Moving towards Seamless Mobility: State of the Art and Emerging Aspects in Standardization Bodies

Marc Emmelmann[*1], *Sven Wiethoelter*[*2], *Andreas Koepsel*[*3], *Cornelia Kappler*[**4], *Adam Wolisz*[*5]

[*]Telecommunication Networks Group
Technical University Berlin, Sekr. FT5–2, Einsteinufer 25, 10587 Berlin, Germany
[1]emmelmann@ieee.org, [2]wiethoel@tkn.tu-berlin.de, [3]koepsel@tkn.tu-berlin.de, [5]awo@ieee.org

[**]Siemens
Communications, Siemensdamm 62, 13623 Berlin, Germany
[4]cornelia.kappler@siemens.com

Abstract

The challenge to provide seamless mobility in the near future emerges key aspects in various standardization bodies. This includes firstly the support of seamless homogeneous handovers. Distinct technologies—such as IEEE 802.11WLANs and 802.16WiMax—have recently amended such support to existing standards. Contrary, cellular networks already included this inherently from their design perspective. Since considerable effort has been made towards coupling of different radio access technologies, the second key aspect includes seamless heterogeneous handovers. IEEE, IETF, as well as 3GPP consider different approaches towards architectures and protocols enabling seamless mobility management. In this work, we discuss recent and on-going standardization activities within IEEE, IETF, and 3GPP towards seamless homogeneous as well as heterogeneous mobility support.

Keywords

Seamless Handover, IEEE 802.11 WLAN, 802.16 WiMAX, IETF, FHMIP, NETLMM, 3GPP.

1. Introduction

Wireless access technologies as well as the number of mobile devices have been continuously growing over the last decades. The importance of mobility support continuously shifts away from mere nomadic networking towards mobile networking. The latter enables users to maintain their application session while moving within a single or among several access technologies.

Even though mobile networking is possible today, it cannot provide service continuity completely. For this seamless mobility, an ongoing application session has to be maintained continuously such that an acceptable quality of service (QoS) perceived by a user is sustained.

As seamless mobile networking is further and further evaluated by its ability to support QoS-sensitive applications, i.e. voice or video conferencing, ongoing standardization efforts focus on three mobility aspects:

1. providing seamless handover for homogeneous technologies,
2. providing seamless handover among different access technologies, and
3. integrating different access networks and technologies under a common IP backbone.

This article provides an overview on most recent and ongoing standardization efforts enabling seamless mobility in both, homogeneous and heterogeneous environments. It will be structured according to standardization efforts: Current work done by IEEE for homogeneous technologies like WLAN 802.11 and WiMax 802.16 is presented in Section 2, while Section 3 summarizes ongoing work in the IETF for seamless services functionality. An interface between the latter two lower and upper layers is presented in Section 4 focusing on media independent handover primitives as specified by IEEE 802.21. Finally, Section 5 lays out mobility support in 3GPP and highlights goals for the Long Term Evolution (LTE) work done. A view on building future 3G networks with IP based mobility schemes concludes the paper.

2. IEEE

2.1. 802.11 WLAN

For 802.11 devices, mobility is only supported in infrastructure mode in which several stations (STA) and an access point (AP) form a basic service set (BSS). In order to enlarge wireless coverage area, a distribution system (DS) may connect several BSSs forming an extended service set (ESS). Moving from one AP's coverage into another's implies detecting the loss or degradation of the current connection, determining an AP to roam to, and establishing a new layer-2 connection with the new AP, i.e. authentication and association. As these steps may last several seconds [1] means to provide seamless mobility support were amended to the standard.

Even though algorithms on how to detect the loss or degradation of an ongoing connection while moving are not standardized, vendors may profit from 802.11k which amends radio resource measurement schemes. The introduced measurement pilot frame, a compact management frame periodically transmitted by an AP at a relatively small interval as compared to the beacon, provides a minimal set of information including its employed transmission power and noise floor at the AP. In combination with the SNR experienced at the receiver, it allows a link margin calculation suitable for transition decisions. Additionally, 802.11k allows the STA's management plane to automatically trigger reports at the MAC, e.g. if the received channel power falls below a certain threshold, as well as to exchange location configuration information both enabling link status- or position-based handover decisions. [2]

The most time-consuming phase during handoff, scanning [1], is significantly reduced by the above mentioned pilot frame and neighborhood information reports. The former's small transmission interval reduces the time spent by a STA on each channel during passive scanning. The latter contains information on validated neighbor APs that are members of ESS and allows scanning on selected frequencies only or even avoids scanning at all. It should be noted that the amendment does not specify means on how to generate that list but reveals one possible approach: a STA scans for APs, builds a local neighbor report, and exchanges it with the AP. [2]

The 802.11r fast BSS transition amendment suggests to employ 802.11k schemes to reduce scanning times and rather optimizes the number of exchanges required to establish an authentication between the STA and new AP. Instead of conducting an authentication ”over the air” as in legacy 802.11, a remote request broker (RBB) is introduced at each AP. Instead of addressing the target AP, the STA directs its authentication request to the RBB which in turn encapsulates and forwards them to the target AP's RBB ”via the DS”. The latter interacts with the new AP's STA management entity to establish authentication. Besides, a STA may request resources at the new AP via the DS using the RBB. This allows the MT to uphold an active communication channel via the old AP and to decide on the AP to switch to according to a successful resource reservation a priori the roam. In advance, 802.11r introduces optimized message exchanges establishing security by key forwarding and distribution which is not covered due to space limitations in this overview. [3]

After the handover, the old AP might still have packets addressed to the MT in its buffer. IEEE 802.11F [4] provided a recommended practice for an inter access point protocol which allowed the new AP to trigger the old AP forcing the latter to forward these packets.[1] Additionally, 802.11r provides a deauthentication via the DS to release resources at the old AP. [3]

2.2. 802.16 WiMax

IEEE 802.16 networks provide centralized broadband wireless access. The BS controls the (mobile) subscriber stations (M)SS employing a combination of time division multiple access (TDMA) and demand assigned multiple access (DAMA). The downstream can be based upon continuous time division multiplexing (TDM) or slotted, TDMA-like bursts. In addition to the handover phases discussed for IEEE 802.11, due to the strictly timed WiMax media access scheme, SSs have to synchronize themselves to the BS and have to adjust the employed transmission power (ranging process). 802.16e amends a mobility support already optimized in terms of reduced handover delays.

To detect the need for handover, BSs may mandate SSs to continuously monitor the carrier-noise-interference-ration (CNIR) and report its mean / std-derivation via a prioritized fast feedback channel. This information may serve as an input for handover algorithms which are not standardized.

To establish a knowledge on neighboring BSs, SSs may periodically scan for neighbor BSs. Therefore, the SS may request a time interval reserved for scanning from its serving BS which in turn may specify, in terms of time interval and metric, how the SS should report the scanning result back to the BS. Apart from SS-initiated scan, the reservation of scan intervals may be transmitted unsolicited by the BS. Based on the feedback from the SS, the BS builds a neighborhood list which is periodically broadcasted. As the latter includes for each neighbor BS information regarding up- and downlink-channel slot assignments, BS id, and PHY synchronization field, these parameters have not to be obtained while switching from one BS to another reducing handover latency.

In order to establish a link layer connectivity with the new BS, the SS has to convey information like its MAC address and capability information to the target BS as well as going through the ranging process. In the traditional way, the SS ”associates without coordination”, i.e. it exchanges these information directly with the target BS over the wireless link. In the second mode, the serving BS coordinates the association by forwarding these information to the target BS via the backbone allowing the SS to immediately start the ranging process in order to adjust its transmission power correctly. The third approach is ”with network assistance”. Additionally to mode two, target and serving BSs exchange the feedback of the ranging algorithm over the backbone and the serving BS provides a single, condensed answer to the SS. This scheme allows the SS to maintain multiple associations at a time reducing the duration of the handover process.

This exchange of security keys between BS and SS is also shifted into the scanning phase a prior the handover. Even a direct communication between serving and target BS is foreseen neglecting the need for authorization via the wireless link during the handoff.

In order to provide a seamless handover support even for higher OSI-layers, 802.16e also amends optional support for fast handover, namely macro diversity handover and fast BS transition. Both cases require strict time synchronization of involved BSs including exchange of MAC state information as well as their operation on the same frequency. For macro diversity handover, involved BSs synchronously transmit downlink data such that diversity combining can be performed by the SS. For the uplink, traffic is received by all involved BSs such that selection diversity can be performed. The information on the up- and downlink slot assignment may be either conveyed by all BSs forming the diversity set or only by a single BS, the so-called anchor. For the fast BS transition approach, only a single BS anchor provides up- and downlink capacity. The continuous monitoring of the BSs' signal levels allows adding and dropping BSs as well as the decision on when to switch to a new anchor. [5]

3. IETF

Handling mobility in IP based networks remains a challenging task. Using IP addresses for routing purposes as well as node identification leads to a problem known as *IP semantic overloading*. Transport protocols use IP addresses for identifying transport endpoints thus changing a node's layer-3 network attachment point (NAP) breaks already established transport connections. MobileIP [6] serves as the standard mobility mechanism for IP based networks and tackles the semantic overload-

[1] 802.11F has expired. Its withdrawal has been voted on by IEEE 802.11 Working Group in November 2005 and was approved by the IEEE SA in March 2006.

ing problem by assigning a node two addresses: one is used for routing purposes (Care-of-Address) and one for node identification (Home-Address). However, other approaches have been proposed: the Host Identity Protocol (HIP) introduces a new, separated namespace of host identifiers (more precisely: IP stack identifiers) for transport endpoint naming. This decoupling of the transport from the network layer allows the latter to evolve independently. Making transport protocols aware of IP address changes is an approach followed by the Stream Control Transport Protocol (SCTP). Here, the transport protocol provides functionality to handle multiple IP addresses and their changes concurrently.

MobileIP, HIP, and SCTP define frameworks for pure end-to-end mobility: two nodes may maintain a connection while moving without relying on further support of the underlying network infrastructure. Nevertheless, the provided mobility is not always seamless if, e.g., two nodes change their NAP concurrently (double jump problem). In this case the interlocutors have to contact their home-agent (MIP) a.k. rendezvous servers (HIP) to re-establish a mutual view on the connection endpoints. While node mobility is typically network controlled in cellular networks (e.g. 3GPP), IP based mobility schemes place their mobility management on the end nodes. Even though the latter approach reduces network complexity, pure end-to-end mobility schemes suffer from high transmission latencies when exchanging mobility related signaling between the two communicating nodes. Therefore, adding mobility support functions to IP based networks have become more attractive in order to decrease or avoid end-to-end signaling. These functions can be grouped into four categories: Management of handover domains, multihoming, rerouting and context transfer, as well as network attachment.

Management of handover domains comprises access network based mechanisms that hide node movement from the end-to-end mobility protocol within limited mobility zones. *Multihoming* tackles issues regarding multi-interface operation and application- and flow-mobility. *Rerouting and Context Transfer* deals with problems related to redirecting misrouted packets already in transit and covers mechanisms for moving IP path assigned state, e.g. maintaining a quality of service context. Finally, *network attachment* covers problems related to detecting network attachment and access to handover decision information.

3.1. Handover Domains

Hiding a node's movement within an access network thus reducing end-to-end mobility signaling is an approach typically followed by layer-2 schemes like IEEE 802.11 or IEEE 802.16. Besides the proposed layer-2 protocols, there has also been work done by the IETF in the area of local area mobility schemes. Work on a new micro mobility protocol has started in the NETLMM working group: an access network provides means to maintain a node's IP address even when crossing subnet boundaries.

A number of proposals for realizing a NETLMM based mobility scheme have been made [7, 8]. However, the NETLMM approach still raises a set of open issues. Handover domains may reduce end-to-end signaling significantly and are a prerequisite for realizing seamless handover services. The impact of layer-3 based schemes on layer-2 based micro mobility schemes and their mutual relationship are still unclear.

3.2. Multihoming

With the deployment of a multitude of heterogeneous wireless access systems the initial assumptions of providing IP based mobility for mobile devices have been altered. As an example, consider the principle of using terminal based mobility: a mobile node always uses a single wireless access link for all its communication needs. With multiple access links a node may spread its communication paths over the wireless links that match application requirements best. Thus terminal based mobility is complemented by application/flow-based mobility. The MONAMI6 (*Mo*bile *N*odes *A*nd *M*ultiple *I*nterfaces in IPv*6*) working group was formed recently. It tackles the issues of interface selection, concurrent use of multiple CoAs, simultaneous location in Home and Foreign Networks, as well as flow redirection.

3.3. Rerouting and Context Transfer

When a node changes its layer-3 network attachment point (parts of) the existing IP path between the two nodes may be altered. This affects state currently assigned to the communication path, e.g., in terms of quality of service. The IETF SEAMOBY working group, though inactive now, has released specifications for transferring context state among network nodes during a handover. Work on an IP paging and alerting protocol has been canceled. For MobileIPv6 a number of proposals exist for reducing the impact of rerouting, e.g., in terms of packet loss: Hierarchical MobileIPv6 [9], Fast MobileIPv6 [10], and their recently proposed combination known as Fast Hierarchical MobileIPv6 [11] reduce packet losses and shorten times required for redirecting the end-to-end IP path.

3.4. Network Attachment

Characteristics of the wireless channel pose special challenges on mobile nodes. Wireless link properties differ significantly from wired systems: boundaries of wireless access links are not strictly defined, nodes may float among different wireless links either in time or space due to radio link variations. A change in the underlying wireless link may also affect a node's IP configuration, e.g., when a node's new network attachment point is part of a different IP subnet. The Detecting Network Attachment (DNA) IETF working group develops standards for handling fast network attachments for moving nodes.

Network attachment may rely solely on layer-3 mechanisms, e.g., for IPv6 based networks, the impact of router advertisements and solicitations on link identification is considered by the DNA working group. Using adequate information from lower layers may improve network attachment operation and thus further enhances pure layer-3 schemes. A unified interface between layer-3 and layer-2 is an approach followed by the IEEE 802.21 working group, enabling a technology-independent information exchange.

4. 802.21

The IEEE 802.21 Working Group (WG) is working on media independent handover services, the first draft version [12] was finished in March 2006. The goal is to optimize handovers between heterogeneous access technologies such that on-going services of end users are not terminated, i.e., services can be continued although a handover takes place. 802.21 covers wired as well as wireless technologies—the complete IEEE 802 group as well 3GPP and 3GPP2.

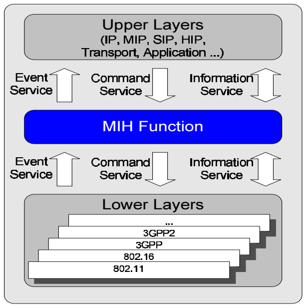

Figure 1: Placement and services of the 802.21 MIH function

802.21 will discover and provide relevant pieces of information for handover decisions to upper layers. This includes signaling of information about QoS support of access networks, network discovery, and network selection. In other words, 802.21 will provide a framework for generic link layer purposes. Handover policies and handover decision entities are thereby out of scope of the 802.21 WG.

For the generic link layer instance, 802.21 introduces a Media Independent Handover (MIH) function between layer-2 and upper layers. The MIH function will define generic SAPs and primitives to higher as well as to lower layers. This may later require an adaptation of technology-specific SAPs of 802.11/16 and of 3GPP/2.

Figure 1 shows the placement of the MIH function and its services. The MIH function is a logical entity which resides on MN as well as on network side. Pieces of information can be exchanged either locally within node's protocol stack by triggers or between MN and an access network entity via MIH-specific messages; for the latter part, 802.21 specifies the MIH Protocol.

The MIH function provides three services: Media Independent Event, Command and Information Service, which are responsible for signaling of state changes at lower layers, coordination and control by higher layers, and information provision about the current and neighbor access networks, respectively. Due to space limitation, we will not describe these services in more detail.

5. 3GPP

3GPP networks so far employ their own mobility solution based on the GPRS Tunneling Protocol (GTP). GTP offers two features that are difficult to achieve with the traditional mobility protocols, e.g. Mobile IP, developed by the IETF: Seamless mobility support and operator-controlled mobility. Both features however are crucial for 3GPP operators.

Mobility in today's 3GPP network works in two modes, connected mode and idle mode. When a mobile terminal is not sending or receiving data, it is in idle mode. In this case it listens to location-specific network broadcasts, and keeps the network informed of its current location area. This way, when a call comes in, the network can page the mobile terminal quickly. When the mobile terminal is engaged in an active session, it is in connected mode. In this case, a tunnel with a specfic QoS is established from the 3GPP Radio Access Network (RAN) up to the gateway of the 3GPP Core Network to external networks. All session data are transported through this tunnel, also known as PDP context. Seamless, operator-controlled mobility is realized by the network deciding, based on measurements of both terminal and base station, when a handover should be performed and then relocating the PDP context accordingly.

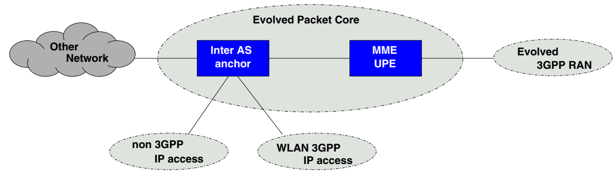

Figure 2: Architecture for an evolved 3GPP network

Currently, 3GPP is developing several possibly fundamental changes to its specifications. Already in 2005, it was standardized how a non-3GPP Radio Access Network, WLAN, could interwork with a 3GPP network. However, a handover between WLAN and 3GPP RAN is not yet possible. At the same time, 3GPP is exploring the feasibility of moving into the general direction of all-IP networks, based on protocols developed by the IETF. Particularly, in the context of SAE (System Architecture Evolution), different options for the evolution of mobility support are being investigated. The goal is to support access to the 3GPP network via multiple non-3GPP access networks, incl. WLAN and WiMAX, and to support handover between these access networks and a 3GPP network.

Fig. 5 shows the current status of the 3GPP architecture debate [13]. An evolved 3GPP RAN is connected to an evolved Packet Core Network. It features an entity handling the user traffic, the User Plane Entity (UPE), and an entity handling the majority of control functions, including intra-3GPP mobility control, the Mobility Management Entity (MME). The Inter AS Anchor controls handover between non-3GPP access systems and 3GPP.

The evolved 3GPP Network will be deployed in parallel to the current 3GPP Network, and seamless handover between them will be specified. For supporting mobility to and from non-3GPP access technologies, several options are currently debated [13] which can roughly be subdivided into two high-level approaches, a more conservative one evolving the current GTP-based approach, and a more progressive one almost completely based on protocols developed by the IETF. The technical details for Connected Mode are still quite open. We therefore describe just current status of the debate on Idle Mode where seamlessness is not an issue.

The *conservative approach* leaves mobility support within 3GPP mostly as-is. To support mobility between non-3GPP access systems and 3GPP, Mobile IP would be employed. Mobile IP is specified to work between the mobile node, i.e. UE, and the home agent, i.e. the Inter AS Anchor. Hence the UE becomes involved with mobility control, which on the one hand takes away control from the network, and furthermore implies existing UEs must be updated. An alternative is the usage of Proxy MIP [14] which is however still in very early draft state. Proxy MIP is an amendment to Mobile IPv6 which allows moving the mobility control from the mobile node to a proxy in the network.

The *progressive approach* would still employ GTP for mobility interworking with the current 3GPP network. Within the evolved 3GPP network and with non-3GPP access networks, however, protocols currently developed by the Network-based Localized Mobility Management (NETLMM) Working Group of the IETF would be utilized [15], which will support network-based mobility control. Furthermore, global and local mobility control shall be separated even more strongly than in Hierarchical MIP [9] by using distinct protocols: an adaptation of Mobile IP (e.g. Proxy MIP) for mobility between "domains", and a distinct, local mobility protocol (to be developed by NETLMM) within a "domain". A single domain can be the evolved RAN of a 3GPP network, or a non-3GPP RAN. However, evolved 3GPP RANs and non-3GPP RANs can also together form a single domain, e.g. when belong to the same operator. It should be noted that there is no IETF protocol for paging which thus is still an open issue.

6. Conclusion

Standardization bodies continuously provide further mechanisms enabling seamless mobility. Each of them focus on their technical application field and have so far not primarily emphasized heterogeneous seamless handover. While IEEE 802.11 and 802.16 have so far focused on layer-2 mechanisms, the IETF inherently provided an technology-agnostic approach. The problem herein is the different understanding of each standardization body of seamlessness: ranging from strictly time-constrained, low-latency handover delays up to enabling nodes to maintain active communication channels on both, IP and transport layer, possibly causing handover latencies in the order of seconds.

Work in the IETF has yielded a number of competing end-to-end mobility schemes which are also under consideration for a future all IP-based 3GPP core (LTE). While new improvements of existing IETF mobility schemes have emerged (FMIP / HMIP) IETF working groups have also seen the necessity for including layer-2 triggers even further reducing the handover latency. This tendency may make use of upcoming, media independent, handover-supporting service primitives, which are under development by IEEE 802.21.

Additionally, the IEEE starts leaving its strictly mobile controlled handover schemes. E.g., the wireless network management working group 802.11v discusses a paradigm shift towards supporting a network directed handover allowing to achieve, e.g., load balancing between APs. [16, 17] Additionally, mechanisms to dynamically adjust individual HO policies at a STA, e.g. by the AP, are under discussion. [18, 19]

This trend in standardization shows that optimal handover performance may not be achieved by a solemnly network controlled handover, i.e. in 3GPP, or mobile controlled handover, i.e. in WiFi. It is an open question how much handover functionality should be either placed in the network or in the mobile terminal. In the future, nodes will face a vast heterogenity of mobility frameworks with access networks ranging from legacy networks with no seamless handover support up to sophisticated networks providing a number of improved seamless handover related mechanisms. This also drives a paradigm change whereas terminal mobility is extended by application- and flow-mobility schemes.

Nevertheless, all standards will only provide mechanisms for mobility support and do not tackle the autonomy of operators: integrating or merging policies to gain and seamlessly maintain network access while on the move is still not covered and is expected to remain a future challenge.

7. References

[1] J. Spilman, "Test methodology for measuring BSS transition time," IEEE 802.11 TGr Fast Base Station Transition, Working Document 802.11-04/0748r1, July 2004.

[2] IEEE Std. 802.11k/D4.0, "Amandment 9 to IEEE 802.11-REVma-D5.2: Radio resource measurement," IEEE, March 2006.

[3] IEEE Std. 802.11r, "Amandment 2 to IEEE 802.11-REVma: Fast BSS transition," IEEE, March 2006.

[4] IEEE Std. 802.11F, "Trial-use recommended practice for multi-vendor access point interoperability via an inter-access point protocol across distribution systems supporting IEEE 802.11," IEEE, June 12 2003.

[5] IEEE Std. 802.16e-2005, "Amendment 2 to IEEE 802.16-2004/Cor1-2005: Physical and medium access control layers for combined fixed and mobile operation in licensed bands," IEEE, February 2006.

[6] D. Johnson, C. Perkins, and J. Arkko, "Mobility support in IPv6," IETF, RFC 3775, June 2004.

[7] I. Akiyoshi and M. Liebsch, "NETLMM protocol: draft-akiyoshi-netlmm-protocol-00.txt," IETF draft, October 2005.

[8] J. Wood and K. Nishida, "Edge mobility protocol (EMP): draft-wood-netlmm-emp-base-00.txt," IETF draft, October 2005.

[9] H. Soliman, C. Castelluccia, K. E. Malki, and L. Bellier, "Hierarchical mobile IPv6 mobility management (HMIPv6)," IETF, RFC 4140, August 2005.

[10] R. Koodli, "Fast handovers for mobile IPv6," IETF, RFC 4068, July 2005.

[11] H. Jung, H. Soliman, and S. Koh, "Fast handover for hierarchical MIPv6 (F-HMIPv6)," IETF, draft, October 2005.

[12] IEEE Std. 802.21/D1.0, "Draft IEEE standard for local and metropolitan area networks: Media independent handover services," IEEE, March 2006.

[13] 3G-Partnership-Project, "3GPP TR 23.882 V0.10.0 (2006-01): 3gpp system architecture evolution: Report on technical options and conclusions (release 7)," 3GPP Technical Specification, January 2006.

[14] S. Gundavelli and K. Leung, "Localized mobility management using proxy mobile IPv6: draft-gundavelli-netlmm-mip6-proxy-00.txt," IETF draft, November 2005.

[15] J. Kempf, K. Leung, P. Roberts, K. Nishida, G. Giaretta, and M. Liebsch, "Problem statement for IP local mobility: draft-kempf-netlmm-nohost-ps-01.txt," IETF draft, January 2006.

[16] V. Anantha and R. Skidmore, "Need for site specific / location specific RF management," IEEE 802.11 TGv Network Management, Working Document 802.11-05/740r0, July 2005.

[17] E. H. Qi, J. Walker, M. Rudolf, J. Kwak, and J. Epstein, "Proposal for load balancing," IEEE 802.11 TGv Network Management, Working Document 802.11-05/1065r2, March 2006.

[18] B. Wang, F. Li, and J. Li, "Solve frequent handover requirement," IEEE 802.11 TGv Network Management, Working Document 802.11-05/1071r3, November 205.

[19] B. Wang, F. Li, J. Li, and T. Weong, "Frequent handover," IEEE 802.11 TGv Network Management, Working Document 802.11-06/1280r2, January 2006.