UNIVERSITY
OF TRENTO

DEPARTMENT OF INFORMATION AND COMMUNICATION TECHNOLOGY

38050 Povo – Trento (Italy), Via Sommarive 14
http://www.dit.unitn.it

A CHANNEL ASSIGNMENT AND HANDOFF STRATEGY FOR MOBILE MULTIMEDIA CELLULAR COMMUNICATION SYSTEMS WITH VARIABLE BANDWIDTH SUPPORT

Bo Li and Roberto Battiti

December 2002

Technical Report # DIT-02-0094

Also: this work has been submitted to the IEEE Transcations on Wireless Communications.

A Channel Assignment and Handoff Strategy for Mobile Multimedia Cellular Communication Systems with Variable Bandwidth Support

Bo Li and Roberto Battiti

Subject: Computer Science (QA075)

(This work has been submitted to
the IEEE Transcations on Wireless Communcations)

December 1, 2002

Abstract

We propose an efficient channel allocation and handoff strategy to guarantee continuous services with QoS guarantees to mobile multimedia users, known as the Variable Bandwidth Channel Allocation Scheme (VBCAS). In order to minimize the forced termination probabilities of ongoing calls, the bandwidth for different types of calls varies in a dynamic way. If there are not sufficient channel resources for a new arriving handoff call, the system adjusts the bandwidth of calls in service and reallocates enough channel resources for the arriving handoff call to avoid interrupting its session. The adjustment of the bandwidth of calls is done so that the basic quality of service for a call with particular service type is guaranteed. The strategy is modelled and analyzed by using a multi-dimensional Markov chain. Performance measures, such as blocking probabilities of new calls, forced termination probabilities of calls, average bandwidth allocated to a call are obtained. Numerical results show that significant improvements on QoS can be obtained by using the channel allocation and handoff strategy proposed in this paper.

KEYWORDS: Cellular Mobile Systems, Channel Allocation, Handoff, Variable Bandwidth Support

1 INTRODUCTION

The new third generation wireless cellular communications networks are expected to carry multimedia traffic, such as voice, video, images, data, or combinations of them. Various issues have to be carefully examined before such systems can be implemented [1]. This paper deals with the problems of channel assignment and handoff control to support multiple types of traffic with different QoS parameters.

In a cellular system, the service area is divided into many cells. Each mobile subscriber (MS) communicates via the base station (BS) of the cell it is currently residing in. Handoff occurs when a MS moves away from its assigned BS (resulting in a degradation of signal strength) and establishes a channel assignment at the target BS [2]. In order to support multimedia applications, the system capacity has to be increased significantly. A common way of achieving this goal is to employ microcells and picocells [3], so that the service area covered by a single BS is reduced and the handoff frequency increases. The widely used handoff method called *cutoff priority scheme* or *guarded channel scheme*, proposed in [4] and [5], has been shown to be effective for such systems. In the cutoff priority scheme, new and handoff calls are treated equally on a FCFS basis for channel allocation until a predetermined channel utilization threshold is reached. At this point, new calls are simply blocked, and only handoff calls are honored. Queuing of new calls as well as handoff calls is considered in [6], which demonstrates that queuing can improve channel utilization and reduce call blocking probabilities. We proposed a handoff scheme adopting preemptive priority in [7]. In the scheme, voice handoff calls can preempt channels of data calls, thus reducing the forced termination probability of voice calls.

There has been increasing interest in the study of channel assignment and handoff schemes in systems supporting multimedia traffic. Multiple call handoff problems with mixed platform types are discussed in [8], where the performance of the system with cutoff priority scheme with or without handoff queuing is analyzed. Li and Chuang present a hybrid cutoff priority handoff scheme for multiple classes of traffic in multimedia wireless networks [9]. There, cutoff priority is used, and each class of calls has a different cutoff threshold. In [10], a call admission control scheme for mobile multimedia communications systems is presented. In the scheme, calls are divided into different classes with different priority levels. Moreover, traffic asymmetry between uplink and downlink in the mobile multimedia environments is also considered. In [11], an adaptive measured-based pre-assignment scheme is proposed with connection-level QoS support. The proposed scheme is different from the cutoff priority based scheme in that it allows the handoff calls to utilize a pre-reserved channel pool before competing with new calls for the shared channels.

In all the above schemes the bandwidth of a call of each service type is constant. That is, the bandwidth allocated to a call does not change during the entire communication. If an arrival handoff call cannot get the required bandwidth, it will be blocked and the call will be forced into termination. Variable bandwidth support for voice services is considered in [12]. In [12], when a handoff call enters a busy cell (no channel available), an occupied full-rate channel can be divided into two half-rate channels to accommodate the arriving handoff call. Actually, for most types of data and multimedia service, the bandwidth occupied by a call can vary from time to time. For example, for some type of video calls, the frame rate for transmission and the encoded data rate per video frame can be reduced if channel resources become scarce. In the compression technique named "Scalable Video Coding (SVC)" [13], a video sequence is compressed into several layers: a base layer and several enhancement layers. The base layer can be independently decoded and it provides basic video quality; the enhancement layers can only be decoded together with the base layer and further refine the quality of the base layer. Therefore, the bandwidth occupied by a video call encoded by SVC can be adjusted according to the variation of the traffic intensity of the system. The reduction of the bandwidth must guarantee the basic communication quality of calls. We propose here a more flexible and general channel assignment and handoff scheme called Variable Bandwidth Channel Allocation Scheme (VBCAS). In order to minimize the forced termination probabilities of calls, variable bandwidth for different types of calls is supported. If there are not enough channel resources to support the arrival of a handoff call , the system adjusts the service rates of calls in service and reallocates enough channel resources for the arriving handoff call to avoid interrupting the session. Therefore, according to the variation of the traffic intensity, the bandwidth allocated to a call in service may vary from time to time. However, the basic quality of service for a call with a particular service type is guaranteed by setting a pre-defined minimum allocated bandwidth. The strategy is modelled and analyzed by using a multi-dimensional Markov chain. Performance measures are obtained, such as blocking probabilities of new calls, forced termination probabilities of calls, average bandwidth allocated to a call.

The remaining part of this paper is organized as follows: The traffic model of the system is presented in section II. Our VBCAS scheme is described in section III. In section IV, the performance of the system is analyzed. In section V, numerical results are presented and discussed to demonstrate that better QoS can be obtained by using VBCAS.

2 TRAFFIC MODELS

In the paper, it is assumed that the whole service area of the system is covered by many homogenous cells. The total bandwidth in each cell is denoted by B.

Each cell in the system supports K types of calls, indexed by $i = 1, 2, \ldots, K$. For a call of type i, the bandwidth allocated to it is between $b_i^{\min}$ and $b_i^{\max}$ at any given time. The value $b_i^{\min}$ corresponds to the lowest endurable communication quality, while $b_i^{\max}$ corresponds to the best communication quality. It is assumed that $b_i^{\min} \geqslant 1$ and $b_i^{\min} \leqslant b_i^{\max}$. In this paper, it is also assumed that mobile users always desire to get the best communication quality. That is, a call of type i always desires to get a bandwidth $b_i^{\max}$ if possible.

The dwell time T_i^D for a mobile terminal of type i in a cell is a negative exponentially distributed random variable having a mean $\overline{T_i^D} = \dfrac{1}{\mu_i^D}$.

Moreover, We define T_{i,b_i}^C as the instantaneous unencumbered call (session) duration. It is the unencumbered call duration for a call of type i, assuming that the bandwidth assigned to it remains unchanged. It is assumed that T_{i,b_i}^C is a negative exponentially distributed random variable, which has a mean $\overline{T_{i,b_i}^C} = \dfrac{1}{\mu_{i,b_i}^C}$. It is evident that $\overline{T_{i,b_i}^C}$ is a function of the type of a call i and of the bandwidth b_i obtained for the call.

The arrival rate of new calls of type i is denoted by λ_i^O. It is assumed that the arriving processes are Poisson processes. The arrival rate of handoff calls of type i is λ_i^H. The arriving processes are also assumed to be Poisson.

3 DESCRIPTION OF THE VBCAS STRATEGY

The system model for our channel allocation and handoff scheme with variable bandwidth support is illustrated in Fig. 1.

Because, for a homogeneous system in statistical equilibrium, the state probabilities of all the cells are assumed to be the same, only the state description of a single cell has to be considered. We now focus attention on a given cell. The cell under observation can be characterized as being, at any given instant, in any one of the finite numbers of states. We define the system state of a cell by a sequence of nonnegative integers. It is convenient to order the states by using an index

$s = 0, 1, 2, \ldots, s_{\max}$. Thus a state s corresponds to a distinct set of nonnegative numbers

$$\begin{pmatrix} n_1(s) & n_2(s) & n_3(s) & \ldots & n_K(s) \\ B_1(s) & B_2(s) & B_3(s) & \ldots & B_K(s) \end{pmatrix} \tag{3.1}$$

where $n_i(s)$ $(i = 1, 2, \cdots, K)$ denotes the number of active calls of type i in the cell, and $B_i(s)$ $(i = 1, 2, \cdots, K)$ is the total bandwidth occupied by these $n_i(s)$ calls. As it is shown later, if we use the bandwidth allocation method proposed in the paper, states can be defined only by the integer set $(n_1(s), n_2(s), \cdots, n_K(s))$.

In our scheme, the cutoff priority scheme is used. That is to say, in each cell, a fraction of the total bandwidth is reserved for handoff calls of any types. A new call of type i is blocked by the system if

$$\sum_{j=1}^{K} n_j(s) b_j^{\max} + b_i^{\max} > B - B_R \tag{3.2}$$

when it arrives to the system at state s. Otherwise, the new call is accepted, and $b_i^{\max}$ channel bandwidth is allocated to it. It should be noted that some schemes have been proposed to adaptively determine the value of the guard channels according to the variation of the system traffic intensity [10][14]. In this paper, the numbers of the guard channels is constant. An extension that can be envisioned for practical applications consists of a periodic tuning of the B_R value depending on the statistical estimation of traffic parameters. Moreover, because different types of calls have different bandwidth requirements, it is reasonable to set different numbers of guard channels for different types of traffic to guarantee that the forced termination probabilities for different types of traffic will be maintained at some reasonable low levels at the same time [9]. For simplicity, in this paper, only one common guard channel value is shared among all types of traffic, which does not influence our understanding on the final results obtained.

In our scheme, a handoff request will be blocked by the system, and the call forced into termination if the following equation holds on the arrival of a handoff call of type i at state s,

$$\sum_{j=1}^{K} n_j(s) b_j^{\min} + b_i^{\min} > B \tag{3.3}$$

Otherwise, if the sum of the minimum bandwidth requirements of all the calls including the arriving handoff call is not more than the total bandwidth B, the handoff request is accepted by the target cell. By using this basic idea, we can see from the numerical results that the forced termination probabilities of calls are decreased significantly.

In the following, we propose a method to allocate the total bandwidth to the different types of calls. It is evident that there are various ways to execute this allocation. However, in order to make our scheme tractable and realistic, we consider the following points:

a. The channel allocation will be only determined by the state of the system. That is, one state corresponds to one determined channel allocation pattern.

b. If on arrival of a handoff call, there are not sufficient channel resources for it, the system will adjust the bandwidth of calls in service and reallocate enough channel resource for the arriving handoff call to avoid interrupting its session. In order to reduce the signaling overhead of the system, it is better that the number of types of calls affected is minimized. Therefore, calls are divided into different priority levels. The bandwidth of calls with the lowest priority level will be first adjusted. If this adjustment is not sufficient to meet the bandwidth requirement of the arriving handoff call, the system will adjust the bandwidth of calls with higher priority levels until the bandwidth requirement is met.

c. When the system tries to adjust the bandwidth of calls with a particular type, it tries to adjust the bandwidth of calls belonging to the same type fairly, which helps to minimize the variation of communication quality for calls of a given type.

Based on the above considerations, without loss of generality, we assume that a call with type i has higher priority over a call of type j, if $i < j$. First, one determines the value of $B_1(s), B_2(s), \cdots, B_K(s)$ by only using $n_1(s), n_2(s), \cdots n_K(s)$, $b_i^{\min}$, $b_i^{\max}$, and B.

The system first tries to allocate channel resources to calls with the highest priority levels to make them obtain their maximum possible bandwidth. Then, if there are still free channel resources, the system tries to allocate the remaining bandwidth to calls with the second priority level. This process will be repeated until the bandwidth allocation for calls with the lowest priority level is completed. To describe the channel allocation scheme in details, we first denote $B_i^{\min}$ and $B_i^{\max}$ as the minimum and maximum bandwidth that can be allocated to all calls of type i in service:

$$B_i^{\min}(s) \equiv n_i(s) b_i^{\min} \tag{3.4}$$

$$B_i^{\max}(s) \equiv n_i(s) b_i^{\max} \tag{3.5}$$

We then define $m(s)$ as the minimum priority level such that the maximum bandwidth cannot be allocated

$$m(s) \equiv \begin{cases} \min\left\{l \mid 1 \leqslant l \leqslant K, \sum_{j=1}^{l}\left[B_j^{\max}(s) - B_j^{\min}(s)\right] > B - \sum_{j=1}^{K} B_j^{\min}(s)\right\} & \\ & \text{if } \sum_{j=1}^{K} B_j^{\min}(s) \leqslant B < \sum_{j=1}^{K} B_j^{\max}(s) \\ K+1 & \text{if } B \geqslant \sum_{j=1}^{K} B_j^{\max}(s) \end{cases} \tag{3.6}$$

Then, under the case of $\sum_{j=1}^{K} B_j^{\min}(s) \leqslant B$, the bandwidths $B_i(s)$ is allocated to the various types of calls by using the following method (described in pseudo-code):

```
BEGIN
    Calculate the value of $m(s)$ based on equation 3.6;
    FOR $i = 1$ TO $m(s) - 1$ DO
        $B_i(s) = B_i^{\max}(s)$;
    FOR $i = m(s) + 1$ TO $K$ DO
        $B_i(s) = B_i^{\min}(s)$;
    $B_{m(s)}(s) = B - \sum_{j=1}^{m(s)-1} B_j^{\max}(s) - \sum_{j=m(s)+1}^{K} B_j^{\min}(s)$;
END
```

The next problem to be solved is to allocate the $B_i(s)$ channels to $n_i(s)$ calls. Here we propose to allocate channels to calls as fairly as possible. When we try to allocate $B_i(s)$ channels to $n_i(s)$ mobile users, we must cope with two cases. If the integer $B_i(s)$ can be divided by $n_i(s)$ equal bandwidth is allocated to all these $n_i(s)$ calls. Otherwise, different bandwidth must be allocated to the $n_i(s)$ calls to take care of the remainder of the division. In detail, we assign $b_i^{low}(s)$ channels to each of the $n_i^{low}(s)$ calls, and $b_i^{upper}(s)$ channels to each of the $n_i^{low}(s)$ calls. $n_i^{low}(s)$ and $n_i^{upper}(s)$ must satisfy

$$\begin{cases} n_i^{low}(s) + n_i^{upper}(s) = n_i(s) \\ n_i^{low}(s) b_i^{low}(s) + n_i^{upper}(s) b_i^{upper}(s) = B_i(s) \end{cases} \tag{3.7}$$

For that purpose, let

$$\begin{cases} b_i^{upper}(s) = \left\lceil \dfrac{B_i(s)}{n_i(s)} \right\rceil \\ b_i^{low}(s) = \left\lfloor \dfrac{B_i(s)}{n_i(s)} \right\rfloor \end{cases} \tag{3.8}$$

We can see that $b_i^{upper}(s) - b_i^{low}(s) = 0$ or 1. Thus when $b_i^{upper}(s) - b_i^{low}(s) = 1$

$$\begin{cases} n_i^{upper}(s) = B_i(s) - n_i(s)b_i^{low}(s) \\ n_i^{low}(s) = n_i(s) - n_i^{upper}(s) \end{cases} \tag{3.9}$$

From above equation, it can be seen that the remainder of the division will be fairly allocated to $n_i^{upper}(s)$ calls. Moreover, when $b_i^{low}(s) = b_i^{upper}(s)$, we set $n_i^{low}(s) = n_i(s)$, $n_i^{upper}(s) = 0$.

Next, we explain our channel allocation scheme when a handoff request with type i ($i = 1, 2, \cdots, K$) arrives. It is assumed that before the arrival of the handoff call, the system is in state s', and after the acceptance of the handoff call, the system state transits to s. From the above definitions about our channel allocation scheme, especially the definition of $m(s)$ made in equation 3.6, we know that $m(s) \leqslant m(s')$. That is, partial channel resources occupied by calls with type j ($m(s) \leqslant j \leqslant m(s')$) will be borrowed by the system and allocated to the newly arrived handoff call. The borrowing of channel resources is based on the principle of borrowing channels occupied by calls with lower priority first and borrowing channels fairly among calls belonging to the same class. If some channel resources are released by some call, these channel resources will be returned back to calls with higher priority first. Bandwidth occupied by calls with type i ($i > m(s')$ or $i < m(s)$) remains unchanged. Therefore, for a handoff call, with type i ($i \geqslant m(s) + 1$), it gets a bandwidth of $b_i^{\min}$. For a handoff call, with type i ($i \leqslant m(s) - 1$), it gets a bandwidth of $b_i^{\max}$. For a handoff call, with type i ($i = m(s)$), it gets a bandwidth of $b_i^{low}(s)$ (see equation 3.8).

4 PERFORMANCE ANALYSIS

A. Driving Processes

As time progresses, the cell under observation changes its state at random instants affected by the driving processes and the system dynamics. The driving processes of the state of the system are:

1. The arrival of a new call of type i.
2. The arrival of a handoff call of type i.
3. The departure of an active call of type i from the current cell.

Any state transition must be caused by one of the events listed above.

B. Flow Balance Equations

After defining the system states and identifying the driving processes, the statistical equilibrium state probabilities remain to be determined. This can be done by writing a flow balance equation for each state, and then solve a set of the resultant $s_{\max}+1$ simultaneous equations for the unknown state probabilities $p(s)$. Among the $s_{\max}+2$ equations

$$\begin{cases} \sum_{j=0}^{s_{\max}} q(i,j)p(j) = 0 \ (i = 0, 1, 2, \cdots, s_{\max}) \\ \sum_{j=0}^{s_{\max}} p(j) = 1 \end{cases} \tag{4.1}$$

$s_{\max}+1$ are independent. Coefficients $q(i,j)(i \neq j)$ represent the transition rate into state i from predecessor state j, and $q(i,i)$ is equal to the total transition rate out of state i. Note that all the states mentioned must be permissible. That is, all the channel constraints must be met. We obtain $q(i,j)$'s as follows.

(i) Probability Flow into a State Transitions into a given state s arise from several other states s', depending on the driving process that cause the transition. The determination of permissible predecessors and transition rates for a given state s is discussed below.

(i).1 *a. Flow in due to an origination of a call of type i* Assume that the given cell is in state s, corresponding to $(n_1(s), n_2(s), \cdots, n_K(s))$. Transition into state s due to the origination of a new call of type i in the cell is possible only from certain other permitted states. Specifically, let $q_{1i}(s,s')$ denotes the transition rate from state s' to state s due to the origination of a new call of type i. The predecessor involved in this event is:

$$s' : (n_1(s), n_2(s), \cdots, n_i(s) - 1, \cdots, n_K(s)) \tag{4.2}$$

if the conditions $n_i(s) > 0$ and $\sum_{j=1}^{K} n_j(s) b_j^{\max} \leqslant B - B_R$ hold. The transition rate can be written as

$$q_{1i}(s,s') = \lambda_i^O \tag{4.3}$$

(i).2 *b. Flow in due to an arrival of a handoff call of type i* Let $q_{2i}(s,s')$ denotes the transition rate from state s' to state s due to the arrival of a handoff call of type i in the cell. The predecessor s' involved in this event is:

$$s' : (n_1(s), n_2(s), \cdots, n_i(s) - 1, \cdots, n_K(s)) \tag{4.4}$$

if the conditions $n_i(s) > 0$ and $\sum_{j=1}^{K} n_j(s) b_j^{\min} \leqslant B$ hold. The transition rate can be written as

$$q_{2i}(s, s') = \lambda_i^H \tag{4.5}$$

Since we assume an equilibrium homogeneous mobility pattern, the mean number of incoming users into the current cell is equal to that of outgoing ones. Therefore, the arrival rate of handoff requests at the marked cell is equal to the departure rate of handoff calls from the cell. It is apparent that

$$\lambda_i^H = \overline{n_i} \mu_i^D \tag{4.6}$$

$\overline{n_i}$ is the average number of calls of type i holding channels in a cell. Its calculation will be presented later.

(i).3 *c. Flow in due to a departure of a call of type i from the current cell* Let $q_{3i}(s, s')$ denotes the transition rate from state s' to state s due to the departure of a call of type i. The predecessor s' is:

$$s' : (n_1(s), n_2(s), \cdots, n_i(s) + 1, \cdots, n_K(s)) \tag{4.7}$$

if the condition $n_i(s) \geqslant 0$ hold. The transition rate can be written as

$$q_{3i}(s, s') = n_i^{low}(s')(\mu^C_{i, b_i^{low}(s')} + \mu_i^D) + n_i^{upper}(s')(\mu^C_{i, b_i^{upper}(s')} + \mu_i^D) \tag{4.8}$$

In summary, the probability flow into a state s is the sum of the probability flow due to each individual driving process. Thus it is given as

$$\text{Probability flow into state } s = \sum_{j=0, j \neq s}^{s_{\max}} q(s, j) p(j) \tag{4.9}$$

<ins>(ii) Probability Flow out of a State</ins> Because any transition out of a state is a transition into some other states, we must have

$$q(s, s) = - \sum_{k=0, k \neq s}^{s_{\max}} q(k, s) \tag{4.10}$$

That is, the total transition rate out of a state s is the sum of the transition rates from s towards any other states, with reversed direction of flow.

Thus far, the coefficients needed in equation 4.1 to form a set of $s_{\max}+1$ simultaneous equations for the unknown state probabilities $p(s)$ have been determined. These equations can be solved by using the method described in the Appendix.

Once the equilibrium state probabilities $p(s)$ $(s = 0, 1, 2, \cdots, s_{\max})$ are found, the system performance measures can be determined as follows.

C. Performance Measures

Based on all the state probabilities $p(s)$ obtained, various performance measures can be readily calculated.

The blocking probability P_i^{BO} of new calls of type i is given by

$$P_i^{BO} = \sum_{s \in \Omega_i^{BO}} p(s) \tag{4.11}$$

where $\Omega_i^{BO} = \Big\{ s | \sum_{j=1}^{K} n_j(s) b_j^{\max} + b_i^{\max} > B - B_R \Big\}$.

The blocking probability P_i^{BH} of handoff calls of type i is

$$P_i^{BH} = \sum_{s \in \Omega_i^{BH}} p(s) \tag{4.12}$$

where $\Omega_i^{BH} = \Big\{ s | \sum_{j=1}^{K} n_j(s) b_j^{\min} + b_i^{\min} > B \Big\}$.

When a mobile user is assigned some channels in a cell, subsequent cell boundary crossings while the call is in progress will necessitate handoffs. The handoff requirement probability of a call of type i is defined as P_i^H, which can be calculated as follows,

$$P_i^H = \frac{\lambda_i^H}{\lambda_i^H (1 - P_i^{BH}) + \lambda_i^O (1 - P_i^{BO})} \tag{4.13}$$

The forced termination probability P_i^{FT} of a call of type i, that is, the probability that a call of type i accepted into the system is forced into termination during its lifetime, is an important measure of the system performance. P_i^{FT} can be expressed as

$$\begin{aligned} P_i^{FT} &= \sum_{l=1}^{\infty} P_i^H P_i^{BH} \left[(1 - P_i^{BH}) P_i^H \right]^{l-1} \\ &= \frac{P_i^H P_i^{BH}}{1 - P_i^H (1 - P_i^{BH})} \end{aligned} \tag{4.14}$$

$\overline{n_i}$ $(i = 1, 2, \cdots, K)$ can be obtained as follows

$$\overline{n_i} = \sum_{s=0}^{s_{\max}} p(s) n_i(s) \tag{4.15}$$

Let $\overline{b_i}$ denote the average bandwidth allocated to a call of type i during the course of its duration (average bandwidth for a call of type i). We have

$$\overline{b_i} = \frac{\sum_{s \in \{s | n_i(s) \geqslant 1\}} \dfrac{B_i(s)}{n_i(s)} \cdot p(s)}{\sum_{s \in \{s | n_i(s) \geqslant 1\}} p(s)} \tag{4.16}$$

$\overline{b_i}$ $(i = 1, 2, \cdots, K)$ is an important performance measure, which represents the average communication quality of calls. The larger the $\overline{b_i}$ is, the better the communication quality for calls of type i is. It is evident that $b_i^{\min} \leqslant \overline{b_i} \leqslant b_i^{\max}$.

In order to measure the average variation of the bandwidth $b_i(s)$ allocated to a call with type i at the state s from $\overline{b_i}$, we define σ_i as

$$\sigma_i = \frac{\sum_{s \in \{s | n_i(s) \geqslant 1\}} \left| \dfrac{B_i(s)}{n_i(s)} - \overline{b_i} \right| \cdot p(s)}{\sum_{s \in \{s | n_i(s) \geqslant 1\}} p(s)} \tag{4.17}$$

Based on equation 4.16 and 4.17, we make the following definitions,

$$\begin{cases} \eta_i \equiv \dfrac{\overline{b_i}}{b_i^{\max}} \\ \delta_i \equiv \dfrac{\sigma_i}{\overline{b_i}} \end{cases} \tag{4.18}$$

In the following discussions, η_i and δ_i are used to demonstrate the communication qualities of calls.

5 NUMERICAL RESULTS

In our numerical examples, two types of traffic are considered, that is to say $K = 2$. For concreteness, one man assumes that one is voice traffic $(i = 1)$, and the other one is video traffic $(i = 2)$. In our numerical experiments, for voice and video traffic, the session duration of a call is independent of the channel bandwidth allocated to it.

In our numerical examples, for voice traffic, we have $b_1^{\min} = b_1^{\max} = 1$, $T_1^D = 300\,\text{sec}$, $T_{b_1}^C = 120\,\text{sec}$. For video traffic, we have $b_2^{\max} = 4$, $T_2^D = 1200\,\text{sec}$ and $T_{b2}^C = 3000\,\text{sec}$. A new call is randomly determined as a voice call with probability of 0.85, and as a video call with probability of 0.15. Moreover, the total bandwidth in a cell is set to be $B = 30$.

The blocking probabilities and the forced termination probabilities for voice traffic versus the total arrival rate of new calls are shown in Fig. 2 and Fig. 3, respectively. In Fig. 4 and Fig. 5, the blocking probabilities and the forced termination probabilities for video traffic versus the total arrival rate of new calls are shown, respectively. In these figures, comparisons have been done among four different cases. In case 1, $B_R = 0$ and $b_2^{\min} = b_2^{\max} = 4$. That is to say, no guard bandwidth is set, and the bandwidth of calls cannot be adjusted. In case 2, $B_R = 5$ and $b_2^{\min} = b_2^{\max} = 4$, which means that guard bandwidth is set for handoff calls, but the variable bandwidth is not supported. In case 3, no guard bandwidth is allocated for handoff calls, but variable bandwidth is supported for video traffic. Therefore, in this case, $B_R = 0$, $b_2^{\min} = 2$ and $b_2^{\max} = 4$. Finally, in case 4, $B_R = 5$, $b_2^{\min} = 2$ and $b_2^{\max} = 4$, both guard bandwidth and variable bandwidth for video traffic are supported.

In case 1, no guard channel is set, and the bandwidth of calls cannot be adjusted so that, no priority is provided to handoff calls over new calls. Therefore, from Figs 3 and 5, we can see that the forced termination probabilities for both voice and video calls are very high. Especially for video calls (refer to Fig. 5), in the case of heavy offered traffic, the forced termination probabilities are higher than 0.1, which means that it is difficult to complete a video session without being forced into termination. From our results, it is evident that the performance of the system cannot be accepted by mobile users without providing some priority to handoff calls. In case 2, some guard bandwidth has been reserved exclusively for handoff calls, which makes the forced termination probabilities for both voice and video calls lower than those in case 1. However, since variable bandwidth is not supported, the performance improvements on the forced termination probabilities for both voice and video calls are still very limited. Again, from Fig. 3 and Fig. 5, it can be seen that the forced termination probabilities for both voice and video calls are still very high under the heavy offered traffic. Of course, in order to further decrease the forced termination probabilities for mobile users, we can increase B_R to reserve more bandwidth for handoff traffic, but this will bring about much larger blocking probabilities for new calls and poor utilization of bandwidth resources. The purpose of our method based on variable bandwidth is to drastically decrease the forced termination probabilities for calls of all types, and at the same time, to avoid increasing the blocking probabilities for new calls in a significant way. The variable bandwidth support proposed and analyzed in this paper is an effective way to increase the bandwidth utilization.

Compared to the case 1 and case 2, in case 3 variable bandwidth is supported for video calls without adopting guard bandwidth, which makes the forced termination probabilities for both voice and video calls decrease drastically (see Fig. 3 and Fig. 5). Moreover, from Fig. 2 and Fig. 4, it can be seen that the blocking probabilities in case 3 are not significantly increased compared to those in case 2 (under the case of lower offered traffic load, the blocking probabilities in case 3 are lower than those in case 2), which indicates that the bandwidth is fully utilized in case 3. From Fig. 2 and Fig. 4, it can be also seen that the blocking probabilities in case 3 are larger than those in case 1. This is because in case 3 many more handoff requests are accepted by the system, which makes it more difficult for a new call to have access to the system. In case 4, both guard bandwidth and variable bandwidth for video traffic are supported. Compared to the case 3, some channel resources are reserved for handoff calls, which decreases the forced termination probabilities and increases the blocking probabilities for new calls.

In Fig. 6 and Fig. 7, the communication quality for video calls is shown. In Fig. 6, η_2 for video calls versus the total arrival rate of new calls is shown. In Fig. 7, δ_2 for video calls is shown. In these two figures, comparisons have been done between two different cases: one is $B_R = 0$, $b_2^{\min} = 2$, $b_2^{\max} = 4$, and the other $B_R = 5$, $b_2^{\min} = 2$, $b_2^{\max} = 4$. From Fig. 6 we can see that even in the case of heavy offered traffic load, η_2 is very near to 1.0, which means that the average bandwidth is very near to the maximum bandwidth requirement of a video call. Fig. 7 illustrates the bandwidth variations from the average bandwidth for video calls. Again, we can see that even in the case of heavy offered traffic load, δ_2 is near to 0.0, which means that the bandwidth variations around the average bandwidth for a video call is not large. From these two Figures, we can see that the communication quality for a video call will be affected to some extent by adopting the variable bandwidth strategy, but this effect is not large, especially for the case of low and medium offered traffic load. Moreover, we can also see that the communication quality for a video call will be further improved by increasing the value of B_R, which means that the communication quality of calls with variable bandwidth can be controlled and improved by adopting suitable call admission control scheme.

6 CONCLUSIONS

In this paper, we propose an efficient channel allocation and handoff strategy to guarantee continuous services with good QoS to mobile multimedia users. In order to minimize the forced termination probabilities of ongoing calls, variable bandwidth for different types of calls is supported. If there are not sufficient channel resources to support an arriving handoff call, the system will adjust the bandwidth of calls in service and reallocate enough channel resources for the arrival handoff call to keep its session going on. The adjustment of the bandwidth of calls is done so that the basic quality of service for a call with particular service type is guaranteed. The strategy is modelled and analyzed by using a multi-dimensional Markov chain. Performance measures, such as blocking probabilities of new calls, forced termination probabilities of calls, average bandwidth allocated to a call are obtained. Numerical results show that significant improvements on QoS can be obtained by using the channel allocation and handoff strategy proposed in this paper.

The choice of the channel allocation strategy depends on the different quality of service requirements for different types of traffic. We plan to extend the basic ideas of our scheme with variable bandwidth support to other channel allocation strategies in the future research.

7 APPENDIX

In order to obtain all the state probabilities, we must solve the flow balance equation 4.1. If we assume that λ_i^H's $(i = 1, 2, \cdots, K)$ are constant, the flow balance equation 4.1 can be solved as a set of linear simultaneous equations. However, actually , λ_i^H's $(i = 1, 2, \cdots, K)$ are not constant, but dependent on the state probabilities $p(s)$'s. Thus we use the following iteration procedure to obtain all the state probabilities:

Step1: In order to make the convergence speed of the calculation faster, we select initial values for $\lambda_i^H \approx \dfrac{\mu_i^D}{\mu_{i,b_i^{\max}}^C} \lambda_i^O$ $(i = 1, 2, \cdots, K)$

Step2: With the values of λ_i^H's, compute all the state probabilities $p(s)$'s in equation 4.1 by using the method of Successive Over-Relaxation (SOR) [15].

Step3: Compute $\overline{n_i}$'s $(i = 1, 2, \cdots, K)$ by using equation 4.15.

Step4: Compute $new\lambda_i^H$ $(i = 1, 2, \cdots, K)$ by using equation 4.6.

Step5: If $\left| \dfrac{new\lambda_i^H - old\lambda_i^H}{old\lambda_i^H} \right| < \varepsilon$, then stop the iteration. Otherwise, let $\lambda_i^H \Leftarrow old\lambda_i^H + \omega(new\lambda_i^H - old\lambda_i^H)$ (in our calculation, ω is set to be 0.5), and then go to **Step2** again. ε is a very small positive number to check the convergence. In our numerical examples, ε is set to 0.00001. On the average, It takes about 15 second for convergence by using a personal computer with Intel P4 CPU (1700MHz system clock).

References

1. M. Schwartz, "Network management and control issues in multimedia wireless networks," IEEE Personal Communications, June 1995, pp.8-16.

2. A. Abutaleb and V.O.K. Li, "Performance evaluation of handoff schemes in personal communication systems," IEICE Transactions on Fundamentals E78-A(7), July 1995, pp.773-784

3. A.S. Acampora and M. Naghshineh, "An architecture and methodology for mobile-executed handoff in cellular ATM networks," IEEE Journal on Selected Areas in Communications, 12(8), Oct 1994, pp.1365-1375.

4. E. C. Posner and R. Guerin, "Traffic Policies in Cellular Radio That Minimize Blocking of Handoff Calls," The 11th ITC, Kyoto, Japan, September 1985, pp. 2.4B.2.1–2.4B.2.5.

5. D. Hong and S.S. Rappaport, "Priority oriented channel access for cellular systems serving vehicular and portable radio telephones," IEE Proceedings 136-I(5), Oct 1989, pp. 339-346.

6. R. Guerin, "Queuing-blocking system with two arrival streams and guarded channels," IEEE Transactions on Communications, 36, Feb 1988, pp. 153-163.

7. B. Li, Q.A. Zeng, K. Mukumoto and A. Fukuda, "A Preemptive Priority handoff scheme in integrated voice and data cellular systems," IEICE Transactions on communications, E82-B(10), Oct 1999, pp. 1633-1642.

8. S.S. Rappaport and C. Purzynski, "Prioritized resource assignment for mobile cellular communication systems with mixed services and platform types," IEEE Transactions on Vehicular Technology, vol. 45, No. 3, Aug 1996, pp. 443-458.

9. B. Li, S.T. Chauson, and C. Liu, "Analysis of a hybrid cutoff priority scheme for multiple classes of traffic in multimedia wireless networks," Wireless Networks, vol. 4, No. 4, 1998, pp. 279-290.

10. W. S. Jeon and D. G. Jeong, "Call admission control for mobile multimedia communications with traffic asymmetry between uplink and downlink," IEEE Transactions On Vehicular Technology, vol. 50, No.1, Jan 2001, pp. 59-66.

11. X.Y. Luo, B. Li, I. L. J. Thng, Y. B. Lin and I. Chlamtac, "An adaptive measurement-based preassignment scheme with connection-level QoS support for mobile networks," IEEE Transactions On Wireless communications, vol. 1, No. 3, July 2002, pp. 521-530.

12. Y. B. Lin, A. Noerpel, and D. A. Harasty, "The sub-rating channel assignment strategy for PCS handoffs," IEEE Transactions On Vehicular Technology, vol. 45, no. 1, Feb 1996, pp. 122-130.

13. D. P. Wu, Y. W. Thomas Hou, Y. Q. Zhang, "Scalable video coding and transport over broad-band wireless networks," Proceedings of the IEEE, vol. 89, No. 1, Jan 2001, pp. 6-20.

14. P. Ramanathan, K. M. Sivalingam, P. Agrawal and S. Kishore, "Dynamic resource allocation schemes during handoff for mobile multimedia wireless networks," IEEE Journal on Selected Areas in Communications, vol. 17, No. 7, July 1999, pp.1270-1283.

15. G. Bolch, S. Greiner, H. de Meer, and K.S. Trivedi, Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications, A Wiley-Interscience Publication, 1998, pp. 140-144.

Fig. 1

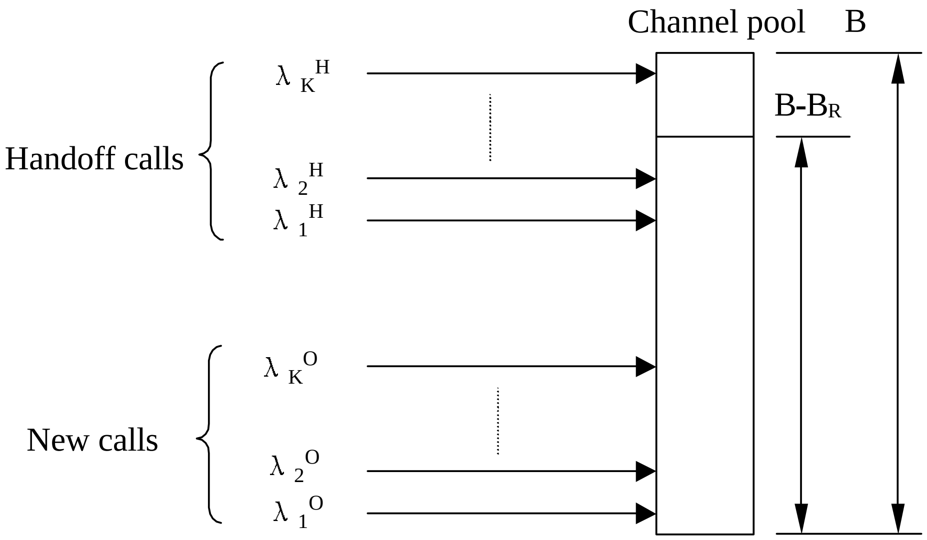

Channel pool
B
B-B_R
Handoff calls
λ_K^H
λ_2^H
λ_1^H
New calls
λ_K^O
λ_2^O
λ_1^O

Fig. 2

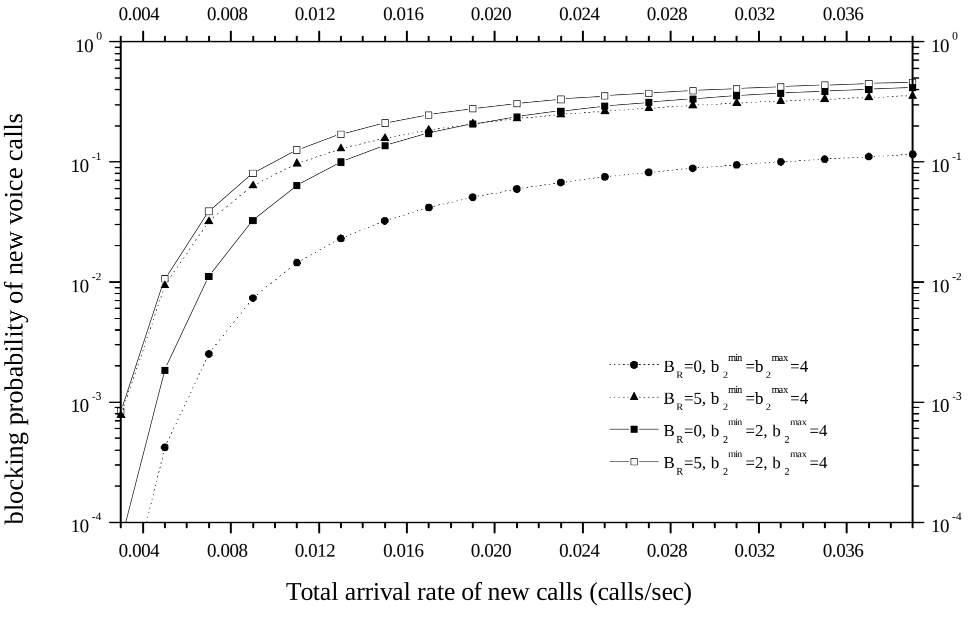

0.004
0.008
0.012
0.016
0.020
0.024
0.028
0.032
0.036
10^0
10^-1
10^-2
10^-3
10^-4
blocking probability of new voice calls
Total arrival rate of new calls (calls/sec)
B_R=0, b_2^min=b_2^max=4
B_R=5, b_2^min=b_2^max=4
B_R=0, b_2^min=2, b_2^max=4
B_R=5, b_2^min=2, b_2^max=4

Fig. 3

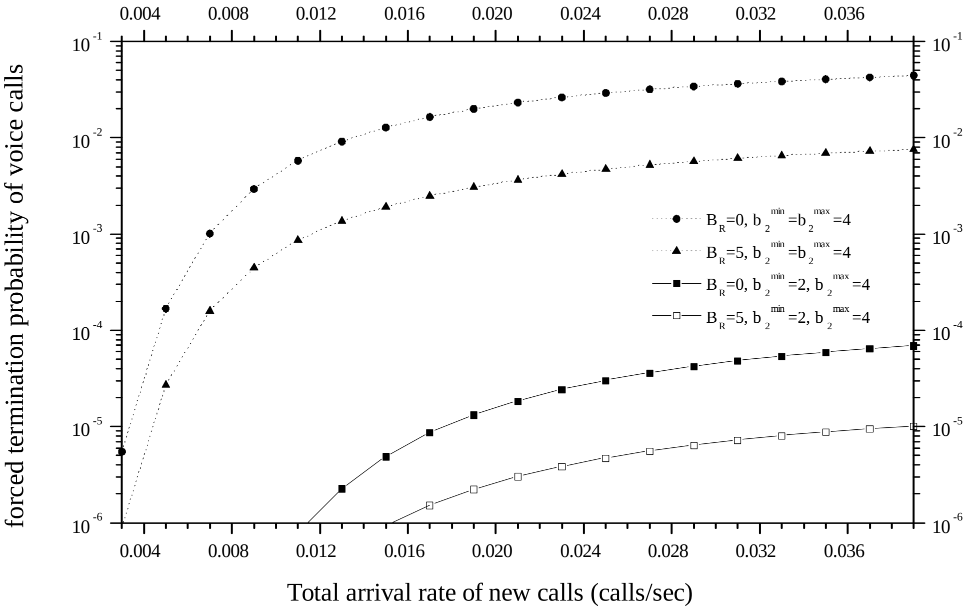

forced termination probability of voice calls
Total arrival rate of new calls (calls/sec)
0.004
0.008
0.012
0.016
0.020
0.024
0.028
0.032
0.036
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
$B_R=0, b_2^{min}=b_2^{max}=4$
$B_R=5, b_2^{min}=b_2^{max}=4$
$B_R=0, b_2^{min}=2, b_2^{max}=4$
$B_R=5, b_2^{min}=2, b_2^{max}=4$

Fig. 4

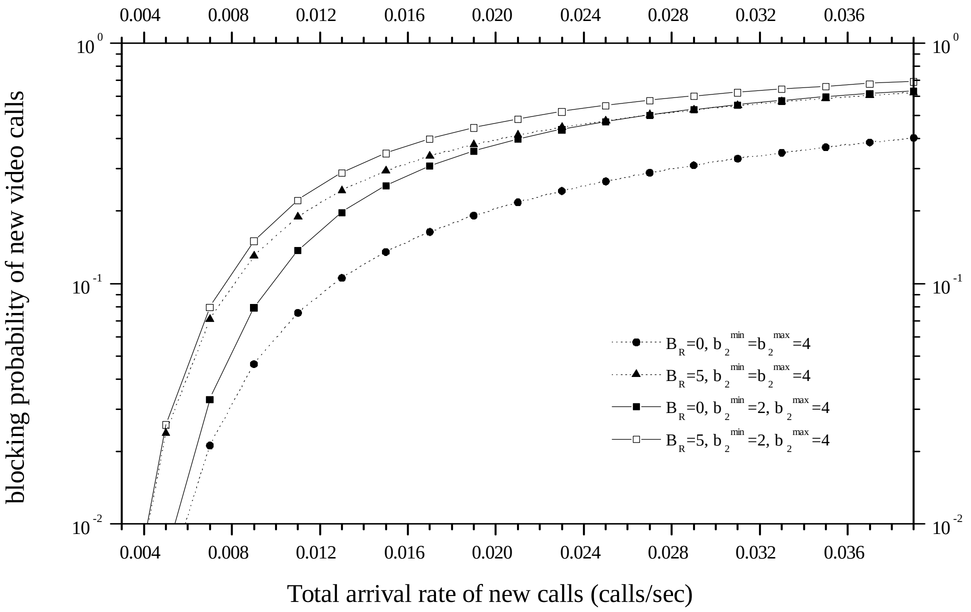
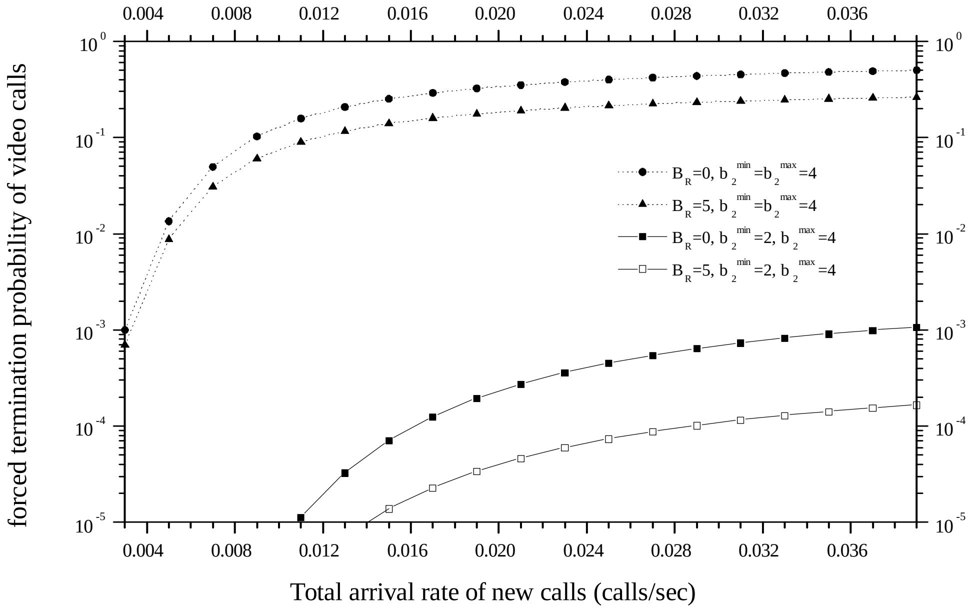
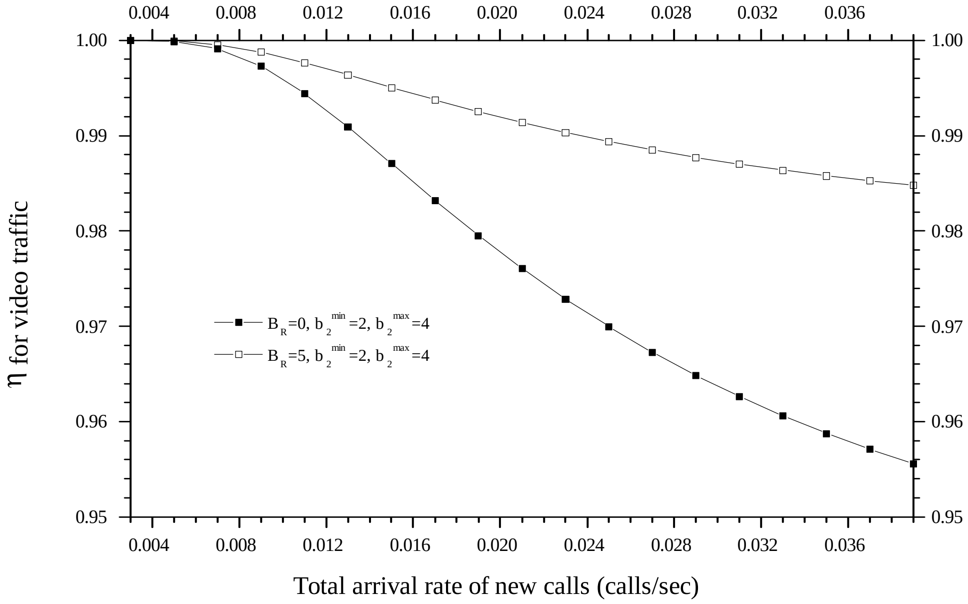

blocking probability of new video calls
Total arrival rate of new calls (calls/sec)
0.004
0.008
0.012
0.016
0.020
0.024
0.028
0.032
0.036
10^{0}
10^{-1}
10^{-2}
$B_R=0, b_2^{min}=b_2^{max}=4$
$B_R=5, b_2^{min}=b_2^{max}=4$
$B_R=0, b_2^{min}=2, b_2^{max}=4$
$B_R=5, b_2^{min}=2, b_2^{max}=4$

Fig. 5



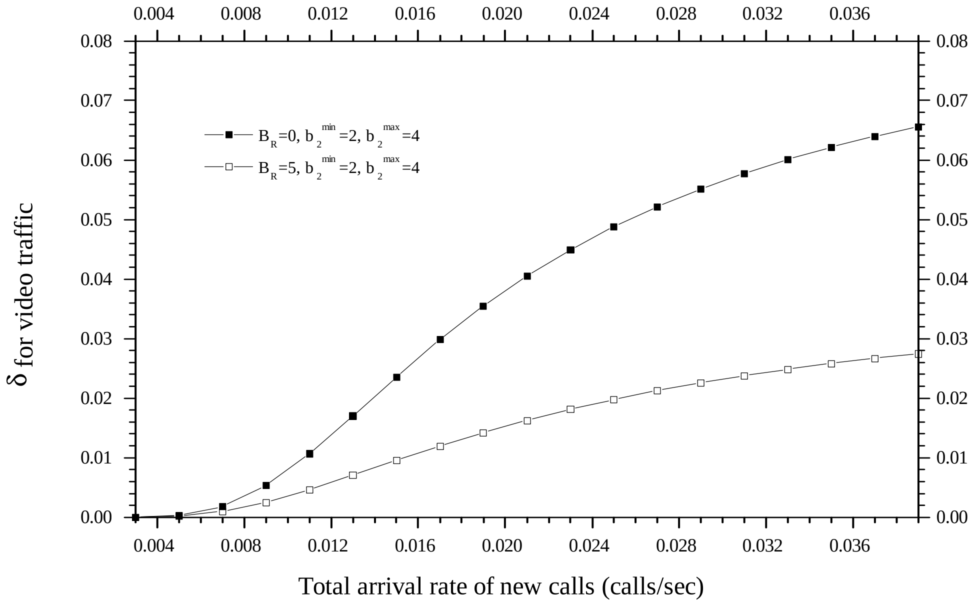
forced termination probability of video calls
Total arrival rate of new calls (calls/sec)
0.004
0.008
0.012
0.016
0.020
0.024
0.028
0.032
0.036
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
$B_R=0, b_2^{min}=b_2^{max}=4$
$B_R=5, b_2^{min}=b_2^{max}=4$
$B_R=0, b_2^{min}=2, b_2^{max}=4$
$B_R=5, b_2^{min}=2, b_2^{max}=4$

Fig. 6



η for video traffic
Total arrival rate of new calls (calls/sec)
0.004
0.008
0.012
0.016
0.020
0.024
0.028
0.032
0.036
1.00
0.99
0.98
0.97
0.96
0.95
$B_R=0, b_2^{min}=2, b_2^{max}=4$
$B_R=5, b_2^{min}=2, b_2^{max}=4$

Fig. 7


0.004
0.008
0.012
0.016
0.020
0.024
0.028
0.032
0.036
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
$B_R=0, b_2^{min}=2, b_2^{max}=4$
$B_R=5, b_2^{min}=2, b_2^{max}=4$
δ for video traffic
Total arrival rate of new calls (calls/sec)