Behavior-Based Mobility Prediction for Seamless Handoffs in Mobile Wireless Networks

Weetit Wanalertlak  
*Oregon State University, wanalewe@eecs.oregonstate.edu*

Ben Lee  
*Oregon State University, benl@eecs.oregonstate.edu*

Chansu Yu  
*Cleveland State University, c.yu91@csuohio.edu*

Myungchul Kim  
*Information and Communications University, mckim@icu.ac.kr*

Seung-Min Park  
*Electronic and Telecommunications Research Institute, minpar@etri.re.kr*

See next page for additional authors

Original Citation

Repository Citation
Wanalertlak, Weetit; Lee, Ben; Yu, Chansu; Kim, Myungchul; Park, Seung-Min; and Kim, Won-Tae, "Behavior-Based Mobility Prediction for Seamless Handoffs in Mobile Wireless Networks" (2011). Electrical Engineering & Computer Science Faculty Publications. 76.  
https://engagedscholarship.csuohio.edu/enece_facpub/76

This Article is brought to you for free and open access by the Electrical Engineering & Computer Science Department at EngagedScholarship@CSU. It has been accepted for inclusion in Electrical Engineering & Computer Science Faculty Publications by an authorized administrator of EngagedScholarship@CSU. For more information, please contact library.es@csuohio.edu.
Authors
Weetit Wanalertlak, Ben Lee, Chansu Yu, Myungchul Kim, Seung-Min Park, and Won-Tae Kim

This article is available at EngagedScholarship@CSU: https://engagedscholarship.csuohio.edu/enece_facpub/76
Behavior-based mobility prediction for seamless handoffs in mobile wireless networks

Weetit Wanalertlak · Ben Lee · Chansu Yu · Myungchul Kim · Seung-Min Park · Won-Tae Kim

Abstract The field of wireless networking has received unprecedented attention from the research community during the last decade due to its great potential to create new horizons for communicating beyond the Internet. Wireless LANs (WLANs) based on the IEEE 802.11 standard have become prevalent in public as well as residential areas, and their importance as an enabling technology will continue to grow for future pervasive computing applications. However, as their scale and complexity continue to grow, reducing handoff latency is particularly important. This paper presents the Behavior-based Mobility Prediction scheme to eliminate the scanning overhead incurred in IEEE 802.11 networks. This is achieved by considering not only location information but also group, time-of-day, and duration characteristics of mobile users. This captures short-term and periodic behavior of mobile users to provide accurate next-cell predictions. Our simulation study of a campus network and a municipal wireless network shows that the proposed method improves the next-cell prediction accuracy by $23\% - 43\%$ compared to location-only based schemes and reduces the average handoff delay down to $24\% - 25\%$ ms.

Keywords Mobility prediction · Fast handoffs · WLANs · WMNs

1 Introduction

Wireless LANs (WLANs) based on the IEEE 802.11 standard have become pervasive in our society. WLANs offer high data transfer rate and are low-cost and easily deployable. These characteristics allow mobile users with portable devices to not only connect to the Internet but also use various services such as Voice over IP (VoIP) and real-time multimedia data transmission, e.g., streaming audio and video. Until recently, deployment and coverage area of WLANs have been limited to variety of isolated hotspots, such as coffee shops, buildings, airport terminals, etc. But now Municipal Wireless Networks [1–4] based on Wireless Mesh Networks (WMNs) [5] allow the coverage area to be extended across a larger geographical area, e.g., a city. Moreover, WLANs will play a crucial role as backbone networks for emerging pervasive computing technologies, such as wearable computers, Wireless Body Area Networks (WBANs) [6], Wireless Personal Area Networks (WPANs), and Vehicular Ad hoc Networks (VANETs) [7]. Nevertheless, achieving seamless, mobile access is a major
challenge because of large handoff delay incurred scanning for available access points (APs) when a mobile station (MS) switches connection from one AP, or cell, to another. Recent studies found that passively scanning for APs during a handoff can be as much as a second [8] and actively scanning for APs requires $350 \sim 500 \text{ ms}$ [8]. This becomes a major concern for mobile multimedia applications such as VoIP where the end-to-end delay is recommended to be not greater than 50 ms [9].

The key to reducing the scanning overhead and thus the handoff delay is to predict the next point-of-attachment of a MS, i.e., mobility prediction. This way a MS can directly authenticate/re-associate with the target AP thereby eliminating the scanning overhead. Numerous efforts have already been made to reduce the scanning overhead in WLANs. These include methods to employ extra hardware, either in the form of additional radios [10] or an overlay sensor network [11], to detect APs, selectively scan channels based on the topological placement of APs [12], and predict the next point-of-attachment based on signal strength [13].

Unfortunately, these techniques neither provide next cell predictions that can eliminate the need to scan for APs nor consider mobility patterns of MSs, which are dictated by the structure of a building or a city block and the past behaviors of MSs. In our previous work [14], we proposed a location-based technique that considers the long-term history of handoff patterns. Although this technique is effective in performing next cell predictions for majority of MSs, it does not consider short-term and periodic behavior of mobile users that cause some MSs to behave differently from the norm and thus suffer mispredictions.

This paper presents a solution called the Behavior-based Mobility Prediction (BMP) technique that provides accurate mobility prediction by considering multifaceted user behavior based the following four factors: location, group, time-of-day, and duration. The location-based mobility prediction is achieved by maintaining the handoff history of all the MSs in the network, and then monitoring direction of movements of MSs relative to the topological placement of cells to predict their next point-of-attachment. In addition, next cell predictions are based on the frequencies of occurrences rather than signal strength. Therefore, it takes into consideration that mobility patterns are dictated by the structure of a building or a city block and the past behaviors of MSs. Moreover, the handoff frequencies are treated as time-series data, thus when next cell predictions fail future predictions are recalibrated based on different groups, where each group of MSs has similar mobility patterns, time-of-day, and duration characteristics.

The paper is organized as follows: Section 2 overviews the 802.11 scanning process. Section 3 presents the proposed BMP method and Sect. 4 evaluates its performance. Section 5 discussed the related work. Finally, Sect. 6 concludes the paper and discusses future work.

2 Background: 802.11 scanning process

In a WLAN, when a MS moves from one cell to another, its network interface senses the degradation of signal quality in the current channel. The signal quality continues to degrade as MS moves further away from the current AP, and a handoff to a new cell is initiated when the signal quality reaches a preset threshold [15]. This process starts with probing for new cells using either passive or active scanning. In passive scanning, a MS switches its transceiver to a new channel and waits for a beacon to be sent by the new AP, typically every 100 ms, or until the waiting time reaches a predefined maximum duration, which is longer than the beacon interval. The time a MS has to wait varies since beacons sent by APs are not synchronized. For these reasons, a recent study has shown that an MS can spend up to 1 s to search all possible channels [8], which results in unacceptable handoff delay.

In active scanning shown in Fig. 1, a MS broadcasts a probe request and waits for a response. If the MS receives a response from an AP, it assumes there may be other APs in the channel and waits for MaxChannelTime. Otherwise, the MS only waits for MinChannelTime. MinChannelTime is shorter than MaxChannelTime to keep the overall handoff delay low, but it should be long enough for MS to receive a possible response. A typical duration for scanning each channel is around 25 ms and $350 \sim 500 \text{ ms}$ for all 11 channels [8].

![Fig. 1 The 802.11 active scanning process](image-url)
After scanning, MS joins the cell with the strongest signal strength, which is done by performing authentication and (re)association. Authentication is the process that a MS uses to announce its identity to a new AP. In the IEEE 802.11 standard, authentication is performed using either open system or shared key. Open system authentication is the default method, and involves the MS sending authentication request frame to the AP containing its source address in the frame header and the type of authentication in the frame body. Then, the AP sends the authentication response frame back to the MS. This frame has the authentication result and the information that indicates the type of authentication.

The next step is (re)association, which allows the distribution system to keep track of the location of each MS so that frames destined for a MS can be forwarded to the correct AP. How (re)association requests are processed is implementation specific, but typically involves allocation of frame buffers and, in the case of re-association, communicating with the old AP so that any frames buffered at the old AP are transferred to the new AP. Finally, the last step involves rerouting the network traffic by updating the Forwarding Table in the switch that connects both the old and new APs.

3 The behavior-based mobility prediction scheme

The basic idea behind BMP is to improve the prediction accuracy by exploiting the behavior of mobile users. This is achieved by identifying temporal segments of handoff history that characterize certain user behaviors, and performing handoff predictions that pertain to those periods.

3.1 User behavior

Behavior of mobile users can be characterized in many different ways. In the proposed method, the four characteristics that define user behavior are location, group, time-of-day, and duration. The following discusses the motivation for using these characteristics.

The location factor, as discussed in Sect. 1, represents the history of mobility patterns that can be either static or dynamic. Static mobility patterns are dictated by fixed structures, such as roads, building structures, and city blocks. On the other hand, dynamic mobility patterns are caused by user behavior as well as frequent and drastic changes in the operating environment of wireless networks due to multipath effects and electromagnetic interference.

The group factor reflects the fact that MSs often behave as groups. For example, MSs in an academic setting can be categorized as students, graduate students, and faculty/staff, and the mobility behaviors of these groups are very different. Moreover, MSs can be associated with specific events that are derived from user habits, e.g., departmental staff spends most of the day near the administrative offices while students congregate in the atrium, classrooms, and computer labs. The group factor can be statically applied during the network registration phase. For example, users in a typical campus network are registered with Unix accounts that are grouped based on their status. Users in large community networks are also registered based on different types of memberships, such as residential, business, free subscriber, etc. In addition to these pre-assigned groups, other groups can be dynamically formed from a set of MSs that suffer from high misprediction rate.

The time-of-day factor indicates that user behaviors change as function of time. For example, mobility patterns observed in an academic setting will change during the course of a day depending on the schedule of classes. There will be bursts of repetitive mobility patterns when students move between classes, and mobility behaviors during the evening will be different from the daytime. Most of MSs in an academic network in the evening are graduate students and, for most part, they tend to move only within limited areas, i.e., graduate student offices, laboratories, and hallways. Similarly, most users in a community network in the evening tend to stay within residential areas. In addition, both environments typically exhibit periodic behaviors such as students attending classes and workers commuting.

The duration factor, which can be categorized as short, medium, and long, directly represents how long a MS is connected to a cell and indirectly represents the speed at which it moves through a cell. Figure 2(a) shows an example of a medium duration, which represents the default case when a MS transits through a cell. For example, the cells $c_x$ and $c_y$ in Fig. 2(a) do not overlap, thus the MS has to transit through $c_y$ before it can reach the destination $c_z$. This is also the case when $c_y$ and $c_z$ overlap but a direct pathway does not exist between the two cells. In WLANs, most handoffs occur with medium duration because the traveling distance of a MS is typically a lot further than coverage area of a cell. Figure 2(a) also illustrates a long duration that represents a MS performing some activity at the destination cell $c_z$. For example, in an academic setting, MS performs a handoff to a cell that covers a classroom and then spends a long period of time listening to a lecture. Similarly, in a community network, a MS performs a handoff in the morning to a cell that covers an office building and stays connected to this cell for many hours.

A short duration often represents an unnecessary or false handoff. For instance, consider a MS moving across three adjacent cells $c_x$, $c_y$, and $c_z$ in Fig. 2(b). As the MS moves from $c_x$ to $c_y$, and then to $c_z$, if the connection duration for $c_y$ is very short then it indicates the three adjacent cells are highly overlapped. This is important because if the degree
of overlapping is sufficiently high then the prediction for the handoff sequence \( c_x \rightarrow c_y \rightarrow c_z \) can be replaced by \( c_x \rightarrow c_z \), which allows MS to move from \( c_x \) directly to \( c_z \), eliminating an extra handoff. Therefore, short duration together with other behavior factors can be used to identify and eliminate these unnecessary handoffs.

Based on our analysis of the simulated networks (see Sect. 4), a typical short duration is few seconds while a long duration lasts more than few minutes. All others are considered as medium duration.

3.2 The proposed method

Figure 3 illustrates a possible configuration for the proposed method, where the BMP scheme is assumed to be implemented by a server and is collocated with the authentication server. Moreover, the next-cell prediction \( c_w \) for the current handoff from \( c_x \) to \( c_w \) is assumed to be available from the last handoff from \( c_y \) to \( c_x \). Therefore, a MS directly authenticates/re-associates with \( c_w \), eliminating the need to perform scanning. A full scan is performed when an MS joins the network for the first time.

When the MS requires a handoff from \( c_x \) to \( c_w \), it sends \( \text{pred}_\text{req}(ID, HS, D) \), where \( ID \) is the ID of the MS, \( HS \) is the handoff sequence for the MS, and \( D \) is the duration of time spent by the MS in the last cell \( c_x \), to the server as a part of the authentication request via \( c_w \) to obtain predictions for the next handoff (Step 1).

The server then performs Next-cell Prediction (Step 2), which involves the following operations illustrated in Fig. 4. First, the last \( k \) handoffs in the handoff sequence, i.e., \( c_x \rightarrow c_w \), are used to search the HS Table for a set of matching entries representing an unordered next-cell Prediction List (uPL). At the same time, \( ID \) is used to index the Group Table to obtain the group ID \( G_i \) in \( \{G_0, G_1, \ldots, G_p\} \), where \( p \) represents the number of groups. Second, \( G_i \) is used to select a particular group’s Time-of-Day characteristic \( ToD_i \). As the name suggests, \( ToD_i \)s model time-of-day characteristics of different groups.
of mobile users. The default period for a ToD, e.g.,
T = 1-day, which means MSs that belong to this group
exhibit similar handoff behavior during the entire period
T. Moreover, a group of MSs that exhibit short-term and
periodic handoff behaviors can be separately modeled as
ToD = \{\tau_0, \tau_1, \ldots, \tau_{q-1}\}, where \tau_j represents a time segment,
q represents the total number of time segments, and
sum_{j=1}^q \tau_j = \mathcal{F}. Thus, the current time (i.e., clock) determines \tau_j
for a particular group \mathcal{G}_i, i.e., (ToD, \tau_j). The importance of
(ToD, \tau_j) is that it uniquely defines a set Integrated Moving
Averages (IMAs) to represent the frequencies of handoff
sequences for a particular group of MSs with specific time-
of-day and duration characteristics. Therefore, (ToD, \tau_j)
together with HS_{match} determine the proper subset of IMAs
be used in applying the priority order for the next-cell
predictions, i.e., ordered Prediction Lists, for both medium
(oPL_{med}) and long (oPL_{long}) durations.

Once the priority is determined, the BMP server returns
pred_respon(oPL_{med}, oPL_{long}) as a part of the authentication
response (Step 3). After receiving the authentication
response, the MS reassociates with the AP in cell \mathcal{C}_w (Step
4). This is followed by Ethernet Switch Update (Step 5)
and Prediction Update operations (step 6), which complete
the handoff.

During the next handoff, the 1st prediction in either
oPL_{med} or oPL_{long} is used as the next cell prediction based
on whether the duration of the MS in the current cell \mathcal{C}_w is
medium or long, respectively. The next-cell prediction
defaults to oPL_{med} if the MS experiences a short duration in
the current cell. If the 1st prediction fails, i.e., authentication/re-association fails, the 2nd prediction is used, and so
on. A full scan is performed when all the predictions in
oPL_{med}/oPL_{long} fails.

The following discusses the detailed operations of BMP.

3.2.1 HS Table

The mobility history of a MS is represented by a handoff
sequence of length l denoted as (c_{n-l-1}, \ldots, c_n), where c_i
indicates the cell ID of i-th visited cell. Figure 5 shows the
HS Table, which is a collection of unique handoff sequences
that represents the global history of mobility patterns in the
network. The mobility prediction based on the HS Table can
be represented by an order-k Markov process
\[
\tilde{P}(X_{n+1} = c_{n+1} | X(n-k+1, n) = (c_{n-k+1}, \ldots, c_n)) = \frac{N(c_{n-k+1}, \ldots, c_n, c_{n+1}, L)}{N(c_{n-k+1}, \ldots, c_n, L)}
\]
(1)
where \(k = l - 1\), \(L\) is the overall observed history of
mobility patterns, and \(N\) is the average frequency of
mobility patterns. Our simulation study in Sect. 4 shows
that \(l = 3\) provides accurate next-cell predictions and yet
the amount of mobility history needed is minimal.

3.2.2 IMA

IMAs are typically used in forecasting time-series data and
can be derived from the more general AutoRegressive
Integrated Moving Average (ARIMA) model. ARIMA is
defined as ARIMA (p, d, q), where p, d, and q refer to the
order of the autoregressive, the differencing, and the
moving average parts of the model, respectively. Exponent­
ial Weighted Moving Average (EWMA) is equivalent to
ARIMA (0, 1, 1) [16, 17] and is much simpler to for­
mulate than the general ARIMA model. The predicted
frequency of each HS for the next period \(t + 1, z_{t+1}\), can be
defined by the following EWMA:
\[
z_{t+1} = (1 - \lambda)z_t + \lambda z_t
\]
(2)
where \(z_t\) is the frequency of the handoff sequence during
the current period \(t\), \(z_t\) is the predicted frequency of the
handoff sequence in the current period \(t\), and \(\lambda\) is the
smoothing factor \(0 < \lambda < 1\). \(t\) represents the minimum
time interval for the time series data (i.e., \(t = 1\) min). The
parameter \(\lambda\) determines the characteristic of the EWMA
model and is typically chosen experimentally. Based on our
analysis of the simulated networks (see Sect. 4), \(\lambda\) for
the time-series data representing the frequency of handoff
sequences is chosen to be 0.1. Although EWMA does not
rely on a full statistical analysis, this simple model gives
results that are relatively close to ones from ARIMA.

The IMA structure is shown in Fig. 6, which is essen­
tially a three-dimensional structure. HS_{match} and ToD_{i}
are used to index the first two dimensions of the IMA structure.
This results in a set of EWMA, i.e., \(z_{t+1}\), which cor­
responds to a particular group’s (\mathcal{G}_i) time-of-day charac­
teristics (ToD). Each element in the last dimension is indexed
by \(\tau_j\) corresponding to a particular time segment containing
a set EWMA’s \(z_{\tau_j+1}^{med}, z_{\tau_j+1}^{long}\), where \(z_{\tau_j+1}^{med}\) and \(z_{\tau_j+1}^{long}\) represent
EWMA for medium and long durations, respectively,
and \(c\) represents the predicted cell. Then, uPL from the HS

<table>
<thead>
<tr>
<th>Past</th>
<th>Current</th>
<th>Next</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>c_y</td>
<td>c_x</td>
<td>c_w</td>
</tr>
<tr>
<td>c_x</td>
<td>c_w</td>
<td>c_y</td>
</tr>
<tr>
<td>c_x</td>
<td>c_w</td>
<td>c_y</td>
</tr>
</tbody>
</table>

Fig. 5 Handoff Sequence Table. This table shows a possible entries
for the example shown in Fig. 3, where the last \(k = l - 1\) visited
cells (i.e., \(c_1 \rightarrow c_n\)) are used to search for the next cell predictions

3.2.3 Dynamic group formation

Separate IMAs are maintained for different user groups. Groups that are formed statically are referred to as parent groups. Child groups can be dynamically formed from a parent group by applying heuristics to a set of MSs that suffer from a high misprediction rate, which can be caused by differences or changes in mobility behaviors within the group. In order to gather the necessary information, each MS keeps track of and submits its average 1st prediction accuracy, $\mu_i$, to the BMP server at the end of each period $T$. If the server detects a set of MSs in a group $G_i$ with their average 1st prediction accuracy lower than the threshold $\rho_i = \mu_i - 0.5\sigma_i$, where $\mu_i$ and $\sigma_i$ represent the average and standard deviation of the 1st prediction accuracy for the group $G_i$, during the last period $T$, a new child group $G_{i+1}$ is formed with its own ToD$_r$. This will improve the prediction accuracy for those set of MSs. However, if a set of MSs within this newly formed child group still suffer from high misprediction rate, i.e., their average 1st prediction accuracies are lower than $\rho_i$, of the parent group, then these MSs rejoin their parent group. After rejoining their parent group, if these MSs as well as other MSs continue to suffer from high misprediction rate, they will form their own child group. This process repeats every period $T$. In order to reduce the processing load, the number of MSs in a group must be more than some threshold $M$ based on the traffic characteristics of the network and capability of the server.

3.2.4 Time-of-day characteristic

The time-of-day factor is applied when the prediction accuracy of a group is lower than some threshold for a period of time. This is done by associating a separate time segment $\tau_j$ with its own IMA. The motivation for generating time-segments is to isolate periods when mobility patterns can be better identified. The BMP server does this by keeping track of the prediction accuracy of each group during a period of $\tau_{\text{min}}$ (e.g., 1 h), which is the minimum time-segment length. If the prediction accuracy of a group $G_i$ is lower than $\rho_i$ during the last $\tau_{\text{min}}$, the server assigns a new IMA to the time-segment. Again, the value of $\tau_{\text{min}}$ is based on the network traffic and server capability.

3.2.5 Update operations

Different update operations are performed depending on different time periods.

- **HS Table Update**: A new HS Table entry is allocated and updated when a new handoff sequence is encountered.
- **Group Table Update**: The following operation is performed after each period $T$:
- **IMA Structure Update**: $z_t$ is determined based on the number of times handoff sequences of MSs match with an entry in the HS Table during the current period $t$. If there are no matches, then $z_t$ is 0. Then, $z_{t+1}$ is calculated for the next period $t+1$ using Eq. 2. Note that $z_t$ is not maintained for a handoff sequence with short durations. This way, any false or unnecessary handoffs are eventually removed from the behavior model.
- **Group Table Update**: A MS is always registered to a parent (i.e., static) group. A MS can also be registered to a child group, and unregistered from the child group when they rejoin the parent group. Thus, a MS registered to a child group will also be registered to a parent group. Once a child group is formed, the priority is given to the child group for determining their time-of-day characteristics.
- **ToD Update**: When IMA values of two consecutive time segments are close together (e.g., closer than 10%) they are combined.

3.3 A distributed implementation

Note that the discussion of BMP thus far has been based on a centralized scheme. However, BMP can also be implemented using a distributed scheme where each AP maintains its own portion of the global mobility history shown in Fig. 4. This can be achieved by relaying next-cell
prediction requests using re-association requests. For example, consider again the handoff between \( c_x \) and \( c_w \) shown in Fig. 3. After authentication, MS sends reassociation request containing \( \text{pred}_{\text{req}}(ID, HS, D) \) to \( c_w \). Then, \( c_w \) performs next-cell prediction and returns \( \text{pred}_{\text{resp}}(oP_{\text{med}}, oP_{\text{long}}) \) as a part of the reassociation response. Finally, \( c_w \) sends the handoff sequence as a part of either Inter-Access Point Protocol (IAPP) [18] or a vendor specific protocol to \( c_x \). After \( AP_x \) receives the handoff sequence, update operations are performed. This allows MSs handing off from \( c_x \) to obtain the next-cell prediction list.

4 Performance evaluation

This section presents the performance evaluation of the BMP scheme and compares its accuracy and delay results against using only location information, such as our prior work on Global Path Cache (GPC) [14], which has been shown to be superior to Neighbor Graph [12] and Selective Scanning with Caching [13].

4.1 Simulation environment

The two network topologies used in the simulation study are shown in Figure 7, which are the first floor of the four-story, 153,000-ft\(^2\) Kelley Engineering Center (KEC) at Oregon State University, and a public WLAN service that covers 2.5-mile\(^2\) area of Portland, Oregon, called MetroFi\(^a\) [1]. The APs in KEC are connected by an Ethernet switch, while APs in the MetroFi\(^a\) network are interconnected by a WMN [5].

The simulated coverage area for KEC contains 6 APs and 450 MSs, and the paths taken by MSs are limited to hallways and the atrium. There are three groups of users, i.e., Students, Graduate Students, and Faculty/Staff, with each having a different type of mobility behavior. For example, Students mostly move between the atrium, the cafe, and the computer lab. In addition, Students move in and out of the classrooms during the last ten minutes of each class hour between 8 AM and 6 PM. In contrast, Graduate Students mainly move between their offices, the atrium, and the computer lab. Finally, Faculty/Staff moves mainly between their offices and the atrium. All MSs move at a pedestrian speed of 1.5 m/s.

The coverage area for Portland (indicated by the dotted line) contains 40 APs and 4,500 MSs and the paths taken by MSs are limited to sidewalks. The results for Portland were generated based on nine different groups of users. Nomadic represents a group of MSs that can move anywhere, any time within the simulated area. The next four groups represent Commuters who work in each of the four quadrants or regions, i.e., C-I, C-II, C-III, and C-IV in Fig. 7(b), and are likely to travel long distances (i.e., 15–20 blocks) to work. Moreover, these groups of MSs only move between 6 AM to 10 AM and 6 PM to 10 PM. The last four groups represent Residents who live in each of the four regions, i.e., R-I, R-II, R-III, and R-IV in Fig. 7(b). These groups of MSs tend to stay at their homes most of time (several hours) but when they move they are likely to only move within few blocks (5–10 blocks) from their homes.

In order to accurately simulate mobility patterns and handoffs in large networks, we developed a simulator that implements a WLAN radio model, generates realistic mobility patterns based on building and city layouts, and supports management frames needed to implement scanning, authentication, and re-association. The structure of the simulator is shown in Fig. 8, which consists of Path...
Generator, Path Finder, Mobility & Network Simulator, and Data Analysis modules.

The Path Generator generates a new destination waypoint based on the Path Graph definition, User Mobility definition, and the current Simulation State for each MS that has completed its trip between the original and destination waypoints. The Path Graph definition is a graphical representation of all the possible paths MSs can traverse within a simulated area, which is similar to the ones proposed in [19, 20]. Figure 9 shows the path graphs for the two network topologies used in the simulation study—KEC and Portland in Fig. 7. They consist of vertices representing waypoints and segments representing paths between adjacent waypoints. The User Mobility definition defines the number of MSs and APs as well how MSs move, including when and where a MS moves to and how long it stays at a waypoint. The Simulation State defines the current time and the locations of all the MSs in the simulated network area. The Path Generator uses these two definitions together with the current state of the simulator to randomly select destination waypoints based on a modified Random Waypoint model [21, 22]. Our modified Random Waypoint model allows a probability distribution to be assigned to sub-areas or regions within a path graph based on different groups of MSs at different times. The Path Finder module then uses the path-finder algorithm [23] to generate the shortest path between the source and destination waypoints. The resulting path consists of multiple segments, which are then fed to the Mobility & Network Simulator.

The Mobility & Network Simulator consists of Mobility module, Radio Model, Handoff Detector, and MAC sub-Layer Management Entity (MLME). The Mobility module simulates the movements of MSs on the segments at one meter resolution. The Handoff Detector monitors each MS’s movement, and based on the Cell Coverage definition and the Radio Model, which is based on log-distance path loss model [24], performs a handoff when the distance between a MS and its associated AP reaches the maximum radius of the coverage area. Handoffs are performed using the MLME module, which supports beacons, probing, authentication, and re-association. Finally, the Data Analysis module records the number of channel switches, the number of times MS has to wait for MaxChannelTime, MinChannelTime, Authentication delay/timeout, and Re-association delay (see Sect. 4.2).

![Fig. 8 Simulation model](image)

![Fig. 9 Path graphs for Fig. 7. a KEC. b Portland](image)

**Table 1 Delay parameters used in the simulation**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Delay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channel switching time ($t_{\text{switch}}$)</td>
<td>11.4 ms</td>
</tr>
<tr>
<td>MinChannelTime ($t_{\text{min}}$)</td>
<td>20 ms</td>
</tr>
<tr>
<td>MaxChannelTime ($t_{\text{max}}$)</td>
<td>200 ms</td>
</tr>
<tr>
<td>Authentication delay ($t_{\text{auth}}$)</td>
<td>6 ms</td>
</tr>
<tr>
<td>Re-association delay ($t_{\text{reassoc}}$)</td>
<td>4 ms</td>
</tr>
</tbody>
</table>
4.2 Simulation delay parameters

The delay parameters used in the simulation are shown in Table 1: **Channel Switching Time** \( (t_{\text{switch}}) \) is the time required to switch from one channel to another; **Min-ChannelTime** \( (t_{\text{min}}) \) is the minimum amount of time a MS has to wait on an empty channel; **MaxChannelTime** \( (t_{\text{max}}) \) is the maximum amount of time a MS has to wait to collect all the probe responses, which is used when a response is received within **MinChannelTime**; **Authentication delay/timeout** \( (t_{\text{auth}}) \) is the time required to perform authentication based on MAC addresses; and **Re-association delay** \( (t_{\text{assoc}}) \) is the time required to perform re-association.

These delay values represent the current off-the-shelf NICs and obtained with an experimental setup that consisted of two laptops with 802.11a/b/g NICs based on Atheros AR 5002X chipset [25] running Linux 2.6 with MadWiFi driver [26], a Sun SPARC Server with Ethernet LAN NIC, and an HP ProCurve Wireless Access Point 420. Measurements were gathered by having one laptop serve as a MS that performs a channel switch and continuously sends UDP traffic to a daemon running on the server while the second laptop sniffs the traffic using **tcpdump** (see [14]).

4.3 Simulation results

All the results in this section were obtained based on the assumption that the HS Table contains a complete history of handoff patterns. This is achieved by first running the simulations for \( 10^4 \) handoffs for KEC and \( 10^6 \) handoffs for Portland to allow the BMP system to “learn” all the handoff sequences [14], and then gathering statistics for up to \( 10^7 \) handoffs.

Figures 10 and 11 show the number of required HS Table entries and the average number of next-cell predictions returned per handoff, respectively, as function of \( l \) (KEC on the left, Portland on the right). As can be seen from Fig. 10, more entries are required to keep track of MSs’ more complex moving paths as \( l \) increases. Moreover, the number of entries needed for Portland is higher and grows significantly faster than KEC due to larger number of APs. Similarly, Portland has, on average, a larger set of next-cell predictions due to a larger number of highly overlapped regions. Note that, for implementation purposes, each entry in the HS Table contains multiple next-cell predictions instead of one prediction entry as illustrated in Fig. 5. Therefore, Figure 10 represents the number of unique handoff sequences of length \( l - 1 \) observed in the two networks. The product of the number of entries in Fig. 10 and the average number of predictions returned per handoff in Fig. 11 represents the total number of unique handoff sequences of length \( l \).

Figure 12 shows the accuracies of individual next-cell predictions based only on location information (starting with the 1st prediction result on the bottom), which are prioritized based only on their frequencies of occurrence. The overall accuracy for both KEC and Portland is 100%, which means one of the next-cell predictions is guaranteed to be correct. However, the accuracy of each prediction is more important because a misprediction incrementally increases the average handoff delay by \( t_{\text{switch}} + t_{\text{auth}} \). For KEC, the 1st next-cell prediction accuracy is 68% and increase slightly as function of \( l \). The 1st next-cell predictions that fail are satisfied by the 2nd next-cell predictions, which represent accuracy rate of 28.5 ~ 25.1% relative to all the next-cell predictions. Similarly, the 3rd next-cell predictions that succeed make up 3.5 ~ 3.1% relative to all the predictions. For Portland, the 1st next-cell predictions accuracy start at 43% and increases as function of \( l \). The 1st next-cell predictions that fail are satisfied by the 2nd next-cell predictions, which represent accuracy rate of 28.5 ~ 25.1% relative to all the next-cell predictions. Similarly, the 3rd next-cell predictions that succeed make up 3.5 ~ 3.1% relative to all the predictions. The combined prediction accuracy beyond the 4th next-cell prediction is 3 ~ 1.43%.
Based on Figs. 10–12, it is clear that suggested size of \( l \) is 3, which provides good 1st prediction accuracy and yet the number of entries required for the HS Table is minimal.

Figure 13 shows the 1st prediction accuracy of BMP, which is important because each misprediction from the \( oPL \) incrementally increases the average handoff delay by \( t_{\text{switch}} + t_{\text{auth}} \).

The average improvement for BMP compared to using only location information for KEC is 23%. The largest improvement comes from Static Group at 12.7%, followed by Dynamic Group, ToD, and Duration with 7, 1.7, and 1.7%, respectively. These improvements also vary for different user groups. For example, BMP improves the 1st prediction accuracy for all three groups in KEC. The largest improvement of 77% comes from Students because their behaviors are dictated by schedule of classes, which causes their handoffs to be bursty and periodic and also results in longer durations. This is followed by Graduate Students with an overall improvement of 61%, which is the result of being able to better capture the fact that they work late at night and on weekends. In contrast, improvement for Faculty/Staff was less at 9.5% because the number of MSs in this group is much less than the other groups and the coverage area affected by their mobility is much more limited.

The average improvement for Portland is much more significant at 43%. The largest improvement came from Static Group at 28%, and this is followed by Duration at 7.7%, ToD at 4.9%, and Dynamic Group at 2%. Again, all of the user groups in Portland resulted in improvements. Among them, Nomadic and Commuter groups (C-I~IV) exhibit large improvements due to short-term surges in handoffs caused by groups of users commuting during rush hour. The Nomadic group is hardly affected by Dynamic Group, ToD, and Duration factors because their mobility behavior is uniformly random across the entire network and independent of time. In contrast, the largest improvement for Commuter groups came from the ToD characteristic. For Resident groups (R-I~IV), the Duration factor provided the most improvement.

Figure 14 compares the average handoff delays based on the parameter set defined in Table 1 (Location-based scheme on the left, and BMP on the right). These results show that BMP provides 10 and 21% improvement for KEC and Portland, respectively. This may appear to be only a small improvement compared to the location-only
5 Related work

Many different mobility prediction techniques have been proposed for a variety of wireless networks, such as cellular [27–32], WLANs [10–13, 33, 34], ad hoc networks [35, 36], and mesh networks [5], and applied to reduce handoff latency [8, 12, 13, 37], provide efficient resource reservation for either GPS or triangulation of signal strengths [29, 30]. The basic idea is to provide predictions by searching the database using user contexts stored in MSs. If a match or multiple matches are found, predictions are provided based on location stored in the context. However, data-mining techniques require a large storage and fast processors to properly analyze long-term mobility behavior of users. In addition, most techniques typically require a GPS device to obtain information about locations and directions of MSs. For systems that rely on signal triangulation, their effectiveness may be limited due to the fact that mobile devices are mainly used for indoors and crowded outdoor areas where the signal strength is highly affected by noise rather than distance [24].
6 Conclusion

This paper presented the BMP technique to improve mobility prediction in WLANs. BMP models mobility patterns not only based on the location information of MSs but also captures their group, time-of-day, and duration characteristics. Therefore, it monitors the next-cell prediction accuracies and readjusts its prediction model based on the dynamic behavior of MSs. Our simulation study shows that BMP provides better predictions than location-only based schemes, and results in much lower handoff delay for all MSs.

For future work, we plan to investigate couple of issues. First, we plan to study the effectiveness of BMP for other wireless networks, such as cellular networks, WiMAX, and WMNs, not only to improve handoff latency but also to improve resource reservation, routing protocols, and energy efficiency. Second, we would like to investigate how BMP can be utilized to speed up vertical handoffs in heterogeneous networks.

Acknowledgments The work described in this paper was supported in part by the NSF under Grant CNS-0831853 and CNS-0821319, and Korean NRF under WCU Grant R31-2008-000-10100-0.

References


