Analysis of a Type II Hybrid ARQ Strategy in a DS-CDMA Packet Transmission Environment

Jordi Pérez-Romero, Associate Member, IEEE, Ramón Agustí, Member, IEEE, and Oriol Sallent, Member, IEEE

Abstract—In this letter, a type II hybrid automatic repeat request scheme is considered as a retransmission strategy in a direct-sequence code-division multiple-access packet mobile radio network. An analysis based on the Equilibrium Point Analysis is presented to model the behavior of the system in a message-based traffic generation model. A simulation approach is introduced to validate the proposed analytical model obtaining results that closely match those derived theoretically.

Index Terms—Automatic repeat request (ARQ), channel coding, code-division multiple access (CDMA), Markov processes, packet radio.

I. INTRODUCTION

■ UARANTEEING transmission reliability at the logical Ink control (LLC) sublayer in a packet radio link is the main task of the automatic repeat request (ARQ) strategies through the retransmission of those packets that have not been successfully received. The so-called hybrid ARQ strategies are located in an intermediate position between ARQ and forward error correction (FEC) strategies, by combining the best of these two techniques. Basically, there are two hybrid ARQ strategies: the type I hybrid ARQ, which makes use of channel coding to protect the information in a similar way as the FEC strategy, but which also carries out retransmissions whenever the code is unable to correct all the errors in the received packet, and the type II hybrid ARQ (from now on referred to as ARQ-II) where an information packet with only error detecting capability is initially sent and, whenever a retransmission is required, redundancy is transmitted instead of repeating the same packet. As a result, the receiver will make use of this redundancy together with the previous data packet to perform error correction ([1], [2]).

Not much effort has been devoted so far in the open literature to assess hybrid ARQ strategies when combined with the multiple access technique scheme direct-sequence code-division multiple access (DS-CDMA) that is emerging as the predominant multiple-access scheme in third-generation mobile communications systems. In [3] a type I hybrid ARQ is considered in a slotted DS-CDMA network in the presence of jamming, while in [4] and [5] a comprehensive analytical model is introduced to evaluate the packet transmission performance in the framework of the combined DS-CDMA ARQ-II technique. However, this model is unable to cope in a manageable

Paper approved by B. Jabbari, the Editor for Wireless Multiple Access of the IEEE Communications Society. Manuscript received May 15, 1999; revised July 15, 2001; May 15, 2002; and February 15, 2003. This work was supported by CICYT under Project TIC 2001-2222.

The authors are with the Department of Signal Theory and Communications, Universitat Politècnica de Catalunya (UPC), Barcelona 08034, Spain (email: jorperez@tsc.upc.es; ramon@tsc.upc.es; sallent@tsc.upc.es).

Digital Object Identifier 10.1109/TCOMM.2003.815081

way with the presence of even moderate buffer sizes, which are required by ARQ-II in the transmitter site. In this letter we introduce a novel analytical model, which overcomes the above difficulties by allowing us to consider any buffer size dimensioning together with a more realistic message-based instead of packet-based analysis. In this context, a message is considered to be composed of a variable number of fixed length packets. The present letter is organized as follows: in Section II the ARQ-II scheme in the uplink of a DS-CDMA environment is described, while in Section III a Markov model is presented in order to analyze the system under perfect feedback conditions. Finally, in Section IV the proposed model is compared with simulation results and the conclusions are summarized in Section V.

II. DESCRIPTION OF ARQ-II IN A DS-CDMA ENVIRONMENT

In order to analyze the performance of ARQ-II in a slotted DS-CDMA packet radio network, a number of users transmitting L-bit packets in the different time slots in the uplink is considered. Each user has a previously assigned spreading code. Perfect power control is assumed to counteract the channel fading. Under these circumstances, the simplified improved Gaussian approximation (SIGA) can be used to model interference [6] and obtain the expression for the bit-error probability p_e as a function of the spreading factor S_f and the number of simultaneous users transmitting a packet in a given slot n_a .

Users generate messages that are divided into K-bit packets which are stored in a buffer. An error detecting code $C_0(L,K)$ is applied to each packet thus obtaining L-bit packets that will be transmitted in the different time slots. A half rate invertible code $C_1(2L, L)$ with error correcting capability t is also applied to each L-bit packet and as a result an L-bit redundancy packet is obtained. Such codes can be obtained from cyclic codes simply by removing some of the bits ([1], [7]). Whenever the feedback information in the downlink indicates that the first L bits contain errors, the second L bits are transmitted in the next slot and then the receiver can make use of these two packets to decode the information based on C_1 . When there are more than t errors within these two packets, retransmission of the first L bits is required, and then correction is performed again based on the last received set of 2L bits. If it is still necessary, the redundancy will be sent again and alternatively original L-bit packet and redundancy will be retransmitted until the information can be decoded successfully. It should be noted that when it is not possible to correct the packet even with the redundancy, retransmissions will be made with probability p_b in the successive slots, which allows the reduction of the number of simultaneous users in the system and therefore of the interference level. Finally, after the

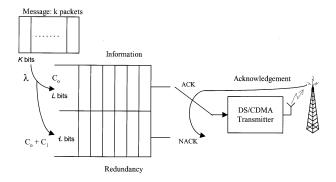


Fig. 1. ARQ-II DS-CDMA system model.

packet has been correctly received, normal transmission continues with the rest of the packets in the buffer. Fig. 1 represents the process explained above.

The *Stop & Wait* technique is considered, and the ideal situation in which the feedback information in the downlink is received error free and instantaneously is assumed, so that each user will know whether a new packet or a redundancy must be transmitted in the next slot. Similarly, it will be assumed that the code C_0 is able to detect all the residual errors in the packet after the correction procedure ([5]).

III. MARKOV MODELING OF ARQ-II DS-CDMA

A. Traffic Generation Model

Several traffic models have been proposed in the literature in order to approximate the data traffic in a realistic way. A common one is an ON/OFF model where the activity (ON) and inactivity (OFF) periods are exponentially distributed. In this case the ON/OFF dynamics can be modeled by a Markov chain, where the transition probability from the ON to the OFF state is $P_{\rm OFF|ON}=1-e^{-1/\beta_{\rm ON}}$, the probability of remaining in the ON state is $P_{\rm ON|ON}=e^{-1/\beta_{\rm ON}}$, the transition probability from the OFF to the ON state is $P_{\rm ON|OFF}=1-e^{-1/\beta_{\rm OFF}}$ and the probability from the OFF to the ON state is $P_{\rm ON|OFF}=1$. ability of remaining in the OFF state is $P_{\rm OFF|OFF}=e^{-1/eta_{\rm OFF}}.$ During the ON periods, each user generates messages according to a Bernouilli process with the probability $1 - e^{-\lambda} = 1 - P_T$ of a new message arriving in the next slot (i.e., an arrival rate of λ messages/slot). Notice that the generation process assimilates a Poisson arrival process by a Bernouilli process, assuming that $\lambda \ll 1$. Message length is geometrically distributed with a mean $(1/\mu)$ bits and, for simplicity, we assume the parity corresponding to code $C_0(L, K)$ to be included in this length. Messages are divided into L-bit packets and if necessary zeros are added in order to have a L-bit multiple. Then, the probability of a new message arriving containing k packets in the next slot during the ON period as a function of $P_0 = e^{-L\mu}$ (i.e., we assimilate an exponential law by a geometrical law by assuming $L\mu \ll 1$) is

$$P(k) = \begin{cases} P_o^{k-1} (1 - P_o)(1 - P_T), & k \ge 1\\ P_T, & k = 0. \end{cases}$$
 (1)

B. Markov State Definition

Let us consider the DS-CDMA system described in Section II, with a total number of U admitted users in the system,

each of them containing a buffer with room for M packets. System dynamics can be modeled by a Markov process where each user can be in a certain state a the beginning of each time slot. A state is denoted by i_X^S , where S represents the status of the source (i.e., $S \in \{ON, OFF\}$), i represents the number of packets waiting for transmission in the buffer $(1 \le i \le M)$ and X denotes the type of transmission to be carried out in the next slot (i.e., X = T if the first packet of the buffer will be transmitted for the first time, X = R if the redundancy corresponding to the first packet will be sent for the first time, or X = B when the packet has not been corrected by the initial redundancy, and therefore, information and redundancy are alternatively retransmitted with probability p_b). The states where the buffer is empty are denoted as 0^{ON} and 0^{OFF} .

Therefore, this approach requires $2 \cdot (3M+1)$ states, which represents an important difference from other proposed models in the literature [4], [5], which require each state to consider also the number of errors in the previously transmitted packet $(0,1,\ldots t-1,\geq t)$ and consequently, they need a number of states proportional to $M \cdot t$, which becomes unmanageable when t increases.

C. Successful Transmission Probability and Successful Packet Decoding Probability

Considering that errors will be equally distributed within the L bits of the packet, the probability $p_{\rm okt}$ of receiving an error-free packet as a function of the bit error probability p_e (that in turn depends on the spreading factor and the number of users transmitting simultaneously n_a according to the expressions given in [6]) is

$$p_{\text{okt}} = (1 - p_e)^L$$
. (2)

The calculation of the probability $p_{
m okr}$ of successfully decoding a packet after redundancy transmission is the key point that allows the number of states to be independent from the code correcting capability t. Strictly speaking, since the number of simultaneous users n_a may vary from slot to slot, this probability needs to consider the number of simultaneous users and the number of errors in the previous transmission (therefore, the states should take into account this number of errors and the number of states would increase with t). However, as the model intends only to derive the system equilibrium points in the steady-state, the approximation to compute $p_{\rm okr}$ relays on considering as the number of simultaneous users n_a the value in the steady state. The validity of this approximation can be checked afterwards by comparing the predicted results with the obtained through simulations. Accordingly, $p_{\rm okr}$ can be calculated as the probability of there existing fewer than t errors in a set of 2L bits (i.e., a set of original packet and redundancy) or there existing no errors in the redundancy packet, given that there is at least one error in the first L bits, as shown in (3) at the bottom of the next page.

D. State Transition Probabilities and Steady-State Distribution

System performance in the steady-state can be evaluated by means of the equilibrium point analysis (EPA) technique, consisting in finding the system equilibrium points characterized by the number of users in each state when the system reaches the steady-state distribution. These numbers are given in vectors $\mathbf{N_s^{ON}} = [N_o^{\mathrm{ON}}, N_{1T}^{\mathrm{ON}}, N_{1R}^{\mathrm{ON}}, N_{1B}^{\mathrm{ON}}, \dots, N_{MT}^{\mathrm{ON}}, N_{MR}^{\mathrm{ON}}, N_{MB}^{\mathrm{ON}}]^{\mathrm{T}}$ and $\mathbf{N_s^{\mathrm{OFF}}} = [N_o^{\mathrm{OFF}}, N_{1T}^{\mathrm{OFF}}, N_{1R}^{\mathrm{OFF}}, N_{1B}^{\mathrm{OFF}}, \dots, N_{MT}^{\mathrm{OFF}}, N_{MR}^{\mathrm{OFF}}, N_{MB}^{\mathrm{OFF}}]^{\mathrm{T}}$.

In the steady state, the expected inflow and outflow must be the same for all the possible states. As a result, the following set of equations are defined:

$$\begin{pmatrix} \mathbf{P}^{\mathrm{ON}} P_{\mathrm{ON|ON}} & \mathbf{P}^{\mathrm{OFF}} P_{\mathrm{ON|OFF}} \\ \mathbf{P}^{\mathrm{ON}} P_{\mathrm{OFF|ON}} & \mathbf{P}^{\mathrm{OFF}} P_{\mathrm{OFF|OFF}} \end{pmatrix} \begin{pmatrix} \mathbf{N}_{\mathbf{s}}^{\mathrm{ON}} \\ \mathbf{N}_{\mathbf{s}}^{\mathrm{OFF}} \end{pmatrix} = \begin{pmatrix} \mathbf{N}_{\mathbf{s}}^{\mathrm{ON}} \\ \mathbf{N}_{\mathbf{s}}^{\mathrm{OFF}} \end{pmatrix}.$$
(4)

Matrix \mathbf{P}^{ON} is related to the state transition probabilities (i.e., the probabilities $p_{i_X^S,j_X^S}$ of going from state \mathbf{i}_X^S to state \mathbf{j}_X^S in a given slot) for initial states in the ON period, while matrix $\mathbf{P}^{\mathrm{OFF}}$ is related to state transition probabilities for initial states in the OFF period. For initial states in the ON period the state transition probabilities are provided in Table I when the final state is also in the ON period. In the case that the final state is in the OFF period, the state transition probabilities would be the same as in Table I simply by substituting $P_{\mathrm{ON}|\mathrm{ON}}$ by $P_{\mathrm{OFF}|\mathrm{ON}}$. In this table, $\Delta(i)$ is the probability of a message arriving with more packets than the maximum number allowed by the buffer capacity M when the buffer contains i packets (i.e., arriving M+1-i packets or more). When such a situation occurs, the whole message is lost.

Regarding $\mathbf{P}^{\mathrm{OFF}}$, it is obtained when substituting in \mathbf{P}^{ON} the values $P(0)=1,\,\Delta(0)=0,\,$ and $P(i)=0,\,\Delta(i)=0$ for $i\neq 0$, meaning that no new messages are generated in the OFF state.

In (4) there are $2 \cdot (3M+1)$ linearly dependent equations and hence, the first row can be suppressed, which leads to another set of equations given by

$$\mathbf{b}N_o^{\mathrm{ON}} + \mathbf{BN} = \mathbf{N} \tag{5}$$

B being a square matrix $(6M+1)\times(6M+1)$, **b** a column vector with (6M+1) rows, and vector **N** is obtained by removing the first component of vector $[\mathbf{N_s^{ON}}^T, \mathbf{N_s^{OFF}}^T]^T$.

The different state transition probabilities depend on the number of simultaneous users n_a in the system. Its value in the steady state can be related with the rest of variables by means of

$$n_a + \mathbf{a}^{\mathbf{T}} \mathbf{N} = 0 \tag{6}$$

where

$$\mathbf{a}^{\mathbf{T}} = \left[\underbrace{-1, -1, -p_b, \dots, -1, -1, -p_b}_{3M \text{ rows}}, 0, \underbrace{-1, -1, -p_b, \dots, -1, -1, -p_b}_{3M \text{ rows}} \right]. \tag{7}$$

TABLE I
STATE TRANSITION PROBABILITIES FOR STATES IN THE ON PERIOD

ī
7
ON
ON

Similarly, by considering the total number of admitted users in the system U and ${\bf u}$ a single column vector formed by (6M+1) "1"s

$$N_o^{\text{ON}} = U - \mathbf{u}^{\mathbf{T}} \mathbf{N}. \tag{8}$$

By combining the last three equations the following relationship can be obtained where the only unknown parameter is n_a :

$$n_a = \mathbf{a}^{\mathbf{T}} \left((\mathbf{B} - \mathbf{I}) - \mathbf{b} \mathbf{u}^{\mathbf{T}} \right)^{-1} \mathbf{b} U.$$
 (9)

After numerically solving this equation in terms of n_a the values of $N_o^{\rm ON}$ and ${\bf N}$ can be derived, and then parameters that deter-

$$p_{\text{okr}} = \frac{\sum_{i=1}^{t} \sum_{j=0}^{t-i} {L \choose j} {L \choose i} p_e^{j+i} (1 - p_e)^{2L - j - i} + \sum_{i=t+1}^{L} {L \choose i} p_e^{i} (1 - p_e)^{2L - i}}{1 - (1 - p_e)^L}$$
(3)

mine system performance can be calculated. For instance, the throughput or number of correctly transmitted packets in each slot will be given by

$$S = \sum_{i=1}^{M} \left(p_{\text{okt}} \left(N_{iT}^{\text{ON}} + N_{iT}^{\text{OFF}} \right) + p_{\text{okr}} \left(N_{iR}^{\text{ON}} + N_{iR}^{\text{OFF}} \right) + p_{b} p_{\text{okr}} \left(N_{iR}^{\text{ON}} + N_{iR}^{\text{OFF}} \right) \right).$$
(10)

The message loss probability, which is the probability of a message arriving containing more packets than the maximum number allowed by the buffer size, is

$$P_{L} = \frac{N_{o}^{\text{ON}} \Delta(0) + \sum_{i=1}^{M} \left(N_{iT}^{\text{ON}} + N_{iR}^{\text{ON}} + N_{iB}^{\text{ON}}\right) \Delta(i)}{(1 - P_{T}) \frac{\beta_{\text{ON}}}{\beta_{\text{ON}} + \beta_{\text{OFF}}} U}.$$
(11)

Denoting x the service time of a packet (i.e., the time between its first transmission and the instant when the packet can be decoded correctly), when a g-packet message arrives in the buffer and it is accepted (i.e., the buffer has enough room for all of its packets), its total delay will have to include the service time for all of its packets, the service time for the packets that are in the buffer and whose service time has not begun yet (N_q) , and the residual service time for the packet that is currently being transmitted ([8]). As a result, the average message delay is given by

$$D_{m} = E[x]E[g] + E[x]E[N_{q}] + (1 - P_{T})(1 - P_{L})$$

$$\times \frac{\beta_{\text{ON}}}{\beta_{\text{ON}} + \beta_{\text{OFF}}} \frac{E[x^{2}]}{2}$$
(12)

where

$$E[x] = p_{\text{okt}} + 2(1 - p_{\text{okt}})p_{\text{okr}} + (1 - p_{\text{okt}})(1 - p_{\text{okr}})\left(2 + \frac{1}{p_b p_{\text{okr}}}\right)$$
(13)

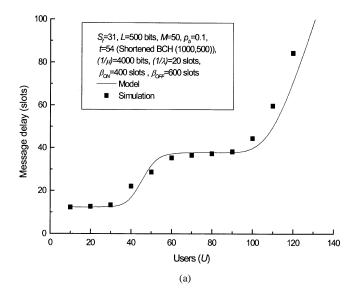
$$E[x^2] = p_{\text{okt}} + 4(1 - p_{\text{okt}})p_{\text{okr}} + (1 - p_{\text{okt}})(1 - p_{\text{okr}}) \times \left(4 + \frac{3}{p_b p_{\text{okr}}} + \frac{2}{(p_b p_{\text{okr}})^2}\right)$$
(14)

$$E[N_q] = \frac{\sum_{i=1}^{M} (i-1) \left(N_{iT}^{\text{ON}} + N_{iR}^{\text{ON}} + N_{iB}^{\text{ON}} \right)}{N_o^{\text{ON}} + \sum_{i=1}^{M} \left(N_{iT}^{\text{ON}} + N_{iR}^{\text{ON}} + N_{iB}^{\text{ON}} \right)}.$$
 (15)

The mean number of packets in a message that is accepted in the buffer, E[g], is the quotient between the mean number of packets N_{TP} and messages N_{TM} that arrive in a given slot, as shown in (16) at the bottom of the page.

IV. MODEL VALIDATION

The proposed model has been validated through computer simulations. They consider a number of users distributed in a single cell that generate messages according to the specified traffic model. Interference is modeled with the SIGA method



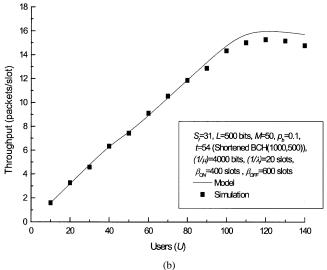


Fig. 2. Comparison of (a) simulated and theoretical message delay and (b) throughput.

[6], neglecting thermal noise. Perfect power control is assumed to model the link layer and propagation and compute those packets that are successfully received. The described type II hybrid ARQ strategy is applied to perform retransmissions until the packets are successfully recovered.

Fig. 2(a) relates the mean message delay with the total number of users in the system for a given arrival rate. Three predominant regions are observed. First of all, when the number of users is small, total system interference is low, so the packets are correctly received after the first transmission and thus the mean message delay equals the average number of packets per message. As the number of users increases, interference grows, and it becomes necessary to send the redundancy for each packet, thus requiring two slots for a successful transmission.

$$E[g] = \frac{N_{TP}}{N_{TM}} = \frac{N_o^{\text{ON}} \sum_{j=1}^{M} jP(j) + \sum_{i=1}^{M} \sum_{j=1}^{M-i} jP(j) \left(N_{iT}^{\text{ON}} + N_{iR}^{\text{ON}} + N_{iB}^{\text{ON}}\right)}{N_o^{\text{ON}} \left(1 - P_T - \Delta(0)\right) + \sum_{i=1}^{M} \left(N_{iT}^{\text{ON}} + N_{iR}^{\text{ON}} + N_{iR}^{\text{ON}}\right) \left(1 - P_T - \Delta(i)\right)}$$
(16)

This corresponds to the second region, where the delay is approximately twice that for a low number of users. Finally, when the number of users is large, the redundancy may not be enough to correct all the errors, and therefore, the delay increases due to an increase in the number of required retransmissions. A good agreement between model and simulations can be observed, with only minor discrepancies in the transitions between the three regions described above, due to the considered approximation to suppress the dependence between the number of states and the correcting capability. The same good agreement has been observed with respect to other measurements from the model such as the throughput in Fig. 2(b) or in results with different parameters, not shown for the sake of brevity.

Another one of the benefits of the proposed modeling relays on detecting those situations in which, depending on system parameters, the system may exhibit multiple equilibrium points. This situation occurs whenever (9) has more than a single solution: this means that the system can stay around one point for a random length of time and then move and stay around another point. Nevertheless, whenever such a situation occurs, only one of the stable equilibrium points corresponds to a good throughput value, while the others correspond to low throughput. As there is no way of controlling when the system will be in the vicinity of each point, this situation with multiple equilibrium points is undesirable. Whether the system exhibits bistable behavior or not depends on the specific values of all the system parameters $(p_b, offered load, U, M, ...)$ and the simplest way to find these regions is by analyzing the behavior of (9) with different parameter values. For example, in Fig. 3 we present the region of bistability depending on p_b and the offered load (offered packets/user/slot) for the case U=50with a buffer length M=50 and for traffic sources that remain always in the ON state. It can be observed that this region occurs for small values of the offered load and high values of p_b .

V. CONCLUSION

An ARQ-II strategy has been studied in a slotted DS-CDMA packet transmission environment. A Markov modeling strategy based on the EPA technique has been presented that can predict the behavior of such a system and quite an accurate match between model and simulations has been shown. This model

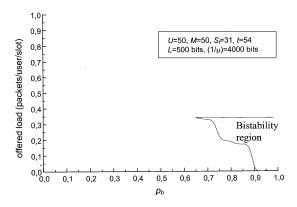


Fig. 3. Region of bistability for the case $U=50,\,M=50$ as a function of the offered load and probability $p_{\rm b}.$

allows the consideration of message oriented statistics with an ON/OFF traffic modeling with Poisson arrivals during the ON periods. Similarly, it allows the determination of those combinations of parameters that can lead the system to an undesirable situation with multiple equilibrium points.

REFERENCES

- S. Lin and P. S. Yu, "A hybrid ARQ scheme with parity retransmission for error control of satellite channels," *IEEE Trans. Commun.*, vol. COM-30, pp. 1701–1719, July 1982.
- [2] Y. M. Wang and S. Lin, "A modified selective-repeat type-II hybrid ARQ system and its performance analysis," *IEEE Trans. Commun.*, vol. COM-31, pp. 593–607, May 1983.
- [3] J. M. Hanratty and G. L. Stüber, "Performance analysis of hybrid ARQ protocols in a slotted direct sequence code-division multiple-access network: jamming analysis," *IEEE J. Select. Areas Commun.*, vol. 8, pp. 562–579, May 1990.
- [4] Q. Zhang, T. F. Wong, and J. S. Lehnert, "Stability of a type-II hybrid ARQ protocol for slotted DS-SSMA packet radio systems," in *Proc. IEEE INFOCOM'98*, vol. 3, San Francisco, CA, Apr. 1998, pp. 1301–1308.
- [5] —, "Performance of a type-II hybrid ARQ protocol in slotted DS-SSMA packet radio systems," *IEEE Trans. Commun.*, vol. 47, pp. 281–290, Feb. 1999.
- [6] R. K. Morrow, Jr., "Accurate CDMA BER calculations with low computational complexity," *IEEE Trans. Commun.*, vol. 46, pp. 1413–1417, Nov. 1998.
- [7] S. Lin and D. J. Costello, Error Control Coding: Fundamentals and Applications. Englewood Cliffs, NJ: Prentice-Hall, 1983.
- [8] L. Kleinrock, Queueing Systems. Volume I: Theory. New York: Wiley, 1975.