Seamless handover within WiMAX and between WiMAX and WLAN

J.R. Damhuis
Faculty of Electrical Engineering, Mathematics and Computer Science
University of Twente, the Netherlands
damhuisjr@cs.utwente.nl

ABSTRACT
This article focuses on seamless handovers within WiMAX and between WiMAX and WLAN. Research has been conducted in this field and solutions are available, but to our knowledge there is no literature study available which summarizes and reviews these solutions. We will provide an overview of ten solutions in total and compare them based on a list of criteria.

Keywords
WiMAX (IEEE 802.16), WLAN (IEEE 802.11), Hard Handover, Soft Handover, Fast Base Station Switching, IP Multicast, ARP, MIH (IEEE 802.21), Mobile IP, FMIPv6, HMIPv6, SIP.

1. INTRODUCTION
Imagine yourself travelling at high speed from A to B in e.g. a train. Would it not be great if you can spend this time working; handle your e-mail, connect to the network at your workplace or even videoconference. In order to do this you need a reliable, high-speed internet connection. When moving from A to B, this connection has to move along with you, along different access points and/or even technologies. Also, you do not want to lose the connection or experience a big delay. As a user you don’t want to notice anything of this handover service. Seamless handover is what you want.

In the past few years there has been an explosive development in wireless access technologies, to fulfill the need of people to be connected at all times. As a result, there are a lot of different technologies, networks, systems, applications and devices. This variety of technology brings a well-known issue to the field of wireless access networks: seamless handover (or: seamless handoff) services.

The most promising new wireless technology on the moment is Worldwide Interoperability for Microwave Access (WiMAX, IEEE 802.16 [IEEE1] [IEEE2]), where Wireless Local Area Network (WLAN, IEEE 802.11 [IEEE3]) is one of the most used wireless technologies nowadays.

WiMAX is a relatively new but very promising standard for wireless communication. Shortly explained it provides the speed of WLAN and the coverage of UMTS (Universal Mobile Telecommunications System). There is literature available covering the most important aspects of WiMAX, like security, frameworks and services. However, we could not find a literature study that provides a comparison between these solutions.

Furthermore, a lot of research is going on about applications, certification, testing and performance, but this is still preliminary. WLAN on the other hand is being developed ever since the mid-eighties. As a result, this standard is far more mature and integrated in today’s society. Personal devices like laptops and mobile phones can use WLAN to establish wireless connections and gain access to, e.g., the internet.

The ideal case would be a combination of both technologies; the wide-spread coverage of WiMAX and the broad applicability of WLAN. This paper describes the conducted research in the area of seamless handover services, both horizontal (handovers within the same technology) and vertical (handovers between different network access technologies or operators) focused on WiMAX and WLAN. The research will eventually provide a clear view of the state of the art in the area of handovers applied within the proposed standards and draw conclusions about the performance. This research will also review which of the available solutions on seamless handover between WiMAX and WLAN satisfy a number of criteria/requirements.

All this is covered by one main research question: “How do the available seamless handover solutions applied within WiMAX (IEEE 802.16e [IEEE2]) and between WiMAX and WLAN (IEEE 802.11 [IEEE3]) perform based on certain criteria/requirements?”

In order to answer this main question correctly and completely, sub questions are proposed. These questions decompose the main question into more tangible subjects:

1. What are the main handover aspects/issues in a wireless communication technology?
2. Which seamless handover solutions can be applied within WiMAX?
3. Which seamless handover solutions can be applied between WiMAX and WLAN?
4. Which requirements/criteria need to be satisfied to provide seamless handover?
5. Can the seamless handover solutions applied in WiMAX and between WiMAX and WLAN satisfy the requirements/criteria?

The research will be based on a literature review and analysis. By extensive reading we will achieve knowledge about the subject in order to elaborate about research questions 1, 2 and 3 and answer them. When we have gathered enough information about the subject, we will analyse the found solutions, described by research question 4 and 5.

The contents of the paper are as follows. In the following section we will describe WiMAX and WLAN and discuss some
main handover issues. By doing so, the first research question will be answered. Section 3 introduces several handover solutions available for WiMAX and between WiMAX and WLAN. In this section, research question 2 and 3 are being answered. Questions 4 and 5 will be discussed in section 4. In this section we will introduce the criteria for comparison and eventually use these to compare the solutions. Section 5 concludes the article and we will list some interesting open issues for future work.

2. WIMAX AND WLAN

When you think ‘wireless’, you almost automatically think ‘radio’. This is obvious as wireless communication is based on radio technology. But when you look beyond the radio waves, you will see that there is a lot more to it. There are a lot of ways to implement functionality and capability into the network that create value for operators and end-users. GSM (Global System for Mobile communications), GPRS (General Packet Radio Service), UMTS, WLAN and WiMAX are some examples of different technologies, each with their own characteristics. As already emphasized, this paper will focus on WiMAX and WLAN. This section will contain a brief overview of WiMAX and WLAN. After this, the main issues concerning seamless handovers will be discussed.

2.1 WiMAX

WiMAX is a wireless telecommunication technology based on the IEEE 802.16 standard [IEEE2]. It uses licensed spectrum to provide high speed wireless data transmissions over long distances in many different ways. There are a lot of different standards, from 802.16a to 802.16m. Nowadays there are two versions which are interesting for common use: 802.16-2004 and 802.16e [SFC05]. More familiar terms for these standards are Fixed WiMAX (802.16-2004 [IEEE1]) and Mobile WiMAX (802.16e [IEEE2]). By definition, Fixed WiMAX does not support handovers and is therefore not interesting for this research. That is why for the rest of the paper, when the phrase WiMAX is used, Mobile WiMAX (802.16e) is meant. The current version of this standard provides mobility support at frequency bands between 2 and 6 GHz, with vehicular speeds up to 120 kilometres per hour. Cell sizes can be up to 70-80 kilometre per cell with average speeds of 30 Mbps in a cell. But speeds of 10 Mbps at 10 kilometres distance are more likely, see [IEEE2] and [INV07]. These specifications describe the air interface between a Base Station (BS) and a Subscriber Station (SS). WiMAX supports point-to-multipoint (PMP) mode and mesh-mode. The difference between these modes is the way SSs connect to BSs and the network. Where in PMP mode every SS makes its own connection to the BS, in mesh-mode this is done by and via other SSs [MET07]. This difference is shown in Figure 1.

WiMAX uses MIMO-SOFDMA radio technology in the physical layer. The first part, MIMO, stands for Multiple Input Multiple Output. This means that both transmitter and receiver use multiple antennas in order to improve coverage, installation, bandwidth efficiency and performance. The second part, SOFDMA, stands for Scalable Orthogonal Frequency Division Multiple Access. This can be described as a division of the frequency band into several sub-carriers. A digital modulation scheme is used to provide every SS with a little piece (a sub-carrier) of the available bandwidth. So, different OFDM sub-channels are assigned to different users. This allows simultaneous data rate transmissions from several SSs. The scalability of the technology provides support for scaling the channel bandwidth in order to keep the spacing of the sub-carriers constant across different bandwidths [SFC05] [SH+07].

Furthermore, WiMAX MAC uses a scheduling algorithm to allocate an access slot to SS. Once obtained, the bandwidth can vary but the slot remains assigned to the SS, so a stable connection is made. The MAC layer consists of three sub-layers, each with their own tasks like bandwidth allocation, system access, connection related mechanisms and security. A more detailed description is not relevant in this context but can be found in e.g. [MET07].

2.1.2 Architecture

The WiMAX Network Architecture defines a framework consisting of several functional entities and interconnections. Figure 2 shows this framework in a simplified manner, followed by a description of each entity [INV07].

- NAP: Network Access Provider
  A business entity that provides WiMAX radio infrastructure.
- NSP: Network Service Provider
  Just like the NAP, the NSP is a business entity. It provides IP connectivity and WiMAX services. The level of services is legally binded through contractual agreements with one or more NAPs.
- SS/MS: Subscriber Station/Mobile Station
  Entity which wants to make a connection to the network.
• ASN: Access Service Network
  This is the point of entry for the SS/MS into the WiMAX network. This entity must support a complete set of functions required to connect a client to the network: authorization, authentication, session management, network discovery, IP-address allocation, QoS etc.

• CSN: Connectivity Service Network
  The CSN is the part of the network which provides IP connectivity services. It consists typically of routers, servers, proxies, and gateways etc. providing functions like Internet access and peer-to-peer services.

Besides these entities the architecture also contains a number of interconnections or reference points. The most important and relevant ones are summarized here.

• R1: Protocols between SS/MS and ASN including PHY and MAC layers as specified by the 802.16 standard.
• R2: Protocols/procedures between SS/MS and CSN concerning authentication, authorization and IP-configuration management.
• R3: Control procedures between ASN and CSN. Provides tunnelling of user data between the two entities.
• R4: Control procedures between ANSs like MS mobility between different ANSs.
• R5: Control procedures for supporting roaming from a home NSP to a visited NSP.
• R8: When switching between different BSs within the same ASN or between different ANSs (which most likely will also involve a switch between BSs) this is an optional reference point to ensure fast and seamless handover through direct transfer of MAC context and data.

Together, the technology and network architecture give a summarized and simplified view of WiMAX networks.

2.2 WLAN

WLAN is a very popular wireless communication technology for short/medium distances nowadays, mostly because it is convenient, easy to deploy, easy to manage and because of its low infrastructure costs. It is based on the 802.11 standard and also knows a lot of different versions; varying from 802.11a to 802.11j (and even to 802.11y in the future) [IEEE3] [IEEE4].

A typical WLAN network consists of Access Points (APs) and several wireless clients, called Stations (STAs). It uses unlicensed spectrum radio waves to communicate. Cell radii are up to 100 meter in open field with a maximum speed of 54 Mbps (802.11g standard). WLAN supports PMP mode and ad-hoc mode. Combinations are also possible; this is a so-called hybrid network. Today, the standard 802.11g, generally referred to as Wi-Fi, has been implemented all over the world. For the remaining part of the paper, when WLAN is used, 802.11g is meant. Just like every other 802.x standard, the MAC and PHY layers are specified. Again, just like WiMAX, de physical layer specifies the modulation scheme used and signalling characteristics for the transmission through the radio frequencies, whereas the MAC layer defines a way accessing the physical layer [MET07].

2.2.1 Technology

WLAN uses radio waves at 2.4 GHz to communicate. It utilizes OFDM modulation technology in the physical layer. OFDM stands for Orthogonal Frequency-Division Multiplexing and is similar to SOFDMA which is used by WiMAX. OFDM also divides the frequency band into subcarriers. Modulation is done by one of four modulation schemes: BPSK, QPSK, 16-QAM and 64-QAM. Details will not be provided here but further reading can be done in e.g. [CFL04]. The difference between OFDM and (S)OFDMA is the way the sub-carriers are assigned to the users. This is shown in figure 3 [SFC05]. With OFDM, all sub-carriers are transmitted parallel with the same amplitude. The same bandwidth is assigned to every user. OFDMA divides the carrier space into groups, each with a number of carriers. Sub-channels are allocated to users depending on the channel conditions and their data requirements [SFC05]. To summarize this: OFDM is contention-based which means that clients have to compete for the use of the facilities of the system. OFDMA uses slots which are assigned to the clients during the time of the connection, so the connection is maintained at all times.

The MAC layer in WLAN utilizes a contention-based scheme called Distributed Coordination Function (DCF) to manage client connections. To avoid collisions, Carrier Sense Multiple Access/Collision Avoidance (CSMA/CA) is used. The layer handles several tasks. Some examples are authentication, encryption, operating modes (infrastructure or ad-hoc) and QoS [MET07].

2.2.2 Architecture

As stated before, a typical WLAN network consists of two entities: APs and STAs. APs are the Base Stations for a wireless network. They communicate wireless with STAs by radio waves. STAs are wireless clients which connect to the APs in order to gain access to the network, e.g. the Internet. Furthermore, the architecture defines service sets. A service set is a set of all stations which can communicate with each other. In WLAN, two service sets can be distinguished, see [MET07]:

• Independent service set.
  This is a service set with no access point. Thus the network has no outside connection, also known as an ad-hoc network.

• Infrastructure service set.
  A service set that can communicate with other service sets using an access point and an access network.

These two service sets are used to easily set up the two most common WLAN-networks; infrastructure and ad-hoc.
2.3 Handovers

Handovers are an important part of a network technology. When moving between different BSs, the connection also has to move. Seamlessness in this paper is defined as follows: the current session, QoS and Service Level Agreements (SLA) must be maintained during and after handover. In other words, a seamless handover is a handover that is seamless to the user. Obviously this also depends on the kind of service the user is requiring. With real-time applications like videoconferencing or streaming media, the user will probably notice a decrease of the connection. On the other hand, while browsing a website or transferring a file, the user does not have to notice anything of the handover process. The latency and packet loss are the two crucial factors for seamless handover. These two factors have to be as small as possible to make the handover seamless [TTL99].

Before discussing several handover issues, the next part will go into detail about handovers in general first. There can be several reasons why and when a handover should be initiated [RAP06]:

- **MS current position and velocity**
  High velocity can result in different handover decisions.
- **Link quality**
  Another BS can deliver a higher quality link (e.g. higher speed, stronger signal, better QoS).
- **Load at a BS**
  When a BS in a subnet is currently overloaded; the network can decide to relocate some MSs.
- **Conserving battery power**
  In order to save battery power, a MS can choose to switch to a closer station to be more energy efficient.
- **Context and requirements**
  When a MS requires different type of service, it can be necessary to switch BS.

There are two types of handovers: horizontal (handovers within the same technology) and vertical (handovers between different network access technologies). Horizontal handovers are Layer-2 handovers (L2HO, also referred to as ‘micro-mobility’). Here, only the BS is changed and IP-information is maintained. Typically this causes small latency and low packet loss. Vertical handovers are Layer-3 handovers (L3HO, also referred to as ‘macro-mobility’), it also changes the IP attachment point and so IP information is changed too. Because of these changes, latency and packet loss are significantly higher [HPF07]. Besides these differences, there are more problems to consider. This will be discussed in the following sections, starting with horizontal handovers.

2.3.1 Horizontal handover

SSs in WiMAX are referred to as Mobile Subscribers (MS). The difference between a SS and a MS is that a MS can change its BS with a handover mechanism when connected to the network. Both BS and MS can initiate the handover. In most of the cases, MS initiates the handover, because of one of the reasons mentioned in section 2.3. The only logical reason why the network (BS) initiates a handover is when a BS is getting overloaded. With this handover, only layer 2 is involved as the technology stays the same and thus the IP address stays unchanged.

Because WiMAX is based on OFDMA technology, a MS basically uses hard handover (HHO, break-before-make handover) when moving to another BS [SG+05]. During handover, all connections are broken at all layers and no context information is shared between BSs. Therefore, the MS cannot receive and/or transmit any data during handover and packet loss occurs. Latency is on the order of around 1000 ms (milliseconds) or more [SG+05]. This is why seamless handover can not be realized in this case. But besides hard handover, recent WiMAX standard also supports soft handover (SHO, make-before-break) and Fast BS Switching (FBSS), both supporting seamless handover. These solutions, and on other, will be clarified in section 3.1.

2.3.2 Vertical handover

These handovers are more complex because they involve both L2Hos and L3Hos. WiMAX and WLAN use different protocols, different air technologies and have different QoS. Besides these differences there are similarities on some level. Both specify the MAC and PHY layers and use IP-technology to identify the network entities. This is why Mobile IP is developed, and used most of the times to support handovers between different technologies [TTL99] [RMB06] [JRJ07].

Because of the switch of technology, the MS enters another subnet, so its IP address changes too. This can be done in several ways, as the number of solutions in section 3.2 shows.

3. WIMAX/WLAN HANDOVER SOLUTIONS

A lot of research is already done in the field of handovers. With the new and promising wireless broadband solutions, the capability to support fast and reliable handover is critical for its success. Most solutions provide extensions and/or improvements of the current standards in order to improve certain aspects.

Hereafter, several solutions will be introduced and briefly explained how they work. Of course, these are only a few of the ones available. The solutions will be divided into their applicability; horizontal or vertical handovers.

3.1 Horizontal handover solutions

The WiMAX standard, see [IEEE2] [HPF07], defines three basic handover protocols: Hard Handover (HHO), Soft Handover (SHO) and Fast BS Switching (FBSS). These will be explained here. After that, another solution will be discussed: Enhanced ARP Handover.

3.1.1 Hard Handover

The WiMAX standard specifies the general HHO process as stated in figure 4 [LKU06]. Before handover initiation, the MS and serving BS conduct network topology acquisition, backed up by the backbone network. BS broadcasts the network topology information for a period of time using MAC message transmissions. These messages contain network information about neighbouring BSs. Then MS is able to select a candidate BS for handover through scanning. After this, MS associates with candidate target BSs. MS records the obtained information so it can be used for future handovers.

After the topology acquisition, the handover process is performed as follows. MS conducts cell reselection based on the acquired information of the network. Then target BS is decided and handover is initiated. MS synchronizes with new downlink and obtains uplink parameters. Initial ranging process is executed; this may be done contention-based or non-contention-based. Finally MS re-enters the network by re-authorizing and re-registering. At last MS terminates the connection with
serving BS [LKU06]. As said, seamless handover is not possible because all connections are broken. But for non-real-time applications this can be sufficient.

### Network Topology Acquisition

#### Before H0
- Network topology advertisement
- MS scanning of neighbor BSs
- Association procedure

#### During H0
- Cell reselection
- HO decision & initiation
- Synchronize with new downlink and obtain parameters
- Obtain uplink parameters
- Ranging and uplink parameter adjustment
- MS re-authorization
- Re-register
- Termination with the serving BS

![Figure 4. WiMAX hard handover process [LKU06]](image)

#### 3.1.2 Soft Handover
SHO (Soft Handover, make-before-break handover) is similar to HHO but with one major difference: with SHO a MS is registered to multiple BSs (the Active Set) at the same time. During SHO two or more BSs are sending and receiving the same information to the MS. This way, seamless handover is possible but it utilizes more resources, see [SG+05].

#### 3.1.3 Fast BS Switching (FBSS)
With FBSS, the MS is only sending/receiving data to/from one of the BSs in the Active Set. The BSs in the Active Set communicate through the backbone network to share context information. Based on this context information, handover decisions are made and handovers initiated. Again, seamless handover is possible but more network resources are used [SG+05]. But compared to SHO, fewer resources are used at the MS, because it is connected with only one BS at a time.

#### 3.1.4 Enhanced ARP Handover
ARP stands for Address Resolution Protocol, see [IETF1], which is a very old draft. During the years, several improvements are proposed. It is a protocol which can be used within a subnet to map an IP address to the corresponding MAC address of the device. Then, this address is used to send packets to. This handover solution is an enhanced version of the original ARP protocol. In WiMAX, BSs can act as an ARP proxy in their cell. The BSs send out beacons and buffers packets for an MS. When an MS receives a stronger beacon from another BS, a handover is initiated. Figure 5 shows the handover [TTL99].

This handover process follows the steps described next:
1. MS sends greet-message to new BS with its own address and the address of the old BS.
2. New BS sends an acknowledgement and creates a routing entry for MS.
3. New BS sends notify-message to old BS together with its address.
4. Old BS deletes MS’s entry and sends buffered packets to new BS along with a notify-ack-message.
5. New BS broadcasts a redirect message to routers to update the ARP cache.

Because of the limitations of this handover solution to the same subnet, it can only be used for horizontal handovers. Nonetheless, it is a very fast handover technique which claims to complete handover in less than 10 ms [TTL99].

![Figure 5. Enhanced ARP Handover [TTL99]](image)

#### 3.2 Vertical handover solutions
As stated before, vertical handovers are more complex compared to horizontal handovers. There are several solutions available in literature, but none of them are standardized yet. In this case, vertical handover solutions are applied to WiMAX/WLAN handovers but can sometimes be used in other handovers as well.

##### 3.2.1 Mobile IP
There are two main types of Mobile IP; Mobile IPv4 [IETF2] and Mobile IPv6 [IETF3]. Mobile IP has three basic functional entities: the mobile node (MS), the Home Agent (HA) and the Foreign Agent (FA). When a MS moves to a foreign network with another technology, it acquires a care-of-address (COA), which uniquely identifies the MS in the foreign network. This COA is registered at the MS’s HA on the MS’s home network. Traffic is then tunnelled by the HA to the COA of the MS. This way, the change of IP address of the MS is hidden, in order to keep all TCP connections (Telnet, FTP etc.) alive.

This can be done in several ways. Mobile IPv4 basically uses triangular routing, see Figure 6. The MS sends data directly to the corresponding host (CH), whereas data for MS is send through the HA and the FA. The total process takes time and suffers from latency and packet loss. Also, because of triangular routing, packets are send along paths which can be much longer and therefore less efficient.

![Figure 6: Mobile IP triangular routing](image)
Mobile IPv6 benefits both from the experiences gained from the development of Mobile IP support in IPv4 and from the opportunities provided by IPv6. Thus, Mobile IPv6 shares many features with Mobile IPv4 but it also offers many improvements. The biggest benefit is the support for route optimization. It uses binding updates to inform the CH about the current IP address of the MS so traffic can be sent directly to the MS. Other improvements are described in [IETF3] but will not further be discussed here.

### 3.2.2 IP Multicast
The IP protocol provides three types of communication: unicast (sending data to a single receiver), broadcast (sending data to all the receivers on a given network) and multicast (deliver the data to a set of selected receivers). In IP Multicast only a single packet is sent by the source. The network duplicates this packet until all the intended BSs receive the packet. The BSs buffer these packets. The serving BS will forward it to the MS it is meant for. When a handover is initiated, a MS messages to both old and new BS and the new BS starts to forward the packets to the MS, see [TTL99] [RMB06].

### 3.2.3 MIH
MIH stands for Media Independent Handover. It is an interworking standard, IEEE 802.21, see [IEEE5], which is being developed to support handovers between any wireless access technologies [DJW07]. MIH defines an extra layer between Layer 2 (IP Layer) and Layer 3 (wireless link layer) in the protocol stack. See Figure 7 for a graphical representation. The MIH layer facilitates messaging among IP and the various wireless technologies in order to select the appropriate network for handover.

![Figure 7. The MIH Function [DJW07]](image)

As can be seen in Figure 6, the MIH layer can convey three types of messages:
- **Events**
  - A lower layer informs a higher layer of an event. Examples are: wireless link quality degrading and handover status.
- **Commands**
  - A higher layer issues a command to a lower layer. Examples are: initiate handover and complete handover.
- **Information**
  - The two layers exchange information. Examples: current QoS, performance information and availability of services.

The standard specifies these messages for different handover combinations. This way, different technologies can communicate efficiently through the MIH layer [DJW07].

### 3.2.4 FMIPv6
One of the disadvantages of Mobile IP handover is high latency. This latency is defined as the time frame in which the MS is unable to send or receive any packets because of link switching and IP reconfiguration. FMIPv6 stands for Fast Mobile IPv6 Handover, see [IETF4]. It is an improved solution of the Mobile IP protocol. Fast Handover provides seamless handover using IPv6 address space and a Layer 2 trigger [PaC03]. Figure 8 shows the operation of the FMIPv6 handover protocol.

Either MS or Old BS may initiate the handover procedure by using the L2 trigger. If MS initiates and old BS receives the trigger, the MS will initiate L3 handover by sending a Router Solicitation for Proxy (RsSolPr) message to the old BS. Otherwise, if old BS initiates, it sends a Proxy Router Advertisement (PrRtAdv) to the MS.

After this, the MS must obtain a new COA, just like with Mobile IP. The difference here is that MS is still connected to the old BS. The old BS validates the new COA and establishes a connection tunnel between old and new BS by sending a Handover Initiate (HI) message. The new BS verifies the new COA and sends a Handover Acknowledgement (HACK).

Before disconnecting, the MS should send a Fast Binding Update (F-BU) to update the binding cache with the new COA. When the old BS receives the F-BU message, it must verify the handover with the new BS. Then it begins forwarding packets from the old COA to the new BS. When the MS connects to the new BS, the new BS starts sending packets to the MS and the handover is completed [PaC03].

![Figure 8. Fast Mobile IPv6 Handover Protocol [PaC03]](image)

### 3.2.5 HMIPv6
HMIPv6 is Hierarchical Mobile IPv6 Mobility Management, see [IETF5]. Also, this is an improved solution of the Mobile IP protocol. Hierarchical schemes reduce handover latency. It does so by employing a hierarchical network structure. This structure separates mobility into micro- and macro mobility (see section 2.3). To support this, a special network entity is placed into the edges of the network: the Mobile Anchor Point (MAP). This entity is a router or a set of routers that maintains a binding between itself and MSs currently in its domain. The MAP acts as the HA of the MS and tunnels all the packets through the COA to the MS. When the MS moves within the local MAP domain (micro mobility), it only needs to register the new on-link address with the MAP because the COA does not change. When the MS moves to a new domain (macro mobility), it acquires a regional COA and an on-link COA. The MS then
uses the new MAP’s address as the regional COA. Then, MS sends a binding update (BU) message to the MAP in order to bind MS’s regional COA to the on-link COA. In addition, the MS sends BUs to its HA and the CNs in order to specify the binding between its home address and the regional COA.

3.2.6 SIP

Session Initiation Protocol, see [IETF6], is an application-layer protocol used to establish and tear down multimedia sessions [WeS99]. Because SIP is a protocol which is handled at the application layer, there is no need for tunnelling of the data stream.

Entities in SIP are users, proxy servers and redirect servers. A user is addressed using an e-mail-like address, e.g. user@host. The user is obviously the user and ‘host’ is the domain name. Data is exchanged during the session in some sort of peer-to-peer mode [JRJ07]. SIP defines a number of methods, see table 1.

A SIP user has two functions: listening for SIP messages and sending SIP messages upon occurring events. The proxy server relays SIP messages, so the use of host names is possible without knowing the IP address. A redirect server returns the location of the host. Both proxy servers and redirect servers accept user registrations, in which the current location of the user is given. If a user goes mobile during a session, it must send a new INVITE to the corresponding host using the same call identifier as in the original call. It should send the new IP address or host name, so the mobile user knows where to send the SIP messages to. Also, the mobile user has to re-register the new location to the proxy- and redirect servers.

Table 1. SIP messages [WeS99].

<table>
<thead>
<tr>
<th>Message name</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>INVITE</td>
<td>Invite user(s) to a session.</td>
</tr>
<tr>
<td>ACK</td>
<td>Acknowledgement of an INVITE message.</td>
</tr>
<tr>
<td>BYE</td>
<td>Release session.</td>
</tr>
<tr>
<td>OPTIONS</td>
<td>Asks server capabilities.</td>
</tr>
<tr>
<td>CANCEL</td>
<td>Cancel pending request.</td>
</tr>
<tr>
<td>REGISTER</td>
<td>Register with a SIP server.</td>
</tr>
</tbody>
</table>

[JRJ07] proposes an improvement of the standard SIP protocol. It uses constant destination identifiers so the user does not have to re-invite the correspondent host. Sessions do not need to be terminated every time. The user updates the registry at the SIP servers and packets from the CH are automatically sent to the user.

4. HANDOVER SOLUTION COMPARISON

In order to compare the solutions, a list of criteria is necessary to review each solution. The following list of criteria will be used to compare the solutions:

- Handover type.
  - Which type of handover is supported by the solution? In this case the only to types relevant are seamless or non-seamless handovers.
- Complexity.
  - How complex is the solution? This is measured in terms of the number of needed entities, protocol layers and/or hardware requirements.
- Dropped packets.
  - During handover, packet loss is inevitable. Of course, there is a difference in the number of dropped packets. Again, obviously, less packet loss is better.
- Duplication of events/Redundancy.
  - Are there duplications of events and/or packets during the handover? More redundancy uses more network resources and is thus less efficient.
- Scalability.
  - Is the solution applicable with a large number of users? How does this scalability influence the performance? Performance must be maintained as much as possible when more users are added to the domain.

It is important to note that the performance results used in the given comparisons are taken from different papers that each described the different solutions separately. This means that the used network topologies and settings are different and therefore, the performance related comparison given in the following sections could be inaccurate.

4.1 Horizontal handover solution comparison

In this section, horizontal handover solutions will be compared based on the proposed criteria. This will be done by discussing each solution. To provide a clear overview of the differences and in order to make a good comparison, a table is introduced. This table is set up as follows: solutions are placed in rows and the criteria will be put in the columns. To compare the solutions and review them, values are introduced: high, medium and low. In this case, there’s no general good or wrong, but each value must be read within the right perspective. For example: the value ‘high’ for the criteria ‘Handover Latency’ is not good, whereas this value for ‘Scalability’ means the solution scores well. For handover type there are two types: non-seamless; N-S and seamless; S.

Table 2 gives an overview of the comparison of horizontal handover solutions. Thereafter, the table will be clarified in the discussion, with an explanation of the choices made.

Table 2: Horizontal handover solution comparison

<table>
<thead>
<tr>
<th>Handover latency</th>
<th>Handover type</th>
<th>Complexity</th>
<th>Dropped packets</th>
<th>Duplication of events</th>
<th>Scalability</th>
</tr>
</thead>
<tbody>
<tr>
<td>HHO</td>
<td>high</td>
<td>N-S</td>
<td>low</td>
<td>high</td>
<td>high</td>
</tr>
<tr>
<td>SHO</td>
<td>med</td>
<td>S</td>
<td>med</td>
<td>med</td>
<td>med</td>
</tr>
<tr>
<td>FBSS</td>
<td>low</td>
<td>S</td>
<td>med</td>
<td>low</td>
<td>low</td>
</tr>
<tr>
<td>ARP</td>
<td>low</td>
<td>S</td>
<td>low</td>
<td>low</td>
<td>high</td>
</tr>
</tbody>
</table>
4.1.1 HHO
According to [LKU06] the HO operation time with HHO and no further improvements lies between 700 ms and 900 ms. This depends mainly on the cell-load, where 700 ms is reached with a cell load of 0% and 900 ms with a cell load of 50%, see [LKU06]. The handover type is non-seamless because it uses a break-before-make handover. Complexity is low because the HO process is pretty straightforward with just two BSs (old and new) and the MS. Because of the break-before-make nature, it is very likely that packets sent to MS are dropped because the MS cannot receive any packets during handover. There is no duplication of events, the BSs do not share any context information and MS just sends/receives packets from one BS. Finally, because of the low complexity of the solution, it is scalable; with its limitations in the number of MSs one BS can serve. But compared to other solutions where BSs have to do a lot more and are overload faster, this solution is scalable.

4.1.2 SHO
Because SHO uses contention-based ranging, handover latency is a lot lower compared to HHO; according to [SG+05] this is approximately 300 ms. Thus, the overall handover latency can be considered as being medium. Because of the contention based ranging, the type is make-before-break and therefore seamless. This seamlessness is merely theoretically because a latency of around 300 ms is relatively large and thus can be noticed by the user. Complexity is also medium, there are more entities involved with the handover. The Active Set contains more BSs compared to the number of BSs with HHO and because the make-before-break nature, the handover process is more complex. Obviously there are less dropped packets because handover is seamless. MS is registered at multiple BSs. Because of this, there is some duplication of events. Last, the solution is scalable just like HHO, but because BSs have to keep more MSs registered, they overload sooner.

4.1.3 FBSS
Fast Base Station Switching is like SHO but because it uses fast-ranging and context sharing, latency is further reduced. According to [SG+05], values vary between 50 and 100 ms, therefore seamless handovers can be supported. Complexity is generally medium, just like SHO, but there is a difference. Complexity at the MS is lower then with SHO, because MS is connected to just one BS. But at the BSs, complexity is slightly higher because they have to share context information about the MSs through the backbone network. But this context sharing combined with lower handover latency also limits packet-dropping. Duplication of events is low because BSs communicate current states of MSs through the backbone network, reducing redundancy. Just like with SHO, scalability depends on the BSs. They tend to get overloaded with too much of MSs connected. When this happens, too much background information has to be shared and performance drops significantly.

4.1.4 Enhanced ARP
According to [TTL99], the total handover latency can be less than 10 ms. This is backed up by experiments and is the lowest handover latency thus far. This provides real-time, seamless handover and will not be noticed by the user. Complexity is low because de BSs serve as ARP routers. The only complexity is the buffer-size of the ARP routers. This value has to be determined based on the desired performance and the number of users; more users mean smaller buffers and thus less performance. Packet loss also depends on the buffer size. But the buffer and the low latency value ensure very limited packet loss. When the old BS forwards packets to the new BS there is a possibility of duplicated packets but because handover latency is low, this is not really an issue. Finally, the solution is highly scalable. Every BS can serve as an ARP router and vary the size of the used buffer based on the number of connected MSs.

4.1.5 Conclusion
From the discussion above and table 2 can be concluded that ARP Handover and FBSS can provide the best performance based on the given criteria. However, as said before, this performance related comparison could be inaccurate.

4.2 Vertical handover solution comparison
This section will elaborate about the comparison of vertical handover solutions. The structure of the section will be the same as section 4.1. First, a table is provided to give a clear overview of the comparison, followed by a detailed discussion to explain how the values are justified.

| Mobile-IP | high | S | med | high | med | med |
| IP-Multicast | low | S | low | low | high | high |
| MIH | low | S | high | low | high |
| FMIp6 | med | S | high | med | high |
| HMIPv6 | high | S | high | med | med |
| SIP | low | S | low | low | high |

4.2.1 Mobile IP
Due to the fact that Mobile IP uses the HA in the data communication, the Mobile IP latency can be in some situations significant. Without any improvements, and according to [HS+02], the handover latency can reach 5000 ms. The handover is seamless because the IP session is maintained, but it will definitely not go unnoticed by the user. Mobile IP uses a rather complex construction with the HA and COA. Because of the high handover latency, and long paths from CH to MS, the packet dropping probability is high. Because all traffic is tunneled through the HA, the COA is added to each packet which causes a lot of overhead, even up to 25%. This is not really ‘duplication of events’ but surely decreases efficiency.
Every MS uses its own HA and acquires its own COA. This might affect the scalability negatively.

### 4.2.2 IP-Multicast

According to [TTL99] and [RMD06], the average delay using IP-Multicast handover is between 20 and 55 ms. This is because it is in basis an uncomplicated solution. When a MS roams to another multicast network, it only has to register with the new BS and the new BS starts forwarding the packets which are destined for the MS. Thus handover is seamless with very few dropped packets. The greatest disadvantage of the solution is that there is a lot of duplication; all BSs receive all packets for the MSs. The solution is in fact scalable, but implementation costs are relatively high because a lot of users require a lot of BSs.

### 4.2.3 MIH

When MIH is used, according to [DJW07], the handover latency is approximately around 200 ms. This is done by the extra added MIH layer which makes it perfect for fast handovers but a rather complex solution because it uses an extra layer which must be implemented between upper- and lower layer. Despite this, packet-loss is minimal because low latency and efficient communication. The MIH layer takes care of communication between the upper and lower layer, resulting in a very efficient solution and minimal duplication of events. Also, the solution is highly scalable since it uses no extra network entities but only an additional layer which is used by all users.

### 4.2.4 FMIPv6

Fast Mobile IPv6 Handover has improvements regarding regular Mobile IPv6. According to [PaC03], it reduces handover latency to 350 ms. Because of smarter routing, the solution itself is less complex compared to Mobile IP. But these improvements come at extra costs in terms of overhead and signalling costs. Thus, the improvements cause more duplicated events. Scalability can be considered medium, just like the other Mobile IP solutions.

### 4.2.5 HMIPv6

Handover latency when using HMIPv6 is seamless and according to [HS+02], takes approximately 700 ms to complete. There is an extra physical entity introduced on the edges of the network, the MAP. Next to the MAP, also two COA addresses are needed. This makes the solution more complex compared to IP Multicast but less complex then the MIH or Mobile IP solutions. The same counts for packet-loss. Because of lower latency and less complexity, packet-loss is medium. During handover, duplication of events can occur. It is possible that the new MAP already starts sending packets to the MS while the old MAP is transferring its last, when MS has not informed the old MAP of re-registration. Scalability in this case depends on the number of MAPs, of course with respect to the implementation costs.

### 4.2.6 SIP

When a MS uses SIP and moves to another domain, the only thing the MS has to do is to re-register at the SIP server. Because of this, the handover delay can be very low. According to [JRJ07], around 50 ms. The direct connection to the CH is maintained through the SIP server, so handover is seamless. SIP is just an application protocol and thus not complex at all. The low latency in this solution means low packet-loss. Scalability is high since redirect servers are used. A redirect server returns the location of the MS. This makes it possible to build highly scalable servers, because these servers only have to return the location of the MS instead of participating in the entire transaction.

### 4.2.7 Conclusion

Again, note that inaccuracy still is possible, due to different information sources. Based on just the table, SIP and FMIPv6 prove to be the best solutions to provide fast and seamless handover. The main problem of SIP is related to the fact that it is only supported at the application layer. So typically it has to cooperate with other solutions, for example with MIH. Another realistic solution is to use a combination of Mobile IP, HMIPv6 and FMIPv6.

Furthermore, the comparison shows that lower latency often comes with more complexity and thus higher implementation costs. A simple, no-nonsense solution has unacceptable high latency, whereas a fast solution is far more complex and difficult to implement. So, settlements must be made between the factors; an acceptable handover latency, low complexity and as little duplication as possible.

Ultimately, it is impossible to point out ‘one winner’ here, because of these considerations.

### 5. CONCLUSIONS AND FUTURE WORK

This paper focused on one of the most important issues of wireless air technologies; seamless handover. As wireless technologies are emerging, the need for constant connectivity is growing. This can not be done without the possibility to roam between different networks and different air technologies. The process which facilitates this is called a handover. Seamless handover is the kind of handover the user does not notice, and thus the preferred one.

Unfortunately, it is not always that easy to support seamless handover. There is a period of time in which a user can not receive or send any data because it has to switch between access points or even change its IP-address. This period of time is called handover latency and is one of the most crucial factors in the success of a handover protocol. Furthermore, there are more issues to consider when reviewing handover solutions: packet loss, complexity and inefficient use of resources can cause serious problems.

As the most promising new wireless access technology, the research first focused on WiMAX. WiMAX promises coverage of UMTS and speeds of WLAN. After exploring the technical details of WiMAX, several horizontal handovers (handovers within the same air technology) were introduced: Hard Handover, Soft Handover, Fast Base Station Switching and Address Resolution Protocol.

Besides this promising technology, another, very popular technology was considered; WLAN. WLAN is applied in lots and lots of devices. A combination of WiMAX and WLAN would be ideal in terms of combining the strong aspects of both technologies. That is why vertical handovers between these two technologies were also researched. Here, vertical handover (handover between different technologies) solutions are: Mobile IP, IP Multicast, Media Independent Handover (MIH), Fast Mobile IPv6 Handover (FMIPv6), Hierarchical Mobile IPv6 Mobility Management (HMIPv6) and Session Initiation Protocol (SIP).
All these solutions were reviewed based on a number of criteria. Those criteria are: handover latency, handover type, complexity, dropped packets, duplication of events and scalability.

As a result of this comparison, several conclusions can be drawn. For horizontal handover, which is less complex compared to its vertical antithesis, it is possible to identify one best solution based on the criteria. In this paper, the ARP Handover has the best overall score. In case of vertical handover, this is not so easy. There are more issues to consider and different solutions can have several advantages. Independently, each solution proposes improvements. This is why there can not be concluded in favour of just one solution. The solution which provides the best handover performance is a combination of several. This is backed up by research papers like [HPF07], [SG+05], [LKU06] and [BMN07]. An example is the MIPv6 combined with HMIPv6 and the fast handover mechanism of FMIPv6.

Further research in this area is, logically, to find out how these solutions can work together in an effective and efficient way. Also, because the solutions place an extra burden on the network entities, it is important to review the efficiency of each combination to justify the extra load on the network. It may be clear that there are some very good-working solutions to provide real seamless, not noticeable handovers.

ACKNOWLEDGEMENTS

Here, I would like to thank dr. ir. G. Karagiannis. As my supervisor he helped me with setting up my research and evaluating the process carefully. With his comments and suggestions I was able to execute the research correctly and completely.

Furthermore, I’d like to thank my review group. They provided useful comments on my work which is very important in order to make an understandable scientific paper.

REFERENCES


Mobility anchor point selection based on user mobility in HMIPv6 integrated with fast handover mechanism, E. Natalizio, A. Scicchitano, S. Marano, IEEE Wireless Communications and Networking Conference vol. 3, March 2005, pp. 1434 – 1439


Handover issues in heterogeneous environments: Vertical handover between WiMAX and WiFi networks, Roberto Albanese, presentation 01-03-2006.


Fixed, nomadic, portable and mobile applications for 802.16-2004 and 802.16e WiMAX networks, November 2005, Prepared by Senza Fili Consulting on behalf of the WiMAX Forum.


