Journal of Mobile Multimedia, Vol. 5, No.3 (2009) 181-202

QUALITY OF SERVICE (QoS) ISSUES IN MULTIMEDIA WIRELESS NETWORK – A SURVEY

SULATA MITRA

Bengal Engineering and Science University, India
sulata@cs.becs.ac.in

MOSA ALI ABU-RGHEFF

University of Plymouth, UK
mosa@plymouth.ac.uk

Received May 25, 2007
Revised April 3, 2009

The advances in multimedia applications over a wide area network have directed considerable research into the quality of service. A comprehensive exposition of the specifications and management of quality of service (QoS) in wireless networks and in distributed computing systems, supporting multimedia applications, are important for both service providers and end users. This article is a survey to explore issues concerning the quality of service in the current and future multimedia services in wireless networks. The survey is done in two phases. In the first phase, a survey of various issues in third generation multimedia applications is presented. The second phase highlights research on QoS issues in the developing technology of the fourth generation wireless networks.

Key words: Multimedia Wireless Network, QoS parameters in 3G Technology, QoS parameters in 4G technology, Current and future multimedia services
Communicated by: I. Khalil & A. Hafid

1 Introduction

Quality of Service (QoS) is a defined measure of performance in a data communication system. For example, to ensure that real-time voice and video are delivered without annoying blips, a traffic contract is negotiated between the customer and service provider that guarantee a minimum bandwidth along with the maximum delay. On the Internet and in other networks, QoS defines physical layer specifications, data path operations and control, basic router mechanism for supporting QoS, as well as end host resource support for QoS.

The physical layer specifications include transmission rates, error rates, probability of blocking and packet loss. Data path operation includes packet classification, shaping and policing, basic router mechanism for supporting QoS, including queue management and scheduling, Path control mechanisms include admission control, policy control and bandwidth brokers and traffic engineering. QoS is of particular concern for the continuous transmission of high bandwidth video and multimedia information. Transmitting this kind of content dependably, is difficult, in public networks using ordinary “best effort” protocols.

The evolution of network systems began way back in the 70s with the design of the analog-voice-oriented First Generation (1G) systems. The transition of digital voice and the data-oriented Second Generation (2G) systems in 1991 were essentially circuit switched which marked the beginning of a multi-service platform that is different from the previous mono service era. The 2.5G systems, General Packet Radio Service (GPRS), were an extension of 2G networks, in that they used circuit switching for voice and packet switching for data transmission. The 2.5G was an interim step between 2G networks and Third Generation (3G) networks, providing enhanced channel capacity, higher data rate and throughput, and optimized packet-data transmission that enhances the Internet access from different wireless devices. The commercialized transition to 3G systems in 2002 marked the beginning of a truly multimedia era where more person-to-machine interactions than person-to-person interactions were prevalent. The 3G networks were proposed to eliminate many problems faced by 2G and 2.5G networks like low speed and incompatible technologies such as time division multiple access/code division multiple access (TDMA/CDMA) which were deployed in different countries. Core packet network systems like CDMA 2000 and wideband CDMA (W-CDMA) provide higher channel capacity with theoretical broadband data up to 2 Mbps, efficient multimedia transmission and global roaming across cellular networks.

In 3G services, the subscriber is able to talk, surf the Web, download pictures, transmit video, etc., all at the same time, with different QoS requirements. The wideband CDMA 3G networks have achieved quite high practical data transmission rates – over 100 kilobytes per second running File Transfer Protocol (FTP) and over 300Kb/s in User Datagram Protocol (UDP), with round-trip packet delays of less than 300msec. Moreover, these good QoS performances are not limited to the laboratory. This era marked the beginning of a fully fledged huge revenue generating multimedia Internet applications such as e-commerce. However in 3G it is very difficult to increase the bandwidth continuously. It is also difficult to achieve the high data rate needed by the multimedia transmission for high quality multimedia services together with the existence of different services needing different QoSs. The 3G has its limitations in the allocated spectrum. It is difficult to roam across distinct service environment in different frequency bands. Furthermore, it can not support seamless mobility among a variety of networks.

In addition, the huge worldwide increase in the number of mobile users each day with emerging demands for new applications (like telemedicine, tele-geoprocessing, virtual navigation, VoIP, seamless global roaming) with ubiquitous coverage cannot be achieved with unhampered QoS support. The 3G systems have started showing limitations with bandwidth availability and interference standards.

Moreover, different short range communication systems like wireless local area network (WLAN), Bluetooth and high performance radio LAN (HIPERLAN) as well as broadcast communication systems make the situation more complicated for 3G systems. Moving beyond 3G requires a significant increase in speed and bandwidth capabilities for wireless networks. Moreover, it is difficult to extend CDMA to higher data rates due to a high spreading clock and possibly excessive interference between services.

The Fourth Generation (4G) networks is a full Internet protocol (IP) based model in which services are provided partly by the mobiles themselves and the service domain while the network domain is responsible for the bit pipe management. The 4G mobile communications will be based on the open wireless architecture (OWA) to ensure that the single terminal can seamlessly and automatically connect to the local high-speed wireless access systems, when available to the users in the offices, homes, airports or shopping centers etc. This converged wireless communication can provide high spectrum efficiency, higher data rate to the wireless, best share of the network resources and channel terminal utilization, optimal management of the service quality and multimedia applications.

From the network point of view, heterogeneity presents the major challenge for mobility management since a user may roam over a series of networks during his global travel. QoS is especially important for the new generation of Internet applications such as VoIP, video-on-demand and other consumer services. Some core networking technologies like Ethernet were not designed to support prioritized traffic or guaranteed performance levels, making it much more difficult to implement QoS solutions across the Internet. QoS parameters include access priority, bandwidth availability, latency, jitter, and packet loss. The real-time compressed and uncompressed voice and video require a fairly immediate network access that guarantees availability of bandwidth, high throughput, low latency, zero jitter, and zero loss. E-mail is at the opposite end of the QoS spectrum, as it is highly tolerant of a low level of priority; high levels of latency, jitter and loss. It does not require any bandwidth availability guarantee during the course of a mail transfer. QoS also must ensure that granting a QoS level to one traffic type or call does not violate the data flow requirements of other traffic types or calls.

The paper is organized as follows: In Section 2 we identify the parameters to measure the QoS and present a brief of some of the published work related to the QoS issues in 3G technologies. The overview of the 4G QoS issues is presented in Section 3. In this section we also present the summary of the published work related to the 4G QoS issues and to identify open research problems. Section 4 concludes the paper.

2 QoS Issues in 3G Technologies

The QoS can be measured and guaranteed in terms of the values of QoS parameters such as congestion and flow control, power control, admission control, resource management, traffic scheduling. In this section the QoS parameters in 3G technologies are identified and published works related to 3G QoS issues are reviewed.

2.1 Congestion and flow control

When too many packets are in the network awaiting transmission, its performance degrades. This situation is called congestion. When the number of packets dumped into the network by the users is within the capacity of the network, they are all delivered. However, as traffic increases beyond the network capacity, the network gets congested which increases the delay in transmission throughout the network. So flow control is required to minimize congestion. The flow can be controlled by maintaining a queue till congestion gets reduced. But such a process of flow control increases the processing delay and if the delay crosses time out it may incur huge packet loss. Evidently, this flow control technique can only be used in case the processing delay and packet loss are not the desired QoS parameters. Otherwise the flow can be controlled by not admitting a new call request in case the desired QoS is not possible to satisfy. However, such process of flow control increases call block and can be used in case the call block is not the desired QoS.

Singh et al. addressed congestion control in [1]. In this scheme, a new data user can be rejected, admitted as an active user or queued in a finite buffer at the base station (BS) depending on the status of the network. New data may enter in a queue and be admitted if sufficient resource is available. The queuing option prevents the new data from being lost to the system. Such queuing of data is useful in case of non real-time services such as e-mail, file transfer, store and forward facsimile etc., but not useful in the case of real time traffic such as voice.

A pricing based congestion control approach is proposed by Hou et al. in [2]. The objective of this scheme is to maximize the utility of wireless resources. Network utility is generally defined as the user's level of satisfaction with perceived QoS. Specifically, for higher user satisfaction, more resources should be allocated to each user. In contrast, to maximize revenue under flat rate pricing, the allocation needs to be degraded to accommodate more users. The optimal point between utility and revenue was determined in terms of the new call arrival rate. The pricing scheme was developed to achieve this optimal effective call arrival rate in the network by changing the cost of a call. It adjusts the fee dynamically by taking the state of the network into account. If the network is congested, it will charge peak hour price, which is higher than the normal hour price. According to this pricing scheme, both the demand function, which describes the reaction of users to the change in price and the peak hour price, is calculated by considering the state of the network. With this pricing scheme, a user may not initiate a call during peak hours, so network congestion can be avoided at the cost of dissatisfaction of such a user.

The queuing schemes can be used to minimize the call dropping and call blocking probability. Schemes in [3,4] help to give high priority to handoff calls and to increase system capacity in case of new calls [5]. The work in [6] considers the general queuing schemes for new call attempts and handoff call attempts. However, none of the previous models consider the effect of the mobility of queued call request. The mobiles with queued call requests are very likely to leave the current cell before being granted a free channel. Hence, a more accurate model to take this into account is needed to evaluate the actual system performance. Vincent et al. proposed in [7] a queuing algorithm to consider the effects of the handoff and queued call requests in case of finite queue length. The proposed queuing algorithm considers two different types of mobile terminals M1 and M2. M1 mobile has a call in progress and granted a channel whereas M2 mobile has a queued call request. A first in first out (FIFO) queue of finite length is used to queue up new call requests. A fixed number of channels are reserved for M1 type of mobiles. As long as the number of available channels in a cell is larger than the fixed number of reserved channels both the new call attempts and M1 handoff could get a channel. However, when the number of available channels is equal to or below the fixed number of reserved channel, only M1 handoff can use the remaining channels, whereas new calls enter into the queue when there is an available entry in the queue, otherwise they are blocked. The queued calls may initiate M2 handoffs to other cells during their wait for free channels. The priority of the M2 handoff arrival is identical to the priority of the new call arrivals. However, the presence of M2 handoffs may increase the blocking probability of the system depending upon the traffic load conditions of the heading cell. Moreover the waiting time in the queue can be reduced by reducing the queue length but it causes only a limited improvement in performance.

In [8] Parekh and Gallager proposed a packet service discipline based on generalized processor sharing and leaky bucket rate control to provide flexible, efficient and fair use of the links. This is an idealized discipline which assumes that the server can serve multiple sessions simultaneously. The authors present a simple packet-by-packet transmission scheme as an approximation to their scheme even when the packets are of variable length. It is a rate-based flow control in which average rate, maximum rate and burstiness are used as parameters of the source's traffic. The user also requests a certain QoS such as the tolerance of worst case or average delay. The network checks to see the accommodation of a new source and if succeeded, reserve transmission links or switching capacity ensure the desired QoS for the new source. The authors concentrate on providing guarantees on throughput and packet delay. However, if the number of sessions sharing the server or the packet sizes are small then it is better to approximate the scheme by weighted round robin.

2.2 Power control

Power is a fundamental resource in wireless communication systems. For existing frequency division multiple access/time division multiple access (FDMA/TDMA) systems, cell planning is based on the worst case scenario where users at the cell boundary can still have acceptable signal quality. Such an arrangement is simple but inefficient, because a user close to BS transmits at unnecessarily high power levels, which not only results in high co-channel interference, but also reduces the lifetime of the battery in the mobile unit. The dynamic adjustment of transmitted power levels can provide adequate quality to each user without causing unnecessary interference to other users.

If all users in the system transmit at the same power level, then the users closer to the BS will have stronger received signals. This may not be a problem for TDMA/FDMA systems because the signals from different users are separated in either frequency or time domain. However, CDMA systems do not have such a separation, and strong signals may flood weak signals, resulting in the so-called `near-far effect'. Thus, power control is required in CDMA systems (3G) to equal the received powers at the BS from all users in the same cell since the received power is signal strength to the desired receiver but is interference to all other receivers and its power that ultimately limits the system capacity. So, the task of power control lies not only in maintaining desired link quality, it also enables us to minimize power consumption, to prolong the battery life time in the mobile unit and to alleviate health concerns about electromagnetic emission.

For portable information devices, the efficient use of the limited power is one of the major challenges. Such power control may be achieved by using the technologies such as source signal compression, channel error control coding and radio transmission. The authors in [9] proposed a minimum total energy of a wireless image transmission system by choosing a coded source bit rate for the image code, redundancy for the Reed-Solomon (RS) coder, transmission power for the power amplifier and the number of fingers in the RAKE receiver [10]. The total energy due to channel codec, transmission and the RAKE receiver is an optimized subject to end-to-end performance of the system. The proposed scheme is simulated for an indoor office environment only subject to path loss and multi-path.

Mercado and Liu in [11] proposed an adaptive QoS for wireless multimedia networks using power control and smart antennas. The authors considered the signal-to-interference and noise ratio (SINR) of each user's channel as the QoS index. The SINR levels of different users are different depending on the requested service types. The proposed algorithm uses an iterative method to adjust the SINR levels as close as possible to the desired levels without deteriorating quality for other users. The authors also proposed a lower complexity algorithm to admit new users into the network without any considerable delay. The overall path gain between base stations and mobile stations necessary to control power is calculated using the optimal weight vectors. The purpose of the power control is to select the transmitting power of each mobile station, to maximize the effective SINR level and to minimize the overall power used by all mobile stations. However an extensive search is needed to find a feasible set of weight vectors.

Lu et al. in [12] proposed a power efficient multimedia communication that minimizes the total power consumption of a mobile transmitter by adjusting the operating parameters of the source coder, channel coder and transmitter while keeping the end-to-end source distortion constant. The authors investigate an abstract class of sources and channels and H.263 as a practical wireless video transmission. They formulate the optimization problem depending upon the models of distortions and power consumptions. For both the abstract and practical source/channel models optimum distortion-power operating points are dependent on the distance of the mobile from the base station. The transmission of each bit for large distance is costly. To reduce the cost of long distance communication, source compression is needed. These highly compressed bits are more error prone, so stronger channel codes are needed for proper error control. Higher transmit energy to combat the channel errors is required in case of large distances. The total power consumption increases with the distance. A video with more significant motion requires a higher bit rate and higher total power consumption, even at lower distances.

Wang in [13] proposed an effective connection admission control algorithm. It is a wide-band time-division-code-division multiple-access control (MAC) protocol. The author developed a minimum-power allocation algorithm to minimize the interference experienced by a code channel and to satisfy the heterogeneous bit error rate (BER) requirements of multimedia traffic. Moreover, from the analysis of the maximum capacity of a time slot, the author concluded that both data rate and BER scheduling are necessary to reach a maximum capacity. A scheduling scheme is proposed to serve packets with heterogeneous BER and QoS requirements in different time slots. An effective connection admission control algorithm is developed for the minimum-power allocation.

Another minimum-power allocation algorithm based on frequency division duplex (FDD) wideband CDMA is proposed by Wang in [14]. The proposed algorithm is used to control the received power levels of simultaneously transmitting users, so that the heterogeneous bit error rate of multimedia traffic is guaranteed. The proposed scheme provides fair queuing to multimedia traffic with a different QoS constraint. It considers the limited number of code channels for each user and the variable system capacity due to interference experienced by users in a CDMA network. A new effective bandwidth connection admission control algorithm is used to determine the admission of real-time connections. However, admission control is not considered for non-real time traffic.

The power management in wireless devices requires to be extended to communication subsystems. In many wireless embedded systems [15], the communication energy dominates the energy consumed for computation. Mariappan and Narayanasamy proposed a dynamic class based power control in [16] to improve QoS requirements with minimal total transmission power. Each user specifies a minimum tolerable QoS in terms of BER or frame error rate (FER), a maximum power limit that they can afford and a minimum data transfer rate that they require. The existing parameters in a CDMA cell are maintained using three vectors as power vector, rate vector, QoS vector. The same three vectors are also maintained for the required limits obtained from the users. The required QoS is achieved either by keeping the existing rate vector constant and by minimizing the existing power vector or by keeping the existing power vector constant and by minimizing the existing rate vector. The rate, QoS requirements and power needed for each user is different due to different type of services provided by

CDMA such as digital picture, email, fax, voice etc. The class based approach is very much efficient as it considers different types of classes for different types of users. The effect of the proposed algorithm on the reverse access channel needs to be explored.

2.3 Admission control

A user's access to a cellular communication system consists of two stages; call set up and call maintenance. During the call set up stage, the system has to decide if there is sufficient resource to accommodate the requesting user. Once the system grants access to the user, it enters the call maintenance stage during which it is the responsibility of the system to provide acceptable service quality. The decision process to whether or not to grant access to a call request is called admission control. A good admission control scheme should admit as many users as possible to maximize revenue of the system while maintaining the quality of ongoing calls.

Leong et al. proposed a call admission control (CAC) policy in [17], where a limited fractional guard channel is used to reserve resources exclusively for potential handoff calls to maintain service quality for admitted users. The voice and data users shared the total capacity in a cell. The voice users have priority access to the cell capacity whereas time variant residue capacity is available for data services. The channel holding time distribution for data users depends upon the allocated bandwidth. The authors exploit the statistical multiplexing among on/off voice calls and between voice and data traffic for high resource utilization. However, the authors assumed that the overall system is homogeneous and in statistical equilibrium.

The predictive QoS-based admission control algorithm is proposed by Epstein and Schwartz in [18]. In this distributed call admission control algorithm, information is exchanged among neighboring cells for resource reservation and admission control, while the admission control decision is made locally. The maximum number of ongoing calls is estimated depending upon the new call blocking and handoff call dropping probability. However, this scheme has huge signaling overheads and did not guarantee high resource utilization.

Zhang et al. proposed an adaptive threshold-based call admission control algorithm in [19], where a new call is admitted if the available resource is greater than a threshold. Otherwise, it is rejected or queued. This algorithm adjusts the threshold dynamically. The scheme has fewer signals overhead than [18] but works successfully under moderate traffic condition only.

Fang and Zhang proposed three call admission control strategies in [20]. In the first scheme, a new call will be blocked if the number of new calls in a cell exceeds a threshold. The handoff call is rejected only when all channels in the cell are busy. This scheme works well when the call arrivals are burst in nature. However when a big burst of calls arrive in a cell, the network may not be able to handle the resulting handoff traffic, which may lead to severe call dropping. In the second scheme proposed by the authors, a new call is accepted if the number of busy channels is less than a threshold. The handoff calls are always accepted if channels are available. This scheme works well if the average channel holding time of new and handoff calls is equal. Finally, the authors proposed a new call thinning scheme where a new call is admitted with certain probability. Thus, when the network is congested, the number of admitted new call reduces.

Wu et al. proposed a call admission control scheme in [21]. In this scheme each BS estimates the number of handoff calls for each class of traffic separately. Each BS uses such estimation to maintain a target handoff call dropping probability. But the implementation complexity of this scheme is very high due to probabilistic estimation for the potential number of handoff calls from neighboring cells.

2.4 Resource management

A user of 3G systems may connect simultaneously to multiple BSs using multiple links. The number of BSs and which BS to connect to are function of the time varying channel, mobility, current bandwidth need and QoS requirements. Once a call is admitted into the network, the resource management strategy will allow the mobile to configure dynamically its connections in such a way to minimize call dropping probability, maximize signal quality and to increase channel resource utilization efficiency. The process to assign resources to users with different bandwidth and QoS requirements as well as to provide a guarantee of negotiated service is called resource management.

Malla et al. proposed a resource allocation protocol in [22]. Here the authors assumed that each BS has some knowledge about the traffic condition in neighboring cells. Such a condition is expressed in terms of the number of incoming calls and handoff calls present in each BS, incoming and handoff calls arrival rates and/or channel holding time for each class of traffic. So to implement this scheme, a volume of communications and processing overhead is required to keep up-to-date information about the state of the neighboring cells.

Wu et al. proposed a time based reservation scheme in [23] using real time mobility predictions. This scheme is extended in [24] by taking into account the deviations of users' mobility model. In this scheme, reservation is based on the prediction of the user's direction of motion where each mobile station (MS) measures its position and orientation periodically and transmits this information to the BS to project further the MS's path. The next cell of the MS may be determined depending upon the projected path; the handoff and the duration of MS residence in the next expected cell may be estimated periodically to adjust the amount of time the resources need to be reserved. Accordingly, the MS sends a reservation request to reserve its required bandwidth during this estimated time interval. But this scheme also has a very high signaling overhead like the scheme in [22]. Moreover, the reservation of the MS may be cancelled if the MS starts to move towards a cell different from that which was computed in the latest prediction.

In [25], Misic and Bun derived arrival rate in the target cell from the finite population of mobile users in the surrounding cells. The bandwidth reservation rate at a target cell depends upon the user population in the first surrounding ring. The BS computes such a reservation rate corresponding to each new/handoff call and sends a bandwidth reservation value to all the surrounding cells. This value can be constant throughout the call duration or may be a function of the number of executed handoffs. When a call leaves the cell, the number of calls in the surrounding cells is updated. Each BS maintains the sum of those numbers associated with the bandwidth requirement of the cell. However, the scheme has very high signal overhead for bandwidth reservation. Moreover, the authors have considered only one class of traffic and a new call arrival rate is assumed constant in case of the non-target cells.

In [26], the authors considered two classes of traffic, real time and non-real time. To process real time traffic such as voice or video, a circuit switching mode is used to reduce the request access overhead and to ensure a smaller processing delay. The packet switching mode is considered to improve the resource utilization and the packet throughput for the non-real time traffic. Although the authors considered two classes of traffic, the dynamic adjustment of transmission rate for the non-real time traffic is not considered.

Jiang et al. proposed efficient resource utilization in [27]. The authors considered CDMA-related QoS provisioning with efficient resource utilization in future cellular networks. The packet scheduling is used to minimize packet loss rate and delay and to maximize throughput. An appropriate power level should be allocated to each mobile node to control interference, to overcome near-far problem and to guarantee a certain level of transmission accuracy at the bit level. The network coordination in the proposed scheme helps to achieve performance gain at the cost of additional complexity and signaling overhead.

Due to the diversity of applications that have different QoS requirements for MSs and due to the dynamic nature of wireless channel quality, the adaptive bandwidth allocation (ABA) would be necessary to improve utilization of wireless network resources. Hsing et al. proposed a dynamic resource reservation approach in [28] by reserving bandwidth dynamically in response to anticipated handoff. But this scheme has huge message exchanging overhead as in [22]. Chou and Shin proposed an ABA algorithm in [29] to allocate a target level of bandwidth to a connection as long as possible but in this scheme a high probability connection may have to wait in a queue.

High priority is given to the more sensitive classes proposed by Wang et al in [30]. It is a new bandwidth adaptation scheme based on per class degradation. When a new user arrives and there are not enough resources in the system, all connections in the lowest priority level is degraded until they reach their minimum acceptable level and if the resources are still not enough, the system continues the same degradation for the next lowest priority level. If all connections that have lower or equal priority to the new connections have degraded to their minimum and still there is not enough resource, this new connection will be rejected. This method would give more fairness between the connections of the same class and give higher priority to the more sensitive classes so that their performance would not be affected by the lower priority classes. But if the high priority class user needs resources for a longer period of time, there may be a dead lock situation.

An adaptive resource reservation scheme for multimedia mobile cellular networks is proposed in [31]. It maintains the handoff dropping probability within limit by using an adaptive bandwidth reservation scheme. In case of a handoff request, if sufficient bandwidth is not available, BS increases the reserved bandwidth for handoff by one bandwidth unit. If the handoff request was successful, the number of successful requests is increased by one. When the number of successful requests is greater than the inverse of the target handoff dropping probability, the reserved bandwidth is reduced by one bandwidth unit. Clearly the proposed algorithm counts the number of handoff failures and successes for a given handoff dropping probability threshold to maintain the handoff dropping probability below its threshold. There are two conditions to admit a new connection request at a cell. The first condition is the sum of the bandwidth being used and the desired bandwidth for a new connection must be less than the difference of cell capacity and reserved bandwidth for handoff at the current cell, where the new connection request arrives. The second condition is the sum of the bandwidth being used and the bandwidth remains reserved in the adjacent cells due to handoff from current cell, must be less than or equal to the cell capacity of the adjacent cells. Hamida and Boukhatem proposed a channel reservation scheme in [32] which is comparable based on same ideas as in [23, 24] but depends upon the mobility information of users. The mobility information includes position, velocity and acceleration parameters of the users. Each user periodically measures its position and orientation. This information is sent to the BS to compute the next cell of the corresponding mobile user. The handoff and residence times in the next expected cell are also estimated periodically to determine the duration of resource reservation. A reservation request is required to be sent by the mobile user to its next cell for the reservation of the desired bandwidth during this estimated time interval. In case of sudden path deviation of a mobile user, it is required to de-allocate the reserved resources. Again the scheme has huge signaling overhead due to the reservation and cancellation of the desired resource.

2.5 Scheduling of traffic

Future wireless multimedia networks will have a mixture of different traffic classes each has its own QoS requirements. Scheduling of the traffic is required to accommodate such heterogeneous and bursty multimedia traffic flows efficiently. The ordering of packet transmission for multimedia traffic has a great impact on the efficiency and performance of the wireless system. The design of a packet scheduler involves the balancing of a number of conflicting requirements such as maximization of throughput, QoS provisioning, scheduling according to the predefined priority structure, and low implementation complexity for packet scheduling in real time.

Kam et al. [33] proposed a credit based scheduling scheme to guarantee throughput and fairness of bursty data traffic in the downlink of a wideband CDMA system. In this scheme, each traffic flow is associated with an amount of credits, which represent the amount of service that the flow is owed, and the service for different traffic flows. The credit of each flow may be increased at each scheduling instant if the allocated bandwidth for the flow is less than the agreed upon bandwidth and can be decreased otherwise. A similar credit based scheduling scheme is proposed in [34] for each uplink channel in a wideband CDMA system. The schemes in [33,34] are simple to implement, since they do not require extensive computation but it is difficult for such schemes to provide tight delay bound, which is crucial for delay sensitive traffic.

The delay and throughput are improved in another credit based scheduling scheme [35] proposed by Stamoulis and Giannakis. The proposed low-complexity generalized processor sharing-based bandwidth scheduling scheme addresses the soft capacity issue, guarantees a minimum bandwidth to each flow according to the generalized processor sharing discipline, and distributes the extra bandwidth to traffic flows in decreasing order of the credits assigned to the users. This scheme also guarantees long-term fairness among all users due to the bounded user credits. However, the actual soft capacity is uncertain due to the various bandwidth demands of the users in the cell.

The dynamic resource scheduling is required to allocate bandwidth on demand that usually changes during the course of the connection. In [36, 37, 38], dynamic resource scheduling is considered by varying the channel rate to support multiple QoS. But the authors did not consider fair scheduling.

The computational complexity is further reduced in [39] proposed by Xu et al. which is a code-division generalized processor sharing fair-scheduling scheme. The scheme employs both dynamic bandwidth allocation and generalized processor sharing to provide fair services in wideband CDMA networks. The proposed scheduler uses both the traffic characteristics in the link layer and the adaptivity of the wideband CDMA physical layer to achieve an efficient utilization of radio resources. The channel rate or service rate of each traffic flow is adjusted on a time-slot by time-slot basis by varying the spreading factor and/or using a multiple of orthogonal code channels, rather than allocating service time for each packet. The scheme provides tightly bounded delay and guaranteed bandwidth provision but the author assumes slow-speed MSs. Moreover, the authors considered the same BER requirement for all users.

To achieve fair scheduling, a time-scheduling approach is proposed in [40,41]. Such approach is suitable for TDMA-based wireless networks. The scheme in [40], proposed by Nandagopal et al. is a generalized processor sharing-based fair scheduling scheme. It is implemented using time scheduling approach, but it needs extensive computation for the virtual time of each packet. Xu et al. proposed a dynamic fair scheduling scheme in [41] based on the generalized processor sharing fair service discipline for a wideband direct-sequence CDMA cellular network to support multimedia traffic. The authors proposed a dynamic rate-scheduling approach to perform fair scheduling in a time-slot-by-time-slot basis. The scheme exploits soft uplink capacity to design an efficient code division generalized processor sharing resource allocation procedure. The authors also proposed a credit-based generalized processor sharing scheme for further improvement of the soft capacity utilization by trading off the short-term fairness. The scheme provides bounded delays, increased throughput, and long-term fairness for both homogeneous and heterogeneous traffic.

3 QoS Issues in 4G Technologies

The different potential challenges in 4G technologies are identified in this section. The user mobility, mobility management, integration and interoperability of diverse networks, streaming multimedia based services, QoS mapping between heterogeneous networks, and call admission control are some of the QoS issues for future 4G technologies. In this section the potential research challenges, the already published works and possible research directions related to 4G technologies are considered for discussion.

3.1 User mobility

In 4G eras users must get a convenient access to the services needed at any given situation. In this context user mobility has become an important aspect in the design of 4G wireless communication systems. Researchers have focused on three models of mobility defined in [42, 43]. The terminal mobility deals with mobility of users having a single device. The session mobility deals with users in a public access network (PAN) having multiple personal devices to provide a live session. The personal mobility concentrates on provision of personalized operating environments for users along with user movements. In personal mobility a service required by the user will be delivered instantaneously, irrespective of the user’s location, device/device location, operator/provider domain, and type of network. So unlike terminal mobility, session and personal mobility concentrate more on user movements rather than terminal movements.

Wang and Abu-Rgheff in [42] discuss the disadvantages in the single-layer mobility approach that maintains strict layer modularity when applied to the emerging wireless systems and argue that such an approach can only provide incomplete mobility functionality. They propose the multi-layer approach to explore contributions from multiple layers for extended mobility functionality and improved performance.

Wang and Abu-Rgheff in [43] proposed an efficient IP-based terminal and personal mobility architecture using integrated Mobile IP (MIP) and Session Initiation Protocol (SIP). The efficiency improvements are evaluated in terms of signalling loads. The analytical and simulation results reveal that the integrated MIP-SIP reduces the signalling loads by 60% compared to its hybrid counterpart.

Thai et al. proposed integrated personal mobility architecture [44]. The scheme uses mobile agents and signaling protocols to facilitate significant optimal usage of the communication channels. This improves the system usability and allows accessibility of the required services anytime globally. This framework suffers from considerable overhead cost. Another framework in [45] facilitates users providing assistance in browsing, accessing emails, accessing files and in FTP related mechanisms.

Much research is needed in the field of session and personal mobility. There is a lot of confusion regarding the choice of core protocol out of Mobile IP (MIP) and SIP. It is also required to explore whether the ideal framework should be network layer based or application layer based for such mobility.

3.2 Mobility management

The 4G mobility management includes moving networks, seamless roaming and vertical handoff. In 4G scenario users expect to be connected to the Internet from “anywhere” at “anytime”, in fixed wireless locations or while on the move, provided that any available access network can be accommodated. For doing so, mobile networks may be multihomed i.e. having multiple points of attachment to the Internet. Moreover a user may have more than one mobile device, say a mobile phone, a laptop and a personal digital assistant (PDA). Each of these devices is likely to have multiple network interfaces that enable them to interconnect with each other as well as with other networks. These devices moving with the user together are an example of a small scale mobile network. The access networks deployed on public transportations such as ships, trains, buses and aircrafts are examples of mobile networks at a larger scale. The existing node mobility arrangement protocols, like MIP protocols [46, 47] cannot support the network mobility as the mobility service should be provided transparently to every node inside the network. A network mobility (NEMO) basic support protocol has been proposed in [48] to support this kind of network. The NEMO basic support protocol is an extension of MIPv6 [47]. In [49] Cho et al. proposed a Home Agent based (HA-based) dynamic load sharing mechanism for multihomed mobile networks. The registered neighbour Mobile Router-Home Agent (MR-HA) tunnels and measured MR-HA tunnel latency is required to provide HA based solution. A dynamic neighbour MR authentication and registration mechanism using the Return Routerability procedure of MIPv6 is considered in this work. The proposed scheme measures tunnel latency using periodic binding update (BU)/binding acknowledgement (BACK) messages and the HAHA protocol [50]. The HA can share traffic load with the neighbour MR-HA tunnel depending upon the measured tunnel latency. In [51] Shima et al. proposed two operational experiments of network mobility. The first experiment is based on NEMO basic support in a real environment. The real environment was the WIDE 2005 autumn camp meeting [51]. At the meeting a wireless network was provided to the attendees. The MR of the proposed mobile network had two network interfaces, one was for external connectivity and the other was used to provide the mobile network. But the result of this experiment shows a serious service disruption problem during handoff. The second network mobility experiment uses the WIDE 2006 spring meeting environment [51]. The multiple Care of Address (CoA) registration mechanism [52] is used in this experiment which helps to use multiple network interfaces concurrently. The MR was equipped with multiple network interfaces. So it is possible to prepare a network interface at the new foreign network, where the MR is going to move, before disconnecting from the old foreign network. The multiple CoA mechanism is useful for seamless handoff of a mobile network and the mobile network is practically usable as a moving network.

In [53], the authors proposed a policy based routing protocol. It extends the prefixes scope binding update (PSBU) message to carry sufficient topology information about nested mobile network to HA. The binding associates the network prefix with the mobile router’s CoA and a sequence of intermediated mobile router’s CoAs. The mobile network prefix identifies the home link within the Internet topology. The same IP prefix is used by all the mobile network nodes. The CoAs of the MRs are the address of the intermediated hops during packet routing into the mobile network. The MR will send a PSBU message with a chain of CoAs to register with HA and core network (CN). The HA and CN builds a binding entry in their binding cache after receiving this message. The HA and CN send packets to mobile networks using an optimal routing path. The proposed routing protocol helps to achieve high throughput.

Support for multihoming in a network mobility environment is crucial since if a MR fails to maintain session continuity this would affect the session preservation of the entire network. The multi-homing support would enhance the load sharing and fault tolerant capabilities of mobile networks. It would be commercially lucrative to provide Internet access to passengers according to the class of travel.

Seamless mobility provides easy, uninterrupted access to information, entertainment, communication, monitoring and control – when, where and how we want regardless of the device, service, network or location. The seamless mobility delivers experiences that span the home, vehicle, office and beyond instead of experiencing disconnection during movement between different devices, environments and networks. Such architecture incorporates a continuum of wide area networks, including CDMA, global system for mobile communication (GSM) and 3G, cable, fiber and digital subscriber line, emerging 4G networks and 802.16 wireless orthogonal frequency division multiplexing based broadband networks such as worldwide interoperability for microwave access. It also encompasses shorter range networks such as 802.11, Bluetooth and ultra-wide band (UWB) wireless that may be deployed in homes, vehicles and public places. All are connected to a common IP core network through gateways.

Future wireless networks must be able to coordinate services within a diverse-network environment. One of the challenging problems for coordination is vertical handoff, which is the decision for a mobile node to handoff between different types of networks. The traditional handoff is based on received signal strength comparisons; whereas vertical handoff must evaluate additional factors, such as monetary cost, offered services, network conditions and user performances.

Ormond et al. proposed a mechanism for user network selection decision in case of non real-time data application [54]. The user's terminal uses an algorithm to predict the current transfer rates of each of the listed radio access networks. It will use the predicted rates and the user's utility function to predict the most suitable network which is able to meet the data time deadline and offer the best value for money. A vertical handoff decision algorithm is proposed in [55]. The proposed algorithm considers two handoff scenarios: handoff from wireless wide area network (WWAN) to WLAN and handoff from WLAN to WWAN. In case of WWAN to WLAN handoff, the multimode mobile node (MN) in the WWAN measures received signal strength indicator (RSSI) of nearby WLAN to detect a high data rate WLAN service. The input data from both user and system are used to select an optimum wireless network for a particular service. The priority order of the preferred user wireless network could be office WLAN, residential LAN, and then public WLAN. The priority order is based on security, throughput, cost and routing performance. In case of WLAN to WWAN handoff an accurate and timely handoff decision is required to avoid the loss of WLAN access of the user during its movement due to the smaller coverage range of WLAN.

Sharma et al. proposed a vertical handoff system [56] which allows mobile users to fall back to cellular networks seamlessly in case WLAN connectivity is not available. The proposed mechanism allows a MN to operate over multiple wireless access networks independent of the end user applications. The authors proposed the design, implementation and evaluation of a fully operational vertical handoff system known as OmniCon. It is the extension of the existing Mobile IP implementation. It enables mobile nodes to switch automatically among WLAN and GPRS in a way completely transparent to the network application. OmniCon considers the issues arising out of network address translation in GPRS networks and the disparity in the link characteristics of these two technologies effectively by using transmission control protocol (TCP) tunneling and traffic prioritization mechanism.

A policy based admission control and handoff decision algorithm for next generation all-IP wireless network is proposed in [57]. It uses policy based admission control to admit a new call and policy based handoff decision algorithm to admit a handoff call. The proposed scheme also uses the network initiated vertical handoff algorithm to perform load balancing [58], to maximize battery power life time of a mobile node, to minimize power consumption, system delay and loss of packet, and finally to maximize the throughput of the network.

The dynamic mobility management for next generation all-IP wireless network is proposed in [59]. The scheme considers an all-IP integration model which provides a global roaming and internetworking facilities among different access technologies as well as seamless service negotiation including mobility, security and QoS. A mobile node receives an IP address from a local agent during initial registration. The mobile node is able to roam across any other network without any interruption in transmission and without any change in future IP address due to handoff. No handoff registration is required from the mobile node which helps to reduce the overhead in terms of power consumption of the mobile unit, location update [58] cost, delay and packet loss. Moreover, two separate interfaces in the mobile node are not required to access the cellular and WLAN domain which helps to reduce the complexity of the mobile node.

The better location coordination among the different diverse sub-networks and the location management of the future mobile software modules that will play a handoff management framework that 4G systems are required to develop. The design of new handoff decision policies and algorithms related to soft, hard and hybrid handoffs is required. The efficient handoff selection policies for optimal cross-layer performance in multimedia environments are also a future research issue.

3.3 Integration and interoperability of diverse networks

The 4G systems intend to facilitate the seamless integration and interoperation of a broad range of existing systems like satellite broadband, 3G systems, wireless local loop, fixed wireless access systems, WLAN, WiMAX, digital video broadcasting (DVB), and multimedia broadcasting multicast service (MBMS).

The intention of such open wireless architectures is to provide unhampered connectivity, full broadband access, with global roaming, and global Internet/data/voice. However, it is difficult to achieve perfect QoS and user controlled services due to the diverse nature of the constituent access technologies in terms of varying bit rates, bandwidth allocation, channel characteristics, fault tolerant levels and handoff management mechanisms.

The future wireless communication systems must offer high data rates, efficient mobility support and seamless communications [60]. Such systems have to utilize a common platform that will unify a variety of evolving access technologies, seamless interworking and interoperability solutions and adaptive multimode user terminals [61]. The evolving 4G wireless technology is a common umbrella that covers and integrates all these requirements. The development of new mobile devices is required to deal with these various network platforms and protocols. The end users will require a simple and efficient solution for transparent and seamless communications under these circumstances. 4G [62] should be exactly that seamless concept which is cost effective, simple, operable and personalized according to the users' needs. It should support the paradigm shift from technology centric to user centric concepts and should provide anytime, anyhow and always connected in a seamless manner. The most relevant 4G IP extension is the Mobile IP protocol [63] which makes the integration between all wireless as well as wirelined network platforms possible. Mobile IP introduces three new network entities (HA-home agent, FA-foreign agent, MN-mobile node). HA acts as a proxy to every registered node, redirecting incoming traffic from the content provider through encapsulation to the most recent registered location of the MN. Mobile IP operates transparently from the underlying network platform. The extension of the protocol can support regional handoffs [64]. Mobile IP allows seamless mobility, even if it occurs between domains with previously incompatible routing models. Also, IPv6 and its mobile counterparts provide advantages regarding routing advertisements and address auto configuration.

Quality enforcement mechanisms, such as Integrated Service (IntServ) and Differentiated Service (DiffServ) techniques can be used to control the IP flows through queuing, marking and dropping the packets [65]. The enforcement techniques should confirm the expectation of the customer, network type and traffic characteristics. QoS management interacts with users, in the form of service level agreement (SLA) [66] and concerns with the propagation of these expectations through the network in the form of network-level and element-level policies. The different QoS metrics are defined as appropriate for the ingredient 4G network components. They may take the form of metapolicies which combine different metrics such as mobility, characteristics of radio channel, cell size and handoff acceptance metrics.

The integrated all-IP environment is an environment of ambient intelligence [67,68]. Ambient intelligence places the user at the center of the information society and supports the shift towards the user-centric paradigm. This integrated all-IP network is lately referred to as Next Generation Wireless Network (NGWN) [69,70].

Though huge research advancement to design more enriched and integrated WiFi-3G cellular networks, WiFi-WiMAX networks, 3G-WiMAX networks and WiFi-Distributed Wireless communication systems are carried out, we still have to wait for truly 4G devices that would be universally compliant with all existing technologies.

3.4 Streaming multimedia based services

The aim of the 4G wireless multimedia communications is to provide efficient transmission of streaming data for video applications such as telemedicine, multimedia video conferencing, three dimensional (3D) virtual reality and virtual navigation. For such applications the constraints like scarce system resources, high QoS, bandwidth requirements, variation in delays and packet losses need to be overcame. The bursting and streaming video services are the popular types of services in 4G systems. The memory requirement for bursting is much larger than streaming whereas streaming should use less bandwidth as it is possible to encode multiple streams within the same file which can be transmitted using the available bandwidth. So a new streaming video application scheme is required to be implemented for optimum use of available bandwidth using limited available memory. The choice of an appropriate protocol is important in this context. The user datagram protocol (UDP) and transport control protocol (TCP) are the two important transport layer protocols for video streaming. However UDP suffers from acute congestion related problems, which may be more conveniently handled by the TCP.

Fitzek and Reisslein proposed an efficient scheme in [71] that activates video streaming at a very high data rate. This scheme shows that significant improvement in cellular capacity and video quality can be achieved for closed loop rate controlled encoded video in cellular multi-code CDMA wireless system using TCP as the transport layer protocol and simultaneous MAC packet transmission (SMPT) techniques. But this scheme is not suitable for open loop video. In [72] both channel variation and burstiness of video traffic is considered. This scheme helps to achieve significant performance for streaming video applications. This scheme is suitable for e-commerce and mobile banking applications.

More TCP friendly video streaming schemes over wireless networks are desirable to make TCP an automatic choice as the transport layer protocol over UDP because UDP may lead to the instability of the Internet with enhanced streaming video applications.

3.5 QoS mapping between heterogeneous networks

The QoS support is essential for multimedia multi-user sessions like video streaming and other multimedia-alike services. For these communication sessions the content is simultaneously destined for multiple users and so these type of sessions are called multi-user sessions. The QoS control for these sessions must be independent of the movement of users. The coordination of the QoS mapping and QoS adaptation controller allows the adaptation of the session to the current network conditions and the dynamic selection of the most suitable network service class to map the session. The multi user session distribution over heterogeneous mobile networks must be performed independently of the QoS model supported by the networks along the session path. For example, DiffServ or IntServ can be implemented to provide QoS assurance for flows of sessions in wired links, while IEEE 802.16 and IEEE 802.11e can be used in wireless links. The network services supported by class based QoS models offer different forwarding behaviors of packets. Hence, each QoS class is defined based on a set of performance metrics such as bandwidth guarantee, tolerance to loss, delay and jitter. In each network, along the end-to-end heterogeneous session path, flow of multi-user sessions with similar QoS requirements must be mapped into the appropriate service classes. In addition, the QoS mapping operations must be accomplished together with QoS adaptation support due to the existence of links with distinct capacities and the dynamic bandwidth behavior of the network resources allocated for service classes. The adaptation support avoids multi-user session blocking, keeps those sessions with acceptable quality level independent of the movement of users and causes re-routing of events in case of link or network agent failure. In a congestion period, a QoS adaptation mechanism must be used to adapt the session to the current network conditions by requesting the re-mapping of the session to a different class or by controlling the quality level of the session by dropping/adding low priority flows.

The existing approach [73] considers the static mapping of unicast session across IntServ and DiffServ QoS models. However, they are dependent on the underlying QoS model. A static mapping between UMTS QoS classes into IP QoS classes following the guideline provided by the International Telecommunication Union-Telecommunication Standardization Sector (ITU-T) Recommendation Y.1541 is discussed in [74]. The mapping of sessions from the DiffServ or IntServ models to the 802.16 QoS model is controlled by several guideline based QoS mapping solution [75]. Furthermore, other solutions require the installation of proprietary modules on the end-hosts [76] or the use of a centralized approach [77] to perform the session QoS mapping among networks with different QoS models which reduce the system flexibility and scalability. Moreover, these QoS mapping proposals assume that all neighbouring networks configure their classes with the same QoS performance metrics. The QoS adaptation control mechanisms are used to adjust the session quality level to the current network conditions. However, existing receiver-based solutions require the implementation of proprietary modules on the end-hosts to join or leave flows of multicast sessions based on notification about the network conditions [78].

Most of the QoS mapping proposals use a static guideline scheme to map the sessions without taking into account the current bandwidth of the service classes. In addition, the existing mapping schemes were developed for networks having specific QoS model or need the implementation of proprietary modules in mobile devices. The analyzed approaches present the drawback of requiring changes on the end-hosts or the installation of devices to modify the content coding within the network. The multi user session control solution is being developed in the QoS for multiuser mobile multimedia (Q3M) project [79].

The QoS mapping among networks having large number of mobile receivers, different QoS models and link capacities is required further study.

3.6 Call admission control (CAC)

Efficient call admission control algorithm is required for 4G wireless networks due to the diverse QoS requirements for multimedia applications and the presence of different wireless access technologies. The call admission control algorithm must be able to handle vertical handoff as discussed in section 3.2. It must be able to accommodate different types of users and applications with different QoS requirements. The system utilization and QoS performance can be improved for multimedia applications by adjusting the bandwidth allocation depending on the state of the network and users' QoS requirements. At both call-level and packet-level, QoS needs to be considered to design call admission control algorithms so that not only the call dropping and call blocking probabilities, but also the packet delay and packet dropping probabilities can be maintained at the target level.

Niyato and Hossain proposed a novel call admission control scheme in [80] for 4G wireless networks. The scheme is divided into two sub-modules, one for the wireless part and the other for the wired part. In the wireless part, the call admission control needs to handle multiple classes of calls as well as calls due to vertical handoff from other types of networks. If the call is used for data transfer, adaptive bandwidth allocation (ABA) can be applied to increase resource utilization. Moreover, call admission control in the wireless part must consider the nature of capacity (i.e., soft or hard) so that resource reservation and admission control can be performed optimally. The call admission control sub-module for the wired part and inter-network with the DiffServ domain is important to reduce the dropping probability for the packets already transmitted. Since the wireless resources are limited in the system, the call admission control sub-module in the wired part must ensure that the wired network can maintain the QoS of traffic from wireless users at the desired level. Both the call and packet levels performance requirements need to be satisfied in the wireless part. In the wired part, packet level QoS performance can be maintained through ABA and proper scheduling mechanisms, whereas the call level performance depends on the resource reservation and admission control strategy. However, in the wired part, only packet level QoS requirements need to be satisfied.

In [81], Song et al. proposed voice and interactive data service. The restricted access mechanism in [82] is used to share the total bandwidth between voice and data services in each network. The priority of voice traffic is considered higher than that of data traffic and occupies up to a certain amount of bandwidth to meet its strict QoS requirements. The remaining bandwidth is dedicated to data traffic. Moreover, to achieve higher resource utilization, all unused bandwidth is shared equally by ongoing data flows. The number of admitted data calls is restricted to satisfy the mean data transfer time. The voice calls are admitted with a preference to the cellular network to minimize the impact of latency and processing overhead by frequent vertical handoff. On the other hand, data traffic has a better rate adaptation capability. The data calls are admitted to the WLAN and because of the larger bandwidth of WLAN, transmission of the data packets will finish sooner, consuming the allocated resources for less time.

A dynamic resource management scheme for next generation all-IP wireless network is proposed in [83]. The scheme uses route selection algorithm, route modification algorithm and network initiated vertical handoff algorithm to select the most appropriate route for the transmission or reception of data packets which helps to perform load balancing, to maximize battery power life time of a mobile node, to minimize power consumption, system delay and loss of packet and finally to maximize the throughput of the network.

The major challenges in designing the CAC schemes for 4G wireless networks are due to heterogeneous wireless access environments, provisioning of QoS to multiple types of applications with different requirements, provisioning for adaptive bandwidth allocation, consideration of both call-level and packet-level performance measures, and consideration of QoS for both the air interface and the wired Internet.

4 Conclusion

The entire survey in the paper is conducted in two phases. The first phase introduces the QoS issues in cellular network using existing 3G technology whereas the second phase highlights the issues that require solutions in 4G technology. The first phase takes care of the QoS parameters such as congestion/flow control, power control, admission control, resource management, and traffic scheduling.

A lot of work has been carried out so far considering such QoS parameters of various levels to satisfy the user requirement at the cost of huge delay or computational complexity. In view of the fact that the cost of QoS varies from country to country, the present techniques may be upgraded to achieve varying bandwidth and transmission rate requirements to satisfy user demand and to ensure the efficient management of heterogeneous networks. The approach to resource allocation should be independent of the underlying architecture, should support varying rate channel characteristics, varying bandwidth and vertical handoffs in heterogeneous network.

After elaborate discussion presented on the work of existing technology in cellular networks, a number of fundamental issues in next generation (4G) cellular network are considered. This is a newly developing field but with the growth of research in this area, an altogether separate survey paper will shortly be necessary to present the new issues.

It follows from the above discussion that the efficient bandwidth utilization and the energy consumption are of prime importance. Literature shows that the scheduling mechanism can be used to improve energy consumption and to achieve efficient bandwidth utilization by allowing equal access to network resources among users. But the challenges of the future wireless multimedia packet switched network include integrated services, high data transmission with speeds greater than 100 Mbps and the support of multimedia services at low per-bit transmission cost. So the existing scheduling schemes, MAC protocol etc. may not be suitable for such future networks. Hence, ongoing research is required to consider the QoS satisfaction of the users in next generation multimedia wireless environment.

References

1. S. Singh, V. Krisnamurty and H. V.Poor, "Integrated Voice/Data Call Admission Control for Wireless DS-CDMA Systems", IEEE Transactions on Signal Proc., vol.50, no.6, pp.1483-1495, 2002.
2. J. Hou, J. Yang and S. Papavassiliou, "Integration of Pricing with Call Admission Control to Meet QoS Requirements in Cellular Networks", IEEE Transactions on Parallel and Distributed System, vol.19, no.9, pp.898-910, 2002.
3. D. Hong and S.S. Rappaport, "Traffic Model and Performance Analysis for Cellular Mobile Radio Telephone Systems with Prioritized and Nonprioritized Handoff Procedures", IEEE Transaction on Vehicular Technology, vol.35, no.3, pp.77-92, 1986.
4. E.C. Posner and R. Guerin, "Traffic Policies in Cellular Radio that Minimize Blocking of Handoff Calls", vol.1, pp.2.4B-2-1-2-4B-2-5, ITC 11, 1985
5. R. Guerin, "Queuing-blocking System with Two Arrival Streams and Guard Channels", IEEE Transaction on Vehicular Technology, vol.36, no.2, pp.153-163, 1988.
6. D. McMillan, "Delay Analysis of a Cellular Mobile Priority Queuing System", IEEE/ACM Transactions on Networking, vol.3, no.3, pp.310-319, 1995.
7. V.K.N. Lau and S.V. Maric, "Mobility of Queued Call Requests of a New Call-Queuing Technique for Cellular Systems", IEEE Transactions on Vehicular Technology, vol.47, no.2, 1998.
8. A.K. Parekh and R.G. Gallager, "A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks:The Single-Node Case", IEEE/ACM Transactions on Networking, vol.1, no.3, 1993.
9. S. Appadwedula, M. Goel, N.R. Shanbhag, D.L. Jones and K. Ramchandran, "Total System Energy Minimization for Wireless Image Transmission", Journal VLSI Signal Processing Systems, vol.27, no.1/2, pp.99-117, 2001.
10. http://www.cdmaonline.com/interactive/workshops/terms1/1035.htm
11. A. Mercado and K.J.R. Liu, "Adaptive QoS for Wireless Multimedia Networks using Power Control and Smart Antenna", IEEE Transactions on Vehicular Technology, vol.51, no.5, pp.1223-1233, 2002.
12. X. Lu, E. Erkip, Y. Wang and D. Goodman, "Power Efficient Multimedia Communication Over Wireless Channels", IEEE Journal on Selected Areas in Commu., vol.21, no.10, pp.1738-1751, 2003.
13. X. Wang, "Wide-Band TD-CDMA MAC with Minimum-Power Allocation and Rate- and BER-Scheduling for Wireless Multimedia Networks", IEEE/ACM Transactions on Networking, vol.12, no.1, pp.103-116, 2004.
14. X. Wang, "An FDD Wideband CDMA MAC Protocol for Wireless Multimedia Networks", IEEE INFOCOM 2003.
15. V. Raghunathan, C. Schungers, S. Park, and M.B. Srivastava, "Energy Aware Wireless Microsensor Networks", IEEE Signal Processing Magazine, vol.19, issue 2, pp.40-50, 2002.
16. V.Mariappan and P.Narayanasamy, "Dynamic Class Based Power Control for Media Streaming in CDMA", Academic Open Internet Journal, vol.15, pp.1-9, 2005.
17. C.W. Leong, W.Zhuang, Y. Cheng and L. Wang, "Call Admission Control for Integrated On/Off Voice and Best-effort Data Services in Mobile Cellular Communications", IEEE Transactions on Communications, vol.52, no.5, pp.778-790, 2004.
18. B. Epstein and M. Schwartz, "Predictive QoS-Based Admission Control for Multi Class Traffic Cellular Wireless Networks", IEEE Journal on Selected Areas in Communications, vol.18, no.3, pp.523-534, 2000.
19. T. Zhang, E. Berg, J. Chennikara, P. Agrawal, C. Jyh-cheng and T. Kodama, "Local Predictive Resource Reservation for Handoff in Multimedia Wireless IP Networks", IEEE Journal on Selected Areas in Communications, vol.19, no.10, pp.1931-1941, 2001.

20. Y. Fang and Y. Zhang, “Call Admission Control Scheme and Performance Analysis in Wireless Mobile Networks”, IEEE Transactions on Vehicular Technology, vol.51, no.2, pp.371-382, 2002.
21. S. Wu, K. Wong and B. Li, “A Dynamic Call Admission Policy with Precision QoS Guarantee using Stochastic Control for Mobile Wireless Networks”, IEEE/ACM Transactions on Networking, vol.10, no.2, pp.257-271, 2002.
22. A. Malla, M. El-kadi, S. Olariu and P. Todorova, “A Fair Resource Allocation Protocol for Multimedia Wireless Networks”, IEEE Transactions on Parallel and Distributed Systems”, vol.14, no.1, pp.63-71, 2003.
23. J.-L.C. Wu, L.-L. Lu and C.-C. Hu, “The Study of Mobility-Dependent Resource Reservation Scheme in Wireless Multimedia Networks”, ISCC, pp.1072-1077, 2003.
24. B. Hamida, L. Boukhatem, M. Neifer and S. Touati, “Efficient Resource Reservation Scheme with Optimal Channel Requests Classification in Cellular Systems”, Personal Indoor and Mobile Radio Communication, IEEE publishers, 2005.
25. J. Misic and T.Y. Bun, “Adaptive Admission Control in Wireless Multimedia Networks under Non-uniform Traffic Condition”, IEEE Journal on Selected Areas in Communications, vol.18, no.11, pp.2429-2442, 2000.
26. J.H. Kim,T.S. Kim, Y.W. Kim and D.K. Sung, “Hybrid Channel Assignment Scheme for Accommodating Voice/Data Traffic in DS-CDMA Cellular Systems”, IEEE Transactions on Vehicular Technology, vol.49, no.5, pp.1566-1577, 2000.
27. H. Jiang, W. Zhuang, X. Shen and Qi Bi, “Quality-of-Service Provisioning and Efficient Resource Utilization in CDMA Cellular Communications”, IEEE Journal on Selected Areas in Communications, vol.24, no.1, pp.1-15, 2006.
28. M.-H. Chiu and M.A. Bassiouni, “Predictive Schemes for Handoff Prioritization in Cellular Networks Based on Mobile Positioning”, IEEE Journal on Selected Areas in Communications, vol.18, No.3, 2000.
29. C.T. Chou and K.G. Shin, “Analysis of Adaptive Bandwidth Allocation in Wireless Networks with Multilevel Degradable Quality of Service”, IEEE Transactions on Mobile Computing, vol.3, no.1, pp.5-17, 2004.
30. X.G. Wang, G. Min, J. Mellor and K. Al-Begain, “A QoS-Based Bandwidth Reservation Scheme in Heterogeneous Wireless Networks”, Intl Journal of Simulation, vol.5, no.1-2, pp.9-17, 2004.
31. H.B. Kim, “An Adaptive Bandwidth Reservation Scheme for Multimedia Mobile Cellular Networks”, IEEE ICC 2005.
32. I.B. Hamida and L. Boukhatem, “A Channel Reservation Scheme using Real-time Users Mobility Information in Cellular Systems”, Wireless Personal Multimedia Communication, 2005.
33. A. C. Kam, T. Minn and K.-Y.S. Siu, “Supporting Rate Guarantee and Fair Access for Bursty Data Traffic in W-CDMA”, IEEE J. on Selected Areas in Commu., vol.19, pp.2121-2130, 2001.
34. O. Sallent, J. Perez-Romero, R. Agusti and J. Sanchez, “Uplink RRM for Conversational and Interactive Services in UTRA-FDD”, in Multi-access, Mobility and Teletraffic for Wireless Communications, X.Lagrange and B.Jabbari, Eds, Norwell, MA:Kluwer, vol.6, 2002.
35. A. Stamoulis and G. Giannakis, “Packet Fair Queuing Scheduling Based on Multirate Multipath-Transparent CDMA for Wireless Networks”, IEEE INFOCOM, vol.3, pp.1067-1076, 2000.
36. S.-J. Oh and K.M. Wasserman, “Dynamic Spreading Gain Control in Multi Service CDMA Networks”, IEEE Journal on Selected Areas in Communication”, vol.17, pp.918-927, 1999.
37. O. Gurbuz and H. Owen, “Dynamic Resource Scheduling Schemes for W-CDMA Systems”, IEEE Communication Magazine, Vol.38, pp.80-84, 2000.
38. V. Huang and W. Zhuang, “QoS-Oriented Packet Scheduling for Wireless Multimedia CDMA Communications”, IEEE Transactions on Mobile Computing, vol.3, no.1, pp.73-85, 2004.
39. L. Xu, X. Shen and J.W. Mark, “Dynamic Bandwidth Allocation with Fair Scheduling for WCDMA Systems”, IEEE Wireless Communications”, vol.9, pp.26-32, 2002.

40. Nandagopal, S. Lu and V. Bharghavan, “A Unified Architecture for the Design and Evaluation of Wireless Fair Scheduling Algorithms”, Wireless networks, vol.7, pp.231-247, 2002.
41. L. Xu, X. Shen and J.W. Mark, “Dynamic Fair Scheduling with QoS Constraint in Multimedia Wideband CDMA Cellular Networks”, IEEE Transactions on Wireless Commu. vol.3, no.1, 2004.
42. Q. Wang and M.A. Abu-Rgheff, “Next-Generation Mobility support”, IEE Communication Engineer, Vol. 1, No. 1, pp. 16-19, 2003
43. Q. Wang and M.A. Abu-Rgheff, “Signalling Analysis of Cost-Efficient Mobility Support by Integrating Mobile IP and SIP in All IP Wireless Networks”, (Wiley) International Journal of Communication Systems, Vol. 19, No. 2, pp. 225-247, 2006.
44. B. Thai, R. Wan, A. Seneviratne and T. Rakotoariyelo, “Integrated Personal Mobility Architecture: Complete Personal Mobility Solution”, Mobile Networks and Applications, pp.27-36, 2003.
45. A.D. Stefano and C. Santoro, “NetChaser:Agent Support for Personal Mobility”, IEEE Internet Computing, vol.4, no.2, pp.74-79, 2000.
46. C. Perkins, “IP Mobility Support for IPv4”, IETF RFC 3344, 2002.
47. D. Johnson, C. Perkins and J. Arkko, “Mobility Support in IPv6”, IETF RFC 3775, 2004.
48. V. Devarapalli, R. Wakikawa, A. Petrescu and P. Thubert, “Network Mobility Basic Support Protocol”, IETF RFC 3963, 2005.
49. S. Cho, J. Na and C. Kim, “A Dynamic Load Sharing Mechanism in Multihomed Mobile Networks”, pp.1459-1463, ICC 2005
50. R. Wakikawa, V. Devarapalli and P. Thubert, “Inter Home Agents Protocol (HAHA)”, IETF Internet Draft, draft-wakikawa-mip6-nemo-haha-01, February 2004.
51. K. Shima, Y. Uo, N. Ogashiwa and S. Uda, “Operational Experiment of Seamless Handover of a Mobile Router using Multiple Care-of Address Registration”, Journal of Networks, vol.1, no.3, 2006.
52. R. Wakikawa, T. Ernst and K. Nagami, “Multiple Care-of Addresses Registration”, IETF, Tech. Rep. draft-wakikawa-mobileip-multiplecoa-05, 2006
53. S.D. Adeniji, S. Khatun, R.S.A. Raja and M.A. Borhan, “Design & Analysis of Resource Management Support Software for Multihoming in Vehicle of IPv6 Network”, pp.13-17, AsiaCSN 2008.
54. O. Ormond, P. Perry and J. Murphy, “Network Selection Decision in Wireless Heterogeneous Networks”, PIMRC, vol.4, pp.2680-2684, 2005.
55. Y. Nkansah-Gyekye and J.I. Agbinya, “Vertical Handoff Decision Algorithm using Fuzzy Logic”, Auswireless 2006.
56. S. Sharma, I. Baek, Y. Dodia and T. Chiueh, “Softwate-Practice and Experience”, pp.779-798, vol.37, no.1, issue 7, 2007.
57. S. Mitra, “Policy Based Admission Control and Handoff Decision Algorithm for Next Generation All-IP Wireless Network”, vol.2, pp.594-599, IEEE ICECE 2008.
58. S. Mitra and S. Das(Bit), “On Location Management and Load balancing in Cellular Mobile Environment-A Probabilistic Approach”, vol.1, pp. 121-126, IEEE ICECE 2008.
59. S. Mitra, “Dynamic Mobility Management for Next Generation All-IP Wireless Network”, AsiaCSN 2008.
60. I. Gavrilovska and R. Prasad, “Adhoc Networking towards Seamless Communications”, Springer, 2006.
61. B. Walke, “IEEE 802 Systems:Protocols, Multi-hop Mesh/Relaying, Performance and Spectrum Coexistence”, John Wiley and Sons, 2007.
62. WWRF-Wireless World Research Forum, URL:http://www.wireless-world-research.org
63. C. Perkins, “IP Mobility Support”, IETF RFC 2002, October 1996.
64. N. Bayer, D. Sivchenko, B. Xu, S. Hischke, V. Rakocevic and J. Habermann, “Integration of Heterogeneous Ad hoc Networks with the Internet”, IWWAN, 2005.

65. A. Iera, A. Molinaro, S. Polito and G. Ruggeri, “End-to-End QoS Provisioning in 4G with Mobile Hotspots”, IEEE Network Magazine, 2005.
66. P. Haraszti, “Inter Operator Interfaces for End-to-End QoS”, Applied IP and Multimedia Services Workshop, Heidleberg, 2000.
67. J. Ahola, “Ambient Intelligence”, ERCIM NEWS, no.47, 2001.
68. L. Fuentes, D. Jimenez and M. Pinto, “Towards the Development of Ambient Intelligence Environments using Aspect-Oriented Techniques”, ACP4IS, 2004.
69. J.-C. Chen and T. Zhang, “IP-based Next Generation Wireless Networks: Systems, Architectures and Protocols”, Wiley, 2004.
70. “NGWNet: Next Generation Wireless Network, EPSRC funded project. URL:http://www.ngwnet.ac.uk/index.php
71. F.H.P. Fitzek and M. Reisslein, “Wireless Video Streaming with TCP and Simultaneous MAC Packet Transmission (SMPT)”, Intl Journal of Communication Systems, vol.17, pp.421-435, 2004.
72. R.S. Tupelly, J. Zhang and E.K.P. Chong, “Opportunistic Scheduling for Streaming Video in Wireless Networks”, Conference on Information Sciences and Systems”, 2003.
73. Z. Mammeri, “Approach for End-to-End QoS Mapping and Handling”, IEEE/IFIP Wireless and Optical Communications Networks, 2005.
74. R. Ben Ali, R.B .Ali, S.Pierre and Y. Lemieux, “UMTS-to-IP QoS Mapping for Voice and Video Telephony Services”, IEEE Network, 2005.
75. J. Chen, W. Jiao and Q. Guo, “An Integrated QoS Control Architecture for IEEE 802.16 Broadband Wireless Access Systems”, IEEE Global Telecommunication Conference, 2005.
76. M. El-Gendy and K.G. Shin, “Paving the First Mile for QoS-dependent Applications and Appliances”, IEEE International Workshop on QoS, 2004.
77. M. Ruy et al., “QoS Class Mapping over Heterogeneous Networks using Application Service Map”, IEEE International Conference on Networking, International Conference on Systems and International Conference on Mobile Communications and Learning Technologies, 2006.
78. Q. Zhang, Q. Guo, Q. Ni, W. Zhu and Y.-Q. Zhang “Sender-adaptive and Receiver-driven Layered Multicast for Scalable Video over the Internet”, IEEE Transaction on Circuits and Systems for Video Technology, vol.15, 2005.
79. E. Cerqueira, L. Veloso, A. Neto, M. Curado, E. Monteiro and P. Mendes, “A Unifying Architecture for Publish-Subscribe Services in the Next Generation IP Networks”, IEEE Global Telecommunication Conference, 2006.
80. D. Niyato and E. Hossain, “Call Admission Control for QoS Provisioning in 4G Wireless Networks: Issues and Approaches”, IEEE Networks, pp.5-11, 2005.
81. W. Song, H. Jiang, W. Zhuang and X. (Sherman) Shen, “Resource Management for QoS Support in Cellular/WLAN Interworking”, IEEE Networks, pp.12-18, 2005.
82. M. Naghshineh and A.S. Acampora, “QoS Provisioning in Microcellular Networks Supporting Multiple Classes of Traffic”, Wireless Networks, vol.2, no.3, pp.195-203, 1996.
83. S. Mitra, “Dynamic Resource Management in Next Generation All-IP Wireless Network”, vol.1, pp.165-170, IEEE INDICON, 2008.