Philippine Engineering Journal (1991) XII(2): 1-10

CHANNEL UTILIZATION OF A HALF-DUPLEX ARQ PROTOCOL FOR PACKET RADIO CHANNELS

Efren F. Abaya, Ph.D.
Assistant Professor
Department of Electrical Engineering
University of the Philippines
Diliman, Quezon City

ABSTRACT

ARQ is a common error control method for terrestrial data transmission. The different types of ARQ are reviewed, and channel utilization characteristics are summarized. A half-duplex go-back-N ARQ method is presented as a model for the AX.25 ARQ protocol used in packet radio, and an equation for channel utilization is derived and analyzed.

INTRODUCTION

Retransmission or ARQ is the most common method for dealing with errors in terrestrial data transmission. Generally there are three types: stop-and-wait ARQ for half-duplex channels, continuous ARQ for full-duplex channels, and selective ARQ.

Over single-frequency radio channels such as HF-SSB, VHF and UHF links, half-duplex data transmissions for simple PC-to-PC file transfer employ various protocols such as the AX.25 amateur packet radio protocol [1,6] or the TCP/IP protocol [5]. This method of communication is convenient and inexpensive, but is limited by the low speeds attainable on radio channels that are prone to signal fading and interference. Thus, it is important to be able to utilize the low-speed channel efficiently.

This paper summarizes some mathematical models that are found in the literature for predicting or analyzing the channel throughput or channel utilization performance of the three general types of ARQ protocols. A half-duplex go-back-N ARQ method is presented as a model for the AX.25 ARQ protocol commonly used in HF, VHF and UHF packet radio systems, and an equation for channel utilization is derived and analyzed.

CHANNEL UTILIZATION OF ARQ PROTOCOLS

For the discussion in this paper, we assume a single sender transmitting data to a single receiver over a noisy channel. Generally the data are sent in small blocks, or frames, that in addition contain a fixed number of overhead bits used for addressing, control and error detection.

In retransmission or ARQ, the sender transmits one or more frames to the receiver. For each frame that arrives, the receiver returns an appropriate acknowledgement that indicates to the sender whether the frame is accepted by the receiver (i.e., error-free) or is rejected because of errors. If a frame is rejected, then the sender retransmits the same frame.

Three general types of ARQ are briefly described below. More extensive descriptions can be found in the literature [4, 5].

The most basic retransmission procedure, commonly utilized on half-duplex channels, is called "stop-and-wait ARQ". The sender transmits one frame of information, then waits for the receiver to return an acknowledgement. If a positive acknowledgement (ACK) is returned, the sender may proceed to send a new frame. If a negative acknowledgement (NAK) is returned, the sender retransmits the errored frame. In some versions of this protocol, NAK is not used and the receiver does not give any reply for an errored frame. Instead, a T second timer at the sender forces retransmission whenever the timeout expires without having received an ACK. This seldom used variation is called "Positive Acknowledgement With Retransmission" (PAR) [5]. The timeout provides protection against deadlock in case frames or acknowledgements can be lost in the channel. In either version, the sender activity alternates between two states--transmitting a frame and waiting for an acknowledgement--with a duty cycle that depends on the length of a frame and the time delays in the channel.

On full duplex channels, "continuous ARQ" or "go-back-N ARG" gives better performance. Here the sender is allowed to transmit frames continuously. The receiver returns an ACK or NAK for each frame received. An ACK response permits the sender to continue sending new frames. A NAK response forces the sender to retransmit all frames in sequence beginning with the rejected frame. (We assume that the sender and receiver have a mechanism such as frame sequence numbers for identifying which frame is being accepted or rejected.) The protocol is characterized by outstanding (unacknowledged) frames since the sender can usually transmit many frames before the acknowledgement for the first frame can return to it. Therefore, when one frame is rejected, all of the frames following it must be retransmitted also, even if these are error-free.

In most go-back-N ARQ implementations, a limit (represented by the variable W defined below) is set on the number of outstanding frames that the sender is allowed to have. In AX.25 and protocols like HDLC [5], this limit can be up to seven frames. If the maximum number of frames has been reached but no acknowledgements have yet been received, the sender has to stop and wait for the acknowledgements to arrive. However, if acknowledgements arrive promptly, the limit is never reached and the sender can transmit frames without pausing.

The most sophisticated retransmission procedure, called "selective ARQ", is similar to continuous ARQ. The difference is that whenever a NAK is returned, only the single errored framed is retransmitted. A limit on the number of outstanding frames that the sender is allowed may also be imposed. This procedure is rarely used in terrestrial data transmission because of its more complicated buffering requirements.

In most cases, the performance measure of interest is the so-called "net data throughput" (NDT) or "Transmission Rate of Information Bits" (TRIB) defined as the average number of information bits accepted by the receiver per unit time (excluding overhead bits, errors and retransmissions) measured in bits per second [3]. Given this figure, the average time to transmit a message of B bits can then be estimated as B/NDT. Therefore, we are interested in maximizing NDT to minimize average transmission time.

For purposes of mathematical analysis, it is convenient to normalize the through put by the Channel data rate to obtain a dimensionless quantity called the "channel efficiency" or "channel utilization" U defined as:

U = (net data throughput)/(channel data rate)

The utilization depends on several factors, such as:

C = channel data rate in bps

D = number of data bits in a frame

H = number of overhead bits in the frame header and trailer, i.e., all non-data bits (assumed constant)

P1 = probability that a data frame is damaged (also called "block error rate" or BLER)

P2 = probability that an acknowledgement is damaged

T = idle time between sender transmissions, including round-trip propagation delay, channel turnaround times, and time to transmit an acknowledgement

W = maximum number of outstanding frames allowed.

For compactness in the equations that follow, it is convenient to define the following variable:

F = D + H = total number of bits in a frame

Equations found in the literature for the channel utilization of the three general types of ARQ procedures will now be summarized below. It is assumed that the unit of transmission is the frame, which includes a header, data and check bits for detecting errors in the frame. For consistency of notation, the formulas have been re-written in terms of the variables that were previously defined.

For stop-and-wait ARQ only one frame at a time may be sent so $W = 1$ in our model. The time T is interpreted as the sender's idle time between frames, or the time for the sender to get an acknowledgement. This idle time includes the time for the receiver to process the frame and to physically transmit an acknowledgement message. The channel utilization for stop-and-wait ARQ is given by [5]:

$$U = D(1-P2)(1-P1)/(F+CT) \qquad \text{Eq.(1)}$$

In this formula, the factor (1-P2)(1-P1) represents the proportion of frames sent that are accepted by the receiver (i.e., excluding frames that get errors, that are lost, or whose acknowledgements get lost). The factor D/(F+CT) represents the proportion of channel time actually used to send data bits, excluding the time spent in sending header bits and idle time wasted in waiting for acknowledgements. The factor CT, which appears in several more equations, is the idle time measured in bits. Similar, though not identical, formulas for stop-and-wait ARQ have been reported by other researchers [2-4].

For continuous ARQ with fast acknowledgements over a full duplex channel where the sender never exceeds the limit W, and thus is allowed to send frames continuously, a lower bound on channel utilization is given by [4]:

$$\begin{aligned} U &\geq (1-P1)/(1 + xP1) \\ x &= [CT'/F] \qquad \text{Eq.(2)} \end{aligned}$$

The symbol $\lfloor \; \rfloor$ stands for the "floor function" or the largest integer not exceeding the quantity enclosed. In this formula, T' represents the time that elapses between the end of one frame and the instant that the acknowledgement for it arrives. In the meantime, the sender continues to transmit more frames. Therefore, in the event that the first frame is rejected, a total of $1+x$ frames have to be retransmitted. It is assumed that $1+x \leq W$. A slightly different formula for the case $W=2$ is found in [2].

For selective ARQ over a full duplex channel where the sender exceeds the limit W, and thus transmits continuously, the channel utilization is given by [4]:

$$U = D(1-P2)(1-P1)/F \qquad \text{Eq.(3)}$$

In the case of selective ARQ where the sender stops for T seconds after transmitting W frames, the channel utilization is [4]:

$$U = D(1-P2)(1-P1)W/(WF+CT) \qquad \text{Eq.(4)}$$

A HALF-DUPLEX PACKET RADIO PROTOCOL

The protocol to be described and analyzed in this section models the ARQ procedure commonly used on packet radio systems that employ the AX.25 amateur radio protocol [1,6]. It is assumed that a single frequency is shared by one sender and one receiver so that the channel is half-duplex. Signal fades, noise and intermittent interference result in transmission errors that require retransmission of frames. The protocol is a variation of the go-back-N ARQ for half-duplex channels. The sender transmits a burst of W frames, then stops for T seconds to receive the acknowledgements from the receiver. If one frame in a burst is rejected, then it and all frames following it must be retransmitted. For example, if the third frame is rejected, then it and all (W-3) frames following it are retransmitted in the subsequent burst. Two new frames may also be included with this next burst.

On radio channels, it is possible for frames to be lost (i.e., the radio signal is obliterated by interference or rapid fading). Acknowledgements from the receiver can also suffer a variety of impairments, such as being lost, hit by errors, or delayed. In case a valid acknowledgement is not received within T seconds after transmitting a frame, the sender retransmits all W frames immediately.

The channel activity from the sender's point of view may be illustrated as follows:

```
                                          retry   and/or   new   frames
 <- - - - - - - -W frames- - - - - - - ->        <- - - - -               - - - - ->
  [FRAME][FRAME][FRAME]                          [FRAME][FRAME][FRAME]
                              <- - -T sec - - ->
```

The channel utilization of this half-duplex go-back-N protocol can be expressed as the product of three factors:

$$U = \begin{array}{c}\text{framing}\\ \text{efficiency}\end{array} \ \text{X}\ \begin{array}{c}\text{duty cycle}\\ \text{efficiency}\end{array} \ \text{X}\ \begin{array}{l}\text{frame}\\ \text{acceptance}\\ \text{rate}\end{array}$$

The framing efficiency measures the proportion of data bits to total bits transmitted in one frame (whether good or errored). Since every frame carries H overhead bits, this factor is computed as:

$$\text{framing efficiency} = D/(D+H)$$

The duty cycle efficiency measures the proportion of channel time actually utilized by the sender in transmitting frames (whether goor or errored). Since the sender transmits a burst of W frames, then stops for T seconds to received an acknowledgement, this factor is computed as:

$$\text{duty cycle efficiency} = WF/(WF+CT)$$

Finally, the frame acceptance rate measures the proportion of frames send that are NOT retransmitted. Here we take the point of view that any frame that is subsequently retransmitted is "wasted" whether it is errored or not. The frame acceptance rate is computed by looking at one transmission burst of W frames. For simplicity, we assume that frame errors occur independently. Since all frames beginning at the first point of error area retransmitted, the number of frames X that will not require retransmission (i.e., are accepted by the receiver), on the condition that a valid acknowledgement is returned, is described by the following truncated geometric probability distribution:

$$\mathrm{P(X=k/acknowledgement\ returned)} \quad = \quad (1-P1)^{k}P1$$
$$k \ = 0,1,2,\ldots(W-1)$$

$$\mathrm{P(X=W/acknowledgement\ returned)} \quad = \quad (1-P1)^{W}$$

The mean of this conditional distribution, which gives the mean number of frames accepted, is

$$\mathrm{E[X/acknowledgement\ returned]} \quad = \quad (1-P1)[1-(1-P1)^{W}]/P1$$

There is also a probability P2 that a valid acknowledgement is NOT returned due to errors in the acknowledgement message itself. This would result in retransmitting all W frames or getting zero accepted. Therefore the overall mean number of frames accepted per burst is

$$E[X] \ = \ (1-P2091-P1)[1-(1-P1)^{W}]/P1$$

The frame acceptance rate is therefore E[X]/W. The reciprocal of this quantity given the average number of tries per successful frame. The effect of transmitting several frames in one transmission burst can be seem from the average number of tries per successful frame, which is given by (ignoring P2):

$$\frac{1}{1-P1} \ \mathrm{X} \ \frac{WP1}{1-(1-P1)^{W}}$$

The first term is the classical number of tries due to transmission errors that hit the frame. The second term accounts for extra transmissions when earlier frames in the same transmission burst are rejected. This term is asymptotic to WP1 as W increases.

Putting all the factors together gives the channel utilization of the half-duplex go-back-N ARQ protocol as:

$$U = D(1-P2)(1-P1)[1-(1-P1)^{W}]/[WF+CT)P1] \qquad \text{Eq.(5)}$$

NUMERICAL COMPUTATIONS

To further analyze the throughput behavior of the half-duplex go-back-N ARQ protocol, a model for the block error rate P1 and acknowledgement error probability P2 is needed. Usually, these will depend on the channel bit error rate E and the number of bits transmitted. If bit errors are random and independent, then

$$P1 = 1 - (1\text{-}E)^F$$

This model will be used in the rest of this section. Other models are described in [3] and [4]. In [4], the model is an empirical fit to experimental measurements. In [3] the model tries to account for burst errors. Since the length of acknowledgement messages is constant, the probability P2 simply becomes a small constant which can be ignored to simplify the analysis.

When W = 1, the channel utilization formula Eq. (5) reduces to that for stop-and-wait ARQ which exhibits the well-known unimodal dependence on the frame size F or number of data bits D exhibited in Fig. 1. A frame size that is very small in comparison to the overhead H and idle time CT results in a low duty cycle efficiency that lowers the channel utilization. On the other hand, a very large frame size incurs a relatively high block error rate P1, resulting in excessive retransmissions that likewise lowers the channel utilization. Consequently, there exists an optimum frame size (dependent on E, CT and H) that maximizes the channel utilization or throughput.

The same qualitative dependence between U and D is present for larger values of W. Figure 1 shows typical curves of U versus D for various values of W. Typical parameters for packet radio transmission over HF-SSB radio are used here: bit error rate E = 1.5×10^{-4} for non-faded conditions; CT = 300 representing a 300 bps channel with one second delay between transmission bursts; AX.25 header bits H = 168. To simplify calculations, errors in acknowledgements (represented by P2) are ignored.

For data sizes below 144 bytes, Fig. 1 shows that two frame bursts (W = 2) or three-frame bursts (W = 3) can give slightly better utilization than simple stop-and-wait ARQ (W = 1). Moreover, the utilization for W = 2 or W = 3 exhibits less sensitivity to data size D than W = 1 over the typical range of data sizes used in packet radio, which may be important when the frame size can vary. The default packet size of 128 bytes used in packet radio is well within this range.

Figure 2a plots channel utilization versus W for the same parameter values specified above. Different curves are shown for values of data size D ranging from 128 to 2048 bits (16 to 256 bytes). This figure shows that there is a rather limited range of data sizes between 512 to 1024 bits (64 to 128 bytes) that gives near-optimum channel utilization.

Given an assumed channel rate of 300 bps for HF-SSB or 1200 bps for UHF, the values of utilization in Fig. 2a yield predicted average transfer rates of 150 bps to 600 bps. This is good enough to transmit one page in about 80 sec. (20 sec. for UHF). Errors in acknowledgements (P2 0) will increase these times slightly.

Fig. 2b shows the effect of varying W when the bit error rate is slightly higher at E = 10^{-3}. This curve shows the same behavior as Fig. 2a. A range of 256 to 512 bits (32 to 64 bytes) with small W gives the best performance.

A more general comparison of stop-and-wait ARQ (single-frame bursts) versus multiple frame bursts can be made by comparing Eq. (1) with Eq. (5) for the case $W=2$. Let $a = CT/F$. For $W=1$

$$U = D(1-P2)/F \times (1-P1)/(1+a)$$

For $W=2$

$$U = D(1-P2)/F \ X \ (1-P1)/P1 \times [(1-P1)^2]/(2+a)$$

From these equations, it may be seen that stop-and-wait ARQ gives better channel utilization if

$$P1 > a/(1+a)$$

Therefore stop-and-wait ARQ is better if CT is small and F large, or if the channel has a high error rate. For example, if $CT=300$ and $F=1200$, the dividing point is $P1 = 0.2$. Since error rates are usually small in condition without fading and interference, this result suggests that multi-frame bursts may have practical applicability in situations where frames are constrained in length.

A rough estimate for the frame length for which $W=1$ is better can be obtained by putting the above criterion in terms of F, with the approximation $P1 = EF$. Then

$$EF(1 + F/CT) > 1$$

and stop-and-wait ARQ has better utilization when

$$F > [CT/E]^{1/2}$$

On the other hand, $W=2$ gives better utilization when

$$F > [CT/E]^{1/2} - CT/2$$

Using the values given previously, these would predict better performance for stop-and-wait ARQ when D 155 bytes.

CONCLUSION

An equation for predicting the channel utilization or channel throughput of a half-duplex go-back-N ARQ procedure has been presented. This mathematical model is useful for studying the performance of packet radio systems using the Ax.25 protocol. Preliminary calculations suggest that transmitting several frames in one burst can give improvements in average speed of transmission. However, other channel condition, such as high error rate, have to be studied before this conclusion can be generalized.

REFERENCES

1. Kirk A. Kleinschmidt, ed., *The ARRL Handbook For the Radio Amateur*, American Radio Relay League, 1989, Ch. 19.

2. Balkovic, M.D. and P.E. Muench, "Effect of Propagation Delay, Caused By Satellite Circuits, On Data Communication Systems That Use Block Retransmission For Error Correction", *IEEE Int. Conf. on Communications Conference Record*, June 1969, pp. 29-31 to 29-36.

3. Boustead, Carl N. and Kirit Mehta, "Getting Peak Performance on a Data Channel", in Harry R. Karp, ed., *Basics of Data Communications*, McGraw-Hill, 1976.

4. Burton, H.O. and D.D. Sullivan, "Errors and Error Control," *Proceedings of the IEEE*, vol. 60, November 1972, pp. 1292-1301; reprinted in Paul E. Green, Jr. and Robert W. Lucky, eds., *Computer Communications*, IEEE Press, 1975, pp, 260-268.

5. Tanenbaum, Andrew S., *Computer Networks*, 2nd. ed., Prentice-Hall, 1989, Ch.4.

6. Waller, Mark, "Making Waves", *Byte*, Dec. 1989, pp. 363-366.

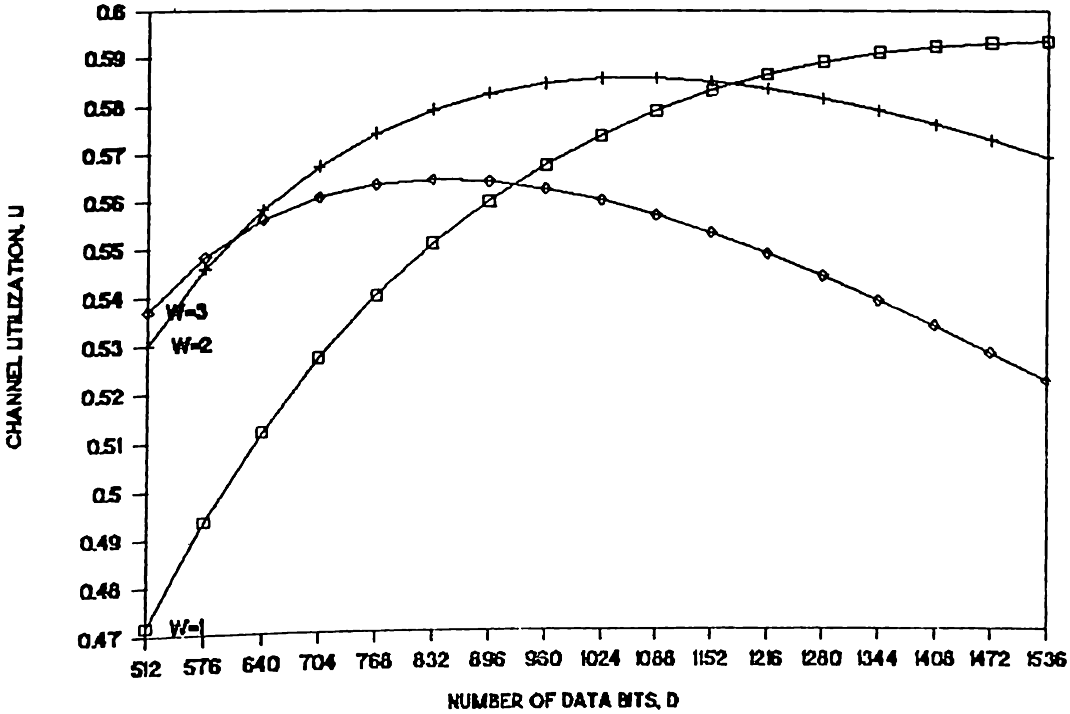

Figure I. Channel Utilization Versus Frame Size

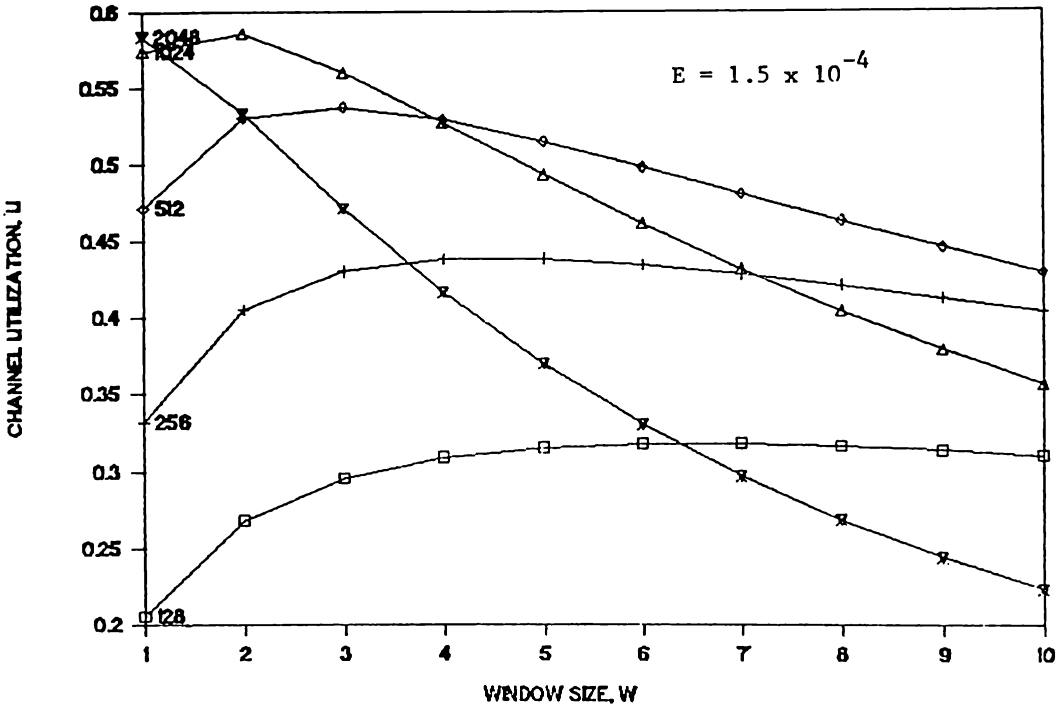

Figure 2a. Channel Utilization Versus Window Size

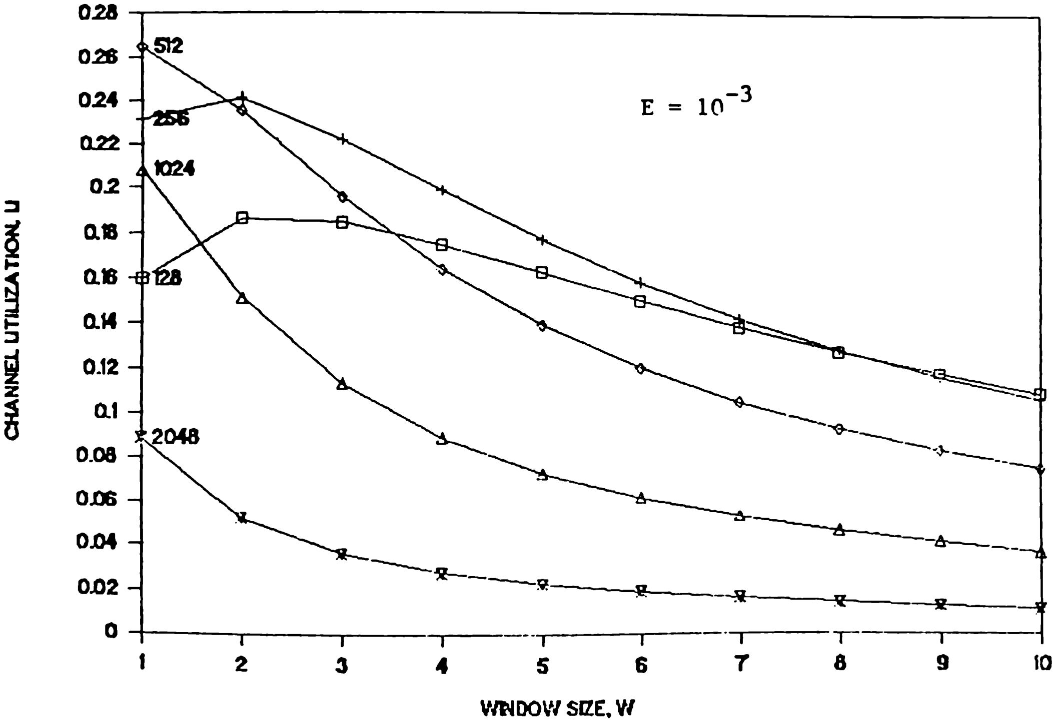

Figure 2b. Channel Utilization Versus Window Size