# PERFORMANCE ANALYSIS OF AMP FOR MOBILITY MANAGEMENT

**Wan H. Hassan** wanh@sunway.edu.my

*School of Computer Technology,*
*Sunway University College,*
*Malaysia*

**Aisha-Hassan A. Hashim** aisha@iiu.edu.my
*ECE Department, Faculty of Engineering,*
*International Islamic University*
*Malaysia*

**Norsheila Fisal** sheila@fke.utm.my
*Telemetries & Optical Eng. Dept.,*
*Faculty of Electrical Eng., UTM,*
*Malaysia*

---

## Abstract

In our previous work [1], a mechanism for handling movement detection using a proposed agent-based architecture for mobility management was described. This architecture, referred here as AMP - Agent-based Mobility Protocol, consists of a collaborative multi-agent system that enhances user/node mobility over an IP-based network. Specifically, mobility agents are placed in the hosts and at the access networks to expedite location and call management requirements. State information of mobile hosts (e.g. location and mobility profile) are relayed to the relevant agents who, in turn, will undertake appropriate tasks to ensure a smooth handover to the next cell(s) during an on-going application session with minimum delay. In this paper, the performance of the AMP architecture and protocol is examined using derived analytical models. Mobile QoS (Quality of Service) may be defined by signaling traffic overhead, handoff latency and packet loss. A comparative analysis is made between the AMP architecture and the IETF's standard mobility management protocol i.e. Mobile IPv6.

---

## 1 INTRODUCTION

Mobility management refers to two main components – location management and handover management. The former refers to the ability of the network to track the location of mobile users between consecutive communications. The latter refers to the process by which the network maintains an active connection for a mobile user as he or she moves from one access point to another. The ability to provide efficient support for mobility management relies on the level of intelligence and the mechanisms by which this intelligence, or control information, is conveyed to the appropriate network nodes or elements, and the corresponding actions taken by those nodes. The Internet differs significantly from the mobile cellular network on how this is accomplished. The notion of mobility in the Internet is typically hidden from the IP network, and any intelligence in facilitating mobility management is restricted to end systems and certain specialised nodes or mobility agents such as the Home Agent (and Foreign Agents) in Mobile IP.

Furthermore, transmission is best effort without a separate mechanism for signaling. The main advantage is simplicity, which translates to lower deployment cost and relative ease of operations. However, the drawbacks include variable delay, contention of resources and consequently, the inability to provide timing and bandwidth guarantees resulting in service degradation. In a sharp contrast, network intelligence is integrated explicitly in cellular networks for mobility management. Extensive collaboration exists between network entities to accommodate and facilitate location and handover management. In addition, signaling and other control information to ensure call delivery and quality of transmission is done out-of-band ensuring efficient and reliable transmission for mobile subscribers. However, the overhead in such systems is complexity and higher cost to service providers and subscribers. In addition, adding new services usually entails costly upgrades in both hardware and software. Henceforth, the research here proposes to address these perplexities by using an approach that takes advantage of both the above, by implementing a mobile-aware node/application working in conjunction with a network that can adapt dynamically for the benefit of the mobile user.

## 2. NETWORK REFERENCE MODEL

For the purpose of analysis, a general network model is used as reference as shown in Figure 1. Here, a mobile host, *MH*, roams into a visited network, Access Network 1. The home network has a mobility agent which maintains the current location of the mobile host. In the proposed AMP architecture, this would refer to the Home Registrar, $Rg_H$, while in Mobile IP, for example, this would typically refer to the Home Agent, *HA*. Each network is connected to the Internet via an access router, *AR*. It is assumed that each network is an autonomous system. In access networks 1 & 2, there are cells for wireless connectivity where each cell represents a particular subnet within the access network. Furthermore, it is assumed that there is a correspondent host, *CH*, which may be in the form of a server, interacting with the mobile host.

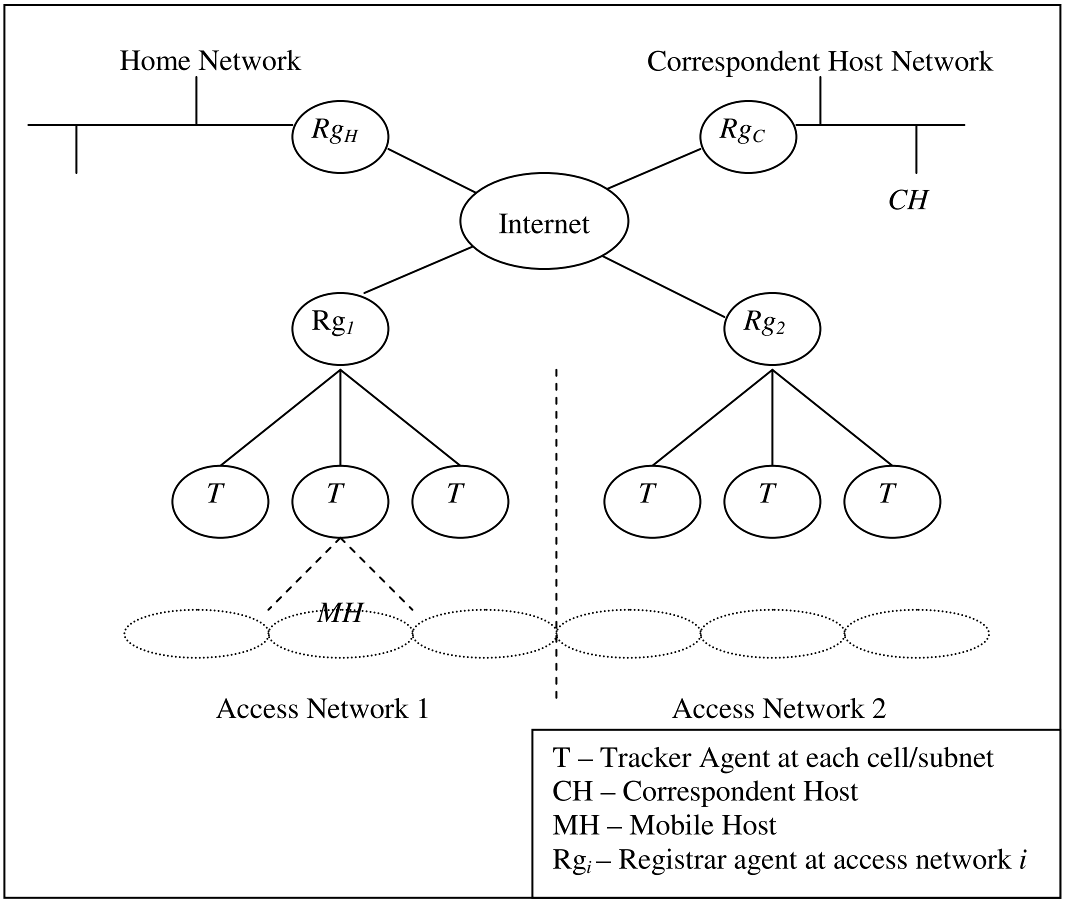


**Figure 1**: AMP reference model

In AMP, a hierarchical architecture is used. In each cell (or subnet), there is a Tracker agent (*T*) to monitor the connection, and maintain the current location and state of each and every mobile host that are directly attached to the cell within the visited access network. Each cell is assumed to be a subnet within an access network, and the tracker agent is connected to a router within the cell/subnet. The tracking mechanism, using flag updates. A Registrar agent manages the tracker agents (for all cells) within a particular access network, as depicted in Figure 1. A Registrar agent is associated with each Access Router (*AR*) in the AMP architecture i.e. access network *i* would have one Registrar agent ($R_i$) associated to one $AR_i$. Registrar agents of neigbouring access networks would have peer-to-peer relationships.

## 3. USER MOBILITY AND TRAFFIC MODELS

For the purpose of evaluation, it is necessary to develop analytical models for user movement and traffic or call/session arrival. The following user mobility and traffic models are based from the works of Baumann [2], [3], [4], [5], [6], and [7], and these have been used in the analysis of mobility management protocols in wireless networks including those in cellular, 3G, PCS and Mobile IP architectures.

It is assumed that all incoming calls or sessions follow the Poisson process where both inter-arrival and inter-session times are exponentially distributed [8], [7]. The traffic models comprise two levels – packet and session/call. The mobile host, *MH*, is modeled by the cell/subnet residence time. For the purpose of evaluation, the following parameters are defined:

| | |
|---|---|
| $t_c$ | random variable for *MH*'s cell/subnet residence time |
| $f_c$ | probability density function of $t_c$ |
| $t_n$ | random variable for the residence time within an access network |
| $f_n$ | probability density function of $t_n$ |
| $t_s$ | inter-session time between two consecutive sessions with PDF $f_s$ |
| $t_{rs}$ | *MH* residual subnet residence time |
| $N_c$ | number of cells/subnets crossing within an access network during intra-network handoff |
| $N_n$ | number of access networks crossing during inter-network handoff |
| $G$ | global binding update cost sent to the home network or correspondent network |
| $L$ | local binding update cost sent to the local registrar |
| $M$ | number of cells/subnets within an access network |
| $K_{CN}$, $K_{CH}$ | number of correspondent networks, or correspondent hosts, having a binding cache entry for an *MH* |
| $h_{p,q}$ | number of hops between nodes *p* and *q* |
| $C_{p,q}$ | transmission cost of control packets between nodes *p* and *q* |
| $PC_p$ | processing cost of control packet at node *p* |
| $C_{hc}$ | binding update cost at home (*h*) and correspondent (*c*)networks |
| $t_{L2}$ | time period between link layer (L2) trigger to link switching. |
| $\mu_c$ | *MH* movement (border crossing) rate out of a cell/subnet |
| $\mu_n$ | *MH* movement (border crossing) rate out of an access network |
| $\mu_l$ | *MH* movement (border crossing) rate in which *MH* still stays within the same access network |

The probability $P_s$ of anticipated handoff signaling success for a particular observed valued $t_T$ may be defined as:

$$P_s = P(\chi_T > t_T) = \int_{t_T}^{\infty} f_T(u,\sigma)du \quad (1)$$

where $\chi_T$ is the random variable for the time between link layer (L2) trigger generation and link down. $\chi_T$ is assumed to be exponentially distributed, $f_T(u,\sigma)$ is the probability density function for successful completion of signaling, of which $\sigma > 0$ is a success rate parameter.

In modeling user mobility, an imbedded Markov chain is used where state $i$ ($i \geq 0$) is defined as the number of cells/subnets that the *MH* has passed. The state transition diagram is as shown in Figure 2.

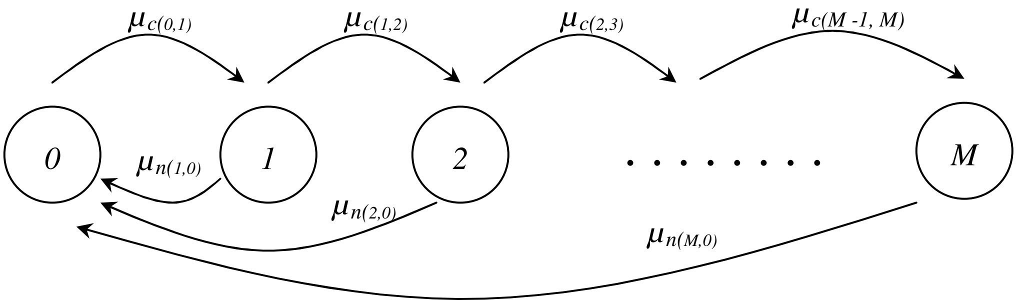


**Figure 2**: State diagram of *MH*'s movement, adapted from [3]

Based on the above Figure 2, state *M* represents the number of cells/subnets that the *MH* has passed within a single access network. The state transition $\mu_{c(i,i+1)}$ ($0 \leq i < M$) represents the rate of *MH*'s movement from one cell/subnet to another. The transitions $\mu_{n(i,0)}$ ($1 \leq i \leq M$) represent the movement to a cell/subnet out of the access network.

A cell or subnet is assumed to be circular in area. It is further assumed that the *MH* residence time within a cell/subnet or an access network follows a Poisson distribution with parameters, $\mu_c$ and $\mu_n$, respectively. From [2], if there are *M* sufficiently large cells/subnets within an access network, then the border crossing rates of *MH* may be defined as follows:

1. Movement rate out of a cell/subnet,

$$\mu_c = 2\frac{v}{\sqrt{\pi A_c}} \quad (2)$$

where $v$ is the average velocity of *MH*, and

$A_c = \pi r^2$, is the cell area with $r^2$ the radius of the cell/subnet

2. Movement rate out of an access network,

$$\mu_n = \frac{\mu_c}{\sqrt{M}} \quad (3)$$

where *M* is the number of cells/subnets within an access network

3. Movement rate in which MH still stays within the same access network,

$$\mu_l = \mu_c - \mu_n = \mu_c - \frac{\mu_c}{\sqrt{M}} = \mu_c \frac{\sqrt{M}-1}{\sqrt{M}} \qquad (4)$$

According to Makaya and Pierre (2008), the cell/subnet crossing probability ($P_c$) and the access network crossing probability ($P_n$) during an inter-session time interval may be stated as follows:

$$P_c = P(t_s > t_c) = \int_0^\infty P(t_s > u) f_c(u) du \qquad (5)$$

$$P_n = P(t_s > t_n) = \int_0^\infty P(t_s > u) f_n(u) du \qquad (6)$$

The following Lemmas from [6] may be used in deriving the analytical models and simplifying the above equations:

Lemma 1: Let $\{N_1(t), t \geq 0\}$ and $\{N_2(t), t \geq 0\}$ be two independent Poisson processes with rate $\lambda_1$ and $\lambda_2$, respectively. Let $t_1$ and $t_2$ denote the times of the first process and the second process, respectively. The probability of one event occurs in the first process before one event occurs in the second process is given as:

$$P(t_1 < t_2) = \delta(\lambda_1, \lambda_2) \equiv \frac{\lambda_1}{\lambda_1 + \lambda_2} \qquad (7)$$

Lemma 2: Let $\{N_1(t), t \geq 0\}$ and $\{N_2(t), t \geq 0\}$ be two independent Poisson processes with rate $\lambda_1$ and $\lambda_2$, respectively. Let $N$ denote the mean number of events occurring in the first process between two events in the second process. Then,

$$N = \frac{\lambda_1}{\lambda_2} \qquad (8)$$

If the session arrival is assumed to follow a Poisson distribution with rate $\lambda$, and the time for residence within a cell or access network is smaller than the session duration i.e. $(t_n \vee t_c) < t_s$, then from Lemma 1, the $MH$ crossing probabilities for both cell/subnet and access network respectively, in relation to the session arrival, may be defined as:

$$P_c = \frac{\mu_c}{\mu_c + \lambda} \qquad (9)$$

$$P_n = \frac{\mu_n}{\mu_n + \lambda} \qquad (10)$$

Let $E(N_i)$ denote the mean number of location bindings or registrations during an inter-session arrival, where $i$ represents the type of border crossing. Then, from Lemma 2, the average number location bindings for cell/subnet crossing is

$$E(N_c) = \frac{\mu_c}{\lambda} \qquad (11)$$

Similarly, the mean number of location binding updates or registrations during an inter-session arrival for access network crossing is:

$$E(N_n) = \frac{\mu_n}{\lambda} \qquad (12)$$

The mean number of location binding updates or registrations during an inter-session arrival while the $MH$ remains in the same network is:

$$E(N_l) = \frac{\mu_l}{\lambda} \tag{13}$$

## 4. SIGNALING COSTS

The total signaling cost may be defined as the cost of transmitting control packets for the purpose of:

1. creating a new address binding or mapping during crossover (or handover) i.e. binding update (*BU*);
2. refreshing an existing address binding upon timer expiry i.e. binding refresh (*BR*); and
3. packet delivery (*PD*).

Thus, the total signaling overhead cost, $C_T$, may be defined as:

$$C_T = C_{BU} + C_{BR} + C_{PD} \tag{14}$$

The signaling cost for binding updates generally may be stated as [7]:

$$C_{BU} = E(N_l)L + E(N_n)G \tag{15}$$

where

$E(N_l)$ is the average number of cell/subnets that an MH crosses but still remains in the same access network,
$L$ and $G$ are the local and global update costs respectively, and
$E(N_n)$ is the average number of access networks that an MH crosses during a particular inter-session arrival.

From (5.12) & (5.13), Equation (5.15) becomes

$$C_{BU} = \frac{1}{\lambda}\left(\mu_l L + \mu_n G\right) \tag{16}$$

From [9] and [3], the notion of call-to-mobility ratio (CMR) is used to reflect the average number of calls or sessions to a *MH* within an access network per unit of time that a *MH* changes cells/subnets per unit of time, or more succinctly, in a state of equilibrium,

$$\text{CMR} = \frac{\lambda}{\mu_c} \tag{17}$$

From Equations (5.3), (5.4) & (5.17), Equation (5.16) may be written as

$$C_{BU} = \frac{\mu_l}{\lambda}L + \frac{\mu_n}{\lambda}G = \frac{1}{\lambda}\left(\frac{\mu_c\left(\sqrt{M}-1\right)}{\sqrt{M}}L + \frac{\mu_c}{\sqrt{M}}G\right)$$

$$C_{BU} = \frac{\mu_c}{\lambda\sqrt{M}}\left[(\sqrt{M}-1)L + G\right]$$

$$C_{BU} = \frac{1}{CMR\sqrt{M}}\left[(\sqrt{M}-1)L + G\right] \tag{18}$$

### Signaling Cost for Binding Updates in AMP

In the proposed AMP architecture, the binding update may be considered either as local for intra-network movement or as global for inter-network movement. Local or global refers to the distance over which the signaling messages are sent. In the former, a binding update will be sent to the local registrar of the visited access network, while in the latter, the update message will be sent across the Internet, typically over multiple hops, to the home network and any correspondent network, where appropriate. In contrast, binding updates in Mobile IP, under versions 4 and 6, are always global regardless of the *MH*'s movement i.e. internally (acquiring a new IP address) within an access network or across to another network under a different administrative domain.

Based on the mobility management operations earlier described ealier, the local and global binding updates/signaling under the AMP architecture may be depicted as in the timeline diagrams shown in Figures 3 and 4 respectively. The local binding update cost in AMP,

$$L = 2\,(C_{MH,T} + C_{T,Rg} + PC_T) + PC_{Rg} \qquad (19)$$

where

$C_{MH,T}$ is the transmission cost between the $MH$ and the Tracker, $T$, in the new cell/subnet;
$C_{T,Rg}$ is the transmission cost between the Tracker and its Registrar, $Rg$;
$PC_T$ is the processing cost at the Tracker which include table lookup/entry/update; and
$PC_{Rg}$ is the processing cost the Registrar which includes table lookup/update and routing

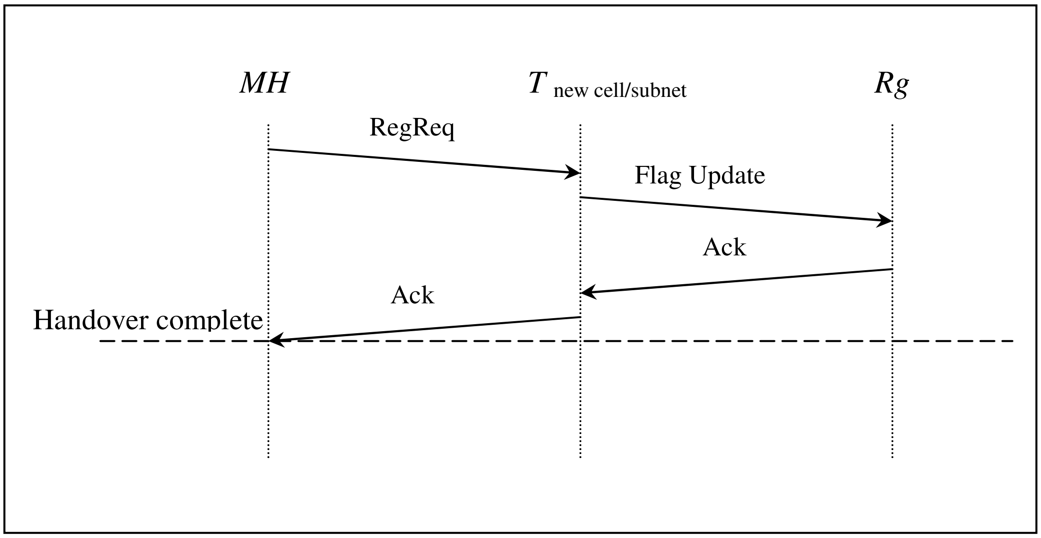


**Figure 3**: Intra-network (local) handover in AMP

From Figure 4, the global binding update cost is

$$G = 3C_{T,Rg} + 2C_{Rgprev,Rgnew} + 2C_{MH,T} + C_{hc} + 3PC_T + 6PC_{Rg} \qquad (20)$$

where

$C_{MH,T}$ is the transmission cost between the $MH$ and the Tracker $T$ in the new cell/subnet
$C_{T,Rg}$ is the transmission cost between the Tracker, $T$, and Registrar, $Rg$ (both at the previous and new access network.
$PC_T$ is the processing cost at the Tracker, $T$, which include table lookup/entry/update
$PC_{Rg}$ is the processing cost the Registrar which includes table lookup/update and routing, and
$C_{hc}$ is the cumulative binding update (BU) cost at the home Registrar agent and all active correspondent registrar agent in other networks.

$C_{hc}$ may defined as the total cost of transmission from the registrar agent in the new access network to the home registrar agent, $C_{Rgnew,RgH}$, the processing cost at the home registrar, $PC_H$, the transmission cost from the home registrar to the $K$ number of active correspondent network registrars, $KC_{RgH,RgC}$, the processing costs at each of the correspondent registrars, $KPC_{RgC}$, the transmission cost of acknowledgment from the home registrar to the new network registrar, $C_{Rgnew,RgH}$, and the transmission cost from acknowledgements from each of the correspondent registrar to the new network registrar agent $KC_{RgC,Rgnew}$:

$$C_{hc} = 2C_{Rgnew,RgH} + PC_{RgH} + KC_{RgH,RgC} + KPC_{RgC} + KC_{RgC,Rgnew} \qquad (21)$$

where the variables are as previously defined.

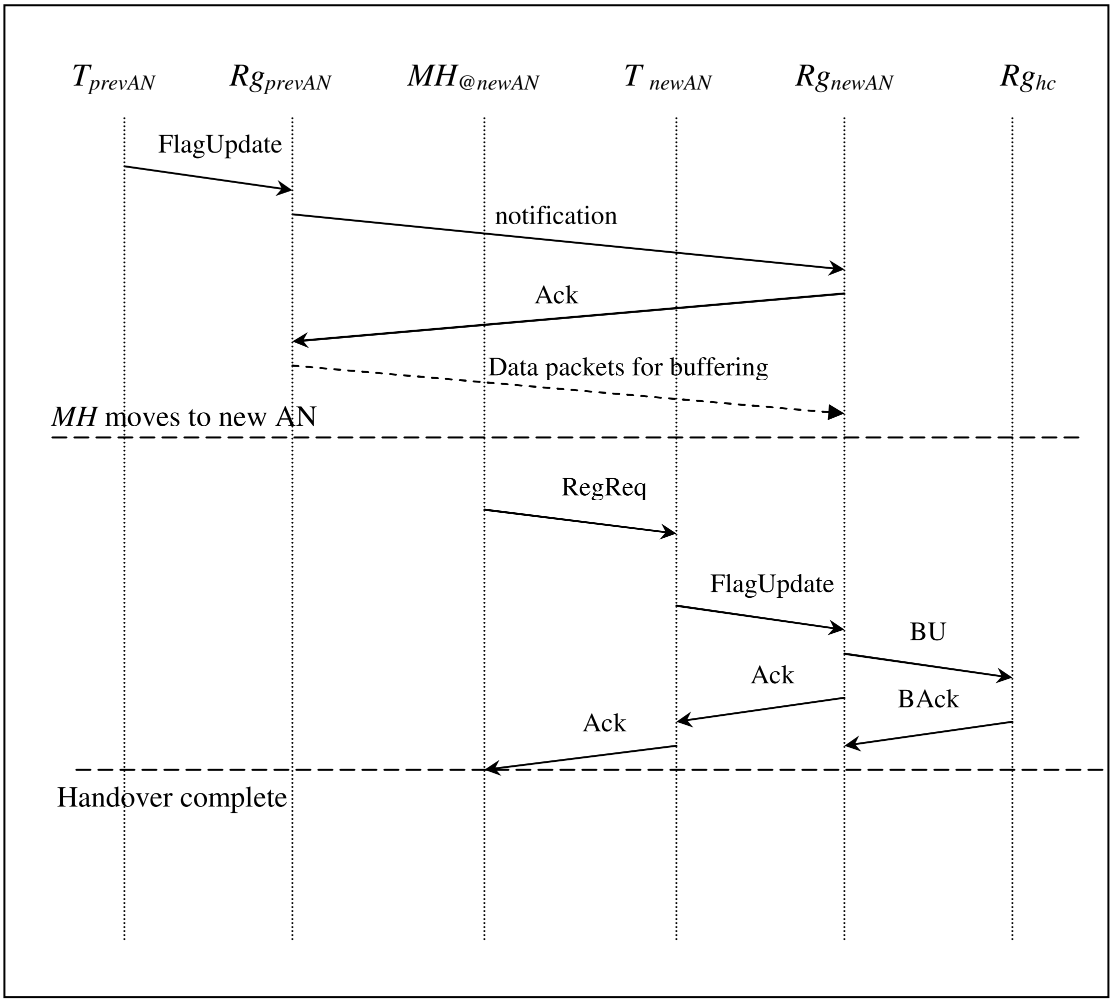


**Figure 4**: Inter-network handover in AMP

**Signaling Cost for Binding Update in Mobile IP**

As a comparison, Figure 5 shows the sequence of signaling messages for Mobile IPv6 (MIP). A mobile host (*MH*) discovers the existence of a new access router (NAR) through Router/Solicitation Advertisement (RS/RA) messages exchange. In addition, a Duplicate Address Detection (DAD) procedure is performed to ensure the *MH* has a unique care-of-address (CoA) by exchanging Neighbour Solicitation/Advertisement (NS/NA) messages. A binding update (BU) is then made to the Home Agent informing the home network of the current address of the *MH*, from which the Home Agent (*HA*) returns a binding acknowledgement (BAck) message. A BU message is also sent to any active correspondent host to enable route optimization in Mobile IP. According to [10], the DAD procedure introduces significant delay in handovers for MIP, attributing to about 78% of the total handover latency.

In MIP, additional security measures are done to ensure that BU messages are properly authenticated and are not sent by malicious *MH*s. The procedure, known as Return Routability (RR), comprise several messages i.e. Home Test Init (HoTI), Home Test (HoT), Care-of-Test Init (CoTI), and Care-of-Test (CoT), and the exchange of these messages, between the *MH*, *HA* and any correspondent host (*CH*), increases further the handover latency. In contrast, in the proposed AMP architecture, a mobile host never needs to acquire a new IP address or send BU messages since this is done securely by the network entities i.e. the Registrar agents, and not end systems or hosts.

global binding update cost is:

$$L^{MIP} = G^{MIP} = 4C_{MH,AR} + 2PC_{AR} + C_{hc}^{MIP} \tag{22}$$

where $C_{MH,AR}$ is the transmission cost of control packets between the *MH* and the access router;

$PC_{AR}$ is the processing cost at the access router; and

$C_{hc}^{MIP}$ is the binding update cost at the Home Agent (*HA*) and at all active correspondent hosts.

The binding update cost at the HA and CHs in Mobile IP is,

$$C_{hc}^{MIP} = 2(C_{MH,HA} + K_{CH}C_{MH,CH}) + PC_{HA} + K_{CH}PC_{CH} + C_{rr} \tag{23}$$

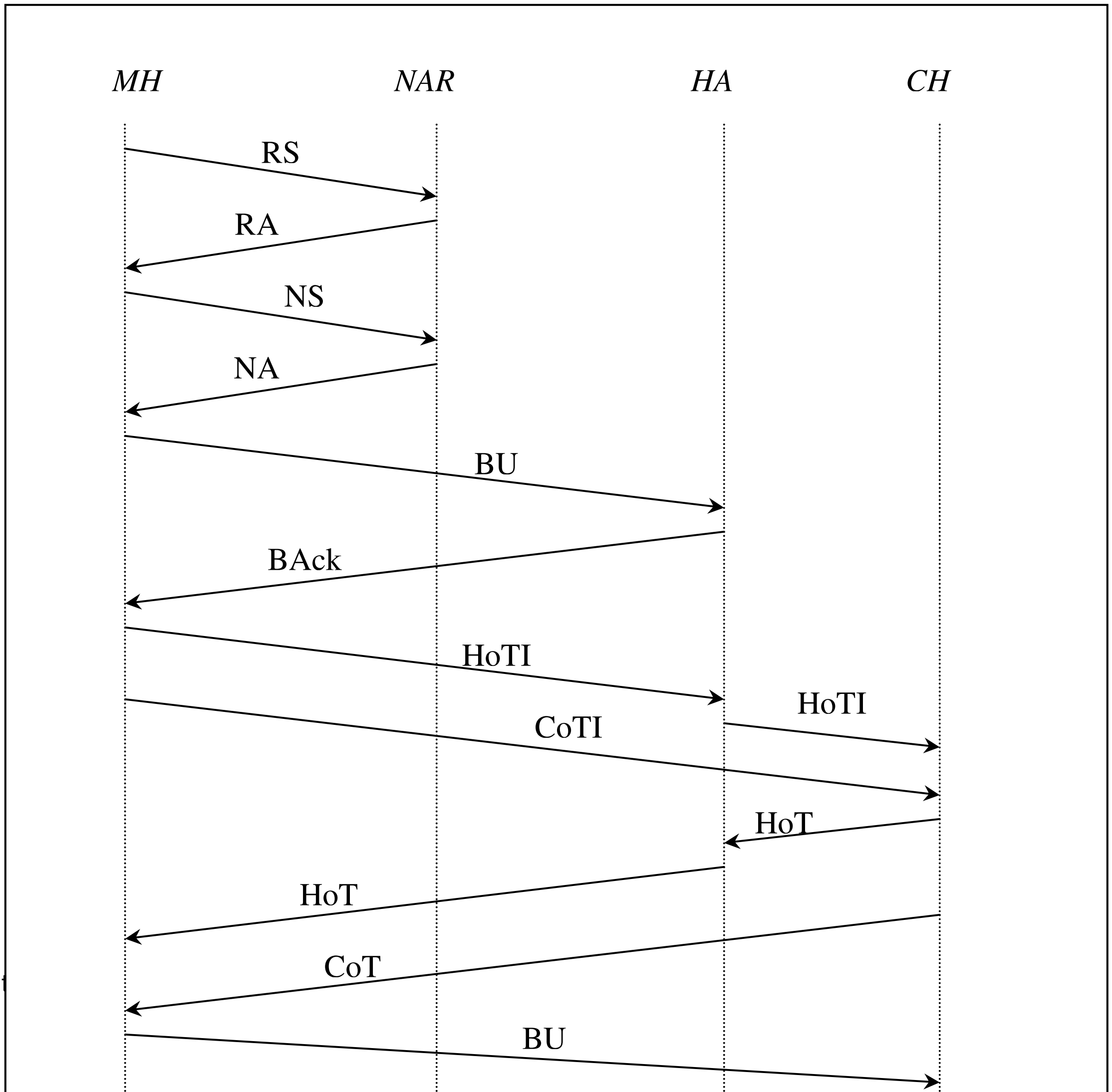
where $C_{MH,HA}$ is the transmission cost between the MH and the HA;

$K_{CH}$ is the number of active correspondent host (CH) with a binding entry for the *MH;*

$C_{MH,CH}$ is the transmission cost between the MH and the CH;

$PC_{HA}, PC_{CH}$ are the processing costs at the HA and the CH respectively; and

$C_{rr}$ is the signaling cost for return routability procedure.



The signaling cost for the return routability procedure is:

$$C_{rr} = 2(C_{MH,HA} + K_{CH}C_{HA,CH} + K_{CH}C_{MH,CH} + PC_{HA} + K_{CH}PC_{CH}) \quad (24)$$

where the variables are as previously defined.

In MIP, both intra-network (local) and inter-network (global) movement would incur the same signaling cost for the binding update procedure.

**Signaling Cost for Binding Refresh**

Binding updates are sent periodically, both Mobile IP and in the proposed AMP architecture. The average rate of sending a binding refresh message is $\lfloor 1/(\mu\Gamma) \rfloor$, where Γ is the binding lifetime period, and $\mu$ is the movement rate out of a particular subnet or access network. Assuming that $\Gamma_N$, $\Gamma_H$, and $\Gamma_C$ are the binding lifetime period at the registrar agents at the visited access network, the home network and the correspondent networks respectively, then the signaling cost for binding refresh in AMP is:

$$C_{BR} = 2\left(\left\lfloor \frac{1}{\mu_c\Gamma_N} \right\rfloor C_{MH,T_N} + \left\lfloor \frac{1}{\mu_n\Gamma_H} \right\rfloor C_{Rg_N,Rg_H}\right) + 2\left\lfloor \frac{1}{\mu_n\Gamma_C} \right\rfloor K_{Rg_C} C_{Rg_H,Rg_C} \quad (25)$$

where the variables are as previously defined. Similarly, the signaling cost for binding refresh in Mobile IP is [7]:

$$C_{BR} = 2\left\lfloor \frac{1}{\mu_n\Gamma_H} \right\rfloor C_{MH,HA} + 2\left\lfloor \frac{1}{\mu_n\Gamma_C} \right\rfloor K_{CH} C_{MH,CH} \quad (26)$$

where the variables are as previously defined.

**Signaling Cost for Packet Delivery in AMP**

In assessing the signaling cost for packet delivery, the factors that contribute towards delay during handovers are firstly considered. According to [11], the handover delay in Mobile IP comprises the link switching delay (L2 handover delay), IP connectivity latency, and the packet reception latency. The IP connectivity delay is the time required for an *MH* to obtain a new IP address and the packet reception delay refers to the time for the packets to be delivered to the new address.

Two types of packet delivery cost are considered in the AMP architecture – one for intra-network movement and another for inter-network movement. In the former, when a *MH* crosses into a new cell/subnet within the same access network, a form of packet re-direction occurs and packets may be lost during this transition period if handover fails. The handover timeline for intra-network movement is shown in Figure 6.

The cost for packet delivery in intra-network movement comprises the cost for packet loss moving from one cell/subnet to another:

$$C_{PD}^{Intra} = \zeta C_{loss} = \zeta\lambda\eta(C_{Rg,T_{prev}} + C_{T_{prev},MH})(t_{L2} + t_{IP} + t_{LU}) \quad (27)$$

where $\zeta$ is the weighing factor for packet loss;

$\lambda$ is the packet arrival rate;

$\eta = l_s/l_d$ , is the ratio of average signaling control packet length, $l_s$, to the average data packet length, $l_d$;

$(C_{Rg,T_{prev}} + C_{T_{prev},MH})$ is the cost of transferring data packets from the correspondent registrar through the tracker agent in the previous cell to the MH when handover fails; and

$(t_{L2} + t_{IP} + t_{LU})$ is the delay for the local update at the registrar agent including the link layer switching delay and the network layer (IP) connectivity delay in the new cell/subnet.

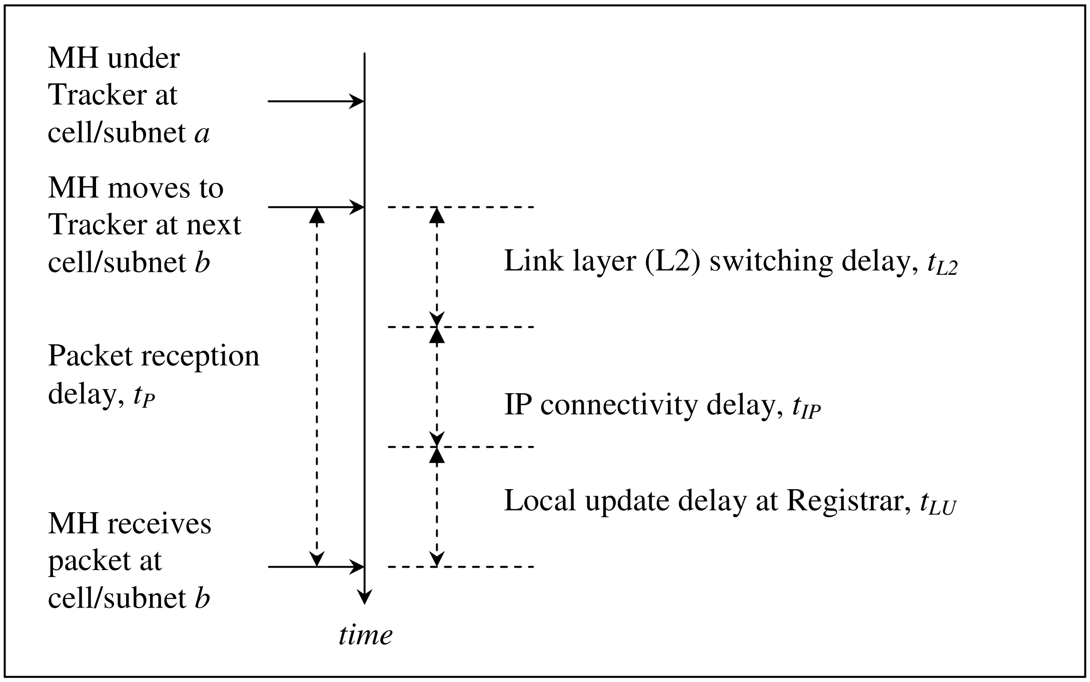


**Figure 6**: Handover timeline in AMP (intra-network)

The handover timeline diagram for inter-network movement in AMP is as shown in Figure 7. The packet delivery cost for inter-network movement considers the buffering of packets from the registrar in the previous access network to the next to mitigate packet loss.

As described earlier, the registrar agent in the next network would normally assign a new address binding for packet delivery to the *MH* before movement to the new access network. This is possible since movement of the *MH* is detected by the tracker agent. If the *MH* is located at a border cell, then pre-registration is done by the registrar agent to its peer agent in the next access network. Furthermore, the registration agent would typically buffer packets to the next registrar agent in the new access network. Hence, in most cases, there would be no packet loss and the packet reception delay would theoretically be lower than in the Mobile IP architecture.

Based on similar works by [5] on Mobile IP, the cost for packet delivery in the AMP architecture may be stated as follows:

$$C_{PD}^{Inter} = \delta C_{fwdg} + \zeta C_{loss}^{Inter} \qquad (28)$$

where $\delta$ and $\zeta$ are weighing factors such that $\delta + \zeta = 1$,

$C_{fwdg}$ is the cost of forwarding duplicate packets to the next registrar network, and

$C_{loss}^{Inter}$ is the cost of packet loss during rapid movement where $t_C \leq t_{RgRg}$

The cost of packet forwarding ($C_{fwdg}$) includes the cost of establishing a ‘link’ between the current registrar agent to its peer in the next access network i.e. pre-registration. This delay between the two peer registrar agents, $t_{RgRg}$, includes notification (and acknowledgement) of impending movement by the *MH* to the new access network, and sending packets for buffering at the new location. The cost of forwarding is proportional to the packet arrival rate, λ, the cost of transmission from the correspondent network through the registrar agents to the *MH*, and the forwarding time:

$$C_{fwdg} = \lambda\eta(C_{Rg_C,Rg_{prev}} + C_{Rg_{prev},Rg_{new}} + C_{Rg_{new},MH})(t_{L2} + t_{IP} + t_U) \qquad (29)$$

where λ and η are as previously defined in Equation (27).

In most cases, packet loss does not occur since buffering is done. However, in cases where the *MH* rate of movement is faster that the time required for pre-registration with the peer agent, then loss may occur before handover may be completed. Hence, the cost of packet loss ($C_{loss}^{Inter}$) is a function of $\lambda\max\{(t_{RgRg} - t_C),0\}$, and the cost of packet transmission to the *MH*:

$$C_{loss}^{Inter} = \lambda\max\{(t_{RgRg} - t_C),0\} \times \eta(C_{Rg_C,Rg_{prev}} + C_{Rg_{prev},MH})(t_{L2} + t_{IP} + t_U) \qquad (30)$$

where λ and η are as previously defined in Equation (27).

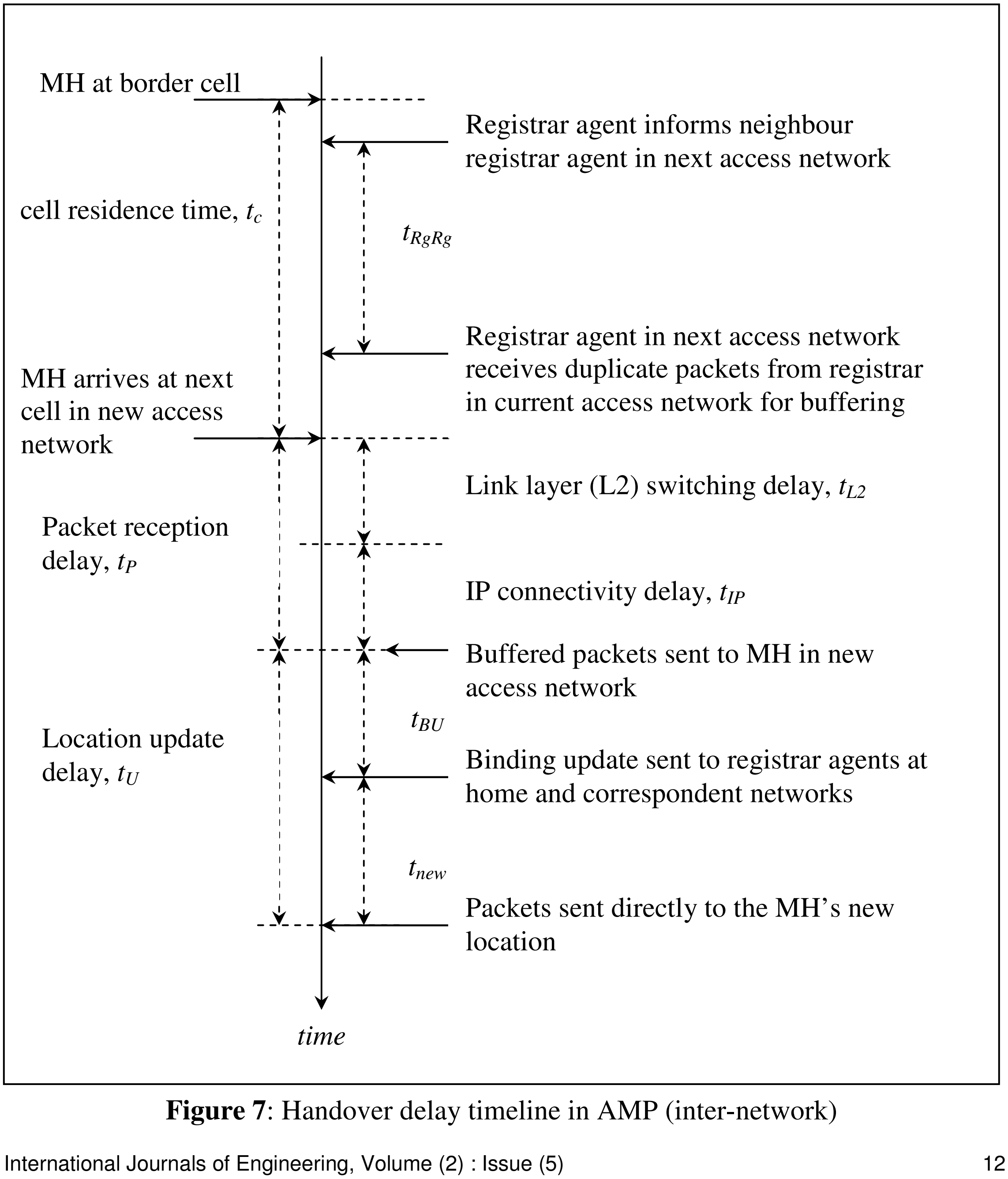


**Figure 7**: Handover delay timeline in AMP (inter-network)

### Signaling Cost for Packet Delivery in Mobile IP

In Mobile IP, there is no packet forwarding or buffering, thus the cost of packet delivery is derived from the cost of packet loss that may occur during handovers [7]:

$$C_{PD}^{MIP} = \zeta C_{loss}^{MIP} = \zeta\lambda\eta(C_{CH,PAR} + C_{PAR,MH})(t_{L2} + t_{IP} + t_{GU}) \quad (31)$$

where $C_{CH,PAR}$ is the transmission cost between the *CH* and the previous access router;

$C_{PAR,MH}$ is the transmission cost between the previous access router and the *MH*;

$\zeta$ is a weighing factor; and

$(t_{L2} + t_{IP} + t_{GU})$ is the total delay for the link layer switching delay, the network layer (IP) connectivity delay in the new network, and the global binding update at the *HA* and *CH*s (including the delay for return routability procedure).

### Handover Latency and Packet Loss

In general, the one-way transmission delay between two nodes *X* and *Y* ($t_{X,Y}$) over wired and wireless links is influenced by the packet size ($l$), link delay ($t_{wired}$, $t_{wireless}$), bandwidths ($R_{wired}$, $R_{wireless}$), the probability of wireless link failure ($p$), the number of hops between *X* and *Y* ($h_{X,Y}$), and the average queueing delay at each router hop ($t_{queue}$) [4]:

$$t_{X,Y}(l) = \frac{1-p}{1+p}\left(\frac{l}{R_{wireless}} + t_{wireless}\right) + \left(h_{X,Y} - 1\right)\left(\frac{l}{R_{wired}} + t_{wired} + t_{queue}\right) \quad (32)$$

For the AMP architecture, the handover latency is considered for both intra- (local) and inter-network (global) movement. For intra-network movement, the handover latency is:

$$D_{Intra}^{AMP} = t_{L2} + 2t_{MH,Rg} \quad (33)$$

where $t_{L2}$ is the link layer switching delay;

$t_{MH,Rg}$ is the one-way packet delay between the *MH* and the registrar agent in the visited access network.

For inter-network movement, the handover latency is:

$$D_{Inter}^{AMP} = t_{L2} + t_{MH,Rg_H} + t_{Rg_H,Rg_C} + t_{Rg_C,MH} \quad (34)$$

where $t_{L2}$ is the link layer switching delay;

$t_{MH,Rg_H}$ is the one-way transmission delay between the MH to the registrar agent in the home network;

$t_{Rg_H,Rg_C}$ is the one-way transmission delay from the home registrar to the correspondent registrar; and

$t_{Rg_C,MH}$ is the one-way transmission delay from the correspondent registrar to MH

As mentioned in the previous section, packets are buffered during inter-network movement. Hence, packet loss may also occur during handover if the buffer space at the access router (where the registrar agent resides) is of insufficient size. The required size of the buffer is proportional to the packet arrival rate, λ, and the handover latency. In addition, the space required for buffering packets would also increase to accommodate the number of mobile hosts performing handovers. Generally, the required space of the buffer, $S_{req}$, considering one mobile host, may be defined as:

$$\begin{aligned} S_{req} &= \lambda D_{Inter} \\ &= \lambda[t_{L2} + t_{MH,Rg_H} + t_{Rg_H,Rg_C} + t_{Rg_C,MH}] \end{aligned} \quad (35)$$

and the packet loss may be computed from:

$$P_{loss}^{AMP} = \max\{(S_{req} - S_{actual}),0\} \quad (36)$$

where $S_{actual}$ is the actual size of the buffer in the access router.

For Mobile IP, the handover latency is given by (Makaya and Pierre, 2008):

$$D^{MIP}_{handover} = t_{L2} + t_{RD} + t_{DAD} + t_{RR} + 2(t_{MH,HA} + t_{MH,CH}) \qquad (37)$$

where $t_{RD}$ is the round-trip delay for router discovery;
$t_{DAD}$ is the delay for the duplicate address detection procedure;
$t_{RR}$ is the delay for router routability procedure;
$t_{MH,HA}$ is the one-way transmission delay between MH and the HA; and
$t_{MH,CH}$ is the one-way transmission delay between MH and the CH.

Since there is no buffering in Mobile IP, packets will be lost during handover operation. Thus, the packet loss in MIP is:

$$\begin{aligned} P^{MIP}_{loss} &= \lambda D^{MIP}_{handover} \\ &= \lambda[t_{L2} + t_{RD} + t_{DAD} + t_{RR} + 2(t_{MH,HA} + t_{MH,CH})] \end{aligned} \qquad (38)$$

## 5. NUMERICAL RESULTS AND DISCUSSION

The network topology in Figure 1 is used for performance evaluation where there is one correspondent host transmitting to a *MH* in a visited access network. The *MH* moves from one cell/subnet to another in Access Network 1 to Access Network 2. All links are assumed to be full duplex with respect to bandwidth and latency. Furthermore, it is assumed that the distance (hop count) between the different autonomous systems is of equal distance to each other. The values or system parameters used are shown in Table 1, and are typical values used in the works of [3], [12], [4], [13], [5], [14], [15] and [7], where appropriate. The scenario assumes that the *MH* moves from one access network to another, crossing 4 cells/subnets with a total of 2 intra-network handovers and 1 inter-network handover (for AMP).

**Table 1**: System parameters

| Parameter | Notation | Value |
|---|---|---|
| Velocity of *MH* | $\nu$ | 5.6 km/h |
| No of correspondent host or network | $K$ | 1 |
| Transmission cost between peer Registrar agents *X,Y* | $C_{X,Y}$ | 10 hops |
| Transmission cost between Tracker agent to Registrar *X* | $C_{X,T}$ | 2 hops |
| Transmission cost between *MH* to Tracker agent | $C_{MH,T}$ | 1 hop |
| Processing cost at Home Registrar | $PC_{RgH}$ | 24 |
| Processing cost at peer Registrar *X* in other networks | $PC_{RgX}$ | 12 |
| Processing cost at correspondent host (in MIP) | $PC_{CH}$ | 4 |
| Number of Tracker agent (cell/subnet) per access network | $M$ | 3 |
| Packet arrival rate | $\lambda$ | 10 packets/s |
| Control or signaling packet size | $l_s$ | 96 bytes |
| UDP data packet size | $l_d$ | 200 bytes |
| Cell/subnet radius | $r$ | 500 m |
| Weighing factor for packet loss | $\zeta$ | 0.8 |
| Weighing factor for packet forwarding | $\delta$ | 0.2 |
| Link layer (L2) switching delay | $t_{L2}$ | 50 ms |
| IP connectivity delay | $t_{IP}$ | 10 ms |
| DAD delay (in MIP) | $t_{DAD}$ | 500 ms |
| Router Discovery delay (in MIP) | $t_{RD}$ | 100 ms |
| Wireless link failure probability | $p$ | 0.5 |
| Wired link bandwidth | $R_{wired}$ | 100 Mbps |
| Wireless link bandwidth | $R_{wireless}$ | 11 Mbps |
| Wired link delay | $t_{wired}$ | 2 ms |
| Wireless link delay | $t_{wireless}$ | 10 ms |
| Average queueing delay | $t_{queue}$ | 5 ms |

Based on the derived analytical models and the system parameters used, the total signaling cost of the AMP architecture and the total signaling cost in Mobile IP are calculated and shown in Figure 8.

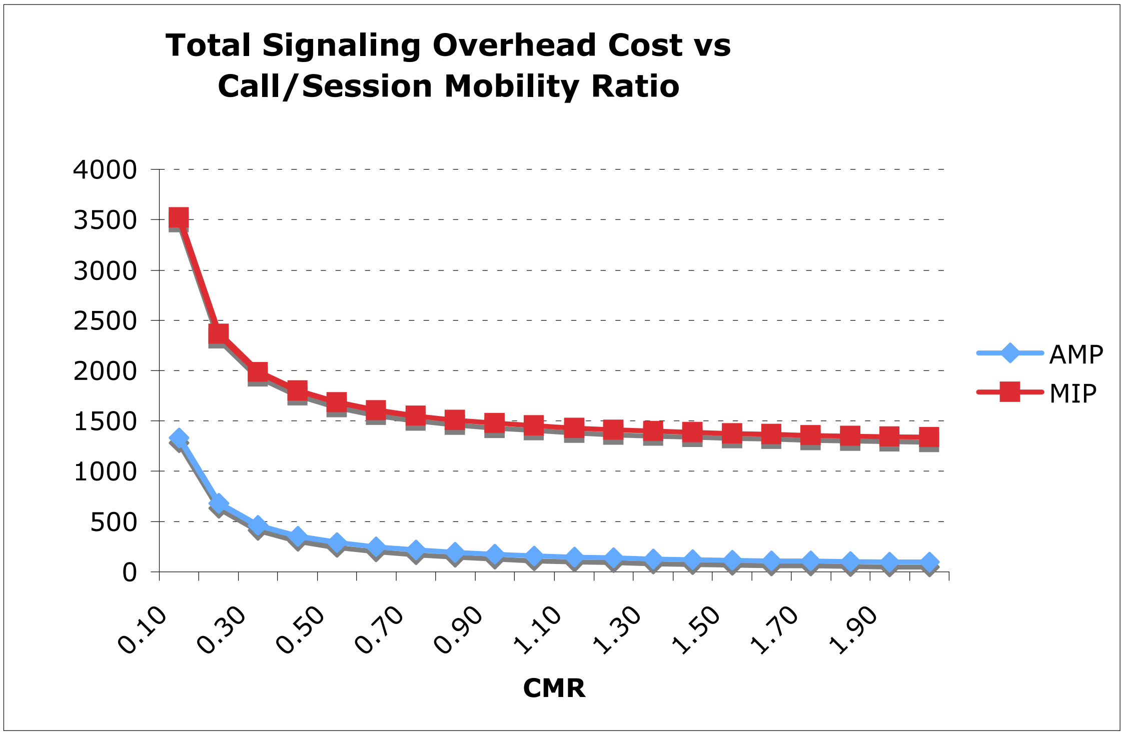


**Figure 8**: Effect of call/session mobility ratio (CMR) on signaling cost

Generally, the signaling overhead is high when the call/session mobility ratio is small. At small values of CMR, the mobility rate is much larger than the call/session arrival rate. Hence, a *MH* crosses over many subnets and induces several handovers. Frequent movement and/or handovers result in higher signaling overhead since binding updates have to be made more often due to the change in location of the *MH*. When the call/session arrival rate is larger than the mobility rate i.e. CMR > 1, less binding updates are performed due to the smaller number of crossovers. The signaling overhead in MIP is much larger than the signaling overhead in AMP. This is because most of the binding updates in AMP are for intra-network movement. Binding updates are localized i.e. sent within the same access network to the local registrar agent due to the hierarchical architecture of AMP. Only when the MH moves to another access network will the binding update sent to the home and correspondent registrar agents. In MIP, in contrast, requires binding updates to be sent to the home and correspondent agents every time the *MH* changes location to a new subnet (globalised update). Hence, the signaling overhead is much higher than in AMP due to the flat architecture in MIP. In addition, binding updates in MIP require several procedures such as DAD, RR, HoT, etc. and these add to the overall latency and processing costs.

The binding refresh cost is a function of the binding lifetime periods at the home, visited and correspondent networks, and in this analysis, it is assumed that these periods are the same. Figure 9 shows the effect of binding lifetime period on the binding refresh cost. Typically, if the binding lifetime period is small, then frequent binding updates need to be sent to refresh the mappings, and this induces additional signaling overhead. Hence, the binding refresh cost is high for short lifetime periods, and small for longer lifetime periods since less binding refresh messages need to be sent. The binding refresh cost is constant during two lifetime period intervals i.e. [0.16, 0.24] and [0.26, 0.30]. During the first interval, the *MH* moves to the adjacent access network before the new binding refresh message takes effect. In the second interval, the average residence time of the *MH* is lower than the binding lifetime period, thus, no binding refresh message is sent resulting in a null value of the binding refresh cost. The binding refresh cost in AMP is slightly lower than the binding refresh cost of MIP since the binding refresh messages are only exchanged between network entities (peer registrars) and the distance is less compared to MIP.

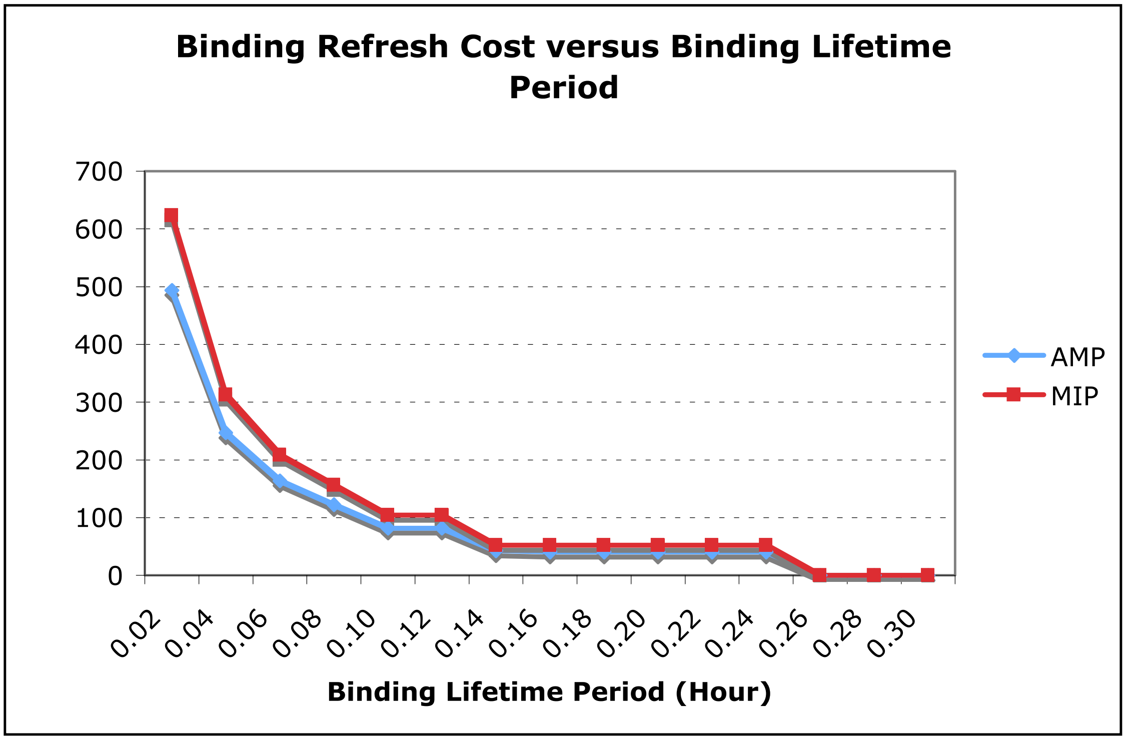


**Figure 9**: Effect of binding lifetime period on binding refresh cost

The cost for packet delivery against the packet arrival rate is shown in Figure 10, and generally, the higher the packet arrival rate, the higher is the packet delivery cost. As depicted in the figure, the packet delivery cost increases significantly with the increase in packet arrival rate in MIP, while in the AMP architecture, the increase rate is linear. The AMP architecture outperforms MIP considerably for a number of reasons – hierarchical architecture with localized signaling for intra-network movement, lower packet loss (if at all) since packets are forwarded and buffered to the *MH* in the next location, and faster handover operations since additional signaling for DAD, RR, HoT are not required in AMP. Thus, AMP would be well suited for real-time applications. However, it must be noted that the effect of buffering my render certain real-time packets useless if the size of the buffer is too large. As such, a more efficient buffering scheme is needed in AMP to support real-time applications with certain timing constraints.

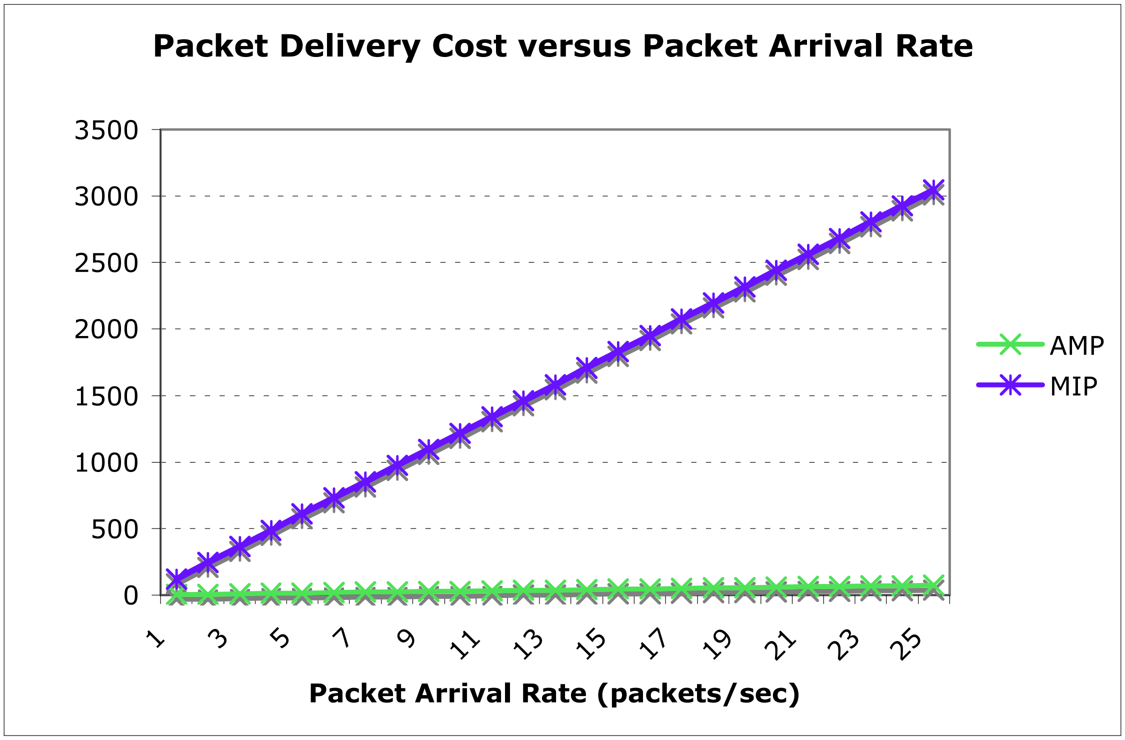


**Figure 10**: Effect of packet arrival rate on packet delivery cost

Figure 11 depicts the handover delay against the wireless link delay. Again, the AMP architecture has a much lower delay than MIP. Generally, the main overheads in MIP increases the cost of handovers especially since the architecture is flat and does not differentiate between local and global movement. In AMP, handovers are supported through localized procedures for intra-network movement and only global handover operations and message exchanges are made when the *MH* moves to another network. Again, this suggests that AMP would be suited to loss-intolerant applications with timing constraints.

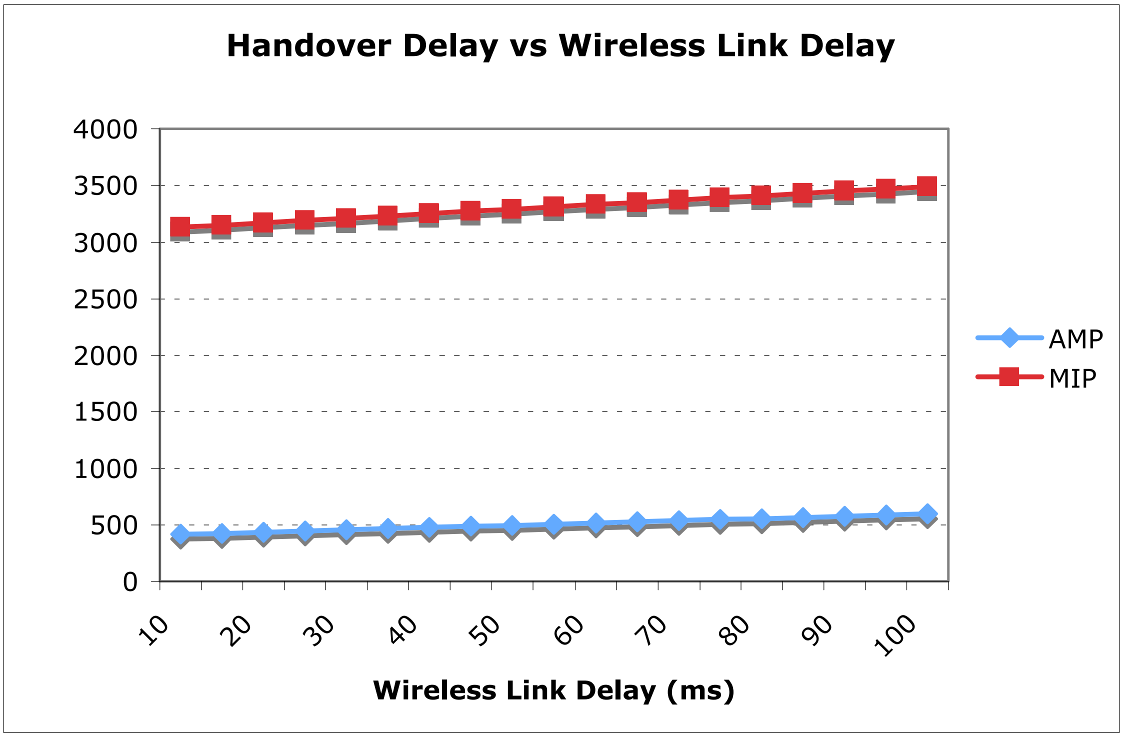


**Figure 11**: Effect of wireless link delay on handover delay

## 6. CONCLUSION AND FUTURE WORK

In this paper, the performance of the AMP architecture and protocol were evaluated using derived analytical models. The key considerations are mobile Quality of Service (mQoS) parameters such as signaling traffic cost, handoff latency and packet delivery cost. Overall, the proposed AMP architecture outperforms Mobile IP – this is mainly due to the fact that the AMP architecture is hierarchical and combines intelligence in the network to perform mobility management via agents. Hence, latencies attributed to encapsulation, tunneling and packet re-routing have been removed, and delays from location registration procedures are significantly reduced. In addition, the architecture supports packet buffering and this reduces packet loss when moving from one subnet to another. However, the size of buffers needs to be carefully weighted against timing considerations for real-time applications. As a future work, the AMP architecture is to be implemented via simulation and its performance compared against Mobile IP via simulated models.

## 7. REFERENCES

1. Hassan, W. H., Hassan, A., Fisal, N., “A Tracking Mechanism Using Agents For Personal Mobility In IP Networks”, *Proc. of the 6th IEEE International Conference on Industrial Informatics (INDIN 2008)*, Daejeong, Korea, July 13-16, 2008.
2. Baumann, F. V., Niemegeers, I. G., “An Evaluation of Location Management Procedures”, *Proc. 3rd Annual International Conference in Universal Personal Communication* (UPC’94), pp. 359-364, September/October 1994.
3. Wang, K., Huey, J. , “A Cost Effective Distributed Location Management Strategy for Wireless Networks”, *Wireless Networks*, Vol. 5, Issue 4, pp.287 -297, Kluwer Academic Publishers, 1999.
4. McNair, J., Akyildiz, I. F., and Bender, M., “Handoffs for Real-Time Traffic in Mobile IP Version 6 Networks”, *Proc. IEEE GLOBECOM*, pp. 3463-3467, Vol. 6, November 2001.

5. Pack S., and Choi Y., “Performance Analysis of Fast Handovers in Mobile IPv6 Networks”, *Proceedings of IFIP Personal Wireless Communications*, LNCS, vol. 2775, pp. 675 -695, September 2003.
6. Xiao Y., Pan Y., Li J, “Design an Analysis of Location Management for 3G Cellular Networks”, *IEEE Transactions on Parallel and Distributed Systems*, Vol. 15, No. 4, April 2004.
7. Makaya, C., Pierre S., “An Analytical Framework for Performance Evaluation of IPv6-Based Mobility Management Protocols”, *IEEE Transactions on Wireless Communications*, Vol. 7, No.3, pp. 972 – 983, March 2008.
8. Fang, Y., “Movement-based Mobility Management and Trade Off Analysis for Wireless Mobile Networks”, *IEEE Trans. Computers*, Vol. 52, No. 6, pp. 791-803, June 2003.
9. Jain R., Lin Y., Lo C., and Mohan S., “A Caching Strategy to Reduce Network Impacts of PCS”, *IEEE Journal on Selected Areas in Communications*, Vol. 12, No. 8, October 1994.
10. Lai W.K and Chiu, J. C., “Improving Handoff Performance in Wireless Overlay Networks by Switching Between Two-Layer IPv6 and One-Layer IPv6 addressing”, IEEE Journal on Selected Areas in Communications, Vol. 23, No. 11, November 2005.
11. Koodli R. and Perkins C. E., “Fast Handovers and Context Transfers in Mobile Networks”, ACM Mobile Computing & Communication Review, Vol. 31, No. 5, pp. 37-47, October 2001.
12. McNair, J., Akyildiz, I. F., and Bender, M., “An Inter-System Handoff Technique for IMT-2000 System”, *Proc. INFOCOMM*, March 2000.
13. Xie, J and Akyildiz, J., “A Novel Distributed Dynamic Location Management Scheme for Minimizing Signaling Costs in Mobile IP”, *IEEE Transactions on Mobile Computing*, Vol. 1, No. 3, July-September 2002.
14. Lai W. K., and Chiu J. C., “Improving Handoff Performance in Wireless Overlay Networks by Switching Between Two-Layer IPv6 and One-Layer IPv6 Addressing, *IEEE Journal in Selected Areas of Communication*, vol. 23, no. 11, pp. 2129-2137, November 2005.
15. Makaya, C., Pierre S., “An Architecture for Seamless Mobility Support in IP-based Next-Generation Wireless Networks”, *IEEE Transactions on Vehicular Technology*, Vol. X, No.X, June 2007.
16. Akyildiz, I.F., Mohanty, S., Jiang, X., “A Ubiquitous Mobile Communication Architecture for Next-Generation Heterogeneous Wireless Systems”, *IEEE Communications Magazine*, Vol. 43, Issue 6, pp. S29 – S36, June 2005.
17. Akyildiz I. A., McNair, J., Ho, J.S.M., Uzunalioglu, H., Wang, W., “Mobility Management in Next-Generation Wireless Systems”, *Proceedings of IEEE*, Vol. 87, No. 8, pp. 1347 – 1384, August 1999.
18. Akyildiz, I. F., and Wang, W., “A Dynamic Location Management Scheme for Next Generation Multitier PCS Systems”, *IEEE Transactions on Wireless Communications*, Vol.1, No. 1, January 2002, pp. 178-189.
19. Akyildiz, I. F., Xie, J. & Mohanty, S. “A Survey on Mobility Management in Next Generation All-IP Based Wireless Systems”, *IEEE Wireless, Communications,*Vol.11, no.4, pp.16-28, August 2004.
20. Johnson D., Perkins C. and Arkko J., “Mobility Support in IPv6”. RFC 3775, June 2004
22. Mohanty. S., Akyildiz F., “Performance Analysis of Handoff Techniques Based on Mobile IP, TCP-Migrate, and SIP”, *IEEE Transactions on Mobile Computing,* vol .I, No.7, July 2007.