DMAP-FR: Integrated Mobility and Service Management with Failure Recovery Support for Mobile IPv6 Systems

Weiping He and Ing-Ray Chen

Department of Computer Science
Virginia Tech
{weiping, irchen}@vt.edu

Abstract

In this paper, we propose and analyze DMAP-FR, a mobility and service management scheme with failure recovery control in Mobile IPv6 systems. The basic idea behind DMAP-FR is to leverage access routers (ARs) as regional mobility anchor points (MAPs) as in 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 MAPs 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.

Key Words: *MIPv6, fault tolerance, failure recovery, mobility management, service management, performance analysis.*

1 Introduction

Mobile IPv6 (MIPv6) [3] 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. 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 and service management scheme with fault tolerance supports. This work extends from [1] to support fault tolerance. The core idea is extend the notion of mobility anchor points (MAPs) in Hierarchical Mobile IPv6 (HMIPv6) [5] 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. 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 which deals with mobility management only. Furthermore, DMAP-FR is scalable. The network cost reduction benefit as a result of applying DMAP-FR is cumulative and proportional to the number of MNs.

The use of MAPs proposed in HMIPv6 can indeed reduce the signaling cost for mobility management. However, failures of 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 MIPv6 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. The research challenge is to ensure service continuity for ongoing service applications despite failures.

2 Related Work

A MN can detect MAP failures in HMIPv6 [5] 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) if its RCoA has changed. However, this recovery process may suffer a long delay.

You, Pack and Choi [6] 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). However, there is a high signaling overhead for sending binding messages to the MAPs, HA, and CNs.

Omar, Saadawi and Lee [4] suggested that a child 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 [2] used passive replication techniques to mask failures of mobile agents in Mobile IPv4. 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 Failure Recovery Design

We make two assumptions in our paper: (1) for each AR, there is an overlapping coverage from other ARs since the failure of an AR will disconnect all MNs attached to it; (2) in the case that a router (not MAP) fails or a link goes down, it can be handled by the recovery mechanism of the routing protocol (e.g., OSPF).

As our DMAP-FR scheme is based on dynamically selecting an AR as the MAP of a MN, it can tolerate two kinds of failures: (1) the current AR can become the MN's DMAP if the DMAP fails; (2) with the enhancement of DMAP-FR, AR failure/recovery can be treated as disconnection/reconnection. The failure of DMAP can be detected by not receiving the announcement message by timeout.

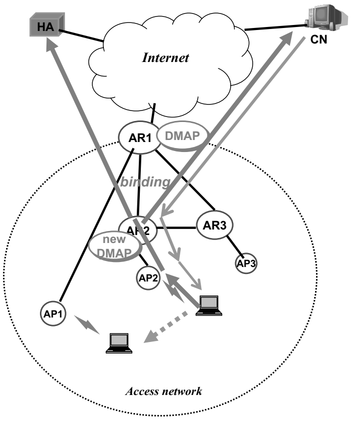

Figure 1: DMAP-FR Failure Recovery.

Figure 1 illustrates failure recovery in DMAP-FR. We consider three cases:

- Failure of the MN's DMAP which is not the current AR: As shown in Figure 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 Figure 1, 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) near by 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.
- Failure of the MN's current AR: As shown in Figure 1, the MN is under AR3 when AR3 fails and the DMAP is on AR1. In this case the DMAP locates another AR, i.e., AR1 or AR2, to replace

AR3. The MN will register with the new AR through a binding message.

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 MN service area can be modeled as consisting of K IP subnets. We aim to determine the optimal service area size in terms of K. 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.

The mobility and service characteristics of a MN and the failure profile of ARs are 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. 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. Loosely, 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. Loosely, 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. 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.

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 [7] assumes that the average number of hops between two communicating processes separated by K subnets is equal to $\sqrt{K}$.

Symbol	Meaning
λ	data packet rate
σ	mobility rate of MN
δ_f	failure rate of DMAP or AR
N	number of server engaged by the MN
K	number of subnets in one service area
τ	1-hop delay per packet in wired networks
α	average distance between HA and DMAP
β	average distance between CN and DMAP
γ	cost ratio for wireless vs. wired network

Table 1: Parameters for DMAP-FR.

4 Performance Analysis

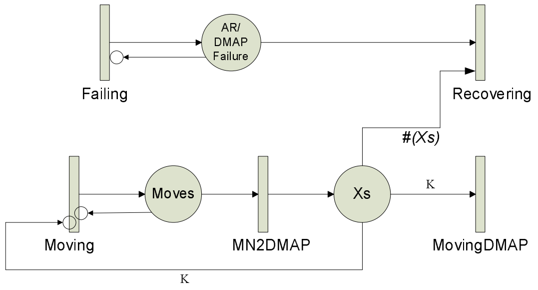

Figure 2: Petri Net Model.

We devise a computational procedure to be executed by the MN to determine the optimal service area size utilizing stochastic Petri net (SPN) techniques. 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 2 shows the performance model based on Stochastic Petri nets. 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 area. We allow it to accumulate to K, at which point we perform a service handoff. Below we explain how the SPN model is constructed to describe the behavior of a MN operating under DMAP-FR:

- 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 GFA) 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 `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 Figure 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 Figure 1 where the MN's current AR and DMAP is AR1 and AR1 fails. In this case, the MN will register with a new AR near by. 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 Figure 1 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` is calculated as $1/\gamma\tau + F(Mark(\mathrm{Xs})+1) \times \tau$ 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. $F(Mark(\texttt{Xs})+1)$ returns the number of hops between the current subnet and the DMAP separated by $Mark(\texttt{Xs})+1$ subnets. The argument of the $F(x)$ function is added by 1 to satisfy the initial condition that $Mark(\texttt{Xs}) = 0$ 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 MAP, the MN will register with a new AR near by. 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 $F(MarkXs())$ hops away from the failed DMAP, the cost can be parameterized as $\{N\left[\beta + F(Mark(\mathrm{Xs}))\right] + \left[\alpha + F(Mark(\mathrm{Xs}))\right] + \gamma\}\,\tau$. The rate to transition `Recovering` is the reciprocal of this quantity.

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,

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

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 i^{th} 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:

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

Let $C_{i,service}$ be the communication overhead for the network to service a data packet when the MN is in the i^{th} 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:

$$C_{service} = \sum_{i=0}^{K} (P_i \times C_{i,service}) \quad (2)$$

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,

$$C_{i,recovery} = \begin{cases} \gamma\tau + F(Mark(i)+1)\tau \\ \qquad \text{if the current AR fails while the DMAP is alive} \\ \gamma\tau + (\alpha\tau + F(Mark(i)+1)\tau) + \\ \quad N\beta\tau + F(Mark(i)+1)\tau) \qquad \text{if the DMAP fails} \end{cases} \quad (3)$$

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

$$C_{recovery} = \sum_{i=0}^{K} (P_i \times C_{i,recovery}) \quad (4)$$

Under the DMAP-FR scheme, a MN and its DMAP would apply the above equations to calculate C_{total} as a function of K and determine the optimal K representing the optimal *service area* size that will minimize the network traffic cost. Below we present numerical results. We vary the values of key parameters to analyze their effects on K_{opt}. The parameter values used are not reported here to save space.

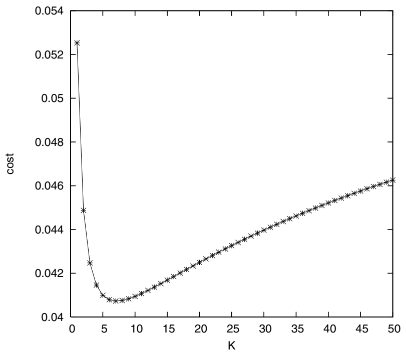

Figure 3: Cost vs. K.

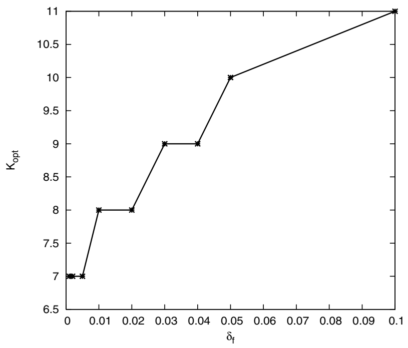

Figure 4: K_{opt} vs. δ_f.

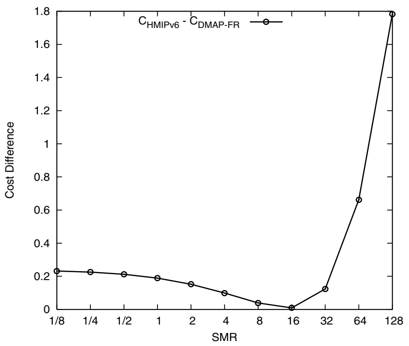

Figure 5: Cost Difference between HMIPv6 and DMAP-FR.

Figure 3 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 Figure 4 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 5 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 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 performs better than HMIPv6, especially when SMR is high.

5 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.

References

[1] I. R. Chen, W. He, and B. Gu. DMAP: A scalable and efficient integrated mobility and service management scheme for Mobile IPv6 systems. *Wireless Personal Communications*, 43(2):711–723, 2007.

[2] R. Ghosh and G. Varghese. Fault-tolerant Mobile IP. Technical Report WUCS-98-11, Washington University, 1998.

[3] D. Johnson, C. Perkins, and J. Arkko. *Mobility Support in IPv6.* http://www.ietf.org/rfc/rfc3775.txt, IETF, Work in Progress, 2004.

[4] H. Omar, T. Saadawi, and M. Lee. Supporting reduced location management overhead and fault tolerance in Mobile-IP systems. In *IEEE International Symposium on Computers and Communications*, pages 347 – 353, 1999.

[5] H. Soliman, C. Castelluccia, K. El-Malki, and L. Bellier. Hierarchical Mobile IPv6 mobility management. *http://www.ietf.org/rfc/rfc4140.txt, IETF, RFC*, 2005.

[6] T. You, S. Pack, and Y. Choi. Robust hierarchical Mobile IPv6 (RH-MIPv6): an enhancement for survivability and fault-tolerance in Mobile IP systems. In *IEEE 58th Vehicular Technology Conference*, pages 2014–2018, 2003.

[7] X. Zhang, J. Castellanos, and A. Campbell. P–MIP: Paging extensions for Mobile IP. *ACM Mobile Networks and Applications*, 7(2):127–141, March 2002.