A Tandem Queueing Model for Delay Analysis in Disconnected Ad Hoc Networks

Ahmad Al-Hanbali, Roland de Haan, Richard J. Boucherie, and Jan-Kees van Ommeren
University of Twente, Enschede, The Netherlands

ABSTRACT

Ad hoc network routing protocols may fail to operate in the absence of an end-to-end connection from source to destination. This deficiency can be resolved by so-called opportunistic networking which exploits the mobility of the nodes by letting them operate as relays according to the store-carry-and-forward paradigm. However, the efficiency of this approach will depend to a large extent on the contact and inter-contact times of node pairs.

In this work, we analyze the delay performance of a small opportunistic network by considering a tandem queueing system. We present an exact packet-level analysis by applying ideas from the polling literature. Due to the state-space expansion, this analysis cannot efficiently be applied for all model parameter settings. For this reason, an analytical approximation is constructed and its excellent performance has extensively been validated. Numerical results on the mean end-to-end delay show that the inter-contact time distribution impacts this metric only through its first two moments. Finally, we study delay optimization under power control.

Keywords: Tandem model, Delay-tolerant networking, Opportunistic networking, Mobile queue , Autonomous server, Performance analysis.

1. INTRODUCTION

End-to-end connectivity is not a natural property of ad hoc networks. For instance, nodes may vary their transmission power, nodes may move, nodes may enter the sleep mode, or nodes may suffer from hardware failures. As a result, the network structure changes dynamically and this may lead to undesired situations of nodes becoming disconnected from parts of the network.

The traditional store-and-forward routing protocols, which require the existence of a connected path between a source and a destination, cannot be employed in highly disconnected ad hoc networks. A solution for this problem is to exploit the mobility of nodes present in the network. The mobile nodes may form in fact bridges which relay traffic between the disconnected parts. This approach has been proposed in the pioneering paper of Grossglauser and Tse [14] as an alternative to the store-and-forward paradigm and it is now known as the store-*carry*-and-forward paradigm in the context of delay-tolerant networking (DTN) [1]. In DTN, the incurred delay to send data between nodes can be very large and unpredictable due to the disconnection problem. Applications of such can be found in, e.g., disaster relief networks, rural networking, environmental monitoring networks, vehicular networks, and interplanetary networks.

A significant amount of research for routing-based approaches in DTN has recently emerged. An important factor in DTN is the so-called contact opportunity between node pairs. Two nodes are in contact if they are within transmission range of one another and thus packet exchange between them is possible. The proposed routing solutions essentially differ on the required knowledge of these contact opportunities. Specifically, depending on whether the contact opportunities are scheduled [16], predicted [20], controlled [33], or opportunistic [25, 27], they can be grouped into different classes. The best performance would be achieved in the full knowledge case of contacts. However, this comes at the expense of a higher complexity both from the implementation and from the maintenance perspectives. In the present work, we will focus on the performance analysis of the opportunistic-based approach where no knowledge is required. For detailed surveys about the different routing-based approaches in DTN we refer to [28, 29].

Another factor that impacts the performance of opportunistic approaches is the inter-contact time which is defined as the time duration between two consecutive contacts of node pairs. The inter-contact time mainly depends on the mobility of the nodes. In [13], simulations showed that for the Random Waypoint [2] and the Random Direction [23] mobility models the distribution of the inter-contact times is exponential when the nodes' transmission range is small. On the contrary, for human mobility, it is shown through experiments that the tail of the distribution of inter-contact times has a power law decay in some finite range [4], and after that it exhibits an exponential decay [17]. In the present work, we will assume that the inter-contact times distribution has a finite first and second moment.

We will analyze the opportunistic approaches in DTN by taking into account, unlike [4, 13, 26], that the transmission of packets may fail due to the short contact time and a retransmission is required. Also, we assume that the source node has a stream of packet arrivals instead of only one packet, like it was considered in [15, 26, 32]. In addition, here we are interested in what happens in a more practical case of small, finite-size networks, rather than in asymptotic cases (see, e.g., [14, 32]). To this end, we adopt the network scenario of a fixed source and destination node and mobile intermediate nodes that serve as relay nodes. As a primary step towards understanding such models, we will study a network model with a single mobile node as a relaying device. Although it is a small model, it contains the main characteristics of an opportunistic network and it is also non-trivial from an analytical perspective.

The network model of our interest is reminiscent of a two-queue tandem model with a single alternating server. Such a tandem model has been analyzed under various servicing strategies (see, e.g., [?]). Typically, these strategies are based on the assumption that the server can be controlled. However, in the mobility-driven model of our interest, the server is autonomous and there is no possibility to control its movement. The research efforts on models with time-limited service periods are also closely related to our work. In a two-queue setting, [5] analyzes the model via boundary value techniques. Unfortunately, the analysis along these lines for more than two queues appears intractable. Time-limited service models have also been studied in the context of polling systems (see, e.g., [10, 22]). However, also in these models, there exists a notion of server control, since it is assumed that whenever a queue becomes empty the server moves to another queue.

In this work, our interest is mainly in the end-to-end delay under opportunistic networking. We assume that data packets arrive according to a Poisson process at the source queue. The mobile node stores the packets received from the source and forwards them to the destination. The source and destination are assumed far apart, so that the mobile node is never in range of both source and destination. The contact times are assumed exponentially distributed. Packets whose transmission is interrupted will be retransmitted during a next contact time.

We study this system at the packet level by considering the tandem queueing model as a particular kind of polling system. That is, a polling system for which customers of one queue move to another queue after being served. This specific polling system is a time-limited polling system extended with the feature that the server remains at a queue even if it becomes empty. We perform an exact analysis for this system by using similar techniques as in [22] and [6].

Due to the state-space expansion, the computation time of the joint queue-length probabilities may grow large for certain model parameters. Therefore, as a complementary tool, we present an analytical approximation for the case that the service requirements at each queue are exponential. The queue-length process at the second queue is then analyzed in isolation as a workload process with Poisson batch arrivals. The Poisson process follows directly from the assumption of exponential contact times. The key element is to approximate the batch size distribution as closely as possible. A similar model has been analyzed by Borst et al. [3]. The authors consider a Poisson batch arrival process for which the batch size depends on the inter-arrival time of the batch. This differs from our model in the sense that batch sizes depend not only on the final inter-arrival period, but also on the previous ones which induce that they are dependent.

Numerical experiments show the excellent performance of the approximation for a broad range of parameter settings. These experiments further show that mean sojourn time is insensitive to third and higher moments of the inter-contact times. Finally, several guidelines are given for delay optimization by power control. In particular, balancing the queues load is not always close to the optimal policy. However, using a simple heuristic based on optimizing the delay of a tandem model of two M/M/1 queues gives nearly optimal results under wide variety of parameter settings.

The main contributions of this article are:

- an analytical model for queue-length and delay analysis in a simple opportunistic network;
- an analytic approximation for the delay in a two-queue tandem model with a single autonomous server;
- third and higher moments of the inter-contact times have negligible impact on mean end-to-end delay;
- load balancing is not an effective tool for delay optimization in DTN.

The rest of the paper is organized as follows. Section 2 gives the description of the opportunistic network model, discusses the stability issues, and finds exact results for sojourn time in the source node and mobile node. Section 3 proposes and analyses an approximation for the sojourn time in the mobile queue. In Section 4, we numerically validate the accuracy of the approximation and we present additional results which give insight on the delay of the network. Section 5 concludes the paper and suggests some research directions.

2. MODEL AND EXACT RESULTS

2.1 Model

We consider a tandem model consisting of 3 first-in-first-out (FIFO) single-server systems with unlimited queue, Q_i, $i = 1, 2, 3$, in which customers arrive to Q_1 and subsequently require service at Q_2 before reaching their destination at Q_3. The special feature of the model is that Q_2 alternates between positions L_1 and L_2 such that the server in Q_1 is available only when Q_2 is at L_1 and the server in Q_2 is available only when Q_2 is at L_2. In addition, Q_2 incurs a switching time from L_i to L_j ($i \neq j,\ j \in \{1,2\}$) during which the server at neither Q_1 nor Q_2 is available. Q_3 is a sink and will not be included in our analysis.

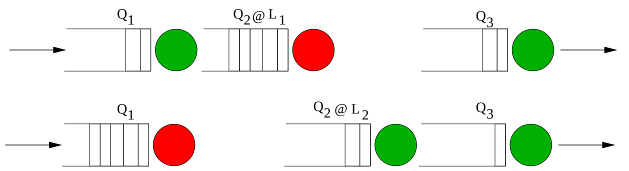

Figure 1: Three queue tandem model with a mobile queue. Top: Q_2 is at L_1, its server is down. Bottom: Q_2 at L_2, its server is up.

Customers arrive to Q_1 according to a Poisson process with arrival rate λ. The service requirement S_i at Q_i has general distribution $S_i(\cdot)$, with Laplace-Stieltjes Transform (LST) $\tilde{S}_i(\cdot)$, and mean $1/\beta_i$. We assume that the service requirements are independent and identically distributed (iid) random variables (rvs).

Movement of Q_2 is autonomous. Q_2 remains at location L_1 (resp. L_2) a (random) time of duration $X_n^{L_1}$ (resp. $X_n^{L_2}$) before it migrates to L_2 (resp. L_1) during its n-th visit. It is important to notice that in the analysis we will use visit (time) rather than contact (time) to refer to (the duration of) a contact opportunity as to be in line with the common practice in the polling literature. After the n-th visit to L_1, Q_2 incurs a switch-over time $C_n^{1,2}$ from L_1 to L_2, and similarly a switch-over time $C_n^{2,1}$ after the n-th visit to L_2. We assume that $C_n^{1,2}$ ($C_n^{2,1}$) is an iid sequence with general distribution $C^{1,2}(\cdot)$ $\left(C^{2,1}(\cdot)\right)$, LST $\tilde{C}^{1,2}(\cdot)$ $\left(\tilde{C}^{2,1}(\cdot)\right)$, and mean $c^{1,2}$ ($c^{2,1}$). We further assume that $X_n^{L_1}$ ($X_n^{L_2}$) is an iid sequence of common exponential distribution with rate α_1 (α_2). Furthermore, we assume $\{X_n^{L_1}, X_n^{L_2}, C_n^{1,2}, C_n^{2,1}\}$ are iid and mutually independent, and also independent at their starting time points of the other rvs in the model (queue length, waiting time, sojourn time, etc.). Therefore, the location of Q_2 is driven by an underlying continuous-time, discrete-state, process $\{L(t) : t \geq 0\}$ of state space $\{-2, -1, 0, 1\}$. More precisely, $L(t) = 1$ ($L(t) = 0$) when Q_2 is at L_1 (resp. L_2) at time t, and $L(t) = -1$ ($L(t) = -2$) when Q_2 switches from L_1 to L_2 (L_2 to L_1). Without loss of generality, let $L(0) = 1$.

During the availability of the server at Q_1 and Q_2, the server may alternate between service and idle periods depending on whether customers are present. It is worth pointing that the term customer throughout this paper will designate packet. When the server is active at the end of a visit of Q_2 to L_1 or L_2, service will be preempted. At the beginning of the next visit of Q_2, the service time will be re-sampled according to $S_i(\cdot)$. This discipline is commonly referred to as *preemptive-repeat-random*. Let $N_i(t)$ denote the number of customers in Q_i, $i = 1, 2$, at time t. Assume $N_i(0) = 0$, $i = 1, 2$.

A word on the notation. $\mathbf{1}_{\{A\}}$ will designate the indicator function of event A ($\mathbf{1}_{\{A\}} = 1$, if A is true, and 0 otherwise), rv will mean random variable, LST Laplace-Stieltjes Transform and p.g.f. probability generating function. Given a rv X, $X(t)$ will denote its distribution function, $\tilde{X}(s)$ its LST.

Our objective is to analyze the sojourn time of a customer in the whole system and at the individual queues Q_1 and Q_2. First, we discuss the stability of the system. Second, we will analyze the sojourn time at Q_1. The model for Q_1 in isolation boils down to a single-vacation model or an on-off server model. Finally, we analyze the sojourn time at Q_2. To this end, we determine the joint queue-length probabilities at a specific instant. These probabilities can be related to the time-equilibrium probabilities for the tandem system. Applying Little's law, the mean sojourn time is then readily found.

2.2 Stability condition

The tandem model is stable if each customer in the system can be served in a finite period of time. We must consider stability on a per-queue basis as service capacity cannot be exchanged between the queues. We say that the system is stable if and only if all the queues in the system are stable.

Let a cycle define the time that separates two consecutive visits to a queue. Due to the independence assumption on our rv's, cycle lengths are iid, with generic rv $C := X^{L_1} + X^{L_2} + C^{1,2} + C^{2,1}$. For an individual queue to be stable, we must have that on average the number of customer arrivals per cycle is smaller than the number of customers that can be served at most per cycle. The latter random variable for Q_i will be denoted by $N^i_{\max}$, $i = 1, 2$, and is geometrically distributed (due to the exponential visit times and preemptive-repeat-random discipline), i.e.

$$\mathbb{P}(N^i_{\max} = k) = p_i(1 - p_i)^k,\ k = 0, 1, 2, \dots\ ,$$

where $p_i = \mathbb{P}(\text{service is preempted at } Q_i) = 1 - \tilde{S}_i(\alpha_i)$, $i = 1, 2$. Thus, the stability condition for Q_i, $i = 1, 2$, reads

$$\rho_i := \frac{\mathbb{E}[\text{arrivals per cycle to } Q_i]}{\mathbb{E}[N^i_{\max}]} = \lambda \mathbb{E}[C] \frac{1 - \tilde{S}_i(\alpha_i)}{\tilde{S}_i(\alpha_i)} < 1, \tag{1}$$

where

$$\mathbb{E}[C] = \frac{1}{\alpha_1} + \frac{1}{\alpha_2} + c^{1,2} + c^{2,1}. \tag{2}$$

Notice that under stability, on average, the arrival rate to Q_2 equals that to Q_1.

2.3 Queue one

Let us recall that we assumed that the server visit process is autonomous and that service is according to the preemptive-repeat-random discipline. It is then easily seen that Q_1 in isolation is an M/G/1 queue with On-Off server with arrival rate λ, service time rate β_1, exponential on-period X^{L_1} with rate α_1, and off-period R^{off} equal to the switch-over times plus the server visit time to Q_2 at L_2, i.e.,

$$R^{off} = C^{1,2} + C^{2,1} + X^{L_2}. \tag{3}$$

By a renewal reward argument, the probability, P_{On}, that the server is on equals

$$P_{On} = \frac{1}{\alpha_1 \mathbb{E}[C]}, \tag{4}$$

and $P_{Off} = 1 - P_{On}$.

The M/G/1 queue with On-Off server has been extensively studied in the literature (see, e.g., [7, 18, 19]). Below, we provide an alternative derivation of the LST of the waiting time of a customer at Q_1 using the idea of waiting-time decomposition. From there, we will deduce the sojourn time and also the time-equilibrium distribution of the number of customers in Q_1. Finally, we determine the p.g.f. of the number of customers at the end of an off-period which we will need later for the analysis of Q_2.

2.3.1 Sojourn time in queue one

Let A_k denote the k-th arriving customer at Q_1. Define the effective service time, S_k^{eff}, of A_k to be the time period which starts when A_k receives service for the first time at Q_1 and which ends when A_k departs from Q_1, i.e.,

$$S_k^{eff} = S_k^* + \sum_{i=1}^{M} C_i, \tag{5}$$

where S_k^* is the conditional service requirement S_k of A_k given that S_k is not interrupted, i.e., $S_k < X^{L_1}$, C_i is a cycle duration distributed as C, and M is the (random) number of off-periods before service completion of A_k. Since the visit time of Q_2 at L_1, X^{L_1}, is exponential and independent of the service requirement, the distribution of M is geometric with parameter $p_1 = P(S_k > X^{L_1}) = 1 - \tilde{S}_1(\alpha_1)$. Note that $\{C_n^{1,2}, C_n^{2,1}, X_n^{L_1}, X_n^{L_2}\}_{n\geq 0}$ are mutually independent. In particular, the lengths of the on-periods are mutually independent and also independent of the lengths of off-periods. Further, the service discipline is preemptive-repeat-random, so that S_k^* is independent of M and C_i, $i = 1, \ldots, M$. As a consequence, we can write the LST of the effective service time as follows.

$$\tilde{S}_k^{eff}(s) = \frac{(\alpha_1 + s)\cdot \tilde{S}_1(\alpha_1+s)}{(\alpha_1+s) - \alpha_1(1-\tilde{S}_1(\alpha_1+s))\cdot \frac{\alpha_2 \tilde{C}^{1,2}(s)\tilde{C}^{2,1}(s)}{\alpha_2+s}}\ , \tag{6}$$

where Re$(s) \geq 0$.

Let $W_{M/G/1}$ denote the waiting time in the $M/G/1$ queue with arrival rate λ and service time S_k^{eff}. The Pollaczek-Khinchine formula for the LST of $W_{M/G/1}$ reads [30, P. 386],

$$\tilde{W}_{M/G/1}(s) = \frac{\big(1 + \lambda(\tilde{S}_k^{eff})'(0)\big)s}{s - \lambda + \lambda \tilde{S}_k^{eff}(s)}, \tag{7}$$

where $(\tilde{S}_k^{eff})'(0)$ is the first derivative of $\tilde{S}_k^{eff}(s)$ at the origin.

The time that the service of an arriving customer to Q_1, say A_k, starts depends on the state of the system at that time. If A_k arrives to a non-empty Q_1, then its service starts at the instant that A_{k-1}, $k \geq 1$, departs from the queue. If A_k arrives to an empty Q_1 and the server is on, then its service starts instantaneously. However, if A_k arrives to an empty Q_1 and the server is off, then it has to wait a residual off time, R_e^{off}, before its service will start. Thus, Q_1 can be seen as an On-Off queue with an exceptional first service time. It is known that for such a queue the waiting time, W_1, of a customer in Q_1 can be decomposed as follows [18, 19]

$$W_1 = W_{M/G/1} + R_e^{off}\mathbf{1}_{\{\text{server off}\}}, \tag{8}$$

where $W_{M/G/1}$ and R_e^{off} are independent. The LST of W_1 is readily found by conditioning on the server's state upon a customer's arrival, which yields

$$\tilde{W}_1(s) = \tilde{W}_{M/G/1}(s)(P_{On} + P_{Off}\tilde{R}_e^{off}(s)), \tag{9}$$

where P_{On} and $\tilde{W}_{M/G/1}(s)$ are given in (4) and (7) respectively, and

$$\tilde{R}_e^{off}(s) = \frac{1 - (\alpha_2/(\alpha_2+s))\cdot \tilde{C}^{1,2}(s)\cdot \tilde{C}^{2,1}(s)}{(c^{1,2} + c^{2,1} + 1/\alpha_2)\cdot s}. \tag{10}$$

The LST of the sojourn time, $D_1 \triangleq W_1 + S_k^{eff}$, of A_k at Q_1 then yields

$$\tilde{D}_1(s) = \tilde{W}_1(s)\tilde{S}_k^{eff}(s). \tag{11}$$

2.3.2 Number of customers in Q_1

The arrival process to Q_1 is Poisson with rate λ. Thus, it follows that the p.g.f. of N_1, which we denote by $F_1(\cdot)$, can be expressed as function of $\tilde{D}_1(\cdot)$ using the so-called functional form of Little's law (see [31] for a general proof for FIFO queues with non-anticipating arrivals) as follows

$$F_1(z) = \tilde{D}_1\big(\lambda(1-z)\big), \quad |z| \leq 1. \tag{12}$$

Next, we will determine $F^{\{-2,1\}}(\cdot)$, the p.g.f. of the number of customers at the end of an off-period, i.e., at the transition of $L(t)$ from -2 to 1. This function will be required later in the analysis for the second queue. To this end, we first compute the p.g.f. of the number of customers in Q_1 given that the server is on. Let T_o denote a random time during an off-period, T_s denote the start time of this off-period, and let $A(x,y)$ denote the total number of arrivals during $(x,y]$ with $y > x$. Thus, $N_1(T_o) = N_1(T_s) + A(T_s, T_o)$. Conditioned on the event 'server on', we may treat the epochs at which the server switches off (and immediately on) as a Poisson arrival stream of batches. Due to the PASTA property, these batches see time average behavior upon arrival. As the system observed by the arriving batches is exactly the system as observed by the server that departs, we have that

$$\begin{aligned} E[z^{N_1}|\text{server off}] &= E\big[z^{N_1(T_s)}\big]E[z^{\Psi}] \\ &= E[z^{N_1}|\text{server on}]E[z^{\Psi}], \end{aligned} \tag{13}$$

where Ψ is the number of arrivals to Q_1 during the age of the off-period. The latter being equal in distribution to the residual time of an off-period. In other words, Ψ is the number of Poisson arrivals of rate λ during R_e^{off}. Thus, conditioned on the state of Q_1's server at a random time, the conditional p.g.f. of N_1 can be written as

$$\begin{aligned} E[z^{N_1}|\text{server on}] &= \frac{F_1(z)}{P_{On} + E[z^{\Psi}]P_{Off}} \\ &= \tilde{W}_{M/G/1}\big(\lambda(1-z)\big)\tilde{S}_k^{eff}\big(\lambda(1-z)\big). \end{aligned} \tag{14}$$

Finally, we can conclude that

$$F^{\{-2,1\}}(z) = E[z^{N_1}|\text{server on}]\tilde{R}^{off}\big(\lambda(1-z)\big)\ . \tag{15}$$

2.4 Queues in tandem

2.4.1 Joint queue-length probabilities at the end of a server visit

In this section, we will determine the queue-length distribution at the end of a server visit at each queue of the tandem of two queues. The analysis builds on the work of Eisenberg [9] and involves setting up an iterative scheme. This iterative approach was introduced by Leung [21] for the study of a probabilistically-limited polling model. Later, this model was extended in [22] to a time-limited polling model and in [6] for a time-limited model in which the server remains at a queue even if it becomes empty. A key role in the iterative scheme is played by the (auxiliary) p.g.f.'s $\phi_k(\mathbf{z})$ and $\phi_k^s(\mathbf{z})$ for $\mathbf{z} := (z_1, z_2)$, which will be explained below. In the final step of the iteration scheme $\gamma_i(\mathbf{z})$, the p.g.f. of the queue-length distribution at the end of a server visit to Q_i, is obtained as a function of $\phi_k^s(\mathbf{z})$.

We consider a tagged queue i and we will leave out the subscript and superscript i whenever it does not lead to ambiguity. Let us introduce the concept of a *service period*. We let a service period be a segment of a visit time such that all service periods together form exactly a visit time. The first service period of a visit starts when the server arrives to the queue. This period ends with either an interruption (due to the departure of the server) or a service completion, whichever occurs first. In the latter case, a second service period will start and this process continues until finally an interruption occurs. Each service period, except for the final service period of a visit, comprises exactly one successfully completed service. Further notice that there need not always be customers present at the start of a service period.

Let us denote by N_k^i the number of customers at the end of the kth service period at Q_i and by κ_i the number of service periods of a visit time of Q_i. We may then define for $k \geq 1$

$$\phi_k^i(\mathbf{z}) := \mathbb{E}[z^{N_k^i} \mathbf{1}_{\{\kappa_i > k\}}] \ . \tag{16}$$

That is, $\phi_k^i(\mathbf{z})$ is the p.g.f. of the number of customers at the queues at the end of the kth service period at Q_i and service period k is not the final service period (i.e., service period k ends with a successful service completion, and service period $k+1$ will occur). Similarly, we define for $k \geq 1$

$$\phi_k^{s,i}(\mathbf{z}) := \mathbb{E}[z^{N_k^i} \mathbf{1}_{\{\kappa_i = k\}}] \ . \tag{17}$$

That is, $\phi_k^{s,i}(\mathbf{z})$ is the p.g.f. of the number of customers at the queues at the end of the kth service period at Q_i and k is the final service period (i.e., service period k will be interrupted, and service period $k+1$ will not occur). Finally, we denote by $\phi_0^i(\mathbf{z})$ the p.g.f. of the number of customers present at the beginning of a visit to Q_i. Let $N(T)$ the number of arrivals during a random period T, I_1 the (exponential) interarrival time of customers at Q_1, and $C^{i,j}(\mathbf{z})$ be the p.g.f. of the number of arrivals during a switch-over time from Q_i to Q_j. Then, by analogy with the results of De Haan et al. [6] for a time-limited polling system, $\phi_k^i(\mathbf{z})$ and $\phi_k^{s,i}(\mathbf{z})$, $i = 1, 2,\ k = 1, 2, \ldots$, are recursively given by

$$\begin{aligned}
\phi_k^1(\mathbf{z}) &= \phi_{k-1}(\mathbf{z})\mid_{z_1=0} \cdot \Big(\mathbb{E}[\mathbf{z}^{N(I_1)} \mathbf{1}_{\{X^{L_1} > I_1\}}] \\
&\times \ \mathbb{E}[\mathbf{z}^{N(S_1)} \mathbf{1}_{\{X^{L_1} > S_1\}}] \cdot z_2\Big) + \Big(\phi_{k-1}(\mathbf{z}) \\
&- \ \phi_{k-1}(\mathbf{z})\mid_{z_1=0}\Big) \cdot \mathbb{E}[\mathbf{z}^{N(S_1)} \mathbf{1}_{\{X^{L_1} > S_1\}}] \cdot \frac{z_2}{z_1} \ , \qquad (18) \\
\phi_k^2(\mathbf{z}) &= (\phi_{k-1}(\mathbf{z}) - \phi_{k-1}(\mathbf{z})\mid_{z_2=0}) \cdot \frac{\mathbb{E}[\mathbf{z}^{N(S_2)} \mathbf{1}_{\{X^{L_2} > S_2\}}]}{z_2} \ , \qquad (19)
\end{aligned}$$

and

$$\begin{aligned}
\phi_k^{s,1}(\mathbf{z}) &= \phi_{k-1}(\mathbf{z})\mid_{z_1=0} \cdot \Big(\mathbb{E}[\mathbf{z}^{N(X^{L_1})} \mathbf{1}_{\{X^{L_1} < I_1\}}] \\
&+ \ z_1 \mathbb{E}[\mathbf{z}^{N(I_1)} \mathbf{1}_{\{X^{L_1} > I_1\}}] \mathbb{E}[\mathbf{z}^{N(X^{L_1})} \mathbf{1}_{\{X^{L_1} < S_1\}}]\Big) \\
&+ \ (\phi_{k-1}(\mathbf{z}) - \phi_{k-1}(\mathbf{z})\mid_{z_1=0}) \\
&\times \ \mathbb{E}[\mathbf{z}^{N(X^{L_1})} \mathbf{1}_{\{X^{L_1} < S_1\}}], \qquad (20) \\
\phi_k^{s,2}(\mathbf{z}) &= \phi_{k-1}(\mathbf{z})\mid_{z_2=0} \cdot \mathbb{E}[\mathbf{z}^{N(X^{L_2})}] + \Big(\phi_{k-1}(\mathbf{z}) \\
&- \ \phi_{k-1}(\mathbf{z})\mid_{z_2=0}\Big) \cdot \mathbb{E}[\mathbf{z}^{N(X^{L_2})} \mathbf{1}_{\{X^{L_2} < S_2\}}] \ , \qquad (21)
\end{aligned}$$

where

$$\begin{aligned}
\phi_0^i(\mathbf{z}) &= \gamma^{3-i}(\mathbf{z}) C^{3-i,i}(\mathbf{z}) \ , \\
\mathbb{E}[\mathbf{z}^{N(I_1)} \mathbf{1}_{\{X^{L_1} > I_1\}}] &= \frac{\lambda}{\lambda + \alpha_1} \ , \\
\mathbb{E}[\mathbf{z}^{N(S_i)} \mathbf{1}_{\{X^{L_i} > S_i\}}] &= \tilde{S}_i(\alpha_i + \lambda(1 - z_1)) \ , \\
\mathbb{E}[\mathbf{z}^{N(X^{L_1})} \mathbf{1}_{\{X^{L_1} < I_1\}}] &= \frac{\alpha_1}{\lambda + \alpha_1} \ ,
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[\mathbf{z}^{N(X^{L_i})} \mathbf{1}_{\{X^{L_i} < S_i\}}] &= \alpha_i \cdot \frac{1 - \tilde{S}_i(\alpha_i + \lambda(1 - z_1))}{\alpha_i + \lambda(1 - z_1)} \ , \\
\mathbb{E}[\mathbf{z}^{N(X^{L_2})}] &= \frac{\alpha_2}{\alpha_2 + \lambda(1 - z_1)} \ .
\end{aligned}$$

Equation (18) can be explained by the following observations. First, the length of the kth service period (and thus also the number of arriving customers) depends on whether at least one customer was present at the end of the previous service period. This explains why the equation consists of two parts. Second, the number of customers at all queues at the end of a service period is equal to the number present at the end of the previous service period plus the ones that arrived during the present service period. Equation (19) consists only of one part due to the fact that once Q_2 is empty no customers will be served anymore during that visit. Along the same lines as Eq. (18), Eqs. (20) and (21) are derived where it should be noticed that the number of arrivals depends on whether a service period is interrupted or not. Finally, we note that $\phi_0^i(1) = 1$, while $\phi_k^i(1) \leq 1$, for all $k = 1, 2, \ldots$, since the kth period completion may not occur at all during a visit to Q_i.

Notice that there is one-to-one relationship between a visit completion and the end of a final service period. Therefore, we can write for the number of customers at the queues at the end of a server visit to Q_i

$$\gamma^i(\mathbf{z}) = \mathbb{E}[z^{N_{\kappa_i}^i}] = \sum_{k=1}^{\infty} \mathbb{E}[z^{N_k^i} \mathbf{1}_{\{\kappa_i = k\}}] = \sum_{k=1}^{\infty} \phi_k^{s,i}(\mathbf{z}) \ . \tag{22}$$

We set up an iterative scheme to obtain $\gamma^i(\mathbf{z})$ numerically. The scheme is constructed in terms of Discrete Fourier Transforms (DFTs) as these appear more convenient for computational purposes. To this end, we replace z_i, $\forall\, i$, in the expressions above by $\omega_i^{k_i}$, where $\omega_i = exp(-2\pi I/J_i)$, so that all expressions become functions of $\mathbf{k} = (k_1, k_2)$. Here I is the imaginary unit and J_i refers to the number of discrete points used for Q_i to determine the joint probabilities. These probabilities that will eventually follow are exact for $J_i \to \infty$, $\forall\, i$. However, the strength of the approach is that in general the probabilities are already close to the exact probabilities for small values of J_i. The pseudo-code of the iterative scheme is presented in Table 1. Notice that we start initially with an empty system. The standard values for the convergence parameters that have been used are $\epsilon = 10^{-6}$ and $\delta = 10^{-9}$. Finally, via the Inverse Fourier Transform, the joint queue-length probabilities at visit completion instants $\gamma^i_{\mathbf{n}}$ are found. These probabilities are only exact for $J_i \to \infty$, $i = 1, 2$ but the strength of the approach is that in general the probabilities are already close to the exact values for small values of J_i. However, it should be noted that when the system load increases, these values J_i must typically be increased to guarantee the accurate computation of the probabilities. Thus, this iterative approach appears mainly applicable to systems with a light to moderate load.

$\gamma^{i_0}(\mathbf{k}) = 1$, $\forall\, i_0$, $\forall\, \mathbf{k}$; (start with an empty system)
FOR $i_1 = 1, 2$
set $i_2 := i_1$;
REPEAT
$\hat{\gamma}^{i_2}(\mathbf{k}) = \gamma^{i_2}(\mathbf{k})$, $\forall\, \mathbf{k}$;
set $j := 0$;
set $\phi_0(\mathbf{k}) = \gamma^{3-i_2}(\mathbf{k}) \cdot C^{3-i_2,i_2}(\mathbf{k})$;
REPEAT
set $j := j + 1$;
compute $\phi_j^{i_2}(\mathbf{k})$, $\forall\, \mathbf{k}$, using (18) and (19);
compute $\phi_j^{s,i_2}(\mathbf{k})$, $\forall\, \mathbf{k}$, using (20) and (21);
compute $\gamma^{i_2}(\mathbf{k}) = \sum_{l=1}^{j} \phi_l^{s,i_2}(\mathbf{k})$, $\forall\, \mathbf{k}$;
UNTIL $1 - \mathrm{Re}(\gamma^{i_2}(\mathbf{0})) < \delta$
set $i_2 := \mathrm{MOD}(i_2, 2) + 1$;
UNTIL $\lvert \mathrm{Re}(\gamma^{i_1}(\mathbf{k})) - \mathrm{Re}(\hat{\gamma}^{i_1}(\mathbf{k}))\rvert < \epsilon$, $\forall\, \mathbf{k}$
END (FOR)

Table 1: Pseudo-code of iterative scheme for determining $\gamma^i(\mathbf{k}), \forall\, i$.

2.4.2 Mean sojourn time

The sojourn time is related to the time-equilibrium queue-length probabilities. These probabilities can be obtained from the queue-length probabilities at visit completion instants due to exponential visit times. We determine these probabilities by conditioning on the position of the server. Notice that the server is either at some queue or switching from one queue to another. Using the same arguments as in Sect. 2.3.2 above Eq. (13), we have that a departing server observes the system in steady-state conditioned on the queue it departs from. Let us further denote the p.g.f. of the number of customers present at a random instant during a switch-over time from Q_{3-i} to Q_i by $C_R^i(\mathbf{z})$. It can readily be found that

$$C_R^i(\mathbf{z}) = \gamma^i(\mathbf{z}) \cdot \frac{1 - \tilde{C}^{3-i,i}\big(\lambda(1-z_1)\big)}{c^{3-i,i} \cdot \lambda(1-z_1)}. \qquad (23)$$

Hence, by conditioning on the position of the server, we may write for $P(\mathbf{z}) := \mathbb{E}[z_1^{N_1} z_2^{N_2}]$, the joint p.g.f. of the time-equilibrium queue lengths,

$$P(\mathbf{z}) = \frac{1}{\mathbb{E}[C]} \sum_{i=1}^{2} \left(\gamma^i(\mathbf{z}) \cdot \frac{1}{\alpha_i} + C_R^i(\mathbf{z}) \cdot c^{3-i,i} \right). \qquad (24)$$

The mean queue length at Q_i, $\mathbb{E}[N_i]$, is then given by

$$\mathbb{E}[N_i] = \sum_{n_1 \ge 0} \sum_{n_2 \ge 0} \mathbb{P}(N_1 = n_1, N_2 = n_2) n_i, \qquad (25)$$

where the probabilities $\mathbb{P}(N_1 = n_1, N_2 = n_2)$ follow immediately from $P(\mathbf{z})$. The mean sojourn time is related to the queue length via Little's law, which then finally provides us with

$$\mathbb{E}[D_i] = \mathbb{E}[N_i]/\lambda. \qquad (26)$$

Remark 1. *We note that using the distributional form of Little's law also higher moments can be obtained for the end-to-end sojourn time. However, this form cannot be applied to the individual sojourn time at Q_2 here, since the arrival process to Q_2 does not satisfy the non-anticipating property [31].*

3. APPROXIMATION

In this section, we present an approximation for $\tilde{D}_2(s)$, the LST of the sojourn time of a customer in the mobile queue Q_2, so that we may also deal with the situations in which the exact approach is no longer computationally feasible. We consider the workload process in Q_2 when $L(t) = 0$, i.e. Q_2 is served. This will be done under the additional assumption that the service time requirements are exponentially distributed at both queues. It turns out that this process corresponds to the workload process in an $M/M/1$ with batch arrivals. The sojourn time of a customer in Q_2 then equals the sum of the waiting time of the batch in this corresponding $M/M/1$ system, the service times of all customers in its batch served up to and including this customer and the time the customer is at Q_2 but $L(t) \neq 0$, i.e. Q_2 is not served. We emphasize that in this case both the preemptive-repeat-random and preemptive-resume disciplines are stochastically identical. For the sake of simplicity, in the following we will consider the preemptive-resume discipline.

3.1 The workload in queue two

To study the workload process at Q_2, we split the time in disjoint intervals which begin at the time instants that the $L(t)$-process jumps from state -2 to 1 (i.e., the start of an on-period at Q_1). Denote the starting points of these intervals by $\{Z_n, n = 1, 2, \cdots\}$, with, by convention, $Z_1 = 0$. Let the n-th cycle of $L(t)$ denote the time interval $[Z_n, Z_{n+1}[$. The duration of the n-th cycle is $Z_{n+1} - Z_n = X_n^{L_1} + C_n^{1,2} + X_n^{L_2} + C_n^{2,1}$, where $X_n^{L_1}$ is the duration of time that customers can arrive at Q_2 ($L(t) = 1$)) and $X_n^{L_2}$ is the duration of time that customers can leave Q_2 ($L(t) = 0$)) during this n-th cycle. Let $V(t)$ denote the workload (i.e., virtual waiting time) of Q_2 present at time t. Without loss of generality, we assume that $V(t)$ is left-continuous, i.e., arrivals are not counted as being in the system until (just) after they arrive. A sample path of the evolution of $V(t)$ as function of $L(t)$ is shown in Figure 2.

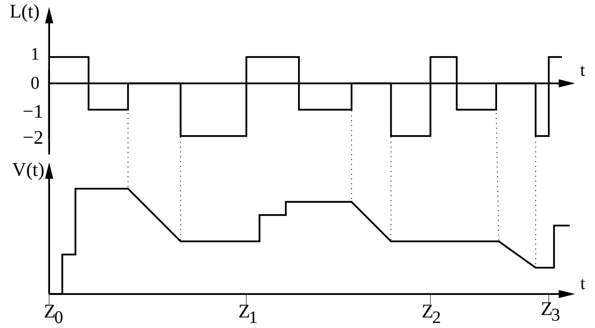

Figure 2: Evolution of $L(t)$ and workload $V(t)$ of queue Q_2.

Let W_n^B denote the workload present in Q_2 at time Z_n. Based on the evolution of $L(t)$, it is easily seen that

$$W_{n+1}^B = \left(W_n^B + \sum_{i=1}^{K_n} S_{2,i} - X_n^{L_2} \right)^+, \quad n \geq 0, \tag{27}$$

where $(.)^+ = \max(.,0)$, K_n is the total number of arrivals to Q_2 (or departures from Q_1) during $X_n^{L_1}$ and $S_{2,i}$ is the service requirement of a customer in Q_2. Note that $S_{2,i}$ is independent of $X_n^{L_2}$ and that K_n depends on $N_1(Z_n)$, the number of customers at Q_1 at time Z_n. Therefore, the rvs K_n, $n = 1, 2, \cdots$ are correlated. For the sake of model tractability it is assumed in the sequel that K_n, $n = 1, 2, \cdots$ are iid and also independent of $\{X_m^{L_2} : m \leq n\}$. By these assumptions, Eq. (27) also represents the relation between the workload seen by the first customer of the n-th batch and of the $(n+1)$-th batch in a queue with Poisson batch arrivals with rate α_2, independent batch size K_n, and exponential service requirement with rate β_2. It is well known that this queue is stable when

$$-\alpha_2 G'(0) = \frac{\alpha_2}{\beta_2} \mathbb{E}[K_n] < 1. \tag{28}$$

Note that this condition (28) is equivalent to the condition in (1) for Q_2. Furthermore, the LST of the steady-state distribution of W_n^B can be written as

$$\tilde{W}^B(s) = \left(1 + \alpha_2 G'(0)\right) \frac{s}{s - \alpha_2 + \alpha_2 G(s)}, \tag{29}$$

where $G(s) := E\left[e^{-s \sum_{i=1}^{K_n} S_{2,i}}\right]$. By conditioning on K_n, we find that

$$G(s) \quad = \quad \mathbb{E}\left[\left(\frac{\beta_2}{\beta_2 + s}\right)^{K_n}\right]. \tag{30}$$

Finally, let $\tilde{V}^j(s)$ denote the LST of the workload seen by the jth customer within a batch upon arrival including the work brought in by himself. Since the service requirement of customers is independent of the workload present in the queue upon arrival and its distribution is exponential with rate β_2, $\tilde{V}^j(s)$ reads

$$\tilde{V}^j(s) \quad = \quad \tilde{V}^{j-1}(s) \frac{\beta_2}{\beta_2 + s}, \quad j = 1, 2, \ldots, \tag{31}$$

with $\tilde{V}^0(s) = \tilde{W}^B(s)$. Moreover, since K_n are iid rvs, $\mathbb{P}(J = j)$, the probability that a customer is the j-th customer within the batch is equal to the fraction of customers who are j-th arrival in their own batch, which gives

$$\mathbb{P}(J = j) = \frac{\mathbb{P}(K_n \geq j)}{\mathbb{E}[K_n]}. \tag{32}$$

Removing the condition on the customer position in a batch, the LST of the sojourn time of an arbitrary customer in the batch arrival queue is given by

$$\tilde{V}(s) \quad = \quad \beta_2 \tilde{W}^B(s) \frac{1 - G(s)}{s \mathbb{E}[K_n]}. \tag{33}$$

Thus, it remains to compute $\mathbb{E}[z^{K_n}]$ in order to find $\tilde{W}_B(s)$. In the following section, we will compute the closed form of $\mathbb{E}[z^{K_n}]$ by using the matrix-geometric approach.

3.2 The p.g.f. of the batch size distribution

As remarked in the previous section, K_n is the total number of departures from Q_1 during the n-th cycle and depends on the queue length of Q_1 at time Z_n. To compute the p.g.f. of K_n, we first assume that Q_1 has a limited queue of $M - 1$ customers. This queue is denoted by Q_1^M. Later, we will let M tend to infinity to get the final results.

As we need the arriving batch size distribution in steady state, we assume that Q_1^M is in steady state at time Z_n. The probability that there are i customers in Q_1^M at Z_n is denoted by $b_M(i)$. Under the assumption that the unlimited Q_1 is stable, $\lim_{M\to\infty} b_M(i) = b(i)$ with $\sum_{i\geq 0} b(i) z^i = F^{\{-2,1\}}(z)$ (see Eq. (15)). Let $b_M = (b_M(0), \cdots, b_M(M-1))$ denote the steady-state distribution of the finite capacity Q_1^M.

Let $(N_1(t), D(t))$ denote the two dimensional, continuous time process with discrete state space $\{0, 1, \cdots, M-1\} \times \{0, 1, \cdots\} \cup \{(M, 0)\}$, where $N_1(t)$ represents the number of customers in Q_1 at time t, and $D(t)$ the number of departures from Q_1 until t. $(M, 0)$ is an absorbing state. We refer to this absorbing Markov chain by **AMC**. The absorption of **AMC** occurs when the server leaves the queue which happens with rate α_1. By setting the probability that the initial state of **AMC** at $t = 0$ is $(i, 0)$ to $b_M(i)$, the probability that the absorption of **AMC** occurs from one of the states $\{(i, k) : i = 0, 1, \cdots, M-1\}$ equals the steady-state batch size distribution $\mathbb{P}(K_n = k)$. The transition state diagram of **AMC** is shown in Figure 3.

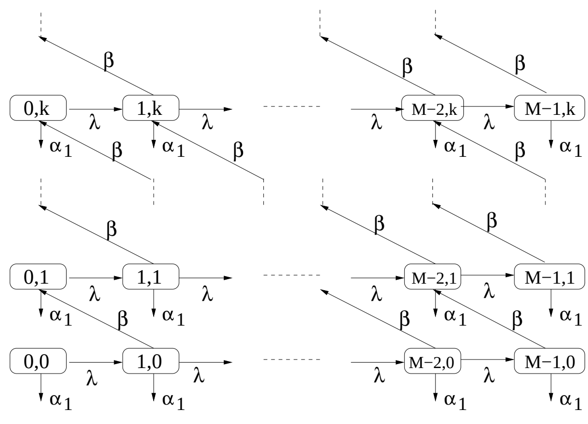

Figure 3: Transition state diagram of AMC.

Now we focus on $\mathbb{P}(K_n = k)$. From Figure 3, we readily seen that the transition matrix $\mathbf{P}$ of **AMC** can be written as

$$\mathbf{P} = \left(\begin{array}{c|c} \mathbf{Q} & \mathbf{R} \\ \hline \mathbf{0} & 0 \end{array} \right),$$

where $\mathbf{Q}$ is an upper bidiagonal square block matrix, $\mathbf{R} = (\alpha_1, \ldots, \alpha_1)^T$, and $\mathbf{0}$ is the row vector with all zero entries. The blocks of $\mathbf{Q}$'s diagonal are all equal to $\mathbf{A}$, the M-by-M bidiagonal matrix $\mathbf{A}$ with diagonal $(-\lambda - \alpha_1, -\lambda - \alpha_1 - \beta_1, \ldots, -\lambda - \alpha_1 - \beta_1, -\alpha_1 - \beta_1)$ and with upper-diagonal $(\lambda, \ldots, \lambda)$. The blocks of $\mathbf{Q}$'s upper-diagonal are all equal to $\mathbf{B}$, the M-by-M lower-diagonal matrix with lower-diagonal $(\beta_1, \ldots, \beta_1)$.

In the sequel, $\mathbb{P}(K_n = k)$ is derived as function of the inverse of $\mathbf{Q}$, that is readily obtained as

$$\mathbf{Q}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} & \mathbf{U}_{0,1} & & \cdots & & \\ & \ddots & \ddots & & & \\ & & \mathbf{A}^{-1} & \mathbf{U}_{m,m+1} & \cdots & \\ & & & \ddots & \ddots & \end{pmatrix}$$

where $\mathbf{U}_{m,l} = \big(- \mathbf{A}^{-1}\mathbf{B}\big)^{l-m}\mathbf{A}^{-1}$ for $m \geq 0$ and $l \geq m$. Note that the matrix $\mathbf{A}$ is invertible since it is upper-bidiagonal with strictly negative diagonal entries.

From the theory of absorbing Markov chains, given that the initial state vector of **AMC** is b_M, the probability that the absorption occurs at one of the states $\{(i,k) : i = 0, 1, \cdots, M-1\}$ is given by (see, e.g., [11], [12, Theorem 11.9])

$$\mathbb{P}(K_n = k) = -\alpha_1 b_M (\mathbf{U}_{0,k}) e = -\alpha_1 b_M \big(- \mathbf{A}^{-1}\mathbf{B}\big)^k \mathbf{A}^{-1} e. \tag{34}$$

where e denote the M-dimensional column vector with all entries equal to one. Thus, the LST of K_n reads

$$E_M[z^{K_n}] = -\alpha_1 b_M (\mathbf{A} + z\mathbf{B})^{-1} e, \tag{35}$$

where $|z| \leq 1$. Therefore, it remains to find $(\mathbf{A} + z\mathbf{B})^{-1}$.

Now, define $\mathbf{Q}(z) := (\mathbf{A} + z\mathbf{B})$, let $u^T = (1, 0, \ldots, 0)$ and let $v^T = (0, \ldots, 0, 1)$. Observe that $\mathbf{Q}(z) = \mathbf{T}(z) + \beta_1 u u^T + \lambda v v^T$, where $\mathbf{T}(z)$ is the M-by-M tridiagonal Toeplitz matrix with diagonal entries equal to $(-\lambda - \beta_1 - \alpha_1)$, upper-diagonal entries equal to λ, and lower-diagonal entries $z\beta_1$. Let t^*_{ij} denote the (i,j)-entry of $\mathbf{T}^{-1}(z)$. By applying the Sherman-Morrison formula [24, page 76] we find that the (i,j)-entry of $\mathbf{Q}^{-1}(z)$ gives

$$q^*_{ij} = m_{ij} - \lambda \frac{m_{iM} m_{Mj}}{1 + \lambda m_{MM}}, \tag{36}$$

with

$$m_{ij} = t^*_{ij} - \beta_1 \frac{t^*_{i1} t^*_{1j}}{1 + \beta_1 t^*_{11}}, \tag{37}$$

for $i = 1, \ldots, M$ and $j = 1, \ldots, M$.

The inverse of a tridiagonal Toeplitz matrix has been computed in closed-form (see [8, Sec. 3.1]). Following that same approach, we obtain

$$t^*_{ij} = \left\{ \begin{array}{ll} -\frac{(r_1^i - r_2^i)(r_1^{M+1-j} - r_2^{M+1-j})}{\lambda (r_1 - r_2)(r_1^{M+1} - r_2^{M+1})} & , \; i \leq j \leq M \\ \frac{(r_1^{-j} - r_2^{-j})(r_1^{M+1} r_2^i - r_2^{M+1} r_1^i)}{\lambda (r_1 - r_2)(r_1^{M+1} - r_2^{M+1})} & , \; j \leq i \leq M \end{array} \right\} \tag{38}$$

where

$$r_{1,2} = \frac{(\lambda + \beta_1 + \alpha_1) \mp \sqrt{(\lambda + \beta_1 + \alpha_1)^2 - 4\lambda\beta_1 z}}{2\lambda}. \tag{39}$$

We take $|r_1| < |r_2|$. Note that $|r_1| < 1 < |r_2|$.

Inserting (36)-(37) into (35) yields that

$$\begin{aligned} E_M[z^{K_n}] \;&=\; -\alpha_1 \sum_{i=1}^{M} b_M(i-1) \sum_{j=1}^{M} \Big[t^*_{ij} - \frac{\beta_1 t^*_{i1} t^*_{1j}}{1 + \beta_1 t^*_{11}} \\ &\quad - \frac{\lambda m_{iM}}{1 + \lambda m_{MM}} \Big(t_{Mj} - \frac{\beta_1 t^*_{M1} t^*_{1j}}{1 + \beta_1 t^*_{11}} \Big) \Big]. \end{aligned} \tag{40}$$

Thus, it remains to let $M \to \infty$ in (40) in order to find $\mathbb{E}[z^{K_n}]$. It is readily seen that

$$\begin{aligned} \lim_{M\to\infty} t_{M,M-j} &= -\frac{1}{\lambda r_2} \lim_{M\to\infty} r_1^j, \\ \lim_{M\to\infty} m_{M-i,M} &= \lim_{M\to\infty} t_{M-i,M} = -\frac{1}{\lambda} \lim_{M\to\infty} r_2^{-(i+1)}, \\ \lim_{M\to\infty} t_{1j} &= -\frac{1}{\lambda} r_2^{-j}, \\ \lim_{M\to\infty} t_{i1} &= \frac{1}{\lambda r_1 r_2} \lim_{M\to\infty} \frac{r_1^{M+1}}{r_2^{M-i+1}} - r_1^i. \end{aligned}$$

Some easy but technical calculus shows that the following limit is equal to zero

$$\lim_{M\to\infty} \alpha_1 \sum_{i=1}^{M} b_M(i-1) \sum_{j=1}^{M} \frac{\lambda m_{iM}}{1 + \lambda m_{MM}} \Big(t_{Mj} - \frac{\beta_1 t_{M1} t_{1j}}{1 + \beta_1 t_{11}} \Big),$$

and that

$$\mathbb{E}[z^{K_n}] = \frac{\alpha_1}{\lambda(1 - r_1)(r_2 - 1)} \Big[1 + \beta_1 \frac{1 - z}{\lambda r_2 - \beta_1} F^{\{-2,1\}}(r_1) \Big], \tag{41}$$

where $F^{\{-2,1\}}(.)$ and $r_{1,2}$ are given in (15) and (39) respectively. Inserting $z = \beta_2/(\beta_2 + s)$ into (41) gives the closed form of $G(s)$, the LST of the service requirement of a total batch (see (30)), which in turn gives the closed form of $\widetilde{W}_B(s)$.

3.3 Sojourn time in queue two

In the beginning of the section, we already remarked that, D_2, the sojourn time of a customer in Q_2 consists of three parts: the waiting time and the service time of a customer in a batch arrival queue, and the time a customer is in Q_2 but Q_2 is not served. Together, the waiting time and service time in the batch arrival queue form the sojourn time of the customer in the batch arrival queue.

Let H_0 denote the sojourn time of a costumer in the batch arrival queue. Let $\{H_t : t \geq 0\}$ denote the remaining sojourn time of a customer in Q_2 if the server would be continuously working at Q_2 from time t onwards. In other words, H_t decreases at rate 1 when $L(t) = 0$ and H_t is constant when $L(t) \in \{-2, -1, 1\}$ at time t. The service at Q_2 is interrupted by the mobility of the queue. Let Y denote the number of service interruptions during the sojourn time of a customer. Figure 4 displays a sample path of the evolution of H_t as a function of t, in this figure the threshold zero is crossed at $Y = 3$.

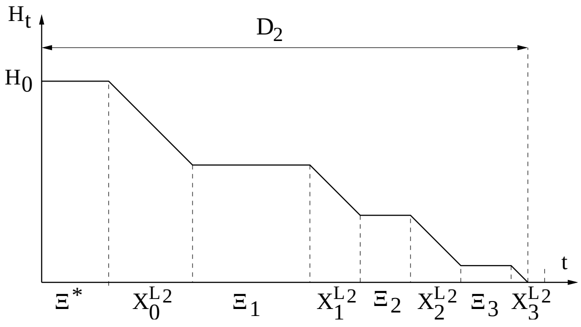

Figure 4: Evolution of H_t as a function of t with $Y = 3$.

The visit periods have an exponential length with rate α_2. Now, given that $H_0 = v$, the number of interruptions has a Poisson distribution with

$$\mathbb{E}[z^Y | H_0 = v] = e^{-\alpha_2 v(1-z)}. \tag{42}$$

The duration of these interruptions are independent and are given by $\Xi = C^{2,1} + X^{L_1} + C^{1,2}$. Furthermore, Ξ^*, the time it takes before H_t actually starts decreasing after time 0, satisfies $\Xi^* = X_e^{L_1} + C^{1,2}$, where $X_e^{L_1}$ is the residual time of X^{L_1}. Note that $X_e^{L_1}$ and X^{L_1} are identically distributed with common exponential distribution.

From Figure 4 it is easily seen that

$$D_2 = \Xi^* + H_0 + \sum_{i=1}^{Y} \Xi_i. \tag{43}$$

By conditioning on H_0 and Y, we find for the LST of D_2,

$$\begin{aligned} \tilde{D}_2(s) &= \mathbb{E}\big[e^{-s\Xi^*}\big]\mathbb{E}\big[e^{-s(\sum_{i=1}^{Y} \Xi_i + H_0)}\big], \\ &= \mathbb{E}\big[e^{-s\Xi^*}\big]\mathbb{E}\big[e^{-sH_0} e^{-\alpha_2 H_0 (1-\tilde{\Xi}(s))}\big]. \end{aligned} \tag{44}$$

where $\tilde{\Xi}(s) = \frac{\alpha_1}{\alpha_1+s}\tilde{C}^{1,2}(s)\tilde{C}^{2,1}(s)$. Since H_0 equals the sojourn time in the batch arrival queue, we find (see, Eq. (33))

$$\tilde{D}_2(s) = \frac{\alpha_1 \tilde{C}^{1,2}(s)}{\alpha_1 + s} \times \tilde{W}^B\big(\Delta(s)\big) \times \frac{\beta_2}{\mathbb{E}[K_n]} \times \frac{1 - G\big(\Delta(s)\big)}{\Delta(s)}, \tag{45}$$

where $\Delta(s) := s + \alpha_2(1 - \tilde{\Xi}(s))$.

4. NUMERICAL EVALUATION

The evaluation of the model will be done in three parts. First, we will extensively validate the accuracy of the approximation. Second, we consider the impact of the switch-over time distribution on the mean sojourn time. Notice that the switch-over times determine to a large extent the inter-contact times. Finally, we study the problem of optimizing the end-to-end delay in the network by adjusting the visit time parameters for a given cycle length. Throughout this section, the distribution of the switch-over times of Q_2, $C^{1,2}$ and $C^{2,1}$, are assumed identically distributed according to an exponential distribution with mean $c^{1,2} = c^{2,1}$.

4.1 Model validation

We validate the approximate model developed in Section 3.3 for the mean sojourn time at Q_2. In this model, K_n, the batch sizes in the batch arrival queue (see, e.g., (27)), were assumed to be mutually independent and independent of all other rvs. The validation will be done by comparing the results with those of the exact model in Section 2.4. We recall that due to the state-space expansion, the computation time for the exact joint queue-length probabilities, and thus also the mean sojourn time, may grow large for certain model parameters. Therefore, in the latter case we will use simulation to determine the mean sojourn time in Q_2.

Now, let us introduce some notation. Let $\mathbb{E}[D_2^{app}]$ (resp. $\mathbb{E}[D_2^{exa}]$) denote the mean sojourn time in Q_2 using the approximate (resp. exact) model given in Sect. 3.3 (resp. in Sect. 2.4.2). Let R_2 denote the absolute relative difference between the approximate and exact mean sojourn time in Q_2, i.e.,

$$R_2 := \left| 1 - \frac{\mathbb{E}[D_2^{app}]}{\mathbb{E}[D_2^{exa}]} \right|.$$

Further, we note that the load at Q_1 and Q_2 can be written as

$$\rho_i \;=\; \frac{\lambda}{\beta_i}\left(\frac{\alpha_1 + \alpha_2}{\alpha_{3-i}} + 2\alpha_i c^{1,2}\right), \quad i = 1, 2.$$

Figure 5 displays R_2 as a function of λ for different values of $c^{1,2}$ with $\beta_1 = \beta_2 = 1$ and $\alpha_1 = \alpha_2 = 0.1$. Thus, in this scenario the load at Q_1 and Q_2 are equal ($\rho_1 = \rho_2$). Observe that R_2 increases with λ and that the approximate model is accurate for $\rho_1 = \rho_2 < 0.5$. This is because the probability that Q_1 is empty upon the departure of the server from Q_1 decreases with λ. For this reason, the auto-correlation of K_n increases with λ. Moreover, Figure 5 shows that R_2 decreases with $c^{1,2}$ for $\rho_1 = \rho_2$ (e.g., for $\rho_1 = 0.5$, $R_2 = 15\%$ when $c^{1,2} = 1sec$ and $R_2 = 8\%$ when $c^{1,2} = 20$). This is because in the case where $\rho_1 = \rho_2$, λ decreases with $c^{1,2}$.

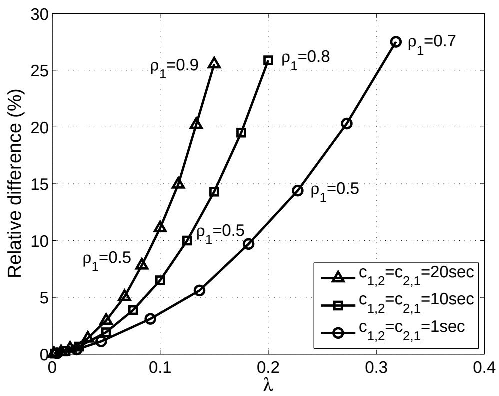

Figure 5: R_2 as a function of λ for different values of $c^{1,2}$ with $\beta_1 = \beta_2 = 1$ and $\alpha_1 = \alpha_2 = 0.1$.

Figure 6 shows the mean sojourn time in Q_2 using the approximate and exact models. Observe that the approximation gives an upper bound for $\mathbb{E}[D_2]$. This observation is in support of the result in [3] which proves that in the correlated $M/G/1$ a positive correlation between the service requirement and the last inter-arrival reduces the mean sojourn time. We should emphasize that also in our model K_n

and the last inter-arrival are positively correlated, i.e., an increase of the last interarrival time induces stochastically an increase of K_n.

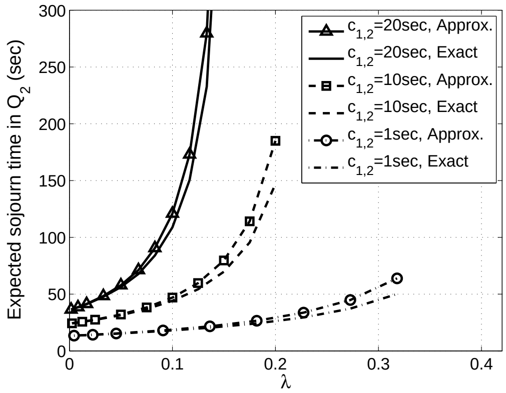

Figure 6: Mean sojourn time in Q_2 calculated from the approximate model and the exact model (resp. simulation for $\lambda > 0.05$) as a function of λ for different values of $c^{1,2}$ with $\beta_1 = \beta_2 = 1$ and $\alpha_1 = \alpha_2 = 0.1$.

Figure 7 shows R_2 as a function of ρ_2 for different values of ρ_1 with λ, β_1, and $\alpha_1 = \alpha_2$ constant. This is done by changing the value of β_2. First, observe that for a given ρ_1, the approximate model is more accurate for small values of ρ_2. This is due to the increase of probability that Q_2 is empty at a batch arrival instant which in turn decreases the correlation between the sojourn times of customers in different batches. Second, for a given ρ_2, the approximate model is more accurate for higher values of ρ_1 (for example when $\rho_2 = 0.6$, $R_2 = 9.6\%$ for $\rho_1 = 0.25$, while $R_2 = 4.72\%$ for $\rho_1 = 0.75$). The reason is that for high values of ρ_1 the queue size of Q_1 is large for most of the time, therefore in this case K_n will only depend on $X_n^{L_1}$ and S_1, the service requirement of a customer in Q_1. Since the sequences $\{X_n^{L_1}\}_{n\geq 0}$ and $\{S_{1,i}\}_{i\geq 0}$ are independent, the auto-correlation of $\{K_n\}_{n\geq 0}$ becomes smaller for higher values of ρ_1.

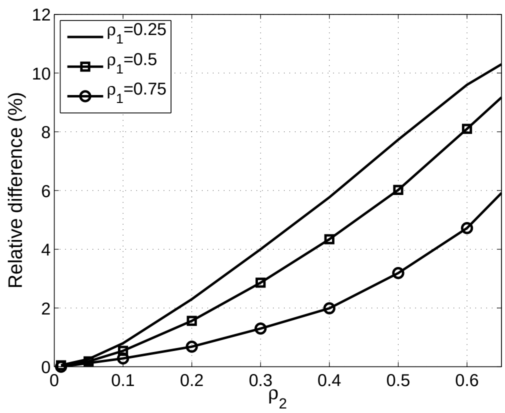

Figure 7: R_2 as a function of ρ_2 for different values of ρ_1 with $\lambda = 0.01$, $\alpha_1 = \alpha_2 = 0.1$, and $c^{1,2} = 10$.

We conclude that the approximate model has the following properties:

- It is accurate for *low and moderate* load at Q_1 and Q_2;

$SCOV_s$	1	5	10	15	20	30
$SCOV_i$	0.33	1.22	2.33	3.44	4.55	6.78
	Hyper-exponential distribution					
$\mathbb{E}[D^{app}]$	45.07	55.85	69.28	82.69	96.06	122.7
$\mathbb{E}[D^{exa}]$	45.02	55.8	69.25	82.66	95.96	121.4
	Coxian distribution					
$\mathbb{E}[D^{app}]$	45.07	55.85	69.31	82.74	96.15	122.9
$\mathbb{E}[D^{exa}]$	45.02	55.81	69.27	82.71	96.12	122.6
	Weibull distribution					
$\mathbb{E}[D^{sim}]$	45.01	55.85	69.25	82.54	95.88	121.9

Table 2: Mean sojourn time in Q_1 and Q_2 as a function of $SCOV$ for the hyper-exponential, Coxian, and Weibull distributions of the switch-over times with $\lambda = 0.01$, $\beta_1 = \beta_2 = 1$, $\alpha_1 = \alpha_2 = 0.1$, and $c^{1,2} = 10$.

- It gives an upper bound for the sojourn time at Q_2;
- It is accurate for *high* load at Q_1 and *moderate* load at Q_2.

4.2 Impact of the switch-over times distribution on sojourn time

We note that in the analysis the distribution of the switch-over time was assumed to be arbitrary. This section studies the impact of the distribution of the switch-over times on the end-to-end sojourn time of a customer. This will be done by considering the following three different distributions of the switch-over times in such way that they share the same first two moments: two-phase hyper-exponential, two-phase Coxian and Weibull distribution.

For $c^{1,2} = 10$, Table 2 displays the mean sojourn time as a function of $SCOV_s := Var(C^{1,2})/(c^{1,2})^2$, the squared coefficient of variation of the switch-over times, and of $SCOV_i$ $:= Var(C^{1,2} + C^{2,1} + X^{L_2})/(2c^{1,2} + 1/\alpha_2)^2$, the squared coefficient of variation of the inter-visit times. This is done using both the approximate and exact models for the hyper-exponential and Coxian distributions. For the Weibull distribution, we used simulation since its LST it is not known in closed form. In our simulation settings, we fixed the confidence interval to be within 1.2% of the mean simulated value. Observe that in Table 2 the mean sojourn time is almost equal for the three different distributions. Hence, we conclude that the mean sojourn time depends on the distribution of the switch-over and inter-visit times through their first two moments. In other words, considering two different distributions of the switch-over time with equal first two moments and different higher moments will induce the same mean sojourn time.

4.3 Insight on the optimal end-to-end sojourn time

In this section we study the evolution of, α_2^{opt}, the optimal value of α_2 that yields the minimum value of the end-to-end sojourn time in Q_1 and Q_2. This will be done under the constraints of zero switch-over time, i.e., $c^{1,2} = 0$, and of constant cycle length, i.e., $\mathbb{E}[C] = 1/\alpha_1 + 1/\alpha_2$ is constant. Moreover, the load at Q_1 and Q_2 should be between zero and one. Note that under these constraints when α_1 increases α_2 should decrease. Since the mean sojourn time in Q_1 (resp. Q_2) increases (resp. decreases) with α_1 (resp. α_2) then

α_2^{opt} exists and it is unique. The adjustment of α_1 and α_2 can be done in practice by controlling the transmission power of the nodes in our model.

In the following, α_2^{opt} will be computed using the approximate mean sojourn time in (45) using the numerical optimization package of MAPLE. This value was validated by verifying that the mean sojourn time using the exact model for $\alpha_2 = \alpha_2^{opt}$ is a local minimum inside $[\alpha_2^{opt} - 10^{-3}, \alpha_2^{opt} + 10^{-3}]$.

In the symmetric case $\beta_1 = \beta_2$, it is found that $\alpha_2^{opt} = \alpha_1 = 2/\mathbb{E}[C]$. In the asymmetric case $\beta_1 > \beta_2 = 1$, Table 3 displays α_2^{opt} as a function of β_1 for $\mathbb{E}[C] = 10$. Observe that in this case α_2^{opt} is smaller than $2/\mathbb{E}[C]$ and that this difference increases with β_1. Table 4 displays α_2^{opt} as a function of β_2 for $\beta_2 > \beta_1 = 1$ and $\mathbb{E}[C] = 10$. In contrast with the previous case, notice that α_2^{opt} is greater than $2/\mathbb{E}[C]$. In fact, the values of α_2^{opt} and α_1 for these two cases are exchanged which is quite surprising since the arrival processes at the queues are essentially different. It is not clear why these values found for α_2^{opt} would indeed lead to the optimal mean sojourn time.

β_1	1.1	2	3	6	11	16
α_2^{opt}	0.197	0.181	0.173	0.165	0.161	0.16
α_1	0.203	0.223	0.236	0.253	0.265	0.27
ρ_1 (%)	4.62	2.79	1.97	1.05	0.6	0.42
ρ_2 (%)	4.91	4.50	4.34	4.12	4.01	3.97

Table 3: α_2^{opt} as a function of β_1 for $\beta_2 = 1$, $\lambda = 0.025$, and $\mathbb{E}[C] = 10$.

β_2	1.1	2	3	6	11	16
α_2^{opt}	0.203	0.223	0.236	0.253	0.265	0.27
α_1	0.197	0.181	0.173	0.165	0.161	0.16
ρ_1 (%)	4.92	4.52	4.34	4.13	4.01	0.42
ρ_2 (%)	4.61	2.79	1.96	1.05	0.60	3.97

Table 4: α_2^{opt} as a function of β_2 for $\beta_1 = 1$, $\lambda = 0.025$, and $\mathbb{E}[C] = 10$.

In practice one might prefer to have a simple rule that provides a value for α_2 which yields a mean sojourn time close to optimal. Therefore, we will discuss two alternative, heuristic optimization approaches. First, we select the values of α_1 and α_2 such that the load is balanced at both queues, i.e., $\rho_1 = \rho_2$. This gives:

$$\alpha_i = \frac{\beta_1 + \beta_2}{\beta_{3-i}} \cdot \frac{1}{\mathbb{E}[C]}, \; i = 1, 2 \; . \tag{46}$$

Second, we choose α_1 and α_2 based on the analysis of a tandem model of two M/M/1 queues with shared service capacity. That means that the servers at both queues are always present, but serving at rate ν at Q_1 and at rate $1-\nu$ at Q_2. Then, the optimal ν, say ν^*, is the one that minimizes the end-to-end sojourn in such a tandem model, which we denote by $\mathbb{E}[D]^{M/M/1}$ and equals simply

$$\mathbb{E}[D]^{M/M/1} = \frac{1}{\beta_1\nu - \lambda} + \frac{1}{\beta_2(1-\nu) - \lambda} \; . \tag{47}$$

We choose the ratio α_1/α_2 equal to $(1-\nu^*)/\nu^*$, such that the fraction the server is at Q_1 in our model equals the optimal rate ν^* in the M/M/1 tandem model.

In Tables 5 and 6, we present the results of this comparison. Here, $\alpha_2^{opt}, \alpha_2^{LB}$ and $\alpha_2^{M/M/1}$ refer to the choice of α_2 in the optimal case, in the load balancing heuristic, and in the M/M/1 tandem heuristic, respectively. Further, we present the relative differences in mean sojourn time using the two heuristics (denoted by ϵ^{LB} and $\epsilon^{M/M/1}$) with respect to the optimal mean sojourn time, $\mathbb{E}[D]^{opt}$. In Table 5, we study the performance of those heuristics when β_1 is increased while β_2, λ and $\mathbb{E}[C]$ are kept constant. We note that for the symmetric case, $\beta_1 = \beta_2$, the heuristics would also give the optimal solution $\alpha_1 = \alpha_2$. The performance using load balancing worsens rapidly when β_1 is increased. Also the M/M/1 tandem heuristic deviates from the optimum, but the relative differences remain small. In Table 6, we investigate the performance of the heuristics when the mean cycle time is varied. The results show that the relative error when using load balancing is almost insensitive to $\mathbb{E}[C]$. We note that in the limit case of the cycle time tending to zero our tandem model approaches the tandem model of two M/M/1 queues. Hence, in this case the M/M/1 tandem heuristic is optimal. This explains the why the relative error increases in $\mathbb{E}[C]$. However, notice that the relative error $\epsilon^{M/M/1}$ is still very small for $\mathbb{E}[C] = 20$.

We can conclude that balancing the load is not a good solution for end-to-end sojourn time optimization unless $\beta_1 \approx \beta_2$. However, using an optimization heuristic based on a simple tandem model of two M/M/1 queues will give nearly optimal results for the mean sojourn time under a wide variety of parameter settings.

β_1	1.1	2	3	6	11	16
α_2^{opt}	0.194	0.174	0.166	0.156	0.150	0.148
α_2^{LB}	0.190	0.150	0.133	0.117	0.109	0.106
$\alpha_2^{M/M/1}$	0.195	0.167	0.154	0.138	0.128	0.123
$\mathbb{E}[D]^{opt}$	14.47	12.82	12.14	11.42	11.08	10.94
ϵ^{LB} (%)	<0.1	3.9	8.6	17.2	23.6	26.3
$\epsilon^{M/M/1}$ (%)	<0.1	0.4	0.9	2.3	3.8	4.7

Table 5: Comparison of α_2 and $\mathbb{E}[D]$ for different optimization approaches for $\beta_2 = 1$, $\lambda = 0.1$, and $\mathbb{E}[C] = 10$.

$\mathbb{E}[C]$	1	2	5	10	20
α_2^{opt}	1.408	0.719	0.300	0.156	0.080
α_2^{LB}	1.167	0.583	0.233	0.117	0.058
$\alpha_2^{M/M/1}$	1.375	0.687	0.275	0.137	0.069
$\mathbb{E}[D]^{opt}$	3.152	4.07	6.83	11.42	20.58
ϵ^{LB} (%)	17.3	17.1	17.1	17.2	17.7
$\epsilon^{M/M/1}$(%)	0.2	0.6	1.4	2.3	3.0

Table 6: Comparison of α_2 and $\mathbb{E}[D]$ for different optimization approaches for $\beta_1 = 6$, $\beta_2 = 1$, and $\lambda = 0.1$.

5. CONCLUSIONS

This study is part of a research effort towards developing analytical models for quantifying the end-to-end delay in a opportunistic network. We consider here a network consisting of a fixed source node, a fixed destination node, and a mobile relay node. A closed-form expression has been derived for the delay at the packet's source node. Next, an iterative approach has been developed for the joint queue-length distribution of the source and the relay node. In addition, a simple approximate model has been proposed for the delay analysis at the relay node. The approximate model has extensively been validated and shows excellent results. Numerical results on the mean end-to-end delay show that the inter-contact time distribution impacts this metric only through its first two moments. Moreover, load balancing is not an effective tool for delay optimization, while the M/M/1 tandem heuristic is near optimal.

In the present work, we have focused on the delay analysis of simple opportunistic networks. As a second step towards understanding these networks, we will study in the future work the scenario where multiple relay nodes coexist in the network. In this case, the packet multi-copy routing-based proposals in DTN will help to reduce the delay. We anticipate that the exact and approximate models can be extended at least to cover the packet single-copy case with only one packet transmission at a time. For the multi-copy case, the help of certain theoretical techniques like customers resequencing and impatient customers might be required to analyze the delay.

6. REFERENCES

[1] Delay tolerant networking research group. Web site: http://www.dtnrg.org.

[2] C. Bettstetter, H. Hartenstein, and X. Pérez-Costa. Stochastic properties of the random waypoint mobility model. *ACM/Kluwer Wireless Networks, Special Issue on Modeling and Analysis of Mobile Networks*, 10(5):555–567, Sept. 2004.

[3] S. Borst, O. Boxma, and M. Combé. Collection of costumers: a correlated M/G/1 queue. *Performance Evaluation*, 20(1):47–59, 1992.

[4] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott. Impact of human mobility on the design of opportunistic forwarding algorithm. In *Proc. of IEEE INFOCOM*, Barcelona, Spain, Apr. 2006.

[5] E. G. Coffman, G. Fayolle, and I. Mitrani. Two queues with alternating service periods. In *Performance '87: Proc. of the 12th IFIP WG 7.3 International Symposium on Computer Performance Modelling, Measurement and Evaluation*, pages 227–239, 1988.

[6] R. de Haan, R. J. Boucherie, and J.-K. van Ommeren. A polling model with an autonomous server. *Research Memorandum, University of Twente*, 2007.

[7] B. Doshi. Queueing systems with vacations-a survey. *Queueing Systems*, 1(1):29–66, 1986.

[8] M. Dow. Explicit inverses of Toeplitz and associated matrices. *ANZIAM J.*, 44(E):E185–E215, Jan. 2003.

[9] M. Eisenberg. Queues with periodic service and changeover times. *Operation Research*, 20(2):440–451, 1972.

[10] I. Frigui and A. Alfa. Analysis of a time-limited polling system. *Computer Communications*, 21(6):558–571, 1998.

[11] D. P. Gaver, P. A. Jacobs, and G. Latouche. Finite birth-and-death models in randomly changing environments. *Advances in Applied Probability*, 16:715–731, 1984.

[12] C. Grinstead and J. Snell. *Introduction to Probability.* American Mathematical Society, 1997.

[13] R. Groenevelt, P. Nain, and G. Koole. The message delay in mobile ad hoc networks. *Performance Evaluation*, 62(1-4):210–228, Oct. 2005.

[14] M. Grossglauser and D. Tse. Mobility increases the capacity of ad hoc wireless networks. *ACM/IEEE Transactions on Networking*, 10(4):477–486, Aug. 2002.

[15] M. Ibrahim, A. Al Hanbali, and P. Nain. Delay and resource analysis in manets in presence of throwboxes. *Performance Evaluation*, 64(9-12):933–947, Oct. 2007.

[16] S. Jain, K. Fall, and R. Patra. Routing in a delay tolerant networking. In *Proc. of ACM Sigcomm*, Aug. 2004.

[17] T. Karagiannis, J.-Y. L. Boudec, and M. Vojnovic. Power law and exponential decay of inter contact times between mobile devices. In *Proc. of MOBICOM*, Montréal, Québec, Canada, Sep. 2007.

[18] T. Katayama. Waiting time analysis for a queueing system with time-limited service and exponential timer. *Naval Research Logistics*, 48:638–651, 2001.

[19] T. Katayama. Level-crossing approach to a time-limited service system with two types of vacations. *Operations Research Letters*, 33(3):295–300, 2005.

[20] J. Leguay, T. Friedman, and V. Conan. Evaluating mobility pattern space routing for DTNs. In *Proc. of IEEE Infocom*, Barcelona, Spain, Apr. 2006.

[21] K. Leung. Cyclic-service systems with probabilistically-limited service. *IEEE Journal on Selected Areas in Communications*, 9(2):185–193, 1991.

[22] K. Leung. Cyclic-service systems with non-preemptive time-limited service. *IEEE Transactions on Communications*, 42(8):2521–2524, 1994.

[23] P. Nain, D. Towsley, B. Liu, and Z. Liu. Properties of random direction models. In *Proc. of IEEE INFOCOM*, Miami, FL, Mar. 2005.

[24] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, 1992.

[25] T. Small and Z. J. Haas. The shared wireless infostation model: A new ad hoc networking paradigm. In *Proc. of ACM MOBIHOC*, Anapolis, MD, USA, 2003.

[26] T. Small and Z. J. Haas. Resource and performance tradeoffs in delay-tolerant wireless networks. In *Proc. ACM SIGCOM Workshop on Delay-Tolerant Networks*, Philadelphia, PA, USA, Aug. 2005.

[27] T. Spyropoulos, K. Psounis, and C. Raghavendra. Spray and wait: An efficient routing scheme for intermittently connected mobile networks. In *Proc. of ACM Sigcomm workshop on Delay Tolerant Networking*, Aug. 2005.

[28] T. Spyropoulos, K. Psounis, and C. S. Raghavendra. Efficient routing in intermittently connected mobile networks: The multi-copy case. *To appear in ACM/IEEE Transactions on Networking (available online)*, 2007.

[29] T. Spyropoulos, K. Psounis, and C. S. Raghavendra. Efficient routing in intermittently connected mobile networks: The single-copy case. *To appear in ACM/IEEE Transactions on Networking (available*

online), 2007.

[30] R. Wolff. *Stochastic Modeling and the Theory of Queues*. Prentice Hall, 1989.

[31] M. Zazanis. A palm calculus approach to functional versions of little's law. *Stochastic Processes and their Applications*, 74:195–201(7), 1998.

[32] E. Zhang, G. Neglia, J. Kurose, and D. Towsley. Performance modeling of epidemic routing. *Computer Networks*, 51(10):2867–2891, July 2007.

[33] W. Zhao, M. Ammar, and E. Zegura. Controlling the mobility of multiple data transport ferries in a delay-tolerant network. In *Proc. of IEEE INFOCOM*, Miami, FL, Mar. 2005.