Good morning!

CS 744: DRF

Shivaram Venkataraman

Fall 2021

ADMINISTRIVIA

- Assignment 2 out! → Pytorch ~2 weeks
- Course Project
  - Project list by Oct 1
  - Form groups and submit project bids by Oct 8
  - Assigned project by Oct 14
  - Introductions due Oct 25

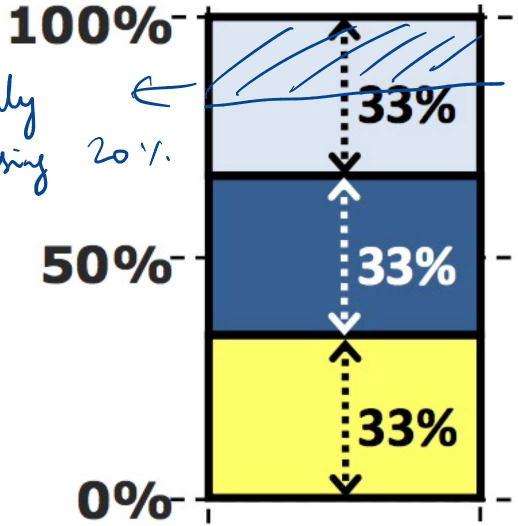
SETTING: FAIR SHARING

→ Background

not work conserving

Equal Share

only using 20%



Max-Min Share

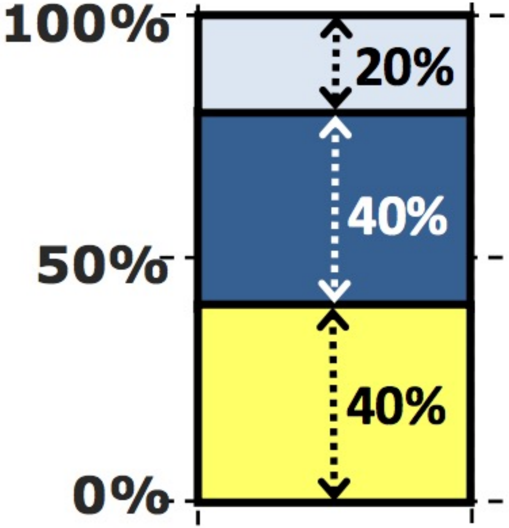
CPU scheduling : Lottery

Network sched

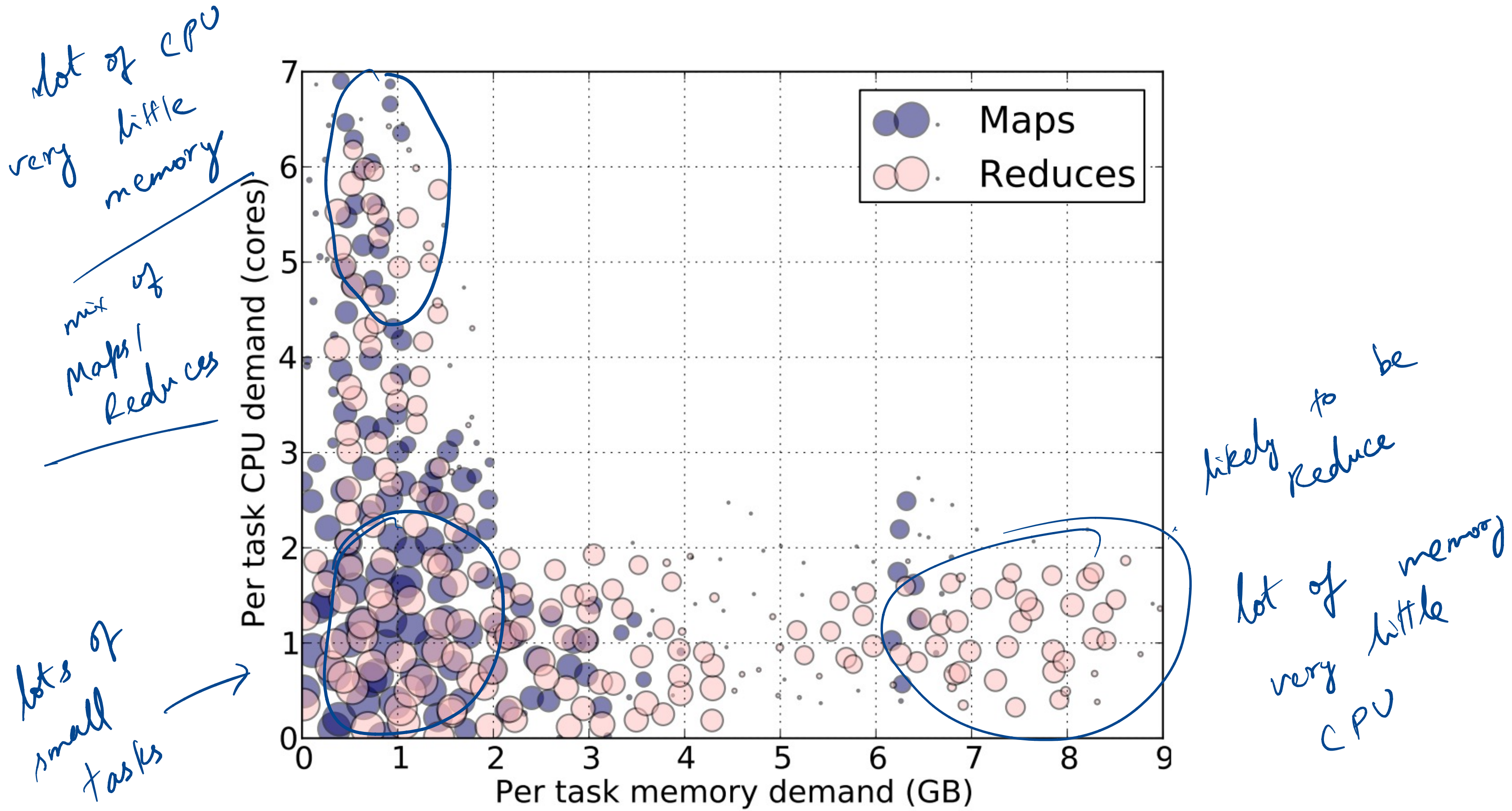
Maximize the allocation for most poorly treated users

max-min fairness

Maximize the minimum



MOTIVATION: MULTI RESOURCES



DRF: MODEL

Users have a demand vector

<2, 3, 1> means user's task needs 2 R1, 3 R2, 1 R3

2 CPUs 3GB 1 disk space

Multi resource

Resources given in multiples of demand vector

i.e., users might get <4,6,2>

→ can run multiple tasks of <2,3,1>

PROPERTIES

Sharing Incentive

Users should be not worse off when sharing the cluster

Strategy Proof

Users cannot get more resources by lying

Pareto Efficiency

Cannot give resources to one user without taking away from another

Envy free

User does not wish for another users allocation

PROPERTIES

Sharing Incentive

User is no worse off than a cluster with

1/n resources

Strategy Proof

User should not benefit by lying about demands

Pareto Efficiency

Not possible to increase one user without decreasing another

Envy free

User should not desire the allocation of another user

DRF: APPROACH

Dominant Resource

Resource user has the biggest share of

Total: <10 CPU, 4 GB> CPU Mem

User 1: <1 CPU, 1 GB> = 1/10 < 1/4

Dominant resource is memory

Dominant Share

Fraction of the dominant resource user is allocated

E.g., for User 1 this is 25% or 1/4

Allocate 2 tasks with <1 CPU, 1 GB> = 2/4

Dominant share

DRF: APPROACH

If User 1 : <2 CPU, 4GB>

Equalize the dominant share of users

Total: <9 CPU, 18 GB>

User1: <1 CPU, 4 GB>
dom res: mem

User2: <3 CPU, 1 GB>
dom res: CPU

Total Used: 9 CPU, 14GB

User	Allocation	Dominant Share
User1	<0 CPU, 0 GB> <1 CPU, 4 GB> <2 CPU, 8 GB> <3 CPU, 12 GB>	0 4/18 = 2/9 8/18 = 4/9 12/18 = 6/9
User2	<0 CPU, 0 GB> <3 CPU, 1 GB> <6 CPU, 2 GB>	0 3/9 6/9

DRF: APPROACH

Total: <9 CPU, 18 GB>

User1: <1 CPU, 4 GB> per task

<3 CPU, 12 GB> for 3 tasks

dom res: mem

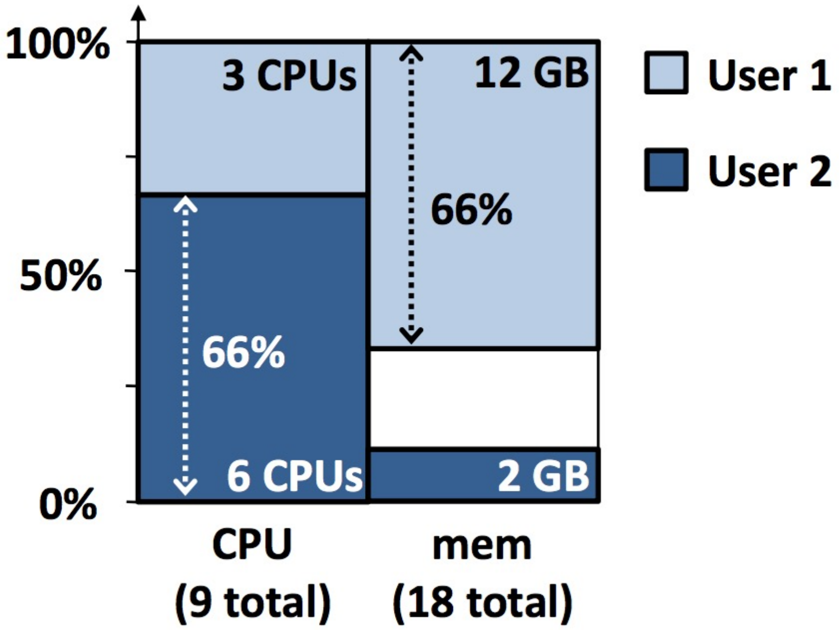
dom share: 12/18 = 2/3

User2: <3 CPU, 1 GB>

<6 CPU, 2 GB> for 2 tasks

dom res: CPU

dom share: 6/9 = 2/3



DRF ALGORITHM

Whenever there are available resources:

Schedule a task to the user with smallest dominant share

DRF ALGORITHM

Algorithm 1 DRF pseudo-code

$R = \langle r_1, \cdots, r_m \rangle$ ▷ total resource capacities → cluster capacity

$C = \langle c_1, \cdots, c_m \rangle$ ▷ consumed resources, initially 0 → Resources used so far

s_i $(i = 1..n)$ ▷ user i's dominant shares, initially 0 → Dominant share updated as we make allocation

$U_i = \langle u_{i,1}, \cdots, u_{i,m} \rangle$ $(i = 1..n)$ ▷ resources given to user i, initially 0

pick user i with lowest dominant share s_i

$D_i \leftarrow$ demand of user i's next task

if $C + D_i \leq R$ **then** → fits within capacity

  $C = C + D_i$ ▷ update consumed vector

  $U_i = U_i + D_i$ ▷ update i's allocation vector

  $s_i = \max_{j=1}^{m}\{u_{i,j}/r_j\}$ → update their dominant share (using … available)

else

  return ▷ the cluster is full

end if

→ a different user can fit in the available resources

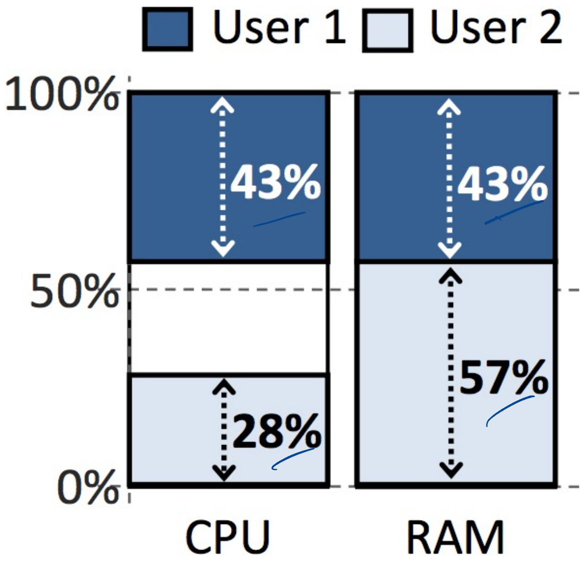
COMPARISON: ASSET FAIRNESS

Asset Fairness: Equalize each user's sum of resource shares

Consider total of 70 CPUs, 70 GB RAM
U1 needs <2 CPU, 2 GB RAM> per task = 4
U2 needs <1 CPU, 2 GB RAM> per task = 3

Asset Fair Allocation:
U1: 15 tasks: 30 CPU, 30 GB (Sum = 60)
U2: 20 tasks: 20 CPU, 40 GB (Sum = 60)

U2: 3, 6 60
U1: 4, 8, ... 60



COMPARISON: ASSET FAIRNESS

Asset Fairness: Equalize each user's sum of resource shares

Violates Sharing Incentive

Prev slide

Consider total of 70 CPUs, 70 GB RAM
U1 needs <2 CPU, 2 GB RAM> per task
U2 needs <1 CPU, 2 GB RAM> per task

U1: 15 tasks
U2: 20 tasks

Sharing incentive?
Half of the cluster is 35 CPU, 35 GB RAM
U1: 17 tasks → is better off with half of cluster
U2: 17 tasks

COMPARISON: CEEI

CEEI: Competitive Equilibrium from Equal Incomes

- Each user receives initially 1/n of every resource,
- Subsequently, each user can trade resources with other users in a perfectly competitive market
- Computed by maximizing product of utilities across users

COMPARISON: CEEI

Utility: Number of tasks user gets

Total: <9 CPU, 18 GB> User1: <1 CPU, 4 GB> User2: <3 CPU, 1 GB>

user 1

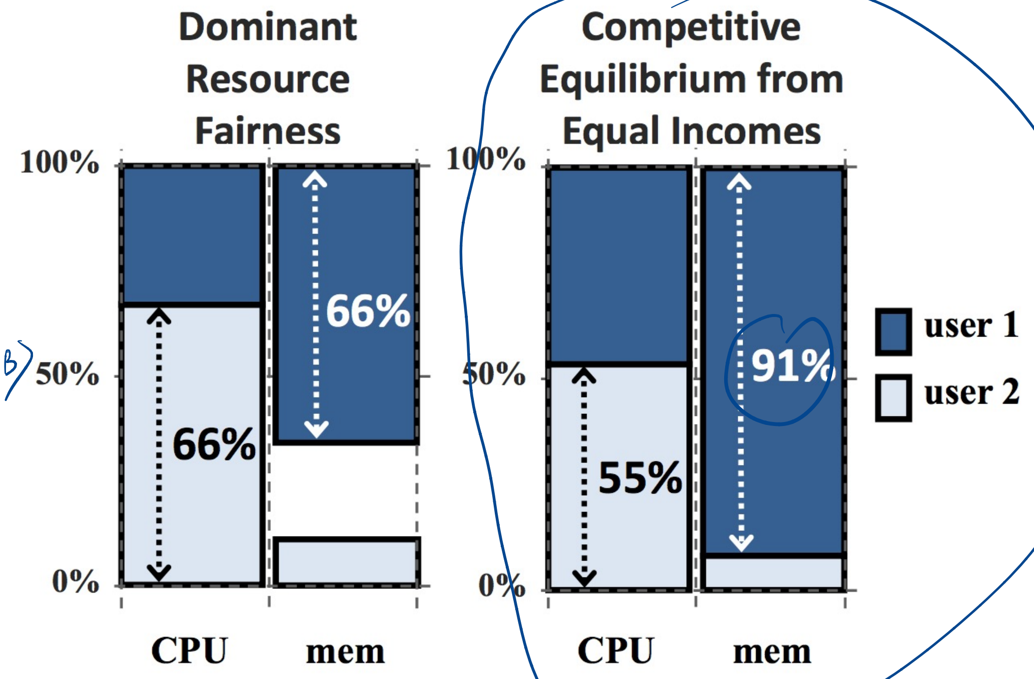
$$\max\,(x \cdot y)$$

subject to

$$\begin{aligned} x + 3y &\leq 9 \quad \text{CPU} \\ 4x + y &\leq 18 \quad \text{Mem} \end{aligned}$$

User1 $x = 4.05$ <4.05, ~16 GB>

$y = 1.62$



CEEI: STRATEGY PROOFNESS

Total: <9 CPU, 18 GB>

User2 Before:
CEEI: 55% CPU, 9% mem

Total: <9 CPU, 18 GB>

User1: <1 CPU, 4 GB>
User2: <3 CPU, 2 GB>

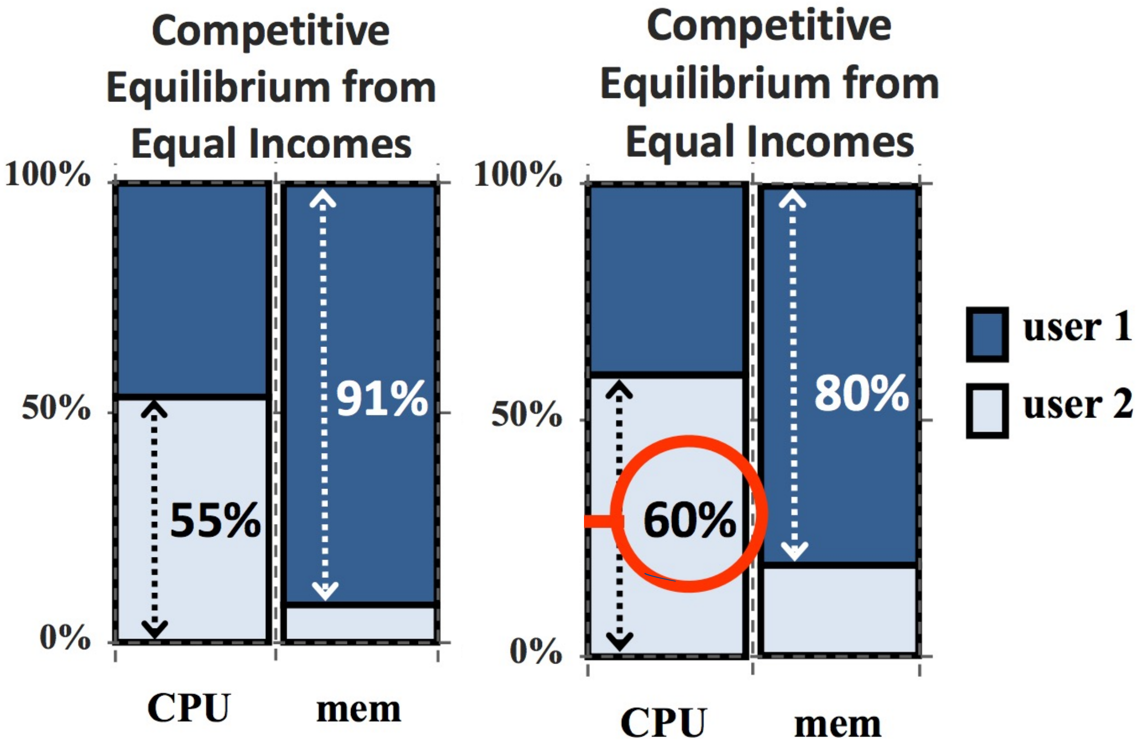
$\max \quad x \cdot y$

$x + 3y <= 9$

$4x + 2y <= 18$

$x = 3.6$

$y = 1.8$



COMPARISON

Property	Allocation Policy Asset	CEEI	DRF
Sharing Incentive		✓	✓
Strategy-proofness	✓		✓
Envy-freeness	✓	✓	✓
Pareto efficiency	✓	✓	✓
Single Resource Fairness	✓	✓	✓
Bottleneck Fairness		✓	✓
Population Monotonicity	✓		✓
Resource Monotonicity			

Table 2: Properties of Asset Fairness, CEEI and DRF.

SUMMARY

DRF: Dominant Resource Fairness

Allocation policy for scheduling

Provides multi-resource fairness

Ensures sharing incentive, strategy proofness

DISCUSSION

https://forms.gle/wdN8bwqxEwjPEcAq7

Consider a system with 40 units of CPU, 20 units of memory and 160 units of disk. Consider three users with the following requirements

Alice (4 CPU, 1 memory, 1 disk)

Bob (1 CPU, 4 memory and 4 disk)

Carol (1 CPU, 2 memory and 16 disk)

List the dominant resource as defined in DRF for Alice, Bob and Carol

Alice $\langle \frac{1}{10}, \frac{1}{20}, \frac{1}{160} \rangle$ CPU

Bob $\langle 1/40, 1/5, 1/40 \rangle$ Memory

Carol $\langle 1/40, 1/10, 1/10 \rangle$ Memory / Disk

What would be the final task allocation in the given cluster for Alice, Bob and Carol ?

Alice	Dom. Share
4, 1, 1	1/10
8, 2, 2	2/10 = 1/5
12, 3, 3	3/10
16, 4, 4	4/10

4 tasks

3 tasks in static 1/3rd cluster

Bob	Dom. Share
1, 4, 4	1/5
2, 8, 8	2/5 = 4/10

1 task in static 1/3rd cluster

Total: 40, 20, 160

1/3rd : 13, 6.3, 53

Carol	Dom. Share
1, 2, 16	1/10
2, 4, 32	1/5
3, 6, 48	3/10
4, 8, 64	4/10

3 tasks in static 1/3rd cluster

What could be one workload / cluster scenario where DRF implemented on Mesos will NOT be optimal?

Placement preferences or locality

NEXT STEPS

Next Week: Machine Learning

Assignment 2 out!