Supporting UMTS QoS in Wireless LANs

Nikos Passas[1], Dimitris Skyrianoglou[1], and Apostolis K. Salkintzis[2]

[1] University of Athens
Department of Informatics & Telecommunications
Communication Networks Laboratory
Panepistimiopolis, 15764, Athens, Greece
{passas, dimiski}@di.uoa.gr
[2] Motorola
Global Telecom Solutions Section
32 Kifissias Av., 15125, Athens, Greece
salki@motorola.com

Abstract. This paper focuses on how the UMTS Quality of Service (QoS) architecture can be supported in Wireless LANs (WLANs), deployed as alternative access networks to the UMTS core. More specifically, the paper addresses the provision of UMTS QoS over WLANs and proposes a solution based on the introduction of the Wireless Adaptation Layer (WAL). The WAL provides means for extending the UMTS QoS in WLAN environments, independently of the underlying WLAN technology. Moreover, the WAL configuration to support the provisioning of UMTS QoS is discussed. In this context, a mapping between UMTS QoS attributes and WAL modules is proposed and justified.

1 Introduction

The Universal Mobile Telecommunications System (UMTS) is expected to allow the operators and service providers to extend the set of available applications offered to the user; nevertheless it will not be capable of supporting all the advanced bandwidth-demanding applications. Even in indoor environments, where a 2 Mbps channel will be offered, applications such as high-quality video will not always be feasible. Users will have to wait for the fourth generation systems (4G), which promise to increase the available bandwidth by offering a variety of broadband access systems. One way to extend the set of applications offered by UMTS to hot-spots and improve the QoS, is the interworking of UMTS with wireless local area networks (WLANs). Technology in the area of WLANs has been evolved during the last few years to offer efficient communications at rates in the range of tens of Mbps. Two of the major standards for WLANs today are the IEEE 802.11 [1] and the ETSI HIPERLAN/2 [2].

WLAN technologies can complement rather than compete with UMTS systems, offering alternative high performance access systems. UMTS/WLAN interworking can combine the wide-area benefits of second- and third- generation mobile systems, including unlimited roaming and mobility, with the additional throughput and capacity of WLANs in strategic local areas. An essential requirement for UMTS/WLAN tight interworking [3] concerns the compatibility of the QoS offered

by the two systems. Different WLANs technologies offer different capabilities for QoS, making it difficult to guarantee the QoS agreed between the user and the UMTS network. In this paper, we address this requirement by introducing an intermediate layer that accepts the UMTS user profiles, as expressed by specific QoS attributes, and performs certain actions in order to guarantee the committed QoS when a WLAN is used as a UMTS access network.

The paper is organized as follows. In section 2, a brief outline of the QoS defined for UMTS is presented. Section 3 discusses issues and alternatives for UMTS/WLAN interworking. In section 4, a solution for QoS interworking is described, focusing on traffic classes and attributes mapping. Finally, section 5 presents our conclusions.

2   QoS in UMTS

The QoS architecture in UMTS is based on the concept of *Bearer Service*, following a layer-based approach. As stated in [4], in order to realize a certain level of QoS, a Bearer Service with clearly defined characteristics and functionality has to be set up from the source to the destination of the service. The QoS architecture of UMTS follows a layer-based approach. Each layer consists of one or more bearer services that are concatenated in order to provide the desired QoS provision on an end-to-end basis (see Figure 1).

The UMTS Bearer Service provides QoS support within the UMTS network, i.e., from the Mobile Terminal (MT) to the Core Network (CN) Gateway. UMTS Bearer Service is further divided into the *Radio Access Bearer Service*, which provides confidential transport between MT and CN Iu Edge Node, and the *Core Network Bearer Service*, which connects the CN Iu Edge Node with the CN Gateway and through this with the external network.

2.1   UMTS QoS Classes and Attributes

Four different QoS classes are defined for the UMTS bearer service, taking into account the nature of the traffic produced by the various applications that are used over UMTS, as well as the restrictions and characteristics of the wireless interface.

The four classes are defined as follows [4]:

- **Conversational Class**. It includes real-time services such as speech and videoconferencing. Services belonging to the conversational class have strict transfer delay constraints.
- **Streaming Class**. This class applies to services where a stream of real-time data is destined to a user (e.g., video or audio streaming). The flow of data is unidirectional while the destination is usually a human. In contrast to the conversational class, delay constraints are not very strict, but the need for preserving the time relation (delay variation) remains important.
- **Interactive Class**. Applications such as web browsing or database retrieval fall into this class. There is an interactive scheme involved with the end-user requesting data from a remote host or equipment. In this case, a key aspect is the round-trip delay, which should be kept within reasonable limits.
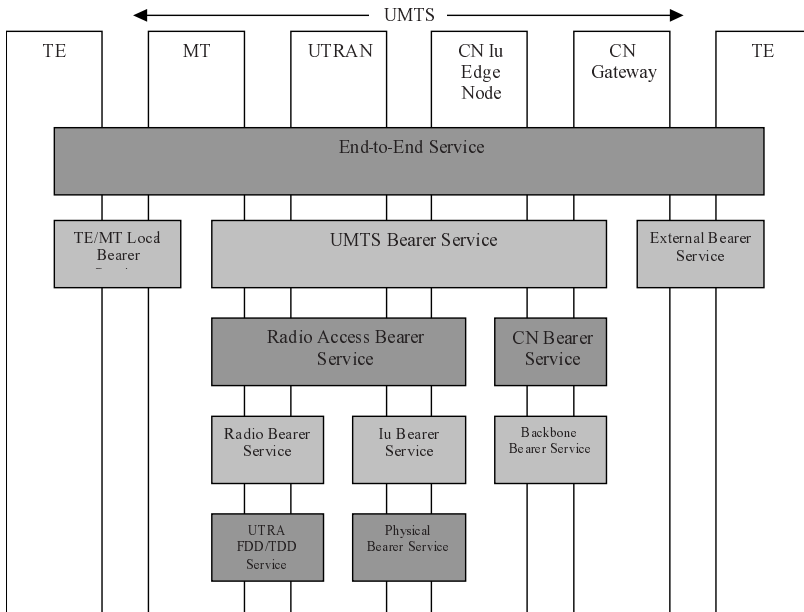
UMTS

TE	MT	UTRAN	CN Iu Edge Node	CN Gateway	TE

End-to-End Service

TE/MT Local Bearer Service

UMTS Bearer Service

External Bearer Service

Radio Access Bearer Service

CN Bearer Service

Radio Bearer Service

Iu Bearer Service

Backbone Bearer Service

UTRA FDD/TDD Service

Physical Bearer Service

Fig. 1. UMTS QoS Architecture

- **Background Class**. This class includes services such as e-mail, or file downloading. All these services are characterized by the fact that the end-user is expecting the data within a large period of time, therefore time constraints do not normally apply. However the content of the data packets should be preserved so the bit error rate should be kept low.

Figure 2 presents the QoS Information Element of UMTS Release 99, used to specify the QoS attributes for UMTS Bearer Service [4]. More specifically, the QoS Information Element includes the following set of QoS attributes:

- **Traffic Class:** It defines the type of application (class) for which the bearer service is optimized. UMTS defines four traffic classes, as described earlier.
- **Maximum Bit-rate:** It is the maximum number of bits delivered by UMTS and to UMTS at a particular Service Access Point (SAP) within a certain period of time divided by the duration of the period. The traffic is conformant with the maximum bit-rate as long as it follows a token bucket algorithm where token rate equals maximum bit-rate and bucket size equals maximum SDU size.
- **Guaranteed Bit-rate:** It is the guaranteed number of bits delivered by UMTS at a particular Service Access Point (SAP) within a certain period of time divided by the duration of the period (provided that there are data to deliver). The traffic is conformant with the guaranteed bit-rate as long as it follows a token bucket algorithm where token rate equals guaranteed bit-rate and bucket size equals Maximum SDU size.
- **Delivery Order:** Indicates whether the bearer service should provide in-sequence delivery of application data packets, referred to as Service Data Units (SDUs).
- **Maximum SDU Size:** Indicates the maximum SDU size.

8	7	6	5	4	3	2	1	
Quality of service IEI								octet 1
Length of quality of service IE								Octet 2
0 0 spare		Delay class			Reliability class			octet 3
Peak throughput				0 spare	Precedence class			octet 4
0 0 0 spare			Mean throughput					octet 5
Traffic Class			Delivery order		Delivery of erroneous SDU			Octet 6
Maximum SDU size								Octet 7
Maximum bit rate for uplink								Octet 8
Maximum bit rate for downlink								Octet 9
Residual BER				SDU error ratio				Octet 10
Transfer delay						Traffic Handling priority		Octet 11
Guaranteed bit rate for uplink								Octet 12
Guaranteed bit rate for downlink								Octet 13

Fig. 2. QoS Information Element in UMTS Release 99

- **SDU Format Information:** Defines a list of possible exact SDU sizes. This information is usually used by the UMTS radio access network in order to achieve spectral efficiency and lower transfer delay.
- **SDU Error Ratio:** It indicates the acceptable fraction of SDUs that can be lost or detected as erroneous.
- **Residual Bit Error Ratio:** Indicates the undetected bit error ratio in the delivered SDUs. If no error detection is requested, residual bit error ration indicates the bit error ratio in the delivered SDUs.
- **Delivery of Erroneous SDUs:** Indicates whether SDUs detected as erroneous shall be delivered or discarded.
- **Transfer Delay:** Defines maximum delay for the 95th percentile of the distribution of delay for all delivered SDUs during the lifetime of a bearer service.
- **Traffic Handling Priority:** Specifies the relative priority of an interactive-class bearer with respect to other interactive-class bearers.
- **Allocation/Retention Priority:** Specifies the relative importance of allocating and retaining a UMTS bearer compared to other UMTS bearers. It is used for admission control during congestion periods.

For each traffic class within UMTS a different set of QoS attributes is applicable. Table 1 shows which parameters are applicable to each traffic class.

Table 1. QoS attributes defined for each Traffic Class

Attributes	Conversational	Streaming	Interactive	Background
Maximum Bitrate	X	X	X	X
Guaranteed Bitrate	X	X		
Delivery Order	X	X	X	X
Maximum SDU Size	X	X	X	X
SDU Format Information	X	X		
SDU Error Ratio	X	X	X	X
Residual Bit Error Ratio	X	X	X	X
Delivery of erroneous SDUs	X	X	X	X
Transfer Delay	X	X		
Traffic Handling Priority			X	
Allocation/Retention Priority	X	X	X	X

3   UMTS Interworking with WLANs

Three types of interworking have been identified by ETSI [5] and the choice between them is mainly a trade-off between minimizing additional functionality and increasing the seamlessness of interworking [6]:

- **No Coupling:** The UMTS and WLAN are used as completely independent networks and the user has to sign up for different contracts in each network. Only Internet services are available to the user in the WLAN, through a standard IP router. This type of interworking enables a rapid introduction of WLAN in UMTS and is suitable for all WLAN technologies. On the other hand, it provides poor transition performance and complete separation between UMTS and Internet services.
- **Loose Coupling:** The main difference with the no coupling case is that the WLAN and UMTS networks share the same subscribers database, stored in the core UMTS network. Authentication, Authorization, and Accounting (AAA) functions in the WLAN access directly the UMTS control functions, in order to perform the

corresponding operations. On the other hand, there is complete separation between the services the terminal can access in the UMTS and the WLAN.

- **Tight Coupling:** The WLAN is considered by the core UMTS network as another access network (such as UTRAN, GERAN, etc.). This enables the QoS and security functions of UMTS to be reused for WLAN, through proper adjustments. Handovers are possible for UMTS connections, while the user only needs to subscribe to one network operator. Access to the Internet is also supported, but through the UMTS core in this case. This type of interworking offers improved transition performance and complete service interworking, but it requires full control and user data translation.

A possible configuration for a UMTS/WLAN interworking architecture is depicted in Figure 3. The WLAN consists of a number of Access Points (APs) that are interconnected through a Distribution System (e.g., an 802.3 Ethernet cable). The WLAN is connected:

- to the UMTS core network through the Inter-Working Unit (IWU), which is responsible for translating and adapting signalling (loose and tight coupling only) and data (tight coupling only) from one system to the other, and,
- to the Internet, through a standard IP router (no and loose coupling only).

A possible scenario could involve a dual mode mobile terminal that leaves the coverage area of UMTS and enters the coverage area of a WLAN (e.g., in a trade center or an airport). It would be desirable for the MT to maintain a QoS level within the coverage area of WLAN and enjoy an extended set of services, with the same or possibly improved level of QoS support, including both UMTS and Internet services.

In the next section, we focus on the tight coupling scenario, trying to provide solutions for mapping of the UMTS traffic classes and QoS attributes to specific WLAN behavior.
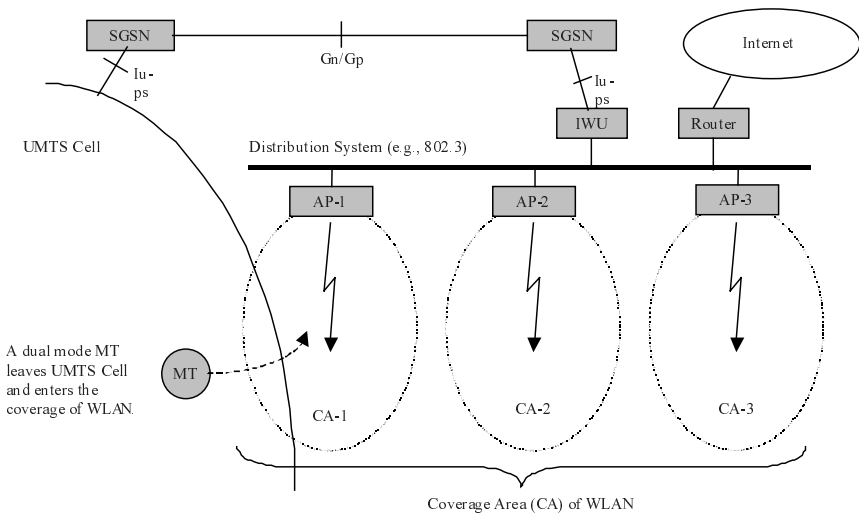

Fig. 3. UMTS-WLAN Tight-Coupling Interworking Configuration

4   The Use of the WAL in UMTS/WLAN Interworking

4.1   The Wireless Adaptation Layer (WAL)

In the communication of WLAN with both the UMTS and the Internet, QoS support is required in order to guarantee a specific end-to-end QoS for all services. Different WLAN technologies offer different or incompatible QoS capabilities, making it essential to design a unique QoS supporting scheme, applicable to all. It is clearly desirable to use the same scheme for both UMTS and Internet communications, in order to minimize complexity.

A unique QoS supporting scheme has already been proposed for WLAN communications with the Internet, in the context of the European Commission funded project WINE, referred to as the Wireless Adaptation Layer (WAL) [8]. The WAL is an intermediate layer, between the IP and the WLAN Data Link Control (DLC) layers, providing a QoS mechanism adaptable to both the capabilities of the underlying WLAN technology and QoS requirements. It provides a uniform interface to IP, while being independent of the underlying wireless network technology. The WAL incorporates a set of functional modules that can be dynamically combined and adapted to the special characteristics of the wireless link and the transport protocol. Here we extend the WAL's usage to a UMTS environment, in order to support the tight coupling interworking.

The concept of the WAL is depicted in Figure 4. The central "intelligence" of the WAL is the **WAL Coordinator**. In the downstream, it accepts data packets from the upper layers and decides on the appropriate WAL chain. In the upstream, the WAL Coordinator receives data packets from the lower layers and passes them through the same WAL chain in the reverse order.

The **QoS** module provides flow isolation and fairness guarantees through traffic shaping and scheduling. Other modules that are considered are:

- **ARQ module (ARQ)**: Can be used to improve packet error rate of non-real-time traffic (for example interactive or background) through retransmissions of erroneous packets.
- **FEC module (FEC)**: Its use is recommended for real-time-traffic (usually conversational and streaming) because it can reduce packet error rate without increasing delays like ARQ.
- **Fragmentation module (FRM)**: Can decide on the recommended fragment length based on channel conditions and traffic parameters.
- **Header Compression module (HC)**: This module can be used to reduce the overall header overhead, leading to increase of the available bandwidth.

The WAL architecture is not restrictive concerning the number or kind of modules. One of the main considerations is openness in order to introduce new modules where necessary. A detailed description of the WAL can be found in [8].

In order to be compatible with the UMTS QoS scheme, UMTS Service Bearers correspond in the WAL layer to the so-called *"WAL Bearers"*. In this way, the Radio Bearer Service of Figure 1 can be divided in two layers, the WAL layer and the standard WLAN DLC layer (e.g., 802.11 DLC). Every WAL Bearer can be treated in a different way in the WAL layer, based on the QoS for the corresponding UMTS Service Bearer and the capabilities of the standard Radio Bearer Service (i.e., WLAN
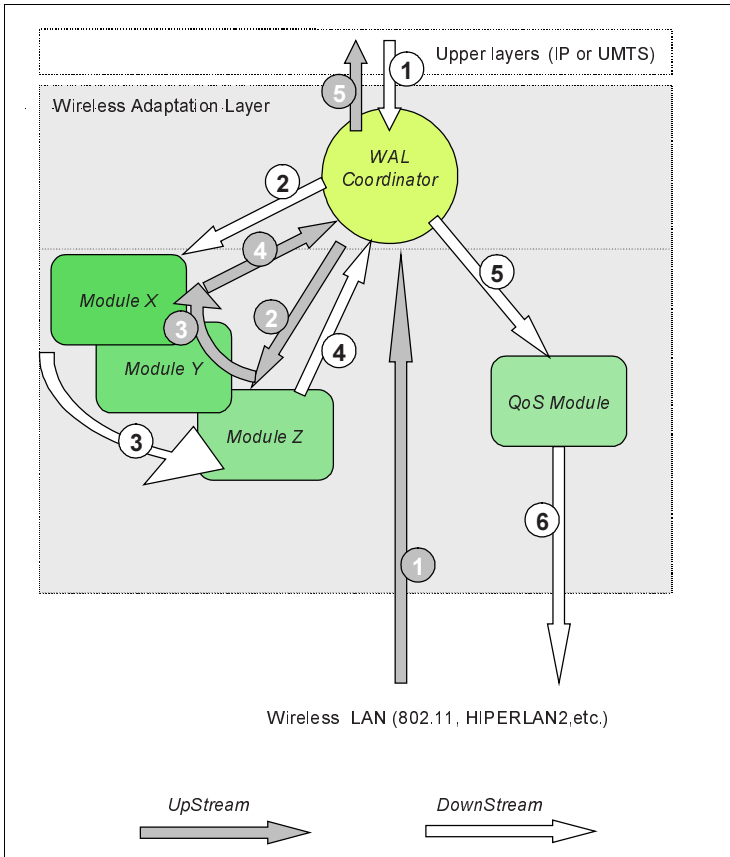
Fig. 4. The WAL design

technology). Treatment is characterised by the sequence of WAL modules X/Y/Z (referred to as a *WAL Chain*) that the packets of a WAL Bearer will pass through, as well as the specific parameters of these modules. These parameters can be adjusted dynamically by the WAL Coordinator, based on time-varying channel or traffic conditions.

4.2   Mapping of UMTS QoS Attributes to WAL Parameters

The WAL can be appropriately configured in order to be compatible with the communication of WLAN with both the UMTS (tight coupling) and Internet (no- or loose-coupling). Here we study the former case of UMTS/WLAN QoS interworking; the latter case has been investigated in [9].

Seen from the UMTS QoS architecture perspective, the WAL (together with the underlying WLAN) should implement a Radio Bearer Service. In this respect, it is necessary to map the QoS attributes of the Radio Bearer Service onto the appropriate semantics and attributes of the WAL modules. Since the WAL has a large degree of

parameterization in terms of modules and module parameters, it leaves enough space for mapping of the UMTS QoS. Below we discuss on how the UMTS QoS attributes described in section 2, can be mapped onto WAL modules. The discussion focuses on the user plane. Future work will include the procedures for configuration of the WAL bearers through the control plane, for example when a PDP context is created. Although we give a high-level approach of the mapping, it is shown that all the UMTS QoS attributes can be utilized by one or more WAL modules, and adjust their performance per bearer. The specific usage is part of each module's operation.

Traffic Class

Each UMTS traffic class is mapped to a specific WAL chain (i.e., sequence of WAL modules) that can efficiently service this class. A possible mapping can be as shown in Figure 5. The choice of the modules in each chain is based on the nature and requirements of each UMTS QoS Class, such as delay and loss requirements. Delay sensitive classes, such as Conversational or Streaming, cannot use ARQ, since this would add extra delay; they instead use FEC for error correction. ARQ is used in the not so demanding Interactive Class. Background Class is not using error control, since the service offered to it is considered as best-effort. Possible errors can be recovered end-to-end through retransmissions. Header Compression and Fragmentation are used in all classes, since they can improve the performance in the wireless medium, provided proper configuration. Additionally, the order of the modules should be chosen so that the preceding modules do not cause conflicts or degrade the operation of the succeeding ones. For example, the QoS module is always placed last in every WAL chain, so as to guarantee that the scheduled packets will not experience any further delays due to processing or queuing.

Maximum Bit-Rate

It can be used within the transmitting WAL QoS module, so as to impose constraints on the maximum transmission rate for a particular WAL Bearer. Knowing the maximum bit-rate of the bearer, the QoS module can ensure that this is never violated by bursts of packets, causing QoS degradations to other bearers. The incorporation of Maximum Bit-rate within the QoS module can be achieved with the use of a token bucket, according to the definition of this attribute in [4].

Guaranteed Bit-Rate

This attribute can be used both at the transmitting and the receiving side of a WAL Bearer. At the transmitting side, it can be mapped on a specific parameter of the QoS module that determines the mean transmission rate of the WAL Bearer that should be policed. Appropriate traffic shaping can also be performed, according to this attribute. At the receiving side, it can be used by the WAL Coordinator, in order to determine the mean bandwidth offered to a specific bearer. In case the WAL fails to provide the Guaranteed Bit-rate for a bearer service (due to excess traffic or low channel quality), the WAL Coordinator at the receiving side may request an adjustment of the parameters of the QoS module at the transmitting side, so as to provide extra bandwidth to the bearer service or change the parameters of other modules (e.g., FEC, ARQ, Fragmentation) in order to improve the performance of the wireless link.
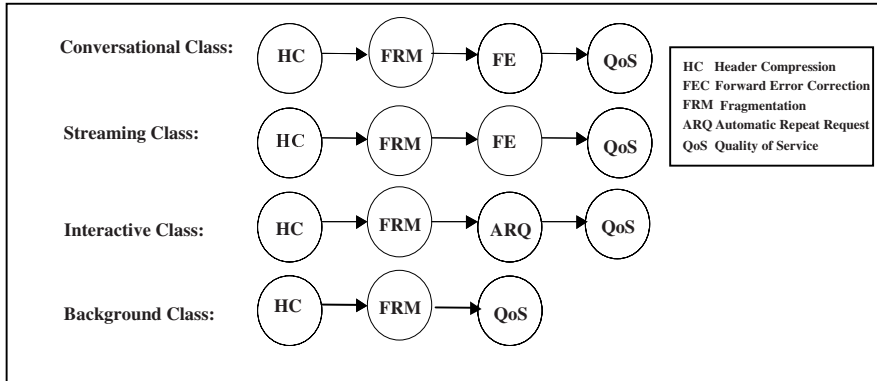
Fig. 5. Possible Module Chains for each UMTS QoS Class

Delivery Order

This attribute can be applied both to the ARQ module and the WAL Coordinator, which is responsible for delivering the packets to the upper layers. In case the Delivery Order attribute is set to "yes", either the WAL Coordinator or the ARQ module (for Interactive Class only) need to maintain a buffer where the out-of-order packets are stored until the preceding packets have been received from the lower layers and delivered to the upper layers. In case the delivery order is set to "no", the packets can be immediately forwarded to the upper layers.

Maximum SDU Size

It can be directly applied to the Fragmentation Module to set the maximum size of fragments that can be applied. The fragmentation module should adapt the fragments' size to the channel conditions without violating this parameter. When the channel is in a good state, the Fragmentation module prefers long SDU sizes, to increase the throughput and decrease the transmission overhead, but should never exceed the maximum SDU size. When the channel condition deteriorates, the Fragmentation module sets shorter SDU sizes, which are less vulnerable to channel errors. Maximum SDU size can also be used by the WAL Coordinator for admission control and traffic policing [4], and, the Header Compression module for determining the optimal compression ratio.

SDU Format Information

Similar to the Maximum SDU Size, this attribute can be utilized by the Fragmentation module and the Header Compression module, in order to optimize their operation by determining the allowed sizes of the SDUs.

SDU Error Ratio

This attribute can be used as a parameter to the FEC or the ARQ module. In order to preserve the agreed SDU Error Ratio, FEC may need to utilize a more powerful CRC scheme, by using longer error correction codes. The ARQ module may have to decide to adapt the parameters of the ARQ scheme (e.g., increase the number of allowed

retransmissions, or enlarge the window size for a specific bearer), in order to maintain error ratio below this value. This attribute can also be utilized by the WAL Coordinator as a threshold for triggering an adjustment of the WAL module parameters.

Residual Bit Error Ratio

Since bit error detection/correction is performed within the FEC module, this attribute can be used to select the appropriate error code. It determines the "undetectable" errors and therefore corresponds to the efficiency of error detecting schemes. As in the SDU Error Ratio case, the Residual Bit Error Ratio can also be used by the WAL Coordinator, in order to trigger adjustment of WAL module parameters.

Delivery of Erroneous SDUs

This attribute can be used by the FEC module or the WAL Coordinator, to decide on how to treat erroneous SDUs. FEC can mark the erroneous SDUs while the WAL Coordinator may decide whether they are forwarded to the upper layers or discarded, based on this attribute.

Transfer Delay

It can be directly used by the QoS module to schedule and prioritize the transmission of SDUs, so as not to violate this tolerance. It can also be utilized by the ARQ module in order to determine the number of retransmissions that are allowed for a specific SDU before it exceeds its delay deadline or set the window size and the duration of the window reset period. SDUs that do not succeed to maintain this value should be discarded.

Traffic Handling Priority

This attribute can be used by the QoS module during scheduling for determining the priority of packets belonging to different WAL bearers. According to [4], it is used as an alternative to absolute guarantees (e.g., Transfer Delay) and thus the two attribute types (relative/absolute) cannot be used together for a single bearer. Based on relative priorities, this attribute may lead to simpler and faster scheduler implementations, reducing the overall cost.

Allocation/Retention Priority

It can be used by the WAL Coordinator upon admission control and instantiation of a new WAL bearer, to determine the importance of this bearer. In situations where resources are reduced (e.g., due to channel quality deterioration), the WAL Coordinator may use this attribute to prioritize bearers with high allocation priority over bearers with lower priorities. If bearer discard is needed, lower priority bearers should be chosen first.

Table 2 shows a possible mapping of all UMTS QoS attributes to WAL modules, based on the above discussion. The specific use of the attributes is part of the operation of the modules.

Table 2. Mapping of UMTS attributes to WAL modules

WAL Modules	WAL Coordinator	ARQ	FEC	Header Compression	Fragment ation	QoS
Maximum Bitrate	X					X
Guaranteed Bitrate	X					X
Delivery Order	X	X				
Maximum SDU Size	X			X	X	
SDU Format Information				X	X	
SDU Error Ratio	X	X	X			
Residual Bit Error Ratio	X		X			
Delivery of erroneous SDUs	X		X			
Transfer Delay		X				X
Traffic Handling Priority						X
Allocation/Retention Priority	X					

5   Conclusions

In this paper, we discussed a framework for QoS interworking between UMTS and WLANs. The approach presented here is based on the WAL, an intermediate shim layer between the upper layers and the WLAN, aiming at improving its performance.

The WAL features a modular architecture that can be easily extended and adapted to meet both the different requirements from the upper layers and the different QoS capabilities of the lower layers. Moreover, the WAL operates transparently to both the upper and lower layers, thus it does not require any modifications or enhancements of the existing technologies. It has also the ability to adapt to the changing channel conditions by the dynamic adaptation of the WAL module parameters. The paper presented a mapping of all the UMTS QoS attributes, as included in the QoS Information Element of UMTS, to specific WAL modules, in order to parameterize their operation per bearer. It was shown that every QoS attribute can be utilized by one or more WAL modules.

Conclusively, the WAL appears as a generic and flexible solution for the UMTS/WLAN QoS interworking. Future work will describe in more detail the operation of each WAL module, based on UMTS QoS attributes, and evaluate the performance of the WAL against no-WAL architectures.

References

1. "IEEE Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications", P802.11, Nov. 1997.
2. ETSI TS 101 761-1 V1.2.1 (2000-11) "Broadband Radio Access Networks (BRAN); HIPERLAN Type2; Data Link Control (DLC) Layer; Part 1: Basic Data Transport Functions", available at http://www.etsi.org.

3. Salkintzis et al., "WLAN-GPRS Integration for Next Generation Mobile Data Networks," IEEE Wireless Communications, vol. 9, no. 5, pp. 112–124, Oct. 2002.

4. 3GPP TS 23.107 V5.5.0, "Universal Mobile Telecommunication System (UMTS; QoS Concept and Architecture", 2002, available at http://www.3gpp.org.

5. ETSI TR 101 957 V1.1.1, "Broadband Radio Access Networks (BRAN); HIPERLAN Type2; Requirements and Architectures for Interworking between HIPERLAN/2 and 3rd Gebration Celluler Systems", 2001, available at http://www.etsi.org.

6. J. Berkwist, J. Engren, O. Eriksson "WLAN as a Complement to UMTS", Master Thesis, available at http://www.s3.kth.se/~matthias/current/Ericsson.pdf.

7. 3GPP, "Universal Mobile Telecommunication System (UMTS; Core Network Protocols – Stage 3", TS 24.008 v.3.11.0, 1999, available at http://www.3gpp.org.

8. P. Mähönen, T. Saarinen, N. Passas, et. al, "Platform-Independent IP Transmission over Wireless Networks: The WINE Approach", IEEE Personal Communications Mag. Vol.8 No.6, Dec. 2001.

9. P. Mähönen, T. Saarinen, N. Passas, et. al, "Platform-Independent IP Transmission over Wireless Networks: The WINE Approach", IEEE Personal Communications Mag. Vol.8 No.6, Dec. 2001.

10. D. Skyrianoglou, N. Passas, and S. Kampouridou, "A DiffServ-based Classification Scheme for Internet Traffic Over Wireless Links", in Proc. ICWLHN 2001, Singapore, Dec. 2001.