Fine-Grained Mobility Characterization: Steady and Transient State Behaviors

Wei Gao and Guohong Cao
Department of Computer Science and Engineering
The Pennsylvania State University
University Park, PA, 16802, USA
{weigao, gcao}@cse.psu.edu

ABSTRACT

Recent popularization of personal hand-held mobile devices makes it important to characterize the mobility pattern of mobile device users, so as to accurately predict user mobility in the future. Currently, the user mobility pattern is mostly characterized at a coarse-grained level, in the form of transition among wireless Access Points (APs). There is limited research effort on the fine-grained characterization of geographical user movement. In this paper, we present a novel approach to characterize the steady-state and transient-state user mobility behaviors at a fine-grained level, based on the Hidden Markov Model (HMM) formulation of user mobility. By applying our approach on both realistic mobility traces and synthetic mobility scenarios, we show that our approach is effective in characterizing user mobility pattern and making accurate mobility prediction. We also experimentally demonstrate that fine-grained user mobility knowledge is more effective to improve the performance of a variety of mobile computing applications.

Categories and Subject Descriptors

C.2.3 [**Network Operations**]: Network monitoring;
C.4 [**Performance of Systems**]: Modeling techniques

General Terms

Algorithms, Measurement, Performance

Keywords

Mobility Characterization, Hidden Markov Model, Mobility Prediction, Localization

1. INTRODUCTION

Recent popularization of personal hand-held mobile devices motivates the design and deployment of a variety of mobile computing applications, where user mobility is considered as one of the major reasons for performance degradation, because devices may lose connection with each other when they move. Hence, the knowledge of user mobility obtained by mobility characterization techniques is of special importance on predicting such disconnection and minimizing its impact on the network performance.

A lot of research efforts have been done on how to characterize the user mobility pattern, in the form of transition among wireless Access Points (APs), based on realistic traces collected from campus or corporate wireless LANs [10, 2, 13, 14, 23]. Earlier studies [10, 2] focused on extracting numeric characteristics from user traces, and some later schemes [13, 14] provided more precise mathematical formulations of the user mobility pattern. In [14], an empirical model is proposed to characterize the space-time joint registration patterns of users, and in [13] the temporal distribution of node AP associations is characterized by exploiting semi-Markov process. In [23], several mobility predictors are analyzed and compared, and the Order-2 Markov predictor is considered the best. This predictor is also used in practical system design [19].

In these schemes, the characterized node mobility pattern is obviously simplified, and only provides coarse-grained prediction on the node[1] movement and location in the future. Such coarse-grained mobility prediction has been proved to be far from enough for many applications, in which more fine-grained node mobility knowledge is required for satisfiable network performance. For example, in mobility-aware clustering [16, 1], a node wants its clusterhead to have the similar mobility pattern with itself, and thus it can avoid the frequent disconnections with its clusterhead. In Delay-Tolerant Networks (DTNs) where only intermittent network connectivity is available [8, 9], fine-grained node mobility knowledge is used to improve the efficiency of data forwarding [17].

In this paper, we present a novel fine-grained mobility characterization approach. In the narrow sense, fine-grained mobility characterization means the accurate prediction of geographical user movement in the future, but we also expand its scope in both temporal and spatial aspects. Our detailed contributions are listed as follows.

- We formulate node mobility at a fine-grained level based on the Hidden Markov Model (HMM).
- We provide fine-grained predictions of node mobility at both steady-state and transient-state time scales.
- We characterize the inter-dependency features of node mobility in both temporal and spatial aspects.

The key component of our approach is the node mobility formulation based on HMM. In this formulation, a node periodically observes its own mobility, and the mobility observations are used as training data for HMM parameter re-estimation, to ensure that the HMM captures the up-to-date node mobility pattern. The general HMM model and the parameter re-estimation algorithm are

[1]In the rest of this paper, without loss of generality, the terms "user" and "node" are used interchangeably.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
MobiHoc'10, September 20–24, 2010, Chicago, Illinois, USA.
Copyright 2010 ACM 978-1-4503-0183-1/10/09 ...$10.00.

extended to bridge the gap between numeric mobility observations and discrete HMM states, and to meet the special challenge in mobile network environments.

The rest of this paper is organized as follows. Section 2 gives an overview of our approach and highlights our motivation of using the HMM mobility formulation. Sections 3 and 4 describe the details of characterizing node mobility based on HMM. The evaluation results are shown in Section 5. Section 6 demonstrates the utilization of fine-grained mobility knowledge for improving the performance of mobility-aware clustering. Section 7 finally concludes the paper.

2. OVERVIEW

2.1 Motivation

The key difference between coarse-grained and fine-grained mobility characterization is illustrated in Figure 1, where the node movement trajectory passes through three APs, and the AP coverage range is indicated by the dashed circles. Coarse-grained mobility characterization schemes characterize node mobility at a discrete level, in the form of AP or geographical area association, and therefore only provide coarse-grained estimation (velocity vectors $\vec{v}_{c1}$ and $\vec{v}_{c2}$) on the actual node movement. Comparatively, vectors $\vec{v}_{f1}$ and $\vec{v}_{f2}$ estimated by fine-grained mobility characterization schemes from the node location records A, B, C and D provide much more accurate estimation on the actual node mobility.

Node mobility is naturally formulated as a discrete-time stochastic process. Markov process has been used for coarse-grained mobility characterization [13, 19], in which there is explicit correspondence between node mobility and the Markov states. In [13] each state corresponds to an AP, and in [19] each state corresponds to a geographical area with a fixed size.

However, for fine-grained mobility characterization, it is hard to find such explicit correspondence between mobility observations which are normally valued in a continuous range, and the discrete Markov states. This is our motivation for using Hidden Markov Model (HMM) for the mobility formulation. Since in a HMM each state is "hidden" and associated with an observation probability density function (PDF), we can exploit these observation PDFs to bridge the gap mentioned above. That is, each Markov state is associated with a specific value range of mobility observations, called a *mobility stage*, via the associated observation PDF of that state. For the HMM of a specific node, we allocate its mobility stages to ensure that they cover the entire value range of the corresponding mobility observations of that node. As a result, we are able to characterize the node mobility by re-estimating the parameters of mobility stages using the mobility observations as training data.

In our approach, node mobility pattern is characterized at both steady-state and transient-state time scales. Steady-state pattern represents the cumulative node mobility features during a long period of time, and is normally represented by certain statistical metrics. Transient-state pattern corresponds to smaller time scale, and is more susceptible to the randomness of node mobility. Therefore, the transient-state pattern can only be represented in the form of probabilistic distribution.

2.2 Hidden Markov Model

Being different from a normal Markov process, which consists of a discrete state space $\mathbf{S} = \{s_1, s_2, ..., s_N\}$, a state transition probability matrix $\mathbf{A} = \{a_{ij}\} \in \mathbb{R}^{N \times N}$ and an initial state distribution $\mathbf{\Pi} = \{\pi_i\}$, a HMM hides its states behind a set of multi-dimensional observation PDFs $\mathbf{B} = \{b_i(\mathbf{X})\}$, where $\mathbf{X}$ is a mobility observation and each $b_i(\mathbf{X})$ is associated with a state s_i.

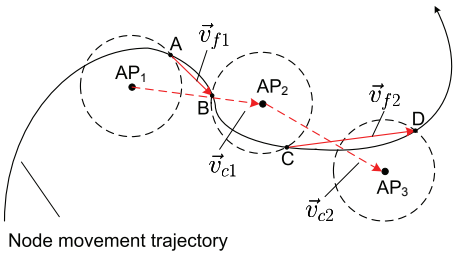


Figure 1: Difference between coarse-grained and fine-grained mobility characterization

For a HMM $\mathcal{H} = (\mathbf{S}, \mathbf{A}, \mathbf{B}, \mathbf{\Pi})$, the occurrence probability of an observation sequence $\mathbf{O} = \mathbf{O}_1\mathbf{O}_2...\mathbf{O}_L$ with the state sequence $\mathbf{I} = i_1 i_2 ... i_L$, is

$$\mathbb{P}(\mathbf{O}|\mathbf{I}, \mathcal{H}) = \prod_{k=1}^{L} b_{i_k}(\mathbf{O}_k),$$

and the cumulative occurrence probability of $\mathbf{O}$ is

$$\begin{aligned} \mathbb{P}(\mathbf{O}|\mathcal{H}) &= \sum_{\mathbf{I}} \mathbb{P}(\mathbf{O}|\mathbf{I}, \mathcal{H}) \cdot \mathbb{P}(\mathbf{I}|\mathcal{H}) \\ &= \sum_{\mathbf{I}} \pi_{i_1} \cdot \prod_{k=1}^{L-1} b_{i_k}(\mathbf{O}_k) a_{i_k i_{k+1}} \cdot b_{i_L}(\mathbf{O}_L). \end{aligned} \quad (1)$$

2.3 Mobility Observation

To characterize the mobility pattern, a node periodically observes its mobility, to which various forms of mobility observations can be applied. In our approach, we use the velocity vector of a mobile node as the form of its mobility observations.

We assume that each node is able to periodically locate itself with satisfiable accuracy via the low-cost GPS devices. When the GPS signal is unavailable due to energy depletion or indoor environment constraint, the node localization can still be done via GPS-free localization techniques [12]. As a result, if a node locates itself at a period T, it calculates its velocity vector for the i-th period as

$$\vec{v}_i = [v_{x_i}, v_{y_i}]^{\mathrm{T}} = \frac{1}{T} \cdot \Big([x_i, y_i]^{\mathrm{T}} - [x_{i-1}, y_{i-1}]^{\mathrm{T}} \Big),$$

where $(\cdot)^{\mathrm{T}}$ indicates matrix transpose, and $[x_i, y_i]^{\mathrm{T}}$ and $[x_{i-1}, y_{i-1}]^{\mathrm{T}}$ are the node locations at time iT and $(i-1)T$, respectively. $\vec{v}_i$ represents the average displacement of a node in unit time, which can be equivalently considered as the node location change.

3. NODE MOBILITY FORMULATION

3.1 Mobility Stage

In our approach, each mobility stage corresponds to a range of the direction of node velocity vectors. An example of a set of 2-dimensional mobility stages is shown in Figure 2, where the number of mobility stages is $N = 8$. Each mobility stage corresponds to a sector-shaped area. Initially, the angle range of the i-th mobility stage is $[\theta_i, \theta_{i+1}]$, where $\theta_i = \frac{i-1}{N} \cdot 2\pi$ for $1 \le i \le N$.

The association of mobility stages to the HMM states is done by setting the mean vectors $\boldsymbol{\mu}_i$ of the observation PDFs $b_j(\mathbf{X})$. The initial mean vectors are shown in Figure 2, where the direction of the mean vector $\boldsymbol{\mu}_i$ of the i-th state is set to $(\theta_i + \theta_{i+1})/2$. The magnitudes of the mean vectors $|\boldsymbol{\mu}_i|$ are uniformly initialized as the average of the first L mobility observations, where L is the size of the training set for HMM parameter re-estimation.

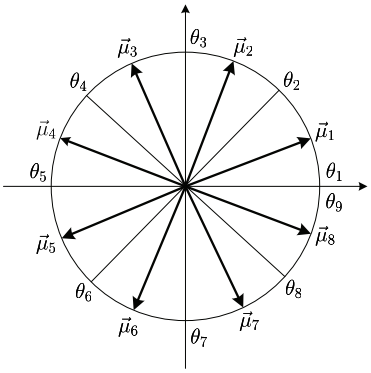


Figure 2: Mobility stages

In our approach, normal distribution is used as the form of observation PDFs, and the above initialization therefore measures the probability of the initial L mobility observations falling into a specific mobility stage. The mean vectors $\boldsymbol{\mu}_i$ and covariance matrices $\boldsymbol{\Sigma}_i$ of $b_i(\boldsymbol{X})$ are then re-estimated based on the mobility observations, and the node mobility pattern is characterized in the form of transition among mobility stages, which are determined by the transition probabilities among the HMM states.

3.2 Parameter Re-estimation

To characterize the up-to-date node mobility pattern, the parameters of mobility stages of a HMM are iteratively re-estimated using the mobility observations as the training sets, so as to maximize the occurrence probabilities of the used training sets. In our approach, we extend the basic HMM parameter re-estimation algorithm by applying weights on the mobility observations in a training set to consider their temporal-dependent validity.

3.2.1 Re-estimation Algorithm

Initially, the initial state distribution and the state transition probabilities in a HMM are all set uniformly, i.e., $\pi_i = 1/N$, $a_{ij} = 1/N^2$, where N is the number of states in the HMM. After a node obtains the first L mobility observations, it uses this first training set to re-estimate the parameters of mobility stages.

For a set of mobility observations $\mathbf{O} = \mathbf{O}_1\mathbf{O}_2...\mathbf{O}_L$, re-estimation for the HMM $\mathcal{H} = (\mathbf{S}, \mathbf{A}, \mathbf{B}, \mathbf{\Pi})$ is to maximize the likelihood $\mathbb{P}(\mathbf{O}|\mathcal{H})$ in Eq. (1). We apply a generalized Expectation-Maximization (EM) algorithm [5] for such re-estimation.

To do such re-estimation, we define two sets of auxiliary probabilities $\xi_t(i,j)$ and $\gamma_t(i)$. First, $\xi_t(i,j)$ denotes the probability of being in state s_i at time t, and in state s_j at time $t+1$, i.e.,

$$\xi_t(i,j) = \mathbb{P}(q_{t+1} = s_j, q_t = s_i|\mathbf{O}, \mathcal{H}),$$

where $t = 1, 2, ..., L-1$, and q_t indicates the current state at time t. Second, $\gamma_t(i)$ denotes the probability of being in state s_i at time t, i.e.,

$$\gamma_t(i) = \mathbb{P}(q_t = s_i|\mathbf{O}, \mathcal{H}).$$

Direct calculation of $\xi_t(i,j)$ and $\gamma_t(i)$ from $\mathbf{A}$ and $\mathbf{B}$ are expensive, and we exploit a forward-backward procedure [3] to calculate them. The forward probability $\alpha_t(i)$ is defined as

$$\alpha_t(i) = \mathbb{P}(\mathbf{O}_1, \mathbf{O}_2, ..., \mathbf{O}_t, q_t = s_i|\mathcal{H}).$$

$\alpha_t(i)$ can be calculated iteratively, such that

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i)a_{ij}\right] b_j(\mathbf{O}_{t+1}), \tag{2}$$

with initial values $\alpha_1(i) = \pi_i b_i(\mathbf{O}_1)$.

Similarly, the backward probability $\beta_t(i)$ is defined as

$$\beta_t(i) = \mathbb{P}(\mathbf{O}_{t+1}, \mathbf{O}_{t+2}, ..., \mathbf{O}_L|q_t = s_i, \mathcal{H}),$$

and can also be calculated iteratively as

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij}b_j(\mathbf{O}_{t+1})\beta_{t+1}(j), \tag{3}$$

with $\beta_L(i) = 1$.

$\xi_t(i,j)$ is then calculated from $\alpha_t(i)$ and $\beta_t(j)$ as

$$\xi_t(i,j) = \frac{\alpha_t(i)a_{ij}b_j(\mathbf{O}_{t+1})\beta_{t+1}(j)}{\sum_{m=1}^{N}\sum_{n=1}^{N} \alpha_t(m)a_{mn}b_n(\mathbf{O}_{t+1})\beta_{t+1}(n)},$$

and $\gamma_t(i)$ is calculated as

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^{N} \alpha_t(j)\beta_t(j)}.$$

Based on $\xi_t(i,j)$ and $\gamma_t(i)$ calculated from the current $\mathcal{H}$, parameters of $\mathcal{H}$ are re-estimated on the training set $\mathbf{O} = \mathbf{O}_1\mathbf{O}_2...\mathbf{O}_L$ using the following set of formulas:

$$\tilde{\pi}_i = \gamma_1(i) \qquad \tilde{a}_{ij} = \frac{\sum_{t=1}^{L-1} \xi_t(i,j)}{\sum_{t=1}^{L-1} \gamma_t(i)}$$

$$\tilde{\boldsymbol{\mu}}_i = \frac{\sum_{t=1}^{L} \gamma_t(i)\cdot\mathbf{O}_t}{\sum_{t=1}^{L} \gamma_t(i)} \qquad \tilde{\boldsymbol{\Sigma}}_i = \frac{\sum_{t=1}^{L} \gamma_t(i)\cdot\mathbf{M}_t(i)}{\sum_{t=1}^{L} \gamma_t(i)} \tag{4}$$

where $\mathbf{M}_t(i) = (\mathbf{O}_t - \boldsymbol{\mu}_i) \cdot (\mathbf{O}_t - \boldsymbol{\mu}_i)^{\mathrm{T}}$.

The computational complexity of the re-estimation algorithm is determined by the parameters L and N. It takes N multiplications to calculate $\alpha_{t+1}(j)$ from $\alpha_t(j)$ using Eq. (2), and the complexity to calculate $\alpha_t(i)$ for $1 \le t \le L$ and $1 \le i \le N$ is therefore $O(L \cdot N^2)$. The same complexity is needed to calculate $\beta_t(i)$ using Eq. (3). For $\xi_t(i,j)$, since the denominator only needs to be calculated once for every t, the complexity is also $O(L \cdot N^2)$. Similarly, the complexity for calculating $\gamma_t(i)$ is $O(L \cdot N)$. Therefore, we conclude that the overall computational complexity is $O(L \cdot N^2)$.

The re-estimation is conducted iteratively until it converges, and it has been proved in [3] that the re-estimation always converges. The convergence rate, i.e., the number of iteration cycles needed to achieve convergence, depends on the size of the training sets and initialization choices, and will be evaluated in Section 5.2.

3.2.2 Weighted Mobility Observations

In the above re-estimation algorithm, each mobility observation in a training set is considered equally. However, a mobility observation in past may be different from the current node mobility, and such difference increases when the mobility observation becomes older. As a result, mobility observations in a training set should be given different weights during the parameter re-estimation, such that more recent ones should be given larger weights.

For a training set $\mathbf{O} = \mathbf{O}_1\mathbf{O}_2...\mathbf{O}_L$, the weight decreasing degree is a constant factor $P = k/L$, $1 \le k \le L$. With a smoothing

factor $\rho = \frac{2}{1+k}$, the weights are

$$\begin{cases} \omega_t = (1-\rho)^{L-k} \cdot \frac{1}{k}, & 1 \le t \le k \\ \omega_t = \rho \cdot (1-\rho)^{L-t}, & k < t \le L \end{cases} \quad (5)$$

It can be verified that $\omega_t \ge 0$, $\sum_{t=1}^{L} \omega_t = 1$. Since $\frac{1}{k} < \rho \le 1$, the first k weights are equal and smaller than the next $L-k$ weights which increase exponentially. The smaller P is, the more weights are placed on the recent $L-k$ mobility observations. Although P can be valued in the range $[1/L, 1]$, in practice we normally use a value of P larger than a pre-determined threshold P_0.

The value of P should be inversely proportional to the node moving velocity. In our approach, the value of P is adaptively adjusted according to the current node velocity v_k, where k indicates the current mobility observation period. Since $P \in [P_0, 1]$, we use

$$P = f(v_k) = e^{-\frac{v_k^2}{a}},$$

where $a = -\frac{{V_{\max}}^2}{ln P_0}$, and $V_{\max}$ is the maximum node velocity in past.

To apply the weights to the parameter re-estimation, we multiply $\xi_t(i,j)$ and $\gamma_t(i,k)$ with the corresponding weight ω_t, i.e.,

$$\bar{\xi}_t(i,j) = \omega_t \xi_t(i,j), \ \bar{\gamma}_t(i) = \omega_t \gamma_t(i),$$

and use the weighted $\bar{\xi}_t(i,j)$ and $\bar{\gamma}_t(i)$ in Eq. (4).

3.3 Impact of Empirical Parameters

The HMM mobility formulation and parameter re-estimation process involve the following empirical parameters:

- T: Period of mobility observation
- L: Size of training set of mobility observations
- N: Number of states in the HMM

Intuitively, T should be inversely proportional to the average node moving velocity, and L is related to the accuracy of parameter re-estimation of mobility stages. Although larger L increases such accuracy, it also takes longer for the re-estimation to be conducted, and thus may prevent the HMM to capture the up-to-date node mobility pattern. N determines the granularity of mobility characterization. With more states in the HMM, there are more mobility stages to cover the value range of mobility observations, and therefore more fine-grained node mobility is characterized. The impact of these parameters on the performance and overhead of mobility characterization are furthermore evaluated in Section 5.

If N is too large and the training data for the parameter re-estimation is insufficient, parameter overfitting may happen. Such overfitting can be avoided by regularization methods [18]. This regularization is generally done by augmenting EM re-estimation algorithm by a penalty term. This penalty term is applied to the observation PDFs $b_i(\mathbf{X})$, and therefore controls the occurrence probability of a specific observation sequence. Moreover, some other approaches [22] have also been proposed to dynamically adjust N during the re-estimation process for better fitting accuracy.

4. NODE MOBILITY BEHAVIORS

Once the node mobility pattern has been characterized from the mobility observations, we are able to investigate the steady-state and transient-state node mobility behaviors from the characterized HMMs. In this section, we will first provide node mobility prediction methods at both steady-state and transient-state time scales, and then develop metrics to evaluate the inter-dependency features of node mobility in both temporal and spatial aspects.

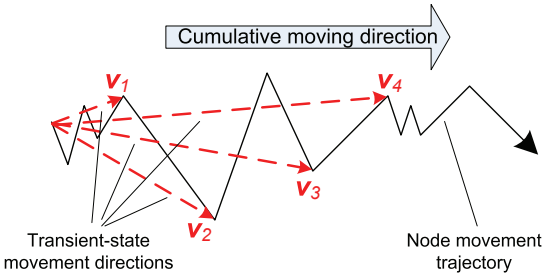


Figure 3: Zig-zag mobility pattern

4.1 Mobility Prediction

4.1.1 Steady-state Prediction

Steady-state mobility prediction estimates the cumulative moving direction of a node during a period of time. This prediction is motivated by the recent investigation on the nature of realistic human mobility [21], which can be analogous to a Levy walk. Since a typical Levy walk consists of many short epochs[2] and a few long epochs, human mobility exhibits a zig-zag movement pattern. As shown in Figure 3, although the transient-state moving directions $\boldsymbol{v}_1 ... \boldsymbol{v}_4$ can be heterogeneous, the cumulative moving direction remains unchanged. Therefore, the cumulative moving direction can be accurately predicted from the characterized mobility pattern.

In our approach, the average direction over all the mobility stages of a HMM is a natural predictor for the node's cumulative moving direction. Let $\mathbb{P}(s_i|\mathcal{H}) = \psi_i$ be the stationary distribution of the HMM $\mathcal{H}$, the vector

$$\boldsymbol{v}_{avg} = \sum_{i=1}^{N} \psi_i \cdot \boldsymbol{\mu}_i$$

acts as the mobility predictor. We are more interested in the direction of vector $\boldsymbol{v}_{avg}$. Its magnitude, i.e., the node moving velocity, is more susceptible to the randomness of node mobility, and is hence estimated at a transient-state time scale.

4.1.2 Transient-state Prediction

Transient-state mobility prediction estimates the node's velocity vector in the near future. This mobility prediction is generally more difficult due to the transient node mobility randomness.

We provide prediction in the form of probabilistic distribution. For a node with the HMM $\mathcal{H} = (\mathbf{S}, \mathbf{A}, \mathbf{B}, \mathbf{\Pi})$, suppose its most recent sequence of mobility observations is $\mathbf{O} = \mathbf{O}_1 \mathbf{O}_2 ... \mathbf{O}_L$, we are able to find the best state sequence $\mathbf{I} = i_1 i_2 ... i_L$ which maximizes $\mathbb{P}(\mathbf{O}|\mathbf{I}, \mathcal{H})$, using the Viterbi algorithm [20]. The probability distribution of the next mobility observation is then estimated as

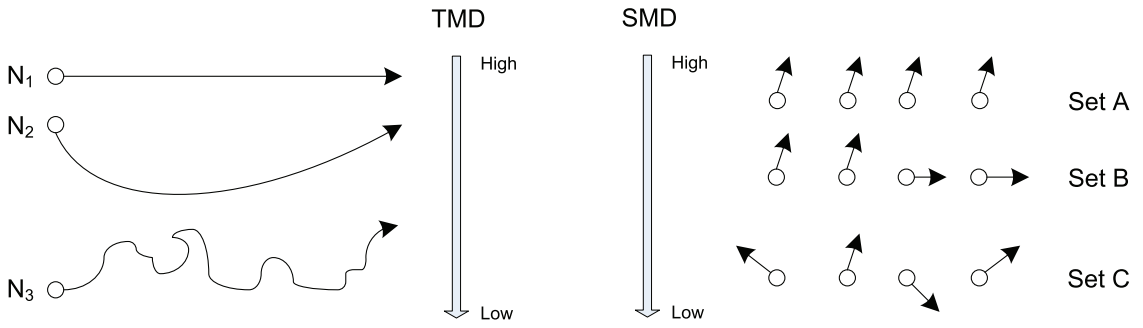
$$p_{t+1}(\mathbf{X}) = \sum_{j=1}^{N} \mathbb{P}(q_{t+1} = s_j | q_t = i_L) \cdot b_j(\mathbf{X}), \quad (6)$$

where $\mathbf{E}p_{t+1}(\mathbf{X})$ can be used as an unbiased predictor of node mobility, under the maximal likelihood criterion.

4.1.3 Discussions

The accuracy of mobility prediction mainly depends on the randomness of node mobility. If a mobile node frequently changes its moving direction at random, it will make the transient-state mobility prediction less accurate. Comparatively, the accuracy of steady-state mobility prediction is less sensitive to such changes, as long

[2]Here an epoch refers to a period of time in which the corresponding object moves straightly to a destination.



(a) Temporal Mobility Dependency (TMD) (b) Spatial Mobility Dependency (SMD)

Figure 4: Illustration of temporal and spatial mobility features

as the node moves in a zig-zag pattern. The randomness of node mobility generally increases if a node moves faster.

The accuracy of mobility prediction is also affected by the error of node localization. Such localization error is classified into two categories, i.e., the system error and the random error, but the impact of either category can be effectively limited in our approach. First, the system error applies universally to every node location record. Since our approach uses node velocity vector as the mobility observation, the system error is eliminated when the displacement between two node location records is calculated. Second, the impact of random error is limited because the parameters of mobility stages are re-estimated in an accumulative manner over the mobility observations. As described in Section 3.2, the parameters are re-estimated based on both the current parameter values and the new training set, and the random error from various mobility observations are therefore neutralized by each other.

4.2 Temporal and Spatial Mobility Features

We propose two metrics to investigate the steady-state node mobility features in the temporal and spatial aspects, corresponding to the following two questions, respectively.

- How is the current node movement related to its past history?
- How is the movement of a node related to others?

4.2.1 Temporal Mobility Dependency

In practice, since node mobility is constrained by physical laws of acceleration, the velocity and direction of node movements can only change gradually. Therefore, the current mobility of a node may depend on its past mobility history, and mobility observations in a sequence are correlated in time. We call such mobility feature as the *temporal mobility dependency (TMD)*.

DEFINITION 1. Temporal Mobility Dependency (TMD)
The TMD of a node N_j at time t with HMM $\mathcal{H}_j^{(t)}$ is defined as

$$TMD_t^{(j)} = \lim_{\Delta t \to 0} \frac{\Delta t}{d_{KL}(\mathcal{H}_j^{(t)}, \mathcal{H}_j^{(t-\Delta t)})}, \tag{7}$$

where $d_{KL}(\mathcal{H}_j^{(t)}, \mathcal{H}_j^{(t-\Delta t)})$ is the Kullback-Leibler (KL) distance measure between two HMMs $\mathcal{H}_j^{(t)}$ and $\mathcal{H}_j^{(t-\Delta t)}$.

In Eq. (7), $d_{KL}(\mathcal{H}_j^{(t)}, \mathcal{H}_j^{(t-\Delta t)})$ can be calculated as

$$\begin{aligned} d_{KL}(\mathcal{H}, \mathcal{H}^*) = & \sum_i \sum_j \mathbb{P}(s_i|\mathcal{H}) \cdot a_{ij} \cdot log\frac{a_{ij}}{a_{ij}^*} \\ & + \sum_i \mathbb{P}(s_i|\mathcal{H}) \cdot d_{KL}(b_i, b_i^*), \end{aligned} \tag{8}$$

where $\mathbb{P}(s_i|\mathcal{H}) = \psi_i$ is the stationary distribution of $\mathcal{H}$. More details about Eq. (8) can be found in the Appendix A.

In Eq. (7), TMD is defined as the reciprocal of the changing rate of $\mathcal{H}_j^{(t)}$ at time t, and has a higher value when $\mathcal{H}_j^{(t)}$ changes slower at time t. The definition of TMD is also illustrated in Figure 4(a). As shown in Figure 4(a), node N_1 has the highest TMD because it moves towards a fixed direction constantly, and its HMM remains unchanged. The TMD of node N_2 is lower because its movement is taking a circular curve. Node N_3 has the lowest TMD because its movement is completely random.

Since it is hard to get a closed form of the limit in Eq. (7), the discrete-time differentiation is used instead, i.e.,

$$TMD_k^{(j)} = (d_{KL}(\mathcal{H}_j^{(k)}, \mathcal{H}_j^{(k-1)}))^{-1}, \tag{9}$$

for each mobility observation period k. In practice, to observe the steady-state mobility behaviors of a node during a long period of time, the average TMD value $\overline{TMD}_k^{(j)}$ of every node N_j is maintained in an iterative manner, i.e.,

$$\overline{TMD}_k^{(j)} = (\overline{TMD}_{k-1}^{(j)} \cdot (k-1) + TMD_k^{(j)})/k, \tag{10}$$

for every mobility observation period k.

4.2.2 Spatial Mobility Dependency

In some realistic scenarios, the movement of a mobile node is not only constrained by the physical laws, but also constrained by the other nodes nearby. For example, on a congested highway, the speed of any vehicle is limited by the vehicles ahead [24]. We call such correlation of mobility among different nodes as the *spatial mobility dependency (SMD)*.

DEFINITION 2. Spatial Mobility Dependency (SMD)
The SMD between two nodes N_i and N_j at time t is defined as

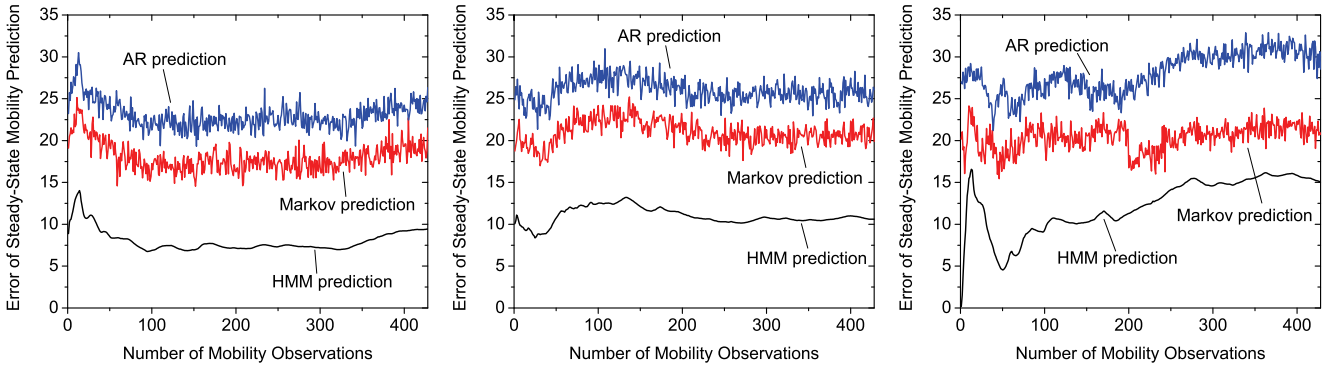
$$SMD_t^{(ij)} = \frac{1}{d_{ij}}, \tag{11}$$

where $d_{ij} = \frac{1}{2}(d_{KL}(\mathcal{H}_i^{(t)}, \mathcal{H}_j^{(t)}) + d_{KL}(\mathcal{H}_j^{(t)}, \mathcal{H}_i^{(t)}))$. Similarly, the SMD among a set $\mathcal{S}$ of mobile nodes is defined as

$$SMD_t^{\mathcal{S}} = \frac{1}{2K(K-1)} \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{S}, j \neq i} \frac{1}{d_{ij}}, \tag{12}$$

where $K = |\mathcal{S}|$.

In Eq. (11), SMD between two nodes is defined as the reciprocal of the KL distance between their HMMs, and therefore has a higher value if the HMMs are more similar with each other. This essentially means that the two nodes tend to move with similar patterns at time t. Such definition is also illustrated in Figure 4(b). In



(a) NCSU trace (b) KAIST trace (c) Disney trace

Figure 5: Accuracy of steady-state mobility prediction

Figure 4(b), the node set A has the highest SMD because all the nodes move towards the same direction. The SMD of node set B is lower, because the moving directions of some nodes are different from others. Node set C has the lowest SMD because all the nodes move towards different directions independently.

In practice, each node calculates and maintains its SMD with respect to all of its neighbor nodes based on Eq. (12). Being similar with Eq. (10), the average SMD values are maintained at mobile nodes over all the past mobility observations in an iterative manner.

5. PERFORMANCE EVALUATION

In this section, we evaluate the performance of our mobility characterization approach, using both realistic human-involved mobility traces and synthetic mobility scenarios with pre-defined parameters. We use realistic traces to evaluate the accuracy of our approach on steady-state mobility prediction and characterizing the TMD of human mobility. Synthetic mobility scenarios are used to evaluate the effectiveness of our approach on characterizing the SMD among a group of mobile nodes, as well as the accuracy of transient-state mobility prediction. Since we can simulate a large number of nodes in synthetic mobility scenarios, we also evaluate the impact of various empirical parameters on the performance of our approach.

5.1 Real Trace Evaluation

We use the NCSU human mobility trace [21], which are collected from various sites, to evaluate the performance of our approach. These sites vary from university campus (NCSU, KAIST) to large theme park (Disney) to cover different scenario sizes and human mobility patterns. Each of these traces is collected from a person carrying a GPS receiver, and thus records the complete movement trajectory of this person during a long period of time. The GPS receiver takes reading of its current position every 10 seconds and records it into a daily track log. In each trace, the radius of the trace is a half of the maximum distance that the person travels during a day. The summary of these traces is shown in Table 1.

Table 1: Trace summary

Site (No. of participants)	No. of traces	Avg. duration (hours)	Avg. radius (km)
NCSU (20)	35	10.19	1.82
KAIST (4)	46	10.62	1.26
Disney (4)	15	8.68	1.67

5.1.1 Accuracy of Steady-state Mobility Prediction

As described in Section 4.1.1, human mobility shows a zig-zag movement pattern, and steady-state mobility prediction aims at identifying cumulative node moving direction. Therefore, we measure the accuracy of steady-state mobility prediction using the error of the predicted node moving direction, and compare the accuracy of our approach with the following schemes:

1. Auto-regressive (AR) process based on linear regression: a L-order AR process is described as [4]

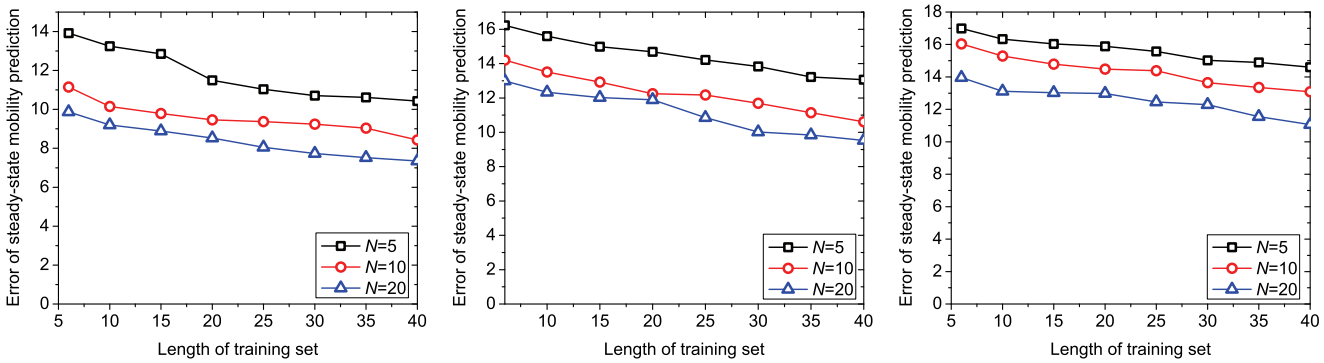
$$x_t = \phi_0 + \phi_1 x_{t-1} + ... + \phi_p x_{t-L} + \varepsilon_t,$$

where ε_t is an error term with distribution $\mathcal{N}(0, \sigma^2)$. The parameters ϕ_i and ε_t are estimated according to past L mobility observations based on the ML criterion.

2. The order-2 Markov prediction method described in [23] is considered as the best among all the Markov-based predictions. We use the modified version of the prediction used in [19] for comparison, such that each Markov state corresponds to a geographical area with fixed size and the prediction is given as a state. Based on the past node location records, the next Markov state is predicted based on the state transition probabilities, and the prediction of node moving direction is calculated from the central point of the corresponding geographical area.

The mobility predictor is trained at every mobility observation period by using the previous available mobility observations as training data. The prediction error, in terms of degrees, is shown in Figure 5, where the X-axis indicates the number of available mobility observations as training data.

According to the training method of mobility predictors, the past mobility history affects mobility prediction in the future, and therefore the average prediction error is gradually stabilized at a constant level. AR prediction is based on linear regression and it has the highest prediction error because the property of linearity may not hold for node mobility in practice. The prediction error of the Order-2 Markov approach is lower, but its average error is still around 15°. This is basically because such prediction is essentially coarse-grained and each Markov state only corresponds to a geographical area. As stated in [19], the size of each geographical area is $110m \cdot 80m$, which is considerably larger than the granularity of GPS localization itself. Reducing the area size improves the mobility prediction accuracy, but will also lead to higher computational overhead for training the predictor.



(a) NCSU trace (b) KAIST trace (c) Disney trace

Figure 6: Impact of empirical parameters on the accuracy of steady-state mobility prediction

These two approaches also show considerable instability in prediction accuracy over time, which means that they are sensitive to mobility randomness. For the AR predictor, such sensitivity is because of the deflection of node mobility from linearity. For the Order-2 Markov predictor, the sensitivity comes from the coarse-grained nature of the predictor.

In contrast, the prediction accuracy of our scheme is generally much better than the two coarse-grained approaches. As shown in Figure 5, when we set $N = 10$ and $L = 20$ and use the recent L mobility observations to re-estimate the HMM parameter, the prediction error is within 10° for the NCSU and KAIST traces, and within 15° for the Disney trace. As the training time increases, more mobility observations are taken into account for the parameter re-estimation of mobility stages, and the steady-state prediction error shows higher temporal stability against the mobility randomness. The amount of mobility observations used as training data, however, does not determine the prediction accuracy itself, which is also affected by the variation of the transient-state node mobility randomness.

We have also evaluated the impact of empirical parameters on the accuracy of steady-state mobility prediction, and the results are shown in Figure 6. With more HMM states or larger training sets, the mobility prediction is more accurate. This result is consistent with the analysis in Section 3.3; i.e., the number of HMM states controls the granularity of mobility characterization, and larger training set increases the accuracy of mobility characterization.

5.1.2 Average Temporal Mobility Dependency

We evaluated the average TMD of human mobility in NCSU traces. The results are shown in Figure 7, which are generally consistent from the investigated human mobility nature in a form of Levy walk in [21]. Since the human movement is towards a given destination during a long period of time, the characterized human mobility pattern show high temporal mobility dependency with its past mobility history. Such dependency is especially high in university campus scenarios, where the density of buildings is higher and the human movement is more constrained.

We also evaluated the effects of the number of HMM states (N) and the length of training sets (L) on the TMD values. As shown in Figure 7, the average TMD values decrease when the values of N and L become larger. The decreasing trend is higher in the NCSU trace and the Disney trace, especially when $N = 20$. The average TMD in the KAIST trace is relatively stable for different values of N and L, but still shows a decreasing trend. Such decreasing trend can be explained by the definition of TMD in Eq. (7). Since TMD is the reciprocal of the KL distance which measures the amount of discriminating information between two HMMs, the TMD value is inversely proportional to N. Besides, since larger L corresponds to a shorter mobility observation period, the difference between $\mathcal{H}_j^{(k)}$ and $\mathcal{H}_j^{(k-1)}$ in Eq. (9) increases when larger L applies, and therefore results in a lower TMD.

5.2 Simulations

The trace-based evaluations are limited by the trace scales, and is therefore difficult to evaluate the performance of our scheme in large-scale networks. In this section, we show the capability of our scheme on characterizing node mobility pattern in large-scale networks by applying it to various mobility scenarios generated from synthetic mobility models. The characterized mobility pattern is compared with the pre-defined scenario parameters to evaluate the characterization correctness. We uniformly deploy 50 mobile nodes in a $1000m \times 1000m$ area. The node communication range is 250m. The pre-set average node velocity varies between 1m/sec to 20m/sec, and the network executes 5000secs. In all the simulations, we set $N = 5$ and $T = 50/v_{avg}$.

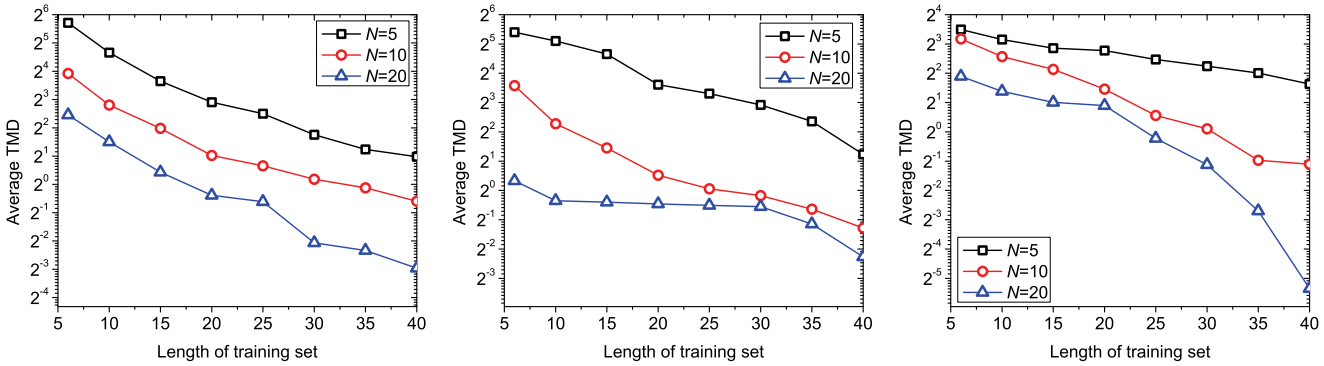
The mobility models used include the well-known Random Way Point (RWP) model, the Gauss-Markov (GM) model [15], and the Reference Point Group Mobility (RPGM) model [11]. GM and RPGM models are representatives that consider the temporal and spatial mobility dependencies, respectively. In the GM model, node velocities correlated over time are modeled as a Gauss-Markov process, such that

$$v_n = \alpha v_{n-1} + (1-\alpha)\mu + \sqrt{1-\alpha^2}x_{n-1},$$

where the correlation level is featured by $\alpha \in [0, 1]$, μ is the asymptotic mean of v_n when $n \to \infty$, and $x_n \sim N(0, \sigma^2)$ is an independent, uncorrelated and stationary Gaussian process. In the RPGM model, mobile nodes are separated into groups, and each group has a leading node determining the mobility behavior of the entire group. A RPGM mobility scenario is featured by the average number of nodes n per group, and the maximal distance r from a mobile node to the leading node of the group.

5.2.1 Impact of Empirical Parameters

We first study the impact of empirical parameters on our scheme. According to Section 4.1.2, such accuracy is evaluated by the average relative error of the magnitude of predicted node velocity vector to the actual node moving velocity. The pre-set average node velocity in all the mobility models is fixed to $v_{avg} = 5m/sec$.



(a) NCSU trace (b) KAIST trace (c) Disney trace

Figure 7: Average Temporal Mobility Dependency (TMD)

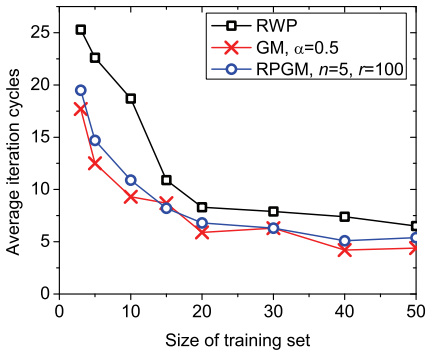
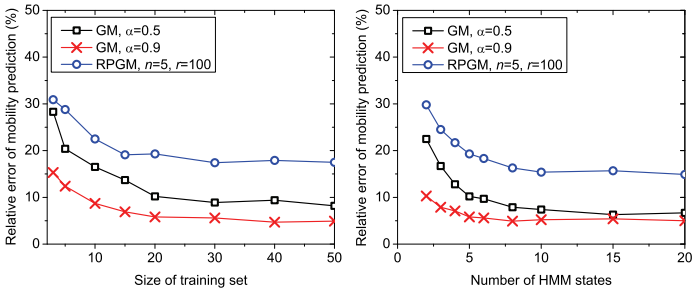


Figure 8: Iteration cycles of HMM parameter re-estimation



(a) Size of training set (b) Number of HMM states

Figure 9: Impacts of empirical parameters on the accuracy of transient-state mobility prediction

As stated in Section 3.2, the proposed algorithm for the HMM parameter re-estimation converges in all cases, but the number of iterations needed for convergence depends on the size of the training set, as shown in Figure 8. The re-estimation process generally converges faster when larger training set is used. However, when the training set is large enough (≥ 15), increasing the size of the training set will not further reduce the convergence time, but will increase the computational overhead of the parameter re-estimation. It is also shown that parameter re-estimation in RWP needs longer time to converge due to the memoryless node movement.

The impact of different parameters on the accuracy of transient-state mobility prediction is shown in Figure 9. As can be seen, all the parameters, including the size of training sets L and the number of HMM states N, show long-tail impact to the accuracy of mobility prediction. When the values of these parameters are too small, the prediction accuracy will be greatly impaired. However, too large values will not further improve the prediction accuracy. In Figure 9(a), when L changes from 2 to 20, the average relative error of mobility prediction is reduced up to 100%, but further increasing L to 50 only leads to a slight reduction of 5%. Similar feature can also be found in Figure 9(b). Since selecting larger values of these parameters will increase the computational overhead of parameter re-estimation, according to the results in Figure 9, we set $L = 20$, $N = 5$, and $T = 50/v_{avg}$ in the following simulations.

5.2.2 Characterization of Temporal and Spatial Mobility Features

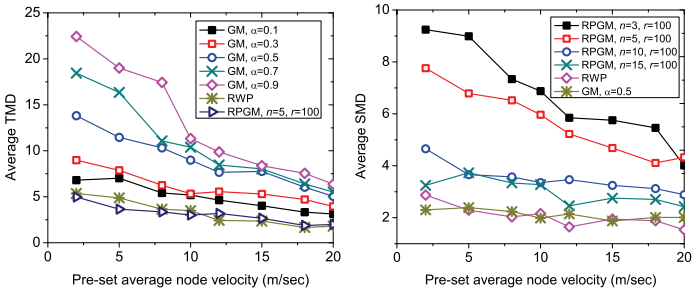
We analyze the steady-state mobility features characterized by our approach in both temporal and spatial aspects, as described in Section 4.2. Figure 10(a) shows the average TMD values for different mobility models. All the GM mobility scenarios have higher TMD values because of the intrinsic temporal correlation of node mobility. Such values also differ among different GM mobility scenarios due to the different correlation levels. When α varies from 0.1 to 0.9, the TMD values accordingly receive an increase around 40% to 200%. RWP and RPGM have much lower TMD values, because of their memoryless node movement pattern.

The average SMD values shown in Figure 10(b) differentiate RPGM mobility scenarios with different sizes of groups. The RPGM scenarios with smaller group sizes have higher SMD values, because of the simplicity to coordinate the mobility behavior of fewer nodes. RWP and GM have very low spatial mobility dependencies.

Since our approach can differentiate various mobility scenarios and get consistent results with their pre-defined parameters, we can claim that our approach is effective in characterizing the underlying temporal and spatial mobility patterns.

5.2.3 The Accuracy of Transient-State Mobility Prediction

Such accuracy of our approach is shown in Figure 11. When the node mobility is correlated in time, our mobility prediction shows higher accuracy, especially for the GM scenarios. For example, the average relative error can be limited below 15% when $\alpha \geq 0.5$. However, when the node mobility is less temporally dependent, the prediction error increases considerably. When α is set to 0.1 in the GM scenario, the prediction error can be up to 20%. Since in



(a) Average TMD (b) Average SMD

Figure 10: Characterization of temporal and spatial mobility features

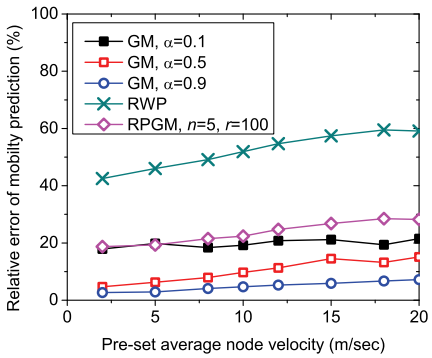


Figure 11: Accuracy of transient-state mobility prediction

RWP there is no underlying mobility pattern we can characterize, the mobility prediction has large error.

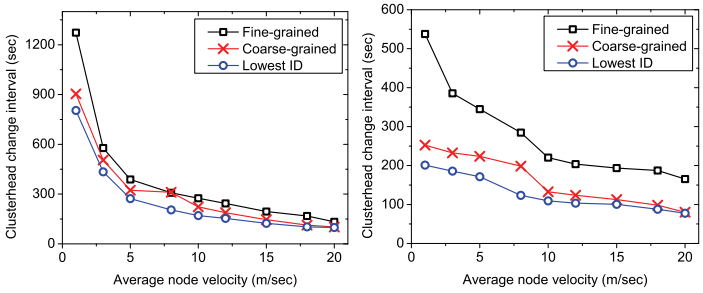
6. APPLICATION

To demonstrate the utilization of fine-grained mobility characterization, we apply the fine-grained node mobility knowledge to mobility-aware clustering, which is one of the most general applications in mobile ad-hoc networks (MANETs), and show the advantage of exploiting the steady-state node mobility features on improving the cluster stability. Our approach is not only compared with traditional clustering schemes, but also compared with methods considering coarse-grained mobility patterns [13]. We use the same network setting as in Section 5.2, i.e., there are 50 mobile nodes uniformly deployed in a $1000m \times 1000m$ square area, and the node communication range is set as 250m.

Here we consider 1-hop clustering, such that each node selects the node among its neighbors with the highest metric value to be its clusterhead, and clustering algorithms are differentiated by the metric used for clusterhead selection. In traditional clustering algorithms such as lowest-ID [6], each node is assigned a randomized ID as the clusterhead selection metric. As a result, a node may frequently disconnect with its randomly selected clusterhead due to their mobility, which leads to low cluster stability and high communication overhead for changing clusterhead. Instead, the chance of disconnection to the clusterhead can be reduced by selecting the clusterhead based on the average SMD of nodes with their neighbors, defined in Eq. (12). A node with higher average SMD is considered as the center of a group of nodes with similar mobility patterns, and therefore is much less likely to be disconnected from its cluster members if being selected as the clusterhead.

For comparison, we also implemented the AP-based mobility characterization approach in [13]. We randomly deploy 10 APs with a communication range of 500m in the area, and characterize the AP affiliation pattern of mobile nodes in the form of semi-Markov chains described in [13]. Being similar with Eq. (12), the SMD of nodes are also measured based on the KL distance between semi-Markov patterns, and are used for clusterhead selection.

The simulation results are shown in Figure 12(a) and 12(b), in which the RWP and RPGM mobility models are used, respectively. The cluster stability is measured by the average time that a node changes its clusterhead due to disconnection. In RWP mobility scenarios, clustering based on fine-grained node mobility knowledge shows a slight advantage of around 10%-20%, and such advantage is much more considerable in the RPGM scenarios due to the group mobility patterns. Note that the clustering stability based on fine-grained node mobility knowledge is considerably better than that



(a) RWP mobility model (b) RPGM mobility model

Figure 12: Performance of mobility-aware clustering

based on coarse-grained knowledge, especially for the RPGM scenarios. This indicates that our approach is more effective and accurate to characterize the underlying spatial mobility dependencies among nodes than the coarse-grained AP-based schemes.

In general, the methodology of utilizing fine-grained mobility knowledge described above can also be used in other mobile computing applications, and the key point is to consider the characterized mobility features as decision metrics. For example, in mobility-aware routing, two nodes whose mobility patterns are spatially correlated are less likely to be out of the communication range of each other, and thus are more suitable for building routes with high stability. In data forwarding in DTNs, a node currently carrying data was usually considered to have good mobility features and has higher chances to meet other nodes. If its mobility has high temporal dependency, it is more likely to keep such mobility features and is still a good choice to be selected as data relay in the future.

In order to be effectively utilized, the information about node mobility patterns has to be disseminated among mobile nodes periodically. Under the HMM mobility formulation, this dissemination has the complexity of $O(N^2)$ due to the state transition probability matrix of HMM with a size $N \times N$, where N is the number of states in a HMM. The transmission overhead in practice also depends on the node mobility itself, such that the node mobility knowledge needs to be updated more frequently if nodes move faster. This knowledge is easier to be updated in MANETs with persistent network connectivity, but may not be maintained up-to-date in DTNs where only intermittent connectivity among mobile nodes can be expected. Essentially, due to the long inter-contact time, mobile nodes need to exchange the information about their mobility patterns every time they contact with each other.

7. CONCLUSIONS

In this paper, we propose a novel fine-grained mobility characterization approach. Node mobility is formulated as a HMM with periodically re-estimated parameters based on node mobility observations, and node mobility is characterized based on such formulation. Node mobility is predicted at both the steady-state and transient-state time scales, and the inter-dependency features of node mobility are investigated in both temporal and spatial aspects. Future work includes the extension of our approach to multi-hop neighbors of a mobile node, and to match these behaviors with some existing mobility models via pattern recognition.

Acknowledgment

We would like to thank Oliver Waldhorst for his insightful comments and suggestions. We would also like to thank the anonymous reviewers for their helpful suggestions. This work was supported in part by the US National Science Foundation (NSF) under grant number CNS-0721479, and by Network Science CTA under grant W911NF-09-2-0053.

8. REFERENCES

[1] B. An and S. Papavassiliou. A mobility-based clustering approach to support mobility management and multicast routing in mobile ad-hoc wireless networks. *Int'l Journal of Network Management*, 11(6):387–395, 2001.

[2] M. Balazinska and P. Castro. Characterizing mobility and network usage in a corporate wireless local-area network. *Proc. MobiSys*, pages 303–316, 2003.

[3] L. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The Annals of Mathematical Statistics*, 41(1):164–171, 1970.

[4] G. Box, G. M. Jenkins, and G. C. Reinsel. *Time Series Analysis: Forecasting and Control*. Prentice-Hall, 3rd edition, 1994.

[5] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.

[6] A. Ephremides, J. E. Wieselthier, and D. J. Baker. A design concept for reliable mobile radio networks with frequency hopping signaling. In *Proc. IEEE*, volume 75, pages 56–73, 1987.

[7] M. Falkhausen, H. Reininger, and D. Wolf. Calculation of distance measures between hidden markov models. *Proc. Eurospeech*, pages 1487–1490, 1995.

[8] K. Fall. A delay-tolerant network architecture for challenged internets. *Proc. SIGCOMM*, pages 27–34, 2003.

[9] W. Gao, Q. Li, B. Zhao, and G. Cao. Multicasting in delay tolerant networks: a social network perspective. In *Proceedings of MobiHoc*, pages 299–308, 2009.

[10] T. Henderson, D. Kotz, and I. Abyzov. The changing usage of a mature campus-wide wireless network. *Computer Networks*, 52(14):2690–2712, 2008.

[11] X. Hong, M. Gerla, G. Pei, and C.-C. Chiang. A group mobility model for ad-hoc wireless networks. In *Proc. ACM MSWiM*, 1999.

[12] K. Langendoen and N. Reijers. Distributed localization in wireless sensor networks: a quantitative comparison. *Computer Networks*, 43(4):499–518, 2003.

[13] J.-K. Lee and J. Hou. Modeling steady-state and transient behaviors of user mobility: formulation, analysis and application. In *Proc. MobiHoc*, 2006.

[14] D. Lelescu, U. Kozat, R. Jain, and M. Balakrishnan. Model T++: an empirical joint space-time registration model. In *Proc. MobiHoc*, 2006.

[15] B. Liang and Z. Haas. Predictive distance-based mobility management for pcs networks. In *Proc. INFOCOM*, 1999.

[16] A. McDonald and T. F. Znati. A mobility-based framework for adaptive clustering in wireless ad hoc networks. *IEEE Journal on Selected Areas in Communications*, 17(8):1466–1487, 1999.

[17] S. Merugu, M. Ammar, and E. Zegura. Routing in space and time in networks with predictable mobility. *Technical Report, GIT-CC-04-7, Georgia Institute of Technology*, 2004.

[18] C. Neukirchen and G. Rigoll. Controlling the complexity of HMM systems by regularization. In *Proceedings of Conference on Advances in Neural Information Processing Systems II*, pages 735–743, 1999.

[19] A. J. Nicholson and B. D. Noble. Breadcrumbs: forecasting mobile connectivity. In *Proceedings of MobiCom*, pages 46–57, 2008.

[20] L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings IEEE*, 77(2):257–286, 1989.

[21] I. Rhee, M. Shin, S. Hong, K. Lee, and S. Chong. On the levy-walk nature of human mobility. In *Proc. INFOCOM*, pages 924–932, 2008.

[22] S. Siddiqi, G. Gordon, and A. Moore. Fast state discovery for HMM model selection and learning. In *Proc. Int'l Conference on Artificial Intelligence and Statistics*, 2007.

[23] L. Song, D. Kotz, and R. Jain. Evaluating location predictors with extensive Wi-Fi mobility data. *Proc. INFOCOM*, 2004.

[24] J. Zhao and G. Cao. VADD: Vehicle-Assisted Data Delivery in Vehicular Ad Hoc Networks. *Proc. INFOCOM*, 2006.

APPENDIX

A. KULLBACK-LEIBLER DISTANCE MEASURE BETWEEN HMMS

The KL distance is defined to evaluate the dissimilarity of two probabilistic distributions $p_1(\mathbf{X})$ and $p_2(\mathbf{X})$

$$d_{KL}(p_1, p_2) = \int_{\mathbf{X}} p_1(\mathbf{X}) log \frac{p_1(\mathbf{X})}{p_2(\mathbf{X})} d\mathbf{X},$$

and the KL distance between two HMMs $\mathcal{H}$ and $\mathcal{H}^*$ with the same state space is defined as

$$d_{KL}(\mathcal{H}, \mathcal{H}^*) = \frac{1}{T} \int_{\mathbf{X}} P(\mathbf{X}|\mathcal{H}) log \frac{P(\mathbf{X}|\mathcal{H})}{P(\mathbf{X}|\mathcal{H}^*)} d\mathbf{X}, \quad (13)$$

where T is the length of input sequence $\mathbf{X} = \mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_T$.

For a given $\mathbf{X}$ with a sufficient large T, if the HMM is ergodic, the KL distance can be approximated using the ergodic theorem [7]

$$\begin{aligned} d_{KL}(\mathcal{H}, \mathcal{H}^*) = & \sum_i \sum_j \frac{|\{t|s_t = i, s_{t+1} = j\}|}{T} log \frac{a_{ij}}{a^*_{ij}} \\ & + \sum_i \int_{\mathbf{x}} \frac{|\{t|s_t = i, \mathbf{X}_t = \mathbf{x}\}|}{T} log \frac{b_i(\mathbf{x})}{b^*_i(\mathbf{x})} d\mathbf{x}. \end{aligned} \quad (14)$$

If T is long enough, it is reasonable to assume that

$$\begin{cases} \frac{|\{t|s_t = i, s_{t+1} = j\}|}{T} \rightarrow a_{ij} P(s_i|\mathcal{H}) \\ \frac{|\{t|s_t = i, \mathbf{X}_t = \boldsymbol{x}\}|}{T} \rightarrow b_i(\boldsymbol{x}) P(s_i|\mathcal{H}), \end{cases}$$

and therefore Eq. (14) comes to be

$$\begin{aligned} d_{KL}(\mathcal{H}, \mathcal{H}^*) = & \sum_i \sum_j P(s_i|\mathcal{H}) a_{ij} log \frac{a_{ij}}{a^*_{ij}} \\ & + \sum_i P(s_i|\mathcal{H}) d_{KL}(b_i, b^*_i). \end{aligned}$$

If $b_i(\mathbf{x}) \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $b^*_i(\mathbf{x}) \sim N(\boldsymbol{\mu}^*_i, \boldsymbol{\Sigma}^*_i)$, we have

$$\begin{aligned} d_{KL}(b_i, b^*_i) = & \frac{1}{2} \Big(\log(\frac{\det \Sigma^*_i}{\det \Sigma_i}) + \mathbf{tr}(\Sigma^{*-1}_i \Sigma_i) \\ & + (\boldsymbol{\mu}^*_i - \boldsymbol{\mu}_i)^{\mathrm{T}} \Sigma^{*-1}_i (\boldsymbol{\mu}^*_i - \boldsymbol{\mu}_i) \Big) - 1, \end{aligned}$$

where $\mathbf{tr}$ gets the matrix trace, and $\det$ gets the matrix determinant.