2-Dominant Resource Fairness: Fairness-Efficiency Tradeoffs in Multi-resource Allocation

Suhan Jiang and Jie Wu

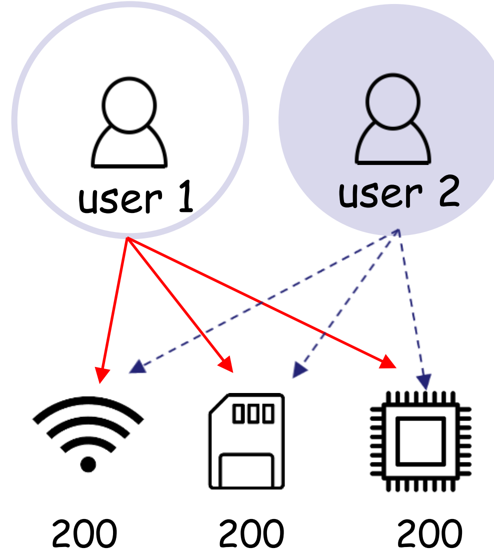
Dept. of Computer and Info. Sciences

Temple University, USA

Road Map

- Introduction
- Motivation
- 2-Dominant Resource Fairness (2-DF)
- Extension (k-DF) and Properties
- Experiment
- Conclusions

1. Introduction



- Multi-resource allocation
  - Sharing more than one type of resource

    Bandwidth, Memory, CPU, etc

  - Users have heterogeneous resource demands

    EX.1 Three resources: Bandwidth, Memory, CPU

    Total resource: < 200, 200, 200 > units

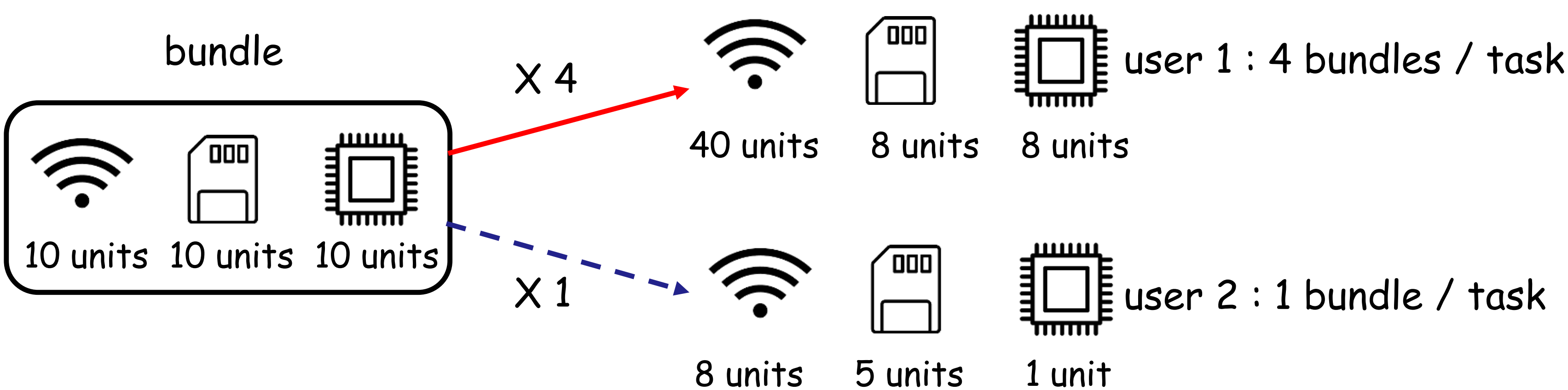
    User 1 requires < 40, 8, 8 > units / task

    User 2 requires < 8, 5, 1 > units / task

  - How to fairly/efficiently allocate all resources among users

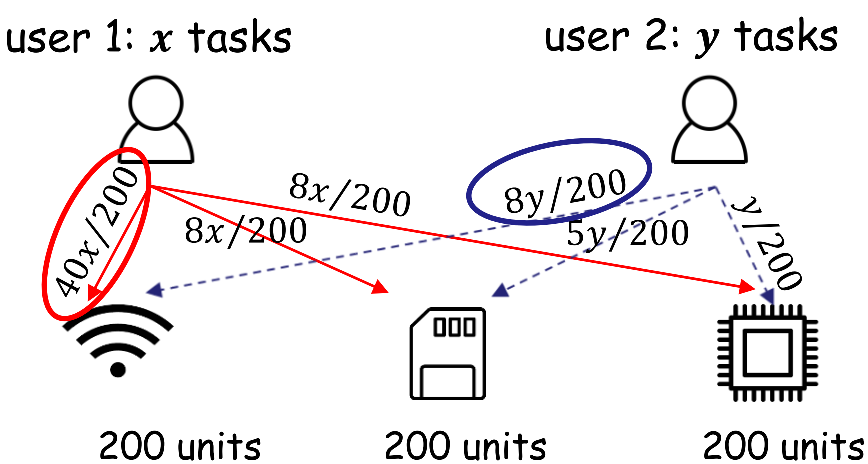
Resource Abstraction

- All resources are partitioned into bundles
  - each bundle has fixed amounts of different resources
  - multiple resources are abstracted as a single resource
- Drawbacks
  - ignore different demands of heterogeneous users
  - cannot always match nicely with users' demands



Dominant Resource Fairness [1] (DRF)

- Dominant resource
  - the resource that a user has the biggest share of
- Dominant share
  - the fraction of the dominant share a user is allocated
- DRF allocation mechanism
  - applying max-min fairness to dominant shares



$$\text{maximize} \quad x, y$$

$$\text{subject to} \quad \begin{cases} 40x/200 = 8y/200 \\ 40x + 8y \le 200 \\ 8x + 5y \le 200 \\ 8x + y \le 200 \end{cases}$$

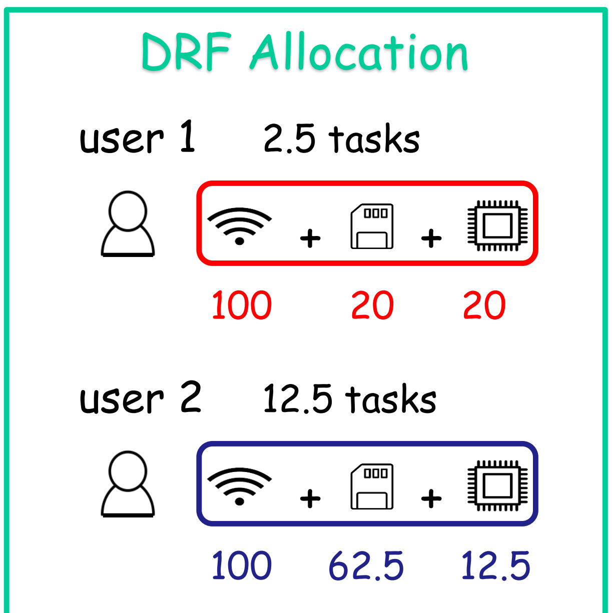
$$x = 2.5 \qquad y = 12.5$$

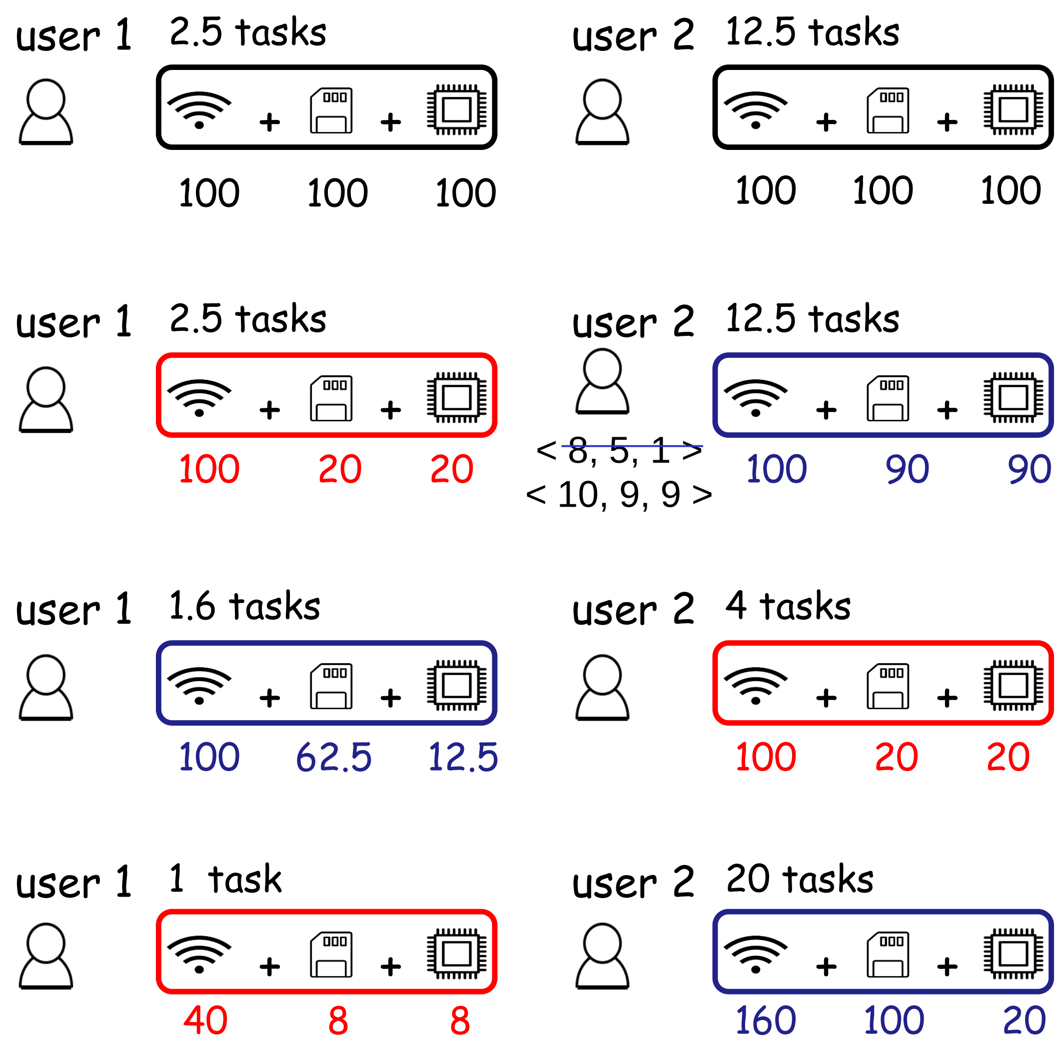
[1] A. Ghodsi, M. Zaharia, B. Hindman, A. Konwinski, S. Shenker, and I. Stoica. '' Dominant resource fairness: fair allocation of multiple resource types.'' In NSDI, 2011.

Dominant Resource Fairness (DRF)

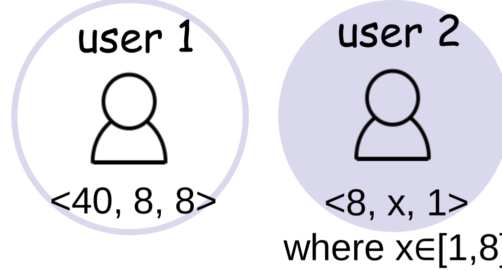
Properties

- sharing incentive
- strategy proof
- envy-free
- Pareto efficient

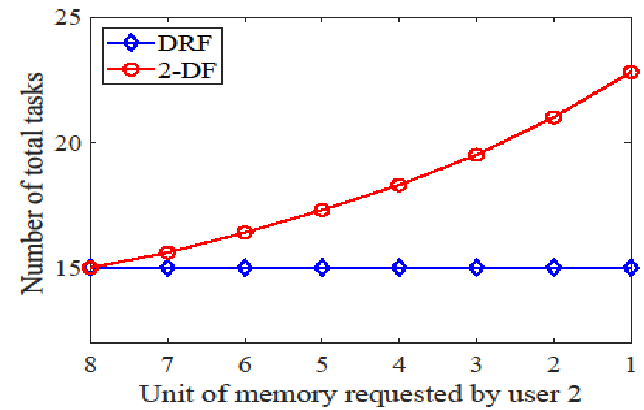




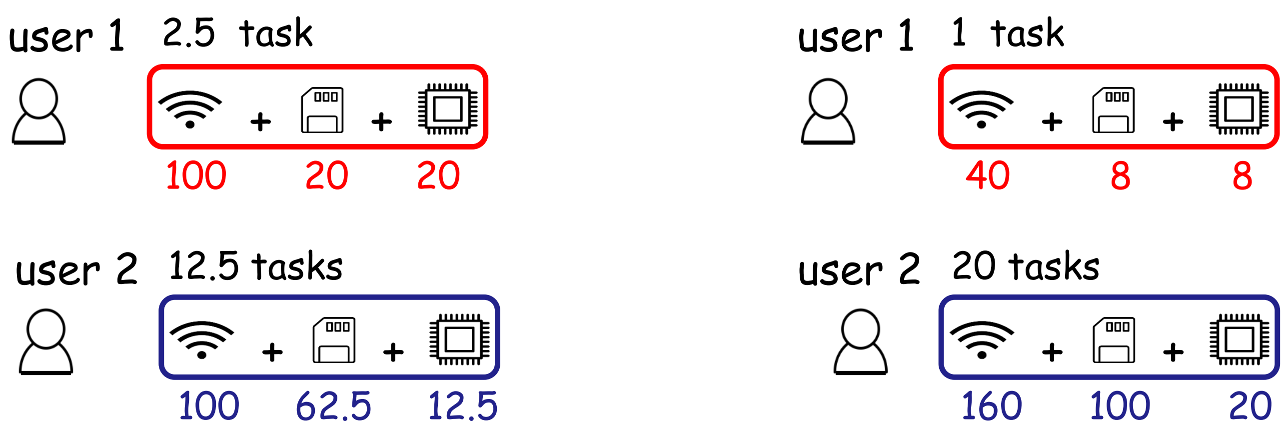
2. Motivation



- Fairness dispute in DRF
  - Focus on one resource



- Efficiency loss in DRF
  - DRF less efficiently uses resources [2]

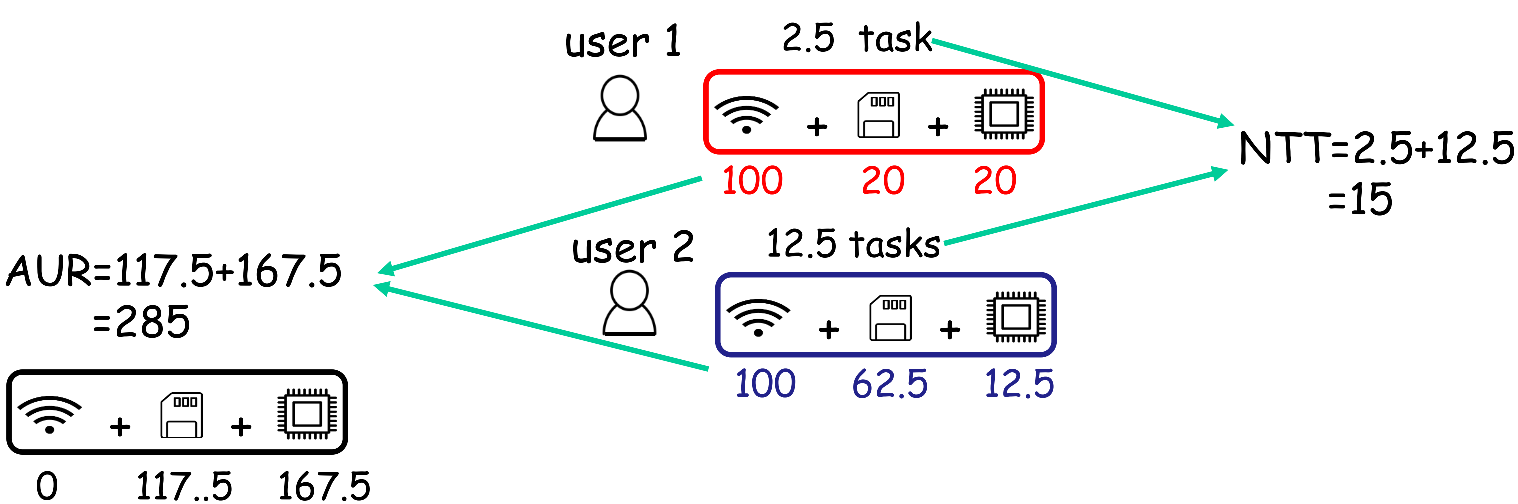


[2] Y. Jin and M. Hayashi. "Efficiency comparison between proportional fairness and dominant resource fairness with two different type resources." in CISS 2016.

Metrics on Fairness and Efficiency

- Fairness
  - Desirable Properties

    sharing incentive, strategy proof, envy-free, Pareto efficient
- Efficiency
  - Two measurements
    1. the number of total tasks completed (NTT)
    2. the amount of unused resources (AUR)



3. 2-Dominant Resource Fairness

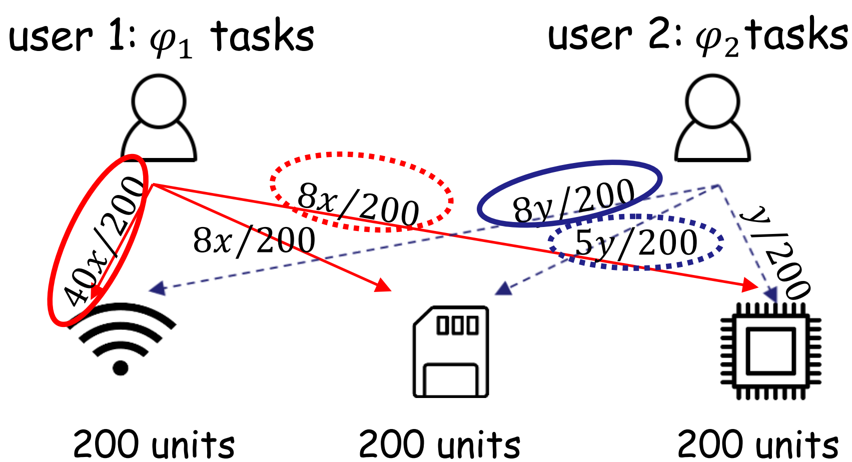
- Model
  - r resources and n users
  - Resource j's capacity: C_j
  - User i's request vector: D_i
  - User i's final allocation vector: A_i
- Objective
  - Design a new fairshare function for max-min fairness for efficiency improvement
  - A fairshare function on
    - multiple resources (instead of dominant resource alone)
    - weighting factors among different resources

2-DF Allocation Mechanism

- 2-DF fairshare function
  - 2-dominant share: $s_i = \varphi_i \cdot d_{i1} \cdot d_{i2}$ where
  - φ_i: number of user i's tasks
  - $d_{i1} = \max\left\{\frac{a_{ij}}{c_j}\right\}$: user i's first dominant ratio among all resources j
  - $d_{i2} = \max\left\{\frac{a_{ij}}{c_j}\right\}$-$\{d_{i1}\}$: user i's second dominant ratio among all resources j
- Allocation mechanism

$s_1 = \varphi_1 \cdot d_{11} \cdot d_{12} = \varphi_1/125$

$s_2 = \varphi_2 \cdot d_{21} \cdot d_{22} = \varphi_2/1000$



$$\text{maximize} \quad \varphi_1, \varphi_2$$

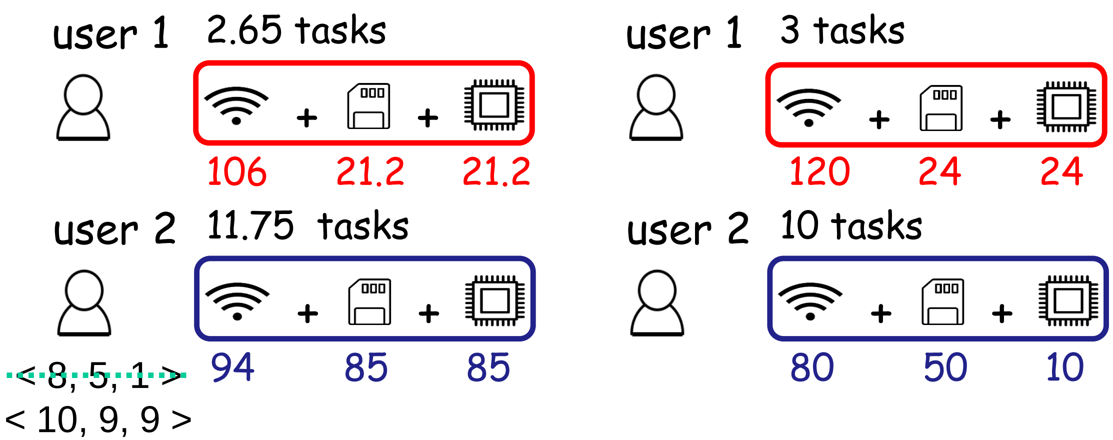
$$\text{subject to} \quad \begin{cases} \varphi_1/125 = \varphi_2/1000 \\ 40\varphi_1 + 8\varphi_2 \le 200 \\ 8\varphi_1 + 5\varphi_2 \le 200 \\ 8\varphi_1 + \varphi_2 \le 20 \end{cases}$$

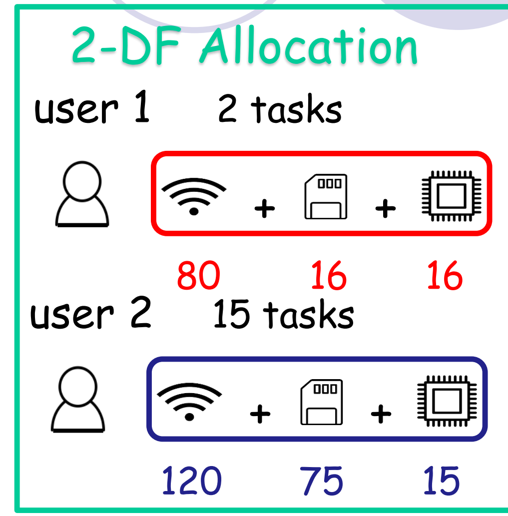
$\varphi_1 = 2 \quad \varphi_2 = 15$

4. Properties and Extension (k-DF)

Properties

- Strategy proof
- Pareto-efficient





k-DF mechanism

- k-dominant share: $s_i = \varphi_i \prod_{l=1}^{k} w_{il} d_{il}$ where w_{il} is a weight
- consider k dimensions of resources

5. Experiment -- First Scenario

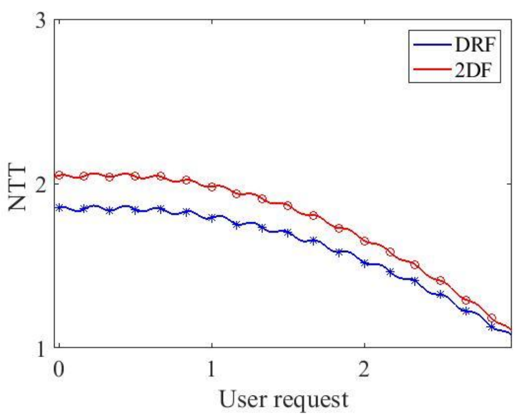
- Setting
  - A data center with 3 resources and 3 users
  - Resource capacity $< \mathrm{C}, \mathrm{C}, \mathrm{C} >$ where $\mathrm{C} \in \{3, 5\}$
  - User 1's request vector $<d_{11},\ d_{12},\ d_{13}>$ where $d_{1i} \in [1,\mathrm{C}]$
  - User 2's request vector $<d_{21},\ d_{22},\ d_{23}>$ where $d_{2j} \in [1,\mathrm{C}]$
  - User 3's request vector $<d_{31},\ d_{32},\ d_{33}>$ where $d_{3k} \in [1,\mathrm{C}]$
- Three comparison algorithms
  - No Fairness Constraints (NFC)
  - Dominant Resource Fairness (DRF)
  - 2-Dominant Resource Fairness (2DF)

First Scenario

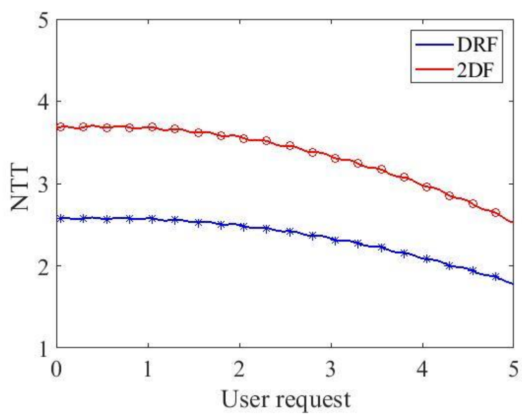
- Efficiency -- NNT
  - Average NTT under different capacities

Capacity	NFC	DRF	2DF
3	1.626	1.342	1.387
5	1.823	1.481	1.545

  - NNT with request vectors increasing



(a) Capacity of 3



(b) Capacity of 5

5. Experiment -- Second Scenario

- Setting
  - A data center with 3 resources and 2 users
  - Resource Capacity $< 1000, 1000, 1000 >$
  - Two user request types: heavy and light

    a request D_i= $<d_{i1},\ d_{i2},\ d_{i3}>$ is heavy if $\forall\ d_{ij} \in \{25x_1, 5x_1, x_1\}$

    a request D_i= $<d_{i1},\ d_{i2},\ d_{i3}>$ is light if $\forall\ d_{ij} \in \{25x_2, 5x_2, x_2\}$

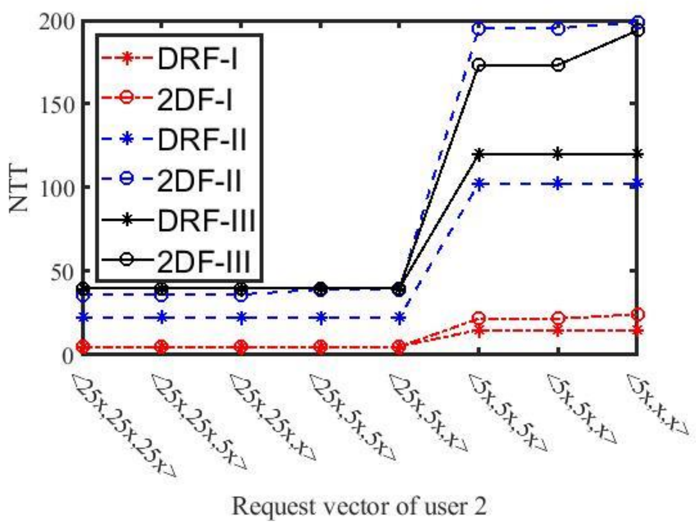
    where $x_1 \sim \mathrm{N}(8, 0)$ and $x_2 \sim \mathrm{N}(1, 0)$
  - Three combinations of user request types

Combination	user 1	user 2
I	heavy	heavy
II	heavy	light
III	light	light

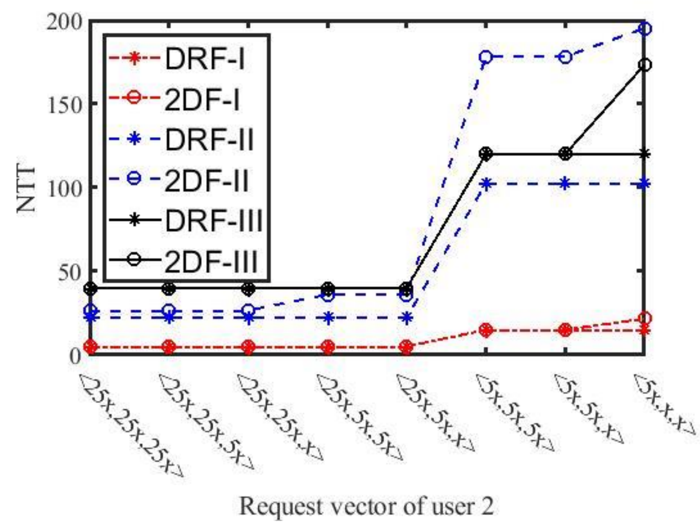
Second Scenario

Efficiency

- NNT : When user 1 has big tasks and user 2 has small tasks, the improvement of NNT is most obvious.
- Average increase: 45%



(a) I: user 1 $<25x_1, 25x_1, 25x_1>$
II: user 1 $<25x_1, 25x_1, 25x_1>$
III: user 1 $<25x_2, 25x_2, 25x_2>$



(b) I: user 1 $<25x_1, 5x_1, x_1>$
II: user 1 $<25x_1, 5x_1, x_1>$
III: user 1 $<25x_2, 5x_2, x_2>$

Second Scenario

- Efficiency
  - AUR : 2DF consumes less resources while yields more tasks

Average AUR	DRF	2DF
I	418	654
II	418	802
III	418	654

Average AUR	DRF	2DF
I	1218	1271
II	1218	998
III	1218	1271

(a) I: user 1 $\langle 25x_1, 25x_1, 25x_1 \rangle$
II: user 1 $\langle 25x_1, 25x_1, 25x_1 \rangle$
III: user 1 $\langle 25x_2, 25x_2, 25x_2 \rangle$

(b) I: user 1 $\langle 25x_1, 5x_1, x_1 \rangle$
II: user 1 $\langle 25x_1, 5x_1, x_1 \rangle$
III: user 1 $\langle 25x_2, 5x_2, x_2 \rangle$

6. Conclusion

- DRF suffers from serious fairness concerns without utility guarantees
- 2-DF seeks to balance fairness and efficiency
  - Extension from 2-DF to k-DF
  - strategy proof and Pareto efficient

Q & A