13ème Atelier en Évaluation des Performances

Toulouse, Dec 2-4, 2024

# Decision-Epoch Matters: Unveiling its Impact on the Stability of Scheduling with Randomly Varying Connectivity


N. Soprano-Loto, U. Ayesta, M. Jonckheere, and I.M. Verloop

INRIA de Paris IRIT-CNRS LAAS-CNRS IRIT-CNRS

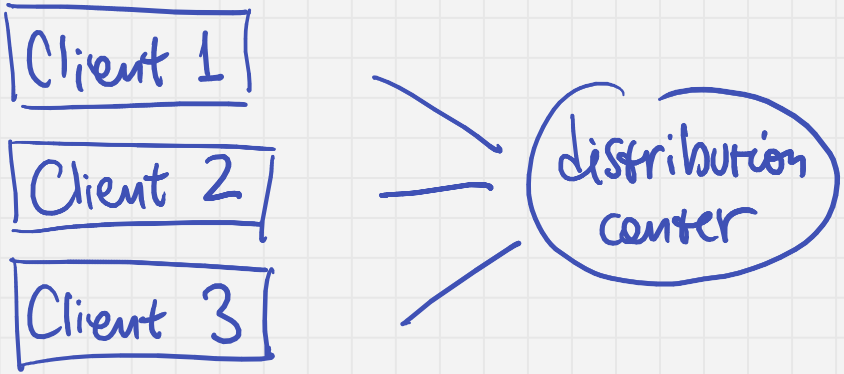
LAAS-CNRS

Application example:
supply chains

queuing systems ; random environment ; timing restrictions over decision making

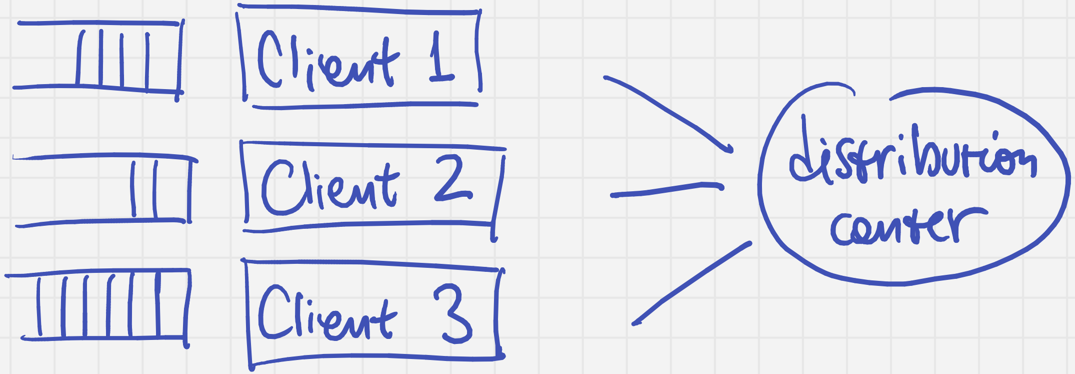
# Application example: supply chains

queuing systems ; random environment ; timing restrictions over decision making



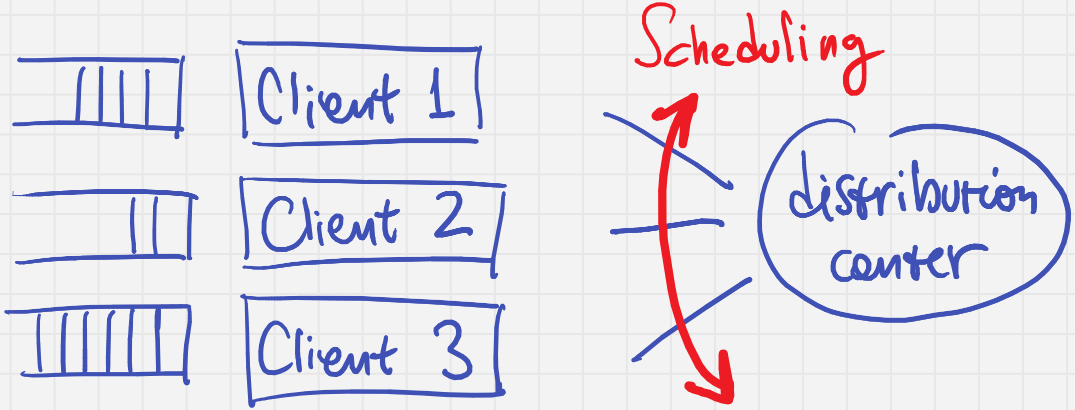
Application example: supply chains

queuing systems ; random environment ; timing restrictions over decision making



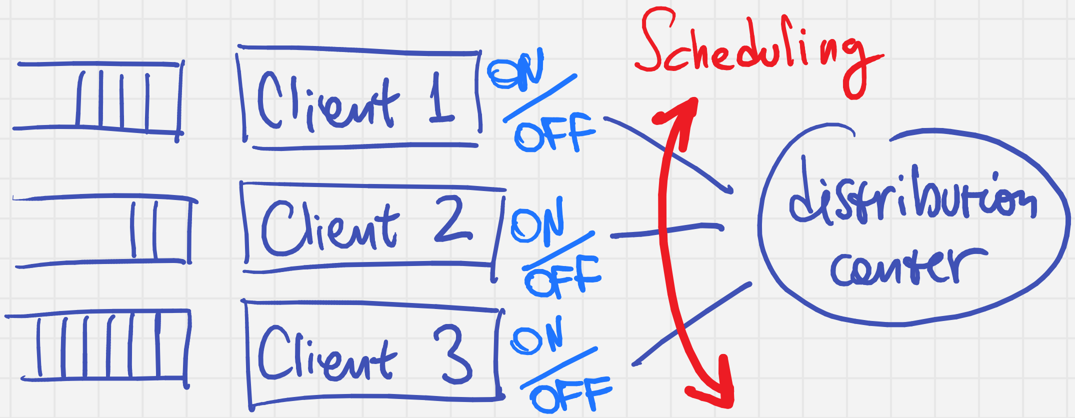
Application example: supply chains

Queuing systems ; random environment ; timing restrictions over decision making



# Application example: supply chains

queuing systems ; random environment ; timing restrictions over decision making

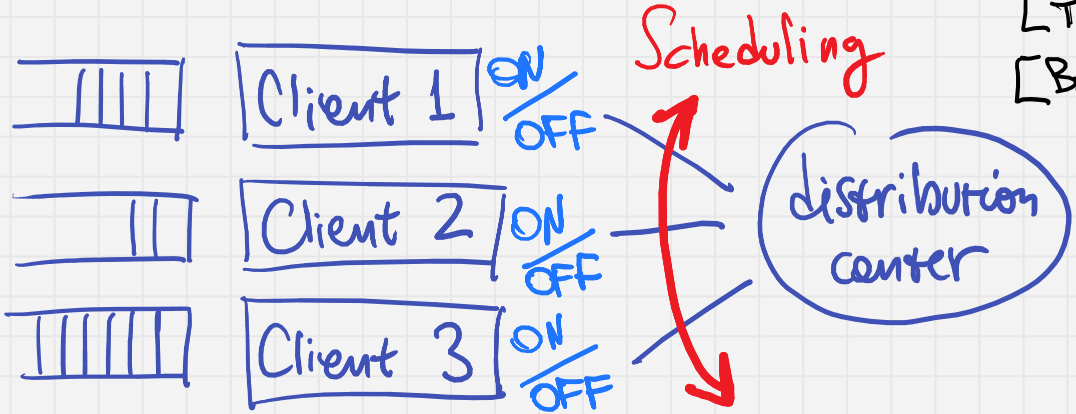


# Application example: supply chains

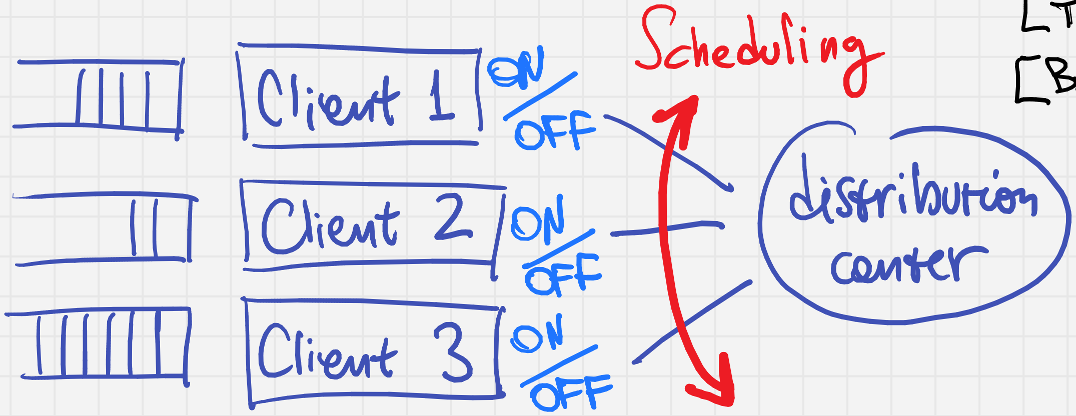
queueing systems ; random environment ; timing restrictions over decision making

[Tassiulas, Ephremides; 1993]

[Bambos, Michailidis; 2005]

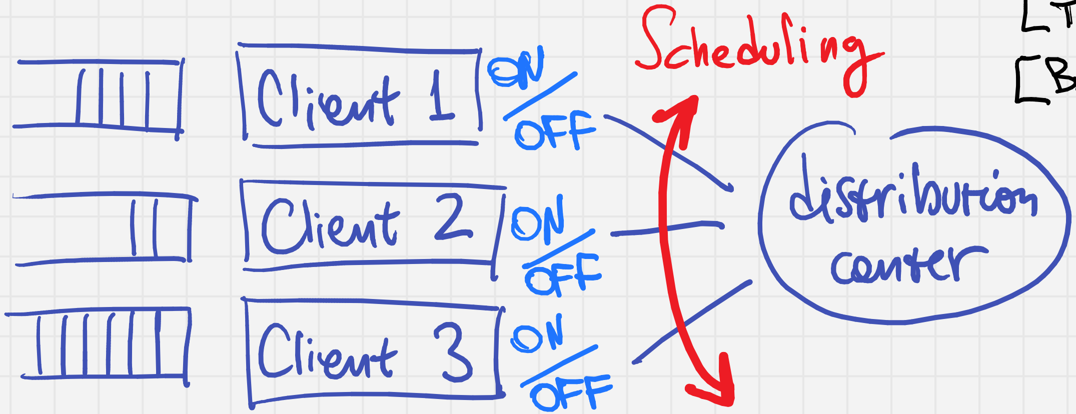


Application example: supply chains




$\underbrace{\text{queuing systems; random environment}}_{\text{[Tassiulas, Ephremides; 1993] [Bambos, Michailidis; 2005]}}$; timing restrictions over decision making

# Application example: supply chains




$\underbrace{\text{queuing systems ; random environment ;}}_{\text{[Tassiulas, Ephremides; 1993] [Bambos, Michailidis; 2005]}}$ timing restrictions over decision making

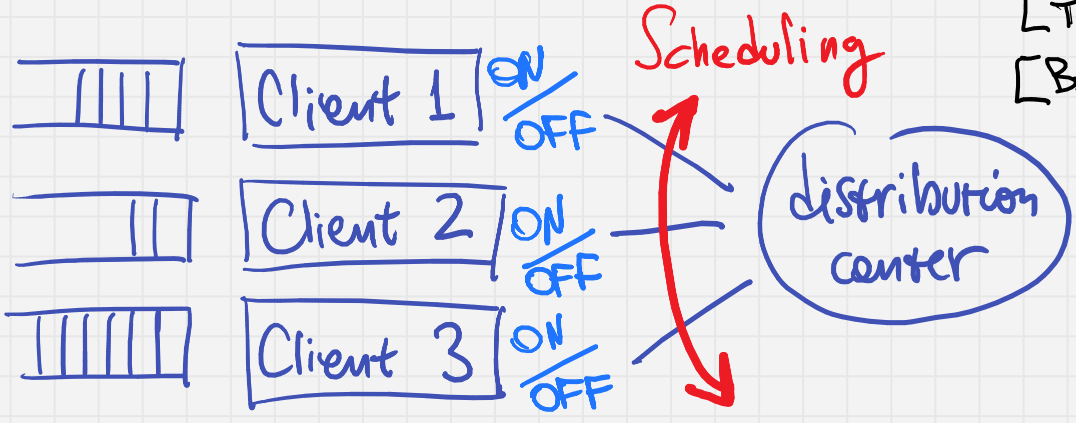
# Application example:

## supply chains

queueing systems ; random environment ; timing restrictions over decision making

[Tassiulas, Ephremides; 1993]

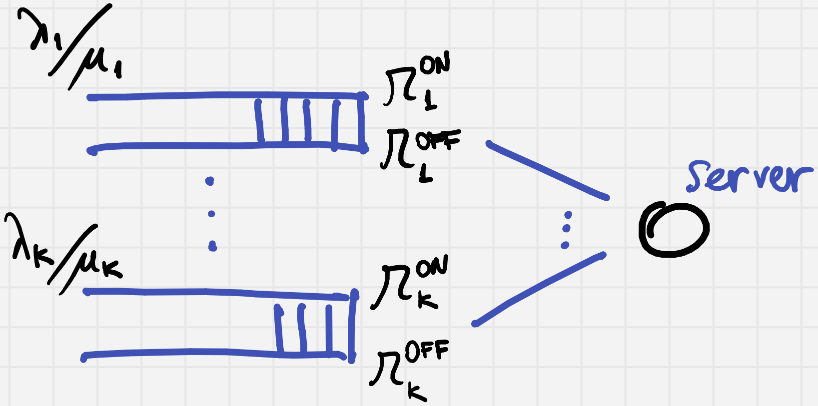
[Bambos, Michailidis; 2005]



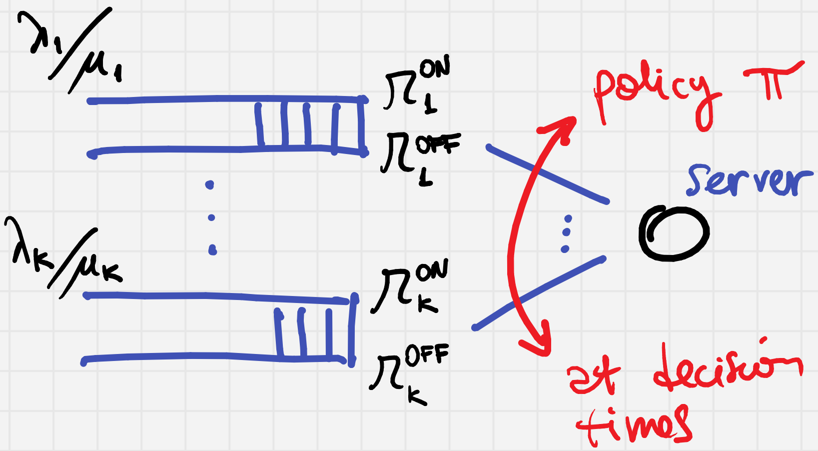

Performance guarantees?

Optimal policies?

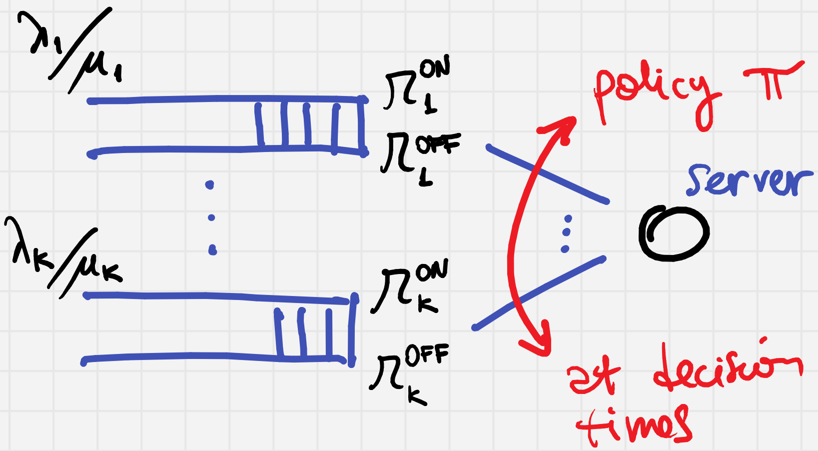
# The mathematical model



# The mathematical model



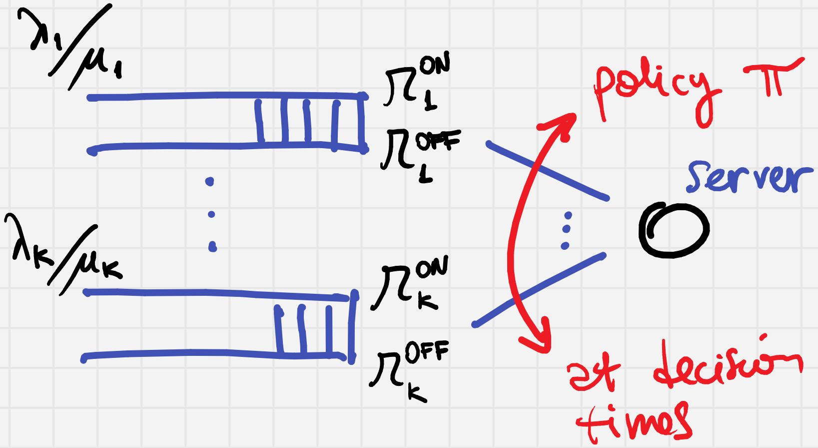
# The mathematical model




Model assumptions:

- commutative state-space
- Markovian (observable) policies
- No processor sharing

# The mathematical model




Timing restrictions:

<u>Setting 1</u>: unconstrained

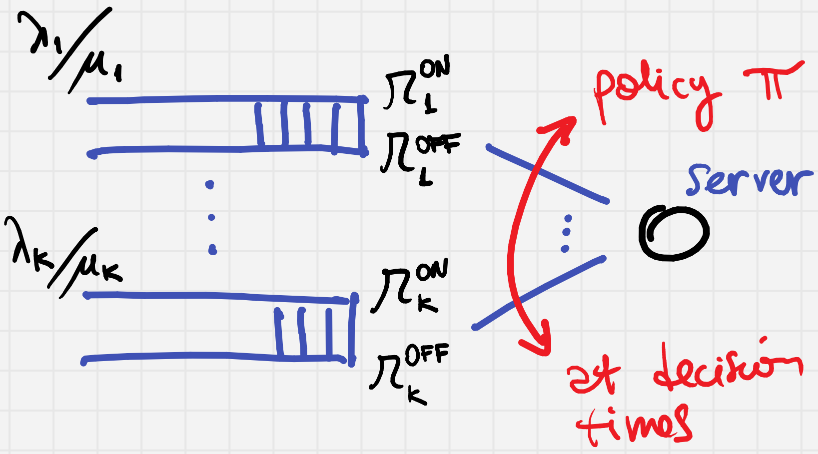
<u>Setting 2</u>: non-preemptive

<u>Setting 3</u>: $\gamma$-rate exponential

Model assumptions:

- commutative state-space
- Markovian (observable) policies
- No processor sharing

# The mathematical model




## Timing restrictions:

Setting 1: unconstrained

Setting 2: non-preemptive

Setting 3: $\gamma$-rate exponentials

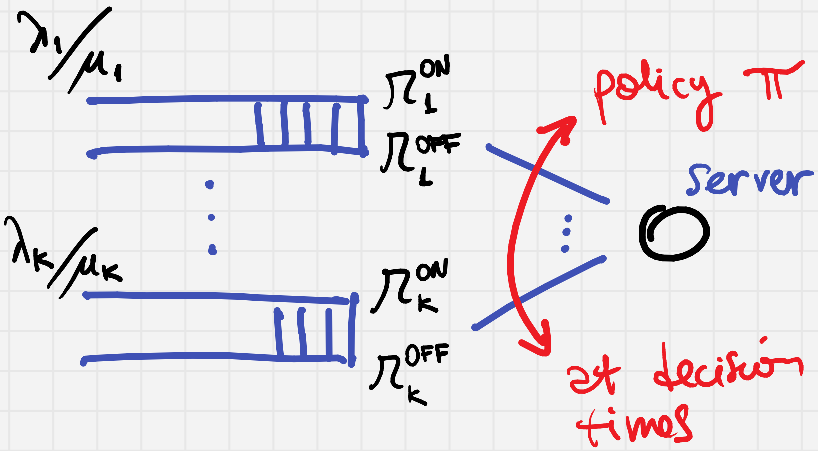
Setting $j$, policy $\pi$ $\rightarrow$ Markov process $\rightarrow$ positive recurrent (stable)

Markov process $\rightarrow$ ~ positive recurrent

## Model assumptions:

- commutative state-space
- Markovian (observable) policies
- No processor sharing

# The mathematical model




## Model assumptions:

- commutative state-space
- Markovian (observable) policies
- No processor sharing

## Timing restrictions:

**Setting 1**: unconstrained

**Setting 2**: non-preemptive

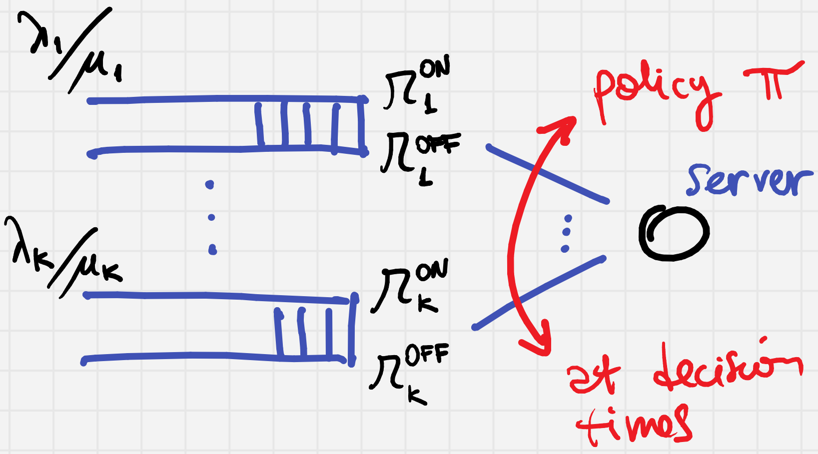
**Setting 3**: $\gamma$-rate exponential

$\left.\begin{matrix}\text{Setting } j \\ \text{policy } \pi\end{matrix}\right\} \rightarrow$ Markov process $\rightarrow$ positive recurrent (stable)

$\rightarrow$ $\sim$ positive recurrent

$$SR_j(\pi) = \left\{ (\underline{\lambda}, \underline{\mu}, \underline{r}^{ON}, \underline{r}^{OFF}) : \pi \text{ is stable} \right\}$$

# The mathematical model




## Model assumptions:

- commutative state-space
- Markovian (observable) policies
- No processor sharing

## Timing restrictions:

__Setting 1__: unconstrained

__Setting 2__: non-preemptive

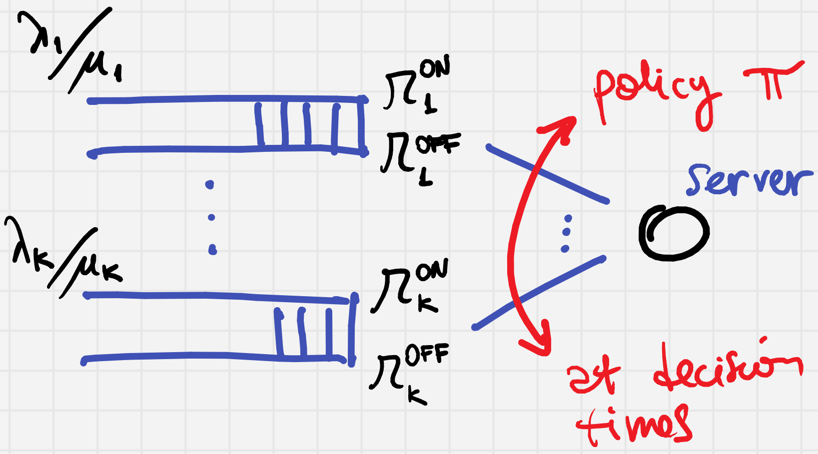
__Setting 3__: $\gamma$-rate exponentially

$$\left.\begin{array}{l}\text{Setting } j\\ \text{policy } \pi\end{array}\right\} \rightarrow \text{Markov process} \begin{array}{l}\nearrow \text{positive recurrent (stable)} \\ \searrow \sim \text{positive recurrent}\end{array}$$

$$SR_j(\pi) = \left\{ (\underline{\lambda}, \underline{\mu}, \underline{r}^{ON}, \underline{r}^{OFF}) : \pi \text{ is stable} \right\}$$

$$MSR_j = \bigcup_{\pi} SR_j(\pi)$$

# The mathematical model




## Model assumptions:

- commutative state-space
- Markovian (observable) policies
- No processor sharing

## Timing restrictions:

**Setting 1**: unconstrained

**Setting 2**: non-preemptive

**Setting 3**: $\gamma$-rate exponential

$\left.\begin{array}{l}\text{Setting } j\\ \text{policy } \pi\end{array}\right\} \rightarrow$ Markov process $\rightarrow$ positive recurrent (stable) / $\sim$ positive recurrent

$$SR_j(\pi) = \left\{ (\underline{\lambda}, \underline{\mu}, \underline{r}^{ON}, \underline{r}^{OFF}) : \pi \text{ is stable} \right\}$$

$$MSR_j = \bigcup_{\pi} SR_j(\pi)$$

$\pi^*$ is throughput optimal if

$$SR_j(\pi^*) = MSR_j$$

# Main theorem

$\rho_i = \dfrac{\lambda_i}{\mu_i}$ , $p_i^{ON} = \dfrac{\%\text{ time ON}}{\text{queue } i} = \dfrac{R_i^{ON}}{R_i^{ON} + R_i^{OFF}} = 1 - p_i^{OFF}$

# Main theorem

$\rho_i = \frac{\lambda_i}{\mu_i}$ , $p_i^{ON} = \frac{\%\text{ time ON}}{\text{queue } i} = \frac{r_i^{ON}}{r_i^{ON} + r_i^{OFF}} = 1 - p_i^{OFF}$

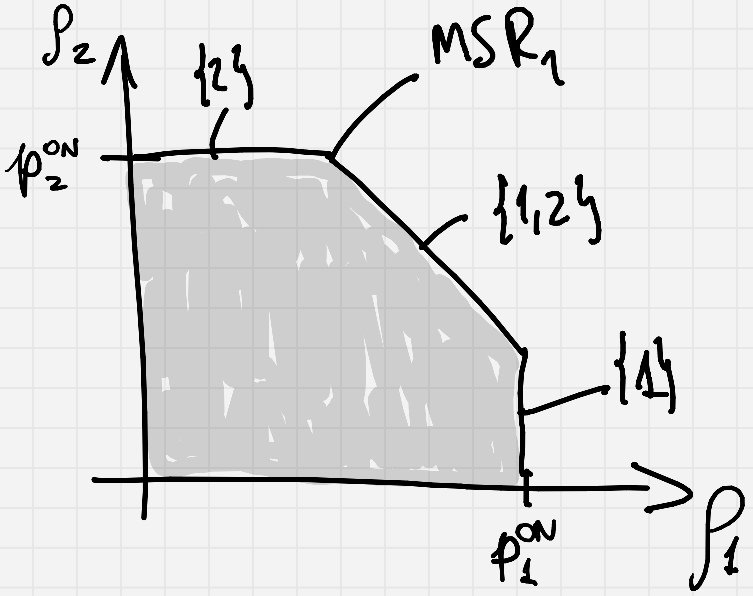
**MSR₁** (unconstrained): $\forall L \subseteq \{1, \dots, K\}$, $\underbrace{\sum_{i \in L} \rho_i}_{\text{load at } L} < \underbrace{1 - \prod_{i \in L} p_i^{OFF}}_{\text{maximum capacity}}$

# Main theorem

$\rho_i = \dfrac{\lambda_i}{\mu_i}$, $p_i^{ON} = \%$ time ON queue $i$ $= \dfrac{r_i^{ON}}{r_i^{ON} + r_i^{OFF}} = 1 - p_i^{OFF}$

$$\text{MSR}_1: \quad \forall L \subseteq \{1, \dots, K\}, \quad \underbrace{\sum_{i \in L} \rho_i}_{\text{load at } L} < \underbrace{1 - \prod_{i \in L} p_i^{OFF}}_{\text{maximum capacity}}$$

(unconstrained)



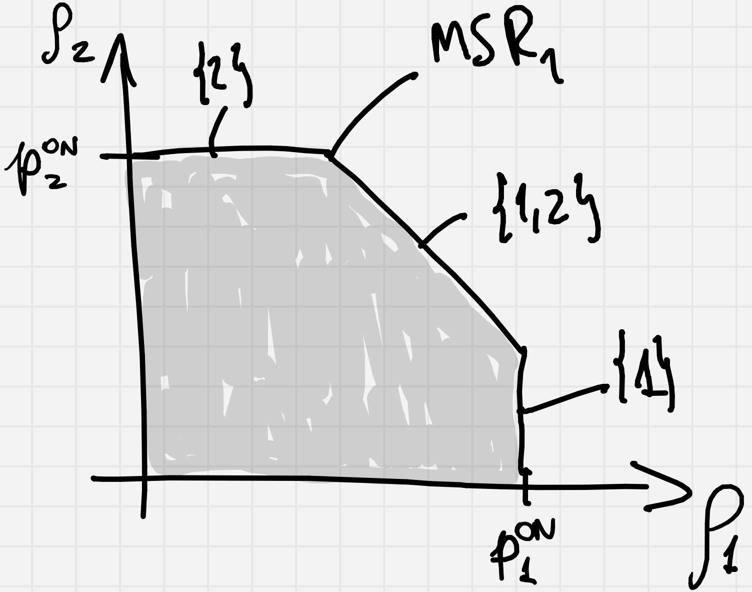
# Main theorem

$\rho_i = \frac{\lambda_i}{\mu_i}$ , $p_i^{ON} = \%$ time ON queue $i$ $= \frac{r_i^{ON}}{r_i^{ON} + r_i^{OFF}} = 1 - p_i^{OFF}$

$MSR_1$ : $\forall L \subseteq \{1, \dots, K\}, \underbrace{\sum_{i \in L} \rho_i}_{\text{load at } L} < \underbrace{1 - \prod_{i \in L} p_i^{OFF}}_{\text{maximum capacity}}$

(unconstrained)

↖ not new



# Main theorem

$\rho_i = \frac{\lambda_i}{\mu_i}$ , $p_i^{ON} = \%$ time ON queue $i$ $= \frac{r_i^{ON}}{r_i^{ON}+r_i^{OFF}} = 1 - p_i^{OFF}$

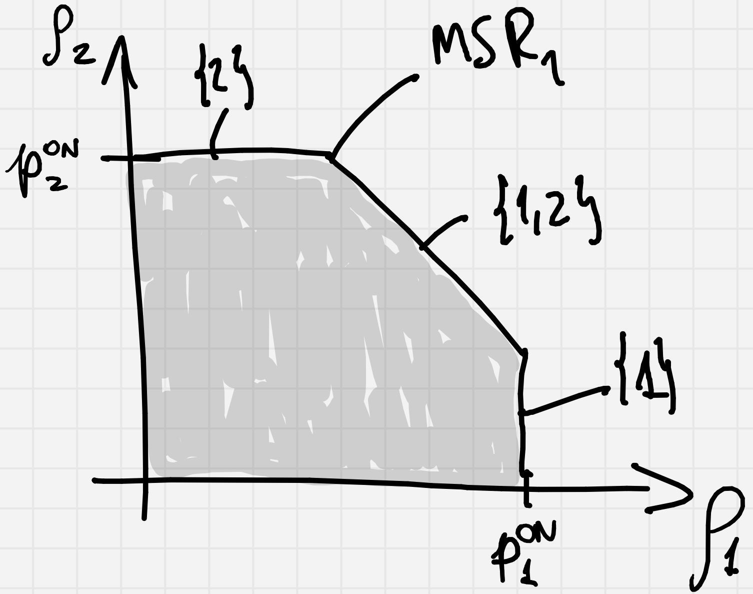
$MSR_1$ : $\forall L \subseteq \{1,\dots,K\}$, $\underbrace{\sum_{i\in L} \rho_i}_{\text{load at } L} < \underbrace{1 - \prod_{i\in L} p_i^{OFF}}_{\text{maximum capacity}}$

(unconstrained)

← not new

$MSR_2$ : $\sum_{i=1}^{K} \frac{\rho_i}{p_i^{ON}} < 1$

(non-preemptive)



# Main theorem

$\rho_i = \frac{\lambda_i}{\mu_i}$ , $p_i^{ON}$ = % time queue $i$ ON $= \frac{r_i^{ON}}{r_i^{ON} + r_i^{OFF}} = 1 - p_i^{OFF}$

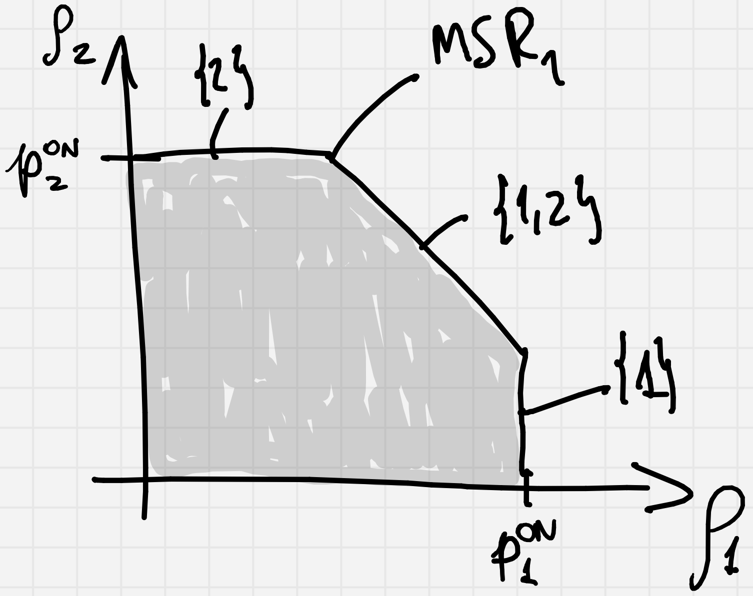
$MSR_1$ : $\forall L \subseteq \{1, \dots, K\}$, $\underbrace{\sum_{i \in L} \rho_i}_{\text{load at } L} < \underbrace{1 - \prod_{i \in L} p_i^{OFF}}_{\text{maximum capacity}}$

(unconstrained)

← not new

$MSR_2$ : $\sum_{i=1}^{K} \frac{\rho_i}{p_i^{ON}} < 1$ ← multi-class M/G/1

(non-preemptive)



# Main theorem

$\rho_i = \frac{\lambda_i}{\mu_i}$ , $p_i^{ON} = \%$ time ON queue $i$ $= \frac{\Omega_i^{ON}}{\Omega_i^{ON}+\Omega_i^{OFF}} = 1 - p_i^{OFF}$

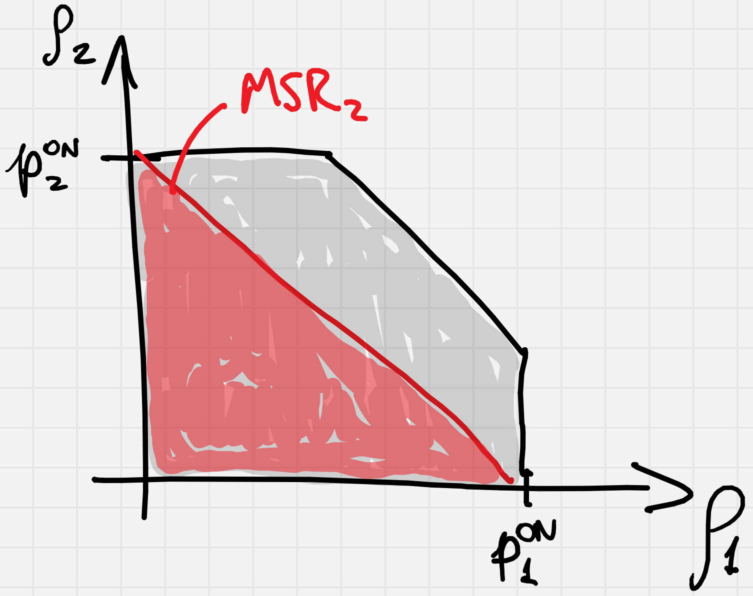
$MSR_1$ : $\forall L \subseteq \{1,\dots,K\}, \underbrace{\sum_{i\in L} \rho_i}_{\text{load at } L} < \underbrace{1 - \prod_{i\in L} p_i^{OFF}}_{\text{maximum capacity}}$

(unconstrained)

← not new

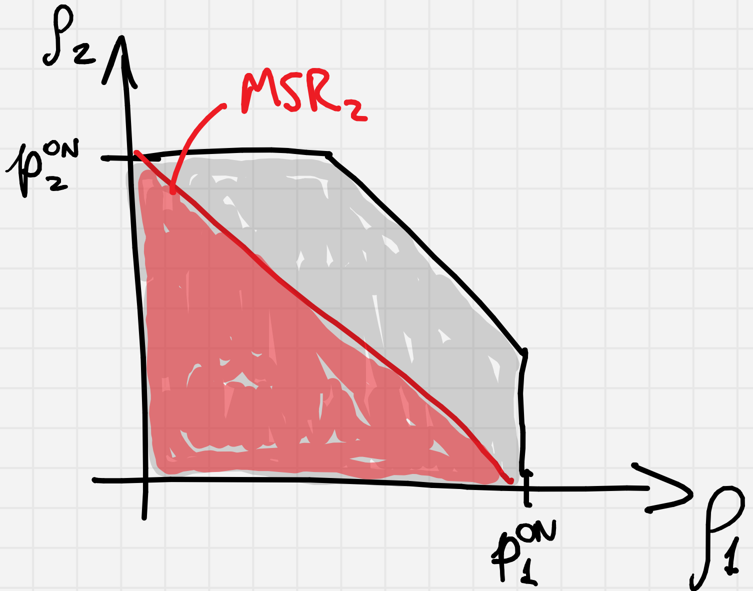
$MSR_2$ : $\sum_{i=1}^{K} \frac{\rho_i}{p_i^{ON}} < 1$ ← multi-class M/G/1

(non-preemptive)



# Main theorem

$\rho_i = \frac{\lambda_i}{\mu_i}$ , $p_i^{ON} = \%$ time queue $i$ ON $= \frac{r_i^{ON}}{r_i^{ON}+r_i^{OFF}} = 1 - p_i^{OFF}$




MSR$_1$ : $\forall L \subseteq \{1,\dots,K\}$, $\underbrace{\sum_{i\in L}\rho_i}_{\text{load at } L} < \underbrace{1-\prod_{i\in L} p_i^{OFF}}_{\text{maximum capacity}}$

(unconstrained)

← not new

MSR$_2$ : $\sum_{i=1}^{K} \frac{\rho_i}{p_i^{ON}} < 1$ ← multi-class M/G/1

(non-preemptive)

MSR$_3(\gamma)$ — CONNECTED POLICIES : $\forall L \subseteq \{1,\dots,K\}$, $\sum_{i\in L} \frac{\rho_i}{\theta_i(\gamma)} < 1-\prod_{i\in L} p_i^{OFF}$

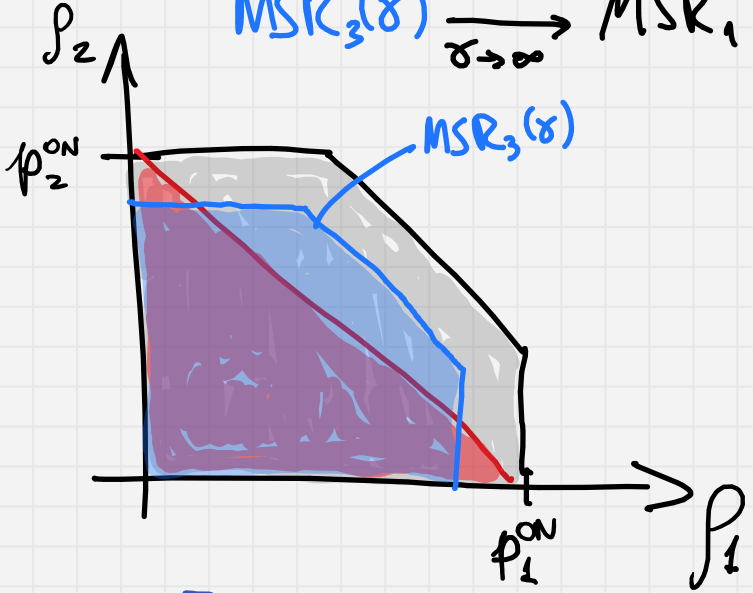
($\gamma$-exponentials)

$\theta_i(\gamma)$: $\text{OFF} \underset{r_i^{OFF}}{\overset{\gamma + r_i^{ON}}{\rightleftarrows}} \text{ON}$

# Main theorem

$\rho_i = \frac{\lambda_i}{\mu_i}$, $p_i^{ON} = \%$ time queue $i$ ON $= \frac{\Omega_i^{ON}}{\Omega_i^{ON} + \Omega_i^{OFF}} = 1 - p_i^{OFF}$

$MSR_3(\gamma) \xrightarrow[\gamma \to \infty]{} MSR_1$




**MSR$_1$** (unconstrained): $\forall L \subseteq \{1, \dots, K\}$, $\underbrace{\sum_{i \in L} \rho_i}_{\text{load at } L} < \underbrace{1 - \prod_{i \in L} p_i^{OFF}}_{\text{maximum capacity}}$

← not new

**MSR$_2$** (non-preemptive): $\sum_{i=1}^{K} \frac{\rho_i}{p_i^{ON}} < 1$ ← multi-class M/G/1

**MSR$_3(\gamma)$** ($\gamma$-exponentials) — CONNECTED POLICIES: $\forall L \subseteq \{1, \dots, K\}$, $\sum_{i \in L} \frac{\rho_i}{\theta_i(\gamma)} < 1 - \prod_{i \in L} p_i^{OFF}$

$\theta_i(\gamma)$: $\text{OFF} \underset{\Omega_i^{OFF}}{\overset{\gamma + \Omega_i^{ON}}{\rightleftarrows}} \text{ON}$

# Main theorem

$\rho_i = \frac{\lambda_i}{\mu_i}$, $p_i^{ON} = \%$ time queue $i$ ON $= \frac{r_i^{ON}}{r_i^{ON} + r_i^{OFF}} = 1 - p_i^{OFF}$

**$MSR_1$:** $\forall L \subseteq \{1, \dots, K\}$, $\underbrace{\sum_{i \in L} \rho_i}_{\text{load at } L} < \underbrace{1 - \prod_{i \in L} p_i^{OFF}}_{\text{maximum capacity}}$

(unconstrained) ← not new

**$MSR_2$:** $\sum_{i=1}^{K} \frac{\rho_i}{p_i^{ON}} < 1$ ← multi-class M/G/1

(non-preemptive)

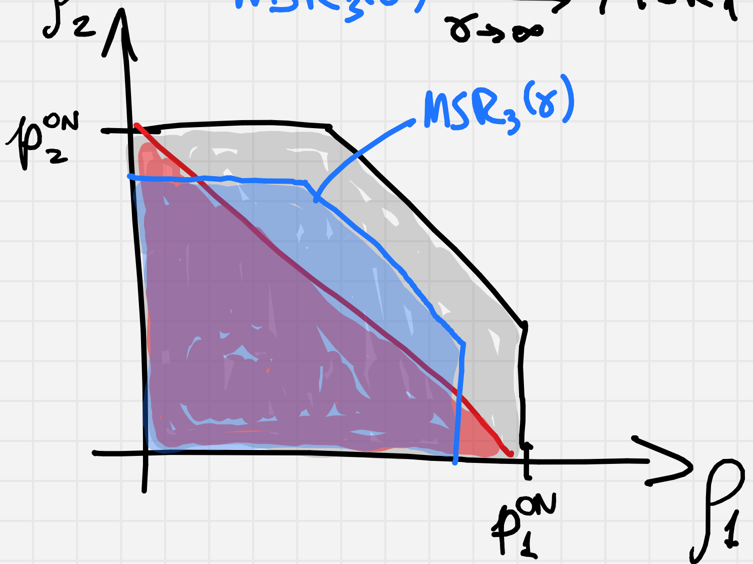
**$MSR_3(\gamma)$** — CONNECTED POLICIES: $\forall L \subseteq \{1, \dots, K\}$, $\sum_{i \in L} \frac{\rho_i}{\theta_i(\gamma)} < 1 - \prod_{i \in L} p_i^{OFF}$

($\gamma$-exponentials)

$\theta_i(\gamma)$: $\text{OFF} \underset{r_i^{OFF}}{\overset{\gamma + r_i^{ON}}{\rightleftarrows}} \text{ON}$

$\forall c_1, \dots, c_K > 0$, $\arg\max_{i \in \{1, \dots, K\}} c_i Q_i \mathbb{1}\{E_i = ON\}$ is optimal.

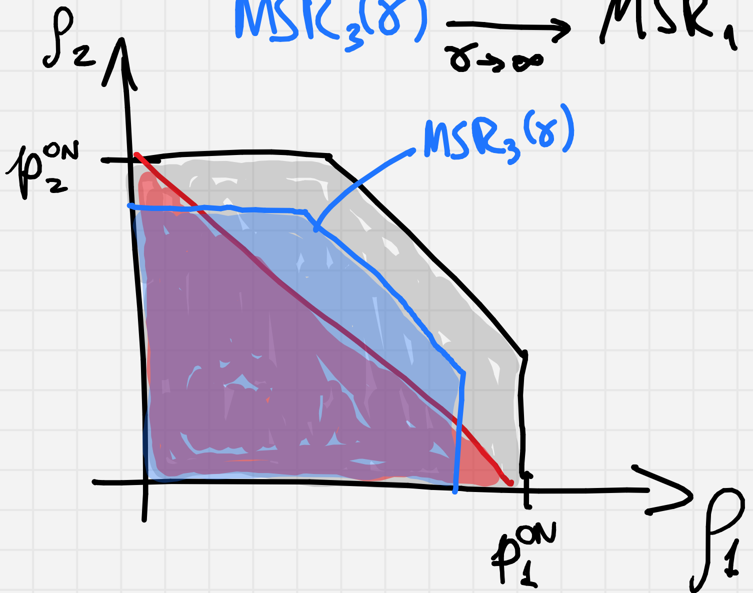
$MSR_3(\gamma) \xrightarrow[\gamma \to \infty]{} MSR_1$



# Main theorem

$\rho_i = \frac{\lambda_i}{\mu_i}$ , $p_i^{ON} = $ % time ON queue $i$ $= \frac{R_i^{ON}}{R_i^{ON}+R_i^{OFF}} = 1 - p_i^{OFF}$

$MSR_3(\gamma) \xrightarrow[\gamma\to\infty]{} MSR_1$




**MSR$_1$** : $\forall L \subseteq \{1,\dots,K\}$, $\underbrace{\sum_{i\in L} \rho_i}_{\text{load at } L} < \underbrace{1 - \prod_{i\in L} p_i^{OFF}}_{\text{maximum capacity}}$

(unconstrained) ← not new

**MSR$_2$** : $\sum_{i=1}^{K} \frac{\rho_i}{p_i^{ON}} < 1$ ← multi-class M/G/1

(non-preemptive)

**MSR$_3(\gamma)$** — CONNECTED POLICIES : $\forall L \subseteq \{1,\dots,K\}$, $\sum_{i\in L} \frac{\rho_i}{\theta_i(\gamma)} < 1 - \prod_{i\in L} p_i^{OFF}$

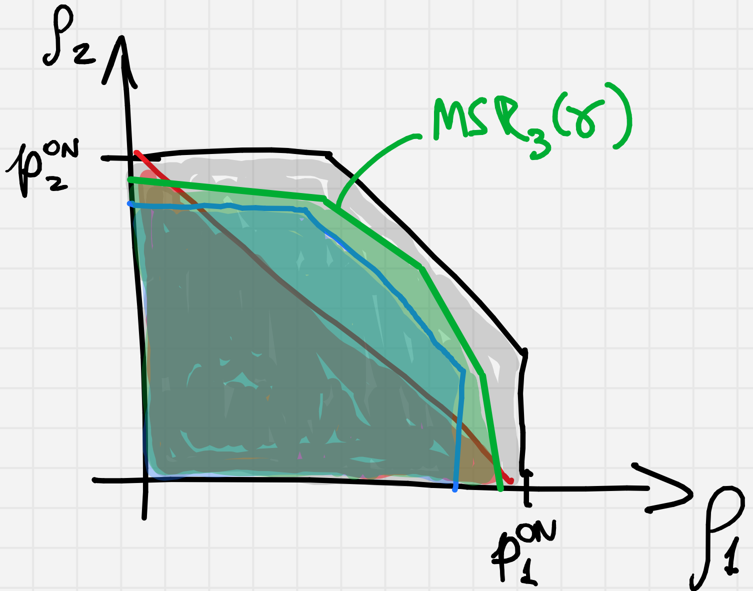
($\gamma$-exponentials)

$\theta_i(\gamma)$: OFF $\xrightarrow{\gamma + R_i^{ON}}$ ON, ON $\xrightarrow{R_i^{OFF}}$ OFF

$\forall c_1,\dots,c_K > 0$, $\arg\max_{i\in\{1,\dots,K\}} c_i Q_i \mathbb{1}\{E_i = ON\}$ is optimal. In particular, SLCQ is optimal.

# Main theorem

$\rho_i = \frac{\lambda_i}{\mu_i}$, $p_i^{ON} = \%$ time ON queue $i$ $= \frac{r_i^{ON}}{r_i^{ON} + r_i^{OFF}} = 1 - p_i^{OFF}$




**$MSR_1$:** $\forall L \subseteq \{1, \dots, K\}$, $\underbrace{\sum_{i \in L} \rho_i}_{\text{load at } L} < \underbrace{1 - \prod_{i \in L} p_i^{OFF}}_{\text{maximum capacity}}$

(unconstrained)

← not new

**$MSR_2$:** $\sum_{i=1}^{K} \frac{\rho_i}{p_i^{ON}} < 1$ ← multi-class M/G/1

(non-preemptive)

**$MSR_3(\gamma)$** — CONNECTED POLICIES : $\forall L \subseteq \{1, \dots, K\}$, $\sum_{i \in L} \frac{\rho_i}{\theta_i(\gamma)} < 1 - \prod_{i \in L} p_i^{OFF}$

($\gamma$-exponentials)

$\theta_i(\gamma)$: OFF $\underset{r_i^{OFF}}{\overset{\gamma + r_i^{ON}}{\rightleftarrows}}$ ON

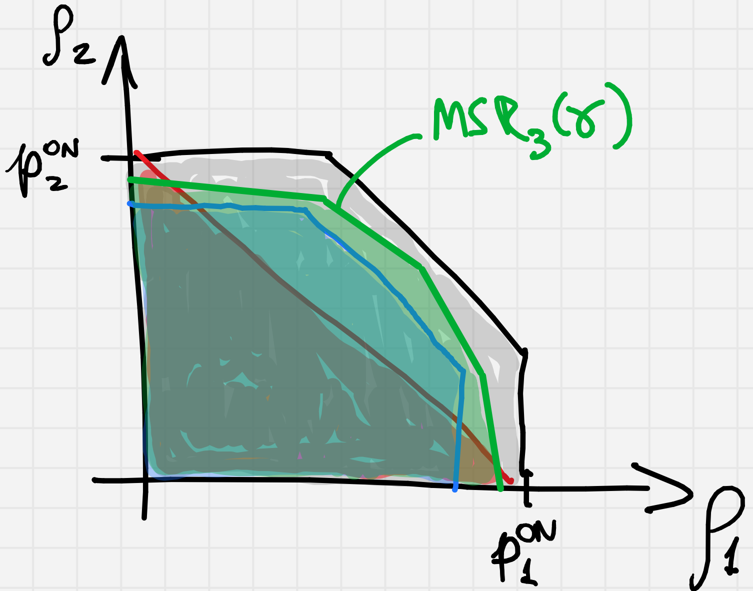
$\forall c_1, \dots, c_K > 0$, $\arg\max_{i \in \{1, \dots, K\}} c_i Q_i \mathbb{1}\{E_i = ON\}$ is optimal. In particular, SLCQ is optimal.

**$MSR_3(\gamma)$** — NON-CONNECTED POLICIES : Minkowski sum of triangles, optimal policies : $\arg\max_{i \in \{1, \dots, K\}} c_i Q_i \left( \mathbb{1}\{E_i = ON\} + \underbrace{K_i}_{\in (0,1)} \mathbb{1}\{E_i = OFF\} \right)$

[Andrews, Kumaran, Ramanan, Stolyar, Vijayakumar, Whiting; 2004]

# Main theorem

$\rho_i = \frac{\lambda_i}{\mu_i}$ , $p_i^{ON} = \%$ time ON queue $i$ $= \frac{r_i^{ON}}{r_i^{ON} + r_i^{OFF}} = 1 - p_i^{OFF}$




**$MSR_1$** (unconstrained) : $\forall L \subseteq \{1,\dots,K\}$, $\underbrace{\sum_{i \in L} \rho_i}_{\text{load at } L} < \underbrace{1 - \prod_{i \in L} p_i^{OFF}}_{\text{maximum capacity}}$

← not new

**$MSR_2$** (non-preemptive) : $\sum_{i=1}^{K} \frac{\rho_i}{p_i^{ON}} < 1$ ← multi-class M/G/1

**$MSR_3(\gamma)$** ($\gamma$-exponentials) — CONNECTED POLICIES : $\forall L \subseteq \{1,\dots,K\}$, $\sum_{i \in L} \frac{\rho_i}{\theta_i(\gamma)} < 1 - \prod_{i \in L} p_i^{OFF}$

$\theta_i(\gamma)$: $\text{OFF} \underset{r_i^{OFF}}{\overset{\gamma + r_i^{ON}}{\rightleftarrows}} \text{ON}$

$\forall c_1, \dots, c_K > 0$, $\arg\max_{i \in \{1,\dots,K\}} c_i Q_i \mathbb{1}\{E_i = ON\}$ is optimal. In particular, SLCQ is optimal.

**$MSR_3(\gamma)$** — NON-CONNECTED POLICIES : Minkowski sum of triangles , optimal policies : $\arg\max_{i \in \{1,\dots,K\}} c_i Q_i \left( \mathbb{1}\{E_i = ON\} + \underbrace{K_i}_{\in(0,1)} \mathbb{1}\{E_i = OFF\} \right)$

Obs: they're all agnostic policies.

[Andrews, Kumaran, Ramanan, Stolyar, Vijayakumar, Whiting; 2004]

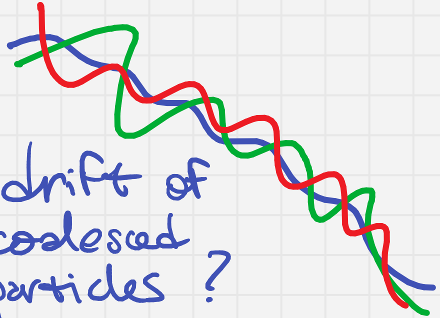
## Proof: test for fluid limits (TFL)

Classical result (Dai, Rybko-Stolyar): stable Fluid model $\Longrightarrow$ stable stochastic model

# Proof: test for fluid limits (TFL)

Classical result (Dai, Rybko-Stolyar): Stable Fluid model $\Longrightarrow$ Stable stochastic model

drift of coalesced particles ?



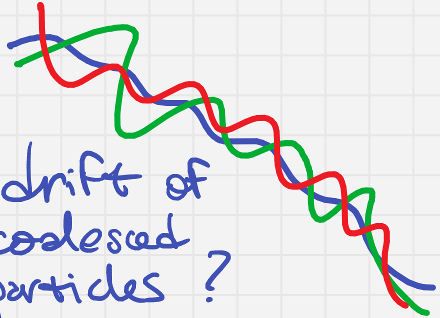
# Proof: test for fluid limits (TFL)

Classical result (Dai, Rybko-Stolyar): Stable fluid model $\Longrightarrow$ Stable stochastic model

$\Uparrow$

Maximum queue length

fluid Lyapunov



drift of coalesced particles ?

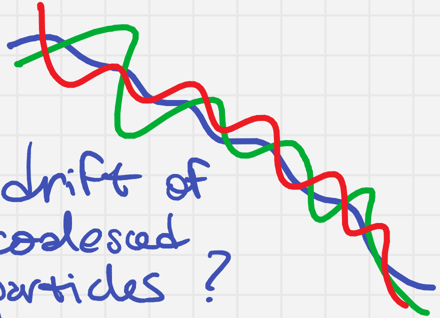
# Proof: test for fluid limits (TFL)

Classical result (Dai, Rybko-Stolyar): Stable Fluid model $\Longrightarrow$ Stable stochastic model

$\Uparrow$

Maximum queue length
Fluid Lyapunov $\Longleftarrow$ $\dfrac{M(t_2)-M(t_1)}{t_2-t_1} \le -\delta$

$\forall\, t_1 < t_2$



drift of coalesced particles ?

# Proof: test for fluid limits (TFL)

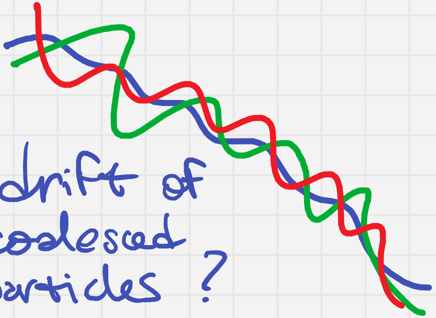
Classical result (Dai, Rybko-Stolyar): Stable Fluid model $\Longrightarrow$ Stable stochastic model

$\Uparrow$

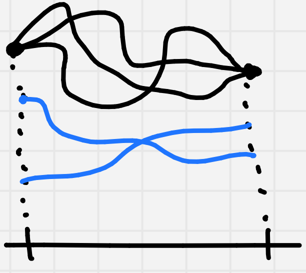
Maximum queue length Fluid Lyapunov $\Longleftarrow$ $\cancel{\forall\, t_1 < t_2}$ $\dfrac{M(t_2)-M(t_1)}{t_2-t_1} \le -\delta$

TFL

For $t_1 < t_2$ s.t.



drift of coalesced particles ?



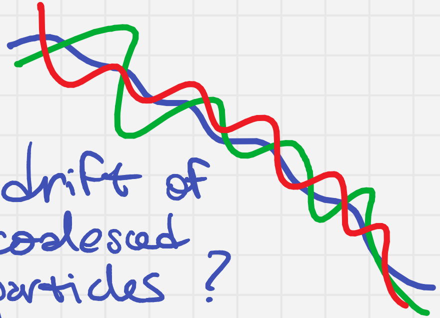
$t_1$ $t_2$

# Proof: test for fluid limits (TFL)

Classical result (Dai, Rybko-Stolyar): Stable Fluid model $\Longrightarrow$ Stable stochastic model

$\Uparrow$

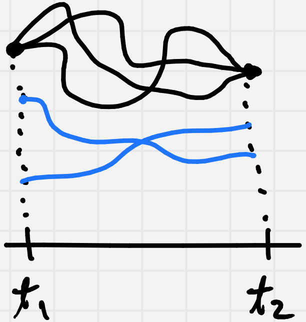
Maximum queue length fluid Lyapunov



drift of coalesced particles?

$\Longleftarrow$ ~~$\forall\, t_1 < t_2$~~ **TFL** For $t_1 < t_2$ s.t.

$$\frac{M(t_2)-M(t_1)}{t_2-t_1} \leq -\delta$$




Under the formal assumption, the control of the slope follows easily!

# Conclusions

- In a queuing network with a random environment, we studied how timing restrictions on decision making affect the stability of the system.
  More precisely, for different timing restriction settings,
  - we explicitly described the MSRs,
  - and we provided a family of MW optimal policies.
- We developed a formal test for fluid limits (TFL) which, if passed, it guarantees the stability of the model.

Thanks !