An Analytical Approach to the Performance Evaluation of Mobility Protocols: The Handoff Delay Case

Stefano Galli, Raquel Morera, Anthony McAuley

Telcordia Technologies, Applied Research, One Telcordia Dr., Piscataway, New Jersey 08854, USA

Abstract[(1)] **– For both commercial and military networks, it is increasingly important to be able to locate and maintain ongoing sessions with a mobile user or node. Although there are a plethora of mobility management solutions proposed, comparing the different solutions is difficult since results depend heavily on the assumptions about the type of mobility, the network characteristics, the cost associated with the mobility mechanisms, and the application requirements. Surprisingly, we found only few analytical approaches reported in the literature. Many more studies are based on simulations and/or experimental results only. One notable exception is found in [1]. Our paper extends and generalizes the approach in [1] and applies it to compare route optimization (as in MIPv6) to using only mobility agents (as in MIPv4). Under few mild assumptions: random walk mobility model, transmission cost proportional to distance, minimum distance routing, we obtain closed form expressions to quantify *exactly* the average handoff delay and the overall signaling cost. Metrics are given as a function of distance between mobile entities and are not tied to a specific scenario. In this paper, we present results for the handoff delay only, whereas a more complete report can be found in [8].**

I. Introduction

Mobility management refers to the problem of locating and providing continuous communication to a node that changes its point of attachment to the network. With the increasing growth in the number of portable and mobile terminals and the need of seamless roaming between wireless networks of different technologies (4G), mobility support (within one single network and among different networks) has become a very much demanded and needed service, in both commercial and military environments. Thus, in recent years there has been a tremendous amount of effort spent on the enhancement of existing and design of new mobility management schemes both within the standard bodies (IETF, 3GPP2, 3GPP, IEEE, etc.) and the research community. As a result, there is a plethora of solutions proposed for different environments. Different solutions are designed to accommodate different degrees of node mobility, traffic characteristics and QoS requirements.

Although some researchers show the benefits of their proposed approach based on analytical analysis ([1]-[3]), we found there is no formal method available for the objective comparison of the many proposed solutions, making more difficult the decision of what solution to implement in a specific environment. Our work is directed to overcome this lack and to provide researchers with a method that allows fair comparison via an analytical approach and not via simulations.

Although the main thrust of the authors of [1] is to show the benefits of adding Mobility Agents to a MIP solution, they also propose an original and useful way to model mobile IP networks. Their model is also used by [2][3] to evaluate different forms of hierarchical mobile IP protocols. We extend on the approach proposed in [1] generalizing it to a broader scope of mobility management schemes. Using the analytical tools here developed we make a comparative analysis of the handoff delay for different mobility management mechanisms and provide asymptotical analysis.

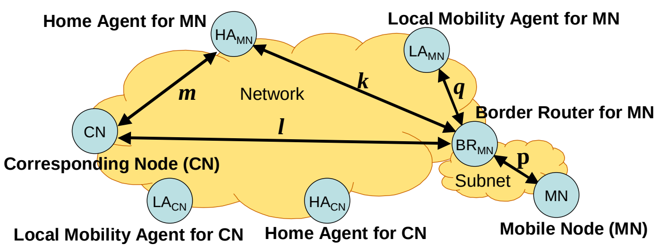
The paper is organized as follows; in section II we derive equations for the handover delay as a function of the "distance" between mobile entities for different mobility management mechanisms under the assumption of minimum distance routing. In section III, we average handover delay among *d* movements under the assumption of a random walk model for the Mobile Node (MN). To test the ubiquity of our approach, in section IV, we compare handover delay of MIPv4, MIPv6 with route optimization and a modified version of the latter. We obtain numerical results based on the equations derived in previous section and provide asymptotic analysis of the ratio between the different mechanisms.

II. Handoff Delay: Single Movement Case

In this section we derive equations for the handoff delay for two mobility management protocols based on Mobile IP (MIP) and composed of different registration and packet delivery mechanisms. The resulting equations are a function of the "distances" between the different elements involved in the communication. The generic elements involved in most MIP mobility management protocols and the distances between these elements are shown in figure 1. We assume the reader is familiar with these elements from protocols such as Mobile IP;

(1) Prepared through collaborative participation in the Communications and Networks Consortium sponsored by the U.S. Army Research Laboratory under the Collaborative Technology Alliance (CTA) Program, Cooperative Agreement DAAD19-2-01-0011. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

those wishing more tutorial material are referred to the books by Perkins [4] and Solomon [5]. Not all elements are used in all protocols and their names are sometimes different. For example, the Border Node is called the Foreign Agent (FA) in Mobile IPv4 and its distance to the MN is always one (p=1). In more general ad hoc network, p could be grater than 1.



⟷	Distance between nodes measured in number of hops
k	Number of hops on shortest path between BR and HA
l	Number of hops on shortest path between BR and CN
m	Number of hops on shortest path between HA and CN
p	Number of hops on shortest path between BR and MN
q	Number of hops on shortest path between BR and LA
s	Number of sessions between MN and all its CNs
w	Number of LA registrations for each HA registration
C_p	Cost to process one message at a node.
C_t	Cost to transmit one message over one hop

Figure 1 Distances.

When traveling to a new subnet and no kind of soft hand over is implemented, a MN is temporarily unreachable until the registration of its new care of address is successful. Then, packets can be routed to the MN's new IP address. As in cellular networks, this time is named handoff delay (HOD) and is defined as the time elapsed between the last received packet (by the MN or the Corresponding Node, CN) when the MN was attached to the old network, i.e. prior the move, and the first received packet (by the MN or CN) when the MN is attached to the new network, i.e. after the move.

When a MN is the receiver, the HOD has mainly two components: a) mobility detection and b) registration/rebinding delay. In this paper we abstract our analysis from any physical layer specific or physical layer interaction approach that can minimize handover delay, concentrating mainly on higher layer mobility management schemes. Of the two components of the HOD, mobility detection can be heavily dependent on lower layer mechanisms and particular implementations. For the sake of comparing exclusively higher layer mechanisms, we assume the same mobility detection procedure for all protocols and do not include this parameter in our analysis. Thus, we express the HOD only as a function of the registration/binding update (BU) delay when the MN is the receiver of the communication. When the MN is the transmitter, the HOD is a function of packet delivery mechanisms in specific protocols, e.g. via triangular routing or route optimization.

The HOD is a function of the processing delay per-packet (C_p) at each network entity and transmission delay per-packet (C_{tx}) between two neighboring nodes one hop apart. We here assume all hops have similar delays (e.g., an *all* wireless network), so sending a packet over multiple (i) hops has delay iC_{tx}. In heavily loaded networks, C_{tx} and C_p are a function of network load, mobility agent and router's capabilities, scheduling scheme mechanisms, etc.; however, as a first order approximation, the variables C_{tx} and C_p can be considered constant in lightly loaded networks.

The IETF has proposed different approaches to support mobility transparently in the network layer. We here derive the equations for the HOD for the two most popular mobility protocols: a) MIPv4 with triangular routing [6] and, b) MIPv6 with route optimization [7]. In order to concentrate on the comparison of higher layer mechanisms we do not consider the differences between the two different IP protocol versions. For example, we obviate the fact that IPv4 and IPv6 have different IP header length due to different IP address size. Furthermore, the MIPv6 protocol (in its current version [6]) has two modes of operation: a) double tunneling and b) route optimization. Due to space constrains and the similarity between the mechanisms used in MIPv6 with double tunneling and MIPv4, we focus on MIPv6 with route optimization option.

A. Home Agent Registration and Triangular Routing: MIPv4

In the MIPv4 framework, the handoff delay (HOD_{MIPv4}) after a move is proportional to the time it takes the MN to send a registration packet to its Home Agent (HA) and get the acknowledgment back. If we assume that a FA acts as an intermediary that is one hop away from the MN and distance k from the HA, then:

$$HOD_{MIPv4}(k) = \begin{cases} 2C_{p,FA} + C_{p,HA} + 2(k+1)C_{tx}, \\ \text{(when MN is roaming)} \\ C_{p,HA} + 2C_{tx}, \\ \text{(when MN is in its home network)} \end{cases} \tag{1}$$

where $C_{p,FA}$ is the processing cost at the FA and $C_{p,HA}$ is the processing cost at the HA.

In MIPv4 with FA, the HOD is the same whether the MN is transmitting or receiving, as the BR/FA will not forward any packets unless the HA registration is complete.

B. HA Registration, Return Routability and Binding Updates with CN: the case of MIPv6

In MIPv6 with route optimization, the MN has to send a HA registration and execute the Return Routability (RR) procedure prior sending the BU to the CN. Thus, registration delay ($REGD_{MIPv6}$) has three components: registration delay ($REGD_{HA}$), RR procedure delay (RRD) and BU delay (BUD), where:

$$REGD_{HA}(k) = C_{p,HA} + 2(k+1)C_{tx} \tag{2}$$

$$RRD_{RR}(k,l,m) = 2\{C_{p,CN} + C_{p,HA} + (k+m+l+2)C_{tx}\} \tag{3}$$

$$BUD(l) = C_{p,CN} + 2(l+1)C_{tx} \tag{4}$$

The difference between equations (1) and (2) is that there is no FA in MIPv6, i.e. there is direct registration with the HA.

In MIPv6 with route optimization, service disruption will only occur when the receive end is moving. The MN sends packets to the CN with its new IP address as soon as it acquires an IP address in its new network. Therefore, assuming that the MN involved in the session is transmitting for half of the time and receiving for the other half, we can express the handoff delay as:

$$HOD_{MIPv6}(k,l,m) = \frac{1}{2} REGD_{MIPv6}(k,l,m) \tag{5}$$

III. Averaging Handoff Delay over *d* Movements

In order to achieve scalability, IP networks are divided into IP subnets. Following [1], in our model we assume that these IP subnets are abstracted to equal-sized, non-overlapping, rectangular cells arranged in a grid configuration. In our abstract model, when the MN crosses the cell boundaries this means that the MN connects to a different subnet and an IP address configuration is required. Note that the abstract representation of IP subnets does not necessary need to match a geographical area, i.e. for a MN moving at a constant velocity, the mobility rate can be different at different cells.

The size of the grid depends on the number of movements the MN is allowed to make. If the MN moves d times, the MN can be located in any cell of a $(2d+1)$x$(2d+1)$ grid. We denote as he *center* of the grid the cell where the MN is initially located before starting to move. If indices of matrix elements, (i,j) for $i=0, \ldots, 2d+1$, $j=0, \ldots, 2d+1$, denote the coordinates of the corresponding locations in the grid, the center cell has coordinates (d,d) (see figure 2).

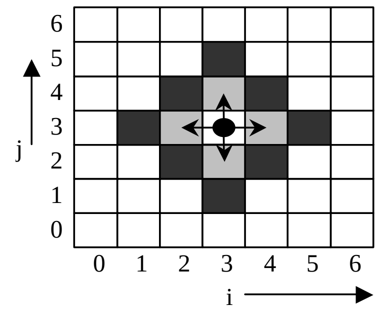
The distance between two cells in the grid is measured by the minimum number of cell boundary crossings required to travel between the cells. If we assume that routes, like the MN movements, can only take horizontal or vertical paths, i.e. diagonal paths are not allowed, the distance between cells C_1 and C_2 with coordinates (x_1,y_1) and (x_2,y_2) is:

$$d_{C_1,C_2} = |x_2 - x_1| + |y_2 - y_1| \tag{6}$$

If a mobile agent is at a distance k away from a generic cell of coordinates (i,j), there are $4k$ possible cell coordinates at distance k from (i,j) in the grid configuration for the mobility agent ($s=0, 1, \ldots, k-1$):

$$\begin{aligned} C_{s,1} &\equiv (i+s,\ j+k-s) \\ C_{s,2} &\equiv (i-s,\ j-k+s) \\ C_{s,3} &\equiv (i-k+s,\ j+s) \\ C_{s,4} &\equiv (i+k-s,\ j-s) \end{aligned} \tag{7}$$

For simplicity, we will here consider the case where the MN is the only mobile entity, whereas CN and HA do not move. On the basis of the assumed random walk mobility model, i.e. the MN may travel to one of the four neighboring cells with equal probability ¼, it is very simple to compute the probability that the MN is in a particular cell. Let us define a matrix L^n where each element $L^n_{i,j}$ is the probability that the MN is located in the cell (i,j) after it moves n times following the random walk model ($0 \le n \le d$). Matrix L^n has size $(2d+1)$x$(2d+1)$, where d is the maximum number of movements the MN is allowed to make.



- The MN is initially in center cell (e.g., at coordinates <3,3>).
- After one move the MN can move to one of four cells (<3,2>, <2,3>, <3,4>, <4,3>).
- After *3* moves, MN will be somewhere in a 7x7 grid
- Cells at distance 1 (2) from center are in light (dark) gray.

Figure 2: General Mobility Model for a MN (d=3).

In the initial state, the MN is in the center cell at a distance k from the mobile entity of interest (HA, CN). Since we want to condition delays to *distances only* and not to the specific location of a mobile entity, we need to consider that there are $4k$ possible locations in the grid for the HA or the CN that are at distance k form the MN. As the MN starts moving from the center cell, the distances between the generic cell (i,j) where the MN is located and all the (fixed) cells where the mobile entity of interest can be located will start changing as well. These $4k$ distances are easily computed ($s=0, 1, \ldots, k-1$):

$$\begin{aligned} D_{s,1}(\xi; i, j) &= |d+s-i| + |d+\xi-s-j| \\ D_{s,2}(\xi; i, j) &= |d-s-i| + |d-\xi+s-j| \\ D_{s,3}(\xi; i, j) &= |d-\xi+s-i| + |d+s-j| \\ D_{s,4}(\xi; i, j) &= |d+\xi-s-i| + |d-s-j| \end{aligned} \tag{8}$$

Now, if the MN arrives in cell (i,j) after d moves, the handoff delay averaged over all the possible locations of the mobile entity (HA, CN) of interest from the MN becomes:

$$\overline{HOD(k \mid \text{MN is in } (i,j) \text{ after } d \text{ moves})}^{D(k;i,j)} = \frac{1}{4k} \sum_{s=0}^{k-1} \sum_{a=1}^{4} HOD\big(D_{s,a}(k;i,j)\big) \tag{9}$$

where $HOD\big(D_{s,a}(k;i,j)\big)$ indicates the basic handoff delay as given in (1) (or in (5)) computed for the distances $D_{s,a}(k;i,j)$ in (8), and $\overline{}^{D(k;i,j)}$ denotes averaging over all the possible distances between the locations of the mobile entity of interest (which are at a distance k from the center cell of the grid) and the MN located in cell (i,j).

The HOD in (9) is conditional to the event that the mobile entity of interest is at a distance k from the MN at the 0^{th} movement (the center cell of the grid) and to the event that the MN arrives in cell (i,j) after d random walk movements. Weighting this conditional delay by the probability that the MN is located in cell (i,j) after d movements, and then summing over all the possible cells where the MN can be located after d moves, gives us the handoff delay at the d-th movement and conditional only to the fact that the mobile entity of interest is at a distance k from the center cell:

$$\overline{HOD(k \mid \text{after } d \text{ moves})}^{D(k)} = \sum_{i=1}^{2d+1} \sum_{j=1}^{2d+1} L_{i,j}^{d} \, \overline{HOD(k \mid \text{MN in } (i,j) \text{ after } d \text{ moves})}^{D(k;i,j)} \quad (10)$$

where $L_{i,j}^{d}$ denotes the probability that the MN arrives in cell (i,j) after d random walk moves starting from the center cell.

We are now able to compute the handoff delay incurred by a MN averaged over d consecutive random walk movements:

$$\overline{HOD}^{(d)}(k) = \frac{1}{d+1}\left[HOD(k) + \sum_{h=1}^{d} \overline{HOD(k \mid \text{after } h \text{ moves})}^{D(k)} \right] \quad (11)$$

In the previous equation, the first term represents the HOD at the 0^{th} movement (when the MN is in the center cell of coordinates (d,d) and is at a distance k from the mobile entity of interest), whereas the second term represents the total delay after d registrations, one at each movement.

IV. Comparative Analysis

In this section we compare the HOD of the mobility mechanisms described in Sect. II. If we carefully look into equations (1) and (5) we can clearly see that the HOD in MIPv4 is always smaller than the HOD in MIPv6 with route optimization. The reason being that, HA registration, RR and BU with CN have to happen sequentially in order to prevent the network from security attacks. The only advantage of MIPv6 with BUs versus MIPv4 is not related to HOD but to other performance parameters such as the load on the HA. This is a strong advantage of MIPv6 vs MIPv4 but given the space limitation its analysis is out of the scope of this paper. If the network had certificate authorities or had built-in very strong security mechanism, HA registration and BU could be performed in parallel and the RR procedure would not be required. This is not possible in current commercial networks but if needed, military networks can be designed to guarantee high security level and therefore sending directly the BU to the CN without waiting for the HA registration or the RR mechanism would be possible.

Thus, let us perform a more interesting comparison in terms of HOD between MIPv4 and an "idealized" version of route optimization, i.e. where BUs can be sent directly to the CN without waiting for the HA registration and RR procedure to be complete. In such protocol, the HOD is:

$$HOD_{BU}(l) = \frac{C_{p,CN}}{2} + (l+1)C_{tx} \quad (12)$$

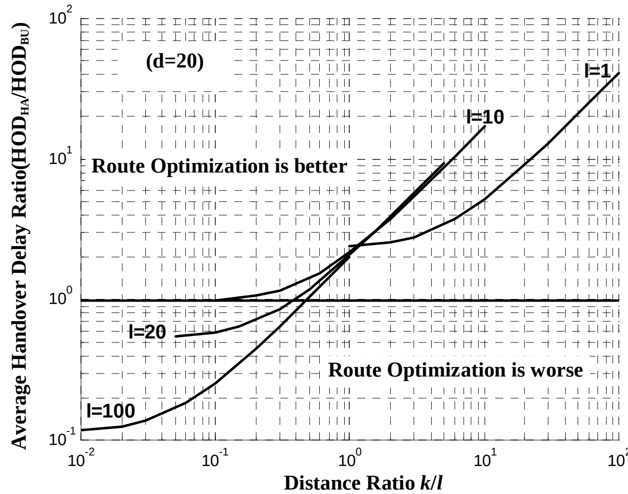


Figure 3: Log-Log plot of the HOD Ratio versus k/l, for several values of l.

The ratio of the average handoff delays in the HA-based and the BU-based mechanisms is

$$HODRatio(k;l) = \frac{\overline{HOD}_{MIPv4}^{(d)}(k)}{\overline{HOD}_{BU}^{(d)}(l)} \quad (13)$$

If the ratio is higher (lower) than one, then using a HA performs worse (better) than using BU. At the crossover point (ratio equal to one), the two mechanisms perform the same.

It is obviously difficult to assign absolute values to the processing and transmission costs. However, by analyzing the ratio of the metrics, the problem of assigning specific values to each cost is by-passed. Moreover, assuming a lightly loaded network, the processing at the FA, at the HA, and at the CN can be assumed the same (i.e. $C_{p,HA}=C_{p,FA}= C_{p,CN}=C_p$).

Figure 3 is a log-log plot of eq.(13) (for $C_{tx}=C_p$) as a function of the ratio of distances k/l, for several values of l and for d=20. As the figure clearly shows, for high values of l there are crossover points in the interval $0.1<k/l<0.5$ (depending on l). These are very interesting results as indicate that there are specific configurations in which the handover delay using HA is lower than the HOD when using route optimization. This fact occurs when the distance between the MN and the CN is much larger than the distance between the MN and the HA. From the graphs, we can also state that: a) the use of HA support *always* yields to a larger HOD when MN and CN are close to each other (l<10) and, b) the use of HA support yields a larger HOD when the two CNs are far apart (l>10) *and* when the distance k from the HA is larger than a fraction α of the distance l between CNs, with α>0.1.

It is also interesting to look at the asymptotic behavior of *RatioHOD*($k;l$) for large values of l. Starting from eq.(13) and passing to the limit for l going to infinity and for k/l fixed, we have found the asymptotic value of the handoff delay ratio:

$$HODRatio^{l\to\infty} = \lim_{\substack{l\to\infty \\ k/l \text{ fixed}}} HODRatio(k;l) \approx 2\frac{k}{l} \tag{14}$$

The interesting result expressed by (14) is that, even if the MN and CN are extremely far apart (very large l), the handoff delay with rebinds can still be lower than the one with HA support ($HODRatio(k;l)>1$) when k is larger than $0.5l$ (independently of m).

For $k=l$, i.e. when the CN is in the home network, the ratio of the HOD between the two mechanisms is plotted in figure 4 as a function of $k=l$, for several values of the averaging value d. We observe that as the distance k increases, the number of movements has less impact on the HODratio.

Assuming that $k>>d$, it can be easily proven that this ratio is asymptotically equal to two (or six) as k and l increase and C_p is negligible or at most similar to C_{tx}, (C_p much larger than C_{tx},). It is easy to prove that eq.(13) boils down to the following expression:

$$HODRatio(k=l) \approx \begin{cases} 2, & \text{for } C_p << C_{tx} \\ \dfrac{4k+10}{2k+3}, & \text{for } C_p \approx C_{tx} \\ 6, & \text{for } C_p >> C_{tx} \end{cases} \tag{15}$$

Now, passing to the limit for $k\to\infty$, we obtain:

$$HODRatio^{k\to\infty}(k=l) = \lim_{k\to\infty} HODRatio(k=l) = \\ = \begin{cases} 6, & \text{for } C_p >> C_{tx} \\ 2, & \text{otherwise} \end{cases} \tag{16}$$

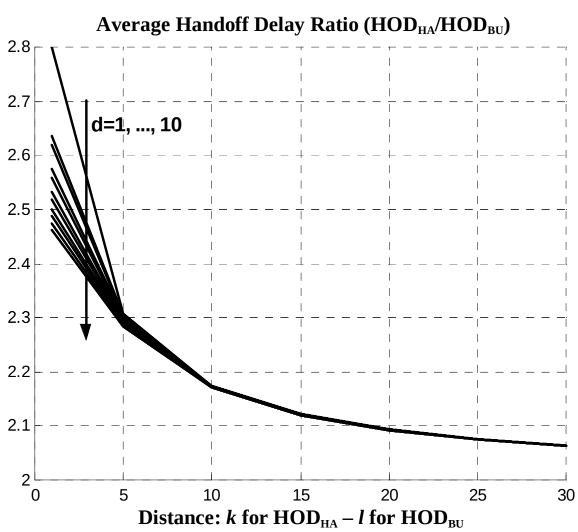


Figure 4: HOD Ratio for several values of the averaging d, versus $k=l$.

This results shows that BU is able to offer substantially lower handoff delays with respect to MIPv4, and that the larger gains are obtained when the processing delay is much larger than the transmission delay. The factor of 2 is due to the fact that MIPv4 has to wait for the registration acknowledgment whether it is receiving or transmitting, while in the case of using BU, the MN can send packets immediately after it gets a new IP address.

V. Conclusions

Under few assumptions (random walk mobility model, transmission cost proportional to distance, minimum distance routing), this paper describes closed form expressions to quantify the handoff delay of two mobility mechanisms, the HA-based and the BU-based ones. This allows us to estimate the performance of different protocols, choose under what conditions each protocol performs best and better tune the protocol parameters to increase performances. The comparative analysis here proposed allows us to quantify the benefits and tradeoffs of each of the approaches and take appropriate decisions when different mobility protocols are at our choice. This is an important result in many applications, as it gives a first order approximation on whether to use the HA option or route optimization option in MIPv6 as a function of the distance of the mobile entities. More extensive results on the handoff delay and on the overall signaling costs can be found in [8]. Although we have focused in this paper basically on two mobility managements mechanisms, the analysis is easily extensible to other approaches(2).

VI. References

[1] Yu Wang, Weidong Chen, J.S.M. Ho, "Performance Analysis of Mobile IP Extended with Routing Agents," *Proceedings of the 2nd European IASTED International Conference on Parallel and Distributed Systems*, July 1998.

[2] I.F. Akyildiz, J.S.M. Ho, Yi-Bing Lin, "Movement Based Location Update and Selective Paging for PCS Networks," *IEEE/ACM Transactions on Networking*, vol.4, no.4, Aug. 1996.

[3] S. Pack, Y. Choi, "Performance Analysis of Hierarchical Mobile IPv6 in IP-Based cellular Networks"

[4] C. Perkins, "Mobile IP Design Principles and Applications," New York: Prentice Hall, 1998.

[5] J. Solomon, "Mobile IP the Internet Unplugged." New York: Prentice Hall, 1998.

[6] C. Perkins, "IP Mobility Support for IPv4," Request for Comments 3344, August2002. <http://www.ietf.org/rfc/rfc3344.txt>

[7] D. Johnson, C. Perkins, J. Arkko, "Mobility Support in IPv6," draft-ietf-mobileip-ipv6-24.txt, June 2003. <http://www.ietf.org/internet-drafts/draft-ietf-mobileip-ipv6-24.txt>

[8] S. Galli, R. Morera, A. McAuley, "An Analytical Approach to the Performance Evaluation of Mobility Protocols: Handoff Delay and Overall Signaling Cost", manuscript in preparation.

(2) The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied of the Army Research Laboratory or the U.S. Government.