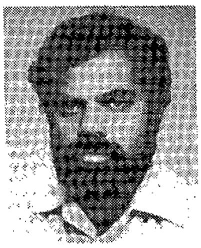
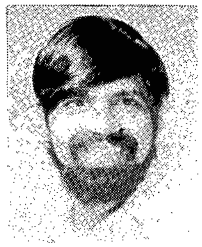
[6] Y. Hatori and H. Yamamoto, "Direct frame to frame coding of NTSC color television signals (simulation results)," Inst. TV Eng., Japan, Monograph, Sept. 25, 1975.
[7] A. N. Netravali and C. B. Rubinstein, "Luminance adaptive coding of chrominance signals," *IEEE Trans. Commun.*, vol. COM-27, pp. 703–710, Apr. 1979.

★



K. A. Prabhu (S'79–M'80) received the Ph.D. degree in electrical engineering from the University of Pittsburgh, Pittsburgh, PA, in 1980.

He is currently with the Television Transmission Department, Bell Laboratories, Holmdel, NJ. His research interests are in picture processing and digital video transmission.



A. N. Netravali (S'67–SM'78) received the B.Tech. (Honors) degree from the Indian Institute of Technology, Bombay, India, in 1967, and the M.S. and Ph.D. degrees from Rice University, Houston, TX, in 1969 and 1970, respectively, all in electrical engineering.

He worked from 1970 to 1972 on problems related to filtering, guidance, and control for the Space Shuttle. Since 1972 he has been at Bell Laboratories, Holmdel, NJ, working on various aspects of signal processing. He is at present Head of the Visual Communication Research Department. He was a Visiting Professor in the Department of Electrical Engineering at Rutgers University, New Brunswick, NJ.

Dr. Netravali serves on the Editorial Board of the PROCEEDINGS OF THE IEEE, and is the Editor for Signal Processing and Communication Electronics for the IEEE TRANSACTIONS ON COMMUNICATIONS. He received the Donald G. Fink award for the best review paper in any IEEE journal. He is a member of Tau Beta Pi and Sigma Xi.

Concise Papers

Stream Traffic Communication in Packet Switched Networks: Destination Buffering Considerations

WILLIAM E. NAYLOR, MEMBER, IEEE, AND LEONARD KLEINROCK, FELLOW, IEEE

Abstract—**In this paper we consider the problem of sending a stream of data (speech, for example) through a packet-switched network which introduces variable source-to-destination delays for different packets of the stream. Ideally, this delay difference should be smoothed so as to preserve the continuity of the stream. We investigate an adaptive destination buffering scheme which may be used to achieve the smoothing of the output stream. The scheme uses delay information, measured for previous streams, in order to compute destination buffering information. Specifically, of the last m packet delays, one discards the largest k and then the range of this partial sample is used for the destination wait time D. We obtain a rule of thumb for choosing m and k, and demonstrate its applicability on some empirical delay distributions from ARPANET measurements. It is, in general, necessary to deal with discontinuities which occur even after smoothing. To this end, we consider two possible playback schemes: method E (time expanded in order to preserve information) and method I (late data ignored in order to preserve timing). The two methods are at opposite ends of a continuum of possible playback schemes. We study the implication of methods E and I on the choice of smoothing parameters and establish a foundation for evaluating all schemes in this continuum.**

Paper approved by the Editor for Computer Communication of the IEEE Communications Society for publication without oral presentation. Manuscript received March 16, 1981; revised April 6, 1982. This work was supported by the Advanced Research Projects Agency of the Department of Defense under Contract MDS 903-77-C-0272.

W. E. Naylor is with the Information Sciences Research Office, The Aerospace Corporation, Los Angeles, CA 90009.

L. Kleinrock is with the Department of Computer Science, University of California, Los Angeles, CA 90024.

I. INTRODUCTION

Let us begin by defining stream traffic. Stream traffic is characterized by the following three properties: 1) small response time and moderate throughput requirements, 2) important timing constraints, and 3) redundant information content. Property 1 allows for the possibility of real-time interactive communication between two or more locations. This property alone makes stream traffic distinguishable from the two classical forms of data communication. Indeed, packet-switched networks have, in general, been designed to carry traffic which has traditionally been classified into two categories: a) LD—low delay (interactive), and b) HT—high throughput (file transfer). As noted by Cohen, Opderbeck, and Kleinrock [10], stream traffic communication falls into yet another category, c) ST—stream traffic, requiring both low delay and moderate throughput. Not only are the transmission requirements of stream traffic unique, but the information itself is of a somewhat different nature than the usual data communication. Property 2 indicates that each unit (bit, if you will) of information has an associated (possibly implied) time stamp and that the relative timing of the information should be preserved as well as possible by the transmission media. If a

0090-6778/82/1200-2527$00.75 © 1982 IEEE

unit of information does not arrive at or before the time that the receiver expects it, a *gap* will occur in the output of the stream. Such gaps are undesirable since, to some extent, they destroy the rhythm of the output and thus hinder the intelligibility of the information. Unlike traditional LD and HT data communication, the information in stream traffic is somewhat redundant. The communication medium may lose a small fraction of the information without seriously affecting the quality.

In communicating stream information via a packet-switched network, rather than the traditional circuit-switched (or dedicated) network, there arise some unique problems which require solutions. This paper discusses methods of smoothing the packet delay variability in the output by destination buffering. By delaying the output of the first message of a stream by an amount D, one may limit the frequency and duration of gaps in the output. The method discussed here corresponds to that of a "null timing information (NTI) device," as described in [1]. We investigate here the adaptive setting of the delay parameter D (T in [1]). For the purpose of discussion here we define *gap probability* G to be the average fraction of packets which arrive at an empty output process. A stream source typically has periods of activity and inactivity (e.g., called "talk-spurts" and "silence" in speech). We shall call a period of activity by a stream source a *sentence*. As the destination buffering delay of the first message is increased, the frequency and duration of gaps decreases. Therein lies the tradeoff which is examined in this paper, namely, gaps versus delay.

II. PLAYBACK METHODS

Unless D is very large, there is a nonzero probability that a gap will still occur. The purpose of this section is to describe two methods for dealing with this eventuality. The first method (*method E*) would *expand* the playout time of the sentence in order to include all packets in the output process. In method E a late packet (i.e., one which arrives to find an empty output process) will result in a gap and delay all successive packets of that sentence. This is similar to the approach taken in [2]. The second method (*method I*) preserves the timing at the expense of *ignoring* some late arriving packets (or partial packets). In method E each late packet results in a gap as the total length of time for a sentence is preserved. A similar scheme is described in [3].

These two approaches lie at opposite ends of a continuum of choices for dealing with gaps. For a constant network delay (e.g., a dedicated point-to-point channel) the two extremes are equivalent. Also, if D is infinite, the extremes coincide. However, with finite D and variable delay, the extremes separate. The separation increases as delay variability increases or as D decreases. With finite delay variability and finite D, one can envision methods which lie between the two extremes. For example, one may wish to discard only those packets which arrive both late and out of order. Such a scheme has properties of both methods E and I. The time axis is expanded when a packet is just late, yet some data may be discarded in order to preserve "reasonable" timing. Another example is to limit the expansion of time and/or the fraction of discarded data to a certain amount and switch methods if a threshold is exceeded.

An important consideration, which shall not be discussed here, is the filling of gaps (i.e., with silence or something else). Several alternatives have been used and are discussed in [3].

III. A DELAY PREDICTOR

In this section we define the m-sample partial range which we have used as a method of adaptively setting the value of D. Estimation of a "good" value for D is the key to the success of this class of algorithm. In both playback methods, if the delay of a packet within a sentence exceeds the delay of the first packet by more than D, at least one gap will occur. Therefore, it would seem reasonable to try to choose D equal to the value by which the maximally delayed packet exceeds the delay of the first packet. The total range of the previous m ($m > 1$) samples provides a pessimistic estimate for this value. The partial range is used to eliminate isolated cases of extremely large delay without significant impact on gap probability.

Let $u(-m)$, $u(-m+1)$, $\cdots$, $u(-1)$ be the m packet delays just prior to the beginning of a sentence. We define the range $D(M, 0)$ of these m packet delays as the maximum difference between delays. We define the partial range $D(m, k)$ as

$$D(m, k) = u(I_{k+1}) - u(I_m), \qquad k = 0, \cdots, m-1$$

where $u(I_k)$ is the kth largest of the $\{u(i)\}$, $i = -m, \cdots, -1$.

It should be noted that D can be computed by the receiver, based only on delay differences. This calculation is possible if packets contain generation time stamps $g(i)$ and if a clock at the destination, running at roughly the same speed as the source clock, marks packets as they arrive at times $a(i)$. The following argument shows that there is no need for the clocks to be synchronized. Let a^* be the instant on the destination time line which corresponds to $g(i-1)$; then

$$u(i) = a(i) - a^*$$

$$u(j) = a(j) - [a^* + g(j-1) - g(i-1)].$$

Therefore,

$$u(i) - u(j) = a(i) - a(j) - [g(i-1) - g(j-1)].$$

The receiver need deal only with the m values $a(-m) - g(-m-1), \cdots, a(-1) - g(-2)$.

IV. ANALYSIS

An analysis of the performance of the adaptive receiving techniques is presented in this section. We begin by considering some assumptions which render the system tractable for subsequent analysis. Our model of the system is pictured in Fig. 1. A period of activity (i.e., a sentence) is initiated by the sender at some time. (This corresponds to the detection of nonsilence in speech, for example.) As time progresses the sentence is broken up into a sequence of segments called packets. A finite amount of time $f(i)$ is required to fill packet i. Also associated with each packet i is a network transit delay $y(i)$. Then, let $u(i) = f(i) + y(i)$ be the source-to-destination delay of packet i. We assume that the $u(i)$ are independent and identically distributed (i.i.d.) with distribution function S. Following the source-to-destination delay, there is a (possibly zero) destination buffering delay for each packet. In particular, this buffering delay takes the value D for the first packet of a sentence. It is assumed that each packet requires $f(i)$ to empty (e.g., a particular packet contains a specific duration of speech). If a packet arrives after its predecessor has completed

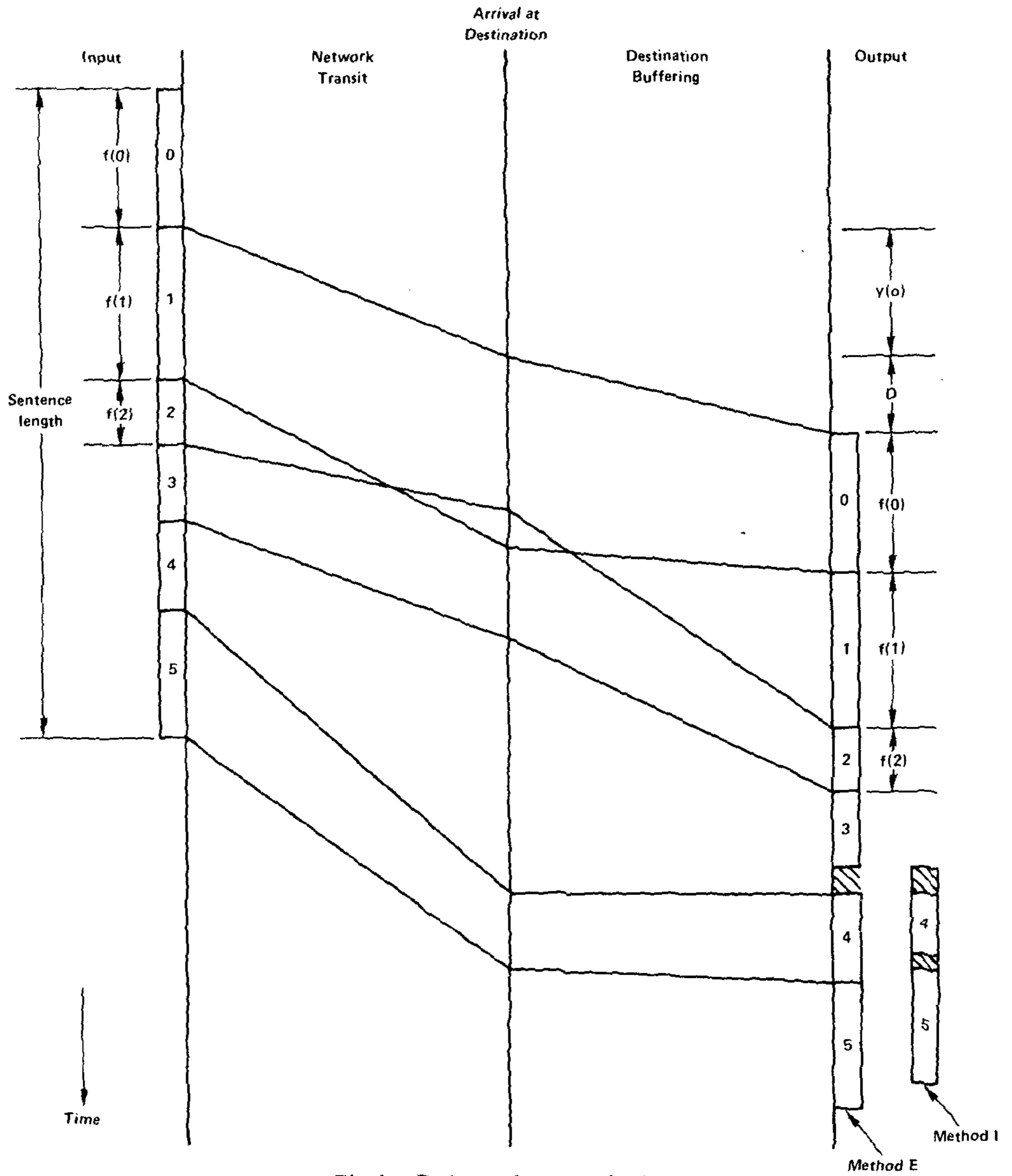


Fig. 1. End-to-end communication.

emptying, then a gap occurs in the output stream (indicated in Fig. 1 by crosshatching at the output).

A. Gap Probability for Playback Method E

A gap occurs in method E at packet i $(i > 0)$ if and only if $u(i) > u(0) + D$ and $u(i) > u(1), u(2), \cdots, u(i-1)$! Therefore, the average fraction of packets at which a gap occurs with method E in a sentence of length n messages is given by

$$\Pr[\text{gap}/n] \approx \frac{1}{n} \sum_{j=0}^{n-2} \int_{D=0}^{\infty} \int_{t=D}^{\infty} S(t-D) \cdot S(t)^j \, dS(t) \, dR(D) \tag{2}$$

where $S(t)$ is the distribution function of $u(i)$ and $R(D)$ is the distribution function of the destination wait time D.

B. Gap Probability for Playback Method I

Since there is no expansion of the time axis in method I, when a message delay exceeds the delay of the first message by more than D, a gap results. Hence, the average fraction of packets at which a gap occurs with method I in a sentence of length n messages is

$$\Pr[\text{gap}\,|\,n] = \frac{n-1}{n} \int_{D=0}^{\infty} \int_{t=D}^{\infty} S(t-D) \, dS(t) \, dR(D). \tag{3}$$

Comparing the right-hand sides of (2) and (3) we notice, as expected, that (3) is larger (identical at $n = 2$). Therefore, in general, the gap probability is lower in method E than in method I.

C. The Distribution of $D(m, k)$

The distribution $R(D\,|\,m, k)$ of the partial range given the distribution $S(t)$ from which the m samples are independently chosen is given by

$$R(D \mid m, k) = m(m-1)\binom{m-2}{k}\int_{t=0}^{\infty}\int_{z=0}^{D}[S(t+z) - S(t)]^{m-k-2}[1 - S(t+z)]^{k} \cdot dS(t+z)\, dS(t). \quad (4)$$

A derivation for the above following that of [4, pp. 98–99] appears in [8].

V. A MEASURE OF COMMUNICATION QUALITY

Before proceeding, we need a basis for comparing the "quality" of communication in our stream traffic context. Among the parameters affecting the quality of communication in the system under discussion are 1) G—the gap probability and 2) $u(i) + D$—the speaker-to-listener (end-to-end) delay. The end-to-end delay and the gap probability are both influenced by the destination wait time D. Ideally one would choose to operate at the point $D = 0$ and $G = 0$. Unfortunately, this cannot be achieved, in general, in a variable delay network. We assume that the $u(i)$ are uncontrollable parameters which determine the value of D chosen. There is a tradeoff between G and D. Experience has shown that we can tolerate nonzero G up to some threshold value [3] beyond which the quality degrades rapidly. For the experiment reported in [3], roughly 6 percent of the packets caused gaps and the quality was considered to be acceptable. It is likely that a similar behavior is true for D (i.e., beyond some value of $u(i) + D$ the communication ceases to be of an interactive nature). Studies regarding the effect of transmission delay on telephone conversations reported in [6] and [7] show that the threshold is above 0.6 s and below 1.8 s.

We choose normalized distance from the origin (i.e., the most desirable operating point) in the DG plane as an example of a measure of communication quality. The Euclidean distance function is normalized by dividing G and D by the threshold values g and d, respectively. We shall define the "region of acceptability" as those operating points which have a normalized distance from the origin of less than or equal to one.

VI. NUMERICAL RESULTS FOR THE EXPONENTIAL DISTRIBUTION

For the purpose of demonstration, we choose the exponential distribution with mean x for delay distribution

$$S(t) = 1 - e^{-t/x}.$$

Admittedly, this is a poor choice for the distribution of network delay [5], [11]! Measurements presented in [8] substantiate that. This distribution, however, has a large coefficient of variation when compared to the measured distributions. It therefore provides an approximate lower bound on performance, which should hold for less tractable but more realistic distributions.

We shall first find $R(D \mid m, k)$. We begin with (4). Substituting for $S(t)$ we have

$$R(D \mid m, k) = x(m-1)\binom{m-2}{k}\sum_{i=0}^{m-k-2}(-1)^{i} \cdot \binom{m-k-2}{i}\frac{1 - e^{-(i+k+1)D/x}}{i+k+1}. \quad (5)$$

A lengthy derivation in [9] yields the mean value of D given m and k.

$$D(m, k) = x\sum_{i=k+1}^{m-1}\frac{1}{i}. \quad (6)$$

Fig. 2 shows the behavior of $D(m, k)$ for m from 2 to 100 and several k. The $k = 0$ curve is logarithmic in shape, having its most rapid ascent at the beginning and then tapering off. This is a useful property in the selection of both m and k. The difference between $D(m, 0)$ and $D(m-1, 0)$ [i.e., $D = 1/(m-1)$] is the curve from which the $D(m, k)$ curves emanate. This shows that the relative difference between successive values of $D(m, k)$ is very small beyond, say, $m = 20$.

Let $GE(n, m, k)$ be the gap probability for method E, given that 1) each sentence contains n messages, 2) D is the partial range of m samples with the k largest samples discarded, and 3) the delay distribution is exponential. Then, (2) yields the following:

$$GE(n, m, 0) = \frac{1}{n}\sum_{j=2}^{n}\frac{1}{j-1} \cdot \left[1 + \frac{1}{j}\sum_{i=1}^{j}(-1)^{i}\binom{j}{i}\frac{1}{\binom{m-2+i}{i-1}}\right]. \quad (7)$$

Fig. 3 shows a family of curves for GE versus m, $k = 0$ and $n =$ 8, 16, 24, 32. For general $k = 0$ we obtain

$$GE(n, m, k) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{j-1} \cdot \left[1 - \left(\frac{1}{j}\right)\left(\frac{m-1}{k}\right)\left[1 - \frac{\binom{m-2}{k}}{\binom{m-2+j}{k}}\right]\right] \quad (8)$$

Fig. 4 shows GE versus m for $n = 16$ and several k.

We define $GI(n, m, k)$, the gap probability for method I, similarly to $GE(n, m, k)$. Beginning with (3) we obtain

$$GI(n, m, k) = \frac{(k+1)(n-1)}{2mn} = (k+1)GI(n, m, 0). \quad (9)$$

Fig. 5 illustrates a family of GI curves, for $k = 0$. Fig. 6 shows GI versus m for various k with $n = 16$.

A. Limiting Performance for Large m

$GE(n, m, 0)$ and $GI(n, m, 0)$ each have limiting values for zero as m approaches infinity. $D(m, 0)$ is unbounded, however. One must allow k to grow large in order to bound D. Suppose we let m and k approach infinity such that k/m approaches the value a $(0 > a > 1)$ Then,

$$\lim_{m \to \infty} GE(n, m, k) = \frac{1}{n}\sum_{j=2}^{n}\left[\frac{1}{j-1}\right] \cdot \left[1 - \left[\frac{1}{j}\right] \cdot \frac{1-(1-a)^{j}}{a}\right]$$

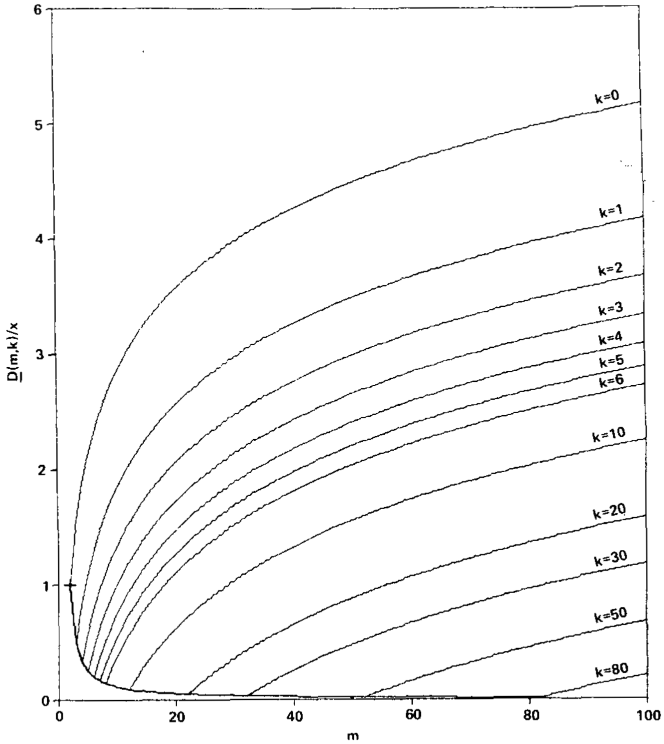


Fig. 2. Mean range versus sample size for exponential distribution.

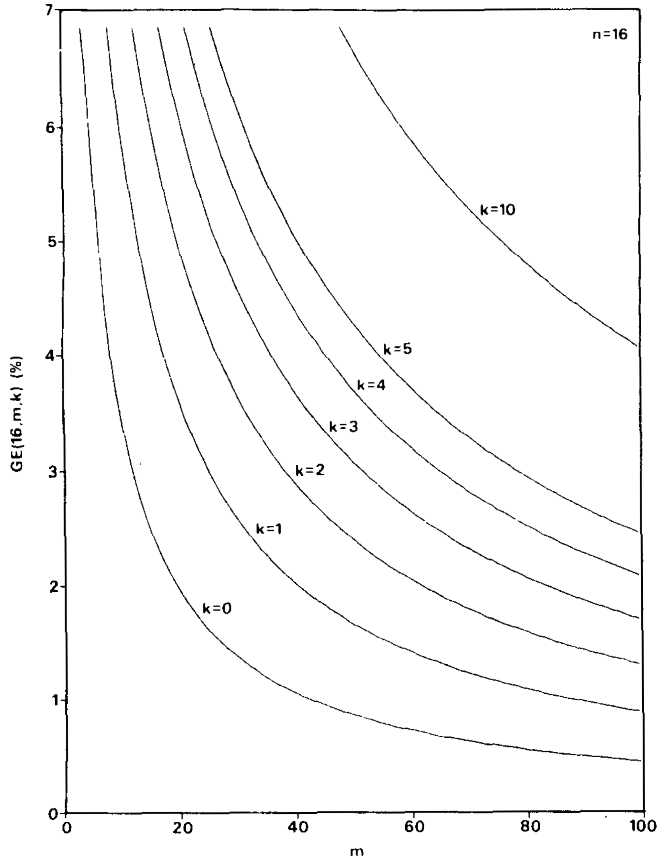


Fig. 4. Gap probability (Method E) versus sample size.

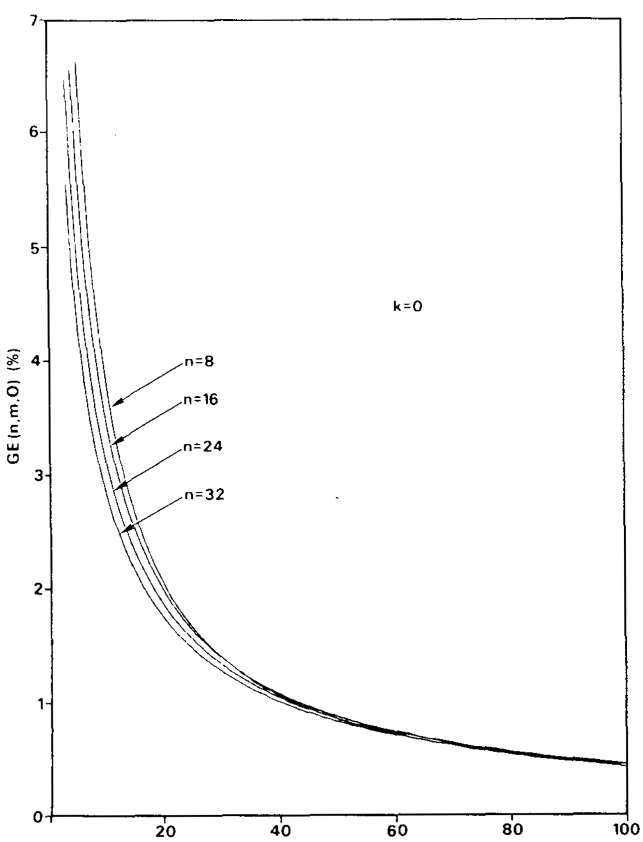


Fig. 3. Gap probability (Method E) versus sample size.

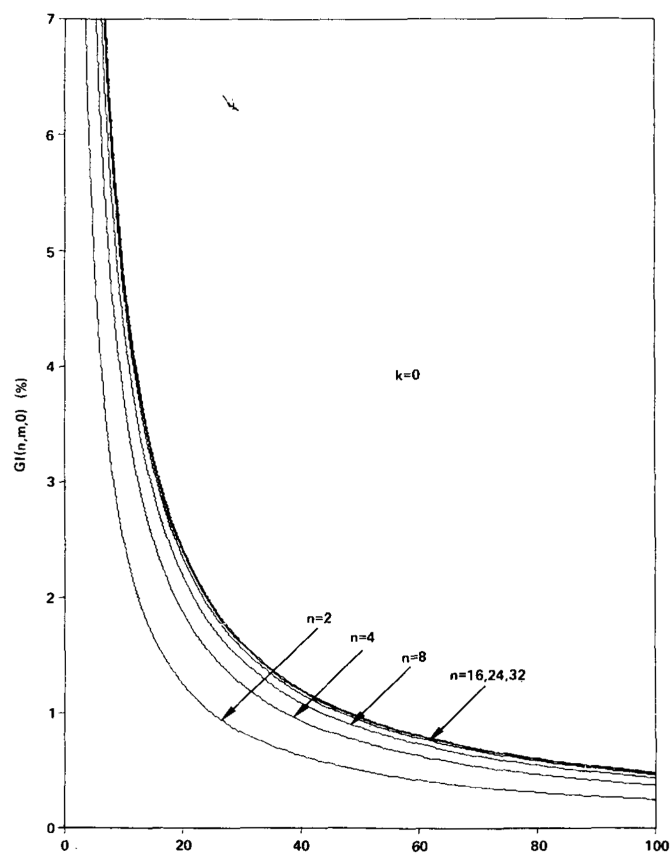


Fig. 5. Gap probability (Method I) versus sample size.

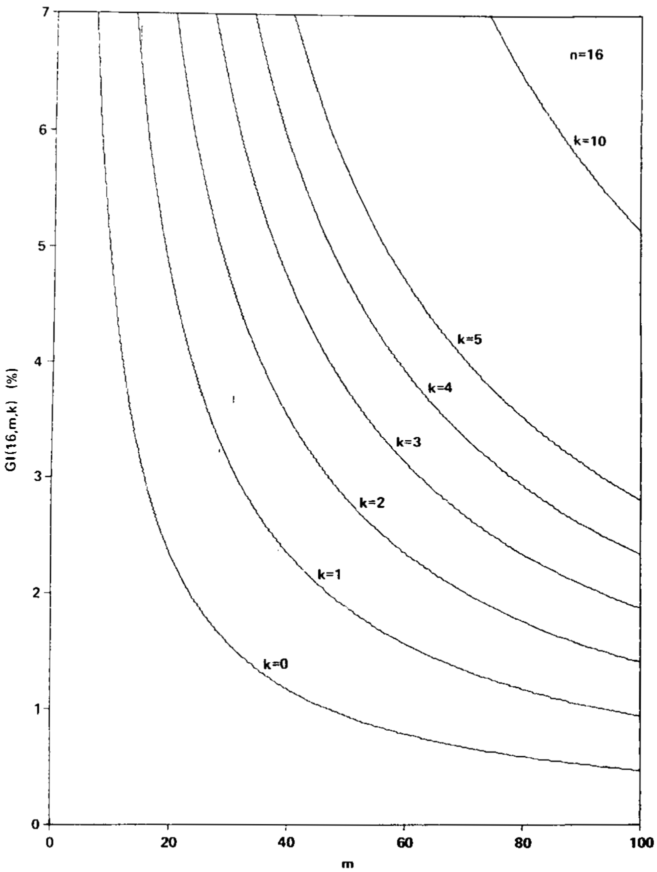


Fig. 6. Gap probability (Method I) versus sample size.

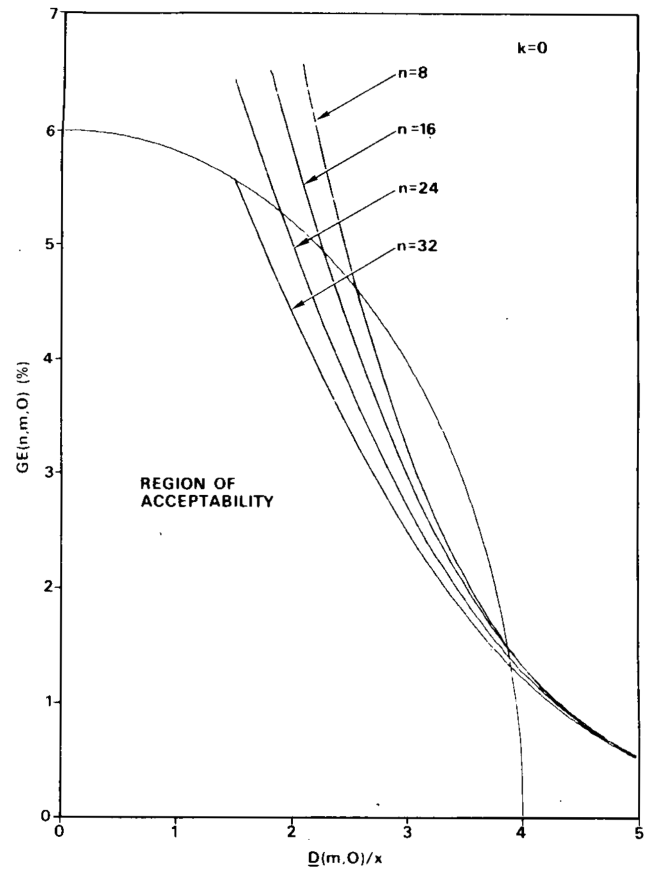


Fig. 7. Gap probability (Method E) versus mean range.

$$\lim_{m\to\infty} GI(n, m, k) = \lim_{m\to\infty} \frac{(k+1)(n-1)}{2mn} = \frac{a(n-1)}{2n}$$

$$\lim_{m\to\infty} D(m, k) = \lim_{m\to\infty} x \sum_{i=k+1}^{m-1} \frac{1}{i} = -x \log (a).$$

The optimum value of a (i.e., that value of a which causes D and G to most closely approach 0) for $n = 16$, $g = 0.06$, and $d = 4x$ is $a = 0.0655$ for method E, and $a = 0.0546$ for method I. This gives corresponding optimum points of $GE = 0.027$ at $D = 2.73x$ and $GI = 0.0256$ at $D = 2.91x$ (where x is the mean network delay). Since one would be required to remember all m values, the process which allows m and k to grow arbitrarily large is computationally infeasible. The knowledge of our proximity to the optimum is, however, quite important.

B. D/G Tradeoff

Let us now study the basic tradeoff between gap probability G and delay D. In Fig. 7 we plot GE versue D for $k = 0$ and various n. The choice of m is quite critical. For the case of $n = 16$ it can be shown that the acceptable range for m is approximately 6–29. If m is smaller, GE is too large; if m is larger, then D is too large. Either extreme performs at 23 percent worse than optimum. With $m = 12$ the performance is within 11 percent of optimum. This is the best possible performance for $k = 0$.

In order to examine the behavior of different values of k, we fix n (at 16) and plot GE versus D in Fig. 8. Notice that performance increases (i.e., normalized distance from the

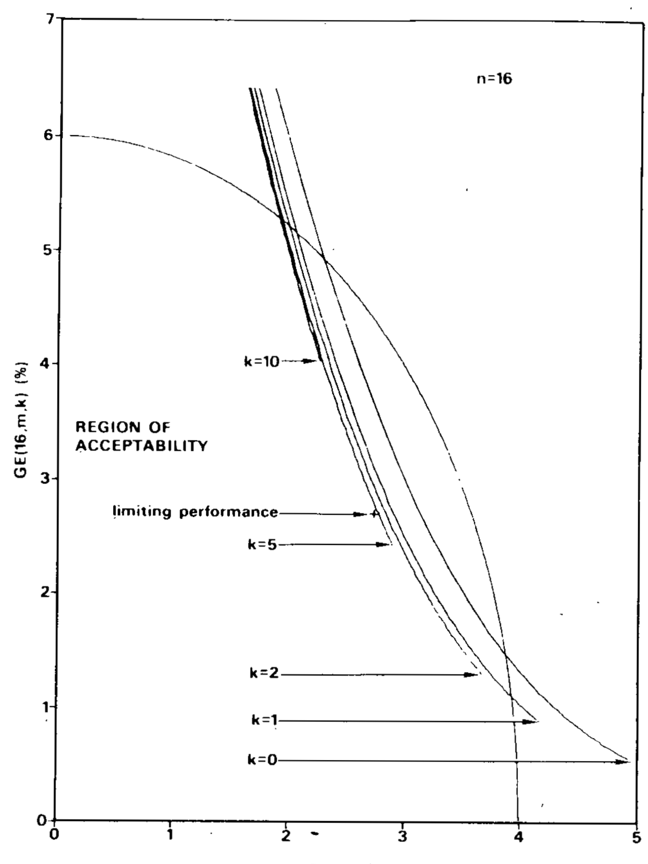


Fig. 8. Gap probability (Method E) versus mean range.

origin decreases) as k increases (with corresponding increase in m). That is, for those curves plotted, for any m, k pair there exists m' such that the performance at m', $k + 1$ exceeds the performance at m, k. Also, for each m there is an optimum value of k (which maximizes the quality function). The parameter a approximates the ratio of the optimal k for each m. Also shown is the limiting performance. The performance approaches the limit quickly at first, and therefore, one may achieve performance quite close to the optimum for small m (and therefore small k).

A rule of thumb for choosing m and k is as follows. Choose $m > 40$ and k approximately 7 percent of m for method E and 5 percent for method I. Using such a rule for the exponential distribution of delay, $n = 16$ and the given region of acceptability achieves performance within 5 percent of the limiting performance. In the next section we relax these assumptions and find that the rule continues to provide good performance.

VII. SIMULATION RESULTS

Some assumptions were made in earlier sections in order to provide a tractable model of the system. Even with these assumptions, analytic results were obtained only for a small class of distributions. We therefore have resorted to simulation in order to relax some of our assumptions. Presented in this section are the results of that simulation.

The simulation program is driven by a sequence of delays. Therefore, it is possible to remove the assumption of statistical independence among samples. This assumption is relaxed by using the actual delay strings from ARPANET measurements gathered at U.C.L.A. and from the simulation (not shown).

It is also possible to remove the assumption of fixed sentence size. Suggested by the early work of Norwine and Murphy [9], we have used a geometric distribution of sentence length (in terms of packets), with a mean which corresponds roughly to the 4.14 s reported in [9] as the average length of a talkspurt.

The simulations were performed as follows. First the delay string is "randomly" divided into sentences. At the beginning of each sentence a selection of D is made based on m and k. Knowing the delay of the first message of the sentence and D, we determine the number of gaps which would have occurred were this the sequence of delays experienced by an actual sentence. This is repeated for each sentence in the string and the results are recorded.

We show only the results using two of the measured delay strings. D is now expressed in ms instead of a multiple of the mean delay. In Figs. 9 and 10 we show the results of the range monitoring techniques using m = 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90 and 100, with k = 0, 1, 2, 3, 4, 5 and 10 ($k > m - 1$). The performance retains the hyperbolic character to a large extent. Notice that a wide range of m, k pairs give performance below $GE = 0.06$ (and $GI = 0.06$, not shown). The $m > 40$, $k = 0.07m$ rule provides good performance here as well, although this is only apparent in Fig. 10.

VIII. CONCLUSION

We have shown that one may devise distribution buffering schemes which effectively balance the frequency (and duration) of output gaps against end-to-end delay. We have produced a framework by which such schemes may be compared. Our approach of looking at each end of a continuum of playback schemes has produced bounds of performance within

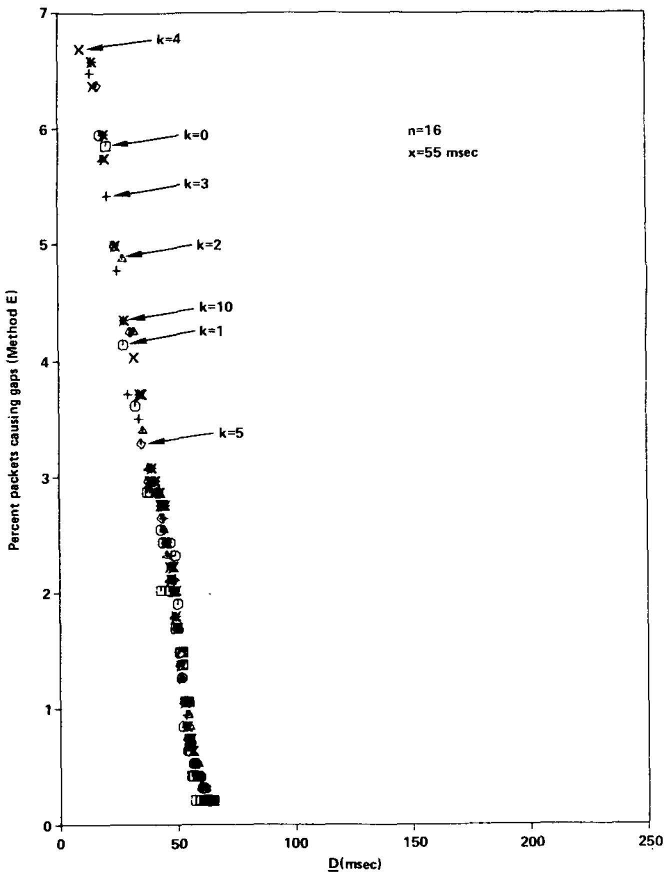


Fig. 9. Simulation results (Method E, measurement 1 hop).

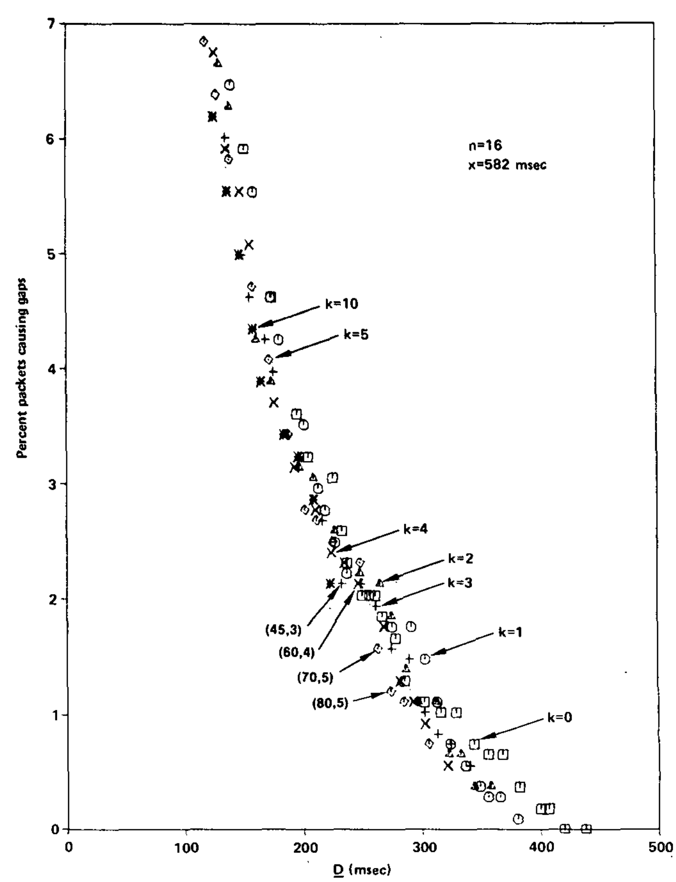


Fig. 10. Simulation results (Method E, measurement 10 hops).

which the performance of all schemes in the continuum must lie. We have gained an understanding of the tradeoff between gaps and delay. The above should provide some useful tools for designing new and better buffering and playback schemes.

The choice between playback methods E (expanded time) and I (late data ignored) is not a clear one. While it is true that method E preserves all information and has a lower gap probability than method I under like delay conditions, method E has the nasty problem that, in general, sentences take longer to output than to input. This can create problems in applications where timing balance between input and output is critical or when the network is subject to loss of packets. Early on we mentioned that the best choice of playback scheme would likely lie between the two extremes. The choice would "optimally" balance timing integrity and information integrity for a particular application. For example, modify method E to include a threshold for the maximum time to wait for any packet. This would allow for some control on the time expansion by eliminating the worst offenders, but would still provide fewer gaps than method I.

ACKNOWLEDGMENT

The authors are indebted to Dr. F. Kamoun who pointed out an error in an earlier version of (2). In addition, we are grateful for the helpful suggestions made by the reviewers.

REFERENCES

[1] G. Barberis and D. Pazzaglia, "Analysis and optimal design of a packet-voice receiver," *IEEE Trans. Commun.*, vol. COM-28, pp. 217–227, Feb. 1980.
[2] D. Cohen, "Specifications for the network voice protocol," Inform. Sci. Inst., Univ. Southern California, Marina del Rey, CA, Network Speech Compression Note 68, Mar. 1976.
[3] J. W. Forgie and C. W. McElwain, "Some comments on NSC note 78 'Effects of lost packets on speech intelligibility,'" Massachusetts Inst. Technol., Lincoln Lab., Network Speech Compression Note 92, Mar. 1976.
[4] E. J. Gumbel, *Statistics of Extremes*, 4th printing. New York: Columbia Univ. Press, 1967.
[5] L. Kleinrock, *Communication Nets: Stochastic Message Flow and Delay*. New York: McGraw-Hill, 1964, reprinted by New York: Dover, 1972.
[6] E. T. Klemmer, "Subjective evaluation of transmission delay in telephone conversations," *Bell Syst. Tech. J.*, vol. 46, pp. 1141–1147, 1967.
[7] R. M. Krauss and P. D. Bricker, "Effects of transmission delay and access delay on the efficiency of verbal communication," *J. Acoust. Soc. Amer.*, vol. 41, no. 2, pp. 286–292, 1967.
[8] W. E. Naylor, "Stream traffic communication in packet switched networks," Ph.D. dissertation, Dep. Comput. Sci., School Eng. Appl. Sci., Univ. California, Los Angeles, Rep. UCLA-ENG-7760, Aug. 1977.
[9] A. C. Norwin and O. J. Murphy, "Characteristic time intervals in telephonic conversation," *Bell Syst. Tech. J.*, vol. 17, pp. 281–291, 1938.
[10] H. Opderbeck and L. Kleinrock, "The influence of control procedures on the performance of packet switched networks," in *Conf. Rec., Nat. Telecommun. Conf.*, San Diego, CA, Dec. 1974, pp. 810–817.
[11] J. Wong, "Distribution of end-to-end delay in message-switched networks," *Comput. Networks*, vol. 2, pp. 44–49, 1978.

Combined Effect of the Carrier Recovery and Symbol Timing Recovery Error on the P_e Performance of QPR and Offset QPR Systems

SHINJIRO OSHITA, MEMBER, IEEE, AND KAMILO FEHER, SENIOR MEMBER, IEEE

Abstract—**In this paper we investigate the *combined* effect of sampling offset and phase error on the probability of error P_e performance of quadrature partial response (QPR) and offset QPR systems. The Gram–Charlier series expansion method is used to get accurate probability of error expressions. As an example, class 1 and class 4 partial response signaling schemes are examined.**

I. INTRODUCTION

Quadrature partial response (QPR) radio systems have been manufactured for a number of years. One of the advantages of these systems is their relative hardware simplicity and tolerance to filtering imperfections [1]-[4], [13]-[15]. In this paper we investigate the *combined* effect of sampling offset and phase error on the probability of error P_e performance of QPR and offset QPR systems.

II. QPR AND OFFSET QPR SYSTEM MODELS

A model of the transmitter and receiver of a QPR system is shown in Fig. 1. From Fig. 1(a) we note that in the transmitter the input NRZ data stream at rate T^{-1} is converted by a serial-to-parallel converter into two random streams labeled $\{a_n'\}$ and $\{b_n'\}$ having a baud rate half that of the original bit rate. Precoding is used to prevent error propagation in the demodulation process of the encoded bits. Streams $\{a_n'\}$ and $\{b_n'\}$ are precoded to produce the new binary bit sequences $\{a_n\}$ and $\{b_n\}$, respectively, as the transmitted signals. For the case of an offset QPR system, the $\{b_n\}$ stream must be delayed by a unit bit duration, in comparison with the other data stream [5]-[6].

For our analysis we assume that all pulse shaping is done by the transmit filter $G(\omega)$, the impulse response of which is $g(t)$. Modulating the baseband signal sequence

$$\sum_{n=-\infty}^{\infty} a_n g(t-2nT) \quad \text{and} \quad \sum_{n=-\infty}^{\infty} b_n g(t-2nT)$$

by the carriers $\cos \omega_c t$ and $\sin \omega_c t$ where ω_c is the center frequency of the carrier, we obtain the following sequences of transmitted signals:

$$\sum_{n=-\infty}^{\infty} a_n g(t-2nT) \cos \omega_c t \tag{1a}$$

Paper approved by the Editor for Radio Communication of the IEEE Communications Society for publication without oral presentation. Manuscript received July 8, 1981; revised February 13, 1982. This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

S. Oshita is with Shinshu University, Matsumoto, Japan.
K. Feher is with the University of Ottawa, Ottawa, Ont. K1N 6N5, Canada, and SPAR Aerospace Limited, Ste-Anne-de-Bellevue, P.Q., Canada.

0090-6778/82/1200-2534$00.75 © 1982 IEEE