



## **USAIA**

# **A Ubiquitous Service Access Internet Architecture**

**Joachim Sokol and Jörg Widmer\***

**TR-01-003**

**Jan. 2001**

### **Abstract**

The next generation Internet will provide high-quality, high-bandwidth connectivity. However, the important aspect of mobility is often neglected. Future Internet users will expect the availability of the full range of Internet applications regardless of the mode of access. We assume that mobile users in particular will use audio-based and video-based applications with specific QoS requirements. The support for these applications that exists in wired networks is therefore also necessary in next generation IP-based wireless networks. In this paper, we present a framework for the seamless integration of QoS and Mobility.

---

\* This material is based upon work supported by Siemens AG, Munich.

# USAIA

## Ubiquitous Service Access Internet Architecture

Joachim Sokol  
Siemens AG  
Corporate Technology

Jörg Widmer  
International Computer Science Institute  
Network Services and Application Group

Version 1.0B  
1/15/2001

### **Abstract**

The next generation Internet will provide high-quality, high-bandwidth connectivity. However, the important aspect of mobility is often neglected. Future Internet users will expect the availability of the full range of Internet applications regardless of the mode of access. We assume that mobile users in particular will use audio-based and video-based applications with specific QoS requirements. The support for these applications that exists in wired networks is therefore also necessary in next generation IP-based wireless networks. In this paper, we present a framework for the seamless integration of QoS and Mobility.

**Keywords** – continuous reservation, Differentiated Services, hierarchical Mobility Management, Integrated Services Model, Mobile IP, MPLS, QoS, RSVP

### **1. Introduction**

Recent initiatives to add mobility to the Internet and packet data services to third generation cellular systems are being considered by mobile service providers as possible candidate technologies for the delivery of IP data to mobile users. However, both of these candidate technologies have shortcomings. For instance, Mobile IP represents a simple and scalable global mobility solution but lacks support for fast handoff control, real-time location tracking, authentication and distributed policy management found in cellular networks today. In contrast, third generation cellular systems offer seamless mobility support but are built on complex and costly connection-oriented networking infrastructures that lack the inherent flexibility, robustness and scalability found in IP networks. Future wireless networks should be capable of combining the strengths of both approaches without inheriting their weaknesses. This motivates us to focus on all-IP networks to provide an integrated solution for the ongoing convergence of IP and mobile telecommunications in wireless networks. The inclusion of IP-centric mobile telecommunications networks presents a number of challenges that we want to address with respect to QoS. The aim is to enhance IP mobility-related solutions to deliver seamless mobility without losing the cost effectiveness, application flexibility and transparency of IP technologies. Broadband wireless router will be available in the near future and are currently being evaluated in field trials (e.g. WT2700 from Cisco [1]). Such devices will provide low-cost cellular wireless broadband access in all IP-based networks. Furthermore, networks with ubiquitous wireless Internet access are already available (e.g. TSUNAMI wireless networking [2]).

The main problem in the area of QoS and Mobility is the mobility management. Mobility management is the capability of the network to route packets to and from mobile nodes (MNs) or mobile networks. Mobility management must be fast enough to avoid noticeable disruptions in the requested service. Mobile IP [3] is not well suited for *all* kinds of movements (e.g. between cells, technologies, topologies or administrative domains) due to latency, scalability and reliability constraints [4]. To enable fast handoffs, a hierarchical mobility management is necessary. The handoff is the procedure to switch the context of a MN to another base station (BS) in case the signal power is falling below a certain threshold value.

The establishment of resource reservations for an arbitrary number of mobile and immobile users is an additional problem. The users must be able to express their required QoS, which can be achieved by a signaling protocol, e.g. RSVP [5]. Standard RSVP can only request resources on the previously established path between the sender and the requesting destination. When the movement of the MN makes it necessary to alter this path, the handoff procedure to the next BS has to be successfully finished before any resource

reservation is possible. This may result in unacceptable delays or degradation of the previously provided QoS.

To solve these problems in all-IP based networks, we have developed a framework for the seamless integration of QoS and Mobility. A major component of that framework is a control entity below IP in the end system. The control entity has to be capable of policy-based selection of a suitable subnetwork layer technology and/or provider. These policies may also take one or more of the following aspects into account:

- connectivity
- type of service requested
- movement pattern (slow, fast)
- communication costs

The added functionality of the control layer supports a novel QoS concept. We propose modifying the approach of *RSVP for mobile Hosts* described in [6]. The scope of RSVP is limited to the first hop within the network. Thus, if the first hop router is a BS, it has to act as an IP router. Our approach avoids the necessity of proxy agents and the specification of a mobile profile thus simplifying the signaling of QoS in advance signaling procedure. The establishment of reservation in advance in a wireless environment must be evaluated from two distinct aspects.

- The first aspect is related to the wireless topology and includes the possibility of contacting several network nodes (BSs) simultaneously at certain locations along the path a mobile node is traversing.
- The second aspect is related to the Call Admission Control (CAC) algorithm. It must provide a mechanism to share the resources so that mobile users with a previous reservation receive preferential treatment compared with users who are initially requesting new resources, thus avoiding distortions of the offered service on these requirements, we are proposing a new CAC procedure to decouple QoS and mobility. Our algorithm aims to offer nearly the same probability of service availability for mobile users along the entire path as compared with immobile users.

The currently available service classes of the Integrated Services Model do not take mobility aspects into account. We have assumed that the Controlled-Load Service Class [7] (CL) is the appropriate service class for mobile users. We suggest extending this service class to cover a range including the absolute lowest service level an application is willing to accept. There are two reasons for this:

- an increasing number of adaptive applications can adapt to a certain extent to different network situations, which is likely to happen more frequently in wireless environments. This is currently not reflected in the specification of the CL service class.
- to provide better network utilization without a complete distortion of the offered service, the CAC must be performed within a range of the service parameters.

To minimize noticeable disruption of the requested service, our framework suggests a new handoff protocol for the cellular level (cells within one subnetwork) based on the protocol described in [8]. We extend this protocol to provide all necessary information with respect to the integration of mobility and QoS. Therefore it is seamlessly aligned with our CAC and RSVP-based QoS signaling procedures. At the administrative domain (AD) level (several subnetworks belonging to one provider), we suggest two solutions. The first is a centralized approach and deals with hierarchical foreign agents; the second is a distributed approach based on MPLS. At the global level, we still see Mobile IP and its optimization as the best solution for mobile communication due to security considerations.

The combination of QoS and Mobility in our framework enables service providers to “advertise” the services they offer in their domain to visiting users. Service location and provisioning with respect to mobile users with an adequate level of QoS is the challenging task our framework focuses on. The study presents the envisioned network architecture and identifies the necessary work. It is organized as follows: The following chapter gives details about the mobility management and the QoS mechanisms. Chapter 3 describes the relevant protocols and the interaction between the network entities. In Chapter 4, we list the work to be done to evaluate our framework. Our concluding remarks are presented in Chapter 5.

## 2. USAIA Architecture

### 2.1 Mobility Management

The USAIA framework provides hierarchical mobility management that interacts with the provided QoS mechanisms in different network areas. We distinguish the following three areas for mobility management and QoS support:

- the cellular level for handoffs between adjacent cells belonging to the same subnetwork (here referred to as local level handoff)
- the administrative domain level for handoffs between different subnetworks under the control of one provider (here referred to as AD level handoff)
- the internetworking level for handoffs between administrative domains (here referred to as global level)

The USAIA framework deals with an all-IP based network over these areas. The overall structure is depicted in Figure 1. Note that BSs are viewed as routers connecting the wireless cellular network to the wired provider network, which itself is connected to the internetwork. At the cellular level, mobility management is performed with a fast handoff protocol between the MN and the BS without involving any agents. Essential QoS information is also provided to the MN to enable the MN to calculate whether sufficient resources are available.

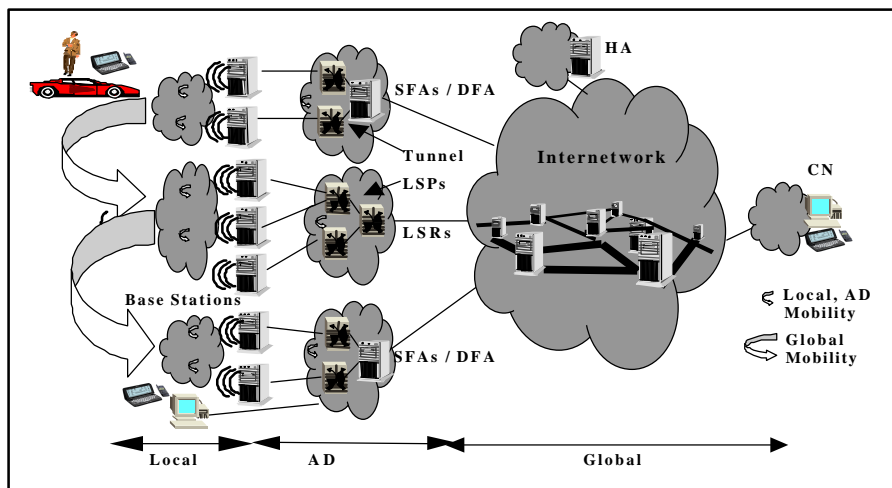


Figure 1: USAIA structure

The local handoff protocol operates as follows: MNs contact the next BS(s) along the movement path based on beacon signals sent from these BSs. The control layer within the MN triggers the handoff procedure by comparing signal strength and possibly other parameters. The handoff is initiated by the MN by transmitting an *MN\_ANNOUNCE* message to the new BS. This message contains the MN's own IP address as well as the address of the "old" BS. The *MN\_ANNOUNCE* message contains an indication as to whether the handoff has to take place immediately or is only used to set up resources in advance. To distinguish between the types we call the latter a *handoff announcement*, because it is not used for the packet data transfer. The BS acknowledges the *MN\_ANNOUNCE* message. If a real handoff occurs, the BS creates a routing table entry for this MN and uses proxy and gratuitous ARP [9] to adjust the routing table entries of all involved nodes within the subnetwork. The handoff protocol is described in more detail in Chapter 3.

Administrative Domain (AD) handoffs are based on the mobility between subnetworks within the same administrative domain (AD). There is no need to expose the type of movement to the home domain of the

MN, as it is required by Mobile IP. To handle the movements within an administrative domain, either a centralized or distributed approach is used.

The centralized approach is based on the concept of hierarchical Foreign Agents with Subnetwork Foreign Agents (SFA) and Domain Foreign Agents (DFA) [8]. As in Mobile IP, each subnetwork has at least one SFA. Furthermore, every domain provides a DFA, whose address is included in the advertisement messages of the SFAs. This address is used by MNs as their care-of address and will not be changed as long as they stay within the domain. The DFA maintains a per MN routing entry, which will be updated whenever a MN moves across subnetworks within its domain.

The distributed approach uses MPLS [10] with several egress routers acting as Label Switched Routers (LSR). A centralized component like the DFA is no longer necessary, thus eliminating a possible single point of failure. Instead of tunnel setup between the DFA and SFAs, the egress LSRs use MPLS to set up the routing information via Label Switched Paths (LSPs). The LSP could be either pre-configured or signaled between the relevant LSRs, which handle mobility management between subnetwork boundaries. This presupposes that the whole domain is MPLS-aware. We use RSVP (e.g. the extension of RSVP-TE) to signal label distribution, because RSVP is still supported on the cell level. With appropriate extensions RSVP can request and distribute MPLS labels, specify a specific path for an LSP, re-route LSP's using a make-before-break strategy, and preempt LSP's. Furthermore it include QoS information to enable resource allocation and LSP establishment to take place simultaneously. This property is used to seamlessly align the QoS requests from the MN via our modified RSVP over the cell level with the available resources on the AD level. We assume that the scalability concern of RSVP, a legitimate concern in IntServ applications, is likely not a factor in MPLS signaling, because the number of LSP's setup (and their frequency of setup and teardown) will likely be orders of magnitude less than per-flow resource control that occurs in similar size IntServ networks.

The next level handles global movement across administrative domains. Mobile IP is likely to be the right candidate. We assume that the Mobile IP implementation in the end system will use standardized extensions such as route optimization [12] and revised protocol operations [13]. Because of the increasing deployment of IP Version 6 and the resulting benefits with respect to mobility, we intend to support IP Version 6 in the end system.

## 2.2 QoS Concept

### 2.2.1 Resource Reservations

New generations of operating systems support RSVP as a signaling protocol to request resources from the network. For this reason, RSVP will be used as the signaling protocol between the MN and the BS. For non-RSVP-aware applications requiring a certain service, a proxy application can be used which deals with the necessary RSVP handling on behalf of those applications. Our modified RSVP is terminated at the first network node (BS)<sup>1</sup>, which is responsible for mapping RSVP to the available QoS mechanism provided at the AD level<sup>2</sup>. There are two reasons for this. Firstly, RSVP is considered to work well only in stub networks [14], which in our case is reduced to the scope of the wireless link. Secondly, the wireless link is likely to be the link with the lowest available bandwidth, so the quantitative description of the QoS parameter within RSVP allows for optimized usage of this scarce resource. We assume that the Integrated Service Model [15] applies to both the MN and the BS to improve the provided concept. Furthermore, even if it is not part of this study, the mapping of RSVP to the QoS capabilities of the underlying wireless technology is an important mechanism in the QoS concept.

Our modified RSVP is able to set up reservations in advance along the path traversed by the MN. These reservations are called passive reservations and the mechanism is based on the *handoff announcement* procedure. To set up resources in advance we distinguish between *active* and *passive associations* of MNs and BSs. Active associations are necessary for data packet exchange and QoS negotiations, whereas passive associations are restricted to QoS in advance announcements. Every MN has only one active association with one BS at any given time, but may have several simultaneous passive associations with

---

<sup>1</sup> Our framework is also applicable for the case, that the BS is not IP aware and RSVP is terminating at the first IP-aware hop in the access network

<sup>2</sup> Note that in our framework RSVP is suggested to support both, label distribution on the AD level (e.g. via RSVP-TE) and resource reservation on the wireless link via our modified RSVP

different BSs. Passive reservations are only possible after a successful handoff announcement in accordance with passive associations.

As stated earlier, the MN is a device that includes a control layer to “measure” signal strength and possibly other parameters. Therefore, the control layer is able to estimate the probability of a handoff. This enables the control layer to inform RSVP when to set up passive reservations at the new BS. For better network utilization these passive reservations can to be used by other flows until the BS receives the message *MN\_Announce* (triggered) from the MN. The control layer has to provide interfaces to RSVP, to the supported subnetwork layer technologies, and the RSVP-Proxy on the application layer as shown in Figure 2.

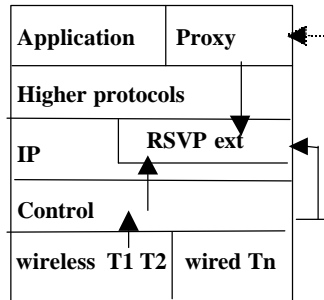


Figure 2: Mobile Node

Our resource reservation model functions as follows. RSVP within the MN is able to set up passive reservations as soon as it receives the notification from the control layer after a successful handoff announcement. These announcements are initiated by the control layer based on the signal strengths of the beacon messages. Beacon messages are periodically transmitted by every BS and the transmission rate is adjusted to the amount of available resources.

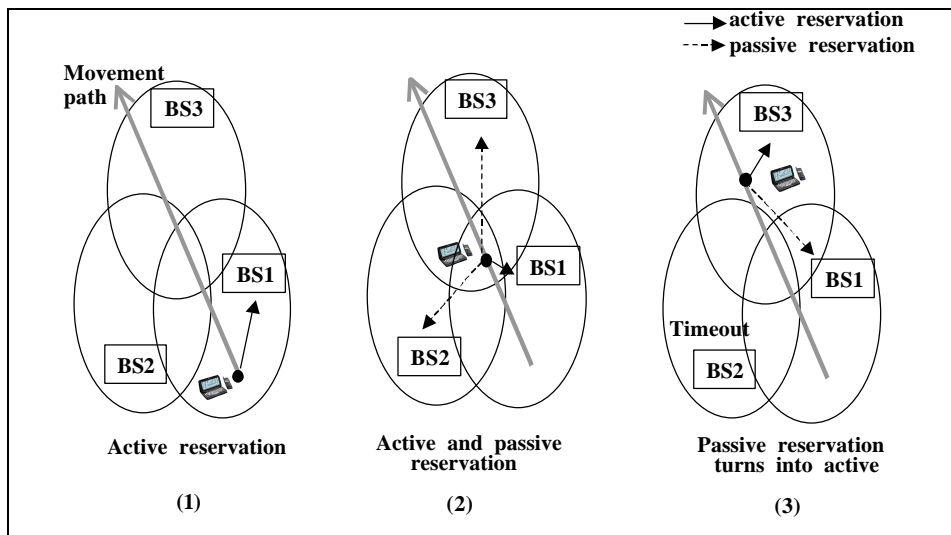


Figure 3: Active and passive reservations

A high rate is necessary to support MNs with real-time traffic, thus enabling them to find neighboring BSs very fast<sup>3</sup> to set up their requests for passive reservations. The rate is reduced when available resources are

<sup>3</sup> fast in the sense, that MNs are able to find at least 1 BS in the cell overlap area to request resources before the initiation of the handoff procedure is necessary based on signal strength constraints

scarce to provide load sharing between the BSs for best effort traffic<sup>4</sup>. Beacon messages contain information about available resources. This information is extracted by the control layer and delivered to RSVP in case an active reservation exists for the active association. After the real handoff occurs, the passive reservation becomes active.

Passive reservations are timed out based on a tunable RSVP retransmission timer. The time interval is usually higher than that of the standard RSVP timer. This can be justified by the fact that the bandwidth set aside for passive reservation is not used by the initiator until the reservation turns active.

Figure 3 shows a typical scenario for dealing with active and passive reservation. At position (1), the MN has an active association with BS1 and maintains an active reservation. No beacon signals are received from other BSs. The MN moves along the movement path and receives beacon signals from BS2 and BS3 after it reaches the cell overlapping area (2). Because the active reservation still exists with BS1, it decides to set up passive reservations to BS2 and BS3 respectively after the *handoff announcement* has taken place. Finally the MN moves towards BS3 and triggers the handoff with BS3. After this procedure is successfully finished, BS3 turns the passive reservation into an active reservation (3). The passive reservation at BS2 times out, since no refresh messages are received.

### 2.2.2 Call Admission Control

We suggest a call admission control procedure which takes into account that QoS must, as far as possible, be decoupled from the potential movement of the MNs. It is necessary to ensure fair sharing of resources between mobile and immobile users. The following metric is used to assess the impact of Mobility on QoS. Let  $Q_{(t_i, L_i, r)} \in \{0; 1\}$  be the QoS a user gets at location  $L_i$  for the time period  $t_i$  with an active association to the BS serving  $L_i$  for request  $r$ .  $Q_{(t_i, L_i, r)} = 1$  means the request can be satisfied,  $Q_{(t_i, L_i, r)} = 0$  means the available QoS is below the requested value  $r$ . Request  $r$  is defined as the absolute minimum level of QoS an application is willing to accept in relation to the initial request  $R$  at location  $L_1$ .  $L_1, L_2, \dots, L_n$  represents the location of BSs of an arbitrary path through the overlapping cell area which will be visited by a MN for the duration of an communication with resource reservation. Then the probability of achieving continuous QoS on location  $L_n$  and all previous visited locations ( $L_2, \dots, L_{n-1}$ ) should be nearly the same as the probability of getting the QoS provided at the first location  $L_1$  without any moving activity.

$$(1) \quad \prod_{i=2}^n E(Q_{(t_i, L_i, r)}) = \alpha E(Q_{(T, L_1, R)}) \quad \text{where } T = \sum_{i=1}^n t_i ;$$

Assuming that the QoS is either available or not at location  $L_i$  for the minimal accepted QoS level represented by request  $r$ , we also can write

$$(2) \quad E(Q_{(t_i, L_i, r)}) = P(Q_{(t_i, L_i, r)} = 1) ;$$

For  $\alpha = 1$  complete decoupling of QoS and Mobility is achieved. In our framework  $\alpha$  is optimized based on the following mechanisms:

- overlapping cell areas
- bandwidth partitioning
- extension of the Controlled-Load Service Class
- load-dependent refresh time interval for *reservations in advance* (see Chapter 3.2)

#### Overlapping cell areas

The framework relies heavily on the fact that at a certain boundary area within each single cell beacon signals of several BSs can be received. Otherwise, no *continuous reservations* can be set up. The optimization of the dimensioning of these overlapping areas for a specific radio technology is subject to simulations.

---

<sup>4</sup> MNs with BE traffic perform the handoff (based on the decreasing signal strength) to the first BS from which they received the beacon signal

### Bandwidth partitioning

Users are classified based on the type of traffic they transmit (real-time with QoS support, best effort) and movement pattern (immobile, mobile). In our framework, this classification is reflected in the partitioning of the bandwidth of the BSs (Figure 4).

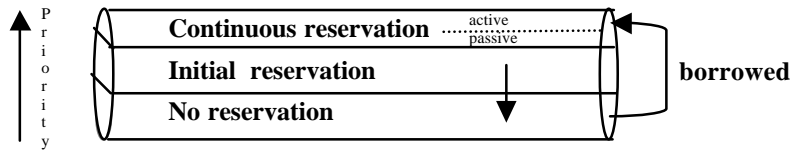


Figure 4: Bandwidth partitioning

- Reservations of users requesting new resources from the network fall into the *initial reservation* class.
- Active and passive reservations of users leaving the scope of their initial BS fall into the *continuous reservation* class.
- Reservations of users sharing the best effort delivery traffic fall into the *no reservation* class, independently of their movement pattern.

Any *initial reservation* can switch to a *continuous reservation* in case a passive reservation is requested for it. Once the handoff has occurred, the *initial reservation* is removed and the *continuous reservation* is changed from passive to active. To improve flexibility, bandwidth bounds do not have to be fixed but can be adjusted to the current load. For better network utilization of the network resources, best effort traffic can use the bandwidth set aside for passive reservations. As soon as a passive reservation is activated, this traffic will be discarded.

### Extension of the Controlled-Load Service Class

We suggest the following extension to the current Controlled-Load Service (CLS) class, motivated by the fact that many multimedia applications will be adaptive. We enhance the CLS class with request  $r$  for the specification of the absolute minimum an application is willing to accept, thus providing a range of QoS parameter values. For applications with hard guarantees the range is set to zero, for very elastic applications the range may cover a wide spectrum. As long as passive reservations can be admitted, they are maintained with request  $r$ . When a new reservation  $m_{new}$  is not accepted due to the lack of resources, every previously admitted passive reservation will be decreased by a relative value (percentage of the  $R$ ) such that the new flow can be admitted. Once enough resources are available again (above a certain threshold value), the reservations will be increased again accordingly. The MN will not be informed of changes of their passive reservations, because first this would require some sort of signaling over the wireless link and second it is unknown, whether the passive reservation will ever turn to an active one. Active reservations remain always unchanged.

### 2.2.3 QoS on the AD-level

Within the core network, we assume that either MPLS [10] or the Differentiated Service [16] approach is used to provide QoS. Depending on the mobility management scheme, the framework uses Differentiated Services coupled with the SFA and DFA concept. With MPLS, we use the inherent capability of MPLS to provide QoS with an appropriate setup of Label Switched Paths (LSP) between the Edge Label Switched Routers (LSRs). The LSPs are set up either in advance or signaled based on the movement of the Mobile Nodes. Note that QoS requirements and policies can be taken into account when setting up the path. Since RSVP is still available at the BS, the extended version RSVP-TE can be used as the signaling protocol for setting up LSPs. Nevertheless, static assignment or a mixture of both may also apply.

The alternative approach of Differentiated Services maps the RSVP messages and their related flows to the appropriate Differentiated Service class according to [17]. On the one hand, the task of dimensioning the Differentiated Services-capable domain is simplified by the use of a single DFA, because it is the ingress point for all traffic from the global level. On the other hand, the DFA is a single point of failure. Scaling issues may apply to the DFA, because the DFA must handle all routing entries efficiently for all MNs.



Which concepts should be used depends on the size of the network. For small networks, the DiffServ approach offers better performance, because there exists no MPLS signaling overhead.

	BS		DFA/MPLS LSR	
	Local	AD	AD	Global
UPSTREAM	IntServ/RSVP	DiffServ	DiffServ	TLA
	IntServ/RSVP	MPLS	MPLS	TLA
Downstream	IntServ/RSVP	DiffServ	DiffServ	TLA
	IntServ/RSVP	MPLS	MPLS	TLA

**Table 1: QoS mapping**

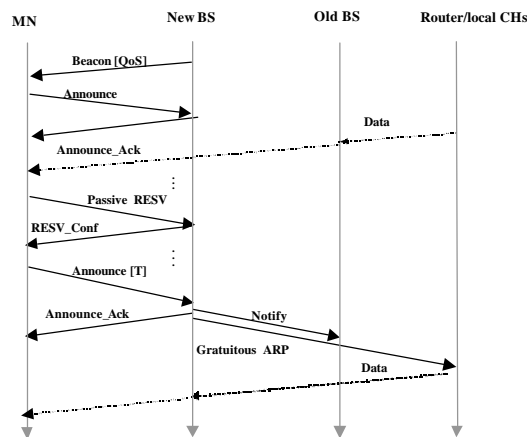
For large networks, we consider the MPLS approach to be more appropriate due to its scaling, dimensioning and reliability characteristics.

Connecting the core network to the global network is subject to Traffic Level Agreements (TLA) between the involved providers and is therefore beyond the scope of this study. In any case, the well-defined location of QoS mapping entities within USAIA framework simplifies the dimensioning process for the provider. The QoS mapping hierarchy with respect to the mobility management hierarchy is depicted in detail in Table 1.

### 3. Protocols

#### 3.1 Local Handoff Protocol

The most critical point in the USAIA architecture is the Local Handoff Protocol (LHP) between the MN and the BS with respect to support real-time traffic. The LHP must provide a very high performance to support fast handoffs. At the same time, it has to be simple and flexible enough to provide all the necessary information for QoS support and the handoff procedure. Figure 5 shows a typical sequence of control messages of the LHP.



**Figure 5: Local handoff protocol**

- The initial message is the beacon signal of a BS, containing the IP address of the BS that sends it. Beacon signals are sent via broadcast. The rate of the beacon signals is correlated to the current load and the available resources of the BS. To support fast handoffs, the beacon signal conveys a provider-defined share of the maximum of resources that can be requested by either during the handoff

procedure as a passive reservation or as an initial reservation, e.g. after the successful termination authentication process<sup>5</sup> after the MN entered the network.

- The control layer of the MN recognizes the overlap of BS regions and triggers the handoff from the current BS to the next one. Based on certain policies, the control layer uses different decision criteria for the kind of handoff. For instance, *smooth handoff* (performing handoff to a new BS and holding the old BS simultaneously) is appropriate if the new BS supports the requested QoS. *Lazy roaming* (performing handoff if the old BS is no longer accessible) could be appropriate if QoS was requested, but the new BS does not support QoS.
- The handoff is initiated by an *MN\_Announce* message to the new BS, carrying the MN's own address as well as that of the "old" BS. The *MN\_Announce* message contains an indication whether the handoff is marked as "triggered" (T-bit set) or whether it is a "handoff announcement" (T-bit not set). The first one initiates the handoff procedure and the latter one provides the possibility to request resources in advance without losing the connectivity to the current BS. Furthermore, the *MN\_Announce* message contains a flag that indicates whether the MN needs a reliable buffer delivery after the handoff to avoid any loss of data during the handoff.
- The BS confirms the *MN\_Announce* message with an *MN\_Announce\_Ack* message.
- If the *MN\_Announce* message is marked as "triggered", then the new BS sends a *Notify* message via the wired link to the old BS to inform it that the MN has moved. This message conveys the address of the new BS. The new BS also creates a routing table entry for the MN.
- Receiving a *Notify* message, the old BS deletes its routing table entry for the MN. If it still has packets for the MN in its retransmission buffer and the *Buffer\_Delivery* indication is set in the *Notify* message, it sends the buffer to the new BS for forwarding them to the MN.
- If the *MN\_Announce* message is not marked as triggered, the BS accepts either a retransmitted *MN\_Announce* message (T-bit set) to initiate the real handoff or QoS in advance requests via RSVP.
- All RSVP messages for reservations in advance are acknowledged by the BS by a *RSVP RESV Confirmation* message.
- The MN sends either a retransmitted *RSVP RESV* message to maintain passive reservations or a retransmitted *MN\_Announce* message marked as "triggered". The latter initializes the real handoff.
- On receiving a *MN\_Announce* message with T-bit set, the new BS turns the passive reservation into an active reservation and transmits a *Notify* message to the old BS (see above).
- Finally, the new BS broadcasts a gratuitous proxy ARP to map the MN IP address to the BS link layer address, thus forcing all involved nodes to update their ARP caches with that information. This mechanism prevents the chaining of several BSs.

### 3.2 RSVP

Communication between the MN and the BS with an active association is in compliance with standard RSVP processing. To support passive resource reservations in RSVP we suggest the following extensions:

- Passive reservations must be accepted by BSs without any previously established PATH state. The scope of these passive reservations is restricted to the MN and the BS. After the passive reservation is admitted, the BS sets the corresponding refresh timer to control the lifetime of the passive reservation. Thus, in the same way as with active reservations, the MN must refresh passive reservations.
- Passive reservations can be activated at a BS without any previously received PATH messages, because the modification of the end-to-end path is restricted only to the link between this BS and the MN.
- Both initial passive reservation requests and refresh messages are acknowledged by the BS with an *RSVP RESV Confirmation* message to ensure that the MN has up-to-date information about available passive reservations.
- We suggest using a range for the refresh time interval to maintain passive reservations. Every BS adjusts this time interval individually to the *current utilization* of the passive reservation part of the *Continuous Reservation Partition*. The reduction of the refresh time interval results in the removal of obsolete passive reservations at an earlier stage, thus allowing new passive reservations to be admitted. An increased utilization results in a reduced timer value when the utilization is between the threshold

---

<sup>5</sup> The necessary AAA procedures for USAIA are out of scope of this document

values  $TPR_{low}$  and  $TPR_{high}$ . Below  $TPR_{low}$  the time interval is set to the maximum value. This interval is longer than the interval for active reservations, because the bandwidth partitions of passive reservations can be used by best effort flows. When the utilization is above  $TPR_{high}$ , the minimum time interval value is used. Timer values of flows with previously passive reservations are adjusted after receipt of the refresh *RSVP RESV* message from the MN. The refresh interval is announced by a *RSVP RESV Confirmation* message containing the *Time Values* object of RSVP.

- The early removal of obsolete passive reservations is achieved at the cost of additional performance and bandwidth overhead due to a higher retransmission rate. Therefore, we control the retransmission rate with an additional threshold value  $TLO_{low}$  based on the *current load* of the BS. A load higher than  $TLO_{low}$  prevents the BS increasing the current value of the refresh timer towards  $TPR_{high}$ . More sophisticated algorithms for controlling the time interval based on the combination of the utilization (passive continuous reservation partition) and the load (active, initial, and no reservation partition) are considered as further optimizations.

In addition, we intend to use optimizations for RSVP which are currently being discussed within the IETF [18]. These optimizations minimize the overhead of exchanged refresh messages and avoid the latency of RSVP signaling.

### 3.3. Additional Protocols

All other protocols used at the AD and global levels, including MPLS, DiffServ and the hierarchical mobility management based on Mobile IP (SFAs, DFA), are considered to fit into our framework without modifying the corresponding standards. The seamless inter-working between these protocols and their appropriate mapping is subject to current standardization efforts.

## 4. Future Work

With the mechanisms described in this study, we intend to provide a platform for seamless Mobility and QoS support. Nevertheless, the whole approach must be evaluated for its applicability. Simulations and partial prototype implementations have to be investigated to achieve results concerning scalability, reliability, service and QoS availability of our approach. The achieved results will be published in other papers. Furthermore the mapping of our modified RSVP to underlying wireless technologies, the seamless integration of our modified RSVP with RSVP-TE in the BS, and the mapping of the requested QoS on the wireless link to appropriate QoS support on the AD level must be investigated.

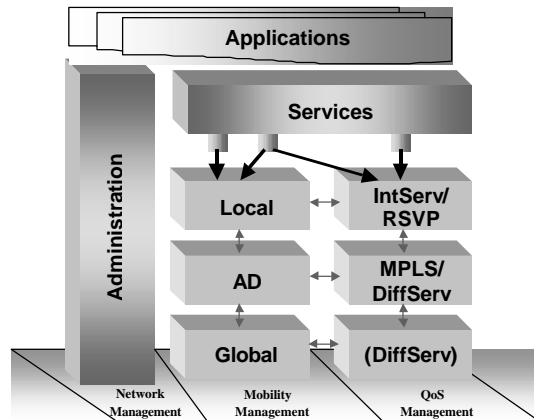
Our intended prototype implementation of the MN (with modified RSVP and control layer) and BS (with modified RSVP and CAC) will be based on *Linux* and wireless LAN according to IEEE 802.11. Furthermore, simulations with *ns* [19] and a Petri net based evaluation with the toolkit TimeNET<sup>6</sup> will be used to examine all mentioned topics in more detail. Attention is drawn to the fact that our concept is subject to change if certain simulation and evaluation results require adjustments.

## 5. Conclusion

In our study, we have defined a framework that permits the provision of QoS independently of the movement of users. On each identified network level (cell, AD, global) we have well-defined QoS mechanisms in conjunction with appropriate mobility management schemes. The interrelationship between these different network levels is depicted in Figure 6. The well-defined QoS mapping entities in the network also simplify QoS provisioning, thus easing the task of providers to realize services within their AD, e.g. providing services according WOS [20]. Furthermore applications will be assisted to adjust in advance their QoS constraints to the available network resources due to the reception of this information in the beacon signals, thus eliminating unnecessary QoS signaling over the wireless link.

---

<sup>6</sup> TimeNET was developed at the Technische Universität of Berlin, Germany



**Figure 6: USAIA planes**

Furthermore USAIA will help to create sophisticated applications with respect to mobility support, because the “feedback channel” via our modified RSVP are able to inform them in the worst case, that no reservations in advance are possible, which could result in appropriate beneficial actions (e.g. the user is informed that the QoS could be subject to degradation in case of further movement). Also the integration of the capabilities of applications to adapt to changing network situations is reflected by USAIA in an efficient and transparent way for better network utilization with the possibility to signal a range of QoS parameters. The framework uses standardized solutions whenever it is possible and we adjust and extend them only when necessary.

## 6. References

- [1] Cisco WT2700 wireless product, “<http://www.cisco.com/warp/public/779/servpro/solutions/wireless/>”
- [2] TSUNAMI Wireless networks, “<http://gotsunami.net/tsunamipress.htm>”
- [3] Perkins et. al., “IP Mobility Support”, RFC 2002, October 1996
- [4] Vatn, Maquire, “The effect of using co-located care-of addresses on macro handover latency”
- [5] Braden et. al., “Resource reSerVation Protocol – Version 1 Functional Specification, RFC 2205 September 1997
- [6] Badrinath, Acharya, “MRSVP: A Resource Reservation Protocol for an Integrated Services Network with Mobile Hosts
- [7] Wroclawski, “Specification of the Controlled-Load Network Element Service”, RFC2211, September 1997
- [8] Caceres, Padmanabhan, “Fast and Scalable Handoffs for Wireless Internetworks”, November 1996
- [9] Plummer, “An Ethernet address resolution protocol”, RFC826, November 1982
- [10] Rosen, et al., “Multiprotocol Label Switching Architecture”, Internet Draft, August 1999
- [12] Perkins, Johnson, “Route Optimization in Mobile IP”, Internet Draft, February 2000
- [13] Perkins, “IP Mobility Support for IPv4, revised”, Internet Draft, January 2000

- [14] Mankin, et al., “Resource ReSerVation Protocol (RSVP) Version 1 Applicability Statement. Some Guidelines on Deployment”, RFC 2208, September 1997
- [15] Clark, Shenker, “Integrated services in the internet architecture: an overview”, RFC 1633, June 1994
- [16] Blake, et al., “An Architecture for Differentiated Services”, RFC 2475, December 1998
- [17] Bernet, et al., “Integrated Services Operation over DiffServ Networks”, Internet Draft, June 1999
- [18] Wang, Terzis, Zhang, “RSVP Refresh Overhead Reduction by State Compression”, Internet Draft, March 2000
- [19] Fall, et al., “ns Notes and Documentation”, VINT Project, UC Berkeley, LBL, USC/ISI, Xerox PARC, October 1999
- [20] Babin, Kropf, Unger, “A Two-Level Communication Protocol for a Web Operating System (WOS), 24<sup>th</sup> Euromico, 1998