Light-traffic interpolation of a multi-class queue with relative priorities

Ane Izagirre

EURO 2013

Joint work with U. Ayesta, A.M. Makowski and I.M. Verloop
Model description

- $M/G/1$ queue with $K$ different classes of customers.
- Class-$k$ customers arrive according to independent Poisson processes with rate $\lambda_k$, $k = 1, \ldots, K$.
- The overall arrival rate: $\lambda = \sum_{k=1}^{K} \lambda_k$.
- Work-conserving
- Service is non-preemptive
Model description

Class $k, k = 1, \ldots, K$, is selected to be served with probability

$$\frac{n_k w_k}{\sum_{j=1}^{K} n_j w_j}$$

where $w_k > 0$ are class-dependent weights, and $n_k$ is the number of class-$k$ customers at decision epoch (service completion).

Intra class policy: random-order-of-service (ROS)
Model description

- $B_k$: Service requirement of class-$k$ customers (i.i.d.)
- $\rho_k = \lambda_k \mathbb{E}[B_k]$: traffic intensity of class-$k$ queue
- Total traffic intensity:

$$\rho = \sum_{k=1}^{K} \rho_k = \sum_{k=1}^{K} \lambda_k \mathbb{E}[B_k] = \lambda \sum_{k=1}^{K} \alpha_k \mathbb{E}[B_k],$$

where $\alpha_k = \lambda_k / \lambda$ is the probability an arrival is of class $k$.

- Stability condition: $\rho < 1$
Outline

- Steady-state waiting time and queue length statistics available.
  Cumbersome expression which are not very insightful.

- Possible way out; Extreme regimes are easier to analyze:
  - Heavy-traffic result \((\rho \uparrow 1 \iff \lambda \uparrow \frac{1}{\mathbb{E}[B]})\)
    
    \[ \mathbb{E} \left[ \lim_{\rho \uparrow 1} (1 - \rho) W_k(\rho) \right] = \frac{\mathbb{E}[B^2]}{2w_k \sum_{j=1}^{K} \frac{\alpha_j}{w_j} \mathbb{E}[B_j]}, \]


  - Light-traffic result \((\rho \downarrow 0 \iff \lambda \downarrow 0)\) Now

- Light-traffic interpolation Now

- Some numeric examples Now

- Monotonicity result Now
Light-traffic result \((\rho \downarrow 0 \Leftrightarrow \lambda \downarrow 0)\)

We have shown

\[
\overline{W}_{LT}^k(\lambda) \simeq \overline{W}_k(0) + \lambda \overline{W}'_k(0) + \frac{\lambda^2}{2!} \overline{W}''_k(0),
\]

where

\[
\begin{align*}
\overline{W}_k(0) &= 0 \\
\overline{W}'_k(0) &= \frac{\mathbb{E}[B^2]}{2} \\
\overline{W}''_k(0) &= \mathbb{E}[B^2]\sum_{l=1}^{K} \alpha_l \frac{\omega_l}{\omega_l + \omega_k} \mathbb{E}[B_l].
\end{align*}
\]

Light-traffic interpolation

Interpolate

\[ T_k(\lambda) := (1 - \rho)W_k(\lambda) = (1 - \lambda \mathbb{E}[B])W_k(\lambda) \]

by the polynomial

\[ \hat{T}_k(\lambda) = g_0 + g_1 \lambda + g_2 \lambda^2 + g_3 \lambda^3. \]

The light-traffic conditions are

\[ \hat{T}_k^{(m)}(0) = T_k^{(m)}(0), \quad \text{for } m = 0, 1, 2, \]

and when \( \lambda \) is small \( \overline{W}_k(\lambda) \simeq \overline{W}_{LT}^k(\lambda) \), where

\[ \overline{W}_{LT}^k(\lambda) \simeq \overline{W}_k(0) + \lambda \overline{W}_k'(0) + \frac{\lambda^2}{2!} \overline{W}_k''(0). \]

The heavy-traffic condition is

\[ \hat{T}_k\left(\frac{1}{\mathbb{E}[B]}\right)^{-} = T_k\left(\frac{1}{\mathbb{E}[B]}\right)^{-} = H_k, \]

where

\[ H_k = \mathbb{E}\left[ \lim_{\rho \uparrow 1} (1 - \rho)W_k \right] = \frac{\mathbb{E}[B^2]}{2p_k \sum_{j=1}^{K} \frac{\alpha_j}{p_j} \mathbb{E}[B_j]}. \]
Light-traffic interpolation result

We undo the normalization being

\[ \overline{W}^{INT}_k(\lambda) := \frac{\hat{T}_k(\lambda)}{(1 - \lambda \mathbb{E}[B])}, \quad 0 \leq \lambda < 1/\mathbb{E}[B] \]

the approximation of \( \overline{W}_k(\lambda) \).

**Proposition**

The mean waiting time of a class-k customer when the arrival rate is \( \lambda \) is approximated by the polynomial

\[
\overline{W}^{INT}_k(\lambda) = \frac{1}{(1 - \lambda \mathbb{E}[B])} \left[ \frac{\mathbb{E}[B^2]}{2} \lambda + \frac{\mathbb{E}[B^2]}{2} \left\{ -\mathbb{E}[B] + \sum_{l=1}^{K} \frac{\alpha_l}{w_l + w_k} \mathbb{E}[B_l] \right\} \lambda^2 \right. \\
+ \left. \frac{\mathbb{E}[B]^3}{2} \left\{ \frac{\mathbb{E}[B^2]}{w_k \sum_{j=1}^{K} \frac{\alpha_j}{w_j} \mathbb{E}[B_j]} - \frac{\mathbb{E}[B^2]^2}{\mathbb{E}[B]^2} \sum_{l=1}^{K} \frac{w_l}{w_l + w_k} \mathbb{E}[B_l] \right\} \lambda^3 \right].
\]

Scaled mean waiting time of class-1 customers and the relative error:

Figure: \( w=[0.1 \ 0.9] \). \( \alpha_1 = 0.415, \alpha_2 = 0.585, E[B_1] = 0.2439, E[B_2] = 0.16667 \) with \( B_1, B_2 \) exponentially distributed.
Scaled mean waiting time of class-1 customers and the relative error:

Figure: \( w=[0.3 \ 0.7] \). \( \alpha_1 = 0.415, \alpha_2 = 0.585, \mathbb{E}[B_1] = 0.2439, E[B_2] = 0.16667 \) with \( B_1, B_2 \) exponentially distributed.
Scaled mean waiting time of class-1 customers and the relative error:

Figure: \( w = [0.9 \ 0.1] \). \( \alpha_1 = 0.415, \alpha_2 = 0.585, \mathbb{E}[B_1] = 0.2439, \mathbb{E}[B_2] = 0.16667 \) with \( B_1, B_2 \) exponentially distributed.
Scaled mean waiting time of class-2 customers and the relative error:

Figure: \( w = [0.1 \ 0.9] \). \( \alpha_1 = 0.415, \alpha_2 = 0.585, \mathbb{E}[B_1] = 0.2439, \mathbb{E}[B_2] = 0.16667 \) with \( B_1, B_2 \) is exponentially distributed.
Scaled mean waiting time of class-2 customers and the relative error:

\[ (1-\rho)E[W_2] \]

\[ \rho \]

Figure: \( w=[0.3 \ 0.7] \). \( \alpha_1 = 0.415 \), \( \alpha_2 = 0.585 \), \( \mathbb{E}[B_1] = 0.2439 \), \( \mathbb{E}[B_2] = 0.16667 \) with \( B_1, B_2 \) is exponentially distributed.
Scaled mean waiting time of class-2 customers and the relative error:

Figure: \( w = [0.9, 0.1] \). \( \alpha_1 = 0.415, \alpha_2 = 0.585, \mathbb{E}[B_1] = 0.2439, \mathbb{E}[B_2] = 0.16667 \) with \( B_1, B_2 \) is exponentially distributed.
Optimal selection of weights

Proposition

Consider two policies with weights $\vec{w}$ and $\vec{v}$, respectively. Without loss of generality assume that $\mathbb{E}[B_{1}] \leq \ldots \leq \mathbb{E}[B_{K}]$. If $\frac{w_{j}}{w_{j+1}} \leq \frac{v_{j}}{v_{j+1}}$, for all $j = 1, \ldots, K - 1$, then

$$\overline{W}_{k}^{INT}(\lambda, \vec{w}) \geq \overline{W}_{k}^{INT}(\lambda, \vec{v}).$$
Proposition

Consider two policies with weights $\mathbf{w}$ and $\mathbf{v}$, respectively. Without loss of generality assume that $\mathbb{E}[B_1] \leq \ldots \leq \mathbb{E}[B_K]$. If $\frac{w_j}{w_{j+1}} \leq \frac{v_j}{v_{j+1}}$, for all $j = 1, \ldots, K-1$, then

$$W_{INT_k}(\lambda, \mathbf{w}) \geq W_{INT_k}(\lambda, \mathbf{v}).$$

Example for 2 classes of customers:

For $K=2$, with $\mathbb{E}[B_1] \leq \mathbb{E}[B_2]$ and $w_1 + w_2 = v_1 + v_2 = 1$, we get

$$W_{INT_k}(\lambda, \mathbf{w}) \geq W_{INT_k}(\lambda, \mathbf{v}) \iff \frac{w_1}{w_2} \leq \frac{v_1}{v_2} \iff w_1 \leq v_1.$$
Conclusions and Future Work

Conclusions:

- Making an interpolation and combining the heavy-traffic and light-traffic results we obtain an approximation of the system for any $\rho$.
- Accuracy of the interpolation result.
- Selection of optimal weights for the performance.

Future work:

- Construct a light-traffic interpolation for a discriminatory-processor-sharing (DPS) discipline.
- Approximate the distribution of the waiting time.
- Explore whether the waiting time can be heavy-tailed.
Conclusions and Future Work

Conclusions:

- Making an interpolation and combining the heavy-traffic and light-traffic results we obtain an approximation of the system for any $\rho$.
- Accuracy of the interpolation result.
- Selection of optimal weights for the performance.

Future work:

- Construct a light-traffic interpolation for a discriminatory-processor-sharing (DPS) discipline.
- Approximate the distribution of the waiting time.
- Explore whether the waiting time can be heavy-tailed.
Thank you for your attention!