A Dynamic and Efficient MAP Selection Scheme for Mobile IPv6 Networks

Tarik Taleb, Tasuku Suzuki, Nei Kato, and Yoshiaki Nemoto
Graduate School of Information Sciences - Tohoku University, Japan
Email: taleb@nemoto.ecei.tohoku.ac.jp,{s-task,kato}@it.ecei.tohoku.ac.jp, nemoto@nemoto.ecei.tohoku.ac.jp

***Abstract*— While mobile communication systems will provide certainly more flexibility to end-users, they present complex mobility management issues. To tackle mobility management issues, the concept of Mobility Anchor Points (MAPs) was introduced and its use was proposed within the framework of the Hierarchical Mobile IPv6 (HMIPv6) protocol. However, due to traffic dynamics, the protocol performance remains critically affected by the selection of MAPs.**

This paper proposes a dynamic and efficient mobility management strategy for the selection of the most appropriate MAP with the lightest traffic load. The MAP selection is based on an estimation of MAP load transition using the Exponential Moving Average (EMA) method. The proposed selection scheme is referred to as *Dynamic and Efficient MAP Selection (DEMAPS)*.

The scheme performance is evaluated through simulations. Simulation results show that the DEMAPS scheme substantially reduces the number of packet drops, guarantees shorter service delays, makes better utilization of the network resources, avoids redundant transmissions, and maintains a fair and efficient distribution of the network load.

I. Introduction

As originally specified, the IP protocol does not support mobility for a number of reasons related to the protocol syntax and semantics. To support global mobility in IP networks, the Mobile IP Working Group within the Internet Engineering Task Force (IETF) proposed a packet-based mobility management protocol, called Mobile Internet Protocol (MIP) [1] [2].

While MIP provides basic mobility support and guarantees seamless data delivery, it is not a suitable solution for environments in which Mobile Nodes (MNs) frequently change their points of attachment to the network. Applying MIP to a wide population of users with relatively high mobility features will result in the generation of a large number of binding update requests, all most likely in a single burst [3]. To process such bursts of binding update requests, a massive amount of network bandwidth and computational load is required. Consequently, the binding update cost becomes extremely huge and the system turns to be unscalable to operate. In case of mobile users roaming far away from their home networks, the system performance gets further aggravated and the signaling delay for binding update becomes longer. This yields to the loss of a significant amount of in-flight packets and ultimately affects the overall QoS of the system.

To reduce the number of location update messages to the Home Agents (HAs) and the correspondent signaling delays, IETF has proposed the Hierarchical Mobile IPv6 (HMIPv6) protocol [4]. The key idea behind the HMIPv6 protocol is to separate local mobility from global mobility. HMIPv6 is based on the deployment of a number of local agents called Mobility Anchor Points (MAPs). MAPs can be located at any level in a hierarchical network of routers. Each MAP administrates a set of Access Routers (ARs) forming a single network domain. Within a MAP network, a mobile node is identified by a Regional Care of Address (RCoA), which refers to the MAP, and an on-Link CoA (LCoA), which is a CoA on the AR the mobile node is attached to. When a mobile node enters into a new MAP site (Inter-MAP handoff), it receives a new RCoA and a new LCoA. The mobile node is then requested to specify a binding between its Home Address (HoA) and RCoA. In case the node moves within the MAP (Intra-MAP handoff), it receives only a new LCoA on its new point of attachment while its RCoA remains unchanged. The mobile host then acknowledges the current MAP of its LCoA. By so doing, during Intra-MAP handoffs, transitions are performed locally and no Binding Update (BU) message is sent (on the entire network) to HAs.

Although HMIPv6 reduces the signaling load and provides optimal routing and fast transition performance, it adds the network management entity and complexity: some MAPs may become congested while others are underutilized. To cope with such an issue, a dynamic MAP management strategy is required. This paper proposes an efficient method for the selection of the most appropriate MAP with the lightest traffic load. The MAP selection is based on an estimation of MAP load transition using the Exponential Moving Average (EMA) method. The proposed selection scheme is dubbed *Dynamic and Efficient MAP Selection (DEMAPS)*.

The remainder of this paper is structured as follows. Section II highlights the major research work recently proposed in the area of mobility management. The proposed scheme is described in detail in Section III. Following this, Section IV portrays the simulation environment and reports the simulation results. Concluding remarks are presented in Section V.

II. Related Work

In general, mobility management techniques can be classified into two categories: Micro-mobility and Macro-mobility. In the former, handoffs are handled locally without any involvement of HAs. Notable examples are Cellular IP [5] and Handoff-Aware Wireless Access Internet Infrastructure (HAWAII) [6]. Cellular IP is specifically designed to support handoff for frequently moving hosts. It is applied on

0-7803-9415-1/05/$20.00 (C) 2005 IEEE
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE GLOBECOM 2005 proceedings.

a local level and can interwork with MIP to support wide area mobility, that is, mobility among Cellular IP networks. The HAWAII protocol divides the network into hierarchies based on domains. The functioning of HAWAII hinges on the assumption that users' mobility is local to domains. For each host, the HA and any Correspondent Node (CN) are unaware of the node's mobility within the host domain. Each domain has a gateway, called the domain router, and each host has an IP address and a home domain. In HAWAII, host based forwarding entries are installed in gateways using a set of specialized path setup schemes. These entries help to reduce both the data path disruptions and the number of binding updates. A major credit of micro-mobility management techniques consists in their reduction of handoff signaling delays.

In macro-mobility, when a mobile node roams to a different network area, the node solicits for a new CoA. A BU message is then sent to the HA. The major issue with macro-mobility pertains to the significant handoff signaling delays for users roaming far away from their home networks. These delays disrupt active connections each time a handoff to a new attachment point of the network is performed. The time required for the establishment of a new connection between MNs and their CNs becomes therefore remarkably long and losses of in-flight packets may be significant.

To cope with packet losses that may occur during handoffs due to the broken data path from the source to the destination, a set of mobility management techniques has been proposed in recent literature. They can be classified in turn into two categories: *caching-based* and *smooth handoff* techniques. In the first category, when a handoff occurs, the old AR caches and forwards the packets to the new AR based on a request to forward the packets. Most pioneering examples that use this technique are Fast Handovers Mobile IP [3] [7] and HAWAII [6]. In the second category, packets are routed to multiple nearby ARs around the MN to ensure delivery of the packets to the node. In addition to the recently proposed multi-path smooth handoff scheme [8], multicast mobility support [9] and bicast used in Cellular IP [10] use this technique. A combination of both smooth handoff and buffering techniques is proposed in [11].

In the sphere of attempts to reduce handoff signaling delays in macro-mobility, a large body of prior work was proposed. The central theme in these pioneering studies pertains to the adoption of hierarchical management strategies using local agents. HMIP6 and TeleMIP for Cellular IP [12] are notable examples. Most proposed protocols employ Foreign Agent (FA) hierarchies to localize the binding traffic. While most hierarchical techniques are intended to reduce the binding update traffic by localizing handoff signaling, they cause additional issues related to network traffic management. Effectively, some local agents get congested with update traffic while others are not efficiently utilized. To overcome this deficiency, the choice of network hierarchies should be performed in a dynamic manner. In this regard, [13] proposes a dynamic and distributed domain-based mobility management scheme. In this scheme, a group of ARs forms a domain. A *domain list* indicating the ARs that belong to the same domain is stored at each AR. Mobile nodes residing in a given domain maintains that *domain list*. If a mobile node changes its point of attachment to a new AR within a different domain, the node will then update its *domain list* to that of the new AR and the latter will serve as a MAP for the node. [14] proposes also another dynamic hierarchical mobility management scheme for MIP networks. The scheme is more efficient and is referred to as DHMIP throughout this paper. In DHMIP, when a mobile host connects to a new subnet via a new AR, the new AR notifies the new CoA of the host to the previous AR. The new AR serves then as a new location management hierarchical level for the node. One major drawback of the two schemes is that they both deliver packets to users via multiple levels of ARs, a fact that leads to long packet delivery delay and congestion of the selected ARs with redundant traffic. One possible solution to this issue is to reduce the size of subnet domains. However, this would lead to frequent inter-domain handoffs and consequently excessive binding update cost.

III. Dynamic and Efficient MAP Selection Scheme

Fig. 1 depicts the major steps in the proposed MAP selection method. This operation is conducted as follows:

- Step 1: Every Δ period of time, each AR receives Router Advertisement (RA) messages from high-layer MAPs similarly to HMIPv6 [15]. Unless otherwise specified, Δ is set to $1s$.
- Step 2: Using information included in RA messages and based on a given computational model, each AR selects the optimum MAP for communication. This MAP is referred to as OMAP. Details on the used computational method will be given later in this section.
- Step 3: Upon performing handoff, a mobile node sends a Router Solicitation (RS) message to the new AR [15].
- Step 4: In response to the RS message, the AR notifies the MN of the OMAP selected in Step 2. It should be emphasized that MNs are notified of OMAPs only upon handoffs. This incurs no additional energy consumption (compared to MIPv6) and shall have no effect on the critical battery life of mobile nodes.
- Step 5: Following this, if the selected OMAP is the same as the Previous MAP (PMAP) used prior to handoff by the MN, this latter judges the handoff as an intra-domain movement and sends a BU message to only the OMAP. This aims to minimize the handoff signaling delay and reduce the signaling traffic for users roaming far away from their home networks. In case the OMAP is different from the PMAP (e.g. inter-domain handoff), the MN sends three BU messages to OMAP, HA, and its CN, respectively.
- Step 6: In response to the BU message, the OMAP acknowledges the MN of a successful binding update via a Binding Acknowledgment (BA) message.

Having a potential number of MNs connected to the same MAP for communication may likely lead to congestion of the MAP in question and result in an inefficient distribution

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE GLOBECOM 2005 proceedings.

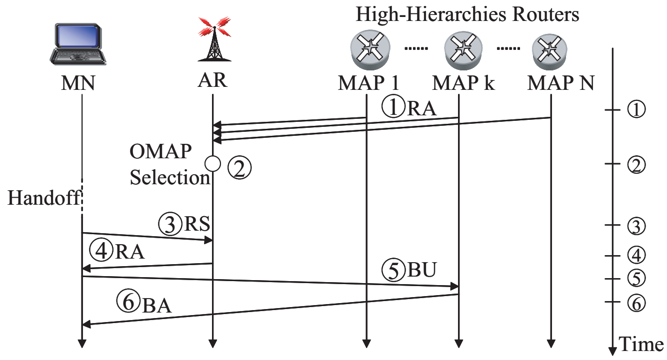


Fig. 1. Major steps in the proposed scheme

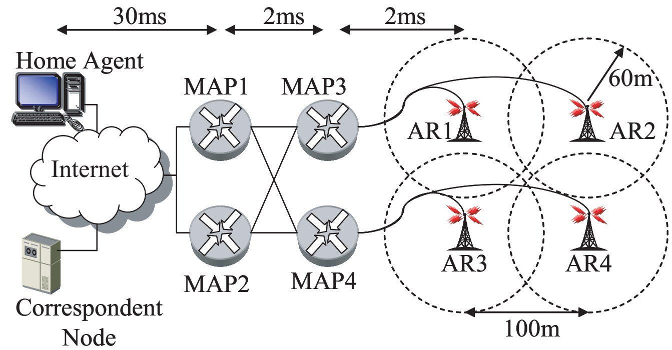


Fig. 2. Simulation environment

of network traffic. To avoid congesting MAPs with traffic, ARs should select the most appropriate MAP with the lightest traffic. ARs should be thus aware of on-going dynamics in network conditions.

To notify ARs of possible changes in network conditions, we use the Exponential Moving Average (EMA) method to predict possible future transitions in the MAP load. The underlying reason beneath the choice of EMA consists in the fact that EMA is a cut-and-dry approach for analyzing and predicting performance, easy to implement, and requires minimal computational load. In the proposed scheme, the traffic load is measured periodically every Δ period of time. Let $M[n]$ and $E[n]$ denote the measured load value and the EMA value of the MAP load at the n^{th} time slot, respectively. By definition, $E[n]$ is expressed as follows:

$$E[n] = \frac{\sum_{i=0}^{\infty}(r^i M[n-i])}{\sum_{i=0}^{\infty} r^i}$$

where r is the exponential smoothing constant ($0 < r < 1$). To give more weight to the latest data, r is set to 0.9 throughout this paper. Considering the fact that ($\sum_{i=0}^{\infty} r^i = \frac{1}{1-r}$), $E[n]$ can be easily computed in a recursive manner as follows:

$$E[n] = (1-r)M[n] + rE[n-1] \quad (1)$$

The key idea behind the proposed method is to use the EMA value to predict the transition tendency of the MAP load. This prediction is based on comparison between the two values $E[n]$ and $M[n]$. In deed, in case ($E[n] < M[n]$), the MAP load has more tendency to increase (Load Increase (LI) tendency), whereas in case ($E[n] > M[n]$), the system load may likely decrease (Load Decrease (LD) tendency). Upon prediction of their load transitions, MAPs notify ARs of this information via the 32-bits option field carried in the packet header of RA messages. In case of LI tendency, MAPs write down the measured load value, $M[n]$, in the RA packet header, whereas in case of LD tendency, the option field is set to null. Based on this information, ARs decide the most appropriate MAPs for future visiting mobile users. High-hierarchies MAPs with LD Tendency are preferably selected as MAPs for communications.

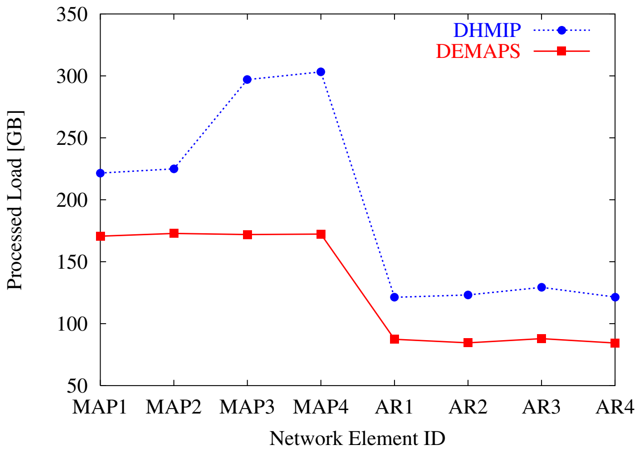
In case of multiple MAPs with LD tendencies, the MAP router at the highest hierarchy is chosen. This aims to create large MAP domains for mobile nodes so as that their future handoffs can be locally handled. This ultimately minimizes the handoff signaling cost. In case all high-hierarchy routers have LI tendencies, ARs select the high-hierarchy MAP router with the minimum traffic load, that is, the lowest value of $M[n]$. This obviously purposes to not further overload the network with signaling traffic.

IV. Performance Evaluation

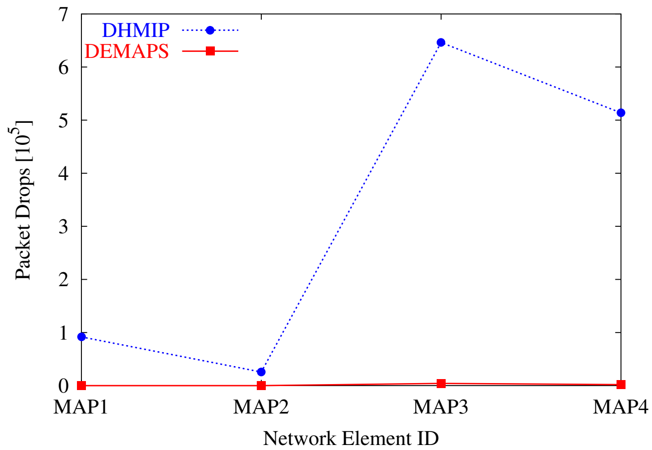
A. Simulation Setup

Having described the details of the proposed scheme, focus is now directed on its performance evaluation through computer simulations. In the performance evaluation, we consider the case of pedestrian mobile users roaming within crowded areas, such as university campuses or Central Business Districts (CBDs). This kind of users are characterized by frequent handoffs at random times in random directions. The mobility pattern of such a population of users can be modeled using the *"Outdoor to Indoor Pedestrian"* model [16]. In this model, upon walking a distance of $5\ meters$, users are assumed to change their moving speed. The users speed follows a normal distribution of an average and a variance value equal to $3km/h$ and $0.3km/h$, respectively. As for the moving directions, the probabilities of users to turn right, turn left, or continue moving straight forward are set to 0.5, 0.25, and 0.25, respectively.

The abstract configuration of the considered network is depicted in Fig. 2. The wireless part of the network consists of four neighboring wireless cells. The coverage radius of each wireless cell is set to 60 meters. The distance between two neighboring ARs is fixed to 100 meters. These parameters are chosen with no specific purpose in mind and do not change any of the fundamental observations about the simulation results. The four ARs are connected to the wired network through a two-layers network made of four MAPs. To form *cross links* among the MAPs, MAPs 1 and 2 are both connected to MAPs 3 and 4. The choice of such a two-layers MAP network with cross links represents a general and simple case [17]. Considering a MAP network with multiple-layers will simply increase the connection RTT and shall have no influence on the overall performance of the proposed scheme. MAP 3



(a) Traffic load processed at each network element.



(b) Number of data packets dropped at high-hierarchies network elements.

Fig. 3. Aggregate performance in terms of transmission efficiency: Avoidance of packet drops and redundant transmissions.

serves ARs 1 and 2, while MAP 4 serves ARs 3 and 4. The MAP network is connected to a HA and a server (CN) via a wired network. All Links are given a capacity of $155Mbps$ (e.g. OC3). The one-way propagation delay over the wired network is set to $30ms$. As for other links, the delay of each is set to $2ms$. Due mostly to its simplicity and its wide usage in today's switches and routers, all routers use Drop-Tail as their packet-discarding policy. All MAPs are assumed to have buffers of $700kB$. A population of 100 nodes is simulated and is randomly scattered over the wireless communication area. Each MN receives UDP packets from CN at a rate of $1.5Mbps$. The UDP packet size is set to $1kB$. All simulations are run for a duration of $30min$, a duration long enough to ensure that the system has reached a consistent behavior. All results are an average of multiple simulation runs.

B. Simulation results

In the performance evaluation, the dynamic hierarchical mobility management scheme (DHMIP) [14] is used as a comparison term. Figs. 3(a) and 3(b) plot the processed traffic load (in bytes) and the number of packet drops at each element in the network. The figures show clearly that the traffic load processed by DHMIP is significantly higher than the processed load in case of DEMAPS, mainly at high-layer MAPs. The difference between the two measures is mostly due to in-flight packets that are redundantly transmitted over same links upon handoff occurrences. Given the limited buffer size of routers, a significant amount of these redundant transmissions is dropped. This justifies the high values of packet drops experienced in DHMIP, as shown in Fig. 3(b). It should be emphasized that the plotted DEMAPS traffic includes both UDP data packets and RA signals. Despite this fact, Fig. 3(a) shows that the overall bandwidth consumption in case of DEMAPS (including signaling packets) is still significantly lower compared to that of DHMIP. This is a strong indication that the additional cost due to signaling packets is minimal. In addition, the obtained performance gains are worthwhile and can be used to advocate the small overhead that may be incurred by RAs.

In the proposed scheme, information on load transition is sent to ARs every Δ time interval. In the simulations conducted so far, Δ was set to $1s$. To investigate the effect of Δ on the DEMAPS performance, we plot the number of packets processed by MAPs for different values of Δ in Fig. 4. The figure demonstrates that setting Δ to higher values results in a poor distribution of network traffic among MAPs: most of the traffic burden is handled by MAP 2. The choice of Δ is a compromise between enhancing the traffic distribution and reducing the frequency of RA messages. In deed, small values of Δ would efficiently distribute the data traffic on the network, whereas large values of Δ would reduce the number of RA signals sent over the communication time.

To illustrate the idea with more clarity, the following index is used:

$$\Phi = 1 - \frac{\sum_{i=1}^{N} |\alpha_i - \bar{\alpha}|}{2\bar{\alpha}(N-1)} \tag{2}$$

where α_i is the number of packets processed by the i^{th} MAP and N is the number of considered MAPs. $\bar{\alpha}$ is the average value of $\{\alpha_i, i = 1 \cdots N\}$. Φ captures the efficiency of traffic distribution over the network and ranges from zero to one. Low values of Φ represent a poor distribution of network traffic and lead to significant packet drops. Fig. 5 graphs the value of Φ for different values of Δ. The figure demonstrates that setting Δ to values larger than $5s$ degrades significantly the traffic distribution over the network. On the other hand, results of Fig. 3(a) show that the system overhead remains minimal when setting Δ to $1s$. It should be emphasized that similar experiments were conducted considering different traffic mobility patterns and identical results were obtained. To conclude, $(\Delta = 1s)$ represents a good tradeoff between an efficient distribution of data traffic and a reduced frequency of RA packets.

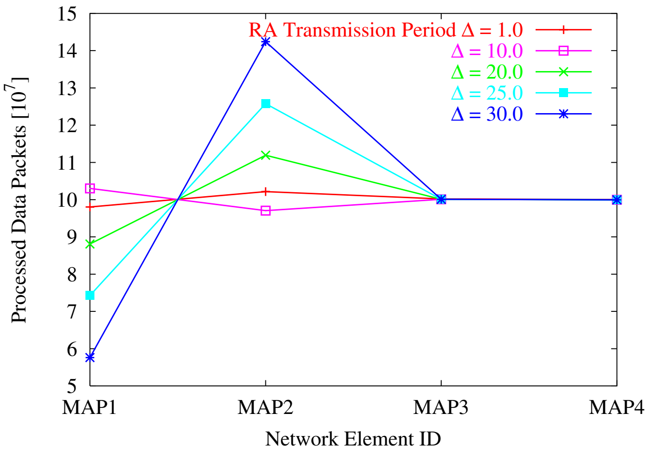


Fig. 4. Number of data packets processed at each MAP for different values of Δ.

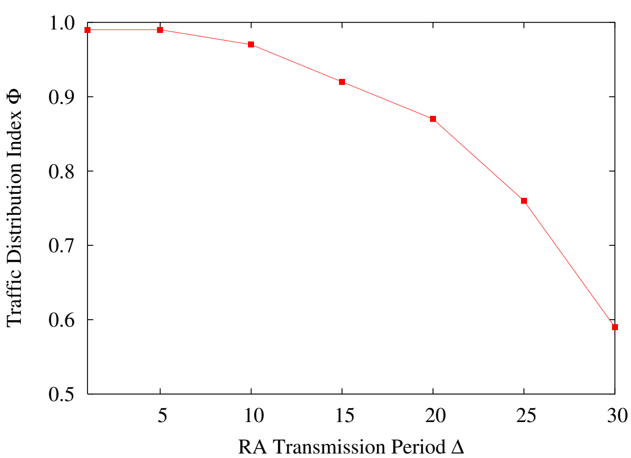


Fig. 5. Traffic distribution index Φ for different values of Δ.

V. Conclusion

This paper represents a significant enhancement to the working of hierarchical mobility management strategies proposed in recent literature. While most of these strategies attempt to solve the macro-mobility issues and provide fast transition performance, they create a complex landscape for network traffic management: some routers are overly congested with redundant transmissions while others are underutilized. To cope with this issue, we proposed a dynamic and efficient technique to select the most appropriate MAP with the lightest traffic load for communication. The MAP selection is based on an estimation of MAP load transition using the EMA method. Information on load transition is notified to ARs via the transmission of RA messages. The proposed scheme is easy to implement and the additional cost required by RA signalings is proven to be minimal.

Simulation results demonstrate that the proposed scheme has the potential of substantially improving the average communication delay, reducing the number of losses, avoiding redundant transmissions, and making better utilization of the network resources. Given the promising future of real-time multimedia services in next-generation mobile networks, the obtained results are highly encouraging and appealing for the provision of such applications in mobile networks. Finally, it should be noted that while this paper focuses solely on the case of Mobile IPv6 networks, with few modifications, this work can be also applied to mobility management over wireless metropolitan networks (e.g Worldwide Interoperability for Microwave Access (WiMax), IEEE 802.16 [18]).

References

[1] C. Perkins, *"IP mobility support"*, Network Working Group, RFC 2002, Oct. 1996.

[2] D. Johnson, C. Perkins, and J. Arkko, *"Mobility support in IPv6"*, Network Working Group, RFC 3775, Jun. 2004.

[3] R. Caceres and V.N. Padmanabhan, *"Fast and scalable handoffs for wireless internetworks"*, In Proc. ACM MOBICOM'96, NY, USA, Nov. 1996.

[4] H. Soliman, C. Catelluccia, K. El Malki, and L. Bellier, *"Hierarchical mobile IPv6 mobility management (HMIPv6)"*, draft-ietf-mipshop-hmipv6-04.txt, work in progress, Dec. 2004.

[5] A.G. Valko, *"Cellular IP: A new approach to Internet host mobility"*, ACM SIGCOMM Comp. Commun. Review, Vol. 29, No. 1, Jan. 1999.

[6] R. Ramjee, K. Varadhan, L. Salgarelli, S. Thuel, S.Y. Wang, and T. La Porta, *"HAWAII: A Domain-Based Approach for Supporting Mobility in Wide-Area Wireless Networks"*, IEEE/ACM Trans. Networking, Jun. 2002.

[7] R. Koodli, *"Fast Handovers for Mobile IPv6 (Work in Progress)"*, Internet Draft, draft-ietf-mipshop-fast-mipv6-03.txt, Oct. 2004.

[8] Y. Pan, M. Lee, J.B. Kim, and T. Suda, *"An End-to-End Multi-Path Smooth Handoff Scheme for Stream Media"*, IEEE J. Select. Areas Commun. Vol. 22, No. 4, May 2004. pp. 653-663.

[9] A. Helmy, *"A Multicast-Based Protocol for IP Mobility Support"*, ACM SIGCOMM 2^{nd} Int. Workshop on Networked Group Communication, Nov. 2000.

[10] A.T. Campbell,J. Gomez, S. Kim, Z. Turanyi, C.Y. Wan, and A. Valko, *"Design, Implementation and Evaluation of Cellular IP"*, IEEE Pers. Commun., Aug. 2000.

[11] C.E. Perkins and K.Y. Wang, *"Optimized smooth handoffs in Mobile IP"*, In Proc. IEEE Symp. Comp. Commun., Red Sea, Egypt, Jul. 1999.

[12] S. Das, A. Misra, and P. Agrawal, *"TeleMIP: Telecommunications-enhanced mobile IP architecture for fast intradomain mobility"*, IEEE Pers. Commun., Vol. 7, No. 4, Aug. 2000. pp. 50-58

[13] C.W. Pyo, J. Li, and H. Kameda, *"Dynamic and distributed domain-based mobility management method for mobile IPv6"*, In Proc. IEEE VTC Fall'03, Orlando, Florida, Oct. 2003.

[14] W. Ma and Y. Fang, *"Dynamic hierarchical mobility management strategy for mobile IP networks"*, IEEE J. Select. Areas Commun. Vol. 22, No. 4, May 2004, pp. 664-676.

[15] T. Narten, E. Nordmark, and W. Simpson, *"Neighbor discovery for IP Version 6 (IPv6)"*, Network working group, RFC 2461, Dec. 1998.

[16] Y. Gwon, J. Kempf, and A. Yegin, *"Scalability and robustness analysis of mobile IPv6, fast mobile IPv6, hierarchical mobile IPv6, and hybrid IPv6 mobility protocols using a large-scale simulation"*, In Proc. IEEE ICC 2004, Paris, France, Jun. 2004.

[17] H.A. Mantar, J. Hwang, I.T. Okumus, and S.J. Chapin, *"A scalable model for inter-bandwidth broker resource reservation and provisioning"*, IEEE J. Select. Areas Commun. Vol. 22, No. 10, Dec. 2004, pp. 2019-2034.

[18] D. Sweeney, *"Wimax operator's manual: Building 802.16 wireless networks"*, Apress; 1^{st} edition, Jun. 2004.