Automatic Retransmission Rather Than Automatic Repeat Request

Abdulaziz Al-Zoman John DeDourek
Bernd Kurz
Faculty of computer science, UNB, P.O. Box 4400,
Fredericton, N.B., E3B 5A3, CANADA
(506) 453-4566, Fax: (506) 453-3566
email: r17g@unb.ca

Abstract

The standard HDLC family of protocols has been designed for low to moderate BERs. But as BER increases frequent losses or corruptions of certain frames will cause extensive timeout recovery. Too many timeouts means low throughput.

A new retransmission scheme is presented which depends on the transmitter to automatically initiate retransmission instead of the conventional way where the receiver automatically requests retransmission. The retransmission strategy is based on two ideas: the receiver can report without any restriction its buffer status by using a multiple acknowledgement frame and the transmitter keeps a record of the transmission order in which frames were sent.

This new approach significantly reduces the use of timeout recoveries, and hence it gives a higher throughput efficiency than the standard and extended HDLC protocols.

1 Introduction

Data communications are becoming widely needed by some of the mobile users. In addition to mobility, cellular radio networks can provide data communication services with some challenges due to their characteristics. The bit error rate (BER) on a cellular radio channel may reach as high as 10^{-1} [1]. In addition to single bit errors, the channel experiences burst errors and dropouts that are due to such phenomena as multipath fading, co-channel interference, handoffs, shadowing, "blank and burst" signaling [2, 3].

Transmission error control can be divided into two basic mechanisms: automatic repeat request (ARQ) and forward error correction (FEC). A combination of both mechanisms (called hybrid ARQ) is also encountered.

In practice, there are two types of ARQ: stop-and-wait (SW) ARQ and continuous ARQ, which employs either a go-back-n (GBN) or selective repeat (SR) retransmission techniques. The SW schemes are very simple but inefficient because of the idle time involved in waiting for the acknowledgement. The idle time is reduced in the continuous ARQ schemes since data blocks (frames) are continuously sent without waiting for the acknowledgement. A NAK (negative-acknowledgement) in GBN forces the transmitter to stop sending new frames, back up, and retransmit frames starting from the erroneous frame. But in the case of SR, a NAK causes only the erroneous frame to be transmitted. Therefore (in many cases), GBN exhibits a lower performance compared with SR, but has a simpler implementation where complex buffer management is not needed.

The error control is provided by the data link level to ensure error-free transmission of information blocks. HDLC [4] is an international standard protocol which defines the operational functions of the second layer of the OSI model. It has been used as a foundation for the development of a number of other widely used data link protocols, such as LAPB, LAPD, etc [5]. The focus of this paper is on the throughput efficiency of a point-to-point data link during the interchange of information. It is defined as the ratio of the time that would be taken to transmit the whole data (without any overhead involved such as control information, retransmission, timeouts, idle time due to buffering, etc.) to the actual time that is taken to correctly transfer the data. It is also assumed that data flow is unidirectional, i.e., the receiver does not transmits data blocks.

The HDLC family of protocols has been designed for low to moderate BERs. However, when BER increases, frequent losses or corruptions of certain frames cause extensive timeout recovery. Too many timeouts means low throughput. It is known that timeout recovery is only an emergency solution and it should not be used extensively.

Therefore, this paper addresses the problem of sensitivity to the loss or corruption of certain frames that causes the extensive usage of timeout recovery. It proposes a new approach that depends on the transmitter to automatically initiate retransmissions instead of the conventional way where the receiver is the one automatically requesting retransmission.

This paper is structured as follows. Section 2 provides the basic background of HDLC protocols. Section 3 highlights some of the problems when using the HDLC protocols in harsh error environments such as cellular radio channels. Also, it lists some existing solutions. The simulation method used to evaluate the investigated protocols is presented in Section 4. A new proposed approach is described in Section 5. Finally, Section 6 states some concluding remarks.

2 Basic HDLC Protocol

HDLC [4] defines various frame types, such as information (I) frames which carry user data, and supervisory (S) frames which perform control functions such as acknowledging frames and requesting retransmissions.

Each side of the data link (referred to as station) maintains a send state variable V(S) which denotes the sequence number of the next in-sequence I-frame to be transmitted, and a receive state variable V(R) which denotes the sequence number of the next in-sequence I-frame expected to be received. (Note that V(R) and V(S) of the transmitter and receiver, respectively, are equal to 0 in the case of unidirectional data flow). Each time a frame is formed and ready to be transmitted, its send sequence number N(S) and receive sequence number N(R) are set to the current values of V(S) and V(R), respectively. As part of the control information, an HDLC frame has a final/poll bit for checkpointing recovery.

To detect errors in transmission, frames carry additional information (referred to as the frame check sequence – FCS). Frames with detectable errors in the form of an FCS-error are discarded by the receiver. As well, sequence errors are detected either by the receiver after receiving a correct but out of sequence frame or by the transmitter using checkpointing or timeout.

When the receiver detects a sequence error it sends either an REJ-frame (GBN) or an SREJ-frame (SR). Then, it initiates an exception condition which can only be cleared by receiving the requested frame (i.e., N(S)=V(R)). During the REJ exception condition all frames with N(S) not equal to V(R) are discarded. But during the SREJ exception condition they are kept in a buffer and passed to the higher layer when the exception condition is cleared. Unfortunately, during the exception condition, another REJ or SREJ frame cannot be issued.

This limitation has a negative effect on the performance since if either one of the REJ-frame, the SREJ-frame, or the retransmitted I-frame are lost or corrupted, an idle time will be encountered in the channel until the erroneous frame is recovered by the transmitter through checkpointing or timeouts. This limitation is the main concern of this paper which tries to eliminate the sensitivity to the loss or corruption of certain frames such as S-frames and/or retransmitted frames. The frequent losses and corruptions of these frames cause extensive timeout recovery which is known to be the least efficient way for recovering lost or corrupted frames [6, 7].

3 Harsh Error Environments

Several studies of variations of HDLC have been carried out with a number of assumptions to simplify the analysis. For instance, the backward channel (which carries acknowledgements or supervisory frames (S-frames)) is assumed to be error free [8, 9, 10, 11]. Therefore, the timeout recovery which is caused by the corruption or loss of S-frames is not included in the analysis. This assumption would be acceptable if the channel experiences low error rate and the size of an S-frame is significantly shorter than the size of an I-frame. Nevertheless, under a harsh error environment as experienced by cellular radio channels, this assumption is not valid. Because S-frames, too, will be subjected to errors, this will lead to timeout recoveries.

Another simplification is to assume that errors are independently distributed [8, 12, 9]. This assumption does not completely reflect the error behaviour of any practical channel [13].

Cellular radio channels experience very harsh error behaviours due to such impairments as multipath fading, radio shadowing, co-channel interference, and handoff. A simulation model [2, 6], based on field measurements, has been developed to describe the behaviour of the channel and is used to study the performance of a number of data link protocols. This is described in Section 4.

3.1 Existing Problems

When the error environment becomes very harsh, the performance of the existing HDLC protocols will be degraded. First, these protocols rely on the receiver to detect out of sequence exception conditions and automatically issue a repeat request for the missing frames (hence the name ARQ). But, at the same time, there are some restrictions imposed on the receiver in issuing another repeat request while in the exception condition state even though the previous request or its reply are in error. Second, the existing protocols are sensitive to the loss or corruption of certain frames (e.g., S-frames, frames with the P/F bit set to one, retransmitted frames) which leads to timeout recovery (this is referred henceforth as the sensitivity problem).

3.2 Existing Improvements

A number of modifications to the basic ARQ protocols have been proposed to improve their throughput performance. For instance, early solutions to overcome the sensitivity problem use bit-by-bit majority voting of multiple copies of the same data block[14, 15]. Another approach transmits multiple copies of the same data block before receiving its acknowledgement [16, 17, 18]. These schemes solve the sensitivity problem but add transmission overhead by sending multiple copies even if they are not needed.

Hybrid ARQ schemes have also been considered [19, 20, 21]. In these approaches, some erroneous frames can be corrected but still the sensitivity problem is not resolved. Also, for the current HDLC frame format, FEC is not effective enough because of the bit stuffing technique [1]. Sato at el [1] proposed a scheme that uses different FEC codes depending on the error environment. Furthermore, they use a multiple rejection scheme in which a single frame (MREJ-frame) requests selectively that multiple frames be retransmitted. However, they have not solved the problem of sensitivity. That is, timeout recovery will be used whenever an MREJ-frame is lost or corrupted.

Among the proposed approaches which addresses the problem of excessive timeout recovery, is the work (multi-rejection) done by Brady [7]. He identified cases where issuing additional copies of REJ or SREJ frames may be permitted while in the exception condition. This scheme has improved the performance dramatically, especially in the case of multi-SREJ with a large window size. However, the sensitivity problem is solved only partially. That is, there are some cases that cannot be detected which lead to timeout recovery [6].

Netravali et al [22] proposed a new transport protocol which addressed the sensitivity problem by exchanging control information periodically and frequently. Control information reports condition states with respect to blocks. Each block consists of a group of packets. Hence, the receiver only acknowledges blocks and not individual packets. Timeout recoveries are used to initiate block retransmissions. Therefore, a single erroneous packet within a block causes the retransmission of the whole block. Under harsh error environment this may cause frequent block retransmissions and unnecessary retransmissions for correctly received packets.

4 Simulation Method

A "testbed" (simulation tool) for simulating and evaluating point-to-point data link protocols has been developed by the authors[6]. Several variations of HDLC protocols are implemented in software modules and incorporated into the testbed to study their behaviour and investigate the sources of their low performance over a cellular radio channel model.

The cellular radio channel model is based on modelling the effect of transmission impairments on the digital level of the channel media. The model considers not only single bit errors and burst errors but also dropouts due to total signal loss. Modelling parameters are obtained by field measurements [2].

Through direct observations of the simulated point-to-point data link transactions, a number of throughput degradation sources in the tested protocols over a cellular radio channel model were observed. For example,

1. Sensitivity to the loss or corruption of certain frames which causes the usage of timeout recovery.
2. Extensive usage of timeout recoveries decreases the throughput performance.
3. S-frames are also subject to corruptions or losses.
4. Error-free (out of sequence) received and discarded frames by the receiver have the chance to be in error in the following retransmission.

The following sections describe a proposed solution which has been designed to reduce the number of times that a timeout recovery is used, and hence, it improves the throughput performance. It relies on the transmitter to automatically retransmit erroneous frames. Therefore, it is called automatic retransmission (ART) error control.

5 Automatic Retransmission Scheme

The transmitter uses a counter associated with the transmission order in which I-frames are transmitted. It is called the Frame Transmission Order (FTO) counter. Also, the transmitter uses an array of integers, called *FrameTxOrder*, in which it records the values of the FTO counter for the transmitted (outstanding and not yet acknowledged) I-frames. That is, each time the transmitter sends or re-sends an I-frame i, the FTO counter value is recorded in an entry of *FrameTxOrder[i]* associated with that I-frame.

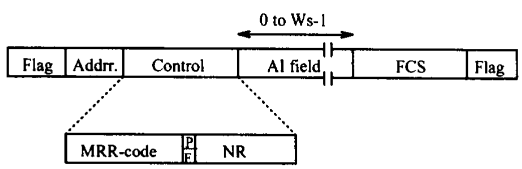

Figure 1: The MRR frame format

The receiver is able to acknowledge all I-frames received using a new multiple acknowledgement scheme. The receiver is assumed to have enough buffer storage to hold at most a window size of data blocks. A multiple acknowledgement frame (referred to as an MRR-frame) consists of two parts: the N(R) and (optionally) the acknowledgement information (AI) fields, shown in Figure 1. The N(R), set to V(R), acknowledges all correctly received I-frames having frame sequence numbers less than N(R). The AI field acknowledges I-frames received out of sequence. AI is a variable length bit pattern indicating which I-frames (with frame sequence numbers higher than V(R)) have been received correctly by the receiver. A "0" is assigned to the bits corresponding to frame sequence numbers that have been received, and a "1" is assigned to bits corresponding to frame sequence numbers that have not been correctly received yet. The AI field reports the reception status of frames with sequence numbers starting from V(R)+1 to the highest received frame sequence number. The maximum length that an AI field may reach is equal to the window size less one. The status of the frame with the sequence number equal to V(R) is not included as part of the AI field report because it would be redundant since the MRR-frame carries an N(R) field set to V(R). The AI field is optional. It is not needed if no frame has been received with a frame sequence number that is higher than V(R).

At the transmitter, when a (multiple) acknowledgement is received, an automatic retransmission is initiated for all frames having FTO less than the (highest) FTO of the frames being acknowledged. That is, the transmitter first marks all the acknowledged frames and notices the highest FTO of these frames. Then it checks the *FrameTxOrder* for outstanding frames with FTO less than the highest FTO of the acknowledged frames. These frames are added to the frame retransmission set. When the transmitter gains the right to send frames it considers the frame retransmission set before sending new frames. Therefore, following this scheme, the transmitter is the one automatically initiating retransmission based on which frames have been received and acknowledged by the receiver.

5.1 Procedure

It is assumed that the transmitter maintains a local state variable, V(A), referring to the oldest outstanding I-frame which has not been acknowledged. That is, V(A) is the lower limit of the transmitter window. Also, it is assumed that the receiver maintains a local state variable, V(H), referring to the highest N(S) of received frames.

It is best to describe the ART procedure by an example. Figure 2 shows an example of ART frame transactions for window size of 8. It is assumed that the example has the following initial values: $V(A) = V(S) = V(R) = V(H) = 10$, and the FTO counter = 40.

Now the transmitter sends frames (10, 11, 12), and records their FTO numbers (40, 41, 42) into the *FrameTxOrder* array. Frame 12 is corrupted or lost. The receiver receives frames (10, 11) correctly and sends "MRR 11" and "MRR 12", respectively. By receiving "MRR 11", the transmitter adjusts its window, continues sending frames (13, 14, 15, 16), and records their FTO numbers (43, 44, 45, 46) in the corresponding entries of *FrameTxOrder*.

By receiving frame 14, the receiver sends "MRR 12 [10]" indicating that all frames with frame numbers less than 12 were received correctly, the next expected frame in-sequence to be received is frame 12, and frame 14 is received correctly.

When the transmitter receives "MRR 12 [10]", it notices that frames (11, 14) are being acknowledged. They have FTO numbers (41, 44) respectively. The highest FTO is 44. Then, the transmitter checks for all outstanding frames with FTO less than 44. It finds frames (12, 13) with FTO number (42, 43). These frame are added to the frame retransmission set. Frames (12, 13) are retransmitted and their new FTO numbers (47, 48) are recorded into the *FrameTxOrder* array.

Every time the receiver receives a frame it sends a (multiple) acknowledgement. Therefore, the task of the receiver includes acknowledging frames that have been received correctly, but does not include explicitly requesting retransmission because it is the task of the transmitter to select frames for retransmission. As can be seen from the example, the retransmission is initiated automatically by the transmitter based on the acknowledgement list reported by a multiple acknowledgement and also by keeping records of the transmission order in which frames were sent. Hence, any received (new) acknowledgement has sufficient information about the receiver buffer status and will assist the transmitter to detect if there are some frames were lost or corrupted.

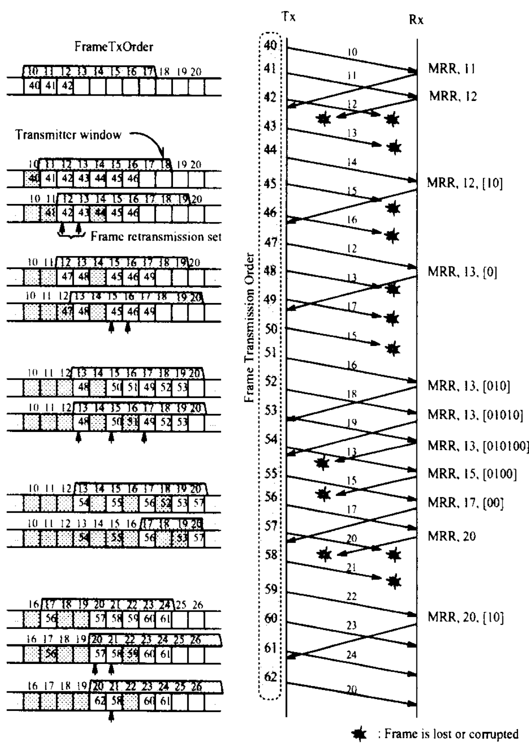

Figure 2: An ART scheme example

Because the (out of sequence) error recovery of the ART scheme does not depend on particular frames, corrupted or lost I or MRR frames do not necessarily cause the occurrence of a timeout recovery. Therefore, the sensitivity problem has been solved to some extent. With this scheme, a frame is not retransmitted unless it is clear that the previous transmission failed. An exception might occur in the case if a timeout. In this case the transmitter initiates retransmission for all outstanding frames. A timeout happens only if the transmitter does not receive any acknowledgement from the receiver for a period of time equal to the time required to transmit a full window of frames. With a large window size and the receiver sending an acknowledgement for each received frame, the usage of timeout recovery becomes very rare.

Name	Comments
REJ	HDLC ARM mode using REJ recovery scheme.
SREJ	HDLC ARM mode using SREJ recovery scheme as described by Brady [7].
SREJ/REJ	HDLC ARM mode using SREJ and REJ recovery schemes.
MultiREJ	HDLC ARM mode using MultiREJ recovery scheme[7].
MultiSREJ	HDLC ARM mode using MultiSREJ recovery scheme[7].

Table 1: List of error recovery schemes under investigation

5.2 Numerical Results

In this section the throughput performance of the ART scheme is compared with the throughput performance of a number of error recovery schemes based on the simulation results. The error recovery schemes under investigations are listed in Table 1.

Figures 3 and 4 depict the throughput and the number of timeouts for the investigated protocols including ART. They are for window sizes of 10, 25 and 60 frames, and one-way propagation delays of 20 and 250 milliseconds.

In general, as the number of timeouts increases the throughput efficiency decreases. For example, referring to Figure 3(a) and (b), the worst throughput efficiency is given by the REJ protocol which has the largest number of timeouts and the best throughput efficiency is given by the ART scheme which has the smallest number of timeouts. However, the number of timeouts is not the only factor in the throughput efficiency, but it has a significant impact. Thus, a protocol with the largest number of timeouts is not necessarily the one with the lowest throughput efficiency. For example the REJ protocol has approximately the largest number of timeouts but it does not give the lowest throughput efficiency as shown in Figure 3(c) and (d).

As can be seen from Figures 3 and 4 the throughput

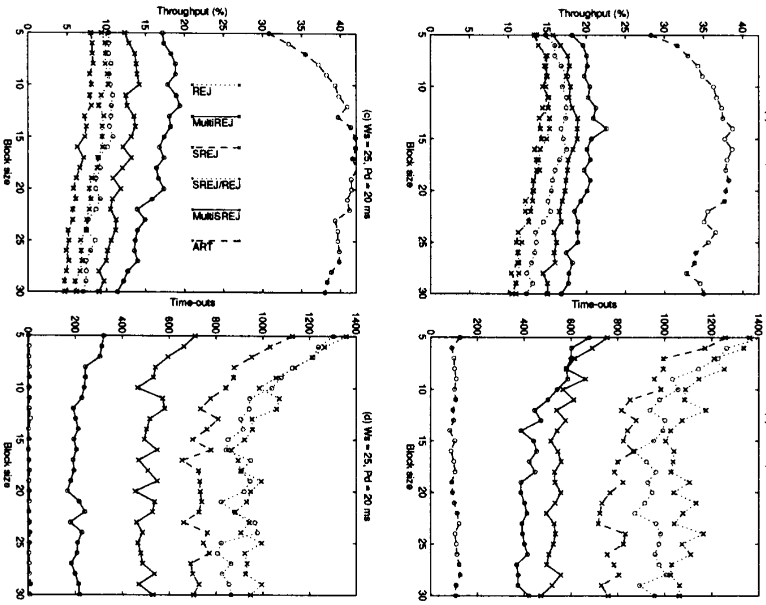

Figure 3: Throughput and number of timeouts: Pd = 20 ms, Ws = 10 and 25

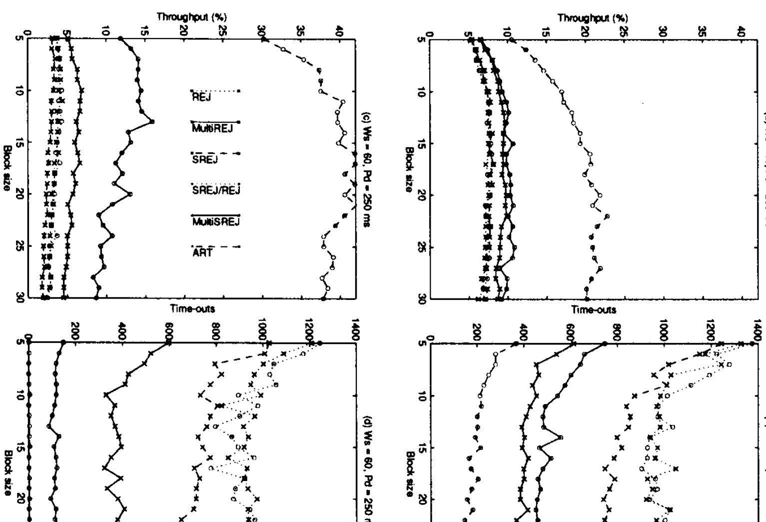

Figure 4: Throughput and number of timeouts: Pd = 250 ms, Ws = 10 and 60

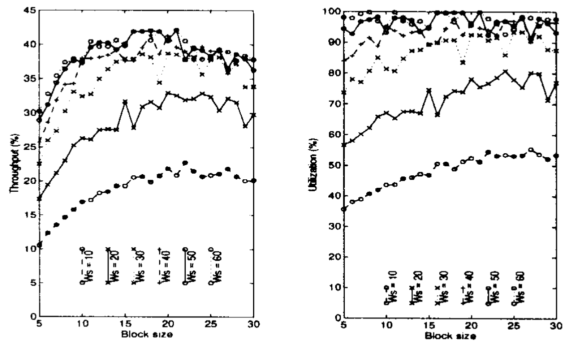

Figure 5: ART throughput and channel utilization: Pd = 250 ms.

of the ART scheme is about 150 to 350 % more than the highest throughput among the investigated protocols, and the optimum block size is quite small about 13 to 25 bytes (for the error condition used in these results).

The throughput efficiency of the ART scheme increases as the window size increases and hence the number of timeouts decreases. However, as the window size increases the throughput of ART increases until it reaches the point that most of the idle time has been utilized and the number of timeouts is almost zero, see Figure 5.

A larger window size is needed in the case of a high propagation delay to maintain the same level of throughput efficiency as in a low propagation delay case. For example, a window size larger than approximately 50 is needed in the case of a propagation delay of 250 ms compared to a window size of 25 in the case of a propagation delay of 10 ms, see Figure 3(c) and Figure 4(c).

6 Conclusions

In harsh error environments (such as cellular radio channels), frequent loses and corruptions of frames are highly expected. Therefore, a retransmission scheme that does not depend on a particular frame for its continuous operation is preferable (i.e., minimizing the idle time caused by timeout recoveries).

The existing standard and extended HDLC protocols have been designed with some assumptions regarding the behaviour of the error patterns. They greatly suffer from loses and corruptions of particular frames (such as S-frames, frames with the P/F bit set to one, and retransmitted I-frames). This is because they assume that these types of frames are not lost or corrupted so often. Therefore, a timeout recovery is used in case of loss or corruption of these frames. Unfortunately, extensive timeout recoveries degrade the throughput efficiency because they introduce significant idle time.

The proposed ART scheme, which depends on the transmitter to automatically retransmit erroneous I-frames, significantly reduces the use of timeout recoveries. Hence, it improves the throughput efficiency. The retransmission strategy of ART is based on the acknowledgement list reported by a multiple acknowledgement (MRR) and also on keeping records of the transmission order in which frames were sent.

The operation of the ART scheme does not depend on a particular frame for its continuous operation. The loss or corruption of any I-frame or MRR-frame is not critical as long as the transmitter receives any MRR-frame before the window is full and the retransmission set is empty. Note that the receiver sends an MRR-frame each time an I-frame is received. Each MRR-frame has sufficient information about the receiver buffer status.

The only time when a timeout recovery may occur in ART is when the transmitter does not receive any acknowledgement from the receiver for a period of time equal to the time required to transmit a full window of frames. With a large window size the usage of timeout recovery becomes very rare.

The retransmission strategy of the ART scheme could be combined with other error control mechanisms (such as FEC and variable block size) to possibly improve the throughput efficiency and adapt to the variability of non-stationary channels.

The behaviour of error patterns on cellular radio channels is unpredictable and variable with the time depending on factors such as the location and speed of the mobile unit. Therefore, a data link protocol with a block size adaptation mechanism would certainly maintain the throughput efficiency to some acceptable level. The adaptation has been added to the ART scheme and it illustrated very good results (not shown in this paper).

References

[1] T. Sato, M. Kawabe, T. Kato, and A. Fukasawa, "Protocol configuration and verification of an adaptive error control scheme over analog cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 41, pp. 69–76, Feb. 1992.

[2] A. Al-Zoman, J. DeDourek, and B. Kurz, "Modelling cellular radio channels for asynchronous data transmissions," in *Proceedings of IASTED International Conference on Modelling, Simulation and Control in the Process Industry, Ottawa, Ont.*, May 1994.

[3] W. Lee, *Mobile cellular telecommunications systems*. McGraw-Hill, 1989.

[4] ISO, *Information processing systems - Data Communication - High-level data link control elements of procedures*. International Standard, ISO:4335, 1987.

[5] U. Black, *Data Link Protocols*. Englewood Cliffs, New Jersey: PTR Prentice Hall, 1993.

[6] A. Al-Zoman, J. DeDourek, and B. Kurz, "Performance of data link protocols over cellular networks," in *Proceedings of the 2nd International conference on telecommunication systems modelling and analysis, Vanderbilt University, Nashville, TN*, March 1994.

[7] P. T. Brady, "Performance evaluation of multireject, selective reject, and other protocol enhancements," *IEEE Transactions on Communications*, vol. com-35, pp. 659–666, June 1987.

[8] M. E. Anagnostou and E. N. Protonotarios, "Performance analysis of the selective repeat protocol," *IEEE Transactions on Communiacations*, vol. com-34, pp. 127–135, Feb. 1986.

[9] K. J. Guth and T. T. Ha, "Adaptive stop-and-wait ARQ strategy for mobile data communications," *Computer Communications*, vol. 14, pp. 405–412, Sept. 1991.

[10] Y. C. J. Liu, "A throughput analysis of SDLC/HDLC link level protocols," in *Proceedings of IEEE Global Communications Conference*, (Tokyo, Jaban), pp. 209–216, 1987.

[11] C. K. Siew and D. J. Goodman, "Packet data transmission over mobile radio channels," *IEEE Transactions on Vehicular Technology*, vol. 38, pp. 95–101, May 1989.

[12] J. C. I. Chuang, "Comparison of two ARQ protocols in a Rayleigh fading channel," *IEEE Transactions on Vehicular Technology*, vol. 39, pp. 367–373, Nov. 1990.

[13] D. J. Bond, "A theoretical study of burst noise," *Br. Telecom Technol*, vol. 5, pp. 51–60, Oct. 1987.

[14] N. S. Jayant, "Data communication through the atmospheric burst error channel," *IEEE Transactions on Communications*, vol. com-15, pp. 383–389, June 1967.

[15] M. R. Karim, "Transmission of digital data over a Rayleigh fading channel," *IEEE Transactions on Vehicular Technology*, vol. vt-31, pp. 1–6, Feb. 1982.

[16] H. Bruneel and M. Moeneclaey, "On the throughput performance of some continuous ARQ strategies with repeated transmissions," *IEEE Transactions on Communications*, vol. com-34, pp. 244–248, Mar. 1986.

[17] T. H. Lee, "Throughput performance of a class of continuous ARQ stagies for burst-error channels," *IEEE Transactions on Vechicular Technology*, vol. 41, pp. 380–386, Nov. 1992.

[18] D. Towsley, "The stutter go back-n ARQ protocol," *IEEE Transactions on Communications*, vol. com-27, pp. 869–875, June 1979.

[19] R. A. Comroe and D. J.Costello, "ARQ schemes for data transmission in mobile radio systems," *IEEE Journal on Selected Areas in Communication*, vol. SAC-2, pp. 472–481, July 1984.

[20] G. Fairhurst, "Operataion of HDLC ABM in harsh error environments: a robust extension," *IEE Proceedings-I*, vol. 138, pp. 291–298, Aug. 1991.

[21] S. B. Wicker, "High-reliability data transfer over the land mobile redio channel using interleaved hybrid-ARQ error control," *IEEE Transactions on Vehicular Technology*, vol. 39, pp. 46–55, Feb. 1990.

[22] A. Netravali, W. Roome, and K. Sabnani, "Design and implementation of a high-speed transport protocol," *IEEE Transactions on Communications*, vol. com-38, pp. 2010–2023, Nov. 1990.