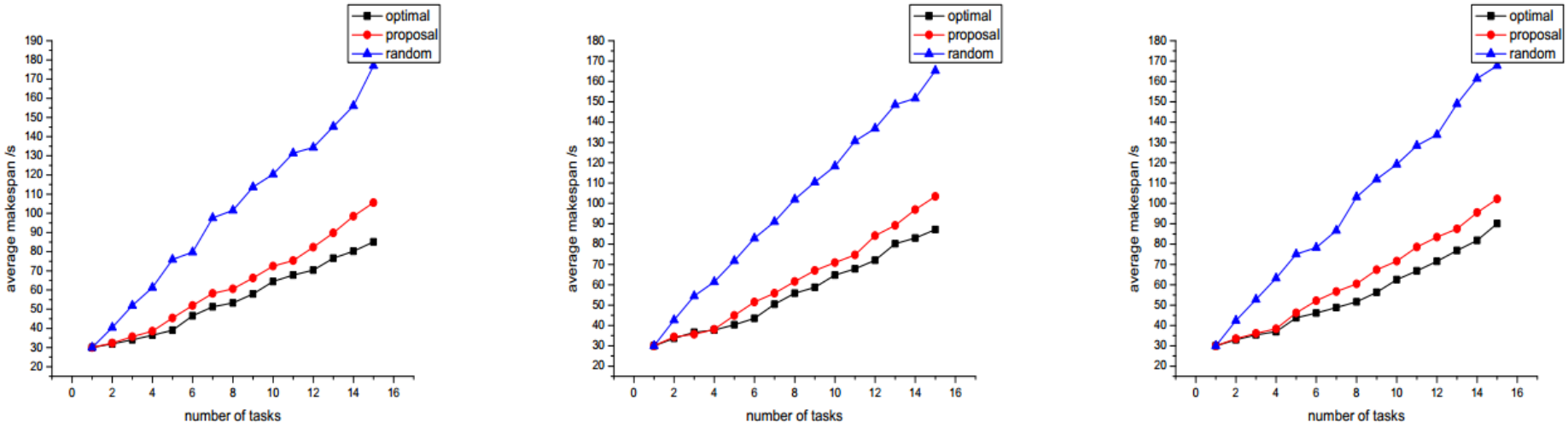


(a) $\alpha = 2, \beta = 2, \gamma = 15$ (b) $\alpha = 15, \beta = 2, \gamma = 15$ (c) $\alpha = 45, \beta = 2, \gamma = 15$

Fig. 5: Performance comparisons of average makespan vs. the value of α

For different pre-computation times (α), as shown in Figure 5, our algorithm perform is very close to optimal.

Simulations(cont'd)

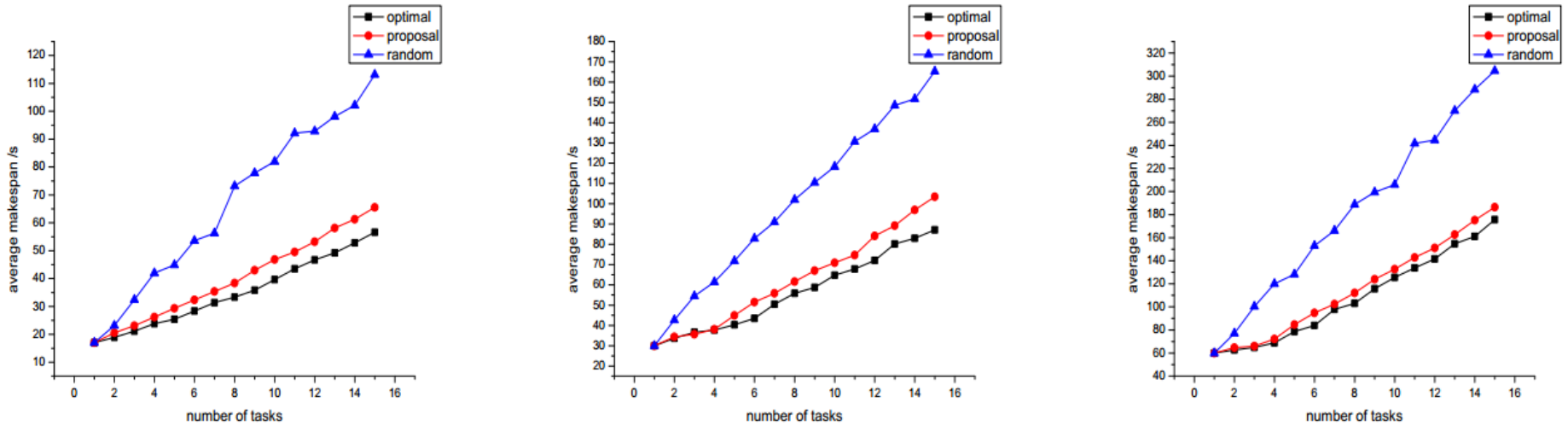


(a) $\alpha = 15, \beta = 1, \gamma = 15$ (b) $\alpha = 15, \beta = 2, \gamma = 15$ (c) $\alpha = 15, \beta = 4, \gamma = 15$

Fig. 6: Performance comparisons of average makespan vs. the value of β

For different communication times (β), as shown in Figure 6, our algorithm perform is very close to optimal.

Simulations(cont'd)



(a) $\alpha = 15, \beta = 2, \gamma = 2$ (b) $\alpha = 15, \beta = 2, \gamma = 15$ (c) $\alpha = 15, \beta = 2, \gamma = 45$

Fig. 7: Performance comparisons of average makespan vs. the value of γ

For different post-computation times (γ), as shown in Figure 7, our algorithm perform is very close to optimal.

Outline

1. Introduction
2. Problem Formulation
3. A One-layer Cluster Study
4. Multi-layer Cluster Study
5. Simulations
6. Conclusions and Future work

Conclusions

- In this paper, we study the classic task allocation problem in data centers.
  - We study the tradeoff between locality and load balancing when do task allocation.
  - We firstly study the one-layer cluster and discuss the optimal solution.
  - After that, we propose our hierarchical task allocation algorithm to deal with the multi-layer cluster.
  - Simulations result validate the efficiency of our algorithm.

Future Work

- In this paper,we only study the homogeneous task model under the semi-homogeneous data center configuration. In our future work, we will:

- Firstly, we will extend the task model into heterogeneous settings. That is, each task will have different values of α , β , and γ .

- Furthermore, the heterogeneous configuration of data centers will also be studied.

- Lastly, We will evaluate the efficiency of algorithm under the above heterogeneous scenarios both theoretically and experimentally.

Thank you!

Questions?

kang.kang.li@temple.edu