Network Coding Decisions for Wireless Transmissions with Delay Consideration

Arupa Mohapatra, Natarajan Gautam, *Senior Member, IEEE,* Srinivas Shakkottai, *Member, IEEE,* and Alex Sprintson, *Member, IEEE*

***Abstract*—We consider a relay node that stochastically receives packets from two opposing flows. Whenever opportunities exist, the relay performs network coding to efficiently transmit packets. However, on one hand, because of the stochastic nature as well as possible asymmetry between the opposing flows, it would not be possible to always code packets. On the other hand, waiting for a coding opportunity could result in excessive latency, and one may be better off transmitting packets without coding. Thus, one needs to decide at each transmission opportunity whether to transmit a packet uncoded or wait for a future transmission opportunity. To enable us to optimally make that decision, we consider costs for transmission and delay, and formulate our problem as a Markov decision process. We show that the optimal policy is threshold type under a sufficient condition, and we compute it by modeling the resulting system as a Markov chain. Through numerical analysis, we show the effectiveness of the threshold policy in the relay node network as well as in a line network scenario. Further, we compare the threshold policy against a number of simple heuristic policies and identify situations where these policies can be effective.**

***Index Terms*—Network coding, energy-delay trade-off, Markov decision processes**

I. Introduction

Over recent years, there has been an increasing interest in the applications of network coding in multihop wireless networks (see [1] for important application areas). Network coding techniques can significantly reduce the transmission load in wireless networks [2]. For example, consider the two-way relay network shown in Fig. 1(a). Here nodes 1 and 2 want to exchange a pair of packets x_1 and x_2 through node 3 which works as a relay node. In the conventional *store-and-forward* approach, node 1 sends its packet x_1 to node 3 which then forwards it to node 2. Similarly packet x_2 is sent from node 2 to node 1 via node 3 in two transmissions. However, in the network coding approach, once the two packets x_1 and x_2 are received at node 3, they are combined by a bit-wise XOR operation, and then the coded packet $x_1 \oplus x_2$ is broadcast to nodes 1 and 2 simultaneously. Now nodes 1 and 2 can get their required packets by decoding the coded packet. Note that, in this case, a total of 3 transmissions are required compared to 4 transmissions in the conventional approach. In another example, consider a line network with two information flows in opposite directions (see Fig. 1(b)). In this case, network

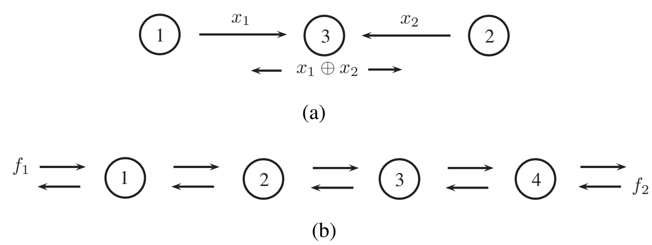


Fig. 1. Wireless network coding in (a) two-way relay network (b) line network

coding at successive nodes (known as "reverse carpooling" [3]) allows both flows to share one common path and achieves significant reduction in the number of transmissions in the network.

Given the ability of network coding to reduce transmission load greatly, high energy savings are possible by using this technique. However, in most wireless networks, data flows on different links vary significantly. Hence, coding opportunities are not always available, and waiting for such opportunities can cause substantial delay in transmission. In fact, the energy savings achieved through network coding may be offset by delays incurred by waiting for coding opportunities. Therefore, though the system needs to take advantage of network coding, it must also ensure an acceptable delay for the packets. For this, it is required to decide whether to transmit a packet without network coding or to wait for a future coding opportunity. To make such decisions, we aim to develop a model to optimally trade-off between low energy consumption and high quality-of-service (i.e. low delay).

In this paper, we consider the energy-delay trade-off issue in network coding in a two-way relay network (see Fig. 1(a)), which is a basic component of larger networks. Our objective is to make network coding decisions at the relay node in such a way that the average energy and delay costs are minimized over the long-run. To achieve this, we formulate our problem as a Markov decision process (MDP) which takes into account the energy and delay costs as well as the uncertainty in packet arrival processes. A policy for this MDP specifies how many coded and uncoded transmissions are to be made at a transmission opportunity based on the queue backlogs at the relay node. We aim to find the optimal policy in this MDP and develop insights into other simpler policies that can be effective. We will also apply these policies in other network settings such as the one in Fig. 1(b).

A. Mohapatra is with Oracle Corporation, Bala Cynwyd, PA 19004, USA (email: arupa.mohapatra@oracle.com). N. Gautam is with the Department of Industrial and Systems Engineering, and S. Shakkottai and A. Sprintson are with the Department of Electrical and Computer Engineering at Texas A&M University, College Station, TX 77843, USA (email: gautam@tamu.edu; sshakkot@tamu.edu; spalex@tamu.edu)

A. Related Work

Network coding has attracted significant interest from the research community since its introduction in the seminal work of Ahlswede *et al.* [4]. However, the energy-delay trade-off issue in network coding has received attention only very recently. He and Yener [5] analyzed this performance trade-off in a two-way relay under a simple a first-come-first-serve (FCFS) policy. Among other similar works that use different models and analyses, Abdelrahman and Gelenbe [6], and Gunasekara *et al.* [7] analyzed the energy-delay performance using waiting-time based policies while Chen *et al.* [8] and Goseling *et al.* [9] considered policies that make uncoded transmissions using certain probabilities. Note that most of these works analyze the energy-delay trade-off under an assumed operating policy and then tune the parameters of the policy to achieve the best performance. There have been very few attempts to derive the operating policy itself to achieve the optimal trade-off. Ciftcioglu *et al.* [10] developed game-based distributed policies to optimize the energy-delay performance in a different relay network setup. In our earlier work [11], we developed an MDP based energy-delay trade-off model in a simpler case with a maximum transmission capacity of one packet per time-slot at the relay node. Compared to [11], we consider in this paper a more general problem with different packet arrival and transmission models, which results in significantly different analysis and algorithms.

Markov decision processes (MDP) serve as efficient methods to optimize cost-performance trade-off in a stochastic environment in different applications [12]. Arapostathis *et al.* [13] provide an extensive survey of works on discrete-time average cost MDPs. In particular, as in our case, it is usually difficult to find an optimal stationary policy (which may also not exist) in average cost MDPs with countably infinite number of states (hereafter we will say 'countable' to mean countably infinite) and unbounded costs. The works by Borkar [14], Cavazos-Cadena [15], Cavasoz-Cadena and Sennott [16], Sennott [17], [18], and Schäl [19] are major contributions to the theory of MDPs with countable state space.

B. Contribution

The main focus of this paper is to make efficient "transmit or wait" decisions for packets at the relay node at every transmission opportunity. The problem that we consider is new as compared to the related works in the literature. The structure of our MDP is also uncommon, and the two-dimensional countable state space makes it quite challenging to solve for the optimal policy. We exploit the structure of the problem and use certain convexity concepts in a novel way to derive structural properties of the optimal policy. We prove that, under a sufficient condition, the optimal network coding policy for the relay node is a threshold based policy. We develop an analytical approach to compute this threshold policy. We also show the effectiveness of this policy in situations where our model assumptions are not satisfied. Moreover, we show that some simple policies (e.g. see "transmit-all" and "rate-based" policies in Section V) can be as efficient as the threshold policy in particular situations. These policies are easy to implement, and there is no computational overhead when system parameters change. This makes these policies particularly attractive for practical implementation.

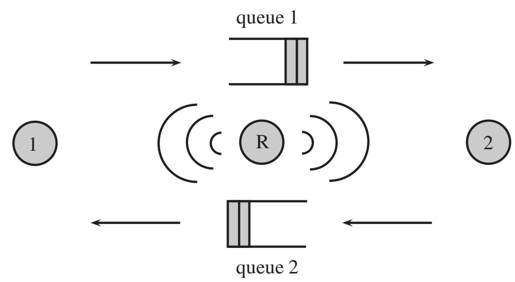


Fig. 2. Two-way relay network

The remainder of the paper is organized as follows. In Section II, we provide MDP formulations of our problem and show that an optimal stationary policy exists for the average cost MDP. In Section III, we derive important structural properties of the optimal policy. In Section IV, we develop an approach to compute the threshold policy which is optimal in certain cases and is effective (and possibly optimal) in other cases. We report our computational results in Section V. In Section VI, we present our concluding remarks and some future research directions.

II. Model Formulation

We consider network coding in a two-way relay network shown in Fig. 2. The relay node R forwards packets or their combinations that belong to two opposite flows. The relay node maintains queues 1 and 2 to store packets received from nodes 1 and 2 respectively (i.e. intended for nodes 2 and 1 respectively). Packets arrive at queues 1 and 2 according to independent Poisson processes with rates λ_1 and λ_2 respectively. The relay node gets opportunities to transmit at fixed time intervals. Let T be the fixed time gap between two consecutive transmission opportunities. The relay node can send any number of packets during a transmission opportunity. We assume low to medium load conditions for this network, i.e. λ_1 and λ_2 values are not very large (note that the relay usually shares the channel resources with the two source nodes, which creates a challenging situation at high loads [20]). We also assume that there is no interaction between the relay and the sources to regulate transmissions. This is reasonable since there is not a lot of competition between the relay and the sources for medium access given the low load conditions and the periodic scheduling of transmissions from the relay.

To save energy, the relay node tries to reduce the number of packet transmissions through network coding. As long as the queues 1 and 2 are nonempty, it sends as many number of coded packets as possible by combining packets from both the queues. When a packet cannot be coded (due to the shortage of a packet in the other queue), the relay can transmit it uncoded or hold it for transmission in the future. To make this decision, we consider some costs associated with transmission and delay. Let c_t be the cost of transmitting a coded or uncoded packet, and let $\bar{c}_h$ be the cost of holding a packet per unit time. The cost of holding a packet from one transmission opportunity to the next is $c_h := \bar{c}_h T$. We will consider only the case $c_h < c_t$ since holding will not be a cost-effective option otherwise. However, our analytical results are derived without this assumption.

When each of the queues 1 and 2 has n packets at a transmission opportunity, it is optimal to transmit n coded packets. When queue 1 has n_1 packets and queue 2 has n_2 ($\neq n_1$) packets, $\min(n_1, n_2)$ coded packets will be sent. However, in this case, since the remaining $|n_1 - n_2|$ packets (that are left in one of the queues) cannot be coded, it is required to decide whether to transmit some of them uncoded. Therefore, our objective is to develop a strategy for the relay to decide how many uncoded packets to transmit (note that the number of coded packets to be sent is already known) at every transmission opportunity so that the average transmission and holding cost is minimized over the long-run. To develop such a strategy, we formulate our problem as a Markov decision process (MDP) which we describe next.

In our MDP setup, the *state* of the system is described by a two-dimensional vector (s_n^1, s_n^2), where s_n^1 and s_n^2 are the number of packets in queues 1 and 2 respectively just before the n-th ($n = 0, 1, \cdots$) transmission opportunity. The state space is $\mathcal{S} = \{(i,j) : i = 0, 1, \cdots, j = 0, 1, \cdots\}$. Based on the state of the system at every stage of the MDP (i.e. at every transmission opportunity), a certain number of coded and uncoded packets are transmitted. Here, the total number of packets transmitted is defined as the *action* which is denoted by a_n in the n-th stage. Note that the action space in state (i,j) is $\mathcal{A}_{i,j} = \{\min(i,j), \cdots, \max(i,j)\}$. The complete action space is $\mathcal{A} = \cup_{(i,j)\in\mathcal{S}} \mathcal{A}_{i,j}$. When an action a is taken in state (i,j), the number of coded and uncoded transmissions are $\min(i,j)$ and $a - \min(i,j)$ respectively. Let p_y^z be the probability of y packet arrivals to queue z ($z = 1, 2$) between two transmission opportunities. Note that p_y^1 and p_y^2 ($y = 0, 1, \cdots$) are probability distributions of Poisson random variables with means $\lambda_1 T$ and $\lambda_2 T$ respectively. Hence, if an action a is taken in state (i,j), the system will be in state (k,l) in the next stage with probability $p_{(i,j)(k,l)}(a) := p^1_{k-(i-a)^+}\, p^2_{l-(j-a)^+}$, where $(x)^+ = \max(x, 0)$.

In every stage of the MDP, a transmission cost is incurred depending on the action selected. Further, there is a cost for holding the remaining packets (after transmission) as well as the new arriving packets until the next transmission opportunity. When an action a is taken in state (i,j), the total cost, denoted by $c'(i,j,a)$, is computed as

$$\begin{aligned} c'(i,j,a) \quad &= \quad c_t a + \bar{c}_h T\left((i-a)^+ + (j-a)^+\right) \\ &\qquad + \bar{c}_h \mathrm{E}\left[\sum_{z=1}^{2} \sum_{y=1}^{N_z(T)} (T - \xi_y^z)\right], \end{aligned} \tag{1}$$

where T is the time between the current and next transmission opportunities, $N_z(T)$ is the random variable indicating the number of new packet arrivals to queue z ($z = 1, 2$) in time T, and ξ_y^z is the time of arrival (measured from the current transmission opportunity) of the y-th packet ($y = 1, \cdots, N_z(T)$) to queue z ($z = 1, 2$). Conditional on $N_z(T) = m$ ($z = 1, 2$), the arrival times $\xi_1^z, \cdots, \xi_m^z$ are distributed as the order statistics of m independent random variables, each uniformly distributed over $(0, T)$ [21]. Using this property of Poisson arrivals, the expression in (1) is simplified as $c'(i,j,a) = c_t a + c_h\left((i-a)^+ + (j-a)^+\right) + \frac{c_h T}{2}(\lambda_1 + \lambda_2)$. Note that the cost component $\frac{c_h T}{2}(\lambda_1 + \lambda_2)$ is a constant and will not have any effect on deciding the action in any state. Hence, we will ignore it in our model and use the following function for cost per stage:

$$c(i,j,a) = c_t a + c_h\left((i-a)^+ + (j-a)^+\right). \tag{2}$$

Having described the components of the MDP, we need to find a policy that will decide the action at every stage in such a way that the average cost over an infinite time horizon will be minimized. A stationary policy for our MDP is a mapping $\theta : \mathcal{S} \to \mathcal{A}$, where $\theta(i,j)$ is the action selected when the state of the system is (i,j). Given the countable state space and infinite time horizon in our MDP, we consider only stationary policies. The long-run average cost under any policy θ is defined as

$$\begin{aligned} &g(\theta) = \\ &\lim_{N\to\infty} \frac{1}{N+1} \mathrm{E}_\theta\left[\sum_{n=0}^{N} c(S_n^1, S_n^2, a_n)\middle| (S_0^1, S_0^2) = (0,0)\right], \end{aligned} \tag{3}$$

where (S_n^1, S_n^2) is the random state in the n-th stage. Note that, though we start the system in state $(0,0)$, the average cost is independent of this choice [12]. Our objective is to find a stationary policy that minimizes the average cost function $g(\theta)$. However, given the countable state space and unbounded costs in our problem, such a policy would exist only under specific conditions.

In the following subsections, we present our analysis to show that an optimal stationary policy exists for our MDP. We now introduce the discounted cost formulation which will later be used to derive results for the average cost problem.

A. Discounted Cost Formulation

In our MDP setup, when a discount factor of β ($0 < \beta < 1$) is considered, the total expected discounted cost incurred under a policy θ is given by

$$v_{\beta,\theta}(i,j) = \mathrm{E}_\theta\left[\sum_{n=0}^{\infty} \beta^n c(S_n^1, S_n^2, a_n)\middle| (S_0^1, S_0^2) = (i,j)\right], \tag{4}$$

where the initial state of the system is (i,j). The optimal discounted cost in an initial state (i,j) is given by the *discounted cost function* $v_\beta(i,j) := \inf_\theta v_{\beta,\theta}(i,j)$. As per the following proposition, $v_\beta(i,j)$ values are finite.

Proposition 1. *The discounted cost function $v_\beta(i,j)$ is finite for every state $(i,j) \in \mathcal{S}$ and discount factor β ($0 < \beta < 1$).*

Proof: See Appendix I-A. □

Proposition 1 implies that the discounted cost function $v_\beta(i,j)$, for $(i,j) \in \mathcal{S}$, satisfies the *discounted cost optimality* *equation* [18]:

$$v_\beta(i,j) = \min_{a\in\mathcal{A}_{i,j}} \Bigg\{ c(i,j,a) + \beta \sum_{k,l} p_k^1 p_l^2 \; v_\beta\left((i-a)^+ + k, (j-a)^+ + l\right) \Bigg\}. \quad (5)$$

Any stationary policy that realizes the minimum in the right side of (5) is discounted cost optimal. We will use properties of the discounted cost optimal stationary policy to characterize the optimal stationary policy in the average cost case.

B. Average Cost Formulation

Recall that our objective is to find a stationary policy that minimizes the long-run average cost defined in (3). In this subsection, we show that an average cost optimal stationary policy exists in our MDP, and it can be computed as the limit of discounted cost optimal stationary policies. We will need the following propositions to prove this main result in Theorem 1.

Proposition 2. *The discounted cost function $v_\beta(i,j)$ is non-decreasing in i and j.*

Proof: See Appendix I-B. □

Proposition 3. *The MDP has a stationary policy inducing an irreducible, ergodic Markov chain with a finite average cost.*

Proof: See Appendix I-C. □

Theorem 1. *(a) There exist a constant $g = \lim_{\beta\uparrow 1} (1-\beta)v_\beta(i,j)$ for every $(i,j)\in\mathcal{S}$, and a function $h(i,j)$ satisfying the average cost optimality inequality:*

$$g + h(i,j) \geq \min_{a\in\mathcal{A}_{i,j}} \Bigg\{ c(i,j,a) + \sum_{k,l} p_k^1 p_l^2 \; h\left((i-a)^+ + k, (j-a)^+ + l\right) \Bigg\}. \quad (6)$$

The constant g is the optimal average cost, and any stationary policy that realizes the minimum in the right side of (6) is average cost optimal.
(b) There exists an average cost optimal stationary policy θ^* that is a limit point of a sequence of discounted cost optimal stationary policies $\{\theta_{\beta_k}\}_{k\geq 1}$, where $\beta_k \to 1$.

Proof: To prove (a) and (b), we will first show that the following conditions (provided by Sennott [18]) are satisfied.

1) The discounted cost function $v_\beta(i,j)$ is finite for every state (i,j) and discount factor β.
2) There exists a nonnegative number N such that $-N \leq h_\beta(i,j)$ for all (i,j) and β, where $h_\beta(i,j) = v_\beta(i,j) - v_\beta(0,0)$.
3) There exist nonnegative numbers M_{ij} such that $h_\beta(i,j) \leq M_{ij}$ for every (i,j) and β. For every state (i,j), there exists an action $a(i,j)$ such that $\sum_{(k,l)\in\mathcal{S}} p_{(i,j)(k,l)}(a(i,j)) \, M_{kl} < \infty$.

Condition 1 is satisfied through Proposition 1. Condition 2 is implied by Proposition 2 and is therefore satisfied. Based on Proposition 5 in [18], Proposition 3 satisfies Condition 3. Now that Conditions 1-3 are satisfied, the result in (a) follows from Theorem 7.2.3 (and its proof) in [22].

Now, let $\{\beta_n\}_{n\geq 1}$ be any sequence of discount factors converging to 1, and let $\{\theta_{\beta_n}\}_{n\geq 1}$ be the corresponding sequence of discounted cost optimal stationary policies. By Lemma 1 in [18], there exists a subsequence of discount factors $\{\beta_{n_k}\}_{k\geq 1}$ (also converging to 1) and a stationary policy θ^* that is a limit point of the subsequence $\{\theta_{\beta_{n_k}}\}_{k\geq 1}$. Further, as Conditions 1-3 are satisfied, the stationary policy θ^* is average cost optimal by Theorem 1 in [18]. □

III. Structural Properties of Optimal Policy

In this section, we derive structural properties of the optimal policy in both discounted cost and average cost problems. In Theorem 1, we showed that an average cost optimal stationary policy in our MDP can be found as a limit point of discounted cost optimal stationary policies. Therefore, the structural properties of the discounted cost optimal stationary policies will remain the same in this average cost optimal stationary policy.

Before formalizing our main results, we would like to introduce certain concepts of convexity of a function defined over discrete points. First, we define a univariate "discrete convex function" for our purpose (hereafter we will use the word "convex" instead of "discrete convex" for discrete functions).

Definition 1. *The function $f:\mathbb{Z}_+\to\mathbb{R}$ is defined to be convex if and only if $f(i+1)-f(i)$ is nondecreasing in i.*

If $f:\mathbb{Z}_+\to\mathbb{R}$ is convex (by Definition 1), a point i in the domain of f is a global minimum if it is a local minimum in the sense that $f(i)\leq\min\{f(i-1), f(i+1)\}$. This function will also satisfy other natural properties of a convex function [23]. Now we extend the idea in Definition 1 to a bivariate discrete function in which we are primarily interested in analyzing convexity in the direction of one variable.

Definition 2. *The function $f:\mathbb{Z}_+^2\to\mathbb{R}$ is defined to be convex in i if and only if $f(i+1,j)-f(i,j)$ is nondecreasing in i for every j. Similarly, f is defined to be convex in j if and only if $f(i,j+1)-f(i,j)$ is nondecreasing in j for every i.*

Note that, if $f(i,j)$ is convex in i (by Definition 2), then for a fixed j (say j_1), the function $f'(i) := f(i,j_1)$ is convex. Similarly, if $f(i,j)$ is convex in j, then for a fixed i (say i_1), the function $f''(j) := f(i_1,j)$ is convex. Additionally, if $f(i,j)$ and $g(i,j)$ are convex in i (resp. in j), the function $c_1 f(i,j) + c_2 g(i,j)$ is convex in i (resp. in j) for $c_1\geq 0, c_2\geq 0$. We will use these properties in the proofs of various results presented in this section.

Remark 1. *Definition 2 helps check convexity of a bivariate discrete function in only one variable (while the other is fixed), which is useful for deriving our results. However, note that, even if a function is convex in each variable separately, it is not sufficient to identify the function as convex unless it is additively separable.*

Now we present in the following Lemma 1 and Proposition 4 which are required to prove the main result of this section in Theorem 2.

Lemma 1. *Suppose $f(i,j) : \mathbb{Z}_+^2 \to \mathbb{R}$ is convex in i and j. For $c > 0$, the function $g(i,j) = \min_{a \in \{\min(i,j),\cdots,\max(i,j)\}} \{ca + f((i-a)^+, (j-a)^+)\}$ is convex in i and j if*

$$\min\{f(1,0) - f(0,0), c\} + \min\{f(0,1) - f(0,0), c\} \geq c. \tag{7}$$

Proof: See Appendix I-D. □

Proposition 4. *If $c_h \geq c_t/2$, the discounted cost function $v_\beta(i,j)$ is convex in i and j.*

Proof: See Appendix I-E. □

Theorem 2. *If $c_h \geq c_t/2$, then (a) there exist constants $L_1^\beta, L_2^\beta \geq 0$ such that the optimal action in state $(i,j) \in \mathcal{S}$ in the β-discounted cost problem is given by*

$$\begin{aligned} a^*(i,j) &= \min(i,j) + \left(i - \min(i,j) - L_1^\beta\right)^+ \\ &\quad + \left(j - \min(i,j) - L_2^\beta\right)^+. \end{aligned} \tag{8}$$

(b) There is an average cost optimal policy $\theta(L_1^*, L_2^*)$ such that $\theta(L_1^\beta, L_2^\beta) \to \theta(L_1^*, L_2^*)$ as $\beta \to 1$, where $\theta(L_1^\beta, L_2^\beta)$ is the β-discounted cost optimal policy described in (8).

Proof: Since $a \in \mathcal{A}_{i,j} = \{\min(i,j), \cdots, \max(i,j)\}$, we can write $a = \max(i,j) - (i-a)^+ - (j-a)^+$. Using this in the expression for $c(i,j,a)$ (see (2)), (5) can be written for every $(i,j) \in \mathcal{S}$ as

$$\begin{aligned} v_\beta(i,j) &= c_t \max(i,j) \\ &\quad + \min_{a \in \mathcal{A}_{i,j}} \Bigg\{ -(c_t - c_h)\left[(i-a)^+ + (j-a)^+\right] \\ &\quad + \beta \sum_{k,l} p_k^1 p_l^2 \; v_\beta\left((i-a)^+ + k, (j-a)^+ + l\right) \Bigg\}. \end{aligned} \tag{9}$$

Since $c_h \geq c_t/2$, the discounted cost function $v_\beta(i,j)$ is convex in i and j by Proposition 4. Therefore the function $f(i,j) := \sum_{k,l} p_k^1 p_l^2 \; v_\beta(i+k, j+l)$ is convex in i and j. Also, it follows from definition that $-(c_t - c_h)(i+j)$ is convex in i and j. Now we can write (9) as

$$v_\beta(i,j) = c_t \max(i,j) + \min_{a \in \mathcal{A}_{i,j}} \left\{ g\left((i-a)^+, (j-a)^+\right) \right\}, \tag{10}$$

where $g(i,j) = -(c_t - c_h)(i+j) + \beta f(i,j)$. Note that $g(i,j)$ is convex in i and j.

Now, we will prove result (a) by considering the following cases: (i) $i \geq j$, and (ii) $i \leq j$. In case (i) $i \geq j$, (10) can be written as

$$\begin{aligned} v_\beta(i,j) &= c_t \max(i,j) + \min_{a \in \{j,\cdots,i\}} g(i-a, 0) \\ &= c_t \max(i,j) + \min_{b \in \{0,\cdots,i-j\}} g_1(b), \end{aligned} \tag{11}$$

where $b = i - a$, and $g_1(b) = g(b, 0)$. Note that $g_1(b)$ is convex. Let $L_1^\beta = \arg\min\{g_1(b) : b \geq 0\}$ be a global minimum of $g_1(b)$. Hence, the value of b that minimizes $g_1(b)$ in (11) is given by

$$b^*(i,j) = \begin{cases} i - j & \text{if } 0 \leq i - j < L_1^\beta, \\ L_1^\beta & \text{if } i - j \geq L_1^\beta. \end{cases}$$

Therefore, in case (i) $i \geq j$, the optimal action in state (i,j) can be found as

$$a^*(i,j) = i - b^*(i,j) = \begin{cases} j & \text{if } 0 \leq i - j < L_1^\beta, \\ i - L_1^\beta & \text{if } i - j \geq L_1^\beta. \end{cases} \tag{12}$$

Similarly, in case (ii) $i \leq j$, it can be shown that there exists a constant $L_2^\beta \geq 0$ such that the optimal action in state (i,j) is given by

$$a^*(i,j) = \begin{cases} i & \text{if } 0 \leq j - i < L_2^\beta, \\ j - L_2^\beta & \text{if } j - i \geq L_2^\beta. \end{cases} \tag{13}$$

Note that (12) and (13) can be combined into one expression as presented in (8). This completes the proof of the result (a). Result (b) follows from Theorem 1(b). □

Theorem 2 shows that, if $c_h \geq c_t/2$, the optimal transmission policy for the relay node is a threshold based policy. Observe that, under such a policy, $\min(i,j)$ coded packets are sent in state (i,j) as in any other policy. However, once the coded packets are sent, packets from the nonempty queue are sent uncoded until the number of remaining packets in the queue reaches an optimal threshold level (which is L_1^* for queue 1 and L_2^* for queue 2 in the average cost case).

Notice that the condition $c_h \geq c_t/2$ is only sufficient, but not necessary, for the optimal policy to be threshold-based. We have not shown whether such threshold policy is optimal when $c_h < c_t/2$. However, in this case, we conjecture that the threshold policy will be very effective, if not optimal. In fact, some heuristic strategies such as the "transmit-all" and "rate-based" policies that we show to be very effective in specific situations even under the condition $c_h < c_t/2$ are actually similar threshold based policies (see Section V) .

Remark 2. *If transmission and holding costs are charged to individual packets, a packet would incur a transmission cost of $c_t/2$ if it is coded, and c_t if it is sent uncoded. Note that, in this situation, a packet cannot reduce its own cost by waiting for a coding opportunity if $c_h + c_t/2 \geq c_t$ or $c_h \geq c_t/2$. Therefore, under the condition $c_h \geq c_t/2$, the "individually optimal" average cost policy is to always transmit (coded if possible, else uncoded) at a transmission opportunity. However, it is clear from Theorem 2 that this policy may not be optimal at the system level.*

IV. Computation of Threshold Policy

Given the countable state space in our MDP, computing the optimal stationary policy (in both discounted cost and average cost problems) by standard value iteration or policy iteration procedures is intractable. However, we know that the average cost optimal policy is threshold-based if $c_h \geq c_t/2$. Also, such a threshold policy will be efficient (and possibly optimal) in the case $c_h < c_t/2$. Therefore, we are primarily interested in computing the threshold policy in all cases. Here, by threshold policy, we mean the best threshold policy, i.e. optimal values of the thresholds are used. In this section, we develop an approach to compute the optimal threshold values L_1^* and L_2^* which completely characterize the threshold policy.

Consider in our MDP an arbitrary threshold policy with threshold values L_1 and L_2 (where $L_1 \geq 0, L_2 \geq 0$). Under this (L_1, L_2) threshold policy, the action in state $(i,j) \in \mathcal{S}$ is given as

$$a(i,j) = \min(i,j) + (i - \min(i,j) - L_1)^+ + (j - \min(i,j) - L_2)^+ . \quad (14)$$

Under this policy, let Z_n^1 and Z_n^2 denote the number of packets in queues 1 and 2 respectively just after all transmissions are completed in the n-th transmission opportunity. Note that the stochastic process $\{(Z_n^1, Z_n^2), n \geq 0\}$ is an irreducible discrete-time Markov chain with state space $\mathcal{S}' = \{(0,L_2), (0, L_2-1), \cdots, (0,1), (0,0), (1,0), \cdots, (L_1-1, 0), (L_1, 0)\}$. The transition probabilities in this Markov chain are given by

$$\begin{aligned}
\bar{p}_{(i,0)(k,0)} &= q_{(k-i)}, && 0 \leq i \leq L_1,\ 0 \leq k < L_1,\\
\bar{p}_{(i,0)(L_1,0)} &= \sum_{m \geq L_1 - i} q_m, && 0 \leq i \leq L_1,\\
\bar{p}_{(i,0)(0,l)} &= q_{-(i+l)}, && 0 \leq i \leq L_1,\ 0 \leq l < L_2,\\
\bar{p}_{(i,0)(0,L_2)} &= \sum_{m \leq -(i+L_2)} q_m, && 0 \leq i \leq L_1,\\
\bar{p}_{(0,j)(0,l)} &= q_{-(l-j)}, && 0 \leq j \leq L_2,\ 0 \leq l < L_2,\\
\bar{p}_{(0,j)(0,L_2)} &= \sum_{m \leq -(L_2-j)} q_m, && 0 \leq j \leq L_2,\\
\bar{p}_{(0,j)(k,0)} &= q_{(j+k)}, && 0 \leq k < L_1,\ 0 \leq j \leq L_2,\\
\bar{p}_{(0,j)(L_1,0)} &= \sum_{m \geq j + L_1} q_m, && 0 \leq j \leq L_2,
\end{aligned}$$

where q_m is the probability of queue 1 receiving m (where m is an integer in $(-\infty, \infty)$) more packets than queue 2 between two transmission opportunities. Note that q_m is the probability distribution of the difference of two independent Poisson random variables, and it is specified by the Skellam distribution [24]:

$$q_m = e^{-(\lambda_1+\lambda_2)T} \left(\frac{\lambda_1}{\lambda_2}\right)^{m/2} I_{|m|}(2T\sqrt{\lambda_1\lambda_2}),$$

where $I_{(\cdot)}(\cdot)$ is the modified Bessel function of the first kind.

Note that the long-run average cost incurred in the considered Markov chain is precisely the average cost under (L_1, L_2) threshold policy in our MDP. The expected cost incurred in state $(i,j) \in \mathcal{S}'$ of the Markov chain is given by

$$\begin{aligned} c_{ij}(L_1, L_2) = \sum_{k,l} p_k^1 p_l^2 \Bigg\{ c_h(i+j) + c_t \Bigg[&\min(i+k, j+l) \\ &+ (i+k - \min(i+k, j+l) - L_1)^+ \\ &+ (j+l - \min(i+k, j+l) - L_2)^+ \Bigg] \Bigg\}. \end{aligned} \quad (15)$$

Now the long-run average cost under (L_1, L_2) threshold policy can be calculated as

$$\bar{g}(L_1, L_2) := \sum_{(i,j) \in \mathcal{S}'} \bar{\pi}_{ij} c_{ij}(L_1, L_2), \quad (16)$$

where $\{\bar{\pi}_{ij} : (i,j) \in \mathcal{S}'\}$ is the set of stationary probabilities of the Markov chain satisfying the equations $\sum_{(k,l) \in \mathcal{S}'} \bar{\pi}_{kl} \bar{p}_{(k,l)(i,j)} = \bar{\pi}_{ij}$, for all $(i,j) \in \mathcal{S}'$, and $\sum_{(i,j) \in \mathcal{S}} \bar{\pi}_{ij} = 1$.

Now the optimal values of the thresholds, L_1^* and L_2^*, can be found by minimizing the discrete function $\bar{g}(L_1, L_2)$ in (16). As a simple approximation method, the global minimum of $\bar{g}(L_1, L_2)$ can be found by evaluating this function over a finite set $\{0, 1, \cdots, N_1\} \times \{0, 1, \cdots, N_2\}$, where N_1 and N_2 are suitably large integers. However, in our numerical experiments, a local minimum of $\bar{g}(L_1, L_2)$ was found to be the global minimum in all cases. Assuming that $\bar{g}(L_1, L_2)$ has this property, any discrete gradient search method can be applied to find the minimum (L_1^*, L_2^*) more efficiently.

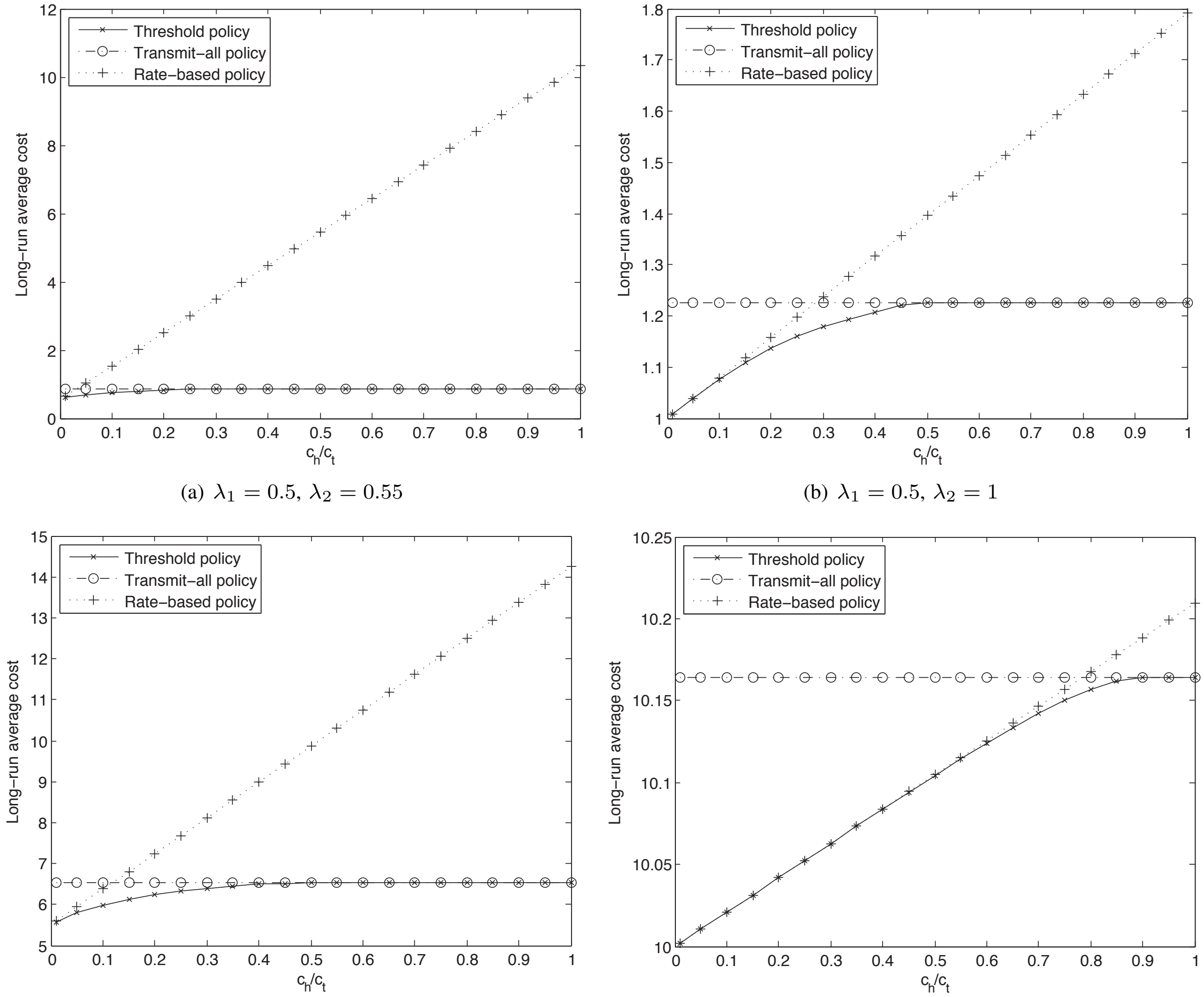
V. Numerical Results

In this section, we present our numerical results to demonstrate the effectiveness of the threshold policy in network coding decisions. Most of the results presented in this section are based on our MATLAB simulations of node/network operation covering a large number of transmission opportunities. The objective in these experiments is to compare the threshold policy against other simple policies, and test these policies in situations where assumptions do not hold. We mainly study the performance of the following policies:

1) *Transmit-all policy*: Under this policy, the relay node transmits all the queued packets at every transmission opportunity. In this case, like every other policy, when the relay has i and j packets in queues 1 and 2 respectively, $\min(i,j)$ coded packets are sent. Then all the remaining packets (that are left in one of the queues) are sent uncoded. Note that this policy is expected to be effective when the holding cost c_h is high.
2) *Rate-based policy*: This policy is based on λ_1 and λ_2 which are the rates of packet arrivals (Poisson distributed) to queues 1 and 2 respectively. Under this policy, upon transmission of coded packets, all the

TABLE I
Computed threshold policy

λ_1	λ_2	c_h	c_t	L_1^*	L_2^*	Avg. cost
5	5	0.05	1	8	8	5.4931
5	5	0.1	1	5	5	5.6875
5	5	0.2	1	3	3	5.9439
5	5	0.4	1	1	1	6.2138
5	5	0.6	1	0	0	6.2455
5	5.5	0.05	1	14	4	5.7952
5	5.5	0.1	1	8	3	5.9814
5	5.5	0.2	1	4	2	6.2331
5	5.5	0.4	1	1	0	6.4996
5	5.5	0.6	1	0	0	6.5422
5	6	0.05	1	21	2	6.1750
5	6	0.1	1	11	1	6.3270
5	6	0.2	1	6	1	6.5510
5	6	0.4	1	2	0	6.7908
5	6	0.6	1	0	0	6.8669
5	7.5	0.05	1	48	0	7.5480
5	7.5	0.1	1	23	0	7.5960
5	7.5	0.2	1	11	0	7.6908
5	7.5	0.4	1	4	0	7.8520
5	7.5	0.6	1	2	0	7.9542



(a) $\lambda_1 = 0.5$, $\lambda_2 = 0.55$

(b) $\lambda_1 = 0.5$, $\lambda_2 = 1$

(c) $\lambda_1 = 5$, $\lambda_2 = 5.5$

(d) $\lambda_1 = 5$, $\lambda_2 = 10$

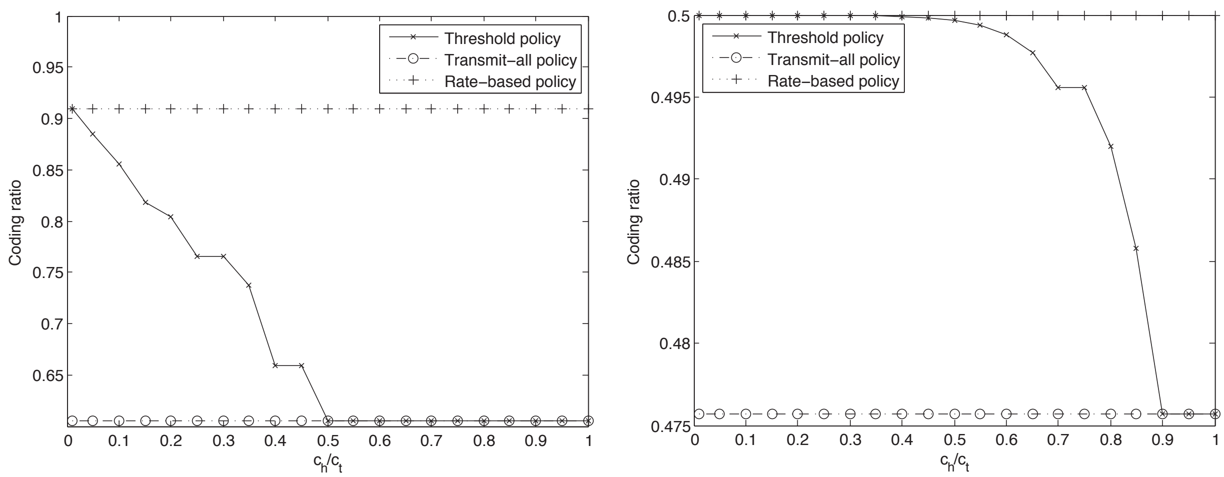
Fig. 3. Comparison of long-run average costs of policies in single relay-node network ($c_t = 1$ and $T = 1$ in all cases)

remaining packets in queue 1 (if it is nonempty) are sent uncoded if $\lambda_1 > \lambda_2$, and are held if $\lambda_1 < \lambda_2$. Similarly, if queue 2 is nonempty following transmission of coded packets, all its remaining packets are sent unocoded if $\lambda_1 < \lambda_2$, and are held if $\lambda_1 > \lambda_2$. We will not use this policy when $\lambda_1 = \lambda_2$. Notice that this policy is particularly attractive when $\lambda_1 << \lambda_2$ or $\lambda_1 >> \lambda_2$.

3) *Threshold policy*: Here we mean the best (L_1, L_2) threshold policy, i.e. optimal values of the thresholds (L_1^* and L_2^*) are used. Recall that this is the optimal transmission policy for the relay node if $c_h \geq c_t/2$.

Note that both "transmit-all" and "rate-based" policies are special instances of the (L_1, L_2) threshold policy (see (14)). In the "transmit-all" policy, we have $(L_1, L_2) = (0, 0)$. In the "rate-based" policy, $(L_1, L_2) = (0, \infty)$ if $\lambda_1 > \lambda_2$, and $(L_1, L_2) = (\infty, 0)$ if $\lambda_1 < \lambda_2$. Since these policies are specific threshold policies, they will not perform any better than the best threshold policy (which we call just the "threshold policy"). However, since these policies are easy to implement (as the values of the thresholds L_1 and L_2 are already known), we would like to find out how efficient they are.

We computed the threshold policy for a relay node using our approach described in Section IV. Table I presents the threshold values (L_1^* and L_2^*) in the computed threshold policy for different values of arrival rates λ_1 and λ_2, and cost parameters c_t and c_h. We use $c_t = 1$ and $T = 1$ in all our results. Notice that $L_1^* > L_2^*$ whenever $\lambda_1 < \lambda_2$. This is because, when $\lambda_1 < \lambda_2$, packets in queue 1 have higher chances of being coded than those in queue 2, and therefore the holding option is more appealing to queue 1 than queue 2. Likewise we have $L_1^* < L_2^*$ if $\lambda_1 > \lambda_2$, though instances of this case are not shown. Observe that, for fixed values of c_t and c_h, as λ_2 increases over λ_1, value of L_1^* increases, and value of L_2^* decreases. Similarly, if λ_1 increases over λ_2, value of L_1^* will decrease, and value of L_2^* will increase. This trend suggests that the threshold policy will be close to the "rate-based" policy when $\lambda_1 << \lambda_2$ or $\lambda_1 >> \lambda_2$. Also observe that, when values of λ_1 and λ_2 are fixed, both L_1^* and L_2^* values decrease as the holding cost quantity c_h increases. This is expected since fewer packets will be held when the holding cost is more. When c_h is considerably high, we have $L_1^* = L_2^* = 0$. In this case, the threshold policy is precisely the "transmit-all" policy.



(a) $\lambda_1 = 5$, $\lambda_2 = 5.5$ (b) $\lambda_1 = 5$, $\lambda_2 = 10$

Fig. 4. Comparison of coding ratios of policies in single relay-node network ($c_t = 1$ and $T = 1$ in all cases)

Based on our simulation results, Fig. 3 presents comparison of the long-run average costs of the policies in different instances of λ_1 and λ_2. Since the holding cost quantity c_h is usually not known explicitly, we compare the average costs of the policies over different possible values of c_h/c_t. Notice that, as expected, the threshold policy always achieves the minimum average cost among the considered policies. However, the "transmit-all" and "rate-based" policies perform as well as the threshold policy in certain situations. When both arrival rates λ_1 and λ_2 are very small, the "transmit-all" policy is very effective at almost all values of c_h (see Fig. 3(a)). When the difference between λ_1 and λ_2 is large, the "rate-based" policy performs very close to the threshold policy at most values of c_h (see Fig. 3(d)). In all other cases of λ_1 and λ_2, the "rate-based" policy is effective at very low values of c_h, and the "transmit-all" policy is very effective at higher values of c_h (see Fig. 3(b) and Fig. 3(c)). Further, when either λ_1 or λ_2 is very large, the overall holding costs are insignificant compared to the transmission costs. Therefore, in such a case (e.g. see Fig. 3(d)), there is little difference in the performances of the considered policies.

It is also important to see how effective our policies are in availing of network coding opportunities. We measure this performance by the "coding ratio", which we define as the long-run proportion of coded packets in the total number packet transmissions. Fig. 4 presents comparison of the coding ratios attained by our policies over different possible values of c_h/c_t in a single relay node network. Notice that the coding ratio of the threshold policy always lies between the coding ratios of the "transmit-all" and "rate-based" policies. The "transmit-all" policy has the lowest coding ratio since it never holds packets for possible network coding opportunities.

Remark 3. *In case of "rate-based" policy, if $\lambda_1 < \lambda_2$, the total number of coded packets over the long-run is equal to the total number of packet arrivals to queue 1. Also, the total number of packet transmissions is equal to the total number of packet arrivals to queue 2. Hence, the coding ratio in this case is expected to be λ_1/λ_2, which can be observed in Fig. 4. Similarly, if $\lambda_1 > \lambda_2$, the coding ratio in the "rate-based" policy will again be a constant and is equal to λ_2/λ_1. It is also important to note that any other policy that tries to achieve a coding ratio higher than the "rate-based" policy will make the system unstable by building up at least one of the queues.*

Finally, to study the performance of our policies at a network level, we used them separately in a 4-node line network (see Fig. 1(b)). In this case, packets in the input flows f_1 and f_2 (to queue 1 of node 1, and queue 2 of node 4 respectively) arrive as per independent Poisson processes with rates λ_1 and λ_2 respectively. Note that the rates of packet arrivals to queues 1 and 2 of each node are also λ_1 and λ_2 respectively. However, the corresponding arrival processes (except for queue 1 of node 1, and queue 2 of node 4) are not Poisson anymore. Now we consider a transmission schedule in which nodes 1 and 3 transmit together during a transmission opportunity, nodes 2 and 4 transmit together during the next opportunity, and this cycle repeats. In this case, if the time period between consecutive transmission opportunities is T for the network, the corresponding time period for an individual node is $2T$. Therefore, the holding cost quantity is $2c_h$ for each node. In this network setup, Fig. 5 presents comparison of the long-run average costs of our considered policies at the network level when they are used locally at each node (e.g. in case of the threshold policy, threshold values are computed at each node assuming it as a single relay node network). Notice that the threshold policy achieves the minimum average cost among the considered policies in most cases. Further, the 'transmit-all" and "rate-based" policies perform as well as the threshold policy in specific situations, and these observations are similar to those we discussed in the single node network case. Notice that the effect of holding cost quantity c_h on the average cost is significantly higher at the network level as compared to the single node case (see Fig. 3). This is primarily because the proportion of the holding costs in the total costs is more in the network level in our considered network case as compared to the single node case.

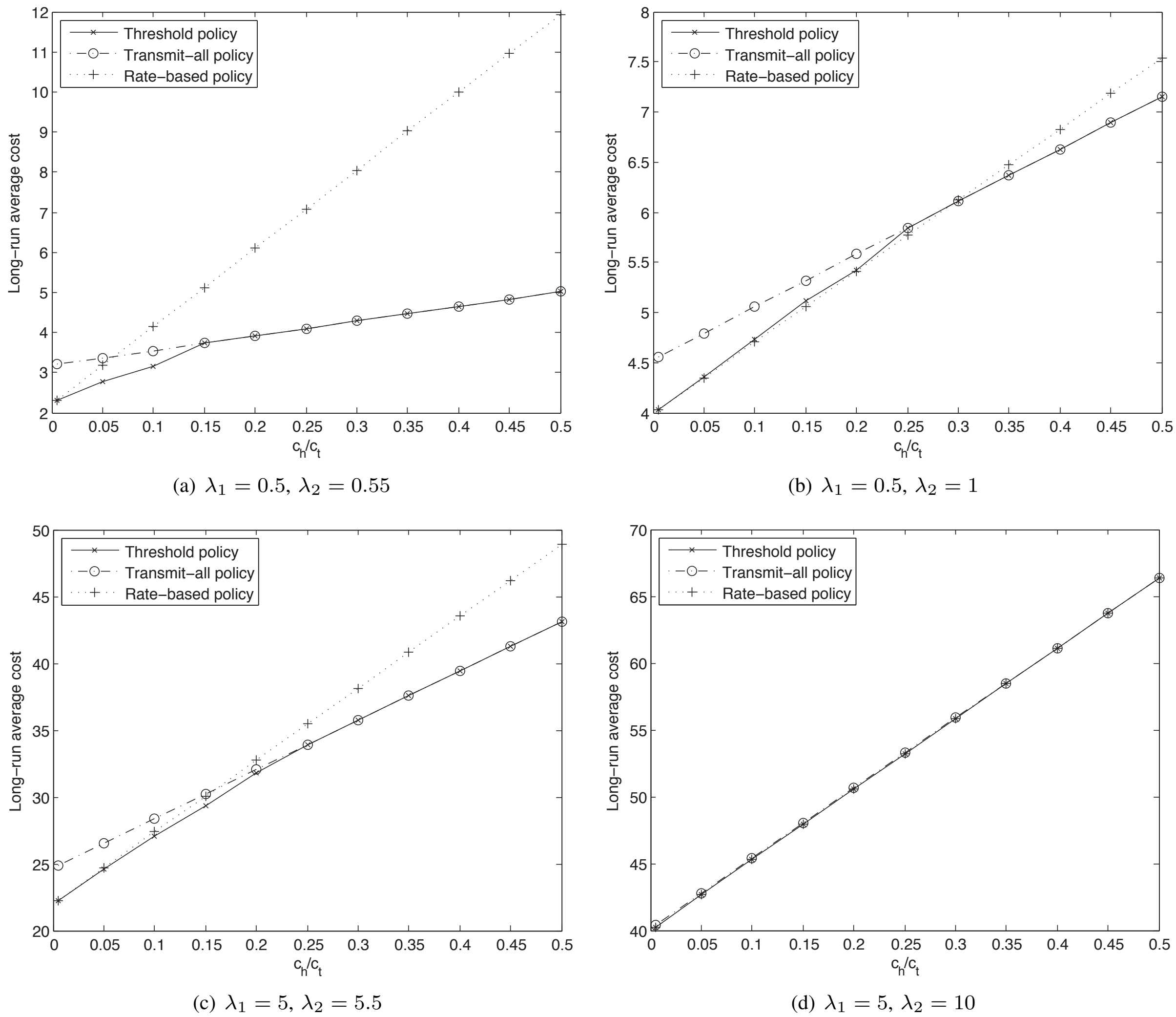


Fig. 5. Comparison of long-run average costs of policies in 4-node line network ($c_t = 1$ and $T = 1$ in all cases)

VI. Conclusion

In this paper, we developed a Markov decision process (MDP) based model to manage energy-delay trade-off in network coding decisions in a two-way relay network. We proved that an optimal stationary policy exists in our average cost MDP. But computing this policy is difficult due to countably infinite number of states in the MDP. However, we showed that, in a certain case, an optimal stationary policy in our MDP is a threshold based policy. We developed a method to compute this threshold policy. We also found such threshold policy to be very effective in other possible cases. Further, based on the structure of the threshold policy, we developed insights into other simple policies which we showed to be efficient in particular situations.

As our numerical results indicate, the threshold policy performs the best among our considered policies in all situations. However, note that the "transmit-all" and "rate-based" policies provide near optimal performance in two different operating regions. These policies are also attractive due to their simplicity and ease in implementation. Therefore, one can consider only the given two heuristic policies and use the one that provides the lower long-run average cost (which can be computed using (16)). This approach would eliminate the need for additional resources for the computation of the threshold policy and is particularly suitable for practical systems. Further, the threshold policy may not be suitable when exact values of the arrival rates (λ_1 and λ_2) and the cost parameters (c_h and c_t) are not known. In this case, based on the learnings from our numerical results, the "transmit-all" or "rate-based" policy can be used when we have some rough information about these parameters. The "transmit-all" policy is mostly effective when the difference between λ_1 and λ_2 is small, and the "rate-based" policy is effective when this difference is considerably large. In other cases of λ_1 and λ_2, the "rate-based" policy performs well when the holding cost c_h is very small compared to the transmission cost c_t, and the "transmit-all" policy performs very well at higher values of c_h.

Many extensions of this work can be considered for future research. The time period T between consecutive transmission opportunities can be considered as a random variable, which is possible in certain applications. It would be also worthwhile to see if our results can be extended to the case of a relay which serves as a connection between multiple pairs of nodes.

Finally, more effective distributed policies can be explored for managing the energy-delay trade-off in network coding in large wireless networks.

Acknowledgment

This material is based upon work partially supported by the NSF under grants CNS-0954153, CNS-0963818, CNS-0904520 and CNS-1149458, by the AFOSR under contract No. FA9550-13-1-0008 and the DTRA under grant HDTRA1-13-1-0030. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsoring agencies.

The authors would also like to thank the associate editor and reviewers for their comments which significantly improved the content and presentation of this paper.

Appendix I

A. Proof of Proposition 1

Consider a policy θ where the decision is to transmit all the queued packets at every transmission opportunity. Note that, under this policy, a total number of $\max(i,j)$ packets will be transmitted in state (i,j). As per (4), the total discounted cost under this policy θ is given by

$$v_{\beta,\theta}(i,j) = c_t \max(i,j) + \mathrm{E}\left[\sum_{n=1}^{\infty} \beta^n c_t \max(A_n^1, A_n^2)\right],$$

where A_n^1 and A_n^2 are the number of packet arrivals to queues 1 and 2 respectively between the $(n-1)$-st and n-th transmission opportunities. Note that

$$\begin{aligned} v_\beta(i,j) &\leq v_{\beta,\theta}(i,j) \\ &\leq c_t \max(i,j) + \mathrm{E}\left[\sum_{n=1}^{\infty} \beta^n c_t (A_1^n + A_2^n)\right] \\ &= c_t \max(i,j) + \frac{\beta c_t T(\lambda_1+\lambda_2)}{1-\beta} < \infty. \end{aligned}$$

Since initial state (i,j) and discount factor β are arbitrary, $v_\beta(i,j) < \infty$ for every (i,j) and β.

B. Proof of Proposition 2

We will prove this result by using induction on the steps of the value iteration algorithm [12]. Based on (5), the discounted cost function in the n-th step of value iteration is given as

$$\begin{aligned} v_{\beta,n}(i,j) &= \min_{a\in\mathcal{A}_{i,j}} \Bigg\{ c(i,j,a) \\ &\quad + \beta \sum_{k,l} p_k^1 p_l^2 \; v_{\beta,n-1}\left((i-a)^+ + k, (j-a)^+ + l\right) \Bigg\}, \\ &\qquad (i,j) \in \mathcal{S}. \end{aligned} \tag{17}$$

At the start of value iteration, $v_{\beta,0}(i,j) = 0$ for every state (i,j). Hence, for $n=0$ case, $v_{\beta,n}(i,j)$ is nondecreasing in i and j.

Now, suppose the result is true for $n-1$, i.e. $v_{\beta,n-1}(i,j)$ is nondecreasing in i and j. Using (17), select an action $a \in \mathcal{A}_{i+1,j}$ such that

$$\begin{aligned} &v_{\beta,n}(i+1,j) = c(i+1,j,a) \\ &\quad + \beta \sum_{k,l} p_k^1 p_l^2 \; v_{\beta,n-1}\left((i+1-a)^+ + k, (j-a)^+ + l\right). \end{aligned} \tag{18}$$

For the same action a, we must have from (17):

$$\begin{aligned} &v_{\beta,n}(i,j) \leq c(i,j,a) \\ &\quad + \beta \sum_{k,l} p_k^1 p_l^2 \; v_{\beta,n-1}\left((i-a)^+ + k, (j-a)^+ + l\right). \end{aligned} \tag{19}$$

Note that $a \notin \mathcal{A}_{i,j}$ when $a = i+1 > j$, but (19) will still hold. Since $v_{\beta,n-1}(i,j)$ and $c(i,j,a)$ (see (2)) are nondecreasing in i, we use (18) and (19) to show that

$$\begin{aligned} &v_{\beta,n}(i+1,j) - v_{\beta,n}(i,j) \\ &\geq c(i+1,j,a) - c(i,j,a) \\ &+ \beta \sum_{k,l} p_k^1 p_l^2 \Bigg\{ v_{\beta,n-1}\left((i+1-a)^+ + k, (j-a)^+ + l\right) \\ &- v_{\beta,n-1}\left((i-a)^+ + k, (j-a)^+ + l\right) \Bigg\} \geq 0. \end{aligned}$$

Hence $v_{\beta,n}(i,j)$ is nondecreasing in i for any fixed j. Similarly we can show that $v_{\beta,n}(i,j)$ is nondecreasing in j for any fixed i. Thus $v_{\beta,n}(i,j)$ is nondecreasing in both i and j. Therefore the discounted cost function $v_\beta(i,j)$ is nondecreasing in i and j, as $v_{\beta,n}(i,j) \to v_\beta(i,j)$.

C. Proof of Proposition 3

Consider again the policy θ where the decision is to transmit all the queued packets at every transmission opportunity. Under this policy, the state of the system $\{(S_n^1, S_n^2), n \geq 0\}$ can be described by an irreducible and ergodic Markov chain with transition probabilities $p_{(i,j)(k,l)} := p_k^1 p_l^2$ for $(i,j),(k,l) \in \mathcal{S}$. Note that the stationary distribution of this Markov chain is $\pi_{i,j} := p_i^1 p_j^2$ for $(i,j) \in \mathcal{S}$. Therefore the long-run average cost under policy θ is given by

$$\begin{aligned} g(\theta) &= \sum_{(i,j)\in\mathcal{S}} \pi_{i,j} \; c_t \max(i,j) \\ &\leq c_t \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p_i^1 p_j^2 (i+j) = c_t(\lambda_1+\lambda_2)T < \infty. \end{aligned}$$

This shows that the average cost in the considered policy is finite.

D. Proof of Lemma 1

When $i \geq j$, we have

$$\begin{aligned} g(i,j) &= \min_{a\in\{j,\cdots,i\}} \{ca + f(i-a,0)\} \\ &= ci + \min_{k\in\{0,\cdots,i-j\}} f_1(k), \end{aligned}$$

where $f_1(k) = -ck + f(k,0)$. Since the functions $-ck$ and $f(k,0)$ are convex (by Definition 1), $f_1(k)$ is convex. Let

$k_1^* = \arg\min\{f_1(k) : k \geq 0\}$ be a global minimum of the function f_1. Using convexity of f_1, we have

$$g(i,j) = \begin{cases} ci + f_1(i-j) & \text{if } 0 \leq i-j < k_1^*, \\ ci + f_1(k_1^*) & \text{if } i-j \geq k_1^*. \end{cases} \quad (20)$$

Similarly, when $i \leq j$, we can show that there exists a $k_2^* \geq 0$ such that

$$g(i,j) = \begin{cases} cj + f_2(j-i) & \text{if } -k_2^* < i-j \leq 0, \\ cj + f_2(k_2^*) & \text{if } i-j \leq -k_2^*, \end{cases} \quad (21)$$

where $f_2(k) = -ck + f(0,k)$. Also note that $f_1(0) = f_2(0) = f(0,0)$.

Now, by Definition 2, the function $g(i,j)$ will be convex in i, if

$$g(i+1,j) - g(i,j) \geq g(i,j) - g(i-1,j), \ \ i \geq 1, j \geq 0. \quad (22)$$

Similarly $g(i,j)$ will be convex in j, if

$$g(i,j+1) - g(i,j) \geq g(i,j) - g(i,j-1), \ \ i \geq 0, j \geq 1. \quad (23)$$

We will show that (22) and (23) hold in all the following cases: (a) $i > j$, (b) $i < j$, and (c) $i = j$.

First, in case (a) $i > j$, we consider two sub-cases: (a1) $0 < i-j < k_1^*$, and (a2) $i - j \geq k_1^*$. In sub-case (a1) $0 < i-j < k_1^*$, using (20) and convexity of f_1, we show in the following that (22) and (23) are satisfied.

$$\begin{aligned} & g(i+1,j) - g(i,j) \\ & = c(i+1) + f_1(i-j+1) - ci - f_1(i-j) \\ & \geq [ci + f_1(i-j)] - [c(i-1) + f_1(i-j-1)] \\ & = g(i,j) - g(i-1,j). \end{aligned}$$

$$\begin{aligned} & g(i,j+1) - g(i,j) \\ & = ci + f_1(i-j-1) - ci - f_1(i-j) \\ & \geq [ci + f_1(i-j)] - [ci + f_1(i-j+1)] \\ & = g(i,j) - g(i,j-1). \end{aligned}$$

Also, in sub-case (a2) $i - j \geq k_1^*$, we show in the following that (22) and (23) are satisfied.

$$\begin{aligned} g(i+1,j) - g(i,j) &= c(i+1) + f_1(k_1^*) - ci - f_1(k_1^*) \\ &= [ci + f_1(k^*)] - [c(i-1) + f_1(k_1^*)] \\ &\geq g(i,j) - g(i-1,j). \\ g(i,j+1) - g(i,j) &\geq [ci + f_1(k^*)] - [ci + f_1(k_1^*)] \\ &= g(i,j) - g(i,j-1). \end{aligned}$$

Thus both (22) and (23) are satisfied in case (a) $i > j$. Similarly, by using (21) and convexity of f_2, we can show that these conditions are also satisfied in case (b) $i < j$. In case (c) $i = j$, note that (22) will hold if

$$\begin{aligned} & g(i+1,i) - g(i,i) \geq g(i,i) - g(i-1,i), \\ \text{i.e.} \quad & \min\{ci + f(1,0), c(i+1) + f(0,0)\} - ci - f(0,0) \\ & \geq ci + f(0,0) - \min\{c(i-1) + f(0,1), ci + f(0,0)\}, \end{aligned}$$

i.e.

$$\min\{f(1,0) - f(0,0), c\} + \min\{f(0,1) - f(0,0), c\} \geq c,$$

which is the condition specified in (7). Similarly it can be shown that, when this condition holds, we also have $g(i,i+1) - g(i,i) \geq g(i,i) - g(i,i-1)$.

Thus, given that (7) holds, $g(i,j)$ satisfies (22) and (23) at all points. Therefore $g(i,j)$ is convex in i and j.

E. Proof of Proposition 4

We will prove this result by induction on the steps of the value iteration algorithm. The discounted cost function in the n-th step of value iteration is given by

$$\begin{aligned} v_{\beta,n}(i,j) \;\; = \;\; & \min_{a \in \mathcal{A}_{i,j}} \Bigg\{ c_t a + c_h \left[(i-a)^+ + (j-a)^+ \right] \\ & + \beta \sum_{k,l} p_k^1 p_l^2 \ v_{\beta,n-1}\left((i-a)^+ + k, (j-a)^+ + l \right) \Bigg\}, \\ & s(i,j) \in \mathcal{S}. \quad (24) \end{aligned}$$

At the start of value iteration, $v_{\beta,0}(i,j) = 0$ for every state (i,j). Hence, for $n = 0$ case, $v_{\beta,n}(i,j)$ is convex in i and j.

Now, suppose the result is true for $n-1$, i.e. $v_{\beta,n-1}(i,j)$ is convex in i and j. Therefore, the function $f_{n-1}(i,j) := \sum_{k,l} p_k^1 p_l^2 \ v_{\beta,n-1}(i+k, j+l)$ is convex in i and j. Now we can write (24) as

$$v_{\beta,n}(i,j) = \min_{a \in \mathcal{A}_{i,j}} \left\{ c_t a + g_{n-1}\left((i-a)^+, (j-a)^+ \right) \right\}, \quad (25)$$

where $g_{n-1}(i,j) = c_h(i+j) + \beta f_{n-1}(i,j)$. Since $c_h(i+j)$ and $f_{n-1}(i,j)$ are convex in i and j, $g_{n-1}(i,j)$ is convex in i and j. Therefore by Lemma 1, in (25), $v_{\beta,n}(i,j)$ is convex in i and j if

$$\begin{aligned} & \min\{g_{n-1}(1,0) - g_{n-1}(0,0), c_t\} \\ & + \min\{g_{n-1}(0,1) - g_{n-1}(0,0), c_t\} \geq c_t. \quad (26) \end{aligned}$$

Since $v_{\beta,n}(i,j)$ is nondecreasing in i and j in every stage of value iteration (see proof of Proposition 2), we have

$$\begin{aligned} & g_{n-1}(1,0) - g_{n-1}(0,0) \\ & = c_h + \beta \sum_{k,l} p_k^1 p_l^2 \ \left[v_{\beta,n-1}(k+1,l) - v_{\beta,n-1}(k,l) \right] \geq c_h, \end{aligned} \quad (27)$$

$$\begin{aligned} & g_{n-1}(0,1) - g_{n-1}(0,0) \\ & = c_h + \beta \sum_{k,l} p_k^1 p_l^2 \ \left[v_{\beta,n-1}(k,l+1) - v_{\beta,n-1}(k,l) \right] \geq c_h. \end{aligned} \quad (28)$$

Using (27) and (28), the sufficient condition for (26) to hold is $2c_h \geq c_t$. However, this is the necessary condition in case $n = 1$ where (27) and (28) are satisfied at equality. Thus $v_{\beta,n}(i,j)$ is convex in i and j if $c_h \geq c_t/2$. Therefore, under the same condition, the discounted cost function $v_\beta(i,j)$ is convex in i and j, as $v_{\beta,n}(i,j) \to v_\beta(i,j)$.

References

[1] M. Médard and A. Sprintson, *Network coding: Fundamentals and applications*. Academic Press, 2011.

[2] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Médard, and J. Crowcroft, “Xors in the air: practical wireless network coding,” *ACM SIGCOMM Computer Communication Review*, vol. 36, no. 4, pp. 243–254, 2006.

[3] Y. Wu, P. A. Chou, and S.-Y. Kung, "Information exchange in wireless networks with network coding and physical-layer broadcast," in *Proceedings of the 39th Annual Conference on Information Sciences and Systems (CISS)*, 2005.

[4] R. Ahlswede, N. Cai, S. R. Li, and R. W. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, 2000.

[5] X. He and A. Yener, "On the energy-delay trade-off of a two-way relay network," in *Proceedings of the 42nd Annual Conference on Information Sciences and Systems (CISS)*, 2008, pp. 865–870.

[6] O. H. Abdelrahman and 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.

[7] J. Charith Gunasekara, A. Alfa, and P. Yahampath, "A queueing theoretic model for opportunistic network coding," in *International Conference on Computing, Networking and Communications (ICNC)*, 2013, pp. 999–1004.

[8] W. Chen, K. B. Letaief, and Z. Cao, "Opportunistic network coding for wireless networks," in *IEEE International Conference on Communications (ICC)*, 2007, pp. 4634–4639.

[9] J. Goseling, R. J. Boucherie, and J.-K. Van Ommeren, "Energy–delay tradeoff in a two-way relay with network coding," *Performance Evaluation*, vol. 70, no. 11, pp. 981–994, 2013.

[10] E. Ciftcioglu, Y. Sagduyu, R. Berry, and A. Yener, "Cost-delay tradeoffs for two-way relay networks," *IEEE Transactions on Wireless Communications*, vol. 10, no. 12, pp. 4100–4109, 2011.

[11] Y.-P. Hsu, N. Abedini, S. Ramasamy, N. Gautam, A. Sprintson, and S. Shakkottai, "Opportunities for network coding: To wait or not to wait," in *Proceedings of the IEEE International Symposium on Information Theory Proceedings (ISIT)*, 2011, pp. 791–795.

[12] M. Puterman, *Markov decision processes: Discrete stochastic dynamic programming*. John Wiley & Sons, Inc., 1994.

[13] A. Arapostathis, V. S. Borkar, E. Fernández-Gaucherand, M. K. Ghosh, and S. I. Marcus, "Discrete-time controlled markov processes with average cost criterion: a survey," *SIAM Journal on Control and Optimization*, vol. 31, no. 2, pp. 282–344, 1993.

[14] V. S. Borkar, "Control of markov chains with long-run average cost criterion: The dynamic programming equations," *SIAM Journal on Control and Optimization*, vol. 27, no. 3, pp. 642–657, 1989.

[15] R. Cavazos-Cadena, "Necessary conditions for the optimality equation in average-reward markov decision processes," *Applied Mathematics and Optimization*, vol. 19, pp. 97–112, 1989.

[16] R. Cavazos-Cadena and L. I. Sennott, "Comparing recent assumptions for the existence of average optimal stationary policies," *Operations Research Letters*, vol. 11, no. 1, pp. 33–37, 1992.

[17] L. I. Sennott, "A new condition for the existence of optimal stationary policies in average cost markov decision processes," *Operations Research Letters*, vol. 5, no. 1, pp. 17–23, 1986.

[18] ——, "Average cost optimal stationary policies in infinite state markov decision processes with unbounded costs," *Operations Research*, vol. 37, no. 4, pp. 626–633, 1989.

[19] M. Schäl, "Average optimality in dynamic programming with general state space," *Mathematics of Operations Research*, vol. 18, no. 1, pp. 163–172, 1993.

[20] M. Hundebøll, J. Ledet-Pedersen, J. Heide, M. V. Pedersen, S. A. Rein, and F. H. Fitzek, "Catwoman: Implementation and performance evaluation of ieee 802.11 based multi-hop networks using network coding," in *IEEE Vehicular Technology Conference (VTC)*, 2012, pp. 1–5.

[21] S. Resnick, *Adventures in stochastic processes*. Birkhäuser Boston, 1992.

[22] L. I. Sennott, *Stochastic Dynamic Programming and the Control of Queueing Systems*, 3rd ed. Wiley Inter-science, New York, 1999.

[23] K. Murota, "Recent developments in discrete convex analysis," in *Research Trends in Combinatorial Optimization*. Springer, 2009, pp. 219–260.

[24] J. Skellam, "The frequency distribution of the difference between two poisson variates belonging to different populations." *Journal of the Royal Statistical Society*, vol. 109, no. 3, pp. 296–296, 1946.

Arupa Mohapatra is a Member of Technical Staff at Oracle Corporation. He received his B.Tech. degree in electrical and electronics engineering from the National Institute of Technology, Tiruchirappalli, India and Ph.D. in industrial and systems engineering from Texas A&M University, College Station.

His research interests are in the areas of performance evaluation and optimization in various networked systems with a focus on wireless communication networks and transportation networks.

Natarajan Gautam is an Associate Professor in the Department of Industrial and Systems Engineering at Texas A&M University with a courtesy appointment in the Department of Electrical and Computer Engineering. Prior to joining Texas A&M University in 2005, he was on the Industrial Engineering faculty at Penn State University for eight years. He received his M.S. and Ph.D. in Operations Research from the University of North Carolina at Chapel Hill, and his B.Tech. from Indian Institute of Technology, Madras.

His research interests are in the areas of modeling, analysis and performance evaluation of stochastic systems with special emphasis on optimization and control in computer, telecommunication and information systems. He is an Associate Editor for the INFORMS Journal on Computing, IIE Transactions, and OMEGA.

Srinivas Shakkottai (S '00-M '08) received his Bachelor of Engineering degree in electronics and communication engineering from the Bangalore University, India, in 2001 and his M.S. and Ph.D degrees from the University of Illinois at Urbana-Champaign in 2003 and 2007, respectively, both in electrical engineering. He was Postdoctoral Scholar at Stanford University until December 2007, and is currently an Assistant Professor in the Department of Electrical and Computer Engineering at Texas A&M University. He has received the Defense Threat reduction Agency Young Investigator Award (2009) and the NSF Career Award (2012) as well as research awards from Cisco (2008) and Google (2010).

His research interests include content distribution systems, pricing approaches to network resource allocation, game theory, congestion control, and the measurement and analysis of Internet data.

Alex Sprintson (S '00-M '03) is an Associate Professor with the Department of Electrical and Computer Engineering at Texas A&M University, College Station. From 2003 to 2005, he was a Postdoctoral Research Fellow with the California Institute of Technology, Pasadena. His research interests lie in the general area of communication networks with a focus on network coding and software defined networks.

He has received the Wolf Award for Distinguished Ph.D. students, the Viterbi Postdoctoral Fellowship, and the NSF CAREER award. Currently, he serves as an Associate Editor for the IEEE Transactions on Wireless Communications. He has been a member of the Technical Program Committee for the IEEE Infocom 2006–2015.