

## Appendix A: Transition rate matrix for subnetwork $i$

Table 1 details the function  $f_{\bar{Q}}$  of Equation (13), i.e., it details the full transition rate matrix of subnetwork  $i$ , which is composed of queues indexed by  $i$ ,  $i + 1$ , and  $i + 2$ . This table enumerates all non-zero and non-diagonal elements of the transition rate matrix. The table considers an initial joint aggregate state  $s = (j_i, j_{i+1}, j_{i+2})$ . Each row considers a possible transition. The first column describes the new state, the second column gives the conditions on the initial state  $s$  under which such a transition can take place. The third column gives the transition rate.

The transitions are grouped into six sets according to the event that triggered the transition. The first three sets consider transitions that arise following external arrivals (i.e., arrivals from outside the subnetwork) to queues  $i$ ,  $i + 1$  and  $i + 2$ , respectively. An external arrival to a given queue can cause the queue to transition from aggregate state 0 (resp. 1) to aggregate state 1 (resp. 2). Upon an external arrival to, for instance, queue  $i$ , the transition from aggregate state 0 to 1 occurs with probability 1 and the transition from aggregate state 1 to 2 occurs if queue  $i$  is in the disaggregate state  $\ell_i - 1$ . The probability of queue  $i$  being in state  $\ell_i - 1$  given that it is in aggregate state 1 is a disaggregation probability. Recall, that the proposed method uses state-dependent disaggregation probabilities. Hence, depending on the states of queues  $i + 1$  and  $i + 2$  the disaggregation probability is given by  $\alpha_{i,1,\ell_i-1}^k$ ,  $\alpha_{i,2,\ell_i-1}^k$  or  $\alpha_{i,3,\ell_i-1}^k$ .

The fourth set considers a service completion at queue  $i$ . Such an event can cause queue  $i$  to transition from aggregate state 1 (resp. 2) to 0 (resp. 1). Upon service completion, the transition from state 2 to 1 occurs with probability 1 and the transition from 1 to 0 occurs if queue  $i$  is in the disaggregate state 1. The latter is captured by the state-dependent disaggregation probability  $\alpha_{i,1,1}^k$ . Additionally, a service completion at queue  $i$  can cause queue  $i + 1$  to transition from aggregate state 0 (resp. 1) to 1 (resp. 2), which occurs with probability 1 (resp.  $\alpha_{i,4,\ell_{i+1}-1}^k$  or  $\alpha_{i,5,\ell_{i+1}-1}^k$ ).

The fifth set considers a service completion at queue  $i + 1$ . The rates are obtained through similar reasoning as for service completion at queue  $i$ . Additionally, if a job at queue  $i$  is being blocked by queue  $i + 1$ , then a service completion at queue  $i + 1$  may trigger a change in the state of queue  $i$ . This is described via the blocking probability  $\beta_{i,1}$ . More specifically, if queue  $i$  is blocked by queue  $i + 1$ , then a service completion at queue  $i + 1$  will:

1. send the job that has completed service at queue  $i + 1$  to queue  $i + 2$ , which may lead queue  $i + 2$  to transition from aggregate state 0 (resp. 1) to 1 (resp. 2);
2. unblock a job at queue  $i$ , which may lead queue  $i$  to transition from aggregate state 1 (resp. 2) to 0 (resp. 1);
3. have no impact on the state of queue  $i + 1$  (since an arrival and a departure occur simultaneously).

The final set considers a service completion at queue  $i + 2$ . The rates are obtained through similar reasoning as for service completions at queue  $i + 1$ . Since queue  $i + 2$  may block both queue  $i + 1$  and queue  $i$ , then a service completion at queue  $i + 2$  may trigger changes in the states of both queues  $i$  and  $i + 1$ . This unblocking is described via the blocking probabilities  $\beta_{i,2}$ ,  $\beta_{i,3}$  and  $\beta_{i,4}$ .

Table 1: Transition rate matrix for subnetwork  $i$  during time interval  $k$ . Enumeration of all possible transitions assuming an initial joint aggregate state  $s = (j_i, j_{i+1}, j_{i+2})$  and a new state  $t$ .

Arrival to queue $i$		
New state $t$	Initial conditions	Rate
$(j_i + 1, j_{i+1}, j_{i+2})$	$j_i = 0$	$\hat{\gamma}_i^k$
$(j_i + 1, j_{i+1}, j_{i+2})$	$j_i = 1, j_{i+1} = \{0, 1\}$	$\hat{\gamma}_i^k \alpha_{i,1,\ell_i-1}^k$
$(j_i + 1, j_{i+1}, j_{i+2})$	$j_i = 1, j_{i+1} = 2, j_{i+2} = \{0, 1\}$	$\hat{\gamma}_i^k \alpha_{i,2,\ell_i-1}^k$
$(j_i + 1, j_{i+1}, j_{i+2})$	$j_i = 1, j_{i+1} = 2, j_{i+2} = 2$	$\hat{\gamma}_i^k \alpha_{i,3,\ell_i-1}^k$
External arrival to queue $i + 1$		
New state $t$	Initial conditions	Rate
$(j_i, j_{i+1} + 1, j_{i+2})$	$j_{i+1} = 0$	$\gamma_{i+1}$
$(j_i, j_{i+1} + 1, j_{i+2})$	$j_{i+1} = 1, j_{i+2} = \{0, 1\}$	$\gamma_{i+1} \alpha_{i,4,\ell_{i+1}-1}^k$
$(j_i, j_{i+1} + 1, j_{i+2})$	$j_{i+1} = 1, j_{i+2} = 2$	$\gamma_{i+1} \alpha_{i,5,\ell_{i+1}-1}^k$
External arrival to queue $i + 2$		
New state $t$	Initial conditions	Rate
$(j_i, j_{i+1}, j_{i+2} + 1)$	$j_{i+2} = 0$	$\gamma_{i+2}$
$(j_i, j_{i+1}, j_{i+2} + 1)$	$j_{i+2} = 1$	$\gamma_{i+2} \alpha_{i,6,\ell_{i+1}-1}^k$
Service completion at queue $i$		
New state $t$	Initial conditions	Rate
$(j_i, j_{i+1} + 1, j_{i+2})$	$j_i = 1, j_{i+1} = 0$	$\mu_i (1 - \alpha_{i,1,1}^k)$
$(j_i - 1, j_{i+1} + 1, j_{i+2})$	$j_i = 1, j_{i+1} = 0$	$\mu_i \alpha_{i,1,1}^k$
$(j_i, j_{i+1} + 1, j_{i+2})$	$j_i = 1, j_{i+1} = 1, j_{i+2} = \{0, 1\}$	$\mu_i (1 - \alpha_{i,1,1}^k) \alpha_{i,4,\ell_{i+1}-1}^k$
$(j_i - 1, j_{i+1}, j_{i+2})$	$j_i = 1, j_{i+1} = 1, j_{i+2} = \{0, 1\}$	$\mu_i \alpha_{i,1,1}^k (1 - \alpha_{i,4,\ell_{i+1}-1}^k)$
$(j_i - 1, j_{i+1} + 1, j_{i+2})$	$j_i = 1, j_{i+1} = 1, j_{i+2} = \{0, 1\}$	$\mu_i \alpha_{i,1,1}^k \alpha_{i,4,\ell_{i+1}-1}^k$
$(j_i, j_{i+1} + 1, j_{i+2})$	$j_i = 1, j_{i+1} = 1, j_{i+2} = 2$	$\mu_i (1 - \alpha_{i,1,1}^k) \alpha_{i,5,\ell_{i+1}-1}^k$
$(j_i - 1, j_{i+1}, j_{i+2})$	$j_i = 1, j_{i+1} = 1, j_{i+2} = 2$	$\mu_i \alpha_{i,1,1}^k (1 - \alpha_{i,5,\ell_{i+1}-1}^k)$
$(j_i - 1, j_{i+1} + 1, j_{i+2})$	$j_i = 1, j_{i+1} = 1, j_{i+2} = 2$	$\mu_i \alpha_{i,1,1}^k \alpha_{i,5,\ell_{i+1}-1}^k$
$(j_i - 1, j_{i+1} + 1, j_{i+2})$	$j_i = 2, j_{i+1} = 0$	$\mu_i$
$(j_i - 1, j_{i+1} + 1, j_{i+2})$	$j_i = 2, j_{i+1} = 1, j_{i+2} = \{0, 1\}$	$\mu_i \alpha_{i,4,\ell_{i+1}-1}^k$
$(j_i - 1, j_{i+1}, j_{i+2})$	$j_i = 2, j_{i+1} = 1, j_{i+2} = \{0, 1\}$	$\mu_i (1 - \alpha_{i,4,\ell_{i+1}-1}^k)$
$(j_i - 1, j_{i+1} + 1, j_{i+2})$	$j_i = 2, j_{i+1} = 1, j_{i+2} = 2$	$\mu_i \alpha_{i,5,\ell_{i+1}-1}^k$
$(j_i - 1, j_{i+1}, j_{i+2})$	$j_i = 2, j_{i+1} = 1, j_{i+2} = 2$	$\mu_i (1 - \alpha_{i,5,\ell_{i+1}-1}^k)$
Service completion at queue $i + 1$		
New state $t$	Initial conditions	Rate
$(j_i, j_{i+1}, j_{i+2} + 1)$	$j_{i+1} = 1, j_{i+2} = 0$	$\mu_{i+1} (1 - \alpha_{i,4,1}^k)$
$(j_i, j_{i+1} - 1, j_{i+2} + 1)$	$j_{i+1} = 1, j_{i+2} = 0$	$\mu_{i+1} \alpha_{i,4,1}^k$
$(j_i, j_{i+1}, j_{i+2} + 1)$	$j_{i+1} = 1, j_{i+2} = 1$	$\mu_{i+1} (1 - \alpha_{i,4,1}^k) \alpha_{i,6,\ell_{i+2}-1}^k$
$(j_i, j_{i+1} - 1, j_{i+2})$	$j_{i+1} = 1, j_{i+2} = 1$	$\mu_{i+1} \alpha_{i,4,1}^k (1 - \alpha_{i,6,\ell_{i+2}-1}^k)$
$(j_i, j_{i+1} - 1, j_{i+2} + 1)$	$j_{i+1} = 1, j_{i+2} = 1$	$\mu_{i+1} \alpha_{i,4,1}^k \alpha_{i,6,\ell_{i+2}-1}^k$
$(j_i, j_{i+1} - 1, j_{i+2} + 1)$	$j_i = 0, j_{i+1} = 2, j_{i+2} = 0$	$\mu_{i+1}$
$(j_i, j_{i+1} - 1, j_{i+2})$	$j_i = 0, j_{i+1} = 2, j_{i+2} = 1$	$\mu_{i+1} (1 - \alpha_{i,6,\ell_{i+2}-1}^k)$
$(j_i, j_{i+1} - 1, j_{i+2} + 1)$	$j_i = 0, j_{i+1} = 2, j_{i+2} = 1$	$\mu_{i+1} \alpha_{i,6,\ell_{i+2}-1}^k$
$(j_i, j_{i+1} - 1, j_{i+2} + 1)$	$j_i = \{1, 2\}, j_{i+1} = 2, j_{i+2} = 0$	$\mu_{i+1} (1 - \beta_{i,1})$
$(j_i, j_{i+1} - 1, j_{i+2})$	$j_i = \{1, 2\}, j_{i+1} = 2, j_{i+2} = 1$	$\mu_{i+1} (1 - \alpha_{i,6,\ell_{i+2}-1}^k) (1 - \beta_{i,1})$
$(j_i, j_{i+1} - 1, j_{i+2} + 1)$	$j_i = \{1, 2\}, j_{i+1} = 2, j_{i+2} = 1$	$\mu_{i+1} \alpha_{i,6,\ell_{i+2}-1}^k (1 - \beta_{i,1})$
$(j_i, j_{i+1}, j_{i+2} + 1)$	$j_i = 1, j_{i+1} = 2, j_{i+2} = 0$	$\mu_{i+1} (1 - \alpha_{i,2,1}^k) \beta_{i,1}$

$(j_i - 1, j_{i+1}, j_{i+2} + 1)$	$j_i = 1, j_{i+1} = 2, j_{i+2} = 0$	$\mu_{i+1} \alpha_{i,2,1}^k \beta_{i,1}$
$(j_i, j_{i+1}, j_{i+2} + 1)$	$j_i = 1, j_{i+1} = 2, j_{i+2} = 1$	$\mu_{i+1} (1 - \alpha_{i,2,1}^k) \alpha_{i,6,\ell_{i+2}-1}^k \beta_{i,1}$
$(j_i - 1, j_{i+1}, j_{i+2})$	$j_i = 1, j_{i+1} = 2, j_{i+2} = 1$	$\mu_{i+1} \alpha_{i,2,1}^k (1 - \alpha_{i,6,\ell_{i+2}-1}^k) \beta_{i,1}$
$(j_i - 1, j_{i+1}, j_{i+2} + 1)$	$j_i = 1, j_{i+1} = 2, j_{i+2} = 1$	$\mu_{i+1} \alpha_{i,2,1}^k \alpha_{i,6,\ell_{i+2}-1}^k \beta_{i,1}$
$(j_i, j_{i+1} - 1, j_{i+2} + 1)$	$j_i = 2, j_{i+1} = 2, j_{i+2} = 0$	$\mu_{i+1} (1 - \beta_{i,1})$
$(j_i, j_{i+1} - 1, j_{i+2} + 1)$	$j_i = 2, j_{i+1} = 2, j_{i+2} = 1$	$\mu_{i+1} \alpha_{i,6,\ell_{i+2}-1}^k (1 - \beta_{i,1})$
$(j_i, j_{i+1} - 1, j_{i+2})$	$j_i = 2, j_{i+1} = 2, j_{i+2} = 1$	$\mu_{i+1} (1 - \alpha_{i,6,\ell_{i+2}-1}^k) (1 - \beta_{i,1})$
$(j_i - 1, j_{i+1}, j_{i+2} + 1)$	$j_i = 2, j_{i+1} = 2, j_{i+2} = 0$	$\mu_{i+1} \beta_{i,1}$
$(j_i - 1, j_{i+1}, j_{i+2} + 1)$	$j_i = 2, j_{i+1} = 2, j_{i+2} = 1$	$\mu_{i+1} \alpha_{i,6,\ell_{i+2}-1}^k \beta_{i,1}$
$(j_i - 1, j_{i+1}, j_{i+2})$	$j_i = 2, j_{i+1} = 2, j_{i+2} = 1$	$\mu_{i+1} (1 - \alpha_{i,6,\ell_{i+2}-1}^k) \beta_{i,1}$

---

Service completion at queue  $i + 2$

---

New state $t$	Initial conditions	Rate
$(j_i, j_{i+1}, j_{i+2} - 1)$	$j_{i+2} = 1$	$\hat{\mu}_{i+2}^k \alpha_{i,6,1}^k$
$(j_i, j_{i+1}, j_{i+2} - 1)$	$j_{i+1} = 0, j_{i+2} = 2$	$\hat{\mu}_{i+2}^k$
$(j_i, j_{i+1}, j_{i+2} - 1)$	$j_{i+1} = \{1, 2\}, j_{i+2} = 2$	$\hat{\mu}_{i+2}^k (1 - \beta_{i,3})$
$(j_i, j_{i+1} - 1, j_{i+2})$	$j_{i+1} = 1, j_{i+2} = 2$	$\hat{\mu}_{i+2}^k \alpha_{i,5,1}^k \beta_{i,3}$
$(j_i, j_{i+1} - 1, j_{i+2})$	$j_i = 0, j_{i+1} = 2, j_{i+2} = 2$	$\hat{\mu}_{i+2}^k \beta_{i,3}$
$(j_i, j_{i+1} - 1, j_{i+2})$	$j_i = \{1, 2\}, j_{i+1} = 2, j_{i+2} = 2$	$\hat{\mu}_{i+2}^k \beta_{i,4}$
$(j_i - 1, j_{i+1}, j_{i+2})$	$j_i = 1, j_{i+1} = 2, j_{i+2} = 2$	$\hat{\mu}_{i+2}^k \alpha_{i,3,1}^k \beta_{i,2}$
$(j_i - 1, j_{i+1}, j_{i+2})$	$j_i = 2, j_{i+1} = 2, j_{i+2} = 2$	$\hat{\mu}_{i+2}^k \beta_{i,2}$

## Appendix B: Marginal finite capacity queueing model

We briefly outline here the formulation of the queueing-theoretic urban network model of Chapter 4 of Osorio (2010). The method is a stationary decomposition technique that decomposes the network into single queue subnetworks, as described in Osorio and Bierlaire (2009). The notation used in this model for a given queue  $i$  is listed in Table 2. Its formulation consists of the System of Equations (1). The model approximates the marginal stationary distribution of each queue. It accounts only for first-order between-queue dependency information.

$\gamma_i$	external arrival rate;
$\lambda_i$	total arrival rate;
$\mu_i$	service rate of a server;
$\tilde{\mu}_i$	unblocking rate;
$\hat{\mu}_i$	effective service rate;
$P_i^f$	probability of being blocked at queue $i$ ;
$p_{ij}$	routing probability from queue $i$ to queue $j$ ;
$\ell_i$	space capacity;
$N_i$	number of vehicles in queue $i$ ;
$P(N_i = \ell_i)$	probability that queue $i$ is full;
$\mathcal{I}_i^+$	set of downstream queues to queue $i$ .

**Table 2** List of variables used in marginal finite capacity queueing model.

$$\left\{ \begin{array}{l} \lambda_i = \gamma_i + \frac{\sum_j p_{ij} \lambda_j P(N_j < \ell_j)}{P(N_i < \ell_i)} \quad (1a) \\ \tilde{\mu}_i = \sum_{j \in \mathcal{X}_i^+} \frac{\lambda_j P(N_j < \ell_j)}{\lambda_i P(N_i < \ell_i) \hat{\mu}_j} \quad (1b) \\ \frac{1}{\hat{\mu}_i} = \frac{1}{\mu_i} + P_i^f \frac{1}{\tilde{\mu}_i} \quad (1c) \\ P_i^f = \sum_j p_{ij} P(N_j = \ell_j) \quad (1d) \\ P(N_i = \ell_i) = \frac{1 - \rho_i}{1 - \rho_i^{\ell_i + 1}} \rho_i^{\ell_i} \quad (1e) \\ \rho_i = \frac{\lambda_i}{\hat{\mu}_i}. \quad (1f) \end{array} \right.$$

The exogenous parameters are  $\gamma_i$ ,  $p_{ij}$ ,  $\ell_i$ , and  $\mu_i$ . All other variables are endogenous. Equation (1a) is a flow conservation equation as applied to a loss (finite capacity) queueing model. It corresponds to Equations (2)-(3) of Osorio and Bierlaire (2009). Equation (1b) defines the unblocking rate of a queue that is blocked, it corresponds to Equation (7) of Osorio and Bierlaire (2009). Equation (1c) defines the effective service rate, which accounts for both (exogenous) service,  $\mu_i$ , and blocking,  $\tilde{\mu}_i$ , it corresponds to Equation (8) of Osorio and Bierlaire (2009). Equation (1d) approximates the probability of being blocked at queue  $i$  by averaging the probabilities of downstream queues being full (which are called blocking probabilities). It corresponds to Equation (4) of Osorio and Bierlaire (2009). The blocking probability of queue  $i$  is given in Equation (1e) by the closed-form expression of the stationary probability of being full of an M/M/1/ $\ell$  queue (e.g., Bocharov *et al.* 2004). Equation (1f) defines the traffic intensity,  $\rho_i$ , of a finite capacity single server queue.

## References

- Bocharov, P. P., D'Apice, C., Pechinkin, A. V., and Salerno, S. (2004). *Queueing theory*, chapter 3, pages 96–98. Modern Probability and Statistics. Brill Academic Publishers, Zeist, The Netherlands.
- Osorio, C. (2010). *Mitigating network congestion: analytical models, optimization methods and their applications*. Ph.D. thesis, Ecole Polytechnique Fédérale de Lausanne.
- Osorio, C. and Bierlaire, M. (2009). An analytic finite capacity queueing network model capturing the propagation of congestion and blocking. *European Journal of Operational Research*, **196**(3), 996–1007.