ELSEVIER

Computer Communications 19 (1996) 763–787

computer communications

Queueing analysis of a hybrid CSMA/CD and BTMA protocol with capture

T. Tsiligirides*

Informatics Laboratory, Agricultural University of Athens, 75 Iera Odos, Athens 11855, Greece

Received 8 February 1995; revised 18 July 1995

Abstract

A performance study of the access channel of a mobile, micro-cellular radio telephone network is presented. The access method used is a combination of the Carrier Sense Multiple Access with Collision Detection (CSMA/CD) and the Busy Tone Multiple Access (BTMA) protocols with modified access and back-off probabilities. Based on an imperfect channel and the likely low connectivity of the mobiles with the base station, two new, semi-Markov renewal models are constructed; the asynchronous time, infinite population model, and the synchronous time, finite population model. The latter uses the mini-slot property of the channel, and it is proven to have a matrix geometric representation. The analysis assumes that capture is due either to slow Rayleigh fading or to power level variations with the distances, and the corresponding capture probabilities are calculated analytically, or by using standard Monte Carlo simulation techniques. With the proposed hybrid protocol applied to both models when the capture is perfect, the performance is similar to the classic CSMA/CD, and thus it is better than the performance obtained by the ALOHA or CSMA algorithms with capture. Nevertheless, the power capture effect does not offer a similar improvement in the maximum achieved throughput, as in the cases of the ALOHA or CSMA contention based protocols. But, if capture due solely to slow fading is present, our hybrid model has considerably better performance than the corresponding CSMA ($\alpha = 0.01$) and BTMA protocols. However, with the proposed algorithm the channel shows a bi-stable behaviour, unless the offered traffic is extremely low. The analysis allowed us to draw some very useful conclusions for both models developed. Thus, in the case of the synchronous time, finite population model and with respect to the throughput–delay performance, the most important parameters, including the attempt rate, the signal-to-noise ratio and the length of the successful and unsuccessful messages, are considered. For suitable parameter values, and independently of the attempt rate, the channel appears to have an acceptable operational point. In the case of the infinite model, interest has been given in real systems with NCFSK and CPSK modulations, operating in a slow Rayleigh fading environment. Note that if only power capture is present, the same model yields an upper bound on the throughput. The analysis shows that the maximum achievable throughput is obtained for the same value of traffic rate, independently of the packet length. Higher values of the average SNR leads the system to approach the nonfading case.

Keywords: CSMA/CDS; BTMA protocols; Capture; Mobile network; Performance analysis; Stability

1. Introduction

In the cellular system the control channel is logically composed of two data streams called Forward Control Channel (FCC) and Reverse Control Channel (RCC), with direction towards the mobile terminals and the base station, respectively. Each stream is of 8 kbps/s, and both are used to maintain contact between the mobiles and the base in the cell when no subscriber connection is in progress, as well as for setting up calls. Thus, the mobiles access the system through the RCC, and an interesting problem, on which we elaborate in this research, is to study and give performance measures for this channel.

So far, a number of analytical methods have been developed studying the performance of the ALOHA and CSMA based protocols in the mobile environment. Almost all of these studies are asynchronous and take into account the capture effect. Since the technique requires all the users to be within range of the base station only, but not necessarily of each other, the conventional CSMA or CSMA/CD protocols are inappropriate for mobile radio. For example, an asynchronous analysis for the CSMA/CD protocol yielding

* Email: tsili@auadec.aua.ariadne-t.gr

PII S0140-3664(96)01107-3

very interesting computational and stability results can be found in [1] and [2], but it is not clear if similar results may be obtained in the cellular environment. However, BTMA and its variations [3–6] are also well known and in use in many packet radio networks (Deutche-Bundespost Telekom network in Germany; ARDIS and RAM Mobile data network in USA; Hutchinson Mobile data network in Japan). These are packet data only networks, and do not provide voice service.

In this work the method used to access the RCC is based on a protocol which is a hybrid between the well known Busy Tone Multiple Access (BTMA) [7] and the non-persistent Carrier Sense Multiple Access with Collision Detection (CSMA/CD) [8] protocols. The synchronous type of analysis, applied first in [9] for the CSMA protocol and extended here, is significant, because it leads the model to a matrix geometric representation form that includes the features of the p-persistent CSMA/CD protocol, the busy-tone idea and the capture effect. Thus, apart from our study [9] that appeared recently, there is no analysis (known at least to the author) which extends the mini-slot idea (synchronous time structure) proposed by Tobagi and Hunt [8] in the mobile radio environment. The proposed analysis includes two schemes, one with a moderate and the other with a large but finite population of mobiles, providing accurate and computationally tractable performance measures. However, to clarify our results the asynchronous time, infinite population model is also considered. This protocol was under investigation by the EEC RACE Mobile Telecommunications project R1043/1991, and it is closely connected to the Total Access Communications System (TACS) [10].

In the mobile packet radio environment the transmission is unsuccessful due only to transmission errors resulting from packet collisions, noise, fading and probably shadowing. If the transmitted packets experience not only noise but also fading, the receiver may fail to detect the faded packet, even though there is no collision [11]. However, the fact that independently faded packets, with a different signal strength, arrive at the receiver makes the capture effect possible (the strongest signal may capture the receiver). As a result, the probability of mutual destruction of packets is reduced [12], and the throughput is increased [13–15]. Note that the primary contribution to capture in the ALOHA channel is the difference from the base station, and that capture due to slow or fast Rayleigh fading level fluctuations plays a lesser role in terms of increasing the system throughput; nevertheless, these fluctuations are sufficient to stabilize the channel.

The present model assumes that capture is due to slow Rayleigh fading and to power level variations with the distance, i.e. errors due to other sources of noise and interference are not considered. If the system is not using power level control, both effects combine to produce capture. However, if the system uses some form of power level control, namely all users are received with the same average power, capture still occurs due solely to the Rayleigh fading process. That is, if two packets collide and one of them travels over a severely faded path, then the unfaded packet may survive the collision. However, in the hybrid protocol we are studying our results are different from those found in the ALOHA channel. We will see that our protocol has a similar performance to the classic CSMA/CD protocol when capture is perfect. Nevertheless, when capture due to slow fading or power level variations is present, there is a degradation from CSMA/CD to the new hybrid protocol. In particular, if capture due solely to slow fading is present, our hybrid protocol has a considerably better performance than the corresponding hybrid CSMA ($\alpha = 0.01$) and BTMA protocols. But if only power capture is present, our result is different. Thus, based on the fact that protocols which control contention better than ALOHA (e.g. CSMA) do not have as dramatic an improvement as ALOHA [9], we will see that protocols which control contention better than CSMA (e.g. CSMA/CD) do not have a similar improvement as CSMA.

The paper is organized as follows. In the next section we describe and analyse the asynchronous time, infinite population renewal model, which corresponds to the general framework of the model discussed in this section. We present some approximate performance measures and examine the stability behaviour of the channel. In the following three sections we describe and analyse our hybrid model as a discrete-time, finite population renewal model. The analysis makes use of the single property of the channel, and it includes both the scheme with the moderate population of mobiles, which is based on a binomial packet generation process, and the scheme with the large but finite population of mobiles, which is based on a Poisson packet arrival process. Specifically, in the third section, the mathematical analysis which has already been introduced [9] and which is extended here is presented; in Section 4 the performance measures are derived explicitly, while in Section 5, the analysis is modified to include the variable length of the successful and unsuccessful messages. In Section 6 the numerical approach is discussed, and the main computational difficulties are explored. Further, in this section the numerical results are presented. Finally, in the last section the conclusions drawn from the comparison of both models are presented.

2. Infinite population, asynchronous time renewal model

2.1. Model description

We consider an infinite population of mobiles which

collectively form an independent Poisson source with an aggregate mean message generation rate of S messages/s. This assumption seems quite reasonable, because a mobile user that senses the channel and finds it busy or detects the unsuccessful transmission of his message delays the retransmission by some random time chosen from the uniform distribution. Thus, if the mean rescheduling delay is sufficiently large compared to the transmission times of the messages, the inter-arrival times of the point process defined by the times the mobiles sense the channel (read the busy/idle bit) are independent and approximately exponentially distributed.

In the model we are studying, an access message begins with a 48-bit precursor sequence used for synchronization and identification of the base station to which the message is addressed. This message ends with the information part, which contains the message itself and consists of one to five words, where each word consists of 240 bits. The base station must have received the entire precursor before it can determine whether the message is destined to that particular cell, and not to an adjacent cell sharing the same set of frequencies. To provide the busy-tone function, after every 10 bits in the FCC an extra bit is inserted which indicates the state of the RCC. If the base is currently receiving a message on the RCC, this bit is set to 1 (busy), otherwise it is set to 0 (idle). A mobile monitoring the FCC can therefore determine with a granularity of 11 bits, or 1.375 ms, whenever some mobile is already transmitting.

The overhead load offered to the FCC is independent of the RCC loading. In addition, the FCC overhead is basically small and includes the busy tone overhead. Thus, there is likely to be considerable spare capacity of the FCC unless significant overhead or excess paging is adopted, elements which are not considered here. The presence of persistent interference or transmission degradation, necessitating multiple copies of a message, could also reverse this conclusion. On teletraffic grounds alone, at specified performance levels a sensible system organization is unlikely to need more than a single RCC and FCC per cell. Since our study concerns the performance of the RCC, the FCC overhead is not discussed further.

A mobile intending to access the RCC for the first time listens to the FCC until it has read a particular identification message. This message is transmitted by the base station whenever the FCC becomes idle, and contains some information parameters of the cell. After receiving this message the mobile waits a uniformly distributed random time to avoid a possible collision with other mobiles also waiting for the above message. This delay is called the *initial delay*. The above randomization periods are sufficiently long compared to an average total RCC message length of some 120 ms (or 960 bits = 4 words × 48 bits/word × 5 transmissions). Then, the mobile senses the FCC again to check the busy/idle bit setting, and to determine whether the RCC is free.

If the busy/idle bit is sensed idle, the mobile initiates the transmission of its message by sending a precursor. Clearly, if the precursor wholly or partially coincides with the transmission of another message and the capture effects are ignored, a collision is assumed to have occurred. In particular, during the transmission the mobile observes the busy/idle bit and expects the indication to change from 0 (idle) to 1 (busy) within 7–13 ms (this time interval corresponds to the range of 56–104 bits which have already been sent by the mobile). If this change does not take place as expected, or if it does but it is out of this range, the mobile assumes that the message has collided with another and the transmission is aborted (the mobile instantly turns off its transmitter). This limits the total duration of a colliding message to a maximum of some 120 bits or 15 ms (13 ms plus 2 ms needed by the transmitter to power-up). Therefore, a realistic assumption is that the length of an unsuccessful message is about 15% of the duration of a successful one. Note that the most probable reasons for an unsuccessful transmission are the erroneous reading of the idle/busy bit or interference with mobiles in adjacent cells using the same set of channel frequencies. In contrast, if the precursor is received correctly, the transmission is assumed to be successful due to the large amount of redundant information in the information part of the message. In this case, the busy/idle bit is set 'busy' and it is reset to 'idle' some time after the reception of the last bit of the access message. Lastly, if the busy/idle bit is sensed busy or an unsuccessful transmission is aborted, the mobile waits for a random time which is uniformly distributed. Then, in both cases, the mobile resenses the busy/idle setting again and repeats the algorithm.

Let us now denote by C bits/s the bit rate of the control channel in both directions, and by β and γ the transmission times (in seconds) of a precursor and an unsuccessful transmission, respectively. Under the steady-state conditions, the throughput rate equals the message generation rate S. Thus, $\bar{U}_s = S(\Theta_{\text{mean}} + \xi)$, where $\bar{U}_s$ is the average channel utilization. In the above, Θ_{mean} is the average message length with transmission time $\Theta_{\text{mean}} = \beta + 240w/C$ seconds. Here, w denotes the average number of words in an arbitrary message, and ξ indicates the additional seconds that the busy/idle bit remains busy after the complete reception of a successfully received message. Note that the busy/idle bit is assumed to be available continuously, and that in calculating the probability of a message collision the propagation delay is neglected.

In this section we review some of the most important results drawn from the non-synchronous analysis of the infinite population model. In particular, we will derive some very useful approximate performance measures, such as the throughput, the average channel utilization, the average time the channel is busy due to collisions, the average time the channel is busy due to an

undetected precursor, the average time the channel is idle, and the mean delay. The aim is to give performance measures and study the stability behaviour of the channel.

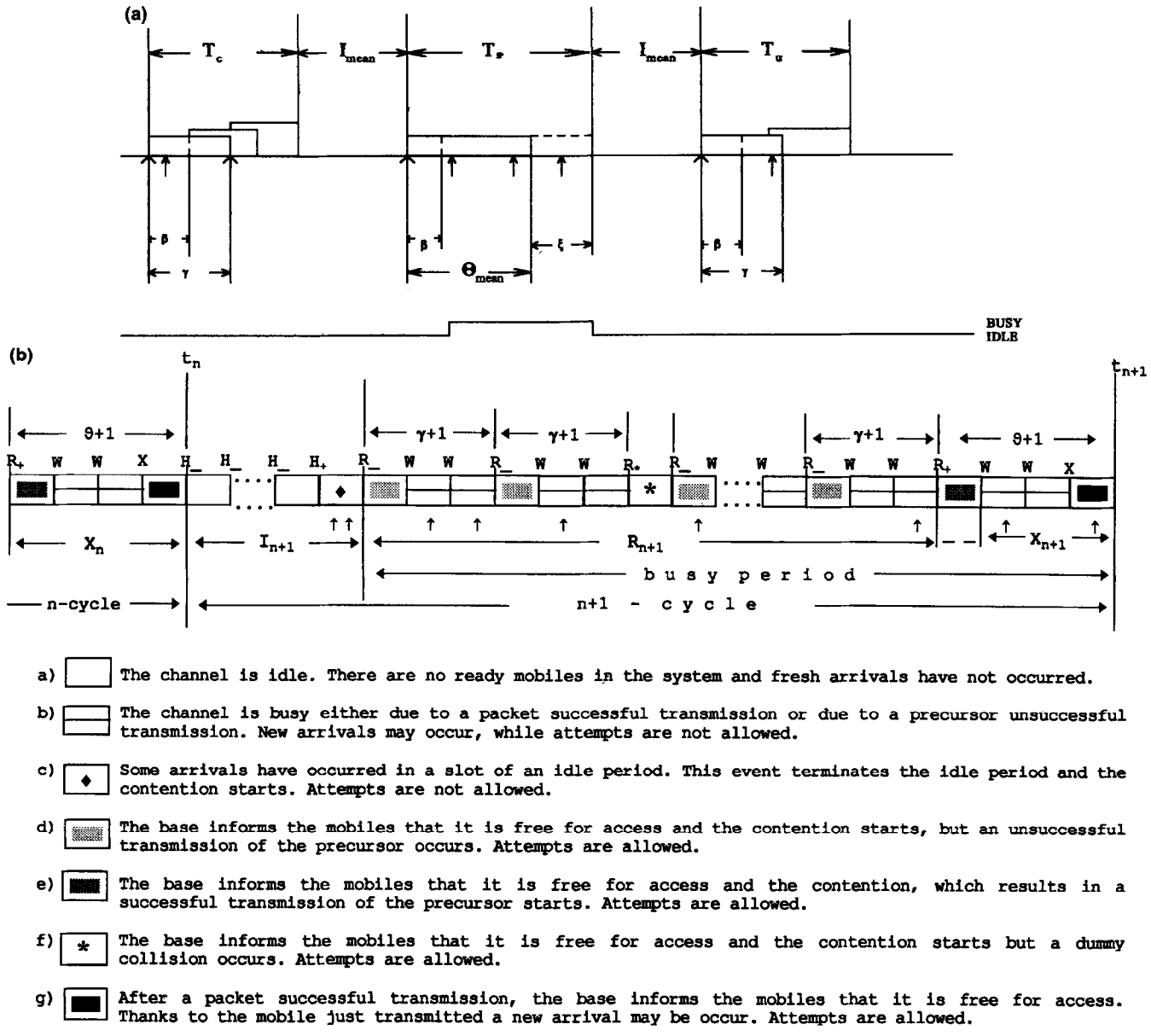
2.2. Model analysis

We start the analysis by identifying that the unslotted time axis has four different periods: these are the idle period and the three busy periods, namely successful, unsuccessful due to a collision, and unsuccessful due to an undetected precursor (see Fig. 1(a)). If we define by P_s, P_c and P_u the probabilities that an arbitrary period is successful, unsuccessful due to a collision, and unsuccessful due to an undetected precursor, respectively, then:

$$P_s = f\exp(-\beta G) \quad P_c = 1 - \exp(-\beta G) \quad \text{and}$$

$$P_u = (1-f)\exp(-\beta G)$$

where G is the sense point rate and f is the capture probability. This is assuming that no power variations with the distance capture occur due solely to Rayleigh fading process. Here, the capture probability is defined as the probability for the base station to successfully receive an un-collided precursor. In Appendix A we examine the effects of the average SNR (Signal-to-Noise Ratio), $\bar{\rho}$ say, on the capture probability f under the slow fading environment. Both the NCFSK (Non-Coherent Frequency Shift Keying) and CPSK (Coherent Phase Shift



a) The channel is idle. There are no ready mobiles in the system and fresh arrivals have not occurred.

b) The channel is busy either due to a packet successful transmission or due to a precursor unsuccessful transmission. New arrivals may occur, while attempts are not allowed.

c) Some arrivals have occurred in a slot of an idle period. This event terminates the idle period and the contention starts. Attempts are not allowed.

d) The base informs the mobiles that it is free for access and the contention starts, but an unsuccessful transmission of the precursor occurs. Attempts are allowed.

e) The base informs the mobiles that it is free for access and the contention, which results in a successful transmission of the precursor starts. Attempts are allowed.

f) The base informs the mobiles that it is free for access and the contention starts but a dummy collision occurs. Attempts are allowed.

g) After a packet successful transmission, the base informs the mobiles that it is free for access. Thanks to the mobile just transmitted a new arrival may be occur. Attempts are allowed.

Fig. 1. (a) The asynchronous time model. (b) The synchronous time model.

Keying) modulations are considered, using as a packet length the 48-bit precursor of our model. The capture probability is calculated as a function of $\bar{\rho}$, giving $f(\bar{\rho}) = \exp(-7.5105/\bar{\rho})$ for the NCFSK modulation and $f(\bar{\rho}) = \exp(-2.637/\bar{\rho})$ for the CPSK modulation, respectively. However, the throughput equation in a fading channel for the non-persistent CSMA and BTMA protocol is given by:

$$S = \{G\exp(-\alpha G)/[G(1+2\alpha) + \exp(-\alpha G)]\}f(\bar{\rho}),$$

where α is the propagation delay. Thus, for a 1000-bit packet length and a 100-bit acknowledgment we obtain (see Appendix A, Figs. 29 and 30):

$$S = G\exp(-\alpha G - 22.5621/\bar{\rho})/[G(1+2\alpha) + \exp(-\alpha G)]$$

using NCFSK modulation, and

$$S = G\exp(-\alpha G - 8.551/\bar{\rho})/[G(1+2\alpha) + \exp(-\alpha G)]$$

using CPSK modulation.

It is interesting to note that the same model may be applied if the power capture is present. Power capture is determined by the probability f_i that one out of i colliding packets is successfully received. By definition $f_0 = 0$, and for a well designed system f_1 is close to one. Since the difficulty in obtaining, analytically, values of the f_i's grows with i, simulation procedures may be adopted. Therefore, capture occurs when $f_i > 0$, $i \geq 2$ and, as will be seen, $f_i \geq f_{i+1}$, $i \geq 1$. In addition, since only the first few f_i's are the most important ones, the realistic assumption, $f_i = f_k$, $i \geq k$ for some k, yields an upper bound on the throughput [16,17]. Here, without loss of generality, we assume that $f_i = f$, fixed for all $i = 2, 3, \ldots$ with $f \in (0, 1)$.

Let us denote by I_{mean} and T_{mean} the average length of the idle, and busy periods respectively. Then:

$$I_{\text{mean}} = 1/G \quad \text{and} \quad T_{\text{mean}} = P_s T_s + P_c T_c + P_u T_u.$$

In the above, T_s, T_c and T_u denote the average lengths of the three types of busy periods, namely the successful and the two types of unsuccessful transmission periods, respectively. These average lengths are calculated as:

$$T_s = \Theta_{\text{mean}} + \xi.$$

$$\begin{aligned} T_c &= E\{Y|Y \leq \beta\} + \{\exp(-\gamma G)/ \\ &\quad [1-\exp(-\gamma G)]\}^{-1} E\{Y|Y \leq \gamma\} + \gamma \\ &= 1/G\exp(-\gamma G) - \beta\exp(-\beta G)/[1-\exp(-\beta G)]. \end{aligned}$$

$$\begin{aligned} T_u &= \beta + (\gamma-\beta)\exp[-(\gamma-\beta)G] \\ &\quad + \{1-\exp[-(\gamma-\beta)G]\}\{E\{Y|Y \leq \gamma-\beta\} \\ &\quad + \{\exp(-\gamma G)/[1-\exp(-\gamma G)]\}^{-1} E\{Y|Y \leq \gamma\} + \gamma\} \\ &= \beta + \{1-\exp[-(\gamma-\beta)G]\}/G\exp(-\gamma G). \end{aligned}$$

In addition, if Y is an exponentially distributed random variable denoting the time interval between two successive sense points with mean value $1/G$, then:

$$E\{Y|Y \leq \rho\} = [1-(1+\rho G)\exp(-\rho G)]/[1-\exp(-\rho G)]G.$$

Clearly, as the load grows, i.e. $G \to \infty$, the average lengths of the unsuccessful transmission periods T_c and T_u tend to infinity.

2.3. Performance measures

Based on the above analysis, we may now introduce some useful measures concerning the performance of this model. Using renewal theory arguments, it can be shown that:

$$\bar{U}_s = S(\Theta_{\text{mean}} + \xi) = P_s T_s/(I_{\text{mean}} + T_{\text{mean}})$$

where $\bar{U}_s$ is the average channel utilization. This also presents the average time of a cycle the channel is busy due to a successful transmission. Thus, the throughput of the system is given by:

$$S = P_s/(I_{\text{mean}} + T_{\text{mean}}) = P_s/(1/G + P_s T_s + P_c T_c + P_u T_u).$$

It is now straightforward to define the mean busy time $\bar{U}_c$ as:

$$\bar{U}_c = P_c T_c/(I_{\text{mean}} + T_{\text{mean}}),$$

namely, the average time of a cycle the channel is busy due to an unsuccessful transmission (collision). In addition, we may define the mean busy time $\bar{U}_u$ as:

$$\bar{U}_u = P_u T_u/(I_{\text{mean}} + T_{\text{mean}}),$$

namely the average time of a cycle the channel is busy due to an undetected precursor. Note that the average time of a cycle the channel is busy is:

$$\bar{U} = \bar{U}_s + \bar{U}_c + \bar{U}_u = T_{\text{mean}}/(I_{\text{mean}} + T_{\text{mean}}).$$

Therefore, the average time of a cycle the channel is idle $\bar{I}$ is given by:

$$\bar{I} = I_{\text{mean}}/(I_{\text{mean}} + T_{\text{mean}}).$$

The average message delay $\bar{D}$ is defined as the time from when a mobile terminal first senses the channel until a successful transmission of its message is initiated. Thus:

$$\begin{aligned} \bar{D} &= (GP_b/S)(t/2) + [G(1-P_b)/S - 1](\gamma + t/2) \\ &= (G/S-1)(t/2) + [(1-P_b)G/S - 1]\gamma \end{aligned}$$

where $P_b = (\Theta_{\text{mean}} + \xi - \beta)P_s/(I_{\text{mean}} + T_{\text{mean}})$ is the probability the busy/idle bit is busy. Note that $G/S - 1$ presents the average number of times the mobile senses the busy/idle bit, excluding the last one. In addition, if an unsuccessful transmission is aborted, or if the busy/idle bit is sensed busy, the mobile waits a uniformly distributed random time $U(0, t)$ seconds before it repeats the

algorithm by resensing the busy/idle bit. Thus, each time the channel is sensed busy, the message suffers an average delay of $t/2$ seconds, while each unsuccessful transmission will add $\gamma + t/2$ seconds to the average delay.

Finally, one additional remark should be made. The above model assumes that the initial delay is zero, that is, the delay is measured from the time instant the mobile first senses the channel. As stated before, the initial delay is long compared with the total average message length, and thus we may regard the message starts as effectively random. However, to include the initial delay in the model one has to take into account the time from when the mobile becomes ready for transmission until it has read the identification message, plus the rescheduling delay, which is assumed uniformly distributed in the interval $(0, \tau)$ seconds. Therefore, the average initial delay is given as $D_{id} = Z + \tau_f + \tau/2$, where τ_f is the time required to read the identification message and Z is the mean residual life of the time between two such successive messages. Although the above approach may be typical [18], the analytical calculation of Z does not seem straightforward, and further work may be required. However, some indicative results may be obtained by means of simulation, and in case some plausible distribution is used.

3. Finite population, synchronous time renewal model

3.1. Model description

We now consider the synchronous time structure renewal model with a finite number of mobiles, say N, requiring access on the RCC. The channel is slotted with a slot duration equal to the transmission time, say T seconds, of the 48-bit precursor (6 ms). All the transmissions (successful or unsuccessful) are normalized with respect to the slot size and are assumed integers. A mobile generates packets with an average packet length equal to $\theta = 1 + (240/48)w = 1 + 5w$ slots, where w again denotes the number of words in an arbitrary packet. This assumption will be relaxed later, and the model will be analysed with a general distribution of the packet length.

A mobile is capable of storing at most one packet in its buffer. If it has no packet in its buffer it is said to be *free*, otherwise, if it has a packet, the mobile is said to be *ready*. A ready mobile attempts to transmit at most one packet in a time slot. Note that a mobile generates (or does not generate) a packet in a time slot independently of the other mobiles, with a probability σ (or $1 - \sigma$). Packets generated from ready mobiles are assumed lost. Thus, packets arrive only from free users at the kT instances; $k = 1, 2, \ldots$. The number of arriving packets is best approximated by the binomial distribution $b(N, j, \sigma) = (N/j)\sigma^j(1-\sigma)^{N-j}$; $j = 1, 2, \ldots, N$. However, if N is large and σ is small, $N\sigma$ takes moderate values (<10). Thus, the number of packet arrivals is best approximated by the Poisson distribution $Ps(N\sigma; j) = \exp(-N\sigma)(N\sigma)^j/j!$; $j = 0, 1, 2, \ldots$.

We assume that the system does not use power level control, and therefore there are power variations with the distance. To consider the power capture effect we define the probabilities f_i that exactly one out of i attempting transmission packets will be successful in a time slot; $i = 0, 1, 2, \ldots, N$ with $f_0 = 0$, $f_1 = 1$ and $f_i \in (0, 1)$; $i = 2, 3, \ldots, N$. The values of f_i are calculated by neglecting shadow fading and Rayleigh fading, and assuming a constant signal during a packet (48-bit precursor) transmission. The introduction of the power capture has the effect that if a collision results in a successful transmission, the successful transmission period of θ time slots starts. Alternatively, if a collision results in no successful transmission, the unsuccessful transmission period of γ time slots starts. γ is called the channel recovery time. In the sequel, γ will be considered fixed, but later (see Section 5) it will be relaxed.

As has been pointed out in the previous section, the idle/busy bit value is assumed to be read by the mobiles continuously, although its status changes only at discrete points in time. By defining $x : (0 < x < 1)$ as the portion of time (in slots) required from the base station to inform the most distant mobile for its status, it appears the time until the channel is sensed idle again by all the mobiles is $(\theta + x)T$ and $(\gamma + x)T$ seconds, respectively. Thus, if a collision with a successful transmission occurs, all the ready mobiles are synchronized to compete again in the next $\theta + 1$ time slot. Alternatively, if a collision with no successful transmission occurs, all the ready mobiles are synchronized to compete again in the next $\gamma + 1$ time slot. Finally, during the period in which the channel is busy, no attempts are made by the ready mobiles, although new arrivals may occur.

A ready mobile receives the information about the status of the RCC from the separate FCC by reading the busy/idle bit and then uses the following procedure: if the status is busy, all the ready mobiles postpone their attempt until the current transmission is finished and the channel is sensed idle again. If the status is idle the contention period starts. During this period, all the ready mobiles attempt to transmit their precursors with probability δ. Inevitably, during this period collisions resulting in no successful transmission may occur, but thanks to the capture effect, one of the colliding precursors may be successfully transmitted. The event of the successful reception of the precursor and detection of the change of the idle/busy bit status (see Fig. 1(b)) terminates the contention period and the transmission period starts immediately.

The above algorithm differs from the one described in the previous section in the following two points. In case the busy/idle bit shows the channel is idle, a ready mobile

attempts to transmit its precursor with the probability $\delta \leq 1.0$, rather than $\delta = 1.0$. On the other hand, in case the indication is busy, a ready mobile does not resense the FCC with some random delay. Instead, the mobile persists in resensing the channel in each time slot until the status of the busy/idle bit changes to idle. Then, the mobile transmits its precursor with probability δ.

Clearly, the proposed algorithm is a combination of the classic CSMA/CD and BTMA, which takes into account the capture effect in mobile radio channels. The modifications made include selection of the access probability and the change of the back off scheme for the ready mobiles. The advantage of the slotted model is, however, that it can be extended to accommodate a finite number of mobiles and a finite average rescheduling delay. Furthermore, the proposed analysis leads to a geometric matrix representation, and thus more efficient and computationally tractable measures compared to the original schemes may be derived. In the case of a large but finite number of mobiles, the computational approach may use the theory, already under development, on parallel algorithms. Lastly, in the case of a small population of mobiles, standard serial algorithms may be adopted. We discuss more about these issues in Section 6, where numerical results are presented.

3.2. Model analysis

The approach undertaken in the analysis of this model is similar to that developed in [9], but the derivations and mathematical manipulations are different. We start by assuming that the slotted time axis is divided into cycles. A cycle is defined as the elapsed time (in slots), between the time instants in which two consecutive successfully transmitted packets end their transmission (see Fig. 1(b)). These time instants are the regenerated points where the state of the system, expressed by the number of ready users left behind, forms a semi-Markov process.

Each cycle consists of the idle, contention and transmission periods. To derive the transition matrix of the system, one has to find the transition matrices of the three corresponding periods mentioned above. This can be achieved by taking advantage of the single property of the channel, and by computing the product of several single slot transition matrices. The state of the system in each single slot is described by the number of ready mobiles. In the sequel, we define the above three periods and derive their transition matrices.

The idle period of a cycle is defined as the time interval (in slots) between the time instant when the successfully transmitted packet of the previous cycle ends its transmission leaving behind no ready mobiles, and the time instant when one or more new packets arrive. If, by the end of the successfully transmitted packet of the previous cycle, one or more ready users are left behind, the idle period is zero and the system immediately enters the contention period.

We assume that the length of the idle period of a cycle is k single slots. Since no packet arrival occurs in the first $k-1$ single slots, the total number of ready mobiles is zero. However, the termination of the idle period indicates that some fresh packet arrivals occur in the last single slot. Let us denote by $\mathbf{H}_- = (h_{ij}^-)$ the one step transition matrix that describes the state of the system in each of the first $k-1$ single slots specified above, and by $\mathbf{H}_+ = (h_{ij}^+)$ the one step transition matrix that describes the state of the system in the last single slot of the above idle period. Then:

$$h_{ij}^- = \begin{cases} b(N,0,\sigma) & \text{if } i=j=0 \\ 0 & \text{otherwise} \end{cases} \quad \text{and}$$

$$h_{ij}^+ = \begin{cases} b(N,j,\sigma) & \text{if } i=0;\ j=1,2,\ldots,N \\ 1 & \text{if } i=j;\ i,j=1,2,\ldots,N \\ 0 & \text{otherwise} \end{cases}$$

and using the single slot property of the channel the transition matrix $\mathbf{H} = (h_{ij})$ of the idle cycle period of a cycle is given as:

$$\mathbf{H} = \sum_{k=1}^{\infty} \mathbf{H}_-^{k-1}\mathbf{H}_+ = (\mathbf{I} - \mathbf{H}_-)^{-1}\mathbf{H}_+$$

with elements:

$$h_{ij} = \begin{cases} b(N,j,\sigma)/\{1-b(N,0,\sigma)\} & \text{if } i=0;\ j=1,2,\ldots,N \\ 1 & \text{if } i=j;\ i,j=1,2,\ldots,N \\ 0 & \text{otherwise} \end{cases}$$

Clearly, the matrix $\mathbf{H}$ has a geometric representation. The matrices $\mathbf{H}$, $\mathbf{H}_-$, $\mathbf{H}_+$ and $\mathbf{I}$ are of $(N+1)\times(N+1)$ dimensions, with $\mathbf{I}$ being the unit matrix. In addition $\mathbf{H1} = \mathbf{1}$, where $\mathbf{1}$ denotes the $(N+1)\times 1$ column vector matrix of one's.

Depending on whether the idle period of a cycle is zero or not, there are two cases in which the contention period can be defined. Thus, if the idle period is zero, the contention period is the time interval (in slots) between the time instant when the successfully transmitted packet of the previous cycle ends its transmission, leaving behind one or more ready users (beginning of a new cycle), and the time instant when the first collision with a successful transmission occurs. Alternatively, if the idle period is not zero, the contention period is defined as the time interval between the time instant when the idle period ends leaving behind one or more ready users, and the time instant when the first collision with a successful transmission occurs. Obviously, the contention period is zero if a collision with a successful transmission occurs in the first time instant.

To proceed in the analysis, we further assume that the contention period of a cycle has m single slots, namely, the ready mobiles compete with probability $\delta \neq 0$ during m single slots. Excluding the last slot, we assume that in m_1 single slots all the attempts result in collisions with no successful transmission, and in m_2 single slots only dummy collisions occur (no attempt is made, although some ready mobiles exist). In the last single slot the termination of the contention period indicates the occurrence of a collision with a successful transmission ($m_1 + m_2 + 1 = m$). Since the unsuccessful attempts occur every $\gamma + 1$ single slots, the length of the contention period is $m_1(\gamma + 1) + m_2$ slots. In this case, collisions with no successful transmission are multiplexed with dummy collisions. Obviously, if these attempts result only in collisions with no successful transmissions, the length of the contention period is $m_1(\gamma + 1)$ slots. A typical example of such a case is the 1-persistent model.

Let $\mathbf{R}_- = (r_{ij}^-)$, $\mathbf{R}_* = (r_{ij}^*)$ and $\mathbf{R}_+ = (r_{ij}^+)$ be the one step transition matrices for the embedded m_1, m_2 and the last one single slots of the contention period, respectively. In the case of 1-persistent ($\delta = 1.0$), no dummy collisions occur. Therefore, the probability elements of the transition matrices $\mathbf{R}_-$ and $\mathbf{R}_+$ are:

$$r_{ij}^- = \begin{cases} b(N-i, j-i, \sigma)\{1 - f_j\} & \text{if } i = 2,3,\ldots,N; \\ & j = i, i+1, \ldots, N \\ 0 & \text{otherwise} \end{cases}$$

and:

$$r_{ij}^+ = \begin{cases} b(N-i, j-i+1, \sigma) f_{j+1} & \text{if } i = 2,3,\ldots,N; \\ & j = i-1, i, \ldots, N-1 \\ 1 & \text{if } i = j;\ i, j = 0, 1 \\ 0 & \text{otherwise,} \end{cases}$$

respectively.

Alternatively, in the case of a δ-persistent ($\delta \neq 1.0$) model, let us denote by ζ the number of attempts made by the ready mobiles in a time slot. Then, if $\zeta = 0$ (dummy collisions), the probability elements of the transition matrix $\mathbf{R}_*$ are given by:

$$r_{ij}^* = \begin{cases} b(N-i, j-i, \sigma) b(j, 0, \delta) & \text{if } i = 1,2,\ldots,N; \\ & j = i, i+1, \ldots, N \\ 0 & \text{otherwise,} \end{cases}$$

while if $\zeta \neq 0$ (collisions with no successful transmission), the probability elements of the transition matrix $\mathbf{R}_-$ are given by:

$$r_{ij}^- = \begin{cases} b(N-i, j-i, \sigma) \sum_{\zeta=1}^{j} b(j, \zeta, \delta)\{1 - f_\zeta\} & \\ & \text{if } i = 1,2,\ldots,N; j = i, \\ & i+1, \ldots, N \\ 0 & \text{otherwise.} \end{cases}$$

In both the above cases, the probability elements of the transmission matrix $\mathbf{R}_+$ are:

$$r_{ij}^+ = \begin{cases} b(N-i, j-i+1, \sigma) \sum_{\zeta=1}^{j+1} b(j+1, \zeta, \delta) f_\zeta & \\ & \text{if } i = 1,2,\ldots,N; \\ & j = i-1, i, \ldots, N-1 \\ 1 & \text{if } i = j = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Finally, let $\mathbf{W} = (w_{ij})$ be the one step transition matrix describing the state of the system when only new arrivals may occur (no attempt is possible) in a single slot. Then, the probability elements of the transition matrix $\mathbf{W}$ are given by:

$$w_{ij} = \begin{cases} b(N-i, j-i, \sigma) & \text{if } i = 1,2,\ldots,N; j = i, \\ & i+1, \ldots, N \\ 1 & \text{if } i = j = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Note that the event of terminating the contention period with N ready mobiles left behind and having a collision with a successful transmission is not possible.

Using the single slot property of the channel the transition matrix of the contention period of a cycle $\mathbf{R} = (r_{ij})$ is computed as follows:

$$\mathbf{R} = \sum_{m=1}^{\infty} \{\mathbf{R}_* + \mathbf{R}_- \mathbf{W}^\gamma\}^{m-1} \mathbf{R}_+$$

$$= (\mathbf{I}_- \mathbf{R}_* - \mathbf{R}_- \mathbf{W}^\gamma)^{-1} \mathbf{R}_+ \quad \text{with } \mathbf{R}\mathbf{1} = \mathbf{1}.$$

The series $\sum_{m=0}^{\infty} \{\mathbf{R}_* + \mathbf{R}_- \mathbf{W}^\gamma\}^m$ converges since all the eigenvalues of the matrix $\mathbf{R}_* + \mathbf{R}_- \mathbf{W}^\gamma$ are absolutely less than one. The matrix $\mathbf{R}$ has a geometric representation and, generally, it is difficult to give an explicit form. All the above described matrices $\mathbf{R}_-$, $\mathbf{R}_+$, $\mathbf{R}_*$, $\mathbf{R}$, $\mathbf{W}$ and $\mathbf{I}$ are of $(N+1) \times (N+1)$ dimensions.

Taking multi-path propagation into account, the capture probabilities f_i, $i = 0, 1, 2, \ldots, N$, are derived in the Appendix B. Note that the signal strength of packets received from terminals a distance away from the receiver is a Rayleigh distributed random variable, with the mean of the Rayleigh distribution determined by the power-law distance relationship. In this case, the conditional probability that any packet succeeds given that there are i packets colliding with an initially transmitting packet does not have a simple closed form. Thus values are readily determined numerically using Monte Carlo simulation. This approach has been suggested by Namislo [16], and has also been used in [9].

Finally, the transmission period of a cycle, having fixed length of $\theta + 1$ time slots, immediately follows the contention period. The first slot of this period is allocated for the successful transmission of the precursor. The one step transition matrix for this single slot is given by the

matrix $\mathbf{R}_+$. Taking into account that, during this period, new arrivals from free mobiles may occur but attempts are not allowed, the one step transition matrix for each of the next $\theta - 1$ single slots is given by the matrix $\mathbf{W} = (w_{ij})$, defined previously. Finally, for the last single slot the one step transition matrix $\mathbf{X} = (x_{ij})$ is given by:

$$x_{ij} = \begin{cases} \sigma b(N-i, j-i, \sigma) + (1-\sigma) & \text{if } i = 1,2,\ldots,N; \\ b(N-i, j-i+1, \sigma) & j = i-1, i, \ldots, N \\ 1 & \text{if } i = j = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Note that in this last time slot of the transmission period an additional arrival, coming from the transmitting mobile, may occur. Therefore, the number of ready users may or may not decrease by one.

As it appears, the chain defined between the consecutive end cycle points is Markovian, but since it takes a random amount of time between changes it is actually semi-Markovian. The system transition probability matrix $\mathbf{P} = (p_{ij})$ is calculated as a product of several single slot transition matrices, namely as:

$$\mathbf{P} = (\mathbf{I} - \mathbf{H}_-)^{-1}\mathbf{H}_+(\mathbf{I} - \mathbf{R}_* - \mathbf{R}_-\mathbf{W}^\gamma)^{-1}\mathbf{R}_+\mathbf{W}^{\theta-1}\mathbf{X}$$
$$= \mathbf{H}\mathbf{R}\mathbf{W}^{\theta-1}\mathbf{X}.$$

The resulting stochastic process is ergodic. Therefore, the steady state probability distribution vector $\boldsymbol{\pi} = (\pi_0, \pi_1, \ldots, \pi_N)$ of the backlogged mobiles at the embedded points can be obtained by solving the following linear system:

$$\left\{ \boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P} \quad \text{with} \sum_{i=0}^{N} \pi_i = 1 \right\}.$$

The above theory has been developed when the number N of mobiles is small. In case this number is large but finite and σ is small enough for it to result in the product $N\sigma$ taking moderate values (usually less than 10), the Poisson distribution for the packet arrivals is more appropriate. To follow the analysis, some modifications in the model have to be made. Thus, keeping the same notation and definitions for the transition matrices used for the idle, contention and transmission periods, and introducing the truncated Poisson distribution with mean λ as:

$$Ps(\lambda; j) = \begin{cases} \exp(-\lambda)\lambda^j/j! & \text{if } j = 0,1,2,\ldots,N-1 \\ 1 - \sum_{j=1}^{N-1} \exp(-\lambda)\lambda^j/j! & \text{if } j = N \end{cases}$$

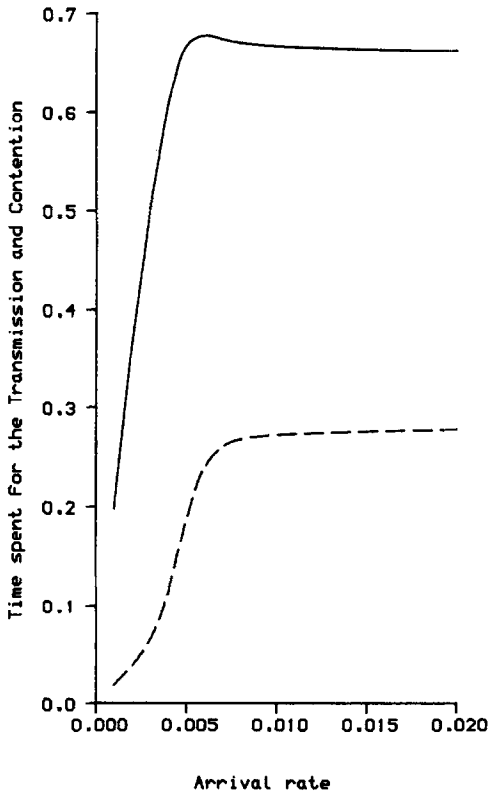


Fig. 2. Arrival rate σ versus the proportion of the time spent in the transmission (throughput S_1: ——) and the contention (U_{con_1}: – – –) period ($N = 19$, $\delta = 1.0$, $\theta = 10$, $\gamma = 3$, S/N = 10 dB).

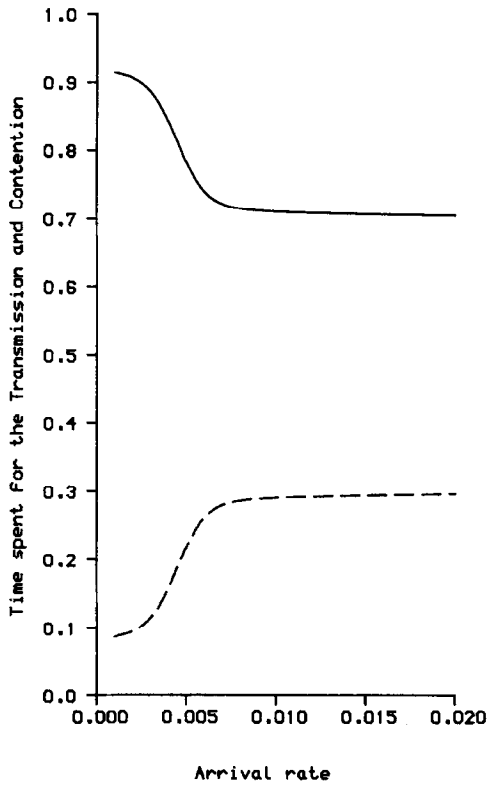


Fig. 3. Arrival rate σ versus the proportion of the time spent in the transmission (throughput S_2: ——) and the contention (U_{con_2}: – – –) period ($N = 19$, $\delta = 1.0$, $\theta = 10$, $\gamma = 3$, S/N = 10 dB). The idle period is excluded.

we may calculate the transition matrix for the idle period as follows:

$$h_{ij}^{-} = \begin{cases} Ps(N\sigma; j) & \text{if } i = j = 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{and}$$

$$h_{ij}^{+} = \begin{cases} Ps(N\sigma; j) & \text{if } i = 0;\ j = 1, 2, \ldots, N \\ 1 & \text{if } i = j;\ i, j = 1, 2, \ldots, N \\ 0 & \text{otherwise.} \end{cases}$$

Therefore:

$$h_{ij} = \begin{cases} Ps(N\sigma; j)/\{1 - Ps(N\sigma; 0)\} & \text{if } i = 0; \\ & j = 1, 2, \ldots, N \\ 1 & \text{if } i = j; \\ & i, j = 1, \ldots, N \\ 0 & \text{otherwise,} \end{cases}$$

and thus:

$$\mathbf{H} = \sum_{k=1}^{\infty} \mathbf{H}_{-}^{k-1}\mathbf{H}_{+} = (\mathbf{I} - \mathbf{H}_{-})^{-1}\mathbf{H}_{+} \quad \text{with} \quad \mathbf{H1} = \mathbf{H},$$

as before.

To calculate the transition matrix for the contention period we assume, without loss of generality, that $f_0 = 0$, $f_1 = 1$ and $f_i = f$, fixed for all $i = 2, 3, \ldots, N$ with $f \in (0, 1)$. Therefore, in the case of 1-persistent ($\delta = 1.0$) we have:

$$r_{ij}^{-} = \begin{cases} Ps((N-i)\sigma; j-i)(1-f) & \text{if } i = 2, 3, \ldots, N; \\ & j = i, i+1, \ldots, N \\ 0 & \text{otherwise} \end{cases}$$

and:

$$r_{ij}^{+} = \begin{cases} Ps((N-i)\sigma; j-i+1)f & \text{if } i = 2, 3, \ldots, N; \\ & j = i-1, i, \ldots, N-1 \\ 1 & \text{if } i = j;\ i, j = 0, 1 \\ 0 & \text{otherwise,} \end{cases}$$

respectively.

Alternatively, in the case of δ-persistent ($\delta \neq 1.0$) and when $\zeta = 0$ (there are no dummy collisions), we have:

$$r_{ij}^{*} = \begin{cases} b(N-i, j-i, \sigma)b(j, 0, \delta) & \text{if } i = 1, 2, \ldots, N; \\ & j = i, i+1, \ldots, N \\ 0 & \text{otherwise,} \end{cases}$$

while, when $\zeta \neq 0$ (there are some dummy collisions), we

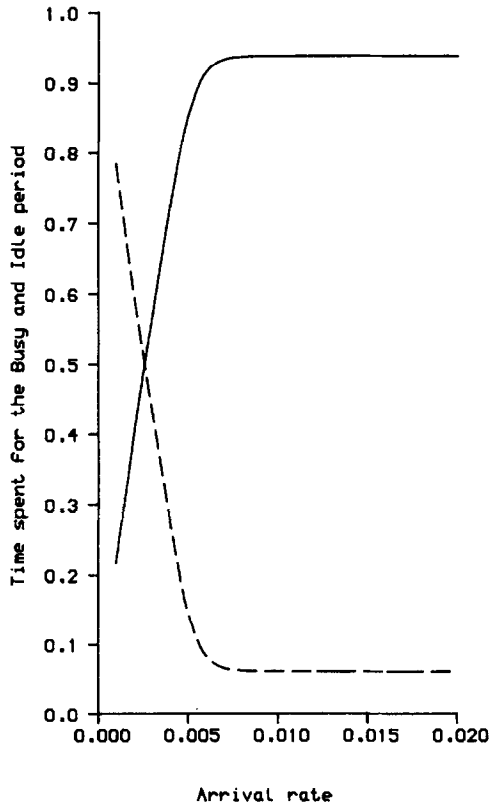


Fig. 4. Arrival rate σ versus the time spent for the busy (U_{bus}: ——) and the idle (U_{idl}: – – –) period ($N = 19$, $\delta = 1.0$, $\theta = 10$, $\gamma = 3$, S/N = 10 dB).

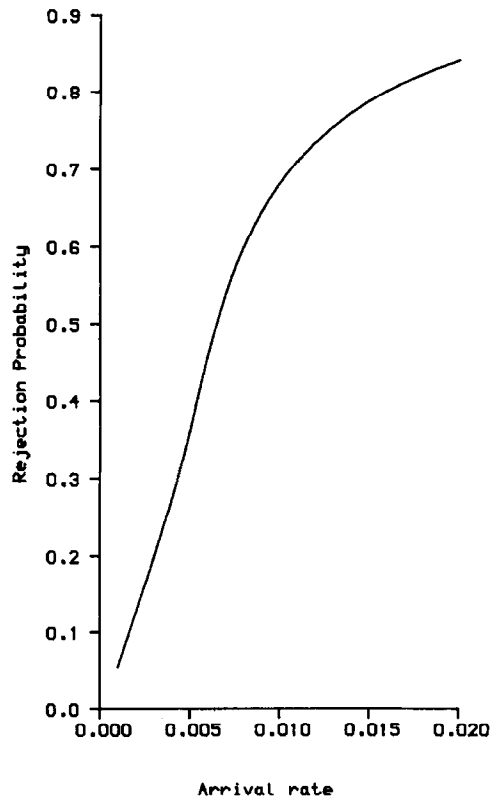


Fig. 5. Arrival rate σ versus the rejection probability Rej ($N = 19$, $\delta = 1.0$, $\theta = 10$, $\gamma = 3$, S/N = 10 dB).

have:

$$r_{ij}^{-} = \begin{cases} Ps((N-i)\sigma; j-i)[1-b(j,0,\delta) & \text{if } i=1, \\ \quad -b(j,1,\delta)](1-f) & \ldots, N \\ & j=i,\ldots, \\ & N-1 \\ 0 & \text{otherwise.} \end{cases}$$

In addition:

$$r_{ij}^{+} = \begin{cases} Ps((N-i)\sigma, j-i+1) & \text{if } i=1,2,\ldots,N \\ \{b(j+1,1,\delta) & j=i-1,\ i,\ldots N-1 \\ \quad +[1-b(j+1,0,\delta) & \\ \quad -b(j+1,1,\delta)]f\} & \\ 1 & \text{if } i=j=0 \\ 0 & \text{otherwise} \end{cases}$$

and:

$$w_{ij} = \begin{cases} Ps((N-i)\sigma; j-i) & \text{if } i=1,2,\ldots,N; \\ & j=i,\ i+1,\ldots,N \\ 1 & \text{if } i=j=0 \\ 0 & \text{otherwise.} \end{cases}$$

Thus, we may calculate the transition matrix of the contention period as:

$$\mathbf{R} = \sum_{m=1}^{\infty}\{\mathbf{R}_{*} + \mathbf{R}_{-}\mathbf{W}^{\gamma})^{m-1}\mathbf{R}_{+} = (\mathbf{I} - \mathbf{R}_{*} - \mathbf{R}_{-}\mathbf{W}^{\gamma})^{-1}\mathbf{R}_{+}$$

with $\mathbf{R1} = \mathbf{1}$.

Finally, in the case of the transmission period we have:

$$x_{ij} = \begin{cases} \sigma Ps((N-i)\sigma; j-i) & \text{if } i=1,2,\ldots,N; \\ \quad +(1-\sigma)Ps((N-i)\sigma; & j=i-1,\ i,\ldots,N \\ \quad j-i+1) & \\ 1 & \text{if } i=j=0 \\ 0 & \text{otherwise,} \end{cases}$$

and the transition matrix **P** may be derived using the same procedures already developed.

4. Performance measures

In this section the most important measures for the system performance will be derived and presented. These measures are the throughput, mean busy time, mean idle time, rejection probability, mean backlog and mean packet delay. The derivation of these measures is based on the mean cycle length, on the mean length of

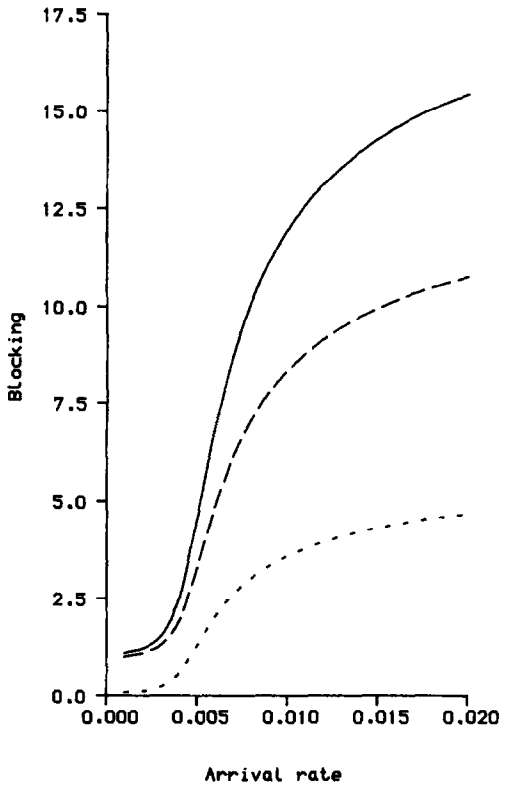


Fig. 6. Arrival rate σ versus the mean queue length of the mobile users during the busy ($\bar{N}_2$: ——), transmission ($\bar{N}_{\text{tra}}$: – – –) and the contention ($\bar{N}_{\text{con}}$: - - - - -) periods ($N = 19$, $\delta = 1.0$, $\theta = 10$, $\gamma = 3$, S/N = 10 dB.

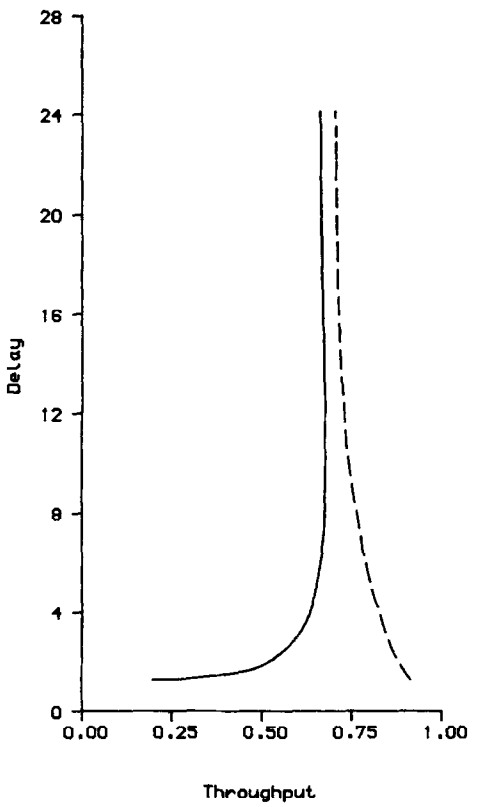


Fig. 7. Throughputs S_1 (——) and S_2 (– – –) versus the mean delay $\bar{D}$ ($N = 19$, $\delta = 1.0$, $\theta = 10$, $\gamma = 3$, S/N = 10 dB).

each period, and on the conditional mean length of the idle and contention periods, given the number of ready users at the beginning of the corresponding period.

To calculate the mean idle period length $\bar{C}_{\text{idl}}$, let $\mathbf{C_I}$ be a $(N+1) \times 1$ column vector matrix with elements of the conditional mean idle period lengths, given that the system had i ($i = 0, 1, 2, \ldots, N$) ready users at the beginning of the idle period. The only non-zero element of this vector corresponds to the state $i = 0$. Since $[\mathbf{H}_-^{k-1}\mathbf{H}_+\mathbf{1}]_{ij}$ is the conditional probability that the idle period length is k slots, given there were i packets at the beginning of this period and j packets at its end, the ith element $[\mathbf{C_I}]_i$ of the column vector $\mathbf{C_I}$ is obtained as follows:

$$[\mathbf{C_I}]_i = \sum_{k=1}^{\infty} k \sum_{j=0}^{N} [\mathbf{H}_-^{k-1}\mathbf{H}_+\mathbf{1}]_{ij} = \sum_{k=1}^{\infty} [k\mathbf{H}_-^{k-1}\mathbf{H}_+\mathbf{1}]_i$$
$$= [\{(\mathbf{I} - \mathbf{H}_-)^{-1}\}^2\mathbf{H}_+\mathbf{1}]_i = [(\mathbf{I} - \mathbf{H}_-)^{-1}\mathbf{H}\mathbf{1}]_i$$
$$= [(\mathbf{I} - \mathbf{H}_-)^{-1}\mathbf{1}]_i.$$

Since $|E(\mathbf{H}_-)| < +\infty$, the matrix $(\mathbf{I} - \mathbf{H}_-)^{-1}$ always exists. Thus, the mean idle period length is given by $\bar{C}_{\text{idl}} = \pi\mathbf{C_I} = \pi(\mathbf{I} - \mathbf{H}_-)^{-1}\mathbf{1}$.

In a similar manner, let $\bar{C}_{\text{con}}$ be the mean length of the contention period and $\mathbf{C_R}$ be a $(N+1) \times 1$ column vector with elements of the conditional mean contention period lengths, given that the system had i ($i = 0, 1, 2, \ldots, N$) ready users at the beginning of the contention period. Obviously, the element corresponding to the state $i = 0$ is zero. The quantity $[(\mathbf{R}_* + \mathbf{R}\_\mathbf{W}^{\gamma})^{m-1}\mathbf{R}_+\mathbf{1}]_{ij}$ represents the conditional probability that the contention period length is m single slots, given there were i packets at the beginning of the period and j packets at its end. Clearly, from the above m slots, the first $m-1$ represent unsuccessful attempts, whereas the last one represents a successful attempt. Therefore, the ith element $[\mathbf{C_R}]_i$ of the column vector $\mathbf{C_R}$ is obtained as:

$$[\mathbf{C_R}]_i = \sum_{m=1}^{\infty} m \sum_{j=0}^{N} [(\mathbf{R}_* + \mathbf{R}\_\mathbf{W}^{\gamma})^{m-1}\mathbf{R}_+\mathbf{1}]_{ij}$$
$$= \sum_{m=1}^{\infty} [m(\mathbf{R}_* + \mathbf{R}\_\mathbf{W}^{\gamma})^{m-1}\mathbf{R}_+\mathbf{1}]_i$$
$$= [\{(\mathbf{I} - \mathbf{R}_* - \mathbf{R}\_\mathbf{W}^{\gamma})^{-1}\}^2\mathbf{R}_+\mathbf{1}]_i$$
$$= [(\mathbf{I} - \mathbf{R}_* - \mathbf{R}\_\mathbf{W}^{\gamma})^{-1}\mathbf{R}_+\mathbf{1}]_i$$
$$= [(\mathbf{I} - \mathbf{R}_* - \mathbf{R}\_\mathbf{W}^{\gamma})^{-1}\mathbf{1}]_i.$$

Since $|E(\mathbf{R}_* + \mathbf{R}\_\mathbf{W}^{\gamma})| < +\infty$, the matrix $(\mathbf{I} - \mathbf{R}_* - \mathbf{R}\_\mathbf{W}^{\gamma})^{-1}$ always exists. Thus, the mean contention period length is given by $\bar{C}_{\text{con}} = \pi\mathbf{C_R} = \pi(\mathbf{I} - \mathbf{R}_* - \mathbf{R}\_\mathbf{W}^{\gamma})^{-1}\mathbf{1}$.

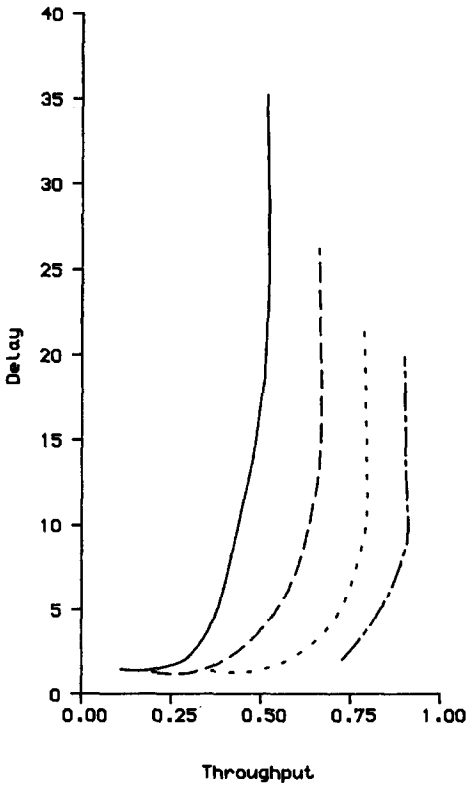


Fig. 8. Throughput S_1 versus the mean delay $\bar{D}$, for $\theta = 5$ (——); $\theta = 10$ (– – –); $\theta = 20$ (- - -); and $\theta = 50$ (– - –). ($N = 19$, $\delta = 1.0$, $\gamma = 10$, S/N = 10 dB).

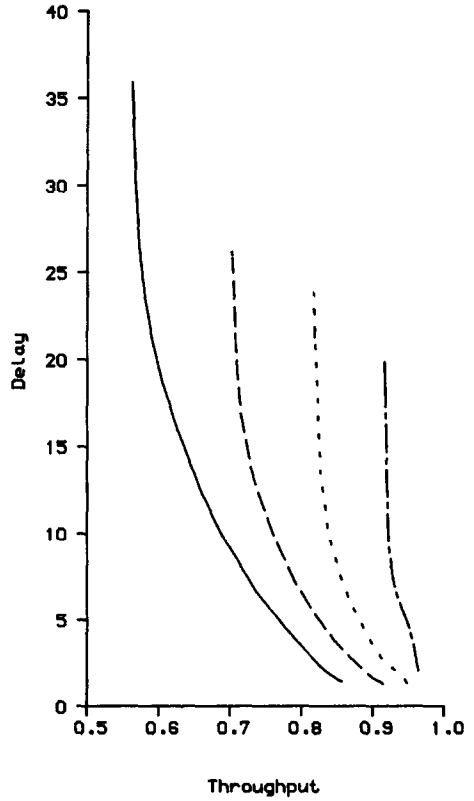


Fig. 9. Throughput S_2 versus the mean delay $\bar{D}$, for $\theta = 5$ (——); $\theta = 10$ (– – –); $\theta = 20$ (- - -); and $\theta = 50$ (– - –). ($N = 19$, $\delta = 1.0$, $\gamma = 10$, S/N = 10 dB). The idle period is excluded.

Finally, the mean cycle length $\bar{C}$ consists of the mean idle period length $\bar{C}_{idl}$, the mean contention period length $\bar{C}_{con}$ and the fixed transmission period of θ time slots. If $\mathbf{C}$ is a $(N+1)\times 1$ column vector such that $\mathbf{C} = \mathbf{C_I} + \mathbf{C_R} + \theta\mathbf{1}$, then,

$$\bar{C} = \pi\mathbf{C} = \pi\{\mathbf{C_I} + \mathbf{C_R} + \theta\mathbf{1}\} = \pi\mathbf{C_I} + \pi\mathbf{C_R} + \pi\theta\mathbf{1}$$
$$= \pi\{(\mathbf{I} - \mathbf{H\_})^{-1} + (\mathbf{I} - \mathbf{R}_* - \mathbf{R\_W}^{\gamma})^{-1} + \theta\}\mathbf{1}.$$

4.1. Throughput

Each cycle carries a successful transmission of θ time slots (a packet without its precursor). Therefore, the throughput of the system is defined as the ratio between the packet transmission time and the mean cycle length $S_1 = \theta/\bar{C}$. Alternatively, the throughput can be defined as $S_2 = \theta/\{\bar{C} - \bar{C}_{idl}\}$, namely, the ratio between the packet transmission time and the mean busy period length.

4.2. Mean busy and mean idle time

The mean busy time of the system, U_{bus_1} say, is defined as the ratio between the mean busy (contention and transmission) period length and the mean cycle length. It consists of two parts. The first part, U_{con_1} say, is the proportion of the time spent in the contention period, while the second is the throughput S_1 of the system. Therefore:

$$U_{bus} = U_{con_1} + S_1 = \bar{C}_{con}/\bar{C} + \theta/\bar{C}$$
$$= \{\pi(\mathbf{I} - \mathbf{R}_* - \mathbf{R\_W}^{\gamma})^{-1}\mathbf{1} + \theta\}/\pi(\mathbf{I} - \mathbf{H\_})^{-1}$$
$$+ (\mathbf{I} - \mathbf{R}_* - \mathbf{R\_W}^{\gamma})^{-1} + \theta\}\mathbf{1}.$$

Alternatively, the proportion of the time spent in the contention period can be defined by excluding the mean idle period length. In this case it will be:

$$U_{con_2} = \bar{C}_{con}/(\bar{C} - \bar{C}_{idl}) = \pi(\mathbf{I} - \mathbf{R}_* - \mathbf{R\_W}^{\gamma})^{-1}\mathbf{1}/$$
$$\pi\{(\mathbf{I} - \mathbf{R}_* - \mathbf{R\_W}^{\gamma})^{-1} + \theta\}\mathbf{1}$$

and $U_{con_2} + S_2 = 1$.

Finally, the mean idle time of the system, U_{idl} say, is obtained as:

$$U_{idl} = 1 - U_{bus} = \bar{C}_{idl}/\bar{C} = \pi(\mathbf{I} - \mathbf{H\_})^{-1}\mathbf{1}/$$
$$\pi\{(\mathbf{I} - \mathbf{H\_})^{-1} + (\mathbf{I} - \mathbf{R}_* - \mathbf{R\_W}^{\gamma})^{-1} + \theta\}\mathbf{1}.$$

4.3. Rejection probability

The rejection probability P_{rej} is obtained as the percentage of users that are not entered into the system.

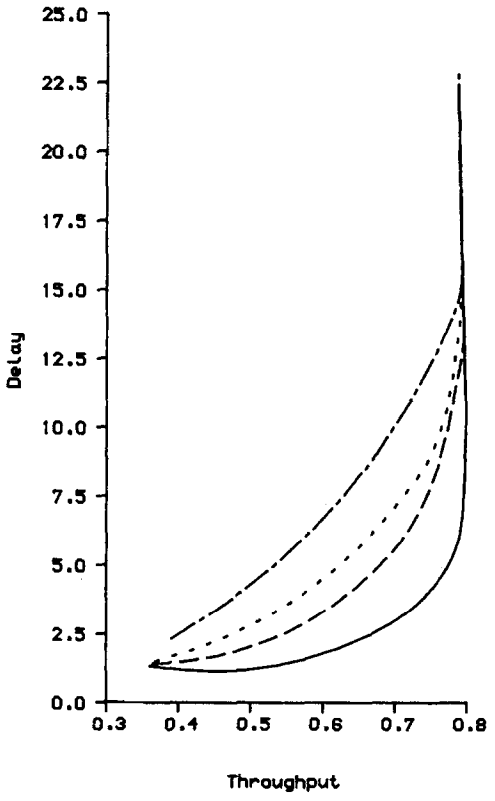


Fig. 10. Throughput S_1 versus the mean delay $\bar{D}$, for $\gamma = 5$ (——); $\gamma = 15$ (– – –); $\gamma = 20$ (- - -); and $\gamma = 35$ (– - –). ($N = 19$, $\delta = 1.0$, $\theta = 20$, S/N = 10 dB).

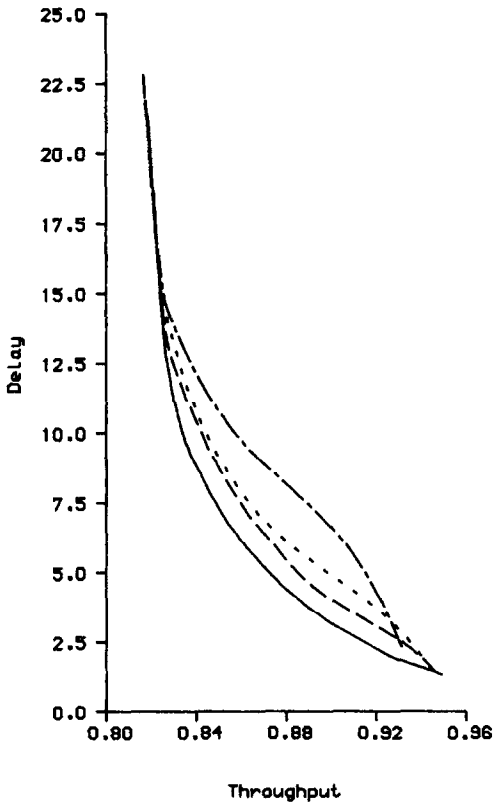


Fig. 11. Throughput S_2 versus the mean delay $\bar{D}$, for $\gamma = 5$ (——); $\gamma = 15$ (– – –); $\gamma = 20$ (- - -); and $\gamma = 35$ (– - –). ($N = 19$, $\delta = 1.0$, $\theta = 20$, S/N = 10 dB). The idle period is excluded.

Therefore, $P_{rej} = (N\sigma - 1/\bar{C})/N\sigma = 1 - 1/N\sigma\bar{C}$.

4.4. Mean backlog

The mean backlog is defined as the mean number of ready users in the system, and it corresponds to the mean queue length, or equivalently, to the mean number of packets in the system. It is computed either as the ratio between the expected number of ready users, $\bar{B}_{bus}$ say, over all slots in the busy period of a cycle, and the mean cycle length $\bar{C}$, or as the ratio between the $\bar{B}_{bus}$ again and the mean busy period length $\bar{C}_{bus} = \bar{C} - \bar{C}_{idl}$. Therefore:

$$\bar{N}_1 = \bar{B}_{bus}/\bar{C} \quad \text{and} \quad \bar{N}_2 = \bar{B}_{bus}/\bar{C}_{bus} \quad \text{with}$$

$$\bar{B}_{bus} = \bar{B}_{con} + \bar{B}_{tra} = \pi\mathbf{B}_\mathbf{R} + \pi\mathbf{B}_\mathbf{X}.$$

Note that $\bar{B}_{con} = \pi\mathbf{B}_\mathbf{R}$ and $\bar{B}_{tra} = \pi\mathbf{B}_\mathbf{X}$ are the expected number of ready users in the contention and transmission periods, respectively. In addition, $\mathbf{B}_\mathbf{R} = ([\mathbf{B}_\mathbf{R}]_i)$ and $\mathbf{B}_\mathbf{X} = ([\mathbf{B}_\mathbf{X}]_i)$; $i = 0, 1, 2, \ldots, N$ are $(N+1) \times 1$ column vectors with elements as the average number of ready users over all slots in the contention or transmission period, respectively, given that the system had i ready users at its beginning. Thus, if $\mathbf{z} = (0, 1, 2, \ldots, N)^T$ is a $(N+1) \times 1$ column vector matrix with elements as the number of ready users corresponding to the state i then:

$$[\mathbf{B}_\mathbf{R}]_i = \sum_{q=0}^{N} [\mathbf{H}]_{iq} \sum_{m=1}^{\infty} \sum_{j=0}^{N} j[(\mathbf{R}_* + \mathbf{R}\_\mathbf{W}^\gamma)^{m-1}]_{qj}$$

$$= \left[\mathbf{H} \sum_{m=1}^{\infty} (\mathbf{R}_* + \mathbf{R}\_\mathbf{W}^\gamma)^{m-1}\mathbf{z}\right]_i$$

$$[\mathbf{B}_\mathbf{X}]_i = \sum_{q=0}^{N} [\mathbf{H}]_{iq} \sum_{r=0}^{N} [\mathbf{R}]_{qr} \sum_{n=1}^{\theta-1} \sum_{j=0}^{N} j[\mathbf{W}^n]_{rj}$$

$$= \left[\mathbf{HR} \sum_{n=1}^{\theta-1} \mathbf{W}^n\mathbf{z}\right]_i.$$

Obviously, the number of ready mobiles during the idle period is zero. The expected number of ready mobiles during the contention and transmission periods is given, respectively, as:

$$\bar{B}_{con} = \pi\mathbf{B}_\mathbf{R} = \pi\mathbf{H}(\mathbf{I} - \mathbf{R}_* - \mathbf{R}\_\mathbf{W}^\gamma)^{-1}\mathbf{z} \quad \text{and}$$

$$\bar{B}_{tra} = \pi\mathbf{B}_\mathbf{X} = \pi\mathbf{HR} \sum_{n=1}^{\theta-1} \mathbf{W}^n\mathbf{z},$$

while the mean backlog in the contention $\bar{N}_{con}$ say, and the transmission $\bar{N}_{tra}$ say, periods can be obtained,

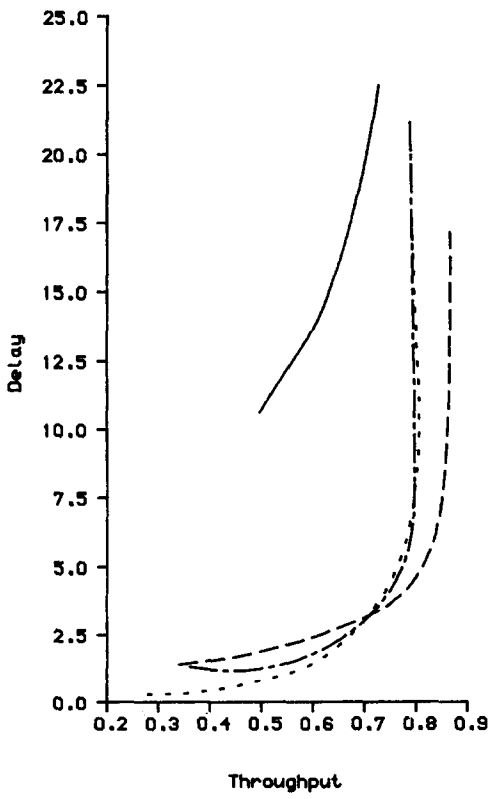


Fig. 12. Throughput S_1 versus the mean delay $\bar{D}$, for $\delta = 0.01$ (——); $\delta = 0.10$ (– – –); $\delta = 0.50$ (- - - -); and $\delta = 1.00$ (– - –). ($N = 19$, $\theta = 20$, $\gamma = 5$, S/N = 10 dB).

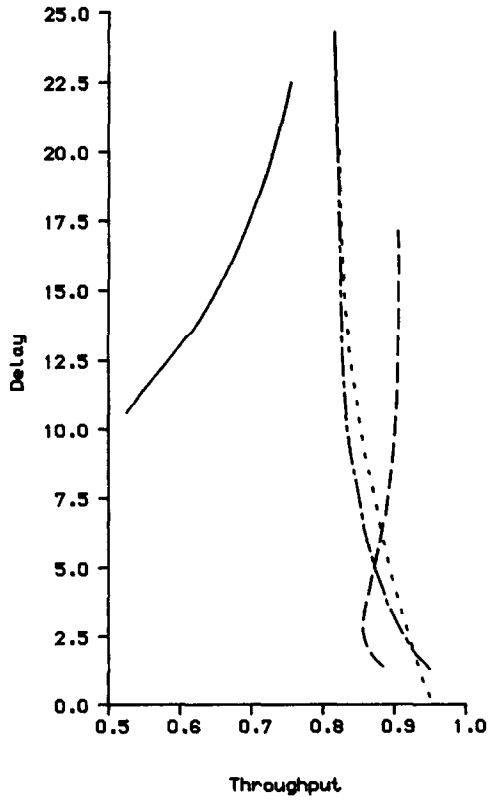


Fig. 13. Throughput S_2 versus the mean delay $\bar{D}$, for $\delta = 0.01$ (——); $\delta = 0.10$ (– – –); $\delta = 0.50$ (- - - -); and $\delta = 1.00$ (– - –). ($N = 19$, $\theta = 20$, $\gamma = 5$, S/N = 10 dB). The idle period is excluded.

respectively, as:

$$\bar{N}_{con} = \bar{B}_{con}/\bar{C}_{bus} \quad \text{and} \quad \bar{N}_{tra} = \bar{B}_{tra}/\bar{C}_{bus} \quad \text{with}$$
$$\bar{N}_2 = \bar{N}_{con} + \bar{N}_{tra}.$$

4.5. Mean packet delay

The mean packet delay, $\bar{D}$ say, can be derived through Little's formula. Since $\bar{N}_1/S_1 = \bar{N}_2/S_2 = \bar{B}_{bus}/\theta$, the mean delay is given by:

$$\bar{D} = \bar{B}_{bus}/\theta = \pi\mathbf{H}(\mathbf{I} - \mathbf{R}_* - \mathbf{R}_-\mathbf{W}^\gamma)^{-1} \times \left[\mathbf{I} + \mathbf{R}_+ \sum_{n=1}^{\theta-1} \mathbf{W}^n\right]\mathbf{z}/\theta.$$

The mean backlogs are given, respectively, by:

$$\bar{N}_1 = \pi\left\{\mathbf{H}(\mathbf{I} - \mathbf{R}_* - \mathbf{R}_-\mathbf{W}^\gamma)^{-1}\mathbf{z} + \mathbf{HR}\sum_{n=1}^{\theta-1}\mathbf{W}^n\mathbf{z}\right\}\Big/\bar{C}$$
$$= \pi\mathbf{H}(\mathbf{I} - \mathbf{R}_* - \mathbf{R}_-\mathbf{W}^\gamma)^{-1}\left\{\mathbf{I} + \mathbf{R}_+\sum_{n=1}^{\theta-1}\mathbf{W}^n\right\}\mathbf{z}/\bar{C}$$
$$\bar{N}_2 = \pi\left\{\mathbf{H}(\mathbf{I} - \mathbf{R}_* - \mathbf{R}_-\mathbf{W}^\gamma)^{-1}\mathbf{z} + \mathbf{HR}\sum_{n=1}^{\theta-1}\mathbf{W}^n\mathbf{z}\right\}\Big/\bar{C}_{bus}$$
$$= \pi\mathbf{H}(\mathbf{I} - \mathbf{R}_* - \mathbf{R}_-\mathbf{W}^\gamma)^{-1}\left\{\mathbf{I} + \mathbf{R}_+\sum_{n=1}^{\theta-1}\mathbf{W}^n\right\}\mathbf{z}/\bar{C}_{bus}.$$

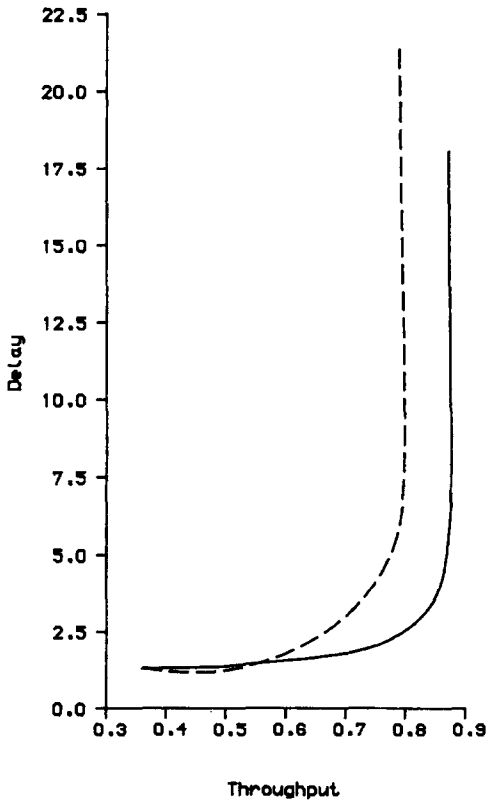


Fig. 14. Throughput S_1 versus the mean delay $\bar{D}$, for S/N = 3 dB (——); S/N = 10 dB (– – –). ($N = 19$, $\theta = 20$, $\gamma = 5$, $\delta = 1.00$).

5. Model relaxation

Up to now the model has been based on the assumption of a fixed packet length θ and channel recovery time γ. To develop the analysis for the case of variable packet length and variable channel recovery time, let Θ and Γ be two discrete random variables with generating functions

$$Q_\Theta(z) = \sum_{\lambda=1}^{\infty} z^\lambda \Pr\{\Theta = \lambda\} \quad \text{and}$$
$$\Phi_\Gamma(z) = \sum_{\nu=1}^{\infty} z^\nu \Pr\{\Gamma = \nu\},$$

respectively. There are three cases to be considered. In the first case, only the packet length varies. Then, the matrix $\mathbf{W}^\Theta$ is replaced by $Q_\Theta(\mathbf{W})$ and the system transmission matrix becomes:

$$\mathbf{P} = (\mathbf{I} - \mathbf{H}_-)^{-1}\mathbf{H}_+(\mathbf{I} - \mathbf{R}_* - \mathbf{R}_-\mathbf{W}^\gamma)^{-1}\mathbf{R}_+Q_\Theta(\mathbf{W})\mathbf{X}$$
$$= \mathbf{HR}Q_\Theta(\mathbf{W})\mathbf{X}.$$

In the second case, only the length γ of the channel recovering time varies. Then the matrix $\mathbf{W}^\Gamma$ is replaced

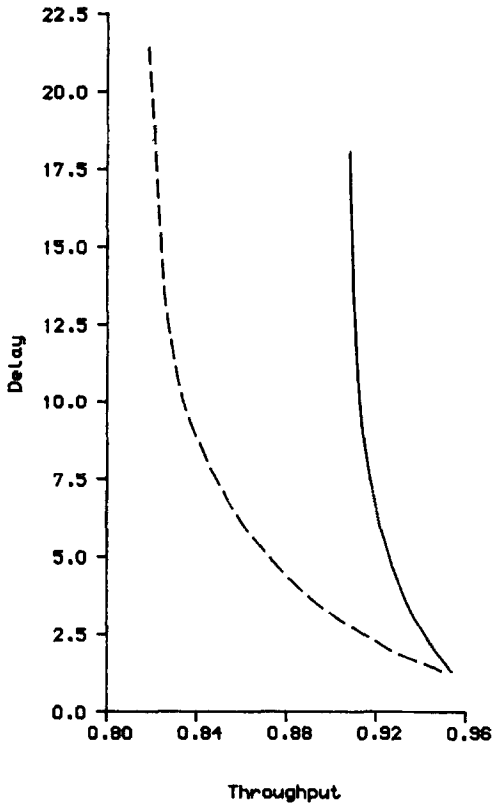


Fig. 15. Throughput S_2 versus the mean delay $\bar{D}$, for S/N = 3 dB (——); S/N = 10 dB (– – –). ($N = 19$, $\theta = 20$, $\gamma = 5$, $\delta = 1.00$). The idle period is excluded.

by $\Phi_\Gamma(\mathbf{W})$, and the system transition matrix becomes:

$$\mathbf{P} = (\mathbf{I} - \mathbf{H}_-)^{-1}\mathbf{H}_+(\mathbf{I} - \mathbf{R}_* - \mathbf{R}_-\Phi_\Gamma(\mathbf{W}))^{-1}\mathbf{R}_+\mathbf{W}^{\theta-1}\mathbf{X}$$
$$= \mathbf{H}(\mathbf{I} - \mathbf{R}_* - \mathbf{R}_-\Phi_\Gamma(\mathbf{W}))^{-1}\mathbf{R}_+\mathbf{W}^{\theta-1}\mathbf{X}.$$

In the last case, both θ and γ are varied. Then, the system transition matrix becomes:

$$\mathbf{P} = (\mathbf{I} - \mathbf{H}_-)^{-1}\mathbf{H}_+(\mathbf{I} - \mathbf{R}_* - \mathbf{R}_-\Phi_\Gamma(\mathbf{W}))^{-1}\mathbf{R}_+Q_\Theta(\mathbf{W})\mathbf{X}$$
$$= \mathbf{H}(\mathbf{I} - \mathbf{R}_* - \mathbf{R}_-\Phi_\Gamma(\mathbf{W}))^{-1}\mathbf{R}_+Q_\Theta(\mathbf{W})\mathbf{X}.$$

All the measures derived in the previous section may now recalculated, with θ and γ being replaced with their means $\bar{\theta} = \sum_{\lambda=1}^{\infty} \lambda \Pr\{\Theta = \lambda\}$ and/or $\bar{\gamma} = \sum_{\nu=1}^{\infty} \nu \Pr\{\Gamma = \nu\}$, according to the three cases examined above.

6. Numerical results

In this section, both the finite and infinite population schemes analysed above are tested, and the results obtained discussed. Although a unified analysis of the finite population scheme has been achieved, the model with large but finite mobile users is excluded from the current tests. Currently under investigation is the derivation of some efficient parallel algorithms for solving linear systems with large but non-symmetric matrices. Details can be found in Ref. [19].

We start with the results found when the synchronous model of moderate population size is considered. This system is tested with $N = 19$ mobiles and for various values of the packet length θ, the channel recovery time γ, and the attempt probability δ. The SNR (signal-to-noise ratio) value is kept fixed either at 10 dB (the typical cellular radio capture ratios for the GSM system is around 9 dB), or at 3 dB (this value is served as a bound; see Appendix B). All the measures defined in Section 4 are examined, and the most important results are reported. The steady-state probabilities are obtained by solving a linear system, using the direct approach of the Gaussian elimination method. The matrix $(\mathbf{I} - \mathbf{H}_-)^{-1}$ is found analytically while the matrices $(\mathbf{I} - \mathbf{R}_* - \mathbf{R}_-\mathbf{W}^\gamma)^{-1}$ and $\mathbf{W}^{\theta-1}$ are calculated by using standard (serial) NAG routines.

In Figs. 2–7 all the measures discussed in the previous

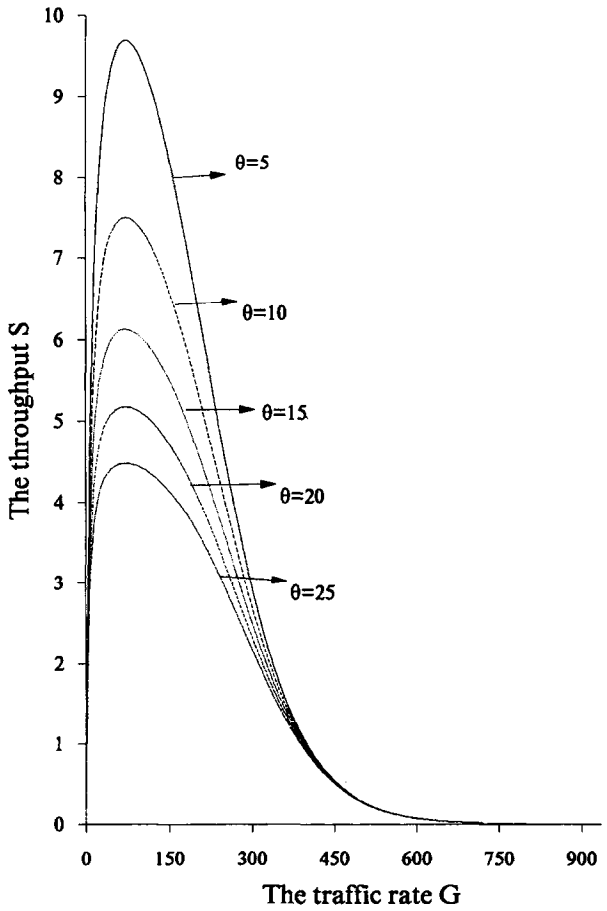


Fig. 16. The throughput S versus the traffic rate G for different values of packet length ($\theta = 5$, 10, 15, 20, 25 slots), when capture is perfect ($f = 1.0$).

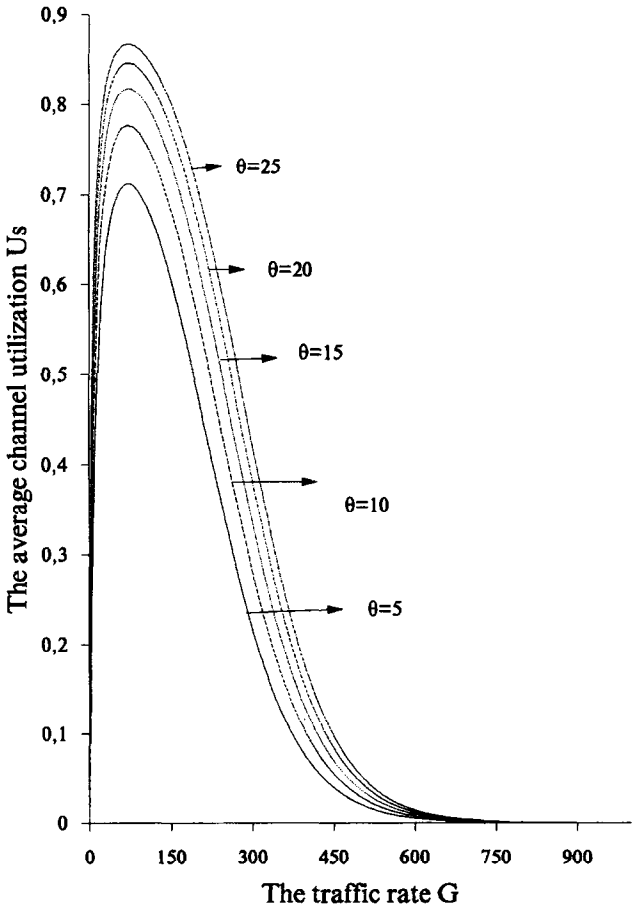


Fig. 17. The average channel utilization U_s versus the traffic rate G for different values of packet length ($\theta = 25$, 20, 15, 10, 5 slots), when capture is perfect ($f = 1.0$).

section are plotted against the arrival rate. Here the packet length is $\theta = 10$, the channel recovery time is $\gamma = 3$, the signal-to-noise ratio value is SNR = 10 dB, and the attempt probability is $\delta = 1.0$ (1-persistent model). In particular, in Fig. 2 (and Fig. 3), the throughput $S_1(S_2)$ and the proportion of time spent in the contention period $U_{\text{con}_1}(U_{\text{con}_2})$ against the arrival rate are presented. In Fig. 4, the proportion of the time spent for the busy U_{bus} and idle U_{idl} periods against the arrival rate is presented, while the rejection probability is examined in Fig. 5. The mean queue lengths of the backlogged mobiles (blocking) during the busy $\bar{N}_2$, the contention $\bar{N}_{\text{con}}$ and the transmission $\bar{N}_{\text{tra}}$ periods against the arrival rate are presented in Fig. 6. Finally, in Fig. 7, the throughput–delay performance is presented.

All the above graphs are consistent, in the sense that with the examined arrival rates the risk of instability is low. Thus, for low arrival rates, almost every packet is successfully transmitted, and therefore the channel appears to have a good operational point. The following results are similar to those presented in [9], where the CSMA protocol with power capture was under investigation. Thus, the values of S_2 and U_{con_2} are higher than those of S_1 and U_{con_1}, respectively, due to the time spent in the idle period (Figs. 2 and 3). In addition, the idle period is longer than the busy period, but as the arrival rate increases the situation is reversed (Fig. 4). Since the transmission period is longer than the other two periods of a cycle, the rejection probability and the blocking increases with the packet length (Figs. 5 and 6). Finally, if the arrival rate is small the delay is small, namely, almost every packet is successfully transmitted (Fig. 7). In this case S_1 takes its lowest values while S_2 takes its maximum, but as the arrival rate increases the situation is reversed. It is expected that as the packet length increases, the throughput S_1 and S_2 tend to obtain the same values.

In order to study the effects of the interaction between the different factors applied, a more comprehensive examination is required. To facilitate the comparison, only the most representative figures of the throughput–delay performance are reported. There are four parameters to be examined. Thus, in Figs. 8 and 9 we present the effect of the packet length variation ($\theta = 5$, 10, 20, 50). Here the attempt rate is $\delta = 1.0$ (1-persistent model),

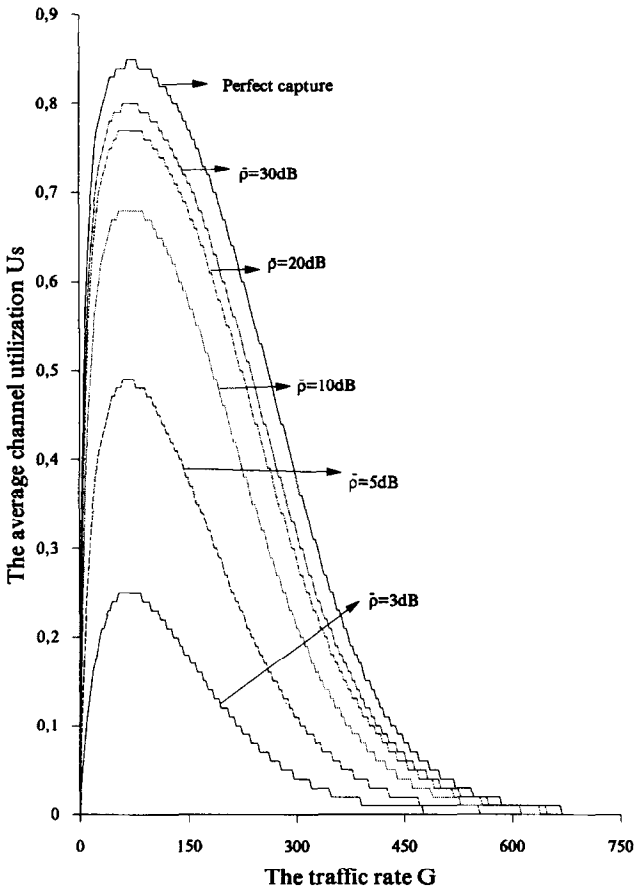


Fig. 18. The average utilization U_s versus the traffic rate G in a slow fading environment: NCFSK modulation, 48-bit precursor, fixed packet length ($\theta = 20$ slots), fixed channel recovery time ($\gamma = 3$ slots), various average SNR values.

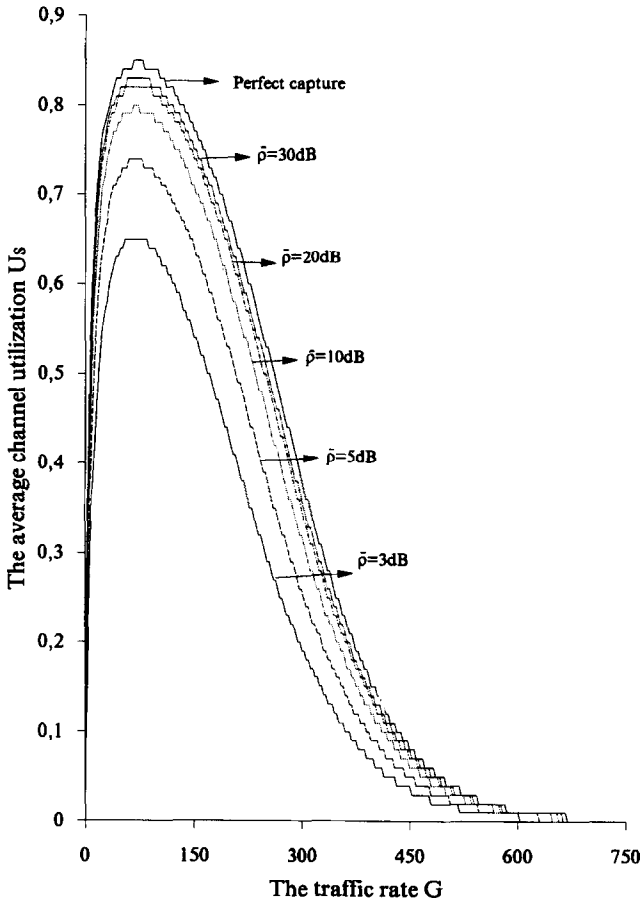


Fig. 19. The average utilization U_s versus the traffic rate G in a slow fading environment: CPSK modulation, 48-bit precursor, fixed packet length ($\theta = 20$ slots), fixed channel recovery time ($\gamma = 3$ slots), various average SNR values.

the channel recovery time is $\gamma = 10$ slots and the signal-to-noise ratio is SNR = 10 dB. Similarly, in Figs. 10 and 11 the effect of the variation in the channel recovery time ($\gamma = 5, 15, 20, 35$) is examined. Here, the packet length is $\theta = 20$, while the rest of the parameters are kept fixed ($\delta = 1.0$, and SNR = 10 dB). In Figs. 12 and 13 the attempt probability variation ($\delta = 0.01, 0.10, 0.50, 1.0$) is examined, with $\theta = 20$ slots, $\gamma = 5$ slots, and SNR = 10 dB. Finally, in Figs. 14 and 15, the effect of the signal-to-noise variation (SNR = 10 dB, 3 dB) is examined, keeping the remaining parameters fixed ($\theta = 20$ slots, $\gamma = 5$ slots and $\delta = 1.0$).

From Figs. 8 and 9, it appears that as the mobiles transmit longer packets, the throughput increases significantly at the expense of the idle and contention period. In addition, the effect of the channel recovery time is significant. The smaller the channel recovery time γ, the better the throughput results (Figs. 10 and 11). However, an interesting result seem to be that, as the arrival rate increases, the influence of γ almost disappears. Figs. 12 and 13 show that, for small values of the attempt probability δ and as the arrival rate increases, packets seem to suffer from an excessive delay. Thus, although the attempt probability plays an important role, especially when the arrival rate is light, it seems to have no further effect for high values of the arrival rate. Values of the attempt probability δ around 0.1 seem to be the most appropriate. It is important to observe that there is an intersection among the curves corresponding to $\delta = 0.1$, 0.5 and 1.0. This is a good operational point, namely, the channel operates with an acceptable level of throughput–delay performance, independently of the value of the attempt rate. Finally, in Figs. 14 and 15, the capture effect, by means of the signal-to-noise variation (SNR = 3 dB, 10 dB) is examined with $\theta = 20$ slots, $\gamma = 5$ slots and $\delta = 1.0$. The results show that the throughput is higher and the risk of instability is lower as the capture ratio is decreased (i.e. the capture probability is increased, and thus the power capture effect becomes more significant; see Appendix B). However, it is expected that the throughput of the channel will obtain a maximum and then, for large values of traffic, it will decrease slowly, improving the stability of the channel.

We now consider the asynchronous model with an

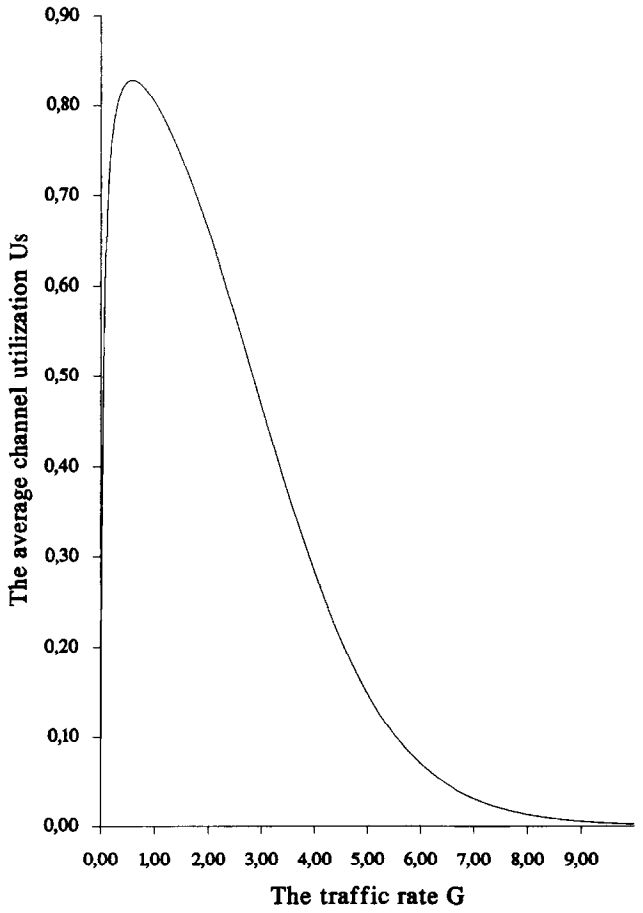


Fig. 20. The average channel utilization U_s (also referred as throughput S in [8]) versus the traffic rate G for the CSMA/CD protocol without capture ($\theta = 20$ slots, $\gamma = 3$ slots).

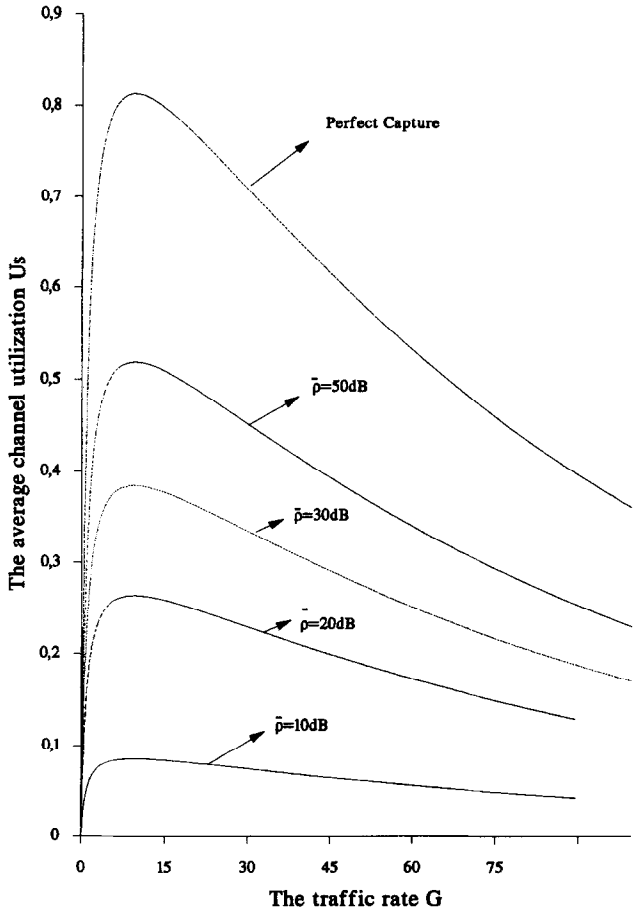


Fig. 21. The average channel utilization U_s (also referred as throughput S in [8]) versus the traffic rate G for nonpersistent CSMA ($a = 0.01$) and BTMA protocol in slow fading: NCFSK modulation, 1000-bit packet 100-bit acknowledgement, various average SNR values.

infinite population of mobiles. Thirteen figures are presented in total (Figs. 16–28). In Fig. 16 we show the relationship between the sense point rate G and the channel utilization achieved (also referred as throughput in [8]) for different packet length values, i.e. w: 5, 4, 3, 2, 1. These values of w correspond to θ: 25, 20, 15, 10, 5 time slots. Without loss of generality, we may assume perfect capture (the case of imperfect capture is examined in the sequel). However, this assumption may also be used to show a bound for performance. We note that $\bar{U}_s$ is a bell-shape function in channel traffic G, that is, $\bar{U}_s$ increases with G until reaching a maximum value, then decreases to zero as G continues to increase. As expected, the average channel utilization is increased with the packet length at the expense of the average time the channel is busy due to collisions or due to an undetected precursor. In addition, the channel shows a bistable behaviour unless the offered traffic is extremely low. However, in Fig. 17, where the average utilization U_s versus the traffic rate G is presented, it is important to observe that the maximum value is obtained for the same value of G, namely, independently of the average length of the message.

In the sequel (Figs. 18–28), we study the performance of our hybrid model, which is produced by using a slot of 48 bits, packet length of $\theta = 20$ slots, and channel recovery time of $\gamma = 3$ slots. We start with the Figs. 18 and 19, where we present the average utilization $\bar{U}_s$ versus the traffic rate G for various average SNR values ($\bar{\rho} = 3$ dB, 5 dB, 10 dB, 20 dB, 30 dB, ∞) using NCFSK and CPSK modulation, respectively. Higher values of $\bar{\rho}$ leads to the system performance approaching the non-fading case. Thus, for an average SNR of about 30 dB the performance is close to the conventional CSMA/CD case (see Fig. 20). This is particularly true for the case of CPSK modulation, which presents higher values of $\bar{U}_s$ compared with those found in the NCFSK case (see Figs. 18 and 19). From the above, it appears that there is no improvement in the performance between our hybrid model and the classic CSMA/CD protocol when the capture is present. In fact, channel utilization is less for the capture case than for CSMA/CD, and therefore there is a degradation from CSMA/CD to the new protocol. However, our hybrid protocol has a similar performance to the classic CSMA/CD protocol in the case of perfect

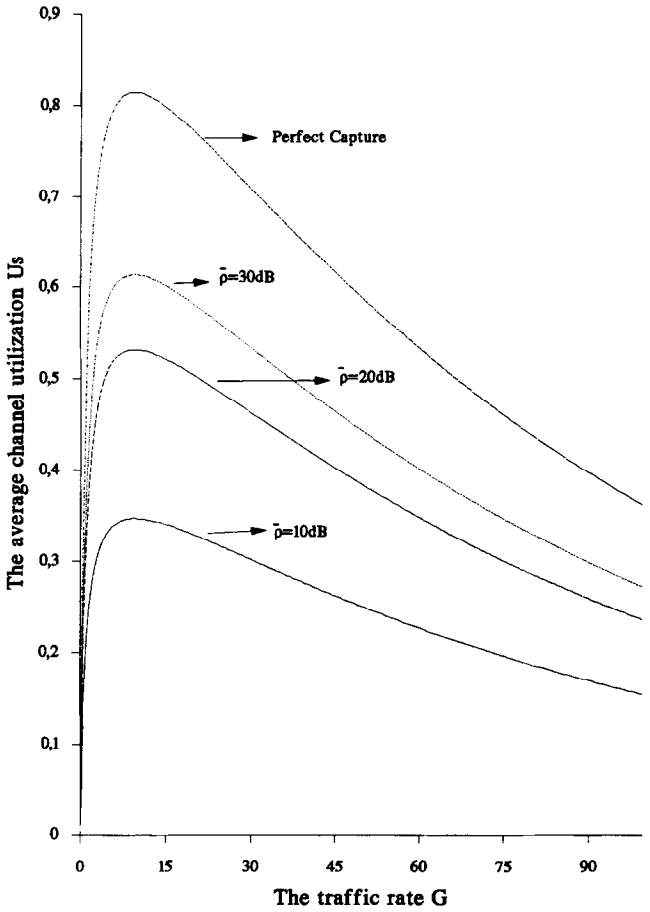


Fig. 22. The average channel utilization U_s (also referred as throughput S in [8]) versus the traffic rate G for nonpersistent CSMA ($a = 0.01$) and BTMA protocol in slow fading: CPSK modulation, 1000-bit packet, 100-bit acknowledgement, various average SNR values.

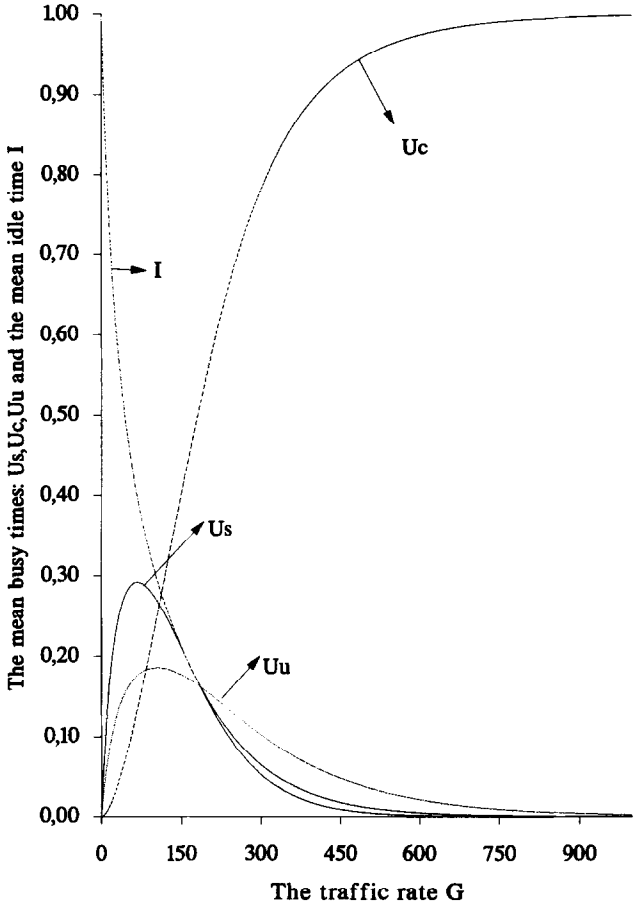


Fig. 23. The traffic rate G versus the average time the channel is busy (U_s, U_c, U_u) or idle (I) in a slow fading environment: capture probability $f = 0.1$ (for NCFSK average SNR: 3.3 dB, for CPSK average SNR: 1.2 dB), $\theta = 20$ slots, $\gamma = 3$ slots.

capture (see Figs. 18, 19 and 20). This result is particularly important, because in the slow fading environment the capture probability (which has been defined as the probability for the cell site to detect an uncollided precursor) is usually high, and therefore a result similar to the classic CSMA/CD performance can be achieved.

In Figs. 21 and 22 we present the average utilization $\bar{U}_s$ (which is considered as throughput in [8]), versus the traffic rate G, for the non-persistent CSMA ($\alpha = 0.01$) and BTMA protocol with NCFSK and CPSK modulations, respectively. Various average SNR values ($\bar{\rho}$: 10 dB, 20 dB, 30 dB, ∞) are examined. We see that the average utilization is remarkably higher in the case of CPSK modulation than the corresponding NCFSK modulation in all the examined values of the average SNR. In both cases, higher values of $\bar{\rho}$ leads to the system performance approaching the non-fading case. Note that $\bar{U}_s$ may theoretically attain 1 by letting $\alpha \to 0$, $G \to \infty$ and $\bar{\rho} \to \infty$. However, a very interesting result appears by comparing our hybrid model (Figs. 18 and 19), and the non-persistent CSMA ($\alpha = 0.01$) and BTMA protocol (Figs. 21 and 22), for different average SNR values. We observe that there is a significant improvement in the performance of our model when capture is present.

In Figs. 23–26 the performance of the mean busy time due to successful transmissions $\bar{U}_s$, the mean busy time due to collisions $\bar{U}_c$, the mean busy time due to an undetected precursor $\bar{U}_u$ and the mean idle time $\bar{I}$ are examined for different values of the capture probability. The examined values of the capture probability f: 1.0, 0.9, 0.5, 0.1 correspond to different average SNR values, namely, $\bar{\rho}$: ∞, 71.3 dB, 10.8 dB, 3.3 dB, using the NCFSK modulation and $\bar{\rho}$: ∞, 25.0 dB, 3.8 dB, 1.2 dB, using the CPSK modulation, respectively. It appears that $\bar{U}_s$ increases with the capture probability f (or equivalently with the average SNR values). In addition, one should also note that $\bar{U} + \bar{I} = (\bar{U}_s + \bar{U}_c + \bar{U}_u) + \bar{I} = 1$.

Finally, Figs. 27 and 28 present the delay $\bar{D}$ versus the average channel utilization $\bar{U}_s$, for various average SNR values ($\bar{\rho} = 3$ dB, 5 dB, 10 dB, 20 dB, ∞), and using

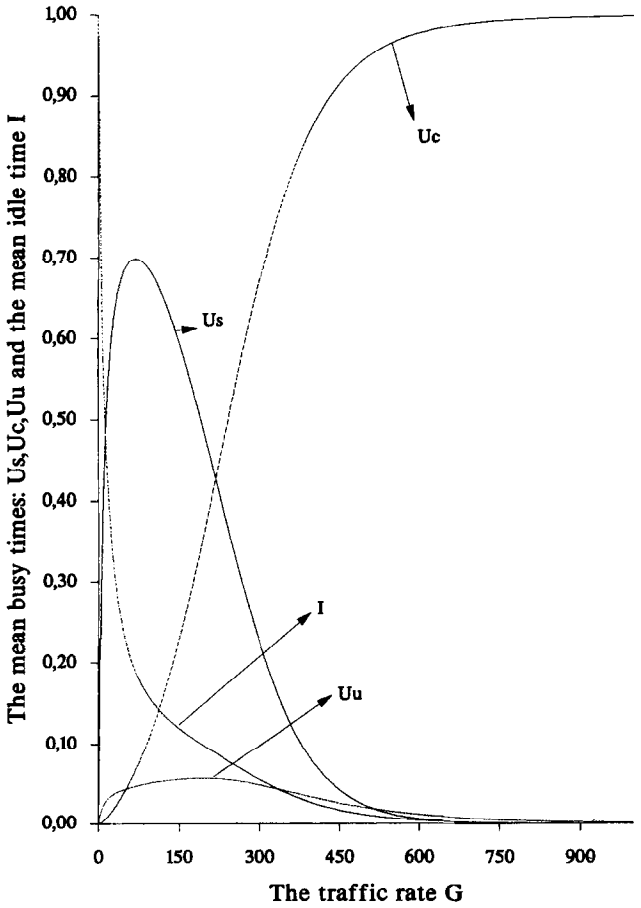


Fig. 24. The traffic rate G versus the average time the channel is busy (U_s, U_c, U_u) or idle (I) in a slow fading environment: Capture probability $f = 0.5$ (for NCFSK average SNR: 10.8 dB, for CPSK average SNR: 3.8 dB), $\theta = 20$ slots, $\gamma = 3$ slots.

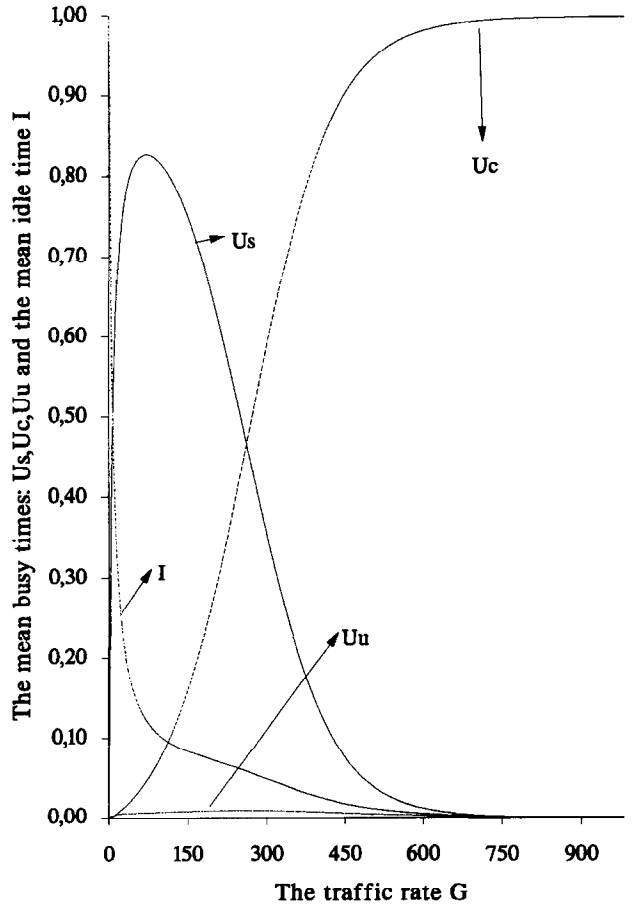


Fig. 25. The traffic rate G versus the average time the channel is busy (U_s, U_c, U_u) or idle (I) in a slow fading environment: Capture probability $f = 0.9$ (for NCFSK average SNR: 71.3 dB, for CPSK average SNR: 25.0 dB), $\theta = 20$ slots, $\gamma = 3$ slots.

NCFSK and CPSK modulation, respectively. They also indicate a higher delay for low average SNR. Clearly, the channel shows a bistable behaviour, with a good operational point obtained when the traffic is low.

However, as stated in Section 2.1, the same model may be applied if the power capture is present. In this case, the realistic assumption $f_i = f_k$, $i \geq k$ for some k yields an upper bound on the throughput. For example, taking $f_i = f$, fixed; $i \geq 2$, we may reproduce Figs. 23–26. However, f presents the power capture, and therefore the interpretation of these graphs is now different. Thus, higher values of f are obtained for lower values of the capture ratio c, or the average SNR (see Appendix B). This result shows that a performance similar to the classic CSMA/CD is achieved when f is close to one. Note that previous work [9] shows that protocols which control contention better than ALOHA (e.g. CSMA) do not have as dramatic an improvement with capture as ALOHA. Our new result is that protocols which control contention better than CSMA (e.g. CSMA/CD) do not have a similar improvement with capture as CSMA.

7. Conclusions

In this work a performance study for the RCC channel on which the mobiles randomly send system access messages has been presented. The access method used for multiplexing the mobiles over the access channel is a hybrid between the CSMA/CD and the BTMA protocols. Thus, a combination of the well known features of the above protocols has been achieved.

Based on the regenerated time instants between successful transmissions and on the slot property of the channel, a new, semi-Markov synchronous time renewal model with finite number of mobiles is constructed, analysed and discussed. A unified analysis is produced, resulting in a geometric matrix representation form model which includes both schemes with a moderate and a large but finite number of mobiles. All the important measures are derived in an efficient and computationally tractable form. Numerical results are obtained in the case of the moderate population scheme using a standard computational approach. It is suggested that in the case of a large but finite number of users, the

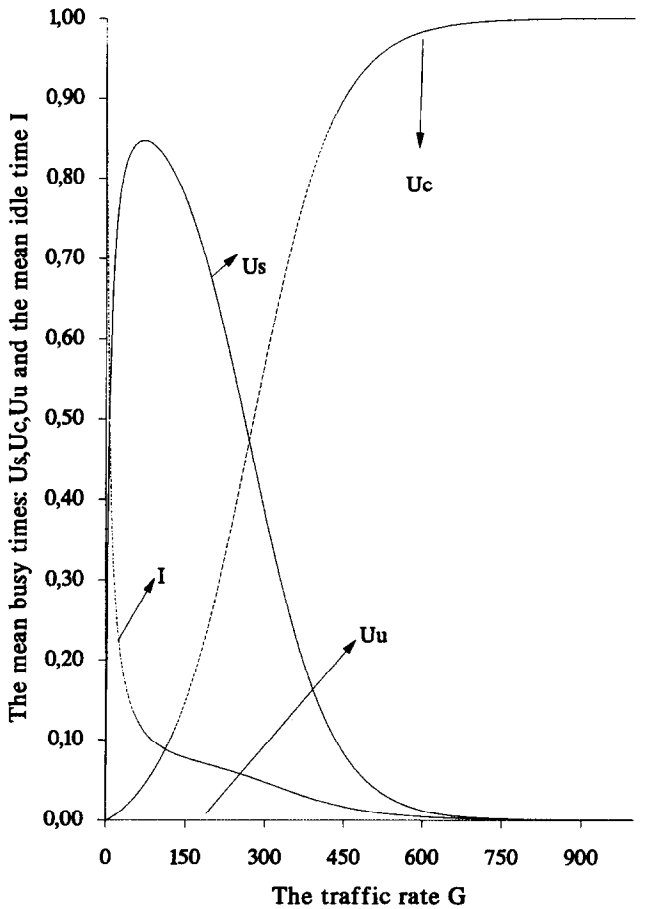


Fig. 26. The traffic rate G versus the average time the channel is busy (U_s, U_c, U_u) or idle (I) in a slow fading environment: Perfect capture (for NCFSK, CPSK modulations), fixed channel recovery time ($\gamma = 3$ slots).

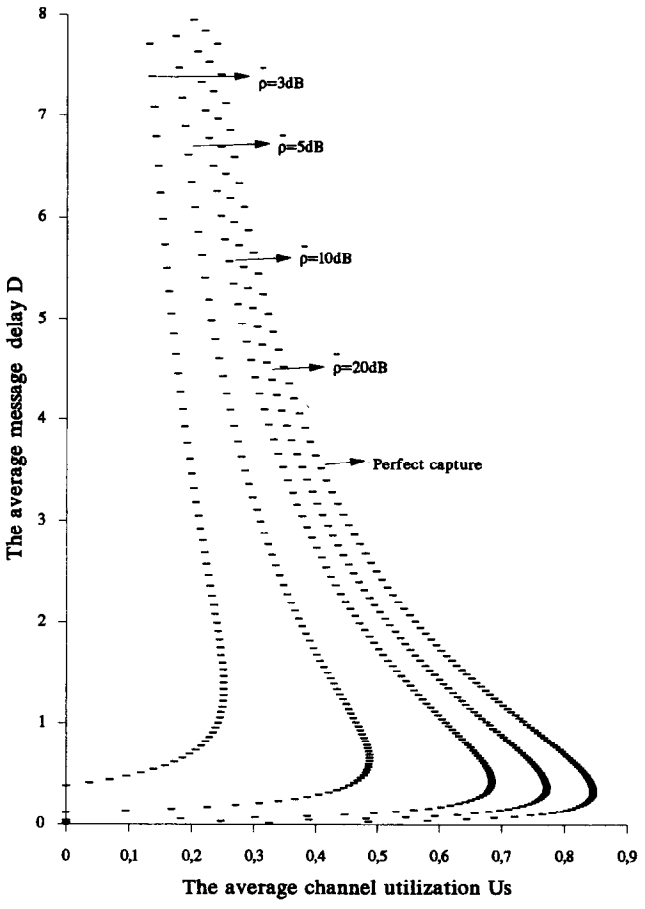


Fig. 27. The average message delay D versus the channel utilization U_s in a slow fading environment: NCFSK modulation, 48-bit precursor, fixed packet length ($\theta = 20$ slots), fixed channel recovery time ($\gamma = 3$ slots), various average SNR values.

computational approach may use the theory on parallel algorithms for solving linear systems with large but non-symmetric matrices. This area is still under investigation, and some results will be presented in future work.

We have considered the most important system parameters, namely the packet length, channel recovery time, the attempt probability and the power capture effect. It is shown that when the traffic is low, such a system exhibits a stable operation with better performance in the throughput–delay trade-off than CSMA with power capture (see [9]). By varying one of the above four parameters at a time, while keeping the rest fixed, we have established the following significant results:

(a) The channel throughput increases as the power capture ratio takes smaller values (see Appendix B).
(b) The average time a cycle spends in a successful transmission increases with the packet length θ, at the expense of the average time of a cycle the channel is idle or busy for due to collisions which result in no successful transmissions.
(c) The average time the channel needs to recover from an unsuccessful transmission is important. The less the channel recovery time γ is, the better the throughput results. However, as the traffic increases, the influence of γ almost disappears.
(d) For small values of the attempt rate δ, packets seem to suffer by excessive delays. Moderate values of δ result in an increase in the channel throughput. However, for suitable values of the system parameters and independently of the attempt rate δ, the channel appears to have a good operational point.

To clarify our results in practice, the infinite population model has also been analysed and discussed. The semi-Markov model constructed has been derived using a non-synchronous time renewal analysis, and the results found are generally consistent with those derived in the synchronous time finite population model. Thus, as in the previous model, the proposed hybrid protocol has a similar performance (when capture is perfect) to the classic CSMA/CD protocol, namely, a performance which is better than the CSMA or the ALOHA algorithms with capture. In addition, the channel shows a bistable behaviour, a well known phenomenon in contention protocols. In particular, the channel gives acceptable results in respect of the throughput–delay trade-off for sufficiently long periods, provided that the offered load is

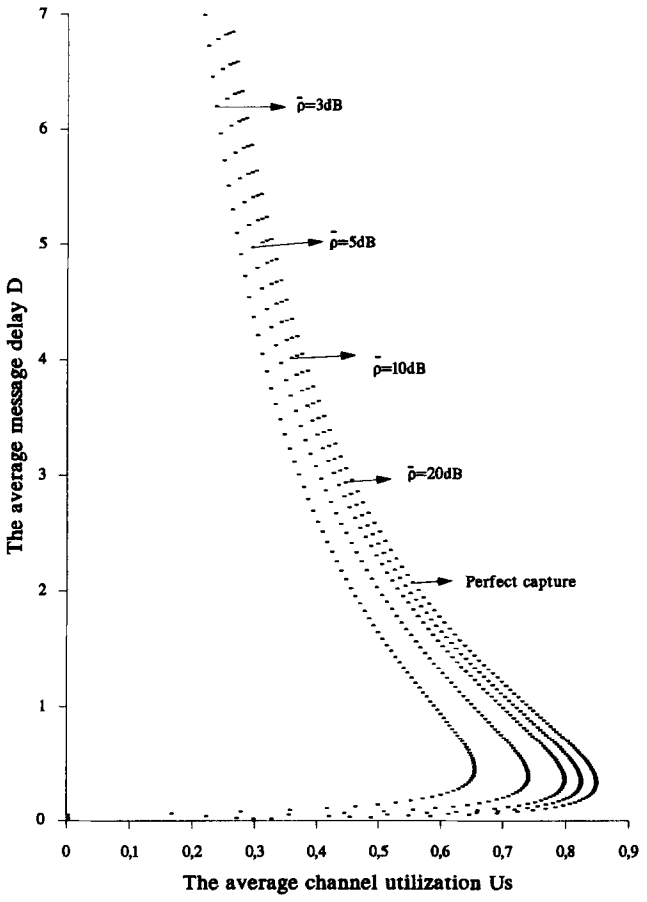


Fig. 28. The average message delay D versus the average channel utilization U_s in a slow fading environment: CPSK modulation, 48-bit precursor, fixed packet length ($\theta = 20$ slots), fixed channel recovery time ($\gamma = 3$ slots), various average SNR values.

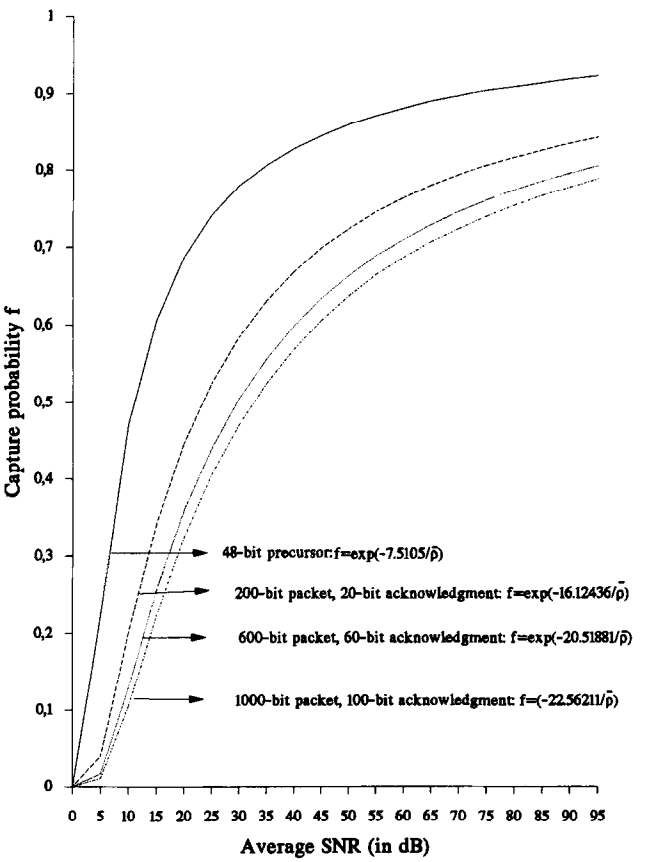


Fig. 29. Effects of average SNR on capture probability f in a slow fading environment: NCFSK modulation, different packet sizes and overheads.

low. However, as the traffic increases, the throughput drops to zero, resulting in excessive delays.

Various performance measures have been examined, and the results found indicate some significant aspects of the protocol. Interest has been given in practical systems for which the packet length θ is around 20 slots, and the channel recovery time γ is around 3 slots. The most interesting results established for various average SNR values are as follows:

(a) The average channel utilization $\bar{U}_s$ increases with the packet length in the expense of the average time the channel is busy, either due to collisions or an undetected precursor.
(b) The maximum achievable throughput S is obtained for the same value of the traffic rate G, independently of the packet length θ.
(c) The capture effect due to slow fading is significant. Higher values of the average SNR (at about 30 dB for the CPSK and 50 dB for the NCFSK) leads to the system approaching the nonfading case. Thus, when capture is perfect a similar performance to the conventional CSMA/CD protocol is obtained. However, when capture is present, a considerably better performance than CSMA ($\alpha = 0.01$) and BTMA may be achieved.
(d) The channel shows a bistable behaviour, unless the offered traffic is extremely low.

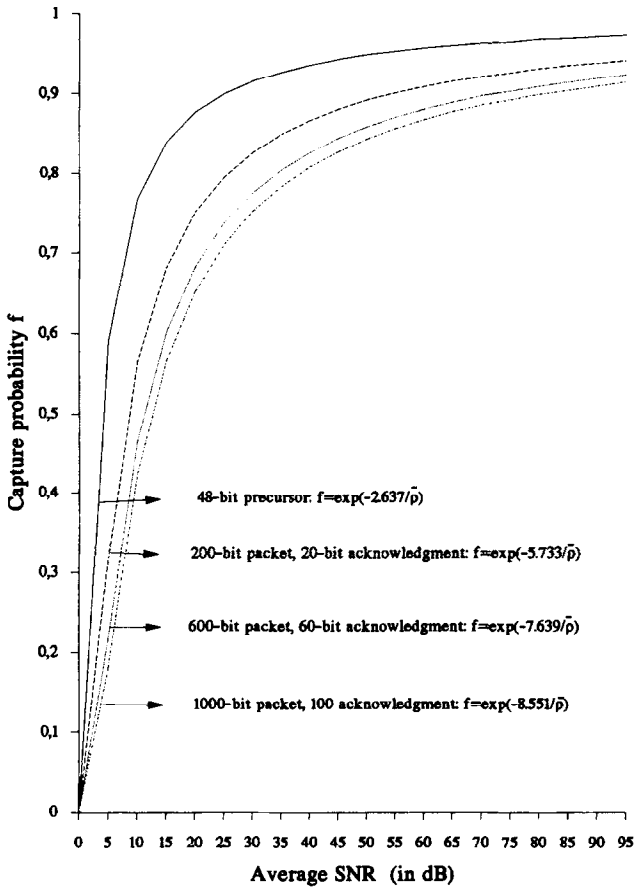


Fig. 30. Effect of average SNR on capture probability f in a slow fading environment: CPSK modulation, different packet sizes and overheads.

Acknowledgements

The author would like to thank Professor D.G. Smith and J. Smith of Strathclyde University, Department of EEE, for their suggestions and strong support during the initial steps of this work. He is also grateful to the General Editor of *Computer Communications* for his suggestions and help.

Appendix A

The model we are studying assumes transmission over a slow fading channel which results in constant signal-to-noise ratio during transmission of any one block. However, the signal strength may vary or fade from block to block. Because of the modulation scheme, independent bit errors are produced with probability $\epsilon(\rho)$ given by:

$$\epsilon(\rho) = (1/2)\exp(-\alpha\rho)$$

$$\begin{cases} \alpha = 1/2 & \text{for NCFSK modulation scheme} \\ \alpha = 1 & \text{for DPSK modulation scheme} \end{cases}$$

and

$$\epsilon(\rho) = (1/2)\,\mathrm{erfc}\,(\sqrt{\alpha\rho})$$

$$\begin{cases} \alpha = 1/2 & \text{for CFSK modulation scheme} \\ \alpha = 1 & \text{for CPSK modulation scheme} \end{cases}$$

where ρ is the instantaneous signal-to-noise power ratio and $\mathrm{erfc}\,(*)$ is the complementary error function defined as $\mathrm{erfc}\,(x) = (1/2\pi)\int_x^\infty \exp(-t^2)\mathrm{d}t$. In the above, NCFSK and CFSK present the non-coherent and coherent frequency-shift-keying modulation, respectively, while DPSK and CPSK present the deferentially and coherent phase-shift-keying modulation, respectively. The probability of block error P_{be} is defined as the probability that a packet is not accepted by the base station due to noise errors, and therefore:

$$P_{be} = \Pr(\text{at least one error})$$
$$= 1 - \Pr(\text{all bits received correctly})$$
$$= 1 - [1 - \epsilon(\rho)]^M,$$

where M is the block length in bits ($M = 48$; the precursor length in our hybrid model). Note that the capture probability f is defined as $f := 1 - P_{be}$. We assume that the system uses power level control such that all users have the same average received power, and thus all users are transmitting at the same power (there are no power

variations with distance). The capture in such a system is due solely to Rayleigh fading.

Assuming constant ρ over the block length, the probability of block error averaged over the fading signal level is given by:

$$\bar{P}_{be} = \int_0^\infty \{1 - [1 - \epsilon(\rho)]^M\}\phi(\rho)\,d\rho$$

where $\phi(\rho)$ is the probability density function of ρ. Further, it is assumed that phase variations are slow enough to be not significant to either a coherent or non-coherent modulation/demodulation scheme we are considering. For the case of Rayleigh fading, the signal power has an exponential probability density function. Hence, the signal-to-noise ratio, say ρ, is also exponentially distributed, namely:

$$\phi(\rho) = (1/\bar{\rho})\exp(-\rho/\bar{\rho}) \quad \rho \geq 0$$

where $\bar{\rho}$ is the average value of ρ.

To show the effect of $\bar{\rho}$ on the capture probability f, we make use of the method already developed in [11]. This method is applicable in both the NCFSK and CPSK modulation schemes, but is not applicable in the case of DPSK because of the dependent nature of the occurrence of errors. However, the analysis does provide an upper bound on DPSK performance.

A.1. Non-coherent signalling: NCFSK modulation

In the case of NCFSK modulation, the probability $\bar{P}_{be}$ is obtained using the bit error probability function $\epsilon(\rho)$ with $\alpha = 1/2$. Thus:

$$\begin{aligned}\bar{P}_{be} &= (1/\bar{\rho})\int_0^\infty \{1 - [1 - (1/2)\exp(-\rho/2)]^M\} \\ &\quad \times \exp(-\rho/\bar{\rho})\,d\rho \\ &= 1 - (1/2)^{-2/\bar{\rho}}(2/\bar{\rho})B_{1/2}(2/\bar{\rho}, M+1) \\ &= 1 - {}_2F_1(2/\bar{\rho}, -M; 1 + 2/\bar{\rho}; 1/2),\end{aligned}$$

where $B_x(a, b)$ is the incomplete beta function and ${}_2F_1(a, b; c; x)$ is the hyper-geometric function. Exact values of $\bar{P}_{be}$ may be obtained only for small values of M using the series representation of the hyper-geometric function. It is interesting to note that for $M = 1$ the block error probability becomes a bit error probability, and using the series representation of the above hyper-geometric function, the average probability of the bit error for NCFSK in Rayleigh fading may be obtained. However, both the incomplete beta and hyper-geometric functions can be evaluated by numerical methods. For large values of M and general values of $\bar{\rho}$ the step function approximation for $\bar{P}_{be}$ yields the following asymptotic formula:

$$\bar{P}_{be} = F(\rho_0) = 1 - \exp(-\rho_0/\bar{\rho}).$$

In the above, $\rho_0 = 2\ln M - 0.23187$ and $F(*)$ is the cumulative function of ρ evaluated at the point $\rho = \rho_0$. Thus, the asymptotic value of $\bar{P}_{be}$ for large $\bar{\rho}$, is $\lim_{\bar{\rho}\to\infty} \bar{P}_{be} = 0$.

As an example of interest, we show in Fig. 29 the effect of the average signal-to-noise ratio $\bar{\rho}$ on the capture probability f for different packet sizes. This is of particular importance, since the message delay due to channel error is usually minimized by optimizing the packet length for a certain set of system parameters. However, since the system parameters vary from one location to the other, a variable packet length is required to keep the delay at its minimum value. Note that a packet size between a few hundred bits to 1000 bits seems a satisfactory choice for most cases of computer communication requirements, including real-time computer speech. Fig. 29 also shows the advantage of using the 48-bit precursor as an access mechanism. For the same value of $\bar{\rho}$, the capture probabilities obtained using large packet lengths are much less compared with the corresponding capture probability obtained using the 48-bit precursor of our model. Remember that if the precursor is received correctly, the transmission is assumed to be successful due to the large amount of redundant information in the information part of the message (see Section 2.1).

A.2. Coherent signalling: CPSK modulation

Using the step function approximation for large M, it has been shown [11] that one has to evaluate the integral numerically:

$$\begin{aligned}\rho_0 &= \int_0^\infty [1 - (1 - (1/2)\,\mathrm{erfc}(\sqrt{\alpha\rho}))^M]\,d\rho \\ &= (1/\alpha)\int_0^\infty [1 - (1 - (1/2)\,\mathrm{erfc}(\sqrt{x}))^M]\,dx.\end{aligned}$$

Note that α takes two different values, according to the modulation scheme adopted. As an example, we worked out the case of CPSK modulation ($\alpha = 1$) following the same steps as in the case of NCFSK modulation. Thus, Fig. 30 shows the effect of the average signal-to-noise ratio $\bar{\rho}$ on the capture probability f for different packet sizes. Similar conclusions to the non-coherent case may be drawn. However, one should note that the capture probabilities are higher than the corresponding ones found in the NCFSK modulation scheme.

Appendix B. Capture probabilities

There are several methods to calculate the capture probabilities f_i; $i = 0, 1, 2, \ldots, N$. The most efficient method requires explicit knowledge of the area covered by the radio network. Using topographical databases, it is possible to create maps showing the mean levels of energy produced at the base station (receiver) by a mobile transmitting from different regions of the area covered.

Information from statistical surveys is also used to estimate the effect of the different attenuations, like the path loss and shadow, or Rayleigh fading. The above information is combined with the relative numbers of mobiles expected in the respective regions to give the probability distribution $F(x) = \Pr[E < x]$, where E is a random variable showing the level of power caused by a mobile.

Namislo [16] indicated that, even in the most simple cases, it is extremely difficult to calculate the exact values of the capture probabilities f_i; $i = 0, 1, 2, \ldots, N$. Alternatively, to calculate them by means of simulation, he assumed that users are spread uniformly over the whole region, and that their distances from the receiver are random variables with a common distribution function given by $G(x) = x^2$; $x \in (0, 1)$. If the level of power E_j caused by a transmitting mobile at the receiver due to path loss is proportional to $1/x^m$, the base station will be able to capture this transmission if $E_j / \sum_{i=1}^{N} E_i \geq c$; $i \neq j$, for some constant $c \geq 1$, and the probabilities f_i can be determined using Monte Carlo simulation. Note that the exponent m is between 2 and about 5, depending on the geography of the region.

The simulation experiment assumes that k identical ready mobiles are uniformly spread in a circular area of radius 1, which approximates the cell, with the base station located at the centre of the circle. It is also assumed that shadow or Rayleigh fading effects are neglected ($m = 2$). Then:

- Set $f_0 = 0.0$ and $f_1 = 1.0$
- For each value of $k = 2, 3, \ldots, N$ representing the number of ready mobiles in the cell:
  - Compute the Cartesian distances d_i and the level of power $E_i = 1/d_i^2$; $i = 1, 2, \ldots, k$ of all the mobiles in the cell from the base station.
  - Check if $\max_i\{E_i\} \geq c \sum_{j=1}^{k} E_j$; $j \neq i$, $i = 1, 2, \ldots, k$.
- Repeat the algorithm.

Then, the capture probabilities f_i are given as the relative number of times the above condition is satisfied. Note that higher SNR values correspond to a higher capture ratio c, which results in a decrease of the power capture probability f_i for every $i = 2, 3, \ldots k$ (i.e. with $c = 2$, 10, 100 the corresponding SNR values are 3 dB, 10 dB and 20 dB, respectively).

References

[1] E.J. Coyle and B. Liu, A matrix representation of CSMA/CD networks, IEEE Trans. Commun., 33(1) (January 1985) 53–64.

[2] S.L. Beuerman and E.J. Coyle, The delay characteristics of CSMA/CD networks, IEEE Trans. Commun., 36(5) (May 1988) 553–563.

[3] K. Mukumoto and A. Fukuda, Idle signal multiple-access (ICMA) scheme for terrestrial packet radio networks, Electr. Commun. Japan, 64-B(10) (1981).

[4] A. Murase and K. Imamura, Idle-signal casting with collision detection (ICMA-CD) for land mobile radio, IEEE Trans. Veh. Technol. 36(2) (1987) 45–50.

[5] F. Adachi, K. Ohno and M. Kitagawa, Performance analysis of ICMA/CD multiple access for a packet mobile radio, Electr. Lett., 24(8) (1988) 469–471.

[6] K. Zdunek, D. Ucci and J. Locicero, Throughput of non persistent inhibit sense multiple access with capture, Electr. Lett., 25(1) (1989) 30–31.

[7] F.A. Tobagi and L. Kleinrock, Packet switching in radio channels: Part II — the terminal hidden problem in carrier sense multiple access and the busy tone solution, IEEE Trans. Commun., 23 (December 1975) 1417–1433.

[8] F.A. Tobagi and V.B. Hunt, Performance analysis of carrier sense multiple access with collision detection, Computer Networks, 4 (October–November 1980) 245–259.

[9] T. Tsiligirides and D.G. Smith, Analysis of a p-persistent CSMA packetized cellular network with capture phenomena, Computer Commun., 14(2) (March 1989) 94–104.

[10] M.S. Appleby and J. Garrett, The Cellnet cellular radio network: Part I — General system overview, Br. Telecomm. Eng., 4 (July 1985) 62–69.

[11] J.A. Roberts and T.J. Haley, Packet radio performance over slow Rayleigh fading channels, IEEE Trans. Commun., 28 (February 1980) 279–286.

[12] J.J. Metzner, On improving utilization in ALOHA networks, IEEE Trans. Commun., 24 (April 1976) 447–448.

[13] F. Kuperus and J. Arnback, Packet radio in a Rayleigh channel, Electr. Lett., 18(12) (June 1982) 506–507.

[14] J. Arnback and W. Blitterswijk, Capacity of slotted ALOHA in Rayleigh-fading channels, IEEE J. Select. Areas Commun., 5 (February 1987) 261–269.

[15] C. Van Der Plas and J.-P.M.G. Linnartz, Stability of mobile slotted ALOHA network with Rayleigh fading, shadowing, and near-far effect, IEEE Trans. Veh. Technol., 39(4) (November 1991) 359–366.

[16] C. Namislo, Analysis of mobile radio slotted ALOHA networks, IEEE J. Select. Areas Commun., 2(4) (July 1984) 583–588.

[17] I.M.I. Habbab, M. Kanehrad and C.-E.E. Sundberg, ALOHA with capture over slow and fast fading radio channels with coding diversity, IEEE J. Select. Areas Commun., 7(1) (January 1989) 79–88.

[18] S.S. Lam, Packet switching in a multi-access broadcast channel with application to satellite communications in a computer network, PhD Dissertation (Technical Report UCLA-ENG-7429), Department of Computer Science, University of California, Los Angeles, CA (April 1974).

[19] I. Gladwell and M. Paprzycki, Parallel solution of almost block diagonal systems on the GRAY Y-MP using level 3 BLAS, J. Computational and Appl. Math., 45 (1993) 181–189.

Theodore A. Tsiligirides received his BSc in mathematics from the University of Athens, Greece, his MSc in probability and statistics from Manchester-Sheffield University, UK, and his PhD in telecoms from the University of Strathclyde, Scotland. In 1981 he joined the Mathematics and Computer Science Division of the Agricultural University of Athens, first as a research/teaching assistant, and then as a lecturer. From 1993 he has worked as an Assistant Professor in the Informatics Laboratory of the same division. His research interest include modelling and traffic control of high speed networks, access methods in mobile and broadband networks, queueing networks and parallel algorithms.