An Adaptive Measured-Based Preassignment Scheme With Connection-Level QoS Support for Mobile Networks

Xiaoyuan Luo, *Member, IEEE*, Bo Li, *Senior Member, IEEE*, Ian Li-Jin Thng, Yi-Bing Lin, and Imrich Chlamtac, *Fellow, IEEE*

***Abstract*—This paper presents a new adaptive bandwidth allocation scheme to prevent handoff failure in wireless cellular networks, known as the measurement-based preassignment (MPr) technique. This technique is particularly useful in micro/pico cellular networks which offers quality-of-service (QoS) guarantee against call dropping. The proposed MPr scheme distinguishes itself from the well-known guarded channel (GC) based schemes in that it allows the handoff calls to utilize a prereserved channel pool before competing for the shared channels with new call arrivals. The key advantage of the proposed MPr scheme is that it enables easy derivation of the number of channels that needs to be reserved for handoff based on a predetermined handoff dropping probability, without the need for solving the often complex Markov chain required in GC schemes, thus, making the proposed MPr scheme simple and efficient for implementation. This is essential in handling multiple traffic types with potentially different QoS requirements. In addition, the MPr scheme is adaptive in that it can dynamically adjust the number of reserved channels for the handoff according to the periodical measurement of the traffic status within a local cell, thus completely eliminating the signaling overhead for status information exchange among cells mandated in most existing channel allocation schemes. Numerical results and comparisons are given to illustrate the tradeoff.**

***Index Terms*—Call admission control, channel allocation, handoff.**

I. INTRODUCTION

THERE HAS BEEN a rapid development in wireless cellular communications. The next generation of networks are expected to eventually carry multimedia traffic—voice, video, images, or data, or combinations of them. Various issues related to this future wireless mobile networks carrying multimedia traffic have to be carefully examined before such systems can be realized [2], [20]. These include issues ranging from the physical layer, channel access protocol, channel allocation, handoff control, and etc. The quality of service (QoS) guarantee is essential in networks supporting multimedia. One of the key elements in providing QoS guarantee is an effective bandwidth allocation policy, which not only has to ensure that the network meets the QoS of the newly arriving calls if accepted, but also guarantees that the QoS of the existing calls does not deteriorate.

This paper addresses issues of the bandwidth allocation in mobile wireless networks. When a mobile moves across the boundary of a cell, handoff dropping can occur, primarily caused by the unavailability of channels in the new cell. Dropping a call in progress is generally considered to have more negative impact from users' perception than rejecting (blocking) a newly requested call. Therefore, one of the key design goals is to minimize that the handoff dropping probability. With cell sizes being systematically decreased to enable better frequency reuse, thus, increases the overall system capacity, this also considerably increases the number of handoffs during a mobile's life time. Therefore, mechanisms that ensure the handoff dropping probability to be maintained at the prespecified target QoS level, independent of the traffic conditions in the system, becomes increasingly critical in micro or picocell wireless networks [1], [20].

There are a number of unique aspects in the next generation of multimedia enabled wireless cellular networks that the design of effective bandwidth allocation schemes needs to take into account, in particular:

- Smaller cells will be employed (microcells or picocells), thus, the number of handoffs during a call's lifetime is likely to be increased; additionally, there is an increased influence from *neighboring cells* [19].
- Possibly different QoS requirements for different traffic types, and potentially more stringent QoS requirements of individual calls mandate a highly precise resource allocation [12].
- Diversified traffic load requires that bandwidth allocation has to be *adaptive* to the changing traffic pattern. Therefore, a dynamic approach is preferred [16].

To reduce handoff failure due to lack of resources in adjacent cells, a basic approach is to reserve resources for handoffs in each cell. The well-known trunk reservation scheme, also known as guard channel (GC) scheme and its numerous variations [6], [17] reserve a fixed number of channels in each cell exclusively for handoffs in a single service environment. Ramjee *et al.* proved that such a scheme is optimal for a linear objective

---

Manuscript received December 26, 2000 accepted August 27, 2001. The editor coordinating the review of this paper and approving it for publication is L.-C. Wang. The work of B. Li was supported in part by Grants from Research Grants Council (RGC) under Contract HKUST6157/98E and Contract HKUST6163/00E.

X. Luo is with the Motorola Electronics PTE Ltd., Singapore 569088 (e-mail: Clarence.Luo@motorola.com).

B. Li is with the Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong (e-mail: bli@cs.ust.hk).

L.-J. Thng is with the National Singapore University, Singapore 119260 (e-mail: eletlj@nus.edu.sg).

Y.-B. Lin is with the National Chiao Tung University, Hsinchu, Taiwan (e-mail: liny@csie.nctu.edu.tw).

I. Chlamtac is with the University of Texas at Dallas, Richardson, TX 75083–0688 USA (e-mail: chlamtac@utdallas.edu).

Publisher Item Identifier 10.1109/TWC.2002.800550.

function of call dropping and new call blocking probabilities [18].

In [3] and [5], two-tier resource allocation schemes for two types of calls were considered, i.e., a narrowband type and a wideband type. Complete sharing (CS), complete partitioning (CP), and hybrid allocation were investigated. Both schemes assume that the narrowband call can request handoffs. In [14], resource management for an integrated mobile network with real-time connections and nonreal-time connections was discussed. Bandwidth is allocated to admitted connections according to their QoS requirement. Nonreal-time connections are assumed not to request for handoff and real-time connections are assumed to have the same bandwidth requirements. Li *et al.* proposed a GC-based scheme to handle multiple streams of traffic, each having potentially different QoS requirements, thus, requiring potentially different channel thresholds [12]. All such policies are *static* in that they do not adapt to changes in the traffic pattern.

A number of recent proposals have made fine attempts to implement dynamic control in the above schemes. The proposed schemes utilize a combination of motion detection and load status broadcasting (to neighboring cells) to support multiservice handoffs. The shadow cluster concept [10] proposed by Levine *et al.* exploited the network capability to predict the motion probability of each mobile. The servicing base station maintains status of each active call, and delivery it to all cells in the shadow cluster, i.e., the cells that the mobile might visit in the future. The efficiency of these schemes is achieved by the network in predicting the potential movement of mobile hosts. The use of load status or number of active channels in neighboring cells, to determine local admission thresholds can be found in [8], [15], [16], and [18]. These schemes create dependencies on neighboring cells for local call admission process, in which frequent cell status information exchange is required among different cells. Li *et al.* in [13], showed that a local estimation algorithm can effectively eliminate the signaling overhead by using a stochastic control mechanism to dynamically adjust the number of channels needed to be reserved. Put it in a nutshell, three problems remain unanswered in supporting handoff calls in a multiservice environment:

- Implementation complexity. Most schemes assume additional function to be performed by the network (e.g., signaling and motion prediction), and require base station to have high computation capability and large buffers to store call status or load status in neighboring cells.
- The assignment of reserved channels is neglected. Most contributions have been focused on determining the total amount of resources to be reserved for all handoff calls. It is not clear which of the reserved channels are for which class of calls. If reserved resources are not distinguished for different call classes, wideband calls can encounter much higher handoff dropping probabilities than narrowband calls. This will be further elaborated in Section V.
- Complicated analytical model. Due to the modeling complexity of the multiservice system, a simple close-form formula to relate the desired QoS with the amount of resource to reserve is currently unavailable.

In this paper, we propose a measurement-based prereservation (MPr) scheme for the support of multiservice handoff with QoS guarantee on handoff dropping. The MPr scheme is adaptive to changing traffic conditions, thus, can dynamically change the amount of reserved resources. The salient features of the proposed MPr schemes are: 1) the reserved resources for each type of call with different QoS requirement can be differentiated; 2) a close-form solution can be derived according to the specified QoS levels with the amount of channels to reserve for each class of calls. This gives the wireless network provider full and easy control of the system even to the extent of changing the QoS level of any call type under widely varying load scenarios; 3) the MPr scheme preserves cell independence. It is a generic scheme that can be implemented in any cell even the adjacent cell is not compliant to the proposed scheme. It does not maintain the motion status for each call, thus, there is no requirement for information exchanges among neighboring cells. Consequently, issues leading to inefficiency of the system like late or delayed arrivals of inter-cell information does not affect an MPr-compliant cell; and 4) finally, the MPr scheme is practical for implementation as the operation for a base station is fairly simple.

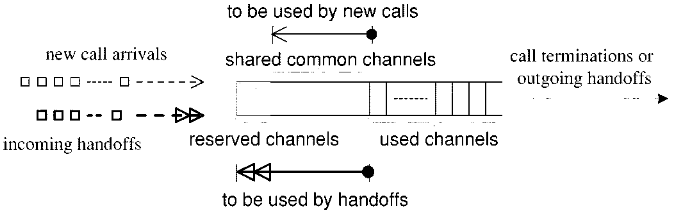


Fig. 1. Post assignment in a guard channel scheme.

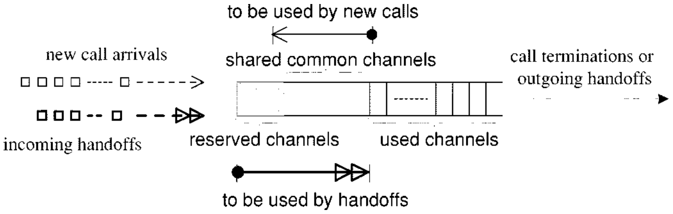


Fig. 2. Preassignment in the proposed MPr scheme.

The rest of the paper is organized as follows. We start from illustration of basic measurement-based prereservation for handoff of a single service environment in Section II and Section III. Thereafter, we present the enhancement to support handoff control in a multiservice environment in Section IV. Section V discusses the numerical results. We present the conclusion and highlight the future work in Section VI.

II. PREASSIGNMENT METHODOLOGY IN MPR SCHEME

A fundamental difference between the GC schemes and the MPr scheme is the methodology of assigning the reserved channels. Figs. 1 and 2 illustrate the resource management of the proposed MPr scheme, in comparison with the GC scheme. In a homogeneous call environment, the GC scheme employs a post-assignment methodology as illustrated in Fig. 1, where handoff calls and new calls are first assigned to the common channels. In the event the common channels are exhausted, new calls are blocked while handoff calls are assigned to the remaining channels. In the preassignment methodology of Fig. 2 used in the MPr scheme, handoff calls are first assigned to the reserved channels. If the reserved channels are used up, handoff calls will then compete with new calls for the common channels. Intuitively, the preassignment strategy can more effectively deal with multiple traffic types, as each traffic type can be preallocated with certain channels exclusively, while the GC scheme does not allow such reservations [12].

By using the prereservation scheme, the MPr scheme enjoys a desirable feature that a stated QoS can be easily "understood" and realized by the resource controller. The MPr scheme computes a reservation pool size for the stated QoS specification under the given loading conditions. It would become clear later that a smooth enhancement of the MPr scheme can be found to support multiple QoS specifications in a multiservice environment. In contrast, for a stated QoS, the number of guard channels to be reserved is cumbersome to determine, as it relies on state-transition queuing analysis as used in [17] or looking up multiple tables with different loading parameters. Although the GC scheme is optimal for a evaluation function constructed by using the linear sum of call blocking probability and handoff dropping probability [18], however, it cannot adapt to changes in the network conditions due to its static nature. In addition, in order to obtain the guard channel number, it has to compute the correspondent Markov chain with respect to the stated QoS measure, this poses considerable difficulty under changing traffic conditions. Furthermore, extending the GC scheme for handoff support a multiservice environment becomes particularly cumbersome, due to the computation complexity in solving the multidimensional Markov chains [12].

Two mechanisms are required for the MPr scheme. First, how to dynamically manage the reservation poo. Second, how to determine the reservation pool size which can adapt to the traffic condition.

The first mechanism is described as follows. Given a target reservation pool size, i.e., R_{RPS}, which provides the required handoff dropping QoS specified for the service, a mechanism is required to maintain the number of reserved channels at time t, $R(t)$, toward R_{RPS}. We refer this as a token update mechanism.

As an illustration, consider the arrival of a handoff call at time t. Due to the preassignment principle, the number of reserved channels will immediately shrink by one as follows:

$$R(t^{+}) = R(t^{-}) - 1 < R_{\mathrm{RPS}}. \tag{1}$$

The second mechanism is to determine R_{RPS} in order to satisfy the desired handoff dropping probability for a given cell. The following issues need to be addressed.

- The mechanism must be adaptive toward changing load conditions, i.e., increasing or decreasing R_{RPS} where necessary to ensure the specified handoff dropping QoS to be maintained for the cell.
- The mechanism can be deployed in the network in a distributed fashion where the mechanism runs independently within each cell.
- For generic and robust implementation, the mechanism should not require information exchanges among neighboring cells.
- The mechanism should also support multiple QoS requirements for different types of services.

In Section III, we present the two mechanisms required to perform MPr in a wireless cellular network that is operating in a homogeneous call environment.

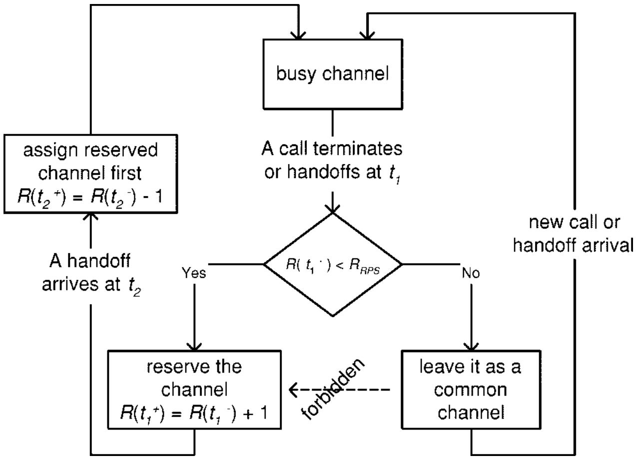


Fig. 3. State transition of a radio channel.

III. MPr FOR A HOMOGENEOUS CALL ENVIRONMENT

A typical homogeneous service environment often assumes that each call, i.e., handoff call or new call, occupies one unit of bandwidth, or a channel. It is also assumed following the convention that new call arrivals and handoff call arrivals are Poisson processes. In addition, the call life time and call dwelling time within a cell are assumed to be exponentially distributed.

A. Token Update of $R(t)$

As mentioned before, a mechanism is required to update $R(t)$ (the number of reserved channels at time t) toward a given R_{RPS} (the reservation pool size). Fig. 3 illustrates the basic principle in token update schemes adopted in the MPr algorithm.

Fig. 3 illustrates a very important property of the MPr scheme which we refer to as the *token update*, specifically, a channel can only be reserved at time t if the following three conditions are true.

- At time t^{-}, the channel was in the busy state, servicing a call.
- At time t^{-}, $R(t^{-}) < R_{\mathrm{RPS}}$.
- At time t, the above channel is released by the call due to termination or handoff.

Simply put, a released channel will be returned to the reservation pool given $R(t^{-}) < R_{\mathrm{RPS}}$. Notice, however, this does not require that a channel be "*tagged*," a counter recording the current value of $R(t^{-})$ will serve the purpose. When a new call arrives, it can only be admitted *if and only if* the number of unused channels is more than $R(t^{-})$. Otherwise, all unused channel(s) if any are reserved for handoff. The token update also indicates the following.

- A channel that is initially designated to the common pool cannot be switched over directly to the reservation pool. It has to become a busy state first before it can be considered for designation in the reservation pool. See Fig. 3.
- If the number of reserved channels have not reached the desired the number, i.e., $R(t^{-}) < R_{\mathrm{RPS}}$, then a channel, upon returning to the system when a call terminates or handoffs, is immediately reserved.
- If the reservation pool is exhausted, i.e., $R(t) = 0$, then a handoff call at time t will have to compete with new calls for an idle channel in the common pool. A strict dropping

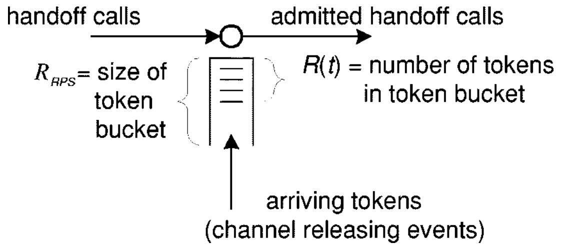


Fig. 4. The analytical model of the MPr scheme.

policy may be employed, which drops a handoff request under such a condition, but is not encouraged for the same reasons as discussed in Section I. Comparison of such a variation with the normal admission condition for handoff calls will be presented in the numerical section.

The token-property update policy holds the key to the simplicity of the MPr scheme and contributes, to a certain extent, some of the favorable features of the MPr scheme. This will be elaborated further in Section IV.

B. Update of $R_{\rm RPS}$

With the token-bucket update policy of $R(t)$ in place, it is now rather easy to develop a token bucket model for updating $R_{\rm RPS}$ as illustrated in Fig. 4. In the figure:

- The size of the token bucket represents the $R_{\rm RPS}$, i.e., the reservation pool size.
- Arriving tokens at the bucket ingress represent events where channels are released (by call handoff or call termination).
- If the bucket is full, i.e., $R(t) = R_{\rm RPS}$, then any arriving tokens are discarded, meaning that a released channel is returned to the common channel pool.
- Arriving handoff calls are assigned reserved channels if tokens are available. If the token bucket is empty (i.e., reservation pool is used up), then there is no guarantee the handoff call will be admitted as it depends on the availability of common channels.

The token bucket can be modeled as a $M/M/1$ queue system with finite capacity $R_{\rm RPS}$. We shall see that the possibility for a handoff to be accepted when there is no channel reserved and is fairly low in the proposed MPr scheme. The argument is that there is no need to control the new call arrival when the system is under light traffic loading. In other words, the stringent QoS of the handoff dropping probability only needs to be ensured under heavy or over-loading case. Such a case requires two conditions to be satisfied. First, the free channel must be released at the time the reservation pool was full so that it was not captured. Second, during the period handoff calls arrive and use up all the reserved channels, no new call attempts within the radio cell utilize that particular free channel. This is because a released channel will be first assigned another reserved channel, especially when the reserved channel pool is near exhaustion. Therefore, the probability that the reservation pool is empty effectively gives a tight upper bound for handoff dropping probability.

If the handoff dropping QoS probability is specified to be $P_{\rm QoS}$, then the MPr scheme has to determine a suitable token bucket size to ensure that

$$P_{\rm QoS} = P(\text{handoff call droppped}) \leq P(\text{token bucket empty}). \tag{2}$$

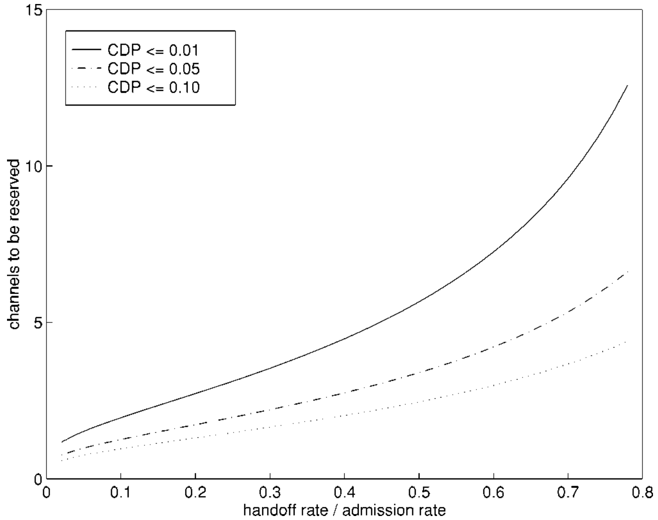


Fig. 5. Channels to be reserved over $1/\rho$.

Employing the $M/M/1/K$ queue principles [11], we note that

$$P(\text{token bucket is empty}) = \frac{1-\rho}{1-\rho^{R_{\rm RPS}-1}} \tag{3}$$

$$\text{where } \rho = \frac{\text{channel releasing rate}}{\text{incoming handoff rate}} = \frac{\mu_r}{\lambda_h}. \tag{4}$$

Incorporating the above equations, we obtain the update condition for minimum $R_{\rm RPS}$ as follows:

$$R_{\rm RPS} \geq \frac{\ln \frac{P_{\rm QoS}+\rho-1}{P_{\rm QoS}}}{\ln \rho} - 1. \tag{5}$$

It is noted that the update for $R_{\rm RPS}$ in (5) only requires local cell metrics μ_r and λ_h—this is precisely the strength of the MPr scheme in terms of robustness and generic deployment in a wireless cellular network where not necessarily all cells are MPr compliant.

C. Fine Tuning of the $R_{\rm RPS}$ *Update*

There are a number of issues regarding the update equation of $R_{\rm RPS}$ in (5), the first is the estimate of the channel releasing rate μ_r which can be replaced with

$$\mu_r = \lambda_a = \text{successful call admittance rate}. \tag{6}$$

Therefore, there is no necessity to incorporate a MPr device to monitor the channel releasing rate as the call admission control device of the cell can indirectly perform that function.

It is now clear that the $1/\rho$ represents the percentage of incoming handoff calls in the total successful call attempts, provided handoff failure is negligible because of the reserved resources. By monitoring ρ in (5), $R_{\rm RPS}$ can be adaptively adjusted. This is better illustrated in Fig. 5 which plots $R_{\rm RPS}$ against $1/\rho$ for various $P_{\rm QoS}$. Notice that as the percentage of handoff rate increases, so does the number of channels to be reserved. Intuitively, the more stringent the QoS requirement, the more channels need to be reserved.

In addition, there is no stringent requirement on the refresh frequency of λ_a and λ_h, neither the identical update period for every cell in [15], nor the periodical flooding of call status to

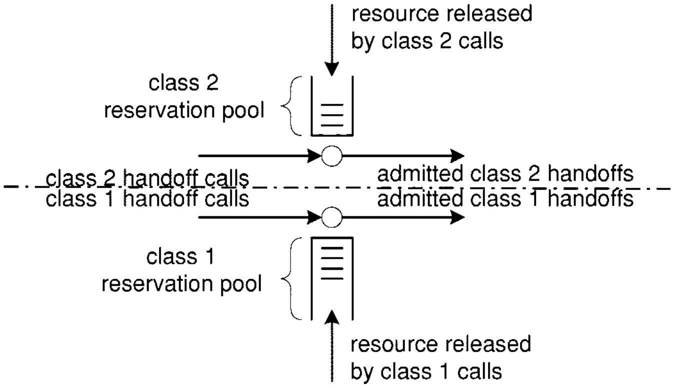


Fig. 6. A complete partitioning multiservice environment.

neighboring cells as in [10]. Once every few minutes seems to be adequate. Chances are that channels are reserved but handoff attempts decrease afterwards, the arrival rate would decrease drastically. In such a case, excessively reserved channels will be released upon the above mentioned measurement, is refreshed and a new size of the reservation pool is recomputed. To smooth out random or drastic measurement inaccuracies, an exponential smoothing average formula can be employed [13] to obtain a moving average λ_{ave} as follows:

$$\lambda_{\text{ave}}(n) = \alpha\lambda_{\text{ave}}(n-1) + (1-\alpha)\lambda_{\text{new}} \tag{7}$$

where $\lambda_{\text{ave}}(n)$ is the average new call arrival rate up to now, and the λ_{new} is the observed call arrival rate in the nth period. More details including the selection of the value of α can be found in [13].

IV. MPR SUPPORT FOR MULTISERVICE HANDOFF CALLS

In a multiservice environment with mixed call types, either of the following reasons require the cell to differentiate handoffs belonging to different call types: 1) different call i has different QoS requirements in terms of handoff dropping, i.e., $P_{\text{QoS}i}$; and 2) different call i has different channel requirement B_i.

In this section, for illustration, we show how the MPr algorithm can be used in both a CP environment and a resource sharing (RF) environment. In principle, the MPr scheme can also apply to other bandwidth partitioning strategies [9], [14].

A. Multiservice With CP Resource Allocation

With CP resource allocation, the allocated bandwidth boundary for each service class is clearly determined. Thus, an incoming class i call consumes bandwidth allocated for class i and an outgoing class i call returns the resources back to its own class. Consequently, the handoff reservation pool for each class is also separated from each other and belongs strictly to each class. In this environment, independent MPr mechanisms outlined in Section III can be used to update the required reservation pool for each class as shown in Fig. 6. Thus, the MPr mechanism in this scenario is just a trivial extension of the one in the homogeneous call environment. A point to note is that (5) should take into consideration the different amount of bandwidth each class of call requires. Thus, (5) should be modified to

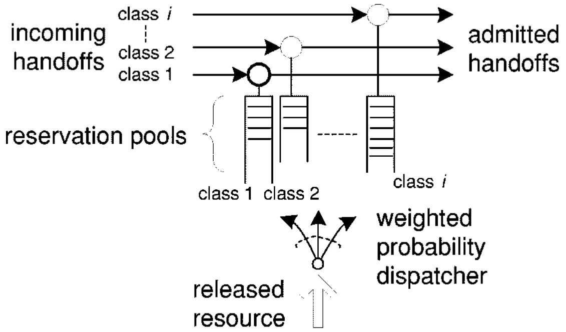


Fig. 7. A resource sharing multiservice environment.

$$R_{\text{RPS}_i} \geq B_i \times \left(\frac{\ln \frac{P_{\text{QoS}_i}+\rho_i-1}{P_{\text{QoS}_i}}}{\ln \rho_i} - 1 \right) \tag{8}$$

where R_{RPS_i} is the number of channels to be reserved in class i and

$$\begin{aligned}\rho_i =& \frac{\mu_{ri}}{\lambda_{hi}} = \frac{\text{call releasing rate for class } i}{\text{incoming handoff rate for class } i}\\ =& \frac{\lambda_{ai}}{\lambda_{hi}} = \frac{\text{successful call admittance rate for class } i}{\text{incoming handoff rate for class } i}.\end{aligned} \tag{9}$$

B. Multiservice Resource Sharing Environment

In a resource sharing multiservice environment, the bandwidth of the cell is not separated and is shared by different call types. In order to support the different handoff requirements, a total of i independent MPr processes are required, each maintaining its own respective reservation pool for each class i as illustrated in Fig. 7. An incoming class i handoff call will immediately be assigned reserved channels corresponding to its class. And as before, whichever released channels not reserved (i.e., the selected token buffer is full) merely becomes part of the pool of common channels.

However, a difficulty is now to decide which token buffer in Fig. 7 deserves the arriving token (arising out of resource releasing events). Specifically, the difficulty lies in characterizing the overall channel releasing process and then designing an appropriate scheduling mechanism that preserves the $M/M/1/K$ model for each reservation pool class. The difficulty can be overcome by the use of a fictitious model rather than the actual model. Then, based on the fictitious model, an appropriate and simple dispatching mechanism can be developed, which is also applicable for use in the original model. Finally, the appropriate update equations for the various reservation pool classes of the original model are provided. We next describe how this fictitious model can be constructed.

1) The Overall Channel Releasing Process: The overall channel releasing process associated with Fig. 7 is not a single Poisson process but a sum of Poisson processes with bulk arrivals. The class i call completion or outgoing-handoff process would return resource to the system at the rate of μ_{ri} with bulk arrival B_i, where B_i is units of bandwidth used by class i calls. Characterizing the overall channel releasing process is, in general, rather difficult. Here, we introduce an approximation by using a fictitious system which will be shown to accurately represent the real system.

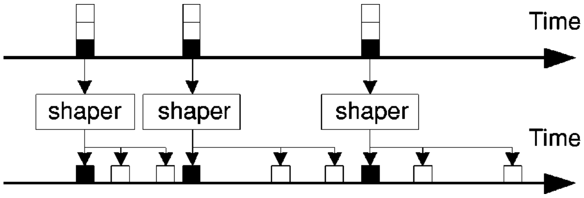


Fig. 8. Shaping the bulk arrival into a more inferior system.

Due to the memoryless nature of the Poisson process, a conventional Poisson process at the rate μ'_{ri} can be constructed from a given bulk arrival Poisson process with rate μ_{ri} and bulk size B_i via shapers, which distribute/delay exponentially (in time) the bulk arriving channels, as illustrated in Fig. 8. If we construct all the bulk arrival channel releasing processes of the cell to be exponentially shaped to conventional Poisson processes, then the overall channel releasing process of the fictitious model is a sum of conventional Poisson processes. Consequently, the overall channel releasing process of the fictitious model is a well characterized conventional Poisson process with rate μ'_r given as follows:

$$\mu'_r = \sum_i \mu_{ri} B_i = \sum_i \lambda_{ai} B_i \tag{10}$$

where λ_{ai} is defined in (6).

The model is an approximation since, in order to distribute the bulk arrivals to exactly construct a conventional Poisson process, shapers ought to have prior knowledge of future bulk arrival events—which is impractical. Nevertheless, the fictitious domain is useful for further analysis, not only because its overall channel releasing process is well characterized, but it is also noted that in the fictitious system, its channel reservation process is inferior compared to the actual system's reservation process since shapers in the fictitious system introduce artificial delays to arriving resources that can otherwise be immediately reserved. Thus, by assuming the fictitious domain characterized by (10), it is clear that the eventual MPr update equations should perform at least equal or better in the actual system since the bulk arriving channel releasing processes are not delayed.

2) Update of R_{RHSi}: The consumption rate of each reservation pools shown in Fig. 7 is easily characterized. Since each arriving handoffs will consume B_i channels, the channel consumption rate λ'_{hi} experienced by the class i reservation pool when class i handoffs are arriving at the rate λ_{hi} is

$$\lambda'_{hi} = B_i \times \lambda_{hi}. \tag{11}$$

With the overall channel releasing process characterized, we designed a simple scheduling algorithm to dispatch the released resources to one of reservation pools for call classes which require handoff support. Referred to as the Weighted Probability Dispatcher, it is described as follows:

> *Weighted Probability Dispatcher*: For each unit of channel returning to the system, the dispatching mechanism chooses token buffer i with probability $B_i\lambda_{hi}/\sum_i B_i\lambda_{hi}$ as the selected token buffer.

If the selected class i reservation pool is not full, the released channel is immediately reserved as a class i channel. Otherwise, the channel is returned to the common channel pool. By using this dispatching mechanism, it is clear that the *channel refilling rate* of the reservation pool for class i handoffs is as follows:

$$\mu'_{ri} = \mu'_r \times \frac{B_i\lambda_{hi}}{\sum_i B_i\lambda_{hi}} \tag{12}$$

where μ'_r, as defined in (10), is the overall rate of the channel releasing process associated with the fictitious system.

For class i handoff calls with bandwidth requirement B_i, the *token refilling rate* is μ'_{ri}/B_i, by counting B_i unit of bandwidth as one token.

Therefore, the appropriate reservation pool size to support class i handoff calls with QoS requirement in terms of handoff failure probability P_{QoS} is

$$R_{\text{RPS}i} = B_i \times \left(\frac{\ln \frac{P_{\text{QoS}_i} + \rho'_i - 1}{P_{\text{QoS}_i}}}{\ln \rho'_i} - 1 \right) \tag{13}$$

where

$$\mu'_{ri}\rho'_i = \frac{\frac{\mu'_{ri}}{B_i}}{\lambda_{hi}} = \frac{\mu'_{ri}}{\lambda_{hi}B_i}. \tag{14}$$

In summary, by the inclusion of the weighted probability dispatcher and the periodic updating of λ_{hi} and λ_{ai}, i.e., traffic loading conditions, the required reservation pool sizes for the various class of services can be dynamically adjusted by (13). This sums up the MPr scheme for handoff support in a resource sharing multiservice environment. Finally, it is noted that in the MPr scheme of Fig. 7, the reservation pools are clearly distinguished for the different class of services.

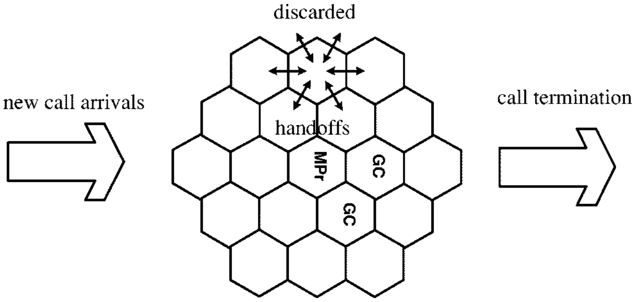


Fig. 9. A cellular system with three rings of radio cells.

V. NUMERICAL RESULTS

The simulation model used here is a hexagonal system with three rings of radio cell clusters (total 9 cells) as shown in Fig. 9. To alleviate the *finite size effects*, we implement periodic connections on the three pairs of opposite sides of the cluster (wraparound). We investigated two scenarios, one for single type of service and the other with two call types. Both the call life time and channel holding time are set to be exponentially distributed. We tag individual cells and obtain performance metrics by simulation. Note the generic feature of the MPr scheme that there is no requirement that all cells must be MPr compliant.

A. A Homogeneous Call Scenario

Fig. 10 shows the handoff dropping probability (CDP) and call blocking probability (CBP) of the MPr scheme with a target CDP below 0.05 for a homogeneous call environment. Each call

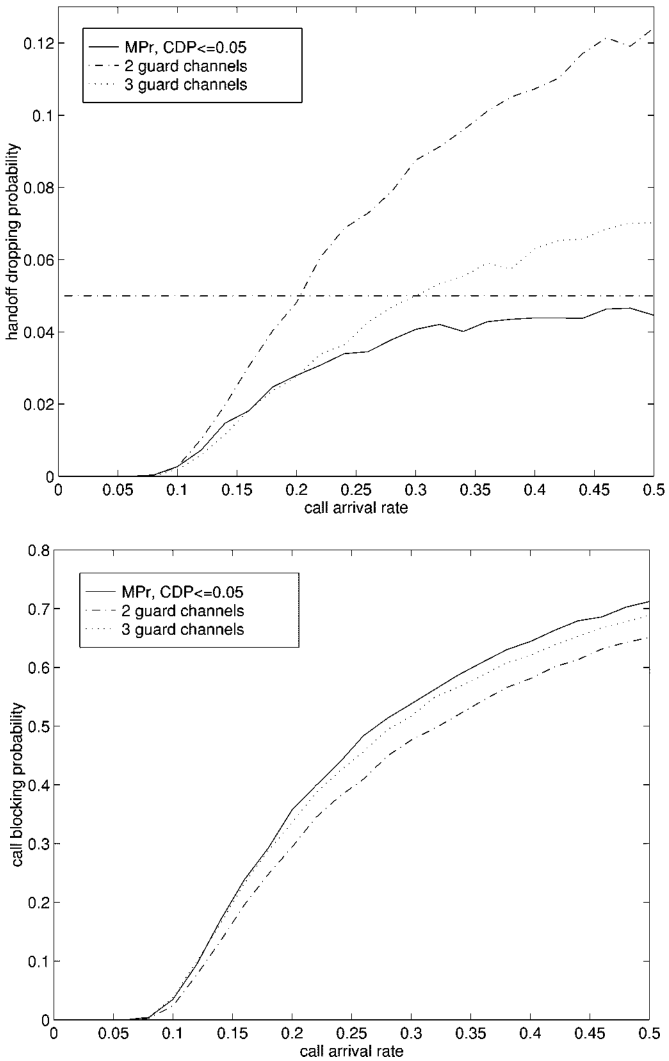
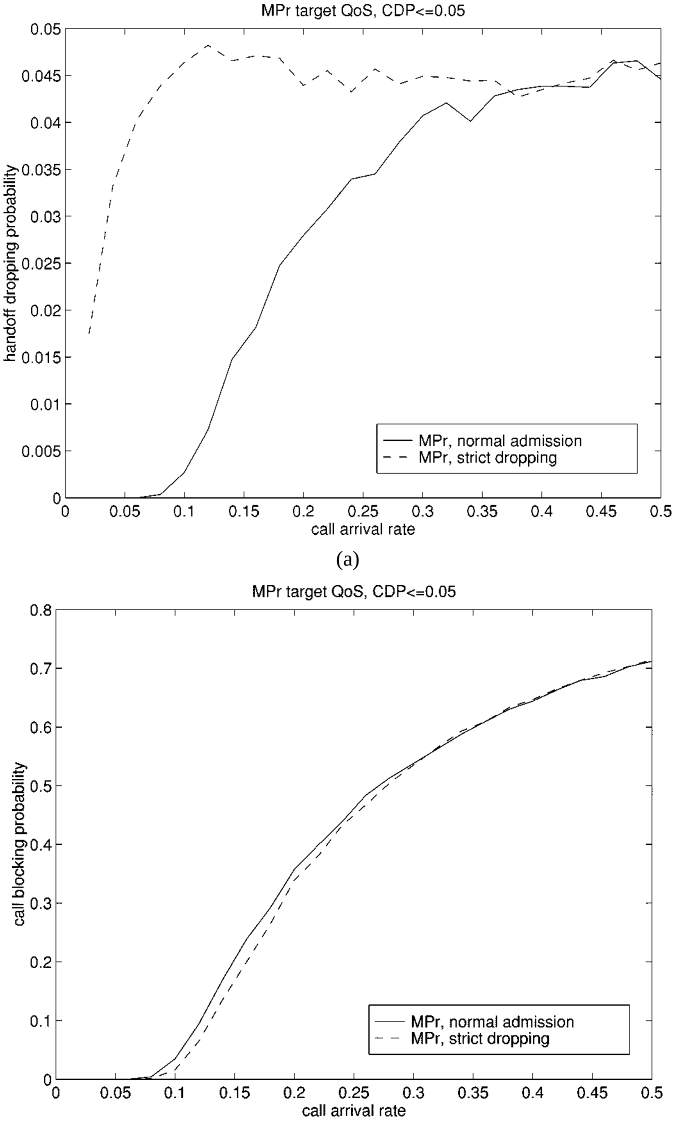


Fig. 10. CDP and CBP versus arrival rate, single service.



(a)

(b)

Fig. 11. The strict dropping policy and normal admission for handoff.

occupies one unit of bandwidth. The cell capacity is 30. The mean call lifetime is 200 s and the mean channel holding time for a call to stay in one cell is 100 s. Under such a setting, a mobile hands off twice on average during its life time.

For comparison, we present the results of the GC scheme with two and three guard channels, as they achieve comparable performance in terms of call dropping probability as well as call blocking probability, under light or mild load condition. When load increases, the CDP of GC scheme soars. However, the MPr scheme can well maintain the CDP below the target level. By ensuring the handoff dropping QoS, a tradeoff occurs where the corresponding blocking probability of the MPr scenario is higher than either guard channel scenarios, because more channels are reserved for handoff calls.

It should also be noted that for a stated QoS specification on call dropping, the MPr scheme is self-driven while the GC scheme requires either table lookup or the calculation of a two dimensional Markov chain as presented in [17].

B. The Strict Dropping Policy

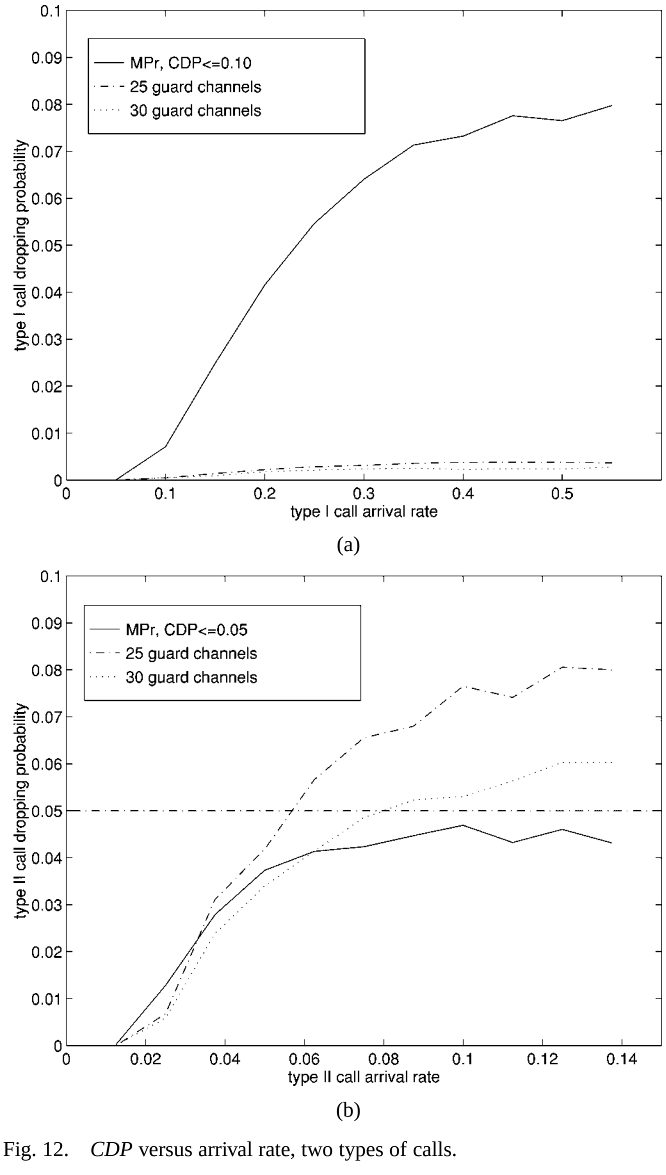
The performance of the strict dropping policy in a homogeneous call environment is also presented for comparison in Fig. 11. In strict dropping policy, handoff calls are accepted only when there are channels available in the reservation pool, whereas in normal dropping policy, handoffs are accepted as long as there are unused channels. The strict dropping shows a worse case than the normal admission condition for handoff calls. It proves that the correct amount of resources is reserved to guarantee the QoS specification against handoff dropping by the MPr scheme. The lower call dropping probability for the normal admission under low load condition results from the competition for unused common channels, which are not guaranteed to be available under all loading conditions. Although it is shown in Fig. 11(a) the expected QoS on call dropping is not violated, the strict dropping policy is not desired for the same reasons as discussed in Section I. However, also observe from the Fig. 11, that both schemes ensure that CDP measures are not violated, and as the traffic loading increase, the difference between the two schemes diminishes.

C. A Multiservice Scenario

We next discuss a multiservice scenario with two call types. Both types of calls may use all the available channels for the cell.

TABLE I
PARAMETERS FOR TWO TYPES OF CALLS

Call Parameters	Type I	Type II
mean call life time	300s	300s
mean dwelling time	100s	150s
bandwidth requirement	1	12
call arrival rate	λ_1	$\lambda_2 = 0.25\lambda_1$
target maximum call dropping	0.10	0.05



(a)

(b)

Fig. 12. *CDP* versus arrival rate, two types of calls.

We only consider the case with resource sharing (RS), as the complete partitioning case, not presented, is a trivial extension of the single-service scenario. The specifications of the two call types are shown in Table I. Note that the wideband call arrival rate is *four times* smaller than that of a narrowband call and its handoff QoS requirements more stringent than the narrowband call. Also note that the two types of calls have different motion characteristics.

Fig. 12 illustrates the call dropping probabilities obtained using the MPr scheme in comparison with GC schemes. We notice the overwhelming favoring of narrowband calls compared to wideband calls in the GC cells, which is one of the key deficiencies employing simple GC scheme. The reason is that under heavy traffic loading, each time when a channel is released, the narrowband traffic will immediately occupy the channel before the wideband calls have a chance. However, observe from Fig. 12 that the MPr compliant cells do not overfavor the narrowband calls, but maintain the stated QoS against handoff dropping for each type of calls, i.e., 0.10 for narrowband calls in Fig. 12(a) and 0.05 for wideband calls for all loading conditions Fig. 12(b). This is anticipated in that each type has its own reservation pool, respectively.

VI. CONCLUSION

In this paper, we have presented a novel measurement-based prereservation scheme, known as MPr technique, to reduce handoff failure in mobile wireless networks with mixed-call types. The proposed MPr scheme significantly reduces the computation complexity in deriving the reserved pool size while at the same time eliminates the need for inter-cell signaling using local status estimation. We further demonstrate that the MPr scheme is capable of supporting multiple connection-level QoS for handoff calls of different types. In addition, the MPr scheme is very simple and efficient for implementation in any heterogeneous cellular network where not all cells are MPr compliant.

The existing work [4], [7] showed that data traffic does not obey Erlang distribution, largely used in voice centric cellular networks. We are currently extending the MPr model to accommodate data traffic with more realistic traffic characteristics assumptions.

REFERENCES

[1] A. Acampora and M. Naghshineh, “Control and quality-of-service provisioning in high speed micro-cellular networks,” *IEEE Pers. Commun.*, vol. 1, pp. 36–43, 2nd quarter 1994.

[2] I. F. Akyildiz, J. McNair, J. Ho, H. Uzunalioglu, and W. Wang, “Mobility management in next generation wireless systems,” *IEEE Proc.*, vol. 87, pp. 1347–1384, Aug. 1999.

[3] J. Chen and M. Schwartz, “Reservation strategies for multimedia traffic in a wireless environment, performance summary,” in *Proc. IEEE PIMRC’95*, Sept. 1995, pp. 1067–1072.

[4] M. Cheng and L.-F. Chang, “Wireless dynamic channel assignment performance under packet data traffic,” *IEEE J. Select. Areas Commun.*, vol. 17, pp. 1257–1269, July 1999.

[5] B. Epstein and M. Schwartz, “Reservation strategies for multimedia traffic in a wireless environment,” in *Proc. IEEE Vehicular Technology Conf.*, 1995, pp. 165–169.

[6] M. Fang, I. Chlamtac, and Y.-B. Lin, “Channel occupancy times and handoff rate for mobile computing and PCS networks,” *IEEE Trans. Comput.*, vol. 47, pp. 679–692, 1998.

[7] Y. Fang and I. Chlamtac, “Teletraffic analysis and mobility modeling of PCS networks,” *IEEE J. Select. Areas Commun.*, vol. 17, pp. 1062–1072, July 1999.

[8] S. Jiang, B. Li, X. Luo, and H. K. Tsang, “A modified distributed call admission control scheme and its performance,” *ACM/Baltzer J. Wireless Networks*, vol. 7, pp. 127–138, Mar./Apr. 2001.

[9] Y. Kim and C. Un, “Analysis of bandwidth allocation strategies with access restriction in broadband ISDN,” *IEEE Trans. Commun.*, vol. 41, pp. 771–781, May 1993.

[10] D. A. Levine, I. F. Akyildiz, and M. Naghshineh, “A resource estimation and call admission algorithm for wireless multimedia networks using the shadow cluster concept,” *IEEE/ACM Trans. Networking*, vol. 5, pp. 1–12, Feb. 1997.

[11] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering*, 2nd ed. Reading, MA: Addison-Wesley.

[12] B. Li, C. Lin, and S. Chanson, "Analysis of a hybrid cutoff priority scheme for multiple classes of traffic in multimedia wireless networks," *ACM/Baltzer J. Wireless Networks*, vol. 4, pp. 279–290, Aug. 1998.

[13] B. Li, L. Yin, K. Y. M. Wong, and S. Wu, "An efficient and adaptive bandwidth allocation scheme for mobile wireless networks based on on-line local parameter estimations," *ACM/Baltzer J. Wireless Networks*, vol. 7, pp. 107–116, Mar./Apr. 2001.

[14] M. Naghshineh and A. Acampora, "QoS provisioning in micro-cellular supporting multimedia traffic," in *Proc. IEEE Infocom*, 1995, pp. 1075–1085.

[15] M. Naghshineh and M. Schwartz, "Distributed call admission control in mobile/wireless networks," *IEEE J. Select. Areas Commun.*, vol. 14, pp. 711–717, May 1996.

[16] C. Oliveira, J. B. Kim, and T. Suda, "An adaptive bandwidth reservation scheme for high-speed multimedia wireless networks," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 858–874, Aug. 1998.

[17] E. Posner and R. Guerin, "Traffic policies in cellular radio that minimize blocking of handoff calls," in *Proc. ITC-11*, Kyoto, 1985, pp. 294–298.

[18] R. Ramjee, R. Nagarajan, and D. Towsley, "On optimal call admission control in cellular networks," in *IEEE Infocom'96*, San Francisco, CA, Mar. 1996, pp. 43–50.

[19] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[20] M. Schwartz, "Network management and control issues in multimedia wireless networks," *IEEE Pers. Commun.*, pp. 8–16, June 1995.

[21] S. Wu, K. Y. M. Wong, and B. Li, "A new, distributed dynamic call admission policy for mobile wireless networks with QoS guarantee," *IEEE/ACM Trans. Networking*, pp. 257–271, Apr. 2002.

[22] "3GPP, 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects QoS Concept," 3GPP, 3G TR 23.907 Version 1.2.0, 1999.

Xiao Yuan Luo (S'97–M'00) received the B.S. degree in electronics and information system from Zhong Shan (Sun Yet-San) University, China, in 1996, and the M.Eng. degree in electrical engineering from the National University of Singapore, in 1999.

He is a System Engineer with Motorola Electronics PTE Ltd., Singapore.

Bo Li (S'89–M'92–SM'99) received the B.S. (*summa cum laude*) and M.S. degrees in computer science from Tsinghua University, Beijing, P. R. China, in 1987 and 1989, respectively, and the Ph.D. degree in computer engineering from the University of Massachusetts at Amherst, MA, in 1993.

Between 1994 to 1996, he worked on high performance routers and ATM switches in IBM Networking System Division, Research Triangle Park, NC. Since then, he has been with the Computer Science Department, Hong Kong University of Science and Technology, Kowloon, China. His current research interests are: wireless mobile networking supporting multimedia, video multicast, and all optical networks using WDM.

Dr. Li is a member of ACM. He has been on the editorial board of *ACM Mobile Computing and Communications Review* (MC2R), *ACM/Baltzer Journal of Wireless Networks* (WINET), IEEE Journal on Selected Areas in Communications (JSAC)—Wireless Communication Series, *SPIE/Kluwer Optical Networking Magazine* (ONM), *KICS/IEEE Journal of Communications and Networks* (JCN), IEEE Transactions on Wireless Communications and IEEE Transactions on Vehicular Technology. He has been a guest editor for *IEEE Communications Magazine* Special Issue on Active, Programmable, and Mobile Code Networking (April 2000), IEEE Journal on Selected Areas in Communications Special Issue on Protocols for Next Generation Optical WDM Networks (October 2000), *ACM Performance Evaluation Review* Special Issue on Mobile Computing (December 2000), and *SPIE/Kluwer Optical Networks Magazine* Special Issue on Wavelength Routed Networks: Architecture, Protocols, and Experiments (January/February 2002). He is currently serving as the lead guest editor for IEEE Journal on Selected Areas in Communications Special Issue on Recent Advance in Service Overlay Networks (to be published in the 4th quarter 2003). In addition, he has been involved in organizing over 30 conferences, in particular IEEE Infocom since 1996.

Ian Li-Jin Thng received the Ph.D. degree from Curtin University, Perth, Australia.

He is currently working at the National University of Singapore as an Assistant Professor. Prior to that, he was working as a Postdoctoral Fellow in the Australian Telecommunications Research Institute, Curtin University. His research interest includes very high speed digital communications, long range dependant traffic, broadband array signal processing, and wireless networking.

Yi-Bing Lin received the B.S.E.E. degree from National Chiao Tung University, Hsinchu, Taiwan, in 1983, and the Ph.D. degree in computer science from the University of Washington, Washington, DC, in 1990.

From 1990 to 1995, he was with the Applied Research Area, Bell Communications Research (Bellcore), Morristown, NJ. In 1995, he was appointed as a Professor of the Department of Computer Science and Information Engineering (CSIE), National Chiao Tung University (NCTU). In 1996, he was appointed as Deputy Director of Microelectronics and Information Systems Research Center, NCTU. During 1997–1999, he was elected as Chairman of CSIE, NCTU. His current research interests include design and analysis of personal communications services network, mobile computing, distributed simulation, and performance modeling.

Dr. Lin is an Associate Editor of *IEEE Network*, Editor of IEEE Journal on Selected Areas in Communications: Wireless Communications Series, Editor of *IEEE Personal Communications Magazine*, Editor of *Computer Networks*, Area Editor of *ACM Mobile Computing and Communication Review*, a Columnist of *ACM Simulation Digest*, Editor of *International Journal of Communications Systems*, Editor of *ACM/Baltzer Wireless Networks*, Editor of *Computer Simulation Modeling and Analysis*, Editor of *Journal of Information Science and Engineering*, Program Chair for the 8th Workshop on Distributed and Parallel Simulation, General Chair for the 9th Workshop on Distributed and Parallel Simulation, Program Chair for the 2nd International Mobile Computing Conference, Guest Editor for the ACM/Baltzer MONET special issue on Personal Communications, Guest Editor for IEEE Transactions on Computers Special Issue on Mobile Computing, and a Guest Editor for *IEEE Communications Magazine* Special Issue on Active, Programmable, and Mobile Code Networking. Dr. Lin is the author of the book, *Wireless and Mobile Network Architectures* (coauthor with Imrich Chlamtac; New York: Wiley). He received 1998 and 2000 Outstanding Research Awards from National Science Council, ROC, and the 1998 Outstanding Youth Electrical Engineer Award from CIEE, ROC. He is an Adjunct Research Fellow of Academia Sinica.

Imrich Chlamtac (M’86–SM’86–F’93) received the B.Sc. and M.Sc. degrees awarded with the Highest Distinction from Tel Aviv University, Israel, and the Ph.D. degree in computer science from the University of Minnesota, Minneapolis, MN.

Since 1997, he holds the Distinguished Chair in Telecommunications at the University of Texas at Dallas, Richardson, TX, where he is also the Director of the Center for Advanced Telecommunications Systems and Services (CATSS). Prior to joining University of Texas at Dallas, he was on the faculty at Technion—The Israel Institute of Technology, Haifa, Israel and the University of Massachusetts, Amherst, MA. He is also the recipient of the honorary titles of Senior Professor at Tel Aviv University, The Bruno Kessler Honorary Professor at the University of Trento, Italy, and University Professor at the Technical University of Budapest. He is the founder of Consip Ltd., Haifa, Israel, cofounder and past CEO of BCN Inc., Boston, MA. He published over two hundred and fifty papers in refereed journals and conferences. He is the coauthor of the first textbook on Local Area Networks (Lexington Books, 1981, 1982, 1984) and of Wireless and Mobile Network Architectures (New York: Wiley, 2000).

Dr. Chlamtac is a Fellow of the ACM societies for numerous contributions to networking, including the concept of lightpath or wavelength routing and contention free channel access methods. He is the *IEEE Network* 2000 Editor’s Choice and Amazon.com engineering best-seller. He is an IEEE Distinguished Lecturer and was a Fulbright Scholar. He is the winner of the 2001 ACM award for contributions to wireless and mobile networks, the Society for Computer Simulation Award for the introduction of network emulators, and has received multiple best paper awards in wireless and optical networking. He serves as the founding Editor in Chief of the *ACM/URSI/Kluwer Wireless Networks* (WINET) and the *ACM/Kluwer Mobile Networks and Applications* (MONET) journals. He is the founding Editor of the *SPIE/Kluwer Optical Networks* (ONM) Magazine and has served on the editorial boards of most major publications in telecommunications. He is also the founder and Steering Committee Chair of ACM/IEEE MobiCom and SPIE/IEEE/ACM OptiComm conferences and the founder and past Chairman of the ACM SigMobile.