Wireless Pers Commun
DOI 10.1007/s11277-012-0583-9

Smart Routers for Cross-Layer Integrated Mobility and Service Management in Mobile IPv6 Systems

Ding-Chau Wang · Weiping He · Ing-Ray Chen

Abstract We propose and analyze a cross-layer integrated mobility and service management scheme called DMAPwSR in Mobile IPv6 environments with the goal to minimize the overall mobility and service management cost for serving mobile users with diverse mobility and service characteristics. The basic idea of DMAPwSR is that each mobile node (MN) can utilize its cross-layer knowledge to choose smart routers to be its dynamic mobility anchor points (DMAPs) to balance the cost associated with mobility services versus packet delivery services. These smart routers are just access routers for MIPv6 systems except that they are capable of processing binding messages from the MN and storing the current location of the MN in the routing table for forwarding service packets destined to the MN. The MN's DMAP changes dynamically as the MN roams across the MIPv6 network. Furthermore the DMAP service area also changes dynamically reflecting the MN's mobility and service behaviors dynamically. Unlike previous mobility management protocols such as HMIPv6 that focus only on mobility management, DMAPwSR considers integrated mobility and service management. We develop an analytical model based on stochastic Petri nets to analyze DMAPwSR and compare its performance against MIPv6 and HMIPv6. We validate analytical solutions obtained through extensive simulation including sensitivity analysis of simulation results with respect to the network coverage model, the MN's residence time distribution and the DMAP service area definition.

Keywords Mobile IP · MIPv6 · HMIPv6 · Smart routers · Mobility management · Service management · Performance analysis

D.-C. Wang
Department of Information Management, Southern Taiwan University, Tainan, Taiwan
e-mail: zh9@mail.stut.edu.tw

W. He · I.-R. Chen (✉)
Department of Computer Science, Virginia Tech, Northern Virginia Graduate Center, Falls Church, VA 22043, USA
e-mail: irchen@vt.edu

W. He
e-mail: weiping@vt.edu

1 Introduction

Mobile IPv6 (MIPv6) [11] is a network protocol for enabling mobility in IPv6 networks. It allows mobile nodes (MNs) to move within IP-based networks while maintaining on-going connections. With the advances of IP-based wireless networks, and the growth in the number of wireless devices, it is widely speculated that MIPv6 will become prevalent in next generation all-IP networks to allow users to maintain service continuity while on the go [21].

Two major sources of traffic in MIPv6 systems are due to mobility management [2,18,19] and service management [6,8,10]. Traditionally mobility management has been considered separately from service management [5,3], as mobility management mainly deals with mobility handoff, location update and location search, while service management mainly deals with data delivery and applications can always send packets to the MN by using the MN's permanent IP address.

For next-generation mobile IPv6 networks, MNs are expected to be very active with significant mobility. The mobility rate with which subnets are crossed by MNs can be high, causing a high signaling overhead for the MN to inform the MN's home agent (HA) and corresponding nodes (CNs) of the address change. There have been approaches [7,15,19,20,24] proposed to mitigate this high volume of network signaling cost, including, most noticeably, IETF work-in-progress MIP Regional Registration (MIP-RR) [9], Hierarchical MIPv6 [22] and IDMP [7]. MIP-RR uses a Gateway Foreign Agent (GFA) to keep track of the MN's current care-of-address (CoA) as long as the MN moves within a region, thereby reducing the network signaling cost when the MN moves within a region. When the MN moves to a new region, it registers with a new GFA whose address is updated to the HA as the current regional CoA. Hierarchical MIPv6 (HMIPv6) [22] is designed to reduce the network signaling cost for mobility management based on the observation that statistically local mobility accounts for more than 60 % of movements made by a MN. In addition to a CoA, a regional CoA (RCoA) is also allocated to a MN whenever the MN enters a new DMAP domain. The HA and CNs ideally only know the MN's RCoA, so whenever the MN moves across a MAP domain and triggers a RCoA address change, the new RCoA address needs to be propagated to the HA and CNs. Whenever a MN moves from one subnet to another but is still within a region covered by a MAP domain, the CoA change is only propagated to the MAP instead of to the HA and CNs, thus saving the signaling cost for mobility management. The number of subnets covered by a MAP domain is static in HMIPv6. That is, MAPs in HMIPv6 are statically pre-configured and shared by all MNs in the system. Here we note that both MIP-RR and HMIPv6 deal with mobility management only without considering service management.

In this paper, we propose and analyze a cross-layer, scalable and efficient integrated mobility and service management scheme, called DMAPwSR, with the goal to minimize the network cost incurred for mobility management and service management in MIPv6 systems. The basic idea of DMAPwSR is that each MN can utilize its cross-layer knowledge to choose its own dynamic mobility anchor points (DMAPs) to balance the cost associated with mobility management versus service management. The cross-layer knowledge we used in this paper is the mobile application's packet arrival rate versus the MN's mobility rate, or service to mobility ratio (SMR). The MN's DMAP changes dynamically as the MN roams across the MIPv6 network. These DMAPs, as in HMIPv6, are smart access routers (ARs). However, there is no pre-configuration of MAPs in the system as in HMIPv6. Rather, every AR is "smart" and can be chosen by a MN to act as the MN's DMAP to reduce the signaling overhead for intra-regional movements. The DMAP domain size, or the number of subnets in a region covered by a DMAP, is based on cross-layer knowledge embedded in the MN regarding the MN's mobility and service characteristics. Here we note that MNs typically have diverse mobility and service characteristics, i.e., diverse SMR values, while engaging in applications ranging from online texting, chatting and shopping mobile applications (which have low SMR) to online audio/video mobile applications (which have high SMR).

Our contribution is that we identify the best DMAP domain size that can minimize the network traffic via model-based evaluation with simulation validation. Unlike HMIPv6 where a MAP has a fixed domain size and serves all MNs registered under it, DMAPwSR considers a *per-user* DMAP domain size which can vary dynamically depending on the MN's runtime mobility and service characteristics. We show that DMAPwSR outperforms HMIPv6 in terms of the overall cost incurred per time unit in mobility and service management in MIPv6 environments.

The rest of the paper is organized as follows. Section 2 describes DMAPwSR for cross-layer integrated mobility and service management in MIPv6 environments. Section 3 develops a mathematical model based on stochastic Petri nets to determine the per-user optimal DMAP service area based on a MN's mobility and service characteristics so as to minimize the network communication cost induced by mobility and service management operations. In Sect. 4, we compare DMAPwSR versus HMIPv6 and present analytical results validated with extensive simulation. Finally, Sect. 5 summaries the paper and outlines some future research areas.

2 DMAPwSR with Smart Routers

Our proposed DMAPwSR protocol provides cross-layer integrated mobility and service management to reduce the network signaling and communication overhead for servicing mobility and service induced operations. The only requirement is that there exist smart access routers (ARs) capable of processing mobility binding messages issued from MNs to store the current AR locations of the MNs who select them to be the DMAP. A smart router is simply a HMIPv6-aware AR capable of serving as a DMAP for a MN with an internal routing table storing the MN's current CoA and thus capable of routing IPv6 packets destined to the MN to the MN's CoA.

Specifically, when a MN crosses a DMAP service area (to be determined by DMAPwSR on a per-user basis), it makes the AR of the subnet just crossed as the DMAP as in HMIPv6. The MN also determines the size of the new DMAP service area (or the DMAP domain) for which the MN makes use of cross-layer knowledge regarding its mobility and service behaviors. Concurrently, it acquires a RCoA as well as a CoA from the current subnet and registers the address pair (RCoA, CoA) to the current DMAP (the AR of the current subnet) in a binding request message. Note that the RCoA could be the same as the CoA upon the MN's entry into a new DMAP domain. The MN also informs the HA and CNs of the new RCoA address change in another binding message so that the HA and CNs would know the MN by its new RCoA address. When the HA and CNs subsequently send packets to the MN, they would use the RCoA as the MN's address. When a packet is routed to the MN, it will come to the DMAP first because all packets with the RCoA as the destination IP address will be routed through the DMAP first. The DMAP will examine the RCoA and, based on the (RCoA, CoA) entry found in the routing table, route the packet by the MN to its current CoA. Here we note that a CoA is an IP address allocated from an AR to a MN when the MN moves into the AR's subnet. A RCoA is the same as a CoA except that it is allocated from a DMAP and thus is updated to the MN's HA and CNs who know the MN only by its RCoA. Consequently whenever a MN moves across a DMAP service area, it must obtain a new RCoA from the new DMAP and update the RCoA to the HA and CNs. It should also

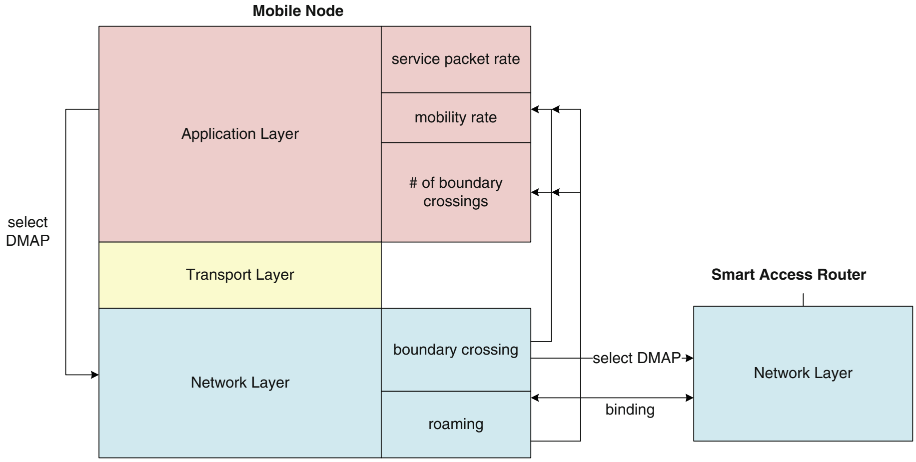

Fig. 1 DMAPwSR cross-layer design

be noted that neither HMIPv6 nor DMAPwSR is applicable to MIPv4 since unlike MIPv6, MIPv4 does not have a mechanism to allow a CN to encapsulate IP packets and store the CoA of a MN without modifying the IP stack of the CN, and consequently MIPv4 does not allow a MN to update its address change to CNs.

By inspecting the address pair (RCoA, CoA) stored in the internal table, the DMAP knows that the MN's address is actually the current CoA and will forward the packet to the MN through tunneling. If the RCoA and CoA are in the same subnet, the DMAP can directly forward the packet to the MN without using tunneling. When the MN subsequently crosses a subnet but is still located within the service area, it would inform the DMAP of the CoA address change without informing the HA and CNs to reduce the network signaling cost. This "DMAP table lookup" design maps RCoA to CoA by having the current DMAP maintain an internal table, so the DMAP can intercept a packet destined for RCoA and forward it to the MN's CoA. It is efficient since the RCoA-CoA routing function can be performed efficiently by DMAPs (which are routers) through simple table lookup operations. All packets destined to a MN under a DMAP will come to the DMAP through IP routing because they have the same IP subnet address. Upon receiving a packet, a smart AR simply looks up its routing table to map the packet's destination address RCoA to CoA for a MN in its DMAP domain. This operation rides for free because an AR has to do a routing table lookup anyway for IP routing whenever it receives a packet. It is scalable because the design is scalable to a large number of MNs by having all ARs in MIPv6 networks DMAP-enabled and randomly spreading the routing and table lookup functions to all ARs in the network. In terms of security and fault tolerance, it can also leverage existing solutions in HMIPv6 because this design is HMIPv6-compliant except that a MN dynamically selects ARs to be MAPs.

As illustrated by Fig. 1, the cross-layer design of DMAPwSR refers to the fact that the MN gains knowledge about its mobility behaviors through network-layer binding messages such as the rate at which it crosses subnets and the number of subnets it has crossed since the last time it registers with a DMAP, and it gains knowledge about its service behaviors through application-layer messages such as the service packet rate. Leveraging cross-layer knowledge regarding its mobility and service characteristics, a MN sitting at the application

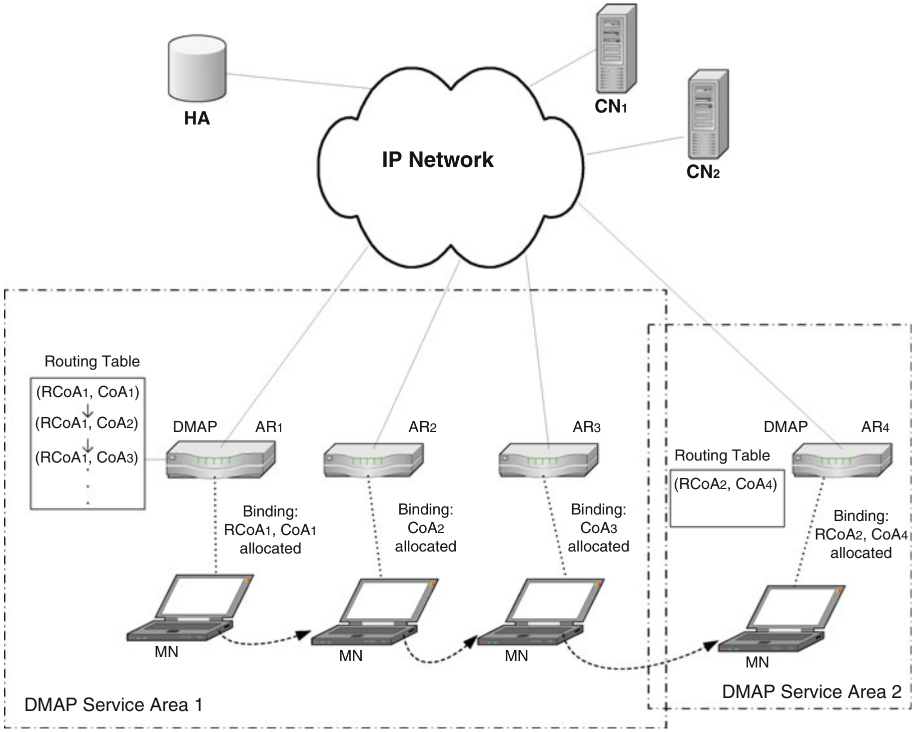

Fig. 2 Example mobility and service management scenarios under DMAPwSR

layer then cooperates with smart ARs sitting at the network layer to minimize the network cost for mobility and service management.

Figure 2 illustrates the DMAPwSR scheme. When the MN enters AR_1 in DMAP service area 1 (the left DMAP area), it selects AR_1 as the DMAP. The MN acquires $RCoA_1$ as well as CoA_1 from AR_1 and an entry ($RCoA_1$, CoA_1) is recorded in the routing table of AR_1 serving as the MN's current DMAP. The HA and CNs are informed of the MN's RCoA address, i.e., $RCoA_1$. When the MN moves across AR_2 but still within DMAP service area 1, the MN only informs the DMAP of the new CoA address (CoA_2) without informing the HA and CNs. After a local binding is made, an entry ($RCoA_1$, CoA_2) is updated in the routing table of the DMAP. A CN knows the MN only by its RCoA. When a CN sends a packet to the MN, it will send it by its RCoA (i.e., $RCoA_1$) which will be intercepted by the DMAP who by examining its routing table will know that the MN is currently in the subnet of AR_2 and will forward the packet to the MN accordingly. If subsequently the MN moves to AR_3 but still within DMAP service area 1, the MN again only informs the DMAP of the new CoA address (CoA_3) without informing the HA and CNs. After a local binding is made, an entry ($RCoA_1$, CoA_3) is updated in the routing table of the DMAP. If subsequently the MN moves to AR_4 in DMAP service area 2, since it crosses the DMAP domain area, it will acquire a new RCoA ($RCoA_2$) as well as a CoA (CoA_4) from AR_4. Then AR_4 becomes the new DMAP which will record in its routing table an entry ($RCoA_2$, CoA_4) for packet routing. The new RCoA (i.e., $RCoA_2$) is sent to the HA and CNs to inform them of the RCoA address change. A CN again knows the MN only by this new RCoA and if it needs to send packets to the MN it will send them with the new RCoA (i.e., $RCoA_2$) as the destination IP address.

In our DMAPwSR scheme, the MN appoints a new DMAP only when it crosses a DMAP service area whose size is determined based on knowledge regarding the MN mobility and service characteristics in the new DMAP service area. Thus, the optimal DMAP service area size is dictated by the MN's mobility and service characteristics. A large DMAP service area size means that the DMAP will not change often. The consequence of not changing the DMAP often is that the service delivery cost would be high because of the triangular routing path CN-DMAP-MN for data communication between the CN and the MN. On the other hand, a small DMAP service area size means that the DMAP will be changed often so it will stay close to the MN. The consequence is that the communication cost for service data delivery would be low because of the short CN-DMAP-MN route. However, a DMAP change involves the cost of informing the HA and CNs of the RCoA address change. There is a trade-off between the large service delivery cost when a large DMAP service area is used versus the large location management cost of informing the HA and CNs of the RCoA address change when a small DMAP service area is used. Hence, an optimal DMAP service area exists. In Sect. 3, we will develop a performance model to analyze this trade-off and identify the optimal service area.

DMAPwSR is movement-based, that is, the DMAP service area size is determined by the number of subnet crossings, say K, the MN moves away from the DMAP. Since a MN may move sideways or even back and forth, a DMAP service area is not necessarily a circular or square area with K subnets as the radius, but an area covering K moves from the last DMAP. The responsibility of determining the best DMAP domain size lies in the MN in cooperation with smart ARs in the system. Specifically, as part of the standard binding process a MN can easily keep track of the number of subnets it has crossed since the last time it registers with a DMAP. When the number of subnet crossings is equal to K, it will request the AR it just moves into to become its new DMAP and update the HA and CNs with its RCoA. Essentially, the optimal DMAP service area, denoted by K_{opt}, depends on the MN's mobility and service behaviors characterized by the MN's SMR. A table can be built based on static analysis listing the best K_{opt} values with SMR as input. Then at runtime A MN can measure its SMR periodically and apply K_{opt} dynamically. A MN with little movement only means that its SMR is large, so K_{opt} should be kept at 1 to minimize the network communication cost for mobility and service management.

3 Performance Model

We develop a mathematical model for analyzing performance characteristics of DMAPwSR based on stochastic Petri net (SPN) techniques. The SPN model resembles that for modeling a MN's mobility behavior in HMIPv6 and DMAP [4]. We choose SPN because of its ability to deal with general time distributions for events, its concise representation of the underlying state machine to deal with a large number of states, and its expressiveness to reason about a MN's behavior as it migrates from one state to another in response to events occurring in the system. The goal is to identify the optimal DMAP service area based on an individual MN's mobility and service behaviors and to compare DMAPwSR with HMIPv6.

The cost metric that we aim to minimize is the *communication cost* incurred *per time unit* due to mobility and service operations. The communication cost includes the signaling overhead for mobility management for informing the DMAP of the CoA changes, and informing the HA and CNs of the RCoA changes, as well as the communication overhead for service management for delivering data packets between the MN and CNs. Table 1 lists a set of

Table 1 Model parameters

Symbol	Meaning
λ	Data packet rate between the MN and CNs
σ	Mobility rate at which the MN moves across boundaries
SMR	Service to mobility ratio (λ/σ)
N	Number of server engaged by the MN
K	Number of subnets in a DMAP service area
τ	1-hop communication delay per packet in wired networks
α	Average distance between HA and DMAP
β	Average distance between CN and DMAP
γ	Cost ratio, wireless versus wired

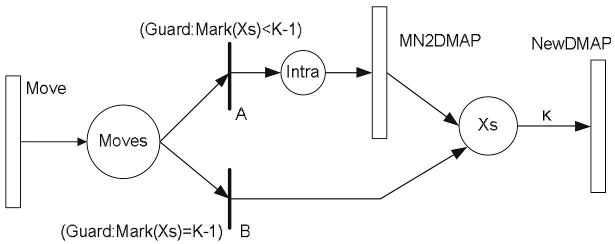

Fig. 3 Performance model based on SPN

identified system parameters that characterize the mobility and service characteristics of a MN in a MIPv6 system.

The SPN model shown in Fig. 3 describes the behavior of a MN operating under DMAP-wSR. We follow the notation used in our earlier work [4] to define the SPN model. The SPN model consists of entities including transitions (`Move`, `A`, `B`, `MN2DMAP` and `NewDMAP`), tokens, places (`Moves`, `Intra` and `Xs`) and arcs. A transition is used to represent the firing of an event. A transition can be a timed transition (e.g., `Move`, `MN2DMAP` and `NewDMAP`) or an immediate transition (e.g., `A` and `B`). A timed transition is fired after an event occurrence time is elapsed, while an immediately transition fires immediately. For example, when the MN moves across a subnet after a residence time in the previous subnet is elapsed, a subnet crossing event occurs. This is modeled by firing transition `Move`. A token is used as a marker; it is used here to represent an event occurrence. For example, when transition `Move` fires, a subnet crossing event occurs, so we place a token in place `Moves` to represent a subnet crossing event. A place is a token holder to contain tokens which represent the number of event occurrences. For example, place `Xs` is used to hold the number of subnet crossing events. Finally, an output arc connects a transition to a place and an input arc connects a place to a transition. An arc is associated with a multiplicity defining the number of tokens that will be moved into the output place (if it is an output arc) or moved out of the input place (if it is an input arc). For example, the arc that connects place `Xs` to transition `NewDMAP` has a multiplicity of K. This means that when transition `NewDMAP` fires, K tokens will be taken out of the input place `Xs`.

The SPN model is constructed as follows:

- Place Xs holds the number of subnet crossings since the last DMAP registration. Initially there is no token in place Xs. By inspecting the number of tokens in place Xs, we will know if the next subnet crossing is an intra-domain move, or an inter-domain move. In the former case, the MN only needs to inform the DMAP of the CoA change. In the latter case, the MN will ask the AR it moves into to serve as its new DMAP, obtain a new RCoA from new DMAP, and inform the HA and CNs of the new RCoA.
- If place Moves holds a token it means that a subnet crossing event just happens.
- When a MN moves across a subnet area, thus incurring a location handoff, a token is put in place Moves. The mobility rate at which location handoffs occur is σ which is the mobility rate of the MN and thus the transition rate assigned to Move.
- If the current move is an intra-domain move, i.e., the number of tokens in place Xs is less than $K - 1$, such that the guard for transition A returns *true*, then the MN will only inform the DMAP of the CoA change. This is modeled by defining a condition associated with transition A to fire transition A when the condition is satisfied, allowing the token in place Moves to move to place Intra. Subsequently, once the MN obtains a CoA from the AR it just moves into, it will inform the DMAP of the new CoA change. This is modeled by enabling and firing transition MN2DMAP. After MN2DMAP is fired, a token in place Intra flows to place Xs, representing that a location handoff has been completed and the DMAP has been informed of the CoA change of the MN.
- If the current move is an inter-domain move, i.e., the number of subnet crossings in place Xs is equal to $K - 1$, such that the guard for transition B returns *true*, then the move will make the MN cross a DMAP service area. This is modeled by enabling and thus firing immediate transition B, allowing the token in place Moves to move to place Xs in preparation for a service handoff event. Note that in an SPN, firing an immediate transition does not take any time.
- If the number of moves, including the current one, in place Xs has accumulated to K, a threshold determined by DMAPwSR representing the size of a DMAP service area, then it means that the MN has just moved into a new DMAP service domain. This is modeled by assigning an enabling function that will enable transition NewDMAP when K tokens have been accumulated in place Xs. After transition NewDMAP is fired, all K tokens are consumed and place Xs contains no token, representing that the number of subnet crossings is reset to zero and the AR of the subnet that the MN just enters has been appointed as the DMAP by the MN in the new DMAP service area.

The stochastic model underlying the SPN model is a continuous-time Markov chain (with the event occurrence time being exponentially distributed) with the state representation of (a, b) where a is the number of tokens in place Moves, b is the number of tokens in place Xs. Let P_i be the steady state probability that the system is found to contain i tokens in place Xs.The steady-state probability P_i, $1 \leq i \leq K$, can be solved easily utilizing numerical method solution techniques such as SOR or Gauss Seidel [23].

Let $C_{i,service}$ be the communication cost for the network to service a data packet given that the MN has moved across i subnets since the last DMAP registration. The communication cost $C_{i,service}$ includes a communication delay between the DMAP and a CN in the fixed network ($\beta\tau$), a delay from the DMAP to the AR of the MN's current subnet in the fixed network ($i\tau$), and a delay in the wireless link from the AR to the MN ($\gamma\tau$). Let $C_{service}$ be the average communication cost to service a data packet weighted by the respective P_i probabilities. Then, $C_{service}$ is calculated as follows:

$$C_{service} = \sum_{i=0}^{K} (P_i \times C_{i,service}) = \gamma\tau + \beta\tau + \sum_{i=0}^{K} (P_i \times i\tau) \tag{1}$$

Let $C_{i,location}$ be the network signaling overhead to service a location handoff operation given that the MN has moved across i subnets since the last DMAP registration. The communication cost $C_{i,location}$ depends on i. If $i < K$, only a minimum signaling cost will be incurred for the MN to inform the DMAP of the CoA address change, i.e., the cost includes a communication delay in the wireless link from the MN to the AR ($\gamma\tau$) and a communication delay from the AR to the DMAP ($i\tau$) for a total cost of $\gamma\tau + i\tau$. On the other hand, if $i = K$, then the location handoff also triggers a DMAP service handoff. A DMAP service handoff will incur a higher communication signaling cost to inform the HA and N CNs (or application servers) of the RCoA address change. The cost includes a communication delay in the wireless link from the MN to the AR ($\gamma\tau$) and a communication delay from the AR to the HA and N CNs ($\alpha\tau + N\beta\tau$) for a total cost of $\gamma\tau + \alpha\tau + N\beta\tau$. Let $C_{location}$ be the average communication cost to service a move operation by the MN weighted by the respective P_i probabilities. Then, $C_{location}$ is calculated as follows:

$$C_{location} = \sum_{i=0}^{K} (P_i \times C_{i,location}) = P_K(\gamma\tau + \alpha\tau + N\beta\tau) + \sum_{i=0}^{K-1} \{P_i(\gamma\tau + i\tau)\} \tag{2}$$

The total communication cost *per time unit* for the Mobile IP network operating under our DMAPwSR scheme to service operations associated with mobility and service management of the MN, denoted by $C_{DMAPwSR}$, is the sum of the communication cost per data packet delivery multiplied with the rate at which data packets are generated between the MN and CNs, plus the communication cost per location update operation multiplied with the MN's mobility rate, i.e., $C_{DMAPwSR}$ is calculated as follows:

$$C_{DMAPwSR} = C_{service} \times \lambda + C_{location} \times \sigma \tag{3}$$

Here λ is the data packet rate between the MN and CNs, and σ is the MN's mobility rate.

Equations (1), (2) and (3) together allow one to calculate $C_{DMAPwSR}$ as a function of K and determine the optimal K, i.e., K_{opt}, representing the optimal "DMAP service area" size that will minimize the network cost $C_{DMAPwSR}$, i.e., after K_{opt} subnet crossings since the last DMAP registration, the MN will cross a new DMAP service domain and will request the AR it just moves into to be the new DMAP. Here we note again that the optimal DMAP size determined from the analytical model is movement-based, as K_{opt} defines the optimal DMAP size. Below we report numerical results obtained from evaluating the analytical model, when given a set of parameter values reflecting the MN's mobility and service behaviors, as well as MIPv6 network conditions.

4 Numerical Results with Simulation Validation

By utilizing Eqs. (1), (2) and (3) from solving the SPN model developed we can obtain the total communication cost incurred per time unit analytically. In this section we report numerical results with simulation validation. We adopt event-driven simulation and use SMPL [17] as our simulation tool for its simplicity and effectiveness. SMPL provides basic constructs to allow one to create events with distinct event types and priorities, schedule events to occur at their occurrence times, and control event processing when an event occurs (e.g., when a mobility event occurs, it could be an intra-domain move or an inter-domain move).

SMPL implicitly maintains the event list, determines the most imminent event to process, and advances the simulation clock when an event occurs. Lastly, SMPL provides basic constructs allowing one to collect data with statistical significance for reporting performance measures of interest.

A MN is characterized by its own mobility rate σ and service rate λ. When a MN moves out of a DMAP service area, it incurs a service handoff and moves into a new DMAP service area with the first AR it moves into being the new DMAP. DMAPwSR is movement-based, that is, the DMAP service area size is determined by the number of movements the MN moves away from the DMAP so it does not have any network model in mind because all it concerns is the number of subnet crossings by the MN. The DMAP service area, however, can be distance-based, i.e., the DMAP service area size is determined by the distance between the current subnet and the DMAP. We compare simulation results of movement-based DMAP service areas versus distance-based DMAP service areas to see if the results are sensitive to the definition of DMAP service areas. We also consider three network coverage models in simulation to test if the results obtained are sensitive to the network coverage model used.

The first network coverage model is a two-dimensional hexagonal-shape network coverage model as shown in Fig. 4. Assume that the MN moves in accordance with random walk [1] by which a MN stays in a subnet for a while and then moves from the current AR to one of the 6 neighbor ARs randomly with equal probability of 1/6. The whole area is wrapped around so the structure can be reused. That is, if the MN moves out of the simulated area, its location will be circled to the other side of the simulated area, i.e., its location will be changed from (x, y) to $(-x, -y)$, thus allowing the simulated area to be reused. The simulation system maintains the locations of all MNs and their DMAPs according to the protocol used, e.g., a MN's DMAP is the first AR in the new DMAP service area upon a service handoff. The hexagonal network model can be used to simulate movement-based or distance-based DMAP service areas. We mark a distance-based DMAP service area in boldface in Fig. 4. Suppose a subnet area is represented by its center location with radius r. A distance-based DMAP service area will have a radius of $(2K - 1)r$ where K represents the DMAP service area size.

The second network model is the mesh network coverage model as illustrated in Fig. 5 [1]. In this model, a MN moves from the current AR to one of the 4 neighbor ARs randomly with equal probability of 1/4 of moving into a neighbor subnet. We mark several distance-based DMAP service areas with distances 1 through 4 in Fig. 5.

The third network model is based on real trace data, i.e., a real world wireless network consisting of access points (APs) on the campus of Dartmouth College. The trace data [12] collected by CRAWDAD (http://craw\discretionary-dad.cs.dart\discretionary-mouth.edu) is a comma-separated list of APs on campus along with their location information expressed in terms of latitude and longitude values. There are 695 APs on campus. The location information of each AP includes the MAC address, AP name, latitude, and longitude. However, some AP actual locations are unknown. We remove those APs whose locations are unknown. Figure 6 shows the distribution of APs on the x–y coordinate system. In the trace-data based network coverage model, we consider two APs as neighbor APs if they are separated in distance in the range of [100, 200 m], taking into account the fact that the AP signal coverage range is about 300 feet (91.4 m) [16]. For mobility events, when a MN leaves an AP, it randomly selects one of its neighbor APs to move into.

Regardless of which network coverage model is being used, in the simulation we keep track of the locations of all MNs and their DMAPs based on the mobility and service management protocols used. Since the simulation program knows the locations of a MN and its DMAP all the time, whenever a mobility or service management event occurs, such as the

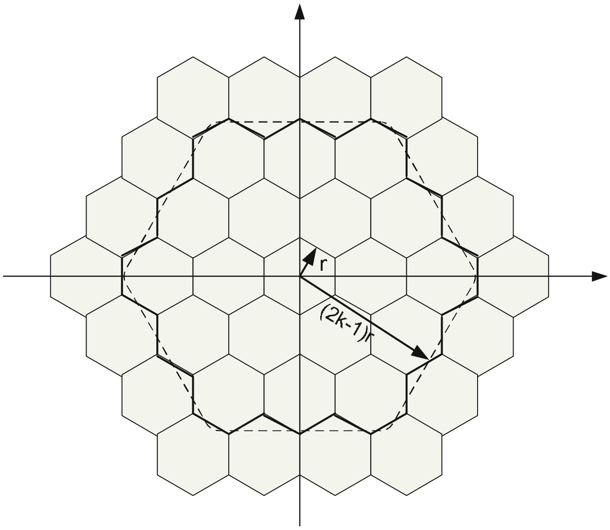

Fig. 4 A service area under hexagonal network coverage model

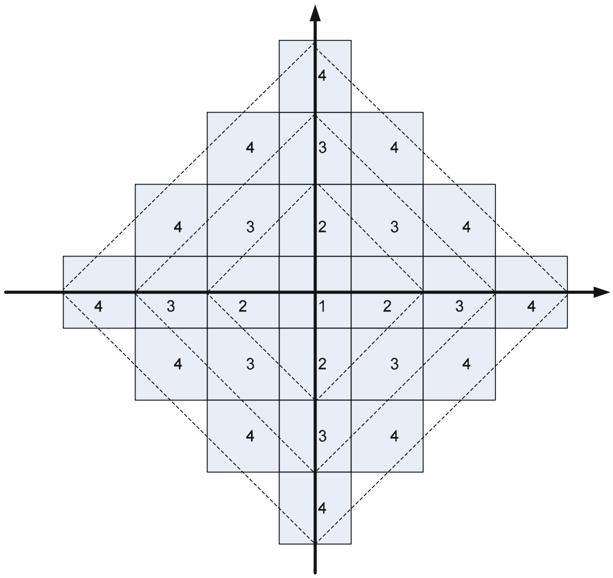

Fig. 5 A service area under mesh network coverage model

MN moves to another AR or the CN generates a packet to the MN, it knows exactly the cost incurred in response to a movement or a service request event.

Below we report analytical results obtained from evaluating the SPN model developed versus simulation results obtained. Table 2 lists the parameter values used in our analysis. The 1-hop communication delay per packet τ accounts for the transmission delay of a packet of size 1,024 bytes over an effective bandwidth 10 Mbps for the wired network, i.e., $\tau = (1,024 \times 8)/(10 \times 10^6)$. The cost ratio of wireless versus wired network γ is 10,

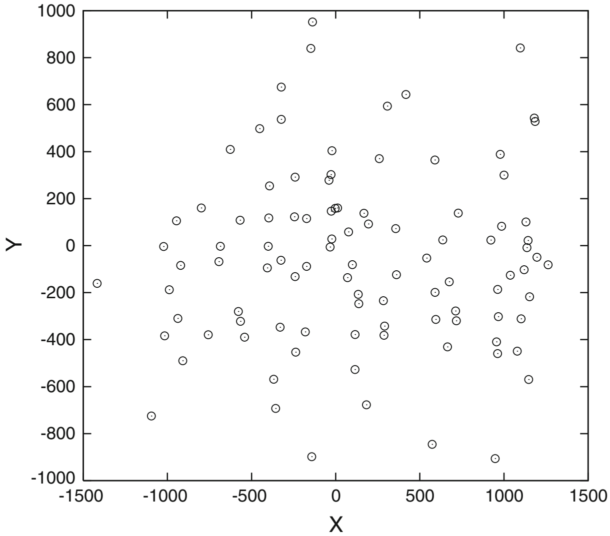

Fig. 6 Access point locations at Dartmouth College campus

Table 2 Parameters and default values used in performance analysis

Parameters	Default value
τ	0.0008 s
N	1
α	{10, 20, 30, 40} hops
β	{10, 20, 30, 40} hops
γ	10
λ	{1/8, 1/4, 1/2, 1, 2, 4, 8, 16, 32, 64}
σ	{1/16, 1/8, 1/4, 1/2, 1}

considering an effectiveness bandwidth of 1 Mbps for the wireless network. The number of CNs that a MN concurrently engages is assumed 1. Finally since the location of the CN and the location of the HA can be anywhere in the network, we assume that the average distance between between the CN and DMAP, denoted by α, and the average distance between the HA and DMAP, denoted by β, can be between 10 and 40 hops.

4.1 Comparison of DMAPwSR with MIPv6 and HMIPv6

We first compare performance characteristics of DMAPwSR versus two baseline schemes including MIPv6 and HMIPv6.

For MIPv6, we calculate its total cost per time unit as follows. The communication cost $C_{service}^{MIPv6}$ for servicing a packet delivery in basic MIPv6 includes a communication delay from the CN to the AR of the current subnet, and a delay in the wireless link from the AR to the MN. Thus,

$$C_{service}^{MIPv6} = \beta\tau + \gamma\tau \tag{4}$$

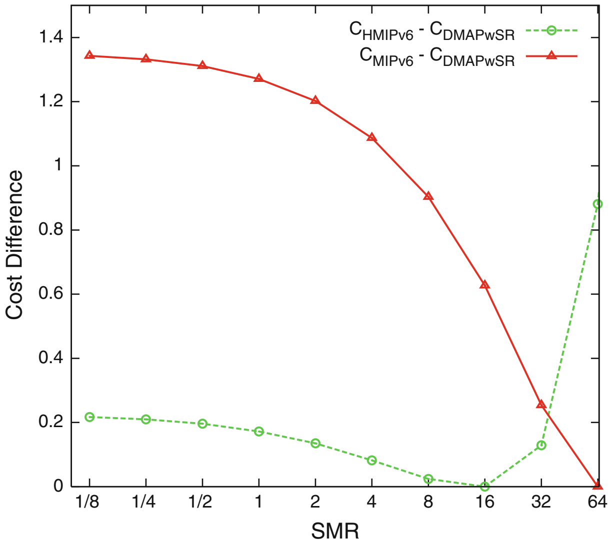

Fig. 7 Cost difference between MIPv6, HMIPv6 and DMAPwSR

The communication cost $C^{MIPv6}_{location}$ for servicing a location handoff consists of a delay in the wireless link from the MN to the AR of the subnet that it just enters into, a delay from that AR to the CNs to inform them of the CoA change, and a delay from that AR to the HA to inform the HA of the CoA change. Thus,

$$C^{MIPv6}_{location} = \gamma\tau + \alpha\tau + N\beta\tau \quad (5)$$

Following Eq. (3), the total cost per time unit for servicing data delivery and mobility management operations under MIPv6 is given by:

$$C_{MIPv6} = C^{MIPv6}_{service} \times \lambda + C^{MIPv6}_{location} \times \sigma \quad (6)$$

For HMIPv6, the placement of MAPs is pre-determined. That is, there are designated ARs that serve as MAPs for MNs. We compare DMAPwSR with an implementation of two-level HMIPv6 in which each MAP covers a fixed-size area, say, covering K_H subnets. We utilize our performance model to obtain performance measures of HMIPv6 with K set to K_H.

Figure 7 compares the communication cost difference incurred per time unit by MIPv6, HMIPv6 and DMAPwSR as a function of SMR. The total cost incurred per time unit by DMAPwSR is computed by Eq. (3) with the DMAP service area set at the optimal K_{opt} values in response to varying SMR values. The total cost incurred per time unit by HMIPv6 is also computed by Eq. (3) with the DMAP service area set at K_H. The total cost incurred per time unit by MIPv6 is given by Eq. (6).

Figure 7 shows two lines. The solid line shows the cost difference between basic MIPv6 and DMAPwSR ($C_{MIPv6} - C_{DMAPwSR}$), and the dotted line shows the cost difference between HMIPv6 and DMAPwSR ($C_{HMIPv6} - C_{DMAPwSR}$), as a function of SMR. We see that DMAPwSR dominates basic MIPv6 when SMR is low. As SMR increases exceeding a threshold (e.g., 64 in this case), K_{opt} approaches 1 under which DMAPwSR degenerates to basic MIPv6. The reason is that when SMR is sufficiently high, the MN's packet arrival rate is much higher than the mobility rate, so the data delivery cost dominates the mobility management cost. Therefore, the MN's DMAP will stay close to the MN to lower the

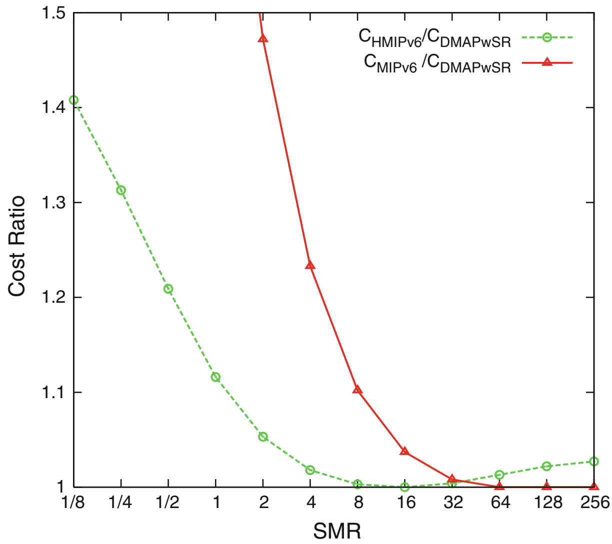

Fig. 8 Cost ratio between DMAPwSR and MIPv6/HMIPv6

data delivery cost, thus making $K_{opt} = 1$ in our DMAPwSR scheme in order to reduce the CN-DMAP-MN triangular routing cost for packet delivery. Next we observe that the cost difference between HMIPv6 and DMAPwSR (the dotted line) initially decreases as SMR increases until K_{opt} coincides with K_H at which point DMAPwSR degenerates to HMIPv6, and then the cost difference increases sharply as SMR continues to increase. We conclude that DMAPwSR performs better than HMIPv6 when SMR is either low and high.

Correspondingly, Fig. 8 shows the cost ratio curves of $C_{HMIPv6}/C_{DMAPwSR}$ and $C_{MIPv6}/C_{DMAPwSR}$ as a function of SMR to better see the percentage of cost increase when HMIPv6 or MIPv6 is used instead of DMAPwSR. When HMIPv6 is used instead of DMAPwSR, the percentage cost increase goes from 40, 20, 10, 0 to 5 % as SMR goes from 1/8, 1/2, 1, 8, to 256. Here we observe that under low SMR (e.g., online texting, chatting and shopping while moving) the cost gain of DMAPwSR over HMIPv6 can go as high as 40 %, while under high SMR (e.g., online audio/video while moving), the cost gain is less than 5 %. We should emphasize that the cost gain is per-MN per time unit, so for a cost gain even as low as 5 %, the cumulative gain over all MNs over a long period of operational time would still be significant.

Lastly Fig. 9 illustrates the effect of α and β. One can see that the cost difference between HMIPv6 and DMAPwSR widens as α or β increases as the cost saving of DMAPwSR compared with HMIPv6 is especially pronounced when the distance between the HA (or CN) and the MN is high.

4.2 Simulation Validation

To ensure statistical significance of simulation results, we use the *batch mean analysis* technique [17] by which the simulation period is divided into batch runs with each batch consisting of 2,000 observations for computing the average value. A minimum of 10 batches were run to compute the grand mean for the *overall network cost* metric. Additional batches are added if necessary until the grand mean is within 95 % confidence level and 10 % accuracy from

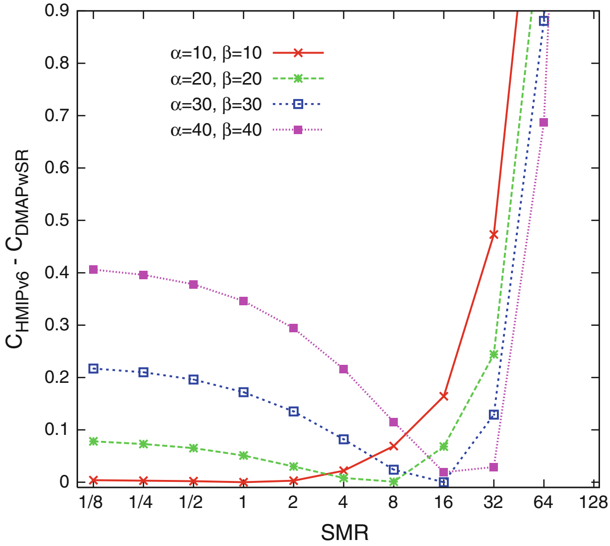

Fig. 9 Effect of α and β on cost difference between HMIPv6 and DMAPwSR

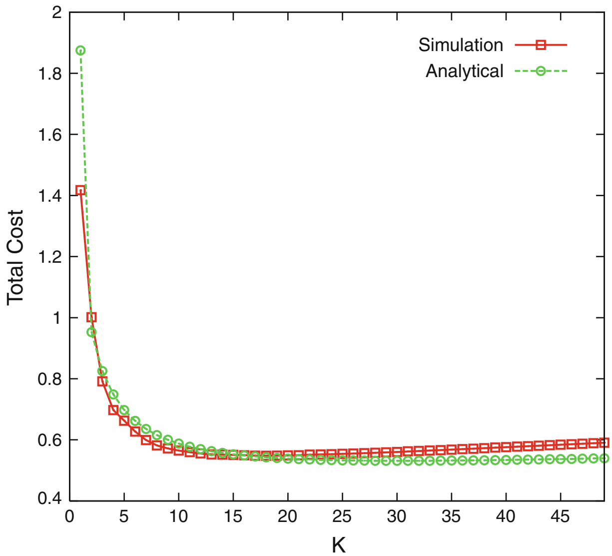

Fig. 10 Simulation versus analytical results for C_{total} versus service area size K in DMAPwSR

the true mean. With the batch mean analysis technique, a simulation run for collecting a data point will not end until the expected confidence level and accuracy are achieved. To achieve the confidence level and accuracy of 0.95 and 0.05, it normally takes more than 20,000 observations.

Figures 10 and 11 compare analytical results versus simulation results for DMAPwSR. Figure 10 shows that an optimal DMAP service area size K_{opt} exists that can best balance the trade-off between the large service delivery cost when a large DMAP service area is

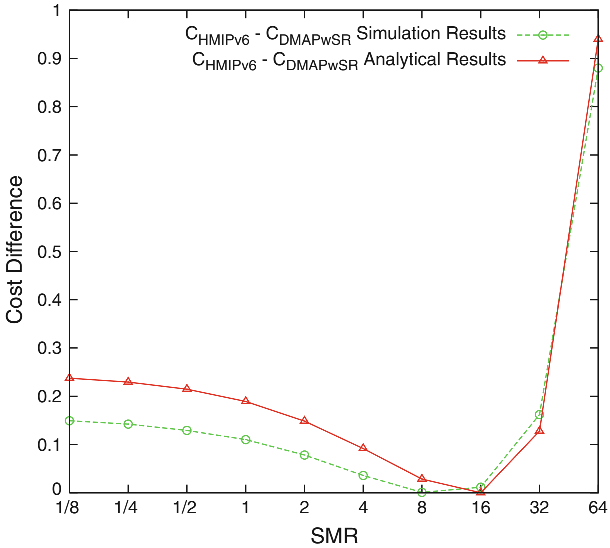

Fig. 11 Simulation versus analytical results: cost difference between HMIPv6 and DMAPwSR

used versus the large location management cost of informing the HA and CNs of the RCoA address change when a small DMAP service area is used. Figure 11 summarizes the cost difference between HMIPv6 and DMAPwSR as a function of SMR. We use solid line for analytical results and dashed line for simulation results. We observe that simulation results obtained match well with analytical results with the same trend exhibited. We conclude that DMAPwSR performs significantly better than HMIPv6 when SMR is either low and high.

The results reported above are based on distance-based DMAP service areas as shown in Figs. 4 and 5. We test the sensitivity of simulation results obtained with respect to the definition of a DMAP service area, i.e., movement-based versus distance-based. The distance-based DMAP service area size is determined by the distance between the current subnet and DMAP. The movement-based DMAP service area size is determined by the number of movements the MN moves away from the DMAP. Note that the analytical model considers only movement based DMAP service areas, which ignores the possibility of back-and-forth movements. Hence, by comparing simulation results based on movement-based DMAP service areas versus those based on distance-based DMAP service areas, we could measure the imprecision introduced by movement-based DMAP service areas.

Figure 12 compares simulation results obtained under movement based service areas versus those under distance-based DMAP service areas for DMAPwSR. We see that the simulation results in general and the trend exhibited in particular are insensitive to the definition of the DMAP service area.

Next we study the sensitivity of simulation results obtained with respect to the residence time distribution, i.e., the distribution of the time a MN stays at a subnet, including the normal, uniform and exponential distributions. Note that all analytical results reported are based on the residence time being exponentially distributed. Figure 13 compares the cost difference between HMIPv6 and DMAPwSR under uniform, normal and exponential residence time distributions. We observe that the trend remains about the same and the results remain valid

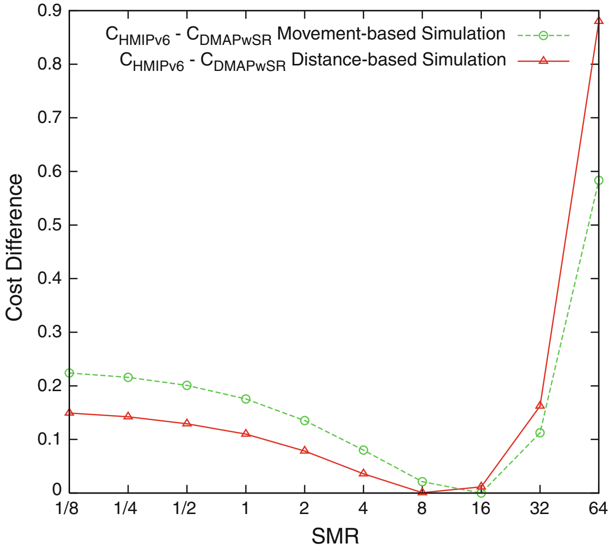

Fig. 12 Cost difference under movement-based versus distance-based service area simulation

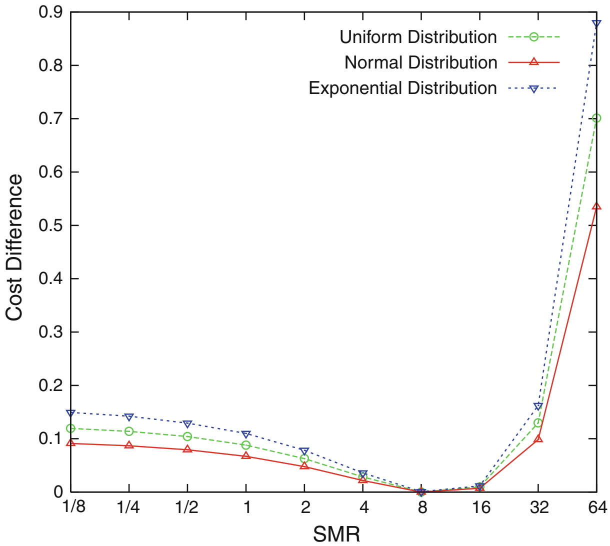

Fig. 13 Cost difference under different residence time distribution

irrespective of the type of time distribution considered for modeling the subnet residence time.

Figure 14 shows optimal K versus SMR under various time distributions (normal, uniform and exponential), We see from this figure that the trend exhibited under these different time distributions is remarkably similar and the results in general are insensitive to the time distribution.

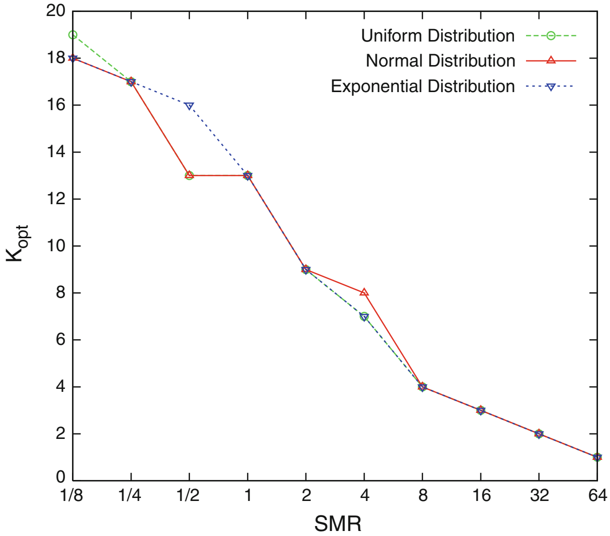

Fig. 14 Optimal K versus SMR under various time distributions

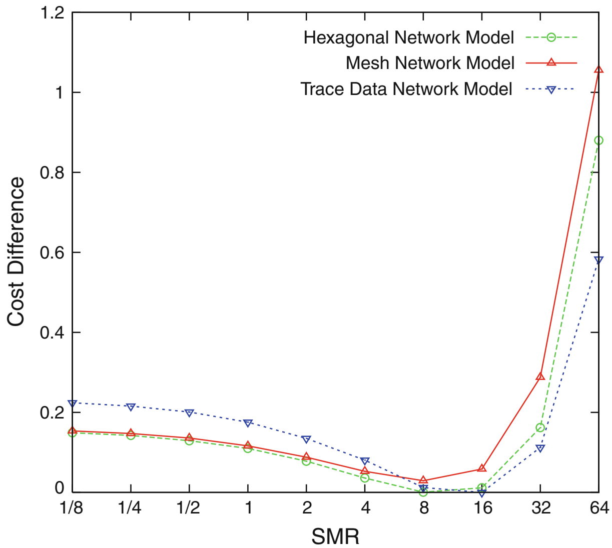

Fig. 15 Cost difference under different network coverage model

So far all results reported are based on the hexagonal network coverage model. Next we test if simulation results would be sensitive to the network coverage model used, including hexagonal, mesh and trace-data based network coverage models introduced in the beginning of this section. Figure 15 compares the cost difference between HMIPv6 and DMAPwSR under hexagonal, mesh and trace-data based network coverage models. We observe that the general trends exhibited by the simulation results obtained under hexagonal, mesh and trace-data based network coverage models are remarkably consistent with each other. Thus we conclude that simulation results obtained are virtually insensitive to the network coverage model used.

5 Conclusion

In this paper, we study the performance of a novel DMAPwSR scheme for integrated mobility and service management with the goal to minimize the overall mobility and service management cost. We develop an analytical model based on stochastic Petri nets to analyze DMAPwSR and compare its performance against HMIPv6. We validate analytical solutions obtained through extensive simulation including sensitivity analysis of simulation results with respect to the network coverage model, the MN's residence time distribution and the DMAP service area definition. Our scheme outperforms HMIPv6 in terms of the network communication overhead, the effect of which is especially pronounced when the SMR is either low or high. The performance gain is in the amount of communication cost saved per time unit per user, so the saving due to a proper selection of the best DMAP service area will have significant impacts since the cumulative effect for all mobile users over a long time period would be significant.

For future work, we plan to extend this research to consider other performance metrics such as QoS, network utilization, and request blocking probability. Also we plan to extend this research to consider DMAP selection issues based on load balancing principles while still being able to maintain the optimal regional service area for each MN to minimize the total network cost incurred to the system. Here we note that the added functionality required out a smart AR is actually quite minimal. A smart AR essentially still performs the same function of a router except that additionally it maintains a (RCoA, CoA) table entry for each MN choosing it as a DMAP. Hence one research direction is to investigate the limitation and feasibility of making all IPv6 routers DMAP-compliant. Lastly, we plan to explore the concept of cross-layer integrated mobility and service management in emerging wireless mesh networks [13,14] by designing a forwarding and resetting protocol under which the length of the forwarding chain corresponds to the DMAP service area size.

References

1. Akyildiz, I., Lin, Y., Lai, W., & Chen, R. (2000). A new random walk model for PCS networks. *IEEE Journal on Selected Areas in Communications, 18*(7), 1254–1260.
2. Akyildiz, I., Xie, J., & Mohanty, S. (2004). A survey on mobility management in next generation all-IP based wireless systems. *IEEE Wireless Communications, 11*(4), 16–28.
3. Chen, I. R., Gu, B., & Cheng, S. (2006). On integrated location and service handoff schemes for reducing network cost in personal communication systems. *IEEE Transactions on Mobile Computing, 5*(2), 179–192.
4. Chen, I. R., He, W., & Gu, B. (2007). DMAP: Efficient integrated mobility and service management for mobile IPv6 systems. *Wireless Personal Communications, 43*(2), 711–723.
5. Chen, I. R., He, W., & Gu, B. (2007). Proxy-based regional registration for integrated mobility and service management for network cost minimization in mobile IP systems. *The Computer Journal, 50*(3), 281–293.
6. Chen, I. R., & Wang, D. C. (2010). Regional registration-based mobile multicast service management in mobile IP networks. *Wireless Personal Communications, 54*(4), 635–649.
7. Das, S., Mcauley, A., Dutta, A., Chakraborty, K., & Das, S. K. (2002). IDMP: An intradomain mobility management protocol for next-generation wireless networks. *IEEE Wireless Communications, 9*, 38–45.

8. Gu, B., & Chen, I. R. (2005). Performance analysis of location-aware mobile service proxies for reducing network cost in personal communication systems. *ACM Mobile Networks and Applications, 10*(4), 453–463.
9. Gustafsson, E., Jonsson, A., & Perkins, C. (2007). *Mobile IPv4 regional registration*. IETF, RFC 4857. http://tools.ietf.org/html/rfc4857.
10. He, W., & Chen, I. R. (2009). Proxy-based integrated cache consistency and mobility management in mobile IP systems. *Journal of Parallel and Distributed Computing, 69*(6), 559–572.
11. Johnson, D., Perkins, C., & Arkko J. (2004). *Mobility support in IPv6*. IETF, RFC 3775. http://www.ietf.org/rfc/rfc3775.txt.
12. Kim, M., Fielding, J., & Kotz, D. (2006). Risks of using AP locations discovered through war driving. In *Proceedings of the fourth international conference on pervasive computing*, Dublin, Ireland (pp. 67–82).
13. Li, Y., & Chen, I. R. (2011). Adaptive per-user per-object cache consistency management for mobile data access in wireless mesh networks. *Journal of Parallel and Distributed Computing, 71*(7), 1034–1046.
14. Li, Y., & Chen, I. R. (2011). Design and performance analysis of mobility management schemes based on pointer forwarding for wireless mesh networks. *IEEE Transactions on Mobile Computing, 10*(3), 349–361.
15. Lim, T., Yeo, C., Lee, F., & Le, Q. (2009). TMSP: Terminal mobility support protocol. *IEEE Transactions on Mobile Computing, 8*, 849–863.
16. Luo, H., & Shankaranarayanan, N. (2004). A distributed dynamic channel allocation technique for throughput improvement in a dense WLAN environment. In *IEEE international conference on acoustics, speech, and signal processing*, Montreal, Canada.
17. MacDougall, M. H. (1987). *Simulating computer systems: Techniques and tools*. Cambridge, MA: MIT Press.
18. Mohanty, S., & Akyildiz, I. (2007). Performance analysis of handoff techniques based on mobile IP, TCP-migrate and SIP. *IEEE Transactions on Mobile Computing, 6*(7), 731–747.
19. Mohanty, S., & Akyildiz, I. F. (2006). A cross-layer (layer 2 + 3) handoff management protocol for next-generation wireless systems. *IEEE Transactions on Mobile Computing, 5*, 1347–1360.
20. Pack, S., Shen, X., Mark, J. W., & Pan, J. (2007). Adaptive route optimization in hierarchical mobile IPv6 networks. *IEEE Transactions on Mobile Computing, 6*, 903–914.
21. Perez-Costa, X., Torrent-Moreno, M., & Hartenstein, H. (2003). A performance comparison of mobile IPv6, hierarchical mobile IPv6, fast handovers for mobile IPv6 and their combination. *SIGMOBILE Mobile Computing and Communications Review, 7*(4), 5–19.
22. Soliman, H., Castelluccia, C., El-Malki, K., & Bellier, L. (2008). *Hierarchical mobile IPv6 mobility management*. IETF, RFC 5380. http://tools.ietf.org/html/rfc5380.
23. Trivedi, K., Ciardo, G., & Muppala, J. (1999). *SPNP version 6 user manual*. Durham, NC: Department of Electrical Engineering, Duke University.
24. Yeh, J., Chen, J., & Agrawal, P. (2009). Fast intra-network and cross-layer handover (finch) for wimax and mobile internet. *IEEE Transactions on Mobile Computing, 8*, 558–574.

Author Biographies

Ding-Chau Wang received the B.S. degree from Tung-Hai University, Taichung, Taiwan, and the M.S. and Ph.D. degrees in computer science and information engineering from National Cheng Kung University, Tainan, Taiwan. He is currently an assistant professor in the Department of Information Management at the Southern Taiwan University, Tainan, Taiwan. His research interests include distributed systems, mobile computing, security, and performance analysis.

Weiping He received his BE degree in Electrical Engineering from Huazhong University of Science and Technology, China, in 1996, and his M.S. and Ph.D. degrees in Computer Science from Virginia Tech in 2002 and 2009, respectively. His research interests include mobile computing, wireless networks, performance analysis, mobile data management, and mobility and service management in mobile computing environments.

Ing-Ray Chen received the B.S. degree from the National Taiwan University, Taipei, Taiwan, and the M.S. and Ph.D. degrees in computer science from the University of Houston. He is a professor in the Department of Computer Science at Virginia Tech. His research interests include mobile computing, wireless systems, dependable computing, security, data management, real-time intelligent systems, and reliability and performance analysis. Dr. Chen currently serves as an editor for *Wireless Personal Communications, Wireless Communications and Mobile Computing, The Computer Journal, Security and Network Communications, IEEE Communications Letters, IEEE Transactions on Network and Service Management, and International Journal on Artificial Intelligence Tools*. He is a member of the IEEE and ACM.