Presented at the American Mathematical Society short course on Computer Communications, Denver, CO, January 3-4, 1983.

January 1983 LIDS-P-1261

INTRODUCTION TO DATA COMMUNICATION

Pierre A. Humblet[†]

PREFACE. These lecture notes are intended as an introduction to the problem of transmitting digital data on analog channels subject to noise and distortion. Modulation, detection and coding theories are reviewed to the extent they apply directly to practical systems. Some non classical problems that arise when a channel is shared by many users are also treated.

I. INTRODUCTION.

The purpose of these notes is to discuss the problem of reliably transmitting digital information on analog circuits, like telephone lines, which are subject to noise and various kinds of distortions.

The need to do this efficiently arose first for defense in the 1950's and grew to the point where commerce and industry are almost as dependent on efficient and reliable digital communication as they are on voice communication.

At the outset engineers had all the essential theoretical ingredients at hand: classical detection theory, Nyquist's work on characterizing waveshapes allowing independent transmissions of a sequence of digits, and Shannon's information theory. It shows the existence of a maximum rate at which information can be transmitted as reliably as desired, provided one is willing to build complex encoders and decoders.

Sections 2, 3 and 4 will give a brief review of these theories, and outline how they are actually implemented.

With the recent advent of computer networks, where data is typically generated in a bursty fashion, it makes economical sense to share communication channels between different "streams" of traffic. Thus a method must be found to decide what stream can use the channel at a given time, and addressing information must also be transmitted on the channel. This leads to new problems that will be described in Section 5.

2. DETECTION THEORY.

The transformation of digital data into a form suitable for transmission

[†]Supported by NSF Grant ECS 79-19880.

on an analog channel is done in a device called a modulator. The reverse operation is done by a demodulator. The two devices are usually combined into a single one, called a "modem".

In its most idealized form a modem takes in a group of n binary digits, with values denoted i, $i = 1,2,\ldots 2^n$, and produces in turn one of the waveforms $s^i(t)$, $i = 1,2,\ldots 2^n$, that is transmitted on a channel.

A simple characterization of the channel is as a linear time invariant filter (introducting deterministic distortion) with impulse response h(t), followed by a source of additive noise n(t). For convenience n(t) will be modelled as a 0 mean stationary Gaussian process with correlation function k(t).

We will denote the convolution of s^i and h by $\tilde{s}^i$, and the channel output by r(t) ($r(t) = \tilde{s}^i(t) + n(t)$, for some i). $S^i(f)$, $\tilde{S}^i(f)$, H(f) and K(f) will denote corresponding Fourier transforms. Note that K(f) is real and non negative, as k(t) is a correlation function.

The demodulator receives r(t), and must produce an estimate of i.

The modern view of this problem is in terms of "signal space": the space $\tilde{S}$ of functions spanned by the $\tilde{s}^i(t)$. Assuming that $\int |\tilde{S}^i(f)|^2/K(f)\, df$ is finite for all i, we can make $\tilde{S}$ a Hilbert space by defining the inner product $\langle x(t), y(t)\rangle = \int X(f)Y^*(f)/K(f)\, df$, where x and y are in $\tilde{S}$, and X(f) and Y(f) are their Fourier transforms. * denotes complex conjugate.

By using Gram Schmidt's orthonormalization procedure one can find an orthonormal basis for $\tilde{S}$ and display the waveforms $\tilde{s}^i(t)$ as points in a finite dimensional space of dimension D, typically smaller than 2^N. $\tilde{s}^i_j$ will denote the jth coordinate of $\tilde{s}^i(t)$.

The demodulator processes the received waveform r(t). It can find its projection on $\tilde{S}$ (r_j denotes its jth coordinate). Elementary computations reveal that, thanks to the choice of innerproduct, the component of r(t) perpendicular to $\tilde{S}$ is independent of i and $(r_1, r_2, \ldots, r_D)$ and is thus irrelevant to the decision process.

Other elementary calculations show that, conditional on i being transmitted, the r_j's are independent Gaussian random variables with unit variance and mean $\tilde{s}^i_j$.

At this point the demodulator has all the statistical knowledge required to make a decision, which need only be based on the r_j's. In particular if all the i's are equally likely and minimum probability of error decoding is desired, one should decide i maximizing the conditional probability density of $(r_1, r_2, \ldots, r_D)$ given i, i.e. decide i such that

$$\sum_{j=1}^{D} (r_j - \tilde{s}^i_j)^2 \qquad (1)$$

is minimum so that the demodulator essentially performs minimum distance decoding.

The resulting probability of error can be shown to depend only on the distances $||\tilde{s}^i - \tilde{s}^j||^2$ between signals at the channel output, i.e. on

$$\int |S^i(f) - S^j(f)|^2 \, |H(f)|^2 / K(f) df,$$

leading to the obvious conclusion that the signals $s^i(t)$ should differ at frequencies where $H(f)$ is large (i.e. in the channel passband).

$K(f)$ is often assumed to be constant in that band, so that the innerproduct $\langle x,y \rangle$ is simply the usual innerproduct

$$\int x(t)y(t)\, dt.$$

The previous channel model is not accurate. Non linear and time varying distortions occur and noise is not always Gaussian. Nonetheless modems typically exhibit the structure just derived: first "filtering" elements and samplers compute the projection of the received waveform. They are followed in turn by a non linear decision device that chooses the signal closest to the received waveform.

The signal space $\tilde{S}$ has usually 1 or 2 dimensions, the basis functions being of the form $x(t)\cos(2\pi f_c t)$ and $x(t)\sin(2\pi f_c t)$, where $x(t)$ has a Fourier transform confined to the low frequencies. The usual telephone lines have a passband between 300 and 3000 Hz, so that f_c is typically 1650 or 1800 Hz. n is 2, 3, 4, or 6 depending on whether data is transmitted at 2400, 4800, 9600 or 14400 bits per second.

3. TRANSMISSION OF SEQUENCES

The number N of binary digits that must be transmitted by the modem is usually very large so that there is an enormous (2^N) number of waveforms, and the theory developed in the previous section cannot be implemented without more structure being added.

The sequence of N binary digits is usually parsed in groups of n (with values $1,2,\ldots 2^n$). If the kth group has value $m(k) = i$ then the waveform $s^i(t-kT)$ is transmitted (this is just the waveform $s^i(t)$ delayed by kT). The received waveform is the sum of the waveforms corresponding to each group plus the noise thus it has the form

$$r(t) = \sum_{k=0}^{\frac{N-n}{n}} s^{m(k)}(t-kT) + n(t).$$

To generate the most likely sequence the demodulator must find the values

$m(0), m(1),\ldots, m(\frac{N-n}{n})$ minimizing (1). This operation can be rewritten as maximizing

$$\sum_{k=0}^{\frac{N-n}{n}} \left(<r(t),\tilde{s}^{m(k)}(t-kT)> - \frac{1}{2}||\tilde{s}^{m(k)}(t-kT)||^2 - \sum_{k'<k} <\tilde{s}^{m(k')}(t-k'T),\tilde{s}^{m(k)}(t-kT)>\right) \quad (2)$$

The only quantities depending on r(t) in this formula are the $\frac{N}{n}\,2^n$ innerproducts $<r(t), s^i(t-kT)>$. For k fixed the demodulator can either evaluate them directly, or first compute the components of r(t) in the space spanned by the $\tilde{s}^i(t-kT)$.

Maximizing (2) would be easy if $<\tilde{s}^i(t-\ell T), \tilde{s}^j(t)>$ were 0 for all i,j and $\ell \neq 0$, i.e. if there were no "intersymbol interference". In that case each m(k) can be decoded independently of the others.

As observed by Nyquist, this condition is met if, for all i,j,

$$\sum_{\ell=-\infty}^{\infty} \tilde{S}^i(f+\tfrac{\ell}{T})\tilde{S}^{j*}(f+\tfrac{\ell}{T})/K(f+\tfrac{\ell}{T})$$

is independent of f.

Waveforms are usually designed to (approximately) satisfy this condition, as the task of the demodulator is then simplified. However it becomes increasingly hard to meet when the passband of the channel is close to $\frac{1}{2T}$, as is the case in high performance modems, or when the channel response H(f) is not known a priori.

If $<\tilde{s}^i(t-k^iT), \tilde{s}^j(t)>$ is negligible for $|k| > L$ then an elegant algorithm is available to maximize (2). It is a form of dynamic programming known to communication engineers as Viterbi's algorithm.

It rests on the observation that the maximum of the first K+1 terms in the outer sum of (2), over all sequences with m(0), m(1),...m(k-L) arbitrary and m(k-L+1),..m(K) given, can be expressed in terms of the maximum of the first K terms over all sequences with m(0), m(1),...m(K-L-1) arbitrary and m(K-L),...m(K-1) given. Thus by iterating on K one can find the sequence maximizing (2) after an amount of computation proportional to $\frac{N}{n}\,2^{(L+1)n}$. This is much less than expected, but is still too large to be commonly used.

The technique that is usually chosen to combat intersymbol interference is called "linear equalization". It is a heuristic method most easily explained when the $s^i(t)$ form a one dimensional signal set, i.e. $s^i(t) = a^i\, s(t)$.

In that case the demodulator computes $r(k) = <r(t), s(t-kT)>$. Instead of processing the r(k) to maximize expression (2), the r(k) are passed through a time invariant linear digital filter with coefficients $c(\ell)$ to yield the sequence r'(k), where

$$r'(k) = \sum_{\ell} r(k-\ell)\, c(\ell)$$

This is readily implementable by using high speed digital logic. The goal is to choose the c(k) to minimize the mean square error

$$E(r'(k) - a^{m(k)})^2$$

and then to decide m(k), such that $a^{m(k)}$ is closest to r'(k). If one defines

$$C(\theta) = \sum_{\ell} c(\ell)\, e^{-j2\pi\ell\theta}$$

$$\Phi(\theta) = \sum_{\ell} \langle \tilde{s}(t), \tilde{s}(t-\ell T) \rangle\, e^{-j2\pi\ell\theta}$$

then the mean square error can be expressed as

$$\int_0^1 d\theta\, (E(a^{m(k)})^2\, |C(\theta)\Phi(\theta)-1|^2 + |C(\theta)|^2\, \Phi(\theta))$$

if one assumes that the $a^{m(k)}$ are zero mean and linearly independent. The first term expresses the effect of intersymbol interference while the second term is due to noise. It is readily found that the optimum $C(\theta)$ is

$$\frac{E(a^{m(k)})^2}{E(a^{m(k)})^2\, \Phi(\theta) + 1}$$

which yields a mean square error equal to

$$\int_0^1 d\theta\, \frac{E(a^{m(k)})^2}{E(a^{m(k)})^2\, \Phi(\theta) + 1} \qquad (3)$$

Note that $\Phi(\theta)$, and thus $C(\theta)$, are constant if Nyquist's criterion is met. In that case only c(0) is non zero. As can also be observed from (3) this technique yields good results as long as $\Phi(\theta)$ does not have nulls. Fortunately this is the case on channels used for commercial data transmission.

A similar minimization can be done when c(·) is restricted to have only finitely many non zero coefficients. It yields a system of linear equations that the filter coefficients must satisfy.

The previous theory is interesting as it allows to quantify the mean square error, but at first sight it does not appear to be practically useful: it assumes that the channel response h(t) is known. If this were the case, we might as well design s(t) to avoid intersymbol interference altogether.

What makes it important, in fact what makes high data rate transmissions over telephone channels possible, is the discovery in the 1960's that the filter coefficients c(ℓ) can be adjusted by the demodulator itself to compensate for the effects of the channel response. This is called "adaptive equalization".

The basic observation is that the partial derivative of the mean square error with respect to c(ℓ) is equal to the expected value of the product

$2(r'(k) - a^{m(k)})\, r(k-\ell)$. r' and r can be measured. $a^{m(k)}$ is either known a priori if a training sequence is sent by the modulator, or can be estimated if the filter coefficients are already such that the probability of error is small. The demodulator can thus estimate a descent direction and update the filter coefficients while data is being transmitted, thus continously adjusting for variations in the channel response

Techniques for adaptive equalization and the study of their rate of convergence and steady state performances are still active topics of research, specially for channels that are rapidly varying or for which $\Phi(\theta)$ has nulls. The previous discussion constitutes only an introduction to the subject.

4. ERROR DETECTING CODES.

Shannon proved the surprising result that information can reliably be transmitted on a channel up to a maximum rate, called the channel capacity. At rates above capacity reliable communication is impossible, but an arbitrarily small probability of error can be achieved at rates below capacity, provided one is willing to jointly encode and decode large numbers of information bits. This requires complex and expensive equipment.

Shannon's theory has had little practical effects on commercial communication systems in that error correcting codes are little used. This is due to two reasons. First, the relationships between signal and noise powers, and data rate and channel bandwidth are such that the probability of error cannot be reduced easily by error correcting codes. Secondly, channels are typically two way. This property does not increase capacity but makes it simple to achieve reliability by using error detecting codes and requesting data retransmissions when errors are detected. The situation is just the opposite on the deep space channel; space probes make successful use of error correcting codes.

We will spend the rest of this section examining the fundamentals of the theory of error detecting codes.

Error detecting codes are usually implemented as polynomial codes, also known as cyclic redundancy codes. The operations described below are extremely easy to implement using digital logic.

A block of N binary digits that is to be protected by L check bits is viewed as a polynomial $M(x)$ of degree $N+L-1$ on the finite field with two elements. The N high order coefficients are the data bits themselves, the L low order coefficients being zero.

Both the transmitter and receiver agree on a generator polynomial $G(x)$ of degree L. The transmitter computes the remainder $R(x)$ of the division of $M(x)$ by $G(x)$ and transmits the $N+L$ coefficients of $M(x)-R(x)$.

The receiver just checks that the received polynomial is divisible by G(x). If it is, no error is assumed to have occurred and the N high order coefficients are retained as data bits. If it is not, then the receiver requests a retransmission of the block. Insuring that blocks are never lost or duplicated requires non trivial protocols between transmitter and receiver. They are beyond the scope of these notes.

We now turn our attention to the kinds of errors that are detected by these codes. An error pattern can also be viewed as a polynomial E(x) of degree N+L-1. It will be detected unless it is divisible by G(x).

Bursts of k errors, $k \leq L$ (including single errors) will always be detected as they have $E(x) = x^i(x^{k-1} \ldots + \ldots + 1)$, where i is the order of the lowest order coefficient in error. Such an E(x) is not divisible by G(x) if G(x) contains an x^0 term.

Similarly if G(1) is zero then all patterns of odd numbers of errors will be detected as they have E(1) equal to one.

Finally all patterns of two errors will be detected if x^i+1 is not divisible by G(x), $i \leq N+L-1$. This will be the case if G(x) contains a primitive polynomial of degree k, with $2^k-1 \leq N+L$. (An irreducible polynomial of degree k is primitive over the modulo 2 field if and only if it divides x^n-1 for no n less than 2^k-1. Primitive polynomials exist of all degrees.)

Generator polynomials used in practice have the three properties just mentioned; they are the product of a primitive polynomial of degree L-1 and (x+1).

Another view to look at the problem is to observe that when many errors occur (and modems often make large numbers of errors when they make any) there is about a chance in 2^L that the error polynomial will be divisible by G(x).

Older international communication standards specify generator polynomials of degree 16, so that a block hit by a large number of errors has about a chance in 65000 to be accepted as correct. This is too high in many applications, and more recent standards specify L equal to 32.

5. MULTI-USER SYSTEMS.

Users of data communication networks are often bursty; they may be silent most of the time! For that reason it makes economical sense to share communication lines between many streams of traffic. Together with each piece of data it is then necessary to send address information specifying to what stream the data belongs. This is the essential justification for the use of packet switched networks.

We introduce two new problems that occur in such systems. They are somewhat imprecise, and whatever is known about their solution does not have enough

structure to be concisely presented in these notes.

First the amount of overhead information carried in packets is far from negligible. It can dwarf the amount of data! The presence of this overhead loads the lines and causes undesirable delays.

How much of this overhead is necessary has not been properly quantified. In the simplest case imagine a group of sources with known data generation statistics. They are connected to a single transmitter and there are some buffers between the sources and the transmitter where data can be stored.

Whenever the line is idle the transmitter can select a source and transmit data contained in its buffer, together with addressing information. The transmitter faces two related questions: should data be transmitted at all, and if so, from what source? Also, how should the fact that data is being transmitted and the source identity be communicated to the receiver?

This formulation lacks an optimality criterion. Reasonable candidates would be to minimize the expected number of overhead bits, or to minimize the expected delay suffered by the data.

Solutions to these questions are not easy, although attempts have been made by using queueing and information theories [7] [8]. Some understanding of the tradeoffs involved is possible. Consider the policy known as synchronous time division multiplexing, where the transmitter divides time in equal slots and scans all the buffers cyclically, transmitting data from a source only during its assigned slots. No explicit address information is communicated to the receiver, but the system causes unnecessarily long delays when there are many sources and traffic is light. Data waits for its slot to come, while typically empty slots are being transmitted!

At the other extreme data could be sent in first come first served order, each being prefixed with the explicit address of its source (essentially forming packets). This scheme results in small delays in light traffic, but can be terrible in heavy traffic due to the effects of the addressing overhead.

From these examples it is clear that overhead information can be traded off for delay. This was pointed out in [7]. Further studies [8] have shown that protocols of the first type should be used when the number M of sources is much smaller than the expected number of bits N waiting in buffers. When M is much smaller than N the second type of protocol is excellent. What to do in the intermediate region is not clear.

Let us now introduce a second problem that has generated much excitement recently. It is known as the ALOHA channel problem, having originated in Hawaii. Again there are M sources producing data and sharing a communication channel. The novelty is that no single device can observe the state of all the buffers. Rather the sources can transmit whenever they wish and can also

observe the outcomes of all transmissions. These may be "idle" if no one transmits, "success", or "collision" if more than one source transmit. All collided messages must be retransmitted. This is a fair model of some radio and cable communication systems that are now in use.

The challenge is to design a way to use feedback information to maximize the rate at which successes occur, while keeping the delay small.

The most successful algorithms [9], [10], [11] are of the "binary splitting" type. A group of sources is allowed to transmit. If a "collision" occurs the group is split in two, each subgroup being allowed to transmit in turn. The process is repeated until all collisions have been resolved.

The excitement has come from the observation that this type of algorithm can achieve success rates in the vicinity of .5 while keeping the delay bounded no matter the number of sources M. Larger rates of success are possible but then the delay is not bounded as M increases.

This threshold effect is reminiscent of the concept of channel capacity, and much effort has been spent using a variety of techniques to characterize it. The threshold is known to be between .48 and .59 (see [12] and [13] and references therein).

REFERENCES

Good accounts of the theories of modulation, detection and coding, together with appropriate credits to their authors, can be found in classical texts like

1. A. J. Viterbi and J. K. Omura, Principles of Digital Communication and Coding, McGraw Hill, New York, New York, 1979.

2. R. G. Gallager, Information Theory and Reliable Communication, John Wiley & Sons, Inc., New York, New York, 1968.

Description of their applications to computer networks appears in

3. A. S. Tanenbaun, Computer Networks, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1981.

Recent papers on equalization, each containing more references to the relevant literature are

4. A. Gersho and T. L. Lim, "Adaptive Cancellation of Intersymbol Interference for Data Transmission", Bell System Technical Journal, 60, No. 11 (1981) (1997-2021).

5. D. Godard, "Channel Equalization Using a Kalman Filter for Fast Data Transmission", IBM J. Res. Develop., May 1974.

6. J. Proakis, "Advances in Equalization for Intersymbol Interference", Advances in Communication Systems, 4, Academic Press, 1975.

Multiuser systems are treated in

7. R. G. Gallager, Basic Limits on Protocol Information in Data Communication Networks, IEEE Trans. Inf. Th., IT-22, No. 4, (1976), 385-398.

8. P. A. Humblet, "Source Coding for Communication Concentrators", LIDS Technical Report ESL-R-798, M.I.T., Dept. of E.E.&C.S., January, 1978.

9. J. F. Hayes, "An Adaptive technique for Local Distribution", IEEE Trans. Commun., COM-26, (1978), 1178-1186.

10. J. Capetanakis, "Tree Algorithms for Packet Broadcast Channels", IEEE Trans. Inform. Theory, IT-2t, (1979), 505-515.

11. P. A. Humblet and J. Mosely, "Efficient Accessing of a Multiaccess Channel", presented at the IEEE Conf. Decision Contr., Albuquerque, NM, Dec. 1980.

12. R. Cruz and B. Hajek, "A New Upper Bound to the Throughput of a Multi-Access Broadcast Channel", IEEE Trans. on Inform. Theory, IT-28, No. 3, (1982) 402-405.

13. B. S. Tsybakov and V. A. Mikhailov, "An Upper Bound to Capacity of Random Multiple Access Systems", Presented at 1981 IEEE Inform. Theory Symp., Santa Monica, CA, Feb. 1981, Probl. Peredach, Inform, 17, No. 1, 90-95 (1981).

DEPARTMENT OF ELECTRICAL ENGINEERING
AND COMPUTER SCIENCE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS 02139