

Future mobile broadband wireless networks: a radio resource management perspective

Shalini Periyalwar,^{1,*†} Bassam Hashem,² Gamini Senarath,¹ Kelvin Au¹ and Robert Matyas¹

¹*Wireless Technology Labs, Nortel Networks, Ottawa, Canada*

²*Adjunct Research Professor, Carleton University, Ottawa, Canada*

Summary

Future wireless evolution envisages high rates, low hierarchy in the network architecture, antenna array processing, multiple access modes and multihop operation as part of the system concept. To exploit the increased capabilities of the systems in conception, efficient resource management strategies need to be developed. The goal of this paper is to examine the key aspects of the evolution which impact radio resource management for the mobile broadband wireless network, and to emphasize the areas that need to be addressed for servicing mobile users with varying quality of service requirements. Copyright © 2003 John Wiley & Sons, Ltd.

KEY WORDS: radio resource management; 4G wireless networks; 3G evolution MAC states; coverage; quality of service; admission control; congestion control; scheduler; coverage; performance evaluation

1. Introduction

Future generation wireless networks (e.g. 4G) are in the process of being defined with plans for extensive studies in various organizations [1,2]. Future wireless networks will be deployed in an environment where wired and wireless infrastructure is already established; thus, it will not replace the current generation and its evolution but will complement and enhance these systems. Any new spectrum allocation that may be identified in the coming years must take into consideration the intended objective of very high data rates over moderate cell sizes. Future wireless evolution envisages low hierarchy in the network architecture with access points (some user deployable) and terminals supporting multiple access modes. Such a system may be conceived as an integration of a number of wireless air interfaces, each optimized to its

environment. Furthermore, communication involving multiple access points and/or terminals in the delivery of the data to the user is seen as a promising means to ensure ubiquity of services, especially at the cell edge.

The vision articulated in international forums [1,2] for future mobile broadband wireless networks entails significant new innovation in the air interface along with the need to support new services with stringent quality of service (QoS) requirements. Future generation air interfaces are expected to offer higher performance, by an order of magnitude or greater improvement in data rates and by more efficient QoS provisioning, over current third-generation evolution proposals. Advanced antenna processing is an integral part of the vision, along with multihop relaying, which will see an increasing role in future systems. The physical layer and medium access control (MAC) layer will offer a range of features, which

*Correspondence to: Shalini Periyalwar, Nortel Networks, Wireless Technology Labs, MS 04391Y30, 3500 Carling Ave, Ottawa, K2H 8E9, Canada.

†E-mail: speriya@nortelnetworks.com

may be mixed and matched to address the QoS needs of services. These and other advanced characteristics of future wireless systems are covered in Section 2 of the paper. To exploit the significantly increased capabilities, efficient resource management strategies are needed. Efficient scheduling, seamless handoff for all types of services, interference avoidance and congestion management schemes are some of the techniques which will need to be significantly enhanced for servicing mobile users with varying QoS requirements in a highly adaptive system framework. These key functions, which are essential to ensure the optimal operation of future mobile broadband wireless networks, are addressed in Section 3 of the paper. A conservative approach to performance evaluation of these new systems is discussed in Section 4, which is followed by the conclusions.

2. Characteristics of Future Wireless Systems

Future wireless systems will be expected to have the following key attributes to enable a high performance seamless user experience equivalent to a wireline:

- Service coverage ubiquity: the expectation is that a user will be able to run any required services anywhere, any time without being hindered by the limitations of the wireless system.
- Optimized connectivity: the user is always connected to the most efficient access network (e.g. macrocellular, wireless LAN) in terms of network resource usage, to cater to the specific QoS and mobility requirements.
- Always on: the user will be connected to the network as long as the terminal power is on and experience minimal access delay.

All of the above will be available to a much larger population of users than current 3G evolution (3G-Ev) rate-controlled systems [3,4] can support, due to significantly higher spectral efficiency achieved with the use of novel air interface techniques, and by employing innovative coverage enhancement techniques. Important characteristics of a system to realize significantly higher data rates and service coverage ubiquity are discussed in the paragraphs below.

2.1. Wireless Access Networks Evolution Model

2.1.1. Integrated wireless access networks

The future integrated wireless access network may comprise of different wireless air interfaces (WLAN, 3G cellular, 4G cellular, peer-to-peer (P2P), multihop relays etc.) with a range of cell sizes supported within an integrated wireless access umbrella, as illustrated in Figure 1. Such a trend is already in evidence today with 3G systems beginning to work seamlessly with wireless LANs to provide ubiquitous access. The concept of 'optimized connectivity' is a main feature of such an architecture. The fundamental premise for assigning a user to any one of these access networks would be to maximize the user experience, while ensuring the most economical use of the radio resources for the operator. The split of users between the different access networks will be driven by system

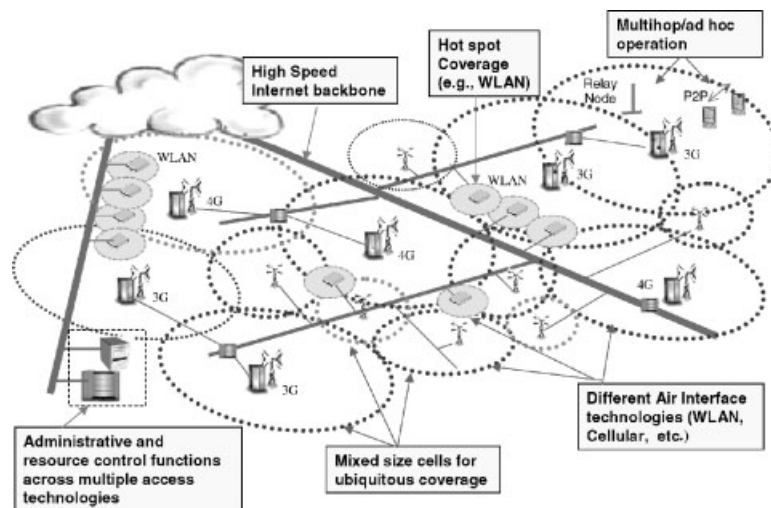


Fig. 1. Characteristics/trends seen in evolving wireless access architectures.

performance, flexible service offerings that are cost effective to users, complexity considerations and terminal capabilities. Other factors that will determine the best air interface for a given user include power constraints, cost and required QoS. The access systems are supported by the high speed Internet backbone, and are unified at the top level with common administrative functions (e.g. billing, AAA) and centralized management of resources.

2.1.2. Relays

Multihop or relaying technologies, shown in Figure 1, can be used to further enhance the data rate throughput/coverage of cellular systems [5]. Relaying can alleviate the problem of 'dead spots' in cellular systems where some areas will have no/poor coverage, by leveraging some of the terminals within a densely populated network as P2P relays. Alternately, relaying can be facilitated by the infrastructure with fixed intelligent relays operating to maximize coverage. Relaying may be preferred for delay-tolerant services, as additional delays are incurred from multihop transmission. Resource management schemes at the access node and at the mobile need to address issues specific to relaying, such as QoS based routing, multihop scheduling, cellular interference and terminal energy management, and co-ordination among nodes.

2.2. Service Implications and QoS

2.2.1. Characterization of future services

Once wireless access capabilities are enhanced to a level comparable to that of wireline systems, more resource intensive applications spanning a wide range of QoS requirements, relative to services now being supported by 3G-Ev, will need to be supported in an efficient manner. Further, the QoS requirements of a given type of service may change over time, for example, although current games are characterised as low rate, low delay (50–100 ms) services [6,7], games with much higher data rate requirements may emerge in the future. The service interface of the future wireless system should have the flexibility to address the changing needs without modifying internal radio system design.

2.2.2. Uplink and downlink symmetry

Unlike web browsing, which is download biased, the potential widespread use of symmetrical applications

such as video conferencing and wireless gaming requires a larger uplink pipe. Moreover, the potential adoption of relaying technologies implies that some mobile devices will also be carrying additional data targeted for another device. Mobile terminals will also be uploading more data due to the use of P2P applications such as file sharing programs. It is expected that the asymmetry of aggregate bandwidth requirements for the uplink and downlink may vary both on a long-term basis and on a short-term basis.

2.2.3. Service coverage and fairness

In order to improve the system's service coverage, special techniques (e.g. relays, interference avoidance) should be developed to enhance service availability in situations where it is inadequate. Service coverage ubiquity can be measured, for example, by the availability of that service over the area for a minimum number of users. The system design also offers the operator the flexibility to maximize the capacity by throttling the access of some specific users to specific services that can adapt to variable throughput; this can be done by specifying a 'fairness' objective (e.g. see 1xEV-DV fairness definition in Reference [8]). In this case, the operator may adapt the system's capabilities to suit his service offering and charging scheme. However, this does not necessarily imply that the system's service coverage has been compromised.

2.3. Air Interface Evolution

A completely new broadband air interface has to be designed for a future mobile wireless system to achieve an order of magnitude increase in required data rate and coverage relative to 3G-Ev systems. The main task is to investigate and develop a new broadband air interface that can deal with high peak rates of the order of 100–200 Mbit/s. Since the available frequency spectrum is limited, high spectral efficiency is a major feature in the new air interface design.

2.3.1. Duplex mode

The asymmetric nature of traffic in the downlink and the uplink should be considered when designing the new air interface. Even though the TDD mode provides more flexibility in splitting the available bandwidth between the downlink and the uplink, frequency division duplex (FDD) could be supplemented with additional bandwidth (if available) in either direction [9], depending on traffic needs.

2.3.2. Multiple access

Most of the standards developed for wireless high rate data transmission in recent years have been based on multicarrier modulation (i.e. orthogonal frequency division multiplexing (OFDM) [10]) as the access mechanism. Among the earliest of these were the digital audio broadcasting (DAB) standard and the terrestrial digital video broadcasting (DVB-T) standard, followed by the wireless local area network (WLANs) standards, namely IEEE 802.11a, ETSI HIPERLAN/2 and MMAC and the recent wireless metropolitan area network (WMAN) standards IEEE 802.16.

A primary reason, among several others [11], for selecting OFDM as an access option for future wireless networks is its robustness to transmission in multipath radio channels due to elimination of intersymbol interference (ISI), enabling higher bit-rates and improved system throughput with uniform coverage. OFDM also offers attractive features directly impacting radio resource management (RRM), for example, the high flexibility in bandwidth allocation to users, enabling simultaneous division of the time-bandwidth space, for ease of data rate and service adaptation.

Other multiple access methods are also under discussion for consideration in a 4G system [12]. The rate-controlled design with variable code rates and QAM modulation levels used in conjunction with OFDM offers a range of performance, which can be tailored to the user's QoS needs.

2.3.3. Antennas and spatial processing

Recent developments in space-time coding and smart antennas [13] have demonstrated the viability of these schemes for the new air interface. The combination of multiple input multiple output (MIMO) with OFDM is a key enabling technology to fundamentally improve the spectral efficiency of the future systems. It can be safely assumed that multiple antenna devices, including PDAs, will be available commercially for use with the future mobile broadband access system.

System performance can also be improved for mobile equipment without requiring multiple terminal antennas by forming a virtual antenna array configuration. For instance, a two transmit and two receive antenna configuration for the downlink is proposed in Reference [14] where the 2nd receive antenna comes from a relaying mobile. From the RRM perspective, this requires some co-ordination of scheduling between multiple users who may transmit/receive at the same time.

In order to reduce the inter- and intra-cell interference on the uplink, directional antennas can be deployed on the mobile. While this technique may not be applicable to all mobile devices and will be more suitable for devices with a larger form factor, research is ongoing towards enabling directional antennas on small form factors.

2.3.4. MAC layer evolution

The MAC layers in the different wireless technologies in use today (1xEV, WLANs, etc.) have been independently designed. If the MAC design of the future wireless system can be optimised to the physical layer parameters of multiple air interfaces (as in 802.16), the physical layer variants can be made transparent to the upper layers of the network, thus simplifying inter-system operation (e.g. handoff between access systems). The MAC design should also take into consideration the possibility that the terminal may also operate as a relay.

With the need to support very high data rates, the energy consumption of the device increases, for example, due to the use of multiple transmit and receive chains to support MIMO, and the need for continuous channel state feedback. MAC protocol overheads and high error rates causing retransmissions also drain energy from the terminal battery. MAC layer protocols should take into account the optimization of energy usage [15] in conjunction with meeting the QoS requirements of the service.

2.3.5. MAC states

The introduction of MAC states allows battery power conservation and efficient use of the limited wireless resource. In the most basic form, a mobile device will either be in an 'Active' state, where it can maintain full communication with the network in both transmit and receive modes and therefore consumes energy, or in an 'Idle' state, where it maintains only intermittent signaling in both transmit and receive modes, therefore conserving battery power with minimal radio resource utilization.

Figure 2 shows a generic MAC state diagram where the two end states are 'Active' and 'Idle'. The transition delay from 'Idle' to 'Active' is due to the signaling delay, which is largely dependent on the system design. The transition from 'Active' to 'Idle' is determined by the amount of time the mobile device is inactive. The multiple states in between the end states represent various degrees of resource utilization

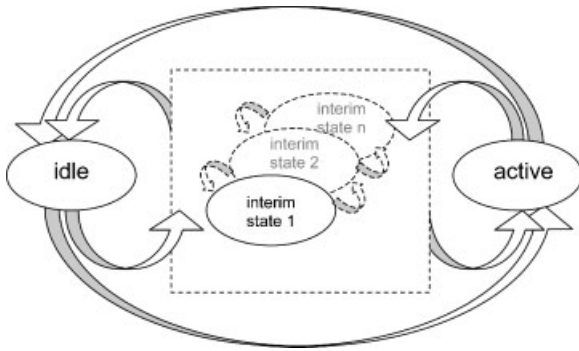


Fig. 2. A generic MAC states model. The arrows represent state transitions which are triggered by predefined conditions.

by a mobile device as deemed necessary by the system design. Such a design is prevalent in current systems [16] to optimize the use of resources (e.g. codes), which may be released when not in use, but quickly reacquired as needed.

Wireless systems employing a rate-controlled architecture [3,4,17] utilize a fast scheduler which enables the exploitation of wireless channel variations among users to increase capacity. In such a system, it is advantageous to be able to quickly transition users into the active state, but also to be able to revert these users to an intermediate low power consumption state when they are not actively transmitting. This implies that MAC states will need to work closely with scheduling in order to provide superior performance.

MAC state transitions and the associated timers (e.g. inactivity timer) are integral to guaranteeing

end-to-end QoS under varying channel conditions. For instance, a transition from an intermediate state to an active state may need to be triggered earlier than normal if the traffic flow is of a higher QoS. In a packet network where applications use TCP for flow control, MAC state design also needs to minimize the delay as seen by the upper layer to avoid falsely triggering the congestion avoidance algorithm of TCP.

Thus, optimizing MAC state design and the associated conditions for triggering transitions among the different MAC states are important topics for future research.

3. Radio Resource Management Evolution

A high performance integrated wireless access design requires RRM functions which utilize the new features in an efficient manner to deliver the high data rates and meet the QoS needs of emerging services. Figure 3 illustrates RRM functions operating at the different anchor points in the future integrated access network. The evolved RRM functions may be loosely segregated into four levels, relating to functions performed: (a) at the point of entry into the integrated wireless access network, (b) at the specific access system, (c) the access node and, (d) the mobile. Some of the functions may be replicated at multiple levels (e.g. location dependent decisions) for serving the RRM needs at those levels. The following paragraphs provide an overview of the RRM architecture.

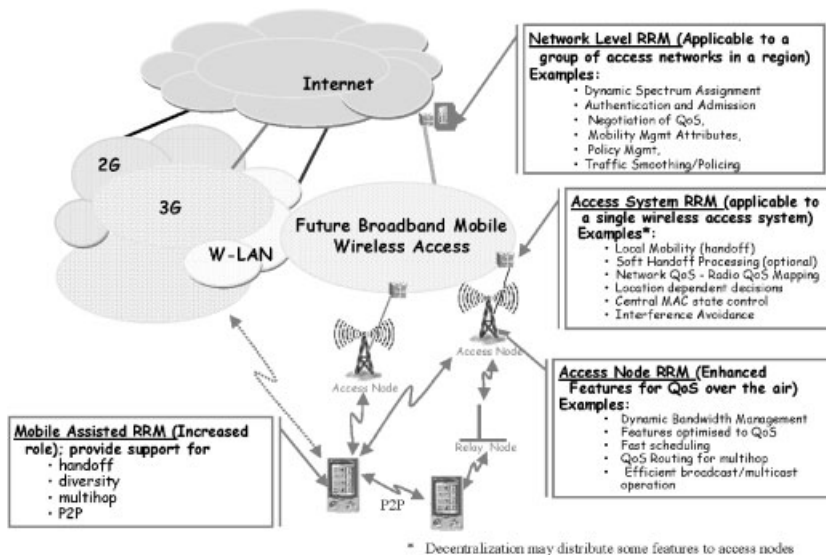


Fig. 3. Radio resource management in future broadband mobile wireless access.

First, the admission of a user (after power up) is performed at a centralized network level RRM entity in the network, common to all the air interfaces. The 'best connected' facility would be provided by dynamically allocating the appropriate access network. Admission at this level may be a simple authorization process including the transfer of Service Level Agreement (SLA)[†] related information to the selected access network entry point. Policy management including policing and shaping of traffic of already admitted sources and higher layer mobility support (i.e. handoff between two access networks) would also be the function of the network RRM entity.

Once admitted into the access network of choice, a user is entitled to 'always on' best effort service as a default. Special services requiring QoS guarantees may need to be negotiated with the network level RRM entity and/or with the access system RRM entity residing in each access network, depending on the mobility and the QoS requirement. The access system RRM unit would carry out RRM functions specific to a particular air interface. The main functions are: network QoS to radio QoS mapping, local mobility functions (i.e. handoff between two access nodes belong to the same access network), macro-diversity support functions such as multicasting,[‡] central MAC state control common to two or more access points, interference avoidance and load balancing. Should the system architecture evolve to a more decentralized network as discussed in Section 3.1., some of these functions may be moved to the access node.

The Access Node RRM entity would carry out local bandwidth management, providing statistical QoS guarantees, performance monitoring, choosing optimized parameters and the suite of air interface technologies (e.g. fast ARQ, fast scheduling, distributed MAC state control specific to the access point) for a given QoS flow/service, supporting the network level and access system level RRM functions, efficient broadcast/multicast operation and supporting functions for multihop networks. The user's geographic location will be another piece of information that is readily available to assist in RRM at the access node and elsewhere as needed.

[†]An SLA is a contract between a wireless service provider and a customer that specifies what specific level of services the subscription will support (e.g. billing, priority, QoS guarantees).

[‡]For the downlink, a mobile may receive IP packets via multiple access nodes, for which purpose the IP packets may be multicast to the multiple radio access points.

The mobile supports the RRM activity by performing and communicating measurements of the radio signal, and in some cases, proactively requesting an action, such as handoff. Thus, the mobile assisted RRM component in Figure 3 provides support for handoff, diversity, P2P communications, RF measurement and multihop operation.

Both the network and the mobile play a critical role in the optimal management of the radio resources. Optimal RRM design will distribute the functions to the most efficient elements of the network, with mobile-assisted and network-assisted RRM working co-operatively to enhance system performance.

While many of these mechanisms are being employed in 3G systems and their planned evolutions, they will need further enhancement with the introduction of new services, and the integrated approach to network access with WLAN, fixed relay and P2P networks. The evolution of selected resource management functions for the support of wide area mobility is elaborated further in the paragraphs below.

3.1. Impact of Architecture Evolution

The evolving macrocellular access system should be able to provide a robust overlay to mobile users who may be alternating between the macrocell and the use of a number of other technologies within the area of the macrocell. While the current macrocellular architecture is optimized for performance, both from the perspective of mobility and coverage enhancement with macrodiversity, it may benefit from further decentralization of functions in future systems supporting very high rate access. A primary reason for decentralization would be to allow decisions requiring a fast response so as to maximize the use of the air interface, to be taken at the nearest entity to the over-the-air link, i.e. the access node. Another reason for decentralization is to minimize the amount of redundant traffic in the access network generated from the transmission of information from the central entity (e.g. BSC) to its nodes (BTS) and *vice versa*. With the access system supporting very high rate transmissions, the redundancy may create a bottleneck. A key aspect to consider in the evolution of the macrocellular architecture is the ability to efficiently deliver the required QoS. The assumptions so far, that the network infrastructure has predetermined interconnectivity to enable rapid transfer of user status at the physical layer in support of seamless handover, and the access node to network connection has sufficient capacity to satisfy virtually all traffic demands, need

to be re-examined in the context of architecture evolution for future networks.

In evolving the cellular architecture towards a new and potentially more decentralized framework, we need to examine approaches to supporting QoS and mobility features with appropriate decentralization. Key areas that need to be addressed are:

- efficient support of mobility for all QoS levels, to enable seamless transition across different wireless access mechanisms, while minimizing the state information at the edge of the wireless network;
- support of ubiquitous service coverage, without the dependency on coordinated transmissions among access nodes for macrodiversity;
- design impact on MAC, scheduling and resource management.

3.2. Congestion Control

Congestion control enables the delivery of statistical QoS guarantees as per the service class and associated operator/customer service level agreement. Congestion control consists of a number of RRM mechanisms working in tandem and spans multiple levels of RRM, with features in the network, the access system and the access node levels. These mechanisms include, amongst others, dynamic spectrum and bandwidth management, admission control, load balancing and QoS-flow based performance monitoring.

Providing QoS guarantees is increasingly important for offering transparent, location and mobility independent service capabilities to the users. However, the delivery of QoS guarantees is hampered by the rapid fluctuations in the radio resource requirements due to changes in mobile locations, where the user at the

edge of the cell may need an order of magnitude or more in resources relative to a user located close to the base, if the same level of service needs to be maintained. Providing hard service guarantees would require more dedicated resources than average resource needs, to prevent service outages. Therefore, only statistical guarantees may be offered by wireless systems and the delivery of these requires specific techniques to manage system overloading.

One of the key challenges in the design of congestion control mechanisms would be to consistently maintain the right balance in resource usage among different classes of services depending on their expected priority levels. In general, radio resource usage of a service depends highly on its QoS requirements and the level of QoS guarantee (e.g. priority) the user expects. As will be seen later in Section 4, for the same amounts of bits sent over the channel, a delay sensitive application (e.g. VoIP) would require more than 10 times the resources than a delay tolerant application (e.g. FTP). Therefore, the radio resource costs are different from service to service depending on the QoS, user requirements and channel conditions. Based on the minimum and average needs of a given service, an operator should be able to specify a minimum proportion of the total bandwidth for specific services or service/user classes. The following mechanisms could be used in combination for congestion control in wireless systems; some of these techniques are illustrated in Figure 4.

3.2.1. Admission control, traffic shaping and policing

Multiple levels of admission control would apply in the case of traffic management between access

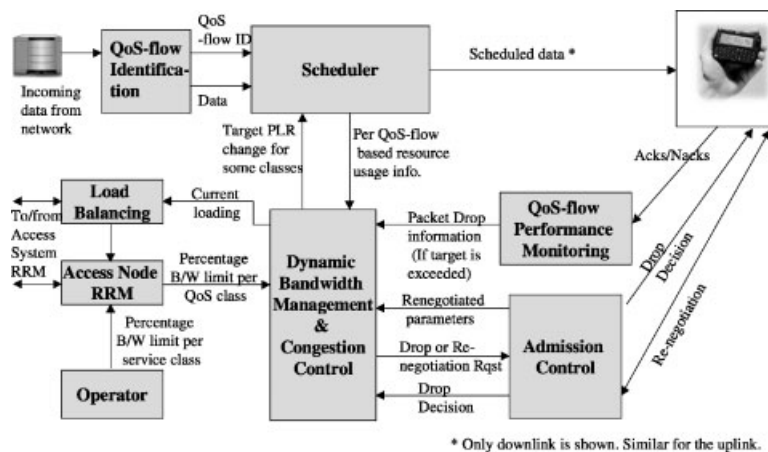


Fig. 4. Congestion control, bandwidth management and QoS guarantee mechanisms.

networks and within each access network. Admission after power-on is performed on the basis of 'best-connected', by interaction between the network level RRM entity and the associated system access RRM entity. Admission of individual sessions which require specific QoS guarantees, is done at the access system RRM level so as to avoid congestion within the access network. Quotas could be evaluated for different QoS classes based on the operator's objectives taking into account the overall effectiveness of congestion control mechanisms in satisfying the QoS requirements. A new session would not be admitted if the quota for that QoS class (see Section 3.4.) has been exceeded and/or if the system is fully loaded. Extensive studies are needed to understand how the mix of traffic impacts the number of users that can be supported for the various QoS classes while maintaining required outage levels. Self-learning techniques (see Section 3.6.) could be used to identify future resource usage trends. Traffic shaping, to limit the rate and the burst size as per the service level agreement, and policing to discard packets that are not in compliance with the negotiated service level, would be implemented prior to sending data via the air interface.

3.2.2. Dynamic spectrum management

With spectrum independently assigned by regulation to different access mechanisms and owned by different entities, it is difficult under current circumstances to envisage dynamic sharing of spectrum at the highest level, say between multichannel multipoint distribution service (MMDS) allocation and macro-cellular allocation. However, such a scenario is envisaged for the future [9,18], where the spectrum could be pooled and dynamically assigned between different access mechanisms. This scheme would reside at the network RRM level, and work with the higher-level admission control scheme to admit users into a network.

3.2.3. Dynamic bandwidth management

At the access network level, although an operator may have a preferred bandwidth proportion for each service class, when sufficient traffic is not available for a given class, its bandwidth quota could be adaptively used for the other QoS classes in order to optimize resource usage. Such schemes may be designed to accommodate both slow and rapid variation in traffic. For example, bandwidth borrowed from a delay sensitive service class (e.g. voice) may need to be

returned at short notice when a new voice session is initiated. In this case, the excess bandwidth could only be allocated to a delay tolerant service class. Additionally, a bandwidth margin could be maintained based on arrival rate, prior to borrowing.

3.2.4. Performance monitoring and congestion management

In order to identify congestion or overload situations, the system should monitor the performance of each service class or session as appropriate at all levels of the network and assess the ability to meet the QoS on a regular basis. Long-term and short-term trends should be evaluated to identify potential congestion situations in the future and appropriate actions taken as described below.

- *Delaying or discarding packets of some flows*: If the system becomes temporarily overloaded with already admitted sessions, selected packets belonging to low priority classes could be delayed or discarded. Delaying or discarding of packets could happen at multiple levels.
- *Session re-negotiation*: Temporary re-negotiation of session parameters of admitted users could be used to control congestion. The session parameters could be changed back to the original values once congestion is relieved.
- *Session discontinuation*: As a last resort, some sessions could be terminated.
- *Load balancing*: Load balancing would occur within the access network among different sectors/cells for congestion avoidance, and possibly between different access schemes in the form of re-direction of traffic. For example, the macrocellular network when overloaded may seamlessly direct or transfer low mobility traffic into a more localized access system (e.g. WLAN), if available.

3.3. Scheduling

Schedulers used in wireless networks [19] range in complexity and performance, from the basic round-robin scheduler to schedulers that use dynamic channel information for optimum performance [20,21]. The latter type of schedulers operate to take advantage of multiuser diversity in the system, scheduling users on 'up-fades' in every transmission time interval (on the order of ms) and thereby maximizing the aggregate system capacity at the expense of a reduced level of service fairness to some users. Such schedulers may operate for both downlink and uplink transmissions,

with the uplink scheduler also located in the access node and providing transmission grants to the active users.

On the uplink, it might sometimes be necessary to transmit data without waiting for the access node to schedule resource grants to the user. In such cases, autonomous scheduling can be employed. The mobile transmits data without notifying the base station scheduler, under a pre-defined assignment of resources such as data rates and channels. This form of scheduling may be particularly suited to services with constant bit rate type of traffic characteristics and for low latency needs.

For future systems, both fast and autonomous scheduling approaches will continue to be employed to allow for highly efficient utilization of radio resources. However, unlike 3G systems, it is expected that the widespread deployment of MIMO schemes in future wireless network may reduce the benefits of scheduling based largely on channel quality. With the use of more antennas, the channel fluctuations due to fast fading are reduced. Therefore capacity gain from scheduling on the up fades is smaller than that in the single-transmit and -receive antenna scenario [22]. However, scheduling based on channel quality is not completely negated because MIMO schemes do not reduce the channel fluctuations due to the slower lognormal shadowing conditions. For high-speed terminals, such fluctuations are more frequent over time. Provided a reasonably good channel prediction algorithm is in place, the scheduling gain is expected to be large. Also, the number of antennas deployed in mobile devices may still be limited due to their physical size and battery power requirements. Therefore, capacity gain by scheduling that exploits channel variations may still be substantial. The scheduler in future systems will operate to jointly optimize transmission in a multidimensional time-frequency-MIMO layer-space.

Maximizing the aggregate cell capacity should only be one of the functions of a scheduler. QoS that reflects end users' satisfaction needs to be taken into account as well. In order to efficiently support users with diverse channel conditions and QoS needs, channel adaptivity as well as QoS adaptivity should be supported [23]. While optimal scheduling prescribes that a user transmit when he/she has a good channel, the stringent QoS requirements (e.g. low delay) of the data awaiting transmission may override the channel condition criterion. The scheduler needs to support the soft statistical guarantee for the delay and the throughput of users for the various classes, in

the varying channel conditions. An adaptive scheduling approach is needed, which takes into account these multiple and sometimes contradicting requirements. As discussed in Section 2.3.5., schedulers should work closely with MAC states to deliver high performance over the wireless link.

With the introduction of virtual antenna arrays involving multiple users in the network as discussed in Section 2.3.3., schedulers would need to correlate the transmission needs and channel conditions of multiple users for simultaneous transmission. Schedulers in multihop networks would need to take into consideration the requirements for transmissions on the connected hops, especially in the event the additional hops are using the same FDD spectrum as the base station. For wireless ad-hoc networks working within a system with some centralized control (e.g. ad-hoc network of mobiles within a macrocellular system), the location of a scheduler may not be limited to the base station alone. A mobile terminal will use a local scheduler to allocate resources and prioritize data for transmission to another mobile terminal. The complexity of scheduling algorithms that run on a mobile device depends on the capability of the terminal (e.g. processing power and battery life).

3.4. Features Optimized to QoS Needs

The radio technologies that are ideal for one application may not be suitable for another application that has vastly different QoS requirements. A simple illustration of this fact can be provided by comparing a delay tolerant service such as audio streaming with a delay sensitive service such as a voice conversation. For the former, which can allow a large packet delay, ARQ retransmission can be used, while for the latter, fast retransmission techniques such as HARQ may need to be employed along with robust FEC. Similarly, a voice conversation could be efficiently served using a power controlled subsystem to ensure a constant rate, while audio streaming can be efficiently served with a rate-controlled system.

System design to match each and every application that emerges in the future is not practical, since the statistical characteristics of the applications are not known a priori. The best approach is to design the system based on a 'tool box'[¶] of air interface features which would provide a flexible mapping to an internal

[¶]The available suite of physical layer and MAC layer air interface features that can be mixed and matched to deliver the QoS requirements.

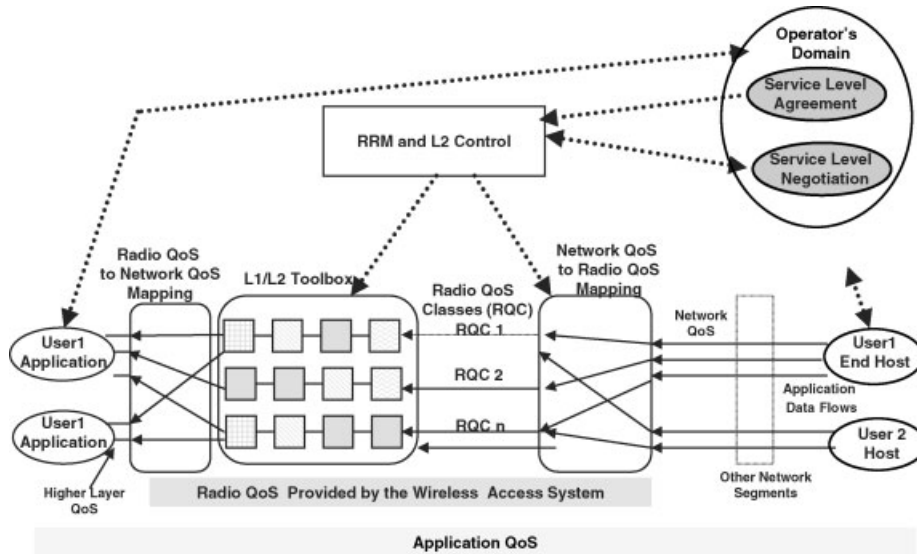


Fig. 5. A robust wireless QoS model (flexible to evolving end-to-end QoS establishment).

QoS classification, and cater to a range of QoS requirements, including the most stringent ones.

In the design of a new access system for future wireless networks, QoS dependent channelization may be provided at the access system level to support different types of services, and to satisfy the QoS performance budgets set for the radio access portion of the end-to-end network. RRM could flexibly use a suite of MAC and physical layer techniques to support the full range of QoS requirements of future services in an efficient manner. As shown in Figure 5, the 'Network QoS', the QoS defined by the immediate wireline network (e.g. IETF QoS establishment protocols), may be explicitly mapped to the access system specific 'Radio QoS', reflecting a set of physical layer and MAC layer capabilities, and the appropriate resources scheduled to service different users in the most efficient manner. QoS flows of applications that require the same set of radio technologies for resource optimization are grouped as a single Radio QoS Class (RQC). Note that this classification is internal to the access system. This would allow a flexible radio design, which is adaptable to changes in the Network QoS classifications, i.e. only the class mapping functions need to be changed if the Network QoS classification is changed; no changes are required within the wireless system. In addition, any future service can be mapped into the appropriate RQC for the most efficient delivery of that service.

The Radio QoS Class that includes per-flow based parameters (e.g. packet loss rate) is supplemented with a set of secondary QoS parameters (e.g. packet

delay), which includes per-packet based characterization. Although two packets may have the same RQC class, they may have different secondary QoS parameters for different treatment in the air interface, for example, prioritization in the schedulers.

With this approach, the access network design will support a range of performance indices having a prescribed granularity.

3.5. Handoff

With the expansion of services to be supported and the integration of multiple access networks to provide seamless service to a user as he migrates from one to the other, the complexity of handoff processing increases as well. Seamless mobility solutions, which can accomplish handoff within milliseconds, within and across networks, are desirable [24]. A combination of conventional cellular handoff and higher layer handoff may be provided depending on the QoS needs of a given service, with fast intra-system handoff needed for low delay services only. Another aspect of research in an all-IP distributed cellular network is the support of soft handoff [25] to improve coverage at the cell edge. Further work is needed both at the research as well as the standardization levels towards an optimal mobility handoff solution for all levels of QoS supported by the network.

3.6. Prediction

The increasing speed of DSP processing will enable prediction-based and self-learning based dynamic

decision making a viable solution for future systems. Prediction can be used for estimating signal strength [26,27], channel quality, the mobile's position, velocity, traffic patterns and subscriber profile.

A priori knowledge of the channel quality could be used to minimize interference or to adaptively select the coding and modulation level for transmission. Results show that channel quality measurement error (primarily caused by reporting delay) has an impact on system capacity [4, p. 33, Tables 7–9]. Thus a significant capacity improvement could be expected from reasonably accurate channel quality prediction schemes. Prediction schemes will need to address means to work with discontinuous feedback information generated by the intermittent transmission of packets rather than the continuous feedback available with circuit switched transmission.

Traffic prediction [28] of both individual and aggregate traffic characteristics could be used for resource management, for example, admission control, dynamic bandwidth management, loading based MAC state control (see Section 2.3.5.) schemes. Self-learning techniques, which correlate current variations of a given parameter to previous records of the same parameter, may be used effectively to identify favourable situations for transmission ahead of time, to identify the target access nodes for handoff, and for system deployment scenarios.

3.7. Interference Avoidance

It is expected that interference avoidance using beam-forming and beam-transmission control would be a key performance enhancement technique for use within the wireless access system of choice. There have been many studies aimed at reducing the interference by directing beams toward a specific user (or users) during the time of data transfer to/from the user(s), thus translating directly into coverage and capacity gains. In addition, these schemes allow for low power transmissions from the terminals, for saving terminal battery power. In such an environment, the user/access node needs to dynamically select the beam that will be used for both uplink and downlink (may not necessarily be the same beam) transmissions. The scheduler needs to make dynamic decisions, with the user accessing resources from different access nodes at different times, for rapidly allocating time slots and setting desired power levels and data rates for transmission. Simultaneous transmission from multiple beams may be possible with some form of co-ordination or with assistance from the

mobile. These areas need further investigation, as there is a large potential for significant capacity and coverage gain from these schemes [30].

4. Performance Evaluation

An essential element to complete the system design, is the validation of the design by evaluating it against a range of service, mobility, propagation and other deployment scenarios. A primary consideration is the definition of the future wireless service environment, which is expected to cover a larger variety of services relative to those included in current 3G-Ev system evaluation methodologies [8,31], while the other parameters, such as mobility and propagation, are not expected to change significantly.

Key generic characteristics of wireless services can be considered as: (1) bandwidth requirement & its variability (e.g. mean data rate, maximum burst size, peak rate, minimum assured rate etc.); (2) burstiness (e.g. packet inter-arrival time, packet size); (3) reliability (e.g. packet loss rate per traffic flow); (4) packet delay (on a packet by packet basis); and (5) compatibility with higher layer protocol(s) (e.g. TCP, UDP). These characteristics impact the system performance in various ways. For example, our representative simulation results shown in Table I indicate that the system throughput is reduced by a factor of 10 for a delay sensitive (packet delay bound = 20 ms) service compared with that of a delay tolerant service (packet delay bound ≥ 2 s), for a downlink rate-controlled system with full queue traffic and a C/I and delay based scheduler. Results also indicate (not shown here) that the capacity of the systems could drop significantly depending on the 'burstiness' of traffic (e.g. ftp vs. http). Although we cannot predict these parameters for each and every application that would emerge in the future, we may, for the purpose of design, safely assume ranges for these characteristics and the performance can be evaluated for a range of service characteristics, rather than for specific services.

It is important to note that conventional means of assessing coverage performance using data rate area

Table I. Impact of delay on system aggregate throughput (normalized to 4 s packet delay bound; full queue traffic is assumed).

Packet delay bound	20 ms	50 ms	100 ms	2 s	4 s
Normalized capacity	0.11	0.3	0.75	~1.0	1.0

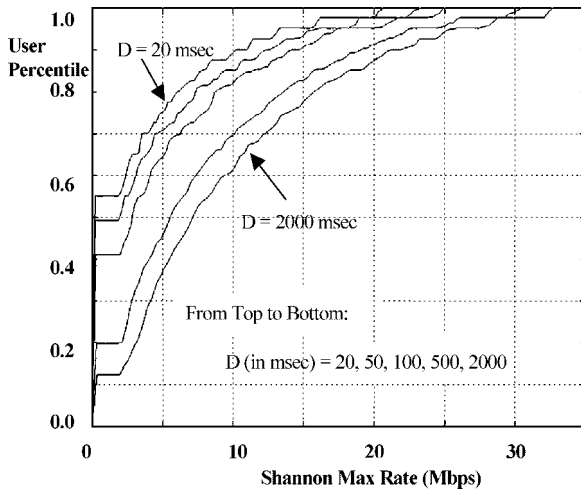


Fig. 6. Service coverage for 2 Mbps rate and different delay bound (D) values.

coverage statistics alone is not sufficient for wireless systems with diverse service offerings. For example, our results presented in Figure 6 for a rate-controlled downlink shows how the coverage of a service (with a minimum rate of 2 Mbps over 99% of the time) changes with the delay requirement. For a delay requirement of 20 ms, the area coverage is about 45%, while for a 2 s delay requirement the area coverage is around 88%. Similar impacts have been observed with other QoS parameters. This raises an interesting requirement that the coverage should be specified for a range of services, i.e. as 'service coverage', in order to derive a knowledge of true system performance. This new metric adds value specifically while evaluating different coverage enhancing solutions.

In order to identify the strengths or deficiencies of the different access layers in the air interface, the RRM schemes and the access system as a whole, the performance is usually evaluated at different levels of complexity, for different mobility situations, throughput fairness levels and different terminal capabilities. The physical layer specific performance is first evaluated with link level simulation, yielding performance metrics such as peak rate, FER (BLER)/data rate versus SINR. The physical layer performance is then combined with the cell averaged C/I distribution generated from multicell simulations to yield the aggregate throughput and the percentage area coverage for different data rates. To capture the impact of the MAC layer and RRM, full-queue traffic simulations are performed in a multiuser, multicell environment with performance metrics that include aggregate throughput and per user throughput distribution.

The simulation of a multiuser system with appropriate traffic models and with upper layer protocols such as TCP, provides the overall system performance for aggregate throughput and per user throughput distribution, and captures the effectiveness of the air interface and RRM schemes under practical conditions, with factors such as acceptable outage rates taken into consideration.

5. Conclusion

This paper has addressed some of the important aspects of resource management in evolving to future mobile broadband wireless access systems, with emphasis on ensuring the delivery of QoS, and at the same time improving system throughput by optimizing resource usage. Aspects relevant to future wireless access RRM design, with emphasis on some items such as MAC states, scheduling, congestion control and performance evaluation, haven been reviewed. Some of the other key aspects, such as handoff and prediction, have been identified to emphasize their importance, but not elaborated on due to brevity of space, while others such as efficient broadcast/multicast operation over the air, self-organization and QoS routing for multihop have not been addressed. By introducing intelligent adaptation at all levels of system design, coupled with reasonable amounts of feedback, radio resource management evolution for future-generation wireless systems will provide the foundation for delivering ubiquitous service access to the user wherever he may be.

References

1. ITU-R WP8F. Preliminary draft new recommendation (PDRN): Vision framework and overall objectives of the future development of IMT—2000 and of systems beyond IMT—2000, ITU-R M. [IMT-VIS] (Rev.1)-E 4 June 2002. Available at: http://www.fcc.gov/wrc03/files/docs/meeting/iwg/iwg_1/wrc03_iwg_1_background_doc_32.pdf
2. Mohr W. WWRF—the wireless world research forum. *Electronics and Communication Engineering Journal* 2002; **14**(6): 283–291.
3. TIA/EIA IS-856. CDMA2000 High Rate Packet Data Air Interface, Oct. 27, 2000. Available at: http://www.3gpp2.org/Public_html/specs/index.cfm#tsgc
4. 3GPP TR 25.848 v. 4.0.0 (2001–03), Physical layer aspects of UTRA High Speed Downlink Packet Access (HSDPA). Available at: <http://www.3gpp.org/ftp/Specs/html-info/25-series.htm>
5. Zadeh AN, Jabbari B, Pickholtz R, Vojcic B. Self-organising packet radio ad hoc networks with overlay. *IEEE Communication Magazine* 2002; **40**(6): 149–157.
6. Fitzek F, Kopsel A, Wolisz A, Krishnam M, Reisslein M. Providing application-level QoS in 3G/4G wireless systems:

a comprehensive framework based on multirate CDMA. *IEEE Wireless Communications* 2002; **9**(2): 42–47.

7. Bangun RA, Dutkiewicz E, Anido GJ. An analysis of multi-player network games traffic. *1999 IEEE 3rd Workshop on Multimedia Signal Processing* 1999; 3–8.
8. 1xEV Evaluation Methodology, Addendum (V6), 1xEV-DV WG5 Evaluation AHG, July 25, 2001. Available at: ftp://ftp.3gpp2.org/TSGC/Working/2001/TSG-C_0108/TSG-C-0801-Portland/WG5/
9. Spectrum Exploration for Mobile Radio. *WWRF Book of Visions*. 2001 Draft Version 1.0; 198–200. Available at: http://www.wireless-world-research.org/general_info/bookofvisions/BoV1.0/BoV/BoV2001v1.1B.pdf
10. Bingham JAC. Multicarrier modulation for data transmission: an idea whose time has come. *IEEE Communications Magazine* 1990; **28**(5): 5–14.
11. R1-02-1029, Nortel Networks, Introduction and benefits of OFDM, 3GPP TSG-RAN-1 Meeting #28, Seattle, USA, 19th–22nd August, 2002. Available at: http://www.3gpp.org/ftp/tsg_ran/WG1_RL1/TSGR1_28/Docs/Zips/
12. Kaiser S, Hunt B, Falconer D, Springer A, Wiebke T, Matsumoto T. New air interface technologies—requirements and solutions. *White Paper, WWRF/WG4/Subgroup on New Air Interfaces*, 25 October 2002.
13. Nossek JA, Utschick W (guest editors). Special Issue on Smart Antennas. *European Transactions on Telecommunications* 200; **12**(5): 393–406.
14. Dohler D, Lefranc E, Aghvami H. Space-time block codes for virtual antenna arrays. *13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications* 2002; **1**: 414–417.
15. Ephremides A. Energy concerns in wireless networks. *IEEE Wireless Communications* 2002; **9**(4): 48–59.
16. So JW, Cho DH. On effect of timer object for sleep mode operation in cdma2000 system. *IEEE International Conference on Communications* 2000; **1**: 555–559.
17. 3GPP2 C.S0001-C, Introduction to cdma2000 Standards for Spread Spectrum Systems, May 2002. Available at: http://www.3gpp2.org/Public_html/specs/index.cfm#tsgc
18. Berezdivin R, Breinig R, Topp R. Next-generation wireless communications concepts and technologies. *IEEE Communications Magazine* 2002; **40**(3): 108–116.
19. Fattah H, Leung C. An overview of scheduling algorithms in wireless multimedia networks. *IEEE Wireless Communications Magazine* 2002; **9**(5): 76–83.
20. Jalali A, Padovani R, Pankaj R. Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system. *Proceedings 51st IEEE Vehicular Technology Conference* 2000; **3**: 1854–1858.
21. Kelly F. Charging and rate control for elastic traffic. *European Transactions on Telecommunications* 1997; **8**: 33–37.
22. Kogiantis AG, Joshi N, Sunay O. On transmit diversity and scheduling in wireless packet data. *IEEE International Conference on Communications* 2001; **8**: 2433–2437.
23. Haas Z (guest editorial). Design methodologies for adaptive and multimedia networks. *IEEE Communications Magazine* (Special Issue) 2001; **39**(11): 106–107.
24. Morand L, Tessier S. Global mobility approach with mobile IP in All IP networks. *IEEE ICC* 2002; **4**: 2075–2079.
25. Zhang T, Agrawal P, Chen J-C. IP based base stations and soft handoff in all-IP wireless networks. *IEEE Personal Communications Magazine* 2001; **8**(5): 24–30.
26. Duel-Hallen A, Hu S, Hallen H. Long-range prediction of fading signals. *IEEE Signal Processing Magazine* 2000; **17**(3): 62–75.
27. Wiscom Technologies. Use of long-range prediction for channel estimation and its application in HSDPA, TSGR1#17(00) 1393, TSG-RAN Working Group 1 Meeting #17, Stockholm, Sweden, 21–24 November 2000. Available at: http://www.3gpp.org/ftp/tsg_ran/WG1_RL1/TSGR1_17/Docs/Zips/

28. Gallardo JR, Makrakis D, Angulo M. Dynamic resource management considering real behaviour of aggregate traffic. *IEEE Transactions on Multimedia* 2001; **3**(2): 177–185.
29. Zekavat SA, Nassar CR. Smart antenna arrays with oscillating beam patterns: characterization of transmit diversity in semi-elliptical coverage. *IEEE Transactions on Communications* 2002; **50**(10): 1549–1556.
30. Chuang J, Sollenberger N. Beyond 3G: wideband wireless data access based on OFDM and dynamic packet assignment. *IEEE Communications Magazine* 2000; **38**(7): 78–87.
31. 3GPP TR 25.896 V0.0.4 (2003-01), Annex A, Technical Report, Feasibility Study for Enhanced Uplink for UTRA FDD (Release 6). Available at: <http://www.3gpp.org/ftp/Specs/html-info/25-series.htm>

Authors' Biographies



Shalini Periyalwar received her Ph.D. degree in Electrical Engineering in 1992 from Dalhousie University, Canada. She held an NSERC-Industry funded position as assistant professor in the Department of Electrical Engineering at Dalhousie University from 1991 to 1994, specializing in research

in the area of coding and modulation, following which she joined Nortel Networks. She has led teams that have contributed to radio resource management features and system capacity evaluation of Nortel Networks' North American TDMA products, and to CDMA and UMTS access system design and standards. She is currently working on the design of high-capacity wireless networks in Nortel Networks' Wireless Technology Labs. She has published several journal and conference papers, and holds three patents. Her research interests include all aspects of system design for cellular and multi-hop networks.



Bassam Hashem was born in Amman, Jordan, in 1968. He received the B.Sc. and M.Sc. degrees in Electrical Engineering from KFUPM University, Dhahran, Saudi Arabia, in 1991 and 1994 respectively. In 1998, he received the Ph.D. in Electrical Engineering from the University of Toronto, Canada. From 1994 to 1995, he was

the Motorola Division Technical Manager at NASCO (Motorola Agent), Riyadh, Saudi Arabia. Dr. Hashem was with Nortel Networks, Ottawa, from 1998 to 2002, where he was involved in defining the 3G cellular systems. His last position at the Wireless Technology Labs of Nortel Networks was that of an advisor. He is currently with Saudi Telecom Company, Saudi Arabia. He is also an adjunct faculty at the Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada. Currently, Dr. Hashem is serving as the secretary of the IEEE Technical Committee on Personal Communications. He has 18 patents (issued and filed) and about two dozen papers in IEEE journals and conferences mainly on radio resource management. His research interests include power control, handoff, admission control and fixed wireless technologies.



Gamini N. Senarath was born in Hambantota, Sri Lanka, in 1958. He received the B.Sc. degree in Electronic and Telecommunications Engineering from the University of Moratuwa, Sri Lanka, in 1980, the Master of Electronic Engineering degree from Philips International Institute, The Netherlands, in 1986, and a Ph.D. in Electrical Engineering from the University of Melbourne, Australia, in 1995. He was with the Sri Lanka Telecom from 1980 to 1990 as the regional head of the western province telecommunications division and a chief engineer in charge of microwave radio backbone and switching development projects. From 1991 to 1992, he worked as a lecturer in Communications Engineering at Ballarat University, Australia. He joined Nortel Networks, Canada, in 1996 and has been involved in various projects ranging from propagation planning and handoff to future generation wireless system design. Currently, he is working as an advisor in the Wireless Technology Labs. His research interests include radio resource management, future generation architectures, media access protocols and fourth-generation system concepts.



Kelvin Au received the B.A.Sc. degree in Engineering Science in 1998 and the M.A.Sc. degree in Electrical Engineering in 2000 from the University of Toronto, Canada. He was the recipient of the Natural Sciences and Engineering Research Council (NSERC) postgraduate scholarship. In 2000, he joined the Wireless Technology Labs of Nortel Networks, Ottawa, Canada. His research interests include Medium Access Control/Radio Resource Management, semi-blind equalization for wireless communications and digital watermarking.



Robert Matyas is with Nortel Networks where he is senior manager responsible for research in wireless communications including third generation evolution and next generation systems. His group has contributed to the development of standards for UMTS and IS-2000 and their evolutions, and developed key technology that has been incorporated into Nortel's wireless infrastructure products. He has extensive experience directing research and development for satellite, digital mobile and signal processing systems for both commercial and military environments. Notable among these are an airborne communications terminal for use with the INMAR-SAT Aeronautical system, an ATM-based VSAT terminal for the European Space Agency, a SARSAT/COSPAS search and rescue satellite system digital telemetry receiver, a high-resolution image frame grabber and a secure digital mobile radio communications system using the world's first LSI MSK modem. He was also responsible for developing the terminal specifications for the North American MSAT mobile satellite system. He is the author of numerous papers in the area of digital communications. His book *Digital Communications by Satellite*, co-authored with Vijay Bhargava, David Haccoun and Peter Nuspl, has been translated into Chinese and Japanese. Bob Matyas holds engineering degrees from McGill University, Montreal and Queen's University, Kingston. He is a fellow of the Engineering Institute of Canada.

He is a fellow of the Engineering Institute of Canada.