

Secondary Publication



Prokopp, Daniel; Matthes, Michael; Drobnik, Oswald; Krieger, Udo

Integration of Mobility-, QoS-, and CAC-Management for Adaptive Mobile Applications

Date of secondary publication: 22.04.2026

Accepted Manuscript (Postprint), Conferenceobject

Persistent identifier: urn:nbn:de:bvb:473-irb-114797x

Primary publication

Prokopp, Daniel; Matthes, Michael; Drobnik, Oswald; Krieger, Udo (2003): Integration of Mobility-, QoS-, and CAC-Management for Adaptive Mobile Applications, in: Wojciech Burakowski, Andrzej Beben, and Berthold Koch (Ed.), Architectures for Quality of Service in the Internet : International Workshop, Art-QoS 2003 Warsaw, Poland, March 24–25, 2003 Revised Papers, Berlin ; Heidelberg: Springer, pp. 29–48, doi: 10.1007/3-540-45020-3_3.

Legal Notice

This work is protected by copyright and/or the indication of a licence. You are free to use this work in any way permitted by the copyright and/or the licence that applies to your usage. For other uses, you must obtain permission from the rights-holders.

This document is made available with all rights reserved.

Integration of Mobility-, QoS-, and CAC-Management for Adaptive Mobile Applications

Daniel Prokopp¹, Michael Matthes¹, Oswald Drobnik¹, and Udo Krieger^{1,2}

¹ Telematics & Distributed Systems, Institute of Computer Science,
J.W. Goethe-University, D-60054 Frankfurt, Germany

{prokopp,matthes,drobnik}@tm.informatik.uni-frankfurt.de

² T-Systems, Technologiezentrum, Am Kavalleriesand 3, D-64295 Darmstadt
udo.krieger@ieee.org

Abstract. We sketch a QoS framework for adaptive mobile multi-media environments that integrates mobility- and quality-of-service management with new connection admission control (CAC) schemes. Referring to measurement-based CAC procedures applied in mobile access networks, we discuss their extension to adaptive mobile applications and describe the integration of corresponding resource reservation and CAC policies in our framework. Furthermore, we evaluate the performance of several CAC schemes by a simulation study.

1 Introduction

Currently, the integration of 3rd generation (3G) wireless networks into high-speed next generation networks (NGNs) based on the IP technology and its corresponding resource reservation and QoS-management mechanisms are some of the most challenging tasks of network design and engineering. These 3G networks like UMTS or IEEE802.11a/b compatible wireless LANs (WLANs) are realized on top of new WCDMA or DSSS transmission technology. The fast convergence of these wireless and wired IP networks coincides with the planned deployment of new integrated location-aware multi-media and interactive Web services. The latter are independent of the realized mobility patterns of the terminals and users and derived from a unified personal mobility concept. At present, there are different proposals to realize such new mobile quality-of-service (QoS) architectures which integrate both wired core networks and wireless segments at the edges. In the IST-project Moby Dick, for instance, a Diffserv approach is extended to the wireless segments. Recently, an alternative methodology has been discussed that is derived from the Intserv service model and its resource management and signaling protocol RSVP (cf. [10], [11]). Since nowadays the closely related MPLS paradigm with RSVP-signaling and tunneling components is used in the core network, we follow the corresponding resource-management concept derived from a mobile RSVP (MRSVP) approach (cf. [12], [20]).

Using the currently available inexpensive IEEE802.11b compatible radio access network (RAN) technology and a Mobile IP based data communication,

we have implemented a new integrated mobility and quality-of-service management concept for mobile adaptive multi-media environments. Referring to measurement-based connection admission control (CAC) schemes applied in radio access networks, we sketch their extension to adaptive mobile applications and the integration of the corresponding resource reservation and CAC policies into our developed framework. Furthermore, we describe several CAC schemes and evaluate their performance by a simulation study. The paper is organized as follows. In Section 2 we discuss mobility and QoS management as well as measurement-based CAC in mobile networks. Section 3 is devoted to the integration of micro-mobility, resource management and admission control policies. In Section 4 we present the results on the performance comparison of the integrated resource reservation and adapted CAC strategies. Finally, some findings are summarized.

2 Mobility- and QoS-Management in Mobile Networks

2.1 Mobility Management in IP Networks

Considering multi-media and advanced Web communication of mobile hosts (MHs), we can distinguish two different mobility scenarios:

- *Macro-mobility*: MHs move among different address domains of IP subnetworks with changes of the address space.
- *Micro-mobility*: MHs move within the same address domain.

Looking at macro-mobility issues, changing the access point (AP) or base station (BS) of the radio access network (RAN) may result in a new RAN operator. This type of movement is called *inter-domain* mobility. Such a change of the RAN operator causes, in general, a change of the IP address of the moving MH, since different Internet (subnetwork) domains are normally managed in distinct RANs. To guarantee the connectivity of an MH, the latter is assigned a local address of the related domain. The corresponding mobility and address management functions are provided by enhanced variants of the IPv4 protocol such as Mobile IP (MIP) or by an integral part of IPv6 not studied here (cf. [16]).

Normally, an MH moving within its home network is registered at a home agent with its permanent address called *home address*. Considering the delivery of datagrams to an MH moving outside its home network, apart from its fixed home address a second temporary local address, called *care-of-address*, is additionally assigned to the MH by the Mobile IP protocol after its handover to a new domain. It is provided by a corresponding mobility management agent, called foreign agent, and registered in its data base together with the other routing information. Then it is used to forward the incoming datagrams to the corresponding MH residing outside of its home network (cf. Fig. 1). Mobile applications identify the mobile clients by the home address. During the movement of an MH address translation is provided in a transparent manner by changing the care-of-address and its registration using a communication over tunnels between

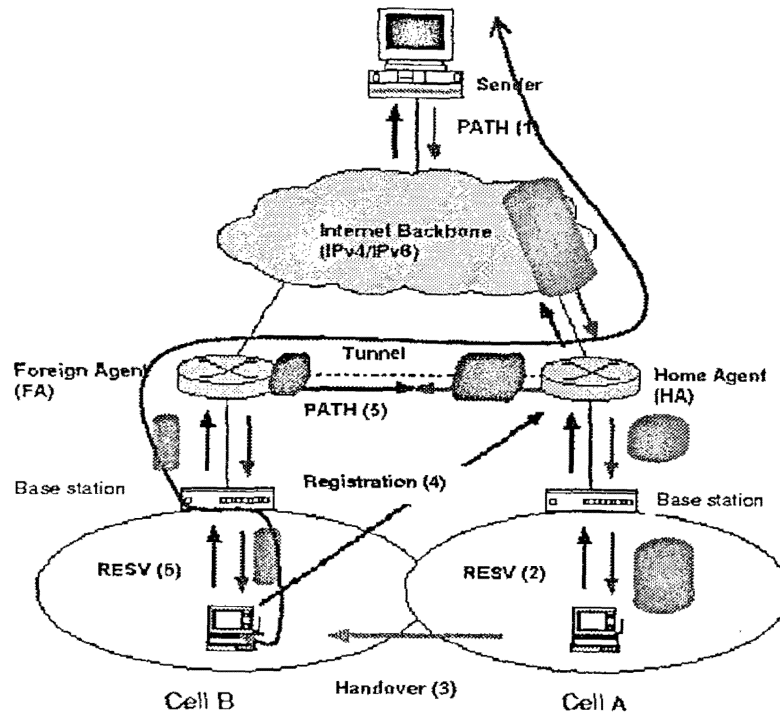


Fig. 1. Roaming and packet forwarding with Mobile IP

the involved home and foreign agents (cf. Fig. 1). However, a large overhead is required to guarantee a transparent communication among the peer instances resulting in a considerable performance degradation during the handover process. Therefore, specialized procedures have been developed to improve the latter processes by effective detection and realization of fast handoff control with QoS and security support (cf. [19], see Fig. 1).

Micro-mobility captures a scenario in a micro-cellular network where an MH detaches from its current AP, is handed over to a new radio cell and attaches to a new AP while residing within the management domain of the same RAN operator (cf. [17,18,19]). This scenario is called *intra-domain* mobility. The MH keeps its IP address assigned after a handoff in the domain of a RAN. The connectivity of the MH is guaranteed within a domain applying specific routing mechanisms (cf. [17,18,19]). Since the mobility management and corresponding registration procedures are performed within the same domain, the required location updates and signaling of address binding information of an MH can be efficiently realized compared to the inter-domain mechanisms used by Mobile IP.

2.2 QoS Management in Mobile Networks

IP communication in wireless segments of advanced mobile networks like the RAN of an IEEE802.11a WLAN with 54 Mbps is substantially different from current inter-networking in the high-speed backbone. Normally, the protocol functions of the error control and MAC sub-layers of the data link layer (DLL)

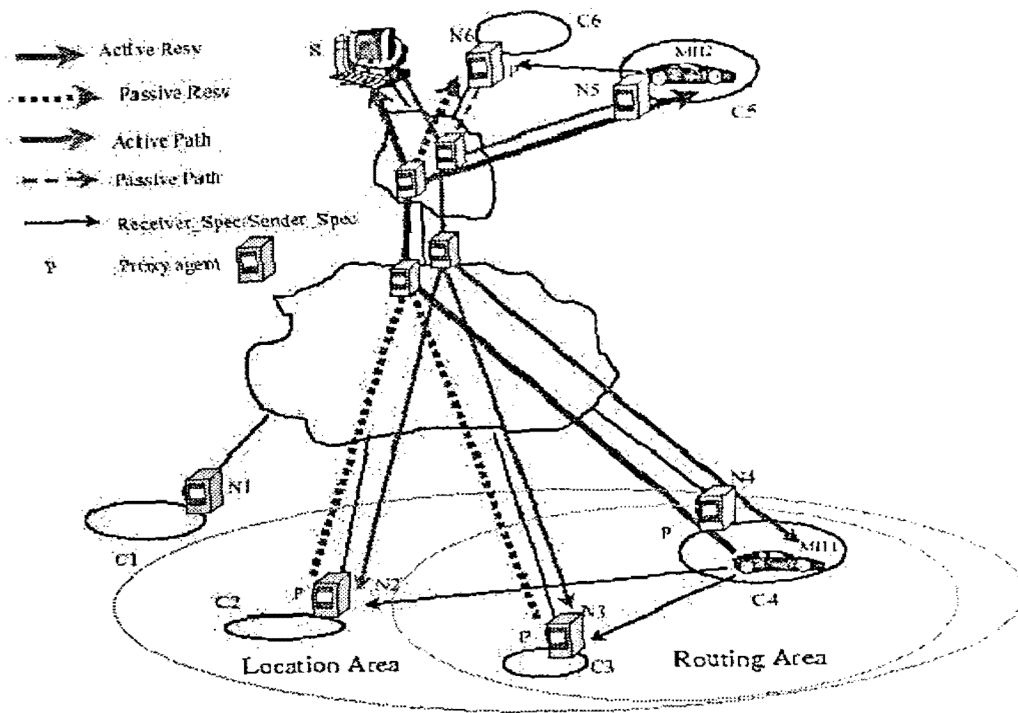


Fig. 2. Active and passive reservation of resources in the RAN

are decoupled of those procedures invoked at upper layers during data transfer. Moreover, distributed environments of contemporary IP networks do not offer any integrated controls that support the information exchange between the network layer and its lower layers. The latter, however, is a common approach during handoffs in 2+ and 3rd generation networks such as GPRS and UMTS where the radio resource management performs corresponding tasks.

To provide an efficient solution of micro-mobility management issues, optimized handoff and QoS management procedures are required (cf. [2,12]). It is the objective to minimize the disruption phase of connectivity while an MH detaches from its current AP and registers at the mobility agents after attaching to a new AP. In an advanced mobile network the corresponding handoff and registration processes affect several layers of the protocol stack (cf. Fig. 3). First, the movement detection functions at the MAC layer of the DLL are invoked to determine the required change of the AP and to identify the destination of a handoff. To maintain the connection and the effective flow of datagrams, registration and control instances at the upper layers are involved. For instance, by protocols like Hawaii, cellular IP or hierarchical Mobile IP in a way similar to Mobile IP appropriate IP address and routing information is generated and routing agents are notified to guarantee a seamless flow of data after completion of a handoff (cf. [17,18,19]). Furthermore, the resources of the transferred connections required in the new cell have to be re-negotiated and the flow rates have to be adjusted. Then the associated mobile applications can adapt to the resource utilization within the new cell (cf. [2,12], see Fig. 3).

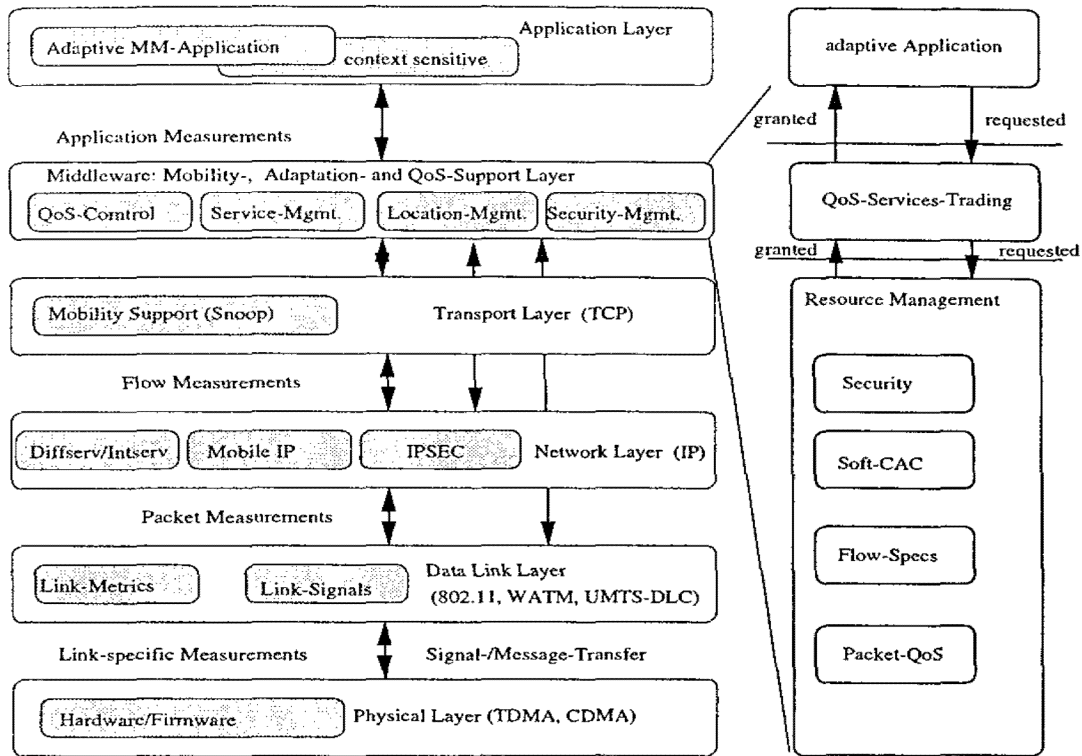


Fig. 3. Protocol stack of an NGN

After a handoff among radio cells the following situations can arise that result in quite different effects on the resource allocation and seizure by active connections of adaptive mobile applications:

- In the visited cell more resources are available than in the left. Hence, the MH can seize this additional capacity, distribute it among the active connections and trigger the invoked applications to increase their sending rates accordingly.
- In the new cell less resources are available than before. In this case, the MH has to reduce its consumption and to notify the active applications on the required adaptation processes.
- After a handoff no changes of the resource allocations are required.

In all cases an information exchange between the MH and a resource management component in the RAN about the required and granted resources and QoS objectives is necessary. Depending on the previous status of the resource allocation to the active connections of an MH corresponding adaptation processes have to be invoked at the transport and application layers (cf. [2,12], see Fig. 3).

2.3 Measurement-Based Adaptive Connection Admission Control

Considering mobile users with their mobile hosts (MHs) roaming within an advanced mobile network, the corresponding connections have different QoS requirements. The latter are characterized in terms of a quadruple: the minimal and maximal required bandwidth, a maximal end-to-end (E2E) delay and an

upper bound on the packet loss. Based on a chosen service model, normally we distinguish between those connections belonging to service class C1 with stringent QoS guarantees, i.e. guaranteed or premium service, a class C2 with soft requirements, i.e. predictive, assured or elastic service, and the best effort class C3 without any QoS requirements.

To maintain the QoS guarantees of active connections of roaming MHs during handover processes and to check whether new connections of fresh and handoff calls can be established without violating the guarantees of existing ones within a specific radio cell, a connection admission control (CAC) scheme constitutes a necessary ingredient of the resource management system (cf. [5,6,7,8,9,14,15, 21,22]). This decision process has to distinguish between both the connections of the sketched three service classes and the type of a call, i.e. whether it is a fresh call just triggering a new connection set-up in a cell or an already active connection requiring further connectivity in a neighbor cell after a handoff. Since the latter process may fail due to a lack of resources or a violation of existing QoS guarantees, a constraint on the corresponding dropping probability can be formulated as additional QoS criterion of a connection. For fresh calls the blocking probability of a connection is the basic QoS term apart from delay and loss constraints for real-time traffic of the premium class C1 and apart from the loss constraints for the elastic traffic of class C2. Furthermore, the utilization of the bandwidth offered by a cell constitutes the basic efficiency metric that has to be maximized by the network operator.

To achieve these performance objectives in the dynamic environment of a RAN, the corresponding CAC components comprise three building blocks: means to handle the traffic specification including the QoS requirements, a bandwidth reservation and admission control policy (ACP) and related decision algorithms as well as measurement functions to monitor the actual load and utilization of the resources managed by a specific resource management entity (RMC) of a RAN (cf. [21,22], see Fig. 4).

In the following we focus on the relationship of micro-mobility management and the corresponding CAC functionality. Regarding the QoS demands at the call level we restrict ourselves to the blocking and dropping probabilities in a single cell since the concatenation (i.e. convolution) of the latter performance indices provides a convenient tool to determine the E2E performance. Therefore, local and distributed control architectures are under consideration as ACP if we consider a target cell and its immediate neighbors.

The simplest ACP disregards the service classes and traffic types as well as the mobility (i.e. handoff) patterns and accepts fresh or handoff calls if there is enough capacity C available in a cell. For that purpose, a continuous monitoring of the load and utilization is required yielding measurement-based CAC schemes. Due to the random movement of the MHs dynamic ACPs are more appropriate policies. Since best-effort connections can be dropped without any constraints, it is sufficient to study the impact of micro-mobility on connections of the premium and elastic classes C1, C2. Since real-time traffic is transported in C1 and non-real-time in C2, we assume that only handoff calls of C1 specify

a maximal dropping probability $IP\{\text{drop}\}$. The latter has to be guaranteed by the ACP during handoff to the new cell whereas C2 handoff calls do not require such treatment due to the elasticity and delay-insensitivity of the corresponding sources. The actual realization of $IP\{\text{drop}\}$ is estimated by the ratio of rejected and the overall number of occurring handoff calls of the corresponding class.

To achieve the required protection of handoff calls compared to fresh calls, two elements of a protection strategy can be applied based on resource pooling and bandwidth reservation for handoff calls: reservation of a bandwidth portion in a specific handoff buffer and different threshold mechanisms (cf. [21,22]). They extend the concepts previously known as guarded channel schemes. Considering the simplest threshold schemes, bandwidth resources (BW) are reserved for calls potentially handed over to a cell. The reserved bandwidth is reflected by a threshold value. The latter is determined by the difference of the BW available in the cell and the required BW reserved for handoffs (i.e. a reservation index). A fresh call is accepted if the sum of the BW occupied by active calls and the applying fresh one does not exceed the threshold level, whereas a handoff call is allowed to exceed the threshold as long as enough capacity is available in the cell. In the case of a buffer scheme the sum of the BW occupied by active calls, the applying fresh one and the junk of bandwidth reserved for handoff calls is not allowed to exceed the cell capacity (cf. [5], see Fig. 2).

3 Integrating Micro-Mobility and Adaptive Connection Admission Controls

3.1 Integrated Mobility-, Resource-, and CAC-Management

In our approach we assume that mobility-aware multi-media and real-time applications are capable to adapt their required bandwidth by means of an advanced coding technology to a certain extent according to control signals emitted by data-stream and resource management components of our new QoS architecture (cf. [2,12]). To satisfy the diverse requirements of co-existing adaptive and standard applications, the developed programming model uses an abstraction with three layers: the standard and adaptive applications, the resource-management and QoS-adaptation level called QoS-services-trading, and the integrated connection and flow layer (see Figs. 3, 4). The latter is called network interface. It provides an abstraction of the transport and network layers and their corresponding signaling and data flows at the mobile nodes and the nodes of the access infrastructure such as the APs. Data flows of the applications are transported across these components. Moreover, the control functionality of these lower layers, e.g. reservation agents and bandwidth broker, traffic classifier, meter and shaper, packet scheduling and marking as well as buffer management, that is implemented in a distributed manner at the network elements to handle the QoS requirements at the connection, flow and packet levels can be adjusted by the network interface.

It is the goal of this mobile QoS (MQoS) management system to allocate, monitor, and manage the resources of the wireless network elements and to distribute their capacities in a fair manner among the competing applications and the associated data flows of their connections. The basic components performing these tasks are called data-stream management at the MNs and resource management (RMC) at the nodes of the access infrastructure (see Fig. 4). They handle the signaling of the requirements, the resource allocation, adaptation and management processes. To get the information required for the adaptation processes, they are supported by a monitoring component called environment monitor (EMC). It is an abstraction of those components already available in mobile terminals of second generation networks, e.g. the control processes monitoring the signal-strength, noise, BER and PDU-loss along the forward and backward paths of data flows over the air interface, and the movement-detection and location-identification processes. By these means the location of a terminal and the potential change of a micro-cell and its association to an AP due to a horizontal, or even vertical, handoff can be identified. Hence, the monitoring component can also provide control information and an interface for location-management and service-management components at higher layers below the applications if location-dependent services are supported (see Fig. 3).

Considering the actual actors at the distributed control instances residing at the session, flow and packet levels of the protocol stack of the MNs and the active nodes of the access network, appropriate control commandos and their parameters are passed by means of an intermediate mediator component called network-interface control. It translates the invocation messages of the management component into corresponding control actions that are transferred to the network-interface. Changes of the allocations, e.g. the addition of new bandwidth after the handoff of a flow to a new AP at the old interface and the reduction of the latter at the new interface among all competing flows due to the fair-share concept, are signalled by the data- and resource-management and translated into control actions of the network interface by the network-interface control.

Evaluating the experience gained in a new WLAN testbed (cf. [2,12,13]), we have realized that a re-negotiation of the resources after the completion of handoff procedures at the physical, MAC and IP layers consumes too much time. Therefore, a measurement-based CAC functionality is incorporated in our MQoS framework as connection access control (CON) component in the RAN (cf. Fig. 4). For this purpose, we have extended the existing ACPs to incorporate the adaptivity of advanced mobility-aware multi-media and real-time applications, e.g. voice services with AMR codecs and video applications with variable bit-rate codecs based on MPEG4 etc. This means that the corresponding handoff calls of C1 can adapt their sending rate or bandwidth requirements based on notification signals that are dynamically generated by the RMC and CON during a handoff process. The corresponding architecture is depicted in Fig. 4. This integration of micro-mobility management, bandwidth reservation and adaptive admission control proceeds as follows. In each cell a variable proportion of resources is assigned in a pro-active manner as bandwidth buffer to handle handoff calls arising

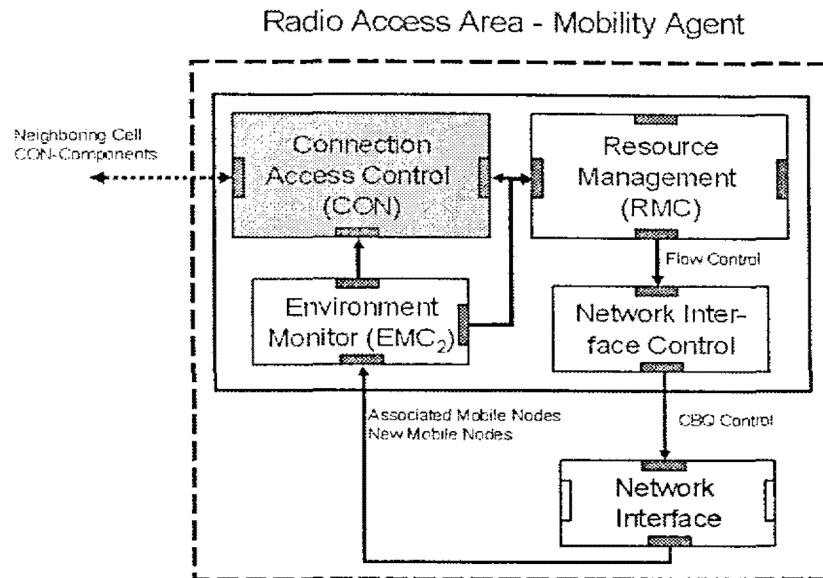


Fig. 4. Integration of mobility and resource management as well as connection admission control

from neighbor cells. The reservation processes are coordinated by the interworking of RMC, EMC and CON evaluating the micro-mobility patterns arising to and from as well as among neighbor cells (cf. [3,4,22,23]). For this purpose, the corresponding instances communicate by control messages with each other and the CON has interfaces to EMC, RMC and other CONs. The RMC determines the requested and granted resources for each call and CON, which implements the ACP, reserves an appropriate number of BW units for a potential handoff call in the neighbor cells.

To implement this process in a way similar to the mobile RSVP model, each connection of a service class is characterized by a bandwidth interval of minimal and claimed BW units (up to a maximal value) that are required to realize a call with sufficient service quality. Dropping of a handoff call is prevented or, at least, relieved by reserving an appropriate number of BW units in potential target cells surrounding the current residence of an MH before a handoff. To avoid the wasteful passive reservation of bandwidth, the movement of MHs is anticipated and a portion of the actually demanded BW units is reserved for all potential calls and put into a BW pool called handoff BW buffer. It is exclusively used for handoff calls. To prevent inefficient BW use and to avoid high blocking probabilities of fresh calls, its size is limited.

Moreover, it is dynamically controlled depending on the experienced dropping probability. If the dropping probability that is monitored in a cell by the EMC and CON components exceeds a prescribed QoS level, more resources are granted in advance for future handoff calls in the cell. If a fresh call arrives and cannot be admitted due to a lack of capacity subject to these BW reservations applying a specified ACP, it is blocked. If a handoff call demands admission and

the required BW units are not available, then a two stage process is initiated. First, its requirements are reduced to the specified minimal value and admission control is applied again based on the specified ACP. Only if it fails, the handoff call is dropped. Thereby, the congestion of handoff calls is relieved exploiting the elasticity of the sources belonging to the premium class.

3.2 Admission Control Policies for Adaptive Mobile Applications

Considering the ACP for calls arising from adaptive mobile applications, we generalize the schemes discussed in [5]. We assume that class C1 comprises MHs running delay-sensitive applications whereas C2 includes delay-insensitive connections of elastic services, e.g. arising from http and plain tcp data transport. Regarding the performance targets of micro-mobility management C1 calls are characterized by maximal dropping probabilities $IP\{\text{drop}\}$ during handoff while no QoS objectives are specified for C2 calls.

Prediction of Mobility. The prediction of local movements of MHs between cells uses a list of handoff patterns observed in the past (cf. [3,4,5,22,23]). They are stored in an appropriate data structure. It comprises a list of quadrupels $Q = (t, k, i, d)$ related to a previous handoff event into the current cell 0 from a predecessor cell k . At time t after the dwell time d in the current cell 0 the MH has been handed over into the successor cell i . Regarding a new call arriving at cell 0 the tasks include the computation of the transition probabilities (t.p.s) p_h and the epochs t of its corresponding handoff events to successor cells i .

For the j -th connection $C_{0,j}$ handed over from a predecessor cell k into the current cell 0 a histogram-type estimate of the corresponding movement-dependent t.p.s can be calculated as follows:

$$p_h(C_{0,j} \rightarrow i) = \frac{\# \text{ quadrupels with predecessor cell } k \text{ and successor cell } i}{\# \text{ quadrupels with predecessor cell } k} \quad (1)$$

Regarding the conditional average dwell-times $t_h(C_{0,j} \rightarrow i)$ of the j -th call $C_{0,j}$ coming from cell k before handoff into cell i a similar empirical mean is used as estimate:

$$t_h(C_{0,j} \rightarrow i) = \frac{\sum_{l=1}^{|Q(k,i)|} d^l}{|Q(k,i)|} \quad (2)$$

Here $Q(k, i)$ denotes the set of all quadrupels with predecessor cell k and successor i and d^l the dwell time of the l -th quadrupel. To limit the computational effort, the order of the list should not exceed an upper bound N_{quad} and outdated quadrupels are substituted by recent ones.

Bandwidth Reservation and Admission Control Strategies. Considering the passive resource management for calls of class C1 by CON, after arrival and admission to a cell 0 a portion b_{pre} of the demanded BW is reserved in the BW pool of C1-handoff calls in its neighbor cells. In cell i this passive BW reservation

of a call $C_{0,j}$ admitted in cell 0 is determined by the product of the seized BW $b(C_{0,j})$ and the t.p. $p_h(C_{0,j} \rightarrow i)$ to the successor cell i :

$$b_{pre} = b(C_{0,j}) * p_h(C_{0,j} \rightarrow i) \quad (3)$$

Together with the expected mean dwell time $t_h(C_{0,j} \rightarrow i)$ as time-to-activate (TTA) value, this amount b_{pre} is signalled by CON to cell i . Receiving the passive reservation message, the RMC of cell i is notified by its CON. It calculates the reservation epoch by $t = t_0 + t_h(C_{0,j} \rightarrow i) - T$ where t_0 is the actual time of a reservation receipt and T denotes a correction term. The latter adjusts the mobility predictions and is dynamically updated.

Following [7] the procedure quoted here has been incorporated in the ACP

```

SP = 1/P(Drop) ; LP = SP;
SH = 0; SHD = 0; LH = 0; LHD = 0;
T = 0;
IF (MH handed over)
{
    SH++; LH++;
    IF (MH connection dropped)
    {
        SHD++; LHD++;
        IF (LHD > 1)
        {
            LP += SP;
            T += STEP_SIZE;
        }
    }
    IF (SH == SP)
    {
        IF (SHD < 1)
            T -= STEP_SIZE;
        SH = 0; SHD = 0;
        IF (LH == LP)
        {
            LH = 0; LHD = 0;
            LP = SP;
        }
    }
}

```

approach to adjust the corresponding correction T of the reservation epoch that influences a passive reservation. Here the variable STEP_SIZE controls the speed of the adaptation. It is required to correct errors of the mobility predictions that yield over- or under-estimates of the handoff BW and, hence, a performance degradation of the admission policy. A positive T will increase the reserved passive BW if the dropping level is exceeded. If one is in a regime far below this level, $T < 0$ shall reduce temporary the current portion of passive reservations.

Storing the capacity requests and TTA values of passive reservations in a list and traversing the latter periodically, CON and RMC of a RAN can determine the actual number of active reservations in a cell i . Activating a passive reservation, the reservation index B_i of a cell i is increased by the corresponding number b_{pre} of BW units. It determines the number of BW units in the handoff buffer serving the calls handed over from other cells.

If the connection $C_{0,j}$ leaves the cell 0, CON notifies the neighbor cells to release the corresponding reservations. If a reservation has not been activated, it is simply deleted from the list. Otherwise, the reservation index is decremented accordingly.

CON has the task to implement an ACP that guarantees a specified QoS level for a fresh or handoff call of classes C1 and C2, respectively. In our approach we use a scheme with this sketched distributed information state and a local

decision process. Regarding fresh calls of C2 and handoff calls of either type the acceptance decision is only determined by the utilization status of a considered cell 0. For fresh calls of C1 it is additionally determined by the state of the neighbor cells. We denote the reservation index of this target cell 0 with capacity $C(0)$ by B_0 . Let C_0 be the set of all connection indices i in cell 0 with bandwidth demands $b(i) = b(C_{0,i})$ and let b_{new} be the BW demand of a new connection. Then the ACP is determined by the following rules:

1. admission of fresh calls of type C_1 :
 - check $\sum_{i \in C_0} b(i) + b_{new} \leq C(0) - B_0$
 - check whether an overshooting of reservations exists in neighbor cells (cf. [5])
2. admission of handoff calls of type C_1 :
 - check $\sum_{i \in C_0} b(i) + b_{new} \leq C(0)$
3. admission of fresh and handoff calls of type C_2 :
 - check $\sum_{i \in C_0} b(i) + b_{new} \leq C(0) - B_0$

Applied Adaptive Admission Control Policies. To identify the most appropriate ACP subject to a micro-mobility regime, several known bandwidth reservation and admission control schemes have been adapted to the new context of adaptive mobile connections arising from the premium and elastic services classes C1 and C2. There are three basic BW reservation schemes inherited from the class library of the tool Cellular Network Simulator (CNS) (cf. [1]):

Scheme A: Neither passive reservations nor CAC are applied. Fresh calls (FCs) are accepted as long as enough BW is available, whereas handoff calls (HOCs) are admitted if the minimal BW requirement can be satisfied.

Scheme B: It is similar to A. However, if BW is missing for a C1-HOC, it is grasped from the contingent of active C2 calls. Only the minimal requirement of a C1-HOC is seized. The reduction is uniformly distributed among all C2 calls.

UBB: This scheme applies passive BW reservations. A C1-FC is accepted if in the current and all neighbor cells there is enough BW applying the ACP therein. The maximal BW of all C1 calls is reserved in the neighbor cells.

A C1-HOC is admitted if enough BW can be provided to satisfy its minimal requirement in the current cell and all passive reservations in neighbor cells are granted. In contrast to C1-FCs or C2 calls, a C1-HOC is allowed and forced to seize a portion of the reserved BW in the handoff buffer. If the latter exceeds the capacity of the buffer partially or completely, the BW can be taken from the common resource pool if possible. If the call leaves the cell, the seized BW is put to the common pool and not added to the handoff pool. The latter is changed by future reservation or deletion requests.

These basic schemes have been combined with the following adjusted variants of some proposed adaptive distributed ACPs (cf. [5,6,7,15]):

UBB_adapt: It is similar to UBB, but applies a dynamic adjustment of the BW pool (cf. [15]). If a prescribed dropping level is exceeded or the utilization of the handoff pool exceeds a threshold, the latter is increased. If these control parameters fall below some thresholds, it is reduced.

Prob_Thres: For each C1 call a random portion of BW is reserved in neighbor cells and reduces the threshold levels (cf. [6]). A C1-FC is blocked if the modified threshold value cannot be granted in a neighbor cell. A C1-HOC can seize the overall available BW. Fresh calls and C2-HOCs are only admitted if the sum of their BW demand and the already seized BW does not exceed the threshold.

AC2: It is a threshold-based scheme with similar ACP like Prob_Thres regarding C2-FCs and C2-HOCs (cf. [7]). For C1-FCs and HOCs no passive reservations are performed in neighbor cells. It is only checked whether the seized BW does not exceed the thresholds there. In this case, the cell is too crowded and C1-calls are rejected.

AC3: It is similar to AC2, but additionally thresholds are adjusted between neighbor cells to increase inter-cell fairness (cf. [7]).

4 Performance Analysis of Adaptive CAC Schemes

4.1 An Enhanced Cellular Network Simulator (CNS)

Using the existing Java class library [1], we have enhanced the corresponding *Cellular Network Simulator (CNS)*. The related tool incorporates the sketched ACP variants of the bandwidth reservation and CAC schemes that have been adapted to adaptive mobile applications of classes C1 and C2, respectively. To enable a comprehensive investigation of these policies and their ability to guarantee both a maximal dropping probability of C1-HOCs and a high utilization of the bandwidth provided by a micro-cellular network subject to different handoff scenarios, the tool has been further enhanced by a unique graphical user interface (see Fig. 5). The latter offers the specification of all relevant network, load and simulation parameters. By these means it is possible to specify the network structure (called *Network Specification*) in terms of its size, i.e. as a grid of N micro-cells with a common capacity of C BW units each. By a *Mobile Host Specification* the parameters of both service classes C1, C2 can be determined. A connection of each class is given in terms of a minimal and a demanded BW level. Furthermore, the movement pattern of the MHs can be set. The attribute DIRECTED characterizes those MHs which select a random direction on the cell grid at connection set-up whereas RANDOM governs MHs selecting new cells at handoff events in a random manner. The grid is wrapped around, i.e. reaching its edge a new cell is selected at random. The *Simulation Specifications* allow to control the simulation runs and the used load model in terms of the arrival rate λ , the mean call durations μ^{-1} , and the mean dwell time γ^{-1} in a cell.

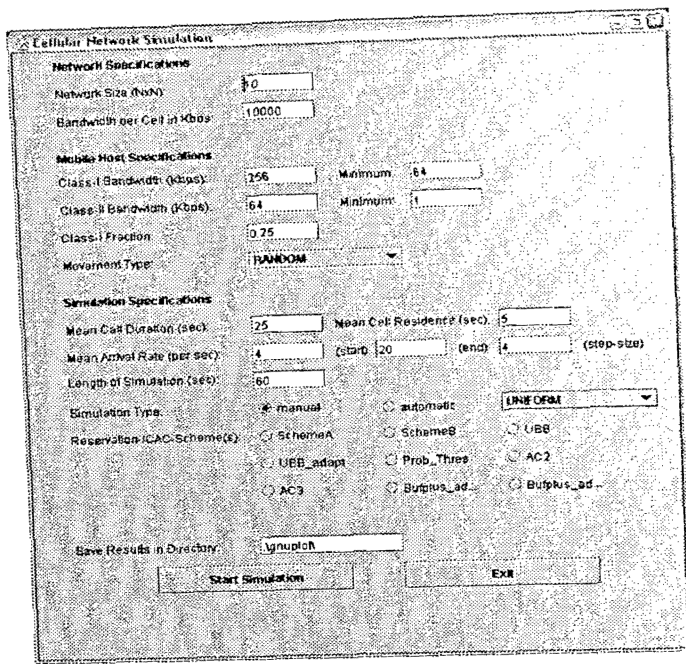


Fig. 5. GUI of a CNS enhancement

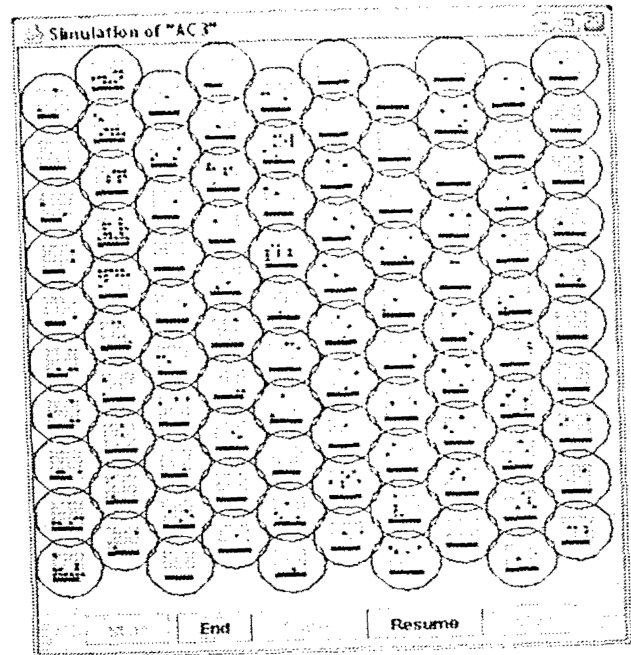


Fig. 6. Results of a simulation of a 10×10 grid

The arrival of MHs to a cell are governed by a Poisson distribution, and the call and dwell times are independent exponentially distributed. Furthermore, three different load scenarios can be studied:

- **UNIFORM:** For each cell the mean arrival rate is identical.
- **SINGLE_SOURCE:** New calls are always initiated in the same cell. Movement patterns can be tracked easily to trace a run.
- **BURSTS:** A target cell and its neighbors are the active zone of the network. After the expiration of a *Burst Cycle Timer* all MHs move simultaneously into the neighbor cells.

After the simulation runs all observed performance metrics, i.e. captured statistics of ACTIVE-, IDLE-, DROPPED- and BLOCKED calls, can be depicted and evaluated (cf. Fig. 6).

4.2 Performance Results of a Comparison

To investigate the performance of the developed ACPs and to determine which scheme should be implemented in the MQoS framework, a simulation study has been executed by the enhanced CNS tool. It uses a grid with 25 cells wrapped around at the edges yielding 6 neighbors of each cell.

The load is specified in terms of the demand of BW units (BU). C1 calls require 4 BU, C2 1 BU. Each cell can carry $C = 100$ BU. Two load scenarios are studied: either an equal number of C1 and C2 calls ($F_{C1} = 0.5$) is generated in a run or only C1 calls ($F_{C1} = 1$) are generated.

Two mobility patterns are studied: high mobility of MHs yielding on average 3.5 handoffs and low mobility resulting in 1.5 handoffs per call. The average call

duration is $\mu^{-1} = 120$ sec, the average dwell time in a cell $\gamma^{-1} = 34$ sec for high mobility and $\gamma^{-1} = 80$ sec for low mobility. The maximal utilization of a cell is determined by:

$$L = (4 * F_{C1} + 1 * (1 - F_{C1})) * \lambda * \mu^{-1} / C \quad (4)$$

It is the objective of the ACP simulation to compare the achieved QoS metrics, i.e. the mean blocking probabilities of calls and dropping probabilities of C1 handoff calls and the gained average utilization of a cell as well as the coordination costs of the control algorithms for adaptive mobile applications.

First, the impact of the sketched adapted local admission control schemes has been studied. Typical results for a scenario with load bursts show that AC2 is the worst scheme regarding the dropping and utilization objectives. The modifications exhibit a higher utilization. However, they cannot guarantee appropriate dropping levels. The second set of experiments compares the uncontrolled Scheme A with AC2 for a uniform load scenario. As shown in Figs. 7, 8 the utilization of Scheme A is only slightly higher than that achieved by AC2. However, Scheme A cannot guarantee the QoS objectives as expected. Hence, it is obvious

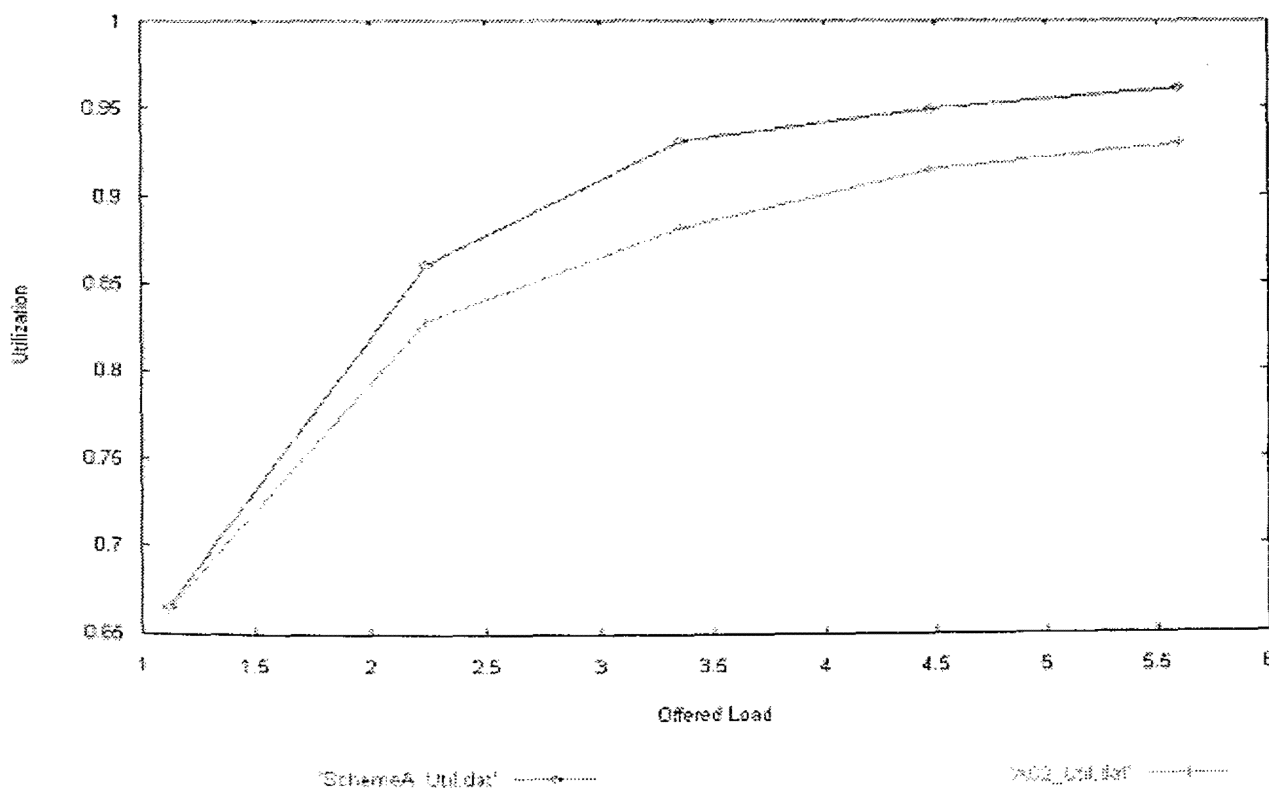


Fig. 7. Cell utilization for a uniform load scenario applying AC2 (lower curve) and Scheme A (upper curve)

that distributed ACPs are required to relieve congestion of handoff calls.

To evaluate further the benefits of those corresponding ACP proposals, moderately loaded cells with utilization $0.3 \leq \rho \leq 0.5$ are studied. Due to space

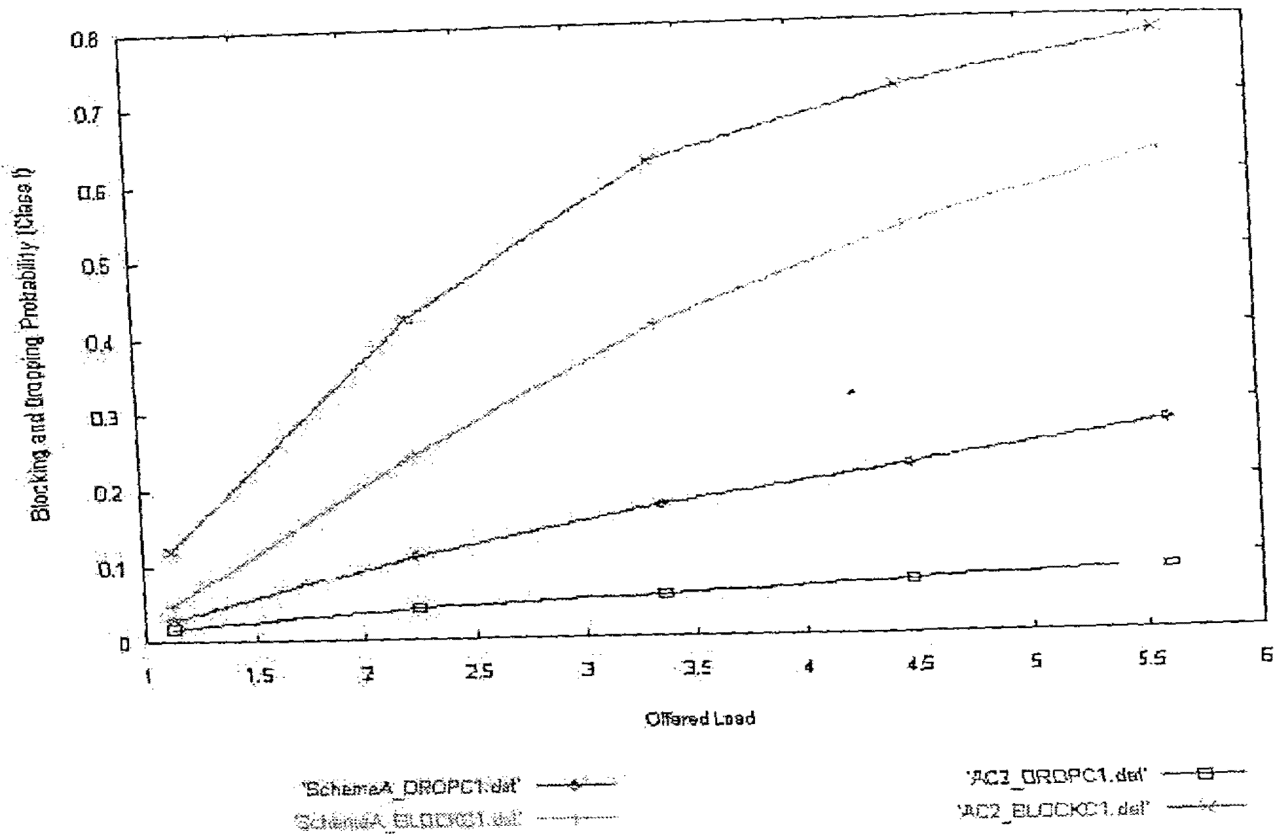


Fig. 8. Blocking and dropping probabilities of C1 calls given a uniform load scenario applying the AC2 and Scheme A

limitations only results of C1 calls are depicted here for the high mobility scenario. AC3 denotes the adapted scheme of Lee et al. [7], CHOI the variant of [5] and CHOLagg_besserHoRes its enhanced version with reduced overhead developed here. While AC3 adjusts the reservation index B_i of a cell i mainly by the monitored dropping levels, and CHOI considers only BW requests of potential incoming handoff calls within a reservation window, CHOLagg_besserHoRes represents a simplification of the latter. It works with the presented T -adjustment for the initial value $STEP_SIZE = 1$.

Focussing on the UNIFORM load model, a typical relationship of the dropping probabilities of C1 for different loads is shown in Fig. 9. We observe that all three schemes can guarantee the maximal dropping value of 0.1 and perform even better. Considering the average utilization of a cell, Fig. 10 illustrates that the proposed scheme CHOLagg_besserHoRes outperforms the others. The reason is that it makes less passive reservations and distributes the saved BW among other calls. Therefore, the resulting dropping probability slightly increases as shown in Fig. 9. Finally, the communication overhead of a typical cell is depicted in Fig. 11. As expected CHOI experiences the highest and AC3 the least cost while the proposed scheme lies in between. The reason is that AC3 and CHOLagg_besserHoRes require a test of the load status of all neighbor cells by the admission algorithm. If a call is completed or a handoff occurs, the latter scheme will cancel all pending reservations in these neighbor cells. Therefore, the communication overhead is twice as high as in AC3.

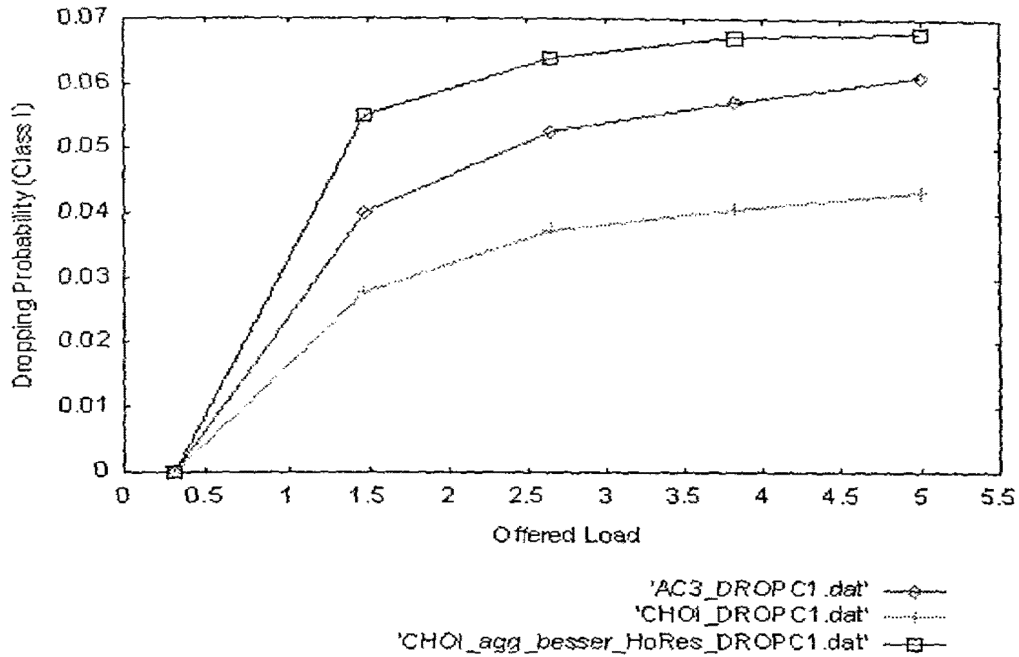


Fig. 9. Dropping probabilities for uniform load

In conclusion, we realize that our proposal of the distributed ACP CHOI_agg_besserHoRes with passive resource reservation can be effectively implemented in our MQoS framework to improve the QoS management of adaptive mobile applications subject to micro-mobility regimes.

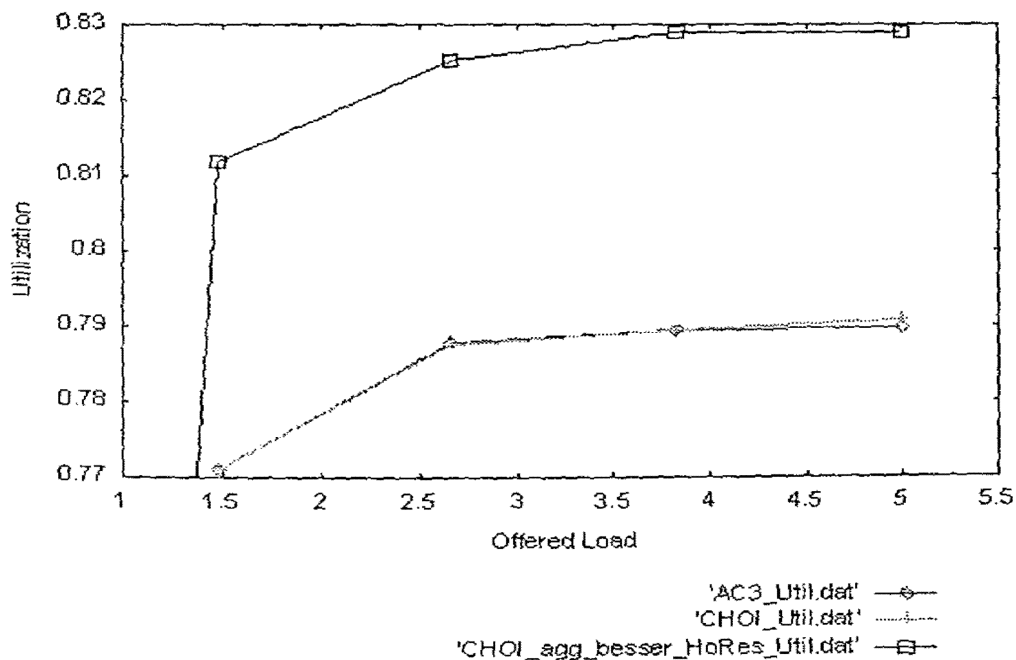


Fig. 10. Average utilization of a cell

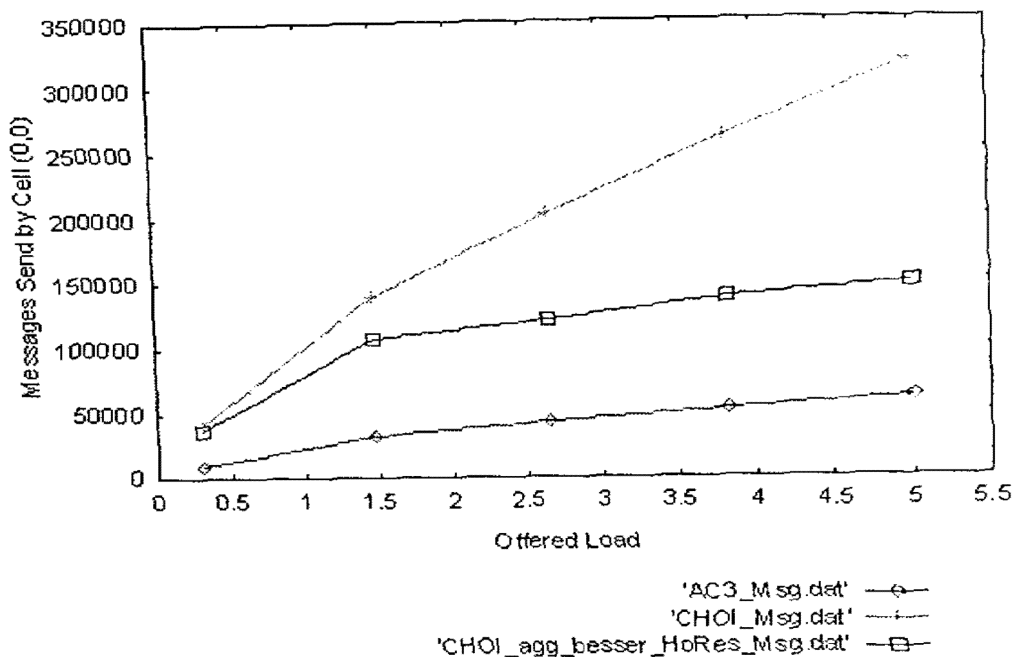


Fig. 11. Communication overhead

5 Conclusion

Currently, the integration of 3rd generation wireless networks such as WLANs and UMTS into high-speed IP core networks and its corresponding resource reservation and QoS-management mechanisms are some of the most challenging tasks of network design and engineering to support the deployment of new integrated location-aware multi-media and interactive Web services.

In our paper we have sketched a mobile QoS framework for adaptive mobile multi-media environments that integrates mobility- and quality-of-service management with new connection admission control (CAC) schemes. Referring to measurement-based CAC procedures applied in mobile access networks, we have discussed their extension to adaptive mobile applications and described the integration of the corresponding resource reservation and CAC policies into our framework. Furthermore, we have evaluated the performance of several CAC schemes by a simulation study. The results provide guidelines for further studies and an implementation in QoS architectures of next generation mobile networks.

Acknowledgment. The authors express their appreciation to J. Bachmann and T. Schönberger whose programming efforts constituted a substantial contribution to the project. Furthermore, U.R. Krieger wishes to express his sincerest gratitude to all those working at the library of T-Systems' Technologiezentrum Darmstadt, who supported him in gathering up-to-date scientific material throughout the past years. He acknowledges the "wisdom" of the trend-setting decision to close the library which reflects, from his personal point of view as a lecturer at J.W. Goethe-University, a mature level of scientific and cultural ignorance.

References

1. Possible Improvements to an Adaptive Bandwidth Reservation Scheme for High-Speed Multimedia Wireless Networks.
<http://www.cs.odu.edu/~elkadi/mobile/final.html>, 1998.
2. J. Bachmann, M. Matthes, U.R. Krieger, and O. Drobnik. Mobility and QoS-management for adaptive applications. In *Proc. of 11. International World Wide Web Conference*, Hawai, May 2002.
3. P. Bahl, V. N. Padmanabhan, and A. Balachandran. Enhancements to the RADAR User Location and Tracking System. *Technical Report, MSR-TR-2000-12, University of California at San Diego and Microsoft Research*, December 2000.
4. J. Chan, S. Zhou, and A. Seneviratne. A QoS Adaptive Mobility Prediction Scheme for Wireless Networks. In *Proc. of Globecom'98*, 1998.
5. S. Choi and K. G. Shin. Predictive and adaptive bandwidth reservation for handoffs in QoS-sensitive cellular networks. In *Proc. of ACM SIGCOMM'98*, September 1998.
6. R. Hutchens and S. Singh. Bandwidth Reservation Strategies for Mobility Support of Wireless Connections with QoS Guarantees. In *Proc. of Twenty-Fifth Australian Computer Science Conference (ACSC2002)*, 2002.
7. J. Y. Lee, S. Bahk, and K. Park. Adaptive Admission Control with Inter-Cell Fairness in QoS-Sensitive Wireless Multimedia Networks. See *<http://citeseer.nj.nec.com/479960.html>*, 2003.
8. 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 (TON)*, vol. 5, no. 1, February 1997.
9. S. Lu and V. Bharghavan. Adaptive Resource Management Algorithms for Indoor Mobile Computing Environments. *ACM SIGCOMM Computer Communication Review*, vol. 26, no. 4, October 1996.
10. I. Mahadevan and K. M. Sivalingam. A Hierarchical Architecture for QoS Guarantees and Routing Wireless/Mobile Networks. *Journal of Parallel and Distributed Computing*, vol. 60(4), pp. 510–520, April 2000.
11. I. Mahadevan and K. M. Sivalingam. Architecture and Experimental Results for Quality of Service in Mobile Networks using RSVP and CBQ. *Wireless Networks*, vol. 6(3), pp. 221–234, June 2000.
12. M. Matthes, J. Bachmann, U.R. Krieger, and O. Drobnik. A QoS management system for adaptive applications supporting mobility and end-to-end guarantees. In *Proc. of the IASTED International Conference, Wireless and Optical Communications (WOC2002)*, Banff, July 2002.
13. M. Matthes, O. Drobnik, and U.R. Krieger. Auswirkung drahtloser Netzsegmente auf die Verkehrsgüte von TCP/IP-Verbindungen. *Kommunikation in Verteilten Systemen (KiVS)*, 12. Fachkonferenz der Gesellschaft für Informatik (GI), Fachgruppe “Kommunikation und Verteilte Systeme” (KuVS), Springer Verlag, Germany, February 2001.
14. M. Naghshineh and M. Schwartz. Distributed Call Admission Control in Mobile/Wireless Networks. *IEEE Journal on Selected Areas in Communications*, vol. 14, no. 4, May 1996.
15. C. Oliveira, J. B. Kim, and T. Suda. An adaptive bandwidth reservation scheme for high-speed multimedia wireless networks. *IEEE Journal on Selected Areas in Communications*, vol. 16(6), August 1998.

16. C.E. Perkins. IP Mobility Support for IPv4. *IETF Working Group: Network, Request for Comments (RFC) 3344*, August 2002.
17. R. Ramjee, T. La Porta, S. Thuel, K. Varadhan, and S.Y. Wang. HAWAII: A Domain-Based Approach for Supporting Mobility in Wide-area Wireless Networks. In *Proc. of Seventh Annual International Conference on Network Protocols*, November 1999.
18. R. Ramjee, T. La Porta, S. Thuel, K. Varadhan, and S.Y. Wang. IP micro-mobility support using HAWAII. *IETF Internet Draft*, 1999.
19. P. Reinbold and O. Bonaventure. A Survey of IP Micro-Mobility Protocols. CiteSeer: Scientific Literature Digital Library, <http://citeseer.nj.nec.com/reinbold02survey.html>, 2002.
20. A. K. Talukdar, B. R. Badrinath, and A. Acharya. Integrated Services Packet Networks with Mobile Hosts: Architecture and Performance. *Wireless Networks*, vol. 5(2), pp. 111–124, March 1999.
21. N. Tang, S. Tsui, and L. Wang. A Survey of Admission Control Algorithms. *Technical Report*, Course Project: A Survey of Admission Control Algorithms for Providing QoS over the Internet, Computer Science Department, University of California (UCLA), September 1998.
22. F. Yu and V.C.M. Leung. A framework of combining mobility management and connection admission control in wireless cellular networks. In *Proc. of IEEE ICC'01*, June 2001.
23. M. M. Zonoozi and P. Dassanayake. User mobility modeling and characterization of mobility patterns. *IEEE Journal on Selected Areas in Communications*, 15(7), September 1997.