Wireless Pers Commun (2009) 48:7–32
DOI 10.1007/s11277-007-9431-8

Multi-access Management in Heterogeneous Networks

Joachim Sachs · Mikael Prytz · Jens Gebert

Published online: 24 November 2007

Abstract Communication networks and mobile devices integrate an increasing number of access technologies. At the same time, new business roles emerge, which lead to new cooperation schemes between access providers providing different types of access connectivity. As a result, a variety of access technologies will be available for users at the same time. In this article we present an architecture and a framework capable of integrating different access systems into a multi-access system and selecting the best suited access for users. A utility-based approach is proposed for the evaluation of different access allocation choices, which is based on user and network policies, the performance of access bearers, and the availability of access resources. We present a general multi-access management framework, which integrates the different multi-access related functions: access detection, access evaluation and access selection, which can then lead to an access handover.

Keywords Heterogeneous networks · Multi-radio access · Access selection

1 Introduction

The area of wireless communications has seen a tremendous expansion in the last 15 years; the number of subscribed users to cellular communication services alone exceeded 3 billion by mid-2007. This growth has been accompanied by an increasing diversity of radio access technologies that are being developed and deployed. For instance, wide-area cellular mobile

J. Sachs (✉)
Ericsson Research, Ericsson Allee 1, 52134 Herzogenrath, Germany
e-mail: Joachim.Sachs@ericsson.com

M. Prytz
Ericsson Research, Torshamnsgatan 23, 164 80 Stockholm, Sweden
e-mail: Mikael.Prytz@ericsson.com

J. Gebert
Alcatel-Lucent, Lorenzstrasse 10, 70435 Stuttgart, Germany
e-mail: J.Gebert@alcatel-lucent.de

communications systems are based on 2G (e.g., GSM, EGPRS) and 3G (e.g., UMTS, HSPA) radio access technologies, with the next evolution (long-term evolution, LTE) already under development. Similarly, the number of wireless local-area access networks has increased enormously. These are mainly based on different versions of the IEEE 802.11 standards and are often operated by private persons.

Further, regional-area networks are being deployed, for example based on IEEE 802.16. In parallel on the fixed access side, a substantial and growing number of households is connected via fixed broadband access to the Internet—often extended with wireless local area networks. While access to packet data networks, like the Internet, is getting more and more ubiquitous, the heterogeneity of access technologies increases constantly, and mobile devices, like laptops or smartphones, integrate a growing number of access technologies. This development requires an increasing effort in managing the complexity of integrating these heterogeneous access technologies into so-called multi-access systems.

For network operators, a key driver motivating multi-access systems is deployment cost. Different access technologies have different characteristics and properties, like capacity and range. In a network with multiple access technologies being jointly deployed, the access properties can be exploited by placing the radio access points of each different access technology where it is the most efficient one. If mobile devices can switch the connectivity between different access technologies, it is not required to provide full coverage with all access technologies. Load balancing between different access systems allows to reduce the margin of spare capacity in each access system, and enables more cost efficient network deployment. The multi-access system extends the service coverage of an access system to areas where other access technologies are available. Lowered deployment costs may be viewed as mainly an operator benefit, but on a competitive market it will ultimately benefit end users. In an environment where connectivity is provided by different business entities, cost savings may be even more important since each actor (business entity) targets smaller segments that may not be financially viable on their own. Several quantitative studies have shown that deployment cost can be reduced substantially in suburban and urban areas with typical geographical traffic demand patterns if a multi-access system is used [1–6].

For end users, the benefit of multi-access systems is not only reduced costs and a choice of selecting access providers but also an increased service experience due to the fact that multi-access provides a continuous service area out of heterogeneous access networks. The user can thus exploit the full access capabilities when they are available. End users are *always best connected* [7] with an *always best experience*.

Standardisation efforts to integrate different access technologies have started in several standardisation fora, like 3GPP [8], IEEE [9], with a focus on architecture design. A general approach towards managing and integrating heterogeneous technologies—denoted as *Ambient Networking*—has been proposed in [10], which develops a common network control framework, the Ambient Control Space, for configuration and management of heterogeneous networks and cooperation and inter-working between different networks. This article describes the part of the Ambient Control Space that manages different available accesses for a user.

The article is structured as follows: in Sect. 2 we introduce the business environment and business scenarios on which a multi-access system is based. In Sect. 3 we present the multi-access system model in which the multi-access control functions operate, and in Sect. 4 we introduce the multi-access architecture. Section 5 presents a general framework to manage different accesses. It is followed in Sect. 6 by a discussion about the objectives of access selection and the evaluation of the utilities of accesses selection choices. The article is summarised in Sect. 7.

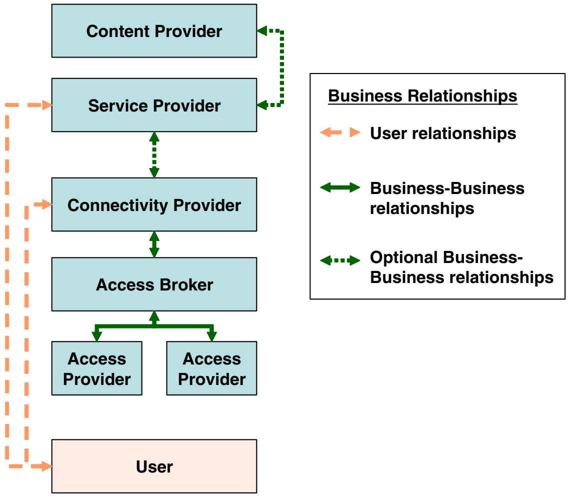

Fig. 1 Business roles and business relationships

2 Business Environment

In a multi-access environment connectivity can be provided by different business entities. For example, different network operators can provide access connectivity to a user. In order to understand the motivation and objectives for performing access selection we have to understand the involved business players and their interests and relationships. In this section we define business roles and discuss how these roles are provided by business entities in different business scenarios. We discuss objectives and utilities for the different business roles.

2.1 Business Roles

Different roles can be distinguished in the service provisioning chain that provides a service to the end-user. In the following we discuss the elementary roles of a communication chain, as depicted in Fig. 1, in the context of enabling a user to utilise the best suited access. The different business roles can be taken by different business entities in varying business scenarios, as we describe later.

The *content provider* provides the content transmitted in a communication session. Content can be either stored or live-generated. Examples for stored data are music download, mobile-TV, video-on-demand, or general data exchange, when data is uploaded to or downloaded from a remote storage (messaging, file access, blogging, WWW-access, data backup). For live-generated content, the communication peers provide the content directly, as for example in voice or video telephony.

The *service provider* provides the communication service to the end-user. It provides content in a useful format for the end-user and manages the communication sessions. One example is telephony service providers that offer audio-/video telephony, telephone conferences, or group communication (e.g., push-over-cellular), which can be enriched with messaging, and file sharing. Another example is the provisioning of mobile-TV, video or audio services to the end-user in suitable formats. The service provider has a direct business relationship with the end-user and it typically also acts as re-seller of the content stemming from third party content providers.

The *connectivity provider* provides connectivity for the end-user to external networks, like the internet, corporate networks or telephony networks. Further it provides reachability for the user, so that a communication session can be established when initiated from communication peers. It coordinates the different accesses available to the user and enables dynamically changing the access in use.

The *access provider* provides the access connectivity to a user. For different types of access technologies the coverage of the access service differs. It can be on local, regional or national level.

The *access broker* facilitates the business cooperation of access providers with connectivity providers. Connectivity providers typically operate on national level, or at least covering a larger region. Access providers can, in contrast, operate in small local areas only, in particular if based on local area radio access technologies. Access brokers bundle a number of access providers into a business relationship with the connectivity provider, and thus circumvent a large number of peered business relationships between access providers and connectivity providers.

The user compensates the received services in monetary form. For the communication connectivity, the user pays the connectivity provider. This compensation can be either per service, per amount of resources used during the session, a flat-rate compensation package, or a combination of all these types. The connectivity provider distributes a part of the income to the access broker, which in turn provides compensation to the access providers. For the communication service, the user compensates the service provider. If the content of the communication service is not for free, the service providers compensate the content provider. These business-to-business compensation schemes can be based on the real contributions of different entities in the provisioning of the end-user service, or based on a compensation package. If a business relationship exists between the service provider and the connectivity provider, a discount on the costs can be provided to the user for the bundled communication and connectivity services.

If different business entities provide complementary functionality they can reduce the costs of their operation. For example, if different access providers supply access in different regions or have only limited capacity in their networks, they can, by cooperation, provide access with wider coverage and increased capacity. Thereby each access provider can save investment costs, which is beneficial in particular for new access providers with high start-up investment costs. On the other hand, every business relationship is associated with a certain overhead. First, some technical functionality is required to enable the cooperation; and second, the setup of business relationships involves certain costs.

2.2 Business Scenarios

In traditional cellular networks the mobile network operator combines and integrates all the business roles described above. Roaming agreements, in addition, provide a means for cooperation between multiple mobile network operators. A visited network, which has a roaming relationship to the home network of the user, takes the access provider role towards the end user. The connectivity provider role remains either in the home network or it can be shared between the visited network and the home network. The development of new communication technologies lead to the appearance of new types of networks. Local and regional access networks are provided by private persons, corporations, municipalities or cities. These access networks can be either wireless networks, for example based on WiFi or WiMAX wireless access technologies, or fixed access, for example based on xDSL or fibre. Also new service providers appear that are decoupled from connectivity and access provisioning, for example, voice-over-IP service providers. The partial disintegration of the communications market leads to some separation of business roles. As a result new business entities can emerge, which provide some of these different roles. The development of business scenarios has been investigated and discussed in [11–15]. To manage the increasing heterogeneity and resulting complexity, convergence of networking technology is required. To facilitate the diversity in business scenarios, dynamic cooperation of networking domains is a pre-requisite, which is a key concept developed in [10,16]. Business relationships can be diverse and complex, covering the complete range from competitive to cooperative scenarios. In this work we limit ourselves to cooperative scenarios. The business scenarios considered in our multi-access setting comprise a single operator with multiple access technologies integrated into a common packet core network. In addition, there can be different networks each comprising one or more access technologies. The cooperation can be similar to today's roaming scenarios between "equal" operators, but can also include cooperation with small access providers. It can also include the case of an untrusted access, without direct cooperation between the access provider and the home connectivity/service provider. These different cooperation scenarios can be realised with the evolved packet core network architecture that is currently developed in 3GPP [8,17,18].

3 Multi-access System Model

3.1 Requirements

This article considers multi-access systems constructed to operate with *any kind of access technology*, which can be very different from each other in their designs. To put this into praxis it is necessary to build the multi-access system around a *generic model* of what access technologies are capable of and how they work. A key component in such a model is a representation of how access technologies keep track of, and manage application data flows. The model needs to capture how an application data flow is characterized from a policy and quality-of-service requirements point of view, how an application data flow is addressed and located, and what application data flow management procedures to use when enforcing multi-access decisions. It is also necessary to model the capabilities of access technologies in relation to the policy and quality-of-service requirements model of an application data flow. This capability model can be used as a first filter when validating the accesses that can be used for some particular application data flow.

Another key component in the model concerns the performance of an access, that is, how good and efficient it is at fulfilling the application data flow requirements. The access performance is relevant both for an access in use as well as for candidate alternative accesses. A multi-access system needs measurements of access performance frequently, at least in the order of application data flow session durations; how often depends on the access performance variability and the application data flow requirements. In many access technologies and access networks the performance is limited by one bottleneck link, for example, the radio link in cellular radio access networks. In such cases the term link performance is often used instead of access performance.

A resource-aware multi-access system also needs a model of access technologies' resource structures. It needs to know how many resources there are, how the resources are distributed geographically and among network nodes, and in what way an application data flow consumes resources. An access technology manages its resources on its own (to varying extents) using Access Resource Management (ARM) procedures, in wireless systems usually called Radio

Resource Management (RRM). The particular ARM/RRM scope of operation needs to be reflected in the multi-access system resource model. It is also advantageous if the multi-access system can interact with the access technology specific ARM/RRM functionality.

In summary, the multi-access systems considered in this article require abstraction models of the application data flow (requirements, addressing) and the access technology characteristics (capabilities, resource structures, performance, control procedures). The following sections describe these abstractions in more detail.

3.2 Connectivity Abstraction

A multi-access system is typically only one part of a larger communication system in which an application data flow is being transmitted. The application data flow will *pass through* the multi-access system, but the end-systems, or at least one of them, are not necessarily visible to the multi-access system. A generic connectivity abstraction model that can be used for modelling multi-access systems has been developed in [16,19,20]. We extend this connectivity abstraction model and apply it to the terminology and connectivity framework used in 3GPP [21–24]. This multi-access connectivity abstraction is presented in Fig. 2.

The *application data flow (ADF)* is the data connection used between two application entities located at the end-systems to transmit application data of an application session. An application data flow is unidirectional, and for most applications bidirectional connectivity is provided by a pair of application data flows. A typical example of an application data flow is an IP flow, that is, the communication connecting two IP sockets in the end-systems. The path through the network is determined by the IP routing infrastructure. Within the network, the application data flow can be transmitted by different means. This depends on the network domains and the technology used therein. In the easiest case the ADF is directly routed end-to-end in an IP infrastructure. In reality, in many network domains on the end-to-end path, transmission is done via different communication technologies in different network domains. Within the multi-access network, multiple application data flows which are governed by the same policy rules and which have similar quality of service requirements are combined into a *service data flow* (*SDF*). The SDF is defined by a template of SDF filters, which characterise the application data flows that are part of the SDF. Such filters operate on the header fields of the IP data packets of the ADF, for example the source and destination addresses, the identification field of the higher layer protocol, the source and destination port numbers, the flow identifier and type of service information (like DiffServ Code Point [25]). The SDF embeds all application data flows which match the template of SDF filters. For every service data flow, quality of service requirements and policy rules are determined by the policy decision function (PDF)—for example the policy control and charging rules function (PCRF) in [22]—and enforced by the policy enforcement function (PEF)—for example the policy and charging enforcement function (PCEF) in [22].

An interaction with an application function (AF) enables the PDF to obtain application level session information, like service requirements. An example for an application function is a session control function in the IP Multimedia Subsystem (IMS), which allows to specify the requirements of user services, like voice-over-IP (VoIP) or multimedia telephony (MMTel) services. Not all application sessions support an application function. The service requirements for those application data flows can be signalled in-band via the RSVP [26] or NSIS [27] protocol, or are marked in the packet headers via DiffServ code points [25]. If the service requirements cannot be determined, best-effort service requirements are assumed as

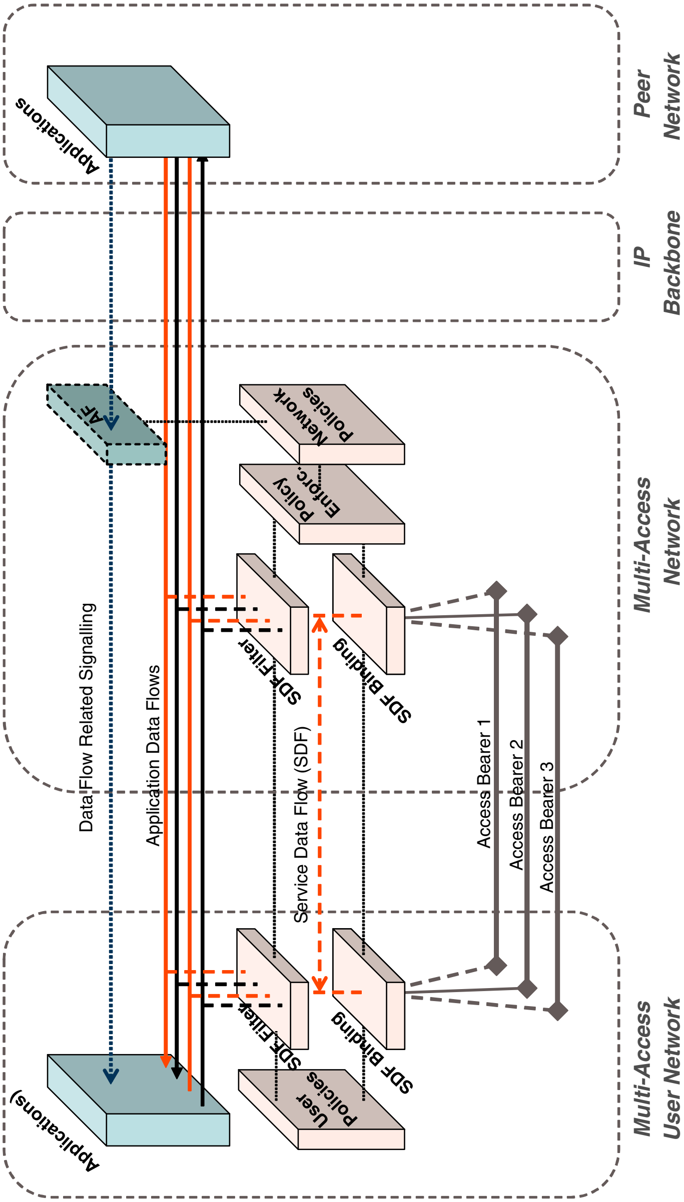

Fig. 2 Connectivity abstraction model for a multi-access system

default. A service data flow is mapped to an *access bearer*.[1] In a multi-access system multiple such access bearers are available. The *SDF binding* determines via which access bearer the SDF is transmitted. In a multi-access system the available accesses may belong to different (access) network domains, and hence, use different schemes for transporting the SDF within access bearers. For example, in order to support user mobility the transmission path of an access bearer is controlled by a mobility protocol (e.g., GTP [28], Mobile IP [29,30], NETLMM [31], PMIP [32,33]). For traffic engineering purposes or for virtual bridging of network domains, routing can be controlled via for example MPLS [34], GRE [35] or IPsec [36]. For the access specific transmission over the last hop, typically access technology specific bearers are used, like radio access bearers as defined for UMTS and LTE [37,38]. Thus, an access bearer can stretch over multiple technology domains, and in each domain it is embedded into a technology-specific bearer. This leads to a hierarchical mapping of access bearers to technology-specific bearers. Figure 3 depicts different access bearer realisations as specified by 3GPP, like a 2G EGPRS access bearer [23], a generic access network (GAN) access bearer [39], a 3G UMTS access bearer [21], an Interworking-WLAN (IWLAN) access bearer [40], a long-term evolution (LTE) access bearer [17] and a non-3GPP access bearer [18] that is based on any other access technology. The multi-access system can manage bearers (setup, monitoring, and tear-down) in general, regardless of the access technology. It is not necessarily the case that all multiple access bearers are maintained at the same time. In some cases the user network has not attached to multiple access networks but has detected the availability of different access networks. In this case, access bearers are not yet established and the multi-access system needs to keep track of this on the user network level.

In Fig. 2 the applications are located in a multi-access terminal (e.g., a mobile phone or a laptop) that has attached to multiple available access networks and established access bearers. However, the application can also be located in other nodes. For example, multiple devices can be connected to a personal area network or moving network. The node which terminates the service data flow and contains the SDF filters and bindings acts as a gateway or bridge for the moving network to the multi-access system. We use the general term user network to describe the user controlled device(s) which connect to the multi-access system.

3.3 Service Data Flow Requirements

A service data flow contains multiple application data flows that have the same quality of service requirements and that abide by the same policy rules. The SDF maintains a generic description of the SDF requirements that are relevant for determining the suitability of an access bearer for the service data flow. The suitability of the access bearer is mainly determined by the performance provided by the access bearer, which is typically governed by the performance of the (bottleneck) access link.
Three main requirement types are defined for a service data flow:

- data rate $b_{\min}$; discrete or elastic,
- delay $d_{\max}$; delay sensitive or delay insensitive,
- (residual) error ratio $e_{\max}$; error sensitive or error tolerant.

All applications have a strictly positive minimum average data rate requirement $\bar{b}_{\min}$ since, even for best effort applications, there is some practical time limit beyond which the communication session becomes obsolete (for example, user patience, device battery life, age of information). The suitability of a service can be expressed by a utility value, which is a function of the performance metrics of the access bearer: the data rate, the delay and the

[1] In 3GPP terminology, an access bearer is denoted as *IP connectivity access network (IP-CAN) bearer*.

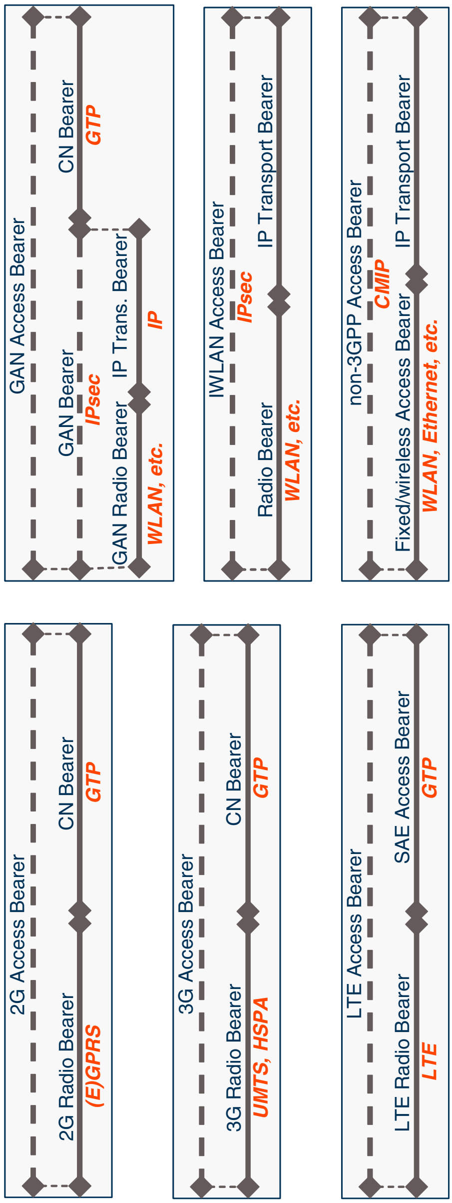

Fig. 3 Different access bearers based on technology-specific realizations

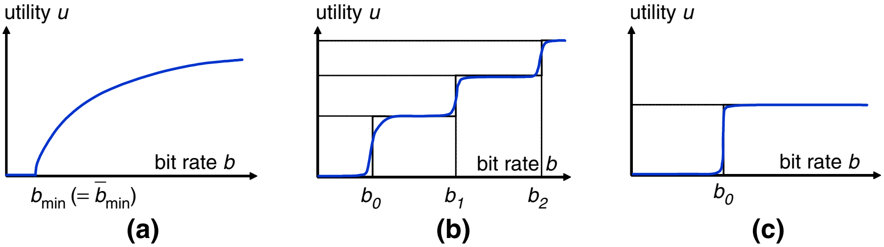

Fig. 4 Utility of a service data flow depending on the data rate of the access bearer for (**a**) elastic traffic, (**b**) discrete services, (**c**) speech telephony

reliability. The service requirements of a SDF determine the shape of the utility functions. A study and evaluation of suitable utility functions is beyond the scope of this article. However, we briefly discuss the characteristics of the utility functions.

The utility of any service data flow increases with increasing data rate. For service data flows carrying *elastic traffic* the utility function is increasing with diminishing marginal increase, i.e., the function is concave as depicted in Fig. 4(a). Examples of applications for elastic traffic are internet applications using the transmission control protocol (TCP), like file transfer and web surfing. The larger the obtained data rate, the larger the utility. Other types of service data flows carry *discrete services*. With discrete services we mean any service that has a number of pre-determined data rates at which the utility of the service data flow increases sharply. Examples of such services are rate-adaptive video telephony or video streaming applications, where the video is encoded in discrete data rates. Whenever, the data rate provided by the access bearer reaches a next higher data rate requirement, the application can switch to the higher-encoded video stream. The corresponding utility for the service data flow is depicted in Fig. 4(b). The utility has discrete service rate values b_0 to b_n at which a significant increase in utility is perceived. Depending on if audio or video data is variable-rate encoded or has a constant service rate, the utility function follows more or less the step function. A special case of a discrete service with a single service rate—like a speech telephony service—is depicted in Fig. 4(c). If a speech service does not achieve the required service rate b_0, the utility drops to zero. In environments with scarce capacity, as typically the case in access systems, admission control is a key resource management function to limit the amount of services sharing the bottleneck resource. If the number of services exceeds a threshold, several of them drop below the critical rate r_0 and the utility drops to zero.

The descriptor of the data rate requirements of the SDF includes an indicator whether it is of the discrete or the elastic type, and the set of service rates b_i for elastic services.

As with data rate, all applications have some (large) upper bound on the maximum acceptable delay $d_{\max}$ beyond which the communication sessions become obsolete. Some applications have much stricter requirements. For delay sensitive applications, like voice/video telephony and other interactive media sessions such as multi-player on-line games, the maximum delay is upper-bounded by roughly 200 ms. The delay requirements of streaming applications are determined by the length of the play-out buffer size, which is in the order of several seconds. Delay insensitive applications, for example, file transfer, have requirements that are orders of magnitude larger than typical transmission delays. As with the data rate requirements types there is a difference between delay sensitive and insensitive applications with respect to the utility of the provided delay: the delay sensitive type shows an approximately step-wise utility profile (similar to Fig. 4(c) but using $1/d$ on the x-axis), while the delay insensitive type is gradually affected by the delay (similar to Fig. 4(a), again using 1/d

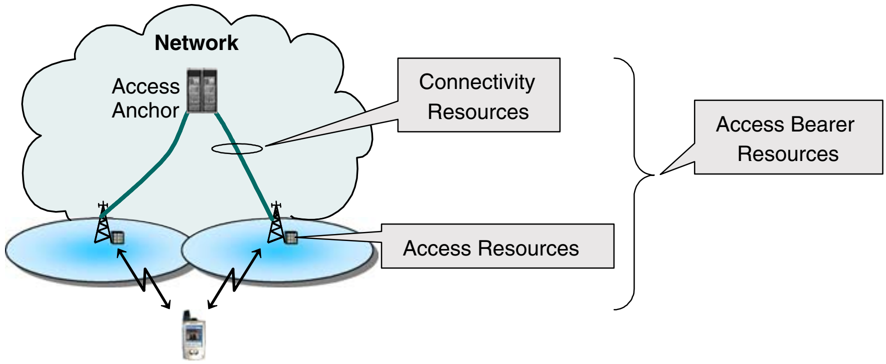

Fig. 5 Access bearer resources

on x-axis). The SDF descriptor includes an indicator whether it is of the delay sensitive or insensitive type (to flag the utility behaviour) and the $d_{\max}$ value.

Finally, applications have requirements on the (remaining) error rate $e_{\max}$ provided by an access bearer. In general application data can be corrupted during communications transmission. Access technologies employ error detection (CRC) and error correction (forward, FEC, and backward, ARQ) mechanisms to control the rate of erroneous data bits (or data packets); still a residual amount of errors can remain and thus become noticeable by the application. Error tolerant applications, for example, voice, audio, and video, can accept a certain amount of bit (or packet) errors, which may lead to some tolerable signal distortion that the end user may not perceive. Setting the error control mechanisms to accept errors at the tolerable rate may allow the access technology to use its resources more efficiently. Error sensitive applications, for example, file transfers have, on the other hand, basically zero tolerance for errors as a single bit error may invalidate a whole file. The SDF descriptor includes an indicator whether it is of the error tolerant or sensitive type and, for error tolerant, the $e_{\max}$ value.

3.4 Access Bearer Resources

An access bearer is provided by different connectivity systems that provide the necessary transmission resources. In general, these resources can be divided into access resources and connectivity resources as depicted in Fig. 5. Access resources provide the direct connectivity for the end user to the multi-access infrastructure, i.e., to the point of attachment. Access resources can provide either fixed connectivity, like DSL or fibre, or they provide wireless connectivity. Connectivity resources connect the point of attachment to the (multi-)access anchor, which serves as mobility anchor and provides connectivity to external networks. The access anchor contains the service data flow filters and bindings (cf. Fig. 2).

3.4.1 Access Resources

An abstract description of an *access resource* (AR) is necessary to capture the resource structure of different access technologies, in part to know the current capabilities (available resources) for handling service data flows, and in part to support operation when there is no service data flow active but an access system has been detected (for example, through reception of a beacon). The AR is a resource on which an access link can be established. This access link is part of the connectivity provided for the access bearer. The access bearer can span further than the access link (for example, to an anchor node) and may use other (non-access) connectivity resources for the remaining (non-access) connectivity. In wireless networks the AR corresponds to the radio resources of a radio cell, where the radio resources are allocated to active access links using some multiple access scheme (for example, TDMA, FDMA, CDMA, SDMA, or some combination thereof). In a fixed network the AR can correspond to, for example, the resources (transport, ports) of a DSL Aggregator/Multiplexing (DSLAM) node.

At the setup of an access bearer for a particular service data flow, admission control can be performed and access resources can be reserved. If resource reservation and admission control is used or not is dependent on the type of access technology and the service requirements associated with the service data flow. For best-effort service data flows typically no reservation is performed. In the case of service data flows with minimum quality of service requirements, 2G, 3G and LTE access bearers (cf. Fig. 3) apply admission control and resource reservations [17,24]. For many other access technologies, like WLAN in a GAN access bearer (cf. Fig. 3), no dedicated resources are reserved.

For evaluating the suitability of an access bearer for a service data flow, the load and availability of access resources are important parameters. When different alternative access bearers with sufficient access bearer performance exist, the resource situation of the different access resources can be used to balance the load between the access systems. A difficulty in this evaluation is the fact that the resources of different access technologies cannot be directly compared. Therefore, a generic abstraction of access resources for different types of access technology is required. For example, the number of available time slots in a TDMA access technology cannot be directly compared to the amount of remaining transmit power in a CDMA access technology. The generic resource abstraction allows to describe the total and available access resources in a common way, and to specify which impact the allocation of a service data flow has on the resource situation.

In addition to a resource abstraction, the performance perceived by the service data flow is a main criterion to assess the suitability of an access bearer. For this the access bearer needs to be characterised, for example based on measurement of the access link quality. It is desirable to derive the utility that the access bearer provides to the service data flow (for a formal description see Sect. 6). Thus the generic performance abstraction needs to characterise the data rate, delay and reliability of the access bearer in a generic way. Generic abstractions of access resources and access performance have been developed in [19,20].

3.4.2 Connectivity Resources

Although the access resources constitute in most cases the bottleneck of an access bearer, there exist other constellations when rather the connectivity resources are the bottleneck. An obvious example is a WLAN access point with a net peak data rate of approximately 27 Mb/s, which is connected to the access anchor via a DSL line with 6 Mb/s peak rate. Even if the capacity of the connectivity network is larger than the capacity of an access resource, congestion can occur due to traffic aggregation of service data flows from a possibly large number of access resources. In these cases the limitations of the connectivity resources determine the suitability of the access bearer. It is not trivial to evaluate the performance and resource situation of the connectivity resources. In some case, it may be a single *connectivity link* that connects the point of attachment of the access resource to the multi-access anchor of the access bearer. In other cases, the point of attachment and the anchor can be connected via a *connectivity path* through a complete connectivity network. In this case traffic engineering methods can be applied in the connectivity network to avoid resource limitations. In the case of cooperating networks, the point of attachment can be even located in a visited network with the anchor being located in the home network; the connectivity path then leads through two connectivity networks connected via a third interconnection network. This plethora of options makes a general description of the resource situation and the performance provided by connectivity resources to a service data flow difficult. A practical solution has been presented in [41], where a bottleneck of the connectivity resources is determined by a constraint value. This constraint could be obtained, either dynamically by network management functions, or more static by operation and maintenance procedures. In some cases it may not be possible to detect a bottleneck of the connectivity path at all. The constraint value of the connectivity resources is used as a weight to the service utility of an access bearer in the assessment of the access bearer suitability.

4 Multi-access Architecture

The multi-access architecture is described in terms of three main entities, as shown in Fig. 6. The *Multi Radio Resource Management (MRRM)* is the key control entity in the multi-access system. It monitors available access bearers for each user network and allocates one or more of these to a service data flow. It thus performs access selection, as well as other MRRM functions like admission control and load management. The access selection algorithm depends on (static and dynamic) input information and how it is obtained, the objectives (selection criteria), and the degree of cooperation in multi-operator cases. A result of access selection is typically either to stay with the current access or that a handover towards another access shall be executed. The *Generic Link Layer—Interface and Context Transfer* (GLL_{I-CT}) provides a generic interface towards MRRM and support functionality for transmission over an access link. GLL_{I-CT} performs the abstraction [19,20] of the resource status and performance of an access bearer, as shown in Fig. 7. Thus GLL_{I-CT} provides MRRM with generic descriptors about access bearer characteristics, which enable MRRM to compare different access bearers independent from the access technology that is used. The reporting from GLL_{I-CT} to MRRM are controlled by certain rules and thresholds in order to perform event classification and filtering. In this manner the granularity and frequency of GLL to MRRM reported link events (triggers) is determined. The control of event classification, filtering and reporting allows to balance the precision (and benefit) of access selection versus signalling overhead and scalability. This trade-off needs to be adapted to every individual multi-access system realisation. In case that a flow is handed over between different GLL entities, it supports link layer context transfer [42,43]. Figure 7 depicts the relationship of the Generic Link Layer with respect to the access technology specific functions and the multi-access functions.

The Multi-Access Anchor (MAA) (see Fig. 6) is a routing decision point that maps service data flows to access bearers. The MAA is the entity where handovers are executed. Each MAA entity needs to store the active mapping (i.e., SDF binding) of service data flows to access bearers. Note that there can be multiple MAAs which are then typically structured in a hierarchical manner. In addition the MAA may be combined with the *GLL context anchor* functionality (GLL_{CA}) to support link layer context transfers where copies of data packets are kept in the GLL_{CA} until the GLL_{I-CT} entities signal that the packets have been successfully transmitted.

The multi-radio access entities can be implemented in different ways to suit different networking scenarios. While the MRRM is often depicted as a single box, it has to be stressed that it generally comprises multiple physical entities that are typically located in different nodes.

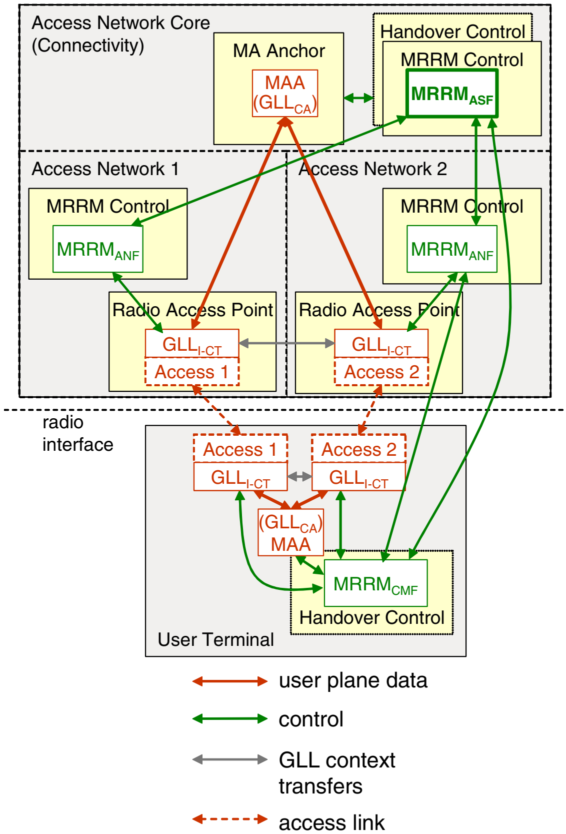

Fig. 6 Multi access architecture

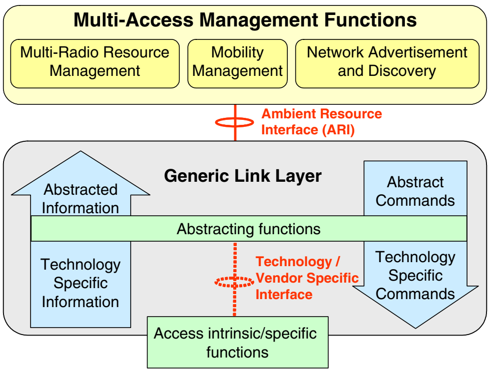

Fig. 7 Generic abstractions of access specific functions and characteristics to support multi-access management

Figure 6 shows an example of a physical implementation of multi-access with a distributed MRRM. The following MRRM entities are included in this architecture:

- $MRRM_{ASF}$: The access selection function which is the master MRRM entity responsible for deciding on the best-suited access bearer for a service data flow. It maintains the sets of useable access bearers and determines the utility of access bearers for SDFs as described in detail in Sects. 5 and 6.

- $MRRM_{ANF}$: The access network control function which configures measurements in the access network via the GLL_{I-CT} and monitors access network related parameters like cell-load.
- $MRRM_{CMF}$: The connection management function which monitors the performance of access bearers for the user network.

Such an approach is in particular useful, in case that the MAA and the $MRRM_{ASF}$ are high up in the network hierarchy, and need to manage a very large number of users and radio cells. For scalability reasons, it is then advantageous if the $MRRM_{ASF}$ receives only limited information. For example, the $MRRM_{ASF}$ does not need to know the exact link quality for every access link of all user networks, but it is only informed by $MRRM_{CMF}$ if a link quality becomes critical or a new link is discovered.

It has to be noted that the initial access selection—when the user terminal/user network tries to establish connectivity to access networks—always takes place in the user terminal/user network. Once the user network is connected, it reports service data flow requirements to the access network as well as measurements on the actual access/link performance. Additional discovered accesses are also reported to the access network. Based on this information and information collected in the access network, access selection takes places as described in the next section.

The multi-access architecture depicted in Fig. 6 can be applied in different business scenarios, as described in Sect. 2. For example, different access networks can belong to different business actors. Depending on the type of cooperation the distribution of functionality may vary. For example, an untrusted access network may not comprise MRRM and GLL functionality; instead the multi-access functionality is only located in the access network core and the user network. The distribution of functionality between network domains of different network operators can be based on pre-established business agreements (e.g., roaming agreements) or establishing dynamic cooperation agreements based on network composition [10,16,44,45]. The distribution of multi-access functionality in different networking scenarios has been discussed for example in [46–49].

5 Management of Accesses

Although the selection of the best suited access is the prime objective of multi-access management, it is part of a larger access management process. Before access selection can be performed, it is required to learn which accesses are available for each user network. Further, information needs to be collected from which the suitability of every access for a data session can be determined. It also has to be determined, when connectivity with an access is established. In a practical scenario, there can be limitations for a user network concerning the capability to monitor or connect to multiple accesses at the same time. These limitations can stem from the implementation of the radio modems. For example, a device based on software defined/reconfigurable radio design [50] has only a single configurable radio front-end; most access technology specific operations are realised by software. Such a device can only connect to a single radio access technology at a time. Before changing to another access, the radio front-end needs to be reconfigured to the new carrier frequency and carrier bandwidth, and the software modules must be reconfigured for the access functions. The access functions include the coding and modulation scheme, multi-antenna configuration and algorithms, as well as, radio protocol functions, like medium access control and scheduling, segmentation and automatic repeat request, ciphering and header compression. As a consequence, the user network first needs to disconnect from one access before it can connect to the new access. Already for making measurements on other accesses, the radio front-end must be temporarily reconfigured. But also terminals with multiple separate implementations of radio modems face limitations. For example, due to interference between RATs in close frequency bands, it may not be possible to connect to two such RATs simultaneously. Also measurements in the terminal of one RAT can be hampered if simultaneously on another RAT in a close frequency band data is transmitted. Finally, even if no further restrictions on simultaneous usage of different RATs remain, simultaneous connectivity and RAT measurements require substantial battery resources; wise usage of measurements and connectivity via multiple RATs is required for battery-powered mobile devices.

5.1 Access Sets

For the management of accesses we propose the usage of *access sets*, as shown in Fig. 8, which extends the approach described in [19,46,51]. These access sets are shared between the MRRM entity in the user network and the MRRM entities in the network. The *detected access set* (*DAS*) contains all access links that are detected by the user network, including those to which it is already connected. The elements are included, when a new cell of an access system is detected by scanning for beacon signals broadcasted in the radio cells. An element is removed from the detected set, if a beacon signal cannot be observed anymore. The *validated access set (VAS)* contains all accesses of the DAS, which are validated by local policies. For example, certain networks can be barred for usage for the user network. When a service is invoked and a service data flow is setup, it has to be decided which access to use for that service data flow. MRRM determines a *candidate access set* (*CAS*), which includes those accesses with capabilities that match the requirements of the service data flow. Also in this process policies may restrict the admitted accesses, for example, if the usage of a particular access for a particular service requires an agreement between the user and the network

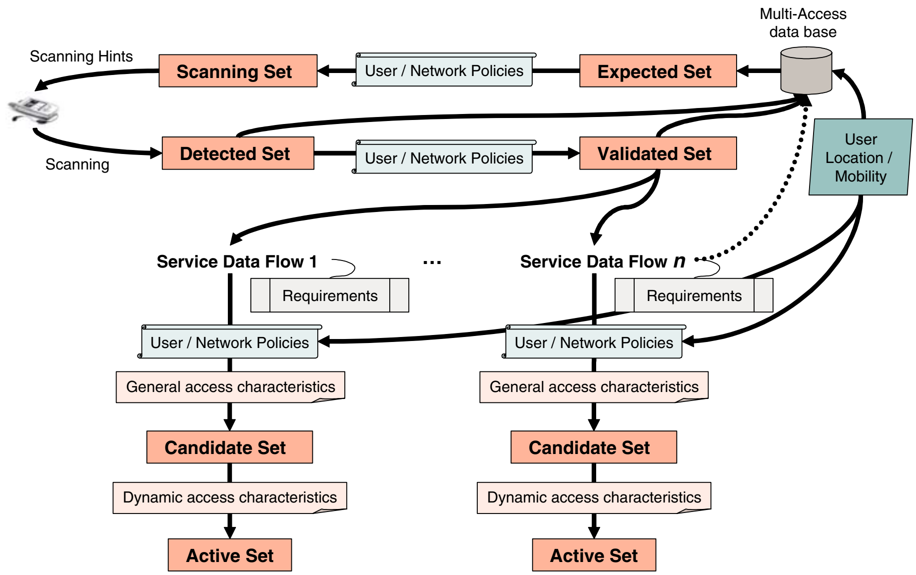

Fig. 8 Access sets for a user network and its service data flows

provider. If the user policy requires that a minimum security level is required for a service data flow, only accesses with sufficient security levels are admitted. Thereby, already active accesses are preferred over new accesses. This policy-based restriction of accesses is equivalent to a filtering process to determine suitable candidates accesses. For all elements of the CAS a utility is derived from the service requirements, as well as dynamic system parameters, like the access performance and load level for an access. More static parameters can also impact the utility. For example, depending on the agreement of the user and network provider, like gold/silver/bronze subscriptions, users can be assigned certain preference values for each access which can also take the cost-of-use into account. The utility can be further weighted with a preference value according to cross-service aspects. For example, if the user is already running some services via some accesses, it may be desirable to limit the total number of RATs that are active simultaneously. More parameters can be considered in the utility function. Different RATs can cope differently with user mobility, for example, the capability of the radio transmission to adopt to time-varying radio channels, or to support efficient handover procedures. Therefore the user mobility—derived from location measurements, handover rates or user indications—combined with an access-specific mobility support knowledge, can be included as parameter to determine the utility of an access. Also the reliability of a RAT with respect to fulfilling the service performance can influence the utility. For example, the grade of coverage can describe the reliability of an access for providing continuous service coverage. Based on this set of parameters, a utility is determined for every access in the CAS for the service data flow. The access selection decision is then to select the access with the highest utility. This access is included in the *active access set* (*AAS*), which contains the access in use for the service data flow. It is, in general, possible to split a service data flow onto several accesses, so the AAS can contain multiple elements. However, typically the AAS contains only a single access. The selection of the best suited access is a dynamic process, since the parameters which determine the utility are time varying. In particular the radio link performance and the cell load can change dynamically and require a re-evaluation of the CAS to determine the best suited access. For an already ongoing session, it has to be considered that any handover between different accesses comprises a signalling overhead and can also temporarily degrade the service experience, depending on the access handover procedure. Therefore, a minimum utility benefit margin is required in the access selection process. An access handover is only triggered, if an access in the CAS exceeds the utility of the active access by at least the utility benefit margin. This provides a hysteresis to avoid ping-pong effects of changing between accesses, and also accommodates for the handover costs implied in the change of access.

In order to accommodate for the limitation of a terminal to scan all available accesses, two extra access sets are used. The *expected access set* (*EAS*) contains the accesses that a terminal is expected to be able to connect to. It is determined from the user position—given geographically or by the radio cell(s) it is currently connected to—and multi-access neighbour list information. A multi-access database containing neighbour cell relationships can be dynamically maintained based on terminal measurements, or by network configuration. A terminal can maintain a local copy of the multi-access database. From the EAS a *scanning access set* (*SAS*) is determined, which contains accesses that the terminal is directed to scan for. The policy rules, according to which the SAS is determined, include terminal capabilities about which accesses can be used by the user network, roaming relationships based on which accesses are provided by cooperating roaming partners, the estimated benefit that can be achieved by detecting a new access and adding it to the existing candidate access sets. The SAS provides hints to the user network to scan for new accesses.

These hints contain, for example, the type of RAT, the carrier frequency and the provider name.

5.2 Access Handover

Once an access selection decision has made the decision to change the active access, it is required that the service data flow is redirected from one access bearer to another one. For this the SDF binding has to be modified in the multi-access anchor and the user network. As discussed in Sect. 3.2, access bearers can be based on different types of access technologies and mobility protocols. Consequently, access bearers are identified by different descriptors, for example, the locators of the access bearer endpoints. Such a locator can be a Mobile IP care-of-address, an IPsec security association or a GTP tunnel-endpoint identifier. Depending on the technology also the handover procedures can vary by which the SDF binding is updated. A multi-access anchor may need to support multiple handover procedures for different access bearers. Multi-radio resource management steers handover execution function to perform the access handover according to the appropriate handover procedure. In [52,53] this handover execution functions is referred to as handover and locator management (HOLM). It keeps track of the locators of an access bearer and the handover procedure/protocol that is required to update the SDF binding. It thus contains a toolbox of handover tools, from which the appropriate tool is selected depending on the source and target access bearer. The handover toolbox can also comprise handover optimization tools. Such tools can support context transfer and data forwarding for seamless and lossless access handover, or network mobility for moving networks. Depending on the capability of the source and target access bearers, such optimization tools can be used to increase the access handover performance. When a handover command is initiated by MRRM to perform a handover between two access bearers, HOLM determines the suitable mobility protocol for the access bearers. It further determines what handover optimization tools can be applied, for example for context transfer. For the handover execution, the SDF binding is updated in the corresponding multi-access anchor and in the user network with the appropriate handover protocol (e.g., GTP, MIP, PMIP).

6 Access Selection

We will now derive a formulation for *access selection*, which is in a sense the key mechanism in a multi-access system: given multiple available accesses, which one should be picked? The central role of access selection is reflected in the large body of literature treating various access selection schemes and algorithms, see, for example, [54–63]. Previous work, however, only considers a single or few parameters as criteria for access selection, like the radio link quality or costs of access. In this section we describe a general framework for access selection which can comprise a large number of criteria to be considered in the evaluation process. This section focuses on the snapshot access selection decision problem: how to choose the best access (or accesses) among the available ones. To be more precise using the terminology of the connectivity abstraction model of Sect. 3, the access selection decision concerns selecting the best access bearer (or access bearers) to use for a particular service data flow. There are two main aspects of this problem: (1) how to define “best” and, (2) how to determine the best access bearer. The first relates to the objective of the decision-making, what is the measure and what information is needed, while the second relates to the algorithm and mechanisms for maximizing (or minimizing) the objective.

6.1 Access Selection—Objectives

There are many parameters that characterize and influence access selection objectives, including different policies (like security), user, operator, and service preferences (for example, a particular service is best handled in a specific access bearer type), the capabilities and current performance of the available access bearers (if the access bearer is capable of meeting the SDF quality of service requirements and how well and/or efficient it is at doing it momentarily), the resource status in different access systems, the cost of using an access (for the user or operator), and whether or not a particular access bearer is already in use for an SDF (changing to another access bearer may incur performance degradation during handover). Some of these parameters may be conflicting with each other, for example, service preferences and current access bearer performance. Parameters may also exist in multiple instances corresponding to different actors with different roles. Apart from the user and operator roles mentioned above there may be multiple operators and other cooperating actors with preferences and policies that would influence a particular access selection decision, possibly in conflicting ways. Some of the parameters have an inclusive/exclusive impact on the access bearer candidates, that is, an access bearer is either allowed or forbidden. Other parameters instead indicate the relative merits of the different available access bearers. In formulating access selection objectives it is necessary to treat the diversity of parameters, the different roles that the involved actors have, and the way that the parameters affect the list of candidate access bearers. A general approach for a decision problem of this type is to pose it as a mathematical optimization problem using utility functions. Utility functions, as already introduced in Sect 3.3, are well suited in the context of access selection since they provide means to quantify the relative merits of satisfying the differennt SDF quality-of-service requirements. For every SDF s the access selection problem can be stated as:

$$\begin{array}{cc} \max & u(x, P, s) \\ \text{subject to} & x \in X(P, s) \end{array} \tag{1}$$

Here the objective function is a utility function $u(x, P, s)$ that returns a real value representing the utility of choosing access bearer x for SDF s given the set of parameters P; the *outcome* of the access selection process is the access bearer that is best suited for SDF s. The feasible set $X(P, s)$ is the set of available (candidate) access bearers for SDF s given the set of parameters P. It can be viewed as a validation filter on available (detected) access bearers for the inclusive/exclusive type of parameters (such as policies and access bearer capabilities), that is:

$$X(P, s) = \{x \in D : g(x, P_e, s) = \text{TRUE}\} \tag{2}$$

where D is the set of available (detected) access bearers, P_e is the set of inclusive/exclusive type of parameters, and $g(x, P_e, s)$ is a boolean function that is TRUE if access bearer x is allowed for SDF s according to the parameters P_e. Note that the filtering "process" is here defined on a per SDF basis starting from an available (detected) set of access bearers. In practice some of the filtering applies to a per user (or user network) basis, which can be performed even before an SDF exists. This filtering is equivalent to the construction of the candidate access set from the detected access set as discussed in Sect. 5.

Returning to the utility function $u(x, P, s)$, it is clear that it will be composed of several different parts. The approach in this article is to structure it at the top level according to the different actors or entities that are involved in the multi-access system. For the business scenario considered with users U having user networks UN connected to an operator having multiple access technologies/networks AN integrated into a common multi-access network

N, possibly in combination with other cooperating multi-access networks (AN and N), the following decomposition is suitable:

$$u(x, P, s) = f\,(u_S, u_U, u_{UN}, u_{AN}, u_N)\,. \tag{3}$$

Here $u_S, u_U, u_{UN}, u_{AN}, u_N$ are separate utility functions (the dependence on x, P, s is suppressed for brevity) that reflect the utility from the service (S), user (U), user network (UN), access network (AN), and overall network (N) perspectives respectively.

The service S utility function $u_S(x, P, s)$ measures how well an access bearer fulfils the quality-of-service requirements for SDF s. Section 3.3 introduced service type, bit rate, delay, and error ratio as the quality-of-service descriptors and discussed the shape of individual utility functions for the latter three (see Fig. 4). The individual bit rate, delay, and error ratio utilities can be combined in different ways—one approach is multiplicatively:

$$u_S\,(x, P, s) = w_S \cdot u_b\,(b\,(x)\,, s) \cdot u_d\,(d\,(x)\,, s) \cdot u_e(e(x), s), \tag{4}$$

where $b(x), d(x), e(x)$ are the bit rate, delay, and error ratio provided by access bearer x respectively, and w_S is a weight factor, which can, for example, be used to represent a service priority. Note that the service type is captured in the shape of the individual utility functions (shown by the dependence on s) and that $b, d, e \in P$.

The user U utility function $u_U(x, P, s)$ captures user satisfaction defined through policies or preferences. An example is where a user requests an elastic service (Fig. 4(a)) and where different access bearers have different costs, corresponding to different degrees of provided quality-of-service. To capture the trade-off between cost and the service utility in $u_S(x, P, s)$ the user utility function $u_U(x, P, s)$ can include an inverse cost dependence.

The user network UN utility function $u_{UN}(x, P, s)$ characterizes how well resources, such as transmission power, are used in the user network. For example, two different access bearers may require different amounts of transmission power in the user network to provide the same service.

The access network AN utility function $u_{AN}(x, P, s)$ measures how well resources are used in the access network. Since the access network is shared between many users it depends on the current distribution of allocated resources between the user networks and on aggregated measures such as the current load level, total data rate, total number of users in the access network, or total resource-usage efficiency.

Finally, the (common multi-access) network N utility function $u_N(x, P, s)$ represents the preferences of the overall network, for example policies on priorities for different users.

6.2 Access Selection—Algorithms and Mechanisms

The access selection algorithm concerns how to find the solution to Eq. 1 above, that is, how to find a feasible access bearer x that maximizes the total utility for SDFs given the set of parameters P. Looking at Eq. 1 from an optimization point of view the problem has a very small feasible set corresponding to the available (candidate) access bearers, typically only a handful, and the goal is to pick one of them. The best algorithm is simply to:

1. Determine the feasible set of available (candidate) access bearers $X(P, s)$ from the set of detected access D bearers according to Eq. 2.
2. Compute the utility values $u(x, P, s)$ for each access bearer in $X(P, s)$ according to Eq. 3.
3. Choose the one with the largest value of $u\,(x, P, s)$, that is:

$$\begin{aligned} \hat{x} = \arg\max \quad & u(x, P, s) \\ \text{subject to} \quad & x \in X(P, s) \end{aligned} \tag{5}$$

This straightforward general scheme becomes in reality more complicated. First, the discussion so far has focused on selecting exactly one access bearer for an SDF, but it may be desirable for the multi-access system to be able to select more than one access bearer at the same time. Multiple access bearers can be used jointly to, for example, increase total data rate or enhance resilience to failures. If this option is included in the system then the utility functions need to be redefined for multiple access bearers $u\ (x_1, \ldots, x_n, P, s)$, the set of available (candidate) access bearers $X\ (P, s)$ needs to be extended to all feasible combinations of access bearers, and the evaluation effort of the combinations increases greatly.

Another aspect is that the access selection algorithm above is defined for one particular SDF only. Even if the proposed structure of the utility function is used, which, for example, includes an access network *AN* component to reflect the effects of multiple users on the resource-situation, it is still a localized decision in a system that must simultaneously manage many SDFs and access bearers. A particular choice for a given SDF could imply that the choice for another SDF should be changed. Ideally the system should solve, in each snap-shot situation, a joint access selection problem where all current SDFs are assigned to access bearers at the same time, but this would likely be too complex. Nevertheless, the individual SDF access selection scheme may be complemented with a joint SDF scheme working on a slower time scale proposing re-allocation to improve overall performance.

A third problem with the algorithm above is architectural and concerns where it is executed in relation to the information that is available to it. Some service utility parameters may easily be collected in the user network, for example, current data rate throughput, whereas access network utility parameters, for example, resource load, is derived in the access network. The collected parameter information needs to be distributed to all the access selection decision points. If the access selection decision function is very distributed, for example, if it is located in the user network, then this signalling may be costly, in particular when it is over radio links. There may also be synchronization problems if multiple decision points (user networks) receive resource load parameter information and then react independently and simultaneously in the same way, which may overall result in a poor solution. In practice the full utility function structure, with service, user, user network, access network, and network components, may only work with a network-based distribution of the access selection decision function.

A fourth problem with the general, single-SDF access selection algorithm arises from the fact that many SDFs are "bundled" to the same user and user network, for example, there may be a simultaneous downlink and an uplink SDF for the application in use, such as in a voice call. The access bearer is a unidirectional abstraction but it may nevertheless be mapped to a bidirectional bearer in an access system (for example, a 3GPP RAN). If this happens for an access selection decision for, say, the downlink SDF, then the subsequent access selection decision for the uplink SDF should clearly choose the same access bearer since the resources are already there. This problem, and similar issues from SDFs that are correlated, can be handled by extending the single-SDF access selection algorithm with a "memory", a list of recently allocated SDFs containing information on how they correlate with other SDFs, and means for modifying the utility function evaluations through a correlation utility value [19].

In spite of the complicating issues discussed above for the single-SDF access selection algorithm, the proposed structure of the utility function still captures the essence of the multi-access system functionality. However, it must be framed in a general machinery for management of accesses to handle all aspects from detecting that an access bearer is available to carrying out the changes needed to enforce an access selection decision.

The access selection function is triggered whenever significant changes to the input parameters occur. The *input parameters* for the access selection are:

(a) Requirements of the service data flow(s),
(b) Policies and preferences of the user, as well as, of the operator(s) that provide access,
(c) Cost of usage for each access,
(d) Capabilities and performance of the available access bearers,
(e) Availability of resources for the available access bearers, and resource efficiency of using an access bearer for the service data flow.

Access selection is triggered from the service requirements (a) at setup and termination of a data session. Also during changes of the service requirement, for example when switching a video transmission to a higher resolution, access selection is triggered. When policies, preferences (b) or cost of usage (c) change, the filtering rules in Eq. 2 that determine the candidate access sets are adapted. Access selection events of type (a–c) occur only infrequently in typical usage scenarios. In contrast, the resource situation (e) and performance (d) of access bearers can change dynamically due to radio channel fluctuations, user mobility and handover, as well as, dynamic changes of the traffic load. The frequency of events to this *dynamic access selection* depends on the event classification and filtering that is performed by GLL for reporting link events to MRRM. It needs to be configured to the topology and realisation of a specific multi-access system as discussed in Sect. 4.

7 Conclusion

Communication networks contain an increasing number of access technologies for both fixed and wireless access. Similarly, end user networks and terminals integrate multiple access technologies to provide connectivity to access networks. This provides a choice of how user networks and terminals are connected to the communication infrastructure. At the same time the number of access networks increase and the characteristics of these access networks diversify depending on the access technology in use. In this article we have presented and discussed the challenges and complexity stemming from such a multi-access system. We have investigated the roles that network providers can play in this diversified market setting and illustrated new business scenarios, which can differ from the roles found in the communication market of today. We have presented approaches to design and create a realistic, functioning multi-access system that integrates various heterogeneous access and networking technologies. For this we have developed a multi-access system model and a multi-access architecture that abstracts and describes access bearers, based on different access technologies, in a common way. We have presented a framework for access selection based on determining a utility for different accesses. This utility-based approach balances the interests of different entities in the communication system. For example, it considers the performance of an access allocation for a data service, but at the same time includes the resource situation in the access network, as well as, user and network preferences and policies. Finally we have presented how the different multi-access functions—access detection, access selection and access handover—are integrated into a general multi-access management framework.

It remains for future work to investigate how this multi-access management framework can be integrated into mobile network architectures.

Acknowledgement This work has been partly performed within the Ambient Networks project, which is supported by the European Commission under grant 027662 within the Sixth Framework Program. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Ambient Networks project or the European Commission.

References

1. Prytz, M., Karlsson, P., Cedervall, C., Bria, A., & Karla, I. (2006). Infrastructure cost benefits of ambient networks multi-radio access. In *Proceedings of IEEE Vehicular Technology Conference (VTC)*. Spring.
2. Johansson, K., Furuskär, A., Karlsson, P., & Zander, J. (2004). Relation between base station characteristics and cost structure in cellular systems. In *Proceedings of IEEE Personal, Indoor, and Mobile Communications (PIMRC)*.
3. Cedervall, C., Karlsson, P., Prytz, M., Hultell, J., Markendahl, J., Bria, A., Rietkerk, O., & Karla, I. (2005). Initial findings on business roles, relations and cost savings enabled by multi-radio access architecture in ambient networks. In *Proceedings of the 14th Wireless World Research Forum*.
4. Furuskär, A., Almgren, M., & Johansson, K. (2005). An infrastructure cost evaluation of single- and multi-access networks with heterogeneous traffic density. In *Proceedings of IEEE Vehicular Technology Conference (VTC)*. Spring.
5. Johansson, K., & Furuskär, A. (2005). Cost efficient capacity expansion strategies using multi-access networks. In *Proceedings of IEEE Vehicular Technology Conference (VTC)*. Spring.
6. Johansson, K. (2005). Cost efficient provisioning of wireless access: infrastructure cost modelling and multi-operator resource sharing. Licentiate Thesis, Royal Institute of Technology (KTH), ISSN 1400-9137.
7. Gustafsson, E., & Jonsson, A. (2003). Always best connected. *IEEE Wireless Communications, 10*, 49–55.
8. 3GPP TR 23.882. 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; 3GPP System Architecture Evolution: Report on Technical Options and Conclusions (Release 7); V1.6.1; November 2006.
9. IEEE 802.21. Draft IEEE Standard for Local and Metropolitan Area Networks: Media Independent Handover Services; D02.00; September 2006.
10. Niebert, N., Schieder, A., Abramowicz, H., Malmgren, G., Sachs, J., Horn, U., Prehofer, C., & Karl, H. (2004). Ambient networks—an architecture for communication networks beyond 3G. *IEEE Wireless Communications, 11*, 14–22.
11. Viktorsson, O. (2004). Business Aspects and Requirements on Ambient Networks. Presented at the 1st Symposium of the Wireless World Initiative (WWI), Brussels, Belgium, 9–10 December, 2004. http://www.wireless-world-initiative.org/ (accessed 14 Feb. 2007).
12. Markendahl, J., Johnsson, M., & Karlsson, P. (2006). Business opportunities and regulatory issues of Ambient Networks. In *Proceedings of 17th European Regional ITS Conference*. Amsterdam, Netherlands, August 22–24, 2006.
13. Markendahl, J., & Johnsson, M. (2006). Ambient networking and related business concepts as support for regulatory initiatives and competition. In *Proceedings of the 5th Conference on Telecommunication and Techno-Economics (CTTE)*. Athens, 8–9 June 2006.
14. Rietkerk, O., Huitema, G., & Markendahl, J. (2006). *Business roles enabled by Ambient Networking to provide access for anyone to any network and service*. Presented at Helsinki Mobility Roundtable, Helsinki, Finland, 1–2 June 2006.
15. Markendahl, J., Poyhoenen, P., Strandberg, O., & Laganier, J. (2006). *Business implications of AN composition framework*. Presented at Helsinki Mobility Roundtable, Helsinki, Finland, 1–2 June 2006.
16. Ambient Networks. (2006) Ambient Networks Draft System Description. Public project deliverable D7-A.2, December 2006. http://www.ambient-networks.org/.
17. 3GPP TS 23.401. (2007). 3GPP System Architecture Evolution: GPRS enhancements for LTE access (Release 8); V0.1.0; January 2007.
18. 3GPP TS 23.402. (2007). 3GPP System Architecture Evolution: Architecture Enhancements for non-3GPP accesses (Release 8); V0.1.0; January 2007.
19. Ambient Networks. (2006). Multi-Access & ARI design and Initial Specification. Public project deliverable D2-C.1, December 2006. http://www.ambient-networks.org/.
20. Sachs, J., Agüero, R., Daoud, K., Gebert, J., Koudouridis, G. P., Meago, F., Prytz, M., Rinta-Aho, T., & Tang, H. (2007). Generic abstraction of access performance and resources for multi-radio access management. In *Proceedings of the 16th IST Mobile & Wireless Communications Summit*. Budapest, Hungary, 1–5 July 2007.
21. 3GPP TS 23.107. (2006). 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Quality of Service (QoS) concept and architecture (Release 6), V6.4.0; March 2006.
22. 3GPP TS 23.203. (2006). 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Policy and charging control architecture (Release 7); V7.1.0; December 2006.
23. 3GPP TS 23.207. (2005). 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; End-to-end Quality of Service (QoS) concept and architecture (Release 6), V6.6.0; September 2005.

24. Ludwig, R., Ekström, H., Willars, P., & Lundin, N. (2006). An evolved 3GPP QoS concept. In *Proceedings of the 63rd Vehicular Technology Conference Spring*. Melbourne, Australia, 7–10 May, 2006.
25. Nichols, K., Blake, S., Baker, F., & Black, D. (1998). Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers. IETF RFC 2474, December 1998.
26. Braden, R., Zhang, L., Berson, S., Herzog, S., & Jamin, S. (1997). Resource ReSerVation Protocol (RSVP). IETF RFC 2205, September 1997.
27. Hancock, R., Karagiannis, G., Loughney, J., & Van den Bosch, S. (2005). Next Steps in Signaling (NSIS): Framework. IETF RFC 4080, June 2005.
28. 3GPP TS 29.060. (2006). 3GPP Technical Specification Group Core Network and Terminals; GPRS Tunnelling Protocol (GTP) across the Gn and Gp interface (Release 7); V7.4.0; December 2006.
29. Perkins, C. (2002). IP Mobility Support for IPv4. IETF RFC 3344, August 2002.
30. Johnson, D., Perkins, C., & Arkko, J. (2004). Mobility Support in IPv6. IETF RFC 3775, June 2004.
31. Levkowetz, H., Giaretta, G., Leung, K., Liebsch, M., Roberts, P., Nishida, K., Yokota, H., & Parthasarathy, M. (2006). The NetLMM Protocol. IETF work in progress, draft-giaretta-netlmm-dt-protocol-02, October 5, 2006.
32. Gundavelli, S., Leung, K., Devarapalli, V., Chowdhury, K., & Patil, B. (2007). Proxy Mobile IPv6. IETF work in progress, draft-sgundave-mip6-proxymip6–02, March 05, 2007.
33. Bedekar, A., Singh, A., Kumar, V., & Kalyanasundaram, S. (2007). A Protocol for Network-based Localized Mobility Management. IETF work in progress, draft-singh-netlmm-protocol-02, March 05, 2007.
34. Rosen, E., Viswanathan, A., & Callon, R. (2001). *Multiprotocol Label Switching Architecture*. IETF RFC 3032, January 2001.
35. Farinacci, D., Li, T., Hanks, S., Meyer, D., & Traina, P. (2000). Generic Routing Encapsulation (GRE). IETF RFC 2784, March 2000.
36. Kent, S., & Seo, K. (2005). Security Architecture for the Internet Protocol. IETF RFC 4301, December 2005.
37. Ekström, H., Furuskär, A., Karlsson, J., Meyer, M., Parkvall, S., Torsner, J., & Wahlqvist, M. (2006). Technical Solutions for the 3G Long-Term Evolution. *IEEE Communications. Magazine, 44*, 38–45.
38. 3GPP TR 25.913. (2006). Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN) (Release 7); March 2006.
39. 3GPP TS 43.318. (2006). 3rd Generation Partnership Project; Technical Specification Group GSM/EDGE Radio Access Network; Generic access to the A/Gb interface; Stage 2 (Release 7); V7.0.0; November 2006.
40. 3GPP TS 23.234. (2006). 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; 3GPP system to Wireless Local Area Network (WLAN) interworking; System description (Release 7); V7.4.0; December 2006.
41. Gunnar, A., Ahlgren, B., Blume, O., Burness, L., Eardley, P., Hepworth, E., Surtees, A., & Sachs, J. (2007). Access and path selection in ambient networks. In *Proceedings of the 16th IST Mobile & Wireless Communications Summit*. Budapest, Hungary, 1–5 July 2007.
42. Sachs, J., Khurana, B. S., & Mähönen, P. (2006). Evaluation of handover performance for TCP Traffic based on generic link layer context transfer. In *Proceedings of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. Helsinki, Finland, 11–14 September 2006.
43. Sachs, J., Kampmann, M., Lundsjö, J., Meago, F., Rinta-Aho, T., & Tharon, B. (2006). Continuous connectivity in ambient networks. In *Proceedings of the IST Mobile and Wireless Communications Summit*. Myconos, Greece, 4–8 June 2006.
44. 3GPP TR 22.980. (2006). 3GPP Technical Specification Group Services and System Aspects; Network composition feasibility study (Release 8); V1.0.0; October 2006.
45. Kappler, C., Poyhonen, P., Johnsson, M., & Schmid, S. (2007). Dynamic network composition for beyond 3G networks: a 3GPP viewpoint. *IEEE Network, 21*, 47–52.
46. Ambient Networks. (2005). Multi-Radio Access Architecture. Technical Report D2–4. December 2005. http://www.ambient-networks.org/ (accessed Dec. 2006).
47. Lundsjö, J., Agüero, R., Alexandri, E., Berggren, F., Cedervall, C., Dimou, K., Gebert, J., Jennen, R., Jorguseski, L., Karimi, H.R., Meago, F., Tang, H., & Veronesi, R. (2005). Multi-radio access architecture for ambient networking. In *Proceedings of the IST Mobile and Wireless Communications Summit*. June 19–23, 2005.
48. Sachs, J., Agüero, R., Berg, M., Gebert, J., Jorguseski, L., Karla, I., Karlsson, P., Koudouridis, G. P., Lundsjö, J., Prytz, M., & Strandberg, O. (2006). Migration of existing access networks towards multi-radio access. In *Proceedings of the IEEE Vehicular Technology Conference (VTC) Fall*. Montreal, Canada, 25–28 September 2006.

49. Magnusson, P., Lundsjö, J., Sachs, J., & Wallentin, P. (2004). Radio resource management distribution in a Beyond 3G Multi-Radio Access Architecture. In *Proceedings of the IEEE Global Telecommunications Conference (Globecom)*. Dallas, USA, Nov. 29–Dec. 3, 2004.
50. Tuttlebee, W. H. W. (1999). Software-defined radio: Facets of a developing technology. *IEEE Personal Communications, 6*, 38–44.
51. Johnsson, M., Sachs, J., Rinta-Aho, T., & Jokikyyny, T. (2006). Ambient networks—a framework for multi-access control in heterogeneous networks. In *Proceedings of the IEEE Vehicular Technology Conference (VTC) Fall*. Montreal, Canada, 25–28 September 2006.
52. Agüero, R, Surtees, A., Eisl, J., Georgiades, M. (2007). Mobility Management in Ambient Networks. In *Proceedings of the Vehicular Technology Conference (VTC)*. Spring, Dublin, Ireland, 23–25. Apr, 2007.
53. Ambient Networks. (2006). Mobility Support: Design and Specification. Public project deliverable D9-B.1, December 2006. http://www.ambient-networks.org/.
54. Koudouridis, G. P., Agüero, R., Alexandri, E., Berg, M., Bria, A., Gebert, J., Jorguseski, L., Karimi, H. R., Karla, I., Karlsson, P., Lundsjö, J., Magnusson, P., Meago, F., Prytz, M., & Sachs J. (2005). Feasibility Studies and Architecture for Multi-Radio Access in Ambient Networks. In *Proceedings of the 15th Wireless World Research Forum Meeting*, 8–9 December 2005, Paris, France.
55. Koo, I., Furuskär, A., Zander, J., & Kim, K. (2004). Erlang capacity of multiaccess systems with service-based access selection. *IEEE Communication Letters, 8*, 662–664.
56. Furuskär, A., Pettersson, J., Simonsson, A., & Yilmaz, O. (2005). Access selection in WCDMA and WLAN multi-access networks. In *Proceedings Vehicular Technology Conference*, May 30–June 1, 2005.
57. Furuskär, A. (2003). Radio Resource Sharing and Bearer Service Allocation for Multi-Bearer Service, Multi-Access Wireless Networks. PhD Thesis, Royal Institute of Technology (KTH), ISSN 1400-9137.
58. Hultell, J. (2007). Access Selection in Multi-System Architectures. Licentiate Thesis, Royal Institute of Technology (KTH), ISSN 1653-6347.
59. Baraev, A., Jorguseski, L., & Litjens, R. (2005). Performance Evaluation of Radio Access Selection Procedures in Multi-Radio Access Systems. In *Proceedings of the 8th International Symposium on Wireless Personal Multimedia Communications (WPMC) 2005*. Aalborg, Denmark, September 17–22, 2005.
60. Magnusson, P., Berggren, F., Karla, I., Litjens, R., Meago, F., Tang, H., & Veronesi, R. (2005). Multi-radio resource management for communication networks beyond 3G. In *Proceedings of the 62nd Vehicular Technology Conference (VTC) Fall*.
61. Berggren, F., & Litjens, R. (2006). Performance Analysis of Access Selection and Transmit Diversity in Multi-Access Networks. In *Proceedings of the 12th Annual International Conference on Mobile Computing and Networking (MobiCom)*.
62. Fodor, G., Furuskär, A., & Lundsjö, J. (2004). On access selection techniques in always best connected networks. In *Proceedings of the 16th ITC Specialist Seminar*, Antwerp, 2004.
63. Sachs, J., & Magnusson, P. (2007). Assessment of the Access Selection Gain in Multi-Radio Access Networks. European Transactions on Telecommunications. doi:10.1002/eff.1254

Author Biographies

Joachim Sachs studied electrical and electronics engineering at Aachen University (RWTH), Germany, ENSEEIHT, France, and the University of Strathclyde, Scotland. After receiving his Dipl-Ing. degree from Aachen University in 1997, he joined Ericsson Research in Aachen. He has managed and participated in several projects on the performance evaluation of Internet services in wireless networks, as well as network architectures for the integration of multiple access technologies. Joachim is a member of IEEE and the German VDE Information Technology Society (ITG). From 1998 to 2001 he contributed to the ITG working group on convergence of telephony networks and the Internet, and since 2002 to the ITG working group on system architecture and traffic engineering. Since 2006 he is member of the technical advisory board of the European network of excellence on next generation Internet Euro-NGI. He has published papers at several conferences and journals, and holds numerous patents. In 2006 he received the Ericsson Inventor of the Year award.

Mikael Prytz received an MSc in Engineering Physics from the Royal Institute of Technology (KTH), Sweden in 1993, an MS in Engineering-Economic Systems and Operations Research from Stanford University, USA in 1998, and a Ph.D. in Optimization Theory from the Royal Institute of Technology (KTH), Sweden in 2002. He worked as a network design engineer at Ericsson Telecom between 1993 and 1996 within the areas of transport network design, topology design, capacity dimensioning, routing, and network dimensioning algorithms and tools. His Ph.D. thesis research was on mathematical optimization algorithms for network design problems with unicast and multicast traffic. He joined Ericsson Research in 2003 where he is currently managing the radio network deployment and spectrum management group. He has worked on network architectures for integration multiple access technologies; multi-standard and optimization based radio resource management mechanisms, algorithms, and architectures; and future radio access technology system concepts. From March–June 2007 he was a visiting researcher within the Communications Networks Research Group at Princeton University, USA.

Jens Gebert received his diploma in electrical engineering in 1993 from the University of Stuttgart. In the same year, he joined Alcatel (now Alcatel-Lucent), working as a software engineer for Mobile Systems. From 1997 to 1999, he worked at Alcatel Network Systems in the USA for mobile switching systems. In 1999, he joined the Mobile Systems Architecture Team in Stuttgart working on various projects for Next Generation Networks and UMTS, where he was also involved in 3GPP standardization. In 2003, he joined Alcatel Research & Innovation, where he is investigating the combination and integration of existing and future access technologies including the development and optimisation of a Multistandard Radio Resource Management and where he is actively contributing to the architecture of Radio Access Networks. Currently, he is managing the Work Package on "Multi Access" in the Ambient Networks Phase 2 Project.