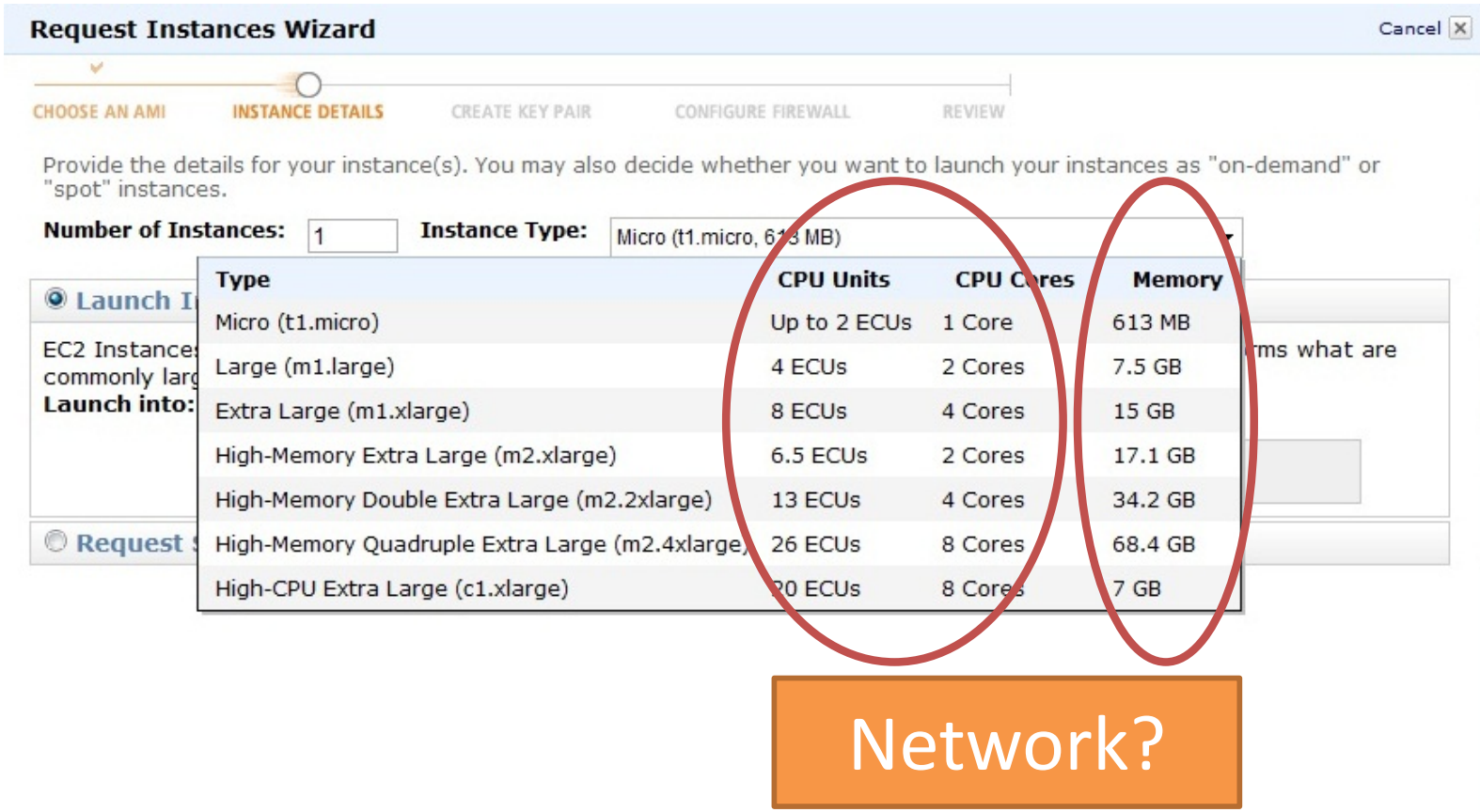
FairCloud

Slides by

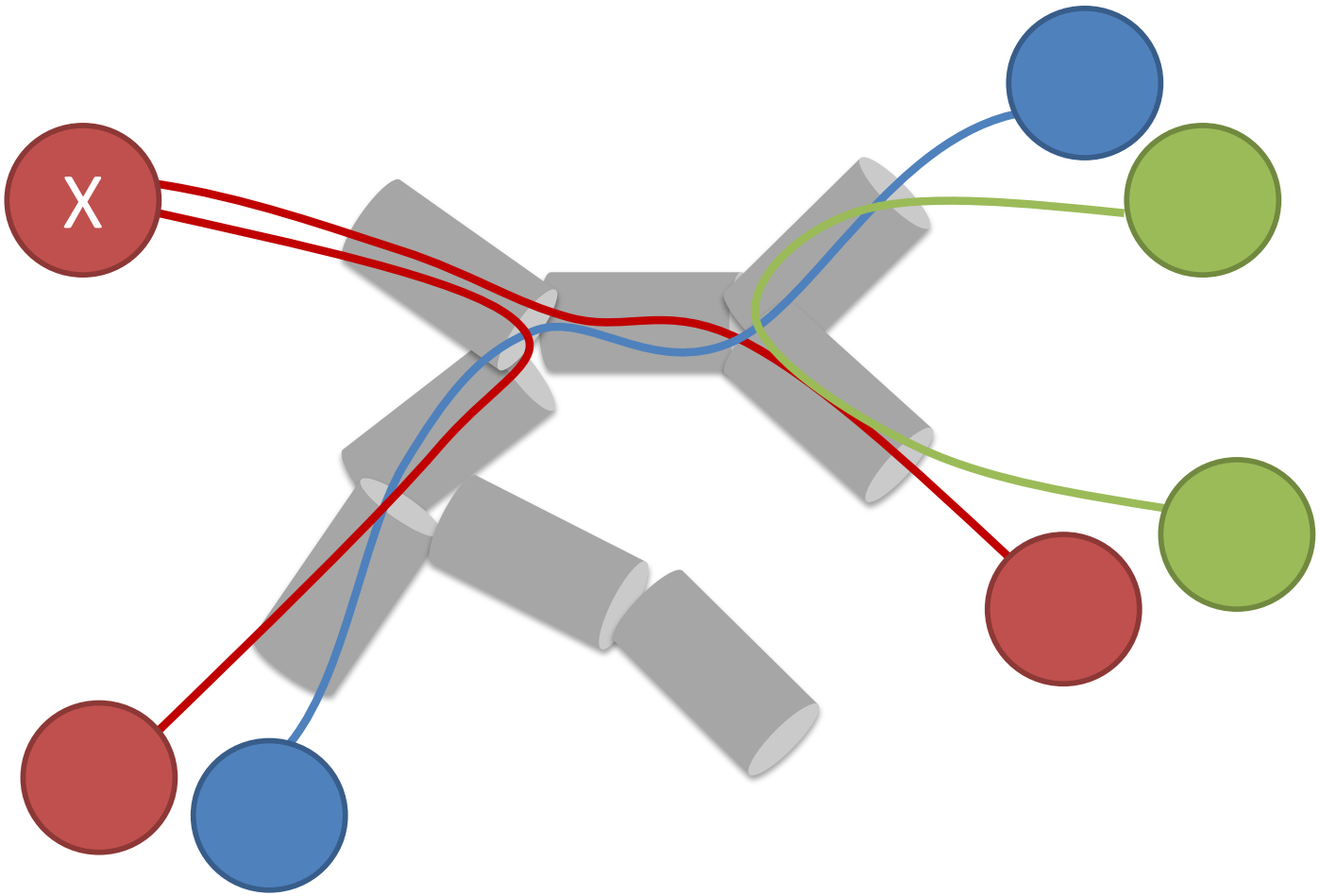
Original Authors

Motivation


Request Instances Wizard
Cancel
CHOOSE AN AMI
INSTANCE DETAILS
CREATE KEY PAIR
CONFIGURE FIREWALL
REVIEW
Provide the details for your instance(s). You may also decide whether you want to launch your instances as "on-demand" or "spot" instances.
Number of Instances:
1
Instance Type:
Micro (t1.micro, 613 MB)
Launch I
EC2 Instances
commonly larg
Launch into:
ms what are
Request
Type
CPU Units
CPU Cores
Memory
Micro (t1.micro)
Up to 2 ECUs
1 Core
613 MB
Large (m1.large)
4 ECUs
2 Cores
7.5 GB
Extra Large (m1.xlarge)
8 ECUs
4 Cores
15 GB
High-Memory Extra Large (m2.xlarge)
6.5 ECUs
2 Cores
17.1 GB
High-Memory Double Extra Large (m2.2xlarge)
13 ECUs
4 Cores
34.2 GB
High-Memory Quadruple Extra Large (m2.4xlarge)
26 ECUs
8 Cores
68.4 GB
High-CPU Extra Large (c1.xlarge)
20 ECUs
8 Cores
7 GB
Network?

Context

Networks are more difficult to share than other resources



Context

- Several proposals that share network differently, e.g.:
  - proportional to # source VMs (Seawall [NSDI11])
  - statically reserve bandwidth (Oktopus [Sigcomm12])
  - …
- Provide specific types of sharing policies
- Characterize solution space and relate policies to each other?

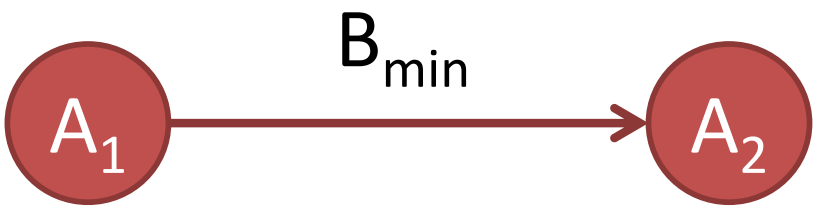
FairCloud Paper

1. Framework for understanding network sharing in cloud computing
   - Goals, tradeoffs, properties

2. Solutions for sharing the network
   - Existing policies in this framework
   - New policies representing different points in the design space

Goals

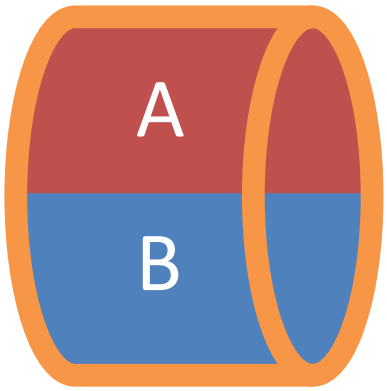
1. **Minimum Bandwidth Guarantees**
   - Provides predictable performance
   - Example: file transfer finishes within time limit



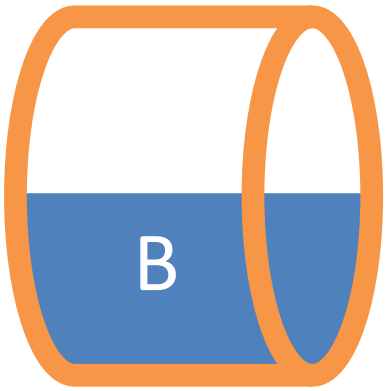
$$Time_{max} = Size / B_{min}$$

Goals

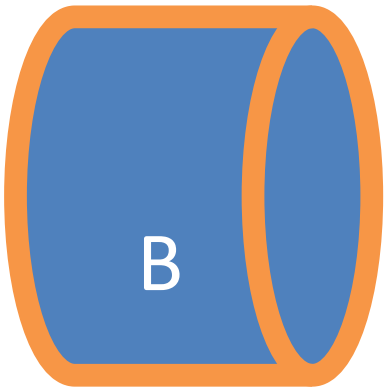
1. **Minimum Bandwidth Guarantees**
2. **High Utilization**
   - Do not leave useful resources unutilized
   - Requires both work-conservation and proper incentives



Both tenants active



Non work-conserving



Work-conserving

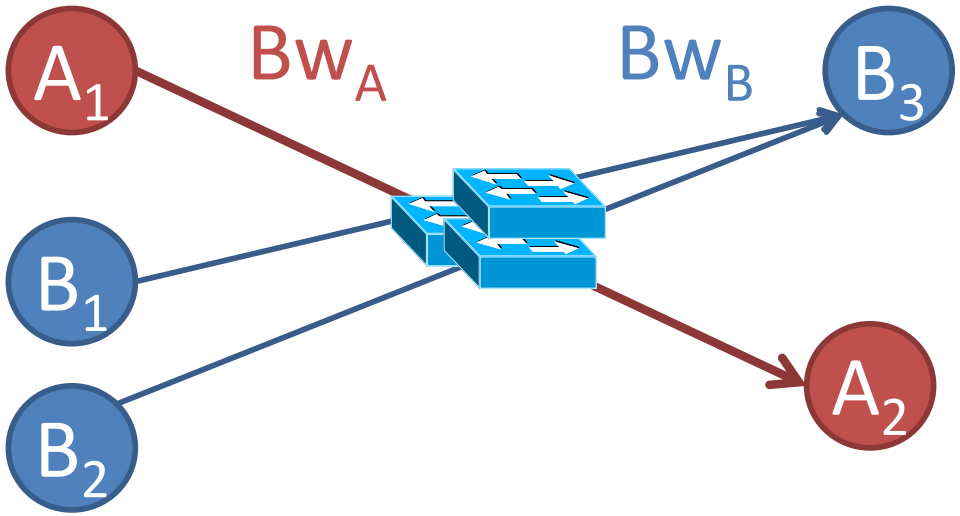
Goals

1. **Minimum Bandwidth Guarantees**
2. **High Utilization**
3. **Network Proportionality**
   - As with other services, network should be shared proportional to payment
   - Currently, tenants pay a flat rate per VM ➔ network share should be proportional to #VMs

     (assuming identical VMs)

Goals

1. **Minimum Bandwidth Guarantees**
2. **High Utilization**
3. **Network Proportionality**
   - Example: A has 2 VMs, B has 3 VMs



$$\frac{Bw_A}{Bw_B} = \frac{2}{3}$$

When exact sharing is not possible use max-min

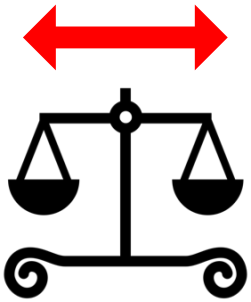
Goals

1. **Minimum Bandwidth Guarantees**
2. **High Utilization**
3. **Network Proportionality**

Not all goals are achievable simultaneously!

Tradeoffs

Min Guarantee



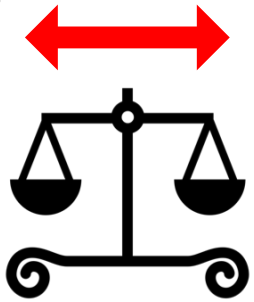
Network Proportionality

High Utilization

Not all goals are achievable simultaneously!

Tradeoffs

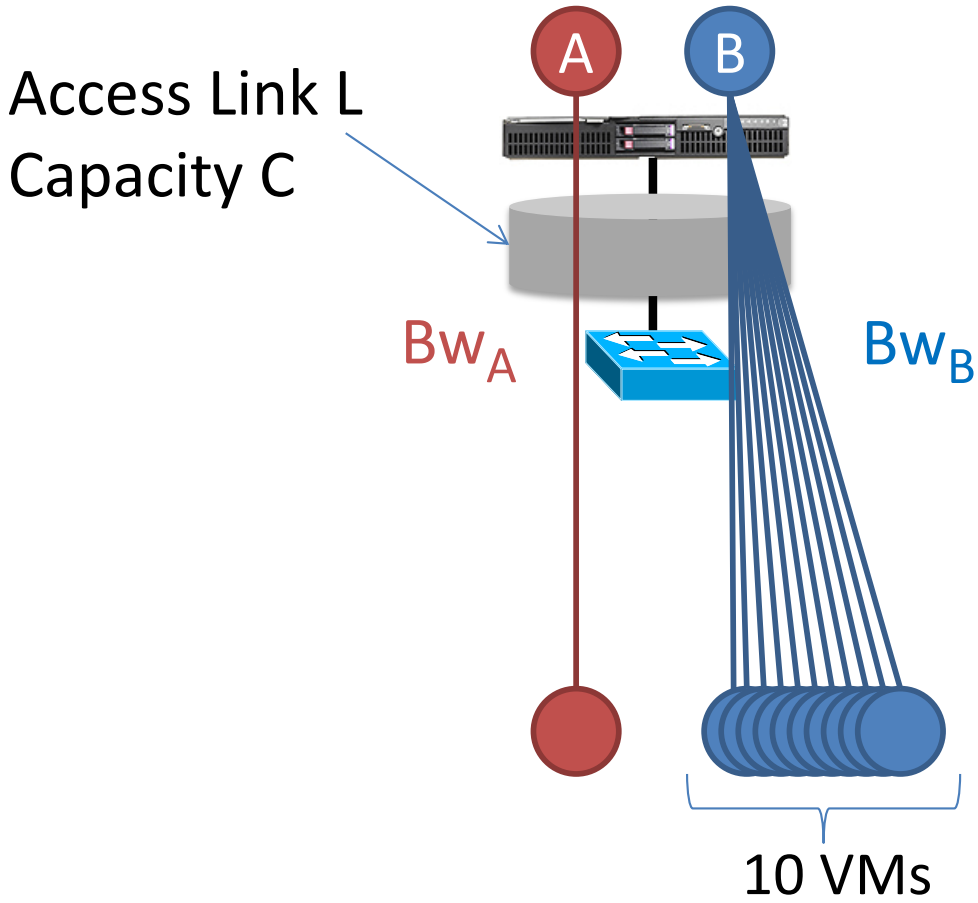
Min Guarantee



Network
Proportionality

Tradeoffs

Min Guarantee ⟷ Network Proportionality



Minimum Guarantee

$Bw_A = 1/2\ C$

$Bw_B = 1/2\ C$

Network Proportionality

$Bw_A = 2/13\ C$

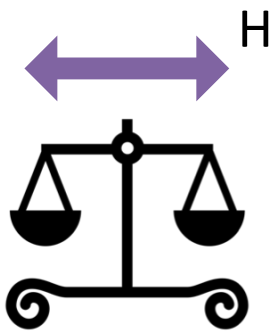
$Bw_B = 11/13\ C$

$Bw_A \approx C/N_T \rightarrow 0$

#VMs in the network

Tradeoffs

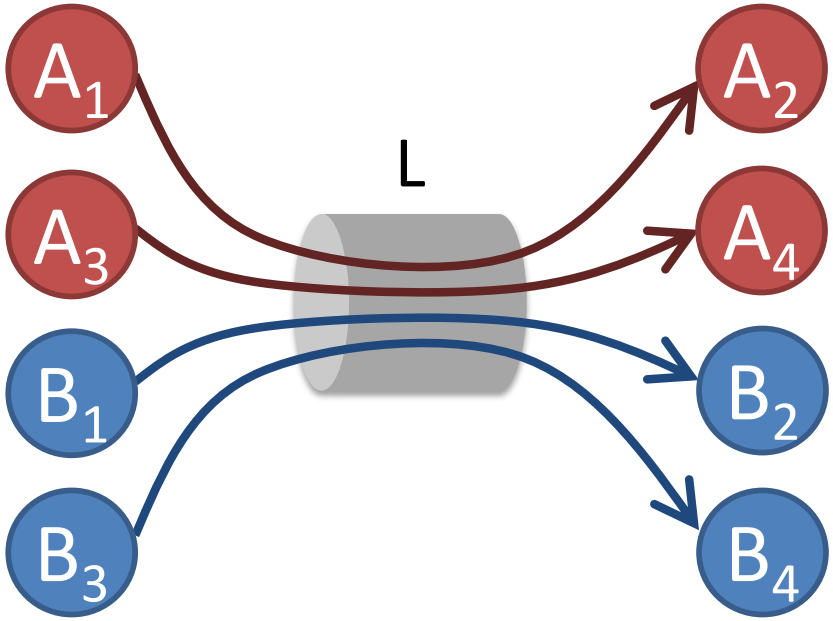
Network Proportionality



High Utilization

Tradeoffs

Network Proportionality ⟷ High Utilization



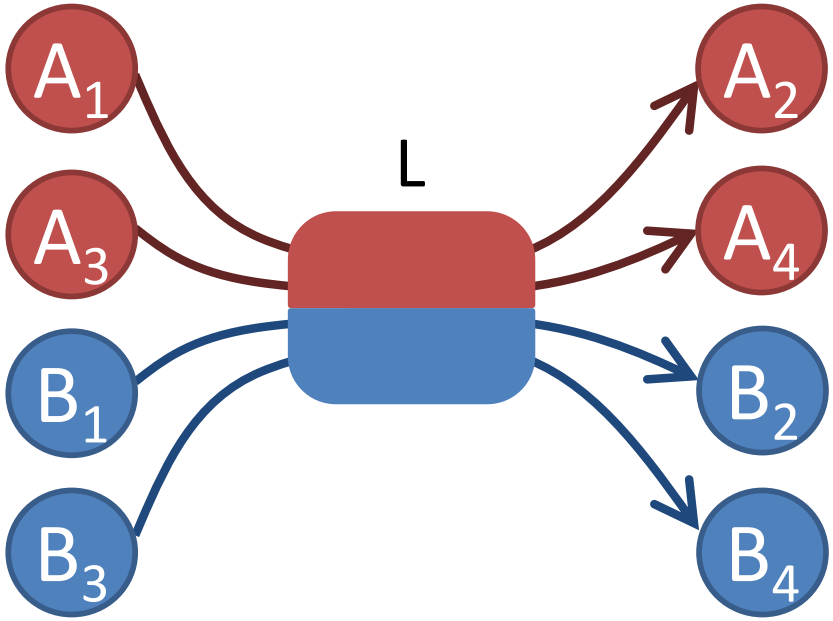
Tradeoffs

Network Proportionality ⟷ High Utilization

Network Proportionality

$Bw_A = 1/2\ C$

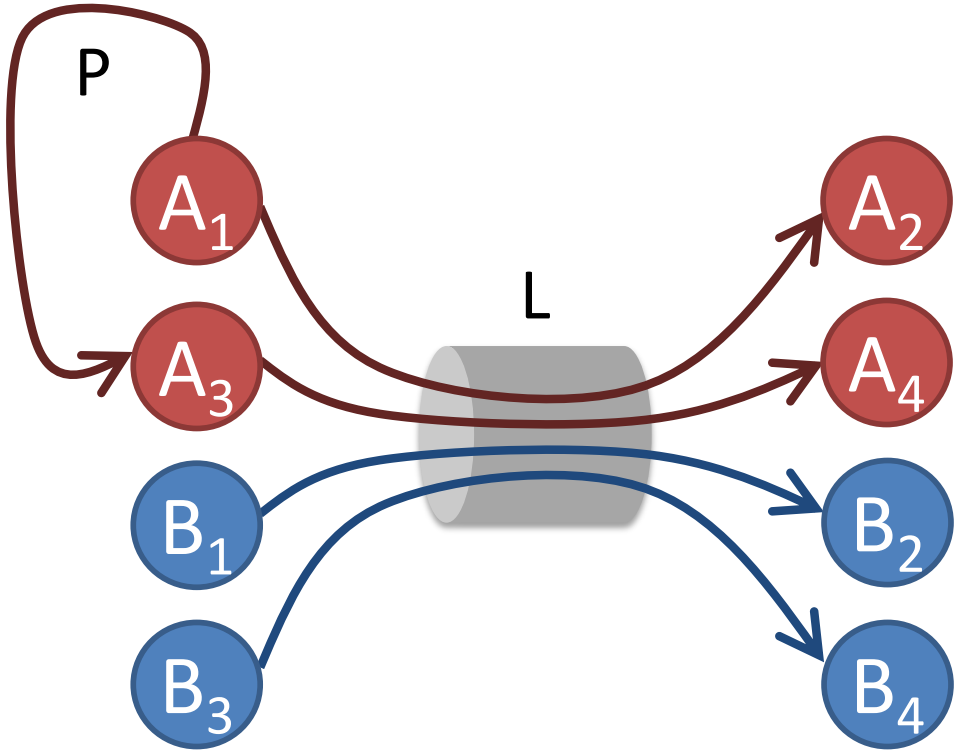
$Bw_B = 1/2\ C$



Tradeoffs

Network Proportionality ⟷ High Utilization

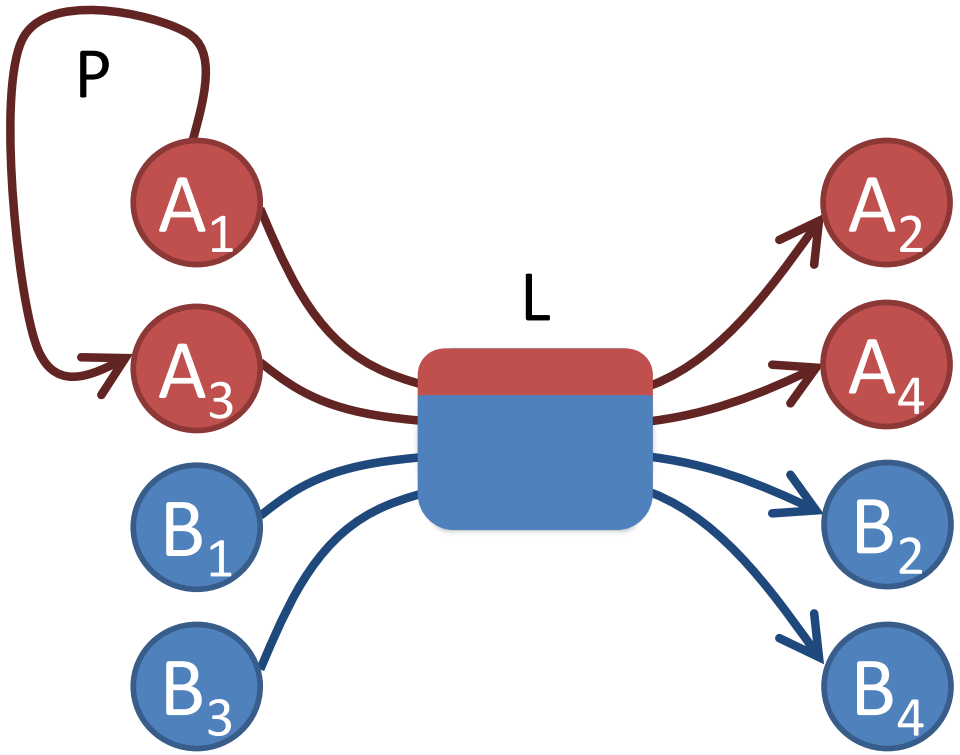
Uncongested path



Tradeoffs

Network Proportionality ⟷ High Utilization

Uncongested path



Network Proportionality

$$Bw_A^P + Bw_A^L = Bw_B^L$$

$$Bw_A^L < Bw_B^L$$

Tenants can be disincentivized to use free resources

If A values $A_1 \rightarrow A_2$ or $A_3 \rightarrow A_4$ more than $A_1 \rightarrow A_3$

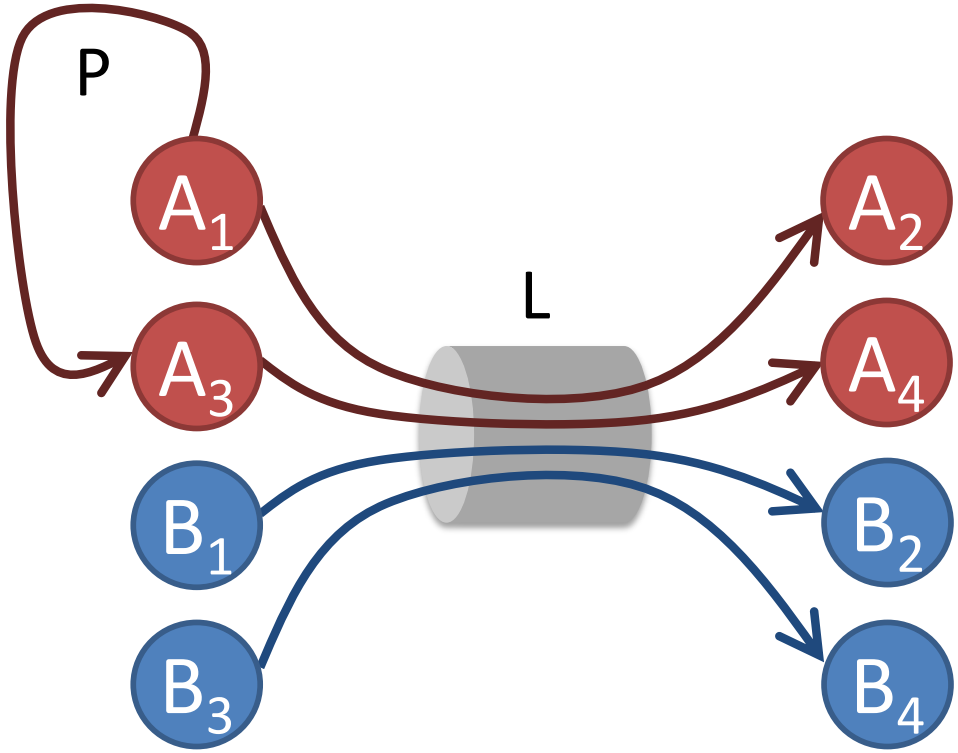
Tradeoffs

Network Proportionality

High Utilization

Congestion Proportionality

Uncongested path



Network proportionality applied only for flows traversing *congested* links *shared* by multiple tenants

Tradeoffs

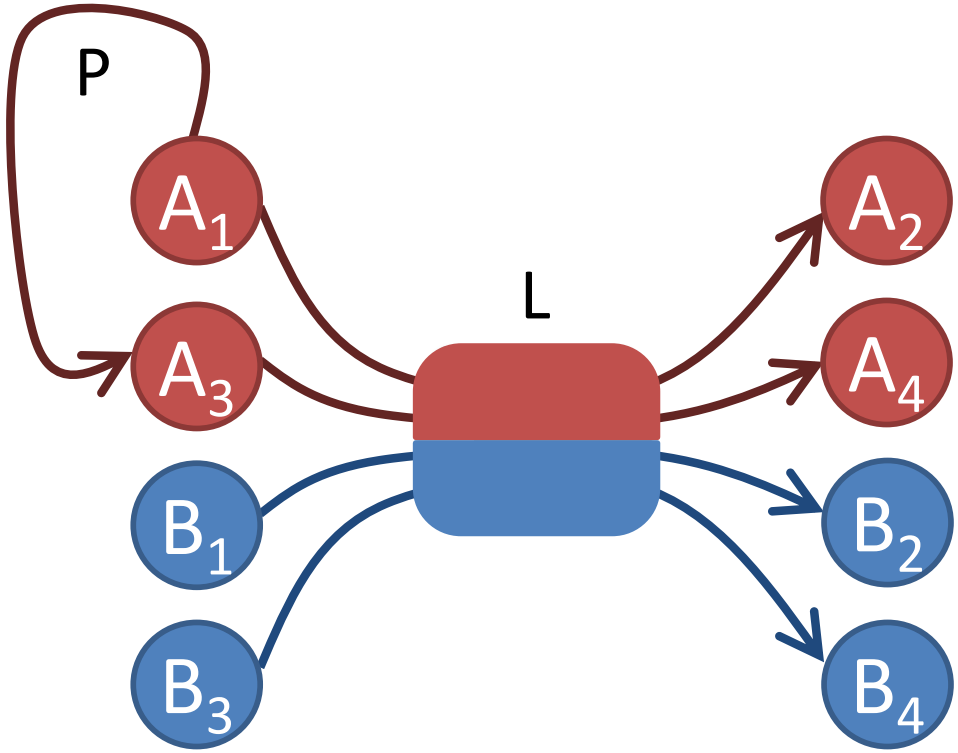
Network Proportionality

⋮

Congestion Proportionality

High Utilization

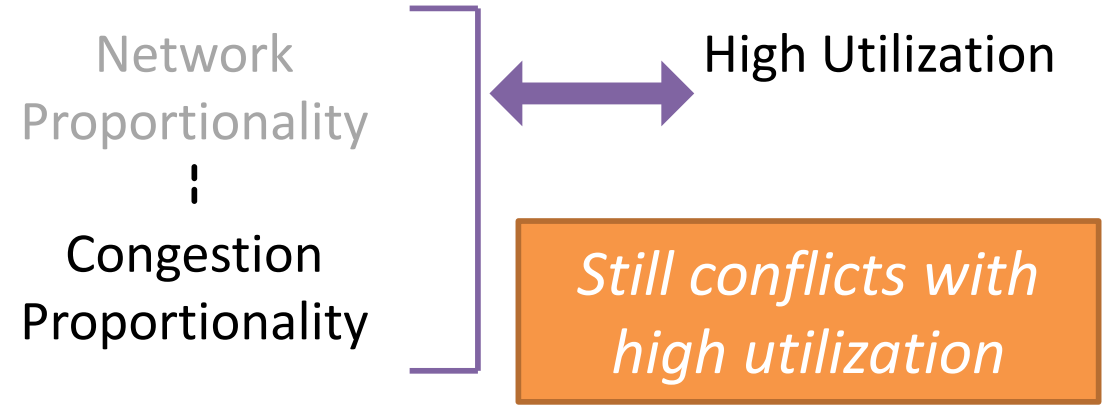
Uncongested path



Congestion Proportionality

$$Bw_A^L = Bw_B^L$$

Tradeoffs

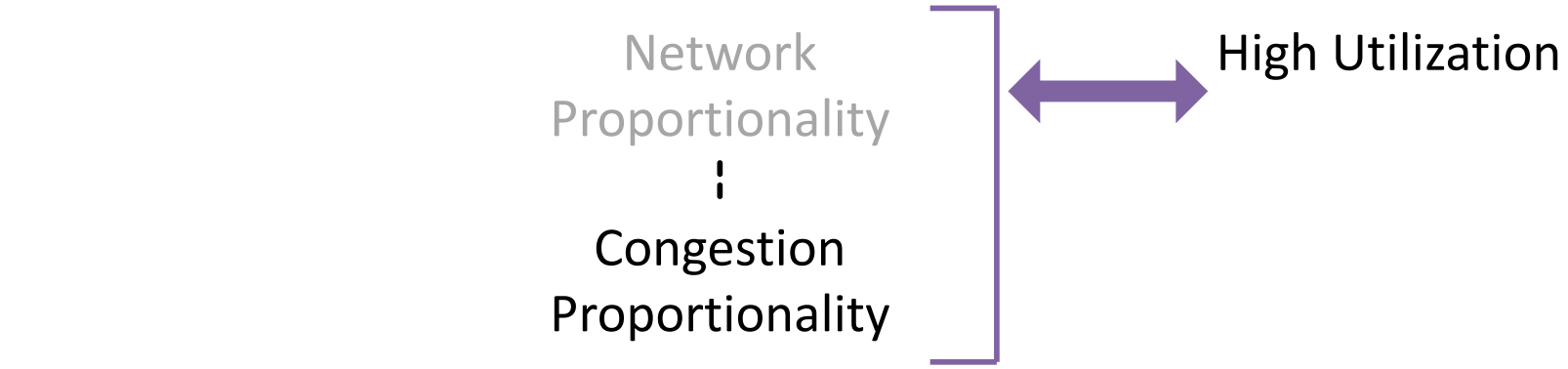

Network Proportionality
Congestion Proportionality
High Utilization
Still conflicts with high utilization

Tradeoffs

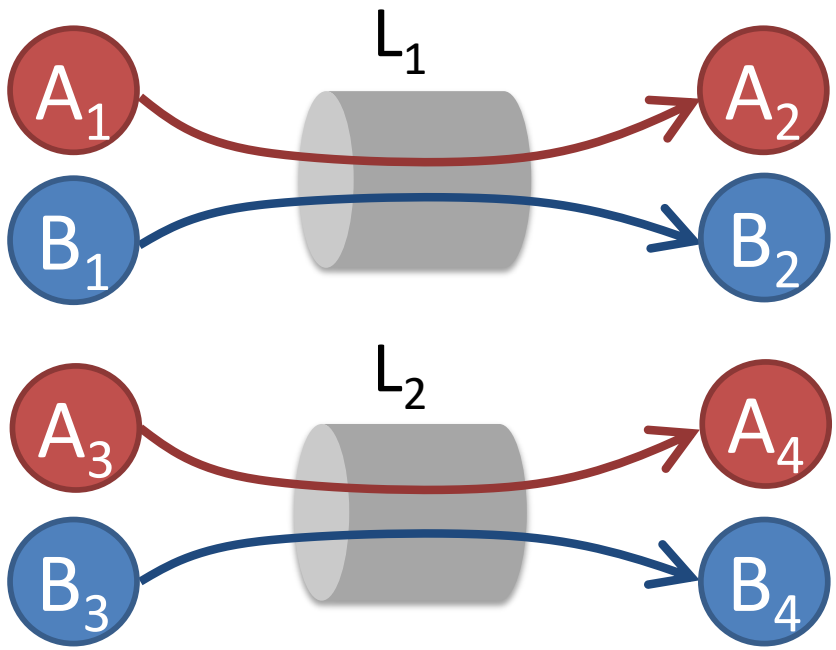
Network Proportionality

⋮

Congestion Proportionality



$C_1 = C_2 = C$

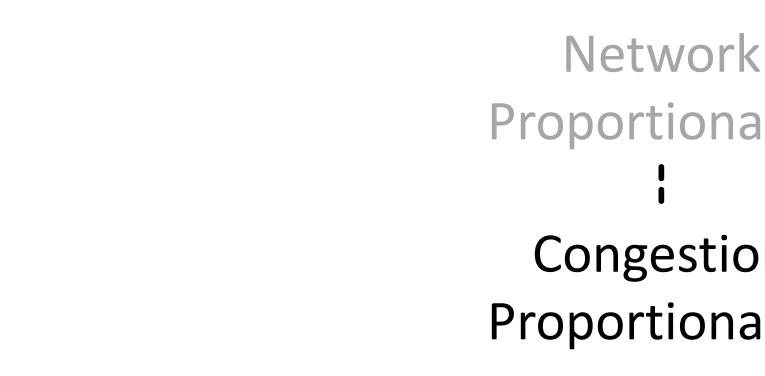


Tradeoffs

Network Proportionality

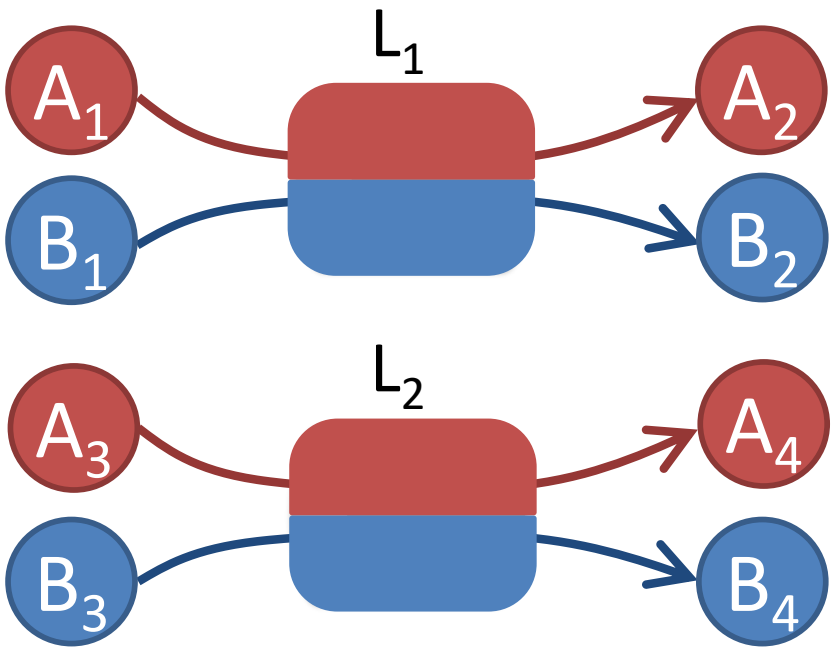
⋮

Congestion Proportionality



High Utilization

$C_1 = C_2 = C$



Congestion Proportionality

$$Bw_A^{L1} = Bw_B^{L1}$$

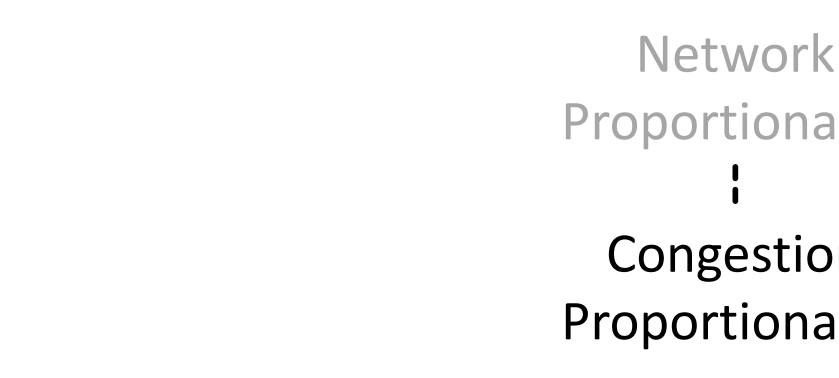
$$Bw_A^{L2} = Bw_B^{L2}$$

Tradeoffs

Network Proportionality

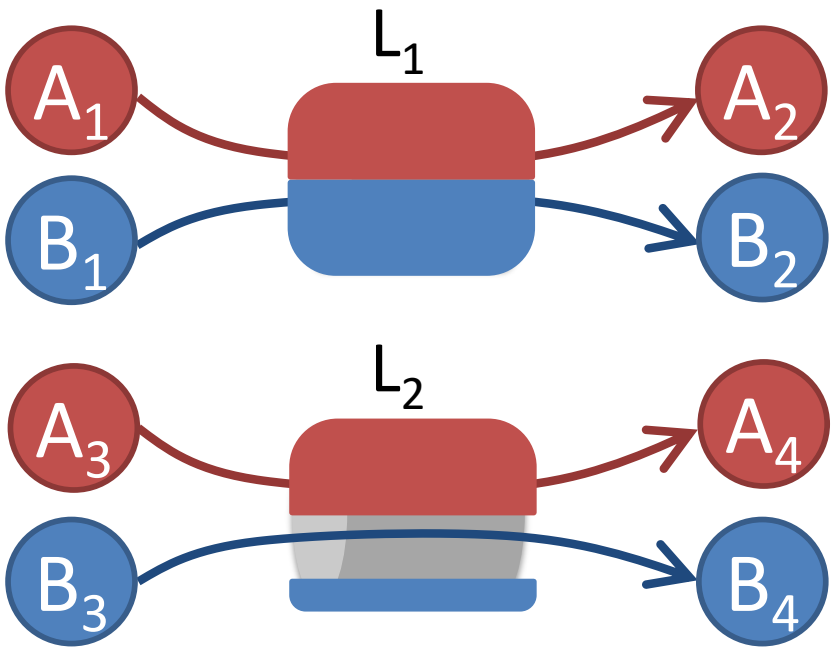
⋮

Congestion Proportionality



High Utilization

$C_1 = C_2 = C$



Demand drops to ε

Tradeoffs

Network Proportionality

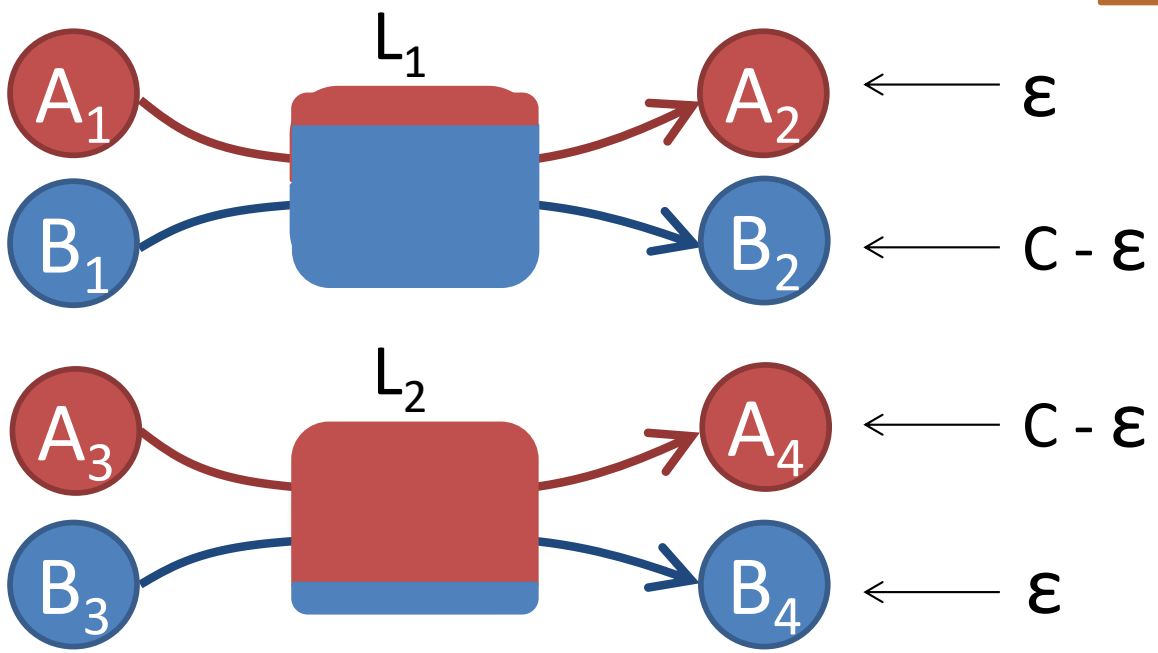
⋮

Congestion Proportionality

High Utilization

Tenants incentivized to not fully utilize resources

$C_1 = C_2 = C$



Tradeoffs

Network Proportionality

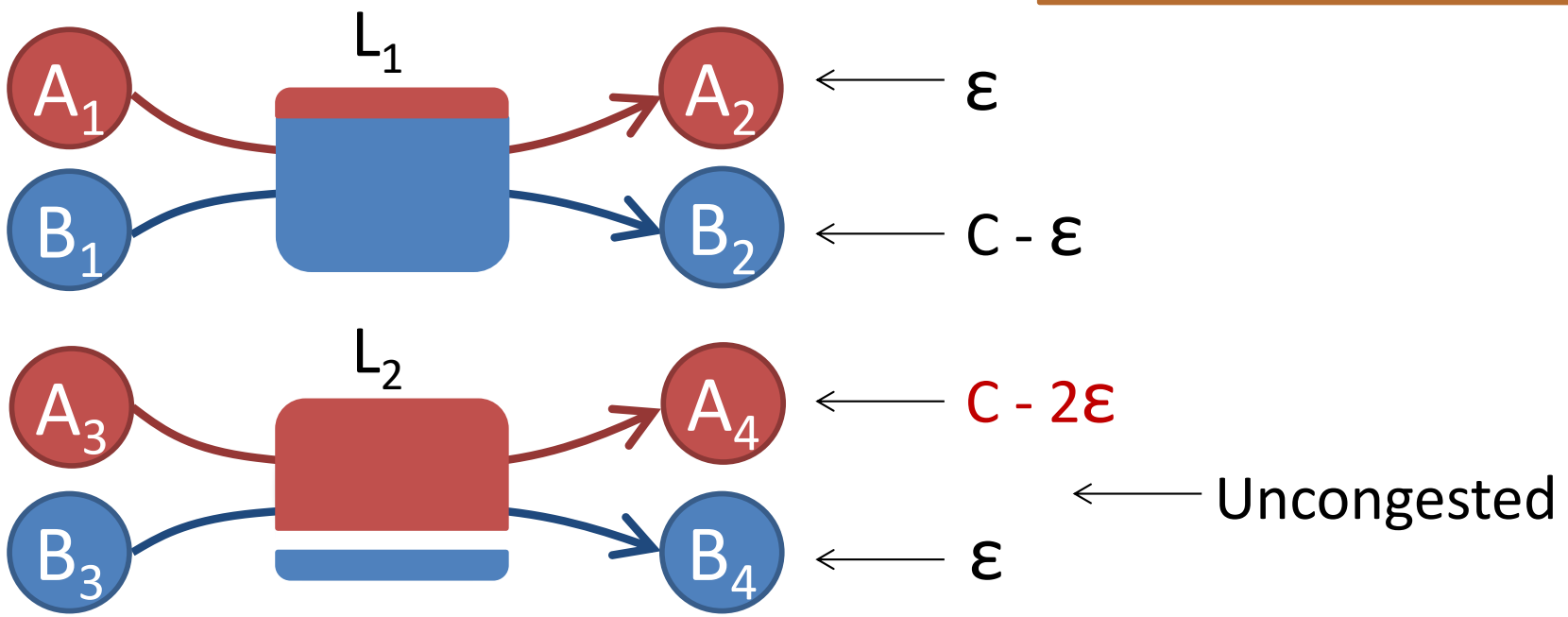
⋮

Congestion Proportionality

High Utilization

Tenants incentivized to not fully utilize resources

$C_1 = C_2 = C$



Tradeoffs

Network Proportionality

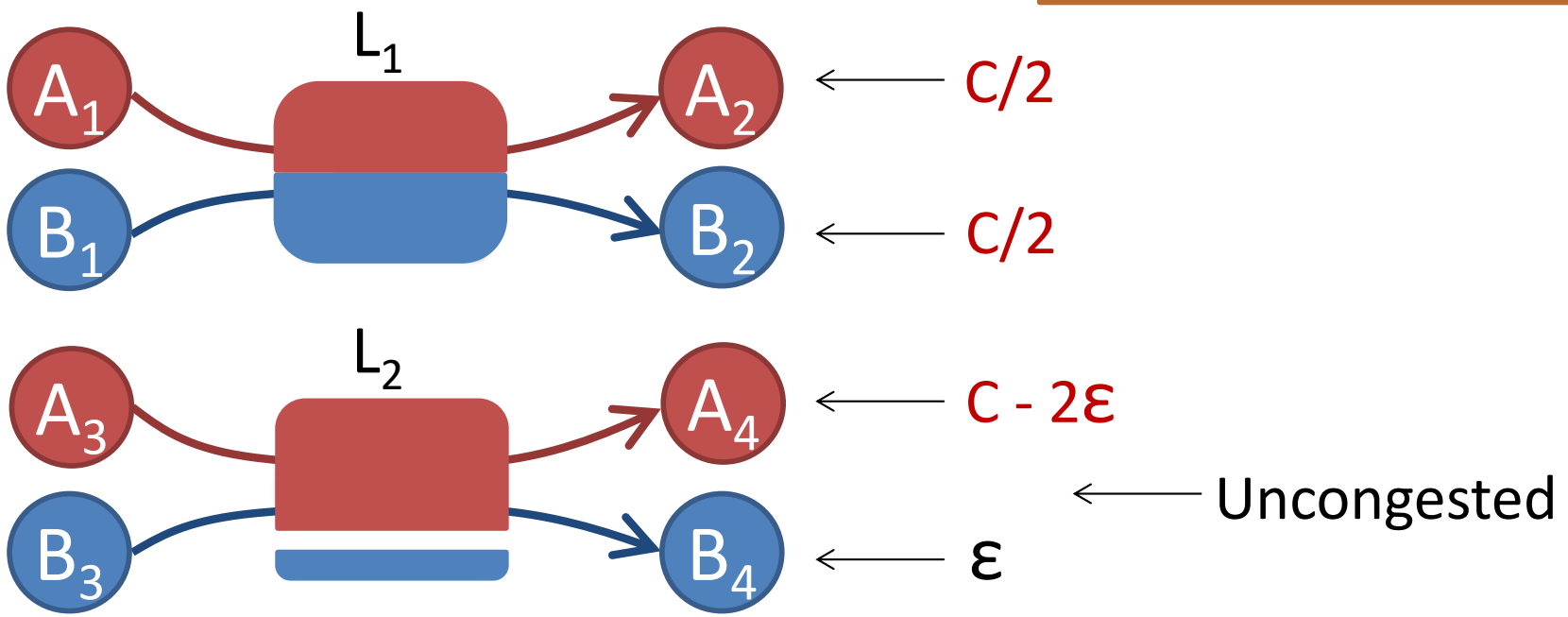
⋮

Congestion Proportionality

High Utilization

Tenants incentivized to not fully utilize resources

$C_1 = C_2 = C$



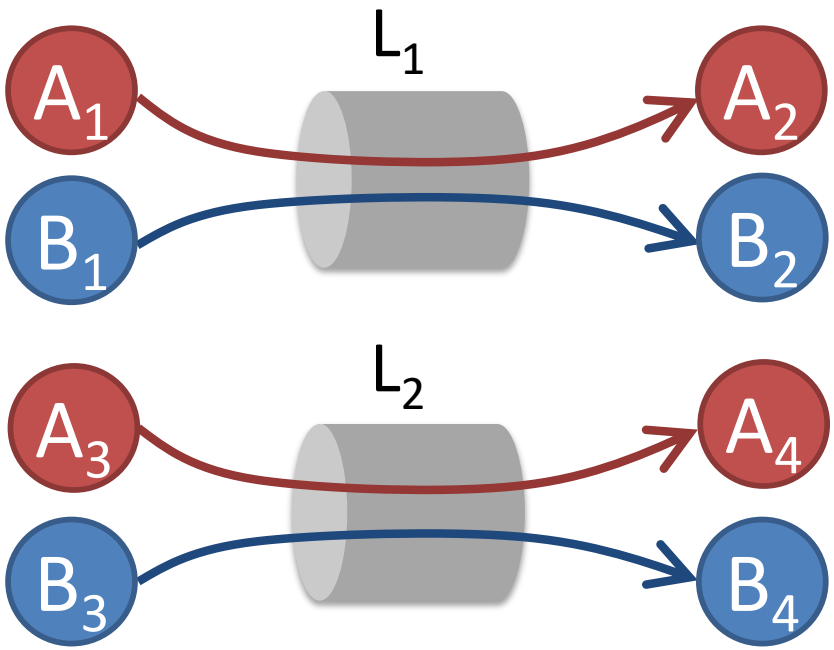
Tradeoffs

Network Proportionality

Congestion Proportionality

Link Proportionality

High Utilization



Proportionality applied to each link *independently*

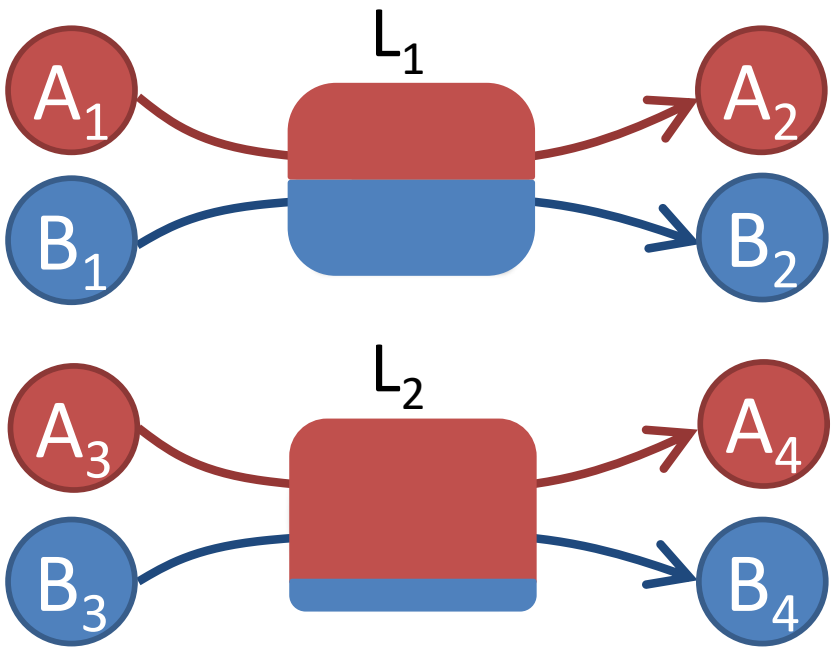
Tradeoffs

Network Proportionality

Congestion Proportionality

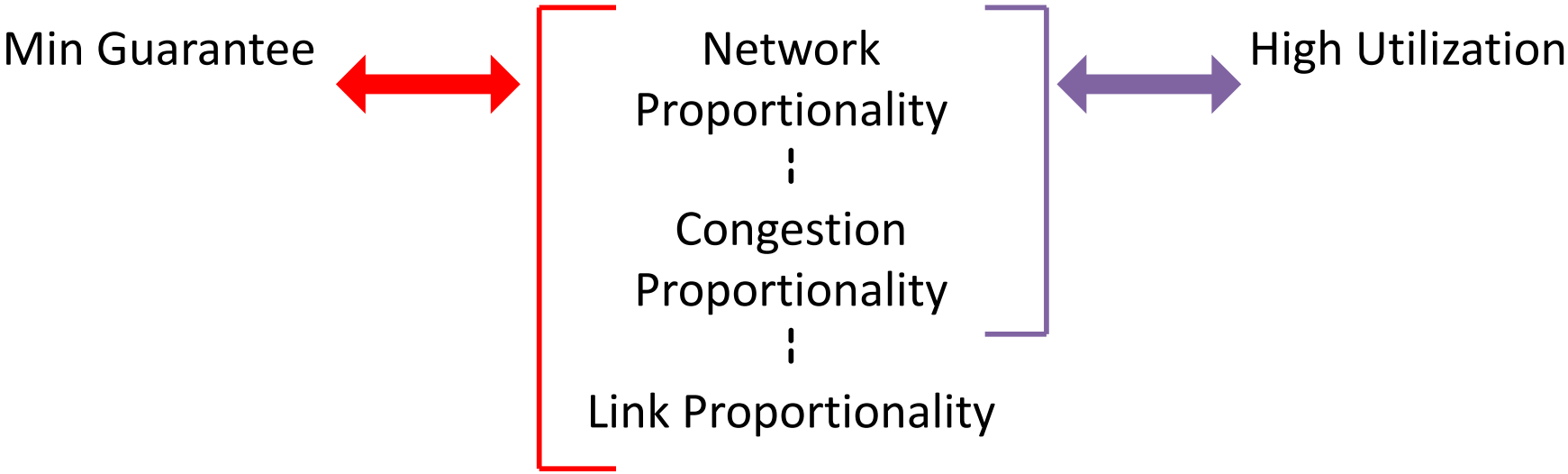
Link Proportionality

High Utilization

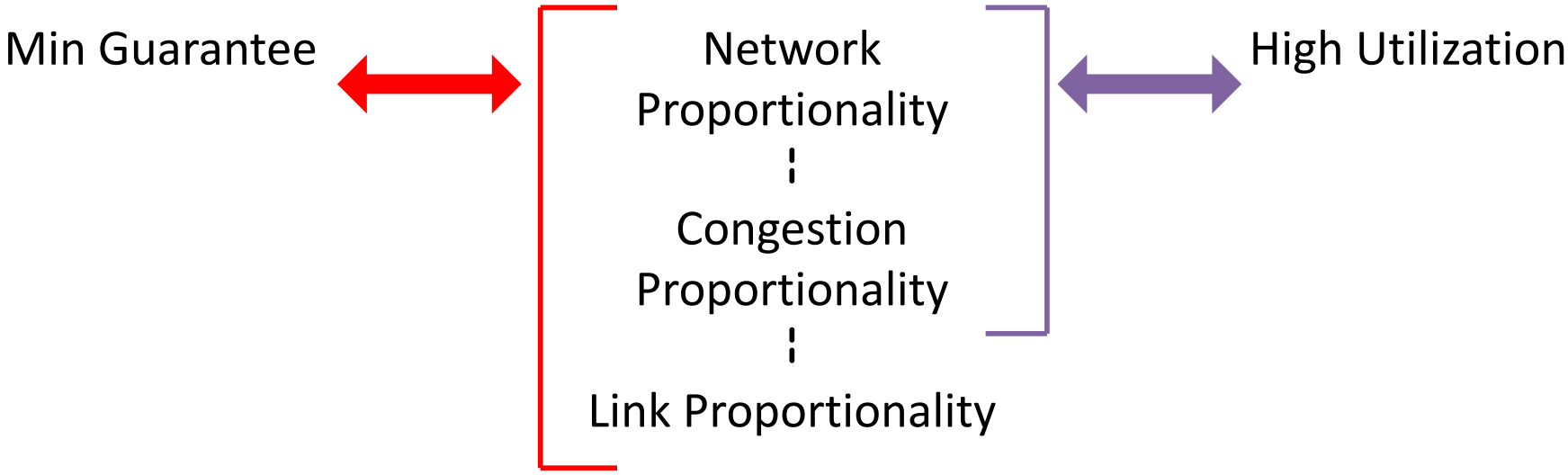


Full incentives for high utilization

Goals and Tradeoffs

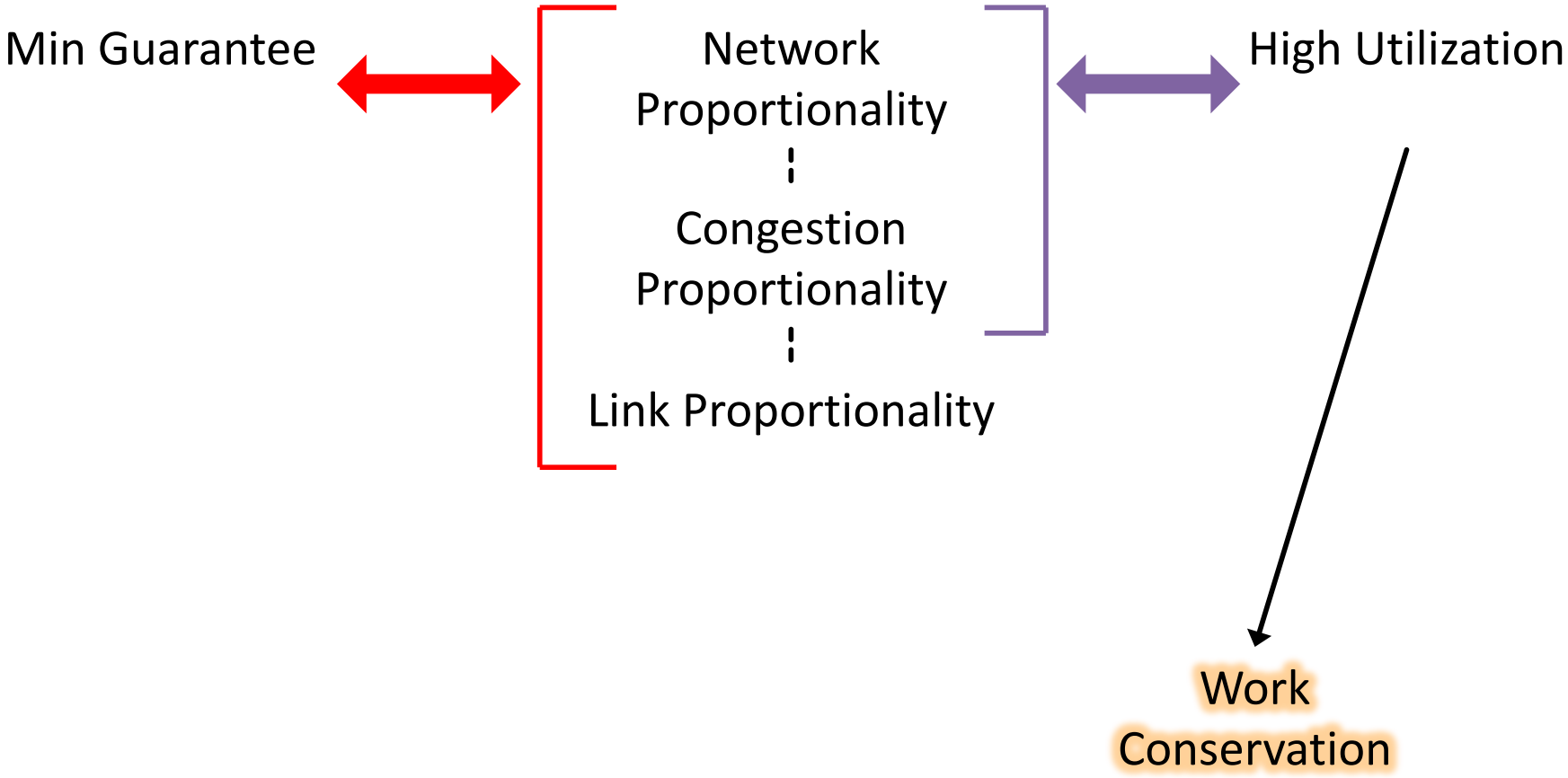


Guiding Properties



Break down goals into lower-level necessary properties

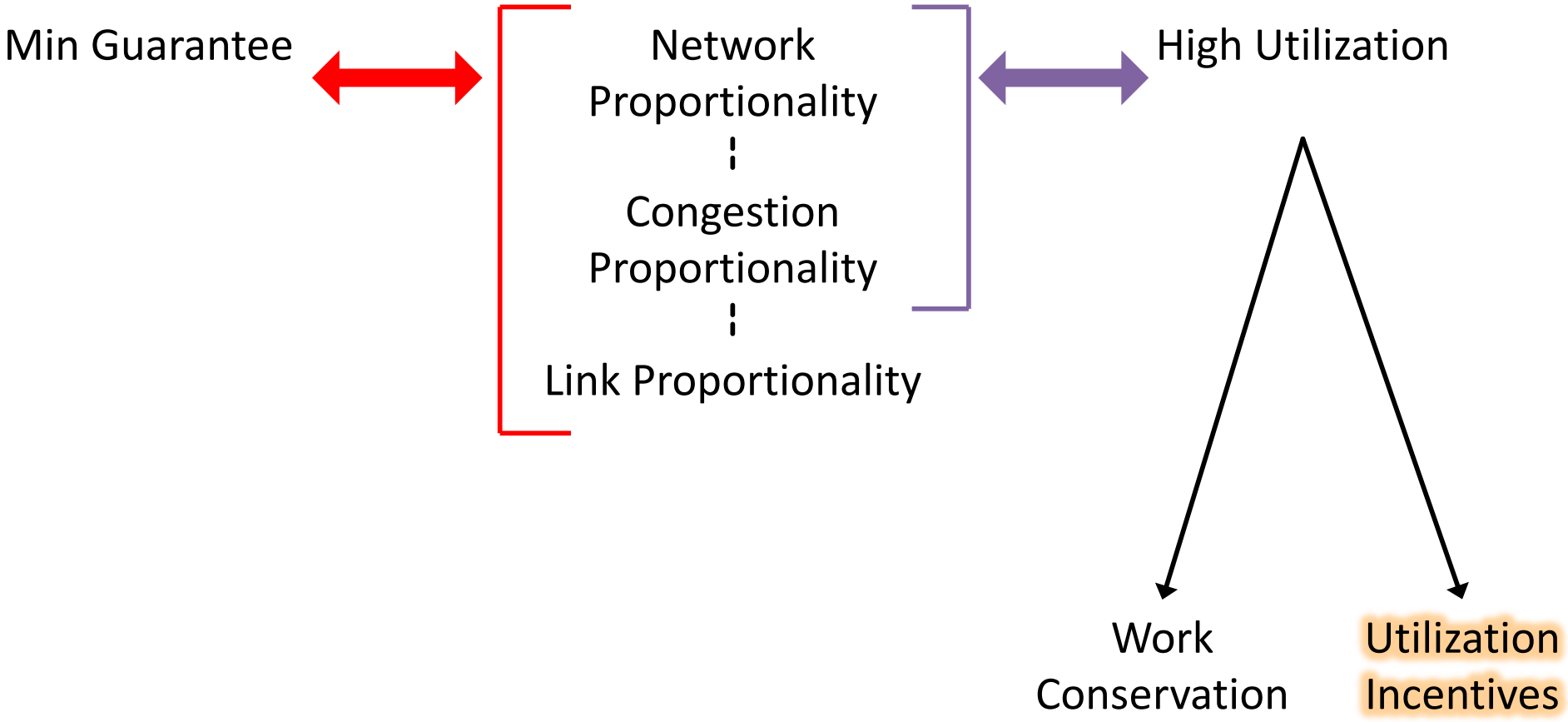
Properties


Min Guarantee
Network Proportionality
Congestion Proportionality
Link Proportionality
High Utilization
Work Conservation

Work Conservation

- Bottleneck links are fully utilized
- Static reservations do not have this property

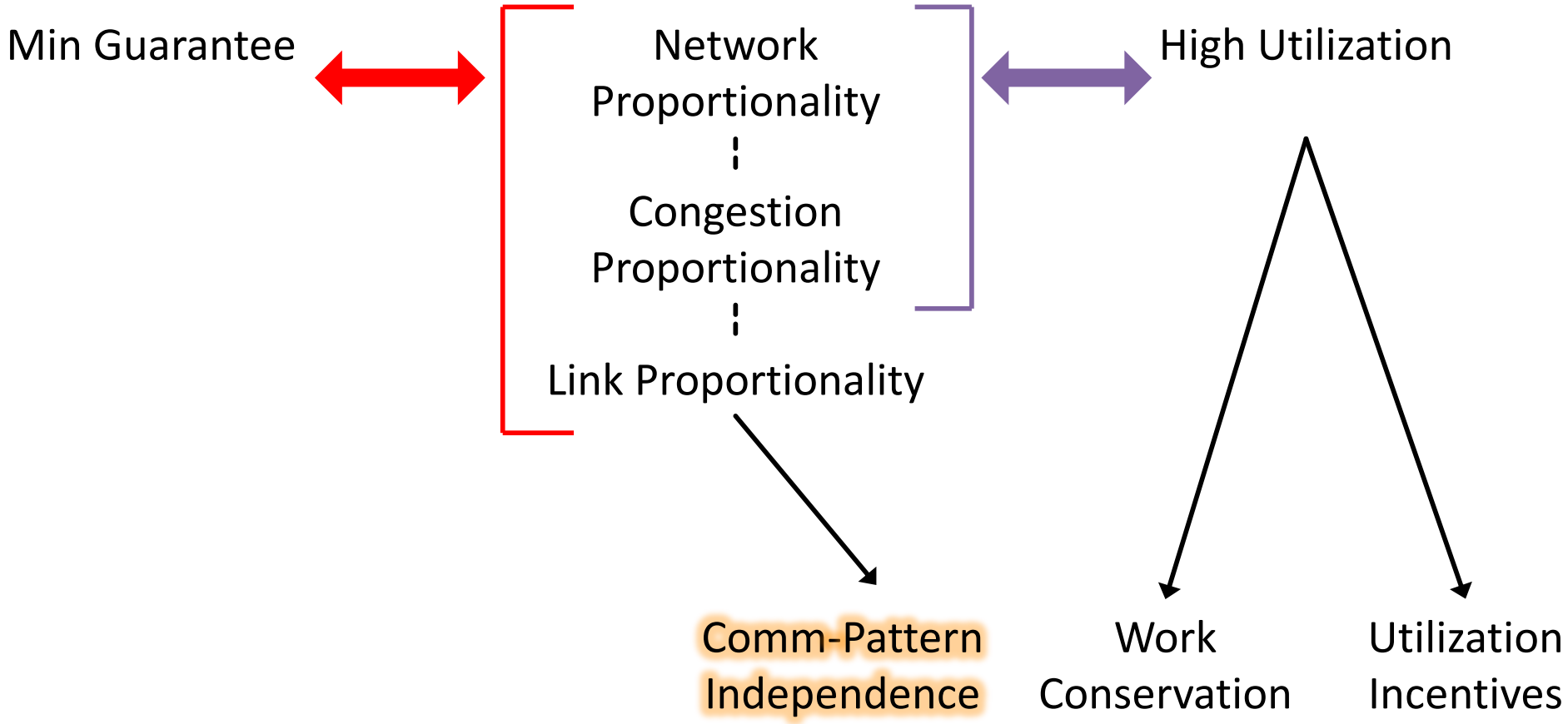
Properties


Min Guarantee
Network Proportionality
Congestion Proportionality
Link Proportionality
High Utilization
Work Conservation
Utilization Incentives

Utilization Incentives

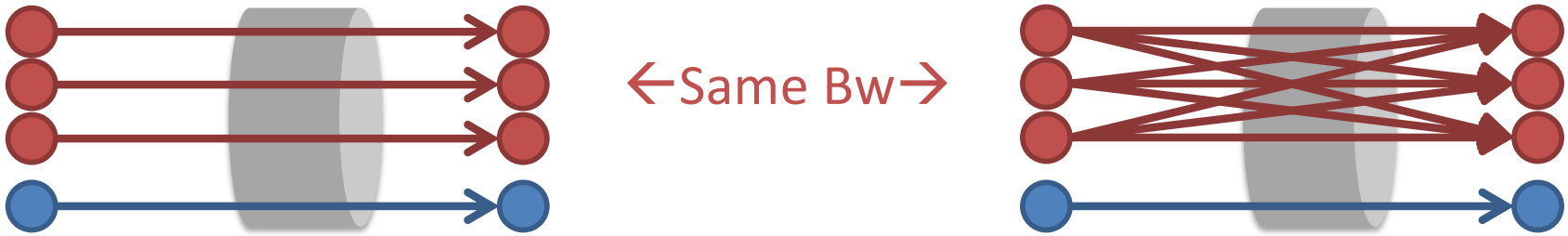
- Tenants are not incentivized to lie about demand to *leave links underutilized*
- Network and congestion proportionality do not have this property
- Allocating links independently provides this property

Properties

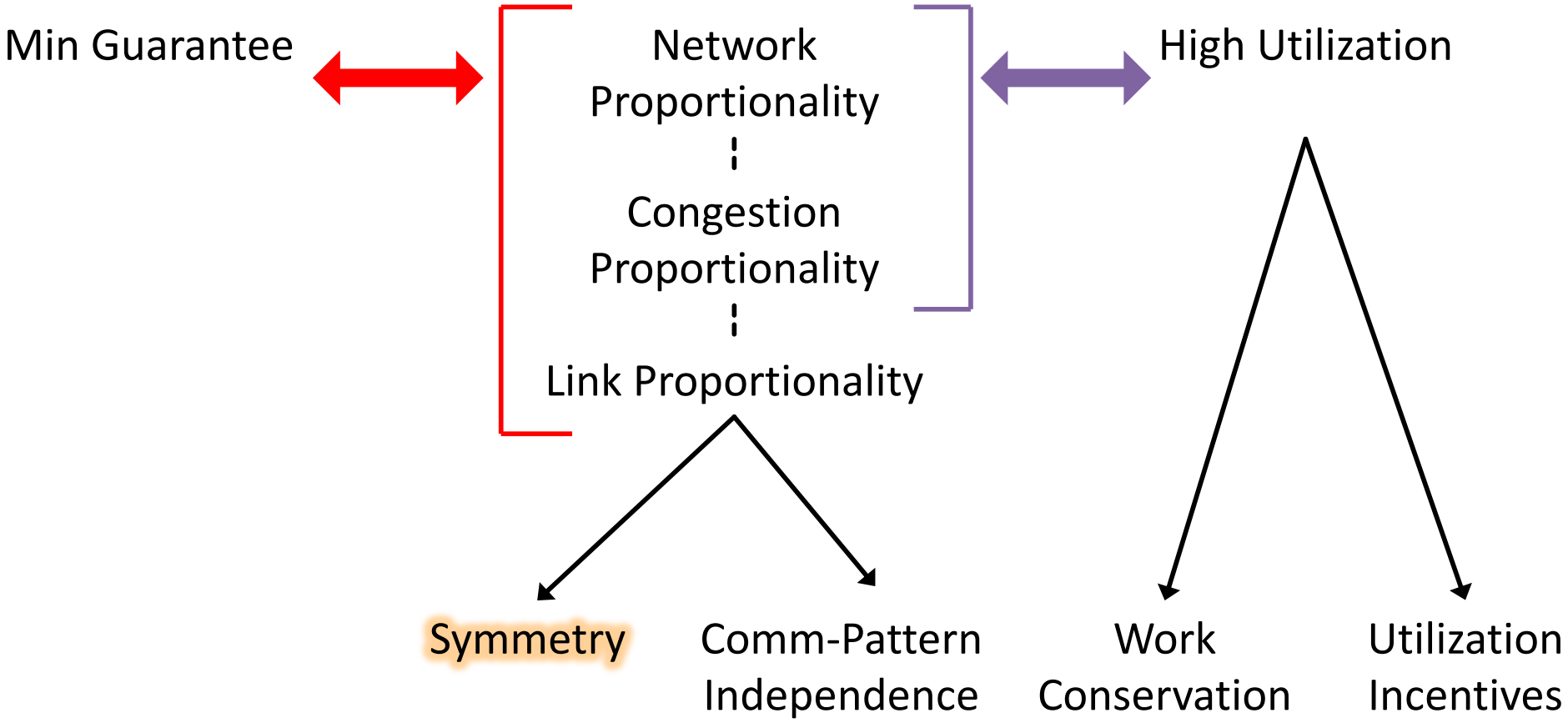

Min Guarantee
Network Proportionality
Congestion Proportionality
Link Proportionality
High Utilization
Comm-Pattern Independence
Work Conservation
Utilization Incentives

Communication-pattern Independence

- Allocation does not depend on communication pattern
- Per flow allocation does not have this property
  - (per flow = give equal shares to each flow)



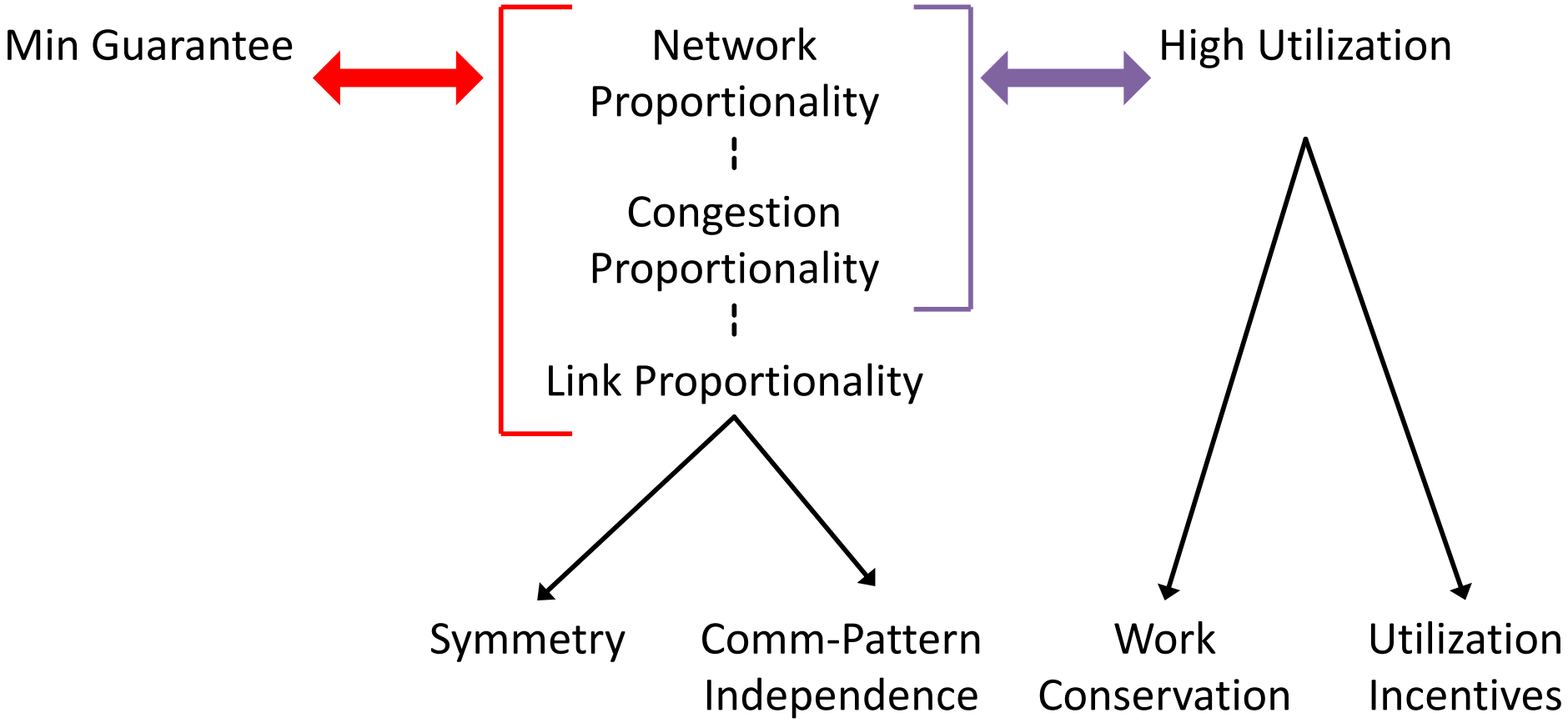
Properties


Min Guarantee
Network Proportionality
Congestion Proportionality
Link Proportionality
High Utilization
Symmetry
Comm-Pattern Independence
Work Conservation
Utilization Incentives

Symmetry

- Swapping demand directions preserves allocation
- Per source allocation lacks this property
  - (per source = give equal shares to each source)

Goals, Tradeoffs, Properties

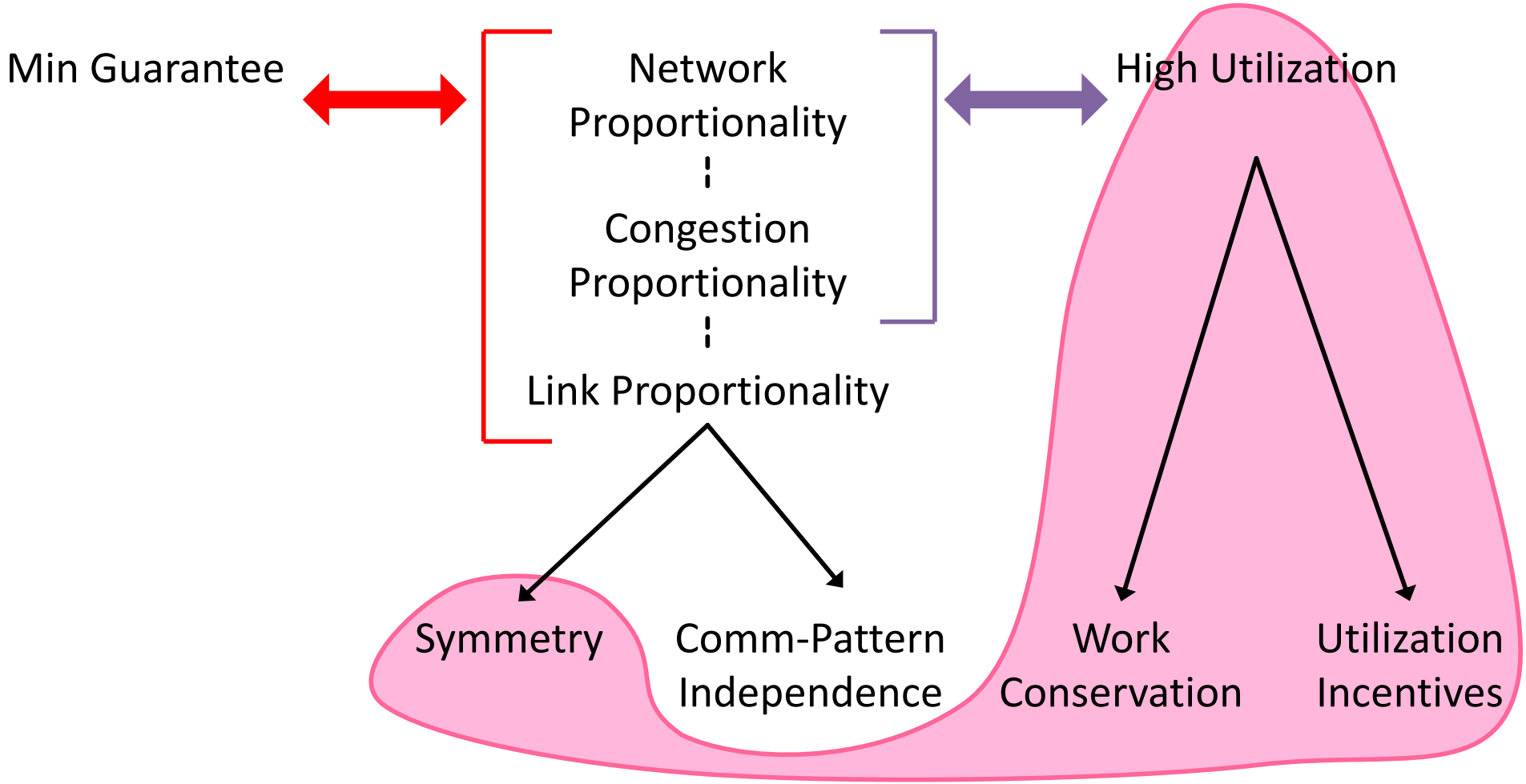


Outline

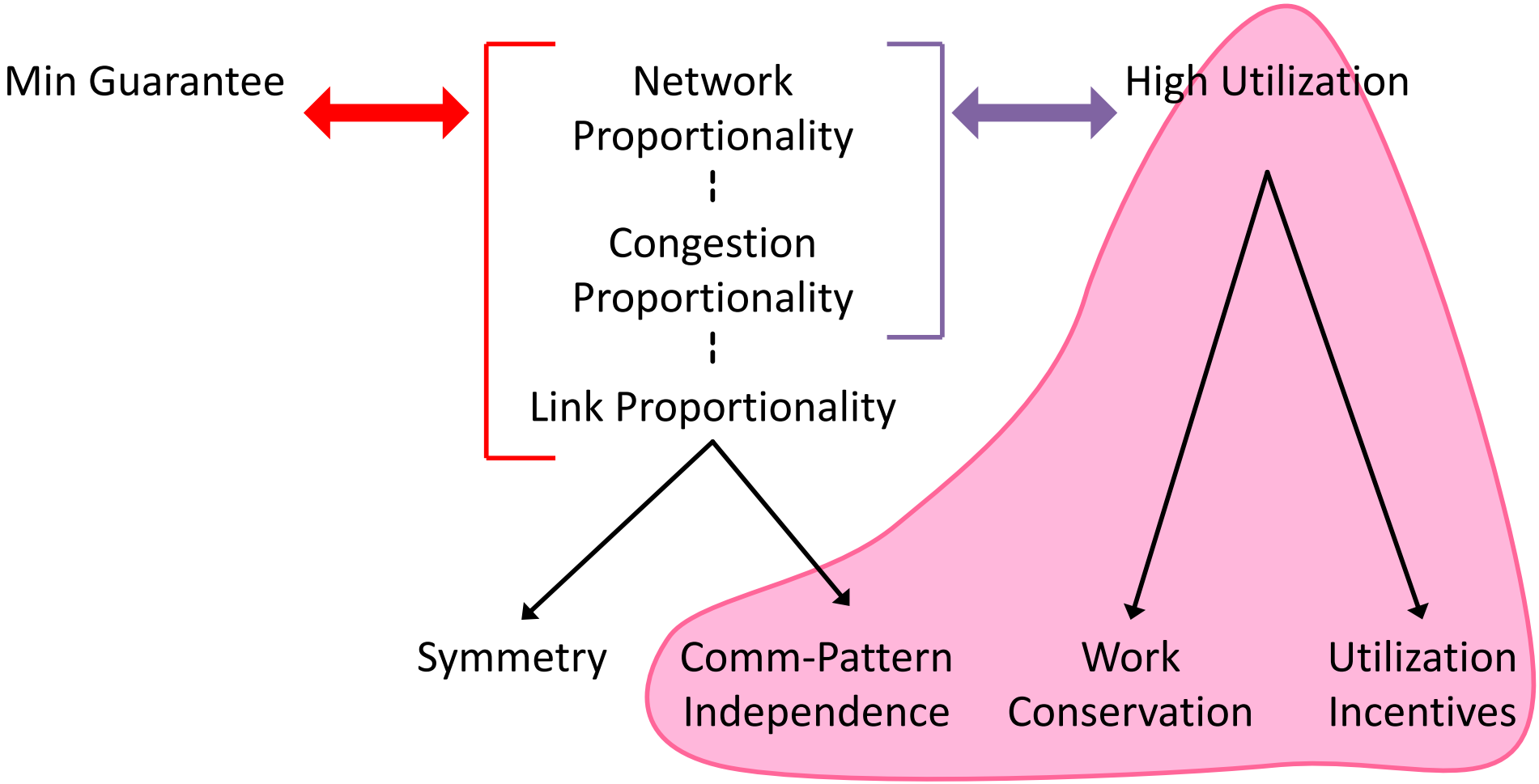
1. Framework for understanding network sharing in cloud computing
   - Goals, tradeoffs, properties

2. Solutions for sharing the network
   - Existing policies in this framework
   - New policies representing different points in the design space

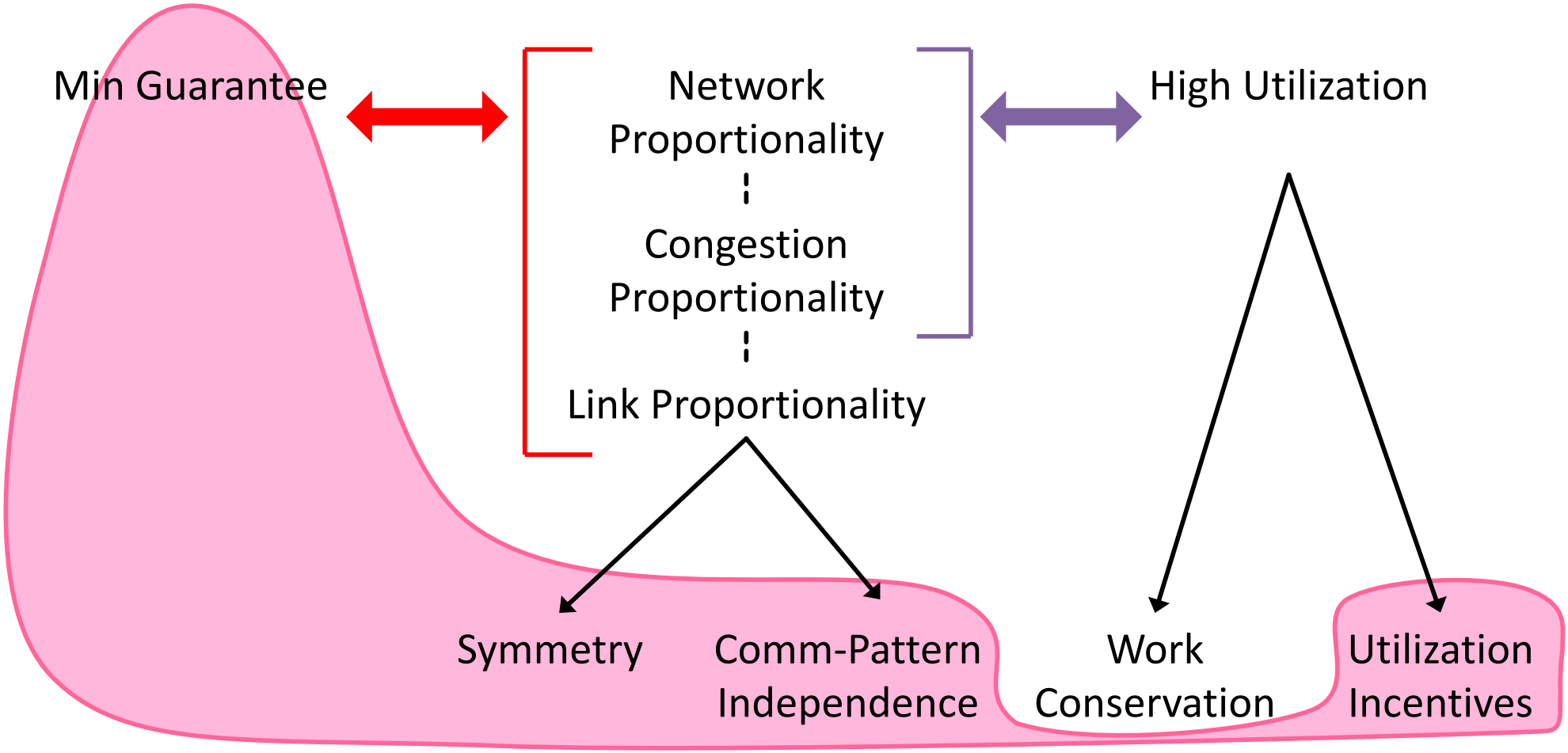
Per Flow (e.g. today)



Per Source (e.g., Seawall [NSDI'11])



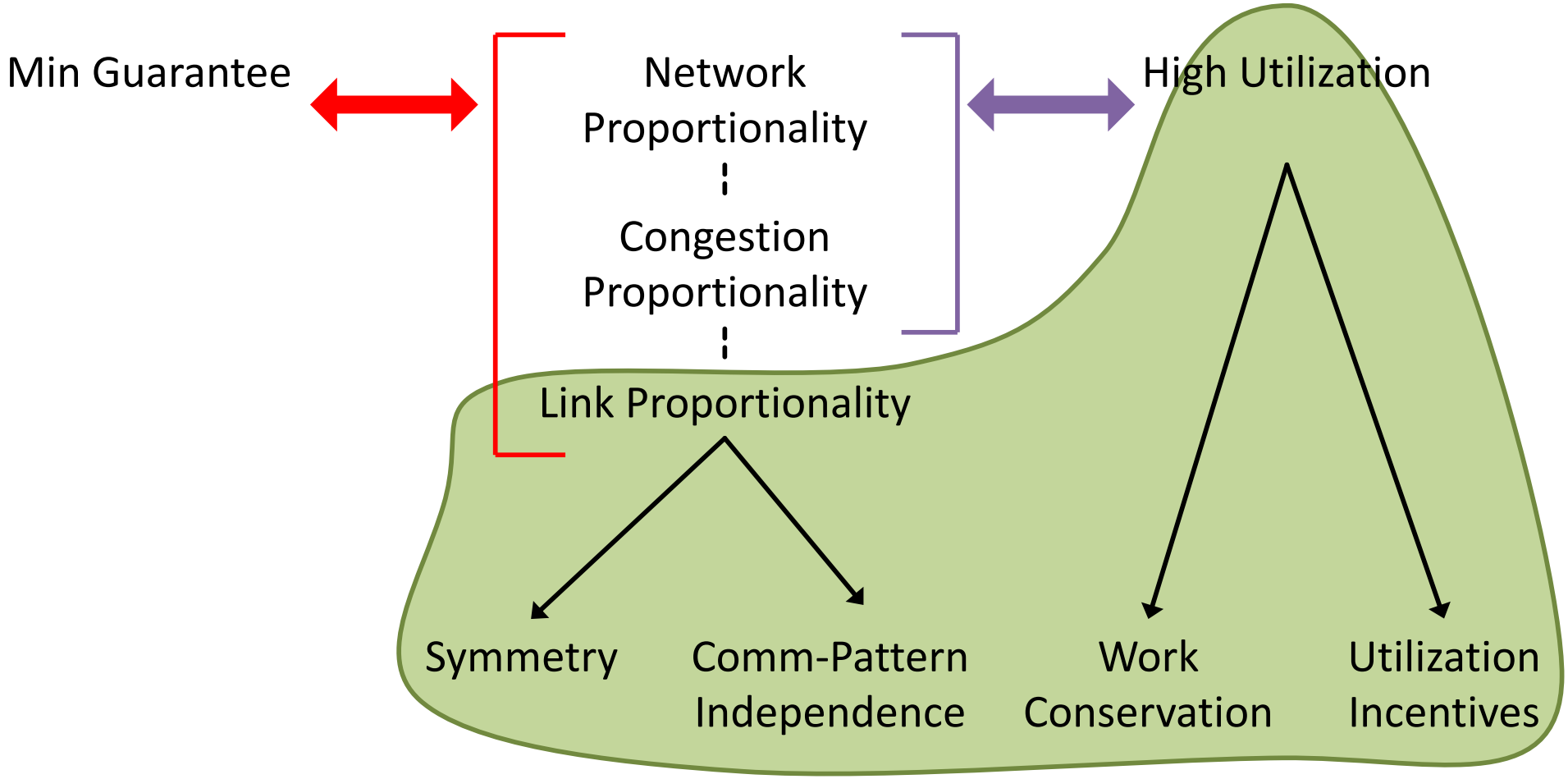
Static Reservation (e.g., Oktopus [Sigcomm'11])



New Allocation Policies

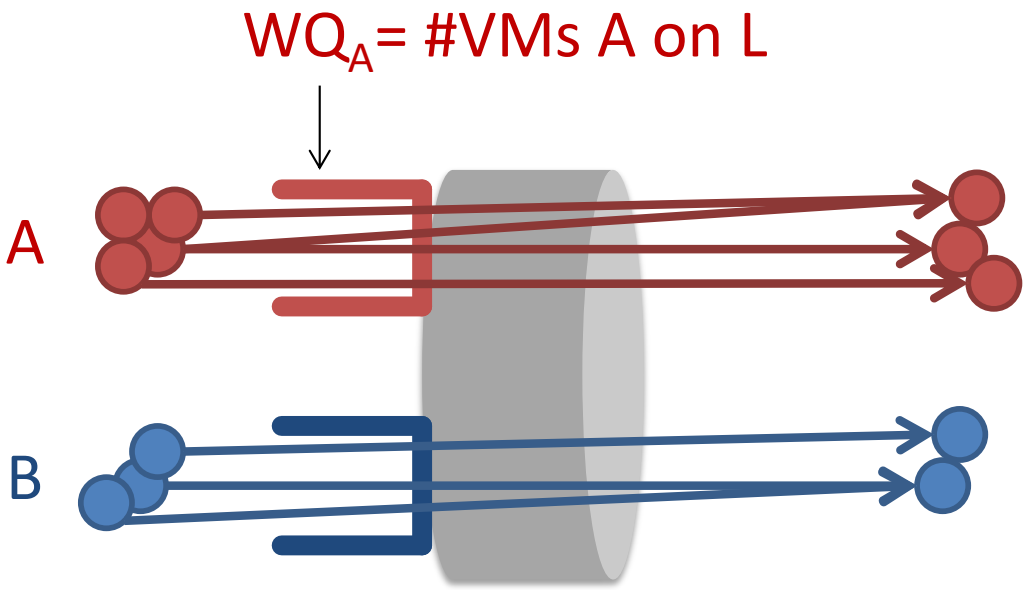
3 new allocation policies that take different stands on tradeoffs

Proportional Sharing at Link-level (PS-L)



Proportional Sharing at Link-level (PS-L)

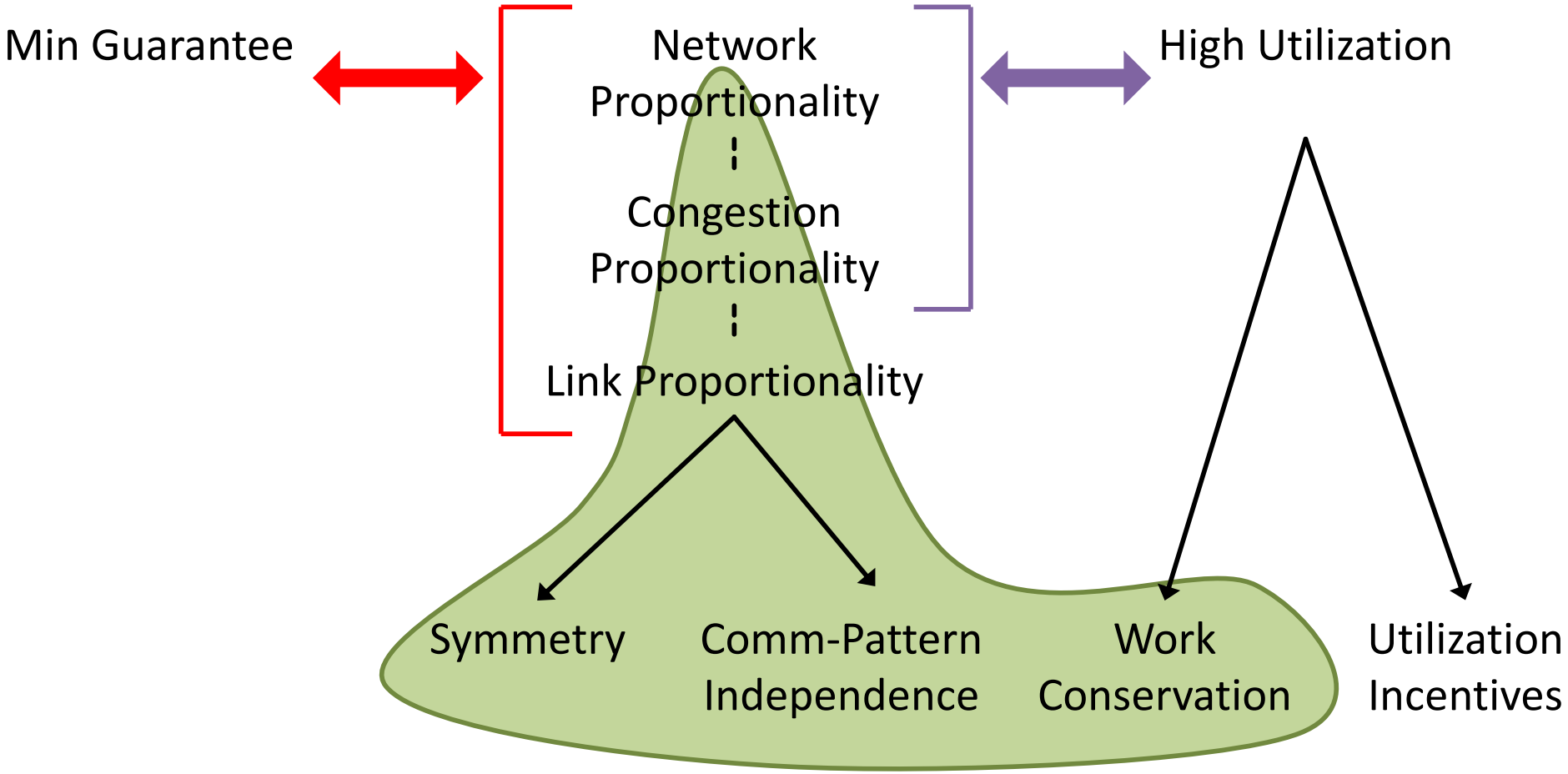
- Per tenant WFQ where weight = # tenant's VMs on link



$$\frac{Bw_A}{Bw_B} = \frac{\#VMs\ A\ on\ L}{\#VMs\ B\ on\ L}$$

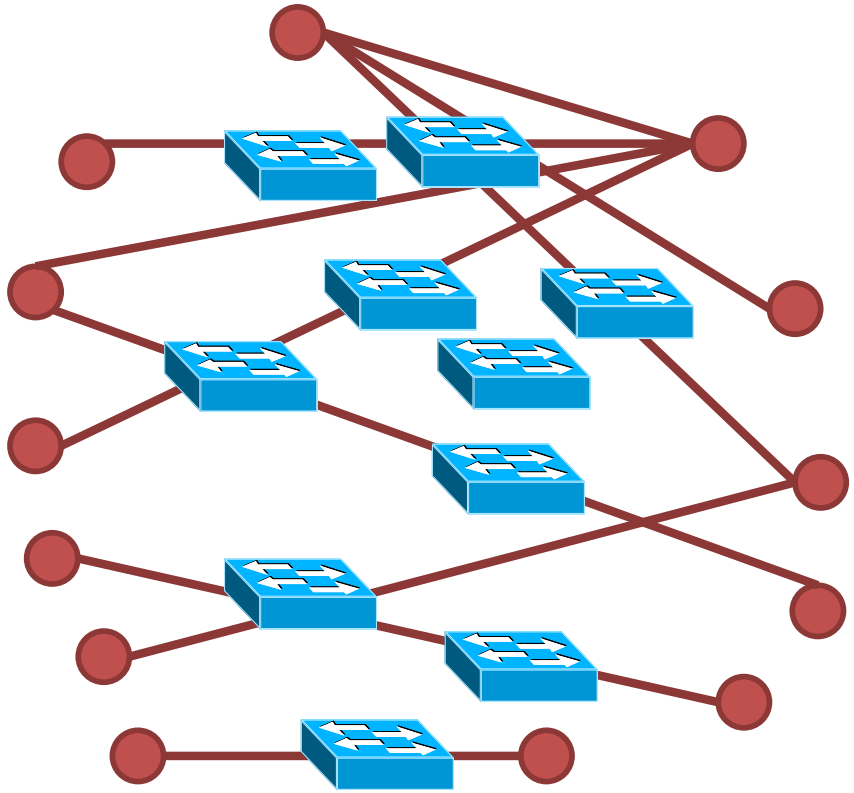
Can easily be extended to use heterogeneous VMs (by using VM weights)

Proportional Sharing at Network-level (PS-N)



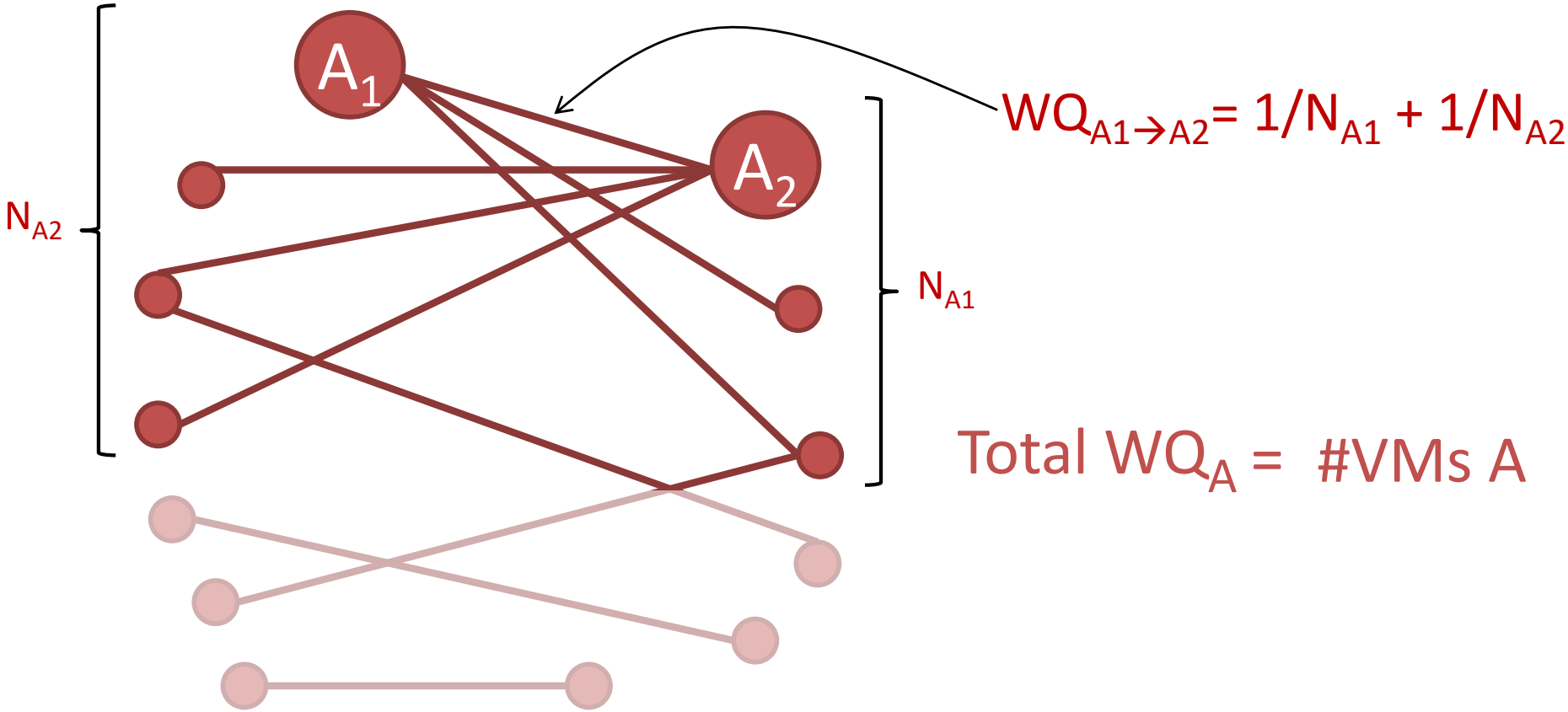
Proportional Sharing at Network-level (PS-N)

- Congestion proportionality in severely restricted context
- Per source-destination WFQ, total tenant weight = # VMs



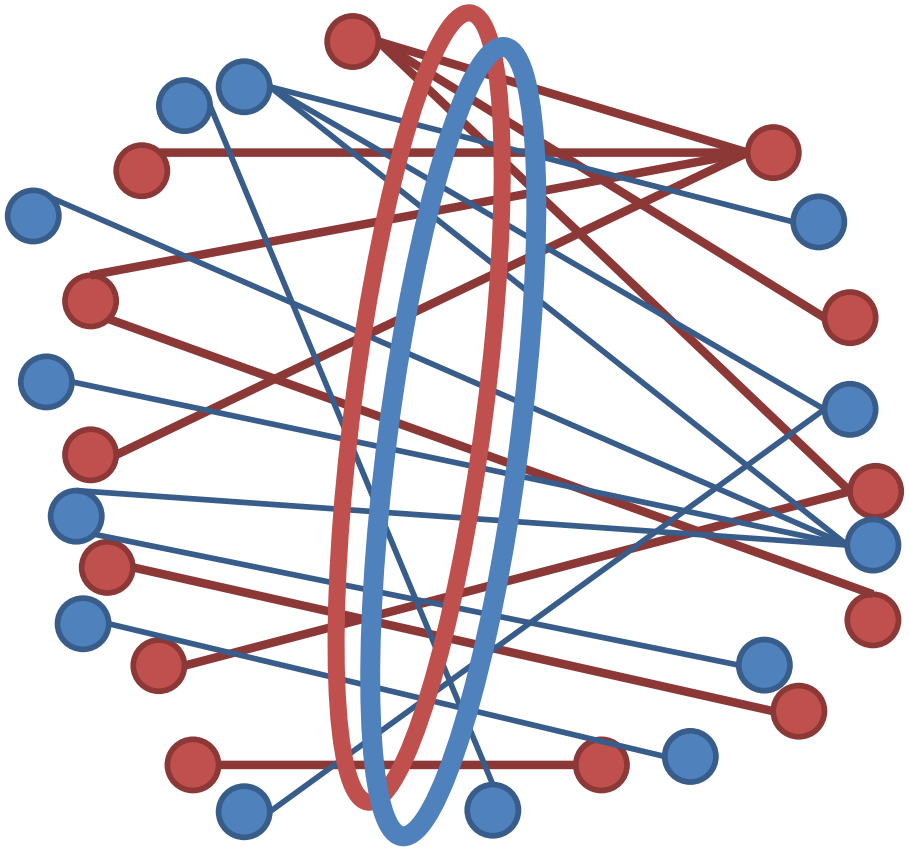
Proportional Sharing at Network-level (PS-N)

- Congestion proportionality in severely restricted context
- Per source-destination WFQ, total tenant weight = # VMs



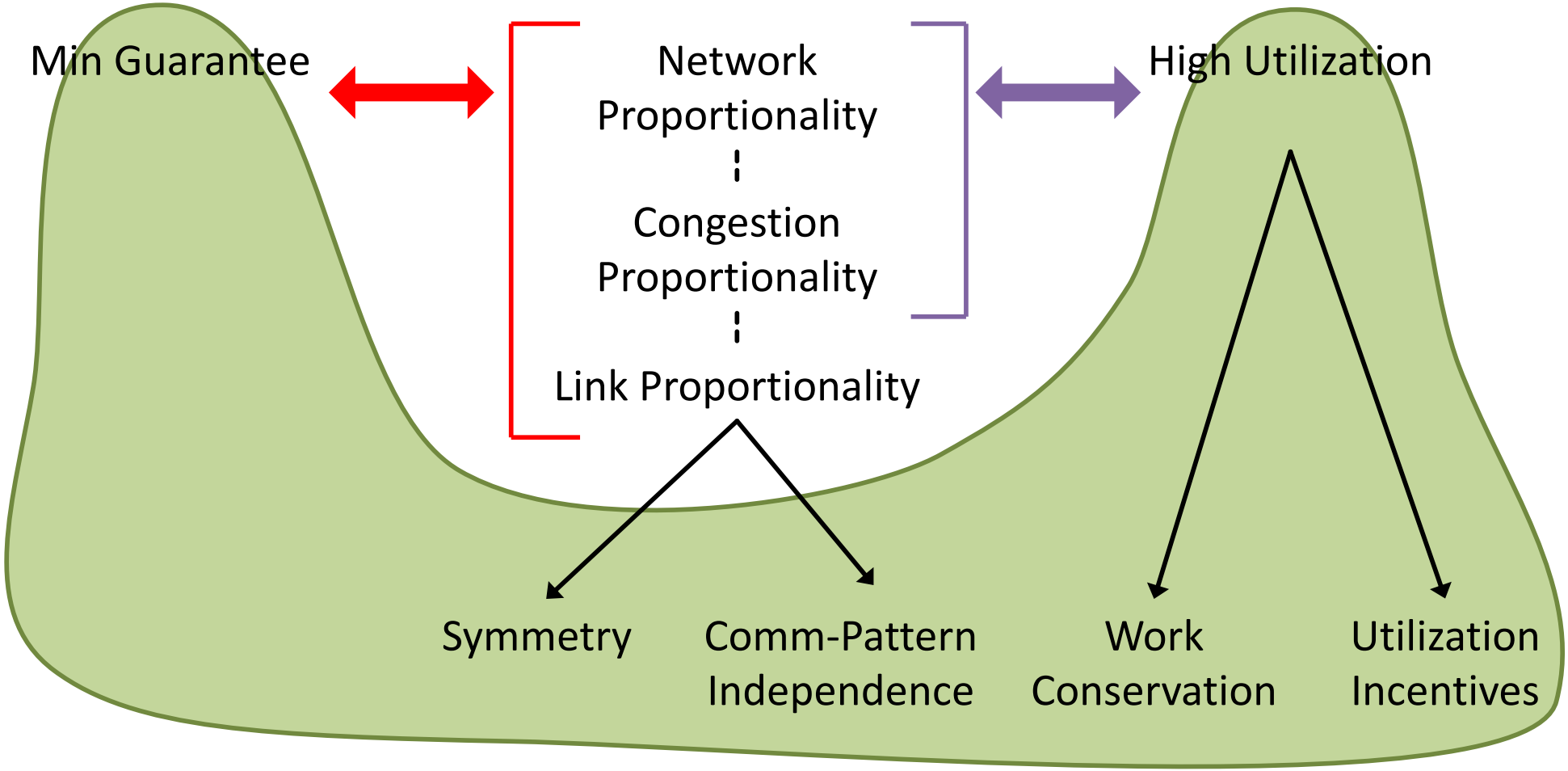
Proportional Sharing at Network-level (PS-N)

- Congestion proportionality in severely restricted context
- Per source-destination WFQ, total tenant weight = # VMs



$$\frac{WQ_A}{WQ_B} = \frac{\#VMs\ A}{\#VMs\ B}$$

Proportional Sharing on Proximate Links (PS-P)



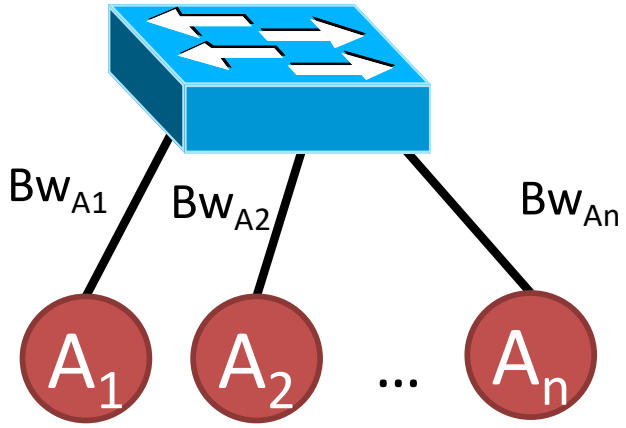
Proportional Sharing on Proximate Links (PS-P)

- Assumes a tree-based topology: traditional, fat-tree, VL2
  (currently working on removing this assumption)

Proportional Sharing on Proximate Links (PS-P)

- Assumes a tree-based topology: traditional, fat-tree, VL2
  (currently working on removing this assumption)

- Min guarantees
  - Hose model
  - Admission control

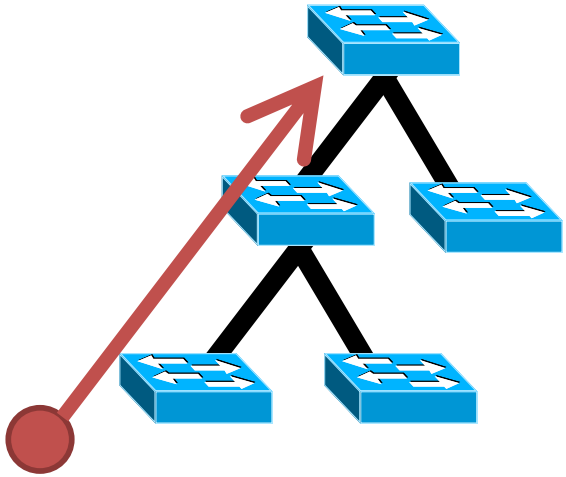


Proportional Sharing on Proximate Links (PS-P)

- Assumes a tree-based topology: traditional, fat-tree, VL2
  (currently working on removing this assumption)

- Min guarantees
  - Hose model
  - Admission control

- High Utilization
  - Per source fair sharing towards tree root

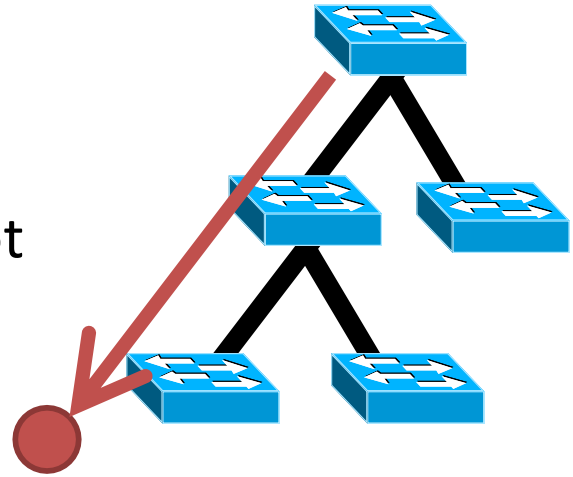


Proportional Sharing on Proximate Links (PS-P)

- Assumes a tree-based topology: traditional, fat-tree, VL2
  (currently working on removing this assumption)

- Min guarantees
  - Hose model
  - Admission control

- High Utilization
  - Per source fair sharing towards tree root
  - Per destination fair sharing from tree root



Deploying PS-L, PS-N and PS-P

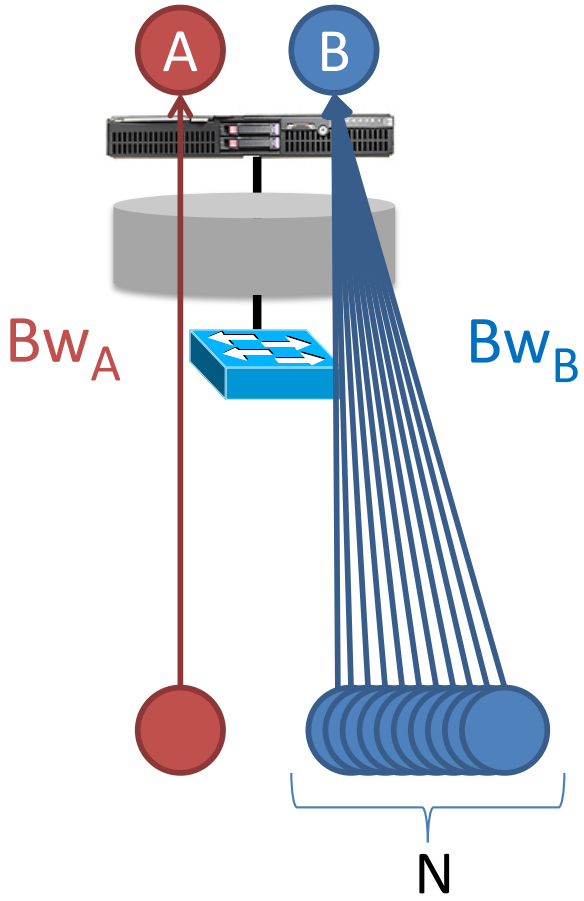
- Full Switch Support
  - All allocations can use hardware queues (per tenant, per VM or per source-destination)
- Partial Switch Support
  - PS-N and PS-P can be deployed using CSFQ [Sigcomm'98]
- No Switch Support
  - PS-N can be deployed using only hypervisors
  - PS-P could be deployed using only hypervisors, we are currently working on it

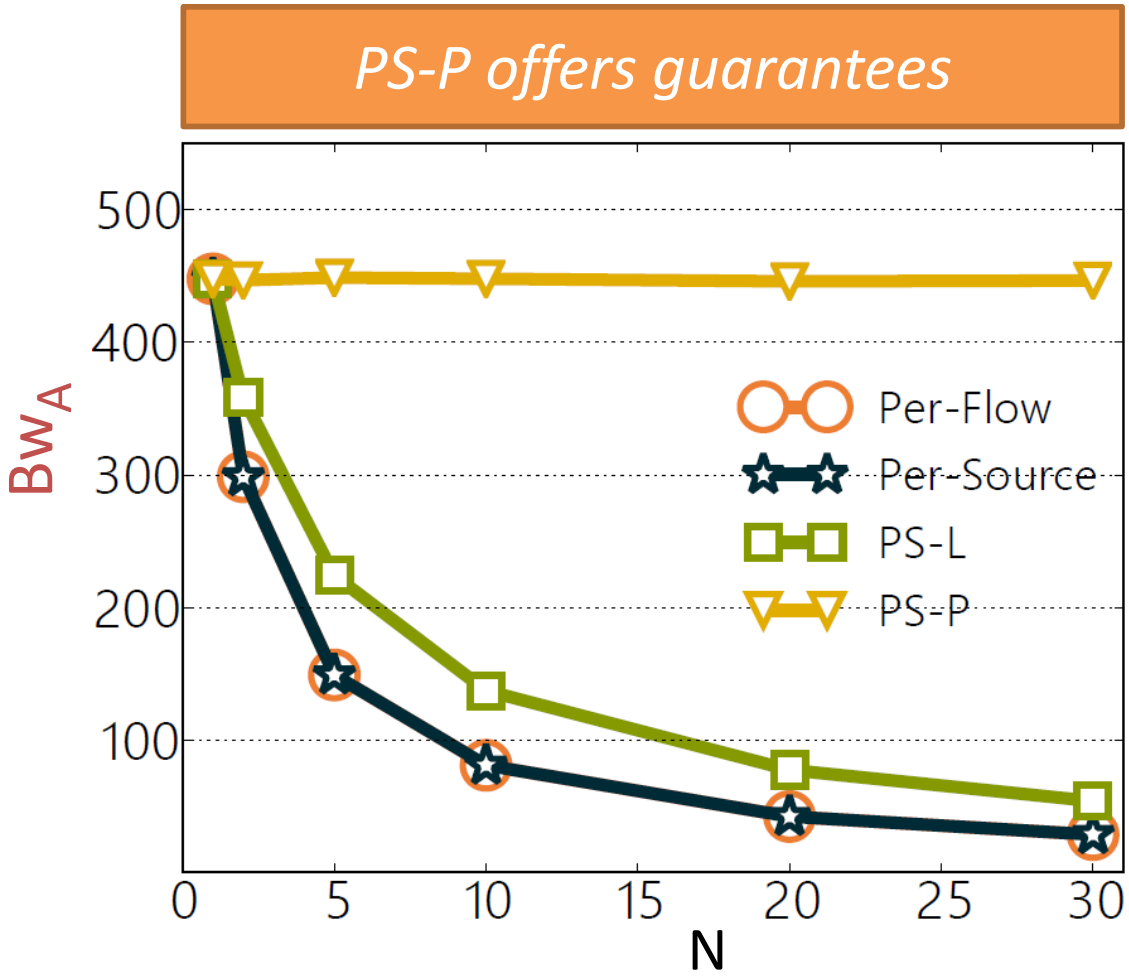
Evaluation

- Small Testbed + Click Modular Router
  - 15 servers, 1Gbps links
- Simulation + Real Traces
  - 3200 nodes, flow level simulator, Facebook MapReduce traces

Many to one

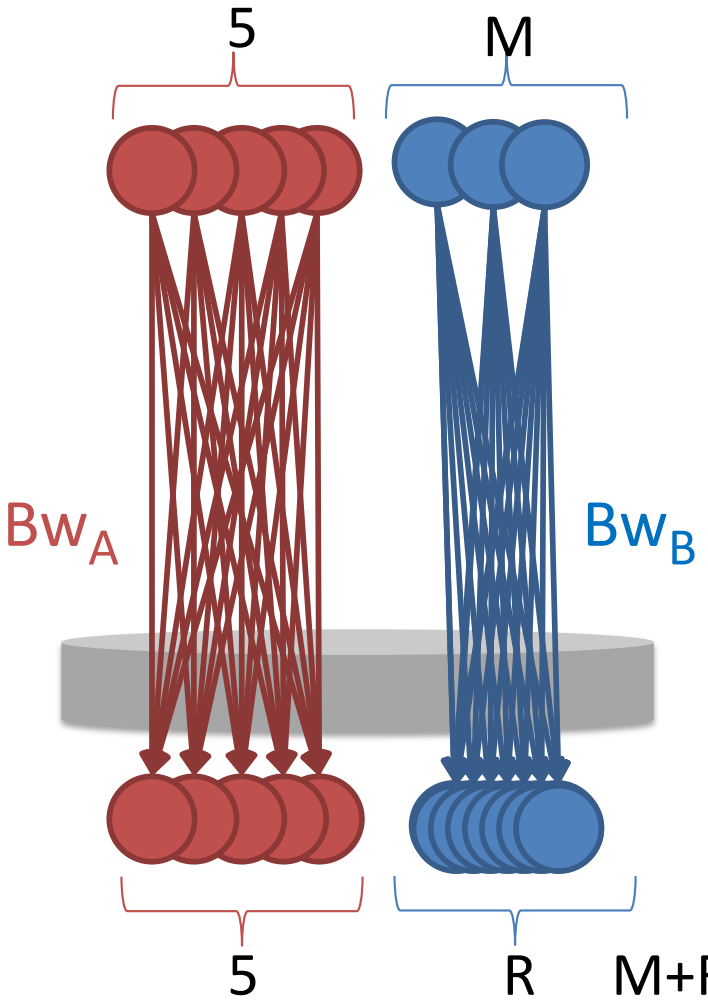
One link, testbed

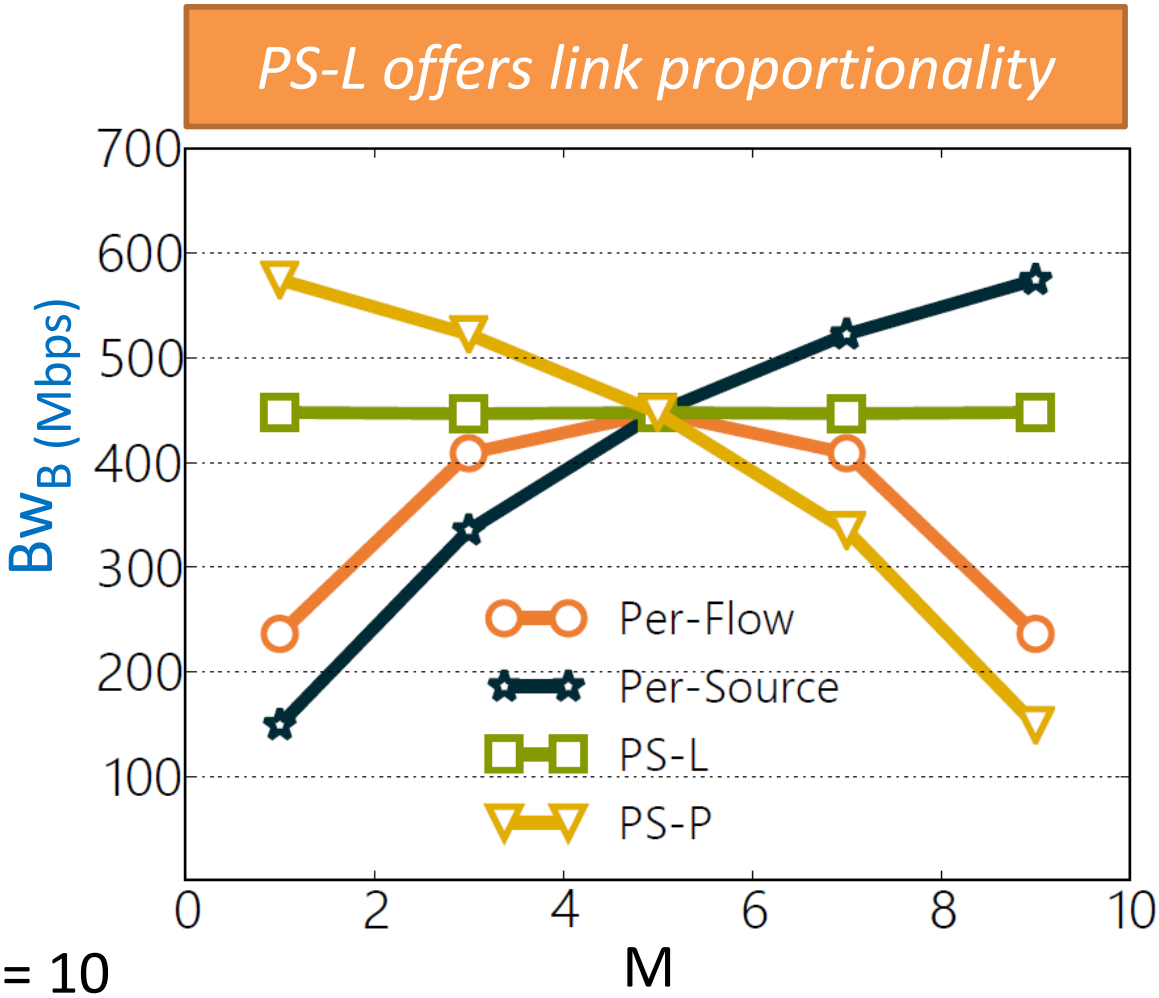




MapReduce

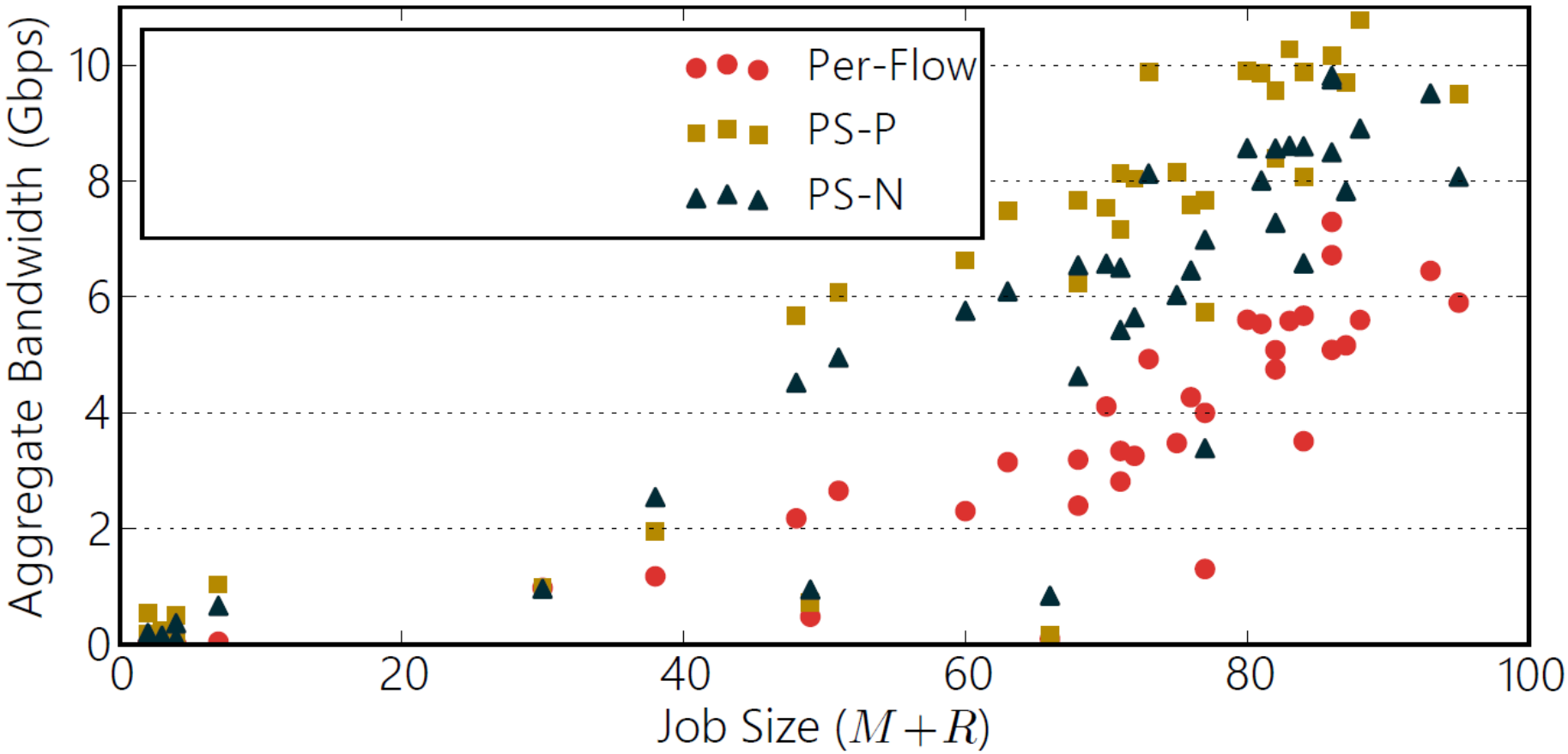
One link, testbed





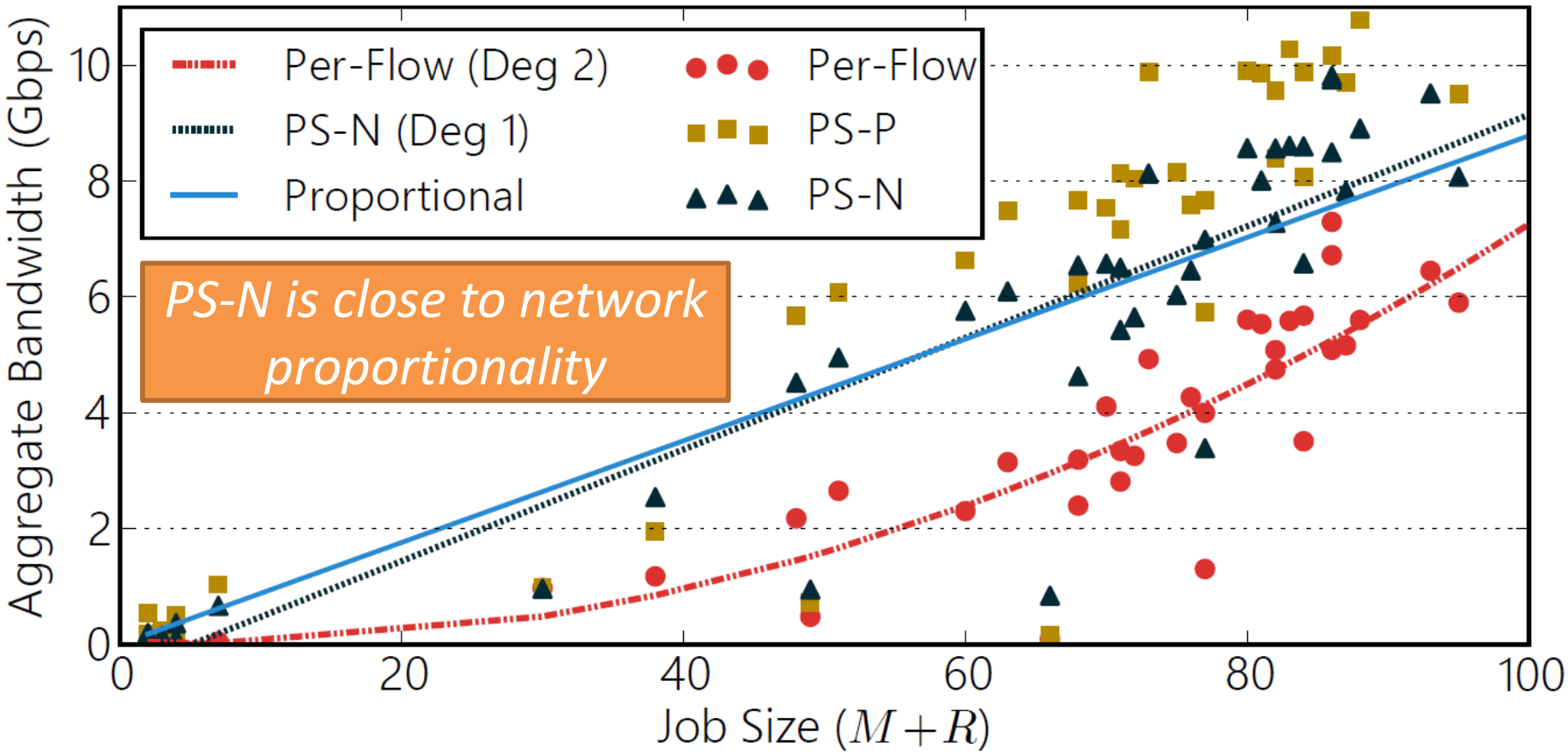
MapReduce

Network, simulation, Facebook trace



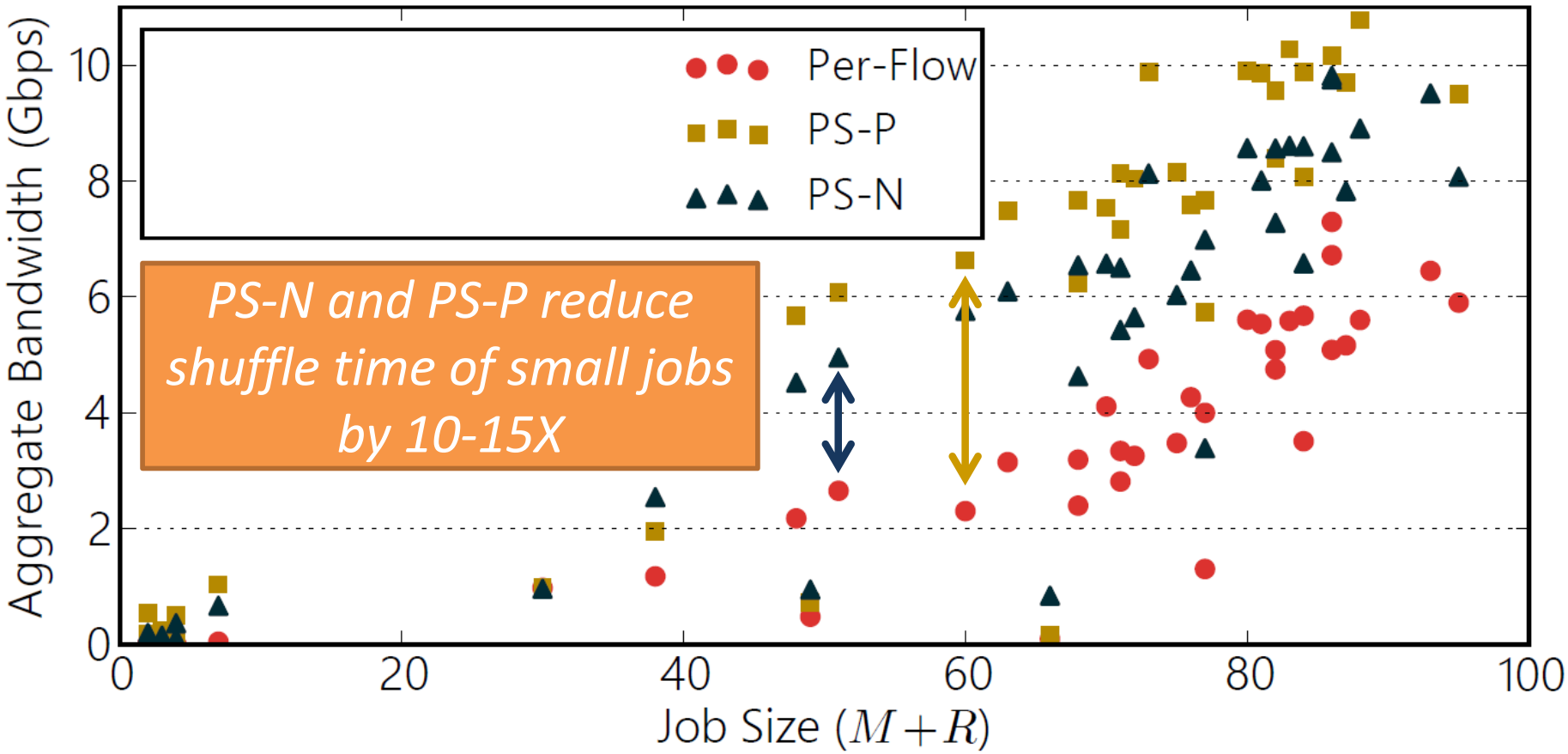
MapReduce

Network, simulation, Facebook trace



MapReduce

Network, simulation, Facebook trace



Summary

- Sharing cloud networks is not trivial
- First step towards a framework to analyze network sharing in cloud computing
  - Key goals (min guarantees, high utilization and proportionality), tradeoffs and properties
- New allocation policies, superset properties from past work
  - PS-L: link proportionality + high utilization
  - PS-N: restricted network proportional
  - PS-P: min guarantees + high utilization
- What are the assumptions, drawbacks?