A Framework of Handoffs in Wireless Overlay Networks Based on Mobile IPv6

Cheng Wei Lee, Li Ming Chen, Meng Chang Chen, and Yeali Sunny Sun

***Abstract*—Although there are various wireless access network technologies with different characteristics and performance level have been developed, no single network that can satisfy the anytime, anywhere, and any service wireless access needs of mobile users. A truly seamless mobile environment can only be realized by considering vertical and horizontal handoffs together. With the advantages of Mobile IPv6, a more comprehensive and integrated framework of heterogeneous networks can be developed. In this paper, we discuss the issues related to handoffs including horizontal and vertical handoffs. We present a scheme for integrating wireless local area network and wide area access networks, and propose a micromobility management method called HiMIPv6+. We also propose a quality-of-service-based vertical handoff scheme and algorithm that consider wireless network transport capacity and user service requirement. Our prototype evaluations and the simulations show that our framework performs as expected.**

***Index Terms*—Hierarchical mobile IPv6 (HiMIPv6), horizontal handoff, mobile IPv6, quality-of-service (QoS), vertical handoff, wireless local area network (WLAN).**

I. Introduction

WHILE commercial third-generation (3G) is on the way to takeoff worldwide, research on fourth–generation/beyond 3G (4G/B3G) [1] is already underway in many countries. Compared with the 3G wireless system defined by IMT-2000, B3G is expected to provide users with convenient global information access capabilities and personalized multimedia wireless communication services. Such services will be characterized by broadband, global availability, immediacy, and high mobility. To realize the goal of B3G, a generally accepted approach is to integrate the broadband Internet and various wireless communication systems to create a heterogeneous multiaccess network. In this network, various air interface techniques, especially, wireless local area network (WLAN) and mobile cellular networks, such as General Packet Radio Service (GPRS) and Universal Mobile Telecommunications System (UMTS), are integrated with Internet protocol (IP)-based networks as an overlay structure. In an all-IP environment, every device would be assigned a unique, static and public IP address to enhance connectivity to the networks. Obviously, the earlier version of the Internet protocol version 4 (IPv4) suffers from the shortage of address naming space. Internet Protocol version 6 (IPv6), on the other hand, is a better choice because of its expanded addressing capacity, which allow a much larger number of addressable nodes to active at the same time. In addition, IPv6 changes the way that IP header options are encoded, allows for more efficient forwarding, less stringent limits on the length of options and greater flexibility for introducing new options in the future.

Because no existing wireless network technology can provide high bandwidth, low latency, low-power consumption, and wide-area data service to a large number of mobile users simultaneously [2], mobile nodes (MNs) must consider three factors when performing horizontal or vertical handoffs to support network applications. The first factor is *mobility*. An MN may move out of its current wireless base stations (BS's) coverage area, hence, the signal strength would decrease. To maintain connectivity, the MN is forced to seek and access the network through another BS. The second factor is that more than one BS (the same or a different wireless access technology) are available simultaneously, or the loading of the connected BS may have reached its capacity limit (e.g., bandwidth and latency), the MN may choose to switch to another access BS to achieve a better *performance* for the current application. Finally, *user preference* must also be considered. A mobile user may have the access preference based on price, power consumption, speed, and other requirements, and may initiate a handoff manually. However, only by considering vertical handoff and the traditional handoff mechanisms (the so-called horizontal handoff) together, can a truly seamless mobile environment can be realized.

Horizontal handoff within a wide-area access network (WAAN) is already quite mature. But a general handoff problem among WLAN domains is the lack of immediate upper layer awareness when the lower layer has performed a handoff to a new access point (AP) in a different subnet. It usually takes several seconds for the upper layer to detect MN movement and complete the reregistration procedure. Many micromobility designs and lower layer supported protocols have been proposed, but there is still room for further improvements. Also, there are additional concerns about vertical handoff, which needs to coordinate heterogeneous network interfaces and mobile devices for smooth handoff. Vertical handoff can be performed anywhere within network domains, not just at the border of a cell, because different wireless networks do not necessarily overlay with each other only at the boundary of transmission domains. In addition, issues such as authentication/authorization, billing, location management, and seamless handoff need to be considered in a heterogeneous multiaccess network.

Manuscript received December 1, 2004; revised June 20, 2005. This work was supported in part by the National Science Council (NSC) under Grant 93-2213-E-001-006 and Grant 93-2213-E-002-058.

C. W. Lee and Y. S. Sun are with the Department of Information Management, National Taiwan University, Taipei, Taiwan 106 R.O.C. (e-mail: jameslee@ntu.edu.tw; sunny@im.ntu.edu.tw).

L. M. Chen and M. C. Chen are with the Institute of Information Science, Academia Sinica, Nankang 11529, Taipei, Taiwan, R.O.C. (e-mail: leeming@iis.sinica.edu.tw; mcc@ iis.sinica.edu.tw).

Digital Object Identifier 10.1109/JSAC.2005.856833

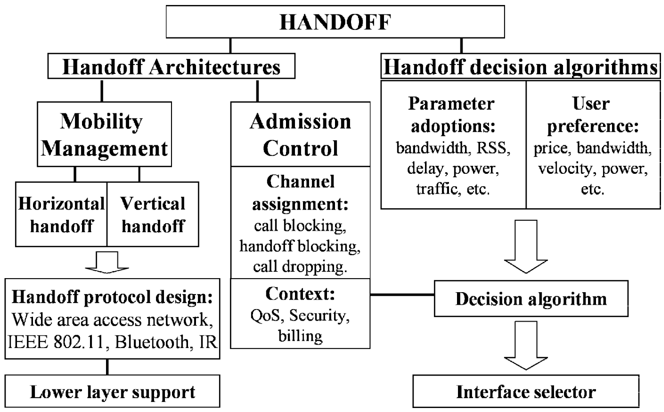

Fig. 1. Handoff architecture and important issues.

In the following section, we discuss important issues about handoff and related works. In Section III, we specify a loosely coupled WLAN-WAAN network architecture for seamless handoffs in a wireless overlay network based on Mobile IPv6. Also, a WLAN bandwidth measurement and a bandwidth-based vertical handoff algorithm are proposed for handoff decision-making. We then describe a micromobility management scheme, hierarchical Mobile IPv6 (HiMIPv6), and an enhanced multilayer HiMIPv6, HiMIPv6+, both of which help minimize the latency and packet loss, thereby supporting seamless handoff. In Section IV, we present some experiment results of HiMIPv6 and HiMIPv6+. We also discuss the design of a network interface selector is described and the results of our simulations. Finally, we then conclude the paper in Section V.

II. Issues and Enabling Technologies

In this section, we review the issues and enabling technologies for handoffs in wireless networks. There are many important issues related to handoff. Phalavan *et al.* made a summary of handoff issues in [3]. Here we specify some further issues and present a modified architecture, as shown in Fig. 1. Akyildiz *et al.* [4] introduces the three-stage handoff process, including handoff decisions, radio link switch and channel assignment, that its important issues are addressed in this article except some admission control issues involved further resource management and AAA (Authentication, Authorization, and Accounting).

A. Mobile IPv6

The Mobile IPv6 [5] standard specifies that a MN should always be identified by its home address, regardless of its current attachment point to the Internet. And while away from its home domain, the MN is also associated with a care-of address (CoA). The protocol provides for registering the CoA with the MN's home agent (HA), which sends each datagram destined for the MN through a tunnel to the CoA. After arriving at the end of the tunnel, each datagram is then delivered to the MN. Once the MN registers with the sender of the datagram, it can communicate directly without the assistance of the HA.

However, the mobile IP protocol has several drawbacks, such as long update latency, high signaling overhead, and a lack of paging support. The interval between router advertisements (RA) is critical to the length of handoff. A long interval is not suitable for the timely movement detection of MN, because in average it waits for half an interval to receive the RA before it can perform new address configuration and registration. Although Mobile IPv6 improves the RA interval from 3 s [6] to 30 $\sim$ 70 ms, it may still not be suitable because unnecessary message broadcasting may not be affordable, as wireless resources are generally scarce.

B. Micromobility Protocols and Horizontal Handoffs

In recent years, several IP-based micromobility management systems have been proposed to complement the mobile IP. To reduce the signaling overhead caused by global mobility management, the concept of a "domain" was introduced. The movement of an MN in a domain is transparent to the HA and the CN. In other words, the MN does not have to send another binding update (BU) to the HA or the corresponding node (CN) when changing resident subnet from one attachment point to another in the same domain. Furthermore, incoming packets can still be efficiently forwarded to the current location of the MN.

HAWAII [7], a domain-based approach for supporting mobility, uses specialized path setup schemes that install host-based forwarding entries in specific routers to support intradomain micromobility. However, there are two weaknesses in these schemes: 1) they assume that BSs have IP routing functionality and 2) they incur a higher processing overhead than mobile IP on limited-power MNs.

Cellular IP [8] maintains a distributed cache for location management and routing purposes. A distributed paging cache coarsely maintains the position of an "idle" MN in a service area. Cellular IP uses this paging cache to pinpoint "idle" MNs that wish to engage in "active" communications. Meanwhile, distributed routing cache maintains the position of active MNs in the service area and dynamically refreshes the routing state in response to the handoff of active MNs. When an active MN is approaching a new BS, it transmits a route-update packet from the old to the new BS. Some wireless network standards, like IEEE 802.11, specify that the search for a new BS can only be triggered when the signal strength of the old BS decreases to a certain level. Hence, the old BS may not successfully receive the route-update packet. On the other hand, transmission of a route-update packet when the old BS signal is still strong is not feasible, as it may duplicate transmission of the packet to the MN, thereby increasing the network's load unnecessarily.

HMIPv6 [9] is a hierarchical architecture based on Mobile IPv6, in which each MN is required to configure two care of adresses (CoAs): a regional CoA and an on-link CoA. When the MN moves to a foreign network, it sends two different binding updates (BUs): one is a Mobile IPv6 BU sent to HA or CN, and the other is a local-BU sent to the mobility anchor point (MAP), which is the root of the hierarchical architecture. When MN moves to another subnet within the same foreign domain, it only updates its on-link CoA by sending a local BU to MAP. If the MN is a portable device, which has poor computing capability and a small memory, it may not be able to perform many computing jobs or store much information. HMIPv6 may be too complicated for a simple MN, since it has to identify which BU is to be sent and also store information about the MAP. In addition, HMIPv6 is a two-level hierarchical architecture (MAP and access router). Thus, whenever a MN moves to a different subnet, it has to register to the same MAP, which places an unnecessary load on MAP.

C. Integrated Architecture of Heterogeneous Networks

With regard to integrated network architecture European Telecommunications Standards Institute (ETSI), specified two basic approaches for Internet working between WLAN and other cellular networks in 2001 [10]. One is tightly coupled architecture while the other is a loosely coupled architecture. In the former, WLAN operates as a virtual radio access network (RAN) in the cellular system. In this case, the data traffic of WLAN must pass through the core network of the cellular system before entering the Internet. The main feature of the tightly coupled architecture is that the handoff and security issues in WLAN are totally controlled by the cellular system, while the main drawbacks are the difficulty of supporting the WLAN interface in serving GPRS support node (SGSN) and the performance overhead of the cellular system.

In a loosely coupled architecture, the WLAN is deployed as an access network that complements the cellular system. In other words, the WLAN bypasses the cellular system and provides direct data access to the Internet. Although the loosely coupled architecture provides a flexible and independent environment for both wireless technologies, it requires mobile IP to support mobility across the two access networks. Further details of both architectures can be found in [11].

Phalavan *et al.* compared five different architectures for implementing vertical handoff between networks based on the IEEE 802.11 standard and networks based on GPRS [3]. Two of the five architectures, Mobility Gateway/Proxy-Based and Mobile IP-Based, have been found to be the most efficient due to their loosely coupled structure. Extending research in [3], Tsao and Lin [12] proposed a gateway approach to integrate WLAN and UMTS networks, in which mobile IP is not required and users can roam between networks using one IP address. The same idea of the gateway approach for the integration of WLAN and CDMA2000 was proposed by Bduuhikot *et al.* [13], which incorporates the concepts of tight integration and loose integration.

D. Handoff Control and Handoff Decision Making Algorithms

There are three strategies for detecting the need for handoff: mobile-controlled handoff (MCHO), network-controlled handoff (NCHO), and mobile-assisted handoff (MAHO). MCHO is used in IEEE 802.11 WLAN networks, where the MN continuously monitors the signal of an AP and initiates the handoff procedure when some handoff criteria are met. In contrast, MAHO has been widely adopted in the current WAAN that the MN measures the signal of surrounding BSs, and the network then decides whether or not to begin the handoff procedure.

One of the advantages of MCHO is the low complexity in network equipment. In addition, the WLAN NIC is designed to communicate via single channel at a time according to the IEEE 802.11 standard, so the cost of deploying and using a WLAN is cheaper. However, current WLAN users suffer from long latency and a large number of packet losses when intersubnet handoff occurs. As the network lacks knowledge about the MN's movements, not until the MN announces its presence in a new subnet can the packets' routing path be changed and directed to the correct location of the MN.

In [14], Koodli suggested that with the aid of link-layer specific mechanisms, a MN can collect the link-layer addresses of surrounding APs. Thus, it can obtain information about new access routers (NAR) from the current access router [i.e., the previous access router (PAR)] and formulate a prospective new care-of address (NCoA). Once the network layer of the MN is notified the occurrence of a link-layer trigger, e.g., radio signal degradation, it sends a fast binding update (FBU) message to current access router. After a series of message exchanges, packets sent to the MN's old CoA are tunneled from the PAR to the NAR and directed to the new location of the MN, which reduces the handoff latency and packet loss. However, this fast handover method suffers from the same dilemma with cellular IP.

Interaccess point protocol (IAPP) [15] provides message exchange between wireless APs with the same ESSID in the same LAN. The AP broadcasts a layer 2 update frame carrying the MAC address of the MN that associates or reassociates with it. All the layer 2 devices (e.g., switches and bridges) in the same LAN receive the update frame and update their switch table and, hence, alter the switching path of packets directed to the MN. IAPP facilitates intrasubnet handoff, but, unfortunately, IAPP can only be used in LAN.

The handoff decision of MAHO is made by the network. Thus, global optimization and coordination among MNs can be supported. But when it comes to handoff between heterogeneous wireless access network technologies, only MNs have the knowledge about what kind of interfaces they are equipped with that prevents MAHO from fully functioning. Even if the network has the knowledge, there may be no way to control another network that the MN is about to handoff to. Therefore, MCHO and further assistance from the networks is more suitable for vertical handoff.

A great deal of researches, such as [16]–[18] uses the received signal strength (RSS) of WLAN beacons as the basic criterion for handoff decision. In general, the radio signal strength fades with the transmission distance. If the RSS is the only factor in the decision of vertical handoff, the handoff only happens at the cell boundary, where the RSS is lower than a predefined threshold. The major difference between our proposed bandwidth-aware vertical handoff technique and those of other approaches is that our method considers the residual bandwidth of WLAN, besides RSS, as the criterion for handoff decisions.

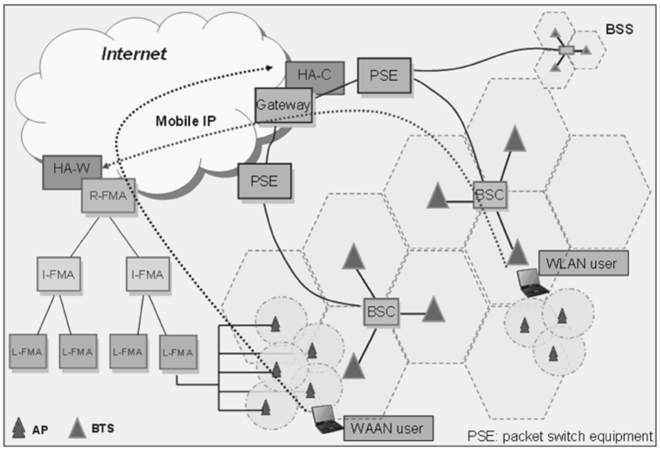

Fig. 2. WLAN-WAAN integrated architecture.

III. System Architecture

A. An Integrated Architecture

In view of the advantages of minimal modification of the original network nodes and network topology and the potential for future growth, we adopt the ETSI loosely coupled architecture and take an integrated WLAN and WAAN (mostly a cellular network) as an example. In Fig. 2, the right-hand side is the architecture of WAAN and the left-hand side is a WLAN domain (e.g., a campus) constructed with HiMIPv6+. The Internet mobile devices, equipped the two network interfaces, select one interface to communicate with other users or connect to the Internet via one of the two wireless networks. Mobility management is provided by the Mobile IPv6 protocol, which is transparent and incorporated in several wireless systems, such as GSM, GPRS, 3G, Bluetooth, WLAN, and wireline systems.

Two kinds of mobile users can roam around this architecture. One is the WLAN user whose HA (HA-W) resides in a WLAN domain. The prefix of the static IP that the node uses belongs to the WLAN domain. The other is the WAAN user whose HA (HA-C) resides in a WAAN network, and the prefix of the static IP belongs to the WAAN domain. The HA can be a specific device added to the network, or it can reside in an existing device like GGSN in the GPRS network, or the root-FMA in our HiMIPv6+.

B. Handoff Decision-Making Algorithm

1) WLAN Bandwidth Measurement: As the bandwidth of WLAN is shared by all MNs, the number of MNs in a domain and their network access patterns affects the available bandwidth (called the residual bandwidth) for an MN about to join the domain. To estimate of the residual bandwidth in the WLAN, we refer to the concept of QBSS (QoS BSS) load defined in IEEE 802.11e [19]. The QBSS is a basic service set [(BSS) i.e., a WLAN domain] that supports quality-of-service (QoS) facility for communication via the IEEE 802.11 wireless medium. The QBSS load is an element information field in the beacon frame sent by an AP and contains information about the station count, channel utilization, and packet loss rate of the QBSS. The station count field indicates the total number of stations currently associated with the QBSS. The channel utilization field is defined as the percentage of time the AP senses the medium is busy using the carrier sense mechanism. The frame loss rate in the QBSS load is defined as the portion of transmitted medium access control (MAC) protocol data units (MPDUs) that require retransmission, or are discarded as undeliverable.

In general, the virtual carrier-sensing function of IEEE 802.11 is provided by the network allocation vector (NAV). The MN sets the value of the duration field of the data frame to the time it expects to use the medium and other MNs update their NAVs accordingly when they sense the 802.11 frame. The NAV is a timer that indicates the amount of time the medium is reserved by other MNs or by an AP. When the NAV is nonzero, the virtual carrier-sensing function indicates that the medium is busy; otherwise, the medium is idle. By using the NAV, MNs can ensure that atomic operations are not interrupted.

In order to calculate the channel utilization of the QBSS load, an AP has to know how busy the channel is. In addition to the NAV reserved by the MNs, the AP also considers the duration of the packet that it will send to MNs as channel busy time. The AP then broadcasts the calculated channel utilization to MNs. However, even when the network is overloaded, the calculated channel utilization of the QBSS load does not reach 100% since there are waiting periods for MNs and the AP to start transmission, including SIFS, DIFS and random backoff periods. When the NAV of a station counts down to zero, the MN or the AP must still wait for a DIFS period and random backoff time before they can send an IEEE 802.11 frame. All the waiting time needs to be considered when measuring the residual bandwidth of a WLAN. When the MN receives the beacon carrying the QBSS load, it can approximately evaluate the residual bandwidth of WLAN by the following formula:

$$\begin{aligned} residual\_bandwidth = (throughput) & \\ &\times (1 - \alpha \times channel\_Utilization) \\ &\times (1 - Frame\_Loss\_Rate)\,. \end{aligned}$$

In the above formula, as the channel utilization of QBSS calculated by AP considers the time that the wireless medium is busy, including all signaling overhead and collision times, a factor α in introduced to reflect the IEEE 802.11 MAC overhead. From empirical results, the upper bound of channel utilization is about 80% in IEEE 802.11b when the WLAN is overloaded. Therefore, α is set at 1.25 to reflect the true ratio of elapsed time. The throughput in the formula is defined as the real throughput that can be shared among mobile devices in WLAN, which is set as 6 Mb/s. The channel utilization of the QBSS load is a moving average, so the evaluated residual bandwidth can be considered as a good prediction of network performance, rather than as an instantaneous value. The residual bandwidth also takes account of the packet loss rate.

In a WAAN, such as GPRS, UMTS, or CDMA2000, an MN acquires radio resources from a BS via a two-phase access procedure. The allocation of channels and time slots is strictly controlled by the BS. Hence, there is no guarantee that a MN will receive as much bandwidth as it desires, or even get information about residual bandwidth beforehand. After the two-phase negotiation, the MN will be assigned a fixed bandwidth for data transmission. For example, an MN may have a packet data rate between 114 and 200 kb/s in GPRS depending on the coding scheme, whereas, in UMTS, it may up to a 384 kb/s for outdoor use and 2 Mb/s for indoor use. In this paper, we assume that via some time slot allocation mechanisms in the BS, the MN will eventually acquire sufficient bandwidth within the maximum rate of the WAAN.

2) Vertical Handoff Algorithm: In addition to the network performance measure, the network interface selector also uses other parameters for the vertical handoff algorithm. We donate the parameters as user profile, network profile, and application profile. The user profile is an important, yet complex, factor based on the user's preferences, charges, and types of service, etc. The user preference network determines the network that the MN connects to in the idle mode. In the network profile, we define the maximum useful bandwidth of WLAN and WAANs to help the algorithm in decision making. For the application profile, we follow the four QoS classes of network applications defined by UMTS [20]. These application traffic types are conversational class, streaming class, interactive class and background class. According to the delay sensitivity characteristics, the first two classes are grouped as real time service, while the other two belong to nonreal time service. Real-time service is sensitive to delay, thus, a guaranteed transmission rate is needed. Nonreal time service, on the other hand, is insensitive to delay, but the correctness of the transmitted data is essential. As the transmission delay of a radio medium is negligible compared to the end-to-end network delay, the handoff decision of real-time and nonreal time services are both based on bandwidth in the vertical handoff algorithm. The parameters in our algorithm are listed in Table I.

TABLE I
PARAMETERS OF VERTICAL HANDOFF DECISION ALGORITHM

Profile parameters	
N_{prefer}	the preference access network of a user
$WLAN.B_{max}$	the maximum useful bandwidth of WLAN
$Cellular.B_{max}$	the maximum useful bandwidth of WAAN
APP	the current application
$APP.T_{class}$	the traffic class (real time/non-real time)
$APP.B_{need}$	the minimum bandwidth required by the application
Other parameters	
RSS	the received signal strength of WLAN beacon
RSS_T	the predefined threshold of RSS when the mobile device crosses the cell boundary of WLAN
N_{RSS}	the number of continuous RSS of WLAN beacons below RSS_T
N_{RSS_T}	the N_{RSS} threshold to determine the handoff
$B_{residual}$	the residual bandwidth of WLAN
B_{now}	the current throughput of the application

The vertical handoff decision algorithm is designed for use in two situations: 1) when the MN connects to WLAN and 2) when the MN connects to a WAAN. Fig. 3 shows the handoff decision flow chart of the two situations. In the algorithm, we consider the RSS of beacons, because it is the most common criterion for deciding the distance between the MN and the AP and for judging if the MN is going to cross the service area of the WLAN.

When a MN connects to a WLAN and receives a beacon, it compares the RSS of the beacon with the RSS threshold. If the RSS falls below the RSST, NRSS is increased by one. The MN determines if it should leave the wireless service area by comparing NRSS_T with the updated NRSS. This algorithm also considers the QoS of the MN. If the MN is in the idle state, it will switch to its preferred access network. When the MN is in the busy state and the current application is a real time service, the algorithm first determines if the bandwidth supported by WLAN is sufficient for the application. If the answer is positive, the MN stays with the same network and does not need to handoff to the WAAN or other overlay networks, even if they can provide higher bandwidth. In this way, the real time service of the MN can avoid unnecessary handoff latency caused by vertical handoff. For nonreal time service, the amount of transmission data is more important than the delay. Therefore, the vertical handoff decision just checks if the WLAN is under drastic contention and if the cellular network can provide higher bandwidth than the WLAN. If both are affirmative, the interface switches to WAAN.

When the MN is in a WAAN, it activates the WLAN interface to check beacon information as infrequently as possible, because a WLAN interface is of higher power consumption. When the WLAN interface receives the beacon, the algorithm is activated. At first, it determines whether the WLAN is nearby. If it is, the algorithm considers the need for vertical handoff. For an idle user, the operation is the same as in the WLAN. If the MN is in the busy state, it measures the residual bandwidth of the WLAN and determines the time to handoff to the WLAN.

C. Radio Link Switch

Horizontal handoff for cellular networks is well developed, but there is still a lot of work to do on WLANs and interheterogeneous networks. We therefore propose a vertical integration of layers 2 and 3 mobility management scheme HiMIPv6+.

1) Hierarchical Mobile IPv6 (HiMIPv6): A foreign network domain comprises a number of routers interconnected as a hierarchy, as shown in Fig. 4. Each router is equipped with an extended foreign agent, referred to as the foreign mobility agent (FMA), which supports hierarchical mobility management. Three physical types and 1 logical type of FMA are responsible for slightly different functions. The former are root-FMA, interior-FMA, and leaf-FMA, and the latter is switching-FMA.

When an MN first moves to a foreign network, it receives a router advertisement (RA) according to the Internet Engineering Task Force (IETF) Mobile IP framework. In our approach, a

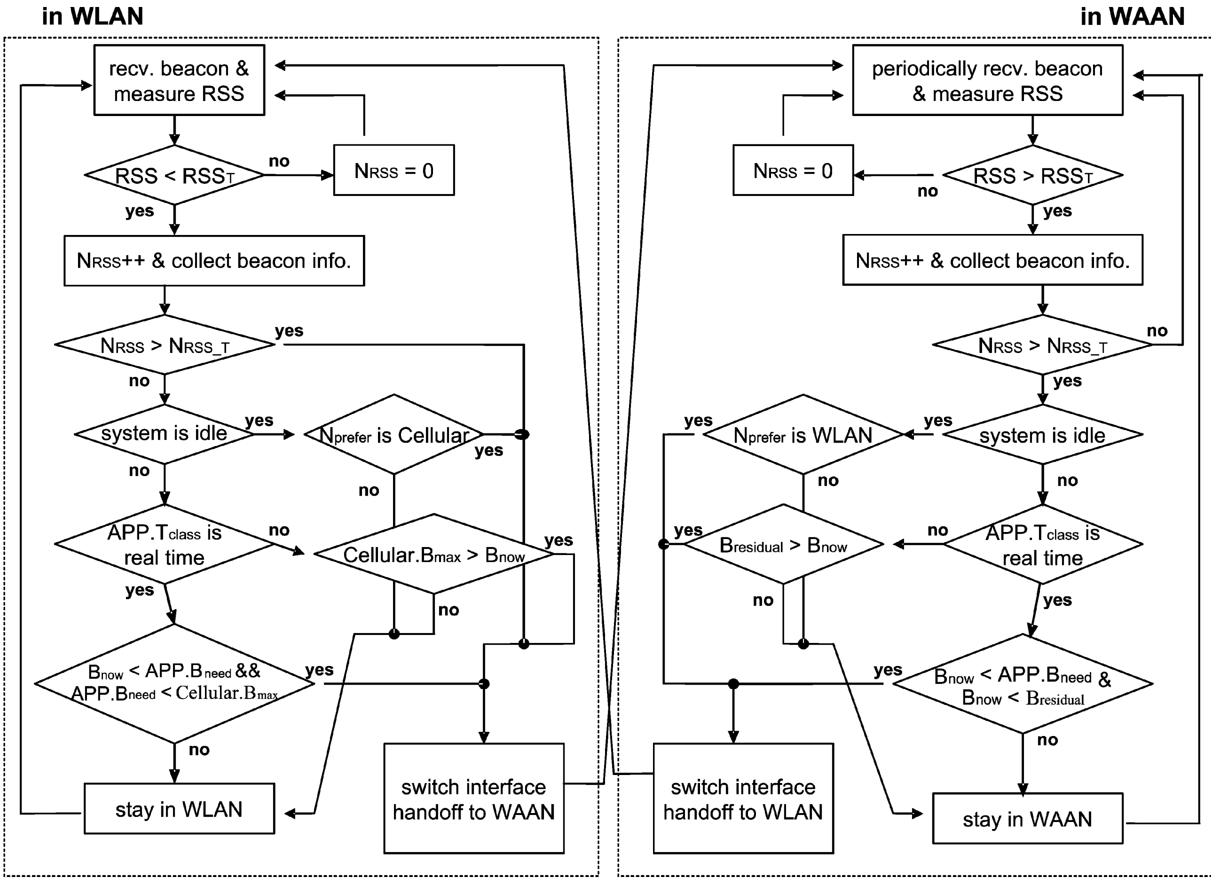

Fig. 3. Vertical handoff decision algorithm.

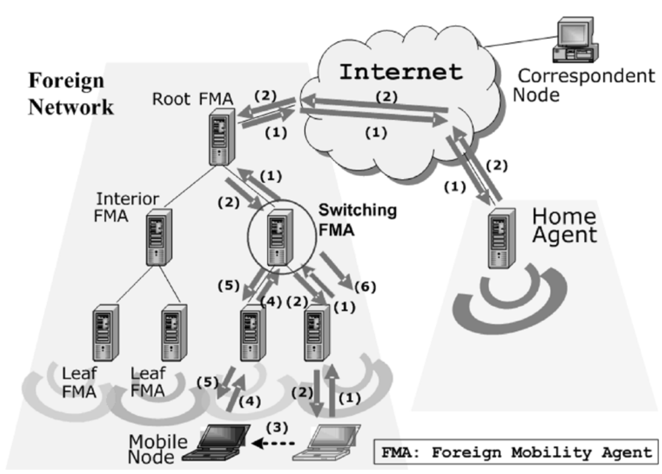

Fig. 4. HiMIPv6 network architecture.

HiMIPv6 RA (a RA with the "Hy" bit set) is used. The MN then sends a Hy-BU (a BU with "Hy" bit set) to its HA and all the CNs with which it is communicating to update its new location (Steps 1 and 2 in Fig. 4). Subsequently, as the MN may roam across different subnets within the domain (Step 3), the Hy-BUs sent by the MN will be identified and intercepted by the switching-FMA (Step 4) to update the "visitor list," without forwarding the Hy-BUs to the HA and CNs. The visitor list is an important data structure used by the FMAs to process HiMIPv6 messages and route packets from/to visiting MNs within the domain. In Step 5, the switching-FMA sends a Hy-binding acknowledgment (Hy-BA) as if it were the HA (there is no difference in the format and values of a Hy-BA and a Mobile IPv6 BA), thereby completing the MNs new registration. At the same time, the switching-FMA sends a Hy-BI (Step 6) to inform all the FMAs on the old registration path that the MN has already left their coverage.

Under such a hierarchical mobile IP architecture, the MNs CoA—known to the HA and CNs—is the address of the root-FMA. Although the MN obtains new CoAs as it roams from one subnet to another, its movement within the HiMIPv6 domain is transparent to the HA and the CNs. With hierarchical-structured FMAs, HiMIPv6 not only localizes the registration messages, thus reducing handoff delay, but also an MNs functionality changes slightly from Mobile IPv6. In fact, an MN only has to detect the "Hy" bit in Hy-RA and send a Hy-BU. In contrast, a MN that wants to operate on the HMIPv6 network has to identify which kind of BU it should send and keep information about the MAP. These operations require more computation power and memory than HiMIPv6. In addition, with multilevel FMA deployment, each HiMIPv6 FMA can become a switching-FMA and handle a MN's handoff without adding more loads to its ancestor FMA. On the contrary, MAP of HMIPv6 is heavier loaded.

2) HiMIPv6+: From the MN's perspective, the handoff time is the time between receiving the last packet from the

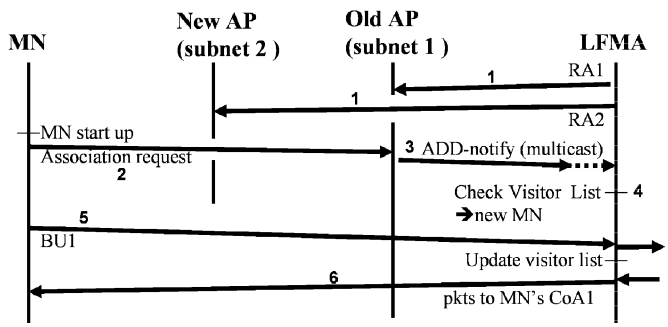

Fig. 5. Message sequence as an MN starts up.

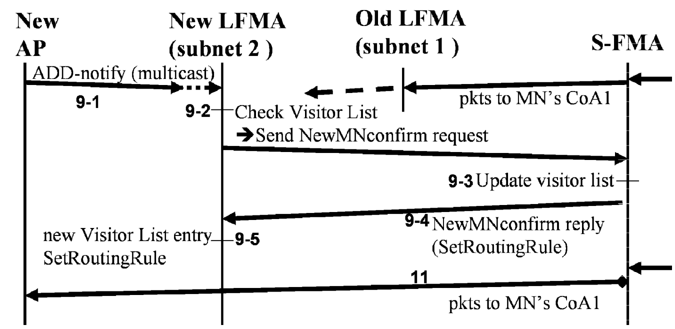

Fig. 7. Message sequence of NewMNconfirm.

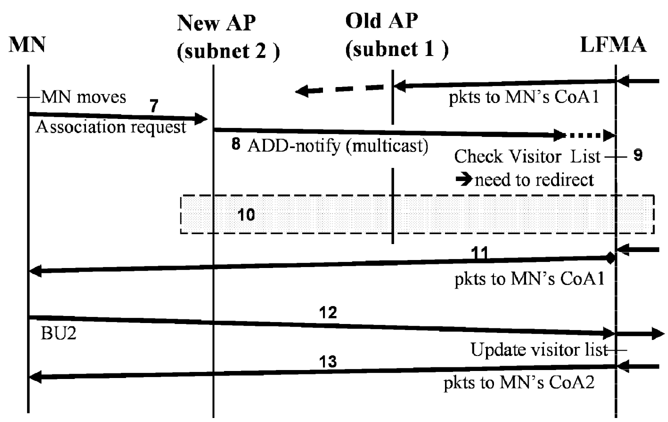

Fig. 6. Message sequence as an MN moves to subnet 2.

old AP and receiving the first packet from the new AP. The main contribution of HiMIPv6 is to localize the reregistration message, but the handoff time is still too long for applications which require real-time transmissions. In HiMIPv6+, we take advantage of the ADD-notify message sent by an AP equipped with IAPP when a MN (re)associates with the AP. By joining the IAPP IPv6 multicast group, the leaf-FMA is aware of the MN's appearance in the new subnet promptly. Thus, the FMAs can take actions to reduce the packet loss caused by the MN's handoff.

MN Startup: All APs under HiMIPv6+ are equipped with IAPP. The following is the message sequence of an MN joining subnet 1, as shown in Fig. 5. The AP that the MN is associated with sends ADD-notify (Step 3). The leaf-FMA then records the MN's information, such as its MAC and the AP's address in "VMAC list," and waits for the Hy-BU (Step 4). The VMAC List provides the relationship between the MN's MAC and its home address, and records information to set the routing rules. When the leaf-FMA detects the corresponding Hy-BU (Step 5), it updates the "visitor list" and the "VMAC list." Note that the packet carrying the BA is not illustrated in Fig. 6. After the registration is completed, packets sent to the MN's CoA1 can be correctly routed to the MN in subnet1 (Step 6).

Handoff Under the Same Leaf-FMA: In Fig. 6, after the MN moves (Step 7) to another subnet (subnet2) in the same foreign domain, the new AP sends an ADD-notify message (Step 8). The leaf-FMA checks its "visitor list" and "VMAC list" according to the MN's MAC address in the ADD-notify message (Step 9) and determines whether the MN has just started up. At this point, the MN has not configured a new CoA or sent a Hy-BU, but it can still receive packets sent to its old CoA. If the leaf-FMA finds that the MN has performed a handoff, it sets a routing rule at the top of its routing table, and redirects packets sent to the MN's old CoA (CoA1) to subnet 2 (Step 11), so that the MN can receive packets. Once the leaf-FMA detects a new Hy-BU (Step 12), it updates the "visitor list" and the "VMAC list" according to the contents of the packet and cancels the routing rule for the MN. The packet carrying the Hy-BA is not illustrated in Fig. 6. After the registration is completed, packets sent to the MN's CoA2 can be correctly routed to the MN in subnet 2 (Step 13).

Handoff Between Different Leaf-FMA: If an MN performs a handoff between different leaf-FMAs, the newly attached leaf-FMA cannot determine whether the MN has just started up or is just performing a handoff when it receives ADD-notify. Hence, every leaf-FMA also has to send an MNinfoUpdate message at Steps 5 and 12. MNinfoUpdate is used to update the ancestor FMAs' visitor list and VMAC list with information about the MN (e.g., the MAC).

When a leaf-FMA receives ADD-notify, it sends a NewMNconfirm request to its ancestor FMA (Step 9-2 in Fig. 7). The FMA that has the information about the MN is a switching-FMA (Step 9-3). It sends a NewMNconfirm reply back (Step 9-4) and sets the status field to "SetRoutingRule." Otherwise, the root-FMA will reply and set the status field to "NewMN." If the NewMNconfirm reply indicates that the MN has performed a handoff, every FMA on the request/reply path (including the switching-FMA) should set the routing rule for the MN (Step 9-5), as described in Section IV-B. At Step 11, the switching-FMA and the FMAs all the way to subnet 2 will redirect packets sent to the MN's CoA1 to subnet 2.

Using HiMIPv6+, packets sent to the MN's old CoA are temporarily redirected to subnet 2 before the MN completes the new registration; thus, packet loss can be minimized during handoff. There is an additional advantage to using HiMIPv6+. Because the leaf-FMA knows the address of the MN's old AP, even though the new AP and old AP belong to different subnets, the new AP can still contact the old AP for the MN's parameters of authentication, authorization, QoS, etc., as illustrated in Step 10 in Fig. 6. In contrast, pure IAPP can only be used in a LAN environment.

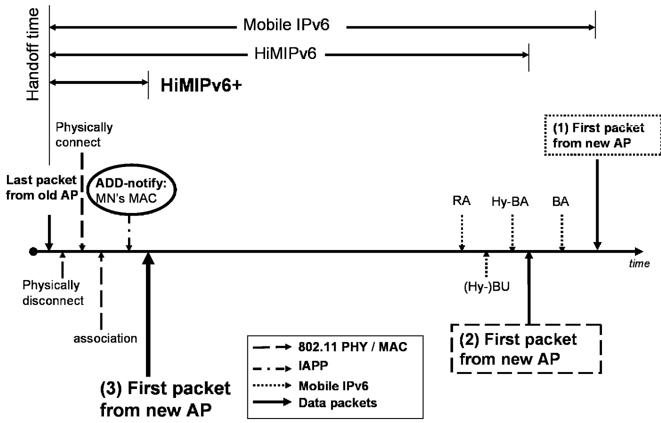

Fig. 8. Handoff times of protocols.

The periods which make up the latency are shown in Fig. 8: 1) for Mobile IPv6, the latency consists of the waiting period of RA and a round-trip time (RTT) of BU and BA; 2) HiMIPv6 has localized the registration message, thus, the handoff time is reduced; and 3) in HiMIPv6+, the leaf-FMA is aware of the MN's appearance in new subnet much earlier. The FMAs can then redirect the packets bound for the MN's old CoA to new subnet and, thus, substantially reduce the packet loss caused by handoff.

IV. Performance Evaluation

A. HiMIPv6 and HiMIPv6+

Having implemented a prototype to evaluate the performance of HiMIPv6+, we now discuss the results. With the RA interval set to 3 $\sim$ 4 s, Fig. 9(a) shows that the latencies of handoffs. The average latency of HiMIPv6 and HiMIPv6+ for intersubnet handoff is 3602 and 455 ms, respectively, i.e., HiMIPv6+ reduces the latency by approximately 87%. The latency time standard deviation of HiMIPv6 and HiMIPv6+ is 908.83 and 280.91, respectively. The former deviates so much because the MN has to wait for the arrival of the RA to configure a new address before completing the handoff procedure, which is not necessary in our proposed approach. The reduction of packet loss during intersubnet handoff can be clearly seen in Fig. 9(b). In HiMIPv6, the MN losses packets from its handoff start time (at $\mathrm{time} = 1$) until the handoff is complete, while in HiMIPv6+, the MN only suffers from a small packet loss right after the handoff (from subnet 1 to 2) at $\mathrm{time} = 1$, and immediately receives packets destined for its old CoA, and consequently, the MN starts to receive packets for new CoA after registration at $\mathrm{time} = 4.2$.

To evaluate how handoff latency affects transmission control protocol (TCP) performance, a TCP traffic flow is generated to the MN. The flow speed is 8 kb/s, and the layer 4 payload size of each packet is 20 bytes. The RA interval is set to be $1 \sim 2$ s. Fig. 10 shows the throughput of MN performing an intersubnet handoff. There is minor improvement from Fig. 10(a) Mobile IPv6 to Fig. 10(b) HiMIPv6, as HiMIPv6 only localizes the register messages. Fig. 10(c) shows that HiMIPv6+ can ease the TCP performance degradation while performing intersubnet handoff.

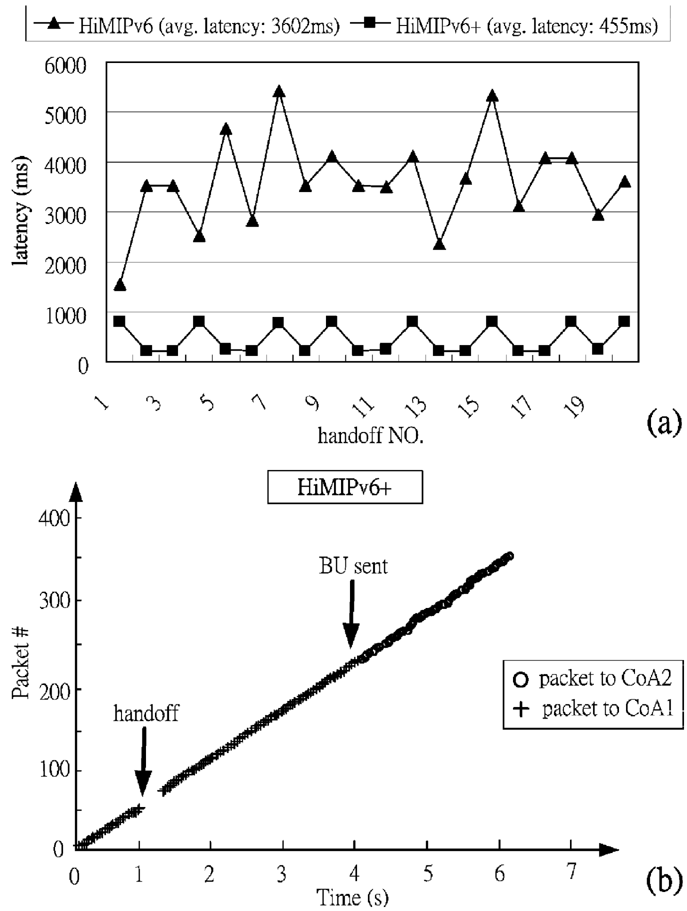

Fig. 9. Comparison of HiMIPv6 and HiMIPv6+.

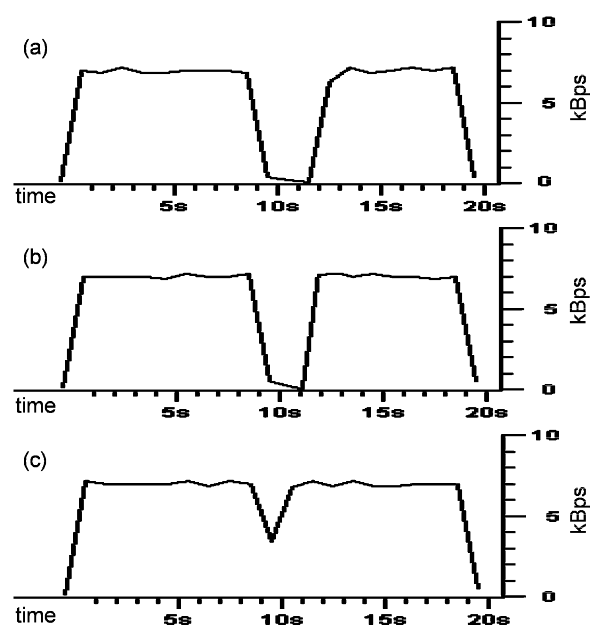

Fig. 10. TCP performances of protocols.

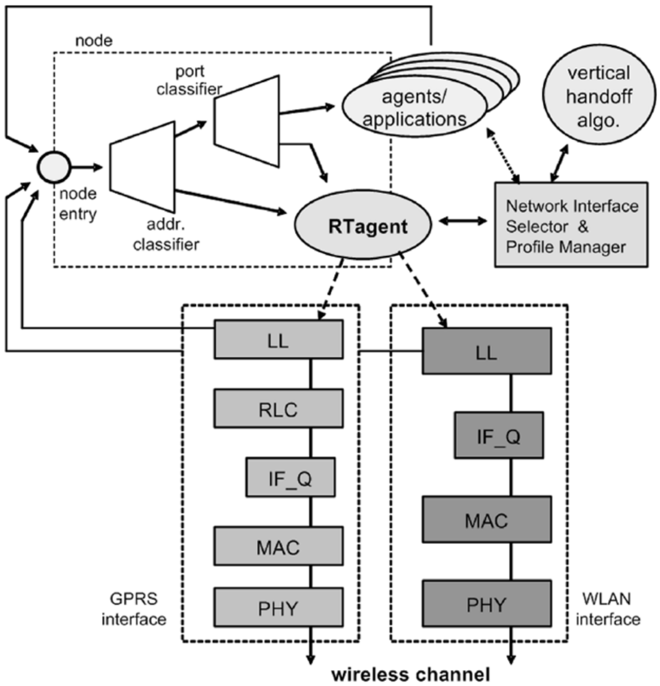

Fig. 11. MN and network interface selector.

B. Network Interface Selector

If an MN has multiple wireless interfaces, it should be able to switch among the interfaces during the communication period. Managing the wireless interfaces and minimizing the impact of switching on the communication quality are important issues in a heterogeneous multiaccess network. In this paper, we refer to the concept of multiple interface management in [21] and design a network interface selector for the MN. We also implement the interface selector in NS2 for later simulation. By using the network interface selector, the MN can choose a suitable wireless interface and decide the right time for vertical handoff by applying the proposed handoff decision algorithm.

Fig. 11 shows a schematic drawing of a WLAN and GPRS dual-interface mobile device with the proposed network interface selector. The network interface selector, positioned between the application layer and the MAC layer, cooperates with the routing agent (RTagent) to make decision about switching network interface and triggers vertical handoff through the vertical handoff algorithm. To achieve a QoS-based vertical handoff, the interface selector needs to collect information about the outer wireless environment and the inner operation modules to use as parameters for the vertical handoff decision algorithm. The immediate information about the outer air interface techniques is collected through each specific network interface and is forwarded to the network interface selector. The information about the inner operation modules includes configurations of the MN and basic rules of the application agents. When vertical handoff occurs, the network interface selector will control the RTagent to switch the interface of selected network so that the original interface may be turned off, or changed to the sleep mode.

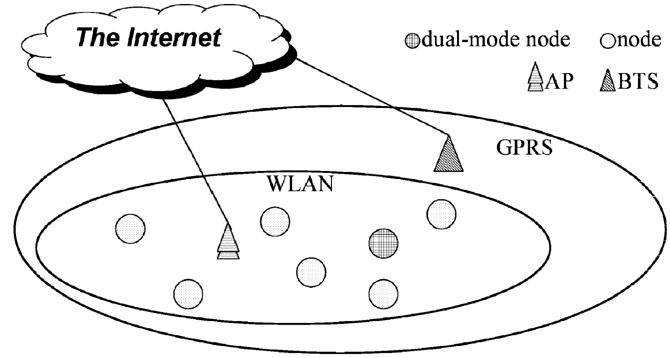

Fig. 12. NS2 simulation topology.

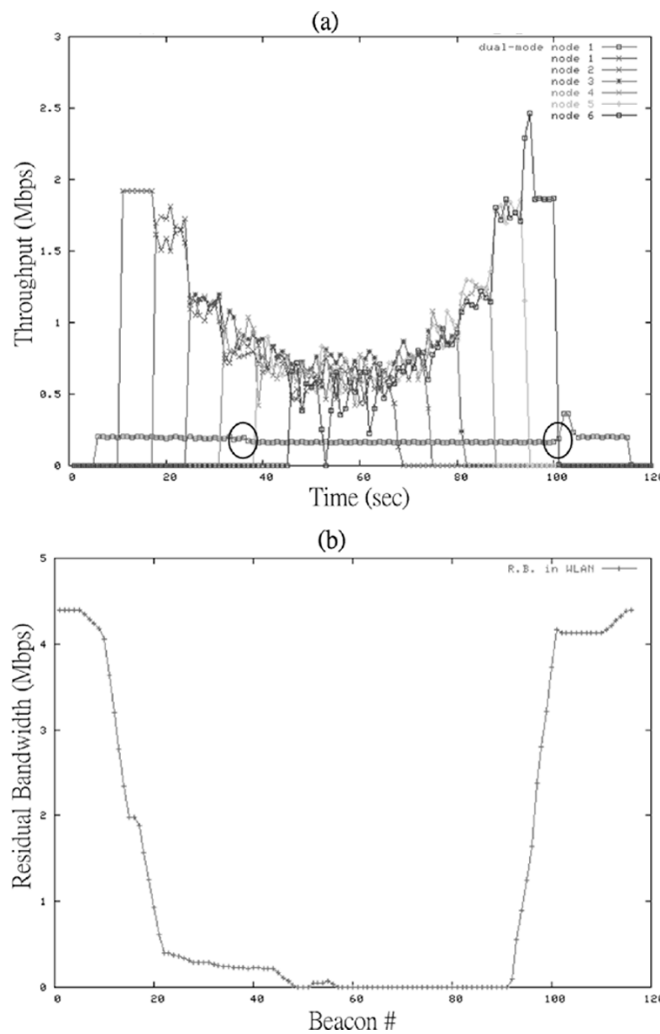

Fig. 13. Vertical handoff due to poor QoS. (a) MN throughput and other WLAN sessions throughputs. (b) Residual bandwidth of WLAN.

We construct the simulation environment with WLAN-GPRS on NS2. The WLAN is in the coverage area of GPRS and the dual-mode MN initially uses the WLAN, as shown in Fig. 12. The MN in our simulation has two network interfaces and vertical handoff is controlled by the network interface selector. Six nodes in the WLAN start user datagram protocol (UDP) traffic flows one by one, and then turn them off one by one.

Fig. 13(a) shows the throughput (the lowest line) of a dual-mode MN which performs handoff to GPRS from the WLAN due to the high contention in the WLAN. The MN performs vertical handoff to GPRS at 36.5 s. As the UDP traffic flows stop, the MN performs handoff back to the WLAN at 100.8 s. Fig. 13(b) shows the calculated residual bandwidth of the WLAN. The Simulation results show that by performing vertical handoff, the MN avoids QoS degradation, and switches to a lower priced network, i.e., to a WLAN, according to its preference when the provided QoS is acceptable.

V. Conclusion

In this paper, we have presented an overview of issues related to horizontal and vertical handoffs. We have also discussed the architecture of integrated WLAN and WAAN networks based on Mobile IPv6. The proposed mobility management scheme, HiMIPv6+, reduces the signaling overhead on the Internet and minimizes packet loss for the MN during handoff. It also eases the loads of the HA and CN, and can interoperate with Mobile IPv6 nodes.

With regard to the handoff decision algorithm, we have proposed a bandwidth measurement of WLAN. The QoS-based vertical handoff algorithm takes wireless network transport capability and user service requirement into account. With the algorithm, the MN can select a suitable access network and make the vertical handoff at the right time. The prototype evaluations and the simulations show that our framework performs as expected. Considering that bandwidth may not be enough for QoS, parameters like RTT, which reflects the degree of congestion, and the packet loss rate [22] can be used to enhance our algorithm in the future. The prototype evaluations and the simulations show that our framework performs as expected.

Acknowledgment

We would like to thank C.-C. Huang and F. Liu of the Department of Information Management, National Taiwan University, for their contributions to the design and prototyping of HiMIPv6.

References

[1] Wireless World Research Forum. [Online]. Available: http://www.wireless-world-research.org/

[2] M. Stemm and R. H. Katz, "Vertical handoffs in wireless overlay networks," in *Proc. ACM MONET*, 1998, pp. 335–350.

[3] K. Pahlavan, P. Krishnamurthy, and A. Hatami, "Handoff in hybrid mobile data networks," *IEEE Pers. Commun.*, vol. 7, no. 2, pp. 34–47, Apr. 2000.

[4] I. F. Akyildiz, J. McNair, J. S. M. Ho, H. Uzunalioglu, and W. Wang, "Mobility management in next-generation wireless systems," *Proc. IEEE*, vol. 87, pp. 1347–1384, Aug. 1999.

[5] D. Johnson, C. Perkins, and J. Arkko, "Mobility support in IPv6," IETF, RFC3775, 2004.

[6] T. Narten, E. Nordmark, and W. A. Simpson, "Neighbor discovery for IP version 6 (IPv6)," IETF, RFC 2461, 1998.

[7] R. Ramjee, K. Varadhan, L. Salgarelli, S. Thuel, S. Y. Wang, and T. La Porta, "HAWAII: A domain-based approach for supporting mobility in wide-area wireless networks," in *Proc. IEEE Int. Conf. Netw. Protocols*, Jun. 1999, pp. 396–410.

[8] A. T. Campbell, J. Gomez, S. Kim, A. G. Valko, C.-Y. Wan, and Z. R. Turanyi, "Design, implementation, and evaluation of cellular IP," *IEEE Pers. Commun.*, vol. 7, no. 4, pp. 42–49, Aug. 2000.

[9] H. Soliman, C. Castelluccia, K. El-Malki, and L. Bellier, "Hierarchical MIPv6 mobility management," *IETF Mobile IP Working Group*, Sep. 2000.

[10] "Requirements and Architecture for Interworking Between HIPERLAN/3 and 3rd Generation Cellular Systems," ETSI, Tech. Rep. ETSI TR 101 957, 2001.

[11] A. Salkintzis, C. Fors, and R. Pazhyannur, "WLAN-GPRS integration for next-generation mobile data networks," *IEEE Wireless Commun.*, vol. 9, no. 5, pp. 112–124, Oct. 2002.

[12] S. Tsao and C. Lin, "VGSN: A gateway approach to interconnect UMTS/WLAN networks," in *Proc. IEEE Int. Symp. Pers., Indoor, Mobile Radio Commun.*, 2002, pp. 275–279.

[13] M. Bduuhikot, G. Chandranmenon, S. Han, Y. W. Lee, S. Miller, and L. Salgarelli, "Integration of 802.11 and third-generation wireless data networks," in *Proc. INFOCOM*, Mar.-Apr. 2003, pp. 504–512.

[14] R. Koodli, "Fast Handovers for Mobile IPv6," IETF, IETF draft-08, 2003.

[15] *Recommended Practice for Multi-Vendor Access Point Interoperability via an Inter-Access Point Protocol Across Distribution System Supporting IEEE 8.2.11 Operation*, IEEE Standard 802.11f/D5, Jan. 2003.

[16] H. Park, S. Yoon, T. Kim, J. Park, M. Do, and J. Lee, "Vertical handoff procedure and algorithm between IEEE802.11 WLAN and CDMA cellular network," in *Proc. CDMA Int. Conf.*, 2002, pp. 103–112.

[17] M. Ylianttila, M. Pande, J. Makela, and P. Mahonen, "Optimization scheme for mobile user performing vertical handoffs between IEEE 802.11 and GPRS/EDGE networks," in *Proc. GLOBECOM*, Nov. 2001, pp. 3439–3443.

[18] S. Sharma, I. Baek, Y. Dodia, and T.-C. Chiueh, "OmniCon: A mobile IP-based vertical handoff system for wireless LAN and GPRS links," *Mobile Wireless Netw.*, pp. 330–337, 2004.

[19] *IEEE802.11 WG, Draft Supplement to Part 11, "Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specification: Medium Access Control (MAC) Enhancements for Quality of Service (QoS),"*, IEEE Standard 802.11e/D4.3, May 2003.

[20] Quality of Service. UMTS. [Online]. Available: http://www.umtsworld.com/technology/qos.htm

[21] F. Andre, J. M. Bonnin, B. Deniaud, K. Guillouard, N. Montavont, T. Noel, and L. Suciu, "Optimized support for multiple wireless interfaces within an IPv6 end-terminal," presented at the Smart Objects Conf., Grenoble, France, May 15–17, 2003.

[22] S. Kashihara, T. Nishiyama, K. Iida, H. Koga, Y. Kadobayashi, and S. Yamaguchi, "Path selection using active measurement in multi-homed wireless networks," *Proc. SAINT*, pp. 273–276, 2004.

Chen Wee Lee was born in Chiayi, Taiwan, in 1979. He received the B.S. and M.S. degrees from the Department of Information Management, National Taiwan University, Taipei, Taiwan, in 2002 and 2004, respectively. Currently, he is working towards the Ph.D. degree in the Department of Information Management, National Taiwan University.

His research interests include mobile Internet, cellular communication systems, and wireless mesh networks.

Li-Ming Chen received the B.S. degree in computer science from National Chiao-Tung University, Hsinchu, Taiwan, in 2002 and the M.S. degree in computer science from National Taiwan University, Taipei, Taiwan, in 2004.

He is a Research Assistant with the Institute of Information Science, Academia Sinica, Taipei, Taiwan. His research interests are in the areas of wireless communication, computer networks, and operation systems.

Dr. Meng Chang Chen received the B.S. and M.S. degrees in computer science from National Chiao-Tung University, Hsinchu, Taiwan, and the Ph.D. degree in computer science from the University of California (UCLA), Los Angeles, in 1989.

He joined AT&T Bell Laboratories in 1989 as Member of Technical Staff, and became an Associate Professor with the Department of Information Management, National Sun Yat-Sen University, Kaohsiung, Taiwan, in 1992. Since 1993, he has been with the Institute of Information Science, Academia Sinica, Taipei, Taiwan. He has held an Associate Research Fellowship since July 1996, and assumed the responsibility of Deputy Director from August 1999 to July 2002. From 2001 to the end of 2003, he served as the Chair of Standards and Technology Transfer Group, National Science and Technology Program for Telecommunications Office (NTPO). His current research interests include quality-of-service (QoS) networking, wireless networks, operating systems, knowledge discovery and management, and multimedia systems and networking.

Yeali Sunny Sun was born in Taipei, Taiwan. She received the B.S. degree in computer science from National Taiwan University, Taipei, Taiwan, in 1982, and the M.S. and Ph.D. degrees from the University of California (UCLA), Los Angeles, in 1984 and 1988, respectively, both in computer science.

From 1988 to 1993, she was with Bell Communications Research, Inc. (Bellcore; now Telcordia), where she was involved in the area of planning and architecture design of information networking, broadband networks, and network and system management. In August 1993, she joined National Taiwan University and is currently a Professor and Chairperson of the Department of Information Management. She was the Director of the Information Networking Group, Computer Center, National Taiwan University, responsible for managing the university campus network and Taiwan Academia Networks (TANet) Northern Regional Point of Presence (POP) and leading an advanced networking research team prototyping and conducting a series of field trials and experiments on policy-based QoS delivery services, bandwidth management, and Internet pricing for congestion control. From 1996 to 2002, she served on the TANet Technical Committee, Steering Committee of the National Broadband Experimental Network (NBEN) and Internet2, and the IP Committee of TWNIC. Her research interests are in the area of mobile Internet, quality-of-service (QoS), content classification, wireless mesh networks, multimedia content delivery, Internet pricing and network management, and performance modeling and evaluation.