Hesham Soliman

INTERNET-DRAFT
Expires: November 2006

Qualcomm-Flarion

May, 2006

Network-based mobility considered harmful
draft-soliman-netlmm-harmful-00.txt

Status of this memo

This document is a submission of the IETF MIP6 WG. Comments should be directed to the IPv6 WG mailing list, mip6@ietf.org.

Abstract

Over the last six years the IETF has developed a number of protocols that are expected to be useful for localized mobility management (LMM), both for IPv4 and IPv6. All of these protocols were designed as extensions to Mobile IPv4 and Mobile IPv6. However, more recently, the IESG has approved the formation of the NETLMM WG, which aims to develop a network-based localized mobility management protocol. This memo discusses the impacts of a network-based mobility management protocol on the Internet and the IETF. The aim of this memo is to generate discussion in this area in order to better critique the network-based approach and show its impacts.

Table of Contents

1. Introduction

[MIPv6] and [MIPv4] were produced by the IETF to address mobility management for IPv6 and IPv4, respectively. In the last 6 years, it became evident that neither protocol is sufficient for managing frequent mobility in a fast and efficient manner. In order to improve the performance of Mobile IP handovers, a large effort started in the Mobile IP WG to address what was then referred to as Localized mobility management (LMM) for IPv4 and IPv6. In order to improve the handover performance, two main components were addressed: 1) Reducing the time it takes for the mobile node to detect that it has moved to a new link, and reducing the resulting address configuration time (specifically for IPv6); and 2) minimizing the amount of signaling required to re-route packets to the mobile node's new location. As a result of this effort the following specifications were produced: [LOWLAT], [HMIPv6] and [FMIPv6]. Other supporting protocols were also produced by the Seamoby WG: [CT] and [CARD].

Recently, the IESG has chartered the NETLMM WG to develop a network-based approach to mobility management. The NETLMM approach excludes the mobile node from any IP mobility decision, which is a distinct departure from all of the mobility protocols mentioned above. This memo critiques the approach used by the NETLMM WG and highlights some of the problems that it can cause.

This memo begins by analyzing the problem statement that led to the establishment of the WG then proceeds to discuss the technical impacts of such protocol on the deployments.

2. Problems in the problem statement

The problem statement used to motivate the NETLMM effort has, in the author's opinion, a number of factually incorrect statements. This memo addresses the perceived problems with the current LMM approach developed in IETF, i.e. the solutions mentioned above. The draft lists four problems with the current approach.

The first perceived problem is that current approaches require updates to the host software, which, however small, limit the broad deployment of those protocols. If this concern were used as a basis for protocol design, we would have to minimize the addition of any new software to hosts in order to ensure wide deployment of new features. As a result, we would maximize the dependency on new features in the network, creating the inevitable, and sad, result of "smart networks and dumb hosts", which is diametrically opposed to the past and current design philosophy used in IETF. Moreover, it is not clear why some software is acceptable in the host (e.g. the

entire TCP/IP stack, resource reservation, session control, security) while other software is not. Finally, what proof exists today that

relying on adding software to a host limits the deployment of new features? The problem statement does not address these questions.

The second perceived problem with current approaches is that other global mobility management protocols may be used in future; therefore relying on Mobile IP extensions may not be appropriate because there is an underlying assumption that Mobile IP will be used for global mobility. The authors of the problem statement cite Mobike, which is not a mobility management protocol, and HIP. Regardless of the merits of those protocols cited as candidates for global mobility, the argument is fundamentally flawed. The entire premise of the existing mobility management protocols (in IPv6) is that they are completely decoupled from the global mobility protocol. It is true that both local and global protocols use [MIPv6], however, the current LMM protocol [HMIPv6] does not require the use of MIPv6 for global mobility. The same can be said for MIPv4-based approaches that use a local HA.

The third perceived problem is that existing LMM solutions do not support both IPv4 and IPv6. This is simply wrong. The MIP6 WG has adopted [DSMIPv6] for this purpose and MIPv4 WG is considering an equivalent approach [DSMIPv4].

The fourth and final perceived problem is that the current LMM protocols require complex security mechanisms. One cannot argue with this point since complexity, like beauty, is in the eye of the beholder. However, it is worth noting that FMIPv6 does not yet have a security mechanism defined, and HMIPv6 relies on IPsec, which happens to be widely deployed and also used for MIPv6.

3. Problems with the NETLMM approach

The NETLMM approach is described on a high level in the current charter. Simply put, NETLMM expects to place a mobility anchor point in the visited network, which acts like a home agent. The mobile node will be configured with an address on the anchor point's link and keep it for the time it is located in the NETLMM domain. The anchor point binds the mobile node's address to its current default router (or Access Router, AR). Hence, whenever the mobile node moves, the new AR will bind its address to the one allocated to the mobile node. Tunneling is done between the anchor point and the AR. Therefore, the AR's address can be seen as a Care-of address for the mobile node. On a more detailed level, several minor changes can be made; however, the overview above gives the general idea.

This approach raises a number of problems that were not discussed in the process of establishing the WG. Some of these concerns are presented in this section.

3.1. Applicability to different link layers

The NETLMM WG charter states that the protocol will be link-layer agnostic by running over IP. This is not completely accurate. It is certainly true that running over IP provides some independence from the link layer technology. However, when it comes to mobility management, simply running over IP is not sufficient for link-layer independence. A mobility management protocol that improves handover performance needs to be able to adapt a sequence of events depending on the capabilities of the link layer. For instance, Fast handovers in IPv4 and IPv6 (FMIP) will run in different modes depending on the underlying link layer. Broadly speaking, in the context of mobility management, there are two types of link-layers:

1) Link-layers that allow the network to be aware of the mobile node's next AR and anticipating its movement. Examples of such link-layers include most cellular networks in existence today (but not all).

2) link-layers that do not provide the knowledge of the mobile node's next AR to the network. Examples of such link-layers include the widely deployed WLAN technologies and some cellular link-layers.

Where the network is aware of the mobile node's next AR, the network may be able to do the signaling on behalf of the mobile in order to perform a predictive handover (more problems with network signaling are discussed in the following sections). However, if the network is not aware of the mobile node's next AR, it will not be able to perform a predictive handover, which leaves it with reactive handovers as the only option, which would typically result in worse performance. Independently of the performance issue, it is important to note that NETLMM is not link-layer agnostic when it comes to mobility management issues.

3.2. Networks with multiple access technologies

The NETLMM problem statement document defines local mobility as follows:

> Local Mobility is mobility over a restricted area of the network topology. Note that, although the area of network topology over which the mobile node moves may be restricted, the actual geographic area could be quite large, depending on the mapping between the network topology and the wireless coverage area.

According to this definition, local mobility may take place between two ARs connected to two different radio technologies. This poses a serious problem for any network-based mobility management scheme. A mobile node may wish to move some or all of its traffic to one access

technology. However, a network-based mobility management scheme cannot be aware of the mobile node's preferences and may force one

technology for all of the mobile node's ongoing, and possibly future, connections. This situation can also cause operators to force a node to be connected to a particular technology, which may not be the preferred choice for the mobile node. This situation is not addressed in the current charter or work done so far in the NETLMM WG. A network-based mobility management scheme cannot handle this situation in a reliable or deterministic manner. The flaws with a network-based approach in this situation are:

(a) If one accepts that global mobility management is going to be Mobile IP based then one accepts the idea that the end node should be able to select between links to different administrative domains (or network operators). Links to different operators can of course be of the same or different technology. If this is a good thing, why do we not want to provide the host with the same flexibility when different links/technologies are available under the same local domain?

(b) Multi-homed end nodes will at some point be able to use different links for different applications depending on link quality/capabilities. It is easy to see that the level of complexity increases significantly when taking into account flow movement. The proliferation of applications and possibility that the end node is enabled with interfaces unknown to any given network-based mobility scheme makes this a difficult problem. How would a network based mobility management system know which flows to move to which link?

(c) Since the coverage area of different technologies is likely to overlap, the decision to use one technology or the other becomes a policy decision. The end nodes will have to deal with making such policy decisions between different networks and they should be able to make the same decisions between different technologies. The network operator should define metrics (like cost, loading etc) but it should let the end host decide what to do. This is not a philosophical point; there are concrete reasons why the host needs to make this policy decision. For instance, the host is most knowledgeable about the applications it runs and what radio technologies are best suited to those applications.

3.3. The robustness argument for end-to-end signaling

End to end signaling is important and necessary in order to maintain the end to end design philosophy of the Internet. When it comes to localized mobility management, the end to end concept remains crucial to the robustness of the mobility management mechanism. Handovers are uncertain by nature and in some cases the new attachment point may change during the handover process. This is due to the volatile nature of the radio link at cell borders, which is typically the case in most cellular technologies. It is also known that mobile nodes can

experience ping-pong movement, or cellular thrashing, during handovers; i.e. the mobile node may quickly move back and forth between two different access points for a short period of time. A

network-based mobility management protocol can cause the mobile node's traffic to be routed to the wrong AR, i.e. the AR that the mobile node was expected to move to, but did not. This can result in packet losses. In contrast, if the IP mobility signaling is initiated from the mobile node, it would be able to discover that the next AR has changed and inform the network of its new choice. When the action is taken by the mobile node it can be done in a quicker manner for predictive or reactive handovers.

3.4. Mobile Vs Network Control

One of the unwritten motivations of NETLMM is that some operators and vendors "believe" that the network must control the handover. Lets explore this belief a bit more. Specifically, what does network control mean? Why is it needed? And how does a network-based mobility management mechanism allow for more control?

One can interpret the words "Network control" to mean that the network needs to authorize the handover of a mobile node to another location. In fact, what is really meant by increasing "Network control" is increasing "Operator control". There maybe reasons why the operator needs to authorize, or at least be informed of such move. The current LMM mechanisms do not eliminate this step. In all of the Mobile IP-based solutions, the mobile node sends a message to a mobility agent that responds positively or negatively to the mobile node. Hence, it is fair to say that current LMM approaches do in fact allow for network control of the handover. Mobile IP-based approaches have gone further; Fast handovers for IPv4 and IPv6 allow the access router/FA to suggest a new link for the mobile node. Hence, a "smarter" network can suggest that a mobile node move to another link, e.g. to better utilize its radio interface resources.

Eliminating the message from the mobile node does not increase the network control it merely eliminates the ability of the mobile node to request to move to a particular location.

3.5. Network dimensioning

Network dimensioning can be more difficult in the case where a network-based approach is adopted. In a large network with millions of mobile nodes, the network is expected to be broken down into several local mobility domains under the same administrative domain. Each local mobility domain would consist of several mobility anchors (i.e. MAPs in IPv6). This is done in order to cope with the amount of signaling and traffic generated by each mobile node. Therefore, a form of load sharing is needed between the mobility anchor points within each mobility domain. One of the factors involved in dimensioning the network must be the number of users that each anchor

point can handle. This can be estimated by the amount of bandwidth available to each anchor point, the rate of signaling it can handle, and the number of tunnels available.

When moving between two different mobility domains, it is strongly recommended that the handover disruption be minimized in a similar manner to intra mobility-domain handovers. This is especially true for mobile nodes with ongoing sessions. Hence, it is indeed likely that mobile nodes may continue to be attached to an anchor point in a different mobility domain for some time after leaving that domain.

The above situation is more difficult to support when a network-based mobility management mechanism is adopted. In particular, the following problem arises. An anchor point may be required to setup a security association with any access router in the network at any time. A network administrator is suddenly forced to consider the impacts on memory capacity and the speed of the security association establishment at the critical handover time. This situation does not arise when the signaling is done end-to-end because in this case only one security association is needed, regardless of the mobile node's location. Furthermore, the security association does not need to be established during the critical handover time.

4. Non-technical impacts

The endorsement of more than one alternative for the same problem needs to be strongly justified. Unfortunately this was not the case for the NETLMM work in IETF. Both NETLMM BoFs clearly stated that the WG will exclude the mobile node from the IP mobility management signaling, which is not a typical requirement related to a problem, but one designed around a solution. This premise was challenged in both BoFs. Despite lack of clear consensus, the IESG decided to form the WG. The problem here is two-fold: How do we decide on new WGs? And what is the impact of solving the same problem in two different standards? Both questions are not specific to the NETLMM WG, however, this WG illustrates the need for a uniform and predictable process.

Developing more than one solution to solve the same problem in different scenarios is certainly possible. However, alternatives need to be strongly justified. In this case, the reasons are not accurate and in most cases factually incorrect. Furthermore, the downside of the NETLMM approach was not even discussed. This is clearly a recipe for a bad outcome.

When justifying alternatives one needs to at least answer the following questions to the satisfaction of the community:

- What is the problem with the current approach?
- Can this problem be solved by extending the current approach?
- What is the alternative?
- What are the pros and cons of such alternative?

As this document states, the first question was inadequately answered and the others were not answered at all.

5. Security Considerations

The author was using a complex, host-based, IP layer security mechanism called IPsec while writing this draft.

6. Acknowledgements

A lot of the points made in this document were discussed with, inspired by, or produced during discussions with (in alphabetical order): Scott Corson, Vince Park, Geroge Tsirtsis. Their input has significantly improved the quality of this document.

7. References

[CARD] CARD Design team, Liebsch, M., Singh, A., Chaskar H. and D. Funato, "Candidate Access Router Discovery", RFC 4066 March 2003.

[CT] Loughney, J. Ed., Nakhjiri, M., Perkins, C. and R. Koodli, "Context Transfer Protocol (CXTP)", RFC 4067 July 2005.

[DSMIPv4] Tsirtsis, G., Soliman, H., and V. Park, " Dual Stack Mobile IPv4 ", draft-tsirtsis-v4v6-mipv4-01/

[DSMIPv6] H. Soliman et al, " Mobile IPv6 for Dual Stack Hosts and Routers (DSMIPv6)", draft-ietf-mip6-nemo-v4traversal-01.

[HMIPv6] Soliman, H., Castelluccia, C., ElMalki, K., and L. Bellier, "Hierarchical Mobile IPv6 Mobility Management (HMIPv6)", RFC 4140, August 2005.

[IPv6] S. Deering and B. Hinden, "Internet Protocol version 6 (IPv6) specification", RFC 2460

[KEYWORDS] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[LOWLAT] K. ElMalki, Ed., "Low Latency handovers for Mobile IP"

[MIP-PB] Tsirtsis, G., and H. Soliman, "Mobility management for Dual stack mobile nodes, A Problem Statement", draft-ietf-mip6-dsmip-problem-01.txt, July 2005.

[MIPv4] C. Perkins, "Mobility Support for IPv4", RFC3344

[MIPv6] D. Johnson, C. Perkins and J. Arkko, "Mobility Support in IPv6", RFC 3775, June 2004.

[NEMO] Devarapalli, V., Wakikawa, R., Petrescu, A., and P. Thubert,

"Network Mobility (NEMO) Basic Support protocol", RFC 3963, January 2005.

[SEC] Arkko, J., Devarapalli, V., and F. Dupont, "Using IPsec to Protoect Mobile IPv6 Signaling between Mobile Nodes and Home Agents", RFC 3776, June 2004.

[SNRIO] Larsson, T., Gustafsson, E., and H. Levkowetz, "Use of MIPv6 in IPv4 and MIPv4 in IPv6 networks", draft-larsson-v6ops-mip-scenarios-01.txt, February 2004.

Authors' Addresses

Hesham Soliman
Qualcomm-Flarion Technologies
E-mail: Hesham@Qualcomm.com

Intellectual Property Statement

Full Copyright Statement

Disclaimer of Validity

This Internet-Draft expires September, 2006.