*Wireless Personal Communications* **26:** 179–202, 2003.


# Internetworking between HIPERLAN/2 and UMTS

KWANG-CHENG CHEN and CHUN-YING WU

*Graduate Institute of Communication Engineering, National Taiwan University, R246 EE Building 2, 1 Roosevelt Rd. Sec. 4, Taipei, 10617 Taiwan, R.O.C.*
*E-mail: chenkc@ieee.org, cywu_jeff@ieee.org*

**Abstract.** Wireless LAN has been widely considered for wireless broadband networks with low mobility, which can be generally considered as a matching part for future all-IP cellular networking. This scenario imposes challenges on providing seamless services among different networks. In this paper, we propose the architecture of internetworking between HIPERLAN/2 and UMTS networks. The framework of the integrated system is proposed and analyzed to evaluate the system performance in terms of scheduling and admission control.

**Keywords:** internetworking, UMTS, HIPERLAN/2, Wireless Local Area Networks.

## 1. Introduction

Wireless Local Area Network (WLAN) has been seriously developed as low-mobility wireless broadband access technology, which can be considered as an integrated part of future all-IP wireless cellular communicatons/networks to support different application scenarios. Current WLAN is widely deployed up to 11 Mbps transmission speed using carrier sense multiple access with collision avoidance (CSMA/CA) under the framestructure of the IEEE 802.11b. IEEE 802.11a/g and ETSI HIPERLAN/2 support up to 54 Mbps orthogonal frequency division multiplexing (OFDM) transmission, while HIPERLAN/2 using a medium access control closer to cellular architecture. An appropriate and efficient integration of WLAN and 3G cellular (such as UMTS) is thus a critical step toward all-IP cellular networking.

The internetworking between two heterogeneous network technologies, especially when one is a wide area circuit-switched network and the other is a local area IP-based network, is sometimes called *overlay networking*, which is the unification of several heterogeneous networks, of varying coverage and performance, into a single logical network that provides coverage that is the union of the networks' coverage with performance corresponding to the best network in range [1]. Roaming between different networks invokes vertical handoff (or *intertech roaming*) procedures [2]. Much of the related works focus on internetworking between IEEE 802.11 WLAN and GPRS (General Packet Radio Service) [3, 4]. In [3], handoff problem between IEEE 802.11 and GPRS is approached in two ways: *mobility gateway/proxy-based architecture* and *Mobile-IP-based architecture*. In the first method, an intermediate server called mobility gateway is placed in the network so that any traffic to and from the mobile terminals is forced to pass through it. The second method employs mobile IP to restructure connections when mobile terminals roams from one data network to another.

In this paper, we propose the architecture to integrate HIPERLAN/2 into UMTS network. The proposed architecture is based on general UMTS system architecture, which consists of three major parts: core network (CN), UMTS terrestrial radio access network (UTRAN) and

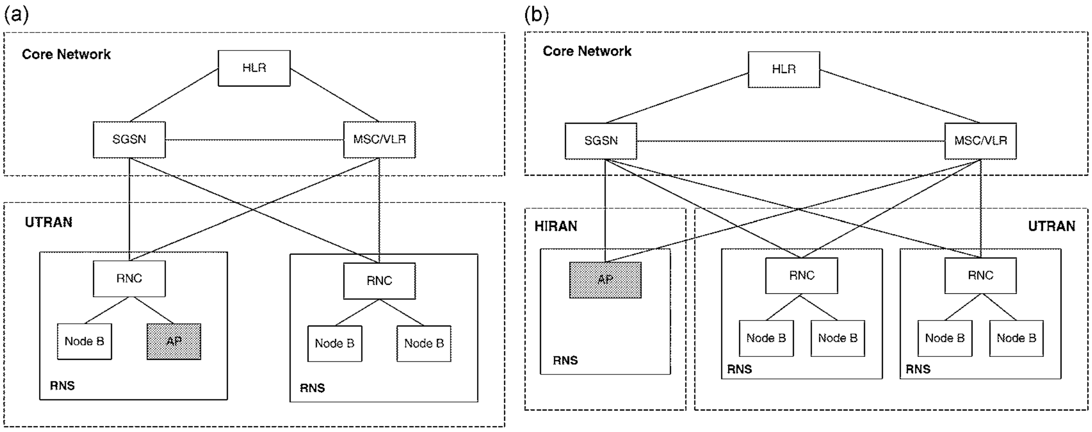


*Figure 1.* Two possible implementations of HIPERLAN/2 as an access network to UMTS core network. (a) AP as a Node B. (b) AP as an RNC.

user equipment (UE). The general UMTS system architecture is flexible in the sense that it allows for different access networks other than UTRAN, as long as the implementations follow the standardized interfaces. For example, to interoperate with the legacy GSM system, GERAN (GSM/EDGE Radio Access Network) is proposed and specified. The hierarchical design of UMTS leaves room for possible upgrades in different parts of the system independently. HIPERLAN/2, developed by ETSI, is a good candidate for high-speed radio access network. In addition to the high data rate (up to 54 Mbits/s at physical layer), the extensible layered structure makes it easy to provide access to a variety of networks including 3G mobile core networks, ATM networks and IP-based networks.

The rest of the paper is organized as follows: in Section 2 we define our internetworking architecture. The integrated system is analyzed and modeled in Sections 3 and 4, respectively. In Section 5, scheduling and admission control schemes are discussed. One important feature in our system is that different RANs may have *overlapped coverage*. This imposes difficulties on efficient resources management since handoff occurs more frequently in our system than in ordinary mobile wireless networks. This problem can be approached by considering scheduling and admission control simultaneously. In Section 6 we show and explain our simulation results. Section 7 presents the conclusion.

## 2. Architecture

There are two possible implementations of HIPERLAN/2 as an access network to UMTS core network, as shown in Figure 1. One is to view the AP as a Node B connected to RNC via Iub interface. The other is to form another RAN (HIRAN, HIPERLAN/2 Radio Access Network) of its own, where the AP plays the role as an RNC. These two options have pros and cons. The advantage of the first method is simplicity. Only one new SSCS for Iub interface needs to be defined. However, this configuration may increase the complexity in radio resources management since the definitions of radio resources may not be the same for both standards, and both the AP and RNC provide functionalities of resources management. The second configuration can separate two networks further in terms of radio resources management. It makes sense to form different RANs for different radio access technologies. Therefore we choose the second method as our candidate.

In ETSI's technical report interworking HIPERLAN/2 and 3G cellular systems [5], two levels of interworking are discussed: *loose interworking* and *tight interworking*. Loose interworking is defined as the utilization of HIPERLAN/2 as a packet based access network complementary to current 3G networks, utilizing the 3G subscriber databases but without any user plane Iu type interface. Tight coupling is a direct integration of the HIPERLAN/2 radio access network into a 3G network, which is closer to our configuration. The major difference lies in the way IWU (InterWorking Unit) should be placed. We think that the IWU should be implemented directly in the convergence layer to avoid redefining the interfaces such as Iubhl2. In [6], the authors pointed out that although tight coupling has the advantage that the mechanisms for mobility, QoS and security of the UMTS core network can be directly reused, interface redefinition may impose implementation complexity. Therefore, the BRAN project decided that this solution would be very complex to analyze, define and standardize and is now considered not to be a priority for HIPERLAN/2–3G interworking. However, the capability of providing QoS guarantee in the IP-based Internet is far from satisfactory. To provide guaranteed services for multimedia traffic, it is better to incorporate HIPERLAN/2 into the UMTS system with well-defined framework of mobility, QoS and security support.

## 3. System Analysis

To manage connections between HIPERLAN/2 and UMTS networks, we identify five major procedures involved: *Association, Paging, Location Update, Authentication*[1] *and Transaction*. In Figure 2, the arrows describe the relationship of dependence. For example, to perform paging procedure the paging MT should be associated to the AP first. An MT is not allowed to use the radio resources before it is associated to the AP. After association, the MT is eligible to establish a transaction. Transaction used here refers to the connection established in the UMTS network. To set up a transaction, the MT first performs paging procedure. If the paged target (either a MT or UE) responds, the MT performs authentication procedure with the AuC (Authentication Center) in the UMTS network side. Then, transaction setup is able to be performed. Location update procedure is performed at times defined by the system configuration.

### 3.1. Association

The basic requirement of any connection for an MT is to be associated to the AP. Association to the AP means that the MT is able to decode and monitor the channel. The association procedure consists of the following sub-procedures: *RBCH association, MAC ID assignment, link capacity negotiation, encryption startup, authentication and info transfer* [7]. MAC ID is the identity to distinguish mobile terminals in the HIPERLAN/2 network. Similarly, IMSI (International Mobile Subscriber Identity) is the identity used in the UMTS network. Therefore, there should be a mapping mechanism between MAC ID and IMSI. The AP should maintain the mapping table. Each time an associated MT connects to the UMTS network, the mapping of the MAC ID and IMSI is established. When IMSI detached, the corresponding mapping entry is deleted.

[1] The authentication state here refers to the one in the UMTS network.

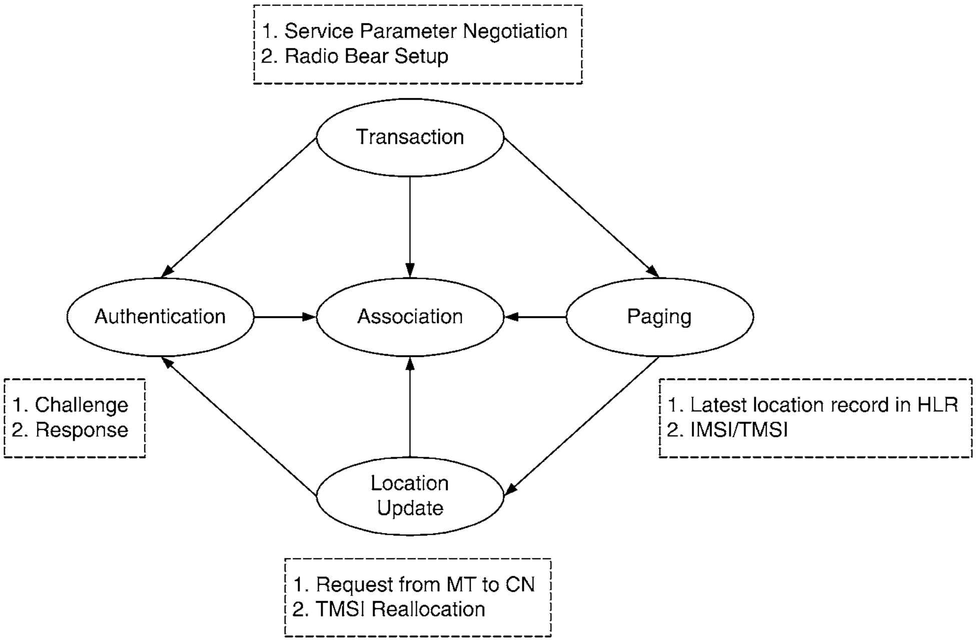


*Figure 2.* Five main procedures of connection management.

### 3.2. Paging

The MT may perform the authentication procedure because of: *paging response, location update* and *connection setup request*. Paging procedure, which is essential to call establishment, is performed with some tricks in our system. Since there is no such 'paging channel' in HIPERLAN/2 system, we can use the *Resource Grant (RG)* IE in the FCCH, which is monitored by every associated mobile terminal, to notify the MT that it is paged. Figure 11 illustrates the paging procedure.

After being paged, the addressed MT should respond in a certain period of time. Paging response can be done either by polling or by contention. The time limit for paged MT to acquire a channel is determined by the AP. If the paged MT is unable to respond by the time limit , the call establishment is terminated and the state remains at association.

### 3.3. Authentication

Before the associated MT can establish a call, some requirements have to be met. It must have radio connection with the AP, i.e. DLC User Connection (DUC). Each DUC is assigned a unique DLCC ID to identify the connection. After the authentication, the MT is eligible to set up a transaction. The procedure of DUC setup is omitted here, and the authentication procedure is illustrated in Figure 12.

### 3.4. Location Update

Location update procedure (Figure 13) is invoked in two situations:

– when the UE has detected that its location area is changed and it is necessary to inform the CN about the new location.

– on a periodic base.

Moving among different APs controlled by the same SHLN (Supporting HIPERLAN/2 Node, the enhanced SGSN node capable of handling HIPERLAN/2 traffic.) does not trigger location update since the location area identifier (LAI) remains the same. However, moving to a new area where the LAI broadcast in the channel is different from the LAI stored in the USIM invokes location update.

### 3.5. Transaction Setup

Transaction setup (Figure 14) involves call control setup and QoS negotiation. Each call has been assigned a unique transaction identifier (TI). When core network receives connection setup request, it checks whether the MT and current subscription have rights to perform the requested operation. If qualified, the core network issues RAB assignment request with given QoS parameters. The AP must map the requested QoS parameters of UMTS network to that of HIPERLAN/2 network. After the DUC is set up, call proceeds.

### 3.6. Transaction Clearing

If the transmission had a user plane active, the user plane should be disconnected first. Either the MT or the CN can invoke the disconnection procedure. After the user plane has been disconnected, the CN issues RAB release to release the DUC. Then the state goes back to association.

## 4. System Model

To evaluate our system, we consider our system in two parts: *control plane* and *user plane*. Control plane consists of procedures for connection establishment and management. User plane is responsible for data transmission. User plane with transport layer can provide reliable transmission, while user plane without transport layer cannot.

### 4.1. Control Plane

The control plane can be specified by the state transition model in Figure 3. As indicated in Figure 2, association with the HIPERLAN/2 access point is the basic requirement of connecting to the UMTS network. The triggers of state transition from association to authentication include *paging, location update* and *connection setup request*. Paging request comes from the UMTS network when an MT-terminated call is trying to set up a transaction, whereas location update and connection setup request are originated from the HIPERLAN/2 network. After authentication is done successfully, two states regarding transaction are possible: *circuit-switched transaction* and *packet-switched transaction*. After completion of the transaction (circuit-switched and packet-switched), the state transits to association. Disassociation request is issued according to the criteria defined in HIPERLAN/2 standard.

#### 4.1.1. *From Disassociation to Association*

Based on the state transition model, we can derive a specific model for performance evaluation. When an MT wants to associate with the AP, it has to contend for opportunity of transmission. The contention process functions in the *Slotted ALOHA* with *Binary Exponential Backoff (BEB)*. Once the contention is successful (which is determined by the AP), the MT

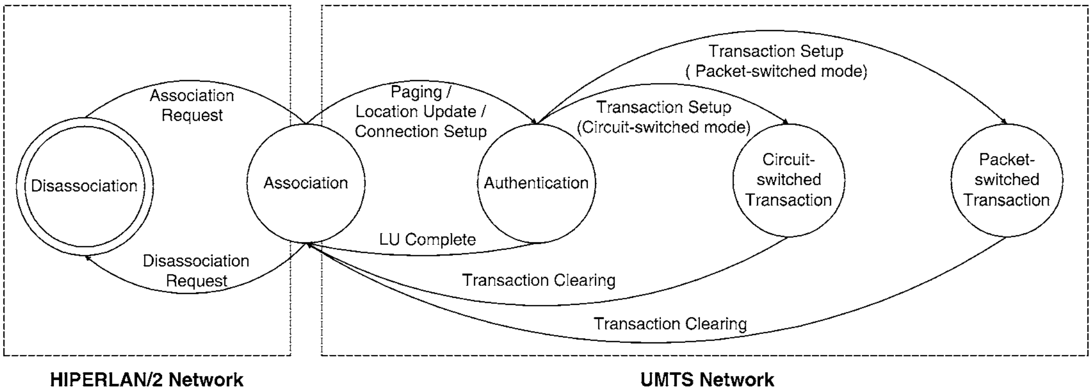


*Figure 3.* State transition model of the IWF.

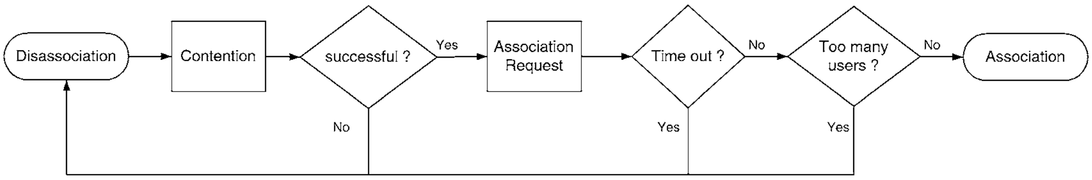


*Figure 4.* Flow chart from Disassociation state to Association state.

sends *Association Request* to the AP. If no response (*Association Accept*) is received before certain time limit,[2] the association request is considered *time out* and the MT must start over from contention. In addition to time out, the association request could be rejected because of

– the number of mobile terminals in association reaches maximum[3]
– call admission control.

We will discuss call admission control later. For the time being, we only consider the first reason for association request rejection.

Since different traffic types are supported in the HIPERLAN/2, different traffic modeling techniques should be applied in the mathematical model. However, for the purpose of description, we assume the aggregate traffic arrival follows Poisson with mean arrival rate $\lambda$. We model the contention process as a system of *infinite servers* since every MT is eligible to contend for transmission. However, the time needed for successful contention, i.e. service time in the infinite servers, is hard to model. For simplicity and to avoid the phenomenon of *first contend, first succeed*, we choose exponential distribution for the successful contention time. The maximum buffer size $N_1$ is determined by the implicit requirement that *maximum queueing delay should not be larger than time-out limit of association request*. Figure 5 shows the equivalent queueing system for the flow chart in Figure 4.

### 4.1.2. *From Association to Authentication*

After association, the MT starts authentication process. Without loss of generality, we assume that authentication process takes *constant time* for every MT. Then we can have a similar queueing model in Figure 6. The sum of buffer sizes $N_2$ and $N_3$ represents the *maximum*

[2] See [7, Annex C] for more information.
[3] See [8, Section 5.5] for more information.

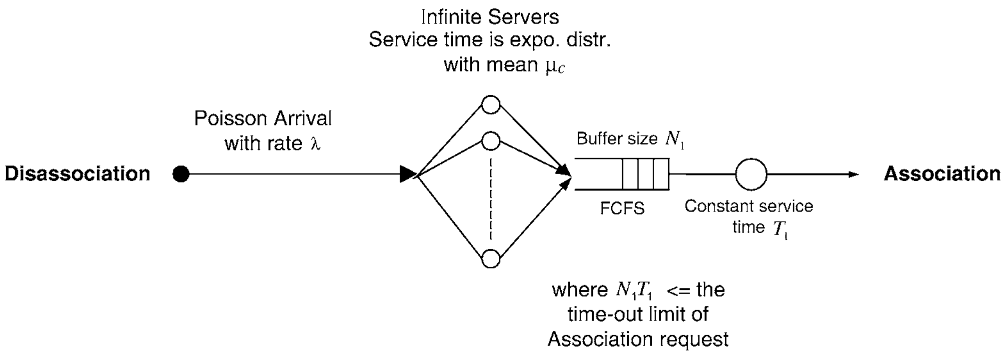


*Figure 5.* Equivalent queueing system from Disassociation to Association.

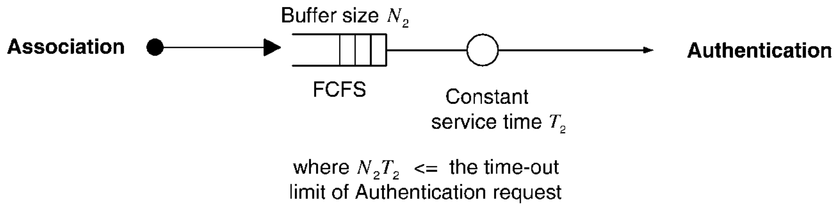


*Figure 6.* Equivalent queueing system from Association to Authentication.

*number of mobile terminals allowed in association state*. An MT with successful contention is blocked if the queue length reaches maximum.

4.1.3. *From Association to Unicast DUC Setup*

In addition to the UMTS traffic, *LAN traffic* should be taken into consideration. What we mean by LAN traffic is the traffic between mobile terminals, where the AP simply relays the traffic from one to another. Transmitters must first associate to the AP, set up the DUC, and then transmit the message.

4.1.4. *From Authentication to UMTS Transaction Setup*

Authentication is followed by transaction setup. There are two kinds of transaction in our system: circuit-switched and packet-switched. Assume at most $M_c$ simultaneous circuit-switched transactions are allowed, and at most $M_p$ packet-switched transaction are allowed. That is, the maximum queue length of packet-switched transaction setup is $M_p$. Both of the transaction setup are assumed to take constant time. An associated MT may be *blocked* due to (1) call admission control and (2) current number of transactions reaches maximum.

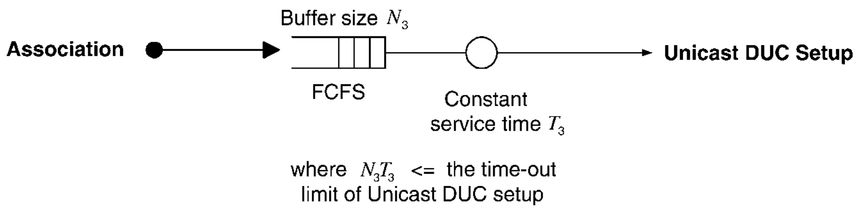


*Figure 7.* Equivalent queueing system from Association to Unicast DUC setup.

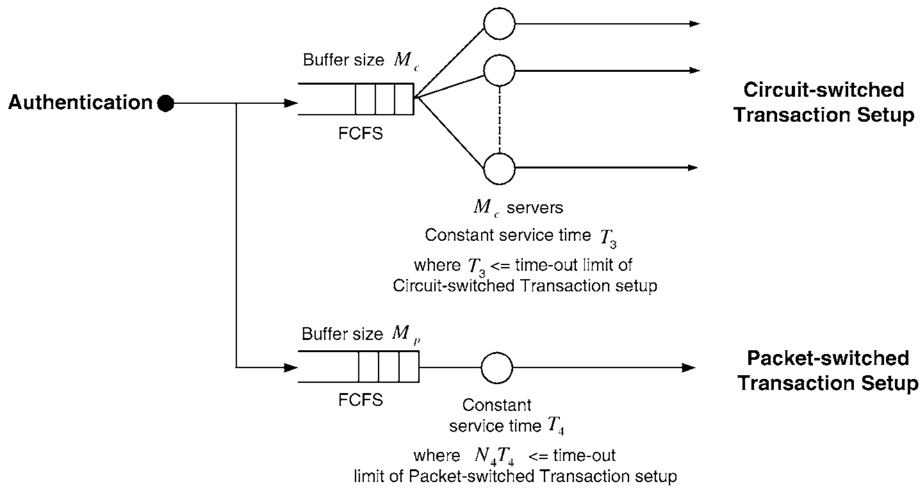


*Figure 8.* Equivalent queueing system from Authentication to Transaction setup.

## 4.2. User Plane

In addition to the traffic *types*, traffic transmission *direction* is important, too. We consider four typical types of traffic: *voice, video, data* and *LAN*, and two directions: *uplink* and *downlink*.

### 4.2.1. *Voice Traffic*

We use the *two-state Markov process* to model the voice source and use the parameters in [9] for our voice model. Assume that the voice packetization period is $T = 16$ ms, the silence periods are exponentially distributed with mean $\beta^{-1} = 650$ ms, and the mean talk spurt periods are exponentially distributed with mean $\alpha^{-1} = 352$ ms.

Since voice traffic is generally bi-directional, for a uplink voice source there is a corresponding downlink voice source. These two voice sources should be correlated. For example, the possibility of one side remaining silent is different under the condition that the other side is talking or listening. For simplicity, we assume that these two sources are *independent*. Thus, for each flow of voice traffic, there are two statistically identical sources in the uplink and downlink, respectively.

### 4.2.2. *Video Traffic*

We generate video sources based on a two-state Markov Modulated Poisson Process (MMPP). The two-state MMPP has four parameters: $\lambda_1$, $\lambda_2$, $\mu_1$ and $\mu_2$, where $\lambda_1$ and $\lambda_2$ are the conditional traffic arrival rates, and $\mu_1$ and $\mu_2$ are conditional transition rates given that the Markov chain is in state 1 and 2, respectively. We set our parameters as follows: $\lambda_1 = \lambda_2 = 10.0$, $\mu_1 = \mu_2 = 1.0$.

We assume that video sources only occur in the downlink direction, which is mostly the case of viewing streaming video from the web.

### 4.2.3. *Data and LAN Traffic*

It is known that the sources generating non-real-time traffic such as WWW and FTP traffic could be effectively modeled as ON/OFF processes with Pareto distribution [10, 11]. However, for simplicity, we model our sources of data and LAN traffic by Poisson arrival processes and exponentially distributed packet lengthes with mean packet size 1500 bytes.

Data traffic occurs in both uplink and downlink directions. LAN traffic is generated in the uplink direction. As described earlier, the AP simply relays the LAN traffic to the target MT. Thus, after successfully transferred in the uplink, the AP schedules the LAN traffic in the downlink.

### 4.3. Complete System Model

From the above discussion, we can build up a complete model for performance evaluation. Figure 15 shows the complete model of control plane, and Figures 16 and 17 show the user plane of different types of traffic.

## 5. Performance Evaluation

### 5.1. Scheduling

In order to deliver guaranteed QoS while at the same time make the most of statistical multiplexing, we propose an effective scheduling scheme called *Urgency First Round-Robin with Priority (UFRRP)*. The proposed scheme aims to improve the well-known EDF (Earliest Deadline First) scheme, which is not quite feasible in the sense of timing synchronization. Instead of *continuous-time deadline*, we use a *discrete-time delay bound* defined in terms of *MAC frames* to avoid the timing synchronization problem. When the MT sends RR message to the AP, it includes the delay bound information before which the requested LCHs must be scheduled for transmission. Smallest delay bound is defined to be 0, which means this RR must be scheduled at next MAC frame. We call this delay bound information *urgency*. To achieve high throughput in ATM network, each cell should be delayed as close to its deadline as possible [12]. Additional non-delay-sensitive cells can be inserted in these extra vacancies resulted from intentionally delayed cells. In this way, higher throughput can be achieved without violating guaranteed delay bound. In HIPERLAN/2 network, however, cells (LCHs) are not individually scheduled. The basic element for scheduling is the PDU train. Nevertheless, following the same methodology, we can delay transmission of requested LCHs to the MAC frame they can be scheduled without violating the QoS guarantee.

#### 5.1.1. *Proposed Scheduling Algorithm*

To transmit data, the MT first issues resource request (RR) to the AP through either polling or contention. Each RR is queued in the corresponding priority queue. In each queue, the RRs are sorted in the order of their urgency, where RR with smaller urgency is placed in front of RR with larger urgency. For each MAC frame, RRs with 0 urgency in each queue are scheduled first. If there are available LCHs (or *residual bandwidth*) left, we check each queue in a round-robin manner to see if there are any RR waiting. If any, we schedule it for transmission until we run out of all available LCHs. To conform to our system model in Section 4, we classify RRs into four priorities: *voice, video, data* and *LAN*. The details can be further explored by Figure 9. The number of available LCHs is difficult to determine. It depends on modulation scheme used. In the HIPERLAN/2 standard, each PDU train is allowed to use different modulation scheme, which is negotiated between the MT and the AP. For simplicity, we assume the modulation scheme used in our system remains the same for all the time. Besides, the number of RCHs in each MAC frame is kept constant since we simply model the contention process as time elapsing (exponentially distributed with mean $\mu_c$), although there could be improvement

For each MAC frame,

1. Determine **the number of schedulable LCHs** in the current MAC frame.
2. Scan through each queue in the order of desceding priority, find those RRs with **0 urgency** and schedule them for transmission.
3. For those RRs with urgency >= 1, **descrease their urgency by 1**.
4. After all RRs with 0 urgency are scheduled, check if there is any LCH available
5. If any, schedule RRs in each queue in a **round-robin** manner until all LCHs are run out.

*Figure 9.* The proposed UFRRP algorithm.

*Table 1.* The number of SCHs for different modulation scheme.

| Modulation scheme | Nominal bit rate (Mbit/s) | Bytes per MAC frame | No. of SCHs per MAC frame |
|---|---|---|---|
| BPSK, 1/2 | 6 | 1500 | 167 |
| BPSK, 3/4 | 9 | 2250 | 250 |
| QPSK, 1/2 | 12 | 3000 | 333 |
| QPSK, 3/4 | 18 | 4500 | 500 |
| 16QAM, 9/16 | 27 | 6750 | 750 |
| 16QAM, 3/4 | 36 | 9000 | 1000 |
| 64QAM, 3/4 | 54 | 13500 | 1500 |

in adaptively assigning RCHs [13]. For example, for BPSK modulation, code rate $1/2$, the nominal rate on top of physical layer is 6 Mbit/s. Thus, there are 6 Mbit/s $*$ 2 ms $=$ 1500 bytes in one MAC frame. The length of BCH is constant and equal to 15 bytes. ACH contains 9 bytes. Assume 6 RCHs is assigned in one MAC frame, where one RCH contains 9 bytes. While BCH, ACH, and even RCHs with additional assumption contain constant number of bytes, the length of FCH depends on scheduling result. If there are $n$ RGs scheduled in the current MAC frame, the length of FCH will be $\lceil \frac{n}{3} \rceil * 27$ bytes. Therefore, at least 27 bytes is necessary to carry the FCH. Assume for now only one RG exists. Then the number of available LCHs equals to $25\frac{5}{6}$. The fractional part comes from the fact that another smaller unit exists: the SCH, which carries control information other than user data. The length of SCH is one sixth of LCH. In the above example, we can count available bandwidth in terms of SCH, which will be 155 SCHs. In Table 1 we list the number of SCHs for different modulation scheme.

## 5.2. Admission Control

There are two main categories of call admission control algorithms: *parameter-based* and *measurement-based*. In parameter-based CAC, users declare traffic descriptors used for admission control at connection setup time. Admission control algorithms use the characterizations of sources to calculate the *worst-case* behavior of all the existing calls in addition to the incoming one. If the impact of addition of the new call is tolerable in the sense of QoS parameters such as cell loss rate (CLR) and cell transfer delay, the new call request is admitted. In this way, real-time services with *hard* bounds on QoS parameters can be guaranteed. However, there are some limitations in parameter-based CAC. First, the use of *a priori* characterization of sources is not quite effective since, for example, the exact average rate is an *a posteriori* parameter. Second, resources utilization under the guaranteed service model is usually acceptable when traffic flows are smooth; when the flows are bursty, guaranteed service inevitably results in low utilization [14]. In addition, parameter-based admission control methods require construction of appropriate traffic descriptors for every new application. This limits the feasibility of the admission control algorithms in an environment of heterogeneous traffic.

The measurement-based CAC uses measurements to characterize those on-going calls that have been in the system for a reasonable duration. It is the incoming calls that are characterized by *a priori* traffic descriptors in order to be evaluated by the admission control algorithm. Hence, resources utilization does not suffer significantly if the traffic descriptions are not tight. However, since it relies on measurements, and source behavior is not static in general, the measurement-based approach to admission control can never provide the completely reliable delay bounds needed for guaranteed services. Furthermore, when there are only a few sources present in the system, the unpredictability of individual source's behavior dictates that the admission control may be very conservative to prevent overload. Therefore a measurement-based admission control algorithm can deliver significant gain in utilization only when there is a high degree of statistical multiplexing [15].

Among all these dazzling algorithms, it is essential to apply the appropriate method for a particular system. For our system, it is not practical to use *a priori* parameters to characterize the traffic since not only the MT-originated traffic presents, there are also network-originated traffic which may not fit into the pre-assumed behavior after multiple nodes. Besides, CAC is not independent of scheduling algorithm. They represent different levels of resource allocation.

In [16], the authors proposed a CAC method based on the measurement of *instantaneous virtual path utilization* in the ATM network, which was defined as the total cell rate of the active virtual channels normalized by the virtual path capacity. A low-pass filter was used to determine the instantaneous virtual path utilization from crude measurements. The residual bandwidth was derived from the maximum instantaneous utilization observed during a monitoring period. Based on [16] and our UFRRP scheduling algorithm, we develop a suitable admission control scheme for our system.

### 5.2.1. *Proposed Admission Control Algorithm*

During the connection setup procedure, the source should address its *peak rate*, which is defined to be the maximum number of requested LCHs in each MAC frame. In the beginning of each monitoring period, the *instantaneous utilization* $\lambda(t)$ is defined to be

$$\lambda(t) = \sum_i \frac{R_i}{SLOTS},$$

where source $i$ is active at time $t$, and *SLOTS* is the number of available LCHs in the current MAC frame, which must be calculated in the UFRRP scheduling algorithm. We update $\lambda(t)$ every MAC frame during the monitoring period with a recursive LPF whose smoothing factor is $\alpha$

$$\lambda(t) = \alpha n(t) + (1 - \alpha)\lambda(t - \tau),$$

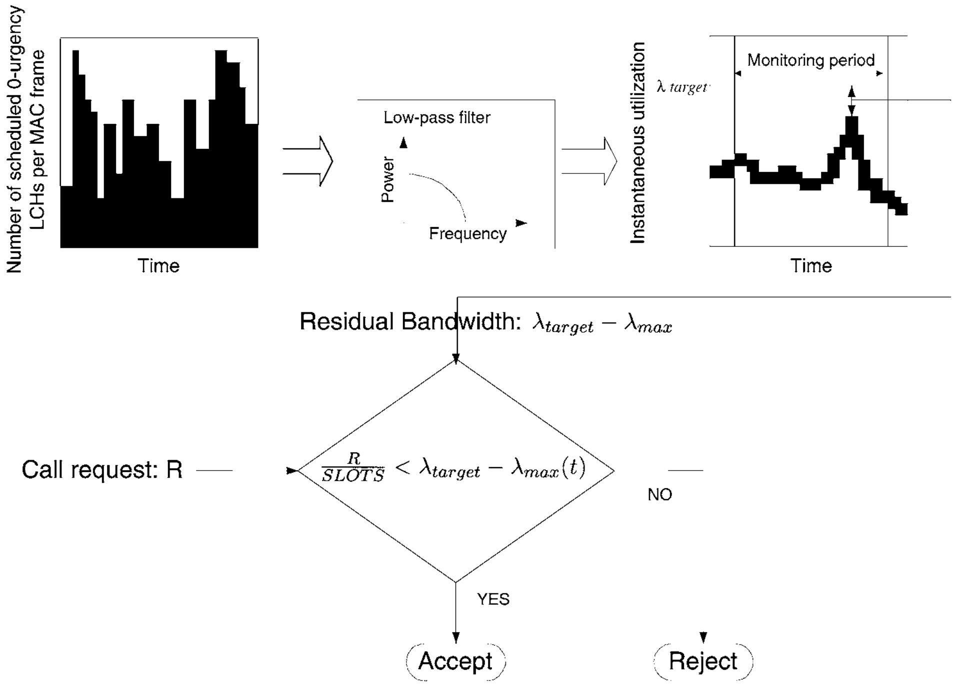


*Figure 10.* Illustration of admission control scheme.

where $n(t)$ is the number of LCHs with zero-urgency scheduled for the current MAC frame and $\tau$ equals to the length of one MAC frame. The maximum instantaneous utilization observed during the monitoring period $\lambda_{max}(t)$ is defined as

$$\lambda_{max}(t) = \max_{t' \in (t-T_m, t]} \lambda(t'),$$

where $T_m$ denotes the length of a monitoring period. Thus, the admission criteria can be written as

$$\frac{R}{SLOTS} < \lambda_{target} - \lambda_{max}(t)$$

for a new call request with peak rate $R$. We define $\lambda_{target} - \lambda_{max}(t)$ to be the residual bandwidth. If the call request is admitted, the instantaneous utilization is increased by an amount equal to the accepted call's peak rate

$$\lambda_{new}(t') = \lambda(t') + \frac{R}{SLOTS}.$$

The above algorithm is illustrated in Figure 10 [16].

## 6. Simulation Results

The simulation is constructed by YACSIM [17], which is a process-oriented discrete-event simulator implemented as an extension of the C programming language. Following the complete system model in Figure 15, we apply proposed scheduling and admission control schemes to evaluate our system performance.

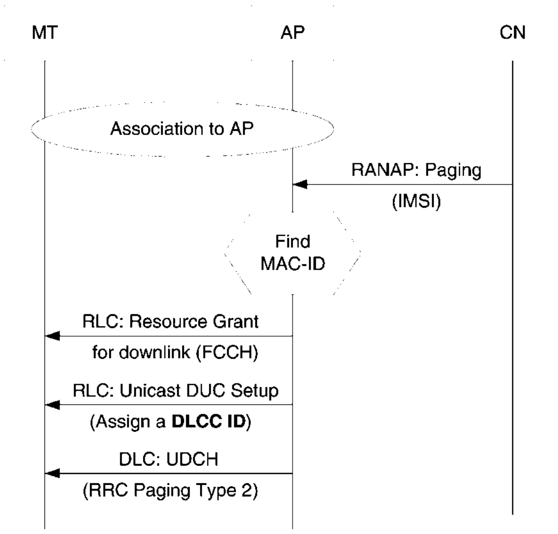


*Figure 11.* Paging procedure.

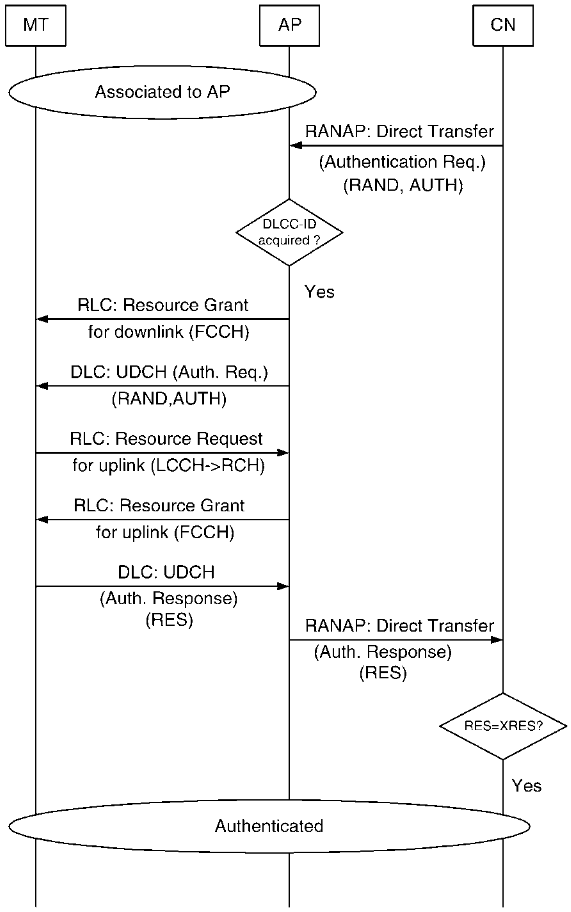


*Figure 12.* Authentication procedure.

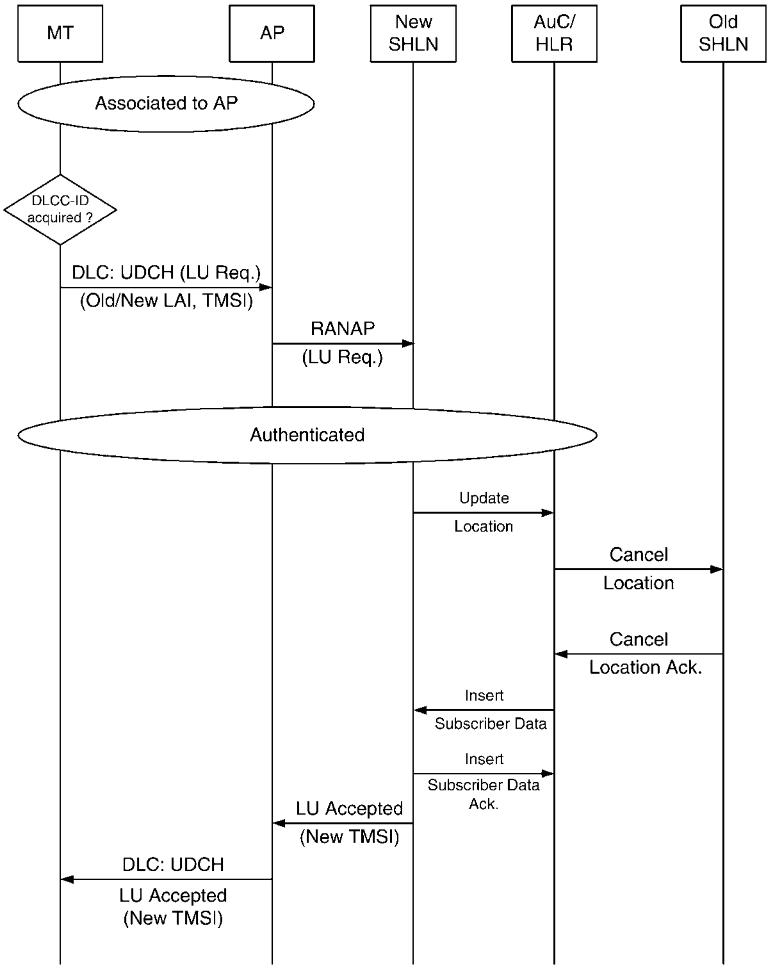


*Figure 13.* Location update procedure.

### 6.1. Results of Proposed Scheduling Scheme

To evaluate our scheduling algorithm, we compare the performance of EDF and FCFS with our UFRRP. We use delay suffered by each burst of message to be the performance metric. The delay bound for voice burst is set to 10 ms, video to 20 ms. Data burst, although not delay-sensitive, is assigned a delay bound 100 ms since UMTS traffic is still delay bounded. Similarly, we assign a delay bound of 1 s to LAN packet burst in order to fit into the UFRRP and EDF algorithms. The simulation assumes BPSK, coding rate 1/2 modulation scheme.

#### 6.1.1. *Increase Voice Traffic Load*

From Figures 18(a–c), we can see that for real-time traffic like voice and video, the performance of UFRRP is comparable to that of EDF, sometimes even better. But for non-delay-sensitive traffic like data and LAN, in Figures 18(d–f), the average delay is close to, but not exceeds its delay bound for UFRRP. This is due to the mechanism we apply in the algorithm that PDUs are delayed as close as possible to its deadline to achieve higher throughput.

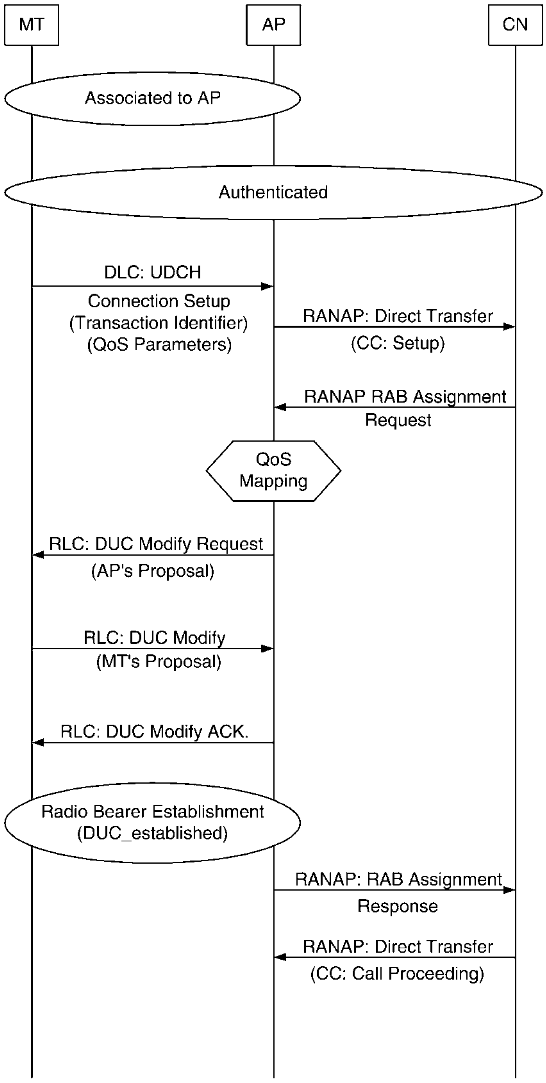


*Figure 14.* Transaction setup procedure.

### 6.1.2. *Increase LAN Traffic Load*

In Figure 19(a), the voice user delay in FCFS rises rapidly as the number of LAN MTs increases from 10 to 50, while in UFRRP and EDF the delays are kept relatively constant. With a closer look at Figure 19(b), we can find that sometime UFRRP achieves lower delay than EDF. This is because in UFRRP, residual bandwidth is redistributed according to priority, which may result in better performance of higher priority traffic than traditional EDF. The same phenomenon can be observed in Figures 20 and 21. Although the average delay of data traffic in UFRRP is higher that that in EDF and FCFS, as seen in Figures 22(a, b), but is kept under the delay bound and quite constant comparing to that in FCFS. The performance of LAN traffic (Figure 22(c)) is sacrificed in UFRRP to achieve better performance of real-time traffic.

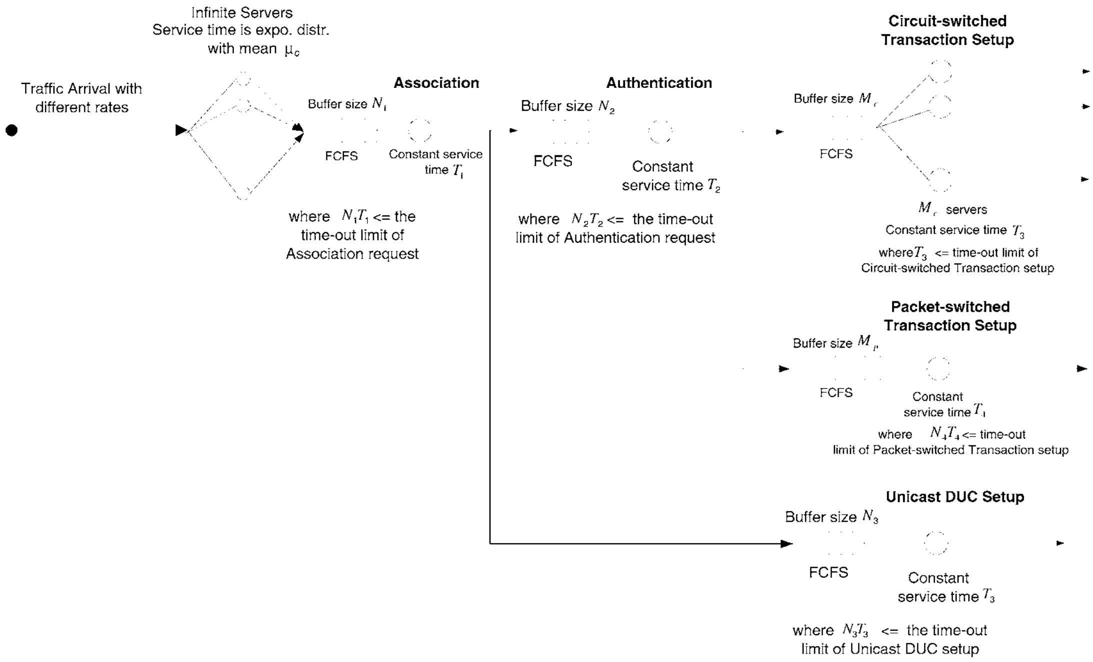


*Figure 15.* Complete model of control plane.

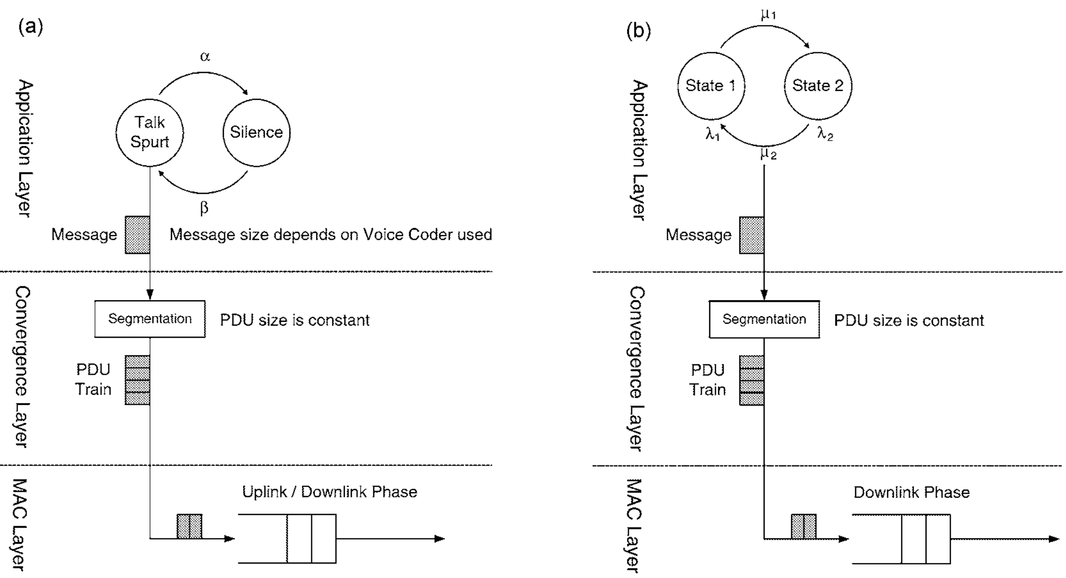


*Figure 16.* Complete model of (a) single voice source; (b) single video source.

## 6.2. Results of Proposed Admission Control Scheme

The admission control is invoked before association. Only calls that pass the criteria could be admitted and allowed to associate with the access point. The parameters of video traffic are changed to increase the load: $\lambda_1 = \lambda_2 = 10.0$, $\mu_1 = 100$, $\mu_2 = 50$. Each call will proceed for a random time which is exponentially distributed with mean 60 s. After terminated, a new call request will be generated after a random time exponentially distributed with mean 1 s.

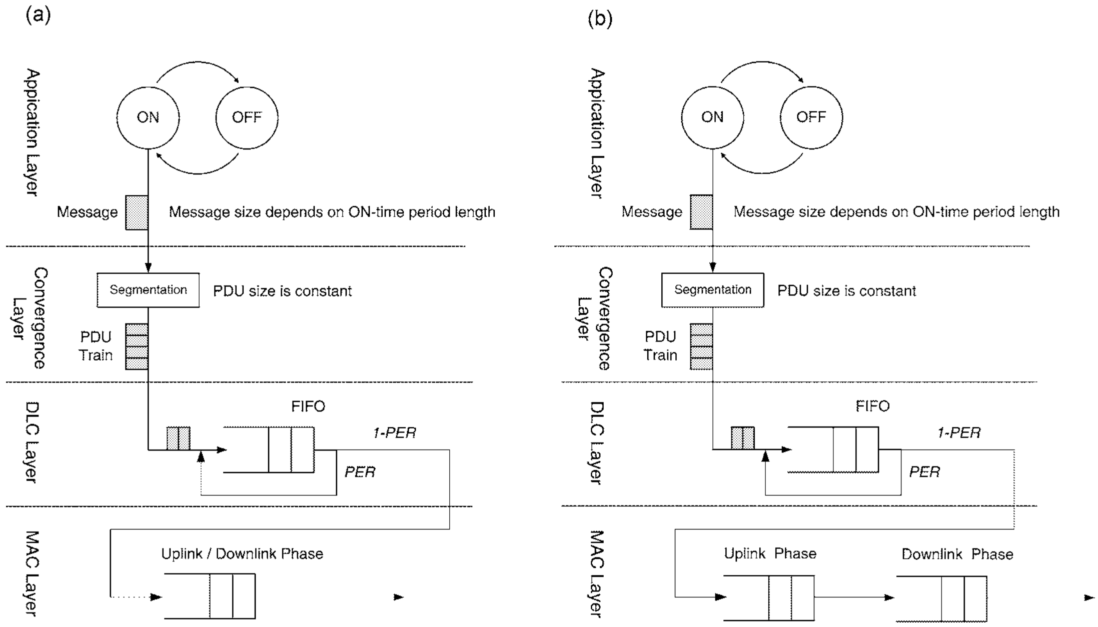


*Figure 17.* Complete model of (a) single data source; (b) single LAN source.

The number of voice mobile terminals are kept under 10, while for video, data, LAN mobile terminals, the maximum numbers are 10, 20, 50, respectively. Smoothing coefficient $\alpha$, which controls the cutoff frequency of the LPF, is set to be 0.001 in the simulation. The simulation assumes BPSK, coding rate 1/2 modulation scheme.

In Figure 23, we can see that varying the length of monitoring period has not much impact on PDU loss rate (PLR) of real-time traffic. Under the same parameters but without admission control, Voice UL PLR $\simeq$ 0.0244, Voice DL PLR $\simeq$ 0.0246, Video DL PLR $\simeq$ 0.0045, which are higher than those with admission control. In Figure 24(a), we investigate the relationship between blocking rate and monitoring period. The result shows that as the length of monitoring period increases, blocking rates tend to increase as well. This may be because the longer the monitoring period, the greater the possibility that when a bursty traffic arrives during the monitoring period, the residual bandwidth is underestimated, resulting in high blocking rate for the rest of the monitoring period. An abrupt rising at the point where monitoring period is equal to 2 s may indicate an feasible operation point under our simulation settings.

The relationship between blocking rate and $\lambda_{target}$ is shown in Figure 24(b). Lowering $\lambda_{target}$ increases the blocking rate. In other words, we can reserve bandwidth by adjusting the value of $\lambda_{target}$: the smaller the value of $\lambda_{target}$, the more bandwidth reserved for different reasons, such as emergent call requests. This bandwidth reservation mechanism is very important to internetworking systems where handoffs between two networks are frequent. Blocking a handoff request is worse than blocking a ordinary call request since we are supposed to provide a *seamless* mobile networking environment. However, the amount of reserved bandwidth depends on system configuration, which is out of the scope of our concern.

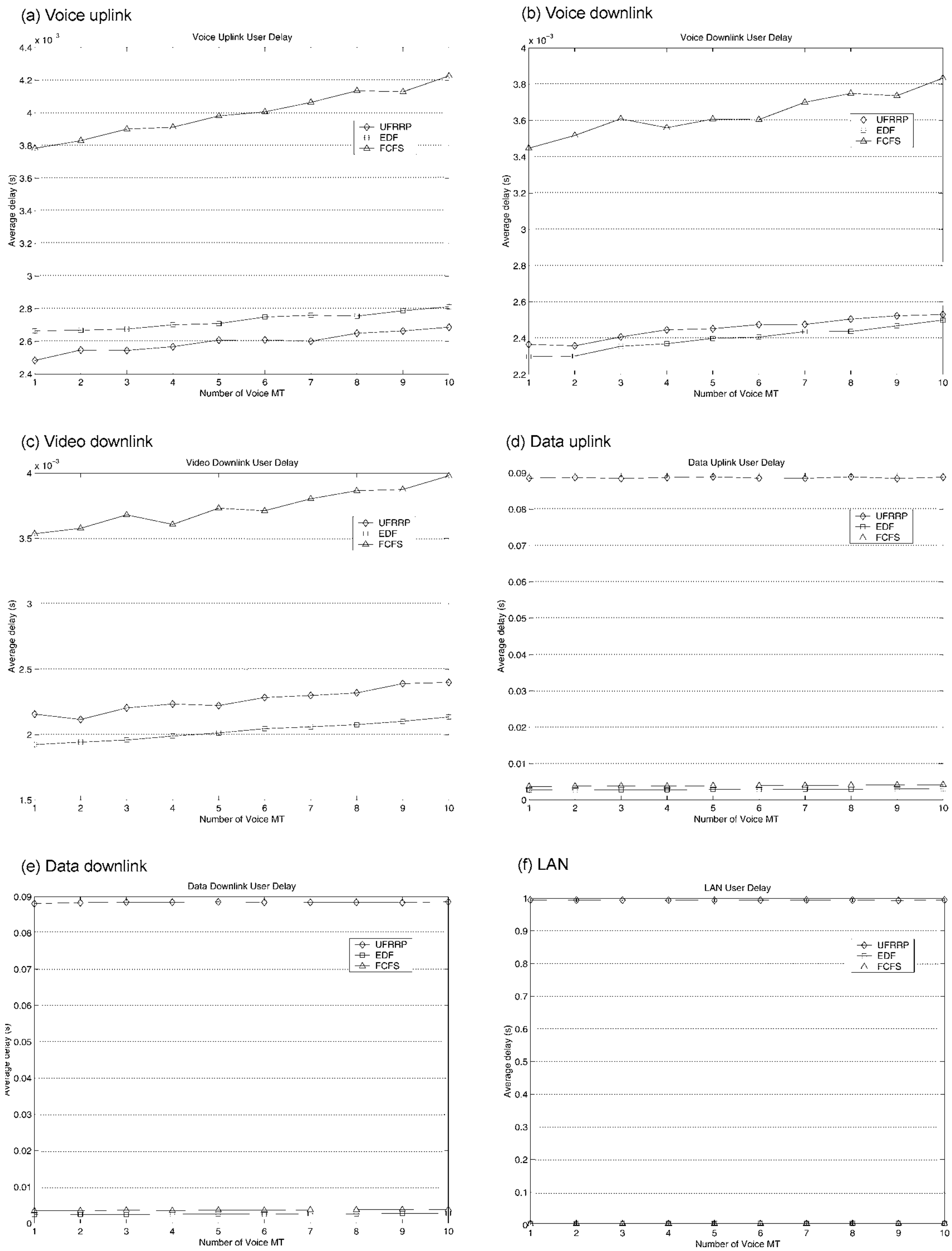


*Figure 18.* Delays of different traffic types with increasing number of voice MTs.

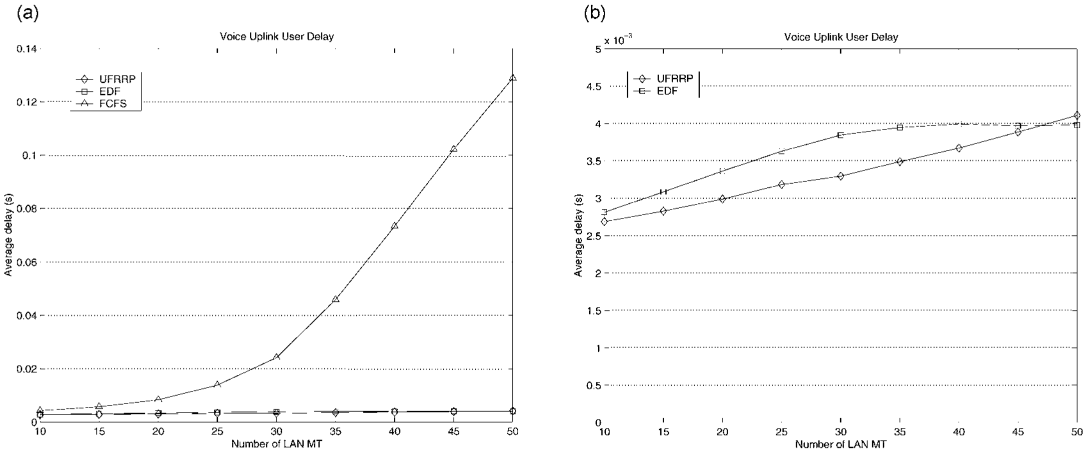


*Figure 19.* Voice uplink user delay with increasing number of LAN MTs.

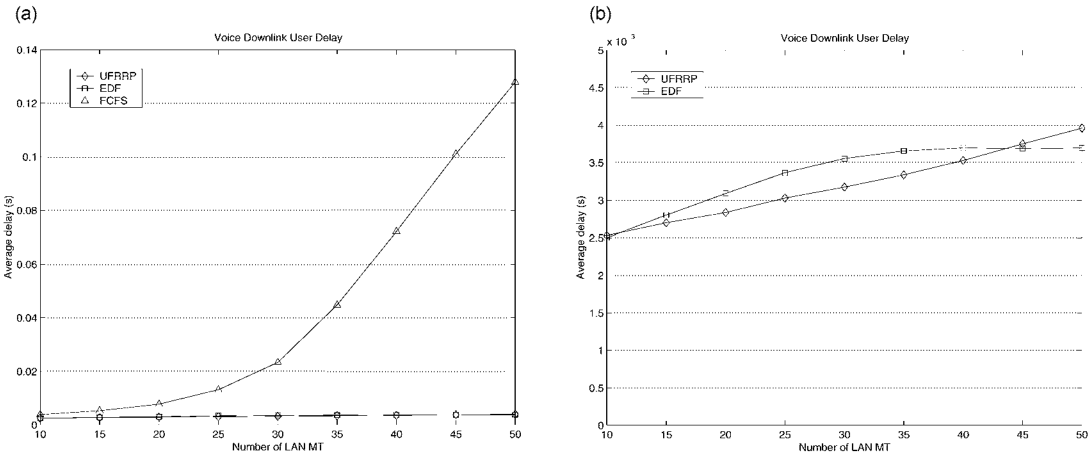


*Figure 20.* Voice downlink user delay with increasing number of LAN MTs.

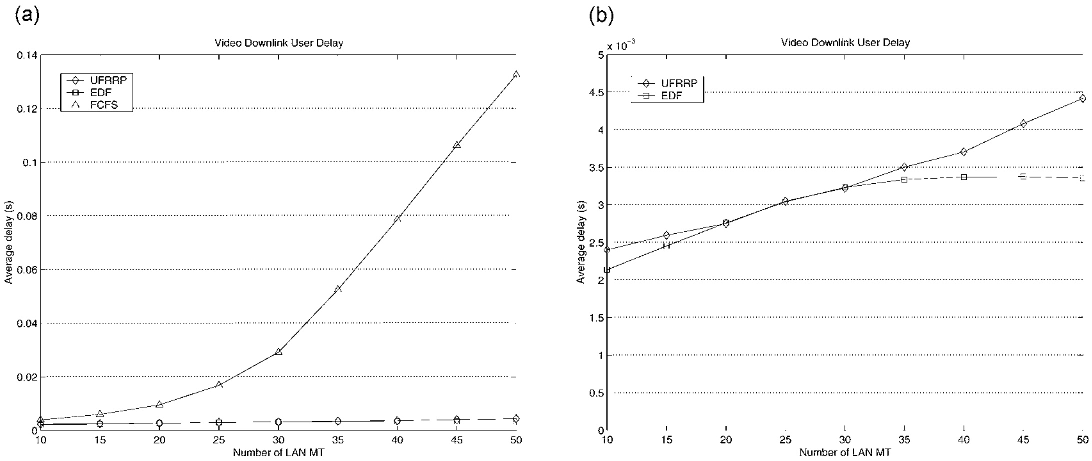


*Figure 21.* Video downlink user delay with increasing number of LAN MTs.

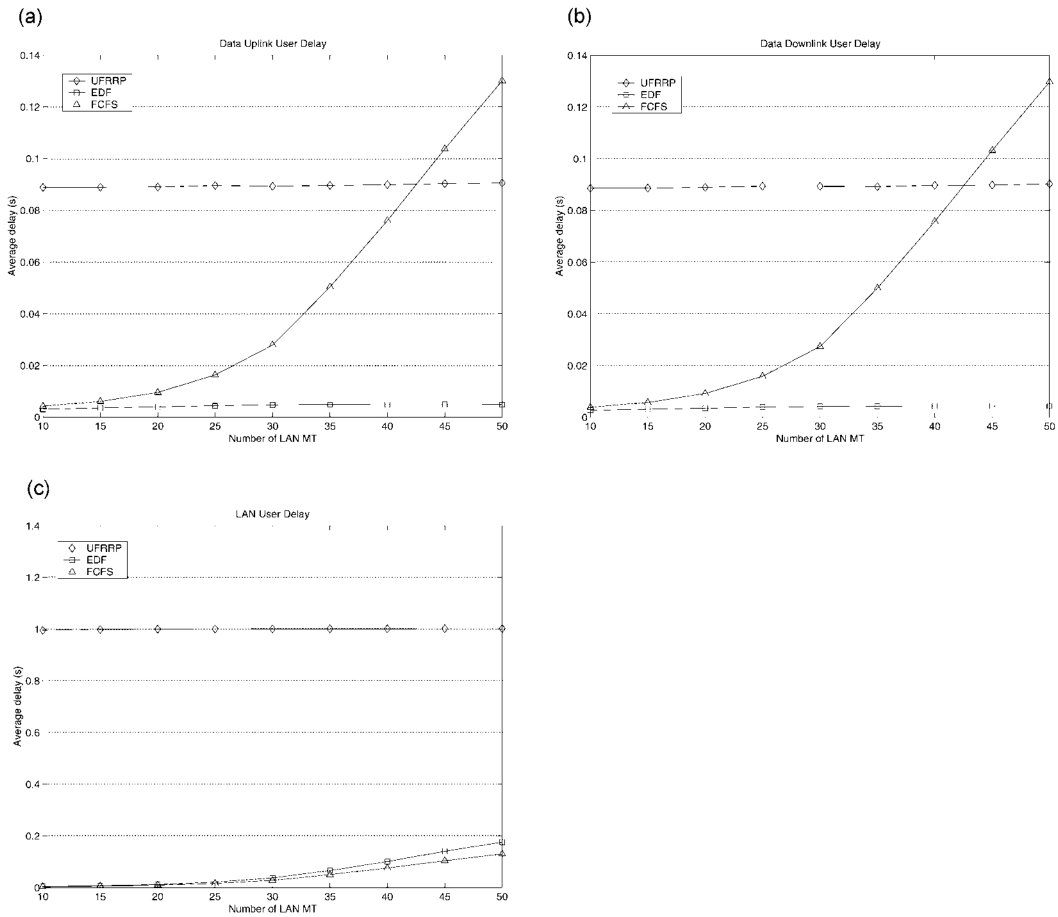


*Figure 22.* (a) Data uplink. (b) Data downlink. (c) LAN traffic delay with increasing number of LAN MTs.

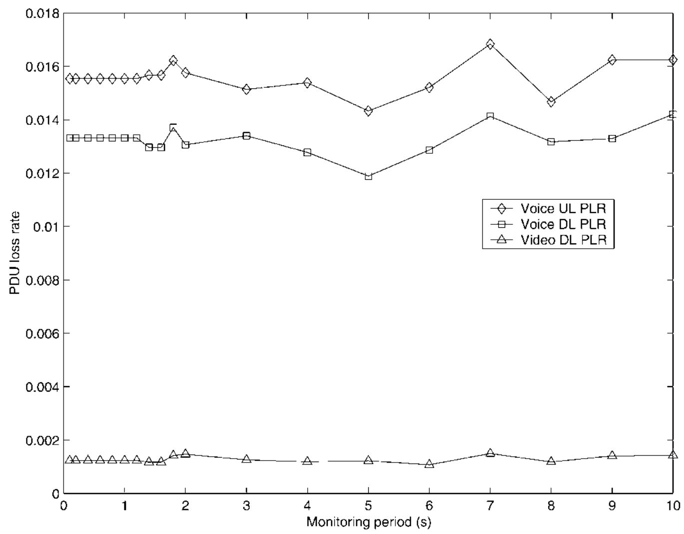


*Figure 23.* PDU loss rate v.s. Monitoring period, where $\lambda_{target} = 1.0$.

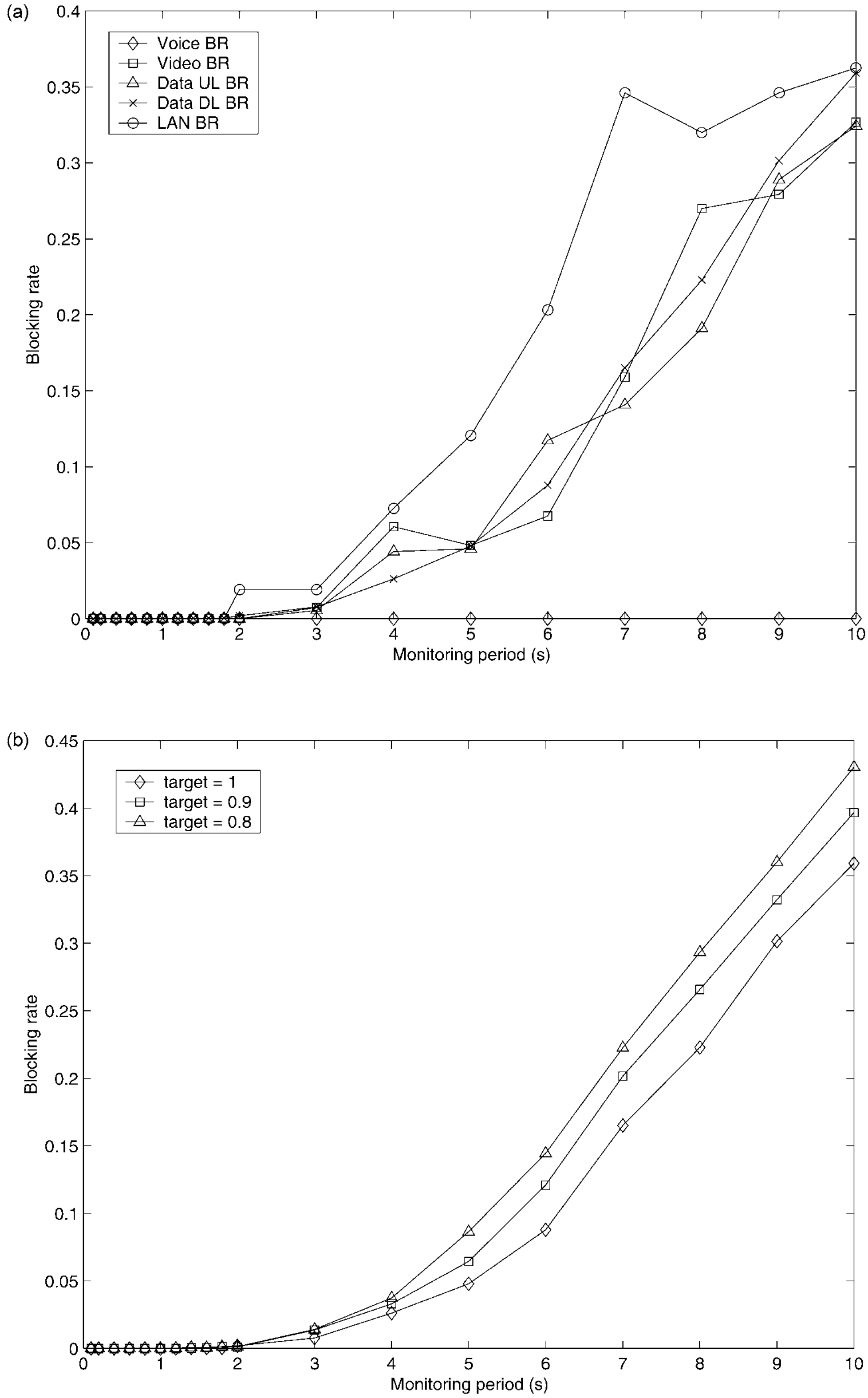


*Figure 24.* Block rate v.s. (a) Monitoring period, where $\lambda_{target} = 1.0$. (b) $\lambda_{target}$.

## 7. Conclusion

In this paper, an internetworking architecture for HIPERLAN/2 and UMTS networks has been proposed. Based on this architecture, we develop a framework for evaluating the performance of such a system by identifying main signalling procedures concerning the interconnection. Then we investigate the potential problems resulting from the interconnection of two independent, heterogeneous networks. In scheduling, we propose an algorithm designed under the constraint of the HIPERLAN/2. Numerical results shows that when using proposed UFRRP algorithm, the delay of real-time traffic can be bounded, and the performance is close to the EDF algorithm at the cost of performance of non-delay-sensitive traffic. A simple and efficient admission criteria is proposed based on the scheduling results. Therefore, our method for interconnecting HIPERLAN/2 and UMTS networks is feasible and efficient with appropriate configuration.

## References

1. E. Brewer, R. Katz, Y. Chawathe, S. Gribble, T. Hodes, G. Nguyen, M. Stemm, T. Henderson, E. Amir, H. Balakrishnan, A. Fox, V. Padmanabhan and S. Seshan, "A Network Architecture for Heterogeneous Mobile Computing", *IEEE Personal Communications*, Vol. 5, No. 5, pp. 8–24, 1998.
2. M. Stemm and R.H. Katz, "Vertical Handoffs in Wireless Overlay Networks", *Mobile Networks and Applications*, Vol. 3, No. 4, pp. 335–350, 1998.
3. K. Pahlavan, P. Krishnamurthy, A. Hatami, M. Ylianttila, J. Makela, R. Pichna and J. Vallstron, "Handoff in Hybrid Mobile Data Networks", *IEEE Personal Communications*, Vol. 7, No. 2, pp. 34–47, 2000.
4. M. Ylianttila, M. Pande, J. Makela and P. Mahonen, "Optimization Scheme for Mobile Users Performing Vertical Handoffs between IEEE 802.11 and GPRS/EDGE", *Global Telecommunications Conference, 2001. GLOBECOM '01. IEEE*, Vol. 6, pp. 3439–3443, 2001.
5. BRAN, "HIgh PErformance Radio Local Area Network (HIPERLAN) Type 2; Requirement and Architectures for Interworking between HIPERLAN/2 and 3rd Generation Cellular Systems", *ETSI Standard TR 101 957 V1.1.1*, 2001.
6. S. McCann and H. Flygare, "HiperLAN/2 Public Access Interworking with 3g Cellular Systems (Invited Paper)", in *Proceeding of European Wireless 2002*, Feb. 2002, available at http://www.ing.unipi.it/ew2002/
7. BRAN, "HIgh PErformance Radio Local Area Network (HIPERLAN) Type 2; Data Link Control (DLC) Layer; Part 2: Radio Link Control (RLC) Sublayer", *ETSI Standard TS 101 761-2 V1.2.1*, April 2001.
8. BRAN, "HIgh PErformance Radio Local Area Network (HIPERLAN) Type 2; Data Link Control (DLC) Layer; Part 1: Basic Data Transport Functions", *ETSI Standard TS 101 761-1 V1.2.1*, Nov. 2000.
9. K. Sriram and W. Whitt, "Characterizing Superposition Arrival Processes in Packet Multiplexers for Voice and Data", *IEEE Journal on Selected Area in Communications*, Vol. SAC-4, No. 6, pp. 833–846, 1986.
10. V. Paxson and S. Floyd, "Wide Area Traffic: The Failure of Poisson Modeling", *IEEE/ACM Transactions on Networking*, Vol. 3, No. 3, pp. 226–244, 1995.
11. M.E. Crovella and A. Bestavros, "Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes", *IEEE/ACM Transactions on Networking*, Vol. 5, No. 6, pp. 835–846, 1997.
12. T. Ling and N. Shroff, "Scheduling Real-Time Traffic in ATM Networks", in *Proc. IEEE INFOCOM'96*, Mar. 1996, pp. 198–205.
13. G.-H. Hwang and D.-H. Cho, "Adaptive Random Channel Allocation Scheme in HIPERLAN Type 2", *IEEE Communications Letters*, Vol. 6, pp. 40–42, 2002.
14. H. Zhang and D. Ferrari, "Improving Utilization for Deterministic Service in Multimedia Communication", in *Proceedings of the International Conference on Multimedia Computing and Systems*, 1994, pp. 295–304.
15. S. Jamin, P. Danzig, S. Shenker and L. Zhang, "A Measurement-Based Admission Control Algorithm for Integrated Services Packet Networks", *IEEE/ACM Transactions on Networking*, Vol. 5, No. 1, pp. 56–70, 1997.

16. K. Siomoto, S. Chaki and N. Yamanaka, "A Simple Bandwidth Management Strategy Based on Measurements of Instantaneous Virtual Path Utilization in ATM Networks", *IEEE/ACM Transactions on Networking*, Vol. 6, No. 5, pp. 625–634, 1998.
17. J.R. Jump, "Rice Parallel Processing Testbed", available at http://www.crpc.rice.edu/softlib/rppt.html, 1993.

**Kwang-Cheng Chen** received B.S. from the National Taiwan University in 1983, M.S. and Ph.D. from the University of Maryland, College Park, United States, in 1987 and 1989, all in electrical engineering. From 1987 to 1991, Dr. Chen worked with SSE, COMSAT, and IBM Thomas J. Watson Research Center in mobile communication networks. During 1991 to 1998, he was with the Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan, ROC. Since 1998, Dr. Chen is a Professor at Institute of Communication Engineering, College of Electrical Engineering, National Taiwan University, Taipei, Taiwan, ROC. He was a visiting scientist with Hewlett-Packard Laboratories in California USA during 1997 and a visiting Professor at the Delft University of Technology, Netherlands, 1998 summer. Dr. Chen is also adjunctly appointed by the Executive Yuan Science and Technology Advisory Group to plan Taiwan's communication and networking technologies, and is also appointed by the Ministry of Transportation and Communications as a member of Telecommunication Commentary Board from 2001 to 2004. Dr. Chen actively involves the technical organization of numerous leading IEEE conferences, including as the Technical Program Committee Chair of *1996 IEEE International Symposium on Personal Indoor Mobile Radio Communications*, and TPC co-chair for IEEE Globecom 2002. He has served editorship with the following prestigious international journals: *IEEE Transaction on Communications, IEEE Communications Letters, IEEE Communication Surveys, IEEE Personal Communications Magazine, International Journal of Wireless Information Networks, IEEE Journal on Selected Area in Communications, ACM/Blatzer Journal on Wireless Networks*. He has been a voting member for IEEE 802.11 (wireless LANs), IEEE 802.15 (Wireless Personal Area Networks), IEEE 802.14 (HFC modem) international standard working groups, and participating US TIA45.5 CDMA Cellular standard, ETSI SMG2 cellular standard, and ITU-R TG8/1 IMT-2000 (3G) standard. He has authored and co-authored over 160 technical papers and 10 granted/pending US patents. Dr. Chen was elected as an *IEEE Senior Member* in 1993, one of Ten *Outstanding Young Engineers* in 1994, one of *Ten Outstanding Young Persons* (the most prestigious achievement award for people under age 40 in Taiwan) in 1996, *NSC Excellent Research Award* in 2000, *ISI Citation Classic Award* for high-impact research in 2001, *Outstanding Engineering Professor* in 2002, and listed in the 15th edition *Marquis Who's Who in the World* in 1998 and *Who's Who in Industry* in 1999, and is the IEEE Communication Society Asia Pacific Board Director during 2002–2003. He led APEC Telecommunication

Working Group WTO Implementation task group. Dr. Chen was invited as a speaker in the United Nation ITU TELCOM 95 Technology Summit and Asia TELCOM 97 Strategy Summit. Dr. Chen's research interests include wireless communications and broadband access networks.

**Chun-Ying Wu** received his B.S. degree in electrical engineering and M.S. degree in communication engineering from National Taiwan University, Taipei, Taiwan, in 2000 and 2002, respectively. His research interests include wireless networks and QoS control in data networks.