How to Improve the Efficiency of IPv6 Handovers in IEEE 802.16 Networks

Tomasz Mrugalski, M.Sc.
tomasz.mrugalski@gmail.com
Gdansk University of Technology
ul. Storczykowa 22B/12
80-177 Gdansk, Poland

Jozef Wozniak, prof.
jowoz@eti.pg.gda.pl
Gdansk University of Technology
ul.Narutowicza 11/12
80-952 Gdansk, Poland

Abstract – **The first generation of fully conformant IEEE 802.16-based networks is being deployed throughout the world. Most of these networks do not support full mobility, due to radio access layer limitations. Newer solutions, based on IEEE 802.16-2005 standard, offer mobility support for subscriber stations. Unfortunately, after quickly changing the point of attachment on the WiMAX layer, very slow and inefficient IPv6 reconfiguration takes place. Delays introduced by DHCPv6 stateless automatic configuration and Mobile IPv6 can easily diminish or even render useless all benefits gained using the efficient WiMAX - data link layer. IPv6 automatic configuration process was not designed with fast reconfiguration in mind. As handover speed is a crucial requirement in mobile cellular environments, reasons behind delays introduced by IPv6 layer mechanisms have to be analyzed and appropriate countermeasures applied. Proposals include novel use of DHCPv6 relays for remote configuration, solving DAD delays, limiting Binding Update procedure in Mobile IPv6, configuring routing through DHCPv6 communication and some other.**

This paper describes all stages of full IPv6 handover in IEEE 802.16 environment, focusing on major reasons of reconfiguration delays. A new metric for assessing impact of every stage on handover efficiency is defined. Several proposed improvements to the IPv6 handover process are evaluated and simulation results are presented. A discussion regarding possible generalization of best improvement proposals and further research areas concludes this paper.

Keywords: mobility, DHCPv6, autoconfiguration, IEEE 802.16, IPv6, WiMAX

I. MOBILITY – WIMAX AND IPV6 PERSPECTIVE

Digital information processing including information transmission, storing and retrieving is becoming more and more popular, being at the same time the main reason of a continuously increasing demand for the amount of data exchanged by users. At the same time, due to miniaturization and advancements in wireless electronics, new services enabling users to perform mobile computing are gaining significant advantage. Wide area wireless data transmission is currently provided mainly by 2G and 3G cellular networks. These networks provide excellent mobility support, but do not address issues of high data rates. The fastest deployed UMTS enhancements – HSDPA and HSUPA – offer peak data rates of 14Mbps and 5,8Mbps per cell, respectively. With Enhanced HSPA, UMTS Release 8 networks will offer peak data rates of 42 Mbps in downlink and 11.5 Mbps in uplink but no such networks have been deployed yet. These capacities are shared by all users operating in a cell. Although such throughput might seem to be enough for some services like web browsing or e-mail, is not adequate for handling applications requiring high data rates, e.g. high resolution video conferences. UMTS technology has some other disadvantages, including complicated architecture and small cell dimensions. The last feature makes UMTS well prepared for providing support in densely populated urban centers. However, in other environments, e.g. town outskirts, small towns or villages, UMTS deployment is often considered to be too expensive. In sparsely populated areas or countries, like Australia, that may prove to be a considerable disadvantage.

In order to solve the aforementioned problems another technology, namely IEEE 802.16, has been developed. An example of a typical 802.16 network infrastructure is presented in Fig.1. The IEEE 802.16 standard, also known under its commercial name WiMAX[1], defines mechanisms which allow Subscriber Stations (SSs) to communicate with Base Stations (BSs). Thanks to the usage of advanced radio access technologies and smart bandwidth management, significant improvements were made in transmission ranges (up to 40-50km) as well as in available throughput values (up to 70Mbps). Lack of mobility support in the initial IEEE 802.16-2004 specification was solved in early 2006, when [1] was published. Numerous mechanisms supporting subscriber mobility were introduced, like Neighbor Advertisement, Scanning and Handover. They will be briefly discussed in Section 3.

After performing data link handover and network reentry in a new location, an IPv6 node, working on the top of the SS stack, is required to make handover in IPv6 and higher layers. After handover every IPv6 node (after power-up, power conservation wakeup or handover) is required to confirm its old or obtain a new address and configuration parameters. That can be arranged using a stateless [2] or stateful [3] automatic configuration (often referred to as autoconfiguration) procedure. Since stateless autoconfiguration does not provide means of configuring any parameters beside those regarding address and routing, it is generally agreed that in any non-trivial network the stateful configuration should be used. See [4] for a detailed discussion regarding this topic. After full configuration is completed, a given node informs its

[1] WiMAX logo can be used by vendors, whose equipment pass specific conformance and interoperability tests.

corresponding nodes[2] (CN) and the home agent about its new location, according to [5]. This procedure concludes the handover and the node becomes fully operational in its new location, regains its full communication capability, and can continue to maintain its communication activities.

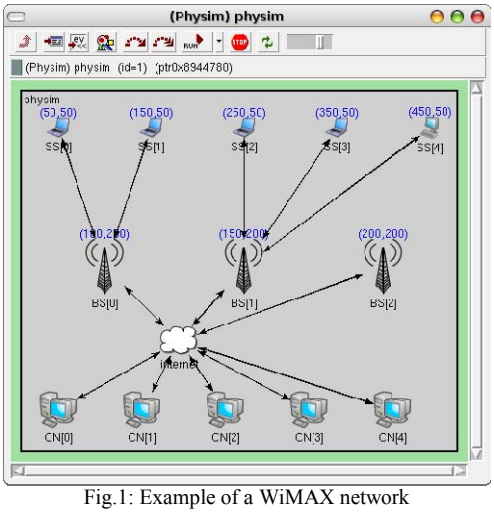

Fig.1: Example of a WiMAX network

This paper describes all stages of full IPv6 handover in IEEE 802.16 environment, focusing on areas causing essential reconfiguration delays. In order to properly evaluate their influence on the overall handover performance a new metric for assessing the impact of every stage is defined. Several novel improvements to the IPv6 handover process are also proposed and evaluated. Simulation results and conclusions are presented in subsequent sections. A discussion, regarding possible generalization of best improvement proposals and further research areas, concludes this paper.

II. PROBLEM STATEMENT

Full scale handover process, described above, is long and complicated. During certain steps, like scanning or IPv6 autoconfiguration, the SS is unable to maintain communication. Since real time data transfer, e.g. videoconferences or voice connections, compose one of major types of applications foreseen for WiMAX, interruptions in data transfer are highly problematic, as they lead to deterioration of service quality. To some lesser extent this disadvantage also affects video on demand streaming. Packet drops and delays in that case can easily be counteracted by data buffering in the SS. However, in turn, interactive multimedia scenarios (e.g. using VoIP) do not allow for extensive data buffering as adding an extra buffering delay becomes a source of service quality degradation.

An approach proposed in this paper consists of two phases. During the first phase, actions or procedures undertaken by network elements are assessed. Procedures that meet following criteria are good candidates for optimizations:

- **Blocking property** – during such an action communication with CNs is not possible;
- **Action necessity** – means that this action cannot be omitted as it is required by the handover process;
- **Action duration** – a considered action introduces a significant delay – time intervals between communication opportunities.

To conveniently assess and compare radically different handover phases (actions), we propose a metric called Handover Delay. It is expressed in milliseconds and specifies how long an IPv6 node does not have full communication capability due to the analyzed method. X is the metric value, while *HD(.)* stands for its symbolic designation.

$$X = HD(\text{action})\ [\text{ms}] \qquad (1)$$

In general, lower scored methods are considered "better", as they introduce shorter delays. If a method (action) allows IPv6 node to communicate immediately, with no handover delay, its HD value is equal to 0ms, so it does not hinder communication in any way and thus requires neither optimization nor improvements.

The second phase of the proposed approach concerns improvements for actions – methods with highest HD metrics. This metric is also used to assess benefits from the proposed improvements. Full handover procedure may also be evaluated in a similar manner. As this paper describes on-going research, improvements are proposed only for selected SS and BS actions.

III. HANDOVER – THE IEEE 802.16 PERSPECTIVE

During normal operation, the SS is associated with one base station - the Serving Base Station (SBS). Because of degraded signal quality, unsatisfactory quality of service or network operator enforcement, the SS may intend to migrate to another base station - the Target Base Station (TBS). The IEEE 802.16-2005 standard provides various mechanisms that make such a handover possible.

A. Neighbor Advertisements and scanning

The Serving Base Station possesses information regarding neighboring base stations. The list of available target base stations is periodically announced to the associated SSes. That list does not provide a complete set of target base stations. SS and TBS might be on the opposite sides of the area covered by an SBS, so communication between them might be impossible. Also one operator will advertise only its own base stations, although there might be other nearby BSs operated by another service provider. As during neighbor advertisements SS maintains full communication capability, HD metric value of this mechanism equals 0ms. Therefore this method does not require optimization.

To verify the current status of all BSs within its range, the SS performs scanning. This procedure includes temporary disassociation with SBS and attempt to receive transmissions from other BSs. This feature allows not associated SSs to receive transmissions and perform signal strength measurement and quality assessment.

The SS cannot maintain communication with its CNs or its BS during the scanning procedure. Duration of the

[2] A peer node with which a mobile node is communicating, as specified in [5].

scanning period is variable as a single scanning period length, number of iterations, as well as interval between consecutive scanning periods are requested by the SS and provided by the BS. Therefore it can adjust scanning procedure parameters to currently supported services. HD metric of this method is difficult to be precisely measured as in theory scanning periods can be arbitrarily long. In practice scanning periods will rarely exceed 20ms. If longer scanning is required, it can be split into several shorter periods. Those facts indicate that scanning does have significant impact on handover delays.

B. Handover

After scanning, the SS has reliable information about possible handover targets. It evaluates all candidates and chooses one, i.e. the TBS, as its new network attachment point. The SS announces its decision to its SBS by requesting handover and after receiving confirmation is ready to detach. Optionally, BS can notify the Target Base Station via backbone network, so necessary preparation can be arranged in advance. The exact moment of detachment is chosen by the SS. SS maintains full communication capability until the HO–IND (Handover Indication) message is sent. therefore this step does not require any optimization.

C. Network reentry

After leaving its SBS, the SS adjusts its radio to receive transmissions from a TBS and performs network reentry at the new location. Normal network entry procedure consists of several steps. *Anonymous* and *handover ranging* are intended for radio tuning, and connection identifiers update. *Negotiate Basic Capabilities* is performed to find the least common denominator between BS and SS capabilities, like ARQ support and list of supported modulations. Optional cryptographic protection is achieved via *Privacy Key Management*. *Registration* performed afterwards specifies additional information about IP protocol version used, vendor specific information etc. The final step is to create connections by using *Dynamic Service Flow* procedure. After successful registration the subscriber is logged into the network but is able to communicate only with its base station for control purposes. To send and receive data traffic, service flows must be created. Since service flows are unidirectional, at least one uplink and one downlink flow must be created.

Fortunately the IEEE 802.16 standard provides numerous optimizations in this procedure, designed with mobility in mind. Therefore, the full procedure as described above is usually performed only once, during the first network entry. Further reentries are less time consuming. Exact duration of reentry process is considered one of the more important parameters of a Base Station and is highly dependant on the amount of information about a given SS that this BS has before the actual reentry takes place. Conservative estimation shows that HD metric value can reach, in some cases, even several hundreds of milliseconds. However, there are also solutions with HD less than 100ms. Further optimizations require detailed analysis of the medium access mechanisms and are outside of the scope of this paper.

IV. DELAYS IN THE IPV6 LAYER

When the SS's data link layer changes the point of attachment, upper layers must also be reconfigured. After moving to its new location, the IPv6 node must obtain a new address and configure routing parameters. Stateless or stateful autoconfiguration is used to obtain those new parameters. In IP protocol, the address serves two goals: equipment identity (nodes are identified by their addresses) and user location (an address determines the node's location). Therefore a mobile node is required to inform its CNs about a new location. As important parts of this reconfiguration process were not designed with mobility in mind, they introduce significant delays.

A. Router Discovery

Routing in IPv6 networks is configured using stateless autoconfiguration mechanism called Router Advertisements [2]. Router periodically announces advertisements describing what prefixes are available directly on the link and what routes are reachable via router. Those advertisements might also be requested by nodes, which are not willing to wait for the next unsolicited announcement. Although this procedure can be executed quickly, its main disadvantage is that no other configuration parameters, except routing, might be obtained by those means. This includes lack of basic parameters like DNS server addresses, VoIP configuration and other. Necessity to transmit extra messages and wait for responses introduces additional delay every time the subscriber moves to a new location. Its duration depends on how fast the base station is able to grant requested bandwidth and how fast the router is able to send responses. HD metric is estimated to be within a 40ms range.

B. DHCPv6

To provide remaining configuration parameters, stateful configuration is used. It can be achieved by using Dynamic Host Configuration Protocol for IPv6 [3]. Although more complex than stateless autoconfiguration, it offers far better functionality: conveys numerous additional configuration options, maintains control over address assignment, provides authentication, etc.

Initial stateful automatic configuration is divided into two phases. The first one is server discovery, during which a client sends a message containing information, how many and what kind of configuration parameters is it interested in. As DHCPv6 protocol offers server redundancy, all available servers respond with advertisements containing their proposed addresses and configuration parameters. As specified in [3], client waits 1 second to allow all servers to generate and send answers. The second phase is the actual configuration. The node chooses one of the available servers and requests configuration, which is granted by the server. Clearly, HD metric of the basic DHCP configuration is

over 1000ms, so this is a major area for possible improvement.

C. Duplicate Address Detection

After new IPv6 address is obtained, according to [2] node is required to verify if this new address is not used. This procedure is mandatory for all IPv6 nodes starting to use new address, regardless of whether they were obtained through stateful or stateless autoconfiguration. When handover (considered a network interface reinitialization) is completed, a node must initiate DAD procedure, which introduces another 1000ms delay. Again, HD metric is over 1000ms.

D. Location update in Mobile IPv6

After successfully obtained and verified addresses, a node must inform its home agent and CNs about its new location. This procedure is also known as *Binding Update*. Although this procedure consists of exchange of only two messages, they are not exchanged locally. Assuming that a message processing time is very small (thus can be neglected), this procedure takes full round trip time from the current mobile node location to its CN. This might be the longest step in the whole handover process, described in the last two sections.

V. AREAS OF IMPROVEMENT

As discussed in the previous sections, there are several mechanisms that introduce significant delays to the handover procedure. To mitigate or, in some cases, even completely eliminate such delays, a number of new improvements are proposed.

A. Remote DHCPv6

During a normal handover procedure, the data link layer (i.e. IEEE 802.16) initiates and performs the handover procedure. After it is completed, the network layer (i.e. IPv6) handover is performed. By doing this in a sequential order, delays introduced by each layer are adding up, resulting in a large overall delay. To avoid this, data gathered by IEEE 802.16 may be used to exercise some preparatory steps before actual switching takes place. Although dealing with horizontal handover (between two locations with the same network access type), 802.21 Media Independent Handover framework proposed by [9] may be used for L2 handover notifications. A subscriber knows its target location, before actual handover occurs. This prior knowledge may be exploited to initialize a connection with a DHCPv6 server, located within the destination network. As all BSs are connected to a common network (e.g., Internet or an ISP's network), it is possible to perform a connection between base stations using a backbone network.

To initiate and maintain such communication, existing DHCPv6 relays may be used, albeit in a modified form. In a classical configuration, relays work as intermediaries between clients and servers. From the client's perspective, direct communication with a server or via relays is indistinguishable. Relays act as representatives of the server. From the server's perspective, a client is connected to the remote link. By modifying relay's behavior, it is possible to use them to forward data from a client to the server and vice versa. In this proposed scenario, a client is aware of the relays. It sends messages to relays and expects them to be forwarded to the specific remote server. Thus relays act as representatives of clients. From the server's perspective, a client is connected directly to the local link. To achieve such operation, clients, relays and servers must support this new mode. As this modification causes DHCP configuration to be performed while still maintaining fully connectivity, it effectively cancels any negative impact on the handover delays. Thus HD metric value of this new improved DHCP configuration is zero.

B. DAD elimination

Every IPv6 node after receiving a new address must perform Duplicate Address Detection procedure. It is intended to detect possible duplicates, i.e. if there are other nodes that use this recently acquired address. According to [2], a node is supposed to wait 1000ms for an unlikely response. Also, as this is the first IPv6 message to be sent from an interface after reinitialization, the node should also delay the transmission by a random delay between 0 and 1000ms.

In a real life environment address duplicates are extremely rare and usually a sign of a severe network misconfiguration or a malicious attack. In the second case – of a malicious attack – to spoof duplicate address, attacker must have already penetrated the network. After link has been compromised, attacker can spoof numerous error conditions and DAD procedure will not prevent him from doing so.

Therefore one proposal to limit the delays is to omit the DAD procedure completely. Such radical proposal lowers HD metric of the DAD phase to zero.

C. Server side DAD

In some cases skipping DAD procedure completely may not be the best course of action. To expedite automatic configuration process, server may maintain a small pool of IPv6 addresses that are checked against duplication. After boot, server selects several addresses and performs DAD procedure for each address added to this pool. After successful validation, address is ready to be leased by clients. It is essential for a server to passively participate in the DAD procedure for those addresses, i.e. to answer for possible DAD messages, sent by other nodes trying to use such address. Before server sends information about particular lease to the client, assigned address should be removed from the server's interface (i.e. server should stop responding to DAD messages sent to this specific address).

To make sure that client supports this enhancement, it should send special suboption to the server in the IA_NA option. It will inform the server that this particular client supports "server side DAD" and it is possible to grant pre-validated address.

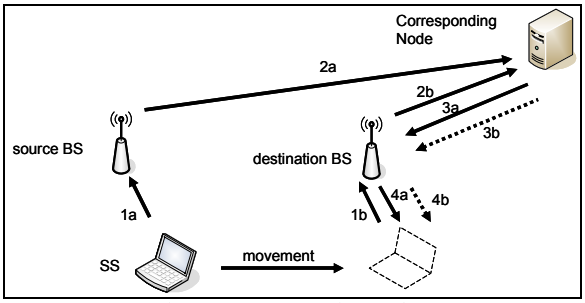

Fig. 2: Remote Location Update

D. Routing configuration

During normal configuration, IPv6 nodes are expected to wait for (or force router to transmit) router advertisements to configure routing. From the mobile node perspective, that requirement is unfortunate as it introduces additional delays. To solve this problem, authors propose a new method for routing configuration. As clients have to exercise message exchange with DHCPv6 server, it may also provide extra information that allows clients to configure routing. To convey this information, a new DHCPv6 option – address parameters – has been proposed and support for this option has been implemented. See [6] for details.

E. Remote Location Update

Usage of the already described Remote DHCPv6 configuration provides access to a whole new class of possible solutions. One of them is a proposed improvement in Location Update, a crucial procedure in Mobile IPv6 protocol [5]. After configuration of a new address the mobile node sends necessary information to all CNs and the home agent. Corresponding nodes and the home agent update their location tables and send confirmation. This procedure often causes significant delays, as it requires quite a long round trip time to complete. Depending on location of notified nodes, that may be several seconds.

Since a mobile node knows its new address before the actual handover takes place, it may send notification to its home agent and CNs before commencing actual handover. If calculated properly, a node will complete reconfiguration at destination location exactly between sending notification and receiving update, thus reducing HD metric to a half of previous value. Unfortunately, it is not possible to reliably measure round trip time as it fluctuates frequently. Also, since handover is imminent, it is fair to assume that transmission conditions for one of the BSs are poor and it is likely that transmitted message will be lost. To avoid a situation, when a mobile node moves to a new location and waits for a message from the corresponding node that may never arrive (due to the lost notification sent by the node, just before handover), it may retransmit update after completing movement at the destination location. The only drawback is a possible duplicate update, which can be safely ignored. This scenario has been presented in Fig. 2.

VI. VALIDATION

There are several ways to assess whether proposed modifications are really improving the handover. First one is to construct a theoretical model that will describe the analyzed scenarios. However it may prove to be extremely difficult to develop an analytical model reliably describing such a complex environment as an IEEE 802.16 network with multiple entities. Even worse, the IPv6 layer provides an additional level of complexity. Therefore this approach appears to be hardly feasible. The second way to assess usefulness of modifications is to develop a simulation environment that will emulate all affected processes. Although complicated, this task is feasible. In general, simulation results can be accepted to prove usefulness of new solutions. Therefore this approach has been selected as a primary validation method. To further reinforce our claims, another verification method has also been used., namely some parts of the proposed improvements were included in a real DHCPv6 implementation.

A. Simulation environment

Currently there are no applicable solutions available, so a new one was implemented for research purposes. To encourage open discussions and contributions to this, the authors have released the simulation environment as an open source project Numbat. The source code is available at the project website [7]. It is a simulation environment that provides implementation of a mobile IEEE 802.16 environment with IPv6 protocols and mechanisms – IPv6, DAD, DHCPv6, Mobile IPv6 and others. For efficiency reasons, parallel, visual and text versions are provided. For detailed description of the Numbat environment see [8] and a project website [7].

B. Simulation results

Ten different scenarios were prepared. First five scenarios contained optimizations provided by existing standards, while scenarios six to ten provided proposed improvements. The following scenarios were simulated: 1. No optimization at all. 2. Maximum optimization provided by 802.16 networks, 3. Skip initial delay in DHCPv6, 4. Use DHCPv6 preference set to 255, 5. Use rapid-commit DHCPv6 option, 6. Skip DAD procedure, 7. Perform server side DAD, 8. Perform remote DHCPv6 autoconfiguration, 9. use address parameters in DHCPv6, 10. Use Remote Location Update. All scenarios contained improvements introduced in previous scenarios plus some extra feature (except scenario 6, which is mutually exclusive with 7).

Numbat simulation environment provides several mobility models. As this research is focused on the handover process itself, rather than decision if and when initiate handover, handover performed after specified timeout was selected. Numerous parameters were observed, like the number of packets transmitted and received by a mobile SS and its corresponding node, average packet delay, number of packets dropped by subscriber (due to lack of communication capability), received bits per second by SS, handover preparation time, 802.16 network reentry time, DHCPv6 configuration time, IPv6 reconfigure time and lack of communication capability periods.

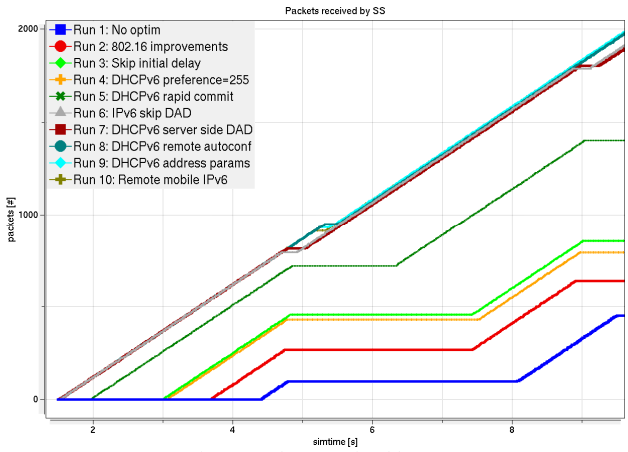

Fig. 3: Packets received by SS

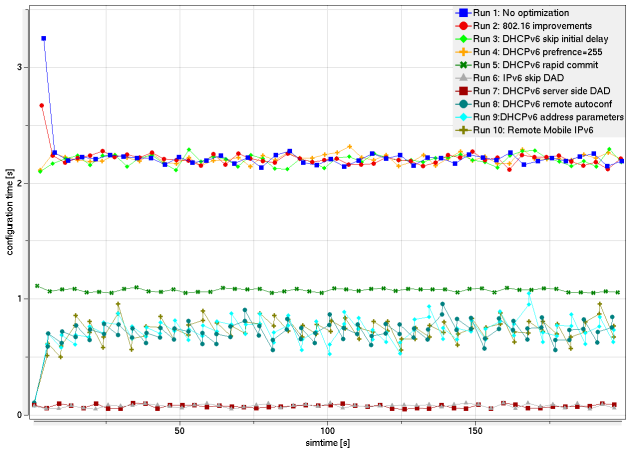

Fig. 4: DHCPv6 configuration time.

It is clear that with enlarging the set of improvements, number of received IPv6 packets increases. That may be considered as a first indication that the proposed mechanisms behave as expected. When handover delay is limited, network is able to deliver more packets in the same time. Handover occurrences can be clearly identified as horizontal lines in the diagram. This relation, for all analyzed modifications has been presented in Fig. 3.

It is also worth noting that the number of bytes, received per second by the SS, has significantly increased. Intervals of transmission inactivity on Run 1 are clearly visible. For the mobility model used, the last scenario not only allows transmitting more data in a continuous manner, but also makes possible to complete a greater number of handovers. (In our mobility model, a next handover is performed 4 seconds after the previous L2 handover was completed.)

As an intermediate value, DHCPv6 configuration time (i.e. time required to complete DHCPv6 message exchange) was also measured (see Fig. 4). In general, with more improvements, this time decreases. There's an exception, though. All scenarios that use remote autoconfiguration last longer to complete. That is understandable, as messages have to be exchanged between BSs. It is also crucial to understand that during remote autoconfiguration, subscriber station maintains full communication capability, thus HD metric is zero.

VII. CONCLUSIONS AND FUTURE RESEARCH

The current IEEE 802.16-2005 standard, covering two lower layers in the ISO/OSI protocol stack, offers quite good mobility support and there are no significant areas of necessary improvements. However, from the mobility perspective, IPv6 standard does not support real-time mobility. The IPv6 protocols (like DAD or DHCPv6) introduce significant, and often unnecessary, delays. To overcome at least some of these delays, several new methods were proposed. All proposals were positively validated and verified, using the Numbat simulation environment. It is worth noting that the proposed HD metric allows estimating essential delays introduced by different actions during handover. The next step to ultimately confirm usefulness of discussed proposals is to implement them in a real DHCPv6 environment and to validate it in a controlled network. Some features, e.g. routing configuration via DHCPv6, was implemented already as part of the Dibbler project [6]. As this appears to be a very reasonable solution, work is in progress to specify an RFC draft and to submit it to IETF, as an independent proposal.

ACKNOWLEDGMENT

This work was supported in part by the Polish National Center for Research and Development under the PBZ grant MNiSW –02/II/2007.

REFERENCES

[1] IEEE working group, "*IEEE 802.16-2004: IEEE Std. for Local and Metropolitan Area Networks – Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems*", IEEE, Dec.2005

[2] S. Thomson, and T. Narten "*IPv6 Stateless Address Autoconfiguration*", RFC2462, IETF, Dec.1998

[3] R. Droms, Ed. "*Dynamic Host Configuration Protocol for IPv6 (DHCPv6)*", RFC3315, IETF, Jul.2003

[4] J. Wozniak, K. Nowicki, T. Mrugalski, "*Mobile users issues, in micro and macro scale, in IP networks*", SIS2004, Lodz, Sep.2004

[5] D. Johnson, C. Perkins, J.Arkko, "Mobility support in IPv6", RFC3775, IETF, June 2004

[6] T. Mrugalski, "Dibbler User's Guide", http://klub.com.pl/dhcpv6/, retrieved July 2008

[7] T. Mrugalski "Numbat – project website", http://klub.com.pl/numbat/, retrieved July 2008

[8] J. Wozniak, T. Mrugalski "Numbat – extensible simulation environment for mobile, IPv6 capable IEEE 802.16 stations", ATNAC'2007, Christchurch, NZ, December 2007

[9] IEEE 802.21 working group, http://www.ieee802.org/21/