Heavy-Traffic Universality of Redundancy Systems with Data Locality Constraints

YEQT 2021

Ellen Cardinaels, Sem Borst, Johan van Leeuwaarden

Heavy-Traffic Universality

of Redundancy Systems

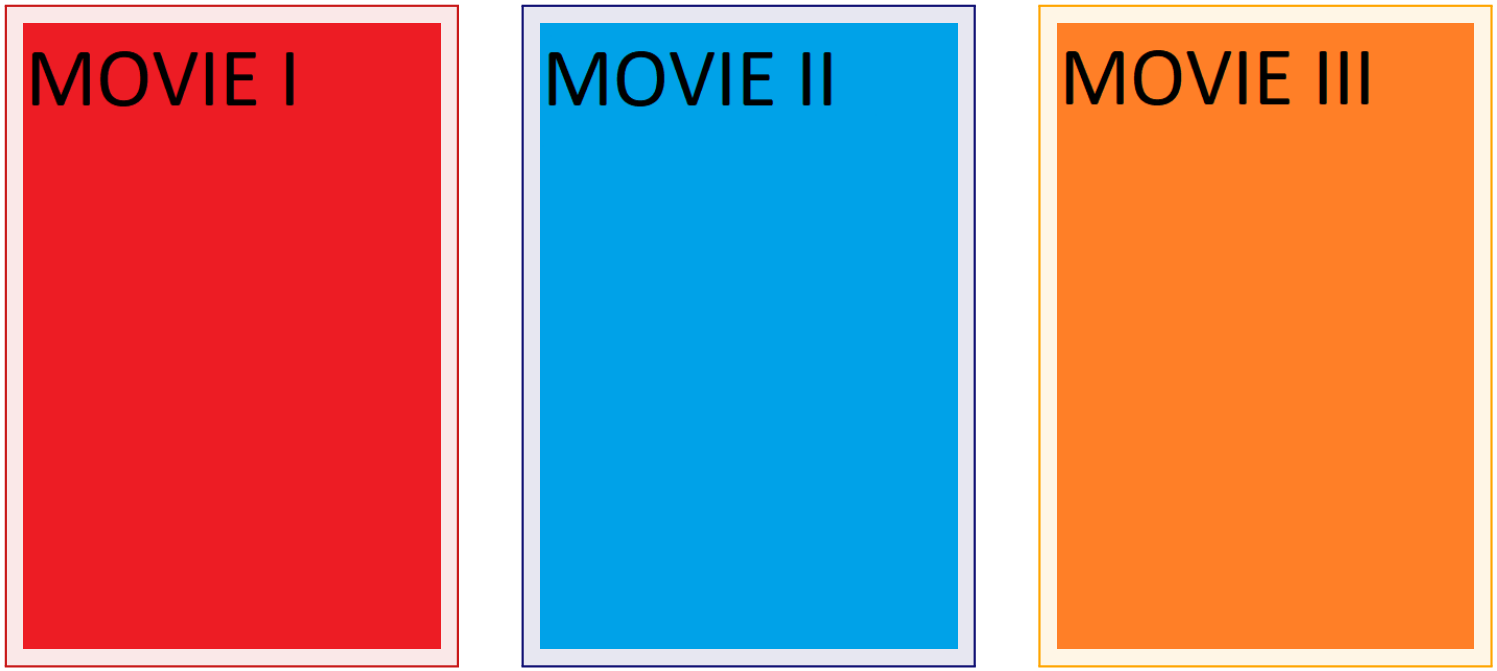
with Data Locality Constraints

Heavy-Traffic Universality (result)

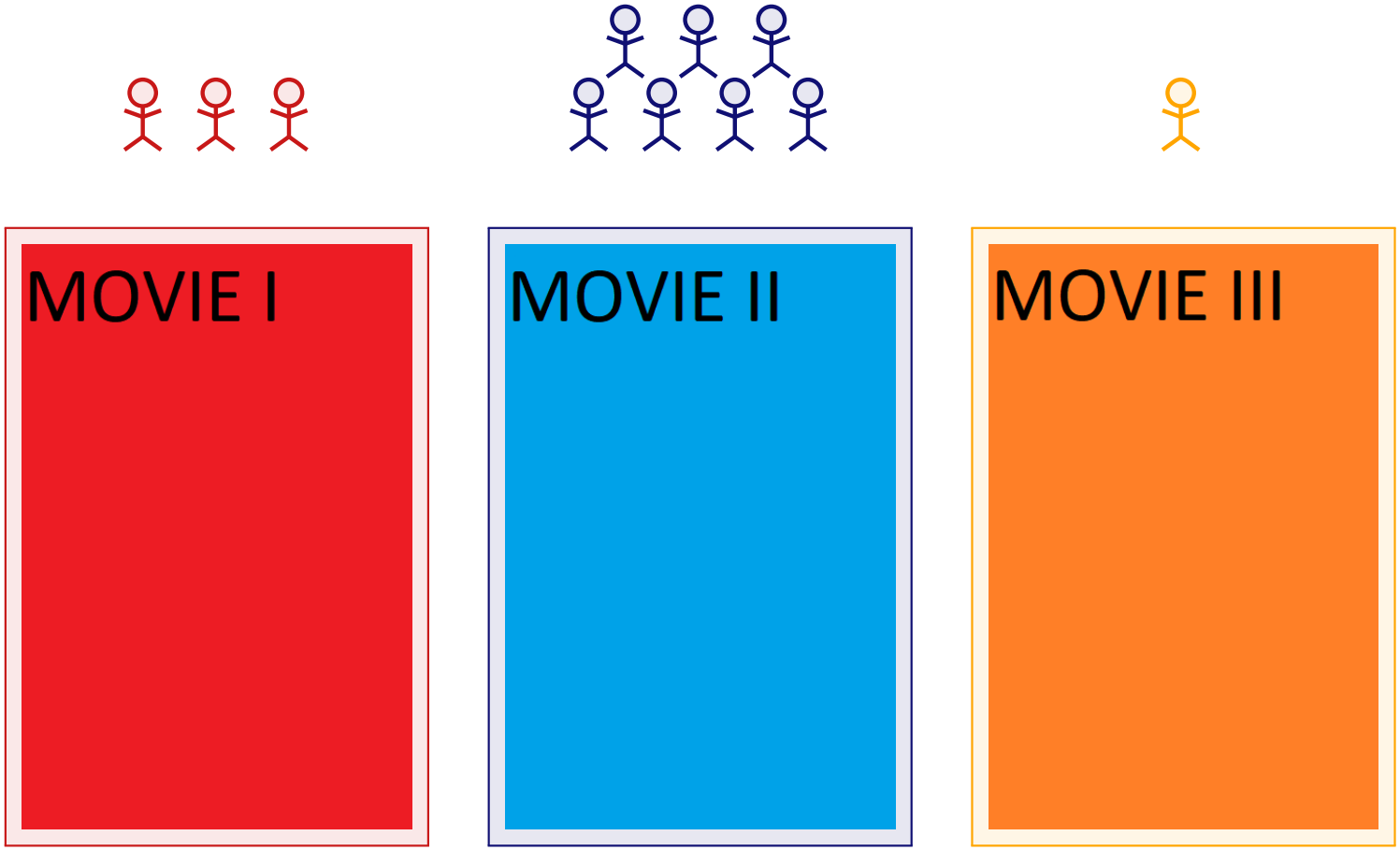
of Redundancy Systems (tool: product-form expressions)

with Data Locality Constraints (motivation)

Motivation



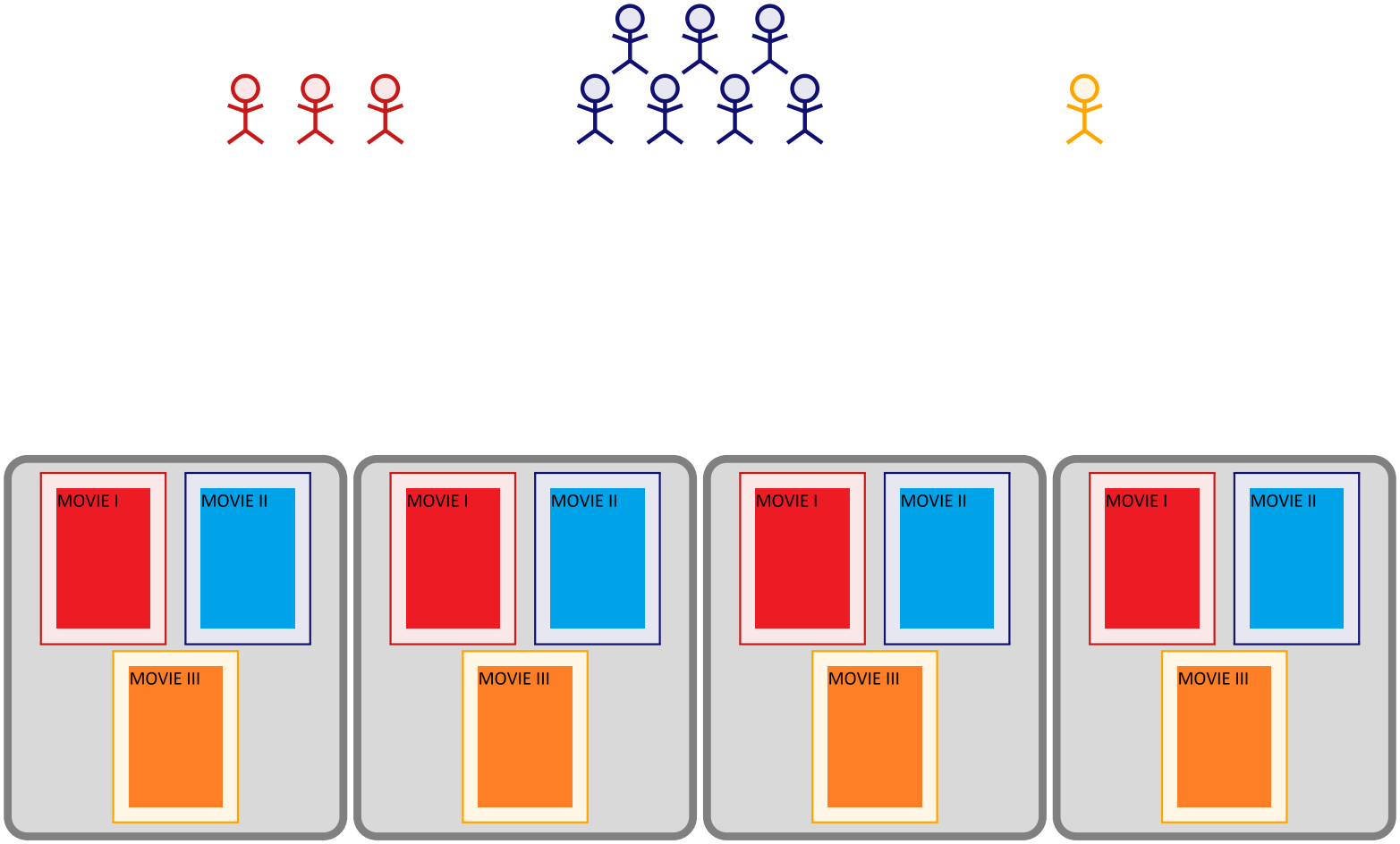
Motivation



Question: How to *model* such a system?

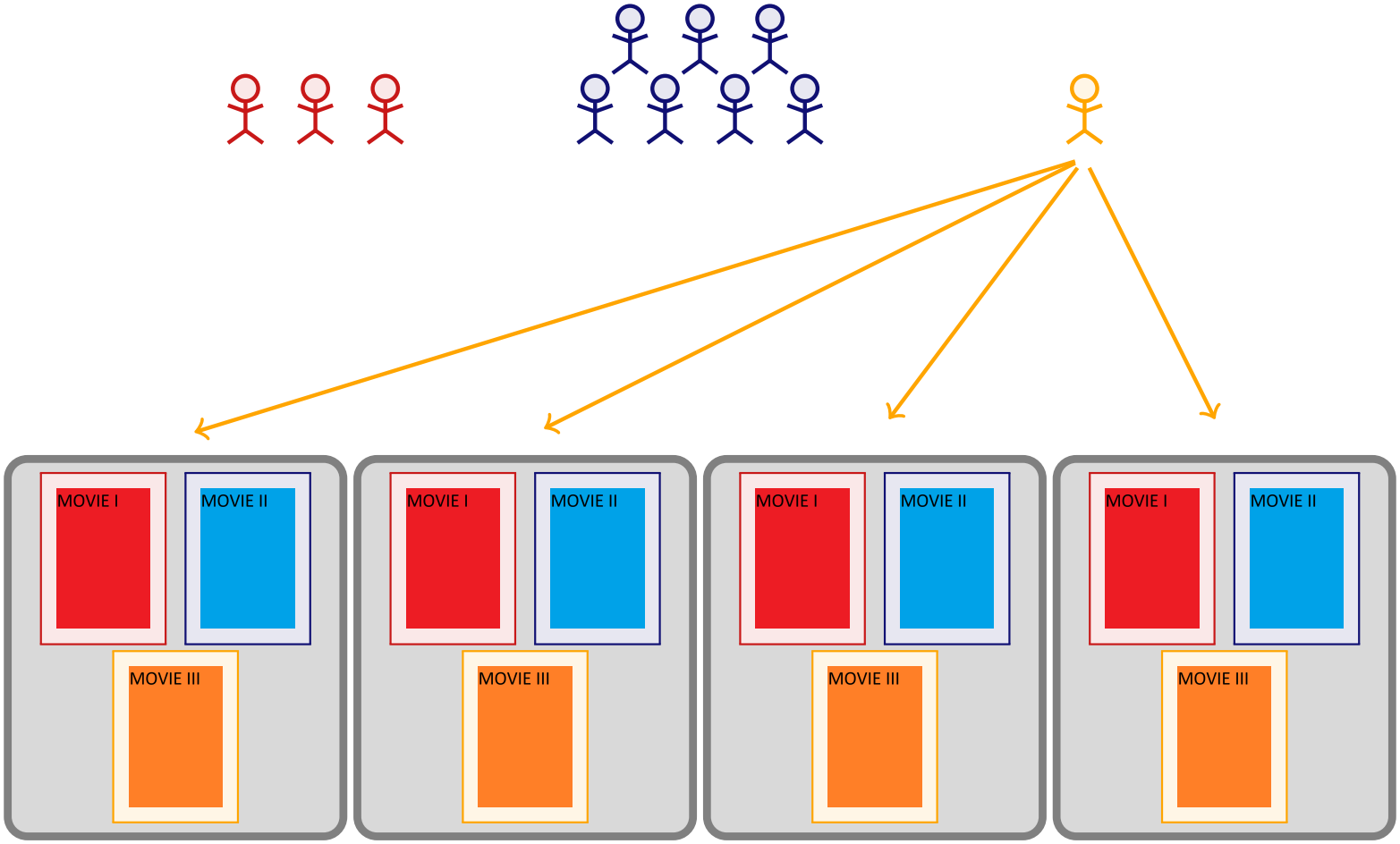
Motivation

Supermarket Model



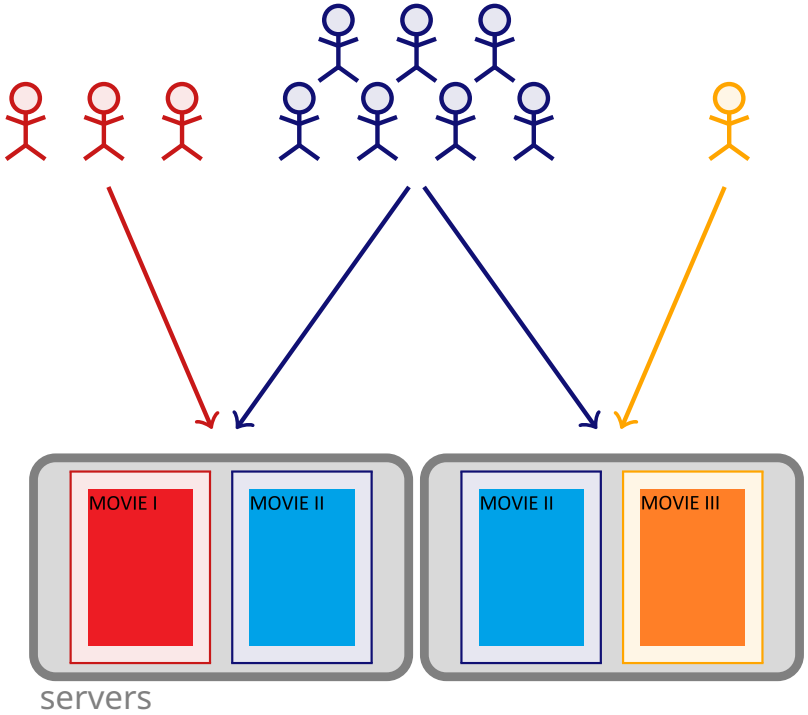
Motivation

Supermarket Model



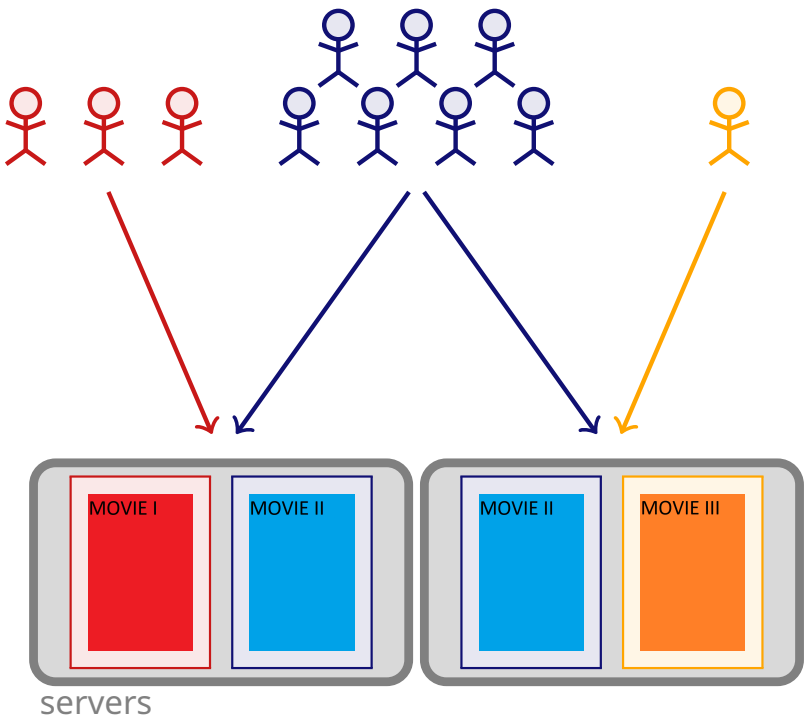
Motivation

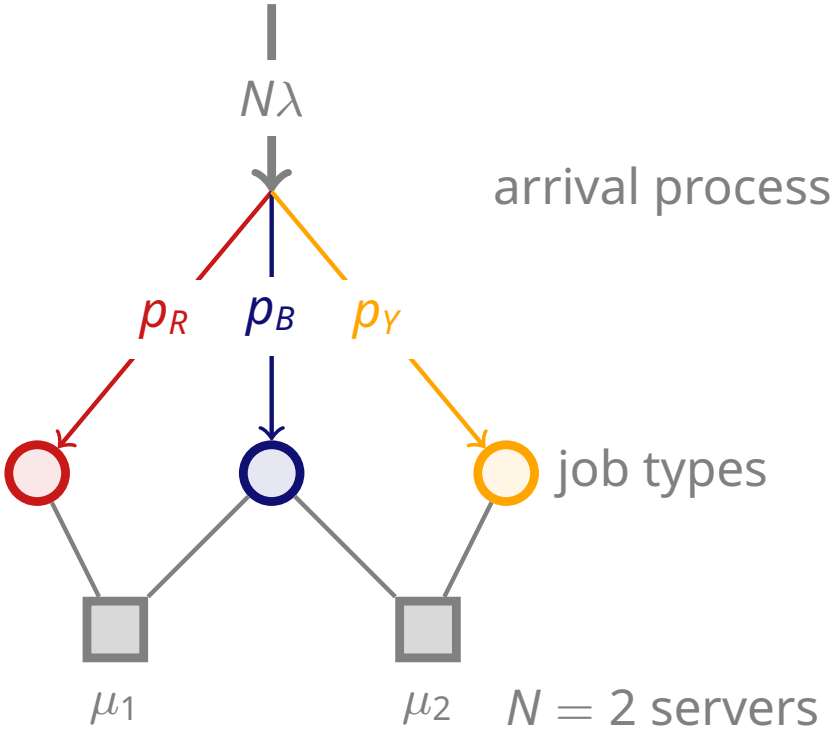
Data Locality Constraints



Motivation

Data Locality Constraints

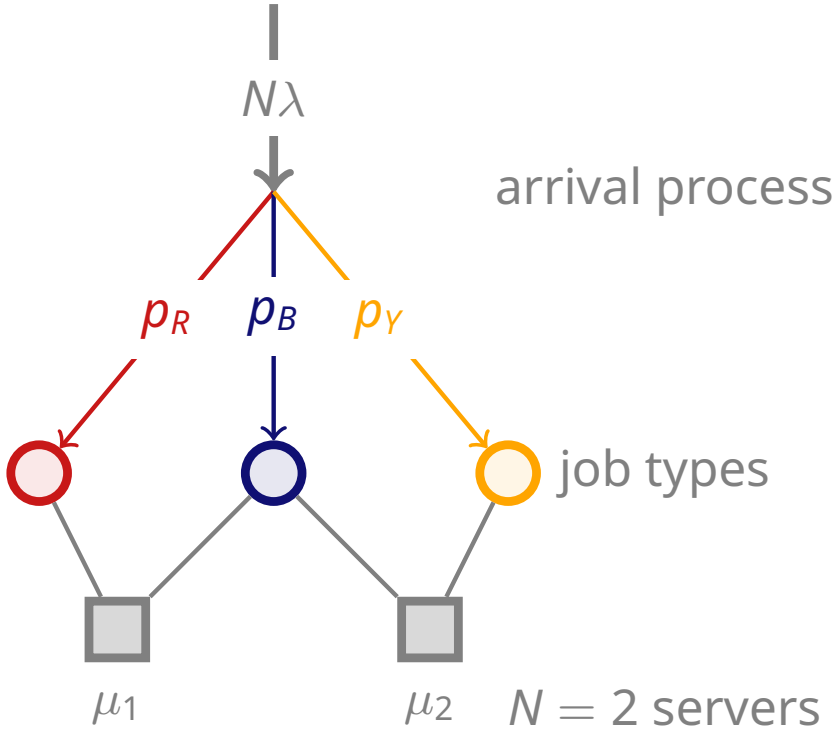




Motivation

Data Locality Constraints

JSQ(d) *
speed aware JSQ/JIQ †
Redundancy
...

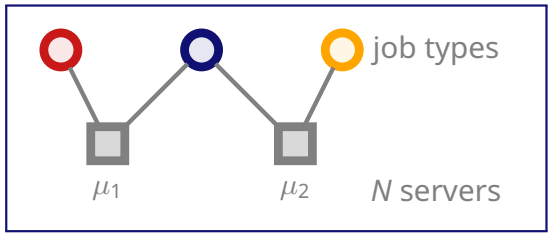


Question: How to *assign* jobs to servers?

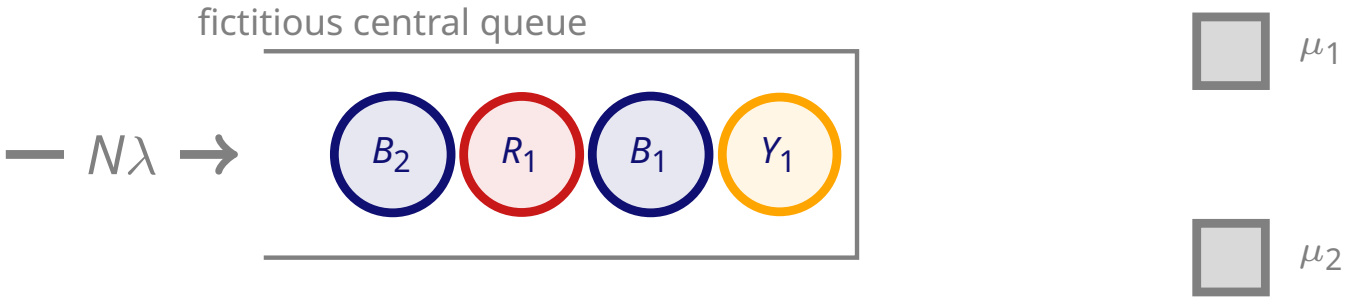
*Rutten and Mukherjee, '20 (large N reg.)
†Weng et al., '20 (large N reg.)

Redundancy Systems

Assignment policy

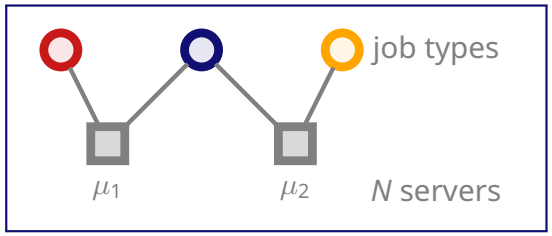


redundancy cancel-on-completion (c.o.c.)

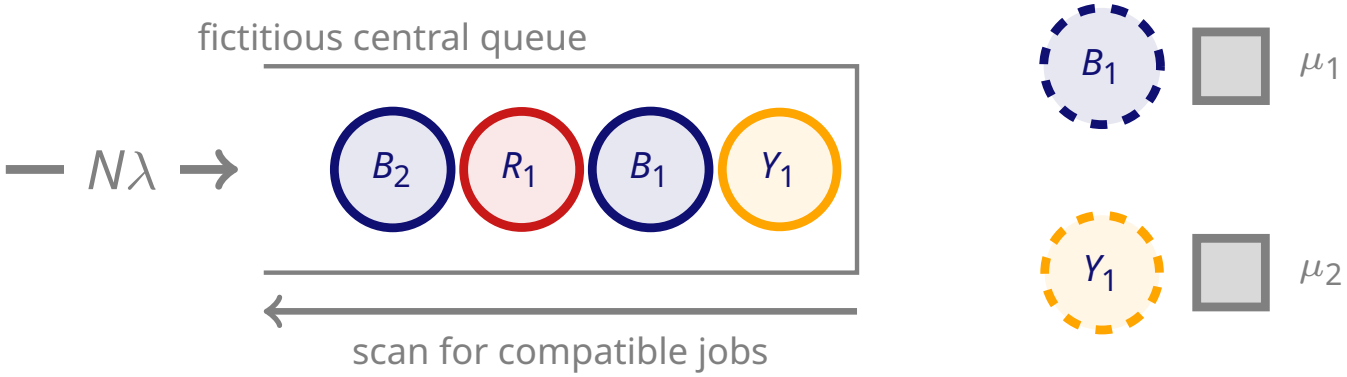


Redundancy Systems

Assignment policy

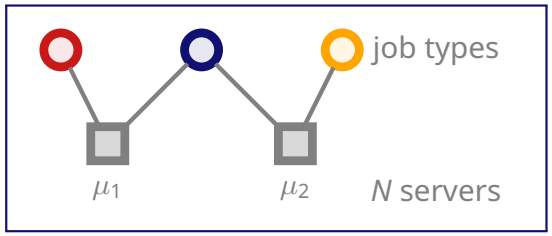


redundancy cancel-on-completion (c.o.c.)

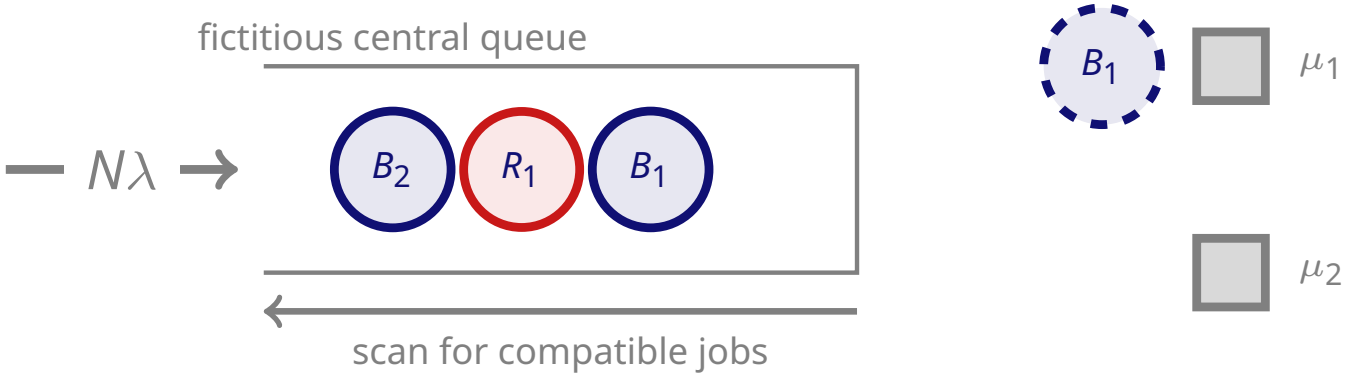


Redundancy Systems

Assignment policy

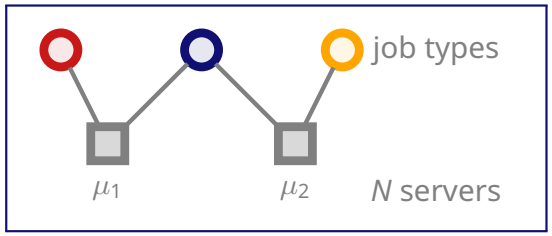


redundancy cancel-on-completion (c.o.c.)

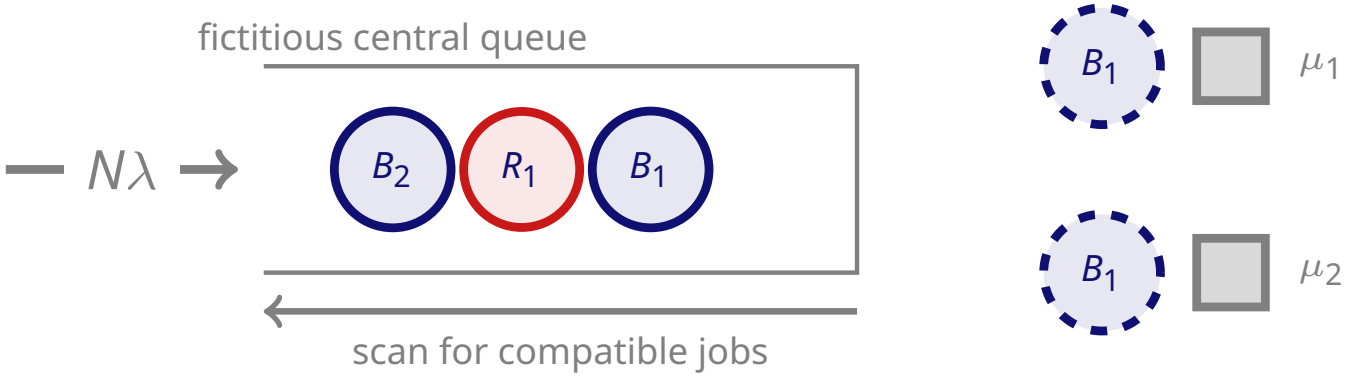


Redundancy Systems

Assignment policy

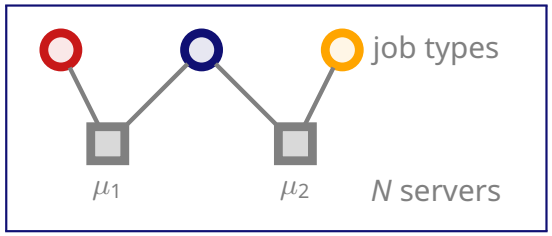


redundancy cancel-on-completion (c.o.c.)

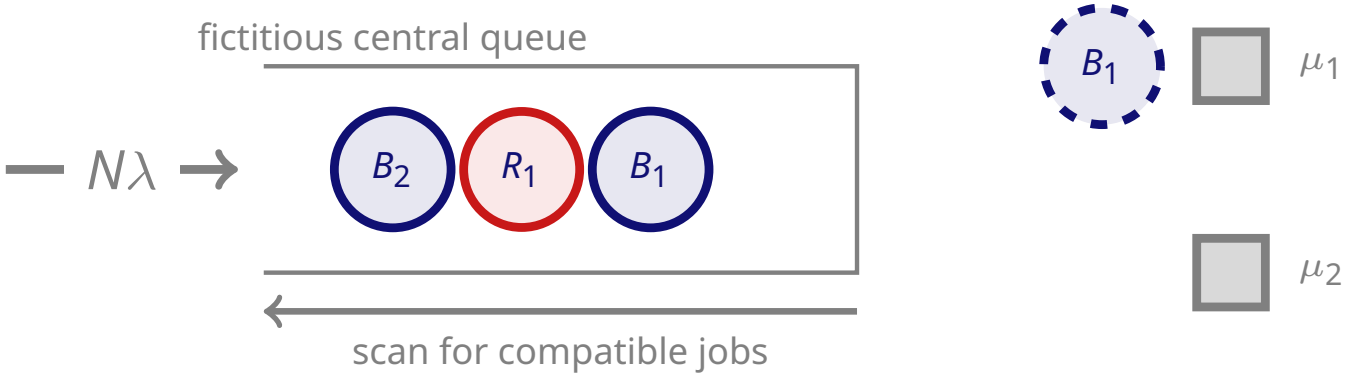


Redundancy Systems

Assignment policy

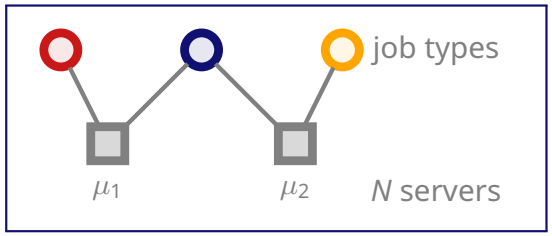


redundancy cancel-on-completion (c.o.c.)

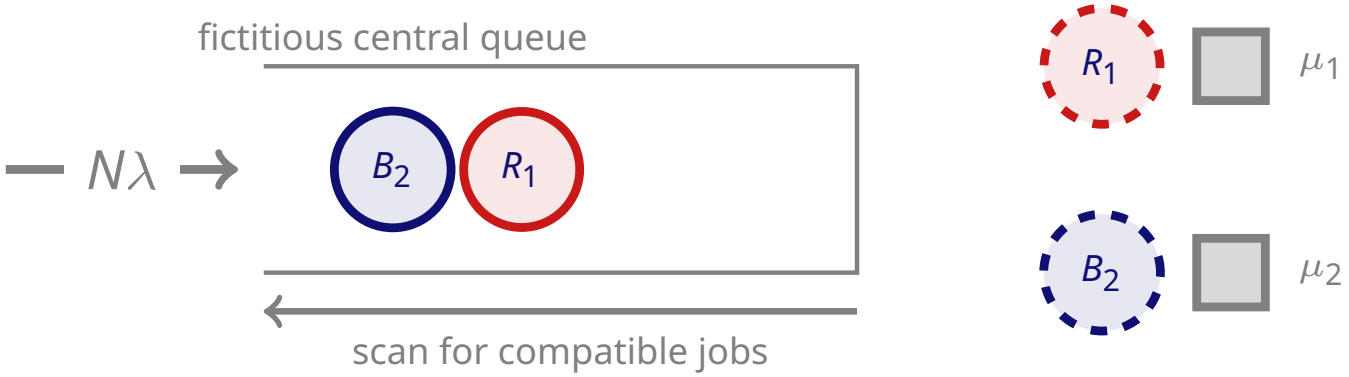


Redundancy Systems

Assignment policy



redundancy cancel-on-completion (c.o.c.)



Question: How to *analyze* such a system?

Redundancy Systems

Product-form expression [‡]

$$\pi(c_1, c_2, \ldots, c_n) = C \prod_{i=1}^{n} \frac{N\lambda p_{c_i}}{\mu(c_1, \ldots, c_i)}$$

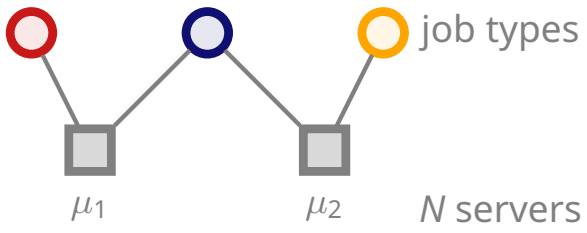
[‡] Gardner et al., ’16; Ayesta et al., ’18

Redundancy Systems

Product-form expression [‡]

$$\pi(c_1, c_2, \ldots, c_n) = C \prod_{i=1}^{n} \frac{N\lambda p_{c_i}}{\mu(c_1, \ldots, c_i)}$$

$$\begin{aligned} & \pi(Y, B, R, B) \\ = \quad & C \left(\frac{N\lambda p_Y}{\mu_2}\right) \left(\frac{N\lambda p_B}{\mu_1+\mu_2}\right) \left(\frac{N\lambda p_R}{\mu_1+\mu_2}\right) \left(\frac{N\lambda p_B}{\mu_1+\mu_2}\right) \end{aligned}$$

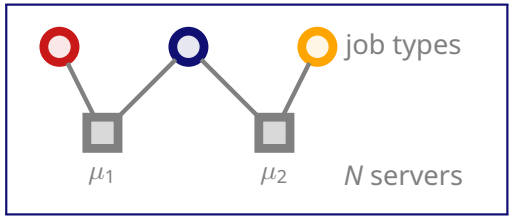


Question: How to *analyze* this system using PF expressions?

[‡] Gardner et al., '16; Ayesta et al., '18

Redundancy Systems

Product-form expression

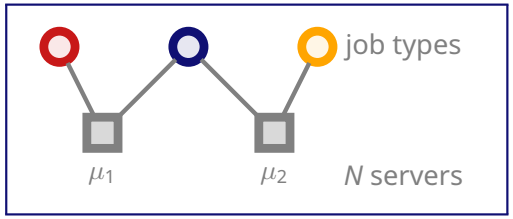


Difficult: aggregate states according to the number of jobs.

$$
\begin{aligned}
& \mathbb{P}\{3 \text{ jobs in the system}\} \\
= \; & \pi(Y, B, R) + \pi(R, R, R) + \pi(B, R, R) + \pi(R, R, B) + \pi(R, Y, B) + \dots
\end{aligned}
$$

Redundancy Systems

Product-form expression



Difficult: aggregate states according to the number of jobs.

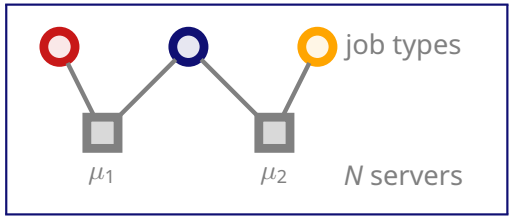
$$\begin{aligned} & \mathbb{P}\{3 \text{ jobs in the system}\} \\ = \; & \pi(Y,B,R) + \pi(R,R,R) + \pi(B,R,R) + \pi(R,R,B) + \pi(R,Y,B) + \dots \end{aligned}$$

$$\begin{aligned} \pi(Y,B,R) &= C\left(\frac{N\lambda p_Y}{\mu_2}\right)\left(\frac{N\lambda p_B}{\mu_1+\mu_2}\right)\left(\frac{N\lambda p_R}{\mu_1+\mu_2}\right) \\ \pi(R,R,R) &= C\left(\frac{N\lambda p_R}{\mu_1}\right)\left(\frac{N\lambda p_R}{\mu_1}\right)\left(\frac{N\lambda p_R}{\mu_1}\right) \end{aligned}$$

different job types

Redundancy Systems

Product-form expression



Difficult: aggregate states according to the number of jobs.

$$\begin{aligned} & \mathbb{P}\{3 \text{ jobs in the system}\} \\ = \; & \pi(Y,B,R) + \pi(R,R,R) + \pi(B,R,R) + \pi(R,R,B) + \pi(R,Y,B) + \dots \end{aligned}$$

$$\begin{aligned} \pi(Y,B,R) &= C\left(\frac{N\lambda p_Y}{\mu_2}\right)\left(\frac{N\lambda p_B}{\mu_1+\mu_2}\right)\left(\frac{N\lambda p_R}{\mu_1+\mu_2}\right) \\ \pi(R,R,R) &= C\left(\frac{N\lambda p_R}{\mu_1}\right)\left(\frac{N\lambda p_R}{\mu_1}\right)\left(\frac{N\lambda p_R}{\mu_1}\right) \end{aligned}$$

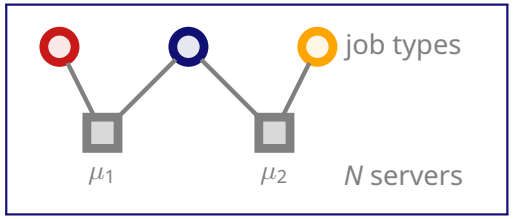
different job types

$$\begin{aligned} \pi(B,R,R) &= C\left(\frac{N\lambda p_B}{\mu_1+\mu_2}\right)\left(\frac{N\lambda p_R}{\mu_1+\mu_2}\right)\left(\frac{N\lambda p_R}{\mu_1+\mu_2}\right) \\ \pi(R,R,B) &= C\left(\frac{N\lambda p_R}{\mu_1}\right)\left(\frac{N\lambda p_R}{\mu_1}\right)\left(\frac{N\lambda p_B}{\mu_1+\mu_2}\right) \end{aligned}$$

order of the jobs

Redundancy Systems

Product-form expression



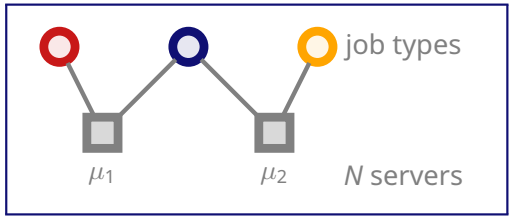
Easier: aggregate states according to the job types *and* their order of first occurrences.

E.g., R and B:

$$\pi(R, B) = C \left(\frac{N\lambda p_R}{\mu_1}\right) \left(\frac{N\lambda p_B}{\mu_1 + \mu_2}\right)$$

Redundancy Systems

Product-form expression



Easier: aggregate states according to the job types *and* their order of first occurrences.

E.g., R and B:

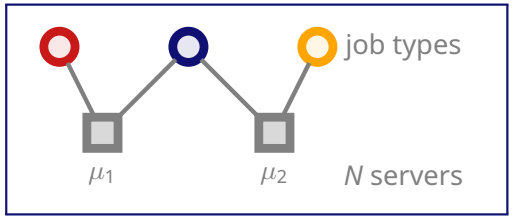
$$\pi(R, B) = C \left(\frac{N\lambda p_R}{\mu_1}\right) \left(\frac{N\lambda p_B}{\mu_1+\mu_2}\right)$$

$$\pi(R, \{R\}^m, B) = C \left(\frac{N\lambda p_R}{\mu_1}\right) \left(\frac{N\lambda p_R}{\mu_1}\right)^m \left(\frac{N\lambda p_B}{\mu_1+\mu_2}\right)$$

$m, n_R, n_B \in \mathbb{N}$

Redundancy Systems

Product-form expression



Easier: aggregate states according to the job types *and* their order of first occurrences.

E.g., R and B:

$$\pi(R, B) = C \left(\frac{N\lambda p_R}{\mu_1}\right) \left(\frac{N\lambda p_B}{\mu_1+\mu_2}\right)$$

$$\pi(R, \{R\}^m, B) = C \left(\frac{N\lambda p_R}{\mu_1}\right) \left(\frac{N\lambda p_R}{\mu_1}\right)^m \left(\frac{N\lambda p_B}{\mu_1+\mu_2}\right)$$

$$\pi(R, \{R\}^m, B, \{R\}^{n_R}, \{B\}^{n_B}) = C \left(\frac{N\lambda p_R}{\mu_1}\right) \left(\frac{N\lambda p_R}{\mu_1}\right)^m \left(\frac{N\lambda p_B}{\mu_1+\mu_2}\right) \left(\frac{N\lambda p_R}{\mu_1+\mu_2}\right)^{n_R} \left(\frac{N\lambda p_B}{\mu_1+\mu_2}\right)^{n_B}$$

$m, n_R, n_B \in \mathbb{N}$

Redundancy Systems

Product-form expression

Probability generating function:

$$\mathbb{E}\left[z_R^{Q_R} z_B^{Q_B} z_Y^{Q_Y}\right]$$

$$= C + C \sum_{m=1}^{3} \sum_{\boldsymbol{s} \in \mathcal{S}_m} \prod_{j=1}^{m} \left(\frac{N\lambda p_{s_j} z_{s_j}}{\mu(s_1,\ldots,s_j)} \right) \left(1 - \frac{N\lambda}{\mu(s_1,\ldots,s_j)} \sum_{i=1}^{j} p_{s_i} z_{s_i} \right)^{-1}$$

$|z_R|, |z_B|, |z_Y| \leq 1.$

Heavy-Traffic Universality

Result

$$\boxed{\left(1 - \frac{\lambda}{\mu}\right)(Q_R, Q_B, Q_Y) \xrightarrow{d} \mathrm{Exp}(1)\,(p_R, p_B, p_Y)}$$

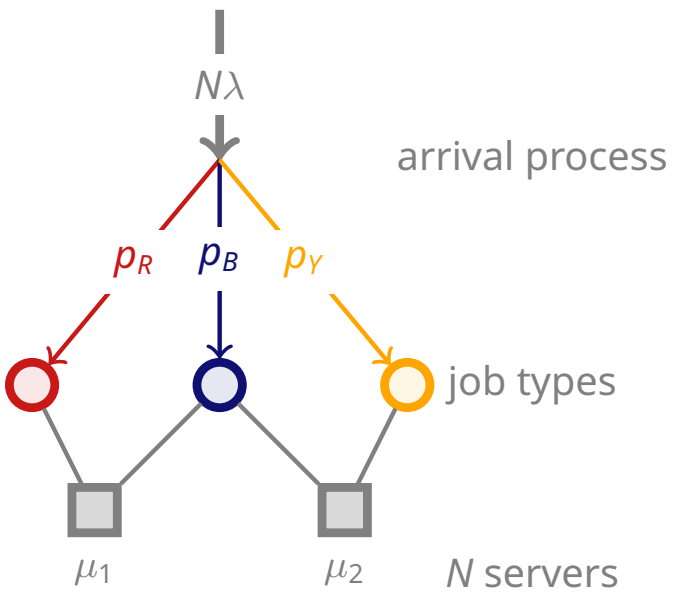
as $\lambda \uparrow \mu$.

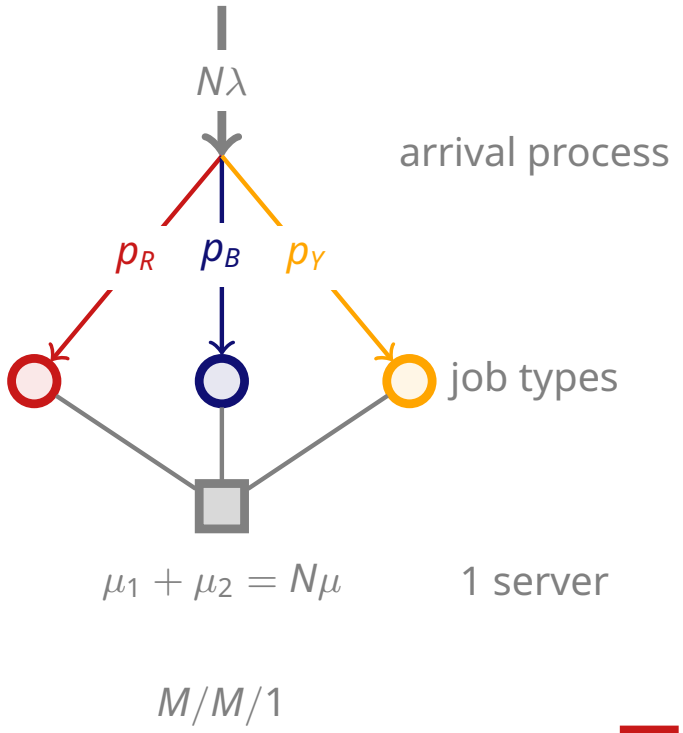
Heavy-Traffic Universality

Result

$$\left(1 - \frac{\lambda}{\mu}\right)(Q_R, Q_B, Q_Y) \xrightarrow{d} \mathrm{Exp}(1)\,(p_R, p_B, p_Y)$$

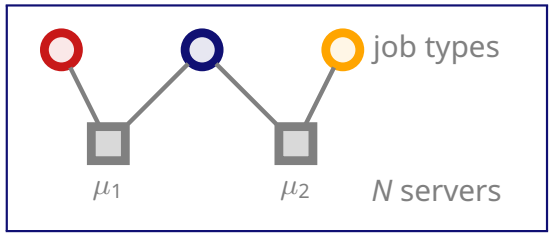
as $\lambda \uparrow \mu$.





Heavy-Traffic Universality

Result



Stability condition for the PF:

$$\begin{aligned}
\lambda N p_R &< \mu_1 \\
\lambda N p_Y &< \mu_2 \\
\lambda N p_B &< \mu_1 + \mu_2
\end{aligned}$$

$$\begin{aligned}
\lambda N (p_R + p_B) &< \mu_1 + \mu_2 = \mu N \\
\lambda N (p_B + p_Y) &< \mu_1 + \mu_2 = \mu N
\end{aligned}$$

$$\lambda < \mu$$

HT result: No local bottlenecks allowed.

Concluding remarks

General result for *all* systems with no local bottlenecks.

Heavy-Traffic Universality (result)
of Redundancy Systems (tool: product-form expressions)
with Data Locality Constraints (motivation)

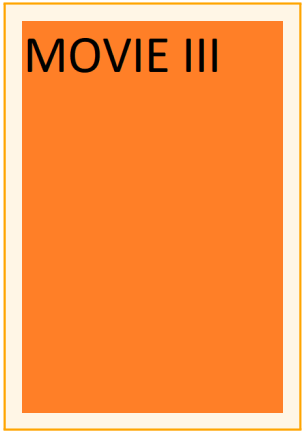
arXiv: 2005.14566
Extensions: Relaxed condition & Redundancy cancel-on-start

Concluding remarks

General result for *all* systems with no local bottlenecks.

Heavy-Traffic Universality (result)
of Redundancy Systems (tool: product-form expressions)
with Data Locality Constraints (motivation)

arXiv: 2005.14566
Extensions: Relaxed condition & Redundancy cancel-on-start



Thank you!

Heavy-Traffic Universality

Systems without local bottlenecks

Stability conditions for PF expressions:

$$N\lambda \sum_{S \in \mathcal{T}} p_S < \sum_{n \in \cup_{S \in \mathcal{T}} S} \mu_n$$

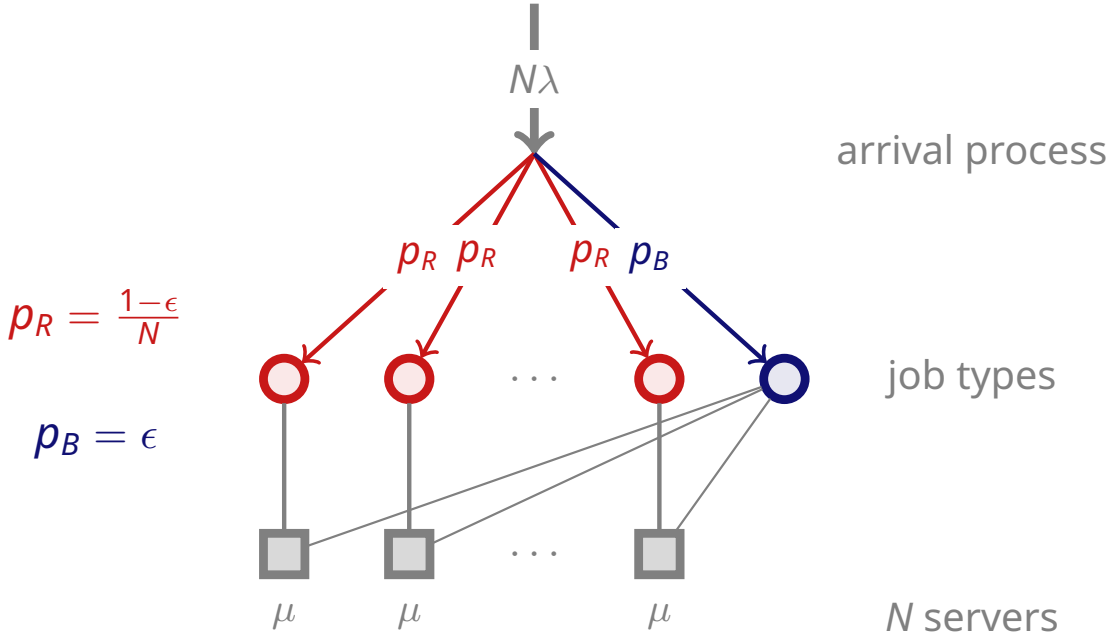
Condition for no local bottlenecks:

$$\sum_{S \in \mathcal{T}} p_S < \frac{1}{N\mu} \sum_{n \in \cup_{S \in \mathcal{T}} S} \mu_n$$

Heavy-Traffic Universality

Systems without local bottlenecks

General result for *all* systems with no local bottlenecks.



Heavy-Traffic Universality

Systems with local bottlenecks

(partial) state space collapse and a truncated system:

$$\left(\left(1-\frac{\lambda}{\lambda^*}\right)(Q_S)_{S\in\mathcal{T}^*},(Q_S)_{S\notin\mathcal{T}^*}\right)\xrightarrow{d}\left(\mathrm{Exp}(1)\left(\frac{p_S}{p_{\mathcal{T}^*}}\right)_{S\in\mathcal{T}^*},(Q_S^*)_{S\notin\mathcal{T}^*}\right)$$

as $\lambda\uparrow\lambda^*$