Appendix
to
Set-Valued Queueing Approximations Given Partial Information

Yan Chen
Industrial Engineering and Operations Research, Columbia University, yc3107@columbia.edu

Ward Whitt
Industrial Engineering and Operations Research, Columbia University, ww2040@columbia.edu

This is an appendix to the main paper. Here is its abstract:

In order to understand queueing performance given only partial information about the model, we propose determining intervals of likely performance measures given that limited information. We illustrate this approach for the steady-state waiting time distribution in the $GI/GI/K$ queue given the first two moments of the interarrival-time and service-time distributions plus additional information about these underlying distributions, including support bounds, higher moments and Laplace transform values. As a theoretical basis, we apply the theory of Tchebycheff systems to determine extremal models (yielding tight upper and lower bounds) on the asymptotic decay rate of the steady-state waiting-time tail probability, as in the Kingman-Lundberg bound and large deviations asymptotics. We then can use these extremal models to indicate likely intervals of other performance measures. We illustrate by constructing such intervals of likely mean waiting times. Without extra information, the extremal models involve two-point distributions, which yield a wide range for the mean. Adding constraints on the third moment and a transform value produces three-point extremal distributions, which significantly reduce the range, yielding practical levels of accuracy.

Key words: performance approximations, queues, multi-server queues, bounds, mean waiting time, extremal queues, October 21, 2019
History: October 21, 2019

---

1. Overview of this Appendix

This appendix provides additional material expanding upon the main paper. It is divided into two parts. First, in §2-§5 we focus on the classic bounds based on the first two moments of the underlying cdf F of an interarrival time U and the cdf G of a service tims V, as in

$$(\mathbb{E}[U], \mathbb{E}[U^2], \mathbb{E}[V], \mathbb{E}[V^2]) \equiv (1, c_a^2, \rho, c_s^2) \quad (1)$$

taken from (4) of the main paper. In these sections we consider the impact of the support bounds, but not yet the Laplace transform values, which play a prominent role in our main method for generating intervals of likely values for the mean in §4 of the main paper.

In §2 we discuss the wide range of possible values of the mean $E[W]$ given only the the first two moments of F and G, without yet introducing the finite support bounds M_a for F and M_s for G. Then in §3 we elaborate on §4.1 of the main paper on how to choose the support bounds. In §4 we supplement Table 1 of the main paper by presenting additional tables studying the direct application of Theorem 5 of the paper, which gives bounds based on adding only support bounds to the parameters in (1). In §5 we relate the support bound constraints to extra third-moment constraints. In particular, we show that for appropriate choices, there is an explicit one-to-one correspondence between the third moment and the support bound. Consequently, provided that we decide to use a support bound, we can specify either the third moment or the support bound, and the other will be determined. However, our approach in this paper is to introduce support bounds that should have only negligible impact on the mean waiting time. Thus, it should not be surprising that support bounds associated with natural third moments tend to take smaller values.

In §6-§7 we present additional results related to the Laplace transform constraints in Theorem 6 of the main paper, which forms the basis of our approach to the mean $E[W]$. In §6 we expand the study in §4.3 and §4.4 of the main paper, complementing Table 3 of the main paper. In §7 we present additional tables for the $M/M/1$ and $M/M/2$ model complementing Tables 2 and 4 of the main paper.

2. A Wide Range for $\mathbb{E}[W]$ Given the First Two moments

The standard way to evaluate approximations such as the heavy-traffic approximation (HTA)

$$\mathbb{E}[W] \approx \frac{\rho^2(c_a^2 + c_s^2)}{2(1-\rho)}. \tag{2}$$

taken from (2) of the main paper is to compare it to simulation estimates for specific cases. This approach is of course excellent if we have a specific model we want to analyze. An alternative approach to obtain a broader understanding is to look at the set of all possible values, given the partial specification by the parameter 4-tuple in (1) when this can be done.

A principle conclusion of this line of work is that the range of possible values for $\mathbb{E}[W]$ given the partial information in (1) is remarkably wide. We now illustrate by providing simple approximation formulas for the absolute and relative errors, obtained by viewing the established bounds in a revealing way. In particular, it is helpful to look at the accuracy of the upper bound (UB) separately from the lower bound (LB), and it is helpful to use the simple HTA in (2) as a reference.

As an approximation for the UB, we use the (non-tight) Daley (1977) UB

$$\mathbb{E}[W] \leq \frac{\rho^2([(2-\rho)c_a^2/\rho]+c_s^2)}{2(1-\rho)}. \tag{3}$$

We could compute the exact tight UB using Chen and Whitt (2019), but we want to produce a simple formula. We use the tight LB

$$\mathbb{E}[W(LB)] = \frac{\rho((1+c_s^2)\rho-1)^+}{2(1-\rho)}. \tag{4}$$

The LB has long been known, see Stoyan and Stoyan (1974), §5.4 of Stoyan (1983), §V of Whitt (1984a), Ott (1987), Theorem 3.1 of Daley et al. (1992) and references there. It is significant that the LB is often 0 for smaller values of ρ; indeed it occurs whenever we can have $P(V-U\leq 0)=1$, which cannot be effectively prevented by moment constraints alone. The main paper shows that a third moment and the transform constraints address this shortcoming.

Let the absolute upper error (AUE) and the relative upper error (RUE) of the heavy-traffic approximation (2) (HTA) be defined by the formulas

$$AUE \equiv UB - HTA \quad \text{and} \quad RUE \equiv \frac{UB-HTA}{HTA}. \tag{5}$$

Similarly, let the absolute lower error (ALE) and the relative lower error (RLE) of the heavy-traffic approximation (2) be defined by the formulas

$$ALE \equiv HTA - LB \quad \text{and} \quad RUE \equiv \frac{HTA-LB}{HTA}. \tag{6}$$

We subtract the smaller from the larger in each case, so that these measures of the possible errors are always positive. We use the HTA in the denominator because it produces more revealing simple formulas. (If we divided by the bound, then the RUE would decrease, but the RLE would increase.)

When ρ and $\mathbb{E}[W]$ are not too small, it seems natural to focus on the relative error; otherwise it may be better to focus on the absolute error.

PROPOSITION 1. (*upper and lower errors for the mean*) *Suppose that we use the non-tight UB for the mean* $E[W]$ *in* (3) *and the tight LB in* (4). *For the LB, assume that* $\rho > 1/(1+c_s^2)$; *otherwise it must be* 0. *Then the upper and lower errors given the parameter four-tuple* $(1, c_a^2, \rho, c_s^2)$ *can be expressed as*

$$
\begin{aligned}
AUE &= \rho c_a^2, \\
RUE &= \left(\frac{2(1-\rho)}{\rho}\right)\left(\frac{c_a^2}{c_a^2+c_s^2}\right) \quad \textit{or} \quad \frac{1-\rho}{\rho} \quad \textit{if} \quad c_a^2 = c_s^2, \\
ALE &= \left(\frac{\rho^2}{2(1-\rho)}\right)\left(c_a^2 + \frac{1-\rho}{\rho}\right) = \left(\frac{\rho^2 c_a^2}{2(1-\rho)}\right) + \frac{\rho}{2} \quad \textit{and} \\
RLE &= \frac{c_a^2 + 1 - \rho}{c_a^2 + c_s^2} \quad \textit{or} \quad \frac{1}{2} + \frac{1-\rho}{c_a^2} \quad \textit{if} \quad c_a^2 = c_s^2.
\end{aligned} \tag{7}
$$

COROLLARY 1. (*monotonicity as functions of the parameters*) *The relative errors RUE and RLE in* (7) *are decreasing in* ρ *and* c_s^2 *but are increasing in* c_a^2.

COROLLARY 2. (*heavy traffic and light traffic*) *The upper errors are asympotically effective in the sense that* $RUE(\rho) \to 0$ *as* $\rho \uparrow 1$, *while* $AUE(\rho) \to 0$ *as* $\rho \downarrow 0$. *In contrast,* $RLE(\rho) \to c_a^2/(c_a^2 + c_s^2)$ *as* $\rho \uparrow 1$.

As in Corollary 1 of Whitt (1984a) for the $GI/M/1$ model, which shows that the overall relative error $(UB - LB)/LB$ for the mean queue length in the $GI/M/1$ model is c_a^2, Proposition 1 and Corollary 2 dramatically show the wide range of possible values. This suggests imposing further constraints on these distributions to concentrate on realistic "typical" cases, as was done in Klincewicz and Whitt (1984) and Whitt (1984b) for the $GI/M/1$ model. This program was extended to phase-type distributions by Johnson and Taaffe (1991, 1993). The main paper carries out the same program for the more general $GI/GI/1$ and $GI/GI/K$ models in a new way (by initially focusing on the asymptotic decay rate).

3. Elaboration on Specifying Appropriate Support Bounds

In this section we elaborate on §4.1 of the main paper, expanding upon the discussion there. As we wrote there, most applications of the $GI/GI/1$ queueing model do not have interarrival-times and service-time distributions with finite support. We introduce the support bounds M_a and M_s as a device to help expose the typical range of possible values of the simple approximations for decay rate θ_W in equation (5) and §3.1 of the main paper.. We propose using values of M_a and M_s that should have negligible impact on the mean waiting time in typical cases of interest, so that the bounds with M_a and M_s give a good indication of the likely set of possible values given the partial information. (§II.5.9 of Cohen (1982) provides theoretical support for this step.) Assumption 1 of the main paper about the critical singularity s^* of the moment generating function $\hat{g}(-s)$ is critical. We show how to construct support bounds that are typical as well ones that are conservative.

3.1. Starting from a Model or Data

Starting from a specific model with unbounded U and V, we suggest choosing the support bounds M_a and M_s so that

$$P(U > M_a\mathbb{E}[U]) = P(U > M_a) = P(V > M_s\mathbb{E}[V]) = P(V > \rho M_s) = \epsilon \tag{8}$$

for a suitably small ϵ such as 0.001. We might take $\epsilon = 0.0001$ to be more conservative or $\epsilon = 0.01$ to narrow the range (but losing confidence in the reliability). With ample data, we would estimate the corresponding empirical complementary cdf (ccdf) of the service time, and use the same criterion in (8).

Example 1. (the $M/M/1$ case)

As a helpful orientation, we first consider the $M/M/1$ queue with arrival rate $\lambda = 1$ and mean service time $\rho < 1$. Notice that the service-time complementary cdf (ccdf) is

$$P(V > x) = e^{-x/\rho}, \quad x \geq 0, \tag{9}$$

so that its decay rate $\theta_V \equiv \lim_{x\to\infty}\{-\log(P(V > x))/x\}$ is independent of x, i.e.,

$$\theta_V = -\log(P(V > x))/x = 1/\rho \quad \text{for all} \quad x. \tag{10}$$

while the associated waiting time ccdf is

$$P(W > x) = \rho e^{-(1-\rho)x/\rho}, \quad x \geq 0, \tag{11}$$

so that its decay rate is

$$\theta_W \equiv \lim_{x\to\infty} -\log(P(W > x)/x = (1-\rho)/\rho. \tag{12}$$

Hence, for the $M/M/1$ model we see that $\theta_W/\theta_V = 1 - \rho < 1$ which quantifies the well-known property that large waiting times ar more likely than large service times, becoming ever more so as the traffic intensity approaches 1.

Moreover, in the present light-tail case, provided that ρ is not too small, large waiting times are likely to be the result of several service times associated with a cluster of arrivals rather than one especially large service time.

For the $M/M/1$ model where U and V have exponential distributions, the target in (8) becomes $e^{-M} = \epsilon$. For $M_a = M_s = M = (4,5,6,7,8,9,10)$, the corresponding values are $\epsilon(M) = (0.0183, 0.0067, 0.0025, 0.0091, 0.0033, 0.00123, 0.00045)$. We use 7 and 9 in our experiments later.

Based on this analysis, we conduct a simulation comparison to show how the support bounds affect the decay rate θ_W of the extremal queues for the case $\rho = 0.7$ and $c_a^2 = c_s^2 = 1$ in Table 1. We implement Monte-Carlo Simulation with $N = 10^8$ and $R = 20$ to simulate the tail probability with different quantities and report 95% confidence interval length (CIL).

Table 1 **Simulation comparison of the waiting time ccdf and delay rate θ_W for two-point extremal models with $M_a = \{1+c_a^2, 5, 10\}$ and $M_s = \{1+c_s^2, 5, 10\}$ under $c_a^2 = c_s^2 = 1, \rho = 0.7$**

$M_a=2, M_s=10$	$x=10$	CIL	$x=12$	CIL	$x=14$	CIL	$x=16$	CIL	$x=18$	CIL
$P(W>x)$	2.90E-02	3.65E-05	3.54E-01	3.63E-05	8.85E-03	2.82E-05	4.83E-03	1.49E-05	2.64E-03	1.16E-05
$-\log(P(W>x))/x$	3.54E-01		3.44E-01		3.38E-01		3.33E-01		3.30E-01	
$M_a=2, M_s=5$	$x=10$	CIL	$x=12$	CIL	$x=14$	CIL	$x=16$	CIL	$x=18$	CIL
$P(W>x)$	1.94E-02	3.02E-05	9.42E-03	3.04E-05	4.56E-03	1.54E-05	2.21E-03	1.20E-05	1.07E-03	8.66E-06
$-\log(P(W>x))/x$	3.94E-01		3.89E-01		3.85E-01		3.82E-01		3.80E-01	
$M_a=2, M_s=2$	$x=10$	CIL	$x=12$	CIL	$x=14$	CIL	$x=16$	CIL	$x=20$	CIL
$P(W>x)$	1.31E-02	2.68E-05	5.77E-03	2.81E-05	2.51E-03	1.55E-05	1.08E-03	7.94E-06	4.58E-04	5.21E-06
$-\log(P(W>x))/x$	4.33E-01		4.30E-01		4.28E-01		4.27E-01		4.27E-01	
$M_a=5, M_s=2$	$x=4$	CIL	$x=5$	CIL	$x=6$	CIL	$x=7$	CIL	$x=8$	CIL
$P(W>x)$	7.50E-02	4.46E-05	4.12E-02	3.00E-05	2.32E-02	2.20E-05	1.27E-02	1.67E-05	6.97E-03	1.47E-05
$-\log(P(W>x))/x$	6.48E-01		6.38E-01		6.28E-01		6.24E-01		6.21E-01	
$M_a=10, M_s=2$	$x=4$	CIL	$x=5$	CIL	$x=6$	CIL	$x=7$	CIL	$x=8$	CIL
$P(W>x)$	2.39E-02	2.13E-05	9.98E-03	2.35E-05	4.24E-03	9.87E-06	1.73E-03	6.92E-06	7.08E-04	5.77E-06
$-\log(P(W>x))/x$	9.33E-01		9.21E-01		9.11E-01		9.08E-01		9.07E-01	

Table 1 shows that there is rapid convergence of $-\log(P(W>x))/x$ to the decay rate θ_W as x increases; it is not necessary to make x extraordinarily large. Notice that the estimated decay rate is monotone in Table 1, with the $M/M/1$ exact value $(1-\rho)/\rho = 0.3/0.7 = 0.4285$ bounded below and above by the values for $(M_a, M_s) = (2,5)$ and $(5,2)$ taken from the last column of Table 1.

4. Direct Application of Theorem 5 to the Mean $\mathbb{E}[W]$

We now elaborate on §4.1 of the main paper by providing additional results about how the extremal UB model $F_0/G_u/1$ and LB model $F_u/G_0/1$ for the decay rate from Theorem 5 of the main paper apply to the mean $\mathbb{E}[W]$ with $K=1$ when we introduce the support bounds M_a and M_s following the prescription in §4.1 of the main paper.

This issue relates strongly to Chen and Whitt (2018), which studied the extremal models for $E[W]$. For the mean, there is strong evidence (but not yet a mathematical proof) that the model $F_0/G_u/1$ directly yields the UB for the mean, for both bounded and unbounded support, just as it does for the decay rate. However, the situation is different for the LB, as discussed in §7 and EC.6 of Chen and Whitt (2018). For unbounded support, the tight LB is given here in (4). It is attained by the $D/A_3/1$ model, where A_3 denotes a three-point distribution, which has all mass on multiples of the deterministic interarrival time. The D interarrival time violates the moment condition, but nevertheless is attained asymptotically. In Chen and Whitt (2018) we present evidence that the $F_u/A_3(u)/1$ model attains the LB, where $A_3(u)$ is a natural analog of A_3

4.1. Elaborating on Table 1 of the Main Paper

Table 1 of the main paper shows how the support bounds reduce the range of the possible values of $\mathbb{E}[W]$. It reports results for the cases $(c_a, c_s^2) = (1.0, 1.0)$, $(4.0, 4.0)$, $(0.5, 0.5)$, $(4.0, 0.5)$, $(0.5, 4.0)$.

We now supplement Table 1 of the main paper by showing the UB and LB for the mean $\mathbb{E}[W]$ with the support bounds chosen to satisfy (8) with targets $\epsilon = 0.001$ and 0.0001 for all four cases of $c_a^2, c_s^2 \in \{0.5, 4.0\}$ for 10 values of ρ.

Table 2 Evaluation of $\mathbb{E}[W]$ for $F_u/G_0/1$ and $F_0/G_u/1$ with (M_a, M_s) for $c_a^2 = c_s^2 = 4$

ρ	Tight LB	$M_a = 39.9$	$M_a = 31.1$	HTA	$M_s = 31.1$	$M_s = 39.3$	Tight UB
0.10	0.000	0.000	0.000	0.044	0.401	0.402	0.422
0.20	0.000	0.026	0.033	0.200	0.867	0.873	0.904
0.30	0.107	0.172	0.186	0.514	1.453	1.463	1.499
0.40	0.333	0.458	0.498	1.067	2.254	2.265	2.304
0.50	0.750	1.013	1.097	2.000	3.419	3.430	3.470
0.60	1.500	2.079	2.282	3.600	5.239	5.251	5.295
0.70	2.917	4.303	4.748	6.533	8.384	8.394	8.441
0.80	6.000	9.829	10.697	12.800	14.856	14.865	14.917
0.90	15.750	28.924	30.239	32.400	34.658	34.671	34.721
0.95	35.625	68.695	70.106	72.200	74.553	74.568	74.621

Table 3 Evaluation of $\mathbb{E}[W]$ for $F_u/G_0/1$ and $F_0/G_u/1$ with (M_a, M_s) for $c_a^2 = c_s^2 = 0.5$

ρ	Tight LB	$M_a = 4.5$	$M_a = 3.5$	HTA	$M_s = 3.5$	$M_s = 3.5$	Tight UB
0.10	0.000	0.000	0.000	0.006	0.050	0.050	0.053
0.20	0.000	0.000	0.000	0.025	0.101	0.101	0.113
0.30	0.000	0.000	0.000	0.064	0.159	0.163	0.184
0.40	0.000	0.000	0.000	0.133	0.243	0.255	0.280
0.50	0.000	0.000	0.000	0.250	0.377	0.388	0.414
0.60	0.000	0.076	0.164	0.450	0.588	0.601	0.637
0.70	0.058	0.410	0.530	0.817	0.966	0.982	1.017
0.80	0.400	1.167	1.311	1.600	1.760	1.774	1.822
0.90	1.575	3.613	3.771	4.050	4.207	4.229	4.295
0.95	4.037	8.596	8.735	9.025	9.185	9.220	9.284

Table 4 **Evaluation of $\mathbb{E}[W]$ for $F_u/G_0/1$ and $F_0/G_u/1$ with (M_a, M_s) for $c_a^2=4, c_s^2=0.5$**

ρ	Tight LB	$M_a=39.9$	$M_a=31.1$	HTA	$M_s=3.5$	$M_s=4.5$	Tight UB
0.10	0.000	0.000	0.000	0.025	0.400	0.400	0.403
0.20	0.000	0.000	0.000	0.113	0.805	0.806	0.816
0.30	0.000	0.000	0.000	0.289	1.253	1.254	1.274
0.40	0.000	0.000	0.000	0.600	1.806	1.808	1.837
0.50	0.000	0.000	0.000	1.125	2.556	2.559	2.595
0.60	0.000	0.005	0.065	2.025	3.669	3.675	3.720
0.70	0.058	0.342	0.450	3.675	5.524	5.533	5.583
0.80	0.400	1.268	1.798	7.200	9.250	9.261	9.317
0.90	1.575	9.075	11.988	18.225	20.469	20.486	20.546
0.95	4.037	32.083	34.934	40.613	42.955	42.970	43.033

Table 5 **Evaluation of $\mathbb{E}[W]$ for $F_u/G_0/1$ and $F_0/G_u/1$ with (M_a, M_s) for $c_a^2=0.5, c_s^2=4.0$**

ρ	Tight LB	$M_a=4.5$	$M_a=3.5$	HTA	$M_s=31.1$	$M_s=39.9$	Tight UB
0.10	0.000	0.000	0.000	0.025	0.064	0.065	0.072
0.20	0.000	0.034	0.046	0.113	0.184	0.188	0.200
0.30	0.107	0.179	0.192	0.289	0.388	0.393	0.409
0.40	0.333	0.462	0.487	0.600	0.720	0.726	0.746
0.50	0.750	0.957	0.988	1.125	1.263	1.270	1.289
0.60	1.500	1.841	1.869	2.025	2.176	2.186	2.212
0.70	2.917	3.464	3.494	3.675	3.841	3.851	3.875
0.80	6.000	6.973	6.985	7.200	7.374	7.379	7.422
0.90	15.750	17.973	17.993	18.225	18.408	18.427	18.470
0.95	35.625	40.183	40.322	40.613	40.811	40.826	40.871

Table 1 of the main paper and Tables 2-5 above show that the support bounds reduce the range of possible value in all cases. The tables also show that the cases differ dramatically. Just as in Corollaries 1 and 2, we see that the relative errors are remarkably small for

$(c_a^2, c_s^2) = (0.5, 4.0)$, but remarkably large for $(c_a^2, c_s^2) = (4.0, 0.5)$, even with the support bounds.

4.2. Using Heavy-Traffic Approximations

To show that we could also start from the HT approximations for the mean $E[W]$ in (2) and for the decay rate θ_W in (8) of the main paper instead of the exact models based on E_2 and H_2 distributions, Table 6 compares the exact values of θ_W and $\mathbb{E}[W]$ to these heavy-traffic approximations. Table 6 shows that the HTA in (2) overestimates the exact value when $c_a^2 = 0.5$, which is consistent with the refinement in (44) and (45) of Whitt (1983a).

Table 6 Decay rates and mean values: exact compared to the HT approximations in (2) and (??)

$c_a^2 = c_s^2 = 0.5$					$c_a^2 = c_s^2 = 4$				
ρ	Exact θ_W	Approximate θ_W	Exact $\mathbb{E}[W]$	HTA	ρ	Exact θ_W	Approximate θ_W	Exact $\mathbb{E}[W]$	HTA
0.5	2.00	2.00	0.195	0.250	0.5	0.244	0.250	2.02	2.00
0.7	0.857	0.857	0.725	0.817	0.7	0.106	0.107	6.61	6.53
0.9	0.222	0.222	3.92	4.05	0.9	0.0278	0.0278	32.6	32.4
$c_a^2 = 4, c_s^2 = 0.5$					$c_a^2 = 0.5, c_s^2 = 4$				
ρ	Exact θ_W	Approximate θ_W	Exact $\mathbb{E}[W]$	HTA	ρ	Exact θ_W	Approximate θ_W	Exact $\mathbb{E}[W]$	HTA
0.5	0.826	0.444	0.882	1.13	0.5	0.311	0.444	1.05	1.13
0.7	0.260	0.190	3.37	3.68	0.7	0.153	0.190	3.56	3.68
0.9	0.0537	0.049	18.0	18.2	0.9	0.0458	0.0494	18.0	18.2

These tables show that the support bounds reduce the range of possible value in all cases. The tables also show that the cases differ dramatically. Just as in Corollaries 1 and 2, we see that the relative errors are remarkably small for $(c_a^2, c_s^2) = (0.5, 4.0)$, but remarkably large for $(c_a^2, c_s^2) = (4.0, 0.5)$, even with the support bounds.

5. Relating Third Moments to Support Bounds

So far, we have obtained a reduced range of possible values of $\mathbb{E}[W]$, by introducing the support bounds M_a and M_s in addition to the model parameters $(1, c_a^2, \rho, c_s^2)$. We can then apply Theorem 5 of the main paper. In §3 we chose M_a and M_s so that the approximate tail probability was suitably small, as in (8). To cover typical distributions, we used the approximate tail probabilities based on the decay rates of typical distributions.

An alternative way is to exploit third moments. For third moments, we might also specify candidate values by looking at candidate distributions with the given parameters $(1, c_a^2, \rho, c_s^2)$. Indeed that was done in §5.1 of Whitt (1983b), and we use the same prescription here. For $c_a^2 \geq 1$, based on the H_2 distribution with balanced means as before, $m_{3,a} = 3c_a^2(1 + c_a^2)$. For $c_a^2 \leq 1$, based on the E_k distribution, let $m_{3,a} = (2c_a^2 + 1)(c_a^2 + 1)$.

We apply these "typical" third moments to go with $(1, c_a^2, \rho, c_s^2)$ by relating the third moments to the support bounds M associated with F_u and G_u. We observe that the third moment of F_u is

$$m_3^U = \frac{c_a^2 M_a^3}{c_a^2 + (M_a - 1)^2} + \frac{(M_a - 1 - c_a^2)^3}{(M_a - 1)(c_a^2 + (M_a - 1)^2)}, \tag{13}$$

while the third moment of F_0 is

$$m_3^L = \frac{c_a^2 (1 + c_a^2)^3}{c_a^2 + c_a^4}. \tag{14}$$

Now observe that the third moment in (13) is a strictly increasing function of M_a, so that we can invert it to obtain M_a as a function of m_3, getting

$$M_a = \frac{-1 - c_a^2 + m_3^U + \sqrt{1 + 6c_a^2 + 9c_a^2 + 4c_a^3 - 2m_3^U - 6c_a^2 m_3^U + (m_3^U)^2}}{2c_a^2}. \tag{15}$$

Hence, given typical values of m_3 associated with any parameter 4-tuple $(1, c_a^2, \rho, c_s^2)$, we can construct a corresponding support bound M_a^* for which we can determine the range of possible mean values. With this approach, we obtain $M_a^* = 13.081$ for $c_a^2 = 4$, $M_a^* = 3.414$ for $c_a^2 = 1$ and $M_a^* = 2.366$ for $c_a^2 = 0.5$. We note that these values are substantially smaller than the values determined in §3. Table 7 presents the numerical ranges of third moment as a function of c^2 and M.

Table 7 **Reasonable ranges in third moment relating to reasonable setting of bounded support**

M	5.25	6.75	7.00	9.00	17.5	22.5	m_3^L
$c^2 = 0.5$	4.566	5.332	5.458	6.469	10.735	13.238	2.25
$c^2 = 1$	8.015	9.576	9.833	11.875	20.439	25.454	4.000
$c^2 = 4$	26.235	33.217	34.333	43.000	78.030	98.256	25.000

We again applied simulation to study the performance of the extremal queues based on the third moments in addition to the basic model parameters in (1). As a first step, we use the support bounds that come from the third moments via (15). Based on Theorem 5 of the main paper, the candidate UB and LB models for the mean $E[W]$, based on the reverse order for θ_W, are $F_0/G_u/1$ and $F_u/G_0/1$. As shown in Chen and Whitt (2018), there is strong evidence that $F_0/G_u/1$ actually yields the tight UB for the mean $E[W]$, but it known that $F_u/G_0/1$ is not actually the tight LB for $E[W]$, although it is close; see §7 and §EC.6 of Chen and Whitt (2018). There it is conjectured that the LB for $E[W]$ is attained by a special three-point distribution, denoted by $A_3(u)$. Table 8 compares the resulting LB and UB extremal queues to the HTA in (2). We include results for both the $F_u/A_3(u)/1$ and $F_u/G_0/1$ candidate LB models. (We again refer to §§7 and EC.6 for background.)

Table 8 Range of mean waiting times after including typical third moments: balanced models

	$c_a^2=4, c_s^2=0.5, M_a=13.1, M_s=2.37$				$c_a^2=4, c_s^2=4, M_a=13.1, M_s=13.1$			
ρ	$\mathbb{E}[W(F_u,A_3(u))]$	$\mathbb{E}[W(F_u,G_0)]$	HTA	$\mathbb{E}[W(F_u,G_0)]$	$\mathbb{E}[W(F_u,A_3(u))]$	$\mathbb{E}[W(F_0,G_u)]$	HTA	$\mathbb{E}[W(F_0,G_u)]$
0.50	0.040	0.154	1.125	2.552	0.917	1.856	2.000	3.336
0.60	0.174	0.737	2.025	3.663	1.750	3.622	3.600	5.148
0.70	1.213	2.577	3.675	5.515	3.886	6.768	6.533	8.289
0.80	5.799	6.572	7.200	9.235	10.591	13.263	12.800	14.758
0.90	17.447	17.988	18.225	20.451	30.510	33.004	32.400	34.555
	$c_a^2=0.5, c_s^2=0.5, M_a=2.37, M_s=2.37$				$c_a^2=0.5, c_s^2=4, M_a=2.37, M_s=13.1$			
ρ	$\mathbb{E}[W(F_u,A_3(u))]$	$\mathbb{E}[W(F_u,G_0)]$	HTA	$\mathbb{E}[W(F_u,G_0)]$	$\mathbb{E}[W(F_u,A_3(u))]$	$\mathbb{E}[W(F_0,G_u)]$	HTA	$\mathbb{E}[W(F_0,G_u)]$
0.50	0.058	0.131	0.250	0.348	0.933	1.015	1.125	1.216
0.60	0.200	0.353	0.450	0.550	1.775	1.886	2.025	2.124
0.70	0.537	0.725	0.817	0.924	3.385	3.516	3.675	3.780
0.80	1.336	1.496	1.600	1.714	6.884	7.019	7.200	7.311
0.90	3.749	3.935	4.050	4.169	17.899	17.994	18.225	18.340

From Table 8, we see that

$$\mathbb{E}[W(F_u,A_3(u))] \leq \mathbb{E}[W(F_u,G_0)] \leq \mathbb{E}[W(F_0,G_u)]$$

in all cases. In addition,

$$\mathbb{E}[W(F_u,G_0)] \leq HTA \leq \mathbb{E}[W(F_0,G_u)]$$

in all cases except $(c_a^2, c_s^2) = (4.0, 4.0)$. In that case, the smaller values of M than produced by (8) makes it important to use the better LB model $(F_u/A_3(u))$.

Most important, we see that our range of possible values of the mean $\mathbb{E}[W]$ is reduced substantially by adding the additional parameters (M_a, M_s) obtained from $(m_{a,3}, m_{s,3})$. To illustrate, note that the range in the case $(\rho, c_a^2, c_s^2) = (0.8, 4, 4)$ is reduced from $[6.000, 14.917]$ to $[10.593, 14.758]$. The change is obviously much greater for F than for G.

6. The Impact of the Laplace Transform Constraints

We now elaborate on §4 of the main paper, which investigates the application of Theorem 6 in the main paper to obtain shorter intervals of likely values for the mean $E[W]$ by exploiting values of the Laplace transform $\hat{f}(s)$ and the moment generating function (mgf) $\hat{g}(-s)$. Recall that the Laplace transform is defined as

$$\hat{f}(s) \equiv \int_0^\infty e^{-st}\, dF(t), \quad s \geq 0 \tag{16}$$

When we look at $\hat{g}(-s)$, it corresponds to the mgf, i.e.,

$$\hat{g}(-s) \equiv \int_0^\infty e^{st}\, dG(t) = E[e^{sV}], \quad s \geq 0. \tag{17}$$

We now show how a direct application of Theorem 6 in the main paper reduces the range. In this section we avoid issues involving the singularity s^* in Assumption 1 of the main paper, by primarily considering case (ii) in (30) of Theorem 6 in the main paper, in which

$$\mu_s < \theta_W < \mu_a, \tag{18}$$

which we achieve by following (7) of the main paper, i.e.,

$$\mu_s \equiv \theta_W / R \quad \text{and} \quad \mu_a \equiv R\theta_W \tag{19}$$

for suitable R. We begin by considering a range of R.

6.1. The Impact of Truncation

We initially truncate the basic models by M_a, M_s because Theorem 6 of the main paper only applies to models with bounded support. In the implementation, we do not want to $\mu_s > s^*$. Thus, if we are considering one of the cases with $\mu_s \geq \theta_W$, then we first check to see if $R\theta_W > s^*$ for our largest value of R, which we take to be $R = 20$. If it is, then we create alternative values of μ_s in the interval (θ_W, s^*). In particular, we use

$$\mu_s \equiv \theta_W + \left(\frac{R}{25}\right)(s^* - \theta_W), \quad 1 \leq k \leq 4, \tag{20}$$

so that the values of R remain in $\{5, 10, 15, 20\}$, but all values are within (θ_W, s^*).

Table 9 shows a careful comparison between parameters under truncation or not.

Table 9 **A numerical comparison of truncated and original Laplace transform values for** $E_2/H_2/1$ $(\theta_W = 0.1527)$ **and** $M/M/1$ $(\theta_W = 0.4286)$

	$E_2/H_2/1$	Truncated	Original	Truncated	Original	$M/M/1$	Truncated	Original	Truncated	Original
	$R_a = R_s$	$\hat{f}(s)$	$\hat{f}(s)$	$\hat{g}(-s)$	$\hat{g}(-s)$	$R_a = R_s$	$\hat{f}(s)$	$\hat{f}(s)$	$\hat{g}(-s)$	$\hat{g}(-s)$
$\mu_a, \mu_s \geq \theta_W$	1	0.8630	0.8632	1.1568	1.1585	1	0.6998	0.7000	1.4293	1.4286
	5	0.5229	0.5238	1.4582	1.5498	5	0.3180	0.3182	4.9983	4.9779
	10	0.3205	0.3216	1.4582	1.5498	10	0.1890	0.1892	4.9983	4.9779
	20	0.1558	0.1566	1.4582	1.5498	20	0.1044	0.1045	4.9983	4.9779
$\mu_a, \mu_s \leq \theta_W$	1	0.8630	0.8632	1.1568	1.1585	1	0.6998	0.7000	1.4293	1.4286
	5	0.9701	0.9701	1.0226	1.0226	5	0.9210	0.9210	1.0639	1.0638
	10	0.9849	0.9849	1.0110	1.0110	10	0.9589	0.9589	1.0310	1.0309
	20	0.9924	0.9924	1.0054	1.0054	20	0.9790	0.9790	1.0152	1.0152

Additionally, the first three moments with and without truncation are close, i.e, $s_2 = 2.44, s_3 = 20.191$ for the truncated model and $s_2 = 2.45, 20.58$ for the original model for $E_2/H_2/1$. Since difference between parameters are negligible, it may suffice to apply the original $m_{a,2}, m_{a,3}$ instead of $m''_{a,2}, m''_{a,3}$ for reducing computation complexity. In other word, we could ignore the truncation effect and simply apply Theorem 6 of the main paper using parameters of the basic models without truncation.

6.2. The Parameter Pair (R_a, R_s)

However, we also report results exploring a more general two-parameter range, using (R_a, R_s) with R_a applying to F and R_s applying to G. In summary, we proceed as follows: Given an initially specified decay rate θ_W, the range vector (R_a, R_s) with $R_s \leq 1 \leq R_a$ and the specified parameters $(1, c_a^2, m_{a,3}, \mu_a, M_a)$ partially characterizing F and $(1, c_s^2, m_{s,3}, \mu_s, M_s)$ partially characterizing G, where $\mu_s \equiv \theta_W/R_s \leq \theta_W < R_a\theta_W$, we identify the set of possible performance measures in two steps.

In the first step, we can determine the extremal distributions F_L, G_L, F_U, G_U by solving n equations in n unknowns for the appropriate n. In the second step, we simulate $\mathbb{E}[W(F_L, G_L)]$ and $\mathbb{E}[W(F_U, G_U)]$ by Monte-Carlo simulation and obtain decay rates by solving equation (6) of the main paper for the LB and UB models F_L/G_L and F_U/G_U.

We now illustrate the results.

6.3. The $H_2/H_2/1$ Model with $c_a^2 = c_s^2 = 4.0$

We use UB (LB) to refer to the minimum (maximum) decay rate, which yields our estimate of the UB (LB) for $E[W]$. Table 10 shows estimates of the UB and LB for the decay rate θ_W and the mean $\mathbb{E}[W]$ in the case $c_a^2 = c_s^2 = 4$ and $\rho = 0.7$ for a range of R_a and R_s varying from 1 to 20, based on the first three moments and LT transforms from model $H_2/H_2/1$ with balanced means, which has exact mean $\mathbb{E}[W(H_2, H_2)] = 6.608$ and exact decay rate $\theta_W = 0.1064$. (See Table 6.)

As indicated above, here we allow R_a and R_s to differ, but we still require that $\mu_s \equiv \theta_W/R_s$ and $\mu_a \equiv R_a\theta$ for $R_a \geq 1$ and $R_s \geq 1$.

Table 10 The improved LB and UB based on information of $H_2/H_2/1$

$c_a^2 = c_s^2 = 4$, $\rho = 0.7$, $\theta_W = 0.1064$ and exact $\mathbb{E}[W(H_2, H_2)] = 6.608$										
	$R_s \backslash R_a$	1	5	10	20		$R_s \backslash R_a$	1	5	10	20
θ_W (UB)	1	0.106	0.105	0.104	0.104	$\mathbb{E}[W]$ (UB)	1	6.292	6.197	6.175	6.708
	5	0.106	0.105	0.104	0.103		5	6.329	6.275	6.163	6.684
	10	0.106	0.105	0.104	0.103		10	6.336	6.278	6.155	6.688
	20	0.100	0.099	0.099	0.098		20	6.884	7.047	7.134	7.225
	$R_s \backslash R_a$	1	5	10	20		$R_s \backslash R_a$	1	5	10	20
θ_W (LB)	1	0.106	0.107	0.108	0.108	$\mathbb{E}[W]$ (LB)	1	6.667	6.516	6.465	6.410
	5	0.107	0.108	0.108	0.108		5	6.692	6.485	6.443	6.387
	10	0.107	0.108	0.108	0.108		10	6.693	6.474	6.448	6.386
	20	0.109	0.110	0.110	0.110		20	6.866	6.664	6.532	6.420

Consistent with part (d) of Theorem 6 in the main paper, Table 10 show that the UB decay rate θ_W is monotone decreasing in R_a and R_s, while the LB decay rate is monotone increasing. Moreover, recall we utilize information from the exact queueing models where F and G have unbounded support, so that we do not expect perfect consistency. On the other hand, there is less order in the corresponding values of $E[W]$. Nevertheless, from Table 10, we conclude that a reasonable range of $E[W]$ can be generated by $R_a = R_s = 20$.

To elaborate further, Table 11 shows the explicit numerical values of the three-point extremal distributions F_L, G_L and F_U, G_U obtained in the case $c_a^2 = c_s^2 = 4, \rho = 0.7$ with $R = R_a = R_s \in \{1, 5, 10, 20\}$, supporting Table 10.

Table 11 Numerical examples of extremal distributions

$R_a=R_s=1$	F			G			$R_a=R_s=5$	F			G		
$F_L/G_L/1$	q_1	q_2	q_3	p_1	p_2	p_3	$F_L/G_L/1$	q_1	q_2	q_3	p_1	p_2	p_3
	0.620	0.370	1.04E-02	0.677	0.317	6.08E-03		0.526	0.459	1.57E-02	0.656	0.336	7.69E-03
	y_1	y_2	y_3	x_1	x_2	x_3		y_1	y_2	y_3	x_1	x_2	x_3
	0	2.21	17.6	0	1.93	14.4		0.0	1.65	15.5	0	1.78	13.4
$F_U/G_U/1$	q_1	q_2	q_3	p_1	p_2	p_3	$F_U/G_U/1$	q_1	q_2	q_3	p_1	p_2	p_3
	0.956	0.0433	2.88E-04	0.965	0.0345	1.73E-04		0.936	0.0639	4.30E-04	0.963	0.0370	2.12E-04
	y_1	y_2	y_3	x_1	x_2	x_3		y_1	y_2	y_3	x_1	x_2	x_3
	0.587	9.86	39.9	0.440	7.86	27.9		0.505	7.99	39.9	0.431	7.54	27.9
$R_a=R_s=10$	F			G			$R_a=R_s=20$	F			G		
$F_L/G_L/1$	q_1	q_2	q_3	p_1	p_2	p_3	$F_L/G_L/1$	q_1	q_2	q_3	p_1	p_2	p_3
	0.451	0.530	1.87E-02	0.654	0.338	7.88E-03		0.358	0.621	2.14E-02	0.653	0.339	7.97E-03
	y_1	y_2	y_3	x_1	x_2	x_3		y_1	y_2	y_3	x_1	x_2	x_3
	0.0	1.37	14.6	0	1.76	13.4		0	1.13	14.0	0	1.75	13.3
$F_U/G_U/1$	q_1	q_2	q_3	p_1	p_2	p_3	$F_U/G_U/1$	q_1	q_2	q_3	p_1	p_2	p_3
	0.917	0.0828	5.02E-04	0.962	0.0374	2.17E-04		0.891	0.108	5.62E-04	0.962	0.0376	2.20E-04
	y_1	y_2	y_3	x_1	x_2	x_3		y_1	y_2	y_3	x_1	x_2	x_3
	0.439	6.97	39.9	0.430	7.50	27.9		0.360	6.08	39.9	0.429	7.48	27.9

Next, Figure 1 plots the extremal Laplace transforms $\hat{f}(s)$ and $1/\hat{g}(-s)$ for UB (LHS) and LB (RHS) for the case $c_a^2 = c_s^2 = 4$ and $\rho = 0.7$. The curves intersect at the decay rate θ_W. The decay rate for $R_a = R_s = 1$ is 0.106, while for $R_a = R_s = 20$ it is 0.098 for the UB and 0.110 for the LB.

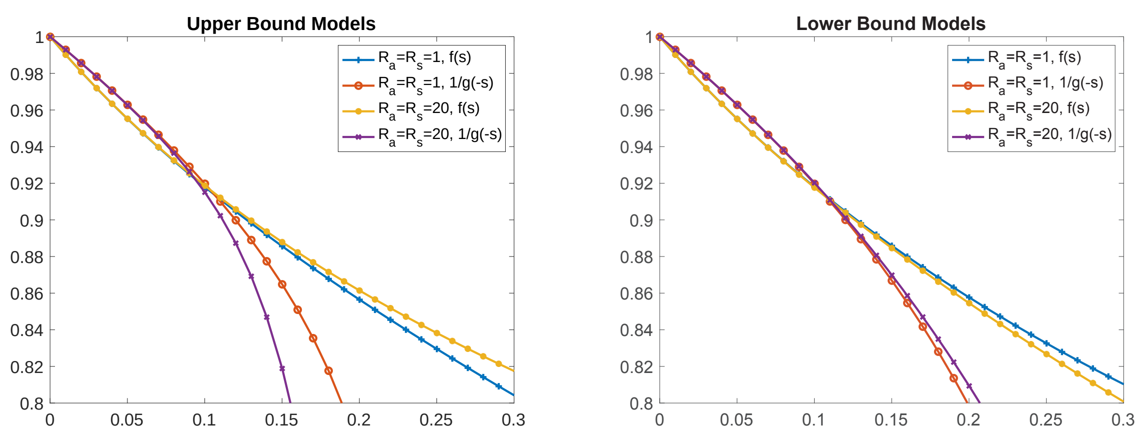

Figure 1 **Display of $\hat{f}(s)$ and $1/\hat{g}(-s)$ for UB (LHS) and LB (RHS) for the case $c_a^2 = c_s^2 = 4$ and $\rho = 0.7$: the decay rate for $R_a = R_s = 1$ is 0.106 and for $R_a = R_s = 20$ in UB is 0.098 and in LB is 0.110**

Next Tables 12 and 13 show the estimated extremal values of θ_W and $E[W]$ as a function of $R_a, R_s \in \{1, 5, 10, 20\}$ based on simulation for $\rho = 0.5, 0.9$ for this same case $(c_a^2, c_s^2) = (4, 4)$.

Table 12 **The improved LB and UB based on information of $H_2/H_2/1$ with $\rho = 0.5$**

$c_a^2 = c_s^2 = 4,\ \rho = 0.5,\ \theta_W = 0.2444, E[W] = 2.02$											
	$R_s \backslash R_a$	1	5	10	20		$R_s \backslash R_a$	1	5	10	20
θ_W (UB)	1	0.244	0.235	0.231	0.227	$\mathbb{E}[W]$ (UB)	1	1.97	2.32	2.48	2.60
	5	0.239	0.231	0.227	0.224		5	1.95	2.02	1.97	2.20
	10	0.239	0.230	0.227	0.224		10	1.96	2.03	1.98	2.20
	20	0.238	0.230	0.227	0.224		20	1.96	2.03	1.99	2.21
	$R_s \backslash R_a$	1	5	10	20		$R_s \backslash R_a$	1	5	10	20
θ_W (LB)	1	0.244	0.251	0.253	0.255	$\mathbb{E}[W]$ (LB)	1	2.09	1.87	1.81	1.75
	5	0.250	0.258	0.261	0.263		5	2.12	1.83	1.80	1.75
	10	0.251	0.258	0.261	0.263		10	2.12	1.84	1.80	1.75
	20	0.251	0.259	0.261	0.264		20	2.12	1.85	1.80	1.75

Table 13 **The improved LB and UB based on information of $H_2/H_2/1$ for $\rho = 0.9$**

$c_a^2 = c_s^2 = 4,\ \rho = 0.9,\ \theta_W = 0.0278, E[W] = 32.6$											
	$R_s \backslash R_a$	1	5	10	20		$R_s \backslash R_a$	1	5	10	20
θ_W (UB)	1	0.0278	0.0277	0.0277	0.0277	$\mathbb{E}[W]$ (UB)	1	32.9	31.4	31.7	32.1
	5	0.0278	0.0277	0.0277	0.0277		5	32.8	31.8	31.5	32.1
	10	0.0278	0.0277	0.0277	0.0277		10	32.8	31.7	31.6	32.2
	20	0.0278	0.0277	0.0277	0.0277		20	33.0	31.6	31.5	32.2
	$R_s \backslash R_a$	1	5	10	20		$R_s \backslash R_a$	1	5	10	20
θ_W (LB)	1	0.0278	0.0278	0.0278	0.0278	$\mathbb{E}[W]$ (LB)	1	32.8	32.6	32.7	33.2
	5	0.0278	0.0278	0.0278	0.0278		5	32.8	32.7	32.9	33.4
	10	0.0278	0.0278	0.0278	0.0278		10	32.7	32.8	32.7	33.4
	20	0.0278	0.0278	0.0278	0.0278		20	32.7	32.9	32.9	32.4

6.4. The $H_2/E_2/1$ Model with $c_a^2 = 4.0, c_s^2 = 0.5$

Next, Tables 14-16 show corresponding results for the case $c_a^2 = 4, c_s^2 = 0.5$, based on the first third moments and LT transform values from the model $H_2/E_2/1$, again using H_2 with balanced means. The exact values for the original $H_2/E_2/1$ model are given in Table 6. The exact values for $\rho = 0.7$ are $\mathbb{E}[W(H_2, E_2)] = 3.368$ and exact decay rate $\theta_W = 0.2602$.

Table 14 **The improved LB and UB based on information of $H_2/E_2/1$**

$c_a^2=4, c_s^2=0.5,\ \rho=0.7,\ \theta_w=0.2602$ and exact $\mathbb{E}[W(H_2,E_2)]=3.368$											
	$R_s\backslash R_a$	1	5	10	20	$R_s\backslash R_a$		1	5	10	20
θ_W (UB)	1	0.260	0.228	0.218	0.210	$\mathbb{E}[W]$ (UB)	1	3.438	3.920	4.091	4.265
	5	0.260	0.228	0.218	0.210		5	3.436	3.925	4.093	4.265
	10	0.260	0.228	0.217	0.210		10	3.436	3.993	4.194	4.335
	20	0.260	0.228	0.217	0.210		20	3.441	3.993	4.197	4.341
	$R_s\backslash R_a$	1	5	10	20		$R_s\backslash R_a$	1	5	10	20
θ_W (LB)	1	0.260	0.291	0.306	0.321	$\mathbb{E}[W]$ (LB)	1	3.343	2.951	2.782	2.638
	5	0.260	0.291	0.306	0.321		5	3.345	2.950	2.789	2.641
	10	0.260	0.292	0.307	0.321		10	3.345	2.965	2.815	2.692
	20	0.260	0.292	0.307	0.321		20	3.345	2.966	2.815	2.692

Table 15 **The improved LB and UB based on information of $H_2/E_2/1$ for $\rho=0.5$**

$c_a^2=4, c_s^2=0.5,\ \theta_w=0.8260, \rho=0.5$											
	$R_s\backslash R_a$	1	5	10	20		$R_s\backslash R_a$	1	5	10	20
	1	0.826	0.531	0.484	0.456	$\mathbb{E}[W]$ (UB)	1	0.93	1.53	1.72	1.84
	5	0.826	0.531	0.484	0.456		5	0.93	1.53	1.72	1.84
θ_W (UB)	10	0.826	0.531	0.484	0.456		10	0.93	1.54	1.72	1.84
	20	0.814	0.530	0.483	0.456		20	0.89	1.60	1.77	1.88
	$R_s\backslash R_a$	1	5	10	20		$R_s\backslash R_a$	1	5	10	20
	1	0.826	1.21	1.44	1.64	$\mathbb{E}[W]$ (LB)	1	0.860	0.498	0.381	0.305
	5	0.827	1.22	1.45	1.66		5	0.856	0.499	0.384	0.309
θ_W (LB)	10	0.827	1.22	1.45	1.66		10	0.854	0.495	0.386	0.307
	20	0.831	1.24	1.49	1.74		20	0.856	0.507	0.399	0.325

Table 16 The improved LB and UB based on information of $H_2/E_2/1$ for $\rho = 0.9$

$c_a^2 = 4, c_s^2 = 0.5,\ \theta_w = 0.0537, \rho = 0.9$											
	$R_s \backslash R_a$	1	5	10	20		$R_s \backslash R_a$	1	5	10	20
delay rate (UB)	1	0.0537	0.0534	0.0531	0.0529	EW (UB)	1	18.0	18.1	18.2	18.3
	5	0.0537	0.0534	0.0531	0.0529		5	18.1	18.2	18.3	18.4
	10	0.0537	0.0534	0.0531	0.0529		10	18.1	18.2	18.3	18.4
	20	0.0537	0.0534	0.0531	0.0529		20	18.1	18.2	18.3	18.4
	$R_s \backslash R_a$	1	5	10	20		$R_s \backslash R_a$	1	5	10	20
delay rate (LB)	1	0.0537	0.0539	0.0540	0.0541	EW (LB)	1	18.0	17.9	17.9	17.8
	5	0.0537	0.0539	0.0540	0.0541		5	18.0	17.9	17.9	17.8
	10	0.0537	0.0539	0.0540	0.0541		10	18.0	17.9	17.9	17.9
	20	0.0537	0.0539	0.0540	0.0541		20	18.0	17.9	17.9	17.9

Again, consistent with part (d) of Theorem 6 in the main paper, these tables show that the UB decay rate θ_W is monotone decreasing in R, while the LB decay rate is monotone increasing. Recall that we utilize information from the exact queueing models where F and G have unbounded support, so that we do not expect perfect consistency.

Next, Table 17 presents the extremal decay rates that go with the associated mean values $E[W]$ in Table 3 of the main paper. We obtain the rates here by solving the key equation (6) of the main paper for the original E_2 and H_2 distributions, so there is good numerical precision, but there is a minor difference from the truncated model, which explains the lack of precise order in a few cases.

Table 17 The decay rates for all basic models under $\rho = 0.7$

$E_2/H_2/1$	$R_s \backslash R_a$	5	10	20	$H_2/E_2/1$	$R_s \backslash R_a$	5	10	20
$\mu_s \le \theta_W \le \mu_a$	UB	0.150	0.150	0.149	$\mu_s \le \theta_W \le \mu_a$	UB	0.228	0.217	0.210
	LB	0.156	0.156	0.164		LB	0.291	0.307	0.321
	$R_s \backslash R_a$	5	10	20		$R_s \backslash R_a$	5	10	20
$\mu_s \ge \theta_W \ge \mu_a$	UB	0.151	0.150	0.143	$\mu_s \ge \theta_W \ge \mu_a$	UB	0.246	0.243	0.242
	LB	0.153	0.153	0.150		LB	0.283	0.286	0.288
	$R_s \backslash R_a$	5	10	20		$R_s \backslash R_a$	5	10	20
$\mu_s, \mu_a \le \theta_W$	UB	0.150	0.149	0.149	$\mu_s, \mu_a \le \theta_W$	UB	0.246	0.243	0.242
	LB	0.156	0.156	0.164		LB	0.283	0.287	0.289
	$R_s \backslash R_a$	5	10	20		$R_s \backslash R_a$	5	10	20
$\mu_s, \mu_a \ge \theta_W$	UB	0.151	0.150	0.143	$\mu_s, \mu_a \ge \theta_W$	UB	0.228	0.218	0.210
	LB	0.153	0.154	0.150		LB	0.291	0.306	0.320
$E_2/E_2/1$	$R_s \backslash R_a$	5	10	20	$H_2/H_2/1$	$R_s \backslash R_a$	5	10	20
$\mu_s \le \theta_W \le \mu_a$	UB	0.842	0.833	0.825	$\mu_s \le \theta_W \le \mu_a$	UB	0.105	0.104	0.098
	LB	0.880	0.889	0.893		LB	0.108	0.108	0.110
	$R_s \backslash R_a$	5	10	20		$R_s \backslash R_a$	5	10	20
$\mu_s \ge \theta_W \ge \mu_a$	UB	0.847	0.841	0.825	$\mu_s \ge \theta_W \ge \mu_a$	UB	0.106	0.105	0.103
	LB	0.861	0.859	0.842		LB	0.107	0.107	0.108
	$R_s \backslash R_a$	5	10	20		$R_s \backslash R_a$	5	10	20
$\mu_s, \mu_a \le \theta_W$	UB	0.849	0.848	0.848	$\mu_s, \mu_a \le \theta_W$	UB	0.106	0.105	0.100
	LB	0.866	0.867	0.867		LB	0.107	0.107	0.111
	$R_s \backslash R_a$	5	10	20		$R_s \backslash R_a$	5	10	20
$\mu_s, \mu_a \ge \theta_W$	UB	0.839	0.826	0.805	$\mu_s, \mu_a \ge \theta_W$	UB	0.105	0.104	0.101
	LB	0.874	0.880	0.863		LB	0.107	0.108	0.108

Remark 1. In general, we cannot claim that the bounds for θ_W yield bounds for $\mathbb{E}[W]$, so the connection is heuristic. From equation (3.205) in §II.5.11 of Cohen (1982), it follows that for the $K_2/GI/1$ model that $\mathbb{E}[W] = A + \theta_W^{-1}$, where A is a constant that depends on the parameters in (1) and F within K_2, but not otherwise on G. As a consequence, for fixed F, $\mathbb{E}[W]$ is a strictly decreasing function of θ_W for given first two moments.

6.5. The Possibility of Using Heavy-Traffic Approximations

Tables 18 and 19 show the improved LB and UB for the mean $E[W]$ starting with the exact decay rates of the base models and the approximation in (8) of the main paper. These show that we could also work with the HT approximations.

Table 18 **The improved LB and UB for $GI/GI/1$ Queues under Exact Decay Rates**

$\rho = 0.5$	$c_a^2 = c_s^2 = 0.5$			$\rho = 0.7$	$c_a^2 = c_s^2 = 0.5$			$\rho = 0.9$	$c_a^2 = c_s^2 = 0.5$		
	5	10	20		5	10	20		5	10	20
UB	0.200	0.204	0.222	UB	0.720	0.719	0.734	UB	3.91	3.92	3.89
LB	0.153	0.145	0.143	LB	0.649	0.625	0.642	LB	3.82	3.94	3.92
$\rho = 0.5$	$c_a^2 = c_s^2 = 4$			$\rho = 0.7$	$c_a^2 = c_s^2 = 4$			$\rho = 0.9$	$c_a^2 = c_s^2 = 4$		
	5	10	20		5	10	20		5	10	20
UB	2.02	1.98	2.21	UB	6.28	6.16	7.23	UB	31.8	31.6	32.2
LB	1.83	1.80	1.75	LB	6.49	6.45	6.42	LB	32.7	32.7	32.4
$\rho = 0.5$	$c_a^2 = 4, c_s^2 = 0.5$			$\rho = 0.7$	$c_a^2 = 4, c_s^2 = 0.5$			$\rho = 0.9$	$c_a^2 = 4, c_s^2 = 0.5$		
	5	10	20		5	10	20		5	10	20
UB	1.53	1.72	1.88	UB	3.92	4.19	4.34	UB	18.2	18.3	18.4
LB	0.499	0.386	0.325	LB	2.95	2.82	2.69	LB	17.9	17.9	17.9

Table 19 The improved LB and UB for $GI/GI/1$ Queues under Approximate Decay Rates

$\rho=0.5$	$c_a^2=c_s^2=0.5$			$\rho=0.7$	$c_a^2=c_s^2=0.5$			$\rho=0.9$	$c_a^2=c_s^2=0.5$		
	5	10	20		5	10	20		5	10	20
UB	0.220	0.238	0.247	UB	0.721	0.720	0.734	UB	3.87	3.93	3.92
LB	0.153	0.145	0.143	LB	0.649	0.625	0.642	LB	3.82	3.94	3.92
$\rho=0.5$	$c_a^2=c_s^2=4$			$\rho=0.7$	$c_a^2=c_s^2=4$			$\rho=0.9$	$c_a^2=c_s^2=4$		
	5	10	20		5	10	20		5	10	20
UB	2.02	1.98	2.19	UB	6.27	6.15	6.70	UB	31.8	33.0	33.1
LB	1.84	1.80	1.75	LB	6.48	6.45	6.42	LB	32.7	32.5	32.5
$\rho=0.5$	$c_a^2=4, c_s^2=0.5$			$\rho=0.7$	$c_a^2=4, c_s^2=0.5$			$\rho=0.9$	$c_a^2=4, c_s^2=0.5$		
	5	10	20		5	10	20		5	10	20
UB	1.35	1.62	1.78	UB	3.84	4.11	4.28	UB	18.2	18.3	18.4
LB	0.629	0.494	0.390	LB	3.01	2.88	2.74	LB	17.9	17.9	17.9

7. More on the $M/M/K$ Model

In this section we present results for $M/M/1$ and $M/M/2$ complementings Table 2 and 4 of the main paper. As above in this appendix, we use the parameter pair (R_a, R_s).

7.1. Results for $M/M/1$

We start by presenting results for the $M/M/1$ model that complement Table 2 of the main paper. First, Table 20 shows results for $M/M/1$ model using case (ii) of (30) in Theorem 6.

Table 20 **The Decay Rates and Set-valued Approximations of $M/M/1$ under Different μ_a, μ_s in case (ii) of (30) in Theorem 6**

$c_a^2 = c_s^2 = 1,\ \theta_w = 0.4286, \rho = 0.7,\ \mathbb{E}[W(M,M)] = 1.63,\ \mu_s \le \theta_W \le \mu_a$											
	$R_s \backslash R_a$	1	5	10	20		$R_s \backslash R_a$	1	5	10	20
	1	0.429	0.423	0.419	0.416		1	1.60	1.56	1.62	1.69
θ_W (UB)	5	0.427	0.421	0.418	0.415	$\mathbb{E}[W]$ (UB)	5	1.61	1.59	1.61	1.68
	10	0.427	0.421	0.418	0.415		10	1.61	1.61	1.62	1.68
	20	0.427	0.421	0.418	0.415		20	1.61	1.58	1.60	1.68
	$R_s \backslash R_a$	1	5	10	20		$R_s \backslash R_a$	1	5	10	20
	1	0.429	0.433	0.435	0.437		1	1.60	1.51	1.55	1.57
θ_W (LB)	5	0.430	0.434	0.437	0.438	$\mathbb{E}[W]$ (LB)	5	1.61	1.53	1.54	1.56
	10	0.430	0.434	0.437	0.439		10	1.61	1.53	1.56	1.56
	20	0.436	0.441	0.444	0.446		20	1.68	1.65	1.63	1.61

Next, Table 21 shows results for $M/M/1$ model using case (i) of (30) in Theorem 6 with $\mu_a, \mu_s \le \theta_W$.

Table 21 **The Decay Rates and Set-valued Approximations of $M/M/1$ under Different μ_a, μ_s in case (i) of (30) in Theorem 6**

$c_a^2 = c_s^2 = 1,\ \theta_w = 0.4286, \rho = 0.7, \mathbb{E}[W(M,M)] = 1.63$											
	$R_s \backslash R_a$	1	5	10	20		$R_s \backslash R_a$	1	5	10	20
	1	0.429	0.427	0.427	0.427		1	1.67	1.67	1.68	1.67
$\theta_W(F_L, G_U)$	5	0.427	0.426	0.425	0.425	UB	5	1.66	1.67	1.68	1.68
	10	0.427	0.426	0.425	0.425		10	1.66	1.67	1.68	1.68
	20	0.427	0.425	0.425	0.425		20	1.66	1.67	1.67	1.67
	$R_s \backslash R_a$	1	5	10	20		$R_s \backslash R_a$	1	5	10	20
	1	0.429	0.430	0.431	0.431		1	1.66	1.65	1.65	1.65
$\theta_W(F_U, G_L)$	5	0.430	0.432	0.432	0.432	LB	5	1.66	1.65	1.65	1.65
	10	0.430	0.432	0.432	0.432		10	1.66	1.65	1.65	1.65
	20	0.437	0.439	0.439	0.439		20	1.55	1.56	1.56	1.56

Table 22 shows results for $M/M/1$ model using case (iii) of (30) in Theorem 6 with $\mu_a, \mu_s \geq \theta_W$.

Table 22 **The Decay Rates and Set-valued Approximations of $M/M/1$ under Different μ_a, μ_s in case (iii) of (30) in Theorem 6**

$c_a^2 = c_s^2 = 1,\ \theta_w = 0.4286, \rho = 0.7, \mathbb{E}[W(M,M)] = 1.63$											
	$R_s \backslash R_a$	1	5	10	20		$R_s \backslash R_a$	1	5	10	20
$\theta_W(F_U, G_L)$	1	0.429	0.423	0.419	0.416	UB	1	1.67	1.71	1.72	1.73
	5	0.422	0.417	0.413	0.411		5	1.68	1.71	1.72	1.71
	10	0.422	0.417	0.413	0.411		10	1.68	1.71	1.72	1.71
	20	0.422	0.417	0.413	0.411		20	1.68	1.71	1.72	1.71
	$R_s \backslash R_a$	1	5	10	20		$R_s \backslash R_a$	1	5	10	20
$\theta_W(F_L, G_U)$	1	0.429	0.433	0.435	0.437	LB	1	1.67	1.65	1.63	1.62
	5	0.429	0.433	0.435	0.437		5	1.67	1.65	1.64	1.62
	10	0.429	0.433	0.435	0.437		10	1.67	1.65	1.64	1.62
	20	0.429	0.433	0.435	0.437		20	1.67	1.65	1.63	1.62

7.2. Corresponding Results for $M/M/2$

Tables 23 and 24 present corresponding results for the $M/M/2$ model in cases (ii) and (iii) of (30) in Theorem 6.

Table 23 **The Set-valued Approximations for $M/M/2$ in case (ii): $\mu_s \leq \theta_W \leq \mu_a$**

$c_a^2 = c_s^2 = 1,\ \theta_w = 0.4286, \rho = 0.7,\ \mathbb{E}[W(M,M)] = 1.35,\ \mu_s \leq \theta_W \leq \mu_a$											
	$R_s \backslash R_a$	1	5	10	20		$R_s \backslash R_a$	1	5	10	20
	1	0.429	0.424	0.420	0.417		1	1.31	1.34	1.40	1.42
θ_W (UB)	5	0.428	0.422	0.418	0.416	$\mathbb{E}[W]$ (UB)	5	1.30	1.34	1.39	1.42
	10	0.427	0.421	0.418	0.415		10	1.30	1.34	1.39	1.41
	20	0.427	0.421	0.418	0.415		20	1.30	1.34	1.39	1.41
	$R_s \backslash R_a$	1	5	10	20		$R_s \backslash R_a$	1	5	10	20
	1	0.426	0.430	0.432	0.434		1	1.33	1.32	1.33	1.34
θ_W (LB)	5	0.430	0.434	0.436	0.438	$\mathbb{E}[W]$ (LB)	5	1.34	1.30	1.31	1.33
	10	0.430	0.434	0.437	0.438		10	1.34	1.30	1.31	1.32
	20	0.430	0.434	0.437	0.439		20	1.34	1.30	1.31	1.33

Table 24 **The improved LB and UB based on information of $M/M/2$ $(\mu_a, \mu_s \geq \theta_W)$**

$c_a^2 = c_s^2 = 1,\ \theta_w = 0.4286, \rho = 0.7$											
	$R_s \backslash R_a$	1	5	10	20		$R_s \backslash R_a$	1	5	10	20
	1	0.426	0.420	0.417	0.414		1	1.40	1.40	1.39	1.37
$\theta_W(F_U, G_L)$	5	0.422	0.417	0.413	0.411	UB	5	1.40	1.41	1.38	1.34
	10	0.422	0.417	0.413	0.411		10	1.41	1.41	1.38	1.34
	20	0.422	0.417	0.413	0.411		20	1.40	1.40	1.38	1.34
	$R_s \backslash R_a$	1	5	10	20		$R_s \backslash R_a$	1	5	10	20
	1	0.429	0.432	0.434	0.437		1	1.35	1.32	1.27	1.25
$\theta_W(F_L, G_U)$	5	0.429	0.433	0.435	0.437	LB	5	1.35	1.32	1.29	1.26
	10	0.429	0.433	0.435	0.437		10	1.36	1.31	1.28	1.26
	20	0.429	0.433	0.435	0.437		20	1.36	1.31	1.29	1.26

Tables 23 and 24 show that the method for producing approximate intervals of likely values for the mean $E[W]$ remain effective for $K = 2$.

References

Chen Y, Whitt W (2018) Extremal $GI/GI/1$ queues given two moments, Columbia University, http://www.columbia.edu/∼ww2040/allpapers.html.

Chen Y, Whitt W (2019) Algorithms for the upper bound mean waiting time in the $GI/GI/1$ queue, Columbia University, http://www.columbia.edu/∼ww2040/allpapers.html.

Cohen JW (1982) *The Single Server Queue* (Amsterdam: North-Holland), second edition.

Daley DJ (1977) Inequalities for moments of tails of random variables, with queueing applications. *Zeitschrift fur Wahrscheinlichkeitsetheorie Verw. Gebiete* 41:139–143.

Daley DJ, Kreinin AY, Trengove C (1992) Inequalities concerning the waiting-time in single-server queues: a survey. Bhat UN, Basawa IV, eds., *Queueing and Related Models*, 177–223 (Clarendon Press).

Johnson MA, Taaffe MR (1991) An investigation of phase-distribution moment-matching algorithms for use in queueing models. *Queieing Systems* 8(1-2):129–148.

Johnson MA, Taaffe MR (1993) Tchebycheff systems for probability analysis. *American Journal of Mathematical and Management Sciences* 13(1-2):83–111.

Klincewicz JG, Whitt W (1984) On approximations for queues, II: Shape constraints. *AT&T Bell Laboratories Technical Journal* 63(1):139–161.

Ott TJ (1987) Simple inequalties for the $D/G/1$ queue. *Opperations Research* 35(4):589–597.

Stoyan D (1983) *Comparison Methods for Queues and Other Stochastic Models* (New York: John Wiley and Sons), translated and edited from 1977 German Edition by D. J. Daley.

Stoyan D, Stoyan H (1974) Inequalities for the mean waiting time in single-line queueing systems. *Engineering Cybernetics* 12(6):79–81.

Whitt W (1983a) Queue tests for renewal processes. *Oper. Res. Letters* 2(1):7–12.

Whitt W (1983b) The queueing network analyzer. *Bell Laboratories Technical Journal* 62(9):2779–2815.

Whitt W (1984a) On approximations for queues, I. *AT&T Bell Laboratories Technical Journal* 63(1):115–137.

Whitt W (1984b) On approximations for queues, III: Mixtures of exponential distributions. *AT&T Bell Laboratories Technical Journal* 63(1):163–175.