Delay Model of Single-Relay Cooperative ARQ Protocols in Slotted Radio Networks with Non-Instantaneous Feedback and Poisson Frame Arrivals

Isabella Cerutti Scuola Superiore Sant'Anna, Pisa, Italy E-mail: i.cerutti@sssup.it Andrea Fumagalli, and Puja Gupta The University of Texas at Dallas Emails: {andreaf,pkg021000}@utdallas.edu

Abstract—In cooperative ARQ protocols, data frame retransmissions may be performed by a neighboring node (the relay) that has successfully overheard the source's frame transmission. One advantage is the diversity provided by the relay. The three-way (source, destination, relay) frame exchange sequence required in the cooperative ARQ protocols may however introduce extra latency when compared to non-cooperative ARQ protocols. To take advantage of cooperative ARQ protocols, it is then necessary to resort to selective repeat solutions.

The focus of the paper is to derive a delay model for cooperative selective repeat ARQ protocols in slotted radio networks. The derived analytical model quantifies, with closed formulas, the queueing and transmission delay experienced by Poisson arriving data frames, whose retransmissions are performed by a single relay.

I. Introduction

As various factors (e.g., path loss, fading, and noise) affect the signal quality in radio networks, automatic repeat request (ARQ) protocols are used to guarantee reliable data delivery over the radio channel. By exploiting the broadcast nature of the radio medium, a node(s), which is within earshot from the source and the destination, may actively help deliver the data frame correctly [1]-[3]. This node is commonly referred to as the *relay* and may be involved in the data frame retransmission process as follows. When the source's data frame transmission is not successful, the relay is invited to take part in the frame retransmission process, typically upon reception of a negative acknowledgment from the destination. By doing so, the destination can rely on data frames that are transmitted by both the source and relay, possibly yielding a better overall reception quality. The ARQ protocols based on the relay's retransmission of the data frame are referred to as *cooperative*.

The common assumption used across the existing body of work is the presence of an instantaneous feedback channel, i.e., the relay(s) can take prompt action, without introducing further latency in the ARQ protocol. Under this assumption, the cooperative stop and wait ARQ protocols studied

This research was supported in part by NSF Grants No. ANI-0082085, ECS-0225528, CNS-0435429. and the Italian Ministry of University (MIUR) under FIRB project "Enabling platforms for high-performance computational grids oriented to scalable virtual organizations" (contract n. RBNE01KNFP).

recently are the simplest efficient solution [4]. Some radio networks, however, may be more accurately modeled with non-instantaneous feedback channel due to signal propagation time, de/coding process latency, hardware switching time, and other ARQ protocol requirements. While non-instantaneous feedback channel is thoroughly studied for non-cooperative ARQ protocols [5], the adverse effects that it may have on cooperative ARQ protocols remain to be studied.

The focus of this paper is to gain further knowledge about cooperative ARQ protocols when they are applied to radio networks with non-instantaneous feedback channel. First, it is demonstrated that, in slotted collision-free (TDMA) radio networks characterized by relatively small channel latency, cooperative stop and wait ARQ protocols may not be efficient, unlike the non-cooperative ones. The reason is that, when the outcome of the relay's retransmission attempt is yet unknown, the source cannot continue with the transmission of the next data frame. This happens when the overall time for requesting and completing the data frame retransmission at the relay and for receiving the final acknowledgment exceeds the time interval between two consecutive transmission slots at the source. For these networks, it is then necessary to resort to the more efficient cooperative selective repeat ARQ protocols.

Simple cooperative selective repeat ARQ protocols, based on (type I) hybrid retransmission techniques [6], are then considered in the paper. A delay model is derived to evaluate, with a closed formula, the expected queueing and transmission delay, i.e., the time spent by a data frame in the source buffer plus the time interval from the start of its first transmission to the complete reception of its positive acknowledgment.

The performance of the cooperative selective repeat ARQ protocols is compared against an efficient non-cooperative ARQ protocols based on stop and wait mechanism, i.e., (type I) Hybrid-ARQ (H - ARQ) [6].

II. SINGLE-SOURCE SINGLE-RELAY COOPERATIVE ARQ PROTOCOLS

A. System Assumptions

In describing and studying the cooperative ARQ protocols, the following simplifying assumptions are made.

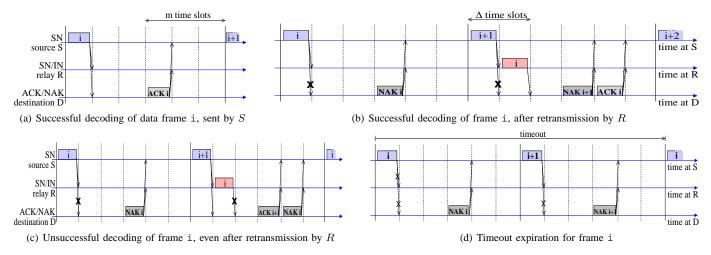


Fig. 1. Time chart of C-ARQ protocol for M=6, m=3, and $\Delta=2$

Consider a network, in which M mobile stations (MS) are in close proximity and share with time division multiplexing two common channels: one for uplink transmission to the base station (BS) and the other for downlink transmission from the BS. Time is divided into time frames of duration T_F . Each time frame is divided into M time slots. Full duplex transmission is used. Each MS is scheduled to transmit (receive) in a given time slot of the uplink (downlink) channel. Uplink time frames are delayed by m < M slots with respect to the downlink time frames, to accommodate for the non-instantaneous feedback channel.

Data frames may carry some degree of redundancy to enable error detection and correction at the receiver. Transmitted data frames are always received. However, the data frame payload may not be decoded correctly at the receiver due to transmission errors. For this reason, ARQ protocols must be used to provide reliable delivery of data frames to the destination.

B. Cooperative ARQ Protocol Description

The cooperative ARQ protocol (C-ARQ) is described assuming that MS source S is scheduled to transmit to the BS destination D^{-1} in the first time slot of each time frame. MS relay R is the only relay assigned to help S (single-relay case)². Note that R must be provided with the capability of "eavesdropping" the transmission from S to D.

Let $\Delta-1$ be the time offset from S to R, i.e., the time interval between the beginning of the time slots assigned to S and the beginning of the time slot assigned to R in the same time frame is $(\Delta-1)T_F/M$. Δ is an integer value in [2,M]. Depending on the value of Δ , the following cases are possible. (Note that further limitations on Δ may apply in half-duplex systems.)

Case 1: if $\Delta \leq m$ and $\Delta \geq M-m+2$, both R and S receive the acknowledgments from D prior to their next scheduled transmission slot. In this case, a cooperative stop and wait ARQ protocol can be efficient, as demonstrated in [4].

Case 2: if $\Delta \geq \max(m+1, M-m+1)$ or $\Delta \leq \min(M-m+1, m+1)$, R can eavesdrop the transmitted acknowledgment from D to S and decide whether or not the data frame must be retransmitted. However, the acknowledgment from D to R is received by S only after the next scheduled transmission slot for S. In this case, to be efficient, the cooperative ARQ protocol must be a selective repeat, whereby up to two data frames may be under transmission at once, one at S and the other at S. This case is further studied in the paper.

In the protocol description that follows, SN indicates the data frame sequence number. ACK/NAK indicates the data frame sequence number successfully/unsuccessfully received.

When it is required, R transmits an exact replica of the data frame received from S. The following four sequences of data frame and ACK exchanges are possible, as shown in Fig. 1, for M=6, m=3, and $\Delta=2$.

- a) Fig. 1(a): D successfully receives the data frame transmitted by S. Data frame SN=i is transmitted by S and acknowledged by D with the transmission of ACK=i. In the next time frame, S may transmit data frame SN=i+1.
- b) Fig. 1(b): D successfully receives the data frame with the help of R. Data frame SN=i is transmitted by S. It is not correctly decoded by D. However, it is correctly received and decoded by R. D sends a (re)transmission request to R using NAK=i. R can transmit data frame SN=i in the next time frame. The frame is correctly received by D, which sends control frame ACK=i to S. Note that while R is retransmitting data frame SN=i, S begins a new sequence and transmit data frame SN=i+1.
- c) Fig. 1(c): D does not receive successfully the data frame due to some transmission error(s) detected in the data frame from R. This sequence begins in a way similar to **b**. This

¹With minor changes, it is possible to derive the case in which the data flow is from the BS to the MS.

²The case in which multiple relays assist the same source is outside the scope of the paper.

time, however, data frame SN=i transmitted by R is not correctly received by D. D sends NAK=i to S which begins a new transmission sequence of data frame SN=i, two time frames after the first transmission of data frame SN=i.

d) Fig. 1(d): timeout expires. For various reasons, S may not receive the ACK from D. In this case, a timeout is used at S to avoid deadlock. The timeout duration is set to be two time frames. In the example shown, data frame SN=i is transmitted by S. It is not successfully decoded by D. D sends a (re)transmission request to R using NAK=i. However, not even R is able to decode successfully the data frame transmitted by S. Thus, it cannot cooperate, and the request from D is discarded. Upon expiration of the timeout, S begins a new transmission sequence of data frame SN=i.

Recall that both S and R are assigned one distinct time slot in every time frame. For each transmission attempt made by S, R may help with its own transmission attempt. Thus, R need not store more than two data frame copies at any time, and each copy need not be stored at R for more than two time frames.

III. ANALYTICAL FRAMEWORK

This section derives the queueing model that is used to estimate the C-ARQ protocol performance.

A. Model Assumptions

For making the protocol analytically tractable, the following assumptions are introduced. The C-ARQ protocol is considered fully reliable, i.e., data frames are retransmitted until correct reception. Both transmitter and receiver buffers at all nodes have unlimited size.

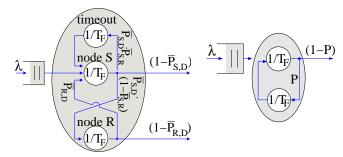
Data frame arrivals follow a continuous-time Poisson process of rate λ . When $\lambda \to 1$, the source buffer is never empty, modeling heavy traffic. Transmissions of data frames or acknowledgment control frames initiate only at the beginning of a time slot and extend over one time slot, T_F/M .

It is assumed that while cooperation takes place, R does not generate traffic of its own and simply helps S deliver data frames to D.

Probability that an acknowledgment control frame is received uncorrectly is negligible. (This last assumption can be removed, if necessary [5].) Also, it is assumed that the frame error probability, $\overline{P}_{i,j}$, i.e., probability that node j unsuccessfully decodes the frame sent by node i, is known.

B. C-ARQ Queueing Model with Three-Stage Service Facility

Fig. 2(a) shows the queueing models of the C-ARQ protocols for evaluating the queueing and transmitting delay up to the next slot for S, i.e., the time interval between a data frame arrival in the buffer and the start of slot for S after its acknowledgment. Three servers are shown, i.e., node S, node R, and timeout. Timeout is required when the data frame is not correctly received by the relay. The service time of all three servers is one time frame, T_F .



(a) Three-stage service facility

(b) Two-stage service facility

Fig. 2. C - ARQ protocol queueing model

Data frames enter with rate λ . At the end of each service, the frame moves to another server or leaves the system as indicated by the transition probabilities shown in the figure.

It is possible to prove that the queue statistics (i.e., queueing delay and buffer occupancy) of the queueing models with the three-stage service facility (Fig. 2(a)) and with the two-stage service facility (Fig. 2(b)) are equivalent if $P = \overline{P}_{S,D} \cdot (1 - \overline{P}_{S,R}) \cdot \overline{P}_{R,D} + \overline{P}_{S,D} \cdot \overline{P}_{S,R}$. The study of the C - ARQ queueing model is carried out in the next section using the queueing model with the two-stage one-exit service facility.

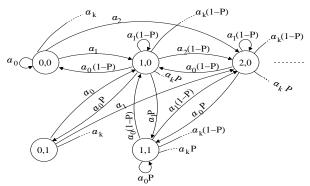


Fig. 3. Embedded Markov chain for two independent servers

1) Embedded Markov Chain: The (discrete-time) embedded Markov chain is derived by sampling the state of the queue shown in Fig. 2(b) at the beginning of each time frame. The state of the chain is the tuple $S_{i,j}$, whereby $i \geq 0$ indicates the number of jobs waiting in queue and in the top server, and $j = \{0,1\}$ indicates the number of jobs in the bottom server. Fig. 3 illustrates the state transition diagram and the transition probabilities of the chain. Note that $a_k = e^{-\lambda T_F} (\lambda T_F)^k / k!$ is the probability of k Poisson arrivals during one time frame.

The utilization of the top server is

$$\rho = \frac{\lambda T_F}{1 - P}.\tag{1}$$

The Markov chain is ergodic if the stability condition is satisfied, i.e., $\rho < 1$. Under this condition, the three steady state probabilities, i.e., $\pi_{i,j} = Prob\{\text{chain is in state } S_{i,j}\}, \forall i+j \leq 1$, can be derived by solving the system of equations

$$\begin{cases}
\pi_{0,0} = a_0 \pi_{0,0} + a_0 (1 - P) \pi_{1,0} \\
\pi_{0,1} = a_0 P \pi_{1,0} \\
\pi_{0,0} + \pi_{0,1} = 1 - \rho.
\end{cases}$$
(2)

2) Expected Queue Size and Server Occupancy: As shown next, the queue size and the top server occupancy is a function of the three state probabilities in (2). The derivation is based on the second moment of the number of jobs awaiting in the queue and in the top server. (This technique is similar to the M/G/1 derivation presented in [7].)

Let v_n be the number of jobs arriving during time frame n. Let q_n be the number of jobs awaiting in the queue and in the top server at the beginning of time frame n. Let U_{q_n} be the shifted unit step function in q_n , i.e.,

$$U_{q_n} = \begin{cases} 1 & \text{if } q_n > 0\\ 0 & \text{otherwise.} \end{cases}$$
 (3)

Let $U_{Pq_{n-1}}$ be the shifted unit step function in q_{n-1} with probability P

$$U_{Pq_{n-1}} = \begin{cases} 1 & \text{with probability } P \text{ if } q_{n-1} > 0 \\ 0 & \text{otherwise.} \end{cases}$$
 (4)

The value of q_{n+1} is given by

$$q_{n+1} = q_n + v_n - U_{q_n} + U_{Pq_{n-1}}. (5)$$

Note that (5) takes into account q_n , i.e., the number of jobs in the system at time n, v_n , i.e., the number of newly generated jobs arrived during time frame n, U_{q_n} , i.e., the departure of a job from the top server, and $U_{Pq_{n-1}}$, i.e., the return of a previously served job to the top server.

The objective now is to evaluate the first moment of q_n when the embedded Markov chain is ergodic, i.e., when

$$\lim_{n \to \infty} E[q_n] = E[\tilde{q}]. \tag{6}$$

 $E[\tilde{q}]$ may be computed from the expectation of the square of (5), i.e.,

$$\begin{aligned} q_{n+1}^2 &= q_n^2 + v_n^2 + U_{q_n}^2 + U_{Pq_{n-1}}^2 + 2q_n v_n + \\ &- 2q_n U_{q_n} + 2q_n U_{Pq_{n-1}} - 2v_n U_{q_n} + \\ &+ 2v_n U_{Pq_{n-1}} - 2U_{q_n} U_{Pq_{n-1}}, \end{aligned} \tag{7}$$

for the limit as $n \to \infty$. Notice that

$$U_{q_n}^2 = U_{q_n} \tag{8}$$

$$U_{Pq_{n-1}}^2 = U_{Pq_{n-1}} \tag{9}$$

$$q_n U_{q_n} = q_n, (10)$$

and that the averages $E[q_nv_n]$, $E[v_nU_{q_n}]$, and $E[v_nU_{Pq_{n-1}}]$ may be rewritten in the product form.

In the presence of ergodicity, the first moment of U_{q_n} and U_{Pq_n} are, respectively,

$$E[U_{\tilde{q}}] = \lim_{n \to \infty} E[U_{q_n}] = Prob\{\text{server busy}\} = \rho$$
 (11)

$$E[U_{P\tilde{q}}] = \lim_{n \to \infty} E[U_{Pq_n}] = P \cdot Prob\{\text{server busy}\} = P\rho.$$
 (12)

As derived in [8],

$$\lim_{n \to \infty} E[q_n U_{Pq_{n-1}}] = PE[\tilde{q}] - P\pi_{0,0} \cdot \rho \cdot (1 - P) - P\pi_{0,1}(1 + \rho(1 - P))$$
(13)

$$\lim_{n \to \infty} E[U_{q_n} - U_{Pq_{n-1}}] = P(\rho - a_0 \pi_{1,0}), \tag{14}$$

where the expression for $\pi_{i,j}$ is given in (2).

One can solve for $E[\tilde{q}]$

$$E[\tilde{q}] = \frac{1}{2(1-\rho)(1-P)} \left[2\rho(1-P) - \rho^2(1-P)^2 \right]$$
 (15)

$$-2P\Big(\rho(1-P)\pi_{0,0}+(1+\rho(1-P))\pi_{0,1}-a_0\pi_{1,0}\Big)\Big].$$

C. C - ARQ Protocol Performance

The following C-ARQ protocol performance metrics can be obtained from the analytical model derived so far:

- \hat{N} : expected buffer occupancy at MS S at the beginning of a time frame,
- T: expected delay, defined as the expected time elapsed between the data frame generation at S and the end of the time slot in which S receives the acknowledgment from D.

 $E[\tilde{q}]$, obtained in the previous sections, represents the expected number of frames, excluding the outstanding frames, in the buffer or being transmitted at MS S (i.e., waiting in the queue or at node S server of the three-stage service facility). The buffer occupancy at S can be derived from $E[\tilde{q}]$, by including the outstanding frames at S (i.e., the frames at node R server and timeout server of the three-stage service facility)

$$\hat{N} = E[\tilde{q}] + \rho \cdot \overline{P}_{S.D}. \tag{16}$$

The expected delay T can be obtained by applying Little's theorem (first term)

$$T = \frac{\hat{N}}{\lambda} + \frac{\lambda T_F}{2} - \tau,\tag{17}$$

where the second term includes the time required by Poisson arrivals to synchronize to the time frame beginning, and τ is the expected interval between the reception of an ACK and the successive time frame beginning, i.e.,

$$\tau = (m-1)\frac{T_F}{M} \cdot \frac{1 - \overline{P}_{S,D}}{1 - P}$$

$$+ \left| m - \Delta \right|_M \cdot \frac{T_F}{M} \cdot \frac{\overline{P}_{S,D}(1 - \overline{P}_{S,R})(1 - \overline{P}_{R,D})}{1 - P}.$$

$$(18)$$

In the above expressions, the frame error probabilities must be replaced with their actual values computed for the $C\!-\!ARQ$ protocol, as explained in Section III-B.

IV. A CASE STUDY

In this section a case study is presented to assess the overall performance gains when using the cooperative ARQ protocols. These gains may originate from spatial diversity, use of R's bandwidth to retransmit S's data frames, and better S-R and R-D channel SNR than S-D channel SNR.

The selective repeat C-ARQ performance results are compared against the non-cooperative stop and wait H-ARQ. The analytical model for the H-ARQ protocols is derived as explained in [8].

The system parameters have been set as follows: $T_F = 1$, M = 8, and m = 5. Unless otherwise indicated, R is at half

distance between S and D, i.e., a good location for successful cooperation. The distance between S and D is taken as a reference distance of 1.

The radio channel is simulated assuming a 60 dB path loss with path loss exponent of 4, and fading effects on both data and incremental redundancy frame transmissions. Frequencyflat, block-Rayleigh fading (quasi-static) is assumed with fading levels that are constant over the duration of an entire data frame transmission, i.e., time slot. The fading levels are statistically independent of the time slot, channel, and space. (These assumptions tend to favor the non-cooperative protocols, as in reality it is expected that cooperative protocols are more robust over non-cooperative protocols when fading is correlated in time, due to their spatial diversity [3].) Payload and CRC comprise 128 bits that are encoded into 256 bit codewords using a rate-compatible punctured convolutional code (RCPC) with rate 1/2, parent code rate of 1/4, puncturing period of 8, memory of 4 and generator polynomials G(23,35,27,33)(octal) [4], [9]. The relation between the channel SNR and the frame error $\overline{P}_{i,j}$ is shown in [10]. Simulation results have confidence interval values of 5% or better, at 95% confidence level. The expected values of the error frame probabilities, obtained through Monte Carlo integration, are used for the time invariant frame error probabilities, required in the analytical model.

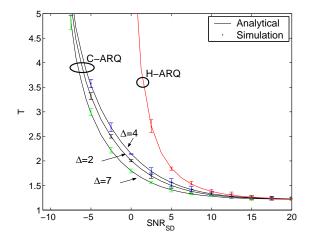


Fig. 4. T vs. $SNR_{S,D}$ (dB), $\lambda = 0.3$, $\Delta = 2, 4, 7$

Fig. 4 plots the expected frame delay, i.e., T, for various SNR values of the S-D channel, i.e., $SNR_{S,D}$. The arrival rate for the two ARQ protocols is $\lambda=0.3$. For the C-ARQ protocol, $\Delta=2$, 4, and 7. Fig. 5 plots T versus λ , for the two ARQ protocols, when $SNR_{S,D}=3$ dB.

In the figures, analytical and simulations results are reported as solid curves and confidence interval bars, respectively. The good match between the analytical results and the simulation results under various conditions supports the correctness of the derived model. The figures reveal that the known superiority of the cooperative over non-cooperative ARQ protocols still holds in the presence of non-instantaneous feedback, especially in the low SNR region and for high arrival rates. Finally, from the plots it is possible to deduce that for the C-ARQ protocols

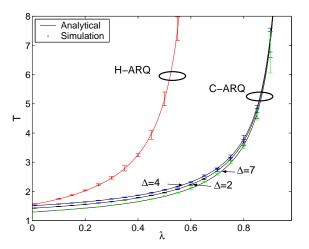


Fig. 5. T vs. λ , $SNR_{S,D}=3$ dB, $\Delta=2,4,7$

the best performance is achieved when $\left| m - \Delta \right|_M$ in (18) is minimum.

V. Conclusion

The paper derived a delay model for cooperative selective repeat ARQ protocols in TDMA-based radio networks with non-instantaneous feedback. The derived model estimates the queueing and transmission delay experienced by Poisson arriving data frames in the single-source and single-relay case.

Numerical results obtained from the model are encouraging and indicate under what conditions the cooperative ARQ protocols are superior when compared to non-cooperative ARQ protocols. Results show that the cooperative advantages — e.g., spatial diversity and relay bandwidth — may fully overcome the extra latency that may originate from the three-way nature of cooperative ARQ protocols.

REFERENCES

- [1] A. Nosratinia, T. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Comm. Mag.*, vol. 42, no. 10, 2004.
- [2] J. N. Laneman, G. W. Wornell, and D. N. C. Tse, "An efficient protocol for realizing cooperative diversity in wireless networks," in *Proc. IEEE* ISIT, 2001.
- [3] B. Zhao and M. C. Valenti, "Practical relay networks: a generalization of hybrid-ARQ," *IEEE J. Select. Areas Commun.*, vol. 23, no. 1, 2005.
- [4] 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. Networking*, to appear.
- [5] L. Badia, M. Rossi, and M. Zorzi, "Queueing and delivery analysis of SR ARQ on Markov channels with non-instantaneous feedback," *Proc. IEEE GLOBECOM*, 2005.
- [6] S. Lin, D. Costello, and M. Miller, "Automatic-repeat-request errorcontrol schemes," vol. 22, no. 12, 1984.
- [7] L. Kleinrock, Theory, Volume 1, Queueing Systems. John Wiley and Sons, 1975.
- [8] I. Cerutti, P. Gupta, and A. Fumagalli, "Cooperative ARQ protocols in slotted radio networks," The University of Texas at Dallas, http://opnear.utdallas.edu/publications/reports/UTD-EE-12-2005.pdf, Tech. Rep. UTD/EE-12/2005, 2005.
- [9] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *IEEE Trans. Commun.*, vol. 36, no. 4, pp. 389–400, 1988.
- [10] M. Tacca, P. Monti, and A. Fumagalli, "Cooperative and non-cooperative ARQ protocols for microwave recharged sensor nodes," in *Proc. EWSN*, 2005.