A Service Model to Provide Quality of Service in Wireless Networks Focusing on Usability

Jörg DIEDERICH (1) and Martina Zitterbart (2)
(1) Institute of Operating Systems and Computer Networks, Technical University of Braunschweig, Germany
Email: dieder@ibr.cs.tu-bs.de
(2) Institute of Telematics, Technical University of Karlsruhe (T.H.), Germany
Email: zit@tm.uka.de

Abstract:
QoS has been a widely explored area of research in recent years. A remaining problem in this area is to create a service model which is easy to use for application programmers or service users. This is an important problem since an overly complex service model with too many QoS parameters to configure is likely not accepted by a user or deployed by a service provider. In wireless mobile networks, this problem becomes even more complex since, for example, the occurance of handoffs render the deployment of QoS more difficult. This paper proposes a service model especially suited for wireless mobile networks. The service model is derived from the currently discussed Differentiated Services approach.

1 Introduction

In recent years, much research has been done in the area of Quality of Service (QoS), especially on the technical level. As an example, the Differentiated Services (DiffServ) approach (Nichols et al. 1998) has been developed within the Internet Engineering Task Force with a special focus on scalability and incremental deployment in order to move QoS from the research labs out into existing networks. A necessary component in a QoS-enabled network is the *service model* which contains the provided services and their QoS parameters. A problem of many proposed service models is their high complexity. Therefore, they are not easy to use for application programmers or the users of the applications. Instead, service models should hide most of the complexity of the technical implementation of QoS and should focus on *usability*. This is already difficult in the context of fixed networks. In case of mobile networks, this becomes additionally complicated since two further aspects have to be considered: First, if a mobile terminal is handed off from one base station to another base station due to terminal mobility, the sessions on this mobile terminal may have to be interrupted due to insufficient resources at the new base station. Such a handoff resource shortage may occur if there are insufficient resources available at the new base station to support the session. Second, the efficiency of resource utilization plays an important role in mobile networks since the available network resources are mostly a magnitude smaller compared to fixed networks.

2 Related Work

Existing proposals for service models can be divided according to the taxonomy in Figure 1 (Diederich 2003).

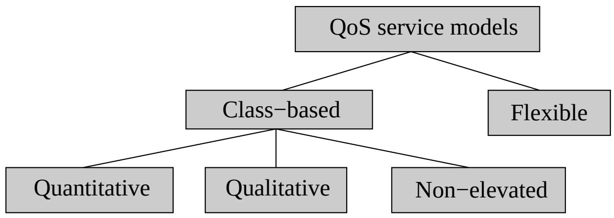

Figure 1: **A service model taxonomy**

At first, service models can be either flexible or class-based. *Flexible service models* (Schmitt & Wolf 1997) represent the service user's wishes who wants the network to adapt to the application's QoS needs. Hence, a flexible service model contains as many QoS parameters as possible for individual customization by the user. This way, flexible service models impose no limits on the applications supported by the underlying network. However, a drawback is that these models are mostly very difficult to configure for an application programmer or a service user.

In contrast, *class-based service models*, such as the legacy DiffServ service model (Zhang et al. 1999), provide only few service classes, each with a limited number of QoS parameters to be tuned individually. In this case, each service class supports only applications with certain QoS requirements, for example, applications requiring a low delay.

Class-based service models can consist of quantitative, qualitative or non-elevated services. Quantitative services give absolute values for assurances on at least a single QoS parameter. They are intended for applications with strict requirements on the respective QoS parameters. For example, mobile telephony requires strict assurances on the delay and the delay jitter to work properly. The services from the legacy DiffServ service model as well as services from the IntServ service model (Braden et al. 1994) belong to the quantitative services. The IntServ service model has many configurable QoS parameters (five per service) which are not intuitively understandable by non-experts. In contrast, the legacy DiffServ service model has a low complexity and is easier to configure. However, there are no means to avoid handoff resource shortages.

Qualitative service models provide services which do not give a strict assurance on QoS parameters. Instead, they describe the service quality for a QoS parameter either qualitatively (e.g., a 'low delay') or in relation to other service classes as in the Proportional Differentiated Services proposal (Dovrolis et al. 2002). Although these service models have only few configurable service parameters, they are also difficult to use for a service user. This is because the resulting QoS for such services is not stable over time and may vary depending on the amount of competing traffic in the network. Therefore, the configuration has to be adapted dynamically in order to enable a predictable service quality.

Non-elevated service models are similar to the qualitative ones in that they do not give strict assurances on QoS parameters. Moreover, a higher relative QoS for one QoS parameter is coupled to a lower relative QoS for a different QoS parameter. For example, a service class with a lower delay might potentially experience a higher packet loss rate at the same time (Hurley et al. 2001). Therefore, these service models are also difficult to use especially for real-time applications with strict QoS requirements although there are only very few QoS parameters to configure.

Further service models have been proposed especially for wireless mobile networks. These models belong either to the flexible models (e.g., the UMTS service model (3GP 2002)) or they contain quantitative services or qualitative services. However, the proposed models either have a very high number of parameters to specify (such as the many QoS parameters in the UMTS service model), or the parameters are difficult to provide in general (e.g., specifying the mobility profile of each mobile terminal in advance as necessary in the MRSVP approach (Talukdar et al. 1999)). Furthermore, the usage of the QoS parameters is not sufficiently specified at all (as for the 12 service classes in the ITSUMO™ approach (Cheng et al. 2000)).

Therefore, none of the proposed service models has a low complexity, a high usability *and* can deal with handoff resource shortages in a resource-efficient way.

3 The Mobile Differentiated Services Model

Within the so-called Mobile Differentiated Services QoS Model (*MoDiQ*) (Diederich et al. 2001, Diederich et al. 2003), we proposed the *MoDiQ service model*. It extends the legacy DiffServ service model to accommodate the specifics of mobile networks (cf., Fig. 2). The *MoDiQ* service model comprises the new services Mobile Premium Service, Portable Premium Service, Best-Effort Low-Delay (BELD) Service, Mobile Olympic Service, and Portable Olympic Service in addition to the legacy Best-Effort Service. It provides support for the typical applications of today's fixed and mobile networks, it can avoid handoff resource shortages with a high probability and its services can be deployed incrementally. In this paper, we focus on how the service model can be used from the viewpoint of application programmers, service users, and network providers.

3.1 Portable Premium Service

Portable Premium Service is a low-delay, low-jitter, low-loss service giving an assurance on the provided maximum bandwidth. The single configurable QoS parameter 'peak-rate' is to be configured by the application programmer (if the rate is fixed for the application) or the service user on a session request. Portable Premium Service is similar to Premium Service from the legacy DiffServ service model (Zhang et al. 1999) and can be built using the Expedited Forwarding per-hop behavior (Davie et al. 2002) although the packet delay will be higher and more variable than in fixed networks. This is because of packet retransmissions on the link layer or other means (e.g., interleaving) to compensate the higher bit error rate on wireless links.

Portable Premium Service is intended for those terminals which do not perform handoffs while being connected to the network. For example, Portable Premium Service can be used to provide a 'home zone' mobile telephony service, which gives no guarantee that the QoS of a session will be sustained if the service user moves out of reach of the corresponding 'home' base station. For service providers, such a service makes network dimensioning easier since the traffic in the mobile network will become more predictable in this case: Similar to fixed networks, the service provider can use the number of 'home zone' subscribers as a basis for dimensioning the capacity of the network.

3.2 Mobile Premium Service

Mobile Premium Service enhances Portable Premium Service in that it additionally gives an assurance that handoff resource shortages are avoided with a certain (high) probability. This assurance is based on a so-called handoff prioritization scheme which pre-reserves a certain amount of bandwidth, the so-called handoff resources. These resources can only be used by mobile terminals performing a handoff into a cell. Simulations have shown, that this probability to avoid handoff resource shortages can be above 95% for a reasonable mobility

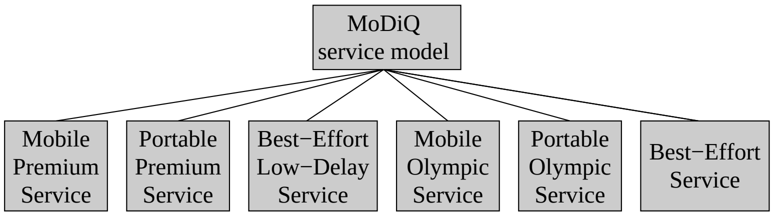

Figure 2: **The MoDiQ service model**

model (Diederich 2003). However, the basic trade-off of such a scheme is that session requests from newly emerging mobile terminals have to be blocked since they are not allowed to utilize the handoff resources. An example application in IP-based wireless mobile networks which can make use of Mobile Premium Service, is high-end mobile telephony.

3.3 Best-Effort Low-Delay Service

Best-Effort Low-Delay (BELD) Service (cf., (Diederich & Zitterbart 2003)) is a low-delay service and, therefore, related to Portable Premium Service and Mobile Premium Service. However, in contrast to the previously described two Premium Services, packet drops may occur in BELD Service for a certain period of time. One objective of BELD Service is to utilize unused Premium Service resources (i.e., Portable and Mobile) to make the resource utilization more efficient. The resource efficiency is important since the Premium Service resources are assumed to provide a large part of the operator's revenue. This is similar to the first class and business class passengers in aviation services. An example application for BELD Services is a mobile telephony application, which has the lowest costs of all low-delay services but still higher costs compared to other services with no assurances on the delay. Such a service is interesting, for example, for parents to provide a cheap mobile telephony service to their children. Because of the lower costs compared to the high-end Mobile Premium Service, the children accept (or have to accept) the varying speech quality and occasional service disruptions. The BELD Service can be implemented using the Expedited Forwarding with Dropping per-hop behavior (Diederich & Zitterbart 1999).

3.4 Portable / Mobile Olympic Service

In the legacy DiffServ service model, Olympic Service (Heinanen et al. 1999) is intended to support service differentiation for bursty data flows which is in contrast to Premium Service. Olympic Service provides no assurances on delay or jitter but an assurance on the negotiated minimal bandwidth. For example, streaming applications using a variable bit rate codec may make use of an Olympic Service class. They require a certain assurance on a minimal bandwidth but can compensate a varying delay by buffering packets on the receiver (up to a certain extent).

In the *MoDiQ* service model, Olympic Service is divided into Portable Olympic Service and Mobile Olympic Service analogously to the differentiation Portable / Mobile Premium Service. However, avoiding handoff resource shortages in Mobile Olympic Service holds only for the traffic which is within the negotiated minimal rate. One of the remaining problems in mobile networks is that the available bandwidth may vary heavily, for example, at an inter-system handoff between cells of different wireless networks such as from a Wireless LAN to a UMTS network. Therefore, the *micro-flow prioritization scheme* has been added to Olympic Service in the *MoDiQ* service model. In this approach, the service user can signal the priority of a micro-flow to the edge of the network using the DiffServ codepoint in the IP header. Introducing such user-provided priorities for micro-flows (Bouch et al. 2000) in the *MoDiQ* proposal may increase the QoS for applications with several micro-flows. As an example, consider a TV news broadcasting application with an audio and a video stream. If both streams are treated the same, both will experience the same packet loss rate under congestion. Above a certain level of congestion, this will make both streams useless. This can be avoided if the user specifies priorities for the streams. In the above news broadcasting application, the audio stream will most likely be more important than the video stream. However, this cannot be generalized and may be different in other scenarios. As an example, the video stream may be more important in a sports event than the live comments from the reporter.

4 Applicability

The *MoDiQ* service model is flexible: A single service is not only usable for a single application, as described with the above examples and summarized in Table 1.

Additionally, the combination of both, loss-sensitive and loss-insensitive services, leads to the creation of a combined service with a relative service differentiation: The low-loss service constitutes the high-priority part and

Table 1: **Single applications using single services**

Mobile service	Application class	Example application
Portable / Mobile Premium Service	QoS-sensitive applications on portable / mobile terminals with low-delay and low-loss requirements	High-end wireless IP telephony (without / with QoS-supported handoffs)
BELD Service	Loss-adaptive applications on mobile terminals with low-delay requirements which can tolerate a certain amount of packet loss	Low-cost wireless IP telephony
Portable / Mobile Olympic Service	Adaptive applications, less sensitive to delay	Audio / Video Streaming (without / with QoS-supported handoffs)
Best-Effort Service	Adaptive applications without QoS requirements	File transfer, email delivery

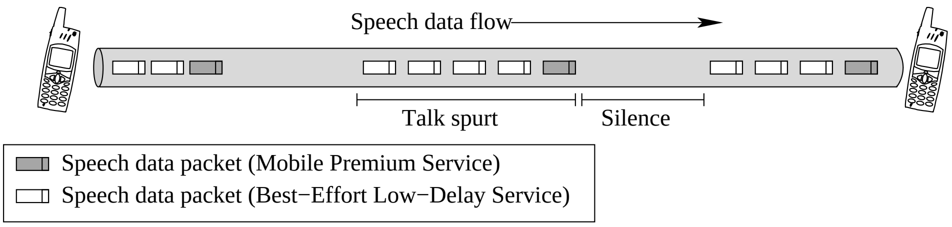

Figure 3: **An example speech service using Mobile Premium Service and BELD Service**

the service with no loss assurance the low-priority part of the resulting combined service.

As an example, current speech codecs such as the AMR codec use error concealment methods to deal with a varying bit error rate or a limited number of lost packets. These methods interpolate the speech data of a lost packet from previously processed packets. For such codecs, Sun, Wade et al. (Sun et al. 2001) have found the location of a lost packet within a speech data flow has an influence on the acoustic speech comprehensibility of mobile telephony applications. As a result, packet loss at the beginning of a talk spurt is much more disturbing than a lost packet at any other place in the speech data flow. In this context, speech applications can be supported by the *MoDiQ* service model in two different ways:

1. A high-quality speech service can use Portable / Mobile Premium Service for all speech data.

2. A low-cost speech service can make use of Portable / Mobile Premium Service for packets at the beginning of a talk spurt and BELD Service for the remaining packets (cf., Fig. 3).

In the second case, the negotiated Premium Service rate is significantly lower than in the first case. However, a remaining problem is to predict the necessary Premium Service rate to be negotiated with the mobile network.

Table 2 summarizes the above mentioned example applications if a single application utilizes loss-sensitive and loss-insensitive services simultaneously.

5 Summary

In this paper, we have described the *MoDiQ* service model, an extension of the legacy DiffServ service model, and how it can be used to provide QoS for wireless mobile networks. The overall complexity of this service model is medium: Although there are only very few QoS parameters to be configured for each service, the resulting number of services is quite high, at least higher than for the underlying legacy DiffServ service model. However, this complexity is necessary to accommodate to the specifics of wireless mobile networks, such as the occurance of handoffs, and to realize a high resource efficiency. The complexity can be decreased through incremental deployment of the *MoDiQ* service model if not all services are necessary in a particular networks. For example, if initially almost all terminals are mobile, no differentiation between the Portable and Mobile Services is necessary from the beginning: Providing the Mobile Services is sufficient initially. For each proposed service, simplicity is achieved by limiting the number of configurable QoS parameters: It is a single one for the Premium Services (the negotiated peak bandwidth) and a minimal band-

Table 2: **Single applications using multiple services**

Mobile service	Application class	Example application
Premium Service + BELD Service	Low-delay applications with two drop precedences	Speech codecs with packet loss location considerations
Olympic Service with micro-flow prioritization	Adaptive applications, less sensitive to delay, with several micro-flows	Video streaming (with audio)

width for the Olympic Services together with an optional drop precedence level to implement the micro-flow prioritization scheme. Therefore, the *MoDiQ* service model reflects a good compromise between a low complexity and the special needs of wireless mobile networks.

Acknowledgements

The authors would like to thank Thorsten Lohmar and Dr. Ralf Keller from Ericsson Eurolab Deutschland GmbH, Aachen, Germany, for their support on the initial design of *MoDiQ*.

References

3GP (2002), *QoS Concept and Architecture (Release 1999)*, Technical report, 3rd Generation Partnership Project: Technical Specification Group Services and System Aspects. URL: ftp://ftp.3gpp.org/specs/latest/23_series/23107-370.zip.

Bouch, A., Sasse, M. & DeMeer, H. (2000), *Of packets and people: A user-centered approach to Quality of Service*, *in* 'Proceedings of 9th International conference on Quality of Service (IWQoS'00)', Pittsburgh, PA, USA.

Braden, R., Clark, D. & Shenker, S. (1994), *Integrated Services in the Internet Architecture: an Overview*, Request for Comments (Informational) 1633, Internet Engineering Task Force.

Cheng, J.-C., McAuley, A., Caro, A., Baba, S., Ohba, Y. & Ramanathan, P. (2000), *QoS Architecture Based on Differentiated Services for Next Generation Wireless IP Networks*, Internet Draft, Internet Engineering Task Force. Work in progress.

Davie, B., Charny, A., Bennet, J., Benson, K., Le Boudec, J.-Y., Courtney, W., Davari, S., Firoiu, V., Kalmanek, C., Ramakrishnam, K. & Stiliadis, D. (2002), *An Expedited Forwarding PHB*, Request for Comments (Proposed Standard) 3246, Internet Engineering Task Force.

Diederich, J. (2003), *Simple and Scalable Quality of Service for Wireless Mobile Networks*, Shaker Verlag, Aachen, Germany. Doctoral thesis, University of Karlsruhe.

Diederich, J., Lohmar, T., Zitterbart, M. & Keller, R. (2001), *A QoS Model for Differentiated Services in Mobile Wireless Networks*, *in* 'Digest of the 11th IEEE Workshop on Local and Metropolitan Area Networks (LANMAN 2001)'.

Diederich, J., Wolf, L. & Zitterbart, M. (2003), *A Mobile Differentiated Services QoS Model*, *in* 'Proc. of the 3rd IEEE Workshop on Applications and Services in Wireless Networks (ASWN)', Berne, Switzerland.

Diederich, J. & Zitterbart, M. (1999), *An Expedited Forwarding with Dropping PHB*, Internet Draft, Internet Engineering Task Force. Work in progress.

Diederich, J. & Zitterbart, M. (2003), *Best-Effort Low-Delay Service*, *in* 'Proc. of the 28th IEEE Conference on Local Computer Networks (LCN)', Bonn, Germany. Accepted for publication.

Dovrolis, C., Stiliadis, D. & Ramanathan, P. (2002), '*Proportional Differentiated Services: Delay Differentiation and Packet Scheduling*', IEEE/ACM Transactions on Networking **10**(1), 12–26.

Heinanen, J., Baker, F., Weiss, W. & Wroclawski, J. (1999), *Assured Forwarding PHB Group*, Request for Comments (Proposed Standard) 2597, Internet Engineering Task Force.

Hurley, P., Le Boudec, J.-Y., Thiran, P. & Kara, M. (2001), '*ABE: Providing a Low-Delay Service within Best-Effort*', IEEE Network **15**(3), 60–69.

Nichols, K., Blake, S., Baker, F. & Black, D. (1998), *Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers*, Request for Comments (Proposed Standard) 2474, Internet Engineering Task Force.

Schmitt, J. & Wolf, L. (1997), *Quality of Service – An Overview*, Technical report, Darmstadt University of Technology.

Sun, L., Wade, G., Lines, B. & Ifeachor, E. (2001), *Impact of Packet Loss Location on Perceived Speech Quality*, *in* 'IP Telephony Workshop', New York, USA, pp. 114–122.

Talukdar, A., Badrinath, B. & Acharya, A. (1999), '*Integrated Services Packet Networks with Mobile Hosts: Architecture and Performance*', ACM Wireless Networks **5**(2), 111–124.

Zhang, L., Jacobson, V. & Nichols, K. (1999), *A Two-bit Differentiated Services Architecture for the Internet*, Request for Comments (Informational) 2638, Internet Engineering Task Force.