Performance Evaluation of Multiple-Relay Cooperative ARQ Strategies for Mobile Networks

Juan J. Alcaraz and Joan García-Haro

Department of Information Technologies and Communications, Technical University of Cartagena
Plaza Hospital, 1, 30202. Cartagena, Spain
juan.alcaraz@upct.es, joang.haro@upct.es

***Abstract*—In Cooperative Automatic Repeat reQuest (C-ARQ) protocols, one or more nodes can act as relays, collaborating in the frame retransmission process between a sender and a destination node. In the framework of a broadband mobile access network, we consider a relay enhanced cell where the sender represents the base station, and an undefined number of relays can cooperate with a single user (destination). In this paper we present a Markov model complemented with a reward model to analyze the throughput performance and the efficiency in the bandwidth utilization at the base station, highlighting a clear tradeoff between them. Both performance metrics are balanced by means of a multi-objective optimization algorithm resulting in a retransmission strategy for the access node that notably improves the bandwidth efficiency while maintaining the throughput very close to its maximum. The benefits of the proposed approach are evaluated in two scenarios. In the first one, the relays are part of the network infrastructure. In the second one, the relays are cooperative users.**

I. Introduction

The use of cooperative relays in future mobile access networks is considered a promising approach to enable the deployment of 4G technology [1] [2] [3]. Additionally, in cooperative diversity environments, Cooperative Automatic Repeat reQuest (C-ARQ) protocols are receiving increasing attention. In this paper we study a C-ARQ system in a time-slotted radio interface, where signals from different nodes do not collide, the frame sending process follows a stop and wait operation and ACKs arrive without errors. This scenario is similar to previous works [4] [5] [6]. We investigate the performance in two different scenarios: first, a relay enhanced cell [7] where the relays are part of the network infrastructure. Second, a situation where some users join into a collaborative group and form a cluster.

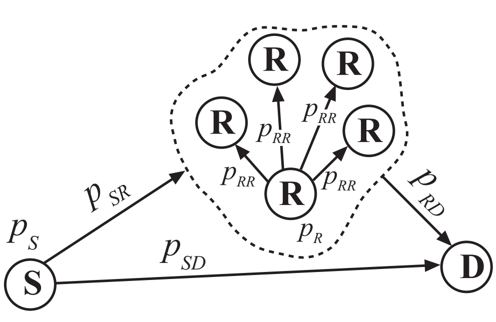

Fig. 1. Schematic diagram of the system analyzed

Our model includes the concept of cooperation group [4], which consists of a certain number, N, of relays (R) assisting one communication link between a sender (S) and a destination (R), as depicted in Fig. 1. This figure shows the probability of incorrect frame reception, or frame error ratio (FER), in the *direct link* (p_{SD}), in the *relay links* (p_{SR}) and in the *access links* (p_{RD}). The innovative aspects of our model are: First, we consider that retransmissions from a relay can be overheard by other relays, which contributes to increase the number of relays with a correct copy of the frame in subsequent time-slots. The FER among relays is p_{RR}. Second, each relay with a copy of the frame retransmits it with certain probability (p_R). Third, the sender also retransmits with certain probability, p_S.

The probabilistic retransmission strategy is the base of our proposal. This paper shows that, by an accurate setting of p_S, considering the FER of the links and the number of relays, it is possible to increase the availability of bandwidth resources at the base station, while achieving a throughput very close to the deterministic strategy ($p_S = 1$). Thereby, in a time-slot not assigned to retransmission, the radio resources allocated to the link between the sender node and the receiver node are released. The sender can re-allocate these resources temporarily to other links, introducing new data in the network, resulting in a more efficient use of the available bandwidth. In many existing and future radio access networks this is possible because resource allocation is done in a slot-by-slot basis. Two examples of this are the High Speed Downlink Packet Access (HSDPA) [8] and the *WINNER* 4G concept [7]

Probabilistic retransmission has been previously considered in a recent work [9] as a strategy to balance cooperation and collision probability in order to achieve smaller latencies, while our work is focused in collision-free networks, which is the most usual case for cellular access environments. Other works like [4], [5] and [6] consider a deterministic retransmission scheme at the source node. In contrast to these previous works, we address the issue of increasing the bandwidth efficiency by reducing the amount of retransmissions performed at the base station, showing the tradeoff between this goal and the user throughput, in a single-destination scenario. We develop a Markov-reward model of the system and combine it with a multi-objective optimization strategy to balance both performance metrics. Numerical results obtained in very different scenarios show that a remarkable retransmission rate reduction can be obtained with an almost negligible decay of the user

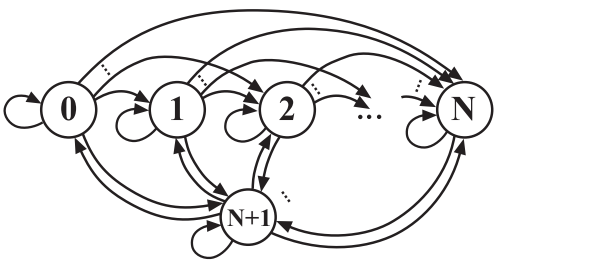

Fig. 2. Discrete Time Markov Chain modeling the operation of the system

throughput, in a single-destination scenario. When considering multiple destinations, the transmission opportunities released by the retransmission rate reduction at the base station can be assigned to the transmission of new frames, increasing the overall throughput of the cell, which is the ultimate goal of our proposal.

The rest of the paper is organized as follows. The Markov model of the system is summarized in Section II. This model is used in Section III to compute the performance metrics of the system. Section IV presents the strategy to balance throughput and retransmission rate at the base station. Numerical results are discussed in Section V. Finally, Section VI outlines the conclusions of this work.

II. System Model

The analysis relies on a Discrete Time Markov Chain (DTMC), in which the state of the system is denoted by the number of relays that have correctly decoded the frame. An additional state is added, which represents the reception of the frame at D. Therefore, the state space of a system with N relays is $\Omega = \{0, 1, \ldots, N, N+1\}$, where $N+1$ is the state representing a correct frame decoding at D. Because this event implies a new frame transmission from S, there may be a transition from state $N+1$ to any other state. However, because it is assumed that relays do not discard their copy of the frame during the retransmission process, the transitions among other states ($i \to j$) can only occur if $i \leq j$.

A diagram of the DTMC model for the system is shown in Fig. 2. In order to compute the transition probabilities, the following events are defined: frame reception at D (D), frame retransmission from S (S), frame retransmission from any relay (L), and their corresponding complementary events ($\overline{D}, \overline{S}, \overline{L}$). In addition, let A_n denote the reception of the frame at n additional relays, i.e. at n relays that have not received the frame in previous time-slots. Transition probabilities can be defined in terms of these events and their computation imply probabilistic and combinatorial analysis. Let us consider first the probability of a generic transition $i \to j$, where $N+1 > j > i > 0$. This probability is given by:

$$\begin{aligned} p_{i,j} &= P\{S \wedge \overline{L} \wedge A_{j-i} \wedge \overline{D}\} \\ &+ P\{S \wedge L \wedge A_{j-i} \wedge \overline{D}\} + P\{\overline{S} \wedge L \wedge A_{j-i} \wedge \overline{D}\} \end{aligned} \quad (1)$$

where the first term is obtained from the probabilities of the system, resulting in:

$$\begin{aligned} &P\{S \wedge \overline{L} \wedge A_{j-i} \wedge \overline{D}\} = \\ &\binom{N-i}{j-i} p_S p_{SD} (1-p_{SR})^{j-i} p_{SR}^{N-j} (1-p_R)^i \end{aligned} \quad (2)$$

the second term is given by:

$$\begin{aligned} &P\{S \wedge L \wedge A_{j-i} \wedge \overline{D}\} = \binom{N-i}{j-i} \cdot \\ &\sum_{k=1}^{i} \binom{i}{k} p_S (1-p_R)^{i-k} p_R^k \left(p_{SR} p_{RR}^k\right)^{N-j} \left(p_{SD} p_{RD}^k\right) \cdot \\ &\left((1-p_{SR}) + p_{SR} \sum_{m=1}^{k} \binom{k}{m} p_{RR}^{k-m} (1-p_{RR})^m\right)^{j-i} \end{aligned} \quad (3)$$

and the third term is obtained from (3) by replacing the term p_S with $(1-p_S)$ and the term p_{SR} with 1:

$$\begin{aligned} &P\{\overline{S} \wedge L \wedge A_{j-i} \wedge \overline{D}\} = \\ &\sum_{k=1}^{i} \binom{i}{k} (1-p_S)(1-p_R)^{i-k} p_R^k p_{RR}^{k(N-j)} p_{RD}^k \cdot \\ &\binom{N-i}{j-i} \left(\sum_{m=1}^{k} \binom{k}{m} p_{RR}^{k-m} (1-p_{RR})^m\right)^{j-i} . \end{aligned} \quad (4)$$

Considering the probability of the transition $i \to i$, where $N+1 > i > 0$, it can be checked that it is given by:

$$\begin{aligned} p_{i,i} &= \left(P\{\overline{S}\} + P\{S \wedge A_0 \wedge \overline{D}\}\right) \cdot \\ &\left(P\{\overline{L}\} + P\{L \wedge A_0 \wedge \overline{D}\}\right) . \end{aligned} \quad (5)$$

The terms in (5) are computed from the probabilities of the system in the following way: $P\{\overline{S}\} = 1 - p_S$, $P\{\overline{L}\} = (1-p_R)^i$, $P\{S \wedge A_0 \wedge \overline{D}\} = p_S p_{SR}^{N-i} p_{SD}$, and

$$P\{L \wedge A_0 \wedge \overline{D}\} = \sum_{k=1}^{i} \binom{i}{k} (1-p_R)^{i-k} p_R^k p_{RR}^{k(N-i)} p_{RD}^k .$$

The transition probability $p_{i,N+1}$, with $0 < i < N+1$ is expressed as:

$$p_{i,N+1} = P\{S \wedge D\} + \left(P\{\overline{S}\} + P\{S \wedge \overline{D}\}\right) P\{L \wedge D\}$$

where the computation of $P\{S \wedge D\}$, $P\{\overline{S}\}$ and $P\{S \wedge \overline{D}\}$ is straightforward, and the last term is given by:

$$\begin{aligned} P\{L \wedge D\} &= \sum_{k=1}^{i} \binom{i}{k} (1-p_R)^{i-k} p_R^k \cdot \\ &\sum_{m=1}^{k} \binom{k}{m} p_{RD}^{k-m} (1-p_{RD})^m . \end{aligned} \quad (6)$$

The transition probabilities from state 0 are obtained with the following expressions:

$$\begin{aligned} p_{0,0} &= p_S p_{SD} p_{SR}^N + (1-p_S) \\ p_{0,N+1} &= p_S (1-p_{SD}) \\ p_{0,j} &= \binom{N}{j} p_S (1-p_{SR})^j p_{SR}^{N-j} p_{SD} \end{aligned} \quad (7)$$

where $j < N+1$. Finally, transition probabilities from state $N+1$ are: $p_{N+1,0} = p_{SR}^N p_{SD}$, $p_{N+1,N+1} = 1 - p_{SD}$ and, when the next state is j, such that $0 < j < N+1$, the transition probability is:

$$p_{N+1,j} = \binom{N}{j}(1-p_{SR})^j p_{SR}^{(N-j)} p_{SD}. \quad (8)$$

Keeping in mind that $p_{i,j} = 0$ if $N+1 > i > j$ (see Fig. 2), we obtain the following DTMC transition matrix:

$$\mathbf{P} = \begin{pmatrix} p_{0,0} & p_{0,1} & \cdots & p_{0,N} & p_{0,N+1} \\ 0 & p_{1,1} & \cdots & p_{1,N} & p_{1,N+1} \\ 0 & 0 & \cdots & p_{2,N} & p_{2,N+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & p_{N,N} & p_{N,N+1} \\ p_{N+1,0} & p_{N+1,1} & \cdots & p_{N+1,N} & p_{N+1,N+1} \end{pmatrix} \quad (9)$$

III. Performance Analysis

Let $\overline{\pi} = \{\pi_0, \pi_1, \ldots, \pi_{N+1}\}$ be the steady-state distribution of the DTMC, where π_i is the steady-state probability of state $i \in \Omega$. $\overline{\pi}$ is obtained by solving the following system of linear equations:

$$\overline{\pi} = \overline{\pi}\mathbf{P}$$
$$\sum_{i\in\Omega} \pi_i = 1 \quad (10)$$

where $\mathbf{P}$ is given by (9). Solving (10) for $\overline{\pi}$ we obtain the following solution:

$$\pi_i = \alpha_i \pi_{N+1} \quad (11)$$

where, obviously $\alpha_{N+1} = 1$. The value of α_i is given by the following recursive equation:

$$\alpha_i = \frac{1}{1-p_{i,i}}\left(p_{N+1,i} + \sum_{k=0}^{i-1} p_{k,i}\alpha_k\right) \quad (12)$$

and π_{N+1}, by the normalization condition, is given by:

$$\pi_{N+1} = \left(\sum_{i=0}^{N+1} \alpha_i\right)^{-1}. \quad (13)$$

The throughput is defined as the average number of frames successfully received in the destination node per time-slot. According to the model, the throughput is the average number of time-slots that the DTMC spends in state $N+1$.

The retransmission rate is defined as the number of retransmissions from the source, normalized by the total number of time-slots. This measurement reflects the amount of resources allocated to retransmissions. In order to compute the retransmission rate, a reward model is constructed. This approach has been previously applied to the analysis of ARQ protocols. See [10] for a description of this technique.

Considering a generic transition $i \to j$, let R_{ij} denote the average number of retransmissions from S associated to this transition (associated reward). Analytically it is expressed as $R_{ij} = P\{S|i \to j\}$. The retransmission rate associated to a reference state, i, is computed with the following expression:

$$R_i = \sum_{j\in\Omega} p_{ij} R_{ij}. \quad (14)$$

Applying the law of total probability, the resulting rates are $R_i = p_S$ for $i < N+1$, and $R_{N+1} = 0$, because the transitions $N+1 \to j$ for $j \in \Omega$ always involve the transmission of a new frame, which obviously can not be considered as a retransmission. The average retransmission rate (R) at the sender is given by a weighted sum of the rates in (14), where the weighting factors are the steady state probabilities of the Markov chain:

$$R = \sum_{i\in\Omega} \pi_i R_i. \quad (15)$$

Because we focus on finding the optimal p_S, we use, for convenience, the following notation: The steady state distribution of the system is denoted by $\overline{\pi}(p_S)$. Then, the throughput is given by $T(p_S) = \pi_{N+1}(p_S)$. It is obvious that, for any given set of values $\{p_{SD}, p_{SR}, p_{RD}, p_N\}$, the maximum throughput is achieved when the source retransmits in every time-slot until an ACK reception ($p_S = 1$). Thus, we define the maximum throughput as $T_M = T(1)$. T_M is taken as a reference to evaluate different values of p_S. Similarly, we can express R in terms of p_S as $R(p_S) = p_S(1-\pi_{N+1}(p_S))$. Therefore, the retransmission rate associated with T_M is $R_M = R(1)$.

IV. Balancing Throughput and Bandwidth Efficiency

There exists a clear tradeoff between the two performance measurements derived in previous section. The maximum throughput, T_M, is achieved by setting $p_S = 1$. However, this can lead to a high retransmission rate, especially if the direct link is highly degraded. On the other hand, if $p_S = 0$, then the retransmission rate reduces to 0. In order to find an optimum balance of both objectives, we use an approach based on a multi-objective optimization technique, known as global criterion method [11]. This method consists of minimizing a global criterion function, defined as:

$$G(p_S) = \sum_{k=1}^{n} \alpha_k \left(\frac{O_k - f_k(p_S)}{O_k}\right)^2 \quad (16)$$

where n is the number of objective functions, $f_k(p_S)$ are the objective functions, O_k are the optimum values for each objective function and α_k are the factors weighting the relative importance assigned to each objective. In the system analyzed, we are balancing two objectives. First, it is desirable that the throughput approaches the maximum throughput as much as possible, i.e. $f_1(p_s) = T(p_s)$ and $O_1 = T_M$. Second, it is also desirable to reduce the retransmission rate of S, therefore $f_2(p_s) = R(p_s) + C$, and $O_2 = C$, where C is an auxiliary non-zero real number required to avoid a division by 0 in the global objective. In our model we choose $C = 1$ because the relative importance of $R(p_S)$ approaching to 0 is already controlled by α_2. Applying these definitions in (16) we obtain:

$$G(p_S) = \alpha_1 \left(\frac{T_M - T(p_S)}{T_M}\right)^2 + \alpha_2 R(p_S)^2. \quad (17)$$

Let p_S^* denote the solution to the multi-objective optimization problem. Because p_S^* is a probability, we must consider the constraint $0 \leq p_S \leq 1$. Therefore we have:

$$p_S^* = \arg\min_{0 \leq p_S \leq 1} \{G(p_S)\}. \quad (18)$$

Let $T_O = T(p_S^*)$ be the optimum throughput in terms of the multi-objective optimization problem. Similarly, let $R_O = R(p_S^*)$ denote the optimum retransmission rate. In the following section we compare T_O and R_O with T_M and R_M in several scenarios. Because our objective is to improve the bandwidth efficiency with the less possible impact on the throughput, the weighting factors chosen are $\alpha_1 = 1$ and $\alpha_2 = 0.2$. The following section shows that this configuration is adequate in the two scenarios under study and discusses the effects of modifying these factors.

V. Numerical Results

The numerical results in this section show the impact of the environmental parameters on the global performance of the system. We consider two scenarios. The first one represents a relay enhanced cell, where the relays are part of the network infrastructure. In the second one, the relays are cooperative users. In both scenarios, we compare the performance of the deterministic C-ARQ configuration ($p_S = 1$) with the optimized one ($p_S = p_S^*$).

A. Infrastructure Relays

Because in this scenario relays are part of the network infrastructure, it is realistic to assume that they have been deployed to have a relatively good *relay link* ($p_{SR} = 0.2$). Another realistic assumption is that these relays always cooperate in the retransmission process ($p_R = 1.0$) and that the user whose communication is assisted by the relay cooperation group has a degraded *direct link* ($p_{SD} = 0.5$) and a good *access link* ($p_{RD} = 0.1$). Otherwise, C-ARQ might not be especially helpful. It is also assumed that the links among the relays of a cooperation group have good propagation conditions ($p_{RR} = 0.1$).

Fig. 3 shows the performance metrics using the deterministic policy (T_M and R_M), and using the optimized policy (T_O and R_O) versus the number of relays. As expected, as the number of relays increases, the optimized algorithm assigns a smaller retransmission probability to the sender (see Fig. 4). It is remarkable that this strategy has an almost negligible impact on the throughput, given that T_O remains very close to T_M. It can be seen in Fig. 3 that, when N is larger than 6, the source retransmits with a probability very close to 0, which means that the sender virtually delegates all the retransmission process to the relay cooperation group.

B. Cooperative Users Acting as Relays

Because the relay cooperation group is formed by users that are close to each other, it is reasonable to assume that their distances to the base station are very similar. Thus, considering relatively degraded channel conditions in their links with the base station, we set $p_{SD} = p_{SR} = 0.4$. In contrast, the distance among them should be small, $p_{RD} = p_{RR} = 0.05$. Users are not necessarily devoted to cooperation, therefore we consider a

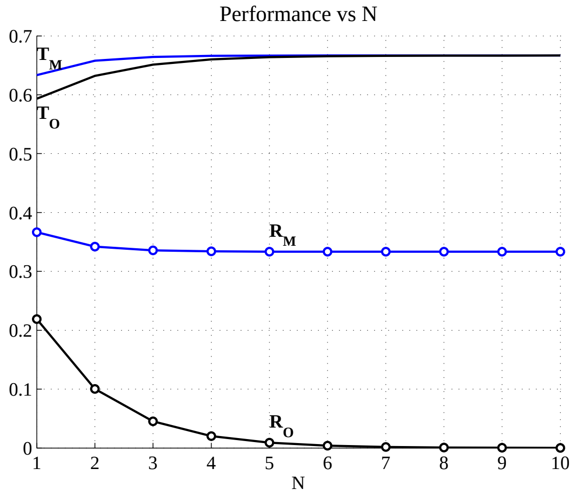

Fig. 3. Performance vs. N in a relay enhanced cell scenario.

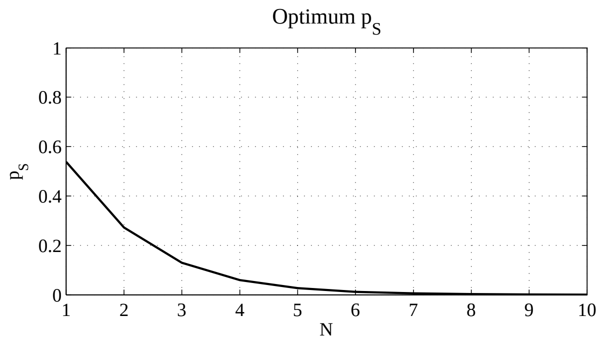

Fig. 4. p_S^* vs. N in a relay enhanced cell scenario.

much lower willingness to cooperate than in previous scenario ($p_R = 0.2$). Fig. 5 shows the performance figures versus the number of cooperative users. The value of p_S^* for the optimized strategy is shown in Fig. 6.

The optimum strategy is also useful to reduce the retransmission rate at the base station. However, in contrast to previous scenario, more relay nodes are required to obtain the maximum throughput with the optimized strategy. As the number of cooperative users increases, T_O gets closer to T_M and p_S decays (see Fig. 6), which obviously reduces R_O even more. It was checked that, as the value of the weighting factor α_2 in (17) is reduced, T_O gets closer to T_M in all values of N. At the same time, R_O also increases, which reveals that, in this scenario, retransmissions from S have more impact on the throughput.

VI. Conclusions

This paper analyzes the performance of a C-ARQ protocol comprising a cooperation group of relays. The system is modeled by means of a Markov-reward model for the computation of the throughput and the retransmission rate. There exists a clear tradeoff between these two metrics. We propose a multi-objective optimization strategy to compute a

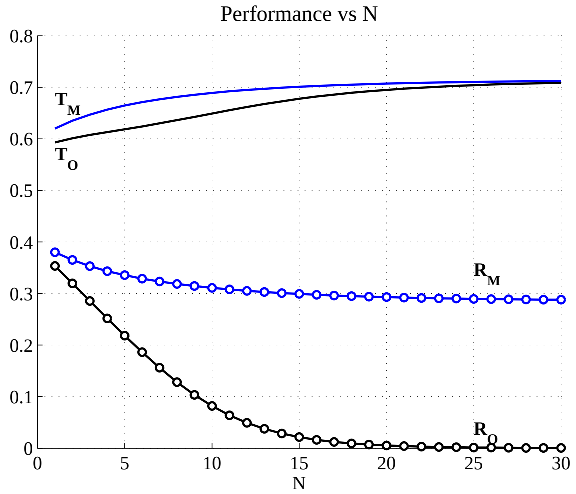

Fig. 5. Performance vs. N when the relays are cooperative users.

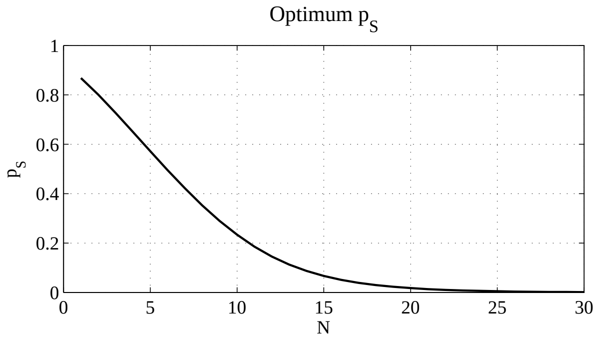

Fig. 6. p_S^* vs. N when the relays are cooperative users.

retransmission probability at the source that balances these two objectives. The system is evaluated in two different scenarios. The first one corresponds to a situations where relays are part of the network infrastructure (relay enhanced cell), and therefore relays are fully devoted to cooperation. The second one corresponds to situations where other users act as relays, forming cooperation groups when certain conditions hold, e.g., when they are close to each other (good relay link) and far from the base station (degraded direct and relay links). Users are considered to cooperate with a small probability. In both scenarios it was observed that the proposed strategy is able to reduce the retransmission rate at the base station with a small reduction of the throughput in the single-destination case. In a multiple-destination scenario, the retransmission rate reduction at the base station allows it to allocate more transmission opportunities to new frames, increasing the overall throughput.

The proposed strategy breaks with the classical and widely adopted policy of giving maximum priority to retransmissions in wireless access nodes. This idea can be applied to the design of scheduling mechanisms, considering the computed retransmission probability p_S^* as a lower bound for the amount of bandwidth reserved for retransmissions. The rest of the bandwidth can be assigned to other communication processes whenever they have available data.

Because the optimization algorithm runs in the sender node, S must be informed of the (average) propagation conditions of each channel, which is very usual in mobile networks. Moreover, it is especially important for the base station to know both the number of relays in the cooperation group and their willingness to cooperate. These two factors have shown a great influence in the value of p_S^* and in the overall performance of the system.

The research activity in relay-based broadband wireless networks is currently very active and in consequence many issues are still open. The ideas developed in this paper open new research lines that require further investigation to consolidate and generalize these initial findings. For example, what is the impact of frame discarding at the relay nodes (Truncated C-ARQ)? Is the protocol also suitable in radio interfaces where collisions can occur? Some related contributions in these scenarios can be found in [12] and [9].

Acknowledgment

This research has been supported by project grant TEC2007-67966-01/TCM (CON-PARTE-1) and it was also developed in the framework of "Programa de Ayudas a Grupos de Excelencia de la Región de Murcia, Fundación Séneca, Agencia de Ciencia y Tecnología de la RM (Plan Regional de Ciencia y Tecnología 2007/2010".

References

[1] A. Nosratinia, T. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Commun. Mag.*, vol. 42, pp. 74–80, Oct. 2004.

[2] D. Soldani and S. Dixit, "Wireless relays for broadband access [radio communications series]," *Communications Magazine, IEEE*, vol. 46, no. 3, pp. 58–66, March 2008.

[3] F. Fitzek and M. Katz, *Cooperation in Wireless Networks: Principles and Applications*. Springer, 2006.

[4] M. Dianati, X. Ling, K. Naik, , and X. Shen, "A node-cooperative arq scheme for wireless ad hoc networks," *IEEE Trans. Veh. Technol.*, vol. 55, pp. 1032–1044, May 2006.

[5] I. Cerutti, A. Fumagalli, and G. Ho, "Saturation throughput gain in fixed multiplexing radio networks with cooperative retransmission protocols," *Communications, 2006. ICC '06. IEEE International Conference on*, vol. 10, pp. 4489–4494, June 2006.

[6] I. Cerutti, A. Fumagalli, and P. Gupta, "Delay models of single-source single-relay cooperative arq protocols in slotted radio networks with poisson frame arrivals," *IEEE/ACM Trans. Netw.*, vol. 2, pp. 371–382, Apr. 2008.

[7] K. Doppler, A. Osseiran, M. Wodczak, and P. Rost, "On the integration of cooperative relaying into the winner system concept," *Mobile and Wireless Communications Summit, 2007. 16th IST*, pp. 1–5, 1-5 July 2007.

[8] A. T. H. Holma, *HSDPA/HSUPA for UMTS: High Speed Radio Access for Mobile Communications*. Wiley, 2006.

[9] L. Xiong, L. Libman, and G. Mao, "Optimal strategies for cooperative mac-layer retransmission in wireless networks," *Wireless Communications and Networking Conference, 2008. WCNC 2008. IEEE*, pp. 1495–1500, March 31 2008-April 3 2008.

[10] M. Zorzi and R. Rao, "On the use of renewal theory in the analysis of arq protocols," *IEEE Trans. Commun.*, vol. 44, pp. 1077–1081, Sep. 1996.

[11] S. Rao, *Engineering Optimization: Theory and Practice. Third Edition*. Wiley, 1996.

[12] I. Krikidis, "Distributed truncated arq protocol for cooperative diversity networks," *Communications, IET*, vol. 1, no. 6, pp. 1212–1217, Dec. 2007.