2

EFFECT OF CHANNEL CONTROL ACCESS SCHEMES AND ERROR CONTROL PROTOCOLS ON SPECTRUM EFFICIENCY OF ADVANCED DIGITAL TRANSMISSION TECHNIQUES OVER LAND MOBILE CHANNELS

by

1
Dr. Robert W. Donaldson, Professor
Department of Electrical Engineering
Faculty of Applied Science
The University of British Columbia

for

Department of Communications
Ottawa, Ontario, Canada

under

Department of Supply and Services Contract Serial No. OST83-00077
Period: April 1, 1983 - March 31, 1984

Rec'd from DEMS J. Boeschin

ABSTRACT

The report summarizes work primarily involving estimates of spectrum efficiency θ and delay D when data is transmitted over conventional line-switched land mobile radio channels. Spectrum efficiency here refers to the fraction of the time that any channel is actually used for transmission of information bits. Delay is measured from the time of message generation to completion of transmission.

The report examines types of messages to be transmitted, including short interactive data messages, real-time speech and file-data. Suitable channel access protocols are briefly described.

A general access control format is proposed, consisting of request (REQ), answer (ANS), message (MSG) and acknowledgement (ACK). An ACK may follow each message segment, or may come only at the end of the entire message. Typical radio environments are discussed in relation to the general access control procedure. Problems considered include multiple independent bases sharing one channel, and access control difficulties resulting from mobiles not being within line of sight of each other. A proposed data transmission format is described.

Considerable effort was directed to obtaining analytical estimates of D vs. θ, initially when the entire REQ/ANS/MSG/ACK sequence is carried on the channel selected for use. For CSMA access control in environments with low propagation delay relative to REQ packet length, perfect scheduling (queueing) results provide reasonably accurate estimates of performance. Such performance was calculated, and for the data transmission format chosen was seen to yield relatively low θ values; typically 0.11 with delays of two message lengths for three-line messages when mobiles operate on one input/output half-duplex channel pair using CSMA access control.

Various options involving the REQ/ANS sequence were examined. One option is to assign some channels solely for transmission of REQ and ANS packets. This approach provides for reduced contention delay by using lightly loaded REQ channels, and also enables mobiles with multichannel operational capability to monitor one channel only for REQ or ANS packets. Perfect scheduling (properly calculated) again provides an estimate of D vs. θ, and estimates were determined for CSMA and pure ALOHA access controls. With operational capability on $m = 4$ message channels, $\theta = 0.28$ and 0.40, respectively, for pure ALOHA and CSMA, with two-message-length delays for three-line messages. It was found that operational capability on more than one channel greatly improves spectrum efficiency, but not much improvement is gained for $m \gtrsim 8$. The best mix of access and message channels was calculated and found to depend on the access control protocol and message length.

The report concludes with a summary of major results, and some suggestions for additional useful work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

I. INTRODUCTION

I-1 Scope and Objectives of the Study

This report presents the results of a Communications Canada Contract study whose motivation and purpose are summarized below, together with an overview of the results obtained.

Motivation

Recent advances in land mobile system development has provided increasing evidence that digital transmission of dispatch data as well as other digitally encoded information will constitute an important part of land mobile communication traffic. The need for the development of guidelines identifying transmission characteristics, requirements, and limitations of digital transmission is becoming increasingly urgent. These guidelines would lead to the development of new regulations and standards governing the technical characteristics and operation of equipment and systems employing digital transmission of information over land mobile communication channels.

The following items are of specific interest:

1. Effect of channel control access schemes on digital land mobile communication systems.
2. Effect of error control protocol on digital land mobile communication systems.
3. For a given digital modulation technique derive the optimum channel control access scheme and error control protocol for the most efficient spectrum utilization.
4. Comparison of spread spectrum techniques with conventional analog FM, over a land mobile radio channel.

Purpose

To produce technical data in support of the development of guidelines for the use and operation of land mobile radio equipment and systems in the 900 MHz band for communication over land mobile channels utilizing digital modulation techniques.

At the outset it was recognized that an important part of the study would involve selection of an approach which would provide useful technical data consistent with the study's motivation and purpose.

I-2 Executive Summary

Chapter II begins with a brief summary of types of messages transmitted on land mobile radio channels. Message types include short data messages typical of enquiry/response environments, long non-real-time file transmissions, and real-time voice messages. Spectrum efficiency (throughput) θ is then defined as the fraction ot the time that channels actually carry information bits. Known results for delay and blocking probability for queueing systems are summarized for subsequent use. Channel access control protocols suitable for use on land mobile radio channels are then described.

Chapter III begins with the presentation of a general channel access control procedure which involves transmission of a request (REQ), answer-to-request (ANS), message (MSG) and acknowledgement (ACK). A single ACK may follow the entire message, or ACK's may occur after each message segment. Contention access delays are examined and existing curves are closely fitted by simple functions, for later use. Typical radio environments and their operational features are discussed in relation to the general access control procedure. Included in the discussion are the effects of multiple bases sharing a channel, as well as the difficulties and possible solutions which ensue because mobiles are not normally within line of sight of each other. The line of sight problem can create serious contention access delays and spectrum efficiency degredations unless a busy tone or paired input/output channels are used. A proposed CCIR mobile data transmission format is described. The effects of retransmissions of all or parts of a message are determined.

Chapter IV involves a consideration of delay vs. throughput performance when the entire REQ/ANS/MSG/ACK sequence is carried on one channel. The problem involves analytical difficulties, since the channel is used first for contenion by REQ packets and then for transmission of the ANS/MSG/ACK sequence, during which time the channel is unavailable for REQ packet transmissions. This unavailable probability was determined, assumed to be totally random, and used to estimate contenion access delay for CSMA and ALOHA REQ access protocols. This latter result was then used to calculate delay vs. throughput performance. However, upon comparison with perfect scheduling access delay, the results were found to be too optimistic, and the independent busy period assumption (which has been used by others in different contexts) had to be set aside. An alternative approach was then developed based on mean busy period duration during the REQ contention period. It was argued that perfect scheduling of the REQ/ANS/MSG/ACK sequence provides a lower bound to delay for any access protocol, and provides a reasonably accurate estimate of delay vs. throughput for CSMA under conditions of small relative propagation delay. Delay vs. throughput results for perfect scheduling were actually calculated for data messages of various mean length under the proposed CCIR format, and were found to be rather low. For example, transmission of three lines of text using mobiles which can operate on $m = 1$ channel only, with a delay of twice the (exponentially distributed) message length implies a spectrum efficiency $\theta = 0.11$, assuming half-duplex input/output channel pairs. Transmission capability on any of four chanels increases θ to 0.245. Delay here refers to the time from message generation to completed transmission.

Chapter V examines issues involving the REQ/ANS sequence. It is shown that for efficient access protocols such as CSMA omission of this sequence can improve spectrum efficiency for short messages. However the REQ/ANS sequence is useful for longer messages or for inefficient access protocols such as pure ALOHA, or when retransmissions due to channel errors are likely. The alternative of using dedicated access channels for transmission of the REQ/ANS sequence is analysed. For short messages or inefficient access protocols this option is beneficial, since REQ contentions occur on dedicated and relatively lightly loaded channels. An important parameter of interest here is the ratio α of the number of access channels to the number of message channels, which ratio beneficially increase as the message length or access protocol efficiecy decreases. Perfect scheduling (properly calculated) again provides a lower bound on delay for any θ value, and is reasonably accurate for small REQ to MSG length ratios. Delay vs. throughput results obtained for perfect scheduling on message channels and maximum loads on access channels show that even on dedicated access channels, channel sensing provides considerably better performance than does pure ALOHA, especially for short messages. One advantage of using dedicated access channels is that mobiles with multichannel capability are required to monitor one channel only for REQ or ANS packets. A typical result: for m = 4 message channels and transimssion of three lines of text with pure ALOHA and np CSMA (a = 0.01) access, θ = 0.40 and 0.28, respectively for delays of less than two message lengths. Here we assume paired input/output access channels and two-way half-duplex message channels. The corresponding value for all transmissions on paired input/output channels with m = 1 channel per mobile is θ = 0.11.

Chapter VI presents a summary of the major results. A very effective way to enhance spectrum efficiency is to provide for multichannel operation on message channels, so that some channels do not remain idle while others are heavily congested. As well, an efficient access protocol is useful to avoid excessive contention delays which inevitably preceed message transmission.

Chapter VI also includes suggestions for further work. Proposed work includes simulations to verify delay-vs-throughput behaviour for various access controls, and comparative evaluations of other system configurations including those utilizing packet switching or spread spectrum signalling.

II. ACCESS CONTROL CONSIDERATIONS

In this chapter some known facts and results relating to mobile radio transmission and access control are summarized, for subsequent use.

II-1 Message Types

Messages on land mobile radio channels, as well as other channels which share voice and data may be categorized, as follows [G1, G2]:

1. Speech is generated in real-time and has much natural redundancy. Speech normally requires transmission as it is generated. In conversational environments, a voice response is expected with a few seconds. Awkward silences are embarrassing!

   The actual information rate for speech is probably less than 100 bits/sec.; conventional 8-bit PCM implies 56 Kb/sec. transmission rates [P1, A1]. Adaptive differential PCM requires at least 10 Kb/sec. Linear predictive coders require on the order of 2 Kb/sec. Because of its natural redundancy, channel bit error rates as low as 10^{-2} are tolerable on random error channels [C1].

   On mobile channels speech is often transmitted via analog frequency modulation, although digital transmission using low bit rate encoding/decoding is of increasing interest [J1].

2. Text or File Data may be generated slowly ($\lesssim$ 25 bits/sec.) by humans at keyboards. Text may also consist of stored messages or data, which is part of a data base. Video data (facsimile) such as line drawings consistute another form of text or file data.

   Text does not normally involve real-time transmission, and may not require immediate channel access. However, when text constitutes a response to an enquiry, it is often desirable for the response to be received within two or three seconds of transmitting the enquiry [M1].

Most text, including English text [S1] and video data [H1, V1] includes natural redundancy some of which can be removed prior to transmission by source coding. English text, for example is approximately 50% redundant [S1].

3. Conversational Data consists of short data blocks, typically one 80-character line (≈500 bits) or less in length. Conversational data requires rapid delivery to its destination (typically ≈1 sec.), as well as a rapid response if the data block represents an inquiry or a short control message.

In summary, speech has low accuracy requirements but involves real-time transmission. Long file transfers involving English text or video data have moderate accuracy requirements, but may require neither immediate channel access nor real-time transmission. Conversational data requires immediate channel access and may have high accuracy requirements.

The mixing of various message types with different characteristics and requirements on one channel involves difficulties and compromises. The problem is receiving continued and active attention [A2, F1, G1, G2, H2, R1, W1]. Reference [G2] includes a good bibliography relative to this problem.

II-2 Spectrum Efficiency

One important measure of land mobile radio system performance is spectrum efficiency, which equals the number of information bits/sec/Hz. carried by a system. Previous work [D1] has shown that for conventional narrowband cellular systems with N cells, G groups of channels, channel spacing Δ Hz. and bit rate R, the spectrum efficiency Q is as follows:

$$Q = (R/\Delta)(N/G)\theta \qquad (2\text{-}1)$$

In (2-1) θ denotes the fraction of the time that any channel is used for transmission of actual information bits. Ideally, $\theta \simeq 1$. In fact, $\theta < 1$, as a result of the transmission of check bits, address bits, retransmissions, requests for access, other overhead bits and channel idle time. The purpose of the work described in this report is to estimate θ for various channel access control schemes.

The term R/Δ depends on the allowable level of adjacent-channel interference, which in turn affects the bit-error-rate p. As R/Δ increases so does p. The term N/G depends on the allowable level of co-channel interference. The way in which R/Δ and N/G affect performance has been considered previously [D1].

II-3 Delay and Blocking Probability

Messages awaiting channel access are often regarded as entities in a queue [K1, K2, J1].

Under the Erlang C message-handling discipline, a message is served by one of m free channels, on a first-come-first-served (FCFS) basis. Once in the queue, the message waits until served.

Under the Erlang B discipline, a message does not wait for service; if all m channels are busy, the message is blocked from transmission.

The Erlang C discipline has been proposed as an appropriate model for data transmission [K2, J1]. The Erlang B case has been proposed for use in voice transmission [J1]. Recent work has involved studies on models which incorporate features of both disciplines [B1, N1].

In this report, we use the Erlang C discipline for voice as well as for data. Our reason is that a voice terminal with a message to transmit would be willing to wait some time for a connection to be completed, and this time really constitutes a queueing delay. Even if the waiting time were excessive, the caller could leave the queue and later return. The results below for queueing delays are applicable to any service discipline which does not depend on the length of waiting messages [K2]. This use of Erlang C throughout permits meaningful comparisons of results involving voice traffic, data traffic, and voice and data traffic on one or more channels.

To determine the queuing delay under Erlang C, we consider a set of m channels, each of which transmits at C bits/sec. Assume that the messges are exponentially distributed, with mean length μ^{-1} bits. The messages arrive with a Poisson arrival rate distribution, with mean rate $m\lambda$ messages/sec. (i.e.; λ messages per channel). We define the per channel utilization ρ as

$$\rho = \lambda / \mu C \qquad (2\text{-}2)$$

Define P_m as the probability that all m channels are busy, w as the average time that a message waits for service, and T as the waiting plus transmission time. It has been shown that [K1, K2]:

$$P_m = p_o (m\rho)^m / [m!\ (1-\rho)] \tag{2-3}$$

$$p_o = \left[\frac{(m\rho)^m}{(1-\rho)m!} + \sum_{k=0}^{m-1} \frac{(m\rho)^k}{k!} \right]^{-1} \tag{2-4}$$

$$w = P_m / m\mu C(1-\rho) \tag{2-5}$$

$$T = (1/\mu C) + w \quad . \tag{2-6}$$

Fig. 2-1 shows P_m vs. ρ and m. Fig. 2-2 shows the normalized delay d = μCT vs. ρ and m. In both cases, one sees that as m increases, ρ can move ever closer to unity without increasing d and P_m.

When message lengths are not exponentially distributed, it is not possible, in general to calculate P_m, w and T. For m = 1 channel, however, $P_m = \rho$, and

$$w = \left(\frac{1}{\mu C} \right) \left(\frac{\rho}{1-\rho} \right) \phi \tag{2-7}$$

where 2ϕ is the ratio of the second moment $\overline{L^2}$ of the message length to the square of the first moment, $\bar{L}^2$;

$$\phi = \overline{L^2}/2\bar{L}^2 \quad . \tag{2-8}$$

These results are used repeatedly throughout the report.

II-4 Channel Access Protocols

In land mobile radio environments, a large number of uncoordinated mobiles and one or more base stations share one or more channels. Specific arrangements for control of channel access for transmission of messages become necessary.

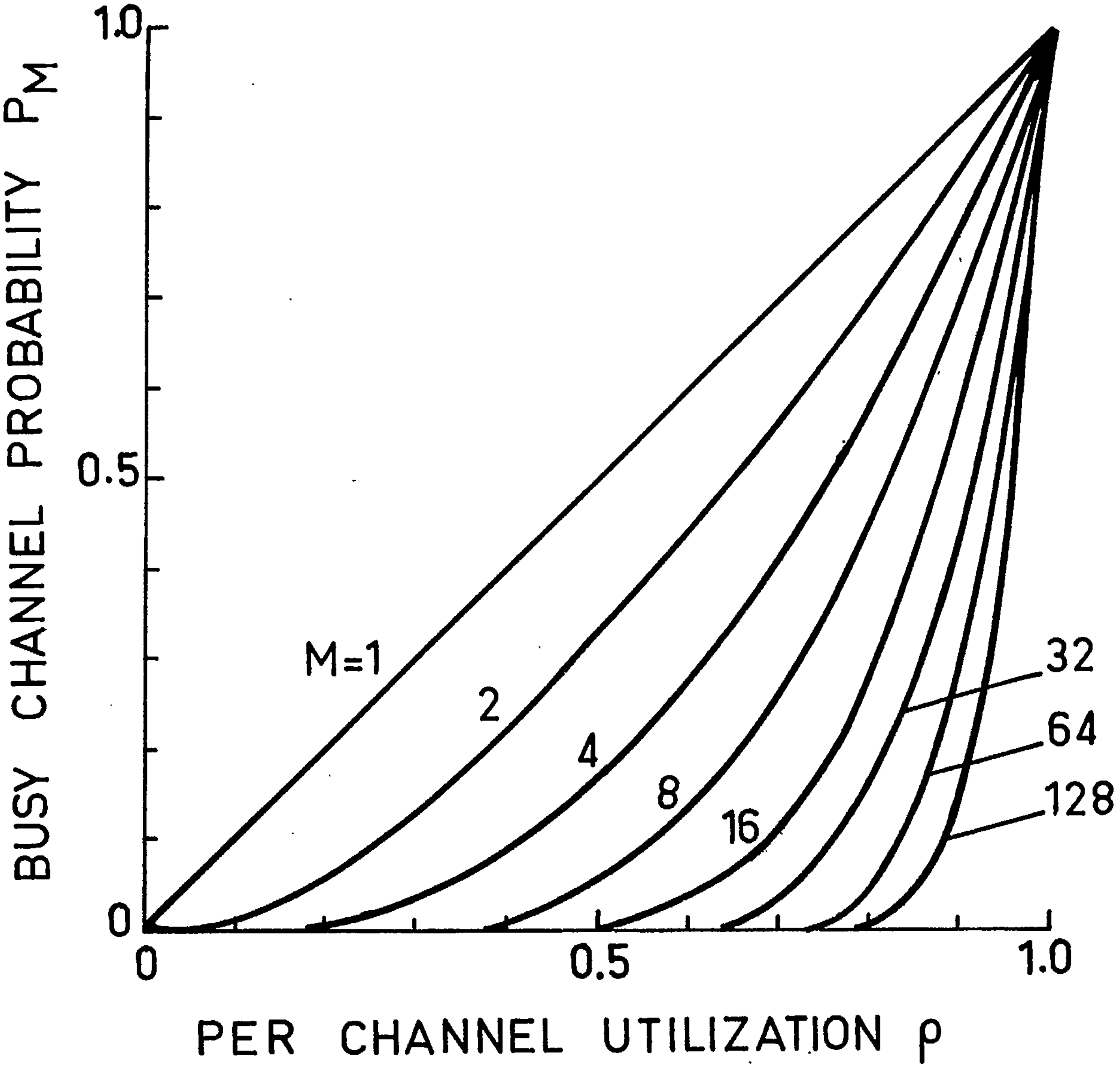


Fig. 2-1 Busy channel probability P_m vs. utilization p and number of channels m (Erlang C)

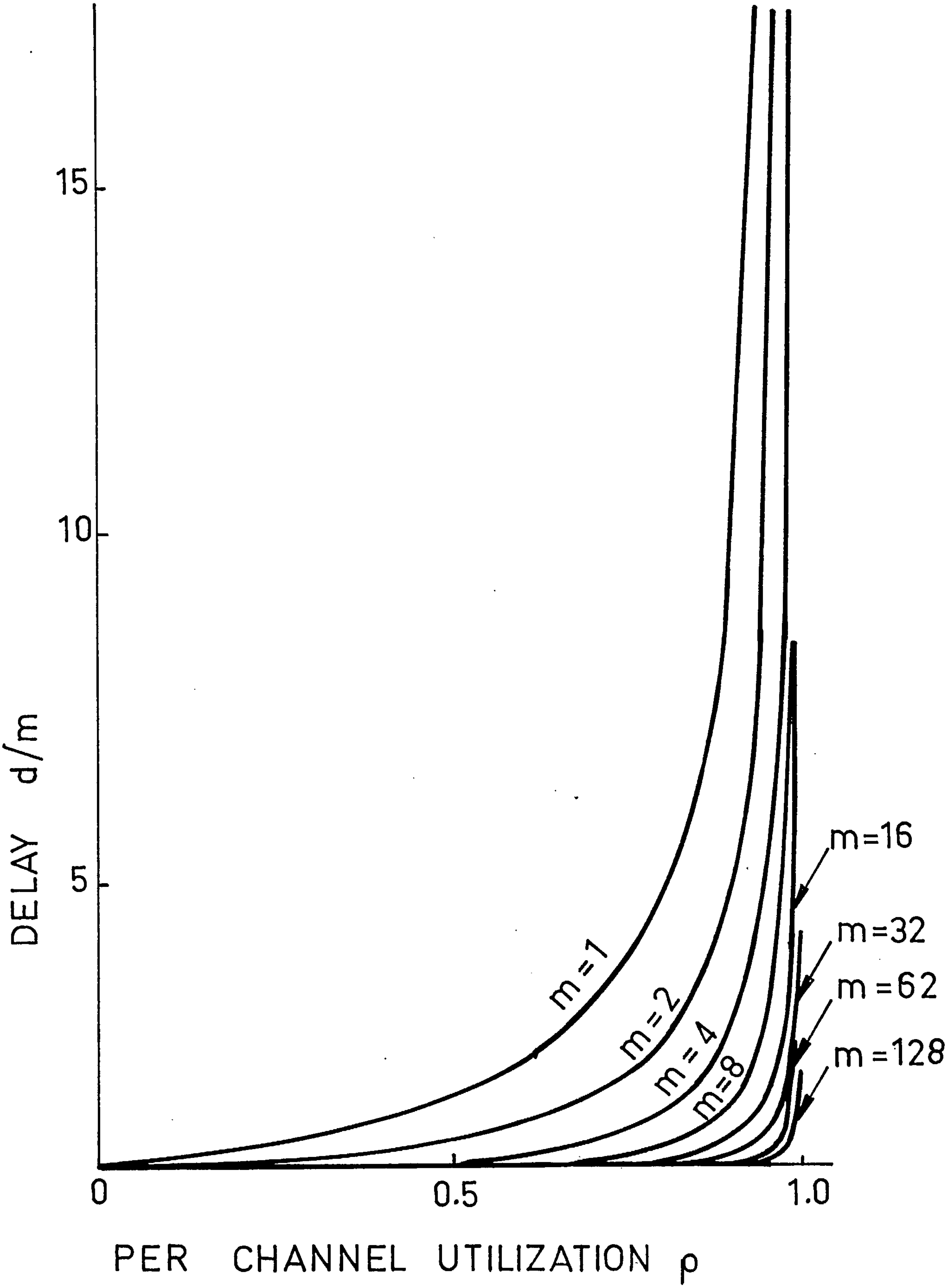


Fig. 2-2 Normalized delay d vs. ρ and m (Erlang C)

One possibility is to require that each mobile not transmit unless polled by the base, which would initiate all transmissions. The difficulty with such a scheme is that excessive time is lost in transmitting control messages, particularly when only a small fraction of mobiles have ready messages at any given time.

An alternative is to use carrier sense multiple access (CSMA) protocols [T1, K2, C2]. Assume for the moment that all mobiles hear each other and the base at all times. Let the maximum round-trip propagation delay between any two mobiles be τ sec., and the length of the data packet to be transmitted equal P sec. Let $a = \tau/2P$ be the normalized one-way delay.

Non-persistent CSMA operates as follows [T1, K2]:

1. A ready terminal senses the channel.
2. If the channel is sensed idle the terminal transmits the packet.
3. If the channel is sensed busy, the terminal reschedules the transmission to some later time, in accordance with a delay distribution. At this later time, it returns to step 2.

The above protocol may be "unslotted", in which case transmission can occur at any time, or "slotted", in which case transmission occurs only at the beginning of a system-wide time-slot of $\tau/2$ sec. duration. Slotting reduces the probability of collision between two packets.

The p-persistent CSMA protocol operates in a slightly different way. The slotted version is as follows:

1. A ready terminal senses the channel.
2. If the channel is sensed idle, then with probability p the terminal transmits; with probability 1-p it waits $\tau/2$ sec., and repeats the above procedure.
3. If the channel is sensed busy, the terminal waits until the channel becomes idle, then repeats the procedure in step 2.

The $\tau/2$ sec. wait is to avoid collisions with other packets seeking access at the same time.

What happens when packet collisions occur? The transmitting terminals fail to receive an acknowledgement to their transmitted packet, and once again seek access. To avoid another collision with the same and possibly other new packets, the retransmission is randomized over an average of $\bar{X}$ (normalized with respect to P) time units. The resulting delay, with $\bar{X}$ optimized has been determined. Fig. 2-3 shows this normalized delay, assuming packets of fixed length μ^{-1} bits, Poisson generation rate λ packets/sec., and channel bit rate C bits/sec. (thus $P = (\mu C)^{-1}$). Utilization ρ is as defined in (2-2).

Clearly, optimum p-persistent CSMA, where p is optimized for each value of ρ is best, although slotted non-persistent CSMA is equally good for $\rho \lesssim 0.45$.

The simplest protocol is pure ALOHA [K2], where any terminal with a packet to transmit does so without first sensing the channel. Collisions are frequent, and the maximum value for ρ is 0.18. A slotted version with P sec. slot duration doubles this maximum value to $\rho = 0.37$.

The ALOHA results in Fig. 2-3 apply for any value of a, whereas the CSMA results apply for a = 0.01. All the results in Fig. 2-3 ignore the effect of acknolwedgement traffic; i.e. acknowledgements (ACK's) are assumed to arrive instantaneously at no bandwidth cost. As well, all errors are assumed to result solely from packet collisions. Much of the remainder of the report deals with the effect of (ACK's) and error control overheads.

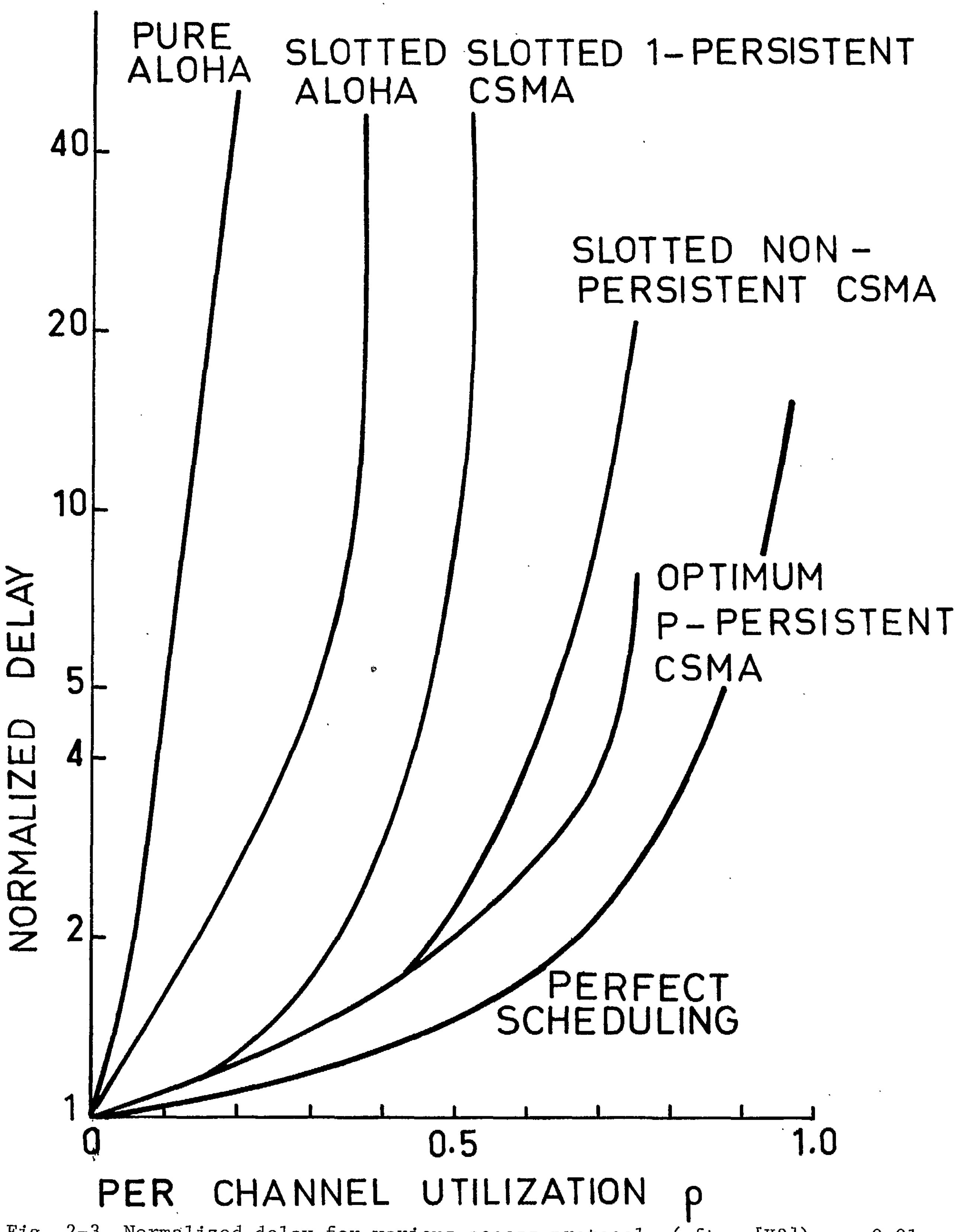


Fig. 2-3 Normalized delay for various access protocols (after [K2]) a = 0.01

In most situations all terminals would not be able to hear all other terminals, although in many cases all terminals would hear the base. In this case, carrier sensing would not be very helpful in reducing collisions. One solution is for a busy tone to be transmitted, from a point within line of sight of the base and each mobile, whenever the channel is occupied. Such a scheme has been called busy-tone multiple access (BTMA) [T1, K2]. The tone reduces the available channel capacity and increases the delay, since some time is required for a terminal to decide whether or not the tone is present. Optimization of BTMA has shown a reduction in ρ by approximately 15% from its slotted non-persistent CSMA value. This matter is discussed in more detail in Section II-3.

III. ANALYSIS OF ACCESS CONTROL PROTOCOLS

III-1. General Access Control Procedures

In this report we consider the general access control procedure diagrammed in Fig. 3-1. The steps are summarized as follows:

1. Following a waiting period, a request (REQ) is sent from a source with a ready message to the intended receiver or sink.
2. Upon receipt of the REQ, an answer (ANS) is returned to the source.
3. Upon receipt of the ANS, the message (MSG) is sent from source.
4. Upon receipt of the MSG an acknowledgement (ACK) is returned by the sink to the source.

The above access control sequence is quite general, and provides for various alternatives. For example, a single channel may carry the REQ, ANS, MSG and ACK. Alternatively, special access control channels may be used for REQ and/or ANS packets. In those cases where messages are short, the REQ packet may include MSG data, and the ANS then becomes the ACK. This case may be viewed as compatible with Fig. 3-1 in which the length of both the MSG and ACK sequence is zero.

Fig. 3-1 implies that ANS, MSG and ACK transmissions are always received once they have been sent. If an ANS is not received within a time-out period, then the REQ is retransmitted, perhaps following a waiting period. If an ACK is not received within a time-out period or if a NACK (negative-ACK) is received then the MSG is retransmitted (see Fig. 3-2).

The above access control protocol is applicable to circuit-switched communication environments, including conventional land mobile radio systems. Once the ANS is sent, the channel becomes available solely to the requesting

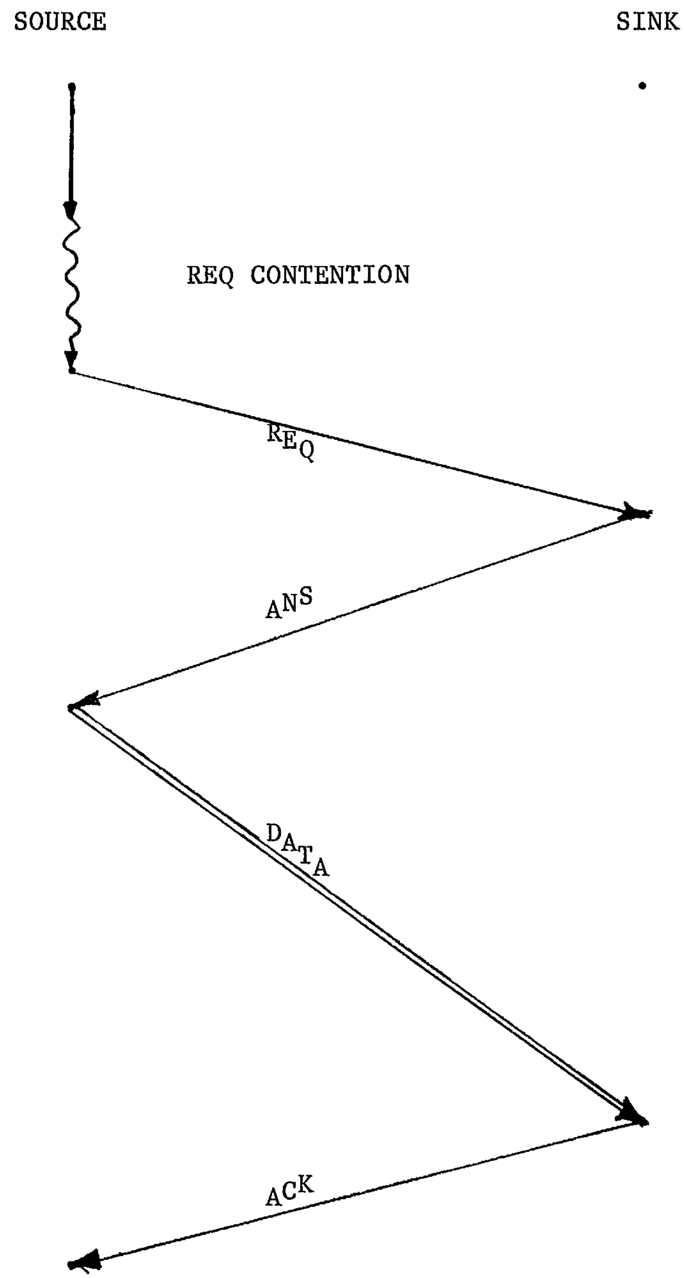


Fig. 3-1 Access Timing Diagram.

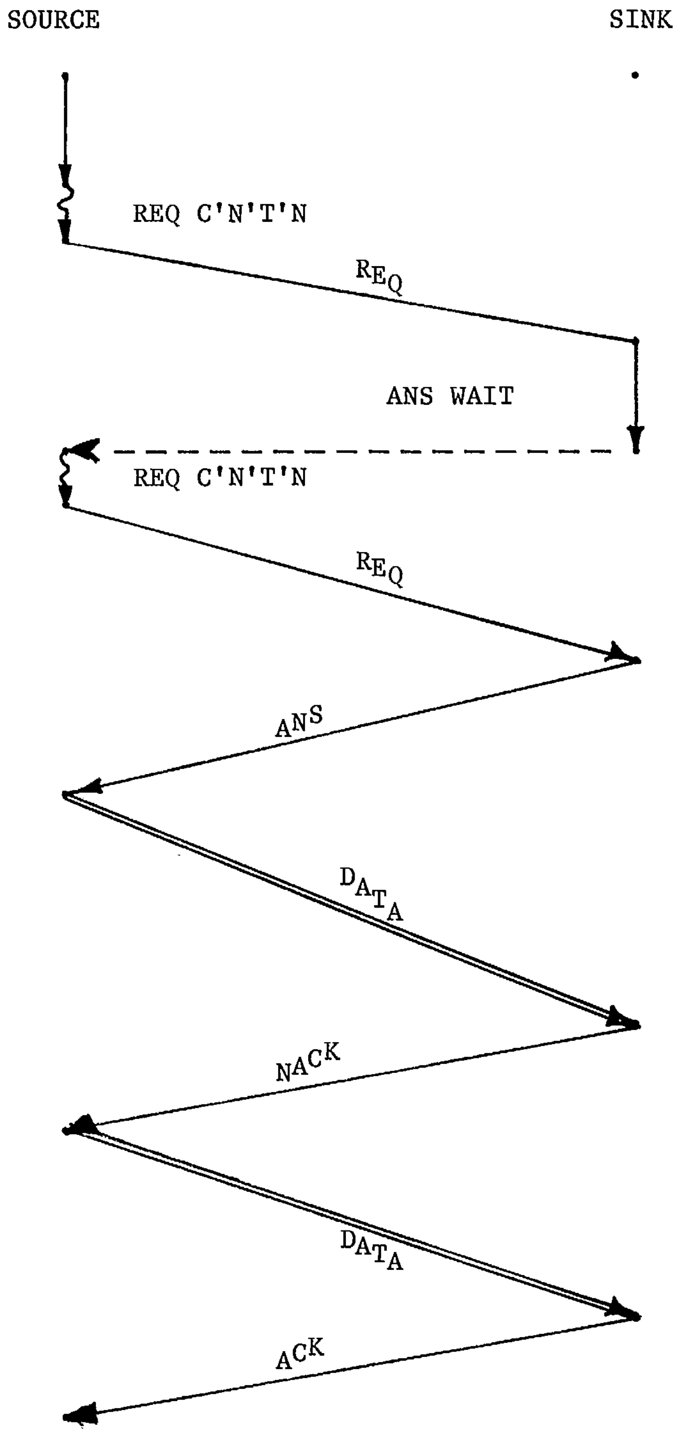


Fig. 3-2 Access timing diagram with retransmissions.

and answering parties until completion of the MSG transmission or expiration of a previously agreed-upon connect time.

Although the MSG is shown as being transmitted in one direction, bi-directional flow of MSG information is possible. As well, the MSG may be broken into pieces, with an ACK following successful transmission of each MSG piece. Another alternative is to follow the last ACK by an end-of-message (end-of-file (EOF)) transmission. We assume without loss in generality that any EOF is included at the end of the MSG, prior to the ACK.

The waiting time referred to in step 1 prior to transmission of the ACK refers to a queueing delay on a data link with a memory buffer, or to a contention delay on a radio link. To analyse various access protocols in terms of delay, it is necessary to quantify the delay throughput behaviour described by Fig. 2-3.

III-2. Contention Access Delays

In the following chapters it will be necessary to use delay-vs-throughput results of the kind displayed in Fig. 2-3. Attempts were therefore made to fit the curves with various functions. The functional form finally selected is as follows, where Y_o is the waiting time, normalized relative to the fixed length of packets contending for channel access (i.e. Y_o is the normalized delay in Fig. 2-3):

$$Y_o = \begin{cases} a\,\rho^n & 0 \leqslant \rho \leqslant \rho_c \qquad (3\text{-}1a) \\ b\rho^k/(1 - \frac{\rho}{\rho_o}) & \rho_c \leqslant \rho \leqslant \rho_o \qquad (3\text{-}1b) \end{cases}$$

Constants a, b, k, n and p_c were selected to fit Tobagi's [T2] data points, and ρ_o is the maximum throughput for the protocol in question. For a pure

ALOHA and slotted non-persistent CSMA with a = 0.01, ρ_o = 0.184 and 0.857, respectively. The form of (3-1) is relatively simple and provides an excellent fit (so good that (3-1) overlays the existing curves) to all data points except for the CSMA case with $\rho = 0.15$. In this case, the estimate for Y_o given by (3-1) is too low; however this is not too serious since the value of Y_o is very low in this region. Eqn. (3-1) provides that $p \to \infty$ as $\rho \to \rho_o$ and fits the perfect scheduling case with $\rho_c = 0$, $\rho_o = 1$, $b = 1/2$ and $k = 1$. The parameters chosen for pure ALOHA slotted np-CSMA (a = 0.01) scheduling are listed in Table 3-1.

In fitting the curves in (3-1) regression techniques were not used because of the relatively low number of data points. Instead, pairs of data points were used for each region fitted as follows: $(\rho, Y_o) = (0.05, 1.2)$; (0.10,4); (0.15,40) for pure ALOHA and $(\rho, Y_o) = (0.3, 0.4)$; (0.56,2.0); (0.75,20) for slotted np CSMA (a = 0.01). The (0.56,2.0) pair fit the curve in Fig. 2-3 rather than Tobagi's [T2] actual data point.

In an earlier report we used the approximation $Y_o = \exp(c\rho) - 1$ with $c = \log_e 80$ for pure ALOHA. This approximation is, in fact, adequate for $Y_o \lesssim 80$, and longer delays would normally be unacceptable in an actual system. However, (3-1) is more satisfying because of the $\left(1 - \frac{\rho}{\rho_o}\right)$ factor in the denominator which gives the correct behaviour as $\rho \to \rho_o$.

Protocol	Parameters					
	a	n	ρ_c	b	k	ρ_o
pure ALOHA	214	1.733	0.10	5014	3.44	0.184
np CSMA (slotted, a = 0.01)	9.12	2.62	0.56	8.84	4.39	0.857
perf. scheduling	-	-	0	0.5	1.0	1.0

$$y_o = \begin{cases} a\ \rho^n & 0 \leqslant \rho \leqslant \rho_c \\ b\ \rho^k/(1 - \frac{\rho}{\rho_o}) & \rho_c \leqslant \rho \leqslant \rho_o \end{cases}$$

Table 3-1 Parameter values for Y_o vs ρ in (3-1), to fit curves in Fig. 2-3.

III-3 Radio Environments and Hidden Terminals

There are various radio traffic environments which might be considered as follows:

Case A:

All terminals generate messages with identical statistics.

Case B:

As in case A, except that some terminals generate many more messages than others.

Case C:

As in Case A, except that one terminal generates as many messages as all others together.

Case A applies when there is no centralized node (base). Most of the existing analytical results apply to this case. Cases B and C apply when some nodes carry more information than others; for example when all communication is between remote terminals (mobiles) and one or more central nodes (bases).

There are also various alternatives regarding radio transmission paths:

Case 1. All terminals (mobiles and bases) are within line of sight of each other.

Case 2. All mobiles enjoy line-of-sight paths with the associated base, but some mobiles are hidden from other mobiles.

Case 3. Line-of-sight communication between all mobiles and all other mobiles or bases is not assured.

In case 1, channel sensing is always possible, and any radio channel can be used to carry the entire REQ/ANS/MSG/ACK sequence, in half-duplex mode.

Alternatively REQ/ANS sequences can be carried on dedicated access channels again in half-duplex mode. Case 1 provides an upper bound obtainable on spectrum effeciency for a given REQ, ANS, MSG, and ACK structure. Unfortunately most mobile radio systems do not meet the universal line-of-sight requirements of Case 1.

In Case 2, channel sensing of inbound transmissions by a mobile with a ready REQ packet is not feasible, and if sent this REQ packet could collide with another inbound REQ, ANS, MSG, or ACK if all of these were to be sent on the same half-duplex channel. There are several ways to handle the difficulty, as follows:

1. If all transmissions (REQ/ANS/MSG/ACK) were on a single half-duplex channel, then a portion of the channel could be used to carry a busy-tone signal from the base site indicating the state of the channel. The presence of such a tone allows channel sensing but reduced the spectrum by approximately 15% [T2,K2].

2. Paired half-duplex channels may be used. One of the pair would then carry all inbound (mobile-to-base) traffic while the other would carry all outbound (base-to-mobile) traffic. As well, the outbound channel could carry a busy tone signal (when not transmitting REQ, ANS, MSG, or ACK's) indicating the status of the inbound channel. Such a scheme would allow channel sensing but would reduce the spectrum efficiency to one-half the value obtainable under Case 1 conditions. If the outbound channel carried information to mobile(s) while receiving inbound information from mobiles (full-duplex) then the spectrum efficiency would lie beetween that which occurs for busy tone operation and paired-

channel half-duplex operation. In this latter case the presence of an outbound signal is not sufficient to indicate the inbound channel status, and "busy-bits" [D2] as some other method must be used to indicate inbound channel status.

3. Use dedicated access channels to carry all REQ and ANS packets. In this case the remaining channels could be used in half-duplex mode to carry MSG's and ACK's. The access channels could operate in half-duplex mode with busy tone signalling or could be paired; either alternative would allow for CSMA transmission of REQ packets. Alternatively, REQ and ANS packets could be carried on the same half-duplex channel in which case an ALOHA acces protocol for REQ packets would be appropriate. Prior to transmission of a REQ any terminal would listen for τ sec to ensure that an ANS was not occupying the channel. The REQ packet would then either be sent or rescheduled as appropriate. There would still remain some possibility of REQ-ANS collisions (due to the propagation delay uncertainty inherent in CSMA), but such collisions would be infrequent, occurring with probability $p = 1 - e^{-aG}$. For $G \simeq 1$ and $a = 0.01$, $p \simeq 10^{-2}$. The result would be to increase the REQ packet transmissions and average delay by the factor $(1-p)^{-1} \simeq 1.01$. ANS packet would always be transmitted immediately without any τ-sec. delay. We consider this matter in more detail in Chapter 5.

In those Case 3 situations where the base is hidden from some of its mobiles, a repeater would be placed within line of sight of all mobiles and bases. Repeater frequencies would be paired, since simultaneous (analog) reception and retransmission would require different frequencies. One could require that all transmissions to the repeater be on one frequency, with all transmissions from the repeater on the other frequency. The signal from the repeater would be available to indicate the state of any channel-pair. The throughput would be equal to 50% of the value obtainable under Case 1 conditions. Access could be either via message channels or via dedicated access channels.

The presence of a repeater would permit mobile-to-mobile transmissions; however we do not consider this possibility is this report.

In some cases all base sites could communicate with the repeater via a dedicated wire line, and the situation would then be similar to Case 2.

III-4 Mobile Data Formats and Retransmissions

With reference to Fig. 3-1 we denote the length of the REQ, ANS, MSG and ACK packets as L_Q, L_A, L_M and L_K, respectively. We denote λ as the (Poisson) message generation rate, and $\rho_i = \lambda L_i$ where index i assumes values Q, A, M or K. Since ρ_i denotes utilization resulting from transmission of various packet types, we include a τ-sec delay in the lengths L_i. To denote combined packet lengths or utilizations, we use the symbol $L_{ij\ldots k}$ (or $\rho_{ij\ldots k}$). For example $L_{QAMK} = L_Q + L_A + L_M + L_K$ and $\rho_{QAMK} = \rho_Q + \rho_A + \rho_M + \rho_K$.

Various formats have been proposed for data transmission. Appendix II details one of these. It follows that distinct REQ, ANS, MSG and ACK packets would all be at least 96 bits in length under this format. At 2400 b/s, the transmission time for a 96-bit REQ packet, for example, would equal 40 ms. A data message consisting of one 80-character line would, at 8 bits/character contain 640 information bits, 640/3 check bits and 33 preamble, synch and word indicator bits for a total of (approximately) 885 bits. At 2400 bits/sec, $L_M = 370$ ms. In this case $L_M/L_Q \simeq 9.2$. Transmission of 11 lines of text would make $L_M/L_Q \simeq 100$.

The beginning of any transmission involves various start-up overheads which depend in part on the equipment employed. Examples of such overheads include the start-up time for transmitters, and the time for receivers to detect the presence of a signal [D2]. The preamble would absorb some of these times, and we assume therefore that these times are either negligible or included in the packet preamble. Explicit inclusion is straight forward once the values of any overheads are known.

In the following chapters it will be necessary to consider the effects of retransmissions in any delay analysis. Let the probability of retransmitting a packet be P_r. Then the mean number of retransmissions is as follows:

$$\begin{aligned} RET &= \sum_{i=1}^{\infty} P_r^{\,i} \\ &= \left(\sum_{i=0}^{\infty} P_r^{\,i}\right) - 1 \\ &= P_r/(1 - P_r) \qquad (3\text{-}2) \end{aligned}$$

Within the data transmission formats in Appendix II some flexibility remains. For example, in a long data message it may be efficient to send ACK's following a certain number of data words, even if the complete message has not been sent, in order to avoid the need to retransmit a long message. Alternatively, ACK's may be sent less frequently and indicate which packets should be transmitted. One could use some of the check bits for forward error correction [F2]. The question of how to best use check bits and ACK's for optimum performance is a subject which would involve detailed and careful study [E1,F2].

IV - ACCESS VIA MESSAGE CHANNELS

IV-1 Introduction

In this chapter we assume that all of the communication diagrammed in Fig. 3-1 is over the same channel or channel pair. We begin with the basic result (Little's formula [K2,K3]):

$$D = RET \cdot Y + \tau + L_{QAMK} \tag{4-1}$$

where RET and Y denote, respectively, the mean number of transmissions and the mean time between retransmissions. For CSMA, the term retransmissions includes any rescheduling of transmissions when the channel is sensed busy. The τ-sec delay occurs prior to transmission of any REQ packet. As explained earlier this wait avoids collisions with non-REQ packets, and gives the ANS/MSG/ACK sequence priority over REQ's.

The channel is selected from among m ($m \geq 1$) channels or channel pairs on which the mobile can transceive. We ignore for now any time required to determine the state of a channel. Formally, the channel selection protocol is as follows:

1. A terminal with a ready REQ searches continuously for a free channel.
2. When a free channel is found a wait of τ sec. occurs. If the channel remains free after τ sec. the channel is deemed available for REQ packets, and the REQ access protocol begins.
3. If the channel becomes busy before completion of the τ-sec. time interval, step 1 begins again (Note that the channel will <u>not</u> become busy as a result of a REQ packet transmission, since the τ-sec. wait is imposed on <u>all</u> REQ packets.)

IV-2 Mean Number of Retransmissions

We now determine the mean number of REQ packet retransmissions RET, in terms of P_{r_o} and P_U; P_{r_o} is the probability of a REQ retransmission resulting from a collision (in the case of ALOHA) or a rescheduling (in the case of CSMA) caused by the presence of other REQ packets, and P_U is the probability that the channel is unavailable for REQ packets because of the presence of the ANS/MSG/ACK sequence.

The total retransmission probability is

$$P_r = P_{r_o}(1 - P_U) + 1 \cdot P_U \tag{4-2}$$

From (3-2) it follows that

$$\begin{aligned} RET &= \frac{P_{r_o}(1 - P_U) + P_U}{1 - [(1 - P_U)P_{r_o} + P_U]} \\ &= \frac{P_{r_o}(1 - P_U) + P_U}{(1 - P_U)(1 - P_{r_o})} \\ &= \frac{P_{r_o}}{1 - P_{r_o}} + \left[\frac{P_U}{1 - P_U}\right]\left[\frac{1}{1 - P_{r_o}}\right] \\ &= RET_o + (RET_o + 1)\, RET_U \end{aligned} \tag{4-3}$$

where

$$RET_o = P_{r_o}/(1 - P_{r_o}) \tag{4-4}$$

$$RET_U = P_U/(1 - P_U) \tag{4-5}$$

The number of retransmissions in (4-3) is a sum of two terms; RET_o equals the mean number of retransmissions when the channel is available for access, and $(RET_o + 1)(RET_U)$ equals the mean number of transmissions which occur as a result of the channel being inaccessible because of ANS/MSG/ACK transmissions.

IV-3 Basic Delay Relations

From (4-1) the delay D can be written as follows

$$D = w \cdot L_Q + L_{QAMK} + \tau \tag{4-6}$$

$$w = Y_o + Y_U \tag{4-7}$$

$$Y_o = Y \cdot RET_o \tag{4-8}$$

$$\begin{aligned} Y_U &= Y \cdot (RET_o + 1)\, RET_U \\ &= Y_o (1 + RET_o{}^{-1}) RET_U \end{aligned} \tag{4-9}$$

where Y_o and Y_U denote delays (normalized relative to L_Q) which arise from retransmissions during access periods and during ANS/MSG/ACK transmission periods, respectively.

The quantity Y has been determined using simulation techniques in which the mean retransmission delay $\overline{X}$ was chosen to minimize Y_o [T2,K2,K3]. Actually, we should add the effect of the τ-sec REQ delay to avoid ANS collisions, but this we ignore in comparison with $T_Q + T_A + 2\tau + \overline{X}$.

From [K3],

$$RET_o = ((G/S) - 1) \tag{4-10}$$

where G/S denotes the ratio of the offered traffic rate G to the throughput rate S [K2,T2]:

$$\frac{G}{S} = \frac{1 - e^{-aG} + a}{ae^{-aG}} \qquad \left(\begin{array}{c}\text{slotted}\\ \text{np CSMA}\end{array}\right) \tag{4-11}$$

$$\frac{G}{S} = \frac{G(1 + 2a) + e^{-aG}}{e^{-aG}} \qquad \left(\begin{array}{c}\text{np}\\ \text{CSMA}\end{array}\right) \tag{4-12}$$

In both case as $a \to \infty$

$$\frac{G}{S} = 1 + G \tag{4-13}$$

For pure and slotted ALOHA, respectively [K2,T1]

$$G/S = e^{G} \qquad \text{(slotted ALOHA)} \tag{4-14}$$

$$G/S = e^{2G} \qquad \text{(pure ALOHA)} \tag{4-15}$$

We can now write the wait w in terms of known quantities, as follows, and the delay is then given by (4-6).

$$w = Y_o \left[1 + \left(\frac{G/S}{(G/S) - 1}\right) RET_U\right] \tag{4-16}$$

$$RET_U = P_M(\rho_{AMK})/(1 - P_M(\rho_{AMK})) \tag{4-17}$$

In (4-16) Y_o is given by the equations in Section 3-2 which fit the experimentally determined delay Y_o vs. throughput S, where in this case in accordance with Appendix I

$$S = \rho_Q/(1 - \rho_{AMK}) \tag{4-18}$$

The busy channel probability P_M depends on ρ_{AMK} in accordance with (2-3), in the case of exponentially distributed packets. For fixed length packets, an analytical expression for P_M exists for m = 1 only, in which case $P_M = \rho_{AMK}$.

IV-4 "Perfect" CSMA ($a \to 0$)

Before obtaining general results we consider some special cases, the first of which occurs when $a \to 0$. This case corresponds to perfect channel sensing. Some of the previous equations simplify, as follows:

$$(G/S)/\left(\frac{G}{S} - 1\right) = \frac{1 + G}{G}$$

$$= 1/S \qquad (a \to 0) \tag{4-19}$$

where S is again given by (4-18). In this case for m = 1 channel,

$$w = Y_o\,[1 + (RET_U/S)]$$
$$= Y_o\left[1 + \left(\frac{\rho_{AMK}}{1 - \rho_{AMK}}\right)\left(\frac{1 - \rho_{AMK}}{\rho_Q}\right)\right]$$
$$= Y_o \cdot (\rho_{QAMK}/\rho_Q) \qquad (a \to 0) \qquad (4\text{-}20)$$

Since $\rho_{QAMK}/\rho_Q = L_{QAMK}/L_Q$, the effect of (4-20) is to normalize the wait with respect to the length of the REQ/ANS/MSG/ACK sequence rather than relative to the REQ sequence length only.

As $a \to 0$ we would expect the result to be identical to that for perfect scheduling; thus we would expect

$$Y_o = (\rho_{QAMK}/2)/(1 - \rho_{QAMK}) \qquad (4\text{-}21a)$$

However, with S given by (4-18),

$$Y_o = (S/2)(1 - S)$$
$$= (\rho_Q/2)(1 - \rho_{QAMK}) \qquad (4\text{-}21b)$$

Since (4-21a) and (4-21b) are identical only if $\rho_{AMK} = 0$, the result for w in (4-16) might well understate the actual waiting time, particularly for $\rho_Q \ll \rho_{AMK}$ when $\rho_{AMK} \to 1$. We note that there is an implicit assumption in deriving w, namely that periods of unavailability given by P_U occur randomly, which is not really the case; busy periods <u>always</u> follow a REQ and continue uninterrupted for L_{AMK} sec. In Section IV-6, we examine these matters further.

For slotted np CSMA simplifications occur for small <u>a</u> values when $G \ll 1$. In this case

$$\frac{G}{S} = \frac{(1+a)\,e^{aG} - 1}{a} \qquad (aG \ll 1) \qquad (4\text{-}24)$$
$$\simeq 1 + (a + 1)G \qquad (aG \ll 1) \qquad (4\text{-}22)$$
$$\frac{G/S}{(G/S) - 1} \simeq 1/(1 + a)S \qquad (aG \ll 1)$$

and

$$w = Y_o \, [\rho_{Q(1+a)AMK} / \rho_{Q(1+a)}] \qquad (aG \ll 1) \qquad (4\text{-}23)$$

where

$$\rho_{Q(1+a)} = (1 + a)\rho_Q \qquad (aG \ll 1) \qquad (4\text{-}24)$$

The effect here is to multiply ρ_Q by the factor $(1 + a)$.

IV-5 Maximum Throughput (∞ Delay)

We next determine the maximum information throughput θ (spectrum efficiency).

There are two inherent constraints on throughput. The first is the limitation on the REQ utilization S. The maximum value for S for ALOHA, CSMA and perfect scheduling is known; for example for pure ALOHA $S \leqslant 1/2e$ and for perfect scheduling $S \leqslant 1$.

If we denote S_m as the maximum permitted value for S, we determine θ by first solving (4-18) for ρ_M:

$$S(1 - \rho_{AK} - \rho_M) = \rho_Q \qquad (4\text{-}25)$$

$$\theta = \beta_M \rho_M \qquad (4\text{-}26)$$

The other constraint on θ results from the requirement

$$\rho_{AMK} \leqslant 1 \qquad (4\text{-}27)$$

The requirement in (4-27) is met automatically if (4-25) and (4-18) are satisfied with $S \leqslant S_m$.

In our analysis we assume equal lengths for REQ, ANS and ACK packets, in which case

$$\rho_i = (k/c)\rho_M \qquad i = A,Q \qquad (4\text{-}28)$$

$$\rho_K = k\rho_M \qquad (4\text{-}29)$$

where $k \leqslant 1$ and $c \geqslant 1$. The constant c allows for an ACK following each of a prespecified number of data packets. If $c = 1$ then no ACK occurs until the end of the message.

Substitution of (4-28) and (4-29) into (4-25) yields

$$S[1 - ((1+c^{-1})k+1)\rho_M] = kc^{-1}\rho_M \tag{4-30}$$

Solution of (4-30) for ρ_M yields

$$\rho_M = S/[kc^{-1} + S(1 + (1+c^{-1})k)] \tag{4-31}$$

Given the maximum allowable value for S, ρ_M and θ can be calculated in terms of c and k.

We now consider the data format in Appendix II, to obtain specific results. To find a general expression for β_M we define d as the number of successive 64-bit data words in one line of a MSG packet. We assume that ANS and ACK packets consist of 96 bits, that an ACK follows each line (unless otherwise stated), that a single REQ/ANS sequence preceeds each transmission and that transmissions are error free. We also assume (unless otherwise stated) that messages consist of full lines. It follows that c equals the number of lines and that

$$k = 96/(96 + 64d) \tag{4-32}$$

Since 16 information bits can be included in an address packet and 47 bits in a data packet (see Appendix II), and since each full MSG packet contains 96 + 64d bits,

$$\beta_M = (16 + 47d)/(96 + 64d) \tag{4-33}$$

These results also apply to single-line messages if we let $c = 1$ and d denote the number of data words in the message. For a very short MSG consisting of one address packet and one data packet, $d = 1$ and $\beta_M = 0.39$.

Table 4-1 shows β_M, k, ρ_M and θ for various values of c and d, with S_m

d	c	k	β_M	S_m					
				0.184	0.37	0.518	0.815	0.857	1.0
1	1	0.60	0.39	0.183	0.262	0.298	0.341	0.345	0.357
6	1	0.20	0.62	0.402	0.515	0.560	0.608	0.612	0.625
12	1	0.11	0.67	0.548	0.657	0.696	0.736	0.740	0.750
12	3	0.11	0.67	0.741	0.801	0.820	0.838	0.839	0.844
12	20	0.11	0.67	0.872	0.883	0.887	0.890	0.890	0.891
240	1	0.006	0.731	0.956	0.971	0.976	0.980	0.981	0.982

(a) ρ_M

1	1	0.60	0.39	0.072	0.103	0.117	0.134	0.136	0.141
6	1	0.20	0.62	0.250	0.320	0.348	0.378	0.380	0.388
12	1	0.11	0.67	0.368	0.441	0.467	0.494	0.497	0.503
12	3	0.11	0.67	0.497	0.538	0.550	0.562	0.563	0.566
12	20	0.11	0.67	0.585	0.593	0.595	0.598	0.598	0.598
240	1	0.006	0.731	0.699	0.710	0.713	0.716	0.717	0.717

(b) θ

Table 4-1 ρ_M and θ vs d, k, β_M and S_m. Access via Message Channels.

corresponding to access protocols as follows:

$$S_m = \begin{cases} 0.184 & \text{pure ALOHA} \\ 0.37 & \text{slotted ALOHA} \\ 0.518 & \text{np CSMA } (a = 0.1) \\ 0.815 & \text{np CSMA } (a = 0.01) \\ 0.857 & \text{slotted NP CSMA } (a = 0.01) \\ 1.0 & \text{perfect scheduling} \end{cases} \qquad (4\text{-}34)$$

In arriving at Table 4-1 we assumed 72-character lines of text (72 x 8 = 576 information bits). Thus, for full-line messages d = 12 corresponding to 580 bits (slightly more than one line) and $\beta_M = 0.67$. One sees a marked difference in spectrum efficiency, dependent on the message length and access protocol.

From (4-28), $kc^{-1} << 1$ and $c >> 1$ for very long messages; (4-29), (4-31) and (4-32) show that with $d = 14$, $k \simeq 0.11$ and $\rho_M \simeq 0.9$ since (4-31) yields in this case

$$\rho_M \simeq (1 + k)^{-1} \qquad (kc^{-1} << 1) \qquad (4\text{-}35)$$

Since $\beta_M = 0.67$, $\theta = (0.9)(0.67) = 0.60$ in this case, independent of the access protocol.

For very long messages where ACK's are infrequent, $k << 1$ and $kc^{-1} << 1$. In this case $\rho_M \simeq 1$ independent of the access protocol. For voice messages, $\beta_M \simeq 1$ and $\theta \simeq 1$, again independent of the access protocol. One could argue that for voice messages not subject to redundancy reduction techniques, there is in fact much inherent redundancy and that $\theta = 1$ is misleading.

For long text or graphic messages with infrequent ACK's, some error control bits for FEC may be included making $\beta_M < 1$ [F2]; in this case $\theta \simeq \beta_M$.

IV-6 Delay vs Throughput Calculation Algorithm

The following algorithm was used to determine the delay D vs throughput θ based on w as calculated using (4-16):

1. Input d and c, and from these calculate k from (4-32) and β_M from (4-33).

2. Input G_o, the channel utilization which maximizes throughput S for the access protocol used. (i.e. G = 0.5 for pure ALOHA and $\simeq$ 10 for np CSMA for a = 0.01). Calculate S for various values of $G < G_o$.

3. Calculate ρ_M from (4-31).

4. Calculate θ from (4-26).

5. Calculate ρ_{AMK} as follows:

$$\rho_{AMK} = [1 + k(1 + c^{-1})]\rho_M \qquad (4\text{-}36)$$

6. Calculate $P_M(\rho_{AMK})$ from (2-3).

7. Calculate RET_U in (4-17).

8. Calculate $(G/S)/[(G/S) - 1]$.

9. Calculate w using (4-16).

10. Calculate D using (4-6).

11. Plot D vs θ.

In executing the above algorithm we found the calculated waiting time to be less than that for perfect scheduling of the REQ/ANS/MSG/ACK sequence for $\rho_Q \ll \rho_{AMK}$ and for utilizations θ approaching the maximum obtainable. The fact that busy periods are not really independent and random causes w in

(4-16) to be understated. Initially, we were somewhat surprised by the result, particularly since we subsequently found that Tobagi [T4] uses an almost identical equation in a somewhat different but substantially equivalent situation, where in effect $\rho_Q = \rho_{AMK}$. Some error in w would result from our approximation of the curves in Fig. 2-3; however we carefully checked to ensure that these inaccuracies could not explain our anomolous results. In fact, the problem is the apparently unacceptable assumption of independent busy periods. In Section IV-8 we propose another method to calculate D vs θ.

IV-7 Delay-vs-Throughput for Perfect Scheduling

The best possible delay-throughput performance occurs under perfect scheduling of the REQ/ANS/MSG/ACK sequence. This case would apply if all (error-free) transmissions originated from the base in a single-based system. In this case the delay is a queueing delay.

For fixed-length messages on m = 1 channel, the delay is

$$D = \left[\frac{1 - (\rho_{QAMK}/2)}{(1 - \rho_{QAMK})}\right] L_{QAMK} + \tau \tag{4-37}$$

where from (4-28) and (4-29)

$$\rho_{QAMK} = \rho_M(1 + 2kc^{-1} + k) \tag{4-38}$$

Fig. 4-1 shows (D/L_M) vs θ, where

$$D/L_M = \left[\frac{1 - (\rho_{QAMK}/2)}{1 - \rho_{QAMK}}\right] [1 + 2kc^{-1} + k] \tag{4-39}$$

where we assume $\tau << L_M$.

The minimum delay in (4-39) is $1 + 2kc^{-1} + k$.

The obtainable throughput θ never exceeds 60% for d = 12 even as $c \to \infty$,

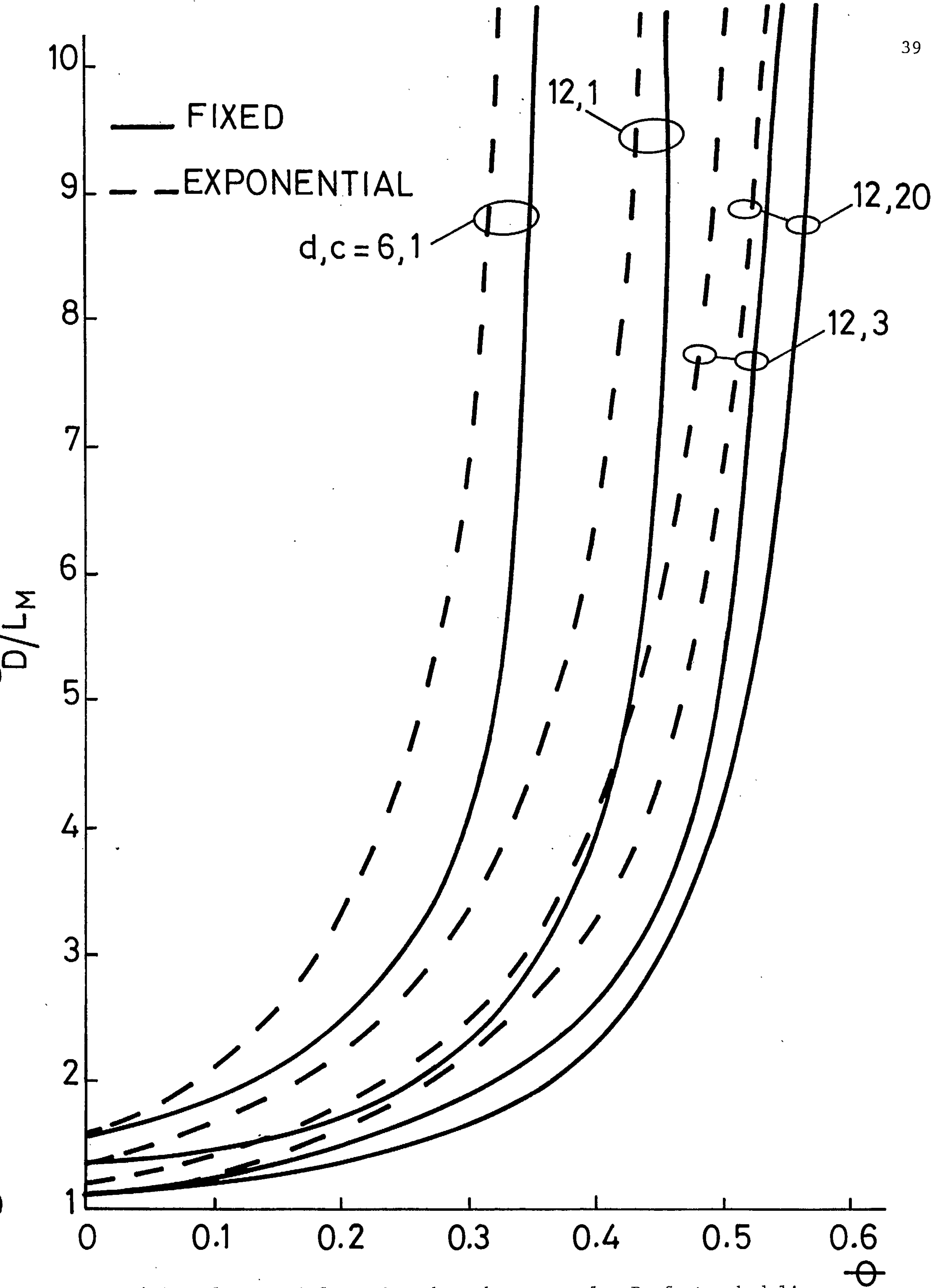


Fig. 4-1 Delay vs. information throughput, m = 1. Perfect scheduling.

because of the preamble synch and check bit overheads. For $d = 6$ and $c = 1$, the upper limit on θ is 39%. (See also Table 4-1).

If one requires that the delay be bounded, the obtainable throughputs are considerably smaller. For example consider the constraint $D/L_M < 3$ which corresponds to a wait/transmit time ratio of 2:1. Then the maximum throughput is $\theta = 25\%$ for a single half-line fixed length message ($d = 6$, $c = 1$); 35% for a full-line message ($d = 12$, $c = 1$); 42.5% for 3 full lines ($d = 12$, $c = 3$) and 45% for 20 full lines ($d = 12$, $c = 20$). If one assumes exponentially distributed REQ/ANS/MSG/ACK sequences then

$$D = (1 - \rho_{QAMK})^{-1} L_{QAMK} + \tau \tag{4-40}$$

where L_{QAMK} is now an average length and

$$D/L_M = (1 + 2kc^{-1} + k)/(1 - \rho_{QAMK}) \tag{4-41}$$

D/L_M vs θ for this case appears on Fig. 4-1 as well, for comparison purposes. The delay is larger than for fixed length packets of the same average length, and θ for D/L_M values of 3 in Fig. 4-1 are 0.18, 0.275, 0.34 and 0.38.

Transmission capability on $m > 1$ channels improves θ for any given D/L_M value. Figs. 4-2 and 4-3 show the improvement for perfect scheduling of exponentially distributed length L_{QAMK}, calculated as follows:

$$D/L_M = (1 + 2kc^{-1} + k)(1 + [P_m(\rho_{QAMK})/m(1 - \rho_{QAMK})]) \tag{4-42}$$

where $P_m(\rho)$ is given by (2-3).

The effects on θ of being able to communicate on more than one channel are clear. For example, with $D/L_M = 3$ and $(d,c) = (6,1)$ θ increases from 0.18 for $m = 1$ to 0.265, 0.312, 0.36 for $m = 2$, 4 and 8 respectively. There

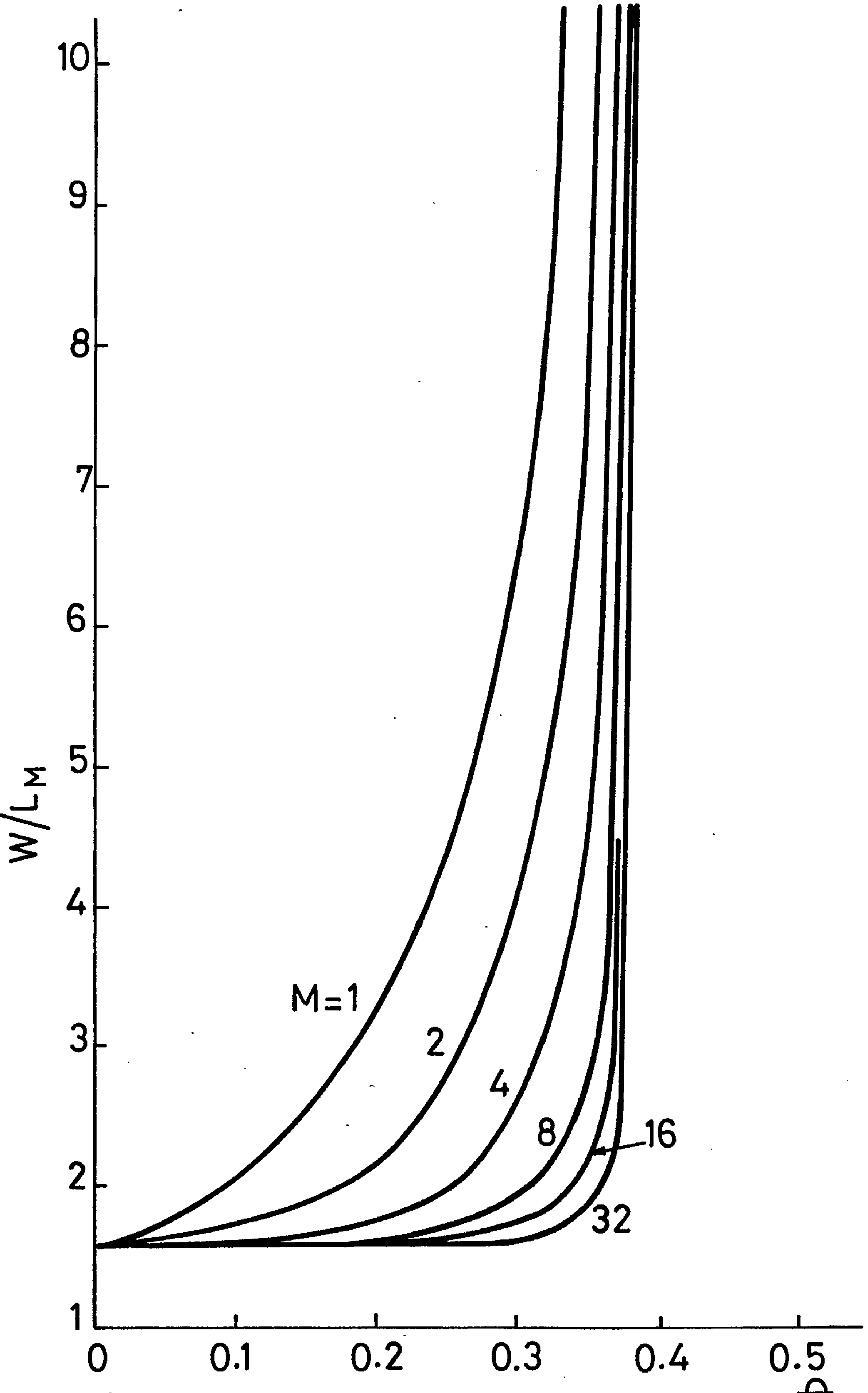


Fig. 4-2 Delay vs. information throughput; m channels. Perfect scheduling (d,c) = (6,1)

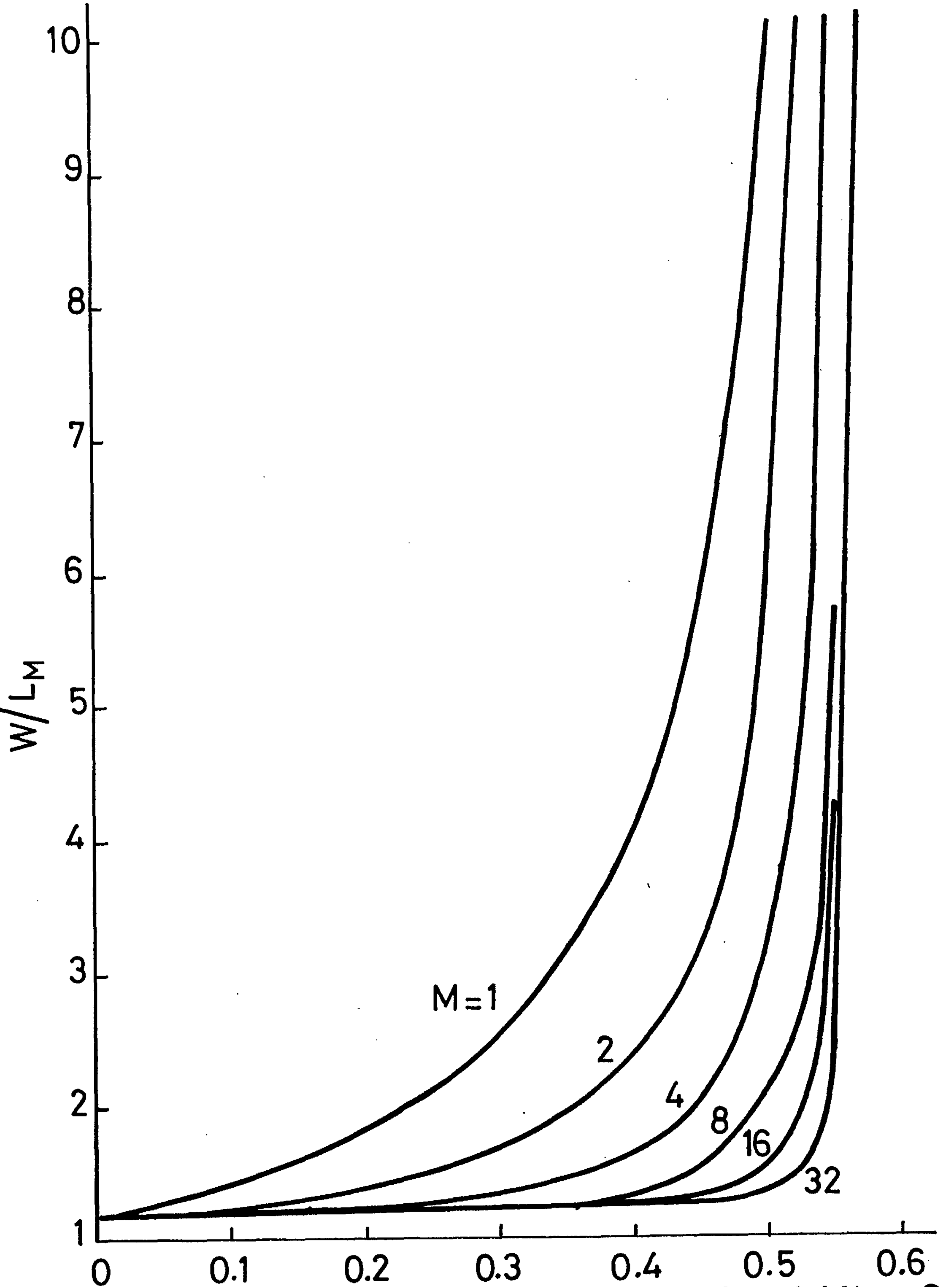


Fig. 4-3 Delay vs. information throughput; m channels. Perfect scheduling.
(d,c) = (12,3)

is not much improvement beyond m = 8.

The above analysis ignores the time needed to switch channels in searching for one which is free and also assumes that all channels can be simultaneously monitored for REQ packets. Mobiles often operate on one channel only because of equipment limitations or regulatory requirements. Thus, the advantages of multi-channel communication are often not directly available. Again, we note that multichannel capability does not imply any improvement in maximum throughput for large delays, but does substantially improve throughput at delays on the order of a few message lengths.

We not that all θ values assume half-duplex channels. For paired channels in half-duplex mode, θ values are reduced by 50%. For busy-tone signalling θ values are reduced by approximately 15% as explained earlier.

IV-8 An Alternate Approach to CSMA Delay

We now develop an alternative approach to delay calculation when access is via message channels.

In Fig. 4-4 we show time-lines as seen at each terminal, including REQ/ANS/MSG/ACK transmissions and periods of contention.

Consider now that packets with lengths L_1 through L_5 in Fig. 4-4 queue and are served in the order shown. Assuming Poission arrivals (not strictly true) the waiting time and delay are again given by the queuing relations in (2-3) to (2-6), since service order is random and independent of message length [K2]. The time added to the front end of any packet equals the time to the beginning of its service from either: (a) the completion of service

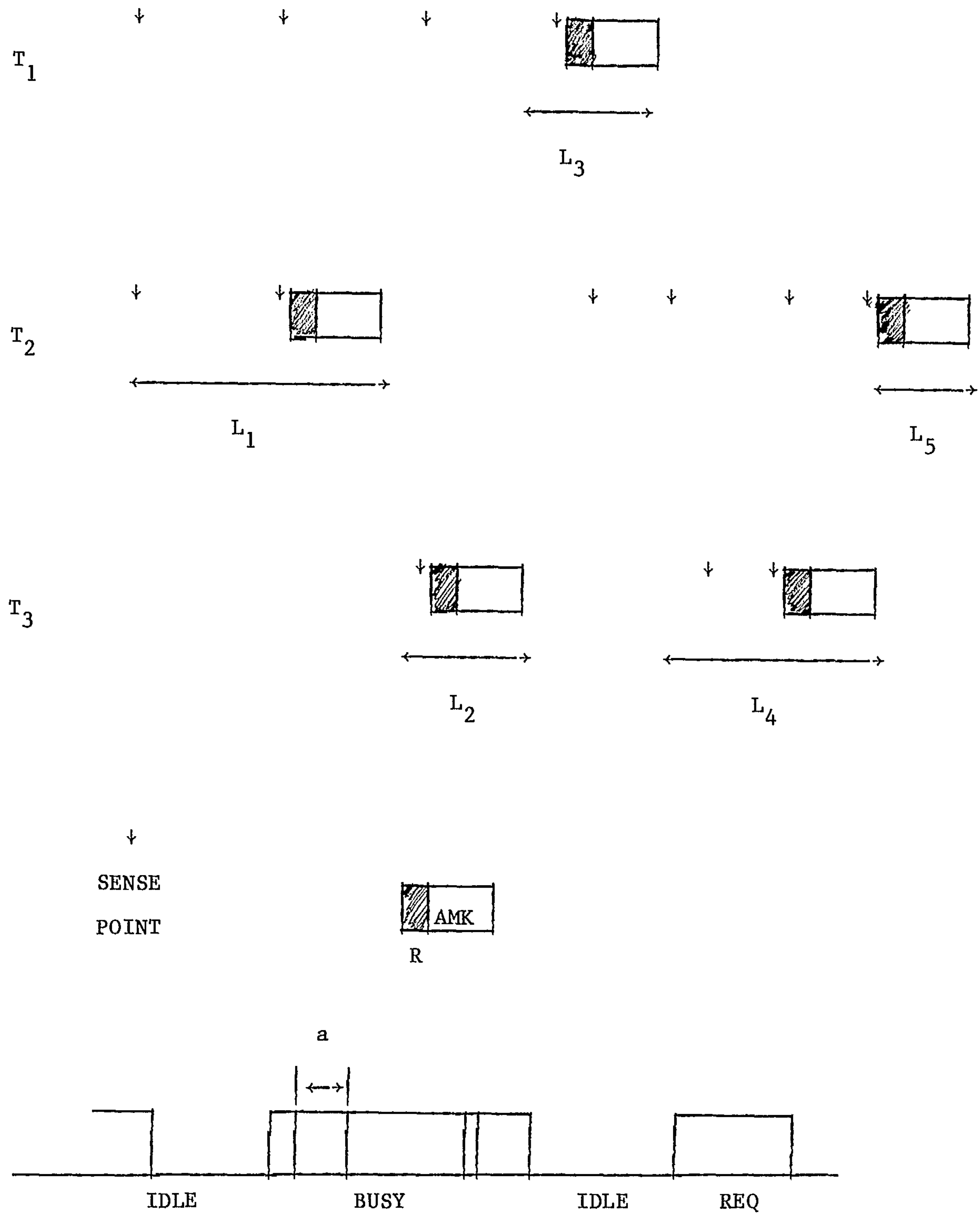


Fig. 4-4 Illustrating packet arrivals at terminals T_i. CSMA access control.

of the previous packet, or (b) the beginning of a contention period. A contention period begins whenever the possibility of mutual interference of packets begins.

To calculate the mean time to beginning of service, we note that the idle periods in Fig. 4-4 average 1/G sec, and from [T1,T2] the busy-period average $\overline{B}$ (illustrated in Fig. 4-4) is, for np CSMA

$$\overline{B} = 1 + 2a - \left(\frac{1 - e^{-aG}}{G}\right) \tag{4-43}$$

Note that $\overline{B}$, $\underline{a}$ and G are <u>normalized relative to L_Q</u>.

Using the fact that the probability of no collision within $\underline{a}$ sec is exp(-aG) we calculate the mean time to first successful REQ transmission during a contention period:

$$T_c/L_Q = \frac{1}{G}\, e^{-aG} + \left(\frac{2}{G} + \overline{B}\right)(1 - e^{-aG})e^{-aG} + \left(\frac{3}{G} + 2\overline{B}\right)(1 - e^{-aG})^2 e^{-aG} + \ldots$$

$$= e^{-aG}\left[\frac{1}{G}\sum_{i=0}^{\infty} (i + 1)(1 - e^{-aG})^i + \overline{B}\sum_{i=0}^{\infty} i(1 - e^{-aG})^i\right] \tag{4-44}$$

We note that the sums above can be evaluated as follows, with

$$x = 1 - e^{-aG}$$

$$\sum_{i=0}^{\infty} (i + 1)x^i = \sum_{j=0}^{\infty} jx^{j-1}$$

$$= \frac{d}{dx}\sum_{j=0}^{\infty} x^j$$

$$= \frac{d}{dx}(1 - x)^{-1}$$

$$= (1 - x)^{-2}$$

$$\sum_{i=0}^{\infty} ix^i = x\sum_{i=0}^{\infty} ix^{i-1}$$

$$= x/(1 - x)^2$$

Thus, during a contention or busy period

$$T_c = e^{aG}\left[\frac{1}{G} + \overline{B}\ (1 - e^{-aG})\right]L_Q \qquad (4\text{-}45)$$

If a packet arrives when no other packets are present or contending then access is immediate and nothing is added to the front end of the packet.

To include the effects of immediately successful transmission of a REQ packet, we multiply T_c in (4-45) by the utilization $\rho = \rho_c + \rho_{QAMK}$ where ρ is the contention plus transmission utliization:

$$\begin{aligned}\rho_c &= T_c \cdot \rho/L_{QAMK} \\ &= (T_c/L_Q)(kc^{-1}/(1 + k + 2kc^{-1}))\rho \qquad (4\text{-}46)\end{aligned}$$

$$\begin{aligned}\rho &= \rho_c + \rho_{QAMK} \\ &= \overline{t}_c\rho + \rho_{QAMK} \\ &= \rho_{QAMK}/(1 - \overline{t}_c) \qquad (4\text{-}47)\end{aligned}$$

where

$$\overline{t}_c = e^{aG}\left[\frac{1}{G} + \overline{B}\ (1 - e^{-aG})\right][kc^{-1}/(1 + k + 2kc^{-1})] \qquad (4\text{-}48)$$

Since $\rho < 1$, ρ in (4-47) is upper bounded as follows:

$$\rho < \overline{t}_c + \rho_{QAMK} \qquad (4\text{-}49)$$

For CSMA with $a \ll 1$ and $G \gtrsim 1$,

$$T_c/L_Q \simeq 1/G \qquad (a \ll 1,\ G \gtrsim 1) \qquad (4\text{-}50)$$

Typically $G \sim 5$ for CSMA with $a = 0.01$, and $G \sim 1$ for $a = 0.1$. For a single line of text, $k = 0.11$ and $c = 1$, in which case $\overline{t}_c = 0.016$ or 0.0843 for $a = 0.01$ or 0.1 respectively. Thus, the increase in ρ appears minimal particularly if a $\approx$ 0.01, and the queueing results in the previous section would provide reasonably good throughput estimates for CSMA access protocols when $a \ll 1$.

We note that if messages were transmitted directly without the REQ/ANS sequence, then $\overline{t}_c$ is given by (4-45), with L_M replacing L_Q; for G = 5 and a $\approx$ 0.01, a 20% increase in utilization would occur, which would require a corresponding decrease in throughput θ by the factor $(1.2)^{-1} = 0.83$ to maintain the delay equal to its queuing value obtained with $\rho = \rho_{MK}$. Thus, a short REQ/ANS sequence is useful in reducing the absolute contention period when moderate or long messages are involved.

The same type of analysis procedures can be used for other access protocols. Known methods [T2] are available to calculate the busy period B. Although accurate generalizations are difficult, our work indicates that a pure ALOHA access protocol results in a busy period many times larger than that of 0.01 CSMA.

V - REQ/ANS SEQUENCE AND SPECTRUM EFFICIENCY

V-1 Introduction

We now examine some aspects of how the REQ/ANS sequence or its absence affects spectrum efficiency. We are particularly interested in comparisons against the m = 1 cases in Fig. 4-1, 4-2 and 4.3 since m = 1 is the operational condition in many situations.

V-2 Omission of the REQ/ANS Sequence

It is intuitively clear that omission of the REQ/ANS sequence would enhance spectrum efficiency when the MSG sequence is short.

When the MSG length is fixed at some constant value, the analysis leading to (4-16) is applicable. In such case (4-16) gives the delay D as follows, assuming that transmission involving any mobile is restricted to m = 1 channel.

$$D = w \cdot L_{MK} + L_{MK} + \tau \tag{5-1}$$

$$w = Y_o(s)[1 + [\frac{G/S}{(G/S) - 1}] RET_U] \tag{5-2}$$

$$= \frac{Y_o(s)}{1 - P_u} [1 - \frac{P_U}{RET_o}] \tag{5-3}$$

where in this case

$$S = \rho_M/(1 - \rho_K) \tag{5-4}$$

$$= \rho_M/(1 - k\rho_M) \tag{5-5}$$

Normally, $\rho_K << \rho_M$ and it is expected that D as given by (5-1) would be reasonably accurate because the assumption regarding the randomness of channel unavailability is more closely approximated than was the case in Chapter IV.

A reasonably good and simple approximation to the mean waiting time w is obtained for $\rho_K \overset{\sim}{<} \rho_M$ as follows:

$$w = \frac{Y_o[\rho_M/(1 - k\rho_M)]}{1 - k\rho_M} \tag{5-6}$$

Use of the approximation leads to Table 5-1 which compares the throughput θ obtainable for D = 3 and m = 1, with and without the REQ/ANS sequence, using np CSMA with a = 0.01.

One sees from Table 5-1 that for message lengths of up to one line of text, omission of the REQ/ANS sequence provides for improved throughput. For longer messages, the REQ/ANS sequence is beneficial.

Another comparison appears in Table 5-2 where maximum θ values (for arbitrarily large delay) are displayed, for the cases (d,c) = (12,1) and (12,20). These cases correspond to a single line and 20 lines of text with an ACK at the end of each line. Various $S = S_m$ values are used corresponding to the various access protocols discussed in conjunction with Table 4-1.

One sees the advantage of the REQ/ANS sequence for access protocols which are less efficient than the np CSMA (a = 0.01) case. In fact, for np CSMA with a = 0.1 and for ALOHA protocols, inclusion of the REQ/ANS sequence provides for an improved throughput even for the single line of text case. Only for the perfect scheduling case is there any significant advantage in omitting the REQ/ANS sequence.

For very short messages of 16 bits or less, a single packet can include the entire MSG. In this case, the REQ/ANS sequence becomes the MSG/ACK sequence, and in the case the REQ/ANS sequence is inherently absent.

d	c	k	β_M	θ;REQ/ANS Included	/ Excluded
6	1	0.20	0.62	0.25	0.30
12	1	0.11	0.67	0.35	0.34
12	3	0.11	0.67	0.425	0.34
12	5	0.11	0.67	0.45	0.34
240	1	0.006	0.73	0.58	0.40

Table 5-1 Spectrum efficiency θ with and without REQ/ANS sequences; np CSMA (a = 0.01) access control protocol. m = 1 channel/mobile. $D/L_M = 3$.

d,c \ S_m	0.184	0.37	0.518	0.815	0.857	1.0
12,1	0.368/	0.441/	0.467/	0.494/	0.597/	0.503/
	/0.12	/0.238	/0.328	/0.501	/0.524	/0.603
12,20	0.585/	0.593/	0.595/	0.598/	0.598/	0.598/

Table 5-2 Maximum θ values for various access protocols REQ/ANS sequence Included/Excluded

In an actual operating environment initial REQ transmissions would sometimes be repeated because of a mobile being located in a "radio hole" or because of a transciever being unintentionally turned off. The effective length of initial transmissions would thereby be increased in accordance with (3-2), and this increase would further favour inclusion of the REQ/ANS sequence.

5-3 Use of Dedicated Access Channels; Analysis Approaches

To prevent long waiting times for access on message channels and to allow for simple access protocols, dedicated access channels may be used. In such case one access channel serves more than one message channel. Another advantage of dedicated access channels is that only one channel needs to be monitored for receipt of REQ or ANS packets. In a multichannel system a REQ or ANS from a central node would include the identity of the MSG channel to be used.

We define α as follows:

$$\alpha = \frac{\text{Number of Access Channels}}{\text{Number of Message Channels}} \tag{5-7}$$

Various arrangements are possible for transmission of the ANS. We employ here a separate ANS channel for each REQ channel. As a result, the overall system spectrum efficiency

$$\theta = \psi/(2\alpha + 1) \tag{5-8}$$

where ψ is the spectrum efficiency on the message channels, which are assumed to carry the MSG and ACK.

$$S = \rho_Q(2\alpha + 1)/\alpha(1 - P_m(\rho)) \tag{5-9}$$

where P_m is given by (2-3) and

$$\rho = (2\alpha + 1)\rho_{AMK} \tag{5-10}$$

Note that ρ_Q and ρ_{AMK} apply to the system consisting of all channels, including MSG plus REQ plus ANS channels. The spectrum efficiency over all channels is again given by (4-26).

Use of (4-28) and (4-29) enables (5-9) and (5-10) to be written as follows:

$$S = kc^{-1}(2\alpha + 1)\rho_M/\alpha(1 - P_m(\rho)) \tag{5-11}$$

$$\rho = (2\alpha + 1)(1 + kc^{-1} + k)\rho_M \tag{5-12}$$

The definition of ρ in (5-10) and (5-12) is based on the assumption that as soon as the ANS transmission begins, the MSG channel identified for use is regarded as busy. (It is no different, for analysis purposes, than if the ANS channel remained unused, with the ANS sent on the designated MSG channel.)

The analysis in Chapter IV, if used, would employ S as given by (5-11) instead of (4-18), and $P_U = P_m(\rho)$. However, the busy periods defining P_U are not independent and random, and the calculated delay would again be optimistic.

We adopt, therefore, the following analysis algorithm, which provides a lower bound to the realizable delay D/L_M vs throughput θ.

1. Specify k, c, and the number m of message channels on which a mobile can operate.

2. Select the REQ access protocol, and note S_m, the maximum utilization allowed.

3. Select ρ and determine the smallest α value such that $S \leq S_m$ in

(5-11).

4. Determine θ and D/L_M as follows:

$$\theta = \beta_M \rho/(2\alpha + 1)(1 + kc^{-1} + k) \tag{5-13}$$

$$D/L_M = \frac{(1 + kc^{-1} + k)P_m(\rho)}{m(1 - \rho)} + (1 + 2kc^{-1} + k) \tag{5-14}$$

The above approach assumes, in effect, that all waiting is queueing delay for one of the m message channels. Because REQ packets are usually short relative to the MSG, only a small fraction of a MSG transmission time is needed to send a REQ packet, and for S not too large, the REQ access wait is also small.

V-4 Dedicated Access Channels: Results

Figs. 5-1 to 5-4 inclusive show D/L_M vs θ (lower bounds) when dedicated channels carry REQ and ANS packets. The cases $S_m = 0.815$ and 0.184 correspond, respectively, to np CSMA with $a = 0.01$ and pure ALOHA access controls. Also shown are the perfect scheduling curves from Chapter IV when access is via message channels.

Figs. 5-1 to 5-4 show a considerable penalty in terms of obtainable spectrum efficiency if pure ALOHA is used instead of np CSMA ($a = 0.01$). For example with (d,c) = (12,3) and m = 4, $\theta_{max} = 0.40$ and 0.28 for the two cases; the maximum value from Table 4-1 for perfect scheduling is $\theta = 0.57$. Similarly, for (d,c) = (6,1) and m = 4, $\theta_{max} = 0.21$ and 0.12, respectively; the maximum value here is $\theta = 0.39$.

Considerable loss in maximum spectrum efficiency can result from using dedicated access channels. In the cases cited above, the ratio of θ_{max} for dedicated access channels with CSMA to θ from the perfect scheduling results

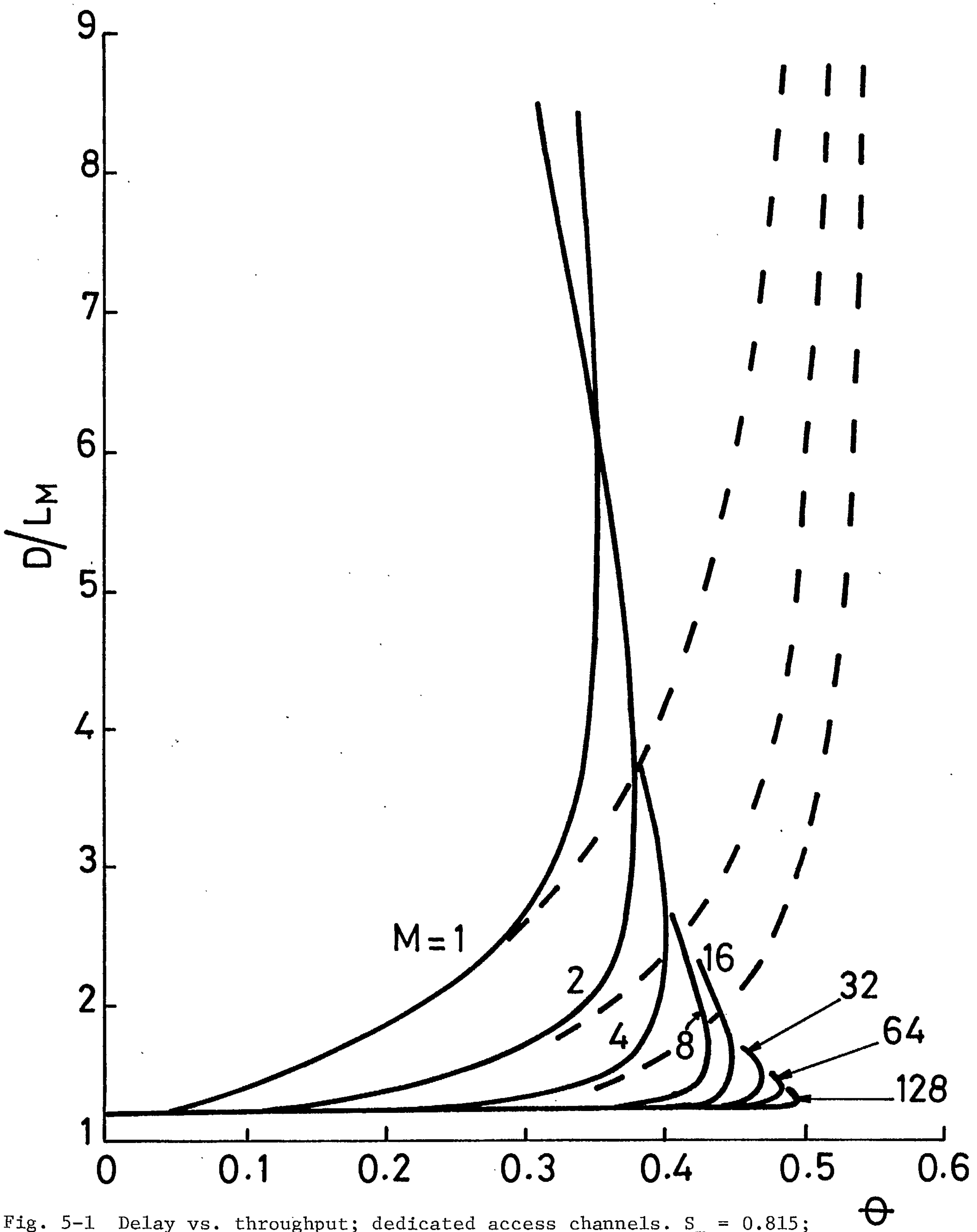


Fig. 5-1 Delay vs. throughput; dedicated access channels. $S_m = 0.815$; $(d,c) = (12,3)$

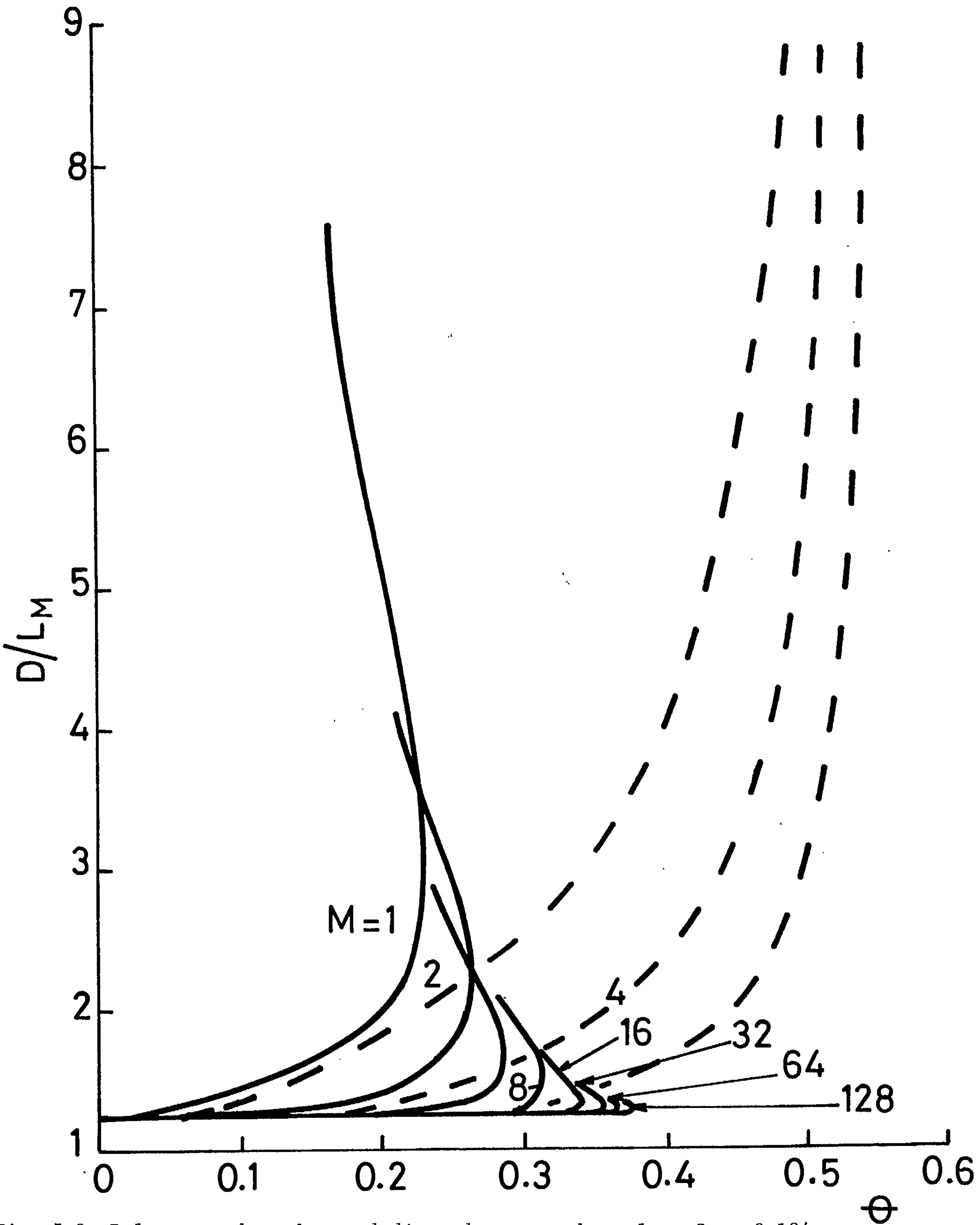


Fig. 5-2 Delay vs. throughput; dedicated access channels. $S_m = 0.184$; (d,c) = (12,3)

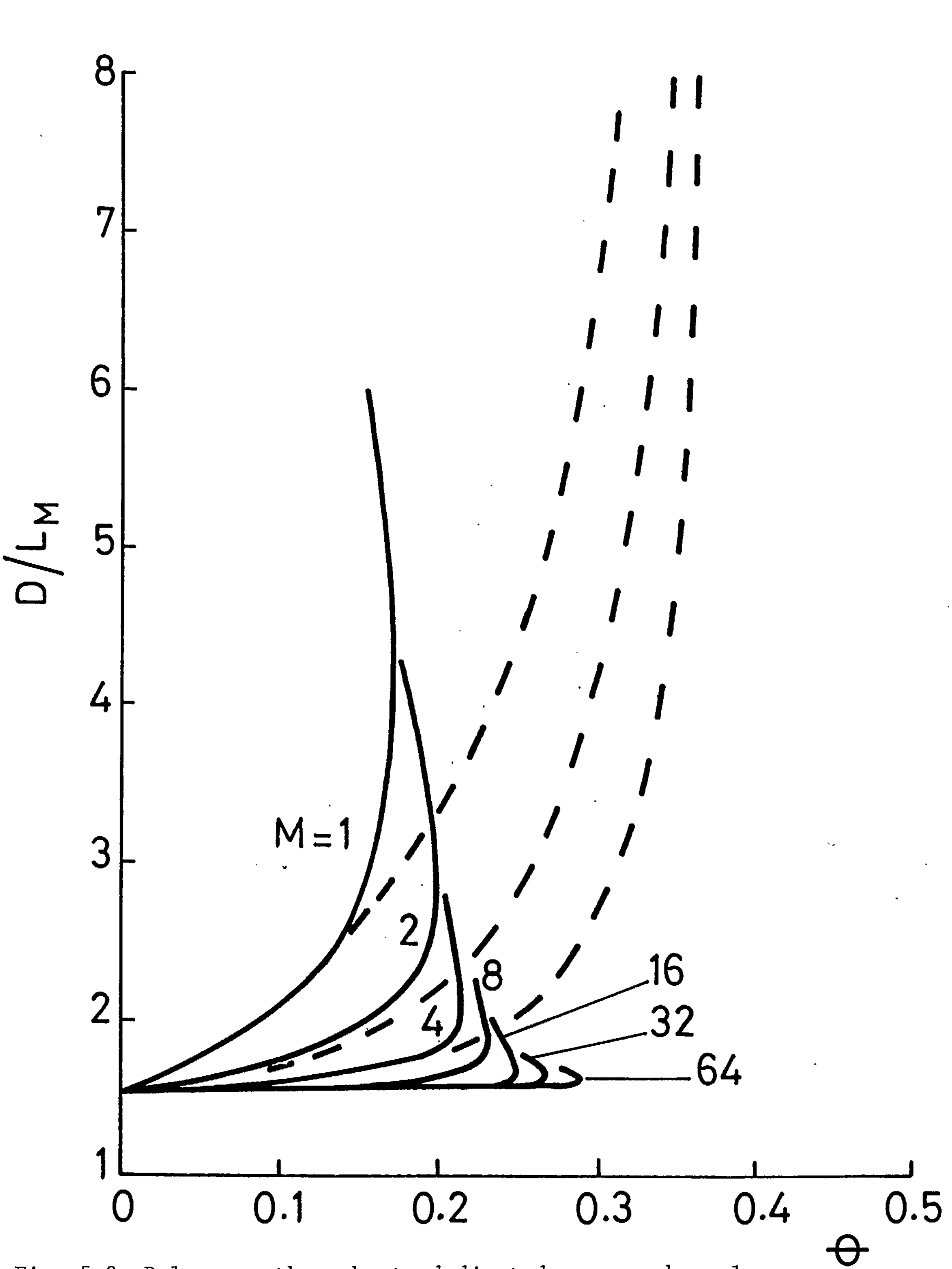


Fig. 5-3 Delay vs. throughput; dedicated access channels.
S_m = 0.815; (d,c) = (6,1)

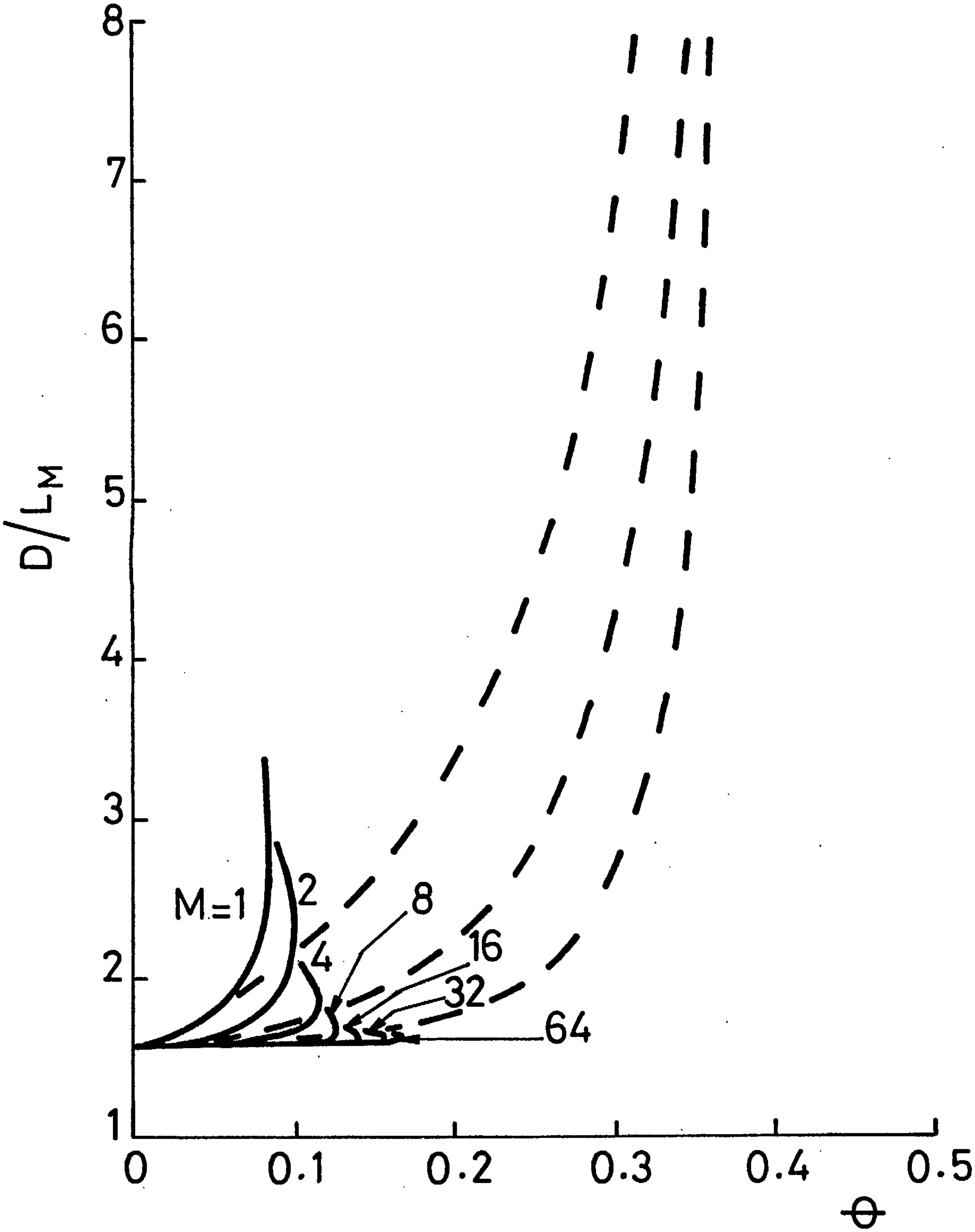


Fig. 5-4 Delay vs. throughput; dedicated access channels. $S_m = 0.184$; $(d,c) = (6,1)$

in Chapter IV are 0.40/0.57 = .70 for (d,c) = (12,3) and 0.21/0.39 = 0.54 for (d,c) = (6,1).

The comparisons above are somewhat unfair, since CSMA with a = 0.01 on dedicated access channels is being compared against CSMA with $a \to 0$ when access is via message channels. If we assume perfect scheduling on dedicated REQ channels, then $S_m = 1$. We do not include graphs for this case, because because they are in fact very little better than for the CSMA case shown in Figs. 5-1 and 5-3. For example, with m = 4 and perfect scheduling $\theta_{max} = 0.42$ rather than 0.40 for (d,c) = (12,3), and 0.23 rather than 0.21 for (d,c) = (6,1). Most of the degradation occurs because dedicated access channels force more of the MSG and ACK traffic onto the remaining message channels. On the other hand, for short messages use of dedicated access channels with pure ALOHA access control would give delay vs throughput performance better than what is obtained when ALOHA access is via message channels.

The use of dedicated access channels was investigated by others [C3] in another context, namely when long messages are sent in an Erlang B environment. In this case blocking probability rather than delay is of interest. These results [C3] indicate reasonably good performance using pure ALOHA on dedicated access channels. It was assumed in effect that ANS and ACK's arrive over separate channels with no delay or cost.

Table 5-3 shown the α values which maximize θ in our present work. These values are not precise (say within ±10%), and performance is not highly sensitive to α provided $S < S_m$. In an actual system, α would be fixed at some specific value. Our results shown that with α as that value which

(d,c)	(12,3)		(6,1)	
S_m	0.184	0.815	0.184	0.815
m				
1	0.32	0.12	0.48	0.22
2	0.22	0.16	0.48	0.22
4	0.27	0.11	0.55	0.24
8	0.18	0.05	0.47	0.14
16	0.26	0.06	0.42	0.15
32	0.20	0.05	0.60	0.17
64	0.18	0.04	0.60	0.15
128	0.16	0.03	0.60	0.15

Table 5-3 Optimum α values at maximum spectrum efficiency.

maximizes θ for given values of (d,c) and m, the delay for throughput θ below θ_{max} does not change very much. However as $\theta \to \theta_{max}$ the delay increases vertically without bound and does not fold back toward the D/L_M axis.

In an actual system, α would be fixed over all conditions, and some reasonable choice would have to be made. Our results indicate for all m values, $\alpha = 0.20$ and 0.60 for ALOHA and CSMA, respectively, would not significantly alter the results for (d,c) = (6,1), and that $\alpha = 0.10$ and 0.20, respectively, would not alter the results significantly for (d,c) = (12,3). The use of the higher α values appropriate to the (6,1) case would reduce the maximum obtainable θ values for the (12,3) case. For example with $\alpha = 0.60$, m = 4 and (d,c) = (12,3), the θ value for CSMA at $D/L_M = 2.5$ is 0.25, as compared with 0.40 in Fig. 5-1. Long messages and efficient REQ access protocols clearly favour small α values, whereas short messages and inefficient REQ access protocols favour larger α values. Clearly some compromise must be reached in choosing α, and no choice is best in all situations.

Referring again to Table 5-3, one sees that α tends to decrease as m increases. This behaviour one expects; ρ_M in (5-11) decreases as m increases, and a smaller value of α is thereby permitted. The same effect was observed in another study [C3].

We see from Figs. 5-1 to 5-4 that increasing m does result in a steady improvement in θ for a given delay value. However most of the improvement has occurred for $m \simeq 8$.

We note that the use of dedicated REQ and ANS channels does indeed provide considerable improvements in delay-throughput performance over that which can occur when a single message channel ($m = 1$ case) is used to carry the entire REQ/ANS/MSG/ACK sequence.

Finally, we note that all results assume that REQ, ANS and MSG/ACK channels are half-duplex. In practise, one might use one channel for inbound REQ's and ANS's and another for the packets outbound. In such case, the delay throughput performance would not be changed much from values calculated using the earlier assumption. However, if MSG/ACK channels were split into an inbound/outbound half-duplex pair then for a given delay the throughput would be reduced by the factor $0.5/(2\alpha + 1)$. For $\alpha \simeq 0.20$ the reduction factor is 0.36.

VI SUMMARY AND DISCUSSION

VI-1 Summary of Main Results

The significant results from the work described in this report are summarized below.

1. A signalling sequence including REQ, ANS, MSG and ACK has been proposed for control of access to land mobile radio channels. The sequence is quite general. Included is the possibility of omitting the REQ/ANS portion (which omission may be appropriate for short messages). Also possible within the analysis framework presented is the transmission of an ACK following a certain number of MSG bits, for example at the end of each line of text.

2. Various REQ contention protocols are described. Channel sensing protocols are potentially very efficient, provided the "hidden terminal" problem is overcome, and provided "a", the propagation delay relative to REQ packet length is small. Means to combat the hidden terminal problem are described together with their effects on spectrum efficiency.

3. Delay vs. throughput results for fixed length packet transmission protocols are available from the work of others. The pure ALOHA and CSMA curves were fitted with relatively simple functions. Delay vs. throughput equations were then derived when the REQ packets do not have sole access to a channel but instead share it with the ANS/MSG/ACK sequence. The performance results calculated were found to be overly optimistic, apparently because the usual assumption (used also by others) of independent busy periods. This rather negative result is actually very important, and indicates that the independent busy period

assumption fails to provide acceptable performance estimates in many cases of practical interest.

4. Another approach was developed to estimate delay vs. throughput performance, based on calculation of the mean busy period during channel contention times. This approach can be used to show that when CSMA is used with small normalized propagation delay, perfect scheduling of the REQ/ANS/MSG/ACK sequence provides a good (lower bound) estimate of performance except when contention loading approaches capacity of the CSMA protocol.

5. The perfect scheduling performance referred to in 4 above was calculated for a proposed CCIR data transmission format (see also [M4]). Important parameters considered include the average message length, frequency of ACK packets and number of channels m on which a mobile can operate. It was found that the value of m greatly affects the spectrum efficiency, particularly when delays are required to be two or three times the message length. In many cases much of the possible improvement was achieved with $m = 4$.

6. The effect of omitting the REQ/ANS sequence was considered. It was found that the maximum obtainable throughput with the proposed CCIR data format was slightly improved for messages up to one line of text in length with a good CSMA ($a \simeq 0.01$) access control protocol. However for pure ALOHA access, the REQ/ANS sequence seems useful, even for short messages. The reason is that the initial contention period should be kept as short as possible and this goal is achieved particularly for inefficient protocols by contending with short REQ packets, rather than with the longer message packets.

7. The use of separate channels dedicated solely for transmission of REQ and ANS packets was considered. Lower bounds on spectrum efficiency were obtained and compared with perfect scheduling results when access is via message channels. An important variable is the ratio α of access channel capacity/message channel capacity. The longer the message or the more efficient the access protocol, the smaller is α to optimize spectrum efficiency. In an actual operational situation some compromise would be needed in selecting α. One primary advantage of dedicated access channels is that only one channel need be monitored for REQ packets, by mobiles having multichannel capability. Mobiles' monitoring of two or more message channels simultaneously is not always feasible.

8. The effect of retransmissions of packets was determined, and was shown to increase the effective packet length by the factor $(1 - P_r)^{-1}$ where P_r is the retransmission probability. This effective increase in length provides further motivation for including the REQ/ANS sequence.

9. The actual spectrum efficiencies calculated were rather low, where spectrum efficiency θ is defined here as the fraction of the time that any channel carries information bits. For example, with operation on $m = 1$ input/output channel pair, transmission of three lines of text under the proposed CCIR digital format with a delay of two message lengths yielded $\theta = 0.11$ under perfect scheduling. Use of dedicated access channels, with $m = 8$ half-duplex message channels per mobile yielded a lower bound of $\theta = 0.32$ using pure ALOHA access, and $\theta = 0.43$ using np CSMA ($a = 0.01$). These results assume an ACK at the end of

each line and error-free transmissions, and would be lowered further by any retransmissions.

10. Although the work reported here was motivated by land mobile radio channel data transmission applications, it is useful in other operational environments where contention access for circuit switched transmissions of voice or data is of interest. Examples include two-way interactive co-axial cable environments and radio system environments with fixed (non-mobile) terminals.

One of the most effective ways to enhance spectrum efficiency is to use mobiles with multichannel capability, and to allow for multichannel operation. This approach avoids the situation where some channels sit idle while others are heavily congested. As well, it is important to use an efficient access control protocol when short messages are involved, to avoid excess access delay.

VI-2 Suggestions for Further Work

Some items on which further study could be beneficial are briefly described below.

1. Additional work could be done to determine delay vs. throughput performance for various access control protocols. The approach developed in Section IV-8 could be used to calculate performance for specific length distrubutions of REQ/ANS/MSG/ACK sequences for single-channel ($m = 1$) operation, or for exponential distributions for $m \geqslant 2$. Simulations to verify calculated results would be very useful. Simulations would allow for full duplex operation of input/output channel pairs. Actual spectrum efficiency would then lie between that

for full- and half-duplex operation over individual channels.

2. Data formats other than the one specifically considered in this report are of interest. It is not evident that all error-control bits should be used for error detection; in fact one study [F2] indicates that some forward error correction is useful in many situations. Also, it is not evident that the CCIR proposal of one bit in four for error control purposes is optimum. A related issue is the best overall policy for handling of ACK's [M2, M3, E1, T2, T3].

3. The effects of mixing voice, data and file message traffic is not fully understood. Whether all channels (other than access control channels) should carry all types of messages, or whether some channels should be dedicated for short data messages needs further consideration. This problem has received some consideration in another context [G1], and it was shown that on mixed voice and data channels that very long waits can occur for data transmissions. This problem may be obviated by priorizing message types or by transmitting data during voice silent periods [F1,W1].

4. An issue which is loosely related to this present work involves the choice of data rate R, which in turn involves choice of channel spacing Δ. There is a tradeoff between R and the amount of error detection/correction capability, as well as between R and Δ. Given Δ, R cannot exceed some limit or else adjacent-channel interference criteria will be violated. In practise, R would probably be compatible with digital rates on telephone lines, but there remains some flexibility and choice. Also, as R increases the preamble or synchronization sequence

at the beginning of any transmission may have to be increased.

5. Other operational system structures should be examined. Two alternatives to narrowband line-switched channel systems of the type implied by our work include spread spectrum transmission and packet switching. In general, packet switching is most appropriate for short interactive messages, and line switching for file traffic [G1,R1,H2,T1,W1]. However, it is not clear whether or not packet mobile radio systems would enhance spectrum efficiency. Some work has been done on performance analysis of spread spectrum systems [C4,H3,H4,G3]. However, definitive comparisons of spectrum efficiency of SSMA and conventional line switched systems with channel reuse are not yet finalized.

REFERENCES

A1 J.P. Agrawal, J.B. O'Neal, Jr., and J.S. Cooper, "The design of an ADPCM/TASI system for PCM speech compression", IEEE Trans. Commun., vol. COM-29, pp. 1393-1398, Sept. 1981.

A2 O.A. Avellaneda, J.F. Hayes and M.M. Nassehi, "A capacity allocation problem in voice-data networks", IEEE Trans. Commun., vol. COM-30, pp. 1767-1775, July 1982.

B1 S.H. Bakry and M.H. Ackroyd, "Teletraffic analysis for multicell mobile radio telephone systems", IEEE Trans. Commun., vol. COM-30, pp. 1905-1909, Aug. 1982.

C1 D.L. Cohn and J.L. Melsa, "The residual encoder - an improved ADPCM system for speech digitization", IEEE Trans. Commun., vol. COM-23, pp. 935-941, Sept. 1975.

C2 I. Chlamtac, N.R. Franta, and K.D. Levin, "BRAM - the broadcast recognizing access method", IEEE Trans. on Commun., vol. COM-27, pp. 1183-1189, Aug. 1979.

C3 G.L. Choudhury and S.S. Rappaport, "Cellular communication schemes using generalized fixed channel assignment and collision type request channels", IEEE Trans. on Veh. Technol., vol. VT-31, pp. 53-65, May 1982.

C4 G.R. Cooper and R.W. Nettleton, "A spread-spectrum technique for high-capacity mobile communications", IEEE Trans. Veh. Technol., vol VT-27, pp. 264-275, Nov. 1978.

D1 R.W. Donaldson, "Frequency Assignment for Land Mobile Radio System in the 900 MHz Band: Suitability of New Modulation Techniques over Land Mobile Channels", Rpt. to Dept. of Communications, Ottawa, Feb. 1982.

D2 P.F. Driessen, "Maximizing the throughput of a mobile radio data system", IEEE Veh. Tech. Conf., 1982.

E1 M.C. Easton, "Design choices for selective - repeat retransmission protocols", IEEE Trans. Commun., vol. COM-29, pp. 944-954, July 1981.

F1 M.J. Fischer, "Data performance in a system when data packets are transmitted during voice silent periods - single channel case", IEEE Trans. Commun., vol. COM-27, pp. 1371-1375, Sept. 1979.

F2 C. Fujiwara, M. Kasahara, K. Yamashita and T. Namekawa, "Evaluation of error control techniques in both independent-error and dependent-error channels", IEEE Trans. Commun., vol. COM-26, pp. 785-794, June 1978.

G1 D.P. Gaver and J.P. Lehoczky, "Channels that co-operatively service a data stream and voice messages", IEEE Trans. Commun., vol. COM-30, pp. 1153-1162, May 1982.

G2 J.G. Gruber, "Delay related issues in integrated voice and data networks", IEEE Trans. Commun., vol. COM-29, pp. 787-800, June 1981.

G3 D.J. Goodman and U. Timor, "Spread-spectrum mobile radio with variable-rate speech transmission", IEEE Trans. Commun., vol. COM-30, pp. 531-538, March 1982.

H1 T.S. Huang and A.B.S. Hussain, "Facsimilie coding by skipping white", IEEE Trans. Commun., vol. COM-23, pp. 1452-1460, Dec. 1975.

H2 E.A. Harrington, "Voice/data integration using circuit switched networks", IEEE Trans. Commun., vol. COM-28, pp. 781-793, June 1980.

H3 B.G. Haskell, "Computer simulation results on frequency hopped MFSK mobile radio-noisless case", IEEE Trans. Commun., vol. COM-29, pp. 125-132, Feb. 1981.

H4 P.S. Henry, "Spectrum effeciency of a frequency-hopped-DPSK spread spectrum mobile radio system", IEEE Trans. Veh. Technol., vol. VT-28, pp. 327-332, Nov. 1979.

J1 W.C. Jakes, Jr., Microwave Mobile Communications. N.Y.: Wiley, 1974.

K1 L. Kleinrock and M.O. Schell, "Packet switching in radio channels: new conflict-free multiple access schemes", IEEE Trans. on Commun., vol. COM-28, pp. 1015-1029, July 1980.

K2 L. Kleinrock, Queueing Systems, Vol. 2: Computer Applications. N.Y.: J. Wiley, 1976.

K3 L. Kleinrock and F. Tobagi, "Packet switching in radio channels: part I - Carrier sense multiple-access modes and their throughput-delay characteristics", IEEE Trans. Commun., vol. COM-23, pp. 1400-1416, Dec. 1975.

M1 J. Martin, Systems Analysis for Data Transmission. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

M2 P.J. Mabey, "Mobile radio data transmission for error-control", IEEE Trans. Veh. Technol., vol. VT-27, pp. 99-110, Aug. 1978.

M3 M.J. Miller and S. Lin, "The analysis of some selective-repeat ARQ schemes with finite receiver buffer", IEEE Trans. Commun., vol. COM-29, pp. 1307-1316, Sept. 1981.

M4 P.J. Mabey, "Predicting the range and throughput of mobile data systems", IEEE Veh. Tech. Conf. 1982.

N1 M. Nesenbergs, "A hybrid of Erlang B and Erlang C formulas and its applications", IEEE Trans. Commun., vol. COM-27, pp. 59-68, Jan. 1979.

P1 J.G. Proakis, Digital Communications. N.Y.: McGraw-Hill, 1983.

R1 M.J. Ross and O.A. Mowafi, "Performance analysis of hybrid switching concepts for integrated voice/data communications", vol. COM-30, pp. 1073-1088, May 1982.

S1 C.E. Shannon and W. Weaver, The Mathematical Theory of Communication. Urbana, Is.: U. of Illinois Press, 1962; also in Bell Syst. Tech. J., July 1949 and Scientific American, July 1949.

T1 F.A. Tobaji, "Multiaccess protocols in packet communication systems", IEEE Trans. Commun., vol. COM-28, pp. 468-488, Apr. 1980.

T2 F.A. Tobagi, "Random Access Techniques for Data Transmission over Packet Switched Radio Networks", Ph.D. dissertation, Comput. Sci. Dept., School of Engr. and Appl. Sc., Univ of Calif. Tech. Rept. UCLA-ENG 7499, Dec. 1974.

T3 R.F. Turney, "An improved stop-and-wait ARQ logic for data transmission in mobile radio systems", IEEE Trans. Commun., vol. COM-29, pp. 68-72, Jan. 1981.

T4 F.A. Tobagi, "Analysis of two-hop centralized packet radio network - part II: carrier sense multiple access", IEEE Trans. Commun., vol. COM-28, pp. 208-16, Feb. 1980.

V1 D.C. Van Voorhis, "An extended run-length encoder and decoder for compression of black/white images", IEEE Trans. Inform. Theory, vol. IT-22, pp. 190-199, Mar. 1976.

W1 C.J. Weinstein and E.M. Hofstetter, "The tradeoff between delay and TASI advantage in a packetized speech multiplexer", IEEE Trans. Commun., vol. COM-27, pp. 1716-1720, Nov. 1979.

APPENDIX I

CAPACITY OF INTERMITTENT CHANNELS

We consider a communication channel which transmits C bits/sec when it is available for use. The probability of it being unavailable at any given time is P_U. Then the actual capacity C_A is as follows:

$$C_A = C\ (1 - P_U) + 0 \cdot P_U$$

$$= C\ (1 - P_U) \qquad \text{(A-I-1)}$$

If the channel services messages of mean (Poisson) arrival rate λ sec.$^{-1}$ and mean length μ^{-1} sec. then the actual utilization ρ_A is:

$$\rho_A = \lambda / \mu C_A \qquad \text{(A-I-2)}$$

$$= \frac{\lambda}{\mu C(1 - P_U)} \qquad \text{(A-I-3)}$$

$$= \rho/(1 - P_U) \qquad \text{(A-I-4)}$$

Examination of (A-I-3) provides a slightly different interpretation of the effect of P_U on capacity. Eq. (A-I-3) can be written:

$$\rho_A = \lambda_A / \mu C \qquad \text{(A-I-5)}$$

where

$$\lambda_A = \lambda/(1 - P_U) \qquad \text{(A-I-6)}$$

The above result can be interpreted as follows: Messages arriving during those periods when the channel is unavailable for use are distributed to those periods when the channel is available, and the utilization ρ_A is given by (A-I-5). The message arrival rate λ_A at times when the channel is available is given by (A-I-6).

PART E

FORMATS FOR DATA TRANMISSION

1. Introduction

This part gives details of some of the data formats that are being used in the land mobile service.

2. The preferred binary format in the United Kingdom

The format is preferred for selective calling, status reporting, precoded messages, vehicle location, monitoring and supervisory systems, direct dialling, control in trunked systems and for mobile terminals (printers and displays).

2.1 *Format definition*

Preamble	Sync. word	Address code word	Optional data code words

Minimum length transmission: 96 bits

FIGURE 17 – *The format*

2.1.1 *Preamble*

16 or more bits "1010 . . . 10" ending with 0.

2.1.2 *Synchronization word*

Every message begins with:

Bit No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Bit Value	1	1	0	0	0	1	0	0	1	1	0	1	0	1	1	1

(Bit number 1 is transmitted first.)

FIGURE 18 – *Synchronization word*

2.1.3 *Code words*

All code words are of 64 bits (including 16 check bits). Short messages consist of a single address code word (which includes some data); longer messages have an address code word followed by data code words.

2.1.4 *Address code word*

Bit No.	1	2 ... 8	9 ... 48	49 ... 64
Number of bits	1	7	40	16
	«1»	user's identity	addresses + data	check bits

FIGURE 19 – *Address code word structure*

Bits:

- 1: always "1" to indicate an address word
- 2-8: user's identity
- 9-20: addressor identity (i.e. to) } optional
- 21-32: addressor identity (i.e. from) } optional
- 33-48: data } optional
- 49-64: check bits (see § 2.1.7)

2.1.5 *Data code word*

As many as are needed for the message.

Bit No.	1	2 … 48	49 … 64
Number of bits	1	47	16
	«0»	data	check bits

FIGURE 20 – *Data code word structure*

Bits:
- 1: always "0" to indicate data
- 2-48: data
- 49-64: check bits (see § 2.1.7)

2.1.6 *Character sets*

Binary coded decimal (BCD) coding can be used for the addressee and addressor identities. The character sets for messages are BCD for numeric-only messages, and the ISO 7-bit data code for alphanumeric messages. Characters are transmitted in reading order and least significant bit (b_1 in ISO code) first.

2.1.7 *Encoding and error checking*

The information bits 1-48 are the coefficients of a polynomial having terms from x^{62} down to x^{15}. This polynomial is divided modulo 2 by the generating polynomial $x^{15} + x^{14} + x^{13} + x^{11} + x4 + x^2 + 1$. The fifteen check bits, code word bits 49-63, correspond to the coefficients of the terms from x^{14} to x^0 in the remainder polynomial.

The final check bit of the code word (bit 63) is inverted to protect against misframing in the decoder.

One bit is appended to provide an even parity check of the whole 64 bit code word.

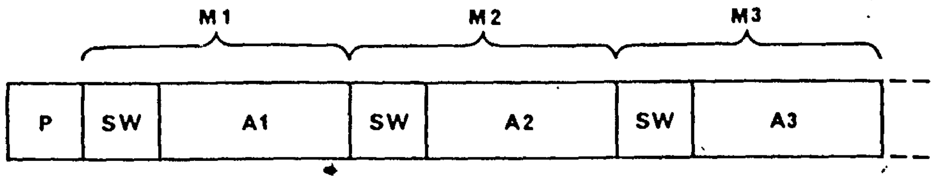
2.1.8 *Concatenated messages*

Figure 21 illustrates how several messages may be sent in one transmission.

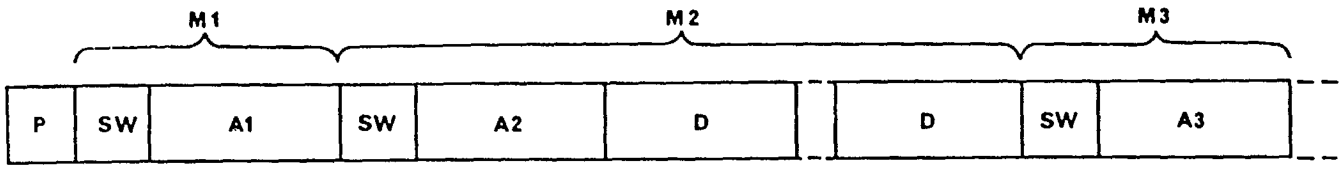
2.2 *Format design*

An error detecting code (which has a distance of 5 bits) was chosen rather than an error correcting code because it has an adequate performance and a simple, fast decoder.

The format does not rely on a data operated squelch circuit to prevent false messages by inhibiting decoding at low signal levels. A low false rate is obtained by coding alone.



a) Short messages



b) Mix of short and arbitrary length messages

FIGURE 21 – *Concatenated messages*

P: preamble
SW: synchronization word
A: address code word
D: data code word
M: message

2.2.1 *Synchronization word*

The use of a synchronization word is the most efficient method of identifying the start of each message, establishing code word framing and ensuring a low false call rate by inhibiting code word decoding at high bit error ratios, without the need for a signal squelch circuit.

The synchronization word must satisfy the following criteria:

- it must have a good success rate so that the messages which follow are not missed;
- the success rate should be about equal to the success rate in decoding address code words. A suitable synchronization word is then 16 bits;
- the synchronization word should have good correlation properties when it is preceded by preamble so that it is not decoded spuriously during the preamble, and so a preamble of 15 bit with an additional bit is used;
- finally, the synchronization word must provide a high security against false messages.

2.2.2 *Error detecting code*

To avoid false messages caused by misframing, the format uses a coset code [Peterson and Weldon, 1972]. Inverting the final bit in the code word is sufficient to ensure that valid code words do not appear for misframing by up to 14 bit positions.

Because a synchronization word may be found falsely at the end of a preceding codeword it is necessary to label data code words with a flag bit to distinguish them from address code words.

2.2.3 *Performance with a steady signal level*

The successful message probability P_s and the false message probability P_f have been calculated for a steady signal level in terms of the bit error ratio p, assuming independent errors.

The successful message probability is $P_s = (1 - p)^{80}$ which is 80% at a bit error ratio of $p = 2.8 \times 10^{-3}$.

The false message probability has been calculated as the probability that errors cause a transmitted address code word to be decoded as a different address code word.

$$P_f \simeq P(0,s) \cdot P(\geq d,n) \cdot 2^{-r} \tag{1}$$

$P(0,s)$ is the probability that the s bit synchronization word ($s = 16$) is received error free. For independent errors $P(0,s) = (1-p)^{16}$.

$P(\geq d,n) \cdot 2^{-r}$ is the conventional expression for the false rate of a cyclic code [Lucky, Salz and Weldon, 1968] where d = minimum distance of the code ($d = 5$), n = code word length ($n = 63$), and r = number of check bits in a code word ($r = 15$). $P(\geq d,n)$ is the probability that an n bit word contains d errors or more.

For independent errors

$$P(\geq d,n) = \sum_{i=d}^{n} \binom{n}{i} p^i (1-p)^{n-i}$$

Because the code guarantees detection of all odd numbers of errors, $P(\geq d,n)$ was evaluated for even numbers of errors only (i = even integer).

P_f is plotted in Fig. 22 which shows that the false message probability is less than 2×10^{-6} per transmitted message.

The values of P_f calculated apply to mobile to base transmissions where the base decoder accepts any valid code word. The false message rate will be lower when some code words remain unused. For base to mobile transmissions the false message probability will be lower because a mobile decoder will only accept messages bearing its own address.

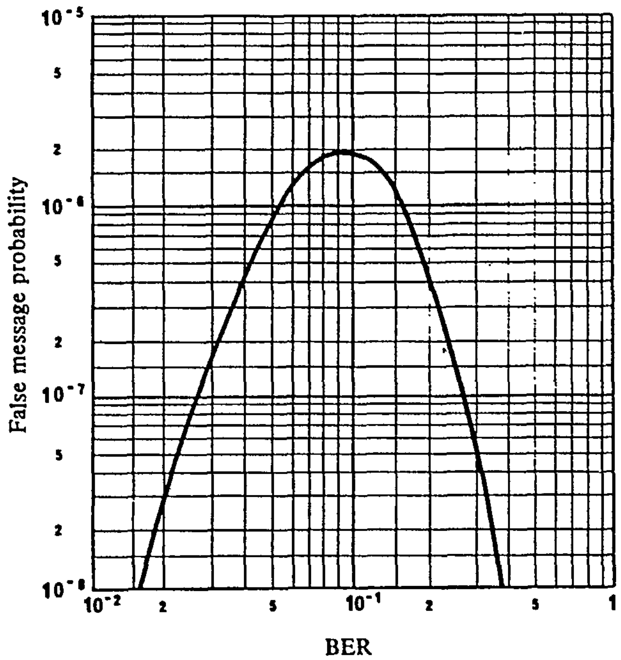


FIGURE 22 – *False message probability with a steady signal*

2.2.4 *Field measured performance*

Measurements made on a route which had Rayleigh fading and some shadowing (σ = 4 dB) with a vehicle speed of about 50 km/h, without ignition noise present, are given in Fig. 23.

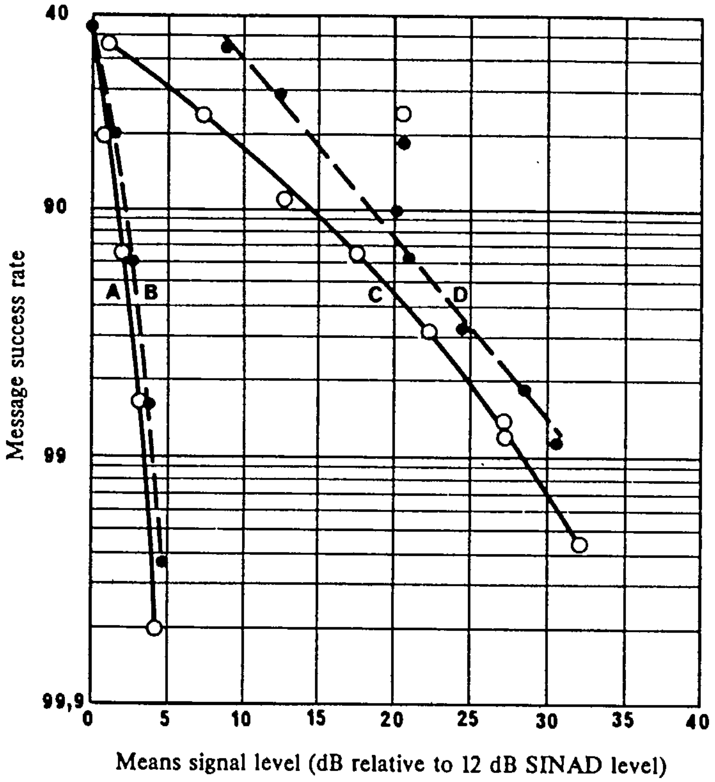


FIGURE 23 – *Field measured performance*

All curves: 1200 bit/s FFSK

A: stationary at 165 MHz FM
B: stationary at 465 MHz FM
C: moving at 50 km/h at 165 MHz FM
D: moving at 50 km/h at 465 MHz FM

REFERENCES

LUCKY, R. W., SALZ, J. and WELDON, E. J. Jr. [1968], *Principles of Data Communication*, McGraw-Hill, New York.

PETERSON, W. W. and WELDON, E. J. Jr. [1972], *Error Correcting Codes*, 2nd edition. MIT Press.