Reprinted from pp. 307-315 of Proceedings of the Int. Conf. on Info. Th. and Systems, NTG-Fachberichte vol. 65, Berlin, Sept. 18-20, 1978

CODING TECHNIQUES FOR DIGITAL DATA NETWORKS

James L. Massey

Department of System Science, University of California, Los Angeles.

1. Introduction

The most common form of error control in the links of terrestial digital data networks has been coding for error-detection only as part of an automatic repeat-request (ARQ) system. Such ARQ systems have generally been preferred over forward-error-correction (FEC) systems for two reasons. First, because a packet of data may be routed through several links so that end-to-end error control is complex and costly of network resources, individual link transmissions are usually required to be virtually error-free (say, an error probability of 10^{-12} or less for a packet of data). Such extremely small error probabilities are much more simply obtainable with an ARQ system than with an FEC system. Second, the necessary feedback channel in an ARQ system is readily available in digital data networks since the communication links are characteristically bi-directional.

Because error-detection is so much simpler than error-correction, both in implementation and in theory, it might seem that there is little new that can be said about the former at this late date in the development of coding theory. In fact, however, there are several common misconceptions about error-detection that need to be dispelled. In Section 2, we review these misconceptions and also present several new results on the undetected error probability for linear codes that should prove useful to the designers of ARQ systems for the links of digital data networks. Another characteristic of digital data networks is that there are many senders of information. This multiplicity of senders leads to mutual interference of their signals. In Section 3, we derive a bound on the "ARQ capacity" of multiple-access communications links and illustrate its application to an idealized, but theoretically interesting, memoryless multiple-access channel.

2. Undetected Error Probability for Linear Codes

2.1. Preliminaries

Let the binary K-tuple $\underline{x}_I$ denote the K information bits to be transmitted in an ARQ system, and let the binary (N-K)-tuple $\underline{x}_p$ denote the N-K redundant digits which are appended to form the transmitted code word $\underline{x} = [\underline{x}_I : \underline{x}_p]$. Let the binary N-tuple $\underline{y}$ represent the received word. We can write

$$\underline{y} = \underline{x} + \underline{e} \tag{1}$$

where $\underline{e} = [\underline{e}_I : \underline{e}_p]$ is the binary error pattern and where the summation in (1), as well as all summations of binary digits hereafter in this section, is modulo-two. The receiver accepts $\underline{y}$ when $\underline{y}$ is a valid code word, otherwise it asks for a repeat of the transmission. Thus, an <u>undetected error</u> occurs if and only if $\underline{y}$ is a code word different from $\underline{x}$ or, equivalently, if and only if $\underline{x} + \underline{e}$ is a code word and $\underline{e} \neq \underline{0}$. Both for ease of encoding and because the resultant codes are best understood, the components of $\underline{x}_p$ are generally chosen as modulo-two sums of selected components of $\underline{x}_I$. (We ignore, without loss of generality in our analysis, the fact that certain components of $\underline{x}_p$ may be complemented before transmission to eliminate $\underline{0}$ as a valid code word, or for other

reasons.) The encoding operation can then be written

$$\underline{x}_P = \underline{x}_I P \tag{2}$$

where P is some K × (N-K) binary matrix. In the usual language of coding theory, such a code is an (N,K) linear systematic code; its generator matrix is

$$G = [I_K : P] \tag{3}$$

where I_K is the K × K identity matrix.

We suppose next that the forward channel is a binary symmetric channel (BSC) with crossover probability ε ($0 \le \varepsilon \le 1$). In this case, the components of $\underline{e}$ are statistically independent binary random variables, each taking on value 1 with probability ε. It is now a simple matter to write the expression for the undetected error probability $P_{ue}(\varepsilon)$; we leave ε as an explicit variable because we shall later be interested in the dependence on ε.

Let A_i, i = 0,1,...,N, be the weight enumerators of the linear code, i.e., A_i is the number of code words with Hamming weight i. Equivalently, A_i is the number of code words at Hamming distance i from any given code word $\underline{x}$. Since $\varepsilon^i(1-\varepsilon)^{N-i}$ is the probability of the error pattern $\underline{e}$ that will cause $\underline{x}$ to be received as any given N-tuple $\underline{y}$ at distance i from $\underline{x}$, we have the well-known (cf. [1,p.21]) expression

$$P_{ue}(\varepsilon) = \sum_{i=1}^{N} A_i \varepsilon^i (1-\varepsilon)^{N-i} \tag{4}$$

The utility of (4) is hindered by the fact that the weight enumerators are known only for a very few linear codes. Nonetheless, there is one case in which (4) can always be evaluated exactly, namely when $\varepsilon = 1/2$. For any linear code $A_0 = 1$, which implies

$$\sum_{i=1}^{N} A_i = 2^K - 1.$$

Thus, (4) yields

$$P_{ue}\left(\frac{1}{2}\right) = 2^{-(N-K)} - 2^{-N} \tag{5}$$

regardless of the particular (N,K) code used.

2.2. The Worst-Case Fallacy

By complementing the channel output when $\varepsilon > 1/2$, one can for most purposes restrict attention to the case $0 \le \varepsilon \le 1/2$ for BSC's. But the capacity of a BSC is $1 - h(\varepsilon)$ where

$$h(\varepsilon) = -\varepsilon \log_2 \varepsilon - (1-\varepsilon)\log_2(1-\varepsilon) \tag{6}$$

is the binary entropy function, so the capacity decreases monotonically from 1 to 0 as ε increases from 0 to 1/2. It is appealing then to argue that $P_{ue}(\varepsilon)$ will be a maximum for $\varepsilon = 1/2$, and hence, from (5), that $2^{-(N-K)}$ will be an upper bound on $P_{ue}(\varepsilon)$ for all ε in the range $0 \le \varepsilon \le 1/2$. The surprising

fact that this "plausibility argument" is incorrect was first pointed out by Leung and Hellman [2].

Example: Consider the (N,2) code with

$$G = \begin{bmatrix} 1 & 0 & 1 & 1 & \dots & 1 \\ 0 & 1 & 0 & 0 & \dots & 0 \end{bmatrix}$$

for which $A_0 = A_1 = A_{N-1} = A_N = 1$ when $N \geq 3$. For $N = 100$ and $\varepsilon = 1/N$, (4) yields

$$P_{ue}(.01) = 2^{-8}$$

which is $2^{90} \approx 10^{27}$ times greater than the "supposed worst case"

$$P_{ue}(1/2) = 2^{-98}.$$

More recently, Leung, Barnes and Friedman [3], by using arguments based on the known weight enumerators, have shown that $2^{-(N-K)}$ is indeed an upper bound on $P_{ue}(\varepsilon)$ for $0 \leq \varepsilon \leq 1/2$ for certain codes, most notably all perfect codes and all double-error-correcting primitive BCH codes. It might then be expected that, for "most codes," $2^{-(N-K)}$ is the worst-case undetected error probability — but we shall see next that this is also not true.

2.3. Average Undetected Error Probability

Consider now the ensemble of all (N,K) linear systematic codes in which all codes are equi-probable. This is equivalent to assigning to each code the probability that its P matrix would be chosen when the K(N-K) components of P are chosen independently and with probability 1/2 of being 1. It is well-known (cf. [1,pp.283-288]) that the average weight enumerators for this ensemble of codes are given by

$$\bar{A}_i = \left[\binom{N}{i} - \binom{N-K}{i}\right] 2^{-(N-K)}, \quad i \neq 0. \tag{7}$$

(The truth of (7) can be seen as follows: For $\underline{x} \neq \underline{0}$ but $\underline{x}_I = \underline{0}$, there are no solutions P of (2). But for any one of the $\binom{N}{i} - \binom{N-K}{i}$ vectors $\underline{x}$ of weight i with $\underline{x}_I \neq \underline{0}$, (2) always has solutions P — in fact, the fraction of P's which solve (2) is then exactly $2^{-(N-K)}$ since, if the j-th component of $\underline{x}_I$ is 1, for any choice of the other components in the m-th column of P, there is exactly one choice of the j-th component in this column so that the m-th components on both sides of (2) coincide.) It appears not to have been previously noticed that one can obtain, by taking averages in (4) and using (7), the average undetected error probability for this ensemble of codes as

$$\overline{P_{ue}(\varepsilon)} = [1-(1-\varepsilon)^K]\, 2^{-(N-K)}, \tag{8}$$

which holds for every ε, $0 \leq \varepsilon \leq 1$.

Equation (8) shows that, for any $\varepsilon > 0$,

$$\overline{P_{ue}(\varepsilon)} \approx 2^{-(N-K)}$$

when K is large. This suggests that, for any $\varepsilon > 0$, the probability of choosing a code with $P_{ue}(\varepsilon)$ greater than $2^{-(N-K)}$ is on the order of one-half! Thus there is good reason to believe that, for most codes, worst-case undetected error probability will not occur for $\varepsilon = 1/2$.

2.4. A Bound on the Worst-Case Undetected Error Probability

The worst-case undetected error probability is of interest for two reasons. First, the designer of the ARQ system often does not know ε in advance (and, in fact, the same system may be used on many channels with different ε's), but still desires to specify an upper bound on the undetected error probability. Second, on real channels, the value of ε itself may slowly vary — in fact, even $\varepsilon \approx 1$ is possible when differential modulation is employed. (When $\varepsilon \approx 1$ is possible, one must avoid using a linear code containing the all-one code word, i.e., a code with $A_N = 1$, for otherwise $P_{ue}(\varepsilon) \approx 1$ when $\varepsilon \approx 1$.) Hereafter, we write

$$P_{wc} = \max_{0 \le \varepsilon \le 1} P_{ue}(\varepsilon) \tag{9}$$

to denote the worst-case undetected error probability for a linear code.

Useful insight into P_{wc} can be obtained by examining the individual probabilities $\varepsilon^i(1-\varepsilon)^{N-i}$ that appear in (4). Elementary calculus shows that

$$\max_{0 \le \varepsilon \le 1} \varepsilon^i(1-\varepsilon)^{N-i} = 2^{-Nh(i/N)} \tag{10}$$

where the maximum is uniquely achieved by $\varepsilon = i/N$. (Equation (10) should be pleasing to information-theorists — it states that the probability of receiving a given code word at Hamming distance i from the transmitted code word is greatest when the required error pattern is a "typical error sequence" for the BSC.) This shows that the N terms in the summation of (4) make their maximum contributions to $P_{ue}(\varepsilon)$ at different values of ε, namely at $\varepsilon = i/N$ for $i = 1,2,\ldots N$.

Overbounding the maximum over ε of the summation in (4) by the sum of the maxima, and making use of (10), we obtain

$$P_{wc} \le \sum_{i=1}^{N} A_i 2^{-Nh(i/N)}, \tag{11}$$

which is reasonably tight (it can be pessimistic by at most a factor of N) upper bound on P_{wc}, but requires knowledge of the weight enumerators. We next average (11) over the same ensemble of codes as before to obtain, with the aid of (7),

$$\overline{P_{wc}} \le \sum_{i=1}^{N} \left[\binom{N}{i} - \binom{N-K}{i}\right] 2^{-(N-K)}\, 2^{-Nh(i/N)}, \tag{12}$$

which can be used numerically to find a reasonably tight upper bound on the average of P_{wc} for the ensemble of all (N,K) linear systematic codes. More insight can be obtained if we weaken (12) by using the bound

$$\binom{N}{i} - \binom{N-K}{i} \le \binom{N}{i} \le 2^{Nh(i/N)} \tag{13}$$

to obtain

$$\overline{P_{wc}} \leq N2^{-(N-K)}. \tag{14}$$

From (14), we draw the reassuring conclusion that most codes will have a worst-case undetected error probability less than N times greater than the "fallacious upper bound" $2^{-(N-K)}$. Thus, the bound $N2^{-(N-K)}$ can confidently be used for the P_{wc} of an ARQ system in which the code has been selected with some care — say, to have large minimum distance d and no code words of Hamming weight greater than N-d. In fact,

$$A_i \approx \binom{N}{i} 2^{-(N-K)}, \quad d \leq i \leq N-d$$

is a common "rule of thumb" for such codes, and considerable evidence supports the use of the bound (14) for any such specific code [1,pp.282-288].

The practical thrust of the above is that, to guarantee some specified worst-case undetected error probability, one will need about $\log_2 N$ parity digits more than the "fallacious upper bound" $2^{-(N-K)}$ indicates. For instance, to guarantee $P_{wc} \leq 10^{-12} \approx 2^{-40}$ with $N = 1000 \approx 2^{10}$, one should use 50 parity digits, rather than 40, in a block.

3. A Bound for Multiple-Access ARQ Systems

3.1. The Noiseless Binary Multiple-Access Adder Channel

Most of the interesting new problems in digital data networks stem from the fact that the network must simultaneously accommodate many senders of information. To indicate the new considerations that this introduces for ARQ systems, we consider the noiseless binary multiple-access adder channel for M users (M>1), which we denote hereafter by NBMAC(M). For this memoryless channel, there are M binary inputs X_i, i = 1,2,...M, and a single (M+1)-ary output Y given by

$$Y = X_1 + X_2 + \ldots + X_M, \tag{15}$$

where here and hereafter the summation is ordinary summation of integers. When Y = 0 or Y = M, the output uniquely determines all M input bits. But for $1 \leq Y \leq M - 1$, the output does not unambiguously determine the input bits; however, the ambiguity arises entirely from the "mutual interference" of the users, rather than from some channel "noise" mechanism. We do not claim that the NBMAC(M) is a realistic model of any practical multiple-access channel, but only that its very simplicity suggests that we should understand it well before going on to tackle more realistic models.

We first consider the use of the NBMAC(M) without the availability of feedback. Let R_i, i = 1,2,...M be the rate (in information bits per use of the channel) at which the i-th user sends information over the channel. The "capacity" for such many-sender channels is actually the boundary of the region of all vectors $(R_1,R_2,\ldots R_M)$ such that, under whatever constraints are imposed upon the senders, arbitrarily reliable communications are simultaneously achievable at these rates for the M senders. However, we shall only be interested in the total rate $R_1 + R_2 + \ldots + R_M$, and we shall say that capacity is the upper limit of this total rate for arbitrarily reliable communications.

For any "multiple-access" channel without feedback, the implied constraint is that there be no cooperation among the users, i.e., that $X_1, X_2, \ldots X_N$ be statistically independent random variables. It is well-known (cf.[4]) that capacity is then given by

$$C = \max I(\underline{X};Y) \tag{16}$$

where $\underline{X} = [X_1, X_2, \ldots X_M]$ and where the maximum is over all probability distributions $P(\underline{x}) = P(x_1)\, P(x_2) \ldots P(x_M)$. For the NBMAC(M),

$$I(\underline{X};Y) = H(Y) - H(Y|\underline{X}) = H(Y) \tag{17}$$

since $\underline{X}$ uniquely determines Y. It is further well-known (cf.[5]) that, for NBMAC(M), H(Y) is maximized when $P(X_i=1) = 1/2$, all i, which causes Y to be binomially-distributed and gives capacity as

$$C_{nc} = \sum_{i=0}^{M} \left[\binom{M}{i} 2^{-M}\right] \log_2 \left[\binom{M}{i} 2^{-M}\right], \tag{18}$$

where we have used the subscript "nc" to underscore the assumption that there is no cooperation among the users.

Suppose that we now remove the "multiple-access" constraint of no cooperation and instead permit the M users to use total cooperation. In this case, we actually have an ordinary one-user channel in which the user has a vector input. Capacity is then still given by (16), except that now the maximum is over an arbitrary probability distribution $P(\underline{x})$. Equation (17) still holds, so we see that now capacity is achieved when we choose $P(\underline{x})$ to make the M + 1 values of Y equally probable. Thus,

$$C_{tc} = \log_2(M+1), \tag{19}$$

where we have used the subscript "tc" to underscore the fact that we now permit total cooperation among the users of the NBMAC(M).

Examples: (1) For M = 2, $C_{nc} = 3/2$ and $C_{tc} = 1.585$.
(2) For M = 3, $C_{nc} = 1.811$ and $C_{tc} = 2$.

3.2. The ARQ Bound

We now consider the use of the NBMAC(M) in the presence of a noiseless feedback channel from which each user learns the value of Y after each transmission; however, the users are ignorant of each others' specific transmissions except as they learn of them via the feedback link. We define an ARQ system for such a situation to be a transmission scheme where, at each use of the channel, either (1) each user sends a new information bit, or (2) each user transmits a digit, entirely dependent on his own previous transmissions and past feedback digits, for use in resolving past ambiguous receptions. Hereafter, we refer to such transmissions as type-(1) or type-(2), respectively. We then define C_{ARQ}, the ARQ-capacity, to be the upper limit of the rate sum $R_1 + R_2 + \ldots + R_M$ for arbitrarily reliable communications.

We now prove a useful upper bound on C_{ARQ}. For each type-(1) transmission, the uncertainty about the input $\underline{X}$ at the receiver is

$$H(\underline{X}\,|Y) = H(\underline{X}) - I(\underline{X};Y)$$
$$= M - C_{nc} \tag{20}$$

because the inputs X_i, $i = 1,2,\ldots M$, are statistically independent and have $P(X_i = 1) = 1/2$ for all i. No matter how cleverly subsequent type-(2) transmissions are made, the average number, $\bar{T}$ of such type-(2) transmissions per type-(1) transmission must satisfy

$$\bar{T} \geq \frac{M - C_{nc}}{C_{tc}} \tag{21}$$

since, at best, the users can cooperate totally in supplying the information required to remove the ambiguities in past receptions. Since M information bits are introduced for each type-(1) transmission but none are introduced for type-(2) transmissions, it follows that

$$C_{ARQ} = \max \frac{M}{1 + \bar{T}},$$

which with the use of (21) gives

$$C_{ARQ} \leq \frac{M}{M + C_{tc} - C_{nc}} C_{tc}. \tag{22}$$

Inequality (22) is our desired <u>ARQ bound</u>; in this form, it applies to any M-user binary memoryless multiple-access channel for which $P(X_i=1) = 1/2$, all i, is the capacity-achieving distribution with the "no cooperation" constraint of statistical independence. For the NBMAC(M), the values of C_{tc} and C_{nc} are given by (19) and (18), respectively.

<u>Examples</u>: For the NBMAC(M), inequality (22) becomes

(1) for $M = 2$, $C_{ARQ} \leq 1.5204$;
(2) for $M = 3$, $C_{ARQ} \leq 1.8816$.

3.3. Tightness of the ARQ Bound

The case $M = 2$ for the NBMAC(M) is exceptional in that, after the feedback of Y, each user always knows exactly what the other sent. This means that the two users can totally cooperate for their type-(2) transmissions in an ARQ scheme, and thus the ARQ bound (18) holds with equality! In fact, Gaarder and Wolf [6] have given an ARQ scheme that approaches $C_{ARQ} = 1.5204$ arbitrarily closely for this case. Since $C_{ARQ} > C_{nc} = 3/2$, Gaarder and Wolf's result showed that the capacity with feedback for the NBMAC(2) is greater than its capacity without feedback — something that is never true for a one-sender memoryless channel. (In general, the "capacity with feedback" is greater than C_{ARQ} since ARQ schemes do not exhaust the ways in which feedback can be exploited.)

The case $M = 3$ is more interesting for the NBMAC(M) since, when $Y = 1$ (or $Y = 2$) for a type-(1) transmission, the two users who sent 0's (or sent 1's) do not know which other user sent the 1 (or sent the 0). Consider, however, this simple retransmission strategy:

$$X_1' = X_1$$

$$X_2' = 0$$

$$X_3' = 1 - X_3.$$

When $Y = X_1 + X_2 + X_3 = 1$, the received digit is $Y' = X_1' + X_2' + X_3' = 2X_1 + X_2$, which thus uniquely determines X_1, X_2 and X_3. When $Y = X_1 + X_2 + X_3 = 2$, the received digit is $Y' = 2X_1 + X_2 - 1$, which again uniquely determines X_1, X_2 and X_3. With probability 3/4, Y will be 1 or 2 on a type-(1) transmission, so that the average number of retransmissions per type-(1) transmission is just $\bar{T} = 3/4$. Hence, this simple ARQ achieves the rate

$$R = \frac{3}{1 + 3/4} = 1.714,$$

which is surprisingly close to the bound of 1.882 given by (22). Moreover, we could do no better even if total cooperation were allowed on the retransmissions, given that we chose to resolve each ambiguous type-(1) reception before making another type-(1) transmission.

During the past year, one of my graduate students, T. Paul, has made an extensive investigation of ARQ schemes for NBMAC(3) of the type where type-(1) transmissions are used until A ambiguous receptions occur, thereafter type-(2) transmissions are used until these ambiguities are fully resolved [7]. Quite surprisingly, Paul found schemes for A = 2 and A = 3 that, like the A = 1 scheme we just described, could not be improved upon even if total cooperation were allowed on the retransmissions! We conjecture that this is true for all A. Interestingly, the A = 3 scheme achieves the rate R = 1.831, which is greater than the capacity without feedback $C_{nc} = 1.811$. Moreover, Paul devised a clearly non-optimum scheme that, for $A \to \infty$, achieves a rate of about 1.86, very close to the ARQ bound of 1.88. It appears to be a safe conjecture that the bound (22) is tight for NBMAC(3), even if the above conjecture proves false.

For $M \geq 4$, it is not hard to see that for Y=1 no ARQ retransmission scheme can resolve A = 1 ambiguity in only one retransmission, as could be done if total cooperation were allowed on retransmissions. Nonetheless, it is not clear that (22) is not an equality. We clearly have a lot further to go in our understanding of ARQ schemes for NBMAC(M) when $M \geq 3$, much less of those for more realistic multiple-access channels.

4. Acknowledgment

The research reported here was performed at the Electronic Systems Laboratory of the Massachusetts Institute of Technology in Cambridge, Massachusetts, and was supported in part by the Office of Naval Research under Contract ONR-N00014-64-C-1183 and in part by the National Science Foundation under Grant NSF-ENG76-2447.

5. References

[1] F. J. MacWilliams and N.J.A. Sloane, The Theory of Error-Correcting Codes (2 volumes). Amsterdam: North-Holland, 1977.

[2] S. K. Leung and M. E. Hellman, "Concerning a bound on undetected error probability," IEEE Trans. Info. Th., vol. IT-22, pp. 235-237, March 1976.

[3] S. K. Leung, E. R. Barnes and D. U. Friedman, "On some properties of the undetected error probability of linear codes," to appear in IEEE Trans. Info. Th.

[4] E. C. van der Meulen, "A survey of multi-way channels in information theory," IEEE Trans. Info. Th., vol. IT-23, pp. 1-37, Jan. 1977.

[5] J. K. Wolf, "Multi-user communication networks," to appear in Proc. NATO Adv. Study Inst., Darlington, England, Aug. 1978.

[6] N. T. Gaarder and J. K. Wolf, "The capacity region of a multiple-access discrete memoryless channel can increase with feedback," IEEE Trans. Info. Th., vol. IT-21, pp. 100-102, Jan. 1975.

[7] T. Paul, "Some efficient ARQ strategies for the multiple-user adder channel," S. M. Thesis, Dept. Elec. and Comp. Engr., M.I.T., Cambridge, Massachusetts, expected completion date: Aug. 1978.