DMAP-FR: Dynamic Mobility Anchor Points for Mobility, Service and Failure Recovery Management in Mobile IPv6 Systems

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

Abstract   In this paper, we propose and analyze DMAP-FR, a mobility and service management scheme with failure recovery (FR) control in Mobile IPv6 systems. The basic idea behind DMAP-FR is to leverage access routers (ARs) running as regional mobility anchor points (MAPs) as in Hierarchical Mobile IPv6 (HMIPv6) for mobility and service management for mobile nodes (MNs). However, unlike HMIPv6, DMAP-FR allows the MAP of each MN to be determined dynamically based on the mobility and service characteristics of the MN and the failure behavior of ARs with the goal to minimize the network traffic. DMAP-FR incorporates fault tolerance mechanisms to allow the system to quickly recover from AR and MAP failures. We identify the best dynamic regional area size for the selection of MAP for each MN such that the overall network traffic due to servicing mobility, service and fault tolerance related operations is minimized. We demonstrate that DMAP-FR outperforms HMIPv6 for the same AR failure rate.

Keywords   MIPv6 · Fault tolerance · Failure recovery · Mobility management · Service management · Performance analysis

1 Introduction

With the success of the Internet, it is widely believed that IP will become the foundation of next generation wireless networks. Mobile IPv6 (MIPv6) [7] is a network protocol for

W. He · I.-R. Chen (✉)
Department of Computer Science, Virginia Tech, Northern Virginia Graduate Center,
7054 Haycock Road, Falls Church, VA 22043, USA
e-mail: irchen@vt.edu

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

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

enabling mobility in IPv6 networks. It allows mobile nodes (MNs) to move within IP-based networks while maintaining on-going connections. MIPv6 has been flagged as the mobility management protocol for future all-IP mobile systems and is expected to have wide deployment. With the anticipated increase in inexpensive, computationally powerful mobile devices, more and more mobile applications will access multimedia and data services over broadband wireless connections based on IPv6. Two issues remained to be solved before wide deployment of MIPv6. One issue is to devise effective mobility and service management schemes to reduce the network traffic. Another issue is to provide fault tolerance for service continuity despite network router failures.

In this paper, we investigate DMAP-FR, an integrated mobility, service and failure management scheme for MIPv6 systems. The core idea is to extend the notion of mobility anchor points (MAPs) in Hierarchical Mobile IPv6 (HMIPv6) [14] with the notion of "dynamic" mobility anchor points (DMAPs) for each individual MN instead of static ones for all MNs. These DMAPs are simply access routers (ARs) chosen by individual MNs to act as a regional router 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 the mobility and service characteristics of each MN and the failure behavior of ARs. The goal is to identify the optimal service area size dynamically. By executing a computational procedure developed, a MN can optimally determine when and where to launch a dynamic MAP so as to minimize the network cost in serving mobility management, service management and fault tolerance related operations. We demonstrate the resulting network signaling cost saving to be significantly better than that provided by HMIPv6 with failure recovery support. Furthermore, DMAP-FR is more scalable than HMIPv6 because (a) unlike HMIPv6 where there are only a handful of fixed MAPs scattered around to serve all MNs, in DMAP-FR every AR can serve as a DMAP, thereby making DMAP-FR more scalable to a large MN population; and (b) the network cost reduction benefit as a result of applying DMAP-FR is cumulative and proportional to the MN population.

While the use of MAPs proposed in HMIPv6 can reduce the signaling cost for mobility management, failures of static MAPs can disrupt services engaged by MNs under failed MAPs. Moreover, having a single MAP serve all MNs under its coverage can overload the MAP and deteriorate the MAP's performance. From the reliability point of view, a wireless network with a hierarchy of MAPs as in HMIPv6 is not robust because there are more failure points. DMAP-FR is based on a two-tier MAP architecture to provide robust fault tolerance to recover failures of ARs and MAPs so as to ensure service continuity for ongoing service applications.

The rest of the paper is organized as follows. Section 2 describes related work. Section 3 describes DMAP-FR in detail. Section 4 develops a computational procedure for determining the optimal DMAP domain size to minimize the network communication cost induced by mobility, service and failure management operations. In Sect. 5, we present numerical results and conduct an extensive simulation study to validate analytical results. Finally, Sect. 6 summaries the paper and outlines some future research areas.

2 Related Work

In MIPv6 [13] a MN is identified by its permanent home IP address. If the MN is not in its home area, it obtains a temporary IP care of address (CoA) to associate with its current location. The MN's home agent (HA) maintains a dynamic mapping between the MN' home address and its CoA. When an application server, called a corresponding node (CN),

sends packets to the MN by the MN's home address, the HA intercepts them and forwards them to the MN. In MIPv6, the MN can also inform the CN of its new CoA to allow the CN to send packets directly to the MN to avoid the overhead of CN-HA-MN triangular routing.

Hierarchical MIPv6 (HMIPv6) [14] is a proposed standard from the Internet Engineering Task Force (IETF) 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. Since its inception in 2001, it has evolved into a well accepted standard (IETF RFC 5380) for mobility management of MIPv6. In addition to a CoA, a Regional CoA (RCoA) is also allocated to a MN whenever the MN enters a new MAP 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. This concept can be applied at multiple levels in a hierarchical manner [12,14] and can be combined with the use of fast handover (IETF RFC 5568) [5,8] utilizing forwarding pointers between the current and next subnets in a hybrid manner.

In HMIPv6, MAPs are statically configured and shared by all MNs in the system. Access routers (ARs) are responsible for announcing their MAP's identity by means of router advertisement packets so that roaming MNs would know if they have crossed a MAP domain and need to perform a RCoA update to the HA and CNs. For fault tolerance, a MN can detect MAP failures in HMIPv6 [14] when it receives a router advertisement containing a MAP option with a lifetime of zero. The MN will then start the MAP discovery process and attempt to register with another MAP. After it has selected and registered with another MAP, it informs the correspondent nodes (CNs) and the home agent (HA) that its RCoA has changed. However, this recovery process may suffer a long delay.

Recently Yeh, Chen and Agrawal proposed a fast intra-network and cross-layer handover protocol called FINCH in MIPv6 environments [17]. FINCH was shown to perform better than MIPv6, Fast-MIP [8], and HMIPv6 in the mobility and packet delivery management cost. However, the issue of failure recovery was not addressed. You, Pack and Choi [18] proposed robust hierarchical mobile IPv6 (RH-MIPv6) to provide fault tolerance and robustness in MIPv6 networks. A MN registers its primary (P-RCoA) and secondary (S-RCoA) regional care of addresses to two different MAPs (primary and secondary) simultaneously. Either the MN or CN can detect the failure of the primary MAP and change their attachment from the primary to the secondary MAP using the Internet Control Message Protocol for IPv6 (ICMPv6) [4]. However, there is a high signaling overhead for sending binding messages to the MAPs, HA, and CNs.

Omar, Saadawi and Lee [11] suggested that a child foreign agent (FA) of a failure FA can bind directly to the HA, or it can use the parent FA to replace the failure FA. It is not clear of the performance impact to the overall traffic. Also the design is for MIPv4 only.

Ghosh and Varghese [6] used passive replication techniques to mask failures of mobile agents in Mobile IP environments. Each mobile agent maintains bindings of all MNs registering within its network coverage. If a mobile agent receives a registration request message from a MN, it processes the message, forwards the message to its peers and then waits for acknowledgment messages from them. This protocol causes a high overhead during failure free periods especially when there are many mobile agents in the network. This protocol is also not scalable to large MNs.

3 DMAP-FR Design

We extend mobility anchor points (MAPs) in HMIPv6 [14] and in [2] with the notion of dynamic MAPs to support integrated mobility, service and fault tolerance management dynamically in Mobile IPv6 environments. These DMAPs are selected by individual MNs instead of being static ones for all MNs as in HMIPv6. More specifically, these DMAPs are access routers (ARs) chosen by individual MNs to act as a regional router to reduce the signaling overhead for intra-regional movements. The DMAP domain size, i.e., the number of subnets covered by a DMAP, is based on the MN's mobility and service characteristics and the failure behavior of ARs.

Under our DMAP-FR protocol, a MN interacts with its HA and CNs (representing the application servers) as in the MIPv6 protocol, but optimally determines when and where to launch its DMAP to minimize the network cost in serving the user's mobility, service and failure management related operations. When a MN starts in a MIPv6 environment, a DMAP is created to run on the current AR to interact with the HA and CNs. The DMAP communicates with CNs on behalf of the MN as if it were the MN. The DMAP moves only when the MN crosses a "service area" thus incurring a *service handoff*. A large service area size means that the DMAP will not move often. The consequence of not moving 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 MN. On the other hand, a small service area size means that the DMAP will be moved 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 move involves the cost of informing the HA and CNs of the DMAP address change, and possibly a service context transfer cost from the last location to the current location. Therefore, there is a trade-off between these two cost factors and an optimal service area exists. We aim to determine the optimal service area size for each individual MN based on the MN's mobility and service characteristics and the failure behavior of ARs. Here we note that the time and space complexity of DMAP-FR would remain the same as that of DMAP without failure recovery [2] because a DMAP/AR failure recovery event will be handled in the same way a service handoff event is being handled, i.e., by selecting a new DMAP/AR.

The mobility and service characteristics of a MN and the failure profile of ARs can be summarized by three parameters. The first parameter is the residence time that the MN stays in a subnet. This parameter can be collected by each MN based on statistical analysis [3]. We expect that future MNs are reasonably powerful for collecting data and doing statistical analysis. The residence time in general would be characterized by a general distribution. We use the MN's mobility rate (σ) to represent this parameter. The second parameter is the service traffic between the MN and server applications. The MN can also collect data statistically to parameterize this. We use the data packet rate (λ) between the MN and CNs to represent this parameter. The ratio of λ over μ is called the service to mobility ratio (SMR) of the MN. The third parameter is the failure time of an AR. We use the AR failure rate δ_f to represent this parameter.

We assume that for each AR, there is an overlapping coverage from other ARs, so when an AR fails, the MNs attached to it can be reassigned to other overlapping ARs. In the case that a router (which is not a MAP) fails or a link goes down, it can be handled by the recovery mechanism of the routing protocol (e.g., OSPF). Lastly since the number of subnets separating two communicating processes would not properly measure the hop-count distance, we let $F(K)$ denote a function that returns the number of hops as a function of the number of subnets K. This function can be periodically and dynamically determined by a MN which
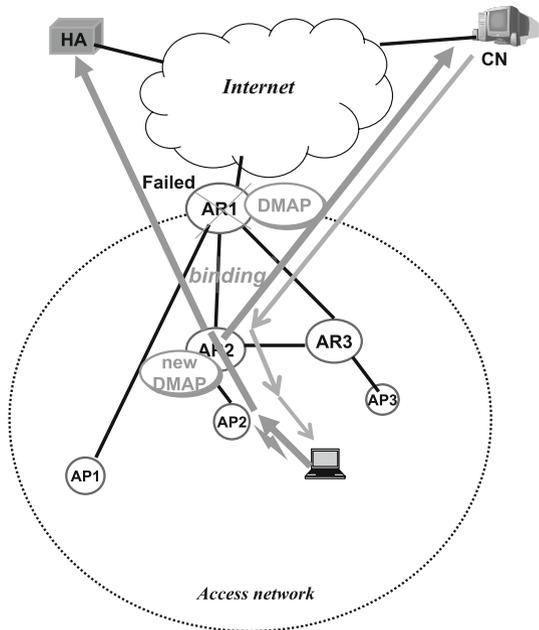
collects statistical data as it roams across subnets. The fluid flow model [19] assumes that the average number of hops between two communicating processes separated by K subnets is equal to $\sqrt{K}$.

As part of the handoff protocol, a MN knows when it enters a new subnet. Therefore, each MN would keep track of the number of subnets it has crossed since the last DMAP was selected. When the number reaches the optimal K value identified by applying the computational procedure devised in this paper, the MN will select the AR of the subnet it just enters as the new DMAP. Here we note that since a MN has its own optimal DMAP domain size and can choose any AR to be its DMAP, it is possible that an AR can serve as the DMAP for multiple MNs, with different DMAP domain sizes for different MNs.

The failure/disconnection of a MN's DMAP can be detected by not receiving any announcement message from the DMAP in a timeout period. The fault tolerance property of DMAP-FR stems from the fact that each MN can dynamically select an AR as its DMAP. Both DMAP and AR failures can be tolerated. Once an AR is recovered, it can be put back to the system seamlessly without affecting the system operation.

The failure recovery procedure in DMAP-FR considers three cases:

- Failure of the MN's DMAP which is not the current AR: As shown in Fig. 1, suppose that the MN is currently under AR2 and the current DMAP is AR1 that has failed. In this case, the current AR (i.e., AR2) becomes the MN's DMAP. AR2 will inform the HA and CNs that it is now the DMAP.
- Failure of the MN's DMAP which is the current AR: As shown in Fig. 2, the MN is under AR1 which is the current DMAP and it fails. In this case, since the wireless coverage area of the current AR is overlapping, the MN could be under radio range of several other subnets. The MN will register with a new AR (i.e., AR2) nearby which will become the MN's DMAP. AR2 will inform the HA and CNs of the RCoA change. In case the original DMAP recovers, the MN can sense a stronger signal and can switch back to the original DMAP.

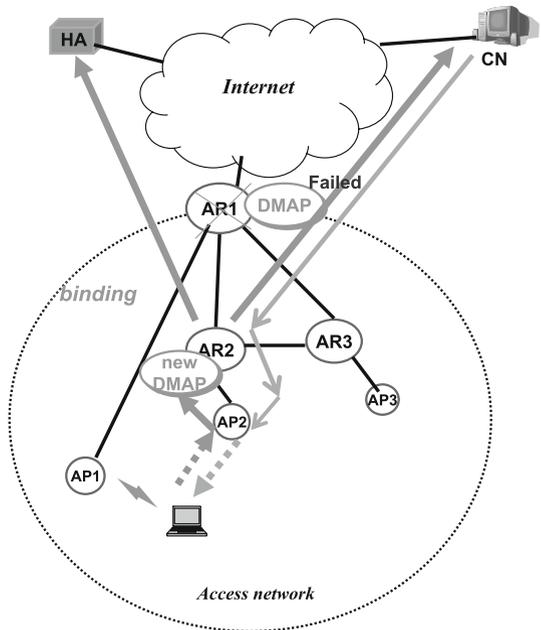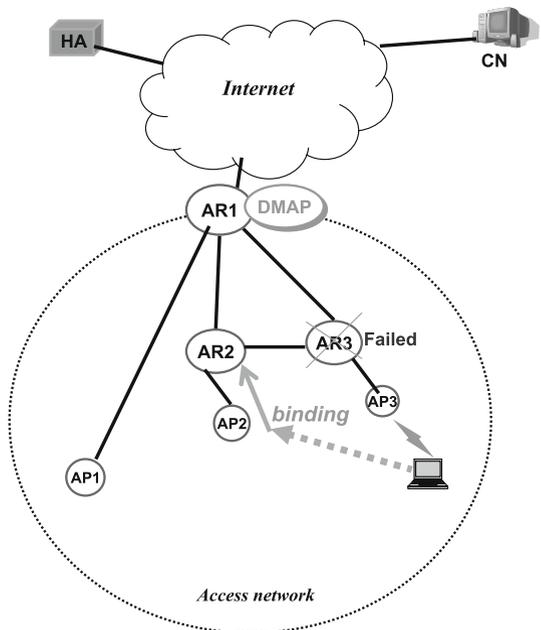**Fig. 2** DMAP-FR failure
recovery case 2

Fig. 3 DMAP-FR failure
recovery case 3

- Failure of the MN's current AR: As shown in Fig. 3, the MN is under AR3 when AR3 fails and the DMAP is on AR1. In this case the DMAP locates another AR, i.e., AR2, to replace AR3. The MN will register with the new AR to the DMAP through a binding message.

Table 1 Parameters for DMAP-FR

Symbol	Meaning
λ	Data packet rate between the MN and CNs
σ	Mobility rate at which the MN moves across subnet boundaries
δ_f	Failure rate of DMAP/AR
N	Number of servers engaged by the MN
K	Number of subnets in one service area
τ	1-hop communication delay per packet in wired networks
α	Average distance between HA and MAP
β	Average distance between CN and MAP
γ	Cost ratio between wireless versus wired network

Table 1 lists a set of identified system parameters that characterize the mobility and service characteristics of a MN and the failure behavior of ARs in a MIPv6 system. Here a server refers to a corresponding node (CN) sitting on the Internet with which the MN is communicating. An AR is just an IP router in the Mobile IP network but can serve as a MN's DMAP if it is selected by the MN.

4 Performance Analysis

We devise a computational procedure to be executed by a MN operating under DMAP-FR to determine the optimal service area size utilizing stochastic Petri net (SPN) techniques [15]. 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.

Figure 4 shows the performance model based on Stochastic Petri nets. Table 2 gives the meaning of places and transitions defined in the model. The function Mark (P) returns the number of tokens in place P. The number of tokens accumulated in place Xs, that is, Mark (Xs), represents the number of subnets crossed by the MN since the MN enters a new service
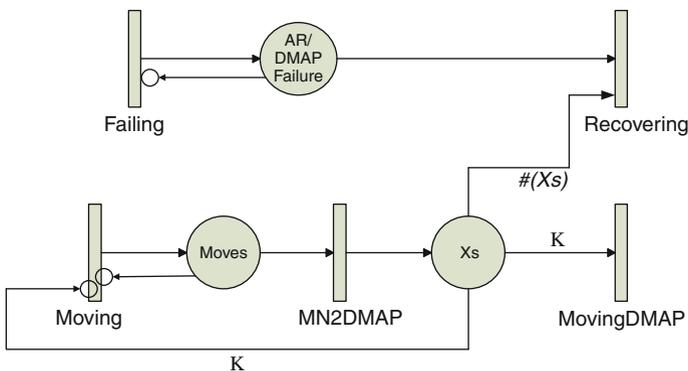

Fig. 4 Stochastic petri net model

Table 2 Meanings of places and transitions in the stochastic petri net model

Symbol	Meanings
`Moves`	Mark `(Moves)`=1 means that the MN just moved into a new subnet area.
`Moving`	A timed transition for the MN to move across a subnet area.
`MN2DMAP`	A timed transition for the MN to communicate with the DMAP.
`Xs`	Mark `(Xs)` indicates the number of subnets crossed.
`MovingDMAP`	A timed transition for the DMAP to move into a new server area.
`K`	number of subnets crossed after which a service handoff will occur.
`Failing`	A timed transition for an AR/DMAP to fail.
`Recovering`	A timed transition for an AR/DMAP to recover.
`AR/DMAP Failure`	Mark `(AR/DMAP Failure)`=1 means that the AR/DMAP just failed.

area. We allow it to accumulate to K, at which point we perform a service handoff. We construct the SPN model to describe the behavior of a MN operating under our proposed integrated mobility, service and failure management scheme:

- The mobility rate at which location handoffs occur is σ which is the transition rate assigned to `Moving`. When a MN moves across a subnet area, a token is put in place `Moves`.
- After moving into a subnet, the MN obtains a new CoA and informs the DMAP (that acts as a gateway foreign agent) of the CoA change. This is modeled by enabling and firing transition `MN2DMAP` while disabling transition `Moving`. After `MN2DMAP` is fired, a token in place `Moves` flows to place `Xs`, representing that a location handoff has been completed and the DMAP has been informed of the CoA address change of the MN.
- If the number of tokens in place `Xs` has accumulated to K, a threshold representing the size of a service area, then it means that the MN has just moved into a new service area and a service handoff ensues. This is modeled by assigning an enabling function that will enable transition `MovingDMAP` after K tokens have been accumulated in place `Xs`. After transition `MovingDMAP` is fired, all K tokens are consumed and place `Xs` contains no token, representing the action that the DMAP has just moved into a new service area. The rate at which transition `MovingDMAP` fires depends on the cost of informing the HA and CNs of the DMAP CoA change.
- The DMAP alternates between "work" and "failure" states. Initially the DMAP is in the work state. After a time is elapsed, the DMAP goes to the failure state. This is modeled by transition `Failing`. Note that if the MN is already in place `AR/DMAP Failure`, transition `Failing` cannot fire.
- While the MN is in failure mode, after a time is elapsed representing the recovery time, the DMAP goes to the work state. This is modeled by firing transition `Recovering`.
- For case 1, the DMAP fails but the current AR is alive, as illustrated in Fig. 1 where the MN is under AR2, the DMAP is at AR1 and fails. In this case, the current AR (i.e., AR2) will become DMAP, the new DMAP (AR2) will inform the HA and CNs of the RCoA change. This is modeled by firing transition `Recovering` with a transition rate reflecting the cost which we will parameterize later. Firing this transition will flush all the tokens in place `Xs` as if a service handoff had happened. This is modeled by a variable input arc from place `Xs` to transition `Recovering`.
- For case 2, the DMAP fails and the current AR happens to be the DMAP, as illustrated in Fig. 2 where the MN's current AR and DMAP is AR1 and AR1 fails. In this case, the

MN will register with a new AR nearby. The new AR (i.e., AR2) will become the MN's DMAP who will then inform the HA and CNs of the RCoA address change. This event is also modeled by firing transition `Recovering`.

- For case 3, the current AR fails but the DMAP is alive, as illustrated in Fig. 3 where the DMAP is AR1 and the MN's current AR is AR3 which has failed. In this case, the MN will register with another AR nearby (e.g., AR1 or AR2). The new AR then only needs to inform the DMAP of the CoA change. This event can also be modeled by transition `Recovering`. Note that the rate to transition `Recovering` depends on the system state. Later we will parameterize the rate to transition `Recovering`.

Below we parameterize transition rates of `MN2DMAP`, `MovingDMAP` and `Recovering` based on the set of base parameters defined in Table 1.

The firing time of transition `MN2DMAP` stands for the communication time of the MN informing the DMAP of the new CoA through the wireless network. This time depends on the number of hops separating the MN and its DMAP. Thus, the transition rate of transition `MN2DMAP` may be calculated as

$$\frac{1}{\gamma\tau + F(Mark(\text{Xs}) + 1) \times \tau}$$

where τ stands for the one-hop communication delay per packet in the wired network and γ is a proportionality constant representing the ratio of the communication delay in the wireless network to the communication delay in the wired network. $F(Mark(\text{Xs}) + 1)$ returns the number of hops between the current subnet and the DMAP separated by $Mark(\text{Xs}) + 1$ subnets. The argument of the $F(x)$ function is added by 1 to satisfy the initial condition that $Mark(\text{Xs}) = 0$ in which the DMAP has just moved into a new service area, so at the first subnet crossing event, the distance between the DMAP and the subnet is one subnet apart. Note that this transition rate is state-dependent because the number of tokens in place `Xs` changes dynamically over time.

When transition `MovingDMAP` fires, the AR of the subnet that the MN moves into will be selected as the DMAP. The communication cost includes that for the MN to inform the HA and CNs of the new RCoA address change, i.e.,

$$(\alpha + N\beta)\tau$$

where α is the average distance in hops between the MN and the HA, β is the average distance in hops between the MN and a CN, and N is the number of CNs that the MN concurrently engages. Consequently, the rate of transition `MovingDMAP` is calculated as

$$1/(\alpha + N\beta)\tau$$

When transition `Recovering` fires, the MN will contact the DMAP. If the DMAP fails and the current AR is the DMAP, the MN will register with a new AR nearby. The new AR will become the DMAP and inform the HA and CNs of the RCoA change. If the DMAP fails while the current AR is still alive, the current AR will become the DMAP. The new DMAP will inform the HA and CNs. In either case, the current AR chosen becomes the new DMAP and the cost involved is to inform the HA and CNs of the DMAP's RCoA address change. Since a new DMAP is $F(Mark(\text{Xs}))$ hops away from the failed DMAP, the rate to transition `Recovering` can be parameterized as

$$\frac{1}{\{N[\beta + F(Mark(\text{Xs}))] + [\alpha + F(Mark(\text{Xs}))] + \gamma\}\tau}$$

A MN and its DMAP determine the service area dynamically to minimize the overall network signaling costs for mobility management, service management and fault tolerance related operations incurred by the MN. There are three costs to be considered: (1) the service cost; (2) the mobility cost; (3) the failure recovery cost. The overall network traffic cost that we aim to minimize is the sum of these three costs *per time unit*.

Let $C_{service}$ be the average communication cost to service a data packet. Let $C_{mobility}$ be the average communication cost to service a location handoff, including one that can trigger a service handoff. Let $C_{recovery}$ be the average communication cost to perform failure recovery. Finally let C_{total} be the overall cost incurred *per time unit*. Then, C_{total} is the sum of the product of the respective communication cost multiplied with the rate at which the respective event occurs, that is,

$$C_{total} = C_{service} \times \lambda + C_{mobility} \times \sigma + C_{recovery} \times \delta_f \tag{1}$$

where λ is the data packet rate between the MN and CNs, σ is the MN's mobility rate and δ_f is the DMAP failure rate.

Below we derive $C_{service}$, $C_{mobility}$ and $C_{recovery}$. The stochastic model underlying the SPN model is a continuous-time semi-Markov chain with the state representation of (a, b, c) where a is the number of tokens in place Moves, b is the number of tokens in place Xs, and c is the number of tokens in place AR/DMAP Failure. Let P_i be the steady state probability that the system is found to contain i tokens in place Xs such that Mark(Xs) = i. P_i's can be obtained by applying numerical analysis methods such as SOR or Gauss Seidel to solve the underlying model.

Let $C_{i,mobility}$ be the network signaling overhead to service a mobility handoff operation given that the MN is in the ith subnet in the service area. If $i < K$, only a minimum signaling cost will be incurred for the MN to inform the DMAP of the CoA address change. On the other hand, if $i = K$, then the mobility handoff also triggers a service handoff. A service handoff will incur a higher communication signaling cost to inform the HA and N CNs (or application servers) of the RCoA address change. Thus, $C_{mobility}$ is calculated as follows:

$$C_{mobility} = \sum_{i=0}^{K}(P_i \times C_{i,mobility})$$

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

Let $C_{i,service}$ be the communication overhead for the network to service a data packet when the MN is in the ith subnet in the service area. The communication overhead includes a communication delay between the DMAP and a CN in the fixed network, a delay from DMAP to the AR of the MN's current subnet in the fixed network, and a delay in the wireless link from the AR to the MN. Thus, $C_{service}$ is calculated as follows:

$$C_{service} = \sum_{i=0}^{K}(P_i \times C_{i,service})$$

$$= \sum_{i=0}^{K} P_i(\gamma\tau + \beta\tau + F(i)\tau)$$

$$= \gamma\tau + \beta\tau + \sum_{i=0}^{K}(P_i \times F(i)\tau) \tag{3}$$

Let $C_{i,recovery}$ be the communication overhead for the network to recover from DMAP or AR failures when the MN is in state i. If the current AR fails while the DMAP is alive, the MN will register with another AR nearby. The communication cost incurred is for the MN to inform the DMAP of the local CoA address change. If the current DMAP fails, the MN will choose the current AR as the new DMAP, and inform the HA and N CNs (or application servers) of the RCoA address change. Therefore,

$$C_{i,recovery} = \begin{cases} \gamma\tau + F(Mark(i) + 1)\tau \\ \quad \text{if the current AR fails while the DMAP is alive} \\ \gamma\tau + (\alpha\tau + F(Mark(i) + 1)\tau) + \\ \quad N\beta\tau + F(Mark(i) + 1)\tau) \quad \text{if the DMAP fails} \end{cases} \quad (4)$$

Thus, $C_{recovery}$ is calculated as follows:

$$C_{recovery} = \sum_{i=0}^{K} (P_i \times C_{i,recovery}) \quad (5)$$

The overall cost C_{total} is a function of K and P_i which can be calculated from the SPN model. Therefore, by varying K values, one can easily find K_{opt} representing the optimal service area size that would minimize the overall cost under the DMAP-FR scheme.

5 Results with Simulation Validation

In this section, we first present analytical results obtained from applying the computational procedure discussed in Sect. 4. Then we analyze the effect of key parameters on the results obtained. Lastly we present simulation validation.

5.1 Analytical Results

The example is based on $F(K) = \sqrt{K}$ [19], $\alpha = \beta = 30$ and $\gamma = 5$. The cost metric is normalized with respect to $\tau = 1$. We vary the values of key parameters to analyze their effects on K_{opt}.

Figure 5 shows that under DMAP-FR there exists an optimal service area size K_{opt} to minimize the overall network traffic cost, when given a set of parameter values characterizing the mobility and service behaviors of the MN and failure behaviors of ARs in Mobile IP networks.

We observe from Fig. 6 that K_{opt} increases as δ_f increases. The reason is that as the failure rate increases, the MN's DMAP likes to stay at a large service area to reduce the location handoff cost such that a location handoff will most likely only involve informing the DMAP of the location change without incurring a service handoff to migrate the DMAP.

Figure 7 summarizes the cost difference between HMIPv6 and DMAP-FR as a function of SMR. We observe that the cost difference between HMIPv6 and DMAP-FR initially decreases as SMR increases until K_{opt} coincides with the fixed regional area size K_H at which point DMAP- FR degenerates to HMIPv6, and then the cost difference increases sharply as SMR continues to increase. We conclude that DMAP-FR outperforms HMIPv6 with failure recovery.
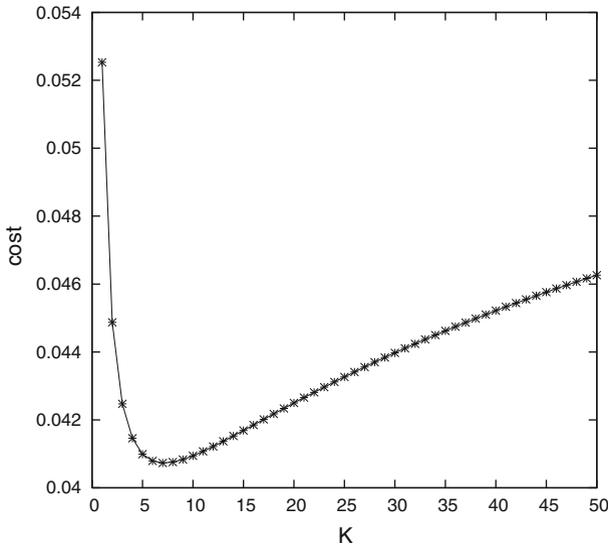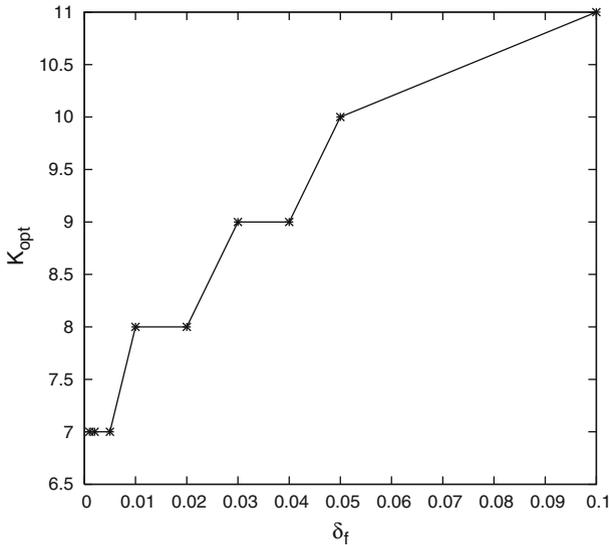
Fig. 5 Cost versus K

Fig. 6 K_{opt} versus δ_f

5.2 Simulation

We conduct an extensive simulation study to imitate operations of DMAP-FR and HMIPv6 with failure recovery support in Mobile IP networks. We first discuss how we set up the simulation environment. Then we present simulation results for simulation validation of analytical results obtained earlier. We adopt event-driven simulation in our simulation study and use SMPL [9] as the simulation software for simplicity and effectiveness. To ensure statistical
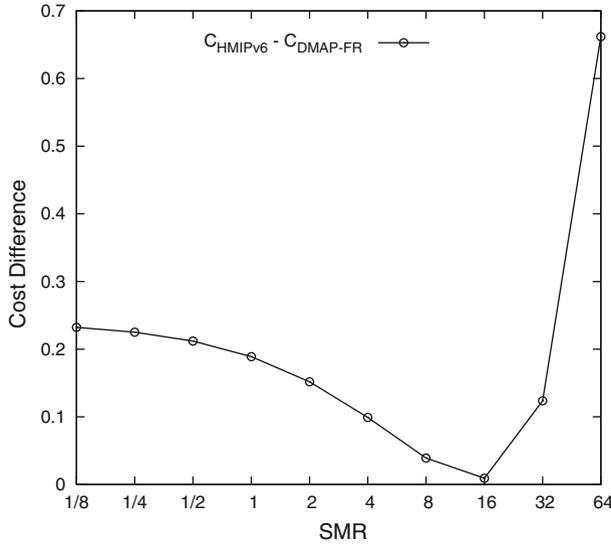
Fig. 7 Cost difference between HMIPv6 and DMAP-FR

significance of simulation results, we use the *batch mean analysis* technique [9] 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 was 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 the true mean.[1] 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 95% confidence level and 10% accuracy, it normally takes more than 20,000 observations.

There are 61 ARs located at the centers of 61 subnets organized in a 5-level hexagonal network coverage model as shown in Fig. 8, with each hexagon representing a subnet. The hexagonal network is wrapped-around, so when a MN moves out of a subnet at one end of the hexagonal network, it actually moves into a subnet at the other end of the network. In effect, it simulates a large mobile IP network with no operational area boundary. A handoff is triggered when the MN moves across a subnet boundary, the rate of which is determined by the mobility rate parameter σ. A handoff is also triggered when the current subnet in which the MN resides fails, the rate of which is determined by the AR failure rate parameter δ_f.

We assume a random walk model [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 as illustrated in Fig. 9.

As illustrated in Fig. 8, a subnet area is represented by its center location with radius r. A service area of size k has radius $(2k - 1)r$ where k represents the service area size in a

---

[1] Given N sample results $Y_1, Y_2, \ldots, Y_N$, assuming all normally and independently distributed since each of which is a batch mean of 2000 observations, the *confidence accuracy* is defined as H/Y, where H is the *confidence interval half-width* and Y is the grand mean of the results ($Y = (Y_1 + Y_2 + \cdots + Y_N)/N$). The *confidence level* $(1-\alpha)$ is defined as the probability that the absolute value of the difference between the Y and μ (the true mean of the sample results) is equal to or less than H, i.e., $Prob(Y - H <= \mu <= Y + H) = 1 - \alpha$. H is defined by $H = t_{\alpha/2; N-1} \times \sigma/\sqrt{N}$, where σ^2 is the sample variance given by $\sigma^2 = \Sigma_i (Y_i - Y)^2/(N-1)$ and t is the standard t distribution [16].
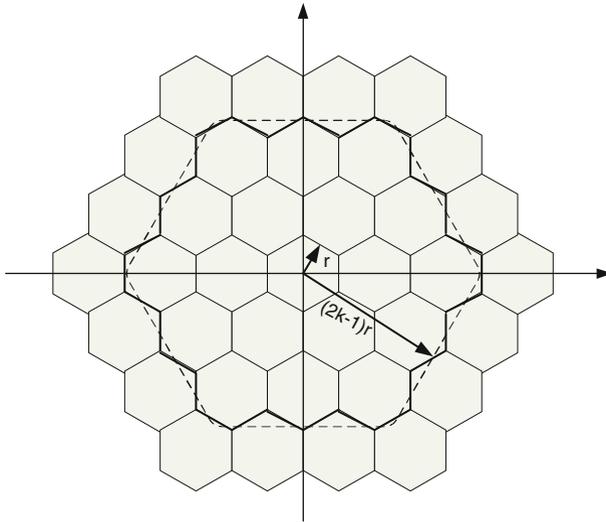
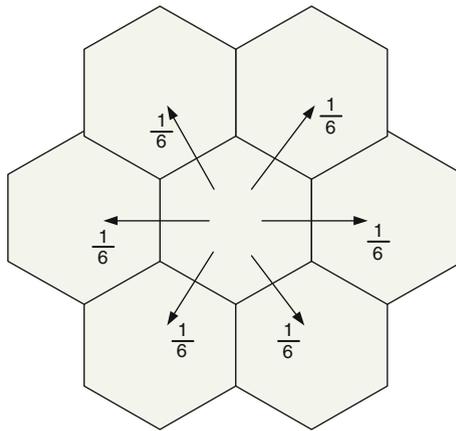**Fig. 8** A service area under hexagonal network coverage model

Fig. 9 Routing pattern in hexagonal network coverage model

regional registration based mobility management scheme such as DMAP-FR and HMIPv6. The center of the simulation area is at (0, 0). Each hexagonal AR area is represented by its center location (x, y). A MN is characterized by its own mobility rate σ and service rate λ. When a MN moves out of a service area, it incurs a service handoff and moves into a new service area with the first AR it moves into being at the center of the service area. 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. At all times, the location of the MN is known. As the simulation program knows the locations of the MN and its service proxy all the time, whenever a movement, an AR failure, or a service request event occurs, it knows exactly the cost incurred in response to the event. These per-event costs are then accumulated to the
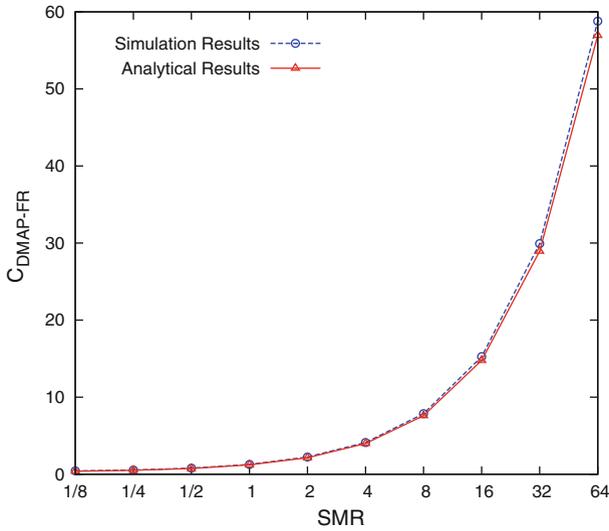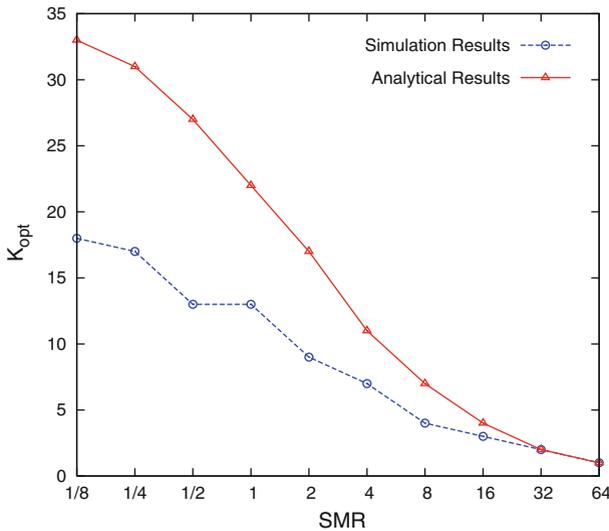
Fig. 10 Cost of DMAP-FR versus SMR

Fig. 11 Optimal DMAP domain size K_{opt} as a function of SMR

overall cost during the course of the simulation. At the end of each simulation batch run, the average cost is computed by dividing the cumulative cost over the simulation period.

Figure 10 shows the overall cost incurred for mobility, service and failure recovery management by the system operating under DMAP-FR as a function of SMR. Both analytical results (solid line) and simulation results (dashed line) are shown for ease of comparison. Figure 11 shows simulation results vs. analytical results for K_{opt} vs. SMR. This validates the analytical result obtained earlier that the optimal service area size K_{opt} decreases as SMR increases. We observe that in general simulation results show very good correlations with
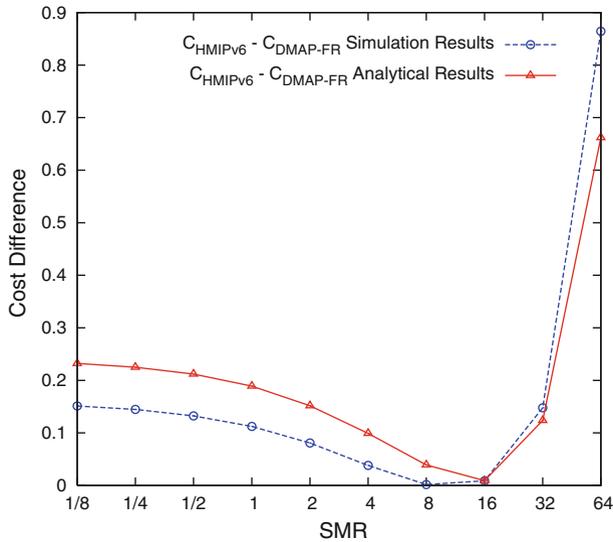
Fig. 12 Simulation versus analytical results: cost difference between HMIPv6 and DMAP-FR

analytical results although a small discrepancy exists in K_{opt} vs. SMR. We attribute this small discrepancy to the fact that the simulation model is based on distance-based service areas, thus ignoring the back-and-forth movement of the MN. Nevertheless, we see that the small discrepancy in K_{opt} vs. SMR does not impact the overall cost much as the analytical curve is almost identical to the simulation curve in Fig. 10.

Figure 12 summarizes the cost difference between HMIPv6 and DMAP-FR as a function of SMR for both simulation and analytical results. This validates the conclusion that DMAP-FR outperforms HMIPv6 over a wide range of SMR and that there indeed exists an optimal DMAP domain size K_{opt} under which the overall cost for mobility, service and failure recovery management is minimized. Here we note that the impact of mobility on protocol performance is reflected in the SMR parameter because SMR in effect determines if the packet delivery cost will dominate the mobility management cost or vice versa, depending on the relative largeness of the mobility rate compared with the service rate.

6 Conclusion

In this paper we have investigated DMAP-FR to provide efficient mobility and service management with failure recovery supports in MIPv6 environments. We devised a computational procedure to compute the optimal service area size that would minimize the overall network traffic cost, when given a set of parameters characterizing the MN's mobility and service characteristics and the network's failure behaviors. We compared our scheme with HMIPv6 scheme and concluded that our scheme outperforms HMIPv6 in terms of the network traffic cost, especially when the service to mobility ratio is high. The performance gain is in the amount of network traffic communication cost saved per time unit per user, so the cost saving due to a proper selection of the best service area dynamically will have significant impacts since the cumulative effect for all mobile users over a long time period would be significant.

By means of mobility management with failure recovery, our proposed DMAP-FR will always keep track of a MN's location, unless the MN is disconnected either voluntarily (e.g., power-down) or involuntarily (e.g., no radio coverage). Once the MN is reconnected to the network, DMAP-FR knows the location of the MN again. Thus, the overall cost incurred per time unit by a MN at time t depends on its mobility and service characteristics at time t. We suggest the computational procedure devised in the paper be applied periodically (by the MN and its DMAP) so as to determine the best K_{opt} dynamically. While a MN is disconnected, the overall cost incurred by the MN is zero (because the location update cost, the service delivery cost, and the failure recovery cost are all zero), so in this special case there is no optimal K_{opt} and the system will have to apply the computational procedure again to determine K_{opt} once the MN is reconnected to the system.

In the future, we plan to consider the implementation issue by building a testbed system based on [10] to further validate the analytical results. We plan to develop test scenarios with variations in the mobility and service characteristics of MNs in this testbed to compare DMAP-FR with MIPv6 and HMIPv6 with failure recovery supports for experimental validation.

References

1. Akyildiz, I. F., Lin, Y. B., Lai, W. R., & Chen, R. J. (2000). A new random walk model for pcs networks. *IEEE Journal on Selected Areas in Communications, 18*(7), 1254–1260.
2. Chen, I. R., He, W., & Gu, B. (2006). DMAP: A scalable and efficient integrated mobility and service management scheme for Mobile IPv6 systems. In *Proceedings of the 31st IEEE conference on local computer networks* (pp. 753–760). Florida U.S.A: Tampa.
3. Chen, I. R., & Verma, N. (2003). Simulation study of a class of autonomous host-centric mobility prediction algorithms for cellular and ad hoc networks. In *Proceedings of the 36th annual simulation symposium* (pp. 65–72). USA: Orlando.
4. Conta, A., Deering, S., & Gupta, M. (2006). Internet control message protocol (ICMPv6) for the internet protocol version 6 (IPv6). http://tools.ietf.org/html/rfc4443 IETF RFC 4443.
5. Dimopoulou, L., Leoleis, G., & Venieris, I. S. (2004). Introducing a hybrid fast and hierarchical MIPv6 scheme in a UMTS-IP converged architecture. In *Proceedings of the 29th annual IEEE international conference on local computer networks* (pp. 42–49). Florida U.S.A: Tampa.
6. Ghosh, R., & Varghese, G. (1998). *Fault-Tolerant Mobile IP*. Technical Report WUCS-98-11. St. Louis, USA: Washington University.
7. Johnson, D., Perkins, C., & Arkko, J. (2004). Mobility support in IPv6. http://www.ietf.org/rfc/rfc3775.txt IETF RFC 3775.
8. Koodli, R. (2009). Mobile IPv6 fast handovers. http://tools.ietf.org/html/rfc5568 IETF RFC 5568.
9. MacDougall, M. H. (1987). *Simulating computer systems: Techniques and tools*. Cambridge, MA: MIT Press.
10. Nelson, R., Daley, G., & Moore, N. (2002). Implementation of Hierarchical Mobile IPv6 for linux. In *Proceedings of the 6th IFIP international symposium on communications interworking* (pp. 289–300). Australia: Perth.
11. Omar, H., Saadawi, T., & Lee M. (1999). Supporting reduced location management overhead and fault tolerance in Mobile-IP systems. In *Proceedings of the IEEE international symposium on computers and communications* (pp. 347–353). Egypt: Red Sea.
12. Pack, S., Choi, Y., & Nam, M. (2006). Design and analysis of optimal multi-level hierarchical mobile IPv6 networks. *Wireless Personal Communications, 36*(2), 95–112.
13. Soliman, H. (2004). *Mobile IPv6: Mobility in a wireless internet*. Reading, MA: Addison-Wesley.
14. Soliman, H., Castelluccia, C., El-Malki, K., & Bellier, L. (2008). Hierarchical Mobile IPv6 mobility management. http://tools.ietf.org/html/rfc5380 IETF RFC 5380.
15. Trivedi, K. S., Ciardo, G., & Muppala, J. (1999). *SPNP version 6 user manual*. Durham, NC: Department of Electrical Engineering, Duke University.
16. Yang, X., Bouguettaya, A., Medjahed, B., Long, H., & He, W. (2003). Organizing and accessing web services on air. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans, 33*(6), 742–757.

17. Yeh, J. H., Chen, J. C., & Agrawal, P. (2009). Fast intra-network and cross-layer handover (FINCH) for wimax and mobile internet. *IEEE Transactions on Mobile Computing, 8*(4), 558–574.
18. You, T., Pack, S., & Choi, Y. (2003). Robust hierarchical mobile IPv6 (RH-MIPv6): an enhancement for survivability and fault-tolerance in Mobile IP systems. In *Proceedings of the IEEE 58th vehicular technology conference* (pp. 2014–2018). Florida, USA: Orlando.
19. Zhang, X., Castellanos, J., & Campbell, A. (2002). P–MIP: Paging extensions for mobile IP. *ACM Mobile Networks and Applications, 7*(2), 127–141.

Author Biographies

Weiping He received his B.E. 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 & Network Communications, and International Journal on Artificial Intelligence Tools. He is a member of the IEEE/CS and ACM.

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.