

A Mobile Differentiated Services QoS Model

Jörg Diederich^a, Lars Wolf^a, Martina Zitterbart^b

^aInstitute of Operating Systems and Computer Networks, Technical University of Braunschweig, Mühlenpfordtstr. 23, D-38106 Braunschweig, Germany, Email: {dieder|wolf}@ibr.cs.tu-bs.de

^bInstitute of Telematics, University of Karlsruhe (T.H.), Postfach 6980, Zirkel 2, D-76128 Karlsruhe, Germany, Email: zit@tm.uka.de

UMTS networks will be based on the Internet Protocol (IP) to provide an efficient support for applications with bursty traffic characteristics, e.g., WWW browsers. Such IP-based networks must include Quality of Service (QoS) mechanisms to enable the usage of real-time applications, such as mobile telephony. Different network services are necessary to satisfy the different needs of applications in wireless mobile networks.

To overcome this problem, this article proposes the *Mobile Differentiated Services QoS Model (MoDiQ)*, a framework to provide QoS in wireless mobile networks. *MoDiQ* is based on the Differentiated Services approach, developed by the Internet Engineering Task Force for wireline networks. An important part of *MoDiQ* is the *MoDiQ* service model which comprises separate services for mobile terminals and non-mobile terminals. This can enhance the efficiency of resource utilization in scenarios where not all mobile terminals are actually moving.

Keywords: Quality of Service, mobile communication, Differentiated Services

1. Introduction

A recent trend in mobile communication is the integration of the Internet Protocol (IP) into upcoming cellular mobile networks, e.g., the Universal Mobile Telecommunications System (UMTS). In such a packet-switched cellular mobile network, applications with bursty data flows, for example, WWW browsers or video coders with a variable bit rate, can share network resources more efficiently than in traditional circuit-switched mobile networks.

To fulfill the requirements of real-time applications, IP-based networks must provide certain assurances on the IP Quality of Service (QoS) parameters such as bandwidth, delay, delay jitter, and packet loss. Current speech coders, for example, require a minimum bandwidth of 4.75 kbit/s for an acceptable speech quality of a telephony session [2], so they cannot use adaptation means if this minimum bandwidth is no longer available. Thus, additional network components to assure the QoS parameters must be integrated into IP-based cellular mobile networks. This comprises

suitable packet scheduling schemes, QoS signaling, and resource management. Basically, this allows for an emulation of a circuit-switched network over a packet-switched IP-based network. This way, the characteristics of circuit-switched networks (e.g., assurances on the QoS parameters such as bandwidth) and the characteristics of IP-based networks, such as a higher resource efficiency, can be combined into a single QoS-enabled IP network. Moreover, QoS-enabled IP networks can provide a larger variety of QoS-enabled services than a circuit-switched network alone.

One of the main differences of *cellular* mobile networks compared to wireline networks is that a mobile terminal can change its point of attachment to the network during an ongoing communication session. This phenomenon, known as a *handoff*, can lead to a resource shortage which means that the negotiated bandwidth for a session is no longer available after a handoff. In such a case of *handoff resource shortage*, the communication session must be terminated or, in case of adaptive applications, a re-negotiation of the network-provided QoS with the application may

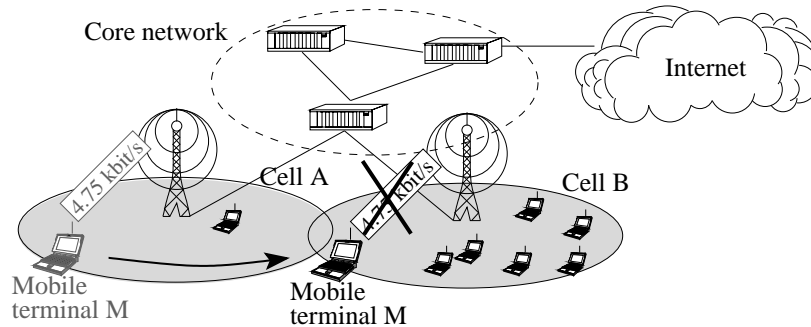


Figure 1. Handoff resource shortage in a cellular mobile network

take place. This constitutes a major problem because the user of a mobile network, in general, expects a QoS-enabled application to work steadily over the lifetime of the communication session. An example for a handoff resource shortage in a cellular mobile network is shown in Figure 1. The mobile terminal M initially resides in the lightly loaded cell A and has successfully requested a telephony session with a bandwidth of 4.75 kbit/s from the mobile network provider. (This bandwidth is a minimum requirement only to ensure a minimum service quality for the session. The mobile terminal M might adapt using more bandwidth if available, for example, if it is currently communicating using a Wireless LAN network.) If this mobile terminal M needs to perform a handoff to the heavily loaded cell B, a handoff resource shortage occurs. The network cannot continue to provide the negotiated bandwidth since the available bandwidth in cell B is completely utilized by other mobile terminals.

To accommodate handoff resource shortages, a mobility-specific QoS parameter must be considered: the handoff success probability. Providing *assurances on the handoff success probability* is crucial, especially for future cellular mobile networks where the cell size is decreased to accommodate more mobile terminals in a given geographical area. In this case, the number of handoffs per session and, thus, the probability for a handoff resource shortage can become high even

if the mobile terminal moves with only a moderate speed.

Furthermore, a specific service model for this wireless mobile network scenario is needed. Such a service model comprises the services, provided by the network, and the corresponding QoS parameters. It determines what kind of applications the network can support and which QoS parameters are available for configuration by the service user. An example for service is the Premium Service, known from the Differentiated Services (DiffServ) approach [24]. Its characteristics are similar to a leased line, i.e., the data is forwarded with low delay, low delay jitter, and with almost no packet loss. The only QoS parameter to be negotiated between the service user and the service provider is the peak rate of the data stream. This service is well-suited for interactive real-time applications such as mobile telephony.

Hence, a framework to provide QoS in wireless mobile networks, the so-called Mobile Differentiated Services QoS Model (*MoDiQ*) [11], is proposed. It consists of three major parts:

- A data forwarding plane based on the Differentiated Services approach.
- A control plane providing assurances on the handoff success probability and the necessary signaling protocols.
- A service model especially suited for usage in wireless mobile networks.

This paper is focused on the *MoDiQ service model* which extends the legacy DiffServ service model [35] to accommodate the specifics of mobile networks. It contains 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. The *MoDiQ* service model provides support for the typical applications of today's wireline and mobile networks and it can give assurances on the handoff success probability if necessary.

This paper is organized as follows: Section 2 lists the requirements on a service model to be applicable in future mobile networks. The *MoDiQ* service model is described in detail in Section 3, followed by a brief simulative evaluation in Section 4. Section 5 concludes the paper and outlines future work.

2. Requirements on Service Models in Mobile Networks

A service model interfaces the QoS-enabled network to the QoS-enabled application. Those QoS parameters which the service model allows to configure, determine how the application can make use of each particular service within the service model. Thus, the usability of a service model is mostly determined by its *simplicity* [28,19,13,27]. Although a service model should include few services only for simplicity reasons, it must support *typical applications* of today's networks. Additionally, it should be possible to enhance the service model to add support for applications in the future.

Handoff constitutes a major difference between wireline networks and mobile networks with regard to QoS. Two different approaches exist to deal with the handoff resource shortage problem: Either the mobile terminal adapts to the varying availability of resources or the network tries to provide *assurances on the handoff success probability* for this mobile terminal. Both approaches have their advantages and drawbacks and will likely coexist in the following way: Applications on mobile terminals can have different operation modes depending on the available QoS. For ex-

ample, the Adaptive MultiRate (AMR) speech codec [2] can operate on different QoS levels with a consumed bandwidth varying from 4.75 kbit/s to 12.2 kbit/s. Within this interval, the codec can adapt the speech quality according to the available service quality. However, below the lower boundary of the bandwidth interval, the codec is no longer able to produce sufficiently legible speech quality. Thus, application adaptation alone is not sufficient to handle QoS variations [7]. The network has to provide QoS assurances with regard to the minimum requirement even in case of adaptive applications in order to support seamless mobile communication.

Reserving resources exclusively for handoffs is a well-known method to provide assurances on the handoff success probability. One of the first proposals, the so-called "Guard channel" scheme [17] uses a static amount of guard bandwidth to be reserved for handoffs exclusively. However, this simple approach can become highly inefficient in networks with highly dynamic mobility patterns, where the traffic load in each cell can vary drastically. These dynamically changing mobility patterns are especially important for networks with small cell sizes, where the number of handoffs is significantly higher than in large-cell networks. Thus, further schemes propose a dynamically adjustable amount of such a guard bandwidth to accommodate variations in the mobility patterns occurring in networks with smaller cells [22,8]. The basic drawback of all schemes is a lower *efficiency* of network utilization compared to networks without such a handoff resource reservation. Thus, it is important to provide assurances on the handoff success probability only if necessary, i.e., for those mobile terminals which really perform handoff. Therefore, a service model suited for mobile networks should provide separate services for *mobile* terminals, which potentially perform handoff, and *portable* terminals which do not perform handoff [5,23,21]. This way, the *efficiency* of resource utilization can be significantly increased due to the reduction of the necessary guard bandwidth. We do not consider the case of vertical handoffs [25] here, where even portable terminals may perform a handoff from one mobile network, for example a UMTS

network, to another mobile network, such as a WirelessLAN, without moving at all.

In summary, *simplicity*, support for *typical applications*, *efficiency*, *support for assurances on handoff success probabilities*, and providing separate services for *mobile/portable* terminals are considered the most important requirements on service models in mobile networks.

Available proposals for service models have at least one of the following problems when used for wireless mobile networks:

- They comprise many configurable QoS parameters which makes the *configuration* of a communication session difficult. Examples are the Guaranteed Service [29] and the Controlled-Load Service [34] from the Integrated Services approach [7]. Furthermore, in the Mobile Integrated Services approach [31], the mobility pattern of each mobile terminal has to be specified in advance at session startup.
- They do not support the above mentioned *typical applications* properly. For example, services with only qualitative service assurances [12,27] are not able to provide assurances on the minimum bandwidth which is necessary for certain real-time applications such as mobile telephony. This is also true for so-called ‘non-elevated service models’ such as Alternative Best-Effort [18] or the Equivalent Differentiated Services model [14].
- They provide no *assurances on the handoff success probability*. For example, neither the above mentioned two services within the Integrated Services approach nor the legacy DiffServ service model incorporate such assurances.
- They do not differentiate between non-mobile terminals, which do not perform handoff, and mobile terminals to increase the *efficiency* of resource utilization. This holds, for example, for the service model of the Integrated Services approach as well as for the legacy DiffServ service model.

3. The MoDiQ Service Model

The *MoDiQ* service model extends the legacy DiffServ service model, which comprises Premium Service, Olympic Service, and the traditional Best-Effort Service (cf., Fig. 2), from three to six service classes as depicted in Figure 3.

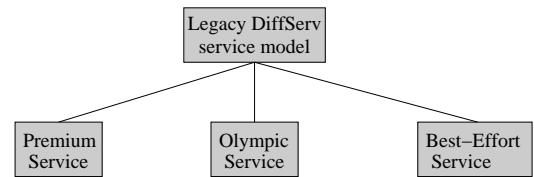


Figure 2. The legacy DiffServ service model

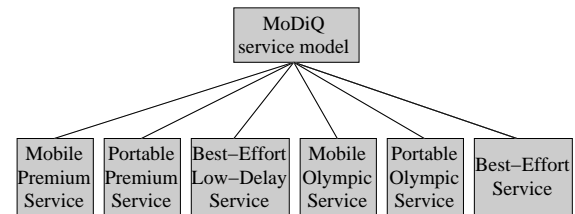


Figure 3. The *MoDiQ* service model

The legacy DiffServ service model considers only QoS sensitivity in general and delay sensitivity as a subclass of QoS sensitivity (cf., Fig. 4): The Best-Effort Service is targeted at elastic applications, Premium Service [35] at delay-sensitive real-time applications, and Olympic Service [16] at delay-insensitive applications with minimal bandwidth demands.

The *MoDiQ* service model adds support for assurances on the handoff success probability for both, Premium Service and Olympic Service. Furthermore, separate services without such an assurance are available for portable terminals (cf.,

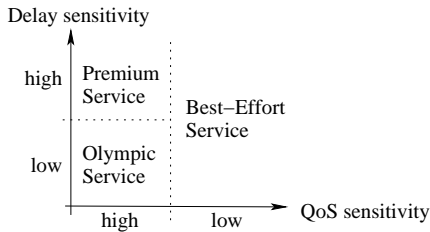


Figure 4. The legacy DiffServ service model with support for delay/QoS sensitivity

Fig. 5) to increase the efficiency of resource utilization. For this reason, the legacy Premium Ser-

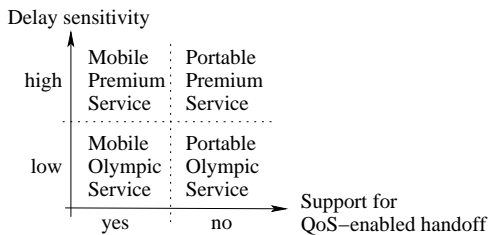


Figure 5. The *MoDiQ* service model with support for QoS-enabled handoff

vice class is divided into two parts: *Mobile Premium Service* provides low-delay packet delivery with support for assurances on the handoff success probability whereas the *Portable Premium Service* is a low-delay service with no such support. Analogously, the Olympic Service class is split into a *Mobile Olympic Service* and a *Portable Olympic Service*. It is reasonable to provide such a service on the network layer and not only on the wireless link, because a handoff resource shortage might also occur within the wired part of the mobile network, for example, on the link from the base station towards the backbone. As for the legacy Premium Service, sharing resources between active sessions within the same class is not

allowed in Mobile Premium Service or Portable Premium Service to achieve the low-delay characteristic. This is in contrast to the two Olympic Service variants where resource sharing is explicitly desired to employ statistical multiplexing and, thus, to increase resource utilization.

The *MoDiQ* service model further supports loss-sensitive and loss-insensitive applications for both types of applications, those with a high delay sensitivity and those with a low delay sensitivity (cf., Fig. 6).

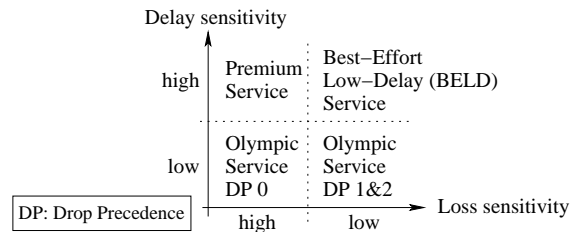


Figure 6. The *MoDiQ* service model with support for loss-sensitivity

It adds the *micro-flow prioritization approach* to the Olympic Service to support different degrees of loss-sensitivity for less delay-sensitive applications. This way, a relative service differentiation is introduced within each Olympic Service class. This is useful for QoS-sensitive applications with several micro-flows since they can indicate that one flow (e.g., a video stream) has a lower importance compared to other flows (e.g., an audio stream) [6]. The technical base for this approach is already available. In Olympic Service the drop precedence level can be set in the DiffServ codepoint of each packet to signal the priority of a packet to intermediate nodes. However, the basic intention of Olympic Service is that traffic conditioning decides about the drop precedence level, not the application or the user, so the legacy DiffServ service model provides no means to specify the relative priority for a flow. Therefore, the *MoDiQ* service model includes an optional specification of a relative priority for a session as QoS

parameter for Olympic Service. In this case, the higher drop precedences of Olympic Service can be viewed as a service with packet loss while the lowest drop precedence constitutes a service (almost) without any packet loss.

For delay-sensitive and loss-tolerant applications, there is currently no service available in the legacy DiffServ service model. Thus, the MoDiQ service model contains a third low-delay service called *Best-Effort Low-Delay (BELD) Service*. This service can take favor of unused Premium Service resources to increase the network utilization. Resource sharing is, therefore, an inherent property of this service.

3.1. Portable Premium Service

Similar to Premium Service, *Portable Premium Service* has a low-delay, low-jitter, low-loss, and assured-bandwidth property. It has a single configurable QoS parameter ‘peak-rate’ which the service user must configure on a session request. Portable Premium Service is intended for portable terminals which are not forced to do handoffs due to mobility while being connected to the network. Such terminals do not need assurances on the handoff success probability, meaning that the service scope is intra-cell.

In contrast to the legacy Premium Service, the special characteristics of the wireless link have to be taken into account for Portable Premium Service [33]. Error control schemes can reduce the high bit error rate of wireless links, but they lead to a higher delay or a higher delay jitter compared to wireline networks. Thus, for both the low-delay and the low-loss service characteristics, the meaning of ‘low’ will be different, i.e., numerically higher, as for the legacy Premium Service. For example, in UMTS the lowest possible delay is planned to be within the range of 80 ms for the radio access network [1]. This article assumes that both, the residual bit error rate and the delay, are small enough for real-time applications using Portable Premium Service.

As an example, a high-end IP-based mobile telephony application without handoff support can utilize Portable Premium Service. To accommodate the wireless link characteristics, the speech data can be coded with speech codecs that

can tolerate a low percentage for the residual bit error rate without compromising the acoustic speech comprehensibility significantly (although the speech quality degrades) [30].

3.2. Mobile Premium Service

Mobile Premium Service is an enhanced Portable Premium Service where the service scope is not limited to a single cell. It provides assurances on the handoff success probability, i.e., the amount of bandwidth assigned in the old cell will be available in the new cell with a certain probability. However, this is only possible if the handoff is performed between two cells of the same radio access network (RAN), a so-called *intra-RAN* handoff. An example application for Mobile Premium Service is a high-end mobile telephony application with intra-RAN handoff support.

Mobile Premium Service can not deal with inter-RAN handoff because of the significantly different availability of resources in different radio access networks. For example, the available resources per terminal in wideband mobile networks such as Wireless LANs are up to 54 Mbit/s in an empty cell. In narrow-band networks, such as GPRS, the available bandwidth per mobile terminal is theoretically up to 171 kbit/s, but on average only 40 kbit/s. Thus, providing assurances on the handoff success probability for a handoff from Wireless LAN to GPRS is almost impossible (although it may work in the opposite direction) for those session with a high bandwidth requirement. Therefore, Mobile Premium Service can only be provided for an intra-RAN scope.

Admission Control for Mobile Premium Service

The admission control algorithm for Mobile Premium Service [9] reserves a certain amount of resources for handoff purposes only. This amount is dynamically adjusted according to the current mobility pattern. It uses information from past handoffs to predict the mobility for the near future on a per-cell level. The algorithm considers not only the resource situation in the current cell as for Portable Premium Service, but takes also the resource situation in the neighbor-

ing cells into account to provide assurances on the handoff success probabilities. This so-called *distributed admission control* is necessary because local schemes cannot give a sufficiently high assurance on the handoff success probability. In contrast to existing admission control schemes, the proposed scheme provides support for assurances on handoff success probabilities for a wide variety of traffic patterns while being easy to administrate at the same time. To enable an incremental deployment, the scheme can be deployed to the bottleneck links initially (such as the wireless links). For end-to-end QoS, a legacy DiffServ resource management (e.g., using a Bandwidth Broker) can be provided in the mobile network which has to be coordinated with the initially deployed admission control and resource management on the bottleneck link in this case.

Portable Premium Service vs. Mobile Premium Service

The advantage of Portable Premium Service is that its users have a higher probability of being admitted in a mobile network with many mobile terminals, a high number of prioritized handoffs and with cells with different resource utilization levels. As an example, consider a cell X serving a city highway and a neighboring cell Y from which the highway can be reached and which also serves some apartments (cf., Fig. 7). If the ‘highway

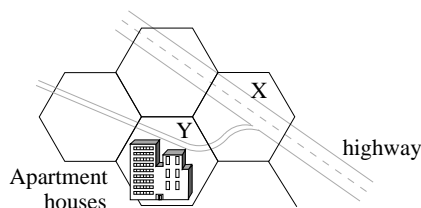


Figure 7. Example: Apartments in the neighborhood of a highway

cell X’ is highly loaded and cell Y is only lightly loaded, Mobile Premium Service requests in cell Y are denied owing to the distributed admission

control algorithm of Mobile Premium Service: It considers not only the resource situation in the current cell, but takes also the resource situation in the neighboring cells into account to provide a sufficiently high assurance on the handoff success probabilities. This is necessary to block those sessions which will enter the highway cell X and, hence, experience handoff drop with a high probability. However, portable users of the mobile network in the apartments do not want to move to the highway cell X. Therefore, Portable Premium Service enables these users to receive resources which increases the resource utilization level of low-delay resources in cell Y.

3.3. Best-Effort Low-Delay Service

Similar to Portable Premium Service and Mobile Premium Service, Best-Effort Low-Delay (BELD) Service [10] belongs to the low-delay services. In contrast to the former two low-delay services, packet drops can occur in BELD Service for a certain period of time. The introduction of BELD Service has two objectives:

1. It utilizes unused Premium Service resources (i.e., Portable and Mobile) to increase the efficiency in the utilization. These resources are assumed to provide a considerable part of the operator’s revenue.
2. It provides a low-delay, but lossy service for delay-sensitive and loss-tolerant applications. Since all low-delay applications can naturally work using Premium Service, BELD Service must be significantly cheaper to attract the interest of potential customers.

For example, a low-cost mobile telephony application based on a loss-tolerant speech codec may make use of the BELD Service. The users of such an application may accept the varying speech quality and occasional service disruptions because of a lower price compared to the high-end Mobile Premium Service. Such a low-cost service is interesting, for example, for parents to provide a cheap mobile telephony service to their children. Furthermore, BELD Service can be used for signaling purposes, where a low delay is required

and occasional packet losses can be handled by retransmissions or periodic transmissions. This is useful for periodical measurements of sensor data where some packet loss can be interpolated from the previous sensor values.

BELD Service traffic is intended to utilize unused Premium Service resources to achieve a low-delay characteristic. These resources are not utilized up to 100% in general for the following reasons:

1. To achieve a high assurance on the handoff success probability even for traffic patterns which have a low probability to occur, the necessary handoff resources to reserve are estimated rather conservatively in *MoDiQ*. Thus, some handoff resources are available for BELD Service on average, i.e., if the mobility pattern does not represent a worst-case scenario so that less handoff resources are necessary than expected.
2. The negotiated Portable/Mobile Premium Service peak rate for a single flow is rarely used up to 100%. For example, silence suppression can reduce the average necessary bandwidth of a telephony application to about 30–50% compared to a speech codec without silence suppression [3]. However, it is not possible to increase the utilization by subscribing to a lower rate than the peak rate of the codec. This would lead to substantial service disruptions in case of long speech periods which is against the ‘low loss’ property of such a ‘Premium Service’.

In the existing DiffServ approach, Best-Effort traffic consumes unused Premium Service capacity so that it is not wasted. However, Best-Effort traffic can have already experienced delay or will experience delay because of packet queuing, so it does not significantly improve the delay characteristic of Best-Effort Service. Therefore, unused Premium Service resources are assigned to BELD Service in the *MoDiQ* service model.

Packet loss in BELD Service occurs, therefore, if the sum of the Premium Service traffic and the BELD Service traffic exceeds the available Premium Service capacity. If packet loss in BELD

Service should be limited, admission control can be used to limit the amount of BELD Service traffic introduced into the network. In this case, the service user must specify a peak-rate at the beginning of a session.

3.4. Portable/Mobile Olympic Service with Micro-Flow Prioritization

In contrast to Premium Service, Olympic Service is intended to support service differentiation for bursty data flows. It provides an assurance on the negotiated minimal bandwidth but no assurances on delay or jitter. Thus, it is well-suited, for example, for streaming applications, which require a certain assurance on a minimal bandwidth and which can compensate a varying delay to a certain extent with buffers at the receiver.

Similar to the differentiation Portable/Mobile Premium Service, Olympic Service is divided into Portable Olympic Service and Mobile Olympic Service in the *MoDiQ* service model. However, the assurance on the handoff success probability in Mobile Olympic Service is only valid 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 inter-RAN handoffs. Adding the micro-flow prioritization scheme to Olympic Service, as proposed in the *MoDiQ* service model, can improve the overall QoS for applications with several micro-flows. The service user can signal the priority of a flow to the first boundary node using the DiffServ codepoint. This implicit signaling mechanism is scalable since a multi-field classification on interior nodes is not necessary to identify those flows which should be affected by packet loss preferentially. This way, user-provided priorities for the micro-flows become possible [6]. Without such a mechanism, all micro-flows from the application are affected by packet loss in the same way, which will likely make the application useless in case of congestion.

4. Simulations

MoDiQ has been evaluated extensively using the network simulator ns-2. To show the advantage of having separate services for portable ter-

minals and mobile terminals, this section examines the gain of using Portable Premium Service in addition to Mobile Premium Service.

4.1. Simulation Model

The network model consists of two parts. The wireless part is composed of nine base stations which are placed onto a rectangular grid. The distance between two base stations is 700 m horizontally and vertically which is a typical distance for mobile networks in a densely populated city area. The cell size is 800 m so the coverage areas of two neighboring base stations overlap up to 100 m to enable soft handoffs without interruptions of connectivity. The handoff control algorithm is based on a hysteresis [32] which can avoid subsequent handoffs between two base stations within a short period of time (the so-called *flip-flop effect*). The wireless network is based on the IEEE 802.11 standard to simulate realistic effects such as collisions on the air interface since it has been found to be important to model these effects to achieve realistic simulation results [15]. The base stations are interconnected via a tree-like topology (cf., Fig. 8) leading towards the root node of the backbone which itself may be connected to the Internet in a real-world scenario. The node connected to the root-node represents

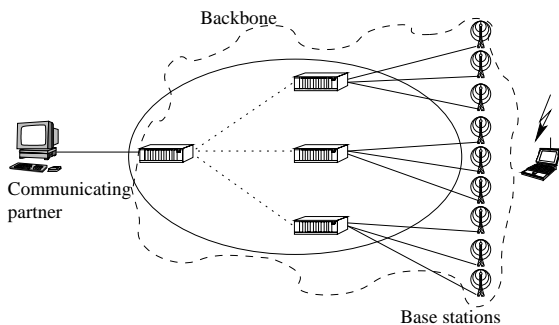


Figure 8. The network model

a node in the Internet and is the communicating partner of all mobile terminals.

The network is designed such that each base station can carry up to 100 sessions simultaneously. The session duration is modeled as an exponential distribution with a mean of 180 s , which models telephony sessions quite accurately [4].

As mobility model, the so-called *Random-Move scheme* is used. It consists of a 3×3 mobility cell scenario (with one base station in the center of each cell) where eight cells (2–9) are in use (cf., Fig. 9). One edge cell is not used to create a higher load in the center cell. In each cell, the

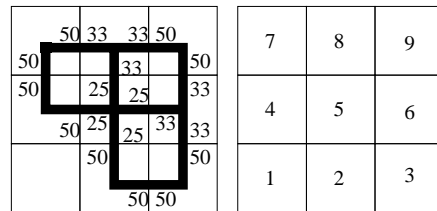


Figure 9. The Random-Move mobility pattern

probability of a mobile terminal to change to one of the neighboring cells is equal as given in Figure 9. Mobile terminals have the same probability to start in any of these eight cells, the average speed is a truncated Gaussian distribution [32] with a mean of 17 m/s (about 60 km/h) and a maximum deviation of $\pm 20\%$. The Random-Move scenario, therefore, represents a scenario with no traffic jams where the terminals can move fluently, for example, on a main-street with coordinated traffic lights. To show the effect of having portable terminals in the network, only 25% of the terminals move in this scenario, the remaining 75% do not move during the simulations (the so-called ‘Static Random-Move scenario’).

4.2. Simulation Results

We propose that Portable Premium Service is useful in areas where cells have different resource utilizations and where new session requests in a cell with free resources are rejected because of a highly loaded neighboring cell. This is true in the

Random-Move scenario where the center cell has a higher resource utilization than the surrounding cells.

Figure 10 depicts the gain of introducing Portable Premium Service in the static Random-Move scenario where 75% of the terminals are non-mobile. In the Mobile Premium Service sim-

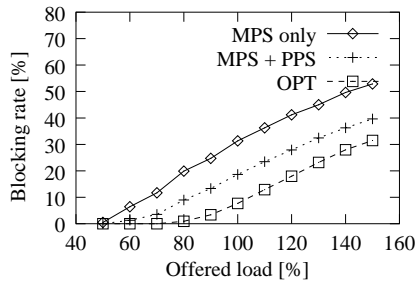


Figure 10. Portable Premium Service simulations: Blocking rate

ulations (‘MPS only’ in the figure), portable terminals request Mobile Premium Service whereas they request Portable Premium Service in the simulations with both Mobile Premium Service and Portable Premium Service (‘MPS + PPS’ in the figure). In the former case, this means that the distributed admission control scheme of Mobile Premium Service is used for all mobile terminals regardless whether they perform handoff or not. This is the only difference between the two simulation runs so that it is ensured that a change in the result comes from the introduction of Portable Premium Service. For comparison, the simulation have been performed for a theoretically optimal scheme [20] (OPT in the figure). The blocking rate for the Portable Premium Service simulations, i.e., the probability that a new session request is denied, is much closer to the optimal approach than the blocking rate of the Mobile Premium Service simulations. Depending on the offered load, the introduction of Portable Premium Service increases the number of accepted sessions up to 25%. At the same time, no ad-

ditional handoff resource shortages occur so that no sessions have to be terminated after a handoff. The introduction of Portable Premium Service can, therefore, provide a significant enhancement of the resource utilization.

5. Summary and Future Work

The Mobile Differentiated Services QoS model (*MoDiQ*) is a simple and scalable solution for providing QoS in wireless mobile networks. This article provides an overview on the *MoDiQ* service model, which is a service model especially adapted for mobile networks.

It enhances the legacy DiffServ service model with support for QoS-enabled handoffs, if necessary, and with separate services for loss-sensitive and loss-tolerant applications. QoS-enabled handoffs are supported in Mobile Premium Service and Mobile Olympic Service: These services give assurances on the handoff success probability for delay-sensitive and less delay-sensitive applications, respectively. Providing separate services without support for QoS-enabled handoff (i.e., Portable Premium Service and Portable Olympic Service) can potentially improve the resource utilization. The Best-Effort Low-Delay (BELD) Service is part of the *MoDiQ* service model for two reasons: To provide a Best-Effort-like service for loss-tolerant delay-sensitive applications, such as low-cost telephony, and to increase the utilization of low-delay resources.

Thus, the *MoDiQ* service model supports important applications in future wireless mobile networks such as mobile telephony, WWW browsing and streaming applications. In contrast to the legacy DiffServ service model, it provides support for assurances on handoff success probabilities, so the probability for handoff resource shortages can be reduced.

Future work includes several areas: The *MoDiQ* proposal is currently focused on providing assurances on the handoff success probability on the bottleneck links only. These links are presumably the wireless link or the last wired mile [26], which connects the base station to the backbone of the mobile network. To achieve end-to-end QoS assurances, it is possible to deploy

a legacy DiffServ resource management (e.g., a Bandwidth Broker) within the backbone of the mobile network. This requires a coordination of the components of *MoDiQ* with this backbone resource management. A challenge is to maintain the high scalability of *MoDiQ* in such an end-to-end scenario.

Further investigations may evaluate the usage of handoff degradation or resource re-allocation for adaptive applications which can work under different resource situations. This way, more sessions can possibly be supported simultaneously and the QoS can be enhanced for adaptive applications if sufficient resources become available. An extension of the *MoDiQ* service model with qualitative or non-elevated service classes may be reasonable in that case.

Acknowledgment

The authors would like to acknowledge Thorsten Lohmar and Dr. Ralf Keller from Ericsson Eurolab Deutschland GmbH, Aachen, Germany, for their insightful comments and discussions.

Information on the Authors



Jörg Diederich received the diploma degree in computer science in 1998 from the Technical University of Braunschweig, Germany, and the doctoral degree from University of Karlsruhe (TH), Germany, in 2002. He was a member of the

Institute of Operating Systems and Computer Networks, Technical University of Braunschweig, from February 1999 to September 2003. Since October 2003, he is a visiting associate professor at the Department of Telematic Engineering, Carlos III University of Madrid, Spain.



Lars C. Wolf received the diploma degree in computer science in 1991 from the University of Erlangen-Nuremberg, Germany, and the doctoral degree from the Technical University of Chemnitz in 1995. From

1991 to 1996 he worked at IBM's European Networking Center in Heidelberg, Ger-

many. There he contributed to distributed multimedia systems in several research and development projects. In 1996 he joined the Technical University of Darmstadt and led a research group working on multimedia networking and quality of service in distributed multimedia systems. Lars Wolf joined University of Karlsruhe (TH), Germany, in 1999 where he was professor in the computer science department and alternate director of the computer center. Since 2002 he is professor for computer science at the Technical University of Braunschweig where he is head of the Institute of Operating Systems and Computer Networks. Lars Wolf serves on several editorial boards and as chair and member of program committees of several workshops and conferences.



Martina Zitterbart is full professor in computer science at the University of Karlsruhe, Germany. She received her doctoral degree from the University of Karlsruhe in 1990.

From 1987 to 1995 she was Research Assistant at the University of Karlsruhe. From 1991–1992 she was on leave of absence as a visiting scientist at the IBM T.J. Watson Research Center, Yorktown-Height, NY. She was visiting professor at the University of Magdeburg and the University of Mannheim and full professor at the Technical University of Braunschweig (1995–2001). Her primary research interests are in the areas of multimedia communication systems, mobile and ubiquitous computing, ambient technologies as well as Elearning. She is member of the IEEE (served on the Board of Governors of the communication society 1995–1998), ACM and the German Gesellschaft für Informatik. In 2002 Martina Zitterbart received the Alcatel SEL research award on technical communication.

REFERENCES

1. QoS Concept and Architecture (Release 1999). Technical report, 3GPP: Technical Specification Group Services and System Aspects, January 2002.
2. AMR speech Codec: General description (Release 4). Technical report, 3GPP: Technical Specification Group Services and System As-

- pects, April 2000.
3. IP telephony white paper. Technical report, ACT Networks, Inc., January 1998.
URL: <http://www.aliancedatacom.com/ip-telephony/ip-telephony-white-paper.html>.
 4. BAKOM: Bundesamt für Kommunikation, Abteilung Telecomdienste. Average voice call durations in Switzerland (1998–2000). Report on Fernmeldestatistik 2000, Biel, Schweiz, November 2001. In German.
 5. V. Bharghavan, K.-W. Lee, S. Lu, S. Ha, J.R. Li, and D. Dwyer. The TIMELY Adaptive Resource Management Architecture. *IEEE Personal Communications Magazine*, 5(4), August 1998.
 6. A. Bouch, M. Sasse, and H.G. DeMeer. 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, June 2000.
 7. R. Braden, D. Clark, and S. Shenker. Integrated Services in the Internet Architecture: an Overview. RFC (Informational) 1633, IETF, June 1994.
 8. S. Choi and K.G. Shin. A comparative study of bandwidth reservation and admission control schemes in QoS-sensitive cellular networks. *ACM Wireless Networks*, 6(4):289–305, 2000.
 9. J. Diederich. *Simple and Scalable Quality of Service for Wireless Mobile Networks*. Shaker Verlag, Aachen, Germany, July 2003. Doctoral thesis, University of Karlsruhe.
 10. J. Diederich, M. Doll, and M. Zitterbart. Best-Effort Low-Delay Service. In *Proceedings of the 28th Conference on Local Computer Networks (LCN 2003)*, Bonn, Germany, October 2003. IEEE.
 11. J. Diederich, T. Lohmar, M. Zitterbart, and R. Keller. 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)*. IEEE, March 2001.
 12. C. Dovrolis, D. Stiliadis, and P. Ramanathan. Proportional Differentiated Services: Delay Differentiation and Packet Scheduling. *IEEE/ACM Transactions on Networking*, 10(1):12–26, February 2002.
 13. P. Ferguson and G. Huston. *Quality of Service: Delivering QoS on the Internet and in Corporate Networks*. Wiley & Sons, New York, USA, January 1998.
 14. B. Gaidioz and P. Primet. The Equivalent Differentiated Services Model. Research Report 2002-09, RESAM, INRIA, France, February 2002.
 15. J. Heidemann, N. Bulusu, J. Elson, C. Intanagonwiwat, K. Lan, Y. Xu, W. Ye, D. Estrin, and R. Govindan. Effects of detail in wireless network simulation. In *Proc. of the SCS Multiconference on Distributed Simulation*, pages 3–11, Phoenix, USA, January 2001.
 16. J. Heinanen, F. Baker, W. Weiss, and J. Wroclawski. Assured Forwarding PHB Group. RFC (Proposed Standard) 2597, IETF, June 1999.
 17. D. Hong and S.S. Rappaport. Traffic Model and Performance Analysis for Cellular Mobile Radio Telephone Systems with Prioritized and Nonprioritized Handoff Procedures. *IEEE Transactions on Vehicular Technology*, 35(3):77–92, August 1986. See also: CEAS Technical Report No. 773, June 1, 1999, College of Engineering and Applied Sciences, State University of New York, Stony Brook, NY 11794, USA.
 18. P. Hurley, J.-Y. Le Boudec, P. Thiran, and M. Kara. ABE: Providing a Low-Delay Service within Best-Effort. *IEEE Network*, 15(3):60–69, May 2001.
 19. G. Huston. Next Steps for the IP QoS Architecture. RFC (Informational) 2990, IETF, November 2000.
 20. R. Jain and E. Knightly. A Framework for Design and Evaluation of Admission Control Algorithms in Multi-Service Mobile Networks. In *Proc. of the IEEE Conference on Computer Communications (IEEE Infocom)*, New York, March 1999.
 21. S. Jiang, B. Li, X. Luo, and D.H.K. Tsang. A Modified Call Admission Control Scheme and Its Performance. *ACM Wireless Networks*, 7(2):127–138, March 2001.

22. D.A. Levine, I.F. Akyildiz, and M. Naghshineh. A Resource Estimation and Call Admission Algorithm for Wireless Multimedia Networks Using the Shadow Cluster Concept. *IEEE/ACM Transactions on Networking*, 5(1):1–12, February 1997.
23. S. Lu, K.-W. Lee, and V. Bharghavan. Adaptive Service in Mobile Computing Environments. In *Proc. of the International Workshop on Quality of Service (IWQoS)*, pages 25–36, New York, USA, May 1997.
24. K. Nichols, S. Blake, F. Baker, and D. Black. Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers. RFC (Proposed Standard) 2474, IETF, December 1998.
25. K. Pahlavan, P. Krishnamurty, A. Hatami, M. Ylianttila, J.-P. Makela, R. Pichna, and J. Vallström. Handoff in hybrid mobile data networks. *IEEE Personal Communications Magazine*, pages 34–47, April 2000.
26. D. Partain, G. Karagiannis, P. Wallentin, and L. Westberg. Resource Reservation Issues in Cellular Radio Access Networks. Internet Draft, IETF, June 2002. Work in progress.
27. J. Ruutu and K. Kilkki. Simple Integrated Media Access – a Comprehensive Service for Future Internet. In *Proc. of the IFIP Conference on Performance of Information and Communications Systems (PICS)*, Lund, Sweden, May 1998.
28. J. Schmitt and L. Wolf. Quality of Service – An Overview. Technical report, Darmstadt University of Technology, April 1997.
29. S. Shenker, C. Partridge, and R. Guerin. Specification of Guaranteed Quality of Service. RFC (Standards Track) 2212, IETF, September 1997.
30. L.F. Sun, G. Wade, B.M. Lines, and E.C. Ifeachor. Impact of Packet Loss Location on Perceived Speech Quality. In *IP Telephony Workshop*, pages 114–122, New York, USA, April 2001.
31. A.K. Talukdar, B.R. Badrinath, and A. Acharya. Integrated Services Packet Networks with Mobile Hosts: Architecture and Performance. *ACM Wireless Networks*, 5(2):111–124, November 1999.
32. N.D. Tripathi, J.H. Reed, and H.F. Van-Landingham. Handoff in Cellular Systems. *IEEE Personal Communications Magazine*, 5(6):26–37, December 1998.
33. A. Veres, M. Barry, L.-H. Sun, and A.T. Campbell. Supporting Service Differentiation in Wireless Packet Networks using Distributed Control. *IEEE Journal on Selected Areas in Communications*, 19(10):2081–2093, October 2001.
34. J. Wroclawski. Specification of the Controlled-Load Network Element Service. RFC (Standards Track) 2211, IETF, September 1997.
35. L. Zhang, V. Jacobson, and K. Nichols. A Two-bit Differentiated Services Architecture for the Internet. RFC (Informational) 2638, IETF, July 1999.