Alexandria Engineering Journal

A decision strategy for the fast handoff protocol based on movement tracking of mobile nodes

Hussein H. Aly and Mustafa Y. El-Nainay
Computer and Systems Eng. Dept., Alexandria University, Alexandria, Egypt
alyh@computer.org, ymustafa@vt.edu

In this paper, a novel handoff decision strategy is proposed. The proposed handoff procedure utilizes the movement information provided by the location tracking devices and information about IP cells topology to predetermine the next Foreign Agent (FA) and the handoff decision time, "where and when to handoff". The proposed handoff procedure complements the mechanism of Fast Handoff described in and Simultaneous Binding described in with the movement tracking of MNs to provide a seamless handoff and decouples layer3 handoff from layer2 handoff. Two approaches are presented to develop the proposed handoff procedure. The first approach is distributed and mobile-controlled (network-assisted) in which the FA provides the MN with the neighbor FAs information and the MN is responsible of taking the handoff decision. On the other hand, the second approach is centralized and network-controlled (mobile-assisted) in which the MN continuously provides its current FA with its movement information and the FA takes the decision of the handoff. The proposed handoff procedure is evaluated through simulation and compared to current handoff methods. Moreover, the overhead of the proposed handoff procedure is discussed through a cost analysis. The simulation results show that the proposed handoff procedure outperforms other handoff procedures with a little increase in signaling overhead. Moreover, the proposed handoff procedure gives layer3 handoff latency similar to that of layer2 handoff.

في القريب العاجل سيكون مصاحبا لكل وحدة متحركة جهاز لتحديد مكان الوحدة. و ذلك بسبب تناقص سعر التكنولوجيا و جاذبية الخدمات المترتبة على تحديد المكان. بالإضافة إلى ذلك فإن تزايد الطلب على تطبيقات الوسائط المتعددة و تطبيقات الوقت الفعلي يحتاج إلى اتصالات عالية الإعتمادية و عدم فقدان اى رسالة (أو عدد قليل) من الرسائل المنقلة. و بالتالي فإن الوحدات المتحركة تحتاج لبروتوكول مناولة سريع للتنقل بين خلايا الشبكة حتى لا يؤثر على كفاءة الاتصالات. في هذا البحث نقترح استراتيجية جديدة لاتخاذ القرار لبروتركول المناولة السريعة اعتمادا على المعلومات المتاحة من أجهزة تتبع حركة الوحدات المتحركة. كما تم اقتراح طريقتين لإنشاء هذا البروتوكول. الطريقة الأولى موزعة و الذي يكون فيه التحكم في اتخاذ قرار المناولة مسئولية الوحدات المتحركة. أما الطريقة الثانية فهي مركزية و يعني أن اتخاذ قرار المناولة مسئولية الشبكة.

Keywords: Handoff, Fast handoff, Movement tracking, Mobile IP, Mobile node

1. Introduction

Handoff is one of the essential means to guarantee the user mobility in a mobile communications network, where the Mobile Node (MN) can move around. Maintaining the traffic connection with a MN is made possible with the help of Mobile IP.

Mobile IP (MIP) [4] describes a global mobility solution. In Internet (IP) environments, when a MN moves and attaches itself to another network, it needs to obtain a new IP address. This changing of the IP address requires all existing IP connections to the MN be terminated and then re-connected. This is necessary as the IP routing mechanisms rely on the topological information embedded in the IP address to deliver the data to the correct end-point. MIP overcomes this by introducing a level of indirection at the network IP layer.

This indirection is provided with the use of network agents and does not require any modification to the existing routers or end Correspondent Nodes (CNs). With MIP, each MN is identified by a static home network address from its home network, regardless of the point of attachment. While a MN is away from its home network, it updates a special entity, a Home Agent (HA), with information about its current IP address. The HA intercepts any packets destined to the MN,

Alexandria Engineering Journal, Vol. 46 (2007), No. 5, 711-732
© Faculty of Engineering, Alexandria University, Egypt.

and tunnels them to the MN's current location. Thus, it is necessary for a MN to register its location at the HA. The time taken for this registration process combined with the time taken for a MN to configure a new network care-of address in the visiting network, amounts to the overall handoff latency. Thus the handoff latency in MIP is primarily due to two procedures, namely, the address resolution and the (home) network registration.

There have been numerous proposals for minimizing the handoff latency of MIP. These can be broadly classified into two groups. The first group aims to reduce the network registration time by using a hierarchical network management structure while the second group attempts to reduce the address resolution time through address pre-configuration. The former is generally referred to as hierarchical handoff and the latter as fast handoff or low latency handoff. IETF drafts [1 and 2] incorporate the concepts of hierarchical and fast handoff mechanisms in the IPv6 network, based on MIPv6 [4]. However, although it has been shown that the combined use of hierarchical handoff and fast handoff improves the performance, it is nonetheless not sufficient in providing a packet lossless handoff environment at IP layer.

In this paper, a novel handoff decision strategy is proposed. The proposed handoff procedure utilizes the movement information provided by the location tracking devices and information about IP cells topology to predetermine the next Foreign Agent (FA) and the handoff decision time, "where and when to handoff". The proposed handoff procedure complements the mechanism of Fast Handoff described in [2] and Simultaneous Binding described in [3] with the movement tracking of MNs to provide a seamless handoff and decouples layer 3 handoff from layer 2 handoff.

Two approaches are presented to develop the proposed procedure. The first approach is distributed and mobile-controlled (network-assisted) in which the FA provides the MN with the neighbor FAs information and the MN is responsible of taking the handoff decision. On the other hand, the second approach is centralized and network-controlled (mobile-assisted) in which the MN continuously provides its current FA with its movement information and the FA takes the decision of the handoff. The rest of this paper is organized as follows: section 2 surveys related works. In section 3, the proposed handoff procedure is described in details with two alternative implementations. In section 4, the simulation model used for evaluating the performance of the proposed handoff procedure is described. Section 5 presents the performance evaluation of the proposed handoff procedure against previous works. Finally, section 6 concludes the paper and discusses the suggested future work.

2. Related works

The conventional MIP handoff procedure suffers from many problems. These problems can be categorized to the following categories:

- *Handoff Latency:* the latency of the conventional MIPv6 handoff procedure is mainly due to two reasons:

1. IP connectivity latency which is divided to the latency of Movement Detection and the latency of Address Configuration.
2. Registration and binding update latency which is the time taken by the MN to notify its new CoA to its HA and CNs.

- *Packets losses:* as a result of changing the point of attachment and the handoff latency, packets destined to the MN in the current connections may get lost during the period of the handoff. This problem increases with the real-time sensitive applications and multimedia applications.
- *Handoff decision time*: the third reason that contributes in the handoff latency is the vague of the handoff decision time "timing ambiguity problem" [3]

Towards smooth handoff, numerous methods have been proposed to tackle the conventional MIP handoff procedure problems. These methods can be classified into the following classes:

1. Methods try to minimize the Home Registration and Binding Updates latency through the "hierarchical structures".
2. Methods try to minimize the address reconfiguration latency through what is called "low latency" or "fast handoffs".

3. Methods try to solve the problem of movement detection or handoff decision time "where and when to handoff".

2.1. Hierarchical structures and protocols

Hierarchical schemes separate mobility management into micro mobility (intra-domain) and macro mobility (inter-domain). They introduce a Mobility Routing Point (MRP) [8] that separates micro from macro mobility. The MRP entity is normally placed at edges of a network, above a set of access/edge routers which constitute the MRP's network domain. The MRP intercepts all packets on behalf of the MN it serves and redirects them to the MN. This enables MNs, which move between access networks that are within the same MRP network domain, to register with the MRP, thus avoiding potential lengthy round-trip delay associated with registration to its home agent. This type of intra-domain mobility is managed by IP Micro-mobility management protocols, such as [9 and 10], while inter-domain or macro mobility is managed using MIP or using Inter-domain Handoff techniques [11].

In the context of Hierarchical Mobile IPv6 (HMIPv6), the Mobility Routing Point is equivalent to the MAP (Mobility Anchor Point) entity and the protocol for micro and macro mobility is achieved using the features of MIPv6. The MAP in HMIPv6 intercepts all packets on behalf of the MN it serves and tunnels them to the MN's on-link care-of address (LCoA).

When a MN moves into a new MAP domain, it acquires a regional address (RCoA) and an on-link address (LCoA). In the simplistic case, the address of the MAP entity is used as the RCoA, while the LCoA address is formed using the Stateless Address Autoconfiguration [5] protocol. Other methods of obtaining the RCoA and LCoA are described in [1]. After obtaining these addresses, the MN sends a Binding Update (*BU*) to the MAP, which will bind the MN's RCoA to the LCoA. If successful, the MAP will return a Binding Acknowledgement (*BAck*) to the MN indicating a successful binding (registration). In addition, the MN must also register its new RCoA with its home agent by sending another *BU* that specifies the binding between its home address and the RCoA, i.e. the MAP's address. When the MN moves to a new access router within the same MAP domain, it simply acquires an LCoA, and updates only the MAP binding.

Hierarchical structures reduce the latency of the Binding Updates. However, they do not address the other handoff problems. Moreover, the Router Advertisement message has to be extended to include the MAPs' addresses.

2.2. Low latency and fast handoffs methods

Low latency address configuration is about configuring an address for the MN in a network that it is likely to move to, before it moves. The Low Latency handoff proposal [12] describes two methods of achieving this, namely *Pre-registration* and *Post-registration.*

With pre-registration handoff, the MN is assisted by the network to perform L3 (layer-3) handoff *before* it completes the L2 (layer-2) handoff. It uses L2 'triggers', which arises as a result of beaconing signals from the network the MN is about to move to, to initiate an IP layer (L3) handoff. Its design however, diverges from the clean separation of L2 and L3 of the base MIPv4 scheme. With Post-registration handoff, L2 triggers are used to setup a temporary bi-directional tunnel between the old FA and the new FA. This allows the MN to continue using its old FA while performing the registration at the same or later time. A combined method is also possible where, if the Pre-registration does not complete in time, the old FA forwards traffic to the new FA using the Post-registration method in parallel.

Fast-Handoff [2] is the Low Latency handoff equivalent for MIPv6 network. It is similar in concept to the combined method described prior. The basic operation of the Fast-handoff is illustrated in the next fig.1. Fast-handoff introduces seven additional message types for use between access routers (foreign agents) and the MN. An access router is the last router between the wired network and the wireless network where the MN is situated. These seven messages are: Router Solicitation for Proxy (*RtSolPr*), Proxy Router Advertisement (*PrRtAdv*), Handoff Initiation (*HI*), Handoff Acknowledgement (*HAck*), Fast

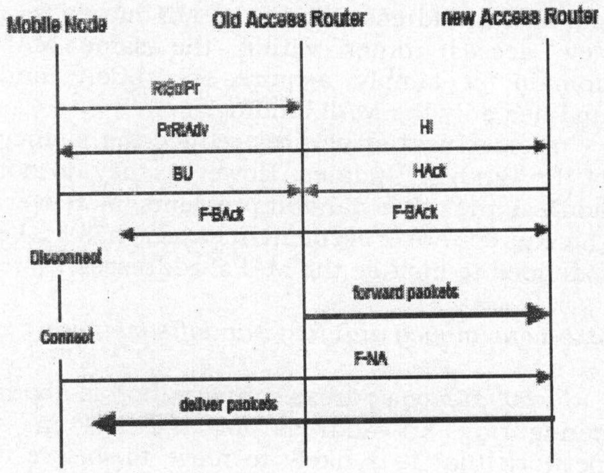


Fig. 1. Fast handoff procedure message interaction.

Binding Acknowledgement (*F-BAck*), Fast Binding Update (*F-BU*) and Fast Neighbor Advertisement (*F-NA*). In addition, the old Access Router (oAR) is defined as the router to which the MN is currently attached, and the new Access Router (nAR) as the router to which the MN is about to move to. Moreover, the Fast Handoff Operation consists of three phases: Handoff Initiation, Tunnel Establishment, and Packet Forwarding.

The *handoff initiation* is started by the L2 trigger based on certain policy rule (unspecified by IETF). This is done by the MN sending a Router Solicitation Proxy (RtSolPr) message to the PAR (Previous Access Router) indicating that it wishes to perform a fast-handoff to a new attachment point. The *RtSolPr* contains the link-layer address of the new attachment point, which is derived form the NAR's (New Access Router) beacon messages. The MN will receive, in response, a Proxy Router Advertisement (PrRtAdv) message from the PAR, with a set of possible responses indicating that the point of attachment is i) unknown, ii) known but connected through the same access router or iii) is known and specifies the network prefix that the MN should use in forming the new CoA. Subsequently, the MN sends a Fast Binding Update (*F-BU*) to the PAR using its newly formed CoA based on the prior *PrRtAdv* response, as the last message before the handoff is executed. The MN receives a Fast Binding Acknowledgement (*F-BAck*) either via the PAR or the NAR indicating a successful binding.

The *tunnel establishment* phase creates a tunnel between the NAR and the PAR. To establish a tunnel, the PAR sends a Handoff Initiation (HI) message (containing the MN's requesting CoA and the MN's current CoA) to the NAR. In response, the PAR receives a Handoff Acknowledgement (HAck) from the NAR. If the new CoA is accepted by the NAR, the PAR sets up a temporary tunnel to the new CoA. Otherwise, the PAR tunnels packets destined for the MN to the NAR, which will take care of forwarding packets to the MN temporarily.

Finally, the *packet forwarding* phase is performed to smoothen the handoff until subsequent registration by the MN to the home agent is completed. The PAR interacts with the NAR to facilitate the forwarding of packets between them, through the previously established tunnel. The initiation of the forwarding is based on an 'anticipation timing interval' heuristic, that is, the network anticipates as to when a MN is likely to handoff and therefore infers the appropriate packet forwarding moment based on the anticipation timing interval. Such an interval is however extremely difficult to generalize, and forwarding too early or too late will result in packet losses, negating the purpose of packet forwarding. Once arriving at the new access network, the MN sends the Fast Neighbor Advertisement (F-NA) message to initiate the flow of packets (to itself) from the NAR.

Low latency and fast handoff methods reduce the latency of address configuration. However, they do not address the other handoff problems.

HMIPv6 with Fast-handoff [1] is another attempt to further reduce the overall handoff latency from what Fast-handoff can offer alone. By combining HMIPv6 with Fast-handoff, latency due to i) address configuration and ii) the subsequent home network/agent registration, can both be reduced.

The MAP can be viewed as the 'local home agent', and in most cases, it is located closer to the MN than the HA. Therefore, the signaling cost saved is the difference between the roundtrip time of the MN to the MAP and the roundtrip time between the MN to the HA, assuming that message processing time within a network node is insignificant in comparison. This combination requires minor modification to the standard HMIPv6 protocol and the Fast-handoff protocol, i.e., relocating the forwarding anchor point from the PAR to the MAP as outlined in [1].

An alternative to the packet forwarding scheme (in Fast Handoff) has also been proposed, namely, the *Simultaneous Bindings* framework [3]. It proposes to reduce packet losses at the MN by n-casting packets for a short period to the MN's current location and to n-other locations where the MN is expected to move to. The n-casting can be carried out by the PAR, the MAP or the HA. The Simultaneous Bindings scheme recognizes the problem of not knowing when the MN is likely to move "the timing ambiguity" and attempts to remove it by 'careful' packet duplication to multiple access networks. It also claims to be able to address the problem associated with ping-pong movement of MNs between two access routers by this packet duplication process, as it is not necessary to re-configure the MN's CoA during ping-pong movement (rapid back and forth movement between two access routers/points).

2.3. Movement detection algorithms

There are two types of Move Detection algorithms, namely Advertisement [6] and Hint Based algorithms [7]. The first rely on periodic broadcasts from MIP mobility agents while the latter require information from the link-layer, termed as hints.

2.3.1. Advertisement based algorithms

For this kind of algorithms, all mobility agents must periodically broadcast advertisements. A MN discovers agents by receiving their advertisements. Moreover, a MN determines its position by evaluating these advertisements. For example, loss of contact with a "discovered" agent for 3 advertisement periods (advertisement/agent lifetime) is perceived as an indication of movement out of a network. In [4] the framework for advertisement based Move Detection algorithms is presented by defining a range of alternative reactions to the events of agent discovery and expiration. Currently, the two most distinct algorithms are the Lazy and Eager Cell Switching Move Detection algorithms [6] that manage opposite behaviors. The aim of Lazy Cell Switching (LCS) is to avoid MIP handoffs until they are absolutely necessary. In LCS, MIP handoff is performed only through expiration of the serving mobility agent or link disruption. This indicates that LCS is always slow to adapt to mobility. While the Eager Cell Switching (ECS) algorithm tends to function in a way opposite to that of LCS. It assumes frequent location changes and therefore pursues immediate handoffs upon discovering a mobility agent.

2.3.2. Hint based algorithms

The performance of Advertisement Based algorithms is directly dependent on the advertisement rate, i.e., optimum performance requires higher rates. However, for bandwidth efficiency purposes, [4] restricts the shortest advertisement period to a minimum of 1 second. [13] has shown for agent advertisements with a minimum size that Move Detection performance may be improved by reducing the advertisement period to 100 ms without a major impact to efficiency. Meanwhile, in hierarchical schemes such as [14] where FAs are required to include in their advertisements all mobility agents above them in the hierarchy branch, advertisements can reach a significant size.

As advertisement periods take on smaller values that lead to optimum Move Detection performance, the link efficiency drops significantly. In order to avoid this trade-off, [7] introduced the Hinted Cell Switching Move Detection algorithm. It suggested that by overruling layer independence and establishing a communication between MIP and lower layers for the communication of events such as link-layer hand-offs, optimum Move Detection performance could be achieved without any effect to efficiency. Hinted Cell Switching (HCS) bases its functionality on link-layer information, termed as hints, with the purpose of indicating when a link-layer hand-off occurs. The receipt of a link-layer hint will cause a MN to broadcast an agent solicitation that in turn forces all adjacent mobility agents to respond with a unicast agent advertisement. As such, the MN is in the position to promptly determine the identity of the local mobility agents regardless of the advertisement period. Fast Hinted Cell Switching (FHCS) expects to determine the identity of the local mobility agents through link-layer hints by extending the amount of information communicated between MIP and the link-layer; thus, avoiding the need for agent discovery and selection.

3. The proposed handoff procedure

3.1. IP cells topology

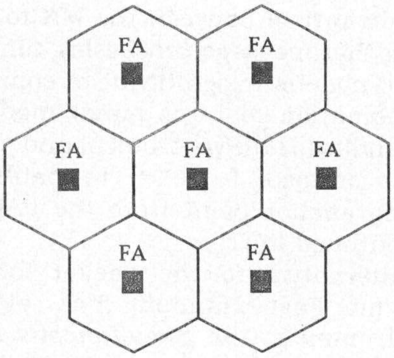
In this section, some assumptions are made on the Access Points (APs) and Foreign Agents (FAs) in order to simplify the description of the proposed handoff procedure in next sections. Later in section 3.5, these assumptions are relaxed.

The following assumptions are considered:

1. The Access Points (APs) in the network are of the same capabilities. This means that every AP covers the same area.
2. Each AP is connected to only one Foreign Agent (FA).
3. Each FA is connected to only one AP or is connected to the same number of APs that are of the same topology.

The above assumptions result in IP cells of homogeneous shapes. For example, each IP cell may form a hexagon, see figure (which is a widely accepted approximation for the IP cells [11, 17-18]). Each IP cell has the same number of neighbors. Moreover, the boundary between two neighbor IP cells is approximated to be a line segment that can be represented by its two end points. The physical location of the end points of the line segments can be calculated based on the positions and coverage areas of the APs. Each of These line segments represents the critical boundary between two IP cells that separates the coverage areas of the two FAs of these IP cells. In other words, if the MN crosses this boundary, an IP handoff must be performed between these two FAs.

Using the boundary information of the IP cells in conjunction with the movement information of the MNs, we can estimate where and when the MN will soon handoff.



Next subsections describe the proposed handoff procedure in details.

3.2. The proposed handoff procedure overview

3.2.1. Overview

The general idea of the proposed handoff procedure is to keep track of the MNs movement information either in the MNs themselves or in their FAs. This movement information is given by the location tracking devices coupled with the MNs. Using this information together with the IP cells topology information, the following fields are calculated:

- Which boundary the MN will cross next. "The expected next FA"
- The time when the MN will cross this boundary. "The expected Time to Cross the next Boundary (TCB)".

Before crossing the boundary with a sufficient time HANDOFF_THRESHOLD_TIME (a system parameter that is between 300ms to several seconds [19]; depending on the number of current MN active connections and the handoff procedure, see the next subsection), the MN or its FA initiates the handoff procedure.

In this section, two alternatives are proposed to design and implement the proposed handoff protocol. One of them is distributed (Mobile controlled- Network assisted) in which each MN has the responsibility of taking its own handoff decisions based on its movement information. The other alternative is centralized (Network controlled- Mobile assisted) in which the MN sends its movement information to its FA and the FA has the responsibility of taking the handoff decision.

In both cases a common data structure is maintained in the FAs namely table" (Neighbor Foreign Agents table) (NFA). This table is similar to the CAR table in CARD protocol [20] except that the boundary information is added. Each FA maintains its own NFA table that contains the following fields for each neighbor FA:

- Boundary Line Segment: The end points of the line segment that separates the coverage area of the current IP cell from the IP cell of that FA.
- Link Layer Address: The link layer address of that FA is needed to save the time of Neighbor Discovery process [4].
- IP Address: The IP Address of that FA.
- Prefix Information: The prefix of the new Subnet of that FA. This field is used for address auto-configuration. It may or may not be the same as the prefix of the IP Address of the FA.

The NFA table is supposed to be configured statically. However, values of some fields may change (i.e. IP address, Subnet Prefix ...). In order to dynamically discover these changes, the FAs have to exchange this information through periodical messages (or on demand). For example, these fields may be added to the Router Advertisements [4]. However, this extension is out of our scope.

In the next two subsections, the two alternatives are discussed in details.

3.2.2. Handoff_threshold_time

HANDOFF_THRESHOLD_TIME is the amount of time needed to complete the handoff procedure before the MN crosses the boundary of its current IP cell to one of its neighbors. The value of HANDOFF_THRESHOLD_TIME should be large enough to include the followings:

- New CoA Registration Time: the time required to formulate and validate the NCoA of the MN. This includes the signaling (HI, HAck) between the current FA and the next one to validate the NCoA (see next section).
- Home Registration and Binding Updates Time: the time required to register the NCoA with the HA and CNs of the MN. This period equals the maximum of the round trip time (BU, Acknowledgment) from the MN to its HA, and half the round trip time (BUs) from the MN to each of its CNs.

Clearly, each MN has a different value of HANDOFF_THRESHOLD_TIME. It is cumbersome to calculate the exact value of HANDOFF_THRESHOLD_TIME. However, since the MN does not have to complete all CNs binding updates before the handoff to the new IP cell, the proposed handoff procedure can depend on an average value of HANDOFF_THRESHOLD_TIME. Moreover, the HANDOFF_THRESHOLD_TIME is considered as a simulation parameter in order to study its effect on the proposed handoff procedure.

3.3. Distributed approach – mobile controlled

In this approach, each MN has to keep track of its movement information and has the responsibility to decide where and when to handoff. Since each MN takes the responsibility of its own handoff decisions, this approach is considered distributed and mobile controlled that relieves FAs from maintaining MNs' movement status and reduces the overall complexity in the network (recommended as a handoff decision strategy for 4th generation mobile networks [21]). However, this approach requires FAs to provide its registered MNs with the IP cell boundaries information (NFA table) and adds some complexity to MNs (see the cost analysis section in section 5).

The distributed approach consists of five phases:

1. Neighbor FAs Information Fetching.
2. Movement Tracking.
3. Handoff Initiation.
4. Home registration/ Binding Updates.
5. Handoff Completion.

Moreover, a recovery phase is needed to recover from "false handoffs". This is the case when a change in either MN direction or speed occurs after initiating the handoff.

3.3.1. Neighbor FAs information fetching

As stated in section 3.2, each FA has to maintain NFA table that contains its neighbor FAs information with its cell boundaries information. When a MN registers itself with a new FA, it sends an "NFA request" message to request the NFA table information from its FA.

The FA responds to the NFA request with an "NFA reply" that contains the NFA table information. Upon receiving the NFA reply, The MN saves the NFA table in a local cache. In case of dynamic discovery of some fields of the NFA, the FA must notify its MNs with these changes.

If the size of the NFA reply message is large, the NFA reply can be divided into more than one message. In this case, the MN receives a number of messages with each one corresponding to one neighbor FA and so it constructs its own NFA table from the information of these messages altogether. Moreover, the NFA reply can be piggybacked with the packets destined to the MN in case that technique (piggybacking) is supported. In case the MN does not receive an NFA reply within a predefined amount of time NFA_REQUEST_RETRY seconds, it uses exponential backoff to retransmit the NFA request.

3.3.2. Movement tracking

While the MN is roaming inside the boundary of one IP cell, it keeps tracking of its movement information through its location tracking device in terms of position (x,y) (in other words latitude and longitude of its current position), movement linear equation (y=ax+b) parameters (a, b) "these parameters can be calculated using two consecutive positions", its direction (D) (either (N)orth or (S)outh) and its speed (S). In order to simplify the calculations, the center of the IP cell can be used as an origin.

Using these movement information in conjunction with the NFA table information acquired from the previous phase, the MN calculates and maintains the following two fields:

- *The expected next FA:* This field can be calculated by solving the MN movement linear equation (y= ax+b) with the line segment information of each FA in the NFA table. This gives us two possibilities/intersection points, with one cell in the north and the other in the south, which in turn reduced to only one using the MN direction information. This field answers the question "where to handoff".
- *The expected Time to Cross the Boundary (TCB):* This field is calculated by first calculating the distance (Dist) between the current MN position and the intersection point from the previous field calculations. Then the TCB is calculated (TCB = Dist/S). This field answers the question "when to handoff".

In case that the MN speed is changed, the next FA remains the same but TCB field must be recalculated. In case the direction is changed, then the next FA may be changed and TCB is changed so both fields must be recalculated.

3.3.3. Handoff initiation

In this phase, four messages from the fast handoff protocol are used. These messages are Handoff Initiation (HI), Handoff Acknowledgement (HAck), Fast Binding Update (F-BU) and Fast Binding Acknowledgement (F-BAck). Moreover, the concept of Simultaneous Binding [3] with Fast Handoff is used but with different processing.

Before TCB with a sufficient time "HANDOFF_THRESHOLD_TIME", the MN uses the information of the estimated next FA from the NFA table to formulate a new Care of Address (NCoA) through stateless address auto-configuration methods [5]. After formulating the NCoA, the MN sends a F-BU to its FA containing the NCoA, IP address and Link Layer address of the next FA, and the current value of TCB with the Simultaneous Binding flag (B) set [3]. The purpose of the F-BU is to allow the current FA to know that the MN will soon (after TCB) leave the current cell.

Upon receiving the F-BU message, the FA checks whether the proposed NCoA is valid or not by sending HI message to the next FA. The HI message contains the current CoA, NCoA, Link layer address of the MN and TCB. In response to HI message the next FA perform the following steps:

1. determines whether NCoA supplied in the HI message is a valid address for use. And in case the NCoA is invalid it assigns another NCoA to the MN.
2. allocates NCoA (either the proposed or the assigned one) for the MN.
3. sends a HAck message to the old FA (current) with the accepted NCoA.

Upon receiving the HAck message from the next FA, the current FA adds a new binding for the MN that maps the current CoA to the

NCoA received in the HAck with a lifetime value equals the lifetime granted by the next FA (in the HAck message). Thus, the arriving packets after TCB amount of time are forwarded to the MN's NCoA. This means that the FA keeps two bindings for the MN but uses the old one for a TCB seconds "the FA delivers the arriving packets to the current CoA", after that amount of time it uses the new one "it tunnels the arriving packets (if any) to the NCoA". This is different from the original Simultaneous Binding in which the packets are bi-casted to both the current and new CoA. In the proposed procedure, knowing TCB eliminates the need to overload the network with these duplicate packets. Following that, the current FA sends FBack message to the MN with the accepted NCoA and its granted life time.

3.3.4. Home registration/ binding updates

After receiving the FBack from its FA, the MN registers this NCoA with its HA [4] by sending Binding Update (BU) message to its HA. This message must have the Home Agent (H) bit set to indicate this as a home registration, and must have the acknowledgement (A) set to request an acknowledgment of the Binding Update from the HA [4]. The Simultaneous Binding flag (B) is set to inform the HA to keep both bindings (the current one and the new one) in its cache. Two binding lifetimes are sent in the BU message; the first lifetime is the valid time for the current CoA which is the current value of TCB, while the other one is the lifetime for the new CoA which is granted by the next FA.

Upon receiving the BU message with the B bit set, the HA:

1. changes the current binding lifetime to the first lifetime of the BU message.
2. adds a new binding for the MN with the NCoA and the second lifetime in the BU message.
3. sends a Binding Acknowledgment to the MN.
4. After that, the FA must forward any packets destined to the MN to its current CoA until the lifetime of this binding is expired then forward the packets to the NCoA. Again, the proposed procedure eliminates the need for bi-casting packets, which is used in the Simultaneous Binding [3].

Moreover, The MN sends Binding Updates to the Correspondent Nodes (CNs) with existing connections with it. The BU is the same as the above case except that the H flag is not set (it is not a home registration) and the A bit may or may not be set. The CN that receives the BU message from a MN processes it with the same manner as the above case.

3.3.5. Handoff completion

As soon as the MN crosses the boundary (TCB =0), it sends a unicast "Router Solicitation" to the next FA to ensure its reachability. Upon receiving "Router Advertisement", it sends an F-NA message [2] to announce itself to its new FA. Upon receiving F-NA, the new FA create a neighbor cache entry for the MN and set it to REACHABLE state [4] then it forwards any buffered packets to the MN. If the new FA is unreachable, the MN assumes that an error occurred and switches to the conventional MIP handoff. After the MN registers itself with the new FA, it starts the whole procedure again (starts requesting the new NFA table and so on).

3.3.6. Recovery from direction/ speed changes

If the speed and/or the direction of the MN are changed after initiating a handoff, the MN recalculates the next FA and TCB fields and performs the following recovery procedure:

- If the next FA is changed, it sends a HCancel (Handoff Cancellation) message to its FA/HA/CNs to cancel the Handoff process (delete the new binding) and continue registering the MN in its Binding Cache with the current binding. The current FA in turn forwards the HCancel to the "previous" next FA to inform it to release the current NCoA allocation. Following that, the MN restarts the current phase of the handoff procedure using the "new" next FA and TCB values.
- In case of the next FA is not changed (i.e. only MN speed is changed, but not stopped), the MN uses the new TCB value to send a TCB_Update message to its FA/HA/CNs to inform it with the new TCB value to change the lifetime of the current binding.

- In the first case (next FA is changed), if the new TCB value is less than HANDOFF_THRESHOLD_TIME, the MN switches to the ordinary fast handoff protocol.
- In the second case (next FA is not changed), if the new TCB value is greater than HANDOFF_THRESHOLD_TIME, the MN sends HCancel message to its FA/HA/CNs to cancel the current handoff. Later, when the TCB value is equal to HANDOFF_THRESHOLD_TIME, the MN starts the handoff procedure again with the new values of next FA and TCB.

3.4. Centralized approach – network controlled

In this approach the MNs provide their FAs with their movement information and the FAs have the responsibility of deciding where and when the MNs should handoff. Since each FA takes the responsibility of the handoff decisions instead of all the MNs registering with it, this approach is considered a centralized approach (all handoff decisions from a single network entity). Moreover, it is a network controlled- mobile assisted approach since the MNs sends their movement information to their FAs and FAs initiate the handoff. On contrary to the previous approach, the network topology and cell boundary information are transparent to MNs. In other words, the MNs have no idea about the IP cells.

The centralized approach consists of five phases:

1. MN Registration.
2. Movement Tracking.
3. Handoff Initiation.
4. Home Registration/ Binding Updates.
5. Handoff Completion.

In addition to the NFA table that is described in section 3.2.1, the FAs, in the current approach, have to maintain the MNs movement information in a new data structure; we call it (MI) Movement Information table. In this table, the FA maintains the following fields for each MN:

- Link Layer Address (LLA): Link layer address of the current link.
- Care of Address (CoA): The current CoA of the MN.
- Position (x,y): The last received position from the MN.
- Direction (D): This is a character that equals "N" for North or "S" for South.
- Movement Linear Equation Parameters (a,b): A tuple (a,b) of the movement linear equation ($y = ax + b$) that represents the MN movement.
- Speed (S): The speed of the MN in meter per second.
- The expected Next FA: The next FA is where the MN will handoff if the current direction remains without changing (and the MN does not stopped). This field is calculated by solving the linear equation of the MN movement with the line segments of the neighbor FAs in the NFA table to get the line segment that intersects with the MN (and on its direction). This field solves the problem of where to handoff. This field is a link to the NFA entry that represents the next FA.
- The expected Time to Cross the Boundary (TCB): The time before the MN crosses the boundary between the current FA and the expected Next FA. This field is used to initiate the handoff before the MN reaches the boundary by HANDOFF_THRESHOLD_TIME.

The initialization and maintenance of the above fields for each MN are described in the following subsections.

3.4.1. MN registration

When a MN registers itself with a new FA, it sends a MIR (Movement Information Registration) message to its FA containing the LLA, CoA, Direction, Speed, Movement linear equation parameters and current position.

Upon receiving the MIR message, the FA adds a new entry for the MN in the MI table with the information in the message. Moreover, it uses these information in conjunction with the information in the NFA table to calculate the next FA and TCB fields as described in the previous approach (section 3.3.2).

3.4.2. Movement tracking

After registering its movement information through sending MIR message, the MN keeps track of its direction and speed. If any or both of them are changed, it must notify its FA with this change to modify the MN information in its MI table.

If the change is in the MN speed only, then the next FA remains the same but the TCB must be recalculated. Thus, the MN sends a Speed Update (SU) message to its FA containing the new value of the MN speed and its current position. Upon receiving the SU message, the FA recalculates the TCB field using the new MN position and speed.

If the MN stops (its speed = 0), the MN sends a SU message to its FA as in the previous case. However, in this case the FA keeps the MN entry for a while then deletes it if no SU messages sent by the MN (no necessary for keep tracking of stationary MNs). Later, when the MN moves, it must send a MIR message to its FA.

If the change is in the MN direction (with or without change in speed), the MN sends a Direction Update (DU) message to its FA containing its new movement information (current position, direction, movement linear equation parameters and speed "if changed"). Upon receiving DU message, the FA updates the MN entry information in the MI table and recalculates both next FA and TCB fields. (even though in some cases the next FA may be not changed).

3.4.3. Handoff Initiation

When the TCB field of one MN equals HANDOFF_THRESHOLD_TIME, the FA initiates the handoff process. First the FA uses the information of the MN from the MI table with the information of the next FA from the NFA table to formulate a NCoA using a stateless address auto-configuration method [5]. Following that the FA sends a HI message to the next FA to validate the NCoA. The HI message contains the current CoA, NCoA, Link layer address of the MN and the current value of TCB.

In response to HI message the next FA perform the following steps:

1. determines whether NCoA supplied in the HI message is a valid address for use. And in case the NCoA is invalid it assigns another NCoA to the MN.
2. allocates NCoA (either the proposed or assigned one) for the MN.
3. sends a HAck message to the old FA (current) with the accepted NCoA.

Upon receiving the HAck message from the next FA, the current FA adds a new binding for the MN that maps the current CoA to the NCoA received in the HAck with a lifetime value equals the lifetime granted by the next FA (in the HAck message).

After validating the NCoA, the FA sends a HI message to the MN with the accepted NCoA, TCB and the life time of the NCoA. This message is equivalent to the FBack message in the previous approach. However, since the sender this time is the FA, we prefer to keep it with Handoff Initiation (HI) name.

3.4.4. Home registration/ binding updates

When a MN receives a HI message from its FA containing its NCoA, TCB and the granted lifetime of the NCoA, it starts registering the NCoA with its HA and its CNs. This phase is the same as the previous approach with no change except in the way it is triggered.

3.4.5. Handoff completion

Again this phase is the same as the previous approach. And again after the completion of the handoff procedure, the MN must register its movement information with its new FA through sending MIR message and the complete procedure starts again.

3.4.6. Recovery from direction/speed changes

As described in the movement tracking phase, if the speed and/or the direction of the MN are changed, the MN sends SU/DU message to its FA. If the FA receives such messages from a MN *after initiating a handoff* for this MN, a recovery procedure should be performed as follows:

- If the FA receives a SU message, it recalculates the TCB value of the MN's entry in the MI table based on the new information in the SU message. With the new TCB value, the FA sends a TCB_Update message to the MN/next FA which changes the previous TCB value. Moreover, the MN forwards this message to its HA/CNs. The HA/CNs update the current MN binding lifetime in their binding cache.
- If the FA receives a DU message, it recalculates both the next FA and TCB fields of the sender.
- If the next FA is changed, it sends a HCancel message to the MN/next FA. Upon receiving the HCancel message, the next FA releases the current allocation of the NCoA.

The MN ,that receives a HCancel message from its FA, cancels the handoff, keeps its current CoA and forwards the message to its HA/CNs to cancel the Handoff process (delete the new binding) and continue registering the MN in its Binding Cache with the current binding.

- If the next FA is not changed and only the TCB value is changed, the FA handles this case as the first one (only speed change).
- In case of the next FA is not changed, if the new TCB value is greater than HANDOFF_THRESHOLD_TIME, the FA cancels the handoff and restarts the handoff later, when the TCB value is equal to HANDOFF_THRESHOLD_TIME.
- In case of the next FA is changed, if the new TCB value is less than HANDOFF_THRESHOLD_TIME, the FA switches to the ordinary fast handoff protocol.

3.5. Assumptions relaxation

In section 3.1, some assumptions are made to simplify the description of the proposed algorithm. In the current section, the required changes to accommodate the general case, when one or more of those assumptions are violated, are discussed.

Violating assumptions 1 and/or 3 lead to IP cells of heterogeneous shapes. This means that each IP cell may have different number of neighbors. Subsequently, the size of NFA table may differ from one IP cell to another one. The solution of this problem is to divide the NFA reply message to a number of messages equals to the number of neighbors. In this case, each NFA reply message has the same size, no matter the number of neighbors.

Violating assumption number 2 leads to more than one FA for each AP. This means that each neighbor of an IP cell may have more than one FA. In this case the NFA table should be extended to include all FAs of each neighbor IP cell. Again, NFA reply message should be divided to a number of messages equals to the number of neighbor FAs. Moreover, if a MN wants to handoff to an IP cell that contains more than one FA, the MN may choose one of them according to preference, capabilities, load, services, .. (This information should be added to the NFA table fields).

4. Simulation model

The performance of the proposed handoff procedure and previous works is evaluated through extensive simulation using Network Simulator 2 "NS2" [22]. The goal of the simulation is to examine the effectiveness of the proposed handoff procedure in L3 (IP layer) handoff latency reduction, over reliable end-to-end communication, namely TCP. In particular, we are interested in examining the bulk data flow rather than the interactive data flow scenario, since bulk data flow is more prone to disruption during a handoff. Moreover, the performance of the proposed handoff procedure is compared to those of previous procedures; namely:

1. The conventional MIP handoff procedure [4].
2. The fast handoff procedure [2, 19].
3. The hierarchical mobility management [1, 19].
4. The hierarchical MIPv6 with fast handoff [1, 26].

Since both approaches proposed in the previous section (Mobile controlled and Network controlled) exhibit the same message exchange sequence from the handoff initiation point, there is no need to implement both approaches. In other words, the differences between the two approaches are in the signaling and processing cost before the handoff initiation. Thus, only the mobile controlled approach is implemented in NS2. Moreover, differences between the two approaches are discussed in the cost analysis in the next section.

4.1. Simulator extensions

In order to be compatible with the performance evaluation of similar works, the NS2 version 2.1b7a is used in the evaluation process. The standard NS2 is patched with NOAH [23], a wireless routing agent that (in contrast to DSDV, DSR, ...) only supports direct communication between access points and MNs. Moreover, it is patched with the MIP, hierarchical mobility and ordinary fast handoff extensions [24]. This is further extended with the implementation of the proposed handoff procedure.

Firstly, the new data structures that are used in the proposed handoff procedure are implemented. The NFA table is added to the original FA node model. Secondly, new and modified messages that are used in the proposed handoff procedure are added to the mobile and FA nodes. Finally, the required functionality in performing the proposed handoff procedure based on movement tracking is added to the MN.

Existing NS2 functions are used for the part of movement information extraction from the location tracking device; since they do the same functionalities.

4.2. Simulated network topology

The next fig. 2 illustrates the network topology used for the simulation experiments. This topology depicts a simplistic version of a typical MIP network topology. This topology is used extensively in MIP performance studies [15, 16, 19].

Both Corresponding Node (CN) and the HA are connected to an (dummy) intermediate node (N1) with 2 milliseconds (ms) delay, 100 Megabits/s (Mb/s) links. The current FA is connected to an intermediate node N3 while the next FA is connected to N4, all with 2ms delay, 1Mb/s links. Both N3 and N4 are connected to an intermediate node N2 with 2ms delay, 10Mb/s links. N1 is connected to N2 with a 100Mb/s link with 50ms delay. This simulates the distant home network. (In the hierarchical mobility management simulation, N2 is replaced by the MAP node).
All links use the Random Early Detection (RED) queue, except links from the intermediate nodes to the FAs (both current and next), which are Droptail (FIFO) queues. The wireless coverage area of the FAs is set to 40 meters in radius. The FAs are set to be 70 meters apart with free space environment in between; this reduces the complexity of results analysis, as we only need to consider signal interference. A Lucent WaveLan card running 802.11 protocols is simulated under NS2.

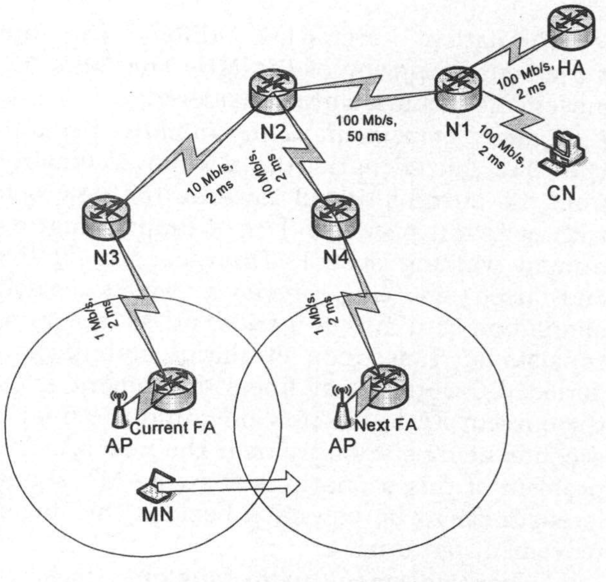


Fig. 2. Simulation network topology.

4.3. Simulation scenarios

Different simulation scenarios are used to evaluate the proposed handoff procedure against previous works. In all scenarios an *ns* TCP source agent is attached to the Corresponding Node (CN) and an *ns* TCP sink agent is attached to the MN (MN). The TCP packet size is set to 512 bytes and window size is 32. The MN is initially positioned near the current FA inside the current IP cell and starts to move towards the new FA in the new IP cell 6 seconds after the simulation starts. This is to enable the establishment of TCP communication and allowing it to stabilize, meaning; TCP is transferring data with a full window. The TCP Tahoe, which follows a 'go back-n model using accumulative positive acknowledgement with slow start, congestion avoidance and fast retransmission' model, was chosen as the default TCP flavor. An FTP session (bulk data transfer) between the MN and the CN is started 5 seconds after the simulation has started. The bulk FTP data traffic flow is from the CN to the MN. Finally, in all simulations, the simplistic case with only one MN initiating one single connection at a time is considered. Moreover, L2 handoff time is set to 200 ms [25] and address resolution time (time to validate the NCoA) is set to 100 ms [15].

Simulation scenarios differ in the movement scenarios of the MN. The following movement scenarios are considered:

- *Linear movement with human walking speed:* in this scenario, the MN moves linearly from the current IP cell towards the new one with constant speed of 1 m/s (approximating human walking speed). This scenario is the one used in the previous works. Each simulation run has a period of 80 seconds, consists of 5 seconds of initial stabilization period, 70 seconds of linear movement from the current FA to the new one, and the final 5 seconds being stationary near the new FA. The problem of this scenario is that the MN speed is assumed to be constant besides the linear movement pattern.
- *Linear movement with constant speed (> 1m/s):* in this scenario, the effect of MN speed on the MIP handoff latency is considered through simulation runs with different speed values 2m/s and 5m/s. Each simulation run consists of 5 seconds of initial stabilization period, time for the linear movement from the current FA to the new one, and the final 5 seconds being stationary near the new FA.
- *Ping pong movement with constant speed:* in this scenario, the MN moves linearly from the current IP cell towards the new one with a constant speed of 1 m/s. Each simulation run has a period of 80 seconds, consists of 5 seconds of initial stabilization period, the next 70 seconds (for the ping-pong case) are similar to that of the linear case, except that at 46 seconds into the simulation the MN reverses its direction of movement back towards the current FA. It reverses its direction every 2 seconds thereafter, until finally heading towards the new FA again, 52 seconds into the simulation, and the final 5 seconds being stationary near the new FA.

4.4. Simulation parameters

The following simulation parameters are varied in order to investigate their impacts on the simulated handoff procedures:

- *Moving speed:* the moving speed of the MN.
- *Round trip time:* the round trip time from the CN to the MN.
- *Handoff_threshold_time.*

The first simulation parameter simulates the degree of the MN mobility while the second parameter simulates the distance and traffic (propagation delay) in links between the CN and the MN.

The first parameter changes as described in the previous subsection (movement scenarios) while the second one is varied from 30 ms to 400 ms by tuning the propagation delays of the link from CN to N1 and the link from N1 to N2.

The default values of simulation parameters that are used in the simulation experiments are given in table 1

4.5. Performance metrics

Since the goal of the proposed handoff procedure is to minimize the effect of the mobile handoff on the existing communications, the following performance metrics are counted during the handoff period:

- *Overall Handoff Latency (msec):* the amount of time when the MN detaches the old network FA till the disrupted communication session is returned to normal operational state at the new network FA.
- *TCP Goodput (Kbytes/sec):* total amount of Kbytes received per second at the MN.
- *Retransmitted Packets:* the number of retransmitted packets due to handoff.
- *TCP cwnd:* the value of the TCP congestion window.

5. Performance evaluation

In this section, the simulation results of the proposed handoff procedure are presented in conjunction with the results of the other handoff procedures. First, the simulation scenarios described in the previous section are considered. These scenarios investigate the effect of different movement behavior of the

Table 1
The default values of simulation parameters

Moving speed	1 m/s.
Round trip time	112 ms.
Handoff_threshold_time	300 msec.

MN on the handoff process. Secondly, the effect of changing the Round Trip Time (RTT) between the MN and the CN is investigated. Finally, the effect of under/overestimating the HANDOFF_THRESHOLD_TIME is investigated. In all handoff procedures, the L3 handoff process starts with L2 handoff process. Moreover, in the fast, hierarchical and the HMIP-Fast procedures the handoff process starts when the MN receives the first advertisement message from the new FA. This gives the best results for these handoff procedures.

5.1. Linear movement with human walking speed

In this scenario, the simulation parameters are set to their default values. The simulation scenario is run for each of the simulated handoff procedures. The next figures show the results for the chosen performance metrics.

The previous fig. 3 illustrates the effect of the handoff process on the TCP connection. This figure shows the CN's TCP sending buffer. The following observations can be made from this figure:

- The L3 handoff process starts with L2 handoff at 40.5s. In this period, the TCP connection is interrupted by the handoff process. The sender (CN) does not receive acknowledgement (ack) messages from the MN.

Overall Handoff Latency

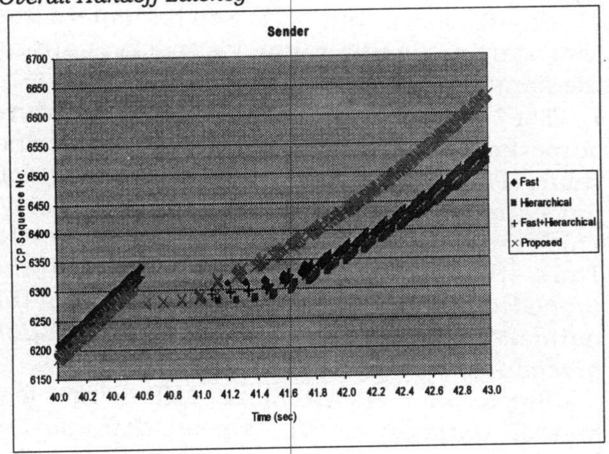


Fig. 3. TCP sequence no. Vs Time for CN's TCP sending buffer.

- In the conventional MIP handoff procedure, the time between the first retransmitted packet sent by the CN (received by the MN) and the first time this packet was sent (and not received by the MN) is 4306 ms. This period is used as a measure for the overall handoff delay. This period is because the conventional MIP procedure waits for 3 advertisement periods to consider the current FA unreachable. The MIP is not included in the above figure because the handoff starts at 48.4s and its results are out of order.
- In the fast handoff case, the handoff delay is 734 ms. This is the period of L2 handoff plus the time of completing the fast handoff. Although the fast handoff saves the time for obtaining the new CoA, the MN has to setup the tunnel between the old FA and the new one. Moreover, the MN has to update its binding at its HA and CN through sending BU messages. The binding update delay is proportional to the round trip time between the MN and the receiver of the BU message. This is the weakness of operating the fast handoff in the flat MIP environment. After the binding update process, the MN sends F-NA message to activate the data flow from the new FA. Even though the fast handoff sets a tunnel to forward the packets from the old to the new FAs, the MN is unable to receive these packets until it sends an F-NA to the new FA.
- In the Hierarchical case, the handoff delay is 730 ms. this is the period of L2 handoff plus the time of completing the handoff process. Although this case reduces the binding update time of the HA and CN to the time needed to update the MAP, the MN still has to obtain a new CoA and completes the handoff process at the new cell.
- The Hierarchical MIP with Fast Handoff (HMIP-Fast) approach gives better results with a handoff delay equals 642 ms. This is because the HMIP-Fast approach saves both the address resolution and binding update time.
- In former handoff procedures, the handoff delay drives the TCP source to reach zero window and stop transmitting. This means that all packets sent from the offered window are lost due to handoff. After the retransmission timer of the first packet of the current window is expired, the TCP source (CN)

enters a slow start mode with congestion avoidance and begins the retransmission of the lost packets. In this time, the MN receives the packet and sends ack message back to the CN.

- In contrast to the above procedures, the proposed handoff procedure results in handoff delay equals 205 ms. Moreover the TCP source does not reach zero window. This delay is mainly due to the L2 handoff (set to 200 ms) which means that the proposed handoff procedure gives a L3 handoff delay of the same order as L2 handoff delay. In this case, the packets sent by the CN during the handoff period are lost due to L2 handoff. After completing L2 handoff, the MN sends an F-NA message to its new FA to activate the data flow from the new FA. The MN receives out of sequence packets, and thus replies with ack messages containing the expected sequence number back to the CN. After the CN receives three (fast retransmission) of such ack messages, the CN enters the slow start mode and begin the retransmission of the lost packets. After the MN receives all of the lost packets, the MN sends ack message to acknowledge all of the received packets before the CN enters the slow start mode. Upon receiving this ack message, the CN stops the retransmission mode and starts the transmission from the new requested packet.

5.1.1. TCP Goodput

Table 2 shows the TCP Goodput in Kbytes per second for the TCP connection during the handoff period. The HMIP-Fast approach outperforms the fast handoff and the hierarchical one with a TCP goodput 63.5 Kbytes/sec. Moreover, the proposed handoff procedure outperforms the HMIP-Fast approach with a TCP goodput 77.5 Kbytes/sec.

5.1.2. Retransmitted packets

The previous table gives the number of retransmitted packets due to the handoff process. For all cases, except the proposed handoff procedure, the handoff delay is large enough to drive the TCP sender to reach a zero window. Thus, in all these cases, the number of retransmitted packets equals the window size (set to 32). The only exception is the proposed handoff procedure in which the number of retransmitted packets equals 17.

Table 2
Linear movement with human walking speed scenario

Handoff procedure	Handoff latency (ms)	TCP goodput (Kbytes/sec)	Number of retransmitted packets
Fast	734	60	32
Hierarchical	730	60.2	32
HMIP-fast	642	63.5	32
Proposed	205	77.5	17

TCP cwnd

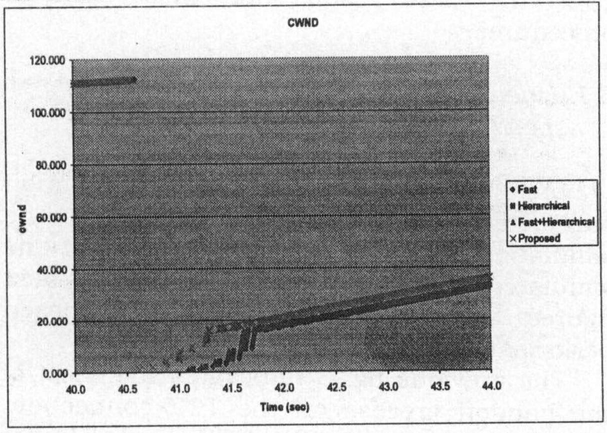


Fig. 4. Cwnd value Vs Time.

The previous fig. 4 illustrates the TCP congestion window (cwnd) plotted against the time during the handoff period. The following observations can be made from the previous figure:

- In all cases, the TCP sender enters the slow start mode due to the lost packets during the handoff period.
- The proposed handoff procedure outperforms the others in terms of the cwnd value. This is because the proposed handoff procedure gives the minimum handoff delay. This results in a faster response from the MN. Thus, the TCP connection, in case of the proposed handoff procedure, returns to the normal operation faster than other handoff procedures.

Since the Hierarchical MIP with fast handoff outperforms the fast handoff and the hierarchical one [26], next comparisons are made only between the proposed handoff procedure and the HMIP-Fast approach.

5.2. Linear movement with constant speed (> 1m/s)

The results of this scenario are similar to the previous section (1 m/s case). Thus, both the proposed handoff procedure and the HMIP-Fast procedure are independent of the MN's speed.

5.3. Ping pong movement with constant speed

The ping pong (back and forth movement) scenario is defined in the previous section. In this section, the results of the ping pong simulation scenario are presented.

Fig. 5 illustrates the effect of the ping pong movement on the TCP connection during the handoff. This figure shows the CN's TCP sending buffer. The following observations can be made from this figure:

- The HMIP-Fast approach is notably disrupted with distinguishable breaks in the communication flow.
- The communication flow disruption is much little in the case of the proposed handoff procedure.
- The maximum delay between two consecutive received packets is 629 ms in the HMIP-Fast approach while it is only 368 ms in the proposed handoff procedure.
- The ping pong movement can be perceived as multiple individual linear handoffs. Thus, the proposed handoff procedure outperforms the HMIP-Fast approach in the ping pong movement just like the linear movement case.

Overall Handoff Latency

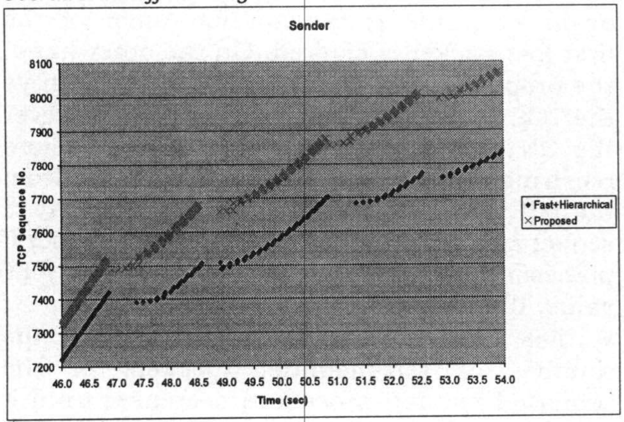


Fig. 5. TCP sequence no. Vs time for CN's TCP sending buffer.

The same observations can be seen from the results of TCP Goodput in table 3 and TCP cwnd in fig. 6.

5.4. The effect of changing round trip time

In this section, the effect of changing RTT is investigated. The RTT value between the CN and the MN is varied from 30 ms to 400 ms and the results of the simulations are in the following fig. 7.

In the previous figure, the handoff latency is plotted against the RTT. The following observations can be made from this figure:

- When the RTT value is 30 ms, the handoff delay of the HMIP-Fast approach drives the TCP source to reach zero window and stop transmitting. However, due to the small value of the RTT, the retransmission starts quickly (retransmission timer depends on the RTT value). Thus, the MN receives the first retransmitted packet quickly. Therefore, the handoff delay is only 531 ms.
- Although the handoff delay of the HMIP-Fast approach is only 531 ms, the proposed handoff procedure outperforms the HMIP-Fast approach with handoff delay equals 216 ms. Moreover the TCP source does not reach zero window.

Table 3
TCP Goodput

Handoff procedure	TCP goodput (Kbytes/sec)
HMIP-Fast	43.75
Proposed	49.63

5.2.3.3 TCP cwnd

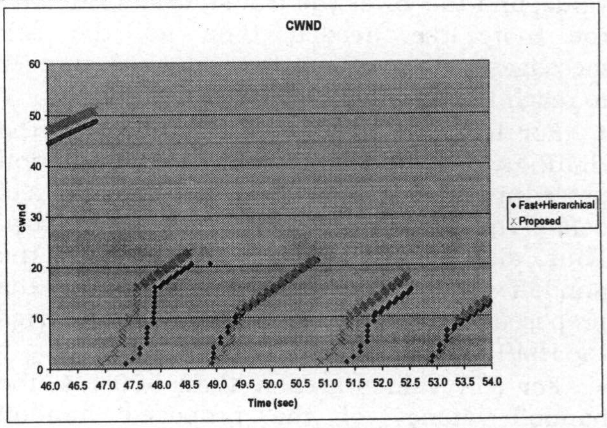


Fig. 6. Cwnd value Vs time.

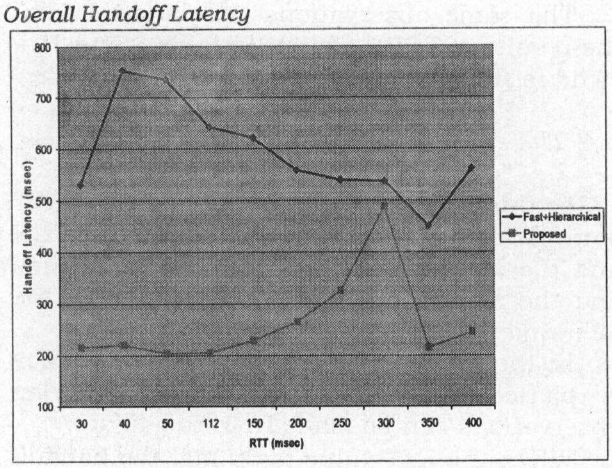


Fig. 7. Overall handoff latency (ms) Vs RTT (ms).

- For RTT values between 40 ms and 150ms (inclusive), the HMIP-Fast approach acts the same as the 30 ms RTT case. The TCP source reaches zero window and stops transmitting until the retransmission timer of the first sent packet of the current window is expired. Upon the expiration of the retransmission timer, the TCP source enters the slow start mode with congestion avoidance and starts the retransmission of the lost packets.
- In contrast to the HMIP-Fast approach, the proposed handoff procedure does not drive the TCP source to reach zero window. However, the TCP source enters the slow start mode due to the reception of ack messages with the expected sequence number after the MN receives out of sequence packets. Furthermore, the more the RTT value, the more the handoff latency. This is because the more the RTT value, and the lower the transmission rate and the more the needed time for the ack messages, with the expected sequence number, to reach the CN.
- For RTT values greater than 150ms, the HMIP-Fast approach completes the handoff procedure and starts receiving the packets before the TCP source reaches zero window. Thus, although the RTT value increases, the handoff latency decreases. However, the proposed handoff procedure still outperforms the HMIP-Fast approach.
- For RTT values greater than 300 ms, the handoff latency of the proposed handoff procedure decreases suddenly. The reason for this sudden decrement is that the time between the sent packets is now enough for the proposed handoff procedure to complete the handoff process. Thus, the MN does not lose any packets (at most one packet).

In fig. 8, the TCP Goodput is plotted against the RTT. The following observations can be made from this figure:

- For both the HMIP-Fast approach and the proposed handoff procedure, the more the RTT value, the less the TCP Goodput value. This is because the more the RTT value, the lower the transmission rate.
- The proposed handoff procedure outperforms the HMIP-Fast approach for all RTT values. However, when the RTT value is 300ms, the HMIP-Fast approach reaches the same TCP Goodput as the proposed handoff procedure. This is because the handoff delays of both approaches are nearly equal at this point.
- The TCP Goodput of the proposed handoff procedure increases suddenly for the RTT values greater than 300 ms. The reason is the reduction of the handoff latency as discussed in the previous subsection.

In fig. 9, the number of retransmitted packets is plotted against the RTT. The following observations can be made from this figure:

- Although both the HMIP-Fast approach and the proposed one have the same number of retransmitted packets for RTT values less than 112 ms which is 32 (= TCP window), the reasons for this are different. The HMIP-Fast approach drives the TCP source to reach zero window. This means that all packets sent from the offered window (=32) are lost due to handoff and the retransmission will not restart again until the retransmission timer of the first lost packet is expired. On the other hand, the proposed handoff procedure does not drive the TCP source to reach zero window. However, the CN retransmits the whole window before receiving the accumulative ack message from the MN which notifies the CN with the out of sequence packets received after the handoff process. This is because the less the RTT value, the higher the transmission rate.
- For RTT values greater than 50 ms, the number of retransmitted packets of the proposed handoff procedure decreases until it becomes zero for RTT values greater than 300 ms.

TCP Goodput

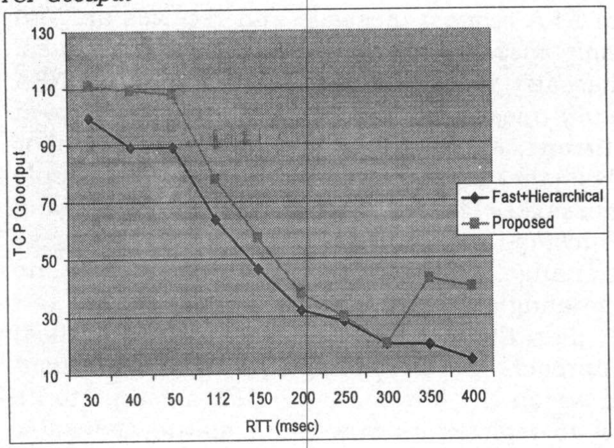


Fig. 8. TCP goodput (Kbytes/sec) Vs RTT (ms).

Retransmitted Packets

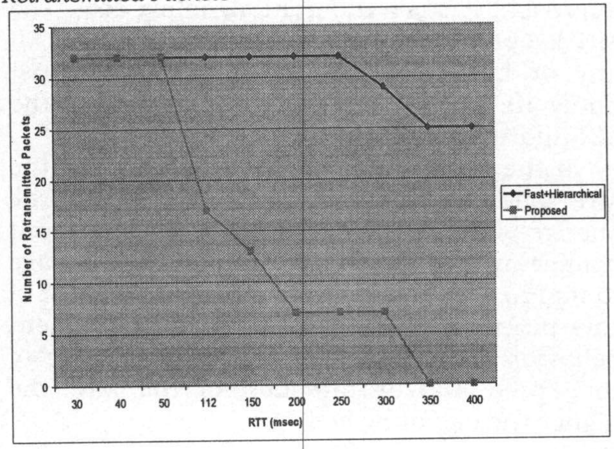


Fig. 9. Number of retransmitted packets Vs RTT (ms).

- For RTT values greater than 150 ms, the HMIP-Fast approach starts receiving packets before the TCP source reaches zero window. Therefore, the number of lost packets becomes less than the TCP window (32). However, for some RTT values the number of retransmitted packets is 32 for the same reason mentioned above.

5.5. The effect of changing HANDOFF_THRESHOLD_TIME

In this section, the effect of the under or over estimation of the HANDOFF_THRESHOLD_TIME is investigated. Since the default value of this parameter in the simulated network is 300 ms, the values of 200 ms and 400 ms are used as indications for the under and over estimation of the HANDOFF_THRESHOLD_TIME respectively.

Fig. 10 illustrates the effect of the HANDOFF_THRESHOLD_TIME on the handoff process. This figure shows the CN's TCP sending buffer. The following observations can be made from this figure:

- The overestimation of the HANDOFF_THRESHOLD_TIME (400 ms) has no effect on the handoff process. This case gives the same results as the 300 ms (good estimation) case. However, if the MN changes its direction or its speed after initiating the handoff procedure, the handoff procedure is cancelled and the performance is degraded. It means that nothing harm from the overestimation of the HANDOFF_THRESHOLD_TIME provided that the MN does not change its movement status after initiating the handoff.
- On the other hand the underestimation of the HANDOFF_THRESHOLD_TIME (200 ms) has considerable effect on the handoff performance. In this case the handoff procedure is started but not completed at the old IP cell. According to the proposed handoff procedure, the MN switches to the fast handoff procedure which degrades the handoff performance to the case of the fast handoff procedure.

The same observations can be seen from the results of TCP Goodput, Number of retransmitted packets and TCP cwnd in table 4 and fig. 11.

Overall Handoff Latency

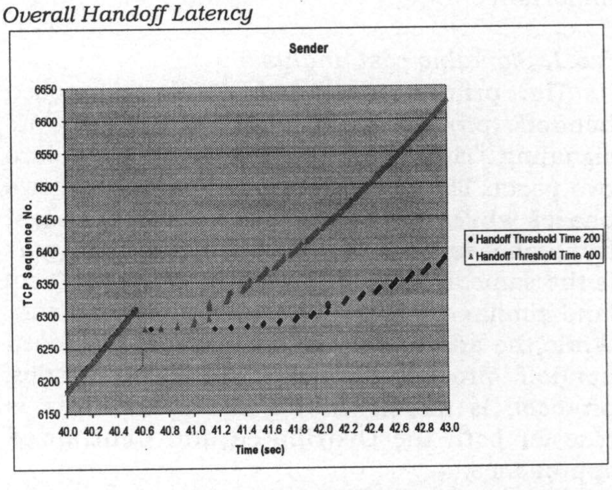


Fig. 10. TCP sequence no. Vs time for CN's TCP sending buffer.

TCP cwnd

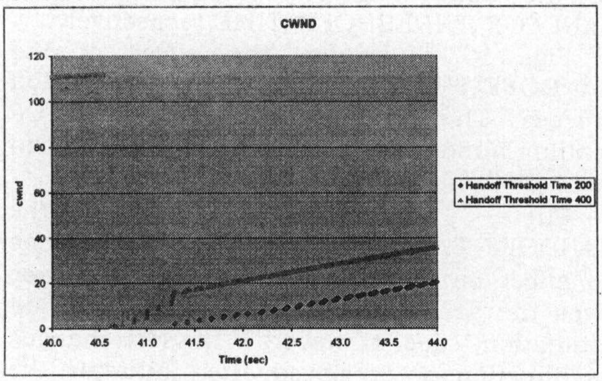


Fig. 11: Cwnd value Vs time.

Table 4
The effect of changing HANDOFF_THRESHOLD_TIME

HANDOFF_ THRESHOLD_ TIME	Handoff latency (ms)	TCP goodput (kbytes/sec)	Number of retransmitted packets
200	750	38.3	26
300	205	77.5	17
400	205	77.5	17

5.6. Cost analysis

In this section, the overhead of the proposed handoff procedure is analyzed in terms of signaling cost and energy (power consumption) cost. Besides signaling and energy cost, there is a money cost because each MN has to be coupled with a location tracking device (e.g. GPS receiver). However, the continuous reduction of the technology prices indicates that the money cost is of no importance.

5.6.1. Signaling cost analysis

The primary overhead of the proposed handoff procedure is associated with its signaling. This signaling can be divided into two parts. The first part includes the first two phases while the second part includes the last three phases. The overhead of the second part is the same as that of the fast handoff protocol (and similar to the HMIP and HMIP with Fast). While the additional overhead of the proposed handoff procedure, over the fast handoff protocol, is mainly in the first part. This is true for both the Distributed and Centralized approaches.

For the Distributed (Mobile Controlled) approach, the first phase is *"Neighbor FAs Information Fetching"* in which the MN sends an NFA *request* message and receives an *NFA reply* message from its FA. The *NFA request* message is of constant size while the NFA *reply* message depends on the number of current cell neighbors. The more the neighbor FAs, the larger the size of the *NFA reply* message. The second phase is *"Movement Tracking"* in which there is no message exchange. Therefore, this phase adds no signaling overhead.

For the Centralized (Network Controlled) approach, the first phase is *"MN Registration"* in which the MN sends an *MIR* message to its FA to register its movement status at its FA. The second phase is *"Movement Tracking"*. While this phase has the same function as the *movement tracking* phase of the distributed approach, it has a different signaling cost. The MN keeps track of its direction and speed. If any or both of them are changed, it must notify its FA with this change to modify the MN information in its MI table.

If the change is in the MN speed only, the MN sends a *SU* message to its FA. If the change is in the MN direction (with or without change in speed), the MN sends a DU message to its FA. This means that the signaling cost of this phase depends greatly on the movement behavior of the MN. The more the change of the speed and/or direction of the MN, the higher the signaling cost.

5.6.2. Energy cost analysis

One major problem with mobile devices is energy consumption. In the proposed handoff procedure, the MN has to continuously update its movement information which means that the MN has to communicate with its location tracking device in a continuous mode. This consumes part of the battery power. Moreover, the *movement tracking* phase of the distributed approach needs recalculating the expected next FA and the expected TCB every time the MN changes its speed and direction. This consumes another part of the battery power. However, the growing demand of multimedia applications pushes many companies to investigate for solutions for this problem such as replacing the traditional battery with the fuel cell.

Generally, the Distributed approach has lower signaling cost than the Centralized approach. While the Centralized approach has lower energy cost than the Distributed approach. Moreover, both approaches add signaling and energy costs over other handoff procedures.

6. Conclusions and future work

In this paper, a novel handoff decision strategy is proposed. The proposed handoff procedure utilizes the movement information provided by the location tracking devices and information about IP cells topology to predetermine the next FA and the handoff decision time, "where and when to handoff".

The following points summarize the advantages of the proposed handoff procedure:

- The proposed handoff procedure eliminates the need for the bi-casting (or n-casting) of the original Simultaneous Binding. Thus, the proposed handoff procedure eliminates the need to overload the network with these duplicate packets.
- The proposed handoff procedure solves the "time ambiguity" problem.
- The proposed handoff procedure outperforms the other handoff procedures in the linear movement in terms of the handoff latency, TCP Goodput, Number of retransmitted packets and cwnd value. This is because the proposed handoff procedure gives the minimum handoff delay. This results in a faster response from the MN. Thus, the TCP connection, in case of the proposed handoff procedure, returns to the normal operation faster than other handoff procedures.
- The ping pong movement can be perceived as multiple individual linear handoffs. Thus, the proposed handoff procedure outperforms the HMIP-Fast approach in the ping pong movement just like the linear movement case.
- The proposed handoff procedure is independent of the MN speed.
- For all RTT values, the proposed handoff procedure outperforms the other handoff procedures.
- The overestimation of the HANDOFF_THRESHOLD_TIME has no effect on the handoff process.

On the other hand, the primary overhead of the proposed handoff procedure is associated with its signaling and its energy consumption. Besides signaling and energy cost, there is a money cost because each MN has to be coupled with a location tracking device. Moreover, the underestimation of the HANDOFF_THRESHOLD_TIME has considerable effect on the handoff performance.

The following points are suggested as research directions for future work:

- Complex simulation experiments have to be performed with multiple connections at the same time. Moreover, more complex topologies have to be considered for future simulation experiments.
- Since the proposed handoff procedure does not drive the TCP source to stop the transmission (zero window), the proposed handoff procedure can be associated with a buffering mechanism to buffer the packets during layer 2 handoff. Thus, the number of retransmitted packets (and lost packets) becomes zero.
- The proposed handoff procedure can be complemented with the context transfer protocol [27] so that the applications running on the MN can operate with minimal disruption during the handoff. Transferring the current state associated with these applications, such as Authentication, Authorization, and Accounting (AAA), header compression, and Quality of Service (QoS), eliminates the need to re-establish the services by performing the necessary signaling flow from scratch.
- The proposed handoff procedure has to be complemented with the necessary security considerations that protect messages exchange from malicious attacks.

References

[1] H. Soliman et al., "Hierarchical Mobile IPv6 Mobility Management", Internet Draft, IETF, Work in Progress, October (2004).

[2] G. Dommety et al., "Fast Handovers for Mobile IPv6", Internet Draft, IETF, Work in Progress, October (2004).

[3] K. Malki and H. Soliman, "Simultaneous Bindings for Mobile IPv6 Fast Handoffs", Internet Draft, IETF, Work in Progress, June (2002).

[4] C. Perkins et al., "Mobility Support in IPv6", Internet Draft, IETF, Work in Progress, June (2003).

[5] S. Thomson and T. Narten, "IPv6 Stateless Address Autoconfiguration", RFC 2562, IETF, December (1999).

[6] C. Perkins "Mobile IP, Design Principles and Practices", Wireless Communication Series, Addison Wesley (1997).

[7] N.A. Fikouras and C. Gorg, "Performance Comparison of Hinted and Advertisement Based Movement Detection Methods for Mobile IP Handoffs", The International Journal of Computer and Telecommunications Networking, ACM, Vol. 37, Issue 1, pp. 55 - 62 (2001).

[8] A.T. Campbell et al., "Comparison of IP Micro-Mobility Protocols", IEEE Wireless Communications Magazine, Vol. 9 (1), February (2002).

[9] R. Ramjee et al., "HAWAII: A Domain-Based Approach for Supporting Mobility in Wide-area Wireless Networks", In Proceedings of ICNP, IEEE (1999).

[10] A. Valko et al., "Design and Performance of Cellular Access Networks", IEEE Personal Communications, Special issue on IP-Based Mobile Telecommunications Networks (2000).

[11] J. McNair et al., "An Inter-System Handoff Technique for the IMT-2000 System", in Proc. of IEEE INFOCOM 2000, Tel Aviv, Israel, March (2000).

[12] K. Malki, "Low Latency Handoffs in Mobile IPv4", Internet Draft, IETF, Work in Progress, June (2002).

[13] R. Cáceres and V. Padmanabhan, "Fast and Scalable Wireless Handoffs in Support of Mobile Internet Audio", ACM, Mobile Networks and Applications, Vol. 3 (4), December (1998).

[14] K. Malki et al., "Fast Handoffs Method for Real-Time Traffic over Scaleable Mobile IP Networks", Internet Draft, IETF, Work in Progress, June (1999).

[15] R. Hsieh et al., "A S-MIP: A Seamless Handoff Architecture for Mobile IP", In Proceedings of INFOCOM, IEEE, San Francisco, USA (2003).

[16] Y. Pan et al., "An End-to-End Multi-Path Smooth Handoff Scheme for Stream Media", in Proceedings of WMASH, ACM, San Diego, California, USA, September (2003).

[17] K. Mitchell et al., "Approximation Model of Wireless Cellular Networks Using Moment Matching", IEEE Journal on Selected Areas in Communications Vol. 19 (11), November (2001).

[18] Z. Xu et al., "A New Adaptive Channel Reservation Scheme for Handoff Calls in Wireless Cellular Networks", In Proceedings of IFIP Networking, Pisa, Italy, (2002).

[19] R. Hsieh et al., "A Comparison of Mechanisms for Improving Mobile IP Handoff Latency for End-to-End TCP", in Proceedings of MobiCom, ACM, San Diego, California, USA, September (2003).

[20] M. Liebsch et al., "Candidate Access Router Discovery", Internet Draft, IETF, Work in Progress, September (2004).

[21] W. Zhang et al., "Design and Evaluation of a Handover Decision Strategy for 4th Generation Mobile Networks", White Paper, University of Stuttgart, Institute of Communication Networks and Computer Engineering (2005).

[22] The UCB/LBNL/VINT, Network Simulator – ns (version 2), http://www.isi.edu/nsnam.

[23] J. Widmer, Extensions to the ns Network Simulator, http://www.informatik.uni-mannheim.de/~widmer.

[24] R. Hsieh, Home Page, http://mobqos.ee.unsw.edu.au/~robert/index.php.

[25] H. Yokota et al., "Link Layer Assisted Mobile IP Fast Handoff Method Over Wireless LAN Networks", in Proceedings of MobiCom, ACM, Atlanta, Georgia, USA, September (2002).

[26] R. Hsieh and Seneviratne, "A Performance Analysis on Hierarchical Mobile IPv6 with Fast-Handoff Over End-to-End TCP", In Proceedings of GLOBECOM, IEEE, Taipei, Taiwan, (2002).

[27] C. Perkins et al., "Context Transfer Protocol", Internet Draft, IETF, Work in Progress, August (2004).

Received May 29, 2007
Accepted September 20, 2007