Mobility and QoS Support in 4G Wireless Networks

Taehyoun Kim and Jaiyong Lee

Abstract: **Fourth-generation (4G) wireless networks will be the IP-based cellular networks integrating Internet with the existing cellular networks. Two important issues should be concerned in the IP-based cellular networks, IP mobility, and quality-of-service (QoS) guarantees. In this paper, we proposed two mechanisms to solve the problems with IP mobility and RSVP-based QoS provisioning. First, virtual-IP (VIP) allocation scheme in areas with a large rate of handoff can minimize the wireless signaling overhead due to IP mobility. The access routers (ARs) create dynamically the VIP zone by using the measured handoff rate derived from the history of the handoff into neighboring ARs. We show that VIP allocation scheme reduces the binding update messages in the wireless link than hierarchical mobile IPv6. Second, the new advance resource reservation protocol called proportional aggregate RSVP (PA-RSVP) can minimize waste of bandwidth and soft state refresh overhead due to IP mobility. It allocates the bandwidth in advance between the mobility anchor point and neighboring ARs using proportional aggregate reservation. We also show that PA-RSVP provides an improved performance over existing protocols.**

Index Terms: **4G wireless networks, advance resource reservation, IP mobility, PA-RSVP, QoS, virtual-IP, wireless signaling.**

I. INTRODUCTION

In the Internet protocol (IP)-based cellular networks, an access router (AR) resides in an access network and is connected to one or more access points (APs). These ARs use IP protocols for data transport and signaling in the fourth-generation (4G) wireless networks. The IP-based cellular networks have the advantages of directly applying IP techniques and applications written for the wired data networks to the wireless networks. Two important issues should be concerned in the IP-based cellular networks: IP mobility and quality-of-service (QoS) guarantees [1], [2].

Mobile IPv6 [3] enables a mobile node (MN) to keep the network connectivity even if the MN changes its point of attachment to the Internet. However, mobile IPv6 suffers from the problem with a large signaling overhead for frequent binding updates. Hierarchical mobile IPv6 (HMIPv6) [4]–[6] is an enhanced mobile IPv6 to minimize the signaling overhead in the wired network link. But, HMIPv6 cannot reduce the signaling overhead in the wireless link. In the mobile Internet, the wireless link has far less available bandwidth resources and limited scalability compared to the wired network link. Therefore, the signaling overhead due to IP mobility has a severe effect on the wireless link [7]. Moreover, each cell becomes smaller and this increases handoff rates yielding more signaling overhead in the wireless link [8]–[10].

Manuscript received February 15, 2005.

The authors are with the Department of Electrical and Electronic Engineering, Yonsei University, 134, Seoul, 120-749, Korea, email: {tuskkim, jyl}@nasla.yonsei.ac.kr.

IPv6 supports DiffServ-style QoS, but real-time services such as streaming audio and video services [11], [12] would be much better served under the RSVP/IntServ model [13]–[15] as they have a relatively constant bandwidth requirement for a known period of time. However, the delay of reserving resources of RSVP along new data paths after handoffs may cause service disruptions for real-time multimedia services. Therefore, in order to provide seamless QoS services, advance resource reservations are necessary on the cells that are likely to be visited by the MN. Recently, many researches have been done to provide QoS in the wireless networks using advance resource reservations. However, while these proposals provide seamless QoS guarantees, they have problems with waste of bandwidth due to advance resource reservations. Also, RSVP suffers from a scalability problem since per-flow reservation states have to be periodically refreshed. Eventually, we need to find new method to provide seamless QoS guarantees while minimizing the use of the pre-provisioning resources.

In this paper, we propose two mechanisms to solve the problems with IP mobility and RSVP-based QoS provisioning in the IP-based cellular networks. First, virtual-IP (VIP) allocation scheme in areas with a large rate of handoff can minimize the wireless signaling overhead due to IP mobility. The access routers (ARs) dynamically create the VIP zone by using the measured handoff rate to neighboring ARs. Even if the handoff of the MNs occurs between the ARs within the same VIP zone, the current care-of-addresses (CoAs) of the MNs are not changed. As a result, local binding updates are not generated in the VIP zone. Therefore, VIP allocation scheme has many benefits: Saving of the wireless network resource, reduction of the power consumption of the MN, and reduction of the interference in the wireless link. Second, the new advance reservation protocol called proportional aggregate RSVP (PA-RSVP) can minimize waste of bandwidth and soft state refresh overhead due to IP mobility. It allocates the bandwidth in advance between the mobility anchor point (MAP) and neighboring ARs using proportional aggregate reservation. Also, PA-RSVP has many benefits: Saving of the network resource, reduction of the maintenance cost, and avoidance of the network congestion.

The remainder of the paper is organized as follows. In Section II , we describe the mobility issues in 4G wireless networks. In Sections III and IV, we describe virtual-IP (VIP) allocation scheme and proportional aggregate RSVP (PA-RSVP) in the IP-based cellular networks, respectively. Finally, some conclusions are summarized in Section V.

II. MOBILITY ISSUES IN 4G WIRELESS NETWORKS

In this section, we present the problems of IP mobility protocol and resource reservation protocol in 4G wireless networks. The purpose of this discussion is to set the stage for the following sections.

A. IP Mobility Protocols

Since IPv6 is becoming a standard for 4G networks, mobile IPv6 [3] has been developed by the IETF with new functionalities for streamlining mobility support that are missing in mobile IPv4. In mobile IPv6, when an MN moves from a network coverage cell to another cell, it gets a care-of address (CoA) from the visited network. After receiving the CoA, the MN registers the association between the CoA and the home address by sending a binding update message to its home agent (HA) or correspondent node (CN). The HA and the CN each keeps a binding cache that is updated when new binding update arrives. In mobile IPv6, when the HA and the MN are far from each other, even small MN movements create the binding update messages that traverse a long the way across the network. This signaling load for the binding update in mobile IPv6 may become very significant as the number of the MNs increases.

To overcome this problem, hierarchical mobile IPv6 (HMIPv6) [4] is proposed to reduce the amount of signaling that is required for managing the MN movement within the mobile networks. HMIPv6 uses a local anchor point called mobility anchor point (MAP). An MN entering a MAP domain will receive the router advertisement containing information on local MAP. The MN can bind its current CoA (on-link CoA; LCoA) with an CoA on the MAP's subnet (regional CoA; RCoA). Acting as a local HA, the MAP encapsulates and forwards them directly to the MN's current address (i.e., LCoA). When the MN moves within a local MAP domain, the MN sends the binding update messages only up to the MAP. This reduces additional signaling cost in the wired network link between the MAP and the CN that exists in mobile IPv6. But, HMIPv6 cannot reduce the binding update messages in the wireless link. In addition, IETF proposed the fast handoff over HMIPv6 [7], [16] that integrates HMIPv6 and the fast handoff mechanism to reduce the handoff latency by address pre-configuration. Since the fast handoff over HMIPv6 inherits the basic signaling structure of HMIPv6, the signaling cost in the wireless link is unchanged from HMIPv6.

B. Resource Reservation Protocols

The simple QoS protocol [17] integrates RSVP tunnel and the mobile IP routing mechanism. This protocol is practical in use and can be implemented easily. When an MN moves to a foreign network, it informs its HA of its new location information. When the HA knows the MN's location, it sets up an RSVP tunnel between itself and the foreign agent (FA), and encapsulates PATH messages from the sender and transmits them to the MN through the tunnel. The simple QoS protocol has advantages of requiring a little modification in basic mobile IP, low bandwidth requirements, and reduced maintenance cost. But, RSVP tunnel still exists in simple QoS protocol and it causes a triangular routing problem. This means the resource reservation path between the sender and the MN always includes the HA. The packets always have to go through the HA and, as a result, the routing path is not optimized. Therefore, the MN may not be able to receive packets because of the handoff delay caused by the non-optimal routing path. Also, service disruptions may occur if there is not enough resource reserved on the new data path. Moreover, the signaling overhead due to RSVP tunnel exists.

Mobile RSVP (MRSVP) [18] protocol reserves resources in advance on the cells that the MN is expected to visit. MRSVP protocol defines mobility specification (MSPEC) which includes a list of CoAs that the MN is expected to visit during the lifetime of the flow. This protocol has two classes—active reservation and passive reservation. An active reservation is used on the current active communication path and is setup from the sender to the MN via the proxy agent. On the other hand, the passive reservations are used on the communication paths that are not currently active and are setup from the sender to the proxy agents in the other locations in its MSPEC. After the setups of the active and the passive reservations, when the MN moves to a new location within MSPEC, the active reservation switches to a passive reservation and the passive reservation on the new location switches to an active reservation. Hence, MRSVP can overcome service disruptions during handoffs by making advance reservations on neighboring cells that are expected to be visited by the MN. However, since this protocol makes reservation on the path to all cells that are expected to be visited by the MN, the network bandwidth is wasted. It also requires the transfer of periodic refresh messages to maintain the soft state which increases the signaling overhead.

In RSVP-RA protocol [19], multiple FAs exist under a RSVP agent (RA). When the sender sends a PATH message to the MN using RSVP, the RA that is managing the MN intercepts the PATH message and sends it to neighboring FAs to establish resource reservations in advance. RSVP-RA protocol has two classes that are similar to those of MRSVP—reserved reservation and prepared reservation. The reserved path is the reserved resource along flow from the sender to the MN. Data packets are actually transmitted through the reserved reservation links to the MN. The prepared path is the reserved resource from an RA to the neighboring cells of an MN's current cell. Although data packets from the sender to the MN are not currently being transmitted through the prepared reservation paths, it is needed to prepare for an MN's handoff to a neighboring cell. After the setups of the reserved and the prepared reservations, when the MN moves to a new location within neighboring cells, the reserved reservation switches to a prepared reservation and the prepared reservation on the new location switches to a reserved reservation. Different from MRSVP protocol, RSVP-RA protocol uses RA to support mobility at the network boundary without leaving any burden to the Internet backbone. RSVP-RA protocol reduces waste of network bandwidth and signaling overhead compared to MRSVP protocol. Yet, as RSVP-RA reserves resources in advance on the access networks, it still has problems of waste of network bandwidth and signaling overhead.

III. VIRTUAL-IP ALLOCATION SCHEME

The network architecture of virtual-IP (VIP) allocation scheme is based on HMIPv6 [4] as shown in Fig. 1. In order to apply VIP allocation scheme in the IP-based cellular networks, we need a set of MNs supporting HMIPv6, the ARs supporting IP routing and connected with several access points (APs), and the MAP that act as a local anchor point in HMIPv6. For a reference, the HA and the CN operate as defined in mobile IPv6. The network entities are defined as follows.

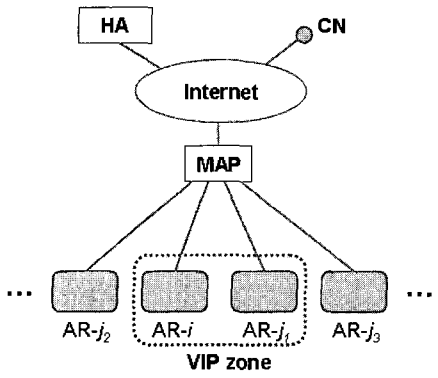


Fig. 1. Reference architecture of VIP allocation scheme.

- *Mobility anchor point (MAP)*: It commands to corresponding ARs to create or release a VIP zone and routes the packets of the MN to the new AR.
- *Access router (AR)*: It counts the number of handoffs to its neighboring ARs during a specified period of time. When it detects that the measured handoff rate exceeds the threshold value or drops below the threshold value, it sends its status and IP address of its neighboring AR into the MAP. According to the command of the MAP, it sends the virtual network prefix or the original network prefix to the MNs.
- *Virtual CoA (VCoA) and local CoA (LCoA)*: The VCoA and the LCoA each is configured on the MN based on the virtual network prefix and the original network prefix advertised by the AR.
- *Regional CoA (RCoA)*: It is configured by the MN when it receives the MAP option.
- *Movement update*: When the MN moves into a new AR within the VIP zone, old AR sends a movement update message to the MAP in order to establish the binding caches among RCoA, VCoA, and new AR IP address.

A. Method of Establishing the Virtual-IP Zone

To explain the concept of the VIP zone, we illustrate the VIP zone between the AR-i and the AR-j_1 in Fig. 1. Each AR monitors the movement status that has a trace of handoff history to its neighboring ARs during active communications and counts the number of handoffs. In this framework, the ARs create the VIP zone by using the measured handoff rate derived from the history of the handoff in the ARs. The VIP zone is created and released by two conditions described as follows. If the condition in (1) is met, the AR-i send a request to the MAP to create the VIP zone with the AR-j_1, and the MAP commands the AR-i and the AR-j_1 to create a VIP zone. Therefore, a VIP zone is formed at the AR-i and the AR-j_1.

$$rateHO(i, j_1) \geq THi(VC), \tag{1}$$

where *rateHO*(i,j_1) is the handoff rate from the AR-i to the AR-j_1 during a certain period. *THi(VC; VIP creation)* is the threshold value of the handoff to create a VIP zone in the AR-i.

If the condition in (2) is met, the AR-i sends a request to the MAP to release a VIP zone with AR-j_1, and the MAP commands the AR-i and the AR-j_1 to release a VIP zone. Therefore, a VIP zone is released at the AR-i and the AR-j_1.

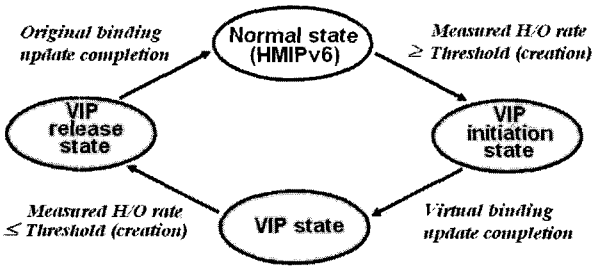


Fig. 2. State diagram of VIP allocation scheme.

$$rateHO(i, j_1) \leq THi(VR), \tag{2}$$

where *THi(VR; VIP release)* is the threshold value of the handoff to release a VIP zone in the AR-i. Here, *THi(VR)* is sufficiently different from *THi(VC)* for some hysteresis (i.e., ensure that the trigger condition for the initiation of a VIP zone is sufficiently different from the trigger condition for the release of a VIP zone) to avoid oscillation in stable condition.

B. Description of Virtual-IP Allocation Scheme

As shown in Fig. 2, virtual-IP (VIP) allocation scheme operates in four states. First, the normal state operates as HMIPv6. If the VIP creation condition is met in the normal state, the state switches to the VIP initiation state. Second, in the VIP initiation state, the ARs send the virtual network prefixes to the MNs in their area and switches to the VIP state. Third, in the VIP state, the VIP zones are created by the ARs with the same virtual network prefix. As a result, local binding updates are not generated and this greatly reduces the signaling cost in the wireless link. At this time, if the VIP release condition is met in the VIP state, the state switches to the VIP release state. Finally, the VIP release state switches to the normal state by sending different original network prefixes in each AR of the VIP zone.

Each router has an original network prefix and a virtual network prefix. The original network prefix is distinctly assigned to every AR within a MAP domain by the IP address class. But, the virtual network prefix is assigned per MAP domain. It is identically assigned to every AR within a MAP domain. In a situation where the AR-j_1 is part of one VIP zone 1, and the AR-j_3 is part of another VIP zone 2 in Fig. 1, if AR-j_1 and AR-j_3 are neighbors ARs, the VIP zone 1 and the VIP zone 2 are merged in one VIP zone because the VIP zone 1 and the VIP zone 2 each is assigned the same virtual network prefix within the MAP domain. As a result, when an MN under AR-j_1 moves to AR-j_3, it receives a same virtual network prefix under AR-j_3 as it did under AR-j_1 and a new binding update is not created. On the other hand, if the VIP zone 1 and the VIP zone 2 use different virtual network prefixes, then the VIP zones do not merge. Consequently, as the MN received new network prefixs when an MN under AR-j_1 moves to AR-j_3, the signaling cost is increased by the new binding updates. Therefore, a better performance is achieved if a common virtual network prefix is assigned to several VIP zones within the MAP domain.

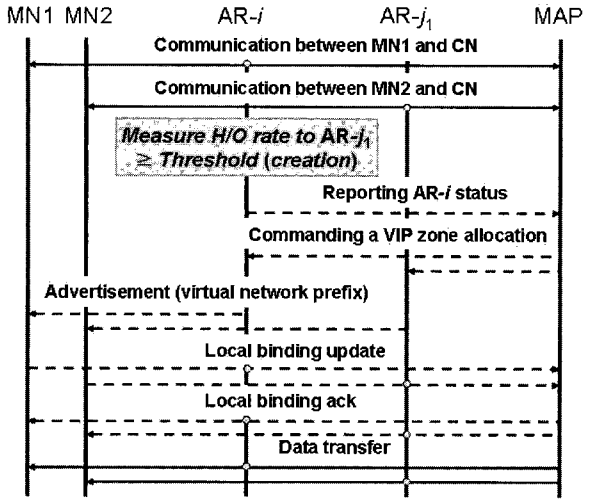


Fig. 3. Message flow in VIP initiation state.

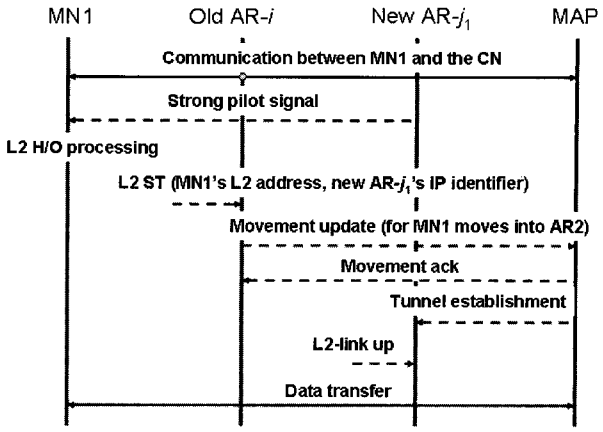


Fig. 4. Message flow in VIP state.

Meanwhile, it is possible to apply VIP allocation scheme statically. That is, the handoffs can be carried out using L2-source trigger in all areas independent of the threshold values (i.e., *THi(VC)* and *THi(VR)*). However, VIP allocation scheme becomes meaningless if the utilization of the wireless link does not reach to the maximum capacity of the wireless link in the AR. Therefore, reducing the signaling cost in the wireless link is meaningful if the VIP zones are dynamically allocated during the times when the MNs are densely populated and the handoff rate increases. For this reason, although VIP allocation scheme can operate statically, it operates dynamically in practice.

B.1 Normal State

In the normal state, it operates as HMIPv6 proposed by IETF until the VIP creation condition is met in the normal state.

B.2 VIP Initiation State

Fig. 3 shows the message flow in the VIP initiation state. The MN1 and the MN2 are communicating with the AR-i and the AR-j_1 in the normal state, respectively. In this state, the binding cache of the MN1 is {RCoA1: LCoA1} and the binding cache of the MN2 is {RCoA2: LCoA2}. Also, the binding caches of the MAP each is {RCoA1: LCoA1} and {RCoA2: LCoA2}. When the AR-i detects that the measured handoff rate exceeds the threshold value for the creation of the VIP zone, the AR-i sends its status and IP address of the AR-j_1 to the MAP. Then, the MAP commands to the AR-i and the AR-j_1 to create a VIP zone. If the AR-j_1 does not have a virtual network prefix, it can reject the creation request of the VIP zone from the MAP. However, these are unusual case that happens during operations. Hence, excluding some exceptional abnormal situations, each AR accepts the request from the MAP. The AR-i and the AR-j_1 each sends the same virtual network prefix instead of existing different original network prefix. The MN1 and the MN2 each receives this new network prefix and compare with its original network prefix, respectively. They recognize an arrival of a new network prefix and generate new virtual CoA1 (VCoA1) and VCoA2 by auto-configuration [20]. This mechanism causes the MN1 and the MN2 to perceive as if they are separately being handed over to a new AR, and causes them to change their current CoA. The MN1 and the MN2 register newly acquired the VCoA1 and the VCoA2 to the MAP. The MAP updates those binding caches (as MN1 - {RCoA1: VCoA1: AR-i}, and MN2 - {RCoA2: VCoA2: AR-j_1}). Through this procedure, the VIP zone is performed on the AR-i and the AR-j_1 and the state then switches to the VIP state.

B.3 VIP State

In the VIP state, the packet data sent from the CN to an MN1 and an MN2 are encapsulated with new VCoA1 and VCoA2 by the MAP and forwarded to the MN1 and the MN2, respectively. Fig. 4 shows the message flow of the handoff in the VIP state. We consider that the MN1 in the AR-i is handovered to new AR-j_1. Here, the binding cache of the MN1 is {RCoA1: VCoA1}. Also, the binding cache of the MAP for the MN1 is {RCoA1: VCoA1: AR-i}. When the MN1 approaches new AR-j_1, it receives a strong pilot signal and performs layer 2 (L2) handoff. Here, old AR-i determines the IP address of new AR-j_1 using L2 source trigger (L2-ST) [16]. L2-ST includes the information such as the L2 identifier of new AR. The L2 identifier is translated into the IP address of the AR-j_1. This source trigger is not newly generated message from the MN. It is transferred using the existing message of the L2 handoff (e.g., it is the identifier of base station in CDMA or the identifier of access point in WLAN used at L2 handoff). Therefore, no additional signaling messages in L2 are generated during the handoff. Old AR-i detects the MN1 moving towards new AR-j_1 and sends a movement update message to the MAP. Then, the MAP updates its binding cache for the MN1 as {RCoA1: VCoA1: AR-j_1} and establishes a tunnel to new AR-j_1. New AR-j_1 sets up a host route for the VCoA1 of the MN1. After that, the MN1 moves to new AR-j_1 and sets up a L2 link after completing a L2 handoff. At the same time, since the MN1 receives the same virtual network prefix with old AR-i from new AR-j_1, it does not perform a binding update. The packet data sent from the CN to the MN1 on the MAP are encapsulated with the VCoA1 and for-

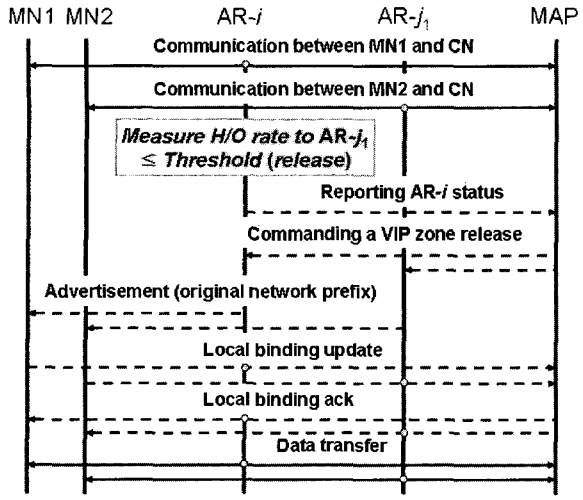


Fig. 5. Messages flow in VIP release state.

ward to the MN1 through new AR-j_1 by the MAP. Even if the MN1 is handed over to new AR-j_1 in the VIP zone, the binding update requests and acks need not to be sent over the wireless link. Hence, the signaling cost with binding update in the wireless link is greatly reduced. If an MN in the VIP zone moves to the AR out of the VIP zone or if an MN outside the VIP zone moves into the AR in the VIP zone, in only these two cases, an MN receives a different network prefix.

B.4 VIP Release State

An MN1 and an MN2 are communicating with the AR-i and the AR-j_1 respectively in the VIP state. In this state, the binding cache of the MN1 is {RCoA1: VCoA1} and the binding cache of the MN2 is {RCoA2: VCoA2}. Also, the binding caches of the MAP each is {RCoA1: VCoA1: AR-i} and {RCoA2: VCoA2: AR-j_1}. Fig. 5 shows the message flow in the VIP release state. When the AR-i detects that the measured handoff rate drops below the threshold value for the release of the VIP zone, the AR-i sends its status and the IP address of the AR-j_1 to the MAP. Then the MAP commands the AR-i and the AR-j_1 to release a VIP zone. Following the procedure, the AR-i and the AR-j_1 independently send different original network prefix instead of the same virtual network prefix, respectively. The MN1 and the MN2 each registers newly acquired LCoA1 and LCoA2 to the MAP. The MAP updates those binding caches as {RCoA1: LCoA1} and {RCoA2: LCoA2}. Through this procedure, the VIP release state is finished at the AR-i and the AR-j_1, and the state switches to the normal state.

C. Performance Evaluation

C.1 Analytic Mobility Model

We adopt the fluid flow model which is commonly used to analyze cell boundary crossing related issues in our case [21], [22]. The topology of the analysis consists of a MAP domain. The AR and the VIP zone are assumed to be square-shaped.

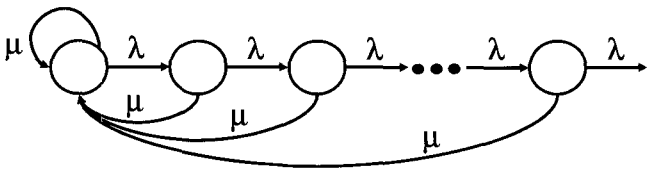


Fig. 6. Imbedded Markov chain for the handoff process.

Under the fluid-flow model, the direction of an MN movement is uniformly distribution in the range of $[0, 2\pi]$. The average crossing rate of the boundary of the AR and the VIP zone, $E[R_{\mathrm{AR}}]$ and $E[R_{\mathrm{VIP}}]$ each is

$$E[R_{\mathrm{AR}}] = \frac{4v}{\pi \cdot \ell_{\mathrm{AR}}}, \quad E[R_{\mathrm{VIP}}] = \frac{4v}{\pi \cdot \ell_{\mathrm{VIP}}}, \tag{3}$$

where v is the average speed of MN in km/hr and ℓ_{AR} and ℓ_{VIP} each is the side length of the AR (m) and the VIP zone (m).

Since an MN crossing the border of the VIP zone must cross the border of the AR, the AR crossing rate of the MN still staying in the same VIP zone is computed as

$$E[R_{\mathrm{AR}\in\mathrm{VIP}}] = E[R_{\mathrm{AR}}] - E[R_{\mathrm{VIP}}] = \frac{4v(\ell_{\mathrm{VIP}} - \ell_{\mathrm{AR}})}{\pi \cdot \ell_{\mathrm{AR}}\ell_{\mathrm{VIP}}} \tag{4}$$

The handoff process of an MN is modeled as an imbedded Markov chain, which illustrates an in Fig. 6. This model defines the handoff process of an MN within a VIP zone to show the number of the AR crossings by an MN before it moves out of the VIP zone. Also, a state of the chain is defined as the number of handoffs between the ARs in the same VIP zone before leaving it. The advantage of VIP allocation scheme is the reduction of the signaling cost when the MNs move between the ARs within the same VIP zone. Hence, in order to verify the performance enhancements by VIP allocation scheme, we need to find the crossing rate of the boundary of the AR until the MN moves out of its VIP zone and the crossing rate of the boundary out of the VIP zone. Here, λ represents the state transition rate at which the MN moves from state k to state $k+1$, $k = 0, 1, 2, \cdots$. In other words, the MN is traveling between the ARs within the same VIP zone at the rate λ. μ denotes the state transition rate at which the MN moves from state k to state 0. It represents the MN moving out of a VIP zone. Therefore, λ and μ each is expressed as

$$\lambda = E[R_{\mathrm{AR}\in\mathrm{VIP}}], \quad \mu = E[R_{\mathrm{VIP}}]. \tag{5}$$

The equilibrium state probability of state k, P_k is expressed as

$$P_k = \left[\frac{\lambda}{\lambda+\mu}\right]^k P_0, \quad P_0 = \frac{\mu}{\lambda+\mu}, \tag{6}$$

where P_0 is the equilibrium state probability of state 0.

Using the expression from (3) to (6), the average number that an MN moves out of the VIP zone, α and the average number that an MN moves within the domain, β each is expressed as ratio of the side of the AR to that of the VIP zone.

$$\alpha = P_0 = \frac{\ell_{\mathrm{AR}}}{\ell_{\mathrm{VIP}}}, \quad \beta = \sum_{k=1}^{\infty} kP_k = \frac{\ell_{\mathrm{VIP}}}{\ell_{\mathrm{AR}}} - 1. \tag{7}$$

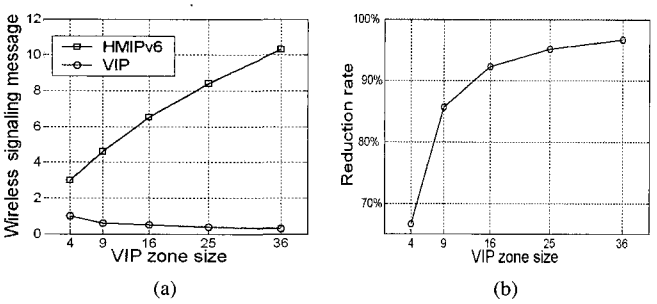


Fig. 7. Effect of VIP zone size on wireless signaling cost; (a) effect of VIP zone size, (b) reduction rate of VIP allocation scheme.

Therefore, the average wireless signaling cost in HMIPv6 and VIP allocation scheme can be expressed respectively as

$$C_{HMIPv6} = M_{WB}\cdot\alpha + M_{WB}\cdot\beta = M_{WB}\left(\frac{\ell_{\mathrm{AR}}}{\ell_{\mathrm{VIP}}} + \frac{\ell_{\mathrm{VIP}}}{\ell_{\mathrm{AR}}} - 1\right), \quad (8)$$

$$C_{HMIPv6} = M_{WB}\cdot\alpha = M_{WB}\cdot\frac{\ell_{\mathrm{AR}}}{\ell_{VIP}}, \quad (9)$$

where M_{WB} is the cost for the local binding update in the wireless link.

C.2 Analytic Results

In the wireless signaling analysis, the performance of HMIPv6 and VIP allocation scheme each is evaluated from the ratio of ℓ_{AR} to ℓ_{VIP}, which can be determined from the side length of the AR and the VIP zone, respectively. The VIP zone size in Fig. 7 represents the number of ARs that forms the VIP zone. Fig. 7(a) shows the analysis result of the wireless signaling cost depending on the VIP zone size. When the VIP zone size is 16 ARs, the wireless signaling cost is 6.5 in HMIPv6 and 0.5 in VIP allocation scheme. Also, the wireless signaling cost is 10.3 in HMIPv6 and 0.3 in VIP allocation scheme when the VIP zone size is 32ARs. The reason for this is whenever the MN changes its point of attachment, the binding update still occurs regardless of the change of the VIP zone size in HMIPv6. But, as the VIP zone size increases, the probability of moving out of the VIP zone gradually decreases in VIP allocation scheme. As a result, Fig. 7(b) shows that as the VIP zone size increases, VIP allocation scheme reduces the wireless signaling cost from 65% up to 95% compared with HMIPv6. Thus, the wireless signaling cost is drastically reduced when VIP allocation scheme is applied in the IP-based cellular networks.

C.3 Simulation Environments

We used ns-2.1b7a mobile IP [23] with a wireless extension for the simulation as shown in Fig. 8. The link characteristics, namely the bandwidth and the delay are also shown in Fig. 8. Constant bit rate sources are used as traffic sources. A source agent is attached to the CN and the sink agents are attached to the MNs. The duration of each simulation experiment is 180 simulation time units (sec). The cells within a MAP domain are divided into either a dense area or a sparse area. A dense area is composed of 4, 9, 16, 25, or 36 cells around a center

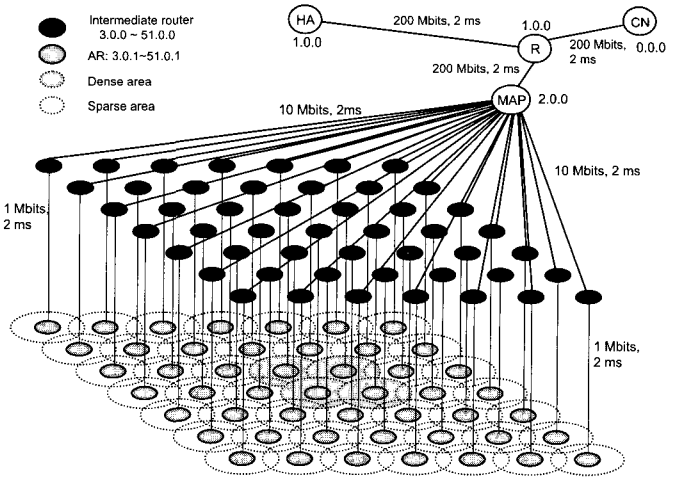


Fig. 8. Simulation networks topology.

cell. A sparse area is composed of the remaining cells except for a dense area in the MAP area. The densities of the MNs within the dense area and the sparse area are 10 MNs/cell and 5 MNs/cell, respectively. The MNs within the dense area move as follows. First, 1 MN, 2 MNs, 3 MNs, 4 MNs, and 5 MNs move randomly within the dense area at time instants 10, 20, 30, 40, and 50 sec, respectively. Second, 6 MNs move randomly within the dense area from 60 sec to 130 sec. Third, 5 MNs, 4 MNs, 3 MNs, 2 MNs, and 1 MN move randomly with the dense area at time instants 140, 150, 160, 170, and 180 sec, respectively. On the other hand, 1 MN moves only randomly during the total simulation time in the sparse area. The *THi(VC)* and the *THi(VR)* of cells in the dense area are 4 MNs/sec and 2 MNs/sec, respectively. Also, the *THi(VC)* and the *THi(VR)* of cells in the sparse area are 2 MNs/sec and 1 MN/sec, respectively. That is, when the MNs handovered to neighboring cell exceed 40% of active MNs within a cell, a VIP zone is created. Also, when the MNs handovered neighboring cell drop below 20% of active MNs with a cell during the period of the VIP zone, a VIP zone is released. Here, The *THi(VR)* was set to a smaller value than the *THi(VC)* to avoid oscillation in stable condition.

C.4 Simulation Results

We record the wireless signaling counts every 10 simulation time units (sec) for all experiments. Fig. 9 shows changes in the wireless signaling messages depending on the simulation time when the VIP zone is formed in 4 cells. We can see from Fig. 9(a) that when the VIP zone is created and released, the wireless signaling messages in VIP allocation scheme increase more impulsively than those in HMIPv6. But, the wireless signaling messages in VIP allocation scheme are much less than those in HMIPv6 during the duration of the VIP zone. In this figure, although VIP allocation scheme uses 23 more messages than HMIPv6 during the creation instant of the VIP zone, about 80 messages are saved using VIP allocation scheme during the duration period of the VIP zone. Fig. 9(b) shows cumulative wireless signaling messages of Fig. 9(a). Therefore, we can see from the Fig. 9(b) that the cost gain during the duration of the VIP zone is much more than the cost loss during the creation and

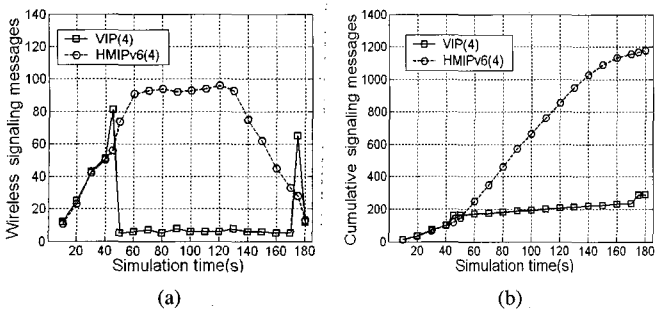


Fig. 9. Variation in wireless signaling cost (when VIP zone is 4 cells); (a) wireless signaling messages, (b) cumulative wireless signaling messages.

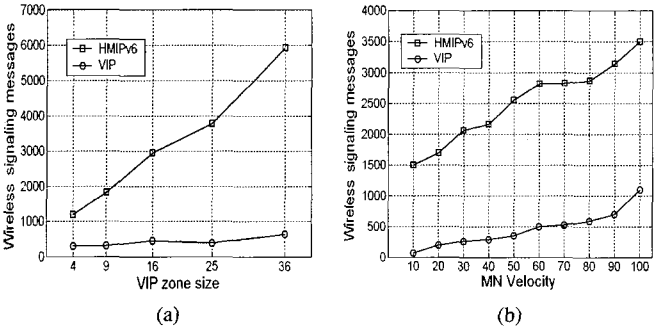


Fig. 10. Effect of MN movement within VIP zone; (a) effect of VIP zone size, (b) effect of MN velocity.

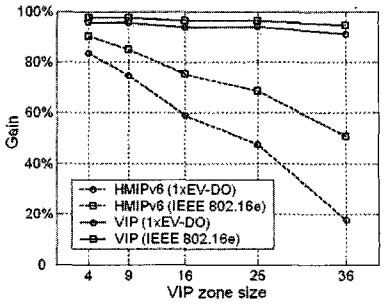


Fig. 11. Effect of overall gain.

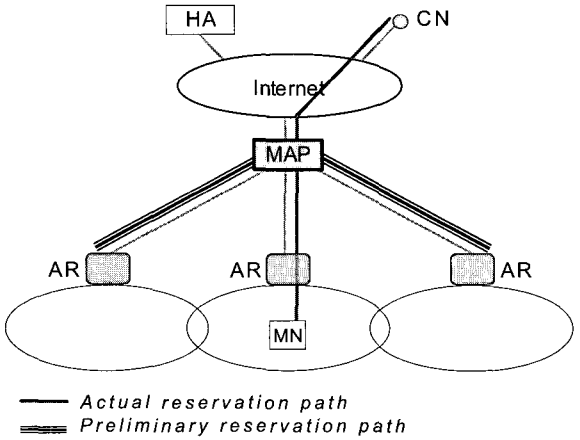


Fig. 12. Network architecture of PA-RSVP protocol.

the release of the VIP zone. Moreover, as the duration period of the VIP zone increases, the cost gain increases greatly. For example, when VIP allocation scheme is applied at the amusement center, the duration period of the VIP zone lasts up some hours because the VIP zone is created early in the evening and is released at midnight. Fig. 10 shows the effect of VIP allocation scheme within the VIP zone. The simulation results in Fig. 10 follow the same trend shown in Fig. 7. These simulation results imply that the performance of VIP allocation scheme with the VIP zone is closely related to the size of the VIP zone and the velocity of the MNs. Fig. 11 shows the result for overall gain. This gain means how much time (%) is available for non-wireless signaling and is expressed as follows.

$$\text{Gain} = \left(\frac{T(SIM) - T(SIG)}{T(SIM)}\right) \times 100\ (\%), \qquad (10)$$

where $T(SIM)$ is the total simulation time and $T(SIG)$ is the cost in time for the wireless signaling. The time required to process one wireless signaling varies in the MAC layer depending on the multiple access technology used in the wireless link. For example, the time required to process one wireless signaling is 25 ms in 1xEV-DO and 15 ms in IEEE 802.16e. Based on these parameter values, we can observe in Fig. 11 that the time excluding the wireless signaling overhead from the total time is much longer in VIP allocation scheme than in HMIPv6. Also, IEEE 802.16e shows a greater benefit by the gain than 1xEV-DO from the difference in the wireless bandwidth.

IV. PROPORTIONAL AGGREGATE RSVP SCHEME

The network architecture of proportional aggregate RSVP (PA-RSVP) is also based on HMIPv6 as shown in Fig. 12. To apply PA-RSVP in the IP-based cellular networks, we need a set of the RSVP-enabled ARs supporting IP routing and the MAP that performs advance resource reservations. The network entities are explained as follows.

- *Mobility anchor point (MAP)*: It functions as an aggregating router and makes resource reservations to neighboring ARs of an MN's current AR in advance using proportional aggregation reservation.
- *Access router (AR)*: It functions as a de-aggregating router and a proxy agent of the MN and monitors movement status into neighboring ARs using handoff history. It calculates the expected the number of MNs that is expected to visit neighboring ARs based on the movement status, and based on these values, it proportionally estimates the aggregate bandwidth by the bandwidth ranking.
- *Correspondent node (CN)* or *home agent (HA)*: It sends a PATH message to the MN using binding update information.

PA-RSVP protocol has three reservation classes—actual reservation, preliminary reservation, and extra reservation. The actual reservation is the resource that is reserved from the CN to the MN through the current AR that is currently being used. The preliminary reservations are the resources that are reserved in advance using proportional aggregation from the MAP to neigh-

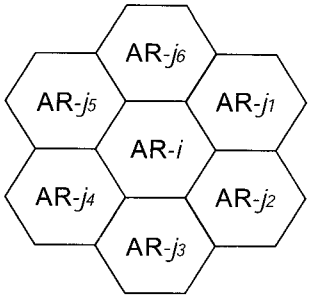


Fig. 13. Example of proportional aggregation reservation.

boring ARs that are expected to be visited by the MN, and are not currently being used. The bandwidth on the preliminary reservation is reserved but not actually allocated. Lastly, the extra reservation temporarily makes use of the inactive bandwidth that is reserved by other flows in the current AR. If the MN moves to a neighboring AR, the resource of the preliminary reservation class at the new location switches to the actual reservation class. Also, the resource of the actual reservation class at the previous location switches to the preliminary reservation class. However, when the MN, initially established in the actual reservation class, moves to a neighboring AR, the actual reservation switches to individual preliminary reservation. If the reservation class information of the MN is not correctly updated by the MAP, there could be problems with changing reservation classes when an MN moves to a neighboring AR.

A. Method of Allocating Proportional Aggregate Resource

Allocating proportional aggregate resource in the preliminary reservation path is carried out in four steps as follows. Here, we consider the coverage areas for the ARs as shown in Fig. 13.

(STEP 1) Each AR calculates the handoff rates of the MNs in active communication to neighboring ARs. That is, the handoff rate from AR-i to AR-j_k, P_{i,j_k} is measured. For example, the handoff rate from AR-i to AR-j_1 is calculated as (11).

$$P_{i,j_1} = \frac{numHO(i,j_1)}{\sum_{jk \in Neigh(i)} numHO(i,j_k)}, \tag{11}$$

where $numHO(i,j_k)$ is the number of active handoffs from AR-i to AR-j_k during certain periods, and $Neigh(i)$ is a set of AR-i's neighboring ARs.

(STEP 2) Each AR stores the handoff rates to neighboring ARs as the following type.

$$(neighboring\ AR's\ address, handoff\ rate\ to\ neighboring\ AR). \tag{12}$$

For example, the handoff rates from AR-i to AR-j_k in Fig. 13 are stored as $(Addj_1, P_{i,j_1})$, $(Addj_2, P_{i,j_2})$, $(Addj_3, P_{i,j_3})$, $(Addj_4, P_{i,j_4})$, $(Addj_5, P_{i,j_5})$, and $(Addj_6, P_{i,j_6})$.

(STEP 3) P_{i,j_k} is periodically updated. Therefore, the number of MNs that is expected to move from AR-i to neighboring AR-j_k, $E[numHO(i,j_k)]$ is calculated as (13).

$$E[numHO(i,j_k)] = roundup[P_{i,j_k} \times numMN(i), 0], \tag{13}$$

Table 1. Example of bandwidth allocation order for MNs in AR-i.

Ranking	Reservation BW	MN's ID
BW_1	5 Mbps	#8
BW_2	5 Mbps	#10
BW_3	3 Mbps	#2
BW_4	3 Mbps	#4
BW_5	3 Mbps	#6
BW_6	1 Mbps	#1
BW_7	1 Mbps	#3
BW_8	1 Mbps	#5
BW_9	1 Mbps	#7
BW_{10}	1 Mbps	#9

Table 2. Example of storage for handoff rates into AR-j_k in AR-i.

Neighboring AR	Handoff rate	Cache type
AR-j_1	$P_{i,j_1} = 0.08$	$(j_1, 0.08)$
AR-j_2	$P_{i,j_2} = 0.1$	$(j_2, 0.1)$
AR-j_3	$P_{i,j_3} = 0.12$	$(j_3, 0.12)$
AR-j_4	$P_{i,j_4} = 0.4$	$(j_4, 0.4)$
AR-j_5	$P_{i,j_5} = 0.23$	$(j_5, 0.23)$
AR-j_6	$P_{i,j_6} = 0.07$	$(j_6, 0.07)$

where $numMN$(i) is the number of MNs in active communication within AR-i.

(STEP 4) Each AR stores the list sorted in descending order by the offered bandwidths for active MNs within its coverage. Then, each AR adds up the bandwidths from the top of the list for $E[numHO(i,j_k)]$. This added bandwidth, Ω_{i,j_k} is calculated as (14) and is allocated in advance between the MAP and neighboring ARs using proportional aggregate reservation.

$$\Omega_{i,j_k} = \sum_{r=1}^{E[numHO(i,j_k)]} BW_r, \tag{14}$$

where r stands for the rank in the bandwidth order, and BW_r represents the r-th bandwidth.

Let us look at an example of allocating proportional aggregate resource to neighboring AR-j_k. In Fig. 13, we assume that there exists 10 MNs that are in active communication under AR-i and the offered bandwidths are 5 Mbps in 20% (#8, #10), 3 Mbps in 30% (#2, #4, #6), and 1 Mbps in 50% (#1, #3, #5, #7, #9) of the MNs. AR-i stores the values of the offered bandwidths of active MNs in its coverage as shown in Table 1. Then, the handoff rates from AR-i to neighboring AR-j_k calculated from (12) are stored as shown in Table 2. Based on these values, AR-i calculates the number of MNs that is expected to move into AR-j_1, $E[numHO(i,j_1)]$ as below.

$$\begin{aligned} E[numHO(i,j_1)] &= roundup[(P_{i,j_1} \times numMN(i)), 0] \\ &= roundup[(0.08 \times numMN(i)), 0] = 1, \end{aligned}$$

then, the numbers of MNs that are expected to move into the rest of neighboring ARs are calculated in the same way using (13)

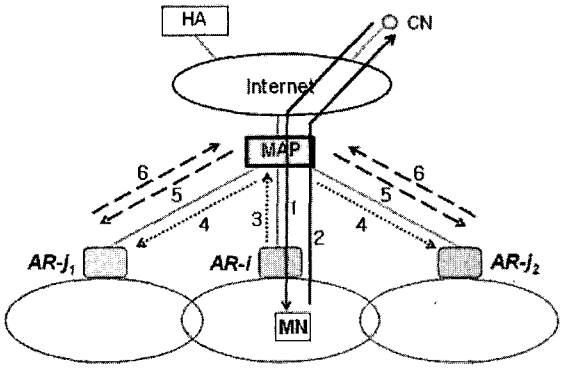


Fig. 14. Setup procedure of PA-RSVP.

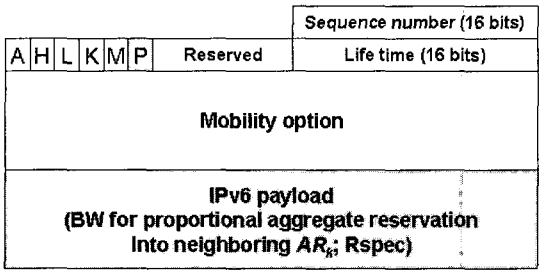


Fig. 15. Message format of proportional aggregate QoS.

as follows.

$$E[numHO(i,j_2)] = 1,\ E[numHO(i,j_3)] = 2,$$
$$E[numHO(i,j_4)] = 4,\ E[numHO(i,j_5)] = 3,$$
$$E[numHO(i,j_6)] = 1.$$

Finally, proportional aggregate bandwidths allocated from AR-i to neighboring AR-j_k is calculated using (14). First, the proportional aggregate bandwidth from AR-i to AR-j_1 is

$$\Omega_{i,j_1} = \sum_{r=1}^{E[numHO(i,j_1)]} BW_r = \sum_{r=1}^{1} BW_r = 5 \text{ Mbps}.$$

Also, the proportional aggregate bandwidths allocated to the rest of the neighboring AR-j_k are calculated as follows.

$$\Omega_{i,j_2} = 5 \text{ Mbps}$$
$$\Omega_{i,j_3} = (5 \text{ Mbps} + 5 \text{ Mbps}) = 10 \text{ Mbps}$$
$$\Omega_{i,j_4} = (5 \text{ Mbps} + 5 \text{ Mbps} + 3 \text{ Mbps} + 3 \text{ Mbps}) = 16 \text{ Mbps}$$
$$\Omega_{i,j_5} = (5 \text{ Mbps} + 5 \text{ Mbps} + 3 \text{ Mbps}) = 13 \text{ Mbps}$$
$$\Omega_{i,j_6} = 5 \text{ Mbps}.$$

B. Setup Procedure

Fig. 14 shows the operation of PA-RSVP setup procedure. We assume that the MN has created within AR-i. The MN has completed power-up registration to the CN through the MAP. First, the CN sends an end-to-end (E2E) PATH message to the MN (1 in Fig. 14). The MN replies with an E2E RESV message to the CN (2 in Fig. 14). An actual reservation is setup between the CN and the MN through these messages. According to the creation of new active MN within AR-i, AR-i calculates proportional aggregate bandwidths that are allocated to neighboring AR-j_k. Then, the calculated values of proportional aggregate bandwidths are included in IPv6 payload within HMIPv6 binding update message as shown in Fig. 15. Also, a new P-flag bit within this message is set to 1. Then, this message is sent to the MAP (3 in Fig. 14). When the MAP receives this message and detects that the P-flag bit is set to 1, it discovers that this message is not the HMIPv6 binding update message but the message that contains QoS information (i.e., Rspec) for proportional aggregate bandwidth of preliminary path. The MAP sends the Rspec for proportional aggregate bandwidth to neighboring AR-j_1 and AR-j_2 (4 in Fig. 14). After that, the MAP sends an aggregate PATH message to AR-j_1 and AR-j_2 to request the allocation for proportional aggregate bandwidth (5 in Fig. 14). As a response, AR-j_1 and AR-j_2 each sends an aggregate RESV message to the MAP based on the Rspec received from the MAP (6 in Fig. 14). Through these procedures, the MAP reserves the preliminary bandwidth with AR-j_1 and AR-j_2. The MN sends a reservation through pre-reserved proportional aggregate resource. The aggregate reservation based on Diffserv [24] is operated at the path between the MAP and AR-j_1. If an active MN existing in AR-i moves to AR-j_1, PA-RSVP protocol carries out following procedures. The class is changed from an actual reservation to a preliminary reservation between the MAP and AR-i. The proportional aggregate bandwidth is reduced between the MAP and AR-j_2 since the MN no longer exists in AR-i.

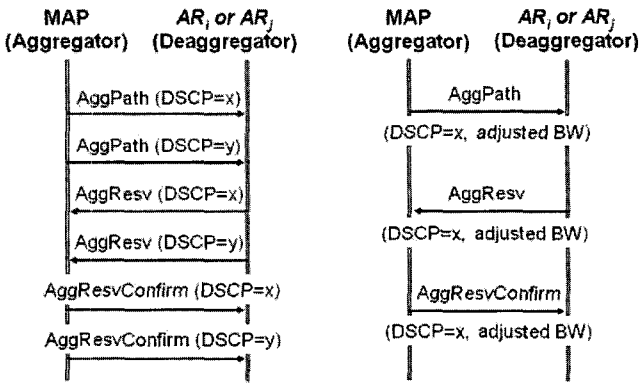


Fig. 16. Signaling flow in preliminary reservation path.

C. Operation in Preliminary Reservation Path

The aggregate reservation allocated in advance between the MAP and AR-j_1/AR-j_2 operates with differentiated service code point (DSCP) [25] of DiffServ on aggregate RSVP. The classifier in Diffserv classifies the packets into different classes in terms of DSCP of the packets. That is, the classification and the scheduling are executed by DSCP within the message of aggregate RSVP on the preliminary reservation (i.e., aggregate reservation) paths between the MAP and AR-j_1/AR-j_2. So, the packets on the aggregate reservation path between the aggregating MAP and de-aggregating AR-j_1/AR-j_2 are processed using

Table 3. Analysis parameters.

Item	Definition
d_{CN}	Number of hops between CN and AR
d_{GW}	Number of hops between gateway and AR
d_{RA}	Number of hops between RA and AR
d_{MAP}	Number of hops between MAP and AR
f	Number of flows in AR-i
m	Number of ARs in MSPEC
k	Number of neighboring ARs
R_A	Resource currently reserved for a flow
R_P	Resource reserved in advance for a flow
BW_r	r-th bandwidth in descending order in AR
$Neigh(i)$	Neighboring ARs for AR-i
P_s	Size of PATH message
R_s	Size of RESV message
T_f	Soft state refresh period

DSCP of DiffServ. If there are changes in the number of MNs in AR-i or the handoff rates to AR-j_1 or AR-j_2, the reservation bandwidths between the MAP and AR-j_1/AR-j_2 are resized with newly calculated proportional aggregation bandwidths. A signaling flow in the preliminary reservation path between the MAP and neighboring ARs is shown in Fig. 16. The first figure in Fig. 16 shows the case where no aggregate reservation path exists. In this case, we consider that two aggregation reservations with different DSCPs in the packet header are initiated. Here, the soft state is updated between the MAP and neighboring ARs aggregately are carried out in n units' resource. Therefore, the signaling overhead can be reduced by this mechanism. Here, the guaranteed service reservation is configured as DSCP Class = x, and the control load service reservation is configured as DSCP Class = y. The second figure in Fig. 16 shows the case where needs to resize when pre-allocated aggregate reservation bandwidth is changed. In this case, the MAP requests to resize aggregate resource bandwidth in neighboring AR-j_1 and AR-j_2 according to the newly calculated value from AR-i.

D. Correlation between VIP and PA-RSVP

Upon handoffs, two protocols operate separately or together depending on the region that they operate on. Within the VIP zone, PA-RSVP and VIP allocation scheme operate together, but in the non-VIP zones, PA-RSVP and HMIPv6 operate together. When an MN moves within the VIP zone, it detects a change in the RSVP path by means of L2-source trigger of VIP allocation scheme and establishes the preliminary reservations between the MAP and the neighboring ARs using PA-RSVP. If an MN moves within the non-VIP zones, it detects a change in RSVP path by HMIPv6 and makes preliminary reservations between the MAP and neighboring ARs using PA-RSVP.

E. Performance Evaluation

E.1 Cost Formulation

We use the mobility model that was adopted in Section III because advance resource reservation protocol transmits the RSVP signaling messages under IP mobility protocol. Hence, the average number that an MN moves out of a MAP domain, γ and the average of number that an MN moves within a MAP domain, δ each is expressed respectively as

$$\gamma = \frac{\ell_{\mathrm{AR}}}{\ell_{DO}}, \quad \delta = \frac{\ell_{DO}}{\ell_{\mathrm{AR}}} - 1, \tag{15}$$

where ℓ_{AR} and ℓ_{DO} is the side length of the AR (m) and the MAP domain (m), respectively. We compare the performances among PA-RSVP, MRSVP, and RSVP-RA with respect to the costs of the resource reservation bandwidth for advance resource reservation and the signaling overhead for the soft state refresh messages on the paths reserved during intra-domain and inter-domain handoffs. Route optimization is completed in each protocol. That is, the data packets from the CN are directly sent to the MN without sending the HA. Also, three service types (2 Mbps: 20%, 1 Mbps: 50%, 0.5 Mbps: 30%) are provided. The parameters used in the evaluation are listed in Table 3.

First, the resource reservation bandwidths in each protocols are expressed as follows.

- *Resource reservation bandwidth for MRSVP*:

$$RB_{\mathrm{MRSVP}} = f\{(d_{\mathrm{CN}}+1)R_A + (d_{\mathrm{CN}} \cdot m \cdot R_P)\}\left(\frac{\ell_{DO}}{\ell_{\mathrm{AR}}} + \frac{\ell_{\mathrm{AR}}}{\ell_{DO}} - 1\right). \tag{16}$$

The above equation represents the bandwidth for the active reservation and passive reservations from MSPEC during intra-domain and inter-domain handoffs.

- *Resource reservation bandwidth for RSVP-RA*:

$$\begin{aligned} RB_{\mathrm{RSVP-RA}} &= f\{(d_{\mathrm{CN}}+1)R_A + (d_{\mathrm{RA}} \cdot k \cdot R_P)\}\left(\frac{\ell_{DO}}{\ell_{\mathrm{AR}}} - 1\right) \\ &\quad + f(d_{\mathrm{CN}}+1)R_A\left(\frac{\ell_{\mathrm{AR}}}{\ell_{DO}}\right). \end{aligned} \tag{17}$$

The first term represents the bandwidth for the reserved reservation and prepared reservation between RSVP agent and neighboring ARs during intra-domain handoffs. The second term represents the bandwidth for the reserved reservation during inter-domain handoffs.

- *Resource reservation bandwidth for PA-RSVP*:

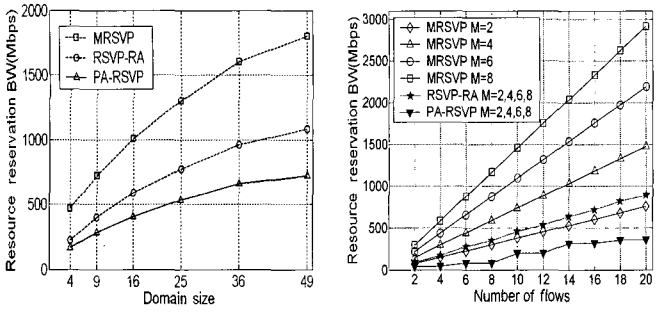
$$\begin{aligned} RB_{\mathrm{PA-RSVP}} &= \{f(d_{\mathrm{CN}}+1)R_A\}\left(\frac{\ell_{DO}}{\ell_{\mathrm{AR}}} - 1\right) \\ &\quad + \{d_{\mathrm{MAP}}(\textstyle\sum_{j_k \in Neigh(i)} \Omega_{i,j_k})\}\left(\frac{\ell_{DO}}{\ell_{\mathrm{AR}}} - 1\right) \\ &\quad + f(d_{\mathrm{CN}}+1)R_A\frac{\ell_{\mathrm{AR}}}{\ell_{DO}}. \end{aligned} \tag{18}$$

The first and the second terms each represents the bandwidth for the actual reservation and proportional aggregate reservations between the MAP and neighboring ARs during intra-domain handoffs. The third term represents bandwidth for the actual reservation during inter-domain handoffs.

Second, the soft state refresh overheads for each protocols are expressed as follows.

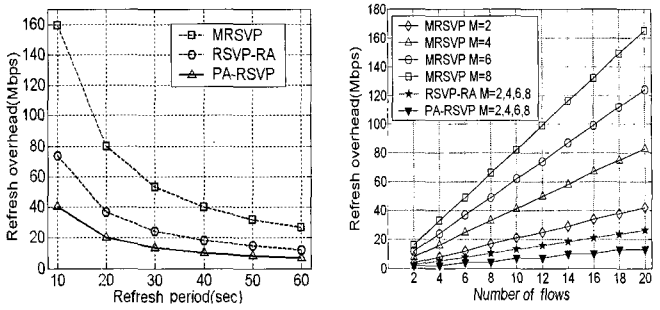
- *Soft state refresh overhead for MRSVP*:

$$SO_{\mathrm{MRSVP}} = f\frac{(P_s+R_s)}{T_f}\{(d_{\mathrm{CN}}+1) + m \cdot d_{\mathrm{CN}}\}\left(\frac{\ell_{DO}}{\ell_{\mathrm{AR}}} + \frac{\ell_{\mathrm{AR}}}{\ell_{DO}} - 1\right). \tag{19}$$



(a) (b)

Fig. 17. Numerical results 1: Resource reservation bandwidth; (a) effect of domain size, (b) effect of number of flows.



(a) (b)

Fig. 18. Numerical results 2: Soft state maintenance overhead; (a) effect of refresh period, (b) effect of number of flows.

The above equation represents the soft state refresh overhead during intra-domain and inter-domain handoffs.

- *Soft state refresh overhead for RSVP-RA*:

$$SO_{\text{RSVP-PA}} = f^{\frac{(P_s+R_s)}{T_f}}\{(d_{\text{CN}}+1)+k\cdot d_{\text{RA}}\}(\tfrac{\ell_{DO}}{\ell_{\text{AR}}}-1) + f^{\frac{(P_s+R_s)}{R_f}}(d_{\text{CN}}+1)(\tfrac{\ell_{\text{AR}}}{\ell_{DO}}). \quad (20)$$

The first term represents the soft state refresh overhead in the reserved reservation and prepared reservation between RSVP agent and neighboring ARs during intra-domain handoffs. The second term represents the soft state refresh overhead in the reserved reservation during inter-domain handoffs.

- *Soft state refresh overhead for PA-RSVP*:

$$SO_{\text{PA-RSVP}} = \tfrac{(P_s+R_s)}{T_f}\{f(d_{\text{CN}}+1)+k\cdot d_{\text{MAP}}\}(\tfrac{\ell_{DO}}{\ell_{\text{AR}}}-1) + f^{\frac{(P_s+R_s)}{R_f}}(d_{\text{CN}}+1)(\tfrac{\ell_{\text{AR}}}{\ell_{DO}}). \quad (21)$$

The first term represents the soft state refresh overhead in the actual reservation and proportional aggregate reservation between the MAP and neighboring ARs during intra-domain handoffs. The second term represents the soft state refresh overhead in the actual reservation during inter-domain handoffs.

E.2 Analytic Results

Fig. 17 shows the comparison results of consumed resource reservation bandwidths of three protocols. The parameter values used in this analysis are $m = 4$, $k = 4$, $f = 10$, d_{GW} (or d_{RA}, d_{MAP}) $= 3$ hops, and $d_{\text{CN}} = 10$ hops. A domain consists of 16 ARs in this analysis. Fig. 17(a) shows that PA-RSVP consumes much less bandwidth overhead than MRSVP and RSVP-RA during advance resource reservation. Also, the differences in the bandwidth overhead among three protocols become larger as the domain size increases. Fig. 17(b) shows the bandwidth overhead depending on the number of flows when the number of cells in MSPEC are 2, 4, 6, and 8. In this figure, RSVP-RA consumes less bandwidth overhead than MRSVP when the number of cells in MSPEC are 4, 6, and 8, while MRSVP shows better performance than RSVP-RA only when the number of cells in MSPEC is 2. It is because the number of advance reservations is also small in MRSVP when the number of cells in MSPEC is small. But, PA-RSVP always consumes much less bandwidth overhead than MRSVP and RSVP-RA when the number of cells in MSPEC are 2, 4, 6, and 8. Thus, these results come from the fact that whereas MRSVP and RSVP-RA reserve individually resources for active MNs into neighboring ARs in advance, PA-RSVP reserves aggregately resources in advance for active MNs with expected movement using proportional aggregation.

Figs. 18(a) and 18(b) each shows the result depending on the soft state refresh period and number of MN flows, respectively. The default values of the parameters used in this analysis are $PATH_{\text{SIZE}} = 64$ kbytes, $RESV_{\text{SIZE}} = 64$ kbytes, $k = 4$ ARs, $f = 10$ flows, and $T_f = 30$ sec. Fig. 18(a) shows that the signaling overhead caused in PA-RSVP protocol is much less than that of MRSVP and RSVP-RA protocols with respect to the refresh period. Also, we can see that Fig. 18(b) show similar results as in Fig. 17(b), and can observe that overhead caused by the soft state refreshes in PA-RSVP protocol is less than that of MRSVP and RSVP-RA protocols. These results come from the fact that whereas MRSVP and RSVP-RA individually manages soft states of all MNs, PA-RSVP aggregately manages soft states of MNs in n units' resource.

E.3 Simulation Environments

We evaluate the performance of PA-RSVP by simulation study to show the mobility impacts on the amount of reserved bandwidth and soft state maintenance. We used the ns-2.26 RSVP simulator [26]. We also compared the performance of PA-RSVP with MRSVP and RSVP-RA. In RSVP-RA, resource reservations are made for the CN to the current location of the MNs and from an RSVP agent to neighboring cells for each MN, while in MRSVP, resource reservations are made from the CN to all locations where the MNs are expected to visit during the lifetime of the connection. In PA-RSVP, resource reservations are made for the CN to the current location of the MNs and from an MAP to neighboring cells through the proportional aggregate. We use the same simulation topology that was implemented in Section III. A source agent is attached to the CN and the sink

agents are attached to the MNs. The total simulation time is 500 simulation time units (sec). There are 5 MNs in each cell and the movement of the MNs is generated randomly. That is, each MN moves to its destinations, stays there for 30 simulation time units and then moves again. When the simulation starts, the CN sends up to one best-effort flow per the MN. These flows are exponential on/off (Expoo) flows. This means that the packets are sent where the burst and idle times are chosen from exponential distributions (with an average of 1 sec and 0.5 sec for the burst time and the idle time, respectively). The packet sizes for best effort flow are constant during the lifetime of a flow, but the sizes of each flow are chosen form an exponential distribution with the average 2500 bytes. The peak rate of these flows is 1 Mbps. Also, the CN sends up to 2 RSVP flows per MN. In addition, 10 RSVP flows are generated per cell because there are 5 active MNs per cell. The Rspec is applied between times of 100–200 sec and 300–400 sec, and it is assigned with the following ratio: 20% of flows (2 out of 10 flows) - 0.5 Mbps, 50% of flows (5 out of 10 flows) - 1 Mbps, and 30% of flows (3 out of 10 flows) - 2 Mbps. The rate for the flow is measured by a simple monitor object at the MN.

E.4 Simulation Results

Fig. 19 each shows the changes in resource reservation bandwidths depending on the domain sizes, the number of flows and the number of hops. Here, the resource reservation bandwidth in Fig. 19 shows the resources used up by each protocol when the Rspecs of each RSVP flows are pre-provisioned. First, Figs. 19(a) and 19(b) each shows that PA-RSVP consumes less bandwidth overhead than MRSVP and RSVP-RA as the domain size and the number of flows increases. This simulation results follow the similar trend shown in Fig. 17. Therefore, the proposed PA-RSVP protocol not only reduces the number of messages by aggregating signaling information, but also reduces the network resources by allocating pre-provisioning resources in aggregates by the statistical evaluation for the MN's handoff rates. Of the two benefits, reducing the pre-provisioning resources, by tracking handoff rates to neighboring ARs, is of more importance. Second, Figs. 19(c) and 19(d) each shows changes in the resource reservation bandwidths depending on the number of hops (i.e., d_{GW}, d_{RA}, and d_{MAP}) between the gateway (or RA, MAP) and the AR, and the number of hops (i.e., d_{CN}) between the CN and the AR. Fig. 19(c) shows that MRSVP is much less sensitive to changes in the number of hops between the gateway and the AR. But, RSVP-RA is more sensitive to changes in the number of hops between the RSVP agent and the AR than other protocols. On the other hand, compared with RSVP-RA, PA-RSVP is less sensitive to changes in the number of hops in access networks. Also, we can observe from Fig. 19(d) that MRSVP is much more sensitive to changes in number of hops between the CN and the MN than RSVP-RA and PA-RSVP. In conclusion, the proposed PA-RSVP is more stable to changes in the number of hops in the access networks and the Internet backbone than MRSVP and RSVP-RA.

Fig. 20 shows the simulation results of the cost of the soft state maintenance among three protocols. In the Figs. 20(a) and 20(b), we can observe that the cost caused by soft state maintenance in PA-RSVP is less than MRSVP and RSVP-RA as the domain size and the number of flows increase. Also, Fig. 20(c) plots the soft state refresh overhead versus the refresh period for three protocols. The simulation result shows that the soft state refresh overhead in PA-RSVP is less than that in MRSVP and RSVP-RA. This simulation results follow the similar trend shown in Fig. 18.

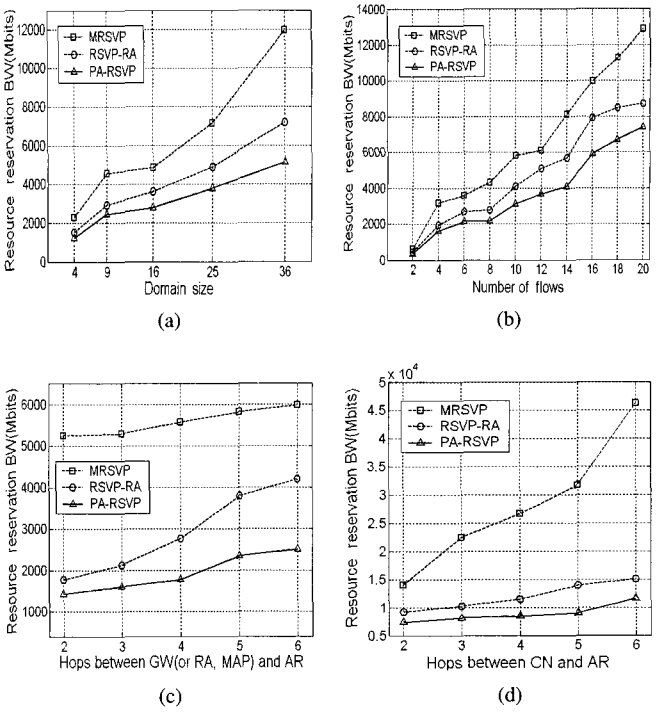


Fig. 19. Simulation results 1: Resource reservation bandwidth; (a) effect of domain size, (b) effect of number of flows, (c) effect of d_{GW} (or d_{RA}, d_{MAP}), (d) effect of d_{CN}.

Fig. 21 shows the simulation results for the average data rate using the MRSVP, RSVP-RA, and PA-RSVP over simulation time. Fig. 21 shows the average data rate with reservations in a chosen MN with 1 Mbps RSVP flow. This simulation results show that PA-RSVP, MRSVP, and RSVP-RA are stable at 1 Mbps at the reservation times (from 100–200 sec, 300–400 sec), respectively. This phenomenon shows that, although PA-RSVP allocates proportional aggregate resource to neighboring ARs, it can still provide a high QoS guarantees for the service quality of the MN as MRSVP and RSVP do. On the other hand, we observe that the average data rate is obviously decreased at the moment a handoff takes place (at 148, 285, and 463 time units with MRSVP, at 106, 159, and 305 time units with RSVP-RA, at 78, 141, and 442 time units with PA-RSVP).

V. CONCLUSION

In this paper, we considered mobility issues that should be concerned in 4G wireless networks and proposed the efficient IP mobility management and QoS provisioning, virtual-IP (VIP) allocation scheme and proportional aggregate RSVP (PA-RSVP). First, we showed that VIP allocation scheme in area

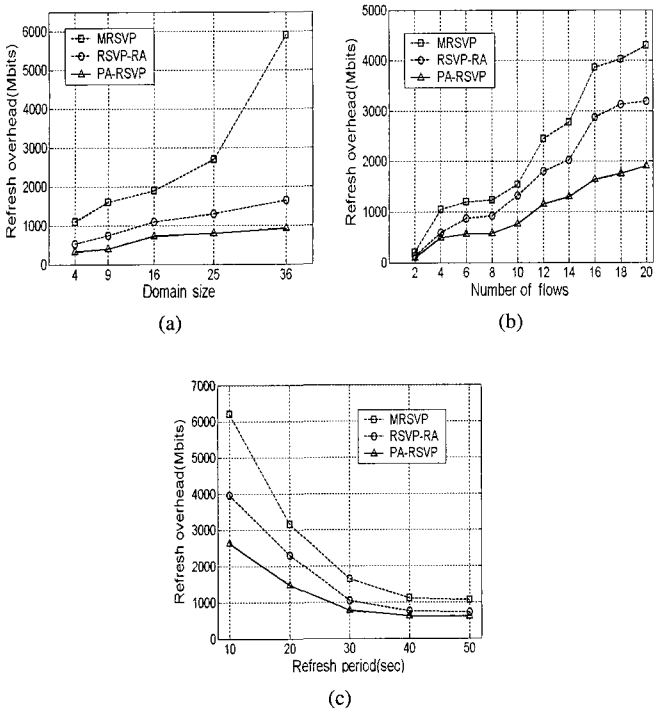


Fig. 20. Simulation results 2: Soft state maintenance overhead; (a) effect of domain size, (b) effect of number of flows, (c) effect of refresh overhead.

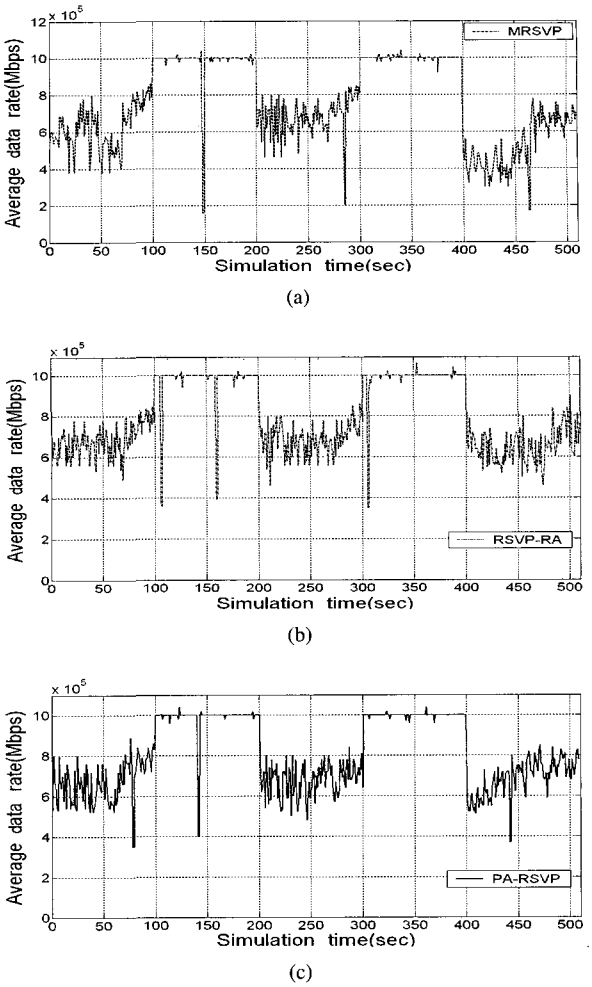


Fig. 21. Average data transmission rate; (a) MRSVP, (b) RSVP-RA, (c) PA-RSVP.

with a large rate of handoff has less wireless signaling overhead due to IP mobility than HMIPv6. Second, we also showed that PA-RSVP can minimize the resource reservation bandwidth and the signaling overhead while providing QoS guarantees over existing advance reservation protocols. As VIP allocation scheme and PA-RSVP are deployed to reduce the resource in the IP-based cellular networks, we expect that mobile users will be offered with more reliable service. In our future work, we will consider QoS guarantees considering such as time-varying wireless channel condition by selecting the wireless channel and the scheduling algorithm.

REFERENCES

[1] G. Chiruvolu *et al.*, "Mobility and QoS support for IPv6-based real-time wireless Internet traffic," in *Proc. IEEE ICC'99*, vol. 1, 1999, pp. 334–338.

[2] A. Lopez and J. Manner, "A study on QoS provision for IP-based radio access networks," *Lecture Notes in Computer Science (LNCS)*, vol. 2170, pp. 139–157, 2001.

[3] D. Johnson and C. Perkins, "Mobility support in IPv6," RFC 3775, June 2003.

[4] C. Castelluccia and L. Bellier, "Hierarchical mobile IPv6," *Internet Draft*, draft-ieft-mipshop-hmipv6-03.txt, Dec. 2004.

[5] S. Park and Y. Choi, "A study on performance of hierarchical mobile IPv6 in IP-based cellular networks," *IEEE Trans. Commun.*, vol. E87-B, no. 3, pp. 462–469, Mar. 2003.

[6] X. Perez-Costa and M. Torrent-Moreno, "A performance study of hierarchical mobile IPv6 from a system perspective," in *Proc. IEEE ICC 2003*, 2003.

[7] R. Hsieh and Z. G. Zhou, "SMIP: A seamless handoff architecture for mobile IP," in *Proc. INFOCOM 2003*, vol. 3, 2003, pp. 1774–1784.

[8] R. Berezdivin and R. Breinig, "Next-generation wireless communication concepts and technologies," *IEEE Commun. Mag.*, vol. 40, no. 3, pp. 108–116, 2002.

[9] G. Evans and K. Baughan, "Visions of 4G," *IEEE Electron. Commun.*, vol. 12, pp. 293–303, Dec. 2002.

[10] A. S. Acampora and M. Naghshineh, "An architecture and methodology for mobile-executed handoff in cellular ATM networks," *IEEE J. Select. Areas Commun.*, vol. 12, no. 8, pp. 1365–1375, Oct. 1994.

[11] D. Clark *et al.*, "Supporting real-time applications in an integrated services packet network architecture and mechanism," in *Proc. SIGCOMM'92*, 1992, pp. 14–26.

[12] W. T. Chen and L. C. Huang, "RSVP mobility support: A signaling protocol for integrated services Internet with mobile hosts," in *Proc. INFOCOM 2000*, vol. 3, 2000, pp. 1283–1292.

[13] R. Braden *et al.*, "Integrated service in the Internet architecture: An overview," RFC 1633, 1994.

[14] R. Braden *et al.*, "Resource rservation protocol (RSVP) - version 1 functional specification," RFC 2205, 1997.

[15] S. Pack *et al.*, "Seamless QoS handling mechanism in macro and micro mobility," *Lecture Notes in Computer Science (LNCS)*, vol. 2344, pp. 62–73, 2002.

[16] R. Koodli, "Fast handoffs for mobile IPv6. Internet draft," *Internet Draft*, draft-ietf-mobileip-fast-mipv6-02.txt, 2004.

[17] A. Terzis *et al.*, "A simple QoS signaling protocol for mobile hosts in the integrated services Internet," in *Proc. IEEE INFOCOM'99*, 1999, pp. 680–688.

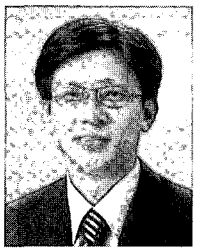
[18] A. K. Talukdar *et al.*, "MRSVP: A reservation protocol for an integrated services packet networks with mobile hosts," TR-337, 1997.

[19] Y.-J. Suh *et al.*, "An efficient resource reservation protocol by QoS agents in mobile networks," *IEICE Trans. Commun.*, vol. E86-B, no. 3, pp. 1094–1101, Mar. 2003.

[20] S. Thomson and T. Narten, "IPv6 stateless address autoconfiguration," RFC 2462, 1998.

[21] T. Brown and S. Mohan, "Mobility management for personal communications systems," *IEEE Trans. Veh. Technol.*, vol. 46, no. 2, pp. 269–278, 1997.
[22] S. Mohan, R. Jain, "Two user location strategies for personal communications services," *IEEE Pers. Commun.*, vol. 1, no. 1, pp. 42–50, 1994.
[23] ns2 simulator, version 2.1b7a. http://www.isi.edu/nsnam/ns, 2004.
[24] F. Baker *et al.*, "Aggregation of RSVP for IPv4 and IPv6 reservations," RFC 3175, 2001.
[25] K. Nichols *et al.*, "Definition of the differentiated services field (DS field) in the IPv4 and IPv6 headers," RFC 2474, 1998.
[26] ns2 simulator, version 2.26, http://www.isi.edu/nsnam/ns, 2004.

Taehyoun Kim received the B.S. in Electronic Engineering from Hong-ik University, Seoul, Korea, in 1996, and the M.S. in Electronic Engineering from Yonsei University, Seoul, Korea, in 1999. From 1996 to 2002, he worked as a senior research engineer in Samsung Electronics Co., Ltd. Since 2002, he has been working towards the Ph.D. degree in Electrical and Electronic Engineering at Yonsei University. His major research interests are QoS provision, mobility management, cross-layer radio resource management, and vertical handoff in heterogeneous networks.



Jaiyong Lee received the Ph.D. degree in Computer Engineering from Iowa State University, USA, in 1987. He has been with ADD (Agency for Defense Development) as a research engineer from 1977 to 1982, and with Computer Science Dept. of POSTECH as an associate professor from 1987 to 1994. Since 1994, he has been a professor in School of EE, Yonsei University. He is also actively serving in the IT areas such as Executive Director of KICS (Korea Institute of Communication Sciences), Executive Vice President of OSIA (Open Standards and Internet Association), and Editor of JCN and ETRI journal. Recently, he has been working for standard activity, especially in IETF, as an IT professional of MIC (Ministry of Information and Communication Republic of Korea) for 2 years. Now, he is performing research works in the areas of QoS and mobility management for supporting 4G architecture, MAC and multicast protocol design for wireless access network, and sensor MAC/routing protocol design.