A Point-to-Multipoint ARQ Scheme with Multicopy Retransmission for High Speed Satellite Communications

Heliomar M. de Lima
Banco do Brasil S. A.
Rio de Janeiro - RJ - Brasil
e-mail: heliomar@coe.ufrj.br

Otto Carlos M. B. Duarte
UFRJ - COPPE/EE - PEE
Rio de Janeiro - RJ - Brasil
e-mail: otto@coe.ufrj.br

Abstract

This work presents and analyses a point-to-multipoint GB(N) ARQ protocol with multicopy retransmission that provides buffering memory at the receivers. Its key ideas is to retransmit m copies ($m \geq 2$) of all negatively acknowledged messages as well as to store the good messages following an erroneously received one. It avoids buffer overflows, reduces the effect of the round trip delay and improves significantly the throughput efficiency in high speed environments. The analysis uses a time-discrete semi-Markov process model. Numerical results show that the proposed strategy presents better performance than all the other known point-to-multipoint GB(N) schemes. Furthermore, the throughput efficiency remains in an usable range even for very high error rate conditions where it outperforms many finite buffer selective repeat schemes.

1 Introduction

In addition to having the important ability to provide point-to-multipoint connections with several geographically dispersed destination stations, satellite links provide bandwidth of various Megabits/second and are of most advantage for distributed systems over difficult terrain such as mountains, large forests or oceans, or where terrestrial communications facilities are not available or too expensive.

The availability of high speed links with broadcast capacities opens up the possibility of new range of point-to-multipoint applications, such as multimedia conference systems, broadcasting of information, document distribution and communication for distributed systems. However, in high speed environments, as the transmission speed increases the protocol processing overhead at the stations and network nodes becomes a performance bottleneck. An approach to overcome this problem is the design of new protocols with mechanisms specifically conceived for high bandwidth communications, such as simple error recovery procedures.

Automatic Repeat Request (ARQ) schemes are widely used guarantee a reliable end-to-end transfer when the network residual error rate does not satisfy the quality of service required by the application. Several different ARQ strategies have been proposed and are normally classified in three basic types: stop-and-wait (SW), go-back-N (GB(N)) and selective-repeat (SR). SW is the easiest to implement but it is not practical in high speed satellite environments because it allows exactly one message to be transmitted at a time and to remain unacknowledged. GB(N) and SR schemes do not impose this restriction and present increasing both implementation complexity and throughput efficiency.

Different versions of ARQ GB(N) [1-3] and SR schemes [4, 5] have been evaluated for point-to-multipoint communications. The last two schemes [4, 5] use the key idea of repeated transmissions, introduced by Sastry [6]. Further, some schemes [2-4] have increased the receiver capabilities, with respect to the GB(N) basic one, providing a buffering memory so that it can store and acknowledge the errorless non-sequenced messages and reorder the retransmitted ones.

The throughput of strategies in which the emitter always transmits m copies ($m > 1$) of each message, even at the first time, is penalized when the channel presents very good error rate, since its maximum efficiency is $1/m$. On the other hand, the drawback of the recovery strategies that transmit (retransmit) continuously repeated copies of the same message is the bandwidth waste of S messages for each transmitted (retransmitted) message, where S is the number of messages that can be transmitted during a round trip delay. Weldon has introduced the key idea of sending a few copies (and not repeating continuously) of a message only after being negatively acknowledged in his SR scheme [7], which has extended for point-to-multipoint environments by Towsley&Mithal [5].

In order to face the high speed processing time bottleneck, this paper looks for simple protocols, without correcting codes. Thus, it presents a point-to-multipoint GB(N) recovery strategy with receiver buffering memory and multicopy retransmission where the additional copies are selectively retransmitted. The proposed protocol, an extension of the point-to-point scheme proposed in [8], does not consume too much resources and is well suited to high speed satellite environment where the channel is good most of the time and noisy only occasionally.

First, the analytical model, based on discrete-time semi-Markov process, and the environment are presented. The following section describes the proposed recovery strategy and analyses its throughput efficiency. Numerical results are provided and compared with other ARQ schemes. Finally, the conclusions are presented.

2 Analytical Model

Throughout this paper it is assumed an environment consisting in one transmitter and M receivers, where all data is transmitted in form of fixed length *information messages* over a point-to-multipoint channel. Each information message, which has L total bits with K useful ones, includes a *cyclic redundancy code* which enables each receiver to detect transmission errors and a *sequence number* which uniquely identifies the message and its position in the transmission sequence. The L bits message transmission time is defined as the time unit.

The probability of an information message arriving with at least one bit error at a given receiver is denoted by P_{be} ($P_{bc} = 1 - P_{be}$). This probability is independent from message to message and from receiver to receiver. Moreover, the round trip delay, measured in time units, is constant and equal to S for all receivers. It is also assumed that there is always an information message waiting to be transmitted at the transmitter.

Furthermore, this work makes the following assumptions: control messages (acknowledgments) are always received without errors (the return channel is noiseless); transmitter and receivers require a $S + 1$ message buffer; all erroneous messages are detected; the processing time at transmitter and receivers is negligible; the modulus sequence number is sufficiently large to allow continuous transmission of messages and messages are identified at the instance of arrival, even the erroneous ones.

The protocol throughput efficiency is defined as the ratio between the number of information bits communicated to the set of receivers and the total number of transmitted and retransmitted bits (including overhead bits). It will be obtained by assigning states to the set of receivers according to the recovery process conditions in at least one receiver and, for each state, computing the throughput efficiency and the probability of the set of receivers being at that state. The final protocol throughput efficiency is the weighted mean value of the throughput efficiency for each state, where the weights are the calculated probabilities. In order to determine these probabilities, the set of receivers behavior is modeled as a discrete-time semi-Markov process, whose state sequence is: $X(1), X(2), \ldots, X(n), \ldots$ where $X(n), n = 1, 2, \ldots$, is a random variable denoting the receiver state after the reception of the n^{th} message.

3 The Proposed Scheme

The proposed strategy, an extension of the point-to-point protocol proposed and evaluated in [8], is a GB(N) scheme supposed to be used in a point-to-multipoint high speed environment characterized by a good bit error rate during most of the time and that occasionally presents a noisy condition. Thus, the scheme is a conventional point-to-multipoint GB(N) with a modified recovery procedure in order to achieve an improved throughput performance in the high bit error rate region. In addition to the classical $(S+1)$-messages retransmission buffer, the scheme requires an equal size buffer at the receivers. Each receiver processes and stores, for posterior user delivery, all the out of sequence messages correctly received during an error recovery procedure. Further, the transmitter uses the basic idea of the Weldon's recovery scheme retransmitting m copies ($m \geq 2$) of all negatively acknowledged messages. In other words, the proposed scheme is a GB(N) protocol with multicopy retransmission in which the additional copies are selectively retransmitted.

The reason to use the go-back-N technique is to reduce the processing overhead. At the same time, the message storage at the receivers avoids the great number of correctly received messages (up to S ones) that are discarded during a conventional GB(N) recovery and improves the scheme performance in the high bit error rate region, because it eliminates the need for successful retransmission of the messages already correctly received and stored at all receivers. The multicopy retransmission aims at reducing the total number of go-back-N retransmissions.

The scheme requires two types of messages. Information messages that are sequentially numbered and control messages. Through control messages each receiver acknowledges the information messages. The control messages provide two acknowledgment fields. The first one is a positive and accumulative acknowledgment that contains a number $N(R)$ indicating the good reception of all messages numbered until $N(R)-1$. The second one (B_{e_i}) is an individual negative acknowledgment that indicates an erroneous reception of the message numbered B_{e_i}. Hereafter in this work, when this field is activate the control message will be called REJ B_{e_i} (*reject*).

Each receiver begins an error recovery procedure when it detects an erroneous message. It sends a retransmission request, reserves a slot in its buffer and continues to receive, process and acknowledge the messages sent in advance. The messages correctly received are stored for posterior delivery while a buffer slot is reserved for each erroneous detected message. The transmitter orders the messages B_{e_i} retransmission requests taking into account the set of receivers as a whole. That is, the message B_{e_1} may correspond to a slot reserved at the receiver j while the message B_{e_2} may correspond to a slot reserved at receiver k (possibly the first slot in this receiver). Therefore, after the REJ B_{e_0} sending, the first message out of those sent in advance detected in error by one or more receivers will be called message B_{e_1}, regardless which and how many receivers have received it in error. The second will be called message B_{e_2}. And so on. An *exception condition* exists in the set of receivers from the time the REJ B_{e_0} is sent to the time all the messages $B_{e_0}, B_{e_0}+1, \ldots, B_{e_0}+S$ are communicated to all receivers. During a recovery procedure, the receivers wait for m erroneous copies of the messages corresponding to a slot in its receiving buffer (messages B_{e_i}) before resending a REJ B_{e_i}.

The transmitter sends sequenced messages and stores them in a retransmission buffer while waiting for their positive

acknowledgment from all receivers. In order to ensure that every message is correctly received by all receivers, the transmitter maintains a state variable $V_i(R)$ for each receiver, indicating the number of the last message acknowledged positively by this receiver. Hence, $V_i(R) = N(R)$ indicates that the receiver i has correctly received all messages with sequence number smaller than N(R). Besides these variables, the transmitter maintains two global state variables: $VG(S)$ that indicates the number of the next message to be transmitted and $VG(R) = \min_i\{V_i(R)\}$ that indicates the last message communicated to all receivers. Thus, $VG(R) = N(R)$ indicates that the messages with sequence number smaller than N(R) have been communicated to the set of receivers and that, therefore, may be removed from retransmission buffer.

On the receipt of a REJ B_{e_i}, the transmitter compares the fields $N(R)$ and B_{e_i}. If $N(R)$ is equal to B_{e_i}, it goes back and retransmits continuously the retransmission buffer, resending m copies of each negatively acknowledged message (messages B_{e_i}). Hereafter this retransmission will be called B_{e_i} *go-back-N retransmission*. Otherwise, if B_{e_i} is greater than $N(R)$, this information indicates the message B_{e_i} must be repeated m times, without causing an additional "go-back-N procedure". Therefore, upon receipt of REJ B_{e_0} ($N(R)=B_{e_0}$), the transmitter accomplishes a B_{e_0} go-back-N retransmission resending m copies of the messages $B_{e_0}, B_{e_1}, \ldots$. This work considers that a go-back-N retransmission begins at the B_{e_0} retransmission instance and ends after the retransmission of the message $B_{e_0} + S$.

During a go-back-N retransmission, the reception of a REJ B_{e_j} with $N(R)=B_{e_j}$ and $j \geq i$ (from other receiver) does not causes an additional go-back-N retransmission. The number N(R) contained in the REJ B_{e_i} messages received in this interval will only be used to updated the variable $V_i(R)$. At the same time, the transmitter uses the number B_{e_i} to retransmit additional copies of the message B_{e_i}. After resending additional copies of a message B_{e_i}, the transmitter disregards additional messages REJ B_{e_i} received from other receivers in the same go-back-N retransmission.

Performance Analysis

The proposed strategy defines two states for the set of receivers: normal and exceptional states. In the normal state, the receivers accept the sequenced correctly received messages and deliver them to their users. There is no messages stored in any receiver buffer and no outstanding error recovery in this state. The set of receiver turns to the exceptional state when an erroneous message is received (message B_{e_0}) and remains in this state while there is an erroneous message to be recovered. Thus, the set of receivers is in the exceptional state when an exceptional condition exists in at least one receiver. It is defined that the transition occurs at the beginning of this message reception.

When the set of receiver goes to the exceptional state, it begins the recovery procedure that it will be classified in cycles. The exceptional state has one *transmission cycle* and one or more *retransmission cycles*. The transmission cycle corresponds to the processing and storing of messages sent in advance. In this cycle each receiver can store up to S+1 messages (including the slots reserved for messages B_{e_i}). Hence, the set of receivers continues to process the messages numbered $B_{e_0}+1, \ldots, B_{e_0}+S$. The number of retransmission cycles corresponds to the number of go-back-N retransmissions needed to recover all messages $B_{e_i}, 0 \leq i < S$. Thus, the set of receivers only returns to the normal state after recovering all these messages B_{e_i}. Each retransmission cycle ends after the retransmission of the message numbered $B_{e_0}+S$.

During the exceptional state, the set of receivers stores only the messages corresponding to the transmission cycle (messages with sequence number between B_{e_0} and $B_{e_0} + S$). Thus, during this state $S + 1$ messages are always communicated to set of receivers. After processing and storing the messages numbered $B_{e_0}, B_{e_0}+1, \ldots, B_{e_0}+S$ (during the transmission cycle), the set of receivers waits for the retransmissions, asks for new retransmissions and goes to the normal state when all messages $B_{e_i}, i < S$ are recovered.

The analysis considers an *inferior scheme* of the protocol described in the preceding subsection. In the proposed scheme when the message B_{e_0} is recovered and still remains one or more unrecovered messages B_{e_i}, the set of receivers sends a REJ B_{e_i} requesting a go-back-N retransmission beginning with the message B_{e_i} and delivers to the users the sequenced messages stored in the receiving buffers. In this analysis, if the message B_{e_0} is recovered in a retransmission cycle, the corresponding released buffer space will not be taken into account. In the following, the throughput efficiency of the this scheme will be derived. As the set of receivers passes from normal to exceptional state at the beginning of the erroneous message reception, only error-free messages are received in the normal state. Then, the throughput efficiency of the protocol in this state is K/L.

As $S + 1$ messages are always communicated during the exceptional state and the total protocol efficiency is the weighted mean value of the efficiency in each state:

$$\eta = \frac{K}{L}\left(P(N) + P(E)\frac{S+1}{E[t_E]}\right) \qquad (1)$$

where $P(N)$ and $P(E)$ are the probabilities of the set of receivers being in normal and exceptional states, respectively. The expression $(S + 1)/E[t_E]$ is the efficiency for the exceptional state and $E[t_E]$ is the mean waiting time at the exceptional state.

Applying the semi-Markov process properties to calculate $P(N)$ and $P(E)$ and combining them with equation 1 the throughput efficiency can be expressed as:

$$\eta = \frac{K}{L}\frac{E[t_N] + S + 1}{E[t_N] + E[t_E]} \qquad (2)$$

where $E[t_N]$, the mean waiting time in normal state, is:

$$E[t_N] = \sum_{i=0}^{\infty} i.(1 - P_{bc}^M).(P_{bc}^M)^i = \frac{P_{bc}^M}{1 - P_{bc}^M}\,. \qquad (3)$$

Substituting this mean waiting time (equation 3) in equation 2, the throughput efficiency expression becomes:

$$\eta = \frac{K}{L}\frac{P_{bc}^M + (1-P_{bc}^M).(S+1)}{P_{bc}^M + (1-P_{bc}^M).E[t_E]} \ . \qquad (4)$$

The mean waiting time at exceptional state is given by:

$$E[t_E] = S + 1 + E[n_e](S+1) + E[t_a], \qquad (5)$$

where $E[n_e]$ is the expected number of retransmission cycles and $E[t_a]$ is the additional mean waiting time due to the additional copies of messages B_{e_i}. The average number of exceptional state retransmission cycles is given by:

$$E[n_e] = \sum_{j=1}^{\infty} j.Pr\{n_i = j\}, \qquad (6)$$

where $Pr\{n_i = j\}$ is the probability that all $S+1$ messages corresponding to the transmission cycle are successfully communicated to all M receivers during the j^{th} retransmission cycle, but not all recovered within the $(j-1)^{th}$ retransmission cycle, which is given by:

$$Pr\{n_i = j\} = Pr\{n_i \leq j\} - Pr\{n_i \leq j-1\} \qquad (7)$$

where

$$Pr\{n_i \leq j\} = (1-P_{be}^{mj+1})^{MS}\frac{(1-P_{be}^{mj+1})^M - P_{bc}^M}{1-P_{bc}^M} \ . \qquad (8)$$

Next, the expression for $E[t_a]$, the mean additional waiting time due to additional copies of messages B_{e_i}, is derived. To calculate $E[t_a]$, it is necessary to compute the additional waiting time due to additional copies of each message. Each message $B_{e_i}(i \geq 0)$ increases in $m-1$ time units the waiting time of the first retransmission cycle. In following cycles, each message not recovered in the previous cycles increases this waiting time in up to $m-1$ time units. Thus, the mean waiting time due to the additional copies of message B_{e_i} is:

$$E[t_{a_i}] \leq (m-1).E[n_{REJB_{e_i}}]\sum_{j=i}^{S}\binom{S}{j}(1-P_{bc}^M)^j\,(P_{bc}^M)^{S-j} \qquad (9)$$

where $E[n_{REJB_{e_i}}]$ is the average number of retransmission cycles until message B_{e_i} is successfully recovered. It corresponds to average number of REJ B_{e_i} answered by the transmitter until recovering this message (with B_{ei} equal to $N(R)$ or not). It is given by:

$$E[n_{REJB_{e_i}}] = \sum_{j=1}^{\infty} j\frac{(1-P_{be}^{m.j+1})^M - (1-P_{be}^{m.(j-1)+1})^M}{1-P_{bc}^M}. \qquad (10)$$

Finally, the mean additional waiting time due to the additional copies of all messages B_{e_i} is

$$E[t_a] = (m-1)E[n_{REJB_{e_i}}]\sum_{j=0}^{S}\sum_{i=j}^{S}\binom{S}{i}(1-P_{bc}^M)^i\,(P_{bc}^M)^{S-i} \ . \qquad (11)$$

Therefore, combining equations 11, 6, 5 and 4, a *lower bound* on the throughput efficiency of the proposed scheme is obtained.

4 Numerical Results

This section discusses numerical results of the proposed protocol and compares them with other published results. Fig. 1 provides the relationship between the throughput efficiency and bit error rate for various values of $m = 1, 2, 3, 5$ and 10 considering 5 receivers ($M = 5$), with message length of 6800 bits ($L = 6800$) over a satellite link at 10 Mbps ($S = 1024$). These figure provides an envelope of optimum performance that specifies the values of m to use for a given bit error rate. There is a narrow range of bit error rates in which a given value of m is better than the other ones. Nevertheless, the optimum performance envelope of the multicopy retransmission improves the throughput efficiency for a large range of bit error rates (BER). It can be observed an efficiency decrease for greater m values in the region of high efficiency value. This can be explained considering that a given m value may be too much for a low BER region and not enough for a high BER region. Thus, the improvement due to greater m values occurs in a region of higher BER, where the throughput efficiency can be low and of no practical interest. Therefore, the results suggest the use of low m values.

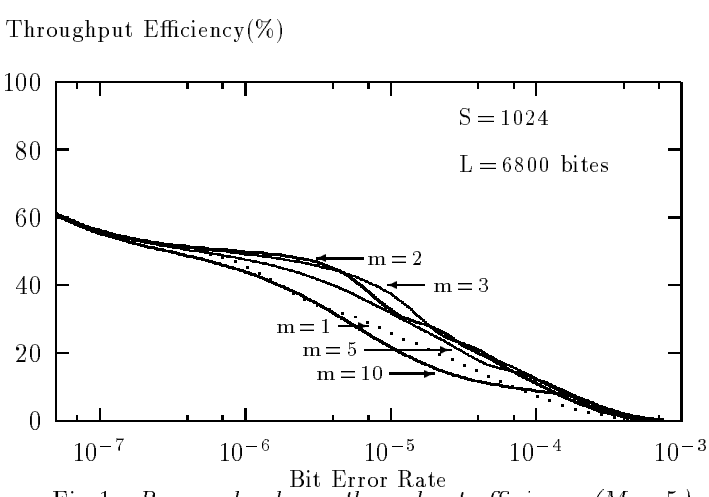

Fig.1 - *Proposed scheme throughput efficiency ($M = 5$)*

Fig. 2 compares the throughput efficiency versus M (number of receivers) of the new scheme ($m = 2$) and those of Gopal&Jaffe's GB(N) [1] (*full memory* and *limited memory*), Towsley's GB(N) [2] with blocks of 1024 ($F = 1024$) messages, Chandran&Lin's SR [4], Towsley&Mithal's SR [5] with $q = 1$ and one copy retransmission ($n_1 = 1$) for bit error rate of 10^{-6}. In the Towsley&Mithal's SR scheme, the transmitter requires a buffer capacity for $2.S + 1$ messages and the receivers, a capacity for $S+1$ ones. Towsley's GB(N) scheme requires buffers for a whole block (F messages) at the receivers and for $F + S$ messages at transmitter. It is is a block oriented GB(N) with selective retransmission: in each block, only the messages not correctly received are retransmitted. Besides outperforming other GB(N) protocols, the proposed scheme performs better than Towsley&Mithal's SR for $M > 27$.

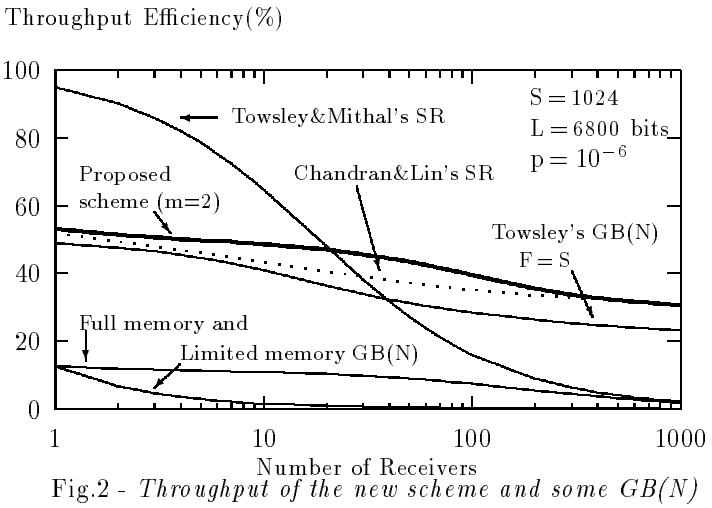

Fig.2 - *Throughput of the new scheme and some GB(N) and SR protocols*

Fig. 3 compares the throughput efficiency versus bit error rate of the above protocols over a point-to-multipoint link with 5 receivers ($M = 5$), message length of 6800 bits ($L = 6800$) and $S = 1024$. Towsley's scheme and Gopal&Jaffe's *full memory* one are the best previous known GB(N) results for point-to-multipoint communications. Thus, the proposed scheme outperforms other GB(N) schemes for all bit error rates. It can be also observed that it outperforms the SR scheme in [4] and that for high error rates, it performs better than the Towsley&Mithal's SR scheme.

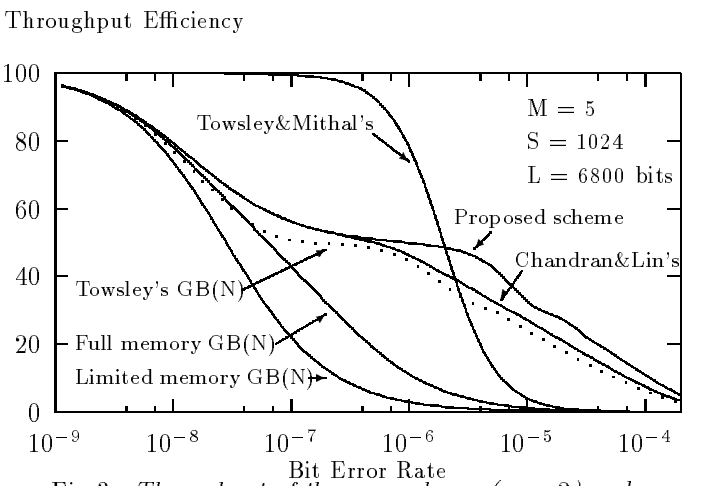

Fig.3 - *Throughput of the new scheme ($m = 2$) and some GB(N) and SR protocols*

5 Conclusion

This paper has presented and analyzed a point-to-multipoint GB(N) scheme with buffers at receivers and multicopy retransmissions where the additional copies are selectively retransmitted. The receivers store the good messages following an erroneously received one and the emitter retransmits m copies ($m \geq 2$) of all negatively acknowledged messages. The proposed strategy makes a best effort to recover the negatively acknowledged messages, reduces the effect of the round trip delay, performs well even under high error rate channel condition and is suitable for high speed satellite environments.

The analytical results show that the new strategy presents a better performance than the best known point-to-multipoint GB(N) results and suggest low values for the number of repeated messages to be retransmitted. Furthermore, the proposed strategy outperforms two finite buffer selective-repeat ARQ schemes over a wide range of middle and high bit error rates and/or large number of receivers. At the same time, for certain condition of bit error rate and number of receivers its performance is close to the best possible one achievable by an ARQ protocol.

References

[1] I. S. Gopal and J. M. Jaffe, "Point-to-multipoint communication over broadcast links," *IEEE Transactions on Communications*, vol. COM-32, pp. 1034–1044, Sep. 1984.

[2] D. Towsley, "An analysis of a point-to-multipoint channel using a go-back-n error control protocol," *IEEE Transactions on Communications*, vol. COM-33, pp. 282–288, March 1985.

[3] H. M. de Lima and O. C. Duarte, "An improved GB(N) ARQ scheme for point-to-multipoint high speed satellite communications," in *IEEE International Symposium on Circuits and Systems ISCAS'94*, (London), May 1994.

[4] S. R. Chandran and S. Lin, "A selective repeat ARQ scheme for point-to-multipoint communications and its throughput analysis," *ACM Computer Communications Review*, vol. 16, pp. 292–301, Aug. 1986.

[5] D. Towsley and S. Mithal, "A selective repeat ARQ protocol for a point to multipoint channel," in *IEEE International Conference on Communications INFOCOM'87*, (San Francisco,CA), pp. 521–526, Apr. 1987.

[6] A. R. K. Sastry, "Improving repeat-request (ARQ) performance on satellite channels under high error rate conditions," *IEEE Transactions on Communications*, vol. COM-23, pp. 436–439, Apr. 1975.

[7] E. Weldon, "An improved selective repeat ARQ strategy," *IEEE Transactions on Communications*, vol. COM-30, pp. 480–486, March 1982.

[8] H. M. de Lima and O. C. Duarte, "A Go-Back-N protocol with multicopy retransmission for high speed satellite communications," in *IEEE International Conference on Communications SUPERCOMM/ICC'94*, (New Orleans), May 1994.