MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Relay Selection and Data Transmission Throughput Tradeoff in Cooperative Systems

Virag Shah, Neelesh Mehta, Raymond Yim

TR2009-087 December 2009

Abstract

A common and practical paradigm in cooperative communications is the use of a dynamically selected 'best' relay to decode and forward information from a source to a destination. Such a system consists of two core phases: relay selection phase, in which the system expends resources to select the best relay, and a data transmission phase, in which it uses the selected relay to forward data to the destination. In this paper, we study and optimize the trade-off between the selection and data transmission phase durations. We derive closed-form expressions for the overall throughput of a non-adaptive system that includes the selection phase overhead, and then optimize the selection and data transmission phase durations. Corresponding results are also derived for an adaptive system in which the relays can vary their transmission rates. Our results show that the optimal selection phase overhead can be significant even for fast selection algorithms. Furthermore, the optimal selection phase duration depends on the number of relays and whether adaptation is used.

Global Telecommunications Conference, 2009

Relay Selection and Data Transmission Throughput Tradeoff in Cooperative Systems

Virag Shah
Dept. of Electrical Communication Eng.
Indian Institute of Science (IISc)
Bangalore, India
Email: virag4u@gmail.com

Neelesh B. Mehta
Dept. of Electrical Communication Eng.
Indian Institute of Science (IISc)
Bangalore, India
Email: nbmehta@ece.iisc.ernet.in

Raymond Yim
Mitsubishi Electric Research Labs.
Cambridge, MA, USA
Email: yim@merl.com

***Abstract*—A common and practical paradigm in cooperative communications is the use of a dynamically selected 'best' relay to decode and forward information from a source to a destination. Such a system consists of two core phases: a relay selection phase, in which the system expends resources to select the best relay, and a data transmission phase, in which it uses the selected relay to forward data to the destination. In this paper, we study and optimize the trade-off between the selection and data transmission phase durations. We derive closed-form expressions for the overall throughput of a non-adaptive system that includes the selection phase overhead, and then optimize the selection and data transmission phase durations. Corresponding results are also derived for an adaptive system in which the relays can vary their transmission rates. Our results show that the optimal selection phase overhead can be significant even for fast selection algorithms. Furthermore, the optimal selection phase duration depends on the number of relays and whether adaptation is used.**

I. Introduction

Relay-based multi-hop cooperation, in which a source node transfers information to the destination with the help of a relay selected from the available nodes has attracted considerable attention in the literature [1]–[4]. The relay is selected depending on its instantaneous channel gains on the basis of a real-valued locally known metric that is a function of the relay-destination (RD) channel gain or the source-relay (SR) channel gain or both, depending on the cooperation protocol. Relay selection has been shown to help the system exploit the spatial diversity afforded by having geographically spaced multiple relays. It improves the symbol error probability [3] or increases the data transmission rate [5]. In general, the extent and nature of the benefits from selection depends on both the selection criteria and the cooperation scheme [6]–[11].

After the source broadcasts its data, these systems typically use two phases to complete the transmission to the destination: (i) a *relay selection phase*, in which the 'best' relay with the highest metric is chosen by a selection mechanism, and (ii) a *data transmission phase*, in which data is transmitted to the destination by the selected relay. The selection phase is needed because the source does not know a priori which relay is the best one. Furthermore, since the metric is a function of local channel gains, each relay knows only its metric, and not that of the others.

The work of the first two authors was partially supported by the Boeing-ANRC research project PC36101.

In most papers, the selection is assumed to be perfect and instantaneous. In effect, it is assumed to incur no time or energy overhead. However, in practice, the system does need to expend time and energy during the selection phase to find the best relay. The simulations in [12], [13], which modeled several practical aspects of a contention-based selection process, indicate that the relative time and energy spent in the relay selection phase can be considerable. This overhead clearly depends on the selection mechanism. For example, in a centralized polling mechanism, the overhead for selection increases linearly with the number of available relays. In [12], a source uses overhead handshaking messages to exhaustively track the rate that each candidate relay can support. The selection phase overhead can be reduced by using distributed mechanisms based on back-off timers [14] or time-slotted splitting algorithms [15]–[17].

The selection phase cannot always be perfect. For example, a practical system may terminate the selection phase after some time even when the best relay has not been selected. This leads to an outage during the subsequent data transmission phase. While increasing the selection phase duration reduces this outage probability, it does so at the expense of the overall throughput due to the less time available for data transmission. Thus, the two phases affect each other, and cannot be optimized in isolation.

This paper conducts a comprehensive system-level analysis and optimization that considers the trade-offs between the two phases. Such a modeling and joint optimization has received limited attention in the literature [11]–[13], [18]. For example, [18] considers outage and throughput, but not the selection phase overhead. We consider a generic system-level model that explicitly models the two phases and develop analytical expressions for the overall system throughput. We first analyze a non-adaptive system in which the data rates and time intervals for the phases are fixed, and then an adaptive system in which the selected relay adapts its transmission rate to improve overall throughput. The optimal trade-off between the two phases will turn out to be quite different for these two systems.

To be able to make precise system-level quantitative statements, a choice needs to be made about the mechanism used for selection. To this end, we consider the splitting algorithm for relay selection [15], [17] given its remarkable speed and scalability. While the time taken by the splitting algorithm to select the best node depends on the specific realizations of their metrics, on average it can find the best node within 2.47 slots even when an asymptotically large number of nodes contend. Even for such an efficient algorithm, the time overhead of the selection phase often turns out to be considerable. Our analysis shows that the optimal selection phase duration is often at least a factor of two or more than the above average in order to ensure a sufficiently large probability of success in selecting the best relay. Thus, our analysis shows that a joint design is also desirable in systems that use other selection mechanisms.

The paper is organized as follows. The system model is set up in Sec. II. The analysis is developed in Sec. III. The design implications are brought out in Sec. IV, and are followed by our conclusions in Sec. V.

II. System Model

Figure 1 shows a schematic of the cooperative relay network that we consider. It contains one source node, one destination node, and n decode-and-forward relays. The channels from the source to the relays as well as from the relays to the destination are assumed to be frequency flat channels that undergo independent fading. Thus, for relay i, the source to relay (SR) channel power gain, h_{si}, and the relay to destination (RD) channel power gain, h_{id}, are independent and identical exponentially distributed random variables with the same mean.

We focus on the identically distributed case in this paper due to space constraints. The analysis can be generalized to handle the case of non-identical SR and RD channels [19]. We assume, without loss of generality (w.l.o.g.), the fading powers means are all normalized to 1. We assume a block fading channel model in which the channel gains remain constant over the selection and data transmission phases (described below). In practice, the system can operate over time-varying channels and thus have to live with partially (though not fully) outdated metrics after selection. Modeling the time-varying nature of the channels and its impact on the overall system performance, and then optimizing the selection mechanism's parameters is an interesting avenue for future work. The noise at each receiving node is additive white Gaussian with zero mean and unit variance. For analytical tractability, we shall assume that the direct source to destination link is weak, as is typically the case [6], [20].[1]

We now describe the cooperation protocol for the baseline non-adaptive system. The adaptive system is discussed next.

1) *SR Data Transmission Phase*: The source broadcasts at a fading-averaged SNR of ρ_s for T_d time units and all the relays listen. The number of bits per transmission per unit time is B. To relate B, h_{si}, and ρ_s, we use the ideal Shannon capacity formula, and assume that a relay i can decode the source's transmission only if

$$B \leq WT_d \log_2(1+h_{si}\rho_s), \textit{i.e.}, h_{si} > \frac{2^{\frac{B}{WT_d}}-1}{\rho_s} \triangleq \gamma_{sr}, \quad (1)$$

where W is bandwidth. Practical coding inefficiencies can also be easily incorporated along the lines of [22], [23].

2) *Relay Selection Phase*: The relays contend for a duration of T_c slots, each of duration t_{slot}. The selection criterion and algorithm is explained below in Sections II-A and II-B.

3) *RD Data Transmission Phase*: At the end of the relay selection phase, the selected relay, if any, transmits data to the destination with a fading-averaged SNR of ρ_r for T_d time units (using the same modulation and coding as the source). The destination decodes the message if $h_{jd} > \left(2^{\frac{B}{WT_d}}-1\right)/\rho_r \triangleq \gamma_{rd}$, where j is the selected relay.

[1]The reader is referred to [21] for an analysis that includes the S-D link. The papers discusses why doing so makes the analysis more involved.

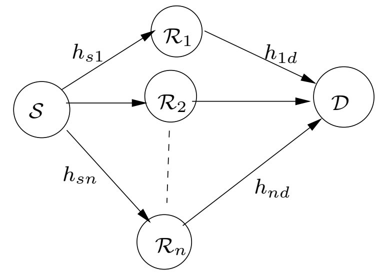

Fig. 1. A cooperative system consisting of a source (S), a destination (D), and n relays, from which the best relay is selected chosen.

We say an *outage* occurs if the destination cannot receive source's message successfully. All time durations such as T_d – and consequently B itself – are normalized w.l.o.g. with respect to a contention slot's duration t_{slot}.

A. Relay Selection Criterion

The goal of the selection algorithm is to select the relay with the highest RD channel gain [20]. Each relay knows its RD channel gain but not that of others. Furthermore, only those relays that decode the source's message and have an RD channel gain large enough to support transmission to the destination participate in the selection process. *Other relays set their metrics to* 0 *and do not contend*, as this wastes energy without improving throughput. Formally, the local metric, μ_i, based on which a relay i contends is

$$\mu_i = \begin{cases} 0, & h_{si} < \gamma_{sr} \text{ or } h_{id} < \gamma_{rd} \\ h_{id}, & \text{otherwise} \end{cases}. \quad (2)$$

The complementary cumulative distribution function (CCDF) of the metric μ_i is denoted by $F_c(\mu) = \Pr(\mu_i \geq \mu)$.

B. Relay Selection Algorithm

The splitting-based selection algorithm only requires that the CCDF and n is known to all relays. It proceeds as follows [15], [17]. At the beginning of a slot k, each relay locally computes two thresholds $H_L(k)$ and $H_H(k)$, and transmits if its metric μ satisfies $H_L(k) < \mu < H_H(k)$. At the end of each slot, the source broadcasts one of three outcomes to all the relays: (i) *idle* when no relay transmitted, (ii) *success* when exactly one relay transmitted, or (iii) *collision* when multiple relays transmitted and collided.[2] The relays update their thresholds for the next slot accordingly.

Formally, the algorithm is specified as follows. Let $H_{\min}(k)$ denote the largest value of the metric known up to slot k above which the best metric surely lies. And, let split $(a, b) \triangleq F_c^{-1}\left(\frac{F_c(a)+F_c(b)}{2}\right)$. This *split* function ensures that, on average, only half the relays involved in the last collision transmit in the next slot. It can be easily shown that

$$F_c(\mu) = \begin{cases} e^{-(\gamma_{sr}+\mu)}, & \mu \geq \gamma_{rd} \\ e^{-(\gamma_{sr}+\gamma_{rd})}, & 0 < \mu < \gamma_{rd} \\ 1, & \mu = 0 \end{cases}. \quad (3)$$

In the first slot, the variables are initialized as follows: $H_L(1) = \max(\gamma_{rd}, F_c^{-1}(1/n))$, $H_H(1) = \infty$, and $H_{\min}(1) = 0$. In the $(k+1)^{\text{th}}$ slot, the variables are updated based on the outcome as follows:

1) If feedback (of the k^{th} slot) is an idle and no collisions have occurred thus far, then $H_H(k+1) = H_L(k)$, $H_{\min}(k+1) = 0$, and $H_L(k+1) = \max(\gamma_{rd}, F_c^{-1}(\frac{i+1}{n}))$.
2) If feedback is a collision, then $H_H(k+1) = H_H(k)$, $H_{\min}(k+1) = H_L(k)$, and $H_L(k+1) =$ split $(H_L(k), H_H(k))$.
3) If feedback is an idle and a collision has occurred in the past, then $H_H(k+1) = H_L(k)$, $H_{\min}(k+1) = H_{\min}(k)$, and $H_L(k+1) =$ split $(H_{\min}(k), H_H(k))$.

The reader is referred to [15], [17] for a detailed explanation of the above steps. We shall call the durations of the algorithm before and after the first non-idle slot as the *idle* and *collision* phases, respectively. During idle phase, the algorithm ensures that only one node, on average, transmits in each slot. Once collision happens, the nodes are *split* in subsequent slots until success happens.

The splitting algorithm is fast because it ensures that one node, on average, transmits in each of the idle slots, regarless of the number of nodes in the system. Furthermore, in the event of a collision, it is very likely that only two nodes have collided. These are then separated quickly by successively splitting the interval.

Comments:

- Note that the formulation above differs slightly from the original algorithm proposed in [15]. During idle phase, we introduce the $\max(.)$ function to prevent relays with metrics below γ_{rd} from transmitting. This is done, since transmission from such relays only results in wastage of energy without increasing throughput as they have either not decoded the source's message or do not have a strong enough RD channel to forward the message to the destination. One implication of this is that no relay will get selected after $\lceil nF_c(\gamma_{rd}) \rceil$ initial idle slots.
- The slot duration of the splitting algorithm depends on the transmission protocol. Each slot needs to allow for two transmissions – one by the nodes and the other by the sink – and necessary gaps, as required, between these two transmissions. For example, in 802.11 systems, each slot's duration can easily exceed 100 μsec [24].

[2]The sink can determine these outcomes based, for example, on the strength of the received power, which is measurable by many receivers today [16].

C. Adaptive System

Since the relay knows its channel gain to the destination, it can reduce its transmission duration by adapting its transmit rate to $\log_2(1 + \rho_r h_{id})$. The relay's transmit SNR, ρ_r, is kept the same as for the non-adaptive system. In addition, the system wastes no time in the RD phase in case the selection phase results in an outage.

It must be noted that other forms of adaptation are certainly possible if the system design allows it. For example, the relay can adjust its transmit power instead of rate to minimize energy consumed. Furthermore, in some systems, even the selection phase can terminated as soon as a success is fed back by the source or when it becomes obvious (after $\lceil nF_c(\gamma_{rd}) \rceil$ idle slots) that no success is possible. These adaptations and others are not analyzed in this paper due to space constraints, and are addressed in detail in [19].

III. Analysis

We first analyze the throughput of the non-adaptive system and then that of the adaptive system.

A. Non-Adaptive System

The destination can successfully decode only when all the following three conditions are satisfied: (i) There is at least one relay that has decoded the message from the source, (ii) Among the relays that have decoded the source message, at least one of them has a good enough link to the destination, and (iii) The selection phase can select the best relay within T_c slots. In our set up, the above conditions together are equivalent to the condition that the selection phase terminates *successfully* within T_c slots.

The throughput, η, for the non-adaptive system is therefore

$$\eta = \frac{BP_s(T_c)}{2T_d + T_c}, \quad (4)$$

where $P_s(T_c)$ is the probability that the selection phase terminates successfully. The denominator in (4) is the total time, including selection phase, taken by the source and relays to transmit B bits (normalized with respect to t_{slot}).

To determine $P_s(T_c)$, we first state an intermediate result about the selection algorithm's behavior.

Lemma *1:* The probability, $p(a, b)$, that exactly b slots are required to resolve a collision involving a relays is

$$p(a,b) = \frac{1}{2^a}\left(p(a, b-1) + \sum_{i=2}^{a} \binom{a}{i} p(i, b-1)\right), \ \forall\ a, b > 1, \quad (5)$$

where $p(a,1) = a/2^a$, $\forall\ a > 1$, and $p(1,b) = 0$.

Proof: The proof is given in Appendix A. ■

The main result on the probability of success now follows.

Theorem *1:* The probability of successful data transfer is

$$P_s(T_c) = \sum_{i=1}^{\min(T_c,r)} \left(1 - \frac{i}{n}\right)^{n-1} + I_{\{T_c>r\}} n\left(\alpha - \frac{r}{n}\right)(1-\alpha)^{n-1} + \sum_{i=1}^{\min(T_c-1,r)} \sum_{k=2}^{n} \binom{n}{k} \left(\frac{1}{n}\right)^k \left(1 - \frac{i}{n}\right)^{n-k} \sum_{j=1}^{T_c-i} p(k,j) + I_{\{T_c>r+1\}} \sum_{k=2}^{n} \binom{n}{k} \left(\alpha - \frac{r}{n}\right)^k (1-\alpha)^{n-k} \sum_{j=1}^{T_c-r-1} p(k,j), \tag{6}$$

where $I_{\{x\}}$ is an indicator function that equals 1 if condition x is true and is 0 otherwise, $\alpha = F_c(\gamma_{rd})$, and $r = \lceil n\alpha \rceil - 1$.

Proof: The proof is given in Appendix B. ■

The above result can be understood as follows. Its first two terms correspond to the first non-idle slot being a success. The third and the fourth term correspond to the first non-idle slot (i^{th} slot) being a collision among $k > 1$ relays that is resolved in the remaining $T_c - i$ slots.

B. Adaptive System

In this system, a relay can now adapt its transmission rate. Equivalently, it adapts its transmission duration – subject to a cap of $T_{\max}$, which is a system parameter. A relay contends only if it can transmit the information to the destination within $T_{\max}$, *i.e.*, it contends when its RD channel gain exceeds the threshold $\gamma_{rd}^{\text{adp}} \triangleq \frac{2^{\frac{B}{WT_{\max}}}-1}{\rho_r}$.

The new throughput (normalized with respect to t_{slot}) then equals

$$\eta = \frac{BP_s(T_c)}{T_d + T_c + T_{\text{adp}}}, \tag{7}$$

where T_{adp} is the average RD transmission phase duration. The probability of successful selection, $P_s(T_c)$, above is again given by (6) with γ_{rd} simply replaced by γ_{rd}^{adp}.[3] Finally, the average RD transmission phase duration, T_{adp}, is then given by the following result.

Theorem *2:* When the RD data transmission duration is adapted, its average equals

$$T_{\text{adp}} = \sum_{i=1}^{\min(T_c,r)} \left(1 - \frac{i}{n}\right)^{n-1} T\left(\frac{i-1}{n}, \frac{i}{n}; 1\right) + \sum_{i=1}^{\min(T_c-1,r)} \sum_{k=2}^{n} \binom{n}{k} \left(\frac{1}{n}\right)^k \left(1 - \frac{i}{n}\right)^{n-k} T\left(\frac{i-1}{n}, \frac{i}{n}; k\right) \sum_{j=1}^{T_c-i} p(k,j) + I_{\{T_c>r\}} \left(\alpha - \frac{r}{n}\right)(1-\alpha)^{n-1} T\left(\frac{r}{n}, \alpha; 1\right) + I_{\{T_c>r+1\}} \sum_{k=2}^{n} \binom{n}{k} \times \left(\alpha - \frac{r}{n}\right)^k (1-\alpha)^{n-k} T\left(\frac{r}{n}, \alpha; k\right) \sum_{j=1}^{T_c-r-1} p(k,j), \tag{8}$$

[3]The threshold γ_{sr} for the SR channel gain is the same as for the non-adaptive case since the source behavior is not modified.

where, $T(x,y;k)$, $0 \le x < y$, is the average time required for data transmission when the relay with the highest metric is selected from among k relays all of whose metrics lie in the interval $(F_c^{-1}(y), F_c^{-1}(x))$. It is given by

$$T(x,y;k) = \frac{B}{W} \frac{ke^{-k\gamma_{sr}}}{(y-x)^k} \sum_{j=0}^{k-1} (-1)^j \binom{k-1}{j} \log(2) \times \left(ye^{\gamma_{sr}}\right)^{k-1-j} \left(\psi_1\left(\frac{\rho_r}{j+1}, \log(1+\rho_r F_c^{-1}(y))\right) - \psi_1\left(\frac{\rho_r}{j+1}, \log(1+\rho_r F_c^{-1}(x))\right)\right), \tag{9}$$

and $\psi_1(a,u) = \int_u^\infty \frac{1}{t} \exp\left(t + \frac{1-e^t}{a}\right) dt$.

Proof: The proof is given in Appendix C. ■

We specify the intervals limits in $T(.,.;k)$ in terms of $F_c^{-1}(.)$ as it helps simplify the notation. This is valid as $F_c^{-1}(.)$ is a one-to-one monotonic mapping for continuous distributions.

IV. Results and System Design Implications

We study the system-level tradeoffs using both the analytical results derived in Sec. III and Monte Carlo simulations with 10^5 samples. The parameter values chosen are: $\rho_s = \rho_r = 6$ dB, and $B = WT_d \log_2(1 + 10^{0.6})$, which implies that a relay or destination can decode successfully if its instantaneous received SNR is at least 6 dB.

For the non-adaptive system, Figure 2 shows that the probability of successful selection, $P_s(T_c)$, increases as the total number of available relays (n), or the selection phase duration (T_c) increases. For large T_c, $P_s(T_c)$ saturates at $1 - (1-\alpha)^n$. This is because $(1-\alpha)^n$ is the probability that an outage occurs because none of the n available relays contend in the selection phase. Notice also that the analytical and simulation results match very well.

While increasing T_c improves the probability of successful selection, it also increases the overall time of the three phases. This important trade-off between the overall system throughput and T_c is shown in Figure 3. For small T_c, increasing T_c improves throughput since the probability of outage decreases. However, for larger T_c, the throughput starts decreasing as $P_s(T_c)$ saturates. The results show that the optimal value for the selection phase duration depends on both n and T_d. For example, for $T_d = 13$, $T_c = 4$ and 5 are optimal for $n = 6$ and 15, respectively. Whereas, for $T_d = 100$, the optimal values of T_c increase to 7 and 8 for $n = 6$ and 15, respectively. Furthermore, the throughput increases as the number of relays increases or as the data transmission duration increases. The latter occurs because the relative overhead of the selection phase decreases. Also, the optimal T_c that maximizes the throughput increases with number of nodes. This is because the system can afford to spend a little more time to find the best node and benefit from increased diversity.

The impact of T_c and the number of relays on the throughput is shown in Figure 4 for the adaptive system. Compared to the non-adaptive case, the throughput is greater. For example, at

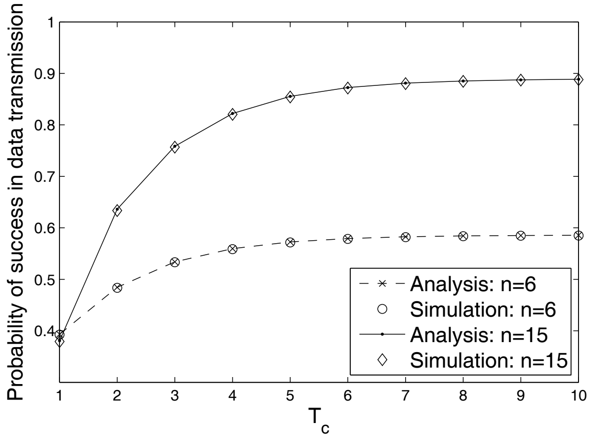

Fig. 2. Probability of successful data transfer as a function of selection phase duration (T_c) and number of relays (n) for the non-adaptive system.

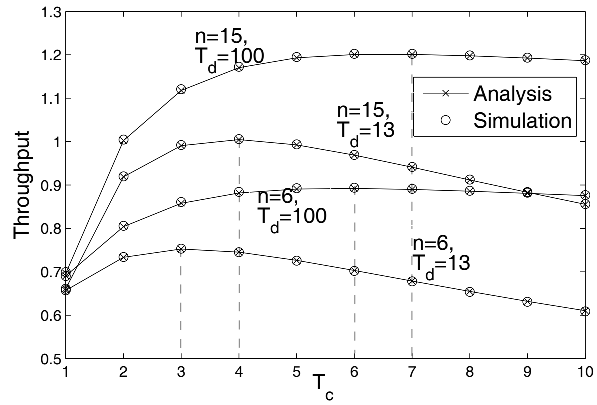

Fig. 4. Adaptive system: Trade-off between overall throughput (normalized with respect to t_{slot} and W) and selection phase duration for different numbers of relays (n) and different source data transmission durations (T_d) for $T_{\max} = T_d$.

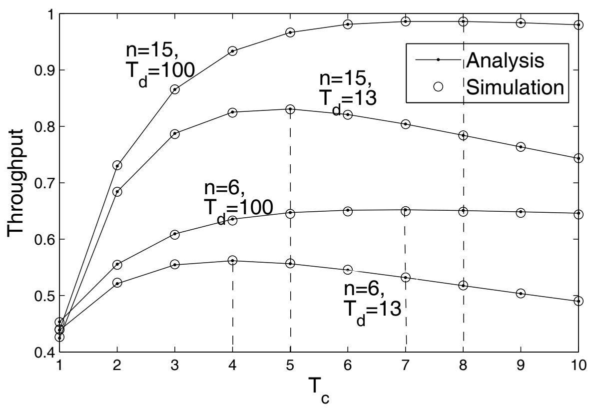

Fig. 3. Non-adaptive system: Trade-off between overall throughput (normalized with respect to t_{slot} and W) and selection phase duration for different numbers of relays (n) and for different source data transmission durations (T_d).

$n = 15$ and $T_d = 100$, the maximum throughput of 1.202 of the adaptive system is 22% more than its non-adaptive counterpart. The optimal selection duration shrinks for the adaptive case. For example, at $T_d = 13$, $T_c = 3$ and 4 are throughput optimal for $n = 6$ and 15, respectively. As in the non-adaptive system, the optimum selection duration increases as T_d increases. As Figures 3 and 4 suggest, the throughput is a concave function of T_c. Thus, fast convex optimization techniques can be used to determine optimum T_c. The optimal T_c typically lies between 2 and 9 (slots), with the exact value depending on T_d.

V. Conclusions

We analyzed the system-level interactions and trade-off between the relay selection and data transmission phases of a cooperative relay system. To this end, we developed analytical expressions for the probability of successful selection and the average system throughput. We saw that even with a fast splitting-based selection algorithm, the relative overhead of the selection phase was non-negligible. In general, the relative overhead of the selection phase decreased as the source transmission duration increased. Also, the overhead increased as the number of available relays increased. We also saw that the optimal system parameter settings for the adaptive and non-adaptive systems are different. The optimum selection phase duration value was lower for the rate adaptive system. The results in this paper show that the time devoted to the selection phase must be carefully chosen in order to maximize the overall system throughput. Future work includes optimizing the energy-efficiency of the systems as well.

Appendix

A. Proof of Lemma 1

The probability that i relays, among the a that collided, transmit in the next slot equals $\binom{a}{i}/2^a$. Thus $p(a,b) = a/2^a$ for $b = 1$. When $b > 1$, the following three cases arise: (i) The next slot is idle: The probability that the collision among a relays is resolved in $b-1$ slots is $p(a, b-1)$; (ii) i $(i > 1)$ relays collide in the next slot: The probability that it is resolved in exactly $b-1$ remaining slots is $p(i, b-1)$; (iii) The next slot is a success: Since the collision is already resolved, the probability that it is resolved in exactly b slots is 0.

B. Proof of Theorem 1

For $i \leq r$, the probability that the first non-idle slot is the i^{th} slot and $k \geq 1$ relays transmit in it is $\binom{n}{k}\left(\frac{1}{n}\right)^k\left(1-\frac{i}{n}\right)^{n-k}$.

The $i = (r+1)$ case is slightly different. From (2), relays whose RD channel gains are below a_r set their metrics to be 0. This occurs with probability $1-\alpha$, where $\alpha = F_c(a_r)$. Now, the probability that the first non-idle slot is $(r+1)^{\text{th}}$ slot with k relays involved in it is the probability that k relays have their metric between $F_c^{-1}\left(\frac{r}{n}\right)$ and a_r, and the remaining $n-k$ relays have metric 0. Therefore, this probability equals $\binom{n}{k}\left(\alpha - \frac{r}{n}\right)^k (1-\alpha)^{n-k}$. No non-idle slot can occur after the $(r+1)^{\text{th}}$ slot. If k relays are involved in the collision in the

i^{th} slot, the probability that the collision is resolved in the remaining $T_c - i$ slots is $\sum_{j=1}^{T_c-i} p(k,j)$.

Depending on T_c and r, the following two cases occur:

1) $T_c \leq r$: Successful data transfer occurs if in the first non-idle slot (i^{th} slot): (i) a success occurs, which happens with probability $\binom{n}{1}\left(\frac{1}{n}\right)\left(1-\frac{i}{n}\right)$, or (ii) $k \geq 2$ relays collide, which happens with probability $\binom{n}{k}\left(\frac{1}{n}\right)^k\left(1-\frac{i}{n}\right)^{n-k}$, and are resolved in $T_c - i$ slots. Thus,

$$P_s(T_c) = \sum_{i=1}^{T_c}\left(1-\frac{i}{n}\right)^{n-1} + \sum_{i=1}^{T_c-1}\sum_{k=2}^{n}\binom{n}{k}\left(\frac{1}{n}\right)^k\left(1-\frac{i}{n}\right)^{n-k}\sum_{j=1}^{T_c-i}p(k,j).$$

2) $T_c > r$: Successful data transfer occurs if during the first non-idle slot ($i \leq r+1$): (i) a success occurs, or (ii) if $k \geq 2$ relays collide and this is resolved in $T_c - i$ slots. If a collision occurs in the $(r+1)^{\text{th}}$ slot, the interval that needs to be split has a probability mass $\alpha - \frac{r}{n}$ since only relays with non-zero metrics contend. The expression for $P_s(T_c)$ becomes

$$P_s(T_c) = \sum_{i=1}^{r}\left(1-\frac{i}{n}\right)^{n-1} + n\left(\alpha-\frac{r}{n}\right)(1-\alpha)^{n-1} + \sum_{i=1}^{r}\sum_{k=2}^{n}\binom{n}{k}\left(\frac{1}{n}\right)^k\left(1-\frac{i}{n}\right)^{n-k}\sum_{j=1}^{T_c-i}p(k,j) + \sum_{k=2}^{n}\binom{n}{k}\left(\alpha-\frac{r}{n}\right)^k(1-\alpha)^{n-k}\sum_{j=1}^{T_c-r-1}p(k,j).$$

Equation (6) compactly expresses the above two results using the indicator function $I_{\{.\}}$.

C. Proof of Theorem 2

First, we derive the expression for $T(x,y;k)$. Given that a metric lies in $(F_c^{-1}(y), F_c^{-1}(x))$, its probability distribution function (PDF) is $f_1(t) = \frac{e^{-\gamma_{sr}}e^{-t}}{y-x}$ and its cumulative distribution function is $F_1(t) = \frac{e^{-\gamma_{sr}}}{y-x}(e^{\gamma_{sr}}y - e^{-t})$. Using order statistics [25], the average total transmit time required when k relays have their metrics in the above interval and the one with the highest metric is selected is

$$\begin{aligned} T(x,y;k) &= \frac{B}{W}\int_{F_c^{-1}(y)}^{F_c^{-1}(x)} \frac{1}{\log_2(1+\rho_r t)} k f_1(t) F_1(t)^{k-1} dt, \\ &= \frac{B}{W}\int_{F_c^{-1}(y)}^{F_c^{-1}(x)} \frac{k e^{-k\gamma_{sr}} e^{-t}(e^{\gamma_{sr}}y - e^{-t})^{k-1}}{\log_2(1+\rho_r t)(y-x)^k} dt. \end{aligned}$$

Expanding the integrand as a binomial series and further simplifying yields (9).

Given that the first non-idle slot is the i^{th} slot and k relays are involved in it, the average time required is $T\left(\frac{i-1}{n}, \frac{i}{n}; k\right)$, if $i \leq r$, and $T\left(\frac{r}{n}, \alpha; k\right)$, if $i = r+1$. The probability of this event can be derived from the proof of Theorem 1. Combining the above results leads to the desired expression.

References

[1] A. Nosratinia, T. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Commun. Mag.*, vol. 42, pp. 68–73, 2004.

[2] D. S. Michalopoulos and G. K. Karagiannidis, "Performance analysis of single relay selection in Rayleigh fading," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 3718–3724, Oct. 2008.

[3] Y. Zhao, R. Adve, and T. J. Lim, "Symbol error rate of selection amplify-and-forward relay systems," *IEEE Commun. Lett.*, vol. 10, pp. 757–759, Nov. 2006.

[4] Y. Jing and H. Jafarkhani, "Single and multiple relay selection schemes and their achievable diversity orders," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 1414–1423, Mar. 2009.

[5] Y. Zhao, R. Adve, and T. J. Lim, "Improving amplify-and-forward relay networks: optimal power allocation versus selection," in *Proc. IEEE Int. Symp. Inform. Theory*, pp. 1234–1238, 2006.

[6] R. Madan, N. B. Mehta, A. F. Molisch, and J. Zhang, "Energy-efficient cooperative relaying over fading channels with simple relay selection," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 3013–3025, Aug. 2008.

[7] J. Luo, R. S. Blum, L. J. Cimini, L. J. Greenstein, and A. M. Haimovich, "Link-failure probabilities for practical cooperative relay networks," in *Proc. VTC (Spring)*, pp. 1489–1493, 2005.

[8] Z. Lin and E. Erkip, "Relay search algorithms for coded cooperative systems," in *Proc. Globecom*, 2005.

[9] V. Sreng, H. Yanikomeroglu, and D. Falconer, "Relayer selection strategies in cellular networks with peer-to-peer relaying," in *Proc. VTC (Fall)*, pp. 1949–1953, 2003.

[10] T. C.-Y. Ng and W. Yu, "Joint optimization of relay strategies and resource allocations in cooperative cellular networks," *IEEE J. Select. Areas Commun.*, vol. 25, pp. 328–339, Feb. 2007.

[11] J. Ai, A. A. Abouzeid, and Z. Ye, "Cross-layer optimal policies for spatial diversity relaying in mobile ad hoc networks," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 2930–2939, Aug. 2008.

[12] P. Liu, Z. Tao, Z. Lin, E. Erkip, and S. Panwar, "Cooperative wireless communications: a cross-layer approach," *IEEE Trans. Wireless Commun.*, vol. 13, pp. 84–92, Aug. 2006.

[13] Z. Zhou, S. Zhou, J.-H. Cui, and S. Cui, "Energy-efficient cooperative communication based on power control and selective single-relay in wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 3066–3078, Aug. 2008.

[14] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," *IEEE J. Select. Areas Commun.*, vol. 24, pp. 659–672, Mar. 2006.

[15] X. Qin and R. Berry, "Opportunistic splitting algorithms for wireless networks," in *Proc. INFOCOM*, pp. 1662–1672, Mar. 2004.

[16] R. Yim, N. B. Mehta, and A. F. Molisch, "Fast multiple access selection through variable power transmission," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 1962–1973, Apr. 2009.

[17] V. Shah, N. B. Mehta, and R. Yim, "Analysis, insights and generalization of a fast decentralized relay selection mechanism," in *Proc. ICC*, 2009.

[18] C. K. Lo, J. R. W. Heath, and S. Vishwanath, "The impact of channel feedback on opportunistic relay selection for hybrid-ARQ in wireless networks," *IEEE Trans. Veh. Technol.*, vol. 58, pp. 1255–1268, Mar. 2009.

[19] V. Shah, N. B. Mehta, and R. Yim, "The relay selection and transmission tradeoff in cooperative communication systems," *Submitted to IEEE Trans. Wireless Commun.*, 2009.

[20] A. Bletsas, H. Shin, and M. Win, "Cooperative communications with outage-optimal opportunistic relaying," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 3450–3460, Sept. 2008.

[21] E. Beres and R. Adve, "Selection cooperation in mutli-source cooperative networks," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 118–127, Jan. 2008.

[22] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 2349–2360, 2005.

[23] T. Starr, J. M. Cioffi, and P. J. Silverman, *Understanding Digital Subscriber Line Technology*. Prentice Hall PTR, 1 ed., 1999.

[24] "Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications," Tech. Rep. IEEE Std 802.11-2007, IEEE Computer Society, June 2007.

[25] H. A. David and H. N. Nagaraja, *Order Statistics*. Wiley Series in Probability and Statistics, 3 ed., 2003.