Fast handoff scheme based on mobility prediction in public wireless LAN systems

S. Pack and Y. Choi

Abstract: Recently, wireless LAN systems have been widely deployed for public mobile Internet services. Public wireless LAN systems can provide high speed Internet connectivity using portable devices such as laptop computers, personal digital assistants (PDAs), etc. In public wireless LAN systems, reliable user authentication and mobility support are essential issues. However, re-authentication during handoff procedures causes long handoff latency and this affects the quality of service in real-time multimedia applications. A fast handoff scheme based on mobility prediction is proposed. In this scheme, a mobile host entering the area covered by an access point (AP) performs authentication procedures for multiple APs, rather than just the current AP. These multiple APs are selected by a prediction method called the frequent handoff region (FHR) selection algorithm, which takes into account users' mobility patterns, service classes, etc. Since a mobile host is registered and authenticated for an FHR in advance, handoff latency resulting from re-authentication can be significantly reduced. Simulation results show that the proposed scheme is more efficient than other schemes in terms of handoff delay and buffer requirements.

1 Introduction

Public wireless LAN systems based on IEEE 802.11 are becoming popular in hot spot areas. Unlike existing wireless Internet services, a public wireless LAN system can provide high-speed Internet connectivity of up to 11 Mbit/s. Originally, the wireless LAN was designed for indoor network solutions, so that deploying a wireless LAN in public areas requires a system of access control for unauthorised users. In addition, public wireless LAN systems should support different user mobility patterns. Therefore, user authentication and handoff support between access points (APs) are two of the most important issues to be considered in the design of public wireless LAN systems. Generally, since user authentication should be performed at each AP, when a mobile host (MH) moves into a new AP area, it should perform a new user authentication procedure and receive a new wired equivalent privacy (WEP) key, which encrypts the transmitted data in the wireless link.

These authentication mechanisms impact on the network performance. Since MHs need to be authenticated during and after handoff, the authentication mechanisms need to be responsive to the handoff time-scale required in micro-mobility environments [1]. Furthermore, since authentication, authorising, and accounting (AAA) servers are located at locations far away from the AP, the current handoff scheme cannot meet certain requirements [2] for real-time multimedia applications.

IEE Proceedings online no. 20040834

doi:10.1049/ip-com:20040834

Paper first received 29th July 2003 and in revised form 18th May 2004. Originally published online: 15th September 2004

The authors are with the School of Computer Science & Engineering, Seoul National University Seoul, Korea

To reduce the handoff latency, several schemes have been proposed in the literature [3, 4]. However, these schemes focused on the reduction of the latency incurred in scanning, channel detection, and execution procedures and they did not consider any authentication latency, which is one of the inevitable latencies in public WLAN systems. In order to reduce the authentication delay, a pre-authentication scheme was proposed [5]. In the current WLAN standard, although MHs must authenticate with an AP before associating with it, no schemes in IEEE 802.11 require that authentication takes place immediately before association. On the other hand, in the pre-authentication scheme, MHs can authenticate with several APs during the scanning process so that when association is required, the MH is already authenticated. As a result of pre-authentication, MHs can reassociate with APs immediately upon moving into their coverage area, rather than having to wait for the authentication exchange. Although the pre-authentication scheme can reduce the authentication handoff delay, it requires a sufficient overlapping area and the modification of MHs. These points can be drawback in the deployment of the pre-authentication scheme.

In this paper, we propose a fast handoff scheme that reduces the re-authentication latency in public wireless LANs. When a MH sends an authentication request, the AAA server authenticates not only the currently used AP, but also multiple APs, and sends multiple WEP keys to the MH. These multiple APs are selected by the frequent handoff region (FHR) selection algorithm based on mobility prediction. The FHR selection algorithm utilises the handoff weight in each link between APs, which is collected and calculated by the centralised system.

2 Fast handoff scheme in public WLAN systems

The objective of the proposed handoff scheme is to reduce the handoff latency caused by authentication procedures at

the new AP. In the proposed scheme, a MH performs authentication procedures not only for the current AP but also for neighbouring APs, when initial registration is performed. The key issue in this scheme is how to select the neighbouring APs to be authenticated in advance. The simplest method is to select all of the APs adjacent to the current AP. However, this is inefficient because it does not consider any movement patterns or the AP's geographical location. Since a wireless LAN is a solution usually employed within an administrative network domain, centralised AP configuration and management are possible. Therefore, we present a centralised neighbour selection algorithm, called the 'FHR selection algorithm'.

2.1 Design principles

The FHR is a subset of adjacent APs, which are likely to move to in the near future. Many handoff prediction algorithms, which are based on the mobility history collected from MHs with regularity, have been proposed [6]. Namely, since most MHs move according to a specific regularity, it is possible to predict the mobility pattern using the previous patterns. However, most MHs in public WLAN systems are visiting or temporary hosts without any regular mobility patterns. Therefore, it is not a feasible solution to predict the mobility pattern from the mobile-specific information. Instead of the mobile-specific prediction scheme, we utilised a network-specific prediction scheme. In our prediction scheme, the handoff probability between APs is estimated using the previous handoff ratio and residence time collected and calculated by the central system.

2.2 FHR selection algorithm

In the proposed scheme, all handoff events are recorded in the event log database. Table 1 shows an example of the event log database. Each handoff event consists of five fields: sequence field, previous AP id, next AP id, in-time, and out-time. For example, the first event shows that a MH is associated with AP2 at 07:54:57 and that it moves to the area of AP4 at 08:14:25. On the other hand, the fourth event indicates that a MH is associated with AP3 and that logout (or power off) occurs at AP3 without any handoff. This event does not have to be counted in the handoff weight decision.

Using these event logs, it is possible to find out the handoff ratio per unit time between APs. The handoff ratio is calculated as in (1)

$$H(i, j) = \sum_{k=1}^{N(i,j)} \frac{1}{R_k(i, j)} \tag{1}$$

where $H(i, j)$ and $N(i, j)$ denote the unit handoff ratio and the number of handoff events from AP(i) to AP(j), respectively. Let $R_k(i, j)$ be the residence time in the kth handoff event from AP(i) to AP(j). The residence time is calculated as follows

$$R_k(i, j) = T_{out}(k) - T_{in}(k) \tag{2}$$

Table 1: Example event log database

Seq	Previous AP id	Next AP id	In-time	Out-time
1	2	4	07:54:57	08:14:25
2	1	2	08:00:55	08:05:18
3	2	5	08:04:23	08:14:03
4	3	0	08:11:02	08:41:31

where $T_{out}(k)$ and $T_{in}(k)$ are in-time and out-time of kth handoff event, respectively. For the FHR selection algorithm, we should draw a weighted bi-directional graph on AP placements. The weight values between APs are determined by the handoff ratio. Equation (3) shows the weight function between AP(i) and AP(j). $w(i, j)$ denotes the link weight value.

$$w(i, j) = \begin{cases} 0 & (i = j) \\ \dfrac{1}{H(i,j)} & (i \neq j,\ AP(i) \text{ and } AP(j) \text{ are adjacent}) \\ \infty & (AP(i) \text{ and } AP(j) \text{ are not adjacent}) \end{cases} \tag{3}$$

As shown in (3), the weight value is reversely proportional to the handoff ratio. The handoff ratio in (1) is influenced only by the handoff events, so that the weight value in the path from AP(i) to AP(j) is set as zero. In addition, if two APs are not adjacent, the weight value is infinite. Owing to traffic asymmetry, $w(i, j)$ and $w(j, i)$ are not equal. To select the FHR, the user's service level also should be considered. To consider the user's service level in FHR selection, we define a weight bound value according to the users' service class. According to the weight bound value, the number of selected APs for each user is limited.

Algorithm 1: Select FHR (W, N, D, M, i)

```
1: initiate W, N, D, M, i;
2: i←0;
3: while i≤N do
4:   M(i) = False;
5:   i++;
6: end while
7: while j≤N do
8:   if w(i, j)<D then
9:     M(i) = True;
10:    while k≤N do
11:      if w(j, k) ≤D−w(i, j) then
12:        M(k) = True;
13:      end if
14:      k + +;
15:    end while
16:  end if
17:  i+ + ;
18: end while
```

Next, the detailed selection algorithm needs to be considered. Using (1) and (3), we can obtain an N by N weight matrix, $\boldsymbol{W}$. Algorithm 1 shows a pseudo-program for the FHR selection algorithm. N denotes the number of APs. $\boldsymbol{W}$ represents the weighted bi-directional graph of AP placements. $\boldsymbol{M}$ is the 1 by N matrix for the result of the FHR selection algorithm. D and H are the weight bound value and maximum hop count, respectively. If D has a high value, the user will authenticate a greater number of neighbouring APs in order to obtain a more seamless service. The value H determines the handoff scope of a MH. In most cases, handoffs occur between two adjacent APs. However, if a MH moves with very high speed, we should consider the handoff possibilities between two non-adjacent APs. For example, if a MH moves from AP(i) to AP(k) via

AP(j) and the residence time at AP(j) is very short, the MH's connection should be maintained at AP(k) as well as AP(i). In other words, fast handoff to AP(k) should be supported. H provides for the possibility of this multi-hop handoff. However, since the multi-hop handoff is a rare case, H is set to 2 in the FHR selection procedure shown in algorithm 1. Each element $M(i)$ of the FHR matrix is initialised as '*false*'. If the sum of the weights of AP(j) is less than or equal to the weight bound value then D, then $M(i)$ is set as '*true*' (namely, AP(j) is selected as the FHR.) This nested loop is repeated according to the maximum hop count, H. Figure 1 shows an example of the FHR selection procedure. The marked point in Fig. 1 means the current AP where a MH is located.

2.3 Key distribution mechanism

After selecting the FHR, a MH performs the authentication procedures. Firstly, the MH sends an 'access request' message to the AAA server. Specific message formats are dependent on the AAA protocols being used [7], such as RADIUS [8], DIAMETER [9] etc. The AAA server receiving an access request message performs an authentication procedure on the MH and sends an 'access reply' message to the corresponding MH. The 'access reply' message contains several pieces of information including the session key, WEP key, and so on.

In our scheme, the key information should be distributed to all APs located in the FHR. Since the current standard supports only one-to-one key distribution, a modified key distribution mechanism is required. Figure 2 shows the initial authentication procedure in the proposed scheme. Although a MH sends only one 'access request' message through the current AP, the AAA server sends multiple 'access reply' messages to all APs belonging to the FHR. After receiving 'access reply' messages, the neighbouring APs, with the exception of the current AP, maintain this authentication information in the soft state. Namely, if there are no handoff events within a specific time period $T(i, j)$), the key information is deleted. Unlike other authentication responses, the response relayed to the MH through to the current AP should contain a variety of information such as the session key, multiple WEP keys, etc. Figure 2 also shows the case of re-authentication after handoff. When a MH hands off to any other AP, since the new AP receives session information in advance, further message exchanges are not needed. Generally, since the AAA server is often located in a remote domain for more scalable service, the delay in the path from the AP to the AAA server is a critical factor in the total handoff latency.

To determine the length of the soft state time period, the handoff probability should be calculated in advance. The handoff probability can be obtained from the handoff ratio presented in (1). In (4), $P_h(i \rightarrow a(l))$ denotes the handoff probability from AP(i) to AP($a(l)$). $a(1), \ldots a(2), \ldots a(m)$ is a set of neighbouring APs to AP(i) and a(l) is an AP

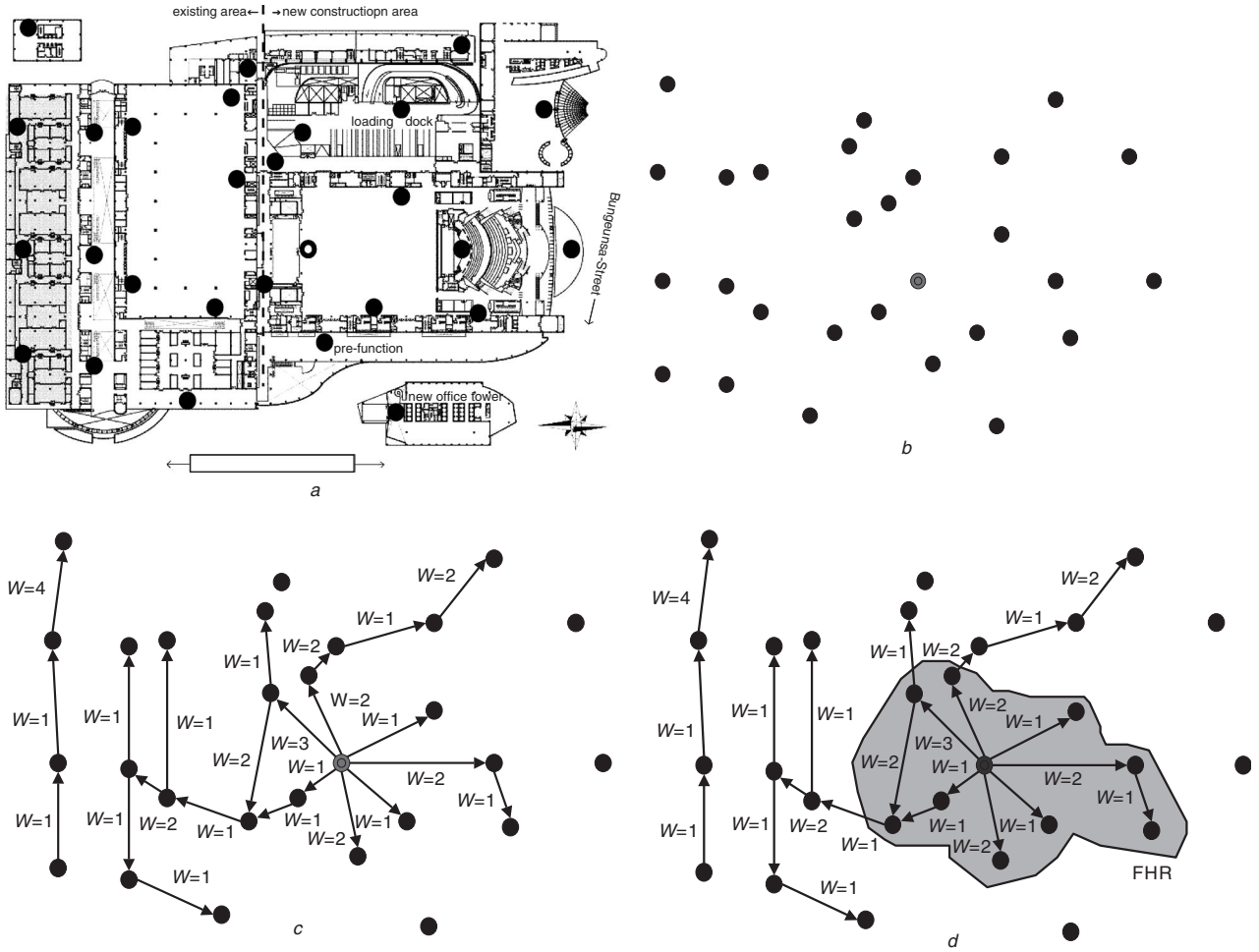

Fig. 1 *FHR selection procedure*
a Initial AP placement
b AP placement
c Weighted directed graph
d FHR selection

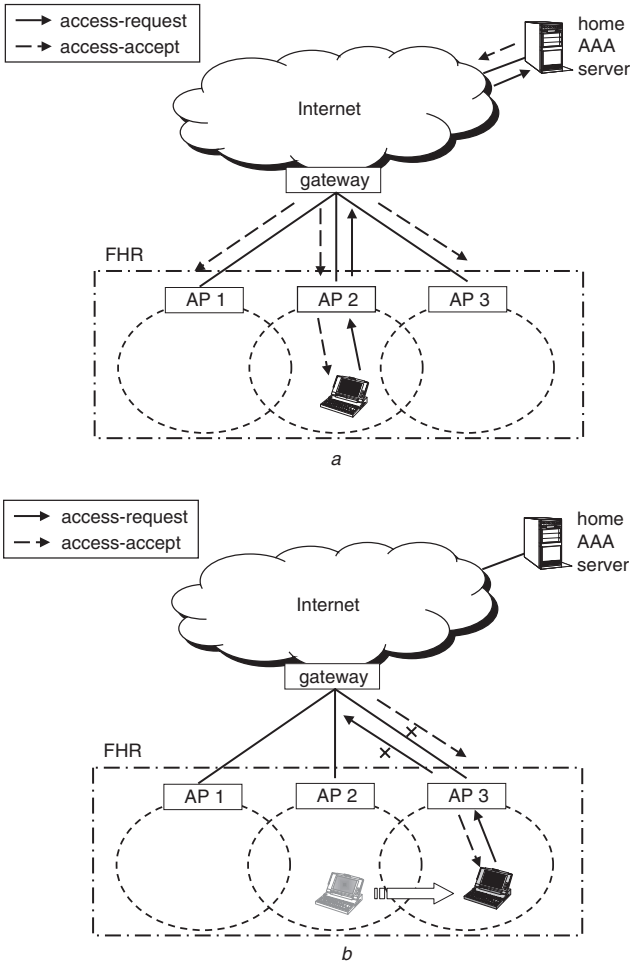

Fig. 2 *Key distribution*
a Initial authentication before handoff
b Re-authentication after handoff

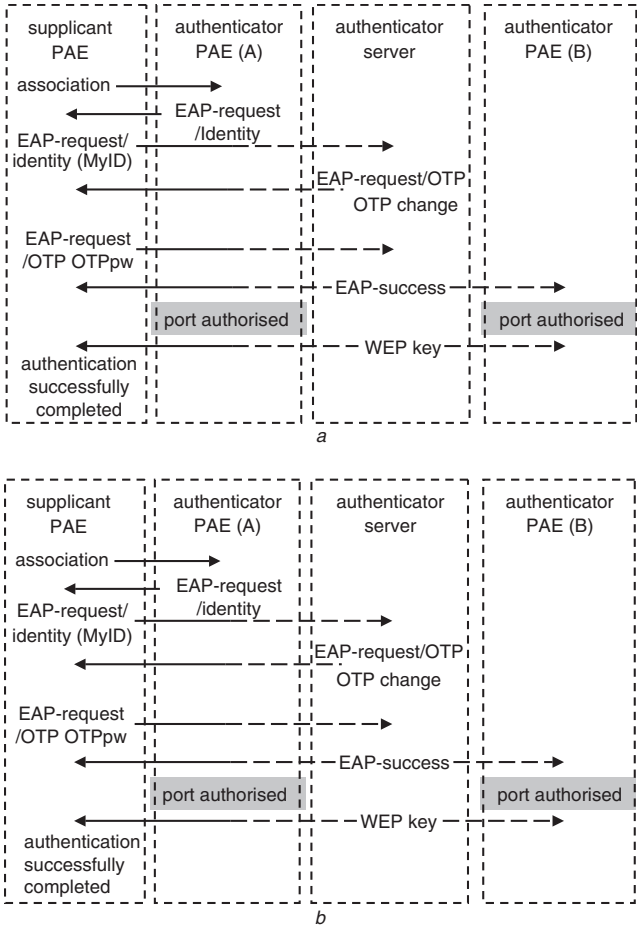

Fig. 3 *Authentication procedure in IEEE 802.1x model*
a Message flow before handoff
b Message flow after handoff

belonging to the set.

$$P_h(i \rightarrow a(l)) = \frac{H(i, a(l))}{\sum_{j=1}^{m} H(i, a(j))} \quad (4)$$

With the handoff probability, the average residence time, $E(R(i, j))$, is calculated as follows:

$$E(R(i, j)) = \frac{\sum_{k=1}^{N(i,j)} R_k(i, j)}{N(i, j)} \quad (5)$$

Using the handoff probability and the average residence time, the soft state timer is obtained as (6) in our scheme. In (6), η is a scaling factor to adjust the timer value.

$$T(i, j) = \eta E(R(i, j)) \times P_h(i \rightarrow j) \quad (6)$$

To support the proposed scheme, each AP should maintain session information in the soft state. To this end, IEEE 802.1x [10] can be utilised. IEEE 802.1x enables authenticated access to IEEE 802 media. IEEE 802.1x provides a network port access control scheme. Figure 3 shows the basic components and the port-based access control mechanism in the 802.1x model. The 'Supplicant' system is an entity at one end of a point-to-point LAN segment that is being authenticated by an 'Authenticator' attached to the other end of that link. The 'Authentication server' system is an entity that provides an authentication service to an authenticator. This service determines, from the credentials provided by the supplicant, whether the supplicant is authorised to access the services provided by the authenticator. Port access entity (PAE) is the protocol entity associated with a port. A given PAE can support the protocol functionality associated with the authenticator, the supplicant, or both. In IEEE 802.1x, there are two types of network ports: the controlled port and the uncontrolled port. The uncontrolled port is used for transmission of the 'access request' and 'access reply' messages. On the other hand, the controlled port is used for data transmission. A MH can obtain access to the controlled port only after performing the user authentication and receiving session and WEP keys. In our scheme, since an authenticated MH possesses multiple session and WEP keys, corresponding to the different APs included in the FHR, and each AP has already received the result of authentication from the AAA server, the MH can obtain access to the controlled port for data transmission without further re-authentication.

3 Performance evaluation

3.1 Simulation environment

To evaluate the performance of the proposed fast handoff scheme, we developed an event-driven simulator using C programming language. Figure 4 shows the simulation environment and the weight values for each link. In this

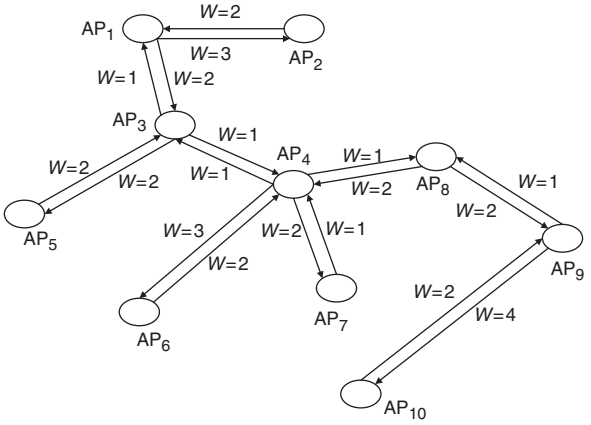

Fig. 4 *Simulation environment*

environment, AP(4) is the current AP receiving an authentication request from a MH. Equation (7) represents the weighted graph in the form of a matrix used in a FHR Selection procedure. The matrices shown in (8) present the results of the FHR Selection procedure in the cases where $D=1$, $D=2$, and, $D=3$, respectively (1 'true', 0 'false'). In terms of scaling factor (η) adjusting the timer value, 1, 2, and 3 are used for class 1, class 2, and class 3 MHs, respectively.

$$\boldsymbol{W} = \begin{pmatrix} 0 & 3 & 2 & \infty & \infty & \infty & \infty & \infty & \infty & \infty \\ 2 & 0 & \infty & \infty & \infty & \infty & \infty & \infty & \infty & \infty \\ 1 & \infty & 0 & 1 & 2 & \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & 1 & 0 & \infty & 3 & 2 & 1 & \infty & \infty \\ \infty & \infty & 2 & \infty & 0 & \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & 2 & \infty & 0 & \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & 1 & \infty & \infty & 0 & \infty & \infty & \infty \\ \infty & \infty & \infty & 2 & \infty & \infty & \infty & 0 & 2 & \infty \\ \infty & \infty & \infty & \infty & \infty & \infty & \infty & 2 & 0 & 4 \\ \infty & \infty & \infty & \infty & \infty & \infty & \infty & \infty & 2 & 0 \end{pmatrix} \quad (7)$$

$$M(1) = [0011000100] \quad M(2) = [1011001100]$$
$$M(3) = [1011111110] \quad (8)$$

3.2 Mobility model

Generally, the random walk model is widely used as the micro-mobility model [11]. In the random walk model, a user moves in one direction in a random manner. However, since we assumed that the mobility weights for each path could be determined by a centralised method, we used the independent and identically distributed (i.i.d.) mobility model. In this model, time is divided into multiple intervals and a MH can make at most one move during one interval. The time interval length (i.e. cell residence time) follows a Gamma distribution shown in (9)

$$f_R(t) = \frac{b^k t^{k-1}}{\Gamma(k)} e^{-bt} \quad (9)$$

where b is equal to $k\lambda_m$ and $\Gamma(k)$ is the Gamma function, which is defined as $\int_0^\infty t^{k-1}e^{-t}dt$. The mean and variance of the Gamma distribution are $1/\lambda_m$ and $1/k\lambda_m^2$, respectively. In simulations, the mean residence time and variance are 100 s and 1000 s, respectively.

If a MH is in cell i at the beginning of a time slot, then during this slot the MH moves to cell $i+1$ with probability p, moves to cell $i-1$ with probability q, or remains in cell i with probability $1-p-q$, independent of its movements in other time slots. Each transition probability, p and q, can be found based on (4). However, we did not assign the weight value in the case that a MH stays at the same AP in the next time slot, so that we used the stability factor, α. If $\alpha=0$, the MH hands off to another AP with probability 1. On the other hand, if $\alpha=1$, the MH stays at the current AP with probability 1. $P(i, j)$ is the transition probability from AP(i) to AP(j) and G is the normalisation constant.

$$P(i,j) = \begin{cases} \frac{1}{G} \times \frac{1}{w(i,j)} & (i \neq j) \\ \frac{1}{G} \times \alpha & (i = j) \end{cases} \quad (10)$$

$$G = \sum_{i \neq j} \frac{1}{w(i,j)} + \alpha$$

3.3 Simulation result

In this Section, we compare the handoff latency in both the proposed fast handoff scheme and the general handoff scheme.

Figure 5 shows the average handoff latency when the AAA server is located in the local domain. Generally, since the latency of a local AAA server is not high, the average latency is also not high. However, we find that the average latency is lower when the proposed scheme is used. Also, since user 3, belonging to class 3, has a larger weight bound value, his latency is lower than that of a user belonging to another class, especially when user mobility is high. Figure 6 shows the simulation result in the case of a remote AAA server. The overall pattern is very similar to that of Fig. 5. However, the average handoff latency is more than 40 ms when the fast handoff scheme is not used. Since the

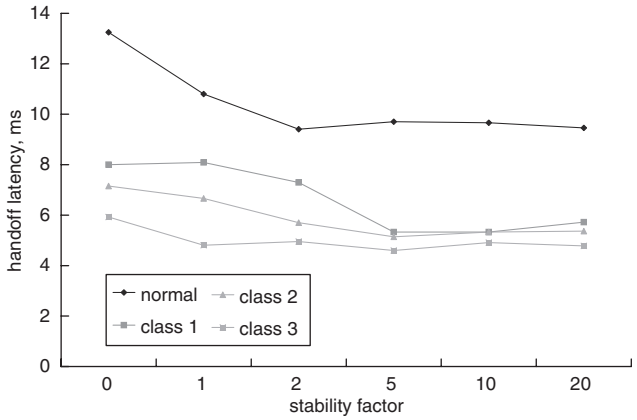

Fig. 5 *Handoff latency (local AAA server)*

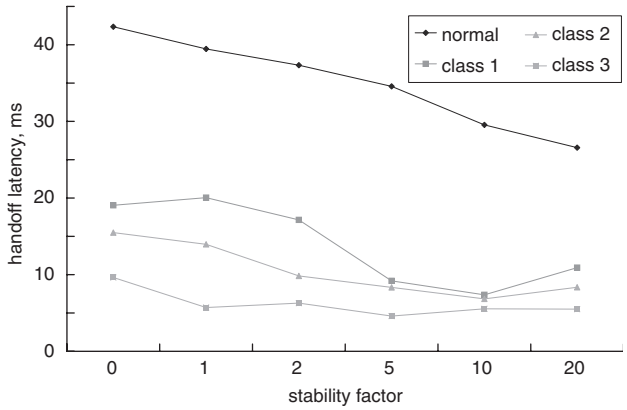

Fig. 6 *Handoff latency (remote AAA server)*

end-to-end latency bound in multimedia applications is generally about 150 ms, this value is not appropriate for multimedia applications. However, the latency obtained in the case of the fast handoff scheme is less than 20 ms, which is more appropriate for multimedia applications and other situations requiring reduced latency times.

Figure 7 shows the effect of the average number of movements. In Fig. 7, the ratio is defined as the number of handoffs towards the selected FHR to the total number of handoffs. When the average number of movements is small (e.g. 2), the ratio is about 0.75–0.88 whereas the ratio is 0.44–0.51 when the average number of movements is large (e.g. 6). This is because the probability, that a MH moves to APs not in the selected FHR, increases as the average number of movements increases. The ratio can be increased by using large η values or weight bound values (D). The previous measurement studies [12, 13] indicate that the number of APs associated with a MH during roaming is only two or three. Therefore, η and D should be carefully determined based on the previous measurement results.

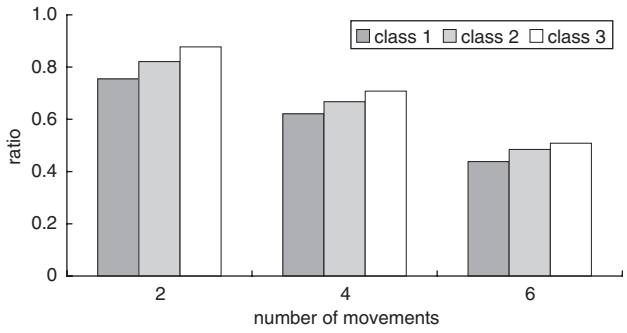

Fig. 7 *Effect of the number of movements*

In mobility management, buffering at the previous AP is necessary for smooth handoff. The buffer size is affected by handoff latency. To determine the required buffer size at the previous AP, we used four traffic sources. Each traffic source has a constant bit rate and packet size. Table 2 shows the parameters for each source. Figure 8 shows the necessary buffer size of each source for the cases where fast handoff is used or not. According to Fig. 8, the buffer size is substantially reduced when the proposed fast handoff scheme is used. Of course, in the case where fast handoff is not used, the buffer size here is less than 25 kbytes. However, this size is only for the case of one MH where each traffic source has a bit rate less than 2 Mbit/s.

Table 2: Four traffic sources

Source	Bit rate, kbit/s	Packet size, bytes
1	64	64
2	384	128
3	1024	512
4	2048	512

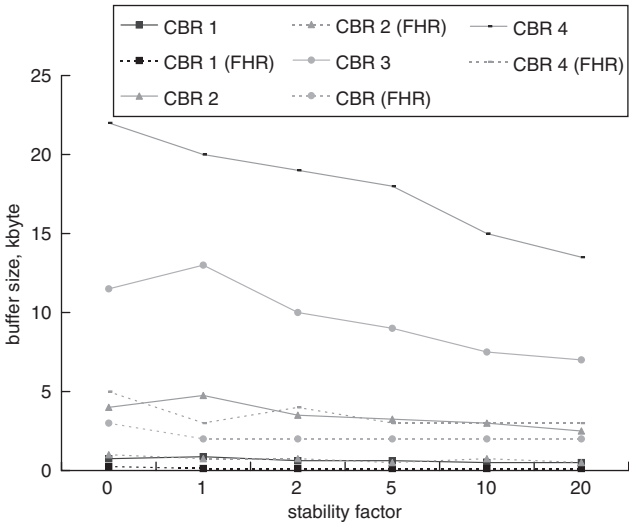

Fig. 8 *Buffer requirement at the previous AP*

3.4 Overhead analysis

In our scheme, since an 'access reply' message that a MH receives through the authentication procedures contains multiple WEP keys, the length of the message is longer than that of the message used in the general scheme. The length of the message is proportional to the number of pre-authenticated APs calculated according to the FHR selection algorithm. However, no other information besides the WEP key needs to be duplicated, so the overhead induced by the increased message length is not a critical factor. Furthermore, the number of selected APs can be adjusted according to the users' service class, and the number of selected APs is usually quite small due to the users' locality in a local network environment. Generally, the resources available in a wired link are much richer than those of the wireless link. Therefore, it is important to reduce the amount of resources being wasted in the wireless link. Although multiple APs are pre-authenticated in our scheme, there is only one message delivery in the wireless link between the MH and the current AP. Therefore, no additional resources are used in the wireless link.

4 Conclusions

Since handoff and authentication procedures are essential in public wireless LAN systems, we have focused our attention on minimising authentication latency during handoff. In our scheme, multiple APs, selected by the SelectFHR algorithm, are authenticated in advance. The SelectFHR algorithm utilises a mobility prediction method based on traffic patterns and user characteristics, which are collected and managed by the centralised system. Simulation results show that the total handoff latency of the proposed handoff scheme is much less than that of the general handoff

scheme. Furthermore, as regards smooth handoff, when the proposed handoff scheme is used, less buffering is required at the previous AP. Therefore, the proposed fast handoff scheme enables public wireless LAN systems to support a variety of multimedia applications.

5 Acknowledgments

This work was supported in part by the Brain Korea 21 project of the Ministry of Education and in part by the National Research Laboratory project of the Ministry of Science and Technology, 2003, Korea.

6 References

1 Campbell, A., and Gomez, J.: 'IP micro-mobility protocols', *ACM Mob. Comput. Commun. Rev.*, 2000, **4**, (4)
2 International Telecommunication Union 'General characteristics of international telephone connections and international telephone circuits' ITU-TG.114, 1998
3 Mishra, A., Shin, M., and Arbaugh, W.: 'An empirical analysis of the IEEE 802.11 MAC layer handoff process', *Comput. Commun. Rev.*, 2003, **33**, (2)
4 Velayos, H., and Karlsson, G.: 'Techniques to reduce IEEE 802.11b MAC layer handover time'. KTH technical report TRITA-IMIT-LCN R 03:02, 2003
5 Aboba, B.: 'IEEE 802.1X pre-authentication preauthentication' IEEE 802.11 TGi draft, June 2002
6 Soh, W., and Kim, H.: 'QoS provisioning in cellular networks based on mobility prediction techniques', *IEEE Commun. Mag.*, 2003, **41**, pp. 86–92
7 Mitton, D., Johns, M., Barkley, S., Nelson, D., Patil, B., Stevens, M., Wolff, B.: 'Authentication, authorization, and accounting: protocol evaluation' IETF RFC 3127, June 2001
8 Rigney, C., Willens, S., Rubens, A., Simpson, W.: 'Remote authentication dial in user service (RADIUS)' IETF RFC 2865, June 2000
9 Calhoun, P., Loughney, J., Guttman, E., Zorn, G., Arkko, J.: 'Diameter base protocol' IETF RFC 3588, September 2003
10 'IEEE standards for local and metropolitan area networks: port based network access control' IEEE Std 802.1x-2001, June 2001
11 Bettstetter, C.: 'Smooth is better than sharp: a random mobility model for simulation of wireless networks'. Proc. ACM MSWiM 2001, Rome, Italy, July 2001
12 Balachandran, A., Voelker, G., Bahl, P., and Rangan, P.: 'Characterizing user behaviour and network performance in a public wireless LAN'. Proc. ACM SIGMETRIC, Los Angeles, CA, USA, June 2002, pp. 195–205
13 Schwab, D., and Bunt, R.: 'Characterising the use of a campus wireless network'. Proc. IEEE INFOCOM, Hong Kong, March 2004