MULTI-RESOURCE ALLOCATION

FAIRNESS-EFFICIENCY TRADEOFFS IN A UNIFYING FRAMEWORK

Carlee Joe-Wong
Princeton University

Soumya Sen
Princeton University

Tian Lan
George Washington University

Mung Chiang
Princeton University

March 28, 2012 **IEEE Infocom**

What is Fairness?

- Politics, economics, sociology, engineering…

How do you allocate a resource to different users?

- Variance, Jain's index, entropy (see TR for references)…
- Isoelastic or α-fairness
- Unifying axiomatic theory of decomposable fairness measures

$$\mathrm{sgn}(1-\beta)\left(\sum_{i=1}^{n}\left(\frac{x_i}{\sum_{j=1}^{n} x_j}\right)^{1-\beta}\right)^{\frac{1}{\beta}}\left(\sum_{i=1}^{n} x_i\right)^{\lambda}$$

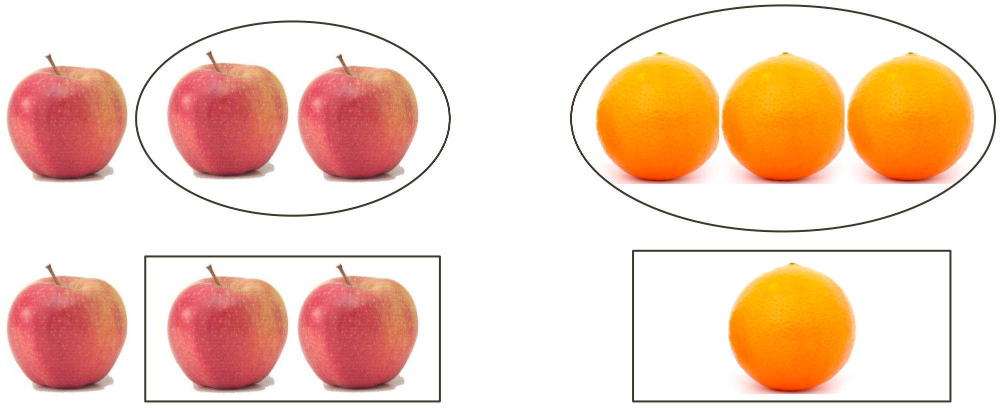
T. Lan, et al. An Axiomatic Theory of Fairness in Network Resource Allocation. IEEE Infocom 2010.

Our Question

- Suppose you have multiple non-substitutable resources.
  - Memory
  - CPU
  - Bandwidth
- They combine to make something…
  - Jobs in a datacenter
- that multiple people want.
  - Different bundles of resource requirements
- But the resources are finite.

How do you allocate the resources to different people?

Two-Resource Example



Generalized Fairness on Jobs (GFJ)

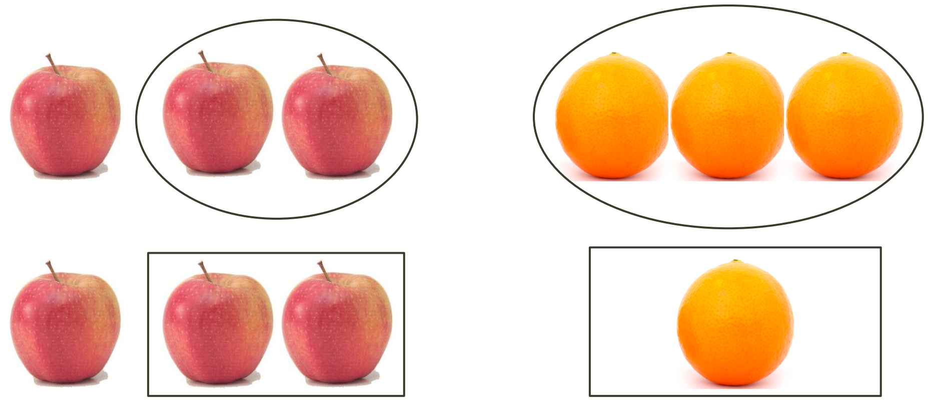
Fairness   Efficiency

$$f_{\beta,\lambda}(\mathbf{x}) = \mathrm{sgn}(1-\beta)\left(\sum_{i=1}^{n}\left(\frac{x_i}{\sum_{j=1}^{n} x_j}\right)^{1-\beta}\right)^{\frac{1}{\beta}} \left(\sum_{i=1}^{n} x_i\right)^{\lambda}$$

- Unique family of functions: β and *λ* parameters
  - β: type of fairness
  - *λ*: importance of efficiency

Defining "Fairness"

- An equal allocation?
  - 1 job for each user
- But not efficient

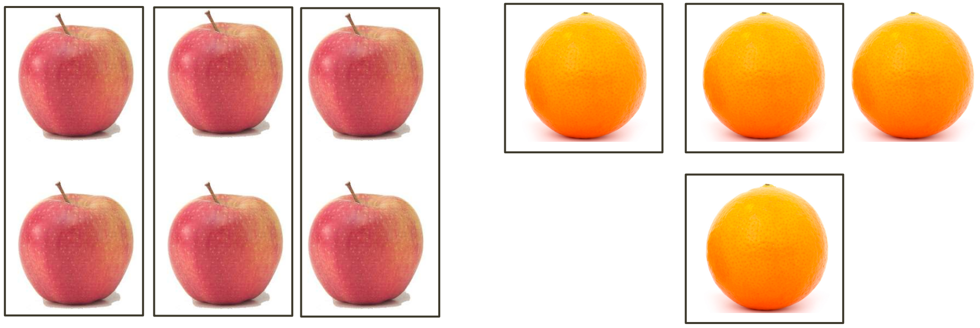


- Ranking the fairness of different allocations

$$\operatorname{sgn}(1-\beta)\left(\sum_{i=1}^{n}\left(\frac{x_i}{\sum_{j=1}^{n} x_j}\right)^{1-\beta}\right)^{\frac{1}{\beta}}$$

Defining "Efficiency"

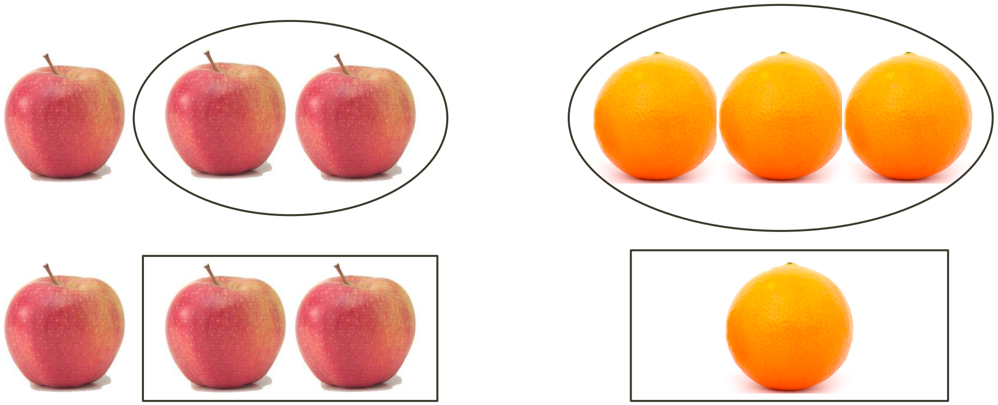
- Maximize the total number of jobs?
  - 0 jobs to user 1
  - 3 jobs to user 2
- But not that fair



- Ranking the efficiency of different allocations

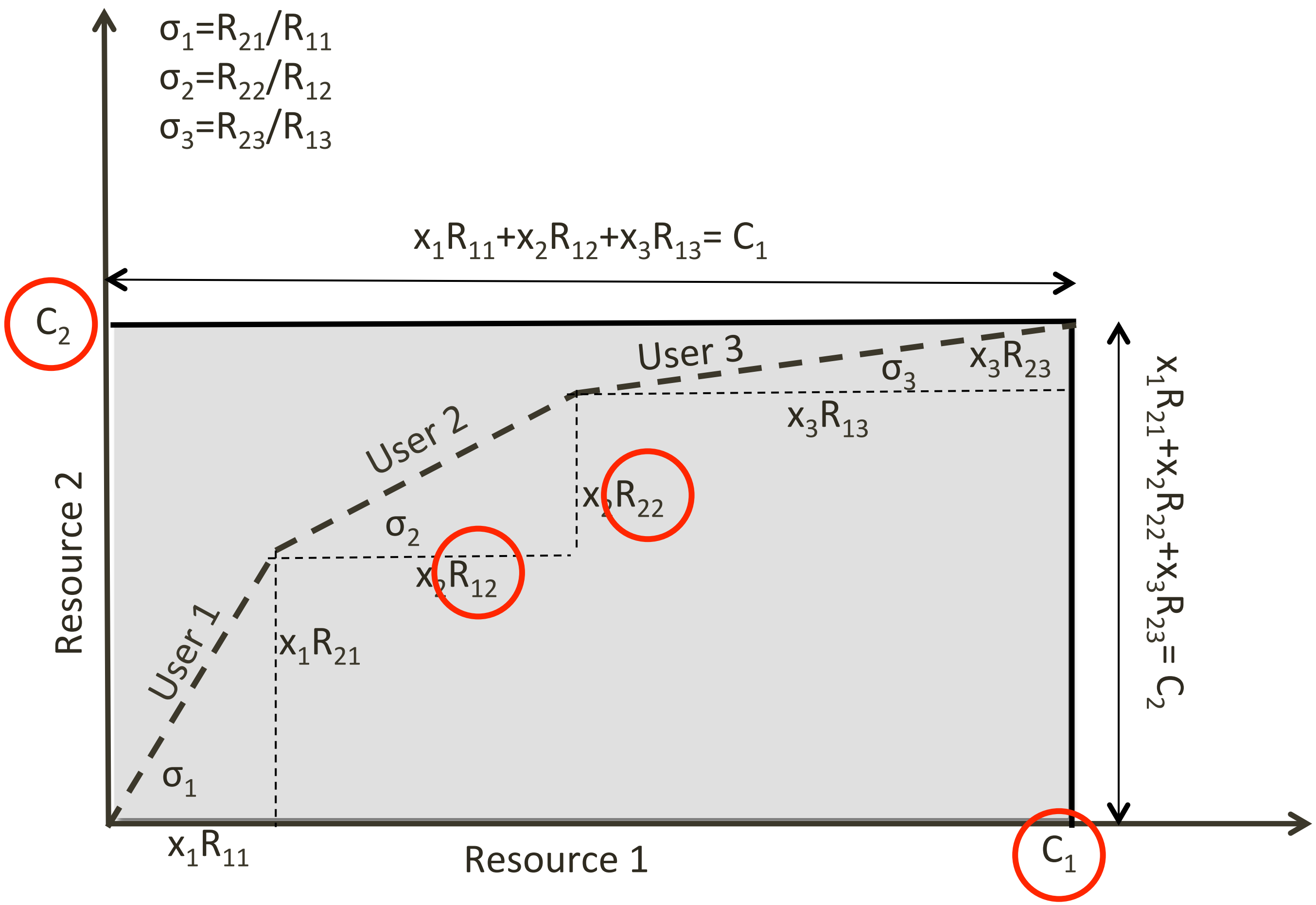
$$\left(\sum_{i=1}^{n} x_i \right)^{\lambda}$$

Heterogeneous Users



- Different users need different mixes of resources…
- Is it fair to treat them the same way?

Visualizing Heterogeneity



3 Users, 2 Resources

Dominant Shares

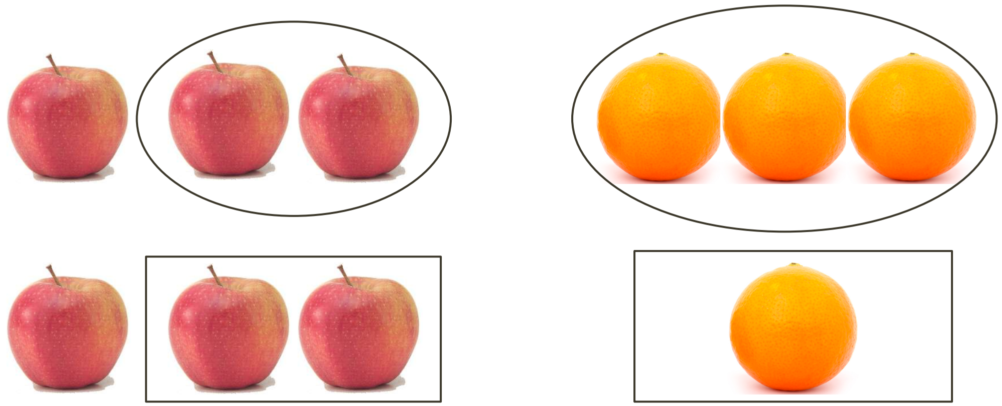
- Dominant shares $\mu_j x_j$ for each user

$$\mu_j = \max_i \left\{ \frac{R_{ij}}{C_i} \right\}$$

- Maximum share of any resource

A. Ghodsi, et al. Dominant Resource Fairness: Fair Allocation of Multiple Resource Types. USENIX NSDI 2011.

Calculating Dominant Shares



$$\mu_j = \max_i \left\{ \frac{R_{ij}}{C_i} \right\}$$

- 2 of 6 apples and 3 of 4 oranges: μ_1 = max (⅓, ¾)
- 2 of 6 apples and 1 of 4 oranges: μ_2 = max (⅓, ¼)

Fairness on Dominant Shares (FDS)

- Use dominant shares instead of number of jobs
- If μ is larger, equal dominant shares for smaller number of jobs

$$\operatorname{sgn}(1-\beta)\left(\sum_{j=1}^{n}\left(\frac{\mu_j x_j}{\sum_{k=1}^{n}\mu_k x_k}\right)^{1-\beta}\right)^{\frac{1}{\beta}}\left(\sum_{j=1}^{n}\mu_j x_j\right)^{\lambda}$$

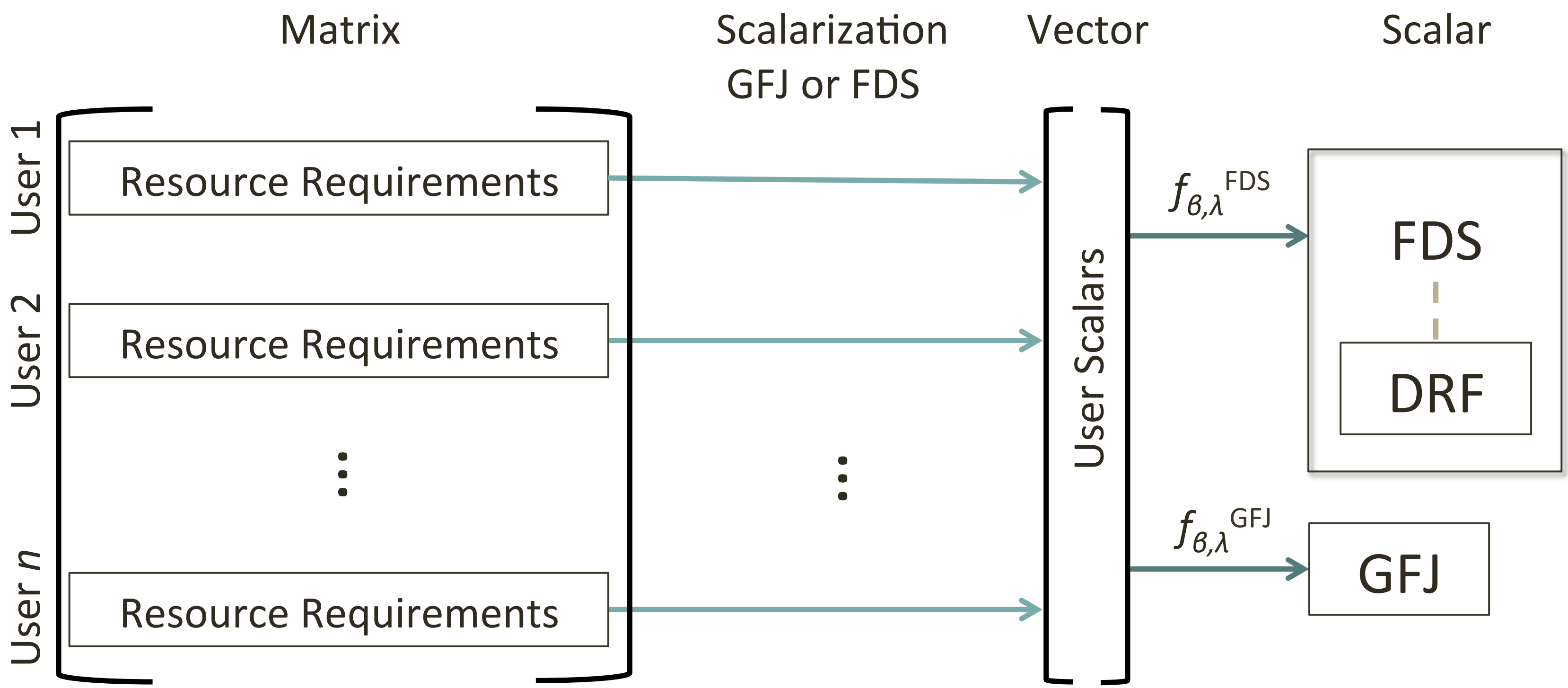
GFJ

Generalized Fairness on Jobs

FDS

Fairness on Dominant Shares

Resource Scalarization



Desirable Properties
Of Fairness Functions

Property 1: Pareto-Efficiency

- *f(x)* > *f(y)* whenever the allocation *x* Pareto-dominates *y*.
  - $x_i \geq y_i$ for all entries *i*, with strict inequality for some *i*
- Not an axiom: needs to be proven
- Does not hold for all parameter combinations

Parameter Conditions

- Necessary and sufficient conditions

$$|\lambda| \geq \left|\frac{1-\beta}{\beta}\right| \qquad \beta > 0$$

- Holds for FDS and GFJ
  - Comes from the same conditions for single-resource fairness
- If $\lambda = \frac{1-\beta}{\beta}$ and $\beta > 0$, fairness becomes α-fairness with $\alpha = \beta$.

T. Lan, et al. An Axiomatic Theory of Fairness in Network Resource Allocation. IEEE Infocom 2010.

Property 2: Sharing Incentive

- Each user receives at least a $^1/_n$ share of some resource.
    - Dominant share is over $^1/_n$
- Users don't want to share the resources equally.
- Does it hold?

Parameter Conditions

- Sufficient conditions:

FDS $\lambda = \frac{1-\beta}{\beta}$ $\beta > 1$

- Counterexamples exist:

$\lambda = \frac{1-\beta}{\beta}$ FDS $0 < \beta < 1$ GFJ $\beta > 0$

Property 3: Envy-Freeness

- A user can process more jobs with his own rather than another user's resource allocation.
  - Users don't want to switch allocations.
- Does it hold?

Parameter Conditions

- Sufficient conditions:

FDS $\lambda = \frac{1-\beta}{\beta}$ $\beta > 1$

- Counterexamples exist:

$\lambda = \frac{1-\beta}{\beta}$ FDS $0 < \beta < 1$ GFJ $\beta > 0$

Sufficient Conditions

Fairness	Pareto-Efficiency	Sharing Incentive	Envy-Freeness
FDS	$\lvert\lambda\rvert \geq \left\lvert\frac{1-\beta}{\beta}\right\rvert,\ \beta > 0$	$\lambda = \frac{1-\beta}{\beta}, \beta > 1$ $\lambda = 0$, any β	$\lambda = \frac{1-\beta}{\beta}, \beta > 1$ $\lambda = 0$, any β
GFJ	$\lvert\lambda\rvert \geq \left\lvert\frac{1-\beta}{\beta}\right\rvert,\ \beta > 0$	–	–

Existence of a Counterexample

Fairness	Sharing Incentive		Envy-Freeness	
FDS	$\lambda = \frac{1-\beta}{\beta}, 0 < \beta < 1$	$\lambda = \infty$, any β	$\lambda = \frac{1-\beta}{\beta}, 0 < \beta < 1$	$\lambda = \infty$, any β
GFJ	$\lambda = \frac{1-\beta}{\beta}, \beta > 0$ $\lvert\lambda\rvert < \frac{\lvert 1-\beta\rvert}{\beta}, \beta > 1$	$\lambda = \infty$ or 0, any β $\lvert\lambda\rvert > \frac{\lvert 1-\beta\rvert}{\beta}, 0 < \beta < 1$	$\lambda = \frac{1-\beta}{\beta}, \beta > 0$ $\lvert\lambda\rvert < \frac{\lvert 1-\beta\rvert}{\beta}, \beta > 1$	$\lambda = \infty$ or 0, any β $\lvert\lambda\rvert > \frac{\lvert 1-\beta\rvert}{\beta}, 0 < \beta < 1$

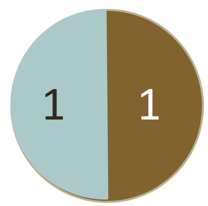
C. Joe-Wong, et al. Multi-Resource Allocation: Fairness-Efficiency Tradeoffs in a Unifying Framework. Tech report, available http://www.princeton.edu/~chiangm/multiresourcefairness.pdf

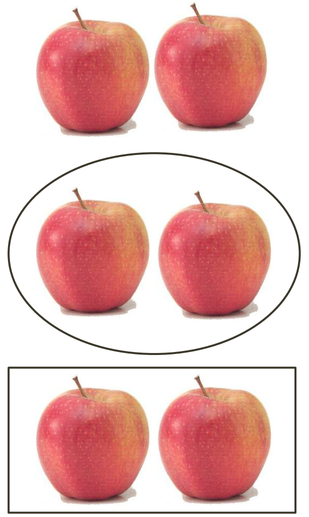
What about Efficiency?

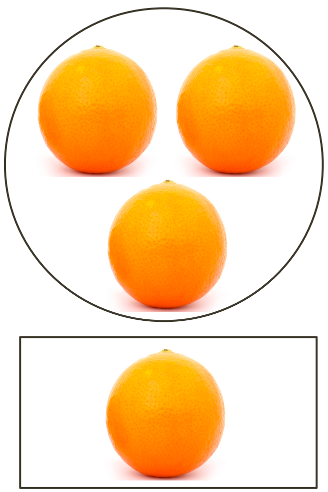
Fair, Efficient, or Both?

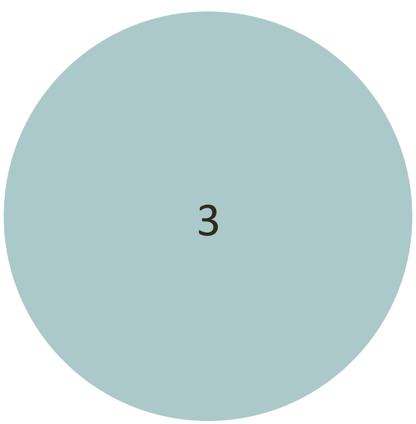
User 1

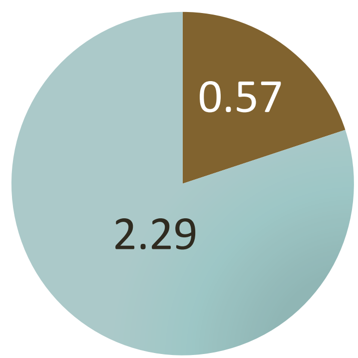
User 2

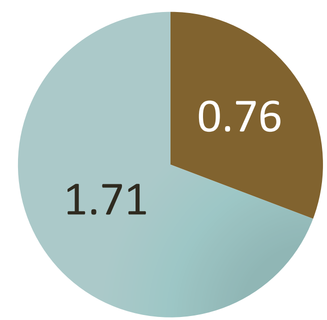








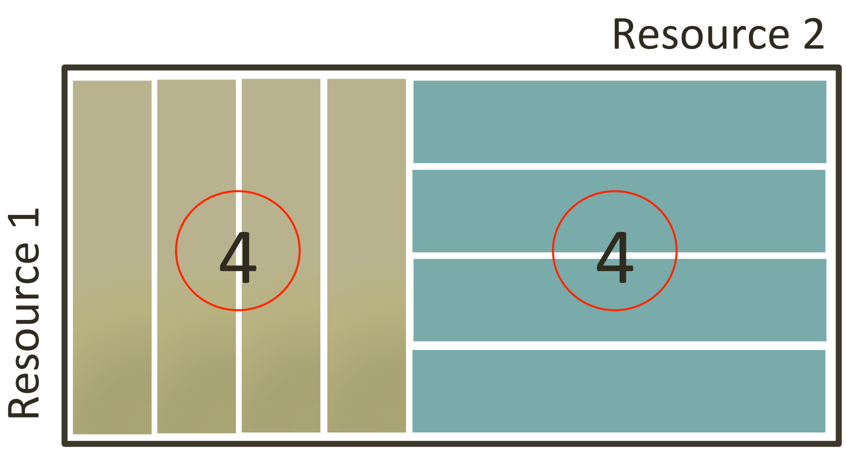




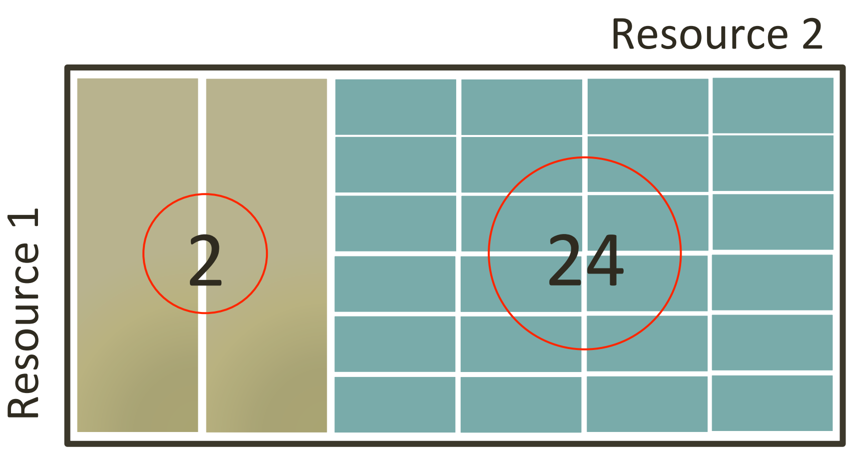
Existence of a Tradeoff

- Nonlinear, non-separable, multidimensional, continuous state-space knapsack problem
  - Maximize fairness function subject to multiple linear capacity constraints
  - Allow fractional jobs

No Tradeoff



Tradeoff Exists



Equal Allocations at Maximum Efficiency

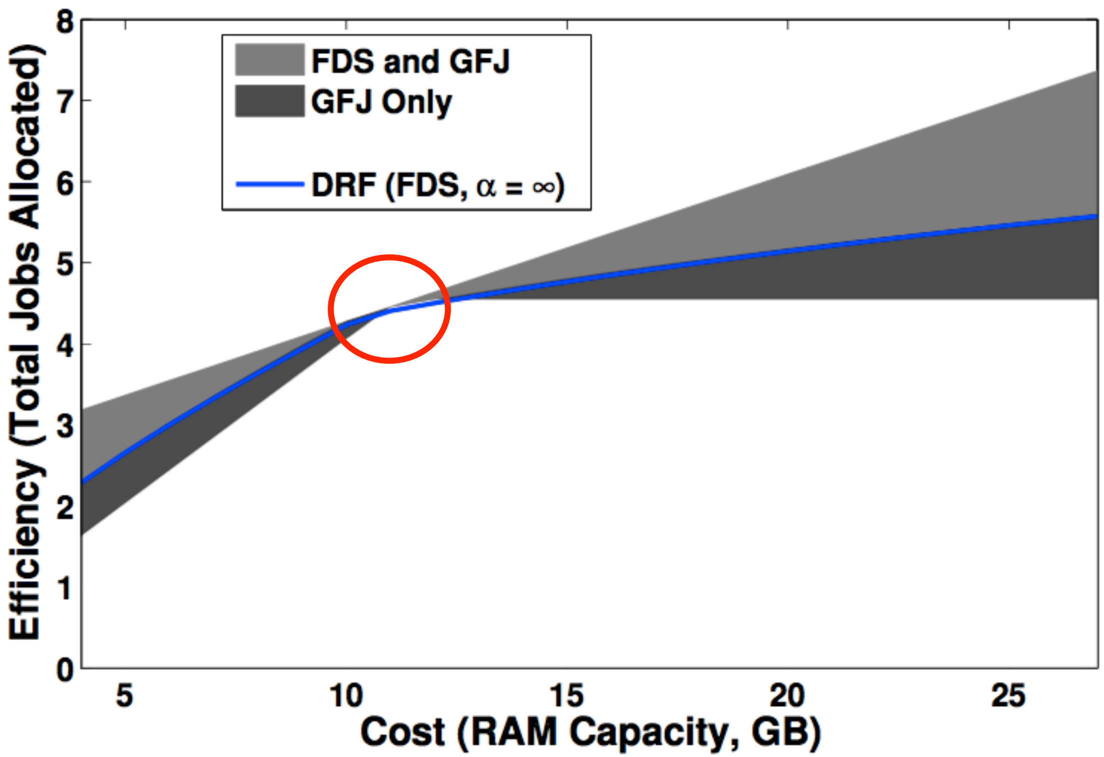
- Number of tight resource constraints = number of users

$$\sum_{j=1}^{n} \gamma_{ij} x_j \leq 1 \, \forall \, i.$$

FDS $$\sum_{j=1}^{n} \frac{\gamma_{ij}}{\mu_j} = \rho$$

GFJ $$\sum_{j=1}^{n} \gamma_{ij} = r$$

Efficiency Operating Range



$$x + 3y \leq 9$$ CPUs

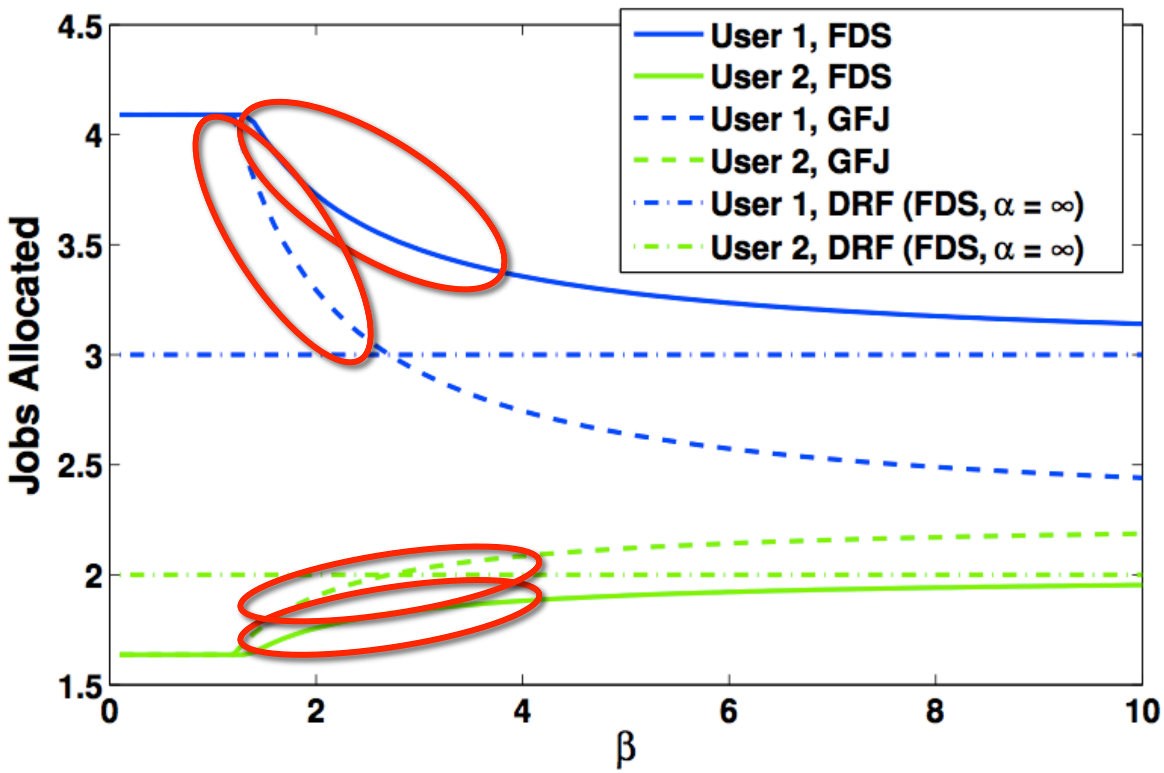
$$4x + y \leq 18$$ GB

Optimal allocations for a range of β and λ

Dominant Resource Fairness (DRF): max-min fairness on dominant shares

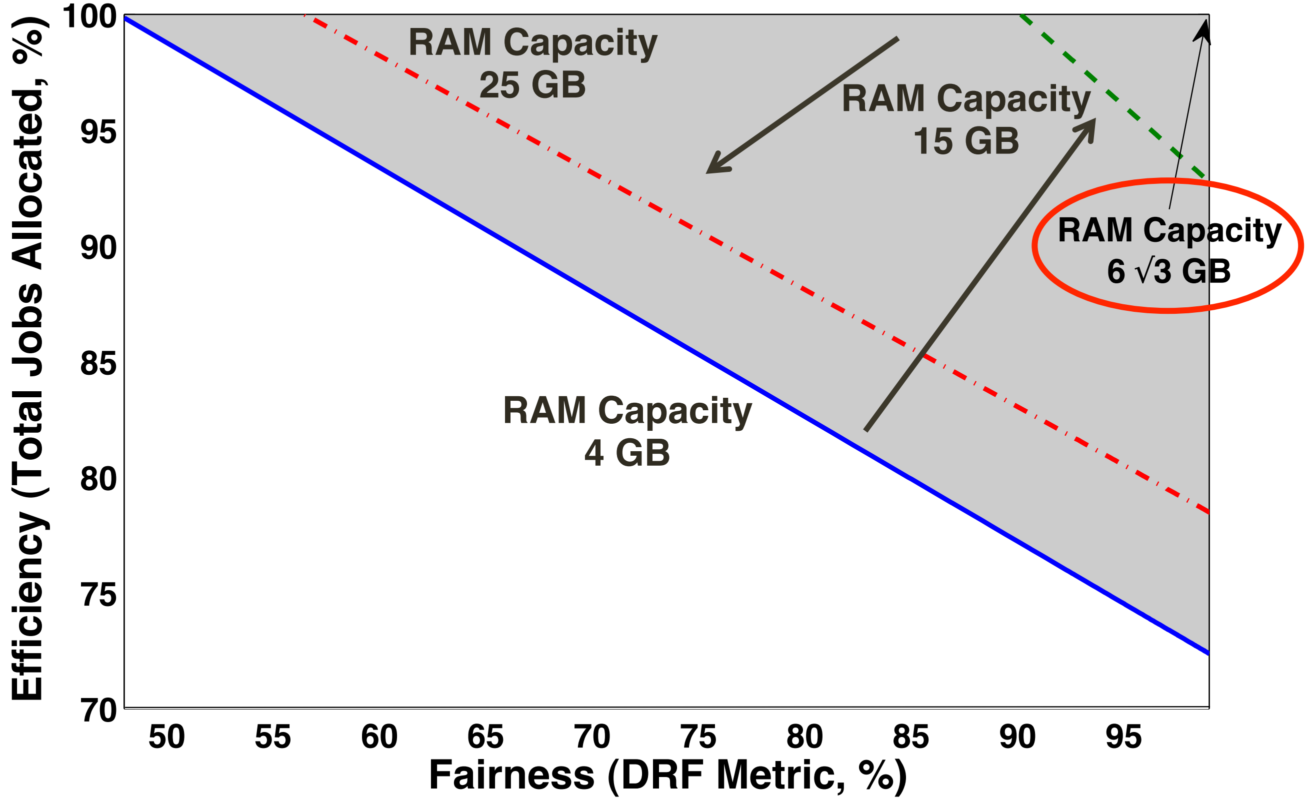
Job Allocation

$$\lambda = \frac{1-\beta}{\beta}$$



Optimal allocations for α = β-fairness

Numerical Example



- Fairness: DRF-fairness divided by maximal DRF value
- Efficiency: Total jobs divided by maximum number of jobs

Psychological Perceptions

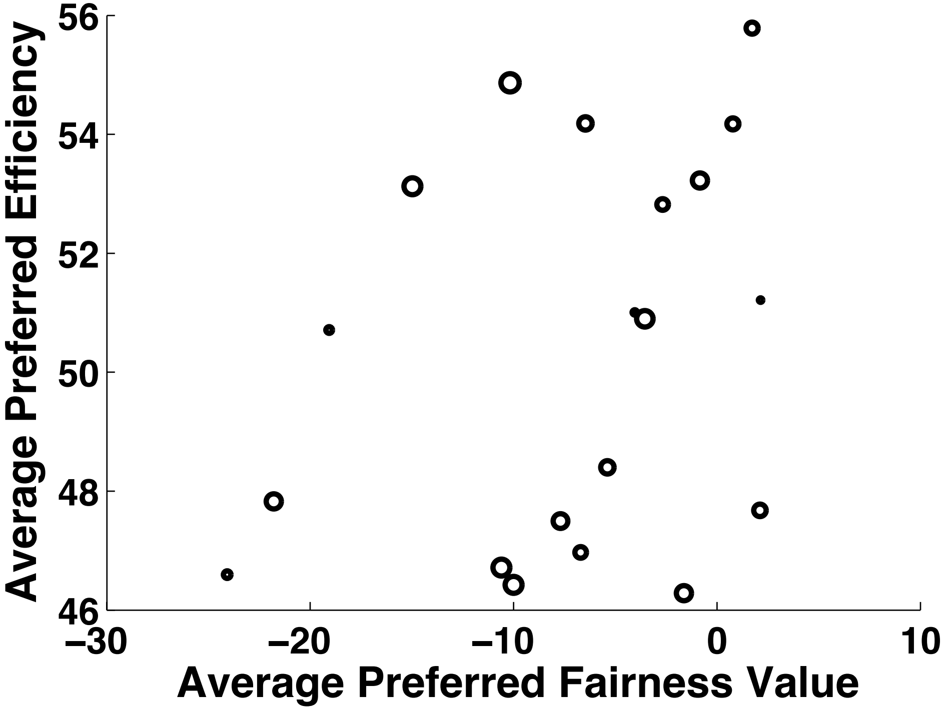
C. Joe-Wong, et al. Multi-Resource Allocation: Fairness-Efficiency Tradeoffs in a Unifying Framework. Tech report, available http://www.princeton.edu/~chiangm/multiresourcefairness.pdf

Parameter Values

- What parameters are compatible with the responses?
  - Do they satisfy Pareto-efficiency, etc.?
- Do people agree with each other?
- Online survey asking people to rank datacenter allocations

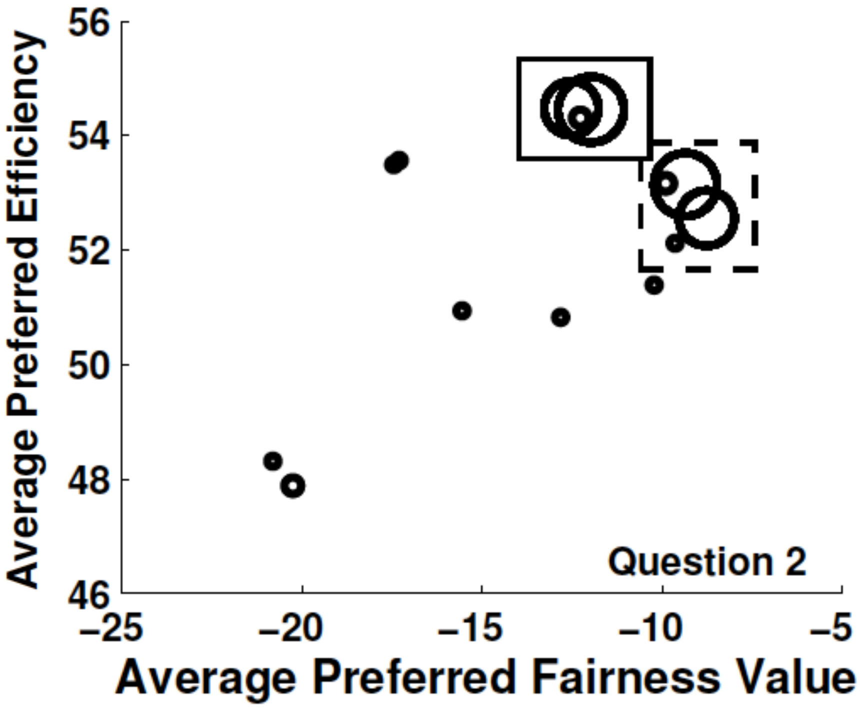
Allocation Options	Allocated to Client A			Allocated to Client B			Total no. of Jobs Completed
	CPU	TB	No. of Jobs Completed for Client A	CPU	TB	No. of Jobs Completed for Client B	
Allocation 1	24	96	**24**	84	28	**28**	**52**

Are People Very Different?

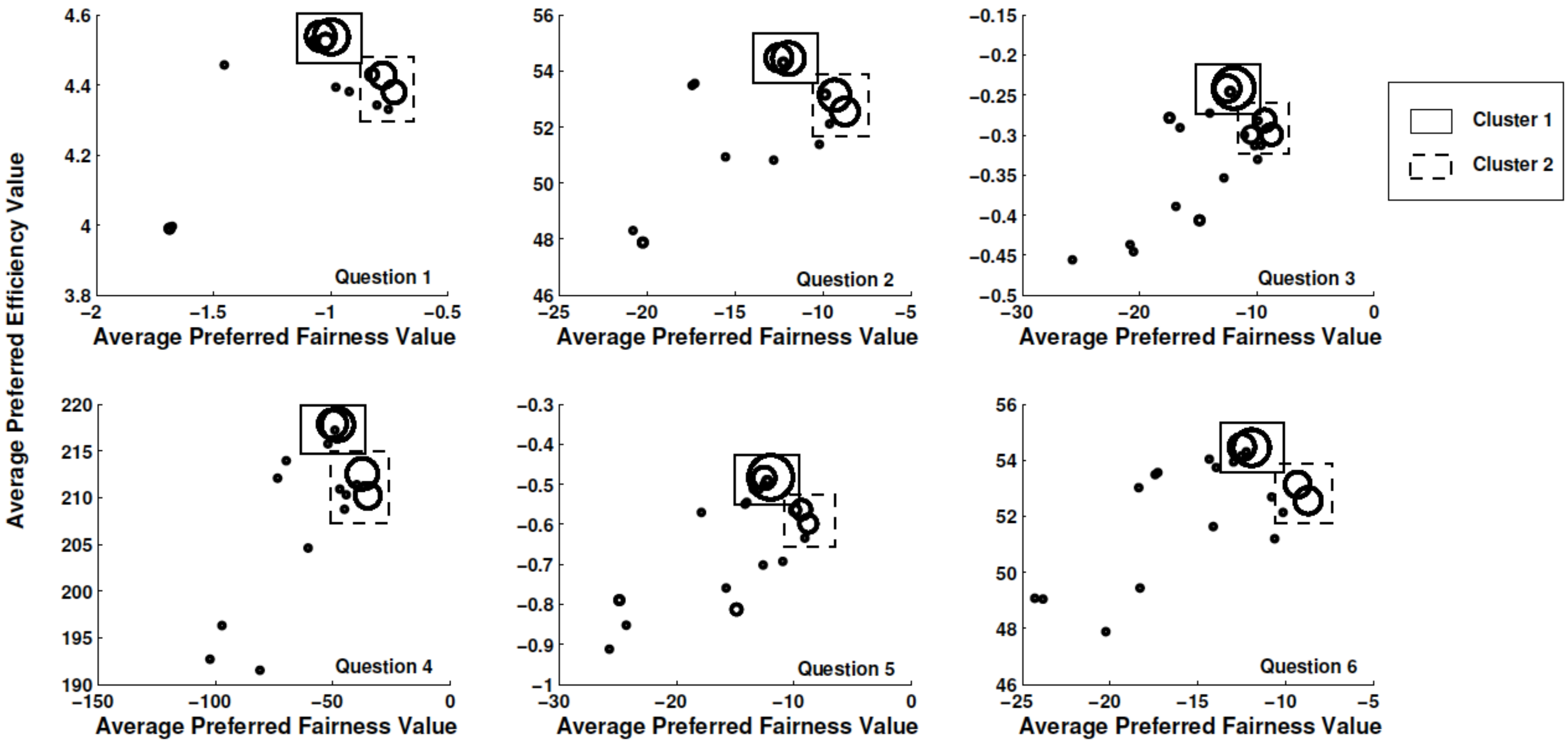


Actually They're Pretty Similar

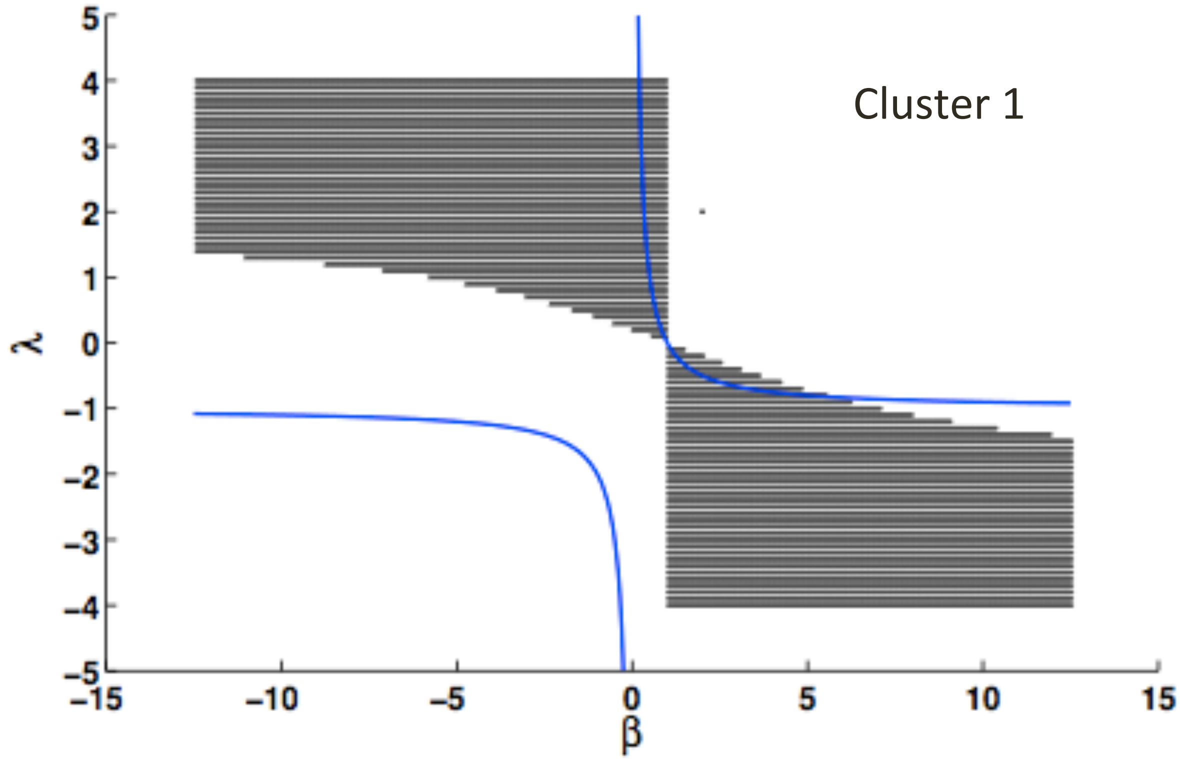
- Cluster 1 prefers efficiency to fairness
- Cluster 2 prefers fairness to efficiency



All Responses



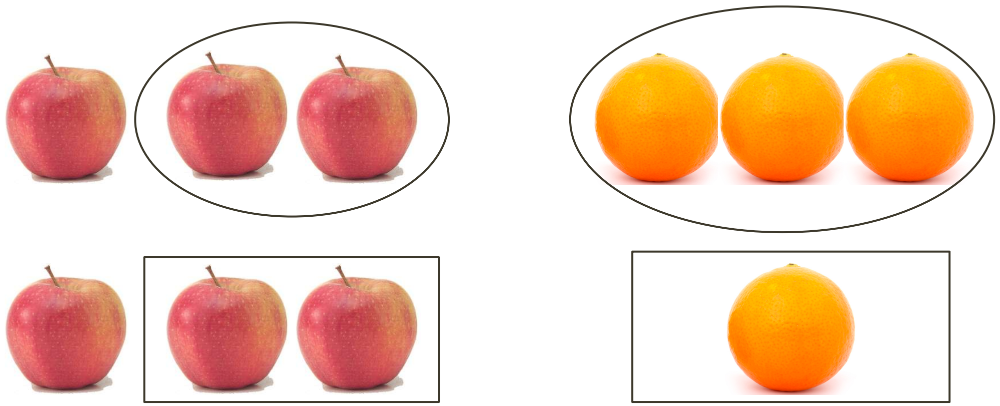
Compatible Parameters (GFJ)



The darker the square, the more participant rankings were compatible.

Lines represent Pareto-efficient frontiers.

Back to the Motivation



Questions Answered

- How do we define fairness?
  - GFJ and FDS
- Are these properties satisfied?
  - Pareto-efficiency
  - Envy-freeness
  - Sharing incentive
- Does a fairness-efficiency tradeoff exist?
- What parameters are consistent with actual preferences?
  - Users fall into 2 clusters

Photo: Cam Barker

Thank you!

Questions?