Energy Minimization in Two-Way Relay Networks with Digital Network Coding

(Invited Paper)

Zhi Chen
Dept. of Elect. & Comp. Eng.
National Univ. of Singapore
E-mail: elecz@nus.edu.sg

Teng Joon Lim
Dept. of Elect. & Comp. Eng.
National Univ. of Singapore
E-mail: eleljt@nus.edu.sg

Mehul Motani
Dept. of Elect. & Comp. Eng.
National Univ. of Singapore
E-mail: elemm@nus.edu.sg

Abstract—In a two-way wireless relay network, two nodes communicate with each other through a relay, and throughput is enhanced by the relay transmitting a linear combination of the messages received from the two sources either at the bit level (digital network coding) or the signal level (analog network coding). In this paper, we solve for the fraction of time resources allocated to each transmission mode in a DNC-based two-way relay network that minimizes total energy while ensuring queue stability at all nodes, for a given pair of random packet arrival rates. We also provide a queueing analysis of the relaying protocol designed.

Index Terms—Two-way relays, network coding, energy efficiency, queue stability

I. INTRODUCTION

Network coding [1] has emerged as a viable means to improve throughput in complex networks. Messages at the packet level are linearly combined at intermediate nodes and forwarded to multiple intended destinations. Using knowledge of the manner in which messages were combined, communicated through some additional overhead bits, as well as knowledge of the message it contributed, a destination can successfully reconstruct the message intended for it. In this way, network throughput is greatly improved. Due to the broadcast nature of wireless communication, a transmission can potentially be decoded by several intermediate nodes, and hence network coding promises throughput improvements in wireless systems with multiple nodes or users.

The two-way relay network [2] exemplifies the use of network coding in wireless communication. This simple network is usually comprised of two source nodes (S_1 and S_2) and one relay node (R), where S_1 and S_2 have information to exchange with each other. A direct link between S_1 and S_2 is assumed to be unavailable. This model applies for instance to communication between a base station and a mobile user in cellular communications, in which the mobile user is in a location shadowed from the base station and coverage is provided in the area by means of a relay. Another typical application is that of bidirectional communication between two nodes located on opposite sides of a large obstruction (such as a hill) with a relay node at the top of it.

With the aid of network coding, we can exchange two messages in two time slots with analog network coding [4], [5] or three slots with digital network coding [2], [3], compared with the four slots needed with pure forwarding. With network coding, the relay first receives messages from both source nodes. It then combines these messages and broadcasts the network-coded packet to both sources. As each source has prior knowledge of messages generated by itself, it can recover the desired messages. The rate region for DNC-based two-way relay networks is explored in [2], [3]. The former also investigated queue stability in the case of random data arrivals.

In addition to throughput, resource allocation was also investigated in [6], [7], [9], [10]. Optimal resource allocation for analog network coding in a MIMO two-way relay network was investigated in [4]. In [6] and [7], optimal resource allocation for an ANC-based system was considered for OFDM systems. In [9], rate fairness was studied with optimal resource allocation. In [10], resource allocation was investigated under the scenario of asymmetric multi-way relay communication over orthogonal channels.

In [8], the fading nature of a wireless channel was taken into consideration and the optimal position for the relay node was investigated. In [11], outage regions of DNC and ANC strategies were derived. It found that with the presence of a direct link, DNC was more promising than ANC in most cases in terms of outage performance.

In this paper, we formulate, simplify and solve the problem of allocation of channel resources to the various transmission modes in a DNC-based two-way relay network to minimize total energy used in transmission. The constraint is that queue stability is maintained at all four queues in the network for a given pair of random packet arrival rates at S_1 and S_2. The scheduler, which we take to be the relay node, is assumed to have full knowledge of the instantaneous channel coefficients in the four flat-fading links. We also design and simulate a scheduling protocol based on the designed resource allocation policy, and demonstrate that a substantial saving in total transmitted energy is accomplished through the proposed resource allocation method.

Notation: Scalars are denoted by lower-case letters and bold-face lower-case letters are used for vectors and bold-face upper-case letters for matrices. In addition, we use \( \mathbf{I} \) and \( \Theta \) to denote the identity matrix and the all-unity matrix respectively. \( \mathbf{I} \) is the vector with all elements equal to unity.

II. SYSTEM MODEL AND PROBLEM FORMULATION

Consider a two-way relay network consisting of two sources S_1, S_2 and one relay node R. S_1 needs to communicate with S_2 and vice versa, but the direct link between them cannot
support the required data rate. We assume that packets arrive at $S_i$ according to a Poisson process at an average arrival rate of $\lambda_i$, $i \in \{1, 2\}$. We seek to maximize the energy efficiency of this system for a given average arrival rate pair $(\lambda_1, \lambda_2)$ while maintaining queue stability, by adjusting the fraction of time allocated to each of the five transmission modes described below.

- **Mode 1:** $S_1$ transmits to $R$ at rate $R_1$ (uplink phase).
- **Mode 2:** $S_2$ transmits to $R$ at rate $R_2$ (uplink phase).
- **Mode 3:** $R$ broadcasts to $S_1$ and $S_2$ at broadcast rate $R_3$ (Network coding in downlink phase).
- **Mode 4:** $R$ transmits only to $S_1$ at rate $R_4$ (one-way forwarding in downlink phase).
- **Mode 5:** $R$ transmits only to $S_2$ at rate $R_5$ (one-way forwarding in downlink phase).

A block flat-fading channel model is assumed for all links. The four channel gains are $g_{1r}$, $g_{2r}$, $g_{r1}$ and $g_{r2}$, corresponding to modes 1, 2, 4 and 5 respectively. Whether $g_{ri} = g_{ri}$ is inconsequential in this work. Noise at each node is assumed to be i.i.d. Gaussian random variables with zero mean and unit variance. The queue lengths (in terms of packets) at $S_1$ and $S_2$ at epoch $t$ are denoted as $Q_1(t)$ and $Q_2(t)$. We also define $Q_{r1}(t)$ and $Q_{r2}(t)$ as the lengths of the two queues maintained at the relay node for $S_1$ and $S_2$, respectively.

The transmit power for each transmission mode is an exponential function of the transmit rate, given by

$$P_i(R_i) = \frac{2R_i - 1}{g_{ri}}$$

for the five modes, respectively, if we ignore the SNR gap to capacity\(^1\). It should be noted that in Mode 3, the rate of $R_3$ is the rate transmitted to each of $S_1$ and $S_2$, due to the use of network coding.

We assume that in each time slot, a fraction $f_i$, $i = 1, \ldots, 5$, is allocated to Mode $i$. Therefore $f_i P_i(R_i)$ is proportional to the energy used for transmission in Mode $i$. To minimize the total energy consumption per time slot while satisfying queue stability constraints therefore requires the solution of optimization problem P1:

$$\min_{f_i, R_i} \sum_{i=1}^5 f_i P_i(R_i)$$

subject to the queue stability constraints

$$\lambda_1 \leq \min(f_1 R_1, f_3 R_3 + f_5 R_5)$$

$$\lambda_2 \leq \min(f_2 R_2, f_3 R_3 + f_4 R_4)$$

and the physical constraints $\sum_{i=1}^5 f_i \leq 1$ and $f_i \geq 0$. In the next section, we show that P1 can be transformed into a simpler convex optimization problem with only $f_i$ as the design parameters and hence can be solved easily.

## III. Problem Solution

Problem P1 may be simplified through the following lemmas.

**Lemma 1:** Under the optimal solution to P1, we have

$$\begin{cases}
  f_4 = 0, & \text{if } \lambda_1 > \lambda_2 \\
  f_5 = 0, & \text{if } \lambda_1 < \lambda_2
\end{cases}$$

and $f_4 = f_5 = 0$ if $\lambda_1 = \lambda_2$.

**Proof:** Denote the optimal values of $f_i$ and $R_i$ by $f_i^*$ and $R_i^*$. Constraints (7) and (8) dictate that

$$f_3^* R_3^* + f_5^* R_5^* \geq \lambda_1$$

$$f_3^* R_3^* + f_4^* R_4^* \geq \lambda_2$$

Assume that $f_3^* R_3^* + f_5^* R_5^* - \lambda_2 = \epsilon > 0$. Then reducing $f_4^*$ by $\epsilon / R_4$ while keeping all other $f_i^*$ and all $R_i^*$ fixed results in constraint (8) still being satisfied, with a lower total energy. Therefore we can assume that $f_3^* R_3^* + f_4^* R_4^* = \lambda_2$. Similarly, $f_3^* R_3^* + f_5^* R_5^* = \lambda_1$.

Suppose that $f_3^* > 0$. If we reduce $f_3^*$ by $\delta_3$, constraint (8) requires that $f_4^*$ be increased by $\delta_3 R_4^* / R_3^*$. This in turn means that $f_3^*$ can be reduced by $\delta_3 R_3^* / R_3^*$. In other words, queue stability is maintained under the following adjustment:

$$f_3^* \rightarrow f_3^* - \delta_3$$

$$f_4^* \rightarrow f_4^* + \delta_3 R_4^* / R_3^*$$

$$f_5^* \rightarrow f_5^* - \delta_3 R_4^* / R_5^*$$

The total energy used is now reduced by

$$\Delta E = \delta_3 \left( \frac{R_4^*}{R_5^*} P_4^* + \frac{P_4^* - R_4^*}{R_3^*} P_3^* \right)$$

where $P_i^*$ is shorthand for $P_i(R_i^*)$. Note that $R_3^*$, $R_4^*$ and $R_5^*$ need only satisfy two equations equivalent to

$$f_4^* R_4^* - f_5^* R_5^* = \lambda_1 - \lambda_2$$

$$f_3^* R_3^* + f_5^* R_5^* = \lambda_1$$

We can thus always find a point in the solution space of (13) to make the term in brackets in (12) positive. So as long as $\delta_4 > 0$ does not lead to a violation of (13), we can reduce total energy by $\Delta E > 0$. If $\lambda_1 > \lambda_2$, this result and (13) show that we should have $f_4^* = 0$, and $f_5^* R_5^* = \lambda_1 - \lambda_2$.

Similarly, if $\lambda_1 < \lambda_2$, then $f_5^* = 0$ and $f_4^* R_4^* = \lambda_2 - \lambda_1$; and if $\lambda_1 = \lambda_2$, then $f_4^* = f_5^* = 0$.

A corollary of Lemma 1 is that, since either one or both of $f_4^*$ and $f_5^*$ must be 0, $f_2^*$ cannot be zero, i.e. the broadcast phase (Mode 3) must be used. This is intuitive because one bit transmitted in Mode 3 is equivalent to two bits delivered, and hence we should devote as much resources as possible to Mode 3. The following lemma links $R_i$ to $f_i$.

\(^1\) SNR gap term $\Gamma$ can be inserted into each expression but this only changes the effective channel gain to $g_{ri} / \Gamma$ and does not affect the following discourse.
Lemma 2: The solution to P1 satisfies
\[ R^*_i = \frac{\lambda_i}{f_i}, \quad i = 1, 2 \] (14)
\[ R^*_3 = \frac{\min(\lambda_1, \lambda_2)}{f_3} \] (15)
\[ R^*_4 = \max \left( \frac{\lambda_2 - \lambda_1}{f_i}, 0 \right) \] (16)
\[ R^*_5 = \max \left( \frac{\lambda_1 - \lambda_2}{f_5}, 0 \right) \] (17)

Proof: By reducing \( f_1 \) and/or \( R_1 \), total energy is reduced, and so the energy-minimization problem must be solved when \( f_1 R_1 \) is at its smallest value, i.e. \( f_1^* R_1^* = \lambda_1 \). Similarly, \( f_2^* R_2^* = \lambda_2 \).

Equations (16) and (17) arise directly from the proof of Lemma 1. Equation (15) comes from noting that if \( \lambda_1 < \lambda_2 \), Lemma 1 states that \( f_3^* = 0 \) and hence \( f_3^* R_3^* = \lambda_1 \) while if \( \lambda_1 > \lambda_2 \), \( f_3^* R_3^* = \lambda_2 \).

Lemma 2 implies that \( \{R_i\} \) can be removed as optimization variables, leaving only \( \{f_i\} \) in an equivalent optimization problem. Lemma 1 (and also Lemma 2) implies that when \( \lambda_1 > \lambda_2 \), Mode 4 (forwarding from the relay to \( S_1 \)) is not necessary, and that when \( \lambda_1 < \lambda_2 \), Mode 5 is not necessary. Hence there is no loss of generality in assuming that \( \lambda_1 < \lambda_2 \) from this point on, so that Mode 5 is no longer discussed and the min and max functions need not be invoked in Lemma 2. We now have
\[ R^*_3 = \frac{\lambda_1}{f_3}, \quad R^*_4 = \frac{\lambda_2 - \lambda_1}{f_4} \quad \text{and} \quad R^*_5 = 0. \] (18)

By substituting the new expressions for \( \{R_i\} \) into P1, assuming \( \lambda_1 < \lambda_2 \), we get the equivalent optimization problem P2:
\[
\begin{align*}
\min_{\{f_i\}} & \quad f_1 P_1 \left( \frac{\lambda_1}{f_1} \right) + f_2 P_2 \left( \frac{\lambda_2}{f_2} \right) + f_3 P_3 \left( \frac{\lambda_1}{f_3} \right) \\
& \quad + f_4 P_4 \left( \frac{\lambda_2 - \lambda_1}{f_4} \right) \\
\text{s.t.} & \quad f_i \geq 0 \\
& \quad \sum_{i=1}^{4} f_i \leq 1
\end{align*}
\] (19)

where the power functions \( P_i(\cdot) \) are given in (1) to (5).

It is not difficult to show that the cost function in P2 is convex in \( \{f_i\} \), and clearly the constraint set is also convex. Therefore the original problem P1 has been turned into an equivalent convex optimization problem P2.

It should be noted too that the cost function in P2 is monotonically decreasing in \( \{f_i\} \) and hence its solution must lie on the boundary of the constraint set. Since \( f_i = 0 \) for all \( i \) is clearly not a viable solution, it must be that \( \sum_i f_i = 1 \) i.e., that all available time resources are fully utilized. This is easily understandable since increasing any \( f_i \) leads to a corresponding reduction in the associated \( R_i \), which means an exponential decrease in required power and hence we would want \( f_i \) to be as large as possible.

Since P2 is a constrained convex optimization problem, its solution is found by solving the the Karush-Kuhn-Tucker (KKT) equations, which are easily derived as
\[
\begin{align*}
& \quad 2 \frac{\lambda_1 - \lambda_2}{f_2} \ln 2 + \beta - 1 = 0 \quad g_{1r} = 0 \quad \text{(22)} \\
& \quad 2 \frac{\lambda_1 - \lambda_2}{f_2} \ln 2 + \beta - 1 = 0 \quad g_{2r} = 0 \quad \text{(23)} \\
& \quad 2 \frac{\lambda_1 - \lambda_2}{f_2} \ln 2 + \beta - 1 = 0 \quad g_{\min} = 0 \quad \text{(24)} \\
& \quad \sum_{i=1}^{4} f_i - 1 = 0 \quad \text{(26)}
\end{align*}
\]

where \( g_{\min} = \min(g_{1r}, g_{2r}) \) and \( \beta \) is the Lagrange multiplier. Since (22)–(25) are transcendental equations, it is infeasible to obtain explicit solutions in general and numerical methods are instead employed to obtain the solution to problem P2.

However, for the case that \( \lambda_1 \) and \( \lambda_2 \) are very small (\( \lambda_i \ll 1 \)), we can use approximations for the exponential functions and obtain a closed form solution, with \( \beta \) obtained through a one-dimensional bisection search:
\[
\begin{align*}
& \quad f_1 = \lambda_1 \left( 1 + \beta - \frac{1}{g_{1r}} \right)^\frac{1}{2} \quad \text{(27)} \\
& \quad f_2 = \lambda_2 \left( 1 + \beta - \frac{1}{g_{2r}} \right)^\frac{1}{2} \quad \text{(28)} \\
& \quad f_3 = \lambda_1 \left( 1 + \beta - \frac{1}{g_{\min}} \right)^\frac{1}{2} \quad \text{(29)} \\
& \quad f_4 = (\lambda_2 - \lambda_1) \left( 1 + \beta - \frac{1}{g_{1r}} \right)^\frac{1}{2} \quad \text{(30)} \\
& \quad f_4 = 1 - f_1 - f_2 - f_3 \quad \text{(31)}
\end{align*}
\]

Finally, we should emphasize that when \( \lambda_1 > \lambda_2 \), the last two terms in the cost function of P2 change to
\[
\begin{align*}
& \quad f_3 P_3 \left( \frac{\lambda_2}{f_3} \right) + f_5 P_5 \left( \frac{\lambda_1 - \lambda_2}{f_5} \right)
\end{align*}
\]

and the problem can be solved with the appropriate substitutions in the steps described in this section.

Note that since data arrivals are stochastic, the queues at the sources and the relay have a non-zero probability of being empty. Thus, the actual energy used for transmission is upper bounded by \( \sum_i f_i P_i(R_i) \).

IV. SCHEDULING PROTOCOL AND QUEUE ANALYSIS

We now consider scheduling strategies which make use of the optimal time-sharing fractions \( \{f_i\} \) (and hence rates \( \{R_i\} \)) for each transmission mode. We describe below an intuitive protocol called the simple random scheduling protocol (SRSP).

1) At the start of a packet interval of duration \( T \), the relay randomly chooses a transmission mode with probability \( P[\text{Mode } i] = f_i^* \).
2) If Mode 1 is selected, the relay will inform \( S_1 \) to transmit. If Mode 2 is selected, the relay will inform
and the stationary distribution is $\pi = \mathbf{1} \cdot (\mathbf{I} - \mathbf{P}_1 + \Theta)^{-1}$. The average delay at $S_1$, via Little’s Theorem, is

$$D_1 = \frac{Q_1}{\lambda_1}. \quad (38)$$

V. SIMULATION

We now present simulations to verify our findings and explore tradeoffs. To ensure reasonable running times and memory requirements, we design for average arrival rates that are slightly larger than the actual rates, i.e., $\lambda_i \rightarrow (1 + \epsilon)\lambda_i$ ($i = 1, 2$), where $\epsilon$ is a small positive value.

In Fig. 1, the data arrival rate at $S_2$ is $\lambda_2 = 1$. Noise at each node is assumed to be with zero mean and unit variance. An iterative sequential quadratic program (SQP) is used to solve $\mathbf{P}_1$ directly, and $\mathbf{P}_2$ is solved through a numerical solution of the KKT equations. It can be observed that the solution to $\mathbf{P}_1$ and the solution to $\mathbf{P}_2$ matches for different channel gain realizations and varying $\lambda_1$. The equivalence of problems $\mathbf{P}_1$ and $\mathbf{P}_2$ is therefore verified.

In Fig. 2, the energy efficiency of digital network coding and the conventional scheme (without network coding) are compared for some specified channel realizations. The energy required for supporting data arrival rate with conventional transmission is obtained by solving the problem below.

$$\min_{f_i} \quad f_1P_1\left(\frac{\lambda_1}{f_1}\right) + f_2P_2\left(\frac{\lambda_2}{f_2}\right)$$
$$+ f_4P_4\left(\frac{\lambda_4}{f_4}\right) + f_5P_5\left(\frac{\lambda_5}{f_5}\right) \quad (39)$$
$s.t.$

$$f_i \geq 0 \quad (40)$$
$$\sum_{i=1}^5 f_i = 1 \quad (41)$$
$$f_3 = 0 \quad (42)$$

It can be observed that energy efficiency is greatly improved by employing network coding. For instance, in the case of unit channel gains, energy is reduced from 31 units to 15 units, at $\lambda_1 = 1.5$.

Next, we show the tradeoff between energy efficiency and other parameters of interest, e.g. queue length. The data arrival process at each source node is assumed to be Poisson. We also set $\lambda_1 = 5$ frames/slot and $\lambda_2 = 10$ frames/slot and each slot spans of 10 ms. The bandwidth is set to be 1kHz. All link gains are assumed to be unity and noise at each node is assumed to be have zero mean and unit variance.

In Fig. 3 and Fig. 4, it can be seen that the analytical result and simulation for both queue length at $S_1$ and frame delay at $S_1$ matches with each other perfectly. It is also observed that with increasing $\epsilon$, the queue length at $S_1$ and the associated frame delay decrease quickly. However, increasing $\epsilon$ results in an increased probability of idle slots as we are over-provisioning the system, and therefore lower bandwidth efficiency. The factor $\epsilon$ thus controls the trade-off between bandwidth efficiency and buffering delay.

VI. CONCLUSION

In this work, we characterized the energy efficiency of a two-way relay network employing digital network coding with stochastic data arrivals. The problem of minimizing total
energy consumption subject to queue stability constraints was formulated. After making some observations on necessary conditions to be energy optimal, we formulated an equivalent convex optimization problem and presented the solution to this problem. An explicit solution was also derived for small data arrival rates at the source nodes. We also introduced a scheduling protocol to implement the proposed resource allocation scheme, and studied queue length and delay performance through simulations under such a protocol. In the future, it would be interesting to study the case of multiple relays, delay constraints, and random energy availability at one or more nodes due to reliance on energy harvesting. Bandwidth efficiency, reflected in the probability of time slots going unused due to empty queues, will also be considered.

REFERENCES