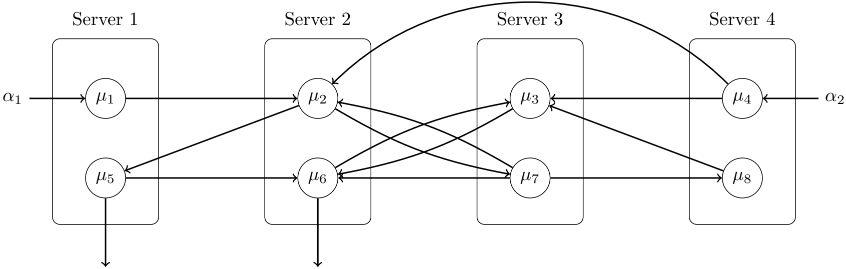
Stability and Instability of Queueing Networks with Infinite Supplies

Yoni Nazarathy
The University of Queensland

Tom Salisbury
York University

Presented at the Banach Center, Bedlewo Poland,
May 31, 2014.

Multi-Class Queueing Networks:
(MCQN, 1990's ...)



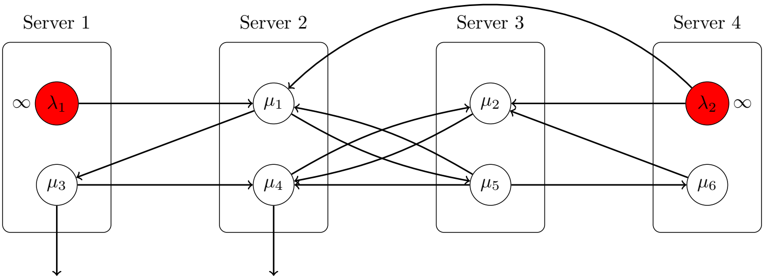
Major Question: How to control for stability?

Bramson, Dai, Foss, Hasenbein, Kumar, Meyn, Stolyar, Weiss, Dieker...

Full utilization $\Rightarrow$ Queue blow-up

Diffusion Limits: Bramson, Dai, Harrison, Reiman, Williams ...

Queueing networks with infinite supplies:
(MCQN-IVQ, Weiss et. al.)



Useful models: Full utilization without queue blow-up

$$\text{classes:} \quad \mathcal{K} = \mathcal{K}_a \cup \mathcal{K}_s$$

$$\text{servers:} \quad \mathcal{I}, \quad \text{with} \;\; \mathcal{C}(i) = \{k \in \mathcal{K} \;:\; k \text{ served by i}\}$$

SLLN for primitive counting processes:

$$S_k(t) \sim \lambda_k t, \;\; k \in \mathcal{K}_a$$
$$S_k(t) \sim \mu_k t, \;\; k \in \mathcal{K}_s$$
$$\Phi_{k,k'}(\ell) \sim p_{k,k'}\ell, \;\; k \in \mathcal{K}, \; k' \in \mathcal{K}_s$$

Dynamics $\Big(Q(\cdot), \; T(\cdot) \Big)$:

$$Q_k(t) = Q_k(0) + \sum_{k' \in \mathcal{K}} \Phi_{k',k}\Big(S_{k'}\big(T_{k'}(t)\big)\Big) - S_k\big(T_k(t)\big), \qquad k \in \mathcal{K}_s$$

$$Q_k(t) = Q_k(0) + \sum_{k' \in \mathcal{K}} \Phi_{k',k}\Big(S_{k'}\big(T_{k'}(t)\big)\Big) - S_k\big(T_k(t)\big), \qquad k \in \mathcal{K}_s$$

constraints:

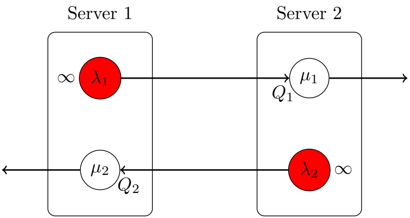
$$\int_0^t \mathbb{1}\{Q_k(s) = 0\}\, dT_k(s) \ = \ 0, \quad k \in \mathcal{K}_s$$

$$\sum_{k \in \mathcal{C}(i)} T_k(t) \leq t, \qquad \forall\, i \in \mathcal{I}$$

For some servers, want full utilization:

$$\sum_{k \in \mathcal{C}(i)} T_k(t) = t$$

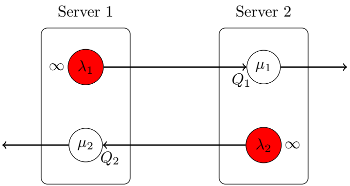
e.g. “push-pull” network:



$$Q_1(t) = Q_1(0) + S_{a,1}\big(T_1(t)\big) - S_{s,1}\big(t - T_2(t)\big)$$
$$Q_2(t) = Q_2(0) + S_{a,2}\big(T_2(t)\big) - S_{s,2}\big(t - T_1(t)\big)$$

$$T_1(t) \geq \int_0^t \mathbb{1}\{Q_2(s) = 0\}\, ds, \qquad T_2(t) \geq \int_0^t \mathbb{1}\{Q_1(s) = 0\}\, ds,$$

Want an adapted policy $\big($yielding $T(\cdot)\big)$, such that $Q(\cdot)$ is stable

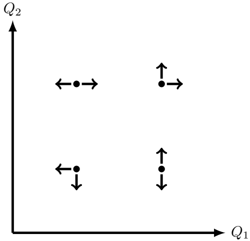


Control policy

$$\mathcal{P} : \mathbb{Z}_+^2 \to \{ \text{“do } \lambda\text{”}, \text{“do } \mu\text{”} \}^2$$

with (axis restrictions):

$$\mathcal{P}(\cdot, 0) = (\text{“do } \lambda\text{”}, \cdot), \quad \mathcal{P}(0, \cdot) = (\cdot, \text{“do } \lambda\text{”})$$



Stability :

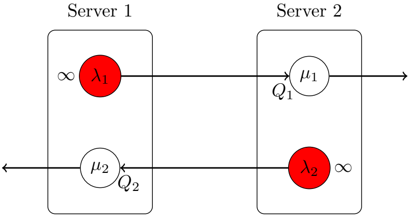
$$\exists\ \mathcal{A} \subset \mathbb{Z}_+^{|\mathcal{K}_s|}, \text{ with } |\mathcal{A}| < \infty \text{ such that for all } Q(0) \in \mathbb{Z}_+^{|\mathcal{K}_s|},$$

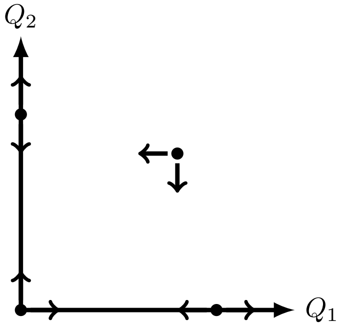
$$\mathbb{E}\Big[\inf\{t \,:\, Q(t) \in \mathcal{A}\}\Big] < \infty$$

Assume $\lambda_i < \mu_i$ and take priority policy:

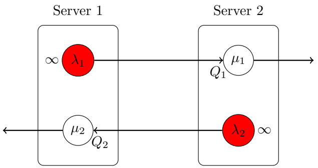
$$T_1(t) = \int_0^t \mathbb{1}\{Q_2(s) = 0\}\, ds, \qquad T_2(t) = \int_0^t \mathbb{1}\{Q_1(s) = 0\}\, ds,$$

In this case the network is stable

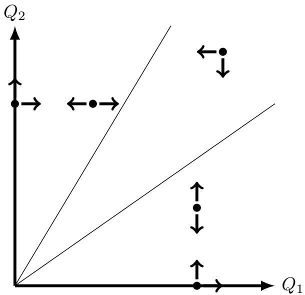

Server 1
Server 2
∞
λ_1
μ_1
Q_1
μ_2
Q_2
λ_2
∞


Q_2
Q_1

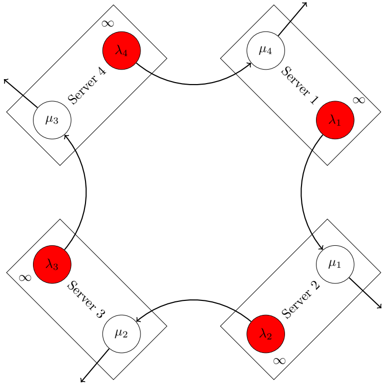
Consider now the case, $\lambda_i > \mu_i$

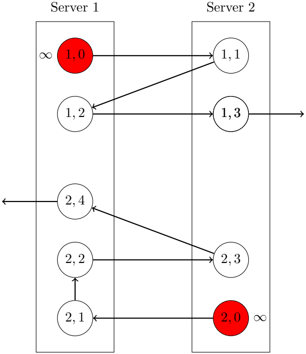


A threshold policy is stabilizing:



Some success with other structured networks also...
(Weiss et. al.)





also Re-entrant lines, etc...

A general question

Static Planning:

$$\max_{u,v} \quad \sum_{i\in\mathcal{K}_a} g_i\,\lambda_i\,v_i$$

$$s.t. \quad (I - P'_{s,s})\ \mathrm{diag}(\mu)\ u - P'_{a,s}\ \mathrm{diag}(\lambda)\ v = \mathbf{0},$$

$$C_s\ u + C_a\ v \le \mathbf{1},$$

$$u,\ v \ge \mathbf{0}.$$

Solution (u^*, v^*) yields load:

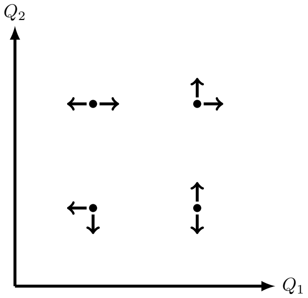
$$\rho := C_s u^* + C_a v^*, \qquad \tilde{\rho} := C_s u^*$$

Question: Assume $\rho = \mathbf{1}$ but $\tilde{\rho} < \mathbf{1}$. Is there a policy that achieves u^*, v^* and is stable?

We believe the answer is almost always "yes",
(ongoing work of Lefeber, Weiss, et. al to find a "universal policy")

In the current work we identify non-stabilizable singular cases

Consider now again the push-pull example with $\lambda_i = \mu_i$
(for simplicity assume $= 1$)



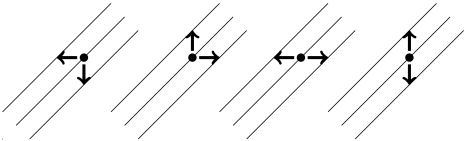
Is there a $\mathcal{P}$ such that the network is stable?

Answer: No

Proof (push-pull exponential processing times)

1. $X(n) = (X_1(n), X_2(n))$ the embedded Markov chain of $(Q_1(t), Q_2(t))$
2. $g(x_1, x_2) = \lambda_2 x_1 - \lambda_1 x_2$ and $Z(n) = g\big(X_1(n), X_2(n)\big)$

   $Z(\cdot)$ is a martingale for any policy $\mathcal{P}$:

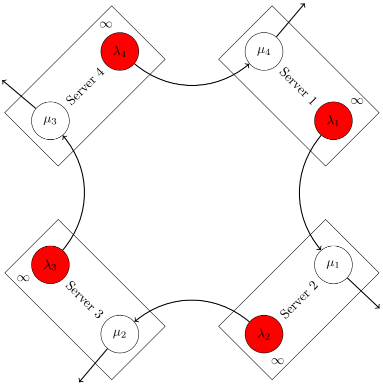


3. Assume $\exists$ positive recurrent $\mathcal{A} \subset \mathbb{Z}_+^2$.
   Take $x, y \in \mathcal{A}$ with $g(x) \neq g(y)$
4. Set $X(0) = x$ and define $\tau = \inf\{n \geq 0 : X(n) = y\}$
5. For $\mathcal{A}$ to be positive recurrent, $\mathbb{E}[\tau] < \infty$ so:

$$g(x) = \mathbb{E}[Z(0)] = \mathbb{E}[Z(\tau)] = g(y),$$

a contradiction.

Idea generalises to other structures also,



YN, Leonardo Rojas–Nandayapa and Tom Salisbury, "*Non-Existence of Stabilizing Policies for the Critical Push-Pull Network and Generalizations*", Operations Research Letters, 41.3 (2013): 265-270.

Fluid intuition (Lefeber)

$$\frac{d\overline{Q}(t)}{dt} = a^* - R^* u(t)$$

with,

$$a^* := P'_{a,s}\text{diag}(\lambda)\,\mathbf{1}$$
$$R^* := (I - P'_{s,s})\ \text{diag}(\mu) + P'_{a,s}\ \text{diag}(\lambda) C_s$$

If R^* is singular with a column-space containing a^* then the system is not controllable

$$\frac{d}{dt}\begin{bmatrix} \overline{Q}_1(t) \\ \overline{Q}_2(t) \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} - \begin{bmatrix} \mu_1 & \lambda_1 \\ \lambda_2 & \mu_2 \end{bmatrix} \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}$$

Main result

Assume:

- Processing times with bounded residual means:

$$\mathbb{E}[U - t \mid U > t] \leq \text{const}$$

- At most one infinite supply per server: $\left|\mathcal{C}(i) \cap \mathcal{K}_a\right| \leq 1$
- Solution to static planning (u^*, v^*) yields $\tilde{\rho} < \mathbf{1}$

Then if $R^* = (I - P'_{s,s})\,\text{diag}(\mu) + P'_{a,s}\,\text{diag}(\lambda) C_s$ is **singular** then there does **not exist** a policy achieving rates,

$$T_k(t) \sim u_k^* t, \quad k \in \mathcal{K}_s, \qquad \text{and} \qquad T_k(t) \sim v_k^* t, \quad k \in \mathcal{K}_a$$

that is **stable**

Proof idea

- Algebraic argument constructing a suitable linear "test-function" $g(x) = \sum_{k \in \mathcal{K}_s} h_k x_k$ based on

$$R^* = (I - P'_{s,s})\ \mathrm{diag}(\mu) + P'_{a,s}\ \mathrm{diag}(\lambda) C_s,$$

  when it is singular

- Then, $Z(t) := g\big(Q(t)\big)$ is a martingale in the exponential processing times case and in the general case is shown to satisfy (under the stability assumption),

$$\Big|\mathbb{E}\big[Z(\tau)\big] - Z(0)\Big| \le \text{const} \sum_k \mathbb{E}\big[S_k\big(T_k(\tau)\big) - \lambda_k T_k(\tau)\big] \le \text{const}$$

  with τ the hitting time of $\mathcal{A}$.

  Thus $\big|Z(0)\big| \le$ const and this is a contradiction.