Poly-Symmetry in Processor-Sharing Systems

Céline Comte[1]

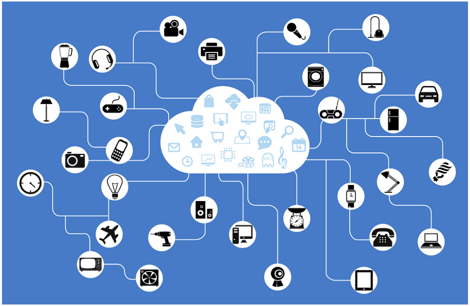
Joint work with Thomas Bonald[1], Virag Shah[2]
and Gustavo de Veciana[3]

[1]Télécom ParisTech, [2]MSR-Inria Joint Center, [3]University of Texas at Austin

April 28, 2017

Motivation

- Data networks



Motivation

- Data networks
- Computer clusters

Motivation

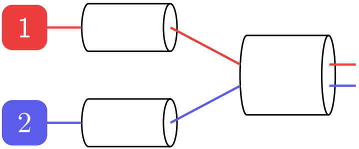
- Data networks
- Computer clusters
- Content delivery networks

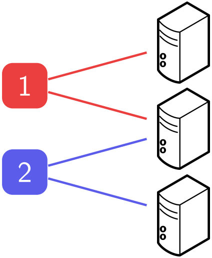
Motivation

- Data networks
- Computer clusters
- Content delivery networks
- ...

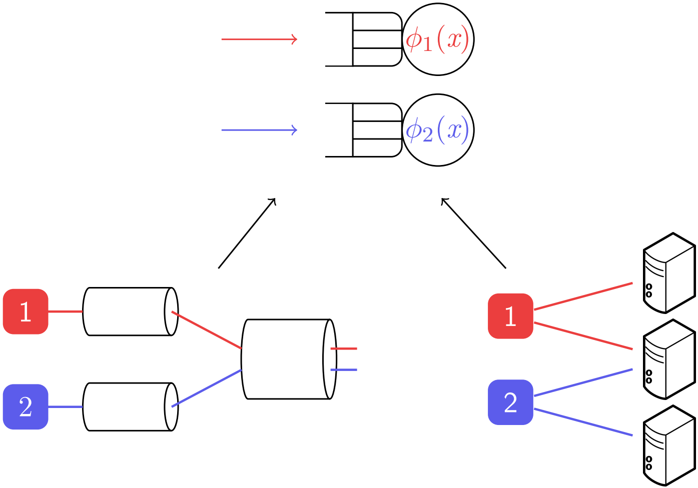
Common features

- **Heterogeneous stochastic demands**
  Random arrival times, variable sizes
- **Concurrent access to limited resources**
  Link bandwidth, server capacity, . . .



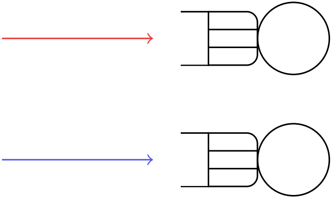


One model to rule them all



Queueing model

Processor-sharing system



Queueing system

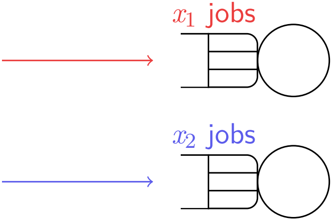
- n processor-sharing queues
- State $x = (x_1, \dots, x_n)$
- Service capacities $\phi(x) = (\phi_1(x), \dots, \phi_n(x))$

Job model

- Poisson arrivals with rate λ_i
- Sizes i.i.d. with mean σ_i

$\rightarrow$ Traffic intensity $\rho_i = \lambda_i \sigma_i$

Processor-sharing system



Queueing system

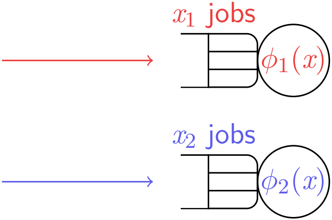
- n processor-sharing queues
- State $x = (x_1, \ldots, x_n)$
- Service capacities $\phi(x) = (\phi_1(x), \ldots, \phi_n(x))$

Job model

- Poisson arrivals with rate λ_i
- Sizes i.i.d. with mean σ_i

$\rightarrow$ Traffic intensity $\rho_i = \lambda_i \sigma_i$

Processor-sharing system



Queueing system

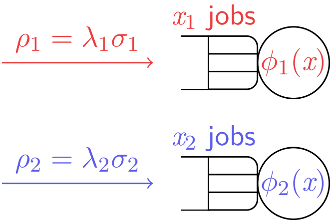
- n processor-sharing queues
- State $x = (x_1, \dots, x_n)$
- Service capacities
  $\phi(x) = (\phi_1(x), \dots, \phi_n(x))$

Job model

- Poisson arrivals with rate λ_i
- Sizes i.i.d. with mean σ_i

$\rightarrow$ Traffic intensity $\rho_i = \lambda_i \sigma_i$

Processor-sharing system



Queueing system

- n processor-sharing queues
- State $x = (x_1, \ldots, x_n)$
- Service capacities $\phi(x) = (\phi_1(x), \ldots, \phi_n(x))$

Job model

- Poisson arrivals with rate λ_i
- Sizes i.i.d. with mean σ_i

$\rightarrow$ Traffic intensity $\rho_i = \lambda_i\sigma_i$

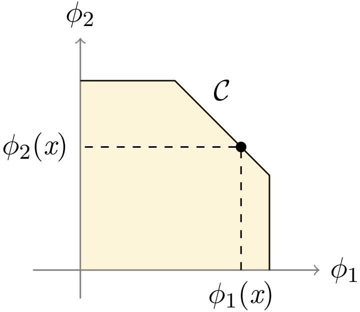
Polymatroid capacity set

Capacity set $\mathcal{C}$

- Set of acceptable vectors ϕ
- Operational constraints of the real system

Assumptions

- Only depends on the set of active queues
- Submodularity



$$\mathcal{C} = \left\{ \phi \in \mathbb{R}_+^n : \ \forall A \subset \{1, \dots, n\}, \ \sum_{i \in A} \phi_i \le \mu(A) \right\}$$

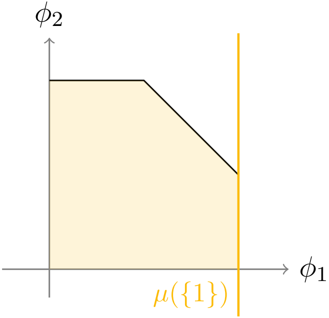
Polymatroid capacity set

Capacity set $\mathcal{C}$

- Set of acceptable vectors ϕ
- Operational constraints of the real system

Assumptions

- Only depends on the set of active queues
- Submodularity



$$\mathcal{C} = \left\{ \phi \in \mathbb{R}_+^n : \forall A \subset \{1, \dots, n\}, \sum_{i \in A} \phi_i \leq \mu(A) \right\}$$

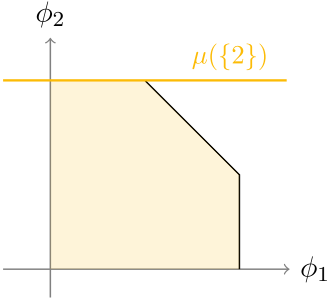
Polymatroid capacity set

Capacity set $\mathcal{C}$

- Set of acceptable vectors ϕ
- Operational constraints of the real system

Assumptions

- Only depends on the set of active queues
- Submodularity



$$\mathcal{C} = \left\{ \phi \in \mathbb{R}_+^n : \; \forall A \subset \{1, \ldots, n\}, \; \sum_{i \in A} \phi_i \le \mu(A) \right\}$$

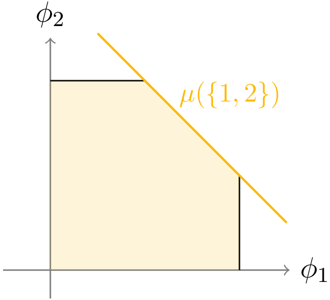
Polymatroid capacity set

Capacity set $\mathcal{C}$

- Set of acceptable vectors ϕ
- Operational constraints of the real system

Assumptions

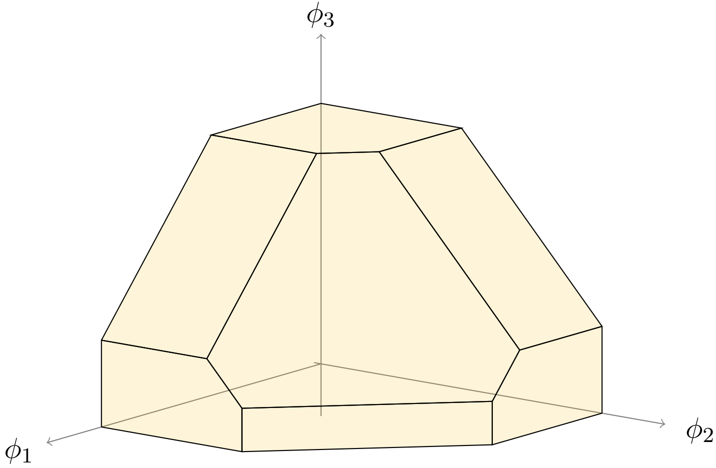
- Only depends on the set of active queues
- Submodularity



$$\mathcal{C} = \left\{ \phi \in \mathbb{R}_+^n : \ \forall A \subset \{1, \ldots, n\}, \ \sum_{i \in A} \phi_i \leq \mu(A) \right\}$$

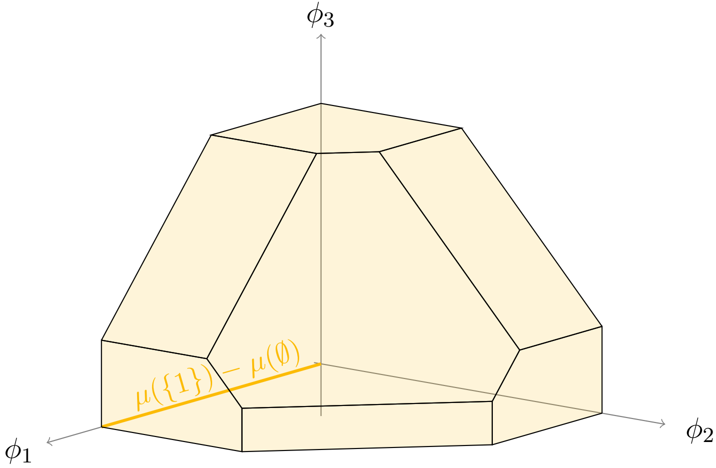
Polymatroid capacity set

Submodularity: $\mu(A \cup \{i\}) - \mu(A)$ decreases when A increases



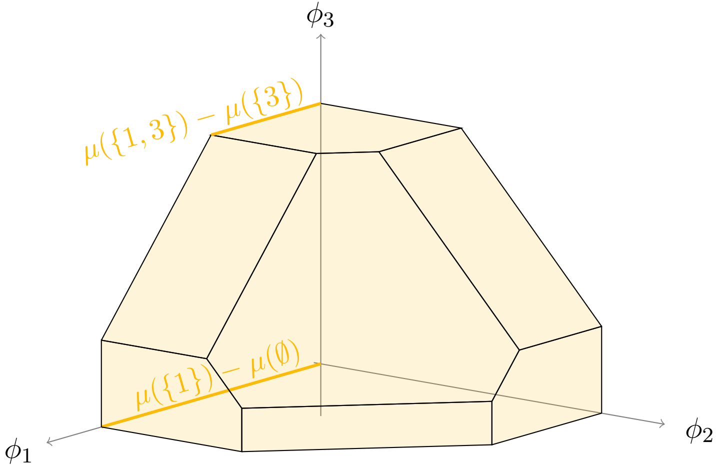
Polymatroid capacity set

Submodularity: $\mu(A \cup \{i\}) - \mu(A)$ decreases when A increases



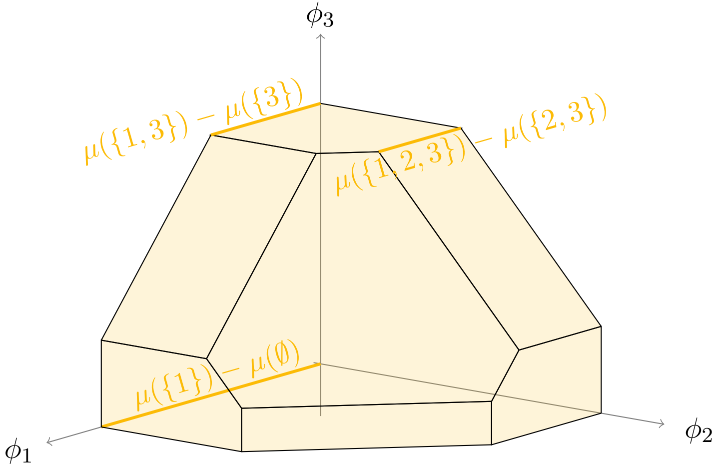
Polymatroid capacity set

Submodularity: $\mu(A \cup \{i\}) - \mu(A)$ decreases when A increases



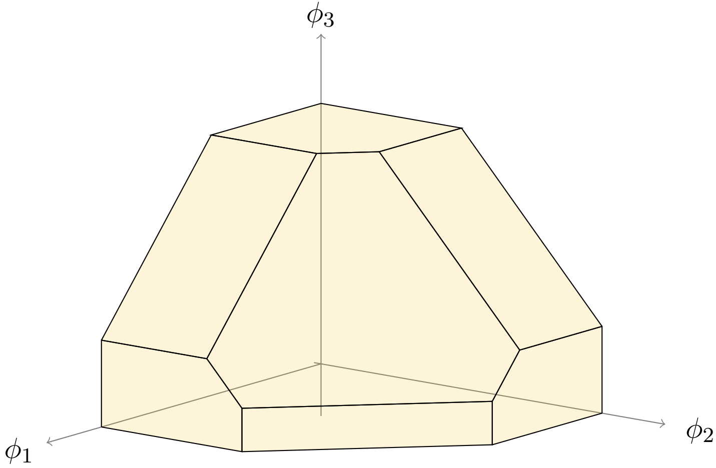
Polymatroid capacity set

Submodularity: $\mu(A \cup \{i\}) - \mu(A)$ decreases when A increases



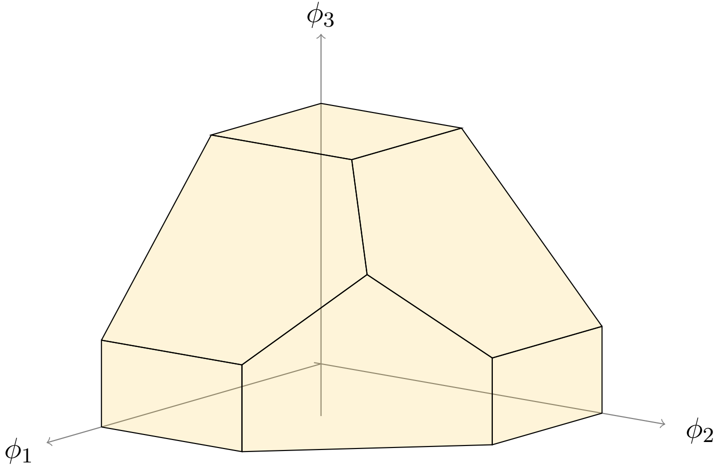
Polymatroid capacity set

Submodularity: $\mu(A \cup \{i\}) - \mu(A)$ decreases when A increases



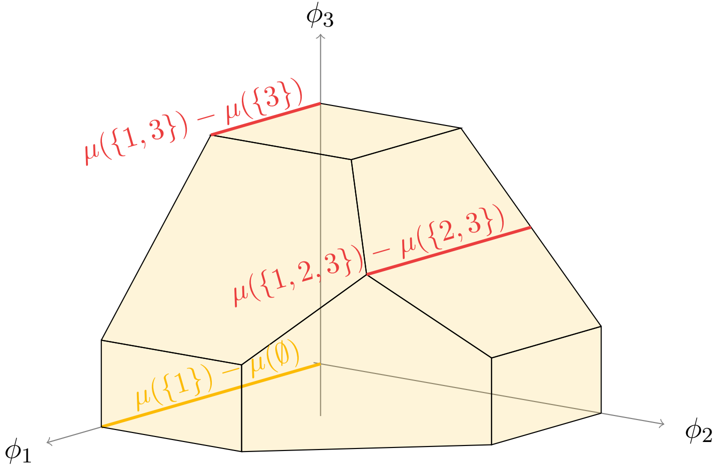
Polymatroid capacity set

Submodularity: $\mu(A \cup \{i\}) - \mu(A)$ decreases when A increases

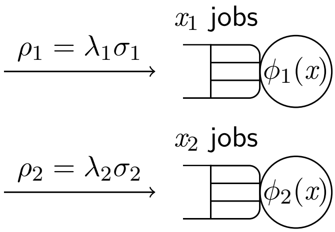


Polymatroid capacity set

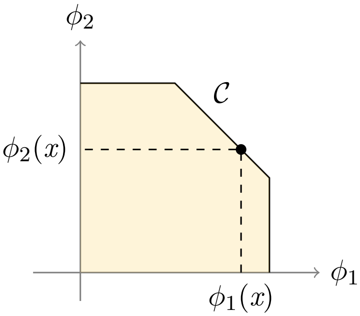
Submodularity: $\mu(A \cup \{i\}) - \mu(A)$ decreases when A increases



Processor-sharing system



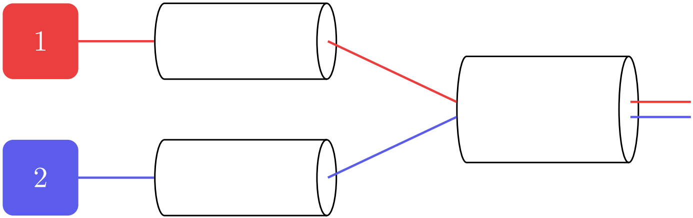
Polymatroid capacity set



Tree data networks

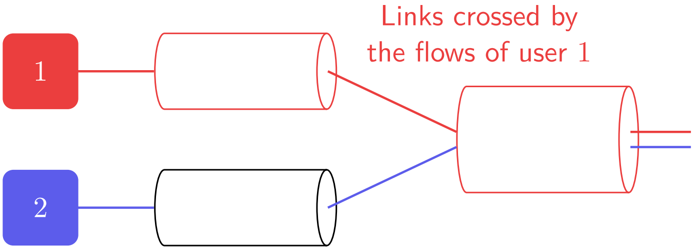
User routes

n users
indexed by I

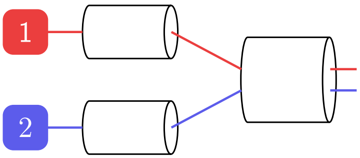


User routes

n users
indexed by I



System dynamics



Resource sharing

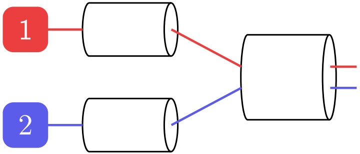
- Divisibility of the link capacity
- All flows of a user receive the same capacity

Flow model

- Poisson arrivals with rate λ_i
- Sizes i.i.d. with mean σ_i

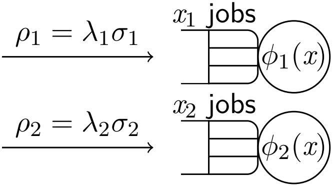
$\rightarrow$ Traffic intensity $\rho_i = \lambda_i \sigma_i$

System dynamics

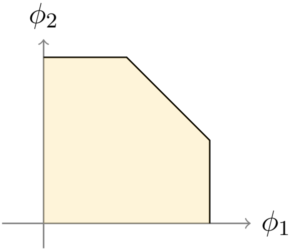


→

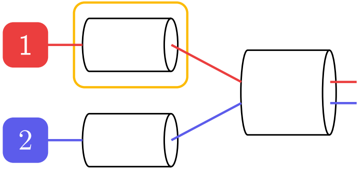
System of PS queues



Capacity region

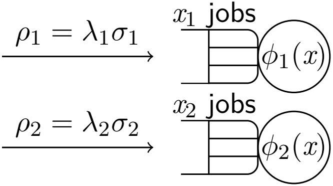


System dynamics

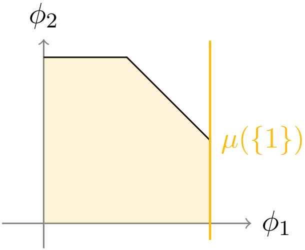


$\rightarrow$

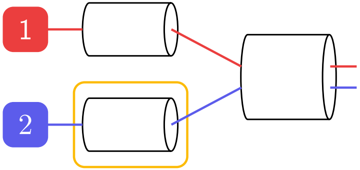
System of PS queues



Capacity region

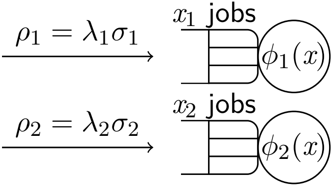


System dynamics

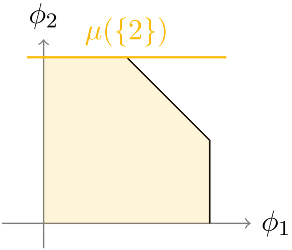


→

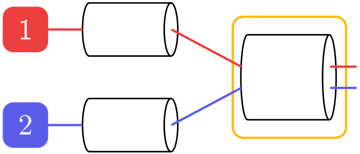
System of PS queues



Capacity region

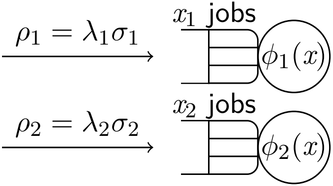


System dynamics

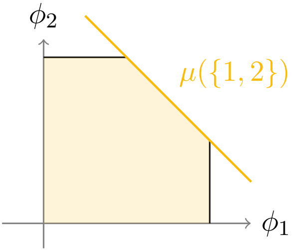


$\rightarrow$

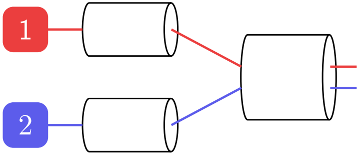
System of PS queues



Capacity region

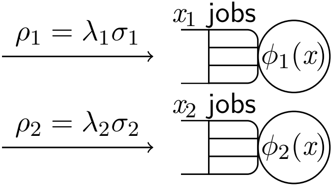


System dynamics

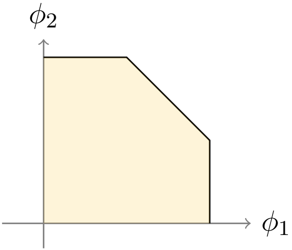


$\rightarrow$

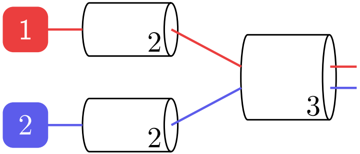
System of PS queues



Capacity region

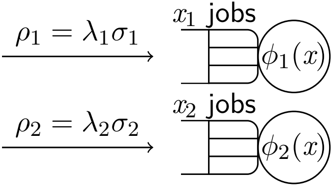


System dynamics

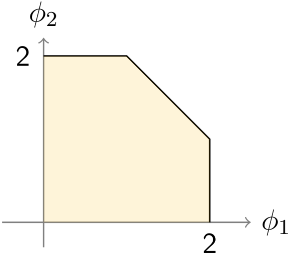


→

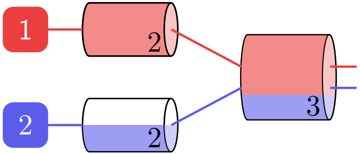
System of PS queues



Capacity region

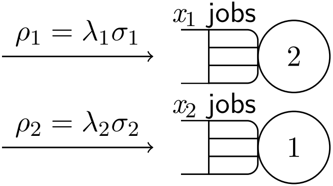


System dynamics

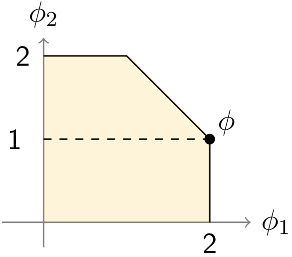


→

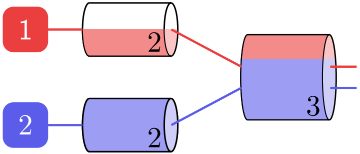
System of PS queues



Capacity region

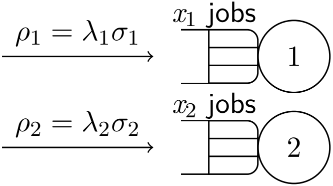


System dynamics

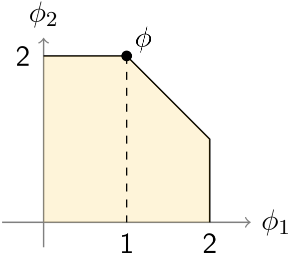


→

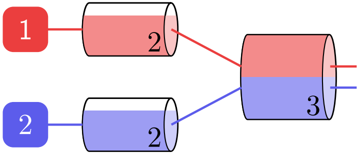
System of PS queues



Capacity region

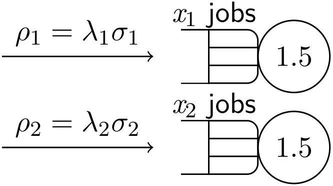


System dynamics

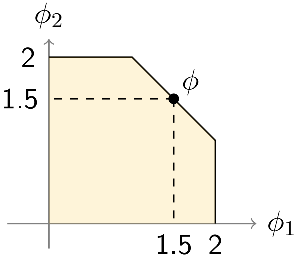


→

System of PS queues



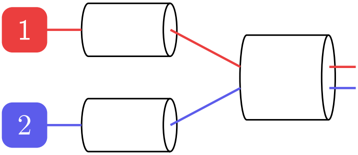
Capacity region

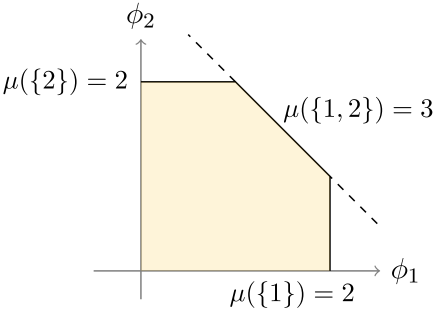


Polymatroid capacity set

$\mu(A)$ = capacity of the most constraining links for the users in A

$$\mathcal{C} = \left\{ \phi \in \mathbb{R}_+^n : \forall A \subset I, \sum_{i \in A} \phi_i \le \mu(A) \right\}$$



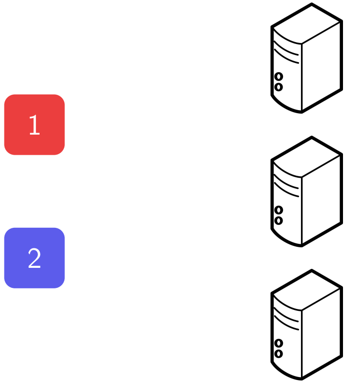


Computer clusters

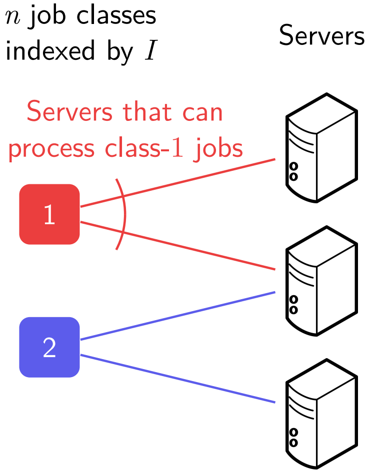
Server assignment - Bipartite graph

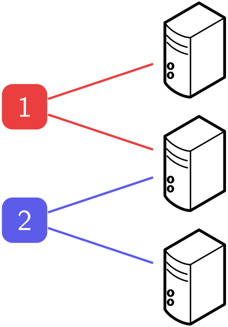
n job classes
indexed by I

Servers



Server assignment - Bipartite graph





Resource sharing

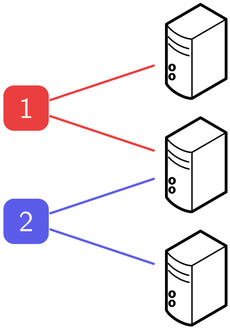
- Divisibility of the server capacity
- Parallel processing
- All jobs of a class receive the same service

Job model

- Poisson arrivals with rate λ_i
- Sizes i.i.d. with mean σ_i

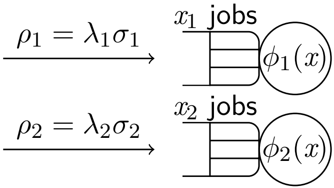
$\rightarrow$ Traffic intensity $\rho_i = \lambda_i \sigma_i$

System dynamics

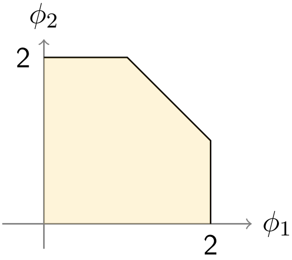


$\rightarrow$

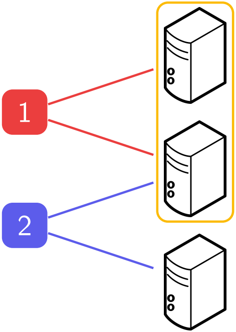
System of PS queues



Capacity region

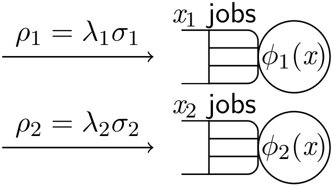


System dynamics

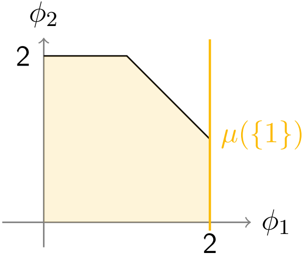


→

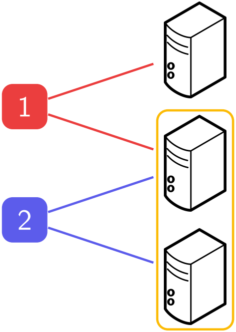
System of PS queues



Capacity region

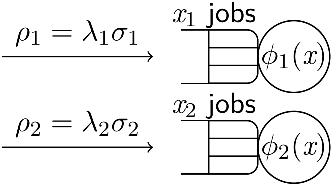


System dynamics

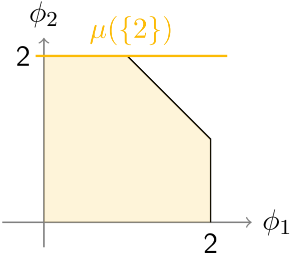


$\rightarrow$

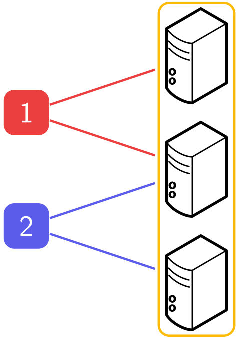
System of PS queues



Capacity region

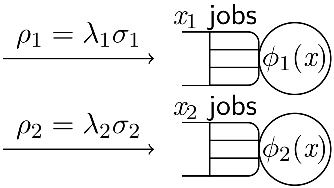


System dynamics

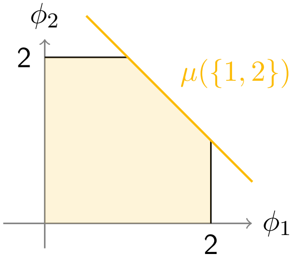


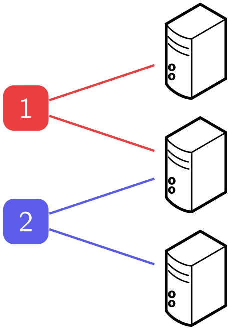
$\rightarrow$

System of PS queues



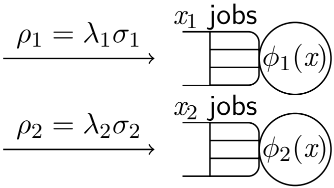
Capacity region



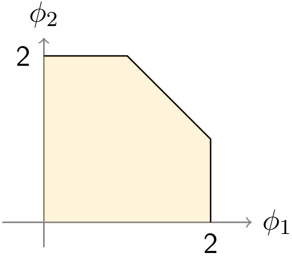


→

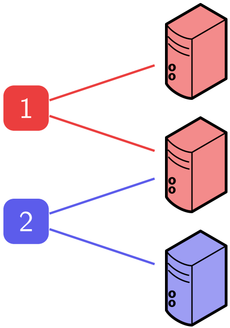
System of PS queues



Capacity region

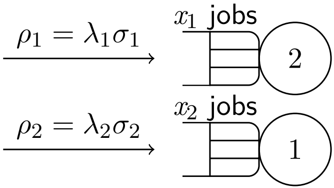


System dynamics

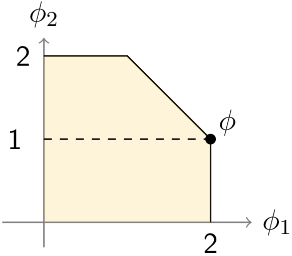


$\rightarrow$

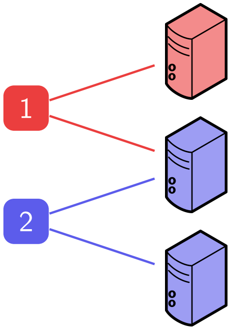
System of PS queues



Capacity region

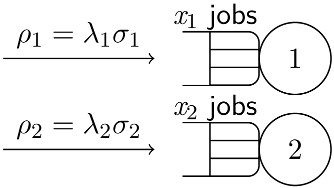


System dynamics

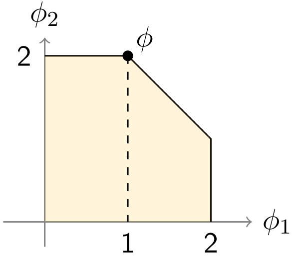


$\longrightarrow$

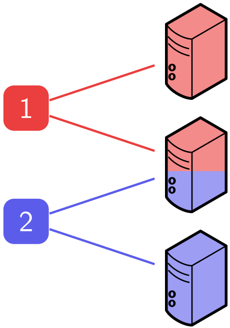
System of PS queues

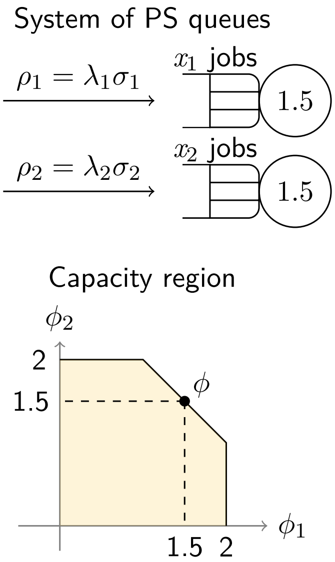


Capacity region



System dynamics

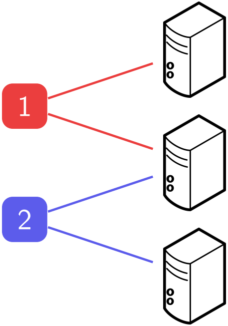


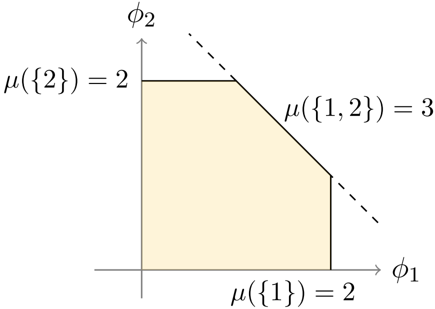


Polymatroid capacity set (Shah and de Veciana, 2015)

$\mu(A)$ = aggregate capacity of the servers of the classes in A

$$\mathcal{C} = \left\{ \phi \in \mathbb{R}_+^n : \ \forall A \subset I, \ \sum_{i \in A} \phi_i \le \mu(A) \right\}$$

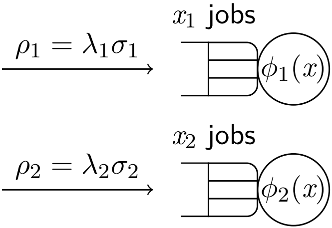




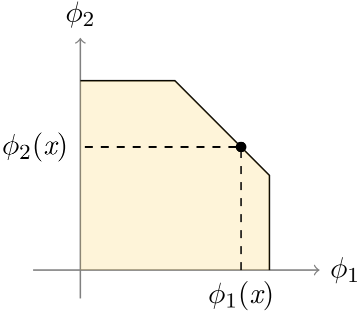
Balanced fairness

Resource allocation

Processor-sharing system



Polymatroid capacity set



How are $\phi_1(x)$ and $\phi_2(x)$ allocated ?

Balanced fairness

The most efficient insensitive resource allocation

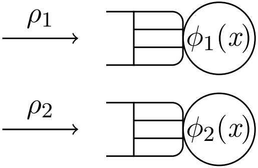
- Introduced for dimensioning data networks (Bonald and Proutière, 2003)
- Good approximation of proportional fairness
- Recently applied to Content Delivery Networks (Shah and de Veciana, 2015 and 2016)

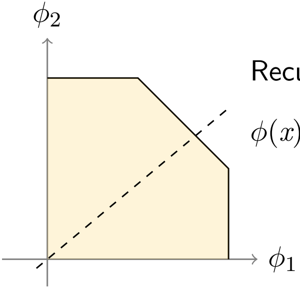
Definition

- Balance property: $\forall i, j \in I(x)$,

$$\frac{\phi_i(x - e_j)}{\phi_i(x)} = \frac{\phi_j(x - e_i)}{\phi_j(x)}.$$

Whittle network



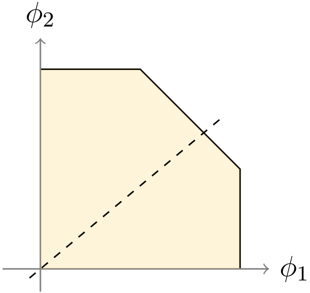


Recursive construction

$$\phi(x) = \begin{pmatrix} \phi_1(x) \\ \phi_2(x) \end{pmatrix} \propto \begin{pmatrix} \phi_1(x - e_2) \\ \phi_2(x - e_1) \end{pmatrix}$$

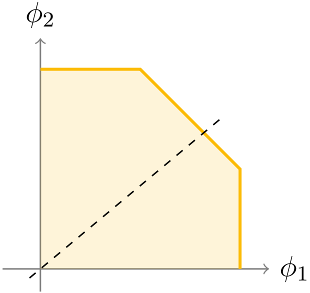
Definition

- Maximize the resource utilization



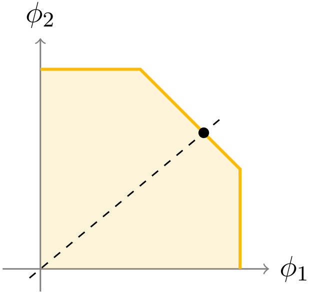
Definition

- Maximize the resource utilization



Definition

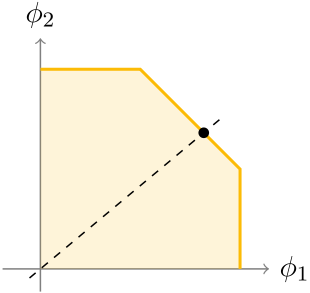
- Maximize the resource utilization



Definition

- Maximize the resource utilization

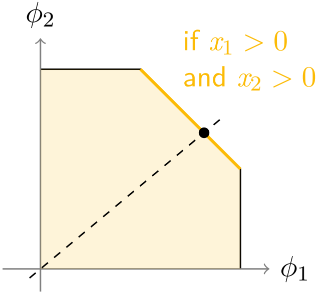
  Pareto-efficiency in polymatroids
  (Shah and de Veciana, 2015)



Definition

- Maximize the resource utilization

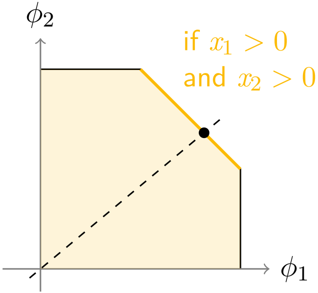
Pareto-efficiency in polymatroids
(Shah and de Veciana, 2015)



Definition

- Maximize the resource utilization

  Pareto-efficiency in polymatroids
  (Shah and de Veciana, 2015)



$\rightarrow$ Explicit recursion formulas for the performance metrics

Performance metrics

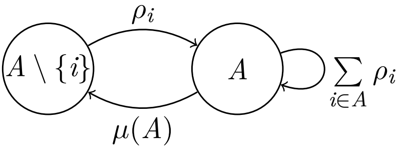
- Law of total expectation

$$\mathbb{E}(\mathbf{X}_i) = \sum_{A \subset I} \pi(A) \times \mathbb{E}\left(\mathbf{X}_i \middle| I(\mathbf{X}) = A\right)$$

with $\pi(A) = \mathbb{P}(I(\mathbf{X}) = A)$

- $\pi(A)$, $\mathbb{E}\left(\mathbf{X}_i \middle| I(\mathbf{X}) = A\right)$ computed recursively

$$\mu(A)\pi(A) = \sum_{i \in A} \rho_i \pi(A \setminus \{i\}) + \sum_{i \in A} \rho_i \pi(A)$$



Interpretation:
Reversibility
+ PASTA property

Poly-symmetry

Symmetry

All classes are exchangeable

Symmetry

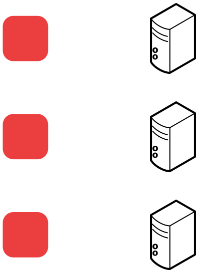
All classes are exchangeable

- Same resource occupation
- Same traffic intensity

Symmetry

All classes are exchangeable

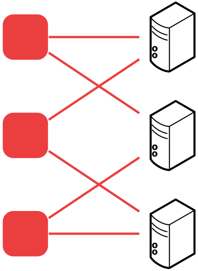
- Same resource occupation
- Same traffic intensity



Symmetry

All classes are exchangeable

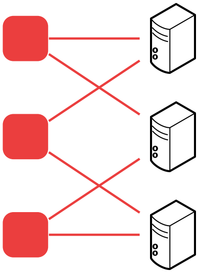
- Same resource occupation
- Same traffic intensity

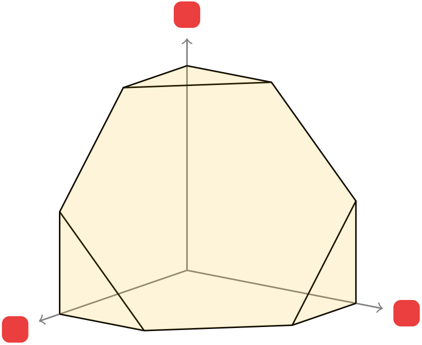


Symmetry

All classes are exchangeable

- Same resource occupation
- Same traffic intensity

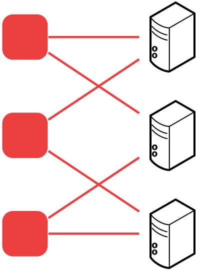


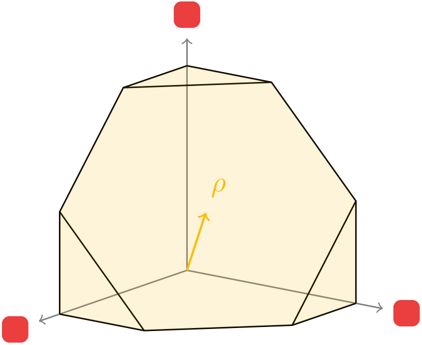


Symmetry

All classes are exchangeable

- Same resource occupation
- Same traffic intensity

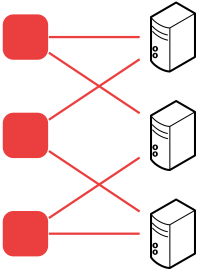


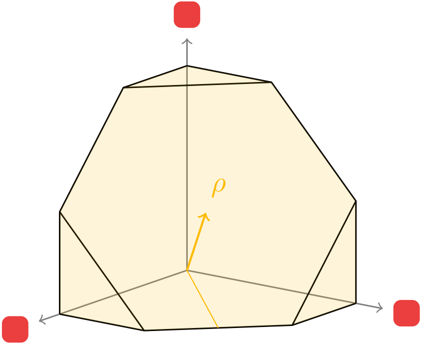


Symmetry

All classes are exchangeable

- Same resource occupation
- Same traffic intensity

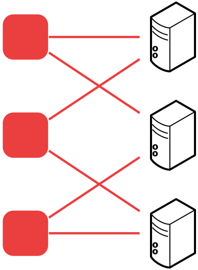


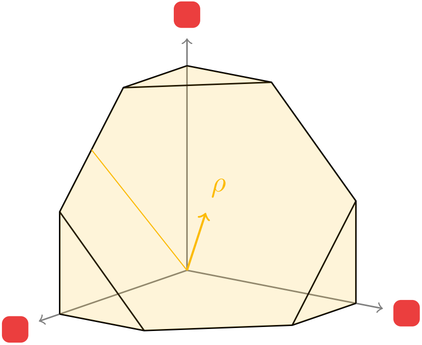


Symmetry

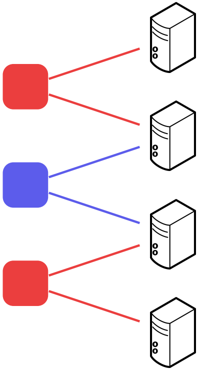
All classes are exchangeable

- Same resource occupation
- Same traffic intensity

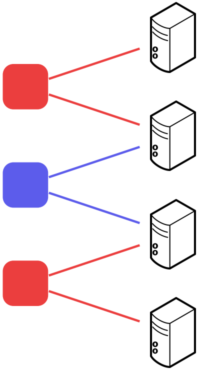


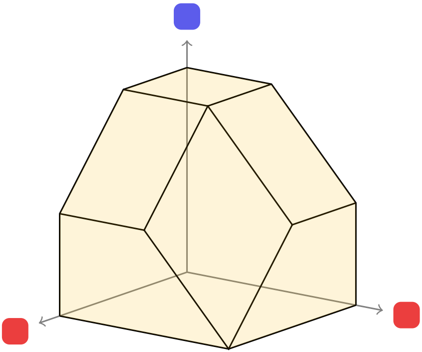


A non-symmetric system

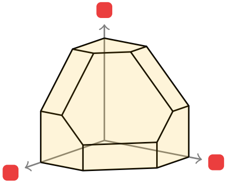


A non-symmetric system



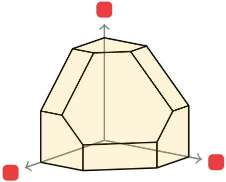


Symmetric



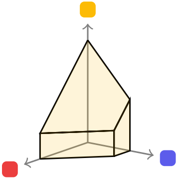
$O(n)$

Symmetric



$O(n)$

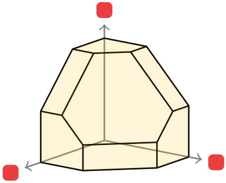
Heterogeneous



$O(2^n)$

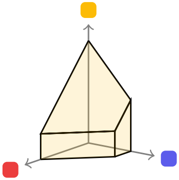
Symmetry

Symmetric



$O(n)$

Heterogeneous

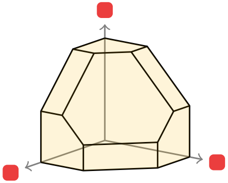


$O(2^n)$

Complexity

Symmetry

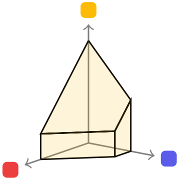
Symmetric



$O(n)$

?

Heterogeneous



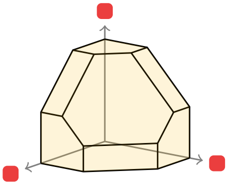
$O(2^n)$

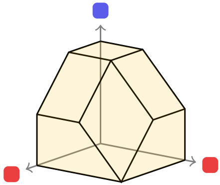
Complexity

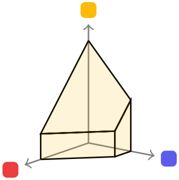
Symmetry

Symmetric

Heterogeneous







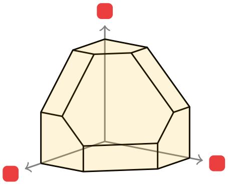
$O(n)$

$O(2^n)$

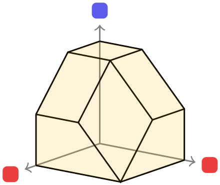
Complexity

Symmetry

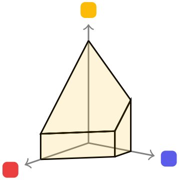
Symmetric | Poly-symmetric | Heterogeneous



$O(n)$



$O\left(n^K\right)$



$O(2^n)$

Complexity

Poly-symmetry

All classes **of the same type**
are exchangeable

Poly-symmetry

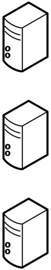
All classes **of the same type**
are exchangeable

- Same resource occupation
- Same traffic intensity

Poly-symmetry

All classes **of the same type**
are exchangeable

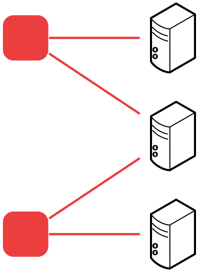
- Same resource occupation
- Same traffic intensity



Poly-symmetry

All classes **of the same type** are exchangeable

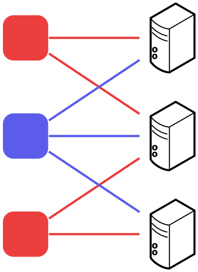
- Same resource occupation
- Same traffic intensity



Poly-symmetry

All classes **of the same type** are exchangeable

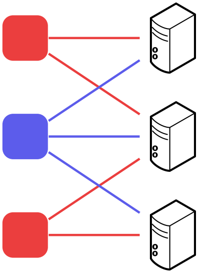
- Same resource occupation
- Same traffic intensity

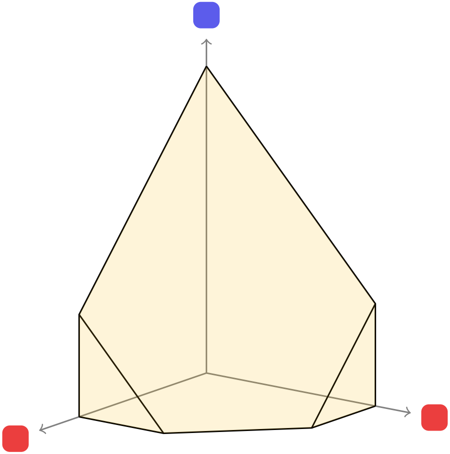


Poly-symmetry

All classes **of the same type** are exchangeable

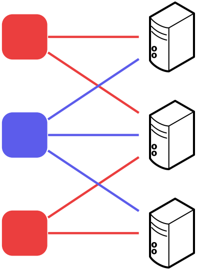
- Same resource occupation
- Same traffic intensity

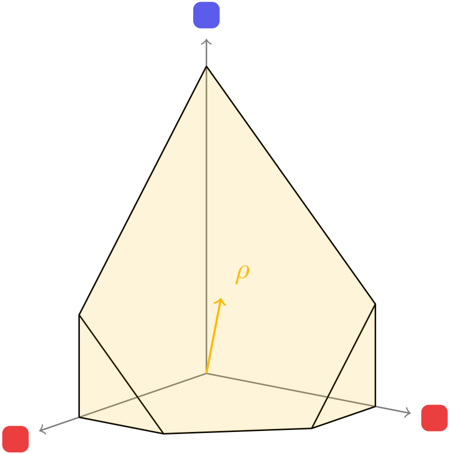




All classes **of the same type** are exchangeable

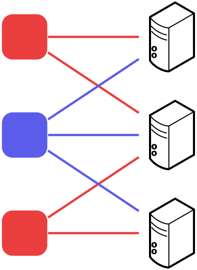
- Same resource occupation
- Same traffic intensity

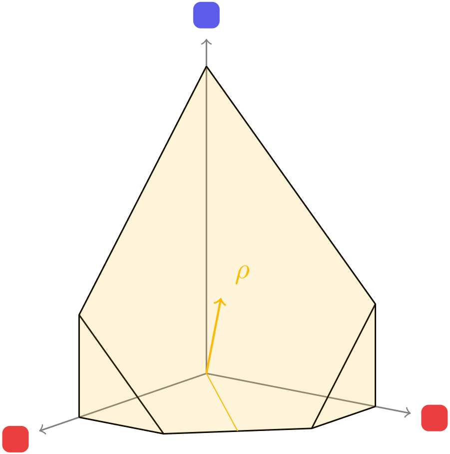




All classes **of the same type** are exchangeable

- Same resource occupation
- Same traffic intensity

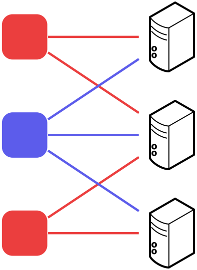


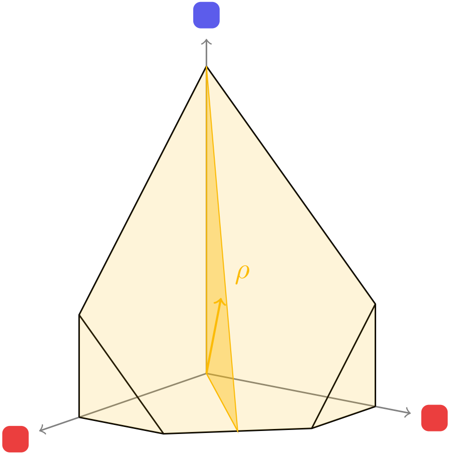


Poly-symmetry

All classes **of the same type**
are exchangeable

- Same resource occupation
- Same traffic intensity





Performance metrics under poly-symmetry

Poly-symmetry with regard to partition $\Sigma = (I_1, \dots, I_K)$
$n = (n_1, \dots, n_K)$ with $n_k = |I_k|$ for $k = 1, \dots, K$

- Law of total expectation

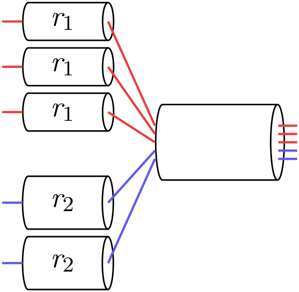
$$\mathbb{E}\left(\sum_{i \in I_k} \mathbf{X}_i\right) = \sum_{a \le n} \pi(a) \times \mathbb{E}\left(\sum_{i \in I_k} \mathbf{X}_i \middle| |I(\mathbf{X})|_\Sigma = a\right)$$

  with $\pi(a) = \mathbb{P}(|I(\mathbf{X})|_\Sigma = a)$

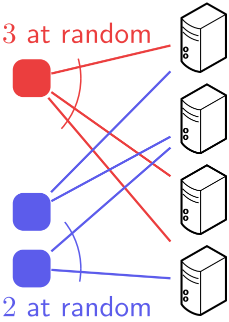
- $\pi(a)$ and $\mathbb{E}\left(\sum_{i \in I_k} \mathbf{X}_i \middle| |I(\mathbf{X})|_\Sigma = a\right)$ computed recursively

Poly-symmetry in the wild

- Access networks of Internet service providers



- Asymptotic poly-symmetry of clusters of servers with random static assignment



Conclusion

- **Processor-sharing systems with polymatroid capacity sets**
  - Model a large variety of real systems
  - Explicit recursion formulas under balanced fairness
- **Poly-symmetry**
  - Combine classes which have the same impact on the system
  - Tractable formulas to compute the performance metrics
- Future works
  - Relax the symmetry assumptions to allow for more flexibility

The paper is available on arXiv.

Bibliography

- T. Bonald and A. Proutière (2003). "Insensitive Bandwidth Sharing in Data Networks". In: *Queueing Syst.* 44.1, pp. 69-100.
- V. Shah and G. de Veciana (2015). "High-Performance Centralized Content Delivery Infrastructure: Models and Asymptotics". In: *IEEE/ACM Transactions on Networking* 23.5, pp. 1674-1687.
- V. Shah and G. de Veciana (2016). "Impact of fairness and heterogeneity on delays in large-scale centralized content delivery systems". In: *Queueing Systems*, pp. 1-37.