EFFICIENT ROAMING OVER HETEROGENEOUS WIRELESS NETWORKS

Hosame Abu-Amara[§1]*, Jeongjoon Lee*[§2]*, Catherine Rosenberg*[§2]*, and Edwin K. P. Chong*[**3]

[1] Telecommunications Engineering Program, University of Texas at Dallas, Richardson, TX 75083-0688

[2] School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907-1285.

[3] Dept. of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO 80523-1373.

ABSTRACT

A *heterogeneous mobile node* has several wireless interfaces (e.g., IEEE 802.11, Bluetooth, GPRS, and satellite). When applied to a heterogeneous mobile node, Mobile IP requires a lot of signaling, and it uses a slow handoff process. In this paper, we first describe these problems and then propose methods to mitigate them through the coordination of the mobile agents from the different wireless networks. We introduce a scheme that simplifies the registration procedure at the Home Agents and reduces the amount of signaling. Next, we assess quantitatively the savings that our scheme introduces in terms of the number of messages transmitted over the air in comparison with the traditional Mobile IP registration procedure.

1. INTRODUCTION

Over the past decade, we have seen a rapid growth in wireless technologies and services. Due to the heterogeneity of wireless technologies, the future mobile node is likely to be equipped with several wireless interfaces (e.g., IEEE 802.11, Bluetooth, GPRS (or other form of cellular data), and satellite), i.e., the mobile node (MN) is multi-homed. Though the use of several interfaces together gives the mobile node more flexibility in communication, the complexity increases because of the need for coordination among these interfaces.

The Internet Protocol (IP) is the dominant internetworking protocol today. As with wired-network users, mobile users want to use IP on their mobile devices, and this motivates efforts to add mobility to the Internet. Mobile IP [2] is a mechanism for maintaining transparent network connectivity to the mobile hosts. Mobile IP (M-IP) allows a mobile host to be addressed by the IP address it uses in its home network, regardless of the network to which it is currently attached. We believe that M-IP will be the glue between heterogeneous technologies. Other researchers have expressed the same view. For example, Pahlavan et al. [3] compared the M-IP architecture with other architectures and showed that the M-IP architecture is the most suitable. Stemm and Katz [4] introduced the concept of horizontal and vertical handoff, where the handoffs were built on top of the mobile routing capabilities of Mobile IP.

In the M-IP context, having several interfaces means that the MN has several IP addresses—typically one IP address for each interface. In this case, we say that the MN is *multi-homed*. A multi-homed MN has many options for communicating with its correspondent node (CN). However, while M-IP has laid the foundations for Internet mobility, there are still many problems to solve when we consider multi-homed MNs. Indeed, while in principle M-IP works with multi-homed MNs, the "vertical" handoff process (between different networks) needs to be improved (i.e., to be made faster) in order to make such a scenario viable. Maintaining the seamlessness during handoffs between the interfaces is one of the important issues. Another important issue with M-IP when applied to a heterogeneous scenario is that it requires a lot of signaling. We describe these problems and propose methods to mitigate them through the coordination of the mobile agents from the different wireless networks.

Our contribution in this paper consists of two parts. First, we introduce the concept of master home agent (MHA), which simplifies the registration procedure and reduces the amount of signaling. Second, we compare our schemes with the traditional M-IP registration procedure and assess quantitatively the savings that our scheme introduces in terms of the number of messages transmitted over the air in comparison with the traditional M-IP registration procedure.

The rest of this paper is organized as follows. We present the problem formulation and notation in Section 2. We then introduce our scheme in Section 3. Quantitative assessment of the savings due to our scheme appears in Section 4. The conclusion follows in Section 5.

2. NOTATION AND MODEL

2.1. Notation

We will use the following notation throughout this paper.

- HN_X : Home subnet of interface X of the MN, where X could be E (Ethernet), W (WLAN), G (GPRS), S (Satellite), or ….
- FN_X : Foreign subnet of X, where the MN is attached.
- IP_X : IP address of the interface X of the MN.
- DIP_X : Dynamic IP address of interface X of the MN.
- HA_X : Home Agent of IP_X (usually resides in its HN_X)
- FA_X : Foreign Agent in the network of FN_X.
- IP_Y^Z : IP address of agent Y in the subnet of the interface Z. For example, the IP address of the FA in the wireless LAN subnet is denoted as IP_{FA}^W.

2.2. Model

We introduce the notions of *usable* and *seized* interfaces. An interface α is said to be usable at time t if α at time t is powered on, within an area covered by the service provider domain for the interface, and has not been disabled by MN at time t. Otherwise, we say that α is unusable at time t. An

interface α is seized at time t if α sends a M-IP registration request message to HA_α at time t or if HA_α is enabled to intercept traffic destined to α at time t because α is registered at HA_α. Otherwise, we say that α is released at time t.

Thus, an interface α can be either usable_seized, usable_released, unusable_seized, or unusable_released. A usable_seized interface can directly transmit and receive IP packets because it is usable, and the interface is associated with a M-IP Care of Address (CoA).

By contrast, unusable_released interfaces have no associated CoA and have no direct access to any network, so all IP flows sent to them are expected to be lost. Unusable_seized interfaces have IP addresses that can be the destination of IP flows. Because unusable_seized interfaces have no access to any network, however, they can not directly receive or transmit IP packets. If MN has such interfaces, then MN may use other interfaces to receive or transmit packets on behalf of the unusable_seized interfaces, as we discuss later. Usable_released interfaces may also have IP addresses that can be the destination of IP flows. Because usable_released interfaces are not registered, however, flows destined to such interfaces can be delivered only if the interfaces are directly connected to their home networks. No M-IP CoA can be assigned to these interfaces.

An interface has a CoA if, and only if, the interface is seized. An interface is directly connected to its home network only if the interface is usable and released. When we say that an interface β sends a message on behalf of another interface α, we mean that α forms the message (including all IP headers), and β transmits the message as is, without any modifications, additions, deletions, or formatting at the network layer. Interface β merely encapsulates the message in the link layer format appropriate for β's transmission medium.

A change of interface α at time t means that α moves from some state s_1 to a state s_2, where s_1 may be the same as s_2, and s_1 and s_2 are states in INTERFACE_STATE. We assume that time is discrete. This is consistent with RFC 3344, in which changes for an interface are assumed not to occur more frequently than once a second. Let I be the number of all interfaces for MN. A system state is a tuple $(s_{\alpha 0}, s_{\alpha 1}, \ldots, s_{\alpha(I-1)})$ in SYSTEM_STATE that specifies the state of each interface α_i, for $0 \le i \le$ I-1. If at sometime there is a change in some interface α, then MN ideally contacts only the HA_α. Formally, let in_msg(t) be the number of messages received by the MN that are related to a change in system state at time t, and let out_msg(t) be the number of messages sent by the MN that are related to the system state. We set in_MSG(t) to $\sum_{i=0}^{t} in\_msg(i)$, where time 0 is the time when MN was powered on. Similarly, we set out_MSG(t) to $\sum_{i=0}^{t} out\_msg(i)$. We also set MSG(t) to in_MSG(t) + out_MSG(t). Our objective is then to create a policy so that MSG(t) is minimized for all t. Let Usable_Seized(ψ) be the set of usable_seized interfaces, Unusable_Released(ψ) be the set of unusable_released interfaces, Usable_Released(ψ) be the set of usable_released interfaces, and Unusable_Seized(ψ) be the set of unusable_seized interfaces in system state ψ.

A *policy* is a specification of a set MESSAGES that contains messages, the set of all allowable system states, a transition function from SYSTEM_STATES×MESSAGES to SYSTEM_STATES, and an output function from SYSTEM_STATES×MESSAGES to MESSAGES. MESSAGES contain the set of all messages that MN receives or transmits. The transition function determines the next system state of MN as a function of current state and received messages. The output function specifies the messages to be sent as a function of current state and received messages. The transitions and output functions are defined only for systems states that are allowable.

Singularity Assumption: In what follows, we assume that no two interfaces can change their states simultaneously. This assumption is for convenience and does not limit the generality of our results.

3. COORDINATION OF HOME AGENTS

In this section, we introduce our scheme. For simplicity, we assume that the MN uses only three interfaces: a wired Ethernet NIC, a WLAN card, and a GPRS modem.

We have some interesting observations regarding the IP addresses the MN possesses and the corresponding HAs:

- Since the IP addresses given to the MN are assumed static, they are seldom if ever changed, especially when the MN is on the move, and therefore the corresponding HAs will not change either.

When a CN initiates a connection to the MN, it must specify the destination address in the header with the IP address of the MN. In practice, the wireless interfaces for the MN are not always on. Therefore, it is undesirable to use any of the IP addresses assigned to these wireless interfaces as the destination address of the MN. We believe that for a multi-homed MN, it is preferable to identify the node in terms of the IP address of one of its wired interfaces. For example, in our scenario, it will be IP_E.

Based on the problem formulation and the above observations, we propose a new scheme by introducing the concept of a *Master Home Agent* (MHA), which is simply a HA corresponding to one of the IP addresses the MN possesses. We will call this IP address the *master* IP address. The other HAs besides the MHA will be called *Slave Home Agents* (SHAs). Similarly, the corresponding IP addresses are called as the *slave* IP addresses. In our example, HA_E is almost certainly the best choice for the MHA among the HAs corresponding to the IP addresses of the MN, because IP_E is well known to the CNs, and the MN is likely to prefer to use the Ethernet interface more than any other, whenever it is available. Making a HA a MHA is simple. We do not use any additional messages or extensions to do this, and in this sense our scheme is simple and robust. We also assume that there is a trusted relationship between all these HAs. Suppose the MN is at the home network of the MHA. It does not have to send any registration message to the MHA since it is at home; however, it has to send registration messages to the SHAs since the network of the MHA is a foreign network from the standpoint of the slave IP addresses. In this case, in contrast to the tradi-

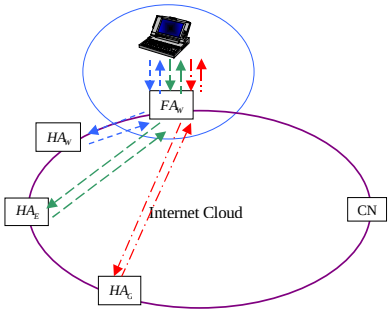

Fig. 2. Traditional M-IP registration when the MN is at FN_W

tional M-IP registration procedure, the MN sends registration messages to the SHAs with the MN's master IP address as the CCoA. Usually, the CCoA is obtained by requesting it from the DHCP server and therefore it is routable. In our case, we do not have to request any IP address from the DHCP server. Since the MN is at home, its IP address (here, master IP address) is inherently routable, and hence there is no difficulty in using the master IP address as the CCoA. The lifetime field in the registration will be set to infinity. Once it is done, the MN does not have to send new registration messages to the SHAs anymore. When the MN moves to a new foreign network, it only needs to send registration messages to the MHA, and by not sending registration messages to the SHAs, we can make the SHAs believe that the MN is still in the home subnet where the MHA resides. As a result, the SHAs tunnel the packets destined for the MN to the MHA, and the MHA then tunnels them to the FA in the network that the MN is currently visiting.

We will examine the registration procedures and the routing paths for our scheme under several scenarios and compare them with the case when we use the traditional M-IP registration.

3.1. Configuring the MHA and SHAs when the MN is at the home subnet of the wired Ethernet interface, HN_E

Making a HA a MHA is done when the MN is at the home subnet, which corresponds to the master IP address chosen by the MN. Suppose the mobile user chooses IP_E as its master IP address. When the MN is at the home subnet of the wired Ethernet, HN_E, it sends registration messages to the SHAs, HA_W and HA_G, with IP_E as the CCoA. Note that the MN can set the lifetime of these registrations as infinite, and then sending one registration message to each SHA is enough; the MN does not necessarily have to send registration messages again to these SHAs. Note also that the MN does not ask for any help from FA_E at this time.

Consider the traditional Mobile IP registration message flow in this case. Since the network associated with the master IP address where the MN currently resides, i.e., HN_E, is a foreign network from the standpoint of the slave IP addresses,

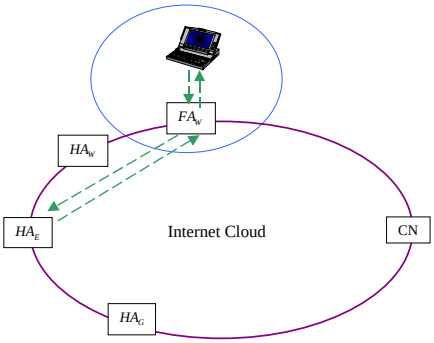

Fig. 1. Registration in our scheme when the MN is at FN_W

the MN has to send registration messages to HA_W and HA_G via FA_E whenever the lifetime is about to expire. In contrast, in our scheme, sending only one registration message with infinite lifetime to each SHA (HA_W and HA_G) is enough.

3.2. The MN is at a foreign subnet of the WLAN network, FN_W

In traditional M-IP registration, the MN needs to send three registration messages separately to all the HAs, as shown in Fig. 2. However, in our scheme, the MN does not have to send registration messages to the SHAs, because the SHAs already have a registration for the MN with an infinite lifetime. Fig. 1 shows the registration message flow to the MHA, HA_E, which is the only registration flow needed. In this way, we can reduce the amount of M-IP signaling when the MN is visiting foreign networks. Once the registration is done, the packets are delivered to the MN as shown in Fig. 4.

While the registration message flow is much simpler in our scheme, the routing path may be longer than in traditional Mobile IP, as can be seen by comparing Fig. 4 with Fig. 3. As we can see in Fig. 4, in our scheme, the packets with the destination address IP_W or IP_G experience slightly longer routing paths because of one extra path from HA_W or HA_G to HA_E. (The packets with the destination address IP_E are routed through HA_E and FA_W, as in an ordinary M-IP triangular delivery.) It should be noted that as the distance between the MHA and the SHAs decreases, the effect of this difference becomes negligible. For example, wired Ethernet and wireless LAN subnets are usually organized by one authority, and the HAs are close compared to the long path from the CN to the HAs.

4. QUANTITATIVE ASSESSMENT

Fig. 5 shows the transitions between the possible states of an interface α, as mapped from Mobile IP RFCs [2] to our model. The set INTERFACE_STATE of possible states comprises us_ac (usable_seized), un_ac (unusable_seized), us_in (usable_released), and un_in (unusable_released). The transition labels use concepts discussed in Mobile IP RFCs [2]. A transition marked **Reg** means that α has to register with HA_α.

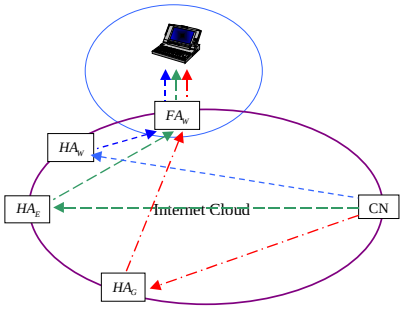

Fig. 3. Traditional M-IP packet delivery when the MN is at FN_W

A transition marked **deReg** means that α has to deregister itself from HA_α. A transition marked **expire** means that α must allow its registration with HA_α to expire. A transition marked **?** means that M-IP does not specify particular procedures for the transition. Note that we allow a transition from state un_in to state un_ac by sending a Reg message for some interface α. This transition is possible only if some other usable interface sends the Reg message on behalf of α. In what follows, we assume that the transitions marked by **X** in the figure are not allowed. This assumption is for convenience only and does not compromise the generality of our results.

4.1. Basic Implementation of Mobile IP

Consider a basic implementation of mobile IP. In this implementation, the policy assumes that every allowable system state contains at most one usable interface. If a system state has a usable interface AC (for ACcess), then every seized interface, other than AC, has a CoA that is equal to IP_{AC}. If a system state has no usable interfaces, then all IP flows to seized interfaces will be lost. This scheme is actually an enhanced version of the mobile IP scheme decribed in the RFCs. As we will show, even with this enhanced version, the scheme is inferior to the one we propose in Section 4.2 below. Consider a transition from some system state ψ1 to another state ψ2≠ψ1. By the Singularity Assumption, exactly one interface α has different states in ψ1 and ψ2. There are several cases, depending on the states of α:

4.1.1. α was unusable in ψ1 and is usable in ψ2:

Then α sends a Reg message on behalf of each interface β that is in Unusable_Seized(ψ1) and whose CoA is not already IP_α.

The Reg message specifies that the CoA is to be set to IP_α, and the message is sent to HA_β. In addition, depending on the new state of α in ψ2, interface α may also need to send a Reg message on behalf of itself. By the operation of mobile IP, each Reg message generates a Reply message that is sent to α. In the worst case, this case requires a total of $2*|\text{Unusable\_Seized}(\psi1)|+2$ messages, where $|\text{Unusable\_Seized}(\psi1)|$ is the size of the Unusable_Seized(ψ1) set.

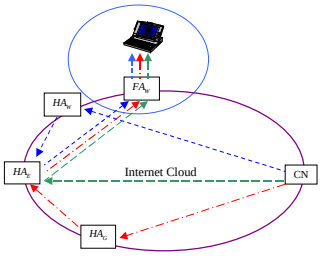

Fig. 4 Packet delivery in our scheme when the MN is at FN_W

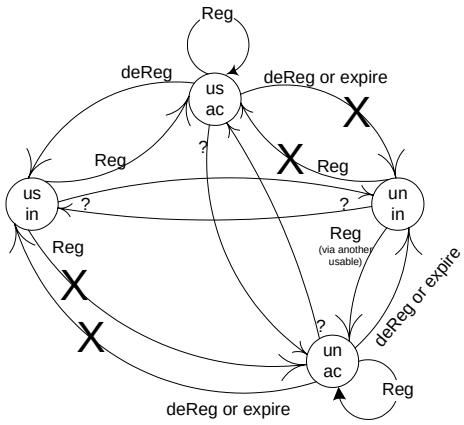

Fig. 5. Transition Function for Interface States

4.1.2. α was us_in in ψ1 and is us_ac in ψ2:

Then α sends a Reg message on behalf of itself. The Reg message specifies a value for CoA, which may be obtained from a DHCP. By the operation of mobile IP, the Reg message generates a Reply message that is sent to α. In the worst case, this case requires a total of 6 messages (2 messages for mobile IP, and 4 more messages for DHCP IP address request).

4.1.3. All other cases:

Due to space limitations, we state that it can be shown that the worst case number of messages in a system state transition is $\max(6, 2*|\text{Unusable\_Seized}(\psi1)|+2)$ messages.

4.2. Our Implementation of Mobile IP

In our implementation, the policy assumes that a specific interface SPEC is chosen so that every seized interface, other than SPEC, has a CoA that is equal to IP_{SPEC}. Consider a transition from some system state ψ1 to another state ψ2≠ψ1. By the Singularity Assumption, exactly one interface α has different states in ψ1 and ψ2. There are several cases, depending on the states of α:

4.2.1. α was unusable in $\psi 1$ and is usable in $\psi 2$:

4.2.1.1. If α is SPEC:

This case requires no new messages to be sent.

4.2.1.2. If α is not SPEC:

If α does not have a DIP, then α is required to request another IP address DIP_α from FA_α, perhaps from a DHCP server attached to FA_α. The reason for the request is as follows. All IP flows destined for are sent to IP_{SPEC}. If SPEC is unusable, then HA_{SPEC} forwards the IP flows to a usable interface. Interface α may be the only usable interface. Hence, α needs an IP address to which HA_{SPEC} can forward IP flows. SPEC can not rely on using IP_α, because α may be seized, so that HA_α may intercept messages sent to IP_α. So, α needs a DIP_α in our implementation. Note that this scheme requires α to maintain a dynamic address in addition to its permanent address. In the worst case, this case requires 4 messages for DHCP IP address request and reply.

4.2.2. α was us_in in $\psi 1$ and is us_ac in $\psi 2$:

Case α = SPEC:

Then SPEC chooses a usable interface β and sends a Reg message via β to HA_{SPEC}. The Reg message specifies that the CoA is to be set either to DIP_β if β is us_ac or to IP_β if β is us_in. By the operation of mobile IP, the Reg message generates a Reply message that is sent to α.
In the worst case, this case requires a total of 2 messages.

Case $\alpha \neq$ SPEC:

Then α sends a Reg message on behalf of itself via some usable interface. The Reg message specifies that the CoA is to be set to IP_{SPEC}. By the operation of mobile IP, the Reg message generates a Reply message that is sent to α. In addition, if the CoA for SPEC was IP_α in $\psi 2$, then SPEC must choose a new CoA, e.g. DIP_α, in the system state that immediately follows $\psi 2$. In the worst case, this case requires a total 2 messages.

4.2.3. α was usable in $\psi 1$ and is unusable in $\psi 2$:

This case requires no new messages to be sent. Nevertheless, there are implications on the internal operations of the MN, as follows.

4.2.3.1. If α was us_ac in $\psi 1$ and is un_ac in $\psi 2$:

Case α = SPEC:

There are no implications on the internal operations of the MN.

Case $\alpha \neq$ SPEC:

If the CoA for SPEC was DIP_α in $\psi 2$, then SPEC must choose a new CoA in the system state that immediately follows $\psi 2$.

4.2.3.2. If α was us_in in $\psi 1$ and is un_in in $\psi 2$:

Case α = SPEC:

Then all IP flows destined to SPEC and to all interfaces of the MN will be lost. In the state that immediately follows $\psi 2$, SPEC must choose a usable interface β and sends a Reg message via β to HA_{SPEC}. The Reg message specifies that the CoA is to be set either to DIP_β if β is us_ac or to IP_β if β is us_in.

Case $\alpha \neq$ SPEC:

Then all IP flows destined to α will be lost. If the CoA for SPEC was IP_α in $\psi 2$, then SPEC must choose a new CoA in the system state that immediately follows $\psi 2$.

4.2.4. All other cases:

Due to space limitations, we state that it can be shown that the worst case number of messages in a system state transition is 4 messages.

5. CONCLUSION

Future mobile nodes will be equipped with multiple interfaces to realize anytime, anywhere, any service, and coordination of these interfaces is a major concern. Maintaining seamlessness during handoffs over heterogeneous wireless networks is an important issue. At the same time, the M-IP signaling issue we raised in this paper is also important.

In this paper, we introduce a scheme that simplifies the registration procedure and reduces the amount of signaling. Second, we compare our scheme with the traditional M-IP registration procedure and assess quantitatively the savings that our scheme introduces in terms of the number of messages transmitted over the air in comparison with the traditional M-IP registration procedure. We show that our scheme uses at most 4 messages per system state transition, compared with at least 6 messages per system state transition in the traditional scheme.

6. REFERENCES

[1] A. T. Campbell, J. Gomez, S. Kim, A. G. Balko, C-Y. Wan, and Z. R. Turanyi, "Design, implementation, and evaluation of Cellular IP," *IEEE Personal Communications*, Aug. 2000, pp. 42-49

[2] C. Perkins, editor, "IP Mobility Support for IPv4", *IETF Network Working Group, Request For Comment 3344*, Aug. 2002

[3] K. Pahlavan, P. Krishnamurthy, A. Hatami, M. Ylianttila, J-P. Makela, R. Pichna, and J. Vallstrom, "Handoff in hybrid mobile data networks," *IEEE Personal Communications*, Apr. 2000, pp. 34-47

[4] M. Stemm and R. H. Katz, "Vertical handoffs in wireless overlay networks," *ACM Mobile Networks and Applications (MONET)*, Vol.3, Issue 4, 1998, pp. 335-350

[5] X. Zhao, C. Castelluccia, and M. Baker, "Flexible network support for mobile hosts," *ACM/Baltzer Journal on Special Topics in Mobile Networks and Applications (MONET)*, Vol. 6, Issue 2, 2001, pp. 137-149

§ This work was supported in part by a grant from Nortel Networks, from the NSF grant No. 0087266 and by the Indiana Twenty- First Century Fund through the Indiana Center for Wireless Communication and Networking.

** This research was supported in part by NSF under grants ECS-0098089, ANI-0099137, and ANI-0207892.