WIRELESS COMMUNICATIONS AND MOBILE COMPUTING
*Wirel. Commun. Mob. Comput.* 2004; **4**:1–15 (DOI: 10.1002/wcm.239)

# Adaptive admission/congestion control policies for CDMA-based wireless internet

U. A. Baroudi[1,*,†] and A. K. Elhakeem[2]

[1]*Computer Engineering, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia*
[2]*Electrical and Computer Engineering, Concordia University, Montreal, Quebec, Canada H3G 1M8*

## Summary

Radio resource management (RRM) is vital for the next generation wireless networks. RRM comprises many functionalities and this paper focuses on the investigation of the performance of several adaptive call admission/congestion control policies based on a window-measurement estimation of the status of the buffer at the base station under the hybrid TDMA/CDMA access scheme. In our study, we interrelate the physical limitations of the base stations (i.e. the number of transmission and reception modems), call and burst level traffic, instantaneous buffer conditions and end-to-end bit error performance in one queuing problem. Subsequently, a window-measurement estimator is implemented to estimate the likelihood of buffer congestion at the base station. Accordingly, the traffic loads shall be controlled. We use event-driven simulation to simulate the multimedia integrated CDMA networks where heterogeneous traffic users are multiplexed into a simple TDMA frames. The simulation results show outstanding performance of the proposed call admission/congestion control policies in guaranteeing QoS requirements. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: RRM; call admission control; CDMA; multiuser detection; traffic control; wireless internet; next generation wireless networks

## 1. Introduction

With the rapid convergence of wireless networks and internet, it is expected that heterogeneous traffic users will traverse these merging networks. Further, the basic resources that these users will compete for are the radio channels. Moreover, as each of these users may request different quality of service (QoS), the issue of managing these resources such that QoS requirements of each user are guaranteed as well as the overall network utilization is maximized is very essential issue and that is the scope of radio resource management (RRM) [1]. The two main functions of RRM are the call admission control (CAC) and the dynamic bandwidth allocation (DBA) policies. Typically, the CAC is invoked only at the registration phase of the incoming call, while the DBA is concerned with the call for its duration until termination. Hence, the CAC criteria are tested only once whereas the DBA procedures require continuous monitoring of the ongoing call parameters and the overall state of the radio access network. The DBA procedures retain an up-to-date reading of the relevant network variables. Adaptive call admission/congestion control policies based on a window-measurement estimating the network activity should significantly help accommodating these services and satisfying their QoS requirements.

*Correspondence to: Uthman A. Baroudi, Computer Engineering, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia.
†E-mail: ubaroudi@ccse.kfupm.edu.sa

Recently, researchers and network service industries have become very active in investigating this problem and finding solutions and answers for it. In this work, several policies have been developed to tackle the above problem and provide a detailed RRM scheme and QoS control function. Looking at the existing literature, many papers addressing this issue have been published (e.g. see Refs. [2–5]). Yet, most of the existing works consider either the uplink or the downlink channel in evaluating the CDMA system under the proposed admission/congestion policies. Further, these studies do not include the error performance in the queuing modeling of the CDMA networks. Nevertheless, the inclusion of the channel performance in the queuing modeling would give more realistic figures of network performance as well as guarantees of the required QoS of the networks.

Our focus in this paper is to simulate and investigate the performance of several adaptive call admission/congestion policies based on a window-measurement estimation of the status of the buffer occupancy at the base station under the hybrid TDMA/CDMA access scheme proposed by the authors in Reference [6]. It is worth noting that these policies can easily lend its self to other schemes or network platforms such as UMTS networks. Further, as the proposed protocols have shown an outstanding performance in guaranteeing QoS parameters and in particular the low error rate over the wireless link, this would significantly help improve the TCP/IP traffic performance by minimizing the fault retransmissions due to data loss (that otherwise, it may be interpreted by the TCP/IP as a network congestion) which is now minimized.

In our study, we interrelate the physical limitations of the base stations (i.e. the number of transmission and reception modems), call and burst level traffic, instantaneous buffer conditions and end-to-end bit error performance in one queuing problem. Subsequently, a window-measurement estimator is implemented to estimate the likelihood of buffer congestion at the base station. Accordingly, the traffic loads shall be controlled. We use event-simulation to simulate the multimedia integrated CDMA networks where heterogeneous traffic users are multiplexed into a simple TDMA frames.

This work differs from the previous work done in that the essence of this work is to introduce an interaction between the physical layer and higher layers, enabling a more practical utilization of multiuser detection and supporting services with different QoS parameters.

The rest of the paper is organized as follows. Section 2 presents the description of the hybrid TDMA/CDMA access protocol. The proposed admission/congestion control policies are discussed in Section 3. Then Sections 4 and 5 discuss in detail the simulation assumptions, parameters and the whole course of simulation of the overall wireless network under consideration. In Section 6, we present and discuss the results and the performance of the proposed policies followed by concluding remarks in Section 7.

## 2. Multiple Access Protocol

In this work, the hybrid TDMA/CDMA is adopted as the access protocol for the uplink traffic where the time domain is divided into frames and each frame is composed of just two time slots (i.e. $T_{s1}$, $T_{s2}$). However, the detection strategy at the base station (BS) for the population of each time slot is different. The conventional receiver (i.e. single-user DS-CDMA receiver) is employed for the demodulation of signals received during $T_{s2}$, while the decorrelator multiuser receiver shall be employed for the users of $T_{s1}$ (i.e. stream traffic). Interested readers may consult Reference [6] for a more detailed discussion of the TDMA/CDMA protocol.

The flow of traffic to each time slot is controlled by the traffic characteristics of each user, its QoS parameters and the detection strategy applied in that time slot. Therefore, the following classification for the traffic population is considered. The traffic is divided into two categories: stream traffic and interactive traffic. Each category has classes of traffic which have common traffic characteristics. The stream traffic includes classes of traffic that have high transmission activities. Nevertheless, it includes different types of users where each one has its own transmission characteristics (i.e. average rate, peak rate, activity of the transmission etc.). On the other hand, the interactive category of traffic comprises all other traffic that are not included in the first category (e.g. signaling, high bursty users) as well as the excess traffic of the stream traffic (i.e. variable bit rate components) which will be explained in the next subsection.

### 2.1. Traffic Flow Control Analysis

Considering the heterogeneous traffic of the future communication networks, and for analysis convenience, various traffic classes should represent their bit rates as $n$ multiple of a basic rate $R_b$; or $1/n$ multiple of $R_b$; $n$ is an integer. For practical rates, users of rate $R_b/n$, the generated packets will be buffered till their gross rate matches with $R_b$. To efficiently utilize the bandwidth at our disposal, and at the same time invest the capability of the CDMA technique, it is suggested that each class of users has its own basic rate, which is chosen according to the QoS requirements. For example, a class of users that can tolerate a $10^{-3}$ bit error rate (BER) might not be effective for their transmission rates to be subdivided by the same reference rate (i.e. basic rate) of another class of traffic that its BER requirement is much smaller than $10^{-3}$. For high rate classes (say class $j$, where the bit rate is $R_j$), when a user needs to transmit with a rate greater than $R_b$, the high-bit stream shall be converted into $n$ low-bit basic streams, where $n = R_j/R_b$. Each new low-bit basic stream has a bit rate equal to $R_b$. Consequently, the low-bit basic stream is packetized into fixed size packets. This strategy of subdividing the high rate users into low-rate streams and then applying the CDMA technique is called multicode CDMA (MC-CDMA).

Figure 1 illustrates the simple TDMA/CDMA access scheme. Basically, $T_{s1}$ is dedicated to the stream traffic users, while, $T_{s2}$ is dedicated to the interactive traffic users. Mobile units are assumed to be synchronized to follow the proposed access scheme. Practically, it is visible and is very similar to what is done in GSM networks. During the signaling period (where the user tries to communicate with BS to get an admission to the system), the stream traffic user $j$ negotiates with BS about (among other things) the average transmission bit rate $R^{s}_{avr(j)}$ that the user shall stick to while it is using $T_{s1}$, though the actual bit rate might be less in some cases. In the case that the instantaneous rate is less than $R^{s}_{avr(j)}$, the user has to stick to the agreement and generate dummy packets. Of course, these dummy packets will affect the bandwidth utilization. However, since this category of users has high transmission activity, it is expected that its role will be minimal.

Now, it is clear that this fixed rate is translated into a fixed number of received packets (i.e. $N^{s}_{avr}$ as in Equation (1)) at the receiver side where a multiuser detector is used. Thus, this simple traffic flow control enables a practical and effective implementation of multiuser detection strategy at the receiver of the stream traffic population since the numbers and identities of overlapping CDMA packets are perfectly known throughout the user call.

$$N^{s}_{avr} = \sum_{j=1}^{N^{s}} N^{s}_{avr(j)} = \sum_{j=1}^{N^{s}} \hat{\xi}^{s}_{j} n^{s}_{j} \qquad (1)$$

where $N^{s}_{avr}$ is the average number of packets emitted from all active stream users to $T_{s1}$, $N^{s}$ is the number of classes of users in the stream traffic category, $n^{s}_{j}$ is the number of active stream users belonging to class $j$, and $\hat{\xi}^{s}_{j} = R^{s}_{avr(j)}/R_b$ is the ratio of a user's average bit-rate to the basic bit-rate. Now, if the user needs to send information using a bit rate higher than the average bit rate (i.e. $R^{s}_{i} > R^{s}_{avr(j)}$), then the excess packets (variable bit rate components) should be queued and directed to the other time slot (i.e. $T_{s2}$) where these signals along with interactive traffic flow shall be detected by the conventional receiver (i.e. single user). Therefore, the assignment of slots to a part or all of user traffic is done *a priori* through an agreement between the base station and the user.

$$N^{s}_{excess} = N^{s}_{inst} - N^{s}_{avr} \qquad (2)$$

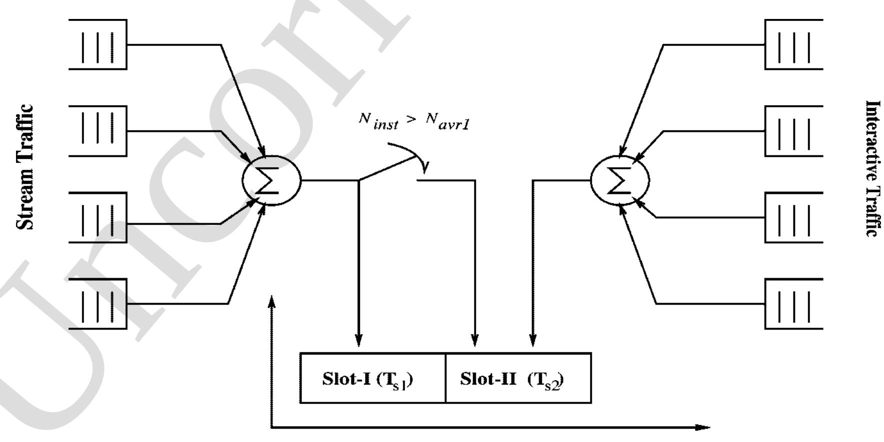



Fig. 1. The TDMA/CDMA multiple access scheme.

where $N^{s}_{inst}$ is the instantaneous number of packets emitted from active stream users.

## 3. Admission/Congestion Policies

Figure 2 shows an ensemble of network activity at BS. We denote the case when the number of packets in the buffer exceeds certain threshold $th_B$ by an upward arrow. Otherwise, it is denoted by a downward arrow and this is what we call the buffer is in 'good' condition. This randomness is due to many factors, such as the traffic burstiness at the packet level, the erroneous channels etc. These activities also vary from window to window due to the above reasons as well as the call activity of each user. It is obvious that these activities could give us valuable information about how the whole network is working. Consequently, traffic loads might be controlled in such a way that the QoS requirements will be met.

Define $S$ as the likelihood that BS buffer is not congested. Consequently, the traffic load must be adapted according to $S$ as shown in Equation (3).

$$\rho_a = \rho \times S \tag{3}$$

Where $\rho_a$ is the average accepted traffic, and $\rho$ is the total offered traffic load (Erlangs). According to Equation (3), $S$ is actually the average throughput of the system. For example, if $S = 0.9$, then only 90% of the traffic should be admitted and the rest must be rejected. By controlling the traffic load as in Equation (3), the following QoS parameters, namely, packet blocking probability, call establishment delay and the packet error will be under control. Therefore, the second phase of the proposed policy (congestion) has two steps. First, the likelihood of buffer congestion ($S$) is estimated. Second, the offered traffic load is adapted accordingly. This adaptation of the offered traffic can be implemented in different ways. For example, the transmission rates of all currently admitted users can be varied such that the accepted traffic load matches with $S$ and the number of admitted calls is kept without change. On the other hand, we can choose to take an opposite strategy that is to keep the transmission rates as they are and instead block some calls following certain policy.

Now, we are in a position to state our general call admission policy; i.e. the new admission rule states that, 'The new call is admitted once the total number of low-bit streams corresponding to the active calls is less than the number of servers and the buffer is not congested as in Equation (??[Q1])'. Now, how are we going to define the term 'congested buffer'? This is what the next subsection shall explain. Q1

### 3.1. Buffer Congestion Definitions

Here, we shall discuss three definitions for buffer congestion state and accordingly we end with three admission/congestion policies.

In the first policy (i.e. Policy-I), we have adopted a stringent admission policy that requires the buffer to stay in the 'good' condition for $q$ consecutive

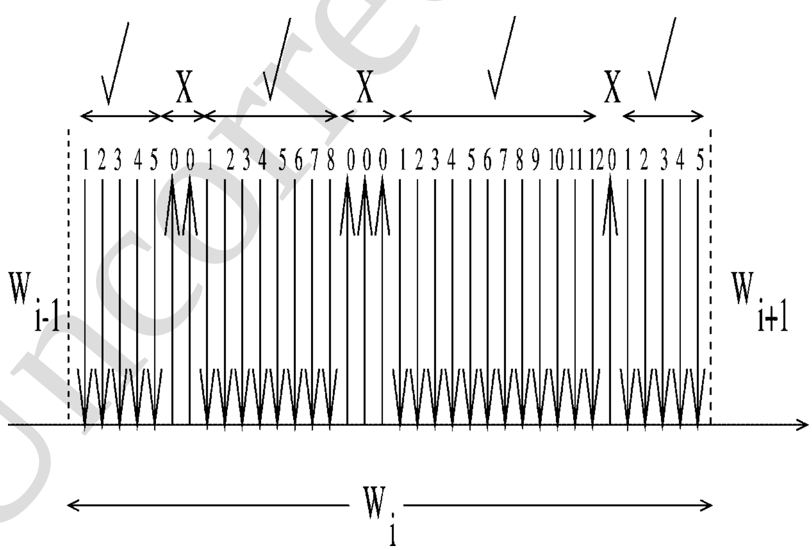



Fig. 2. Illustration of buffer activities at the base stations (BS); window size = 36 packets.

packet-time units. For instance, applying this policy for $W_i$ in Figure 2, we find that frequency of occurrence of this event is

$$S = \frac{\text{Number of events}}{\text{Maximum number of possible events}} \quad (4)$$

where the maximum number of possible events for a window of size $W$ and $q$ consecutive packets is given by

$$\Psi(W, q) = \sum_{i=0}^{W-q+1} \left\lfloor \frac{W-i}{q} \right\rfloor \quad (5)$$

where $\lfloor \cdot \rfloor$ is the greatest integer. Then, considering the example at hand (illustrated in Figure 2), we can observe several small windows where the buffer is not congested for at least $q$ consecutive packets. The first small window has a size of five packet-time units and the buffer is during this small period not congested. For this example, this small window corresponds to two events and so on.

$$S = \frac{2\sum_{i=0}^{2}\lfloor (5-i)/4 \rfloor + \sum_{i=0}^{5}\lfloor (8-i)/4 \rfloor + \sum_{i=0}^{12}\lfloor (12-i)/4 \rfloor}{\sum_{i=0}^{33}\lfloor (36-i)/4 \rfloor} = 0.163 \quad (6)$$

This above policy can be modified to obtain a moderate admission policy (i.e. Policy-II) whose performance lies between the above policies. The modification is as follows. We just count the events where the buffer occupancy exceeds certain threshold. Then, we find how frequent this event happens throughout the whole measurement window, i.e.

$$S = 1 - \frac{\text{Number of event occurrences}}{\text{Measurement window size}} \quad (7)$$

From Figure 2, we find for the example above, that $S = 1 - (6/36) = 0.833$, which means that Policy-II is a bit relax compared to Policy-I.

Thirdly, the admission of new users is dependent on the buffer condition at the end of each measurement window regardless of what was going on throughout the window. Of course, this policy is relax and weak. We call this policy the basic admission/congestion control policy or policy-III.

The question is now how these policies would be employed to control the flow of traffic such that the QoS requirement is guaranteed? This what the next section shall discuss.

### 3.2. Traffic Adaptation

Let $\rho_o^i$ be the total accepted traffic in window $i$ ($W_i$), i.e.

$$\rho_o^i = \rho_a^{i-1} + \rho_{\text{new}}^i \quad (8)$$

where $\rho_a^{i-1}$ is the accepted offered traffic arriving during $W_{i-1}$, and $\rho_{\text{new}}^i$ is the new offered traffic during $W_i$. According to the original proposed admission policy (see Section ??[Q2]), the accepted traffic will be a ratio of the offered traffic as follows, Q2

$$\rho_a^i = S \times \left(\rho_a^{i-1} + \rho_{\text{new}}^i\right) \quad (9)$$

Here, we modify the original admission policy such that a priority discipline is included. The old accepted calls shall be given higher priority than the new traffic and this is because of the fact that these excess packets belong to a call that is already active in the other slot. So, it is preferable to block a new call rather than terminating an active one. Therefore, it is easy to modify the admission policy accordingly and get the following relation:

$$\rho_a^i = S' \times \rho_{\text{new}}^i + S'' \times \rho_a^{i-1} \quad (10)$$

We set $S'' = 1$ and by solving Equations (9) and (10), we obtain the relation between the modified $S$ and the traffic intensity from each traffic category, i.e.

$$S' = (1+\nu) \times S - \nu \quad (11)$$

where $\nu$ is the ratio of the old accepted load to the new traffic load (i.e. $\nu = \rho_a^{i-1} = \rho_{\text{new}}^i$). Figure 3 shows the relation between the original and modified admission policies. We note from Figure 3 that there is a threshold for the values of $S'$ if it is exceeded, $S'$ is not valid any more. It is very easy to find this threshold in terms of $\nu$,

$$S \geq \frac{\nu}{1+\nu} \quad (12)$$

If the above condition is not satisfied, then all new traffic shall be blocked (i.e. $S' = 0$) as well as a

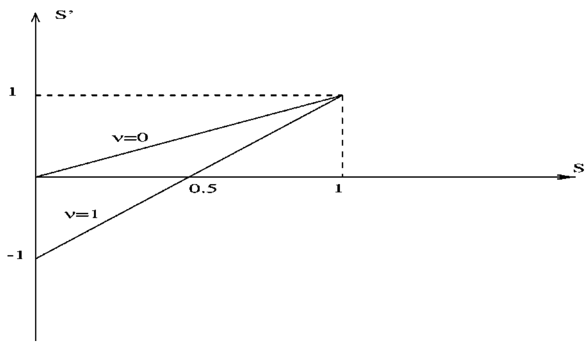



Fig. 3. The relation between the original and modified admission policies.

portion of the old accepted calls. Hence, we should solve Equation (10) for $S''$, i.e.

$$\rho_a^i = S'' \times \rho_a^{i-1} = S \times \left(\rho_a^{i-1} + \rho_{\mathrm{new}}^i\right) \qquad (13)$$

Therefore,

$$S'' = S \times \left(\frac{1+\nu}{\nu}\right) \qquad (14)$$

Here, we have a new performance measure that should be evaluated, that is the probability of call termination, i.e.

$$C_T = \frac{\text{number of terminated calls}}{\text{total accepted calls}} \qquad (15)$$

## 4. System Description

We use event-driven simulation to simulate the multimedia integrated CDMA networks where heterogeneous traffic users are multiplexed into a simple TDMA frames. Each frame is composed of two slots as explained before. In this simulation, our emphasis is on the network performance issues such as packet delay, packet delay jitter, packet losses, call blocking etc. Therefore, the only system parameters that are going to be simulated and randomly generated throughout the simulation program are those parameters which have direct relation to the network performance. For instance, packet error performance will not be simulated. However, the event-driven simulation is supplemented by analytical computation of packet error probability for uplink stream packets received by decorrelator receiver at BS. On the other hand, the packet error probabilities for the uplink slot-II users as well as downlink users from both slots are supplemented from computational analysis given by Reference [9].

The simulation concentrates on the following 'packet level' random events: (1) call initiation, (2) packet generation (i.e. $\mathcal{B}$), (3) successfully transmitted packets via uplink channel $\hat{\mathcal{B}}$, (4) BS operation (i.e. packet dropping, buffer overflow, busy servers on downlink channel) $\mathcal{D}$ and (5) successfully transmitted packets via downlink channel $\hat{\mathcal{D}}$.

A uniform random generator generates a new call request with a certain activity factor. For example, $\theta_{i,\mathrm{on}}^s$ is the activity factor for class $i$ belonging to stream traffic and $\theta_{j,\mathrm{on}}^{\mathrm{Int}}$ is the activity factor for class $j$ belonging to interactive traffic. The aggregated calls generated from all independent users yield a multinomial random process. We assume the call duration is geometrically distributed.

In some previous studies (e.g. Ref. [6]), traffic sources are assumed to be heterogeneous and each source is modeled as a single ON/OFF source for the sake of computational simplicity. Yet, this assumption does not take into consideration the correlation between generated packets. Hence, in this simulation, our traffic sources of interest are assumed to have variable bit rates. Each source has its own activity (i.e. $\alpha, \beta$) and by which its bit variations are controlled. Nevertheless, a constant bit rate user still can be modeled as variable bit rate source with $\alpha = 0$.

In evaluating the performance of the aforementioned admission/congestion policy, we consider the following system specifications. We assume that the uplink traffic and the downlink traffic are using different channels. As in WCDMA technology, we assume that the available bandwidth is 5 MHz. More, the basic data rate ($R_{\mathrm{b}}$) is 10 kbps. However, as a two-slot TDMA protocol is applied, the effective data rate over one slot is 20 kbps. Consequently, the source stream bits are modulated using BPSK-DS-CDMA with a processing gain of 255 using different Gold codes. The low-bit average stream traffic bits are transmitted via the uplink without encoding (multiuser decorrelator receivers are used at BS) while these bits shall be encoded at BS by convolutional codes for transmission on the downlink channel as will be explained later. Moreover, it is assumed that these average stream packets transmitted via the uplink will only be corrupted by AWGN ($E_{\mathrm{b}}/N_o = 12\,\mathrm{dB}$) besides the mutual interference from other intracell users.

On the other hand, the excess traffic bits as well as the interactive traffic bits are encoded by convolutional



Table I. Traffic characteristics.

| | Stream traffic (number of classes = 4) | Interactive traffic (number of classes = 2) |
|---|---|---|
| Class 1 (packet level) | $\xi_1^s = 3$, $\hat{\xi}_1^s = 2$ $N_1^s = 20$, $\theta_{1,\text{on}}^s = 1.0$ | $\xi_1^I = 7$ $N_1^{\text{Int}} = 10$, $\theta_{1,\text{on}}^{\text{Int}} = 0.5$ |
| Class 2 (packet level) | $\xi_2^s = 2$, $\hat{\xi}_2^s = 2$ $N_2^s = 25$, $\theta_{2,\text{on}}^s = 0.75$ | $\xi_2^I = 5$ $N_2^{\text{Int}} = 10$, $\theta_{1,\text{on}}^{\text{Int}} = 0.35$ |
| Class 3 (packet level) | $\xi_3^s = 4$, $\hat{\xi}_3^s = 3$ $N_3^s = 10$, $\theta_{3,\text{on}}^s = 0.90$ | |
| Class 4 (packet level) | $\xi_4^s = 5$, $\hat{\xi}_4^s = 4$ $N_4^s = 5$, $\theta_{4,\text{on}}^s = 0.80$ | |

codes. Further, the uplink bits ($E_\text{b}/N_o = 8.2\,\text{dB}$) are encoded by a convolutional code where $k = 9, r = 1/3$, while the downlink bits ($E_\text{b}/N_o = 8.45\,\text{dB}$) are encoded by another convolutional code with $k = 9, r = 1/2$. It is assumed that the packets on the uplink channel are received non-coherently at BS and the transmitted packets via the downlink channel are assumed to be coherently received by the mobile stations. Both links are assumed to be under the influence of two paths Rayleigh multipath fading. Table I summarizes the traffic characteristics and system parameters used in evaluating the proposed traffic control approaches.

## 5. Course of Simulation

At any time, initiated call requests are either accepted or rejected. The waiting users shall be served first if it is possible according to the buffer status. Then, we check if there are any new call requests that could be served. If the buffer status shows that the probability of buffer congestion is high, we have the following scenarios. First, we keep the priority for those users who are waiting (of course, for a certain 'time out' period) and block any new call requests. The second scenario is to add the new calls to the waiting list.

Each simulation run is composed of $K$ equal measurement windows. At the beginning of each window, new users can initiate admission requests to establish communication with the network. As soon as a user generates a call request and could not be admitted solely because of buffer condition, his call establishment delay (CED) counter is initialized and it shall be incremented through the simulation program until the call is accepted. The final value of this counter is an indication of the number of windows that the concerned user had to wait until its call is accepted. So, the average CED ($j$) for all system users during the $j$th simulation run (assuming $N$ active users, and $M$ accepted calls) is

$$\text{CED}(j) = \frac{\text{CED per user} \times N}{M} \times \text{WindowSize} \quad (16)$$

Then, the BS should check the capacity needed by each new user and compare it to the availability of servers on the uplink as well as the buffer condition at the BS. The peak rate of each user is used an indicator for the capacity needed by each user. Notice that this policy has to be followed in both slots. For example, assume that there are three new call requests belonging to the first class of stream traffic, and there are other two call requests from the first class of interactive traffic. Moreover, assume that queuing is not possible at the user side. Hence, it is essential to accommodate these users even at their maximum rates. Now, we have to calculate three parameters, namely, the maximum possible packets emitted from interactive users, average packets and maximum possible excess packets from stream users. Referring to Table I, we find the following: $\text{Av\_stream} = 3\xi_1^s = 21$, $\text{Ex\_stream} = 3(\xi_1^s - \hat{\xi}_1^s) = 9$, $\text{Intr} = 2\xi_1^I = 20$. Therefore, the expected traffic from these new users for slot-I is 21 packets while for slot-II is 29 packets. Now, assume that the demands of previously admitted calls are as follows: 20 servers on slot-I and 40 servers on slot-II. Now, if the stream traffic has higher priority than the interactive traffic, then, all new stream traffic can be accepted directly, while new interactive shall be rejected.

The second phase in the admission policy is to examine the buffer status. If it is found that the buffer is in a 'good' condition, then these new calls can be admitted to the network. Here, we define the 'good' condition as the state when the buffer occupancy is less than the maximum number of servers on the downlink.

On the other hand, if the first condition is guaranteed but the buffer is not in a 'good' condition (the second criterion), we suggest the following policy. This caller shall be registered in a temporary record (i.e. Reg-temp) such that he is granted higher priority (during the next window) than later callers and a 'time-wait' counter is initiated to control the period in which this user is granted higher priority. The motivation behind this policy is to trade off between the call blocking probability ($C_\text{b}(j)$) and the call establishment delay (CED ($j$)). Further, by accepting a longer delay to establish a call we could reduce $C_\text{b}$ probability by allowing the active users to statistically be multiplexed.

Now, it is clear that the call-block counter (i.e. $C_b$ will be incremented under three circumstances. Firstly, if this call cannot be severed due to a shortage in servers on the uplink, the call will be blocked. Failing to get access to the network (after a time-out period) due to buffer congestion is the second cause of call blocking. Thirdly, $C_b$ counter will also be updated under another circumstance which is due to network congestion control policy. Under this policy the network operator may choose to terminate the new calls after one measurement window because otherwise the QoS of currently carried calls would be degraded. Therefore, at the end of the simulation time (ST), the $C_b(j)$ is given by

$$C_b(j) = \frac{C_b}{\text{total call requests}} \tag{17}$$

It is important to measure how many packets the user would lose during the course of their call. The impairments of the uplink and downlink channels as well as the buffer congestion at the BS are the main causes for packet loss. In this simulation, we have defined the following loss probabilities. First, the probability of packet loss due to unsuccessful transmission via uplink channel during the $j$th simulation run (i.e. $l_{up}(j)$) is given by (assuming $PKT_{up}^{success}$ as the total packets received successfully at the BS and $PKT_{up}^{Tx}$ as the total packets transmitted via uplink channel)

$$l_{up}(j) = 1 - \frac{PKT_{up}^{success}}{PKT_{up}^{Tx}} \tag{18}$$

Second, the probability of packets loss due to unsuccessful transmission via downlink channel during the $j$th simulation run (i.e. $l_d(j)$) is given by (assuming $PKT_d^{success}$ as total packets received successfully by the other mobile user and $PKT_d^{Tx}$ as the total packets transmitted *via* downlink channel)

$$l_d(j) = 1 - \frac{PKT_d^{success}}{PKT_d^{Tx}} \tag{19}$$

Third, define $PKT_{drop}$ as the packets forced to drop because of the BS buffer overflow. Then, the packet losses due to buffer congestion during the $j$th simulation run (i.e. $\text{packet}_b(j)$) which has a crucial influence on the network performance and is given by

$$\text{packet}_b(j) = \frac{PKT_{drop}}{PKT_{up}^{success}} \tag{20}$$

From the user point of view, it is very vital to measure the average overall packet losses. In other words, we should find the end-to-end packet throughput, which can be defined as the ratio of total number of end-to-end successfully received packets to the total generated packets. Moreover, it is important from the network design point of view to measure the throughput ($j$) of the network for the $j$th simulation run which can be defined as follows:

$$\text{throughput } (j) = \frac{\text{total accepted traffic}}{\text{average offered traffic}} \tag{21}$$

where total accepted traffic is the actual number of packets generated by all accepted users. On the other hand, the offered traffic is the expected number of packets which would be generated by all accepted users if their calls had not been terminated due to the application of the admission/congestion policy. Finally, due to the impairments of the down link channel and the priority granted for interactive users, some stream packets (i.e. $PKT_{waiting}^{stream}$) might reside in the buffer for a certain period of time. Hence, these buffered packets should be counted and we can easily find the average packet delay $D(j)$ for the $j$th simulation rum, i.e.

$$D(j) = \frac{PKT_{waiting}^{stream}}{PKT_{buffered}^{stream}} \times \text{frame time} \tag{22}$$

More, this performance measure could also be used to control the flow of traffic to assure that the delay requirements by stream users are not violated.

## 6. Results and Discussion

Having configured the simulation setup for our problem, we have carried out many simulation experiments to explore the system performance of the considered network under wide range of traffic load and system parameters. Further, both policies mentioned in Section 3 will be compared when it is applicable to the basic admission policy (i.e. Policy-III). This policy is a primitive one and it just gives an indication about how the traffic burstiness at BS has been smoothened after a certain period of time (i.e. window). We shall consider the system performance under this policy as a reference for the other two policies when it is applicable.

The simulation results shall explore the network performance under different admission/congestion policies for fixed base station buffer size (i.e.



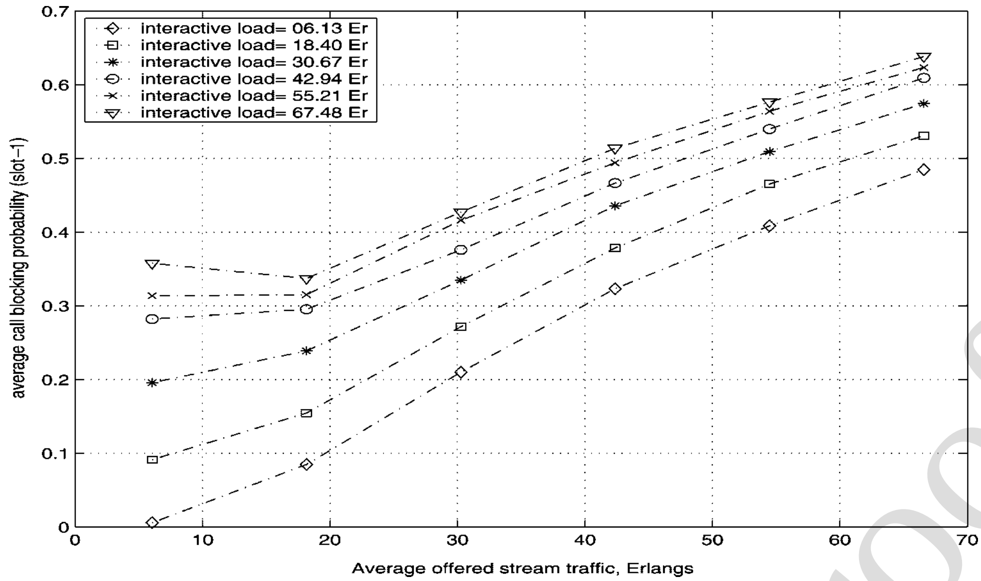



Fig. 4. The call blocking probability under Policy-I ($W = 150$, $q = 90$ and $B = 80$).

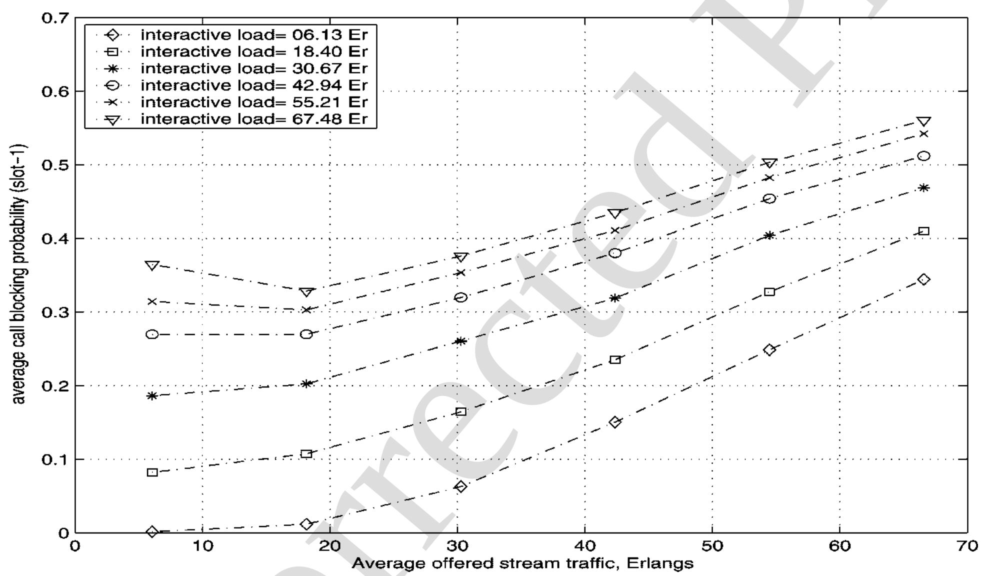



Fig. 5. The call blocking probability under Policy-II ($W = 150$, $q = 90$ and $B = 80$).

$B = 80$) and fixed measurement window size (i.e. $W = 150$).

Figures 4–6 show the average call blocking probability for stream traffic users under different interactive traffic loads for the three admission/congestion policies. The higher the interactive traffic, the higher is the stream call blocking probability. This is a natural result of the proposed hybrid MC-TDMA/CDMA, because many resources (i.e. receivers) are reserved for high bit rate interactive users even though their activities are low. Hence, many stream call requests will be blocked because the excess traffic could not be handled together with interactive traffic on slot-II. Therefore, to alleviate this problem, resources should not be reserved for high bursty users depending on worst case (i.e. peak transmission rate), but on other schemes such as average bit rate where smart queuing is provided at the user's site.

More, the other factor that increases the call blocking probability is the call termination policy. As we

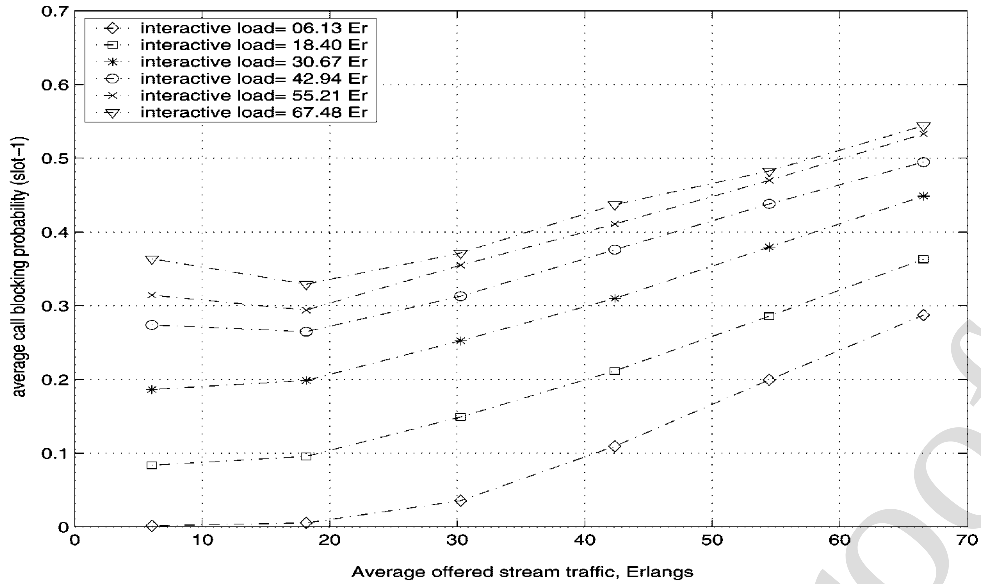



Fig. 6. The call blocking probability under Policy-III ($W = 150$, $q = 90$ and $B = 80$).

can observe from the system performance under Policy-I and Policy-II (see Figures 7 and 8), the call termination is high especially under low interactive traffic and high stream traffic. For instance, about 15% of accepted calls have been terminated under Policy-II (under $\rho_{\mathrm{Int}} = 6.13$ Erlang; $\rho_{\mathrm{s}} = 55.21$ Erlang), while more than 60% have been terminated under Policy-I. On the other hand, Policy-III causes no single call to be terminated.

Nevertheless, it is worth emphasizing that during the course of simulation, we assume that all currently admitted users are requesting the same QoS (i.e. buffer not congested for $q$ consecutive frames). In practice, however, it is not necessary that all users ask for the same QoS. Consequently, the call termination probability will be less. Alternatively, the rejection action can be relaxed by allowing the system performance to degrade for a longer period of time (e.g.

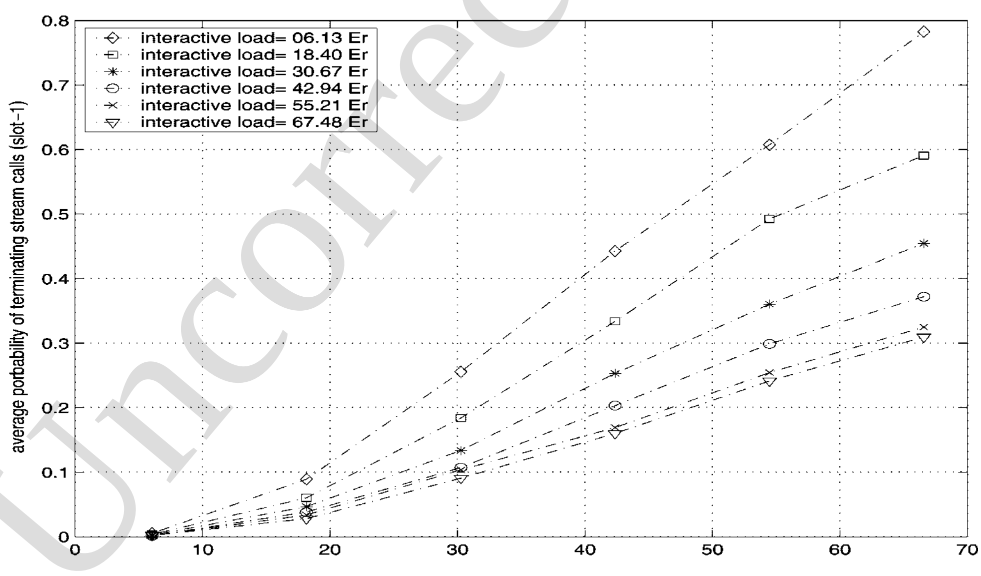



Fig. 7. The call termination probability under Policy-I ($W = 150$, $q = 90$ and $B = 80$).



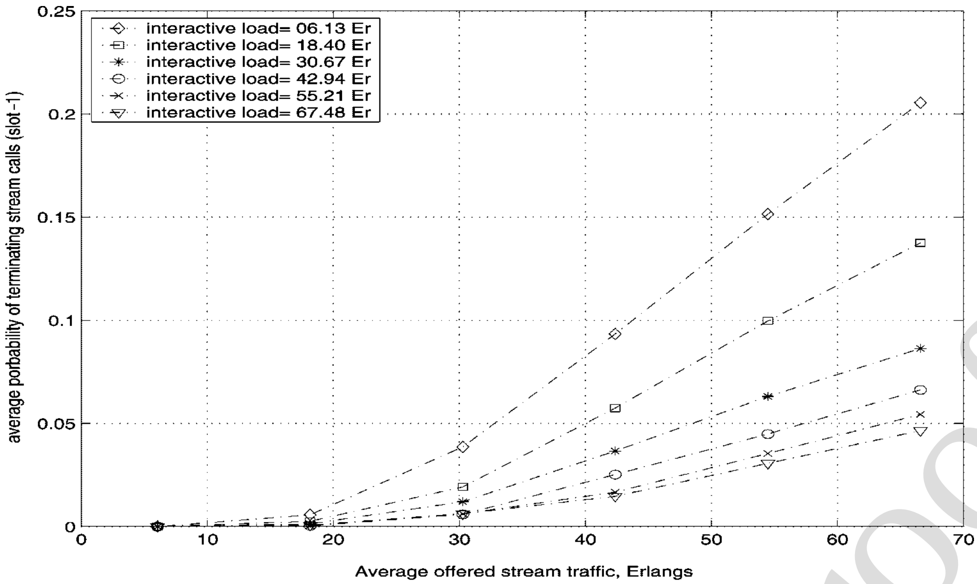



Fig. 8. The call termination probability under Policy-II ($W = 150$, $q = 90$ and $B = 80$).

three windows) hoping the traffic shall statistically be multiplexed. On the other hand, the network operator may choose to adapt the transmission rate (traffic shaping) of some users according to certain flow control policy instead of rejecting the calls completely. Here, of course, it is assumed that the end user can support the large volume of burst packets.

Figures 9–12 illustrate some of the advantages that the stream users will gain as a result of the high price paid in call blocking probability. First, the packet drop (which is the complement of packet throughput as it was defined before) under Policy-I is very low compared to the performance under Policy-II or Policy-III. Industry's standards (e.g. IEEE 802.16.1) request that the packet blocking should not exceed 1%. Considering this specification, we find that Policy-I can guarantee this QoS for up to 40 Erlang stream traffic load under a wide range of interactive load. On the other

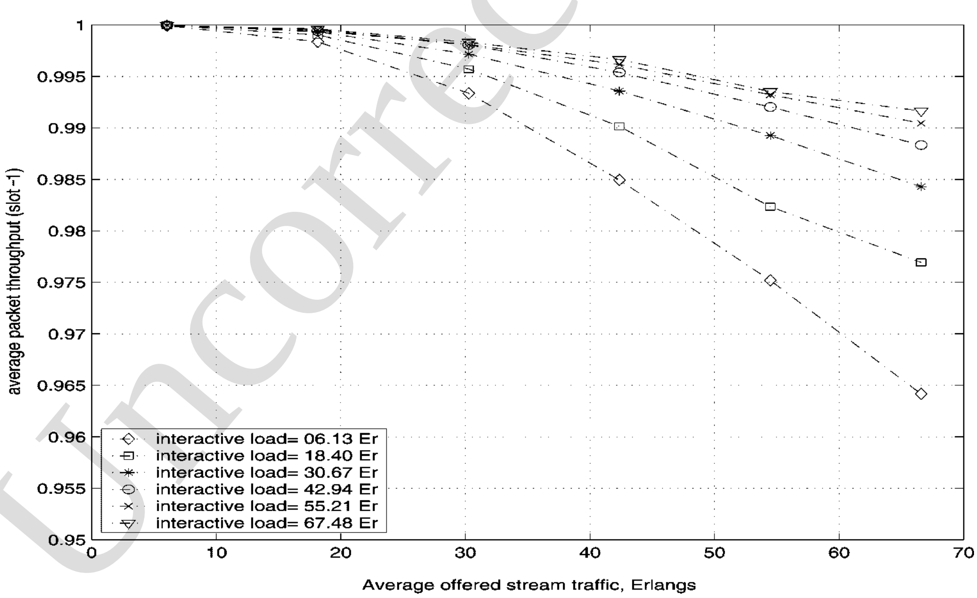



Fig. 9. The packet throughput under Policy-I ($W = 150$, $q = 90$ and $B = 80$).



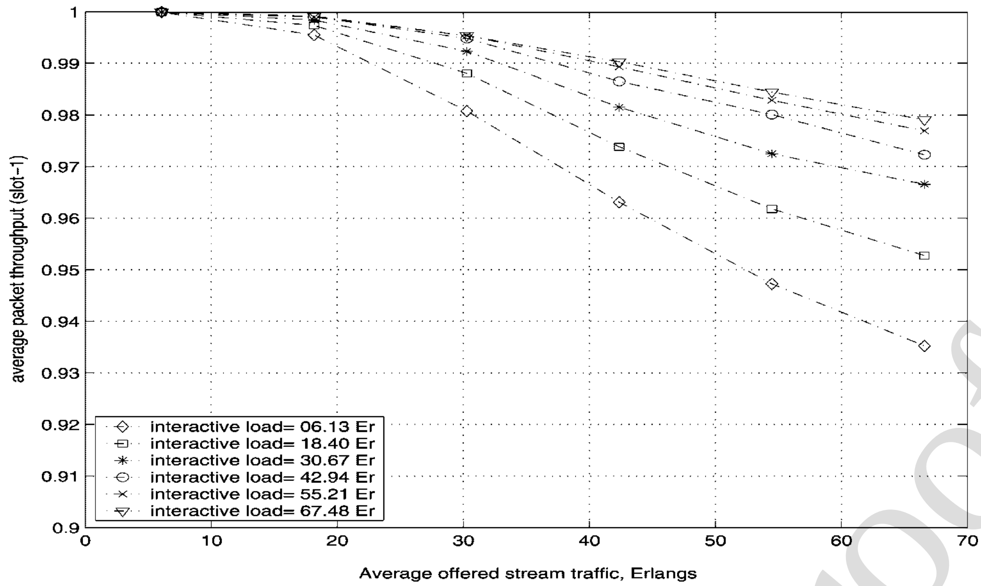



Fig. 10. The packet throughput under Policy-II ($W = 150$, $q = 90$ and $B = 80$).

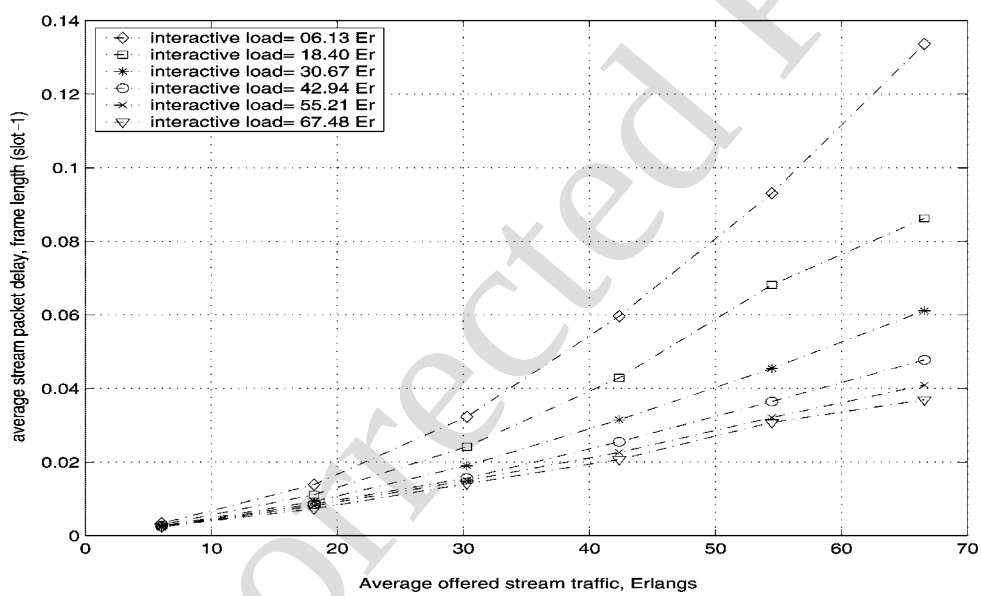



Fig. 11. The average packet delay under Policy-I ($W = 150$, $q = 90$ and $B = 80$).

hand, Policy-II and Policy-III can support up to 33 and 30 Erlang stream loads respectively.

Second, Policy-I has effectively limited the packet delay for stream users. Under the same traffic load, Policy-I limits the packet delay to one-third of the delay under Policy-III and to one half the delay experienced by the stream packets under Policy-II. Yet, Policy-I shows a low system throughput compared with the two policies. This is due to high call termination under this policy which is imposed to guarantee a very stringent QoS that is the active stream should find the buffer not congested for a consecutive $q$ packet-time units.

Considering the performance of interactive users along the excess stream traffic using slot-II, the simulation results show no buffer congestion, and



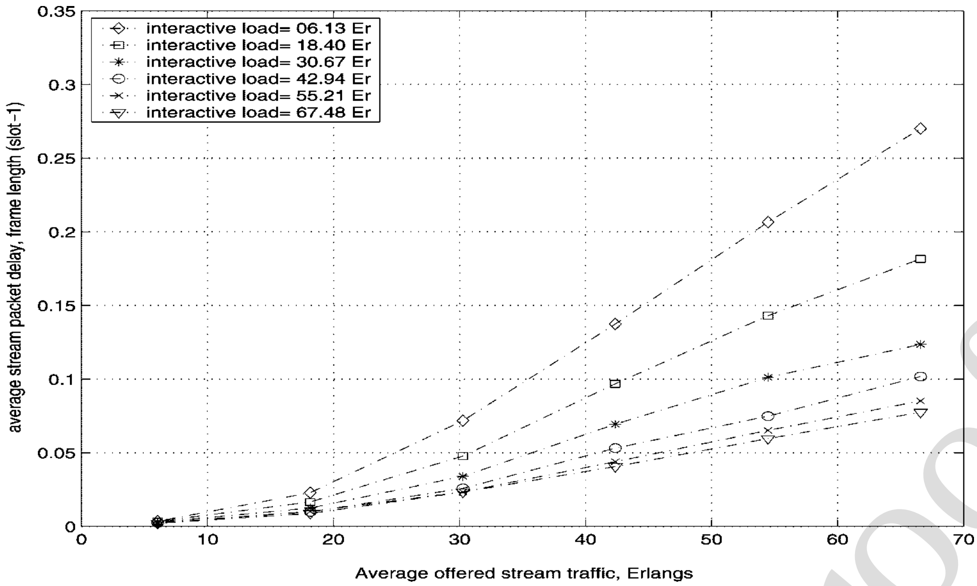



Fig. 12. The average packet delay under Policy-II ($W = 150$, $q = 90$ and $B = 80$).

very small packet delay. However, the two performance parameters that should be examined are the call blocking probability of interactive users requests and the end-to-end interactive packets losses, because these measures have a great impact on the quality of these services. Simulation results illustrate that the stream traffic load plays a little role regarding these parameters especially under Policy-I. The reason for this is the network design where the interactive callers have granted higher priority than the stream callers. Moreover, the call blocking is too high, since the admission policy reserves resources for active interactive users depending on their peak rates.

Indeed, we have run many simulation experiments [8] and due to the limited space, we just present some snapshots of these experiments. Tables II and III show comparison summaries for all simulated admission/congestion policies. The '+' means a gain from the network performance standpoint, while '−' means a loss from the network performance standpoint. For example, Table II shows a fivefold increment in call blocking under Policy-I compared to Policy-III, while Policy-II enhances the call blocking by only 77%. These enhancements in the call blocking are actually losses from the network standpoint.

On the other hand, packet blocking probability drops by 67.4% when we apply Policy-I and by only 10% when Policy-II is applied compared to Policy-III as shown in Table III. Of course, this drop in packet blocking is an improvement in the network performance. Further, from Table III, we can observe an interesting result that is at high traffic load, Policy-I outperforms other policies at the expense of 15% increment in call blocking and at the same time maintaining its stringent QoS requirement.

Table II. Comparison of the network performance under the three proposed Policies, $\rho_{\text{Int}} = 6.13$ Erlang, $\rho_{\text{S}} = 30.27$ Erlang.

| Performance measure | Policy-I ($q = 90$) | Policy-II | Policy-III | Gain/loss (Policy-I) | Gain/loss (Policy-II) |
|---|---|---|---|---|---|
| Call blocking (stream) | 0.2101 | 0.0630 | 0.03563 | 489% (−) | 77% (−) |
| Call blocking (interactive) | 0.0078 | 0.0207 | 0.0095 | 17.9% (+) | 118% (+) |
| Packet blocking | 0.00662 | 0.0192 | 0.02469 | 73% (+) | 22% (+) |
| Packet delay | 0.0323 | 0.0719 | 0.0893 | 63.8% (+) | 19.5% (+) |
| Congested buffer | 0.8962 | 0.9906 | 0.9879 | okay | okay |
| Throughput | 0.8137 | 0.9606 | 1.0000 | 18.6% (−) | 3.9% (−) |
| Packet throughput | 0.9934 | 0.9807 | 0.9953 | 1.85% (+) | 0.5% (+) |
| Call termination | 0.2556 | 0.03862 | 0 | (−) | (−) |

Table III. Comparison of the network performance under the three proposed policies, $\rho_{\text{Int}} = 67.48$ Erlang, $\rho_{\text{S}} = 30.27$ Erlang.

| Performance measure | Policy-I ($q = 90$) | Policy-II | Policy-III | Gain/loss (Policy-I) | Gain/loss (Policy-II) |
|---|---|---|---|---|---|
| Call blocking (stream) | 0.4271 | 0.3760 | 0.3715 | 15% (−) | 1.2% (−) |
| Call blocking (interactive) | 0.7002 | 0.7031 | 0.7003 | 0% | 0% |
| Packet blocking | 0.0017 | 0.0046 | 0.0051 | 67.4% (+) | 10.3% (+) |
| Packet delay | 0.0141 | 0.0234 | 0.0248 | 43.1% (+) | 5.8% (+) |
| Congested buffer | 0.9694 | 0.9983 | 0.9982 | okay | okay |
| Throughput | 0.92753 | 0.99616 | 1.0000 | 7.2% (−) | 0.38% (−) |
| Packet throughput | 0.9983 | 0.9954 | 0.9949 | 0.35% (+) | % (+) |
| Call termination | 0.0908 | 0.0062 | 0 | (−) | (−) |

## 7. Conclusions

In this study, three adaptive admission/congestion policies to accompany the hybrid MC-CDMA/TDMA access protocol have been evaluated using event-driven simulation. Basically, these policies depend in its operation on estimating the buffer occupancy at BS where this estimate will be used to control the call admission as well as to adapt the traffic flow throughout the call duration. The simulation results have shown outstanding performance of Policy-I compared to other admission policies in guaranteeing the requested QoS parameters. It is worth noting that this policy can lend itself easily to other access schemes as well as other network platforms.

In practice, the adaptation process for those newly admitted users can be done through one of the following ways: termination, bite rate variation or both. The former way means that some or all new calls should be terminated to maintain certain QoS requirements. On the other hand, a new user may be requested to lower his transmission bit rate such that the system performance under the new load could be acceptable. It is obvious that we could have many scenarios and the matter of which one should be applied depends either on the kind of service offered by the network or on the kind of service paid for by the user.

Further, this work has focused on evaluating these policies among themselves assuming similar QoS requirements for all users. As a point of research for future studies, we should consider different QoS for different classes of users and also compare their performance against other access schemes.

## Acknowledgments

The authors would like to thank in advance the anonymous reviewers for their comments and suggestions.

## References

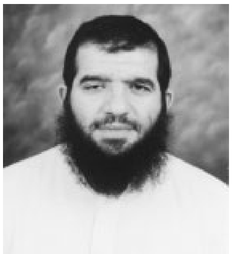
1. Huang D, Shi JJ. Radio resource management for quality of service in third generation mobile communications. In *Proceedings of IEEE VTC'2002*, 2002; pp. 1724–1727.
2. Jeon WS, Jeong DG. Call admission control for CDMA mobile communications systems supporting multimedia services. *IEEE Transactions on Wireless Communications* 2002; **1**(4): 649–659.
3. Singh S, Krishnamurthy V, Poor HV. Integrated voice/data call admission control for wireless DS-CDMA systems. *IEEE Transactions on Signal Processing* 2002; **50**(6): 1483–1495.
4. Soroushnejad M, Geraniotis E. Multi-access strategies for an integrated voice/data CDMA packet radio network. *IEEE Transactions on Communications* 1995; **43**(2/3/4): 934–945.
5. Liu T-K, Silvester JA. Joint admission/congestion control for wireless CDMA systems supporting integrated services. *IEEE Journal of Selected Areas in Communications* 1998; **16**(6): 845–857.
6. Baroudi UA, Elhakeem AK. A hybrid TDMA/MC-CDMA utilizing multiuser detection for integrated wireless networks. *IEICE Transactions on Communications* 2000; **E83-B**(6):[Q4] Q4
7. Lu Z, Karol MJ, ElZarki M, Eng KY. Channel access and interference issues in multi-code DS-CDMA wireless (ATM) networks. *Wireless Networks* 1996; **2**: 173–193.[Q5] Q5
8. Baroudi UA. Performance of Adaptive Admission/Congestion Control Policies on A Hybrid MC-CDMA/TDMA Integrated Multimedia Networks, Ph.D. Dissertation, Concordia University, Montreal, Canada, 2000.
9. Viterbi AJ. *CDMA: Principle of Spread Spectrum Multiple Access Communication.* Addison-Wesley: Reading, MA, 1995.



## Authors' Biographies[Q3]

Q3



# Author Query Form (WCM/239)

**Special Instructions: Author please write responses to queries directly on Galley proofs and then fax back. Alternatively please list responses in an e-mail.**

**Q1: Author: Please provide the equation number.**

**Q2: Author: Please provide section number.**

**Q3: Author: Please provide authors biographies.**

**Q4: Author: Please provide page number.**

**Q5: Author: Reference not cited in text**

**Q6: Author: Please check the short title.**