Online Flow Scheduling with Deadline for Energy Conservation in Data Center Networks

Biyu Zhou[1], Jie Wu[2], Lin Wang[3], Fa Zhang[1], and Zhiyong Liu[1]
[1]Institute of Computing Technology, Chinese Academy of Sciences
[2]Temple University [3]Technische Universitat Darmstadt

Outline

- Backgrounds
- Problem
- Solutions
- Evaluation

Backgrounds

an annual growth of 12%

(U.S.Environmental Protection Agency)

Benefits:

- Economics:

  Reduce bills for energy consumption.

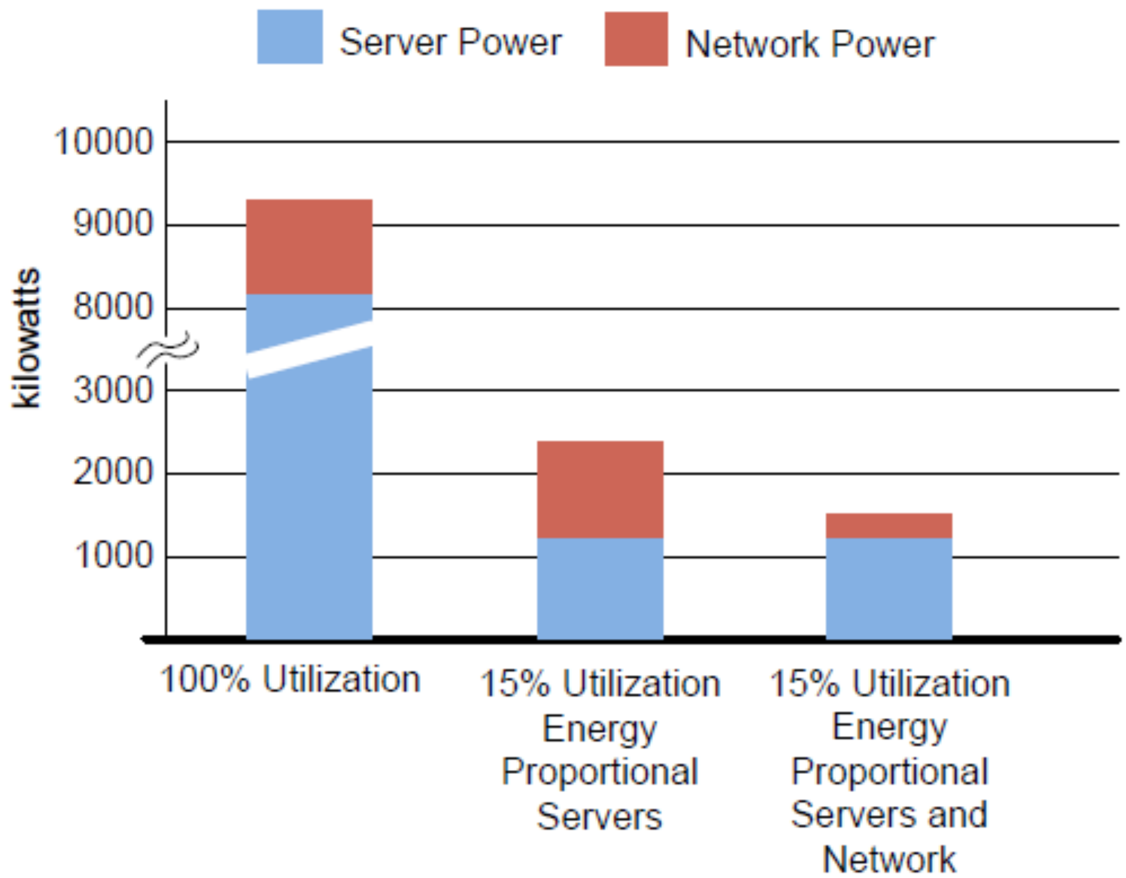
- Environmental Protection:

  Reduce carbon emission.

Data Centers

中国科学院计算技术研究所
INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES

Backgrounds



Reducing the energy consumption has becoming a key component in building green datacenters!

Backgrounds

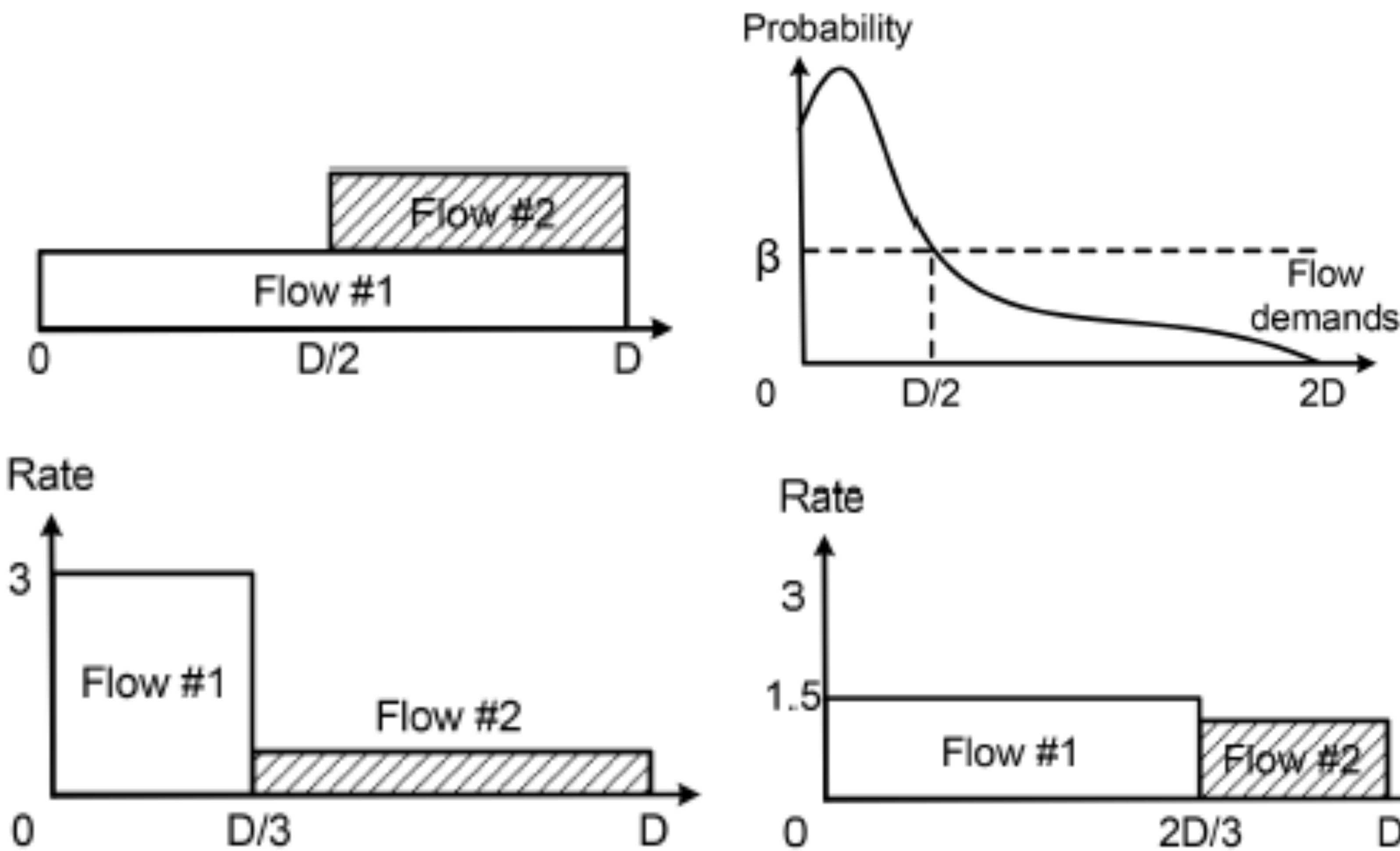
- Q1： how to reduce energy?
- A1： speed scaling.

- Q2： how to guarantee quality of service?
- A2： ensuring flow deadline.

Challenges

- Online scheduling
  - Make decisions continuously for the arriving new flows.
- Incorporating future information
  - The predicted future information should be incorporated in order to improve the quality of the scheduling.
- Efficient in practice
  - Target the average cases rather than devoting themselves to the worst-case inputs.

Motivation example

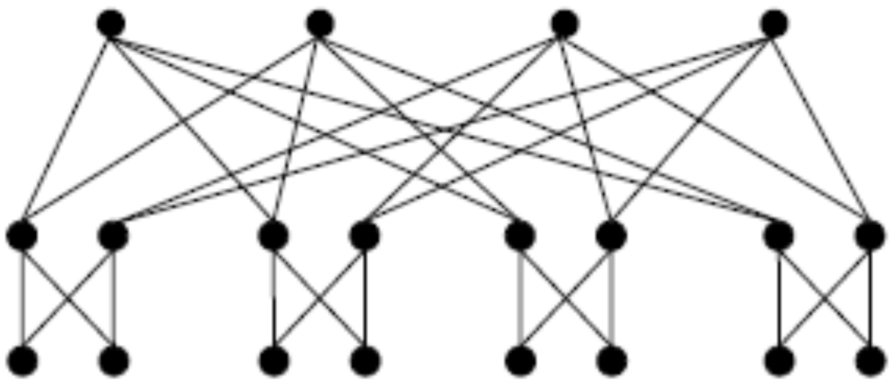


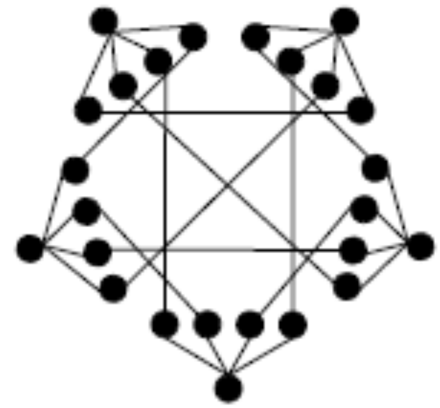
Outline

- Background
- Problem
- Solutions
- Evaluation

Data center flow scheduling

- DCN topology can be abstracted as a undirected graph.





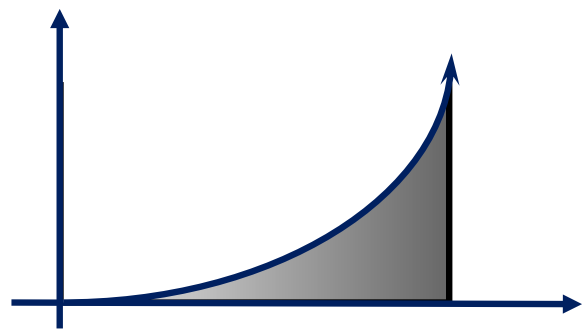
- A flow is associated with four parameters.
  - Flow demand, release time & deadline, the source & destination, the routing path

A schedule is called ”feasible” if every flow can be accomplished within its deadline following this schedule.

中国科学院计算技术研究所
INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES

Models

- Energy saving mechanism
  - Speed scaling
  - $g(x_e) = \sigma + \mu x_e^\alpha$



- Prediction and uncertainty
  - probability-based model

    $$w_{t,t_1} \sim \widehat{f}_{t_p,t,t_1}(w).$$

  - $$\widehat{\mathbf{w}}_\beta(t_p, t) = \{\widehat{w}^\beta_{t_p,t,t'} | t' \in [t, T]\},$$
    $$\widehat{w}^\beta_{t_p,t,t'} \in \{w | \widehat{f}_{t_p,t,t_1}(w) \geq \beta\}.$$

Abstracting the problem

- The Deadline-constrained Energy-efficient Flow Scheduling (DEFS) problem
  - Can we find a feasible online scheduling such that the total energy consumption is minimized?

$$\Phi_f(\mathcal{S}) = T \cdot |\mathcal{E}_a| \cdot \sigma + \int_0^T \sum_{e \in \mathcal{E}_a} \mu (x_e(t))^\alpha dt.$$

- Solving the problem is NP-hard!

Outline

- Background
- Problem
- Solutions
- Evaluation

Framework

Start

Compute routing path

- Using typical routing, e.g., shortest path routing, and earliest deadline first scheduling policy.

Compute optimal offline

- Using Yao's algorithm

Compute scaling factor

- Combining the prediction information and some basic computation framework.

End

The computation of scaling factor

Lemma 1. *([24]) An online algorithm, A, is feasible iff for all $Z \in \mathcal{Z}$ and all $t_1 \le t_2$, $t_1, t_2 \in \mathcal{T}$, the following inequality holds:*

$$\sum_{t=t_1}^{t_2} \sum_{s=t}^{t_2} w_{t,s} \le \int_{t_1}^{t_2} x_e(t)dt. \tag{8}$$

Combining $x_e(t) \le \sigma_{t_1,t_2} x_e^*(t)$ with Lemma 1,

$$\sum_{t=t_1}^{t_2} \sum_{s=t}^{t_2} w_{t,s} \le \int_{t_1}^{t_2} x_e(t)dt \le \sigma_{t_1,t_2} \int_{t_1}^{t_2} x_e^*(t).$$

The computation of scaling factor

According to prediction information and by applying Yao's algorithm, we have

$$x_e^*(Z_t^\beta) = \inf_{Z' \in \mathcal{Z}_\beta, Z'_t = Z_t^\beta} x'_e(Z_t^\beta).$$

Substituting $x_e^*(t)$ with $x_e^*(Z_t^\beta) \le x_e^*(t)$ for inequation

$$\sigma_{t_1,t_2} \ge \frac{\sum_{t=t_1}^{t_2} \sum_{s=t}^{t_2} w_{t,s}}{\int_{t_1}^{t_2} x_e^*(Z_t^\beta)}.$$

By solving the following optimization problem,

$$\sigma_{t_1,t_2} = \sup_{Z \in \mathcal{Z}_\beta} \frac{\sum_{t=t_1}^{t_2} \sum_{s=t}^{t_2} w_{t,s}}{\int_{t_1}^{t_2} x_e^*(Z_t^\beta)}$$

Outline

- Background
- Problem
- Solutions
- Evaluation

The evaluation setting

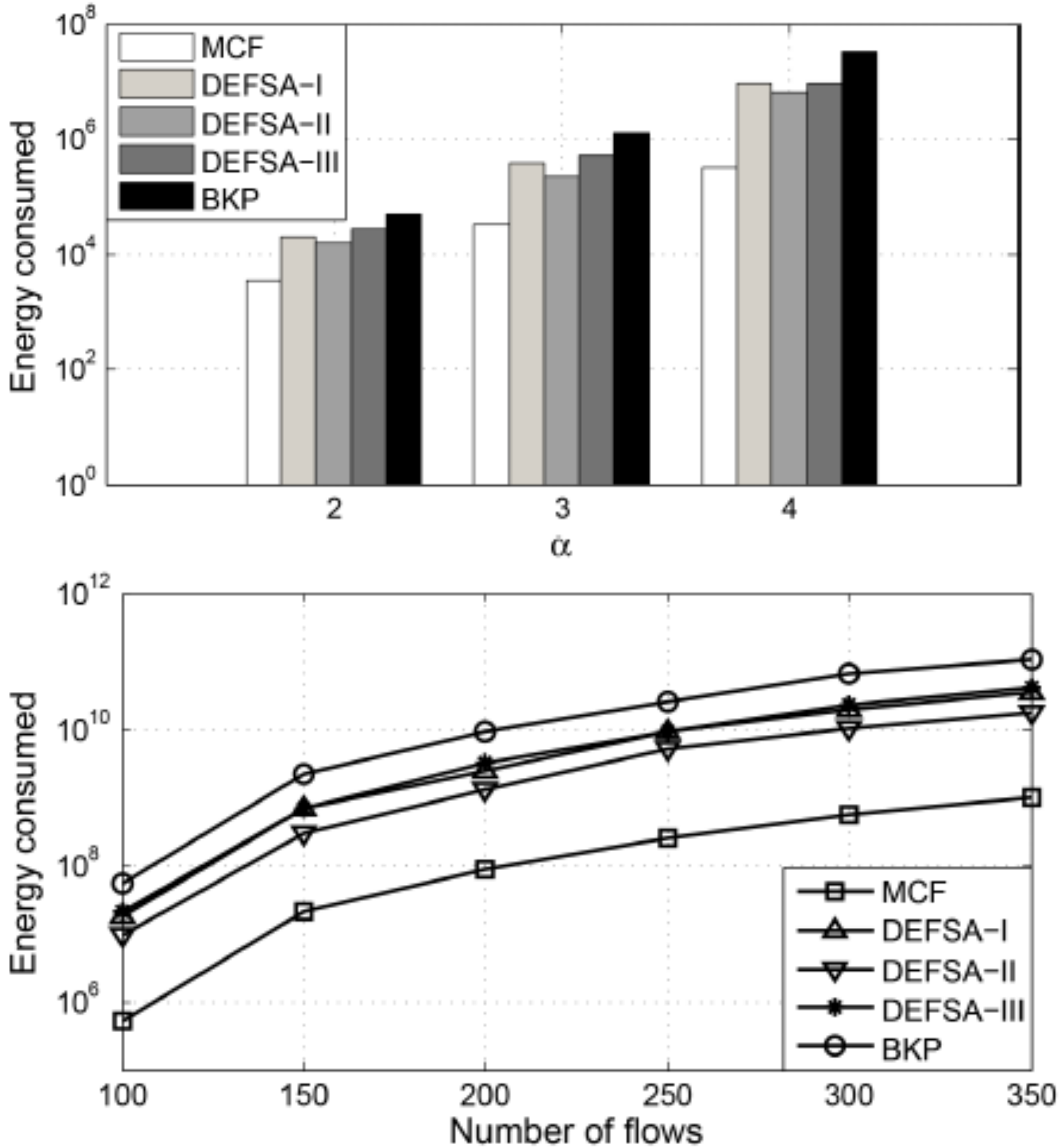
- Data center topology
  - A 8-ary Fat-Tree with 128 servers
- Benchmarks
  - Our algorithm (TRO)
  - Most-Critical-First (MCF)
  - Online scheduling without prediction (BKP)
- Evaluation metric
  - Total energy consumption.

Single VC request scenario

Case	MCF	DEFSA-I	DEFSA-II	DEFSA-III	BKP
100#	3	91	63	90	336
350#	10171	405494	205976	325494	1086340

Case	DEFSA-L0	DEFSA-L30	DEFSA-L60
100#	282	63	23
350#	995307	205976	91342

Multiple VC requests scenario



Thanks for your attention!