PROBABILISTIC RETRANSMISSION STRATEGY FOR SINGLE-RELAY COOPERATIVE ARQ

Juan J. Alcaraz, Joan García-Haro and Fernando Cerdán

Department of Information Technologies and Communication, Technical University of Cartagena
Plaza Hospital 1, Cartagena, Spain

Keywords: Automatic repeat request (ARQ), cooperative diversity, multi-objective optimization.

Abstract: In wireless networks with cooperative automatic repeat request (C-ARQ) protocols, a relay node, placed within the range of the sender node and the destination node, assists the sender in the process of frame retransmission. In a collision-free scenario, the sender and the relay use different physical channels for retransmissions. This paper highlights the tradeoff between throughput increment and efficiency in the use of radio resources. By using a probabilistic retransmission strategy, the sender only retransmits in some time-slots after a frame error notification. At the other time-slots, the source can assign the radio resources to other communication processes, resulting in more efficient use of the bandwidth. The retransmission probability must be carefully adjusted according to the network parameters. In this paper we propose a Markov model to compute the throughput performance and a complementary reward model to compute the retransmission rate of the source. In order to keep the throughput at a high value and to reduce the retransmission rate at the same time, we present a multi-objective optimization method that is capable of balancing both objectives in any scenario. It is shown that the increment in bandwidth efficiency can be very high, especially for degraded links, compensating the small throughput reduction associated with a probabilistic retransmission at the source.

1 INTRODUCTION

Cooperative automatic repeat request protocols, C-ARQ, are attracting increasing research attention. These type of protocols apply the concept of cooperative communication to improve the performance of link-layer protocols in wireless networks. In cooperative communications, each node not only transmits and receives data for its own application, but can also act as a relay node providing an alternative path for other pairs of communicating nodes (A. Nosratinia, 2004). This idea is known as cooperative diversity and is commonly presented as an extension of the spatial diversity concept of multiple-input multiple-output (MIMO), where each one of the multiple antennas are located at each cooperative node instead of in a single node. Future evolutions of mobile access networks are expected to make use of cooperative diversity using relay nodes to improve the connectivity of the terminal nodes, see (R. Pabst, 2004), (F. Fitzek, 2006) or (K. Doppler, 2007).

Classical (non-cooperative) ARQ protocols are widely used in wireless links to increase the reliability of the data frame delivery process between a sender node and a destination node. Several factors of the wireless channel, e.g. path loss, fading and noise, can degrade the quality of the received signal so that received frames can not be correctly decoded at the destination node. ARQ protocols specify how to retransmit data frames when frame losses are detected. C-ARQ exploits the broadcast nature of the wireless channel, involving additional nodes (relay nodes) in the retransmission process. A relay node is located within the transmission ranges of both the sender node and the destination node. If the relay node overhears a frame that the destination is unable to decode, it can assist the sender by retransmitting the same copy of the lost frame. Previous works (M. Dianati, 2006), (L. Xiong, 2008), (I. Cerutti, 2007) have shown that C-ARQ increases the probability of successful retransmission, resulting in higher throughput between the sender and the receiver nodes.

In general, C-ARQ protocols are evaluated in a slotted radio channel, where the destination node can receive simultaneously from the sender and the relay nodes. These nodes use different physical channels to communicate with the destination node, and
therefore signals do not collide. In CDMA-based networks, this requires the use of different scrambling codes; in OFDM networks, different orthogonal frequencies and, in TDMA, different time-slots. C-ARQ applied to these types of networks provides higher reliability at the cost of a more extensive use of wireless resources, assuming that both the sender and the relay node retransmit the lost frame.

The idea proposed in this paper is that, in certain circumstances, it may be beneficial that the sender does not retransmit in every time-slot after a frame loss notification. This way, in a time-slot not assigned to retransmission, the radio resources assigned to the link between the sender node and the receiver node are released. Therefore, the sender node can re-assign this resources temporarily to other links, introducing new data in the network and making a more efficient use of radio resources. In many existing and future radio access networks this is possible because resource allocation is done slot by slot. Two examples of this are the high speed downlink data access (HS-DPA) (H. Holma, 2006) and the WINNER 4G concept (K. Doppler, 2007).

There exists a clear tradeoff between retransmission probability and efficiency in bandwidth use. Our approach consists of adjusting the retransmission probability of the sender node in order to reduce its retransmission rate, while trying to keep the throughput close to its maximum. The amount of resources that this strategy is able to release compensates the slight reduction of the throughput compared to the deterministic strategy, especially when the link between the sender and the destination is highly degraded.

This paper shows how to find an optimal working point that balances throughput and resource efficiency according to the parameters that characterize the network. The proposed strategy provides a new view of cooperative diversity, in which the relay node not only assists the sender node in the retransmission process, but it also allows the sender to release radio resources increasing the overall utilization of the bandwidth.

Probabilistic retransmission has been previously considered in a very recent work (L. Xiong, 2008) as a strategy to balance cooperation and collision probability in order to achieve smaller latencies. In contrast, our work is focused in collision-free networks, and therefore its results are applicable to mobile access networks. Other works like (M. Dianati, 2006), (I. Cerutti, 2007) and (I. Cerutti, 2006) consider a deterministic retransmission scheme at the source node.

The contributions of this paper are:

- We develop an analytical model based on a discrete time Markov chain (DTMC), useful to compute the throughput. The simplicity of this model allows a closed-form solution, which is of great utility for the optimization analysis.
- The Markov model is complemented with a reward model for the derivation of the retransmission rate of the sender node.
- We propose a multi-objective optimization approach to adjust the retransmission probability balancing bandwidth efficiency and throughput.
- It is shown that it is possible to achieve a notable reduction of the retransmission rate of the sender node while keeping the throughput very close to the deterministic scheme.

The rest of the paper is organized as follows. In Section 2 we describe the system under study and its model as a DTMC. This model is used in Section 3 to analyze the performance of the system in terms of throughput and retransmission rate. The multi-objective optimization approach is presented in Section 4, and numerical results are discussed in Section 5. Finally, the implications and future research lines derived from this work are outlined in Section 6.

2 SYSTEM MODEL

The system under study, illustrated in Figure 1, consists of a sender node (S) that transmits data frames to a destination node (D), and a relay node (N) that receives the data frames directed to D, so it can assist S in retransmissions of lost frames. The physical layer consists of slotted radio interface. A time-slot is defined as the time from a frame transmission to the completion of its ACK/NACK. Slots are of fixed duration and synchronized at all the nodes. This kind of radio interface is characteristic of mobile access networks.

![Figure 1: C-ARQ system.](image_url)

We assume a simple channel model, similar to (L. Xiong, 2008), where the channel can be in one of two states: either “on”, in which transmitted signals arrive with sufficient power to be decoded without error, or “off”, in which a transmitted signal can not be decoded. The probability of the channel being “on”
or “off” in a given time-slot is independent of its state in previous time-slots, and independent of the channel states of different node pairs. The probability of being in the "off" state is denoted by \( p_{S} \) for the channel between S and D (direct link), \( p_{D} \) for the channel between S and N, and \( p_{C} \) for the channel between N and D (relay channel).

It is assumed that ACK/NACKs are not lost in any channel, therefore both S and N are informed of the reception status of the frame in D. This assumption is common in performance evaluation of wireless networks, because the shorter length of control messages make it feasible to protect them with longer error correction codes. In the subsequent time-slots after the original transmission, if the relay node possesses a copy of the packet, it may decide to make a cooperative retransmission over its relay channel.

The probability that the relay node performs a retransmission in a time-slot is denoted by \( p_{N} \). This probability depends on the amount of resources that N can assign to the communication between S and D given its traffic load, its scheduling policy and its processing limitations. In the numerical examples of section 5, we set \( p_{N} = 1 \), assuming a relay node fully devoted to cooperation, in line with recent proposals for mobile access networks where relay nodes are deployed as part of the radio cell infrastructure. However, the analysis is valid for \( p_{N} < 1 \), and the conclusions can be generalized to this case as well. The model includes a probability of retransmission in the sender node, denoted by \( p_{S} \). The analysis of the influence of this parameter in the overall performance and its optimization is one of the main goals of this paper. Note that this implies that the node computing the optimum \( p_{S} \) (in general, the sender node) must be informed of the link qualities and the retransmission probability of the relay node.

For successful reception, the destination node must decode without errors at least one frame sent by either S or N. The radio interface is collision-free, as explained in the introduction. We assume a stop-and-wait operation of the protocol, similarly to (M. Di-anati, 2006). However, due to the probabilistic retransmission scheme at the sender, it may transmit new frames to the receiver in time slots not assigned to retransmissions. Each frame transmission is then handled by independent parallel processes in a way similar to HSDPA, (H. Holma, 2006).

The states of the Markov chain describing the operation of the system are combinations of the individual states of N and D. Let \( N_f \) represent the state of the relay node. \( N_f = true \) (or simply \( N_f \)) if the relay node has successfully decoded the frame, and \( N_f = false \) (or simply \( \overline{N_f} \)) otherwise. Because the relay node is assumed not to discard frames, \( N_f = true \) in every time-slot between the successful reception of a frame at N and the reception of this frame at D. The states of the destination node are: \( D_f \), if the frame is correctly decoded, and \( \overline{D_f} \) otherwise.

The DTMC comprises the following three states:

- State 0: \( \overline{N_f}, \overline{D_f} \)
- State 1: \( N_f, \overline{D_f} \)
- State 2: \( D_f \)

The transition diagram of the DTMC is depicted in Figure 2. In order to describe analytically the transition probabilities let us define the following events:

- \( S \): S retransmits a frame.
- \( N \): N retransmits a frame.
- \( N_f \): N successfully decodes a frame sent by S.
- \( D_f \): D successfully decodes a frame sent by N.
- \( D \): D successfully decodes a frame sent by S.
- \( S, \overline{N}, \overline{N}_f, \overline{D}, \overline{D}_f \): are the complementary of the events above.

Making use of the description of the C-ARQ protocol and the definition of the events, the transition probabilities from state 0 are given by:

\[
\begin{align*}
 p_{00} &= P\{ S \land \overline{D} \land \overline{N} \} \\
 p_{01} &= P\{ S \land D \land \overline{N} \} \\
 p_{02} &= P\{ S \land \overline{D} \} 
\end{align*}
\]

Applying the channel state and retransmission probabilities of the model we obtain the following values:

\[
\begin{align*}
 p_{00} &= p_{PA} p_{PB} + (1 - p_{S}) \\
 p_{01} &= p_{PA} (1 - p_{B}) \\
 p_{02} &= p_{S} (1 - p_{A}) 
\end{align*}
\]  

When the system is in state 1, it can not enter into state 0, because the relay node does not discard frames (\( p_{10} = 0 \)). The system remains in the same state with the following probability:

\[
\begin{align*}
 p_{11} &= P\{ (N \land \overline{D}_f) \land (S \land \overline{D}) \} \\
 &= (p_{PN} p_{C} + (1 - p_{N}))(p_{PS} p_{A} + (1 - p_{S}))
\end{align*}
\]

Figure 2: Discrete Time Markov Chain model of the C-ARQ protocol.

PROBABILISTIC RETRANSMISSION STRATEGY FOR SINGLE-RELAY COOPERATIVE ARQ
The transition matrix of the DTMC is given by
\[
P = \begin{pmatrix}
p_{00} & p_{01} & p_{02} \\
p_{20} & p_{21} & p_{22}
\end{pmatrix}
\] (4)

3 PERFORMANCE ANALYSIS

In this section we use the proposed model to obtain two important performance metrics of the C-ARQ scheme: the throughput and the retransmission rate of the source. The throughput is defined as the number of time-slots that the source can transmit frames per time-slot. According to the model, the throughput is the average number of time-slots that the DTMC spends in state 2. 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. Therefore, for an efficient use of the bandwidth, the goal is to reduce this rate. For the computation of the retransmission rate, a reward model is constructed.

3.1 Throughput Performance

Let \( \pi = \{ \pi_0, \pi_1, \pi_2 \} \) be the steady-state distribution of the DTMC, where \( \pi_i \) is the steady-state probability of state \( i \in \{0,1,2\} \). \( \pi \) is obtained by solving the following system of linear equations:

\[
\pi = \pi P \sum_{i=0}^{2} \pi_i = 1
\] (5)

where \( \Omega = \{0,1,2\} \), and \( P \) is given by (4). Solving (5) for \( \pi \) we obtain the following solution:

\[
\pi_2 = \left[ 1 + \frac{p_{20}}{1 - p_{00} p_{01} p_{21}} + \frac{p_{21}}{1 - p_{00} (1 - p_{11})} \right]^{-1}
\]

\[
\pi_1 = \left[ \frac{p_{01} p_{21}}{(1 - p_{00}) (1 - p_{11})} + \frac{p_{21}}{1 - p_{11}} \right] \pi_2
\]

\[
\pi_0 = \frac{p_{20}}{1 - p_{00}} \pi_2
\]

From the transition probabilities obtained in Section 2 we can compute \( \pi \) as \( \pi(p_A, p_B, p_C, p_s, p_N) \). Because we focus on finding the optimal \( p_s \), we use, for convenience, a simpler notation: \( \pi(p_s) \). The throughput of the system is denoted by \( T(p_s) = \pi_2(p_s) \). It is obvious that, for any given set of values \( \{p_A, p_B, p_C, p_N\} \), the maximum throughput is obtained when the source retransmits in every time-slot after a frame loss notification \( p_s = 1 \). Therefore we define the maximum throughput as \( T_M = T(1) \). \( T_M \) is taken as a reference to evaluate different values of \( p_s \).

3.2 Retransmission Rate of the Source

In this subsection we develop a reward model to compute the average retransmission rate in the source node. This approach has been previously applied to the analysis of ARQ protocols. See (M. Zorzi, 1996) for a description of this technique.

In the DTMC considered, let \( X_i \to X_j \) represent a transition from state \( i \) to state \( j \). Let \( R_{ij} \) be the reward associated to this transition. In our context, \( R_{ij} \) represents the average number of retransmissions from \( S \), in the time-slots when \( X_i \to X_j \). Analytically, it is expressed as \( R_{ij} = r P \{ S | X_i \to X_j \} \), where \( r \) is the number of frames in a retransmission time-slot. In the system under study, \( r = 1 \) because only one frame can be transmitted per time-slot.

The reward associated to a transition is 1 if the transition can only take place when the source retransmits the frame. Therefore, any transition from state \( 0 \) to a different state involves a single reward:

\[
R_{01} = R_{02} = 1
\]

(7)

The first transmission of a frame is not considered a retransmission, therefore the transitions from state 2 involve a null reward:

\[
R_{20} = R_{21} = R_{22} = 0
\]

(8)

The reward associated to \( X_0 \to X_0 \) is given by

\[
R_{00} = P \{ S | X_0 \to X_0 \}
\]

which, making use of the definition of conditional probability can be expressed as

\[
P \{ S | (X_0 \to X_0) \} = \frac{P(S \wedge D_{00} \wedge D_{0})}{P(X_0 \to X_0)}
\]

Applying the probabilities of each event, we obtain the following value:

\[
R_{00} = \frac{p_s p_A p_B}{p_{00}}
\]

(9)

Similarly, \( R_{11} \) equals the following conditional probability:

\[
P \{ S | (X_1 \to X_1) \} = \frac{P(S \wedge D_{01} \wedge ((N \wedge D_{02}) \vee N_5))}{P(X_1 \to X_1)}
\]
which, according to the model results in
\[ R_{11} = \frac{pSP_p(p_p p_c + 1 - p_N)}{p_{11}} \] (10)

Finally, \( R_{12} \) is given by
\[ P[S|X_1 \rightarrow X_2] = \frac{P[\{S \wedge D_S\} \cup \{N \wedge D_N\}]}{P[X_1 \rightarrow X_2]} \]

which results in
\[ R_{12} = \frac{p_S(1 - p_A) + p_S p_A p_N(1 - p_C)}{p_{12}} \] (11)

The retransmission rate associated to a reference state, \( i \), is computed with the following expression:
\[ R_i = \sum_{j \in \mathcal{A}} p_{ij} R_{ij} \] (12)

The average retransmission rate from S is given by a weighted sum of the rates obtained in (12), where the weighting factors are the steady state probabilities of the Markov chain:
\[ R = \sum_{i \in \Omega} \pi_i R_i \] (13)

Applying the rewards (7), (8), (9), (10) and (11) in (12), it can be easily checked that \( R_0 = R_1 = p_S \) and \( R_2 = 0 \). For convenience, we can write R in terms of \( p_S \) as \( R(p_S) = p_S(1 - \pi_2(p_S)) \). At the maximum throughput, obtained with \( p_S = 1 \), the retransmission rate is \( R_M = R(1) \).

4 MULTI-OBJECTIVE OPTIMIZATION STRATEGY

There exists a clear tradeoff between the two performance measurements derived in previous section. The maximum throughput, \( T_M \), is achieved 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 between both objectives, we propose an approach based on a multi-objective optimization technique, known as global criterion method (Rao, 1996). This method consists of minimizing a global criterion function, defined as:
\[ G(p_S) = \sum_{k=1}^{n} \alpha_k \left( \frac{O_k - f_k(p_S)}{O_k} \right)^2 \] (14)

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 \( \alpha_k \) 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 its maximum, \( T_M \), 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 \( \alpha_2 \).

Making use of these definitions in (14) we obtain:
\[ G(p_S) = \alpha_1 \left( \frac{T_M - T(p_S)}{T_M} \right)^2 + \alpha_2 R(p_S)^2 \] (15)

Let \( p_S^* \) denote the solution to the multi-objective optimization problem. Because \( p_S^* \) is a probability, the problem is subject to the constraint \( 0 \leq p_S \leq 1 \). Therefore we have:
\[ p_S = \arg \min_{0 \leq p_S \leq 1} \{ G(p_S) \} \] (16)

Let \( T_0 = T(p_S^*) \) be the optimum throughput in terms of the multi-objective optimization problem. Similarly, let \( R_0 = R(p_S^*) \) denote the optimum retransmission rate. In the following section we compare \( T_0 \) and \( R_0 \) with \( T_M \) and \( R_M \) in several scenarios. The weighting factors chosen are \( \alpha_1 = 1 \) and \( \alpha_2 = 0.2 \). Additionally, the effect of \( \alpha_2 \) is also discussed in the following section.

5 RESULTS

In this section we investigate the performance of the probabilistic retransmission approach numerically. It is shown that, with the computed probability, the system achieves a notable reduction in the retransmission rate of the source with a relatively small reduction of the throughput. Moreover, we show that the two objectives are well-balanced in many different combinations of link quality conditions.

In the framework of cooperative mobile access networks, relay nodes are considered to be part of the cell infrastructure, specifically deployed to enhance the connectivity of mobile users. In this scenario, a relay node is always willing to cooperate. Assuming that there are enough resources available, \( p_N = 1 \). The frame error probability in the source-relay link can be a network design parameter, and therefore its value can be relatively small. In the numerical evaluations of this section we set \( p_B = 0.2 \).

First, we study the effect of \( p_B \) in the performance of the proposed strategy, and compare it with the maximum throughput configuration \( (p_S = 1) \). For this
In order to evaluate the effect of the weighting factors, we can set $\alpha_1 = 1$ and compute the performance for a range of values of $\alpha_2$. Let consider the previous configuration of the system, with $p_A = 0.4$. Figure 5 plots the performance metrics obtained with $\alpha_2$ in the range $10^{-2} \leq \alpha_2 \leq 10$. While it was expectable that both $T_O$ and $R_O$ decrease as $\alpha_2$ is set to higher values, it is surprising to check that the throughput does not decay dramatically when $\alpha_2 = 1$ or even at higher values. From the values of $p_S^*$ shown in Figure 6, we observe that a reduction of $p_S$ from 0.9 to 0.4 (more than 50%), causes a reduction of about 10% in the throughput. This fact suggests that, in practice, the system tolerates certain inaccuracy in the computation of $p_S$, without much impact on the throughput.

Because of the mobility of the user terminals, many different combinations of $p_A$ and $p_C$ are possible. Figure 7 shows the computed $p_S^*$ for different settings of both parameters. In order to focus on the performance improvement obtained with $p_S^*$ compared to $p_S = 1$, Figure 8 plots the difference between $R_M$ and $R_O$, and Figure 9 depicts the difference between $T_M$ and $T_O$. These figures show that, while the retransmission rate is highly reduced, especially in poor channel conditions, the throughput never decays noticeable below the maximum. We can observe that the proposed algorithm provides a good balance between the two objectives pursued, even under different combinations of link qualities.

6 CONCLUSIONS

The analysis done in this paper shows that, in wireless networks with cooperative ARQ, a probabilistic retransmission policy in the source node provides notable benefits in terms of efficiency in the use of radio resources, especially in situations of poor propagation.
conditions in the direct link. The optimal retransmission probability for the source node is obtained by means of a multi-objective optimization technique, known as global criterion method. It is shown that this strategy reduces the retransmission rate of the source node with a negligible reduction of the throughput.

This new concept breaks with the classical and widely adopted policy of giving maximum priority to retransmissions in wireless links. Consider, as an example, the Radio Link Protocol of 3G access networks (H. Holma, 2004). As a consequence, the results of this paper can be useful in the design of scheduling mechanisms for cooperative wireless networks. The idea is to consider the computed retransmission probability, \( p^*_S \), as a lower bound for the amount of time-slots (i.e. bandwidth) reserved for retransmissions at the base station. The rest of the bandwidth can be assigned to other communication processes whenever they have available data.

The model described can be extended to include the concept of cooperation group defined in (M. Dianati, 2006). With this enhancement, the cooperation involves an undefined number of relays. Another future line of research consists in the addition of Hybrid ARQ features e.g. incremental redundancy and chase combining, in the model.

**ACKNOWLEDGEMENTS**

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 Seneca, Agencia de Ciencia y Tecnología de la RM (Plan Regional de Ciencia y Tecnología 2007/2010)”.

**REFERENCES**


