Introduction

This **Philosophiæ Doctoral** (PhD) thesis faces some problems about the shared resources allocation in wireless environments, through a new algorithm that takes into account both users mobility and wireless channel degradation level; in addition, a prediction reservation policy for passive resources is also presented, in order to guarantee a certain Quality of Service (QoS) level to mobile services. An analytical analysis of the wireless channel model has also been carried out.

Wireless technologies represent a rapidly emerging area of growth and importance for providing ubiquitous access to the network for all of the community. Students, faculty and staff increasingly want un-tethered network access from general-purpose classrooms, meeting rooms, auditoriums, and even the hallways of campus buildings. There is interest in creating mobile computing labs utilizing laptop computers equipped with wireless cards. Recently, industry has made significant progress in resolving some constraints to the widespread adoption of wireless technologies. Some of the constraints have included disparate standards, low bandwidth, and high infrastructure and service cost. Wireless is being adopted for many new applications: to connect computers, to allow remote monitoring and data acquisition, to provide access control and security, and to provide a solution for environments where wires may not be the best solution. The most important peculiarity of wireless communications is represented by their capability of ensure the requested services to moving hosts in the coverage domain.

A telecommunications network can be considered as a structure that is able to establish, after a user service request, a connection for the communication with another user or a multicast group. The main functionalities of a network are: the choice of the appropriate level of bandwidth that must be assigned to the requesting user, the transmission over the channel the effective desired information, the administration of the resources in order to allow the correctness of the transmissions. Supplying an appropriate service by a network needs the cooperation of different entities of different network nodes, connected through physical communication means. In every kind of telecommunication, each entity coordinates itself with the other one by the *communication protocols*, generally based on the *layer* concept: the functionalities set is subdivided into a stack of overlapped layers, where each of them offers a subset of the total needed functionalities, in order to allow the communication with other systems. Among the all reference models, the ISO-OSI (International Standard Organization – Open System Interconnection) is the most famous, based on a 7-layers stack.

There are two different categories of LAN (Local Area Network): the wired one (wired-LAN) and the wireless one (wireless-LAN or WLAN) that uses radio waves as transmitting media. WLANs minimize the need of wired connections, combining the high grade of connectivity with the mobility of users, which can access to shared resources without the need of finding a physical point where the device must be connected; in addition, network administrators can expand system dimensions without install or moving cables. Other advantages of WLANs are the quick installation, high scalability and very bounded costs.

As all the telecommunication systems, there are also some disadvantages as the employment of radio waves for information transmissions (that leads to the presence of a high amount of electromagnetic phenomena, intrinsic in waves propagation), the physical obstacles (that are not perfectly penetrable by radio waves), the limited bandwidth availability and the frequency reuse.

For the Internet real-time traffic the ReSerVation Protocol (RSVP) is used as a resource reservation protocol; the main aim of the RSVP is the exchange, among the system nodes, of the *reservation state*, i.e. the reservation state of the resource request of the single connection. Users' has a heavy impact on the QoS parameters of the real-time applications. The existing architectures for the management of real-time services on a network with fixed hosts (no users' mobility) are not adequate for mobility support, in fact some additional functionalities are needed to the RSVP, in order to face such kind of problems. For example, the effects of users nobility can be reduced by "pre-reserving" the resources on the locations (cells) that a mobile host will visit during its active connection (as it will be explained, this is called "*passive reservation*" policy). When a mobile host moves among different coverage areas with an active session, the packet delivery delay may vary (the congestion level on a new path with a new set of involved routers can be different if compared with the previous path); if the new coverage node, where the mobile host has moved, is overloaded, the available bandwidth of the new location may be scarce or not able to satisfy the service request: in this case the mobile host must adapt its QoS request. In addition, during a hand-off event, the mobile host may have some temporary service degradations (in some cases the call may be also dropped). In order to avoid these problems, a certain admission control scheme must be executed in the routers and in the Access Points (APs, like the coverage nodes) that belong to the path from the source to the locations that the mobile host may visit during the whole connection duration. A flow is admitted into the network only if all the admission controls reply with a positive answer.

All the operations and funcionalities above are allowed if the Mobile RSVP (MRSVP) protocol is employed, that has been formulated as an extension of the RSVP for the hosts mobility support expecially in the (ISPNs – Integrated Services Packet Networks): two kinds of reservations are used, the *active* reservation (between the sender and the current host coverage cell) and the *passive* one (between the sender and all the future cells specified in the user mobility profile). In this way, the right amount of bandwidth can be "in-advance" reserved needed by the flow in order to avoid service degradations during hand-off events.

Another extension of the RSVP is the Dynamic RSVP (DRSVP), able to offer dynamic QoS in a network with variable bandwidth, where some wireless links are present (their BER performances vary during an active connection) with mobile intermediate nodes (network topology varies, modifying the quality of a generic path between two nodes); in this scenario, the DRSVP limits the assigned amount of bandwidth, by executing an acccurate analysis of the network *bottlenecks*: when a request is admitted into the system, reserving high amounts of bandwidth in some nodes has none-sense if it cannot be used because of the presence of some bottlenecks; if the "over-reservation" is avoided, then there will be a resource availability gain.

As earlier introduced, the advantages of the wireless networks (portability, high connectivity and low costs) are accompanied by some undesired propagation effects: in the WLAN environments the *path-loss* is the outstanding effect, expecially in the free-space communications, where the Line Of Sight (LOS) communication is affected by the distance between transmitter and receiver; in the urban or rural environments the presence of obstacles and buildings smooths the transmitted signal while creating the such called "shadow zones", that block the signal propagation. There are different propagation effects more evident in urban environments (evanescence, trasmission, refraction and shadowing), but the most important is the *reflection* of the electromagnetic beam, that generates additional paths over which the reflected signal propagates itself; this is called *multipath-fading*. Under the multipath-fading effects, the electromagnetic beam is affected by a delayed diffusion, that is to say the reflected signal components arrive to destination at different time instants, creating a shattered version of the original transmitted signal. While in the satellite or micro-wave point-to-point communications the multipath effects are negligible, in the mobile telephony environments many obstacles and object are present, so the destination station will receive many copies of the original signal, which arrives with different delays.

The *fading* term refers to the temporal variation of the received signal power, caused by changes in the transmission mean or in the different paths. In a static environment (low or null grade of node mobility) the fading is supplied by the variations in atmospheric conditions (an example can be the presence of rain); in a mobile scenario, where the receiving/transmitting antenna position changes during an active session, complex transmission phenomena can be observed: if the same digital signal is transmitted twice at two different and well separated time instants, the receiver will observe two different signals, although the wireless media is the same for the two transmissions; it is due to the continuous aleatory evolutions of the wireless channel physical characteristics (the wireless channel is said to have an aleatory and time-variant impulse response). This is verified with the transmission of a short-duration signal (ideally an impulse) through a wireless multipath channel: the received signal will be an impulses train, with different shapes.

So, the only way to describe a wireless link behaviour is represented by a statistical model, which can take into account the time-variant nature of the link; our research has been based on the Markov stochastic process, able to describe, with a certain grade of precision, the trend of the wireless channel during time, in terms of transmission errors and degradation). The most popular fading channels are: the *Rayleigh* one (heavily used when there are many indirect paths from the transmitter to the receiver, without a dominant component) and the *Rice* one (when a direct and dominant path is present in the communication). The Rice and Rayleigh models are often used in the outdoor and indoor environments respectively. In order to relate the stochastic process with the channel model the Signal-to-Noise Ratio (SNR) must be partitioned in a finite set of intervals, associating each one of them to the states of the Markov Chain (in this way the physical characteristics and the time-variant behaviour of the channel are taken into account). In particular, in this thesis the channel model study has been focused on the IEEE802.11 standards, under a *slow fading* condition (the wireless link evolution can be considered "slow" if compared with the packet transmission time).

So, from the discussion above, it is clear that in the wireless communications there is not the "ideality" concept when considering the link between transmitter and receiver and the wastage of bandwidth due to the inevitable presence of positive values of Bit Error Rate (BER) must be taken into account: as in the ISPNs, when a host specifies a minimum level of QoS in its service request, the network must ensure the required level and, if a wireless link is involved, a major amount of bandwidth must be assigned, so the received *effective bandwidth* will respect the desired minimum QoS threshold.

QoS guarantees are the main aims of the modern telecommunication systems: adequate Call Admission Control (CAC) and Rate-Adaptation schemes must be employed, because they ensures to the intermediate and/or coverage nodes that the overload effect probability will be null or below a fixed threshold, so the minimum QoS levels will be always respected. A CAC policy is generally designed for the management of the maximum number of users that can be admitted into the system but, in the same time, it must ensure high system utilization with low congestion levels. A new service request is admitted into the system only if it passes the CAC over the the whole set of involved cells.

The rate-adaptation policies dynamically change the delivered flow-rates, respecting the requested bounds (like the *packet-delay* and *delay-jitter*) and some principles, as the *fairness*, *minimum overhead* and *high system utilization*. In this work a cellular coverage area has been considered, where the geographic area is subdivided in a certain number of regions, called *cells*. Each cell is covered by an AP that must ensure to mobile hosts a global wireless access to the system. In these environments, bandwidth reallocations are necessary when a *hand-in* (a new flow is admitted in the current cell or it arrives from the previous coverage area) or *hand-out* (a flow departs from the current cell) event occurs; in the first case, if the cell is high utilized, a degradation operation is necessary, in order to reduce the bandwidth level assigned to the current users; in the second case, the released bandwidth is redistributed with an upgrade operation. In the considered applications each flow can work with different bandwidth levels and the upgrade/degrade operations consist on increasing or decreasing the assigned rate with constant step. Different algorithms have been proposed in the literature in order to face these problems: some of them are based on the QoS parameters statistics, like the Degradation Ratio (DR, the ratio between the period of time when the host receives a degraded service and the cell residence time) and the Upgrade/Degrade Frequency (UDF, the frequency of switch between full service and degraded service of an admitted call), derived from an accurate analysis of the system evolution through Markov chains; other algorithms try to maximize the system utilization by defining and fixing some *Thresholds*, offering different admission grades, based on the current congestion level and on the arriving calls priorities. Other types of reallocation algorithms are based on the dynamics of some parameters (utilization, number of admitted flows, etc.), taking into account the evolution of the system and the traffic behaviour: in this way the overestimation of bandwidth requests is avoided, as well as the low system utilization. In order to make dynamic evaluations of the system status the *time windows* are used (that is to say certain time periods when the system parameters are sampled and observed). In this work, a "utility-oriented" algorithm has been used in order to manage the different service requests, taking into account both wireless channel conditions and users satisfaction level (through the employment of utility functions): the employed scheme is called "utility-oriented" because it takes into account the level of satisfaction obtained for the received bandwidth; utility functions are able to describe the QoS users requests and how a user can be satisfied by the obtained service. The concept of utility function has been introduced, as an indicator of the user satisfaction level. If the perceived utility must be maximized, then there must be a way to describe how a user is satisfied when the received bandwidth level varies. Utility functions are useful to solve such kind of problems and it has been shown that there are many works in literature that describe the best trends of utility functions, appropriate for the specific application (tolerant, intolerant, etc.).

After a description of some important concepts of utility functions, a new bandwidth allocation protocol has been introduced, with the aim of having a new scheme that can be applicable in the ISPNs systems. Since different applications can be introduced in an ISPN system, the proposed idea takes care of considering the specific utility function, as well as the wireless channel modelling. In this way some important goals can be reached: fairness among users belonging to the same class; high system utilization and QoS guarantees. A complexity analysis has been given for the UB CAC and BAG schemes and the importance of a dynamic bandwidth allocation scheme has been demonstrated, through an addicted campaign of simulations. The obtained results have shown that the introduction of a dynamic scheme for bandwidth management increases system performance, in terms of utilization and number of admitted flows. In addition, it has been shown that the introduction of a channel model is mandatory if channel degradations must be taken into account when dimensioning a wireless system or while serving MDP requests.

In addition, the algorithm must ensure the following criteria:

a) *QoS requirements*: the *utility-outage* (the received utility falls below a lower bound) event must be managed in an adequate manner, ensuring that the outage probability does not exceed a fixed threshold;

b) *Fairness*: for each user in the system an index must be defined (for example, referring it to the received utility) and, in the long term, the algorithm must ensure that there will not be high difference between different indexes of different users;

c) *High system utilization*: the offered utility is the criterion to measure the average utilization of the available bandwidth.

After the definition of the general behaviour of a rate-adaptation algorithm and the relative CAC module, they must be integrated with the ISPNs and their predictive service classes; in particular, in this work the Mobility Independent Predictive (MIP - users that request this kind of services want to avoid mobility effects) and Mobility Dependent Predictive (MDP - users that request this kind of services may suffer the effects of mobility, like degradations or connection droppings). The MRSVP guarantees service continuity a MIP user by making passive reservations over the APs that will serve the user during its connection; this kind of management may cause a resource wastage, because of the amount of passive and unused bandwidth that will switch into active resource only when a MIP user makes a hand-in; the obtained low system utilization is more evident when the MDP flows cannot have the access to the passive bandwidth.

The undesired resource wastage due to the passive reservations of MIP users cannot be avoided but it can be efficiently reduced if the following consideration is taken into account: the passive reservations may not be made over all the cells in the system, because some of them will never be reached by a mobile host with a "relative slow" mobility behaviour; if the reservation scheme takes into account the average *Cell Stay Time* (CST – the time spent in a cell by a user) then the number of possible (and more probable) future visited cells can be obtained, having the knowledge of the average *Call Holding Time* (CHT – the duration of a connection); in a 1-dimensional (1D) environment the CST analysis is able to full characterize the passive prediction policy. In a 2D environment, considering only the CST will not give all the prediction information that are necessary: this time the number of future visited cells must be accompanied by the identification of every single cell where, with high probability, the user will move. It is evident that an additional statistical treatment is necessary in order to consider the directional behaviour of mobile hosts, under a certain mobility model; in the proposed thesis the Random WayPoint Mobility Model (RWPMM) has been considered, as well as the Smooth Random Mobility Model (SRMM); the latter makes hosts movements smoother and more realistic than the various proposed mobility models in the literature. In particular, directional information must be added to the knowledge of the CST in order to make possible the passive reservation prediction in a 2D environment.

Many simulation campaigns have been carried out: first of all the performances of the proposed CAC and rate-adaptation algorithm have been analysed; different "monitor" campaigns have been carried out in order to evaluate the directional behaviour of users in the 2D system as well as the CST distribution, then the passive reservation policy has been introduced; the enhancements have been appreciated, through the analysis of the obtained curves. The passive reservation is able to guarantee service continuity to MIP users, but it makes the system low utilized because of the large amount of passive and unused bandwidth. In order to avoid the disadvantage, two ideas have been applied:

a) make MDP users able to reuse the passive bandwidth, with a certain dropping probability if the passive reservation must be switched into an active one, after a MIP hand-in event;

b) make MIP users able to multiplex their passive reservations: the CST knowledge is very important in order to obtain an estimation of the hand-in event for each user and for each probably visited cell; in this way it is possible to multiplex the pre-reserved bandwidth until the user makes its own hand-in in the considered cell.

The importance of mobility prediction for wireless systems (such as WLANs) has been outlined, introducing some schemes for passive reservations enhancements. Two mobility models have been considered, RWPMM and SRMM, but the proposed schemes are completely uncorrelated with the employed mobility model. The introduction of the proposed schemes has started with the analysis of a simplified 1D scenario, in order to give some knowledge about the CST evaluation and analysis, in terms of statistical distribution. After the quantitative analysis of CST distribution, an extension has been introduced, by considering the complete 2D space and the directional behaviour of mobile hosts. So after a circular reservation, a directional one has been proposed. The obtained CST distributions have been shown, in terms of mean and standard deviation parameters. In addition, a polynomial regression has been shown to be a good way to obtain a CST evaluation by simply introducing system and mobility parameters. The static scheme and the dynamic one have been proposed and formally described by pseudo-code. Obviously, the performance of the proposed schemes must be evaluated; this will be made in the next chapter.

In this PhD thesis the enhancements of the introduction of the proposed reservation scheme and CAC are shown, with their performances analysis. A gain in terms of QoS and system utilization has been obtained by the integration of the rate-adaptation scheme with a passive-reservation algorithm. If the passive bandwidth multiplexing is also activated, the system reaches a utilization level that is near to the saturation state.

The PhD thesis is concluded with the performance evaluation of the proposed ideas. Initially the 1D CST-based prediction scheme has been investigated and good results have been obtained regarding the MIP QoS guarantees: system utilization can considerably increase if passive reservations are made in an adequate manner and service continuity is always guaranteed for MIP service requests. However the simulated 1D scenario is too simply if compared with real mobile environments, so it has been extended with a 2D clustered simulation scenario, where users can move according to the Random WayPoint or Smooth Random Mobility Models. The same CST-based prediction scheme of the 1D case has been employed in the 2D scenario but, although the prediction error is negligible, too many resources are wasted, because of the high number of C_r cells which are interested by passive reservations; so additional and directional information has been introduced in the prediction algorithms in order to make them more selective. Optimal results have been obtained for some combinations of input parameters.