



**HAL**  
open science

# One-dimensional Service Networks and Batch Service Queues

Philippe Nain, Nitish K Panigrahy, Prithwish Basu, Don Towsley

► **To cite this version:**

Philippe Nain, Nitish K Panigrahy, Prithwish Basu, Don Towsley. One-dimensional Service Networks and Batch Service Queues. 2020. hal-03066179v1

**HAL Id: hal-03066179**

**<https://inria.hal.science/hal-03066179v1>**

Preprint submitted on 15 Dec 2020 (v1), last revised 22 Mar 2021 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# One-dimensional Service Networks and Batch Service Queues

Philippe Nain\*    Nitish K. Panigrahy†    Prithwish Basu‡    Don Towsley§

December 15, 2020

## Abstract

A proliferation of smart devices, computational and storage resources is predicted to continue aggressively in the near future. Such “networked” devices and resources which are distributed in a physical space and provide services are collectively referred to as a *distributed service network*. Assigning users or applications to available resources is important to sustain high performance of the distributed service network. In this work, we consider a one-dimensional service network where both users and resources are located on a line, and analyze a unidirectional assignment policy *Move To Right* (MTR), which sequentially assigns users to resources available to their right. We express the communication cost for a user-resource assignment as an increasing function of the *distance traveled by the user request (request distance)* and analyze the expected communication cost for the service network when locations of users and resources are modeled by different spatial point processes. We use results from literature that map the request distance of an assigned user in a one-dimensional service network to the sojourn time of a customer in an exceptional service accessible batch queueing system. We compute the Laplace-Stieltjes transform of the sojourn time distribution for this queueing system for Poisson distributed users with general inter-resource distance distributions and in the process also generate new results for batch service queues. Unlike previous work [22] our framework not only captures the first-order moment of the request distance, but also the *request distance distribution* itself, thus allowing us to compute the expected communication cost under different cost models.

## 1 Introduction

The distributed computing paradigm endeavors to connect a large number of end-users and resources or services distributed across a physical space. We collectively call such a set of

---

\*Inria, France, philippe.nain@inria.fr

†University of Massachusetts, Amherst, MA, USA, nitish@cs.umass.edu

‡Raytheon BBN Technologies, Cambridge, MA, USA, prithwish.basu@raytheon.com

§University of Massachusetts, Amherst, MA, USA, towsley@cs.umass.edu

resources/services a *distributed service network*. In many such distributed networks, for example: *internet of things* (IoT), both the number of end-users and resources including the machines, smart sensors, computational and storage units continue to grow steadily. For instance, International Data Corporation (IDC) estimates 41.6 billion connected IoT resources, or “things”, by the year 2025 [23]. To ensure optimal use of resources and guarantee quality of service to end users at such a large scale, efficient allocation of resources to users is essential.

Optimal resource allocation in a distributed service network has been widely studied. For example, consider the problem of assigning users to resources, all located in a  $d$ -dimensional euclidean space. When  $d = 2$  or  $3$  and resources have unit service capacity, i.e., a resource can support at most one concurrent user request, the optimal allocation policy that minimizes the expected *distance traveled by a user request* (request distance) can be modeled as an Euclidean bipartite matching problem [2],[17], [24]. The problem with general service capacity constraints can be solved by running the *network simplex algorithm* [20] on the flow graph associated with the corresponding bipartite graph [3] of the user–service network. In the case when the network has a very low-dimensional geometric structure, e.g.,  $d = 1$ , more efficient algorithms have been proposed for the user-to-resource assignment problem [22].

Although assigning users to resources in higher dimensional networks is generally important, many networks exhibit lower dimensional structure. Examples of a one-dimensional structure include ride-hailing on a one-way street, trucks on a highway, boats on a river, deployment of wireless sensors for intrusion detection [19]. Assignment policies in such a one-dimensional network can primarily be divided into two categories based on directional restrictions [22],[1]. The class of policies that allocate users only to resources to their right (or left) come under the “unidirectional” category whereas policies without any directional restriction in allocation belong to “bidirectional”. Abadi et al. [1] first proposed a unidirectional policy known as Unidirectional Gale-Shapley (or *UGS*) that minimizes expected request distance. Panigrahy et al. [22] analyzed another unidirectional policy “Move to Right” (or *MTR*) that sequentially allocates users to available resources located to their right and showed its equivalence to UGS with respect to (wrt) expected request distance. Both [1] and [22] exploited the likeness of unidirectional space to time, and mapped a one-dimensional service network to a classical queueing system when locations of users and resources are modeled by statistical point processes. They mapped the request distance of an assigned user in the service network to the *sojourn time* of a customer in the corresponding queueing system, thereby computing closed form expressions for the expected request distance using queueing theoretic tools.

While previous works mainly focus on obtaining expressions for expected absolute request distance, little attention has been given to evaluating the expected *communication cost* associated with the allocation policies. In many service networks, such as IoT, the users are battery powered with limited operation time and connected wirelessly to the resources. In such service networks, the signal attenuation on a wireless link connecting a user to its allocated resource typically varies as an increasing function of corresponding request distance [9]. Therefore the transmission power needed to transmit data over a certain distance varies as a

positive power of the latter, with the exponent typically being greater than 2. Thus the cost to communicate data over a wireless link can be expressed as a function of the transmission power at the user needed for successful reception at its assigned resource node, which is an increasing function of the corresponding request distance. In other service networks such as transportation ride hailing services, the cost can be interpreted as the loss of utility for the user. As the request distance increases, it takes longer for a vehicle to arrive at the user, and this latency is typically super-linear in the distance due to road congestion. Moreover, the degree of a user's impatience may itself be a nonlinear increasing function of the latency. Taking such non-trivial cost models into account while making allocation decisions can lead to a larger number of satisfied users than the baseline scenario of using just the average request distance as the performance metric.

One major implication of considering communication cost can be observed for MTR and UGS policies. While both policies are optimal<sup>1</sup> wrt expected request distance, UGS has a higher variance for request distance as compared to MTR for certain inter-user and inter-resource distance distributions [22]. Thus a more complete characterization of assignment policies wrt communication cost is vital for better performance of the service network.

In this paper, we compute the expected communication cost associated with MTR in a one-dimensional service network. We assume users to be distributed according to a Poisson process<sup>2</sup>. This one-dimensional network setting applies to various service networks. For example, consider vehicular wireless ad-hoc networks on a one-lane roadway [13, 18], where users are in vehicles and resources are attached to fixed infrastructure such as lamp posts. We use results from the literature that map the request distance of an assigned user in a one-dimensional service network to the sojourn time of a customer in an *exceptional service accessible batch queueing* system. We compute the Laplace-Stieltjes transform of the sojourn time distribution for this queueing system and in the process also generate new results for batch service queues. The work closest to the problem we tackle here is [22]; we rely and expand upon this work in this paper. For example, expressions for the expected request distance were obtained in [22] and the results obtained in this paper are complementary to it. Thus, as far as possible, in this paper, we adopt the notation used in [22].

Our major contributions can be summarized as follows:

1. We consider a simple unidirectional allocation policy MTR for Poisson distributed users with general inter-resource distance distribution and compute the Laplace-Stieltjes transform (LST) of the request distance distribution for the network;
2. Using the LST of the request distance distribution, we obtain expressions for the expected communication cost for the network. Further simplified expressions for the expected communication cost is obtained for a path loss based cost model;

---

<sup>1</sup>Both MTR and UGS are optimal wrt expected request distance among all unidirectional allocation policies.

<sup>2</sup>[13] confirms that vehicle location distribution on the streets in Central London can be closely approximated by a Poisson distribution.

3. While computing the request distance distribution, we also generate new results for the batch service queueing systems.

The paper is organized as follows. The next section contains some technical preliminaries. We briefly overview the mapping of a service network to an exceptional service accessible batch queue in Section 3. In Section 4, we derive for the LS transform of the sojourn time in an M/G/1 queue with accessible batches and exceptional first batch in a busy period; this quantity corresponds to the request distance distribution in the 1-D service network. Section 4 is self-contained and can be read independently of the other sections. We compute the expected communication cost and discuss a path loss based cost model in Section 5. We provide some numerical results in Section 6. Section 7 discusses related work. We conclude the paper in Section 8.

## 2 Technical Preliminaries

Let  $R$  and  $S$  denote set of users and servers respectively. Define capacity of a server to be the maximum number of requests that it can process. Denote  $\mathbb{Z}^+ = \{0, 1, 2, \dots\}$  and  $\mathbb{R}^+ = [0, \infty)$ . We assume that each server  $j \in S$  has equal capacity  $c_j = c \in \mathbb{Z}^+$ . Users and servers are located on a line  $\mathbb{R}^+$ . More precisely,  $r : R \rightarrow \mathbb{R}^+$  and  $s : S \rightarrow \mathbb{R}^+$  are the location functions for users and servers, respectively, such that a distance  $d(r, s)$  is well defined for all pairs  $(r, s) \in R \times S$ .

### 2.1 User and Server Spatial Distributions

Let  $\{r_i\}_{i \geq 1}$  and  $\{s_i\}_{i \geq 1}$  represent user and server locations with  $0 \leq r_1 \leq r_2 \leq \dots$  and  $0 \leq s_1 \leq s_2 \leq \dots$  respectively. Denote  $X_j = s_j - s_{j-1}, j \geq 1, s_0 = 0$ , the inter-server distances and  $Y_i = r_i - r_{i-1}, i \geq 1, r_0 = 0$ , the inter-user distances. We assume  $\{X_j\}_{j \geq 1}$  to be a renewal process with cumulative distribution function (cdf)

$$\mathbb{P}(X_j \leq x) = F_X(x). \quad (1)$$

We also assume  $\{Y_i\}_{i \geq 1}$  to be exponentially distributed with cdf  $F_Y(x)$ , i.e.,

$$\mathbb{P}(Y_i \leq x) = F_Y(x) = 1 - e^{-\lambda x}. \quad (2)$$

We denote  $\alpha_X = 1/\mu$  and  $\sigma_X^2$  to be the mean and variance associated with  $F_X$ . Similarly let  $\alpha_Y = 1/\lambda$  and  $\sigma_Y^2$  be the mean and variance associated with  $F_Y$ . We let  $\rho = \lambda/\mu$  and assume that  $\rho < c$ . Denote by  $F_X^*(s) = \int_0^\infty e^{-sx} dF_X(x)$  and  $F_Y^*(s)$  the Laplace-Stieltjes transform (LST) of  $F_X$  and  $F_Y$  with  $\Re(s) \geq 0$ . We consider different inter-server distance distributions, including exponential and deterministic.

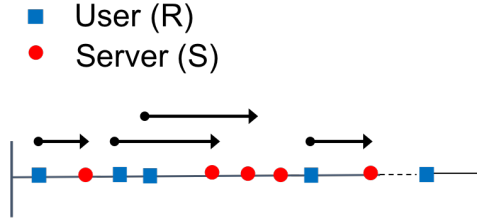


Figure 1: Allocation of users to servers under MTR policy [22].

## 2.2 User Allocation Policy

We evaluate the expected communication cost associated with a unidirectional policy *Move To Right* (MTR) [22]. In MTR, starting from the left, each user is allocated sequentially to the nearest available server to its right as shown in Figure 1.

## 3 Mapping to Queueing Theory

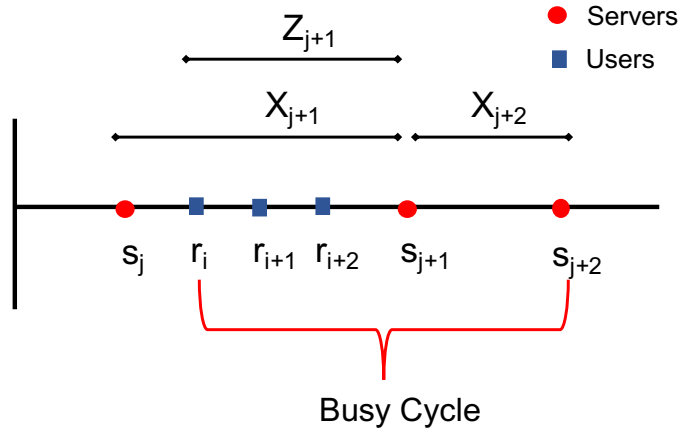


Figure 2: Mapping of service network to an exceptional service accessible batch queueing system with  $c = 2$  [22].

We briefly overview the notion of exceptional service and accessible batches queue (ESABQ) as applied to the one-dimensional service network below, referring the interested reader to [22] for further details.

### 3.1 ESABQ and One-dimensional Service Network

Under MTR, the service network can be modeled as batch service queue as shown in Figure 2. The distance between two consecutive users in the service network can be thought of as

inter-arrival time between customers in the batch service queue. The distance between two consecutive servers maps to a batch service time. Also, customers can join an existing service if there is room at the server which is an example of accessible batch ( $\{r_i, r_{i+1}\}$ ) as shown in Figure 2. Note that, the service time for first customer (user  $r_i$ ) that initiates a busy cycle, is described by the random variable  $Z_{j+1}$  which is different or *exceptional* compared to service times of successive batches. Both users  $r_i$  and  $r_{i+1}$  will leave the system at time  $s_{j+1}$ ; user  $r_{i+2}$  will leave the system at time  $s_{j+2}$  and the system will be empty after her departure.

Because the process  $\{r_i\}_i$  is Poisson, the sequence  $\{Z_j\}_j$  is a renewal sequence with cdf  $F_Z(x)$ . It is shown in [22, Eq. (11)] that

$$F_Z(x) = \frac{\lambda e^{\lambda x} [B_X(\infty) - B_X(x)] - F_X^*(\lambda)}{1 - F_X^*(\lambda)}, \quad (3)$$

with  $B_X(x) := \int_0^x F_X(u) e^{-\lambda u} du$ . Note that  $B_X(\infty) = \lambda^{-1} F_X^*(\lambda)$ . An explicit form for  $F_Z(x)$  is obtained in [22, Section 13.3] when  $F_X(x)$  is the cdf of an exponential, uniform, or deterministic random variable. In particular,  $F_Z(x) = 1 - e^{-\mu x}$  when  $F_X(x) = 1 - e^{-\mu x}$ , a result that also follows from the PASTA property. It is shown in Appendix 10.1 that

$$\alpha_Z := \mathbb{E}[Z] = \frac{1}{\mu(1 - F_X^*(\lambda))} - \frac{1}{\lambda}. \quad (4)$$

In particular,  $\alpha_Z = \alpha_X = \frac{1}{\mu}$  when  $F_X(x) = 1 - e^{-\mu x}$ .

Having established an analogy, ESABQ can now be formally defined in its most general form as follows.

**ESABQ:** *Consider a queueing system where customers are served in batches of maximum size  $c$ . A customer entering the queue and finding fewer than  $c$  customers in the system joins the current batch and enters service at once, otherwise it joins a queue. After a batch departs leaving  $k$  customers in the buffer,  $\min(c, k)$  customers form a batch and enter service immediately. There are two different service times cdfs,  $G_1(x)$  (exceptional batch) and  $G_2(x)$  (regular batch). A batch is exceptional if its oldest customer entered an empty system, otherwise it is a regular batch. When the service time expires, all customers in the server depart at once, regardless of the nature of the batch (exceptional or regular).*

The sojourn time associated with a random customer in the M/G/1 queue described above maps to the request distance for a random user in the corresponding one-dimensional service network upon setting  $G_1 = F_Z$  and  $G_2 = F_X$ . Thus the request distance distribution for a service network can be obtained by calculating the sojourn time distribution for the corresponding M/G/1 queue.

The LST of the customer sojourn time in the above queueing system is derived in Section 4. Section 4 is self-contained and can be read independently from Sections 1-3.

## 4 M/G/1 Queue with Accessible Batches and Exceptional First Batch in a BP

Consider an M/G/1 queue with arrival rate  $\lambda > 0$  where customers are served in *accessible* (see below) batches of maximum size  $c$ , with  $c$  a strictly positive integer. The cdf of the service time of the first batch in a busy period (BP) is  $G_1$  (with mean  $1/\mu_1$ ,  $\mu_1 > 0$ , pdf  $g_1$ , and LST is  $G_1^*$ ) and the cdf of the subsequent batch service times in a BP is  $G_2$  (with mean  $1/\mu_2$ ,  $\mu_2 > 0$ , pdf  $g_2$ , and LST  $G_2^*$ ). The first batch in a BP is called an *exceptional batch* and starts with the arrival of a customer in an empty system. The other batches in a BP are called *regular or batches*. Both exceptional and regular batches are accessible in the sense that if a customer enters the system when there are less than  $c$  customers in the system then this customer joins the server. When the service time of either type of batch expires all customers in the server leave the system at once.

See Section 7 for related works in this area. The work the closest to this work is [22] where the  $z$ -transform for the number of customers is derived for the model described above. In this section we focus on the sojourn time of an arbitrary incoming customer.

Define  $\rho_1 = \lambda/\mu_1$  and  $\rho_2 = \lambda/\mu_2$ . We claim that  $\rho_2 < c$  is the stability condition of this queueing system. Indeed, it is known that  $\rho_2 < c$  is the stability condition of an M/G/1 queue in which customers are served in batches of maximum size  $c$  [6, Section 4.1]. This result was shown for non-accessible batches but clearly the stability condition does not depend on whether or not a batch is accessible. The difference between the latter queueing system and ours is that in ours the service time cdf of the first batch served in a busy period (given by  $G_1$ ) is different from the service time cdf (given by  $G_2$ ) of the subsequent service times in that busy period. However, as far as the stability is concerned, the cdf of the first service duration in a busy period does not matter as long as the expected service time (given by  $1/\mu_1$ ) is finite, thereby validating why  $\rho_2 < c$  is the stability condition of our system.

From now on we assume that the stability condition  $\rho_2 < c$  holds.

We will need  $p(0)$ , the stationary probability that the system is empty. This quantity was computed in [22, Section 6.3.1] when  $z^c - G_2^*(\lambda(1-z))$  has a radius of convergence larger than one (i.e. is analytic in  $\{|z| \leq \nu$  with  $\nu > 1$ ), and is given by

$$p(0) = \frac{1}{\lambda} \sum_{k=1}^c a_k, \quad (5)$$

where the  $c$  constants  $a_1, \dots, a_c$  are the solutions of the linear system of equations

$$\sum_{k=1}^c a_k \frac{d^j}{dz^j} \left( G_1^*(\lambda(1-z)) - z^{k-c-1} G_2^*(\lambda(1-z)) \right) \Big|_{z=\xi_i} = 0, \quad 0 \leq j \leq n_l - 1, 1 \leq l \leq q - 1,$$

$$\sum_{k=1}^c a_k (c(1 + \rho_1) - \rho k) = \lambda(c - \rho),$$



with  $\xi_1, \dots, \xi_c$  the  $c$  zeros of  $z^c - G_2^*(\lambda(1-z))$  in  $\{|z| \leq 1\}$  counting their multiplicity (i.e. if  $\xi_1$  is a zero of multiplicity  $n_1 \geq 1$  then  $\xi_1 = \dots = \xi_{n_1}$ , etc.). The condition that  $z^c - G_2^*(\lambda(1-z))$  has a radius of convergence larger than one is satisfied for all phase-type distributions (exponential, Erlang, Coxian, etc.), Gamma distribution, etc. [22, Section 13.2.3]. In particular, all zeros of  $z^c - G_2^*(\lambda(1-z))$  in  $\{|z| \leq 1\}$  have multiplicity one when  $G_2(x) = 1 - e^{-\mu_2 x}$  – see [22, 13.2.2] for a more general discussion.

Let  $T$  be the sojourn time of the tagged customer  $C$ , with LST  $T^*(s)$ . Define

$$d_l(m) = \int_0^\infty \frac{(\lambda u)^m}{m!} e^{-\lambda u} dG_l(u) \quad (6)$$

$$h_l(m, z) = \int_0^\infty e^{-\lambda u} \frac{(\lambda z u)^m}{m!} dG_l(u), \quad (7)$$

for  $l = 1, 2$  and  $m \geq 0$ . Note that  $d_l(m) = h_l(m, 1)$  and  $d_l(0) = h_l(0, z) = G_l^*(\lambda)$  for  $l = 1, 2$ .

We are now ready to state and prove the main result of this section.

**Proposition 1 (LST of customer sojourn time distribution)**

Let  $\rho_2 < c$ . We assume that  $z^c - G_2^*(\lambda(1-z))$  has  $c$  zeros in  $\{|z| \leq 1\}$  (see discussion above the proposition). Let  $\xi_1, \dots, \xi_q$  be the distinct zeros of  $z^c - G_2^*(\lambda(1-z))$  in  $\{|z| \leq 1\}$ , with  $n_1, \dots, n_q$  their respective multiplicity with  $n_1 + \dots + n_q = c$ . Note that  $z = 1$  is a zero of multiplicity one under the condition  $\rho_2 < c$ . Let  $\xi_q = 1$  by convention (with  $n_q = 1$ ).

Then,

$$\begin{aligned} T^*(s) &= p(0)G_1^*(s) + p(0)\frac{\lambda}{\lambda-s} \sum_{k \geq 0} \left(\frac{\lambda}{\lambda-s}\right)^k G_2^*(s)^{\lfloor \frac{k+1}{c} \rfloor} I_1(s, k) \\ &\quad + (1-p(0)(1+\rho_1))\frac{\mu_2}{\lambda-s} \sum_{k, m \geq 0} \left(\frac{\lambda}{\lambda-s}\right)^k G_2^*(s)^{\lfloor \frac{k+m}{c} \rfloor} \mathbb{P}(N_r = m) I_2(s, k), \quad \Re(s) \geq 0 \end{aligned} \quad (8)$$

where  $p(0)$  is given in (5), and

$$I_l(s, k) := G_l^*(s) - \sum_{j=0}^k \int_0^\infty e^{-\lambda x} g_l(x) \frac{((\lambda-s)x)^j}{j!} dx, \quad l = 1, 2. \quad (9)$$

In (8),  $N_r$  is the stationary number of customers in the system (i.e. including customer(s) in the server) at the start of a service of a regular batch. Its  $z$ -transform  $N_r(z) = \sum_{m \geq 0} \pi_r(m) z^m$ , with  $\pi_r(m) := \mathbb{P}(N_r = m)$ , is given by

$$N_r(z) = \frac{J(z)}{G_2^*(\lambda(1-z)) - z^c}, \quad |z| \leq 1, \quad (10)$$

with

$$J(z) := \sum_{i=1}^c \pi_r(i) \left( z^i \sum_{m=0}^{c-i} h_2(m, z) - z \left( G_1^*(\lambda(1-z)) - \sum_{m=0}^{c-1} h_1(m, z) \right) \frac{\sum_{m=0}^{c-i} d_2(m)}{1 - \sum_{m=0}^{c-1} d_1(m)} \right). \quad (11)$$

The  $c$  unknown constants  $\pi_r(1), \dots, \pi_r(c)$  in (11) solve the system of  $c$  linear equations defined by

$$\frac{d^k}{dz^k} J(z)|_{z=\xi_j} = 0, \quad k = 0, \dots, n_j - 1, \quad j = 1, \dots, q \quad (12)$$

$$\frac{d}{dz} J(z)|_{z=1} = \rho_2 - c. \quad (13)$$

The proof of Proposition 1 is given in Sections 4.1-4.2. It is inspired by Harrison's calculation of the LST of the customer sojourn time in a standard M/G/1 queue [11].

When  $c = 1$ ,  $T^*(s)$  in (8) should reduce to the LST of the customer sojourn time in an M/G/1 queue with exceptional service time, a result first derived by Welch [25]. In addition, if  $G_1^* = G_2^*$  then  $T^*(s)$  in (8) should be equal to the LST of the customer sojourn time in a standard M/G/1 queue (e.g. see [11]). Corollaries 1 and 2 below show that this is indeed the case.

**Corollary 1 (M/G/1 queue with exceptional first service)**

When  $\rho_2 < 1$  and  $c = 1$ ,

$$T^*(s) = \frac{(1 - \rho_2)(sG_1^*(s) + \lambda(G_2^*(s) - G_1^*(s)))}{(1 - \rho_2 + \rho_1)(s - \lambda(1 - G_2^*(s)))}, \quad \Re(s) \geq 0. \quad (14)$$

We retrieve the result in [25, Theorem 4]. The proof can be found in Appendix 10.2.

**Corollary 2 (Standard M/G/1 queue)**

When  $G_1^* = G_2^* := G^*$ , with  $\rho_1 = \rho_2 := \rho < 1$ , (14) becomes

$$T^*(s) = \frac{(1 - \rho)sG^*(s)}{s - \lambda(1 - G^*(s))}, \quad \Re(s) \geq 0.$$

We retrieve the LST of the customer sojourn time in a standard M/G/1 queue with arrival rate  $\lambda$ , LST of the service times given by  $G^*$ , and traffic intensity given by  $\rho$ .

We now proceed with the proof of Proposition 1.

#### 4.1 Proof of Proposition 1: Proof of (8)

We focus on a tagged customer  $C$  arriving at time  $\tau_C$  when the system is in steady-state. When  $C$  joins a non-empty system, we denote by

- $U$  the completed service time
- $V$  the residual service time
- $N_r$  the number of customers in the system (i.e. including customers in the server) at the beginnings of a service time of an regular customer
- $Y_e$  the number of arrivals during  $U$  if this is an exceptional service time
- $Y_r$  the number of arrivals during  $U$  if this is an regular service time.

Let  $X(t)$ ,  $R(t)$  and  $I(t)$  be the number of customers in the system including the one(s) in the server if any, the residual batch service time, and the type of batch at time  $t \geq 0$ . The assumptions placed on the system imply that  $\{(X(t), R(t), I(t)), t \geq 0\}$  is a Markov process on  $\{0, 1, \dots, \} \times [0, \infty) \times \{1, 2\}$ . Denote by  $X$ ,  $R$ , and  $I$  the stationary versions of  $X(t)$ ,  $R(t)$ , and  $I(t)$ , respectively, which do exist when  $\rho_2 < c$ . Also define  $p(n, x; i) = \mathbb{P}(X = n, R < x, I = i)$ ,  $n \geq 1, x > 0, i = 1, 2$ . Recall that (cf. (5))  $p(0) = \mathbb{P}(X = 0)$

Because arrivals are Poisson, PASTA applies and upon arrival the tagged customer  $C$  sees the system in steady-state. Therefore, the probability that upon arrival  $C$  sees an empty system is given by  $p(0) = \mathbb{P}(X = 0)$  and the probability that  $C$  sees a non-empty system and a batch of type  $i = 1, 2$  in the system is  $\mathbb{P}(X > 0, I = i)$ . Conditioning on these three types of events gives

$$\begin{aligned} T^*(s) = & p(0)G_1^*(s) + \mathbb{E}[e^{-sT} | X > 0, I = 1]\mathbb{P}(X > 0, I = 1) \\ & + \mathbb{E}[e^{-sT} | X > 0, I = 2]\mathbb{P}(X > 0, I = 2), \quad \Re(s) \geq 0. \end{aligned} \quad (15)$$

##### 4.1.1 Calculation of $q_i := \mathbb{P}(X > 0, I = i)$ , $i = 1, 2$

We can evaluate  $q_i$  as

$$q_i = \mathbb{P}(X > 0, I = i) = \sum_{n \geq 1} \mathbb{P}(X = n, R < \infty, I = i) = \sum_{n \geq 1} p(n, \infty; i),$$

$$i = 1, 2.$$

Introduce  $F_i(z, s) := \sum_{n \geq 1} z^n \int_0^\infty e^{-sx} p(n, x; i) dx$  for  $|z| \leq 1$  and  $\Re(s) \geq 0$ . Then,

$$q_i = \lim_{s \rightarrow 0} sF_i(1, s), \quad i = 1, 2, \quad (16)$$

from a standard Tauberian result. We now derive an equation for  $F_1(z, s)$ . Writing the Chapman-Kolmogorov equations for the Markov process  $\{(X(t), R(t), I(t)), t \geq 0\}$  and letting

$t \rightarrow \infty$  one can show that (see e.g. [15, Exercice 5.7, pp. 233-234] and [22, Section 6.3.1] for similar derivations)

$$0 = \frac{\partial}{\partial x} p(n, x; 1) - \lambda p(n, x; 1) - \frac{\partial}{\partial x} p(n, 0; 1) + \lambda p(n-1, x; 1) \mathbf{1}(n \geq 2) + \lambda p(0) G_1(x) \mathbf{1}(n=1). \quad (17)$$

Multiplying both sides of (17) by  $z^n e^{-sx}$ , integrating over  $[0, \infty)$ , and summing over  $n \geq 1$  gives

$$\begin{aligned} 0 &= \sum_{n \geq 1} z^n \int_0^\infty e^{-sx} \frac{\partial}{\partial x} p(n, x; 1) dx - \lambda F_1(z, s) - \sum_{n \geq 1} z^n \int_0^\infty e^{-sx} \frac{\partial}{\partial x} p(n, 0; 1) \\ &\quad + \lambda z p(0) \int_0^\infty e^{-sx} G_1(x) dx + \lambda z \sum_{n \geq 2} z^{n-1} \int_0^\infty e^{-sx} p(n-1, x; 1) \\ &= s F_1(z, s) - \lambda F_1(z, s) - \sum_{n \geq 1} z^n \int_0^\infty e^{-sx} \frac{\partial}{\partial x} p(n, 0; 1) + \lambda z F_1(z, s) + \lambda z p(0) \frac{G_1^*(s)}{s} \\ &= (s - \lambda(1-z)) F_1(z, s) - \frac{1}{s} \sum_{n \geq 1} z^n \frac{\partial}{\partial x} p(n, 0; 1) + \lambda z p(0) \frac{G_1^*(s)}{s}. \end{aligned}$$

Hence,

$$s(\lambda(1-z) - s) F_1(z, s) = - \sum_{n \geq 1} z^n \frac{\partial}{\partial x} p(n, 0; 1) + \lambda z p(0) G_1^*(s).$$

The l.h.s vanishes for  $s = \lambda(1-z)$ , which implies that

$$\sum_{n \geq 1} z^n \frac{\partial}{\partial x} p(n, 0; 1) = \lambda z p(0) G_1^*(\lambda(1-z)).$$

Therefore,

$$s(\lambda(1-z) - s) F_1(z, s) = \lambda z p(0) (G_1^*(s) - G_1^*(\lambda(1-z))). \quad (18)$$

Combining (16) and (18), we obtain

$$q_1 = \lambda p(0) \lim_{s \rightarrow 0} \frac{1 - G_1^*(s)}{s} = \lambda p(0) \frac{1}{\mu_1} = p(0) \rho_1, \quad (19)$$

where we have used L'Hôpital's rule to derive the second equality.

The probability  $q_2$  is then obtained from the identity  $p(0) + q_1 + q_2 = 1$ , giving

$$q_2 = 1 - p(0)(1 + \rho_1). \quad (20)$$

#### 4.1.2 Calculation of $\mathbb{E}[e^{-sT} | X > 0, I = 1]$

Let  $\{\sigma_1^{(j)}, j \geq 1\}$  be iid rvs with cdf  $G_1$ , pdf  $g_1$ , and LST  $G_1^*$ . We have

$$\begin{aligned} \mathbb{E}[e^{-sT} | X > 0, I = 1, U, V] &= \sum_{k \geq 0} \mathbb{E}[e^{-sT} | X > 0, I = 1, U, V, Y_e = k] \mathbb{P}(Y_e = k | U) \\ &= e^{-\lambda U} \sum_{k \geq 0} \mathbb{E}[e^{-sT} | X > 0, I = 1, U, V, Y_e = k] \frac{(\lambda U)^k}{k!}. \end{aligned}$$

Conditioned on  $I = 1, U, V$  and  $Y_e = k$ , the sojourn time of the tagged customer is  $V + \sum_{j=1}^i \sigma_1^{(j)}$  when  $ic \leq Y_e + 1 < (i+1)c$  for  $i \geq 0$ , yielding

$$\begin{aligned} \mathbb{E}[e^{-sT} | X > 0, I = 1, U, V, Y_e = k] &= \begin{cases} \mathbb{E}[e^{-sV}] & \text{if } k+1 < c \\ \mathbb{E}[e^{-sV}] G_2^*(s)^i & \text{if } ic \leq k+1 < (i+1)c, i \geq 1. \end{cases} \\ &= \mathbb{E}[e^{-sV}] G_2^*(s)^{\lfloor \frac{k+1}{c} \rfloor}. \end{aligned} \quad (21)$$

Therefore,

$$\mathbb{E}[e^{-sT} | X > 0, I = 1, U, V] = e^{-\lambda U} \mathbb{E}[e^{-sV}] \sum_{k \geq 0} G_2^*(s)^{\lfloor \frac{k+1}{c} \rfloor} \frac{(\lambda U)^k}{k!}.$$

Given that  $I = 1$ , the joint density of  $(U, V)$  at point  $(u, v)$  is  $\mu_1 g_1(u+v)$  (well-known result - e.g. see [12, p. 153]). With this, putting everything together gives

$$\begin{aligned} \mathbb{E}[e^{-sT} | X > 0, I = 1] &= \mu_1 \int_0^\infty \int_0^\infty \mathbb{E}[e^{-sT} | X > 0, I = 1, U = u, V = v] g_1(u+v) dudv \\ &= \mu_1 \sum_{k \geq 0} G_2^*(s)^{\lfloor \frac{k+1}{c} \rfloor} \int_0^\infty \int_0^\infty e^{-\lambda u} e^{-sv} g_1(u+v) \frac{(\lambda u)^k}{k!} dudv \\ &= \mu_1 \sum_{k \geq 0} G_2^*(s)^{\lfloor \frac{k+1}{c} \rfloor} \int_{x=0}^\infty e^{-sx} g_1(x) \int_{u=0}^x e^{-(\lambda-s)u} \frac{(\lambda u)^k}{k!} dudx. \end{aligned}$$

Elementary algebra yields

$$(\lambda - s) \int_0^x e^{-(\lambda-s)u} \frac{(\lambda u)^k}{k!} du = \left( \frac{\lambda}{\lambda - s} \right)^k - e^{-(\lambda-s)x} \sum_{i=0}^k \left( \frac{\lambda}{\lambda - s} \right)^i \frac{(\lambda x)^{k-i}}{(k-i)!}.$$

Therefore,

$$\begin{aligned}
\mathbb{E}[e^{-sT} | X > 0, I = 1] &= \frac{\mu_1}{\lambda - s} \sum_{k \geq 0} G_2^*(s)^{\lfloor \frac{k+1}{c} \rfloor} \\
&\quad \times \int_0^\infty e^{-sx} g_1(x) \left[ \left( \frac{\lambda}{\lambda - s} \right)^k - e^{-(\lambda-s)x} \sum_{i=0}^k \left( \frac{\lambda}{\lambda - s} \right)^i \frac{(\lambda x)^{k-i}}{(k-i)!} \right] dx \\
&= \frac{\mu_1}{\lambda - s} \sum_{k \geq 0} G_2^*(s)^{\lfloor \frac{k+1}{c} \rfloor} \left[ \left( \frac{\lambda}{\lambda - s} \right)^k G_1^*(s) - \sum_{i=0}^k \left( \frac{\lambda}{\lambda - s} \right)^i \int_0^\infty e^{-\lambda x} g_1(x) \frac{(\lambda x)^{k-i}}{(k-i)!} dx \right] \\
&= \frac{\mu_1}{\lambda - s} \sum_{k \geq 0} \left( \frac{\lambda}{\lambda - s} \right)^k G_2^*(s)^{\lfloor \frac{k+1}{c} \rfloor} \left[ G_1^*(s) - \sum_{j=0}^k \int_0^\infty e^{-\lambda x} g_1(x) \frac{((\lambda - s)x)^j}{j!} dx \right]. \quad (22)
\end{aligned}$$

#### 4.1.3 Calculation of $\mathbb{E}[e^{-sT} | X > 0, I = 2]$

Let  $\{\sigma_2^{(j)}, j \geq 1\}$  be iid rvs with cdf  $G_2$ , pdf  $g_2$ , and LST  $G_2^*$ . We have

$$\begin{aligned}
&\mathbb{E}[e^{-sT} | X > 0, I = 2, U, V] \\
&= \sum_{k, m \geq 0} \mathbb{E}[e^{-sT} | X > 0, I = 2, U, V, Y_r = k, N_r = m] \mathbb{P}(Y_r = k, N_r = m | U) \\
&= e^{-\lambda U} \sum_{k, m \geq 0} \mathbb{E}[e^{-sT} | X > 0, I = 2, U, V, Y_r = k, N_r = m] \frac{(\lambda U)^k}{k!} \mathbb{P}(N_r = m), \quad (23)
\end{aligned}$$

since the rvs  $Y_r$  and  $N_r$  are independent.

Conditioned on  $I = 2, U, V, Y_r = k$  and  $N_r = m$ , the sojourn time of the tagged customer is  $V + \sum_{j=1}^i \sigma_2^{(j)}$  when  $ic \leq Y_r + N_r < (i+1)c$  for  $i \geq 0$ , yielding (derivation similar to (21))

$$\mathbb{E}[e^{-sT} | X > 0, I = 2, U, V, Y_r = k, N_r = m] = \mathbb{E}[e^{-sV}] G_2^*(s)^{\lfloor \frac{k+m}{c} \rfloor}. \quad (24)$$

Therefore,

$$\begin{aligned}
&\mathbb{E}[e^{-sT} | X > 0, I = 2, U, V] \\
&= e^{-\lambda U} \sum_{k, m \geq 0} \mathbb{E}[e^{-sV}] G_2^*(s)^{\lfloor \frac{k+m}{c} \rfloor} \frac{(\lambda U)^k}{k!} \mathbb{P}(N_r = m),
\end{aligned}$$

and

$$\begin{aligned}
& \mathbb{E}[e^{-sT} \mid X > 0, I = 2] \\
&= \mu_2 \int_0^\infty \int_0^\infty e^{-\lambda u} \sum_{k,m \geq 0} \mathbb{E}[e^{-sV}] G_2^*(s)^{\lfloor \frac{k+m}{c} \rfloor} \frac{(\lambda u)^k}{k!} \mathbb{P}(N_r = m) g_2(u+v) du dv \\
&= \mu_2 \sum_{k,m \geq 0} G_2^*(s)^{\lfloor \frac{k+m}{c} \rfloor} \mathbb{P}(N_r = m) \int_{x=0}^\infty e^{-sx} g_2(x) \int_{u=0}^x e^{-(\lambda-s)u} \frac{(\lambda u)^k}{k!} du dx \\
&= \frac{\mu_2}{\lambda-s} \sum_{k,m \geq 0} \left( \frac{\lambda}{\lambda-s} \right)^k G_2^*(s)^{\lfloor \frac{k+m}{c} \rfloor} \mathbb{P}(N_r = m) \left[ G_2^*(s) - \sum_{j=0}^k \int_0^\infty e^{-\lambda x} g_2(x) \frac{((\lambda-s)x)^j}{j!} dx \right], \tag{25}
\end{aligned}$$

by using (22). In summary, cf. (15), (19), (20), (22), and (25),

$$\begin{aligned}
T^*(s) &= p(0) G_1^*(s) \\
&+ p(0) \frac{\lambda}{\lambda-s} \sum_{k \geq 0} \left( \frac{\lambda}{\lambda-s} \right)^k G_2^*(s)^{\lfloor \frac{k+1}{c} \rfloor} \left[ G_1^*(s) - \sum_{j=0}^k \int_0^\infty e^{-\lambda x} g_1(x) \frac{((\lambda-s)x)^j}{j!} dx \right] \\
&+ (1-p(0))(1+\rho_1) \frac{\mu_2}{\lambda-s} \sum_{k,m \geq 0} \left( \frac{\lambda}{\lambda-s} \right)^k G_2^*(s)^{\lfloor \frac{k+m}{c} \rfloor} \mathbb{P}(N_r = m) \\
&\times \left[ G_2^*(s) - \sum_{j=0}^k \int_0^\infty e^{-\lambda x} g_2(x) \frac{((\lambda-s)x)^j}{j!} dx \right],
\end{aligned}$$

which proves (8) upon using definition (9).

## 4.2 Proof of Proposition 1: Proof of (10)-(11)

Let  $N_r(k)$  be the number of customers in the system at the beginning of the  $k$ -th regular batch. Clearly,  $\{N_r(k)\}_{k \geq 1}$  is an homogeneous, irreducible and aperiodic Markov chain on  $\{1, 2, \dots\}$ , with steady-state  $N_r$ .

Define  $\mathbb{P}(i, j) = \mathbb{P}(N_r(k+1) = j \mid N_r(k) = i)$  for  $i, j \geq 1$ . Recall that  $\pi_r(i) = \mathbb{P}(N_r = i)$  for  $i \geq 1$ . First, let us identify  $\mathbb{P}(i, j)$ .

Denote by  $Y_r(k)$  the number of customers that have arrived during the  $k$ th regular batch. Recall that  $Y_r$  denotes the steady-state version of the sequence  $\{Y_r(k)\}_k$ .

At the end of the  $k$ th regular batch the next service is an regular batch if  $N_r + Y_r > c$  and an exceptional batch otherwise; in the latter case, the next regular bath occurs when at least  $c$  customers arrive during an exceptional batch. First, notice that

$$\mathbb{P}(i, j) = 0 \quad i > c, 1 \leq j < i - c.$$

From now on we assume that  $1 \leq i \leq c$  or that  $i > c$  and  $j \geq i - c$ .

Clearly,

$$\mathbb{P}(Y_e = m) = d_1(m) \quad (26)$$

$$\mathbb{P}(Y_r = m) = d_2(m), \quad (27)$$

for all  $m \geq 0$ , where  $d_l(m)$  is defined in (6).

Take  $1 \leq i \leq c$  and  $j \geq 1$ . Denote by  $Y_e^{(1)}, \dots, Y_e^{(l)}$  the number of arrivals during  $l$  successive exceptional batches,  $l \geq 1$ . Clearly,  $\{Y_e^{(l)}\}_l$  is an iid sequence. We have

$$\begin{aligned} \mathbb{P}(i, j) &= \mathbb{P}(Y_r + i - c = j) \\ &+ \sum_{l \geq 0} \mathbb{P}(i + Y_r \leq c, Y_e^{(1)} \leq c - 1, \dots, Y_e^{(l)} \leq c - 1, Y_e^{(l+1)} = c + j - 1) \\ &= \mathbb{P}(Y_r = c - i + j) + \mathbb{P}(Y_r \leq c - i) \mathbb{P}(Y_e = c + j - 1) \sum_{l \geq 0} \mathbb{P}(Y_e \leq c - 1)^l \\ &= \mathbb{P}(Y_r = c - i + j) + \mathbb{P}(Y_r \leq c - i) \frac{\mathbb{P}(Y_e = c + j - 1)}{\mathbb{P}(Y_e \geq c)}, \end{aligned} \quad (28)$$

by using the independence of  $Y_r$  and of the iid sequence  $\{Y_e^{(l)}\}_l$ . With (26)-(27), (28) rewrites

$$\mathbb{P}(i, j) = d_2(c - i + j) + \frac{d_1(c + j - 1)}{\sum_{m \geq c} d_1(m)} \sum_{m=0}^{c-i} d_2(m),$$

for  $1 \leq i \leq c$  and  $j \geq 1$ .

Take now  $i > c$  and  $j \geq i - c$ . In this case, the  $k$ th and the  $(k + 1)$ st regular batches are next to each other (i.e. corresponding to two consecutive services). Hence,

$$\mathbb{P}(i, j) = \mathbb{P}(Y_r + i - c = j) = \int_0^\infty e^{-\lambda u} \frac{(\lambda u)^{j+c-i}}{(j+c-i)!} g_2(u) du = d_2(j + c - i),$$

for  $i > c$  and  $j \geq i - c$ . In summary,

$$\mathbb{P}(i, j) = \begin{cases} 0 & \text{if } i > c, 1 \leq j < i - c \\ d_2(c - i + j) + \frac{d_1(c+j-1)}{\sum_{m \geq c} d_1(m)} \sum_{m=0}^{c-i} d_2(m) & \text{if } 1 \leq i \leq c, j \geq 1 \\ d_2(j + c - i) & \text{if } i > c, j \geq i - c. \end{cases} \quad (29)$$



From the balance equations  $\pi = \pi P$  and (29), we get

$$\begin{aligned}
N_r(z) &= \sum_{j \geq 1} \pi_r(j) z^j \\
&= \sum_{j \geq 1} \sum_{i=1}^c \left( d_2(c-i+j) + \frac{d_1(c+j-1)}{\sum_{m \geq c} d_1(m)} \sum_{m=0}^{c-i} d_2(m) \right) \pi_r(i) z^j \\
&\quad + \sum_{i \geq c+1} \sum_{j \geq i-c} d_2(j+c-i) \pi_r(i) z^j.
\end{aligned} \tag{30}$$

Recall the definition of  $h_l(m, z)$  in (7). We have

$$\begin{aligned}
\sum_{j \geq 1} \sum_{i=1}^c d_2(c-i+j) \pi_r(i) z^j &= \int_0^\infty e^{-\lambda u} g_2(u) \sum_{j \geq 1} \sum_{i=1}^c \frac{(\lambda u)^{c-i+j}}{(c-i+j)!} \pi_r(i) z^j du \\
&= \int_0^\infty e^{-\lambda u} g_2(u) z^{-c+i} \sum_{i=1}^c \sum_{m=c-i+1}^\infty \frac{(\lambda u z)^m}{m!} \pi_r(i) du \\
&= \int_0^\infty e^{-\lambda u} g_2(u) z^{-c+i} \sum_{i=1}^c \left( e^{\lambda u z} - \sum_{m=0}^{c-i} \frac{(\lambda u z)^m}{m!} \right) \pi_r(i) du \\
&= z^{-c} G_2^*(\lambda(1-z)) \sum_{i=1}^c \pi_r(i) z^i - z^{-c} \sum_{i=1}^c \pi_r(i) z^i \left( \sum_{m=0}^{c-i} \int_0^\infty e^{-\lambda u} g_2(u) \frac{(\lambda u z)^m}{m!} du \right) \\
&= z^{-c} \sum_{i=1}^c \pi_r(i) z^i \left[ G_2^*(\lambda(1-z)) - \sum_{m=0}^{c-i} h_2(m, z) \right],
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{j \geq 1} \sum_{i=1}^c \frac{d_1(c+j-1)}{\sum_{m \geq c} d_1(m)} \sum_{m=0}^{c-i} d_2(m) \pi_r(i) z^j \\
&= z^{1-c} \sum_{i=1}^c \pi_r(i) \int_0^\infty e^{-\lambda u} \sum_{j \geq 1} \frac{(\lambda z u)^{c+j-1}}{(c+j-1)!} dg_1(u) \frac{\sum_{m=0}^{c-i} d_2(m)}{\sum_{m \geq c} d_1(m)} \\
&= z^{1-c} \left( G_1^*(\lambda(1-z)) - \sum_{m=0}^{c-1} h_1(m, z) \right) \sum_{i=1}^c \pi_r(i) \frac{\sum_{m=0}^{c-i} d_2(m)}{\sum_{m \geq c} d_1(m)},
\end{aligned}$$

and

$$\begin{aligned}
\sum_{i=c+1}^{\infty} \sum_{j=i-c}^{\infty} d_2(j+c-i) z^j \pi_r(i) &= \int_0^{\infty} e^{-\lambda u} g_2(u) \sum_{i=c+1}^{\infty} \sum_{j=i-c}^{\infty} \frac{(\lambda u)^{j+c-i}}{(j+c-i)!} \pi_r(i) z^j du \\
&= \int_0^{\infty} e^{-\lambda u} g_2(u) \sum_{i=c+1}^{\infty} z^{i-c} \pi_r(i) \left( \sum_{l \geq 0} \frac{(\lambda u z)^l}{l!} \right) du \\
&= z^{-c} \sum_{i=c+1}^{\infty} \pi_r(i) z^i \int_0^{\infty} e^{-\lambda u(1-z)} g_2(u) du \\
&= \left( N_r(z) - \sum_{i=1}^c \pi_r(i) z^i \right) z^{-c} G_2^*(\lambda(1-z)).
\end{aligned}$$

Putting everything together and using the identity  $\sum_{m \geq c} d_1(m) = 1 - \sum_{m=0}^{c-1} d_1(m)$  yields

$$\begin{aligned}
N_r(z) &= \frac{1}{G_2^*(\lambda(1-z)) - z^c} \sum_{i=1}^c \pi_r(i) \left[ z^i \sum_{m=0}^{c-i} h_2(m, z) \right. \\
&\quad \left. - z \left( G_1^*(\lambda(1-z)) - \sum_{m=0}^{c-1} h_1(m, z) \right) \frac{\sum_{m=0}^{c-i} d_2(m)}{1 - \sum_{m=0}^{c-1} d_1(m)} \right], \tag{31}
\end{aligned}$$

which establishes (10)-(11). For  $j = 1, \dots, q-1$ ,  $\xi_j$  is a zero of multiplicity  $n_j$  of  $G_2^*(\lambda(1-z)) - z^c$  (notice that  $\xi_j \neq 1$  for  $j = 1, \dots, q-1$  since we have assumed, by convention, that  $\xi_q = 1$ ). Therefore, for  $N_r(z)$  to be analytic in  $\{|z| \leq 1\}$  the numerator in (10), defined to be  $J(z)$  (see (11)), must vanish  $n_j$  times at  $z = \xi_j$ . This gives rise to the conditions (12). On the other hand,

$$1 = N_r(1) = \lim_{z \rightarrow 1} \frac{J(z)}{G_2^*(\lambda(1-z)) - z^c} = \lim_{z \rightarrow 1} \frac{\frac{d}{dz} J(z)}{-\lambda \frac{d}{dt} G_2^*(t)|_{t=\lambda(1-z)} - cz^{c-1}} = \frac{\frac{d}{dz} J(z)|_{z=1}}{\rho_2 - c},$$

where the second equality follows by L'Hôpital's rule which applies since  $J(1) = 1$ , as this can easily be checked. This gives condition (13) and concludes the proof of Proposition 1.

### 4.3 Application: Exponential service times

Define  $r_l = \frac{\rho_l}{1+\rho_l}$ ,  $l = 1, 2$ , where we recall that  $\rho_l = \frac{\lambda}{\rho_l}$ .

Assume that  $g_i(x) = \mu_i e^{-\mu_i x}$  for  $i = 1, 2$  (exponential service times). By (8)-(9), we obtain

$$\begin{aligned}
T^*(s) &= p(0) \frac{\mu_1}{\mu_1 + s} + p(0) \frac{\mu_1}{\mu_1 + s} \sum_{k \geq 0} r_1^{k+1} \left( \frac{\mu_2}{\mu_2 + s} \right)^{\lfloor \frac{k+1}{c} \rfloor} \\
&\quad + \frac{1 - p(0)(1 + \rho_1)}{1 + \rho_2} \sum_{k, m \geq 0} \mathbb{P}(N_r = m) r_2^k \left( \frac{\mu_2}{\mu_2 + s} \right)^{\lfloor \frac{k+m}{c} \rfloor + 1}. \tag{32}
\end{aligned}$$

Denote by  $f_T(t)$  the pdf of the sojourn time  $T$ . We have  $T^*(s) = \int_0^\infty e^{-st} f_T(t) dt$ . Let us identify  $f_T(t)$  from the value of  $T^*(s)$  given in (32). To this end, we need to find functions  $h(t)$ ,  $h_{1,k}(t)$  and  $h_{2,k}(t)$  such that

$$\begin{aligned} \int_0^\infty e^{-st} h(t) dt &= \frac{\mu_1}{\mu_1 + s} \\ \int_0^\infty e^{-st} h_{1,k}(t) dt &= \left( \frac{\mu_2}{\mu_2 + s} \right)^k, \quad k \geq 1 \\ \int_0^\infty e^{-st} h_{2,k}(t) dt &= \frac{\mu_1}{\mu_1 + s} \cdot \left( \frac{\mu_2}{\mu_2 + s} \right)^k, \quad k \geq 0. \end{aligned}$$

Clearly  $h(t) = \mu_1 e^{-\mu_1 t}$  and  $h_{1,k}(t) = \mu_2 (\mu_2 t)^{k-1} e^{-\mu_2 t} / (k-1)!$ . When  $\mu_1 = \mu_2$ ,  $h_{2,k}(t) = h_{1,k+1}(t)$  for  $k \geq 0$ . It remains to find  $h_{2,k}(t)$  when  $\mu_1 \neq \mu_2$ . From the partial fraction decomposition

$$\frac{1}{\mu_1 + s} \cdot \frac{1}{(\mu_2 + s)^k} = \frac{(-1)^k}{(\mu_1 - \mu_2)^k} \cdot \frac{1}{\mu_1 + s} + \sum_{i=1}^k \frac{(-1)^{k+i}}{(\mu_1 - \mu_2)^{k+1-i}} \cdot \frac{1}{(\mu_2 + s)^i},$$

and the values found for  $h(t)$  and  $h_{1,k}(t)$ , we obtain

$$h_{2,k}(t) = (-1)^k \left( \frac{\mu_2}{\mu_1 - \mu_2} \right)^k \mu_1 e^{-\mu_1 t} + \frac{\mu_1}{\mu_2} \sum_{i=1}^k (-1)^{k+i} \left( \frac{\mu_2}{\mu_1 - \mu_2} \right)^{k+1-i} \frac{(\mu_2 t)^{i-1}}{(i-1)!} \mu_2 e^{-\mu_2 t},$$

when  $\mu_1 \neq \mu_2$  and  $k \geq 0$ . Therefore,

$$f_T(t) = p(0)h(t) + p(0) \sum_{k \geq 0} r_1^{k+1} h_{1, \lfloor \frac{k+1}{c} \rfloor + 1}(t) + \kappa \sum_{k, m \geq 0} \mathbb{P}(N_r = m) r_2^k h_{1, \lfloor \frac{k+m}{c} \rfloor + 1}(t), \quad (33)$$

if  $\mu_1 = \mu_2$ , and

$$f_T(t) = p(0)h(t) + p(0) \sum_{k \geq 0} r_1^{k+1} h_{2, \lfloor \frac{k+1}{c} \rfloor}(t) + \kappa \sum_{k, m \geq 0} \mathbb{P}(N_r = m) r_2^k h_{1, \lfloor \frac{k+m}{c} \rfloor + 1}(t), \quad (34)$$

if  $\mu_1 \neq \mu_2$ , with  $\kappa := \frac{1-p(0)(1+\rho_1)}{1+\rho_2}$ .

The  $n$ -order moment of the sojourn time  $T$  can be obtained from (33)-(34) as shown below. By the equality  $\int_0^\infty \nu t^n \frac{(\nu t)^{k-1}}{(k-1)!} e^{-\nu t} dt = \frac{(k+n-1)!}{(k-1)! \nu^n}$  which holds for any  $\nu \in (0, \infty)$ ,  $k \in \mathbb{Z}^+ - \{0\}$ ,  $n \in \mathbb{Z}^+$ , routine algebra gives

$$\begin{aligned} \mathbb{E}[T^n] &= \int_0^\infty t^n f_T(t) dt \\ &= p(0) \frac{n!}{\mu^n} + \frac{p(0)}{\mu^n} \sum_{k \geq 0} r^{k+1} \frac{(\lfloor \frac{k+1}{c} \rfloor + n)!}{\lfloor \frac{k+1}{c} \rfloor!} + \frac{\kappa}{\mu^n} \sum_{k, m \geq 0} \mathbb{P}(N_r = m) r^k \frac{(\lfloor \frac{k+m}{c} \rfloor + n)!}{\lfloor \frac{k+m}{c} \rfloor!}, \end{aligned} \quad (35)$$

if  $\mu_1 = \mu_2 := \mu$  with  $\rho := \lambda/\mu$  and  $r := r_1 = r_2$ , and

$$\begin{aligned} \mathbb{E}[T^n] &= p(0) \frac{n!}{\mu_1^n} \\ &+ p(0) \sum_{k \geq 0} r_1^{k+1} \left[ (-1)^k \left( \frac{\mu_2}{\mu_1 - \mu_2} \right)^k \frac{n!}{\mu_1^n} + \frac{\mu_1}{\mu_2^{n+1}} \sum_{i=1}^k (-1)^{k+i} \left( \frac{\mu_2}{\mu_1 - \mu_2} \right)^{k+1-i} \frac{(n+i-1)!}{(i-1)!} \right] \\ &+ \frac{\kappa}{\mu_2^n} \sum_{k, m \geq 0} \mathbb{P}(N_r = m) r_2^k \frac{(\lfloor \frac{k+m}{c} \rfloor + n)!}{\lfloor \frac{k+m}{c} \rfloor!}, \end{aligned} \quad (36)$$

if  $\mu_1 \neq \mu_2$ .

Let us now calculate  $\mathbb{P}(N_r = m)$ . From the identity  $\int_0^\infty \theta e^{-\theta u} u^m du = m!/\theta^m$  for  $m \geq 0$ , we find (cf. (6)-(7))

$$h_l(m, z) = \mu_l \int_0^\infty e^{-(\lambda+\mu_l)u} \frac{(\lambda z u)^m}{m!} du = \frac{\mu_l (\lambda z)^m}{m! (\lambda + \mu_l)} \times \frac{m!}{(\lambda + \mu_l)^m} = \frac{1}{1 + \rho_l} \left( \frac{z \rho_l}{1 + \rho_l} \right)^m,$$

and  $d_l(m) = h_l(m, 1) = \frac{1}{1 + \rho_l} \left( \frac{\rho_l}{1 + \rho_l} \right)^m$ . With that,  $N_r(z)$  in (10) becomes

$$\begin{aligned} N_r(z) &= \frac{1 + \rho_2(1-z)}{1 - z^c - \rho_2(1-z)z^c} \sum_{i=1}^c \pi_r(i) \left[ \frac{z^i}{1 + \rho_2} \sum_{m=0}^{c-i} \left( \frac{z \rho_2}{1 + \rho_2} \right)^m \right. \\ &\quad \left. - z \left( \frac{1}{1 + \rho_1(1-z)} - \frac{1}{1 + \rho_1} \sum_{m=0}^{c-1} \left( \frac{z \rho_1}{1 + \rho_1} \right)^m \right) \frac{1}{1 + \rho_2} \sum_{m=0}^{c-i} \left( \frac{\rho_2}{1 + \rho_2} \right)^m \right. \\ &\quad \left. - \frac{1}{1 + \rho_1} \sum_{m \geq c} \left( \frac{\rho_1}{1 + \rho_1} \right)^m \right]. \end{aligned} \quad (37)$$

From the identities

$$\begin{aligned} \frac{1}{1 + \rho_2} \sum_{m=0}^{c-i} (z r_2)^m &= \frac{1 - (z r_2)^{c-i+1}}{1 + \rho_2(1-z)} \\ \frac{1}{1 + \rho_2} \sum_{m=0}^{c-i} r_2^m &= 1 - r_2^{c-i+1} \\ \frac{1}{1 + \rho_1} \sum_{m=0}^{c-1} (z r_1)^m &= \frac{1 - (z r_1)^c}{1 + \rho_1(1-z)} \\ \frac{1}{1 + \rho_1} \sum_{m \geq c} r_1^m &= r_1^c, \end{aligned}$$

we get from (37)

$$\begin{aligned}
N_r(z) &= \frac{1}{1 - z^c - \rho_2(1 - z)z^c} \\
&\quad \times \sum_{i=1}^c \pi_r(i) \left[ z^i - z^{c+1} r_2^{c-i+1} - z^{c+1} \left( \frac{1 + \rho_2(1 - z)}{1 + \rho_1(1 - z)} \right) (1 - r_2^{c-i+1}) \right].
\end{aligned} \tag{38}$$

Define  $\Psi(z) = 1 - z^c - \rho_2(1 - z)z^c$ . Since  $\Psi(z)$  is a polynomial of degree  $c + 1$  it has  $c + 1$  zeros, counting their multiplicity.

When  $G_2(x) = 1 - e^{-\mu_2 x}$  we know (see discussion at the beginning of Section 4) that  $\Phi(z) := G_2^*(\lambda(1 - z)) - z^c = \frac{1}{1 + \rho_2(1 - z)} - z^c$  has  $c$  *distinct* zeros in  $\{|z| \leq 1\}$  all with multiplicity one. Let  $\xi_1, \dots, \xi_c$  be these zeros. From the identity

$$\Psi(z) = (1 + \rho_2(1 - z))\Phi(z), \tag{39}$$

we conclude from the above that  $\xi_1, \dots, \xi_c$  are also zeros of  $\Psi(z)$ . Denote by  $\xi_{c+1}$  the  $(c + 1)$ -st zero of  $\Psi(z)$ . This zero also has multiplicity one since it cannot be equal to one of the  $\xi_1, \dots, \xi_c$  as otherwise  $\Phi(z)$  would have  $c + 1$  zeros in  $\{|z| \leq 1\}$  which would be incorrect. Since all zeros of  $\Psi(z)$  have multiplicity one, we can write  $\Psi(z)$  as

$$\Psi(z) = \rho_2 \prod_{i=1}^{c+1} (z - \xi_i),$$

yielding

$$\frac{1}{\Psi(z)} = \frac{1}{\rho_2} \sum_{i=1}^{c+1} \frac{d_i}{z - \xi_i},$$

where  $d_i := \frac{1}{\prod_{m=1, m \neq i}^{c+1} (\xi_i - \xi_m)}$ . From

$$\frac{1}{z - \xi_i} = -\frac{1}{\xi_i} \sum_{m \geq 0} \frac{z^m}{\xi_i^m},$$

for  $|z| < |\xi_i|$ , we finally obtain

$$\frac{1}{1 - z^c - \rho_2(1 - z)z^c} = \sum_{m \geq 0} z^m w_m, \quad \forall |z| < \min_{i=1, \dots, c+1} |\xi_i|, \tag{40}$$

with  $w_m := -\frac{1}{\rho_2} \sum_{i=1}^{c+1} \frac{d_i}{\xi_i^{m+1}}$ . Note that  $w_0 = 1$  since the l.h.s. of (40) is equal to one at  $z = 0$

(see Remark 1). Introducing (40) into (38) gives

$$\begin{aligned}
N_r(z) &= \sum_{m \geq 0} \sum_{i=1}^c z^m w_m \pi_r(i) \left( z^i - z^{c+1} r_2^{c-i+1} - z^{c+1} (1 - r_2^{c-i+1}) \left( \frac{1 + \rho_2(1-z)}{1 + \rho_1(1-z)} \right) \right) \\
&= \sum_{m \geq 0} \sum_{i=1}^c z^m w_m \pi_r(i) \left( z^i - z^{c+1} r_2^{c-i+1} - z^{c+1} (1 - r_2^{c-i+1}) \frac{(1 + \rho_2(1-z))}{1 + \rho_1} \sum_{k \geq 0} z^k r_1^k \right),
\end{aligned} \tag{41}$$

by using the identity  $\frac{1}{1+\rho_1(1-z)} = \frac{1}{1+\rho_1} \sum_{m \geq 0} z^m \left( \frac{\rho_1}{1+\rho_1} \right)^m = \frac{1}{1+\rho_1} \sum_{m \geq 0} z^m r_1^m$ .

For evaluating  $T^*(s)$  in (32) we need to identify the probabilities  $\pi_r(j)$  for  $j \geq 1$ . By definition of  $N_r(z)$  in (30),  $\pi_r(j)$  is obtained as the coefficient of  $z^j$  in (30), that is, by the coefficient of  $z^j$  in the r.h.s. of (41). We already know that  $\pi_r(1), \dots, \pi_r(c)$  are obtained as the solution of (12)-(13). We are left with identifying  $\pi_r(j)$  for  $j \geq c+1$ .

To this end, introduce

$$\begin{aligned}
S_1(z) &:= \sum_{m \geq 0} \sum_{i=1}^c z^{m+i} w_m \pi_r(i) = \left( \sum_{m \geq 0} z^m w_m \right) \left( \sum_{i \geq 0} z^i \pi_r(i) \mathbf{1}_{1 \leq i \leq c} \right) \\
&= \sum_{j \geq 1} z^j \sum_{i=1}^{\min(c,j)} w_{j-i} \pi_r(i) \\
S_2(z) &:= \sum_{m \geq 0} \sum_{k \geq 0} z^{m+k} w_m r_1^k = \sum_{j \geq 0} z^j \sum_{l=0}^j w_l r_1^{j-l}.
\end{aligned}$$

Then,

$$\begin{aligned}
N_r(z) &= S_1(z) - \sum_{i=1}^c r_2^{c-i+1} \pi_r(i) \cdot \sum_{m \geq 0} z^{m+c+1} w_m \\
&\quad - S_2(z) \cdot \frac{(1 + \rho_2)z^{c+1} - \rho_2 z^{c+2}}{1 + \rho_1} \cdot \sum_{i=1}^c (1 - r_2^{c-i+1}) \pi_r(i) \\
&= \sum_{j \geq 1} z^j \sum_{i=1}^{\min(c,j)} w_{j-i} \pi_r(i) - A \sum_{m \geq 0} z^{m+c+1} w_m - B \sum_{j \geq 0} z^j \sum_{l=0}^j w_l r_1^{j-l} ((1 + \rho_2)z^{c+1} - \rho_2 z^{c+2}),
\end{aligned} \tag{42}$$

by setting  $A := \sum_{i=1}^c r_2^{c-i+1} \pi_r(i)$  and  $B := \frac{1}{1+\rho_1} \sum_{i=1}^c (1 - r_2^{c-i+1}) \pi_r(i)$  to make notation more compact.

For  $j \geq c + 1$  the coefficient of  $z^j$  in (42), equals to  $\pi_r(j)$ , is given by

$$\pi_r(j) = \sum_{i=1}^c w_{j-i} \pi_r(i) - A w_{j-c-1} - (1 + \rho_2) B \sum_{l=0}^{j-c-1} w_l r_1^{j-c-1-l} + \rho_2 B \sum_{l=0}^{j-c-2} w_l r_1^{j-c-2-l}, \quad (43)$$

where by convention  $\sum_{j=0}^{-1} \bullet = 0$ .

**Remark 1** *The coefficient of  $z^j$  in (42) is given by  $\sum_{i=1}^j w_{j-i} \pi_r(i)$  for  $j = 1, \dots, c$ . However, it turns out that  $w_1 = \dots = w_{c-1} = 0$  so that, since  $w_0 = 1$ , the coefficient of  $z^j$  reduces to  $\pi_r(j)$  for  $j = 1, \dots, c$ , as expected.*

## 5 Computing communication cost

We define the cost of an allocation as the communication cost associated with an allocated user-resource pair. Consider communication cost as a function  $\mathcal{C}$  of the request distance. Then the expected communication cost across the service network is given as

$$\bar{C} = E[\text{cost}] = \int_{d=0}^{\infty} \mathcal{C}(d) dF_D(d), \quad (44)$$

where  $F_D$  is the request distance distribution. By the equivalence of the one-dimensional distributed service network and corresponding ESABQ,  $F_D$  can be obtained by inverting the LST  $T^*(s)$  in Proposition 1.

One such cost model is widely used in wireless ad hoc networks is [9]

$$\mathcal{C}(d) = t_0 d^\beta. \quad (45)$$

where  $\beta$  is the path loss exponent and  $t_0$  is a constant. When  $\beta \in \mathbb{Z}^+$  the expected communication cost can be computed as

$$\bar{C} = t_0 (-1)^\beta \left. \frac{d^\beta T^*(s)}{ds^\beta} \right|_{s=0}. \quad (46)$$

For the case when  $\beta \in \mathbb{R}^+ \setminus \mathbb{Z}^+$ , let  $\beta = n - \delta$  with  $n \in \mathbb{Z}^+$  and  $0 < \delta < 1$ . Using results by Cressie et al. ([7], Theorem 1) under certain conditions we have

$$\bar{C} = t_0 (-1)^n \Gamma(\delta)^{-1} \int_{-\infty}^s (s-z)^{\delta-1} \frac{d^n T^*(z)}{dz^n} dz \Big|_{s=0}, \quad (47)$$

where  $\Gamma(\cdot)$  is the gamma function. Note that, the conditions in ([7], Theorem 1) are met if the  $(n+1)^{st}$  moments of  $X$  and  $Z$  are finite.

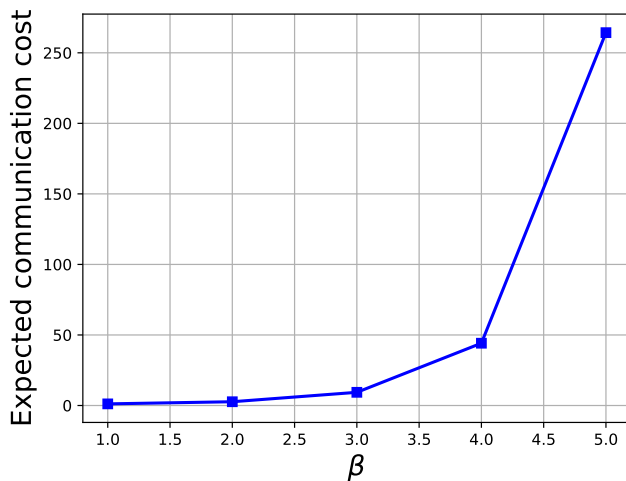


Figure 3: Expected communication cost under path loss based cost model for different values of  $\beta$ .

## 6 Performance Evaluation

In this section, we examine the effect of  $\beta$  on the expected communication cost obtained through analysis for path loss based cost model. We consider the mean user density  $\lambda = 0.5$ . We assume the servers to be distributed according to an exponential distribution with density  $\mu = 1$ . We assume  $c = 2$  and  $t_0 = 1$ . We use numerical methods to compute cost defined in Section 5.

We first calculated the probabilities  $\pi_r(1), \dots, \pi_r(c)$  by solving the set of linear Equations (12)-(13). We then computed the probabilities  $\pi_r(j), j > c$  using (43). We finally plugged these values into (35) by doing some truncation in the infinite single and double sum. In particular, the number of terms considered in the calculations of the single and double sum in (35) were both 10.

We plot the expected communication cost for different values of  $\beta$  in the path loss based cost model as shown in Figure 3. We observe that with increase in the value of  $\beta$ , the expected communication cost increases.

## 7 Related Literature

Holroyd et al. [14] first obtained upper and lower bounds on the expected request distance under stable matching policy for the case when users and resources are distributed according to a Poisson process on a  $d$ -dimensional spatial network. Abadi et al. [1] were the first to derive bounds on expected request distance for a unidirectional policy in a one-dimensional



service network under Poisson distributed users and resources. They assume server capacity to be one in their analysis. Panigrahy et al. [22] derived expressions for expected request distance in a more general setting where either users or resources are Poisson distributed and server capacities are greater than one. Our work complements and extends [22] in the sense that we derive expressions for request distance distribution for a one-dimensional service network.

In the queueing theory field, there exists a vast volume of literature on exceptional service queueing systems. For example, [25], [21], [4], [8] modeled the exceptional queueing system under different inter-customer arrival times and service time distributions. However, these works do not focus on batch service systems. For literature on non-accessible batch service queueing models with regular (non-exceptional) service, refer to [10], [5]. [16] discussed queue length distribution for an accessible regular batch service queueing system for different arrival and service time distributions. All of the above mentioned models either analyze exceptional services or accessible batch service queueing systems individually but not both. While authors in [22] first derived the queue-length distribution for an exceptional service accessible batch queueing system, we derive sojourn time distribution for such a system under Poisson customer arrivals and general service time distribution.

## 8 Conclusion

In this work we analyzed the expected communication cost associated with MTR allocation policy in a one-dimensional distributed service network. We derived expressions for the LS transform of request distance distribution for Poisson distributed users with general inter-resource distance distributions and discussed few special cases. Using the request distance distribution, we computed the expected communication cost for a path loss based cost model. Going further we expect to analyze the service network for other inter-user and inter-server distributions under different cost models.

## 9 Acknowledgment

This research was sponsored by the U.S. ARL and the U.K. MoD under Agreement Number W911NF-16-3-0001 and by the NSF under Grant CNS-1617437. This document does not contain technology or technical data controlled under either the U.S. International Traffic in Arms Regulations or the U.S. Export Administration Regulations. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the National Science Foundation, U.S. ARL or the U.K. MoD.

## References

- [1] H. K. Abadi and B. Prabhakar. Stable Matchings in Metric Spaces: Modeling Real-World Preferences using Proximity. *arXiv:1710.05262*, 2017.
- [2] P. Agarwal, A. Efrat, and M. Sharir. Vertical Decomposition of Shallow Levels in 3-Dimensional Arrangements and Its Applications. *11th Annual ACM Symposium on Computational Geometry*, pages 39–50, 1995.
- [3] R. Ahuja, T. Magnanti, and J. Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice-Hall, Inc, 1993.
- [4] Y. Baba. On M/G/1 Queues with the First n Customers of Each Busy Period Receiving Exceptional Services. *J. of Operations Research Society of Japan*, 42(4):490–500, 1999.
- [5] N. T. J. Bailey. On Queueing Processes with Bulk Service. *J. R. Stat. Soc.*, 16:80–87, 1954.
- [6] M. L. Chaudhry and J. G. C. Templeton. *A First Course in Bulk Queues*. John Wiley and Sons, 1983.
- [7] N. Cressie and M. Borkent. The Moment Generating Function has its Moments. 13:337–344, 1986.
- [8] Do Le Minh. Analysis of the Exceptional Queueing System By the Use of Regenerative Processes and Analytical Methods. *Mathematics of Operations Research*, 5(1):147–159, 1980.
- [9] S. Doshi and S. Bhandare. An On-demand Minimum Energy Routing Protocol for a Wireless ad-hoc Network. *ACM Mobile Computing and Communications Review*, 6(3), 2002.
- [10] F. Downton. Waiting Time in Bulk Service Queues. *Journal of the Royal Statistical Society: Series B (Methodological)*, 17(2):256–261, jul 1955.
- [11] P. G. Harrison. Teaching M/G/1 Theory with Extension to Priority Queues. *IEEE Proceedings - Computers and Digital Techniques*, 147(1), 2000.
- [12] P. G. Harrison and N. M. Patel. Performance Modeling of Communication Networks and Computer Architectures. *Addison-Wesley*, 1993.
- [13] I. W. H. Ho, K. K. Leung, and J. W. Polak. Stochastic Model and Connectivity Dynamics for vanets in Signalized Road Systems. *IEEE/ACM Transactions on Networking*, 19(1):195–208, 2011.

- [14] A. E. Holroyd, R. Pemantle, R. Peres, and O. Schramm. Poisson Matching. *Annales de l'IHP Probabilites et Statistiques*, 45:266–287, 2009.
- [15] L. Kleinrock. *Queueing Systems, Volume 1: Theory*. Wiley & Sons, New York, 1975.
- [16] A. Krishnamoorthy and P. V. Ushakumari. A Queueing System with Single Arrival Bulk Service and Single Departure. *Mathematical and Computer Modelling*, 31(2-3):99–108, 2000.
- [17] H. W. Kuhn. The Hungarian Method for the Assignment Problem. *Naval Research Logistics Quarterly*, 2:83–97, 1955.
- [18] K. K. Leung, W. A. Massey, and W. Whitt. Traffic Models for Wireless Communication Networks. *IEEE Journal on Selected Areas in Communications*, 12(8):1353–1364, 1994.
- [19] L. Li, B. Zhang, and J. Zheng. A Study on One-Dimensional k-Coverage Problem in Wireless Sensor Networks. *Wirel. Commun. Mob. Comput.*, 13:1–11, 2013.
- [20] J. Orlin. A Polynomial Time Primal Network Simplex Algorithm for Minimum Cost Flows. *Mathematical Programming*, 78:109–129, 1997.
- [21] A. G. Pakes. On the Busy Period of the Modified GI/G/1 Queue. *J. Appl. Prob.*, 10:192–197, 1973.
- [22] N. K. Panagahy, P. Basu, P. Nain, D. Towsley, A. Swami, K. S. Chan, and K. K. Leung. Resource allocation in one-dimensional distributed service networks with applications. *Performance Evaluation*, 142, September, article 102110 2020.
- [23] M. Shirer. The Growth in Connected IoT Devices. 2019.
- [24] P. M. Vaidya. Geometry Helps in Matching. *SIAM J. Computing*, 18:1201–1225, 1989.
- [25] P. Welch. On a Generalized M/G/1 Queueing Process in Which The First Customer of Each Busy Period Receives Exceptional Service. *Operations Research*, 12:736–752, 1964.

## 10 Appendix

### 10.1 Calculation of $\alpha_Z$

We have

$$\begin{aligned}
\mathbb{E}[Z] &= \int_0^\infty (1 - F_Z(x)) dx \\
&= \frac{1}{1 - F_X^*(\lambda)} \int_0^\infty \left(1 - \lambda e^{\lambda x} (B_X(\infty) - B_X(x))\right) dx \quad \text{by (3)} \\
&= \frac{1}{1 - F_X^*(\lambda)} \int_{x=0}^\infty \left(1 - \lambda e^{\lambda x} \int_{u=x}^\infty e^{-\lambda u} F_X(u) du\right) dx \\
&= \frac{1}{1 - F_X^*(\lambda)} \int_{x=0}^\infty \int_{u=x}^\infty (1 - F_X(u)) \lambda e^{-\lambda(u-x)} du dx \\
&= \frac{1}{1 - F_X^*(\lambda)} \int_{u=0}^\infty (1 - F_X(u)) e^{-\lambda u} \int_{x=0}^u \lambda e^{\lambda x} dx du \\
&= \frac{1}{1 - F_X^*(\lambda)} \int_0^\infty (1 - F_X(u)) e^{-\lambda u} (e^{\lambda u} - 1) du \\
&= \frac{1}{1 - F_X^*(\lambda)} \int_0^\infty (1 - F_X(u)) du - \frac{1}{1 - F_X^*(\lambda)} \int_0^\infty (1 - F_X(u)) e^{-\lambda u} du \\
&= \frac{\alpha_X}{1 - F_X^*(\lambda)} - \frac{1}{1 - F_X^*(\lambda)} \left( \frac{1}{\lambda} - \frac{1}{\lambda} \int_0^\infty e^{-\lambda u} F_X'(u) du \right) \quad \text{by doing an integration by part} \\
&= \frac{\alpha_X}{1 - F_X^*(\lambda)} - \frac{1}{1 - F_X^*(\lambda)} \left( \frac{1}{\lambda} - \frac{1}{\lambda} F_X^*(\lambda) \right) \\
&= \frac{\alpha_X}{1 - F_X^*(\lambda)} - \frac{1}{\lambda},
\end{aligned}$$

with  $\alpha_X = \frac{1}{\mu}$ .

### 10.2 Proof of Corollary 1

When  $c = 1$  it is easy to derive from (5) that  $p(0) = \frac{1-\rho_2}{1-\rho_2+\rho_1}$ . Letting  $c = 1$  in (8) gives

$$\begin{aligned}
T^*(s) &= \frac{(1 - \rho_2)G_1^*(s)}{1 - \rho_2 + \rho_1} + \frac{\lambda(1 - \rho_2)G_2^*(s)}{(1 - \rho_2 + \rho_1)(\lambda - s)} \sum_{k \geq 0} \left( \frac{\lambda G_2^*(s)}{\lambda - s} \right)^k I_1(s, k) \\
&\quad + \frac{\lambda \rho_1 N_r(G_2^*(s))}{(1 - \rho_2 + \rho_1)(\lambda - s)} \sum_{k \geq 0} \left( \frac{\lambda G_2^*(s)}{\lambda - s} \right)^k I_2(s, k). \tag{48}
\end{aligned}$$

We have by (9)

$$\begin{aligned}
\sum_{k \geq 0} \left( \frac{\lambda G_2^*(s)}{\lambda - s} \right)^k I_l(s, k) &= \frac{(\lambda - s)G_l^*(s)}{\lambda(1 - G_2^*(s)) - s} - \int_0^\infty e^{-\lambda x} g_l(x) \sum_{j \geq 0} \frac{((\lambda - s)x)^j}{j!} \sum_{k \geq j} \left( \frac{\lambda G_2^*(s)}{\lambda - s} \right)^k dx \\
&= \frac{(\lambda - s)G_l^*(s)}{\lambda(1 - G_2^*(s)) - s} - \frac{\lambda - s}{\lambda(1 - G_2^*(s)) - s} \int_0^\infty e^{-\lambda x} g_l(x) \sum_{j \geq 0} \frac{(\lambda G_2^*(s)x)^j}{j!} dx \\
&= \frac{(\lambda - s)(G_l^*(s) - G_l^*(\lambda(1 - G_2^*(s))))}{\lambda(1 - G_2^*(s)) - s}, \quad l = 1, 2. \tag{49}
\end{aligned}$$

Introducing (49) into (48) yields

$$\begin{aligned}
T^*(s) &= \frac{(1 - \rho_2)G_1^*(s)}{1 - \rho_2 + \rho_1} + \frac{\lambda}{(1 - \rho_2 + \rho_1)(\lambda(1 - G_2^*(s)) - s)} \tag{50} \\
&\quad \times [(1 - \rho_2)G_2^*(s)(G_1^*(s) - G_1^*(\lambda(1 - G_2^*(s)))) + \rho_1 N_r(G_2^*(s))(G_2^*(s) - G_2^*(\lambda(1 - G_2^*(s))))].
\end{aligned}$$

It remains to calculate  $N_r(z)$  in (10)-(11) when  $c = 1$ . We have (recall that  $h_l(0, z) = d_l(0) = G_l^*(\lambda)$  for  $l = 1, 2$ )

$$\begin{aligned}
N_r(z) &= \frac{z}{G_2^*(\lambda(1 - z)) - z} \left[ h_2(0, z) - (G_1^*(\lambda(1 - z)) - h_1(0, z)) \frac{d_2(0)}{1 - d_1(0)} \right] \pi_r(1) \\
&= \frac{zG_2^*(\lambda)}{G_2^*(\lambda(1 - z)) - z} \left[ 1 - \frac{G_1^*(\lambda(1 - z)) - G_1^*(\lambda)}{1 - G_1^*(\lambda)} \right] \pi_r(1) \\
&= \frac{zG_2^*(\lambda)(1 - G_1^*(\lambda(1 - z)))}{(G_2^*(\lambda(1 - z)) - z)(1 - G_1^*(\lambda))} \pi_r(1). \tag{51}
\end{aligned}$$

The constant  $\pi_r(1)$  is obtained from (13). We easily find

$$\pi_r(1) = \frac{(1 - \rho_2)(1 - G_1^*(\lambda))}{\rho_1 G_2^*(\lambda)},$$

so that by (51)

$$N_r(z) = (1 - \rho_2) \frac{z(1 - G_1^*(\lambda(1 - z)))}{\rho_1 (G_2^*(\lambda(1 - z)) - z)}. \tag{52}$$

Finally, by (50)-(52)

$$\begin{aligned}
T^*(s) &= \frac{(1 - \rho_2)G_1^*(s)}{1 - \rho_2 + \rho_1} - \frac{\lambda(1 - \rho_2)G_2^*(s)(1 - G_1^*(s))}{(1 - \rho_2 + \rho_1)(\lambda(1 - G_2^*(s)) - s)} \\
&= \frac{(1 - \rho_2)(sG_1^*(s) + \lambda(G_2^*(s) - G_1^*(s)))}{(1 - \rho_2 + \rho_1)(s - \lambda(1 - G_2^*(s)))},
\end{aligned}$$

which concludes the proof of Corollary 1.