Performance Evaluation 70 (2013) 981–994

Contents lists available at ScienceDirect

Performance Evaluation

journal homepage: www.elsevier.com/locate/peva

Energy–delay tradeoff in a two-way relay with network coding

Jasper Goseling [a,b,*], Richard J. Boucherie [a], Jan-Kees van Ommeren [a]

[a] *Stochastic Operations Research, Department of Applied Mathematics, University of Twente, P.O. Box 217, 7500 AE, Enschede, The Netherlands*

[b] *Department of Intelligent Systems, Delft University of Technology, P.O. Box 5031, 2600 GA, Delft, The Netherlands*

ARTICLE INFO

Article history:
Received 15 June 2012
Received in revised form 13 March 2013
Accepted 3 August 2013
Available online 11 August 2013

Keywords:
Network coding
Energy
Delay
Positive and negative customers
Markov reward models
Error bounds

ABSTRACT

A queueing model for a relay in a communication network that is employing network coding is introduced. It is shown that communication networks with coding are closely related to queueing networks with positive and negative customers. The tradeoff between minimizing energy consumption and minimizing delay for a two-way relay is investigated. Analytical upper and lower bounds on the energy consumption and the delay are obtained using a Markov reward approach. Exact expressions are given for the minimum energy consumption and the minimum delay that are attainable.

1. Introduction

We consider a two-way relay communication network in which the relay is performing network coding. The two-way relay receives packets from two different connections that need to be forwarded. Network coding, *cf.*, [1,2], is based on the observation that in addition to simply forwarding packets it can be useful to do additional processing at the relay and combine different packets before forwarding them. The network coding operation performed by the two-way relay is to combine pairs of two packets, one from each connection, by taking the elementwise exclusive-or of bits in both packets, and forward only the combined packet. Since less packets are transmitted, energy consumption is reduced. The influence on packet delay, *i.e.*, queue length, is not obvious. It might be reduced since overall less packets are transmitted. However, it might also be increased since the synchronization required for coding introduces delays. In the current paper we analyze in more detail the energy consumption and the packet delay in the relay and compare these with the performance of a relay that simply forwards all packets and is not performing network coding.

Network coding in a two-way relay has been extensively studied. Initial work, for instance [3–5], focused on the assumption of saturated queues, *i.e.*, on the assumption that the relay always has packets from both connections available. More recent work is dealing with stochastic arrivals of packets at the relay in which case it can happen that there are no packets available for one of the connections. This implies that the relay can now decide to transmit an uncoded packet or to wait for a coding opportunity provided by the arrival of a new packet. In [6] a queueing model is developed in which the decision to transmit an uncoded packet may depend on the number of packets in the queues. The impact of these decisions

* Corresponding author at: Department of Intelligent Systems, Delft University of Technology, P.O. Box 5031, 2600 GA, Delft, The Netherlands. Tel.: +31 534893369.

E-mail addresses: j.goseling@utwente.nl (J. Goseling), r.j.boucherie@utwente.nl (R.J. Boucherie), j.c.w.vanommeren@utwente.nl (J.C.W. van Ommeren).

0166-5316/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.peva.2013.08.002

on the energy–delay tradeoff is analyzed in [7]. In [8] an analysis is made for the case that packets are never transmitted uncoded. The scenario that packets are always transmitted uncoded if opportunity arises is studied in [9,10].

The analysis of stochastic arrivals at a relay with network coding has been extended in many directions: the additional delay incurred by processing is taken into account in [11] — more general network structures have been considered in [12] — the influence of ALOHA type medium access protocols is considered in [13] — the impact of a limited buffer size at the relay and optimal buffer allocation policies are analyzed in [14] — in [15] the stability region under QoS constraints is analyzed. The current work is most closely related to [7] in the sense that we consider the energy–delay tradeoff arising from transmitting uncoded packets. Before discussing the contributions of the current work and the relation to [7] we discuss the model and the methods that have been used.

We model the two-way relay as a two-dimensional continuous-time Markov process. In the model we keep two queues, one for packets from each of the sessions. Packets from the two sessions arrive independently. We model the transmission of a combination of packets as the simultaneous departure of two packets, one packet from each queue. As observed in [7,8], and formulated precisely in Section 4, in order to keep the system stable, it is necessary to also transmit uncoded packets. In particular, if packets are present in one queue while the other queue is empty, uncoded packets will need to be transmitted. We will, therefore, allow for operating policies in which packets from one queue can be transmitted uncoded if the other queue is empty. In particular we analyze policies that, given the opportunity to transmit a packet from the one queue, while the other queue is empty, transmit this packet with some fixed probability. These policies form a generalization of the policies considered in [7] where this uncoded packet would either always or never be transmitted.

Our queueing model has similar properties as a queueing network with positive and negative customers [16]. Observe that packets arrive at the relay one by one, but that packets depart from two queues simultaneously. This is the typical behavior of queueing networks with positive and negative customers. However, due to the behavior when one of the queues is empty, there is not always a one-to-one correspondence. It is demonstrated in the current paper that if the operating policy that decides when to transmit an uncoded packet is chosen carefully, our model is a queueing network with positive and negative customers. In this case the system is known to have a geometric product-form stationary distribution. For other operating policies this is not the case. Therefore, exact results on performance cannot always readily be obtained and we will resort to finding analytical performance bounds. These bounds will be obtained from Markov reward error bounding techniques [17] by relating the performance of our model to that of a perturbed model with a product-form stationary distribution, *i.e.*, a network with positive and negative customers.

Networks with positive and negative customers have previously been applied for the analysis of communication networks. In [18–20] this was done for energy-aware routing, in [21] for energy management in the cloud and in [22] for resource allocation in multimedia systems. The application of networks with positive and negative customers to relays with network coding is new. A survey on networks with positive and negative customers is given in [23] and a collection of literature in the area is provided in [24].

The contributions of the current work are:

1. The relation between a queueing model for a two-way relay with network coding and a network with positive and negative customers is demonstrated.
2. Analytical upper and lower bounds on the energy consumption and the delay in the two-way relay are given.
3. Exact expressions for the minimum possible energy consumption and minimum possible delay are given.

As will be discussed in Section 6, performing network coding at the relay in order to reduce energy consumption might lead to long delays. This was already observed in [8] and it is confirmed by our analytical upper and lower bounds. As discussed above, our policies form a generalization of some of the policies from [7]. The current paper extends our work presented in [25], where only energy consumption was considered. The bounds on energy consumption that were provided in [25] have been significantly strengthened.

The remainder of this paper is organized as follows. In Section 2 we specify the continuous-time Markov chain that will be analyzed and the performance measures of interest. Section 3 is devoted to discussing some of the preliminaries that are required later in the paper. In particular we discuss queueing networks with negative customer and Markov reward error bounds. In Section 4 we give necessary and sufficient conditions for ergodicity of our model. Performance bounds on expected queue size and energy consumption are presented in 5. Numerical examples of the results obtained in the paper are given in Section 6. Finally, in Section 7 we discuss the results presented in this paper and offer suggestions for future work.

2. Model and problem statement

We consider a single node in a wireless network that is acting as a relay for two sessions and develop two different continuous-time queueing models. The classical case without network coding is covered by the first model. In the second model network coding is used. Packets from both sessions arrive at the node according to independent Poisson processes with rate λ_1 and λ_2. The time required to transmit a packet, *i.e.*, to provide service for a packet, is exponentially distributed with rate μ.

The uncoded system is modeled as a single server operating on a single queue using a FIFO policy. Hence the uncoded system is an M/M/1 queue with arrival rate $\lambda_1 + \lambda_2$ and service rate μ.

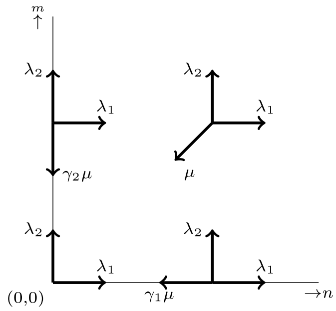

Fig. 1. Transition diagram for Q^{γ_1,γ_2}, the Markov process of the coded system.

In the coded system a separate queue is kept for each session, leading to a two-dimensional model in which the state variables N and M denote the number of packets contained in each of the queues. Network coding is employed by transmitting linear combinations of two packets, one packet from each queue in a combination. This means that a service completion will reduce the number of packets in both queues by one. If only one queue has a packet it is transmitted uncoded and a service completion will remove only one packet from a queue. Since transmitting an uncoded packet is unfavorable in terms of, for instance, energy consumption, we allow for an operating policy in which uncoded packets will not always be transmitted if opportunity arises.

If there is the opportunity to transmit a packet from the first queue, while the second queue is empty, this packet will be transmitted with probability γ_1. Similarly, packets from the second queue will be transmitted uncoded with probability γ_2.

The above description leads to a continuous-time Markov chain Q^{γ_1,γ_2} on state space $\mathbb{N}_0^2$ with transition rates q^{γ_1,γ_2} defined as

$$q_{n,m}^{\gamma_1,\gamma_2}(i,j) = \begin{cases} \lambda_1, & \text{if } i=1, j=0, n\geq 0, m\geq 0,\\ \lambda_2, & \text{if } i=0, j=1, n\geq 0, m\geq 0,\\ \mu, & \text{if } i=-1, j=-1, n>0, m>0,\\ \gamma_1\mu, & \text{if } i=-1, j=0, n>0, m=0,\\ \gamma_2\mu, & \text{if } i=0, j=-1, n=0, m>0,\\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

where $q_{n,m}^{\gamma_1,\gamma_2}(i,j)$ denotes the transition rate from state (n, m) to state $(n+i, m+j)$. To ensure irreducibility of the chain we assume $\lambda_1 > 0$, $\lambda_2 > 0$ and $\mu > 0$. Remember that γ_1 and γ_2 denote probabilities and take values in the interval [0, 1]. The transition structure is depicted in the transition diagram of Fig. 1. To simplify the notation in the remainder of the paper we introduce

$$\rho_1 = \frac{\lambda_1}{\mu}, \qquad \rho_2 = \frac{\lambda_2}{\mu}, \qquad \gamma_1^* = \frac{\rho_1-\rho_2}{1-\rho_2}, \qquad \gamma_2^* = \frac{\rho_2-\rho_1}{1-\rho_1}. \tag{2}$$

In Section 4 it will become clear that γ_1^* and γ_2^* are useful for expressing stability criteria.

Our interest is in two different steady-state performance measures of Q^{γ_1,γ_2}. To introduce notation, first consider an arbitrary cost/reward function $f_{\beta_1,\beta_2} : \mathbb{N}_0^2 \to [0,\infty)$, depending on parameters β_1 and β_2. Let

$$F_{\beta_1,\beta_2}^{\gamma_1,\gamma_2} = \mathbb{E}^{\gamma_1,\gamma_2}[f_{\beta_1,\beta_2}(N,M)], \tag{3}$$

where the expected value is over the stationary distribution of the process Q^{γ_1,γ_2}. In the remainder we will denote by capital letters, the expected value of the cost function defined through the corresponding lower-case letter.

The first performance measure that we consider is the expected energy consumption per unit time. The energy consumed by transmitting a packet is μ per unit time. Therefore, the energy consumed per unit time is

$$\begin{aligned} &0, \quad \text{if } n=0, m=0,\\ &\gamma_1\mu, \quad \text{if } n>0, m=0,\\ &\gamma_2\mu, \quad \text{if } n=0, m>0,\\ &\mu, \quad \text{if } n>0, m>0, \end{aligned} \tag{4}$$

where it is taken into account that a packet is transmitted with probability γ_1 (γ_2) if there is a packet in the first (second) queue while the second (first) queue is empty. It follows that the expected energy consumption of Q^{γ_1,γ_2} equals

$$C_{\gamma_1,\gamma_2}^{\gamma_1,\gamma_2} = \mathbb{E}^{\gamma_1,\gamma_2}\left[c_{\gamma_1,\gamma_2}(N,M)\right], \tag{5}$$

where the cost function c has additional parameters β_1 and β_2 and is defined as

$$c_{\beta_1,\beta_2}(n,m) = \beta_1\mu\mathbb{1}_{\{n>0,m=0\}} + \beta_2\mu\mathbb{1}_{\{n=0,m>0\}} + \mu\mathbb{1}_{\{n>0,m>0\}}. \tag{6}$$

The energy consumption will be bounded in terms of $C^{\hat{\gamma}_1,\hat{\gamma}_2}_{\beta_1,\beta_2}$, for values of $\hat{\gamma}_1$, $\hat{\gamma}_2$, β_1 and β_2, not necessarily equal to γ_1 or γ_2. We will sometimes denote $C^{\gamma_1,\gamma_2}_{\gamma_1,\gamma_2}$ as C^{γ_1,γ_2}, omitting the subscripts.

The second performance measure of interest is the expected delay, *i.e.*, the expected sojourn time of a packet in the system. Without loss of generality we will consider the delay of packets of the first connection and denote it by D^{γ_1,γ_2}. By Little's law it follows that the expected delay is $D^{\gamma_1,\gamma_2} = \mathbb{E}[N]/\lambda_1$, leading to cost function

$$d(n, m) = \frac{1}{\lambda_1} n. \tag{7}$$

Remember, that the uncoded system is an M/M/1 queue with arrival rate $\lambda_1 + \lambda_2$ and service rate μ. Therefore, the expected energy consumption in the uncoded system is

$$C_{\text{uncoded}} = \lambda_1 + \lambda_2 \tag{8}$$

and the expected delay is

$$D_{\text{uncoded}} = \frac{1}{\mu - \lambda_1 - \lambda_2}. \tag{9}$$

3. Preliminaries

Before starting the analysis of the process Q^{γ_1,γ_2} in the next section we will provide some background on the techniques that will be used. First, we discuss queueing networks with positive and negative customers, their relevance for the work at hand and the stationary distribution of such networks. Next, we provide results on Markov reward error bounding techniques.

3.1. Queueing networks with positive and negative customers

We start this section with an interpretation of Q^{γ_1,γ_2} for the case that $\gamma_1 + \gamma_2 = 1$. Under this condition the network can be interpreted as having two dedicated servers, one for each queue, operating at rates $\gamma_1\mu$ and $\gamma_2\mu$. In addition, if a packet is leaving from one of the queues and there is a packet in the other queue, that packet is also removed from the queue. This type of queueing network was first studied by Gelenbe [16]. The networks considered by Gelenbe in [16] are very similar to Jackson networks, with the additional feature that there are two types of customers: positive and negative. Positive customers, upon arriving at a node, require service and are placed in the queue. Negative customers, upon arriving at a queue, do not require service and instead, remove a positive customer from the queue. Upon completing service, there are three possible actions for a positive customer: (1) it leaves the system, (2) it enters another queue in the system as a positive customer, or (3) it enters another queue in the system as a negative customer. The customer chooses randomly, with a fixed probability distribution, which action to take and/or which queue to join. It is shown in [16,26] that these networks have a product-form stationary distribution. The parameters of the distribution are the solution of a set of polynomial equations that can be given for any network of the form described above. We give the resulting set of equations for the system Q^{γ_1,γ_2} under the condition that $\gamma_1 + \gamma_2 = 1$.

Theorem 1 (*Gelenbe [16]*)**.** *Consider the Markov process Q^{γ_1,γ_2} with $\gamma_1 + \gamma_2 = 1$. If the system of equations in σ_1 and σ_2 given by*

$$\gamma_1\sigma_1 + \gamma_2\sigma_1\sigma_2 = \rho_1, \qquad \gamma_2\sigma_2 + \gamma_1\sigma_1\sigma_2 = \rho_2, \tag{10}$$

has a unique solution satisfying $0 < \sigma_1 < 1$ and $0 < \sigma_2 < 1$, then the stationary distribution $\pi(n, m)$ is given by

$$\pi(n, m) = (1 - \sigma_1)\, \sigma_1^n\, (1 - \sigma_2)\, \sigma_2^m.$$

The above result is a special case of the result by Gelenbe, for a network of two queues with service rates $\gamma_1\mu$ and $\gamma_2\mu$, external arrival of positive customers with rates λ_1 and λ_2, no external arrivals of negative customers, and customers leaving one queue entering the other queue as negative customers.

Remark 1. The above theorem provides an expression for the stationary distribution under the condition that a unique solution $0 < \sigma_1 < 1$ and $0 < \sigma_2 < 1$ exists. Note that this corresponds exactly to the condition that the process is stable, *i.e.*, ergodic. Necessary and sufficient conditions for stability are given in [27] in terms of the properties of a fixed point of a continuous function derived from (10). This fixed point can, in general, not be given in explicit form. In Section 4 we will obtain explicit stability criteria for Q^{γ_1,γ_2} based on the theory of two dimensional random walks in the positive orthant.

3.2. Markov reward error bounds

Since for $\gamma_1 + \gamma_2 \neq 1$ the steady-state distribution of the queueing process cannot be obtained in a tractable analytical form, we will use the Markov reward approach to obtain analytical approximations on the performance of Q^{γ_1,γ_2}. This technique, developed by van Dijk [28,29], is based on relating the steady-state performance of the process to the cumulative reward structure in the discrete-time uniformized process. An introduction to this technique is given in, for instance, [17]. Throughout the remainder of this section we omit the dependence on γ_1 and γ_2 in the notation.

Let $f : \mathbb{N}_0^2 \to [0, \infty)$ be an arbitrary performance measure and denote by $\mathbb{E}[f(N, M)]$ the expected performance of Q under the (unknown) stationary distribution π. In addition consider a second Markov process $\bar{Q}$ on the same state space $\mathbb{N}_0^2$, but with different transition rates $\bar{q}$. Finally, consider a second performance measure $\bar{f} : \mathbb{N}_0^2 \to [0, \infty)$. Assume that the stationary distribution for $\bar{Q}$, $\bar{\pi}$ is known. We will approximate $\mathbb{E}[f(N, M)]$ in terms of $\bar{\mathbb{E}}\left[\bar{f}(N, M)\right]$, the expected value of the perturbed measure under distribution $\bar{\pi}$.

Assume that Q and $\bar{Q}$ can both be uniformized and let h be a suitable uniformization constant for both processes, *i.e.*, let h satisfy

$$h \leq \left(\sum_{i,j} q_{n,m}(i, j)\right)^{-1} \quad \text{and} \quad h \leq \left(\sum_{i,j} \bar{q}_{n,m}(i, j)\right)^{-1}, \tag{11}$$

for all $(n, m) \in \mathbb{N}_0^2$.

Let P denote the discrete-time Markov process obtained from uniformization of Q. For P the probability of jumping from (n, m) to $(n + i, m + j)$, $p_{n,m}(i, j)$, is defined as

$$p_{n,m}(i, j) = \begin{cases} hq_{n,m}(i, j), & \text{if } (i, j) \neq (0, 0), \\ 1 - h \sum\limits_{(i,j)\neq 0} q_{n,m}(i, j), & \text{if } (i, j) = (0, 0). \end{cases} \tag{12}$$

On P consider a one step reward $hf(n, m)$ whenever the system is in state (n, m). This leads to expected cumulative reward $F(k, n, m)$, incurred by the uniformized process at time step k when starting from state (n, m) at time 0, defined as

$$F(k + 1, n, m) = hf(n, m) + \sum_{i,j} p_{n,m}(i, j)F(k, n + i, m + j), \tag{13}$$

for $k > 0$ and $F(0, n, m) = 0$.

Using the above notation and definitions we are ready to state the results from [17] that will be used to obtain bounds on $\mathbb{E}[f(N, M)]$.

Theorem 2 (*van Dijk [17]*)**.** *Let Q and $\bar{Q}$ be two continuous-time Markov processes on the same state space $\mathbb{N}_0^2$, with transition rates q and $\bar{q}$, respectively. In addition consider the cost functions f and $\bar{f}$ from $\mathbb{N}_0^2$ to $[0, \infty)$. Suppose that there exists a function $\xi : \mathbb{N}_0^2 \to [0, \infty)$ such that for all $(n, m) \in \mathbb{N}_0^2$ and $k \in \mathbb{N}_0$,*

$$\left| f(n, m) - \bar{f}(n, m) + \sum_{i,j} \left[q_{n,m}(i, j) - \bar{q}_{n,m}(i, j)\right] \cdot [F(k, n + i, m + j) - F(k, n, m)] \right| \leq \xi(n, m). \tag{14}$$

Then

$$\left|\mathbb{E}[f(N, M)] - \bar{\mathbb{E}}\left[\bar{f}(N, M)\right]\right| \leq \sum_{n,m} \bar{\pi}(n, m)\xi(n, m). \tag{15}$$

Theorem 3 (*van Dijk [17]*)**.** *Let Q and $\bar{Q}$ be two continuous-time Markov processes on the same state space $\mathbb{N}_0^2$, with transition rates q and $\bar{q}$, respectively. In addition consider the cost functions f and $\bar{f}$ from $\mathbb{N}_0^2$ to $[0, \infty)$. Suppose that for all $(n, m) \in \mathbb{N}_0^2$ and $k \in \mathbb{N}_0$,*

$$f(n, m) - \bar{f}(n, m) + \sum_{i,j} \left[q_{n,m}(i, j) - \bar{q}_{n,m}(i, j)\right] \cdot [F(k, n + i, m + j) - F(k, n, m)] \leq 0. \tag{16}$$

Then

$$\mathbb{E}[f(N, M)] \leq \bar{\mathbb{E}}\left[\bar{f}(N, M)\right]. \tag{17}$$

Remark 2. The key step in applying Theorems 2 or 3 is to bound terms of the form

$$F(k, n + i, m + j) - F(k, n, m). \tag{18}$$

We will refer to terms of this type as bias terms.

Remark 3. Clearly, if a function $\xi : \mathbb{N}_0^2 \to [0, \infty)$ can be found such that both (14) and (16) hold then Theorems 2 and 3 can be combined to

$$-\sum_{n,m} \bar{\pi}(n, m)\xi(n, m) \le \mathbb{E}[f(N, M)] - \bar{\mathbb{E}}\left[\bar{f}(N, M)\right] \le 0. \tag{19}$$

Remark 4. If, in (16), $\le$ is replaced by $\ge$ then Theorem 3 will provide a lower bound on $\mathbb{E}[f(N, M)]$ instead of an upper bound, *i.e.*, $\mathbb{E}[f(N, M)] \ge \bar{\mathbb{E}}[f(N, M)]$.

4. Stability

The continuous-time Markov process Q^{γ_1,γ_2} is stable if and only if the corresponding uniformized process is stable. The uniformized process, is a two-dimensional homogeneous random walk in the positive quadrant, a type of process that has been extensively studied; see, for instance, [30,31] and the references therein. Necessary and sufficient conditions for ergodicity are given in [30]. For completeness we present the stability conditions for Q^{γ_1,γ_2} in the next theorem. These conditions demonstrate the purpose of introducing γ_1^* and γ_2^* in (2) as $\gamma_1^* = (\rho_1 - \rho_2)/(1 - \rho_2)$ and $\gamma_2^* = (\rho_2 - \rho_1)(1 - \rho_1)$.

Theorem 4. *The process Q^{γ_1,γ_2} is ergodic if and only if $\rho_1 < 1, \rho_2 < 1, \gamma_1 > \gamma_1^*$ and $\gamma_2 > \gamma_2^*$.*

Proof. Directly from [30, Theorem 1.2.1], by taking into account that γ_1 and γ_2 denote probabilities and hence lie in the interval [0, 1]. □

Intuitively, if $\rho_1 > \rho_2$ ($\rho_1 < \rho_2$), some packets from the first (second) queue will have to be transmitted uncoded. Theorem 4 quantifies the minimum fraction of packets that need to be transmitted uncoded. Note, that even if $\rho_1 = \rho_2$, it is not possible to have $\gamma_1 = \gamma_2 = 0$. This was also observed in [7,8]. In the remainder of this section we discuss how this result can be obtained directly without resorting to the theory developed in [30] or the methods used in[7,8].

Consider an alternative representation of Q^{γ_1,γ_2} in which the state variables are N, the number of packets in the first queue, and $K = N - M$, the difference between the number of packets in the first and the second queue. Observe, that if $\gamma_1 = \gamma_2 = 0$, the only changes to K occur from arrivals of packets, *i.e.*, K increases by one with rate λ_1 and decreases with one with rate λ_2. The corresponding discrete-time process, obtained after uniformization, is a random walk on $\mathbb{Z}$ and is not ergodic. If $\rho_1 = \rho_2$, the process is null recurrent.

5. Performance

5.1. The perturbed process

All performance bounds in this section will be obtained by perturbing some of the transition rates along the boundary of the state space. More precisely, we obtain bounds on the performance of our process of interest Q^{γ_1,γ_2} in terms of the performance of the perturbed process $\bar{Q}^{\alpha,1-\alpha}$, where $0 \le \alpha \le 1$ is a free parameter and where the bar notation is used to emphasize the role of the second process. Note that for the perturbed process we have the following transition rates

$$\bar{q}_{n,m}^{\alpha,1-\alpha}(i,j) = \begin{cases} \alpha\mu, & \text{if } i = -1, j = 0, n > 0, m = 0, \\ (1-\alpha)\mu, & \text{if } i = 0, j = -1, n = 0, m > 0, \\ q_{n,m}^{\gamma_1,\gamma_2}(i,j), & \text{otherwise.} \end{cases} \tag{20}$$

The effect of the perturbation is that along the vertical axis the rate towards the origin is changed from $\gamma_1\mu$ to $\alpha\mu$. Along the vertical axis the rate towards the origin changes from $\gamma_2\mu$ to $(1-\alpha)\mu$. In order to apply Theorems 2 or 3 we need to obtain the sign of the LHS of (16) or a bound on the LHS of (14), respectively. Since, $q_{n,m}(i,j) = \bar{q}_{n,m}(i,j)$ unless $n > 0, m = 0, i = -1$ and $j = 0$ or $n = 0, m > 0, i = 0$ and $j = -1$, we only need to obtain bounds on the following two bias terms

$$F(k, n, 0) - F(k, n-1, 0), \tag{21}$$

$$F(k, 0, m) - F(k, 0, m-1). \tag{22}$$

These bounds will be given for the specific performance measures of interest in Sections 5.2 and 5.3.

The parameter α can be chosen freely, but the process $\bar{Q}^{\alpha,1-\alpha}$ should be ergodic. The next theorem states that given λ_1, λ_2 and μ satisfying $\lambda_1 < \mu$ and $\lambda_2 < \mu$, a suitable α always exists. Moreover it gives the stationary distribution of $\bar{Q}^{\alpha,1-\alpha}$ as given by Theorem 1.

Theorem 5. *The system $\bar{Q}^{\alpha,1-\alpha}$ is ergodic iff*

$$\gamma_1^* < \alpha < 1 - \gamma_2^*, \tag{23}$$

in which case it has steady-state distribution

$$\bar{\pi}_\alpha(n, m) = [1 - \sigma_1(\alpha)]\,\sigma_1(\alpha)^n\,[1 - \sigma_2(\alpha)]\,\sigma_2(\alpha)^m, \tag{24}$$

where $\sigma_1(\alpha)$ and $\sigma_2(\alpha)$ are the unique solution of

$$\begin{aligned}&\alpha\sigma_1(\alpha)+(1-\alpha)\sigma_1(\alpha)\sigma_2(\alpha)=\rho_1,\\&(1-\alpha)\sigma_2(\alpha)+\alpha\sigma_1(\alpha)\sigma_2(\alpha)=\rho_2,\end{aligned} \tag{25}$$

satisfying $0<\sigma_1(\alpha)<1$ and $0<\sigma_2(\alpha)<1$. Given $\rho_1<1$ and $\rho_2<1$ it is always possible to choose α such that $\bar{Q}^{\alpha,1-\alpha}$ is ergodic.

Proof. The stability condition (23) follows directly from Theorem 4; the stationary distribution (24) follows from Theorem 1. For the last statement note that besides condition (23), we need $0\le\alpha\le 1$. Therefore, we need to prove that

$$(\gamma_1^*,1-\gamma_2^*)\cap[0,1]\neq\emptyset. \tag{26}$$

First we show that $(\gamma_1^*,1-\gamma_2^*)\neq\emptyset$ by proving that

$$\gamma_1^*=\frac{\rho_1-\rho_2}{1-\rho_2}<\frac{1-\rho_2}{1-\rho_1}=1-\gamma_2^* \tag{27}$$

for $0<\rho_1<1$ and $0<\rho_2<1$. This follows directly by rewriting the inequality as

$$0<\left(\rho_1-\frac{1}{2}\right)^2+\left(\rho_2-\frac{1}{2}\right)^2+\left(\rho_1\rho_2+\frac{1}{2}\right). \tag{28}$$

Finally, (26), follows from the observation that $\gamma_1^*<1$ and $1-\gamma_2^*>0$. □

Remark 5. The above theorem gives the stationary distribution in implicit form, *i.e.*, $\sigma_1(\alpha)$ and $\sigma_2(\alpha)$ are given as the solutions of a system of quadratic equations. The explicit solution is

$$\sigma_i(\alpha)=\frac{-b_i+\sqrt{b_i^2-4a_ic_i}}{2a_i}, \tag{29}$$

where

$$\begin{aligned}&a_1=\alpha, \qquad b_1=1-\alpha+\rho_2\frac{1-\alpha}{\alpha}-\rho_1, \qquad c_1=-\rho_1\frac{1-\alpha}{\alpha},\\&a_2=1-\alpha, \qquad b_2=\alpha+\rho_1\frac{\alpha}{1-\alpha}-\rho_2, \qquad c_2=-\rho_2\frac{\alpha}{1-\alpha}.\end{aligned}$$

It is readily verified that (29) provides the roots of the system of equation that satisfy $0<\sigma_i(\alpha)<1$, $i=1,2$. Details are provided in [32].

5.2. Delay

The first performance measure of interest is D^{γ_1,γ_2}, the expected delay of packets of the first session. We established in Section 2 that $D^{\gamma_1,\gamma_2}=\mathbb{E}[N]/\lambda_1$. Our first result gives the exact distribution of N if $\gamma_1=1$.

Theorem 6. *Let $P(N=n)$ denote the probability that $N=n$ in steady state. If $\gamma_1=1$ then*

$$P(N=n)=(1-\rho_1)\rho_1^n. \tag{30}$$

Proof. Consider the one-dimensional continuous Markov process in N, the number of packets in the first queue. The transition rates of this process are independent of the number of packets in the second queue. The process is equivalent to a M/M/1 queue with load ρ_1. □

We will make use of the comparison result of Theorem 3. The next technical lemma provides the required signs of the bias terms. The proof of the lemma is tedious, but mostly mechanical and therefore omitted here. It is based on induction in k, full details are provided in [32].

Lemma 1. *Let $d:\mathbb{N}_0^2\to[0,\infty)$, $d(n,m)=n/\lambda_1$. For all $(n,m)\in\mathbb{N}_0^2$ and $k\in\mathbb{N}_0$*

$$D^{\gamma_1,\gamma_2}(k,n+1,m)-D^{\gamma_1,\gamma_2}(k,n,m)\ge 0, \tag{31}$$

$$D^{\gamma_1,\gamma_2}(k,n,m+1)-D^{\gamma_1,\gamma_2}(k,n,m)\le 0. \tag{32}$$

The first use of the above lemma is in establishing a monotonicity result.

Theorem 7. *The expected delay of packets of the first session of Q^{γ_1,γ_2} is monotone in γ_1 and γ_2. More precisely,*

$$D^{\gamma_1,\gamma_2} \le D^{\tilde{\gamma}_1,\gamma_2}, \quad \text{if } \gamma_1 > \tilde{\gamma}_1, \tag{33}$$

$$D^{\gamma_1,\gamma_2} \ge D^{\gamma_1,\tilde{\gamma}_2}, \quad \text{if } \gamma_2 > \tilde{\gamma}_2. \tag{34}$$

Proof. Follows directly from Lemma 1 and Theorem 3 by observing that d does not depend on γ_1 and γ_2. □

Let

$$D^* = \inf\{D^{\gamma_1,\gamma_2}; \gamma_1 > \gamma_1^*, \gamma_2 > \gamma_2^*\}. \tag{35}$$

From Theorems 6 and 7 we directly obtain the value of D^*.

Corollary 1. *The minimum delay is $(\mu - \lambda_1)^{-1}$, i.e., $D^* = (\mu - \lambda_1)^{-1}$.*

Proof. From Theorem 7 it follows that the minimum of D^{γ_1,γ_2} is attained at $\gamma_1 = 1$. The result follows directly from Theorem 6. □

Next, we provide bounds on D^{γ_1,γ_2}. The fact that $D^{\gamma_1,\gamma_2}(k, n+1, 0) - D^{\gamma_1,\gamma_2}(k, n, 0)$ and $D^{\gamma_1,\gamma_2}(k, 0, m+1) - D^{\gamma_1,\gamma_2}(k, 0, m)$ have different signs provides the opportunity to obtain both upper and lower bounds on D^{γ_1,γ_2} using only the comparison result of Theorem 3. In particular, we can either choose $\alpha = \gamma_1$ or $\alpha = 1 - \gamma_2$ for the perturbed system, leading to an upper respectively a lower bound, or vice versa depending on the value of $\gamma_1 + \gamma_2$ as will become clear in Theorem 8. Since the perturbed system needs to be stable, *i.e.*, $\gamma_1^* \le \alpha \le 1 - \gamma_2^*$ from Theorem 5, it is not possible to obtain both upper and lower comparison bounds for all values of γ_1 and γ_2.

Theorem 8. *The expected delay of packets of the first session of Q^{γ_1,γ_2} is bounded as*

$$D^{\gamma_1,\gamma_2} \ge \frac{\sigma_1(\gamma_1)}{\lambda_1[1 - \sigma_1(\gamma_1)]}, \quad \text{if } \begin{cases} \gamma_1 + \gamma_2 \ge 1, \text{ and} \\ \gamma_1 < 1 - \gamma_2^*, \end{cases} \tag{36}$$

$$D^{\gamma_1,\gamma_2} \le \frac{\sigma_1(1-\gamma_2)}{\lambda_1[1 - \sigma_1(1-\gamma_2)]}, \quad \text{if } \begin{cases} \gamma_1 + \gamma_2 \ge 1, \text{ and} \\ \gamma_2 < 1 - \gamma_1^*, \end{cases} \tag{37}$$

$$D^{\gamma_1,\gamma_2} \le \frac{\sigma_1(\gamma_1)}{\lambda_1[1 - \sigma_1(\gamma_1)]}, \quad \text{if } \begin{cases} \gamma_1 + \gamma_2 \le 1, \text{ and} \\ \gamma_1 < 1 - \gamma_2^*, \end{cases} \tag{38}$$

$$D^{\gamma_1,\gamma_2} \ge \frac{\sigma_1(1-\gamma_2)}{\lambda_1[1 - \sigma_1(1-\gamma_2)]}, \quad \text{if } \begin{cases} \gamma_1 + \gamma_2 \le 1, \text{ and} \\ \gamma_2 < 1 - \gamma_1^*. \end{cases} \tag{39}$$

Proof. First consider $\alpha = \gamma_1$, which from condition (23) in Theorem 5 can be used only if $\gamma_1 < 1 - \gamma_2^*$. The performance of the perturbed system is given by

$$\bar{\mathbb{E}}^{\gamma_1,1-\gamma_1}[d(N, M)] = \bar{\mathbb{E}}^{\gamma_1,1-\gamma_1}[N/\lambda_1] = \frac{\sigma_1(\gamma_1)}{\lambda_1[1 - \sigma_1(\gamma_1)]}. \tag{40}$$

To decide whether the above expression provides an upper or a lower bound note that

$$q_{0,m}^{\gamma_1,\gamma_2}(0, -1) - \bar{q}_{0,m}^{\gamma_1,1-\gamma_1}(0, -1) \ge 0, \quad \text{if } \gamma_1 + \gamma_2 \ge 1,$$
$$q_{0,m}^{\gamma_1,\gamma_2}(0, -1) - \bar{q}_{0,m}^{\gamma_1,1-\gamma_1}(0, -1) \le 0, \quad \text{if } \gamma_1 + \gamma_2 \le 1.$$

The above inequalities, together with Lemma 1 and Theorem 3, lead to the bounds given in (36) and (38).

Next consider $\alpha = 1 - \gamma_2$. The ergodicity condition (23) reduces to $\gamma_2 < 1 - \gamma_1^*$. Finally, the inequalities

$$q_{n,0}^{\gamma_1,\gamma_2}(-1, 0) - \bar{q}_{n,0}^{\gamma_1,1-\gamma_1}(-1, 0) \ge 0, \quad \text{if } \gamma_1 + \gamma_2 \ge 1,$$
$$q_{n,0}^{\gamma_1,\gamma_2}(-1, 0) - \bar{q}_{n,0}^{\gamma_1,1-\gamma_1}(-1, 0) \le 0, \quad \text{if } \gamma_1 + \gamma_2 \le 1,$$

lead to bounds (37) and (39). □

Conjecture 1. *For all values $(n, m) \in \mathbb{N}_0^2$, $|D^{\gamma_1,\gamma_2}(k, n+1, 0) - D^{\gamma_1,\gamma_2}(k, n, 0)|$ and $|D^{\gamma_1,\gamma_2}(k, 0, m+1) - D^{\gamma_1,\gamma_2}(k, 0, m)|$ are bounded uniformly in k.*

Remark 6. The bounds of Theorem 8 are only valid for specific ranges of values for γ_1 and γ_2. If Conjecture 1 holds it is possible to obtain error bounds on the expected delay using Theorem 2. The bounds would be valid for all parameter ranges and possibly better in the ranges already covered by Theorem 8.

5.3. Energy consumption

The second performance measure for which we obtain bounds is the expected energy consumption $C^{\gamma_1,\gamma_2} = C^{\gamma_1,\gamma_2}_{\gamma_1,\gamma_2} = \mathbb{E}^{\gamma_1,\gamma_2}[c_{\gamma_1,\gamma_2}(N, M)]$ with cost function c_{γ_1,γ_2} defined in (6). In addition to the sign of the bias terms we are also able to obtain bounds on their size. As for the proof of Lemma 1 the proof follows straightforwardly using induction in k. Details are omitted here, but provided in [32].

Lemma 2. *Let* $c_{\gamma_1,\gamma_2} : \mathbb{N}_0^2 \to [0, \infty)$,

$$c_{\gamma_1,\gamma_2}(n, m) = \gamma_1\mu\mathbb{1}_{\{n>0,m=0\}} + \gamma_2\mu\mathbb{1}_{\{n=0,m>0\}} + \mu\mathbb{1}_{\{n>0,m>0\}}.$$

For all $(n, m) \in \mathbb{N}_0^2$ *and* $k \in \mathbb{N}_0$

$$0 \le C^{\gamma_1,\gamma_2}_{\gamma_1,\gamma_2}(k, n+1, m) - C^{\gamma_1,\gamma_2}_{\gamma_1,\gamma_2}(k, n, m) \le 1, \tag{41}$$

$$0 \le C^{\gamma_1,\gamma_2}_{\gamma_1,\gamma_2}(k, n, m+1) - C^{\gamma_1,\gamma_2}_{\gamma_1,\gamma_2}(k, n, m) \le 1. \tag{42}$$

The signs of the bias terms can be used to establish the following monotonicity result.

Theorem 9. *The energy consumption of* Q^{γ_1,γ_2} *is monotone in* γ_1 *and* γ_2. *More precisely,*

$$C^{\gamma_1,\gamma_2}_{\gamma_1,\gamma_2} \ge C^{\tilde{\gamma}_1,\gamma_2}_{\tilde{\gamma}_1,\gamma_2}, \quad \text{if } \gamma_1 > \tilde{\gamma}_1, \tag{43}$$

$$C^{\gamma_1,\gamma_2}_{\gamma_1,\gamma_2} \ge C^{\gamma_1,\tilde{\gamma}_2}_{\gamma_1,\tilde{\gamma}_2}, \quad \text{if } \gamma_2 > \tilde{\gamma}_2. \tag{44}$$

Proof. We use Theorem 3 to compare Q^{γ_1,γ_2} under cost function c_{γ_1,γ_2} with $Q^{\tilde{\gamma}_1,\gamma_2}$ under cost function $c_{\tilde{\gamma}_1,\gamma_2}$, where $\gamma_1 > \tilde{\gamma}_1$. Note that contrary to the monotonicity result of Theorem 7, the reward function in the perturbed model is different from the original reward function. Using Lemma 2 we obtain, for $n > 0$

$$\begin{aligned}\sum_{i,j}\left[q_{n,0}(i,j) - \bar{q}_{n,0}(i,j)\right][C(k, n+i, j) - C(k, n, 0)] &= (\gamma_1 - \tilde{\gamma}_1)\mu\,[C(k, n-1, 0) - C(k, n, 0)] \\ &\ge -(\gamma_1 - \tilde{\gamma}_1)\mu.\end{aligned} \tag{45}$$

Therefore,

$$c_{\gamma_1,\gamma_2}(n, 0) - c_{\tilde{\gamma}_1,\gamma_2}(n, 0) + \sum_{i,j}\left[q_{n,0}(i,j) - \bar{q}_{n,0}(i,j)\right]\cdot[C(k, n+i, j) - C(k, n, 0)] \ge 0, \tag{46}$$

and (43) follows from Theorem 3. Monotonicity in γ_2 follows in similar fashion. □

Let

$$C^* = \inf\{C^{\gamma_1,\gamma_2}; \gamma_1 > \gamma_1^*, \gamma_2 > \gamma_2^*\}. \tag{47}$$

An exact expression for C^* is given by the next result.

Theorem 10.

$$C^* = \max\{\lambda_1, \lambda_2\}. \tag{48}$$

Proof. First, assume that $\lambda_1 > \lambda_2$. Under this assumption $\gamma_1^* > 0$ and $\gamma_2^* < 0$. By Theorem 9 it follows that

$$C^* = \lim_{\gamma_1 \to \gamma_1^*} C^{\gamma_1,0}_{\gamma_1,0}. \tag{49}$$

Let $0 < \epsilon < 1 - \gamma_1^*$ and consider the sequence of processes $\left\{Q^{\gamma_1(l),0}\right\}_{l\in\mathbb{N}}$, $\gamma_1(l) = \gamma_1^* + \epsilon^l$. For each $l \in \mathbb{N}$ we give an approximation on $C^{\gamma_1(l),0}_{\gamma_1(l),0}$, the energy consumption of $Q^{\gamma_1(l),0}$. In particular, we show that $C^{\gamma_1(l),0}_{\gamma_1(l),0} \to \lambda_1$ as $l \to \infty$. Therefore, consider the sequence of perturbed processes $\left\{\bar{Q}^{\gamma_1(l),1-\gamma_1(l)}\right\}_{l\in\mathbb{N}}$. It follows from Theorem 11 that

$$\bar{C}^{\gamma_1(l),1-\gamma_1(l)}_{\gamma_1(l),0} - \delta^{\gamma_1(l),1-\gamma_1(l)}_{\gamma_1(l),0} \le C^{\gamma_1(l),0}_{\gamma_1(l),0} \le \bar{C}^{\gamma_1(l),1-\gamma_1(l)}_{\gamma_1(l),0} + \delta^{\gamma_1(l),1-\gamma_1(l)}_{\gamma_1(l),0}, \tag{50}$$

where $\delta^{\gamma_1(l),1-\gamma_1(l)}_{\gamma_1(l),0}$ is defined in Theorem 11. It is readily verified that

$$\sigma_1\left(\gamma_1(l)\right) \longrightarrow 1, \quad \text{and} \quad \sigma_2\left(\gamma_1(l)\right) \longrightarrow \rho_2, \tag{51}$$

as $l \to \infty$. Therefore, $\delta^{\gamma_1(l),1-\gamma_1(l)}_{\gamma_1(l),0}$ vanishes as $l \to \infty$. From (50) and (51) it follows that

$$C^* = \lim_{l\to\infty} \bar{C}^{\gamma_1(l),1-\gamma_1(l)}_{\gamma_1(l),0} = \gamma_1^*\mu(1-\rho_2) + \mu\rho_2 = \lambda_1. \tag{52}$$

Next, we need to consider the cases $\lambda_1 < \lambda_2$ and $\lambda_1 = \lambda_2$. In similar fashion as the first case it follows for $\lambda_1 < \lambda_2$ that

$$C^* = \lim_{\gamma_2\to\gamma_2^*} C^{0,\gamma_2}_{0,\gamma_2} = \lim_{\gamma_2\to\gamma_2^*} \bar{C}^{1-\gamma_2,\gamma_2}_{0,\gamma_2} = \lambda_2 \tag{53}$$

and for $\lambda_1 = \lambda_2 = \lambda$ that

$$C^* = \lim_{\gamma\to 0} C^{\gamma,\gamma}_{\gamma,\gamma} = \lim_{\gamma\to 0} \bar{C}^{\gamma,1-\gamma}_{\gamma,\gamma} = \lambda. \quad \square \tag{54}$$

Since we have the signs as well as a bound on the value of the bias terms we can employ both the comparison result of Theorem 3 and the error bound result from Theorem 2. Some care needs to be taken in choosing the value of α for the perturbed model $\bar{Q}^{\alpha,1-\alpha}$. We will see that by restricting α to the range $[\min\{\gamma_1, 1-\gamma_2\}, \max\{\gamma_1, 1-\gamma_2\}]$, it is possible to employ Theorem 3.

Theorem 11. *Let* $\min\{\gamma_1, 1-\gamma_2\} \le \tilde{\alpha} \le \max\{\gamma_1, 1-\gamma_2\}$. *Then*

$$\bar{C}^{\tilde{\alpha},1-\tilde{\alpha}}_{\gamma_1,\gamma_2} - \delta^{\tilde{\alpha},1-\tilde{\alpha}}_{\gamma_1,\gamma_2}\mathbb{1}_{\{\gamma_1+\gamma_2\ge 1\}} \le C^{\gamma_1,\gamma_2}_{\gamma_1,\gamma_2} \le \bar{C}^{\tilde{\alpha},1-\tilde{\alpha}}_{\gamma_1,\gamma_2} + \delta^{\tilde{\alpha},1-\tilde{\alpha}}_{\gamma_1,\gamma_2}\mathbb{1}_{\{\gamma_1+\gamma_2\le 1\}}, \tag{55}$$

where

$$\bar{C}^{\tilde{\alpha},1-\tilde{\alpha}}_{\gamma_1,\gamma_2} = \gamma_1\mu\sigma_1(\tilde{\alpha})\left[1-\sigma_2(\tilde{\alpha})\right] + \gamma_2\mu\left[1-\sigma_1(\tilde{\alpha})\right]\sigma_2(\tilde{\alpha}) + \mu\sigma_1(\tilde{\alpha})\sigma_2(\tilde{\alpha}), \tag{56}$$

$$\delta^{\tilde{\alpha},1-\tilde{\alpha}}_{\gamma_1,\gamma_2} = |\tilde{\alpha}-\gamma_1|\mu\sigma_1(\tilde{\alpha})\left[1-\sigma_2(\tilde{\alpha})\right] + |1-\tilde{\alpha}-\gamma_2|\mu\left[1-\sigma_1(\tilde{\alpha})\right]\sigma_2(\tilde{\alpha}). \tag{57}$$

Proof. The expected energy consumption of Q^{γ_1,γ_2} will be bounded in terms of

$$\bar{C}^{\tilde{\alpha},1-\tilde{\alpha}}_{\gamma_1,\gamma_2} = \mathbb{E}^{\tilde{\alpha},1-\tilde{\alpha}}\left[c_{\gamma_1,\gamma_2}(N,M)\right], \tag{58}$$

the value of which can be easily computed based on the stationary distribution of $\bar{Q}^{\tilde{\alpha},1-\tilde{\alpha}}$ given in Theorem 5.

If $\gamma_1+\gamma_2 \le 1$ then $\gamma_1 - \tilde{\alpha} \le 0$ and $\gamma_2 - (1-\tilde{\alpha}) \le 0$. Therefore, by Lemma 2 and Theorem 3, $C^{\gamma_1,\gamma_2}_{\gamma_1,\gamma_2} \ge \bar{C}^{\tilde{\alpha},1-\tilde{\alpha}}_{\gamma_1,\gamma_2}$. If, on the other hand, $\gamma_1+\gamma_2 \ge 1$, then $\gamma_1-\tilde{\alpha} \ge 0$ and $\gamma_2-(1-\tilde{\alpha}) \ge 0$, and it follows that $C^{\gamma_1,\gamma_2}_{\gamma_1,\gamma_2} \le \bar{C}^{\tilde{\alpha},1-\tilde{\alpha}}_{\gamma_1,\gamma_2}$.

It remains to show that

$$\bar{C}^{\tilde{\alpha},1-\tilde{\alpha}}_{\gamma_1,\gamma_2} - \delta^{\tilde{\alpha},1-\tilde{\alpha}}_{\gamma_1,\gamma_2} \le C^{\gamma_1,\gamma_2}_{\gamma_1,\gamma_2} \le \bar{C}^{\tilde{\alpha},1-\tilde{\alpha}}_{\gamma_1,\gamma_2} + \delta^{\tilde{\alpha},1-\tilde{\alpha}}_{\gamma_1,\gamma_2}. \tag{59}$$

We will use Theorem 2 with $\bar{c} = c_{\gamma_1,\gamma_2}$ and hence are required to find a function $\xi : \mathbb{N}_0^2 \to [0,\infty)$ that satisfies

$$\left|\sum_{i,j}\left[q^{\gamma_1,\gamma_2}_{n,m}(i,j) - \bar{q}^{\tilde{\alpha},1-\tilde{\alpha}}_{n,m}(i,j)\right]\cdot\left[C(k,n+i,m+j) - C(k,n,m)\right]\right| \le \xi(n,m) \tag{60}$$

and

$$\sum_{n,m}\bar{\pi}(n,m)\xi(n,m) = \delta^{\tilde{\alpha},1-\tilde{\alpha}}_{\gamma_1,\gamma_2}. \tag{61}$$

From Lemma 2 and the definitions of Q^{γ_1,γ_2} and $\bar{Q}^{\tilde{\alpha},1-\tilde{\alpha}}$ it follows that

$$\xi(n,m) = \begin{cases} |\gamma_1-\tilde{\alpha}|, & \text{if } n>0, m=0,\\ |\gamma_2-(1-\tilde{\alpha})|, & \text{if } n=0, m>0,\\ 0, & \text{otherwise} \end{cases} \tag{62}$$

satisfies (60). A simple computation, using the product-form distribution $\bar{\pi}(\tilde{\alpha})$, shows that (61) is also satisfied. $\square$

Remark 7. We have limited $\tilde{\alpha}$ to the interval $[\min\{\gamma_1, 1-\gamma_2\}, \max\{\gamma_1, 1-\gamma_2\}]$, making sure that Theorem 3 can be used. Obviously, $\tilde{\alpha}$ also needs to satisfy $0 \le \tilde{\alpha} \le 1$ and $\gamma_1^* < \tilde{\alpha} < 1-\gamma_2^*$. It is readily verified that there always exists an $\tilde{\alpha}$ that satisfies all constraints, *i.e.*, Theorem 11 provides upper and lower bounds for all values of the process parameters.

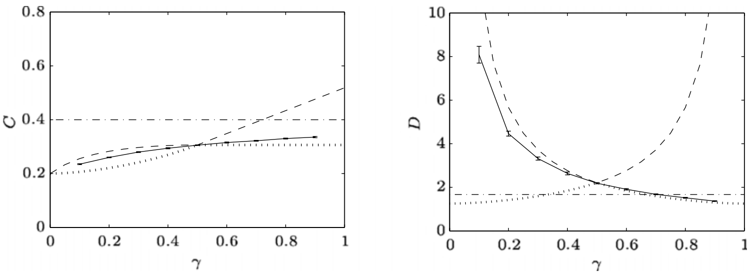

Fig. 2. Performance of a symmetric system $Q^{\gamma,\gamma}$ under low load ($\mu = 1, \lambda_1 = \lambda_2 = 0.2, \gamma_1 = \gamma_2 = \gamma$). Depicted are the analytical lower (dotted lines) and upper bounds (dashed lines), the simulation results (solid lines), and the performance of the corresponding uncoded system (dashed–dotted lines).

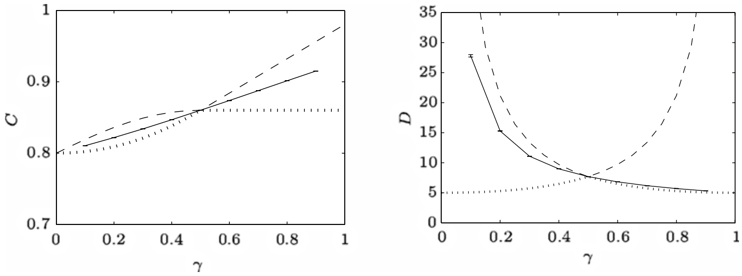

Fig. 3. Performance of a symmetric system $Q^{\gamma,\gamma}$ under high load ($\mu = 1, \lambda_1 = \lambda_2 = 0.8, \gamma_1 = \gamma_2 = \gamma$). Depicted are the analytical lower (dotted lines) and upper bounds (dashed lines), and simulation result (solid lines).

Remark 8. Outside the interval $[\min\{\gamma_1, 1-\gamma_2\}, \max\{\gamma_1, 1-\gamma_2\}]$ there are values of $\hat{\alpha}$ that still satisfy $\gamma_1^* < \hat{\alpha} < 1-\gamma_2^*$. For these values of $\hat{\alpha}$, Theorem 3 cannot be used, but Theorem 2 is still valid. This would lead to bounds of the form

$$\bar{C}^{\hat{\alpha},1-\hat{\alpha}}_{\gamma_1,\gamma_2} - \delta^{\hat{\alpha},1-\hat{\alpha}}_{\gamma_1,\gamma_2} \le C^{\gamma_1,\gamma_2}_{\gamma_1,\gamma_2} \le \bar{C}^{\hat{\alpha},1-\hat{\alpha}}_{\gamma_1,\gamma_2} + \delta^{\hat{\alpha},1-\hat{\alpha}}_{\gamma_1,\gamma_2}. \tag{63}$$

For clarity of exposition bounds of this type have been omitted in the current work.

Remark 9. The bounds given in Theorem 11 depend on $\tilde{\alpha}$. In practice one will want to use the $\tilde{\alpha}$ that provides the tightest bound. By distinguishing between $\gamma_1+\gamma_2 \le 1$ and $\gamma_1+\gamma_2 \ge 1$ as well as lower and upper bounds, we obtain optimization problems that can be readily solved. For the upper bounds in the case that $\gamma_1+\gamma_2 \le 1$ we need to find, for instance,

$$\min\left\{\bar{C}^{\tilde{\alpha},1-\tilde{\alpha}}_{\gamma_1,\gamma_2} + \delta^{\tilde{\alpha},1-\tilde{\alpha}}_{\gamma_1,\gamma_2} : \gamma_1 \le \tilde{\alpha} \le 1-\gamma_2\right\}. \tag{64}$$

This corresponds to finding the minimum of the function

$$\begin{aligned}&\gamma_1\mu\sigma_1(\tilde{\alpha})\left[1-\sigma_2(\tilde{\alpha})\right] + \gamma_2\mu\left[1-\sigma_1(\tilde{\alpha})\right]\sigma_2(\tilde{\alpha}) + \mu\sigma_1(\tilde{\alpha})\sigma_2(\tilde{\alpha})\\ &\quad + (\tilde{\alpha}-\gamma_1)\mu\sigma_1(\tilde{\alpha})\left[1-\sigma_2(\tilde{\alpha})\right] + (1-\tilde{\alpha}-\gamma_2)\mu\left[1-\sigma_1(\tilde{\alpha})\right]\sigma_2(\tilde{\alpha}),\end{aligned} \tag{65}$$

in the interval $\gamma_1 \le \tilde{\alpha} \le 1-\gamma_2$, which is tedious, but simple calculus.

6. Numerical examples

We provide some examples of application of Theorems 8 and 11 for specific parameter values. In addition to the values of the analytical bounds we provide numerical results obtained through simulation. The simulation results have been obtained within 99% confidence intervals. These confidence intervals are depicted in all figures, but they are sometimes so small that they are not visible. Throughout this section we assume $\mu = 1$. Remember that the performance of the uncoded system is given in (8) and (9).

First we consider a symmetric system with $\lambda_1 = \lambda_2 = \lambda$ and $\gamma_1 = \gamma_2 = \gamma$. We fix λ and consider the performance of $Q^{\gamma,\gamma}$ as a function of γ. We consider $Q^{\gamma,\gamma}$ under two scenario's. The first scenario is that of a relatively low load of $\lambda = 0.2$, the results of which are depicted in Fig. 2. In addition to the analytical bounds and the simulation result, we provide the performance of the uncoded system. It is interesting to note that the delay in the coded system is only smaller than that of the uncoded system for large values of γ. For these values of γ the energy savings are significantly smaller than the 50% that are theoretically possible. The second scenario that we consider for $Q^{\gamma,\gamma}$ is that of a relatively high load of $\lambda = 0.8$. The results of which are depicted in Fig. 3. Note, that for $\lambda = 0.8$, the uncoded system is not stable.

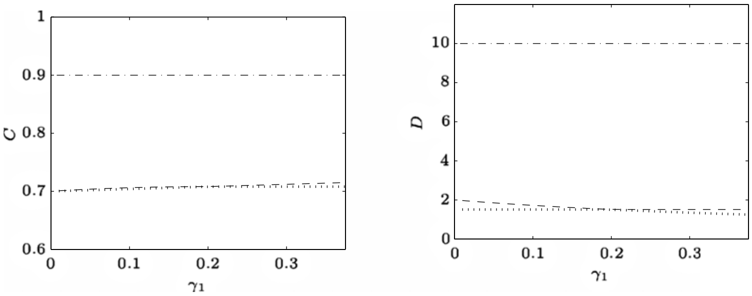

Fig. 4. Performance of an asymmetric system ($\mu = 1, \lambda_1 = 0.2, \lambda_2 = 0.7, \gamma_2 = 0.8$). Depicted are the analytical lower (dotted line) and upper bounds (dashed line), and the performance in an uncoded system (dashed–dotted line).

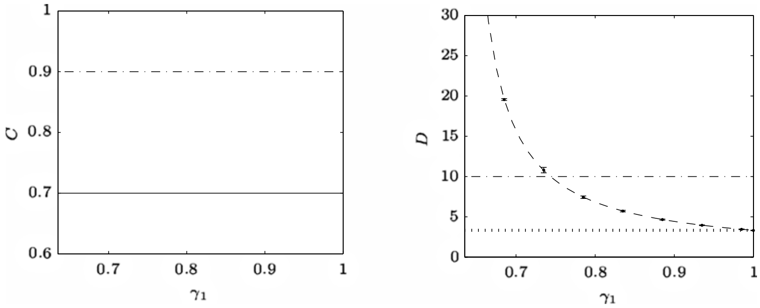

Fig. 5. Performance of an asymmetric system ($\mu = 1, \lambda_1 = 0.7, \lambda_2 = 0.2, \gamma_2 = 0$). The upper and lower bounds on the energy consumption coincide and are depicted in a solid line. The lower and upper bounds on the delay are depicted in a dotted and a dashed line, respectively. The simulation result of the delay nearly coincides with the upper bound, hence only the confidence intervals are depicted. The performance of the uncoded system is depicted in dashed–dotted lines.

Next, we consider a system in which $\lambda_1 < \lambda_2$. In particular, we consider $\lambda_1 = 0.2$, $\lambda_2 = 0.7$ and $\gamma_2 = 0.8$, and analyze the influence of γ_1. In Fig. 4 we give the analytical performance bounds of $Q^{\gamma_1,0.8}$ and the performance of the uncoded system. Since, the upper and lower bounds nearly coincide, we have omitted the simulation results. It can be observed that the coded system is performing significantly better, in terms of energy consumption as well as delay.

Finally, we consider a system in which $\lambda_1 > \lambda_2$. In particular, we consider $\lambda_1 = 0.7$, $\lambda_2 = 0.2$ and $\gamma_2 = 0$, and analyze the influence of γ_1. In Fig. 5 we give the analytical performance bounds of $Q^{\gamma_1,0}$ and the performance of the uncoded system.

7. Discussion

We have provided a queueing analysis of a two-way relay in which network coding is employed. We have compared the energy consumption and the delay in the relay with that of a relay in which network coding is not used by deriving analytical upper and lower bounds on the performance of the coded system. It is shown that different operating policies can be used to tradeoff energy consumption against delay. Exact results have been obtained on the minimum possible energy consumption and the minimum possible delay.

The queueing model that we have studied in this work has similar properties as a queueing network with positive and negative customers. As potential future work it is of interest to consider generalizations to other network configurations with more relays and more connections. In that case it might be necessary to consider generalizations of networks with positive and negative customers, see, for instance, [23,33–36].

Acknowledgment

This work was supported in part by NWO grant 612.001.107.

References

[1] C. Fragouli, J.-Y. Le Boudec, J. Widmer, Network coding: an instant primer, ACM SIGCOMM Computer Communication Review 36 (1) (2006) 63–68.
[2] M. Médard, A. Sprintson, Network coding: Fundamentals and Applications, Academic Press, 2011.
[3] Y. Wu, P.A. Chou, S.-Y. Kung, Information exchange in wireless networks with network coding and physical-layer broadcast, in: Proc. 39th Annual Conference on Information Sciences and Systems, CISS, 2005.
[4] P. Larsson, N. Johansson, K.-E. Sunell, Coded bi-directional relaying, in: IEEE 63rd Vehicular Technology Conference, VTC 2006-Spring, 2006, pp. 851–855.

[5] S.J. Kim, P. Mitran, V. Tarokh, Performance bounds for bidirectional coded cooperation protocols, IEEE Transactions on Information Theory 54 (11) (2008) 5235–5241.
[6] W. Chen, K. Letaief, Z. Cao, Opportunistic network coding for wireless networks, in: IEEE International Conference on Communications, ICC'07, IEEE, 2007, pp. 4634–4639.
[7] Y.E. Sagduyu, A. Ephremides, Cross-layer optimization of MAC and network coding in wireless queueing tandem networks, IEEE Transactions on Information Theory 54 (2) (2008) 554–571.
[8] X. He, A. Yener, On the energy-delay trade-off of a two-way relay network, in: 42nd Annual Conference on Information Sciences and Systems, CISS 2008, 2008, pp. 865–870.
[9] C.-H. Liu, F. Xue, Network coding for two-way relaying: rate region, sum rate and opportunistic scheduling, in: IEEE International Conference on Communications, ICC'08, 2008, pp. 1044–1049.
[10] O. Abdelrahman, E. Gelenbe, Queueing performance under network coding, in: IEEE Information Theory Workshop on Networking and Information Theory, ITW 2009, 2009, pp. 135–139.
[11] O. Abdelrahman, E. Gelenbe, Performance trade-offs in a network coding router, in: Proceedings of 19th International Conference on Computer Communications and Networks, ICCCN, 2010, pp. 1–6.
[12] Y. Yuan, K. Wu, W. Jia, Y. Peng, On the queueing behavior of inter-flow asynchronous network coding, Computer Communications 35 (13) (2012) 1535–1548.
[13] D. Umehara, T. Hirano, S. Denno, M. Morikura, T. Sugiyama, Wireless network coding in slotted ALOHA with two-hop unbalanced traffic, IEEE Journal on Selected Areas in Communications 27 (5) (2009) 647–661.
[14] S. Chieochan, E. Hossain, T. Issariyakul, D. Niyato, Opportunistic network coding and dynamic buffer allocation in a wireless butterfly network, in: IEEE Global Telecommunications Conference, GLOBECOM, 2009, pp. 1–6.
[15] P. Parag, J.-F. Chamberland, Queueing analysis of a butterfly network for comparing network coding to classical routing, IEEE Transactions on Information Theory 56 (4) (2010) 1890–1908.
[16] E. Gelenbe, Product-form queueing networks with negative and positive customers, Journal of Applied Probability 28 (3) (1991) 656–663.
[17] N.M. Van Dijk, Error bounds and comparison results: the Markov reward approach for queueing networks, in: R.J. Boucherie, N.M. Van Dijk (Eds.), Queueing Networks: A Fundamental Approach, in: International Series in Operations Research & Management Science, vol. 154, Springer, 2011.
[18] E. Gelenbe, C. Morfopoulou, Power savings in packet networks via optimised routing, Mobile Networks and Applications 17 (1) (2012) 152–159.
[19] E. Gelenbe, C. Morfopoulou, A framework for energy-aware routing in packet networks, The Computer Journal 54 (6) (2011) 850–859.
[20] E. Gelenbe, Steps toward self-aware networks, Communications of the ACM 52 (7) (2009) 66–75.
[21] E. Gelenbe, Energy packet networks: adaptive energy management for the cloud, in: Proceedings of the 2nd International Workshop on Cloud Computing Platforms, 2012.
[22] E. Gelenbe, H. Shachnai, On G-networks and resource allocation in multimedia systems, European Journal of Operational Research 126 (2) (2000) 308–318.
[23] E. Gelenbe, The first decade of G-networks, European Journal of Operational Research 126 (2) (2000) 231–232.
[24] T.V. Do, Bibliography on G-networks, negative customers and applications, Mathematical and Computer Modelling: An International Journal 53 (1–2) (2011) 205–212.
[25] J. Goseling, R.J. Boucherie, J.C.W. van Ommeren, Energy consumption in coded queues for wireless information exchange, in: Workshop on Network Coding, Theory, and Applications, 2009, NetCod'09, IEEE, 2009, pp. 30–35.
[26] R.J. Boucherie, N.M. van Dijk, Local balance in queueing networks with positive and negative customers, Annals of Operations Research 48 (5) (1994) 463–492.
[27] E. Gelenbe, R. Schassberger, Stability of product form G-networks, Probability in the Engineering and Informational Sciences 6 (3) (1992) 271–276.
[28] N.M. van Dijk, Perturbation theory for unbounded Markov reward processes with applications to queueing, Advances in Applied Probability 20 (1) (1988) 99–111.
[29] N.M. van Dijk, M.L. Putterman, Perturbation theory for Markov reward processes with applications to queueing systems, Advances in Applied Probability 20 (1) (1988) 79–98.
[30] G. Fayolle, R. Iasnogorodski, V. Malyshev, Random Walks in the Quarter Plane: Algebraic Methods, Boundary Value Problems, and Applications, Springer-Verlag, 1999.
[31] J.W. Cohen, O.J. Boxma, Boundary Value Problems in Queueing System Analysis, North-Holland, 1983.
[32] J. Goseling, R.J. Boucherie, J.C.W. van Ommeren, Energy-delay tradeoff in wireless network coding, Technical Report 1942, Department of Applied Mathematics, University of Twente, 2011. URL http://eprints.eemcs.utwente.nl/20173/.
[33] R. Serfozo, B. Yang, Markov network processes with string transitions, The Annals of Applied Probability 8 (3) (1998) 793–821.
[34] R.J. Boucherie, X. Chao, Queueing networks with string transitions of mixed vector additions and vector removals, Journal of Systems Science and Complexity 14 (4) (2001) 337–355.
[35] N. Bayer, O.J. Boxma, Wiener-hopf analysis of an M/G/1 queue with negative customers and of a related class of random walks, Queueing Systems 23 (1) (1996) 301–316.
[36] Y.W. Shin, Multi-server retrial queue with negative customers and disasters, Queueing Systems 55 (4) (2007) 223–237.

Jasper Goseling received a Ph.D. degree from Delft University of Technology (The Netherlands) in 2010. He is currently employed as a postdoctoral researcher at the University of Twente (The Netherlands) and at Delft University of Technology. His general research interest is in the areas of information theory, communication networks and queueing theory.

Richard J. Boucherie is a Professor of Stochastic Operations Research at the University of Twente (The Netherlands). After an ERCIM (European Research Consortium for Informatics and Mathematics) fellowship at INRIA-Sophia Antipolis, and CWI, Amsterdam, in 1992–1993, and a HCM (Human Capital and Mobility) fellowship again at INRIA-Sophia Antipolis in 1994, he was awarded a KNAW (Royal Netherlands Academy of Arts and Sciences) fellowship in 1994, which he spent at the Universiteit van Amsterdam, department of Operations Research. Richard's research interests are in various aspects of (wireless) telecommunications networks, logistics systems, and health care, and in (applications of) stochastic processes such as Petri nets, and queueing networks.

Jan-Kees van Ommeren is an Assistant Professor in the Stochastic Operations Research group at the University of Twente (The Netherlands). His research interests are in stochastic modelling, queuing theory and approximation algorithms, both theoretically and applied on telecommunication systems and logistic systems.