A SURVEY STUDY OF QoS IN NEXT GENERATION MOBILE NETWORKS

Ján PAPAJ
Department of Electronics and Multimedia Communications, Faculty of Electrical Engineering and Informatics,
Technical University of Košice, Letná 9, 042 00 Košice, Slovak Republic, tel.: +421 55 602 2296, E-mail: jan.papaj@tuke.sk

SUMMARY

Current trends in communication networks point to an aggregation of all kind of traffic (data, voice, etc) in the same transport technology. The emergence of real-time applications and the widespread use of wireless and mobile devices have generated the need to provide quality of service (QoS) support in wireless and mobile networking environments. QoS present a key problem of today's mobile networks and plays an important role in QoS provisioning for mobile networks. Supporting appropriate QoS for mobile networks is a complex and difficult issue because of the dynamic nature of the network topology and generally imprecise network state information, and has become an intensely active area of research in the last few years. Although lots of research has been done on supporting QoS in the Internet and other networks but they are not suitable for mobile networks and still QoS support for such networks remains an open problem. In this article we introduce basic definition of QoS in mobile networks and discuss trends in QoS for next generation mobile networks.

Keywords: *QoS, GSM, UMTS, MANET, WLAN, MATM*

1. INDRODUCTION

Providing QoS guarantees to various applications is an important objective in designing the next-generation mobile networks. Different applications can have very diverse QoS requirements in terms of data rates, delay bounds, and delay bound violation probabilities, among others. For mobile networks, since the capacity of a wireless channel varies randomly with time, an attempt to provide deterministic QoS (i.e., requiring zero QoS violation probability) will most likely result in extremely conservative guarantees.

2. CLASSIFICATION OF MOBILE NETWORK

The history and status of mobile communications are shortly listed in the following, together with the respective evaluations on the chief contributions [1]:

a) Traditionally, wireless systems were considered as an auxiliary approach that was used in regions where it was difficult to build a connection by wireline.
b) 1G was based on analogy technique and deployed in the 1980s. It built the basic structure of mobile communications and solved many fundamental problems, e.g. cellular architecture adopting, multiplexing frequency band, roaming across domain, non-interrupted communication in mobile circumstances, etc. Speech chat was the only service of 1G.
c) 2G was based on digital signal processing techniques and regarded as a revolution from analogy to digital technology, which has gained tremendous success during 1990s with GSM as the representative. The utilization of SIM (Subscriber Identity Module) cards and support capabilities for a large number of users were 2G's main contributions.
d) 2.5G extended the 2G with data service and packet switching methods, and it was regarded as 3G services for 2G networks. Under the same networks with 2G, 2.5G brought the Internet into mobile personal communications. This was a revolutionary concept leading to hybrid communications.
e) 3G is deploying a new system with new services instead of only providing higher data rate and broader bandwidth. Networks combine a choice of QoS levels with increased bandwidth, enabling bandwidth-intensive applications and paving the way towards QoS-aware applications for the mobile user. Based on intelligent DSP techniques, various multimedia data communications services are transmitted by convergent 3G networks (UMTS [2], MANET [4], WLAN [4], MATM [10]).
f) The 4G concept supports the provisioning of multiple types of services, ranging from simple network access to complex multimedia virtual reality, including voice communication services, which are themselves a challenge in packet-based mobile communications environments (WLAN [6], MANET [8], MATM [6]).

3. TAXANOMY OF QoS PARAMETERS

QoS parameters for typical applications cover various aspects, such like bounds for bandwidth, packet delay, packet loss rate, jitter and much more. Certain additional parameters that deal with problems unique to wireless and mobile networks are required. Generally, they can be enumerated as below [5]:

- **Mobility problem -** mobility will enable seamless handover between multi-access networks. IP telephony or such services will become seamless even in multi-access environment.
- **High bit error problem -** Packet loss in a mobile environment is an important issue to be considered because of the limited bandwidth of a wireless network and the possible fading and

blackout situations that can occur when a mobile moves from one cell to another.

- **Bandwidth limitation problem -** the bandwidth of the wireless link connecting an MN to the static segment of the network is significantly lower than the one of the wired links between static hosts.
- **Resource constraint problem -** always hold in the cases of memory and storage capacity of the mobile device and in terms of their computational power.
- **IP tunneling problem -** problem arising from this tunnel-based communication is that different data flows addressed to the same terminal (same IP address) are treated in the same manner.

During years, considerable research interests have been devoted in the mobile QoS field. As a result, some definitions concerning the QoS issues in the mobile Internet have appeared in the literature. To better understand the latest achievements, we outline some new but promising terms [5]:

1. **Mobile QoS reference point -** it is stated that QoS measures are only quantifiable at a service access point.
2. **Mobile QoS components -** A connection involving at least one mobile user can be viewed as the concatenation of fixed and wireless links.
3. **Mobile QoS parameters -** QoS objectives should include appropriate metrics. A clear distinction is made between network performance parameters that can be objectively measured and subjective QoS parameters depending on user perception.
4. **Loss profiles -** gives applications an opportunity to choose between a bursty loss and distributed loss in case of an overloaded situation.
5. **Power Level -** parameter informs the base station (BS) about the battery power situation in the mobile and the BS changes the way it schedules packets based on the power profiles.
6. **QoS Object -** is introduced depending on the context; the QoS Object is included as a Destination Option or a Hop-by-Hop Option in IPv6 packets carrying Binding Update and Binding Acknowledgment messages.
7. **Probability of seamless communication -** defines the nature of breaks that can be allowed in the service.
8. **Shadow cluster -** The "shadow cluster“ concept is a concept used to improve resource allocation and call admission in ATM-based wireless networks.
9. **Mobiware -** is a novel QoS-aware middleware platform, which operates between the application and radio ATM link layers.

4. QoS FOR MOBILE NETWORKS

GSM

The traditional telecommunication networks (GSM) guarantee a high and fixed QoS by using circuit switching for real-time applications, which consumes a lot of system capacity. This is due to the fact that a link is reserved for the entire lifetime of a connection and the capacity is provided even for times where no data is transferred. On the other hand, packet switching allow for more efficient use of the system capacity, user idle time and volume charging policy [6]. Also in terms of packet-loss and packet-delay the packet switching (PS) is the technology of choice over the radio interface. Finally, packet switching allow for fast development of complex services thus making the PS the main evolution area in the telecommunication field.

UMTS

Universal Mobile Telecommunications System (UMTS) represents an evolution in terms of capacity, data speeds and new service capabilities from second generation mobile networks [2]. The evolution of wireless systems to third generation (3G) has an important driving factor, which is the need to provide new services that increase the average return per subscriber. Compared to the Global System for Mobile Communications (GSM) and other existing mobile networks, the UMTS provides a new and important feature- namely, it allows negotiation of the properties of a radio bearer [4]. Attributes that define the characteristics of the transfer may include throughput, transfer delay, and data error rate. To be a successful system, UMTS has to support a wide range of applications that possess different QoS requirements. Network services are considered end-to-end, or user-to-user. An end-to-end-service may have a certain QoS, which is provided for the user of a network service. To realize a certain network QoS, a bearer service with clearly defined characteristics and functionalities must be set up from the source to the destination of a service.

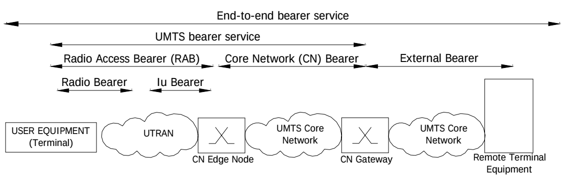

Fig. 1 UMTS QoS Architecture

In order to allow an easier development of QoS guaranteeing algorithms all services will be divided in four defined QoS classes [7]:

- **Conversational class -** which is intended for real-time traffic. This class is very delay sensitive, but it can stand bit errors and packet losses. Good examples of such services are voice and video telephony where speech/video codec’s can conceal errors. The short delay is the most essential feature for the users (human) of the service.
- **Streaming class -** which is very similar to the conversational class with the exception that more delay, is tolerated. The increased delay provides a larger variety of means for achieving lower

error rate. The class is suitable for cases where one end of the connection is human and the other is a machine.

- **Interactive class -** which is intended for the traffic that allows delay variation while requiring reasonably low response time. An example would be Web browsing: the channel can be unused for long periods of time but when a user makes a request for a new page the response time should be reasonably low. Due to less stringent delay requirements, the error rates can be improved by having better channel coding and applying re-transmissions.
- **Background class -** which is described by its name. This class will get service with lowest priority when there are resources to be utilized. Background class is cheap and suitable for applications such as e-mail. This class requires to that the packets should be transmitted with a low bit error rate.

There are many QoS parameters/attributes defined for UMTS [7]: maximum bitrate (kbps), guaranteed bit rate (kbps), delivery order (y/n), maximum SDU (service data unit) size (octets), SDU format information (bits), SDU error ratio, residual bit error ratio, Delivery of erroneous SDUs (y/n/-), transfer delay (ms), traffic handling priority, allocation/retention priority, source statistics descriptor ('speech'/'unknown'), and Signaling indication (yes/no).

The UMTS QoS general requirements define the constraints the set of attributes characterizing the QoS should meet [2]:

- QoS attributes negotiation between UE and UMTS CN Gateway node should be possible as well as renegotiating the QoS for active sessions. The UE and 3G CN Gateway node should be able to indicate the QoS properties to the application layer.
- Interoperability with previous existing QoS schemes should be assured. The overall complexity generated by the QoS mechanism should be low.
- Mapping between the application QoS attributes and the UMTS services should be done by the UMTS QoS mechanisms.
- The QoS mechanisms should assure different levels of QoS using the UMTS mechanisms independent of QoS mechanisms of other networks.
- It should be possible to have different QoS attributes for multiple streams of a session. [7] A session is considered to be a progression of events devoted to a particular activity. A stream provided to a session is a distinct service with its own QoS attributes. For example, for a given session. Simultaneous voice and data transfer should be possible, and each of the different streams should be provided with different QoS.
- Asymmetric bearers (with different QoS for uplink and downlink) should be supported.

The main challenges that the UMTS QoS architecture has to overcome are [7]:

Translation parameters and mechanisms - Service differentiation based on a set of traffic classes needs a simple and reliable translation mechanism between the different domains involved.

UMTS QoS Management - The network should be monitored and managed to assure the implementation of the user agreements. Negotiation and modification of the QoS available from the network should be possible.

MANET

A mobile ad hoc network (MANET) is an autonomous system of mobile routers (and associated hosts) connected by wireless links [8]. The routers are free to move randomly and organize themselves arbitrarily; thus, the network's wireless topology may change rapidly and unpredictably. Such a network may operate in a stand-alone fashion, or may be connected to the larger Internet MANETs are useful in many applications because they do not need any infrastructure support. Collaborative computing and communications in smaller areas (building organizations, conferences, etc.) can be set up using MANETS. Communications in battlefields and disaster recovery areas are further examples of application environments. With the evolution of Multimedia Technology, Quality of Service in MANETs became an area of great interest. Besides the problems that exist for QoS in wire-based networks, MANETS impose new constraints. This is due the dynamic behaviour and the limited resources of such networks.

QoS is usually defined as a set of service requirements that needs to be met by the network while transporting a packet stream from a source to its destination. The network needs are governed by the service requirements of end user applications. The network is expected to guarantee a set of measurable prespecified service attributes to the users in terms of end-to-end performance, such as delay, bandwidth, probability of packet loss, delay variance (jitter), etc. Power consumption is another QoS attribute which is more specific to MANETs. In the literature, the research on QoS support in MANETs spans over all the layers in the network:

- QoS models specify an architecture in which some kinds of services could be provided. It is the system goal that has to be implemented.
- QoS Adaptation hides all environment-related features from awareness of the multimedia-application above and provides an interface for applications to interact with QoS control.
- Above the network layer QoS signaling acts as a control center in QoS support. The functionality of QoS signaling is deter-mined by the QoS model.
- QoS routing is part of the network layer and searches for a path with enough resources but does not reserve resources.

- QoS MAC protocols are essential components in QoS for MANETs. QoS supporting components at upper layers, such as QoS signaling or QoS routing assume the existence of a MAC protocol, which solves the problems of medium contention, supports reliable communi-cation, and provides resource reservation.

QoS architecture includes all networking layers from the application layer to the MAC layer (Fig. 2) [8]. Each layer's features are detailed below.

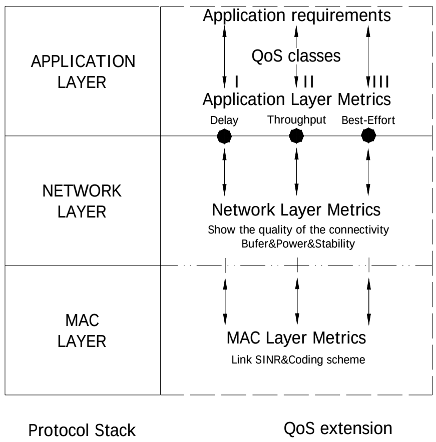

Fig. 2 MANET QoS Architecture

Application layer [8] **-** Applications can be categorized into real-time and non-real-time applications based on their sensitivity to packet delay. Real-time applications have strict requirements on the packet delay. Therefore, packet retransmission is not allowed. On the other hand, for non-real-time applications such as Email and FTP, packet delay is not a big issue, and packet delivery is guaranteed by explicit acknowledgements in the transport layer. The network should be designed to meet the different packet delay requirements of these two types of applications. At the application layer, we propose to classify the QoS requirements into a set of QoS priority classes with their corresponding application layer metrics (ALMs). For example, we classify application requirements into three QoS classes, I, II, & III, and map them to appropriate metrics. Class I corresponds to applications that have strong delay constraints, such as voice. This class is mapped to the delay metric at ALMs. Class II is suitable for applications requiring high throughput such as video or transaction-processing applications. Similarly, we map this class to the throughput metric at the ALMs. Class III has no specific constraints, and it is mapped to best-effort at the ALMs. This mapping is shown in Fig. 2.

Transport layer [8] **-** UDP and TCP are two transport layer protocols widely used in wired networks. UDP has no congestion control scheme to react to network congestion. Applications that use UDP as the underlying transport protocol to transmit packets can easily overwhelm the network with data, which results in a considerable amount of wasted energy and bandwidth in transmitting packets that will be dropped due to congestion. Therefore, some information from the packet queue and the routing layer should be sent to the transport layer for performance optimization.

Network layer [8] **-** To support QoS, the routing protocol should have an embedded scheme such as call admission or adaptive feedback that is designed to support QoS. Therefore, two cross-layer designs should be implemented in QoS-aware routing. Overall, QoS-aware routing should have the following features that traditional routing does not support:

- obtain resource information from lower layers;
- offer bandwidth information to applications;
- incorporate resource reservation schemes; and predict route breaks.

At the network layer, we recommend to use nodes' power state, buffer state, and stability state to characterize the quality of network (see Fig. 2), and they call them network layer metrics (NLMs). The power level represents the amount of available battery over time (i.e. energy). The buffer state stands for the available unallocated buffer. The stability means the connectivity variance of a node with respect to its neighbouring nodes over time. To compute the quality of a path, concave or/and additive functions have to be used in order to represent the NLMs of a path given the value of these metrics for individual nodes on that path. The network layer metrics of a particular node can also reveal whether the node is forced to be selfish or not. In the selfish mode, a node can cease to be a router and acts only as a host due to its poor quality.

Link layer [8] **-** needs to discriminate the different priority packets and schedule packet delivery according to priority levels. The service differentiation should be completed in the packet queue through queue management and in the MAC layer through a MAC discriminator and priority classifier.

- **Queue Management -** The aim of queue management is to schedule the different priority packets.
- **MAC Discriminator -** The main function of the MAC discriminator is to differentiate data packets and control packets that arrive from the wireless channel.
- **Priority Classifier and Packet Scheduler -** To offer service differentiation in a distributed ad hoc network, real-time packets should be granted higher priority to capture the channel.

WLAN

IEEE 802.11 wireless LAN (WLAN) [9] is one of the most deployed wireless technologies all over the world and is likely to play a major role in next generation wireless communication networks. WLAN is a shared-medium communications network that broadcasts information over wireless

ISSN 1335-8243 © 2006 Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

links for all stations to receive. In the original 802.11 standard published in 1997, the MAC (Medium access control) and PHY (Physical layer) definitions are described (Fig. 3). The other 802.11 standards are either the enhancements to the original MAC for QoS and security, or the extension to the original PHY for high-speed data transmission.

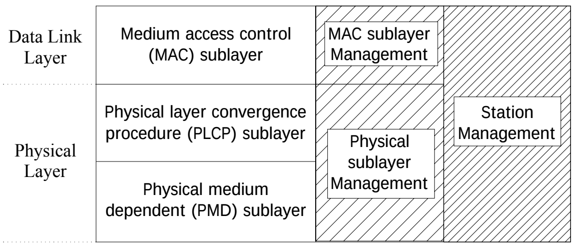

Fig. 3 WLAN QoS Architecture

However, multimedia applications require some qoS, delay, jitter and error rate. Guaranteeing those QoS requirements in 802.11 WLAN is very challenging due to the QoS-unaware functions of its MAC layer and the noisy and variable physical (PHY) layer characteristics.

The 802.11 MAC supports two basic medium access protocols: **contention-based distributed coordination function (DCF)** and **optional point coordination function (PCF)**. DCF is used normally for non real- time services and the centralized access mode. PCF is for real-time services to provide higher priority than for non-real time services.

Most existing QoS mechanisms for 802.11 can be classified into three categories [9]:

- **Service differentiation**
- **Admission control and bandwidth reservation**
- **Link adaptation**

Research has been done to provide certain DCF-based QoS enhancements that mainly address effective support of service differentiation. In fact, these mechanisms do not provide any QoS guarantee, only better than best effort services. Basically, service differentiation is achieved by two main methods: priority and fair scheduling. While the former binds channel access to different traffic classes by prioritized contention parameters, the latter partitions the channel bandwidth fairly by regulating wait times of traffic classes in proportion according to given weights.

Service differentiation is helpful in providing better QoS for multimedia data traffic under low to medium traffic load conditions [3]. However, due to the inefficiency of IEEE 802.11 MAC, service differentiation does not perform well under high traffic load conditions. In this case, admission control and bandwidth reservation become necessary in order to guarantee QoS of existing traffic. Otherwise, the extremely large saturation delay may lead to failure to support multimedia applications.

MATM

The goal of a mobile ATM network is to provide transport for ubiquitous mobile services with different radio technologies and QoS requirements [10]. Since mobile ATM is a compatible upgrade to a conventional ATM network, it can support both mobile and fixed hosts on the same infrastructure. In order to support mobile services, the ATM network protocols must be enhanced to support three basic mobility functional components: wireless access, location management and handoff control.

Five service categories have been defined under ATM [10]. These categories are differentiated according to whether they support constant or variable rate traffic, and real-time or non-real-time constraints. The service parameters include a characterization of the traffic and a reservation specification in the form of QoS parameters. Also, traffic is policed to ensure that it conforms to the traffic characterization, and rules are specified for how to treat nonconforming traffic. ATM provides the ability to tag nonconforming cells and specify whether tagged cells are policed (and dropped) or provided with best-effort service. The service categories are constant bit rate (CBR), real-time variable bit rate (rt- VBR), non-real-time variable bit rate (nrt-VBR), unspecified bit rate (UBR) and available bit rate (ABR).

QoS in wireless ATM requires a combination of several mechanisms acting in concert. QoS mechanisms are needed:

- **At the radio interface -** A QoS-capable medium access control (MAC) layer is required. The mechanisms here are resource reservation and allocation for ATM virtual circuits under various service categories, and scheduling to meet delay requirements. Furthermore, an error control function is needed to cope with radio link errors that can otherwise degrade the link quality. Finally, a CAC mechanism is required to limit access to the multiple access radio link in order to maintain QoS for existing connections.
- **In the network -** ATM QoS mechanisms are assumed in the network. In addition, a capability for QoS renegotiation will be useful. This allows the network or the mobile terminal (MT) to renegotiate the connection QoS when the existing connection QoS cannot be maintained during handover. Renegotiation may also be combined with soft QoS mechanisms, as described later. Finally, mobility management protocols must include mechanisms to maintain QoS of connections rerouted within the network during handover.
- **At the MT -** The MT implements the complementary functions related to QoS provisioning in the MAC and network layers. In addition, application layer functions may be implemented to deal with variations in the available QoS due to radio link degradation and/or terminal mobility. Similar functions may be implemented in fixed terminals communi-cating with MTs.

MAC Layer Functions - The radio link in a wireless ATM system is typically a broadcast multiple access channel shared by a number of MTs. Different multiple access technologies are possible, for instance frequency, time, or code division multiple access. A combination of FDMA and dynamic TDMA is popular in wireless ATM implementations. That is, each radio port (RP) operates on a certain frequency band and this bandwidth is shared dynamically among ATM connections terminating on multiple MTs using a TDMA scheme. ATM QoS is achieved under dynamic TDMA using a combination of a resource reservation/allocation scheme and a scheduling mechanism [10].

Network and Application Layer Functions - Wireless broadband access is subject to sudden variations in bandwidth availability due to the dynamic nature of the service demand (e.g., terminals moving in and out of RPs coverage area, variable bit-rate interactive multimedia connections) and the natural constraints of the physical channel (e.g, fading and other propagation conditions). QoS control mechanisms should be able to handle efficiently both the mobility and the heterogeneous and dynamic bandwidth needs of multimedia applications. In addition, multimedia applications themselves should be able to adapt to terminal heterogeneity, computing limitations, and varying availability of network resources [10].

5. CONCLUSION

Current trends in communication networks point to an aggregation of all kind of traffic (data, voice, etc.) in the same transport technology. The advancement of IT that began with the Internet has converged with mobile communications, to create a new paradigm of mobile convergence. Up to now, these technological advancements have centered around intelligent, multifunction terminals, and these all-in-one terminals have lead advancements in network convergence to a limited degree. The challenge of providing Quality-of-Service is an open problem and remains relatively uncharted territory [6]. Providing a complete QoS solution mobile environment requires the interaction and cooperation of several components. The emergence of real-time applications and the widespread use of wireless and mobile devices have generated the need to provide quality-of-service (QoS) support in wireless and mobile networking environments. This paper provides a survey of current research concerned with the problem of providing QoS in a mobile network architectures.

REFERENCES

[1] J.Z. Sun at all: Features in future: 4G Vision From a Technical Perspective, in IEEE 0-7803-7208-5/01, 2001.
[2] Ľ. Doboš, J. Dúha, S. Marchevský, V. Wieser.: Mobilné rádiové siete. (Mobile Radio Networks). EDIS, Žilinská univerzita v Žiline, December 2002. ISBN 80-7100-936-9 (in Slovak)
[3] Ľ. Doboš, F. Jakab: Multimédia v mobilných sieťach, Elfa. S.r.o., Košice 2005, 185 strán. ISBN 80-8086-000-9
[4] Ľ. Doboš, F. Jakab: Multimadia Delivery in Mobile Environments. Elfa. s.r.o., Košice 2005. 133 strán. ISBN 80-89066-98-4
[6] S. Maniatis at all: End-to-end Quality of Service Issues Over Next Generation Mobile Internet. In Communications and Vehicular Technology.SCVT-200, Symposium, 19. Oct 2000, pp.150-154.
[7] 3GPP. 3rd Generation Partnership Project: Technical Specification Group Services and System Aspects; Quality of Service (QoS) concept and architecture, TS 23.107 Version 5.9.0, (2003–6), 2003.
[8] L. Chen and W. Heinzelman: Network Architecture to Support QoS in Mobile Ad Hoc Networks, in Technical Report of University of Rochester, 2003.
[9] A. Lindgren: Evaluation of Quality of Service Schemes for IEEE 802.11 Wireless LANs, Proc. 26th Annual IEEE Conf. Local Comp. Nets., Tampa, FL, Nov. 2001.
[10] A. Čižmár, Ľ. Doboš, R. Palitefka: Mobile ATM Networks. Mobilné ATM siete. Elfa 1999, ISBN 80-88964-08-3, pp.167.

BIOGRAPHY

Ján Papaj was born in Liptovský Mildláš, Slovakia in 1977. In 2001 he graduated at the Department of Computers and Informatics of the Faculty of Electrical Engineering and Informatics at Technical University in Košice. Since 2001 he is a external PhD. student at the faculty. His research interests are in the fields of dialogue and multimodal systems.