Control of Queueing Systems

in Heavy Traffic

PhD Presentation, 2007

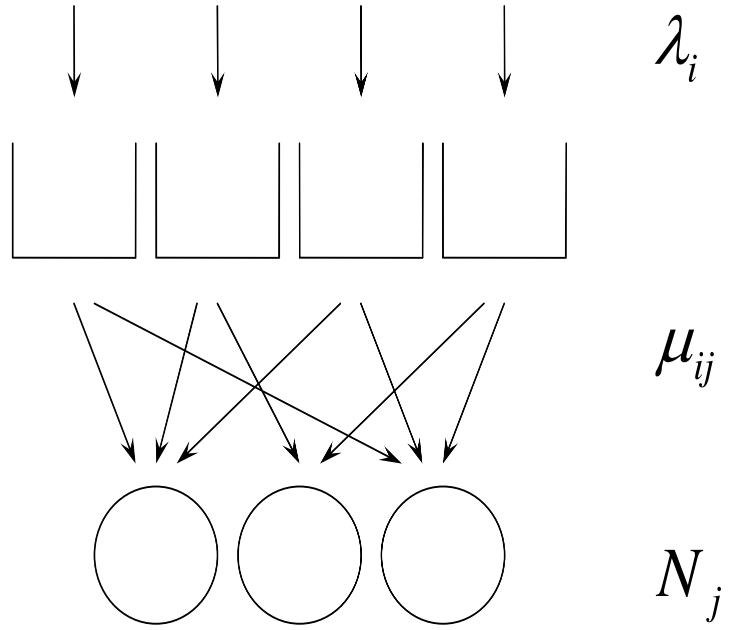
Gennady Shaikhet

Technion, Israel

Advisors: R. Atar and A. Mandelbaum

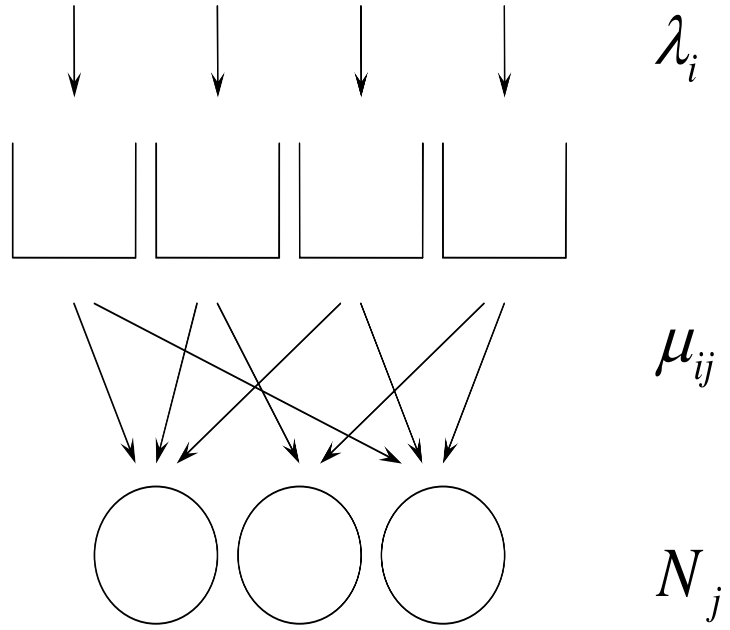
Queueing Model

- $I \geq 1$ customer classes
- $J \geq 1$ service stations
- Arrivals for class i:
  renewal processes, rate λ_i
- Servers in station j:
  N_j (stat. identical)
- Service of class-i by server-j:
  exponential, rate μ_{ij}



Queueing Model

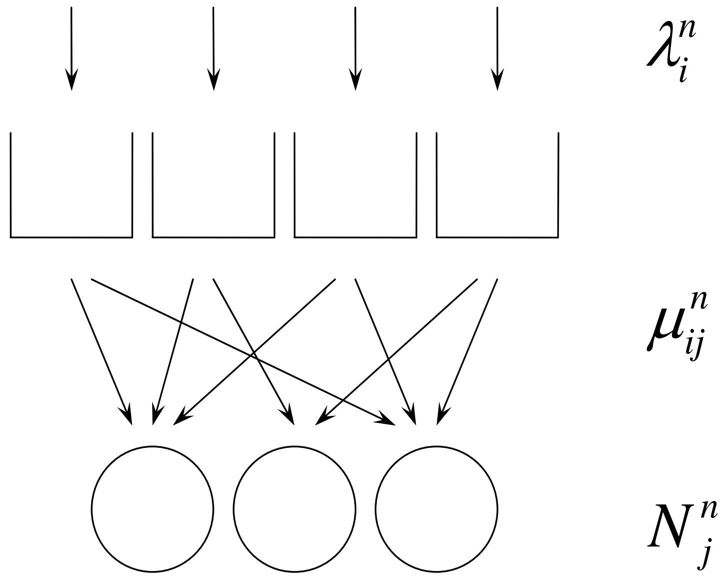
- $I \geq 1$ customer classes
- $J \geq 1$ service stations
- Arrivals for class i:
  renewal processes, rate λ_i
- Servers in station j:
  N_j (stat. identical)
- Service of class-i by server-j:
  exponential, rate μ_{ij}



- Control: has to be specified to complete the description:
  Routing customers   Scheduling servers

Heavy Traffic Regime

- Consider the sequence of systems, indexed by $n \uparrow \infty$
- $\lambda_i^n = n\lambda_i + O(\sqrt{n})$
- $n\mu_{ij}^n = n\mu_{ij} + O(\sqrt{n})$
- $N_j^n = n\nu_j + O(\sqrt{n})$



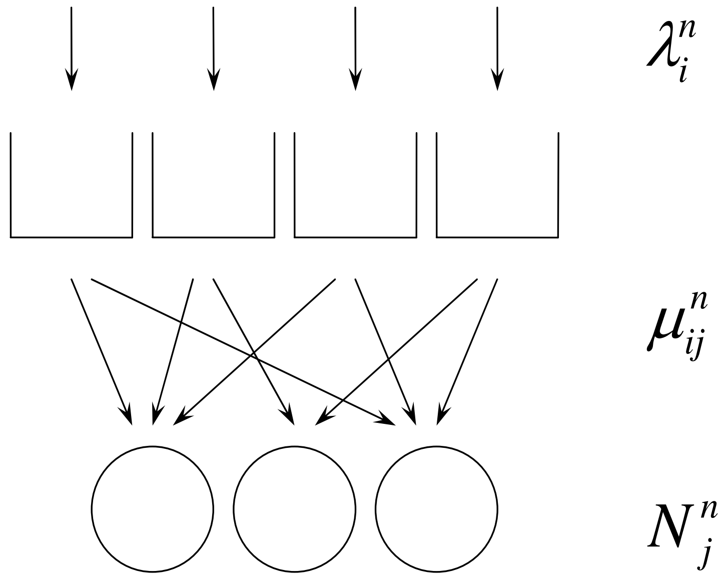
Heavy Traffic Regime

- Consider the sequence of systems, indexed by $n \uparrow \infty$

- $\lambda_i^n = n\lambda_i + O(\sqrt{n})$

- $n\mu_{ij}^n = n\mu_{ij} + O(\sqrt{n})$

- $N_j^n = n\nu_j + O(\sqrt{n})$



- The fluid (order n) level parameters λ, μ, ν guarantee that the system is critically loaded (busy on the fluid level).

Diffusion Scaling

Define:

$X_i^n(t)$ = number of class-i customers in the system at time t,

$Y_i^n(t)$ = number of class-i customers in the queue at time t,

$Z_j^n(t)$ = number of idle servers in station j at time t,

$\Psi_{ij}^n(t)$ = number of class-i customers in service in station j at time t,

Scale them around the static fluid: ψ_{ij}^* and x_i^*:

$$\hat{X}_i^n(t) = n^{-1/2}(X_i^n(t) - nx_i^*), \quad \hat{\Psi}_{ij}^n(t) = n^{-1/2}(\Psi_{ij}^n(t) - n\psi_{ij}^*).$$

$$\hat{Y}_i^n(t) = n^{-1/2} Y_i^n(t), \qquad \hat{Z}_i^n(t) = n^{-1/2} Z_i^n(t).$$

First Observation: Diffusion Model

- The following relation holds for all $t \geq 0$:

$$\hat{X}^n(t) = \hat{X}^n(0) + \hat{W}^n(t) + \int_0^t b(\hat{X}^n(s), U^n(s))ds + \sum_{c \in \mathcal{C}} m_c^n \int_0^t \hat{\Psi}_c^n(s)ds$$

Here U^n is a process with values in some compact space.

Also $0 \leq \hat{\Psi}_c^n \leq kn^{1/2}$ for some $k > 0$.

First Observation: Diffusion Model

- The following relation holds for all $t \geq 0$:

$$\hat{X}^n(t) = \hat{X}^n(0) + \hat{W}^n(t) + \int_0^t b(\hat{X}^n(s), U^n(s))ds + \sum_{c \in \mathcal{C}} m_c^n \int_0^t \hat{\Psi}_c^n(s)ds$$

  Here U^n is a process with values in some compact space.

  Also $0 \leq \hat{\Psi}_c^n \leq kn^{1/2}$ for some $k > 0$.

- As $n \to \infty$, the diffusion model can be rewritten as

$$X(t) = X(0) + W(t) + \int_0^t b(X(s), U(s))ds + \sum_{c \in \mathcal{C}} m_c \eta_c(t)$$

  For each c, η_c is nondecreasing with $\eta_c(0) \geq 0$.

First Observation: Diffusion Model

- The following relation holds for all $t \geq 0$:

$$\hat{X}^n(t) = \hat{X}^n(0) + \hat{W}^n(t) + \int_0^t b(\hat{X}^n(s), U^n(s))ds + \sum_{c \in \mathcal{C}} m_c^n \int_0^t \hat{\Psi}_c^n(s)ds$$

  Here U^n is a process with values in some compact space.

  Also $0 \leq \hat{\Psi}_c^n \leq kn^{1/2}$ for some $k > 0$.

- As $n \to \infty$, the diffusion model can be rewritten as

$$X(t) = X(0) + W(t) + \int_0^t b(X(s), U(s))ds + \sum_{c \in \mathcal{C}} m_c \eta_c(t)$$

  For each c, η_c is nondecreasing with $\eta_c(0) \geq 0$.

- Controlled diffusion with drift and singular control.

Effect of Singular Component

- Consider a singular controlled diffusion

$$X(t) = X(0) + W(t) + \int_0^t b(X(s), U(s))ds + \sum_{c \in \mathcal{C}} m_c \eta_c(t).$$

Effect of Singular Component

- Consider a singular controlled diffusion

$$X(t) = X(0) + W(t) + \int_0^t b(X(s), U(s))ds + \sum_{c \in \mathcal{C}} m_c \eta_c(t).$$

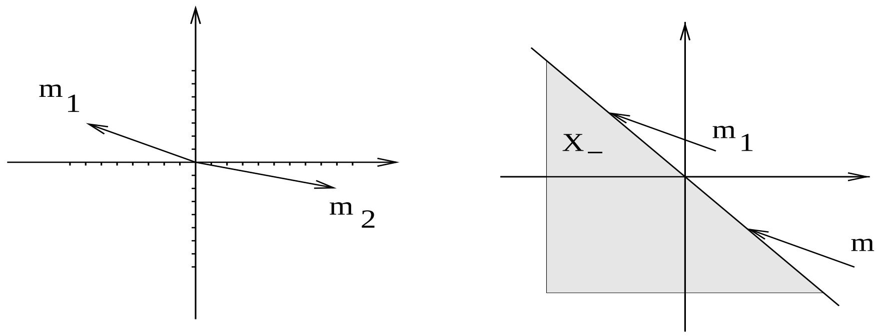
- The singular term η can restrict X to a certain domain.

Effect of Singular Component

- Consider a singular controlled diffusion

$$X(t) = X(0) + W(t) + \int_0^t b(X(s), U(s))ds + \sum_{c \in \mathcal{C}} m_c \eta_c(t).$$

- The singular term η can restrict X to a certain domain.
- It can happen that X can be restricted to a domain, corresponding to all queues being empty

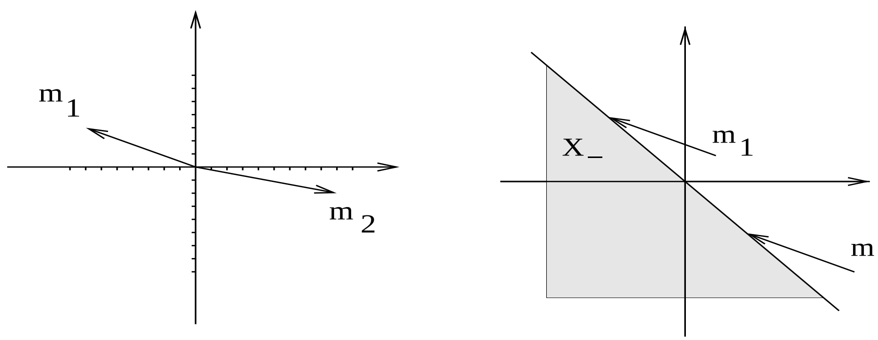


Effect of Singular Component

- Consider a singular controlled diffusion

$$X(t) = X(0) + W(t) + \int_0^t b(X(s), U(s))ds + \sum_{c \in \mathcal{C}} m_c \eta_c(t).$$

- The singular term η can restrict X to a certain domain.
- It can happen that X can be restricted to a domain, corresponding to all queues being empty



- It happens when $e \cdot m_c < 0$ for some c.

Connection to Original (prelimit) Model

- Goal: Find a policy,
  that asymptotically (large n) achieves empty queues.

  For two types of control policies:

Connection to Original (prelimit) Model

- Goal: Find a policy,

  that asymptotically (large n) achieves empty queues.

  For two types of control policies:

- Preemptive (P) regime:

  a service to a customer can be interrupted and resumed at a later time (possibly in a different station).

Connection to Original (prelimit) Model

- Goal: Find a policy,

  that asymptotically (large n) achieves empty queues.

  For two types of control policies:

- Preemptive (P) regime:

  a service to a customer can be interrupted and resumed at a later time (possibly in a different station).

- Non-preemptive (NP) regime:

  service to a customer can not be interrupted before it is completed

Asymptotic Null Controllability

- **Null controllability:** There exist a sequence of policies (both P and NP), s.t. for any given $0 < \varepsilon < T < \infty$,

$$\lim_{n\to\infty} P\Big(Y^n(t) = 0 \text{ for all } t \in [\varepsilon, T]\Big) = 1.$$

Asymptotic Null Controllability

- **Null controllability:** There exist a sequence of policies (both P and NP), s.t. for any given $0 < \varepsilon < T < \infty$,

$$\lim_{n\to\infty} P\Big(Y^n(t) = 0 \text{ for all } t \in [\varepsilon, T]\Big) = 1.$$

- Under weaker conditions, we have

  Weak null controllability: There exist a sequence of P policies, under which for any fixed $0 < T < \infty$,

$$\int_0^T 1_{\{e\cdot Y^n(s)>0\}} ds \to 0 \quad \text{in probability, as } n \to \infty,$$

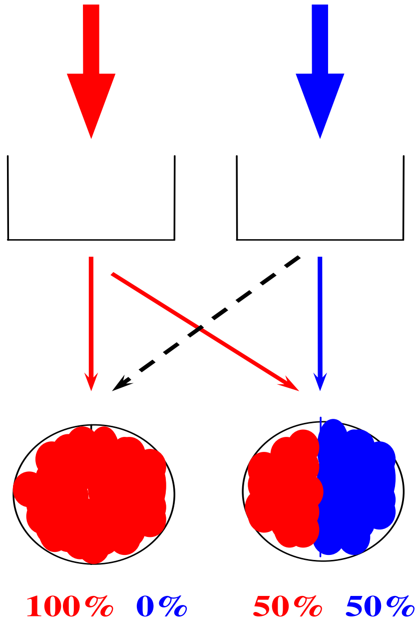
Critically Loaded System. Fluid View

- An example of critically loaded system:

$$\lambda_1 = 7.5, \quad \lambda_2 = 2$$

$$\mu_{11} = 4, \quad \mu_{12} = 7$$
$$\mu_{21} = 2, \quad \mu_{22} = 4$$

$$\nu_1 = 1, \quad \nu_2 = 1$$



$$\xi^*_{11} = 1, \ \xi^*_{12} = 0.5$$
$$\xi^*_{21} = 0, \ \xi^*_{22} = 0.5$$

$$\psi^*_{ij} = \nu_j \xi^*_{ij}.$$

- Any reallocation will cause some of the classes to explode.

Basic and non-basic activities

Activities: pairs (i, j), with $\mu_{ij} > 0$

Activities can be:

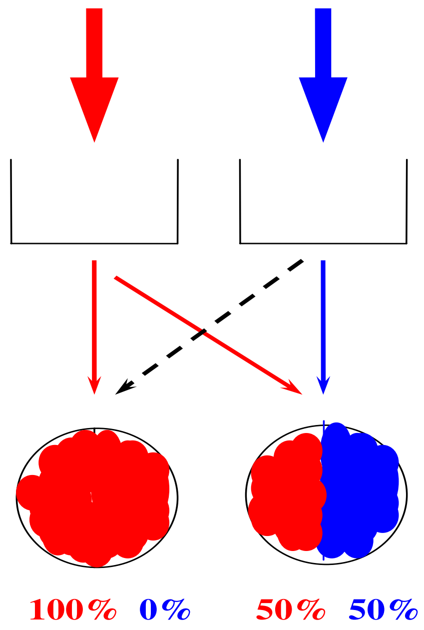
basic (BA), if $\xi^*_{ij} > 0$

non-basic, if $\xi^*_{ij} = 0$

In the example :

basic : $(1, 1)$, $(1, 2)$, $(2, 2)$

non-basic : $(2, 1)$



Basic and non-basic activities

Activities: pairs (i, j), with $\mu_{ij} > 0$

Activities can be:

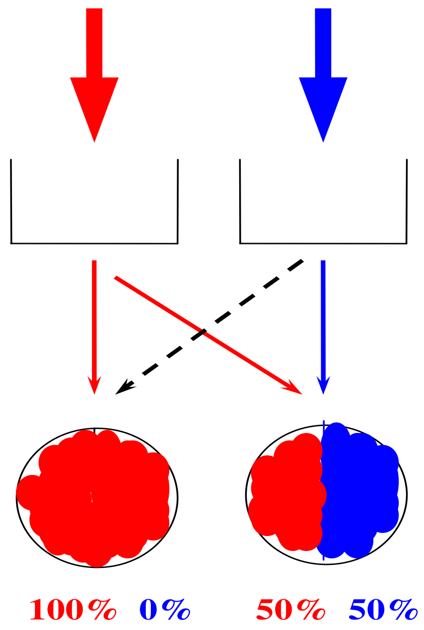
basic (BA), if $\xi^*_{ij} > 0$

non-basic, if $\xi^*_{ij} = 0$

In the example :

basic : $(1, 1)$, $(1, 2)$, $(2, 2)$

non-basic : $(2, 1)$



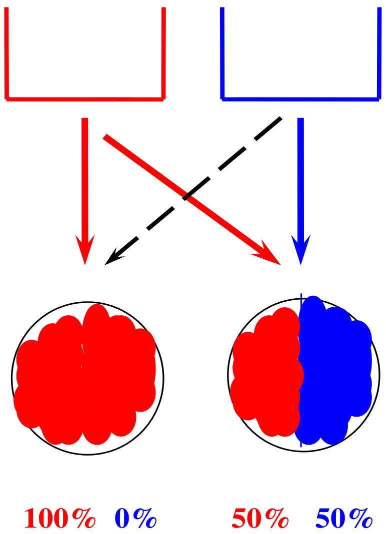
- **Usage of non–basic activities is a reason for a new behaviour.**

Reallocation via the non–basic activity

- Consider the following massive (order n) customers transfers:

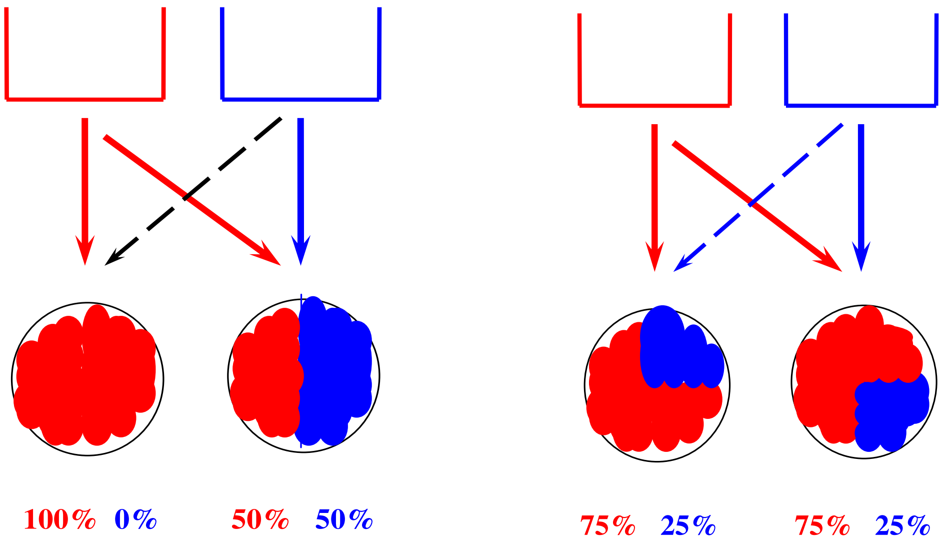
Reallocation via the non–basic activity

- Consider the following massive (order n) customers transfers:



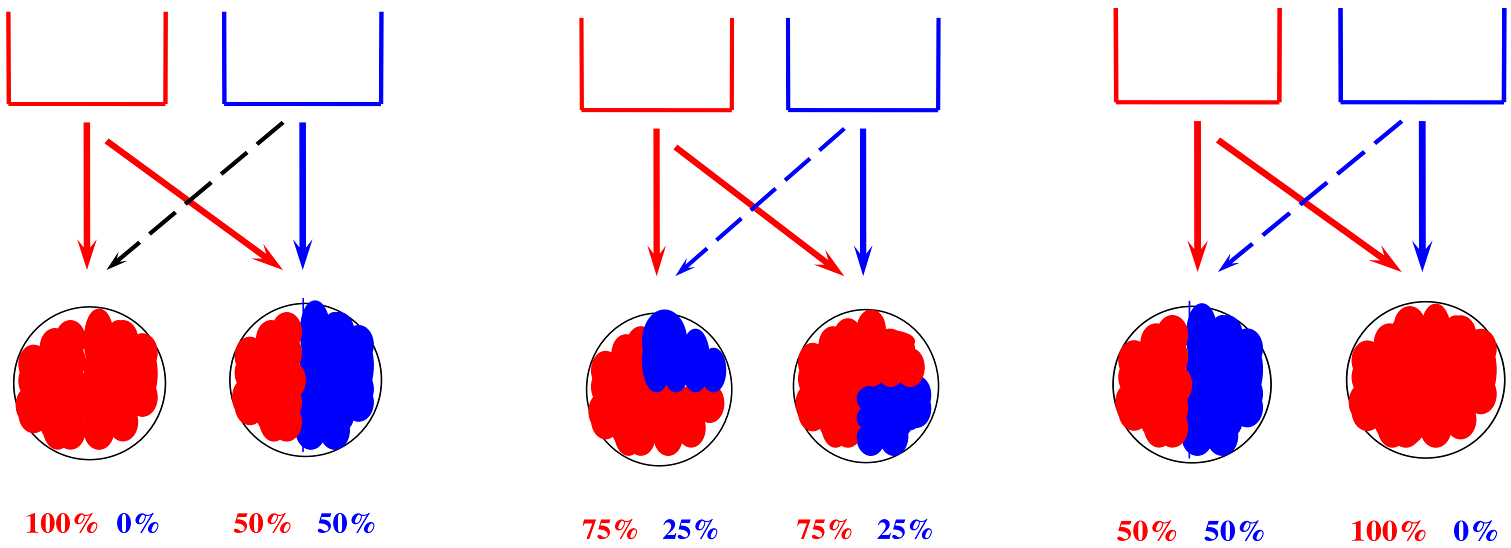
Reallocation via the non–basic activity

- Consider the following massive (order n) customers transfers:



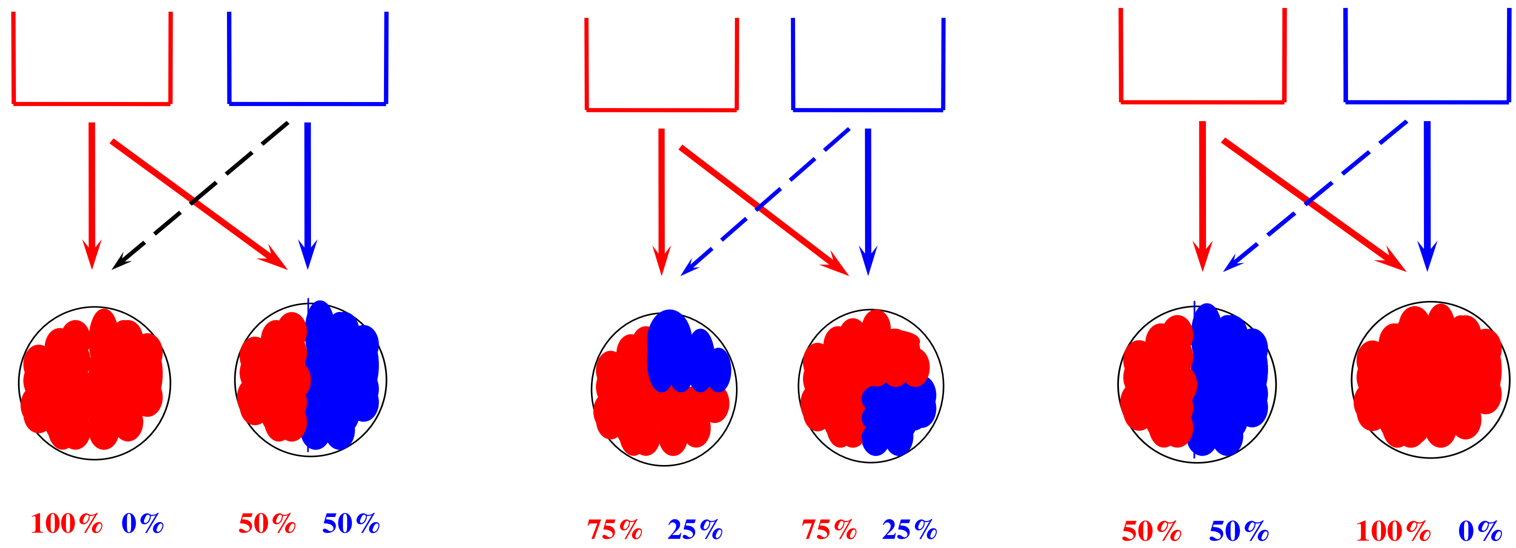
Reallocation via the non–basic activity

- Consider the following massive (order n) customers transfers:



Reallocation via the non–basic activity

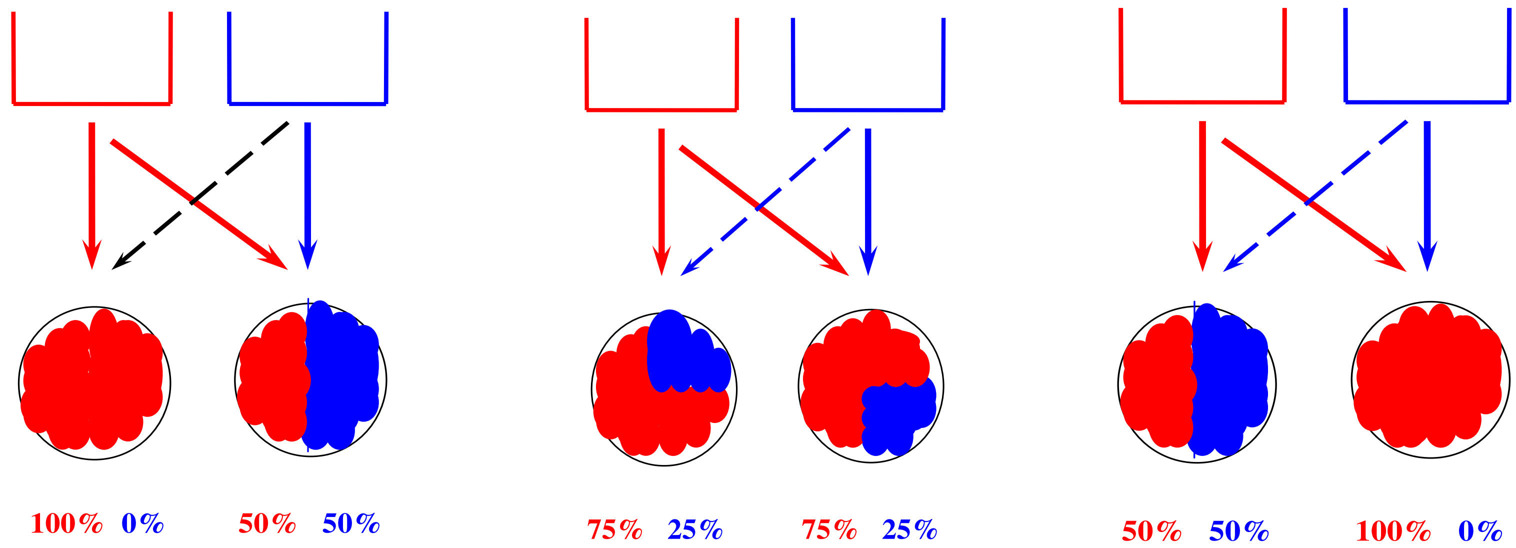
- Consider the following massive (order n) customers transfers:



- Performed instantaneously, such transfers may result in abrupt change of a total service rate.

Reallocation via the non–basic activity

- Consider the following massive (order n) customers transfers:



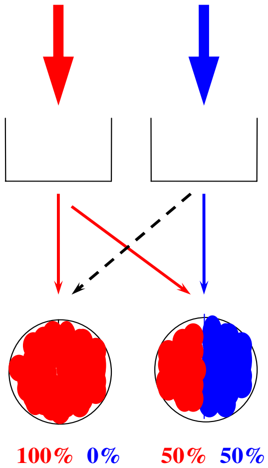
- Performed instantaneously, such transfers may result in abrupt change of a total service rate.

- The above reallocation does not generate immediate queues.
  The reallocation is performed via the closed simple path (simple cycle).

  Closed simple path - a cyclic graph, with one non–basic activity, the rest are basic.

Changing the Fluid Throughput

$\lambda_1 = 7.5, \lambda_2 = 2, \mu_{11} = 4, \mu_{12} = 7, \mu_{21} = 2, \mu_{22} = 4$



Total incoming rate: $7.5 + 2 = 9.5$

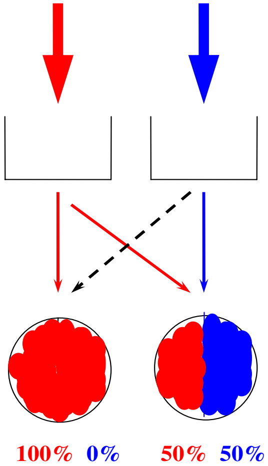
Total processing rate:

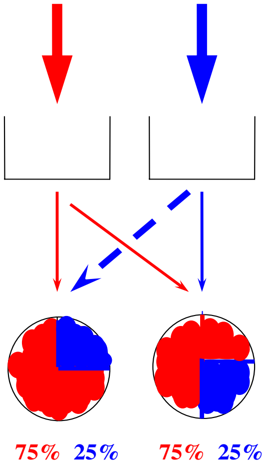
$$4 \cdot 1 + 7 \cdot 0.5 + 4 \cdot 0.5 = 9.5$$

(Total) output equals to input.

Changing the Fluid Throughput

$\lambda_1 = 7.5, \lambda_2 = 2, \mu_{11} = 4, \mu_{12} = 7, \mu_{21} = 2, \mu_{22} = 4$





Total incoming rate: $7.5 + 2 = 9.5$

Total processing rate:

$$4 \cdot 1 + 7 \cdot 0.5 + 4 \cdot 0.5 = 9.5$$

(Total) output equals to input.

Total incoming rate: $7.5 + 2 = 9.5$

Total processing rate:

$$4 \cdot 0.75 + 7 \cdot 0.75 + 2 \cdot 0.25 + 4 \cdot 0.25 = 9.75.$$

(Total) output is greater than input.

Changing the Fluid Throughput

$$\lambda_1 = 7.5, \lambda_2 = 2, \mu_{11} = 4, \mu_{12} = 7, \mu_{21} = 2, \mu_{22} = 4$$

Total incoming rate: $7.5 + 2 = 9.5$

Total processing rate:

$$4 \cdot 1 + 7 \cdot 0.5 + 4 \cdot 0.5 = 9.5$$

(Total) output equals to input.

Total incoming rate: $7.5 + 2 = 9.5$

Total processing rate:

$$4 \cdot 0.75 + 7 \cdot 0.75 + 2 \cdot 0.25 + 4 \cdot 0.25 = 9.75.$$

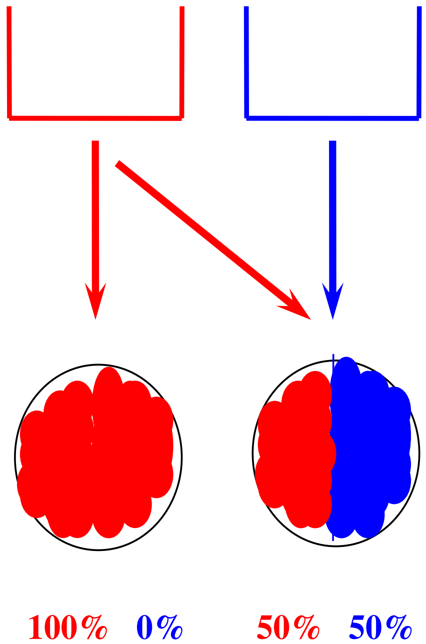
(Total) output is greater than input.

- **The existence of a closed simple path, that increases the throughput, implies (strong) null controllability.**

Activities and non–activities

Activities: pairs (i, j), with $\mu_{ij} > 0$

In the example :



Activities : $(1, 1)$, $(1, 2)$, $(2, 2)$

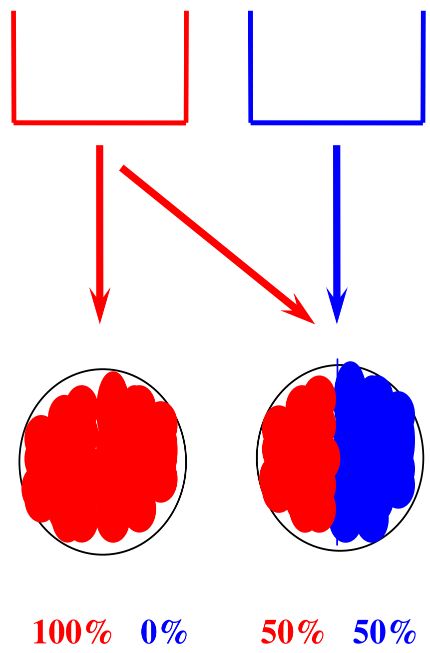
Non–activity : $(2, 1)$

Activities and non–activities

Activities: pairs (i, j), with $\mu_{ij} > 0$

In the example :

Activities : $(1, 1)$, $(1, 2)$, $(2, 2)$
Non–activity : $(2, 1)$



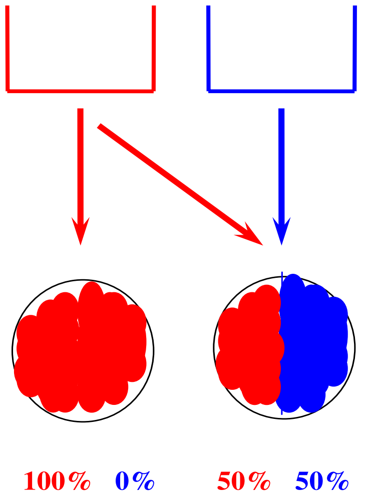
- **"Usage" of <u>non–activities</u> may also imply a new behaviour.**

Reallocation via the non–activity

- Consider the following massive (order n) customers transfers:

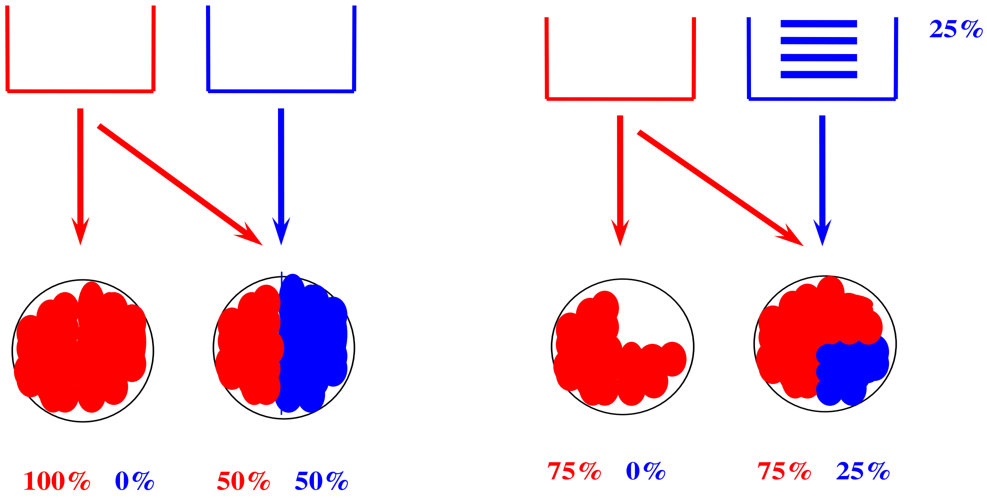
Reallocation via the non–activity

- Consider the following massive (order n) customers transfers:



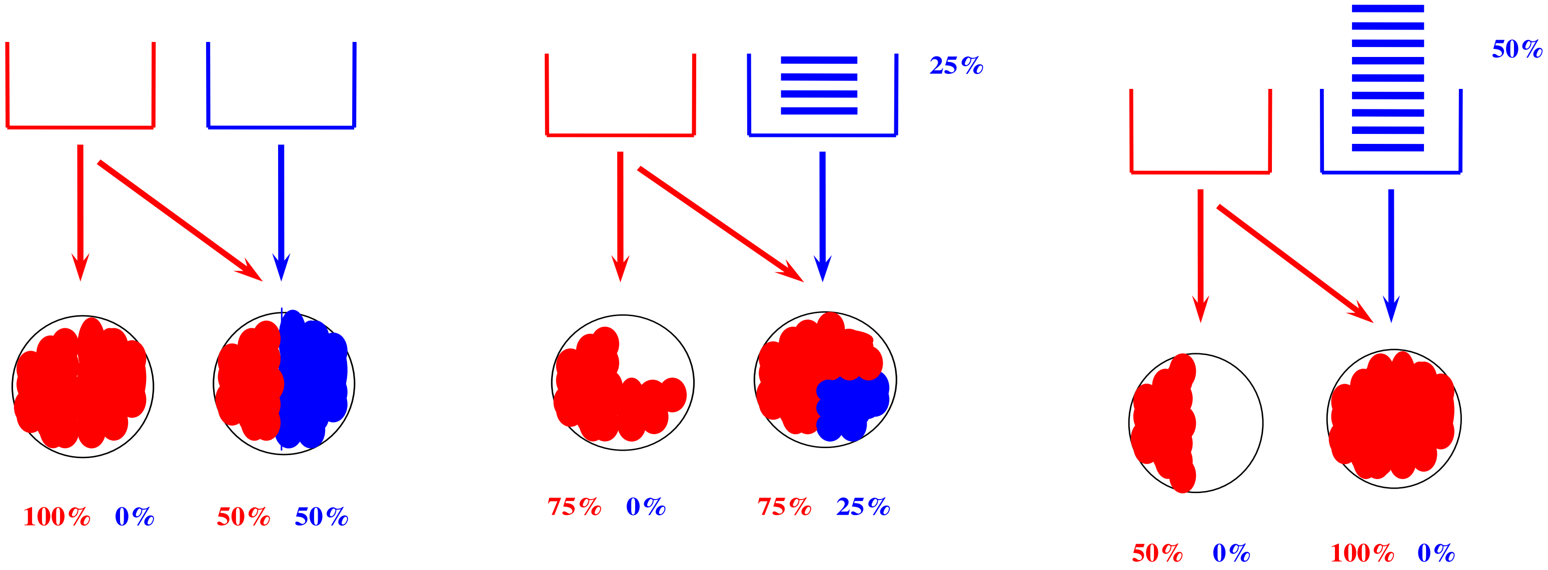
Reallocation via the non–activity

- Consider the following massive (order n) customers transfers:



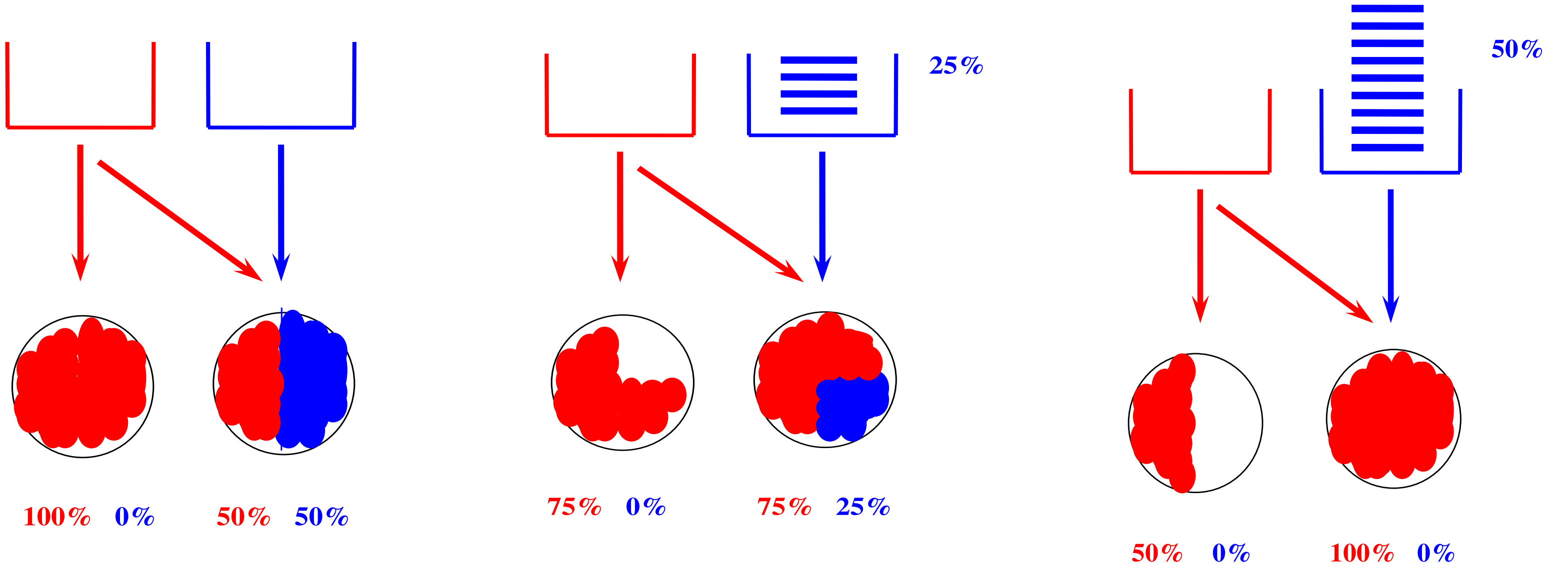
Reallocation via the non–activity

- Consider the following massive (order n) customers transfers:



Reallocation via the non–activity

- Consider the following massive (order n) customers transfers:

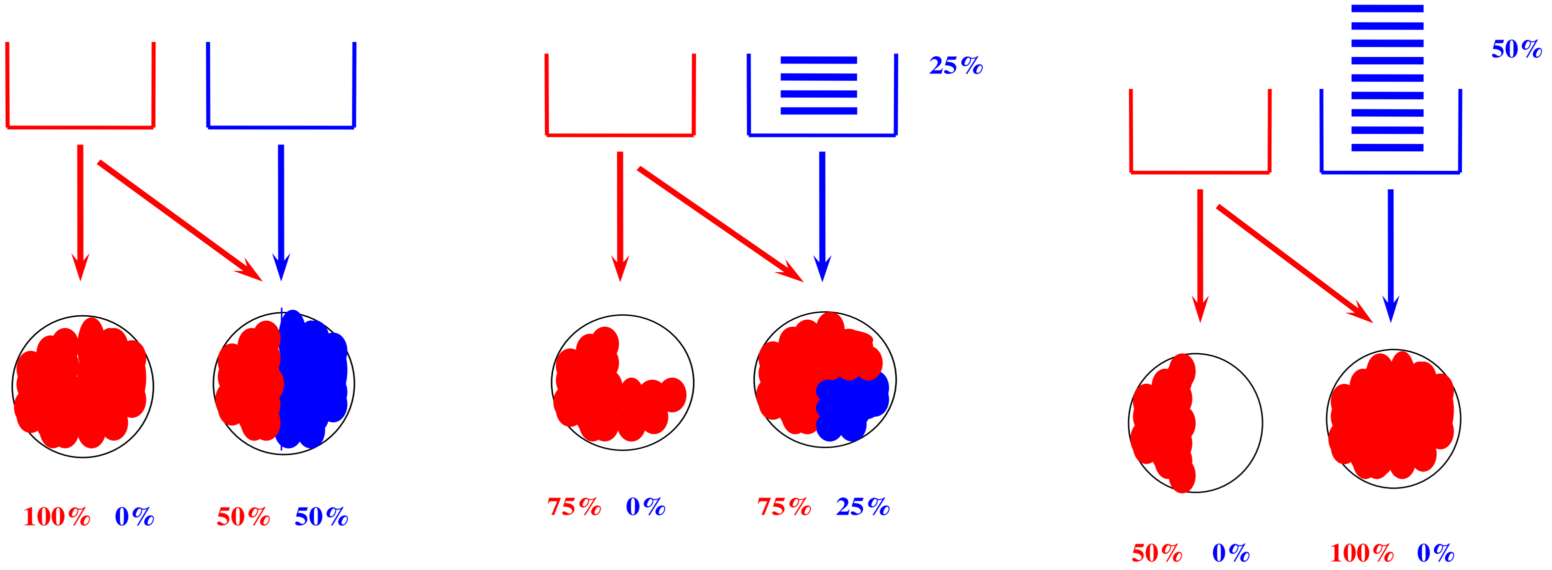


- The above reallocation generates immediate queues.
  The reallocation is performed via the open simple path (imaginary cycle).

  Open simple path - a cyclic graph, with one non–activity, the rest are basic activities.

Reallocation via the non–activity

- Consider the following massive (order n) customers transfers:



- The above reallocation generates immediate queues.
  The reallocation is performed via the open simple path (imaginary cycle).

  Open simple path - a cyclic graph, with one non–activity, the rest are basic activities.

- **The existence of an open simple path, that increases the throughput, implies weak null controllability.**

Throughput Optimality

- Recall the Heavy Traffic requirements:

$$\sum_i \xi^*_{ij} = 1, \quad \forall j \in \mathcal{J}, \quad \sum_j \mu_{ij} \nu_j \xi^*_{ij} = \lambda_i, \quad x^*_i := \sum_{j \in \mathcal{J}} \nu_j \xi^*_{ij} \quad \forall i \in \mathcal{I}.$$

Throughput Optimality

- Recall the Heavy Traffic requirements:

$$\sum_i \xi^*_{ij} = 1, \quad \forall j \in \mathcal{J}, \quad \sum_j \mu_{ij}\nu_j\xi^*_{ij} = \lambda_i, \quad x^*_i := \sum_{j\in\mathcal{J}} \nu_j\xi^*_{ij} \quad \forall i \in \mathcal{I}.$$

- We will say that the static fluid model is throughput optimal if

Whenever $\sum_i \xi_{ij} \leq 1, \quad \forall j \in \mathcal{J}$ and $\sum_{j\in\mathcal{J}} \nu_j\xi_{ij} \leq x^*_i, \quad \forall\, i \in \mathcal{I},$ one has

$$\sum_{(i,j)\in\mathcal{E}} \mu_{ij}\nu_j\xi_{ij} \leq \sum_{i\in\mathcal{I}} \lambda_i.$$

Throughput Optimality

- Recall the Heavy Traffic requirements:

$$\sum_i \xi^*_{ij} = 1, \quad \forall j \in \mathcal{J}, \quad \sum_j \mu_{ij}\nu_j\xi^*_{ij} = \lambda_i, \quad x^*_i := \sum_{j\in\mathcal{J}} \nu_j \xi^*_{ij} \quad \forall i \in \mathcal{I}.$$

- We will say that the static fluid model is throughput optimal if

Whenever $\sum_i \xi_{ij} \le 1, \quad \forall j \in \mathcal{J}$ and $\sum_{j\in\mathcal{J}} \nu_j \xi_{ij} \le x^*_i, \quad \forall\, i \in \mathcal{I},$ one has

$$\sum_{(i,j)\in\mathcal{E}} \mu_{ij}\nu_j\xi_{ij} \le \sum_{i\in\mathcal{I}} \lambda_i.$$

- **Theorem:** The following statements are equivalent:

Throughput Optimality

- Recall the Heavy Traffic requirements:

$$\sum_i \xi^*_{ij} = 1, \quad \forall j \in \mathcal{J}, \quad \sum_j \mu_{ij}\nu_j\xi^*_{ij} = \lambda_i, \quad x^*_i := \sum_{j\in\mathcal{J}} \nu_j\xi^*_{ij} \quad \forall i \in \mathcal{I}.$$

- We will say that the static fluid model is throughput optimal if

Whenever $\sum_i \xi_{ij} \le 1, \quad \forall j \in \mathcal{J}$ and $\sum_{j\in\mathcal{J}} \nu_j\xi_{ij} \le x^*_i, \quad \forall\, i \in \mathcal{I},$ one has

$$\sum_{(i,j)\in\mathcal{E}} \mu_{ij}\nu_j\xi_{ij} \le \sum_{i\in\mathcal{I}} \lambda_i.$$

- **Theorem:** The following statements are equivalent:

1. The static fluid model is not throughput optimal;

Throughput Optimality

- Recall the Heavy Traffic requirements:

$$\sum_i \xi_{ij}^* = 1, \quad \forall j \in \mathcal{J}, \quad \sum_j \mu_{ij}\nu_j\xi_{ij}^* = \lambda_i, \quad x_i^* := \sum_{j\in\mathcal{J}} \nu_j\xi_{ij}^* \quad \forall i \in \mathcal{I}.$$

- We will say that the static fluid model is throughput optimal if

Whenever $\sum_i \xi_{ij} \le 1, \quad \forall j \in \mathcal{J}$ and $\sum_{j\in\mathcal{J}} \nu_j\xi_{ij} \le x_i^*, \quad \forall\, i \in \mathcal{I},$ one has

$$\sum_{(i,j)\in\mathcal{E}} \mu_{ij}\nu_j\xi_{ij} \le \sum_{i\in\mathcal{I}} \lambda_i.$$

- **Theorem:** The following statements are equivalent:

1. The static fluid model is not throughput optimal;
2. There exists a throughput increasing simple path (either open or closed).

Pool–Dependent Service Rates

- A multi–dimensional controlled diffusion:

$$X(t) = X(0) + W(t) + \int_0^t b(X(s), U(s))ds + \sum_{c \in \mathcal{C}} m_c \eta_c(t), \qquad X \in R^I$$

Pool–Dependent Service Rates

- A multi–dimensional controlled diffusion:

$$X(t) = X(0) + W(t) + \int_0^t b(X(s), U(s))ds + \sum_{c \in \mathcal{C}} m_c \eta_c(t), \qquad X \in R^I$$

- Can be reduced to a 1–dimensional

$$\breve{X}(t) = x_e + W_e(t) + \mu_{min} \int_0^t \breve{X}^-(s)ds - \int_0^t [\theta \cdot u(s)]\breve{X}^+(s)ds, \quad \breve{X} \in R$$

Pool–Dependent Service Rates

- A multi–dimensional controlled diffusion:

$$X(t) = X(0) + W(t) + \int_0^t b(X(s), U(s))ds + \sum_{c \in \mathcal{C}} m_c \eta_c(t), \qquad X \in R^I$$

- Can be reduced to a 1–dimensional

$$\check{X}(t) = x_e + W_e(t) + \mu_{min} \int_0^t \check{X}^-(s)ds - \int_0^t [\theta \cdot u(s)]\check{X}^+(s)ds, \quad \check{X} \in R$$

- In particular cases, asymptotically optimal control policies may be explicitly obtained.

Future direction: singular control

- Extend the existing theory to cover the controlled diffusions, arising from queues:

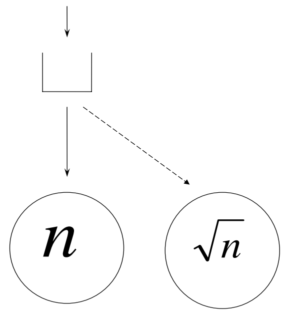
$$X(t) = X(0) + W(t) + \int_0^t b(X(s), U(s))ds + \sum_{c \in \mathcal{C}} m_c \eta_c(t)$$

Future direction: singular control

- Extend the existing theory to cover the controlled diffusions, arising from queues:

$$X(t) = X(0) + W(t) + \int_0^t b(X(s), U(s))ds + \sum_{c \in \mathcal{C}} m_c \eta_c(t)$$

- Study the models with relatively small stations, like:



$$X(t) = X(0) + W(t) + \mu_1 \int_0^t (X(s) - \Psi_2(s))^- ds - \mu_2 \int_0^t \Psi_2(s)ds$$

$$\Psi_2(t) = \Psi_2(0) - \mu_2 \int_0^t \Psi_2(s)ds + B(t), \qquad 0 \leq \Psi_2(t) \leq 1.$$

Future direction: pool–dependent service

- 2 classes, 2 stations + abandonments (θ_1 and θ_2).

  Consider the problem of minimizing linear combinations of queues:

$$V(x) = \inf_{\pi} E_x^{\pi} \int_0^{\infty} e^{-\gamma t}[c_1 Y_1(t) + c_2 Y_2(t)]dt,$$

Future direction: pool–dependent service

- 2 classes, 2 stations + abandonments (θ_1 and θ_2).

  Consider the problem of minimizing linear combinations of queues:

$$V(x) = \inf_{\pi} E_x^{\pi} \int_0^{\infty} e^{-\gamma t}[c_1 Y_1(t) + c_2 Y_2(t)]dt,$$

- The case $c_1 > c_2$ and $\theta_1 > \theta_2$ introduces an interesting trade–off:

Future direction: pool–dependent service

- 2 classes, 2 stations + abandonments (θ_1 and θ_2).

  Consider the problem of minimizing linear combinations of queues:

$$V(x) = \inf_{\pi} E_x^{\pi} \int_0^{\infty} e^{-\gamma t}[c_1 Y_1(t) + c_2 Y_2(t)]dt,$$

- The case $c_1 > c_2$ and $\theta_1 > \theta_2$ introduces an interesting trade–off:

  1. Keep class 1 in queue: high abandonment rate, but also high cost rate; or...

Future direction: pool–dependent service

- 2 classes, 2 stations + abandonments (θ_1 and θ_2).

  Consider the problem of minimizing linear combinations of queues:

$$V(x) = \inf_{\pi} E_x^{\pi} \int_0^{\infty} e^{-\gamma t}[c_1 Y_1(t) + c_2 Y_2(t)]dt,$$

- The case $c_1 > c_2$ and $\theta_1 > \theta_2$ introduces an interesting trade–off:

  1. Keep class 1 in queue: high abandonment rate, but also high cost rate; or...
  2. Keep class 2 in queue: low cost rate, but also lower abandonment rate.

Future direction: pool–dependent service

- 2 classes, 2 stations + abandonments (θ_1 and θ_2).

  Consider the problem of minimizing linear combinations of queues:

$$V(x) = \inf_{\pi} E_x^{\pi} \int_0^{\infty} e^{-\gamma t}[c_1 Y_1(t) + c_2 Y_2(t)]dt,$$

- The case $c_1 > c_2$ and $\theta_1 > \theta_2$ introduces an interesting trade–off:

  1. Keep class 1 in queue: high abandonment rate, but also high cost rate; or...
  2. Keep class 2 in queue: low cost rate, but also lower abandonment rate.

- A reduction to one–dim control problem: **bang–bang policy** (state–dependent)

  keep all the queue either in class 1, or in class 2

Future direction: pool–dependent service

- 2 classes, 2 stations + abandonments (θ_1 and θ_2).

  Consider the problem of minimizing linear combinations of queues:

$$V(x) = \inf_{\pi} E_x^{\pi} \int_0^{\infty} e^{-\gamma t}[c_1 Y_1(t) + c_2 Y_2(t)]dt,$$

- The case $c_1 > c_2$ and $\theta_1 > \theta_2$ introduces an interesting trade–off:

  1. Keep class 1 in queue: high abandonment rate, but also high cost rate; or...
  2. Keep class 2 in queue: low cost rate, but also lower abandonment rate.

- A reduction to one–dim control problem: **bang–bang policy** (state–dependent)

  keep all the queue either in class 1, or in class 2

- Explicit solution???

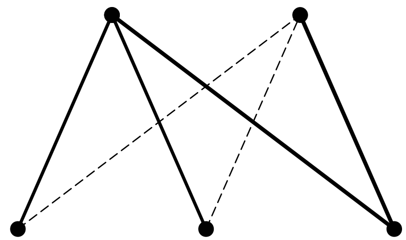
Future direction: staffing issues

- Consider 2×3 queueing system with $\lambda = \begin{pmatrix} 8 \\ 4 \end{pmatrix}, \quad \mu = \begin{pmatrix} 3 & 10 & 1 \\ 1 & 4 & 2 \end{pmatrix}.$

Future direction: staffing issues

- Consider 2×3 queueing system with $\lambda = \begin{pmatrix} 8 \\ 4 \end{pmatrix}, \quad \mu = \begin{pmatrix} 3 & 10 & 1 \\ 1 & 4 & 2 \end{pmatrix}.$

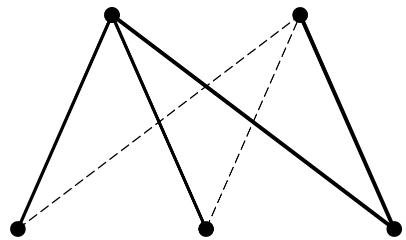
- Staffing of $\nu = (0.3, 0.3, 6.1)'$: throughput is optimal (hence, no null controllability):



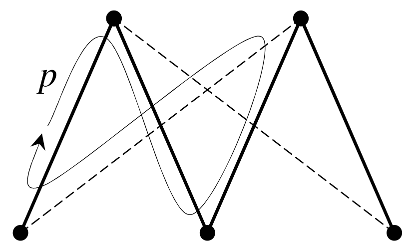
Future direction: staffing issues

- Consider 2×3 queueing system with $\lambda = \begin{pmatrix} 8 \\ 4 \end{pmatrix}, \quad \mu = \begin{pmatrix} 3 & 10 & 1 \\ 1 & 4 & 2 \end{pmatrix}.$

- Staffing of $\nu = (0.3, 0.3, 6.1)'$: throughput is optimal (hence, no null controllability):



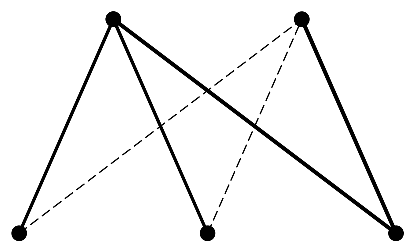
- Staffing of $\nu = (1, 1, 1)'$: throughput is not optimal (as a result, null controllability):



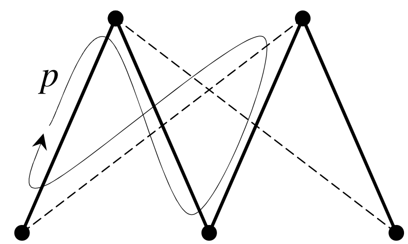
Future direction: staffing issues

- Consider 2×3 queueing system with $\lambda = \begin{pmatrix} 8 \\ 4 \end{pmatrix}, \quad \mu = \begin{pmatrix} 3 & 10 & 1 \\ 1 & 4 & 2 \end{pmatrix}.$

- Staffing of $\nu = (0.3, 0.3, 6.1)'$: throughput is optimal (hence, no null controllability):



- Staffing of $\nu = (1, 1, 1)'$: throughput is not optimal (as a result, null controllability):



- How to characterize a null controllability staffing?