A fluid model for closed queueing networks with PS stations

Juan F. Pérez Giuliano Casale

Department of Computing
Imperial College London

October 14, 2013

Abstract

This technical report introduces a closed multi-class queueing network (QN) model with class-switching, where the service rates are defined to represent multi-processor stations with a processor-sharing (PS) allocation policy. These transition rates are also able to consider traditional delay nodes, and therefore a QN model with these transition rates is well-suited for multi-threaded software applications. In this report, we define the QN model and use the results in [1] to show that the transient sample paths of the QN model converge to the solution of a system of ordinary differential equations (ODEs). As the size of the ODE system grows linearly with the number of stations and job classes in the QN model, solving the ODE system becomes a scalable alternative to Markov chain representations.

1 The QN model

In our context, a fluid model is a continuous-time dynamical system described by a set of ordinary differential equations (ODEs) that approximates the evolution of a stochastic system with a Markovian description. Here the stochastic system is a closed class-switching QN with M stations and R job classes. In this type of QN the jobs can switch among the R classes, and therefore the total number of jobs in each class is not fixed. The state of the QN at time t is given by the vector $X(t) = \{X_{i,r}(t), 1 \leq i \leq M,\ 1 \leq r \leq R\}$, where $X_{i,r}(t)$ is the number of class-r jobs in station i at time t. We make use of the double index (i,r) for any vector to refer to its $(i-1)K+r$ entry, i.e., the one corresponding to station i and class-r jobs. We assume an initial population of N_r class-r jobs in the QN, and a total of $N = \sum_{r=1}^{R} N_r$ jobs. The system state is modified by events, which in our closed QN are limited to service completions. Once the service of a class-r job terminates in station i, the job proceeds to station j as a class-s job with probability $P_{i,j}^{r,s}$, triggering a transition from state $\boldsymbol{x}$ to state $\boldsymbol{x} + \boldsymbol{e}_{j,s} - \boldsymbol{e}_{i,r}$, where $\boldsymbol{e}_{i,r}$ is a vector of zeros with a one in entry (i,r).

We assume the service centers in the QN follow a PS scheduling policy. That is, the service rate is equally divided between all the jobs present at the station, which for station i is $x_i = \sum_{l=1}^{R} x_{i,l}$. Assuming that class-$r$ jobs are served at station i with rate $\mu_{i,r}$ by a *single* PS server, the transition rate associated to a jump $\boldsymbol{e}_{j,s} - \boldsymbol{e}_{i,r}$ would be

$$g(\boldsymbol{x}, \boldsymbol{e}_{j,s} - \boldsymbol{e}_{i,r}) = \mu_{i,r} P_{i,j}^{r,s} \frac{x_{i,r}}{x_i} \mathbb{1}\{x_i > 0\}. \quad (1)$$

This transition rate however poses both numerical and analytical issues, due to the discontinuity at $x_{i,k} = 0$ when $\sum_{l=1, l \neq k}^{K} x_{i,l} = 0$. Instead, we assume m_i servers in station i, and adopt the rate function

$$f(\boldsymbol{x}, \boldsymbol{e}_{j,s} - \boldsymbol{e}_{i,r}) = \mu_{i,r} P_{i,j}^{r,s} \frac{x_{i,r}}{x_i} \min\{m_i, x_i\}. \tag{2}$$

This rate function is such that, when the number of jobs is less than the number of servers ($x_i \leq m_i$), each of the jobs is assumed to be assigned to a different processor, such that each job class receives an effective processing rate proportional to the number of jobs in process, i.e., $\mu_{i,r} P_{i,j}^{r,s} x_{i,r}$. On the other hand, if there are more jobs than servers ($x_i \geq m_i$), the effective processing rate becomes that of a PS "super-processor" with processing rate $m_i \mu_{i,r}$, i.e., $m_i \mu_{i,r} P_{i,j}^{r,s} \frac{x_{i,r}}{x_i}$. Notice that, in addition to representing PS nodes, this transition rate also captures the dynamics of delay nodes. This is achieved by assigning N servers (as many as jobs in the QN) to the delay node, such that the service completion rate for class-r jobs in this station is $\mu_{i,r} P_{i,j}^{r,s} x_{i,r}$, and never adopts the PS form. Therefore, the result in the next section holds for a QN with delay and PS service centers. For later reference, we will use $f(\boldsymbol{x}, i, j, r, s)$ as shorthand for $f(\boldsymbol{x}, \boldsymbol{e}_{j,s} - \boldsymbol{e}_{i,r})$.

2 The fluid model

Up to this point we have described traditional closed class-switching QNs with a slight modification in the transition rates to account for both multi-server PS and delay stations. We now introduce a sequence of QN models, indexed by v, such that when $v \to \infty$, the sample paths of the QN models tend to that of an ODE system, which can be used to approximate the transient behavior of the QN. Let $\{X_v(t)\}_{v \in \mathbb{N}_+}$ be a sequence of QN models such that $X_1(t) = X(t)$ (the QN model defined in the previous section), and $X_v(t)$ for $v \geq 2$ is defined as $X_1(t)$ with an initial population of vN_r class-r jobs (vN jobs in total), and vm_i servers in station i. The state space of $X_v(t)$ is $\{\boldsymbol{x} \in \mathbb{N}^{MR} : \sum_{i=1}^{M} \sum_{r=1}^{R} x_{i,r} = vN\}$. With this definition, we can specify the number of servers in station i, for any QN model $X_v(t)$, as a fraction $0 < c_i \leq 1$ of the total number of jobs. Thus, the number of servers in station i in the QN model $X_v(t)$ is $vm_i = c_i vN = c_i \sum_{h=1}^{M} \sum_{r=1}^{R} x_{h,r}$, for any state $\boldsymbol{x}$ in the state space of $X_v(t)$. The transition rates of $X_v(t)$ in state $\boldsymbol{x}$ are thus given by

$$f_v(\boldsymbol{x}, i, j, r, s) = \mu_{i,r} P_{i,j}^{r,s} \frac{x_{i,r}}{x_i} \min\left\{ c_i \sum_{h=1}^{M} \sum_{r=1}^{R} x_{h,r}, x_i \right\}, \tag{3}$$

and can be written as

$$\begin{aligned} f_v(\boldsymbol{x}, i, j, r, s) &= v\mu_{i,r} P_{i,j}^{r,s} \frac{\frac{x_{i,r}}{v}}{\frac{x_i}{v}} \min\left\{ \frac{c_i}{v} \sum_{h=1}^{M} \sum_{r=1}^{R} x_{h,r}, \frac{x_i}{v} \right\} \\ &= v f_1(\boldsymbol{x}/v, i, j, r, s) = v f(\boldsymbol{x}/v, i, j, r, s). \end{aligned}$$

This property, that the rates of $X_v(t)$ in state $\boldsymbol{x}$ can be written as v times the rates of $X_1(t)$ in state $\boldsymbol{x}/v$, and the fact that the functions $f(\boldsymbol{x}, i, j, r, s)$ are continuous for all $\boldsymbol{x} \in \mathbb{R}^{MR}$, makes the sequence $\{X_v(t)\}_{v \in \mathbb{N}_+}$ a *density-dependent* family of processes [1].

Theorem 3.1 in [1] shows that, under certain conditions, the sample paths of the *normalized* sequence $\{X_v(t)/v\}_{v \in \mathbb{N}_+}$ converge in probability to a deterministic ODE system. For any $\boldsymbol{x} \in \mathbb{R}^{MR}$, let $F(\boldsymbol{x})$ be the drift of $X(t)$ in state $\boldsymbol{x}$, that is

$$F(\boldsymbol{x}) = \sum_{i=1}^{M} \sum_{j=1}^{M} \sum_{r=1}^{R} \sum_{s=1}^{R} (\boldsymbol{e}_{j,s} - \boldsymbol{e}_{i,r}) f(\boldsymbol{x}, i, j, r, s), \tag{4}$$

and let $\boldsymbol{x}(t) \in \mathbb{R}^{MR}$ be the state of a deterministic system that evolves according to the ODE

$$\frac{d\boldsymbol{x}(t)}{dt} = F(\boldsymbol{x}(t)), t \geq 0, \tag{5}$$

and has initial state $\boldsymbol{x}_0$. Theorem 3.1 in [1] states that the sample paths of the sequence $\{X_v(t)/v\}_{v\in\mathbb{N}_+}$ converge to the deterministic limit $\boldsymbol{x}(t)$ as $v \to \infty$, i.e., that $\lim_{v\to\infty} X_v(0)/v = \boldsymbol{x}_0$ implies that for every $\delta > 0$

$$\lim_{v\to\infty} \mathrm{P}\left(\sup_{s\leq t}\left|\frac{1}{v}X_v(s) - \boldsymbol{x}(s)\right| > \delta\right) = 0.$$

This holds for every finite t if $\boldsymbol{x}(s) \in E$ for $0 \leq s \leq t$, where E is an open set $E \subset \mathbb{R}^{MR}$ such that

$$|F(\boldsymbol{x}) - F(\boldsymbol{y})| < M_E\,|\boldsymbol{x} - \boldsymbol{y}|\,,\ \boldsymbol{x}, \boldsymbol{y} \in E, \tag{6}$$

$$\sup_{\boldsymbol{x}\in E}\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{r=1}^{R}\sum_{s=1}^{R}|\boldsymbol{e}_{j,s} - \boldsymbol{e}_{i,r}|\,f(\boldsymbol{x}, i, j, r, s) < \infty, \tag{7}$$

$$\lim_{d\to\infty}\sup_{\boldsymbol{x}\in E}\sum_{(i,j,r,s)\in S(d)}|\boldsymbol{e}_{j,s} - \boldsymbol{e}_{i,r}|\,f(\boldsymbol{x}, i, j, r, s) = 0, \tag{8}$$

where $S(d)$ is the set $\{(i,j,r,s) : |\boldsymbol{e}_{j,s} - \boldsymbol{e}_{i,r}| > d\}$, and M_E is a constant. Notice that condition (6) implies the Lipschitz continuity of $F(\cdot)$ which ensures the existence of a unique solution to the ODE (5).

Theorem 1. *The sequence of QNs* $\{X_v(t)\}_{v\in\mathbb{N}_+}$ *verifies conditions* (6)*,* (7)*, and* (8)*.*

Proof. As the QN model $X(t)$ is closed, the entries $X_{i,r}(t)$ are bounded above by N, a condition that also holds for every $X_v(t)/v$ as well as for $\boldsymbol{x}(t)$. As a result the set E can be chosen as the smallest open set that contains the set $\{\boldsymbol{x} \in \mathbb{R}^{MR} : \boldsymbol{x} \geq 0, \sum_{i=1}^{M}\sum_{r=1}^{R} x_{i,r} = N\}$, as the sample paths of $\boldsymbol{x}(t)$ never leave this set.

To verify condition (6) we consider $\boldsymbol{x}, \boldsymbol{y} \in E$ and the 1-norm to write

$$\begin{aligned}
&|F(\boldsymbol{x}) - F(\boldsymbol{y})| \\
&\quad= \left|\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{r=1}^{R}\sum_{s=1}^{R}(\boldsymbol{e}_{j,s} - \boldsymbol{e}_{i,r})f(\boldsymbol{x}, i, j, r, s) - \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{r=1}^{R}\sum_{s=1}^{R}(\boldsymbol{e}_{j,s} - \boldsymbol{e}_{i,r})f(\boldsymbol{y}, i, j, r, s)\right| \\
&\quad\leq 2\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{r=1}^{R}\sum_{s=1}^{R}|f(\boldsymbol{x}, i, j, r, s) - f(\boldsymbol{y}, i, j, r, s)|\,,
\end{aligned}$$

since $|\boldsymbol{e}_{j,s} - \boldsymbol{e}_{i,r}| \leq 2$. We can therefore focus on each of the terms within the sum, for which we have four cases depending of whether $\sum_{r=1}^{R} x_{i,r}$ and $\sum_{r=1}^{R} y_{i,r}$ are greater or less than m_i.

Case 1: $\sum_{r=1}^{R} x_{i,r} > m_i$ and $\sum_{r=1}^{R} y_{i,r} > m_i$.

$$\begin{aligned}
|f(\boldsymbol{x},i,j,r,s) - f(\boldsymbol{y},i,j,r,s)| &= \left|\mu_{i,r} P_{i,j}^{r,s} m_i\right| \left| \frac{x_{i,r}}{\sum_{l=1}^{R} x_{i,l}} - \frac{y_{i,r}}{\sum_{l=1}^{R} y_{i,l}} \right| \\
&< \frac{\mu_{i,r} P_{i,j}^{r,s}}{m_i} \left| x_{i,r} \sum_{l=1}^{R} y_{i,l} - y_{i,r} \sum_{l=1}^{R} x_{i,l} \right| \\
&= \frac{\mu_{i,r} P_{i,j}^{r,s}}{m_i} \left| x_{i,r} \left(\sum_{l=1}^{R} y_{i,l} - \sum_{l=1}^{R} x_{i,l} \right) - (y_{i,r} - x_{i,r}) \sum_{l=1}^{R} x_{i,l} \right| \\
&\leq \frac{N \mu_{i,r} P_{i,j}^{r,s}}{m_i} \sum_{l=1}^{R} \left| x_{i,l} - y_{i,l} \right|.
\end{aligned}$$

Case 2: $\sum_{r=1}^{R} x_{i,r} > m_i$ and $\sum_{r=1}^{R} y_{i,r} \leq m_i$.

$$\begin{aligned}
|f(\boldsymbol{x},i,j,r,s) - f(\boldsymbol{y},i,j,r,s)| &= \left|\mu_{i,r} P_{i,j}^{r,s}\right| \left| \frac{x_{i,r}}{\sum_{l=1}^{R} x_{i,l}} m_i - y_{i,r} \right| \\
&= \mu_{i,r} P_{i,j}^{r,s} \left| \frac{m_i x_{i,r} - y_{i,r} \sum_{l=1}^{R} x_{i,l}}{\sum_{l=1}^{R} x_{i,l}} \right| \\
&< \frac{\mu_{i,r} P_{i,j}^{r,s}}{m_i} \left| m_i x_{i,r} - y_{i,r} \sum_{l=1}^{R} x_{i,l} \right| \\
&< \frac{\mu_{i,r} P_{i,j}^{r,s}}{m_i} \left| m_i x_{i,r} - m_i y_{i,r} \right| \\
&= \mu_{i,r} P_{i,j}^{r,s} \left| x_{i,r} - y_{i,r} \right|.
\end{aligned}$$

Case 3: $\sum_{r=1}^{R} x_{i,r} \leq m_i$ and $\sum_{r=1}^{R} y_{i,r} > m_i$.

$$\begin{aligned}
|f(\boldsymbol{x},i,j,r,s) - f(\boldsymbol{y},i,j,r,s)| &= \left|\mu_{i,r} P_{i,j}^{r,s}\right| \left| x_{i,r} - \frac{y_{i,r}}{\sum_{l=1}^{R} y_{i,l}} m_i \right| \\
&= \mu_{i,r} P_{i,j}^{r,s} \left| \frac{x_{i,r} \sum_{l=1}^{R} y_{i,l} - m_i y_{i,r}}{\sum_{l=1}^{R} y_{i,l}} \right| \\
&< \frac{\mu_{i,r} P_{i,j}^{r,s}}{m_i} \left| x_{i,r} \sum_{l=1}^{R} y_{i,l} - m_i y_{i,r} \right| \\
&= \frac{\mu_{i,r} P_{i,j}^{r,s}}{m_i} \left| x_{i,r} \left(\sum_{l=1}^{R} y_{i,l} - \sum_{l=1}^{R} x_{i,l} \right) + x_{i,r} \sum_{l=1}^{R} x_{i,l} - m_i y_{i,r} \right| \\
&\leq \frac{\mu_{i,r} P_{i,j}^{r,s}}{m_i} \left| m_i \left(\sum_{l=1}^{R} y_{i,l} - \sum_{l=1}^{R} x_{i,l} \right) + m_i (x_{i,r} - y_{i,r}) \right| \\
&\leq \mu_{i,r} P_{i,j}^{r,s} \sum_{l=1}^{R} \left| x_{i,l} - y_{i,l} \right|.
\end{aligned}$$

Case 4: $\sum_{r=1}^{R} x_{i,r} \leq m_i$ and $\sum_{r=1}^{R} y_{i,r} \leq m_i$.

$$|f(\boldsymbol{x},i,j,r,s) - f(\boldsymbol{y},i,j,r,s)| = \mu_{i,r} P_{i,j}^{r,s} \left| x_{i,r} - y_{i,r} \right|.$$

We therefore have that, using the 1-norm, $|f(\boldsymbol{x}, i, j, r, s) - f(\boldsymbol{y}, i, j, r, s)| \leq \mu_{i,r} P_{i,j}^{r,s} N/m_i |\boldsymbol{x} - \boldsymbol{y}|$, as $m_i \leq N$, and therefore condition (6) is satisfied.

Condition (7) can be readily verified by noticing that $|\boldsymbol{e}_{j,r} - \boldsymbol{e}_{i,r}| \leq 2$ and

$$f(\boldsymbol{x}, i, j, r, s) \leq \mu_{i,r} P_{i,j}^{r,s} x_{i,r} \leq \mu_{i,r} P_{i,j}^{r,s} N.$$

Finally, condition (8) is verified by observing that the set $S(d)$ is empty for every $d > 2$. □

References

[1] Thomas G. Kurtz. Solutions of ordinary differential equations as limits of pure jump Markov processes. *Journal of Applied Probability*, 7:49–58, 1970.