Good morning!

CS 744: DRF

Shivaram Venkataraman
Fall 2020
- Assignment 2 out!  → Piazza
- Course Project
  - Form groups?  ~ 3 students
  - Project list by Monday (9/28)  → Google Form
  - Submit project bids by Thursday (10/1)
  - Assigned project by Friday (10/2)
  ~ 2 months to work on the project
SETTING: FAIR SHARING

Equal Share

Max-Min Share

Maximize the allocation for most poorly treated users

Maximize the minimum

3 users = \( \frac{1}{3} \) of resources

Networking OS → lottery scheduling

different demands across users

earlier

MOTIVATION: MULTI RESOURCES

- Memory
- CPU
- Lots of tasks are CPU-heavy
- Empty?
- Most of tasks here are reduce tasks
- Memory usage is not correlated

Diagram:
- Per task CPU demand (cores)
- Per task memory demand (GB)
- Maps
- Reduces

Tuple (CPU, Mem)
DRF: MODEL

Users have a demand vector

\(<2, 3, 1>\) means user’s task needs 2 R1, 3 R2, 1 R3

\(\rightarrow\) one task

Resources given in multiples of demand vector
i.e., users might get \(<4, 6, 2>\)

\(\leftarrow\) 2 tasks with their demand

slot based model

6 cores
10 GB

4 Map slots = 1 core
1.5 GB of mem
2 reduce slots

No containers
cgroups
Linux
PROPERTIES

Sharing Incentive

User should get at least 1/n of resources.
No worse off than their own cluster with \(\frac{1}{n}\) resources.

Strategy Proof

You can't lie about what you need to get more.
Incentivizing truth telling.

Pareto Efficiency

If you allocate more for one user, you need to take away from others.
Utilization

Envy free

Users should not envy allocation of another user.
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>
User 1: <1 CPU, 1 GB>
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
**DRF: APPROACH**

Equalize the dominant share of users

<table>
<thead>
<tr>
<th>User</th>
<th>Allocation</th>
<th>Dominant Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>User 1</td>
<td>&lt;0 CPU, 0 GB, &lt;1 CPU, 4 GB, &lt;2 CPU, 8 GB, &lt;2 CPU, 12 GB&gt;</td>
<td>0 2/9 4/9 6/9</td>
</tr>
<tr>
<td>User 2</td>
<td>&lt;0 CPU, 0 GB, &lt;3 CPU, 1 GB, &lt;6 CPU, 2 GB&gt;</td>
<td>0 3/9 6/9</td>
</tr>
</tbody>
</table>

**Total used:** 9 CPU, 14 GB

**Total:** <9 CPU, 18 GB>

User1: <1 CPU, 4 GB>

dom res: mem

User2: <3 CPU, 1 GB>

dom res: CPU
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 GPU, 2 GB> for 2 tasks
  dom res: CPU
  dom share: 6/9 = 2/3
Whenever there are available resources:
Schedule a task to the user with smallest dominant share
Algorithm 1 DRF pseudo-code

\[
R = \langle r_1, \cdots, r_m \rangle \quad \triangleright \text{total resource capacities}
\]
\[
C = \langle c_1, \cdots, c_m \rangle \quad \triangleright \text{consumed resources, initially 0}
\]
\[
s_i \ (i = 1..n) \quad \triangleright \text{user } i \text{'s dominant shares, initially 0}
\]
\[
U_i = \langle u_{i,1}, \cdots, u_{i,m} \rangle \ (i = 1..n) \quad \triangleright \text{resources given to user } i, \text{ initially 0}
\]

**pick** user \( i \) with lowest dominant share \( s_i \)

\( D_i \leftarrow \text{demand of user } i \text{'s next task} \)

_if_ \( C + D_i \leq R \) _then_

\[
C = C + D_i \quad \triangleright \text{update consumed vector}
\]
\[
U_i = U_i + D_i \quad \triangleright \text{update } i \text{'s allocation vector}
\]
\[
s_i = \max_{j=1}^{m} \{u_{i,j}/r_j\}
\]

_else_

\[\text{return}\]

\[\triangleright \text{the cluster is full}\]

_end if_
COMPARISON: ASSET FAIRNESS

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

Violates Sharing Incentive

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

Asset Fair Allocation:
U1: 4, 8, 12 ... 60 = 15 tasks for U1
U2: 3, 6, 9, 12 ... 60 = 20 tasks for U2
COMPARISON: ASSET FAIRNESS

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

Violates Sharing Incentive

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

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

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

\[
\begin{align*}
\text{CEEI} & \quad \max (x \cdot y) \\
\text{subject to} & \quad x + 3y \leq 9 \\
& \quad 4x + y \leq 18 \\
\end{align*}
\]

\[\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 4.05 \\ 1.62 \end{bmatrix} \]

Dominant Resource Fairness

Competitive Equilibrium from Equal Incomes
CEEI: STRATEGY PROOFNESS

Total: <9 CPU, 18 GB>

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

3.6 tasks ??
L discrete number of tasks ??

Total: <9 CPU, 18 GB>

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

\[ n + 3y \leq 9 \]
\[ 4x + 2y \leq 18 \]
\[ \max \ x \cdot y \]
\[ x = 3.6 \]
\[ y = 1.8 \]

\[ y = 1.62 \]
\[ x = 4.05 \]
<table>
<thead>
<tr>
<th>Property</th>
<th>Allocation Policy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Asset</td>
</tr>
<tr>
<td>Sharing Incentive</td>
<td></td>
</tr>
<tr>
<td>Strategy-proofness</td>
<td>✓</td>
</tr>
<tr>
<td>Envy-freeness</td>
<td>✓</td>
</tr>
<tr>
<td>Pareto efficiency</td>
<td>✓</td>
</tr>
<tr>
<td>Single Resource Fairness</td>
<td>✓</td>
</tr>
<tr>
<td>Bottleneck Fairness</td>
<td></td>
</tr>
<tr>
<td>Population Monotonicity</td>
<td></td>
</tr>
<tr>
<td>Resource Monotonicity</td>
<td>✓</td>
</tr>
</tbody>
</table>

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

DRF: Dominant Resource Fairness
Allocation policy for scheduling
Provides multi-resource fairness → generalizes max-min fairness
Ensures sharing incentive, strategy proofness
DISCUSSION

https://forms.gle/i7m7xXxKhtfvL9UD9
Consider a system with 100 units of CPU, 50 units of memory and 200 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: CPU 2/50
Bob: Memory 4/50
Carol: Disk 4/50
What would be the final task allocation in the given cluster for Alice, Bob and Carol?

\[
\begin{align*}
4x &= \frac{4y}{50} = \frac{16z}{200} \\
4x + y + z & \leq 100 \\
x + 4y + 2z & \leq 50 \\
x + 4y + 16z & \leq 200 \\
\end{align*}
\]

Every time Alice is allocated \( \frac{8}{200} \)
Bob and Carol \( \frac{16}{200} \)

Alice get two turns
every 2 turns 1 turn each for Bob, Carol

Alice: 12
Bob: 6, Carol: 6

Either (14, 6, 6) or (12, 6, 7)
What could be one workload / cluster scenario where DRF implemented on Mesos will NOT be optimal?

- If there aren't enough resources
  - if at least one task can run

- Heterogeneous tasks
  - [over time?]

- Locality pref?
Next Week: Machine Learning
Assignment 2 out!

**NEXT STEPS**

Mesos:
- resource Offer: (9 CPU, 18 GB)
  → task (1 CPU, 3 GB) at 6x1

**DRF Starvation** → one long running task

& very highly contended cluster

**Strategy proof:** assuming rational actors

- DRF starvation
- One long running task
- Very highly contended cluster

**Optimization problem**

\[
\begin{align*}
\text{max} & \quad \text{min} \quad \text{dominance} \\
\text{subject to} & \quad \text{resource constraints}
\end{align*}
\]

\[
\begin{align*}
\frac{3x}{21} = \frac{4y}{21}
\end{align*}
\]

\*

\[
\begin{align*}
D_1: & \quad \frac{3}{21} \quad \frac{6}{21} \\
D_2: & \quad \frac{4}{21} \quad \frac{8}{21} \\
D_1: & \quad \frac{6}{21} \quad \frac{12}{21} \\
D_2: & \quad \frac{8}{21} \quad \frac{16}{21} \\
\end{align*}
\]