CS 744: DRF

Shivaram Venkataraman
Fall 2019
ADMINISTRIVIA

- Assignment 1 details
- Assignment 2 out tonight
- Project groups
SETTING: FAIR SHARING

Equal Share

Max-Min Share

Maximize the allocation for most poorly treated users

Maximize the minimum
MOTIVATION: MULTI RESOURCES

- CPU intensive
- Maps
- Reduces
- Classic Hadoop sched.
- 4 map - shots
- 2 reduce - shots
- < 2 CPUs, 7 GB
- 4 CPUs, 4 CPUs
- 16 GB, 16 GB
- ~ 2 GB, 2 CPUs
- Memory intensive
Users have a demand vector
\(<2, 3, 1>\) means user’s task needs 2 R1, 3 R2, 1 R3

Resources given in multiples of demand vector
i.e., users might get \(<4, 6, 2>\) \(\geq 2\) tasks with \(<2, 3, 1>\)
PROPERTIES

Sharing Incentive
Each user should at least get at least $n$ users

Strategy Proof
No user should be able to get more resources by lying

Pareto Efficiency
No user can get more without reducing another user's allocation

Envy free
No user would prefer the allocation of another user
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 \( \text{wrt total resources available} \)

Total: \(<10 \text{ CPU, 4 GB}>\)
User 1: \(<1 \text{ CPU, 1 GB}>\)
Dominant resource is **memory**

\[
\frac{1}{10} \text{ CPU} \quad \frac{1}{4} \text{ memory}
\]

**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>User1</td>
<td>&lt;0 CPU, 0 GB&gt;</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>&lt;1 CPU, 4 GB&gt;</td>
<td>2/9</td>
</tr>
<tr>
<td></td>
<td>&lt;2 CPU, 8 GB&gt;</td>
<td>4/9</td>
</tr>
<tr>
<td></td>
<td>&lt;3 CPU, 12 GB&gt;</td>
<td>2/3</td>
</tr>
<tr>
<td>User2</td>
<td>&lt;0 CPU, 0 GB&gt;</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>&lt;3 CPU, 1 GB&gt;</td>
<td>1/3</td>
</tr>
<tr>
<td></td>
<td>&lt;6 CPU, 24 GB&gt;</td>
<td>2/3</td>
</tr>
</tbody>
</table>

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

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

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

**Total:** <9 CPU, 14 GB>
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

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

**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**

**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 ... 9 ... 12 ... 12
U2: 3 ... 6 ... 12 ...

Total: 70 24 120
**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>

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

\[
\begin{align*}
\gamma & = 1.62 \\
x & = 4.05
\end{align*}
\]
CEEI: STRATEGY PROOFNESS

Total: <9 CPU, 18 GB>

User 2 Before:
- DRF: 66% CPU, 16% mem
- CEEI: 55% CPU, 9% mem

Total: <9 CPU, 18 GB>

User 1: <1 CPU, 4 GB>
User 2: <3 CPU, 2 GB>

Dominant Resource Fairness

Competitive Equilibrium from Equal Incomes

\[
\begin{align*}
\text{max } x \cdot y \\
x + 3y \leq 9 \\
4x + 2y \leq 18
\end{align*}
\]

\[
y = 1.8 \\
x = 3.6
\]
<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>X</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.
DISCUSSION

https://forms.gle/s9nm7Gr1uz8Xsn3s5
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
Bob: Memory
Carol: Disk
What would be the final task allocation in the given cluster for Alice, Bob and Carol?

A: \( \langle 4, 1, 1 \rangle \)
B: \( \langle 1, 4, 4 \rangle \)
C: \( \langle 1, 2, 16 \rangle \)
T: \( \langle 100, 50, 2007 \rangle \)

Total

\( \text{Total} = 10 + 8 + 22 \)

\( 60, 48, 132 \)
What could be one workload / cluster scenario where DRF implemented on Mesos will NOT be optimal?

1. Heterogeneous task lengths
2. Fragmentation vs. Packing
3. Placement preferences
NEXT STEPS

Next Week: Machine Learning
Assignment 2 out tonight!
Course projects: Office hours