This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2010 proceedings

Jointly Optimal Power and Resource Allocation for Orthogonal NDF Relay Systems with QoS Constraints

Rooholah Hasanizadeh and Timothy N. Davidson
Dept. Elec. & Comp. Eng., McMaster University,
Hamilton, ON, Canada

***Abstract*—One of the advantages of relaying is that it offers the potential for a reduction the power required to achieve a specified level of quality-of-service (QoS). However, the problem of optimizing the available resources so as to minimize this power is often difficult to solve. In this paper we consider the case of a point-to-point link assisted by an orthogonal non-regenerative decode-and-forward (NDF) relay that is allocated a fraction of the time block. We assign prices to the powers of the source and the relay, and we consider the problem of jointly optimizing the source power, the relay power and the fraction of the time block so as to minimize the total cost of the power required to achieve a specified target rate. The natural formulation of that problem is not convex, but by analyzing the structure of the constraints we obtain a quasi-closed-form expression for the optimal solution that, at most, requires the solution of a simple one-dimensional zero-crossing problem for a monotonic function. This enables the problem to be efficiently solved, and clearly identifies when relaying is superior to direct transmission. Our numerical results illustrate the extent of the gains over regenerative decode-and-forward relaying, in which the resource allocation is, by definition, constrained to be equal.**

***Index Terms*—half-duplex relaying, non-regenerative decode-and-forward, quality-of-service, resource allocation.**

I. Introduction

Conventional wireless communication networks are based on direct communication between nodes. However, collaboration between nodes, or with dedicated relay stations, offers the potential for substantial gains in a variety of important performance metrics; e.g., [1]. For example, it may be possible to increase the achievable data rate of the links or increase the coverage region of the network without requiring additional power from the source, and it may be possible to increase the spatial diversity of the link without using additional antennas. In response to this potential, a large number of collaboration schemes have been proposed; e.g., [1]. These range from fully cooperative systems incorporating code books that encode shared messages, to systems in which the nodes collaborate by simply relaying messages for each other.

In this paper, we will focus on a point-to-point link that is assisted by a half-duplex relay. The source and relay transmit in separate fractions of the available time block, and the orthogonality thus obtained greatly simplifies the decoding process at the destination. The chosen relaying strategy is non-regenerative decode-and-forward (NDF). This enables the flexibility to partition the time block arbitrarily, and the resulting resource allocation parameter represents an additional degree of freedom for the designer.

There are several approaches that can be taken to the design of such a system. A popular approach is that of rate maximization, in which the powers used by the source and the relay, and the channel resource allocation, are jointly optimized so as to maximize the achievable data rate from source to destination, subject to appropriate power constraints; e.g., [2]–[4]. In this paper, we will address a different problem, namely that of minimizing the cost of the power required to achieve a specified quality-of-service (QoS). In particular, we (jointly) optimize the power and resource allocation so as to minimize the cost of the power required to achieve reliable communication from the source to the destination at a specified rate. In this paper, that design will be performed for the case of channels with long coherence times and systems in which perfect channel state information is available, but the simplicity of the obtained solution offers the potential for further development.

In recent work, we addressed the above QoS problem for the case of regenerative decode-and-forward (RDF) relaying [6]. In that case, the time block must be partitioned equally, and the remaining power allocation problem is convex and can be analytically solved. In the case of NDF relaying, the partitioning is an extra degree of design freedom, but the design problem is not convex. However, by analyzing the structure of the problem we obtain a quasi-closed-form expression for the optimal solution that requires, at most, the solution of a simple one-dimensional convex optimization problem. Our numerical example will illustrate the extent of the gain that can be obtained by taking advantage of the extra degree of design freedom.

II. System Model

We consider the simple relaying system illustrated in Fig. 1, in which the destination, D, receives signals directly from the source, S, and also receives a relayed signal from the relay, R. The system operates either in a direct transmission mode or in a relay-assisted mode. In the relay-assisted mode, orthogonal half-duplex signalling is employed, with orthogonality being

This work is supported in part by the Natural Sciences and Engineering Research Council of Canada and the Canada Research Chairs program.

978-1-4244-6404-3/10/$26.00 ©2010 IEEE

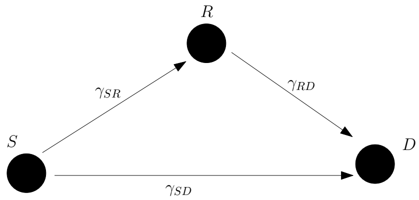

Fig. 1. A simple relay-assisted link.

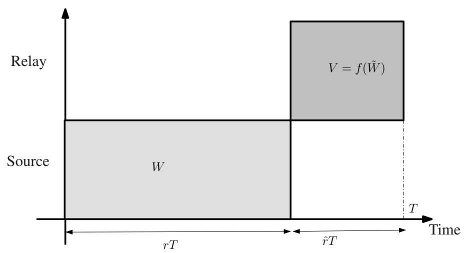

Fig. 2. Frame partitioning in relay-assisted transmission mode.

synthesized by time division, and with the relaying strategy being non-regenerative decode-and-forward (NDF). In particular, frames of length T seconds are partitioned into two time slots of length rT and $\hat{r}T$, respectively, where $0 \leq r \leq 1$ and $\hat{r} = 1 - r$. In the first time slot, the source transmits a codeword W and the relay and destination listen. The relay decodes the codeword and in the second time slot re-encodes the message using a different code book and transmits the codeword V to the destination. This mode of operation is illustrated in Fig. 2, which highlights the advantage of NDF relaying over regenerative decode-and-forward (RDF) relaying. In NDF relaying, the partitioning of the frame, r, becomes a design variable, whereas in RDF relaying the relay employs the same code book as the source and hence the fraction r is fixed at $1/2$. (As an aside, in amplify-and-forward relaying, r is also fixed at $1/2$.) In addition to the relay-assisted mode, we also consider a direct transmission mode, in which $r = 1$, which means that the relay is silent and the source communicates directly with the destination.

We will consider a scenario in which the source, relay and destination nodes each have a single antenna and we consider a narrowband channel model with long coherence times. As a result, the channel can be deemed to be flat in frequency and constant over a frame. Under this model, at each channel use, the received signal at each node takes the form

$$y = hs + n, \tag{1}$$

where the noise n is a sample from a white Gaussian process with zero mean and variance σ^2, and we have implicitly assumed perfect synchronization. As the capacity of a channel of the form in (1) depends on the ratio of the power gain to the receiver noise variance, for each of the links in Fig. 1 we define the effective channel gain

$$\gamma = \sqrt{|h|^2/\sigma^2}. \tag{2}$$

Since the coherence time is long, channel state information (CSI) can be obtained by the receivers at a comparatively low cost, and hence they can perform coherent detection. (In this paper, we will assume that this CSI is perfect) In the development of our design algorithms we will also assume that information for all three channels is available to the node at which the power and resource allocation takes place, and that this allocation can be perfectly communicated to the transmitters.

III. Problem Statement and Solution Strategy

From Fig. 2 it is clear that if the source employs a power level P_S while it is turned on, its average power over a frame is rP_S. Similarly, the average power consumed by the relay over a frame is $\hat{r}P_R$. In practice, the costs to the network of a unit of power from the source and relay are unlikely to be the same, and we will use λ to denote the relative cost of the relay power to that of the source. If the source is battery operated and the relay is connected to the power grid, then λ will be small. If the relay is another battery-operated node in the network that has a low battery level, then λ may be large.

Our design goal is to minimize the cost of the power required to enable reliable communication from source to destination a specified target rate R_t; i.e.,

$$\min_{P_S, P_R, r} rP_S + \lambda \hat{r} P_R \tag{3}$$

subject to the rate R_t being achievable, r lying between zero and one, and the powers being non-negative. We will measure the rate in terms of bit per (complex) channel use and hence all logarithms will be with respect to base 2.

Before we seek to optimize the power and resource allocation in the relay-assisted mode, we ought to assess the cost of direct transmission. Given the channel model in (1), and the presumption of the availability of ideal codes, the achievable rate of direct transmission is $\log(1 + \gamma_{SD}^2 P_S)$, and hence the power required to achieve the rate R_t is

$$P_{S,D} = (2^{R_t} - 1)/\gamma_{SD}^2. \tag{4}$$

Since direct transmission implicitly involves setting $r = 1$ and $P_R = 0$, this is also the cost of direct transmission.

For the relay-assisted mode, we note that for orthogonal NDF relaying with the source being allocated a fraction r of the time frame, the achievable rate is, e.g., [4], [5],

$$\min \big\{ r \log(1 + \gamma_{SR}^2 P_S), \\ r \log(1 + \gamma_{SD}^2 P_S) + \hat{r} \log(1 + \gamma_{RD}^2 P_R) \big\}. \tag{5}$$

Therefore, the problem of optimizing the source and relay powers, and the partitioning of the frame, so as to minimize the cost of the power required to achieve a rate R_t can be written as

$$\min_{P_S, P_R, r} \quad rP_S + \lambda \hat{r} P_R \tag{6a}$$

$$\text{subject to} \quad r \log(1 + \gamma_{SR}^2 P_S) \geq R_t, \tag{6b}$$

$$r \log(1 + \gamma_{SD}^2 P_S) + \hat{r} \log(1 + \gamma_{RD}^2 P_R) \geq R_t, \tag{6c}$$

$$P_S \geq 0, \quad P_R \geq 0, \quad 0 < r < 1. \tag{6d}$$

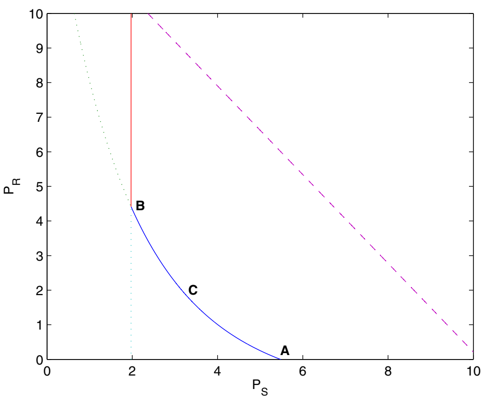

Fig. 3. The feasible region for the problem in (6) for a given value of r.

The lower inequality on r is strict because when $r = 0$ the source does not transmit, and the upper inequality is strict because the case of $r = 1$ corresponds to direct transmission and hence the requirement in (6b) that the relay be able to decode the message does not apply.

Our first observation is that the problem in (6) is not jointly convex in P_S, P_R and r. For instance, the objective is a sum of bilinear functions. As such, this problem may appear to be difficult to solve directly. We will overcome that difficulty by adopting a hierarchical approach in which we first consider the problem in (6) for a fixed value of r. There are three critical points of that problem and a closed-form expression is found for each one of them. One of these critical points will be shown to be inherently inferior to direct transmission, and hence we need only consider the other two. We then optimize these two critical points over r and determine the better of the resulting solutions. In both cases, the optimization problem over r is convex. Furthermore, our analysis reveals that one of these optimized points is subsumed by the combination of the other optimized point and direct transmission, and hence we obtain our quasi-closed-form expression for the optimal solution. Evaluating this expression requires, at most, the solution of a simple one-dimensional convex optimization problem.

IV. Optimal Allocation

As mentioned above, the first step in our derivation of the optimal allocation is to consider the problem of minimizing (6) over P_S and P_R for a given (fixed) value of r. The feasible set for that problem is illustrated in Fig. 3. It is to the right of vertical constraint imposed by (6b), and above the curved constraint imposed by (6c). The critical points for this problem are marked as A, B, and C. At point A, the relay power goes to zero, whereas at point B, both of the constraints in (6b) and (6c) are active. Point C is the point at which the slope of the curve defined by equality in (6c) is the same as that of the objective. Depending on the value of λ, the slope of the objective function (shown with the dashed line) will change and any one of the three points could be optimal.

A. Closed-form expression for critical points for given r

At point A, the relay power goes to zero and hence the source communicates directly with the destination in the allocated fraction r of the frame. From (6c), this requires

$$P_{S,A} = (2^{R_t/r} - 1)/\gamma_{SD}^2, \tag{7}$$

and hence the cost of solution is $rP_{S,A} = r\left(2^{R_t/r} - 1\right)/\gamma_{SD}^2$. However, this solution leaves the channel idle for a fraction $\hat{r} = (1-r)$ of the frame, and, as one might intuitively expect, this incurs a greater cost than direct transmission; cf (4). As a result, for any $r \in (0,1)$, point A is inferior to direct transmission and we need not consider this point any further.

Now, let us examine point B; the point at which the constraints in (6b) and (6c) intersect. Using (6b),

$$P_{S,B}(r) = (2^{R_t/r} - 1)/\gamma_{SR}^2, \tag{8a}$$

and hence, using (6c),

$$P_{R,B}(r) = \frac{1}{\gamma_{RD}^2}\left(\frac{2^{R_t/\hat{r}}}{\left(1 + \frac{\gamma_{SD}^2}{\gamma_{SR}^2}\left(2^{R_t/r} - 1\right)\right)^{r/\hat{r}}} - 1\right). \tag{8b}$$

We note that $P_{S,B}(r)$ is always non-negative, but that $P_{R,B}(r)$ is non-negative if and only if $\gamma_{SR} \geq \gamma_{SD}$. If $\gamma_{SR} < \gamma_{SD}$, then, as intuition would suggest, direct transmission is optimal.

Finally, let us examine point C. At this point, the slope of the curve defined by equality in (6c) is equal to the slope of the objective, $-r/(\hat{r}\lambda)$. Equality in constraint (6c) yields the curve $P_R = g_r(P_S)$, where

$$g_r(P_S) = \frac{1}{\gamma_{RD}^2}\left(\frac{2^{R_t/\hat{r}}}{(1 + \gamma_{SD}^2 P_S)^{r/\hat{r}}} - 1\right). \tag{9}$$

Ignoring, for the moment, (6b) and (6d), solving

$$\frac{dg_r(P_S)}{dP_S} = -\frac{r}{\hat{r}\lambda},$$

yields the following closed-form expression for critical point C

$$P_{S,C}(r) = \frac{1}{\gamma_{SD}^2}\left(2^{R_t}\left(\lambda\frac{\gamma_{SD}^2}{\gamma_{RD}^2}\right)^{\hat{r}} - 1\right), \tag{10a}$$

$$P_{R,C}(r) = \frac{1}{\gamma_{RD}^2}\left(\frac{2^{R_t}}{\left(\lambda\frac{\gamma_{SD}^2}{\gamma_{RD}^2}\right)^{r}} - 1\right). \tag{10b}$$

We will examine the feasibility of this point in the next section.

B. Optimization of critical points over r

As we have already argued, critical point A is always inferior to direct transmission, and need not be considered.

Now let us consider the case in which $\gamma_{SR} > \gamma_{SD}$ so that the intersection of (6b) and (6c) yields a feasible solution. The objective value at this critical point (point B), is $f_B(r) = rP_{S,B}(r) + \hat{r}\lambda P_{R,B}(r)$. To simplify the expression for $f_B(r)$ we define the constants $u = 2^{R_t}$, $v = \lambda\gamma_{SD}^2/\gamma_{RD}^2$,

$p = 1/\gamma_{SD}^2$, and $q = \gamma_{SD}^2/\gamma_{SR}^2$. By using the expressions in (8), the objective can be written as

$$f_B(r) = pqru^{1/r} - rp(q-v) - pv + \frac{pvu^{1/\hat{r}} - rpvu^{1/\hat{r}}}{\left(1-q+qu^{1/r}\right)^{r/\hat{r}}}. \quad (11)$$

The optimization problem that remains is

$$\min_{r\in(0,1)} \quad f_B(r). \quad (12)$$

By analytically evaluating $d^2 f_B(r)/dr^2$ it can be shown that $f_B(r)$ is convex in r, and hence the problem in (12) can be efficiently solved using a simple one-dimensional search over $r \in (0,1)$. For example, one can obtain an analytic expression for $df_B(r)/dr$ and search for its zero-crossing on $(0,1)$.

For point C, the objective function becomes $f_C(r) = rP_{S,C}(r) + \hat{r}\lambda P_{R,C}(r)$, with $P_{S,C}(r)$ and $P_{R,C}(r)$ given by (10). Therefore, $f_C(r)$ can be written as

$$f_C(r) = upv^{\hat{r}} - rp(1-v) - pv. \quad (13)$$

To find the optimal r for the critical point C, we must solve

$$\min_{r} \quad f_C(r) \quad (14a)$$
$$\text{subject to} \quad r\log(1+\gamma_{SR}^2 P_S) \geq R_t, \quad (14b)$$
$$0 < r < 1, \quad P_{S,C}(r) \geq 0, \quad P_{R,C}(r) \geq 0. \quad (14c)$$

As shown in the Appendix, the optimal solution of (14) lies on the boundary of the feasible set, with either (14b) being active, or the constraint $P_{R,C}(r) \geq 0$ being active. The latter case corresponds to a solution at point A, and hence is inferior to direct transmission. The former case results in both (6b) and (6c) being active, and hence corresponds to a solution at point B. Therefore, the optimization of point C over r is subsumed by that of points A and B.

C. The optimal solution

The above analysis shows that there are two candidate solutions for problem in (3), namely: direct transmission with $(r, P_S, P_R) = (1, P_{S,D}, 0)$, where $P_{S,D}$ is given by (4); and relay-assisted transmission with $(r, P_S, P_R) = (r_B^\star, P_{S,B}(r_B^\star), P_{R,B}(r_B^\star))$ corresponding to the optimized point B, where $r_B^\star$ is the solution to (12), and $P_{S,B}(r)$ and $P_{R,B}(r)$ were given in (8). As discussed after (8), direct transmission is optimal if $\lambda \geq \gamma_{RD}^2/\gamma_{SD}^2$ or $\gamma_{SD} \geq \gamma_{SR}$. When that is not the case, we compute the optimal relay-assisted solution and compare its cost to that of direct transmission.

We are now ready to provide a formal statement of the optimal solution to the problem in (3). In Table I we state the solution in the form of an algorithm, as this enables us to avoid computing the solution of (12) when that is not necessary.

TABLE I
OPTIMAL SOLUTION TO THE PROBLEM IN (3).

Given the desired rate, R_t, the effective channel gains, γ_{SD}, γ_{SR} and γ_{RD}, and the relative cost of the relay power, λ, the optimal solution to (3) can be obtained in the following way:

if $\lambda \geq \gamma_{RD}^2/\gamma_{SD}^2$ or $\gamma_{SD} \geq \gamma_{SR}$ **then**
  Perform Direct Transmission with $(r, P_S, P_R) = (1, P_{S,D}, 0)$
else
  Solve (12) for $r_B^\star$
  if $f_B(r_B^\star) \geq P_{S,D}$ **then**
    Perform Direct Transmission with $(r, P_S, P_R) = (1, P_{S,D}, 0)$
  else
    Perform Relay-Assisted Transmission with $(r, P_S, P_R) = \left(r_B^\star, P_{S,B}(r_B^\star), P_{R,B}(r_B^\star)\right)$
  end if
end if

V. Numerical Example

In this section, we present a simple numerical example in which we compare the cost of the power required to achieve reliable communication at a specified rate using NDF relaying, to the cost incurred by RDF relaying. In the NDF scheme, the partitioning of the channel resource, r, can be optimized, whereas in the RDF scheme the resource partition is fixed to $1/2$. (Due to the nature of the NDF scheme, it incurs a lower cost than the RDF scheme even when the optimal value for r is $1/2$.) We consider an environment in which each link is modeled as an independent Rician block-fading channel with additive white circular complex Gaussian noise of unit variance. The Rician 'K-factor' was chosen to be $K = 5$, and the average power gains of each link were chosen to be $\alpha^2 K/(K+1)$, with $\alpha_{SR} = 1$, $\alpha_{RD} = 0.5$ and $\alpha_{SD} = 0.25$. The target rate was set to 1 bit per (complex) channel use.

In Fig. 4 we plot the average value of the cost, $rP_S + \lambda\hat{r}P_R$, of the power required to achieve the target rate against the relative cost of the relay power, λ. In Figs 5 and 6 we provide the corresponding plots for the source and relay powers. (The averages were computed over 10^6 channel realizations.) For each relaying scheme and each channel realization we compare the cost of relaying to that of direct transmission and choose the approach with the lower cost. Therefore, the considered schemes might be better described as "NDF or direct" and "RDF or direct", respectively. Indeed, this is one of the reasons why the costs of the schemes converge as λ increases; as the relative cost of the relay power increases, direct transmission will be chosen more often. The figures demonstrate that when the relay power is cheap, the cost of the "NDF or direct" scheme is about 1.3 dB lower than that of the "RDF or direct" scheme, and the power required by the source is about 3.6 dB lower. However, as the cost of the relay power increases, the gain of the "NDF or direct" scheme decreases. Since the implementation of NDF relaying is significantly more involved than that of RDF relaying, these observations help to quantify the trade-offs involved in the selection of the relaying protocol for QoS-based service on a relay-assisted link.

VI. Conclusions

In this paper, we have considered the problem of minimizing the cost of the power required to achieve a specified target rate on a point-to-point link that is assisted by an orthogonal non-regenerative decode-and-forward relay. We assumed that perfect channel state information was available, and we ob-

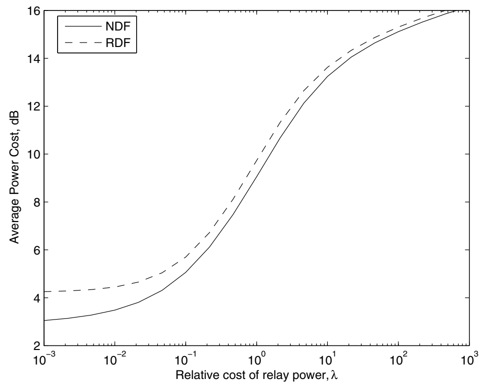

Fig. 4. Average cost of the power, on a decibel scale, against the relative cost of relaying, for the "NDF or direct" and "RDF or direct" schemes. For reference, the average cost of direct transmission is 21.7 dB.

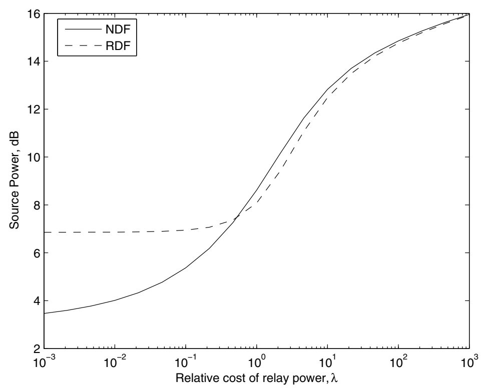

Fig. 5. Average value of the source power, P_S, against the relative cost of relaying, for the "NDF or direct" and "RDF or direct" schemes. For reference, the average power for direct transmission is 21.7 dB.

tained a quasi-closed form expression for the jointly optimal power and channel resource allocation. Given the simplicity of the obtained expression for the optimal solution, there are many potential extensions and generalizations of this work. A simple extension is to include peak power constraints on the source and relay (in addition to the objective of minimizing the cost of the average power). The solution of the resulting optimal allocation problem takes a similar form to that in Section IV-C, but with some additional conditions. Other, more substantial, generalizations include that to compress-and-forward relaying, the incorporation of statistical models for uncertainty in the channel state information, the development of power and resource allocation policies for fading channels with partial channel state information, and the extension to two-way relaying systems.

APPENDIX

The problem in (14) is one dimensional, and the critical points are those at which a constraint is active and any feasible point for which the derivative of $f_C(r)$ is zero. One can show, analytically, that on $r \in (0, 1)$, $d^2 f_C(r)/dr^2 > 0$ and hence

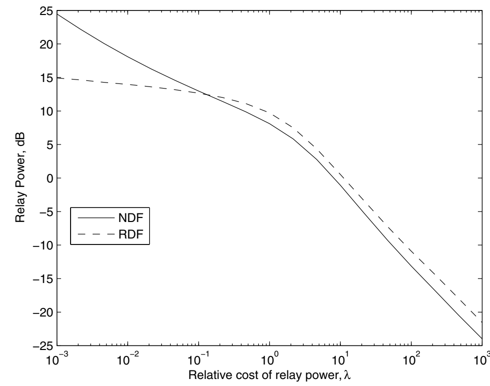

Fig. 6. Average value of the relay power, P_R, against the relative cost of relaying, for the "NDF or direct" and "RDF or direct" schemes.

$f_C(r)$ is strictly convex and has at most one stationary point. From (13) we have that $df_C(r)/dr = -up \ln v \times v^{\hat{r}} - p(1-v)$, and hence the stationary point is

$$\tilde{r}_C = \Big(R_t + \log\big(\frac{v \ln v}{v-1}\big)\Big)/\log v. \quad (15)$$

Now, let us examine whether this point is feasible. First consider the case in which $\lambda > \gamma_{SD}^2/\gamma_{RD}^2$, in which case $v > 1$. Inspection of (15) and (10b) yields a sequence of implications which shows that $P_{R,C}(\tilde{r}_C) < 0$, and hence that $r = \tilde{r}_C$ does not generate a feasible point when $\lambda > \gamma_{SD}^2/\gamma_{RD}^2$. Now, if $\lambda < \gamma_{SD}^2/\gamma_{RD}^2$, then $v < 1$, and in this case inspection of (15) and (10a) yields a different sequence of implications which shows that $P_{S,C}(\tilde{r}_C) < 0$, and hence that $\tilde{r}_C$ does not generate a feasible solution in this case, either. Therefore, regardless of the choice of $\lambda \neq \gamma_{RD}^2/\gamma_{SD}^2$, the solution generated by (15) lies outside the feasible region of (6). For $\lambda = \gamma_{RD}^2/\gamma_{SD}^2$, any value of $r \in (0, 1)$ is optimal and the optimal values for the powers are found to be $P_{S,C} = (2^{R_t} - 1)/\gamma_{SD}^2$, and $P_{R,C} = (2^{R_t} - 1)/\gamma_{RD}^2$, and hence $f_C(r) = (2^{R_t} - 1)/\gamma_{SD}^2 = P_{S,D}$. This is the same objective value as that achieved by direct transmission.

REFERENCES

[1] G. Kramer, I. Maric, and R. D. Yates, "Cooperative communications," *Foundations and Trends in Networking*, vol. 1, no. 3-4, pp 271-425, 2006.
[2] A. Høst-Madsen and J. Zhang, "Capacity bounds and power allocation for the wireless relay channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2020–2040, June 2005.
[3] Y. Liang and V. V. Veeravalli, "Gaussian orthogonal relay channels: Optimal resource allocation and capacity," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3284–3289, Sep. 2005.
[4] W. Mesbah and T. N. Davidson, "Power and resource allocation for orthogonal multiple access relay systems," *EURASIP J. Adv. Signal Processing*, vol. 2008, Article ID 476125, 15 pages, 2008.
[5] S. Serbetli and A. Yener, "Relay assisted F/TDMA ad hoc networks: Node classification, power allocation and relaying strategies," *IEEE Trans. Commun.*, vol. 56, no. 6, pp. 937–947, June 2008.
[6] F. Ahmed and T. N. Davidson, "Optimal power allocation for orthogonal relay systems with quality of service constraints," in *Proc. Asilomar Conf. Signals, Systems, Computers*, Pacific Grove, CA, Nov. 2009.