Utility-Based Predictive Services for Adaptive Wireless Networks With Mobile Hosts

Floriano De Rango, *Member, IEEE*, Peppino Fazio, and Salvatore Marano, *Member, IEEE*

***Abstract*—Over the past few years, there has been a growing interest for research in wireless adaptive networking and resource management to efficiently handle wireless communications between mobile hosts and base stations. Moreover, in wireless environments, the bandwidth of an ongoing multimedia flow can dynamically be adjusted, so there must be an efficient bandwidth allocation scheme to ensure quality-of-service (QoS) guarantees and high system utilization. In this paper, the attention is focused on the management of mobile services. In particular, two classes of service, i.e., *mobility independent predictive* (MIP) and *mobility dependent predictive* (MDP), have been considered, as defined in integrated service packet networks. A utility-based rate adaptation algorithm has been considered, and an admission control has been proposed, taking into account channel conditions, through a slow-fading channel model for wireless LAN 802.11b. The valued algorithm is based on a user utility function, and the admission control can use the prereservation phase among potentially visited cells from MIP hosts while only considering the bandwidth availability on current cells for MDP services. We analyzed the MIP users' mobility along coverage areas to reduce passive resource reservations on cells that users will probably never visit through the knowledge of some mobility parameters. A prediction technique is also proposed. The performances of the wireless system have been evaluated in terms of total bandwidth utilization for MIP services, average user perceived utility, and system outage probability.**

***Index Terms*—Adaptive bandwidth allocation, admission control, fading channels, Markov model, mobility management, partial prereservation, utility function, wireless networks.**

I. Introduction

Recently, a rapid growth of mobile devices has been observed, and the demand for wireless communications has rapidly increased. The adaptive multimedia networking paradigm can play an important role to mitigate the highly varying resource availability in wireless/mobile networks, where users require different levels of quality-of-service (QoS). Wireless networking determines special problems like limited bandwidth and high error rates due to fading and mobility [1]–[3] effects.

This paper takes into consideration both multipath fading and mobility effects. Fading has been considered through a Markov channel modeling, accounting the slow-fading effects, whereas users' mobility has been taken into account through the management of *handoff* events among the visited cells. To offer an adaptive QoS to mobile hosts, an architecture capable of reserving bandwidth levels and of offering guaranteed services is used. To accomplish this task, the Integrated Service (IS) network with mobile hosts, while the Mobile Reservation Protocol (MRSVP) is applied to exchange state information in the wireless environment [4], [5]. It is based on active and passive reservations, and it is capable of prereserving a certain amount of bandwidth for mobility independent predictive (MIP) (for tolerant real-time applications that can allow some bounded data packet delay variations) flows in the current cell and in those that the mobile host will probably visit to guarantee the desired QoS during handoff events while serving mobility dependent predictive (MDP) requests (for applications that can suffer continuous QoS degradations or connection droppings). This way, the effects of mobility for MIP connections are minimized. By an analysis of users' mobility, a prediction can be made for MIP flows to evaluate the number of cells that a flow will probably visit, achieving higher system utilization. Moreover, different users may experience different link capacities due to different locations, and the bandwidth should be allocated in an adaptive and link-state-dependent way. To consider the heterogeneity of different applications and to have a consistent performance measure, we adopt utility functions in our adaptive QoS model [6]–[8]. In this paper, we propose a utility-oriented bandwidth allocation scheme and an admission control policy, which accounts for users' QoS requirements and actively adapts to the dynamics of the physical channel. There has been much work on wireless resource management, focusing on multiple access and channel allocation, but there is less research on adding explicit adaptive mechanisms to bandwidth allocation schemes to deal with the variations of wireless channels.

This paper is organized as follows. Section II gives a brief overview of related work on wireless adaptive network study focusing on the rate adaptation algorithms, resource reservation protocols, and pattern-prediction techniques. Section III presents a short view of the MRSVP protocol, channel, model, and provided service classes in Integrated Service Packet Networks. Section IV describes the proposed algorithms of bandwidth allocation and admission-control schemes. A prediction technique is presented in Section V. In Section VI, the analytical expression for the cell stay time (CST) adopted in the prediction model is introduced. The simulation results are also shown in Section VI, whereas Section VII concludes this paper.

II. Related Works

In the wired network environment, several techniques have been proposed in the academia and industrial community to provide QoS to applications. Resource reservation mechanisms such as *resource ReServaVation Protocol* (RSVP) [9], [10] and packet scheduling algorithms have been used to satisfy throughput and end-to-end delay of the applications. However, mobile computing environments require different and more flexible mechanisms because they should address some other specific issues. First of all, wireless channels are subject to bursty and location-dependent errors, the wireless channel condition is location dependent, and mobile users can move from a lightly loaded cell to a heavily loaded cell. These aforementioned issues determine an adaptive nature of wireless networks, and it is also important to have a resource-reservation mechanism that is able to provide resources not only on the current cell to offer QoS during the user's movement. This means that in the literature, a lot of rate adaptation mechanisms, signaling, and resource-reservation protocols are proposed.

In [5], [11], and [12], the authors propose some rate adaptation schemes based on some specific QoS indexes such as fairness, degradation degree, etc., to dynamically adjust the bandwidth of the ongoing calls, particularly in overloaded situations. So in wireless adaptive networks, the adaptation of applications at multiple bandwidth levels becomes a key issue.

In IS networks, each flow can receive different QoSs, which must be negotiated at the beginning of sessions, between flows and net, by the RSVP protocol [9], [10] or by the MRSVP or DRSVP protocol in mobile scenarios [4], [13]–[15]. In addition, another signaling protocol, i.e., *Next Step In Signaling* (NSIS), was proposed in the literature. In 2001, the Internet Engineering Task Force created the NSIS Working Group to solve new signaling needs [16]. The *QoS-NSIS Signaling Layer Protocol* (QoS-NSLP) is used to signal QoS reservations. The specific information to each QoS model is encapsulated in a *QoS Specification* (QSPEC) field that permits to define the *QoS desired* and the *minimum QoS*.

Since mobility and resource management are critical to provide QoS in wireless networks, it is very important to accurately describe movement patterns of mobile users in wireless cells. Two prediction-based resource reservation techniques are proposed in [17]. These techniques consider the Wiener prediction theory and the time series analysis to make a predictive resource reservation under non-Poisson and/or nonstationary arrival processes, arbitrary distributed call and channel holding time, and arbitrary per-call resource demands. In [18], a hierarchical user mobility model based on an appropriate pattern matching and Kalman filtering is presented. This approach permits obtaining the necessary information for advance resource reservation and optimal route establishment in wireless ATM networks.

In [20], a framework to estimate service patterns and to track mobile users is proposed. The paper is based on historical records and predictive patterns of mobile users that permit estimation of the next cells into which a mobile user will possibly move. The same authors proposed a new location management scheme for the mobile terminals (MTs) roaming across multitier personal communication systems (PCSs) with different technologies or protocols [21].

In [23], some studies on the call holding time (CHT) and cell residence time (CRT) of the novel PCS networks have been led out. The authors show that classical assumptions of exponentially distributed CHT and CRT are not appropriate in a real context. They propose some more realistic distributions that describe the CHT and CRT trend of mobile users.

A scheme for resource reservation and call admission control algorithm has been proposed in [24]. In the paper, the authors make use of handoff prediction to deploy bandwidth resources to mobile users among the visited cells. The proposed reservation scheme is based on the location estimation of the mobile user, the instantaneous variation of the speed, and the direction of mobile stations.

Our previous works addressed the issue of mobile adaptive wireless networks. In particular, in [25], the management of predictive independent (MIP) and dependent services (MDP) from mobility in adaptive multimedia networking is considered. To offer an adaptive QoS (soft QoS) that increases the total wireless system utilization, a utility-based rate adaptation algorithm is considered. In [25]–[27], a prediction technique based on the CST evaluation of a mobile user under a random waypoint mobility model (RWPMM) is proposed. In the paper, a formula that binds speed, cell radius, and variation around the average speed is calculated, and it is utilized in the paper, as explained in Section V.

This paper, instead, presents a unified approach where a rate adaptation algorithm that considers the channel state, a prediction technique based on the CST evaluation, and a resource reservation protocol (MRSVP) are jointly applied and integrated to offer hard-QoS and soft-QoS services to the mobile user's applications. In the following, the key elements of our proposal are presented.

In particular, the main contributions of this paper are listed as follows:

1) channel modeling and implementation through a discrete-time Markov chain (DTMC) to account for the fading effect and to consider the wireless channel fluctuations (Section III-C);
2) implementation of the MRSVP protocol to manage two user types, i.e., users that need some QoS guarantees during their movements and users that can also be dropped if there is not enough available resources (Section III-B);
3) proposal of a call-admission control algorithm that permits accepting different traffic classes associated with the MRSVP protocol (Section IV);
4) proposal of a rate-adaptation scheme based on utility functions and on the wireless channel state to optimize users' utility and system utilization (Section IV);
5) proposal of a reservation scheme combined with a prediction technique to offer soft-QoS guarantees during user movements. Different from classical approaches in the literature where a cell-by-cell reservation was realized, our approach tries to make in-advance bandwidth reservations on more cells simultaneously to offer enough resources to mobile users in the admission phase,

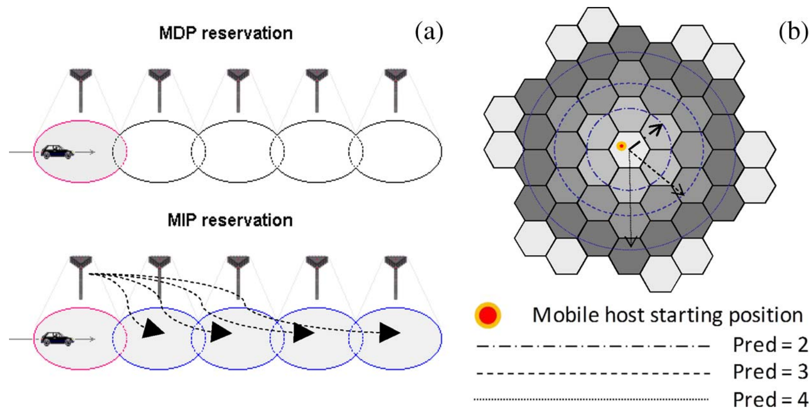


Fig. 1. Different kinds of wireless passive reservations. (a) Unidimensional coverage. (b) Bidimensional coverage.

reducing the QoS degradation or call dropping probability (Section V).

III. Service Classes, MRSVP, Link, and Mobility Modeling

A protocol that permits resource-reservation supporting node mobility is considered in this paper. It allows two flexible bandwidth reservation modalities associated with two traffic classes, i.e., MIP and MDP services. They obey the paradigm of soft QoS that is sometimes more appropriate in mobile adaptive wireless networks, such as the wireless LAN (WLAN) considered in this paper. MRSVP has been applied to a WLAN cluster where an 802.11b channel model has been implemented, as explained in the following.

A. Mobility-Independent and -Dependent Service Classes

Internet best-effort service does not offer any guarantee about available bandwidth, network propagation delays, jitter, and packet delivery. As a consequence, there have been different research groups that tried to define some service models to deal with applications variety in packet networks. IS networks are the results of such kinds of works, as described in [4] and [5]. In a real network, resource reservations can be made by protocols to satisfy QoS requirements and to offer to mobile hosts a service "better than best effort," accounting for the inherent time-varying environmental conditions evident in radio communications (e.g., fading). In IS networks, each flow can receive different QoSs, which must be negotiated at the beginning of sessions, between flows and net, by the RSVP protocol [9], [10] or the MRSVP protocol in mobile scenarios [4]. There are three provided service classes [5]: mobility independent guaranteed (MIG, for hard and intolerant applications that need absolute guarantees on packet delays), MIP (for tolerant real-time applications that can suffer from limited data packet delays), and MDP (for applications that can experiment continuous QoS degradations or connection droppings). In this paper, only MDP and MIP classes have been considered. The MRSVP protocol is used for exchanging the state information of wireless networks, and it can offer soft QoS (adaptive QoS) for MIP and MDP services.

B. Mobile RSVP

To handle user mobility and offer guaranteed services (independent from mobility), the ReSerVation Protocol [9], [10] has been extended with the MRSVP [4]. This way, the handoff events can be managed in an adequate manner, and the mobile users can make reservation requests over more than one cell by their *proxy agents*, e.g., *local proxy agents* (which handle active reservations) and *remote proxy agents* (which deal with passive reservations). An active reservation is made by a user only on the current access point (for MDP class, as we see later), whereas passive reservations are made only on the remote cells that the user will visit during its connection (users belonging to the MIP class requests passive reservations). An MRSVP connection starts with a *proxy discovery protocol* phase in which the user can know the addresses of its remote agents. Then, a resource request can be made, which will reach the net sender, to begin data packet transmission. After the proxy addresses are discovered, users send active_RESV messages to their local access points and passive_RESV messages to their remote access points, so the system must effect an admission control (as explained in next sections) to accept or refuse the users' requests. When a user moves from a coverage area to another one, the handoff event is managed by making a new request (MDP class) or by a reservation switch (MIP class). The reserved resources in the old access point are released in both cases, and if the user belongs to the MIP class, the passive resources can be assigned by switching to an active reservation. For more details about MRSVP, see [4]. Fig. 1 describes the main difference between MDP reservation request (only on the current active cell) and MIP reservation request (on the current and on the passive ones). Information about the cells that a user will visit during its active connection is carried out by the MRSVP through the exchange of the mobility specification message [4], but the use of a prediction algorithm is necessary when users move among a 2-D set of cells. The dotted lines in Fig. 1(a) represent passive reservation requests.

So, the same considerations must be extended in the 2-D environment, where the directional behavior of mobile hosts must be considered. In this paper, as shown in Section V, a circular reservation policy is introduced. If a user will visit C_e cells, then the passive reservations are made on a circle of cells with a radius of C_e cells. Fig. 1(b) shows the mapping of MIP passive reservations in a 2-D environment for the cases of $C_e = 2$, $C_e = 3$, and $C_e = 4$. Details about the circular reservation are given in Section V.

C. Radio Link Model

In this paper, we employed a Markov chain model to describe the behavior of a radio link between users and access points, as proposed in [3] and [28]. We needed to introduce the chain model to consider more realistic conditions in wireless communications and the fluctuations in the received signal level due to the various propagation phenomena during a generic connection (shadowing, refraction, fading, etc.). As we will see, each chain state has an associated ratio, which represents the received percentage of corrupted bits. The model can only be used under the assumption of *slow fading*.

Letting $S = \{s_0, s_1, \ldots, s_{K-1}\}$ denote a finite set of states and $\{S_n\},\ n = 0, 1, 2, \ldots$ be a constant Markov process with the property of *stationary* transitions, then the transition probability is independent of the time index n and can be written as

$$t_{j,k} = P_r(S_{n+1} = s_k | S_n = s_j) \tag{1}$$

for all $n = 0, 1, 2, \ldots$, and $j, k \in \{0, 1, 2, \ldots, K-1\}$. We can define a $K \times K$ state transition probability matrix $\boldsymbol{T}$ with elements $t_{j,k}$. Moreover, with the stationary transition property, the probability of state k without any state information at other time indexes can also be defined as $p_k = Pr(S_n = s_k)$, where $k \in \{0, 1, 2, \ldots, K-1\}$, so a $K \times 1$ steady probability vector $\boldsymbol{p}$ can be defined with its element p_k. To complete the description of the chain model, we require additional information on the channel quality for each state so we can define a $K \times 1$ crossover probability vector $\boldsymbol{e}$ with its elements e_k, $k \in \{0, 1, 2, \ldots, K-1\}$. Now, the finite state Markov chain (FSMC) is completely defined by $\boldsymbol{T}$, $\boldsymbol{p}$, and $\boldsymbol{e}$. More details on the modeling and a graphical representation can be found in [3].

Under the hypothesis of a Rayleigh-distributed received signal envelope, we can derive a relationship between the physical channel and its finite-state model by partitioning the range of the received signal-to-noise ratio (SNR) into a finite number of intervals. Letting $0 = A_0 < A_1 < A_2 < \cdots < A_k = \infty$ be the thresholds of the received SNR, then the Rayleigh fading channel is said to be in state s_k, $k = 0, 1, 2, \ldots, K-1$, if the received SNR is in the interval $[A_k, A_{k+1})$. Associated with each state is a crossover probability e_k, and given a specific digital modulation scheme, the average error probability is a function of the received SNR (the value of e_k is the average error probability on transmitting a bit when the received SNR falls in the kth interval). The elements of $\boldsymbol{p}$ and $\boldsymbol{e}$ can be written as

$$\begin{aligned} p_k &= \int_{A_k}^{A_{k+1}} \frac{1}{\rho} e^{-\frac{a}{\rho}} da \\ e_k &= \left[\int_{A_k}^{A_{k+1}} \frac{1}{\rho} e^{-\frac{a}{\rho}} P_e(a) da \right] \Big/ p_k \end{aligned} \tag{2}$$

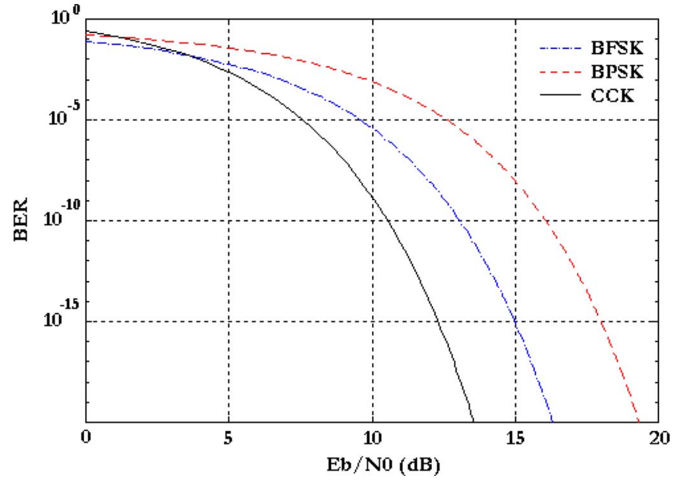


Fig. 2. SNR and BER for some digital modulation schemes.

where $P_e(a)$ depends on the digital modulation scheme chosen. In our simulations, we considered the complementary code keying (CCK) modulation (as in the standard IEEE 802.11b) [29], and $P_{e(\mathrm{CCK})}(a) = 12Q(2\sqrt{a})$.

Due to the nonlinearity between SNR and e_k, the SNR intervals may have to be nonuniform to be useful. Fig. 2 illustrates the course of bit error rate (BER) versus SNR for some kind of useful digital modulation schemes (*binary frequency shift keying*, *binary phase shift keying*, and *complementary code keying*).

The thresholds A_k, with $k = 0, 1, \ldots K-1$, can be calculated by choosing the desired values of crossover probabilities directly dependent from the degradation percentage of each state by an entropy function and solving the previous expression for e_k (in our paper we proceeded in numerical way). We referred to [28] using Jake's level-crossing analysis with the help of Mitra's producer–consumer queuing model [30]. The entire modeling technique is based on the knowledge of the channel state information (CSI). We assumed that, given the channel capacity, we can obtain the related values of crossover probabilities by using an entropy function. In particular, when the CSI is available, the channel capacity is the average capacity over all the states

$$C = \sum_{k=0}^{K-1} p_k \left[1 - h(e_k)\right] \tag{3}$$

where $h(\cdot)$ is the binary entropy function defined as

$$h(e) = e \log \frac{1}{e} + (1-e) \log \frac{1}{1-e}. \tag{4}$$

To calculate the transition probabilities $t_{j,k}$, we first assumed that the Rayleigh fading channel is slow enough that the received SNR remains at a certain level for the time duration of a channel symbol. Furthermore, the channel states associated with consecutive symbols are assumed to be neighboring states. If f_d is the maximum Doppler shift introduced by the user mobility and T the symbol transmission time, then we say that the slow-fading condition is verified if $f_d * T \ll 1$.

D. Mobility Modeling

The choice of an appropriate mobility model plays an important role in bandwidth assignments and network dimensioning. Many works in the literature face this problem, but most of them are based on some simplifications about users' behavior and do not lead to any analytical expression. So, the choice of mobility model has a heavy impact on the obtained results, which can be unrealistic if the model is not appropriate. This paper is based on the RWPMM [19], [34] for a 2-D set of cell clusters, and for statistical analysis, the smooth random mobility model (SRMM) [22] has also been taken into account.

The random waypoint model (RWPM) model is a simple and straightforward stochastic model that describes the movement behavior of a mobile network node in a given system area as follows. A node randomly chooses a destination point (waypoint) in the area and moves with a constant speed on a straight line to this point. After waiting a certain pause time, it chooses a new destination and speed, moves with constant speed to this destination, and so on. The movement of the mobile node can be modeled as a stochastic process $\{(p_i, v_i, \tau_{p,i})\}_{i \in N}$ (where N is the set of natural numbers). A movement period can completely be described by the vector $(p_{i-1}, p_i, v_i, \tau_{p,i})$, such as referred in [19] and [34]. When a single random variable of a process is considered, it is possible to omit the index i and just to indicate P, V, and T_p as the next selected point where to move, movement speed, and pause time, respectively. The values of P (x and y in the 2-D case) are selected by a uniform distribution in the range, respectively, $[0, x_{\max}]$ and $[0, y_{\max}]$ (where $x_{\max}$ and $y_{\max}$ determine the grid size), and speed is uniformly selected in the range $[v_{\min}, v_{\max}]$ with $v_{\min} > 0$ and $v_{\max} < \infty$, and T_p is uniformly selected in the range $[0, \tau_{\max}]$. More details about monodimensional (1-D) and bidimensional (2-D) RWPM can be found in [19] and [34].

On the other hand, the SRMM model makes users' movements smoother and more realistic than previous random models because it relates speed and direction changes. The main concepts of the SRMM are two stochastic processes for direction φ and speed v. Their values are correlated to previous ones to avoid unrealistic patterns and speed/direction changes; e.g., if a user is moving with high speed, a direction change cannot have high φ variations. Speed and direction changes follow two Poisson processes, and different typical patterns or environments can be modeled by setting some parameters, like the set of preferred speeds. This model is also based on a set of preferred speeds in the range $[v_{\min}, v_{\max}]$, and a mobile host moves with constant speed until a new target speed $v*$ is chosen by the stochastic process, so it accelerates/decelerates to reach v^*. Another key parameter is the p_φ value. It represents the probability of direction change and plays an important role in characterizing the users' behavior. The set of preferred speeds $\{v_{\text{pref0}}, v_{\text{pref1}}, \ldots, v_{\text{prefn}}\}$ is also defined to obtain a nonuniform speed distribution, with $p(v_{\text{pref}}) = p(v_{\text{pref0}}) + p(v_{\text{pref1}}) + \cdots + p(v_{\text{prefn}}) < 1$, $v_{\text{pref0}} < v_{\text{pref1}} < \cdots < v_{\text{prefn}}$, $v_{\max}$ is a fixed threshold, and $p(v_{\text{prefx}})$ is the probability that a mobile host moves with speed v_{prefx}. Let $t*$ denote the time at which a speed change event occurs, and a new target speed $v* = v*(t*)$ is chosen. Now, an acceleration $a(t^*) \neq 0$ must be set (or a is set to 0 if $v*(t*) = v(t*)$) considering other two variables, i.e., $a_{\max}$ and $a_{\min}$. The first one represents the maximum possible acceleration and the second one the maximum possible deceleration. In discrete instant times, the new speed $v(t)$ is changed according to the uniformly accelerated motion as $v(t) = v(t - \Delta t) + a(t)$ until $v(t)$ achieves $v*(t)$. For more details about SRMM, see [22].

IV. Bandwidth Allocation and Admission Control Schemes

We used a *utility-oriented* algorithm for rate adaptation and admission control, considering the time-varying nature of links, between hosts and access points [6]. In our case, we used monotonically nondecreasing utility functions [7], describing how the perceived utility changes with the amount of effective bandwidth received by the user. In our service model, each user i can signal its utility function $U_i(r)$ to the network, where r is the amount of effective bandwidth received by the user and the bandwidth allocated to a flow can take its discrete value from the set $B = \{l_1, l_2, \ldots, l_t\}$, where $l_i < l_{i+1}$ for $i = 1, \ldots, t-1$. It is assumed that service requests can belong to MIP or MDP class, and all of them take bandwidth values from the same set B.

As described earlier, the communication link of each user can be modeled by a K-state Markov chain. We can indicate the average state holding time of each state m with t_m and the bandwidth degradation ratio of the state with D_m, where $0 \leq D_m < 1\ \forall 1 \leq m \leq K$.

If, at a particular time instant, r_i is the amount of bandwidth that the network is allocating to user i, we define the *received instant utility* as $u_i = U_i((1 - D_{i,m}) * r_i)$. One of the objectives of the bandwidth allocation scheme is to guarantee the minimum utility level for each user i. If we define *utility outage* as the event that user i's instant utility level falls below its minimum level, the scheme should guarantee that the probability of a utility outage is smaller than a certain threshold p_{outage}. In addition, the fairness criterion should also be based on utility. Considering users i and j with average utility $u_{i,\text{avg}}$ and $u_{j,\text{avg}}$, respectively, we can define the *normalized gap* of the average utility received and the minimum level $u_{*,\min}$ as $G_i = (u_{i,\text{avg}} - u_{i,\min})/u_{i,\min}$, so we want all users to have the same normalized gap in the long run ($G_i \approx G_j, \forall i, j$).

The proposed reallocation algorithm is carried out by each access point when the bandwidth must be redistributed after a channel link quality variation, a user admission, or a user call termination. To manage different classes of service, the bandwidth allocation can differently be made for users belonging to MIP or MDP classes. Once the amount of bandwidth that must be reallocated is determined, the algorithm determines the set of users that will be interested in the reallocation. All admitted users in an Access Point are sorted in a normalized gap list. When a user's link *degrades*, it may surrender some bandwidth to another user, with a smaller normalized gap. When a link *upgrades*, the user may receive some bandwidth from another one, with a larger normalized gap. This way, there is a net gain in the combined instant utility. If, at a particular time, user i''s link state changes to state p, the following steps are performed, obtaining a sorted normalized gap list: 1) All users' average utility level and normalized gap are updated. 2) Users are sorted in increasing order of normalized gap. 3) If the instant utility level of user i is below the minimum (outage event), some users' bandwidth will be reduced and reallocated to user i to meet its $u_{i,\min}$. 4) If there is no step 3, user i may give up part of its bandwidth to another user if the link degrades, whereas it may receive some bandwidth if the link upgrades. We call the user who gives up part of its bandwidth to others the *benefactor*, and the user who receives bandwidth from others the *beneficiary*. In the third step, to satisfy user i's $u_{i,\min}$, the scheme searches for *benefactor(s)* starting from the user with the largest normalized gap. Suppose that the user with the largest normalized gap is user j whose link is currently in state q and it is above $u_{j,\min}$. User j will yield

$$\min\left(\frac{r_{i,\min}}{1-D_{i,p}} - r_i, r_j - \frac{r_{j,\min}}{1-D_{j,q}}\right) \tag{5}$$

amount of bandwidth to user i, where r_i and r_j are the bandwidth allocated to users i and j, respectively, before the link state transition. This procedure will be repeated until $u_{i,\min}$ is reached or all the users have been checked.

When user i's link degrades, the scheme will search for an appropriate *beneficiary*, checking the users in increasing order of normalized gap. When the *beneficiary* is found, the scheme decides the amount of bandwidth to transfer between the users, trying to maximize their combined utility. This procedure is repeated until one *beneficiary* is found or all users with smaller normalized gap than user i's have been checked. Similarly, when user i's link upgrades, user i becomes the beneficiary, and users with larger normalized gap are the candidates for *benefactor*. The scheme checks the candidates in decreasing order of normalized gap, and when the *benefactor* is found, the scheme decides the amount of bandwidth to exchange, maximizing the combined utility of the two users. In addition to the link state changes, adjustments in bandwidth allocation are also needed when a user arrives (new user) or departs. If r is the amount of bandwidth that needs to be collected from current users because of a user arrival, user j with the largest normalized gap G_j is to give up

$$\min\left(\max\left(0, r_j - \frac{r_{j,\min}}{1-D_{j,q}}\right), r\right) \tag{6}$$

amount of bandwidth, where q is the current link state of user j. Similarly, if there is surplus bandwidth, users with the first k smallest normalized gaps are chosen to receive the surplus bandwidth. Each user can increase its effective bandwidth up to the maximum effective bandwidth level. The time complexity of the bandwidth reallocation algorithm can be evaluated. Let us hypothesize that there are n admitted users. First of all, an update of the normalized gap must be made [its complexity is $\theta(n)$]. Then, the list of users must be sorted [the best performance can be obtained in $O(n \log n)$]. At this time, the exchange of bandwidth can be made. The benefactor and beneficiary are found with a single list scan in $O(n)$ time, so the algorithm performs with a time complexity of $O(n \log n)$ in the "worst case." To guarantee the users' minimum utility level, an admission control policy should be enforced to limit the number of users in the system. When a user's instant utility falls below its minimum level, there is a utility *outage* for the user, and the probability p_0 of such an event at any time is

$$p_0 = P_r\left\{\sum_{i=1}^{n} \frac{r_{i,\min}}{1-D_{i,m_i}} > R_a\right\} \tag{7}$$

where m_i is user i's link state at the time instance, n is the total number of users including the new one, and R_a represents the bandwidth associated to a wireless cell a. Modeling the wireless channel through FSMC, it is possible to know the value of p_0 in the worst case, accounting the channel state conditions in the following way:

$$\begin{aligned} p_0 &= \sum_{A} \prod_{1\le i\le n} p_{m_i} \\ A &= \Bigg\{ m_1, m_2, \ldots, m_n | 1 \le m_1, \ldots, m_n \le K \\ &\qquad \sum \frac{r_{l,1}}{1-D_{l,m_l}} > R_a \Bigg\} \end{aligned} \tag{8}$$

where p_{mi} is the probability of user i's link to being in state m at a particular time. The admission control algorithm works differently for two classes of service (MIP and MDP). For the MIP class, the flow is admitted if

$$\sum_{c=1}^{C} p_{0,c} \le C \cdot p_{\text{outage}} \tag{9}$$

where C is the number of cells that the mobile host will visit, and p_{outage} is the outage probability of the wireless system (for the MDP class, the condition must be verified only for the current cell). So, when a new user arrives, the scheme calculates $p_{0,c}$ (p_0 for cell c), and if $p_{0,c} \le p_{\text{outage}}$ for each cell, the new user is admitted; otherwise, it is rejected.

If a user j is admitted, it is initially allocated $r_{j,\min}/(1-D_{j,q})$, where q is the user j's current link state. The assigned amount of bandwidth to j is contributed by the users currently in the network following the algorithm we previously described. Fig. 3 resumes all the phases of the admission control algorithm.

Equation (8) indicates the evaluation that must be performed to decide whether a new user can be admitted into the system. The product of the n terms [time complexity $\theta(n)$] must be repeated for all the state combinations for which the condition $\sum(r_{l,1}/1-D_{l,m_l}) > R_a$ is verified, that is, n^K in the "*worst case*" (supposing that the disequation is verified for all combinations), where n is the number of admitted users (including the new one), and K is the number of states of the chain model. So the time complexity of the call admission control (CAC) scheme for an active reservation admission is $O(n \cdot n^K) = O(n^{K+1})$. For passive reservation, there is the multiplicative factor C that can be disregarded in the "*worst-case*" complexity analysis. It is noticed that the temporal complexity is

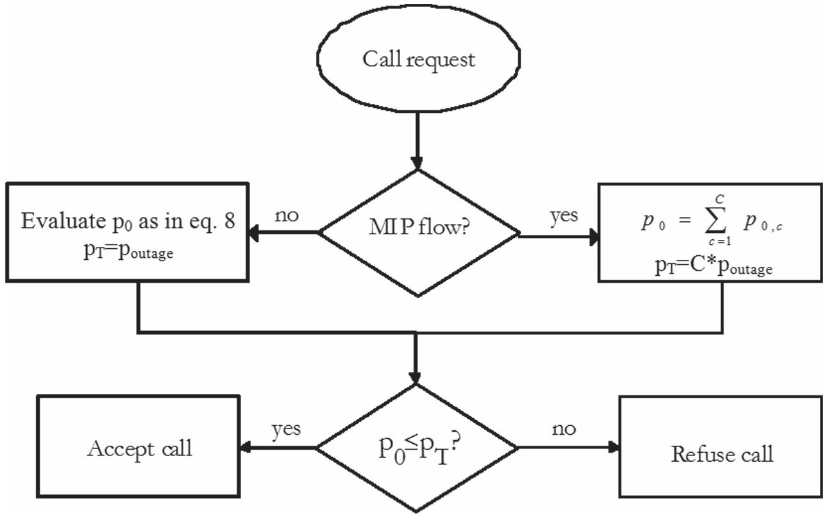


Fig. 3. Call admission control data flow diagram.

proportional to the number of states of the chain model and the number of current admitted users, so the needed time increases when n or K increases. The needed time may become unacceptable, so the on-the-fly evaluation of (8) must be avoided. If the number of chain states K is known, as well as the t bandwidth levels $B = \{l_1, l_2, \ldots, l_t\}$, the evaluation of (8) can be *a priori* made and stored in a CAC matrix, where the columns indicate the number of current admitted users, and the rows indicate the amount of available bandwidth as explained in the performance evaluation section. This way, the space complexity is slightly increased, but the admission can be made in a constant time (only a selection on the matrix and a comparison must be led out).

V. Bandwidth Partial-Reservation Policy

As exposed in the previous sections, passive reservations represent a good way to guarantee and maintain QoS to MIP users during handoff events. This policy is based on "in-advance making" bandwidth reservations over all the cells in the network without evaluating neither the average host's speed nor CHT. In this section, we propose a new criterion for increasing system utilization while maintaining the prereservation policy, improving the performances of the WLAN system. This time, passive reservations are made after evaluating the average hosts' speed and call duration. If the host moves in cells with extremely slow speed, prereserving over all system base stations is not necessary because the user will probably never visit all of them. Moreover, if two mobile hosts have the same speed, the number of visited cells may vary in function of call duration. So, a model for estimating the CST has been evaluated with the help of lots of "monitor simulations" (simulations dedicated to the observation of some system parameters, without any performance evaluation), and this information has been used to calculate the number of cells visited by mobile hosts. This information can be used to make passive reservations only on the cells that a host will effectively visit, leaving the bandwidth availability in other cells. For deriving this model, a *Poisson arrival time distribution* and an *exponentially distributed CHT* have been considered for any mobile host.

A. Constant Speed Versus Variable Speed

In the first simulations, the campaign users' speed has been fixed to a constant value v_{avg}; that is to say, all users move along coverage areas with the same speed, so they only have different connection times. This kind of simulation has been employed to evaluate the model validity under simplified conditions (constant speed). As shown in next section, in this case, low prediction errors are suffered.

In a second campaign of simulations, the average speed has uniformly been selected in a range of $[v_{\mathrm{avg}} - \alpha, v_{\mathrm{avg}} + \alpha]$, where α is a variable percentage of v_{avg} (from 5% to 50%). As shown in the next section, the estimation error increases for high values of α, but it maintains itself in acceptable bounds.

In both cases (constant or variable average speed), for any fixed value of v_{avg}, a probability density function (pdf) of the average CST of mobile hosts has been derived to make a predictive evaluation of visited cells.

An analytical framework to evaluate the cell residential time called CST of a mobile user in a WLAN coverage area is presented in this section. The cell residential time is an important parameter in wireless networks because it permits the evaluation of how long a user will stay in a cell during its CHT and how many cells he will visit. This can be useful in resource reservations for an environment supporting node mobility [25]–[27].

In this paper, the random waypoint model (RWPMM) *for a 2-D environment* was considered [34] as a method to describe users' mobility, but first of all, many simulation runs have been carried out to obtain a certain number of samples of the average CST with an average speed v and a variation coefficient α (this way, the considered users' speed is uniformly distributed in the interval $[v - \alpha; v + \alpha]$). The number of runs that have been led out is 1000. This way, with the availability of 1000 CST samples, the distribution can well be observed, and after a result analysis, a CST distribution has been obtained, like

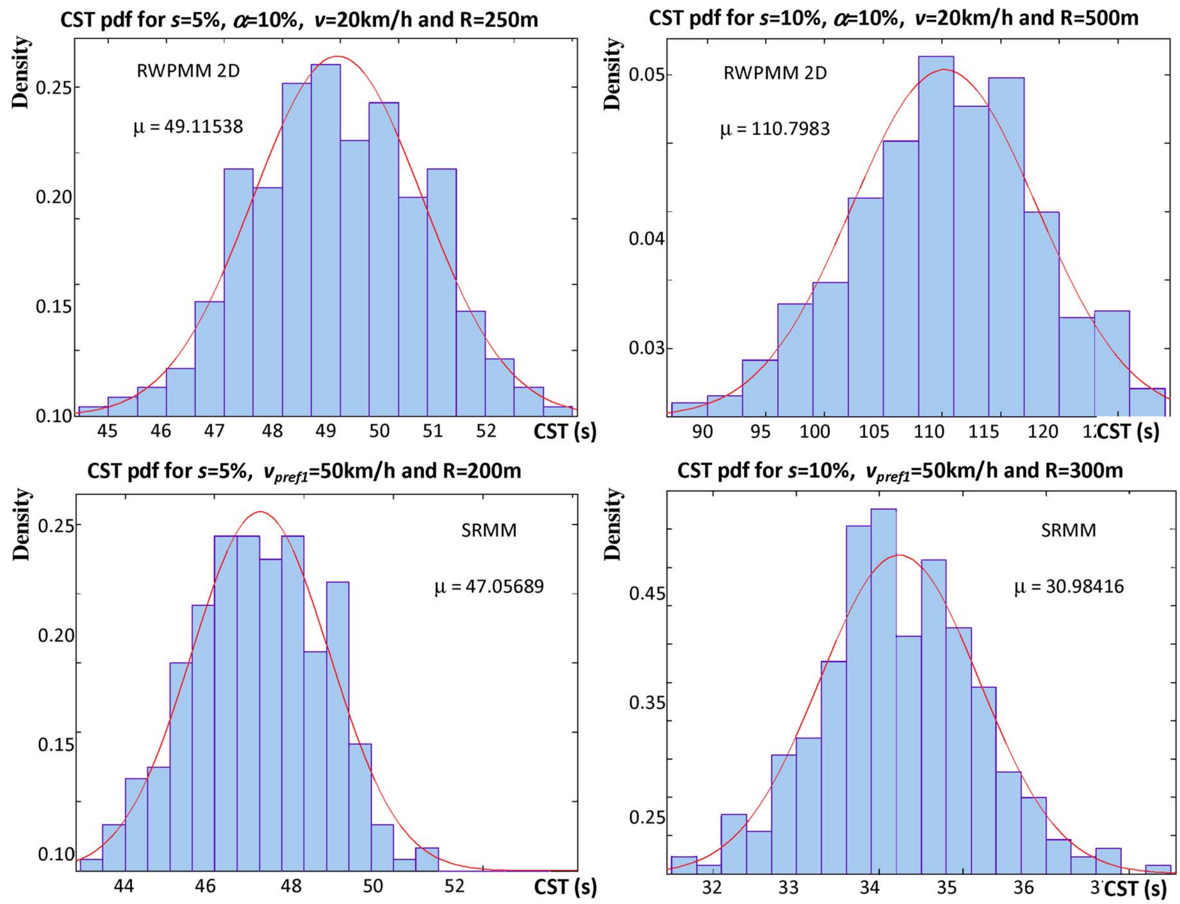


Fig. 4. Distributions and pdfs for two different mobility models and different system parameters.

the one depicted in Fig. 4, with a Gaussian approximation for fixed values of speed and variation. So the general expression of the CST pdf is

$$f_{X_{\mathrm{CST}}}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{10}$$

where $\mu = \mu_{\mathrm{CST}}(v, \alpha, R)$ and $\sigma = \sigma_{\mathrm{CST}}(v, \alpha, R)$ are, respectively, the average and standard deviations of the Gaussian distribution. R represents the cell radius; thus, it is possible to evaluate the error of the considered CST and to make a CST prediction based on confidence intervals and confidence levels considering the worst-case cell outage probability (COP). It is possible to select a CST T_{cst} for a mobile host so that $Prob(X < T_{\mathrm{cst}}) < 1 - \mathrm{COP}$, where X is normally distributed. T_{cst} is called a $(1 - \mathrm{COP}) * 100\%$ *upper confidence bound* for X. If the average CHT T_{cht} is known, it is possible to consider the term C called C_p (C partial) as $C_p = T_{\mathrm{cht}}/T_{\mathrm{cst}}$. So it is possible to use the C_p value to make the prereservation of MIP flows to leave more bandwidth availability in the not visited cells for new MIP flows. As an example, the same analysis has been led out for the SRMM, and a comparison with the RWPMM is shown in Fig. 4. It gives a description of how the CST is distributed under RWPMM and SRMM for different coverage radius R. In the RWPMM, α represents the variation around the average speed v, whereas for the SRMM, v_{pref1} represents the second preferred speed of users, as explained in Section III-D. In both cases, s represents the overlapping of

TABLE I
CST Distribution Mean and Standard Deviation for Different Mobility Parameters ($s = 5\%$, $v_{\mathrm{pref1}} = v_{\max} = 50$ km/h, $R = 200$ m)

Parameters	μ_{CST}	σ_{CST}
p_φ=0.1, p_{vpref1}=0.9	47.057	0.08153
p_φ=0.1, p_{vpref1}=0.1	51.182	0.1253
p_φ=0.9, p_{vpref1}=0.5	61.286	0.2259

adjacent cells, and the continuous line is the obtained Gaussian approximation of the CST pdf (all the diagrams are obtained by the execution of 1000 simulation runs). As previously exposed when users move in a 2-D environment, the circular reservation policy is mandatory. The predicted value of C_p can only be used to make passive reservations in a circular way, i.e., around the current cell [where the call has been admitted, as in Fig. 1(a)], so the number of required passive reservations C_r for MIP services increases with polynomial trend, such as $C_r = 3 \cdot (C_p \cdot ((C_p - 1)$. Table I illustrates some values of CST parameters. Our attention is not focused on the determination of all the possible CST distribution parameters because the proposed approach is independent of the chosen mobility model; however, when the value of p_φ is increased, a higher variance of the CST is observed, as will be confirmed later. The value of p_{vpref1} in the table represents the probability that a generic host assumes the value of v_{pref1} during its movements. The assumption of a

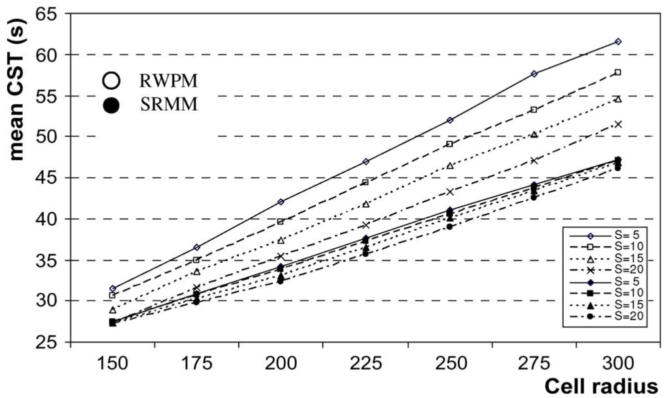


Fig. 5. μ_{CST} for different radius R and overlapping s values.

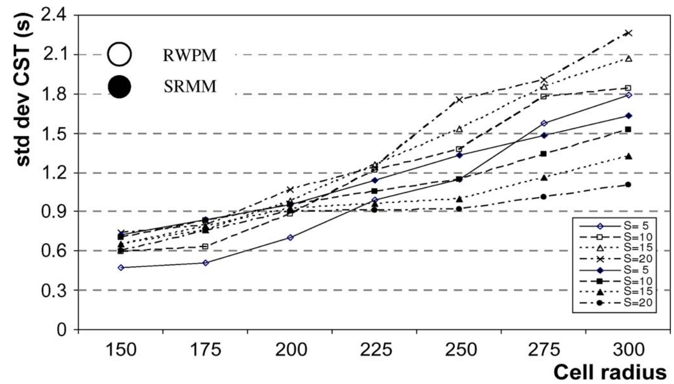


Fig. 6. Standard deviation of CST trend for different radius R and overlapping s values.

normally distributed CST, with different means and standard deviations depending on the fixed mobility parameters, has been verified through the Kolmogorov–Smirnov (KS) normality test [31]. Different p-*values* (a p-*value* for a comparison test represents the likelihood, under the assumption that the null hypothesis is true, that the data would yield the obtained results) have been obtained, showing that there is a negligible error for a CST Gaussian approximation (for RWPMM or SRMM). In the next section, analytical expressions for μ and σ are derived with the application of the regression theory. The *cumulative distribution function* (cdf) of the CST from (10) is

$$F(x) = P(X_{\text{CST}} \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt \tag{11}$$

and the probability that CST is lower than a value x with a fixed error threshold ε is given from

$$\begin{aligned} P(X_{\text{CST}} \leq x) &= P\left(Z \leq \frac{x-\mu_{\text{CST}}}{\sigma_{\text{CST}}}\right) \\ &= \Phi\left(\frac{x-\mu_{\text{CST}}}{\sigma_{\text{CST}}}\right) = 1-\varepsilon \end{aligned} \tag{12}$$

where $Z = (X_{\text{CST}} - \mu_{\text{CST}})/(\sigma_{\text{CST}})$, X_{CST}'s are random variables, and $\Phi(\cdot)$ is the standard Gaussian distribution. After the normalization of X_{CST} through the tabular values of the standard normal distribution, it is possible to obtain the CST estimation for a given threshold ε, such as referred in [32] and [33]. Thus, knowledge of μ_{CST} and σ_{CST} is necessary to obtain a good estimation of CST, depending on v, α, and R. Fig. 5 shows the mean CST distribution for different coverage area dimensions. As can be seen, it increases for higher radius values (because of the higher space to be crossed) and decreases for higher values of s (a higher value of cell overlapping reduces the single cell area). No big differences in the trend are observed between the considered mobility models.

Nevertheless, it must be underlined that the obtained values for RWPMM are higher than those of the SRMM because of the lower determinism of the model. A directional behavior is not manageable in the RWPMM, whereas the parameter p_φ has been fixed to 0.2 for the SRMM, so users have a lower chance to change direction, increasing the CST, as also observed in Table I.

Fig. 6, on the other hand, depicts the trend of the standard deviation of the CST distribution. The same course of the mean is obtained due the higher space to be crossed for higher radius (R) values. Users have a higher probability of changing direction and so a higher probability of remaining for a longer time in the same cell coverage.

In addition, in this figure, the higher randomness of RWPMM is evident for higher values of cell radius (more time to spend into a cell, so more chances to change direction) to obtain Figs. 5 and 6, and the users' average speed (RWPMM) and $v_{\max}$ (SRMM) have been fixed to 45 km/h.

B. CST Evaluation

To relate CST values with mobility parameters (v and α) and coverage radius R (three degrees of freedom), a regression analysis was performed under Matlab application, and the minimum observed value of the *determination coefficient* R^2 (this term is used to discriminate the determination coefficient from the cell radius R) over all obtained polynomial functions is 0.9898. The final formulas for the mean and standard deviation of CST are

$$\begin{aligned} \mu_{\text{CST}}(\bar{v}) &= [n_0, n_1, \ldots, n_4] \cdot [1 \quad v \quad v^2 \quad v^3 \quad v^4]^T \\ &= \langle n \rangle^T \cdot \langle \bar{v} \rangle^{n=4} \end{aligned} \tag{13}$$

$$\begin{aligned} \sigma_{\text{CST}}(\bar{v}) &= [m_0, m_1, \ldots, m_4] \cdot [1 \quad v \quad v^2 \quad v^3 \quad v^4]^T \\ &= \langle m \rangle^T \cdot \langle \bar{v} \rangle^{n=4} \end{aligned} \tag{14}$$

where the notation $\langle \cdot \rangle$ is used to represent a vector, and $\langle \cdot \rangle^T$ is the transpose operator applied to the vector. In (13) and (14), $\langle v \rangle^i = [1 \quad v \quad \ldots \quad \ldots \quad v^i]$ is a $(i+1)x1$ vector. Details on the polynomial regression technique can be found in [26].

VI. Performance Evaluation

To evaluate the goodness estimation of the CST and the performance of the rate-adaptation scheme, call admission control for MIP and MDP, and conformance to QoS parameters (outage probability, minimum received utility, and high system utilization), different simulations have been led out. Our

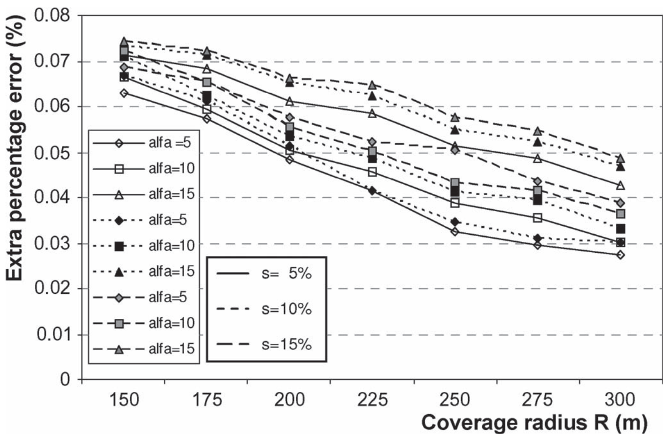


Fig. 7. Percentage number of extra visited cells in comparison with predicted cells.

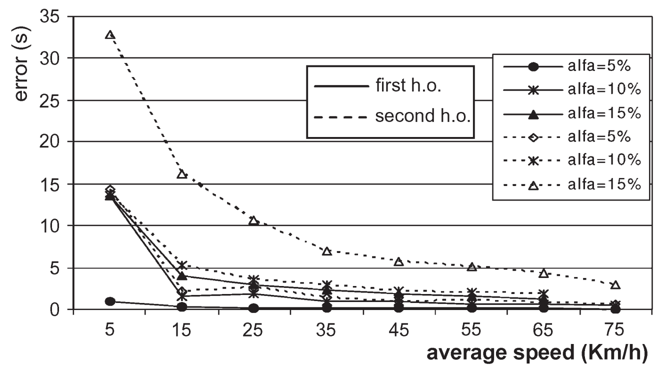


Fig. 8. Prediction error gap (expressed in seconds).

TABLE II
MARKOV CHAIN PARAMETERS

	p_i	e_i	D_i (%)	t_{mi} (s)
0	0.19681	0.303033	61.34	8.89e-3
1	0.49364	0.135814	39.73	8.572e-3
2	0.30954	2.97476e-4	0.2722	2.3e-1

simulated net consists of 49 wireless cells, each one covered by an access point [as illustrated in Fig. 1(b)]. In the simulation scenario, the access point bandwidth capacity is 5.5 Mb/s. The access points are wired connected by a Switching Subnet (SS) to the net sender.

To appreciate the accuracy of the proposed model, some simulation campaigns were carried out. The simulation scenario is based on a coverage radius length changing in the range [150–300] m. The mobile hosts move inside the coverage areas in a thoroidal way following a 2-D RWPMM with zero stop time (there are no pauses when the host decides to change its speed) [26]. The proposed model is employed in our WLAN-simulated net to predict the number of cells that the mobile host will probably visit. These values can be very important for prediction purposes.

An example of a possible application of the CST for prediction purposes is the reservation of bandwidth among cells that the mobile users will probably visit during the CHT. Fig. 7 shows the ratio in percentage value between the number of users that do not find the available resources and the number of total users that move among the WLAN cells (extra percentage error). The curves show that the percentage error is under 7.5%, confirming that the number of cells $n = \lceil \mathrm{CST}/\mathrm{CHT} \rceil$ predicted during the CHT represents a good estimation of the CST. Different values of α and s have been considered to appreciate the prediction algorithm performance in function of some system parameters.

Other simulations have been made to estimate the prediction errors introduced by our model. Different values of α have been considered, and Fig. 8 shows the difference between real and predicted handoff time instants for the first and second handoff events. It can be seen that the predicted times are more accurate for higher speed values because of the lower amount of time that a user can spend under a cell coverage, so the standard deviation of the CST decreases, giving the chance to make a more accurate prediction (curves are obtained for $R = 250$ m and $s = 10\%$).

In addition, it can be observed how the error increases as the allowed speed interval becomes wider (the variance increases because of the higher difference between average and real speeds). Obviously, the minimum error occurs when the lowest value of α is chosen ($\alpha = 5\%$). For the second handoff event, the same considerations can be made, but the error level increases due to the predicted error of the previous handoff event. The employment of a Markov chain represents a powerful way to describe the channel fluctuations during mobile communications. The main issue for this kind of analytical model is the partitioning criterion to obtain an efficient received SNR range partition. Each wireless link is described by a four-state Markov chain, with the parameters obtained as explained in Section III-C.

The set of possible discrete bandwidth levels (in kilobits per second) is $B = \{512, 640, 768, 896\}$, and the utility function is a discrete-level curve, so the set of instant utility values is $U = \{1, 2, 3, 4\}$ for all users (Table II).

In our simulations, the traffic load is composed by MIP and MDP flows in variable percentage. The bandwidth is managed by the policies illustrated in Section IV, and the outage threshold is the same for both flow classes. The mobile host can move with the average speed selected uniformly in the range [5,75] km/h. As explained in Section IV, the CAC is performed through the use of a predetermined matrix, as illustrated in Table III. As can be seen, the outage probability is nondecreasing for a higher number of admitted flows and for lower available bandwidth.

So, the number of admissible users depends on the chosen p_{outage} threshold. (The calculus of the 44 × 12 matrix in Table III takes about 1.8 s on an Intel Core 2 Duo at 2.0 GHz for an AP capacity of 5.5 Mb/s. The employed time increases for lower bandwidth granularity, higher AP capacity, and higher number of Markov chain states. For example, doubling the AP capacity to 11 Mb/s, it takes about 1 h and 27 min.) Maintaining the same bandwidth levels and the same granularity (128 kb/s), the algorithm can be applied "on-the-fly" when the considered AP capacity C is 5.5 Mb/s or lower, whereas for higher values of C, an "*a priori*" evaluation of the CAC matrix must be stored in the CAC module, indicated with $m = (C/granularity)$ and $n = (C/min\_liv)$, where C is the AP capacity, *granularity*

TABLE III
CALL ADMISSION CONTROL MATRIX

Av. Bandwidth	NUMBER OF ADMITTED USERS											
(kbps)	0	1	2	3	4	5	6	7	8	9	10	11
5632	0	0	0	0	0	0	0	0.01623	0.31279	0.83346	1	1
5504	0	0	0	0	0	0	0	0.03456	0.41601	1	1	1
5376	0	0	0	0	0	0	0	0.06521	0.52638	1	1	1
5248	0	0	0	0	0	0	0.00076	0.11156	0.63655	1	1	1
5120	0	0	0	0	0	0	0.0022	0.17962	0.73675	1	1	1
4992	0	0	0	0	0	0	0.00716	0.26652	0.82233	1	1	1
4864	0	0	0	0	0	0	0.01896	0.37031	0.88963	1	1	1
4736	0	0	0	0	0	0	0.03919	0.48754	1	1	1	1
4608	0	0	0	0	0	0	0.07785	0.60478	1	1	1	1
4480	0	0	0	0	0	0	0.13654	0.71489	1	1	1	1
4352	0	0	0	0	0	0.00253	0.21569	0.811	1	1	1	1
4224	0	0	0	0	0	0.00649	0.3221	0.88448	1	1	1	1
4096	0	0	0	0	0	0.01954	0.44202	1	1	1	1	1
3968	0	0	0	0	0	0.04821	0.56707	1	1	1	1	1
3840	0	0	0	0	0	0.09061	0.69211	1	1	1	1	1
3712	0	0	0	0	0	0.16609	0.79728	1	1	1	1	1
3584	0	0	0	0	0	0.26791	0.87916	1	1	1	1	1
3456	0	0	0	0	0.00836	0.38552	0.93867	1	1	1	1	1
3328	0	0	0	0	0.01886	0.52807	1	1	1	1	1	1
3200	0	0	0	0	0.05173	0.66377	1	1	1	1	1	1
3072	0	0	0	0	0.11779	0.77947	1	1	1	1	1	1
2944	0	0	0	0	0.19762	0.87826	1	1	1	1	1	1
2816	0	0	0	0	0.32895	0.94129	1	1	1	1	1	1
2688	0	0	0	0	0.48126	1	1	1	1	1	1	1
2560	0	0	0	0.02763	0.62025	1	1	1	1	1	1	1
2432	0	0	0	0.05368	0.76873	1	1	1	1	1	1	1
2304	0	0	0	0.13115	0.87844	1	1	1	1	1	1	1
2176	0	0	0	0.27153	0.94166	1	1	1	1	1	1	1
2048	0	0	0	0.39659	0.98499	1	1	1	1	1	1	1
1920	0	0	0	0.58468	1	1	1	1	1	1	1	1
1792	0	0	0	0.76232	1	1	1	1	1	1	1	1
1664	0	0	0.0914	0.86427	1	1	1	1	1	1	1	1
1536	0	0	0.14885	0.95713	1	1	1	1	1	1	1	1
1408	0	0	0.31065	1	1	1	1	1	1	1	1	1
1280	0	0	0.57029	1	1	1	1	1	1	1	1	1
1152	0	0	0.70063	1	1	1	1	1	1	1	1	1
1024	0	0	0.8775	1	1	1	1	1	1	1	1	1
896	0	0	1	1	1	1	1	1	1	1	1	1
768	0	0.30233	1	1	1	1	1	1	1	1	1	1
640	0	0.39733	1	1	1	1	1	1	1	1	1	1
512	0	0.65	1	1	1	1	1	1	1	1	1	1
384	0	1	1	1	1	1	1	1	1	1	1	1
256	0	1	1	1	1	1	1	1	1	1	1	1
128	0	1	1	1	1	1	1	1	1	1	1	1

represents the distance between a bandwidth level and an adjacent one, and min_liv is the minimum allowed bandwidth level. The dimension of the CAC matrix will be $m \times n$. Assuming that the elements of the CAC matrix are represented by double values (8 B each), then there will be $((m \times n) * 8)$ bytes of needed space. In our case, for $C = 5.5$ Mb/s, the CAC matrix will occupy $(44 \times 11 \times 8)$ bytes, that is, 3872 B, whereas for $C = 11$ Mb/s, the needed space will be 15 488 B. The following curves illustrate the performances of the utility-oriented algorithm for different values of outage threshold and mobile host speed in the absence of any predictive policy. The improvement in resource allocation for MIP users can be observed in Fig. 9 by increasing their traffic percentages from 20% to 80%. In Fig. 9 and in the following figures, the first two columns indicate the fixed outage probability for MIP and MDP traffic, whereas the last column indicates MIP%–MDP% traffic percentages. For higher MIP traffic, more users belonging to this class can enter the system, more frequently preempting MDP flows and degrading MDP reservations. In addition, for high MIP traffic percentages and increasing outage threshold, a decrease in allocated bandwidth can be observed. In this situation, there are more MIP users sharing the same cell capacities, so they must perceive a lower amount of bandwidth.

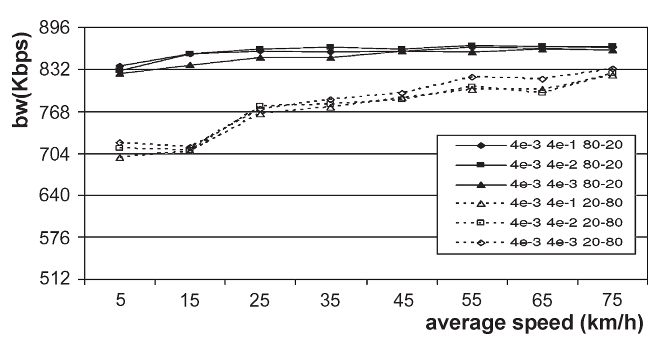


Fig. 9. Average allocated bandwidth for MIP users.

In addition, when the MIP traffic percentage is low (around 20%), the effects of MDP user mobility reflect on the MIP bandwidth, so it increases from 700 to 830 kb/s if the average speed is increased.

Fig. 10 shows the average system utilization when a MIP–MDP traffic percentage of 20%–80% is set. The system is lower utilized by increasing the hosts' speed. For high-speed values, there are more link variations, and consequently, the system must handle a larger number of bandwidth reallocations. This causes utilization wastage, which can reach a magnitude

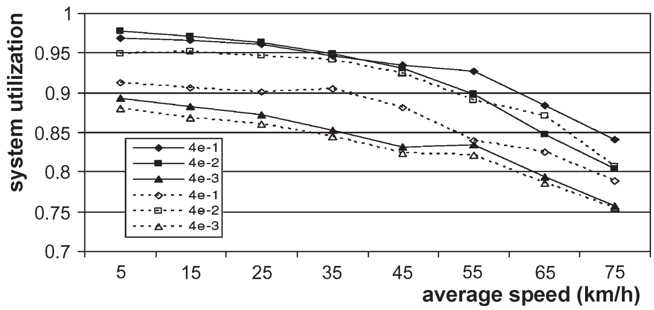


Fig. 10. Average system utilization.

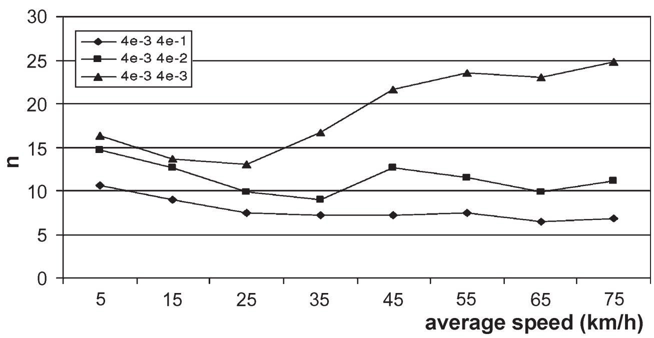


Fig. 11. Average MIP admitted flows.

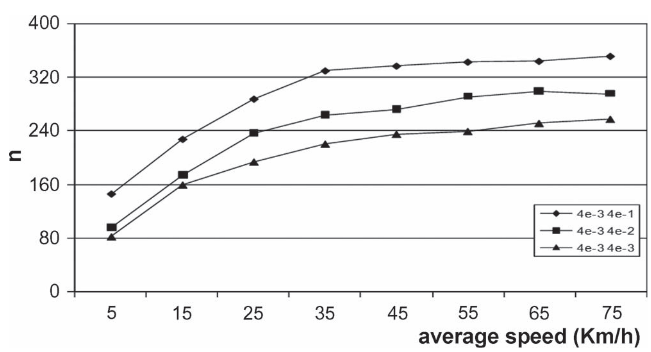


Fig. 12. Average MDP admitted flows.

of 15%–20%. Varying the outage threshold, there are different observed values of resource utilization for the same reason earlier discussed, that is to say, the admission control is less selective for higher threshold values, so more users can enter the network and a higher utilization can be reached. The bandwidth wastage is not evident because of the presence of MDP flows in the system. When the traffic is only composed by MIP flows, the system is underutilized because of the unused passive pre-reserved bandwidth. The difference between continuous curves and dashed ones is the chance for MDP to reutilize (continuous) or not (dashed) the passive bandwidth. MIP flows, after a hand-off, can obviously preempt MDP users. It is evident that there is a gain in the system with the multiplexing of passive bandwidth, particularly for high values of p_{outage}; higher threshold values allow higher MDP in the system, which can reuse the available passive bandwidth.

The minimum and maximum values of admitted flows can be observed in Figs. 11 and 12 (curves are obtained for the 20%–80% traffic percentages). For both cases, they depend on the chosen $p_{\mathrm{threshold}}$ value for MDP traffic (in those figures, the

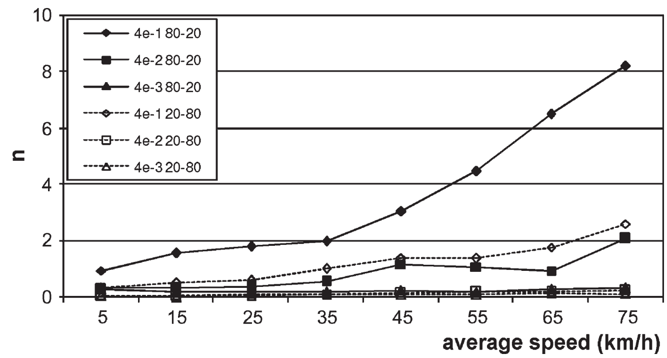


Fig. 13. Average MIP dropped flows.

TABLE IV
STATISTICAL PARAMETERS FOR $\alpha = 0$ AND $\alpha = 5\%$

	α=0 (constant speed)		α=5% (var. speed)	
Speed (km/h)	μ	σ	μ	σ
5	323.6451	1.046545	354.2467	8.9789
15	105.07847	0.0834	119.3434	0.1954
25	60.5112	0.03567	66.4554	0.1688
35	40.83243	0.03148	52.4518	0.0597
45	30.23874	0.02545	40.4576	0.05018
55	26.2372	0.02045	30.3212	0.04045
65	22.323445	0.01344	25.3545	0.0354
75	19.37627	0.00884	21.6966	0.0256

threshold for MIP users has been fixed to $4 * 10^{-3}$ to guarantee a good level of outage avoidance). As illustrated in Fig. 1(a), MDP users make requests only on current cells, whereas MIP users make reservations over all system cells, so they are subject to a more strictly admission policy, and the probability of a system entering is lower than MDP. For increasing speed, some observations must be made. For the MDP case, as the average speed increases, the cell stay time of each user decreases, so more users can find bandwidth availability. For the MIP case, for middle–high ($4 * 10^{-2}$, $4 * 10^{-1}$), the effect of MDP does not impact the MIP admission numbers. Obviously, if the MDP threshold decreases, a higher number of MIP users can be admitted because of the higher bandwidth availability.

From Fig. 13, it can be observed that high values of outage threshold (like 0.4) cannot be suffered by MIP flows because there are too many dropped flows (eight for an average speed of 75 km/h), whereas for low values (like 0.004), the phenomenon can be disregarded. The second column represents the MIP–MDP traffic percentages, whereas the first column represents the chosen $p_{\mathrm{threshold}}$ values. To evaluate the improvements introduced by the proposed prediction policy, a campaign of "monitor simulations" (as previously explained, 1000 runs have been led out) has been executed to obtain statistical prediction model parameters [see (13) and (14)] by varying the average hosts' speed. The obtained results are shown in Table IV under 2-D RWPMM for $R = 250$ m and $s = 10\%$.

Fig. 14 shows the obtained results for the average system utilization by MIP users with different values of MDP outage probability: $4 * 10^{-1}$, $4 * 10^{-2}$, $4 * 10^{-3}$ (defined in Section III), α and s fixed to 10%, $R = 250$ m, and a MIP–MDP traffic percentage of 80%–20%.

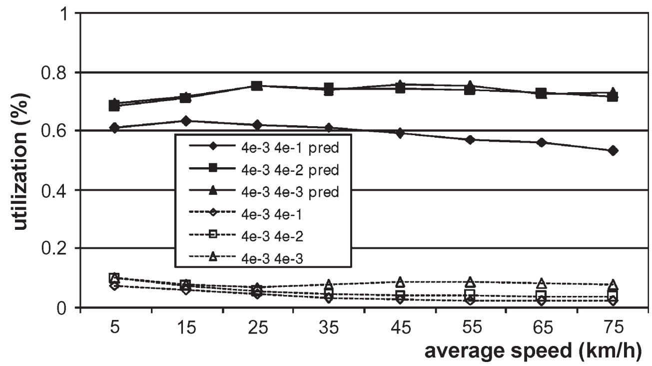


Fig. 14. Average system utilization by MIP users.

From the preceding tables, it can be observed that the standard deviation σ (and, so, the variance σ^2) increases for higher values of α. This reflects bigger variations in the considered speed interval that becomes wider for high values of α.

It is evident that in both cases of full reservation (dotted lines with passive reservations over all network cells) and predictive reservation (continuous lines with passive reservation over a cells circle of radius C_e), the performances are not influenced by the mobile host speed (the slight decreasing course is due to the higher presence of MDP-admitted flows for higher speeds). An improvement near 60% can be reached by choosing an appropriate set of passive cells.

VII. Conclusion

In this paper, we have analyzed the performances of the proposed utility-oriented algorithm, and we have also seen that there is the need to consider the physical link variability to give more effectiveness to bandwidth allocation schemes and to take into account the absence of ideality during radio connections. From the simulation results, we have observed that the channel degradation becomes more evident for high-speed values and the handoff number increases, but the prereservation of MIP classes guarantees a full compliance to QoS parameters (outage probability and maximized user utility function) for low values of outage threshold, so the mobility effects are minimized. A new approach for network improvements is proposed and integrated with the MRSVP. In particular, by simulation results, it has been shown that a circular partial prereservation policy can ensure either wireless QoS during handoff events or higher system utilization than a full prereservation policy. We have also shown that our model presents a prediction error (which is more evident for low speed values and high values of α) that causes negligible effects on QoS guarantees that MIP users require for their connections. In addition, an analytical study of the CST in WLAN networks has been considered, where the CST has been evaluated as a function of the average speed of the mobile user and its variation around the average speed. A dependence of CST on the cell radius and overlapping coverage has also been analyzed, and a regression analysis has been performed to calculate an analytical expression for the CST that has been used in the calculation of the number of future visited cells. Simulation results show the expression of CST distribution parameters as a function of percentage variation α around the average speed, the cell radius R, and the overlapping factor s, permitting the prediction of the right number of visited cells with a percentage error under 7.2%.

References

[1] V. Bharghavan, K.-W. Lee, S. Lu, S. Ha, J.-R. Li, and D. Dwyer, "The TIMELY adaptive resource management architecture," *IEEE Pers. Commun.*, vol. 5, no. 4, pp. 20–31, Aug. 1998.

[2] A. Alwan, R. Bagrodia, N. Bambos, M. Gerla, L. Kleinrock, J. Short, and J. Villasenor, "Adaptive mobile multimedia networks," *IEEE Pers. Commun.*, vol. 3, no. 2, pp. 34–51, Apr. 1996.

[3] H. S. Wang and N. Moayeri, "Finite-state Markov channel—A useful model for radio communication channels," *IEEE Trans. Veh. Technol.*, vol. 44, no. 1, pp. 163–171, Feb. 1995.

[4] A. K. Talukdar, B. R. Badrinath, and A. Acharya, "MRSVP: A resource reservation protocol for an integrated services network with mobile hosts," *Wirel. Netw.*, vol. 7, no. 1, pp. 5–19, 2001.

[5] A. K. Talukdar, B. R. Badrinath, and A. Acharya, "On accommodating mobile hosts in an integrated services packet network," in *Proc. IEEE INFOCOM*, Apr. 2007, pp. 1046–1053.

[6] Y. Cao and V. K. Li, "Utility-oriented adaptive QoS and bandwidth allocation in wireless networks," in *Proc. IEEE ICC*, 2002, pp. 3071–3075.

[7] R.-F. Liao and A. T. Campbell, "A utility-based approach for quantitative adaptation in wireless packet networks," *Wirel. Netw.*, vol. 7, no. 5, pp. 541–557, Sep. 2001.

[8] J. Gomez and A. T. Campbell, "Havana: Supporting application and channel dependent QOS in wireless packet networks," *Wirel. Netw.*, vol. 9, no. 1, pp. 21–35, Jan. 2003.

[9] F. Baker, B. Braden, S. Bradner, M. O'Dell, A. Romanow, A. Weinrib, and L. Zhang, *Resource ReSerVation Protocol (RSVP)—Version 1 Applicability Statement Some Guidelines on Deployment*, Sep. 1997. RFC2208.

[10] R. Braden and L. Zhang, *Resource ReSerVation Protocol (RSVP)—Version 1 Message Processing Rule*, Sep. 1997. RFC2209.

[11] F. De Rango, G. Aloi, and S. Marano, "A fair rate adaptation algorithm based on the degradation factor for integrated wireless networks with mobile nodes," in *Proc. 13th IEEE LANMAN*, Mill Valley, CA, Apr. 2004, pp. 133–138.

[12] F. De Rango, G. Aloi, and S. Marano, "An efficient rate adaptation scheme in wireless mobile networks," in *Proc. WTS*, Pomona, CA, May 2004, pp. 88–93.

[13] A. K. Talukdar, B. R. Badrinath, and A. Acharya, "Rate adaptation schemes in networks with mobile hosts," in *Proc. ACM/IEEE MOBICOM*, 1998, pp. 169–180.

[14] G.-S. Kuo and P.-C. Ko, "Dynamic RSVP protocol," *IEEE Commun. Mag.*, vol. 41, no. 5, pp. 130–135, May 2003.

[15] M. Mirhakkak, N. Schult, and D. Thomson, "Dynamic bandwidth management and adaptive applications for a variable bandwidth wireless environment," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 10, pp. 1984–1997, Oct. 2001.

[16] X. Fu *et al.*, "NSIS: A new extensible IP signaling protocol suite," *IEEE Commun. Mag.*, vol. 43, no. 10, pp. 133–141, Oct. 2005.

[17] T. Zhang *et al.*, "Local predictive resource reservation for handoff in multimedia wireless IP networks," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 10, pp. 1931–1941, Oct. 2001.

[18] T. Liu, P. Bahl, and I. Chalmtac, "Mobility modeling, location tracking, and trajectory prediction in wireless ATM networks," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 6, pp. 922–936, Aug. 1998.

[19] C. Bettstetter, H. Hartenstein, and X. Pérez-Costa, "Stochastic properties of the random waypoint mobility model," *Wirel. Netw.*, vol. 10, no. 5, pp. 555–567, Sep. 2004.

[20] I. F. Akyildiz and W. Wang, "The predictive user mobility profile framework for wireless multimedia networks," *IEEE/ACM Trans. Netw.*, vol. 12, no. 6, pp. 1021–1035, Dec. 2004.

[21] I. F. Akyildiz and W. Wang, "A dynamic location management scheme for next-generation multitier PCS systems," *IEEE Trans. Wireless Commun.*, vol. 1, no. 1, pp. 178–189, Jan. 2002.

[22] C. Betstetter, "Mobility modeling in wireless networks: Categorization, smooth movement, and border effects," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 5, no. 3, pp. 55–66, Jul. 2001.

[23] Y. Fang, I. Chlamtac, and Y.-B. Lin, "Modeling PCS networks under general call holding time and cell residence time distributions," *IEEE/ACM Trans. Netw.*, vol. 5, no. 6, pp. 893–906, Dec. 1997.

[24] L.-L. Lu, J.-L. C. Wu, and W.-Y. Chen, "The study of handoff prediction schemes for resource reservation in mobile multimedia wireless networks," *Int. J. Commun. Syst.*, vol. 17, no. 6, pp. 535–552, Aug. 2004.

[25] F. De Rango, P. Fazio, and S. Marano, "Mobility independent and dependent predictive services management in wireless/mobile multimedia network," in *Proc. IEEE VTC—Fall*, Los Angeles, CA, Sep. 2004, pp. 2596–2600.
[26] F. De Rango, P. Fazio, and S. Marano, "Cell stay time analysis under random way point mobility model in WLAN," *IEEE Commun. Lett.*, vol. 10, no. 11, pp. 763–765, Nov. 2006.
[27] F. De Rango, P. Fazio, and S. Marano, "Cell stay time prediction for mobility independent predictive services in wireless networks," in *Proc. IEEE WCNC*, New Orleans, LA, Mar. 13–17, 2005, pp. 1792–1797.
[28] M. Hassan, M. M. Krunz, and I. Matta, "Markov-based channel characterization for tractable performance analysis in wireless packet networks," *IEEE Trans. Wireless Commun.*, vol. 3, no. 3, pp. 821–831, May 2006.
[29] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Higher-Speed Layer Extension in the 2.4 GHz Band,* IEEE Std. 802.11b, 1999.
[30] D. Mitra, "Stochastic theory of a fluid model of producers and consumers coupled by a buffer," *Adv. Appl. Prob.*, vol. 20, pp. 646–676, 1988.
[31] C. Montgomery, *Applied Statistics and Probability for Engineers*, 3rd ed. Hoboken, NJ: Wiley, 2003.
[32] J. Banks, J. S. Carson *et al.*, *Discrete-Event System Simulation*, 3rd ed. Englewood Cliffs, NJ: Prentice–Hall, 2001.
[33] M. A. Stevens and R. B. D'Agostino, *Goodness of Fit Techniques*. New York: Marcel Dekker, 1986.
[34] W. Navidi and T. Camp, "Stationary distributions for the random waypoint mobility model," *IEEE Trans. Mobile Comput.*, vol. 3, no. 1, pp. 99–108, Jan./Feb. 2004.

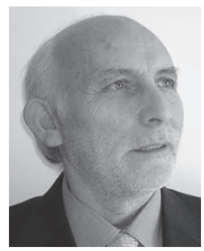
Floriano De Rango (M'06) received the Laurea degree in computer science engineering and the Ph.D. degree in electronics and communications engineering from the University of Calabria, Arcavacata di Rende, Italy, in October 2000 and January 2005, respectively

From January 2000 to October 2000, he was a visiting scholar student with the Telecom Research LAB C.S.E.L.T., Turin, Italy. From March 2004 to November 2004, he was a Visiting Researcher with the University of California at Los Angeles (UCLA). Since November 2004, he has been with the D.E.I.S. Department, University of Calabria, where he was a Research Fellow until September 2007 and is currently an Assistant Professor. He has coauthored more than 100 papers in international journal and conferences proceedings. His interests include satellite networks, IP QoS architectures, adaptive wireless networks, ad hoc networks, and pervasive computing.

Dr. De Rango was the recipient of the Young Researcher Award in 2007. He has served as Reviewer and TPC member for many international conferences such as IEEE VTC, ICC, WCNC, Globecom, Med Hoc Net, SPECTS, WirelessCOM, and WinSys, and has been a Reviewer for many journals such as the IEEE COMMUNICATION LETTERS, the IEEE JOURNAL OF SELECTED AREAS IN COMMUNICATIONS, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, *Computer Communication, Eurasip JWCN, WINET*, etc.

Peppino Fazio received the Laurea degree in computer science engineering and the Ph.D. degree in electronics and communications engineering from the University of Calabria, Arcavacata di Rende, Italy, in May 2004 and January 2008, respectively.

He is currently a Research Fellow with the D.E.I.S. Department, University of Calabria, and is also collaborating with the DISCA Department, "Universidad Politecnica de Valencia," Valencia, Spain. His research interests include mobile communication networks, QoS architectures and interworking wireless and wired networks, mobility modeling for WLAN environments, and mobility analysis for prediction purposes.



Salvatore Marano (M'97) received the Laurea degree in electronics engineering from the University of Rome, Rome, Italy, in 1973.

In 1974, he was with Fondazione Ugo Bordoni. Between 1976 and 1977, he was with the ITT Laboratory, Leeds, U.K. Since 1979, he has been an Associate Professor with the University of Calabria, Arcavacata di Rende, Italy. His research interests include performance evaluation in mobile communication systems, satellite systems, and 3g/4G networks. He has published more than 130 papers in international conference proceedings and journals.

Dr. Marano has been a Reviewer for many journals such as the IEEE COMMUNICATION LETTERS, the IEEE JOURNAL OF SELECTED AREAS IN COMMUNICATIONS, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, the IEEE TRANSACTION ON WIRELESS COMMUNICATIONS, and the *European Transactions on Telecommunications Journal.*