# Online Companion for Paper: "The restricted Erlang-R Queue: Finite-size effects in service systems with returning customers"

Galit B. Yom-Tov
Faculty of Data and Decision Sciences, Technion — Israel Institute of Technology, 32000 Haifa, Israel, gality@technion.ac.il

Fiona Sloothaak
Department of Mathematics and Computer Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB, Eindhoven, {f.sloothaak}@tue.nl

Johan S.H. van Leeuwaarden
Tilburg School of Economics and Management, Tilburg University, Tilburg, Netherlands, {J.S.H.vanLeeuwaarden@tilburguniversity.edu

## Appendix A: Description of the QBD process

### A.1. The QBD-process

We consider the QBD-process $X = \{N, Q_1\}$ in stationarity. Let $\nu(i) = \min\{i, s\}\mu$. To determine the (outgoing) transition rates of the process $X$ we distinguish between the following cases:

- *Transitions from* $(0,0)$*:* There are no patients in the Emergency Department and thus the only possible occurrence is when a new patient arrives. This results in a transition to $(1,1)$ and occurs with rate $\lambda$.
- *Transitions from* $(i,0), 1 \leq i < n$*:* There are exactly $i$ patients assigned to a bed of which none are seen by a nurse. Then either one of those patients becomes needy, or a new patient arrives at the Emergency Department that can immediately be seen by a nurse. The first results in a transition to $(i,1)$ and occurs at rate $i\delta$, and the second results in a transition to $(i+1,1)$ and occurs with rate $\lambda$.
- *Transitions from* $(i,0), i \geq n$*:* Again, the only possible transitions arises from either a newly arrived patient or a patient assigned to a bed becoming needy. However, a newly arrived patient finds all beds occupied and needs to wait. Thus, with rate $\lambda$ we have a transition to $(i+1,0)$ and with rate $n\delta$ a transition to $(i,1)$.
- *Transitions from* $(i,i), i < n$*:* In this case all patients assigned to a bed are in need of service. With rate $\lambda$ a new patient arrives at the Emergency Department. She joins the (possible) queue to be seen by a nurse immediately, so this results in a transition to $(i+1,i+1)$. Moreover, since there are only $s < n$ nurses, a service completion occurs with rate $\nu(i)$. With probability $p$ the patient turns to the holding phase, so in total we still have $i$ patients with one patient less in queue for a nurse. With probability $1-p$ the patient leaves the Emergency Department, decreasing both $N$ and $Q_1$ by one. In other words, with rate $p\nu(i)$ we have a transition to $(i,i-1)$ and with rate $(1-p)\nu(i)$ we have a transition to $(i-1,i-1)$.

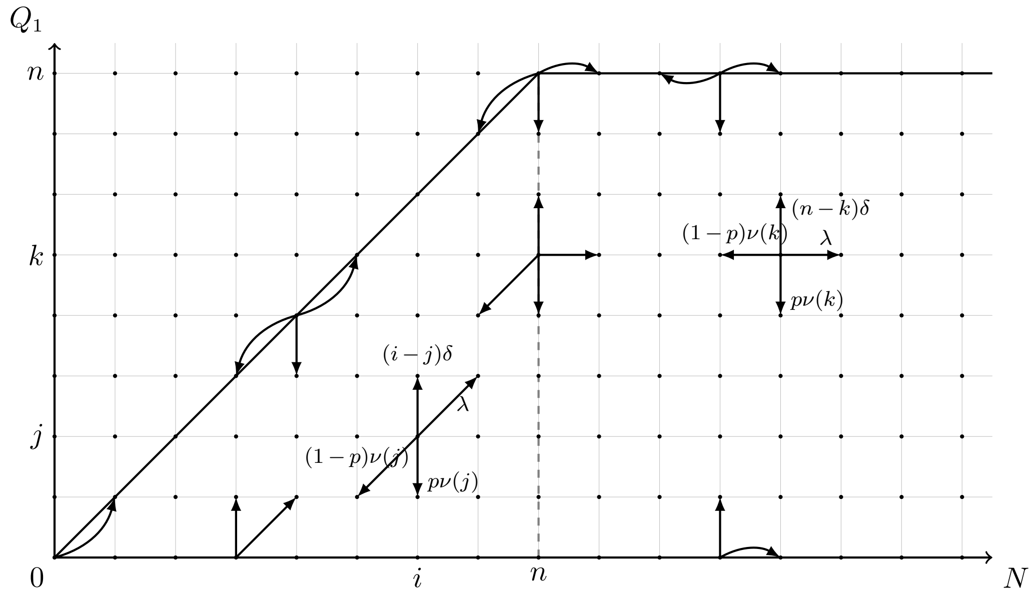


**Figure 1 Illustration of state space and the transitions for the Erlang-R model with holding.**

- *Transitions from* $(n,n)$*:* Similar to the previous case, we have a transition to $(n,n-1)$ with rate $ps\mu$ and with rate $(1-p)s\mu$ we have a transition to $(n-1,n-1)$. In this case however, a newly arrived patient finds all beds occupied, resulting in a transition to $(n+1,n)$ with rate $\lambda$.
- *Transitions from* $(i,n), i>n$*:* We have a transition to $(i+1,n)$ with rate $\lambda$ and a transition to $(i,n-1)$ with rate $ps\mu$. In case that a patient leaves the Emergency Department there are $i-n>0$ patients in the holding room waiting for an available bed. Thus, one of the $i-n$ patients in the holding room is assigned to the available bed in need of service. That is, with rate $(1-p)s\mu$ we have a transition to $(i-1,n)$.
- *Transitions from* $(i,j), 1\le j<i<n$*:* There are four possible transitions. First, with rate $\lambda$ there is a new arrival which results in a transition to $(i+1,j+1)$. Second, with rate $(i-j)\delta$ a patient in one of the beds becomes needy, which results in a transition to $(i,j+1)$. Third, with rate $p\nu(j)$ a patient turns to the content state after service completion, which results in a transition to $(i,j-1)$. Last, with rate $(1-p)\nu(j)$ a patient leaves the Emergency Department after service completion, which results in a transition to $(i-1,j-1)$.
- *Transitions from* $(n,j), 1\le j<n$*:* This case is similar to the previous one. The only difference arises when a new patient arrives, since all $n$ beds are already occupied. Thus, with rate $\lambda$ we have a transition to $(n+1,j)$.
- *Transitions from* $(i,j), i>n, 1\le j\le n$*:* This case is the previous one, except when a patient leaves the Emergency Department after service completion. Then one of the $(i-n)$ patients in the holding room will be assigned to a bed in need of service. This results in a transition to $(i-1,j)$ with rate $(1-p)\nu(j)$.

The state space and transition rates of the Erlang-R model with holding are illustrated in Figure 1.

The state space can be partitioned according to its levels, where level $i$ corresponds to a total queue length $N=i$ patients. This results in an infinite-sized matrix consisting of blocks, where each block corresponds to the transition flow from one level to another. Since the only transitions allowed are within the same level or between two adjacent levels in a QBD-process, we obtain a tridiagonal block structure. Each block consists of elements representing the transition rate of one state to another, and therefore each block is a matrix of size at most $(n+1)\times(n+1)$.

For the Erlang-R model with holding this gives the following result. Let $P$ denote the transition matrix of the process $\{N(t), Q_1(t)\}$. We have the boundary levels $\{1,2,...,n\}$ and $P$ is of the form

$$P=\begin{pmatrix} B_{00} & B_{01} & & & & & & & & \\ B_{10} & B_{11} & B_{12} & & & & & & & \\ & B_{21} & B_{22} & B_{23} & & & & & & \\ & & \ddots & \ddots & \ddots & & & & & \\ & & & & & B_{n\,n-1} & & & & \\ & & & & B_{n-1\,n} & B_{nn} & A_0 & & & \\ & & & & & A_2 & A_1 & A_0 & & \\ & & & & & & A_2 & A_1 & A_0 & \\ & & & & & & & \ddots & \ddots & \ddots \end{pmatrix},$$

where $B_{ii}\in\mathbb{R}_{\nVdash}{}^{(i+1)\times(i+1)}$, $B_{i\,i-1}\in\mathbb{R}_{\nVdash}{}^{(i+1)\times i}$, $B_{i-1\,i}\in\mathbb{R}_{\nVdash}{}^{i\times(i+1)}$, and $A_0, A_1, A_2\in\mathbb{R}_{\nVdash}{}^{(n+1)\times(n+1)}$. The matrices of transition rates for the boundary states are given by

$$B_{00}=(-\lambda),\ B_{i-1\,i}=\begin{pmatrix} 0 & \lambda & & \\ & \ddots & \lambda & \\ & & \ddots & \ddots \\ & & 0 & \lambda \end{pmatrix},\ B_{i\,i-1}=\begin{pmatrix} 0 & & & \\ (1-p)\mu & 0 & & \\ & (1-p)\nu(2) & \ddots & \\ & & \ddots & 0 \\ & & & (1-p)\nu(i) \end{pmatrix},$$

and

$$B_{ii}=\begin{pmatrix} -(\lambda+i\delta) & i\delta & & & \\ p\mu & -(\lambda+\mu+(i-1)\delta) & (i-1)\delta & & \\ & \ddots & \ddots & \ddots & \\ & & p\nu(i-1) & -(\lambda+\nu(i-1)+\delta) & \delta \\ & & & p\nu(i) & -(\lambda+\nu(i)) \end{pmatrix}.$$

Moreover, the transition rates are given by

$$A_0=\begin{pmatrix} \lambda & & & \\ & \lambda & & \\ & & \ddots & \\ & & & \lambda \end{pmatrix},\ A_2=\begin{pmatrix} 0 & & & & & & \\ & (1-p)\mu & & & & & \\ & & 2(1-p)\mu & & & & \\ & & & \ddots & & & \\ & & & & s(1-p)\mu & & \\ & & & & & \ddots & \\ & & & & & & s(1-p)\mu \end{pmatrix},$$

and

$$A_1 = \begin{pmatrix} -(\lambda+n\delta) & n\delta & & & & & \\ p\mu & -(\lambda+\mu+(n-1)\delta) & (n-1)\delta & & & & \\ & \ddots & \ddots & \ddots & & & \\ & & sp\mu & -(\lambda+s\mu+(n-s)\delta) & (n-s)\delta & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & sp\mu & -(\lambda+s\mu+\delta) & \delta \\ & & & & & sp\mu & -(\lambda+s\mu) \end{pmatrix}.$$

### A.2. Stability condition

From the general theory of QBD processes (Neuts 1981) follows that the Markov process $\{N(t), Q_1(t)\}$ is ergodic (stable) if and only if

$$\pi A_0 e < \pi A_2 e, \tag{1}$$

where $e$ is the all one column vector and $\pi = (\pi_0, ..., \pi_n)$ is the equilibrium distribution of the Markov process with generator $A_0 + A_1 + A_2$. In other words, $\pi$ is such that

$$\pi(A_0 + A_1 + A_2) = 0, \; \pi e = 1, \tag{2}$$

and

$$A_0 + A_1 + A_2 = \begin{pmatrix} -n\delta & n\delta & & & & & \\ p\mu & -(p\mu+(n-1)\delta) & (n-1)\delta & & & & \\ & \ddots & \ddots & \ddots & & & \\ & & sp\mu & -(ps\mu+(n-s)\delta) & (n-s)\delta & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & ps\mu & -(ps\mu+\delta) & \delta \\ & & & & & ps\mu & -ps\mu \end{pmatrix}.$$

Then $\pi$ must satisfy the balance equations

$$\begin{aligned} -n\delta\pi_0 + p\mu\pi_1 &= 0, \\ (n-j+1)\delta\pi_{j-1} - (p\nu(j) + (n-j)\delta)\pi_j + p\nu(j+1)\pi_{j+1} &= 0, \\ \delta\pi_{n-1} - ps\mu\pi_n &= 0, \end{aligned}$$

with $\nu(j) = \min\{j, s\}\mu$, and the normalization condition

$$\sum_{i=0}^{n} \pi_i = 1.$$

It is readily verified that

$$\pi_i = \begin{cases} \pi_0 \binom{n}{i} \left(\frac{\delta}{p\mu}\right)^i & \text{for } 0 \le i \le s, \\ \pi_0 \binom{n}{i} \frac{i!}{s!} s^{s-i} \left(\frac{\delta}{p\mu}\right)^i & \text{for } s+1 \le i \le n \end{cases} \tag{3}$$

with

$$\pi_0 = \left( \sum_{i=0}^{s} \binom{n}{i} \left(\frac{\delta}{p\mu}\right)^i + \sum_{i=s+1}^{n} \binom{n}{i} \frac{i!}{s!} s^{s-i} \left(\frac{\delta}{p\mu}\right)^i \right)^{-1}.$$

satisfies the balance equations and the normalization condition.

Proposition 1. *The distribution of the closed two-node Jackson network illustrated in Figure 2a is given by*

$$\hat{\pi}_i = \begin{cases} \hat{\pi}_0 \binom{n}{i} \left(\frac{\delta}{p\mu}\right)^i & \text{for } 0 \le i \le s, \\ \hat{\pi}_0 \binom{n}{i} \frac{i!}{s!} s^{s-i} \left(\frac{\delta}{p\mu}\right)^i & \text{for } s+1 \le i \le n \end{cases} \tag{4}$$

*with*

$$\hat{\pi}_0 = \left[ \sum_{i=0}^{s} \binom{n}{i} \left(\frac{\delta}{p\mu}\right)^i + \sum_{i=s+1}^{n} \binom{n}{i} \frac{i!}{s!} s^{s-i} \left(\frac{\delta}{p\mu}\right)^i \right]^{-1}.$$

*W* e have a two-node closed Jackson network, with probability transition matrix

$$P = \begin{pmatrix} 1-p & p \\ 1 & 0 \end{pmatrix}.$$

Let $r_i(m)$ denote the rate of service when there are $m$ patient at queue $i$, so $r_1(m) = \min\{m, s\}$ and $r_2(m) = m$. The throughput vector $\gamma = (\gamma_1, \gamma_2) \in \mathbb{R}_{\not\vdash}{}^2$ must satisfy $\gamma = \gamma P$ and we find that $\gamma = (p, 1)$ suffices. From the general theory of Jackson networks, see Jackson (1963), it follows that the stationary distribution is given by

$$\pi_i = G^{-1} g_1(i) g_2(n-i)$$

with

$$g_1(i) = \frac{(\gamma_1/\mu)^i}{\prod_{m=1}^{i} r_1(m)},\; g_2(n-i) = \frac{(\gamma_2/\delta)^{n-i}}{\prod_{m=1}^{n-i} r_2(m)},$$

and normalization constant $G = \sum_{i=0}^{n} g_1(i) g_2(n-i)$. Then,

$$g_1(i) = \begin{cases} \frac{1}{i!\mu^i} & \text{for } 0 \le i \le s, \\ \frac{1}{s! s^{i-s} \mu^i} & \text{for } s+1 \le i \le n, \end{cases}$$

$$g_2(n-i) = \frac{1}{(n-i)!} \left(\frac{p}{\delta}\right)^n \left(\frac{\delta}{p}\right)^i,$$

and rewriting the expressions yields (4).

### A.3. Stationary distribution

Assuming that the stability condition is satisfied, we can determine the unique stationary distribution of the Markov process $\{N(t), Q_1(t)\}$. The vector $\pi_i$ can be written as $\pi_{n+i} = \pi_n G^i$ for $i = 0, 1, ...$, where $G$ is the minimal nonnegative solution of the non-linear matrix equation

$$A_0 + G A_1 + G^2 A_2 = 0. \tag{5}$$

The balance equations can be written as

$$\pi_{i-1} A_0 + \pi_i A_1 + \pi_{i+1} A_2 = 0,\; i = n+1, n+2, ...$$

and using $\pi_{n+i} = \pi_n G^{i-n}$ for $i = 0, 1, ...$, this find

$$\pi_n G^{i-n-1} (A_0 + G A_1 + G A_2) = 0,\; i = n+1, n+2, ....$$

Moreover, we have the boundary equations

$$\begin{aligned}
\pi_0 B_{00} + \pi_1 B_{10} &= 0 \\
\pi_0 B_{01} + \pi_1 B_{11} + \pi_2 B_{21} &= 0 \\
\pi_1 B_{12} + \pi_1 B_{22} + \pi_2 B_{32} &= 0 \\
&\vdots \\
\pi_{n-2} B_{n-2\,n-1} + \pi_{n-1} B_{n-1\,n-1} + \pi_n B_{n\,n-1} &= 0 \\
\pi_{n-1} B_{n-1\,n} + \pi_n B_{nn} + \pi_{n+1} A_2 &= 0,
\end{aligned}$$

along with the normalization equation

$$1 = \sum_{i=0}^{\infty} \pi_i e = \sum_{i=0}^{n-1} \pi_i e + \pi_n (I - G)^{-1} e,$$

where we slightly abuse notation by using $e$ as the all ones vector of appropriate size. We note that the matrix $G$ has a spectral radius less than one and therefore $(I - G)$ is invertible.

These equations provide the tools for finding the equilibrium probabilities. Although it is hard to solve $G$ analytically from Equation (5), it is easy to solve numerically by using the following algorithm (matrix-geometric method). Rewriting (5) gives

$$G = -(A_0 + G^2 A_2) A_1^{-1},$$

where $A_1$ is invertible, since it is a transient generator matrix. Let

$$G_{k+1} = -(A_0 + G_k^2 A_2) A_1^{-1},$$

starting with $G_k = 0$. We note that $G_k \uparrow G$ as $k$ grows to infinity (Neuts 1981). Once $||G_{k+1} - G_k||_2$ is below a certain preset threshold, we approximate $G$ by $G_{k+1}$.

## Appendix B: Proof of Proposition 2

First, note that by definition of the Erlang-R model with holding, in which no more that $n$ patients can be admitted in the ED simultaneously, that $Q_1^h(t) + Q_2^h(t) \le n = Q_1^J(t) + Q_2^J(t)$ follows directly. Therefore, we only consider the relation between the states in the blocking and holding variants Erlang-R model.

As noted Section 3.1, the model with holding can be characterized as a three-dimensional Markov chain $X^h(t) = (H(t), Q_1^h(t), Q_2^h(t))$ in which the components denote the number of holding, needy and content patients respectively. The Erlang-R model with blocking similarly admits a Markov process description, but with two dimensions, namely $X^b(t) = (Q_1^b(t), Q_1^b(t))$.

We prove the result by constructing a coupling between the Markov processes $X^h$ and $X^b$. Let $Z(t) := \left(\hat{X}^h(t), \hat{X}^b(t)\right) = \left(\hat{H}(t), \hat{Q}_1^h(t), \hat{Q}_2^h(t), \hat{Q}_1^b(t), \hat{Q}_2^b(t)\right)$.

We first define the transition rates of this five-dimensional Markov process, which naturally only depend on the current state of the system. After that we show that the transition rates relevant to $\hat{X}^h(t)$ and $X^h(t)$ coincide with those of either $X^h(t)$ or $\hat{X}^b(t)$, respectively. The latter implies that the marginal transitions of $\hat{X}^h(t)$ and $X^b(t)$ (and $\hat{X}^b(t)$ and $X^h(t)$) are equal, and hence so are their probability distribution of the Markov processes.

Let $Z(t) = (h, q_1^h, q_2^h, q_1^b, q_2^b)$. While defining the reachable states from this state and associated transition rates, we distinguish four transition types, and further differentiate the transition rates depending on the current state.

**Arrival.** Arrivals to occur in both models simultaneously, but are handled differently according to the current queue lengths.

1. If $q_1^h + q_2^h < n$ and $q_1^b + q_2^b < n$,

$$(h, q_1^h + 1, q_2^h, q_1^b + 1, q_2^b) \qquad \text{with rate } \lambda, \tag{6}$$

2. if $q_1^h + q_2^h = n$ and $q_1^b + q_2^b < n$,

$$(h+1, q_1^h, q_2^h, q_1^b + 1, q_2^b) \qquad \text{with rate } \lambda, \tag{7}$$

3. if $q_1^h + q_2^h < n$ and $q_1^b + q_2^b = n$,

$$(h, q_1^h + 1, q_2^h, q_1^b, q_2^b) \qquad \text{with rate } \lambda, \tag{8}$$

4. if $q_1^h + q_2^h = n$ and $q_1^b + q_2^b = n$,

$$(h+1, q_1^h + 1, q_2^h, q_1^b, q_2^b) \qquad \text{with rate } \lambda, \tag{9}$$

**Departure.** Basically, we align service completions in the two models, but allow a completion occurring solely in either of one of the two models, only if the queue length in this model is strictly larger than in the other one.

1. If $q_1^h \geq q_1^b$ and $h > 0$

$$\begin{cases} (h-1, q_1^h, q_2^h, q_1^b - 1, q_2^b) & \text{with rate } (q_1^b \wedge s)(1-p)\mu, \\ (h-1, q_1^h, q_2^h, q_1^b, q_2^b) & \text{with rate } [(q_1^h \wedge s) - (q_1^b \wedge s)](1-p)\mu. \end{cases} \tag{10}$$

2. If $q_1^h < q_1^b$ and $h > 0$

$$\begin{cases} (h-1, q_1^h, q_2^h, q_1^b - 1, q_2^b) & \text{with rate } (q_1^h \wedge s)(1-p)\mu, \\ (h, q_1^h, q_2^h, q_1^b - 1, q_2^b) & \text{with rate } [(q_1^b \wedge s) - (q_1^h \wedge s)](1-p)\mu. \end{cases} \tag{11}$$

3. If $q_1^h \geq q_1^b$ and $h = 0$

$$\begin{cases} (0, q_1^h - 1, q_2^h, q_1^b - 1, q_2^b) & \text{with rate } (q_1^b \wedge s)(1-p)\mu, \\ (0, q_1^h - 1, q_2^h, q_1^b, q_2^b) & \text{with rate } [(q_1^h \wedge s) - (q_1^b \wedge s)](1-p)\mu. \end{cases} \tag{12}$$

4. If $q_1^h < q_1^b$ and $h = 0$

$$\begin{cases} (0, q_1^h - 1, q_2^h, q_1^b - 1, q_2^b) & \text{with rate } (q_1^h \wedge s)(1-p)\mu, \\ (0, q_1^h, q_2^h, q_1^b - 1, q_2^b) & \text{with rate } [(q_1^b \wedge s) - (q_1^h \wedge s)](1-p)\mu. \end{cases} \tag{13}$$

**Become content** The differentiation between transitions is similar to those in the *departure* transition type.

1. If $q_1^h \geq q_1^b$,

$$\begin{cases} (h, q_1^h - 1, q_2^h + 1, q_1^b - 1, q_2^b + 1) & \text{with rate } (q_1^b \wedge s)p\mu, \\ (h, q_1^h - 1, q_2^h + 1, q_1^b, q_2^b) & \text{with rate } [(q_1^h \wedge s) - (q_1^b \wedge s)]p\mu. \end{cases} \tag{14}$$

2. If $q_1^h < q_1^b$,

$$\begin{cases} (h, q_1^h - 1, q_2^h + 1, q_1^b - 1, q_2^b + 1) & \text{with rate } (q_1^h \wedge s)p\mu, \\ (h, q_1^h, q_2^h, q_1^b - 1, q_2^b + 1) & \text{with rate } [(q_1^b \wedge s) - (q_1^h \wedge s)]p\mu. \end{cases} \tag{15}$$

**Become needy**

1. If $q_2^h \geq q_2^b$,

$$\begin{cases} (h, q_1^h + 1, q_2^h - 1, q_1^b + 1, q_2^b - 1) & \text{with rate } q_2^b\delta, \\ (h, q_1^h + 1, q_2^h - 1, q_1^b, q_2^b) & \text{with rate } (q_2^h - q_2^b)\delta, \end{cases} \tag{16}$$

2. If $q_2^h < q_2^b$,

$$\begin{cases} (h, q_1^h + 1, q_2^h - 1, q_1^b + 1, q_2^b - 1) & \text{with rate } q_2^h\delta, \\ (h, q_1^h, q_2^h, q_1^b + 1, q_2^b - 1) & \text{with rate } (q_2^b - q_2^h)\delta, \end{cases} \tag{17}$$

This set of transitions defines the dynamics of the Markov process $Z(t) = (\hat{X}^h(t), \hat{X}^b(t))$. Let us now restrict our attention to the transitions in which (at least one of the) first three coordinates of $Z(t)$ changes, that is, the marginal transitions of the process $\hat{X}^h$. Let $\hat{X}^h(t) = (h, q_1^h, q_2^h)$, then according to the transition scheme above, $\hat{X}^h$ moves to state

1. If $q_1^h + q_2^h < n$ (and hence necessarily $h = 0$),

$$\begin{cases} (0, q_1^h + 1, q_2^h) & \text{with rate } \lambda, \\ (0, q_1^h - 1, q_2^h) & \text{with rate } (q_1^h \wedge s)(1-p)\mu, \\ (0, q_1^h - 1, q_2^h + 1) & \text{with rate } (q_1^h \wedge s)p\mu, \\ (0, q_1^h + 1, q_2^h - 1) & \text{with rate } q_2^h\delta. \end{cases}$$

2. if $q_1^h + q_2^h = n$ and $h = 0$,

$$\begin{cases} (1, q_1^h, q_2^h) & \text{with rate } \lambda, \\ (0, q_1^h, q_2^h) & \text{with rate } (q_1^h \wedge s)(1-p)\mu, \\ (0, q_1^h - 1, q_2^h + 1) & \text{with rate } (q_1^h \wedge s)p\mu, \\ (0, q_1^h + 1, q_2^h - 1) & \text{with rate } q_2^h\delta. \end{cases}$$

3. if $h > 0$ (and hence necessarily $q_1^h + q_2^h = n$),

$$\begin{cases} (h + 1, q_1^h, q_2^h) & \text{with rate } \lambda, \\ (h - 1, q_1^h, q_2^h) & \text{with rate } (q_1^h \wedge s)(1-p)\mu, \\ (h, q_1^h - 1, q_2^h + 1) & \text{with rate } (q_1^h \wedge s)p\mu, \\ (h, q_1^h + 1, q_2^h - 1) & \text{with rate } q_2^h\delta. \end{cases}$$

One can check that these transitions indeed coincide with the transitions in the original holding model, hence $\hat{X}^h(t) \stackrel{d}{=} X^h(t)$.

Similarly, when the focusing on transitions of $Z(t)$ that are relevant for $\hat{X}^b(t)$, we deduce the following transition scheme. If $\hat{X}^b(t) = (q_1^b, q_2^b)$, then the next move according to the transitions of $Z(t)$ is

$$\begin{cases} (q_1^b + 1_{\{q_1^b + q_2^b < n\}}, q_2^b) & \text{with rate } \lambda, \\ (q_1^b - 1, q_2^b) & \text{with rate } (q_1^b \wedge s)(1-p)\mu, \\ (q_1^b - 1, q_2^b + 1) & \text{with rate } (q_1^b \wedge s)p\mu, \\ (q_1^b + 1, q_2^b - 1) & \text{with rate } q_2^b\delta. \end{cases}$$

These transition rates clearly coincide with the original Erlang-R model with blocking, and also hence $\hat{X}^b(t) \stackrel{d}{=} X^h(t)$.

Next, we show that under this coupling scheme we have that if $\hat{H}(0) = 0$, $\hat{Q}_1^h(0) = \hat{Q}_1^b(0)$ and $\hat{Q}_1^h(0) = \hat{Q}^b(0)$ then for all $t \geq 0$, $Z(t)$ satisfies the hypothesis:

(i) $\hat{Q}_1^b(t) + \hat{Q}_2^b(t) \leq \hat{Q}_1^h(t) + \hat{Q}_2^h(t)$,

(ii) $\hat{Q}_2^b(t) \le \hat{Q}_2^h(t)$,

(iii) $\hat{Q}_1^b(t) \le \hat{Q}_1^h(t) + H(t)$.

We do so by induction on the next state reached after a transition of the joint Markov process $Z = (\hat{X}^h, \hat{X}^b)$. First of all, $Z(0)$ clearly satisfies (i)-(iii). Next, assume $Z(t^-) = (h, q_1^h, q_2^h, q_1^b, q_2^b)$ satisfies the hypothesis and a transition occurs at $t$. We show that under the specified coupling scheme, the state reached after the next transition, $Z(t)$ must satisfy (i)-(iii) as well. To do so, we differentiate between the four types of transitions that could occur: arrival, departure, become content and become needy.

**Arrival.**

Recall that under our coupling scheme an arrival always occurs in both the holding and blocking model simultaneously, see (6)–(9). Furthermore, $q_2^h$ and $q_2^b$ are unchanged during this transition, rendering (ii) trivial.

By hypothesis $q_1^b + q_2^b \le q_1^h + q_2^b$, hence the event $q_1^h + q_2^h < n$ and $q_1^h + q_2^b = n$, with resulting state $(0, q_1^h + 1, q_2^h, q_1^b, q_2^b)$, can be excluded from our analysis We check the conditions for the remaining three cases.

1. If $Z(t) = (0, q_1^h + 1, q_2^h, q_1^b + 1, q_2^b)$, then $q_1^b + q_2^b < n$ and $q_1^h + q_2^h < n$.
   (i) $\hat{Q}_1^b(t) + \hat{Q}_2^b(t) = q_1^b + q_2^b + 1 \overset{(i)}{\le} q_1^h + q_2^h + 1 = \hat{Q}_1^h(t) + \hat{Q}_2^h(t)$.
   (iii) $\hat{Q}_1^b(t) = q_1^b + 1 \overset{(iii)}{\le} q_1^h + 1 = \hat{Q}_1^h(t) = \hat{Q}_1^h(t) + \hat{H}(t)$.
2. If $Z(t) = (h+1, q_1^h, q_2^h, q_1^b + 1, q_2^b)$, then $q_1^b + q_2^b < n$ and $q_1^h + q_2^h = n$.
   (i) $\hat{Q}_1^b(t) + \hat{Q}_2^b(t) = q_1^b + q_2^b + 1 \le n = q_1^h + q_2^h = \hat{Q}_1^h(t) + \hat{Q}_2^h(t)$.
   (iii) $\hat{Q}_1^b(t) = q_1^b + 1 \overset{(iii)}{\le} q_1^h + 1 = \hat{Q}_1^h(t) + \hat{H}(t)$.
3. If $Z(t) = (h+1, q_1^h, q_2^h, q_1^b, q_2^b)$, then $q_1^b + q_2^b = q_1^h + q_2^h = n$.
   (i) $\hat{Q}_1^b(t) + \hat{Q}_2^b(t) = q_1^b + q_2^b \overset{(i)}{\le} q_1^h + q_2^h = \hat{Q}_1^h(t) + \hat{Q}_2^h(t)$.
   (iii) $\hat{Q}_1^b(t) = q_1^b \overset{(iii)}{\le} q_1^h + h < q_1^h + h + 1 = \hat{H}(t)$.

**Departure.** By carefully examining the possible state transitions of $Z(t)$ following a departure, we list six reachable states. However, by (iii), we have that if $h = 0$, then $q_1^b \le q_1^h$, which excludes the state $(0, q_1^h, q_2^h, q_1^b, q_2^b)$ in (13) from the reachability graph. We check the remaining states for conditions (i)–(iii). Again, during a departure, $q_2^b$ and $q_2^h$ are unchanged, so (ii) is automatically satisfied by the induction hypothesis.

1. If $Z(t) = (h-1, q_1^h, q_2^h, q_1^b - 1, q_2^b)$, then $h > 0$.
   (i) $\hat{Q}_1^b(t) + \hat{Q}_2^b(t) = q_1^b + q_2^b - 1 \overset{(i)}{\le} q_1^h + q_2^h - 1 < q_1^h + q_2^h = \hat{Q}_1^h(t) + \hat{Q}_2^h(t)$.
   (iii) $\hat{Q}_1^b(t) = q_1^b - 1 \overset{(iii)}{\le} q_1^h + h - 1 = \hat{Q}_1^h(t) + \hat{H}(t)$.
2. If $Z(t) = (h-1, q_1^h, q_2^h, q_1^b, q_2^b)$, then $h > 0$ and $q_1^h \ge q_1^b$ (*).
   (i) $\hat{Q}_1^b(t) + \hat{Q}_2^b(t) = q_1^b + q_2^b \overset{(i)}{\le} q_1^h + q_2^h = \hat{Q}_1^h(t) + \hat{Q}_2^h(t)$.
   (iii) $\hat{Q}_1^b(t) = q_1^b \overset{(*)}{\le} q_1^h - 1 \le q_1^h + h - 1 = \hat{Q}_1^h(t) + \hat{H}(t)$.
3. If $Z(t) = (h, q_1^h, q_2^h, q_1^b - 1, q_2^b)$, then $h > 0$ and $q_1^h < q_1^b$ (*).
   (i) $\hat{Q}_1^b(t) + \hat{Q}_2^b(t) = q_1^b + q_2^b - 1 < q_1^b + q_2^b \overset{(i)}{\le} q_1^h + q_2^h = \hat{Q}_1^h(t) + \hat{Q}_2^h(t)$.
   (iii) $\hat{Q}_1^b(t) = q_1^b - 1 < q_1^b \overset{(*)}{\le} q_1^h + h = \hat{Q}_1^h(t) + \hat{H}(t)$.
4. If $Z(t) = (h, q_1^h - 1, q_2^h, q_1^b - 1, q_2^b)$, then $h = 0$.
   (i) $\hat{Q}_1^b(t) + \hat{Q}_2^b(t) = (q_1^b - 1) + q_2^b - 1 <\overset{(i)}{\le} (q_1^h - 1) + q_2^h = \hat{Q}_1^h(t) + \hat{Q}_2^h(t)$.

(iii) $\hat{Q}_1^b(t) = q_1^b - 1 \overset{(iii)}{\leq} q_1^h - 1 = \hat{Q}_1^h(t) + \hat{H}(t)$.

5. If $Z(t) = (0, q_1^h - 1, q_2^h, q_1^b, q_2^b)$, then $h = 0$ and $q_1^h > q_1^b$ (*).
   (i) $\hat{Q}_1^b(t) + \hat{Q}_2^b(t) = q_1^b + q_2^b \overset{(i)}{\leq} (q_1^h - 1) + q_2^b \overset{(ii)}{\leq} (q_1^h - 1) + q_2^h = \hat{Q}_1^h(t) + \hat{Q}_2^h(t)$.
   (iii) $\hat{Q}_1^b(t) = qy \overset{(*)}{\leq} q_1^h - 1 = \hat{Q}_1^h(t) + \hat{H}(t)$.

**Content start.** On the event of a patient becoming content, it is clear that the sums $\hat{Q}_1^h(t) + \hat{Q}_2^h(t)$ and $\hat{Q}_1^b(t) + \hat{Q}_2^b(t)$ and $H(t)$ are unaffected. This means that (i) is directly satisfied by the induction hypothesis. According to (14)–(15), three states can be reached.

1. If $Z(t) = (h, q_1^h - 1, q_2^h + 1, q_1^b - 1, q_2^b + 1)$,
   (ii) $\hat{Q}_2^b(t) = q_2^b + 1 \overset{(ii)}{\leq} q_2^h + 1 = \hat{Q}_2^h(t)$.
   (iii) $\hat{Q}_1^b(t) = q_1^b - 1 \overset{(iii)}{\leq} q_1^h + h - 1 = \hat{Q}_1^h(t) + \hat{H}(t)$.
2. If $Z(t) = (h, q_1^h - 1, q_2^h + 1, q_1^b, q_2^b)$, then $q_1^h > q_1^b$ (*),
   (ii) $\hat{Q}_2^b(t) = q_2^b \overset{(ii)}{\leq} q_2^h < q_2^h + 1 = \hat{Q}_2^h(t)$.
   (iii) $\hat{Q}_1^b(t) = q_1^b \overset{(iii)}{\leq} q_1^h + h < q_1^h + 1 + h = \hat{Q}_1^h(t) + \hat{H}(t)$.
3. If $Z(t) = (h, q_1^h, q_2^h, q_1^b - 1, q_2^b + 1)$, then $q_1^b > q_1^h$ and hence by (iii) $h > 0$. The latter is only possible if $q_1^h + q_2^h = n$ (*),
   (ii) $\hat{Q}_2^b(t) = q_2^b + 1 \leq n - q_1^b + 1 = (q_1^h + q_2^h) - q_1^b + 1 \overset{(*)}{\leq} q_2^h = \hat{Q}_2^h(t)$.
   (iii) $\hat{Q}_1^b(t) = q_1^b - 1 < q_1^h + h - 1 \overset{(*)}{\leq} q_1^h + h = \hat{Q}_1^h(t) + \hat{H}(t)$.

**Become needy.**

Just as in the event of content start, the sums $\hat{Q}_1^h(t) + \hat{Q}_2^h(t)$ and $\hat{Q}_1^b(t) + \hat{Q}_2^b(t)$ and $H(t)$ are unaffected, whereby (i) is directly satisfied by the induction hypothesis. By (ii), we have $q_2^h \geq q_2^b$. This excludes the state $(h, q_1^h, q_2^h, q_1^b + 1, q_2^b - 1)$ from being reached, see (17). We check the remaining to possibilities.

1. If $Z(t) = (h, q_1^h + 1, q_2^h - 1, q_1^b + 1, q_2^b - 1)$.
   (ii) $\hat{Q}_2^b(t) = q_2^b - 1 \overset{(ii)}{\leq} q_2^h - 1 = \hat{Q}_2^h(t)$.
   (iii) $\hat{Q}_1^b(t) = q_1^b + 1 \overset{(iii)}{\leq} q_1^h + h + 1 = \hat{Q}_1^h(t) + \hat{H}(t)$.
2. If $Z(t) = (h, q_1^h + 1, q_2^h - 1, q_1^b, q_2^b)$, then $q_2^h > q_2^b$ (*).
   (ii) $\hat{Q}_2^b(t) = q_2^b \overset{(*)}{\leq} q_2^h - 1 = \hat{Q}_2^h(t)$.
   (iii) $\hat{Q}_1^b(t) = q_1^b \overset{(iii)}{\leq} q_1^h + h < q_1^h + 1 + h = \hat{Q}_1^h(t) + \hat{H}(t)$.

Hence, the state reached after any feasible transition under the coupling scheme satisfies the conditions (i)–(iii). Thus we conclude that the joint process $(\hat{H}(t), \hat{Q}_1^h(t), \hat{Q}_2^h(t), \hat{Q}_1^b(t), \hat{Q}_2^b(t))$ adheres to (i)–(iii) for all $t$. Consequently, we have that (i) implies

$$\begin{aligned}
\mathbb{P}\left(Q_1^b(t) + Q_2^b(t) \geq k\right) &= \mathbb{P}\left(Q_1^b(t) + Q_2^b(t) \geq k\right) \\
&= \sum_{j=0}^{n} \mathbb{P}\left(\hat{Q}_1^b(t) + \hat{Q}_2^b(t) \geq k, \hat{Q}_1^h(t) + \hat{Q}_2^h(t) = j\right) \\
&= \sum_{j=k}^{n} \mathbb{P}\left(\hat{Q}_1^b(t) + \hat{Q}_2^b(t) \geq k, \hat{Q}_1^h(t) + \hat{Q}_2^h(t) = j\right) \\
&\leq \sum_{j=h}^{n} \mathbb{P}\left(\hat{Q}_1^h(t) + \hat{Q}_2^h(t) = j\right) \\
&= \mathbb{P}\left(Q_1^h(t) + Q_2^h(t) \geq k\right) = \mathbb{P}\left(Q_1^h(t) + Q_2^h(t) \geq k\right).
\end{aligned}$$

The other two orderings follow similarly.

REMARK 1. Note that under this coupling scheme we cannot get the ordering $\hat{Q}_1^h(t)(t) \geq \hat{Q}_1^b(t)(t)$ for all $t \geq 0$. A minimal counter example occurs for $s = n = 1$. Let $Z(0) = ((0,0,0),(0,0))$. First, two arrivals occur, such that state $((1,1,0),(1,0))$ is reached, followed by a departure transition, yielding $((0,1,0),(0,0))$. Next, the one patient left in the model with holding system becomes content, so that we obtain $((0,0,1),(0,0))$. At this stage, if an arrival occurs, the arriving patient will be put in the holding queue in the model with holding, and admitted to nurse queue in the model with blocking. Hence we end up in state $((1,0,1),(1,0))$, in which $\hat{Q}_1^h(t) < \hat{Q}_1^b(t)$.

## Appendix C: Proof of Theorem 1 - Performance measures of Erlang-R with blocking

For convenience we state the theorem here again.

THEOREM 1. *Let $s$ and $n$ scale as in (11) with $\beta,\gamma > 0$ as $\lambda \to \infty$. Then, if $\beta \neq 0$*

$$g^b(\beta,\gamma) := \lim_{\lambda\to\infty} \mathbb{P}^b(\text{delay}) = \left(1 + \frac{\beta \int_{-\infty}^{\beta} \Phi\left(\frac{\gamma - t\sqrt{r}}{\sqrt{1-r}}\right) d\Phi(t)}{\phi(\beta)\Phi(\eta) - \phi(\sqrt{\beta^2+\eta^2})\mathrm{e}^{\frac{1}{2}\omega^2}\Phi(\omega)}\right)^{-1}, \tag{18}$$

$$f^b(\beta,\gamma) := \lim_{\lambda\to\infty} \sqrt{R_1}\cdot\mathbb{P}^b(\text{block}) = \frac{\sqrt{r}\phi(\gamma)\Phi(-\omega\sqrt{r}) + \phi(\sqrt{\beta^2+\eta^2})\,\mathrm{e}^{\frac{1}{2}\omega^2}\Phi(\omega)}{\int_{-\infty}^{\beta} \Phi\left(\frac{\gamma - t\sqrt{r}}{\sqrt{1-r}}\right) d\Phi(t) + \frac{\phi(\beta)\Phi(\eta)}{\beta} - \frac{\phi(\sqrt{\beta^2+\eta^2})}{\beta}\mathrm{e}^{\frac{1}{2}\omega^2}\Phi(\omega)}, \tag{19}$$

$$h^b(\beta,\gamma) := \lim_{\lambda\to\infty} \sqrt{R_1}\cdot\mathbb{E}[W] = \frac{\frac{\phi(\beta)\Phi(\eta)}{\beta^2} + \left(\frac{\beta}{r} - \frac{\gamma}{\sqrt{r}} - \frac{1}{\beta}\right)\frac{\phi(\sqrt{\eta^2+\beta^2})}{\beta}\,\mathrm{e}^{\frac{1}{2}\omega^2}\,\Phi(\omega) - \sqrt{\frac{1-r}{r}}\frac{\phi(\beta)\phi(\eta)}{\beta}}{\int_{-\infty}^{\beta} \Phi\left(\frac{\gamma - t\sqrt{r}}{\sqrt{1-r}}\right) d\Phi(t) + \frac{\phi(\beta)\Phi(\eta)}{\beta} - \frac{\phi(\sqrt{\beta^2+\eta^2})}{\beta}\mathrm{e}^{\frac{1}{2}\omega^2}\Phi(\omega)}, \tag{20}$$

*and if $\beta = 0$,*

$$g_0^b(\gamma) := \lim_{\lambda\to\infty} \mathbb{P}^b(\text{delay}) \left(1 + \frac{\int_{-\infty}^{0} \Phi\left(\frac{\gamma - t\sqrt{r}}{\sqrt{1-r}}\right) d\Phi(t)}{\sqrt{\frac{1-r}{r}}\frac{1}{\sqrt{2\pi}}\left(\eta\,\Phi(\eta) + \phi(\eta)\right)}\right)^{-1} \tag{21}$$

$$f_0^b(\gamma) := \lim_{\lambda\to\infty} \sqrt{R_1}\cdot\mathbb{P}^b(\text{block}) = \frac{\sqrt{r}\,\phi(\gamma)\Phi(-\omega\sqrt{r}) + \frac{1}{\sqrt{2\pi}}\Phi(\eta)}{\int_{-\infty}^{\beta} \Phi\left(\frac{\gamma - t\sqrt{r}}{\sqrt{1-r}}\right) d\Phi(t) + \sqrt{\frac{1-r}{r}}\frac{1}{\sqrt{2\pi}}\left(\eta\,\Phi(\eta) + \phi(\eta)\right)}, \tag{22}$$

$$h_0^b(\gamma) := \lim_{\lambda\to\infty} \sqrt{R_1}\cdot\mathbb{E}[W] = \frac{1}{2\mu}\,\frac{(\gamma^2/r + 1)\,\Phi(\eta) + \eta\phi(\eta)}{\frac{r}{1-r}\sqrt{2\pi}\int_{-\infty}^{0} \Phi\left(\frac{\gamma - t\sqrt{r}}{\sqrt{1-r}}\right) d\Phi(t) + \sqrt{\frac{r}{1-r}}\left(\eta\Phi(\eta) + \phi(\eta)\right)}, \tag{23}$$

*where $\eta = \frac{\gamma - \beta\sqrt{r}}{\sqrt{1-r}}$ and $\omega := \frac{\gamma - \beta/\sqrt{r}}{\sqrt{1-r}}$.*

In this appendix we prove the heavy-traffic approximations of the system-measures introduced in Theorem 1. As a first stage we present and prove four lemmas.

LEMMA 1. *Let the variables $\lambda$, $s$ and $n$ tend to $\infty$ simultaneously and satisfy the QED scaling conditions in (11) with $\beta \neq 0$. Define $B_1$ as the expression*

$$B_1 = \frac{e^{-R}}{s!}R_1^s\frac{1}{1-\rho}\sum_{l=0}^{n-s-1}\frac{1}{l!}R_2^l e^{-R_2}.$$

*Then*

$$\lim_{\lambda\to\infty} B_1 = \frac{\phi(\beta)\Phi(\eta)}{\beta}.$$

Lemma 2. *Let the variables $\lambda$, $s$ and $n$ tend to $\infty$ simultaneously and satisfy the QED scaling conditions in (11) with $\beta \neq 0$. Define $B_2$ as the expression*

$$B_2 = \frac{e^{-(R_1+R_2)}}{s!} R_1^s \frac{\rho^{n-s}}{1-\rho} \sum_{l=0}^{n-s-1} \frac{1}{l!} \left( \frac{R_2}{\rho} \right)^l .$$

*Then*

$$\lim_{\lambda\to\infty} B_2 = \frac{\phi(\sqrt{\eta^2+\beta^2})}{\beta} e^{\frac{1}{2}\omega^2} \Phi(\omega).$$

Lemma 3. *Let the variables $\lambda$, $s$ and $n$ tend to $\infty$ simultaneously and satisfy the QED scaling conditions in (11). Define $A$ as the expression*

$$A = \sum_{\substack{i,j|i\leq s,\\ i+j\leq n-1}} \frac{1}{i!j!} R_1^i R_2^j e^{-(R_1+R_2)}. \tag{24}$$

*Then*

$$\lim_{\lambda\to\infty} A = \int_{-\infty}^{\beta} \Phi \left( \eta + (\beta - t) \sqrt{\frac{\delta}{\mu p}} \right) d\Phi(t).$$

Lemma 4. *Let the variables $\lambda$, $s$ and $n$ tend to $\infty$ simultaneously and satisfy the QED scaling conditions in (11) with $\beta = 0$. Define $B$ as the expression*

$$B = e^{-(R_1+R_2)} \frac{1}{s!} {R_1}^s \sum_{j=0}^{n-s-1} \frac{1}{j!} {R_2}^j \sum_{i=0}^{n-s-j-1} \rho^i .$$

*Then*

$$\lim_{\lambda\to\infty} B = \sqrt{\frac{\mu p}{\delta}} \frac{1}{\sqrt{2\pi}} \left( \eta \Phi(\eta) + \phi(\eta) \right).$$

### C.1. Proof of Lemma 1

B y using Stirling's formula $\left(s! \approx \sqrt{2\pi s} \left(\frac{s}{e}\right)^s\right)$, and QED assumption that $\sqrt{s}(1 - R_1/s) \to \beta$ as $\lambda \to \infty$, one obtains for $B_1$:

$$\begin{aligned} B_1 &\approx \frac{e^{s-R}}{\sqrt{2\pi s}} \rho^s \frac{\sqrt{s}}{\beta} \sum_{l=0}^{n-s-1} \frac{1}{l!} (R_2)^l e^{-R_2} = \frac{e^{s-R}\rho^s}{\sqrt{2\pi}\beta} \sum_{l=0}^{n-s-1} \frac{1}{l!} R_2^l e^{-R_2} \\ &= \frac{e^{s(1-\rho)}}{\sqrt{2\pi}\beta} \rho^s P(X_\lambda \leq n-s-1) \end{aligned}$$

where $\rho = \frac{\lambda}{(1-p)s\mu}$, and $X_\lambda$ is a random variable with the Poisson distribution with parameter $R_2$. When $\lambda \to \infty$, $R_2 \to \infty$ too, since $p$, and $\delta$ are fixed. Note that

$$\mathbb{P}(X_\lambda \leq n-s-1) = \mathbb{P}\left( \frac{X_\lambda - R_2}{\sqrt{R_2}} \leq \frac{n-s-1-R_2}{\sqrt{R_2}} \right)$$

Now we need to find the limit for the following fraction

$$\frac{n-s-R_2}{\sqrt{R_2}}$$

as $\lambda \to \infty$ using assumption that $\sqrt{s}(1 - R_1/s) \to \beta$ as $\lambda \to \infty$.

$$\lim_{\lambda\to\infty} \frac{n-s-R_2}{\sqrt{R_2}} = \lim_{\lambda\to\infty} \frac{n - \frac{R_1}{r} - s + R_1}{\sqrt{\frac{R_1}{r} - R_1}} = \frac{\gamma\sqrt{\frac{R_1}{r}} - \beta\sqrt{R_1}}{\sqrt{\frac{R_1}{r} - R_1}} = \frac{\gamma - \beta\sqrt{r}}{\sqrt{1-r}}. \tag{25}$$

Hence, define $\eta = \frac{\gamma - \beta\sqrt{r}}{\sqrt{1-r}}$. Thus, when $\lambda \to \infty$, by the Central Limit Theorem (Normal approximation to Poisson) we have

$$\left( \frac{X_\lambda - R_2}{\sqrt{R_2}} \right) \Rightarrow N(0,1)$$

and due to assumption QED (i) of the lemma we get

$$\mathbb{P}(X_\lambda \leq n-s-1) \to \mathbb{P}(N(0,1) \leq \eta) = \Phi(\eta), \quad as\ \lambda \to \infty \tag{26}$$

where $N(0,1)$ is a standard normal random variable with distribution function $\Phi(\cdot)$. It follows thus that

$$B_1 \approx \frac{e^{s(1-\rho)}}{\sqrt{2\pi}\beta} \rho^s \Phi(\eta) = \frac{e^{s(1-\rho+\ln\rho)}}{\sqrt{2\pi}\beta} \Phi(\eta).$$

Making use of the expansion

$$\ln\rho = \ln(1-(1-\rho)) = -(1-\rho) - \frac{(1-\rho)^2}{2} + o(1-\rho)^2, \quad (\rho \to 1)$$

one obtains

$$B_1 \approx \frac{e^{s(1-\rho-(1-\rho)-\frac{(1-\rho)^2}{2})}}{\sqrt{2\pi}\beta} \Phi(\eta) = \frac{e^{-\frac{s(1-\rho)^2}{2}}}{\sqrt{2\pi}\beta} \Phi(\eta)$$

by QED assumption that $\sqrt{s}(1-R_1/s) \to \beta$, it is clear that $s(1-\rho)^2 \to \beta^2$, when $\lambda \to \infty$. This implies

$$\lim_{\lambda\to\infty} B_1 = \frac{\phi(\beta)\Phi(\eta)}{\beta}$$

where $\phi(\cdot)$ is the standard normal density function, and $\Phi(\cdot)$ is the standard normal distribution function. This proves Lemma 1.

### C.2. Proof of Lemma 2

Again according to Stirling's formula, and the QED assumption that $\frac{n-s-R_1-R_2}{\sqrt{R_1+R_2}} \to \eta$ as $\lambda \to \infty$, one obtains for $B_2$:

$$\begin{aligned} B_2 &\approx \frac{e^{s-R_1-R_2}}{\sqrt{2\pi s}} \frac{\rho^n}{1-\rho} \sum_{l=0}^{n-s-1} \frac{1}{l!} \left(\frac{R_2}{\rho}\right)^l = \frac{e^{s(1-\rho)-R_2}}{\sqrt{2\pi s}} \frac{\sqrt{s}\rho^n}{\beta} e^{\frac{R_2}{\rho}} \sum_{l=0}^{n-s-1} \frac{1}{l!} \left(\frac{R_2}{\rho}\right)^l e^{-\frac{R_2}{\rho}} \\ &= \frac{e^{s(1-\rho)+R_2\left(\frac{1-\rho}{\rho}\right)}}{\sqrt{2\pi}\beta} \rho^n \mathbb{P}(Y_\lambda \leq n-s-1), \end{aligned}$$

where $\rho = \frac{\lambda}{(1-p)s\mu}$, and $Y_\lambda$ is a random variable with the Poisson distribution with parameter $\frac{R_2}{\rho}$. Note that

$$\mathbb{P}(Y_\lambda \leq n-s-1) = \mathbb{P}\left( \frac{Y_\lambda - \frac{R_2}{\rho}}{\sqrt{\frac{R_2}{\rho}}} \leq \frac{n-s-1-\frac{R_2}{\rho}}{\sqrt{\frac{R_2}{\rho}}} \right).$$

Now we need to find the limit for the following fraction $\frac{n-s-\frac{R_2}{\rho}}{\sqrt{\frac{R_2}{\rho}}}$ as $\lambda \to \infty$ using assumption QED (i).

$$\begin{aligned} \lim_{\lambda\to\infty} \frac{n-s-\frac{R_2}{\rho}}{\sqrt{\frac{R_2}{\rho}}} &= \lim_{\lambda\to\infty} \frac{\eta\sqrt{R_2}+R_2-\frac{R_2}{\rho}}{\sqrt{\frac{R_2}{\rho}}} = \lim_{\lambda\to\infty} \eta\sqrt{\rho} + \frac{\sqrt{R_2}(\rho-1)}{\sqrt{\rho}} \\ &= \eta - \lim_{\lambda\to\infty} \sqrt{\frac{sp\mu}{\delta}}(1-\rho) = \eta - \sqrt{\frac{p\mu}{\delta}}\beta. \end{aligned} \tag{27}$$

Denote $\omega=\eta-\beta\sqrt{\frac{p\mu}{\delta}}$. Thus, when $\lambda\to\infty$, by the Central Limit Theorem (Normal approximation to Poisson) we have

$$\left(\frac{Y_\lambda-\frac{R_2}{\rho}}{\sqrt{\frac{R_2}{\rho}}}\right)\Rightarrow N(0,1)$$

and

$$\mathbb{P}(Y_\lambda\le n-s-1)\to\mathbb{P}(N(0,1)\le\omega)=\Phi(\omega),\ \text{as}\ \lambda\to\infty,$$

where $N(0,1)$ is a standard normal random variable with distribution function $\Phi$. It follows thus that

$$B_2\approx\frac{e^{s(1-\rho)+R_2\left(\frac{1-\rho}{\rho}\right)}}{\sqrt{2\pi}\beta}\rho^n\Phi(\omega)=\frac{e^{s(1-\rho)+R_2\left(\frac{1-\rho}{\rho}\right)+n\ln\rho}}{\sqrt{2\pi}\beta}\Phi(\omega).$$

Making use of the expansion

$$\ln\rho=\ln(1-(1-\rho))=-(1-\rho)-\frac{(1-\rho)^2}{2}+o(1-\rho)^2,\quad(\rho\to1)$$

and using our assumptions that as $\lambda\to\infty$: $\rho\to1$, $s\approx R_1+\beta\sqrt{R_1}$, $n\approx\frac{R_1}{r}+\gamma\sqrt{\frac{R_1}{r}}$, and $n-s\approx R_2+\eta\sqrt{R_2}$, one obtains

$$\begin{aligned}
&s(1-\rho)+R_2\left(\frac{1-\rho}{\rho}\right)+n\ln\rho=s(1-\rho)+R_2\left(\frac{1-\rho}{\rho}\right)-n\left(1-\rho+\frac{(1-\rho)^2}{2}\right)\\
&=-\left(n-s-\frac{R_2}{\rho}\right)(1-\rho)-\frac{n(1-\rho)^2}{2}\approx-\left(\eta\sqrt{R_2}+R_2-\frac{R_2}{\rho}\right)(1-\rho)-\frac{n(1-\rho)^2}{2}\\
&=\left(\frac{R_2}{\rho}-\frac{n}{2}\right)(1-\rho)^2-\eta\sqrt{R_2}(1-\rho)\approx\left(\frac{R_2}{\rho}-\frac{1}{2}\left(\frac{R_1}{r}+\gamma\sqrt{\frac{R_1}{r}}\right)\right)(1-\rho)^2-\eta\sqrt{R_2}(1-\rho)\\
&=\left(\frac{R_2}{\rho}-\frac{R_1}{2r}\right)(1-\rho)^2-\frac{1}{2}\gamma\sqrt{\frac{R_1}{r}}(1-\rho)^2-\eta\sqrt{R_2}(1-\rho)\\
&=\left(\frac{R_2}{\rho}-\frac{R_1}{2r}\right)\frac{\beta^2}{R_!}-\frac{1}{2}\gamma\sqrt{\frac{R_1}{r}}\frac{\beta^2}{R_1}-\eta\sqrt{R_2}\beta\sqrt{\frac{1}{R_1}}\\
&\approx\frac{p\mu}{\delta\rho}\beta^2-\frac{1}{2}\left(\frac{p\mu}{\delta}+1\right)\beta^2-\eta\beta\sqrt{\frac{p\mu}{\delta}}=-\frac{1}{2}(\eta^2+\beta^2)+\frac{1}{2}\omega^2.
\end{aligned}$$

Therefore,

$$\lim_{\lambda\to\infty}B_2\approx\frac{e^{s(1-\rho)+R_2\left(\frac{1-\rho}{\rho}\right)+n\ln\rho}}{\sqrt{2\pi}\beta}\Phi(\omega)\approx\frac{e^{-\frac{1}{2}(\eta^2+\beta^2)+\frac{1}{2}\omega^2}}{\sqrt{2\pi}\beta}\Phi(\omega)=\frac{\phi(\sqrt{\eta^2+\beta^2})}{\beta}e^{\frac{1}{2}\omega^2}\Phi(\omega).$$

This proves Lemma 2.

### C.3. Proof of Lemma 3

We will find the asymptotic behavior of $A$ by finding its lower and upper bounds. Let us consider a partition $\{s_h\}_{h=0}^{l}$ of the interval $[0,s]$.

$$s_h=s-h\tau,\quad h=0,1,...,\ell;\quad s_{\ell+1}=0$$

where $\tau=\left[\epsilon\sqrt{R_1}\right]$, $\epsilon$ is an arbitrary non-negative real and $\ell$ is a positive integer.

If $\lambda$ and $s$ tend to infinity and satisfy the QED assumption that $\frac{n-s-R_1-R_2}{\sqrt{R_1+R_2}}\to\eta$ as $\lambda\to\infty$, then $\ell<\frac{s}{\tau}$ for $\lambda$ big enough and all the $s_h$ belong to $[0,s]$; $h=0,1,...,\ell$. Emphasize that the length $\tau$ of every interval

$[s_{h-1}, s_h]$ depends on $\lambda$. The variable $A$ is given by the formula (24). Let us consider a lower estimate for $A$ given by the following sum:

$$\begin{aligned} A \geq A_1 &= \sum_{h=0}^{\ell} \sum_{i=s_{h+1}}^{s_h} \frac{1}{i!} R_1^i e^{-R_1} \cdot \sum_{j=0}^{n-s_h-1} \frac{1}{j!} R_2^j e^{-R_2} \\ &= \sum_{h=0}^{\ell} \sum_{i=s_{h+1}}^{s_h} \frac{1}{i!} R_1^i e^{-R} \mathbb{P}(Y_n \leq n - s_h - 1) \\ &= \sum_{h=0}^{\ell} \mathbb{P}(s_{h+1} \leq X_n \leq s_h) \mathbb{P}(Y_n \leq n - s_h - 1) \end{aligned} \tag{28}$$

where $X_n$ and $Y_n$ are independent Poisson random variables with parameters $R_1$ and $R_2$, respectively.

If $\lambda \to \infty$ then $R_1 \to \infty$, since $p$ and $\mu$ are fixed. Note that

$$\mathbb{P}(s_{h+1} \leq X_n \leq s_h) = \mathbb{P}\left( \frac{s_{h+1} - R_1}{\sqrt{R_1}} \leq \frac{X_n - R_1}{\sqrt{R_1}} \leq \frac{s_h - R_1}{\sqrt{R_1}} \right).$$

Thus, when $\lambda \to \infty$, by the Central Limit Theorem (Normal approximation to Poisson) we have

$$\frac{X_n - R}{\sqrt{R_1}} \Rightarrow N(0,1).$$

Since

$$\lim_{\lambda \to \infty} \frac{s_h - R_1}{\sqrt{R_1}} = \lim_{\lambda \to \infty} \frac{s - h\epsilon\sqrt{R_1} - R_1}{\sqrt{R_1}} = \lim_{\lambda \to \infty} \frac{R_1 - \beta\sqrt{R_1} - h\epsilon\sqrt{R_1} - R_1}{\sqrt{R_1}} = \beta - h\epsilon$$

we obtain:

$$\begin{aligned} &\mathbb{P}(s_{h+1} \leq X_n \leq s_h) = \Phi(\beta - h\epsilon) - \Phi(\beta - (h+1)\epsilon), \; h = 0, .., \ell - 1 \\ &\mathbb{P}(0 \leq X_n \leq s_\ell) = \Phi(\beta - \ell\epsilon). \end{aligned} \tag{29}$$

Similarly, if $\lambda \to \infty$ then $R_2 \to \infty$, since $p$,$\delta$ and $\gamma$ are fixed. Note that

$$\mathbb{P}(Y_n \leq n - s_h) = \mathbb{P}\left( \frac{Y_n - R_2}{\sqrt{R_2}} \leq \frac{n - s_h - R_2}{\sqrt{R_2}} \right).$$

Thus, when $\lambda \to \infty$, by the Central Limit Theorem (Normal approximation to Poisson) we have

$$\frac{Y_n - R_2}{\sqrt{R_2}} \Rightarrow N(0,1).$$

Since

$$\lim_{\lambda \to \infty} \frac{n - s_h - R_2}{\sqrt{R_2}} = \lim_{\lambda \to \infty} \frac{n - s - h\epsilon\sqrt{R_1} - R_2}{\sqrt{R_2}} = \lim_{\lambda \to \infty} \frac{R_2 + \eta\sqrt{R_2} - h\epsilon\sqrt{R_1} - R_2}{\sqrt{R_2}} = \eta - h\epsilon\frac{\sqrt{R_1}}{\sqrt{R_2}} = \eta - h\epsilon\sqrt{\frac{\delta}{p\mu}},$$

we obtain:

$$\mathbb{P}(Y_n \leq n - s_h) = \Phi\left( \eta - h\epsilon\sqrt{\frac{\delta}{p\mu}} \right), \; h = 0, .., \ell \tag{30}$$

It follows from (28), (29), and (30) that

$$\lim_{\lambda \to \infty} A \geq \sum_{h=0}^{\ell-1} (\Phi(\beta - h\epsilon) - \Phi(\beta - (h+1)\epsilon)) \Phi\left( \eta - h\epsilon\sqrt{\frac{\delta}{p\mu}} \right) + \Phi(\beta - \ell\epsilon) \Phi\left( \eta - \ell\epsilon\sqrt{\frac{\delta}{p\mu}} \right) \tag{31}$$

which is the lower Riemann-Stieltjes sum of the integral

$$-\int_0^\infty \Phi\left( \eta + x\sqrt{\frac{\delta}{p\mu}} \right) d\Phi(\beta - x) = \int_{-\infty}^\beta \Phi\left( \eta + (\beta - t)\sqrt{\frac{\delta}{p\mu}} \right) d\Phi(t) \tag{32}$$

corresponding to the partition $\{\beta - h\epsilon\}_{h=0}^{\ell}$ of the semi axis $(-\infty, \beta)$. Similarly, let us take the upper estimate for $A$ as the following sum:

$$
\begin{aligned}
A \le A_2 &= \sum_{h=0}^{\ell} \sum_{i=s_{h+1}}^{s_h} \frac{1}{i!} R_1^i e^{-R_1} \sum_{j=0}^{n-s_{h+1}-1} \frac{1}{j!} R_2^j e^{-R_2} \\
&= \sum_{h=0}^{\ell} \sum_{i=s_{h+1}}^{s_h} \frac{1}{i!} R_1^i e^{-R} \mathbb{P}(Y_n \le n - s_{h+1} - 1) \qquad (33)\\
&= \sum_{h=0}^{\ell} \mathbb{P}(s_{h+1} \le X_n \le s_h) \mathbb{P}(Y_n \le n - s_{h+1} - 1)
\end{aligned}
$$

where $X_n$ and $Y_n$ are the same random variable as before. Using the same calculation that were computed for the upper boundary we obtain

$$
\lim_{\lambda\to\infty} A \le \sum_{h=0}^{\ell-1} \left(\Phi(\beta - h\epsilon) - \Phi(\beta - (h+1)\epsilon)\right) \Phi\left(\eta - (h+1)\epsilon\sqrt{\frac{\delta}{p\mu}}\right) + \Phi(\beta - \ell\epsilon) \qquad (34)
$$

which is the upper Riemann-Stieltjes sum for the integral (32). When $\epsilon \to 0$ the boundaries (31) and (34) lead to the following equality

$$
\lim_{\lambda\to\infty} A = \int_{-\infty}^{\beta} \Phi\left(\eta + (\beta - t)\sqrt{\frac{\delta}{p\mu}}\right) d\Phi(t) = \int_{-\infty}^{\beta} \Phi\left(\frac{\gamma - t\sqrt{r}}{\sqrt{1-r}}\right) d\Phi(t)
$$

This proves Lemma 3.

### C.4. Proof of Lemma 4

$F$ irst, we will rewrite Equation (24):

$$
B = e^{-(R_1+R_2)} \frac{1}{s!} {R_1}^s \sum_{j=0}^{n-s-1} \frac{1}{j!} {R_2}^j \sum_{i=0}^{n-s-j-1} \rho^i = e^{-(R_1+R_2)} \frac{1}{s!} {R_1}^s \sum_{j=0}^{n-s-1} \frac{1}{j!} {R_2}^j \frac{1-\rho^{n-s-j}}{1-\rho}.
$$

When $\beta = 0$, as $\lambda \to \infty$, by the QED assumption that $\sqrt{s}(1-\rho) \to \beta$, $\rho = 1$. Therefore,

$$
\sum_{i=0}^{n-s-j-1} \rho^i = n - s - j.
$$

When $\beta \to 0$, $\rho \to 1$ but still $\rho \ne 1$, the expression $\frac{1-\rho^{n-s-j}}{1-\rho}$ can be approximated by $\frac{1-\rho^i}{1-\rho} \approx i$. Thus,

$$
\lim_{\rho\to 1} \sum_{i=0}^{n-s-j-1} \rho^i = n - s - j,
$$

which is the same phrase as when $\rho = 1$. Hence,

$$\begin{aligned}
B &\approx e^{-(R_1+R_2)} \frac{1}{s!} {R_1}^s \sum_{j=0}^{n-s-1} \frac{1}{j!} {R_2}^j (n-s-j) \\
&= e^{-(R_1+R_2)} \frac{1}{s!} {R_1}^s \left( (n-s) \sum_{j=0}^{n-s-1} \frac{1}{j!} {R_2}^j - \sum_{j=0}^{n-s-1} \frac{1}{j!} {R_2}^j - \sum_{j=0}^{n-s-1} \frac{j}{j!} {R_2}^j \right) \\
&= e^{-(R_1+R_2)} \frac{1}{s!} {R_1}^s \left( (n-s) \sum_{l=0}^{n-s-1} \frac{1}{l!} R_2^l - \sum_{j=0}^{n-s-2} \frac{1}{j!} {R_2}^j - R_2 \sum_{j=0}^{n-s-2} \frac{1}{j!} {R_2}^j \right) \\
&= e^{-(R_1+R_2)} \frac{1}{s!} {R_1}^s \left( (n-s) \sum_{l=0}^{n-s-1} \frac{1}{l!} R_2^l - \sum_{l=0}^{n-s-2} \frac{1}{l!} R_2^l - R_2 \sum_{l=0}^{n-s-2} \frac{1}{l!} R_2^l \right) \\
&= e^{-(R_1+R_2)} \frac{1}{s!} {R_1}^s \left( (n-s-R_2) \sum_{l=0}^{n-s-1} \frac{1}{l!} R_2^l + \frac{R_2^{n-s}}{(n-s-1)!} \right) \\
&\approx e^{s-R} \frac{1}{\sqrt{2\pi s}} \rho^s \left( (n-s-R_2) \sum_{l=0}^{n-s-1} \frac{1}{l!} R_2^l e^{-R_2} + \frac{R_2^{n-s} e^{-R_2}}{(n-s-1)!} \right) \\
&= \frac{1}{\sqrt{2\pi s}} \left( (n-s-R_2) \sum_{l=0}^{n-s-1} \frac{1}{l!} R_2^l e^{-R_2} + \frac{R_2^{n-s} e^{-R_2}}{(n-s-1)!} \right).
\end{aligned}$$

As seen in Equation (26)

$$(n-s-R_2) \sum_{l=0}^{n-s-1} \frac{1}{l!} R_2^l e^{-R_2} \approx \eta \sqrt{R_2} \Phi(\eta).$$

By using Stirling's formula:

$$\begin{aligned}
\frac{R_2^{n-s} e^{-R_2}}{(n-s-1)!} &= \frac{(n-s) R_2^{n-s} e^{-R_2}}{(n-s)!} \approx \frac{(n-s) e^{n-s-R_2}}{\sqrt{2\pi(n-s)}} \left( \frac{R_2}{n-s} \right)^{n-s} \\
&= \frac{(n-s) e^{n-s-R_2+(n-s)\ln\left(\frac{R_2}{n-s}\right)}}{\sqrt{2\pi(n-s)}} = \sqrt{\frac{n-s}{2\pi}} e^{(n-s)\left(1-\frac{R_2}{n-s}+\ln\left(\frac{R_2}{n-s}\right)\right)}.
\end{aligned}$$

By assuming that $\frac{n-s-R_1-R_2}{\sqrt{R_1+R_2}} \to \eta$ when $\lambda \to \infty$

$$\begin{aligned}
&(n-s)\left(1-\frac{R_2}{n-s}+\ln\left(\frac{R_2}{n-s}\right)\right) = (n-s)\left(1-\frac{R_2}{n-s}-\left(1-\frac{R_2}{n-s}\right)-\frac{1}{2}\left(1-\frac{R_2}{n-s}\right)^2\right) \\
&= -\frac{n-s}{2}\left(1-\frac{R_2}{n-s}\right)^2 = -\frac{1}{2}\frac{(n-s-R_2)^2}{n-s} \approx -\frac{1}{2}\frac{\left(\eta\sqrt{R_2}\right)^2}{\eta\sqrt{R_2}+R_2} \approx -\frac{1}{2}\frac{\left(\eta\sqrt{R_2}\right)^2}{\eta\sqrt{R_2}+R_2} \approx -\frac{\eta^2}{2}.
\end{aligned}$$

Therefore, by the QED assumption that $\frac{n-s-R_1-R_2}{\sqrt{R_1+R_2}} \to \eta$ .

$$\sqrt{\frac{n-s}{2\pi}} e^{(n-s)\left(1-\frac{R_2}{n-s}+\ln\left(\frac{R_2}{n-s}\right)\right)} \approx \sqrt{\frac{\eta\sqrt{R_2}+R_2}{2\pi}} e^{-\frac{\eta^2}{2}} = \sqrt{\eta\sqrt{R_2}+R_2}\phi(\eta) \approx \sqrt{R_2}\phi(\eta).$$

Combining the above approximations and the assumption that $\beta = 0$ and therefore $s = R_1$ yields

$$\begin{aligned}
B &\approx \frac{1}{\sqrt{2\pi s}} \left( (n-s-R_2) \sum_{l=0}^{n-s-1} \frac{1}{l!} R_2^l e^{-R_2} + \frac{R_2^{n-s} e^{-R_2}}{(n-s-1)!} \right) \\
&\approx \frac{1}{\sqrt{2\pi s}} \left( \eta\sqrt{R_2}\Phi(\eta) + \sqrt{R_2}\phi(\eta) \right) = \frac{\sqrt{R_2}}{\sqrt{2\pi s}} \left( \eta\Phi(\eta) + \phi(\eta) \right) \\
&= \frac{\sqrt{R_2}}{\sqrt{2\pi R}} \left( \eta\Phi(\eta) + \phi(\eta) \right) = \sqrt{\frac{p\mu}{\delta}} \frac{1}{\sqrt{2\pi}} \left( \eta\Phi(\eta) + \phi(\eta) \right) = \sqrt{\frac{1-r}{r}} \frac{1}{\sqrt{2\pi}} \left( \eta\Phi(\eta) + \phi(\eta) \right).
\end{aligned}$$

This proves Lemma 4.

### C.5. Approximation of the Probability of Delay

The first approximation will be for the measure: the probability of waiting or the probability of delay.

THEOREM 2. *Let the variables $\lambda$, $s$ and $n$ tend to $\infty$ simultaneously and satisfy the QED scaling conditions in (11). Then if $\beta \neq 0$*

$$\lim_{\lambda\to\infty} \mathbb{P}(W>0) = \left(1 + \frac{\beta\int_{-\infty}^{\beta}\Phi\left(\eta+(\beta-t)\sqrt{\xi}\right)d\Phi(t)}{\phi(\beta)\Phi(\eta)-\phi(\sqrt{\eta^2+\beta^2})e^{\frac{1}{2}\omega^2}\Phi(\omega)}\right)^{-1},$$

*and if $\beta = 0$,*

$$\lim_{\lambda\to\infty} \mathbb{P}(W>0) = \left(1 + \frac{\int_{-\infty}^{0}\Phi\left(\frac{\gamma-t\sqrt{r}}{\sqrt{1-r}}\right)d\Phi(t)}{\sqrt{\frac{1-r}{r}}\frac{1}{\sqrt{2\pi}}\left(\eta\Phi(\eta)+\phi(\eta)\right)}\right)^{-1},$$

*where $\xi = \frac{R_1}{R_2} = \frac{\delta}{p\mu}$, $\omega = \eta - \beta\sqrt{\xi^{-1}}$.*

*Proof:* By the Arrival theorem for closed networks,

$$\mathbb{P}_n(W>0) = \mathbb{P}_{n-1}(Q_1(\infty)\geq s) = \sum_{m=s}^{n-1}\sum_{i=s}^{m}\pi_{n-1}^{b}(i,m-i) = \pi_0^{n-1}\sum_{m=s}^{n-1}\sum_{i=s}^{m}\frac{1}{s!s^{i-s}}R_1^i\frac{1}{(m-i)!}R_2^{m-i},$$

where

$$\pi_0^{n-1} = \left(\sum_{l=0}^{n-1}\frac{1}{l!}\left(R_1+R_2\right)^l + \sum_{m=s}^{n-1}\sum_{i=s}^{m}\left(\frac{1}{s!s^{i-s}}-\frac{1}{i!}\right)\frac{1}{(m-i)!}R_1^iR_2^{m-i}\right)^{-1}.$$

Thus,

$$\mathbb{P}_n(W>0) = \left(1+\frac{A}{B}\right)^{-1},$$

where

$$\begin{aligned} A &= \sum_{l=0}^{n-1}\frac{1}{l!}\left(R_1+R_2\right)^l e^{-(R_1+R_2)} - \sum_{m=s}^{n-1}\sum_{i=s}^{m}\frac{1}{i!(m-i)!}R_1^iR_2^{m-i}e^{-(R_1+R_2)} \\ &= \sum_{\substack{i,j|i\leq s,\\ i+j\leq n-1}}\frac{1}{i!j!}R_1^iR_2^je^{-(R_1+R_2)}, \end{aligned}$$

$$\begin{aligned} B &= \sum_{m=s}^{n-1}\sum_{i=s}^{m}\frac{1}{s!s^{i-s}}R_1^i\frac{1}{(m-i)!}R_2^{m-i}e^{-(R_1+R_2)} = \sum_{j=0}^{n-s-1}\sum_{i=s}^{n-j-1}\frac{1}{s!s^{i-s}}R_1^i\frac{1}{j!}R_2^je^{-(R_1+R_2)} \\ &= \sum_{j=0}^{n-s-1}\sum_{i=0}^{n-s-j-1}\frac{1}{s!s^i}\frac{1}{j!}R_1^{i+s}R_2^je^{-(R_1+R_2)} = \frac{1}{s!}R_1^se^{-(R_1+R_2)}\sum_{j=0}^{n-s-1}\frac{1}{j!}R_2^j\sum_{i=0}^{n-s-j-1}\rho^i, \end{aligned} \tag{35}$$

and $\rho = \frac{\lambda}{(1-p)s\mu} = \frac{R_1}{s}$.

Then under the QED assumption that $\sqrt{s}(1-\rho)\to\beta$, $-\infty<\beta<\infty$, if $\beta\neq 0$ as $\lambda\to\infty$, we can rewrite the right-hand side in the following way:

$$\begin{aligned} B &= \frac{1}{s!}R_1^se^{-(R_1+R_2)}\sum_{j=0}^{n-s-1}\frac{1}{j!}R_2^j\frac{1-\rho^{n-s-j}}{1-\rho} \\ &= \frac{1}{s!}R_1^se^{-(R_1+R_2)}\frac{1}{1-\rho}\sum_{j=0}^{n-s-1}\frac{1}{j!}R_2^j - \frac{1}{s!}R_1^se^{-(R_1+R_2)}\frac{\rho^{n-s}}{1-\rho}\sum_{j=0}^{n-s-1}\frac{1}{j!}\left(\frac{R_2}{\rho}\right)^j = B_1 - B_2. \end{aligned} \tag{36}$$

Applying Lemmas1-3, if $\beta \neq 0$ we get

$$\lim_{\lambda\to\infty} \mathbb{P}(W>0) = \left(1 + \frac{\beta \int_{-\infty}^{\beta} \Phi\left(\eta + (\beta - t)\sqrt{\frac{\delta}{p\mu}}\right) d\Phi(t)}{\phi(\beta)\Phi(\eta) - \phi(\sqrt{\eta^2+\beta^2}) e^{\frac{1}{2}\omega^2} \Phi(\omega)}\right)^{-1}.$$

Applying Lemma 3 and 4 when $\beta = 0$ we get

$$\lim_{\lambda\to\infty} \mathbb{P}(W>0) = \left(1 + \frac{\int_{-\infty}^{0} \Phi\left(\frac{\gamma - t\sqrt{r}}{\sqrt{1-r}}\right) d\Phi(t)}{\sqrt{\frac{1-r}{r}} \frac{1}{\sqrt{2\pi}} \left(\eta\Phi(\eta) + \phi(\eta)\right)}\right)^{-1}.$$

This proves Theorem 2.

### C.6. Approximation of the Expected Waiting Time

In this appendix we will prove the approximation for the expected waiting time, stated in Section 4.2. The exact measure was defined in Section 3.2. For convenience we state the theorem here again. The first theorem gives the approximation for the case where $\beta \neq 0$.

THEOREM 3. *Let the variables $\lambda$, $s$ and $n$ tend to $\infty$ simultaneously and satisfy the QED scaling conditions in (11) with $\beta \neq 0$. Then*

$$\lim_{\lambda\to\infty} \sqrt{s}\mathbb{E}[W] = \frac{1}{\mu} \frac{\left(\frac{1}{\xi} - \frac{\eta}{\beta\sqrt{\xi}} - \frac{1}{\beta^2}\right) \phi(\sqrt{\beta^2+\eta^2})\, e^{\frac{1}{2}\omega^2}\, \Phi(\omega) + \frac{\phi(\beta)\Phi(\eta)}{\beta^2} - \frac{\phi(\beta)\phi(\eta)}{\sqrt{\xi}\beta}}{\int_{-\infty}^{\beta} \Phi(\eta + (\beta - t)\sqrt{\xi})\, d\Phi(t) + \frac{\phi(\beta)\Phi(\eta)}{\beta} - \frac{\phi(\sqrt{\beta^2+\eta^2})}{\beta}\, e^{\frac{1}{2}\omega^2}\, \Phi(\omega)},$$

*where $\xi = \frac{R_1}{R_2} = \frac{\delta}{p\mu}$, $\omega = \eta - \beta\sqrt{\xi^{-1}}$.*

*Proof:* It follows from (6) of the manuscript that the expectation of the waiting time is given by

$$\begin{aligned}
\mathbb{E}[W] &= \int_0^\infty p_n(s;t)dt = \frac{1}{\mu s} \sum_{m=s}^{n-1} \sum_{i=s}^{m} \pi_{n-1}(i, m-i)(i-s+1) \\
&= \frac{1}{\mu s} \sum_{m=s}^{n-1} \sum_{i=s}^{m} \pi_{n-1}(i, m-i)(i-s) + \frac{1}{\mu s} \sum_{m=s}^{n-1} \sum_{i=s}^{m} \pi_{n-1}(i, m-i) \qquad (37) \\
&= \frac{1}{\mu s} \sum_{m=s}^{n-1} \sum_{i=s}^{m} \pi_{n-1}(i, m-i)(i-s) + \frac{1}{\mu s}\mathbb{P}(W>0) = C + D,
\end{aligned}$$

where $D$ is given by

$$D = \frac{1}{\mu s}\mathbb{P}(W>0) = \frac{1}{\mu s} \frac{B}{A+B}$$

and $A$ and $B$ were defined in (24) and (35), respectively, and $C$ is given by,

$$\begin{aligned}
C &= \frac{1}{\mu s} \sum_{m=s}^{n-1} \sum_{i=s}^{m} \pi_{n-1}(i, m-i)\,(i-s) \\
&= \frac{1}{\mu s}\, \pi_0 \sum_{m=s}^{n-1} \sum_{i=s}^{m} \frac{R_1^i}{s!s^{i-s}} \frac{R_2^{m-i}}{(m-i)!} (i-s) = \frac{1}{\mu s} \frac{Ge^{-(R_1+R_2)}}{A+B}.
\end{aligned}$$

We will rewrite $G$ in the following way:

$$\begin{aligned}
G &= \sum_{m=s}^{n-1} \sum_{i=s}^{m} \frac{R_1^i}{s!s^{i-s}} \frac{R_2^{m-i}}{(m-i)!} (i-s) = \sum_{j=0}^{n-s-1} \sum_{i=s}^{n-j-1} \frac{R_1^i}{s!s^{i-s}} \frac{R_2^j}{j!} (i-s) \\
&= \sum_{j=0}^{n-s-1} \sum_{i=0}^{n-s-j-1} \frac{R_1^{i+s}}{s!s^i} \frac{R_2^j}{j!}\, i = \frac{R_1^s}{s!} \sum_{j=0}^{n-s-1} \sum_{i=0}^{n-s-j-1} i\, \rho^i \frac{R_2^j}{j!}
\end{aligned}$$

Using the formula

$$\sum_{l=0}^{M} l\,\rho^l = \rho\Big(\sum_{l=0}^{M}\rho^l\Big)' = \rho\left(\frac{1-\rho^{M+1}}{1-\rho}\right)' = -M\frac{\rho^{M+1}}{1-\rho} + \rho\,\frac{1-\rho^M}{(1-\rho)^2}, \tag{38}$$

we can rewrite $G$ as a sum $G = G_1 + G_2$, where

$$G_1 = -\frac{R_1^s}{s!}\sum_{j=0}^{n-s-1}(n-s-j-1)\,\frac{\rho^{n-s-j}}{1-\rho}\,\frac{R_2^j}{j!},$$

and

$$G_2 = \frac{R_1^s}{s!}\sum_{j=0}^{n-s-1}\rho\,\frac{1-\rho^{n-s-j-1}}{(1-\rho)^2}\,\frac{R_2^j}{j!}.$$

Therefore,

$$\begin{aligned}
G_1 &= -\frac{R_1^s}{s!}\sum_{j=0}^{n-s-1}(n-s-j-1)\,\frac{\rho^{n-s-j}}{1-\rho}\,\frac{R_2^j}{j!}\\
&= -(n-s-1)\,\frac{R_1^s}{s!}\sum_{j=0}^{n-s-1}\frac{\rho^{n-s-j}}{1-\rho}\,\frac{R_2^j}{j!} + \frac{R_1^s}{s!}\sum_{j=0}^{n-s-1} j\,\frac{\rho^{n-s-j}}{1-\rho}\,\frac{R_2^j}{j!}\\
&= -(n-s-1)\,\frac{\rho^{n-s}}{1-\rho}\,\frac{R_1^s}{s!}\sum_{j=0}^{n-s-1}\frac{(R_2/\rho)^j}{j!} + \frac{R_1^s}{s!}\,\frac{\rho^{n-s}}{1-\rho}\sum_{j=0}^{n-s-1} j\,\frac{(R_2/\rho)^j}{j!}\\
&= -(n-s-1)\,\frac{\rho^{n-s}}{1-\rho}\,\frac{R_1^s}{s!}\sum_{j=0}^{n-s-1}\frac{(R_2/\rho)^j}{j!} + \frac{R_1^s}{s!}\,\frac{R_2}{\rho}\,\frac{\rho^{n-s}}{1-\rho}\sum_{j=0}^{n-s-2}\frac{(R_2/\rho)^j}{j!}\\
&= -\left(n-s-R_2/\rho-1\right)\,\frac{\rho^{n-s}}{1-\rho}\,\frac{R_1^s}{s!}\sum_{j=0}^{n-s-1}\frac{(R_2/\rho)^j}{j!} - \frac{R_1^s}{s!}\,\frac{1}{1-\rho}\,\frac{R_2^{n-s}}{(n-s-1)!}\\
&= -\left(n-s-R_2/\rho-1\right)\,\mathrm{e}^{R_1+R_2}\,B_2 - \frac{1}{1-\rho}\,\frac{R_1^s}{s!}\,\frac{R_2^{n-s}}{(n-s-1)!},
\end{aligned}$$

where $B_2$ was defined in Lemma 2. For $G_2$ we have,

$$\begin{aligned}
G_2 &= \frac{R_1^s}{s!}\,\frac{\rho}{(1-\rho)^2}\sum_{j=0}^{n-s-1}\left(1-\rho^{n-s-j-1}\right)\frac{R_2^j}{j!} = \frac{R_1^s}{s!}\,\frac{\rho}{(1-\rho)^2}\sum_{j=0}^{n-s-1}\frac{R_2^j}{j!} - \frac{R_1^s}{s!}\,\frac{\rho^{n-s}}{(1-\rho)^2}\sum_{j=0}^{n-s-1}\frac{(R_2/\rho)^j}{j!}\\
&= \mathrm{e}^{R_1+R_2}\,\frac{1}{1-\rho}\,(\rho B_1 - B_2),
\end{aligned}$$

with $B_1$ as in Lemma 1. Note that

$$\begin{aligned}
n-s-\frac{R_2}{\rho}-1 &= R_2 - \eta\sqrt{R_2} - \frac{R_2}{\rho} - 1 = R_2\,(1-1/\rho) + \eta\sqrt{R_2} - 1\\
&= R_2\left(\frac{R_1-s}{\sqrt{R_1}}\right) + \eta\sqrt{R_2} - 1 = R_2\cdot -\frac{\beta\sqrt{R_1}}{R_1} + \eta\sqrt{R_2} - 1 \approx \sqrt{R_1}\left(\eta\sqrt{\xi} - \beta\xi\right) \approx \sqrt{s}\left(\eta/\sqrt{\xi} - \beta/\xi\right),
\end{aligned}$$

for $s$ large, where $\xi = R_1/R_2 = \delta/p\mu$. Furthermore,

$$\begin{aligned}
\frac{1}{1-\rho}\,\frac{R_1^s}{s!}\,\frac{R_2^{n-s}}{(n-s-1)!} &= \frac{n-s}{1-\rho}\,\frac{R_1^s}{s!}\,\frac{R_2^{n-s}}{(n-s)!}\\
&= \frac{n-s}{1-\rho}\,\mathrm{e}^{R_1+R_2}\,\mathbb{P}\left(\mathrm{Pois}(R_1)=s\right)\,\mathbb{P}\left(\mathrm{Pois}(R_2)=n-s\right)\\
&= \frac{R_2+\eta\sqrt{R_2}}{\beta/\sqrt{R_1}}\mathrm{e}^{R_1+R_2}\mathbb{P}\left(\mathrm{Pois}(R_1)=s\right)\,\mathbb{P}\left(\mathrm{Pois}(R_2)=n-s\right)\\
&= \sqrt{R_2}\,\mathrm{e}^{R_1+R_2}\,\frac{1}{\beta}\,\sqrt{R_1}\mathbb{P}\left(\mathrm{Pois}(R_1)=s\right)\cdot\sqrt{R_2}\mathbb{P}\left(\mathrm{Pois}(R_2)=n-s\right))\\
&\approx \sqrt{R_1}\,\mathrm{e}^{R_1+R_2}\,\frac{1}{\sqrt{\xi}\beta}\,\phi(\beta)\phi(\eta),
\end{aligned}$$

where we used that $\sqrt{R}\,\mathbb{P}(\text{Pois}(R)=R+x\sqrt{R})\to\phi(x)$ as $R\to\infty$. Hence, we have for $G_1$

$$\frac{1}{\sqrt{s}}\mathrm{e}^{-(R_1+R_2)}\,G_1\to-\left(\eta/\sqrt{\xi}-\beta/\xi\right)\cdot\frac{\phi(\sqrt{\beta^2+\eta^2})}{\beta}\,\mathrm{e}^{\frac{1}{2}\omega^2}\,\Phi(\omega)-\frac{1}{\sqrt{\xi}\beta}\,\phi(\beta)\phi(\eta)$$

and for $G_2$

$$\begin{aligned}\frac{1}{\sqrt{s}}\mathrm{e}^{-(R_1+R_2)}\,G_2&=\frac{1}{\sqrt{s}}\frac{1}{1-\rho}\,(\rho B_1-B_2)\approx\frac{1}{\beta}\,(\rho B_1-B_2)\\&\to\frac{\phi(\beta)\Phi(\eta)}{\beta^2}-\frac{\phi(\sqrt{\beta^2+\eta^2})}{\beta^2}\,\mathrm{e}^{\frac{1}{2}\omega^2}\,\Phi(\omega),\end{aligned}$$

as $s\to\infty$. In total, this gives

$$\frac{1}{\sqrt{s}}\mathrm{e}^{-(R_1+R_2)}\,G\to\left(\frac{1}{\xi}-\frac{\eta}{\beta\sqrt{\xi}}-\frac{1}{\beta^2}\right)\phi(\sqrt{\beta^2+\eta^2})\,\mathrm{e}^{\frac{1}{2}\omega^2}\,\Phi(\omega)+\frac{\phi(\beta)\Phi(\eta)}{\beta^2}-\frac{\phi(\beta)\phi(\eta)}{\sqrt{\xi}\beta}. \tag{39}$$

We can conclude,

$$\begin{aligned}\sqrt{R_1}\,\mathbb{E}[W]&=\frac{\sqrt{R_1}}{\mu s}\frac{Ge^{-(R_1+R_2)}}{A+B}+\frac{\sqrt{s}}{\mu s}\,\mathbb{P}(W>0)=\frac{\sqrt{s}}{\mu s}\,\frac{Ge^{-(R_1+R_2)}}{A+B}+\frac{1}{\mu\sqrt{s}}\,\mathbb{P}(W>0)\\&\to\frac{1}{\mu}\,\frac{\left(\frac{1}{\xi}-\frac{\eta}{\beta\sqrt{\xi}}-\frac{1}{\beta^2}\right)\phi(\sqrt{\beta^2+\eta^2})\,\mathrm{e}^{\frac{1}{2}\omega^2}\,\Phi(\omega)+\frac{\phi(\beta)\Phi(\eta)}{\beta^2}-\frac{\phi(\beta)\phi(\eta)}{\sqrt{\xi}\beta}}{\int_{-\infty}^{\beta}\Phi(\eta+(\beta-t)\sqrt{\xi})\,d\Phi(t)+\frac{\phi(\beta)\Phi(\eta)}{\beta}-\frac{\phi(\sqrt{\beta^2+\eta^2})}{\beta}\,\mathrm{e}^{\frac{1}{2}\omega^2}\,\Phi(\omega)},\end{aligned}$$

as $s\to\infty$.

The second theorem gives the approximation for the case where $\beta=0$.

Theorem 4. *Let the variables $\lambda$, $s$ and $n$ tend to $\infty$ simultaneously and satisfy the QED scaling conditions in (11) with $\beta=0$. Then*

$$\lim_{\lambda\to\infty}\sqrt{s}\mathbb{E}[W]=\frac{1}{2\mu}\frac{\xi^{-1}\left((\eta^2+1)\Phi(\eta)+\eta\phi(\eta)\right)}{\sqrt{2\pi}\int_{-\infty}^{0}\Phi\left(\eta-t\sqrt{\xi}\right)d\Phi(t)+\sqrt{\xi^{-1}}\left(\eta\Phi(\eta)+\phi(\eta)\right)},$$

*where $\xi=\frac{R_1}{R_2}=\frac{\delta}{p\mu}$, $\omega=\eta-\beta\sqrt{\xi^{-1}}$.*

*Proof:* As before

$$\mathbb{E}[W]=\frac{1}{\mu s}\frac{Ge^{-(R_1+R_2)}+B}{A+B}. \tag{40}$$

Since $\beta=0$, we have $\rho=1$ so that

$$\begin{aligned}G&=\frac{R_1^s}{s!}\sum_{j=0}^{n-s-1}\sum_{i=0}^{n-s-j-1}i\,\frac{R_2^j}{j!}=\frac{1}{2}\frac{R_1^s}{s!}\sum_{j=0}^{n-s-1}(n-s-j)(n-s-j-1)\,\frac{R_2^j}{j!}\\&=\frac{1}{2}\,\frac{R_1^s}{s!}\,(n-s-1)\sum_{j=0}^{n-s-1}\frac{R_2^j}{j!}\,(n-s-j)-\frac{1}{2}\,\frac{R_1^s}{s!}\sum_{j=0}^{n-s-1}j\,\frac{R_2^j}{j!}\,(n-s-j)\\&=\frac{1}{2}\,\frac{R_1^s}{s!}\,(n-s-1)\sum_{j=0}^{n-s-1}\frac{R_2^j}{j!}\,(n-s-j)-\frac{1}{2}\,\frac{R_1^s}{s!}\,R_2\sum_{j=0}^{n-s-2}\frac{R_2^j}{j!}\,(n-s-j-1)\\&=\frac{1}{2}\,\frac{R_1^s}{s!}\,(n-s-1)\sum_{j=0}^{n-s-1}\frac{R_2^j}{j!}\,(n-s-j)-\frac{1}{2}\,\frac{R_1^s}{s!}\,R_2\sum_{j=0}^{n-s-1}\frac{R_2^j}{j!}\,(n-s-j-1)\\&=\frac{1}{2}\,\frac{R_1^s}{s!}\,(n-s-1)\sum_{j=0}^{n-s-1}\frac{R_2^j}{j!}\,(n-s-j)-\frac{1}{2}\,\frac{R_1^s}{s!}\,R_2\sum_{j=0}^{n-s-1}\frac{R_2^j}{j!}\,(n-s-j)+\frac{1}{2}\frac{R_1^s}{s!}\,R_2\sum_{j=0}^{n-s-1}\frac{R_2^j}{j!}\\&=\frac{1}{2}\,\frac{R_1^s}{s!}\,(n-s-R_2-1)\sum_{j=0}^{n-s-1}\frac{R_2^j}{j!}\,(n-s-j)+\frac{1}{2}\frac{R_1^s}{s!}\,R_2\sum_{j=0}^{n-s-1}\frac{R_2^j}{j!}.\end{aligned}$$

Here,

$$
\begin{aligned}
\mathrm{e}^{-R_2}\sum_{j=0}^{n-s-1}\frac{R_2^j}{j!}(n-s-j) &= \mathrm{e}^{-R_2}(n-s)\sum_{j=0}^{n-s-1}\frac{R_2^j}{j!} - \mathrm{e}^{-R_2}R_2\sum_{j=0}^{n-s-2}\frac{R_2^j}{j!} \\
&= \mathrm{e}^{-R_2}(n-s-R_2)\sum_{j=0}^{n-s-1}\frac{R_2^j}{j!} + \mathrm{e}^{-R_2}(n-s)\frac{R_2^{n-s}}{(n-s)!} \\
&= \eta\sqrt{R_2}\,\mathbb{P}(\mathrm{Pois}(R_2)\le n-s-1) + (n-s-1)\mathbb{P}(\mathrm{Pois}(R_2)=n-s) \\
&= \eta\sqrt{R_2}\,\mathbb{P}(\mathrm{Pois}(R_2)\le n-s) + (R_2-1)\mathbb{P}(\mathrm{Pois}(R_2)=n-s) \\
&\approx \sqrt{R_2}\left(\eta\,\Phi(\eta)+\phi(\eta)\right) = \sqrt{\xi^{-1}}\sqrt{R_1}\left(\eta\,\Phi(\eta)+\phi(\eta)\right).
\end{aligned} \tag{41}
$$

Furthermore

$$
n-s-1-R_2 = \eta\sqrt{R_2}-1 \approx \eta\sqrt{\xi^{-1}}\sqrt{R_1} \tag{42}
$$

and

$$
\mathrm{e}^{-R_1-R_2}\frac{R_1^s}{s!}R_2\sum_{j=0}^{n-s-1}\frac{R_2^j}{j!} = \xi^{-1}\sqrt{R_1}\sqrt{R_1}\mathbb{P}(\mathrm{Pois}(R_1)=s)\,\mathbb{P}(\mathrm{Pois}(R_2\le n-s-1) \approx \sqrt{R_1}\xi^{-1}\phi(0)\Phi(\eta). \tag{43}
$$

Combining (41)-(43), we find

$$
\begin{aligned}
\mathrm{e}^{-R_1-R_2}G &\approx \frac{1}{2}\xi^{-1}\eta\sqrt{R_1}\,\mathbb{P}(\mathrm{Pois}(R_1)=s)\ (\eta\Phi(\eta)+\phi(\eta)) + \frac{1}{2}\sqrt{R_1}\phi(0)\Phi(\eta) \\
&\approx \frac{1}{2}\xi^{-1}\eta\sqrt{R_1}\left(\eta\,\Phi(\eta)+\phi(\eta)\right) + \frac{1}{2}\xi^{-1}\sqrt{R_1}\,\phi(0)\Phi(\eta) = \frac{1}{2}\sqrt{R_1}\,\xi^{-1}\phi(0)\left[(\eta^2+1)\Phi(\eta)+\eta\,\phi(\eta)\right],
\end{aligned}
$$

for $R_1$ large. Hence, we can conclude

$$
\begin{aligned}
\sqrt{R_1}\,\mathbb{E}[W] &= \frac{\sqrt{R_1}}{\mu s}\frac{Ge^{-R_1-R_2}}{A+B} + \frac{\sqrt{R_1}}{\mu s}\mathbb{P}(W>0) \approx \frac{1}{\sqrt{R_1}\mu}\frac{\mathrm{e}^{-R_1-R_2}G}{A+B} + \frac{1}{\sqrt{R_1}\mu}\mathbb{P}(W>0) \\
&\to \frac{1}{2\mu}\frac{\xi^{-1}\left[(\eta^2+1)\Phi(\eta)+\eta\,\phi(\eta)\right]}{\sqrt{2\pi}\int_{-\infty}^{0}\Phi(\eta-t\sqrt{\xi})d\Phi(t)+\sqrt{\xi^{-1}}(\eta\Phi(\eta)+\phi(\eta))},
\end{aligned}
$$

where we used that $\phi(0)=1/\sqrt{2\pi}$, and $\eta=\frac{\gamma-\beta\sqrt{r}}{\sqrt{1-r}}$.

### C.7. Approximation of the Blocking Probability

We prove here the approximations given in Theorem 1 of the probability of blocking. For convenience we repeat here the relevant theorem.

THEOREM 5. *Let the variables $\lambda$, $s$ and $n$ tend to $\infty$ simultaneously and satisfy the QED scaling conditions in (11) with $\beta\neq 0$. Then*

$$
\lim_{\lambda\to\infty}\sqrt{s}\mathbb{P}(block) = \frac{\sqrt{r}\phi(\gamma)\Phi(-\omega\sqrt{r})+\phi(\sqrt{\eta^2+\beta^2})e^{\frac{\omega^2}{2}}\Phi(\omega)}{\int_{-\infty}^{\beta}\Phi\left(\eta+(\beta-t)\sqrt{\xi}\right)d\Phi(t)+\frac{\phi(\beta)\Phi(\eta)}{\beta}-\frac{\phi(\sqrt{\eta^2+\beta^2})}{\beta}e^{\frac{1}{2}\omega^2}\Phi(\omega)} \tag{44}
$$

*where $\eta=\frac{\gamma-\beta\sqrt{r}}{\sqrt{1-r}}$, and $\omega=\eta-\frac{\beta\sqrt{1-r}}{\sqrt{r}}=\frac{\gamma-\beta/\sqrt{r}}{\sqrt{1-r}}$.*

*Proof:* It follows from (5) that the probability of blocking is given by

$$
\mathbb{P}_n = \pi_0\left(\sum_{i=0}^{s}\frac{1}{i!}R_1^i\frac{1}{(n-i)!}R_2^{n-i}+\sum_{i=s+1}^{n}\frac{1}{s!s^{i-s}}R_1^i\frac{1}{(n-i)!}R_2^{n-i}\right) = \frac{\delta_1+\delta_2}{\tilde{\xi}+\tilde{\zeta}_1-\tilde{\zeta}_2},
$$

where

$$\delta_1 = \sum_{i=0}^{s} \frac{1}{i!} R_1^i \frac{1}{(n-i)!} R_2^{n-i} e^{-(R_1+R_2)}$$
$$\delta_2 = \sum_{i=s+1}^{n} \frac{1}{s! s^{i-s}} R_1^i \frac{1}{(n-i)!} R_2^{n-i} e^{-(R_1+R_2)}$$
$$\tilde{A} = \sum_{\substack{i,j|i\le s,\\ i+j\le n}} \frac{1}{i!j!} R_1^i R_2^j e^{-(R_1+R_2)}$$
$$\tilde{B_1} = \frac{1}{s!} R_1^s \frac{1}{1-\rho} \sum_{l=0}^{n-s} \frac{1}{l!} R_2^l e^{-(R_1+R_2)}$$
$$\tilde{B_2} = \frac{1}{s!} R_1^s \frac{\rho^{n-s+1}}{1-\rho} \sum_{l=0}^{n-s} \frac{1}{l!} \left(\frac{R_2}{\rho}\right)^l e^{-(R_1+R_2)}.$$

Note that by Lemmas 1,2 and 3

$$\lim_{\lambda\to\infty} \tilde{B_1} = \lim_{\lambda\to\infty} B_1 = \frac{\phi(\beta)\Phi(\eta)}{\beta},$$
$$\lim_{\lambda\to\infty} \tilde{B_2} = \lim_{\lambda\to\infty} B_2 = \frac{\phi(\sqrt{\eta^2+\beta^2})}{\beta} e^{\frac{1}{2}\eta_1^2}\Phi(\eta_1),$$
$$\lim_{\lambda\to\infty} \tilde{A} = \lim_{\lambda\to\infty} A = \int_{-\infty}^{\beta} \Phi\left(\eta + (\beta - t)\sqrt{\frac{\delta}{p\mu}}\right) d\Phi(t),$$

$$\begin{aligned}
\delta_1 &= \sum_{i=0}^{s} \frac{1}{i!} R_1^i \frac{1}{(n-i)!} R_2^{n-i} e^{-(R_1+R_2)} = \frac{1}{n!} e^{-(R_1+R_2)} \sum_{i=0}^{s} \frac{n!}{i!(n-i)!} R_1^i R_2^{n-i} \\
&= \frac{1}{n!} e^{-(R_1+R_2)} (R_1+R_2)^n \sum_{i=0}^{s} \frac{n!}{i!(n-i)!} \left(\frac{R_1}{R_1+R_2}\right)^i \left(\frac{R_2}{R_1+R_2}\right)^{n-i} \\
&= \frac{1}{n!} e^{-(R_1+R_2)} (R_1+R_2)^n \sum_{i=0}^{s} \mathbb{P}(X_\lambda = (i, n-i)) \\
&= \mathbb{P}(Y_\lambda = n) \sum_{i=0}^{s} \mathbb{P}(X_\lambda = (i, n-i)) = \mathbb{P}(Y_\lambda = n)\mathbb{P}(X_\lambda^1 \le s)
\end{aligned}$$

where $X_\lambda$ is a random variable with Multinomial distribution with parameters $(n, p_i, p_j)$, $p_i = \frac{R_1}{R_1+R_2}$, $p_j = \frac{R_2}{R_1+R_2}$, $Y_\lambda$ is a random variable with Poisson distribution with parameter $R_1 + R_2$, and $X_\lambda^1$ is a random variable with Binomial distribution with parameters $(n, p_i)$. By the CLT and the use of (27)

$$\begin{aligned}
\mathbb{P}(X_\lambda^1 \le s) &= \Phi\left(\frac{s - np_i}{\sqrt{np_i(1-p_i)}}\right) = \Phi\left(\frac{s - n\frac{R_1}{R_1+R_2}}{\sqrt{n\frac{R_1}{R_1+R_2}(1-\frac{R_1}{R_1+R_2})}}\right) = \Phi\left(\frac{s - n\frac{R_1}{R_1+R_2}}{\sqrt{n\frac{R_1}{R_1+R_2}\frac{R_2}{R_1+R_2}}}\right) \\
&= \Phi\left(\frac{s(R_1+R_2) - nR_1}{\sqrt{nR_1R_2}}\right) = \Phi\left(\sqrt{\frac{R_1}{R_2}} \frac{s\frac{R_1+R_2}{R_1} - n}{\sqrt{n}}\right) = \Phi\left(\sqrt{\frac{R_1}{R_2}} \frac{s(1+\frac{R_2}{R_1}) - n}{\sqrt{n}}\right) \\
&= \Phi\left(\sqrt{\frac{R_1}{R_2}} \frac{s + \frac{R_2}{\rho} - n}{\sqrt{n}}\right) = \Phi\left(-\sqrt{\frac{R_1}{R_2}}\sqrt{\frac{R_2}{n\rho}} \cdot \frac{n - s - \frac{R_2}{\rho}}{\sqrt{\frac{R_2}{\rho}}}\right) = \Phi\left(-\sqrt{\frac{s}{n}} \cdot \frac{n - s - \frac{R_2}{\rho}}{\sqrt{\frac{R_2}{\rho}}}\right) \\
&\approx \Phi\left(-\sqrt{\frac{R_1}{R_1+R_2}} \cdot \left(\eta - \beta\frac{R_2}{R_1}\right)\right) = \Phi\left(\frac{\beta\frac{R_2}{R_1} - \eta}{\sqrt{1+\frac{R_2}{R_1}}}\right) = \Phi\left(\frac{\beta\sqrt{\frac{p\mu}{\delta}} - \eta}{\sqrt{1+\frac{p\mu}{\delta}}}\right).
\end{aligned} \tag{45}$$

By the normal approximation of the Poisson distribution:

$$\mathbb{P}(Y_\lambda = n) \approx \frac{1}{\sqrt{R_1+R_2}}\phi\left(\frac{n-(R_1+R_2)}{\sqrt{R_1+R_2}}\right) \approx \frac{1}{\sqrt{s}\sqrt{1+\frac{p\mu}{\delta}}}\phi\left(\frac{\eta\sqrt{\frac{p\mu}{\delta}}+\beta}{\sqrt{1+\frac{p\mu}{\delta}}}\right). \tag{46}$$

We based on the following equivalences (as $\lambda$ tends to $\infty$) to develop Equations (45) and (46):

$$R_1+R_2 = \frac{\lambda}{(1-p)\mu} + \frac{p\lambda}{(1-p)\delta} \approx s + (1-p)s\mu\left(\frac{p}{(1-p)\delta}\right) = s\left(1+\frac{\mu p}{\delta}\right) = s\left(1+\frac{R_2}{R_1}\right);$$
$$n-(R_1+R_2) \approx s + R_2 + \eta\sqrt{R_2} - (R_1+R_2) = s + \eta\sqrt{R_2} - R_1 \approx s + \eta\sqrt{R_2} - s + \beta\sqrt{s}$$
$$\approx \eta\sqrt{\frac{s\mu p}{\delta}} + \beta\sqrt{s};$$
$$\frac{n-(R_1+R_2)}{\sqrt{R_1+R_2}} \approx \frac{\eta\sqrt{\frac{\mu p}{\delta}}+\beta}{\sqrt{1+\frac{\mu p}{\delta}}}.$$

Following Equations (45) and (46) we get

$$\delta_1 = \mathbb{P}(Y_\lambda = n)\mathbb{P}(X_\lambda^1 \le s) \approx \frac{1}{\sqrt{s}\sqrt{1+\frac{\mu p}{\delta}}}\phi\left(\frac{\eta\sqrt{\frac{\mu p}{\delta}}+\beta}{\sqrt{1+\frac{\mu p}{\delta}}}\right)\Phi\left(\frac{\beta\sqrt{\frac{\mu p}{\delta}}-\eta}{\sqrt{1+\frac{\mu p}{\delta}}}\right). \tag{47}$$

Now lets find an approximation for $\tilde{\delta_2}$.

$$\delta_2 = \sum_{i=s+1}^{n} \frac{1}{s!s^{i-s}}R_1^i \frac{1}{(n-i)!}R_2^{n-i}e^{-(R_1+R_2)} = \frac{e^{-(R_1+R_2)}}{s!s^{-s}}\sum_{i=s+1}^{n}\rho^i\frac{1}{(n-i)!}R_2^{n-i}$$
$$= \frac{e^{-(R_1+R_2)}}{s!s^{-s}}R_2^n\sum_{j=0}^{n-s-1}\frac{1}{j!}\left(\frac{\rho}{R_2}\right)^{n-j} = \frac{e^{-(R_1+R_2)}}{s!s^{-s}}\rho^n\sum_{j=0}^{n-s-1}\frac{1}{j!}\left(\frac{R_2}{\rho}\right)^{j}.$$

When comparing $\delta_2$ to $B_2$ form Equation (36), we observe that

$$\delta_2 = (1-\rho)B_2 \approx \frac{\beta}{\sqrt{s}}B_2.$$

Therefore, based on the approximation of $B_2$ from Lemma 2 we get

$$lim_{\lambda\to\infty}\delta_2 = \frac{\phi(\sqrt{\eta^2+\beta^2})}{\sqrt{s}}e^{\frac{1}{2}\omega^2}\Phi(\omega).$$

This proves Theorem 5.

The next theorem gives the approximation for the case where $\beta = 0$.

THEOREM 6. *Let the variables $\lambda$, $s$ and $n$ tend to $\infty$ simultaneously and satisfy the QED scaling conditions in (11) with $\beta = 0$. Define and $\xi = \frac{R_1}{R_2} = \frac{\delta}{p\mu}$, then*

$$\lim_{\lambda\to\infty}\sqrt{s}\mathbb{P}(block) = \frac{\sqrt{r}\phi(\gamma)\Phi(-\frac{\gamma\sqrt{r}}{\sqrt{1-r}}) + \frac{1}{\sqrt{2\pi}}\Phi(\eta)}{\int_{-\infty}^{0}\Phi\left(\eta - t\sqrt{\xi}\right)d\Phi(t) + \sqrt{\frac{1-r}{r}}\frac{1}{\sqrt{2\pi}}\left(\eta\Phi(\eta)+\phi(\eta)\right)}, \tag{48}$$

*where $\eta = \frac{\gamma-\beta\sqrt{r}}{\sqrt{1-r}}$.*

*Proof:* It follows from (5) of the manuscript that the probability of blocking is given by

$$\mathbb{P}_n = \frac{\delta_1+\delta_2}{\tilde{A}+\tilde{B}},$$

where

$$\delta_1 = \sum_{i=0}^{s} \frac{1}{i!} R_1^i \frac{1}{i!} R_2^{n-i} e^{-(R_1+R_2)}$$
$$\delta_2 = \sum_{i=s+1}^{n} \frac{1}{s! s^{i-s}} R_1^i \frac{1}{(n-i)!} R_2^{n-i} e^{-(R_1+R_2)}$$
$$\tilde{A} = \sum_{\substack{i,j|i\le s,\\ i+j\le n}} \frac{1}{i!j!} R_1^i R_2^j e^{-(R_1+R_2)}$$
$$\tilde{B} = \frac{1}{s!} R_1^s \frac{1}{1-\rho} \sum_{l=0}^{n-s} \frac{1}{l!} R_2^l e^{-(R_1+R_2)} - \frac{1}{s!} R_1^s \frac{\rho^{n-s+1}}{1-\rho} \sum_{l=0}^{n-s} \frac{1}{l!} \left(\frac{R_2}{\rho}\right)^l e^{-(R_1+R_2)}.$$

Note that by Lemmas 3 and 4

$$\lim_{\lambda\to\infty} \tilde{A} = \lim_{\lambda\to\infty} A = \int_{-\infty}^{\beta} \Phi\left(\eta + (\beta - t)\sqrt{\frac{\delta}{\mu p}}\right) d\Phi(t)$$
$$\lim_{\lambda\to\infty} \tilde{B} = \lim_{\lambda\to\infty} B = \sqrt{\frac{\mu p}{\delta}} \frac{1}{\sqrt{2\pi}} \left(\eta\Phi(\eta) + \phi(\eta)\right).$$

In addition, the approximations for $\delta_1$ and $\delta_2$ are the same as the proof of Theorem 5.

$$\lim_{\beta\to 0} \lim_{\lambda\to\infty} \sqrt{s}\delta_2 = \frac{1}{\sqrt{2\pi}} \Phi(\eta)$$
$$\lim_{\beta\to 0} \lim_{\lambda\to\infty} \sqrt{s}\delta_1 = \frac{1}{\sqrt{1+\frac{\mu p}{\delta}}} \phi\left(\frac{\eta}{\sqrt{1+\frac{\delta}{\mu p}}}\right) \Phi\left(-\frac{\eta}{\sqrt{1+\frac{\mu p}{\delta}}}\right).$$

This proves Theorem 6.

## Appendix D: Proof of Proposition 3

Define

$$A(s,n) = \sum_{k=0}^{s} \frac{k}{s} \binom{n}{k} b^k, \qquad B(s,n) = \sum_{k=s+1}^{n} \frac{k!}{s!} \binom{n}{k} s^{s-k} b^k, \qquad C(s,n) = \sum_{k=0}^{s} \binom{n}{k} b^k,$$

where $b = \delta/p\mu = r/(1-r)$. Then

$$\rho_J(s,n) = \frac{A(s,n) + B(s,n)}{C(s,n) + B(s,n)}.$$

Proving that $\rho_{\max}(s,n) \to 1$ as $R \to \infty$ with $s$ and $n$ as in (11) is equivalent to showing that

$$1 - \rho_{\max}(s,n) = \frac{C(s,n) - A(s,n)}{C(s,n) + B(s,n)} = \frac{(1+b)^{-n}[C(s,n) - A(s,n)]}{(1+b)^{-n}[C(s,n) + B(s,n)]} \to 0. \tag{49}$$

First, we rewrite

$$(1+b)^{-n} A(s,n) = (1+b)^{-n} \sum_{k=1}^{s} \frac{n}{s} \binom{n-1}{k-1} b^k = \frac{n}{s} \left(\frac{b}{1+b}\right) \sum_{k=0}^{s-1} \binom{n-1}{k} \left(\frac{b}{1+b}\right)^k \left(\frac{1}{1+b}\right)^{n-1-k}$$
$$= \frac{rn}{s} \sum_{k=0}^{s-1} \binom{n-1}{k} r^k (1-r)^{n-1-k} = \frac{rn}{s} \mathbb{P}(\mathrm{Bin}(n-1,r) \le s-1)$$
$$= \frac{rn}{s} \mathbb{P}\left(\frac{\mathrm{Bin}(n-1,r) - (n-1)r}{\sqrt{nr(1-r)}} \le \frac{s-1-(n-1)r}{\sqrt{nr(1-r)}}\right) \to \Phi\left(\frac{\beta - \gamma\sqrt{r}}{\sqrt{1-r}}\right),$$

since $nr/s = 1 + O(1/\sqrt{R_1})$. Also,

$$(1+b)^{-n}C(s,n) = \sum_{k=0}^{s} \binom{n}{k} \left(\frac{b}{1+b}\right)^k \left(\frac{1}{1+b}\right)^{n-k} = \sum_{k=0}^{s} \binom{n}{k} r^k (1-r)^{n-k}$$
$$= \mathbb{P}(\mathrm{Bin}(n,r) \leq s) \to \Phi\left(\frac{\beta - \gamma\sqrt{r}}{\sqrt{1-r}}\right).$$

Therefore, we have $(1+b)^{-n}[C(s,n) - A(s,n)] \to 0$ as $\lambda \to \infty$. For the remaining term,

$$(1+b)^{-n}B(s,n) = (1+b)^{-n} \sum_{k=s+1}^{n} \binom{n}{k} \frac{k!}{s!} s^{s-k} b^k = (1+b)^{-n} \frac{n!}{s!} s^s \sum_{k=s+1}^{n} \frac{1}{(n-k)!} \left(\frac{s}{b}\right)^{-k}$$
$$= (1+b)^{-n} \frac{n!}{s!} s^s \left(\frac{b}{s}\right)^n \sum_{k=s+1}^{n} \frac{1}{(n-k)!} \left(\frac{s}{b}\right)^{n-k} = r^n \frac{n!}{s!} s^{s-n} \sum_{m=0}^{n-s-1} \frac{1}{m!} \left(\frac{s}{b}\right)^m$$
$$= \left(\frac{r}{s}\right)^n \frac{n!}{s!} s^s \, \mathrm{e}^{s/b} \, \mathbb{P}(\mathrm{Pois}(s/b) \leq n-s-1),$$

in which

$$\mathbb{P}(\mathrm{Pois}(s/b) \leq n-s-1) = \mathbb{P}\left(\frac{\mathrm{Pois}(s/b) - s/b}{\sqrt{s/b}} \leq \frac{n-s-1-s/b}{\sqrt{s/b}}\right) \to \Phi\left(\frac{\gamma - \beta/\sqrt{r}}{\sqrt{1-r}}\right),$$

as $\lambda \to \infty$. By Stirling's approximation,

$$\left(\frac{r}{s}\right)^n \frac{n!}{s!} s^s \, \mathrm{e}^{s/b} \sim \left(\frac{r}{s}\right)^n \sqrt{\frac{n}{s}} \frac{n^n \mathrm{e}^{-n}}{s^s \mathrm{e}^{-s}} s^s \, \mathrm{e}^{s/b}$$
$$= \left(\frac{rn}{s}\right)^n \sqrt{\frac{n}{s}} \mathrm{e}^{-n+s+s/b} = \left(\frac{rn}{s}\right)^n \sqrt{\frac{n}{s}} \mathrm{e}^{-n+s/r}.$$

Since,

$$\frac{rn}{s} = 1 + \frac{\gamma\sqrt{r} - \beta}{\sqrt{R_1}} + O(1/R_1),$$

we find $\sqrt{n/s} = 1/\sqrt{r} + O(1/\sqrt{R_1})$ and

$$\log\left[\left(\frac{rn}{s}\right)^n \sqrt{\frac{n}{s}} \mathrm{e}^{-n+s/r}\right] = n \log\left[\frac{rn}{s}\right] - n + \frac{s}{r}$$
$$= -n\left[\left(1 - \frac{rn}{s}\right) + \frac{1}{2}\left(1 - \frac{rn}{s}\right)^2 + O(1/R^{3/2})\right] + \frac{s}{r}\left(1 - \frac{rn}{s}\right)$$
$$= \frac{s}{r}\left(1 - \frac{rn}{s}\right)^2 - \frac{n}{2}\left(1 - \frac{rn}{s}\right)^2 + O(1/\sqrt{R_1})$$
$$= \frac{(\gamma\sqrt{r} - \beta)^2}{2r} + O(1/\sqrt{R_1}),$$

as $R \to \infty$ and hence,

$$(1+b)^{-n}B(s,n) \to \phi\left(\frac{\gamma\sqrt{r} - \beta}{\sqrt{r}}\right) \Phi\left(\frac{\gamma - \beta/\sqrt{r}}{\sqrt{1-r}}\right).$$

Hence, we conclude that the denominator of (49) converges to a constant value as $\lambda$ grows, and hence the $1 - \rho_{\max}(s,n) \to 0$ as $\lambda \to \infty$.

## Appendix E: Numerical results on accuracy of the asymptotic approximations

To see how this heuristic approach performs under different settings, and particularly if $R_1 \to \infty$, we compare here the approximated delay probability in the Erlang-R model with holding as solution of the fixed-point procedure to the outcomes of simulation experiments.

In this section, we examine to accuracy of the asymptotic approximations of Theorem 1 and the heuristic method in Section 4.3. We perform our numerical experiments for three case instances, with parameter settings as in Table 1.

| | $\mu$ | $\delta$ | $p$ | $r$ |
|---|---|---|---|---|
| Case 1 | 1 | 0.10 | 0.90 | 0.10 |
| Case 2 | 1 | 0.25 | 0.75 | 0.25 |
| Case 3 | 1 | 0.50 | 0.50 | 0.50 |

**Table 1 Parameter settings for numerical experiments.**

### E.1. Restricted Erlang-R model with blocking

| | $\beta=1,\ \gamma=1$ | | | $\beta=1,\ \gamma=2$ | | | $\beta=2,\ \gamma=1$ | | |
|---|---|---|---|---|---|---|---|---|---|
| $R$ | $\mathbb{P}$(d) | $\sqrt{R_1}\mathbb{P}$(b) | $\sqrt{R_1}\mathbb{E}[W]$ | $\mathbb{P}$(d) | $\sqrt{R_1}\mathbb{P}$(b) | $\sqrt{R_1}\mathbb{E}[W]$ | $\mathbb{P}$(d) | $\sqrt{R_1}\mathbb{P}$(b) | $\sqrt{R_1}\mathbb{E}[W]$ |
| 5 | 0.1270 | 0.0900 | 0.2283 | 0.1553 | 0.0212 | 0.1085 | 0.0237 | 0.0868 | 0.0282 |
| 10 | 0.1340 | 0.0910 | 0.1919 | 0.1628 | 0.0206 | 0.1205 | 0.0206 | 0.0872 | 0.0188 |
| 25 | 0.1981 | 0.0945 | 0.1614 | 0.2356 | 0.0216 | 0.2145 | 0.0277 | 0.0876 | 0.0123 |
| 50 | 0.1513 | 0.0963 | 0.1588 | 0.1830 | 0.0205 | 0.1496 | 0.0185 | 0.0913 | 0.0116 |
| 100 | 0.1880 | 0.0956 | 0.1532 | 0.2231 | 0.0224 | 0.2055 | 0.0232 | 0.0888 | 0.0103 |
| 250 | 0.1797 | 0.0971 | 0.1399 | 0.2143 | 0.0219 | 0.2057 | 0.0203 | 0.0905 | 0.0093 |
| | *0.1767* | *0.0981* | *0.1437* | *0.2108* | *0.0217* | *0.1947* | *0.0188* | *0.0914* | *0.0084* |

**Table 2 Numerical results for Erlang-R model with blocking for Case 1.**

| | $\beta=1,\ \gamma=1$ | | | $\beta=1,\ \gamma=2$ | | | $\beta=2,\ \gamma=1$ | | |
|---|---|---|---|---|---|---|---|---|---|
| $R_1$ | $\mathbb{P}$(d) | $\sqrt{R_1}\mathbb{P}$(b) | $\sqrt{R_1}\mathbb{E}[W]$ | $\mathbb{P}$(d) | $\sqrt{R_1}\mathbb{P}$(b) | $\sqrt{R_1}\mathbb{E}[W]$ | $\mathbb{P}$(d) | $\sqrt{R_1}\mathbb{P}$(b) | $\sqrt{R_1}\mathbb{E}[W]$ |
| 5 | 0.0911 | 0.1538 | 0.0479 | 0.1431 | 0.0345 | 0.0909 | 0.0130 | 0.1484 | 0.0044 |
| 10 | 0.1010 | 0.1498 | 0.0560 | 0.1520 | 0.0326 | 0.1025 | 0.0121 | 0.1432 | 0.0042 |
| 25 | 0.1594 | 0.1509 | 0.1058 | 0.2192 | 0.0405 | 0.1785 | 0.0182 | 0.1383 | 0.0070 |
| 50 | 0.1201 | 0.1506 | 0.0726 | 0.1697 | 0.0381 | 0.1248 | 0.0119 | 0.1415 | 0.0043 |
| 100 | 0.1514 | 0.1539 | 0.1001 | 0.2088 | 0.0398 | 0.1704 | 0.0154 | 0.1413 | 0.0059 |
| 250 | 0.1459 | 0.1524 | 0.0957 | 0.2003 | 0.0397 | 0.1618 | 0.0136 | 0.1403 | 0.0051 |
| | *0.1429* | *0.1569* | *0.0940* | *0.1976* | *0.0391* | *0.1617* | *0.0126* | *0.1445* | *0.0048* |

**Table 3 Numerical results for Erlang-R model with blocking for Case 2.**

| | $\beta=1,\ \gamma=1$ | | | $\beta=1,\ \gamma=2$ | | | $\beta=2,\ \gamma=1$ | | |
|---|---|---|---|---|---|---|---|---|---|
| $R_1$ | $\mathbb{P}$(d) | $\sqrt{R_1}\mathbb{P}$(b) | $\sqrt{R_1}\mathbb{E}[W]$ | $\mathbb{P}$(d) | $\sqrt{R_1}\mathbb{P}$(b) | $\sqrt{R_1}\mathbb{E}[W]$ | $\mathbb{P}$(d) | $\sqrt{R_1}\mathbb{P}$(b) | $\sqrt{R_1}\mathbb{E}[W]$ |
| 5 | 0.0547 | 0.1945 | 0.0221 | 0.1181 | 0.0604 | 0.0617 | 0.0034 | 0.1888 | 0.0009 |
| 10 | 0.0579 | 0.2158 | 0.0237 | 0.1325 | 0.0526 | 0.0746 | 0.0030 | 0.2093 | 0.0008 |
| 25 | 0.1113 | 0.2086 | 0.0544 | 0.1959 | 0.0641 | 0.1311 | 0.0070 | 0.1937 | 0.0020 |
| 50 | 0.0813 | 0.2050 | 0.0363 | 0.1523 | 0.0562 | 0.0933 | 0.0043 | 0.1946 | 0.0011 |
| 100 | 0.1060 | 0.2146 | 0.0509 | 0.1873 | 0.0632 | 0.1250 | 0.0061 | 0.1999 | 0.0017 |
| 250 | 0.1006 | 0.2179 | 0.0475 | 0.1820 | 0.0596 | 0.1214 | 0.0052 | 0.2037 | 0.0014 |
| | *0.1011* | *0.2185* | *0.0478* | *0.1792* | *0.0605* | *0.1199* | *0.0052* | *0.2039* | *0.0014* |

**Table 4 Numerical results for Erlang-R model with blocking for Case 3.**

### E.2. Restricted Erlang-R model with holding

| | $\beta=1,\ \gamma=1$ | | $\beta=1,\ \gamma=2$ | | $\beta=2,\ \gamma=1$ | |
|---|---|---|---|---|---|---|
| $R_1$ | $\mathbb{P}$(d) | $\sqrt{R_1}\mathbb{E}[W]$ | $\mathbb{P}$(d) | $\sqrt{R_1}\mathbb{E}[W]$ | $\mathbb{P}$(d) | $\sqrt{R_1}\mathbb{E}[W]$ |
| 5 | 0.1532 | 0.1031 | 0.1628 | 0.1216 | 0.0310 | 0.0121 |
| 10 | 0.1622 | 0.1272 | 0.1697 | 0.1331 | 0.0267 | 0.0123 |
| 25 | 0.2340 | 0.2116 | 0.2413 | 0.2342 | 0.0348 | 0.0171 |
| 50 | 0.1817 | 0.1468 | 0.1890 | 0.1678 | 0.0240 | 0.0108 |
| 100 | 0.2199 | 0.1931 | 0.2304 | 0.2269 | 0.0293 | 0.0143 |
| 250 | 0.2110 | 0.1852 | 0.2176 | 0.2230 | 0.0256 | 0.0120 |
| | *0.2076* | *0.1777* | *0.2187* | *0.2050* | *0.0229* | *0.0104* |

**Table 5 Simulated and approximated probability of delay in Erlang-R model with holding for Case 1.**

| | $\beta=1,\ \gamma=1$ | | $\beta=1,\ \gamma=2$ | | $\beta=2,\ \gamma=1$ | |
|---|---|---|---|---|---|---|
| $R_1$ | $\mathbb{P}$(d) | $\sqrt{R_1}\mathbb{E}[W]$ | $\mathbb{P}$(d) | $\sqrt{R_1}\mathbb{E}[W]$ | $\mathbb{P}$(d) | $\sqrt{R_1}\mathbb{E}[W]$ |
| 5 | 0.1327 | 0.0740 | 0.1620 | 0.1096 | 0.0219 | 0.0079 |
| 10 | 0.1446 | 0.0894 | 0.1683 | 0.1207 | 0.0199 | 0.0073 |
| 25 | 0.2204 | 0.1631 | 0.2442 | 0.2203 | 0.0283 | 0.0128 |
| 50 | 0.1694 | 0.1122 | 0.1888 | 0.1507 | 0.0190 | 0.0078 |
| 100 | 0.2098 | 0.1524 | 0.2322 | 0.2111 | 0.0244 | 0.0097 |
| 250 | 0.2033 | 0.1534 | 0.2190 | 0.1979 | 0.0214 | 0.0083 |
| | *0.1840* | *0.1277* | *0.2109* | *0.1759* | *0.0169* | *0.0066* |

**Table 6 Simulated and approximated probability of delay in Erlang-R model with holding for Case 2.**

| | $\beta=1,\ \gamma=1$ | | $\beta=1,\ \gamma=2$ | | $\beta=2,\ \gamma=1$ | |
|---|---|---|---|---|---|---|
| $R$ | $\mathbb{P}$(d) | $\sqrt{R_1}\mathbb{E}[W]$ | $\mathbb{P}$(d) | $\sqrt{R_1}\mathbb{E}[W]$ | $\mathbb{P}$(d) | $\sqrt{R_1}\mathbb{E}[W]$ |
| 5 | 0.0977 | 0.0413 | 0.1521 | 0.0851 | 0.0072 | 0.0019 |
| 10 | 0.1070 | 0.0469 | 0.1648 | 0.1028 | 0.0067 | 0.0018 |
| 25 | 0.1926 | 0.1076 | 0.2421 | 0.1874 | 0.0148 | 0.0043 |
| 50 | 0.1431 | 0.0727 | 0.1876 | 0.1342 | 0.0092 | 0.0025 |
| 100 | 0.1855 | 0.1012 | 0.2282 | 0.1714 | 0.0132 | 0.0038 |
| 250 | 0.1775 | 0.0963 | 0.2217 | 0.1765 | 0.0114 | 0.0033 |
| | *0.1442* | *0.0711* | *0.1981* | *0.1354* | *0.0078* | *0.0022* |

**Table 7 Simulated and approximated probability of delay in Erlang-R model with holding for Case 3.**

## References

Jackson, J.R. 1963. Jobshop-like queueing systems. *Management Science* **10**(1) 131–142.

Neuts, M.F. 1981. *Matrix-Geometric Solutions in Stochastic Models*. The John Hopkins University Press, Baltimore.