On the benefits of heterogeneous networking and how cellular mobile operators can help

Marc Danzeisen*†, Torsten Braun, Isabel Steiner
Institute of Computer Science and
Applied Mathematics*
University of Berne
Neubrückstr. 10, CH-3012
Email: danzeis|braun|steiner@iam.unibe.ch

Daniel Rodellar
Swisscom Innovations†
Swisscom AG
Berne, Switzerland
Email: daniel.rodellar@swisscom.com

***Abstract*— Many research efforts in the domain of spontaneous networking are aiming at providing means to enable devices to communicate with little or no knowledge of its users about the underlying technology and its configuration. The establishment of communication channels should happen in an ad-hoc and convenient manner for the user. It should also be possible to connect at any place, at any time, with anyone using always the most appropriate radio system available. Unfortunately, the different communication technologies require often different settings and therefore a certain level of knowledge is needed to successfully connect them. This is especially true for direct node to node communication technologies like WLAN, where no centralized system is present to manage the communication setup. Several parameters have to be set before a communication can occur. When talking about secure communication the procedure gets even more complicated due to the additional key negotiation and management.**

Furthermore, depending on the movement of the nodes, the application that is used, the initially chosen communication technology might become suboptimal or even useless. Hence, a seamless handover to another technology would be necessary to allow to maintain the session. The first part of this paper mainly focuses on the benefit of session handovers between infrastructure based communication technologies and direct node to node communication. An implementation architecture for such a heterogeneous session management is proposed in the second part.

***Index Terms*— heterogeneous networking, spontaneous networking, ad-hoc, seamless handover, session management**

I. Introduction

The advances in the domain of portable computing and wireless communication are promising an exciting future of mobile networking. To further satisfy the demand on high bandwidth, low latency and cheap radio communication, the technologies become more specific to certain use cases. To offer the maximum speed for low distance and stationary communication like one would like to have at airport launches, meeting rooms and offices, rather simple communication technologies like WLAN were developed. One of the major reason why WLAN became so popular in the last few year is probably the fact that it is was designed to be simple and cheap to instal and operate. Even if there is still no real mobility management or QoS support deployed in all the WLAN networks, it is still the most appropriate communication technology for users that just want to connect to the Internet. WLAN has proven once more, that users are not willing to pay for features they do not need. Despite all the advantages that a rather simple technology like WLAN may offer in those specific use cases, it will most probably newer replace mobile networks like UMTS because it was designed for completely different applications. UMTS offers full mobility and QoS support at lower bandwidth and at much higher latency levels. WLAN and UMTS are only one example how different communication technologies can be complementary. Further technologies like Bluetooth were designed with other specific use cases in mind and might therefore be the best choice in a specific situation. Unfortunately, users can often not benefit from this complementariness. Despite of big efforts done by the manufacturer to make the handling of these different technologies simpler and more user friendly, the increasing number of different communication technologies makes it nearly impossible for user to dominate.

II. Session Mobility

To allow users to profit from this heterogeneous communication technologies, this variety has to be hidden. Users should not realize when the communication technology is changing. Ongoing communication sessions have to be transparently moved to the best available technologies.

A. Infrastructure based networks

Nowadays 3G networks are already about to be extended with different access technologies (like GPRS, EDGE, UMTS, WLAN, WiMAX) and the end-user device is equipped with multiple network access technologies. An important requirement for roaming is to make it a seamless experience for end-users. The end-user notices as little as possible when changes occur at the network level and he is not interrupted while conducting a communication session (data, voice or video). This requirement is already fulfilled in today's cellular networks where an end-user making a voice call on his cellular handset will not notice a network handoff when he happens to move to another cell. The challenge is to implement the same concept across administrative domains, heterogeneous networks and services. One of the prerequisites

of seamless roaming is transparent end-user authentication and security across different access network technologies. The end-user should not be bothered with technology specific mechanisms such as providing username/password or filling in an access code. Furthermore, additional measures may be necessary if uninterrupted connectivity is required. For example, connection-oriented applications like video streaming clients cannot cope with constantly changing connection endpoints. To resolve this issue the solution should include a form of session mobility. Solutions like Swisscom Mobile's "Unlimited" [1] are bundling various access technologies in a transparent way for the end user. The mobile device (laptop) gets connected to the best available network (in this case the choice is between GPRS, UMTS or WLAN) in terms of signal quality and available capacity. Most of these solutions are realized based on Mobile IP [2], which allows the client node to keep the same session IP address when moving across different IP access networks. In combination with IPSec [3] secure IP mobility can be achieved ([4] [5]). The problem of session mobility is based in the routing mechanisms that are used in the Internet [6]. The current IP architecture has an implicit assumption that hosts in the network are stationary. However, Internet hosts have become mobile with the advent of laptops and PDAs with a wireless Internet connection. The Internet protocol stack was not designed with host mobility in mind. Internet addresses are bound to the physical equipment making up the Internet, and are thus bound to physical locations. When an Internet host (e.g., a laptop) moves to a new location, it has to use a new address. This does not have to be an issue since there are automated ways of configuring a new address, e.g., DHCP [7]. However, in the case where end-user devices move between the networks in the middle of ongoing sessions and the Internet address changes, all TCP and UDP sessions will break down. Mobile IP solves this in an elegant way by making sure the mobile host can keep its address while visiting different network locations.

B. Direct communication links

Additionally to the infrastructure based networks, there are also communication technologies that support direct communication between mobile nodes. WLAN for instance offers an *ad-hoc mode* to form an ad-hoc network among nodes that are within the radio range. In contrast to the *infrastructure mode*, where the access point interacts with a complete infrastructure offering user authentication, key management, address assignment and billing, the *ad-hoc mode* treats all interacting nodes equally. Therefore, these nodes have to agree on several settings before they can securely communicate with each other. Whenever two or more nodes want to interconnect using WLAN *ad-hoc mode*, at least one node has to chose a so-called *service set identifier* (SSID), which can be considered as the name of the ad-hoc network. This SSID is then broadcasted so that the other nodes can easily scan for that specific SSID and connect to that ad-hoc network. Nodes sharing that SSID can communicate with each other on the MAC layer, not yet starting TCP/IP sessions. Hence, the nodes have to agree on IP addresses. The whole setup procedure becomes really complicated, when the connection has to be secured. Even if WLAN used with IPSec offers enough protection, authentication and key management has to be handled properly.

When using Bluetooth to interconnect mobile nodes the connection setup process is somehow more user friendly. Bluetooth offers service detection functionality which reduces the user interaction to key management. Whenever nodes want to securely connect using Bluetooth a PIN has to be entered on all nodes. This PIN is then used for shared secret authentication and to derive a session key for traffic encryption. So Bluetooth basically delegates the key exchange problem to the user, which might severely weakens the security level. Most of the users do even disable this security feature to make the usage of Bluetooth more simple.

As explained in [8] and [9] reusing the cellular network to page, authenticate the nodes and exchange configuration and security parameters enables a simple and user friendly establishment of direct communication links. The cellular network offers a secured channel between the participating nodes to exchange sensitive keying information. By reusing the security association each mobile node has with its mobile network operator, the operator can help to build up a trust chain among participating nodes. Further technical details on a possible implementation can be found in section IV.

C. Using all available networking technologies

As discussed in the previous sections there are several efforts going on to simplify the usage of the different available communication technologies and therefore increase the benefit of heterogeneous networks for the end-user.

In the infrastructure based networks solutions like EAP-SIM [10] allow strong security and simple configuration to access public WLAN hotspots. Combined solutions like Swisscom Mobile's Unlimited [1] using Mobile IP, enable a seamless usage of heterogeneous access networks and make therefore a first step towards an always best connected experience. The 3GPP [11] is addressing the integration of heterogenous access networks into the existing cellular networks.

In the domain of spontaneous ad-hoc networking there is a lot of research going on to allow the users to profit from high bandwidth communication that short range communication technologies can offer. Radio technologies for Wireless Personal Area Networks [12] like ZigBee [13] or UWB [14] promise simple short range communication.

To combine these two networking paradigms a certain interaction is required. But the fact that both, infrastructure and ad-hoc based networks offer IP connectivity makes the interaction in terms of session management a lot easier. Protocols like Mobile IP allow a seamless IP session handover. With the route optimization feature of Mobile IPv6 sessions can even be routed directly between nodes using direct ad-hoc links. When combining bootstrapping concepts like the *Cellular Assisted Heterogenous Networking* approach explained in detail in section IV with Mobile IP route optimization, a seamless

handover between infrastructure and ad-hoc based networks can be realized.

Having such an seamless session management in place offering the ability to switch transparently between communication technologies, the handover decision has to be optimized to guarantee always the best connection.

D. Handover decision optimization

In Mobile IP the handover decision is done within the mobile node. Due to missing synchronization with the different access networks this handover decision can not be done in an optimized way. To decide for the best available network some context information is required. The choice of the access technology is strongly dependent on the application that is used. A simple messaging application does not require a broadband access technology, but a low bandwidth and always connected GSM link would be enough. Furthermore, the operators of heterogeneous networks would like to influence this handover decision based on economical considerations. Different access technologies have different deployment and operational costs that have to be somehow reflected in the resource allocation algorithms (in terms of network operation cost, the WLAN network provides a cheaper access than the UMTS network and is therefore preferably allocated whenever possible).

When analyzing scenarios where two or more mobile nodes are connected via different access technologies, the choice of the appropriate communication technology becomes even more complex. For example, in the case where two mobile nodes are connected through two different access technologies (like UMTS and WLAN) the one being connected to the lower capacity access network is limiting the maximum transfer speed. Hence, allocation of expensive resources for the peering node does not increase the overall connection performance, but may result in waste of network resources. If the peering node would be informed about the limited capacity of the other node, it could downgrade its connection and spare capacity reservations.

To cope with these context conditions, several interesting approaches have been proposed in [15] or in [16]. Most of these approaches base on agent technologies to handle the complex handover decision.

This handover decision has to consider best the interests of the operators and the users. Which is not always easy, especially when taking into account that most direct links are free of charge. So it is an absolute must, that these direct links are used whenever possible, even if the operators can not charge for the transferred data. The motivation for the operators is the ability to seamlessly take over the session, whenever direct communication is not possible like shown in figure 2 at t_2.

The next section describes our vision of a user friendly heterogeneous networking.

III. Secured Ad-hoc And Cellular Communications

When users want to establish a secured connection between their mobile nodes, several actions have to be taken. The

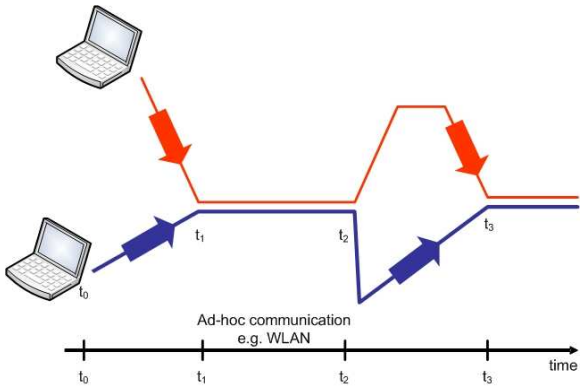

Fig. 1. Usage Scenario

users have to be aware of the capabilities of their devices and often also require certain knowledge about the communication technologies supported by these devices. To show the benefit of a user-friendly seamless connectivity we are considering a scenario were two users want to share some data between their mobile devices (like laptops or PDA). Both users have also cellular phones that can be securely connected to their mobile devices (via Bluetooth, for example). They can be reached on their cellular phones to start a voice conversation, and this paper will show how they could be reached also thanks to their cellular phones to have a data exchange. The scenario can be described as follows: the two users are far from each other (in different cities) but they come close in a given time (for example they could both meet to take the same train) and hence their mobile devices are reachable for ad-hoc connectivity for a lapse of time (between t_1 and t_2). Then the two users diverge (at a given station they take different trains to go work at different places). Some time later the two users could meet again (maybe on their way back home). Figure 1 represents the described scenario.

In today's state of the art networking, users have mainly three possibilities to exchange their data. The first one is to send several emails with all data included. This is what we call an offline centralized data sharing. The second possibility is to stock the data on any server where both users are authenticated and could have access. This is also an offline centralized data sharing capability. The third case is a distributed and online case, which is using the peer to peer capabilities of their devices to establish a connection (for example using the WLAN ad-hoc capabilities). In this later case, if the data to transfer exceeds the time the two users are together they will require finishing the transfer the next time they get close in range. This paper proposes a distributed and online solution that has no distance constraints. A combination of the already existing seamless connectivity products in the market with some novel architecture enabling spontaneous networking capabilities is proposed. Both together they enrich the data communications between users, making a data connectivity as simple as a phone call and they also deliver the data transfer faster than today's capabilities. If we suppose the two users

having the possibility to seamlessly roam from one technology to another, the cellular network can be used whenever the nodes are not in the range of direct WLAN links. Such a heterogeneous session is visualized in figure 2.

In this case the data transfer could be started before the nodes meet each other. The required signaling messages to set up the session can be easily exchanged using SMS (Short Message Service) or USSD (Unstructured Supplementary Service Data). Then the transfer starts first on a cellular technology that is infrastructure based (like UMTS) and will be finished maybe on the same technology or maybe on the ad-hoc WLAN spontaneous networking link in the train when the two user meet together, or maybe later depending on the size of the data exchanged. For all data sizes, the transfer using heterogeneous session will finish earlier than using a pure homogeneous session. Figure 3 visualizes this benefit of using heterogenous networks. All three cases refer to the scenario depicted in figure 2, where a certain amount on data has to be transferred between two nodes. The data session starts at t_0 and ends at t_{finish}. Between t_1 and t_2 and after t_3 the nodes are close enough to directly communicate using the WLAN ad-hoc mode. The first case in figure 3 reflects the data transfer using only WLAN in ad-hoc mode whenever possible (i.e. between t_1 and t_2 and after t_3). When using only UMTS the same data transfer requires even more time. The third case allows the usage of both communication technologies, which has a clear benefit on the required transmission time.

This solution is not only faster for data transfer than today's homogeneous solution, but it also provides added benefits in the communication experience of the user. Due to time based billing and energy constraints users do not stay always connected. Consequently, they only go online when they have to send some data and are hence not online when there is some data that has to be received. Simulations like presented in [17] show that there is a big advantage, when having the ability to use expensive and energy consuming broadband communication technologies only if really required. In other words, it does not make sense to be always connected to WLAN or UMTS when receiving one email an hour. But whenever a email is waiting to be delivered to the mobile node it would be beneficial to trigger that node to connect to WLAN or UMTS to get the email downloaded.

Having such a trigger mechanisms in place would allow to be "virtually" always connected to broadband technologies.

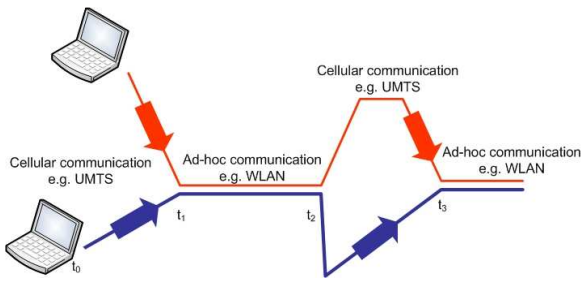

Fig. 2. Heterogeneous Connection

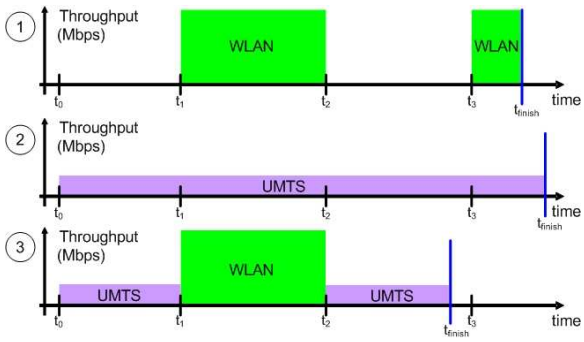

Fig. 3. Heterogeneous Data Transfer

When using protocols like proposed in [8] and [9] over low power carriers like SMS or USSD both the operator and the users could save a lot of resources without loosing the advantage of being always reachable.

The following chapters presents the concepts, architecture and protocols that enable a *Cellular Assisted Heterogeneous Networking*.

IV. Cellular Assisted Heterogeneous Networking

The first requirement to make heterogeneous networking convenient is that each user has his own individual identifier, for instance his mobile subscriber phone number (MSISDN). The sender does not know what type of device the receiver has. Hence, it is up to the receiver to decide which of his devices should be involved in the specific application (for example the data file can be either sent to the laptop or the PDA). This abstraction of the destination node to one statically existing personal identifier helps to solve the problem of temporary identifiers of the different destination nodes. Most of nowadays broadband wireless connections are charged based on time and therefore connected on demand having only temporary valid identifiers like leased IP addresses. Whenever the sender can reach the receiver by a static, personal identifier, the receiver can trigger the appropriate device to become temporarily connected to the broadband access and fulfill the requested transaction. The second main requirement to make heterogeneous networking convenient is probably the automatic choice of the most suitable communication technology to cover the needs of a certain service at the lowest possible costs. This is of high importance when the involved nodes are interconnected by the help of access providers.

The Cellular Assisted Heterogeneous Networking (CAHN) architecture and protocol introduced in [8] and [9] allows the transfer of context information required to choose the optimal communication technology. Moreover, the CAHN protocol allows exchanging information about the networking capabilities of the interacting nodes. Therefore, it is possible to exchange the required configuration and security parameters to interconnect nodes in a very user friendly way using the most appropriate communication technology. The CAHN system hides all the different network devices and their complex configuration from the user. Peers are identified based on their

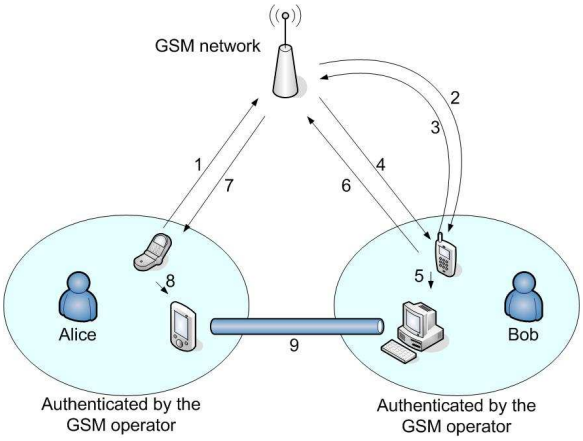

Fig. 4. Setup Process

mobile phone number (aka MSISDN) consequently the connection establishment becomes as easy as setting up a phone call: The user sends a request to the destination nodes and the system determines the optimal links for the interconnection of all participating nodes. This message flow can be seen in Figure 4, where the messages are exchanged as follows:

1) Connection request from Alice to Bob's MSISDN, sent via SMS. The request includes the communication address (i.e. IP address) of Alice's PDA
2) GSM paging to locate Bob's GSM device
3) GSM paging response
4) Delivery of the connection request from Alice via SMS
5) Relay of the connection request from the mobile phone to Bob's computer
6) Connection response including the communication address (i.e. IP address) of Bob's computer and the connection and security parameters to Alice's MSISDN via SMS
7) Connection response of Bob via SMS
8) Relay of the connection response to Alice's PDA
9) Secured link establishment between Alice's PDA and Bob's computer

Note that the cellular network offers the required transport channels (SMS) for the needed information exchange, including authenticated identifier/address resolution and paging mechanisms. Depending on the available links, this connection might be a direct connection using Bluetooth, WLAN (in adhoc mode) or indirect links using any other access technology like WLAN (in infrastructure mode), GPRS, EDGE, UMTS or any other IP capable network like 802.16 or 802.20.

V. Architecture And Protocol Requirements

To use the public SMS service the new protocol messages have to be converted into SMS compliant messages. This conversion could happen on the device, which is intended to be used for the final communication, or on the cellular device. To demonstrate the concept we have chosen the second option and we used an ordinary cell phone as the interface to the GSM network. The control of the mobile phone is done over a serial connection with help of AT commands. The message conversion is realized in the communication device (i.e. laptop). But in the future this capability could be on the GSM device rather than on the laptop. To allow future migration the proposed architecture must be flexible enough to be adapted easily. For that reason we decided to isolate the conversion function from the main application in order to make the main logic independent of the underlying message transport system. This isolated component is called Adapter. For each GSM message delivery mechanism, e.g. USSD, that can be used, a separate Adapter can be written with regard to its characteristics.

The same applies also to the interaction with the communication technology that is used to establish a data connection. The devices can have several communication technologies available, like Bluetooth and WLAN. Therefore, also this part was isolated and the resulting component is called Connector. This component is responsible to apply the parameters agreed on during the communication setup negotiations to the respective network interface card (NIC) and to handle related requests and responses. It is the responsibility of the main logic to choose the Connector in charge for the current communication technology.

This main logic in this architecture is called Communication Module (CM) and it is mainly responsible for the management of the different messages. The CM relays messages to the related component, i.e. to the Adapter, if the messages have to be sent over the GSM network, or to the Connector, if the messages have to be handled locally. Last but not least, the CM offers a standard socket interface to the Adapter, which can also be used by a user interface. For our purpose, we decided to implement a graphical user interface (GUI) to enable a convenient spontaneous networking. With help of this GUI, the user can invoke connection requests and configure his application. Figure 5 shows the schematic structure of the application that was implemented.

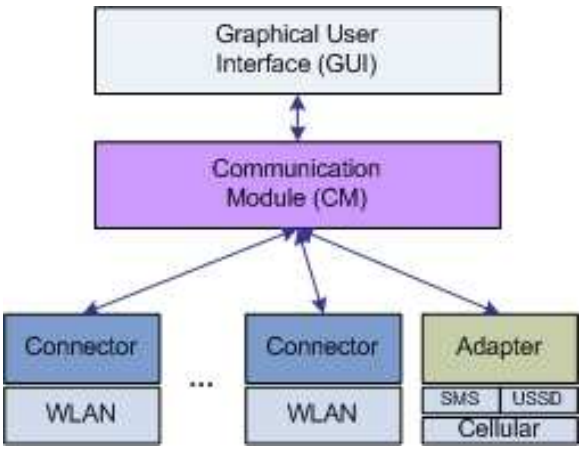

Fig. 5. CAHN Architecture

With this implementation design assures the necessary flexibility to adopt the implementation to support additional GSM message delivery mechanisms and also future communication

technologies. The protocol description is presented in [9] with two implementation scenarios. The first one uses the WLAN technology and the second one is based on Bluetooth. In both cases the GSM network is used to exchange the messages via SMS.

The integration with Mobile IP and its route optimization is ongoing work. It mainly combines standard Mobile IP based seamless access to GPRS, EDGE, UMTS and WLAN and prepares with the help of CAHN the direct links between the users, whenever within the vicinity. Depending on the priority of the user settings, it triggers a Mobile IP route optimization to use the direct link instead of the infrastructure based access network. Preferably, the system keeps the infrastructure based link alive as a fallback channel in case of lost of the direct link.

VI. Conclusion And Outlook

The first part of this paper tried to elaborate on the trends of heterogeneous networks and how its users could benefit from this variety of communication technologies. It seems to a fact, that future communication networks will become even more heterogeneous to meet the requirements of all the different applications. This heterogeneity is fascinating for researchers and engineers, but at the end of the day, users have o be able to handle all these new technologies as well to profit from this heterogeneous networking environment. The business user carrying his laptop, which is enabled with GPRS, EDGE, UMTS, Bluetooth, WLAN needs a system that helps him to chose and configure the device that meets the most his actual requirement.

In the second part a system is proposed that might help to step into that direction of making the use of heterogeneous networks easier. The system is mainly acting as a bootstrapping mechanism to enable the successful use of the different communication technologies. It provides a platform for users to make heterogeneous communication sessions as simple as making phone calls.

The authors are working on further implementation of the described vision and use cases. A special simulation tool is being developed to proof and quantify the actual benefit of cellular assisted heterogenous networking.

References

[1] Swisscom-Mobile, "Mobile unlimited," 2004. [Online]. Available: http://www.swisscom-mobile.ch/bus_asp/bus_home.asp?nid=1675&UserLanguage=E

[2] "Ip routing for wireless/mobile hosts (mobileip)," The Institute of Electrical and Electronics Engineers, Inc., Piscataway, NJ, USA, 2002. [Online]. Available: http://www.ietf.org/html.charters/mobileip-charter.html

[3] "Ip security protocol (ipsec)," The Institute of Electrical and Electronics Engineers, Inc., Piscataway, NJ, USA, 2004. [Online]. Available: http://www.ietf.org/html.charters/ipsec-charter.html

[4] M. Danzeisen and T. Braun, "Secure mobile ip communcation," in *Proceedings of 26th Annual IEEE Conference on Local Computer Networks (LCN'2001)*, 2001. [Online]. Available: http://www.iam.unibe.ch/~rvs/research/publications/wln14.pdf

[5] ——, "Access of mobile ip users to firewall protected vpns," in *Proceedings of WLAN/GIWS*, 2001. [Online]. Available: www.iam.unibe.ch/~rvs/research/publications/secmip_gi.pdf

[6] M. Zivkovic, K. Lagerberg, and J. van Bemmel, "Secure seamless roaming over heterogeneous networks," 2004. [Online]. Available: http://www.ist-albatross.org/

[7] "Dynamic host configuration protocol, rfc 2131," Internet Engineering Task Force (IETF).

[8] M. Danzeisen, T. Braun, D. Rodellar, and S. Winiker, "Heterogeneous networking establishment assisted by cellular operators," in *Proceedings of MWCN*, 2003. [Online]. Available: http://www.iam.unibe.ch/~rvs/research/publications/mwcn03.pdf

[9] ——, "Implementation of a cellular framework for spontaneous network," in *Proceedings of IEEE WCNC*, 2005. [Online]. Available: http://www.iam.unibe.ch/~rvs/research/publications/WCNC05_Danzeisen.pdf

[10] N. H. Haverinen and C. J. Salowey, "Extensible authentication protocol method for gsm subscriber identity modules (eap-sim)," The Institute of Electrical and Electronics Engineers, Inc., Piscataway, NJ, USA, 2004. [Online]. Available: http://www.ietf.org/internet-drafts/draft-haverinen-pppext-eap-sim-16.txt

[11] (2004) 3gpp. The 3rd Generation Partnership Project (3GPP). [Online]. Available: http://www.3gpp.org/Default.htm

[12] IEEE 802.15 working group for wireless personal area networks (WPAN). The Institute of Electrical and Electronics Engineers, Inc., Piscataway, NJ, USA. [Online]. Available: http://www.ieee802.org/15/

[13] (2004) 802.15.4/zigbee. ZigBee Alliance. [Online]. Available: http://www.zigbee.org/en/

[14] (2004) Ultra-wideband (uwb) technology. Intel corp. [Online]. Available: http://www.intel.com/technology/comms/uwb/

[15] B. B. Xiaoxin, "Integrating heterogeneous wireless technologies: A cellular aided mobile ad hoc network (cama)." [Online]. Available: citeseer.ist.psu.edu/681769.html

[16] M. Calisti, T. Lozza, and D. Greenwood, "An agent-based middleware for adaptive roaming in wireless networks," July 2004. [Online]. Available: www.whitestein.com/resources/papers/ubiaamas04.pdf

[17] R. Lagadec, "Optimizing customer experience in mobile data services, tel.con.04, september 22, wienna," 2004. [Online]. Available: http://www.arsenal.ac.at/