

AD A056345

AD No. _____
DDC FILE COPY

June 1978

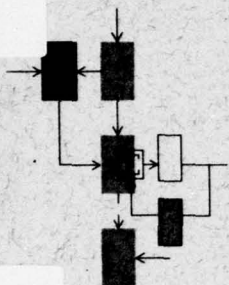
ESL-TH-825

12

LEVEL II

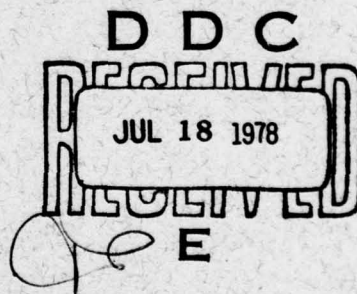
Research Supported By:

ARPA Contract ONR-N00014-64-C-1183



A LOWER BOUND TO PROTOCOL INFORMATION IN DATA NETWORKS

Charles Edward Rohrs



DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited

Electronic Systems Laboratory

MASSACHUSETTS INSTITUTE OF TECHNOLOGY, CAMBRIDGE, MASSACHUSETTS 02139

78 07 12 025

June, 1978

Report ESL-TH-825

A LOWER BOUND TO PROTOCOL INFORMATION
IN DATA NETWORKS

BY

Charles Edward Rohrs

This report is based on the unaltered thesis of Charles Edward Rohrs submitted in partial fulfillment of the requirements for the degree of Master of Science at the Massachusetts Institute of Technology in June, 1978. This research was conducted at the M.I.T. Electronic Systems Laboratory, with support extended by the Advanced Research Projects Agency under Contract ONR-N00014-64-C-1183.

DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited

Electronic Systems Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER <i>Masters thesis</i>
4. TITLE (and Subtitle) A LOWER BOUND TO PROTOCOL INFORMATION IN DATA NETWORKS		5. TYPE OF REPORT & PERIOD COVERED <i>Technical Report</i>
7. AUTHOR(s) <i>10</i> Charles Edward Rohrs		6. PERFORMING ORG. REPORT NUMBER <i>14</i> ESL-TH-825
9. PERFORMING ORGANIZATION NAME AND ADDRESS Massachusetts Institute of Technology Electronic Systems Laboratory Cambridge, Massachusetts 02139		8. CONTRACT OR GRANT NUMBER(s) ARPA Order No. 3045/5-7-75 <i>15</i> ONR N00014-64-C-1183 ARPA Order - 3445
11. CONTROLLING OFFICE NAME AND ADDRESS Defense Advanced Research Projects Agency 1400 Wilson Boulevard Arlington, Virginia 22209		10. PROGRAM ELEMENT PROJECT, TASK AREA & WORK UNIT NUMBERS Program Code No. 5T10 ONR Identifying No. 049-383
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Office of Naval Research Information Systems Program Code 437 Arlington, Virginia 22217		12. REPORT DATE <i>11</i> Jun 1978
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		13. NUMBER OF PAGES 53 <i>10</i> 56p.
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
18. SUPPLEMENTARY NOTES		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Concentrators Computer Networks Protocols Rate-Distortion Theory		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) At the present time, data communication networks transmit much information besides the actual user's messages. This 'extra' information is called protocol information. This thesis extends Gallager's initial work in providing an information-theoretic lower bound to how much of this protocol information is absolutely necessary for the proper operation of a network. The lower bound is a function of the average amount of time messages are allowed to be delayed before being transmitted. The bound suggests that the strategies considered by Gallager are close to optimal.		

A LOWER BOUND TO PROTOCOL INFORMATION
IN DATA NETWORKS

by

CHARLES EDWARD ROHRS

B.S., University of Notre Dame
(1976)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF

MASTER OF SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

(JUNE 1978)

Signature of the Author.....*Charles E. Rohrs*.....
Department of Electrical Engineering and Computer Science
May 12, 1978

Certified by.....*Robert P. Bellager*.....
Thesis Supervisor

Accepted by.....
Chairman, Department Committee

ACCESSION for	
NTIS	White Section <input checked="" type="checkbox"/>
DDC	Buff Section <input type="checkbox"/>
UNANNOUNCED	<input type="checkbox"/>
JUSTIFICATION.....	
BY.....	
DISTRIBUTION/AVAILABILITY CODES	
Dist.	AVAIL. and/or SPECIAL
A	

A LOWER BOUND TO PROTOCOL INFORMATION
IN DATA NETWORKS

by

CHARLES EDWARD ROHRS

Submitted to the Department of Electrical Engineering
and Computer Science

on May 12, 1978 in partial fulfillment of the requirements
for the Degree of Master of Science

ABSTRACT

At the present time, data communication networks transmit much information besides the actual user's messages. This "extra" information is called protocol information. This thesis extends Gallager's initial work in providing an information-theoretic lower bound to how much of this protocol information is absolutely necessary for the proper operation of a network. The lower bound is a function of the average amount of time messages are allowed to be delayed before being transmitted. The bound suggests that the strategies considered by Gallager are close to optimal.

Name and Title of Thesis Supervisor:

Robert G. Gallager

Professor of Electrical Engineering

TABLE OF CONTENTS

ABSTRACT	2
TABLE OF CONTENTS	3
ACKNOWLEDGEMENTS	4
CHAPTER I INTRODUCTION AND BASIC CONCEPTS	5
CHAPTER II EXAMPLES OF NETWORKS	8
CHAPTER III GALLAGER'S PROBLEMS AND RESULTS	17
CHAPTER IV THE NEW LOWER BOUND	23
CHAPTER V DISCUSSION AND SUGGESTIONS FOR FURTHER RESEARCH	47
REFERENCES	52

ACKNOWLEDGEMENTS

First and foremost, I thank Professor Robert Gallager, who not only provided the ground work for this thesis but also guided me through many obstacles along the way. I also thank Professor James Massey and Doctor Pierre Humblet for providing a sounding board for my ideas this last year and also helped immeasurably with the actual preparation of this document.

On a more personal level, I thank my wife, Cathy, who provided much needed support in more ways than one. I thank my parents, who provided more of my education than any school. I would also like to acknowledge certain members of the electrical engineering faculty at my alma mater, Notre Dame, who cared enough to push me towards my capabilities.

Finally, I thank Michelle Thompson for her dedication and patience in typing this thesis.

This work was supported by Bell Laboratories, by a research assistantship from the Electronic Systems Laboratory with funds provided by the Advance Research Projects Agency, and by a Vinton Hayes Fellowship.

Chapter I

Introduction and Basic Concepts

In the past decade, much of the work in communications has dealt with various problems involved with the implementation of data communication networks. In its general form a data communication network may be considered as a finite collection of nodes interconnected by communication links. Each node may have a finite number of sources and/or receivers connected to it. A source produces messages which must be transmitted through the network to a specified receiver.

We define a message as a finite sequence of binary digits. The type of source which is of greatest practical interest is that which produces messages whose length are stochastic, generally short compared to the length of the idle periods between messages, and whose starting times are random. We will call such a source sporadic, and will model the message starting times as a Poisson process. (Fuchs and Jackson, 1969)

It is the job of the nodes to route messages through the network. There are many options available as to exactly how these nodes will accomplish their task and much has been written on various aspects of these options.

One obvious fact is that more information must be transferred from the source to the nodes than just the data to be communicated to the receiver, e.g., there must be information that tells the node what is the destination of

the message. This thesis extends the work of (Gallager, 1976) and establishes a lower bound on how much of this extra information must be sent. We refer to this "extra information" as protocol information. The term "protocol" is often used in connection with any control information in the network, but we will reserve this term for that control information which is absolutely necessary for the network to transmit messages properly.

There are many problems in data networks whose solutions generate control information that will not be considered here. We now mention some of these problems and show that their needed information is separable from the basic protocol information. This will enable us to define more precisely that control information which we are considering to be protocol information.

First of all, we assume that the network topology is static, that is, we do not deal with the problems of adding or removing sources, receivers, or links. Control information normally generated to handle such problems can be viewed as ordinary messages generated by special sources in the network and need not concern us here. Other control information which we regard in the same manner is that necessary to insure user privacy, that used to control the flow of messages into the network, and that required to deal with flexible routing strategies. We also consider all error correcting and error detecting information to be

imbedded within the message itself; the major problem such error control procedures present to the network is that of the variable delay for processing and retransmission. We choose to ignore this problem and other problems which cause variable delay. We therefore assume that our messages are transmitted across the network with some fixed delay.

Now that we have mentioned that control information which is not of interest in this thesis, we next consider the types of control information which do come under our definition of protocol. Any information which the receiver learns simply by receiving the message is necessarily sent whether intentionally or not. Thus protocol information includes certain amounts of addressing information (where the message has arrived), starting time information, and message length information. In the next chapter we shall look at particular examples of networks and identify issues surrounding such protocol information.

Chapter II

Examples of Networks

Since we are trying to discover the amount of protocol required by every network, we first envision the simplest possible network, reasoning that this network will require the least amount of protocol. A very simple system would be that in which each source is allowed to communicate with only one receiver and this communication takes place over a dedicated communication link (or, more precisely, over a fixed fraction of the capacity of each link on the path between the source and receiver.)

The fact that each source can communicate with only one receiver is only a conceptual simplification. If a source actually wants to transmit to N receivers, it can conceptually be partitioned into N sources, each of which transmits only to one receiver. A receiver that wants to receive from more than one source can be similarly partitioned.

We shall call such a system a multiplex system because of its assignment of fixed fractions of each link capacity to each source-receiver pair.

At first glance, it would appear that this system uses no protocol information at all. Indeed, if each source were to produce messages which were always of the same length and produces these messages at regular intervals then this scheme requires no protocol information. Of course, this contradicts our assumption that the sources are sporadic.

We now consider the multiplex system with sporadic sources. The problem here is that the receiver must be able to tell when there is a message on the link and when the link is idle. Otherwise, it might interpret "noise" on the channel as a spurious message. One method of accomplishing this is for the transmitter to have a special idle character or "flag" which is repeatedly sent when it has no message to send and which cannot appear at the beginning of a message. The first absence of the idle character signals the start of a message and the next appearance signals the end of a message. Alternatively, the transmitter can use a special character or flag to signal explicitly the beginning of a message. Information must then also be provided about the length of each message by either prefixing the message with a header or by providing another (possibly different) flag at the end of the message. All such "extra" information is protocol information, necessary for proper operation of the network. Indeed, in a multiplex system, the amount of protocol information for each source-receiver pair is determined a priori as the amount of channel capacity assigned to the pair in excess of the source rate. If the rate of the source is known, we may set the channel capacity as close to the source rate as we like but, since the source is sporadic, the result is a queueing delay which increases as the channel capacity decreases. We note finally that the explicit protocol information in

the multiplex system consists of message starting time and message length information. No specific address information is required.

A second type of system, more commonly used than multiplex systems in data communication networks, is a message switching system. In this latter type of system, messages are preceded by an encoding of the receiver's address and of the message length. These "header" digits represent protocol information. Each intermediate node looks at the protocol bits and learns from this header where it must send the message in order to move it across the network. (How the node decides where to send the message once it has the receiver's address is a large and interesting problem unto itself. This problem, called the routing problem, will not concern us here.) Note that, to do its job, the node must also learn the message length so that it will know how many of the bits following the header should be sent toward the destination. When the message reaches its destination the receiver is alerted to this fact by recognizing his own address. Thus, for the receiver, the address in a switching system performs the same role as the starting flag in a multiplex system.

It is easy to see that, for a network with a large number of receivers, the information contained in the header will be much larger than that contained in a flag; thus, from a protocol standpoint, a multiplex system using starting

time protocols is much more efficient than a message switching system using address protocols. The problem is that, in order to reduce the protocols in the multiplex system, we must incur large delays.

We now explore a third type of system which, when the number of sources per node becomes very large and the capacity of the channel increases appropriately, can be made to exhibit negligible queueing delay. This system has an added advantage in that, if one chooses to allow delay, one can reduce the protocol information. This system is a variation of the multiplex system; we will call it a statistical multiplexing system.

Consider a large number n of sources which want to communicate over a common link. They can share the capacity of the link in the following manner:

Let each source have a queue associated with it and, in addition, let there be a central queue at the node. The transmitting node services the source queues cyclicly, servicing each queue every T_s seconds. Servicing a source queue corresponds to transferring the contents of the source queue to the central queue. (see Fig. 1) The contents of the central queue are then transmitted to the receiver with protocol information added to tell the receiving node from which source the message is coming and how long the message is.

Assume that each source is Poisson and on the average emits α messages per second. Given any service time T_s , if

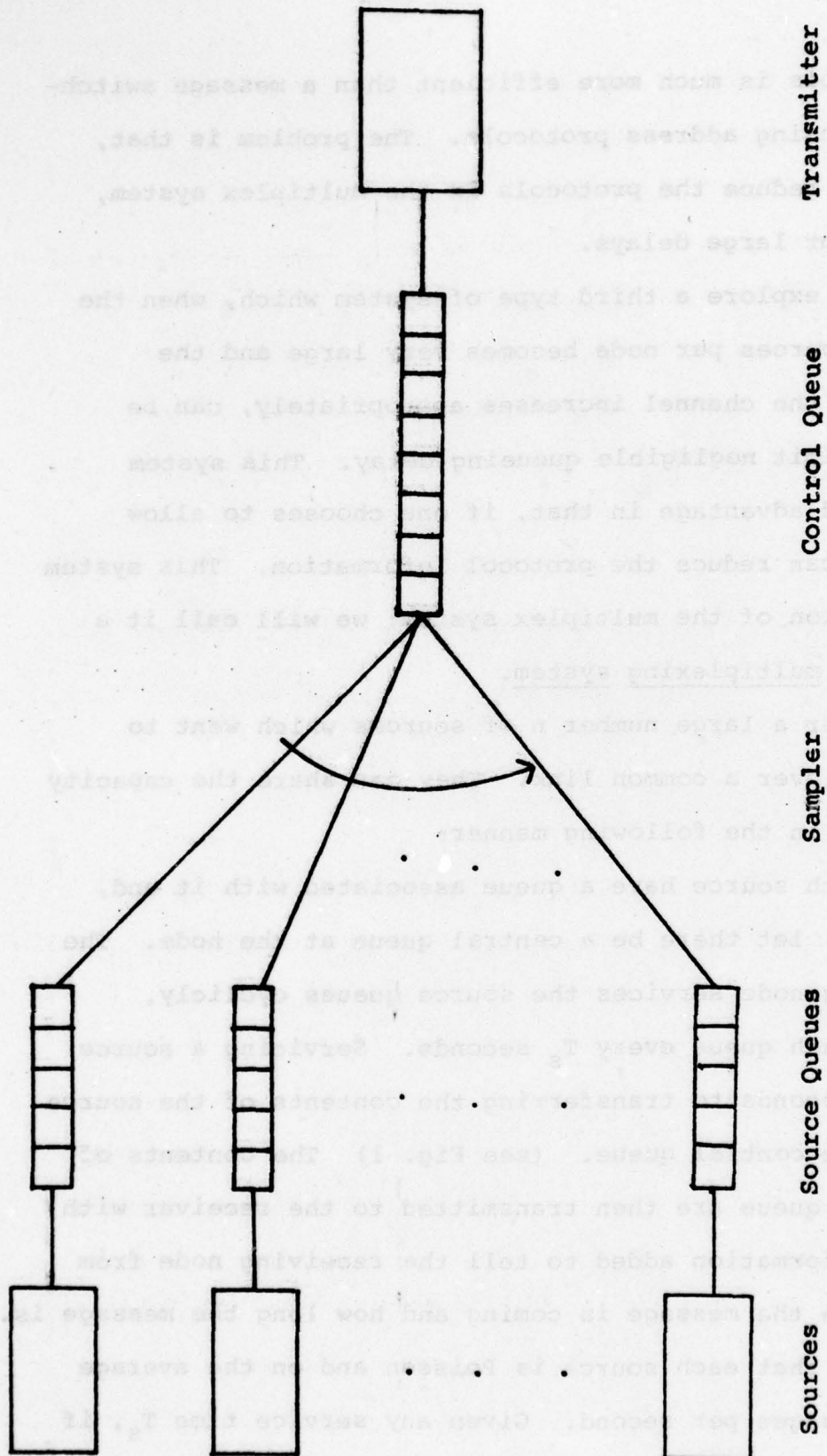


Figure 1: A Statistical Multiplexing System

the number of sources n is large enough, the law of large numbers implies that with high probability, the number of messages entering the central queue every T_s seconds divided by the number of sources is very close to αT_s . In order to be able to keep pace, the channel must have the capacity only a slight percentage in excess of that needed to transmit $n\alpha$ messages and their associated protocol information every second. No more than $\log_2 n$ bits per message are needed to specify the message origin. If the message length is a random variable M , then very close to $H(M)$ bits per message will be necessary to specify the message lengths, where $H(M)$ is the entropy of M . If we assume that bits in a message are zero or one with equal probability, then very close to $E(M)$ bits per message will be necessary to specify the messages themselves. We have again invoked the law of large numbers in these last two statements. We can therefore set the capacity to be only a slight percentage in excess of $n\alpha(E(M) + H(M) + \log_2 n)$ bits per second. Because we need exactly $\alpha(E(M) + H(M) + \log_2 n)$ bits per second per source with probability close to one we don't need much margin in order to guarantee that with high probability no message stays in the central queue more than T_s seconds.

For the statistical multiplexing system, the average delay per message in the central queue is $\frac{T_s}{2}$ seconds. The average delay per message in each source queue is also $\frac{T_s}{2}$ seconds,

so the total average delay per message is T_s seconds. As n increases without limit we can thus make the queuing delay T_s arbitrarily small if we so desire -- but we shall see that we also have the option of increasing the delay in order to reduce the necessary protocol information.

If T_s is small, the probability of finding a message in a source queue is small each time the source queue is sampled. Because of this, the messages in the central queue will be from a completely random selection of sources and we can do no better than to label each message with $\log_2 n$ bits per message to indicate its origin (or, equivalently, its destination.) If we allow T_s to increase, we can decrease this protocol information.

If T_s is large, there will with high probability be more than one message from each source entering the queue during each sampling period. The transmitting strategy can then be the following: transmit all the messages in the queue from source 1, with a header telling how many such messages there are and the message lengths; then do the same for source 2, and continue for each source in turn, returning to source 1 T_s seconds later after sending the messages from the last source. This cycle will also have a period of T_s seconds when there is no excess capacity. If there is excess capacity, idle symbols will need to be sent to fill out the T_s second period, but, since the amount of excess capacity needed is only a small percentage, these

idle symbols represent a negligible amount of protocol information per message. The protocol information in this situation is the message length information which we can do nothing about, and the n independent random variables indicating how many messages from each source arrived in the central queue in a period T_s . These random variables are Poisson-distributed with mean αT_s , so the message starting time protocol per message is $\frac{nH(P_{\alpha T_s})}{n\alpha T_s}$ where

$H(P_{\alpha T_s})$ is the entropy of a Poisson random variable with mean αT_s . We will show later that, for fixed α and large T_s , $H(P_{\alpha T_s}) \geq 1.41 + \frac{1}{2} \ln(\alpha T_s)$ nats so the starting time protocol per message goes to zero as T_s goes to infinity. In fact, this protocol decreases monotonically with increasing T_s . We see that in this statistical multiplexing system there is a direct tradeoff possible between starting time protocols and message delay. The nature of this tradeoff is the subject of this thesis. In the sequel, we will develop a lower bound on the amount of starting time protocol information necessary to operate within a certain constraint on the average message delay. This bound will hold for all networks with the proper source characteristics and will be independent of the particular network strategy used.

Note that in the above example, despite the fact that the sources are sharing the central queueing facilities, the

amount of protocol information per message in the network is equal to the sum of the protocol per message of each source divided by the number of sources, i.e., to the average protocol information per message for a single source. Thus, if we assume identical statistics for each source, then we need analyze only one source-receiver pair to establish a lower bound on protocol information.

Chapter III

Gallager's Problem and Results*

In this section we present the original problem as formulated by Gallager and review his results.

We assume that each source in the network emits messages at random times. The message starting times are assumed to form a stationary Poisson process. As mentioned in the previous section, we need to consider only one source-receiver pair. Let the Poisson process for this source have parameter α ; that is, α is the expected number of message emissions per second, and $1/\alpha$ is the mean interarrival time. We assume further that the entire message arrives instantaneously at a processor associated with the source. This processor has the option of holding the message for an unspecified length of time and then sends it along the network. This processor may be equivalently thought of either as part of the source or as part of the first node. The message is then transmitted instantaneously across the network to the receiver.

The messages from the source have independent lengths (M) described by a probability mass function $P_M(m)$. Introducing such delay does not change the necessary amount of protocol about messages lengths. The transmitted protocol information per message must be at least $H(M) = -\sum_{m=1}^{\infty} P_M(m) \log P_M(m)$.

* All references to Gallager refer to (Gallager, 1976) unless otherwise noted.

If we further assume that the message lengths are distributed geometrically with parameter ϵ (i.e., the average message is $1/\epsilon$), this expression reduces to $H(M) = 1/\epsilon H(\epsilon)$ where $H(x) = -x \log x - (1-x) \log(1-x)$ the binary entropy function. Let X_i , $i=1, 2, \dots$, be the message arrival times and Y_i , $i=1, 2, \dots$, be the times at which the intermediate processor sends the messages on their way. Since transmission is assumed to be instantaneous, Y_i is also the time the i th message arrives at the receiver. Let $D_i = Y_i - X_i$ be the delay for the i th message. The Poisson arrival assumption implies that the interarrival times, $T_i = X_i - X_{i-1}$, $i=1, 2, \dots$, are independent and each has a probability density

$$p_{T_i}(t) = \alpha e^{-\alpha t}.$$

For any given network, for any given scheme for transmitting messages across that network, and for any given N , there is a joint probability distribution $P_N(X^N; Y^N)$ on $X^N = (X_1, X_2, \dots, X_N)$ and $Y^N = (Y_1, Y_2, \dots, Y_N)$. This joint distribution must satisfy certain constraints. The marginal distribution on X^N must be consistent with the Poisson assumption. Also, the delay D_i must be non-negative with probability one for each i . $P_N(X^N; Y^N)$ defines a mutual information $I_{P_N}(X^N; Y^N)$ between X^N and Y^N . $I_{P_N}(X^N; Y^N)$ gives the information about the arrival times X^N that the receiver learns just by receiving the messages at times Y^N . This information is sent along with the messages whether the

network wants to send it or not! Therefore, this information serves as a lower bound to the amount of starting time protocol information necessarily sent. Note that we are not saying that this information is necessarily in a usable form or that the receiver desires to have it, but only that it must be sent whether we want to send it or not. (In the message switching network of the previous chapter, we need not explicitly send message starting time information, but such information is present nonetheless under the guise of addressing information.)

$P_N(X^N; Y^N)$ also establishes another quantity of interest, the expected delay per message, \bar{D}_N . This quantity is defined by

$$\bar{D}_N = \frac{1}{N} \sum_{i=1}^N E(D_i) \quad (1)$$

where $E(D_i)$ is the expected value of $D_i = Y_i - X_i$ with respect to the probability distribution $P_N(X^N; Y^N)$. Let $P_N(d)$ be the class of probability distribution $P_N(X^N; Y^N)$ which satisfy the above constraints and have $\bar{D}_N \leq d$. We want to find the minimum protocol information in a network with an expected delay per message less than or equal to d , so we minimize $I_{P_N}(X^N; Y^N)$ over $P_N(X^N; Y^N) \in P_N(d)$. This is the classical rate-distortion problem. Let

$$R_N(d) = \inf_{P_N(X^N; Y^N) \in P_N(d)} \frac{1}{N} I_{P_N}(X^N; Y^N) \quad (2)$$

be the N th order rate-distortion function. The lower bound we are looking for then is the rate-distortion function

$$R(d) = \lim_{N \rightarrow \infty} \inf R_N(d) \quad (3)$$

Unfortunately we have not been able to compute $R(d)$.
Gallager proved the following theorem which provides a lower bound to $R(d)$.

"Gallager's Lower Bound: $R_1(d)$ (as given by (2) with $N=1$), is a lower bound to $R_N(d)$, for all $N > 1$, to $R(d)$, and to the average protocol information per message about message arrival times between a source-receiver pair for Poisson message arrivals of rate α and expected delay d . Furthermore, $R_1(d)$ is given by

$$R_1(d) = -\log_2(1-e^{-\alpha d}) \text{ bits/message.} \quad (4)$$

The probability measure P_1 that achieves $R_1(d)$ is defined implicitly by

$$Y_1 = \max(X_1, d) + Z \quad (5)$$

where Z is a non-negative random variable, independent of X_1 , with probability density $P_Z(z) = (\alpha + \rho) \exp(-(\alpha + \rho)z)$ where ρ is given by

$$\rho = \frac{\alpha e^{-\alpha d}}{1 - e^{-\alpha d}} \quad (6)$$

Notice that Gallager's bound $R_1(d)$ on $R(d)$ is a function of αd . A little thought reveals that $R(d)$ itself will be a function of αd .

Gallager introduced two strategies for sending messages which, when there are a large number of sources at each

node, seem to be close to optimal. Strategy 1, which is better for small values of αd , yields the following upper bound for attainable protocol information per message.

$$I^{S1}(d) \leq \frac{H(P_{2\alpha d})}{2\alpha d} \text{ bits/message.} \quad (7)$$

$$\text{where } H(P_{2\alpha d}) = \sum_{n=0}^{\infty} \frac{e^{-2\alpha d} (2\alpha d)^n}{n!} \log \frac{n! e^{2\alpha d}}{(2\alpha d)^n} \quad (8)$$

Strategy 2, which is better for large αd yields;

$$I^{S2}(d) \leq \begin{cases} 2H\left(\frac{1}{e(e-1)}\right) & N=1 \\ 2H\left(\frac{1}{2N(N+4/3)}\right) & N \geq 2 \end{cases} \quad (9)$$

where N is chosen by

$$\alpha d \leq \left\lceil N - \frac{5N-3}{12N+16} \right\rceil \quad (10)$$

$R_1(d)$, $I^{S1}(d)$, and $I^{S2}(d)$ are plotted in Figure 2.

The behavior of $I^{S2}(d)$, together with some other information, led Gallager to conjecture that, for large αd , $R(d)$ goes to zero as $(\alpha d)^{-2}$ instead of as $e^{-\alpha d}$ like $R_1(d)$. In the next section, we find an improved lower bound to $R(d)$ which does indeed approach zero as $(\alpha d)^{-2}$ for large αd .

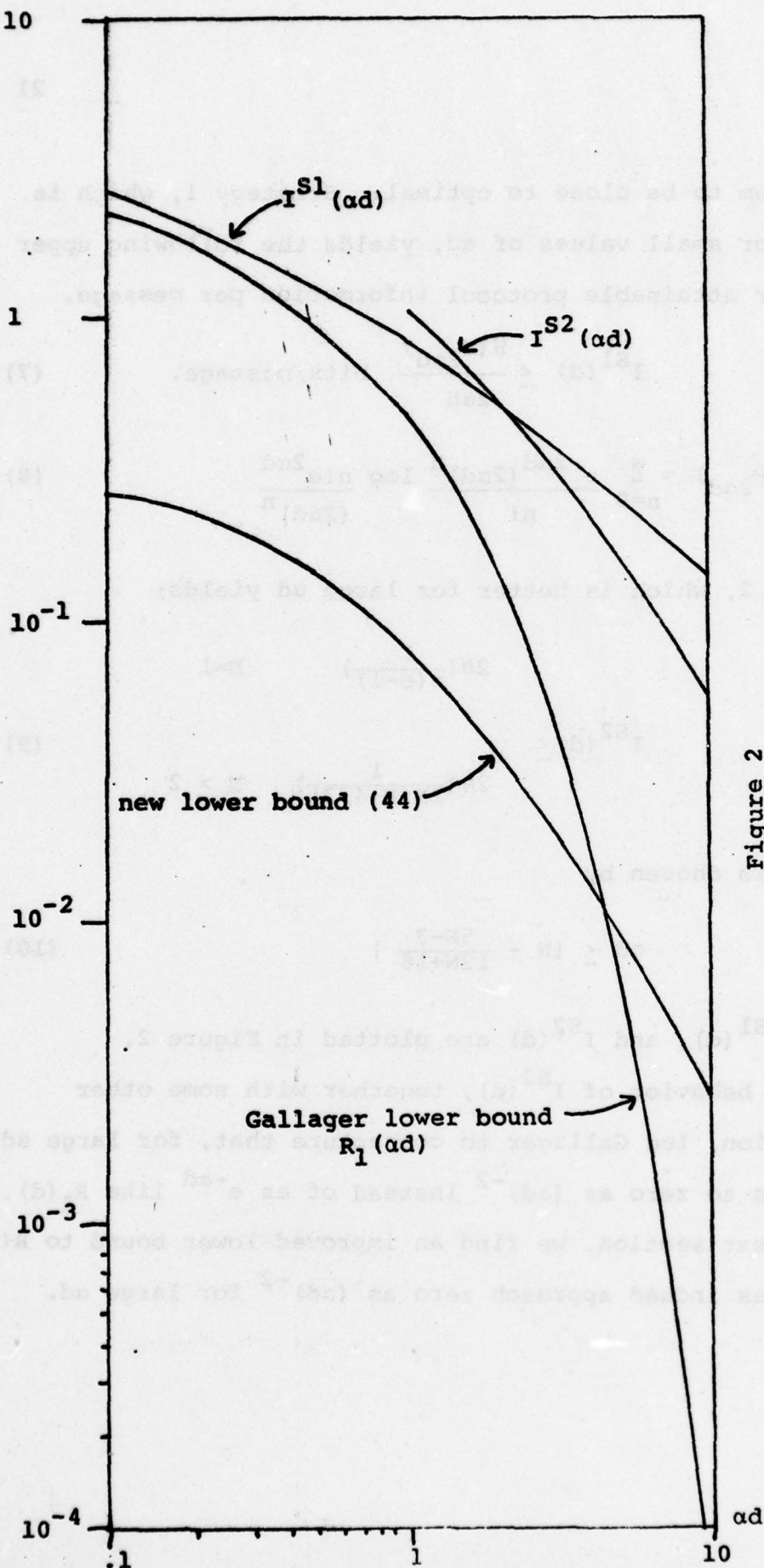


Figure 2

$R_1(ad)$, $I^{S1}(ad)$, $I^{S2}(ad)$, and the new lower bound (44) in nats

Chapter IV

The New Lower Bound

Our attack on the problem will be different from that of (Gallager, 1976) in that we will transform the system into a discrete-time system by sampling, use an information-theoretic approach similar to that used by (Humblet, 1978) on the fully discrete-time problem, and then return to the original continuous problem by letting the sampling period go to zero.

To turn the Poisson source into a discrete source, we make use of the fact that, at any time t , the probability of a message arrival in the next Δ seconds is independent of what happened before t . If the Poisson source has parameter α and if Δ is small, the probability of an arrival in any time period of length Δ is $\alpha\Delta$ while the probability of two or more arrivals in that period is negligible so the probability of no arrivals is $1 - \alpha\Delta$. Thus, if a Poisson source is checked every Δ seconds (with Δ small), the source can be viewed as a discrete-time source which at each time instant emits a one (there is a message available) with probability $\alpha\Delta$ and a zero (there is no message available) with probability $1 - \alpha\Delta$. The source output is thus a binary sequence $\{x_{\Delta}, x_{2\Delta}, \dots\}$ where $x_{n\Delta}=1$ indicates a message arrival in the time interval between $(n-1)\Delta$ and $n\Delta$.

We drop hereafter the Δ from our notation. We further introduce the notation $X_i^j = (x_i, x_{i+1}, \dots, x_{j-1}, x_j)$.

The message arriving each time instant from the source will be placed in a queue of unbounded length. At each time instant, the processor serving the queue will decide either to send a message (represented by $y_i = 1$) from the queue or not to send a message (represented by $y_i = 0$) in queue. When the decision is to send a message, the processor transmits the first message in the queue.

The number of messages in the queue at the end of time interval i , i.e., after a message has been placed in the queue if $x_i = 1$ and after a message has been removed from the queue and sent if $y_i = 1$, will be called the state of the queue at time $i\Delta$ and denoted by S_i . We have the following formula for updating the state:

$$S_{i-1} + x_i - y_i = S_i \quad (11)$$

We also have the constraint

$$S_i \geq 0 \quad (12)$$

which states the obvious fact that no message can be sent before it has arrived from the source.

The discrete-time constraint that results from the continuous-time average delay constraint is a constraint

upon the average length of time W that each message can spend in the queue. This constraint proves to be awkward when used directly. Thus, we invoke Little's formula (Kleinrock, 1975) to transform it into a constraint on the average number of one's allowed in the queue. Little's formula asserts that (under very weak conditions which are satisfied in our system)

$$L = \lambda W$$

where L is the average number of messages in the queue, λ is the arrival rate of the messages and W is the average length of time each message spends in the queue. Relating the quantities to our problem we have:

λ = number of messages arriving per time interval = $\alpha \Delta$

W = average number of time intervals each message

waits = d/Δ

Then L = average number of messages in the queue = λW

$$= \frac{\alpha \Delta d}{\Delta} = \alpha d$$

Thus, the discrete-time problem becomes:

$$\text{minimize } \liminf_{N \rightarrow \infty} \frac{1}{N} I(x_1^N; y_1^N | s_0) \quad (13)$$

where
$$\Pr(X_1^N = \underline{k}) = \prod_{i=1}^N \Pr(x_i = k_i) \quad (14)$$

and where $\Pr(x_i = 0) = 1 - \alpha\Delta$ and $\Pr(x_i = 1) = \alpha\Delta$, subject to the constraints

$$S_i \geq 0, \text{ for all } i \quad (12)$$

$$\limsup_{N \rightarrow \infty} \sum_{j=1}^N E[S_j - \alpha\Delta] \leq 0 \quad (15)$$

and

$$\begin{aligned} &\Pr(x_{in+r}^{in+r+n-1} | s_0, x_1^{in+r-1}, y^{in+r-1}) \\ &= \Pr(x_{in+r}^{in+r+n-1}). \end{aligned} \quad (16)$$

for each i, n , and $r, 0 < r \leq n$.

Equation (16) is a causality-type constraint which states implicitly that the decisions made by the processor to determine the y_i within a block of length n may not depend on x_j for any j which is larger than the largest i in the block. This allows the processor to "look ahead" but only up to n time intervals before deciding what the first y_i in the block should be. This constraint is weaker than a full causality constraint would be and thus using it will provide us with a lower bound to the actual rate-distortion problem as it should be posed.

Before beginning our derivation of the lower bound we introduce one more item of notation. The Hamming weight function of a binary vector is the number of ones in the vector, and we write $W(x_i^j)$ for the Hamming weight of x_i^j .

The expression (15),

$$\limsup_{N \rightarrow \infty} \sum_{j=1}^N E[S_j - \alpha d] \leq 0$$

implies that, for every n , there is a r , $0 < r \leq n$ such that

$$\limsup_{m \rightarrow \infty} E\left[\sum_{i=0}^m (S_{in+r} - \alpha d)\right] \leq 0 \quad (17)$$

for, if the left side of (17) were greater than zero for all r , $0 < r \leq n$, then the left side of (15) would also be greater than zero. We can now establish

Theorem 1: If S_0 is a non-negative, integer-valued random variable with $E(S_0) < \infty$ and if $\{x_i, y_i\}$, $i = 1, 2, \dots$, is a random sequence (possibly dependent on S_0), satisfying (12), (15) and (16), then

$$\begin{aligned} & \liminf_{N \rightarrow \infty} \frac{1}{N} (X_1^N; Y_1^N) \\ & \geq \max_n \inf_{\text{feasible distributions on } (\underline{X}, \underline{Y}, S)} \frac{1}{N} I(\underline{X}; \underline{Y} | S) \end{aligned} \quad (18)$$

where \underline{X} and \underline{Y} are binary random vectors of length n , S is a non-negative random variable, and where a feasible distribution is defined as some joint distribution

$$\Pr(\underline{X} = (j_1, j_2, \dots, j_n), \underline{Y} = (k_1, k_2, \dots, k_n), S = i)$$

satisfying

$$i. \Pr(\underline{X} = \underline{k}) = \prod_{i=1}^n \Pr(x_i = k_i) \text{ where}$$

$$\Pr(x_1 = 0) = 1 - \alpha\Delta \text{ and } \Pr(x_1 = 1) = \alpha\Delta;$$

$$ii. E(S) \leq \alpha d;$$

$$\text{and } iii. E[W(\underline{X}) - W(\underline{Y}) + S] \leq \alpha d.$$

Proof: The proof of the theorem is accomplished by a chain of inequalities which we will first present together and then explain one at a time.

$$\liminf_{N \rightarrow \infty} \frac{1}{N} I(\underline{X}_1^N; \underline{Y}_1^N | S_0) \stackrel{(1)}{\geq} \liminf_{m \rightarrow \infty} \frac{1}{nm} \sum_{i=0}^m I(\underline{X}_{in+r}^{in+r+n-1}; \underline{Y}_{in+r}^{in+r+n-1} | S_0, \underline{X}_1^{in+r-1}, \underline{Y}_1^{in+r-1})$$

(This is true for all $n, n \geq 1$ and all $r, 0 < r \leq n$)

$$= \liminf_{m \rightarrow \infty} \frac{1}{nm} \sum_{i=0}^m H(\underline{X}_{in+r}^{in+r+n-1} | S_0, \underline{X}_1^{in+r-1}, \underline{Y}_1^{in+r-1})$$

$$- H(\underline{X}_{in+r}^{in+r+n-1} | \underline{Y}_{in+r}^{in+r+n-1}, S_0, \underline{X}_1^{in+r-1}, \underline{Y}_1^{in+r-1})$$

$$\stackrel{(2)}{\geq} \liminf_{m \rightarrow \infty} \frac{1}{nm} \sum_{i=0}^m H(\underline{X}_{in+r}^{in+r+n-1} | S_{in+r-1})$$

$$- H(\underline{X}_{in+r}^{in+r+n-1} | \underline{Y}_{in+r}^{in+r+n-1}, S_{in+r-1})$$

$$= \liminf_{m \rightarrow \infty} \frac{1}{n} \sum_{i=0}^m \frac{1}{m} I(\underline{X}_{in+r}^{in+r+n-1}; \underline{X}_{in+r}^{in+r+n-1} | S_{in+r-1})$$

$$\stackrel{(3)}{\geq} \frac{1}{n} I(\underline{X}; \underline{Y} | S) \text{ where } \underline{X}, \underline{Y} \text{ and } S \text{ will be defined in the sequel,}$$

$$\stackrel{(4)}{\geq} \inf_{\substack{\text{feasible distributions} \\ \text{on } (\underline{X}, \underline{Y}, S)}} \frac{1}{n} I(\underline{X}; \underline{Y} | S)$$

Inequality 1

$$\liminf_{N \rightarrow \infty} \frac{1}{N} I(X_1^N; Y_1^N | S_0) \geq \liminf_{m \rightarrow \infty} \frac{1}{nm} \sum_{i=0}^m I(X_{in+r}^{in+r+n-1}; Y_{in+r}^{in+r+n-1} | S_0, X_1^{in+r-1}, Y_1^{in+r-1}) \quad (19)$$

Let X_1^k, X_{k+1}^N partition X_1^N in any way and Y_1^k, Y_{k+1}^N partition Y_1^N in the same way. Then

$$\begin{aligned} I(X_1^k X_{k+1}^N; Y_1^k Y_{k+1}^N | S_0) &= H(Y_1^k, Y_{k+1}^N | S_0) - H(Y_1^k Y_{k+1}^N | X_1^k X_{k+1}^N S_0) \\ &= H(Y_{k+1}^N | Y_1^k S_0) + H(Y_1^k | S_0) - H(Y_{k+1}^N | Y_1^k X_1^k X_{k+1}^N S_0) - H(Y_1^k | X_1^k X_{k+1}^N S_0) \\ &\geq H(Y_1^k | S_0) - H(Y_1^k | X_1^k S_0) + H(Y_{k+1}^N | X_1^k Y_1^k S_0) \\ &\quad - H(Y_{k+1}^N | Y_1^k X_1^k X_{k+1}^N S_0) \\ &= I(X_1^k; Y_1^k | S_0) + I(X_{k+1}^N; Y_{k+1}^N | S_0, X_1^k Y_1^k) \end{aligned}$$

As k is arbitrary, we can apply this argument repeatedly to get

$$\begin{aligned} \frac{1}{N} I(X_1^N; Y_1^N | S_0) &\geq \frac{1}{r + mn + (N - mn - r)} \\ &[I(X_1^r; Y_1^r | S_0) + \sum_{i=0}^m I(X_{in+r}^{in+r+n-1}; Y_{in+r}^{in+r+n-1} | S_0, X_1^{in+r-1}, Y_1^{in+r-1}) \\ &+ I(X_{mn+r}^N; Y_{mn+r}^N | S_0, X_1^{mn+r-1}, Y_1^{mn+r-1})] \end{aligned}$$

where m is greatest integer equal to or less than $\frac{N}{n}$.

Taking the limits of this expression as $N \rightarrow \infty$, and $m \rightarrow \infty$ and

noting that the first and last terms of the right side of it are finite and that $(N - mn - r) \leq n$, yields (19).

Inequality 2

$$\begin{aligned}
 & \liminf_{m \rightarrow \infty} \frac{1}{nm} \sum_{i=0}^m H(X_{in+r}^{in+r+n-1} | S_0, X_1^{in+r-1}, Y_1^{in+r-1}) \\
 & \quad - H(X_{in+r}^{in+r+n-1} | Y_{in+r}^{in+r+n-1}, S_0, X_1^{in+r-1}, Y_1^{in+r-1}) \\
 & \geq \liminf_{m \rightarrow \infty} \frac{1}{nm} \sum_{i=0}^m H(X_{in+r}^{in+r+n-1} | S_{in+r-1}) - H(X_{in+r}^{in+r+n-1} | \\
 & \quad Y_{in+r}^{in+r+n-1}, S_{in+r-1})
 \end{aligned} \tag{20}$$

S_{in+r-1} is determined from S_0, X_1^{in+r-1} , and Y_1^{in+r-1} , so the inequality on the second term on each side of (20) follows from the fact that removing conditioning cannot decrease entropy. The inequality on the first term uses the partial causality constraint (16), i.e., the fact that $X_{in+r}^{in+r+n-1}$ is independent of $S_{in+r-1}, S_0, X_1^{in+r-1}, Y_1^{in+r-1}$.

Inequality 3

$$\begin{aligned}
 & \liminf_{m \rightarrow \infty} \frac{1}{n} \sum_{i=0}^m \frac{1}{m} I(X_{in+r}^{in+r+n-1}; Y_{in+r}^{in+r+n-1} | S_{in+r-1}) \\
 & \geq \frac{1}{n} I(X; Y | S)
 \end{aligned} \tag{21}$$

with X, Y and S still to be defined. We know that the \liminf on the left side of (21) exists because the series for

each m is bounded by $H(x_i)$ above and by zero below. This implies that there is a subsequence which converges to the limit of the original sequence. Let Q_{m_v} , $v = 1, 2, \dots$, be that convergent subsequence; i.e.,

$$Q_{m_v} = \frac{1}{n} \sum_{i=0}^{m_v} \frac{1}{m_v} I(x_{in+r}^{in+r+n-1}, y_{in+r}^{in+r+n-1} | S_{in+r-1}) \text{ and}$$

$\lim_{v \rightarrow \infty} Q_{m_v} = \text{left side of (21)}$. For each j , let

$\Pr(X_{m_v} = j)$, $v = 1, 2, \dots$, be a sequence such that

$$\Pr(X_{m_v} = j) \triangleq \sum_{i=0}^{m_v} \frac{1}{m_v} \Pr(x_{in+r}^{in+r+n-1} = j) = \Pr(x_{in+r}^{in+r+n-1} = j)$$

because the x_i are independent and identically distributed.

Let $\Pr(y_{m_v} = k | X_{m_v} = j)$, $v = 1, 2, \dots$, be a sequence for each k and j such that

$$\Pr(y_{m_v} = k | X_{m_v} = j) \triangleq \sum_{i=0}^{m_v} \frac{1}{m_v} \Pr(y_{in+r}^{in+r+n-1} = k | x_{in+r}^{in+r+n-1} = j).$$

Finally, let $\Pr(S_{m_v} = \ell | X_{m_v} = j, y_{m_v} = k)$, $v = 1, 2, \dots$, be a sequence for each ℓ , j , and k combination such that

$$\Pr(S_{m_v} = \ell | X_{m_v} = j, y_{m_v} = k) = \sum_{i=0}^{m_v} \frac{1}{m_v} \Pr(S_{in+r-1} = \ell) \text{ where}$$

the dependence on X_{m_v} and y_{m_v} is implicit through the state update equation (11).

We now consider the sequence of probabilities

$p_{m_v} \triangleq \Pr(X_{m_v} = \underline{j}; X_{m_v} = \underline{k}; S_{m_v} = \underline{\ell}) = \Pr(S_{m_v} = \underline{\ell} | X_{m_v} = \underline{j}; Y_{m_v} = \underline{k}) \cdot \Pr(Y_{m_v} = \underline{k} | X_{m_v} = \underline{j}) \Pr(X_{m_v} = \underline{j})$ for $v = 1, 2, \dots$. This sequence is bounded between zero and one, so there exists at least one subsequence which converges. Let $\Pr(S_{m_{v_w}} = \underline{\ell}, X_{m_{v_w}} = \underline{j}, Y_{m_{v_w}} = \underline{k})$ be such a subsequence. Note now that not only does $p_{m_{v_w}}$ converge, but it must also converge to a probability distribution. Let \underline{x} , \underline{y} and \underline{s} be random vectors of length n , such that, for each $(\underline{j}, \underline{k}, \underline{\ell})$ combination,

$$\Pr(\underline{X} = \underline{j}, \underline{Y} = \underline{k}, \underline{S} = \underline{\ell}) = \lim_{w \rightarrow \infty} \Pr(X_{m_{v_w}} = \underline{j}, Y_{m_{v_w}} = \underline{k}, S_{m_{v_w}} = \underline{\ell}).$$

Since Q_{m_v} is a convergent sequence in v , every subsequence, in particular $Q_{m_{v_w}}$, converges to the same limit so $\lim_{w \rightarrow \infty} Q_{m_{v_w}}$ is also equal to the left side of (21).

We may now use the fact that the mutual information is convex with respect to the transition probabilities (Gallager, 1968) to get (21) with \underline{X} , \underline{Y} , and \underline{S} as defined above.

Inequality 4

$$\frac{1}{n} I(\underline{X}; \underline{Y} | \underline{S}) \geq \inf_{\text{feasible distribution on } (\underline{X}, \underline{Y}, \underline{S})} \frac{1}{n} I(\underline{X}; \underline{Y} | \underline{S}) \quad (22)$$

To show that (22) is true, we need only show that the distribution $\Pr(\underline{X}, \underline{Y}, \underline{S})$ as given in the section on Inequality 3 satisfies the three conditions for a feasible

distribution given in the theorem statement.

Condition i: $\Pr(\underline{X} = \underline{k}) = \prod_{i=0}^n \Pr(x_i = k_i)$ where $\Pr(x_i = 0)$

$= 1 - \alpha\Delta$ and $\Pr(x_i = 1) = \alpha\Delta$. This is satisfied trivially by the definition of X .

Condition ii: $E(S) \leq \alpha d$:

We first choose $\rho = r + 1$ where r , $0 < r \leq n$ is such that (17) is satisfied, i.e., such that

$$\limsup_{m \rightarrow \infty} E\left[\sum_{i=0}^m (S_{in+\rho-1} - \alpha d)\right] \leq 0$$

But, since the \limsup of a subsequence is at most equal to the \limsup of the original sequence,

$$\limsup_{w \rightarrow \infty} E\left[\sum_{i=0}^{m_{vw}} (S_{in+\rho-1} - \alpha d)\right] \leq 0. \quad (23)$$

$$\rightarrow E\left\{\limsup_{w \rightarrow \infty} \left(\sum_{i=0}^{m_{vw}} \frac{1}{m_{vw}} S_{in+\rho-1}\right) - \frac{m_v \alpha d}{m_{vw}}\right\} \leq 0$$

$$\rightarrow E[S - \alpha d] \leq 0$$

$$\rightarrow E[S] \leq \alpha d$$

Condition iii: $E[W(X) - W(Y) + S] \leq \alpha d$.

We start with (23), change the index to $j = i+1$ and divide

m_{vw} to get

$$\limsup_{w \rightarrow \infty} E \left[\left(\sum_{i=0}^{m_{vw}} \frac{1}{m_{vw}} S_{jn+\rho+n-1} - \frac{\alpha d}{m_{vw}} \right) + \frac{1}{m_{vw}} (S_{\rho-1} - \alpha d) \right] \leq 0$$

Since, in the limit, the separated term goes to zero, we have

$$\limsup_{w \rightarrow \infty} E \left[\sum_{j=0}^{m_{vw}} \frac{1}{m_{vw}} S_{jn+\rho+n-1} - \frac{\alpha d}{m_{vw}} \right] \leq 0 \quad (24)$$

By using the state update equation (11) repeatedly, we also have

$$S_{jn+\rho-1} + W(X_{jn+\rho}^{jn+\rho+n-1}) - W(Y_{jn+\rho}^{jn+\rho+n-1}) = S_{jn+\rho+n-1}.$$

Substituting this into (24) gives

$$\begin{aligned} & \limsup_{w \rightarrow \infty} E \left[\sum_{j=0}^{m_{vw}} \frac{1}{m_{vw}} (S_{jn+\rho-1} + W(X_{jn+\rho}^{jn+\rho+n-1}) \right. \\ & \quad \left. - W(Y_{jn+\rho}^{jn+\rho+n-1})) - \frac{\alpha d}{m_{vw}} \right] \leq 0 \\ & \rightarrow E \left[\limsup_{w \rightarrow \infty} \sum_{j=0}^{m_{vw}} \frac{1}{m_{vw}} S_{jn+\rho-1} + \frac{1}{m_{vw}} W(X_{jn+\rho}^{jn+\rho+n-1}) \right. \\ & \quad \left. - \frac{1}{m_{vw}} W(Y_{jn+\rho}^{jn+\rho+n-1}) - \frac{\alpha d}{m_{vw}} \right] \leq 0 \end{aligned}$$

$$\text{So } E[S + W(X) - W(Y) - \alpha d] \leq 0.$$

At this point we have shown $\liminf_{N \rightarrow \infty} \frac{1}{N} I(X_1^N; Y_1^N | S_0)$

$$\geq \inf_{\substack{\text{feasible distributions} \\ \text{on } (\underline{X}, \underline{Y}, S)}} \frac{1}{N} I(\underline{X}; \underline{Y} | S) \quad (25)$$

Since (25) is true for all n , we can take the maximum of the right side over n and obtain

$$\liminf_{N \rightarrow \infty} \frac{1}{N} I(X_1^N; Y_1^N | S_0) \geq \max_n \inf_{\text{feasible distributions on } (\underline{X}, \underline{Y}, S)} \frac{1}{n} I(\underline{X}; \underline{Y} | S) \quad (18)$$

which proves the theorem.

Hereafter we will abbreviate "feasible distributions on $(\underline{X}, \underline{Y}, S)$ " to "feas. dist."

We now present a lemma due to (Humblet, 1978) which will enable us to underbound the right side of (18).

Lemma 1: If \underline{X} and \underline{Y} are random vectors with some joint probability measure and thus some mutual information $I(\underline{X}; \underline{Y})$ and if $f(\cdot)$ and $g(\cdot)$ are deterministic functions, then

$$I(\underline{X}; \underline{Y}) \geq I(f(\underline{X}); g(\underline{Y})) \quad (26)$$

$$\begin{aligned} \text{Proof: } I(\underline{X}; \underline{Y}) &= H(\underline{X}) - H(\underline{X} | \underline{Y}) \geq H(\underline{X}) - H(\underline{X} | g(\underline{Y})) \\ &= I(\underline{X}; g(\underline{Y})) = H(g(\underline{Y})) - H(g(\underline{Y}) | \underline{X}) \geq H(g(\underline{Y})) - H(g(\underline{Y}) | f(\underline{X})) \\ &= I(f(\underline{X}); g(\underline{Y})). \end{aligned}$$

Applying (26) to (18),

$$\begin{aligned} \liminf_{N \rightarrow \infty} \frac{1}{N} I(X_1^N; Y_1^N | S_0) &\geq \max_n \inf_{\text{feas. dist.}} \frac{1}{n} I(\underline{X}; \underline{Y} | S) \\ &\geq \max_n \inf_{\text{feas. dist.}} \frac{1}{n} I(W(\underline{X}); W(\underline{Y}) | S) \end{aligned} \quad (27)$$

$$\begin{aligned} &= \max_n \inf_{\text{feas. dist.}} \frac{1}{n} (H(W(\underline{X}) | S) - H(W(\underline{X}) | W(\underline{Y}), S)) \\ &= \max_n \inf_{\text{feas. dist.}} \frac{1}{n} [H(W(\underline{X})) - H(S')] \end{aligned} \quad (28)$$

where $S' = W(\underline{X}) - W(\underline{Y}) + S$.

Equation (28) is true because, given $W(\underline{Y})$ and S , $H(\underline{X})$ and S' are identically distributed and so they have the same entropy. In addition,

$$\begin{aligned} & \max_n \inf_{\text{feas. dist.}} \frac{1}{n} [H(W(\underline{X})) - H(S')] \\ & \geq \max_n \left[\inf_{\text{feas. dist.}} \frac{1}{n} H(W(\underline{X})) - \sup_{\text{feas. dist.}} \frac{1}{n} H(S') \right] \quad (29) \end{aligned}$$

To overbound the last term of the right side of (29) and thus underbound the whole expression, we use the fact that S' is constrained by $E(S') \leq \alpha d$ and that S' is non-negative. We then maximize the entropy of S' given these constraints. The result must be at least equal to the sup over feasible distributions.

We now make use of a well-known result, (Gallager, 1968), but give the proof as it is not explicitly available in the literature.

Lemma 2: If S' is a non-negative random variable with $\Pr(S' = k) = p_k$ and if $E(S') \leq \alpha d$, then

$$H(S') \leq (1 + \alpha d) H\left(\frac{1}{1 + \alpha d}\right) \quad (30)$$

where $H(\cdot)$ is the binary entropy function.

Proof: We note that the maximum will occur when $E(S') = \alpha d$

so we wish to find $\max \sum_{k=0}^{\infty} -p_k \ln p_k = -\min \sum_{k=0}^{\infty} p_k \ln p_k$

subject to $\sum_{k=0}^{\infty} p_k = 1$, $\sum_{k=0}^{\infty} k p_k = \alpha d$, and $p_k \geq 0$ for all k . We

temporarily ignore the inequality constraint and form the Lagrangian:

$$L(p, \lambda, \mu) = \sum_{k=0}^{\infty} p_k \ln p_k + \lambda \left(\sum_{k=0}^{\infty} p_k - 1 \right) + \mu \left(\sum_{k=0}^{\infty} k p_k - \alpha d \right).$$

$$0 = \frac{dL}{dp_i} = 1 + \ln p_i + \lambda + \mu i$$

$$p_i = e^{-(1 + \lambda + i\mu)} \quad (31)$$

so all $p_i \geq 0$ and our inequality constraint will be satisfied.

$$0 = \frac{dL}{d\lambda} = \sum_{k=1}^{\infty} p_k - 1 = \sum_{k=1}^{\infty} e^{-(1 + \lambda + k\mu)} - 1$$

$$\sum_{k=1}^{\infty} e^{-k\mu} = e^{1+\lambda}$$

$$\frac{1}{1 - e^{-\mu}} = e^{1+\lambda} \quad (32)$$

$$0 = \frac{dL}{d\mu} = \sum_{k=0}^{\infty} k p_k - \alpha d - \sum_{k=0}^{\infty} k e^{-(1 + \lambda + k\mu)} - \alpha d$$

$$e^{-(1+\lambda)} \sum_{k=0}^{\infty} k (e^{-\mu})^k = \alpha d$$

$$\frac{e^{-\mu}}{(1 - e^{-\mu})^2} = e^{(1+\lambda)} \alpha d \quad (33)$$

Substituting (32) into (33) yields

$$\frac{e^{-\mu}}{(1 - e^{-\mu})^2} = \frac{1}{(1 - e^{-\mu})} \alpha d$$

$$e^{-\mu} = \frac{\alpha d}{1+\alpha d}$$

$$\mu = \ln \left(\frac{1+\alpha d}{\alpha d} \right)$$

$$e^{1+\lambda} = \frac{1}{1 - \frac{\alpha d}{1+\alpha d}} = 1 + \alpha d$$

$$\text{Thus,} \quad p_k = \frac{1}{1+\alpha d} \left(\frac{\alpha d}{1+\alpha d} \right)^k \quad (34)$$

$$\text{and } H(S) = \sum_{k=0}^{\infty} \frac{1}{1+\alpha d} \left(\frac{\alpha d}{1+\alpha d} \right)^k \left[\ln(1+\alpha d) + k \ln \frac{1+\alpha d}{\alpha d} \right]$$

$$= \ln(1+\alpha d) + \alpha d \ln \left(\frac{1+\alpha d}{\alpha d} \right) \quad \text{nats}$$

$$= (1+\alpha d) H \left(\frac{1}{1+\alpha d} \right) \quad \text{bits.}$$

We are now left with the problem of underbounding $\inf_{\text{feas. dist.}} H(W(\underline{X}))$. Note that the infimum is meaningless since the feasibility constraints determine the distribution on \underline{X} completely and therefore, they also determine the distribution on $W(\underline{X})$. But, since nothing is gained by trying to find $H(W(\underline{X}))$ in the discrete-time setting rather than the continuous-time setting, we will apply our limiting operation on the sampling time interval and return to the continuous-time problem.

To this point, we have shown that, for any Δ ,

$$\liminf_{N \rightarrow \infty} \frac{1}{N} I(X_{1\Delta}^{N\Delta}; Y_{1\Delta}^{N\Delta})$$

$$\geq \max_n \frac{1}{n} [H(W(\underline{X})) - \ln(1+\alpha\Delta) - \alpha\Delta \ln(\frac{1+\alpha\Delta}{\alpha\Delta})] \text{ nats per time interval (35)}$$

where we have returned Δ to our notation on the left side of (35) in anticipation of returning to the continuous-time problem, and where $W(\underline{X})$ is binomially distributed with parameters $p = \alpha\Delta$ and $N = n$, i.e.

$$\Pr(W(\underline{X}) = k) = \binom{n}{k} (\alpha\Delta)^k (1 - \alpha\Delta)^{n-k}.$$

Now in order to recover our original Poisson source, we must let $\Delta \rightarrow 0$, $n \rightarrow \infty$ and $N \rightarrow \infty$ so that $n\Delta \rightarrow t$ and $N\Delta \rightarrow T$.

We first divide in (35) by $\alpha\Delta$ to get

$$\liminf_{N\Delta \rightarrow \infty} \frac{1}{N\Delta\alpha} I(X_{\Delta}^{N\Delta}; Y_{\Delta}^{N\Delta})$$

$$\geq \max_n \frac{1}{n\Delta\alpha} [H(W(\underline{X})) - \ln(1+\alpha\Delta) - \alpha\Delta \ln(\frac{1+\alpha\Delta}{\alpha\Delta})] \text{ nats per message}$$

We now let $\Delta \rightarrow 0$, $n \rightarrow \infty$, $n\Delta \rightarrow t$, $N\Delta \rightarrow T$ as described above.

$$\liminf_{T \rightarrow \infty} \frac{1}{\alpha T} I(X^T; Y^T) \geq \max_t \frac{1}{\alpha t} [H(W(X^t)) - \ln(1+\alpha\Delta) - \alpha\Delta \ln(\frac{1+\alpha\Delta}{\alpha\Delta})] \text{ nats per message (36)}$$

where X^T represents a Poisson process with parameter α from time 0 to T , Y^T is some other stochastic process from time 0 to T and $W(X^t)$ is a Poisson random variable associated

with the Poisson process at time t . $W(X^t)$ is distributed as follows:

$$\Pr(W(X^t) = k) = \frac{(at)^k e^{-at}}{k!}.$$

It is the entropy of this random variable that we must lower bound, and we use the properties of the gamma function to assist us.

Lemma 3: The gamma function

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt, \quad x \geq 1 \quad (37)$$

has the properties that $\Gamma(1+x) = x!$ when x is a non-negative integer and, for $(x \geq 1)$, $\ln[\Gamma(x)]$ is convex U in x .

Proof: The first property is well-known (See Feller, 1968)

To prove convexity, we show that $\frac{d^2 \ln \Gamma(x)}{dx^2} \geq 0$ as follows:

$$\frac{d \ln \Gamma(x)}{dx} = \frac{\int_0^\infty (\ln t) t^{(x-1)} e^{-t} dt}{\int_0^\infty t^{(x-1)} e^{-t} dt}$$

$$\frac{d^2 \ln \Gamma(x)}{dx^2} = \frac{[\int_0^\infty t^{(x-1)} e^{-t} dt][\int_0^\infty (\ln t)^2 t^{(x-1)} e^{-t} dt]}{(\int_0^\infty t^{(x-1)} e^{-t} dt)^2}.$$

Defining $f(t) = (t^{(x-1)} e^{-t})^{\frac{1}{2}}$ and $g(t) = \ln t (t^{(x-1)} e^{-t})^{\frac{1}{2}}$ and using the Schwartz inequality

$$(\int_0^\infty f(t)g(t)dt)^2 \leq [\int_0^\infty f^2(t)dt][\int_0^\infty g^2(t)dt],$$

we have that $\frac{d^2 \ln \Gamma(x)}{dx^2} \geq 0$, so the Gamma function is convex U.

We now return to our entropy underbounding.

$$H(W(X^t)) = \sum_{n=0}^{\infty} \frac{(\alpha t)^n e^{-\alpha t}}{n!} [-n \log \alpha t + \alpha t + \log n!]$$

$$= -\alpha t \log \alpha t + \alpha t + \log \Gamma(n+1).$$

We now use the Stirling approximation (usually used for $n!$ but also true for non-integer values of the gamma function) (Feller, 1968).

$$\Gamma(1+x) \geq (2\pi)^{\frac{1}{2}} e^{-x} x^{(x+\frac{1}{2})}$$

$$\text{So } H(W(X^t)) \geq -\alpha t \log \alpha t + \alpha t + \frac{1}{2} \log(2\pi)$$

$$+ \alpha t \log \alpha t - \alpha t + \frac{1}{2} \log \alpha t$$

(38)

$$= \frac{1}{2} \ln 2\pi + \frac{1}{2} \ln \alpha t \text{ nats.}$$

This lower bound together with the true value of the entropy (for which there is no convenient analytical expression) are plotted against $\ln \alpha t$ in Figure 3.

This plot is on a semilog scale with (X) representing the true entropy and (0) representing the lower bound. When viewed as a function of $\log \alpha t$, the bound is a straight line with slope $1/2$. Investigation reveals that the difference between $H(W(X^t))$ and the lower bound reaches a minimum of approximately .355 at $\alpha t \sim .504$. In addition, when $H(W(X^t))$ is viewed as a function of $\ln \alpha t$, its slope is greater than $1/2$ for all $\alpha t > .504$. This means, that given any $\alpha t_0 > .504$,

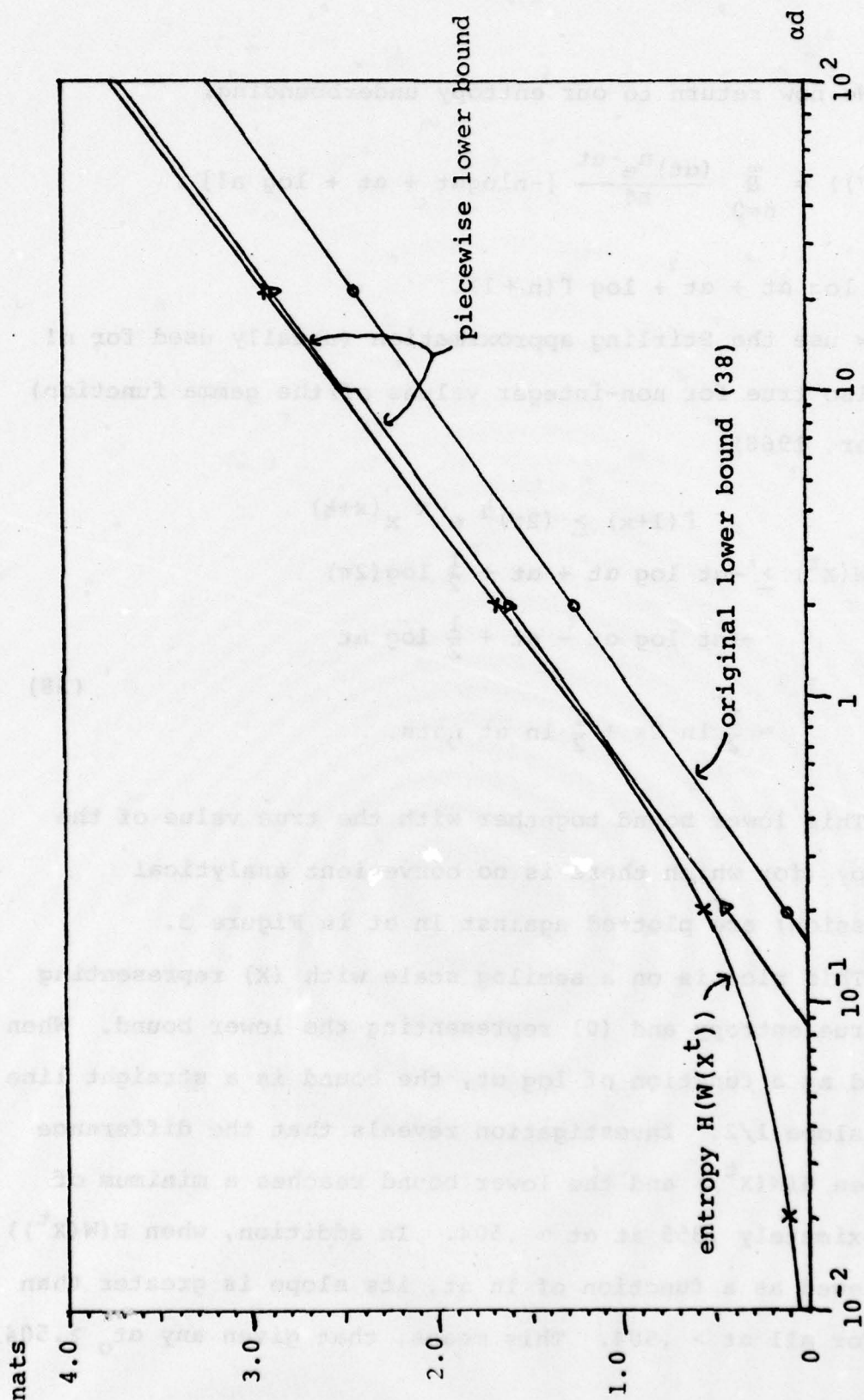


Figure 3

$H(W(x^t))$, Eq.(38), Eq.(39) uv. log at

we can bound $H(W(X^t))$ for all $at \geq at_0$ by adding $1/2 \ln at$ to $[H(W(X^{t_0})) - \frac{1}{2} \ln at_0]$. We note that $H(W(X^t))$ at $at = 10$ has a slope (when viewed as a function $\log at$) of approximately .509 and that $\lim_{at \rightarrow \infty} \frac{dH(W(X^t))}{d(\log at)} = .5$ with the limit approached monotonically from above for $at > 10$. This suggests that $at = 10$ is a legitimate breakpoint for a piecewise lower bound to $H(W(X^t))$ given by

$$H(W(X^t)) \geq \begin{cases} 1.274 + \frac{1}{2} \ln at \text{ nats for } 0 \leq at < 10 \\ 1.410 + \frac{1}{2} \ln at \text{ nats for } at \geq 10 \end{cases} \quad (39)$$

This bound is denoted by a (Δ) in Figure 3.

To show that this lower bound is quite tight, we make use of an upper bound on the entropy of an integer-valued random variable given its variance. This bound is due to Massey (private communication).

Lemma 4: If S is an integer-valued discrete random variable with variance σ^2 , then its entropy satisfies

$$H(X) \leq \frac{1}{2} \log(2\pi e(\sigma^2 + \frac{1}{12})).$$

Proof: We first form a continuous density function from the discrete distribution by

$$f_Y(y) = \Pr(X = x), \quad y \in (x - 1/2, x + 1/2).$$

The continuous entropy of the new random variable Y is equal to the discrete entropy of the original random variable

X, since

$$\begin{aligned} -\int_{-\infty}^{\infty} f_Y(y) \log f_Y(y) dy &= -\sum_{i=-\infty}^{\infty} \int_{i-\frac{1}{2}}^{i+\frac{1}{2}} f_Y(y) \log f_Y(y) dy \\ &= -\sum_{i=-\infty}^{\infty} \Pr(X=i) \log \Pr(X=i). \end{aligned}$$

Using Shannon's bound for continuous random variables (Gallager, 1968) we have

$$H(X) = H(Y) \leq \frac{1}{2} \log(2\pi e \sigma_Y^2) \quad (40)$$

Now $E(X) = E(Y) \triangleq m$. Moreover,

$$\begin{aligned} \sigma_Y^2 &= \sum_{k=-\infty}^{\infty} \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} (t-m)^2 p_k(t) dt \\ &= \sum_{k=-\infty}^{\infty} p_k \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} (t-m)^2 dt \\ &= \sum_{k=-\infty}^{\infty} p_k \left[(k-m)^2 + \frac{1}{12} \right] = \sigma_X^2 + \frac{1}{12}. \end{aligned}$$

Substituting in (40) gives the result.

In our case, the bound of Lemma 4 becomes

$$H(W(X^t)) \leq \frac{1}{2} \log(2\pi e(at + \frac{1}{12})) \approx 1.42 + \frac{1}{2} \log(at + \frac{1}{12}) \text{ nats} \quad (41)$$

showing that our lower bound (39) is very tight for large values of at .

We summarize our results as

Theorem 2: If $W(X^t)$ is a Poisson random variable with mean at , then

$$H(W(X^t)) \geq \begin{cases} 1.274 + \frac{1}{2} \ln at \text{ nats } 0 \leq at < 10 \\ 1.41 + \frac{1}{2} \ln at \text{ nats } at \geq 10 \end{cases} \quad (39)$$

$$\text{and } H(W(X^t)) \leq 1.42 + \frac{1}{2} \ln(at + \frac{1}{12}) \text{ nats} \quad (42)$$

Using the bound (39) in

$$\begin{aligned} \liminf_{T \rightarrow \infty} \frac{1}{\alpha T} I(X^T; Y^T) &\geq \max_t \frac{1}{\alpha t} [H(W(X^t)) - \ln(1+\alpha d) \\ &\quad - \alpha d \ln(\frac{1+\alpha d}{\alpha d})] \text{ nats per message} \end{aligned} \quad (36)$$

we obtain the bound,

$$\begin{aligned} \liminf_{T \rightarrow \infty} \frac{1}{T} I(X^T; Y^T) &\quad (42) \\ &\geq \max_t \frac{1}{t} \left(\begin{cases} 1.274 \\ 1.41 \end{cases} + \frac{1}{2} \ln at - f(\alpha d) \right) \text{ nats per message for } \begin{cases} 0 \leq at < 10 \\ at \geq 10 \end{cases} \end{aligned}$$

where $f(\alpha d) = \ln(1+\alpha d) + \alpha d \ln(\frac{1+\alpha d}{\alpha d})$.

Let c equal 1.274. We can then see that if we maximize (42) with respect to at the maximum occurs at

$$at = e^{2(f(\alpha d) - c) + 1}$$

This leaves us with:

$$\liminf_{T \rightarrow \infty} \frac{1}{\alpha T} I(X^T; Y^T) \geq \frac{\frac{1}{2e} e^{2c}}{e^{2f(\alpha d)}} \quad (43)$$

If we use the fact that $f(\alpha d) \leq \ln(e(1+\alpha d))$ and evaluate the numerator of (43), we arrive at

Theorem 3: $R(d)$ as given by Equation (2) is lower bounded by

$$R(d) \geq \frac{.318}{(1+\alpha d)^2} \quad \text{nats per message} \quad (44)$$

This bound (44) can be improved for $\alpha d \geq 10$ by using $C = 1.41$ instead of $C = 1.274$.

Note also that for large αd (44) approaches zero as $\frac{1}{(\alpha d)^2}$. The bound given by (44) is plotted in Figure 3

along with Gallager's lower bound (4), and the performance of Gallager's best strategy (minimum of (7) and (9) which provides an upper bound to the actual rate distortion function which is the solution to the minimum protocol problem.

Chapter V

Discussion and Suggestions for Further Research

Let us review what we have accomplished at this point. We have found a lower bound (44) to the continuous-time version of the problem constructed on page 25, i.e., we lower bounded the solution to Gallager's rate-distortion problem (page 18) with the additional constraint (corresponding to (16)):

For every i , the process x_{it}^{it+t} is independent of the processes x_{it-t}^{it} , y_{it-t}^{it} and the random variables S_0 and S_{it} , (45)
 where x_u^v represents the message arrival Poisson process from time u to time v , y_u^v represents the stochastic process for the message sending times from time u to time v and S_v represent the number of messages which have arrived but have not yet been sent at time v .

As pointed out in Chapter IV, this added constraint is a partial causality constraint. If we divide time up into blocks of t seconds, then (45) allows us to observe the X process for an entire block before deciding when to send messages in that block. It does not allow us to observe the X process for any time past the end of present block so the largest length of time that we are allowed to look ahead is t seconds.

Since we eventually would like to find the minimum protocol information required we would really like to solve

the following problem which we shall call the minimum protocol problem:

$$\text{Find } R(d) = \liminf_{N \rightarrow \infty} R_N(d)$$

$$\text{where } R_N(d) = \inf_{P_N \in \mathcal{P}_N(d)} \frac{1}{N} I_{P_N}(X^N; Y^N)$$

where $\mathcal{P}_N(d)$ is the class of probability measures P_N which satisfy the constraints for $\mathcal{P}_N(d)$ given in Chapter II (Gallager's constraints) and also satisfies the following constraint.

The complete causality constraint: Assume that, at any time t , i messages have arrived from the source and $j \leq i$ messages have been sent. If Y_{j+1} is the time the $(j+1)^{\text{st}}$ message is sent, then $\Pr(t \leq Y_{j+1} \leq t+\delta)$ must be independent of $(X_k, k = i+2, i+3, \dots)$ where X_k is the time the k^{th} message arrives from the source. But, $\Pr(t \leq Y_{j+1} \leq t+\delta)$ may depend on X_{i+1} only through the event $\{X_{i+1} > t\}$.

It can be readily seen that the complete causality constraint is a very difficult constraint to incorporate in a precise mathematical framework. We note that in the discrete-time framework the constraint becomes much simpler. If x_i equal one when a message arrives in the i^{th} time interval and zero if no message arrives in the interval and if y_i equals one if a message is sent in the i^{th} time interval and zero if no message is sent in the interval,

then the discrete-time complete causality constraint is:

y_i must be independent of $\{x_j, j = i+1, i+2, \dots\}$ and

$$\sum_{j=1}^i x_j + S_0 \geq \sum_{j=1}^i y_j, \text{ for all } i.$$

The fact that the partial causality constraint is a weaker constraint than the complete causality constraint means that the solution to the partially causal problem is a minimization over a larger set than the solution to minimum protocol problem and thus the partially causal solution is a lower bound to the minimum protocol solution. We also note that Gallager's optimal strategy ((9) and (11)) are completely causal and they provide an upper bound to the minimum protocol solution.

We have achieved a lower bound (44) to the partial causality problem and thus to the minimum protocol problem. As ad goes to infinity, this bound approaches zero as $(ad)^{-2}$. As can be seen from figure 3, there is still a gap between the lower bound (44) and the upper bound to the minimum protocol solution, i.e., the performance of the best attainable strategy (9). The upper bound (9) approaches zero as $\frac{\ln ad}{(ad)^2}$. Due to the fact that the lower bound was found using a relaxed causality constraint, we conjecture that the upper bound is closer to the actual minimum protocol solution than the lower bound.

In reviewing our derivation we see that the first time that we might violate a complete causality constraint was in applying the Hamming weight function W to our X and Y sequence in equation (27). We first note that this step is necessary to finding a bound on the second term, so our additional analysis could not be applied unchanged if the complete causality constraint were added.

The way that the complete causality constraint may be violated by applying the W function is that we are allowed to look ahead n time units at the beginning of a block before we must pick our Y vector for this block. What saves our bound from triviality is that we then maximize over n . For n small, the inequalities used to find the bound are not very tight. As n increases, the inequalities get better but eventually the effect of the non-causality begins to dominate.

There are two ways to approach this problem. One way is to start with the results from theorem 1 and try to proceed without using the Hamming weight function. This approach does not appear to be fruitful.

A second way is to try to extend Gallager's approach. It can be shown that $R_N(d)$ as given by Equation (2) provides a lower bound to $R(d)$ for any N . We have expended considerable effort to no avail trying to find $R_2(d)$ in the continuous-time case using standard rate-distortion

methods. These methods lead to integral equations involving the probability density, $W(Y_1, Y_2)$. We have been able to prove that, unlike $W(Y_1)$ in the $R_1(d)$ problem, $W(Y_1, Y_2)$ is ill-behaved. It seems that a more promising approach is to try to solve the $R_2(d)$ problem for the discrete case in order to obtain more insight before renewing the attack on the continuous-time problem. This approach has the advantage of allowing the use of computer methods and it also allows a simpler introduction of the complete causality constraint.

REFERENCES

- ABRAMSON, N. and F.F. Kuo, eds., 1973,
Computer Communication Networks. Englewood Cliffs. N.J.:
 Prentice-Hall, 1973.
- CAMRASS, R.J. and R.G. GALLAGER, 1976,
 "Encoding Message Length for Data Transmission," Report
 ESL-P-687, 1976.
- CERF, V.G., 1974,
 "An Assessment of ARPANET Protocols," Network Information
 Center, #04889, Stanford Research Institute, Menlo Park,
 CA, April 1974.
- DAVIES, D.W. and D.L. BARBER, 1973,
Communication Networks for Computers. New York: Wiley, 1973.
- FELLER, W., 1966,
An Introduction to Probability Theory and Its Applications,
 vol. II, New York: Wiley, 1966.
- _____, 1968,
Idem, vol. I, 3rd edition. New York: Wiley, 1968.
- FUCHS, E. and P.E. JACKSON, 1969,
 "Estimates of Distributions of Random Variables for Certain
 Computer Communications Traffic Models," Proc. ACM Symp.
Problems on the Optimization of Data Communication Systems,
 Pine Mountain, GA, pp. 202-225, Oct. 1969.
- GALLAGER, R.G., 1968,
Information Theory and Reliable Communication. New York:
 Wiley, 1968.
- _____, 1976,
 "Basic Limits on Protocol Information in Data Communication
 Networks," IEEE Trans. Inform. Theory, vol. IT-22,
 pp 385-398, July 1976.
- _____, 1978,
 "Variations on a Theme by Huffman" submitted to IEEE Trans.
Inform. Theory.
- _____, and D.C. VAN VOORHIS, 1975,
 "Optimal Source Codes for Geometrically Distributed Integer
 Alphabets," IEEE Trans. Inform. Theory, vol. IT-21, pp. 228-
 230, March 1975.

- HUFFMAN, D.A., 1952,
"A Method for the Construction of Minimum Redundancy Codes,"
Proc. IRE, vol. 51, pp. 251-252, Sept. 1952
- HUMBLET, P.A., 1978,
"Source Coding for Communication Concentrators,"
Ph.D. Dissertation, Dept. of Elec. Eng. and Comp. Science,
M.I.T., Cambridge, MA, 1978.
- KLEINROCK, L., 1965,
Queueing Systems, vol. I: Theory. New York: Wiley, 1976.
- _____, 1976,
Idem, vol. II: Computer Applications. New York: Wiley,
1976.
- _____, NAYLOR, W.E. and H. OPDERBECK, 1976,
"A Study of Line Overhead in the ARPANET," Commun.Ass.
Computing Machinery, vol. 19, pp. 3-13, Jan. 1976.
- _____, and H. OPDERBECK, 1977,
"Throughout in the ARPANET - Protocols and Measurement,"
IEEE Trans. Communications, COM-25, pp. 95-103, Jan. 1977.
- LITTLE, J.D.C., 1961,
"A Proof of the Queueing Formula: $L = \lambda W$," Operations
Research, vol. 9, pp. 383-389, 1961.
- RUBIN, I., 1976,
"Data Compression for Communication Networks: The Delay-
Distortion Function," IEEE Trans. Inform. Theory, vol. IT-22,
pp. 655-664, Nov. 1976.
- RUDIN, W., 1966,
Real and Complex Analysis. New York: McGraw-Hill, 1966.
- SCHWARTZ, M., 1977,
Computer Communication Network Design and Analysis. Englewood
Cliffs, N.J.: Prentice-Hall, 1977.
- VANDERMEY, J.E., 1976,
"The Architecture of a Transparent Intelligent Network,"
Proc. National Telecommunication Conference, 1976.

Distribution List

Defense Documentation Center Cameron Station Alexandria, Virginia 22314	12 copies
Assistant Chief for Technology Office of Naval Research, Code 200 Arlington, Virginia 22217	1 copy
Office of Naval Research Information Systems Program Code 437 Arlington, Virginia 22217	2 copies
Office of Naval Research Code 1021P Arlington, Virginia 22217	6 copies
Office of Naval Research Branch Office, Boston 495 Summer Street Boston, Massachusetts 02210	1 copy
Office of Naval Research Branch Office, Chicago 536 South Clark Street Chicago, Illinois 60605	1 copy
Office of Naval Research Branch Office, Pasadena 1030 East Green Street Pasadena, California 91106	1 copy
New York Area Office (ONR) 715 Broadway - 5th Floor New York, New York 10003	1 copy
Naval Research Laboratory Technical Information Division, Code 2627 Washington, D.C. 20375	6 copies

Dr. A. L. Slafkosky Scientific Advisor Commandant of the Marine Corps (Code RD-1) Washington, D.C. 20380	1 copy
Office of Naval Research Code 455 Arlington, Virginia 22217	1 copy
Office of Naval Research Code 458 Arlington, Virginia 22217	1 copy
Naval Electronics Laboratory Center Advanced Software Technology Division Code 5200 San Diego, California 92152	1 copy
Mr. E. H. Gleissner Naval Ship Research & Development Center Computation and Mathematics Department Bethesda, Maryland 20084	1 copy
Captain Grace M. Hopper NAICOM/MIS Planning Branch (OP-916D) Office of Chief of Naval Operations Washington, D.C. 20350	1 copy
Mr. Kin B. Thompson Technical Director Information Systems Division (OP-91T) Office of Chief of Naval Operations Washington, D.C. 20350	1 copy
Advanced Research Projects Agency Information Processing Techniques 1400 Wilson Boulevard Arlington, Virginia 22209	1 copy